GENETIC DETERMINANTS OF VACCINE RESPONSES

Daniel O'Connor

St. Catherine's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2014

Genetic determinants of vaccine responses

Daniel O'Connor, St. Catherine's College, D.Phil, Trinity 2014

Abstract

Vaccines have had a profound influence on human health with no other health intervention rivalling their impact on the morbidity and mortality associated with infectious disease. However, the magnitude and persistence of vaccine immunity varies considerably between individuals, a phenomenon that is not well understood. Recent studies have used contemporary technologies to correlate variations in the genome and transcriptome to complex phenotypic traits, and these approaches have started to provide fresh insight into the intrinsic factors determining the generation and persistence of vaccine-induced immunity.

This thesis aimed to describe the relationship between genomic and transcriptomic variations, and the immunogenicity of childhood immunisations. Candidate gene and genome-wide genotyping was conducted to evaluate the influence of genetic variants on vaccine-induced immunity following childhood immunisation. Furthermore, contemporary methodologies were used to assess non-coding and coding gene transcript profiles following vaccination, to further dissect the molecular systems involved in vaccine responses.

Key findings from this thesis include the description of the first genome-wide association studies into the persistence of immunity to three routine childhood immunisations: capsular group C meningococcal (MenC) conjugate vaccine, *Haemophilus influenzae* type b (Hib) conjugate vaccine and tetanus toxoid (TT) vaccine. Genome-wide genotyping was completed on over 2000 participants, with an additional 1000 participants genotyped at selected genetic markers.

Genome-wide significant associations ($p<5\times10^{-8}$) were described between single-nucleotide polymorphisms (SNPs) in two genes, *CNTN6* and *ENKUR*, and the persistence of serological immunity to MenC following immunisation of children 6-15 years of age. In addition, genome-wide significant associations were described between SNPs within an intergenic region of chromosome 10 and the persistence of TT-specific IgG concentrations following childhood immunisations. Furthermore, a number of variants in loci with putative involvement in the immune system such as *FOXP1*, the human leukocyte antigen locus and the lambda light chain immunoglobulin locus, were shown to have suggestive associations ($p<1\times10^{-5}$) with the persistence of vaccine-induced serological immunity. The fundamental challenge will be to describe functional mechanisms associated with these findings, and to translate these into innovative and pragmatic strategies to develop new and more effective vaccines.

Declaration

The research presented in this thesis is original; work not conducted by the author is clearly stated in the text. This thesis includes data generated from clinical trials lead by the Oxford Vaccine Group as well as third party organisations, where applicable appropriate permissions have been obtained. The author acknowledges the work of the clinical trials staff, at the Oxford Vaccine Group, for the collection of the sample cohorts. Furthermore, laboratory staff at the Oxford Vaccine Group supported the DNA extractions and immunological assays detailed in this thesis. Experiments that were outsourced to specialist facilities are clearly stated; these included microarray experiments that were conducted by the Genome Institute of Singapore, Wellcome Trust Centre for Human Genetics and Oxford Gene Technology. The author conducted the data analysis presented in this thesis; however, the author acknowledges those within the Hill group (Wellcome Trust Centre for Human Genetics, Oxford) and WTCHG core facility that shared their high-throughput genomics and transcriptomics data analysis scripts.

Acknowledgements

I would like to thank Professor Andrew J Pollard for his supportive supervision and guidance with this project. In addition, I thank everyone at the Oxford Vaccine Group involved in the study recruitment, enrolment, visits and sample processing, without whom the collection of the sample set assessed in this thesis would not have been possible.

I would like to thank everyone who contributed to the analysis described in this thesis. In particular my colleagues at the Wellcome Trust Centre for Human Genetics, Professor Adrian Hill, Dr Tara Mills, Dr Kate Elliot and Dr Anna Rautanen, for their advice.

I would like to thank Professor Martin L. Hibberd and Dr Chiea Chuen Khor at the Genome Institute of Singapore for facilitating the genome-wide genotyping described in this project, as well as contributing to the analytical strategy.

I would like to thank the European Union's Seventh Framework Programme under EC-GA no. 279185 (EUCLIDS) for funding much of the work described in this thesis. I would also like to thank the EUCLIDS' consortium for their intellectual contribution to this work, in particular Professor Michael Levin at Imperial College London.

Finally, I also am grateful to the study participants and their parents for consenting to this research.

Publications

Published

D. O'CONNOR, C.E. MOORE, M.D. SNAPE, T. JOHN, A.V.S. HILL & A.J. POLLARD. **Exonic single nucleotide polymorphisms within *TLR3* associated with infant responses to serogroup C meningococcal conjugate vaccine**. *Vaccine*, 32(27):3424-30, 2014.

D. O'CONNOR, J. TRÜCK, R. LAZARUS, E.A. CLUTTERBUCK, M. VOYSEY, K. JEFFERY & A.J. POLLARD. **The effect of chronic cytomegalovirus infection on pneumococcal vaccine responses**. *J Infect Dis*, 209(10):1635-41, 2014.

D. O'CONNOR, & A.J. POLLARD. **Characterizing vaccine responses using host genomic and transcriptomic analysis**. *Clin Infect Dis*, 57(6):860-9, 2013.

J. TRÜCK, **D. O'CONNOR**, T.C. DARTON, T.M. JOHN, M.D. SNAPE, & A.J. POLLARD. **Genetic material should be routinely collected in clinical vaccine trials–high consent rates can be achieved across all age groups**. *Vaccine*, 31(25):2744-8, 2013.

G. BLANCHARD-ROHNER, E. KIBWANA, H. PARKS, S. BRAR, **D. O'CONNOR**, K. FORD, T. JOHN, D.F. KELLY, M.D. SNAPE, P. DULL & A. J. POLLARD. **Seroprevalence and placental transmission of maternal antibodies specific for *Neisseria meningitidis* serogroups A, C, Y and W135 and influence of maternal antibodies on the immune response to a primary course of MenACWY-CRM$_{197}$ vaccine in the United Kingdom**. *Pediatr Infect Dis J*, 31(25):2744-8, 2013.

G. BLANCHARD-ROHNER, M.D. SNAPE, D.F. KELLY, **D. O'CONNOR**, T. JOHN, E.A. CLUTTERBUCK, B. OHENE-KENA, C.L. KLINGER, T. ODRLJIN & A.J. POLLARD. **The B-cell response to a primary and booster course of MenACWY-CRM197 vaccine administered at 2, 4 and 12 months of age**. *Vaccine*, 31(20):2441-8, 2013.

C.E. MOORE, B.J. HENNIG, K.P. PERRETT, J.C. HOE, S.J. LEE, H. FLETCHER, D. BROCKLEBANK, **D. O'CONNOR**, M.D. SNAPE, A.J. HALL, S. SEGAL, A.V.S. HILL & A.J. POLLARD. **Single nucleotide polymorphisms in the Toll-like receptor 3 and CD44 genes are associated with persistence of vaccine-induced immunity to the serogroup C meningococcal conjugate vaccine.** *Clin Vaccine Immunol*, 19(3):295-303, 2012.

Submitted or in preparation

R. KANDASAMY, M. VOYSEY, F. MCQUAID, K. DE NIE, R. RYAN, O. ORR, U. UHLIG, C. SANDE, **D. O'CONNOR**, X. DE LAURIE, A. RESTREPO & A.J. POLLARD. **Non-specific immunological effects of selected routine childhood immunisations: a systematic review.** *Article submitted.*

D. O'CONNOR, C.C. KHOR, M.D. SNAPE, D.F. KELLY, E.A. CLUTTERBUCK, C.E. MOORE, M. LEVIN, A.V.S. HILL, M.L. HIBBERD & A.J. POLLARD. **Genome-wide association study of childhood immunity to routine immunisations**. *In preparation.*

D. O'CONNOR, C. WHEELER, M. MCANDREW, J. ANSON, M.D.SNAPE & A.J. POLLARD. **Serum microRNA following childhood immunisation with pandemic influenza vaccine**. *In preparation.*

D. O'CONNOR, M.D.SNAPE & A.J. POLLARD. **Transcriptomic profiles following a single dose of trivalent inactivated seasonal influenza vaccine**. *In preparation.*

D. O'CONNOR, E.A. CLUTTERBUCK, J. TRÜCK, C.J. BLOHMKE, M.N. RA-

MASAMY, D.F. KELLY & A.J. POLLARD. **Transcriptomic responses following immunisation with quadrivalent meningococcal vaccine**. *In preparation.*

Contents

List of Figures

List of Tables

Glossary

ACE Additive genetic, common and random environment model

Ad5 Adenovirus serotype 5

Alum Aluminium hydroxide or aluminium sulfate

ANOVA One-way analysis of variance

APC Allophycocyanin

ARMS-PCR Amplification-refractory mutation system polymerase chain reaction

AS01 Monophosphoryl lipid A, QS21, and liposomes

ASCs Antibody-secreting cells

BCG Bacillus Calmette-Guérin

BCR B cell receptor

BF Bayes factor

BHI Brain heart infusion

BHI-HS Brain heart infusion horse serum

bp Base pair

BTMs Blood transcription modules

CI Confidence interval

CIES Carrier-induced epitope suppression

CNV Copy number variation

CpG non-methylated CpG oligonucleotide

CRM_{197} or CRM Cross reactive material-197 (mutant diphtheria toxoid)

D-PBS Dulbecco's phosphate buffered solution

dbSNP NCBI single nucleotide polymorphism database

ddH20 Deionised water

DEGs Differentially expressed genes

DNA Deoxyribonucleic acid

DT Diphtheria toxoid

DTaP/Hib Diphtheria-tetanus-acellular pertussis/ *Haemophilus influenzae* type b

DTaP/IPV/Hib Diphtheria-tetanus-acellular pertussis/ *Haemophilus influenzae* type b/inactivated poliovirus

DTP Diphtheria-tetanus-pertussis

DTwP/Hib Diphtheria-tetanus-whole cell pertussis/*Haemophilus influenzae* type b

DZ Dizygotic

EDTA Ethylenediaminetetraacetic acid

ELISA Enzyme-linked immunosorbent assay

ELISPot Enzyme-linked immunosorbent spot

EM Expectation-maximisation

EPI Expanded Program on Immunisation

eQTL expression quantitative trait loci

FC Fold-change

FDR False-discovery rate

FHA Filamentous haemagglutinin adhesin

FIM Fimbriae

FITC Fluorescein isothiocyanate

FPCA Functional principal component analysis

FWER Family-wise error rate

GAPT GRB2-binding adaptor protein, transmembrane

GC Germinal centre

GCSAM Germinal centre-associated, signalling and motility

GMC Geometric mean concentration

GMT Geometric mean titre

GRB2 Growth factor receptor-bound protein 2

GWA Genome-wide association

GWAS Genome-wide association study

GWASs Genome-wide association studies

HA Haemagglutinin

HAI Haemagglutination inhibition

HAV Hepatitis A vaccine

HBsAg Hepatitis B surface antigen

HBV Hepatitis B vaccine

HHE Hypotonic hyporesponsive episodes

Hib *Haemophilus influenzae* type b

Hib-MenC Combination conjugate vaccine containing capsular group C meningococcal and *Haemophilus influenzae* type b polysaccharides

Hib-TT *Haemophilus influenzae* type b tetanus toxoid conjugate

HIV Human immunodeficiency virus

HLA Human leukocyte antigen

HMM Hidden Markov model

hSBA human complement serum bactericidal assay

HSE Herpes simplex encephalitis

HWE Hardy-Weinberg Equilibrium

IBD Identity-by-descent

IFN-γ Interferon-gamma

IgG Immunoglobulin G

IGHV Immunoglobulin heavy chain variable

IGKV Immunoglobulin kappa variable

IGL Immunoglobulin lambda locus

IL-4 Interleukin-4

IPA Ingenuity® Pathway Analysis

IPV Inactivated polio vaccine

IQR Interquartile range

kb kilobases

LAIV Live attenuated influenza vaccine

LD Linkage disequilibrium

lincRNAs Long intergenic non-coding RNAs

LLQ Lower limit of quantitation

LPS Lipopolysaccharide

LRR Leucine-rich repeats

LVS *Francisella tularensis* vaccine strain

M1 Matrix protein 1

MAF Minor allele frequency

MAP Molecule activity predictor

MCMC Markov chain Monte Carlo

MDS Multi-dimensional scaling

MEM Minimum essential medium

MenACWY Quadrivalent capsular groups A, C, W and Y meningococcal conjugate vaccine

MenC Capsular group C meningococcus

MenCC Monovalent capsular group C meningococcal conjugate vaccine

MHC Major histocompatibility complex

mHSA methylated human serum albumin

miRNAs MicroRNAs

ml Millilitres

MMR Measles-mumps-rubella

MPL Monophosphoryl lipid-A

mRNA Messenger RNA

MZ Monozygotic

NBBS New born bovine serum

NCBI National Center for Biotechnology Information

NGS Next-generation sequencing

NK Natural killer

NP Nucleoprotein

NS1 Nonstructural protein 1

OPV Oral polio vaccine

PBMCs Peripheral blood mononuclear cells

PBS Phosphate buffered solution

PC Principal component

PCA Principal component analysis

PCR Polymerase chain reaction

PCV7 7-valent pneumococcal conjugate vaccine

PE Phycoerythrin

PEG Polyethylene glycol

PerCP Peridinin chlorophyll protein

PHA Phytohemagglutinin

Poly(I:C) Polyinosinic:polycytidylic acid

Polyphen Polymorphism phenotyping

PPD Purified protein derivative

PPMCC Pearson product-moment correlation coefficients

PRP Polyribosyl ribitol phosphate

PT Pertussis toxin

Q-Q Quantile-quantile

QC Quality control

R0 R0 culture medium (500 ml RPMI, 5 ml penicillin-streptomycin, 5 ml L-glutamine)

R10 R10 culture medium (450 ml RPMI (Sigma), 50 ml heat inactivated new born bovine serum, 5 ml penicillin-streptomycin, 5 ml L-glutamine)

R848 Resiquimod

RIN RNA Integrity Number

RISC RNA-induced silencing complex

RIVM Rijksinstituut voor Volksgezondheid en Milieu

RNA Ribonucleic acid

RNA-seq RNA sequencing

rpm Revolutions per minute

rRNA Ribosomal RNA

rSBA rabbit complement serum bactericidal assay

RSN robust spline normalisation

SAM Significance Analysis of Microarrays

SBA Serum bactericidal assay

SEB Staphylococcal enterotoxin B

SFUs Spot-forming units

SNPs Single-nucleotide polymorphisms

TAS Trait-associated single-nucleotide polymorphisms

TBS Tris buffered solution

TIV Trivalent inactivated influenza vaccine

TLR Toll-like receptors

TMB 3,3′,5,5′-tetramethylbenzidine

TSB Tryptic soy broth

TT Tetanus toxoid

UK United Kingdom

US United States

UTR Untranslated region

VEU Vaccine Evaluation Unit

VSN Variance stabilisation and normalisation

WGS Whole-genome sequencing

YF-17D live-attenuated yellow fever 17D

Chapter 1

Introduction

Since Edward Jenner's epochal demonstration in 1796 that inoculation with the vaccinia virus induced resistance to the related and highly virulent variola virus, billions of doses of vaccines have been administered. Vaccines have had a profound influence on health particularly in childhood, averting an estimated 2.5 million child deaths per year (1). However, inter-individual variation in vaccine responses remains both perplexing and problematic, with a considerable proportion not achieving the putative level of protection following licensed vaccines; for example, 10-15% of adults fail to respond to three doses of hepatitis B vaccine (HBV) (2).

Heterogeneous vaccine responses in infancy are even more noteworthy as immune responses are generally poorer in the young and immunity can wane rapidly (3). A number of factors including age, sex, nutrition, ultraviolet light exposure, smoking and infectious disease have been shown to influence vaccine responses (4). Importantly, a large proportion of the variation seen in responses is thought to be attributable to factors present in the vaccinees' genome. Recent advances in the 'omics', particularly genomics and transcriptomics present highly amenable avenues to explore and dissect the mechanisms involved in complex biological systems, such as vaccine-induced immunity (Figure 1.1).

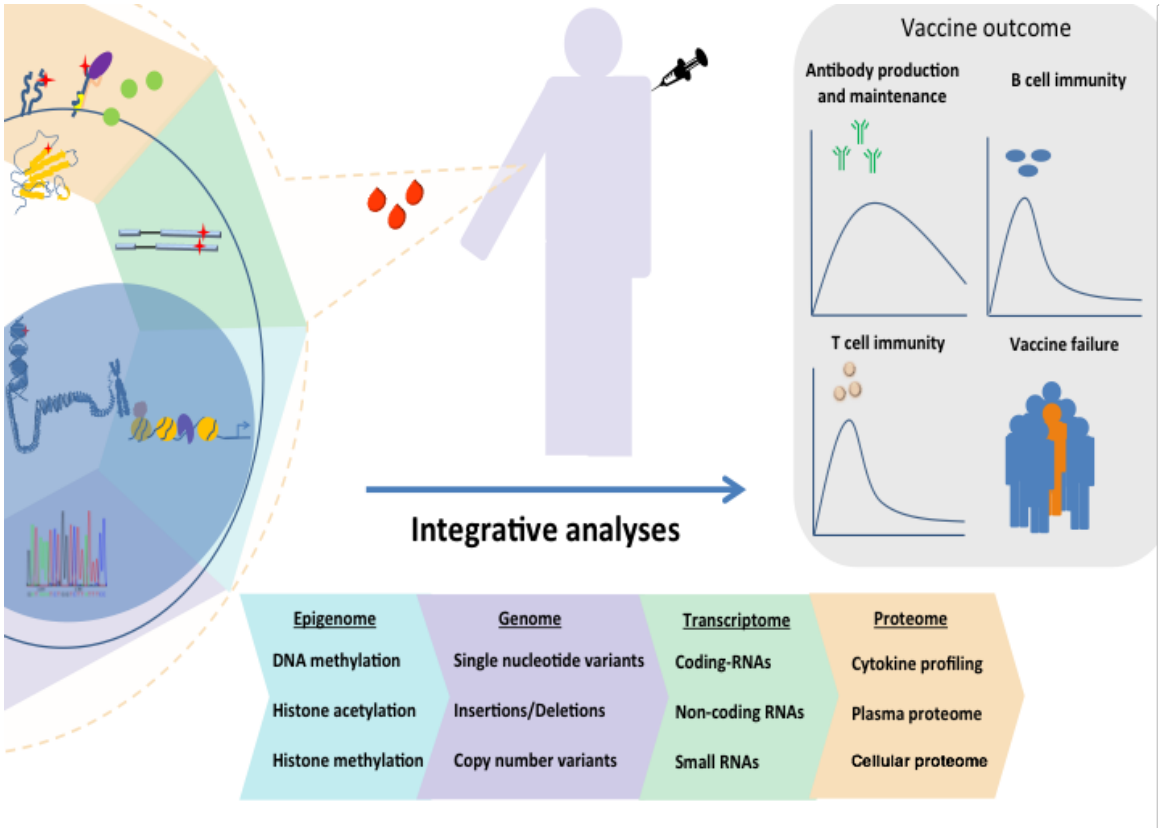


Figure 1.1: Overview of molecular approaches to describe and predict vaccine responses. This figure was reproduced from O'Connor and Pollard 2013 (5), with licence from Oxford University Press.

1.1 Heritability of vaccine responses

1.1.1 Heritability calculations

A number of studies in twins have evaluated the role of genetic variation on vaccine responses, showing these to be 36-85% heritable (Table 1.1 and Figure 1.2) (6, 7, 8, 9, 10). Heritability can either be measured as the broad-sense heritability H^2 or narrow-sense h^2. Broad-sense heritability measures the full contribution of genes on a trait, effectively giving the predictive value of genotype on phenotype. Whereas, narrow-sense heritability measures the additive contribution of genes on a trait. In twin studies heritability can be assessed by comparing phenotypic concordance between monozygotic (MZ) and dizygotic (DZ) twins, as both twin-pairs are likely to be exposed to similar environmental factors but share 100% and 50% of

their segregating DNA sequence, respectively. To date, the principal method of estimating vaccine response heritability has been to use Mx path analysis, under the additive genetic, common and random environment (ACE) model, as displayed in appendix Figure 7.1 (11, 12). The Mx statistical package uses maximum likelihood for estimation of structural equation models, producing estimates of narrow-sense heritability along with confidence intervals under the assumption of normally distributed estimators (12).

1.1.2 Narrow-sense heritability calculation caveats

Human genetic association studies of common phenotypic traits have frequently accounted for a modest proportion of the total genetic contribution inferred from heritability studies (13, 14). While much of this 'missing heritability' may be due to undiscovered variants, heritability estimates may also be inflated (15). Total narrow-sense heritability calculations assume no epistasis among loci. Characterising genetic interactions in humans is challenging; however, non-linear molecular interactions are commonplace in biology and it is not justifiable to assume genetic effects will be purely additive. The absence of gene-gene and gene-environment interactions in heritability calculations may result in overestimation of narrow-sense heritability (15). The genetic architecture of extensive allele sharing between close relatives makes disentangling genetic interactions problematic. Alternative methods of calculating heritability in nearly unrelated individuals in a population have been suggested to overcome the problem of genetic interactions (15, 16).

1.1.3 Vaccine-induced immunity heritability estimates

An early study of Danish adult twins following a single dose of 23-valent pneumococcal polysaccharide vaccine examined the influence of genetic factors on serotype-specific IgG and IgG subclass responses (9). This study found greater correlation in the serotype-specific total IgG 4 weeks post-vaccination in monozygotic than dizygotic twin pairs for all eight pneumococcal serotypes assessed (1, 2, 4, 7F, 9N, 14, 19F and 23F). Interestingly, the genetic contribution to responses to pneumococcal polysaccharides appeared to be serotype dependent, with only IgG

responses to serotypes 1, 2, 4 and 19F reaching statistical significance (9). Of the six serotypes (1, 2, 4, 7F, 9N and 14) assessed greater IgG2 correlation in monozygotic compared to dizygotic twins was observed for serotypes 1, 9N and 14. The disparate genetic contribution to antibody responses to differing pneumococcal polysaccharides may be genetically determined; however, it may also be affected by variable environmental exposure to this widely carried organism in this cohort aged 21-65 years (17, 18).

Another twin study in German adults assessed the heritability of vaccine responses 4 weeks after three doses of the combined hepatitis B vaccine (HBV) and hepatitis A vaccine (HAV) (6). Antibody concentrations to hepatitis B virus surface antigen (HBsAg) were found to be 61% (95% CI 0.41-0.81%) heritable, and antibody concentrations to HAV were found to be 36% heritable but with a lower 95% confidence interval (CI) estimate crossing zero (6). Analysis of the small subset of human leukocyte antigen (HLA)-identical dizygotic twin pairs estimated 25% of the phenotypic variance in anti-HBsAg antibodies was explained by the *HLA-DRB1* locus, although the 95% CIs for intraclass correlations overlapped (6). Furthermore, a cross-sectional study in monozygotic and dizygotic children who had received at least one dose of measles-mumps-rubella (MMR) vaccine found specific-antibody levels to all three of these vaccine antigens to be heritable, but with estimates ranging from 36% for mumps to 89% for measles (19).

A prospective study of a birth cohort of 207 Gambian twin-pairs assessed the heritability of both humoral and cellular responses to a number of vaccine antigens following the Expanded Program on Immunisation (EPI) schedule (7, 8). This study found significantly greater correlation within monozygotic twin pairs for all specific-antibody responses, with heritability ranging from 44-77% for hepatitis B, polio, tetanus, diphtheria and *Haemophilus influenzae* type b (Hib) (7, 8). Furthermore, total serum IgG levels were also found to be 78% (95% CI 67-85%) heritable at 5 months of age (7). This study also assessed cellular immune responses, finding IFN-γ release following peripheral blood mononuclear cell (PBMC) stimulation with mycobacterial purified protein derivative (PPD), killed Mycobacterium tuberculosis, pertactin and filamentous haemagglutinin, as well as IL-13 release following PPD, mycobacterial 65-kDa

heat shock protein, pertussis toxin (PT) and tetanus toxoid (TT), to be heritable (7). Analysis of *HLA-DRB1* identical dizygotic twin pairs estimated 15% of the phenotypic variance in anti-HBsAg antibodies to be explained by this locus. However, the lower 95% confidence interval for phenotypic variance attributable to *HLA-DRB1* overlapped zero for antibody responses for all vaccines assessed (7). In contrast, the estimated contribution of the *HLA-DRB1* locus on IFN-γ (interferon-gamma) responses to PPD did not overlap zero, with this single locus accounting for 76% (61%-85%) of the phenotypic variance (7).

A further infant vaccination study investigated anti-HBsAg antibody responses in Han and Ugyur Chinese twins (10). This study found greater correlation in anti-HBsAg concentrations in monozygotic than dizygotic twins at 1 year of age following immunisation at 0, 1 and 6 months of age with HBV (10). Furthermore, when HBV responses were coded as ordinal variables for IgG concentrations $\geq$100, 10-100 and $<$10mIU ml^{-1}, genetic factors were seen to have a dominant role, accounting for 91% of the phenotypic variance (10).

Table 1.1: Summary of key twin studies investigating the heritability of vaccine responses

Reference	Vaccine	Population	Response	Heritability	CI
(6)	Inactivated HAV and recombinant HBsAg	192 monozygotic and 190 dizygotic German adult twins aged 18-65 years	Anti-HAV antibody Anti-HBsAg antibody	36% 61%	(0-73%) (41-81%)
(9)	Pneumococcal vaccination	48 monozygotic and 36 dizygotic Caucasian adult twins aged 21-65 years	IgG and IgG2 serotype specific antibodies	Variable	
(8)	Hib conjugate vaccine	86 monozygotic and 294 dizygotic Gambian 5 months old infant twins	Anti-PRP IgG	51%	(32-66%)
(7)	Recombinant HBsAg	96 monozygotic and 318 dizygotic Gambian 5 months old infant twins	Anti-HBsAg antibody	77%	(63-85%)
	Oral polio vaccine		Anti-poliovirus antibody	60%	(43-73%)
	DTP		Anti-TT antibody Anti- DT antibody	44% 49%	(16-70%) (17-77%)
	BCG, 0.05 ml		Interferon-γ responses to PPD	41%	(10-71%)
(10)	Recombinant HBsAg vaccine	164 monozygotic twins and 180 dizygotic twins Healthy 1 year old Chinese infants	Anti-HBsAg antibody	91%	(76-97%)
(19)	Live attenuated MMR vaccine	90 monozygotic and 110 dizygotic Caucasian twins (196/200) age 2-18 years	Anti-measles antibody Anti-mumps antibody Anti-rubella antibody	89% 39% 46%	(52%-NR)* (2%-NR)* (5% -NR)*

NR=not reported, *= One-sided confidence interval, MMR= Measles-Mumps-Rubella, HBsAg =Hepatitis B surface antigen
HAV= Hepatitis A virus, TT= Tetanus toxoid, DT= Diphtheria toxoid, DTP = Diphtheria-Tetanus-Pertussis, BCG =Bacille Calmette-Guérin (BCG)

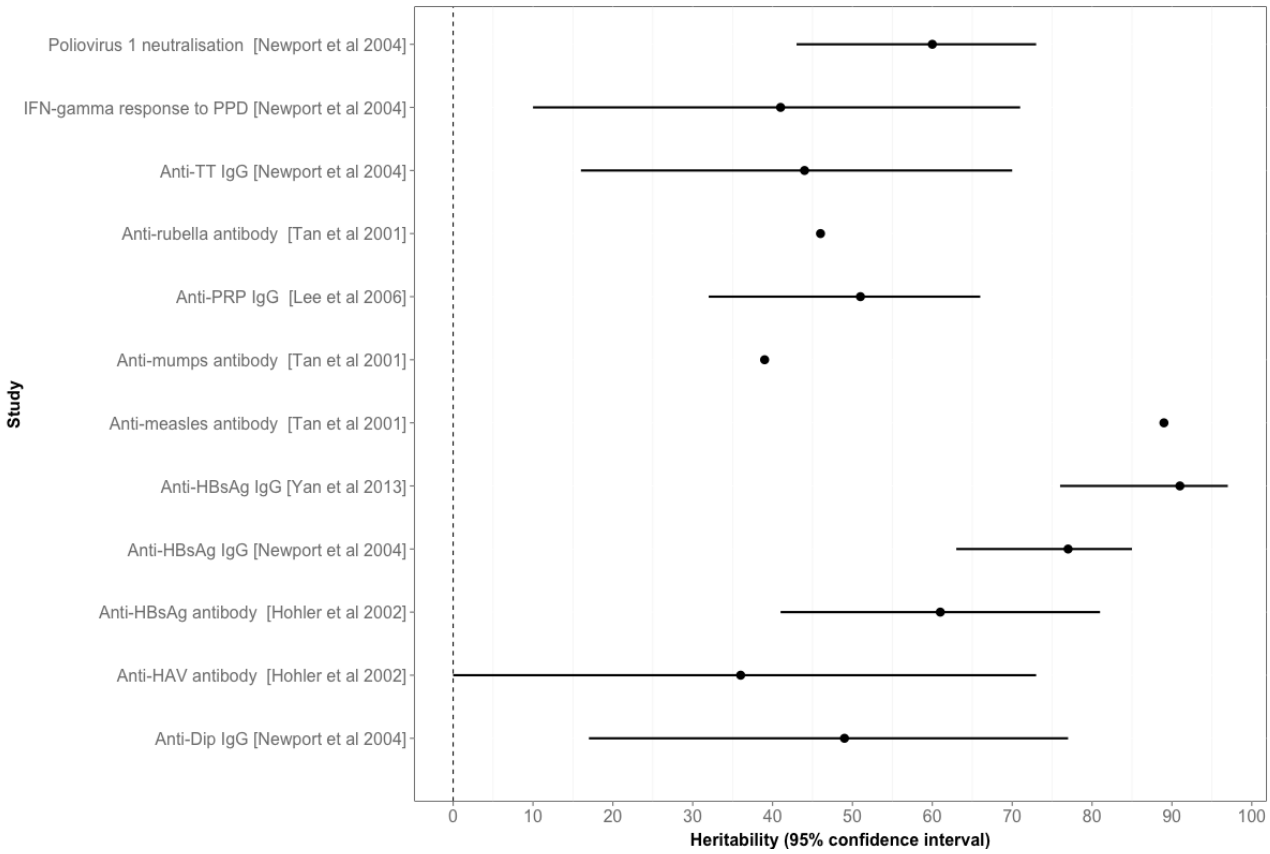


Figure 1.2: Forest plot of the heritability of vaccine-induced immune responses, determined by twin studies. Spots signify the point estimate of heritability and the lines extend to the 95% confidence intervals. PPD= purified protein derivative (*Mycobacterium tuberculosis*),TT= tetanus toxoid, PRP= polyribosylribitol phosphate *Haemophilus influenzae* type B, HBsAg= hepatitis B surface antigen, HAV= hepatitis A virus, Dip= Diphtheria toxoid, IgG= Immunoglobulin G.

1.1.3.1 Variable heritability estimates

The twin data reviewed in the previous section consistently demonstrated a relationship between heritable factors and vaccine responses. Interestingly, the estimates on vaccine heritability varied considerably between studies, and even between vaccines given concomitantly in the same study population (6, 7, 10). Although data suggest that the majority of current vaccines work by evoking antibody mediated protection, the molecular pathways evoked following vaccination may be highly disparate (20, p. 19) (21). Therefore, it is logical to assume some genetic factors may have relatively generic influence on vaccine responses, and others will be vaccine-specific leading to disparate additive genetic contributions. Interestingly, antibody responses to protein toxoids, TT and DT, correlate and show similar heritability, potentially indicating

shared genetic determinants (7). However, antibody responses to HBV and OPV show no correlation and disparate heritability, suggesting distinct genetic factors are involved (7).

In addition to vaccine-specific determinants, participant age, ethnicity, nutrition status as well as the time and type of the immunogenicity measure have all been highlighted as important factors that influence heritability estimates (10). It has also been suggested that the heritability of vaccine responses may be greatest early in life, with greater heritability exhibited in specific antibody to HBV in infants (77% and 91%) compared to adults (61%) (6, 7, 10). The effect of age on observed heritability may be particularly apparent in the case of vaccines that contain antigens that individuals might frequently encounter, such as influenza and pneumococci.

The heritability measures of vaccine responses are also likely to be context and population specific, complicating interpretation and comparison of reported heritability measures. Ethnicity may contribute to the difference seen in the heritability of HBV responses seen in Gambian (77%) and Chinese infants (91%) (6, 10). In addition, experimental factors that influence overall phenotypic variance, such as measurement assay variability, will also affect heritability estimates.

1.2 Genetic studies of immunological responses to vaccines

Despite the evident heritability of vaccine responses the particular genetic determinants involved in vaccine responses remain largely uncharacterised. Several candidate gene studies and more recently genome-wide association studies (GWASs) have investigated the effect of genetic variation on immunological and physiological responses to vaccination.

1.2.1 Genome-wide association studies

A genome-wide association study (GWAS) of 3644 healthy Indonesians following HBV, found 47 single nucleotide polymorphisms (SNP) that associated ($P \leq 1.1 \times 10^{-7}$) with vaccine responses (22). In this study participants were grouped into low responders (0-99.9 IU/l anti-

HBsAg), medium responders (100-999.9 IU/l anti-HBsAg) and high responders (anti-HBsAg $\geq$1000 IU/l) and analysed using ordered logistic regression. Stepwise conditional logistic regression of the significantly associated SNPs demonstrated that these represented three independent signals across the HLA region. The strongest of these independently segregating signals was rs3135363 (p=6.53×10^{-22}) a SNP tagging a haplotype block that includes HLA-DR. The other two index SNPs were in the 3′ untranslated region (UTR) of *HLA-DPB1* and in a HLA class III gene-rich region (22). Intriguingly, the *HLA-DPB1* tagging SNP rs9277535 has previously been associated with chronicity of hepatitis B infection in several Asian subpopulations (23, 24). Clearance of hepatitis B virus coincides with the production of anti-HBsAg antibodies; however, chronically infected individuals fail to produce these antibodies despite detectable HBsAg antigen in their circulation. The antigenic component of HBV is HBsAg; therefore, the concordance between the HBV response and hepatitis B viral clearance genetic association studies suggests individuals are predisposed to chronic infection by a genetically determined inability to produce anti-HBsAg antibodies, and that part of this genetic susceptibility lies proximal to the *HLA-DPB1* locus (22, 23).

A further GWAS assessed 77 Chinese Han individuals who failed to respond to a booster dose of HBV (anti-HBsAg IgG <10 mIU/ml after six doses) with 108 primary HBV high-responders (anti-HBsAg IgG $\geq$1000 mIU/ml after three doses) (25). This study identified a SNP (rs477515) located $\sim$12 kb upstream of *HLA-DRB1* that had an odds ratio of 3.59 and surpassed the level of genome-wide significant association (p=4.81×10^{-8}) (25). The authors of this study were then able to replicate these findings in a larger cohort of 420 primary HBV non-responders (anti-HBsAg IgG <10 mIU/ml after three doses) and 1336 primary high-responders, with rs477515 reaching a combined p-value=2.63×10^{-19} (25). This study also replicated rs3135363, but demonstrated the association between rs477515 to be largely independent of rs3135363 (22, 25). Direct sequencing of the *HLA-DRB1* region, in the cohort of 77 HBV booster non-responders versus the 108 primary HBV responder, associated the HLA-DRB1*0701 allele with non-responders with an odds ratio of 3.69 (p=1.5×10^{-6}) (25).

The two GWASs of HBV responses have shown the peak association to be in the *HLA-DRB1*

locus, which is consistent with the twin data suggesting 15-25% of the heritability of HBV IgG responses could be accounted for by the *HLA-DRB1* locus (6, 7, 22, 25). However, twin data also suggested that the majority of the genetic determinants of HBV antibody responses are outside the *HLA-DRB1* locus, and to date these factors are largely unknown (6, 7, 22, 25).

Another published vaccine GWAS assessed vaccinia-specific neutralisation titres in a multi-ethnic cohort of 1014 healthy adults 1 month to 4 years following pustule forming smallpox vaccination (26). This study included three distinct ethnic groups: Caucasians (n=580), African-Americans (n=217) and Hispanics-Americans (n=217). Association analysis was performed separately for each of the ethnic subgroups, describing a single SNP within the promoter region of a pseudogene that surpassed the authors proposed level of significance ($p< 5 \times 10^{-7}$) in the Caucasian cohort (26). Analysis of the other ethnic groups found eight and five SNPs that associated with vaccinia neutralising antibody in the African-Americans and Hispanics-Americans, respectively (26). In addition to humoral responses, the GWAS data generated in this study were also compared to cellular immunity following smallpox vaccination, measured by $CD8^+$ and total IFN-γ enzyme-linked immunospot (ELISpot) assays, as well as IFN-γ enzyme-linked immunosorbent assay (ELISA) on supernatant from PBMCs stimulated with inactivated vaccinia virus (27). Twenty-five SNPs were found to be associated with vaccinia-specific $CD8^+$ IFN-γ ELISpots, 10 SNPs were associated with total IFN-γ ELISpots and 11 SNPs were associated with secreted IFN-γ (27).

A recent GWAS explored rubella-specific cytokine responses in a cohort of Caucasian adolescents following two childhood doses of MMR vaccine (28). This study found seven SNPs with genome-wide significant associations ($p<5\times10^{-8}$) with rubella-specific IL-6 cytokine secretion (28). Six of these SNPs were within an intergenic region on the short arm of chromosome 9 ($\sim$1.5 Mb from a coding gene). The other genome-wide significantly associated SNP was within *PTPRD* a gene that encodes protein tyrosine phosphatase receptor type D, a member of a family of proteins involved in processes such as cell growth and differentiation (28). This SNP had a minor allelic frequency (MAF) of 0.01 and represented the only SNP within this gene to have a p-value $<1\times10^{-5}$. While these parameters may raise questions about marker reliability, the

SNP MAF did match that of the 1000 Genomes Project Europeans. This study also assessed rubella-specific IFN-γ secretion, but did not find any SNPs to be significantly associated with this measure (28). This study assessed a modest number of participants (n=897) given the need for stringent correction for multiple testing and it may be the case that many SNPs that did not surpass the level of genome-wide significance may influence rubella-specific cytokine responses. The authors stated that replication studies were being undertaken to validate their findings.

GWASs provide a genome-wide assessment of many hundreds of thousands of genetic variants and have the benefit of not requiring exhaustive prior knowledge of the genes involved in a particular trait. However, it is important to consider the implications of this expansive approach, particularly the inherent need for stringent statistical corrections. Using Bonferroni correction, for approximately a million independent variants in the European-American genome, gives a threshold for statistical significance of $p<5\times10^{-8}$ (29). This conservative threshold makes it necessary to work with large sample sizes to provide statistical power and may result in the exclusion of many markers with small to moderate effect sizes.

1.2.2 Candidate gene studies

Until recently, the candidate gene approach had been the mainstay of investigations into the genetic determinants of vaccine-induced immunity. Selecting genes with putative roles in immunological responses to vaccines has the clear disadvantage of dependency on a priori knowledge. Conversely, candidate gene studies typically require less burdensome statistical corrections, as they perform fewer tests, than genome-wide approaches; therefore, can be more powerful in the moderate sample size setting. A variety of genes have been investigated but for simplicity these can be separated into three broad themes: antigen processing and presentation, innate recognition and cellular signalling.

1.2.2.1 Antigen processing and presentation

HLA molecules facilitate the presentation of peptide antigens to T-cell receptors. HLA genes are highly polymorphic, introducing a diverse ability to bind and present peptide within the population. Comparison of vaccine responses within *HLA-DRB1* matched and *HLA-DRB1* disparate DZ twins have estimated that ∼40% of the genetic variation influencing specific-antibody responses to HBV may be attributable to this locus alone (6). Several candidate gene studies have found HLA polymorphisms to be associated with immunological responses to a number of vaccines including hepatitis B, measles, rubella, influenza and anthrax vaccines (30, 31, 32, 33, 34).

HLA polymorphisms represent a considerable portion the catalogue of genetic variants associated with vaccine responses contributed by candidate gene studies (35). However, as HLA polymorphisms account for a large proportion of the candidates assessed in these studies these may be enriched for false positives, making replication particularly pertinent. The associations between HLA-DPA1*0201, HLA-B*2705 and HLA-DPB1*0401 alleles and rubella antibody responses have been replicated, as has the association between HLA-DQA1*0201 and measles vaccine responses (31, 32). A meta-analysis of published associations between *HLA-DRB1* and *HLA-DQB1* alleles, and antibody responses to hepatitis B vaccine, found several alleles to be consistently associated with responses (36). Another meta-analysis found four HLA alleles (DRB1*07, DQA1*02:01, DQB1*02:01, and DQB1*03:03) were associated with lesser, and two HLA alleles (DRB1*13 and DRB1*13:01) were associated with greater, antibody responses to MMR, HBV and influenza vaccines (35). While the meta-analysis of published data may not completely negate false-positives due to publication bias, alleles with consistent directionality of association are more likely to represent true positives.

1.2.2.2 Innate recognition receptors

Innate recognition receptors instigate immunological responses to antigens based upon the recognition of conserved microbial motifs and/or intrinsic host cell 'danger' signals. Polymorphisms in genes in a number of the innate receptor pathways including toll-like receptors (TLRs)

have been associated with variable vaccine responses (37, 38, 39, 40, 41, 42, 43, 44, 45).

In particular, SNPs within the *TLR3* gene locus have been associated with measles-specific antibody persistence following MMR in adolescents (37). Furthermore, *TLR3* SNPs have been associated with the persistence of capsular group C meningococcal (MenC)-specific antibody concentrations in children and adolescents following immunisation with MenC-conjugate vaccine (38). Interestingly, the *TLR3* Leu412Phe polymorphism (rs3775291) implicated in both measles and MenC vaccine responses has been associated with differential IFN-α/β and TLR3 protein expression, following stimulation of PBMCs with the synthetic TLR3 agonist polyI:C (polyinosinic:polycytidylic acid) (46). In addition, a SNP within *TIRAP* (rs1893352) an essential adaptor molecule that binds to both TLR2 and TLR4 has been associated with Hib vaccine failure, with the recessive homozygote genotype being 7.6 times more common in the bacteraemic study cohort than healthy controls ($p=1.3\times10^{-6}$) (47).

1.2.2.3 Cellular signalling

Cytokines are small intercellular signalling proteins with pleiotropic roles in humoral and cellular immunity. Polymorphisms in numerous cytokine genes have been associated with immune responses to vaccines, with some cytokine genes being implicated in responses to a variety of vaccines (41, 48, 49, 50, 51, 52). For example, polymorphisms within *IL-10* have been associated with concentrations of specific-antibody to diphtheria toxoid (DT), HAV, HBV and pneumococcal serotype 6B capsular polysaccharide (48, 49, 53). The minor allele in the promoter region on *IL-10* (rs1800896) has been associated with higher antibody responses to both DT and HAV vaccines; another *IL-10* promoter SNP (rs1554286) has been associated with Hib vaccine failure (47, 48, 49). SNPs in the IL-12 receptor subunits *IL-12Rβ1* and *IL-12Rβ2* have also been associated with antibody responses to measles, mumps as well as pneumococcal 9V and 19F serotypes (41, 48, 54). These data imply a generic and non-redundant role for some cytokine polymorphisms in influencing antibody responses to several vaccine antigens.

1.2.3 Lymphocyte receptor polymorphisms

B and T cell receptors are unique in that they are encoded by multiple genomic loci that, within the relevant cell type, undergo somatic recombination and in the case of the former somatic hypermutation (55, p. 146-151). Lymphocyte receptor loci require special attention in genetic association analysis as their repeated structure makes them susceptible to frequent insertion and deletion events; consequently, the complex structural genetic variation at these loci are difficult to capture with contemporary high-throughput genotyping techniques (56). Despite the vast repertoire of receptor gene segments available, B cells with a highly restricted antibody gene segment usage are directed to Hib and pneumococcal capsular polysaccharides following vaccination (57, 58, 59). Similarly, highly restricted antigen-specific T cell repertoires have also been observed (60, 61). These data have led to speculation that the availability of lymphocyte receptor gene segments, or allelic variants within them, could influence an individual's clonal repertoire following vaccination and ultimately the robustness of vaccine response.

Non-synonymous mutations occur disproportionately in the complementarity determining regions of some immunoglobulin heavy chain variable (IGHV) genes, suggesting these may be under positive selection (62). For example, the IGHV3-23*03 allele has been shown to more effectively bind Hib polysaccharide compared to the more frequently found IGHV3-23*01 allele (63). Additionally, the frequency of the IGHV3-23*03 allele and copy-number of *IGHV3-23* varies across populations, which may support underlying selective pressure (64). Polymorphisms in the germline immunoglobulin sequence have also been associated with immunoglobulin gene expression; for example, the IGKV2D-29*02 allele that results in a single nucleotide variant in a recombination signal of the *IGKV2-29* (immunoglobulin kappa variable 2-29) gene, dramatically decreases the expression of this gene (65, 66). Interestingly, *IGKV2-29* is expressed following vaccination with Hib polysaccharide vaccine and the IGKV2D-29*02 allele has been associated with increased susceptibility of the Navajos population to Hib disease, although a direct causal link has yet to be demonstrated (65, 66, 67). There are few data exploring the relationship between germline polymorphisms and copy-number variations on lymphocyte receptor repertoire expression (68, 69, 70). However, germline polymorphisms and copy-number varia-

tions in lymphocyte receptor genes may be important in repertoire expression and ultimately the robustness of vaccine response and disease susceptibility.

1.2.4 Replication and functional validation

There are an accumulating number of genetic variants being implicated in vaccine responses; however, few have been assigned any functional consequence. The major 'A' allele at rs9277535 in the 3′ UTR of *HLA-DPB1* associated with clearance of hepatitis B virus and HBV antibody responses, has recently been linked with the expression of *HLA-DPB1* mRNA in both liver and lymphoblastoid cell lines (22, 23, 71). These data suggest that variants within the 3′ UTR of *HLA-DPB1* influence responses to hepatitis B virus and HBV via *HLA-DPB1* expression. There is a strong case for a functional role of the non-synonymous *MAL/TIRAP* variant Ser180Leu associated with Hib vaccine failure, this SNP lies within the proposed functional interface with MyD88 (Myeloid differentiation primary response gene 88) and may result in a partial loss of function (47, 72).

More studies providing functional validation for associated genetic loci are needed in order to enable robust inferences to be formed about the genetic determinants of vaccine responses. In the absence of functional validation, replication in an independent population is critical to corroborate 'true' signals and eliminate false positives. Replication is particularly important as associations seldom surpass conservative thresholds for statistical significance, due to the need for corrections for multiple testing and inadequate sample sizes. 'Tiered' or 'two-stage' study designs, with an initial 'discovery' cohort genotyped broadly that is used to produce a smaller subset of variants to be followed-up in a 'replication' cohort have been shown to be both powerful and cost-effective (73). For example, the GWAS of HBV responses adopted this tiered approach genotyping 455,508 SNPs in a discovery set of 1683 individuals and pursuing a subset of 1706 SNPs in a replication cohort of 1931 individuals, resulting in 47 significant SNPs after corrections for multiple testing (22).

1.3 Genetic studies of adverse responses to vaccines

Vaccine associated adverse events are important in the public acceptance of vaccines. Although severe reactions are very rare, both the public and the media often have an inflated perceived risk (74). On the other hand, mild reactions such as pain, swelling and fevers are relatively common and may also be important in the public perception of vaccine safety. The determinants of vaccine associated adverse reactions or reactogenicity are not well understood, and there are currently no published twin data explicitly examining the heritability of these responses. However, higher rates of febrile events following measles vaccination have been reported in some ethnic populations, implying genetics factors have a part to play in susceptibility to vaccine associated adverse reactions (75).

To date, the published studies of the genetic determinants of adverse vaccine responses have focused on responses to smallpox vaccine (76, 77). The first of which assessed the relationship between polymorphisms in 19 candidate genes, selected based upon putative involvement in the control of poxviruses, and fever following immunisation with smallpox vaccine (77). A total of 357 SNPs, which represented all the SNPs in the candidate genes known at the time, were assessed in 346 healthy adult volunteers, 176 of whom were vaccinia-naïve. Individuals with fever ($\geq$ 37.7°C) were compared to individuals without any recorded fever. To account for multiple testing a permutation test was used to determine the statistical cutoff ($p<0.01$). Haplotypes were constructed using an expectation-maximisation (EM) algorithm and tested simultaneously in the regression model to identify haplotypes independently predictive of fever. Eight haplotypes in four different genes *IL1A*, *IL1B*, *IL1R1*, and *IL18* were associated with risk of fever following smallpox vaccination (77). Furthermore, a single *IL4* haplotype was associated with fever following smallpox vaccination in the vaccinia-naïve subset, whereas a *IRF1* haplotype was associated with fever in those who had previously received a smallpox vaccine (77). Notwithstanding, candidate genes were selected for putative involvement in viral control, the fact that haplotypes within the genes encoding IL1 a cytokine that is pyrogenic in humans at subnanomolar concentrations and IL18 a pro-inflammatory cytokine, were associated with fever biologically corroborates the statistical associations (77, 78).

A further study assessed for genetic susceptibility to adverse reactions following smallpox vaccination in healthy vaccinia-naïve adult volunteers (76). This study assessed 1442 SNPs across 386 candidate genes in two cohorts of 85 and 46 vaccinees. Within the two groups subjects with systemic adverse events such as fever, rash, or regional lymphadenopathy, were compared to individuals who did not experience these reactions. SNPs with a $p<0.05$ in both these groups were considered statistically significant (76). Despite the modest sample size the study found a SNP in *MTHFR* (methylenetetrahydrofolate reductase) and two SNPs in *IRF1* (interferon regulatory factor-1) were associated with systemic adverse events following administration of smallpox vaccine, with odds-ratios in the region of threefold (76). Interestingly, the associated *MTHFR* SNP was non-synonymous and has been demonstrated to be thermolabile, affecting both the quantity and activity of the enzyme, which itself has previously been linked with adverse reactions to other pharmacological agents (79, 80). *IRF1* is a member of the interferon regulatory transcription factor (IRF) family and acts as transcription activator of genes induced by interferons α, β, and γ; haplotypes within this gene have now been associated with fever following smallpox vaccination in two studies (76, 77).

The safety profile and target population of the smallpox vaccine is not typical of the majority of vaccines currently in use worldwide, as it has a higher rate of serious adverse events and it is only recommended for particular populations, such as some laboratory staff and specific workers at identifiable risk (81, 82). However, fever following immunisations is not an infrequent event; for example, in a multi-centre European trial of 1885 infants 36% recorded a fever of $\geq 38°C$ following their second dose of routine infant immunisations (at 3, 4 or 5 months of age), which increased to 59% when the newly licensed capsular group B meningococcal vaccine (Bexsero®, Novartis) was given concomitantly (83). Further studies are needed to investigate the genetic determinants of reactogenic responses following vaccination.

1.4 Vaccine response transcriptomics

Contemporary transcriptomic analysis has revolutionised the way in which complex biological systems can be interrogated. Recently, transcriptomic analysis has been utilised to describe and predict responses to licensed vaccines, as well as in early vaccine development and pre-clinical research (Table 1.2 and Table 1.3) (84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95).

1.4.1 The use of transcriptomic analysis in clinical trials

1.4.1.1 In vivo genome-wide transcriptomic analysis of responses to licensed vaccines

Transcriptomic analysis can be used in vaccine clinical trials to, 1) characterise gene expression changes evoked following vaccination, 2) correlate transcriptional changes to measures of immunogenicity and/or tolerability, and 3) described novel biomarkers of vaccine immunogenicity or efficacy. A seminal study in this area assessed the PBMC transcriptome of healthy adult volunteers: 0, 1, 3, 7 and 21 days after vaccination with live-attenuated yellow fever 17D (YF-17D) vaccine (84). This study initially assessed 15 YF-17D vaccinees, describing 97 differentially expressed genes (false-discovery rate [FDR] <0.05) in the week following YF-17D vaccine (84). Sixty-five of these differentially expressed genes (DEGs) were consistently modulated in an independent replication cohort of 10 volunteers, with gene ontology analysis showing enrichment of genes related to immune response, innate immune response and response to virus (84). This study was also able to show a correlation between the expression of particular genes and specific $CD8^+$ T cell and antibody responses; most intriguingly demonstrating that the expression of these genes could predict immunogenicity in an independent replication cohort with a high degree of accuracy (84). Of particular note a gene set containing the *TNFRSF17* gene, which encodes a receptor for the B cell growth factor BLyS-BAFF, was highly predictive of antibody responses (84).

At a similar time another study was published, which looked at the transcriptome of peripheral whole blood samples collected in PAXgene® RNA stabilisation tubes: 0, 3, 7, 10, 14,

28, and 60 days after vaccination of healthy adults with YF-17D vaccine (95). This study initial assessed 15 YF-17D vaccinees finding 594 DEGs ($p<0.05$ and fold-change [FC] >1.3), the majority of which occurred within the first 10 days. Many of the gene induced in the first 10 days post-vaccination were interferon-induced genes and several B-cell related genes were significantly differentially expressed 14 days following YF-17D vaccine, including *TNFRSF17* (95). Further analysis of 43 of the DEGs, selected to represent transcriptional nodes, in an additional cohort of 23 YF-17D vaccinees using quantitative PCR provided corroboration for the majority of these DEGs (95).

A further study analysed the transcriptome of PBMCs following either smallpox vaccine or YF-17D vaccine in healthy adult volunteers (89). This study assessed PBMCs collected 0, 2-4, 5-7 and 50-60 days following vaccination of 24 individuals with smallpox vaccine; and 0 and 4-7 days following vaccination of 20 individuals with YF-17D vaccine (89). Analysis of PBMCs 4-7 days following vaccination found 475 and 615 DEGs (FDR <0.05) in the smallpox and YF-17D vaccine groups, respectively (89). Eighty genes were differentially expressed following both vaccines, characterised by interferon-induced genes as well as genes involved in proteolysis and antigen presentation, implying generic roles in response to live-attenuated viral vaccines (89). Vaccinia specific responses were characterised by repression of B and T cell associated genes and the induction of monocyte/macrophage associated genes, and YF-17D specific responses were associated with downregulation of genes associated with protein biosynthesis (89).

In addition to detecting systemic transcriptional responses to replicating live-attenuated vaccines, studies have also been able to detect transcriptional responses to non-replicating vaccines. A study of the PBMCs in the week following influenza vaccination highlighted disparities between gene profiles evoked following the live-attenuated influenza vaccine (LAIV) given intranasally and the intramuscularly administered trivalent inactivated influenza vaccine (TIV) (85). Although a core set including genes involved in inflammatory and antimicrobial responses was altered similarly, the majority of the DEGs (FDR<0.05 and FC >1.41 in at least 60%) were vaccine-specific. Recipients of the LAIV differentially expressed several interferon-related genes 3 days after vaccination, consistent with observations following YF-17D, poten-

tially reflecting the replication competency of these live-attenuated viral vaccines (84, 85). Furthermore, the analysis of sorted cells implied that DEGs following TIV predominantly originated from the myeloid dendritic cells (mDC) and B cells, whereas plasmacytoid dendritic cells (pDC) contributed the greatest DEGs following LAIV (85). These data imply that the two vaccine strategies, which differ both in composition and route of administration, evoke very divergent molecular responses that are propagated by different cell types. Disparities between these vaccines have previously been described at the serological and epidemiological level, TIV evokes greater haemagglutination inhibition (HAI) serum antibody titres but LAIV has been shown to be more effective at preventing childhood culture-confirmed influenza disease (96).

A study of the transcriptome of whole blood samples from 85 healthy child aged 12-35 months 7-10 days after receiving a single dose of LAIV or TIV vaccination described 265 DEGs, with only 20 being induced by both vaccines (97). This study found that the live-attenuated vaccine evoked five times as many DEGs as the inactivated vaccine, including a more pronounced type 1 IFN-inducible gene signature (97). A further study assessed the transcriptomic profile in whole blood samples from healthy children aged 6 months to 14 years: 1, 7 and 30 days after LAIV or TIV (98). This study described differences in transcriptional kinetics between these two vaccination strategies. While both vaccines perturbed IFN-signalling genes, the kinetics of these responses were disparate with TIV recipients inducing IFN genes early at day 1 and the LAIV inducing this gene module at day 7 post-vaccination (98). These findings match those of the previously mentioned studies that detailed IFN related gene responses in LAIV but not TIV recipients a week after childhood or adult influenza vaccination (85, 97, 98). These data modify the interpretation of differences between the transcriptional responses evoked by these two vaccines and highlight the added value of kinetics data in evaluating transcriptional responses. Interestingly, in the aforementioned study the IFN gene signature induced by LAIV at day 7 appeared to be age dependent and was only seen in those <5 years of age and not those 5-14 years of age (98). Nevertheless, for both vaccines IFN gene upregulation was positively correlated with post-vaccination H3N2 antibody titres (98). Consistent with previous data this study also found *TNFRSF17* to be among the 10 most upregulated genes

7 days following TIV (98). In contrast to previous adult and infant data comparing TIV and LAIV, this study identified more DEGs 7 days following TIV than LAIV (85, 97, 98). However, there were a number of analytical differences between these studies, such as the threshold for DEG classification, and these findings may also be confounded by uneven vaccine group sizes.

A recent study found the expression of a module of plasmablast related genes, which included *TNFRSF17*, correlated with HAI titres and serotype-specific antibodies concentrations following TIV and 23-valent pneumococcal vaccine, respectively (87). A further study found the expression of *TNFSF13B* encoding the B cell growth factor BLyS-BAFF, a ligand for *TNFRSF17*, correlated with post-TIV HAI titres (88). Transcriptomic studies have also alluded to novel roles for genes with previously unappreciated contributions to immune regulation. For example, *CAMK4* expression was shown to inversely correlate with HAI titres following TIV, a finding that was functionally substantiated by the demonstration that *CAMK4*-deficient mice have enhanced TIV antibody responses (85).

A further study explored more detailed kinetics of transcriptomic changes induced following TIV by collecting samples daily, for the first 10 days following vaccination (99). This study performed RNA-sequencing on PBMCs and B cells collected from five participants following a single dose of TIV (99). Per individual time point analysis on the B cell transcriptomics data found vastly differing numbers of DEGs (FDR <0.05) between participants (2-5580 DEGs), of which 742 were common between at least 60% of participants (99). Functional principal component analysis (FPCA) found 90% of transcriptome variance could be accounted for by a single eigenfunction for three individuals and by two eigenfunctions in the remaining two participants (99). As the former reported history of influenza vaccination, the authors suggested the differences were due to 'predominantly recall' compared to 'predominantly de novo' responses (99). Of note the temporal pattern and peak of these eigenfunctions (when displayed graphically) varied, reflecting time-varying responses and highlighting the loss of information that results from inevitable restrictions on sampling frequency.

A further study looked at whole blood transcriptional biomarkers of immunological responses to TIV in 30 young (20-30 years of age) and 59 older (60 to >89 years of age) generally healthy

adults (100). This study assessed baseline parameters including gene expression modules, serum proteins, immune cell subsets, participant demography, cytomegalovirus and Epstein-Barr virus serological status (100). In total 241 variables were included in feature selection and prediction of post-vaccination HAI titres, with cross-validation being completed using the elastic net algorithm. A model with all these features, nine of which were considered relevant, better correlated with post-vaccination HAI titres than a model based only upon participant age ($p= 1.2 \times 10^{-5}$). Only one of the features in this first model displayed a positive relationship with post-vaccination HAI titres, this was a baseline gene module enriched for genes involved in apoptosis (100). In addition to age and pre-existing HAI titres, a gene module composed of gene involved in the regulation of B-cell proliferation was shown to have a negative relationship with post-vaccination HAI titres (100).

A recent study evaluated transcriptional changes following immunisation with either the conjugated or plain-polysaccharide quadrivalent (capsular groups A, C, W and Y) meningococcal vaccine (86). This study examined the transcriptome of PBMCs from healthy adult volunteers 0, 3 and 7 days after vaccination, with either the conjugated (n=17) or polysaccharide (n=13) meningococcal vaccines. In contrast to what has previously been observed following YF-17D, LAIV and TIV, few genes were differentially expressed 3 days after vaccination with either of the meningococcal vaccines, implying these vaccines induce subtle early systemic transcriptional responses (84, 85, 86). Interestingly, 1150 differentially expressed genes ($p<0.001$) were seen 7 days after the meningococcal conjugate vaccination compared to just 105 genes surpassing this threshold 7 days after the polysaccharide vaccine (86). In addition to differing in the dose of total meningococcal polysaccharide content and route of administration, the meningococcal glycoconjugate vaccine is believed to stimulate a T-cell dependent immune responses via the diphtheria toxoid carrier protein (55, p. 646-647). Moreover, as this study found a far greater proportion of plasmablasts induced by the meningococcal conjugate vaccine secreted antibodies to the carrier diphtheria protein than the meningococcal polysaccharides, the day 7 transcription profile may be largely evoked by the carrier protein. Despite the transcriptional differences seen 7 days post-vaccination, the expression of the 'dendritic surface signature' module was

found to significantly correlate with antibody responses to meningococcal polysaccharides following both of the conjugate and polysaccharide meningococcal vaccine (86).

Table 1.2: Summary of clinical studies investigating transcriptomic responses to vaccines

Reference	Vaccine	Summary
(89)	Vaccinia or Yellow-fever (YF) (Live-attenuated)	◦ 44 healthy adults (24 receiving vaccinia and 20 receiving YF vaccine) ◦ Sampling prior to and 4-7 days following vaccination. ◦ Direct transcriptomic analysis of PBMCs ◦ Agilent human 1 cDNA or Oligo Microarray Kit microarray
(84)	Yellow fever (YF) (Live-attenuated)	◦ 25 healthy adults ◦ Sampling 0, 1, 3, 7, 21 days post vaccination ◦ Direct transcriptomic analysis of PBMCs ◦ Affymetrix Human U133 Plus 2.0 Array
(88)	Influenza (Inactivated)	◦ 119 healthy adult males ◦ Sampling 0, 1, 3 and 14 days post-vaccination ◦ Direct transcriptomic analysis of whole blood ◦ Illumina Human HT-12v3 Expression BeadChips
(85)	Influenza (Live-attenuated or inactivated)	◦ 56 healthy adults (28 receiving each vaccine) ◦ Sampling 0, 3 and 7 days post-vaccination ◦ Direct transcriptomic analysis of PBMCs and sorted cell populations ◦ Affymetrix Human U133 Plus 2.0 Array
(90)	Malaria (Protein subunit)	◦ 39 healthy adults ◦ Sampling 0, 1, 3 and 14 days post-vaccination, as well as 5 days post-challenge ◦ Direct transcriptomic analysis of PBMCs ◦ Affymetrix Human U133 A 2.0 or U133 Plus 2.0 microarray
(91)	*Francisella tularensis* (Live-attenuated)	◦ 5 healthy young adults ◦ Sampling 0, 0.75, 2, 8 and 14 days following vaccination ◦ Direct transcriptomic analysis of PBMCs ◦ Affymetrix Human Genome U133+2 array
(95)	Yellow fever (YF) (Live-attenuated)	◦ 40 healthy adults ◦ Sampling 0, 3, 7, 10, 14, 28, and 60 days following vaccination ◦ Direct transcriptomic analysis of whole blood ◦ Illumina Human RefSeq-8 BeadChips v2
(97)	Influenza (Live-attenuated or inactivated)	◦ 85 children (43 received LAIV and 42 TIV) ◦ Sampling 0 and 7-10 days following vaccination ◦ Direct transcriptomic analysis of whole blood ◦ Affymetrix Human U133 A 2.0 and U133 Plus 2.0 arrays
(100)	Influenza (Inactivated)	◦ 89 adults (30 young and 59 older subjects) ◦ Sampling 0 and 28 days following vaccination. ◦ Direct transcriptomic analysis of PBMCs ◦ Illumina HT-12v3 microarray
(86)	Quadrivalent meningococcal vaccine (conjugate or plain polysaccharide)	◦ 30 healthy adults (17 received conjuage and 13 received vaccine) ◦ Sampling 0, 3 and 7 days following vaccination ◦ Direct transcriptomic analysis of PBMCs ◦ Affymetrix Human U133 A 2.0 and U133 Plus 2.0 arrays
(101)	Ad5 containing HIV-1 Gag/Pol/Nef	◦ 10 healthy young adults (7 Ad5 seronegative and 3 Ad5 seropositive) ◦ Sampling 0, 0.25, 1, 3 and 7 days following vaccination ◦ Direct transcriptomic analysis of PBMCs ◦ Affymetrix Human Exon ST 1.0 microarray
(87)	Influenza or pneumococcal (Inactivated and plain-polysaccharide)	◦ 46 healthy adults ◦ Sampling 0, 1.5, 3, 6, 9, 12, 15, 24, 36, and 48 hours as well as 3, 7, 10, 14 and 21 days after vaccination ◦ Direct transcriptomic analysis of whole blood ◦ Illumina HT-12v3 microarray

Continued on next page

Reference	Vaccine	Summary
(99)	Influenza (Inactivated)	◦ 14 healthy adults ◦ Sampling 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 days following vaccination ◦ Direct transcriptomic analysis of PBMCs ◦ TruSeq RNA kits Illumina Genome Analyzer IIx
(98)	Influenza (Live-attenuated or inactivated)	◦ 32 healthy children (18 received LAIV and 14 TIV) ◦ Sampling 0, 1, 7 and 30 days following vaccination ◦ Direct transcriptomic analysis of whole blood ◦ Illumina WG-6 V4 beadchips

1.4.1.2 In vivo genome-wide transcriptomic analysis of responses to unlicensed vaccines

Gene expression analysis has also been undertaken in studies investigating vaccines for which the correlates of protection have yet to be determined, such as malaria, HIV and tularaemia. One of the first studies to apply a high-throughput approach to assess genes modulated following vaccination investigated the transcriptome of PBMCs collected from five healthy adults: 0, 0.75, 2, 8 and 14 days following inoculation with live *Francisella tularensis* vaccine strain (LVS) (91). LVS induces consistent gene expression signatures, enabling accurate separation of pre-vaccination, early (0.75 and 2 days) and late post-vaccination (8 and 14 days) samples based upon the first two principal components of the gene expression data (91). The greatest number of differentially expression genes ($p \leq 0.01$ with FC ≥ 2) were observed 2 days after LVS (91). This study also explored the kinetics of gene expression following LVS. Firstly, probes with an F-score of 10 (i.e. $p < 0.001$) using a one-way analysis of variance (ANOVA) across the five time point and a Pearson's correlation coefficient >0.95 with one of the five primary principal trends of the data set were selected. Probes fulfilling these criteria were then clustered into groups using k-means Euclidean distance clustering (k=10), plots of these clusters revealed four prevalent patterns 'down early' , 'up early', 'sustained up' and 'up late' (91). The 'down early' and 'up early' patterns showed the greatest enrichment for genes annotated as immune related and included regulators of proliferation, pro-inflammatory cytokine response, mast cells, dendritic cell, neutrophils, macrophages, granulocytes, and monocytes (91). The 'up late' pattern was characterised by biosynthetic and metabolic genes; among the genes within

the 'sustained up' pattern were regulators of endocytosis/phagocytosis, granule exocytosis, chemotaxis, and inflammatory cytokines (91). Little is known about the determinants required for protection against tularemia in humans; however, this study described a gene-level view of the host response to the attenuated pathogen.

Another study of gene expression following an investigational vaccine collected PBMCs from 39 healthy adult volunteers: 0, 1, 3 and 14 days after a three-dose schedule of the novel AS01 (monophosphoryl lipid A, QS21, and liposomes) adjuvanted RTS,S malaria (*Plasmodium falciparum*) vaccine (90). Sixty-three inflammatory response genes were differentially expressed (FDR <0.01) in PMBCs 1 day after RTS,S but returned to pre-vaccination levels by day 3 (90). Similar kinetics were observed with 79 genes in the protein kinase cascade family and 107 genes in the apoptosis family that were differentially expressed at day 1 and resolved to pre-vaccination expression by day 3 after vaccination (90). When vaccinees were subsequently challenged with *P. falciparum* 6219 genes were upregulated and 5670 downregulated, compared to just 461 upregulated and 626 downregulated genes in vaccine naïve individuals (90). Despite the increased breadth of responses in vaccine recipients, two-thirds were not protected; however, a 393 gene classifier set 5 days after challenge accurately predicted parasitaemia following infectious challenge (90).

A further study assessed the PBMCs gene expression profile 0, 0.25, 1, 3 and 7 days after vaccination with the investigational Merck HIV-1 Gag/Pol/Nef adenovirus serotype 5 (MRKAd5/HIV) vaccine in HIV-1 uninfected healthy adults (101). This study profiled the transcriptome of seven adenovirus serotype 5 (Ad5) seronegative (Ad5 neutralising antibody titre $\leq$18) individuals finding the peak of DEGs (FDR <0.05) to occur 1 day after MRKAd5/HIV, with 1026 upregulated and 1048 downregulated genes (101). Three days after MRKAd5/HIV a smaller subset of 132 upregulated and 11 downregulated genes were observed, with no statistically significant DEGs detected 7 days post-vaccination (101). Modular analysis found the peak of responses to occur 1 day after MRKAd5/HIV (24 enriched modules), narrowing by day 2 (two enriched modules) and apparently returning to baseline by day 7 (101). The modules induced by MRKAd5/HIV included modules annotated as involved in immune response and

inflammation, such as 'Inflammation', 'Interferon response', 'Myeloid lineage'; and modules such as 'Lymphoid lineage', 'T cells' and 'Cytotoxicity' were suppressed (101). Analysis of three Ad5 seropositive (Ad5 neutralising antibody titre $\geq$200) individuals appeared to show attenuated transcriptional responses following MRKAd5/HIV, consistent with the reduced immunogenicity seen in vaccinees with neutralising antibody titres $\geq$200 (101, 102). Interestingly, these data contradict the hypothesis that enhanced systemic innate immune activation due to pre-existing Ad5 immunity may be the mechanism behind the reported increased risk of HIV-1 acquisition in Ad5 seropositive individuals following MRKAd5/HIV (103). Analysis of transcriptional data from all 10 MRKAd5/HIV vaccinees found a significant relationships between HIV-1 gag protein specific $CD8^+$ T cell responses and transcriptional responses measured at 3 days post-vaccination; 88 and 121 genes were positively and negatively associated with $CD8^+$ T cell responses, respectively (101).

1.4.2 In vitro genome-wide transcriptomic analysis

There are a number of studies where direct assessment of peripheral blood transcriptome following vaccination may not have been possible, optimal or safe. In such cases, in vitro stimulation of peripheral blood with vaccine antigens has typically been used as a surrogate of in vivo responses. For example, a study assessed the transcriptional response of PBMCs from healthy adult volunteers following a primary dose of smallpox vaccine, re-stimulated in vitro with inactivated vaccinia for 18 hours compared to unstimulated PBMCs (104). This study collected PBMCs from 1076 participants 16.2 months (interquartile range [IQR], 9.0-34.6 months) after receiving smallpox vaccine and conducted transcriptomic analysis on a cohort of 200 vaccinees at the 'extreme' of the spectrum of immune responses (i.e. high and low responders) (104). Two thousand one hundred and three DEGs (FDR <0.05) were observed upon culture with inactivated vaccinia virus (104). Pathway analysis using the MetaCore^{TM} software identified four enriched pathways (FDR <0.05) upon viral stimulation: cytokine production by T-helper 17 cells, immune response (antiviral actions of interferon), immune response (innate immune response to RNA viral infection) and immune response (PGE2 signalling in immune response) (104). Com-

parison of the mRNA profiles of smallpox vaccinees who produced the highest and lowest titres of neutralising antibodies did not find any genes with statistically significant relationships with neutralising antibodies (104). However, pathway analysis of high and low antibody responders did find two pathways significantly associated with neutralising antibodies: immune response (antiviral actions of interferons) and immune response (interferon α/β signalling pathway).

A further study assessed the transcriptome of PBMCs from five healthy 10 month old infants, immunised with Bacille Calmette-Guérin (BCG) at birth, following in vitro culture with BCG, purified protein derivative (PPD) or medium alone for 12 hours (105). This study found 411 and 291 DEGs (FDR <0.01) in BCG and PPD stimulated PBMCs compared to medium alone, respectively (105). Two hundred and fifty-eight genes were differentially expression following either BCG or PPD stimulation. Gene ontology analysis found the DEGs induced by both BCG and PPD to be enriched for genes related to responses to external stimuli and stress (105). Furthermore, downregulation of cell adhesion molecules was found to be specific to BCG.

Another study assessed the transcriptome of PBMCs from healthy children and young adults (11-19 years of age), following two-doses of MMR vaccine, re-stimulated in vitro with live rubella virus for 48 hours compared to unstimulated PBMCs (106). This study collected PBMCs from 738 participants a median of 5.8 years after receiving MMR. Twenty-five vaccinees representing the 'extremes' of rubella antibody responses, 12 high responders (median 138 IU/mL rubella-specific IgG) and 13 low responders (median 10 IU/mL rubella-specific IgG) were selected for whole transcriptome mRNA sequencing (mRNA-seq) (106). To minimise confounders all selected study subjects were female non-Hispanic Caucasians. A total 1080 DEGs (FDR<$1.00E^{-14}$) were seen in rubella stimulated PBMCs compared to media alone (106). Five genes showed statistically significant differences (FDR<0.05) in responses between high responders and low responders (106). Ingenuity® pathway analysis (IPA) of the genes with the most statistical evidence (p<0.05) of differential expression between high responders and low responders identified two immune pathways: antigen presentation pathway and complement system pathway (106).

In an attempt to dissociate regulation of gene expression from changes in peripheral blood

cellular constituents, a further study compared in vivo and in vitro transcriptional responses to MRKAd5/HIV and MRKAd5 empty vector, respectively (101). Interestingly, 8 of 13 modules induced by MRKAd5/HIV in vivo at 24 hours were also induced after 24 hour in vitro culture with MRKAd5 empty vector (101). Furthermore, some of modules that were discordant between in vivo and in vitro responses were associated with myeloid, lymphoid, T cell, B cell lineages; supporting the hypothesis that discrepancies may originate in part from the absence of cell trafficking in the in vitro setting (101).

1.4.3 The use of transcriptomic analyses in pre-clinical studies

Preclinical vaccine studies have also made use of transcriptomic approaches, particularly to gain insight into the workings of vaccine adjuvants (92, 93, 94). A study investigating the effects of intramuscular injection with MF59 oil-in-water emulsion, non-methylated CpG oligonucleotide (CpG) or alum (aluminium hydroxide) in mice found that these adjuvants modulated a core set of 168 genes in muscle specimens. This core set was enriched for genes involved in cytokine-cytokine receptor interaction, host-pathogen interaction and defence immunity protein activity (92). MF59 induced nearly three times as many genes as the other adjuvants, preferentially inducing the transcription of genes involved in complement activation, prostaglandin synthesis, IL-1 signalling and leukocyte transendothelial migration (92). The authors of this study then went on to assess gene expression following influenza subunit and TT vaccines, adjuvanted with either MF59, alum, CpG, resiquimod (R848) or Pam3CSK4 (93). MF59 was the strongest modulator of transcriptional events at the injection site, particularly genes involved in leukocyte transendothelial migration and cytokine binding implying this adjuvant may be a strong recruiter of cells from the bloodstream (93). However, R848 was shown to be the strongest modulator of gene expression in the draining lymph nodes, which was characterised by enrichment of type I and II interferon genes as well as T_H1 cytokine genes suggesting this adjuvant is a potent activator of lymphoid tissue (93).

A further preclinical study investigated early inflammatory signals induced following intraperitoneal injection of monophosphoryl lipid-A (MPL) liposomes and alum, representing T_H1

and T_H2 adjuvants respectively (94). Although both adjuvants upregulated a common set of 415 genes, MPL liposomes upregulated more than twice as many genes as alum, and induced a wider range of NK and T cell attracting C-X-C chemokine ligands (94).

Table 1.3: Summary of preclinical studies utilising transcriptomic analysis to assess immune responses to vaccine components.

Reference	Model	Assessment	Summary
(92)	Mouse (BALB/c)	Intramuscular: MF59 oil-in-water emulsion, CpG or alum	◦ 6-8 weeks of age mice, 3 per vaccine and time point ◦ Sampling 3, 6, 12, 24, 48, and 96 hours post-vaccination ◦ Whole homogenized quadriceps muscle ◦ Agilent 44k Whole Mouse Genome Microarray
(93)	Mouse (BALB/c)	Intramuscular inactivated influenza and TT alone or adjuvanted with: MF59, CpG, R848, or Pam3CSK4	◦ 6-8 weeks of age mice, four mice per vaccine group ◦ Sampling 6 hours post-vaccination ◦ Whole homogenized quadriceps muscle and draining lymph nodes ◦ Agilent 44k Whole Mouse Genome Microarray
(94)	Mouse (BALB/c or C57BL/6)	Intraperitoneal: MPL liposomes or Alum	◦ <6 months of age mice, 6 mice per vaccine group ◦ Sampling 26 hours post-vaccination ◦ Peritoneal exudate cells ◦ Affymetrix GeneChip mouse genome 430A 2.0 arrays

MPL= monophosphoryl lipid A, TT= tetanus toxoid , CpG= non-methylated CpG oligonucleotide

1.4.4 Comparative transcript profiling

There are now published in vivo transcriptomics data from human vaccine trials following live-attenuated vaccinia, tularaemia, yellow fever and influenza vaccines, inactivated influenza vaccine, pneumococcal and meningococcal polysaccharide vaccines, meningococcal conjugate vaccine, recombinant malaria protein subunit vaccine, as well as a recombinant HIV-1 adenovirus vectored vaccine (84, 85, 86, 87, 88, 89, 90, 91, 95, 101). Supplementing these in vivo data are the data generated from in vitro and pre-clinical studies (92, 93, 94, 101, 104, 105, 106). However, extracting functionally and ultimately clinically useful knowledge from these high-throughput data remains a considerable challenge. In addition to the generation of larger and better defined high-throughput datasets integrative analysis of existing data, much of which is publicly available, has also been highlighted as a path to the discovery of novel biology. Given the fact that most vaccines mediate protection via antibody there may be generic gene signatures associated with, or even predictive of, robust vaccine responses (21). Conversely, vaccine medi-

ated protection may not be solely proportional to concentrations or titres of specific-antibody, for example LAIV is more effective at preventing childhood culture-confirmed influenza than TIV despite evoking lower titres of systemic neutralising antibody (96). While localised mucosal immunity is likely to play an important role in this protection, it may be possible to use comparative transcriptional analyses to help elucidate the molecular origins of these epidemiological and immunological findings.

Recent analyses have been undertaken to compare the transcriptional signatures evoked by different vaccine compositions (85, 86, 87). The aims of such studies were, 1) to assess whether a 'universal signature' of a desired immunological response exists that could predict robustness of vaccine responses, 2) describe gene expression profiles that are indicative of particular vaccines or routes of administration, 3) elucidate molecular differences in the immunogenicity and/or reactogenicity seen between vaccine compositions.

A recent study used a systematic framework to compare transcriptomic signatures following five human vaccines: live-attenuated YF-17D, live-attenuated influenza vaccine, trivalent-inactivated influenza vaccine, quadrivalent-meningococcal polysaccharide vaccine and quadrivalent-meningococcal conjugate vaccine (86). This study used three methods to compare genes induced by these five vaccines, 1) gene interaction data interactome and bibliome to annotate the biological context of the observed DEGs, 2) robust positional test in gene set enrichment analysis to form molecular pathways based upon the DEGs, 3) a novel blood transcription module approach (86).

Using the first approach day 7 gene profiles following conjugated meningococcal and inactivated influenza vaccine overlapped in a network of genes expressed by antibody-secreting cells (ASCs) (86). Transcriptional responses to YF-17D and LAIV shared a network enriched for genes encoding molecules associated with IFN-related genes and T cell receptor signalling (86). The second approach identified several pathways associated with cell proliferation and correlated the expression of genes in pathways associated with the B cell receptor signalling pathway with antibody responses to TIV and meningococcal conjugate vaccine (86).

The last approach was a novel method, which identified context-specific blood transcription

modules (BTMs) from reconstructed gene networks derived from publicly available microarray data (86). This analysis identified distinct gene profiles that correlated with antibody responses to DT and meningococcal polysaccharide components of the meningococcal conjugate vaccine; however, considerable overlap was seen in profiles correlating with antibody responses to DT and TIV (86). Furthermore, a common dendritic surface signature module correlated with antibody responses to meningococcal polysaccharide components following both the conjugate and polysaccharide meningococcal vaccine (86). Interestingly, T cell activation modules at day 7 post-vaccination negatively correlated with antibody responses to TIV and meningococcal polysaccharide vaccine, despite the latter classically being thought to evoke T cell independent immune responses (86). These data provide some early results comparing and contrasting transcriptional responses following different vaccines, and the relationship between these profiles and immunological endpoints.

1.5 Genomic and transcriptomic indicators of vaccine efficacy

Hitherto, only two published studies have attempted to relate genomic or transcriptomic analyses directly to vaccine efficacy. The first, described an association between variants in *MAL/TIRAP* and *IL-10* with Hib vaccine failure, showing individuals homozygous for the *MAL/TIRAP* Ser180Leu polymorphism to be 5.6 times more at risk of invasive, non-meningitis Hib disease (47). Vaccine failures may be a particularly informative cohort to study in depth as it is logical to believe that individuals with 'extreme phenotypes' also harbour genetic variants with larger effects (107). A second study described gene expression profiles after the novel RTS,S malaria vaccine that associated with protection following infectious challenge (90). This study showed peripheral blood from protected individuals 14 days post-vaccination was enriched for genes in the proteasome degradation pathway, which is involved in the efficient processing of the peptides for HLA presentation (90). While this finding requires further functional and mechanistic investigations, it highlights the potential use of transcriptomic analysis to elucidate novel

correlates of protection.

1.6 Current genomic limitations

Genetic association analyses are complicated by the vastness of genetic variation at the population level, with approximately four million bases of the human genome varying between two randomly chosen individuals (108). This complexity is further compounded by the possibility of multiple variants having only modest individual but considerable synergistic and/or epistatic influences. Alternatively, rare variants with strong effect sizes may be critical but these are not well covered by current DNA microarrays. Moreover, the heritability seen in vaccine responses may not be entirely accounted for by variation in the genomic DNA sequence. Empirical observations have shown the heritability of several complex traits to be more dynamic and context dependent than would be expected by the transmission of stable DNA sequence, leading to speculation that heritable epigenetic modifications may also be important (109).

Technical and particularly biological replication is necessary to validate transcriptomic datasets, as these data are especially vulnerable to artefact. Peripheral blood presents an easily accessible source of circulating immune cells; consequently, it is the most utilised tissue in vaccine related transcriptomic studies. However, a multitude of temporal and physiological factors, as well as technical aspects of sampling can affect transcript measurements (110). Sample preparation and cellular composition can strongly influence transcript levels, for example whole blood differs drastically from PBMC samples [39]. Whole blood samples may decrease the detection sensitivity of transcripts originating from cells other than reticulocytes; conversely, the manipulation of blood samples to ascertain PBMCs induces artifactual changes indicative of prolonged handling (110). Furthermore, a multitude of genetic and environmental factors may impact on both gene expression and vaccine responses; therefore, it is likely to be advantageous to assess populations uniform for age, sex, ethnicity and vaccination history. It has also been suggested that pathway based methodologies, using a priori biological knowledge to aid analysis, are less prone to spurious results (111).

To date, early transcriptional responses to vaccines have been best characterised, corresponding to innate immune responses. These responses correspond to large fold-changes in gene expression in the first hours and days post-vaccination (84, 85, 88, 90). On the other hand, adaptive immune responses may result in broad but subtle global changes in the peripheral blood transcriptome and have not yet been well characterised (88). Further investigations are necessary to describe genes involved in the adaptive immune response to vaccination, these may require large numbers of participants and/or the enrichment of antigen-specific cells to increase power to detect subtle differential expression.

Microarrays have become the routine platform for assessment of transcriptomic changes following vaccination. However, microarrays have a number of intrinsic limitations including constraint to known coding genes, cross-hybridisation of similar sequences, variable hybridisation of polymorphic sequences, PCR-amplification bias as well as a limited dynamic range due to high backgrounds and signal saturation (112, 113). Furthermore, disparity between different array designs, particularly between single and dual channel microarrays, can make comparisons between microarray data problematic (114).

1.7 Genomic forecast

Ever declining costs, along with rapidly advancing analytical tools, mean that next-generation sequencing (NGS) technologies will become the mainstay of future genomic studies into vaccine responses. Whole-genome sequencing (WGS) will enable more comprehensive capture of common and rare variants. Large sample sizes will be needed to power single rare variants analysis in 'healthy' vaccinees. However, until sequencing costs become inexpensive enough, individuals at the ends of the phenotypic distribution are likely to be targeted. In vaccine studies these extremes may include vaccine failures or persistent non-responders, these individuals may be enriched for functional rare variants with large effect sizes. RNA sequencing (RNA-seq) will also transform the characterisation and quantification of the transcriptomic changes induced by vaccination. NGS has several advantages over microarrays, such as the identification of tran-

scripts not previously annotated and a superior dynamic range (112, 113, 114). Furthermore, sequencing enables the assessment of alternative splice variants, allelic expression, non-coding RNAs and small RNA species (115, 116).

1.8 Conclusion

Variations at both the genomic and transcriptomic level have been shown to correlate with measures of vaccine-induced immunity. These analyses have been shown to be highly amenable to all stages of vaccine development, from the pre-clinical models through to the evaluation of licensed vaccines. Understanding factors that influence immune response has vast translational implications, potentially leading to directed and rational development of new and more efficacious vaccine formulations, with better immunogenicity and safety profiles.

Chapter 2

Candidate gene studies of immunological responses to childhood vaccines

2.1 Overview of chapter

The candidate gene approach to studying the genetic determinants of vaccine response was introduced in subsection '1.2.2 Candidate gene studies'. Candidate gene studies use existing biological knowledge to deduce a list of genes believed to be involved in the phenotype of interest. Typically the first step in a candidate gene study is to examine published literature for factors or genes involved in the study phenotype. This search is likely to review data from many sources: genetic disorders, model organisms and expression studies. Next, variants within these candidate genes will be selected for genotyping, variants known to have functional consequences (e.g. non-synonymous) may be preferentially selected, alternatively polymorphisms that efficiently cover the genetic variation at the candidate loci may be selected (117).

A major criticism of the candidate gene approach to genetic association studies has been the lack of replication (118). This chapter will detail two studies to attempt to replicate findings

between polymorphisms previously associated with childhood responses to capsular group C meningococcal and pertussis vaccines. Some of the data presented in this chapter have been published and a licence agreement is in place for the re-use of tables and figures in this document, where applicable (119).

2.1.1 Chapter aims

i) Assess the relationship between polymorphisms in *TLR3* and *CD44*, and the magnitude of infant immune responses to capsular group C meningococcal conjugate vaccine.

ii) Functional evaluation of exonic *TLR3* haplotype associated with capsular group C meningococcal conjugate vaccine responses.

iii) Assess the relationship between a non-synonymous polymorphism (rs4986790) in *TLR4* and childhood booster responses to acellular pertussis vaccine.

2.2 The relationship between polymorphisms in *TLR3* and *CD44*, and the magnitude of infant immune responses to capsular group C meningococcal conjugate vaccine

2.2.1 Background

2.2.1.1 Meningococcus

Neisseria meningitidis is a Gram-negative diplococcus and a major cause of invasive bacterial infection worldwide, with peak disease incidence occurring in children aged under 2 years of age (120, 121). Invasive meningococcal disease is an infrequent outcome of *N. meningitidis* carriage, which is typically asymptomatic even for epidemic strains (122, 123). Meningococcal disease can be challenging to diagnose at presentation and even with appropriate medical care

the case fatality rate is approximately 6%, and higher in meningococcal sepsis (124). Twelve capsular groups of meningococci have been described, based on distinct capsular antigenicities, with the majority of invasive disease worldwide attributable to capsular groups A, B, C, W, X, and Y (82). Prior to the introduction of the capsular group C meningococcal conjugate vaccine, groups B and C were the major causes of invasive meningococcal disease in Europe and the Americas (125).

2.2.1.2 Meningococcal disease epidemiology and vaccines

Susceptibility to invasive meningococcal disease is not completely understood; however, an inverse relationship with the prevalence of complement-dependent serum bactericidal assay (SBA) titres has been demonstrated (126). These data led to SBA titres being widely adopted as a surrogate marker of capsular group specific protection (127). SBA is used to assess serum for antibodies that are functionally capable of fixing complement, the immunological mechanism believed to mediate protection from meningococcal disease (127). An exogenous complement source, obtained from either a suitable human source (hSBA) or from baby rabbits (rSBA), is used to make the assay more reproducible. In the pre-vaccination era, reciprocal capsular group C meningococcal (MenC) hSBA titres of $\geq$4 were shown to correlate with protection from capsular group C disease (126).

Heat killed bacterial cultures were initially assessed as prophylactic vaccines during meningococcal epidemics in the early 1900s, but these preparations were highly reactogenic and not efficacious (128). Conversely, high molecular weight capsular group A (MenA) and MenC capsular polysaccharide vaccines were shown to be non-reactogenic and to evoke bactericidal antibody (129). Subsequently, the MenC capsular polysaccharide vaccine was shown to reduce capsular group C disease by 87-89% in United States army recruits (130, 131). However, MenC capsular polysaccharide vaccines were shown to provide only short-term protection in adults and older children, which was attributed to plain-polysaccharides stimulating T-cell independent responses that do not drive the formation of immunological memory (132). Furthermore, the MenC polysaccharide vaccine was poorly immunogenic in infants, offering no protection

against disease in those under 2 years of age and has been associated with hypo-responsiveness following a booster dose of vaccine (133, 134, 135). Conversely, immunisation with MenC polysaccharide conjugated to a protein carrier was shown to be immunogenic in infants, and to evoke immunological memory (136). Following the licensure of the MenC conjugate vaccine, reciprocal rSBA titres of $\geq$8 were found to correlate with vaccine efficacy in United Kingdom (UK) toddlers 1 month post-vaccine (127).

Following the introduction of the MenC conjugate vaccine a drastic decline in capsular group C disease was observed (Figure 2.1) (137, 138, 139). However, the magnitude and persistence of immunity following MenC vaccination was shown to vary considerably between individuals, a phenomenon that is characteristic of vaccine responses but is not well understood (3, 140). This heterogeneity is particularly noteworthy in the young where immunity wanes rapidly, such that most older children no longer have protective levels of serum bactericidal activity following infant MenC vaccination (3). Consequently, the United Kingdom recently introduced an adolescent booster dose of MenC vaccine to provide protection into early adulthood (82).

A tetravalent meningococcal glycoconjugate vaccine containing capsular group A, C, W and Y meningococcal polysaccharides conjugated to CRM_{197} has also been shown to be immunogenic in infants (141). However, as the capsular group B meningococcal (MenB) polysaccharide does not elicit bactericidal antibodies in humans, it is not a feasible vaccine candidate (142). Recently, a MenB vaccine (Bexsero®, Novartis) comprising of three recombinant neisserial proteins formulated with outer-membrane vesicles (OMVs) from *N. meningitidis* capsular group B strain NZ98/254 has been licensed by the European Medicines Agency (`http://www.ema.europa.eu/ema/`).

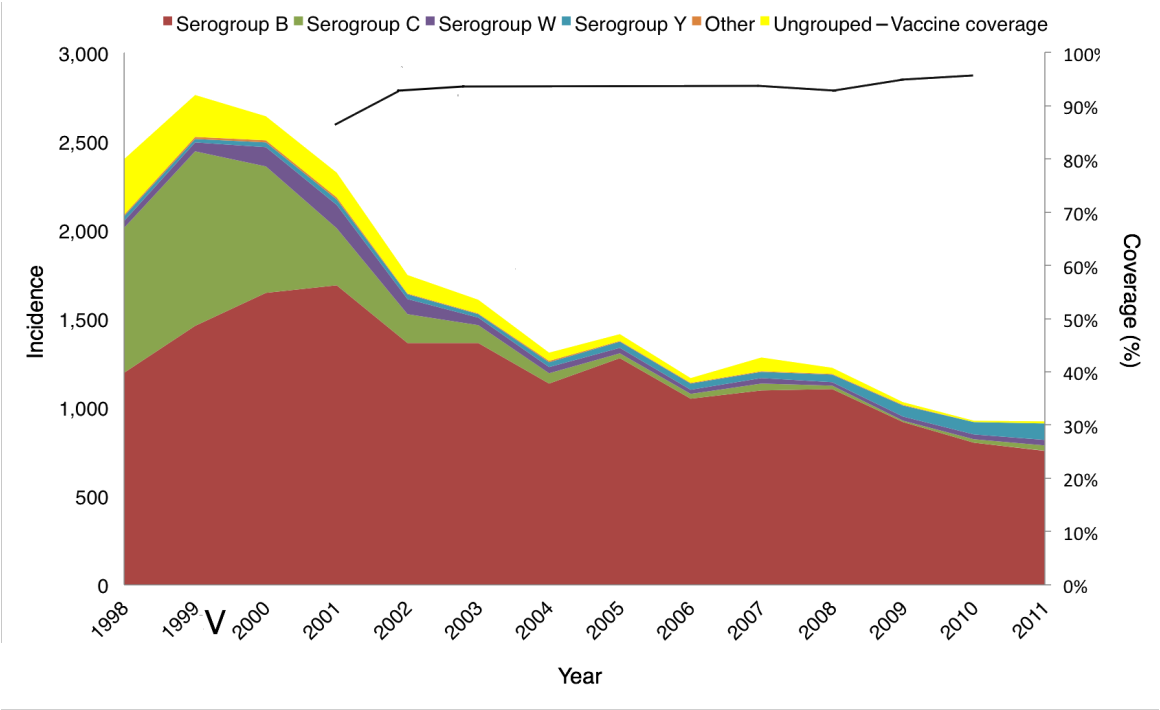


Figure 2.1: Incidence of meningococcal disease in the United Kingdom (UK) from 1998-2011, by capsular group (Data from Health protection agency). V= the capsular group C meningococcal conjugate vaccine was introduced into the routine UK childhood immunisation schedule in November 1999.

2.2.1.3 Genetic studies of meningococcal vaccine responses

A previous study used a candidate gene approach to study the persistence of vaccine-induced immunity following childhood immunisation with MenC conjugate vaccine (38). This study selected nine candidate genes from an in vitro gene expression study of PBMCs cultured with MenC conjugate vaccine, and 122 candidate genes identified from a literature search of genes involved in vaccine-induced immunity, innate immunity, B or T cell immunity (38). A total of 768 single nucleotide polymorphisms (SNPs) were chosen in these genes, based upon frequency and distribution across the loci of interest, with preferential selection of coding changes (38). Participants were categorised into responders and non-responders for both concentrations of MenC-specific IgG (non-responders $<2\,\mu$g/ml and responders $\geq 2\,\mu$g/ml) and reciprocal titres of functional MenC-specific SBA (non-responders <8 rSBA or <4 hSBA, responders ≥ 8 rSBA or ≥ 4 hSBA). In the initial part of this study, a logistic regression model was fitted with vaccine

response status, measured 3.8-6.3 years after primary vaccination of 905 Caucasian teenagers with MenC conjugate vaccine as the response variable, and genotype (heterozygote versus major homozygote and minor homozygote versus major homozygote) as well as non-genetic factors as predictors (38). In the second part of the study, six genes with >1 SNP associated ($p<0.05$) with MenC-specific IgG or SBA in the initial study were examined in a further cohort of 155 children and 196 infants who had received MenC conjugate vaccine 5.9-7.3 years and 8 months previously, respectively (38). Combined analysis found SNPs within *TLR3* and *CD44* to be associated with the persistence of MenC-specific IgG following immunisation with MenC vaccine (38).

While the aforementioned study replicated gene associations in a validation cohort, the associated SNPs were not consistent (38). Furthermore, as SNPs were preferentially selected based upon functional annotation without taking into account locus architecture, variation at these candidate gene loci may not have been optimally captured. The following section presents data assessing systematically derived tagging SNPs (tagSNPs) within *TLR3* and *CD44*, and the magnitude of serological responses following infant immunisation with MenC conjugate vaccine.

2.2.2 Aim

The aim of this section was to assess the relationship between polymorphisms in *TLR3* and *CD44*, and the magnitude of infant immune responses to capsular group C meningococcal conjugate vaccine.

2.2.3 Material and methods

2.2.3.1 Participants

Informed consent was obtained from parents/guardians of 355 infants enrolled in three vaccine trials conducted by one centre in Oxford, United Kingdom. All participants received two or three doses of MenC conjugate vaccine at 2 and 4 or 2, 3 and 4 months of age. Participants received one of three MenC vaccines: monovalent capsular group C meningococcal conjugate

vaccine (MenCC), quadrivalent meningococcal capsular group A, C, W and Y conjugate vaccine (MenACWY) or a combination MenC conjugate vaccine with *Haemophilus influenza* type b (Hib-MenC). Oxfordshire research ethics committees approved these studies (B07/Q1605/41, AO4/Q1604/28 and B07/Q1605/34). An Oxfordshire research ethics committee also approved the genetics study detailed in this section (10/H0504/25).

2.2.3.2 Vaccines

MenCC was a monovalent *N. meningitidis* capsular group C glycoconjugate vaccine containing 10 μg of capsular group C oligosaccharide conjugated to a non-toxic form of diphtheria toxin (CRM_{197}), with aluminium phosphate as an adjuvant. Hib-MenC vaccine contained 5 μg of Hib polyribosylribitol phosphate (PRP) conjugated to 12.5 μg of tetanus toxoid (TT) and 5 μg of *N. meningitidis* capsular polysaccharide C coupled to 5 μg of TT and did not contain an adjuvant. MenACWY consisted of *N. meningitidis* capsular groups A, C, W, and Y capsular polysaccharides (10 μg of capsular group A; 5 μg of each of capsular groups C, W, and Y) individually conjugated to a CRM_{197} with aluminium phosphate as an adjuvant.

2.2.3.3 Immunological measurements

MenC-specific antibody concentrations and SBAs were measured at 5 months of age, 1 month after immunisation with MenC conjugate vaccine (Figure 2.2). The details of the serological assays are summarised below and are published elsewhere (141, 143, 144, 145).

SBAs were performed to detect functional antibody against capsular group C meningococci, using either human complement (hSBA) at Novartis Vaccines, Germany; or rabbit complement (rSBA) at GlaxoSmithKline Biologicals, Belgium. In brief, participant serum was heated to 56 °C for 30 minutes, to inactivated endogenous complement. Capsular group C *N. meningitidis* strain C11 (stock culture stored at -70 °C) was streaked for colony isolation and incubated overnight at 37 °C with 5% CO_2 on a brain-heart infusion 1% horse serum (BHI-HS) agar plate. This was then subcultured for 3-4 hours over the entirety of another BHI-HS agar plate and incubated for 3-4 hours at 37°C with 5% CO_2. After 4 hours these cells were suspended in

DulbeccoÕs PBS (D-PBS) and diluted to 50 to 60 colony forming units/12.5 μl. Next, 50 μl of heat-inactivated serum samples (pre-diluted 1 in 2 for hSBA and 1 in 4 for rSBA) were serially twofold diluted in D-PBS across a 96-well, flat-bottom, tissue culture plate. Equal volumes of cell suspension and complement source (pooled baby rabbit or suitable human serum, rSBA and hSBA respectively) were gently mixed with diluted sample serum. Plates were then incubated at 37 °C without CO_2 for 60 minutes and then 100 μl of tryptic soy broth (TSB) containing 0.9% Noble agar was added to each well and incubated overnight at 37 °C with 5% CO_2. Results were reported as the reciprocal of the highest titre producing a $\geq$50% reduction in bacterial colonies compared to control wells.

Specific IgG concentrations to capsular group C meningococci capsular polysaccharide were measured using an enzyme-linked immunosorbent assay (ELISA). ELISAs were either completed in-house or at GlaxoSmithKline Biologicals, Belgium. The ELISAs completed at GlaxoSmithKline Biologicals used a standardised assay with a lower limit of quantification (LLQ) of 0.3 μg/mL, with values below this being reported as half the LLQ. The in-house capsular group C meningococcal-specific IgG ELISAs were performed using a standardised and quality controlled protocol, calibrated against the international standard reference serum CDC1992 (NIBSC, 99/706), using a previously described method (146, 147). In brief, 1 mg of lyophilised MenC capsular polysaccharide (NIBSC, 07/318) was reconstituted in 1 ml of H_2O and diluted to a concentration of 5 μg/ml in phosphate buffered solution (PBS) (pH 7.2-7.4) containing 5 μg/ml of methylated human serum albumin (mHSA) (NIBSC, 99/592) to assist the binding of polysaccharide to the surface of the polystyrene microtitre plate. Capsular polysaccharides were incubated overnight at 2-8 °C, the next day plates were washed with tris buffered saline (TBS) (0.335 % Brij) and appropriately diluted participant serum, CDC standard 1992 and quality control serum were added and incubated at 2-8°C overnight. The next day plates were washed with TBS (0.335 % Brij) and a secondary mouse anti-human IgG (Stratech Scientific Ltd) antibody conjugated to horse radish peroxidase was added. After 2.5 hours at room temperature plates were washed and a solution containing 3,3′,5,5′-tetramethylbenzidine (TMB) (Sigma, T5525-100TAB) substrate was added, enzymatic substrate cleavage was stopped with

2 M H_2SO_4 after half an hour and the optical density measured at 450 nm. The LLQ for this assay was 0.00167 μg/mL, with values below this being reported as half the LLQ.

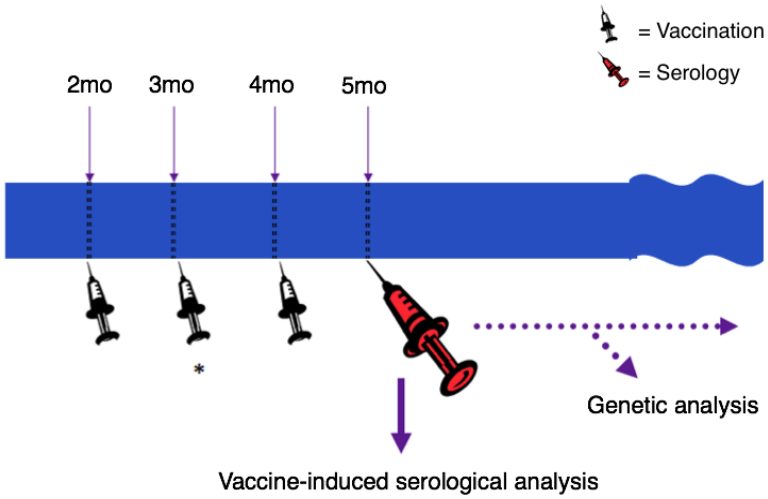


Figure 2.2: Schematic of study into the relationship between polymorphisms in *TLR3* and *CD44*, and the magnitude of infant immune responses to capsular group C meningococcal conjugate vaccine. *= Participants who received the monovalent MenC conjugate vaccine did not receive a dose at 3 months (mo) of age.

2.2.3.4 DNA extraction

Genomic deoxyribonucleic acid (DNA) was extracted using the Maxwell® 16 System (Promega, Wisconsin), following overnight incubation of a 1 ml blood clot in 1 ml of nuclei lysis solution (Promega, Wisconsin), 250 μl of EDTA and 25 μl of Proteinase K (Qiagen, Hilden, Germany) at 56 °C. Extracted DNA was quantified using the Quant-iT™ PicoGreen® assay (Invitrogen, California) and stored at -20 °C.

2.2.3.5 SNP selection

The tagging SNP (tagSNP) approach was adopted to efficiently select SNPs within *TLR3* and *CD44* to optimally capture polymorphisms at these loci (117). TagSNPs within *TLR3* and *CD44* were selected from the HapMap version 3 release 27 dataset (http://www.hapmap.org/) using pairwise tagging, set for a correlation coefficient (r^2) $\geq$0.8 and a minor allelic frequency

(MAF) of ≥0.05. Using the aforementioned parameters, 8 tagSNPs within *TLR3* and 32 tagSNPs in *CD44* were selected for genotyping (Table 2.1).

Table 2.1: List of single nucleotide polymorphisms in *TLR3* and *CD44* selected for genotyping in the candidate gene study of immune responses to capsular group C meningococcal conjugate vaccine.

TLR3	*CD44*		
rs7657186	rs11033008	rs353634	rs7110737
rs13126816	rs2785172	rs353626	rs10768114
rs7668666	rs2553809	rs4141971	rs7105890
rs3775291	rs932703	rs112762	rs1547060
rs3775290	rs11033010	rs11033013	rs7945310
rs3775292	rs353619	rs4756195	rs10734436
rs11732384	rs353618	rs996076	rs11607862
rs5743312	rs353615	rs7934770	rs12419062
	rs353612	rs10128562	rs8193
	rs353644	rs7947433	rs13347
	rs353639	rs11033021	

2.2.3.6 Genotyping

Genotyping was performed using either, an amplification refractory mutation system polymerase chain reaction (ARMS-PCR), or the MassARRAY iPLEX® (Sequenom, San Diego) system (148, 149).

ARMS-PCR genotyping was performed on genomic DNA using a tetra-primer set designed using Primer3 (v.0.4.0) from allelic sequences retrieved from the National Center for Biotechnology Information (NCBI) SNP database. Two allele specific 'inner primers', one forward and one reverse were designed to be amplification refractory to the alternative allele by the introduction of an additional mismatch at the penultimate 3′ base, as described by Little 2001 (148). Common forward and reverse primers were designed at appropriately divergent distances from the allelic position to enable size discrimination of the subsequent amplicons on an agarose gel (appendix Table 7.1). PCR amplifications were performed in a final volume of 25 μl, containing 0.5 μl (10 to 50 ng/μl) of genomic DNA, 0.125 μl Taq polymerase (Qiagen), 5 μl of 5×Q solution (Qiagen), 2.5 μl of 10× buffer (Qiagen), 15 mM $MgCl_2$, primer concentrations were optimised to enable amplification of both common and allele-specific amplicons, and made up to volume

with H_2O. PCR settings were as follows, 95 °C for 5 minutes, (95 °C for one minute, 55 °C for one minute, 68 °C for one minute) × 35 cycles, 68 °C for 10 minutes. PCR amplicons were separated by electrophoresis in a 3% agarose gel stained with ethidium bromide, bands were visualised using a UV transilluminator.

Sequenom MassARRAY® iPLEXs were designed to genotype tagSNPs in *TLR3* and *CD44* (appendix Table 7.2). The 100 bp genomic DNA sequence flanking either side of tagSNPs were retrieved from the NCBI Single Nucleotide Polymorphism Database (dbSNP) (`http://www.ncbi.nlm.nih.gov/projects/SNP/dbSNP.cgi?list=rslist`). The Sequenom Assay Design Suite (`https://www.mysequenom.com/Home`) was used to design iPLEXs, using the high multiplexing pre-set. Pre-genotyping QC eliminated SNPs with high dimer potential, high hairpin potential, unsuitable primer melting temperature (Tm), non-unique PCR and/or extension primers, or where suitable primers could not be found. Sequenom MassARRAY® assays were completed at the Wellcome Trust Centre of Human Genetics core facility (Oxford, UK).

2.2.3.7 Re-sequencing

In addition to tagSNP genotyping, the five exons of *TLR3* were sequenced using the Sanger sequencing methodology (150). Primers were designed to amplify the five *TLR3* exons using primer 3, and assessed for specificity (`http://genome.ucsc.edu/cgi-bin/hgPcr?command=start`) and self-annealing (`http://www.basic.northwestern.edu/biotools/oligocalc.html`) (appendix Table 7.3). PCR amplifications were performed in a final volume of 25 μl, containing 0.5 μl of forward primer, 0.5 μl of reverse primer, 0.5 μl (10 to 50 ng/μl) of genomic DNA, 0.125 μl Taq polymerase (Qiagen), 5 μl of 5× Q solution (Qiagen), 2.5 μl of 10× buffer (Qiagen), and made up to volume with 14.9 μl of H_2O.

PCR amplicons were purified prior to sequencing, 60 μl of PEG (Polyethylene glycol, molecular weight 8000) /2.5 M NaCl was added incubated at room temperature for 30 minutes, and centrifuged at 2750×g_0 for one hour at 4 °C. Excess was removed by inverted centrifugation at 500×g_0 for 1 minute. The DNA pellet was then washed with 150 μl of ice-cold 70% ethanol,

centrifuged at 2750 $\times g_0$ for 10 minutes and excess removed (as previously stated), this step was repeated and the amplicons were re-suspended in H_20. Exon 1 PCR amplicons were purified using the ExoSap methodology as set-up experiments found the ethanol method to result in poor quality sequencing data, possibly due to these amplicons being relatively small and potentially not forming suitable DNA pellets. In this case, 4 μl PCR amplicons were added to 4 μl ExoSap mix: containing 1 μl Shrimp Alkaline phosphatase (Promega), 0.4 μl Shrimp alkaline phosphatase buffer (Promega), 0.1 μl Exonuclease I (New England Biolabs) and 2.5 μl water, and incubated for 30 minutes at 37 °C. The enzymes were then deactivated by heating to 80 °C for 15 minutes.

Sanger dideoxynucleotide sequencing reactions were performed on 2 μl purified PCR product, with the addition of 8 μl of sequencing master mix [1.875 μl of 5× sequencing buffer (Applied Biosystems), 3.875 μl H_20, 2 μl of sequencing primer and 0.25 μl of BigDye® v3.1 Terminator (Applied Biosystems)]. Sequencing reactions were conducted with 25 cycles of, 96 °C for 10 seconds followed by 5 seconds of 50 °C and 4 minutes of 60 °C. Sequencing reaction amplicons were purified by the addition of 50 μl of 100% ethanol, 2 μl EDTA (ethylenediaminetetraacetic acid) (125mM) and 2 μl sodium acetate (3 M, pH 5.5) and incubated at room temperature for 10 minutes. This mixture was then centrifuged at 3000$\times g_0$ at 4 °C, for 30 minutes. Excess was removed with a 5 second inverted centrifugation (~150$\times g_0$), and 70 μl of 70% ethanol was added and the centrifuged at 1750$\times g_0$ at 4 °C for 15 minutes. The ethanol was removed with a one minute inverted centrifugation at 180$\times g_0$.

Purified sequencing reactions were separated by capillary electrophoresis (ABI PRISM® 3700 Genetic Analyzer, Applied Biosystems, California). Sequence chromatograms were assembled and analysed, using the Staden package version 1.6 (`http://staden.sourceforge.net/`). The mutation scanner module was used to compare sequence traces to a reference trace to detect mutation, as described in Bonfield et al 1998 (151). SNPs called by this algorithm were then manually verified by inspection of the forward and reverse strand sequencing traces.

2.2.4 Statistical Analysis

2.2.4.1 Linear regression

Additive and genotypic linear regression models were used to test for associations between tagSNPs and MenC vaccine responses, using the statistical package 'Genetics' in R (152, 153). Genotypic models used the major homozygote as the reference genotype. Reciprocal MenC-specific SBA titres and IgG concentrations were $\log_{10}$ transformed for analysis. Vaccine administered and complement source (for SBA analysis), were included as covariates in regression models.

2.2.4.2 Regional imputation

Bayesian IMputation-Based Association Mapping (BIMBAM) software was used to infer additional untyped genotypes in the two candidate genes, using the HapMap Phase II-III release 28 dataset (forward strand) as a reference panel (154, 155). The Bayes factor (BF) for each SNP was computed under a linear model (Equation 2.1).

$$Y_i = \mu + aX_i + dI(X_i = 1) + \epsilon_i \tag{2.1}$$

Where Y_i denotes the residuals of the $\log_{10}$ transformed reciprocal MenC-specific SBA titres or IgG concentrations regressed for covariates (previously mentioned), X_i represents genotype and ϵ_i denotes an error term. SNPs were filtered to exclude SNPs with $\geq$0.2 missingness and/or $\leq$0.05 MAF. SNP p-values were computed by assessing the proportion of permutations with SNP BFs that exceeded that of the observed data.

Haplotype analysis

Haplotype based analysis was undertaken using PLINK (`http://pngu.mgh.harvard.edu/~purcell/plink/`) (156). An expectation-maximisation algorithm was used to predict phase: 2, 3, 4 and 5 SNP sliding-windows were examined.

2.2.5 Results

2.2.5.1 Demographics

DNA samples were available from 355 infants enrolled in MenC vaccine trials, 318 (90%) were Caucasian and included in the association analysis (Figure 2.3). Serological measures of vaccine immunity were assessed one month after receiving MenC conjugate vaccine, at 5 months of age. MenC-specific SBA titres were available on all 318 participants; however, MenC-specific IgG concentrations were only available in a subset of 135 (42%) participants.

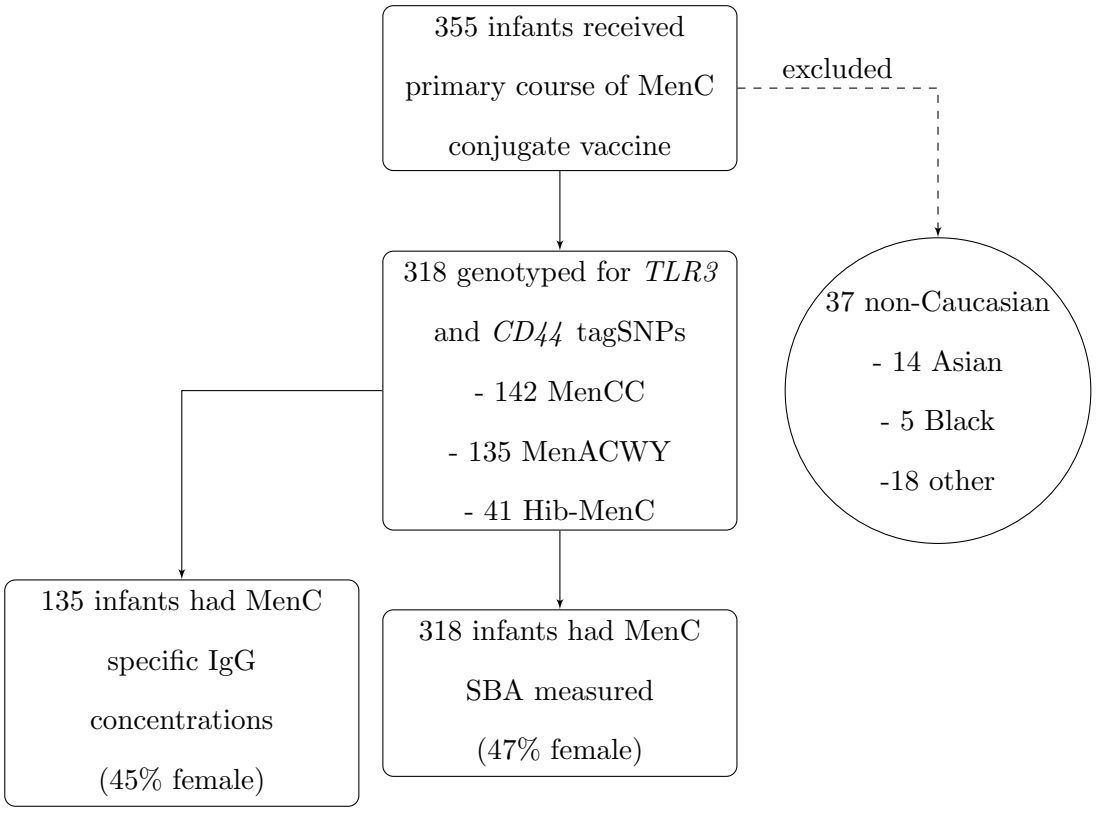


Figure 2.3: Flow diagram of participants included in the study into the relationship between polymorphisms in *TLR3* and *CD44*, and the magnitude of infant immune responses to capsular group C meningococcal conjugate vaccine. SBA = serum bactericidal assay, tagSNP = tagging single-nucleotide polymorphisms. This figure was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

2.2.5.2 Serology

Serological results are summarised in Figure 2.4 and Figure 2.5, and have been published elsewhere (141, 143, 144, 145). Formal statistical comparisons between the MenC-specific SBA titres following TT and CRM conjugated MenC vaccines were not undertaken in this section, as all SBAs conducted on the TT subgroup used rabbit complement, intrinsically more bactericidal than the human complement used in the CRM subgroup. MenC-specific IgG concentrations were statistically higher at 5 months of age following the TT conjugated MenC vaccine (Figure 2.5). While bias may have been introduced by the MenC-specific IgG concentrations having been measured in different laboratories for the TT and CRM conjugated MenC vaccine subgroups, these data are consistent with previous data suggesting the TT conjugated MenC vaccine is more immunogenic in children (157).

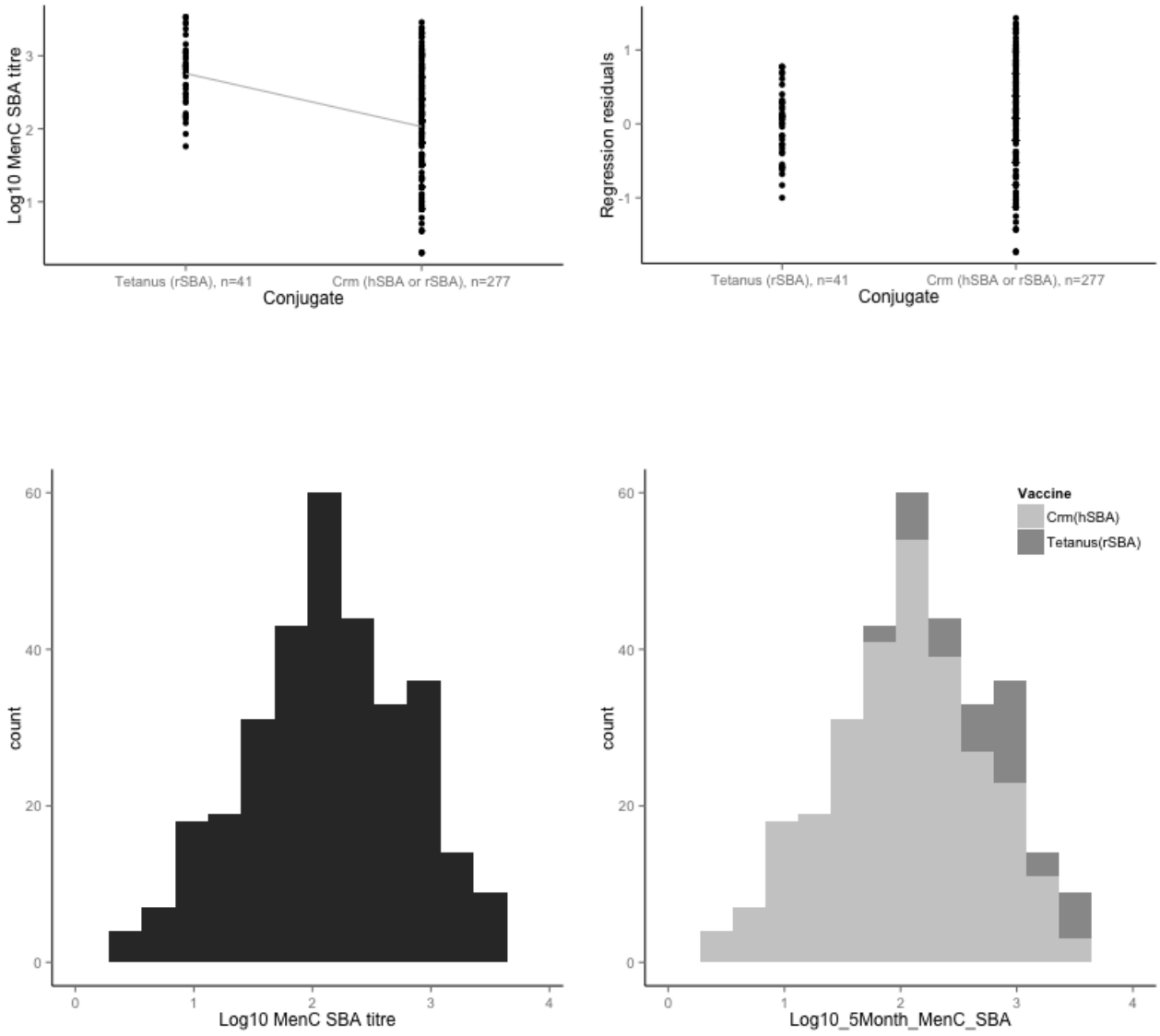


Figure 2.4: Capsular group C meningococci (MenC)-specific serum bactericidal assay titres at 5 months of age, 1 month after receiving MenC conjugate vaccine. Top-left, $\log_{10}$ transformed MenC-specific SBA titres separated by protein-carrier conjugated to MenC capsular polysaccharide. Top-right, plot of the residuals of linear model with $\log_{10}$ MenC-specific SBA titres as a response variable and protein-carrier conjugated to MenC capsular polysaccharide as the predictor variable. Bottom-left, histogram of $\log_{10}$ MenC-specific SBA titres. Bottom-right, histogram of $\log_{10}$ MenC-specific SBA titres coloured by protein-carrier conjugated to MenC capsular polysaccharide. Note: rSBA denotes rabbit complement and hSBA denotes human complement, was used in the respective SBA.

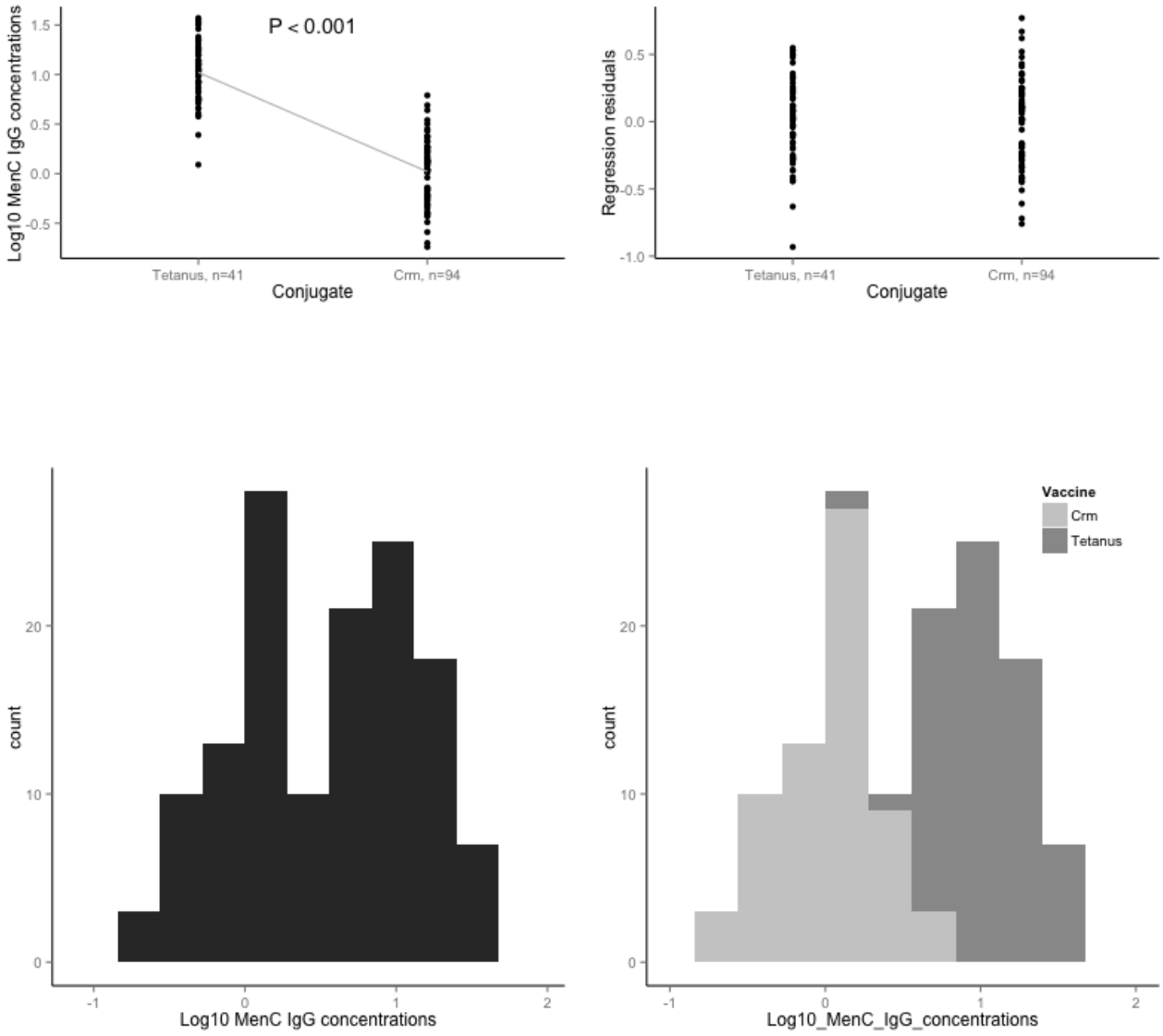


Figure 2.5: Capsular group C meningococci (MenC)-specific IgG concentrations at 5 months of age, 1 month after receiving MenC conjugate vaccine. Top-left, $\log_{10}$ transformed MenC-specific IgG concentrations separated by protein-carrier conjugated to MenC capsular polysaccharide. Top-right, plot of the residuals of linear model with $\log_{10}$ MenC-specific IgG concentrations as a response variable and protein-carrier conjugated to MenC capsular polysaccharide as the predictor variable. Bottom-left, histogram of $\log_{10}$ MenC-specific IgG concentrations. Bottom-right, histogram of $\log_{10}$ MenC-specific IgG concentrations coloured by protein-carrier conjugated to MenC capsular polysaccharide.

2.2.5.3 Genotyping

Three *TLR3* tagSNPs were genotyped using the ARMS-PCR method (appendix Table 7.1). Genotypes were manually called from ethidium bromide stained DNA bands on agarose gels, using an UV transilluminator (supplementary Figure 7.2). Sequenom MassARRAY® iPLEXs were designed to genotype the remaining 37 tagSNPs, two of these failed primer design (one in *TLR3* and one in *CD44*). This Sequenom project formed two iPLEXs, the first containing 27 SNPs and the second containing only 8 SNPs (appendix Table 7.2). Only the first iPLEX design was carried through to the genotyping stage, this iPLEX contained 4 *TLR3* tagSNPs and 23 *CD44* tagSNPs. Seven of the tagSNPs failed post-genotyping quality control (QC) metrics.

2.2.5.4 Hardy-Weinberg equilibrium

TagSNPs were tested for departure from Hardy-Weinberg Equilibrium (HWE), which can be indicative of genotyping error, using Plink (`http://pngu.mgh.harvard.edu/~purcell/plink/`). Of the 23 tagSNPs that passed genotyping QC metrics, 1 significantly deviated from HWE and was excluded from association analysis (Table 2.2).

Table 2.2: Results of quality control performed on single-nucleotide polymorphisms selected in the study of the relationship between genetic variations in *CD44* and *TLR3*, and serological responses to capsular group C meningococcal conjugate vaccine.

SNP	Gene	Sequenom QC	HWE p-value	HWE QC
rs3775292	*TLR3*	Passed	0.804	Passed
rs11732384	*TLR3*	Passed	1.000	Passed
rs5743312	*TLR3*	Passed	0.745	Passed
rs3775291	*TLR3*	Passed	0.073	Passed
rs7657186	*TLR3*	Passed	0.456	Passed
rs13126816	*TLR3*	Passed	0.288	Passed
rs7668666	*TLR3*	Passed	0.113	Passed
rs10734436	*CD44*	Passed	0.714	Passed
rs11033008	*CD44*	Passed	0.723	Passed
rs11033010	*CD44*	Passed	0.113	Passed
rs11033021	*CD44*	Passed	0.526	Passed
rs112762	*CD44*	Passed	0.092	Passed
rs12419062	*CD44*	Passed	1.000	Passed
rs2785172	*CD44*	Passed	0.126	Passed
rs353615	*CD44*	Passed	0.114	Passed
rs353619	*CD44*	Passed	0.599	Passed
rs353626	*CD44*	Passed	0.664	Passed
rs353639	*CD44*	Passed	0.320	Passed
rs7105890	*CD44*	Passed	0.260	Passed
rs7945310	*CD44*	Passed	1.000	Passed
rs7947433	*CD44*	Passed	0.737	Passed
rs932703	*CD44*	Passed	0.008	Failed
rs996076	*CD44*	Passed	0.563	Passed
rs353644	*CD44*	Failed	N/A	N/A
rs4756195	*CD44*	Failed	N/A	N/A
rs11033013	*CD44*	Failed	N/A	N/A
rs4141971	*CD44*	Failed	N/A	N/A
rs7934770	*CD44*	Failed	N/A	N/A
rs8193	*CD44*	Failed	N/A	N/A
rs353634	*CD44*	Failed	N/A	N/A

QC= quality control, SNP= single-nucleotide polymorphism
HWE= Hardy-Weinberg equilibrium

2.2.5.5 Linear regression

The results of the additive and genotypic linear regression models, assessing the relationship between tagSNPs and MenC-specific SBA titres following infant immunisation with MenC conjugate vaccine are shown in Table 2.3 and Table 2.4. Individuals homozygous for the minor allele at rs11732384 in *TLR3* demonstrated significantly higher MenC-specific SBA titres (Table 2.3). No statistically significant relationship between *TLR3* tagSNPs and MenC-specific SBA titres were observed using an additive model.

Table 2.3: Genotypic and allelic association analysis between *TLR3* single-nucleotide polymorphisms and capsular group C meningococci (MenC)-specific serum bactericidal assay titres, 1 month after infant immunisation with MenC conjugate vaccine. This table was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

SNP	Allele	n	GMT	CI (95%)	p-value
rs3775292	C/C	12	247	142-431	0.542
	G/C	117	134	99.9-181	0.959
	G/G	182	131	103-166	-
	Allelic 'C'				0.769
rs11732384	A/A	6	443	93.7-2095	**0.027***
	G/A	83	103	72.9-145	0.156
	G/G	228	142	115-175	-
	Allelic 'A'				0.891
rs5743312	T/T	7	95.5	28.6-319	0.328
	C/T	75	167	118-235	0.271
	C/C	227	128	103-159	-
	Allelic 'T'				0.704
rs3775291	T/T	23	211	92.6-480	0.153
	C/T	119	115	86.6-153	0.562
	C/C	162	131	103-168	-
	Allelic 'T'				0.533
rs7668666	A/A	17	105	53.1 -210	0.410
	C/A	125	131	100-172	1.000
	C/C	171	135	104-174	-
	Allelic 'A'				0.615
rs7657186	A/A	11	150	51.3-437	0.936
	G/A	119	132	97.3-179	0.908
	G/G	186	133	106-167	-
	Allelic 'A'				0.895
rs13126816	A/A	12	185	71.5-479	0.391
	G/A	119	123	91.8-166	0.615
	G/G	185	137	109-173	-
	Allelic 'A'				0.930
	All	318	133	111-159	-

GMT= geometric mean titre; CI = confidence interval
* = significant, with a p-value less than 0.05

Homozygous individuals for the 'C' allele at rs112762 in *CD44* had significantly lower MenC-specific SBA titres, and a significant allelic association was also evident at this SNP (Table 2.4). Furthermore, homozygous individuals for the 'A' allele at rs7105890 in *CD44* had significantly lower SBA titres (Table 2.4).

Table 2.4: Genotypic and allelic association analysis between *CD44* single-nucleotide polymorphisms and capsular group C meningococci (MenC)-specific serum bactericidal assay titres, one month after infant immunisation with MenC conjugate vaccine. This table was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

SNP	Allele	n	GMT	CI(95%)	p-value
rs10734436	A/A	37	104	63.7-171	0.4953
	G/A	146	148	115-189	0.4977
	G/G	129	131	97.3-176	-
	Allelic 'A'				0.8282
rs11033008	C/C	12	192	73-507	0.5482
	C/T	95	112	80.3-158	0.2387
	T/T	209	141	113-175	-
	Allelic 'C'				0.6282
rs11033010	A/A	38	177	115 -271	0.3819
	G/A	125	122	90.3-164	0.2394
	G/G	152	133	103-172	-
	Allelic 'A'				0.6749
rs11033021	T/T	32	102	57.7-182	0.3201
	T/C	146	133	101-176	0.7337
	C/C	138	142	110 -183	-
	Allelic 'T'				0.3729
rs112762	C/C	81	101	69.5-146	**0.0471***
	T/C	141	140	108 -181	0.46527
	T/T	91	164	118 -228	-
	Allelic 'C'				**0.0487***
rs12419062	G/G	34	94.9	51 -176	0.1739
	G/A	138	137	105 -179	0.8667
	A/A	144	140	108-183	-
	Allelic 'G'				0.2801
rs2785172	G/G	54	135	85.2 - 215	0.3982
	G/A	137	107	82.3 - 140	0.0243
	A/A	125	168	127 - 222	-
	Allelic 'G'				0.1687
rs353615	T/T	36	112	65.6 - 191	0.7538
	T/C	123	150	114 - 197	0.3690
	C/C	157	127	97.3 - 165	-
	Allelic 'T'				0.8542
rs353619	A/A	11	155	53.6 - 449	0.9703
	G/A	104	105	77 - 144	0.0693
	G/G	201	149	119 - 187	-
	Allelic 'A'				0.1802
rs353626	C/C	22	197	108 - 359	0.2737
	C/T	117	130	96.4 - 176	0.9756
	T/T	177	129	101 - 164	-
	Allelic 'C'				0.4531
rs353639	G/G	26	125	70.4 - 220	0.8599
	G/T	116	153	112 - 209	0.2392
	T/T	173	122	96.1 - 155	-
	Allelic 'G'				0.4764
rs7105890	A/A	7	42.2	4.49 - 396	**0.0443***
	G/A	100	143	105 - 196	0.9295
	G/G	204	135	109 - 168	-
	Allelic 'A'				0.3854
rs7945310	T/T	53	161	105 - 246	0.5071
	C/T	153	120	92.2 - 157	0.5243
	C/C	110	140	105 - 189	-
	Allelic 'T'				0.7047
rs7947433	G/G	79	93	66.4 - 130	0.0639
	G/A	155	151	116 - 197	0.9252
	A/A	82	149	106 - 208	-
	Allelic 'G'				0.0662
rs996076	A/A	72	112	76.2 - 165	0.4746
	A/G	151	146	112 - 189	0.6055
	G/G	90	130	93 - 180	-
	Allelic 'A'				0.5349
	All	318	133	111 - 159	

GMT= geometric mean titre; CI = confidence interval
* = significant, with a p-value less than 0.05

The results of the additive and genotypic linear regression analyses between tagSNPs and MenC-specific IgG concentrations, following infant immunisation with MenC conjugate vaccine, are shown in Table 2.5 and appendix Table 7.4. Heterozygous individuals for the *TLR3* SNP rs11732384 had significantly lower post-vaccination MenC-specific IgG concentrations (Table 2.5). However, the additive linear regression model did not show any statistically significant relationships between *TLR3* tagSNPs and post-vaccination MenC-specific IgG concentrations. Furthermore, none of the *CD44* tagSNPs showed evidence for association with post-vaccination MenC-specific IgG concentrations, using either the genotypic or additive models (appendix Table 7.4).

Table 2.5: Genotypic and allelic association analysis between *TLR3* single-nucleotide polymorphisms and capsular group C meningococci (MenC)-specific IgG concentrations, one month after infant immunisation with MenC conjugate vaccine. This table was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

SNP	Allele	n	GMC	CI(95%)	p-value
rs3775292	C/C	6	5.07	1.74-14.7	0.229
	C/G	47	4.15	2.74-6.29	0.867
	G/G	82	3.04	2.25-4.09	-
	Allelic 'C'				0.411
rs11732384	A/A	3	10.10	0.21- 483	0.120
	A/G	34	2.87	1.86-4.43	**0.009****
	G/G	97	3.57	2.69-4.74	-
	Allelic 'A'				0.194
rs5743312	T/T	4	3.38	0.207-55	0.848
	C/T	34	3.42	2.26-5.17	0.302
	C/C	96	3.48	2.61-4.65	-
	Allelic 'T'				0.366
rs3775291	T/T	14	2.84	0.97-8.35	0.860
	C/T	44	3.86	2.59-5.75	0.602
	C/C	77	3.38	2.51-4.54	-
	Allelic 'T'				0.694
rs7668666	A/A	5	4.03	0.65-25.1	0.938
	C/A	57	2.70	1.93-3.77	0.943
	C/C	66	4.52	3.17-6.45	-
	Allelic 'A'				0.923
rs7657186	A/A	5	6.13	1.2-31.2	0.887
	A/G	49	3.39	2.32-4.98	0.544
	G/G	75	3.57	2.57-4.94	-
	Allelic 'A'				0.703
rs13126816	A/A	3	3.53	0.01-2321	0.458
	A/G	46	3.91	2.56-5.99	0.890
	G/G	80	3.39	2.54-4.54	-
	Allelic 'A'				0.819
	All	135	3.46	2.75 -4.37	-

GMC= geometric mean concentration; CI = confidence interval
* = significant, with a p-value less than 0.05
** = significant, with a p-value less than 0.01

2.2.5.6 Re-sequencing

Five exonic variants were identified through the sequencing of the *TLR3* exons, all of which were referenced on dbSNP (`http://www.ncbi.nlm.nih.gov/SNP/`). A single exonic variant rs3775290 was significantly associated with post-vaccination MenC-specific IgG concentrations, using additive and genotypic linear models (Table 2.6). None of the exonic *TLR3* SNPs showed evidence for association with MenC-specific SBA titres (appendix Table 7.5).

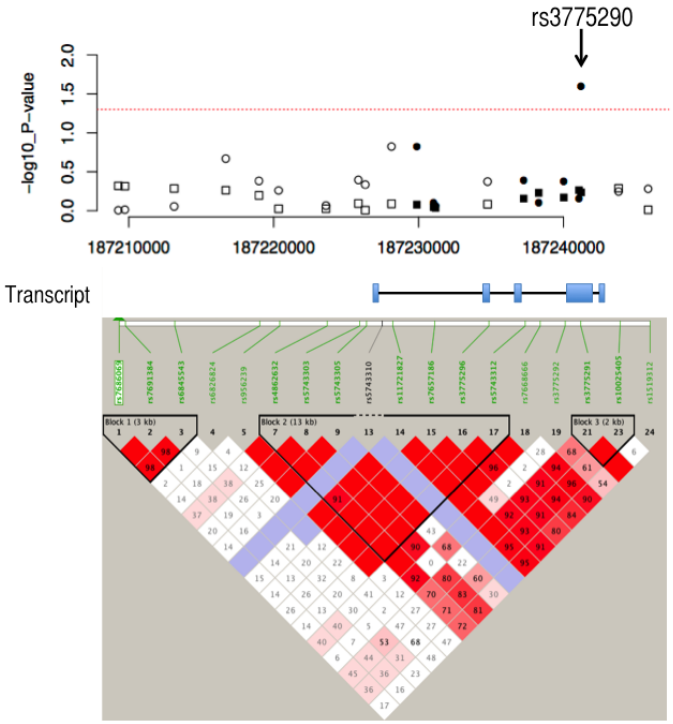
Table 2.6: Genotypic and allelic association analysis between exonic *TLR3* variants and capsular group C meningococci (MenC)-specific IgG concentrations, one month after infant immunisation with MenC conjugate vaccine. This table was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

SNP	Allele	n	GMC	CI(95%)	p-value
rs3775290	T/T	18	2.97	1.63-5.43	0.091
	C/T	51	2.99	2.03-4.41	**0.021***
	C/C	65	4.12	2.91-5.82	-
	Allelic 'T'				**0.025***
rs3775291	T/T	14	2.84	0.97-8.35	0.860
	C/T	44	3.86	2.59-5.75	0.602
	C/C	77	3.38	2.51-4.54	-
	Allelic 'T'				0.694
rs3775296	T/T	2	1.43	0.28-7.36	0.571
	C/T	27	2.3	1.44-3.66	0.953
	C/C	104	4.07	3.11-5.33	-
	Allelic 'T'				0.753
rs73873710	T/T	0	-	-	-
	C/T	3	9.44	4.25-21	0.787
	C/C	130	3.49	2.76-4.42	-
	Allelic 'T'				0.787
rs116729895	T/T	0	-	-	-
	C/T	3	9.44	4.25-21	0.787
	C/C	130	3.49	2.76-4.42	-
	Allelic 'T'				0.787
	All	135	3.46	2.75-4.37	-

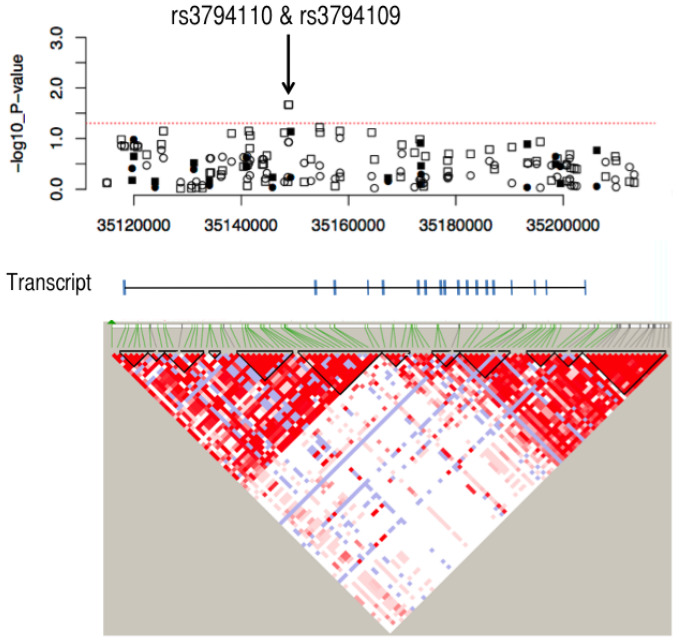
GMC= geometric mean concentration; CI = confidence interval
* = significant, with a P-value less than 0.05

2.2.5.7 Regional imputation

Gene-based imputation resulted in the capture of an additional 13 *TLR3* SNPs and 71 *CD44* SNPs. Using an additive model none of the imputed *TLR3* SNPs reached statistical significance (Figure 2.6a). However, two imputed SNPs within *CD44*, rs3794109 (p-value = 0.021) and rs3794110 (p-value = 0.022), significantly associated with post-vaccination MenC-specific SBA titres (Figure 2.6b).



(a)



(b)

Figure 2.6: Statistical association (-$\log_{10}$ p-values) between single nucleotide polymorphisms in *TLR3* and *CD44*, and the magnitude of serological responses to capsular group C meningococcal conjugate vaccine, along with a graphic of the linkage disequilibrium structure of the respective genetic regions. a) *TLR3* gene locus, b) *CD44* gene locus. Filled icons are directly genotyped and unfilled icons are imputed SNPs; circles are associations with MenC-specific IgG concentrations and squares are associations with MenC-specific SBA titres. This figure was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

2.2.5.8 Haplotype analysis

Haplotype analyses showed a *TLR3* haplotype defined by two exonic SNPs (rs3775290 and rs3775291) to be significantly associated with MenC-specific IgG concentration following MenC vaccination (p-value = 0.01) (Figure 2.7). These two exonic SNPs lie within the fourth exon of *TLR3*, with the C-C haplotype associated with lower post-vaccination MenC-specific IgG concentrations. A single haplotype within *TLR3* defined by five SNPs was significantly associated with MenC-specific SBA titres (p-value =0.040) (appendix Table 7.6).

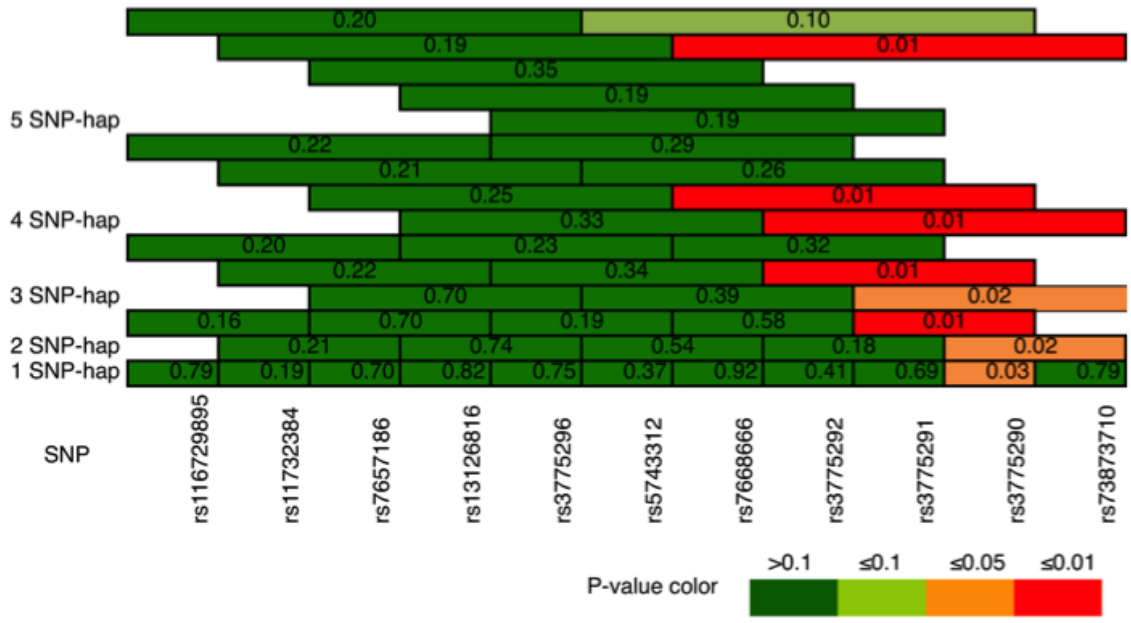


Figure 2.7: *TLR3* haplotypes with the greatest statistical support of association with specific-IgG concentrations to capsular group C meningococcal (MenC) polysaccharide, one month after infant immunisation with MenC conjugate vaccine. The number contained within each box corresponds to the p-value of the most statistically significant haplotype (frequency >5%) for each haplotype window. This figure was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

CD44 haplotypes with the greatest statistical evidence for association with post-vaccination MenC-specific SBA titres and IgG concentrations are shown in Figure 2.8 and Figure 2.9. The

most statistically significant association with post-vaccination MenC-specific SBA titres was evident with a haplotype defined by three *CD44* SNPs (rs112762, rs996076 and rs7947433; p-value = 0.01), with the T-A-A haplotype being associated with greater MenC-specific SBA titres (Figure 2.8).

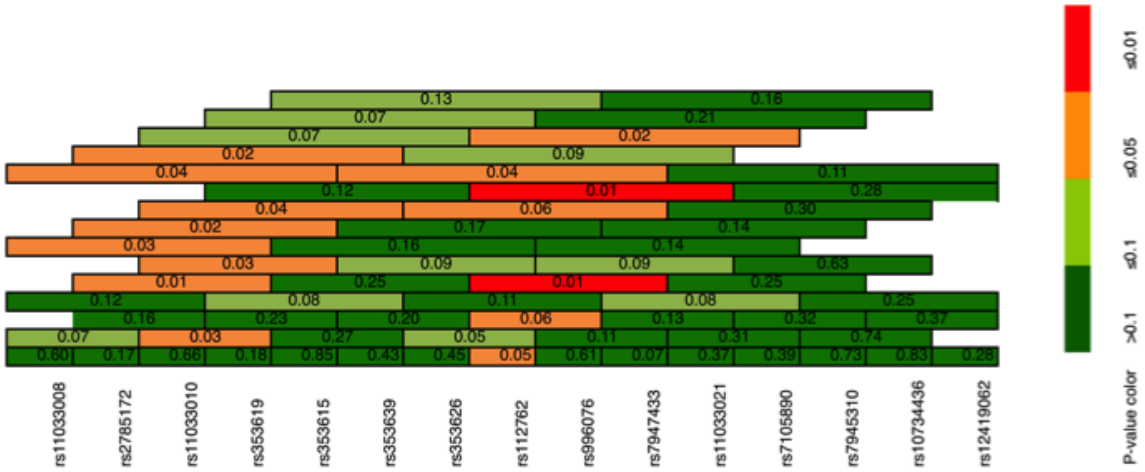


Figure 2.8: *CD44* haplotypes with the greatest statistical support of association with capsular group C meningococci-specific serum bactericidal assay titres, one month after infant immunisation with MenC conjugate vaccine. The number contained within each box corresponds to the p-value of the most statistically significant haplotype (frequency >5%) for each haplotype window. This figure was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

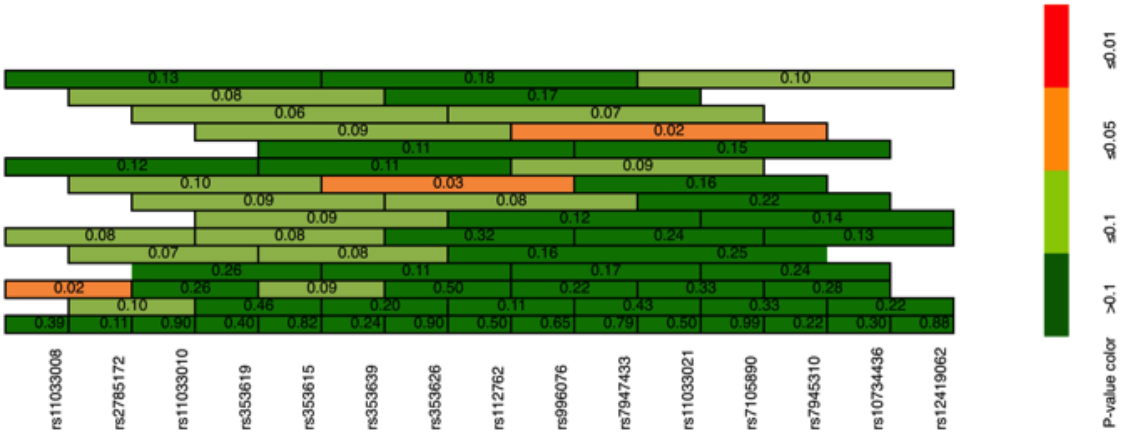


Figure 2.9: *CD44* haplotypes with the greatest statistical support of association with specific-IgG concentrations to capsular group C meningococcal (MenC) polysaccharide, one month after infant immunisation MenC conjugate vaccine. The number contained within each box corresponds to the p-value of the most statistically significant haplotype (frequency >5%) for each haplotype window. This figure was reproduced from O'Connor et al 2014 (119), with licence from Elsevier.

2.2.6 Discussion

Immunological responses to several vaccines have been shown to be highly heritable, particularly in early infancy (7, 8). However, the genetic determinants of these responses have not yet been well characterised. In this study, polymorphisms within two candidate genes were assessed for their influence on the magnitude of vaccine-specific immune responses, following a primary course of MenC conjugate vaccine. Single-SNP analysis showed an association between an exonic SNP (rs3775290) in *TLR3* and post-vaccination MenC-specific IgG concentrations. In addition, three intronic SNPs (rs112762, rs3794109 and rs3794110) in *CD44* were associated with MenC-specific SBA titres.

Haplotype analysis found a haplotype tagged by two exonic *TLR3* SNPs (rs3775291-rs3775290) to be associated with the magnitude of MenC-specific IgG responses. While this haplotype included a missense mutation (rs3775291) residing within the ectodomain of this protein, this SNP is unlikely to be the causal variant as it was not associated in single-SNP analysis. Alternatively, this haplotype may function at the genomic level by influencing mRNA processing,

regulation, expression or stability. Of note, none of the *TLR3* haplotypes significantly associated with post-vaccination MenC-specific SBA titres. This may hint at disparate, or additional, genetic factors being crucial in the production of high-avidity bactericidal antibodies, particularly as MenC-specific IgG concentrations and SBA titres correlate poorly in vaccinees under 2 years of age (158).

Mammalian TLR3 has been linked with immunity to viral infection, and double-stranded RNA (dsRNA) produced during viral replication is an agonist of this predominantly endosomal receptor (159). A mechanistic role for TLR3 in MenC vaccine responses has yet to be demonstrated. However, SNPs in *TLR3* have previously been associated with the persistence of MenC-specific and measles-specific antibodies following vaccination (37, 38). Furthermore, endogenous RNAs from necrotic cells have been shown to be ligands for TLR3 (160). Therefore, it is possible that TLR3 is stimulated at the site of vaccine delivery due to localised tissue necrosis and may consequently have a role in innate and early immune responses.

Haplotype analysis also showed statistical associations between *CD44* haplotypes and both MenC-specific SBA titres and IgG concentrations. CD44 is a cell-surface glycoprotein with multiple protein isoforms encoded by alternative splicing of a single gene locus (`http://www.ncbi.nlm.nih.gov/gene/`). The principal role of CD44 is to maintain tissue structure via cell-cell and cell-matrix adhesion; however, certain isoforms are also involved in lymphocyte activation, adhesion and homing (161). The haplotype that showed the greatest evidence for association in this dataset was tagged by three SNPs (rs112762, rs996076 and rs7947433). This haplotype spans a genomic region of 24 kb and includes exons 2-5 that encode both the ubiquitously expressed isoform as well as variant isoforms of *CD44* (162). Interestingly, haplotypes proximal to this region of *CD44* have previously been associated with responses to HBV (163).

Immunological responses to a number of vaccines have been shown to be heritable (6, 7, 8, 9, 19). However, rare variants or variants of modest effect size may account for a large proportion of this heritability and these present considerable statistical challenges. In order to adequately power association analyses of such variants large sample sizes are required and these are infrequently available. Consequently, there are an accumulating number of genetic variants

that do not surpass conservative thresholds for statistical testing and replication studies are imperative to validate findings.

2.2.7 Conclusion

These data replicate previous findings of an association between SNPs in the *TLR3* and *CD44*, with serological responses to MenC vaccine. Further work is required to demonstrate the functional relevance of these statistical associations.

2.3 Functional evaluation of exonic *TLR3* polymorphisms

2.3.1 Background

Many genetic associations have been described in studies of complex disease and quantitative traits; however, pinpointing causal variants and constructing mechanistic models has been challenging. In section 2.2, the two allele haplotype (rs3775291-rs3775290) was associated with the magnitude of MenC-specific IgG concentrations following infant immunisation with MenC conjugate vaccine. While the mechanistic role of TLR3 in MenC conjugate vaccine responses is uncertain, endogenous RNA is a TLR3 ligand (160). Therefore, a plausible hypothesis is that TLR3 is stimulated at the site of vaccine delivery due to localised tissue necrosis and may consequently have a role in orchestrating innate and early immune responses to vaccines. Furthermore, fibroblasts treated with poly(I:C) (polyinosinic:polycytidylic acid), a synthetic dsRNA analogue and known agonist of TLR3, have been shown to rapidly increase expression of many cytokine mRNA transcripts; including *TNFSF13B*, the expression of which in PBMCs following influenza vaccination has been associated with HAI titres (88, 164).

The rs3775291-rs3775290 haplotype may function at the genomic or protein level, or both. At the genomic level exonic base changes could influence mRNA processing, regulation, expression or stability. This haplotype may also be functioning at the protein level; the two exonic SNPs defining this haplotype encode amino acids that reside within the ectodomain of TLR3, in a region of leucine-rich repeats (LRR) (Figure 2.10). LRRs are versatile structural

frameworks that are frequently involved in protein-protein interactions (165). While rs3775290 is a synonymous mutation, rs3775291 is a nonsynonymous mutation predicted by Polyphen to be damaging (`http://genetics.bwh.harvard.edu/pph/`). This section describes an in vitro study to assess the influence of the two allele haplotype, rs3775291-rs3775290, in *TLR3* on PBMC responses to stimulation with poly(I:C).

2.3.2 Aim

This section tested the hypothesis that the association described between polymorphisms in *TLR3* and immune responses to MenC conjugate vaccine is linked to genetically determined differences in this receptors ability to respond to its agonist poly(I:C).

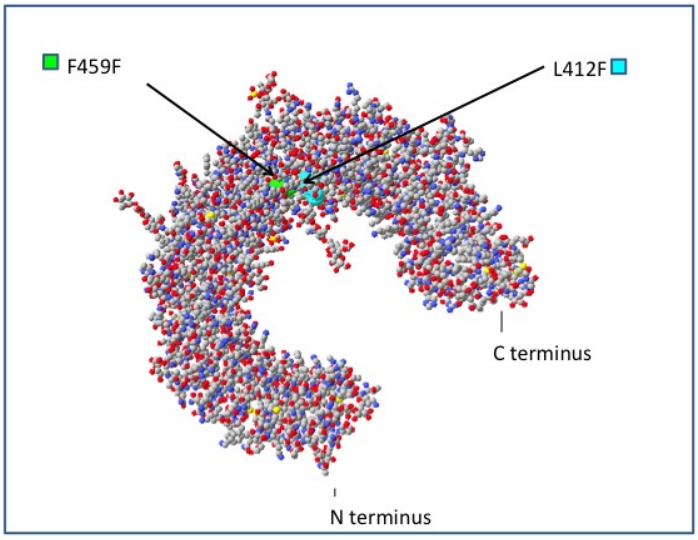


Figure 2.10: Diagram of the position of amino acids, coded by the single-nucleotide polymorphisms rs37775291 and rs3775290, within the *TLR3* ectodomain (protein data bank ID code 2A0Z). This schematic was designed using Swiss-PdbViewer (`http://spdbv.vital-it.ch/`). L412F represents a change in the 412^{th} amino acid, leucine (CTC) to a phenylalanine (TTC), due to a SNP at rs3775291. F459F is a synonymous base substitution TTC (phenylalanine) for TTT (phenylalanine) the SNP in this case is rs3775290.

2.3.3 Materials and Methods

2.3.3.1 Participants

A bank of frozen PBMCs from 60 healthy adult vaccine trial participants (Oxford Research Ethics Committee A 06/Q1604/121) were genotyped at SNPs rs3775291 and rs3775290. To avoid ambiguity, those who inherited the C-C haplotype from both parents (n=8) were compared to a control group (n=8) with neither a maternal nor paternal copy of this haplotype (i.e. homozygote 'T' genotype at rs3775291 and/or rs3775290). A flow diagram of the study design is shown in Figure 2.11.

2.3.3.2 DNA extraction and genotyping

DNA was extracted from cellular pellets available from these vaccine participants as previously described in subsection '2.2.3.4 DNA extraction'. To elucidate the rs3775291-rs3775290 haplotype Sanger sequencing was performed as previously described in subsection '2.2.3.7 Resequencing'.

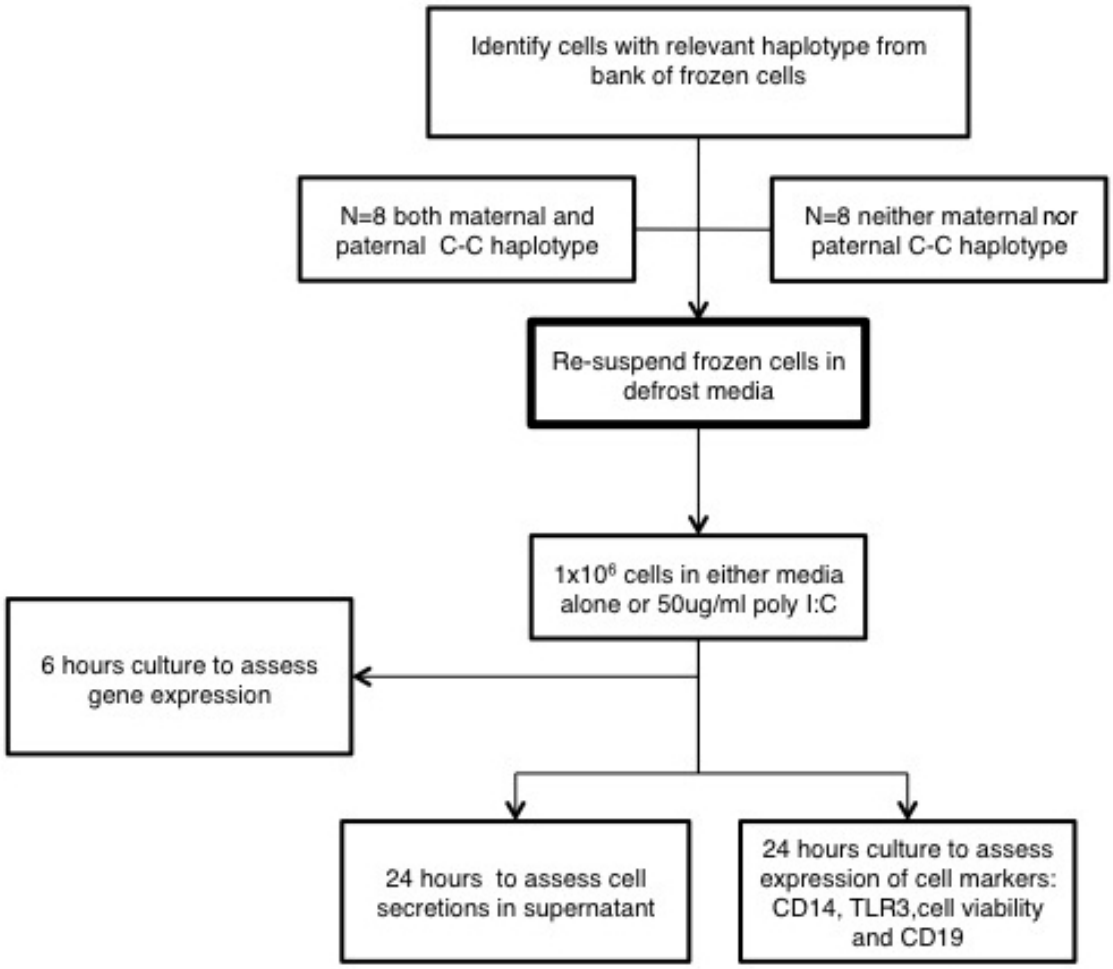


Figure 2.11: Study design for the functional evaluation of the *TLR3* rs3775291-rs3775290 haplotype. A bank of 60 samples were evaluated for the presence of a maternal and paternal copy of the C-C haplotype, eight individuals were identified and an equal number of controls lacking this haplotype were also identified. PBMCs were cultured with or without poly(I:C). After 6 hours of cell culture PBMCs were harvested for gene expression analysis. After 24 hours cell culture supernatant and PBMCs were retrieved for cytokine quantification and phenotyping, respectively.

2.3.3.3 Cell culture

PBMCs had been stored in 1 ml aliquots of 2×10^6 cells in R0 [500 ml RPMI (Sigma), 5 ml penicillin 5,000 units/ml -streptomycin 5 mg/ml (Sigma), 5 ml L-glutamine 200 mM (Sigma)], for between 2 and 3 years. Frozen PBMCs were defrosted at 37 °C and transferred into 10 ml of 37 °C defrost media [450 ml RPMI HEPES modification (Sigma), 50 ml heat inactivated new born bovine serum (Sigma), 5 ml penicillin 5,000 units/ml -streptomycin 5 mg/ml (Sigma), 5 ml L-

glutamine 200 mM (Sigma), 5 ml minimum essential medium (MEM) non-essential amino acids 10mM (Invitrogen), 5 ml MEM sodium pyruvate 100 mM (Invitrogen), 0.5 ml 2-mercaptoethanol 50 mM (Invitrogen)]. Samples were centrifuged at 1800 rpm for 6 minutes, and the pellet was re-suspended in 10 ml of 37 °C defrost media. The centrifugation step was repeated and the resultant pellet re-suspended in 1 ml of 37 °C R10 culture medium [450 ml RPMI (Sigma), 50 ml heat inactivated new born bovine serum (Sigma), 5 ml penicillin 5,000 units/ml -streptomycin 5 mg/ml (Sigma), 5 ml L-glutamine 200 mM (Sigma)]. Cells were counted on a haemocytometer and 500 μl of 2×10^6 cells were pipetted into cell culture plates (12-well Nunc Multidish Nunclon™). An additional 500 μl of R10 (i.e. media alone) or 500 μl of R10 plus 100 μg/ml poly(I:C) (Sigma) was added, and plates were incubated with 5% CO_2 at 37 °C. The poly(I:C) cell culture conditions were based upon a previous publication (166).

2.3.3.4 Gene expression

Gene expression analysis was conducted on PBMCs retrieved after 6 hours in cell culture. Non-adherent cells were retrieved by aspiration, and an additional 1 ml of R10 was added to the culture wells and a cell scraper (Corning) was used to dislodge adherent cells and this step was repeated. Retrieved cell suspensions were centrifuged for 15 minutes at 1800 rpm, and the cell pellet was re-suspended in RNAlater (Invitrogen). Total RNA was extracted using the RNeasy mini kit (Qiagen), following the extraction manual September 2010 'Purification of Total RNA from Animal Cells Using Spin Technology'. RNA was reverse-transcribed with superscript II™ reverse transcriptase (protocol MAN0001342 Rev. Date: 20 May 2010, Invitrogen) using random primers (Promega).

Quantitative PCR was performed on a MyiQ2 (Bio-Rad) system, 25 μl PCR reactions contained 12.5 μl SYBR green supermix (Bio-Rad), 1 μl of cDNA, 1 μl of forward primer (100 μM), 1 μl of reverse primer (100 μM) and 9.5 μl ddH20. Thermocycling settings were as follows; cycle 1 95 °C for 10 minutes, cycle 2 95 °C for 15 seconds followed by 60 °C for one minute repeated 40 times, cycle 3 was a melting curve analysis of an incremental increase of 0.5 °C every 30 seconds starting at 55 °C and ending at 95 °C. A set of nine immune response genes were assessed: *IL-6*,

CXCL10, *TNF*, *CD86*, *CD83*, *IFNA1*, *IFNB1*, *TRIF* and *TLR3* as well as two housekeeping genes, *ACTB* and *RN18S1* (appendix Table 7.7). Quantitative PCR amplification curves and melting curves for these primer sets are shown in appendix Figure 7.3 and appendix Figure 7.4.

2.3.3.5 Cell phenotyping and secretion of soluble proteins

Cell phenotyping and analysis of secreted proteins was undertaken on samples retrieved after 24 hours cell culture. Cell culture supernatant and non-adherent cells were retrieved by aspiration, this suspension was centrifuged for 15 minutes at 2000 rpm and the supernatant collected for the analysis of soluble proteins, the cellular pellet was re-suspended in 1 ml of R10. Cell culture wells were washed twice with 1 ml of R10 and adherent cells displaced with a cell scraper. Dislodged cells were then added to the non-adherent cell fraction and centrifuged at 2000 rpm for 15 minutes. The resultant pellet was washed twice in 10 ml PBS 2 mM EDTA, by centrifugation at 2000 rpm for 15 minutes. Cells were then re-suspended in 50 μl of PBS 2 mM EDTA and pipetted into a 96 well V-bottom plate (Nunc), 2 μl of anti-human CD19-APC antibody (Becton Dickinson), 1 μl of anti-human CD14-FITC antibody (eBiosciences) and 2 μl of Viaprobe-PerCP (Becton Dickinson), was added per well and cells were incubated in the dark at 4 °C for 30 minutes. The cells were then centrifuged at 2000 rpm for 2 minutes and washed twice in PBS 2 mM EDTA by centrifugation at 2000 rpm for 2 minutes. Next, cells were permeabilised by the addition of Fix/Perm (Becton Dickinson [BD]) for 20 minutes at 4 °C. These cells were then washed twice with 200 μl perm/wash (BD) by centrifugation at 2000 rpm for 2 minutes and re-suspended in 50 μl of perm/wash (BD) and 1 μl of anti-human TLR3-PE (eBiosciences) was added per well. Cells were finally incubated for 30 minutes at 4 °C and analysed on a FACSCalibur™ (Becton Dickinson).

Cell culture supernatant was assessed using a Cytometric Bead Array, for multiplexed soluble protein quantification (BD). A four-plex array was designed to evaluate IL-6 (cat.558276), TNF (cat. 558273), IP10 (cat. 558280) and IFN-α (cat. 560379). A 96 well V-bottom plate (Nunc) was pre-wetted with 0.2 μm filtered wash buffer (BD), then 50 μl mixed capture beads and 50 μl of neat cell culture supernatant was added, mixed at 500 rpm on a plate shaker in

the dark, and incubated at room temperature for one hour in the dark. Following this, 50 μl of mixed phycoerythrin (PE) detection reagent (BD) was added and mixed again at 500 rpm on a plate shaker in the dark. Samples were incubated at room temperature for 2 hours and then centrifuged at $200\times g_0$ for 5 minutes. The bead pellet was then re-suspended in 150 μl of 0.2 μm filtered wash buffer (BD) and 1200 events were collected on the FACSCalibur™, using settings optimised for these analyses.

2.3.4 Results

2.3.4.1 Sequencing

Sixty participants with stored PBMCs were genotyped at rs3775291 and rs3775290. Eight individuals with two copies of the C-C haplotype were selected for functional analysis, as well as eight control individuals with no copies of this haplotype, the latter group was comprised of seven individuals with two copies of the T-C haplotype and one individual with two copies of the C-T haplotype.

2.3.4.2 Gene expression

Fold-changes (FCs) in gene expression upon stimulation with poly(I:C) were determined using the $\Delta\Delta Ct$ method (equation 2.2). FCs are graphically illustrated as heat maps in Figure 2.12 and appendix Figure 7.5, normalised by housekeeping genes *RN18S1* and *ACTB*, respectively. Considerable interindividual differences in gene FCs were observed, but there was little sign of differences between haplotype groups. The median FCs were lower for carriers of the C-C haplotype for 8/9 and 9/9 genes, normalised to *ACTB* and *RN18S1*, respectively (appendix Table 7.9 and appendix Table 7.8). However, none of the gene FCs statistically differed between the groups.

$$Fold\ change = \frac{2^{\Delta Ct} target(culture\ only\ -\ polyIC\ treated)}{2^{\Delta Ct} housekeeping(culture\ only\ -\ polyIC\ treated)} \tag{2.2}$$

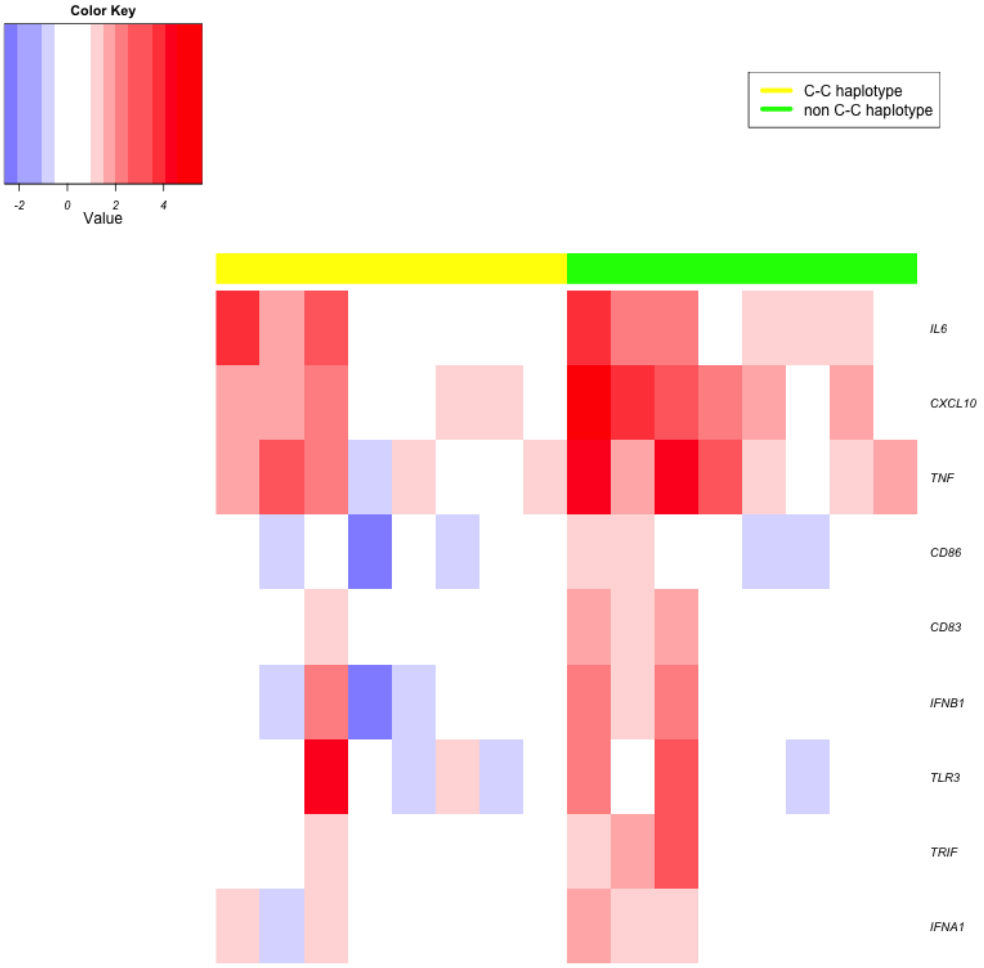


Figure 2.12: Heat map of gene expression following 6 hours of stimulation with poly(I:C) compared to media alone, normalised by *RN18S1* (18s rRNA). Each column is a participant and each row is a gene. Columns are arranged to separate samples based upon the rs3775291-rs3775290 haplotype. The yellow column refers to the C-C haplotype group (n=8) and the green columns to the non C-C haplotype group (n=8). The coloured boxes refer to the $\log_2$ normalised $\Delta\Delta Ct$ values, scaled from blue for downregulated genes, through white for no change, to red for upregulated genes.

2.3.4.3 Secreted soluble proteins

The results of the four-plex cytometric bead assay on supernatant from poly(I:C) and media only cultured PBMCs are shown in Figure 2.13. Culturing with poly(I:C) increased the concentration of pro-inflammatory cytokines IL-6 and TNF-α in the supernatant compared to cells cultured with media alone (Figure 2.13). However, no significant differences in media only subtracted poly(I:C) stimulated supernatant cytokine levels were observed between haplotype groups (Figure 2.13).

2.3.4.4 Phenotyping

The gating strategy for phenotyping is displayed in Figure 2.14. After 24 hours culture, cell viability was on average 85%, with no differences between culture conditions or haplotypes (Table 2.7). $CD14^+$ cells were seen to have a higher geometric mean index of total TLR3 expression (intracellular and surface), compared to bulk PBMCs or the other subsets assessed ($CD19^+$ and $CD14^-$) (Table 2.7). However, the level of expression of TLR3 in viable, $CD14^+$ cells was found to be similar between haplotypes, in samples cultured in media alone or poly(I:C) (Figure 2.15). Furthermore, no difference in TLR3 expression was seen between samples cultured in media alone or poly(I:C) (Figure 2.15).

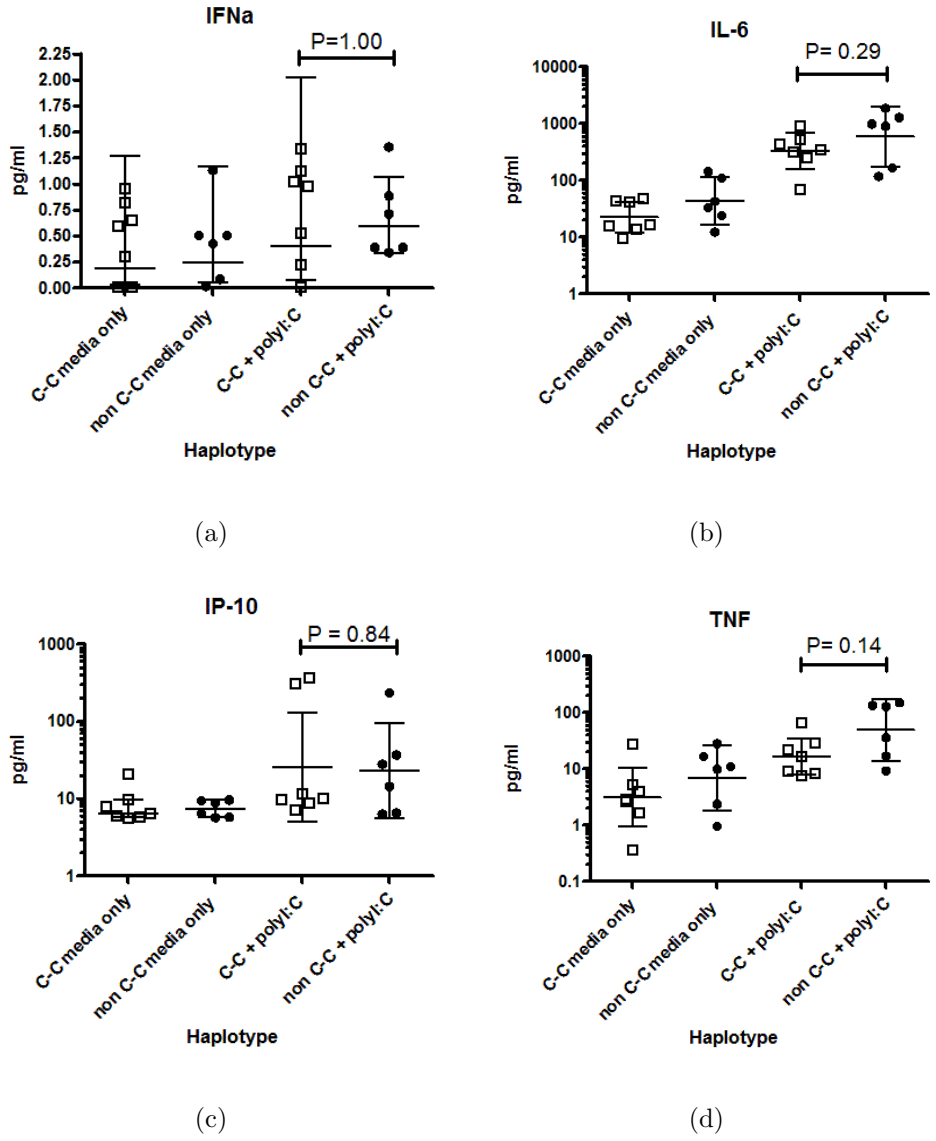


Figure 2.13: Concentrations of a panel of cytokines and chemokines secreted into the supernatant after 24 hours cell culture, separated by rs3775291-rs3775290 haplotype. The horizontal lines represent the geometric mean and 95% confidence interval. The unfilled squares represent PBMCs from individuals with two copies of the C-C haplotype, and the filled circles represent individuals with no copies of this haplotype (non C-C). Along the x-axis are the culture conditions and along the y-axis is the concentration of the respective analyte (a) IP-10, (b) IL-6, (c) IFN-alpha and (d) TNF. P = p-value, calculated using a Mann-Whitney test comparing media only subtracted poly(I:C) stimulated supernatant cytokine levels between haplotype groups. N=7 for C-C haplotype group and n=6 for the non C-C haplotype.

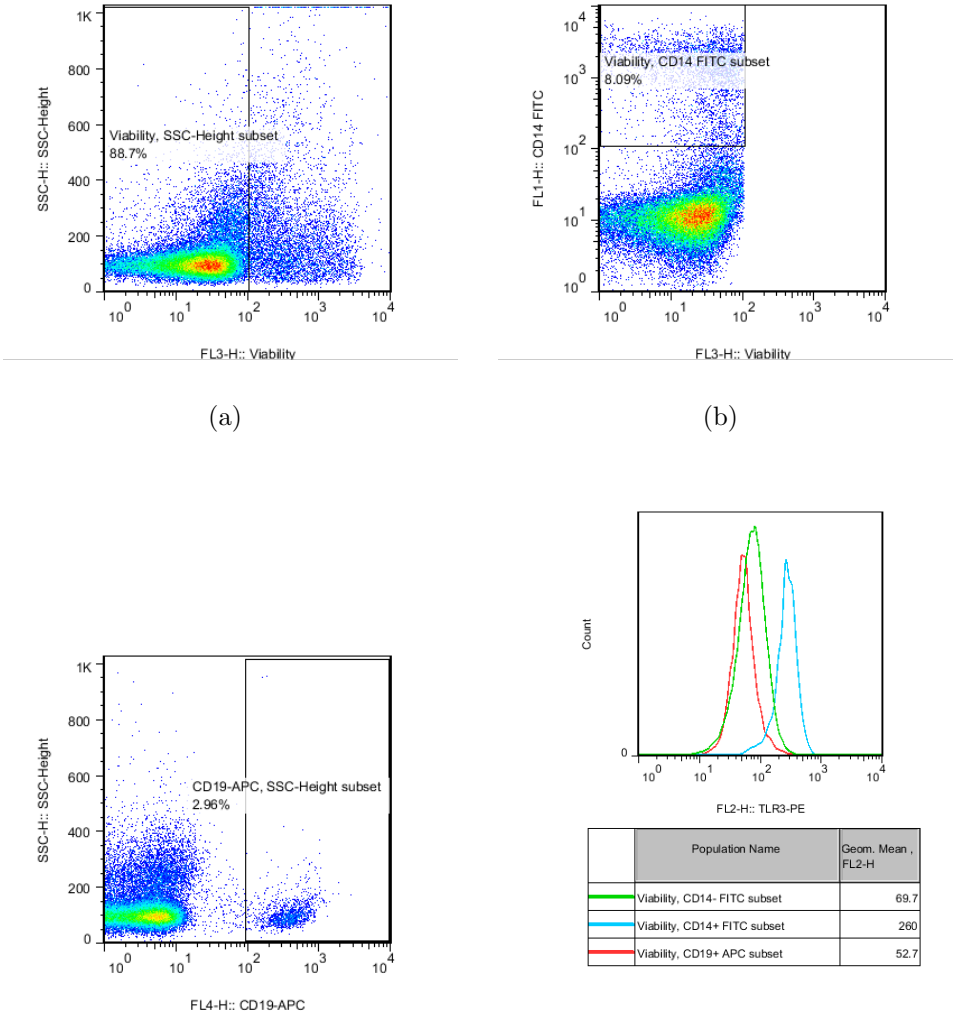


Figure 2.14: Gating strategy for the phenotypic analysis of PBMCs stimulated with poly(I:C). a) cells were gated for lack of staining with Via-probe, a nucleic acid dye that identifies nonviable cells, b) Viable cells were gated on $CD14^{+}$, c) Viable cells were gated on $CD19^{+}$, d) histogram of TLR3 geometric mean intensity of $CD19^{+}$, $CD14^{+}$ and $CD14^{-}$ populations

Table 2.7: Phenotypic data collected on peripheral blood mononuclear cells cultured for 24 hours with poly(I:C) or with media alone.

Sample	Haplotype	Viable %	Viable TLR3 GMI	$CD19^+$ %	$CD19^+$ TLR3 GMI	$CD14^-$ %	$CD14^-$ TLR3 GMI	$CD14^+$ %	$CD14^+$ TLR3 GMI
1 Media	CC	74.2	73.8	7.2	51.0	93.8	68.3	5.9	252
1 Poly(I:C)	CC	78.2	72.6	8.1	52.7	96.7	69.7	3.1	260
2 Media	CC	85.1	72.9	7.5	40.1	83.0	58.0	16.2	227
2 Poly(I:C)	CC	84.4	74.3	8.3	42.3	83.0	59.6	16.0	227
3 Media	CC	88.7	51.8	3.0	36.9	91.2	46.2	8.1	176
3 Poly(I:C)	CC	88.2	66.8	3.8	47.0	88.3	56.6	10.9	242
4 Media	CC	86.9	61.7	12.3	38.9	96.2	59.3	3.2	199
4 Poly(I:C)	CC	88.7	59.2	14.5	38.3	95.6	56.5	4.0	177
5 Media	CC	78.9	77.2	6.3	49.3	84.7	62.7	14.8	250
5 Poly(I:C)	CC	79.3	84.0	7.2	58.6	87.2	70.8	12.1	278
6 Media	CC	89.1	59.9	9.7	36.3	89.4	52.8	10.0	180
6 Poly(I:C)	CC	89.8	68.7	10.8	47.1	89.6	60.7	9.7	210
7 Media	CC	80.4	87.6	3.9	48.1	75.1	62.4	23.6	255
7 Poly(I:C)	CC	79.4	95.8	4.9	58.8	83.5	76.1	15.8	320
8 Media	TC	86.2	60.6	6.5	34.0	80.6	47.2	18.2	175
8 Poly(I:C)	TC	87.4	84.3	7.3	53.3	79.9	65.7	18.7	231
9 Media	TC	88.3	78.7	7.8	47.1	82.1	63.1	16.3	224
9 Poly(I:C)	TC	89.8	120.0	8.9	65.5	75	87.1	22.9	319
10 Media	TC	80.4	84.1	9.9	52.0	86.5	70.1	13.0	276
10 Poly(I:C)	TC	83.9	84.5	11.4	59.3	90.4	73.8	9.1	314
11 Media	CT	93.0	52.0	12.4	39.4	93.3	48.3	6.0	159
11 Poly(I:C)	CT	94.5	55.6	13.5	44.5	92.7	51.4	6.5	162
12 Media	TC	78.6	68.2	11.0	52.4	92.0	60.5	7.6	284
12 Poly(I:C)	TC	81.6	63.6	11.3	52.7	94.5	58.7	5.0	269
13 Media	TC	91.8	53.0	5.8	43.2	96.1	50.6	3.3	200
13 Poly(I:C)	TC	90.9	63.6	6.8	53.1	94.9	59.9	4.5	214
	Geometric mean		70.64		47.1		60.69		229.08
	95%CI		65.09-76.67		43.94-50.48		57.03-64.58		210.61-249.18

CC = C-C rs3775291-rs3775290 haplotype, TC or CT = non C-C rs3775291-rs3775290 haplotypes, CI= confidence interval, GMI = geometric mean index

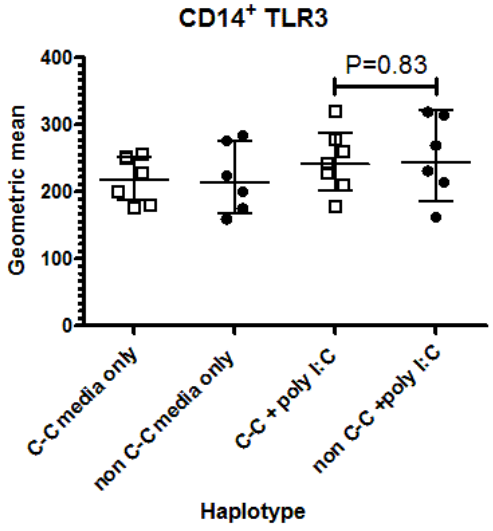


Figure 2.15: The geometric mean index of fluorescently labelled anti-TLR3 antibodies on viable, $CD14^+$ cells, following 24 hours cell culture. The horizontal lines represent the geometric means and 95% confidence intervals. The unfilled squares represent PBMCs from individuals with two copies the C-C haplotype, and the filled circles represent individuals without a copy of this haplotype (non C-C). Along the x-axis are the culture conditions and along the y-axis is the geometric mean fluorescence intensity for labelled TLR3. The p-value was calculated using a Mann-Whitney test comparing media only subtracted poly(I:C) stimulated supernatant cytokine levels haplotype groups

2.3.5 Discussion

There are an accumulating number of genetic variants being implicated in vaccine responses; however, few have been assigned any functional consequence. In this section a functional study was undertaken to assess whether the C-C haplotype (rs3775291-rs3775290) in *TLR3*, previously associated with the magnitude of MenC conjugate antibody responses in infancy, has a demonstrable effect on in vitro responses to poly(I:C). The median fold-change of 9/9 gene transcripts and 3/4 secreted proteins assessed were lower for individuals with the C-C haplotype following stimulation with poly(I:C). However, none of these differences were statistically significant. Cell phenotyping showed $CD14^+$ cells to have higher total TLR3 expression than total PBMCs, $CD19^+$ or $CD14^-$ subsets. However, the level of expression of TLR3 by viable, $CD14^+$ cells was similar between haplotypes and did not change following stimulation with poly(I:C).

There were a number of limitations to this study. Firstly, TLR3-independent responses to dsRNA have been demonstrated via MDA5 and RIG-I, and it has been suggested that TLR3 is largely redundant in responses to dsRNA in various leukocytes (164, 167, 168). Therefore, functional differences in TLR3 may be masked by TLR3-independent responses. Secondly, data from a previous study of PBMCs from an individual with two loss-of-function *TLR3* alleles suggested TLR3-dependent responses to poly(I:C) were time-dependent (164). The study detailed in this section may not have optimally captured TLR3-dependent differences due to the limited number of time points assessed. Thirdly, as the response measurements are downstream of receptor-agonist interactions these would be influenced by numerous downstream modifiers, including other genetic variants in *TLR3* and downstream factors. Fourthly, experimental noise may have been introduced by the freeze-thaw process. Although, total cell viability was high (median 86.6%, IQR 80.4-89.0%) following thawing and culture, this does not guarantee that all cell subsets equally survived the freeze-thaw cycle. Finally, the scope of this study was modest, both in terms of the number of participants and analytes (gene, proteins, cell phenotypes) assessed.

TLR3 is a pattern-recognition receptor (PRR) and member of a family of toll-like receptors

(TLRs) that recognise pathogen associated molecular patterns (PAMPs). Considerable redundancies between receptor-ligand interactions have been described for many PRRs (169, 170). For instance, a moderately high frequency of defect causing mutations in *TLR5* are seen in the population, without any overt clinical consequence (171). Nevertheless, TLR polymorphisms have been associated with a number of phenotypical traits, for example loss-of-function mutations in *TLR3* have been associated with herpes simplex encephalitis (HSE) (164, 171, 172). Furthermore, variants within intracellular TLR genes appear to be under purifying selection suggesting these are non-redundant, in an evolutionary context (173). Interestingly, while the loss of functional TLR3 has been associated with susceptibility to HSE, it has not been associated with the clinical course of a range of other viral infections (164, 172). In addition, variants in genes involved in downstream TLR signalling, such as *IRAK4* and *MYD88*, have been linked with relatively selective infectious susceptibilities (174, 175).

There are over 400 variants in *TLR3*; including 122 coding single-nucleotide variants, comprising of one frameshift, one nonsense and 68 missense mutations (`http://www.ncbi.nlm.nih.gov/SNP/`). Five *TLR3* variants within dbSNP have been cited in PubMed articles: rs121434431 (missense), rs5743316 (missense), rs3775291 (missense), rs5743311 (synonymous), rs3775290 (synonymous), with an additional PubMed cited coding 2236G>T (nonsense) mutation that is not currently present in dbSNP (164). The missense rs121434431 and nonsense 2236G>T mutations are rare loss-of-function mutations that have been associated with HSE (164, 172). The missense rs3775291 variant is seen at a high frequency (MAF ∼25%) in individuals of European and Asian ancestry, and has been associated with susceptibility to a number of infectious diseases as well as immune responses to MenC and measles vaccines (37, 38, 176, 177).

There are some data relating the non-synonymous rs3775291 variant within *TLR3* with variable responses to poly(I:C) in vitro; however, some of these data appear contradictory (176, 178, 179, 180). A study of PBMCs isolated from healthy volunteers, six major homozygotes and eight heterozygotes at rs3775291, stimulated with poly(I:C) for 6 hours found that heterozygotes had increased expression of CD69 and increased production of pro-inflammatory

cytokines, such as IL-6 and CCL3 (176). Conversely, a study of primary human lung fibroblast cell lines from patients with idiopathic pulmonary fibrosis; three major homozygotes, six heterozygotes and three minor homozygotes at rs3775291, stimulated with poly(I:C) for 24 hours found the expression of *TLR3*, *RANTES* and *IFN-β* transcripts to be higher in the major homozygotes (178). A further study of cord blood; 69 major homozygotes, 69 heterozygotes and 14 minor homozygotes at rs3775291, stimulated with poly(I:C) for 18 hours found IFN-γ production to be greater in major homozygotes (181). Furthermore, HEK293 transfected cell line reporter assays have also implied the minor allele results in partially defective TLR3 activation, following stimulation with poly(I:C) (179, 180). However, a further study of PBMCs from adults infected with herpes simplex virus-2, three major homozygotes and three minor homozygotes at rs3775291, stimulated with poly(I:C) did not see differences in secretion of type I or type II interferons between genotypes (182). There are a number of possible explanations for these disparities, importantly in some settings TLR3-independent mechanisms have been shown to compensate for even a functional loss of *TLR3* suggesting the role of TLR3 in responses to dsRNA are dependent on experimental conditions such as duration of stimulation (164).

2.3.6 Conclusion

No functional differences were described between the *TLR3* haplotypes assessed in this section. While the median fold-change of 9/9 gene transcripts and 3/4 secreted proteins assessed were lower for individuals with the *TLR3* C-C (rs3775291-rs3775290) haplotype following stimulation with poly(I:C), none of these relationships were statistically significant.

2.4 Candidate gene study of the relationship between a non-synonymous polymorphism in *TLR4* and childhood booster responses to pertussis vaccine

2.4.1 Background

2.4.1.1 Pertussis

Bordetella pertussis is a highly infectious Gram-negative bacterium responsible for the majority of cases of whooping cough (pertussis) (82). In adults and older children pertussis may result in mild symptoms such as catarrh and an irritating cough, which may progress into the characteristic 'whooping' cough often accompanied by vomiting (82). However, in young infants the characteristic whooping cough may never develop but coughing spasms may be followed by apnoea (82, 183). Severe complications and deaths occur most frequently in those under 6 months of age, with over 90% of infants with laboratory confirmed pertussis being admitted to hospital (183, 184). Pertussis cases remain infectious for 3 weeks after the onset of clinical symptoms and the infection can persist for 6-8 weeks even with administration of antibiotics (82, 184).

2.4.1.2 Pertussis disease epidemiology and vaccines

Prior to the introduction of the pertussis vaccine, in excess of 100,000 cases of pertussis were reported in the United Kingdom each year (Figure 2.16). Following the introduction of the whole-cell pertussis vaccine an inverse relationship between vaccine coverage and pertussis notifications was observed. In the late 70s and early 80s, following anxieties over the safety of the pertussis vaccine uptake declined and a resurgence of disease was observed (Figure 2.16) (184).

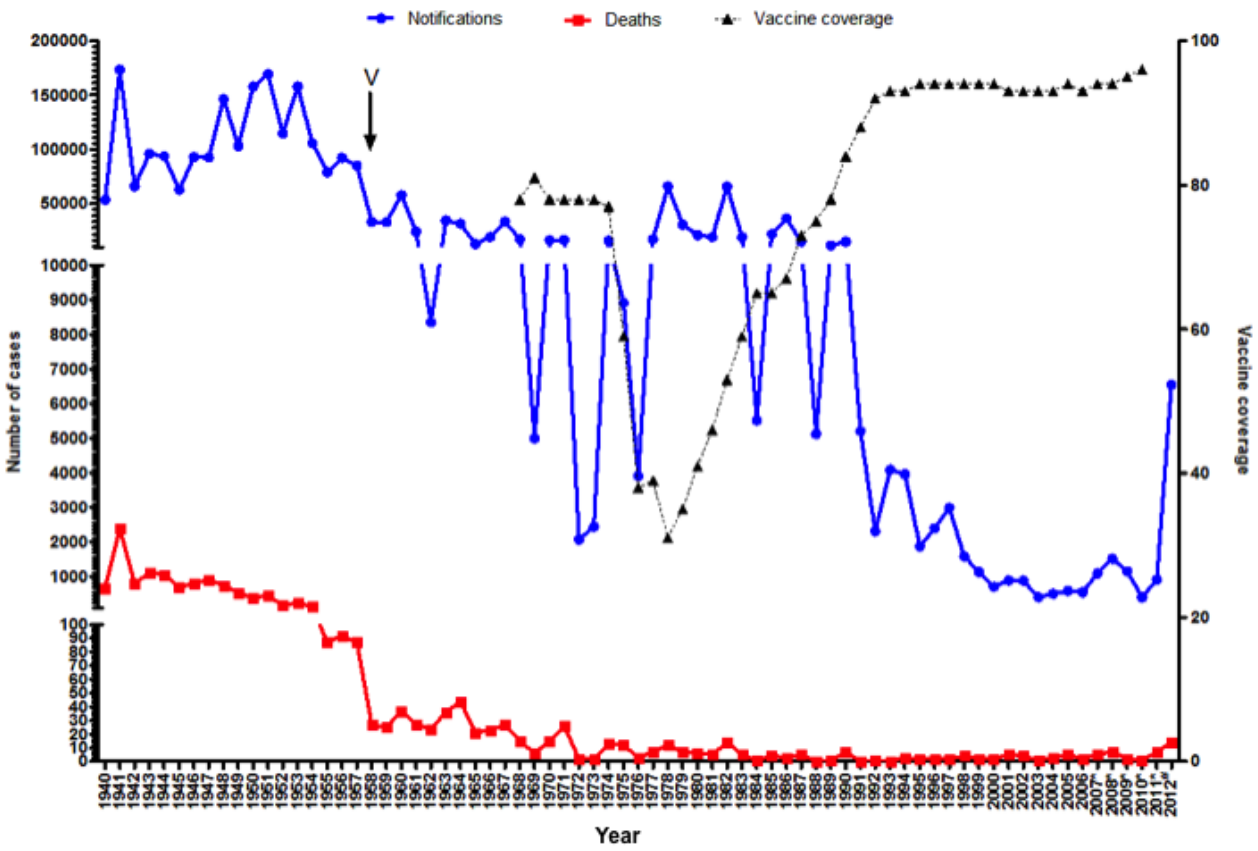


Figure 2.16: United Kingdom epidemiological data of pertussis notifications, deaths, and childhood (at 2 years of age) vaccine coverage (Data from Health Protection Agency, `http://www.hpa.org.uk`). V= pertussis vaccine introduced, 2007-2011* =enhanced diagnostic methods were introduced, 2012$^{\#}$ = provisional data.

The whole-cell vaccine introduced in 1958 was a suspension of killed organisms that caused considerable local and systemic reactions, and was associated with rare cases of infant febrile seizures and hypotonic hyporesponsive episodes (HHE) (185, 186). Estimates of whole-cell vaccine efficacy vary, but a study of US pre-school children (1-4 years of age) exposed to a household case of pertussis found infant immunisation to be 64%, 81% and 95% effective at preventing mild cough, paroxysmal cough and severe disease, respectively (187). Vaccine efficacy wanes following vaccination and is higher for preventing severe disease than infection (188).

The immunological correlates of protection against pertussis are unclear. High levels of

serum antibody capable of agglutinating *B. pertussis* have been associated with protection from disease following the whole-cell vaccine (189). Pertussis agglutinogens include pertactin (PRN) and fimbriae-2 and -3 (FIM) (stronger agglutinogens), as well as pertussis toxin (PT) and filamentous haemagglutinin adhesin (FHA) (weaker agglutinogens) (189). As *B. pertussis* is a mucosal infection, that causes disease without systemic invasion, serum antibody is seen as a surrogate measure of mucosal antibody (189). However, circulating anti-PT antibody neutralises this toxin and has a role in preventing severe disease (189). FHA is believed to facilitate binding of *B. pertussis* to the respiratory epithelium (190). PRN is thought to be involved in cell attachment and entry, and antibodies to PRN are produced following pertussis infection or the whole-cell vaccine (190). High titres of anti-FIM antibodies have been associated with protection against pertussis infection (190). Furthermore, the presence of serum anti-PRN, anti-FIM and anti-PT antibodies have been shown to correlate with protection from clinical disease in children exposed to household *B. pertussis* (191, 192).

Various acellular vaccines, with purified or recombinant pertussis antigens, were developed with the aim of producing an effective vaccine with a better reactogenicity profile than the whole-cell vaccine. A Cochrane review of the efficacy of acellular pertussis vaccines found formations with $\geq$3 components were more effective than those with 1 or 2 proteins (190). The whole-cell pertussis vaccine is still administered to most infants worldwide, but was replaced in the UK infant immunisation schedule by acellular vaccines in 2004 (184, 185). The acellular pertussis vaccines introduced into the UK immunisation program contained either 5 (PT, FIM-2, FIM-3, FHA and PRN) or 3 (FHA, PRN and PT) purified proteins, and have been shown to have similar efficacy in preventing culture-confirmed pertussis as the whole-cell vaccine but with lower local and systemic reactogenicity (193, 194, 195).

Current pertussis vaccines do not induce lifelong immunological protection against pertussis and reinfections occur even after natural infection (184). Vaccine mediated protection is estimated to last 4-12 years and infection acquired immunity wanes after 4-20 years (196). However, pertussis infections in vaccinated or previously infected individuals tend to be mild, and may often be subclinical (197, 198). Despite high vaccine coverage, intermittent epidemics

of pertussis occur every few years as population immunity wanes and transmission rate of residual infection increases, as the population is boosted this transmission rate decreases and this process is repeated in cycles (199).

Adults and older children are often the source of infection to infants who are too young to be fully vaccinated (184, 200). A pre-school pertussis vaccine was introduced in 2001 to boost immunity following infant doses at 2, 3, and 4 months of age, to reduce infection in older children as well as reduce transmission to unvaccinated young infants (184). In late 2011, and 2012 an increased incidence of pertussis was seen, particularly amongst young children (<3 months), older children (10-14 years) and older teens/adults (>15years) (Figure 2.17). In 2012 a temporary programme of immunisation of pregnant women (≥28 weeks gestation) was introduced in the UK, the aim of which was to provide transplacental antibody to protect infants prior to their routine immunisations as well as block mother to infant transmission (184).

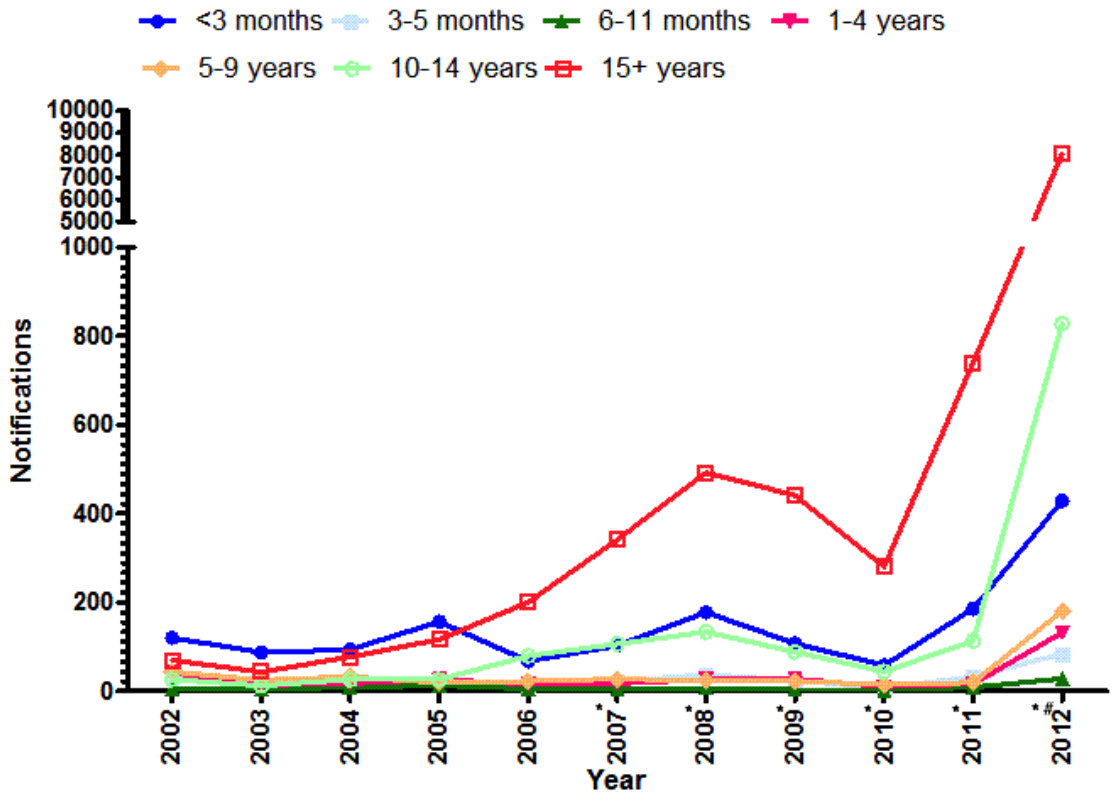


Figure 2.17: Notification of pertussis in the United Kingdom from 2000-2012 (Data from Health protection agency, `http://www.hpa.org.uk`).* =enhanced diagnostic methods were introduced in 2007, # = provisional data.

2.4.1.3 Genetic determinants of pertussis vaccine responses

Considerable variability in the magnitude and persistence of responses to pertussis vaccine has been observed, the determinants of which are largely unknown (201). However, a study of Gambian twins found the IFN-γ release assay to PRN and FHA following infant vaccination to be heritable, 53% (95% CI 35-67%) and 65% (95%CI 65-76%), respectively (7). Candidate gene studies of pertussis vaccine responses have found associations between polymorphisms in *TLR4* and associated downstream signalling molecules (39, 43, 44). TLR4 has been shown to recognise *B. pertussis* lipopolysaccharide (LPS) and PT, and a functional role for TLR4 in vaccine responses is supported by mouse data showing impaired immunity following whole-cell pertussis vaccine in $TLR4^{-/-}$ mice (202, 203). A recent study found an association between a non-synonymous SNP (rs4986790) in *TLR4* and the persistence of serological immunity following a booster dose of acellular pertussis vaccine administered to 74 adolescents (44). Heterozygous individuals, at rs4986790, had lower anti-PT IgG concentrations 3 years after vaccination, as well as poorer maintenance of anti-PT, anti-FHA and anti-PRN IgG 10 years after vaccination (44). Furthermore, studies have associated polymorphisms within the promoter of *TLR4* with responses to whole-cell pertussis vaccine responses, and *TLR4* transcript expression has also been associated with susceptibility to pertussis infection (43, 204).

2.4.2 Aim

The aim of this section was to assess the relationship between a non-synonymous polymorphism (rs4986790) in *TLR4* and childhood booster responses to acellular pertussis vaccine.

2.4.3 Materials and methods

2.4.3.1 Participants

DNA samples were available and informed consent obtained from parents/guardians of 82 children aged 3.5-5.0 years, enrolled in a study to assess the immunogenicity and reactogenicity of a pre-school booster dose of acellular pertussis vaccine. All participants had received whole-cell

pertussis vaccine as infants at 2, 3, and 4 months of age and were boosted with one of three acellular pertussis vaccines. Oxfordshire research and ethics committee (CO1.83) and Milton Keynes ethics committee (MKLREC/14/02) approved the original clinical trial. An Oxfordshire research ethics committee approved the genetics study detailed in this section (10/H0504/25).

2.4.3.2 Vaccines

Participants were randomised into three groups: to receive Td5aP-IPV (REPEVAX™, Aventis Pasteur MSD), Td5aP (COVAXIS™, Aventis Pasteur MSD), or DT2aP-IPV (TETRAVAC™, Aventis Pasteur MSD). Td5aP-IPV and Td5aP contained 2.5 μg of PT, 5 μg of FHA, 5 μg of FIM and 3 μg of PRN. DT2aP-IPV contained 25 μg of PT and 25 μg of FHA.

2.4.3.3 Immunological measures

Concentrations of anti-PT, anti-PRN, anti-FIM and anti-FHA IgG were measured pre-booster, as well as 4-6 weeks (28-42 days) and 3 years (32-39 months) after a booster dose of acellular pertussis vaccine. Aventis Pasteur laboratories (Toronto, Canada) performed ELISAs, using in-house reference standards and the serological results are published (205). The LLQ was 3 EU/ml for anti-FHA and anti-PRN, 17 EU/ml for FIM and 5 EU/ml for anti-PT. All values below the LLQ were set to an arbitrary value of 2.

2.4.3.4 DNA extraction

Genomic DNAs were extracted using the Maxwell® 16 System (Promega, Wisconsin), as described in subsection '2.2.3.4 DNA extraction'.

2.4.3.5 Genotyping

Genotyping was performed using the ARMS-PCR method described in subsection '2.2.3.6 Genotyping'. The primers used are detailed in appendix Table 7.10.

2.4.4 Statistical Analysis

2.4.4.1 Sample size calculation

Sample size calculations were conducted based upon data reported by Gröndahl-Yli-Hannuksel et al 2011 (44). GMCs and 95% CIs reported by Gröndahl-Yli-Hannuksel et al 2011, were used to calculate an estimate of SNP effect size. A Wilcoxon-Mann-Whitney test incorporating the estimated SNP effect size, with α-error= 0.05 and β-error = 0.2 (80% power) and ratio of major homozygous to heterozygous genotypes of 10:1, was conducted using G*Power version 3.1 (206). The results of the sample size calculation are shown in Figure 2.18, with >80% power achieved with a sample size of 86 (major 78 homozygotes and 8 heterozygotes).

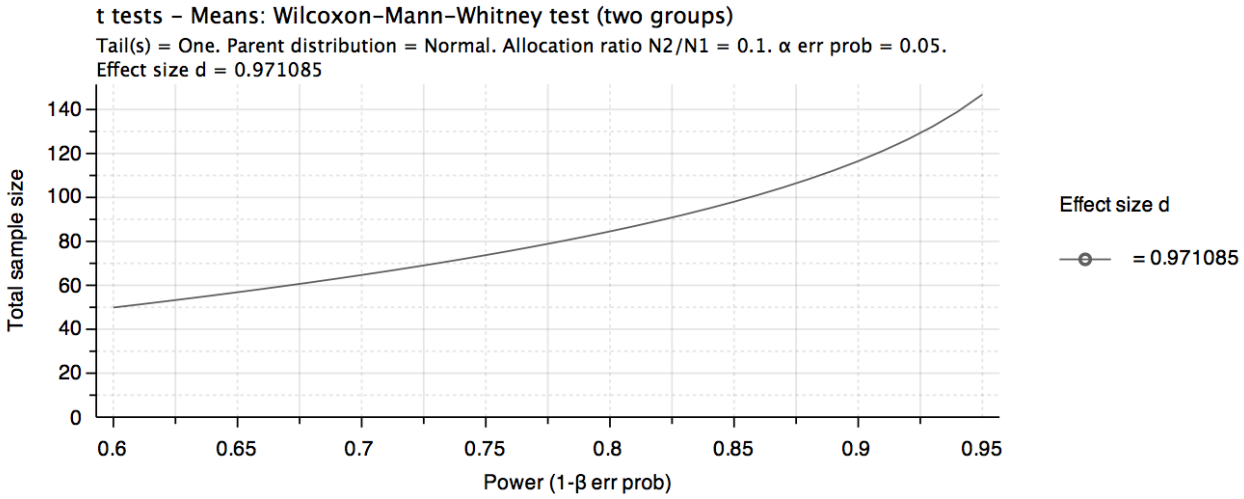


Figure 2.18: Statistical power against sample size for the association study between rs4986790 and booster responses to pertussis vaccine. The effect size parameter used in these calculations was derived from a previous study demonstrating an association between rs4986790 and booster responses to pertussis vaccine (44). Allocation ratio is the number of rs4986790 heterozygous individuals divided by homozygous individuals. This plot was produced using G*Power version 3.1 (206).

2.4.4.2 Association analysis

A Mann-Whitney test was used to compare pertussis serology between rs4986790 genotypes, using GraphPad Prism, version 5.0 (GraphPad Software). A p-value <0.05 was considered

statistically significant.

2.4.5 Results

2.4.5.1 Demographics

Table 2.8: Demographics of children in the candidate gene study of the genetic determinants of childhood booster responses to pertussis vaccine.

Demographic	Details	Value#
Ethnicity		
	Caucasian(%)	81 (99)
	Other (%)	1 (1)
Sex		
	Male (%)	39 (48)
	Female (%)	43 (52)
Pertussis vaccine		
	Td5aP-IPV (%)	21 (26)
	Td5aP (%)	32 (39)
	DT2aP-IPV (%)	29 (35)
Age		
	Days (IQR)	1410*(1353-1480)

= number of children unless specified,* = median, IQR= interquartile range

2.4.5.2 Genotyping

Genotypes at rs4986790 were called for all 82 vaccinees, 74 (90%) were major homozygotes, 8 (10%) heterozygotes and 0 were minor homozygotes.

2.4.5.3 Hardy-Weinberg equilibrium

No deviation from HWE was found for rs4986790 (P=1), calculated using the 'genetics' R package (152).

2.4.5.4 Pre-booster pertussis immunity

Pre-booster pertussis immunity measured by anti-PT, anti-PRN, anti-FIM and anti-FHA IgG concentrations were available on all 82 participants. Heterozygote individuals had higher

(p=0.03) pre-boost anti-FIM antibody concentrations, but no significant differences for any of the other serological measures were observed (Figure 2.19).

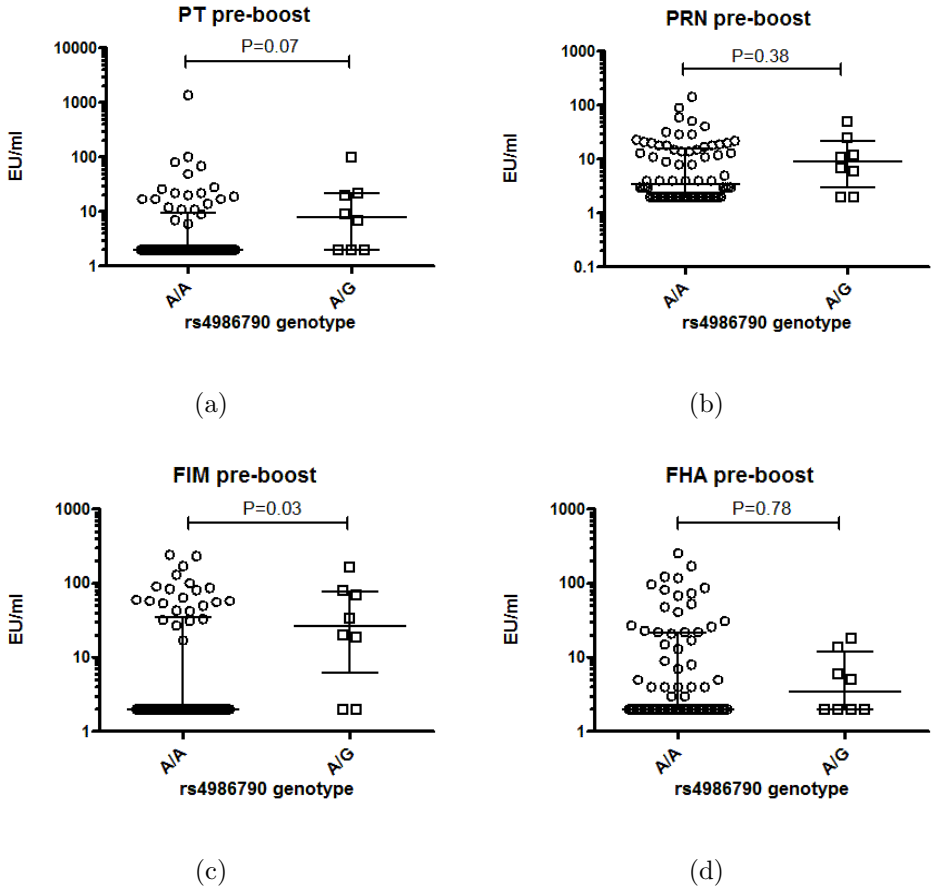


Figure 2.19: Level of pertussis immunity prior to administration of a booster dose of acellular pertussis vaccine to pre-school children, separated by rs4986790 genotype. a) anti-pertussis toxin (PT) IgG concentrations, b) anti-pertactin (PRN) IgG concentrations, c) anti-fimbriae-2 and anti-fimbriae-3 (FIM) IgG concentrations, and d) anti-filamentous haemagglutinin (FHA) IgG concentrations. P= p-value from a Mann-Whitney test between genotype groups. Summary lines represent the median and interquartile range.

2.4.5.5 Pertussis vaccine booster responses

Pertussis immunity 4-6 weeks after a booster dose of acellular pertussis vaccine was measured. No statistically significant differences were seen in the fold-rise (post-booster/pre-booster immunity) in pertussis-specific serological immunity between the rs4986790 genotype groups (Figure 2.20).

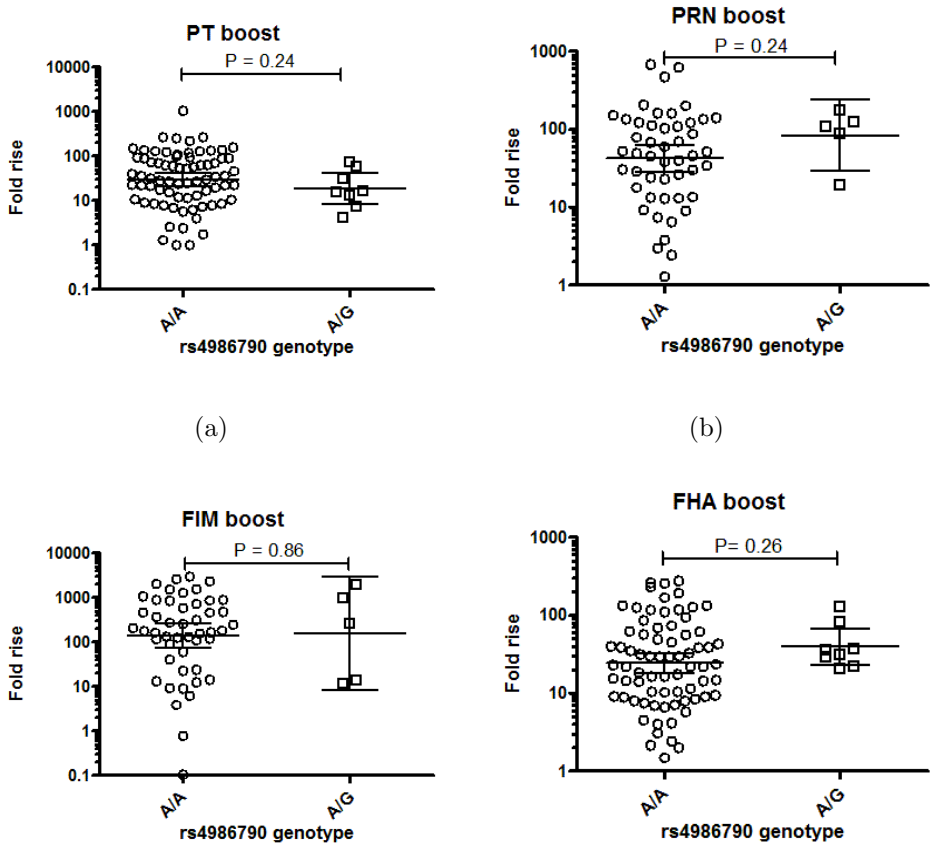


Figure 2.20: Fold-rise in pertussis immunity following the administration of a booster dose of acellular pertussis vaccine to pre-school children, separated by rs4986790 genotype. a) anti-pertussis toxin (PT) IgG fold-rise, b) anti-pertactin (PRN) IgG fold-rise, c) anti-fimbriae-2 and anti-fimbriae-3 (FIM) IgG fold-rise, and d) anti-filamentous haemagglutinin (FHA) IgG fold-rise. P= p-value from a Mann-Whitney test between genotype groups. Summary lines represent the geometric mean and associated 95% confidence intervals. Note: a third of participants received a booster pertussis vaccine that did not contain PRN or FIM, and these were omitted from PRN and FIM analyses.

2.4.5.6 Persistence of pertussis immunity following booster vaccine

Data on the persistence of anti-FIM, anti-PRN and anti-PT IgG 3 years after the booster dose of acellular pertussis vaccine was available for 69 (84%) participants and anti-FHA IgG persistence was measured on 67 (82%) participants. No statistically significant differences were seen in the fold-persistence (3 years post-booster / pre-booster immunity) of pertussis-specific serological immunity between the rs4986790 genotype groups (Figure 2.21).

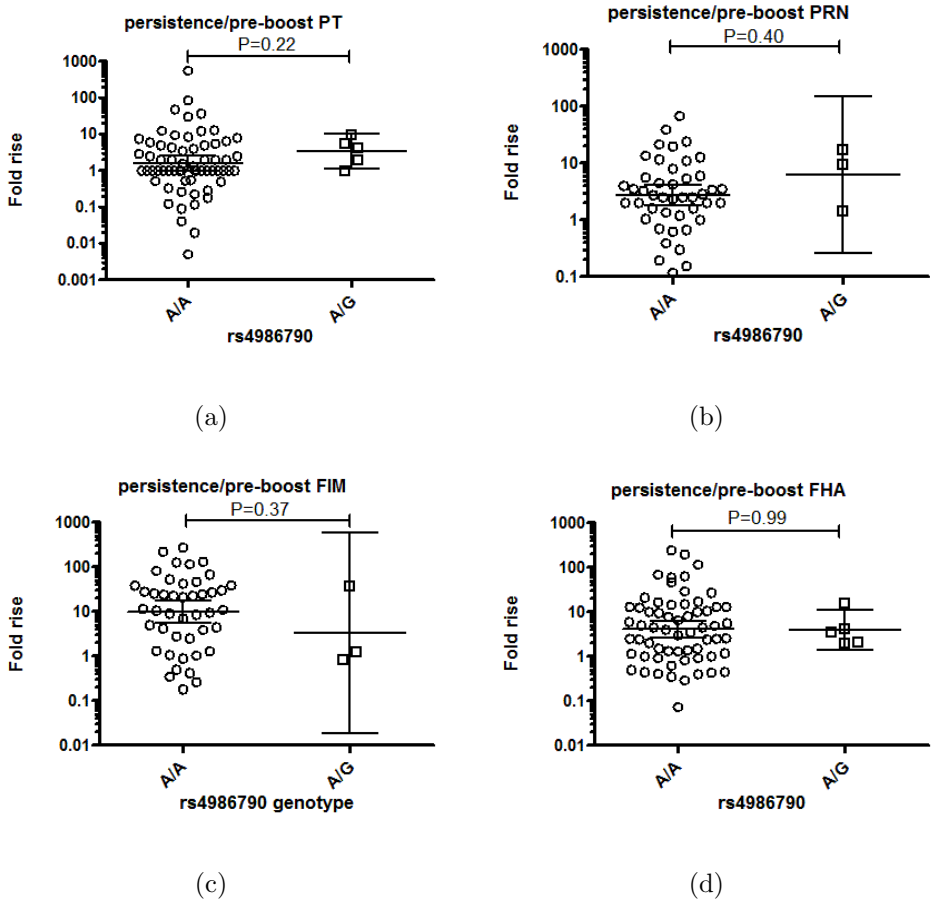


Figure 2.21: Fold-persistence of pertussis immunity 3 years after a pre-school booster dose of acellular pertussis vaccine, separated by rs4986790 genotype. a) anti-pertussis toxin (PT) IgG fold-persistence, b) anti-pertactin (PRN) IgG fold-persistence, c) anti-fimbriae-2 and anti-fimbriae-3 (FIM) IgG fold-persistence, and d) anti-filamentous haemagglutinin (FHA) IgG fold-persistence. P= p-value from a Mann-Whitney test between genotype groups. Summary lines represent the geometric mean and associated 95% confidence intervals. Note: a third of participants received the pertussis booster vaccine that did not contain PRN or FIM, and these were omitted from PRN and FIM analyses.

2.4.6 Discussion

Twin studies have shown immunological responses to pertussis vaccine to be heritable; however, the genetic determinants of these responses have not been well characterised (7). Immunological responses to vaccines are complex, but early innate responses orchestrate subsequent adaptive immunity. TLRs are innate recognition receptors that are known to contribute to early innate immune responses (207). TLRs present on the cell-surface such as TLR1, TLR2, TLR4, TLR5 and TLR6, are thought be particularly important in responses to bacterial ligands (208).

TLR4 has been implicated in responses to the whole-cell pertussis vaccine in both mice and humans (39, 43, 203). More recently, a common non-synonymous SNP (rs4986790) within *TLR4* has been associated with the persistence of antibody following adolescent booster with the acellular pertussis vaccine (44). In this section the relationship between rs4986790 and pertussis vaccine immunity was further explored in a cohort of pre-school children boosted with acellular vaccine. Although pre-boost anti-FIM IgG concentrations were significantly higher in heterozygous individuals, this study did not find evidence of an association between rs4986790 and the magnitude or persistence of vaccine-induced immunity following a booster dose of pertussis vaccine.

There were a number of limitations to this study. Firstly, the sample size calculations were based upon a previous study of Finnish adolescents following the low dose three protein acellular pertussis vaccine (dTaP). The disparities between these studies may result in inaccurate estimates of variance and SNP effect size, potentially resulting in an underpowered study. Secondly, analysis involving few SNPs is intrinsically limited in the evaluation of complex biological systems with redundancies and epistasis, unless variants under investigation have profound effects. Finally, as *B. pertussis* remains in circulation within the population causing epidemics every 3-4 years, the assessment of immunity may be confounded by natural boosting.

TLR4 has been shown to bind PT, and $TLR4^{-/-}$ mice have also been shown to have reduced responses to LPS-free acellular pertussis vaccine, highlighting a potential LPS-independent relationship between TLR4 and immunological responses to pertussis vaccine (202, 209). The non-synonymous SNP, rs4986790, results in the substitution of an aspartic acid with a glycine at amino acid position 299, which resides within the extracellular portion of the receptor. The minor allele at rs4986790 has been associated with susceptibility to Gram-negative bacterial infections and attenuated responses to LPS (210, 211).

A recent study found an association between rs4986790 genotype and the persistence of anti-pertussis antibody following adolescent booster with acellular pertussis vaccine, a finding that was not replicated in the study detailed in this section (44). The reasons for this discrepancy are unclear; however, there were a number of important differences between these studies,

in particular the age of vaccinees. A consistent finding between these studies was a lack of association between rs4986790 and the magnitude of serological response, measured 4-6 weeks after acellular pertussis vaccination (44). Furthermore, these data are consistent with a previous study of 515 one year olds that did not find a significant relationship between rs4986790 and anti-PT concentrations one month after a fourth dose of childhood whole-cell pertussis vaccine (43). While the data associating rs4986790 with the persistence of pertussis antibody following vaccination may imply that this SNP influences persistence of immunity but not the magnitude of immune response, this association has yet to be replicated in an independent population and the data presented in this section do not support this hypothesis.

2.4.7 Conclusion

In contrast to a previous study, no significant association between rs4986790 and the persistence of vaccine-induced serological immunity following a booster dose of pertussis vaccine was demonstrated. Further and more expansive studies, in larger numbers of individuals, are needed to scrutinise the relationship between genetic variants and the magnitude and persistence of pertussis vaccine-induced immunity.

2.5 Chapter summary

Previous data have shown immunological responses to a number of vaccines to be heritable, but few of the underlying genetic variants have been well characterised. Hitherto, the majority of the studies into the genetic determinants of vaccine responses have utilised a candidate gene approach, and few findings have been independently replicated or functionally validated. This chapter detailed two studies designed to study genetic variants that have previously been associated with MenC and pertussis vaccine-induced immunity. This chapter details support for a previous association between variants in *TLR3* and serological responses to MenC conjugate vaccine. However, this chapter also highlights many limitations in traditional candidate gene studies conducted using modest sample sizes.

Chapter 3

Genome-wide association study of capsular group C meningococcal vaccine immunity

3.1 Overview of chapter

The GWAS approach to assessing the genetic determinants of vaccine response has a number of potential benefits over the candidate gene approach detailed in the previous chapter (Chapter 2). Firstly, GWASs provide information on hundreds of thousands of genetic variants in a manner that does not rely upon existing knowledge of genes involved in a particular phenotype. In addition to the potential of this assumption-free approach to uncover novel biology, it may also avoid the intrinsic bias introduced during the selection of candidate genes. Secondly, the breadth of information collected from GWASs facilitates the assessment of gene-gene interactions and epistasis, and provides valuable quality control information to enable the identification of spurious data or outlier samples. Finally, the statistical concepts used in GWASs are typically more robust than those used in candidate genes studies, due to stringent adherence to rigorous corrections for multiple testing.

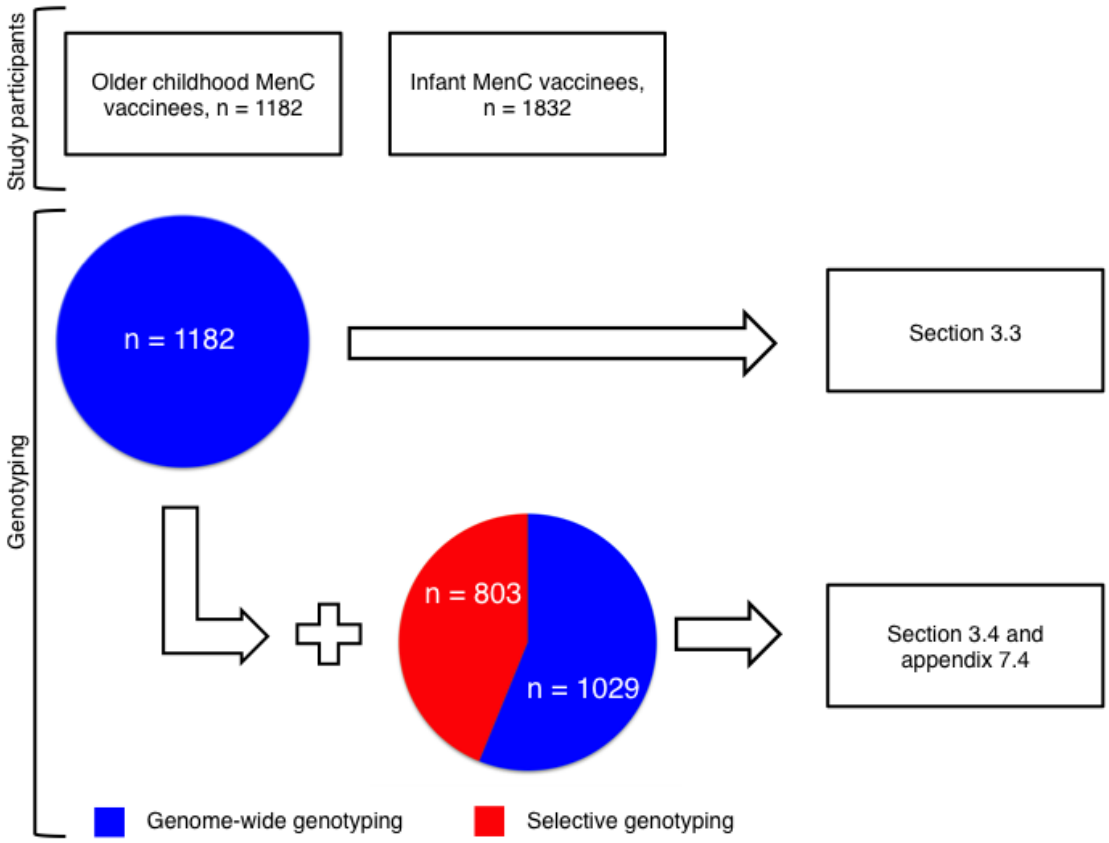


Figure 3.1: Schematic of participants included in a genome-wide association study to assess the genetic determinants of the persistence of capsular group C meningococci (MenC)-specific serological immunity following immunisation with MenC conjugate vaccine. This chapter includes 1182 and 1832 participants vaccinated with MenC conjugate vaccine as older children and infants, respectively. Initial analysis was conducted exclusively on individuals who received MenC conjugate vaccine as older children, all of which were genome-wide genotyped (section 3.3). Further analysis combined the older childhood and infant vaccinees, with a subset of the latter being selectively genotyped at markers found to have the greatest statistical association with the persistence of MenC-specific immunity from the genome-wide genotyping data (section 3.4 and appendix 7.4).

This chapter describes a GWAS of the genetic determinants of the persistence of capsular group C meningococci (MenC)-specific IgG concentrations and SBA titres following immunisation with MenC conjugate vaccine. Initially, the relationship between genetic variants and the persistence of MenC-specific serological immunity in a cohort of adolescents, vaccinated as older children (6-15 years of age) was explored (Figure 3.1). This GWAS was then expanded to include a cohort of individuals immunised with MenC vaccine as infants ($\leq$1 year of age) (Figure 3.1). This stratified approach was adopted as considerable differences in the persistence of post-vaccination serological measures of MenC-specific immunity have been observed be-

tween younger and older children, presumably relating to immune maturation (212, 213, 214). Consequently, it was not known whether the genetic factors influencing the maintenance of MenC-specific immunity are age-dependent. This chapter concludes with the description of a second stage study that assessed a subset of variants identified in the GWAS in an additional independent population of MenC vaccine recipients.

3.1.1 Chapter aims

i) Conduct a genome-wide association study of the persistence of serological immunity to capsular group C meningococcal vaccine following older childhood immunisation

ii) Describe a two-stage genetic association study of the persistence of serological immunity to capsular group C meningococcal vaccine following childhood immunisation

3.2 Introduction to contemporary genome-wide association study methodologies

3.2.1 Human reference genome and catalogue of human variation

The human reference genome was first published in 2001, enabling the systematic identification and mapping of human genes, and serving as a scaffold for sequence assembly (215, 216, 217). The human genome has considerable influence on many physiological and disease traits, but locating causal trait loci is complicated by the vastness and diversity of the human genome (108). Furthermore, the cost of whole-genome sequencing large numbers of individuals to assess genetic markers associated with complex phenotypic traits has, to date, been prohibitive in the majority of study settings. However, the International HapMap project, and more recently the 1000 genomes project, have extensively documented correlations between alleles at proximal polymorphic loci. This information, along with advances to microarray technology have made it feasible to test the majority of common variants in large study cohorts (108, 218). Consequently, genome-wide association studies have successfully produced over 1300 publications,

documenting genetic variants associated (p-value $<10^{-7}$) with a range of traits (219, 220).

3.2.2 Phasing and imputation

3.2.2.1 Phasing

Phasing is the process of determining on which of the two copies of a particular chromosome (in a diploid organism) an allele resides. A haplotype is a combination of adjacent alleles on a chromosome. Information on haplotype phase is used during genotype imputation and can also be used in haplotype based association analysis (221, 222). Experimental approaches of deciphering phase are considerably more expensive and time-consuming than contemporary high-throughput genotyping techniques; therefore, phase is more often inferred statistically.

A number of phasing methodologies have been developed that have their own advantages and disadvantages. IMPUTEv2 (`https://mathgen.stats.ox.ac.uk/impute/impute_v2.html`) and MACH (`http://www.sph.umich.edu/csg/abecasis/MACH/tour/imputation.html`), use a hidden Markov model (HMM) to model haplotypes of $\mathbf{G}$ (vector of genotypes at M markers for an individual) as an imperfect mosaic of haplotypes in $\mathbf{H}$ (set of $\mathbf{K}$ known haplotypes from a sample of individuals at a set of $\mathbf{M}$ markers). Sampling suitable haplotypes for $\mathbf{G}$ using a forward-backward algorithm for HMM has quadratic complexity, $O(MK^2)$. To alleviate this complexity IMPUTEv2 and MACH take a subset of N haplotypes (223, 224). In BEAGLE (`http://faculty.washington.edu/browning/beagle/beagle.html`), $\mathbf{K}$ haplotypes in $\mathbf{H}$ are represented in a graph, the edges of which are taken as states of the HMM. As complexity does not increase at the rate of $\mathbf{K}$ this controls the computational demands (225). FastPHASE (`http://stephenslab.uchicago.edu/software.html`) is based upon a cluster model for haplotypes. A HMM is used with state space involving a set of parameterised core states estimated as the model is fitted. The parsimonious number of states controls for computational complexity (155). SHAPEIT (`https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html`) uses HMM calculations carried out on a graph $\mathbf{H_g}$ of collapsed $\mathbf{K}$ haplotypes. This approach also uses a method of sampling compatible haplotypes for G with linear complexity (226). SHAPEIT2 combines SHAPEIT1 and IMPUTEv2, to purport-

edly enhance performance (221). In brief, this method utilises the SHAPEIT1 Markov model, representing space of haplotypes consistent with a given individuals genotypes across a whole chromosome. Transition probabilities are estimated by applying the IMPUTEv2 'surrogate family' phasing approach in local windows of **W**, with each window **K** informative haplotypes are chosen to update the transition probabilities. Therefore, this method generalises the IMPUTEv2 method to be applied over whole chromosomes with linear scaling (221).

3.2.2.2 Imputation

Genotype imputation is a technique to use information from a densely genotyped reference panel to infer unobserved genotypes in a less densely genotyped, but genetically similar, study population (227). Imputation can offer a number of advantages in genetic association studies. Firstly, imputing SNPs expands the set of SNPs that can be tested for an association with the study phenotype, increasing statistical power (228). Secondly, imputation enables meta-analysis of non-overlapping markers genotyped on different platforms. Thirdly, the increase in marker density following imputation aids fine-mapping association signals to causal variants. Finally, sporadically missing genotypes can also be imputed, increasing non-missing data.

Several imputation methods have been proposed and important considerations when weighing up the benefits of each of these techniques include; 1) the accuracy with which haplotypes are phased at typed SNPs will determine how well they match haplotypes in the reference panel, 2) accounting for unknown phase at typed SNPs is computationally expensive, 3) accuracy will increase with n, meaning algorithms than include several reference panels will be advantageous. Most imputation methods adopt a two-stage design, starting with an initial step to predict untyped SNPs without taking into account phenotypic information. Secondly, imputed genotypes are tested, incorporating an uncertainty value, for an association with the phenotype of interest. This approach has the advantage of not requiring imputation to be repeated for every phenotypic trait.

IMPUTEv1 uses a HMM of an individual's vector of typed SNPs, conditional on the set of reference haplotypes and specified parameters. IMPUTEv2, uses related methodology to

IMPUTEv1, but estimates haplotypes at typed SNPs using a HMM and then imputes alleles at untyped SNPs conditional on current estimated haplotypes, with Markov chain Monte Carlo (MCMC) iterations at the stages accounting for phase uncertainty. The fastPHASE approach is a method of estimating haplotypes and imputing SNPs that is incorporated within the BIMBAM software framework (`http://www.haplotype.org/bimbam.html`). This method capitalises upon the observation that haplotypes cluster with related or similar haplotypes. The HMM modelling used by fastPHASE differs from IMPUTEv1 in that states are represented by clusters rather than reference haplotypes; an expectation-maximisation (EM) algorithm is used to fit the model and impute missing genotypes conditional on the parameter estimates using a forward-backward algorithm. MACH uses a similar HMM approach as IMPUTEv1, iteratively updating the phase of each individual's genotype data conditional upon all haplotype estimates. BEAGLE utilises a graphical model of haplotype sets, iteratively fitting the set of estimated haplotypes and resampling new estimates for individuals based upon the model fit; missing haplotypes are derived from the model fitted to the final iteration.

The majority of imputation methods that utilise a HMM attempt to estimate missing genotypes and integrate over unknown phase of typed SNPs simultaneously. On the contrary, the IMPUTEv2 algorithm separates the phasing and imputing tasks, with more computational effort on phasing and maximising the information being incorporated in this step. IMPUTEv2 attempts to improve imputation accuracy at untyped SNPs by improving phasing accuracy at typed SNPs. Haplotype 'guesses' are then assumed to be true, removing the complexity of computing uncertainty at this stage. The increase in computational requirement for the phasing step is leveraged against quick haploid imputation. This algorithm can also handle multiple large and unphased next generation datasets, the use of which increases accuracy and helps with analysis of rare SNPs.

In this report IMPUTEv2 was the imputation methodology of choice, based upon literature showing its superior accuracy (223, 229, 230). Furthermore, a pre-phasing step in SHAPEIT2 was implemented as this has been shown to outperform other approaches in terms of switch error rate and mean distance to error, with computational comparability (221).

3.3 Genome-wide association study of the persistence of serological immunity to capsular group C meningococcal vaccine following older childhood immunisation

3.3.1 Background

In 1999, the MenC vaccination campaign to immunise those <19 years of age was initiated in the UK in response to increasing rates of group C disease, which accounted for 35-40% of disease isolates at the time (184). The peak incidence of MenC disease was seen in those under 1 year of age, but a second peak was seen in those aged 15-19 years (184) (Figure 3.2). Furthermore, teenagers and young adults represented an important reservoir of meningococci, with the highest rate (∼25%) of *N. meningitidis* carriage in those aged 15-19 years (184). Therefore, it was hoped that this vaccination programme would also benefit unvaccinated individuals by reducing carriage and interrupting transmission of this organism.

During the 1999-2001 MenC vaccination campaign, the MenC conjugate vaccine was introduced as a single-dose for those aged 1-18 years and incorporated into the routine infant immunisation schedule as a three-dose regimen (184). A year after the introduction of the MenC vaccine more than 70% of the target population had been vaccinated, with the vaccine showing 88% and 96% efficacy in preventing disease in children and adolescents, respectively (137). In addition, the MenC vaccine showed 75% effectiveness against capsular group C carriage and indirect protection of unvaccinated individuals, presumably via herd immunity (231, 232). The efficacy of the MenC vaccine in preventing disease in young children, who make modest serological responses that wane rapidly, may have been largely the result of vaccine efficacy against carriage in older children and teenagers (231).

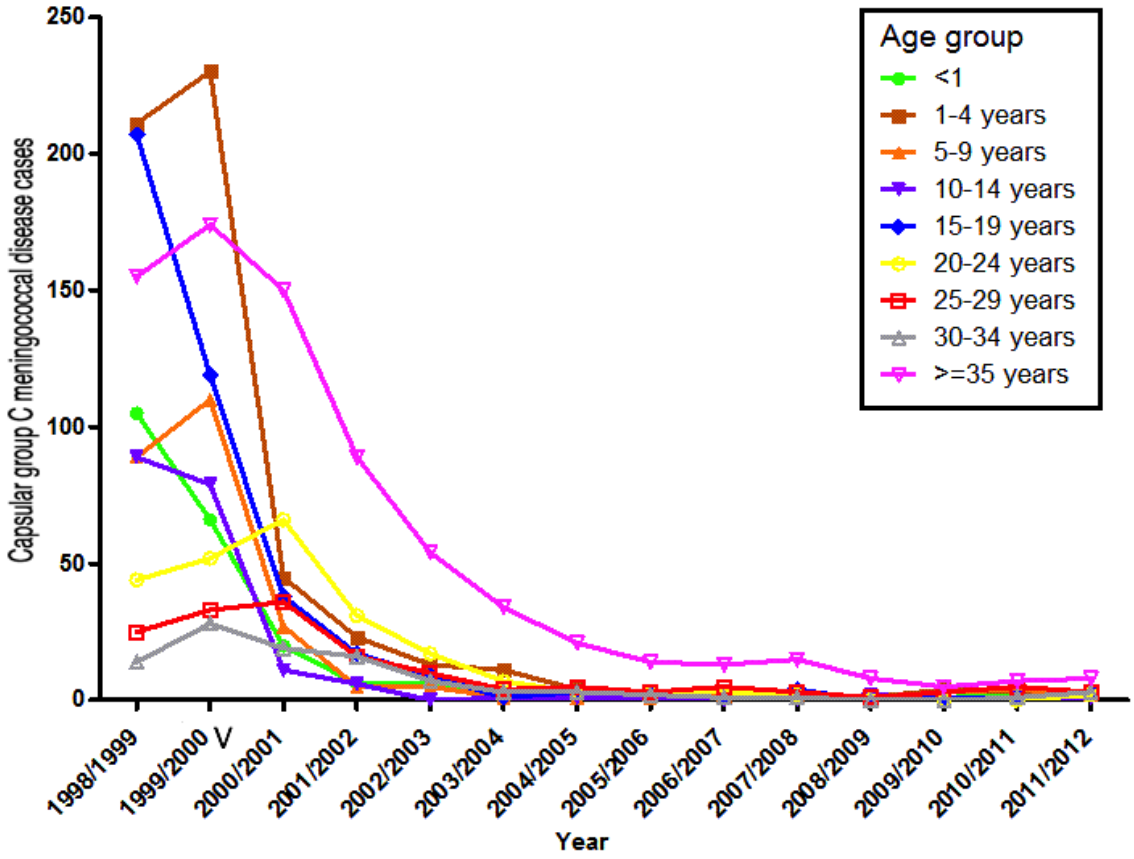


Figure 3.2: The number of cases of capsular group C meningococcal disease by age group, in the United Kingdom (UK) in years 1998-2012 (Data from Health protection agency). V= the capsular group C meningococcal conjugate vaccine was introduced into the routine UK childhood immunisation schedule in November 1999.

Adolescents immunised during the MenC vaccination campaign showed sustained levels of functional antibody to MenC for >5 years after vaccination (214). However, the persistence of functional antibody was variable, and >15% of 11-20 year olds had lost detectable levels of functional antibody 5 years after vaccination (214). Age at vaccination has been shown to be an important factor in the persistence of serological immunity and efficacy, with children immunised with MenC vaccine in their second decade maintaining functional antibody better than younger children (212, 214). However, age at vaccination does not account for all the variation observed, and the mechanisms behind the maintenance of serological immunity are not well understood (233).

Vaccinee genetic factors are likely to account for some the variation seen in the maintenance

of vaccine-induced protection against MenC disease and carriage. The genetic factors influencing carriage of MenC have not been investigated, and the low rates of MenC carriage, even prior to the introduction of the MenC vaccine (0.45% in 15-17 year olds), may be prohibitive (234). The genetic determinants influencing the persistence of MenC-specific antibody levels have been investigated using a candidate gene approach, genotyping 768 SNPs in 122 candidate genes in 905 adolescents 3.8-6.3 years after vaccination (38). In this section a genome-wide genotyping approach was utilised to further explore the genetic factors influencing the persistence of vaccine-induced MenC-specific immunity in adolescents immunised as older children.

3.3.2 Aim

The aim of this section was to conduct a genome-wide association study of the persistence of serological immunity to capsular group C meningococcal vaccine following older childhood immunisation.

3.3.3 Materials and methods

3.3.3.1 Adolescent participants

Healthy adolescents aged 11-20 years old, who had received a single dose of MenC vaccine as part of the UK vaccination campaign of 1999-2001, were recruited from Buckinghamshire and Oxfordshire. Informed consent was obtained from parents/guardians of adolescents under 16 years of age and participants over 16 years of age provided consent. Participants were recruited as part of two trials conducted to assess the persistence of immunity following MenC immunisation (213, 214). The first of these trials (hereinafter referred to as 'cohort 1') examined the persistence of serological immunity to MenC in adolescents aged 13-15 years, more than 3 years after immunisation with the MenC conjugate vaccine (213). A further trial (hereinafter referred to as 'cohort 2') assessed the persistence of serological immunity to MenC in adolescents aged 11-20 years, immunised between 6 and 15 years of age with MenC conjugate vaccine (214). An Oxfordshire research ethics committee approved these studies (approval number C02.328).

An Oxfordshire research ethics committee also approved the genetic study detailed in this section (10/H0504/25). In total 1182 participants were eligible for this study. Nine hundred and five of these adolescents were previously described in a candidate gene study of the persistence of MenC-specific antibody levels following MenC conjugate vaccine (38).

3.3.3.2 Vaccines

Participants' immunisation records were obtained from the child health computer department. Participants received one of three vaccines administered during the MenC vaccine campaign: two of which contained MenC capsular polysaccharide conjugated to CRM_{197} and the other was conjugated to tetanus toxoid. These vaccines were Menjugate (Novartis Vaccines and Diagnostics, Siena), containing 10 μg of MenC polysaccharide conjugated to 12.5-25.0 μg CRM_{197} and adjuvanted with 0.3-0.4 μg aluminium hydroxide (Al^{3+}), Meningitec (Wyeth Vaccines, New York) containing 10 μg of MenC polysaccharide conjugated to 15 μg CRM_{197} and adjuvanted with 0.125 μg aluminium phosphate (Al^{3+}), and NeisVac-C (Baxter Vaccines, Maryland) containing 10 μg of MenC polysaccharide conjugated to 10-20 μg tetanus toxoid and adjuvanted with 0.5 μg aluminium hydroxide (Al^{3+}).

3.3.3.3 Immunological measurements

The Vaccine Evaluation Unit (VEU), Public Health England (Manchester), measured MenC-specific rSBA titres using rabbit complement and Novartis Vaccines (Marburg, Germany) performed MenC-specific hSBA titres using exogenous human complement as described in subsection '2.2.3.3 Immunological measurements'. MenC-specific IgG concentrations were measured by ELISA, in-house or by Novartis Vaccines. The in-house assay was described in Chapter 2 subsection '2.2.3.3 Immunological measurements'. The ELISAs completed by Novartis Vaccines used a standardised assay with a LLQ of 0.3 μg/mL. For the purposes of analysis IgG concentrations below the LLQ were assigned of half the respective assay LLQ. The LLQ for the hSBA assays was a reciprocal titre of 4 and for the rSBA assay this was 8, sera without detectable bactericidal antibody were assigned an arbitrary value of two. All serological values were $\log_{10}$

transformed prior to statistical analysis.

3.3.3.4 DNA extraction

DNA was extracted from whole blood samples collected in EDTA tubes, using the GeneCatcher™ gDNA kits (Invitrogen™, California, United States), according to standard human blood protocol. DNA concentrations were quantified using Quant-IT PicoGreen® dsDNA Assay Kit (Invitrogen™, California, United States) and diluted to 50 ng/μl for genotyping.

3.3.3.5 Genotyping

Genotyping was performed at the Genome Institute of Singapore using routine laboratory protocols, based on manufacturer's specifications (Illumina®, California, United States). In brief, DNA samples were whole-genome amplified and enzymatically cleaved into 300-600 bp fragments. Samples were then purified using isopropyl precipitation and re-suspended in hybridisation buffer. Samples were hybridised to an Illumina® OmniExpress12v1 microarray and an allele-specific enzymatic single-base extension step was performed, to add a biotin-labelled (guanine or cytosine) or dinitrophenol-labelled (adenine or thymine) base. Sequential rounds of staining were performed with green-fluorescent streptavidin and biotin-labelled antibodies, and red-fluorescent anti-dinitrophenol antibodies and dinitrophenol-labelled antibodies. Following staining, the arrays were read on an Illumina® scanner and genotypes were called using Illumina® BeadStudio software.

3.3.3.6 Quality control

A number of quality control (QC) checks were undertaken on the genotyping data before statistical association testing, the QC pipeline utilised was similar to that described by Anderson et al 2010, using the PLINK whole genome association analysis toolset (`http://pngu.mgh.harvard.edu/~purcell/plink/`) (156, 235). The first step was to identify any discrepancies between the sexes reported on participant information data and those deduced from the genotyping data. The homozygosity rate of X chromosome markers was used to assign genotypic

sex, using the '--check-sex' PLINK command (156). Next, sample outliers based upon number of missing genotypes and/or genome-wide heterozygosity rates were identified, using the '--missing' and '--het' PLINK commands, respectively (156). This was followed by the identification of related individuals. Relatedness was calculated by first removing areas of the genome with high linkage disequilibrium (LD), based upon regions published by Anderson et al 2010, and then conducting pairwise identity-by-descent (IBD) analysis, using the '--genome' PLINK command (156, 235). Next, individuals of divergent ancestry were identified by assessing the underlying structure of the genotype data by multi-dimensional scaling (MDS) on the IBD matrix, using the '--cluster' PLINK command (156).

QC was also conducted on array markers. Missing genotype rates were calculated for samples passing the previously described QC thresholds. Marker minor allele frequency (MAF) and Hardy-Weinberg equilibrium (HWE) p-values were calculated, using '--maf' and '--hwe' PLINK commands, respectively (156). The thresholds imposed for these QC measures are presented in the results subsection '3.3.4.3 Quality control'. Post-association QC was also conducted to check for inflation of test-statistics, using the 'GenABEL' R package, under the assumption that the vast majority of the genetic markers assessed would not be associated with the phenotype (236). Violation of this assumption can be indicative of population stratification and it is routine to control for the inflation factor when performing association testing (237).

3.3.3.7 Quantitative trait loci analysis

A linear regression model was used to assess the relationship between SNP allele dosage and the persistence of $\log_{10}$ transformed MenC-specific IgG concentrations, with age at vaccination, time since vaccination and the first three MDS components as covariates. A linear regression model was also used to assess the relationship between SNP allele dosage and the persistence of $\log_{10}$ transformed MenC-specific SBA titres, with SBA complement source, age at vaccination, time since vaccination and the first three MDS components as covariates. Furthermore, as there was a strong relationship between the complement source and SBA titres, linear regression analyses were also performed separately for the rSBA and hSBA subsets and meta-analysed.

Linear regression analysis of genotype data and MenC-specific immunity was conducted using PLINK (156).

3.3.3.8 Imputation

Imputation was performed to increase the breadth of genotype coverage, the theory of which was described in subsection '3.2.2 Phasing and imputation'. In brief, strand information corresponding to the genotype data were updated to build 37 of the human reference genome, using the orientation tool in GTOOL and the Illumina OmniExpress12v1 strand file provided by Will Rayner (`http://www.well.ox.ac.uk/~wrayner/strand/`). Segmented HAPlotype Estimation and Imputation Tool (SHAPEIT) version 2 was used to pre-phase each chromosome of the genotype data individually (221). Imputation was performed using IMPUTE version 2 on 5 Mb chunks of this pre-phased data, applying the 1000 Genomes Phase I integrated variant set release (March 2012) as the variant reference set (223). SNPTEST version 2.4.0 was utilised to test for allelic associations between imputed SNPs and the persistence of MenC-specific serological immunity, incorporating the covariates previously mentioned in subsection '3.3.3.7 Quantitative trait loci analysis' (230). Imputation and subsequent association analysis was performed across multiple cores of the CLUSTER-3 computing resource at the Wellcome Trust Centre for Human Genetics (WTCHG).

3.3.3.9 Power calculations

Power calculations were undertaken under varying effect and sample size parameters. Calculations were based upon a linear regression model $Y_i = a + bX_i + \epsilon_i$; where Y denotes a dependent variable (MenC-specific SBA titre or IgG concentration), X denotes an independent variable (SNP allele dosage) and ϵ are normally distributed random errors. The null hypothesis was specified as H0: $b - b_0 = 0$ and the alternative hypothesis H1: $b - b_0 \neq 0$. Power simulation were performed with various SNP effect size estimates, using the R package 'PWR'. (Figure 3.3) (238). A sample size of 1182 was shown to give >80% power to detect the slope $b_0 = 0.23$ (corresponding to a 1.7 fold-change in MenC-specific IgG concentrations

per allele) with $\alpha < 5{\times}10^{-8}$ (Figure 3.3). The α level for power calculations was based upon the assumption of 1,000,000 independent tests (equal to the estimated number of independent regions in the Caucasian genome), which is likely to be conservative level of statistical correction (29). The effect size (f^2) was derived from the partial R^2, as described in the G*Power manual (`http://www.gpower.hhu.de/en.html`). In brief, $R^2 = p^2$ from $b = p \cdot \sigma_y/\sigma_x$, implying $p^2 = (b \cdot \sigma_x/\sigma_y)^2$. Where σ_y denotes the standard deviation of $\log_{10}$ transformed MenC-specific IgG concentrations and σ_x denotes the standard deviation of allele counts for a theoretical SNP, genotyped in a 1000 individuals, with equal frequencies of homozygous genotypes. The parameters used in the power calculation were as follows, $\sigma_y = 0.78$, $\sigma_x = 0.71$, $b = 0.23$, $\alpha = 5 \times 10^{-8}$, n=1182 and the number of predictors = 6 (genotype, age at vaccine, time since vaccination and MDS components 1-3).

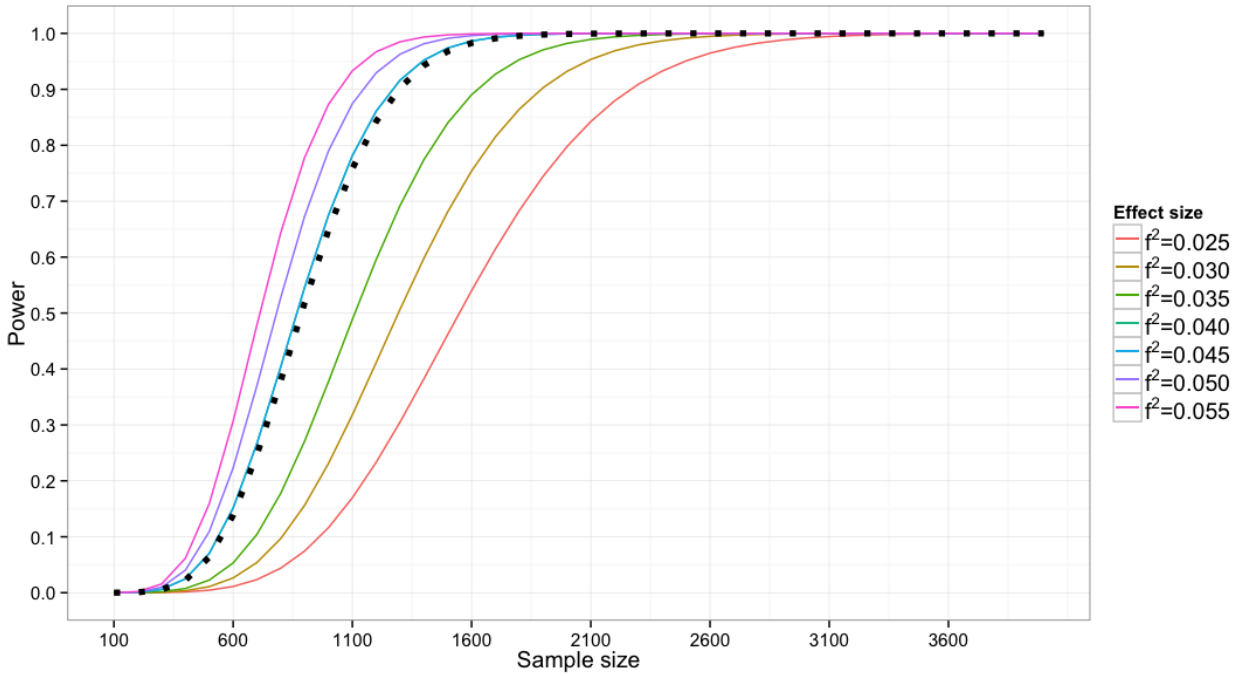


Figure 3.3: Power calculations for the genome-wide association study of the persistence of capsular group C meningococci (MenC)-specific immunity following older childhood immunisation with MenC conjugate vaccine. Calculations were based upon a linear regression model $Y_i = a + bX_i + \epsilon_i$; where Y = MenC-specific immunity, X = allele dosage and ϵ = errors, and performed using the R package 'PWR' (238). The null hypothesis was specified as H0: $b - b_0 = 0$ and the alternative hypothesis H1: $b - b_0 \neq 0$. Each line represents an effect size (f^2) in the simulated power calculations, the black dots representing an effect size corresponding to the slope (b_0) = 0.23 (1.7 fold-change in MenC-specific IgG concentration) with allele dose. The fixed parameters were α=$5{\times}10^{-8}$ and number of predictors =6.

3.3.4 Results

3.3.4.1 Adolescent demographics

An overview of participant demography is displayed in Table 3.1. An independent 2-group Mann-Whitney U Test found Cohort 1 to be older at MenC vaccination ($P<0.05$), with fewer days since immunisation ($P<0.05$).

Table 3.1: Summary of participant demographics in the study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following older childhood immunisation with MenC vaccine. Study cohorts were described in subsection '3.3.3.1 Adolescent participants'

Study	Male/Female	Time since vaccination days* (IQR)	Age at vaccination days* (IQR)	Reference
Cohort 1	122/140	1380.5 (1279.3 - 1423.0)	3832.5 (3633.5 - 4073.3)	(213)
Cohort 2	449/378	1777 (1679 - 1856)	3514 (3038 - 4078)	(214)

IQR= interquartile range, * = median.

3.3.4.2 Serological results

The serological data presented here were generated from two studies conducted to assess the persistence of MenC-specific serological immunity following older childhood immunisation with a single-dose of MenC conjugate vaccine (213, 214). SBAs were performed using either exogenous human (cohort 1, n=262) or rabbit (cohort 2, n=828) complement. The density distribution of $\log_{10}$ reciprocal SBA titres performed using rabbit complement was to the right of the corresponding distribution when exogenous human complement was used, displaying the relative intrinsic bactericidal activity of rabbit complement (Figure 3.4) (239). However, the distributions of SBA titres, around their respective means, were comparable between the two complement sources. Furthermore, both distributions displayed enrichment of values at the left tail, reflecting serum without detectable bactericidal activity (Figure 3.4). For analytical purposes it was assumed that $\log_{10}$ transformed reciprocal SBA titres were sampled from a normal distribution.

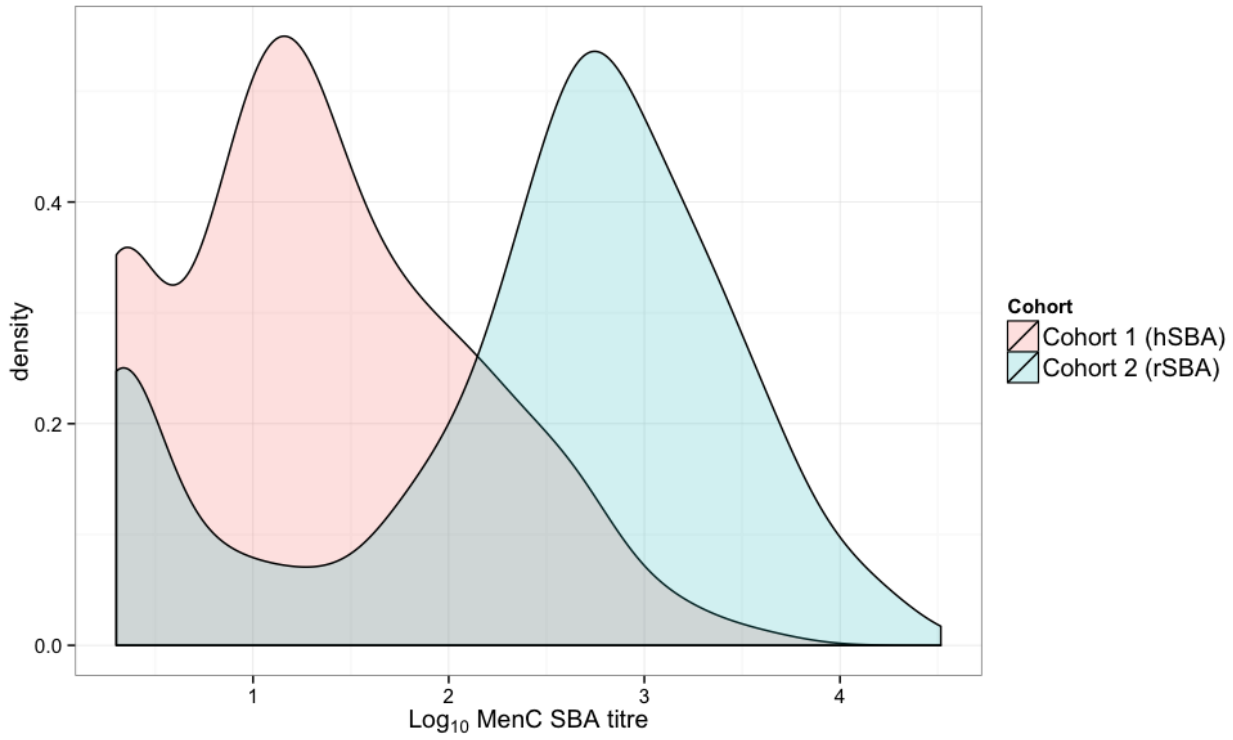


Figure 3.4: Density plots of the persistence of $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following older childhood MenC vaccination. Data are separated by study of origin; hSBA= human complement used in SBA, rSBA= rabbit complement used in SBA. Cohort 1 n=262 and cohort 2 n=828. Density plots were created from a 1d kernel density estimate, using the R package 'ggplot2' (240).

The distribution of $\log_{10}$ transformed MenC-specific IgG concentrations approached the normal distribution, with a slight positive skew (Figure 3.5). Similar ELISA protocols were used to measure MenC-specific IgG concentrations in both study cohorts; however, assays were conducted in different laboratories, as described in subsection '3.3.3.3 Immunological measurements'.

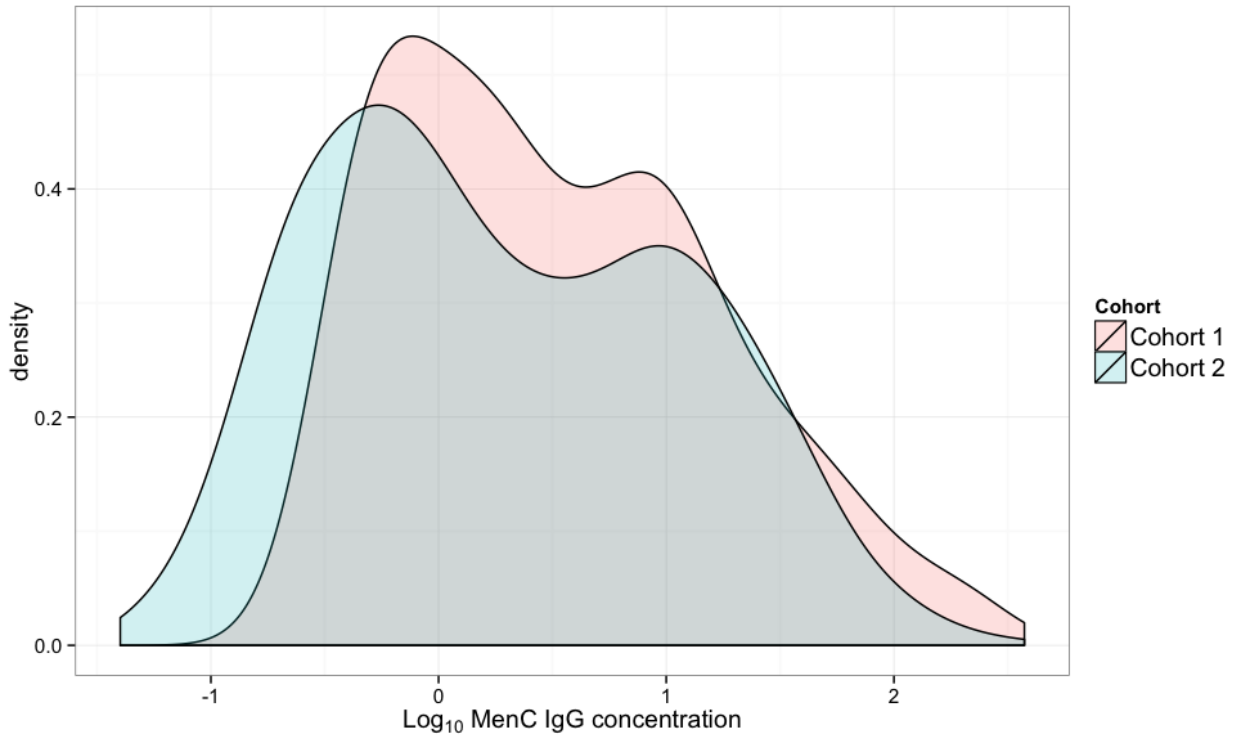


Figure 3.5: Density plots of the persistence of $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following older childhood MenC vaccination. Data are separated by study of origin. Cohort 1 n =270 and cohort 2 n =856. Density plots were created from a 1d kernel density estimate, using the R package 'ggplot2' (240).

3.3.4.3 Quality control

Following DNA quantification, 1126 (99%) of the DNA samples were found to be of adequate quality (>50 ng/μl) for genome-wide genotyping. One participant did not have sex recorded on participant information data and 66 samples were found to have discrepancies between sexes reported in participant information and genotype data. As DNA samples were anonymised it was not possible to troubleshoot these disparities by referring to case report forms, therefore these participants were eliminated from subsequent analysis.

The proportions of missing genotypes and heterozygosity rates for participants are shown in Figure 3.6. The cut-off for acceptable rate of genotype heterozygosity was set to the mean heterozygosity rate +/− two standard deviations, 30 samples failed this QC metric. None of the samples failed QC due to the proportion of missing genotypes, the acceptable cut-off was set to ≤0.03, as described by Anderson et al 2010 (235).

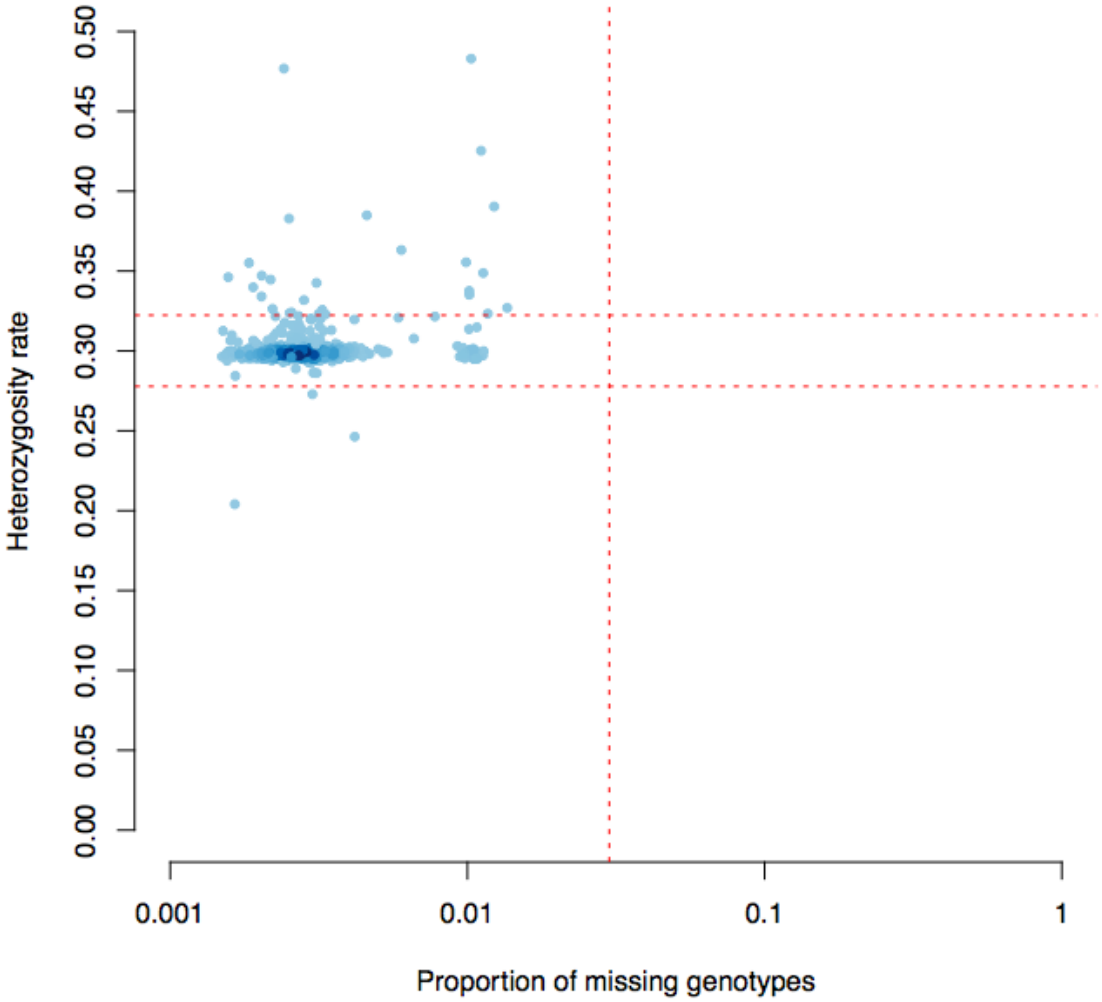
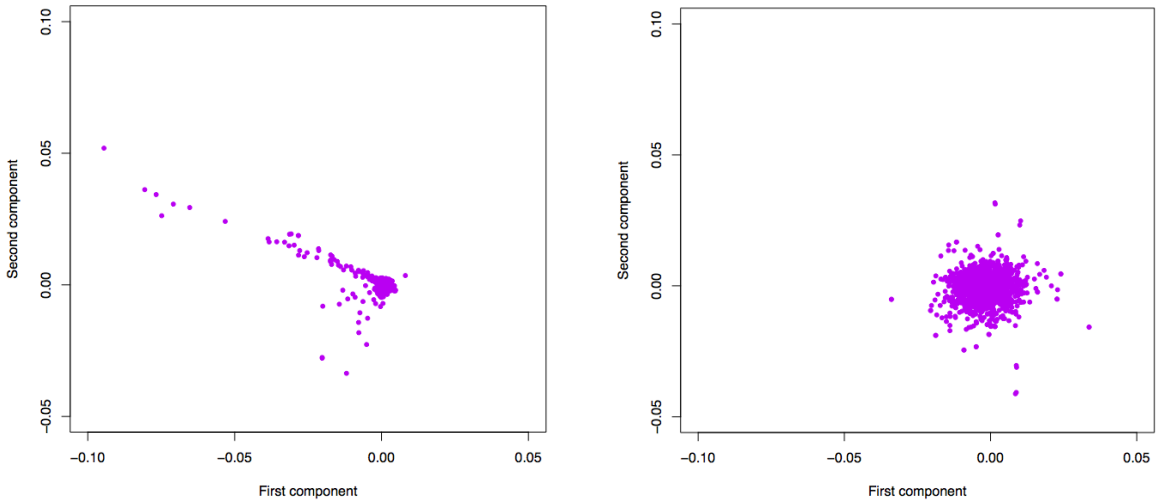


Figure 3.6: Plot of heterozygosity rate compared to the proportion of missing genotypes in the study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following older childhood immunisation with MenC vaccine. Shading is used to highlight coordinates that are densely populated. The horizontal lines represent the mean heterozygosity rate +/− two standard deviations and the vertical line represents a proportion of missing genotypes of 0.03.

Identity-by-descent (IBD) analysis identified 117 full-siblings or monozygotic twins within the genotyped cohort; only the participant with the lowest missing genotype proportion from related participant sets were included in further analysis, resulting in the exclusion of 83 samples. MDS was performed on the IBD matrix and the first two MDS components are shown in Figure 3.7a. Thirty-six outlier samples were eliminated based upon apparent divergent ancestry and MDS was repeated (Figure 3.7b).



(a) (b)

Figure 3.7: The first and second components of multi-dimensional scaling analysis of the identity-by-descent matrix of genotype data in the study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following older childhood immunisation with MenC vaccine. a) all participants, b) filtered to exclude participants with apparent divergent ancestry.

A histogram of the rate of missing genotypes is shown in Figure 3.8. The acceptance cut-off for missing data was set at ≤0.03, resulting in the elimination of 2743 (0.4%) markers. SNPs with a minor allele frequency of <0.01 and/or HWE p-value < 1×10^{-5} were also eliminated, leaving 631987 (87%) markers.

3.3.4.4 Association analysis

Following quality control 1011 individuals and 631,987 markers remained for association analysis. Quantile-quantile (Q-Q) plots were used to visualise the distribution of association p-values, compared to a null distribution (Figure 3.9 and Figure 3.10). The inflation factor (λ) of the distribution of p-values was estimated by regression of observed onto expected p-values, using the 'estlambda' function in the 'GenABEL' R package (236).

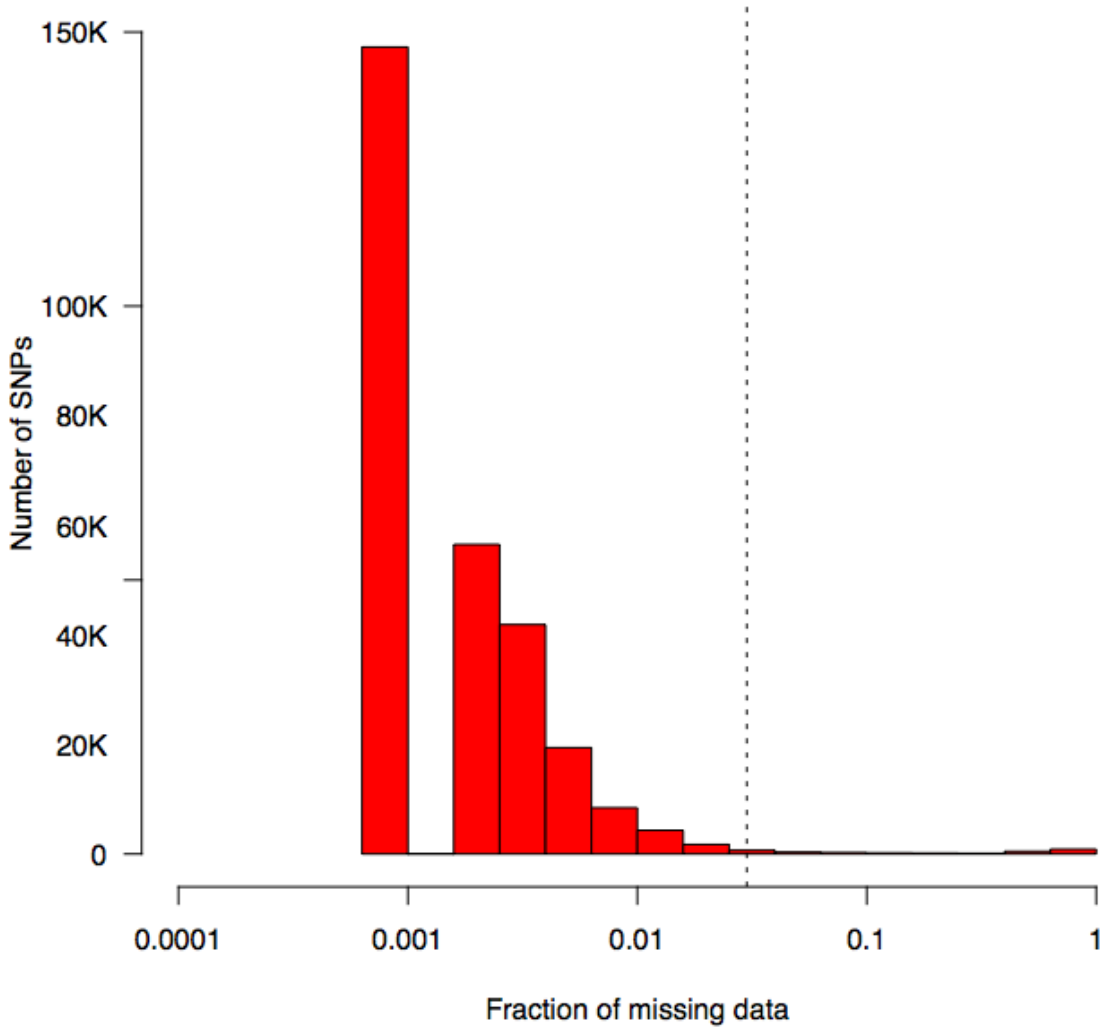


Figure 3.8: Histogram of the rate of missing genotypes from individuals that passed sample based quality control metrics, in the genome-wide association study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following older childhood immunisation with MenC vaccine. The dashed vertical line denotes the threshold (3%) were genetic markers were removed from further analysis due to excessive failure rate.

The Q-Q plot for association p-values resulting from the regression analysis between SNP allele dose and $\log_{10}$ transformed MenC-specific IgG concentrations did not show signs of excessive inflation, and had an inflation value (λ) of 1.0257 (Figure 3.9). Some enrichment of smaller p-values ($<1\times10^{-4}$) was observed; however, these did not deviate from the 95% confidence intervals of the null distribution (Figure 3.9).

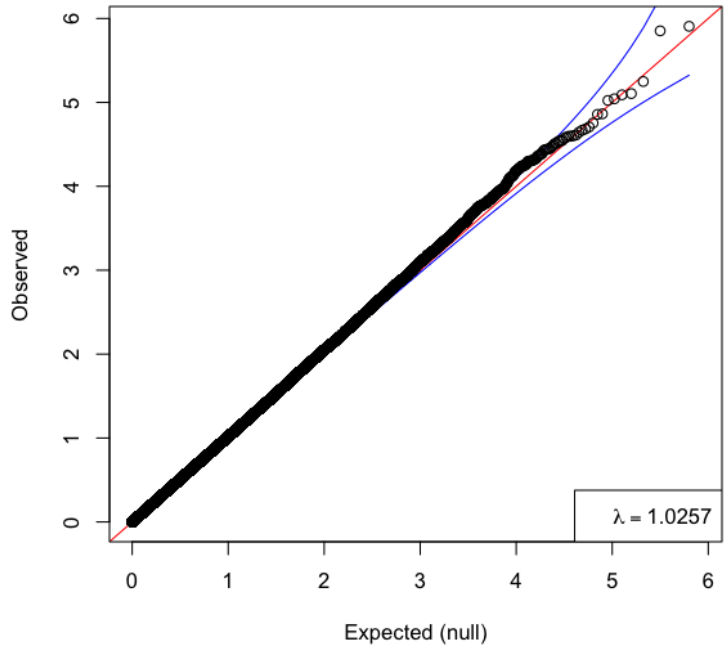
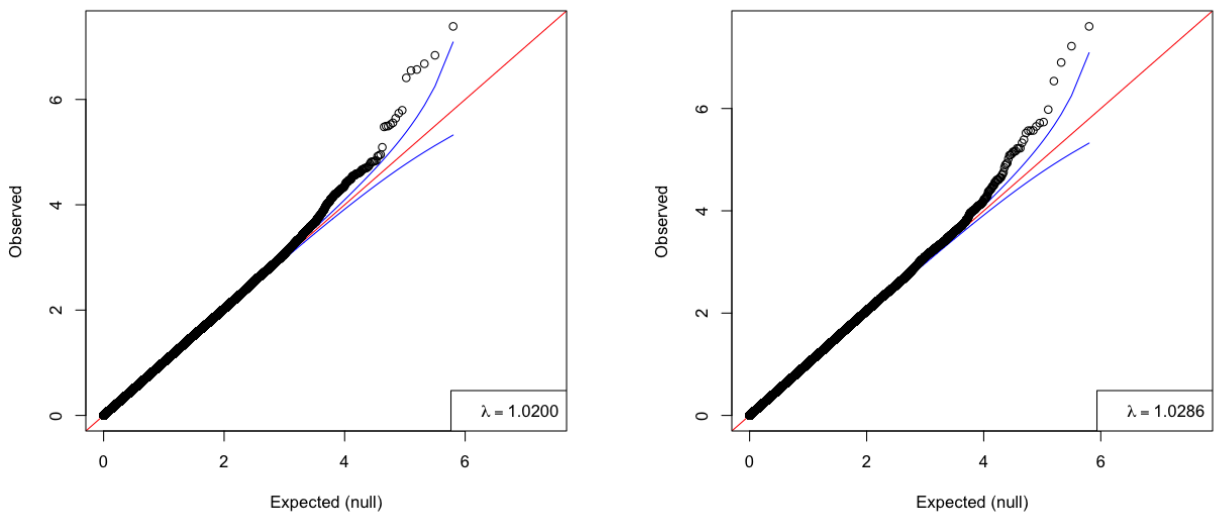


Figure 3.9: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following older childhood immunisation with MenC vaccine. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= lambda inflation factor.

As detailed in the subsection '3.3.3.7 Quantitative trait loci analysis', regression analyses between SNP allele dose and $\log_{10}$ transformed MenC-specific SBA titres were conducted either with the complement source included as a covariate or data separated by complement source were meta-analysed. The Q-Q plots of association p-values resulting from these regression analyses did not show signs of excessive inflation (Figure 3.10). However, for both these analyses there was enrichment of smaller p-values ($< 1\times10^{-4}$), which deviated from the 95% confidence intervals of the null distribution (Figure 3.10).



(a) (b)

Figure 3.10: Quantile-quantile plots of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms (SNPs) and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following older childhood immunisation with MenC vaccine. a) linear regression analysis with complement included as a covariate, b) meta-analysis of regression of SNPs against $\log_{10}$ transformed MenC-specific rSBA and hSBA titres. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= lambda inflation factor.

An overview of the results of the regression analysis between SNP allele counts and $\log_{10}$ transformed MenC-specific IgG concentrations are shown in Figure 3.11. Seven SNPs on three chromosomes surpassed the statistical threshold for a suggestive association ($p < 1 \times 10^{-5}$), four of these SNPs were within three genes: *SRSF7*, *DHX57* and *ENKUR* (Table 3.2).

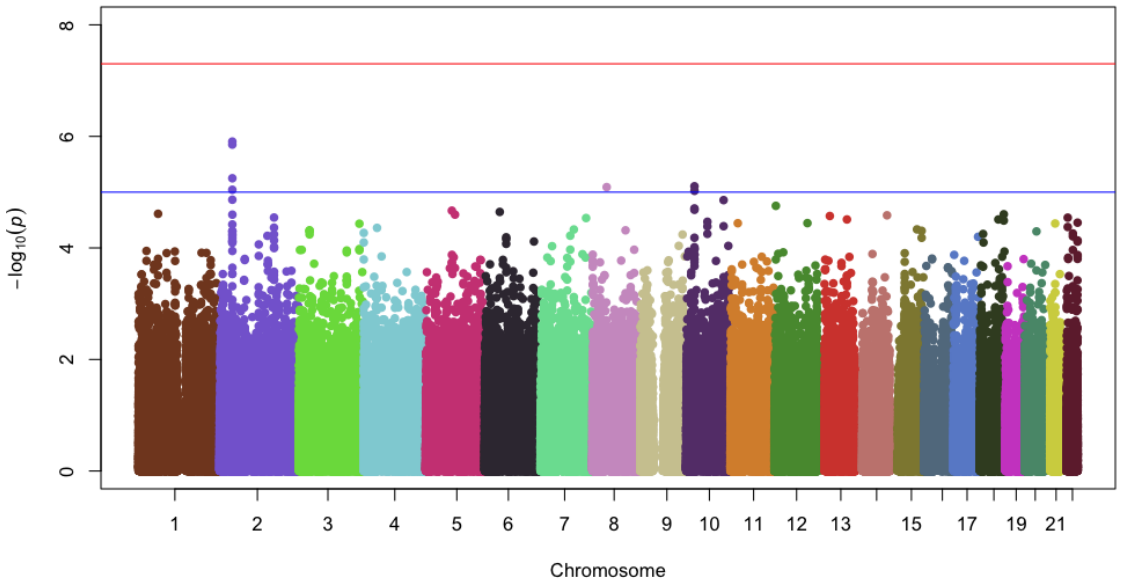


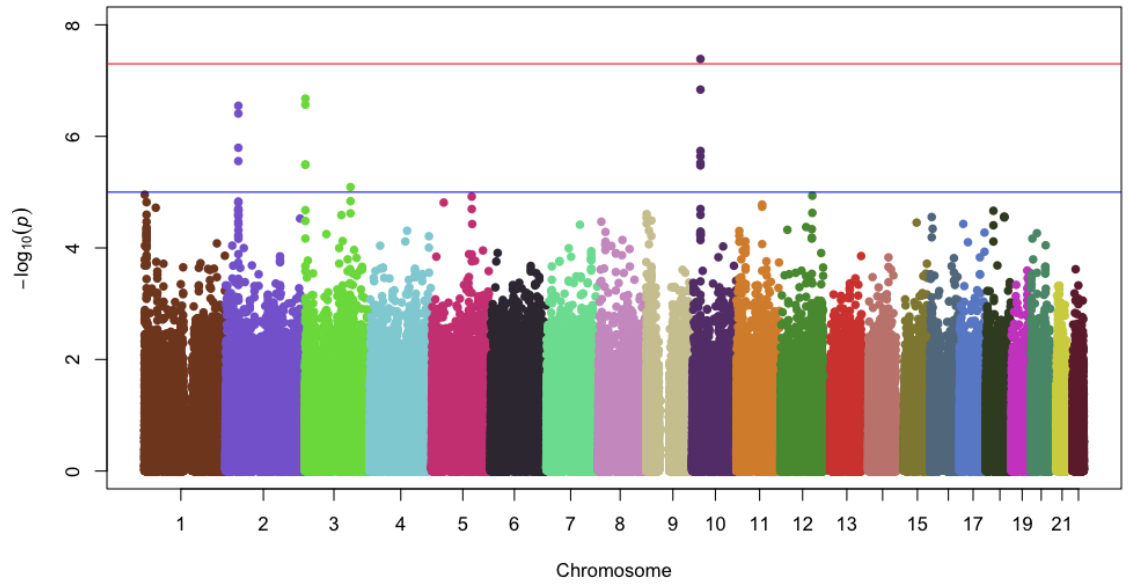
Figure 3.11: Manhattan plot of p-values from regression analysis between single-nucleotide polymorphisms and $\log_{10}$ transformed capsular group C meningococci (MenC)-IgG concentrations following older childhood immunisation with MenC vaccine. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$). The R code for this Manhattan plot is presented in appendix '7.4.3.6 Manhattan plot'.

Table 3.2: Single-nucleotide polymorphisms with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci-specific IgG concentration following older childhood immunisation with MenC conjugate vaccine. Gene information and neighbouring genes were derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241). A1= Allele 1

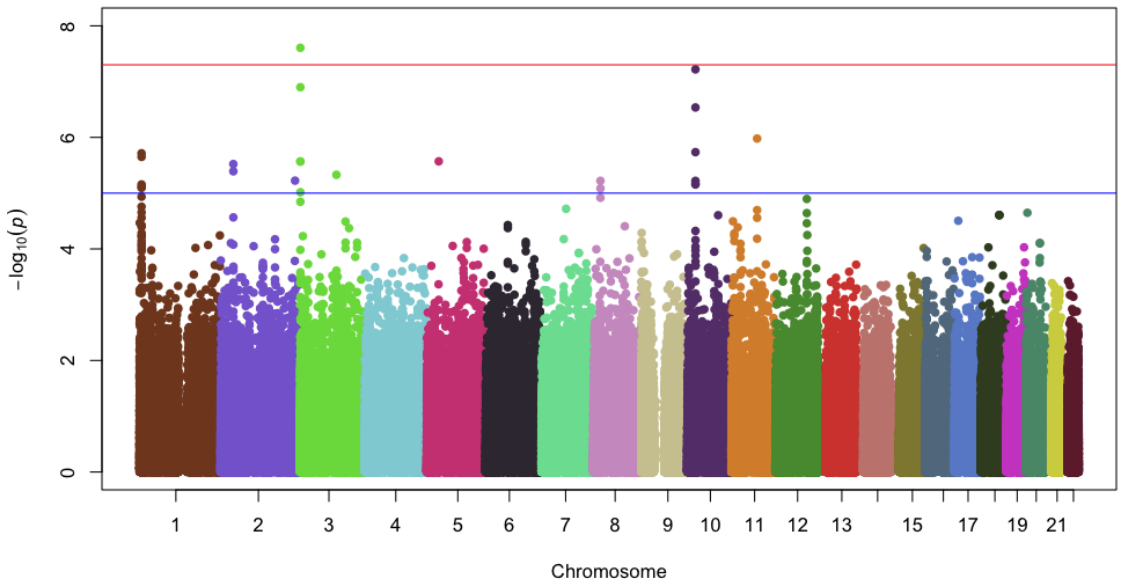
Chr	SNP	Position	A1	β	P	Gene	Neighbouring genes (<100kb)
2	rs3097713	38964531	G	0.21	9.07E-06		*ARHGEF33,LOC375196,SOS1,MORN2*
2	rs3134628	38977612	A	0.22	1.40E-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	rs13024811	38980231	A	0.22	1.24E-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	rs3099968	39033590	G	0.21	5.64E-06	*DHX57*	*ARHGEF33,LOC375196,SOS1*
8	rs8185884	43626988	A	-0.59	8.12E-06		
10	rs2279547	25273988	A	-0.2	7.85E-06	*ENKUR*	*GPR158-AS1,GPR158*
10	rs943043	25394337	G	-0.2	9.48E-06		*GPR158*

An overview of the results of the regression analysis between SNP allele counts and $\log_{10}$ transformed MenC-specific SBA titres are shown in Figure 3.12. Fifteen SNPs across three chromosomes showed p-values suggestive of association with MenC-specific SBA titres when SBA complement source was included as a covariate. Ten of these SNPs were within four genes: *SRSF7*, *CNTN6*, *CLSTN2* and *ENKUR* (Table 3.3). Meta-analysis of linear regression models for rSBA and hSBA titres found 24 SNPs, across seven chromosomes, with suggestive associations (Figure 3.12). Eighteen of these SNPs were within seven genes: *CAMTA1*, *SRSF7*, *CNTN6*, *GCSAM*, *SLC1A3*, *ENKUR*, and *TENM4* (Table 3.4).

At the genome-wide significance threshold of p-value $<5\times10^{-8}$, a single SNP (rs2279547) within *ENKUR* was found to be associated with MenC-specific SBA titres, when SBA complement source was included as a covariate in the linear regression model (Figure 3.12 and Table 3.3). This SNP narrowly missed the genome-wide significance threshold when linear regression was conducted separately based upon complement source and meta-analysed; however, another SNP (rs17785924) within *CNTN6* did surpass the level of genome-wide significance following meta-analysis (Table 3.4). Regional association plots for the two loci containing SNPs at the level of genome-wide significance are shown in appendix Figure 7.6 and appendix Figure 7.7.



(a)



(b)

Figure 3.12: Manhattan plots of p-values from regression analysis between single-nucleotide polymorphisms (SNPs) and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following older childhood immunisation with MenC vaccine. a) linear regression analysis with complement source as covariate, b) meta-analysis of regression models of SNPs against rSBA and hSBA titres. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).

Table 3.3: Single-nucleotide polymorphisms (SNPs) with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following older childhood immunisation with MenC vaccine, when complement source was included as a covariate in regression analysis. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	SNP	Position	A1	β	p-value	Gene	Neighbouring genes (<100kb)
2	rs3097713	38964531	G	0.23	2.78E-06		*ARHGEF33,LOC375196,SOS1,MORN2,DHX57*
2	rs3134628	38977612	A	0.23	3.89E-07	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	rs13024811	38980231	A	0.24	2.83E-07	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	rs2037875	38980922	G	0.21	1.60E-06		*ARHGEF33,LOC375196,SOS1,MORN2*
3	rs11928188	1388589	A	-0.28	3.21E-06	*CNTN6*	*CNTN6*
3	rs17038032	1389117	A	-0.28	3.21E-06	*CNTN6*	*CNTN6*
3	rs17785924	1389720	A	-0.35	2.70E-07	*CNTN6*	*CNTN6*
3	rs10510209	1405829	G	-0.33	2.10E-07	*CNTN6*	*CNTN6*
3	rs9820731	139750888	G	-0.26	8.09E-06	*CLSTN2*	*NMNAT3*
10	rs274311	25244919	G	-0.21	2.28E-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	rs274304	25247006	A	-0.21	3.32E-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	rs274288	25254151	A	-0.24	1.45E-07		*GPR158-AS1,RP11-80K21.3,GPR158*
10	rs2307047	25271270	A	-0.22	2.99E-06	*ENKUR*	*GPR158-AS1,GPR158*
10	rs2279547	25273988	A	-0.25	4.10E-08	*ENKUR*	*GPR158-AS1,GPR158*
10	rs10764519	25321244	A	-0.22	1.83E-06	*ENKUR*	*GPR158*

Chr = Chromosome, A1= Allele 1, β = the beta coefficient of regression analysis.

Table 3.4: Single-nucleotide polymorphisms with p-values suggestive of association with capsular group C meningococci-specific SBA titres following older childhood immunisation, when regression models were conducted separately for rabbit and human complement assays and meta-analysed. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	Position	SNP	A1	OR	p-value	I^2	Gene	Neighbouring genes (<100kb)
1	6936272	rs7546903	G	1.26	1.93e-06	56.37	*CAMTA1*	*CAMTA1*
1	6986109	rs4908574	G	1.23	8.05e-06	62.50	*CAMTA1*	*LOC100129476,CAMTA1*
1	7011805	rs10864251	C	1.21	6.98e-06	0.00	*CAMTA1*	*LOC100129476,CAMTA1*
1	7032168	rs6673655	A	1.25	2.24e-06	65.97	*CAMTA1*	*LOC100129476,CAMTA1*
1	7110334	rs4908590	C	1.21	7.71e-06	0.00	*CAMTA1*	*LOC100129476,CAMTA1*
2	38977612	rs3134628	A	1.23	4.06e-06	47.43	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	38980231	rs13024811	A	1.23	3.01e-06	50.41	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	228254272	rs4075148	A	0.77	5.97e-06	0.00		*LOC646794,SPHKAP*
3	1369428	rs265774	G	0.75	9.63e-06	0.00	*CNTN6*	*CNTN6*
3	1388589	rs11928188	A	0.76	2.71e-06	7.87	*CNTN6*	*CNTN6*
3	1389117	rs17038032	A	0.76	2.71e-06	7.87	*CNTN6*	*CNTN6*
3	1389720	rs17785924	A	0.70	2.48e-08	0.00	*CNTN6*	*CNTN6*
3	1405829	rs10510209	G	0.73	1.26e-07	0.00	*CNTN6*	*CNTN6*
3	111841303	rs1492489	G	1.25	4.69e-06	0.00	*GCSAM*	*PLCXD2,PHLDB2*
5	36668022	rs3776585	A	0.81	2.69e-06	0.00	*SLC1A3*	*RPL39P22,SLC1A3*
8	21466249	rs1383570	G	1.34	8.21e-06	0.00		
8	21470887	rs2010282	A	1.34	6.01e-06	0.00		
10	25244919	rs274311	G	0.82	6.02e-06	56.97		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25247006	rs274304	A	0.82	6.77e-06	49.30		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25254151	rs274288	A	0.80	2.91e-07	33.18		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25271270	rs2307047	A	0.81	1.84e-06	0.00	*ENKUR*	*GPR158-AS1,GPR158*
10	25273988	rs2279547	A	0.79	6.04e-08	0.00	*ENKUR*	*GPR158-AS1,GPR158*
10	25321244	rs10764519	A	0.82	7.01e-06	68.08	*ENKUR*	*GPR158*
11	78794102	rs4944221	A	0.81	1.05e-06	0.00	*TENM4*	*TENM4*

Chr= chromosome, I^2=percentage of total variation accounted for by heterogeneity (242). A1= Allele 1, OR = odds ratio.

3.3.4.5 Imputation

The imputed dataset contained a total of 34,026,329 genetic variants. Imputed data were filtered for minor allele frequencies >0.02, HWE p-values $>1\times10^{-9}$ and information scores >0.8. Subsequently, 6,704,216 and 6,703,388 SNPs remained for linear regression models of MenC-specific IgG concentrations and SBA titres, respectively.

The Q-Q plots and λ inflation values for the imputed data are shown in Figure 3.13 and Figure 3.15. The Q-Q plot of association p-values resulting from the analysis of imputed allele-doses and $\log_{10}$ MenC-specific IgG concentrations did not show signs of excessive inflation (Figure 3.13). However, some deflation of smaller p-values ($< 1\times10^{-4}$) was evident, which can be caused by imputed SNPs in low LD with genotyped SNPs leading to uncertainty and decreased variance of allele dosage (243).

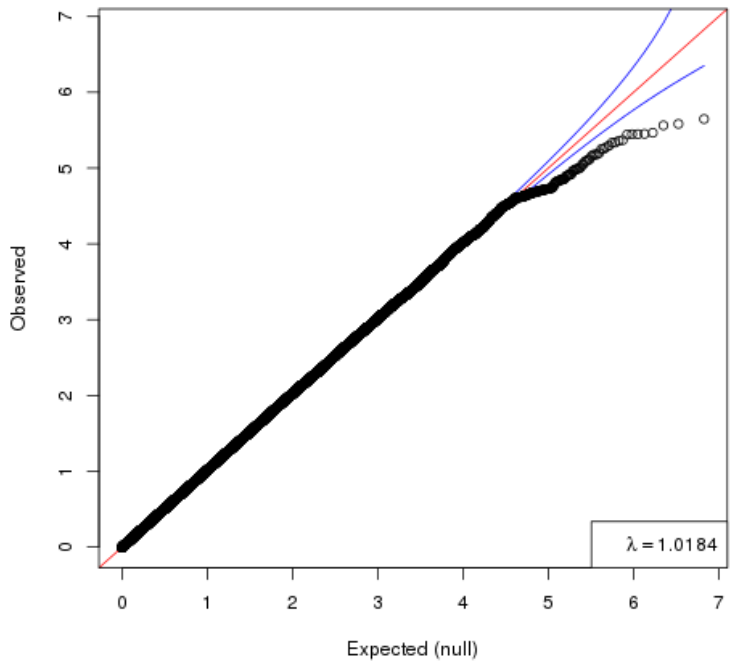


Figure 3.13: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci-specific (MenC)-IgG concentrations following older childhood immunisation with MenC vaccine. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= lambda inflation factor.

A Manhattan plot of p-values resulting from association analysis between imputed SNPs and MenC-specific IgG concentrations is displayed in Figure 3.14. A total of 28 SNPs in seven genes *SRSF7*, *GEMIN6*, *DHX57*, *MORN2*, *SENP2*, *VTI1A* and *DOCK1*, surpassed the level of suggestive association with MenC-specific IgG concentrations in the imputed dataset (Figure 3.14 and appendix Table 7.11).

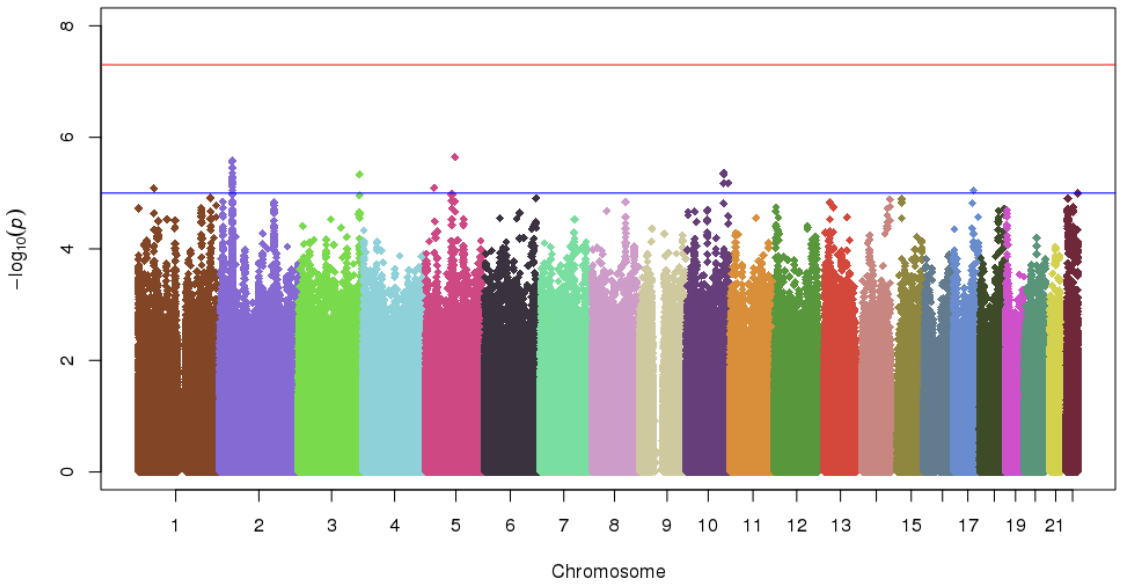
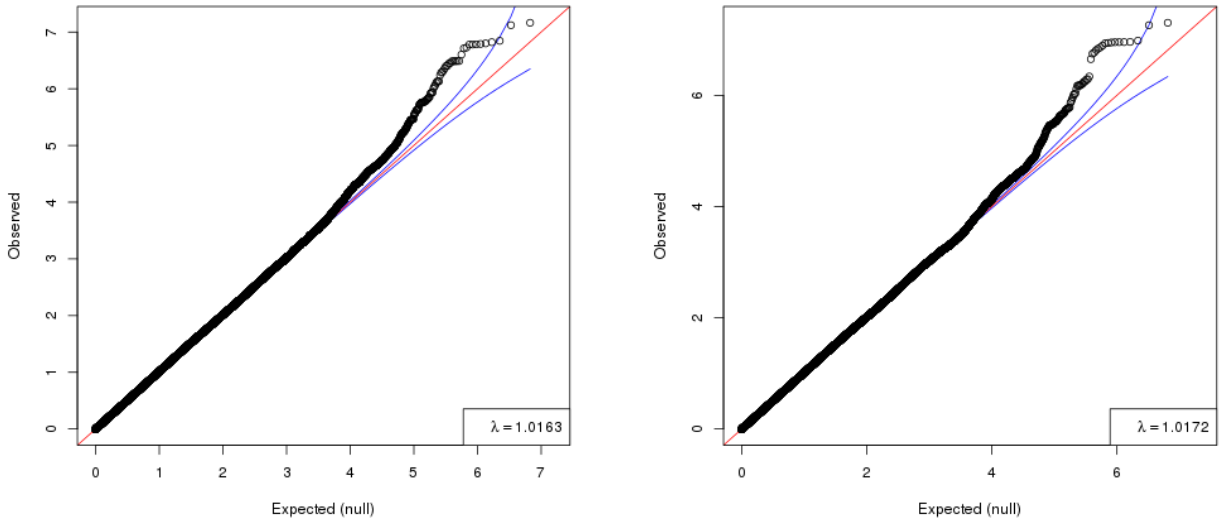


Figure 3.14: Manhattan plots showing p-values from regression analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following older childhood immunisation with MenC vaccine. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).

The Q-Q plots of association p-values resulting from the analyses of imputed allele-doses and $\log_{10}$ MenC-specific SBA titres are shown in Figure 3.15. While the overall distribution was not excessively inflated, enrichment of smaller p-values ($< 1\times10^{-4}$) was evident, some of which deviated from the 95% confidence intervals of the null and may signify loci truly associated with this phenotype.

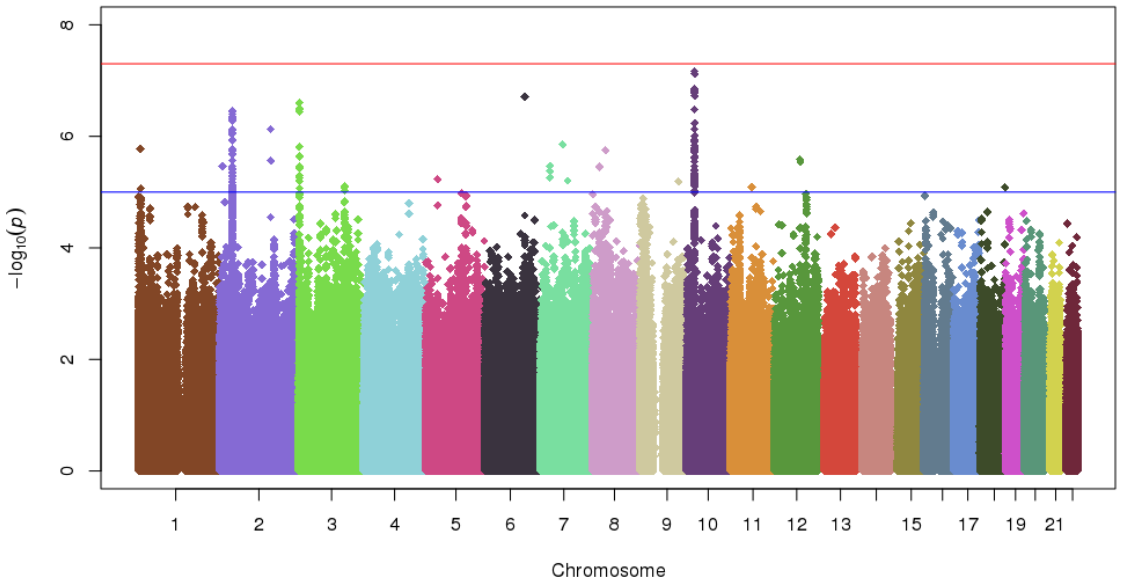


(a) (b)

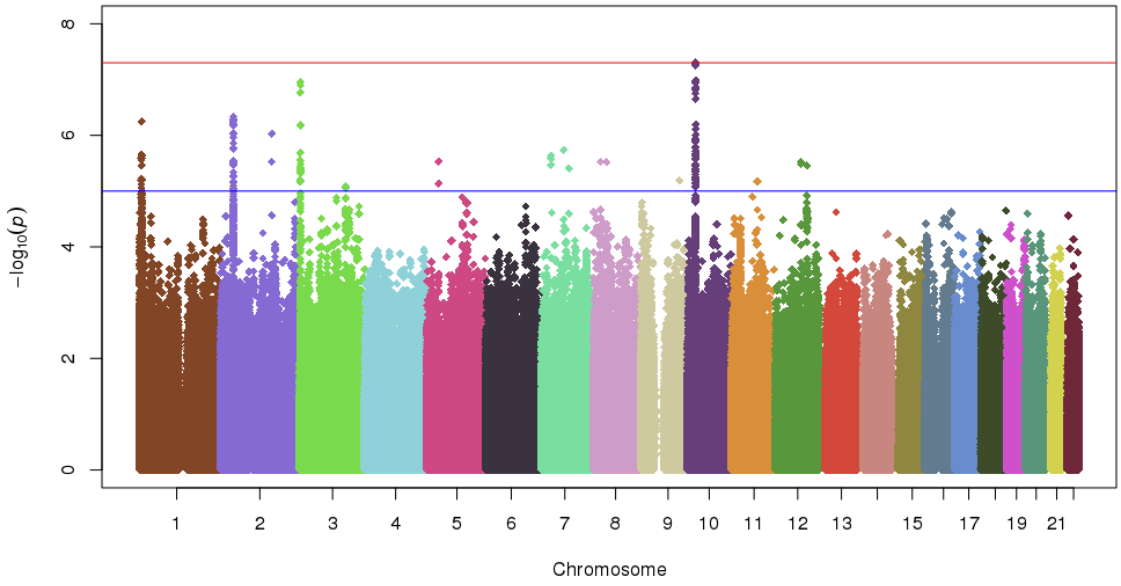
Figure 3.15: Quantile-quantile plots of -$\log_{10}$ p-values observed in association analyses between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay titres following older childhood immunisation with MenC vaccine. a) linear regression analysis with complement as covariate, b) meta-analysis of separate rabbit and human complement SBA regression analyses. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

Manhattan plots of p-values resulting from association analyses between imputed SNPs and MenC-specific SBA concentrations are displayed in Figure 3.16. A total of 132 SNPs in 15 genes surpassed the level of suggestive association with MenC-specific SBA titres in the imputed dataset, when SBA complement source was included as a covariate (Figure 3.16a and appendix Table 7.12). The results of meta-analysis of separate linear regression model for rSBA and hSBA were similar to the joint analysis, with 123 SNPs in 15 genes having p-values with suggestive association with MenC-specific SBA titres (Figure 3.16b and appendix Table 7.13).

Regional plots of the two genetic loci containing the most statistically significant SNPs associated with MenC-specific SBA titres are displayed in Figure 3.17 and appendix Figure 7.8. Figure 3.17a, displays an associated region spanning ∼100kb of *CNTN6*; this region exhibits high LD and is bordered by areas of increased recombination, distal of which the association signal dissipates. Figure 3.17b, shows an extended region ∼200kb of associated SNPs spanning three annotated genes: *PRTFDC1*, *ENKUR* and *THNSL1*.

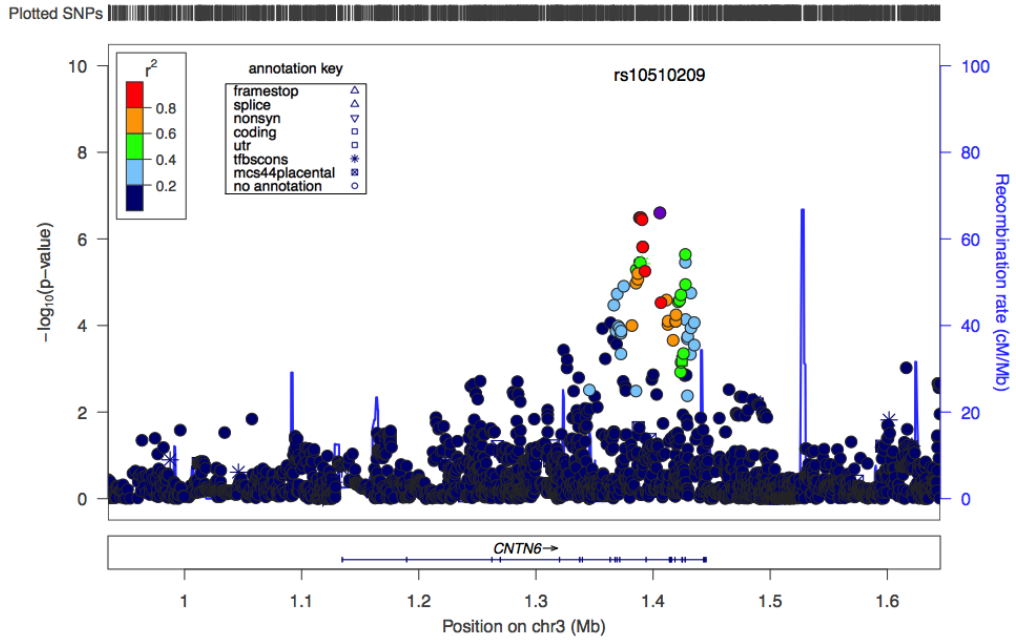


(a)

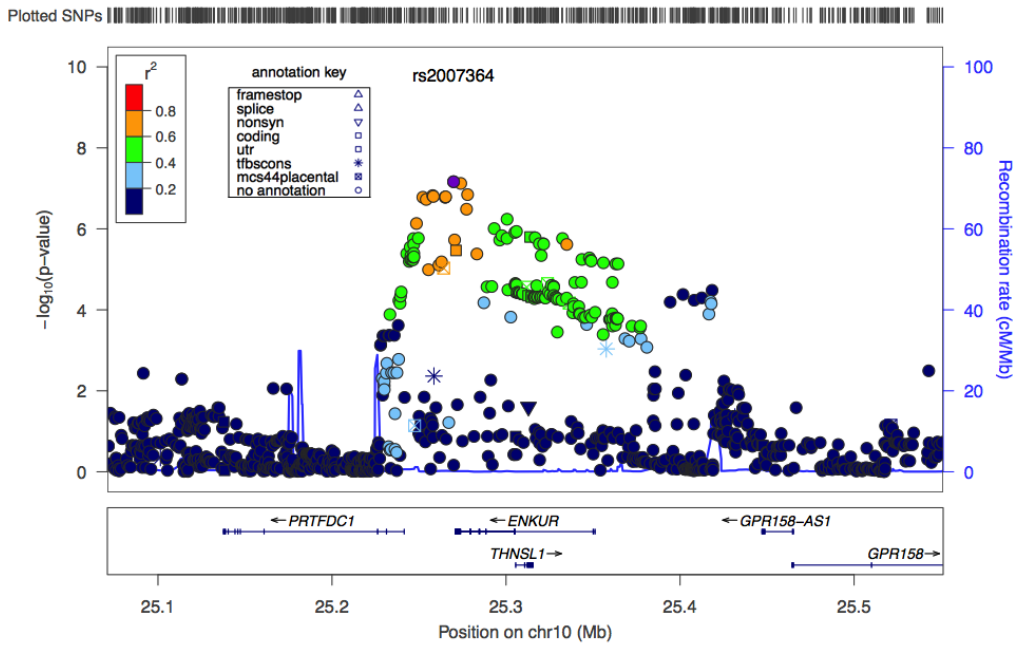


(b)

Figure 3.16: Manhattan plots of p-values from regression analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following older childhood immunisation with MenC vaccine. a) linear regression analysis with complement as covariate, b) meta-analysis of regression of allele count against rSBA and hSBA titres. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).



(a)



(b)

Figure 3.17: A regional plot of the imputed chromosome 3 and 10 loci associated with capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following older childhood MenC vaccination, with SBA complement source included as a covariate. a) chromosome 3 loci associated with MenC-specific SBA titres, b) chromosome 10 loci associated with MenC-specific SBA titres. The local gene structure is shown below the plot, with thick lines denoting exons joined by intronic regions. The statistical association (p-value) is shown on the primary y-axis (left) and the recombination rate is shown on the secondary y-axis (right). The SNP with the most statistical evidence of association is shown in purple and remaining SNPs are coloured according to their linkage disequilibrium (r^2) with this 'lead' SNP. SNPs with particular features are assigned a symbol, as denoted by the 'annotation key'. These plots were created using the LocusZoom tool (244).

3.3.5 Discussion

This section details a genome-wide association study assessing the relationship between genetic polymorphisms and the persistence of vaccine-specific serological immunity following a single-dose of MenC conjugate vaccine administered to children aged 6-15 years. This study is the first genome-wide assessment of the genetic determinants of MenC-specific serological immunity and described two novel genome-wide associated loci, within two annotated genes *CNTN6* and *ENKUR*.

CNTN6 encodes for the contactin 6, a neuronal membrane protein involved in cell adhesion, present at medium to high levels in several tissues including lymph node (`http://www.proteinatlas.org/`). Contactins are a family of six immunoglobulin superfamily cell adhesion molecules (IgCAMs), primarily annotated as being involved in brain development and behaviour (245). Genetic variants in *CNTN6* have been associated with a number of phenotypic traits including anorexia nervosa, bipolar disorder, propensity to store fat viscerally as compared to subcutaneously, and plasma low-density lipoprotein cholesterol levels (246, 247, 248). Genetic variants in *CNTN6* have also been associated with plasma levels of intercellular adhesion molecule-1, a ligand involved in leukocyte adhesion and inflammation (249). Furthermore, murine studies have shown involvement of ICAM-1 in the retention of B cells within splenic marginal zone, germinal centre (GC) B cell fitness and responses to soluble antigens (250, 251).

ENKUR is expressed by several cell types including leukocytes (`http://www.ebi.ac.uk/gxa/home`) and encodes for enkurin, a protein that interacts with transient receptor potential canonical cation channels and calmodulin (252). Phenotypic traits that have been associated with variants in *ENKUR* include, glycosylated haemoglobin A levels and diabetic nephropathies (`http://www.ncbi.nlm.nih.gov/gap/phegeni`) (253). Furthermore, variants in a neighbouring gene *GPR158* have recently been associated with neutralising antibody responses following smallpox vaccine (26).

The functional mechanisms underlying the association between *CNTN6* and *ENKUR* variants, and the persistence of MenC-specific SBA titres are unclear. If these association signals are true, they may indicate a direct role for the protein products of these genes in the gen-

eration and/or persistence of serological immunity to MenC conjugate vaccine; alternatively these signals may be proxies for other functional elements. Previous genome-wide association studies of complex phenotypes have suggested that a large proportion of genetic associations are driven by non-coding variants (254). Fine-mapping association signals to causal variants has represented a considerable challenge for genetic association studies, and methodologies to improve this process continue to be developed (255, 256).

There were a number of limitations to this study. Firstly, this study was powered to detect an additive allele effect of 1.7-fold in MenC-specific IgG concentrations, which represents a relatively large effect size, therefore variants with smaller effect sizes may not have surpassed the stringent level of statistical correction. Secondly, low frequency (0.5-5%) and rare (<0.5%) variants are not well captured by contemporary GWAS microarrays, consequently their contribution may not be well evaluated (257). Thirdly, the analysis of immunological data generated by two laboratories, using differing protocols, is likely to have increased assay variability and reduced power. Fourthly, while the majority of study participants received the CRM_{197} conjugated MenC vaccine, some received the tetanus toxoid conjugated vaccine and there may be important differences in the way these carrier proteins recruit T-cell help that could result in disparate genetic associations. Finally, natural exposure to capsular group C meningococci may have confounded this study. However, very low rates of MenC carriage (<0.15%) have been reported in the UK since the introduction of the MenC vaccine, meaning this is unlikely to be a major factor (234).

Previous studies have found a strong relationship between sample size and the number of loci identified at genome-wide significance (258). As this study represents a modest sample size, variants with small to moderate effect sizes are unlikely to surpass the conservative genome-wide level of statistical significance. When SNPs with $p<1\times10^{-5}$, a threshold commonly regarded in GWASs as 'suggestive' of association were interrogated, a number of other genes are implicated in the persistence of MenC-specific immunity following vaccination (259). In particular, *GCSAM* (germinal centre-associated, signalling and motility) is noteworthy, as it encodes a protein involved in the regulation of B cell receptor signalling and lymphocyte

migration (`http://www.ncbi.nlm.nih.gov`). The expression of *GCSAM* has been shown to be predictive of survival in diffuse large B-cell lymphoma and Hodgkin's lymphoma (260, 261). Moreover, *GCSAM* transgenic mice have been shown to develop polyclonal B-cell lymphoid hyperplasia and hypergammaglobulinemia (260, 262). The *GCSAM* transcript is highly expressed in germinal centre lymphocytes and is induced in B cells by interleukin-4 (IL-4) (260). As IL-4 promotes the differentiation of germinal centre (GC) B cells into memory B cells, it has been suggested that GCSAM may also be involved in B cell differentiation (260, 263). Moreover, in vitro experiments have shown GCSAM to enhance BCR signalling by binding and increasing Syk activation, further suggesting GSCAM involvement in humoral immunity (262).

The remaining genes reaching the suggestive level of statistical association with persistence of MenC-specific immunity have not been annotated as being involved in immunity. However, this does not preclude their functional relevance in the complex phenotype of vaccine-induced immunity. Furthermore, these associations may be driven by regulatory elements affecting proximal as well as distal genes, that may even span chromosomes (219, 264, 265). For example *CAMTA1*, which contains several SNPs with suggestive association with the persistence of MenC-specific SBA titres, also harbours several cis and trans expression quantitative trait loci (eQTL) for genes such as interleukin 17 receptor C (*IL17RC*), colony stimulating factor 1 (*CFS1*) as well as an open reading frame in the MHC class III locus (C6orf25). Furthermore, variants within *CAMTA1* have been associated with serum levels of IL-6, a cytokine that promotes differentiation of B cells and is produced by a range of cells including T cells, monocytes and fibroblasts (249).

Hitherto, two studies have assessed the genetic determinants of the persistence of immunity following immunisation with a meningococcal conjugate vaccine (38, 266). The first, Moore et al 2012, was conducted on a subset of 905 participants described in this section. This study assessed 721 SNPs in 131 candidate genes, finding 59 SNPs in 37 genes to be associated with the persistence of MenC-specific IgG concentrations and 56 SNPs in 36 genes to be associated with the persistence of MenC-specific SBA titres (38). The second, Spector et al 2013, assessed the relationship between 14 SNPs in five genes and the fold-rise in capsular groups A, C, W and

Y-specific SBA, 72 weeks after immunisation with the quadrivalent meningococcal conjugate vaccine, in a cohort of 198 HIV-infected 11-24 year olds (266). This study described a polymorphism in *TLR9* (rs5743836) and a non-synonymous variants in *FCγRII* (rs1801274) that associated with the persistence of a composite measure of fold-rise in SBA titres to meningococcal capsular groups A, C, W and Y (266). However, rs5743836 and rs1801274 were not significantly associated with the fold-rise in MenC-specific SBA titres alone (266). Furthermore, rs1801274 was also not significantly associated with MenC-specific SBA titres in the study described by Moore et al 2012 (38, 266). Additionally, while the HIV-positive cohort were on stable or no antiretrovirals, they had lower fold-rises in meningococcal capsular group-specific SBA titres than HIV-negative individuals and it is not clear how generalisable associations from HIV-infected individuals are to healthy individuals (266, 267). Moreover, the HIV-positive cohort were composed of multiple ethnic subgroups: 44% black non-Hispanics, 41% Hispanics, 13% white non-Hispanics and 2% other, and the analysis of these disparate ancestral groups may be confounded by population stratification (266, 268). The most statistically significant findings of the genome-wide association study described in this section were not assessed in the aforementioned candidate gene studies, illustrating the potential benefit of genome-wide hypothesis-free approaches.

3.3.6 Conclusion

Two genes, *CNTN6* and *ENKUR*, contained SNPs associated at the level of genome-wide significance with the persistence of MenC-specific SBA titres following older childhood immunisation with MenC conjugate vaccine. In addition, a number of SNPs showed suggestion of association with the persistence of MenC-specific SBA titres including SNPs within *GCSAM*, a gene highly expressed in germinal centre lymphocytes and induced in B cells by interleukin-4 treatment. While causal variants have not yet been demonstrated, these associated loci make promising candidates for further functional evaluation.

3.4 Genome-wide association study of the persistence of serological immunity to capsular group C meningococcal vaccine following childhood immunisation - Stage 1

3.4.1 Background

As detailed in section 3.3, the peak incidence of MenC disease prior to the introduction of the MenC conjugate vaccine occurred in those under 1 year of age (184). Following the MenC vaccination campaign of 1999-2000, the incidence of MenC disease decreased drastically (Figure 3.2). However, MenC vaccine-induced immunity wanes rapidly in young children, even after repeated doses (3, 212). Despite the demonstrable induction of immunological memory following infant vaccination, MenC vaccine efficacy was shown to be short-lived and conferred little protection more than a year after vaccination, with antibody levels decreasing to pre-vaccination levels by 4 years of age (212, 269). These findings were consistent with observations made following infant immunisation with *Haemophilus influenzae* type b (Hib) conjugate vaccine (270). While conjugate vaccines generate immunological memory in infants, in the absence of serological immunity this may not be sufficient to protect against rapidly invading organisms (140, 271).

The mechanisms underlying the magnitude and persistence of serological immunity following vaccination are not well understood; however, age has been shown to be an important factor for a number of vaccines (3, 212, 213, 214). The age-related features affecting immune responses are largely unknown but are not limited to vaccine-induced immunity, disparities in susceptibility to infectious and immune-related diseases are also associated with age (272, 273). Various qualitative and quantitative differences in immunological parameters are associated with chronological age, with the immune system early in life having an apparent 'T_H2 bias' (274, 275). The immune system is not completely functional at birth, which is believed to be advantageous in maintaining immune tolerance during pregnancy and enabling acclimatisation to an antigen

rich external environment (276). Full immune competence is thought to occur around school age for most infants, a feature speculated to be the basis of the preponderance of preschool children to infectious disease (274, 275). Moreover, delayed postnatal maturation of 'T_H1' function has been linked with atopy (277). The mechanisms involved in the maturation of the immune system and the transition from 'T_H2 ' to 'T_H1' skewed immunity have not been fully characterised. However, there may be a relationship between genetic variants and the rate of immunological maturation. Conceivably, these variants may account for some of the variability in infant immune responses to vaccines, in an age-dependent manner.

Age-dependent genetic factors may not be limited to early in life. The immune system is subjected to immunologically perturbing hormones during the puberty, a physiological event that coincides with increased incidence of several immune-related diseases (278). In addition, quantitative and qualitative immunological changes are also seen in old age, a process referred to as 'immunosenescence' (273). Age-dependent factors are often overlooked in genetic association studies; however, these have been demonstrated in immune-related phenotypes such as asthma and type-1 diabetes (279, 280, 281). The discernment of age-dependent and age-independent factors influencing immunological phenotypes may also elucidate genes involved in physiological processes such as the maturation of the immune system. To date, the interactions between age and genetic factors determining vaccines responses have not been explored.

The following section describes the first stage of a two-stage GWAS of the persistence of serological immunity to MenC following childhood immunisation with MenC conjugate vaccine. This section also details age-stratified analysis to assess for potential age-dependent and age-independent genetic factors influencing the persistence of serological immunity following immunisation with MenC conjugate vaccine.

3.4.2 Aim

The aim of this section was to describe stage 1 of a two-stage genetic association study of the persistence of serological immunity to capsular group C meningococcal vaccine following childhood immunisation.

3.4.3 Materials and methods

3.4.3.1 Stage 1 participants

The subsequent section includes healthy children recruited into four clinical trials, in addition to the adolescents described in subsection '3.3.3.1 Adolescent participants'. In brief, study 1 assessed the immunogenicity of the Hib-MenC conjugate vaccine in infancy, with persistence of MenC-specific immunity measured at 1 year of age (145). Study 2 evaluated the immunogenicity of the 2009 pandemic H1N1 influenza vaccine; however, data were also collected on routine childhood immunisations and sera were available for the assessment of MenC-specific immunity at 6 months to 12 years of age (282). Study 3 evaluated immunological memory following infant MenC conjugate vaccine administered concomitantly with routine immunisations in consistent or alternating limbs, with persistence of MenC-specific immunity measured at 1 year of age (283). Study 4 examined the immunogenicity of the tetravalent (capsular groups A, C, W and Y) meningococcal conjugate vaccine in infancy, with persistence of MenC-specific immunity measured at 1 year of age (141).

Informed consent was obtained from parents/guardians of participants included in these clinical studies, and explicitly for the genetic research detailed in this section. Oxfordshire research ethics committees approved these studies and the genetic analysis presented in this section (07/Q1605/34, 09/H0604/107, 10/H0604/7, 04/Q1604/28,10/H0504/25). In total 2211 participants were recruited into this study: study 1 n=75, study 2 n=454, study 3 n=301, study 4 n=199 and adolescents n=1182 (described in the subsection'3.3.3.1 Adolescent participants').

3.4.3.2 Vaccines

Participants described in this section received at least one dose of a vaccine containing MenC polysaccharide conjugated to a protein carrier, either in a clinical trial setting or as part of their routine clinical care. Vaccines included monovalent MenC conjugated to CRM_{197} (Meningitec or Menjugate) or tetanus toxoid (NeisVac-C), combined Hib-MenC vaccine conjugate to tetanus toxoid (Menitorix), or quadrivalent MenACWY vaccine conjugated to CRM_{197}. These

vaccines were described in subsections '2.2.3.2 Vaccines' and '3.3.3.2 Vaccines'. MenC immunisation information was obtained from clinical study data, or from the child health computer department.

3.4.3.3 Immunological measurements

The Vaccine Evaluation Unit (VEU), Public Health England, Manchester, measured MenC-specific hSBA titres and IgG concentrations for study 2 and study 3. The lower limit of quantification (LLQ) for the VEU MenC-specific IgG ELISA was 0.12μg/ml and values below this were reported as half the LLQ. GSK Biologicals, Belgium, performed MenC-specific rSBA and IgG measurements for study 1. The LLQ for the GSK MenC IgG ELISA was 0.3μg/ml and values below this were reported as half the LLQ. Novartis Vaccines, Germany, measured MenC-specific hSBA titres for study 4. The LLQ for the hSBA assay was a reciprocal titre of 4. The LLQ for the rSBA assay performed by VEU was a reciprocal titre of 4. The LLQ for the rSBA assay performed by GSK was a reciprocal titre of 8. Sera without detectable bactericidal antibody were assigned an arbitrary value of 2. The immunological methods utilised for the adolescents immunised as older childhood were previously described in subsection '3.3.3.3 Immunological measures'. All serological values were $\log_{10}$ transformed for statistical analysis.

3.4.3.4 DNA extraction

Frozen blood clots (remaining from serum separated peripheral blood) were thawed at 37 °C, placed in Clotspin® baskets (Qiagen, Hilden, Germany) and centrifuged at 2000×g_0 for 5 minutes to disperse the clot. Then, 5 ml of red blood cell lysis solution (16.52 g NH_4Cl, 2 g $KHCO_3$ and 0.074 g DNA in 2 l of H_2O) was added, samples were vortexed for 3 seconds, and shaken for 5 minutes on an orbital mixer at 120-200 rpm. Samples were vortexed for 3 seconds and centrifuged at 2000×g_0 for 5 minutes. Supernatant was discarded and an additional 5 ml of red blood cell lysis solution was added, vortexed for 3 seconds, before repeating the orbital mixer step. Samples were centrifuged at 2000×g_0, supernatant discarded and vortexed. Next, 1 ml of Cell Lysis solution (1 ml nuclei lysis solution, Promega, Wisconsin, United States; 25 μl

Proteinase K, Qiagen, Hilden, Germany; and 250 μl 0.5 M ethylenediaminetetraacetic acid) was added and vortexed, before being incubated at 56 °C for 2 hours. Following this incubation, DNA was extracted according to the standard spin protocol for whole blood, using the Qiamp® midi kit (Qiagen, Hilden, Germany). DNA concentrations were quantified using Quant-iT PicoGreen® dsDNA Assay Kit (Invitrogen™, California, United States) and diluted to 50 ng/μl for genotyping.

3.4.3.5 Genotyping

Genotyping was performed using either Illumina® OmniExpress12v1 or OmniExpress-12v1.1 genotyping microarrays (Illumina, California, United States), at the Genome Institute of Singapore, as described in subsection '3.3.3.5 Genotyping'. A common set of SNPs on both of these platforms was used to perform association analysis (n=717,991). Strand designation and base-pair positions were based upon build 37 of the Genome Reference Consortium's human genome assembly, with genotypes being annotated relative to the positive strand (284).

3.4.3.6 Quality control

The genotyping quality control methods used in this section were analogous to those described in subsection '3.3.3.6 Quality control'. However, in contrast to the previously described QC procedures, SMARTPCA software was used to identify ancestral outliers (`http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html`) (237). In brief, this involved pruning genotype data to a set of variants present in the HapMap genotype data from Europe, Asia and Africa. Then, the first two principal components (PCs) were used to compare study participants to these three reference populations and outliers were excluded as described by Anderson et al 2010 (235).

3.4.3.7 Quantitative trait loci analysis

A linear regression model was used to assess the relationship between SNP allele dosage and the persistence of $\log_{10}$ transformed MenC-specific IgG concentrations, with age at vaccination,

time since vaccination, study identifier (i.e. into which study a participant was recruited), and the first three principal components as covariates. A linear regression model was also used to assess the relationship between SNP allele dosage and the persistence of $\log_{10}$ transformed MenC-specific SBA titres, with SBA complement source, age at vaccination, study identifier, and the first three principal components as covariates. Linear regression analysis of genotype data and MenC-specific immunity were conducted using PLINK (156).

3.4.3.8 Heterogeneity analysis

A fixed-effect meta-analysis was undertaken on data stratified by age group at vaccination (i.e. infant or older childhood), to assess the consistency of relationships between SNPs and the persistence of immune responses following infant and older childhood MenC vaccination. An I^2 value was calculated to assess the heterogeneity between these vaccination age groups, using equation 3.1.

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100 \tag{3.1}$$

Where Q is Cochran's Q test statistic and df = degrees of freedom.

3.4.3.9 Imputation

The imputation method utilised in this section was analogous to that described in subsection '3.3.3.8 Imputation'.

3.4.3.10 Power calculations

An analytical approach was used to calculate statistical power, employing the R software package 'QpowR' (285). Power calculations were based upon genetic variants with an allelic effect on the persistence of MenC-specific immunity following childhood immunisation with MenC conjugate vaccine. The statistical model and the parameters selected can be found in appendix section '7.3.2 Power calculations'. The sample size was limited to the number of suitable participants (n=3014) recruited by the Oxford Vaccine Group at the time this study was initiated. The

type I error was Bonferroni corrected for a million independent tests, resulting in α=5×10^{-8}. As the effect size was unknown a priori, this was assessed at a range of values. A two-stage design was incorporated into power calculations, and the effect of varying the proportion of individuals assessed in the 'stage 1' and the 'stage 2' is displayed in Figure 3.18. Statistical power increases with the proportion of individuals genotyped in stage 1, but only a marginal gain in power is seen when ≥50% of individuals are genotyped in stage 1 (Figure 3.18). A two-stage design was opted for in this study with 65% of the study cohort assigned to stage 1, the trade-off between power and cost is displayed in Figure 3.19.

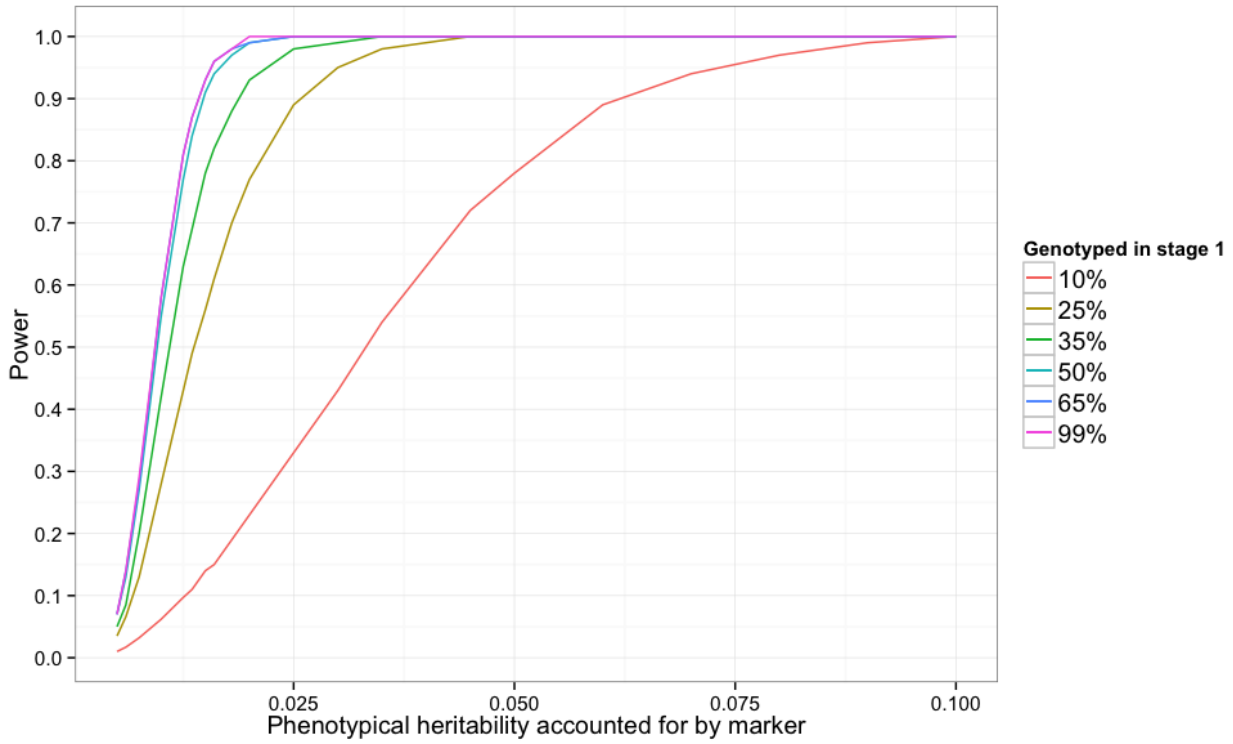


Figure 3.18: Power calculations for a two-stage genome-wide association study. Calculations were based upon an α=5×10^{-8} and 0.1% of markers being carried forward to stage 2, using R software package 'QpowR' (285).

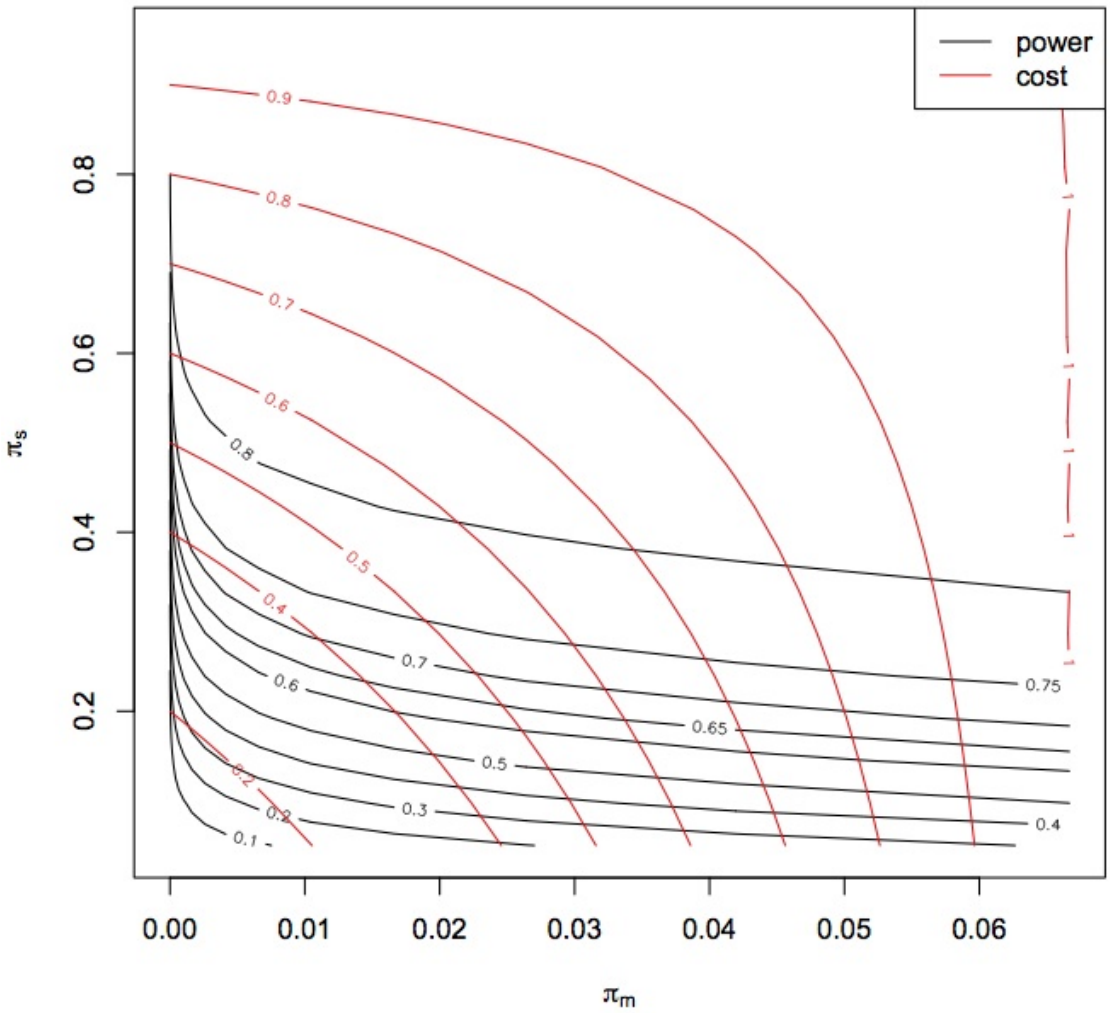


Figure 3.19: Tradeoff between power and cost in a two-stage study design. Calculations based upon the heritability (h^2) accounted for by genetic marker of 0.0125 and genotyping being 14 times more expensive for stage 1 than stage 2. π_s, proportion genotyped in stage 1; π_m, the proportion of markers carried through to stage 2. This plot was created using the R software package 'QpowR' (285)

3.4.4 Results

3.4.4.1 Stage 1 demographics

An overview of stage 1 demographics is displayed in Table 3.5.

Table 3.5: Demographics of participants in stage 1 of the genome-wide association study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following childhood immunisation with MenC vaccine.

Study	Male/Female	Time since vaccination days* (IQR)	Age at vaccination days* (IQR)	Reference
†Cohort 1	122/140	1380.5 (1279.3 - 1423.0)	3832.5 (3633.5 - 4073.3)	(213)
†Cohort 2	449/378	1777 (1679.0 - 1856.0)	3514 (3038.0 - 4078.0)	(214)
‡Study 1	39/27	275.5 (263.3 - 287.8)	123 (118.0 - 128.0)	(145)
‡Study 2	151/120	740 (341.5 - 2336)	365 (120.0 - 365.0)	(282)
‡Study 3	138/129	256 (248.5 - 264)	122 (118.0 - 127.0)	(283)
‡Study 4	76/82	261 (249.3 - 274)	125 (121.0 - 131.0)	(141)

IQR= interquartile range, * = median. †= immunised as older children (6-15 years), ‡= immunised as infants (<1 year of age)

3.4.4.2 Serology

The serological results of stage 1 participants vaccinated as infants are summarised in Figure 3.20 and Figure 3.21. Serological data from participants vaccinated as older children were previously shown in Figure 3.4 and Figure 3.5. Higher MenC-specific IgG concentrations and SBA titres were seen for individuals vaccinated in older childhood than those vaccinated as infants (Figure 3.20, Figure 3.21, Figure 3.4 and Figure 3.5).

The distribution of study 1 MenC-specific SBA titres is to the right of the other studies, which is likely to reflect the use of rabbit complement and human complement, respectively (Figure 3.20). Even between studies that used human complement there were considerable differences in the distributions of MenC-specific SBA titres. In particular, study 2 had a broader distribution, probably relating to the larger range in time since vaccination in this study (Figure 3.20 and Table 3.5). Of note, a large proportion of serum from participants immunised in infancy had lost detectable levels of bactericidal activity, represented by the peak at the left-hand side of the distribution.

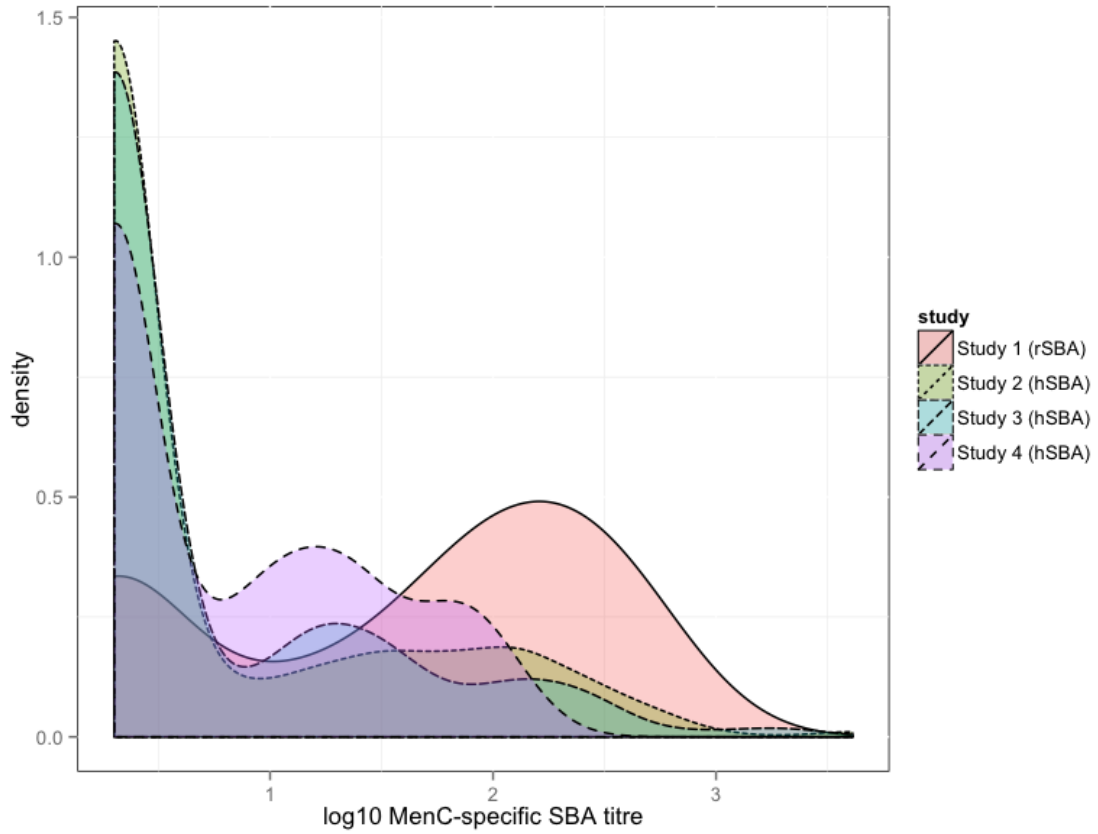


Figure 3.20: Density plot of $\log_{10}$ transformed capsular group C meningococci (MenC)-specific SBA titres following infant immunisation with MenC conjugate vaccine. Data are separated by vaccine study. rSBA= rabbit complement serum bactericidal assay, hSBA= human complement serum bactericidal assay. Density plots were created from a 1d kernel density estimate, using the R package 'ggplot2' (240).

MenC-specific IgG concentrations were assessed following infant vaccination for two of the four studies described in this section. When log-transformed the IgG concentrations from both these studies approached a normal distribution, with enrichment of values below the lower limit of detection. Study 2 had a greater proportion of samples below the lower limit of detection, despite the use of a more sensitive assay, as well as peak density at a lower MenC-specific IgG concentration (Figure 3.21). This is likely to reflect the fact that the time since vaccination was considerably longer for study 2 than for study 1 (Table 3.5).

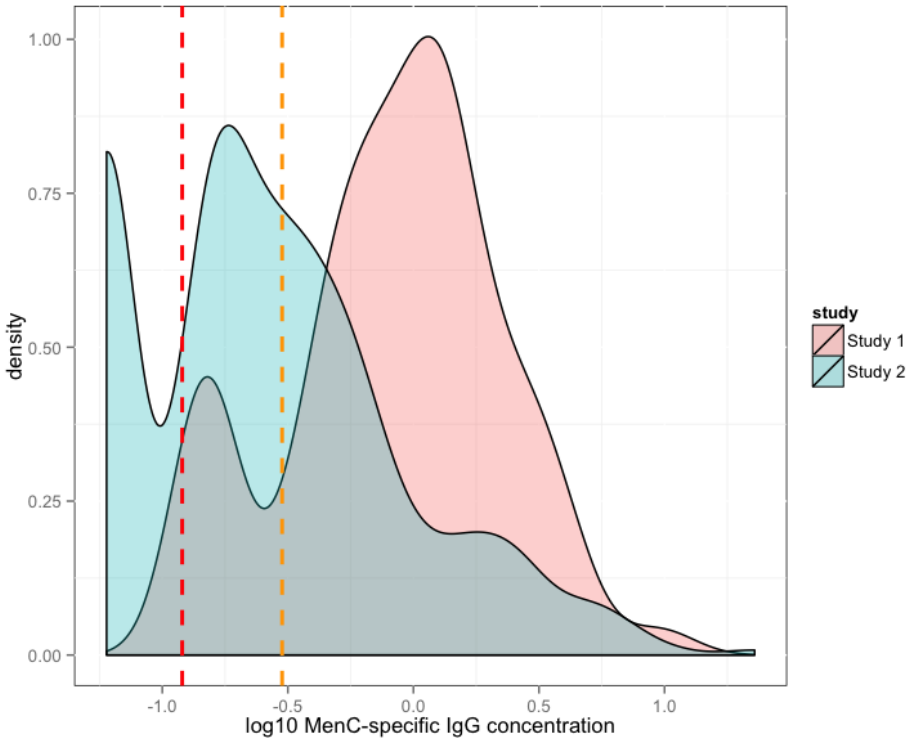


Figure 3.21: Density plot of $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations, following infant immunisation with MenC vaccine. Data are separated by vaccine study. Density plots were created from a 1d kernel density estimate, using the R package 'ggplot2' (240). The vertical dashed orange and red lines denote the lower limit of quantitation (LLQ) for study 1 and study 2, respectively.

3.4.4.3 Quality control

Following pre-genotyping DNA quantification 935 (91%) of the samples, from the participants immunised with MenC conjugate vaccine in infancy, were of adequate quantity (>50 ng/μl) for genome-wide genotyping. These participants were combined with the adolescents described in subsection '3.3.3.1 Adolescent participants' to make a total of 2061 individuals for QC analysis. Genotyping was conducted either on the Illumina® (Illumina, San Diego) OmniExpress12v1 (n =1317) or OmniExpress-12v1.1 microarray (n =744).

Eighty-one samples were excluded from analysis due to discrepancies between sex reported in participant information and genotype derived sex. The proportion of missing genotypes and heterozygosity rates for participants are shown in Figure 3.22. The cut-off for acceptable rate of genotype heterozygosity was set to the mean heterozygosity rate +/− two standard deviations, 28 samples failed this QC metric. Sixteen of the samples failed QC due to the proportion of missing genotypes, the cut-off was set to ≤0.03.

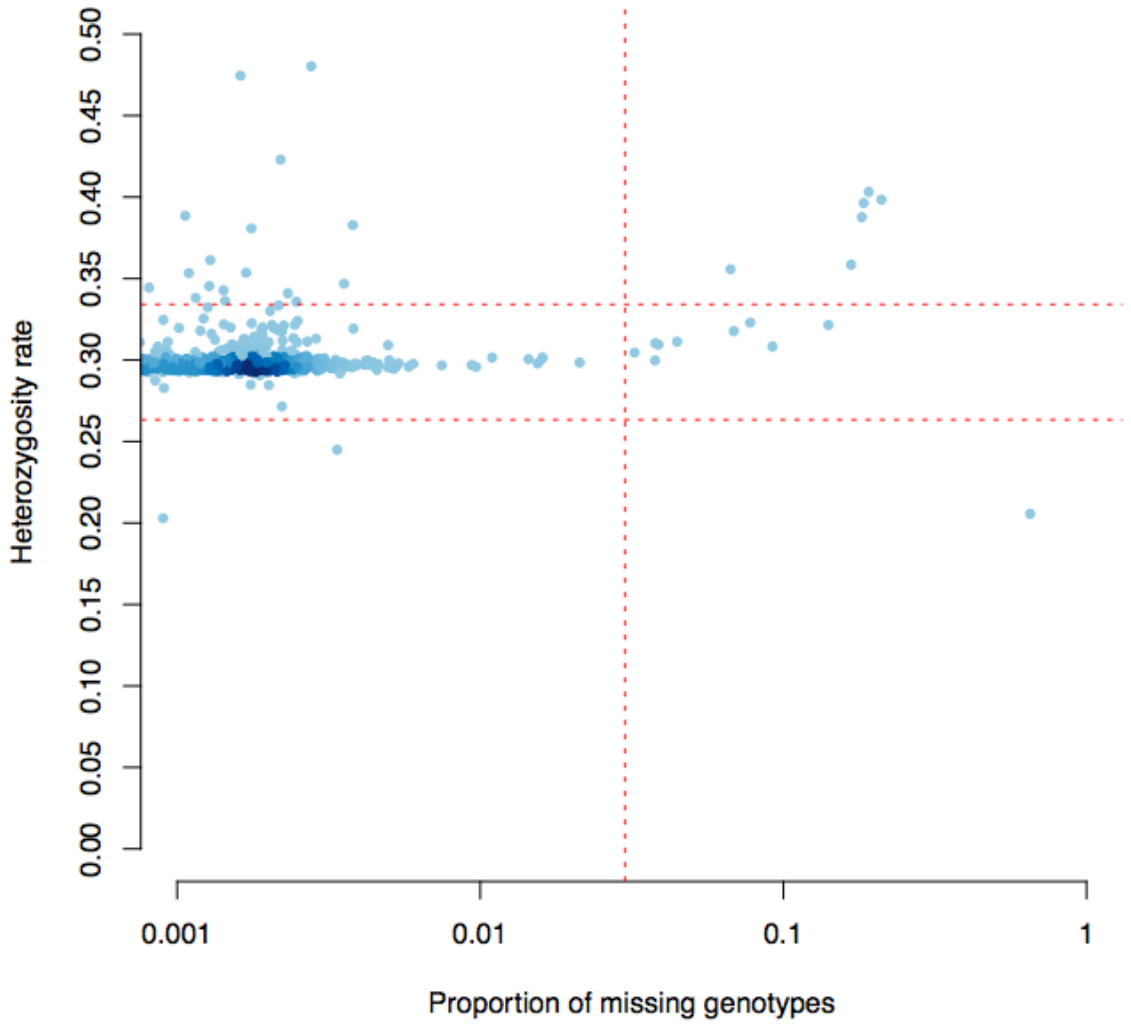
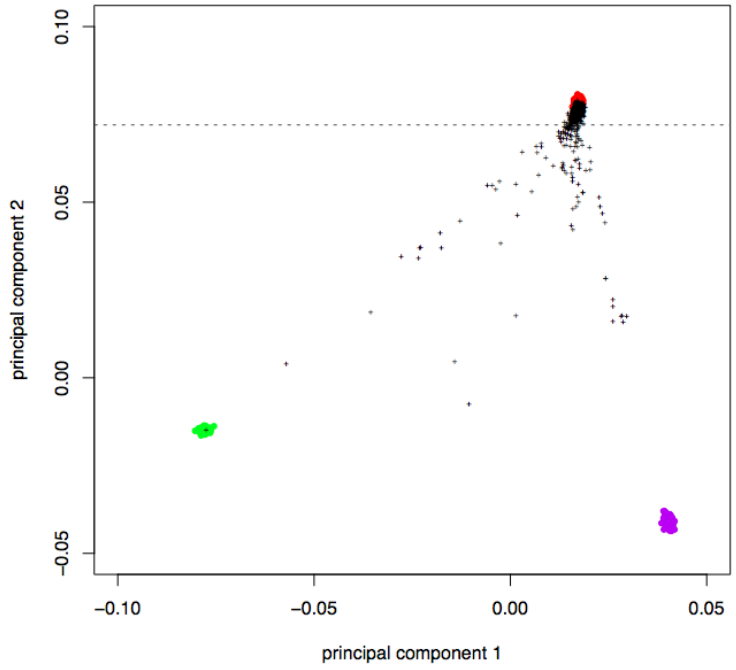
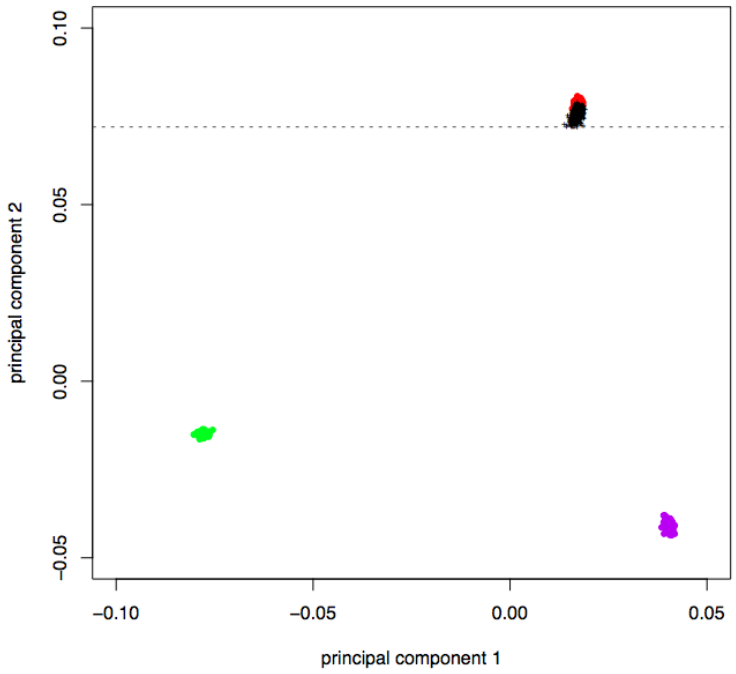


Figure 3.22: Plot of heterozygosity rate compared to the proportion of missing genotypes in the study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following childhood immunisation with MenC conjugate vaccine. Shading is used to highlight coordinates that are densely populated. The horizontal lines represent the mean heterozygosity rate +/− two standard deviations and the vertical line represents a proportion of missing genotypes of 0.03.

IBD analysis identified full-siblings and monozygotic twins within the genotyped cohort, only the participants with the lowest missing genotype proportion from these related participant sets were included in further analysis, this step resulted in the exclusion of 212 samples. Principal component analysis was performed on the IBD matrix and the first two PCs are shown in Figure 3.23. One hundred and thirty-two outlier samples were eliminated, using the PC2 threshold of divergent ancestry described by Anderson et al 2010 (235) (Figure 3.23).



(a)



(b)

Figure 3.23: Clustering of participants by the first and second principal components of the identity-by-descent matrix of genotype data in the study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following childhood immunisation with MenC vaccine. a) All participants. b) Filtered to exclude participants with a second principal component score <0.072 (horizontal grey dashed line). The HapMap3 reference samples are Han Chinese in Beijing and Japanese in Tokyo (purple), Yoruba in Ibadan (green) and Utah residents of northern and western Europe ancestry (red).

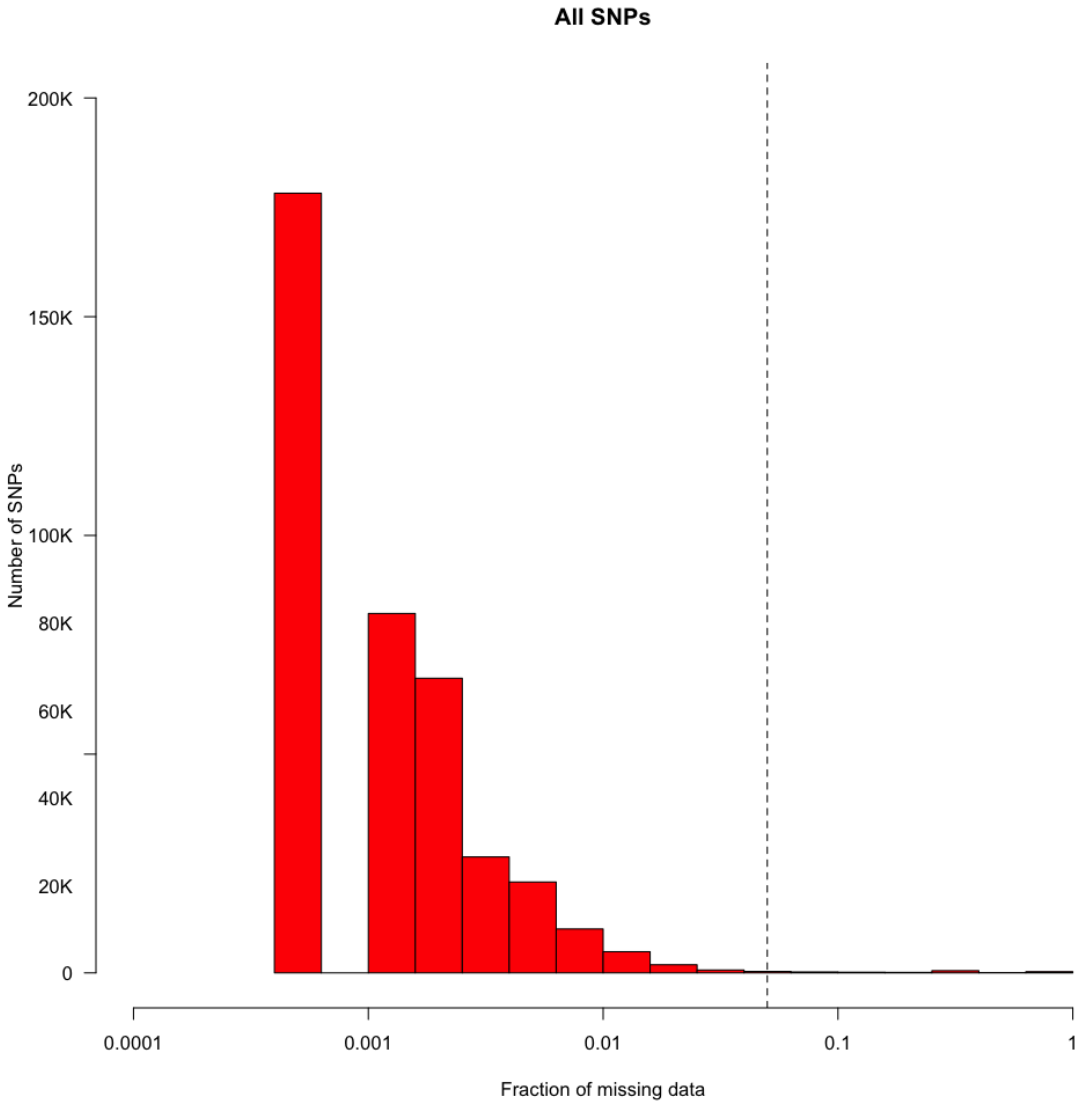


Figure 3.24: A histogram of the rate of missing genotypes from individuals that passed sample based quality control metrics in the genome-wide association study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following childhood immunisation with MenC conjugate vaccine. The dashed vertical line denotes the threshold (3%) for removal of SNPs from further analysis due to an excessive genotyping failure rate.

A histogram of the rate of genotype missingness is shown in Figure 3.24. The acceptance cut-off for missing data was set at $\leq$0.03, resulting in the elimination of 1984 (0.3%) markers. SNPs with a minor allele frequency of <0.01 and/or HWE p-value $< 1\times10^{-5}$ were also eliminated, leaving 622332 (87%) markers.

3.4.4.4 Association analysis

Following quality control 1683 individuals and 622332 markers remained for association analysis. MenC-specific SBA titres were available for 1585 and MenC-specific IgG concentrations were available for 1203 participants. Q-Q plots and associated λ inflation values for p-values from allelic association tests with MenC-specific serological measures did not show signs of excessive inflation (Figure 3.25 and Figure 3.26). There was a slight enrichment for p-values with smaller p-values ($< 1\times10^{-4}$) from the MenC-specific SBA titres regression model, potentially highlighting loci-specific associations (Figure 3.26). In contrast, the distribution of p-values resulting from the MenC-specific IgG concentration regression was comparable to the null (Figure 3.25).

None of the genotyped SNPs were associated with the persistence of MenC-specific IgG concentrations or SBA titres at the level of genome-wide significance ($p < 5\times10^{-8}$). However, 4 and 10 SNPs were associated with MenC-specific IgG concentrations and SBA titres at the level of suggestive association ($< 1\times10^{-5}$), respectively (Figure 3.27 and Figure 3.28). None of the SNPs with p-values suggestive of association with the persistence of MenC-specific IgG concentrations were within genes (Table 3.6). However, two of these SNPs were within 100 kb of *SYF2*, a cell cycle regulator containing eQTLs for immune-mediated diseases, such as ankylosing spondylitis and celiac disease (286). These SNPs were also within 100 kb of *RHD*, a highly polymorphic gene encoding the clinically significant Rhesus blood group D antigen. A single variant on chromosome 18 surpassed the suggestive association threshold and was within 100 kb of *ZNF407*, a zinc finger protein of unknown function. The variant on chromosome 8 that surpassed the suggestive association threshold was >100 kb from an annotated gene.

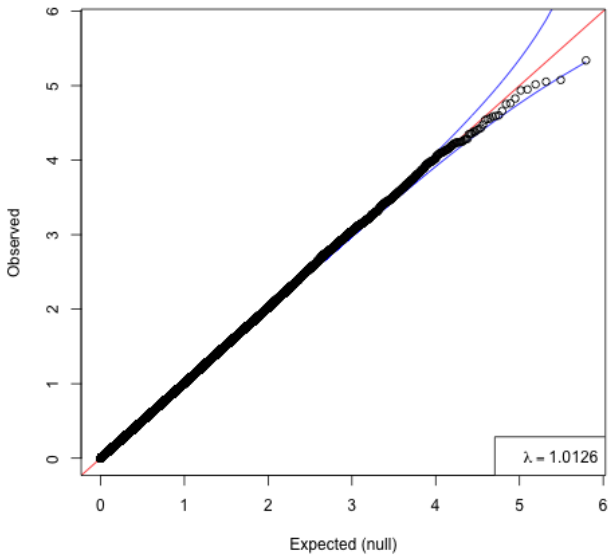


Figure 3.25: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation with MenC vaccine. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

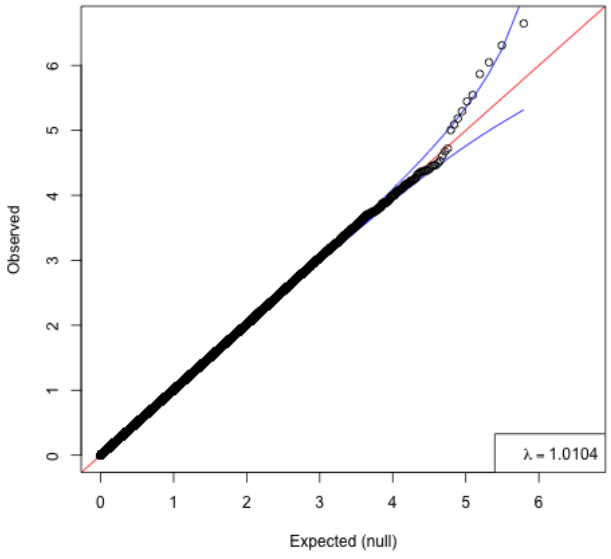


Figure 3.26: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay titres following childhood immunisation with MenC vaccine. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

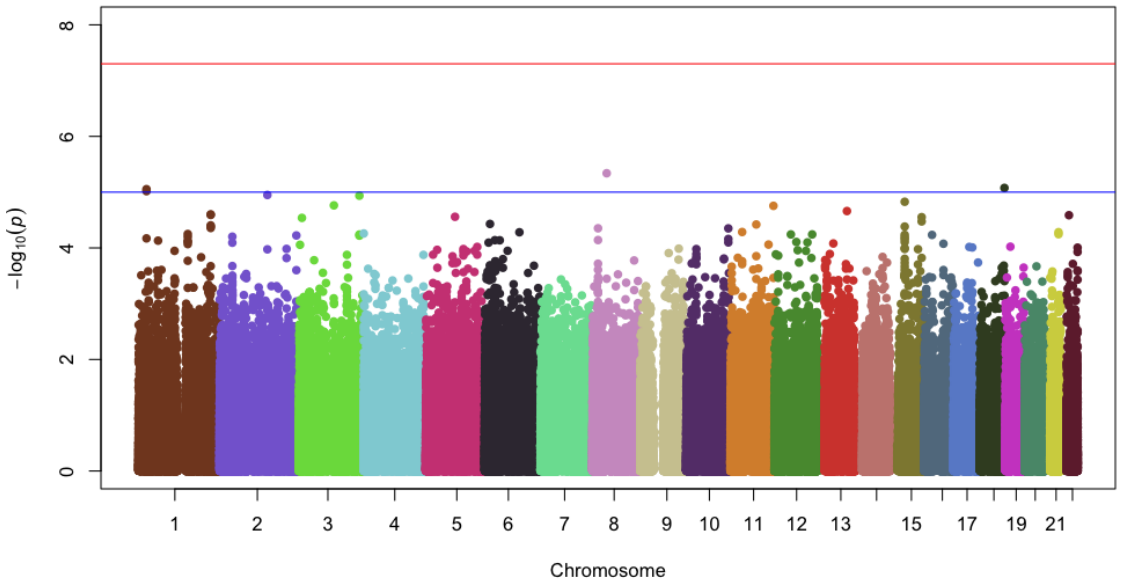


Figure 3.27: Manhattan plot of p-values from regression analysis between single-nucleotide polymorphisms and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation with MenC vaccine. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$)

Table 3.6: Single-nucleotide polymorphisms (SNPs) with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation with MenC vaccine. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	SNP	Position	A1	β	p-value	Gene	Neighbouring genes (<100kb)
1	rs7523543	25186716	G	0.14	8.85E-06		*SYF2,LOC391020,RSRP1,RHD*
1	rs2361150	25194299	G	0.14	9.60E-06		*SYF2,SDHDP6,LOC391020,RSRP1,RHD*
8	rs8185884	43626988	A	-0.4	4.58E-06		
18	rs1557400	74947450	G	0.43	8.40E-06		*ZNF407*

Chr= Chromosome, A1= Allele 1

Ten SNPs with p-values suggestive of association with the persistence of MenC-specific SBA titres were within five gene loci: *ALK*, *CNTN6*, *PDLIM1*, *NPAS3* and *SIRPG* (Table 3.7). Of these genes, only *CNTN6* was previously identified, at the level suggestive association, in the analysis of MenC-specific SBA titres following older childhood immunisation with MenC vaccine (section 3.3). *SIRPG* is annotated as having immune-related functions, including the positive regulation of T cell activation and leukocyte migration (287). *NPAS3* encodes a transcription factor and contains eQTLs for plasma levels of C-reactive protein (`http://www.ncbi.nlm.nih.gov/gap/phegeni`). The two remaining genes implicated in this analysis were *ALK*, encoding anaplastic lymphoma receptor tyrosine kinase a member of the insulin receptor superfamily, and *PDLIM1* encoding PDZ and LIM domain 1 a cytoplasmic protein associated with the cytoskeleton (`http://www.ncbi.nlm.nih.gov/gene/`).

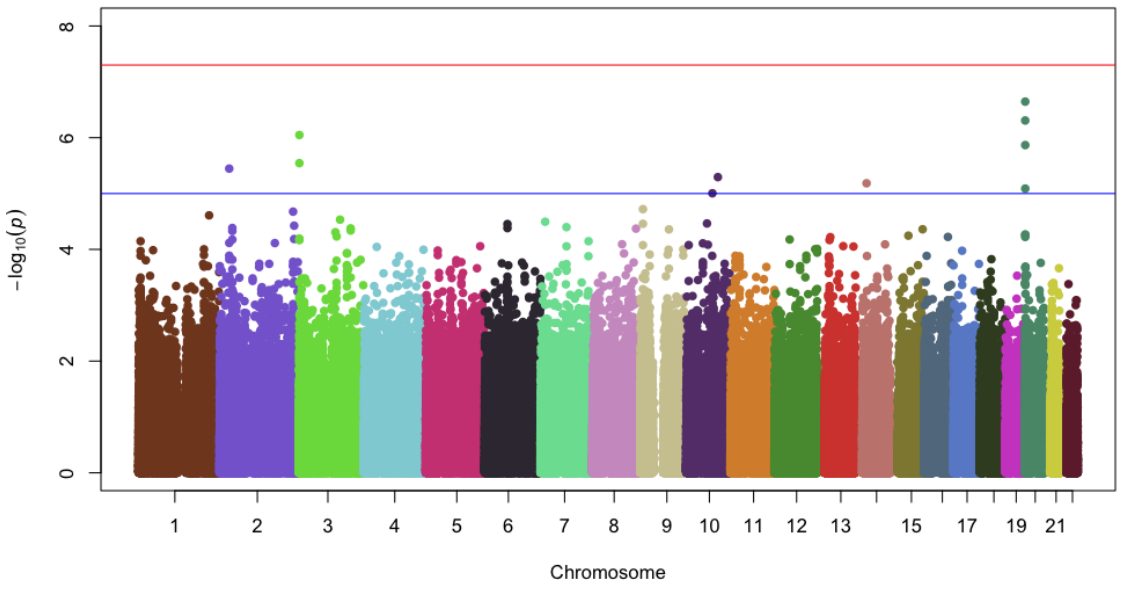


Figure 3.28: Manhattan plot of p-values from regression analysis between single-nucleotide polymorphisms and $\log_{10}$ transformed capsular group C meningococci (MenC)-SBA titres following childhood immunisation with MenC vaccine. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).

Table 3.7: Single-nucleotide polymorphisms with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following childhood immunisation with MenC vaccine. Gene information was derived from a NCBI search of SNP reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	SNP	Position	A1	β	p-value	Gene	Neighbouring genes (<100kb)
2	rs1881417	29511480	A	-0.2	3.58E-06	*ALK*	*ALK*
3	rs17785924	1389720	A	-0.23	2.86E-06	*CNTN6*	*CNTN6*
3	rs10510209	1405829	G	-0.22	8.95E-07	*CNTN6*	*CNTN6*
10	rs1369752	80248755	G	-0.18	9.89E-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs10509685	97021389	A	0.27	5.08E-06	*PDLIM1*	*SLIT1,LCOR*
14	rs1953447	33442259	A	0.15	6.55E-06	*NPAS3*	*NPAS3*
20	rs13045212	1612700	G	0.18	4.92E-07	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs17770331	1615308	A	0.18	2.26E-07	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs6135736	1687450	A	0.17	1.36E-06		*SIRPB3P,LOC101929010,SIRPB1,SIRPG*
20	rs4814511	1710272	G	0.16	8.16E-06		*SIRPB3P,SIRPB1,SIRPG*

Chr= Chromosome, A1= Allele 1

3.4.4.5 Heterogeneity analysis

The SNPs with p-values suggestive of associations with MenC-specific IgG concentrations displayed considerable effect size heterogeneity between infant and older childhood MenC conjugate vaccine recipients (Table 3.8). While all these SNPs showed concordant β values, between these subgroups, effect size estimates were smaller for the infant vaccine cohort. Two of these SNPs had β value 95% CIs that were non-overlapping between infant and older childhood vaccinees. Furthermore, the effect of these SNPs on the persistence of MenC-specific IgG may be exclusive to older children, as the 95% CIs for effect estimates from the infant vaccine subgroup crossed zero; the caveat being the infant vaccine cohort (n=312) was considerably smaller than the older childhood vaccinees (n=891). When a meta-analysis was conducted between vaccine age subgroups a different set of SNPs, than those identified in combined analysis, surpassed the suggestive level of statistical association (appendix Table 7.16).

Table 3.8: The degree of age-related heterogeneity observed in single-nucleotide polymorphisms with suggestive association (p $<1\times10^{-5}$) with the persistence of capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation. Meta p-values were calculated using fixed effect meta-analysis of individuals immunised with MenC conjugate vaccine in infancy (n=312) and in older childhood (n=891).

Chr	SNP	Position	Combined P	Older childhood β (95% CI)	Infant β (95% CI)	Meta P	I^2
1	rs7523543	25186716	8.85E-06	0.174 (0.09686,0.2511)	0.01444 (-0.06332, 0.0922)	0.0006809	87.74
1	rs2361150	25194299	9.60E-06	0.1782 (0.101,0.2555)	0.003367 (-0.07383, 0.08056)	0.001126	89.84
8	rs8185884	43626988	4.58E-06	-0.4663 (-0.6776, -0.2551)	-0.08994 (-0.3072, 0.1273)	0.000246	83.13
18	rs1557400	74947450	8.40E-06	0.5009 (0.257, 0.7448)	0.1907 (-0.03087, 0.4122)	7.61E-05	70.65

I^2= percentage of total variation accounted for by heterogeneity between the subgroups (242), Chr=Chromosome, P= p-value.

Several of the SNPs described to have suggestive association with the persistence of MenC-specific SBA titres showed substantial heterogeneity between individuals who received MenC as infants or older children (Table 3.9). While two SNPs showed low levels of heterogeneity (I^2 = 0−40%), three showed moderate (I^2 = 30−60%) and five showed substantial heterogeneity (I^2 >60%) (288). All SNP β values showed concordant directionality. However, the β estimate 95% CIs for three of these SNPs (rs17785924, rs10510209 and rs1369752) were non-overlapping, between the infant and older childhood vaccinees. Furthermore, the β estimate 95% CIs for these three SNPs crossed zero in the infant vaccine cohort. This could be interpreted as disparate effects for these SNPs between age groups, and potentially no effect in infancy. The caveat being the infant cohort (n=673) was smaller than the adolescent cohort (n=912), which influences the selection of the most significant SNPs from the overall regression model as well as the width of the confidence interval around the estimates made from the subgroup analysis. This analysis may support the generalisability of the SNPs that showed low levels of heterogeneity between these age subgroups, in particular two SNPs (rs10509685, rs1953447) had I^2 values equal to zero, which implies homogenous effects in infant and older childhood vaccinees. When a meta-analysis was conducted between age subgroups eight SNPs surpassed the suggestive level of association, four of which also exceeded this level in the overall regression (appendix Table 7.17). The four SNPs that had suggestive association in both analyses were within three genes: *PDLIM1*, *NPAS3* and *SIRPG*.

Table 3.9: The degree of age-related heterogeneity observed by single-nucleotide polymorphisms with suggestive association (p $<1\times10^{-5}$) with the persistence of capsular group C meningococci (MenC)-specific SBA titres. Meta p-values were calculated using fixed effect meta-analysis of subgroups immunised with MenC conjugate vaccine in infancy (n=673) and in older childhood (n=912).

Chr	SNP	Position	Combined P	Older childhood β (95% CI)	Infant β (95% CI)	Meta-P	I^2
2	rs1881417	29511480	3.58E-06	-0.2496 (-0.3743, -0.1248)	-0.1192 (-0.226, -0.01234)	2.54E-05	58.7
3	rs17785924	1389720	2.86E-06	-0.4014 (-0.5408, -0.262)	-0.00815 (-0.1265, 0.1102)	0.000173	94.37
3	rs10510209	1405829	8.95E-07	-0.3509 (-0.4788, -0.223)	-0.05167 (-0.1617, 0.05836)	2.59E-05	91.73
10	rs1369752	80248755	9.89E-06	-0.2805 (-0.3982,-0.162)	-0.05653 (-0.1586, 0.04556)	0.0001029	87.41
10	rs10509685	97021389	5.08E-06	0.2621 (0.09085, 0.4332)	0.2846 (0.1478, 0.4214)	4.22E-07	0
14	rs1953447	33442259	6.55E-06	0.138 (0.04335, 0.232)	0.1549 (0.07435, 0.2354)	2.33E-06	0
20	rs13045212	1612700	4.92E-07	0.2285 (0.1258, 0.3312)	0.09763 (0.0132, 0.1821)	6.19E-06	73.13
20	rs17770331	1615308	2.26E-07	0.2356 (0.133, 0.338)	0.09937 (0.01516, 0.1836)	3.48E-06	75.27
20	rs6135736	1687450	1.36E-06	0.2079 (0.1044, 0.3113)	0.1057 (0.02006, 0.1913)	1.22E-05	55.04
20	rs4814511	1710272	8.16E-06	0.1819 (0.08045, 0.2833)	0.09614 (0.0141, 0.1782)	6.44E-05	39.76

I^2 describes the percentage of total variation accounted for by heterogeneity between the subgroups (242).

3.4.4.6 Imputation

A total of 34,026,329 genetic variants were included in genotype imputation analysis. These data were filtered for SNPs with a minor allele frequency >0.02, HWE p-values $>1\times10^{-9}$ and an information score >0.8. Subsequently, 6,699,406 and 6,699,405 SNPs surpassed these thresholds for MenC-specific IgG concentrations and MenC-specific SBA titres regression analysis, respectively.

The Q-Q plot of p-values and corresponding λ inflation value, resulting from the regression of allele dose and MenC-specific IgG concentration did not appear to deviate excessively from the null distribution (Figure 3.29). However, there was a slight deflation of smaller p-values, which can be indicative of imputed SNPs in low LD with genotyped SNPs leading to uncertainty and decreased variance of allele dosage.

A Manhattan plot of p-values resulting from association analysis between imputed SNPs and MenC-specific IgG concentrations is displayed in Figure 3.30. Twenty-four SNPs had suggestive associations with the persistence of MenC-specific IgG concentrations following genotype imputation, thirteen of which were within six genes: *SENP2*, *NBEA*, *LRFN5*, *MCTP2*, *GALR1* and *GGT5* (appendix Table 7.14).

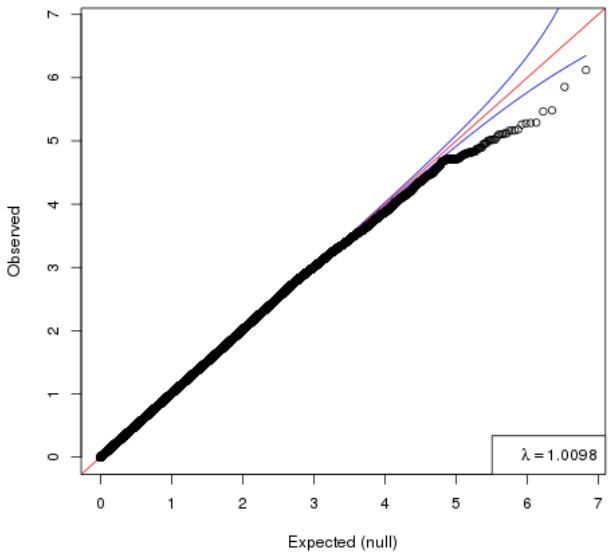


Figure 3.29: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation with MenC vaccine. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

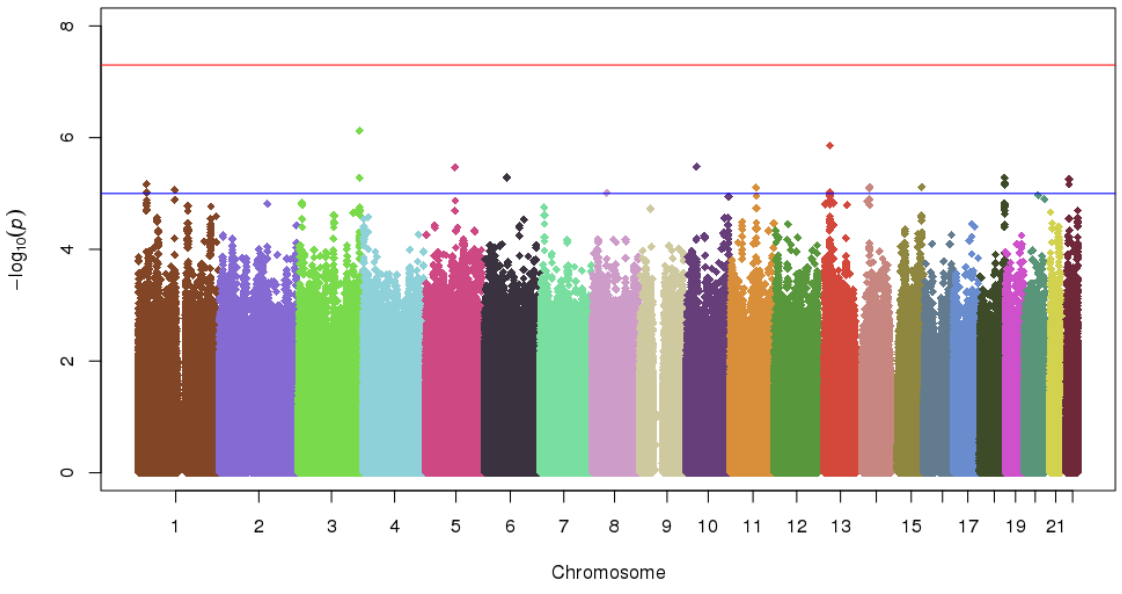


Figure 3.30: Manhattan plots showing genome-wide p-values of regression analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation with MenC vaccine. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).

The Q-Q plot of p-values resulting from regression analysis between imputed SNP allele dose and the persistence of MenC-specific SBA titres showed no signs of excessive inflation (Figure 3.31). However, there appeared to be slight enrichment for smaller p-values. None of the SNP p-values from this regression analysis surpassed the level of genome-wide significance ($p < 5 \times 10^{-8}$).

A Manhattan plot of p-values resulting from association analysis between imputed SNPs and MenC-specific SBA titres is displayed in Figure 3.32. Seventy-four SNPs had suggestive association with the persistence of MenC-specific SBA titres following imputation (Figure 3.32 and appendix Table 7.15). The two particularly noteworthy loci on chromosome 10 and 20, that contain the smallest SNP regression p-values, correspond to gene loci *PDLIM1* and *SIRPG*, respectively (Figure 3.32).

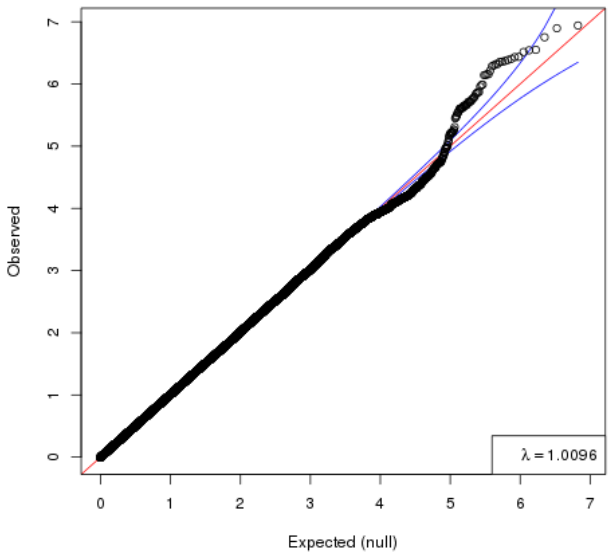


Figure 3.31: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci-specific (MenC)-IgG concentrations following childhood immunisation with MenC vaccine. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

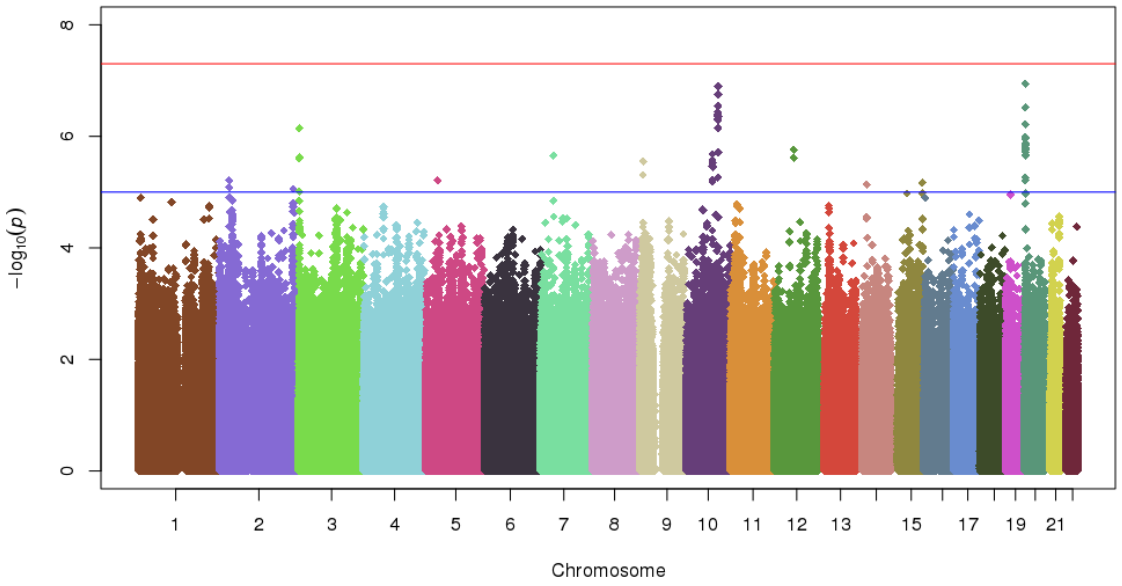


Figure 3.32: Manhattan plots showing p-values of regression analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation with MenC vaccine. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).

3.4.5 Discussion

This section described a GWAS to assess the relationship between genetic variants and the persistence of vaccine-specific serological immunity following childhood immunisation with MenC conjugate vaccine. While no genetic markers were found to surpass the level of genome-wide statistical significance ($p < 5\times10^{-8}$), 24 and 74 SNPs displayed p-values suggestive of association ($p < 1\times10^{-5}$) with the persistence of MenC-specific IgG concentrations and MenC-specific SBA titres, respectively. These associations highlighted a number of variants within genes, including genes with annotated immunological function, such as *SIRPG* and *GGT5*.

Twenty-four SNPs had regression p-values suggestive of association with the persistence of MenC-specific IgG concentrations, of which thirteen resided within six genes: *SENP2*, *NBEA*, *LRFN5*, *MCTP2*, *GALR1* and *GGT5*. Of these genes, *GGT5* was the only for which the protein product has been assigned an overtly immunological function, gamma-glutamyltransferase 5

(GGT5) converts leukotriene C4 to leukotriene D4 and has an important role in the acute inflammatory response (289).

Seventy-four SNPs had regression p-values suggestive of association with the persistence of MenC-specific SBA titres, of which fifty resided within nine genes: *ALK*, *CNTN6*, *FAM124B*, *LINC00265*, *NPAS3*, *PDLIM1*, *PTPRD*, *SIRPG* and *SLC1A3*. Two loci, present on chromosome 10 and 20, contained SNPs with the smallest p-values in regression analysis of the persistence of MenC-specific SBA titres. The first of these was within *PDLIM1*, a gene that encodes PDZ and LIM domain 1, a cytoplasmic protein associated with the cell cytoskeleton (`http://www.ncbi.nlm.nih.gov`). The protein product of this gene is strongly stained in a number of tissues, including lymphoid tissue (`http://www.proteinatlas.org`). The other locus was within *SIRPG*, which encodes signal-regulatory protein gamma a protein involved in T cell activation and leukocyte migration. In addition, *SIRPG* contains variants associated with susceptibility to type 1 diabetes, an immune-mediated disease (290).

Interestingly, SNPs with regression p-values suggestive of association with MenC-specific SBA titres were also demonstrated within *PTPRD*, a gene recently shown to harbour variants associated with rubella specific IFN-γ responses following MMR vaccine (28). *PTPRD* encodes for protein tyrosine phosphatase receptor type D, which is a signalling protein associated with the regulation of cellular processes such as cell growth and differentiation. This gene has yet to be assigned any immunological functions and the gene transcript appears to be expressed at low levels in whole blood and lymphocytes (`http://www.gtexportal.org`). Nevertheless, the protein product of this gene is stained in many tissues such as lymph nodes, bone marrow and spleen (`http://www.proteinatlas.org/`).

Considerable disparity is seen in both the magnitude and the persistence of serological immunity following infant or older childhood immunisation with MenC conjugate vaccine (3, 212, 213, 214). A potential limitation introduced by the inclusion of infant and older childhood vaccinees is the requirement for the effects of genetic variants to span the maturational changes to the immune system occurring in early life (275). Heterogeneity analysis conducted in this section described inconsistencies in SNP effect estimates between infants or older childhood

vaccine recipients, for several of the SNPs with the greatest overall statistical associations. While this heterogeneity may be a function of age of vaccination, they may also be chance findings. Despite various age-related phenotypic differences in immune function, the paucity of age stratified data on gene expression and regulation, makes the interpretation of these findings challenging. On the contrary, associations with low levels of heterogeneity between age subgroups may have generalisable effects on the persistence of serological immunity following childhood immunisation with MenC conjugate vaccine.

There were a number of further potential limitations to this study. Firstly, this section described the relationship between genetic variants and the persistence of MenC immunity 7 months to 12 years after vaccination; limiting the range of time since vaccination would likely have increased power substantially. Secondly, participants received doses of one, or more, of five MenC conjugate vaccine formulations. While these vaccines are broadly similar in terms of polysaccharide content and adjuvantation, the use of two disparate carrier proteins could be relevant. Conjugation of polysaccharides to carrier proteins results in vaccine formulations that can induce affinity maturation and immunological memory, characteristics of T cell-dependent immune responses. The precise mechanisms involved in immune responses to glygoconjugates have not been fully elucidated, but the covalent linkage of polysaccharide to the protein carrier has been shown to be essential (291). Initially it was thought that the protein carrier resulted in 'bystander' T cell help due to proximal recruitment of activated T cells, with data from both animal and human studies correlating carrier and polysaccharide immune responses (292, 293). More recently, there are data showing carrier-mediated presentation of polysaccharide on MHC class II and the induction of polysaccharide-specific T cells (294, 295). The genetic determinants influencing interactions to different carrier proteins such as tetanus toxoid and CRM_{197} may be disparate, introducing the possibility of conjugate-dependent genetic associations with polysaccharide-specific immunity.

Thirdly, participants included in this study received one of five possible vaccine regimens; two or three infant doses with or without a booster dose at 1 year of age, or a single dose of MenC in older childhood (6-15 years of age). Given the variety of regimens and vaccine doses

it is necessary to make the assumption that genetic determinants influencing the persistence of immunity following the last dose are consistent regardless of the prior dosing regimen.

Hitherto, association studies of immune responses to conjugate vaccines have adopted candidate gene genotyping approaches, and few of the findings have been replicated. Recently, genome-wide approaches have been used to assess immunity to three vaccines: HBV, smallpox vaccine and MMR vaccine (22, 25, 26, 27, 28). The aforementioned hepatitis B vaccine was a recombinant vaccine containing hepatitis B surface antigen (HBsAg) adsorbed onto aluminium hydroxide, administered intramuscularly. The MMR vaccine was a trivalent vaccine containing live-attenuated viruses administered subcutaneously and the smallpox vaccine was a live vaccinia virus administered by puncturing the skin multiple times with a bifurcated needle dipped into the vaccine solution. The contribution of common genetic variants and pathways to responses to different classes of vaccines is unclear. However, GWASs of immune-related diseases have shown a high degree of pleiotropy between seemingly disparate immunological diseases (258).

Studies following HBV have demonstrated genome-wide statistically significant associations between polymorphisms in HLA loci and post-vaccination HBsAg-specific IgG concentrations in two independent Asian populations (22, 25). In particular, these studies found significant associations between alleles proximal to *HLA-DRB1*, which is consistent with twin data suggesting that around 40% of the heritability of HBV IgG responses may be accounted for by variants at this gene locus. However, these studies did not show any genome-wide associations outside the HLA region, despite twin data suggesting that the majority of the genetic contribution to post-vaccination HBsAg-specific IgG concentrations were non-HLA (6).

The GWAS of MMR vaccine responses assessed rubella-specific cellular immune responses (28). This study found genome-wide significant associations between variants in *PTPRD* and rubella virus induced IFN-γ release from PBMCs, derived from adolescents and young adults approximately six years after MMR (28). All the genome-wide significant findings in this study were within *PTPRD* and were low-frequency variants, minor allelic frequency (MAF) of $\sim$0.01 (28). Often GWAS data are filtered to remove variants of a MAF less than 1-2%, as low-frequency

variants may inflate false-positive results; although there are some suggestions that the removal of SNPs with low MAFs may not be appropriate (296). The paucity of variants with associations at the level of genome-wide significance in this study may be the result of the modest number of individuals genotyped (n = 897).

The GWAS of immunity following smallpox vaccine was carried out in a mixed ethnic population. In the Caucasian subgroup (n = 580), the largest single ethnic group, no SNPs were associated with neutralising antibody titres at the level of genome-wide significance; however, five SNPs did surpass this threshold in relation to smallpox-specific $CD8^+$ IFN-γ ELISpots (26, 27). The genes closest to the significant SNPs were *PSD3*, *RASA1*, *RGS6* and *TRNAQ32*. Interestingly, the SNPs with the greatest statistical association with neutralising antibody and cellular measures of immunity to vaccinia virus did not appear to be consistent (26, 27). While this disparity may highlight genetic determinants with differing roles in humoral and cellular responses to smallpox vaccine, a modest number of participants were analysed and only data on the most statistically significant ($p < 5 \times 10^{-7}$) SNPs were publicly available for comparison. In contrast to the HBV GWAS no genome-wide statistically significant associations were observed in the HLA region, for either smallpox-specific humoral or cellular immunity, despite previous findings from candidate gene studies implicating HLA alleles in determining the magnitude of immunity to smallpox vaccine (297).

Interestingly, variants in *PTPRD* were found to be associated with rubella specific IFN-γ responses following MMR (28) and the persistence of MenC-specific SBA titres (this chapter). While the most statistically significant variants identified in the MMR and MenC vaccine studies did not overlap, the fact that variants within this gene have been highlighted in immunity to two vaccines suggests some underlying role in vaccine immunity.

Hitherto, there are few data correlating immunological responses to different vaccines, correlations that could indicate shared genetic determinants. Comparison of immune responses to HBV and live oral polio vaccine in Gambian infants showed no correlation between post-vaccination HBsAg-specific IgG concentrations and poliovirus neutralising antibody titres (7). Conversely, the magnitude of IgG responses following two toxoid vaccines, tetanus (TT) and

diphtheria (DT), showed a strong positive correlation in the same cohort (7). Furthermore, a weak but statistically significant correlation was seen between anti-*Haemophilus Influenzae* type b IgG and, anti-TT and anti-DT IgG concentrations post-vaccination (8). Further GWASs are needed to gain insight into common and vaccine-specific genetic variants involved in immunity to vaccines.

3.4.6 Conclusion

Despite the effectiveness of the MenC conjugate vaccine, there is considerable variability in childhood immune responses to this vaccine. Of particular importance vaccine-induced immunity wanes rapidly in the young, and understanding the determinants of this decline could have important implications. In this section a GWAS was conducted highlighting a number of genetic variants that may influence the persistence of vaccine-induced MenC-specific serological immunity. This section represents stage 1 of a two-stage genetic association study, and was used to select a subset of genetic markers for genotyping in stage 2 (appendix 7.4).

3.5 Chapter summary

Initially, this chapter described the relationship between genetic variants and the persistence of MenC-specific serological immunity in a cohort of adolescents, vaccinated as old children. Two genes, *CNTN6* and *ENKUR*, contained SNPs associated at the level of genome-wide significance with the persistence of MenC-specific SBA titres following older childhood immunisation with MenC conjugate vaccine. Furthermore, a number of additional SNPs showed suggestion of association with the persistence of MenC-specific SBA titres including SNPs within *GCSAM*, a gene highly expressed in germinal centre lymphocytes, and product of which is involved in B cell receptor signalling and lymphocyte migration. Analysis was then expanded to include individuals vaccinated with MenC conjugate vaccine as young children, in a two-stage design. In stage 1, vaccinees were densely genotyped; these data were then used to select a subset of genetic markers for genotyping in stage 2. While stage 2 of the study has not yet been

completed, a number of SNPs with p-values suggestive of association with the persistence of MenC-specific immunity, following childhood vaccination, were highlighted in stage 1. These included SNPs within genes with annotated immunological function such as *SIRPG* and *GGT5*. The completion of stage 2 of this study may provide the power to demonstrate statistical associations at the level of genome-wide significance, and such findings would make excellent candidates for functional evaluation.

Chapter 4

Genome-wide association study of *Haemophilus influenza* type B and tetanus toxoid vaccine immunity

4.1 Overview of chapter

Chapter 3 described a genome-wide association study of the persistence of MenC-specific serological immunity following childhood immunisation with MenC conjugate vaccine. In this chapter, this approach was expanded to assess the genetic determinants of the persistence of serological immunity following two other vaccines routinely administered to children in the United Kingdom, *Haemophilus influenzae* type b (Hib) and tetanus toxoid (TT). The participants described in this section are a subset of those described in chapter 3, for whom relevant serology was available. A two-stage GWAS design was adopted. Pleiotropy in serological responses to disparate vaccine antigens was also explored.

4.2 Chapter aims

i) Describe a two-stage genetic association study of the persistence of serological immunity to *Haemophilus influenza* type b following childhood immunisation

ii) Describe a two-stage genetic association study of the persistence of serological immunity to tetanus toxoid following childhood immunisation

iii) Describe pleiotropic loci involved in immunological responses to childhood immunisations

4.3 Genome-wide association study of the persistence of serological immunity to *Haemophilus influenzae* type b following immunisation - Stage 1

4.3.1 Background

4.3.1.1 *Haemophilus influenzae*

Haemophilus influenzae is a Gram-negative coccobacillus that can cause severe disease, particularly in young children (82, 184, 302). Six serotypes, classified by their capsular polysaccharide (a-f) cause the majority of disease, although non-typeable (non-encapsulated) strains occasionally cause invasive disease (82, 303). In the pre-vaccination era the majority of invasive strains isolated from patients in the UK were type b, which are encapsulated with polyribosyl ribitol phosphate (PRP) (184, 304). The most common clinical presentation of invasive *Haemophilus influenzae* type b (Hib) is meningitis (>50%), often with bacteraemia (82, 304). Other invasive presentations include epiglottitis (15%), bacteraemia without other infection (10%), septic arthritis, osteomyelitis, cellulitis, pneumonia and pericarditis. Hib meningitis case fatality is 4-5%, with 8-11% of survivors having permanent neurological sequelae (305, 306).

4.3.1.2 *Haemophilus influenzae* type b disease epidemiology and vaccines

Prior to the introduction of the Hib vaccine the incidence of invasive Hib disease in those age <5 years was 35.5 per 100,000 (307). Hib disease was rarely seen in those under 4 months of age but increased thereafter with the peak incidence at 6-7 months coinciding with the low levels of serum anti-PRP antibody, presumably due to the loss of maternal antibody (308, 309). The first Hib vaccine was a plain PRP polysaccharide formation, which was not effective in those under 18 months of age (310). Conjugation of PRP oligosaccharide to a protein carrier resulted in an immunogenic vaccine that was efficacious in infants as young as 2 months (311). The Hib conjugate vaccine was introduced into the routine UK infant immunisation schedule, as a three-dose regimen, in 1992 (82). Following the introduction of the Hib vaccine cases of invasive disease in individuals under 5 months of age fell by over 95% (82). From 1999, a gradual increase in Hib disease was observed, which was attributed to the use of a less immunogenic combination vaccine and waning population immunity (312, 313, 314). Interestingly, carriage rates remained low despite the increase in disease, with a cross-sectional study of 384 infants attending day care nurseries, in 2002, not detecting a single Hib isolate (315). A Hib vaccine booster campaign was initiated in 2003 for children aged 6 months to 4 years, and a booster dose of Hib vaccine was introduced into the routine UK immunisation schedule at 1 year of age in 2006 to address concerns about waning immunity (82).

4.3.1.3 Genetic studies of *Haemophilus influenzae* type b vaccine immunity

Disparities in Hib vaccine efficacy estimates have been observed between ethnic populations, and a twin study demonstrated serological responses to Hib vaccine to be heritable (8, 316, 317). Despite the apparent genetic contribution to Hib vaccine responses, there are few data describing the genetic determinants underlying these observations (318, 319, 320, 321). Hitherto, the studies of genetic factors influencing serological immunity following Hib vaccine have focused on immunoglobulin (Ig) GM and KM allotypes, which are determined by Ig γ (heavy) and κ (light) chains, respectively (318, 319, 320). Several studies have associated immunoglobulin GM and KM allotypes with immune responses to purified polysaccharide vaccines such as Hib polysac-

charide (318, 319, 320). The mechanisms underlying the association between Ig allotypes, which are linked predominantly to variants in the constant regions of immunoglobulin, with immune responsiveness to polysaccharides have not been clearly elucidated. It may be that these allotypes are proxies for variable region variants, which primarily determine antigen-specificity and affinity. Alternatively, variants in the constant regions may directly influence responses via effector functions such as Fc binding, complement activation and serum half-life (322). Moreover, there are data suggesting that the constant region also influences immunoglobulin affinity and specificity (323).

A study investigating the genetic factors influencing Hib vaccine failure investigated a small number of candidate functional variants finding two variants, within *MAL/TIRAP* and *IL-10*, to be associated with Hib vaccine failure (47). Hib vaccine failure, invasive Hib disease despite vaccination, is extremely rare making replication of this study challenging. However, it is conceivable that the genetic determinant influencing vaccine failure may overlap with those determining the quantity of vaccine-induced immunological endpoints. The following section describes stage 1 of a two-stage GWAS of the persistence of PRP-specific serological immunity following childhood immunisation with Hib vaccine.

4.3.2 Aim

The aim of this section was to describe stage 1 of a two-stage genetic association study of the persistence of serological immunity to *Haemophilus influenza* type b following childhood immunisation.

4.3.3 Materials and methods

4.3.3.1 Stage 1 participants

Participants included in this section were a subset of those described in chapter 3, for whom post-vaccination PRP-specific IgG concentrations, or sera, were available. This subset comprised of adolescents from cohort 2, as well as children from studies 1, 2 and 3, described in

subsection '3.4.3.1 Stage 1 participants'.

4.3.3.2 Vaccines

Participants described in this section received at least one dose of a vaccine containing Hib capsular polysaccharide conjugated to a protein carrier, either in a clinical trial setting or as part of their routine clinical care. The Hib tetanus toxoid conjugate vaccine (Hib-TT) was initially introduced into the UK immunisation schedule in 1992, administered concomitantly with diphtheria-tetanus-whole cell pertussis (DTwP) vaccine and oral polio vaccine (OPV). In 1996, this schedule was modified to include the combination DTwP/Hib vaccine. The diphtheria-tetanus-acellular pertussis/Hib (DTaP/Hib) vaccine was increasingly used from 1999, accounting for ∼50% of Hib combination vaccines administered between 2000 and 2002 (270). Furthermore, the combination diphtheria-tetanus-acellular pertussis/ inactivated poliovirus/Hib (DTaP/IPV/Hib) vaccine became available in the UK in 2004.

The children described in this section received three doses of Hib capsular polysaccharide conjugated to tetanus toxoid: at 2, 3 and 4 months of age in the form of Hib-TT or combination DTwP/Hib, DTaP/Hib or DTaP/IPV/Hib vaccines, depending on vaccine availability for their respective birth cohort. A subset of these children also received a booster Hib conjugate vaccine prior to serological assessment. This subset included participants who were 6 months to 4 years of age in 2003 that received a booster dose DTwP-Hib or DTaP-Hib vaccine, as well as children who were 12 months since 2006 that received a booster dose of Hib-MenC-TT (Menitorix). The adolescent cohort received either a three dose infant Hib conjugate vaccine schedule at 2, 3 and 4 months of age or a single dose as toddlers during the catch-up campaign of 1992. Hib immunisation information was obtained from clinical study data, or from the child health computer department.

4.3.3.3 Immunological measurements

The VEU, Manchester, measured PRP-specific IgG concentrations for study 2 and study 3, using a multiplexed bead assay in accordance with an internal protocol. This assay was based

on a Triplex fluorescent bead assay designed to quantitatively measure serum IgG to PRP, TT and diphtheria toxoid (DT). The lower limit of quantitation (LLQ) for this assay was 46 ng/ml, with measures below this recorded as half the LLQ. GSK Biologicals, Belgium, measured PRP-specific IgG concentrations for study 1, using a standardised ELISA based method to assess serum IgG to PRP. The LLQ for this assay detection was 150 ng/ml with measures below this recorded as half the LLQ. Serological assessment of PRP-specific IgG concentration was completed in-house (Oxford Vaccine Group) for the adolescents (cohort 2), using a standardised ELISA based assay (324). The LLQ for this assay was 100 ng/ml, with measures below this recorded as half the LLQ.

4.3.3.4 Genotyping and imputation

Genotyping was performed as described in subsection '3.4.3.5 Genotyping'. Imputation was completed as described in subsection '3.3.3.8 Imputation'.

4.3.3.5 Quantitative trait loci analysis

A linear regression model was used to assess the relationship between SNP allele dosage and the persistence of PRP-specific IgG concentrations, with age at vaccination, time since vaccination, MDS components 1-3, and $K-1$ dummy variables coding for study cohort as covariates. Linear regression analysis assessed directly typed and imputed SNPs, with posterior probability of each genotype being used to incorporate uncertainly for the latter.

4.3.4 Results

4.3.4.1 Stage 1 demographics

An overview of stage 1 participant demography is displayed in Table 4.1.

Table 4.1: Demographics in stage 1 of the genome-wide association study of the persistence of polyribosyl ribitol phosphate-specific IgG concentrations following childhood immunisation with *Haemophilus influenzae* type b vaccine. Participants were previously described in subsection '3.4.3.1 Stage 1 participants'

Study	Male/Female	Time since vaccination days* (IQR)	Age at vaccination days* (IQR)	Reference
Cohort 2	221/197	4551 (4518 - 4607)	524.0 (165.5 - 1055.5)	(214)
Study 1	37/24	275.5 (263.3 - 287.8)	123 (118 - 128)	(145)
Study 2	141/109	763.0 (400.5 - 2562.5)	365 (120 - 365)	(282)
Study 3	121/117	256 (248.5 - 264)	122 (118 - 127)	(283)

IQR= interquartile range, * = median.

4.3.4.2 Serology

The distributions of $\log_{10}$ transformed PRP-specific IgG concentrations approached normality (Figure 4.1). Inter-study assay variability is an important factor in interpreting these data, in particular the LLQ varied between studies. The LLQ was 150 ng/ml for the assay completed in study 1, 100 ng/ml for the assays utilised for cohort 2 and 46 ng/ml for the assays conducted in studies 2 and 3.

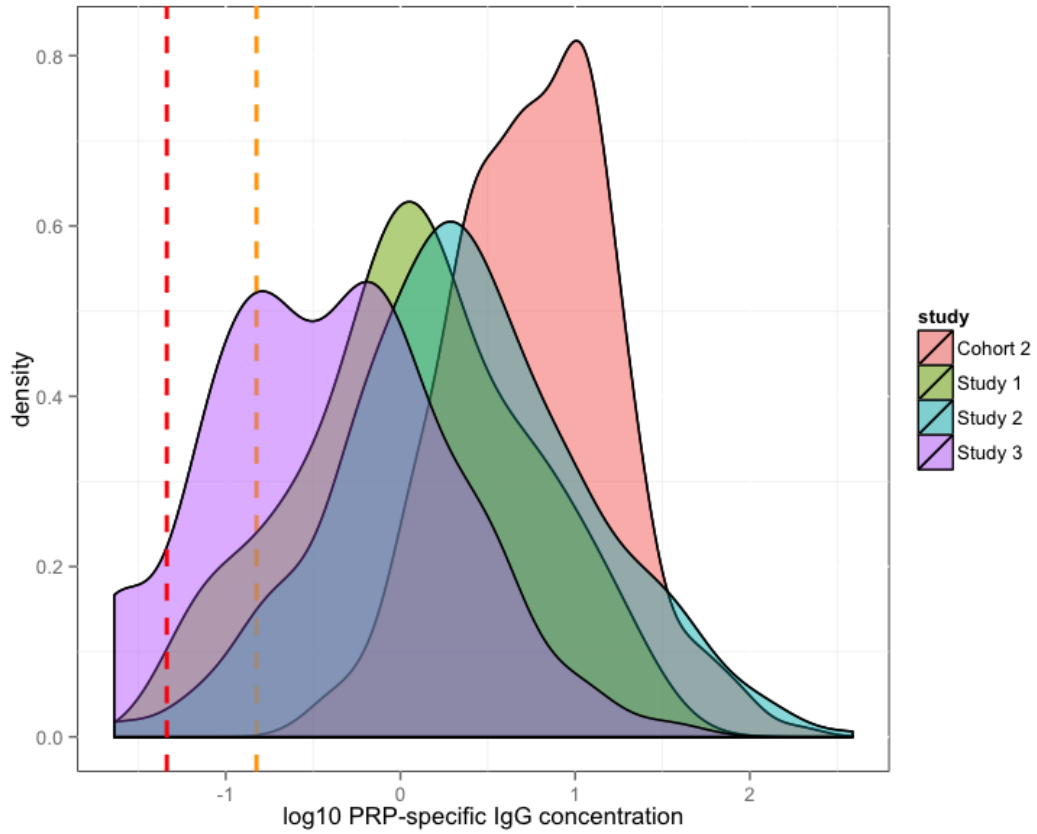


Figure 4.1: Density plots of $\log_{10}$ transformed polyribosyl ribitol phosphate-specific IgG concentrations for stage 1 participants in the genome-wide association study of persistence of serological immunity to *Haemophilus influenzae* type b following immunisation. The orange vertical dashed line represents a lower limit of quantitation (LLQ) of 150 ng/ml used in study 1, and the red vertical dashed line represent a LLQ of 46 ng/ml used in studies 2 and 3. The LLQ of 100 ng/ml for the assays completed on the cohort 2 is not displayed as all these participants surpassed this value. Density plots were created from a 1d kernel density estimate, using the R package 'ggplot2' (240).

4.3.4.3 Association analysis

The imputed data of 34,026,329 variants were filtered for SNPs with a MAF >0.02, HWE $>1\times10^{-9}$ and an information score >0.8, resulting in a set of 6,697,753 SNPs for regression analysis. Of the 1683 participants passing QC (subsection '3.4.4.3 Quality control'), 967 had PRP-specific IgG measures and were included in regression analysis.

The Q-Q plot of p-values resulting from the regression of the imputed SNP dataset and PRP-specific IgG concentrations did not deviate considerably from the null distribution (Figure 4.2). However, there was notable excess in the tail ($p< 1\times10^{-4}$) of the distribution, which may reflect enrichment by loci with true relationships with this phenotype (Figure 4.2).

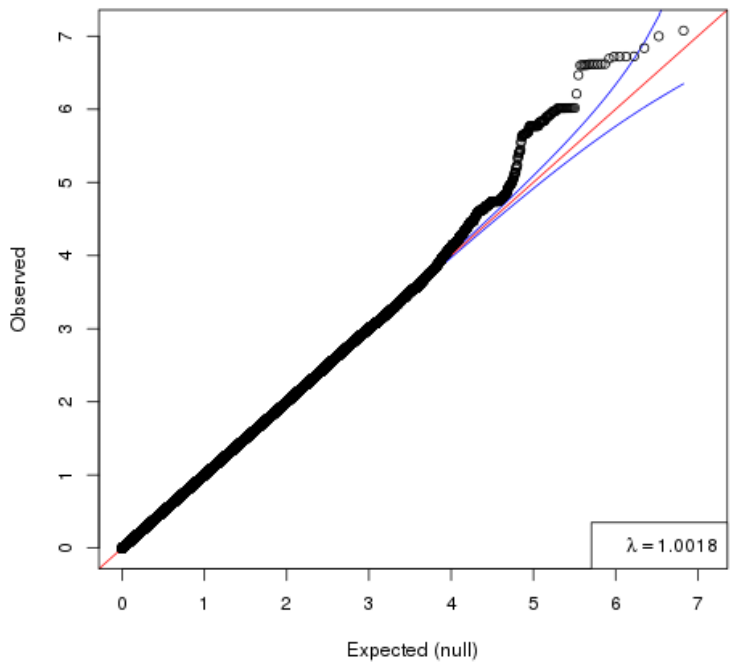


Figure 4.2: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed polyribosyl ribitol phosphate-specific IgG concentrations following immunisation with *Haemophilus influenzae* type b vaccine. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

A Manhattan plot of association p-values between SNPs in the imputed dataset and the persistence of PRP-specific IgG concentrations are shown in Figure 4.3. None of the genetics markers exceeded the level of genome-wide significance ($< 5\times10^{-8}$). However, there was a noteworthy locus on chromosome 13 that contained SNPs with the lowest p-values in the regression model. These SNPs were not within an annotated gene but were <100kb from *FRY*, a gene encoding the furry homolog protein involved in mitosis.

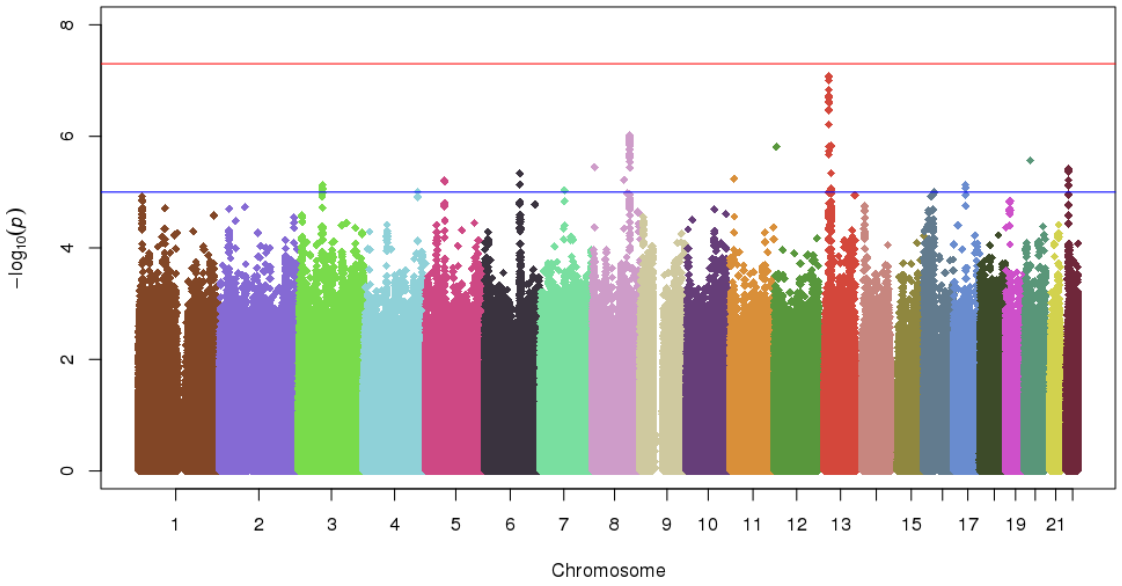


Figure 4.3: Manhattan plot of p-values of regression analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed polyribosyl ribitol phosphate-specific IgG concentrations following immunisation with *Haemophilus influenzae* type b vaccine. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).

One hundred and twenty-two SNPs had regression p-value suggestive of association ($p<1\times10^{-5}$) with the persistence of PRP-specific IgG concentrations. These SNPs were within four annotated genes: *FOXP1*, *PDSS2*, *MLXIPL*, *CSMD3*, as well as the immunoglobulin lambda locus (IGL) (Table 4.2). Of these, *FOXP1* and the IGL, have existing data supporting a role in immune function. *FOXP1* encodes for the transcription factor FOXP1 (forkhead box P1), which is essential in B cell and thymocyte development, as well as monocyte function (325, 326, 327). IGL contains the germline gene segments encoding the lambda light chain portion of immunoglobulin (328). The closest functional lambda segment to this association signal was IGLV3-27. However, the repetitive structure of this locus results in a high frequency of insertions and deletions over an evolutionary timeframe; consequently, the variation at this locus is poorly covered by contemporary genotyping microarrays making fine-mapping of association signals challenging.

The other genes containing SNPs with p-values suggestive of association with PRP-specific

IgG concentrations have not yet been annotated with functions supporting involvement in immunity. *PDSS2* encodes prenyl (decaprenyl) diphosphate synthase subunit 2, an enzyme that synthesises the prenyl side-chain of ubiquinone a key component of the respiratory chain. *MLX1PL* encodes MLX interacting protein-like, a basic helix-loop-helix leucine zipper transcription factor of the Myc/Max/Mad superfamily and *CSMD3* encodes CUB and Sushi multiple domains 3, a potential transmembrane protein that is expressed in several tissues.

Table 4.2: Single-nucleotide polymorphisms (SNPs), in the imputed dataset, with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed polyribosyl ribitol phosphate-specific IgG concentration following immunisation with *Haemophilus influenzae* type b vaccine. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package on build 37 of the human genome (241).

Chr	Position	SNP	p-value	Gene	Neighbouring genes (<100kb)	SNP Class
3	71589755	rs2036281	9.0416e-06	*FOXP1*	*MIR1284,FOXP1*	intron-variant
3	71594837	rs11719972	8.4518e-06	*FOXP1*	*MIR1284,FOXP1*	intron-variant
3	71595258	rs11720121	7.4775e-06	*FOXP1*	*MIR1284,FOXP1*	intron-variant
4	165848324	rs79424296	9.8987e-06		*RPL21P51,LOC100420385,LOC100420386 APELA,LOC101059914,FAM218A,NACA3P LOC391710,LOC391711,TRIM61,LOC391713 RPL26P16,TRIM60P14*	
5	57608276	rs72755777	6.5293e-06			
5	57611114	rs72755778	6.1312e-06			
6	107648957	rs6935797	7.2328e-06	*PDSS2*	*RPS24P12,PDSS2*	intron-variant
6	107666030	rs8180652	4.6599e-06	*PDSS2*	*PDSS2*	intron-variant
7	73019074	rs799158	9.2508e-06	*MLXIPL*	*WBSCR22,VPS37D,TBL2,MLXIPL,STX1A DNAJC30,BAZ1B,BCL7B,*	intron-variant
8	5590355	rs859822	3.5602e-06			
8	96901876	rs920200	5.9992e-06		*LOC100500773,C8orf37-AS1*	
8	113962029	rs78664784	2.0308e-06	*CSMD3*	*CSMD3*	intron-variant
8	114051924	rs117039924	2.4204e-06	*CSMD3*	*CSMD3*	intron-variant
8	114117272	rs117665504	2.1503e-06	*CSMD3*	*CSMD3*	intron-variant
8	114120512	rs142042874	2.1475e-06	*CSMD3*	*CSMD3*	intron-variant
8	114204798	rs116957087	2.1855e-06	*CSMD3*	*CSMD3*	intron-variant
8	114215020	rs10955651	1.4619e-06	*CSMD3*	*CSMD3*	intron-variant
8	114215413	rs10955652	1.4607e-06	*CSMD3*	*CSMD3*	intron-variant
8	114217500	rs118032472	1.4537e-06	*CSMD3*	*CSMD3*	intron-variant
8	114218985	rs117747966	1.449e-06	*CSMD3*	*CSMD3*	intron-variant
8	114222319	rs117575429	1.3809e-06	*CSMD3*	*CSMD3,*	intron-variant
8	114224402	rs118140786	1.2593e-06	*CSMD3*	*CSMD3,*	intron-variant
8	114227683	rs117400618	1.1175e-06	*CSMD3*	*CSMD3*	intron-variant
8	114231015	rs117609106	1.1817e-06	*CSMD3*	*CSMD3*	intron-variant
8	114232302	rs118059686	1.0667e-06	*CSMD3*	*CSMD3*	intron-variant
8	114232362	rs79564895	1.0713e-06	*CSMD3*	*CSMD3*	intron-variant
8	114232418	rs117180292	1.1665e-06	*CSMD3*	*CSMD3*	intron-variant
8	114233355	rs74337126	1.0611e-06	*CSMD3*	*CSMD3*	intron-variant
8	114239131	rs117771716	9.6814e-07	*CSMD3*	*CSMD3*	intron-variant
8	114241285	rs74988398	9.6591e-07	*CSMD3*	*CSMD3*	intron-variant
8	114245946	rs117363533	9.648e-07	*CSMD3*	*CSMD3*	intron-variant
8	114250490	rs79913445	9.637e-07	*CSMD3*	*CSMD3*	intron-variant
8	114251115	rs75701636	9.6314e-07	*CSMD3*	*CSMD3*	intron-variant
8	114252131	rs79624576	9.6314e-07	*CSMD3*	*CSMD3*	intron-variant
8	114253073	rs117328495	9.6259e-07	*CSMD3*	*CSMD3*	intron-variant
8	114256135	rs118123935	9.6204e-07	*CSMD3*	*CSMD3*	intron-variant
8	114258129	rs116977460	9.6149e-07	*CSMD3*	*CSMD3*	intron-variant
8	114265392	rs116902520	9.593e-07	*CSMD3*	*CSMD3*	intron-variant
8	114266286	rs75851420	9.593e-07	*CSMD3*	*CSMD3*	intron-variant
8	114266742	rs78407522	9.593e-07	*CSMD3*	*CSMD3*	intron-variant
8	114268090	rs117450413	9.5875e-07	*CSMD3*	*CSMD3*	intron-variant

Continued on next page

Chr	Position	SNP	p-value	Gene	Neighbouring genes (<100kb)	SNP Class
8	114268944	rs118070483	9.5875e-07	*CSMD3*	*CSMD3*	intron-variant
8	114272966	rs117860097	1.0127e-06	*CSMD3*	*CSMD3*	intron-variant
8	114274722	rs74608707	1.0669e-06	*CSMD3*	*CSMD3*	intron-variant
8	114276384	rs118159724	1.1894e-06	*CSMD3*	*CSMD3*	intron-variant
8	114277203	rs118100628	1.2637e-06	*CSMD3*	*CSMD3*	intron-variant
8	114279220	rs79993933	1.2551e-06	*CSMD3*	*CSMD3*	intron-variant
8	114279801	rs118138964	1.2857e-06	*CSMD3*	*CSMD3*	intron-variant
8	114282379	rs142738486	1.4417e-06	*CSMD3*	*CSMD3*	intron-variant
8	114282814	rs145256102	1.4717e-06	*CSMD3*	*CSMD3*	intron-variant
8	114285499	rs74659793	1.7224e-06	*CSMD3*	*CSMD3*	intron-variant
8	114286854	rs118019897	1.6956e-06	*CSMD3*	*CSMD3*	intron-variant
8	114295786	rs116865016	2.8864e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114301540	rs78448726	1.6956e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114301726	rs118045407	1.6956e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114302932	rs117665578	1.6951e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114303407	rs117458454	1.6952e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114306741	rs75216652	1.6877e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114308432	rs189127506	1.6877e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114313241	rs75574552	1.6877e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114314803	rs150211122	1.6877e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114321334	rs142974216	1.6877e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114323236	rs117493493	1.6879e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114324350	rs117822339	1.6879e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114328193	rs117605863	1.6888e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114329863	rs116897834	1.6888e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114330504	rs12155675	1.6888e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114333666	rs117988440	1.6888e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114335123	rs150037541	1.6888e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114344097	rs78480520	2.0602e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114347504	rs79988843	2.0602e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114349292	rs79528563	2.0602e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114356251	rs117025390	2.2277e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114365068	rs117632465	2.1375e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114365081	rs116981190	2.1387e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114366041	rs118045868	2.3658e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114366307	rs117790062	2.2429e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114377312	rs77576975	2.258e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114378474	rs59196916	2.261e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114380270	rs117874236	2.2618e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
8	114393484	rs117237989	3.6181e-06	*CSMD3*	*RPL18P7,CSMD3*	intron-variant
11	11044677	rs11043061	5.7681e-06			
12	5233825	rs2159402	1.5317e-06		*LOC100507560,KCNA5,LOC390282*	
13	32560995	rs1001852	1.4544e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32561254	rs1001853	8.4019e-08		*FRY-AS1,FRY,EEF1DP3*	
13	32561356	rs1001854	1.8225e-06		*FRY-AS1,FRY,EEF1DP3*	
13	32563266	rs2188631	3.4055e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32564226	rs9566944	1.0029e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32565152	rs9566947	2.0235e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32565369	rs9562389	1.9054e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32565539	rs9566949	1.9139e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32566897	rs35640732	1.8866e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32567495	rs4483729	1.9024e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32568452	rs35908390	6.1494e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32574229	rs7987752	1.5004e-06		*FRY-AS1,FRY,EEF1DP3*	
13	32579551	rs7992130	2.503e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32580110	rs11617881	2.4993e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32582430	rs12585416	2.4842e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32582476	rs34609713	2.4842e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32586408	rs1411481	2.4199e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32587062	rs17591215	2.4186e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32589357	rs66609886	2.4122e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32589361	rs67989666	2.4122e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32589490	rs12856583	2.1433e-06		*FRY-AS1,FRY,EEF1DP3*	
13	32589677	rs68119519	2.4099e-07		*FRY-AS1,FRY,EEF1DP3*	
13	32589954	rs34858589	2.3976e-07		*FRY-AS1,FRY,EEF1DP3*	
13	39651501	rs9576725	8.5533e-06		*NXT1P1,STOML3,NHLRC3,PROSER1*	
13	39678827	rs55713408	4.5636e-06		*NXT1P1,NHLRC3,PROSER1*	
13	39678987	rs7987816	1.4907e-06		*NXT1P1,NHLRC3,PROSER1*	
13	39721225	rs9548645	9.1452e-06		*NXT1P1, NHLRC3*	

Continued on next page

Chr	Position	SNP	p-value	Gene	Neighbouring genes (<100kb)	SNP Class
16	29263902	rs2370287	9.9971e-06		*LOC100996309,LOC101928188,LOC102723693 LOC102723708,SNX29P2*	
17	36766280	rs8068625	8.4223e-06		*C17orf96,MIR4734,LOC101929513 LOC101929533,MLLT6,ARHGAP23,SRCIN1*	
17	36770130	rs9889300	7.394e-06		*C17orf96,MIR4734,LOC101929513 LOC101929533,MLLT6,SRCIN1*	
20	16744979	rs17687121	2.7241e-06		*RPL7AP13,OTOR,SNRPB2*	
22	22923333	rs5758812	7.7634e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	
22	22924377	rs61063966	7.4193e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	
22	22924813	rs5758826	3.8732e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	
22	22924995	rs5758828	4.3068e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	
22	22925548	rs5758831	3.9316e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	
22	22925702	rs5758832	3.7685e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	
22	22927301	rs5751324	4.0314e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	
22	22928199	rs5758855	6.1311e-06		*ZNF280A,LOC129026,ZNF280B,PRAME POM121L1P,IGLV3-32,IGLV3-31,IGLV3-30 IGLV3-29,IGLV3-27,IGLV3-26,IGLV2-34 IGLV2-33,IGLV2-28,IGL,BCRP4,LOC648691 IGLVVI-25-1,GGTLC2*	

Chr= Chromosome

4.3.5 Discussion

The previous section described a GWAS to assess the relationship between genetic variants and the persistence of PRP-specific IgG, following childhood immunisation with Hib conjugate vaccine. While none of SNPs surpassed the level of genome-wide significance, 122 SNPs had p-values suggestive of association with the persistence of PRP-specific IgG concentrations. Furthermore, these included SNPs within loci that have previously been annotated with bio-

logical functions that support the plausibility of these associations, such as *FOXP1* and the immunoglobulin light chain lambda locus.

FOXP1 encodes a transcription factor that is essential in lymphocyte development, and is also involved in regulating germinal centre B cell function (329). FOXP1 protein has been histologically stained in a number of lymphoid and non-lymphoid tissues, and is a transcription factor that is likely to have multiple roles (`http://www.proteinatlas.org`). Data from tonsillar tissue have suggested a role for FOXP1 in regulating the transition of quiescent follicular B cells to activated germinal centre B cells (329). Murine conditional knockouts of *FOXP1* show thymocyte dysfunction, premature acquirement of an activated phenotype by $CD4^+$ and $CD8^+$ T cells and preponderance for cytokine production and apoptosis upon T-cell receptor ligation (327). Variants in *FOXP1* have been associated at the level of genome-wide significance with vitiligo, a condition believed to be immune-mediated (330). *FOXP1* is highly expressed in a subset of diffuse large B cell lymphoma patients and is frequently a target for genomic aberrations in these individuals, and has also been identified as a translocation partner for immunoglobulin heavy chain genes in a subset of mucosa-associated lymphoid tissue lymphomas (331).

In this section, a number SNPs with p-values suggestive of association with PRP-specific IgG concentrations were seen within the immunoglobulin light chain lambda locus; of which the most statistically associated resided between IGLV2-34 and IGLV2-33, with the closest functional lambda segment being IGLV3-27. Furthermore, this signal was <200kb from the IGLV2-14, a segment that has been cloned from heterohybridomas secreting human anti-PRP antibodies (332). However, the genomic structure of the immunoglobulin loci are not well characterised and variants within the immunoglobulin genes are poorly captured by standard GWAS microarray platforms (56). For example, IGLV2-14 was not well tagged by any directly genotyped or imputed markers within the SNP set evaluated in this section. Despite the association signal seen in the IGL locus, no association signals were seen in either the immunoglobulin heavy chain or κ light chain loci. While this may highlight the importance of the λ light chain in PRP-specific responses, this may also reflect the poor coverage of genetic variants in the

immunoglobulin loci.

A previous candidate gene study found two variants in *MAL/TIRAP* (rs1893352) and *IL-10* (rs1554286) to be associated with Hib vaccine failure in a recessive model (47). However, these SNPs were not associated with the persistence of PRP-specific IgG concentration in the study described in this section. There are a number of possible explanations for this apparent disparity. Firstly, the associations found in the vaccine failure study were based upon a recessive model, whereas the analysis conducted in this section used an additive allelic model. Importantly, additive models assume the heterozygous genotype an intermediate of the two homozygous genotypes, and may give disparate results to recessive or genotypic models when this assumption is violated. Secondly, a large proportion of the individuals in the vaccine failure study had underlying medical conditions that were associated with increased risk of low post-vaccination anti-PRP IgG concentrations and Hib disease; therefore, the effect of genetic variants identified in these participants may not be generalisable to a healthy cohort (47, 333).

With the exception of the aforementioned candidate gene study, investigations of the genetic determinants of Hib immunity have been limited to the immunoglobulin gene loci. Several studies have shown PRP-specific antibodies, following Hib disease or vaccination, to be oligoclonal with highly restricted immunoglobulin gene usage and V(D)J combinations (332, 334, 335, 336). Furthermore, a large proportion of PRP-specific antibodies appear to use unmutated germline sequence, which places more importance on the germline repertoire (332). Variants in both the light and heavy chain genes have been correlated with the PRP-specific antibody concentrations following Hib vaccination (318, 319, 320).

While several SNPs in this section had p-values suggestive of association with the persistence of PRP-specific IgG concentrations, none of these surpassed the statistical threshold of genome-wide significance. There were a number of factors that may have limited the power to detect genetic variants influencing serological immunity following Hib vaccination such as participant heterogeneity: in terms of age at vaccination, Hib vaccine regimen and composition, as well as time since vaccination.

4.3.6 Conclusion

Maintaining serological immunity is crucial in preventing the Hib disease. In this section a GWAS was conducted highlighting a number of genetic variants that may influence the persistence of Hib-specific serological immunity following vaccination. Furthermore, this section represents stage 1 of a two-stage GWAS, and the most promising genetic markers are further evaluated in stage 2 (appendix 7.5).

4.4 Genome-wide association study of persistence of serological immunity to tetanus toxoid following immunisation - Stage 1

4.4.1 Background

4.4.1.1 Tetanus

Tetanus is an acute and life-threatening disease, characterised by skeletal muscle rigidity and spasms, caused by tetanus toxin produced by the bacterium *Clostridium tetani*. *C. tetani* is noncommunicable, but spores are commonly present in the environment and can be introduced into the body via skin lesions. In the UK, tetanus became notifiable in 1968; data collected in the ten years following showed the highest incidence in those over 65 years and in young males aged 15-44 years (337). Data from 1930-1970 showed a decline in deaths due to tetanus, which was attributed to a number of social developments as well as the infant and military immunisation campaigns (337). Surveillance data from 1984-2000 showed a further decline in tetanus cases, with the highest incidence occurring in the >65 years (338). In this survey the majority of cases were associated with recent injuries occurring in the garden or home (61%) and were predominantly in unvaccinated (63%) individuals (338). Globally, neonatal tetanus has a significant health impact with the World Health Organisation estimating 50,000 newborn deaths from neonatal tetanus in 2008 (339).

4.4.1.2 Tetanus toxoid vaccine

C. tetani spores cannot be eradicated from environment making vaccine-mediated protection essential in preventing disease. The tetanus vaccine was introduced as a routine childhood immunisation in 1961. The incubation of period of *C. tetani* is ∼10 days meaning the tetanus vaccine can also be given as post-exposure prophylaxis for individuals with tetanus-prone wounds. Tetanus vaccines contain tetanus toxoid (TT), a formaldehyde treated form of tetanus toxin, adsorbed onto aluminium phosphate or aluminium hydroxide to improve immunogenicity.

4.4.1.3 Genetic studies of tetanus toxoid vaccine immunity

Twin data have shown antibody responses to tetanus toxoid to be heritable; however, there are few data describing the determinants underlying the genetic component (7). To date, two published studies have described genetic variants associated with the magnitude of tetanus-specific antibody following DTaP vaccine, both of which were conducted on the same cohort (n=141) of infant vaccinees (48, 340). The first assessed 19 polymorphisms suspected to alter the function of 11, cytokine or cytokine receptor, genes (48). This study found an association between a missense variant (rs1805010) in *IL4R* and the magnitude of tetanus-specific IgG following vaccination, at an uncorrected p=0.037 (48). The second of these studies assessed variants in the HLA region, assessing 1856 SNPs in 154 genes, finding a single variant (rs3130454) in *PSORS1C1* to be associated with tetanus-specific IgG responses at the authors' statistical significance threshold of $p<0.001$ (340). The study presented in this section describes a two-stage genome-wide association study of the persistence of tetanus-specific IgG following immunisation with tetanus toxoid containing vaccines.

4.4.2 Aim

The aim of this section was to describe stage 1 of a two-stage genetic association study of the persistence of serological immunity to tetanus toxoid following childhood immunisation.

4.4.3 Materials and methods

4.4.3.1 Stage 1 participants

Participants included in this section were a subset of those described in chapter 3, for whom post-vaccination TT-specific IgG concentrations, or sera, were available. This subset comprised of children from studies 1, 2 and 3, described in subsection '3.4.3.1 Stage 1 participants'.

4.4.3.2 Vaccines

All participants described in this section received three infant doses of TT, in the form a combination vaccine at 2, 3 and 4 months of age. This combination vaccine was either DTwP, DTwP/Hib, DTaP/Hib or DTaP/Hib. In addition, TT is used as the protein carrier in the Hib conjugate vaccine as well as some MenC conjugate vaccine formulations, administered in the infant immunisation schedule. Children who were 12 months of age since 2006 would have also received an additional dose of TT in their Hib-MenC booster vaccine. Furthermore, school aged participants would have received their preschool combination DTaP/IPV vaccine. TT immunisation information was obtained from clinical study data, or from the child health computer department.

4.4.3.3 Immunological measurements

The VEU, Manchester, measured TT-specific IgG concentrations for study 2 and study 3, using a multiplexed bead assay, in accordance with an internal protocol. The LLQ for this assay was 2 (IU×1000/ml), with measures below this recorded as half the LLQ. GSK Biologicals, Belgium, measured TT-specific IgG concentrations for study 1, using a standardised ELISA based method. The LLQ for this assay was 0.1 IU/ml, with measures below this recorded as half the LLQ.

4.4.3.4 Genotyping and imputation

Genotyping was performed as described in subsection '3.4.3.5 Genotyping'. Imputation was completed as described in subsection '3.3.3.8 Imputation'.

4.4.3.5 Quantitative trait loci analysis

A linear regression model was used to assess the relationship between SNP allele dosage and the persistence of TT-specific IgG concentrations; with age at vaccination, time since vaccination and MDS components 1-10, and $K-1$ dummy variables coding for study cohort as covariates.

4.4.3.6 Imputation of classical human leukocyte antigen alleles

The HLA*IMP framework was used to impute classical HLA alleles from the genotype data (341, 342). This algorithm uses a probabilistic approach referred to as 'LDMhc' to assess the relatedness between a densely genotyped sample set chromosomes with unknown HLA alleles and a reference set of HLA-typed chromosomes, based upon markers typed in both sets (342). The reference dataset utilised by this algorithm included over 2500 samples from European ancestry with dense SNP and classical HLA genotypes, and has purported imputation accuracy of 92-98% for alleles with call rates of 95-99% (342). HLA imputation was performed using the HLA*IMP:01 Front-End program and the online Back-End software, as described by the software manual (`https://oxfordhla.well.ox.ac.uk/hla/tool/main`). HLA alleles were called if attributed a greater than 0.7 posterior probability. Linear regression analysis was then performed using an additive allelic dose model with the covariates previously specified in subsection '4.4.3.5 Quantitative trait loci analysis'.

4.4.4 Results

4.4.4.1 Stage 1 demographics

An overview of stage participant demography is displayed in Table 4.3.

Table 4.3: Demographics in stage 1 of the genome-wide association study of the persistence of tetanus toxoid (TT)-specific IgG concentrations following childhood immunisation with TT containing vaccines.

Study	Male/Female	Time since vaccination days* (IQR)	Age at vaccination days* (IQR)	Reference
Study 1	36/25	275.5 (263.3 - 287.8)	123 (118 - 128)	(145)
Study 2	141/109	616.0 (299.0 - 1497.5)	1185 (365 -1185)	(282)
Study 3	121/117	256 (248.5 - 264)	122 (118 - 127)	(283)

IQR= interquartile range, * = median

4.4.4.2 Serology

The distributions of $\log_{10}$ transformed TT-specific IgG concentrations approached normality (Figure 4.4). All measurement made on sera from study 2 and study 3 surpassed their respective assay LLQ, and 3/61 (5%) of sera assessed in study 1 were below the assay LLQ and were reported as half the LLQ.

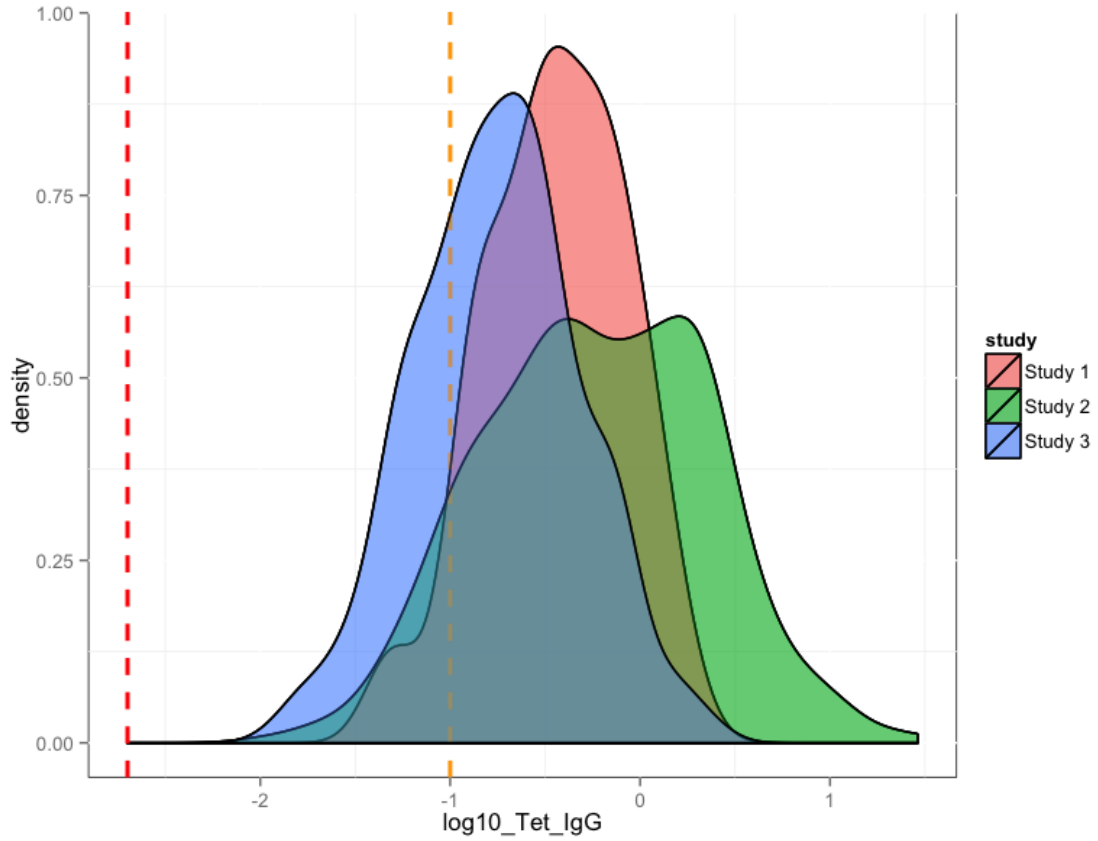


Figure 4.4: Density plots of $\log_{10}$ transformed tetanus toxoid (TT)-specific IgG concentrations for stage 1 participants in the genome-wide association study of persistence of serological immunity to TT following vaccination. The orange vertical dashed line represents a lower limit of quantitation (LLQ) of 0.1 IU/ml used in study 1, and the red vertical dashed line represent a LLQ of 0.002 IU/ml used in Study 2 and Study 3.

4.4.4.3 Association analysis

The imputed data of 34,026,329 variants were filtered for SNPs with a MAF >0.02, HWE $>1\times10^{-9}$ and an information score >0.8, resulting in a set of 6,706,188 SNPs for regression

analysis. Of the 1683 participants passing QC (subsection '3.4.4.3 Quality control'), 549 had TT-specific IgG measures and were included in the regression analysis.

The Q-Q plot of p-values resulting from the regression of imputed allele doses and TT-specific IgG concentrations did not deviate considerably from the null distribution, with modest enrichment of p-values at the tail ($p< 1\times10^{-4}$) of the distribution (Figure 4.5). However, it was necessary to included PCs 1-10 in the regression model as genomic inflation was observed when only the first three PCs were included (appendix Figure 7.14).

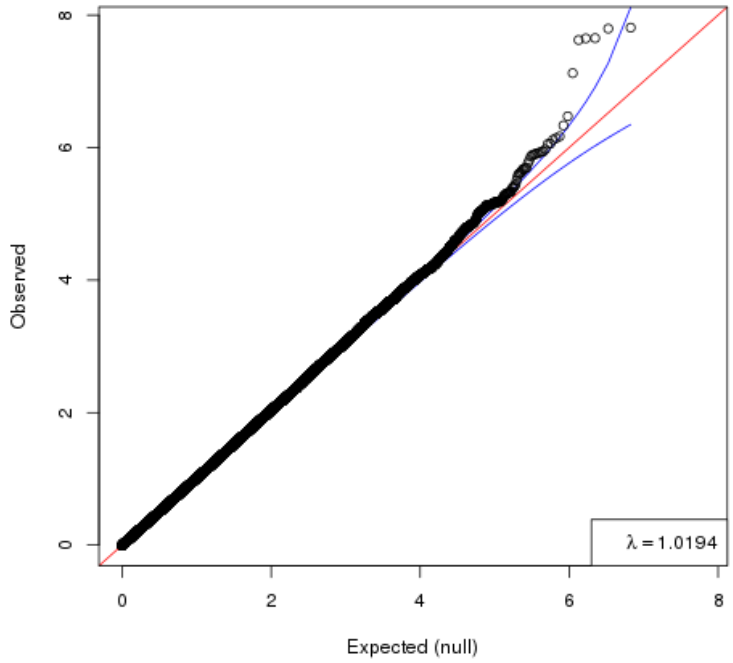


Figure 4.5: Quantile-quantile plot of -$\log_{10}$ p-values observed in association analysis between imputed single-nucleotide polymorphisms in the imputed dataset and $\log_{10}$ transformed tetanus toxoid (TT)-specific IgG concentrations following immunisation with TT containing vaccines. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

A Manhattan plot of association p-values between SNPs in the imputed dataset and the persistence of TT-specific IgG concentrations is shown in Figure 4.6. Five SNPs in an intergenic region of chromosome 10 surpassed the level of genome-wide significance ($p< 5\times10^{-8}$); the closest gene (<250kb) to these associations was *TCERG1L*. Furthermore, 111 SNPs in 10 genes surpassed the level of suggestive association ($p< 1\times10^{-5}$); the corresponding genes were *CDH9,*

AC011343.1, MICB, HLA-DQA1, AC005062.2, DOCK4, DENND1A, LHPP, TMEM135 and *RBFOX3* (Table 4.4). Of these genes, *MICB* and *HLA-DQA1* are annotated as having immune-related functions. *MICB* encodes for the MHC class I polypeptide-related sequence B, a heavily glycosylated protein ligand of NKG2D that is induced by stress and can activate natural killer (NK) cells, CD8 $\alpha\beta$ T cells and $\gamma\delta$ T cells in a peptide independent manner. Furthermore, the most statistical significant SNP within *MICB*, rs2894221, is an eQTL for the expression of this gene (`http://www.gtexportal.org`). *HLA-DQA1* encodes HLA class II alpha chain and plays a central role in the presentation of exogenous peptides by antigen presenting cells to T cells. The most statistically significant SNP within *HLA-DQA1*, rs9272374, is also an eQTL for the expression of this gene (`http://www.gtexportal.org`).

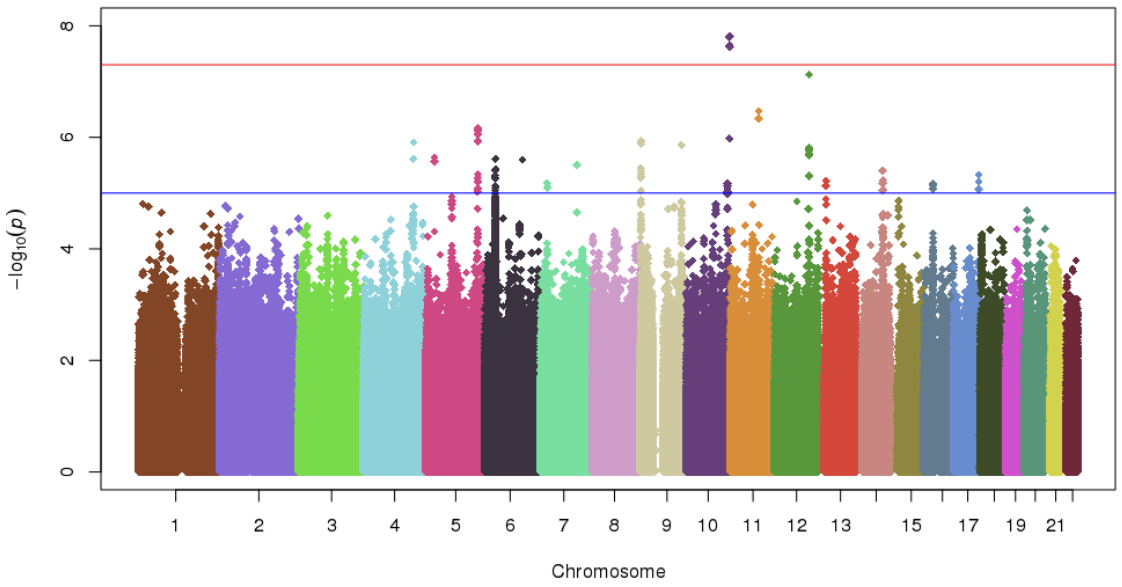


Figure 4.6: Manhattan plot of p-values from regression analysis between single-nucleotide polymorphisms, in the imputed dataset, and $\log_{10}$ transformed tetanus toxoid (TT)-specific IgG concentrations in stage 1 of the genome-wide association study of the persistence of vaccine-induced serological immunity to TT. The horizontal blue line represents the level of suggestive statistical association ($p<1\times10^{-5}$), and the horizontal red line denotes the level of genome-wide statistical significance ($p<5\times10^{-8}$).

Table 4.4: Single-nucleotide polymorphisms (SNPs), in the imputed dataset, with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed tetanus toxoid (TT)-specific IgG concentration in stage 1 of the genome-wide association study of persistence of serological immunity to tetanus toxoid. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package on build 37 of the human genome (241).

Chr	Position	SNP	p-value	Gene	Neighbouring genes (<100kb)	SNP Class
4	152783376	rs72732116	2.405e-06		*LOC100505685,LOC102724700,*	
4	153626438	rs76463503	1.2327e-06		*LOC100421721,LOC102724121,TIGD4,TMEM154 ARFIP1,RPS14P6*	
5	26903163	rs138455466	2.2748e-06	*CDH9*	*CDH9*	intron-variant
5	26914480	rs186330563	2.7039e-06	*CDH9*	*CDH9*	intron-variant
5	159296102	rs7720980	7.5687e-07	*AC011343.1*	*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	non-coding
5	159305271	rs13169166	8.7188e-07		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159308795	rs891902	4.5402e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159317303	rs17674132	5.3631e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159317416	rs6556460	1.1473e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159320525	rs13179580	8.8237e-07		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159323554	rs11745333	6.254e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159325407	rs7729196	7.1797e-07		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159326237	rs10072006	6.7538e-07		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159326688	rs56084650	8.2835e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159327344	rs4921535	9.4582e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159328747	rs11746691	9.3282e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159329910	rs891901	8.6355e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B*	
5	159336790	rs71603644	1.1906e-06		*GAPDHP40,LOC101927766,LOC101927790 ADRA1B,TTC1*	
6	31461771	rs2894221	4.8705e-06	*MICB*	*RPL15P4,MICA,ATP6V1G2-DDX39B LOC101929072,LINC01149,HCP5,NCR3,HLA-X HCG26,MCCD1,LTA,LTB,MICB,NFKBIL1 ATP6V1G2,PPIAP9,SNORD84,SNORD117,TNF DDX39B,LST1*	upstream-2KB
6	32392591	rs116587141	4.7884e-06		*HNRNPA1P2,C6orf10,HLA-DRA,HLA-DRB5 HLA-DRB9,HCG23,BTNL2*	
6	32394913	rs115365653	4.7873e-06		*C6orf10,HLA-DRA,HLA-DRB5,HLA-DRB9,HCG23 BTNL2*	
6	32435955	rs186229361	3.7376e-06		*C6orf10,HLA-DRA,HLA-DRB5,HLA-DRB6 HLA-DRB9,HCG23,BTNL2*	
6	32435983	rs192153994	8.0197e-06		*C6orf10,HLA-DRA,HLA-DRB5,HLA-DRB6 HLA-DRB9,HCG23,BTNL2*	
6	32436185	rs139370718	7.4672e-06		*C6orf10,HLA-DRA,HLA-DRB5,HLA-DRB6 HLA-DRB9,HCG23,BTNL2*	
6	32441679	rs147882356	3.9298e-06		*HLA-DRA,HLA-DRB5,HLA-DRB6,HLA-DRB9 HCG23,BTNL2*	
6	32445270	rs145763900	7.4005e-06		*HLA-DRA,HLA-DRB5,HLA-DRB6,HLA-DRB9 HCG23,BTNL2*	
6	32449523	rs149028172	5.3847e-06		*HLA-DRA,HLA-DRB1,HLA-DRB5,HLA-DRB6 HLA-DRB9,HCG23,BTNL2*	
6	32583831	rs114822165	8.9161e-06			
6	32604669	rs9272374	2.4229e-06	*HLA-DQA1*	*HLA-DQA1,HLA-DQB1,HLA-DRB1,HLA-DRB6 MTCO3P1*	upstream-2KB
6	114933532	rs72943836	2.5197e-06			
7	20132499	rs59229257	6.6847e-06	*AC005062.2*	*MACC1-AS1,LOC101927668,MACC1,RPL21P75*	intron-variant
7	20142890	rs10259621	7.9131e-06	*AC005062.2*	*MACC1-AS1,LOC101927668,MACC1,RPL21P75*	intron-variant
7	111748533	rs55894510	3.123e-06	*DOCK4*	*ZNF277,DOCK4,*	intron-variant
9	1849530	rs7020165	4.1369e-06			
9	1849664	rs7019385	1.1637e-06			
9	1850244	rs6475215	4.6416e-06			
9	1850836	rs7860262	1.2233e-06			
9	1851231	rs10114691	4.932e-06			
9	1852121	rs7025509	1.2865e-06			
9	1852175	rs7028857	1.2909e-06			
9	1852281	rs7040599	5.1158e-06			
9	1853645	rs1117162	5.2988e-06			
9	1858042	rs73638310	3.5599e-06			
9	1867296	rs73389636	9.121e-06			

Continued on next page

4.4 Genome-wide association study of persistence of serological immunity to tetanus toxoid following immunisation - Stage 1

Chr	Position	SNP	p-value	*Gene*	*Neighbouring*	SNP Class
9	126234674	rs113925093	1.3734e-06	*DENND1A*	*MIR7150,LOC102723374,CRB2,DENND1A MIR601*	intron-variant
10	126198328	rs1869658	6.8427e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126199289	rs3740542	6.5573e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126199573	rs7082588	6.5672e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126200421	rs9663728	6.6449e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126200427	rs9665724	7.2172e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126200957	rs11245248	6.7218e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126201047	rs73365885	9.2113e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126201135	rs75612732	9.9434e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126201926	rs12569933	7.3977e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126202455	rs10901743	8.3687e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126204194	rs11245251	8.2673e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126204581	rs11245252	7.4949e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126205537	rs11245253	6.695e-06	*LHPP*	*RPS10P18,NKX1-2,OAT,OATP1,LHPP*	intron-variant
10	126209578	rs10794144	9.8783e-06	*LHPP*	*RPS10P18,NKX1-2,OATP1,LHPP,FAM53B*	intron-variant
10	126210850	rs56681095	9.2225e-06	*LHPP*	*RPS10P18,NKX1-2,OATP1,LHPP,FAM53B*	intron-variant
10	126211692	rs11245258	9.0857e-06	*LHPP*	*RPS10P18,NKX1-2,OATP1,LHPP,FAM53B*	intron-variant
10	126212220	rs66509841	9.423e-06	*LHPP*	*RPS10P18,NKX1-2,OATP1,LHPP,FAM53B*	intron-variant
10	126212235	rs12219151	9.5926e-06	*LHPP*	*RPS10P18,NKX1-2,OATP1,LHPP,FAM53B*	intron-variant
10	132633601	rs114783405	2.2392e-08			
10	132633808	rs115548138	2.2204e-08			
10	132636305	rs79812992	2.367e-08			
10	132642718	rs79697632	1.059e-06			
10	132643632	rs75727401	1.5423e-08			
10	132645646	rs77724803	1.5959e-08			
11	86956746	rs193290451	4.6185e-07	*TMEM135*	*PSMA2P1,LOC100420680,LOC101929104 TMEM135*	intron-variant
11	86960870	rs145850522	3.3568e-07	*TMEM135*	*PSMA2P1,LOC100420680,LOC101929104 TMEM135*	intron-variant
12	105960157	rs1543280	2.0709e-06			
12	105961001	rs2694402	2.0003e-06			
12	105961210	rs2694403	1.7298e-06			
12	105962957	rs10861460	2.0654e-06			
12	105967003	rs1526856	1.5561e-06			
12	105968328	rs1526855	2.1148e-06			
12	105978253	rs2711740	7.5022e-08			
12	105988797	rs12809123	4.9094e-06			
13	24093467	rs9507111	7.3597e-06		*SACS-AS1,LINC00327,LOC101928839,SACS TNFRSF19*	
13	24097260	rs9510764	7.3597e-06		*SACS-AS1,LINC00327,LOC101928839,SACS TNFRSF19*	
13	24097919	rs9507112	7.3921e-06		*SACS-AS1,LINC00327,LOC101928839,SACS TNFRSF19*	
13	24104109	rs9510765	6.0502e-06		*LINC00327,LOC101928839,SACS,TNFRSF19*	
13	24106628	rs56365165	7.3594e-06		*LINC00327,LOC101928839,SACS,TNFRSF19*	
13	24107140	rs67515305	7.3505e-06		*LINC00327,LOC101928839,SACS,TNFRSF19*	
13	24108342	rs34835073	7.3431e-06		*LINC00327,LOC101928839,TNFRSF19*	
13	24108666	rs34858081	7.3415e-06		*LINC00327,LOC101928839,TNFRSF19*	
13	24116122	rs9317815	7.3597e-06		*LINC00327,LOC101928839,TNFRSF19*	
13	24116554	rs4414331	7.3605e-06		*LINC00327,LOC101928839,TNFRSF19*	
13	24118105	rs35017503	7.5179e-06		*LINC00327,LOC101928839,TNFRSF19*	
14	82758463	rs709915	8.6409e-06			
14	82768157	rs709918	9.0227e-06			
14	82768387	rs1068199	8.9662e-06			
14	82778603	rs1457990	6.4433e-06			
14	82779730	rs1457991	6.6371e-06			
14	82782068	rs1379682	6.6735e-06			
14	82783803	rs8022433	5.854e-06			
14	82785464	rs2199009	6.5582e-06			
14	82785587	rs2199010	6.6402e-06			
14	82787076	rs12897855	3.9545e-06			
14	82787254	rs8023135	6.6556e-06			
14	82795671	rs1457983	6.689e-06			
16	26986280	rs12149060	7.2247e-06		*C16orf82*	
16	26986654	rs4494544	7.2061e-06		*C16orf82*	
16	26987784	rs9746000	6.8217e-06		*C16orf82*	
16	26987999	rs10775282	6.8725e-06		*C16orf82*	
16	26992254	rs12935709	8.2389e-06		*C16orf82*	
16	26993216	rs6497981	8.5802e-06		*C16orf82*	
17	77135561	rs12944841	8.4734e-06	*RBFOX3*	*LOC102723900,C1QTNF1,RBFOX3,ENGASE*	intron-variant
17	77136112	rs12952828	6.2842e-06	*RBFOX3*	*LOC102723900,C1QTNF1,RBFOX3,ENGASE*	intron-variant
17	77136597	rs7214274	4.6637e-06	*RBFOX3*	*LOC102723900,C1QTNF1,RBFOX3,ENGASE*	intron-variant

Chr= Chromosome

4.4.4.4 Classical human leukocyte antigen alleles

One hundred and fifty-four classical HLA alleles were imputed at four-digit resolution. The results of association analysis between imputed HLA alleles and TT-specific IgG concentrations are displayed in appendix Table 7.26 and the top 10 HLA alleles based upon p-value are presented in Table 4.5. None of the imputed HLA alleles were associated with TT-specific IgG concentrations at the level for genome-wide significance. However, two alleles, HLA-DQB*0201 and HLA-DRB*0301, surpassed the level of statistical correction when accounting only for HLA alleles tested (i.e. $p<0.05/154$).

Table 4.5: Top ten imputed human leukocyte antigen (HLA) alleles, based upon p-values, associated with tetanus toxoid (TT)-specific IgG concentrations in stage 1 of the genome-wide association study of the persistence of serological immunity to TT.

HLA allele	Estimate	p-value
HLA-DQB1*0201	-0.1741	**0.00010**†
HLA-DRB1*0301	-0.1656	**0.00021**†
HLA-B*3906	1.8050	0.00046
HLA-DRB1*1501	0.1454	0.00103
HLA-B*0801	-0.1342	0.00210
HLA-DQB1*0602	0.1324	0.00276
HLA-C*0702	0.1340	0.00375
HLA-DRB1*1101	0.2917	0.00443
HLA-DQA1*0102	0.1073	0.00593
HLA-B*0702	0.1238	0.00837

†= significant, Bonferroni corrected for number of HLA alleles, Estimate = β coefficient

4.4.5 Discussion

This section describes a GWAS to assess the relationship between genetic variants and the persistence of TT-specific IgG, following childhood immunisation with TT containing vaccines. Genome-wide significance was achieved by five intergenic SNPs on chromosome 10. Furthermore, 111 SNPs in 10 genes displayed p-values suggestive of statistical association including SNPs within *MICB* and *HLA-DQA1*, which are annotated as having immune-related functions.

Genome-wide significant associations were found between SNPs present in an intergenic region of chromosome 10 and the persistence TT-specific IgG concentrations following childhood immunisations with TT-containing vaccines. Currently, the closest annotated transcribed RNA

to this association signal is miRNA378C (∼115 kb) and the closest protein-coding transcript is *TCERG1L* (<250 kb). Intergenic regions are assumed to have the smallest ratio of functional to total DNA sequence; nevertheless, they can contain important regulatory sequences and 43% of trait-associated SNPs (TAS) discovered from GWASs have been intergenic (219). Recent data have shown the human genome to be pervasively transcribed and an increasing number of functional non-coding RNAs are being described (343, 344, 345). Moreover, lincRNAs (long intergenic non-coding RNAs) are enriched for TAS compared to non-transcribed intergenic regions (345). It may be the case that an uncharacterised non-coding RNA(s) underlie this association; alternatively, this signal may originate from DNA based regulatory elements or represent a false positive.

One hundred and eleven SNPs in 10 genes surpassed a suggestive level of association, these included SNPs within *MICB* and *HLA-DQA1*, which are genes with annotated putative involvement in immunity. *MICB* encodes MICB a stress-induced molecule that interacts with natural killer (NK) cells, $CD8^+$ $\alpha\beta$ T cells and $\gamma\delta$ T cells in manner independent of antigen processing (346). Variants in *MICB* have previously been associated with immune-related phenotypes, type I diabetes and Behçet's syndrome (290, 347). *HLA-DQA1* encodes a HLA class II alpha chain. HLA molecules present peptide to T-cells, resulting in the clonal expansion of antigen-specific T cells and the generation of stimulatory factors required for T cell-dependent B cell activation (55, p. 369-370). The HLA region is highly polymorphic and several GWASs have found variants within this locus to be associated with immune-related phenotypes including vaccine immunity (`http://www.genome.gov/gwastudies/`). The complex structure of the HLA region can make unequivocal fine-mapping of causal variants challenging (348, 349). Association signals may be due to one or more HLA allelic variants; alternatively, variants within regulatory elements influencing HLA expression may be causal. Moreover, the HLA region is gene rich and variants in non-HLA genes may also underlie association signals (349).

Imputation of classical HLA alleles did not elucidate any associations at the level of genome-wide significance. However, two HLA alleles surpassed locus-specific correction for multiple testing, HLA-DQB1*0201 and HLA-DRB1*0301. HLA-DRB1*0301 is distinct from the other

major *HLA-DRB1* alleles both genetically and immunologically. HLA-DRB*0301 is unusual in that this allele appears to have arisen from a recombination event between the β1 and β3 DR loci, and the hypervariable region more closely resembles products of the latter (350). The HLA molecule encoded by HLA-DRB1*0301 has been shown to recognise a peptide motif that is distinct from that recognised by the other major *HLA-DRB1* alleles, which was demonstrated by disparate contact residues to HLA-DR restricted peptides such as tetanus toxoid epitope 830-843 (350).

A study conducted in a small cohort of TT vaccinees, using tetramer-guided epitope mapping of a panel of 106 overlapping TT heavy-chain peptides, found heterogeneous T-cell responses for different class II alleles, with HLA-DRB1*0301 presenting the fewest antigenic TT peptides of the five DR alleles assessed (351). While the aforementioned study did not claim HLA-DRB1*0301 to be less effective at presenting TT, the previous section described an association between this allele and lower TT-specific IgG that could be explained by this allele recruiting less TT-specific T-cell help. The HLA-DRB1*0301 allele has also been associated with increased susceptibility to several autoimmune and allergic phenotypes, as well as lower HBsAg-specific IgG concentrations following hepatitis B vaccine, supporting a generalisable difference in the peptide binding preferences (352, 353, 354).

Within the HLA region, the SNP with the greatest statistical evidence for association with persistence of TT-specific IgG concentrations was an eQTL for *HLA-DQA1*. Furthermore, data from a HBV GWAS associated an eQTL for *HLA-DPB1* with the magnitude of HBV-specific IgG concentrations following vaccination (22, 71). These data implicate regulatory elements within the HLA region in determining immune responses to vaccines.

Previous published data have associated variants within *IL4R* (rs1805010) and *PSORS1C1* (rs3130454) with the magnitude of TT-specific IgG concentrations following infant vaccination (48, 340). In the study presented in this section, neither rs1805010 nor rs3130454 had regression p-values suggestive of association with TT-specific IgG concentrations. However, it is was worth noting that the rs3130454 regression statistic was p=0.001, with concordant effect directionality to the aforementioned study. The disparity between these studies may relate to

lack of statistical power due to modest sample sizes, differences in genetic marker set evaluated and/or differences in the ancestries of the populations evaluated.

Twin study data have implied that the majority of the genetic determinants of TT-specific immunity are non-HLA (7). While none of the non-HLA genes that contained SNPs with p-values suggestive of association with TT-specific IgG concentrations were annotated with immunological functions, this does not preclude their involvement in TT vaccine-induced immunity. Furthermore, these genes are unlikely to be comprehensively annotated and will have functions that have not yet to be experimentally demonstrated. Moreover, SNP associations may not relate to proximal protein-coding gene but may influence distal genes or non-coding RNAs.

There are a number of factors that may have limited the power of this study to describe SNPs influencing the persistence of TT-specific IgG concentrations following vaccination. In particular, participant heterogeneity in terms of age at vaccination, TT vaccine regimen and composition, as well as time since vaccination is likely to have considerably reduced power. Furthermore, only a modest number of participants were analysed in this section, given the level of statistical correction needed to account for multiple testing.

4.4.6 Conclusion

This section demonstrated genome-wide significant associations between SNPs present in an intergenic region of chromosome 10 and TT-specific IgG concentrations following childhood immunisation with TT-containing vaccines. Furthermore, a number of SNPs were found to have suggestive levels of association with TT-specific IgG concentrations, including eQTLs for *MICB* and *HLA-DQA1*. Furthermore, this section represents stage 1 of a two-stage GWAS, and the most promising genetic markers are further evaluated in stage 2 (appendix 7.6).

4.5 Pleiotropic loci associated with immunity to childhood immunisations

4.5.1 Background

The determinants of immunological responses to vaccines are not well understood. However, correlations between serological responses to certain vaccine antigens suggest some factors may have relatively generic effects (7, 355). Infant responses to antigens present in the combination DTaP-Hib vaccine have been shown to positively correlate, an observation that was thought to be due to age-dependent immunological maturation (355, 356). However, disparities in the correlations between antigens argued against a simple 'responsiveness' trait due to immunological maturation. The aforementioned studies, found a particularly close relationship between IgG concentrations to Hib and its protein carrier TT, implicating T cell help related to TT in determining the magnitude of immune response to TT and Hib (8, 355, 356).

In a further study, a strong correlation between TT- and DT-specific IgG concentrations was observed following infant immunisations (7). Moreover, TT- and DT-specific IgG concentrations were shown to have a similar degree of heritability, leading to speculation that factors responsible for their heritabilities may be shared between these two toxoid vaccines. Conversely, despite serological responses to both HBV and OPV being highly heritable (>60%) these measures poorly correlated, which could be interpreted as differing genetic factors influencing responses to these disparate vaccines (7). This section compares serological responses to three routine immunisations, and explores potential pleiotropy in serological immunity to these vaccines.

4.5.2 Aim

The aim of this section was to explore pleiotropic loci involved in immunological responses to different childhood immunisations.

4.5.3 Materials and methods

4.5.3.1 Partipants

The subsequent section includes participants described in chapter 3. Participants were included in vaccine-immunity correlation analysis if two or more of the following were available: MenC-specific SBA titres, MenC-specific IgG concentration, Hib-specific IgG concentration and TT-specific IgG concentrations.

4.5.3.2 Vaccine immunity correlations

Pearson product-moment correlation coefficients (PPMCC) were calculated on each paired combination of $\log_{10}$ transformed serological measures of vaccine-induced immunity: MenC-specific SBA titres, MenC-specific IgG concentration, Hib-specific IgG concentration and TT-specific IgG concentrations. Association tests assessed the alternative hypothesis that the test statistic (PPMCC) was not equal to zero.

4.5.3.3 Pleiotropic analysis

SNPs associated with the persistence of serological immunity to MenC, Hib or TT at a $p<1\times10^{-4}$ were collated (subsections '3.3.4.5 Imputation', '4.3.4.3 Association analysis' and '4.4.4.3 Association analysis'). Genetic loci for comparative analysis were defined by gene boundaries; therefore, analysis was restricted to coding loci.

4.5.4 Results

4.5.4.1 Serology correlations

The correlations between the four serological measures of vaccine-induced immunity assessed in this section are displayed in Figure 4.7. Pairwise analysis of these serological measures showed statistically significant, positive, correlations between all serological measures. The two serological measures with the largest correlation coefficient were PRP- and TT-specific IgG

concentrations with r=0.59. The two measures of MenC-specific immunity, SBA and IgG, had a correlation coefficient of r=0.57.

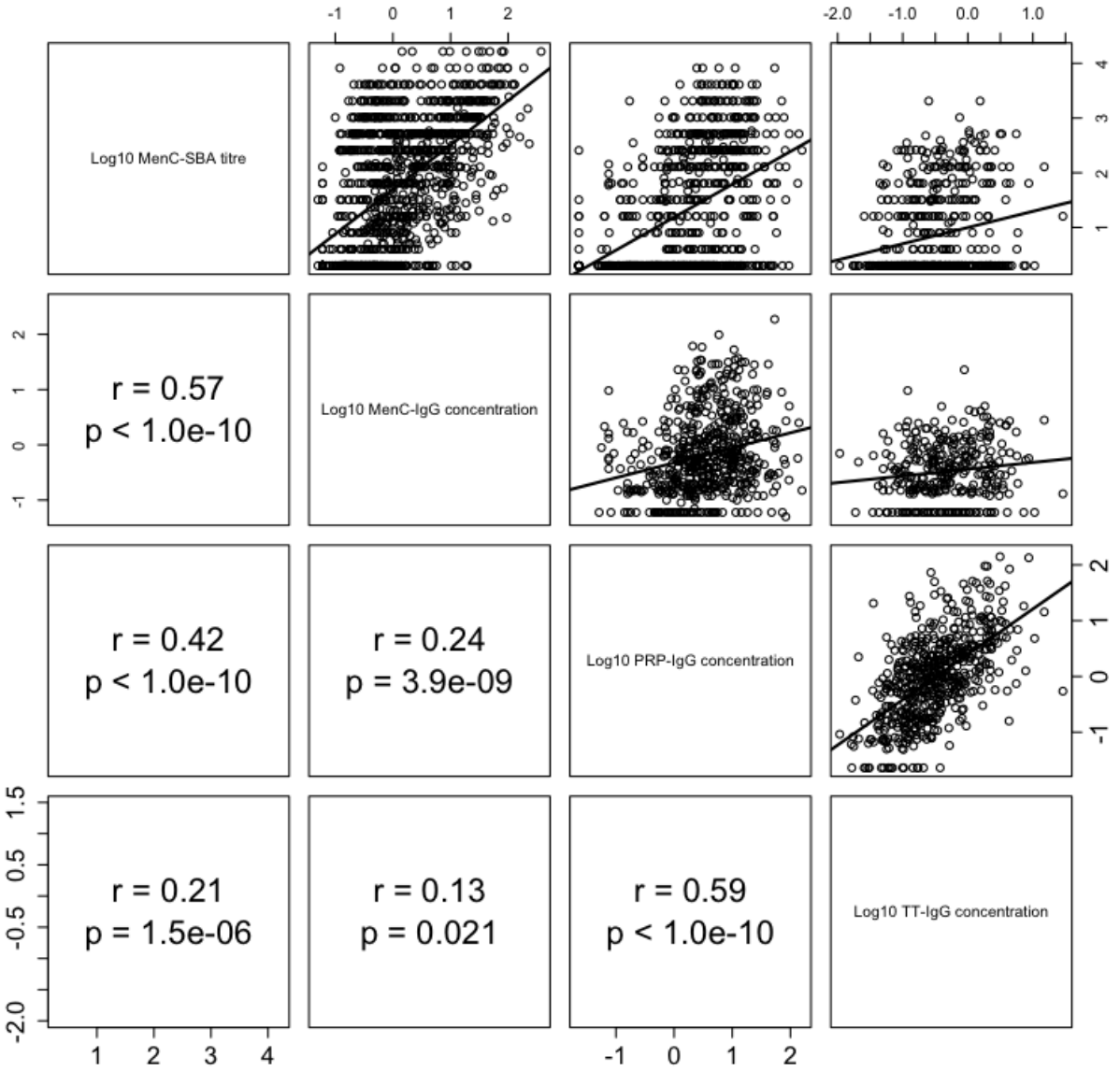


Figure 4.7: Matrix of scatter diagrams of the four serological measures of vaccine-induced immunity. On the upper-panel are the scatterplots and on the lower-panel are the respective correlation coefficients and p-values.

4.5.4.2 Pleiotropic loci

The 301 genes containing SNPs associated ($p<1\times10^{-4}$) with one or more vaccine-induced serological measure are shown in Figure 4.8. Nine of these genes were associated with more than one vaccine-induced immunological measure (Figure 4.8). MenC-specific SBA titres and MenC-specific IgG concentrations had the most shared associated genes (n=3). The gene set associated with TT-specific IgG concentrations had the fewest number of links (n=2) with the other vaccine-induced serological measures.

Of the genes associated with more than one vaccine-induced immunological measure only two, *SRSF7* and *RP11-115H18.1*, displayed the same associated SNPs between these serological measures (appendix Table 7.28). The associated SNPs within, *SRSF7* and *RP11-115H18.1*, displayed the same regression coefficient directionality and can be considered 'correlated and concordant'. If the statistical threshold is relaxed, so that SNPs with a $p<1\times10^{-3}$ in a second serological measure are considered associated; two further genes, *PIWIL1* and *DENND1A*, share SNPs in a correlated and concordant manner (appendix Table 7.28). The associated SNPs within remaining five genes did not overlap between serological measures, even at a lenient threshold of $p<0.05$ (appendix Table 7.28). Therefore, these genes appear to contain SNPs that are 'non-correlated' between serological measures and implicate different haplotypes with the corresponding association signals.

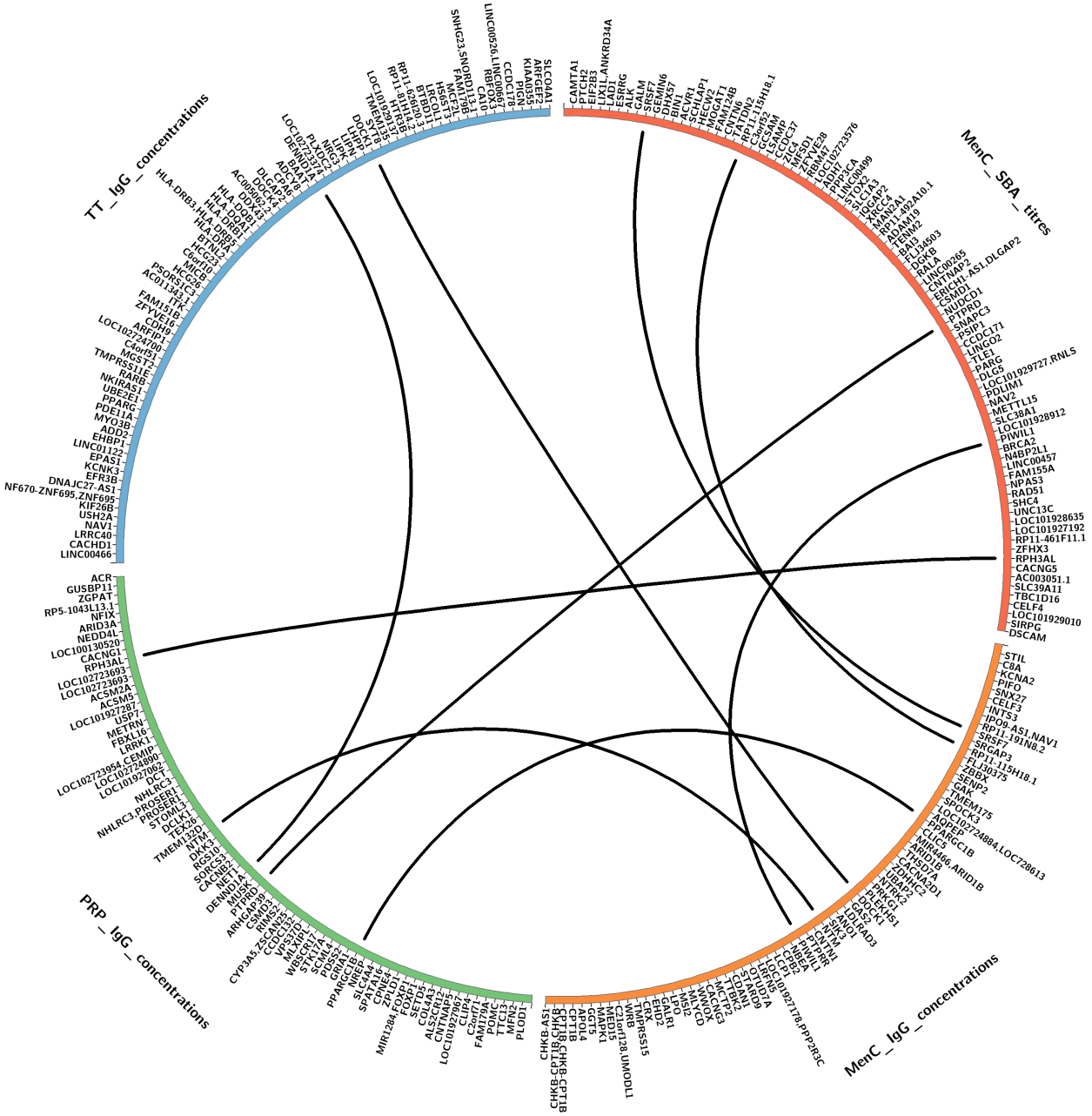


Figure 4.8: Circos plot illustrating genes containing single-nucleotide polymorphisms associated ($p < 1 \times 10^{-4}$) with one or more of four vaccine-induced serological measures. Data are included from the stage 1 genome-wide association studies of persistence of capsular group C meningococci specific (MenC)-specific IgG concentration, MenC-specific serum bactericidal titres, polyribosyl ribitol phosphate (PRP)-specific IgG concentrations and tetanus toxoid (TT)-specific IgG concentration. Genes that are associated with more than of the serological measures assessed are highlighted with links. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' function in the 'NCBI2R' R package on build 37 of the human genome (241). This diagram was generated using Circos (`http://circos.ca/`). Note: MenC SBA and MenC IgG are two measures of serological immunity to capsular group C meningococci.

4.5.5 Discussion

In this section, statistically significant correlations were observed between MenC-specific SBA titres, MenC-specific IgG concentrations, PRP-specific IgG concentrations and TT-specific IgG concentrations, following childhood immunisations. Particularly close correlations were observed between PRP- and TT-specific IgG concentrations, as well as between the two measures of MenC-specific serological immunity. Nine genes contained SNPs associated with two or more vaccine-induced serological measures; however, for the majority of these genes the associated SNPs did not correlate.

The GWASs of MenC-specific SBA titres and MenC-specific IgG concentrations displayed the greatest number of shared associated genes; however, this was only a minority of genes associated with each of these serological measures. This observation may be partly due to heterogeneity introduced by the MenC-specific IgG GWAS data arising from a subset of the participants included in the MenC-specific SBA GWAS. Furthermore, the correlation between MenC-specific IgG concentrations and MenC-specific SBA titres was modest, an observation that is consistent with previous studies (158, 357). These data highlight the differences between these two measures of MenC-induced serological immunity, which may be reflected in their genetic determinants.

Despite the close correlation between PRP- and TT-specific IgG concentrations only one gene, *DENND1A*, contained SNPs associated with both these measures. SNPs within *PTPRD* were associated with both MenC-specific SBA titres and PRP-specific IgG concentrations. Although this gene has not been annotated as having immunological functions, SNPs within this gene have also been associated with rubella specific IFN-γ responses following MMR vaccination (28). Furthermore, the putative murine homolog of *PTPRD* is expressed in the thymic medulla and by B cells (358).

The genes containing the most statistically significant associated ($p < 1 \times 10^{-5}$) SNPs within each of the GWASs did not overlap between vaccine-induced serological measures. This finding is consistent with data from a large study into pleiotropic loci among immune-related diseases, which suggested SNPs with the largest effect sizes tended to be phenotype-specific (258). The

interpretation of potentially pleiotropic loci is complicated by whether the same SNPs are responsible for the associations or whether independent haplotypes underlie associations. Interestingly, despite the positive correlation between vaccine-induced serological measures, the majority of SNPs within shared associated genes did not overlap.

This section utilised a relatively liberal association p-value threshold ($p < 1 \times 10^{-4}$), for inclusion in this exploratory analysis. However, as study sample sizes were modest, and statistical power is likely to have varied considerably between vaccine GWASs, it is not possible to say with certainty that loci that were not shared are not pleiotropic. Conversely, the statistical cut-off used was not stringent enough to say with a high degree of confidence that shared genes were not chance findings. Furthermore, comparative analysis was limited to SNPs within gene boundaries, potentially missing pleiotropic non-coding functional elements.

There are data suggesting interactions between coadministered vaccines, which result in altered immunological responses to one of more the components (359). A seminal study in this area, demonstrated lower PRP-specific IgG concentrations following Hib-TT when coadministered with a tetravalent pneumococcal polysaccharide vaccine also conjugated to TT (359). This effect was shown to be TT dose dependent, with increasing TT content in the pneumococcal vaccine resulting in decreasing PRP-specific IgG concentrations (359). This observation raised concerns about immune interference due to the use of a common protein carrier, potentially in a manner analogous to the phenomenon of carrier-induced epitope suppression (CIES). CIES is defined as interference in antibody responses to a hapten conjugated to a carrier protein due to pre-existing immunity to the carrier protein (360). There are several theories of the mechanisms underlying CIES; however, most of these data have been generated in animal models using peptide haptens. In addition to pre-existing immunity to the carrier proteins, infant immunisations present a scenario where conjugate and unconjugated carrier proteins are delivered simultaneously, and it is not known whether analogous processes to CIES occur in this context (360).

Paradoxically, enhancement due to the use of a common protein carrier has also been observed (201, 361). Concomitant use of MenC-TT conjugate vaccine with DTP-Hib vaccine

has been associated with increased PRP-specific IgG concentrations, compared to MenC-CRM coadministration (201, 361). Coadministration of vaccines can result in unpredictable immunological interactions, even when given at separate sites (201). Interactions between vaccines add another level of complexity to the determinants influencing vaccine responses, and pleiotropic genes may also be involved in the mechanisms underlying these immune interactions.

4.5.6 Conclusion

This section detailed correlations between four measures of serological immunity to three vaccine antigens. Comparisons of GWAS data assessing the genetic determinants of serological immunity to these vaccines highlighted nine genes with potentially pleiotropic roles in vaccine immunity. Further studies are needed to elucidate pleiotropic and vaccine-specific genetic factors influencing vaccine-induced immunity.

4.6 Chapter summary

This chapter described two-stage GWASs of the persistence of Hib- and TT-specific IgG concentrations following childhood immunisation. This section detailed SNPs that were associated at the level of genome-wide significance with the persistence of vaccine-induced immunity following childhood immunisation. Furthermore, while stage 2 of this study has not yet been completed, SNPs with p-values suggestive of association with the persistence of vaccine-induced immunity were also described in number of loci that have putative involvement in immunity: *FOXP1*, *MICB*, *HLA-DQA1* and the immunoglobulin light chain lambda locus. The completion of stage 2 of this study may provide the power to demonstrate further statistical associations at the level of genome-wide significance, and such findings would make excellent candidates for functional evaluation. This chapter also highlighted a number of genes that may have pleiotropic roles in the persistence of immunity to childhood immunisations.

Chapter 5

Transcriptomic analysis following childhood influenza vaccination

5.1 Overview of chapter

Influenza is an acute infection of the respiratory tract caused by three distinct segmented RNA viruses: influenza A, B and C. The clinical course of infection is usually self-limiting; however, more serious manifestations include bronchitis, secondary bacterial pneumonia or otitis media in children, and in rare cases meningitis, encephalitis or meningoencephalitis (82). In the UK, influenza vaccine is currently offered to those ≥65 years of age, aged 2 or 3 years, as well as those ≥6 months of age in at risk groups (82). The aim of this programme is to protect those at highest risk of serious illness directly by vaccination as well as indirectly by reducing viral transmission. The efficacies of influenza vaccines differ considerable by population and between influenza seasons (362, 363). Current influenza vaccine recommendations vary by population characteristics, e.g. the live attenuated vaccine is advocated for children aged 2 or 3 years, and the inactivated vaccine is prescribed for those ≥65 years (82). Enhanced understanding of the determinants of influenza vaccine immunogenicity in both 'risk' and 'transmitting' popula-

tions may aid in development of improved vaccination strategies to reduce influenza associated morbidity and mortality.

During the initial wave of the 2009 pandemic influenza A (H1N1) virus, up to 1 in 3 children were infected in some regions (364). In addition to having the highest household transmission rates, children were four times as likely to be admitted to hospital with suspected pandemic influenza (365, 366, 367). The UK Department of Health purchased two monovalent pandemic influenza vaccines for a national immunisation campaign, an AS03B adjuvanted split virion vaccine and a non-adjuvanted whole virion vaccine. Early clinical trial data found the AS03B adjuvanted influenza vaccine to be more immunogenic, but also more reactogenic, than the non-adjuvanted vaccine in healthy children aged 6 months to 12 years (282). The pandemic influenza vaccine was initially recommended as a two-dose regimen for children at high risk but was subsequently reduced to a single-dose for all children aged 6 months to 5 years, following review of early reactogenicity data (282).

In August 2010, anxieties were raised in Finland and Sweden regarding a possible link between the AS03B adjuvanted pandemic influenza vaccine and narcolepsy among children and adolescents (368, 369). Retrospective analysis of sleep centres and paediatric neurology centres in England supported an association between the AS03B adjuvanted vaccine and an increased risk of narcolepsy in childhood (370). While there remains a risk of confounders such as more rapid referral of vaccinated individuals, these data suggest the AS03B adjuvanted pandemic influenza vaccine was associated with an increased incidence of narcolepsy (369, 370, 371).

Contemporary transcriptomic approaches have transformed the way in which complex biological systems can be interrogated. The transcriptome contains the intermediary information between protein coding DNA and protein, as well as a multitude of transcribed regulators. Transcriptomic analysis has the potential to dissect the mechanisms of complex biological systems, as well as identify biomarkers of physiological or disease states. This chapter explores coding and non-coding RNA profiles following childhood immunisation with influenza vaccine, and attempts to relate these to vaccine immunogenicity and the mode of action of vaccine adjuvants.

5.1.1 Chapter aims

i) Characterise gene expression profiles in peripheral blood from children prior to and 21 days after a single dose of trivalent inactivated influenza vaccine.

ii) Relate gene expression profiles following trivalent inactivated influenza vaccine to immunogenicity.

iii) Compare expression profiles following trivalent inactivated influenza vaccine in those who previously received the AS03B adjuvanted and non-adjuvanted pandemic influenza vaccine.

iv) Characterise microRNA profiles in sera of healthy children prior to and 21 days following immunisation with pandemic influenza vaccine.

v) Assess microRNA profiles as a biomarker of immunogenicity following immunisation with pandemic influenza vaccine.

vi) Compare microRNA profiles following AS03B adjuvanted and non-adjuvanted pandemic influenza vaccines.

5.2 Gene expression following a single dose of seasonal trivalent inactivated influenza vaccine

5.2.1 Background

A number of studies have investigated gene transcript profiles following influenza vaccination, primarily focusing on the early gene signature evoked in peripheral blood following adult immunisation with either LAIV (live attenuated influenza vaccine) or trivalent inactivated influenza vaccine (TIV) (85, 87, 88, 97, 99). These studies have demonstrated a core set of differentially expressed genes (DEGs) involved in inflammatory and antimicrobial responses that are altered by both LAIV and TIV, as well as a number of genes that are specifically affected by one of these vaccines (85, 87, 88, 99).

To date, gene expression studies have focused on the first few days after immunisation with influenza vaccine, corresponding to innate immune responses; however, transcriptional events occurring at later time points, coinciding with the induction of adaptive immunity, have not been characterised (85, 87, 88, 99). Additional studies are needed to describe genes involved in these adaptive immune responses. This section describes a study assessing gene expression profiles 21 days after a single dose of TIV in a cohort of healthy children, who had received a 2009 pandemic influenza vaccine the previous year.

5.2.2 Aim

The aim of this section was assess gene expression profiles in peripheral blood following TIV, and to relate these to measures of vaccine immunogenicity.

5.2.3 Materials and methods

5.2.3.1 Participants

This study included healthy children (17 months and 13 years of age) enrolled in a clinical trial to assess antibody persistence 1 year after receiving either the AS03B adjuvanted or non-adjuvanted pandemic influenza (H1N1) vaccine, as well the immunogenicity and reactogenicity of a subsequent dose of seasonal TIV vaccine (372). Peripheral blood was collected into PAXgene™ RNA tubes pre-vaccination and 21 days following TIV (Figure 5.1). Thirty of the 323 children recruited into this clinical trial were selected for gene expression analysis (subsection '5.2.4.1 Power calculations').

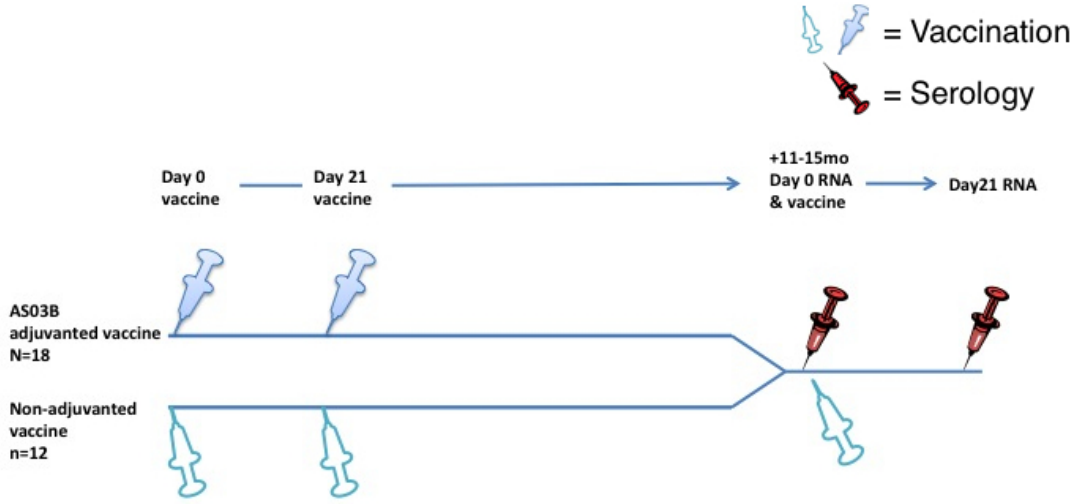


Figure 5.1: Study design for the assessment of differential gene expression following immunisation with seasonal trivalent inactivated influenza vaccine (TIV). Participants had received two doses of either AS03B adjuvanted or non-adjuvanted pandemic influenza (H1N1) vaccine the previous year. Gene expression profiles were assessed pre-vaccination and 21 days after TIV.

5.2.3.2 Vaccine

Participants received a 0.5 ml dose of Fluarix™ (GlaxoSmithKline Biologicals, UK), containing inactivated split virion influenza virus propagated in fertilised hens' eggs. This vaccine contained 15 μg of haemagglutinin antigen for each of the three seasonal influenza strains recommended by the World Heath Organisation: A/California/7/2009 (H1N1)-derived strain, A/Perth/16/2009 (H3N2)-like strain, and B/Brisbane/60/2008.

5.2.3.3 Haemagglutination inhibition assay

Haemagglutination inhibition (HAI) assays were performed by the Centre for Infections Virus Reference Department, Public Health England (London, UK), according to a standardised method, using the NIBRG121 (NIBSC, UK) a reassortant vaccine strain from an A California/7/2009 virus (372, 373, 374). Sera were pretreated with receptor destroying enzyme for 18 hours at 36 °C, followed by heat-inactivation for 1 hour at 56 °C. Sera were diluted in eight twofold dilutions starting from 1:8, and incubated with haemagglutinin (HA) antigen suspension (4 haemagglutination units/25 μl) for 1 hour at room temperature. Then 25 μl of 0.5%

turkey red blood cell suspension was added and further incubated for half an hour at room temperature. Titres were assigned as the highest reciprocal dilution that induced complete inhibition of haemagglutination. Samples that did not inhibit haemagglutination were assigned titres of 4. Sera were tested in duplicate and the geometric mean value for each pair reported.

5.2.3.4 T cell assays

Ex vivo interferon-gamma (IFN-γ) enzyme-linked immunosorbent spot (ELISpot) were performed using fresh PBMCs, by Professor Sarah C. Gilbert's laboratory, University of Oxford. The methods and results of these ELISpots are described by Lambe et al 2012 (375). In brief, 15 to 20-mer peptides pools, spanning the whole of pandemic 2009 H1N1 nucleoprotein (NP), matrix protein 1 (M1) and nonstructural protein 1 (NS1) proteins, were used to stimulate PBMCs. R10 was used as a negative control, and phytohemagglutinin (PHA) and staphylococcal enterotoxin B (SEB) were used as positive controls. IFN-γ spot-forming cells were counted after incubation at 37 °C for 18-20 hours. Spot-forming units (SFUs) per million PBMCs were calculated by subtracting the mean R10 negative control SFUs, and plates with fewer than 1000 SFUs in the positive control wells were excluded.

5.2.3.5 Whole blood RNA processing

Venepuncture was performed to collect 2.5 ml of peripheral blood into a PAXgene™ vacutainer, and this was gently inverted several times. PAXgene™ tubes were left for a minimum of 2 hours at room temperature for RNA stabilisation, and then frozen within 24 hours at −20 °C to −70 °C. Total RNA was extracted from using the PAXgene™ Blood miRNA and total RNA Kit (preanalytix, Switzerland), according to the kit handbook (May 2009 manual). Total RNA concentrations were measured using a Nanodrop® (Thermo Fisher scientific, Massachusetts), and diluted to between 25-100 ng/μl. RNA QC was performed using a Bioanalyser® (Agilent, Berkshire UK); the criteria for sample inclusion were RNA Integrity Number (RIN) >7, 28S/18S ribosomal RNA (rRNA) ratio >1.6 and an absorbance 260/280 nm ratio of 1.8-2.0.

5.2.3.6 Microarray analysis

RNA samples were randomly allocated to a microarray on an Illumina® Human HT12v4.0 Expression BeadChip, each of which contains 12 microarrays. RNA was converted into biotin labelled cRNA and hybridised to a BeadChip microarray. Hybridised microarrays were scanned using an Illumina® iScan scanner. A gene expression project was created using the Illumina® GenomeStudio software, for further quality control. Microarray hybridisation and scanning was conducted by the Wellcome Trust Centre for Human Genetics core facility (Oxford, UK). Sample probe profile raw data were extracted from Illumina® GenomeStudio version 1.9.0. Negative background intensities for each array were subtracted from their respective probe intensities prior to normalisation.

5.2.3.7 Normalisation

RNA microarrays simultaneously measure thousands of transcripts by quantifying the fluorescent intensity of hybridised dye-labelled transcripts and assume these to be proportional to the abundance of gene transcripts. However, due to differences in sample handling, quantities of starting RNAs and labelling efficiency, microarrays are not directly comparable. Therefore, it is standard procedure to normalise microarray data to minimise the variance introduced by factors that are not under investigation. Numerous microarray normalisation methods have been developed and the optimal normalisation method is highly dependent on experiment conditions (376). Four established normalisation methodologies were assessed in this section: Variance Stabilisation and Normalisation (VSN), Robust spline normalisation (RSN), median normalisation and quantile normalisation. Microarray normalisation is described in more detail in appendix subsection '7.7.1.2 Gene expression microarray normalisation'.

5.2.3.8 Probe filtering

The variability of measured intensity values increases as the fluorescent intensity decreases and approaches background. Therefore, it is necessary to assess only probes that are confidently above background. Microarray data were filtered to return probes that were significantly dif-

ferent from their local background (detection p-value of less than 0.05) in at least 65% of all the samples assessed, and only these probes were included in statistical analysis.

5.2.4 Statistical analysis

5.2.4.1 Power calculations

The power of a gene expression study is influenced by a number of factors including sample size, distribution of effect sizes, study group variance, proportion of differentially expressed genes and the acceptable error rate. Contemporary cDNA microarrays assess thousands of transcripts; therefore, stringent correction for multiple testing needs to be undertaken. Typically the error rate in microarray experiments is corrected by using a family-wise error rate (FWER) or a false-discovery rate (FDR) approach. The FWER is the probability of making one or more false positive errors, rejecting the null where an individual test has a type I error of α divided by the number of tests, as proposed by Bonferroni. While this approach controls for type I error at level α in a strong sense, it can be overly conservative and results in loss of power (377). FDR, proposed by Benjamini and Hochberg, is the expected proportion of false positives among the tests that rejected the null hypothesis and it is argued to be a more sensible and powerful approach in high-throughput data analysis (377). Consequently, the FDR approach is more commonly used in gene expression studies than the FWER method.

In this report a procedure described by Liu et al 2007 was used to calculate required study sample size (378). In brief, estimates of the proportion of non-differentially expressed genes and FDR are used to determine the 'critical value' with varying sample size. For a desired power a required sample size can be calculated. Sample size calculations were computed in the 'ssize.fdr' R package (Figure 5.2) (379). A sample size of 30, paired pre- and post-vaccination samples, was calculated to have $\geq$90% power to detect a 1.25 fold-change in gene expression following vaccination with a false discovery rate of 0.05 (380). The variance value used in this power analysis was calculated from a publicly available childhood gene expression dataset downloaded from the Gene Expression Omnibus, further details are presented in appendix subsection '7.7.1.1 Power calculations'. The 75^{th} percentile of the standard deviation of transcripts

detected in this dataset was used as the generalisable standard deviation estimate, as suggested by Wei et al 2004 (381).

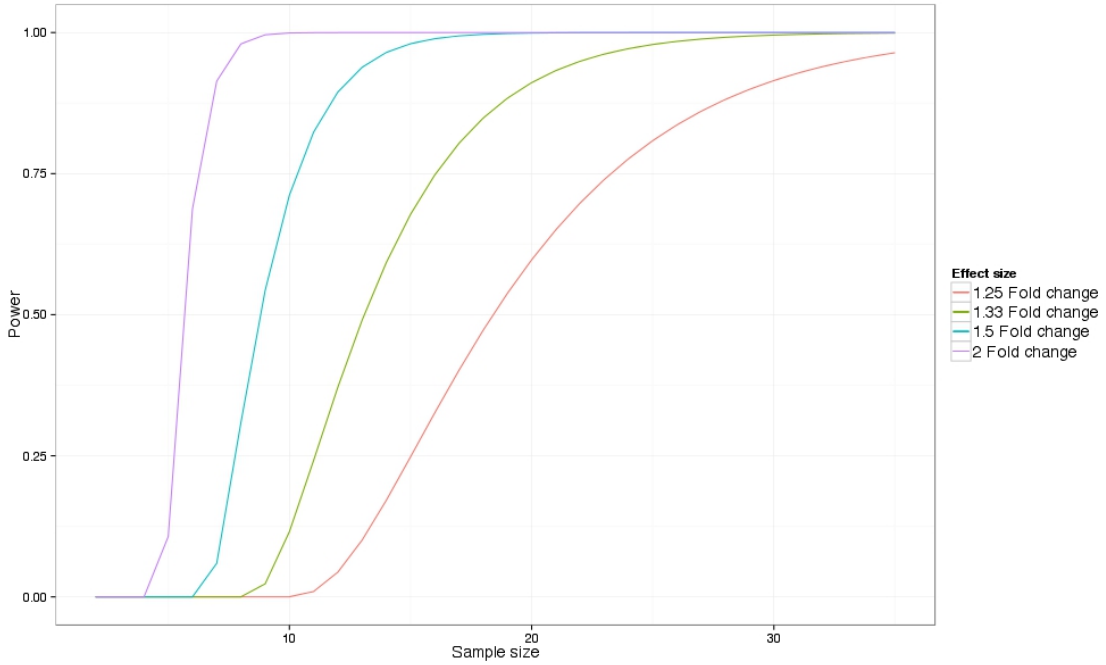


Figure 5.2: Effect of varying sample size and fold-change on the statistical power to detect differentially expressed genes following trivalent inactivated influenza vaccine.

5.2.4.2 Linear regression analysis

A linear model was fitted with normalised expression of each probe as the response variable, and each paired sample and vaccination status as predictors (Equation 5.1). The empirical Bayes method was then used to generate moderated t-statistics, moderated F-statistics, and log-odds of differential expression using the 'eBayes' function in the Limma R Package (382).

$$lmFit(expression\,of\,probe \sim sample\,pair + vaccination\,status) \tag{5.1}$$

5.2.4.3 Significance Analysis of Microarrays

Significance Analysis of Microarrays (SAM) is a method of identifying differentially expressed genes, proposed by Tusher et al 2001 (383). Essentially this method assigns a score to each probe based upon the differential expression relative to the standard deviation of the repeated

measurements. Permuting the repeated measurements of each gene enables estimates of the FDR. Analysis was completed within the 'siggenes' R package (384). The threshold for statistical significance was set as a FDR <0.05.

5.2.4.4 Gene Set Enrichment Analysis

The results of microarray data can easily be ordered by statistical evidence. However, interpreting this ranked list can be challenging; particularly if none of the genes surpass the statistical significance threshold, or if there is a long list of statistically significant genes that embody no obvious biological theme. Furthermore, single-gene analysis is unlikely to detect changes in gene expression <20%, in the sample sizes typically investigated. Nevertheless, modest individual gene expression changes may amount to considerable changes at the gene set level if several of these genes are within a particular pathway. In addition, it may be more informative to compare gene sets between studies than individual genes as these may be more resistant to methodological and analytical differences than single-gene analysis (385).

Gene Set Enrichment Analysis (GSEA) focuses on gene sets that share biological function or regulation (385, 386). The aim of this analysis is to aid interpretation of microarray results by taking into consideration existing biological knowledge. GSEA uses published biological data collated from pathway and co-expression experiments to form an inventory of gene sets (Molecular Signature Database, MSigDB). GSEA assesses whether genes within gene sets show a relationship with the list of differentially expressed genes, ranked by differential expression (385). GSEA was carried out on the list of probes ranked by differential expression, from most upregulated to most downregulated, using the Java based desktop application of GSEA v2.0.14 (385, 386).

5.2.4.5 Ingenuity Pathway Analysis

Ingenuity® Pathway Analysis (IPA) is a commercially available software package (Ingenuity® Systems, California) designed to aid interpretation of transcriptomics data. This pathway and network analysis uses the Ingenuity® Knowledge Base, which is a large manually curated

collection of data from various experimental contexts (387). IPA was conducted to assess potential biological relevance of gene expression data, using the web-based software package (`http://www.ingenuity.com`).

5.2.4.6 Relationship between differential gene expression following trivalent inactivated influenza vaccine and previous pandemic influenza vaccine

To assess whether differential gene expression following TIV was influenced by which of the pandemic influenza vaccines was administered the previous year, a linear model was fitted with $\log_2$ fold-changes of each probe as the response variable and the previous years pandemic influenza vaccine as a predictor variable (Equation 5.2). The empirical Bayes method was used to generate moderated t-statistics using the 'eBayes' function in the Limma R Package (382).

$$lmFit\Big(log_2(Transcript\ fold\text{-}change) \sim pandemic\,vaccine\,type\Big) \tag{5.2}$$

5.2.4.7 Relationship between differential gene expression following trivalent inactivated influenza vaccine and immunological responses to TIV

The relationship between differential gene expression following TIV and the level of immunological response was interrogated using linear regression. A linear model was fitted with $\log_2$ fold-changes of each probe as the response variable and the post-vaccination $\log_{10}$ transformed immunogenicity measures, baseline $\log_{10}$ transformed immunogenicity measures, and the previous years 2009 pandemic influenza vaccine, as predictor variables. These linear models are summarised in Equations 5.3, 5.4, 5.5, and 5.6. The empirical Bayes method was used to generate moderated t-statistics using the 'eBayes' function in the Limma R Package (382).

$$lmFit\Big(log_2(Transcript\ fold\text{-}change) \sim log_{10}(post\,vaccination\,Haemagglutination\,Inhibition\,titres) + log_{10}(baseline\,Haemagglutination\,Inhibition\,titres) + pandemic\,vaccine\Big) \tag{5.3}$$

$$lmFit\Big(log_2(Transcript\ fold\text{-}change) \sim log_{10}(post\ vaccination\ nucleoprotein\ ELISpots) + log_{10}(baseline\ nucleoprotein\ ELISpots) + pandemic\ vaccine\Big) \tag{5.4}$$

$$lmFit\Big(log_2(Transcript\ fold\text{-}change) \sim log_{10}(post\ vaccination\ matrix\ protein\ 1\ ELISpots) + log_{10}(baseline\ matrix\ protein\ 1\ ELISpots) + pandemic\ vaccine\Big) \tag{5.5}$$

$$lmFit\Big(log_2(Transcript\ fold\text{-}change) \sim log_{10}(post\ vaccination\ nonstructural\ protein\ 1\ ELISpots) + log_{10}(baseline\ nonstructural\ protein\ 1\ ELISpots) + pandemic\ vaccine\Big) \tag{5.6}$$

5.2.5 Results

5.2.5.1 Participant demographics

Three hundred and two participants were vaccinated as part of a multi-centre study to assess antibody persistence 1 year after receiving either the AS03B adjuvanted or non-adjuvanted pandemic influenza (H1N1) vaccine, as well the immunogenicity and reactogenicity of a subsequent dose of seasonal TIV vaccine (372). Ninety-nine participants were recruited by the Oxford Vaccine Group. Participants were grouped into those aged 6 months to less than 3 years (younger group) and 3 to 12 years (older group) at the time they received the pandemic influenza vaccine, as differences in immunogenicity and reactogenicity were observed between these age-stratified groups (282). Paired pre- and post-vaccination RNA PAXgene™ tubes were available for 17 of the younger cohort and 40 of the older cohort. Therefore, microarray analysis was restricted to the older cohort as this was a larger group and mixing the two groups may have introduced greater heterogeneity. Of the 40 paired RNA PAXgene™ samples 23 had received the adjuvanted vaccine and 17 the non-adjuvanted vaccine. Nine samples failed RNA QC metrics, five from the adjuvanted and four from the non-adjuvanted group. Thirty of the remaining participants were randomly chosen for microarray analysis, composed of 18 participants who had received the adjuvanted vaccine and 12 participants who had received the non-adjuvanted vaccine. Study demographics are displayed in Table 5.1 and appendix Table 7.30.

Table 5.1: Participant demographics in the study of peripheral blood gene expression profiles following immunisation with seasonal trivalent inactivated influenza vaccine.

Demographic	Details	Value[#]
Ethnicity		
	Caucasian (%)	30(100)
Sex		
	Male (%)	14 (47)
	Female (%)	16 (53)
Previous pandemic influenza vaccine	Adjuvanted (%)	18 (60)
	Non-adjuvanted (%)	12 (40)
Age		
	Months (IQR)	122* (94-151)
Time since vaccination		
	Days (IQR)	21* (20-21)

[#] = number of children unless specified,* = median, IQR= interquartile range

5.2.5.2 Gene transcript microarray quality control

Rigorous QC protocols are needed to ensure the quality and reproducibility of microarray data. As these experiments represent many thousands of data points, assessing the quality of each datum individually is impractical. Therefore, several analytical methods are used to assess performance and identify possible outliers. The QC described in this section were based on variance-based statistical methods, whereby each datum is compared to the distribution of the larger population.

Of the 47231 probes on the RNA microarray 16298 were detected in 65% of the whole blood samples assessed and were included in statistical analysis. Box and whisker plots can be used to visualise probe intensity distributions on a per microarray basis, and are useful in assessing microarray normalisation. Box and whisker plots following each of the normalisation methods evaluated are displayed in appendix subsection '7.7.1.2 Gene expression microarray normalisation'. It is a prerequisite for many statistical techniques, such as linear modelling, that probe variance is stable over different expression values (376). Therefore, an optimally normalised

microarray experiment should show little or no variance to mean expression dependence. Plots of probe standard deviations versus ranked probe means following each of the normalisation methods are presented in appendix subsection '7.7.1.2 Gene expression microarray normalisation'. Following review of box and whisker and standard deviations plots using four different normalisation methods, robust spline normalisation (RSN) was selected as the method of choice. The box and whisker, and standard deviations plots for RSN normalised probe intensities are shown in Figure 5.3 and Figure 5.4, respectively.

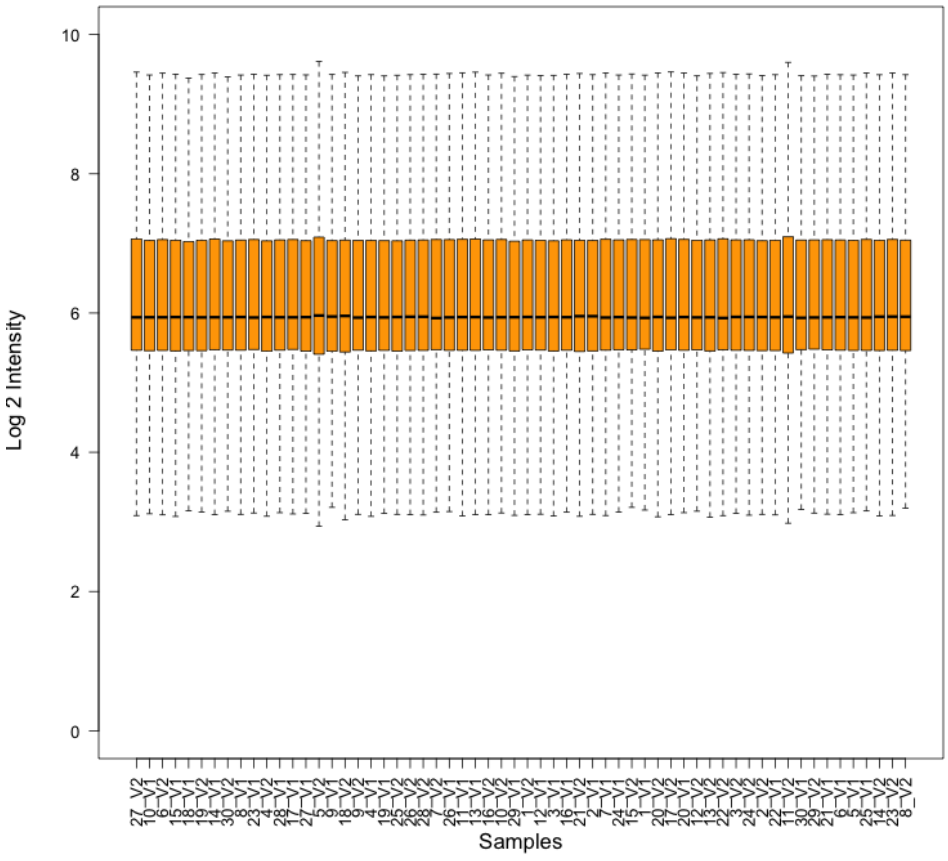


Figure 5.3: Box and whisker plots showing the distribution of robust spline normalised RNA feature intensities, from the study of gene expression profiles following immunisation with trivalent inactivated influenza vaccine. This plot summarises data from all 47231 probes. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. V1= pre-vaccination; V2= 21 days post-vaccination. (Note sample numbers have been anonymised $1 - 30$)

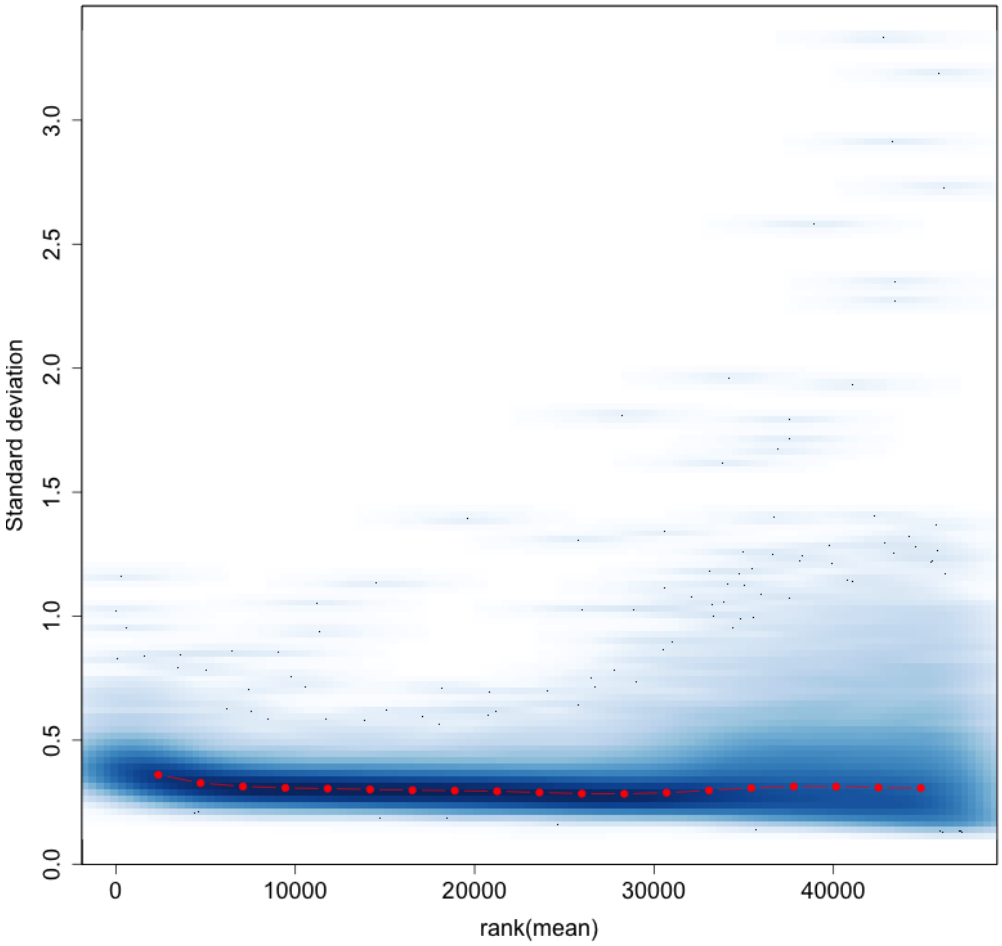


Figure 5.4: Standard deviations against the ranked mean probe intensities for robust spline normalised RNA data, from the study of gene expression profiles following immunisation with trivalent inactivated influenza vaccine. The red dots show the running median estimator, which is approximately horizontal as there is little evident of standard deviation-mean dependence.

In addition to giving a broad view of sample normalisation, Figure 5.3 shows that two of the post-vaccination samples (5_$V2$ and 11_$V2$) have a wider interquartile range of probe intensities. Furthermore, while the density plots displayed in Figure 5.5 shows most of the samples follow a similar distribution, the two samples highlighted as having disparate box and whisker plots display a relative abundance of lower intensity probes.

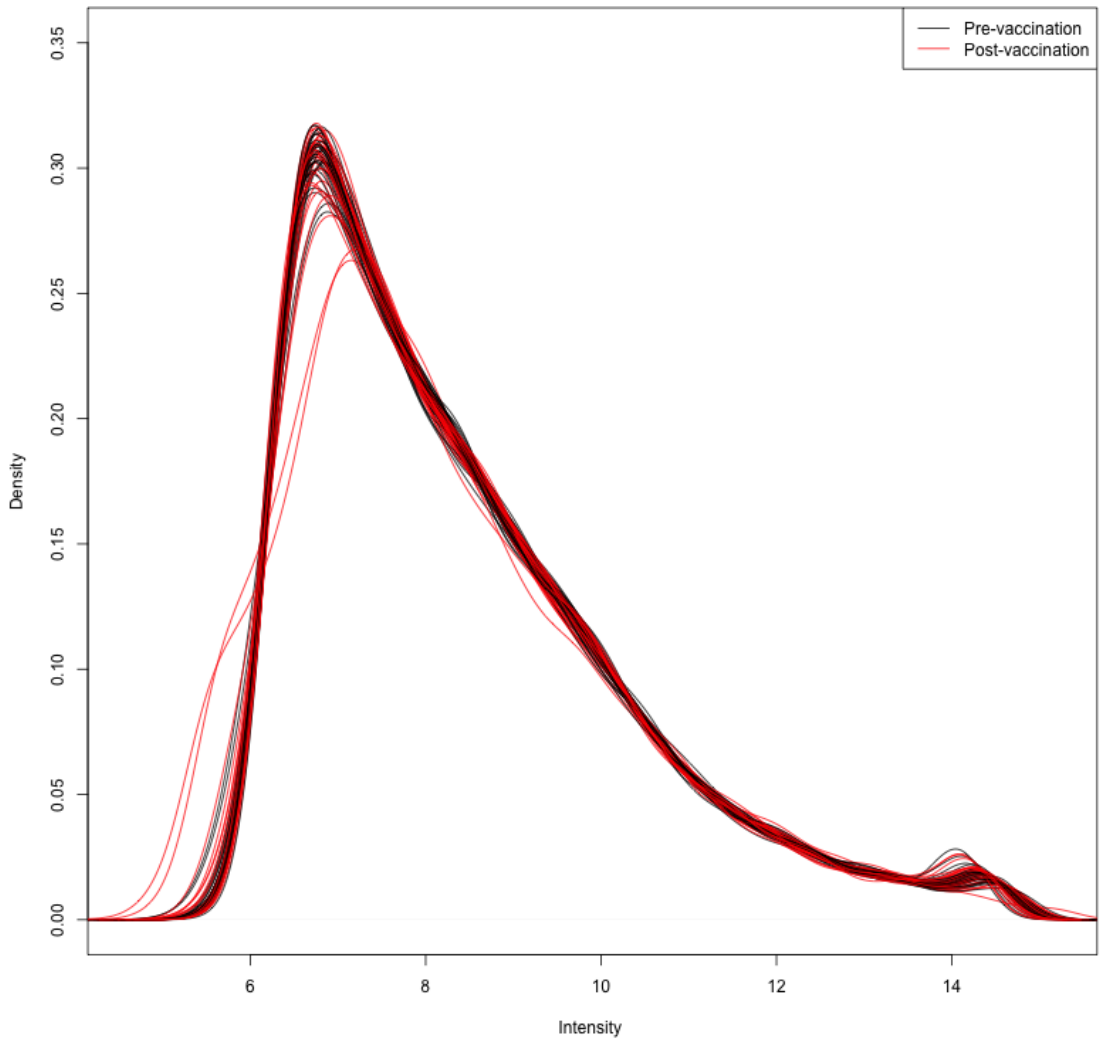


Figure 5.5: Density of transcripts with varying robust spline normalised expression values for each sample from the study of gene expression profiles following immunisation with trivalent inactivated influenza vaccine. Pre-vaccination samples are coloured black and post-vaccination are coloured red. These data have been filtered as described in subsection '5.2.3.8 Probe filtering'

Hierarchical clustering analysis was conducted by the complete-linkage method using the 'hclust' function of the R 'stats' package (Figure 5.6) (153). This agglomerative clustering approach assigns each sample to its own cluster and then iteratively joins the two most similar clusters until there is only one cluster remaining. The tighter cluster is ordered to be on the left of the subtree. A noteworthy feature of Figure 5.6 is that for only six of the participants did the pre- and post-vaccination samples cluster together. This is somewhat surprising as samples from the same individual might be expected to display greater similarity than samples

from other individuals. There was little evidence of samples clustering by treatment group, so vaccination does not appear to explain this observation. Furthermore, there was no sign of clustering by BeadChip (data not shown). The two samples (11_V2 and 5_V2) previously identified as possible outliers again showed the greatest degree of dissimilarities, relative to the rest of the samples.

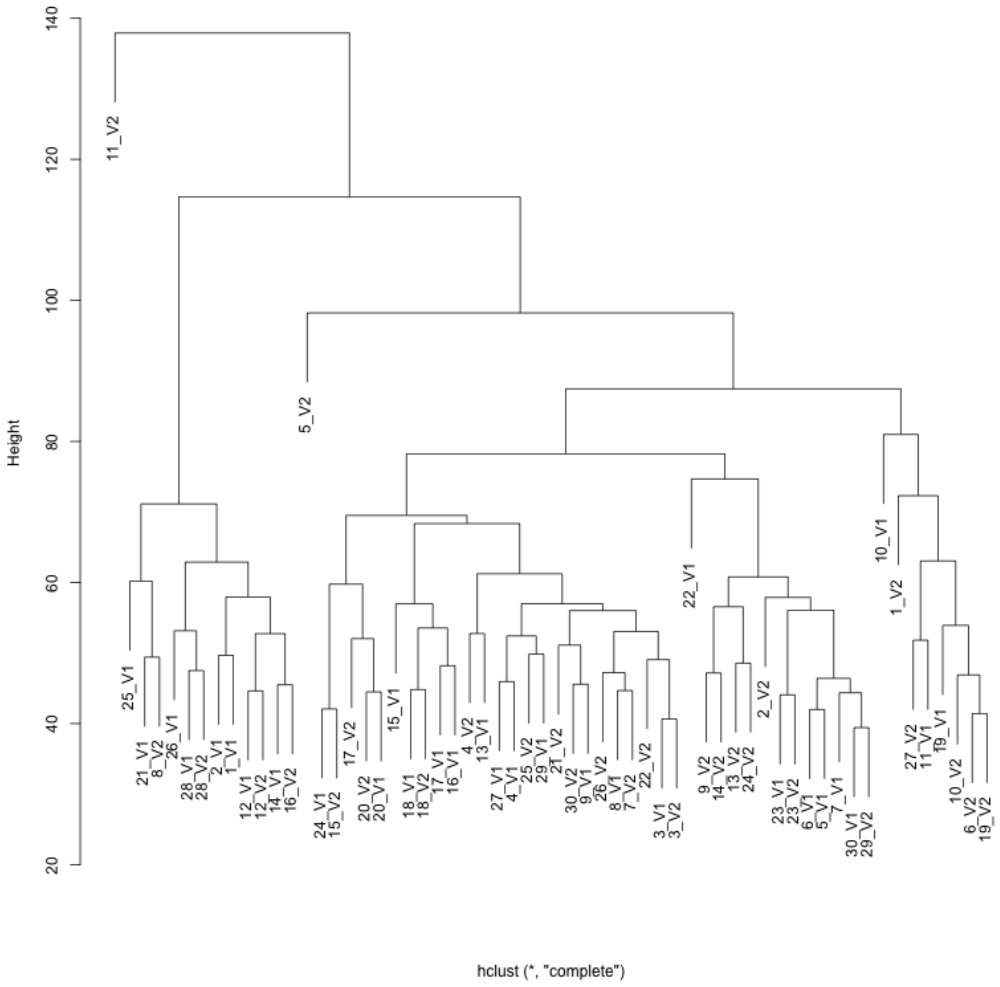


Figure 5.6: Hierarchical clustering of samples by robust spline normalised gene expression data, in the study of gene expression profiles following trivalent inactivated influenza vaccine. These data have been filtered as described in subsection '5.2.3.8 Probe filtering'. V1= pre-vaccination; V2= 21 days post-vaccination. (Note sample numbers have been anonymised $1 - 30$)

The first and second principal components of RSN normalised gene expression data are plotted in Figure 5.7. These were derived from principal components analysis (PCA) conducted using the 'prcomp' function in the R 'stats' package (153). PCA is a method of finding pro-

jections of maximum variability. The first PC accounts for the most variance followed by the second PC and so on, with each subsequent PC being uncorrelated with previous PCs (388, p. 304-305). Figure 5.7 shows that the post-vaccination sample 11 lies some distance from the other samples based on both the first and second principle component. The post-vaccination sample 5 seems to be an outlier only on PC1 and clusters with the rest of the samples in PC2. There is no clear separation of samples based up vaccination status.

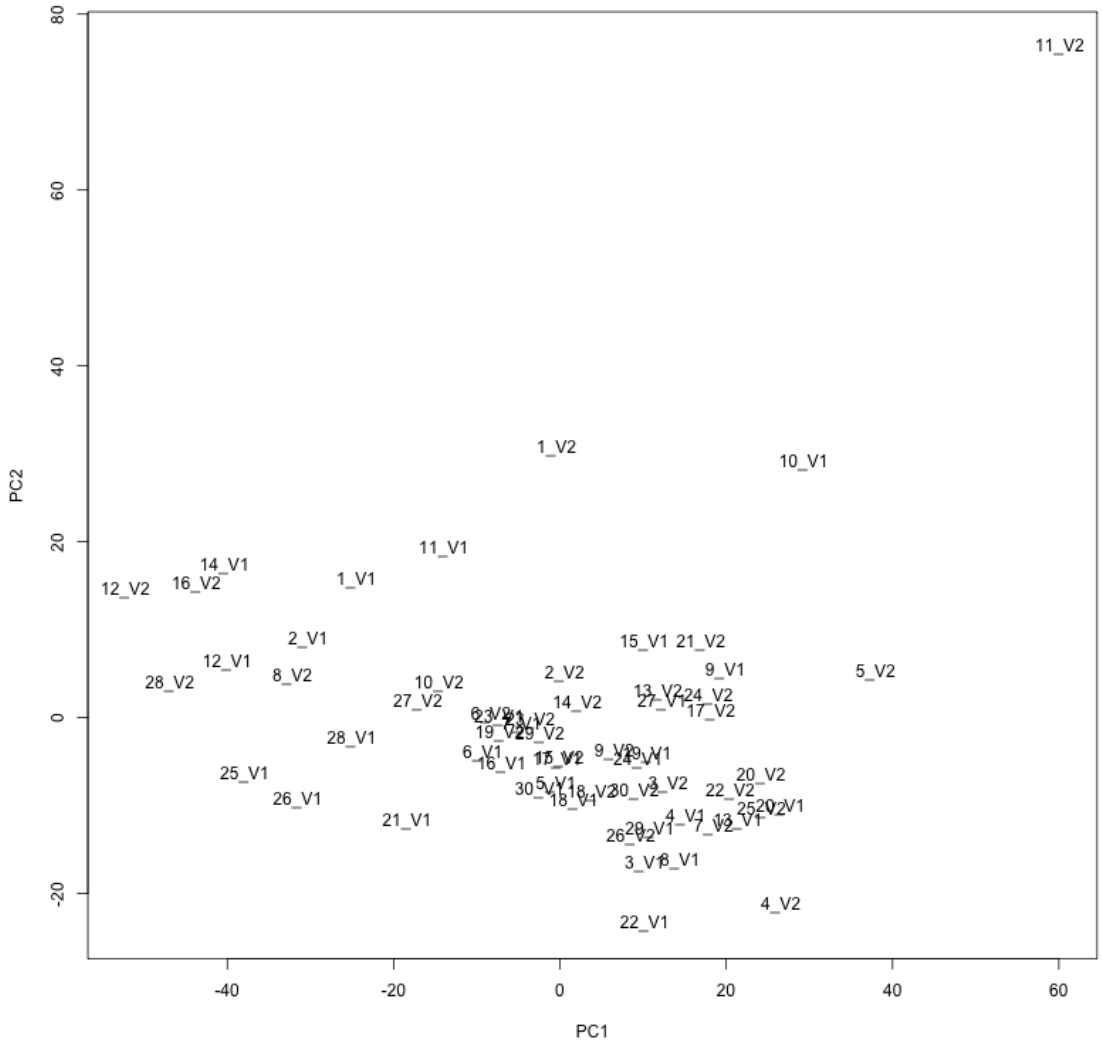


Figure 5.7: The first two principal components of robust spline normalised gene expression data in the study of gene expression profiles following immunisation with trivalent inactivated influenza vaccine. These data have been filtered as described in subsection '5.2.3.8 Probe filtering'. V1= pre-vaccination; V2= 21 days post-vaccination. (Note sample numbers have been anonymised $1-30$).

A heat map of between sample correlation coefficients is shown in Figure 5.8, using the 'heatmap.2' function of 'gplots' R package (389). While the expression values of all samples are highly correlated, the post-vaccination sample 11 is noticeably less well correlated with the rest of the samples. As the QC procedure described in this section are derived solely from these data, it is not easy to discern a cut-off where 'outlier' samples merit exclusion. Therefore, rather than excluding samples, sensitivity analysis was conducted with and without the potential outlier samples 5 and 11.

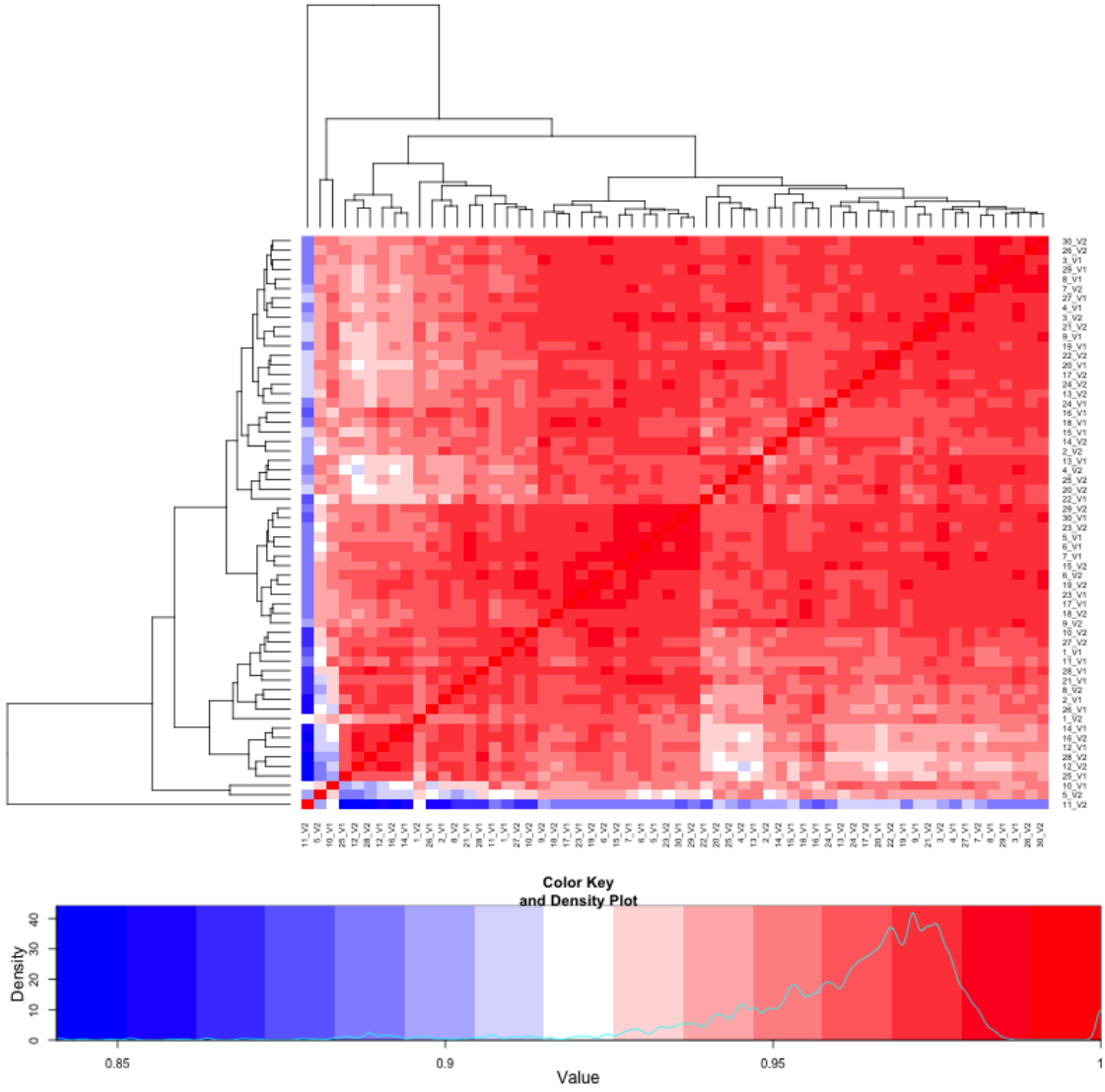


Figure 5.8: Heat map of correlation coefficients between robust spline normalised gene expression data for all of the samples assessed in the study of gene expression profiles following immunisation with trivalent inactivated influenza vaccine. These data were filtered as described in subsection '5.2.3.8 Probe filtering'. V1= pre-vaccination; V2= 21 days post-vaccination. (Note sample numbers have been anonymised $1-30$).

5.2.5.3 Linear regression analysis

A linear model was used to test for differential expression between participant paired pre- and post-vaccination samples. A model matrix was created with contrasts coded using 'dummy' variables for each pair of samples and for vaccination status. A linear model was fitted to the expression data for each probe with the design matrix variables as predictors by generalised least squares, using the 'lmFit' function of the 'limma' R Package (382). The empirical Bayes method was then used to generate moderated t-statistics, moderated F-statistic, and log-odds of differential expression using the 'limma' R package (382). The top 25 transcripts, ranked in order of evidence of differential expression are shown in Table 5.2. For sensitivity analysis, the linear model was retested with the two possible outlier samples excluded, the top 25 transcripts following this analysis are shown in appendix Table 7.29 and were similar to the results from the total dataset, with nine of the top 25 transcripts being shared.

Table 5.2: The top 25 transcripts, ranked in order of statistical evidence for differential expression, 21 days following immunisation with trivalent inactivated influenza vaccine.

Gene ID	Log_2FC	FC	p-value	FDR	Definition
GAPT	-0.35	0.79	0.000019	0.31	GRB2-binding adaptor protein
GAPT	-0.26	0.83	0.000312	0.67	GRB2-binding adaptor protein
SIT1	0.21	1.16	0.000345	0.67	signaling threshold regulating transmembrane adaptor 1
C14ORF4	0.25	1.19	0.000368	0.67	chromosome 14 open reading frame 4
PGRMC1	-0.27	0.83	0.000474	0.67	progesterone receptor membrane component 1
XK	-0.37	0.77	0.000522	0.67	X-linked Kx blood group (McLeod syndrome)
HS.282800	-0.28	0.82	0.000551	0.67	cDNA clone GLCBPH07 3
TOR2A	0.19	1.14	0.000668	0.67	torsin family 2, member A
ARRDC3	-0.23	0.85	0.000679	0.67	arrestin domain containing 3
WDR74	0.21	1.15	0.000710	0.67	PREDICTED: WD repeat domain 74
TLE4	-0.23	0.85	0.000749	0.67	transducin-like enhancer of split 4
MRPS5	0.18	1.13	0.000784	0.67	mitochondrial ribosomal protein S5
TSPAN2	-0.25	0.84	0.000836	0.67	tetraspanin 2
LOC387882	-0.23	0.85	0.001130	0.67	hypothetical protein
NBN	-0.21	0.86	0.001131	0.67	nibrin
LOC100132317	0.30	1.23	0.001150	0.67	PREDICTED similar to C20orf69 protein
RPL23AP13	0.26	1.20	0.001156	0.67	ribosomal protein L23a pseudogene 13
TMEM173	0.19	1.14	0.001203	0.67	transmembrane protein 173
P2RY10	-0.23	0.85	0.001268	0.67	purinergic receptor P2Y, G-protein coupled, 10
RGS18	-0.35	0.78	0.001445	0.67	regulator of G-protein signaling 18
TRIP4	-0.17	0.89	0.001516	0.67	thyroid hormone receptor interactor 4
TXN2	-0.17	0.89	0.001521	0.67	thioredoxin 2
FCER1A	-0.45	0.73	0.001527	0.67	Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide
PPM1A	-0.28	0.82	0.001570	0.67	protein phosphatase magnesium-dependent 1A
LOC729580	0.22	1.16	0.001570	0.67	PREDICTED: hypothetical LOC729580 mRNA.

FC= Fold change, FDR= False discovery rate, ORF= open reading frame

The results of linear regression analysis show that none of the probes surpassed the level of statistical significance (FDR <0.05) when corrected for multiple testing (Table 5.2). However, the number of transcripts with a p-value <0.05 was 1198, which is 7.3% of the 16298 tests conducted showing a small enrichment of smaller p-values. Thirteen probes had a p-values <0.001 and 30 had a fold-change >1.25 (Figure 5.9).

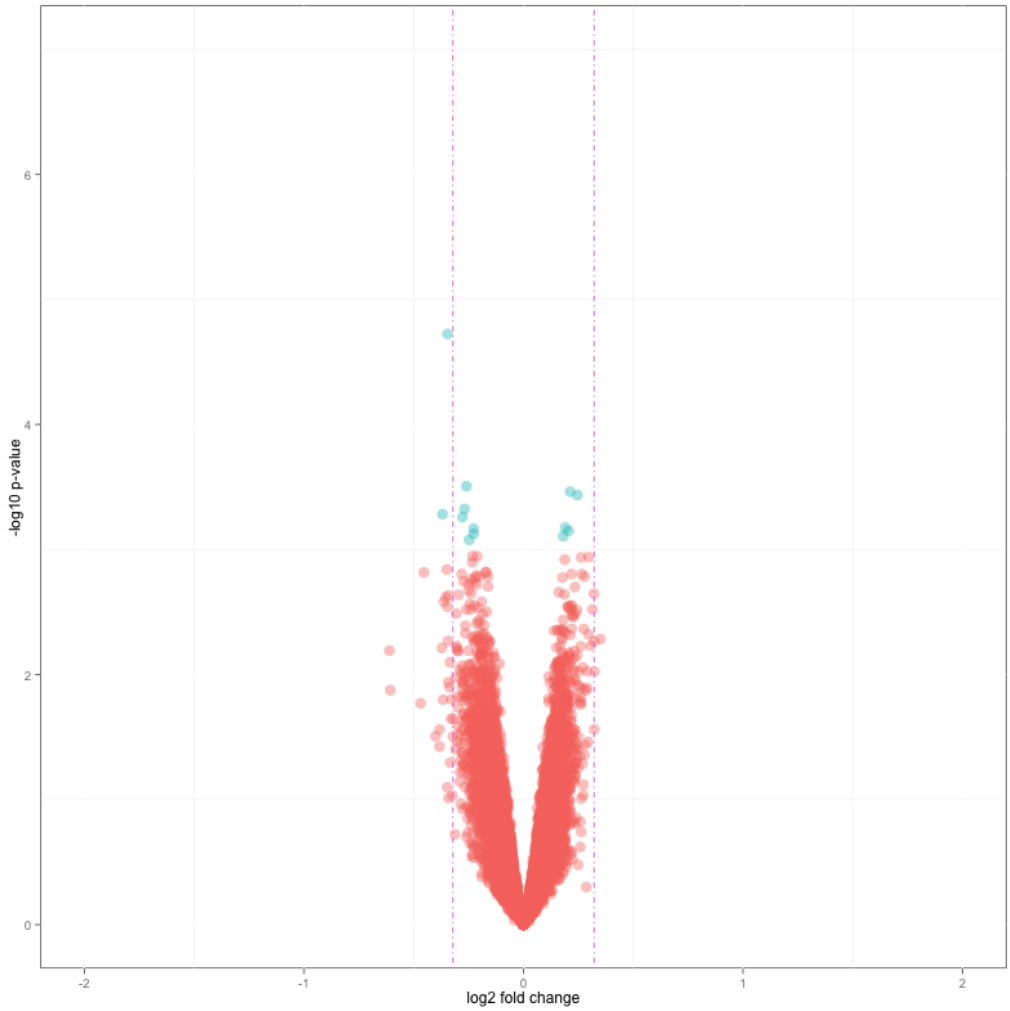


Figure 5.9: Volcano plot of fold-change and p-value of gene expression changes from baseline to 21 days post-vaccination with trivalent inactivated influenza vaccine. The violet lines show probes with a fold-change of greater than 1.25 and the green dots indicate differentially expressed genes a p-value <0.001

5.2.5.4 Significance Analysis of Microarrays

Significance Analysis of Microarrays (SAM) is an alternative method for accounting for the multiple corrections in microarray data analysis (383). SAM analysis was completed on paired pre- versus post-vaccination samples, using the 'sam' function in the 'siggenes' R Package (384). The results of this analysis are displayed in Figure 5.10, a single probe surpassed the level of statistical significance (FDR <0.05) following corrections for multiple testing. This probe corresponded to the *GAPT* gene, which was also the gene with the greatest statistical evidence of differential expression from the linear regression analysis conducted using limma.

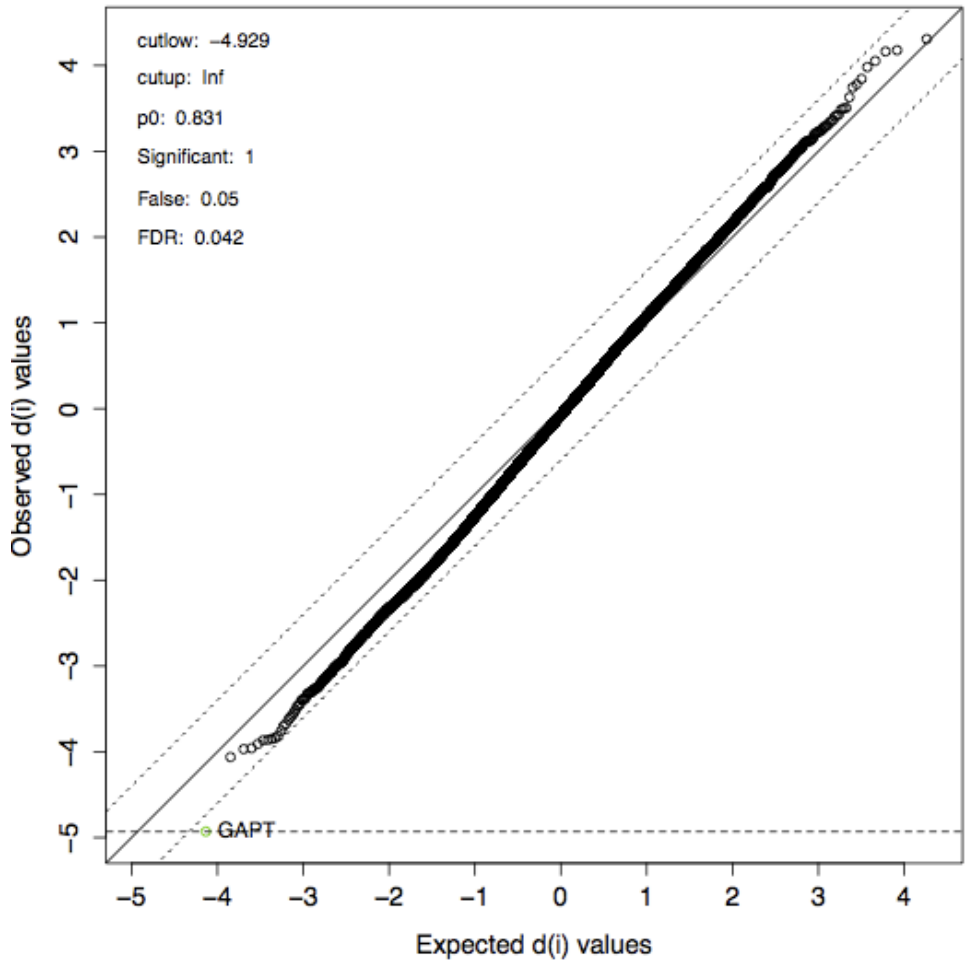


Figure 5.10: Identification of differentially expressed genes in peripheral blood expression 21 days after immunisation with trivalent influenza vaccine using the Significance Analysis of Microarrays method (383). Displayed is the distribution of expected $d(i)$ versus observed $d(i)$, where $d(i)$ = relative differences. The hashed lines mark the delta (distance between observed and expected test scores) corresponding to a FDR of 0.042. Probes with a FDR of <0.05 are highlighted in green

5.2.5.5 Gene Set Enrichment Analysis

GSEA was undertaken on the entire list of filtered probes, ranked by their $\log_2$ fold-change values. This analysis was carried out using 'GseaPreranked' tool in the Java based desktop application of GSEA v2.0.14 (385, 386). Analysis was completed using the 1910 gene sets in the 'c7:Immunological signature' database identified from microarray experiments of gene expression in immunological studies (`http://www.broadinstitute.org/gsea/msigdb/index.jsp`).

GSEA found 12/1910 and 148/1910 gene sets were upregulated and downregulated 21 days post-vaccination (FDR <0.05), respectively. The top 25 upregulated and downregulated gene sets from this GSEA are displayed in Table 5.3 and Table 5.4. The top 4 enrichment plots for differentially upregulated and downregulated gene sets are presented in appendix Figure 7.21 and appendix Figure 7.22, respectively.

The gene sets included in GSEA were derived from a variety of immunological experiments and not all members may be involved in a particular biological process. Genes shared between many of the high-scoring gene sets may be the most informative. Leading-edge analysis extracts the genes in each of the gene sets that contribute to its enrichment score (385). Leading-edge genes shared between four or more of the top 25 upregulated and downregulated genes sets (ranked by their enrichment score) following TIV are shown in Figure 5.11 and Figure 5.12, respectively. Eighty-six leading edge genes were present in four or more of the upregulated gene sets (Figure 5.11), only 18 genes were shared to this degree in the downregulated gene sets (Figure 5.12).

Table 5.3: Top 25 upregulated gene sets 21 days post-vaccination with trivalent inactivated influenza vaccine, generated using Gene Set Enrichment Analysis. False discovery rates and p-values were deduced by comparing the enrichment score of each gene set on the ranked experimental data set to 1000 permutations of the unranked experimental data set.

Gene set name	Size	FDR q-value	FWER p-value
GSE22886: TCELL VS BCELL NAIVE UP	182	0.052	0.086
GSE9006: TYPE 1 VS TYPE 2 DIABETES PBMC AT DX UP	172	0.038	0.123
GSE13485: DAY1 VS DAY3 YF17D VACCINE PBMC UP	20	0.053	0.239
GSE18791: UNSTIM VS NEWCATSLE VIRUS DC 1H UP	37	0.042	0.246
GSE17974: CTRL VS ACT IL4 AND ANTI IL12 0.5H CD4 TCELL UP	62	0.037	0.266
GSE6269: FLU VS E COLI INF PBMC UP	138	0.036	0.298
GSE6269: HEALTHY VS FLU INF PBMC DN	134	0.032	0.307
GSE29618 LAIV VS TIV FLU VACCINE DAY7 MONOCYTE DN	117	0.030	0.328
GSE360: L DONOVANI VS B MALAYI HIGH DOSE MAC DN	157	0.031	0.372
GSE6269: E COLI VS STREP AUREUS INF PBMC DN	148	0.034	0.433
GSE22886: NAIVE CD4 TCELL VS MEMORY TCELL UP	98	0.033	0.461
GSE13485: DAY1 VS DAY7 YF17D VACCINE PBMC UP	26	0.043	0.569
GSE22886: CD4 TCELL VS BCELL NAIVE UP	182	0.045	0.615
GSE37416: CTRL VS 0H F TULARENSIS LVS NEUTROPHIL UP	74	0.047	0.666
GSE2706: LPS VS R848 AND LPS 2H STIM DC UP	70	0.061	0.784
GSE3982: DC VS NEUTROPHIL LPS STIM DN	114	0.067	0.834
GSE24634: TEFF VS TCONV DAY5 IN CULTURE DN	166	0.072	0.875
GSE1341:1 IGM VS SWITCHED MEMORY BCELL UP	64	0.069	0.882
GSE9006: HEALTHY VS TYPE 2 DIABETES PBMC AT DX UP	172	0.067	0.884
GSE20715: WT VS TLR4 KO 24H OZONE LUNG DN	127	0.064	0.885
GSE13306: RA VS UNTREATED TCONV UP	117	0.062	0.890
GSE26495: NAIVE VS PD1LOW CD8 TCELL UP	118	0.060	0.892
GSE3982: BCELL VS BASOPHIL UP	129	0.061	0.907
GSE22886: CD8 TCELL VS BCELL NAIVE UP	186	0.060	0.908
GSE17974: CTRL VS ACT IL4 AND ANTI IL12 1H CD4 TCELL UP	117	0.065	0.934

Gene sets are detailed in the Molecular Signatures Database (`http://www.broadinstitute.org/gsea/msigdb/genesets.jsp?collection=C7`)
FDR= false discovery rate, FWER= family wise-error rate

Table 5.4: Top 25 downregulated gene sets 21 days post-vaccination with trivalent inactivated influenza vaccine, generated using Gene Set Enrichment Analysis. False discovery rate and p-values were deduced by comparing the enrichment score of each gene set on the ranked experimental data set to 1000 permutations of the unranked experimental data set.

Gene set name	Size	FDR q-value	FWER p value
GSE13485: DAY1 VS DAY21 YF17D VACCINE PBMC DN	138	0.001	0.001
GSE34205: RSV VS FLU INF INFANT PBMC UP	93	0.048	0.067
GSE13485: CTRL VS DAY1 YF17D VACCINE PBMC UP	142	0.053	0.110
GSE26928: EFF MEM VS CENTR MEM CD4 TCELL DN	87	0.043	0.119
GSE2706: R848 VS R848 AND LPS 8H STIM DC DN	103	0.071	0.234
GSE15750: DAY6 VS DAY10 EFF CD8 TCELL UP	94	0.064	0.252
GSE3982: NEUTROPHIL VS BASOPHIL DN	159	0.057	0.260
GSE339: EX VIVO VS IN CULTURE CD4POS DC DN	135	0.050	0.261
GSE9037: WT VS IRAK4 KO BMDM UP	125	0.048	0.275
GSE1460: CD4 THYMOCYTE VS NAIVE CD4 TCELL ADULT BLOOD UP	151	0.049	0.304
GSE22886: NAIVE CD4 TCELL VS MEMORY TCELL DN	159	0.047	0.319
GSE3982: BASOPHIL VS EFF MEMORY CD4 TCELL UP	163	0.043	0.321
GSE17721: POLYIC VS CPG 0.5H BMDM DN	132	0.048	0.377
GSE29617: CTRL VS DAY7 TIV FLU VACCINE PBMC 2008 UP	122	0.045	0.379
GSE1460: NAIVE CD4 TCELL CORD BLOOD VS THYMIC STROMAL CELL DN	141	0.043	0.383
GSE1460: CD4 THYMOCYTE VS NAIVE CD4 TCELL CORD BLOOD UP	163	0.040	0.383
GSE3982: MAST CELL VS NKCELL UP	133	0.039	0.388
GSE10239: NAIVE VS DAY4.5 EFF CD8 TCELL DN	133	0.037	0.393
GSE26928: CENTR MEMORY VS CXCR5 POS CD4 TCELL DN	113	0.036	0.398
GSE29615: DAY3 VS DAY7 LAIV FLU VACCINE PBMC DN	86	0.034	0.398
GSE39820: CTRL VS TGFBETA1 IL6 IL23A CD4 TCELL UP	143	0.032	0.400
GSE8515: CTRL VS IL6 4H STIM MAC UP	115	0.031	0.400
GSE14000: TRANSLATED RNA VS MRNA 4H LPS DC UP	154	0.030	0.407
GSE17974: CTRL VS ACT IL4 AND ANTI IL12 0.5H CD4 TCELL DN	112	0.030	0.412
GSE20366: TREG VS NAIVE CD4 TCELL DEC205 CONVERSION DN	97	0.029	0.418

Gene sets are detailed in the Molecular Signatures Database (`http://www.broadinstitute.org/gsea/msigdb/genesets.jsp?collection=C7`)
FDR= false discovery rate, FWER= family wise-error rate

The results of the leading-edge analysis of genes upregulated following TIV vaccination highlighted a high degree of overlapping genes in a group of experiments classifying genes that were upregulated in T cells compared to myeloid cells, monocytes, NK cells and naïve B cells (Figure 5.11). Analysis of genes downregulated following TIV showed sharing of leading-edge genes between gene sets downregulated in naïve T cells compared to $CD4^+$ T cells and $CD8^+$ effector T cells. Furthermore, another cluster of leading-edge genes were shared between a number of gene sets catalogued by human vaccine studies (Figure 5.12).

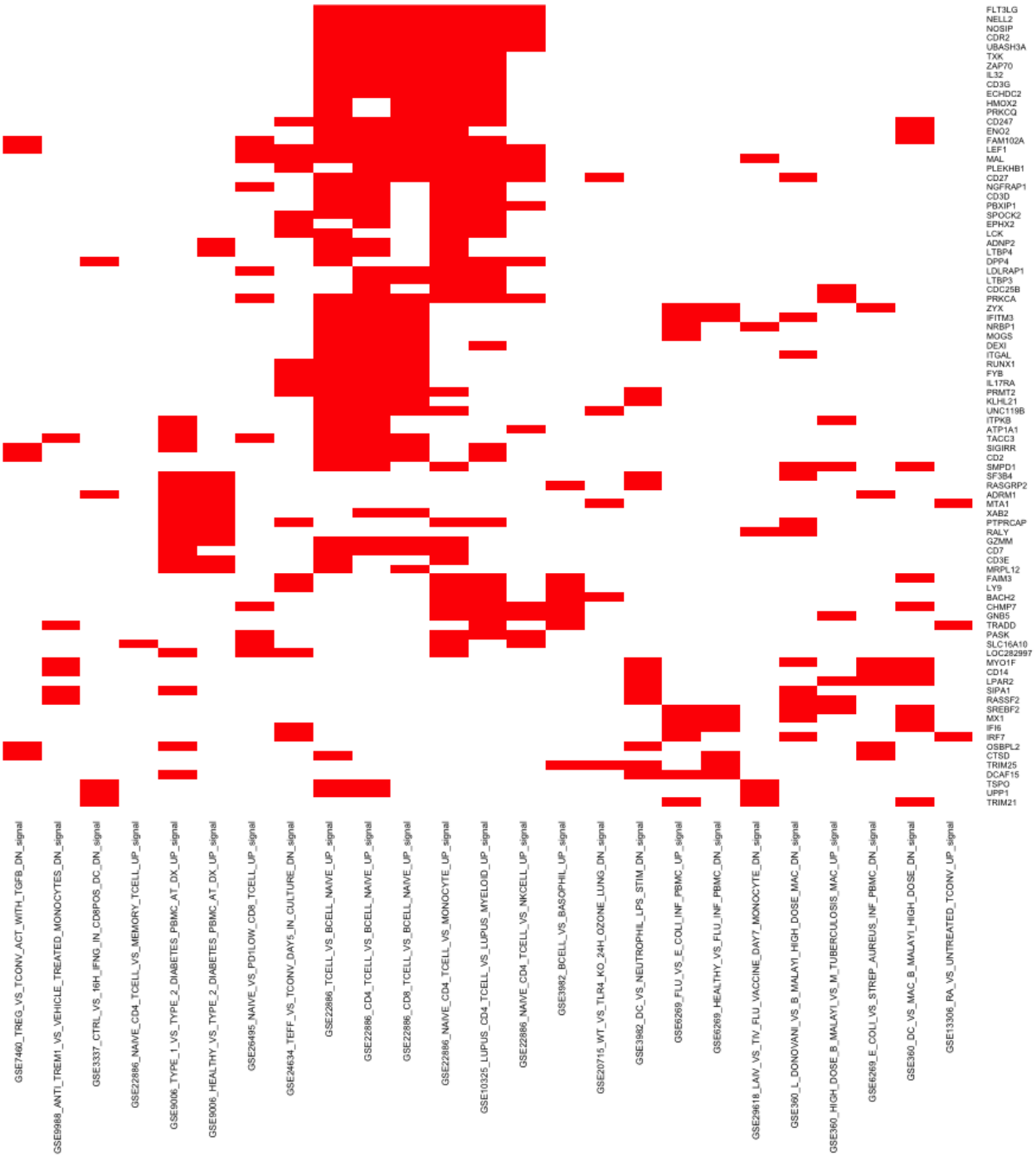


Figure 5.11: Genes contributing to the leading edge of four or more of the top 25 upregulated gene sets 21 days after immunisation with trivalent inactivated influenza vaccine. Red squares are used to denote relationship between leading edge gene and gene set.

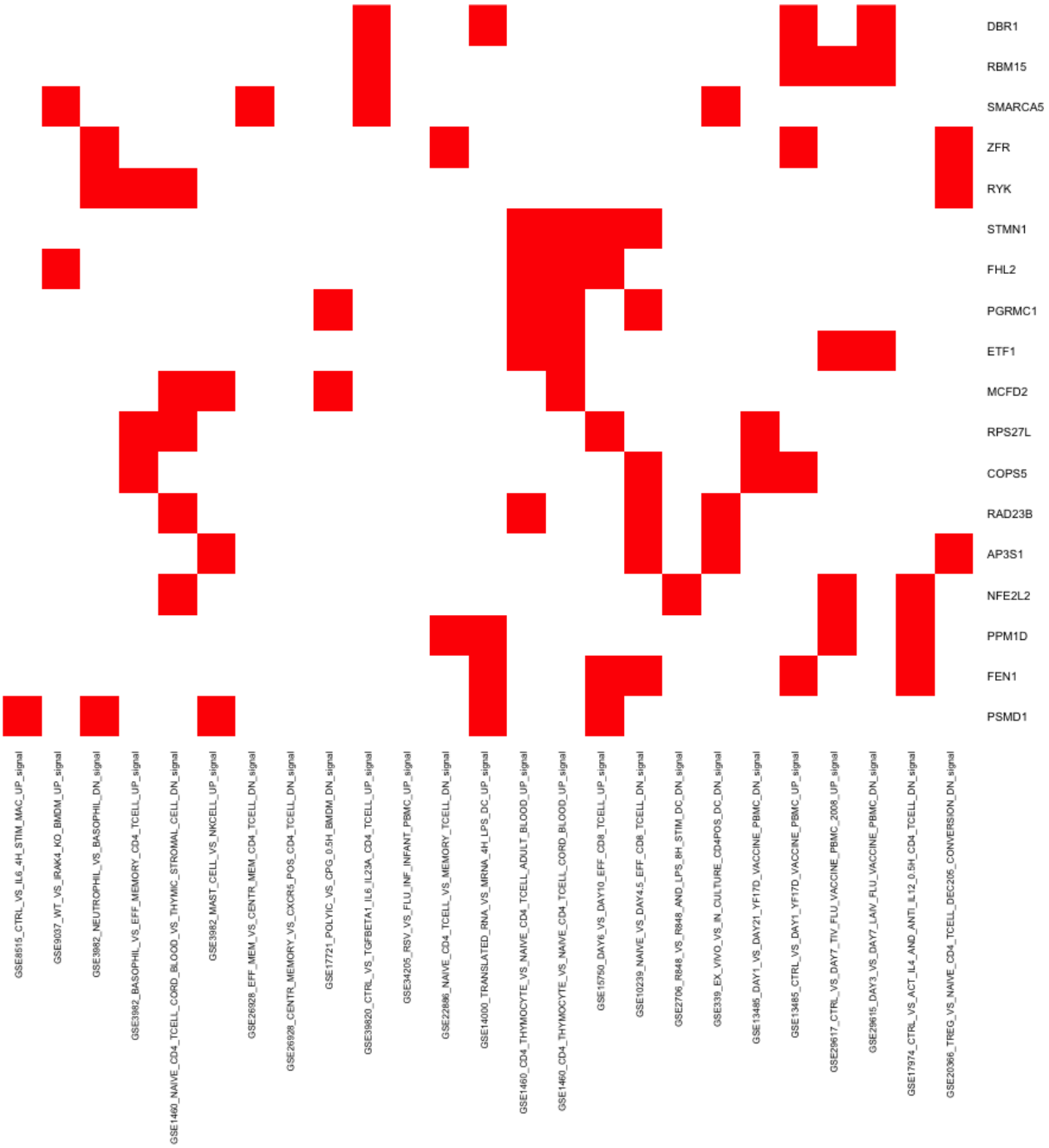


Figure 5.12: Genes contributing to the leading edge of four or more the top 25 downregulated gene sets 21 days after immunisation with trivalent inactivated influenza vaccine. Red squares are used to denote relationship between leading edge gene and gene set.

5.2.5.6 Ingenuity Pathway Analysis

The results of the linear regression analysis were filtered for probes with p-value <0.015, a threshold selected to result in a suitable number of genes for the IPA (n=286). This probe list was loaded into the IPA (Ingenuity® Systems, California) software package, mapped to the IPA Knowledge Base® and used to form de novo networks and assess for overrepresentation of members of canonical pathways. Networks were formed from the uploaded 'seed' genes using interconnected molecules assembled by the IPA network algorithm. The most highly connected molecules were consolidated into networks, the details of which are found in Table 5.5. Several of these networks contained overlapping genes and the relatedness of these networks is shown in appendix Figure 7.23. The top three networks are related to cellular development and proliferation, haematological development and function, and the humoral immune response.

Table 5.5: Top networks identified using Ingenuity® Pathway Analysis on a list of probes with p-values <0.015 for differential expression 21 days after immunisation with trivalent inactivated influenza vaccine.

Rank	Score	Molecules	Description
1	38	23	Cellular Development, Cellular Growth and Proliferation, Cell Cycle
2	36	22	Haematological System Development and Function, Haematopoiesis, Tissue Morphology
3	32	20	Humoral Immune Response, Protein Synthesis, Cellular Function and Maintenance
4	31	20	Cancer, Haematological Disease, Organ Development
5	27	18	Cell-To-Cell Signalling and Interaction, Haematological System Development and Function, Haematopoiesis
6	27	18	Lipid Metabolism, Small Molecule Biochemistry, Vitamin and Mineral Metabolism
7	25	17	Dermatological Diseases and Conditions, Endocrine System Disorders, Haematological Disease
8	25	17	Cell Cycle, Visual System Development and Function, Cell Death and Survival
9	24	17	Carbohydrate Metabolism, Lipid Metabolism, Small Molecule Biochemistry
10	23	16	Hereditary Disorder, Neurological Disease, Psychological Disorders
11	23	16	Connective Tissue Disorders, Developmental Disorder, Gastrointestinal Disease
12	21	15	Metabolic Disease, Neurological Disease, Organismal Injury and Abnormalities
13	21	15	Cellular Development, Cellular Growth and Proliferation, Haematological System Development and Function
14	20	15	Nervous System Development and Function, Organ Morphology, Developmental Disorder
15	19	16	Embryonic Development, Organ Development, Organismal Development
16	14	11	Cell Morphology, Cell-mediated Immune Response, Cellular Movement
17	9	8	Cellular Movement, Immune Cell Trafficking, Haematological System Development and Function

The top three networks were merged and the IPA Molecule Activity Predictor (MAP) tool was used to predict potential downstream effects using the changes in expression seen in the 'seed' genes (Figure 5.13). This merged network contained a number of immunological genes and the IPA MAP tool predicted effects on various downstream immunological complexes including the CD8 receptor, B cell receptor (BCR) complex, IgM, IgA, IgG3 and IgE (Figure 5.13).

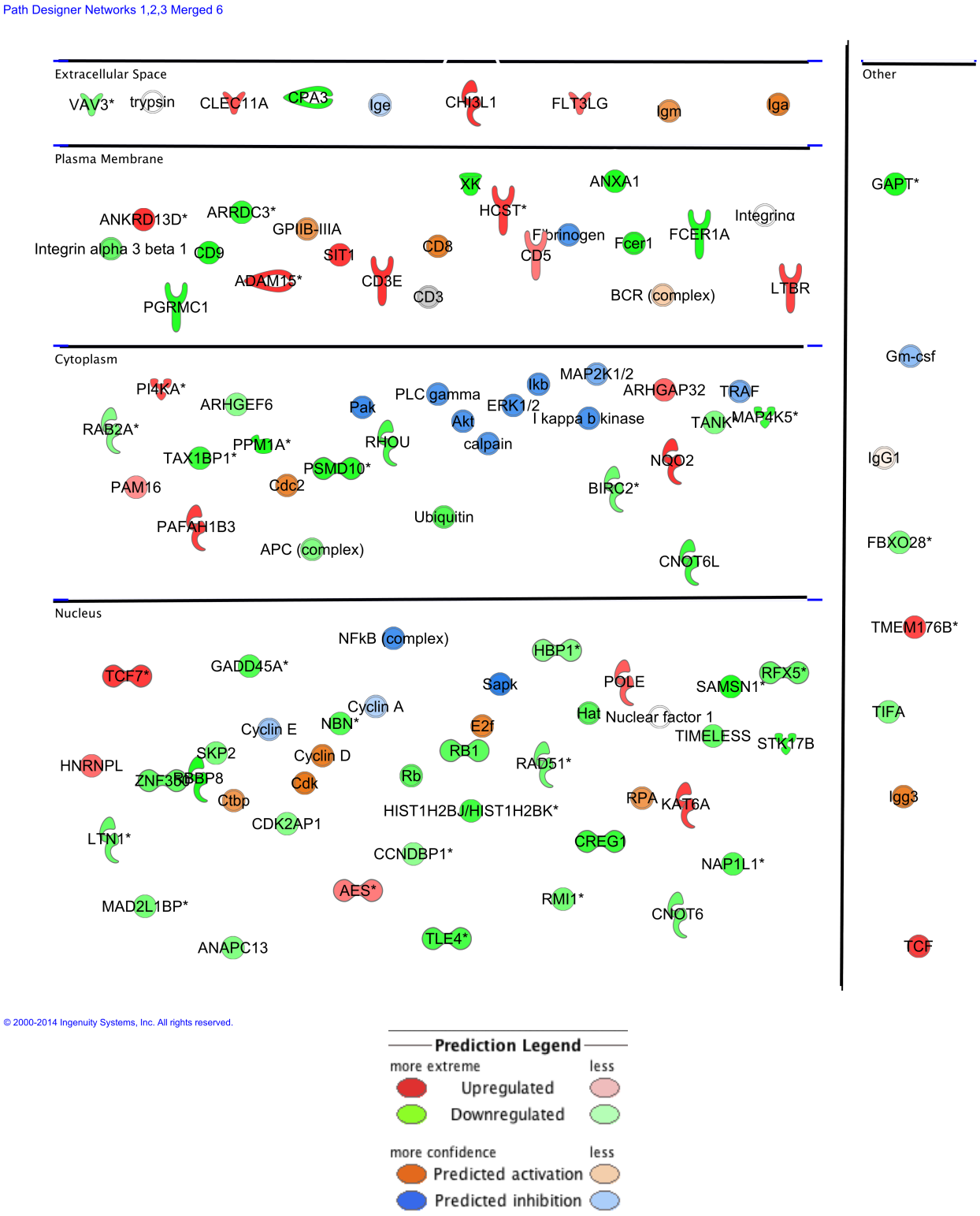


Figure 5.13: Composite network created by merging the top three networks following Ingenuity® Pathway Analysis (IPA) of peripheral blood gene expression profiles 21 days after immunisation with trivalent inactivated influenza vaccine. The Molecule Activity Predictor tool of the IPA package was used to predict downstream effects. The relationships between components have been removed for ease of viewing.

Fisher's exact test was used to assess the relationship between DEGs (p-value <0.015) following TIV and gene pathways annotated in the Ingenuity® Knowledge Base (Table 5.12). Generally the theme of functional annotations associated with these data relate to the quantity of haematopoietic progenitors, as well as leukocytes and lymphocytes trafficking and apoptosis (Table 5.6).

Table 5.6: The 25 Ingenuity® Knowledge Base functional pathways with the greatest enrichment scores in the list of genes differentially expressed following immunisation with trivalent inactivate influenza vaccine at p <0.015.

Category	Disease or functional annotation	p-value	z-score	Molecules
Cancer	precursor T-cell lymphoblastic leukemia-lymphoma	0.00004		7
Haematological Disease	precursor T-cell lymphoblastic leukemia-lymphoma	0.00004		7
Immunological Disease	precursor T-cell lymphoblastic leukemia-lymphoma	0.00004		7
Haematological Development and Function	quantity of haematopoietic progenitor cells	0.00005	0.731	20
Haematopoiesis	quantity of haematopoietic progenitor cells	0.00005	0.731	20
Tissue Morphology	quantity of haematopoietic progenitor cells	0.00005	0.731	20
Cell Death and Survival	apoptosis of blood cells	0.00006	0.287	21
Cancer	tumourigenesis of thyroid gland tumour	0.00010		3
Endocrine System Disorders	tumourigenesis of thyroid gland tumour	0.00010		3
Cell Death and Survival	apoptosis of fibroblast cell lines	0.00011	-0.893	16
Cell Morphology	cell spreading of lymphocytes	0.00018	-0.152	4
Cell Death and Survival	apoptosis of leukocytes	0.00018	0.753	19
Cell Death and Survival	apoptosis of mononuclear leukocytes	0.00018	0.792	16
Cell Death and Survival	cell death of blood cells	0.00020	0.311	25
Haematological Development and Function	quantity of mononuclear leukocytes	0.00023	1.073	26
Tissue Morphology	quantity of mononuclear leukocytes	0.00023	1.073	26
Cell Death and Survival	cell death of mononuclear leukocytes	0.00027	0.398	17
Cellular Assembly and Organisation	fusion of vesicles	0.00029		6
Haematological Development and Function	quantity of lymphocytes	0.00029	0.930	25
Tissue Morphology	quantity of lymphocytes	0.00029	0.930	25
Cancer	incidence of adenoma	0.00031	1.741	6
Cancer	tumourigenesis of liver tumour	0.00031	0.816	6
Gastrointestinal Disease	tumourigenesis of liver tumour	0.00031	0.816	6
Hepatic System Disease	tumourigenesis of liver tumour	0.00031	0.816	6
Cancer	arrest in G1 phase of tumour cells	0.00034		3

5.2.5.7 Relationship between differential gene expression following trivalent inactivated influenza vaccine and previous pandemic influenza vaccine

A linear model was used to assess the relationship between the fold-change in gene expression following TIV and the pandemic influenza vaccine received the previous year. There were no statistically significant relationships between differential gene expression following TIV and which of the pandemic vaccines was administered the previous year, when corrected for multiple testing (Table 5.7).

The IPA networks analysis generated using genes with p-values less than <0.015 in the linear regression model between gene fold-change following TIV and the pandemic influenza vaccine received are shown in appendix Table 7.38, none the IPA canonical pathways showed statistically significant enrichment.

Table 5.7: The 25 transcripts with the most statistical evidence of a relationship between fold-change following immunisation with trivalent inactivated influenza vaccine and which of the pandemic vaccines was administered the previous year (adjuvanted versus non-adjuvanted).

Gene	t-value	p-value
OPN3	3.84	0.000507
TNFRSF12A	-3.78	0.000593
IFT88	3.64	0.000894
HS.583888	3.49	0.001358
FLJ45032	-3.28	0.002377
ZNF839	-3.23	0.002731
RAD50	-3.18	0.003071
N6AMT1	3.12	0.003683
HIST3H2A	3.10	0.003808
DSCAM	-3.10	0.003882
LOC729793	3.09	0.003969
SCNN1D	-3.08	0.004055
MANEA	3.07	0.004129
UBTD2	3.02	0.004733
ANTXR2	-3.01	0.004925
HS.572538	2.99	0.005076
TUFT1	-2.97	0.005359
TP53BP1	-2.92	0.006119
LRRC57	-2.91	0.006302
NFATC3	-2.89	0.006634
LOC649169	2.89	0.006665
PHF20	-2.87	0.006924
LOC653344	2.87	0.006935
HS.156773	2.87	0.006971
HLA-C	-2.85	0.007260

5.2.5.8 Relationship between differential gene expression following trivalent inactivated influenza vaccine and immunological responses to TIV

A linear model was used to assess for a relationship between fold-changes in gene expression following TIV and post-vaccination HAI titres. The 25 genes with the most statistical evidence for a relationship with post-vaccination HAI titres are shown in Table 5.8. IPA networks were generated using genes with fold-changes that associated (p <0.015) with post-vaccination HAI titres are shown in appendix Table 7.31 and the enriched canonical pathways are shown in Figure 5.14. The most statistically associated IPA canonical pathway was the antigen presentation

pathway; within this pathway the downregulation of *HLA-DRB4*, *HLA-DQB1*, *HLA-DRB3*, *HLA-DRB1*, *HLA-F* and *PSMB9* were associated with greater post-vaccination HAI titres.

Table 5.8: The 25 genes with the most statistical evidence of a relationship between fold-change 21 days following immunisation with trivalent inactivated influenza vaccine and haemagglutination inhibition assay titres.

Gene	t-value	p-value	Description
HSP90AA1	3.89	0.000464	heat shock protein 90kDa alpha (cytosolic), class A member 1
SCAND1	-3.70	0.000794	similar to basic-leucine zipper transcription factor MafG
C9ORF78	3.44	0.001597	chromosome 9 open reading frame 78
SLC35B4	3.44	0.001598	solute carrier family 35 (UDP-xylose/UDP-N-acetylglucosamine transporter), member B4
C9ORF85	3.40	0.001799	chromosome 9 open reading frame 85
HLA-DRB3	-3.40	0.001800	major histocompatibility complex, class II, DR beta 3
CETN3	3.39	0.001837	centrin, EF-hand protein, 3
TAX1BP1	3.36	0.001990	Tax1 (human T-cell leukemia virus type I) binding protein 1
TNFSF15	-3.32	0.002260	tumor necrosis factor (ligand) superfamily, member 15
ACSF2	-3.30	0.002379	acyl-CoA synthetase family member 2
TDRD1	-3.28	0.002473	tudor domain containing 1
LOC644132	-3.26	0.002628	similar to basic-leucine zipper transcription factor MafG
RGS18	3.25	0.002701	regulator of G-protein signaling 18
C9ORF89	-3.22	0.002889	chromosome 9 open reading frame 89
HS.576633	-3.21	0.002999	
GIYD1	-3.20	0.003064	SLX1 structure-specific endonuclease subunit homolog A
ZNF577	-3.17	0.003334	zinc finger protein 577
LOC100131644	-3.15	0.003490	
PPM1K	-3.15	0.003492	protein phosphatase, Mg2+/Mn2+ dependent, 1K
TAX1BP1	3.14	0.003581	Tax1 (human T-cell leukemia virus type I) binding protein 1
USP16	3.14	0.003581	ubiquitin specific peptidase 16
LOC653079	3.12	0.003792	RPL21P98 ribosomal protein L21 pseudogene 98
DHRS9	-3.11	0.003834	dehydrogenase/reductase (SDR family) member 9
DLEU7	3.11	0.003855	deleted in lymphocytic leukemia, 7
HLA-DRB1	-3.09	0.004120	major histocompatibility complex, class II, DR beta 1

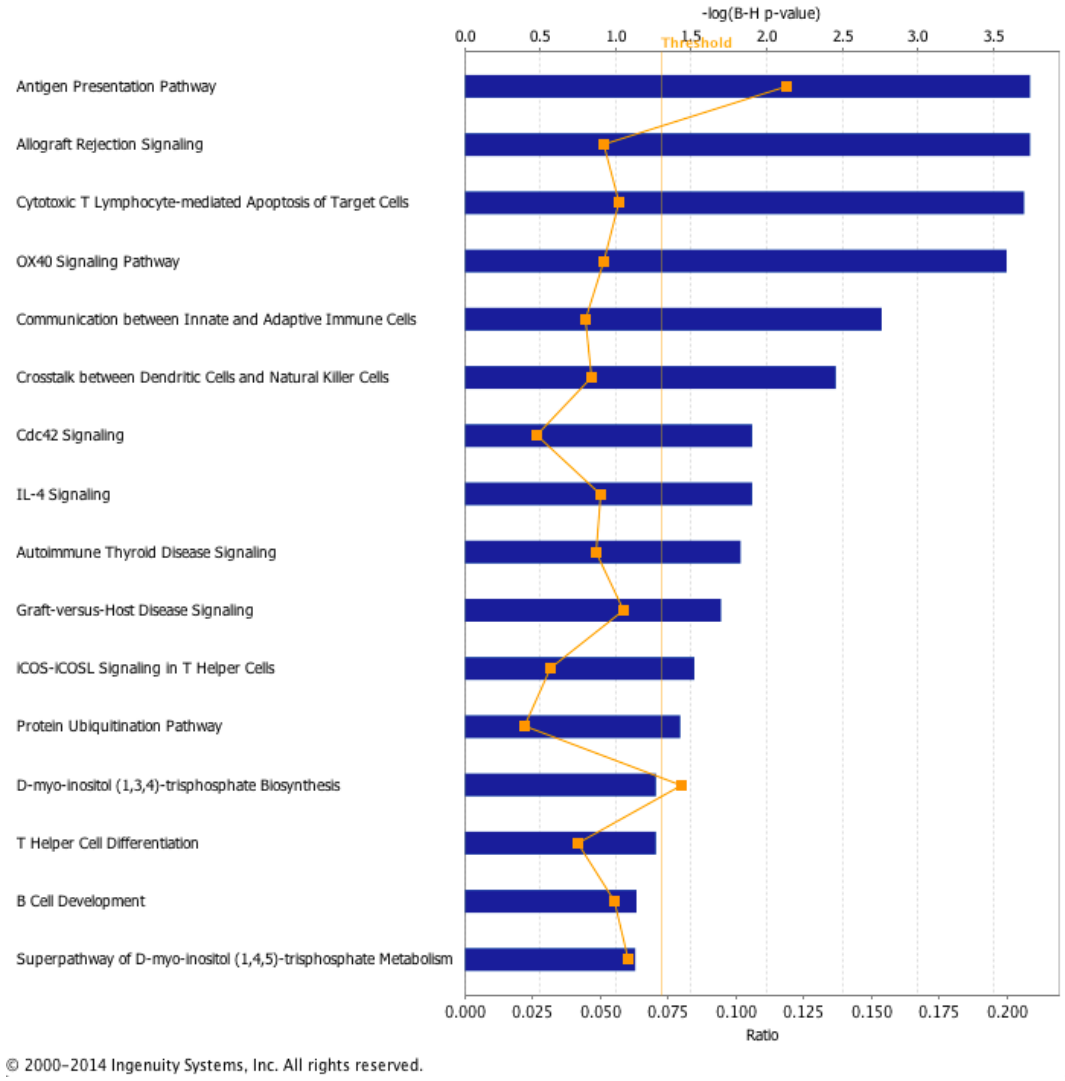


Figure 5.14: The most enriched canonical pathways following Ingenuity Pathway Analysis (IPA) of genes fold-changes 21 days following trivalent inactivated influenza vaccine associated ($p<0.015$) with haemagglutinin assay inhibition titres following immunisation. The threshold displayed is based upon a false-discovery rate of <0.05. B-H p-value = Benjamini-Hochberg p-value, ratio= ratio IPA entry list genes in canonical pathway compared to the full list of genes annotated within the canonical pathway.

A linear model was used to assess for a relationship between changes in gene expression following TIV and post-vaccination T cell responses. The 25 genes with the most statistical evidence of a relationship with T cell responses to each of the three influenza protein assessed are shown in appendix Tables 7.32, 7.34 and 7.36. IPA networks generated using genes with fold-changes that associated ($p<0.015$) with post-TIV T cell responses are shown in appendix Tables 7.33, 7.35 and 7.37.

5.2.6 Discussion

This section assessed whole blood RNA profiles pre- and post-vaccination with TIV, in a cohort of healthy children who received either the AS03B adjuvanted or non-adjuvanted 2009 pandemic influenza (H1N1) vaccine the previous year. Individual gene based analysis displayed little evidence of DEGs 21 days after TIV, with only one gene surpassing the level of statistical significance. On the other hand, gene set and pathway based analysis implied considerable differences between pre- and post-vaccination gene regulation. No statistically significant relationships were demonstrated between gene profiles 21 days after TIV and which of the pandemic influenza vaccines were administered the previous year.

There were two particular limitations to this study that may explain the paucity of statistically significant DEGs observed in the single-gene analysis. Firstly, while T cell data from these children showed significant increases in influenza-specific T cell responses 21 days following vaccination and previous studies have shown increases in influenza-specific memory B cells 3 weeks post-TIV, the transcriptional changes associated with these adaptive immune responses may be relatively subtle in unstimulated whole blood samples (375, 390). Secondly, transcriptional changes in whole blood samples may be masked by the noise associated with this complex fluid. While assessing whole blood may be beneficial to minimise sample handling artefact caused by processing steps such as gradient or magnetic separation, amplified RNA extracted from whole blood contains a high proportion of globin messenger RNA that can reduce the efficiency of microarrays by interfering with the detection of less abundant transcripts (391, 392). For example *TNFRSF17*, the expression of which in PBMC has previously been shown to correlate with TIV and YF-17D responses, was not detectable in whole blood in this study (84, 85). There are some data suggesting that reduction of globin mRNA levels prior to microarray hybridisation improves the sensitivity of microarray expression profiling assays using whole blood samples (393).

To date, the majority of vaccine transcriptomic studies have focused on early transcriptional profiles associated with innate immunity, and few have assessed later onset transcriptional responses coinciding with adaptive immunity (84, 85, 87, 88, 90, 91, 97, 99, 104, 394). However,

a study evaluating peripheral whole blood 1, 3 and 14 days after TIV found the greatest number of DEGs at day 14 (88). Furthermore, another study described a differentially regulated gene module in whole blood 21 days after TIV (87).

In the study detailed in this section, *GAPT* was the only DEG (FDR < 0.05) in single-probe analysis. This gene encodes GRB2-binding adaptor protein transmembrane (GAPT), which is expressed by B cells and myeloid cells (395). This adapter protein binds growth factor receptor-bound protein 2 (GRB2), which links surface growth factor receptors and the RAS signalling pathway (396). GRB2 has been shown to be involved in a number of biological processes including T and B cell receptor signalling (395, 396, 397). GAPT is not phosphorylated upon BCR ligation, but constitutively binds GRB2 via a proline-rich region on its cytoplasmic tail (395). In mice, deletion of the *GAPT* does not affect RAS/MAPK activation upon BCR stimulation; however, B cells from GAPT$^{-/-}$ mice show hyperresponsiveness following BCR stimulation (395). GAPT is believed to negatively regulate B cell proliferation following stimulation via the BCR and might also have a role in the maintenance of marginal zone B cells (395). Interestingly, *GAPT* has also been shown to be differentially expressed in peripheral whole blood following adult immunisation with TIV (88).

The results of the individual probe based analysis showed little evidence of DEGs, suggesting differences between baseline and day 21 post-vaccination may be modest relative to experimental noise. GSEA is an alternative to individual gene based approaches, looking for non-random relationships between gene set membership, determined by common biological function, and the ranked statistical distribution (385). This approach may be advantageous if changes at the single-gene level are subtle but many genes within a gene set show changes in expression. Unlike single-gene based analysis, GSEA suggested considerable transcriptional changes in whole blood 21 days after vaccination with TIV. GSEA found more downregulated than upregulated gene sets following vaccination, 148 and 12 respectively. Leading-edge analysis identified a core set of upregulated genes associated with T cells compared to myeloid cells, NK cells and naïve B cells. Furthermore, leading-edge analysis found a core set of downregulated genes that have previously been shown to be downregulated in PBMCs taken from human volunteers within a

week of immunisation with TIV, LAIV or YF-17D (84, 85).

Pathway analysis is another way of incorporating existing biological knowledge of gene-gene interactions and molecular pathways to increase power of transcriptomic studies. The de novo pathways formed using the IPA MAP tool suggested the gene expression changes induced by TIV could have downstream effects on the CD8 receptor, BCR complex, IgM, IgA, IgG3 and IgE formation. The list of genes with most evidence of differential expression were enriched for genes with functional annotations related to the quantity of haematopoietic progenitors, leukocytes and lymphocytes, as well as their trafficking and apoptosis.

There were no genes with fold-changes that had statistically significant associations with HAI titres post-vaccination with TIV. However, IPA suggested a relationship between fold-changes in genes involved in several immunological canonical pathways including antigen presentation pathway, communication between innate and adaptive immune cells and OX40 signalling pathway, with post-TIV HAI titres. The most significant IPA canonical pathway highlighted was the antigen presentation pathway, with the downregulation of *HLA-DRB4*, *HLA-DQB1*, *HLA-DRB3*, *HLA-DRB1*, *HLA-F* and *PSMB9* 21 days after TIV being associated with greater HAI titres. Interestingly, early upregulation of the antigen presentation pathway has previously been associated with greater HAI titres following TIV (88).

5.2.7 Conclusion

While this section only demonstrated one gene that was statistically differentially expressed 21 days after immunisation with TIV, gene set and pathway based analysis alluded to more substantial transcriptional changes at this time point. These data suggest differential gene expression in whole blood 21 days after childhood immunisation with TIV is subtle at the individual gene level, compared to previous studies assessing earlier time points in adults. Research with greater power to detect changes at later time points associated with adaptive immune responses to vaccines may require larger numbers of participants, recall stimulus or the characterisation of enriched cell subsets.

5.3 Assessment of serum microRNA following childhood immunisation with pandemic influenza vaccine

5.3.1 Background

MicroRNAs (miRNAs) are a family endogenous 21-25 nucleotide non-coding RNAs that bind the 3′ untranslated region (UTR) of mRNA and have been shown to regulate gene expression in a sequence specific manner (398, 399). The current model of miRNAs synthesis includes two processing events: 1) nascent pri-miRNA transcripts are processed into ∼70-nucleotide pre-miRNA by Drosha and DiGeorge syndrome critical region 8 (DGCR8) within the nucleus, 2) pre-miRNA is transported to the cytoplasm and cleaved by Dicer to generate ∼21-25 nucleotide miRNA:miRNA* duplexes, one of which is preferentially assembled (denoted by *) into the RNA-induced silencing complex (RISC). MiRNAs functions by guiding RISC to the 3′ UTR of the target mRNAs, decreasing mRNA stability and/or inhibiting translation (400, 401).

Hundreds of highly conserved miRNAs have been identified in the human genome, each of which may suppress hundreds of target mRNAs such that most mRNAs are targets of miRNAs (401, 402, 403). The expression of miRNA is highly regulated at three levels: transcription, processing and subcellular localisation (401). There are data showing these regulatory processes are perturbed by immunological challenge and inflammation (401, 402). For example, the transcription of several miRNAs expressed by immune cells, such as miR-155 and miR-146a, have been shown to be differentially expressed following stimulation with TLR ligands or pro-inflammatory cytokines (402, 404, 405). The expression of miRNAs is dynamic, and important in innate and adaptive immunity (401). An interesting feature of miRNA biology is their stability, such that they can be isolated from serum and other bodily fluids, making miRNAs particularly appealing candidates in clinical biomarker discovery (406).

There are data suggesting three origins of circulating miRNAs: 1) passive leakage due to cellular injury or cell turnover, 2) active secretion via microvesicles or 3) active microvesicle-free secretion associated with RNA-binding proteins (407). Furthermore, there are some data suggesting circulating miRNAs may function as intercellular communicators (408, 409, 410).

Circulating miRNAs have been shown to be altered by a number of pathological conditions, but serum miRNA profiles in immunological responses to antigenic stimuli have not been thoroughly explored in humans (407, 411). This section will assess levels of miRNAs in sera from children prior to and following a two-dose regimen of the 2009 pandemic H1N1 influenza vaccine, using a high-throughput miRNA microarray.

5.3.2 Aim

The aim of this section was assess serum miRNA profiles following pandemic influenza vaccine, and to relate these to vaccine adjuvantation and immunogenicity.

5.3.3 Participants

This study included healthy children (6 months to 12 years) enrolled in a study to compare the safety and immunogenicity of an AS03B adjuvanted split virion versus a non-adjuvanted whole virion pandemic influenza (H1N1) vaccine (282). Sera were separated from peripheral blood samples pre-vaccination and 21 days following the second of a two-dose schedule of pandemic influenza vaccine (Figure 5.15). A subset of 22 children were selected for serum miRNA expression analysis (subsection '5.3.7.1 Power calculations').

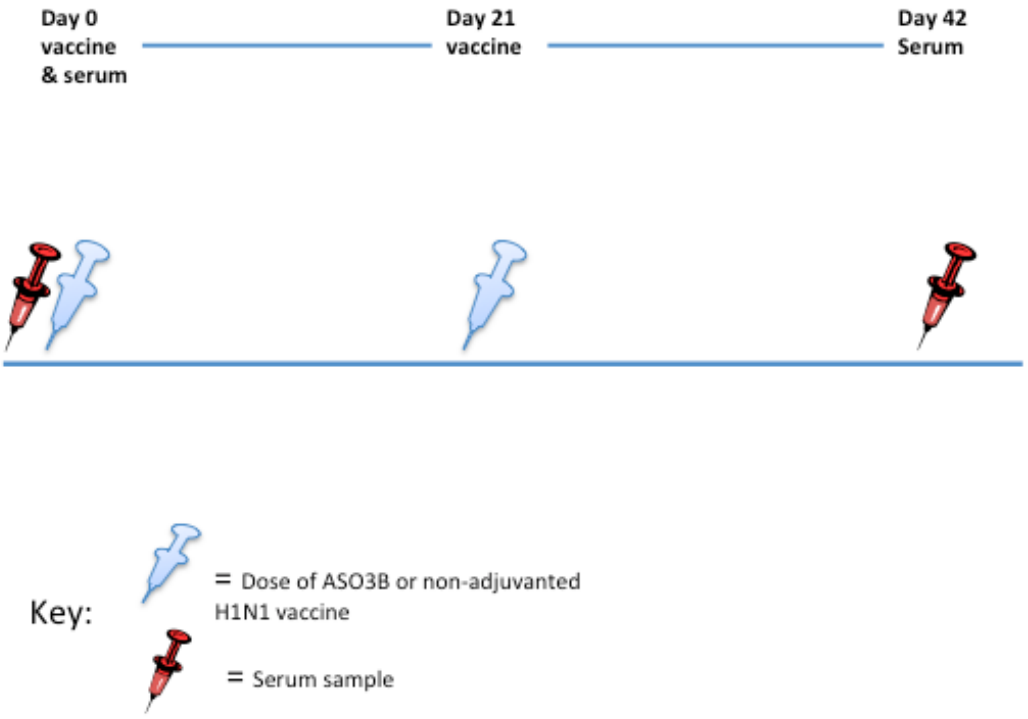


Figure 5.15: Schematic of study profiling serum microRNA following immunisation with H1N1 pandemic influenza vaccine. Serum was collected before and 21 days after a two-dose regiment of, either adjuvanted (ASO3B) split virion vaccine or non-adjuvanted whole virion, 2009 H1N1 pandemic influenza vaccine.

5.3.4 Vaccine

Participants were randomised to receive two doses of either 0.25 ml AS03B adjuvanted split virion vaccine Pandemrix™, containing 1.875 μg of haemagglutinin antigen from H1N1 influenza A/California/07/2009 (GlaxoSmithKline Vaccines, Rixensart, Belgium) or 0.5 ml non-adjuvanted whole virion vaccine Celvapan™, containing 7.5 μg of haemagglutinin from H1N1 influenza A/California/07/2009 (Baxter Vaccines, Vienna)

5.3.5 Immunological measures

Haemagglutination inhibition assays were performed at the Centre for Infections Virus Reference Department, Public Health England (London, UK), according to standardised methods,

as described in subsection '5.2.3.3 Haemagglutination inhibition assay'.

5.3.6 MicroRNA assessment

5.3.6.1 Serum microRNA isolation

RNAs were extracted from serum using a total RNA extraction kit (Norgen Biotek, Thorold, Ontario, Canada; Cat: 17200). In brief, 600 μl lysis buffer (containing 1% β-mercaptoethanol) was added to 200 μl of serum. After vortexing, 800 μl 100% (molecular-grade) ethanol was added and samples were vortexed. The samples were then processed as per Norgen Biotek protocol (`https://norgenbiotek.com/product_resources/total_rna_purification_kit_total_rna_purification_kit__protocol_17200_1503.pdf`) with centrifugation at 14,000$\times g_0$, unless otherwise stated.

5.3.6.2 Microarray analysis

The concentration of the RNA was measured using a NanoDrop® ND-8000 UV-Vis spectrophotometer (Thermo Scientific, Massachusetts, USA). For each sample 100 ng total RNA (at 50 ng/μl) was taken and used in the miRNA microarray labelling procedure (Agilent Technologies, Santa Clara, USA; Cat: 5190-0456). This procedure followed the Agilent miRNA protocol (Version 2.4 September 2011, `http://www.chem.agilent.com/library/usermanuals/public/g4170-90011_mirna_complete_2.4.pdf`); with the exception that after the labelling with Cy3-pCp molecule samples were dried and 'Step 2' was omitted. Serum miRNA levels were assessed using the Human miRNA Microarray 60k (Agilent® Technologies). Samples were hybridised for 40 hours to increase specific signal against background. MiRNA extraction and microarray hybridisation was carried out by Oxford Gene Technology (Oxford, UK).

5.3.6.3 MicroRNA microarray normalisation

MiRNA microarray data normalisation remains a challenge, unlike mRNA the assumption that the majority of transcripts will be constant may not be valid, as few miRNAs may be expressed in a particular sample and the proportion of these that are differentially expressed following

an intervention may be relatively large (412, 413, 414). RNAs such as U6 and SNORD have been used to normalise miRNA data; however, these RNAs are highly variable in sera (414). Acknowledging these caveats miRNA data were normalised using methods analogous to those used in mRNA microarray normalisation. Four different types of normalisation methods were assessed, as detailed in appendix subsection '7.7.2.1 MicroRNA microarray normalisation'.

5.3.6.4 Probe filtering

Sample probe profile raw data were extracted using Agilent Feature Extraction Software Version 10.7.3.1. Expression values were 'patched' such that undetected probes were set to half the minimum value detected for that probe. A filter was imposed to return probes that were detected (Agilent feature extraction metric) in at least 65% of all the samples assessed.

5.3.7 Statistical analyses

5.3.7.1 Power calculation

At the time this study was initiated there were few publicly available data on serum miRNA profiles in humans to base a formal power calculation. A sample size of 22 participants was selected following discussions with Oxford Gene Technology, who provided insight from their ongoing unpublished work. This sample size was felt to be suitable to enable the generation of important data on the variance of childhood serum miRNAs facilitating formal power calculations of future studies, as well as having reasonable power to address the study objectives.

5.3.7.2 Linear regression analysis

A linear model was fitted with normalised expression of each miRNA as the response variable, and each paired sample and vaccination status as predictor variables (Equation 5.7). The empirical Bayes method was used to generate moderated t-statistics, moderated F-statistic, and log-odds of differential expression using the 'limma' R Package (382).

$$lmFit(expression\,of\,miRNA \sim sample\,pair + vaccination\,status) \tag{5.7}$$

5.3.7.3 Ingenuity Pathway Analysis

The Ingenuity® Knowledge Base has a curated collection of data experimentally demonstrated and predicted miRNA targets. IPA (Ingenuity® Systems, California) was conducted to assess potential biological relevance of the serum miRNA results using the web-based IPA software package (`http://www.ingenuity.com`).

5.3.7.4 Comparison of differential miRNA expression following either AS03B adjuvanted or non-adjuvanted 2009 pandemic influenza vaccine

To compare differential serum miRNA levels following either AS03B adjuvanted or non-adjuvanted pandemic influenza vaccine a linear model was fitted with $\log_2$ fold-changes of each probe as the response variable and the pandemic influenza vaccine administered as a predictor variable (Equation 5.2). The empirical Bayes method was then used to generate moderated t-statistics using the 'eBayes' function in the 'limma' R Package (382).

$$lmFit\Big(log_2(miRNA\ fold\text{-}change) \sim pandemic\ vaccine\Big) \tag{5.8}$$

5.3.7.5 Relationship between differential miRNA expression and immunological responses to 2009 pandemic influenza vaccine

The relationship between differential serum miRNA expression following pandemic influenza vaccine and the level of immunological response was interrogated using linear regression analysis. A linear model was fitted with $\log_2$ fold-changes of each probe as the response variable and the baseline $\log_{10}$ transformed H1N1 HAI titres, post-vaccination $\log_{10}$ transformed H1N1 HAI titres and the pandemic influenza vaccine received, as predictor variables (Equation 5.9).

$$\begin{aligned} lmFit\Big(&log_2(miRNA\ fold\text{-}change) \sim log_{10}(post\ vaccination\ Haemagglutination\ Inhibition\ titres) \\ &+ log_{10}(baseline\ Haemagglutination\ Inhibition\ titres) + pandemic\ vaccine\Big) \end{aligned} \tag{5.9}$$

5.3.8 Results

5.3.8.1 Participant demographics

Of the 937 participants who received two-doses of pandemic influenza vaccine as part of the parent study, 274 were recruited by the Oxford recruitment site (Oxford Vaccine Group, UK) (282). Two hundred and thirty-one of these participants had available paired pre- and post-vaccination serum samples. Twenty-two participants were selected at random and their stored pre- and post-vaccination samples were assessed for serum miRNAs. Participant demographics are displayed in Table 5.9 and appendix Table 7.39.

Table 5.9: Participant demographics in the study of serum microRNA profiles following immunisation with pandemic influenza vaccine.

Demographic	Details	Value[#]
Ethnicity		
	Caucasian(%)	20(91)
	Mixed, Asian-Caucasian (%)	2(9)
Sex		
	Male (%)	10(45)
	Female (%)	12(55)
Pandemic influenza vaccine		
	Adjuvanted (%)	9(41)
	Non-adjuvanted (%)	13(59)
Age		
	Months (IQR)	68.3*(38.4.0-112.6)
Time since vaccination		
	Days (IQR)	25*(21-25.8)
Serum processing time		
	Minutes (IQR)	85*(125-154)

[#] = number of children unless specified,* = median, IQR= interquartile range

5.3.8.2 MicroRNA microarray quality control

QC protocols used to assess the miRNA microarray data were analogous to those described in subsection '5.2.5.2 Gene transcript microarray quality control'. Of the 1368 probes on the microarray 167 miRNAs were detected in at least 65% of the serum samples assessed and were included in statistical analysis, a list of these miRNA is displayed in appendix Table 7.40. Box and whisker plots and standard deviations versus ranked probe means plots were used to compare four normalisation methods, appendix subsection '7.7.2.1 MicroRNA Microarray normalisation'. Following review of these plots the VSN was selected as the normalisation method of choice. The box and whisker plot for VSN normalised data does not show any

particular outlier samples (Figure 5.16). The standard deviation plot shows no sign of standard deviation-mean dependence (Figure 5.17). Density plots displayed in appendix Figure 7.29 show that the samples follow a similar intensity distribution.

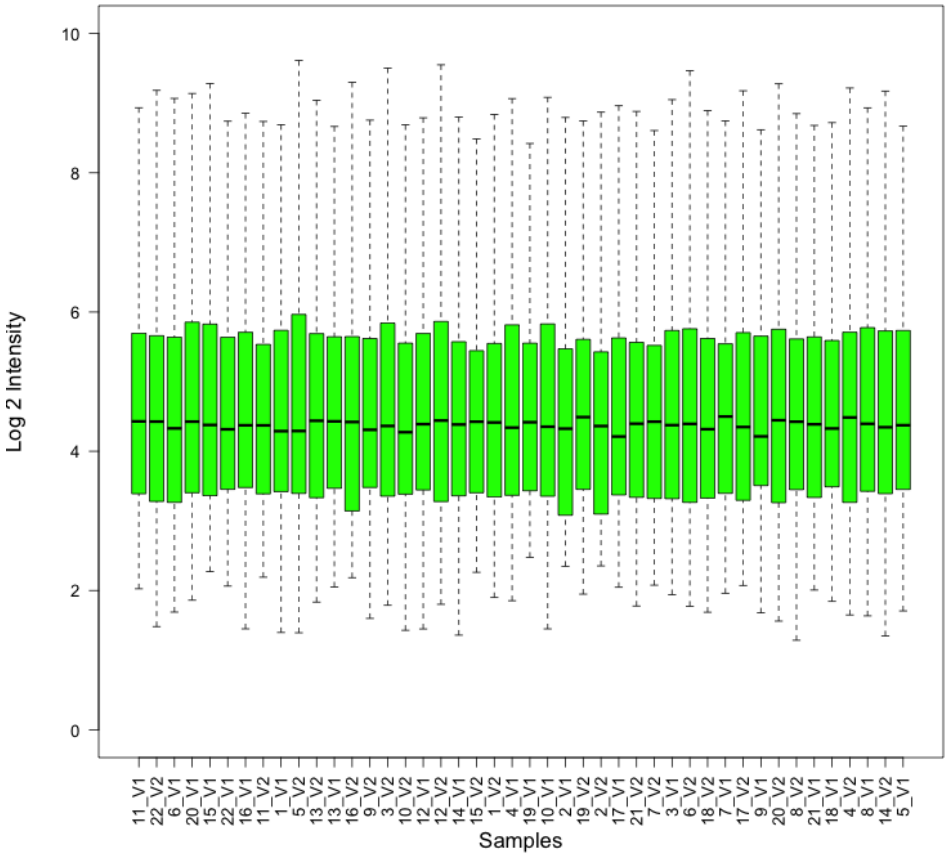


Figure 5.16: Box and whisker plot of variance stabilising normalised microRNA feature intensities, in the study of serum microRNA following immunisation with pandemic influenza vaccine. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. V1= baseline; V2= 3 weeks post-vaccination. (Note sample numbers have been anonymised $1 - 22$).

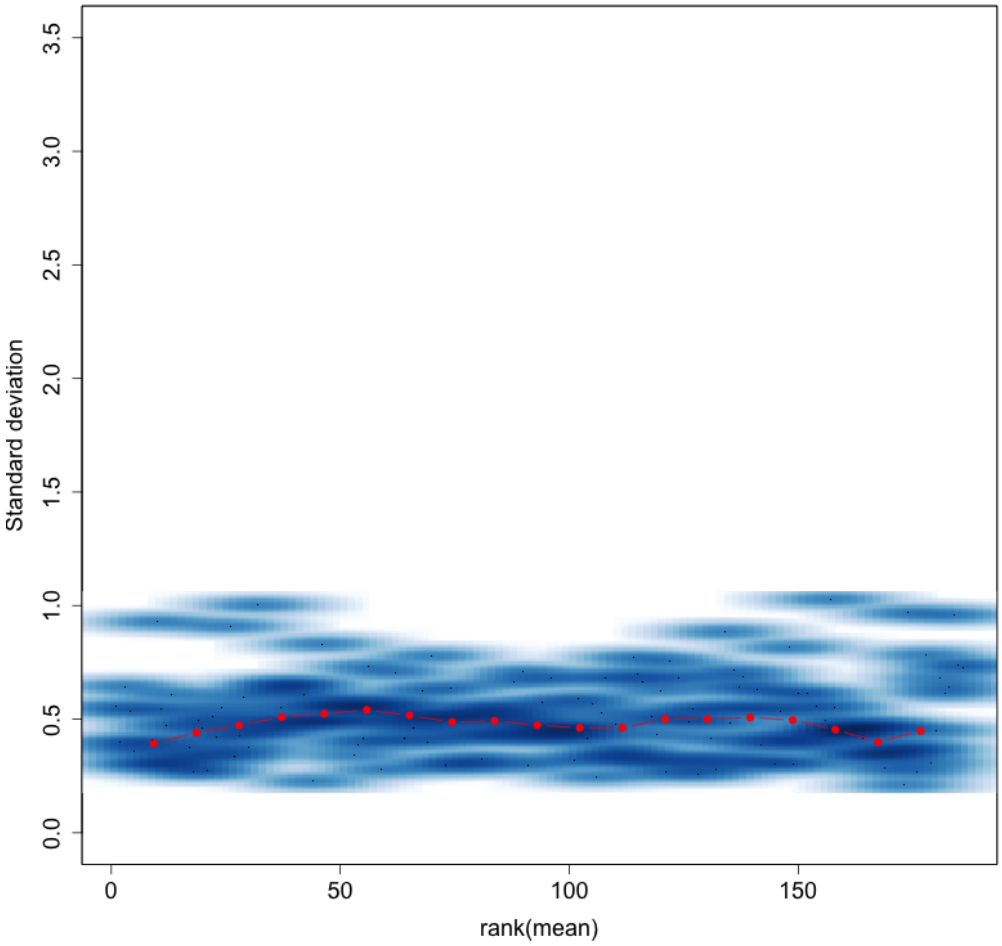


Figure 5.17: Standard deviations against the ranked mean probe intensities for variance stabilising normalised microRNA data, in the study of serum microRNA following immunisation with pandemic influenza vaccine. The red dots show the running median estimator, which is approximately horizontal as there is little evidence of standard deviation-mean dependence.

Hierarchical clustering shows that while participant samples do not cluster together, there is suggestion of clustering by vaccination status (Figure 5.18). PCA did not show any clear underlying data structure or sample outliers (Figure 5.19).

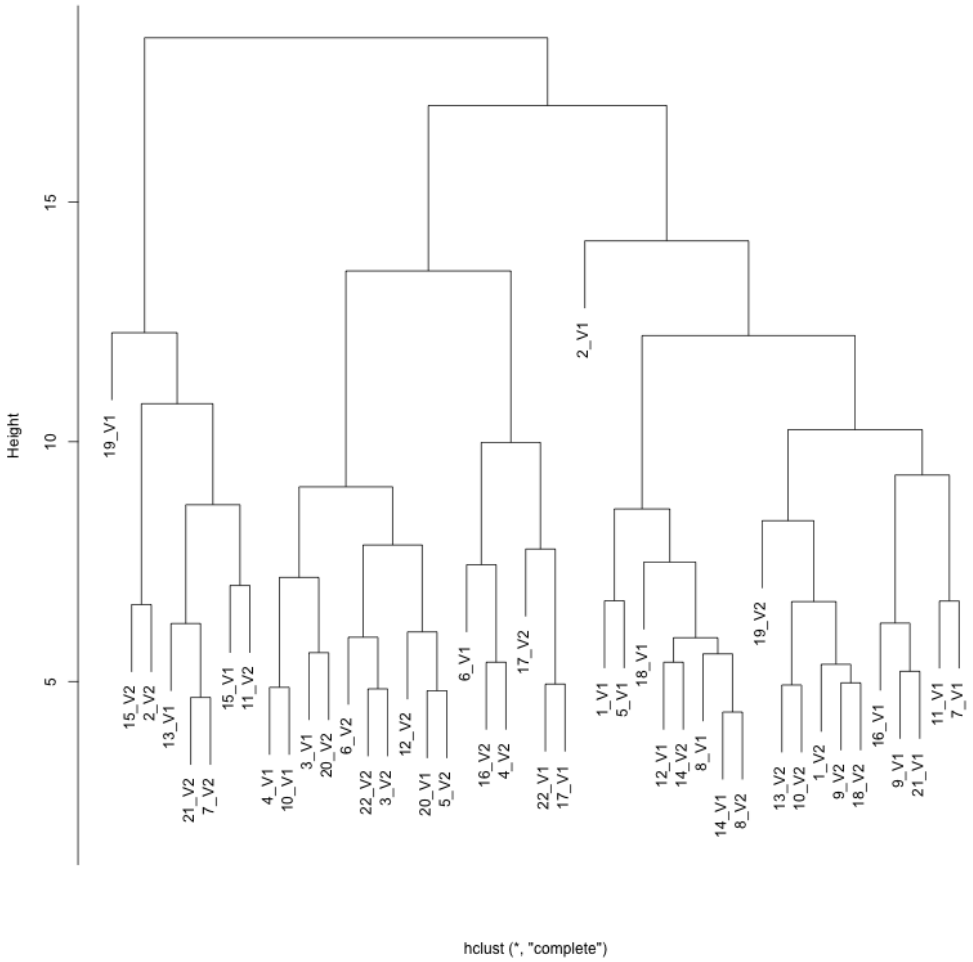


Figure 5.18: Hierarchical clustering of samples by variance stabilising normalised microRNA data in the study of serum microRNA following immunisation with pandemic influenza vaccine. V1= pre-vaccination; V2= 3 weeks post-vaccination. (Note sample numbers have been anonymised $1-22$)

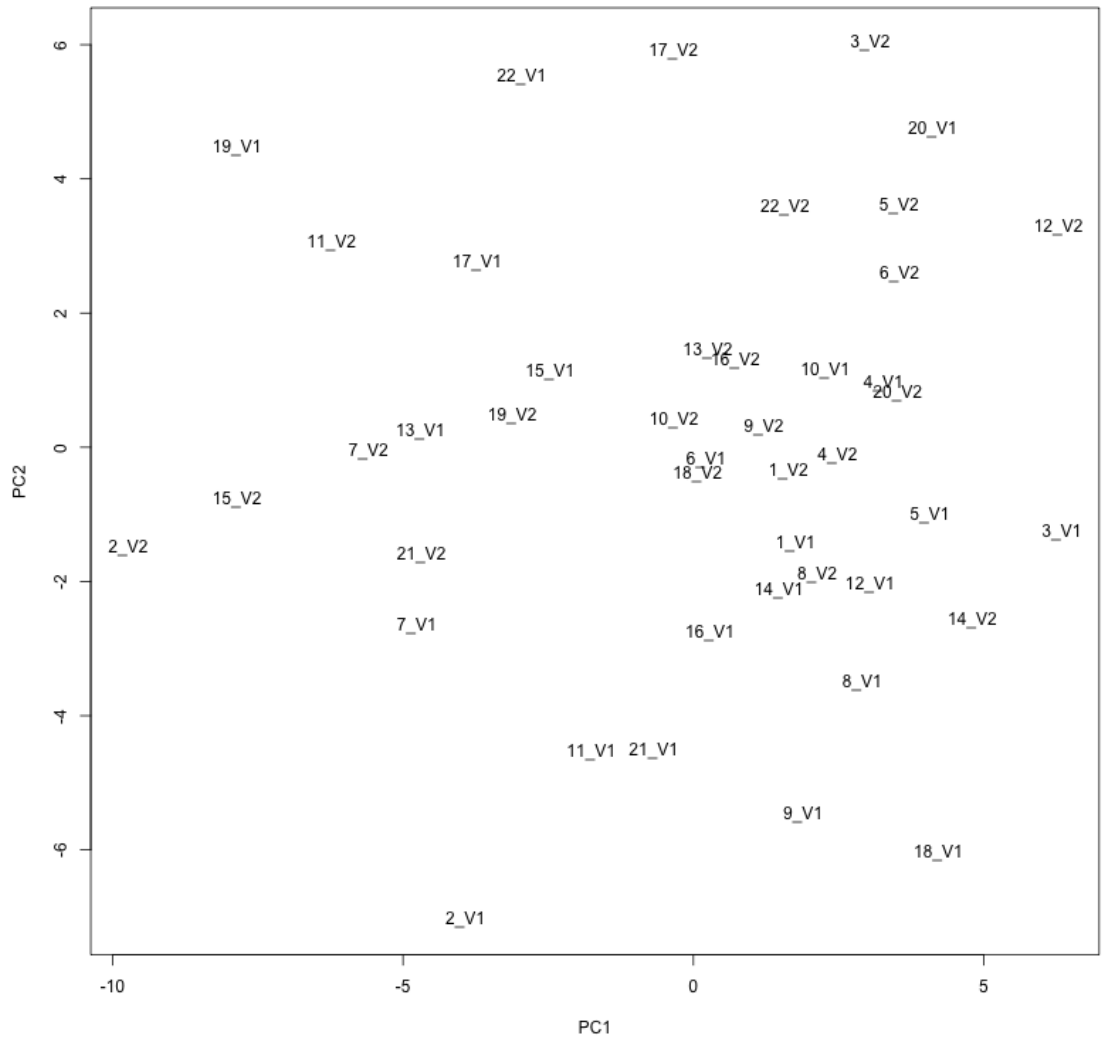


Figure 5.19: The first two principal components of the microRNA microarray data in the study of serum microRNA following pandemic influenza vaccine. V1= baseline; V2= 3 weeks post-vaccination. (Note sample numbers have been anonymised $1 - 22$).

A heat map representation of the matrix of the between sample correlations is shown in Figure 5.20. While the expression values of all samples are highly correlated, the expression values from the pre-vaccination sample from participant 19 and both the pre and post-vaccination samples from participant 2 were noticeably less correlated with the rest of the samples (Figure 5.20).

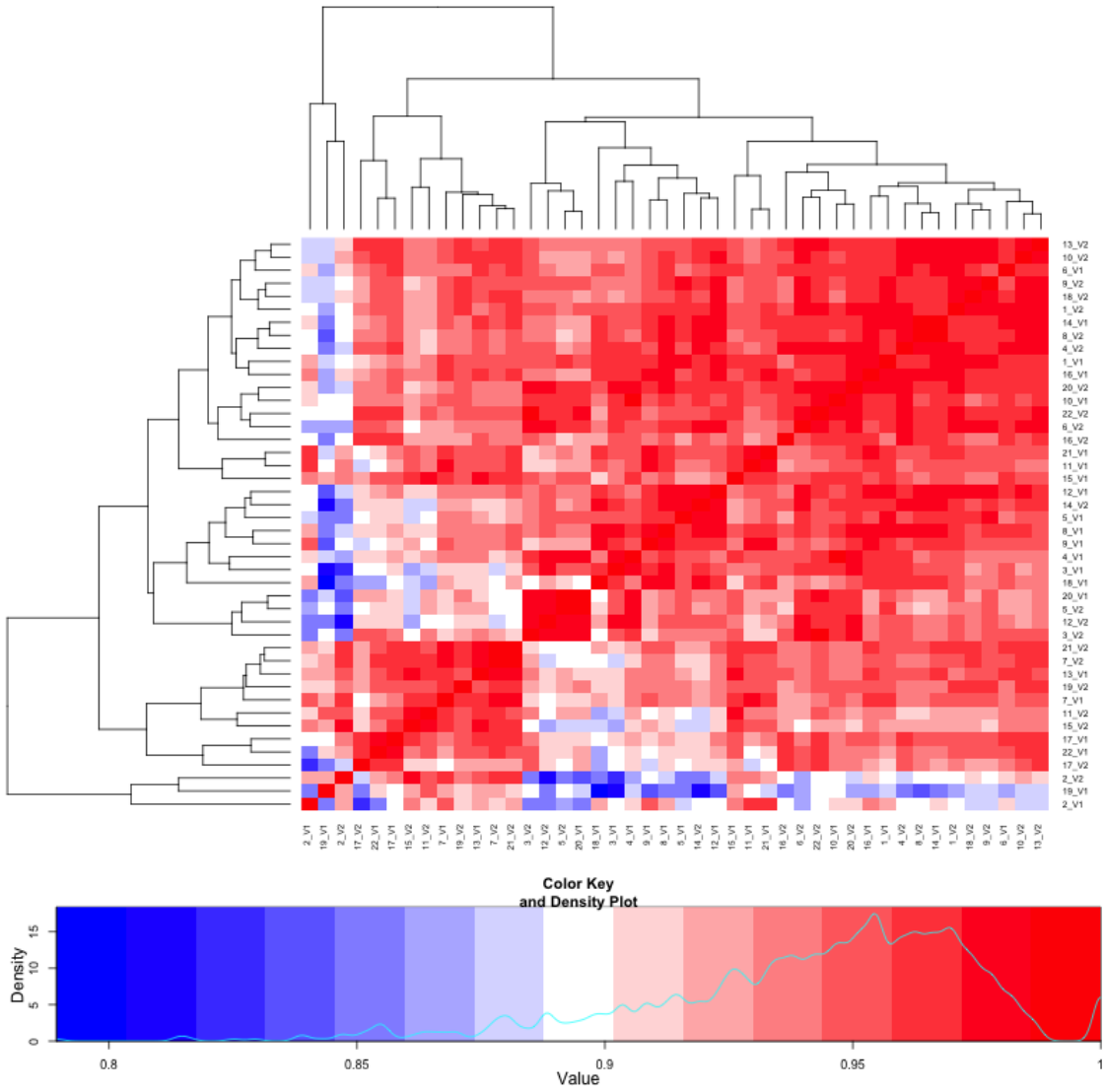


Figure 5.20: Heat map of correlation coefficients between variance stabilising normalised microRNA microarray expression data for all of the samples assessed in the study of serum microRNA profiles following immunisation with pandemic influenza vaccine. V1= baseline; V2= 3 weeks post-vaccination. (Note sample numbers have been anonymised $1-22$).

5.3.8.3 Statistical sensitivity analysis

As there were no suitable data available on the variance of miRNAs within childhood serum samples prior to the initiation of this study a formal power calculation was not conducted. However, a retrospective sensitivity analysis was conducted to assess the power given the variance observed in these empirical data. Based upon the empirical 75^{th} percentile of probe intensity standard deviations a sample size of 19 gives >80% power to detected a 1.5 fold-change in paired

miRNA samples at a FDR of 0.05, under the assumption that 95% of miRNAs will be non-differentially expressed (Figure 5.21). As the proportion of differentially expressed miRNAs was found to be higher than 0.05%, which is an estimate originating from mRNA expression studies, power was simulated with a varying non-differentially expressed value (π_0) (Figure 5.21).

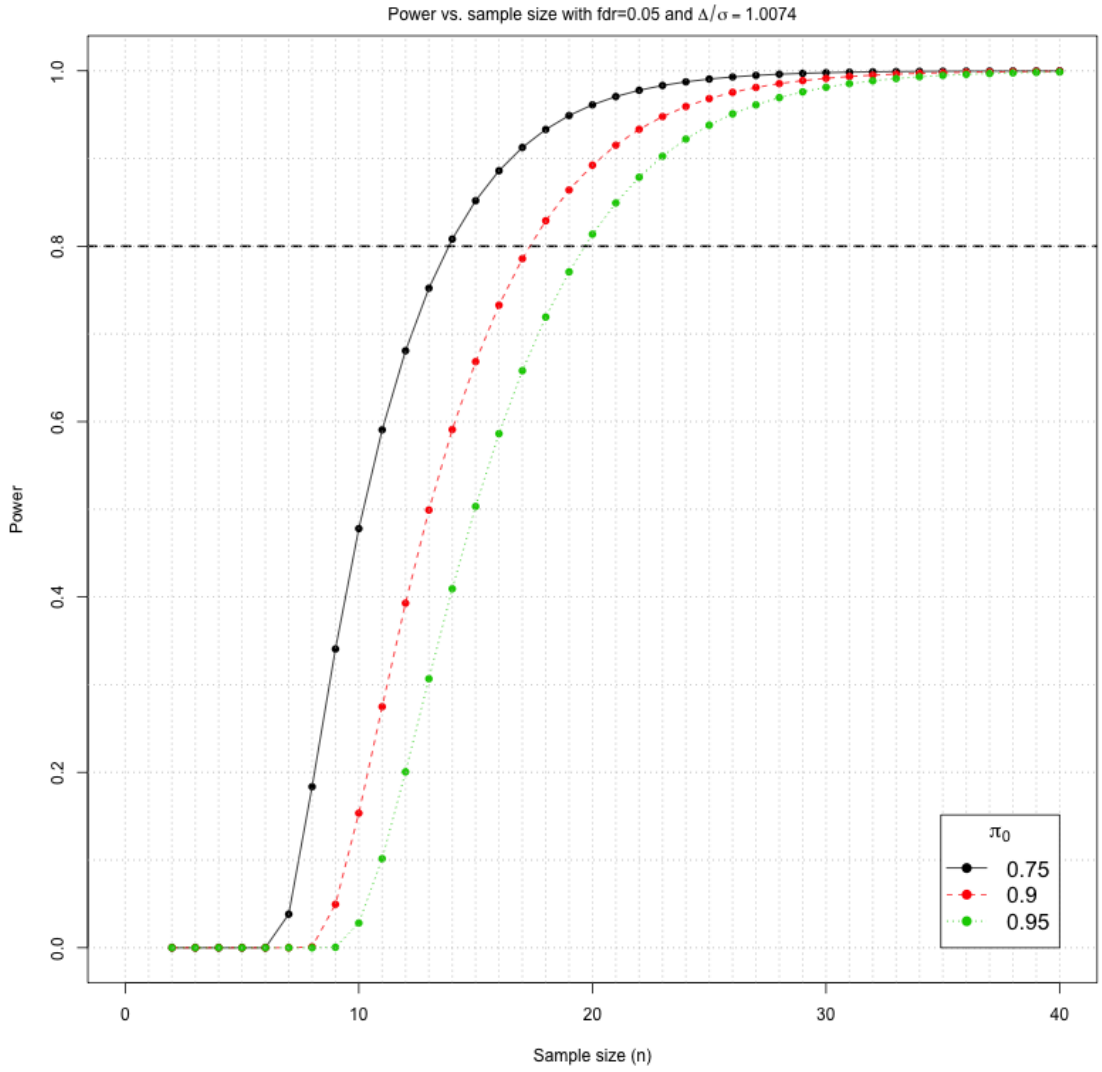


Figure 5.21: Retrospective power calculations to detect a 1.5 fold-change (Δ, $\log_2$ change of 0.58) in miRNA expression at day 21 following pandemic influenza vaccination at a false-discovery rate of 0.05, given the empirical 75^{th} percentile of probe intensity standard deviations. σ= 0.575. π_0 = percentage of non-differentially expressed miRNAs.

5.3.8.4 Linear regression analysis

A linear model was used to test for differential expression of miRNAs in paired serum samples, prior to and 21 days after immunisation with pandemic H1N1 influenza vaccine, using the 'lmFit' function of the 'limma' R Package (382). The top 25 miRNAs in order of statistical evidence of differential expression are shown in Table 5.10. Twenty-two miRNAs, 13.2% of the miRNAs detected in serum, were differentially expressed (FDR <0.05) following vaccination. MiRNA fold-changes and associated false-discovery rates are displayed in Figure 5.22. Of the 22 differentially expressed miRNAs 18 were upregulated following immunisation with pandemic H1N1 influenza vaccine.

Table 5.10: Top 25 differential expressed microRNAs in serum 21 days following immunisation with pandemic influenza vaccine.

MicroRNA identification	Normalised expression	Log_2 Fold-change	Fold-change	p-value	False discovery rate
hsa-miR-575	5.27	0.64	1.55	5.00E-08	8.00E-06
hsa-miR-1207-5p	7.48	0.40	1.32	8.33E-06	4.50E-04
hsa-miR-1202	7.91	0.40	1.32	1.02E-05	4.50E-04
hsa-miR-4270	7.06	0.50	1.41	1.08E-05	4.50E-04
hsa-miR-638	8.52	0.36	1.28	5.71E-05	1.88E-03
hsa-miR-483-5p	4.98	0.46	1.37	7.47E-05	1.88E-03
hsa-miR-2861	8.87	0.34	1.27	8.19E-05	1.88E-03
hsa-miR-3679-5p	5.96	0.43	1.34	9.83E-05	1.88E-03
kshv-miR-K12-3	5.76	0.32	1.25	1.01E-04	1.88E-03
hsa-miR-3141	4.86	0.35	1.27	4.15E-04	6.93E-03
hsa-miR-3196	4.33	0.32	1.25	5.31E-04	8.00E-03
hsv1-miR-H17	5.04	0.26	1.20	5.75E-04	8.00E-03
hsa-miR-30b	3.34	-0.32	0.80	6.67E-04	8.57E-03
hsa-miR-3656	8.17	0.27	1.20	9.63E-04	1.15E-02
hcmv-miR-UL70-3p	3.99	0.35	1.27	1.62E-03	1.71E-02
hsa-miR-1915	7.54	0.23	1.17	1.64E-03	1.71E-02
hsv2-miR-H25	3.70	0.33	1.26	4.17E-03	3.73E-02
hsa-miR-642b	5.11	0.26	1.20	4.35E-03	3.73E-02
hsa-miR-142-3p	5.92	-0.37	0.77	4.38E-03	3.73E-02
hsv2-miR-H10	8.77	0.21	1.15	4.55E-03	3.73E-02
hsa-miR-671-5p	6.01	-0.36	0.78	4.69E-03	3.73E-02
hsa-miR-129-3p	2.80	-0.26	0.84	5.01E-03	3.81E-02
hsa-miR-634	2.29	-0.26	0.84	7.11E-03	5.16E-02
hsa-miR-3648	5.32	0.31	1.24	7.85E-03	5.22E-02
hsa-miR-1225-5p	7.30	0.26	1.19	7.88E-03	5.22E-02

hcmv =human cytomegalovirus, hsv= herpes simplex virus, kshv= karposi sarcoma-associated herpesvirus, hsa= homo sapien miR= mature microRNA sequence, 5p= $5'$, 3p= $3'$

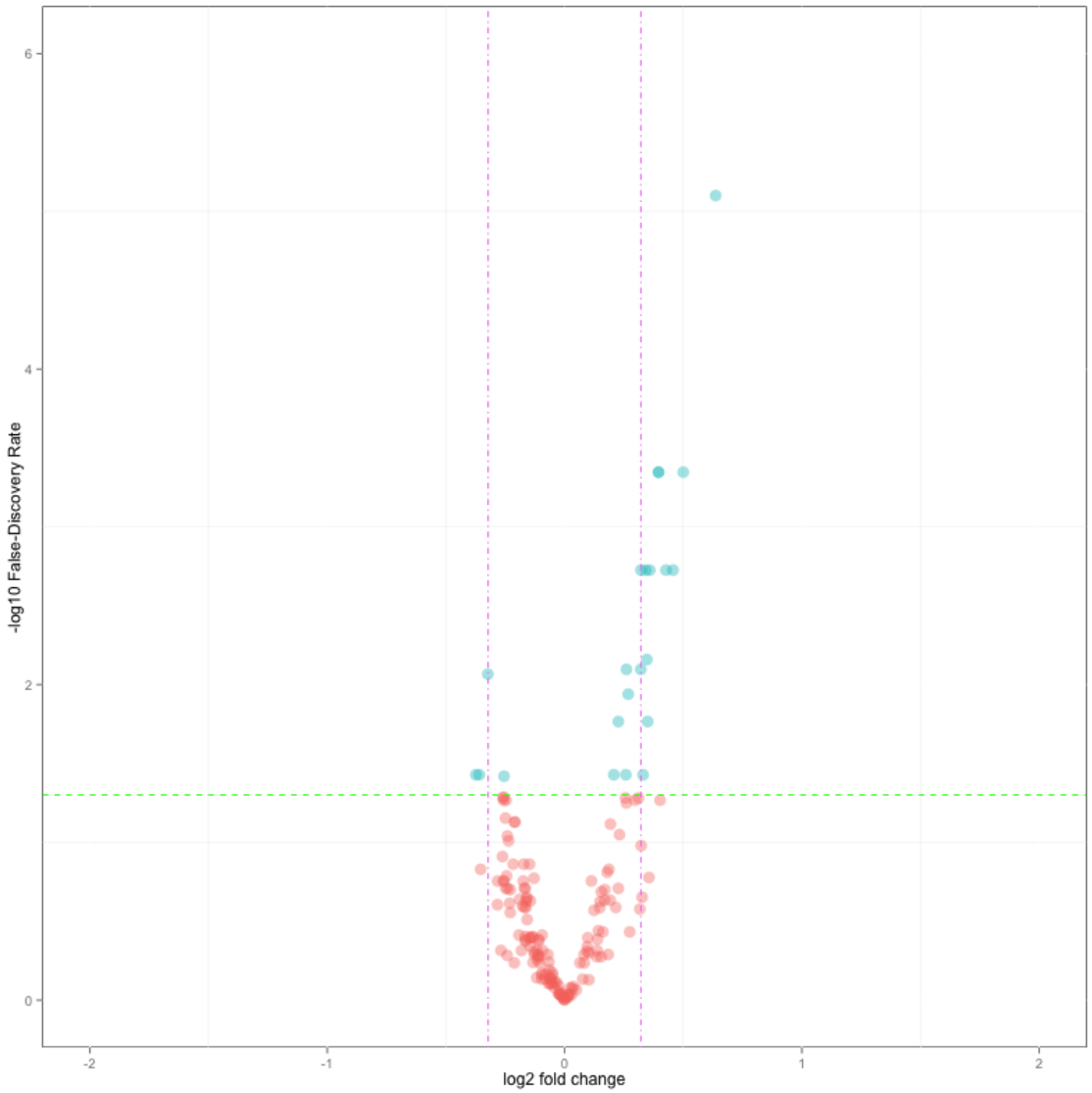


Figure 5.22: Volcano plot of -$\log_{10}$ false discovery rate and $\log_2$ fold-change in serum microRNAs 21 days following immunisation with pandemic influenza vaccine. The violet lines show probes with a fold-change of greater than 1.25 and the green line indicates differentially expression a false-discovery rate <0.05.

The 22 differentially expressed serum miRNAs (FDR <0.05) following pandemic influenza vaccination are graphically illustrated as a heat map in Figure 5.23. The top four differentially expressed miRNAs are shown as line graphs of pre- and post-vaccination normalised expression levels in Figure 5.24.

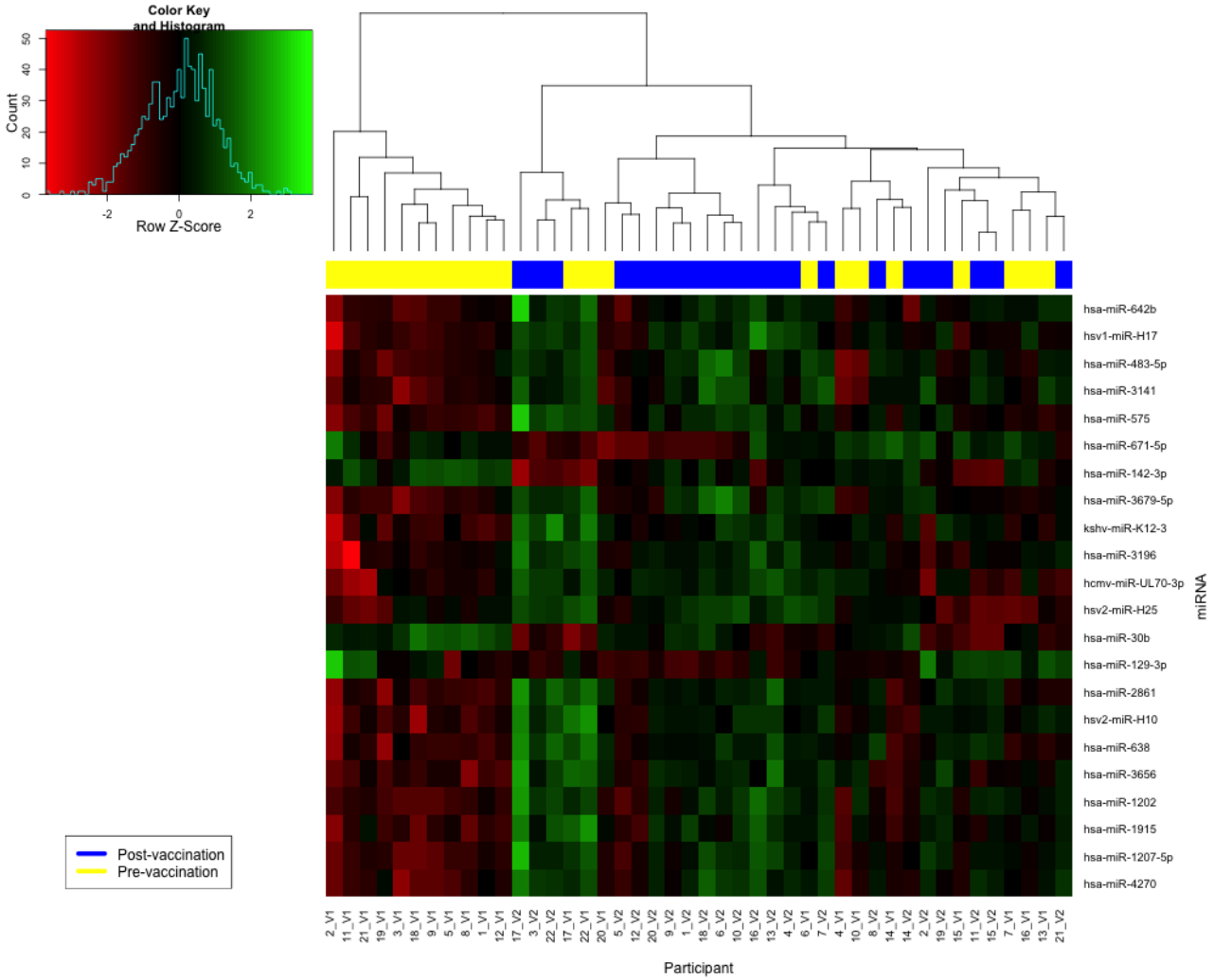


Figure 5.23: Heat map representation of z-scores of the 22 differentially expressed serum microRNA 21 days following immunisation with pandemic influenza vaccine. Hierarchical clustering was carried out on the column data (samples), as displayed in the dendrogram.

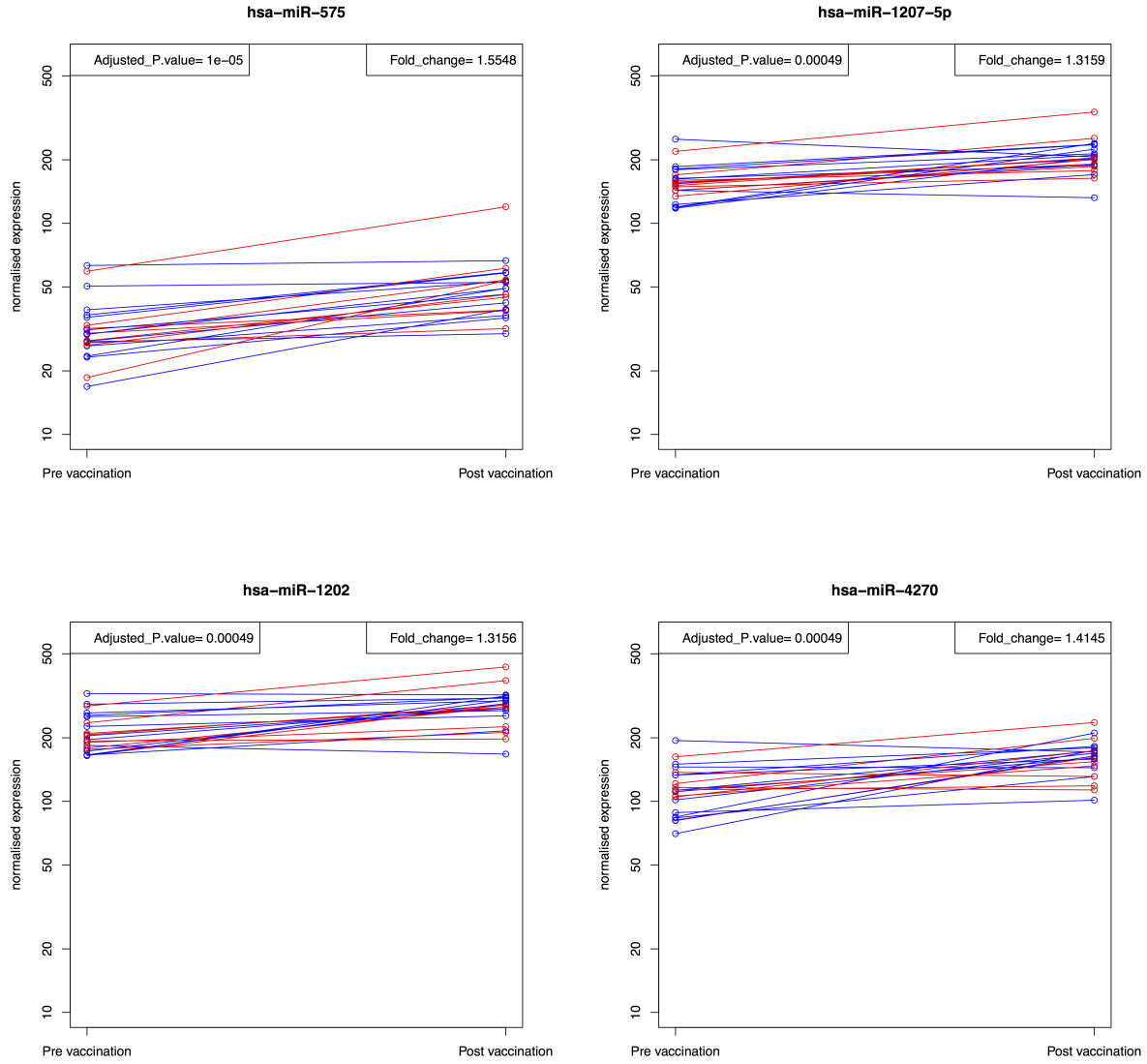


Figure 5.24: Line graphs of the four most differential expression miRNAs, based upon false-discovery rate, following immunisation with pandemic influenza vaccine. In red are the individuals who received the adjuvanted pandemic vaccine and in blue are the individuals who received the non-adjuvanted influenza vaccine.

5.3.8.5 Ingenuity Pathway Analysis

A particular miRNA may target hundreds of messenger RNAs and potentially have pleiotropic biological effects. IPA was conducted to assess whether the 22 differential expressed miRNAs were indicative of any particular biological pathways. The top networks from the IPA are shown in Table 5.11. The Molecule Activity Predictor tool was used to incorporate predicted downstream effects into the top network and is illustrated in Figure 5.25. Furthermore, using the IPA miRNA target filter four of the differentially expressed miRNAs, miR-3670-5p, miR-142-3p, miR-30b and miR-642, were found to have experimental data suggesting regulation of mRNAs involved in cellular and/or humoral immune response pathways (appendix Table 7.41).

Table 5.11: The top networks identified using Ingenuity® Pathway Analysis on the list of 22 differentially expressed serum microRNAs (FDR <0.5) following immunisation with pandemic influenza vaccine.

Rank	Score	Focus molecules	Diseases or functions (if annotated)
1	22	9	Cellular Development, Cellular Growth and Proliferation, Cell-To-Cell Signaling and Interaction
2	3	1	Cancer, Endocrine System Disorders, Organismal Injury and Abnormalities
3	3	1	
4	3	1	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
5	3	1	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
6	3	1	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
7	3	1	Hematological Disease, Immunological Disease
8	3	1	
9	3	1	

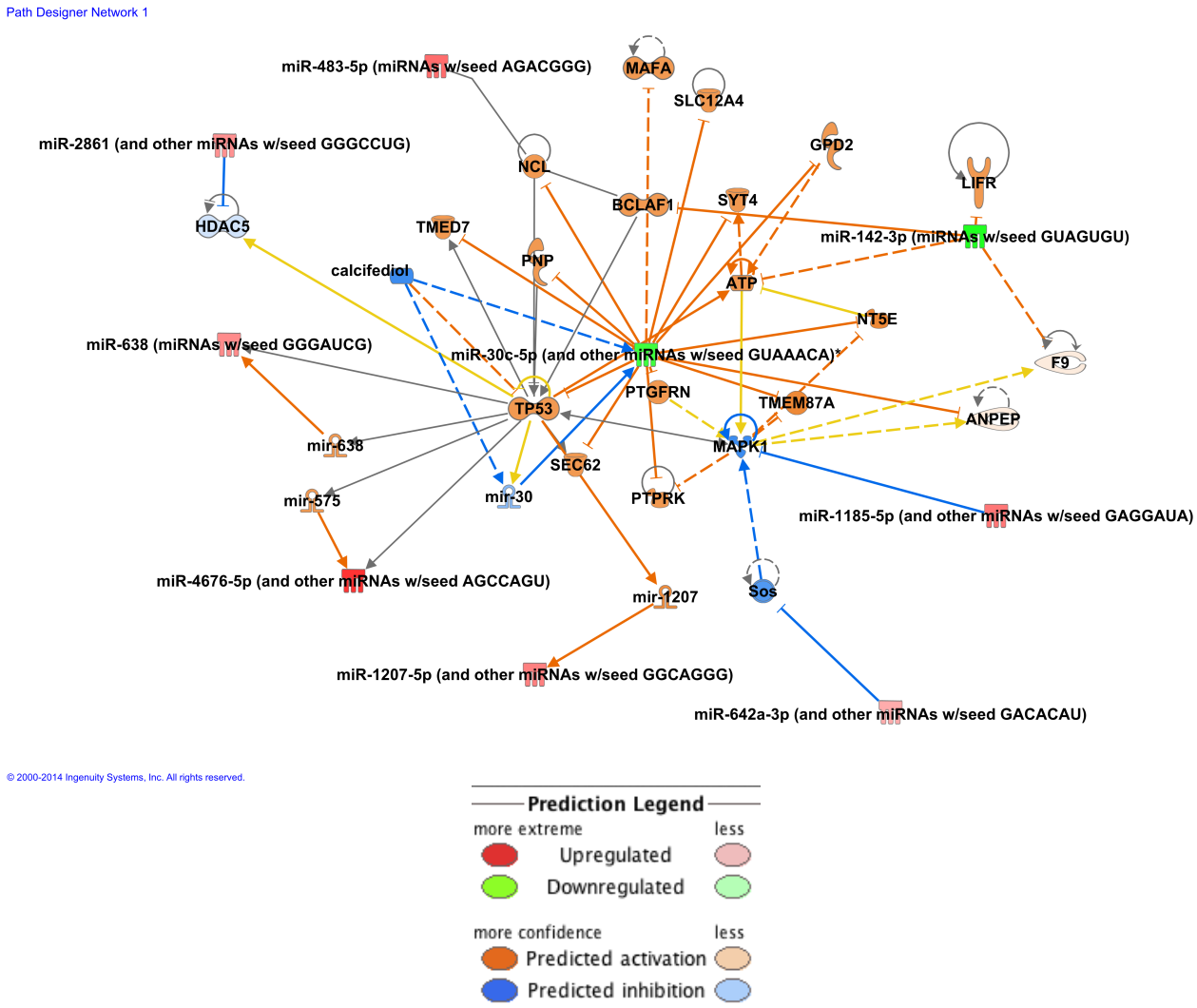


Figure 5.25: The top network from Ingenuity® Pathway Analysis conducted on the 22 serum microRNAs that were differentially expressed (FDR <0.5) following immunisation with pandemic influenza vaccine. The Molecule Activity Predictor tool was used to predict the downstream effector molecules that are included in this network.

5.3.8.6 Comparison of differential miRNA expression following either AS03B adjuvanted or non-adjuvanted 2009 pandemic influenza vaccine

A linear model was used to compare differential expression of serum miRNAs following either AS03B adjuvanted split virion or non-adjuvanted whole virion 2009 pandemic influenza vaccine. The 25 miRNAs with the most statistical evidence of being differentially regulated by these vaccines are shown in Table 5.12; however, none of these surpass the threshold for statistical significance (FDR <0.5).

Table 5.12: The 25 microRNA with the most statistical evidence (false discovery rate) of being differentially regulated in serum 21 days following AS03B adjuvanted compared to the non-adjuvanted pandemic influenza vaccine.

MicroRNA identification	t-value	p-value	False discovery rate
hsa-miR-4313	3.33	0.0027	0.29
hsa-miR-550a	3.22	0.0035	0.29
hsa-miR-4257	2.55	0.0171	0.48
hsa-miR-574-3p	-2.51	0.0187	0.48
hsa-miR-4259	2.49	0.0198	0.48
hsa-miR-574-5p	-2.45	0.0213	0.48
ebv-miR-BART12	-2.42	0.0231	0.48
hsa-miR-197	-2.41	0.0236	0.48
hsa-miR-1260	-2.36	0.0261	0.48
hsa-miR-634	2.32	0.0287	0.48
hsa-miR-30d	2.24	0.0338	0.48
hsa-miR-595	-2.24	0.0344	0.48
ebv-miR-BART16	-2.17	0.0393	0.50
hsa-miR-498	2.06	0.0497	0.59
hsa-miR-466	-1.95	0.0620	0.69
hsa-miR-3149	-1.87	0.0726	0.75
hsa-miR-933	1.78	0.0866	0.75
hsa-miR-32*	-1.78	0.0867	0.75
hsa-miR-494	1.72	0.0985	0.75
hsa-miR-296-5p	1.66	0.1084	0.75
hsa-miR-3679-5p	-1.64	0.1135	0.75
hsa-miR-146a	1.63	0.1158	0.75
hsa-miR-3610	-1.62	0.1174	0.75
hsa-miR-1274b_v16.0	1.59	0.1233	0.75
hsa-miR-572	1.59	0.1238	0.75

miR= mature microRNA sequence, *= the minor product of the miRNA hairpin
ebv= Epstein-Barr virus, hsa= homo sapien, 5p= 5′, 3p= 3′

5.3.8.7 Relationship between differential miRNA expression and immunological responses to pandemic H1N1 influenza vaccine

A linear model was used to test for a relationship between changes in serum miRNA expression following pandemic influenza vaccine and post-vaccination HAI titres. The 25 miRNAs with the most statistical evidence for a relationship with post-vaccination HAI titres are shown in Table 5.13; however, none of these surpass the threshold for statistical significance (FDR <0.5).

Table 5.13: The 25 serum microRNAs with the most statistical evidence of a relationship between fold-change and haemagglutination inhibition assay titres following immunisation with pandemic influenza vaccine.

MicroRNA identification	t-value	p-value	False discovery rate
kshv-miR-K12-8*	-2.12	0.0444	0.91
hsa-miR-3180-5p	-1.88	0.0716	0.91
hsa-miR-3195	1.88	0.0720	0.91
hsa-miR-634	-1.85	0.0761	0.91
hsv1-miR-H8	-1.84	0.0786	0.91
hsa-miR-1225-3p	-1.83	0.0800	0.91
hsa-miR-342-3p	1.81	0.0830	0.91
hsv1-miR-H7*	-1.81	0.0834	0.91
hsa-miR-575	1.74	0.0941	0.91
hsa-miR-3676	-1.73	0.0963	0.91
hsv1-miR-H1*	1.73	0.0968	0.91
hsa-miR-328	1.68	0.1061	0.91
hsa-miR-129*	-1.67	0.1073	0.91
hsa-miR-1275	1.64	0.1137	0.91
hsa-miR-129-3p	-1.60	0.1221	0.91
hsa-miR-1246	-1.57	0.1297	0.91
hsa-miR-572	1.56	0.1325	0.91
hsa-miR-466	1.52	0.1425	0.91
hsa-miR-638	1.47	0.1540	0.91
hsa-miR-1260	1.46	0.1580	0.91
hsa-miR-630	1.45	0.1593	0.91
hsa-miR-1914*	-1.43	0.1646	0.91
hsa-miR-4259	-1.39	0.1773	0.91
hsa-miR-32*	1.38	0.1795	0.91
hsv1-miR-H6-5p	1.36	0.1874	0.91

hsv= herpes simplex virus, kshv= karposi sarcoma-associated herpesvirus
hsa= homo sapien, miR= mature microRNA sequence, 5p= 5′, 3p= 3′
*= the minor product of the miRNA hairpin

5.3.9 Discussion

In this section miRNAs were assessed in serum from healthy children pre- and post-vaccination with a two-dose regimen of either adjuvanted or non-adjuvanted pandemic influenza (H1N1) vaccine. One hundred and sixty-seven miRNAs were detected in at least 65% of the serum samples, and of these 22 were differentially expressed 21 days post-vaccination. Eighteen of these differential expressed miRNAs were upregulated 21 days after vaccination. As miRNAs can bind to several mRNA targets, their differential expression may affect a multitude of different biological pathways. The top de novo network formed by IPA of differentially expressed miRNAs was annotated as involved in cellular development, proliferation and signalling. There were no statistically significant differences in post-vaccination serum miRNA expression between individuals who received the AS03B adjuvanted vaccine compared to the non-adjuvanted

pandemic influenza vaccine.

There were a number of limitation to this study. Firstly, while miRNAs are relatively stable molecules compared to other RNA species, their expression may be influenced by differences in serum processing such as time taken before centrifugation (415). Nevertheless, there was no evidence of systematic differences in processing times between pre- and post-vaccination samples, and the majority of the serum samples were processed within 2.5 hours.

Secondly, normalisation of miRNA microarray data is problematic, the assumption that the majority of transcripts will be constant may not be valid, as few miRNAs may be expressed in a particular sample and the proportion of these that are differentially expressed following an intervention may be relatively large (412, 413, 414). There are several normalisation and quality control protocols that use spike-in controls but these were not included in the microarrays performed in this study (412, 414).

Finally, while miRNAs are detectable and stable in serum, the role of these circulating nucleic acids remains unclear. There are some data suggesting that miRNAs are actively secreted but their downstream effects are unclear (408, 409, 410). MiRNA levels in cellular compartments are far higher than in the extracellular compartment and more detailed characterisation of both these compartments may be biological informative. Nevertheless, there are several reasons why serum miRNAs may make useful clinical biomarkers, including their stability in various bodily fluids and their relatively low level of complexity (414). To date, there are few high-throughput data characterising serum miRNAs in human health and disease (414). A recent study assessed the expression of miR-155 in serum 4 to 6 weeks after three doses of hepatitis B vaccination using quantitative real-time PCR, finding this miRNA to be expressed at higher levels in non-responders (anti-HBs antibody levels < 10 mIU/ml) than responders (anti-HBs antibody levels ≥ 10 mIU/ml) (411). However, there are currently no published data using high-throughput approaches to assess miRNAs expression following vaccination.

A single miRNA may have hundreds of mRNA targets making biological interpretations of the differentially expression of these molecules following vaccination challenging. Several of the miRNAs shown to be differentially expressed in this study have previously been shown

to be differentially expressed in other clinical contexts; for example, hsa-miR-575 and hsa-miR-638 have been shown to be upregulated in lupus nephritis patients, and hsa-miR-1207-5p and hsa-miR-4270 were found to be down regulated in T cells from individuals with immune thrombocytopenia (416, 417). Furthermore, using the IPA miRNA target filter four of the 22 differentially expressed miRNAs had experimental data showing regulation of mRNAs involved in cellular and/or humoral immune response pathways.

The AS03B adjuvanted vaccine has been shown to be more immunogenic and reactogenic than the non-adjuvanted pandemic (H1N1) vaccine in this cohort of children (282). However, in the subset that were analysed for serum miRNA expression no statistically significant differences were seen in miRNAs levels following immunisation with the adjuvanted compared to the non-adjuvanted vaccine. Furthermore, while a number of miRNAs were differentially expressed following immunisation with pandemic influenza vaccine, no statistically significant relationships were observed between their regulation and post-vaccination HAI titres. It may be the case that this subset lacked power to detect such differences and more research is needed to assess the biological significance of differential serum miRNA expression following vaccination.

5.3.10 Conclusion

One hundred and sixty-seven miRNAs were detected in sera from children enrolled in a pandemic H1N1 influenza vaccine study; 22 miRNAs were differentially expressed 21 days after immunisation. Several of the differentially expressed miRNAs have previously been shown to regulate mRNAs involved in cellular and humoral immunity. However, as miRNAs may target hundreds of different mRNAs there are many biological pathways that these miRNAs may be affecting. Further research is needed to describe the origin of serum miRNAs and the biological significance of their differential expression following vaccination.

5.4 Chapter summary

Transcriptomic approaches have transformed the way in which complex biological systems can be interrogated and it is hoped that these approaches may soon directly contribute to vaccine development and evaluation. In this chapter high-throughput transcriptomic approaches were used to assess both mRNA and miRNA following childhood influenza vaccination. A number of miRNAs were differentially expressed in serum 21 days after immunisation with pandemic influenza vaccine. Furthermore, a number of gene pathways were seen to be differentially regulated in whole blood following seasonal TIV, some of which were associated with the magnitude of post-vaccination HAI titres.

Chapter 6

Discussion

Vaccines have had a profound influence on human health; however, the mechanisms involved in the generation and persistence of vaccine-induced immunity are not well understood. There are now substantial data supporting the contribution of human genetic variation in determining the heterogeneity seen in several clinical phenotypes, including vaccine-induced immunity. Recent technological advances have afforded unprecedented resolution and throughput to study genetic variation. This thesis used contemporary methodologies to describe the relationship between genomic and transcriptomic variation, and vaccine-induced immunity following childhood immunisations.

Hitherto, the majority of studies assessing the genetic factors influencing vaccine-induced immunity have adopted a candidate gene approach. While these studies have purported a number of associations between SNPs and vaccine-induced immunity, they have typically employed liberal statistical corrections for multiple testing and few associations have been replicated in independent populations. In this work, associations between variants in *TLR3* and *CD44*, and MenC-specific serological immunity following childhood immunisation were replicated in an independent population of infant vaccinees (38). These data implicated exonic variants, and a particular *TLR3* haplotype, with the magnitude of MenC-specific immune responses.

This thesis described the first GWASs assessing the genetic contribution to three routine childhood immunisations: MenC conjugate vaccine, Hib conjugate vaccine and TT vaccines.

Initially, SNPs in two loci were shown to be associated, at the level of genome-wide significance, with the persistence of MenC-specific SBA titres following a single-dose of MenC conjugate vaccine administered to children aged 6-15 years. These loci were in two annotated genes, *CNTN6* and *ENKUR*, suggesting a novel biological role for these genes in immunity. Furthermore, a number of variants showed suggestive statistical associations with the persistence of MenC-specific SBA titres; such as variants within *GCSAM* a gene highly expressed in human lymph node, which encodes a protein involved in the regulation of B cell receptor signalling and lymphocyte migration (260, 262).

The study of persistence of MenC-specific serological immunity following childhood immunisation was then expanded to include individuals vaccinated as young children, in a two-stage design. In stage 1, participants were densely genotyped and these data were used to select a subset of markers for genotyping in an independent stage 2 cohort. A number of SNPs with p-values suggestive of association with the persistence of MenC-specific immunity following childhood vaccination were highlighted in stage 1, including SNPs within genes with annotated immunological function such as *SIRPG* and *GGT5*. *SIRPG* encodes signal-regulatory protein gamma, a protein involved in T cell activation and leukocyte migration (287). *GGT5* encodes gamma-glutamyltransferase 5, a protein that converts leukotriene C4 to leukotriene D4 and has a role in the acute inflammatory response (289). In addition, SNPs with p-values suggestive of association with MenC-specific SBA titres were demonstrated within *PTPRD*, a gene recently shown to harbour variants associated with rubella specific IFN-γ responses following immunisation with MMR vaccine (28).

This thesis then described two-stage GWASs assessing the genetic determinants of the persistence of immunity to two other routine childhood immunisations, Hib and TT vaccine. A number of SNPs with p-values suggestive of association with the persistence of PRP-specific IgG concentrations following childhood Hib vaccination were highlighted, including SNPs within loci with annotated immunological function such as *FOXP1* and IGL. *FOXP1* is a transcription factor involved in the regulation of germinal centre B cells and the IGL locus contains the germline gene segments encoding the lambda light chain portion of immunoglobulin (328, 329).

Genome-wide significant associations were found between SNPs present in an intergenic region of chromosome 10 and the persistence TT-specific IgG concentrations following childhood immunisations with TT-containing vaccines. The closest annotated transcribed RNA to this association was miRNA378C (∼115kb) and the closest protein-coding transcript was *TCERG1L* (<250kb). The human genome has been shown to pervasively transcribed and this association signal may represent a functional non-coding RNA; alternatively, it may tag a DNA based regulatory element (343, 344, 345). While the genome-wide significant SNPs at this locus have not been annotated as eQTLs on GTExPortal (`http://www.gtexportal.org`), a recent study showed a significant proportion of variants associated with complex phenotypes display context-specific eQTL activity that may only be identifiable upon triggering of immune responses (265). A number of SNPs displayed p-values suggestive of association with the persistence of TT-specific IgG concentrations; these included SNPs within genes with annotated roles in immunity such as *MICB* and *HLA-DQA1*. The lead SNPs in both *MICB* and *HLA-DQA1* have been annotated as eQTLs, potentially alluding to the mechanism by which these associations have arisen. Furthermore, classical HLA allele imputation associated two alleles, HLA-DQB1*0201 and HLA-DRB1*0301, with the persistence of TT-specific IgG concentrations. Interestingly, HLA-DRB1*0301 has an unusual hypervariable region and is distinct, both genetically and immunologically, from other *HLA-DRB1* alleles (350). The HLA molecule encoded by HLA-DRB1*0301 presents a peptide motif that is distinct from that recognised by the other major *HLA-DRB1* alleles (350). Furthermore, HLA-DRB1*0301 has been associated with susceptibility to several autoimmune and allergic phenotypes, as well as lower HBsAg-specific IgG responses following HBV (352, 353). These data suggest differences in peptide binding preferences for this allele; however, the mechanisms underlying these observations and their relevance to vaccine-induced immunity need to be further evaluated.

Statistically significant correlations were observed between the persistence of MenC-specific SBA titres, MenC-specific IgG concentrations, PRP-specific IgG concentrations and TT-specific IgG concentrations, following childhood immunisations. Nevertheless, comparisons of the GWASs conducted on each of these vaccines described only nine genes that contained variants that

were associated with immunity to more than one of these vaccine antigens. Interestingly, SNPs within *PTPRD* were highlighted as being associated with both MenC-specific SBA titres and PRP-specific IgG concentrations, this gene has also been associated with rubella-specific IFN-γ responses following MMR vaccination (28). These data suggest a novel role for *PTPRD* in vaccine-induced immunity and highlight the potential of assumption-free GWASs to elucidate novel immunobiology.

Transcriptomic analysis have transformed the way complex biological systems can be interrogated. In this thesis, coding and non-coding transcriptional profiles were assessed following childhood influenza vaccination. A single gene, *GAPT*, was found to be differentially expressed in peripheral whole blood 21 days after immunisation with TIV. The protein product of *GAPT* is thought to regulate B cells following stimulation via the BCR and may also have a role in the maintenance of marginal zone B cells (395). While single-gene analysis highlighted only one DEG, gene set and pathway analysis implied more substantial alterations to gene regulation in peripheral blood 21 days after childhood immunisation with TIV. GSEA found a number of gene sets that were differentially regulated 21 days after TIV; identifying a core set of upregulated genes associated with T cells as well as a core set of downregulated genes associated with early (<7 days) responses to TIV, LAIV and YF-17D. IPA suggested transcriptional profiles 21 days after TIV were associated with downstream effects on the formation of the CD8 receptor, BCR complex, IgM, IgA, IgG3 and IgE. Furthermore, IPA suggested a relationship between fold-changes in genes involved in antigen presentation and the HAI titres following TIV. Previous data have shown the antigen presenting pathway to be enriched for genes with differential expression associated with HAI titre following adult immunisation with TIV (88). Interestingly, the aforementioned study associated early upregulation of genes involved in antigen presentation with greater HAI titres, whereas the data presented in this thesis associated late downregulation of these genes with greater post-vaccination HAI titres (88). These data highlight the importance of understanding the kinetics of gene expression in describing the mechanisms involved in determining the magnitude of vaccine-induced immunity.

This thesis also described a novel study using microarray technology to assess serum miRNAs

following childhood immunisation with pandemic influenza vaccine. This study described 22 miRNAs that were differentially expressed in serum 21 days after childhood immunisation with pandemic influenza vaccine. IPA of these differentially expressed miRNAs highlighted their involvement in cellular development, proliferation and signalling. Furthermore, four of these miRNAs have previously been shown to be involved in cellular and/or humoral immunity.

Technological and analytical advances mean that complex systems, such as the immune system, can be evaluated in a more comprehensive manner than ever before. This thesis described the use of contemporary microarray platforms to assess the genome and transcriptome in a broad and high-throughput manner. These represent just two of an increasing number of 'omic' approaches that are now available for utilisation in clinical research; others include epigenomics, proteomics, metabolomics and microbiomics. Indeed, the determinants of vaccine immunity represent interactions between the variables captured in 'omics' data, and other physiological and environmental variables such as age, nutritional status and chronic disease. Therefore, in addition to using more 'omic' approaches there may be considerable benefit from integrative analysis of systems data. The integration of genomic and transcriptomic analysis has been used to suggest the function of numerous genetic variants associated with complex phenotypes (264, 265). Furthermore, recent data have suggested many of these interactions are context-specific and may only be identifiable upon triggering of immune responses (265). Given these data, future studies may choose to map eQTLs associated with differential gene expression following human vaccinations; in an analogous manner to that described by Fairfax et al 2014, where context-specific eQTLs were identified using innate stimuli (265). The use and integration of system biology techniques hold promise for the deconvolution of complex biological systems such as vaccine immunity (418, 419).

This thesis applied systems biology analyses to dissect the mechanisms involved in vaccine-induced immunity. This work yielded novel insights into genetic determinants involved in vaccine-induced immunity. While these findings require further work to refine the mechanisms underlying these observations, they provide important data to aid the design of pragmatic functional experiments.

Chapter 7

Appendix

7.1 Appendix: Introduction

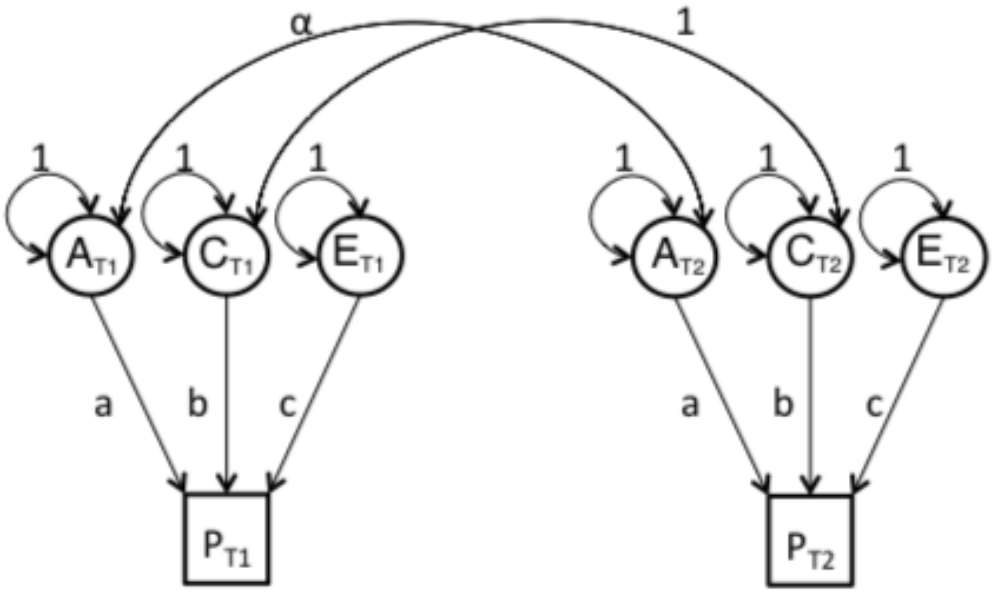


Figure 7.1: Parameterisation of the ACE genetic model (reproduced from Neale et al 1997) for twin data (12). A = additive genetic, C= common environment, E= specific environment, P= phenotype, T1= twin 1 and T2 =twin 2. The parameter α is fixed at 1 for monozygotic and 0.5 for dizygotic twins, latent variables have variance of 1.

7.2 Appendix: Candidate gene studies of immunological responses to childhood vaccines

7.2.1 The relationship between *CD44* and *TLR3* polymorphisms, and infant immune responses to capsular group C meningococcal conjugate vaccine

Table 7.1: Primers and PCR settings for the amplification-refractory mutation system PCR genotyping of *TLR3* single-nucleotide polymorphisms, in the candidate gene study of the magnitude of capsular group C meningococci (MenC)-specific serological immunity following infant immunisation with MenC conjugate vaccine.

SNP	Inner primer forward	Inner reverse	Outer primer forward	Outer primer reverse
rs3775292	1.75μl 'C' allele	1.75μl 'G' allele	0.5μl	0.5μl
Sequence (5′-3′)	CCATGGGCAGAGCAT--TTCTCCCCCTTCCGC	GAACCGAGTAAGGAA--GGACTCGTGCATTCC	GATGCAGTTCTTTACTC--CATCTCCGCTCAC	CCAGGAGTGACTTCAC--GAGCAGTTTATTCA
rs11732384	1.75μl 'A' allele	1.75μl 'G' allele	0.5μl	0.5μl
Sequence (5′-3′)	GAGAACTAAATTAGA--ATATTAGTAAAATAA	TAGAATTGTTCTAATA--TACTTAAAAATACC	TACAATGATTAGTGCCA--GAGTTGTAAAGGC	CATTGTCACACTTTAT--TTTGGTGGGATGTG
rs5743312	0.5μl 'C' allele	0.5μl 'T' allele	0.5μl	0.5μl
Sequence (5′-3′)	GTGGATAGTCCCTAT--CTGTGTCACATACTC	AGGCAGAGGCTAGAGA--GCATTACATTCCGA	CTCAGGTTCAGCTGGAA--AATCTCCAAGAGC	TTGTTAAAAGACAAAG--CCCAGAGACCAAGC

SNP, Single nucleotide polymorphism

Table 7.2: Sequenom MassARRAY® iPLEXs designed to assess the relationship between *TLR3* and *CD44* SNPs, and the magnitude of capsular group C meningococci (MenC)-specific serological immunity following infant immunisation with MenC conjugate vaccine.

i	SNP	Gene	PCR primer 2	PCR primer 1	Dir	Extension primer	A1	A1 mass	A2	A2 mass
1	rs2785172	*CD44*	ACGTTGGATGGCAACAAGAGCTCCTTTCCT	ACGTTGGATGATGACATTGTCAGAGGGC	R	TCTCCCCTTCGCCAC	G	4647.00	A	4727.00
1	rs7657186	*TLR3*	ACGTTGGATGAGGTCTTGTTCTAGGCACAG	ACGTTGGATGTGCTTCTCACTAGAATGTGAAGC	F	GGCACAGACAACACC	A	4811.20	G	4827.20
1	rs7105890	*CD44*	ACGTTGGATGTGAGGCTTTCACATCCAAGG	ACGTTGGATGGTAATACTGACACTGCTGGG	R	TCCTCCAGAGCAGTAG	G	5113.40	A	5193.30
1	rs7945310	*CD44*	ACGTTGGATGAAACCAGGAAAAGGTGAGGC	ACGTTGGATGGAGGCATGGAAAGTCTCATT	F	GGTGAGGCAAAGGTAG	C	5282.50	T	5362.40
1	rs353644	*CD44*	ACGTTGGATGAGGTGGTGGAAACTTTCCTC	ACGTTGGATGATCCCCCACCTGGCACCTT	F	tcGGGCCTCCAAGGCCA	C	5403.50	T	5483.50
1	rs353639	*CD44*	ACGTTGGATGTGCACATAAGAAGCAGGGTC	ACGTTGGATGGTGTAACTCCCCAAACATC	R	aCTTCGCAGTCCAAGAAC	G	5699.80	T	5723.80
1	rs12419062	*CD44*	ACGTTGGATGCCAGCCTAATGCAATGCTTC	ACGTTGGATGTCCTTCCTCTAAGCATACGTC	R	CTTCCTCTTTCCTGTAACA	G	5911.90	A	5991.80
1	rs11033010	*CD44*	ACGTTGGATGGTATCTGGTCAGATGACAGC	ACGTTGGATGGTACTATATGGGAGATGCGG	R	GATAATTGCCCAAAGAGAC	G	6077.00	A	6156.90
1	rs4141971	*CD44*	ACGTTGGATGACCACCCAAAGACAACCTAC	ACGTTGGATGAGAATCTGGAGACGTCTGGT	R	tttCCAAAGAGAGGGCACT	T	6108.00	C	6124.00
1	rs353619	*CD44*	ACGTTGGATGCCCTTGAGGCTTCCTTATCT	ACGTTGGATGGTCTGATAACAGGGCATGTC	R	GGCTACATAATTCCTAACAC	G	6292.10	A	6372.10
1	rs8193	*CD44*	ACGTTGGATGTGCCCTTCCATAGCCTAATC	ACGTTGGATGGTAACTAGTACACCCCAACC	F	ggggTAATCCCTGGGCATTG	C	6436.20	T	6516.10
1	rs11033008	*CD44*	ACGTTGGATGGTATGTTACCTACCAGTGCC	ACGTTGGATGGGGCCTTTTAGAAGGAACGG	F	acTGTCTTTGGGCATTCTGAC	C	6650.30	T	6730.30
1	rs7934770	*CD44*	ACGTTGGATGAGTAACCATGTCGAAGGTGC	ACGTTGGATGTGTGTTCGCTAACATCCGAC	F	AAGGTGCTTATCATTATGGTA	G	6762.40	T	6802.30
1	rs932703	*CD44*	ACGTTGGATGCCCATGTTAGACTCCAATGC	ACGTTGGATGATGGCCACGTATAGCTTCCAAAG	F	CTTATGCTCCTTCAATTCATTA	C	6857.50	T	6937.40
1	rs996076	*CD44*	ACGTTGGATGCCAGGTTGAATTTCACAGAG	ACGTTGGATGTGTCCACTTGTGCAAACACC	F	ttTTCTGTATTTAGACCTGCTG	A	6968.60	G	6984.60
1	rs10734436	*CD44*	ACGTTGGATGAAGGACAGGATGACTTGTTG	ACGTTGGATGTCCTTTCAAAGAAGCCTAC	R	ttACTTGTTGAATCTTGTAGATG	G	7321.80	A	7401.70
1	rs353634	*CD44*	ACGTTGGATGTGGAGTATGCGAACCAGATG	ACGTTGGATGCTGCCTGACTAGTATAGTCC	F	cataTGTCCTGGGAAGAACAAAT	A	7351.80	G	7367.80
1	rs13126816	*TLR3*	ACGTTGGATGTCCATCAACCATAGCAACGG	ACGTTGGATGGTCTTGAAGTTCAGTGAGCG	F	gAGGCAATCTAGAAGAAGAGCAA	A	7434.90	G	7450.90
1	rs353626	*CD44*	ACGTTGGATGCAATCACTTTGGGTGCAGTC	ACGTTGGATGCTAGACTGACCAAGCTAAGG	F	cccgGTGCAGTCCACCCAGAGC	C	7530.90	T	7610.80
1	rs353615	*CD44*	ACGTTGGATGGACCTTTCATAACCTTGCAG	ACGTTGGATGTGCAGCCCTCATATTAACATCCG	R	aatcTCATCCCATACAGGAGGCCA	T	7562.00	C	7578.00
1	rs112762	*CD44*	ACGTTGGATGAATGCCAACAGCAAGAGCAG	ACGTTGGATGAGTGCCTAAAATAGAGCCTG	R	GGCTAATATTTATTGTGCATTCATT	T	7923.20	C	7939.20
1	rs11033021	*CD44*	ACGTTGGATGACCCTGCATGATCTTAGAC	ACGTTGGATGGAGGAAAATAGTGCTCATGT	R	GCATGATCTTAGACATTGTGATATG	T	7982.20	C	7998.20
1	rs3775291	*TLR3*	ACGTTGGATGCCCAACCAAGAGAAACGATC	ACGTTGGATGTATCACTTGCTCATTCTCCC	R	gaggTTTTATTCTTGGTTAGGTTGA	T	8035.30	C	8051.30
1	rs7668666	*TLR3*	ACGTTGGATGTCCCAGTAGTCTTATAGCCTG	ACGTTGGATGTTCCCTAAATAGGACTCTGC	R	cccctTCTTATAGCCTGGAACAAAAT	C	8161.40	A	8201.20
1	rs4756195	*CD44*	ACGTTGGATGGTTCAGTTTAGGAGACTC	ACGTTGGATGTGCCAAAGTGCTGGGATTGTAG	R	aaaaTTTAGGACAGACTCCTTAAAAAG	G	8546.60	A	8626.50
1	rs11033013	*CD44*	ACGTTGGATGCTGTCCTTCTCTAAGAAGAC	ACGTTGGATGCTAACAACAAGGAGTGGTGC	R	tcctcCTCTAAGAAGACATGTTGACATC	G	8754.70	C	8794.80
1	rs7947433	*CD44*	ACGTTGGATGCCAATGTCATATAAGGCAAAC	ACGTTGGATGTCAAAATAACTAGGCATTC	R	GTAAAGATAATTTTTCCATGTTTATTA	G	8837.80	A	8917.70
2	rs13347	*CD44*	ACGTTGGATGTCGATCTCAAGGCGTAACTC	ACGTTGGATGGCCCAGGTCACCTCAAAATC	F	GCACAGCTTGACTACA	C	5097.40	T	5177.30
2	rs353618	*CD44*	ACGTTGGATGCTAGTATGCGTTGAGTGAAC	ACGTTGGATGTTTGACTTGCCCCAGACTCTC	R	TGTTGACTCAACTCACA	T	5400.60	C	5416.60
2	rs2553809	*CD44*	ACGTTGGATGCCCATACAAAGACTGGACTC	ACGTTGGATGCTCTTGTTGCCCCATTCTG	R	CCTTTCCAAAGTTTCCTC	T	5631.70	C	5647.70
2	rs1547060	*CD44*	ACGTTGGATGAAGCCTTGCATCTGAAAGCC	ACGTTGGATGCAAGCTCTTTAATCACTCTGG	F	cGCAGTGTCCCAATTTAAC	C	5994.90	A	6019.00
2	rs7110737	*CD44*	ACGTTGGATGCATGATCCTTCTGTTAGGTG	ACGTTGGATGGCAGCACCCTGATATTGTCAC	F	CCTTCTGTTAGGTGCCATCTC	A	6321.10	T	6377.00
2	rs353612	*CD44*	ACGTTGGATGGCTCTGGGAAGCTACTTTTG	ACGTTGGATGGCAGCCACTTTTGTATTCTGC	R	GTTTTCCAGAGTTTTTTAACAA	G	6657.40	A	6737.30
2	rs10768114	*CD44*	ACGTTGGATGGTATCGCTAAAACATATGG	ACGTTGGATGCCTGAGCAGCTAGAGCTTTC	F	TCGCTAAAACATATGGCTATTC	G	6980.60	T	7020.50
2	rs11607862	*CD44*	ACGTTGGATGCTGTTTGACTGTAACATGGC	ACGTTGGATGCAGCCCATACAGTCTCTAAT	R	ggACTTGTACCAGCGTAATTATT	T	7324.80	C	7340.80

i= Sequenom MassARRAY® iPLEX, A1= Allele 1, A2= Allele 2, Dir= Direction, F= Forward, R= Reverse.

Table 7.3: PCR and sequencing primers used for *TLR3* exon sequencing in the candidate gene study of capsular group C meningococci (MenC)-specific immunity following infant immunisation with MenC conjugate vaccine.

Exon	PCR primers	PCR setting	Sequencing primers
1	5′-GTTTTCTCCCTTTGCCCTTC-3′ 5′-GAGACCCTGCCCAGTAAGAA-3′	1	5′-GTTTTCTCCCTTTGCCCTTC-3′ 5′-GAGACCCTGCCCAGTAAGAA-3′
2	5′-GCATTTGAAAGCCATCTGCT-3′ 5′-GAGAAAGCGAGAGAGGCAAA-3′	1	5′-GCATTTGAAAGCCATCTGCT-3′ 5′-GAGAAAGCGAGAGAGGCAAA-3′
3	5′-AGAGGTGTTGATTTTGCTGTTG-3′ 5′-GGAAGGATTGCTGGAAGACA-3′′	1	5′-AGAGGTGTTGATTTTGCTGTTG-3′ 5′-GAAGGATTGCTGGAAGACA-3′
4 (Part1)	5′-GAAAGTGTGCCGTCACCG-3′ 5′-CCTGTGAGTTCTTGCCCAAT-3′	2	5′-GAAAGTGTGCCGTCACCG-3′′ 5′-GGCGGATCATGAGGTCAG-3′ 5′-GTTGTATTGCTGGTGGTGGA-3′ 5′-TCCACCACCAGCAATACAAC-3′ 5′-CCTGTGAGTTCTTGCCCAAT-3′
4 (Part2)	5′-GCTCATTCTCCCTTACACAT-3′ 5′-GGTTTTTAACAGCCAGAAGCA-3′	2	5′-GCTCATTCTCCCTTACACAT-3′ 5′- ACCCTGGTGGTCCCATTTAT-3′ 5′-AAGAGTTCAAAGGGGGCACT-3′ 5′-AGTGCCCCCTTTGAACTCTT-3′ 5′-GGTTTTTAACAGCCAGAAGCA-3′
5	5′-CTGGCCTCAGGTGAGCTTC-3′ 5′-CAAATGTTTTCTGCCCTTGG-3′	2	5′-CTGGCCTCAGGTGAGCTTC-3′ 5′-CAAATGTTTTCTGCCCTTGG-3′

PCR setting1: Step1 = 1X Denature 95 °C 5mins, Step2 = 29X (95 °C 30s , 58 °C 30s, 60 °C 1min) Step3= 1X 60 °C 10mins

PCR setting2: Step1 = 1X Denature 95 °C 5mins, Step2 = 29X (95 °C 30s , 57 °C 30s, 59 °C 3min) Step3= 1X 60 °C 10mins

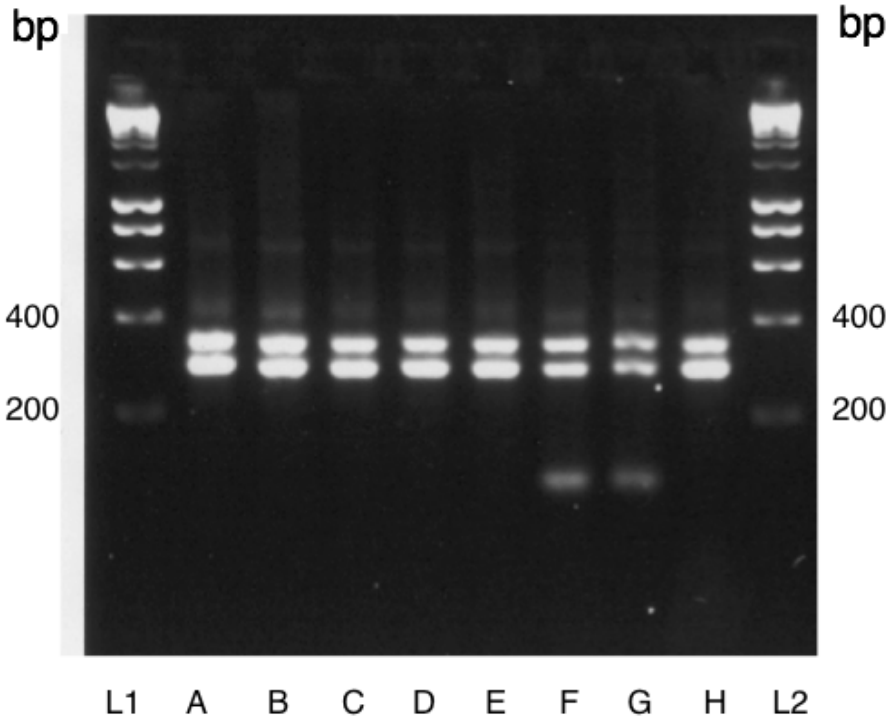


Figure 7.2: A representative agarose gel used for amplification-refractory mutation system PCR genotype calling, in the study assessing the relationship between *TLR3* and *CD44* single-nucleotide polymorphisms, and the magnitude of capsular group C meningococci (MenC)-specific serological immunity following infant immunisation with MenC conjugate vaccine. bp = base pairs, L1 & L2 = DNA ladder, lanes A-E & H are homozygous and lanes F & G are heterozygous at this locus.

Table 7.4: Results of genotypic and allelic association analysis between *CD44* single-nucleotide polymorphisms (SNPs) and specific-IgG concentrations to capsular group C meningococcal (MenC) polysaccharide 1 month after infant immunisation with MenC conjugate vaccine.

SNP	Allele	n	GMC	CI(95%)	p-value
rs10734436	A/A	15	2.09	0.89-4.89	0.2102
	G/A	59	4.31	3.1-6	0.8424
	G/G	51	3.68	2.48-5.47	-
	Allele 'A'				0.2970
rs11033008	C/C	8	3.25	0.67-15.6	0.2308
	C/T	38	2.78	1.78-4.34	0.9524
	T/T	83	4.05	3.03-5.41	-
	Allele 'C'				0.3461
rs11033010	A/A	17	2.97	1.34-6.55	0.4964
	G/A	50	3.49	2.38-5.11	0.3239
	G/G	62	3.84	2.71-5.43	-
	Allele 'A'				0.8734
rs11033021	T/T	18	3.08	1.69-5.63	0.5674
	T/C	55	3.71	2.48-5.56	0.5763
	C/C	56	3.61	2.54-5.14	-
	Allele 'T'				0.4997
rs112762	C/C	32	3.75	2.34-6.01	0.5523
	T/C	51	3.12	2.13-4.57	0.2945
	T/T	43	4.33	2.76-6.8	-
	Allele 'C'				0.5007
rs12419062	G/G	14	1.81	0.73-4.49	0.3110

Continued on next page

7.2 Appendix: Candidate gene studies of immunological responses to childhood vaccines

SNP	Allele	n	GMC	CI(95%)	p-value
	G/A	59	4.41	3.12-6.23	0.2205
	A/A	56	3.4	2.38-4.86	-
	Allele 'G'				0.8793
rs2785172	G/G	22	2.63	1.5-4.6	0.0945
	G/A	45	3.05	2.01-4.61	0.5642
	A/A	62	4.47	3.14-6.38	-
	Allele 'G'				0.1084
rs353615	T/T	20	3.43	1.75-6.72	0.8753
	T/C	45	4.36	2.82-6.74	0.3157
	C/C	64	3.15	2.27-4.36	-
	Allele 'T'				0.8188
rs353619	A/A	6	2.07	0.28-15.2	0.1205
	G/A	42	3.23	2.09-5	0.8854
	G/G	81	3.92	2.92-5.26	-
	Allele 'A'				0.3972
rs353626	C/C	8	4.25	1.62-11.2	0.9403
	C/T	47	3.64	2.34-5.64	0.7940
	T/T	74	3.47	2.55-4.72	-
	Allele 'C'				0.9039
rs353639	G/G	12	3.82	1.4-10.4	0.8064
	G/T	48	4.47	2.97-6.7	0.0594
	T/T	69	3.03	2.2-4.15	-
	Allele 'G'				0.2382
rs7105890	A/A	4	0.774	0.09-6.6	0.3826
	G/A	47	4.4	3.02-6.41	0.5869
	G/G	75	3.55	2.61-4.84	-
	Allele 'A'				0.9865
rs7945310	T/T	17	4.93	2.23-10.9	0.1975
	C/T	60	3.31	2.28-4.79	0.6377
	C/C	52	3.52	2.48-4.99	-
	Allele 'T'				0.2291
rs7947433	G/G	31	3.04	1.76-5.25	0.8138
	G/A	66	3.35	2.39-4.7	0.3314
	A/A	32	4.78	3.04-7.5	-
	Allele 'G'				0.8049
rs996076	A/A	31	3.27	2.1-5.11	0.5225
	A/G	59	4.39	2.99-6.44	0.6751
	G/G	38	2.96	1.93-4.54	-
	Allele 'A'				0.5712
	All	135	3.46	2.75-4.37	

GMC= geometric mean concentration; CI = confidence interval
* = significant, with a P-value less than 0.05

Table 7.5: Results of genotypic and allelic association analysis between *TLR3* exonic variants and serum bactericidal titres against capsular group C meningococci (MenC) 1 month after infant immunisation with MenC conjugate vaccine.

Exonic variant	Allele	n	GMT	CI (95%)	P-value
rs3775290	C/C	149	135	103 -176	-
	C/T	126	127	96.2-169	0.929
	T/T	32	115	71.3-186	0.421
	Allelic 'C'				0.582
rs3775291	C/C	162	131	103-168	-
	C/T	119	115	86.6-153	0.562
	T/T	23	211	92.6-480	0.153
	Allelic 'G'				0.533
rs3775296†	C/C	101	246	172-352	-
	C/T	25	243	135-438	0.9022
	T/T	2	89.3	2.32-3433	0.4861
	Allelic 'C'				0.8106
rs73873710†	C/C	125	237	174-322	-
	C/T	3	567	126-2547	0.3886
	Allelic 'C'				0.3886
rs116729895†	C/C	125	237	174-322	-
	C/T	3	567	126-2547	0.3886
	Allelic 'C'				0.3886
	All	318	133	111-159	

GMT= geometric mean titre; CI = confidence interval
† = Sequenced in subset

Table 7.6: The most statistically significant *TLR3* haplotypes associated with serum bactericidal titres against capsular group C meningococci (MenC) 1 month after infant immunisation with MenC conjugate vaccine.

SNPs	Haplotypes	First haplotype SNP	Last haplotype SNP	Most significant haplotype	Beta	p-value
1	2	rs11732384	rs11732384	A	-0.029	0.712
1	2	rs7657186	rs7657186	G	-0.007	0.919
1	2	rs13126816	rs13126816	G	0.003	0.968
1	2	rs5743312	rs5743312	C	-0.054	0.499
1	2	rs7668666	rs7668666	A	-0.034	0.606
1	2	rs3775292	rs3775292	G	-0.058	0.407
1	2	rs3775291	rs3775291	C	-0.031	0.624
1	2	rs3775290	rs3775290	C	0.039	0.517
2	3	rs11732384	rs7657186	AG	-0.029	0.720
2	3	rs7657186	rs13126816	AG	0.007	0.919
2	3	rs13126816	rs5743312	GT	0.078	0.356
2	4	rs5743312	rs7668666	CA	-0.104	0.206
2	4	rs7668666	rs3775292	CC	0.067	0.385
2	4	rs3775292	rs3775291	CT	0.258	0.230
2	3	rs3775291	rs3775290	CT	-0.045	0.463
3	4	rs11732384	rs13126816	AGG	-0.031	0.700
3	4	rs7657186	rs5743312	GGT	0.084	0.311
3	5	rs13126816	rs7668666	GCA	-0.100	0.217
3	6	rs5743312	rs3775292	CAG	-0.109	0.188
3	6	rs7668666	rs3775291	CCT	0.239	0.262
3	6	rs3775292	rs3775290	CTC	0.154	0.450
4	5	rs11732384	rs5743312	GGGT	0.081	0.335
4	6	rs7657186	rs7668666	GGCA	-0.104	0.197
4	8	rs13126816	rs3775292	GCAG	-0.105	0.203
4	8	rs5743312	rs3775291	CAGC	-0.105	0.215
4	7	rs7668666	rs3775290	AGCT	-0.054	0.442
5	7	rs11732384	rs7668666	GGGCA	-0.101	0.213
5	11	rs7657186	rs3775292	GGCAG	-0.097	0.252
5	11	rs13126816	rs3775291	GCCCT	0.674	**0.040***
5	9	rs5743312	rs3775290	CAGCT	-0.131	0.118

SNP= Single-nucleotide polymorphism, Beta= beta coefficient

7.2.2 Functional assessment of exonic *TLR3* polymorphisms

Table 7.7: Quantitative PCR primer sets used in the functional evaluation of *TLR3*.

Gene primer	Sequence	Reference
RN18S1 RNA forward	TCAAGAACGAAAGTCGGAGG	Designed using Primer3 (`http://primer3.ut.ee`)
RN18S1 RNA reverse	GGACATCTAAGGGATCACA	Designed using Primer3 (`http://primer3.ut.ee`)
ACTB forward	CCACGAAACTACCTTCAACTCC	Designed using Primer3 (`http://primer3.ut.ee`)
ACTB reverse	TCATACTCCTGCTGCTTGCTGATCC	Designed using Primer3 (`http://primer3.ut.ee`)
IL6 forward	GGTACATCCTCGACGGCATCT	(420)
IL6 reverse	GCCTCTTTGCTGCTTTCAC	(420)
CXCL10 forwards	GGCATTCAAGGAGTACCTCTCTC	(420)
CXCL10 reverse	GACAAAATTGGCTTGCAGGA	(420)
TNF forward	CGAGTGACAAGCCTGTAGC	(421)
TNF reverse	GGTGTGGGTGAGGAGCACAT	(421)
CD86 forward	TATGGGCCGCACAAGTTTT	(420)
CD86 reverse	TCCTGTGGGCTTTTTGTGAT	(420)
CD83 forward	CCCCAATGAAAGGCCCTATT	(420)
CD83 reverse	TTAGGTTCTCTGCCCATCCG	(420)
IFNA1 forward	GTGAGGAAATACTTCCAAAGAATCAC	Designed using Primer 3 (`http://primer3.ut.ee`)
IFNA1 reverse	TCTCATGATTTCTGCTCTGACAA	Designed using Primer3 (`http://primer3.ut.ee`)
IFNB1 forward	AGCTGCAGCAGTTCCAGAAG	Designed using Primer 3 (`http://primer3.ut.ee`)
IFNB1 reverse	AGTCTCATTCCAGCCAGTGC	Designed using Primer3 (`http://primer3.ut.ee`)
TLR3 forward	ACAACTTAGCACGGCTCTGGA	(420)
TLR3 reverse	ACCTCAACTGGGATCTCGTCA	(420)
TRIF forward	AGCGCCTTCGACATTCTAGGT	(420)
TRIF reverse	AGAACCATGGCATGCAGGA	(420)

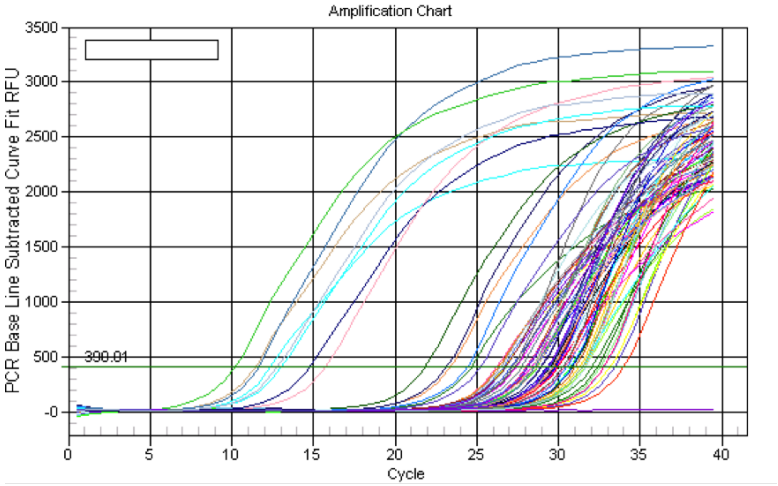


Figure 7.3: Amplification plots for the 11 quantitative PCR (QPCR) primer sets used in the functional evaluation of exonic *TLR3* polymorphisms. This graphic displays the 11 QPCR primer sets assessed on four samples performed in duplicate. Fluorescence intensity increases with the incorporation of SYBR® green with number of PCR cycles, read using an iQ5 real-time PCR machine (Bio-Rad®). The colours of the curves were randomly assigned, from those available in the iQ5 software palette, such that shared colours do not necessarily represent duplicates. The non-amplifying curves are cDNA-less controls.

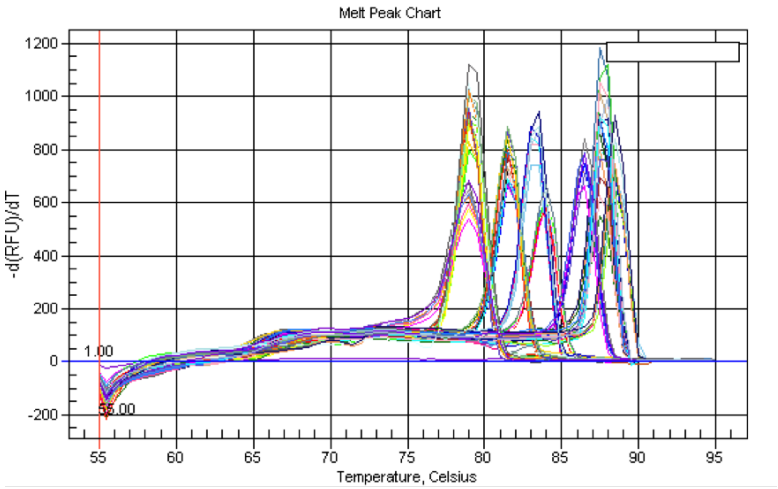


Figure 7.4: Melting curve plots for the 11 quantitative PCR (QPCR) primer sets used in the functional evaluation of exonic *TLR3* polymorphisms. This graphic displays the 11 QPCR primer sets assessed on four samples performed in duplicate, read using an iQ5 real-time PCR machine (Bio-Rad®). The colours of the curves were randomly assigned, from those available in the iQ5 software palette, such that shared colours do not necessarily represent duplicates. The non-amplifying curves are cDNA-less controls.

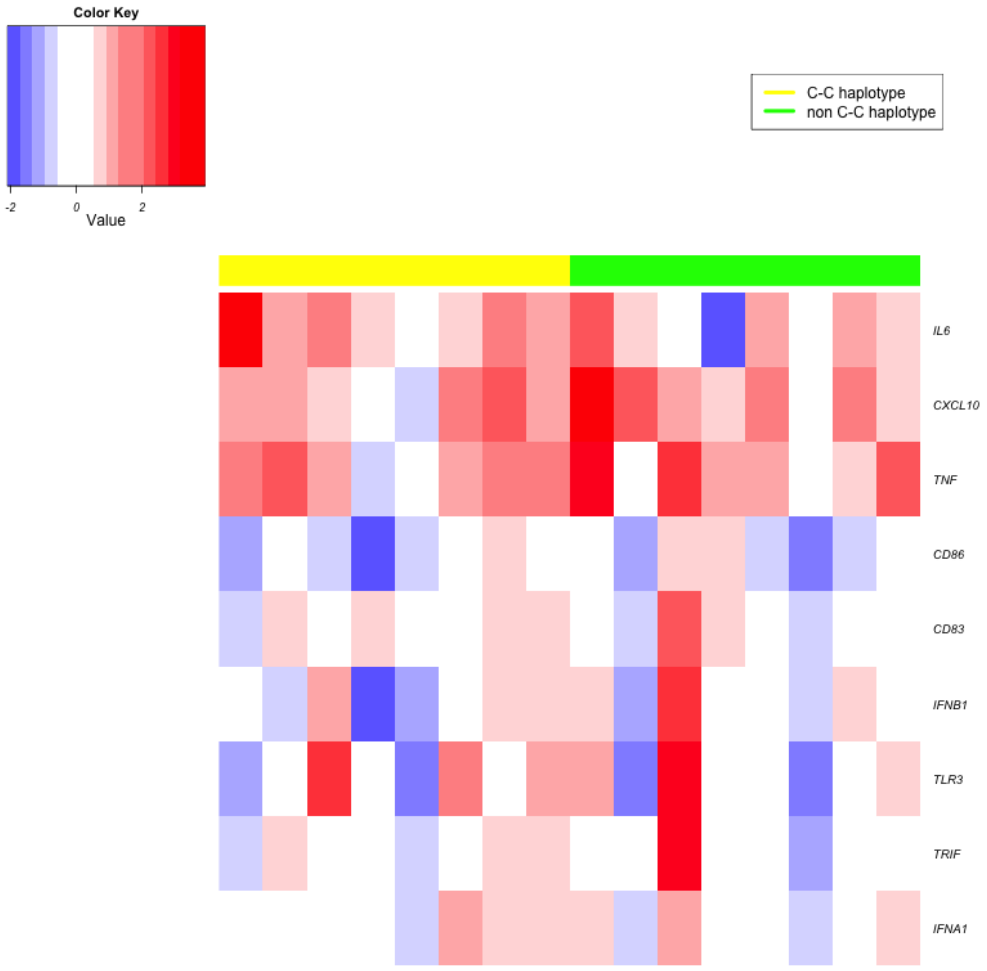


Figure 7.5: Heat map of gene expression following 6 hours of stimulation with Poly(I:C) compared to media alone, normalised by *ACTB* (beta-actin rRNA). Each column is a participant and each row is a gene. Columns are arranged to separate samples based upon the rs3775291-rs3775290 haplotype. The yellow columns refer to the C-C haplotype group (n=8) and the green columns to the non C-C haplotype group (n=8). The coloured boxes refer to the $\log_2$ normalised $\Delta\Delta Ct$ values, scaled from blue for downregulated genes, through white for no change, to red for upregulated genes.

Table 7.8: The relative fold-change in gene expression determined using ΔΔCt method (equation 2.2) for each of the nine immune genes assessed following 6 hours poly(I:C) treatment, normalised by *RN18S1* expression.

Gene	1CC	2CC	3CC	4CC	5CC	6CC	7CC	8CC	1NCC	2NCC	3NCC	4NCC	5NCC	6NCC	7NCC	8NCC	ΔmedianΔΔCt	p-value
IL6	12.3	4.0	9.6	1.1	1.0	1.0	1.6	1.4	17.3	4.8	4.4	0.8	2.4	1.6	2.3	1.2	-2.874	0.889
CXCL10	3.9	4.0	5.3	0.7	0.8	1.9	1.9	1.4	48.5	13.2	8.2	5.6	3.4	1.4	3.0	1.2	-8.615	0.194
TNF	2.7	9.3	5.5	0.5	1.8	1.4	1.4	2.1	29.9	3.4	19.7	8.1	2.1	1.3	1.8	2.9	-6.685	0.188
CD86	0.8	0.5	1.5	0.2	0.7	0.5	0.9	0.8	1.9	1.8	1.3	1.4	0.5	0.5	0.7	0.7	-0.376	0.158
CD83	1.2	0.9	2.2	1.2	1.2	0.8	0.9	1.2	2.9	2.5	3.6	1.1	1.2	0.9	1.2	0.9	-0.620	0.165
IFNB1	1.6	0.4	5.7	0.2	0.6	0.6	0.8	1.0	4.3	1.9	5.4	0.9	0.7	0.9	1.5	0.8	-1.371	0.437
TLR3	0.7	0.7	19.3	0.6	0.5	2.0	0.5	1.3	6.4	1.3	7.4	0.7	0.8	0.5	1.1	1.3	-1.725	0.769
TRIF	0.9	0.9	2.3	0.9	0.9	0.6	0.9	1.0	2.2	3.0	7.1	1.0	1.0	0.8	1.1	1.0	-1.252	0.201
IFNA1	2.3	0.5	2.6	0.6	0.7	1.4	0.8	1.0	3.9	2.6	2.0	1.0	0.7	1.1	1.0	1.2	-0.758	0.371

ΔmeanΔΔCt = The median of gene ΔΔCt of C-C haplotype individuals minus the ΔΔCt of non C-C haplotype individuals, p-values were calculated from a t-test comparing the mean gene ΔΔCt of C-C haplotype and non C-C haplotype individuals.
CC = C-C rs3775291-rs3775290 haplotype samples, NCC= non C-C rs3775291-rs3775290 haplotype samples

Table 7.9: The relative fold-change in gene expression determined using ΔΔCt method (equation 2.2) for each of the nine immune genes assessed following 6 hours poly(I:C) treatment, normalised by *ACTB* expression.

Gene	1CC	2CC	3CC	4CC	5CC	6CC	7CC	8CC	1NCC	2NCC	3NCC	4NCC	5NCC	6NCC	7NCC	8NCC	ΔmedianΔΔCt	P-value
IL6	15.1	2.1	3.5	1.5	0.8	1.7	3.2	2.2	4.9	1.7	1.2	0.3	2.6	1.0	2.3	1.6	0.205	0.321
CXCL10	2.3	2.1	1.9	0.9	0.6	3.2	4.0	2.2	13.7	4.7	2.3	1.7	3.8	0.8	3.0	1.7	-1.821	0.266
TNF	3.3	4.9	2.0	0.7	1.4	2.4	2.9	3.4	8.5	1.2	5.5	2.5	2.3	0.8	1.9	4.1	-0.719	0.493
CD86	0.5	0.8	0.5	0.3	0.5	0.8	1.9	1.2	0.8	0.5	1.5	1.9	0.5	0.3	0.7	1.0	-0.241	0.736
CD83	0.7	1.5	0.8	1.7	0.9	1.3	1.8	1.8	1.1	0.6	4.4	1.5	0.9	0.6	1.2	1.3	-0.063	0.750
IFNB1	0.9	0.7	2.1	0.2	0.4	1.1	1.5	1.6	1.7	0.5	6.6	1.2	0.8	0.6	1.5	1.2	-0.776	0.376
TLR3	0.4	1.1	7.1	0.9	0.4	3.2	1.0	2.0	2.6	0.3	9.1	0.9	0.9	0.3	1.1	1.8	-1.074	0.933
TRIF	0.5	1.4	0.8	1.2	0.7	1.0	1.8	1.6	0.9	0.8	8.8	1.3	1.1	0.5	1.1	1.4	-0.847	0.435
IFNA1	1.3	0.9	0.9	0.9	0.5	2.3	1.7	1.6	1.5	0.7	2.5	1.4	0.8	0.6	1.0	1.6	-0.148	0.997

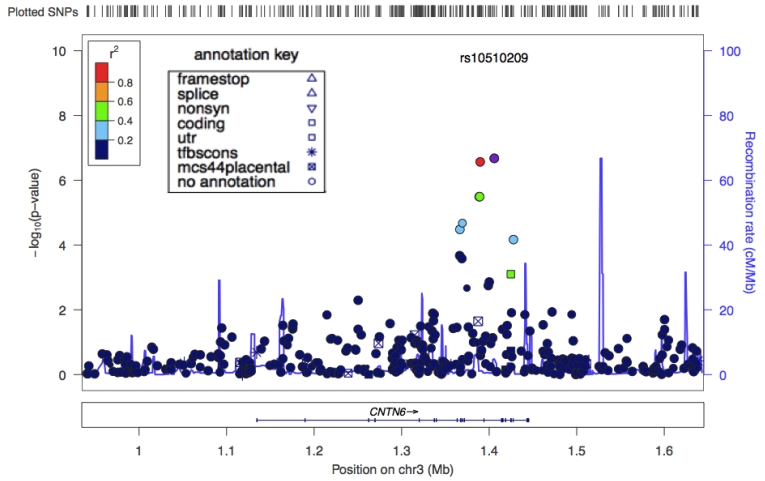
ΔmeanΔΔCt = The median of gene ΔΔCt of C-C haplotype individuals minus the ΔΔCt of non C-C haplotype individuals, p-values were calculated from a t-test comparing the mean gene ΔΔCt of C-C haplotype and non C-C haplotype individuals.
CC = C-C rs3775291-rs3775290 haplotype samples, NCC= non C-C rs3775291-rs3775290 haplotype samples

7.2.3 Candidate gene study of the genetic determinants of childhood booster responses to pertussis vaccine

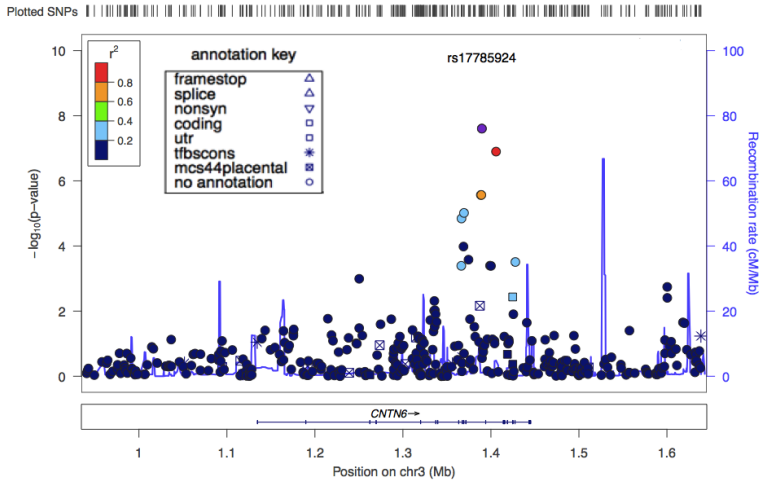
Table 7.10: Primers and PCR settings for the amplification-refractory mutation system PCR genotyping of rs4986790, for analysis between this single-nucleotide polymorphism (SNP) and serological responses to a booster dose of pertussis vaccine.

SNP	Inner primer forward	Inner reverse	Outer primer forward	Outer primer reverse
rs4986790	CATACTTAGACT-	GTCAAACAATTA-	AGTCCATCGTTT-	AAGCATTCCCAC-
Sequence (5′-3′)	-ACTACCTCGATTA	-AATAAGTCAATAATGC	-GGTTCTGG	-CTTTGTTG

7.3 Appendix: Genome-wide association study of capsular group C meningococcal vaccine immunity

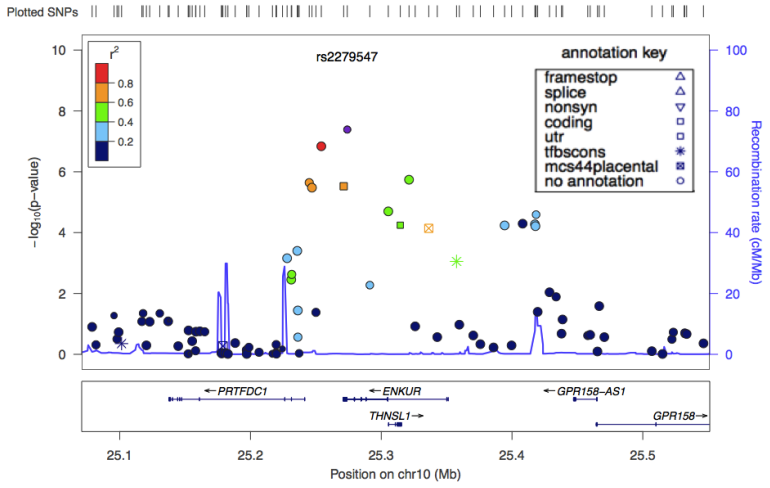


(a)

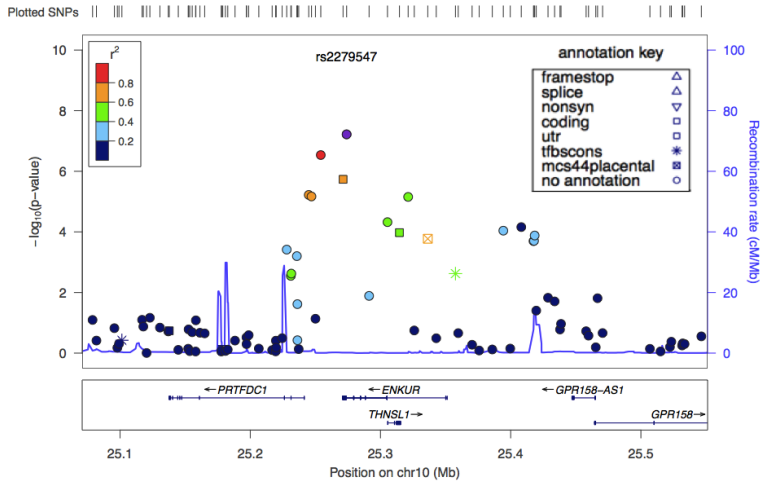


(b)

Figure 7.6: Regional plot for the locus on chromosome 3 associated with the persistence of capsular group C meningococci (MenC)-specific serum bactericidal titres, following older childhood immunisation with MenC conjugate vaccine. The local gene structure is shown below the plot, with thick lines denoting exons joined by intronic regions. The statistical association (p-value) is shown on the primary y-axis (left) and the recombination rate is shown on the secondary y-axis (right). The single-nucleotide polymorphism (SNP) with the most statistical evidence of association is shown in purple and remaining SNPs are coloured according to their linkage disequilibrium (r^2) with this 'lead' SNP. SNPs with particular features are assigned a symbol, as denoted by the 'annotation key'. a) linear regression analysis with complement as covariate, b) meta-analysis of regression analyses conducted separately for SBA using rabbit and human complement. These plots were created using the LocusZoom tool (244).



(a)



(b)

Figure 7.7: Regional plot for the locus on chromosome 10 associated with the persistence of capsular group C meningococci (MenC)-specific serum bactericidal titres, following older childhood immunisation with MenC vaccine. The local gene structure is shown below the plot, with thick lines denoting exons joined by intronic regions. The statistical association (p-value) is shown on the primary y-axis (left) and the recombination rate is shown on the secondary y-axis (right). The single-nucleotide polymorphism (SNP) with the most statistical evidence of association is shown in purple and remaining SNPs are coloured according to their linkage disequilibrium (r^2) with this 'lead' SNP. SNPs with particular features are assigned a symbol, as denoted by the 'annotation key'. a) linear regression analysis with complement as covariate, b) meta-analysis of regression analyses conducted separately for SBA using rabbit and human complement. These plots were created using the LocusZoom tool (244).

7.3.1 Genome-wide association study of the persistence of serological immunity to capsular group C meningococcal vaccine, following childhood immunisation - Stage 1

7.3.2 Power calculations

Power was computed for the two-stage association study design using the R software package 'QpowR' (285). Power calculations were based upon an additive linear fixed effect model (equation 7.1)

$$y_i = \mu + b \cdot x_i + e_i \tag{7.1}$$

Where x takes the value -1, 0 or 1 for minor homozygote, heterzygote and major homozygote genotypes, respectively. Assuming an asymptotic test, a z-test of an association between the SNP genotype and the phenotype under the null hypothesis $z \sim \Phi_{0,1}$, a standard normal distribution, can be approximated. Under the alternative hypothesis, the test statistic follows a normal distribution with mean λ.

$$\lambda = \sqrt{N \frac{h^2}{1 - h^2}} \tag{7.2}$$

$$h^2 = \frac{V_b}{V_b + \sigma_e^2} \tag{7.3}$$

Where h^2 = narrow sense heritability, V_b = variance due to the additive effects of the alleles, σ_e^2 = residual phenotypical variance.

The critical value (C) and the power $(1 - \beta)$ of the z-test are computed as follows.

$$C = \pm\Phi^{-1}(1 - \frac{\alpha_m}{2}) \tag{7.4}$$

$$1 - \beta = 1 - \Phi(C - \lambda) + \Phi(-C - \lambda) \tag{7.5}$$

Where Φ is the normal cumulative density function with mean m and variance V. Φ^{-1} is the inverse-normal cumulative density function and α_m is the experiment-wise error rate. In a two-stage design, samples are split into $N_1 = \pi_s N$ and $N_2 = (1 - \pi_s)N$, where N = total sample size and π_s =proportion of samples typed in the first stage. A combined statistic for significance can be calculated as follows.

$$z_j = \sqrt{\pi_s z_1} + \sqrt{(1 - \pi_s) z_2} \tag{7.6}$$

Conditional on the first stage statistic this forms a Gaussian distribution. The tail area probability can be calculated using equation 7.7, equated to α_m and solved for C_j assuming the null distribution for both test statistics. Power is obtained by computing the integrals with the obtained C_j value and assuming the distribution under the alternate hypothesis.

$$\begin{aligned}
&P(|z_j| > C_j; |z_1| > C_1) \\
&= \int_{-\infty}^{-C_j} \left(\left(1 - \Phi_{z_2}\Big(\frac{C_j - \sqrt{\pi_s x_1}}{\sqrt{1 - \pi_s}} \Big) + \Phi_{z_2}\Big(\frac{-C_j - \sqrt{\pi_s x_1}}{\sqrt{1 - \pi_s}} \Big) \right) \varphi_{z_1}(x_1) dx_1 \right) \\
&+ \int_{C_j}^{\infty} \left(\left(1 - \Phi_{z_2}\Big(\frac{C_j - \sqrt{\pi_s x_1}}{\sqrt{1 - \pi_s}} \Big) + \Phi_{z_2}\Big(\frac{-C_j - \sqrt{\pi_s x_1}}{\sqrt{1 - \pi_s}} \Big) \right) \varphi_{z_1}(x_1) dx_1 \right)
\end{aligned} \tag{7.7}$$

7.3.3 Results

Table 7.11: Single-nucleotide polymorphisms (SNPs) with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following older childhood immunisation with MenC conjugate vaccine. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	Position	SNP	SNP class	p-value	Gene	Neighbouring genes (<100kb)
1	48061318	rs2457077		8.2273e-06		*CYP46A4P,TRABD2B,SKINTL*
2	38977371	rs3112183	intron-variant	3.5327e-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	38977402	rs3134630	intron-variant	3.4197e-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	38977505	rs3134629	intron-variant	3.5713e-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	38977612	rs3134628	intron-variant	2.6117e-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	38978764	rs12621103	upstream-2KB	3.5681e-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	38980231	rs13024811	upstream-2KB	2.7364e-06	*SRSF7*	*ARHGEF33,LOC375196,SOS1,MORN2*
2	38981922	rs12474770		5.5983e-06		*ARHGEF33,LOC375196,SOS1,MORN2*
2	38982036	rs4670892		7.0077e-06		*ARHGEF33,LOC375196,SOS1,MORN2*
2	38985347	rs11685192		5.0409e-06		*ARHGEF33,LOC375196,SOS1*
2	38994000	rs3112186		7.7985e-06		*ARHGEF33,LOC375196,SOS1*
2	39004215	rs3099952	upstream-2KB	6.686e-06	*GEMIN6*	*ARHGEF33,LOC375196,SOS1*
2	39007822	rs4670894	intron-variant	4.3582e-06	*GEMIN6*	*ARHGEF33,LOC375196,SOS1*
2	39007909	rs3099960	intron-variant	5.5867e-06	*GEMIN6*	*ARHGEF33,LOC375196,SOS1*
2	39008468	rs1562855	intron-variant	8.8491e-06	*GEMIN6*	*ARHGEF33,LOC375196,SOS1*
2	39016966	rs12468732		5.2632e-06		*ARHGEF33,LOC375196,SOS1*
2	39030302	rs12712621	intron-variant	9.7491e-06	*DHX57*	*ARHGEF33,LOC375196,SOS1*
2	39037813	rs3112215	intron-variant	3.6167e-06	*DHX57*	*ARHGEF33,LOC375196,SOS1*
2	39104425	rs9808133	intron-variant	7.6547e-06	*MORN2,DHX57*	*CDKL4,SOS1*
2	39112166	rs11124653		6.3129e-06		*CDKL4,SOS1*
3	185308823	rs111307985	intron-variant	4.611e-06	*SENP2*	*MIR5588,EHHADH,EIF2S2P2,MAP3K13*
5	25717507	rs400196		8.0033e-06		
5	89572699	rs60357858		2.2521e-06		
10	114557647	rs4554813	intron-variant	4.623e-06	*VTI1A*	*ABLIM1*
10	114558538	rs941825	intron-variant	4.2773e-06	*VTI1A*	*ABLIM1*
10	114562675	rs4633389	intron-variant	6.5842e-06	*VTI1A*	*ABLIM1*
10	129015164	rs61874056	intron-variant	6.6262e-06	*DOCK1*	
17	60903397	rs6504129		8.9615e-06		*KRT18P61,HMGN1P28,BCAS3*

Chr= chromosome

Table 7.12: Single-nucleotide polymorphisms (SNPs), in the imputed dataset, with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay titres following older childhood immunisation with MenC conjugate vaccine, when SBA complement source was included as a covariate. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	Position	SNP	SNP class	p-value	Gene	Neighbouring genes (<100kb)
1	7014444	rs7554752	intron-variant	1.68e-06	*CAMTA1*	*LOC100129476,CAMTA1*
1	7054979	rs4400633	intron-variant	8.54e-06	*CAMTA1*	*LOC100129476,CAMTA1*
2	8914351	rs7608773	intron-variant	3.42e-06	*KIDINS220*	*MBOAT2,KIDINS220*
2	38949275	rs67716743	intron-variant	3.79e-06	*GALM*	*ARHGEF33,SOS1,MORN2,DHX57*
2	38954419	rs11124646	intron-variant	4.46e-06	*GALM*	*ARHGEF33,SOS1,MORN2,DHX57*
2	38964531	rs3097713		2.71e-06		*ARHGEF33,SOS1,MORN2,DHX57*
2	38977371	rs3112183	intron-variant	4.74e-07	*SRSF7*	*ARHGEF33,SOS1,MORN2*
2	38977402	rs3134630	intron-variant	5.1e-07	*SRSF7*	*ARHGEF33,SOS1,MORN2*
2	38977505	rs3134629	intron-variant	5.18e-07	*SRSF7*	*ARHGEF33,SOS1,MORN2*
2	38977612	rs3134628	intron-variant	4.35e-07	*SRSF7*	*ARHGEF33,SOS1,MORN2*
2	38978764	rs12621103	upstream-2KB	3.97e-07	*SRSF7*	*ARHGEF33,SOS1,MORN2*
2	38979627	rs12472454	upstream-2KB	7.88e-07	*SRSF7*	*ARHGEF33,SOS1,MORN2*
2	38980231	rs13024811	upstream-2KB	3.52e-07	*SRSF7*	*ARHGEF33,SOS1,MORN2*
2	38980922	rs2037875		1.76e-06		*ARHGEF33,SOS1,MORN2*
2	38981207	rs10197412		1.7e-06		*ARHGEF33,SOS1,MORN2*
2	38981922	rs12474770		7.42e-07		*ARHGEF33,SOS1,MORN2*
2	38982036	rs4670892		8.69e-07		*ARHGEF33,SOS1,MORN2*
2	38985347	rs11685192		3.95e-07		*ARHGEF33,SOS1*
2	38986773	rs11124647		9.27e-06		*ARHGEF33,SOS1*

Continued on next page

Chr	Position	SNP	SNP class	p-value	Gene	Neighbouring genes (<100kb)
2	38987617	rs6715866		9.35e-06		*ARHGEF33,SOS1*
2	38989961	rs12991609		9.72e-06		*ARHGEF33,SOS1*
2	38990460	rs4630748		6.63e-06		*ARHGEF33,SOS1*
2	38991956	rs10865143		9.75e-06		*ARHGEF33,SOS1*
2	39004215	rs3099952	upstream-2KB	1.18e-06	*GEMIN6*	*ARHGEF33,SOS1*
2	39007822	rs4670894	intron-variant	3.47e-06	*GEMIN6*	*ARHGEF33,SOS1*
2	39007909	rs3099960	intron-variant	1.16e-06	*GEMIN6*	*ARHGEF33,SOS1*
2	39008468	rs1562855	intron-variant	2.08e-06	*GEMIN6*	*ARHGEF33,SOS1*
2	39023490	rs3112209		1.84e-06		*ARHGEF33,SOS1*
2	39037813	rs3112215	intron-variant	9.1e-07	*DHX57*	*ARHGEF33,SOS1*
2	39038725	rs3112216	intron-variant	9.87e-06	*DHX57*	*ARHGEF33,SOS1*
2	39099918	rs2373473	intron-variant	5.79e-06	*DHX57*	*CDKL4,SOS1*
2	39100760	rs3099945	intron-variant	9.36e-06	*DHX57*	*CDKL4,SOS1*
2	39101105	rs3112170	intron-variant	9.26e-06	*MORN2,DHX57*	*CDKL4,SOS1*
2	39101827	rs3112172	intron-variant	5.1e-06	*MORN2,DHX57*	*CDKL4,SOS1*
2	39101885	rs2060743	intron-variant	7.69e-06	*MORN2,DHX57*	*CDKL4,SOS1*
2	39104425	rs9808133	intron-variant	9.14e-06	*MORN2,DHX57*	*CDKL4,SOS1*
2	39112166	rs11124653		3.92e-06		*CDKL4,SOS1*
2	39129935	rs11681583		7.42e-06		*CDKL4,SOS1*
2	39139916	rs1868867		8.32e-06		*CDKL4,SOS1*
2	39150901	rs10204999	intron-variant	9.08e-06	*ARHGEF33*	*CDKL4,SOS1,MAP4K3*
2	156982876	rs114516289		7.43e-07		
2	156989907	rs115836001		2.72e-06		
3	1385724	rs155400	intron-variant	5.13e-06	*CNTN6*	*CNTN6*
3	1387043	rs55957631	intron-variant	8.53e-06	*CNTN6*	*CNTN6*
3	1387057	rs61456930	intron-variant	6.31e-06	*CNTN6*	*CNTN6*
3	1388418	rs11920256	intron-variant	3.26e-07	*CNTN6*	*CNTN6*
3	1388589	rs11928188	intron-variant	3.54e-06	*CNTN6*	*CNTN6*
3	1388596	rs11916754	intron-variant	3.23e-07	*CNTN6*	*CNTN6*
3	1388726	rs11916814	intron-variant	3.54e-06	*CNTN6*	*CNTN6*
3	1389117	rs17038032	intron-variant	3.54e-06	*CNTN6*	*CNTN6*
3	1389584	rs17785918	intron-variant	3.25e-07	*CNTN6*	*CNTN6*
3	1389720	rs17785924	intron-variant	3.26e-07	*CNTN6*	*CNTN6*
3	1390329	rs66502721	intron-variant	3.72e-06	*CNTN6*	*CNTN6*
3	1390626	rs17785930	intron-variant	3.61e-07	*CNTN6*	*CNTN6*
3	1391139	rs4684123	intron-variant	1.54e-06	*CNTN6*	*CNTN6*
3	1391159	rs4684124	intron-variant	1.54e-06	*CNTN6*	*CNTN6*
3	1393109	rs73006880	intron-variant	5.57e-06	*CNTN6*	*CNTN6*
3	1405829	rs10510209	intron-variant	2.5e-07	*CNTN6*	*CNTN6*
3	1427566	rs72997748	intron-variant	2.28e-06	*CNTN6*	*CNTN6*
3	1427625	rs9820933	intron-variant	3.47e-06	*CNTN6*	*CNTN6*
3	139750465	rs9874196	intron-variant	7.74e-06	*CLSTN2*	*NMNAT3*
3	139750888	rs9820731	intron-variant	8.19e-06	*CLSTN2*	*NMNAT3*
3	139779455	rs2162212	intron-variant	9.33e-06	*CLSTN2*	
5	36667579	rs10491374	intron-variant	5.89e-06	*SLC1A3*	*RPL39P22,SLC1A3*
6	122561446	rs138125116		1.93e-07		*PKIB,SERINC1*
7	28893972	rs28481569		3.43e-06		*CREB5,TRIL*
7	28895777	rs11772521		4.21e-06		*CPVL,CREB5,TRIL*
7	28897686	rs10276611		5.53e-06		*CPVL,CREB5,TRIL*
7	68036803	rs4718787		1.4e-06		*RP11-358M3.1*
7	68041389	rs75351549		1.41e-06		*RP11-358M3.1*
7	83524245	rs62475844		6.26e-06		*SEMA3E*
8	21471008	rs2010278		3.48e-06		
8	39933560	rs55956871		1.77e-06		*IDO2,ADAM2,IDO1*
9	116967338	rs41306492	intron-variant	6.49e-06	*COL27A1*	*ASTN2*
10	25242947	rs274313	upstream-2KB	4.09e-06	*PRTFDC1*	*GPR158-AS1,RP11-80K21.3,GPR158*
10	25244392	rs274312		6.36e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25244919	rs274311		2.83e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25245208	rs274310		5.83e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25245394	rs182130		4.61e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25245721	rs200730		5.63e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25246496	rs274309		5.83e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25246509	rs274308		5e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25246820	rs274307		1.74e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25246950	rs274306		4.99e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25246980	rs274305		2.43e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25247006	rs274304		4.01e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25247057	rs274303		4.92e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25248540	rs274301		7.38e-07		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25249606	rs612482		1.7e-06		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25252205	rs274289		1.66e-07		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25254151	rs274288		1.87e-07		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25257870	rs2818289		1.51e-07		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25258244	rs2031600		1.58e-07		*GPR158-AS1,RP11-80K21.3,GPR158*
10	25261446	rs274290		7.86e-06		*GPR158-AS1,GPR158*
10	25262946	rs274291		6.53e-06		*GPR158-AS1,GPR158*
10	25264155	rs274292		9.33e-06		*GPR158-AS1,GPR158*
10	25265047	rs274293		1.66e-07		*GPR158-AS1,GPR158*
10	25265078	rs274294		1.65e-07		*GPR158-AS1,GPR158*
10	25265303	rs166670		1.62e-07		*GPR158-AS1,GPR158*
10	25269851	rs2007364		6.85e-08		*GPR158-AS1,GPR158*
10	25270441	rs902903	downstream-500B	1.88e-06	*ENKUR*	*GPR158-AS1,GPR158*

Continued on next page

Chr	Position	SNP	SNP class	p-value	Gene	Neighbouring genes (<100kb)
10	25271270	rs2307047	utr-variant-3′	3.41e-06	*ENKUR*	*GPR158-AS1,GPR158*
10	25273988	rs2279547	intron-variant	7.56e-08	*ENKUR*	*GPR158-AS1,GPR158*
10	25277315	rs4556443	intron-variant	3.27e-07	*ENKUR*	*GPR158*
10	25277917	rs10764515	intron-variant	1.43e-07	*ENKUR*	*GPR158*
10	25283248	rs12784769	intron-variant	4.17e-06	*ENKUR*	*GPR158*
10	25293195	rs10764516	intron-variant	9.83e-07	*ENKUR*	*GPR158*
10	25296379	rs10764517	intron-variant	1.89e-06	*ENKUR*	*GPR158*
10	25297598	rs950609	intron-variant	1.47e-06	*ENKUR*	*GPR158*
10	25300544	rs4748993	intron-variant	1.72e-06	*ENKUR*	*GPR158*
10	25300609	rs4748994	intron-variant	5.78e-07	*ENKUR*	*GPR158*
10	25305139	rs987274	intron-variant	1.22e-06	*ENKUR,THNSL1*	*GPR158*
10	25305991	rs7082131	intron-variant	1.16e-06	*ENKUR,THNSL1*	*GPR158*
10	25313818	rs7086282	synonymous-codon	1.59e-06	*ENKUR,THNSL1*	*GPR158*
10	25316755	rs12356765	intron-variant	1.66e-06	*ENKUR*	*GPR158*
10	25319146	rs7393543	intron-variant	2.33e-06	*ENKUR*	*GPR158*
10	25320223	rs6482467	intron-variant	4.57e-06	*ENKUR*	*GPR158*
10	25321244	rs10764519	intron-variant	2.32e-06	*ENKUR*	*GPR158*
10	25321589	rs11014357	intron-variant	2.37e-06	*ENKUR*	*GPR158*
10	25332461	rs1352982	intron-variant	1.73e-06	*ENKUR*	*GPR158*
10	25334906	rs6482468	intron-variant	2.43e-06	*ENKUR*	*GPR158*
10	25343606	rs2367507	intron-variant	5.7e-06	*ENKUR*	*GPR158*
10	25347926	rs11014367	intron-variant	5.21e-06	*ENKUR*	*GPR158*
10	25348695	rs10828750	intron-variant	5.96e-06	*ENKUR*	*GPR158*
10	25348970	rs10828751	intron-variant	6.22e-06	*ENKUR*	*GPR158*
10	25355952	rs10828752		6.88e-06		*GPR158*
10	25363128	rs11014382		7.32e-06		*GPR158*
10	25364341	rs1907790		7.31e-06		*GPR158*
11	64842300	rs2370165		8.14e-06		*RPS16P6,RASGRP2*
12	78675543	rs12303583		2.62e-06		
12	78677265	rs1008694		2.9e-06		
18	76499543	rs2941772		8.23e-06		*LINC00908,FLJ44313,ZNF516*

Chr= chromosome

Table 7.13: Single-nucleotide polymorphisms (SNPs) with p-values suggestive ($<1\times10^{-5}$) of association with the persistence of capsular group C meningococci (MenC)-specific serum bactericidal assay (SBA) titres following older childhood immunisation with MenC conjugate vaccine, when regression models were conducted separately for SBAs using rabbit and human complement, and meta-analysed. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	Position	SNP	p-value	I^2	Gene	Neighbouring genes (<100kb)
1	6936272	rs7546903	8.7e-06	82.37	*CAMTA1*	*RPL37P9,CAMTA1*
1	6948204	rs8918	7.56e-06	81.95	*CAMTA1*	*RPL37P9,CAMTA1*
1	6975221	rs11120785	6.06e-06	82.79	*CAMTA1*	*RPL37P9,CAMTA1*
1	6979712	rs10864246	6.39e-06	74.67	*CAMTA1*	*RPL37P9,CAMTA1*
1	6982350	rs10864247	2.19e-06	81.77	*CAMTA1*	*RPL37P9,CAMTA1*
1	6987405	rs2412219	2.77e-06	81.19	*CAMTA1*	*RPL37P9,CAMTA1*
1	6988244	rs1567401	2.57e-06	81.06	*CAMTA1*	*RPL37P9,CAMTA1*
1	6989303	rs1828309	3.45e-06	79.70	*CAMTA1*	*RPL37P9,CAMTA1*
1	6990863	rs7519035	2.33e-06	79.61	*CAMTA1*	*RPL37P9,CAMTA1*
1	7013328	rs11120795	9.94e-06	84.47	*CAMTA1*	*LOC100129476,CAMTA1*
1	7014444	rs7554752	5.69e-07	18.02	*CAMTA1*	*LOC100129476,CAMTA1*
2	38949275	rs67716743	9.41e-06	74.06	*GALM*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38954419	rs11124646	6.9e-06	39.04	*GALM*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38964531	rs3097713	5.43e-06	67.54		*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38977371	rs3112183	6.17e-07	0.00	*SRSF7*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38977402	rs3134630	6.64e-07	0.00	*SRSF7*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38977505	rs3134629	6.71e-07	0.00	*SRSF7*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38977612	rs3134628	5.64e-07	0.00	*SRSF7*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38978764	rs12621103	5.13e-07	0.00	*SRSF7*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38979627	rs12472454	1.68e-06	66.17	*SRSF7*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38980231	rs13024811	4.58e-07	0.00	*SRSF7*	*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38980922	rs2037875	3.39e-06	60.23		*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38981207	rs10197412	3.3e-06	60.63		*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38981922	rs12474770	9.2e-07	0.00		*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38982036	rs4670892	1.09e-06	0.00		*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	38985347	rs11685192	5.17e-07	0.00		*GALM,ASS1P2,SRSF7,TTC39DP,GEMIN6,DHX57*
2	39004215	rs3099952	1.72e-06	0.00	*GEMIN6*	*GALM,ASS1P2,SRSF7,TTC39DP,MORN2,GEMIN6,DHX57*
2	39007822	rs4670894	4.6e-06	0.00	*GEMIN6*	*GALM,ASS1P2,SRSF7,TTC39DP,MORN2,GEMIN6,DHX57*
2	39007909	rs3099960	1.75e-06	5.90	*GEMIN6*	*GALM,ASS1P2,SRSF7,TTC39DP,MORN2,GEMIN6,DHX57*
2	39008468	rs1562855	2.93e-06	1.47	*GEMIN6*	*GALM,ASS1P2,SRSF7,TTC39DP,MORN2,GEMIN6,DHX57*
2	39023490	rs3112209	3.19e-06	34.33		*GALM,ASS1P2,SRSF7,TTC39DP,MORN2,GEMIN6,DHX57*

Continued on next page

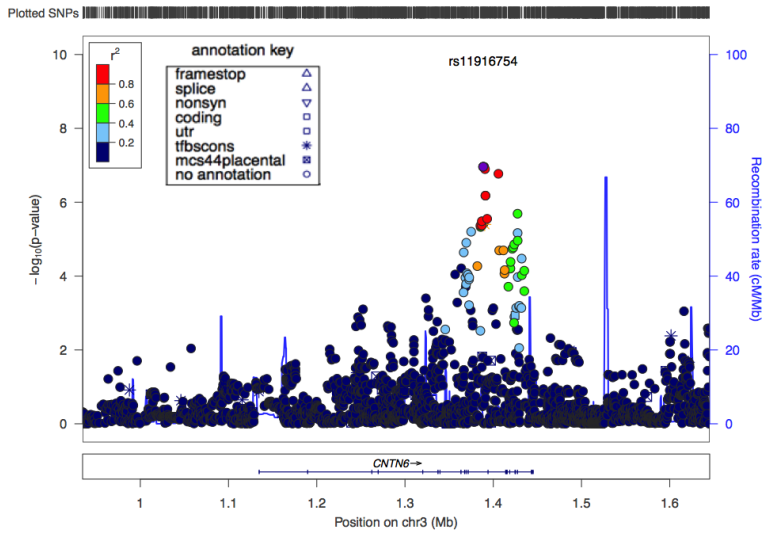
Chr	Position	SNP	p-value	I^2	Gene	Neighbouring genes (<100kb)
2	39037813	rs3112215	1.37e-06	6.82	*DHX57*	*GALM,ASS1P2,SRSF7,TTC39DP,MORN2,GEMIN6,DHX57*
2	39099918	rs2373473	9.14e-06	21.78	*DHX57*	*ARHGEF33,LOC375196,ASS1P2,MORN2,GEMIN6,DHX57*
2	39101827	rs3112172	8.2e-06	23.60	*MORN2*	*ARHGEF33,LOC375196,ASS1P2,MORN2,GEMIN6,DHX57*
2	39112166	rs11124653	3.24e-06	0.00		*ARHGEF33,LOC375196,ASS1P2,SOS1,MORN2,DHX57*
2	156982876	rs114516289	9.42e-07	62.86		*LOC100128759*
2	156989907	rs115836001	2.98e-06	68.23		*LOC100128759*
3	1374976	rs10510170	6.3e-06	0.00	*CNTN6*	*CNTN6*
3	1385477	rs61351715	4.29e-06	0.00	*CNTN6*	*CNTN6*
3	1385724	rs155400	4.68e-06	0.00	*CNTN6*	*CNTN6*
3	1387043	rs55957631	4.25e-06	0.00	*CNTN6*	*CNTN6*
3	1387057	rs61456930	3.26e-06	0.00	*CNTN6*	*CNTN6*
3	1388418	rs11920256	1.1e-07	0.00	*CNTN6*	*CNTN6*
3	1388589	rs11928188	3.74e-06	0.00	*CNTN6*	*CNTN6*
3	1388596	rs11916754	1.1e-07	0.00	*CNTN6*	*CNTN6*
3	1388726	rs11916814	3.74e-06	0.00	*CNTN6*	*CNTN6*
3	1389117	rs17038032	3.74e-06	0.00	*CNTN6*	*CNTN6*
3	1389584	rs17785918	1.11e-07	0.00	*CNTN6*	*CNTN6*
3	1389720	rs17785924	1.11e-07	0.00	*CNTN6*	*CNTN6*
3	1390329	rs66502721	3.99e-06	0.00	*CNTN6*	*CNTN6*
3	1390626	rs17785930	1.27e-07	0.00	*CNTN6*	*CNTN6*
3	1391139	rs4684123	6.63e-07	0.00	*CNTN6*	*CNTN6*
3	1391159	rs4684124	6.64e-07	0.00	*CNTN6*	*CNTN6*
3	1393109	rs73006880	2.81e-06	0.00	*CNTN6*	*CNTN6*
3	1405829	rs10510209	1.7e-07	0.00	*CNTN6*	*CNTN6*
3	1427566	rs72997748	2.05e-06	0.00	*CNTN6*	*CNTN6*
3	1427625	rs9820933	6.82e-06	65.09	*CNTN6*	*CNTN6*
3	139750465	rs9874196	8.33e-06	0.00	*CLSTN2*	*CLSTN2*
3	139750888	rs9820731	8.83e-06	0.00	*CLSTN2*	*CLSTN2*
3	139779455	rs2162212	8.68e-06	0.00	*CLSTN2*	*CLSTN2*
5	36667579	rs10491374	2.94e-06	0.00	*SLC1A3*	*RPL39P22,SLC1A3*
5	36668022	rs3776585	7.3e-06	0.00	*SLC1A3*	*RPL39P22,SLC1A3*
7	28893972	rs28481569	2.31e-06	0.00		*CREB5,TRIL*
7	28895777	rs11772521	2.62e-06	0.00		*CREB5,TRIL*
7	28897686	rs10276611	3.41e-06	0.00		*CREB5,TRIL*
7	68036803	rs4718787	1.82e-06	0.00		*LOC100419458*
7	68041389	rs75351549	1.82e-06	0.00		*LOC100419458*
7	83524245	rs62475844	3.94e-06	0.00		*SEMA3A*
8	21471008	rs2010278	3.01e-06	0.00		*GFRA2*
8	39933560	rs55956871	3.04e-06	0.00		*LOC100420944,IDO2,C8orf4*
9	116967338	rs41306492	6.47e-06	7.82	*COL27A1*	*MIR455,COL27A1*
10	25242947	rs274313	5.04e-06	0.00	*PRTFDC1*	*ENKUR,PRTFDC1,THNSL1*
10	25244392	rs274312	8.33e-06	0.00	*PRTFDC1*	*ENKUR,PRTFDC1,THNSL1*
10	25244919	rs274311	4.04e-06	9.83		*ENKUR,PRTFDC1,THNSL1*
10	25245208	rs274310	8.47e-06	8.70		*ENKUR,PRTFDC1,THNSL1*
10	25245394	rs182130	5.55e-06	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25245721	rs200730	7.38e-06	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25246496	rs274309	8.47e-06	8.50		*ENKUR,PRTFDC1,THNSL1*
10	25246509	rs274308	7.32e-06	13.24		*ENKUR,PRTFDC1,THNSL1*
10	25246820	rs274307	2.33e-06	2.28		*ENKUR,PRTFDC1,THNSL1*
10	25246950	rs274306	7.29e-06	12.95		*ENKUR,PRTFDC1,THNSL1*
10	25246980	rs274305	3.5e-06	10.88		*ENKUR,PRTFDC1,THNSL1*
10	25247006	rs274304	5.52e-06	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25247057	rs274303	7.19e-06	13.08		*ENKUR,PRTFDC1,THNSL1*
10	25248540	rs274301	6.35e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25249606	rs612482	9.76e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25252205	rs274289	1.54e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25254151	rs274288	1.78e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25255447	rs12761282	8.02e-06	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25257870	rs2818289	1.37e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25258244	rs2031600	1.45e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25261446	rs274290	5.75e-06	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25262946	rs274291	4.61e-06	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25264155	rs274292	6.03e-06	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25265047	rs274293	1.16e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25265078	rs274294	1.15e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25265303	rs166670	1.12e-07	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25269851	rs2007364	5.48e-08	0.00		*ENKUR,PRTFDC1,THNSL1*
10	25270441	rs902903	1.18e-06	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25271270	rs2307047	2.11e-06	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25273988	rs2279547	4.96e-08	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25277315	rs4556443	2.23e-07	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25277917	rs10764515	1.04e-07	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25283248	rs12784769	2.64e-06	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25293195	rs10764516	1.28e-06	6.87	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25296379	rs10764517	2.18e-06	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25297598	rs950609	1.91e-06	9.64	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25300544	rs4748993	2.37e-06	9.42	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25300609	rs4748994	7.69e-07	0.00	*ENKUR*	*ENKUR,PRTFDC1,THNSL1*
10	25305139	rs987274	1.69e-06	22.29	*ENKUR,THNSL1*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25305991	rs7082131	1.64e-06	27.60	*ENKUR,THNSL1*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25313818	rs7086282	2.15e-06	18.92	*ENKUR,THNSL1*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25316755	rs12356765	2.23e-06	18.04	*ENKUR*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*

Continued on next page

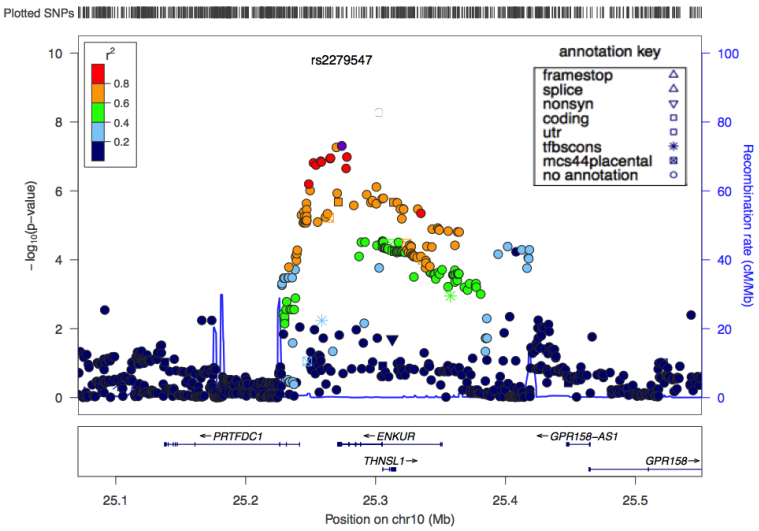
7.3 Appendix: Genome-wide association study of capsular group C meningococcal vaccine immunity

Chr	Position	SNP	p-value	I^2	Gene	Neighbouring genes (<100kb)
10	25319146	rs7393543	3.36e-06	27.88	*ENKUR*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25320223	rs6482467	6.44e-06	33.35	*ENKUR*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25321244	rs10764519	3.35e-06	27.89	*ENKUR*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25321589	rs11014357	3.42e-06	28.63	*ENKUR*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25332461	rs1352982	3.38e-06	59.66	*ENKUR*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
10	25334906	rs6482468	4.48e-06	59.75	*ENKUR*	*LOC101929025,ENKUR,PRTFDC1,THNSL1*
11	78794102	rs4944221	6.68e-06	4.77	*TENM4*	*LOC101928921,TENM4*
12	78675543	rs12303583	2.99e-06	46.70		*NAV3*
12	78677265	rs1008694	3.22e-06	46.44		*NAV3*
12	96161265	rs7977346	3.49e-06	10.49	*NTN4*	*PGAM1P5,CCDC38,NTN4,SNRPF*

Chr= chromosom



(a)



(b)

Figure 7.8: Regional plots of the imputed chromosome 3 and 10 loci associated with the persistence of capsular group C meningococci (MenC)-specific serum bactericidal assay titres following older childhood immunisation, generated from the meta-analysis of regression models conducted separately for assays using rabbit and human complement. The local gene structure is shown below the plot, with thick lines denoting exons joined by intronic regions. The statistical association (p-value) is shown on the primary y-axis (left) and the recombination rate is shown on the secondary y-axis (right). The single-nucleotide polymorphism with the most statistical evidence of association is shown in purple and remaining SNPs are coloured according to their linkage disequilibrium (r^2) with this 'lead' SNP. SNPs with particular features are assigned a symbol, as denoted by the 'annotation key'. a) associated chromosome 3 locus and b) associated chromosome 10 locus. These plots were created using the LocusZoom tool (244).

7.3 Appendix: Genome-wide association study of capsular group C meningococcal vaccine immunity

Table 7.14: Single-nucleotide polymorphisms (SNPs), in the imputed dataset, with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentration following childhood immunisation with MenC vaccine. Gene information was derived from a NCBI search of SNP reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	SNP	SNP Class	Position	p-value	Gene	Neighbouring genes (<100kb)
1	rs4074406		25190803	9.8351e-06		*SYF2,LOC391020,RSRP1,RHD*
1	rs2056957		25192642	9.5882e-06		*SYF2,LOC391020,RSRP1,RHD*
1	rs12058690		25193278	6.832e-06		*SYF2,LOC391020,RSRP1,RHD*
1	rs75137290		111763173	8.7048e-06		*FAM212B-AS1,KCND3-IT1,DDX20,KCND3,FAM212B,RAP1A*
3	rs111307985	intron-variant	185308823	7.5718e-07	*SENP2*	*MIR5588,EHHADH,EIF2S2P2,MAP3K13*
3	rs113473966	intron-variant	185308832	5.2599e-06	*SENP2*	*MIR5588,EHHADH,EIF2S2P2,MAP3K13*
5	rs60357858		89572699	3.4147e-06		
6	rs12212900		67623577	5.132e-06		
8	rs8185884		43626988	9.6819e-06		
10	rs2994657		31130598	3.2872e-06		*ZNF438*
11	rs4278546		79217391	7.8357e-06		*TENM4*
13	rs17052192	intron-variant	35963116	9.9697e-06	*NBEA*	*DCLK1*
13	rs17052202	intron-variant	35968822	1.3978e-06	*NBEA*	*TRNAQ50P,DCLK1*
13	rs61948739	intron-variant	35973962	9.61e-06	*NBEA*	*TRNAQ50P,DCLK1*
13	rs61948743	intron-variant	35978696	9.5508e-06	*NBEA*	*TRNAQ50P,DCLK1*
14	rs76107923	intron-variant	42338009	7.608e-06	*LRFN5*	
14	rs193070711	intron-variant	42340444	7.8692e-06	*LRFN5*	
14	rs117048587	intron-variant	42341777	7.9866e-06	*LRFN5*	
15	rs7171527	intron-variant	94934598	7.5968e-06	*MCTP2*	
18	rs62105211	intron-variant	74969284	5.1958e-06	*GALR1*	*ZNF407*
18	rs146824711		74991578	6.5641e-06		*ZNF407*
18	rs137908407		74997878	6.8862e-06		*ZNF407*
22	rs9608250	intron-variant	24617271	5.4967e-06	*GGT5*	*GGT1,LRRC75B,UPB1,ARL5AP4,CRIP1P4,BCRP3,POM121L10P, SNRPD3,GUCD1*
22	rs5996675	intron-variant	24618331	6.8514e-06	*GGT5*	*GGT1,LRRC75B,UPB1,ARL5AP4,CRIP1P4,BCRP3,POM121L10P, SNRPD3,GUCD1*

Chr= chromosome

Table 7.15: Single-nucleotide polymorphisms, in imputed dataset, with p-values suggestive ($<1\times10^{-5}$) of association with the persistence of $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay titres following childhood immunisation with MenC conjugate vaccine. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	SNP	SNP Class	Position	p-value	Gene	Neighbouring genes (<100kb)
2	rs1881417	intron-variant	29511480	6.1222e-06	*ALK*	*ALK*
2	rs2272409	intron-variant	29519622	8.2104e-06	*ALK*	*ALK*
2	rs7600162	intron-variant	225252746	8.6913e-06	*FAM124B*	
3	rs11920256	intron-variant	1388418	2.4814e-06	*CNTN6*	*CNTN6*
3	rs11916754	intron-variant	1388596	2.4715e-06	*CNTN6*	*CNTN6*
3	rs17785918	intron-variant	1389584	2.4793e-06	*CNTN6*	*CNTN6*
3	rs17785924	intron-variant	1389720	2.4792e-06	*CNTN6*	*CNTN6*
3	rs17785930	intron-variant	1390626	2.3862e-06	*CNTN6*	*CNTN6*
3	rs4684123	intron-variant	1391139	9.7e-06	*CNTN6*	*CNTN6*
3	rs4684124	intron-variant	1391159	9.7238e-06	*CNTN6*	*CNTN6*
3	rs10510209	intron-variant	1405829	7.1515e-07	*CNTN6*	*CNTN6*
5	rs10491374	intron-variant	36667579	6.163e-06	*SLC1A3*	*RPL39P22,SLC1A3*
7	rs13231325	intron-variant	39778210	2.2006e-06	*LINC00265*	*CICP22,LINC00265,RALA,RWDD4P2*
9	rs10815843	intron-variant	8385306	4.9115e-06	*PTPRD*	*PTPRD*
9	rs10815844	intron-variant	8386530	2.7977e-06	*PTPRD*	*PTPRD*
10	rs10824619		80248052	6.5014e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs1369752		80248755	6.2757e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs57705961		80251088	5.9232e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs11002578		80251436	5.9245e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs2163990		80252106	5.9273e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs7074159		80257477	3.457e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs55895329		80258089	3.2205e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs2016514		80258886	2.9432e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs2016695		80259067	2.8857e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs11002580		80259663	3.5555e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs7074960		80259796	2.6422e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs1583344		80260293	3.3343e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*

Continued on next page

Chr	SNP	SNP Class	Position	p-value	Gene	Neighbouring genes (<100kb)
10	rs11002582		80264319	2.077e-06		*DYDC1,ANXA11,MAT1A,RPS12P2,EIF5AP4,DYDC2*
10	rs45605435	intron-variant	97002331	5.4438e-06	*PDLIM1*	*HMGN2P35,SLIT1,LCOR*
10	rs45498798	intron-variant	97012373	5.181e-07	*PDLIM1*	*HMGN2P35,SLIT1,LCOR*
10	rs79551183	intron-variant	97013529	5.0358e-07	*PDLIM1*	*HMGN2P35,SLIT1,LCOR*
10	rs45554238	intron-variant	97015625	7.2018e-07	*PDLIM1*	*HMGN2P35,SLIT1,LCOR*
10	rs113904229	intron-variant	97016791	4.1013e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs45493292	intron-variant	97019278	6.9857e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs45518536	intron-variant	97020219	6.9626e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs10509685	intron-variant	97021389	1.9243e-06	*PDLIM1*	*SLIT1,LCOR*
10	rs45591738	intron-variant	97021727	4.7905e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs45562941	intron-variant	97022157	4.7533e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs17453729	intron-variant	97026530	4.3849e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs17453764	intron-variant	97026751	4.3634e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs17526000	intron-variant	97027086	3.6577e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs45533831	intron-variant	97028360	3.9425e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs79587021	intron-variant	97029366	2.8184e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs17453855	intron-variant	97030377	1.2688e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs45555036	intron-variant	97031101	1.7706e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs45444597	intron-variant	97031118	4.2524e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs45447693	intron-variant	97033775	2.8725e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs1854338	intron-variant	97034273	3.6944e-07	*PDLIM1*	*SLIT1,LCOR*
10	rs79126036	intron-variant	97039749	1.9295e-06	*PDLIM1*	*SLIT1,LCOR*
12	rs34633617		58303487	1.7325e-06		*RPL21P103*
12	rs76031679		58315476	2.4177e-06		*RPL21P103*
14	rs1953447	intron-variant	33442259	7.2953e-06	*NPAS3*	*NPAS3*
15	rs78869717		96672831	6.7951e-06		*SPATA8-AS1,FAM149B1P1*
20	rs2181872	intron-variant	1612323	1.8991e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs13045212	intron-variant	1612700	2.1977e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs6034235	intron-variant	1614914	3.0537e-07	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs13037155	intron-variant	1614915	1.1446e-07	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs13045074	intron-variant	1615105	1.362e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs17770331	intron-variant	1615308	1.1008e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs2277761	synonymous	1617069	6.1006e-07	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs6079931	intron-variant	1622255	1.0209e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs13037249	intron-variant	1622755	1.0368e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs4813190	intron-variant	1626069	1.6702e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs4813191	intron-variant	1626479	6.1376e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs3761270	intron-variant	1632653	1.5962e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs4813199	upstream-2KB	1640267	1.4628e-06	*SIRPG*	*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs4814411		1641146	5.4412e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs4814413		1642302	1.3947e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs4814415		1643499	1.3441e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs2023567		1644269	2.2334e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs2023566		1644291	2.233e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs2273334		1645830	2.1835e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs2273332		1650472	1.9387e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*
20	rs3818168		1651969	1.7846e-06		*SIRPB3P,SIRPB1,SIRPD,SIRPG*

Chr= chromosome

Table 7.16: Single-nucleotide polymorphisms with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations following meta-analysis of regression models fitted separately for younger and older childhood MenC vaccine recipients. Gene information was derived from a NCBI search of SNP reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package on build 37 of the human genome (241).

Chr	SNP	Position	p-value	I^2	Gene	Neighbouring genes (<100kb)
1	rs209741	22899011	1.851e-06	40.40	*EPHA8*	*MIR4253,EPHB2,LACTBL1*
1	rs4466696	48556330	8.088e-06	58.35		*RPL21P25,SPATA6,AGBL4*
3	rs17744749	9291441	7.789e-06	0.00	*SRGAP3*	*SRGAP3-AS3,LOC101927416,THUMPD3,PGAM1P4, THUMPD3-AS1,SRGAP3*
5	rs4916794	89573928	9.264e-06	0.00		
5	rs31318	123614481	1.87e-06	42.44		*CSNK1G3,KRT18P16*
5	rs149473	123620836	7.864e-06	0.00		*CSNK1G3,KRT18P16*
12	rs16945151	115639912	6.495e-06	0.00		*UBA52P7*
15	rs13379920	83756393	1.675e-06	0.00		*ADAMTSL3*

Chr= chromosome

Table 7.17: Single-nucleotide polymorphisms with p-values suggestive ($<1\times10^{-5}$) of association with $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay titres, following meta-analysis of regression models fitted separately for younger and older childhood MenC vaccine recipients. Gene information was derived from a NCBI search of reference SNP (rs) numbers, using 'GetSNPInfo' and 'AnnotateSNPList' functions in the 'NCBI2R' R package (241).

Chr	SNP	Position	p-value	I^2	Gene	Neighbouring genes (<100kb)
8	rs10096657	125357041	3.227e-06	82.91	*TMEM65*	*TRIB1,NSMCE2*
8	rs17297352	125359272	3.039e-06	83.08	*TMEM65*	*TRIB1,NSMCE2*
10	rs10509685	97021389	4.217e-07	0.00	*PDLIM1*	*LOC100505540,LOC102724002,C10orf12,SLIT1,LCOR*
13	rs206070	32896846	5.383e-06	0.00	*BRCA2*	
13	rs543304	32912299	6.714e-06	0.00	*BRCA2*	*LINC00423,TOMM22P3*
14	rs1953447	33442259	2.332e-06	0.00	*NPAS3*	*NPAS3*
20	rs13045212	1612700	6.187e-06	73.13	*SIRPG*	*SIRPB3P,LOC101929010,SIRPB1,SIRPD,SIRPG*
20	rs17770331	1615308	3.482e-06	75.27	*SIRPG*	*SIRPB3P,LOC101929010,SIRPB1,SIRPD,SIRPG*

Chr= chromosome

7.4 Appendix: Genome-wide association study of the persistence of serological immunity to capsular group C meningococcal vaccine following childhood immunisation - Stage 2

7.4.1 Background

Two-stage genetic association study designs, with an initial sample set densely genotyped (stage 1) followed by genotyping a subset of the genetic markers in an additional sample set (stage 2), can considerably reduce costs with only a small reduction in power compared to a one-stage approach of broadly genotyping all participants (73). Two-stage designs have increased efficiency since markers with little statistical evidence of association are eliminated after the first stage. The statistical power of a two-stage design is influenced by a number of parameters that can be controlled by the investigator such as the division of participants into stages 1 and 2, the density of genotyping in stage 1 and the proportion of markers carried forward to stage 2 (73). The power of the two-stage approach is also influenced by the analytical method of association testing. Classically, stage 2 has been treated as a replication cohort with focus being on the

findings that reach statistical significance when stage 2 is examined alone. This approach was perceived to be advantageous as the threshold for statistical significance (α) would only need to be corrected for the number of tests conducted in stage 2. The alternative approach is joint analysis of both stages, which includes more data but necessitates more stringent correction for multiple testing.

Joint analysis of the two-stages has been shown to be more powerful than a replication based analysis in the majority of simulations, particularly when a large proportion ($\geq$30%) of individuals are included in stage 1 (73). The superior power of joint analysis is logical as this approach incorporates the strength of evidence from stage 1, whereas replication analysis simply includes markers achieving a somewhat arbitrary threshold. Similarly, the inclusion of data from stage 1 also assists with analysis in the presence of linkage disequilibrium and polygenic effects (73). A scenario in which replication analysis can be marginally more powerful is when there is heterogeneity between stages 1 and 2, with the latter demonstrating larger effect size (73). Conversely, when the effect size is larger in stage 1 a joint analysis can be far more powerful (73).

Several recent genetic association studies have utilised two-stage designs, including the studies of HBV induced immunity (22, 25). In section '3.4', stage 1 of a two-stage genetic association study of the persistence of serological immunity following MenC vaccine was described. This section describes stage 2 of this genetic association study including participant demographics, SNP selection and analysis methodology; however, genotyping was ongoing at the time of writing this report.

7.4.2 Materials and methods

7.4.2.1 Stage 2 participants

This section includes healthy children recruited into four clinical trials (282, 298, 299, 300). In brief, study I was an additional cohort recruited in the clinical trial described as study 2 in the previous section (subsection '3.4.3.1 Stage 1 participants') (282). Study II examined the immunogenicity of infant immunisation with tetravalent meningococcal conjugate vaccine, with

persistence of MenC-specific immunity measured at 1 year of age (298). Study III was a study assessing the immunogenicity of 7- and 13-valent pneumococcal conjugate vaccine alongside routine infant immunisations, with persistence of MenC-specific immunity measured at 1 year of age (301). Study IV was a longitudinal study assessing kinetics of the decline in MenC-specific antibody following early childhood (1-4 years of age) MenC immunisation, with persistence of MenC-specific immunity measured at 3.5-5.0 years of age (300).

7.4.2.2 Vaccines

Participants described in this section received MenC conjugate vaccine either in a clinical trial setting or as part of their routine clinical care. Vaccines included monovalent MenC conjugated to CRM_{197} (Meningitec or Menjugate) or tetanus toxoid (NeisVac-C), a quadrivalent MenACWY vaccine conjugated to CRM_{197}, or a combination Hib-MenC (Menitorix) with the MenC polysaccharide conjugated to tetanus toxoid. These vaccines were previously described in subsection '2.2.3.2 Vaccines'.

7.4.2.3 Immunological measurements

The VEU, Manchester, measured MenC-specific hSBA titres for study I, and rSBA titres for study III and study IV. Novartis Vaccines, Marburg, measured hSBA titres for study II. MenC-specific IgG concentrations for study I were measured by the VEU, as described in subsection '3.4.3.3 Immunological measurements'. MenC-specific IgG concentrations for study II were measured in-house (Oxford Vaccine Group) by ELISA, as described in subsection '2.2.3.3 Immunological measurements'.

7.4.2.4 DNA extraction

DNA was extracted from clotted blood samples as described in subsection '3.4.3.4 DNA extraction'.

7.4.2.5 SNP selection

A subset of SNPs from stage 1 of the GWAS were selected for genotyping in stage 2. These SNPs were selected to represent the most credible association signals in an efficient manner. To this end, SNPs were selected based on statistical support for association with persistence of MenC-specific IgG concentrations and/or SBA titres in stage 1, taking into account LD between SNPs.

The R code for the algorithm devised to select SNPs for stage 2 is presented in appendix '7.4.3.5 Linkage disequilibrium pruning'. In brief, genotyped SNPs from stage 1 with association p-values $<1\times10^{-4}$, as well as imputed SNPs with p-values $<1\times10^{-5}$, in the regression models of MenC-specific IgG concentrations or SBA titres were selected. A different statistical cut-off for genotyped and imputed SNPs was imposed as imputed SNPs have an inevitable reduced certainty. These SNPs were separated into 500 kb chromosomal windows and windows with <3 SNPs were eliminated as these are more likely to represent genotyping and/or imputation errors. Next, to avoid redundant genotyping of markers in high LD the SNP list was pruned by filtering SNPs with an R^2 >0.95 with the lead SNP within each 500kb window. Then, the next most significant SNP remaining in each window was selected and LD filtering repeated iteratively, until all SNPs had been interrogated.

7.4.2.6 Genotyping

Sequenom MassARRAY® iPLEXs were designed as described in subsection '2.2.3.6 Genotyping'. Sequenom MassARRAY® iPLEX assays were completed at the Genome Institute of Singapore. Genotyping of the replication SNPs was conducted on all samples, including samples that were previously genotyped using the microarray platform in stage 1. These data were used to validate the genotypes typed on the microarray, as well as the imputed SNPs.

7.4.2.7 Quantitative trait loci analysis

Genotyping was ongoing at the time of writing this report. Joint analysis of stages 1 and 2 will be conducted in two ways. Firstly, regression analysis will be conducted using the SNPs

genotyped in both stages in a single analysis, including age at vaccination, study identifier, SBA complement source (where applicable) and time since vaccination as covariates. Secondly, as stage 2 did not have broad enough genotyping data to control population stratification, meta-analysis of stages 1 and 2 will be conducted to assess heterogeneity.

7.4.3 Results

7.4.3.1 Stage 2 demographics

Participant demographics of the two-stage GWAS of the persistence of MenC-specific serological immunity following childhood immunisation with MenC conjugate vaccine are displayed in Table 7.18.

Table 7.18: Demographics of the two-stage genome-wide association study of the persistence of capsular group C meningococci (MenC)-specific serological immunity following childhood immunisation with MenC conjugate vaccine.

Stage	Study	Male/Female	Time since vaccination days* (IQR)	Age at vaccination days* (IQR)	Reference
1	†Cohort 1	122/140	1380.5 (1279.3 - 1423.0)	3832.5 (3633.5 - 4073.3)	(213)
	†Cohort 2	449/378	1777 (1679 - 1856)	3514 (3038 - 4078)	(214)
	‡Study 1	39/27	275.5 (263.3-287.8)	123 (118-128)	(145)
	‡Study 3	138/129	256 (248.5-264)	122 (118-127)	(283)
	‡Study 4	76/82	261 (249.3 -274)	125 (121 -131)	(141)
1/2	‡Study 2 & Study I	413/375	987.0 (422.5 - 2689.3)	120 (120-365)	(282)
2	‡Study II	57/51	253 (246 - 260)	123 (118 - 128)	(298)
	‡Study III	29/34	247 (237 - 257)	123 (117 - 131)	(301)
	‡Study IV	45/53	651.0 (537.5 - 782.5)	776.5 (665.0 - 864.0)	(300)

IQR= interquartile range, * = median.†= immunised as older children (6-15 years), ‡= immunised as younger children (≤3 years of age)

7.4.3.2 Serology

Density plots of the persistence of MenC-specific SBA titres following MenC conjugate vaccine are shown in Figure 7.9, and were analogous to those seen in stage 1 (Figure 3.20). A large proportion of participants had lost detectable levels of bactericidal antibody following immunisation with MenC vaccine in early life, and SBAs measured using rabbit complement formed a distribution to the right of those using human complement (Figure 7.9).

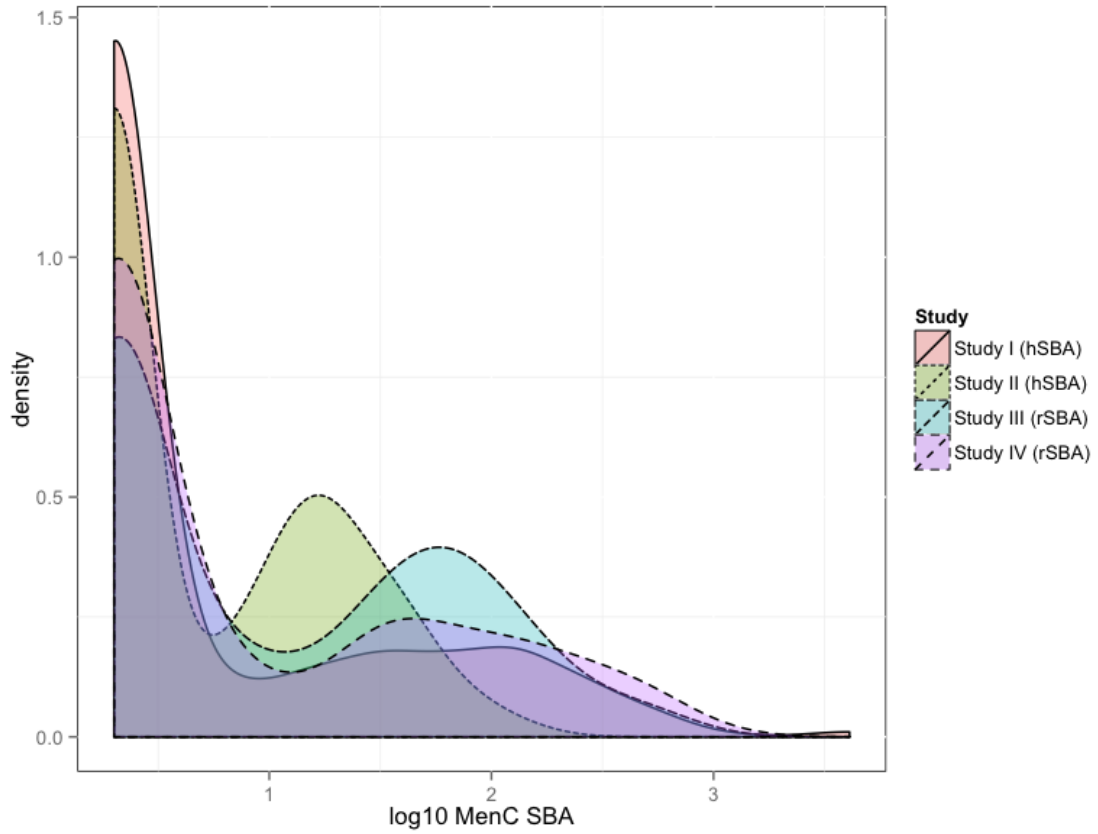


Figure 7.9: Density histogram of $\log_{10}$ transformed capsular group C meningococci (MenC)-specific serum bactericidal assay titres for participants included in stage 2 of the genome-wide association study of the persistence of MenC-specific serological immunity following childhood immunisation. rSBA= rabbit complement serum bactericidal assay, hSBA= human complement serum bactericidal assay. Density plots were created from a 1d kernel density estimate, using the R package 'ggplot2' (240).

Density plots of $\log_{10}$ transformed MenC-specific IgG concentrations are shown in Figure 7.10, and are similar to those seen in stage 1 (Figure 3.21). Study I has an enriched lower tail reflecting values below the assay LLQ (Figure 7.10).

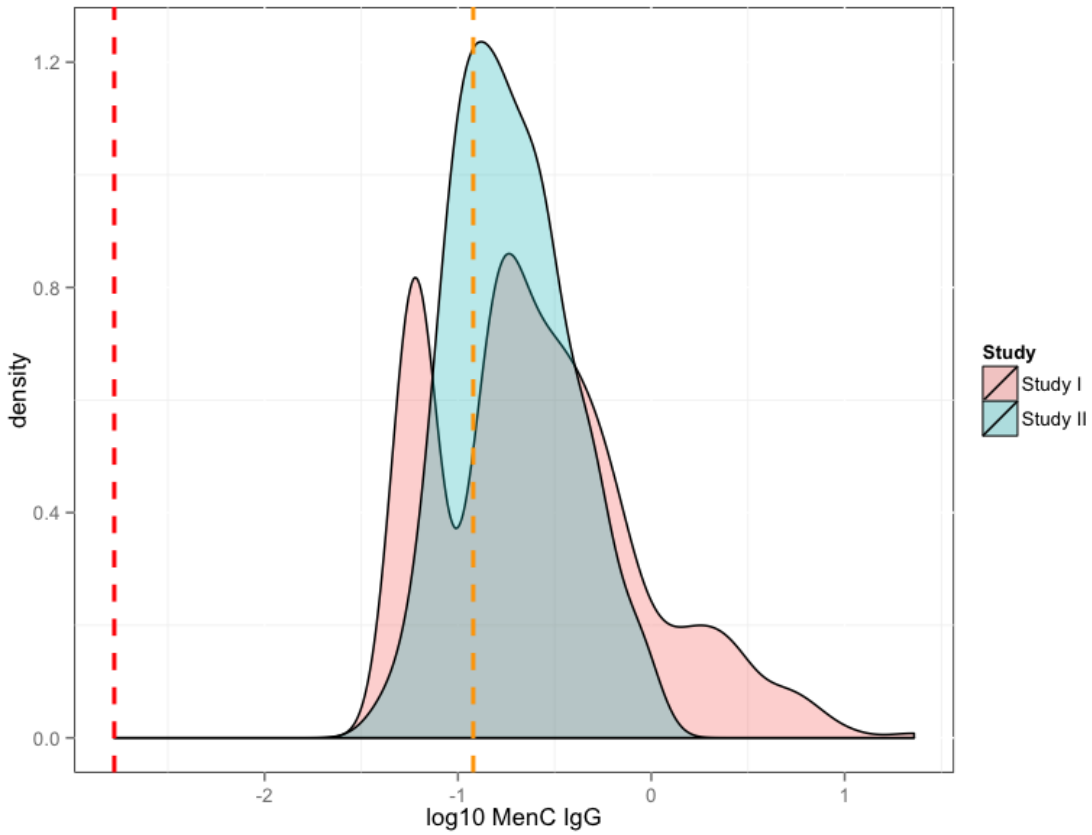


Figure 7.10: Density histogram of $\log_{10}$ transformed capsular group C meningococci (MenC)-specific IgG concentrations for participants included in stage 2 of the genome-wide association study of the persistence of MenC-specific serological immunity following childhood immunisation. Density plots were created from a 1d kernel density estimate, using the R package 'ggplot2' (240). The vertical dashed orange and red lines denote the lower limit of quantitation (LLQ) for study I and study II, respectively. Note, MenC-specific IgG concentrations were not measured for Study III and Study IV.

7.4.3.3 SNP selection

Five hundred and twenty-nine SNPs from stage 1 (section 3.4) were associated with the persistence of MenC-specific SBAs titres with a p-value $<1\times10^{-4}$ (Figure 7.11a). Sixty-nine variants remained following the SNP filtering algorithm described in subsection '7.4.2.5 SNP selection' (Figure 7.4.2.5b). Nine of these SNPs failed Sequenom design, leaving 60 SNPs for Sequenom genotyping in stage 2 (Table 7.19).

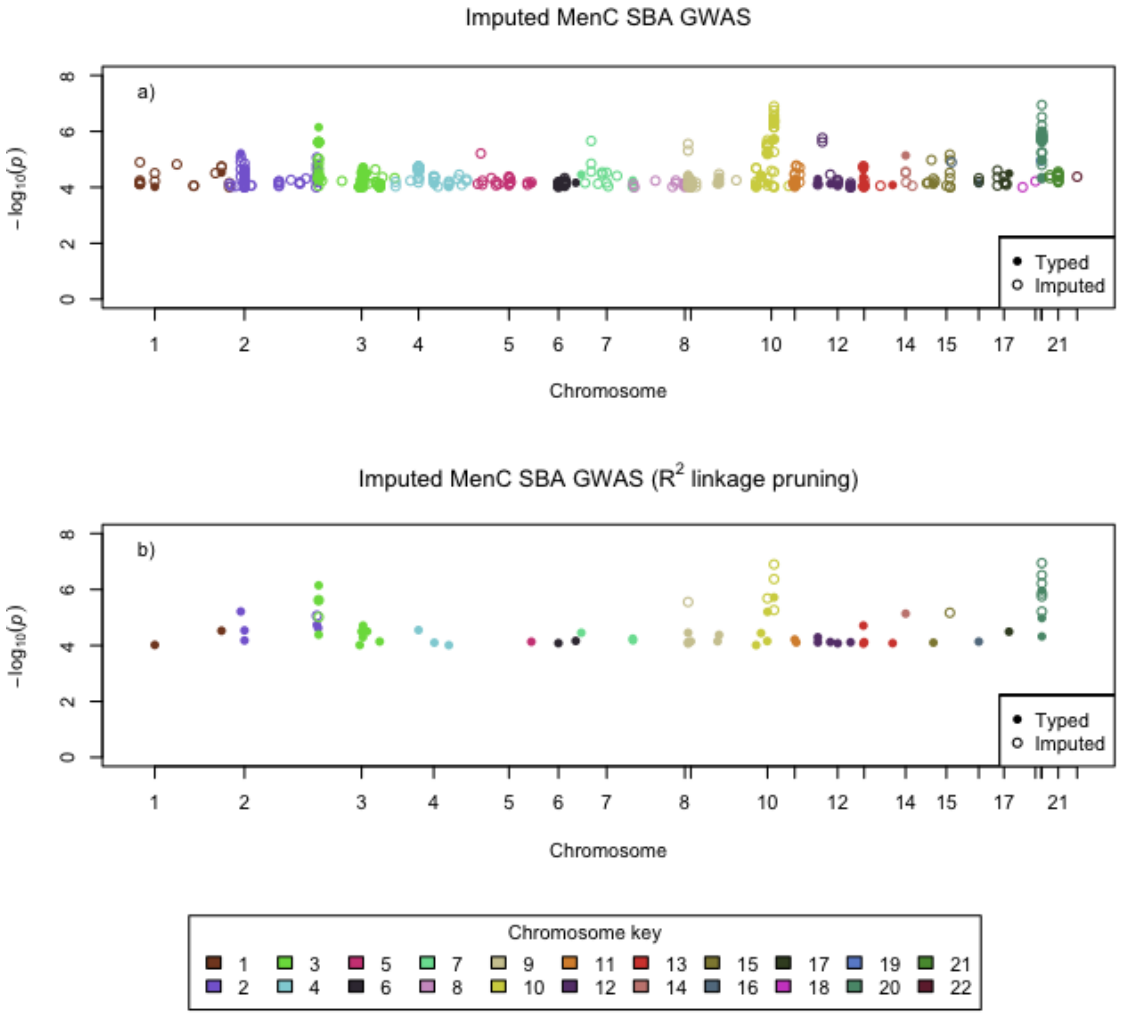


Figure 7.11: Manhattan plots illustrating the selection of stage 1 single-nucleotide polymorphisms (SNP) for stage 2 genotyping, in the genome-wide association study of persistence of capsular group C meningococci (MenC)-specific SBA titres following childhood immunisation with MenC vaccine. a) All SNPs in the imputed dataset with p-values $<1\times10^{-4}$ in stage 1 of the genome-wide association study of the persistence of $\log_{10}$ MenC-specific SBA titres following vaccination, b) SNPs selected for stage 2 analysis using linkage disequilibrium pruning, as described in subsection '7.4.2.5 SNP selection'.

Five hundred and two SNPs from stage 1 (section 3.4) were associated with the persistence of MenC-specific IgG concentrations with a p-value $<1\times10^{-4}$ (Figure 7.12a). Fifty-seven of these variants remained after SNP filtering, described in subsection '7.4.2.5 SNP selection' (Figure 7.12b). Twelve SNPs failed Sequenom design, leaving 45 SNPs for stage 2 analysis (Table 7.20).

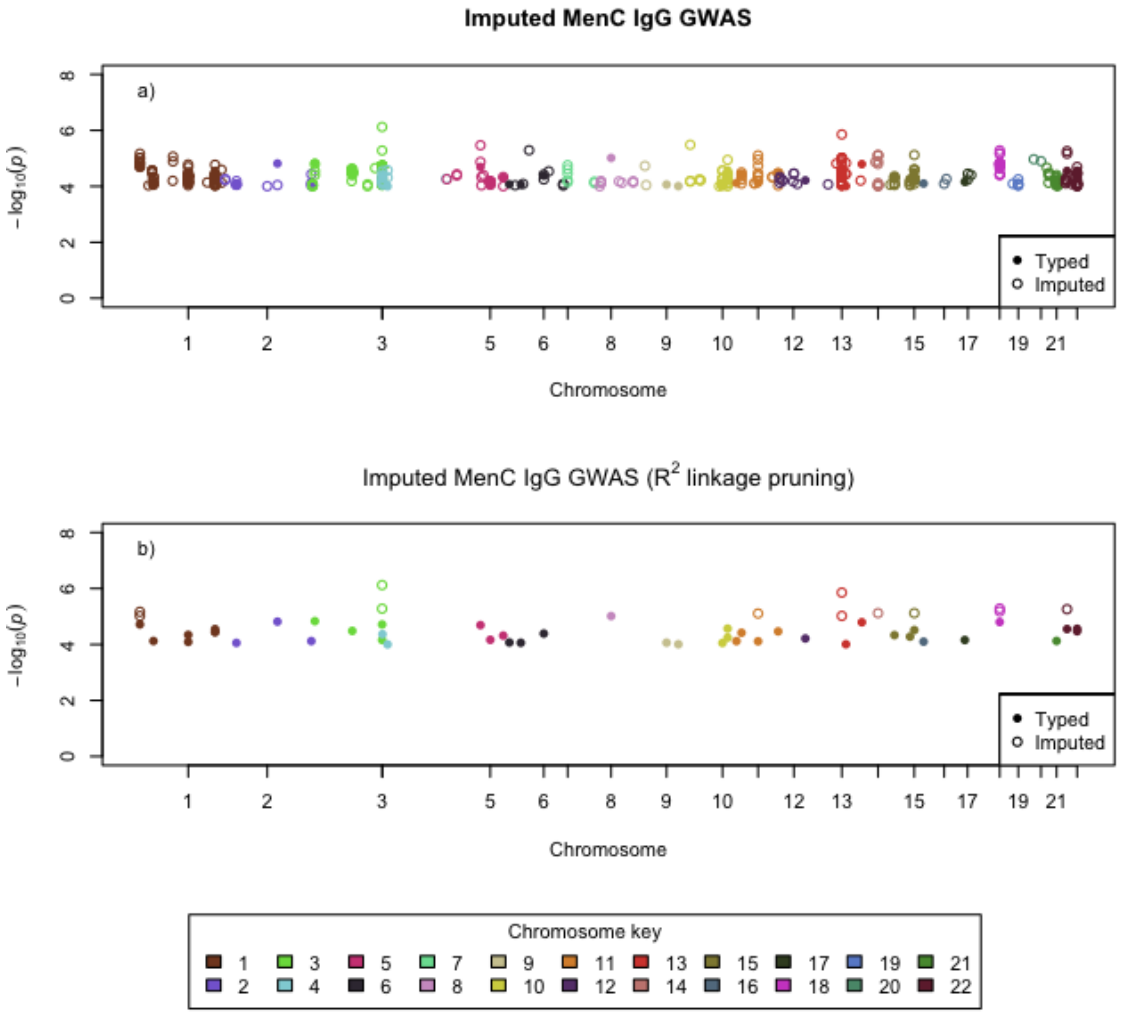


Figure 7.12: Manhattan plots illustrating the selection of stage 1 single-nucleotide polymorphisms (SNPs) for stage 2 genotyping in the genome-wide association study of persistence of capsular group C meningococci (MenC)-specific IgG concentrations following childhood immunisation with MenC vaccine. a) All SNPs in the imputed dataset with p-values $<1\times10^{-4}$ in stage 1 of the genome-wide association study of the persistence of $\log_{10}$ MenC-specific IgG concentrations following vaccination, b) SNPs selected for stage 2 analysis using linkage disequilibrium pruning, as described in subsection '7.4.2.5 SNP selection'.

7.4.3.4 Association analysis

Genotyping was in progress at the time this report was written.

7.4.3.5 Linkage disequilibrium pruning

```
function (dataframe, file_path, rep_LD_matrix, r2)
{
    d = dataframe
    file_path = file_path
    g = dataframe
    rep_LD_matrix = rep_LD_matrix
    j = 0
    r = r2
    dchunk_P <- d[order(d[, "chunk"], d[, "PValue"]), ]
    leading_SNPs <- dchunk_P[!duplicated(dchunk_P$chunk), ][,
        "SNP"]
    repeat {
        j = j + 1
        assign(paste("d_IT", j, sep = ""), subset(d, R2 < r)[,
            c("SNP", "PValue", "SnpType", "Chr", "Position",
                "chunk")])
        d$chunk <- factor(d$chunk)
        assign(paste("d_IT", j, "_chunk_P", sep = ""), get(paste("d_IT",
            j, sep = ""))[order(get(paste("d_IT", j, sep = ""))[,
            "chunk"], get(paste("d_IT", j, sep = ""))[, "PValue"]),
            ])
        assign(paste("d_IT", j, "_chunk_P_unique", sep = ""),
            get(paste("d_IT", j, "_chunk_P", sep = ""))[!duplicated(get(paste("d_IT",
                j, "_chunk_P", sep = ""))$chunk), ][, "SNP"])
        l = 1
        z = NULL
        for (i in 1:nrow(get(paste("d_IT", j, sep = "")))) {
            if (as.numeric(as.factor(get(paste("d_IT", j, sep = ""))$chunk[i])) ==
                l) {
```

```
            z[i] <- c(rep_LD_matrix[as.character(as.factor(get(paste("d_IT",
              j, "_chunk_P_unique", sep = ""))[l])), as.character(as.factor(get(paste
              ("d_IT", j, sep = ""))[i, "SNP"]))])
        }
        else {
            l <- l + 1
            z[i] <- c(rep_LD_matrix[as.character(as.factor(get(paste("d_IT",
              j, "_chunk_P_unique", sep = ""))[l])), as.character(as.factor(get(paste
              ("d_IT", j, sep = ""))[i, "SNP"]))])
        }
    }
    assign(paste("d_IT", j, sep = ""), `[[<-`(get(paste("d_IT",
        j, sep = "")), "R2", value = z))
    d <- get(paste("d_IT", j, sep = ""))
    if (j == 1) {
        assign(paste("Loop_list_of_rep_SNP_R2_", j, "_iterations",
            sep = ""), as.character(get(paste("d_IT", j,
            "_chunk_P_unique", sep = ""))))
    }
    else {
        assign(paste("Loop_list_of_rep_SNP_R2_", j, "_iterations",
            sep = ""), c(as.character(get(paste("d_IT", j,
            "_chunk_P_unique", sep = ""))), get(paste("Loop_list_of_rep_SNP_R2_",
            j - 1, "_iterations", sep = ""))))
    }
    assign(paste("test", j, sep = ""), subset(d, R2 < r)[,
        c("SNP", "PValue", "SnpType", "Chr", "Position",
            "chunk")])
    if (nrow(get(paste("test", j, sep = ""))) == 0)
        break
}
j = j + 1
assign(paste("Loop_list_of_rep_SNP_R2_", j, "_iterations",
    sep = ""), c(as.character(leading_SNPs), as.character(d$SNP),
```

```
        get(paste("Loop_list_of_rep_SNP_R2_", j - 1, "_iterations",
            sep = "")))
    print(length(unique(get(paste("Loop_list_of_rep_SNP_R2_",
        j, "_iterations", sep = "")))))
    write.table(g[g$SNP %in% as.character(unique(get(paste("Loop_list_of_rep_SNP_R2_",
        j, "_iterations", sep = "")))), ], file_path, quote = F,
        row.names = FALSE, sep = "\t")
    list(number.iterations = j)
}
```

7.4.3.6 Manhattan plot

```
function (dataframe, colours = c("#824626", "#856AD4", "#79DB4B",
    "#8FD1D8", "#CE4884", "#3A333F", "#79DFA0", "#CE9CC9", "#CFC99F",
    "#D2D24E", "#D98F39", "#663E79", "#D4473B", "#C7867F", "#8F873F",
    "#627B8E", "#3D4B29", "#CE50CA", "#698CCF", "#599679", "#58973C",
    "#6F273A"), ymax = "max", limitchromosomes = 1:22, ...)
{
    d = dataframe
    colours = colours
    palette()[1:6]
    if (!("Chr" %in% names(d) & "Position" %in% names(d) & "PValue" %in%
        names(d)))
        stop("Missing Chr, Position, or PValue")
    if (any(limitchromosomes))
        d = d[d$Chr %in% limitchromosomes, ]
    d = subset(na.omit(d[order(d$Chr, d$Position), ]), (PValue >
        0 & PValue <= 1))
    d$logp = -log10(d$PValue)
    d$pos = NA
    ticks = NULL
    lastbase = 0
    colours <- rep(colours, max(d$Chr))[1:max(d$Chr)]
    if (ymax == "max")
        ymax <- ceiling(max(d$logp))
```

```
if (ymax < 8)
    ymax <- 8
numchroms = length(unique(d$Chr))
if (numchroms == 1) {
    d$pos = d$Position
    ticks = floor(length(d$pos))/2 + 1
}
else {
    for (i in unique(d$Chr)) {
        if (i == 1) {
            d[d$Chr == i, ]$pos = d[d$Chr == i, ]$Position
        }
        else {
            lastbase = lastbase + tail(subset(d, Chr == i -
              1)$Position, 1)
            d[d$Chr == i, ]$pos = d[d$Chr == i, ]$Position +
              lastbase
        }
        ticks = c(ticks, d[d$Chr == i, ]$pos[floor(length(d[d$Chr ==
            i, ]$pos)/2) + 1])
    }
}
if (numchroms == 1) {
    with(d, plot(pos, logp, ylim = c(0, ymax), ylab = expression(-log[10](italic(p))),
        xlab = paste("Chromosome", unique(d$Chr), "position"),
        ...))
}
else {
    with(d, plot(pos, logp, ylim = c(0, ymax), ylab = expression(-log[10](italic(p))),
        xlab = "Chromosome", xaxt = "n", type = "n", ...))
    axis(1, at = ticks, lab = unique(d$Chr), ...)
    icol = 1
    for (i in unique(d$Chr)) {
        with(d[d$Chr == i, ], points(pos, logp, col = colours[icol],
```

```
            pch = c(1, 16, NA)[subset(d, Chr == i)$SnpType],
            ...))
        icol = icol + 1
        legend("bottomright", legend = c("Typed", "Imputed"),
            pch = c(16, 1), col = "black")
    }
  }
}
```

Table 7.19: Single-nucleotide polymorphisms (SNPs) selected for stage 2 of the genome-wide association study of the persistence of capsular group C meningococci (MenC)-specific serum bactericidal assay titres, following childhood immunisation with MenC conjugate vaccine.

SNP	Stage 1 p-value	Type	Chr	Position	MAF	NCHROBS	Sequenom design
rs263955	9.60E-05	typed	1	45340703	0.4483	3366	Passed
rs11117717	2.98E-05	typed	1	217062140	0.1227	3366	Passed
rs1881417	6.12E-06	typed	2	29511480	0.1518	3366	Passed
rs3134628	2.90E-05	typed	2	38977612	0.3197	3366	Passed
rs13029406	6.62E-05	typed	2	39030074	0.3969	3366	Passed
rs2042086	1.83E-05	typed	2	225222900	0.1892	3366	Passed
rs7600162	8.69E-06	imputed	2	225252746	0.1717	3116	Passed
rs1467180	2.39E-05	typed	2	228461771	0.03179	3366	Passed
rs17038032	4.12E-05	typed	3	1389117	0.1566	3366	Passed
rs17785924	2.48E-06	typed	3	1389720	0.1147	3366	Passed
rs17785930	2.39E-06	imputed	3	1390626	0.1148	3354	Failed
rs4684123	9.70E-06	imputed	3	1391139	0.1209	3358	Passed
rs10510209	7.15E-07	typed	3	1405829	0.1373	3366	Passed
rs4597706	9.70E-05	typed	3	107175934	0.3491	3366	Passed
rs1492489	3.17E-05	typed	3	111841303	0.2365	3366	Passed
rs16823858	5.08E-05	typed	3	115278682	0.05882	3366	Passed
rs7638681	1.95E-05	typed	3	115649821	0.3999	3366	Passed
rs2271630	3.14E-05	typed	3	126138778	0.07516	3366	Failed
rs6441226	7.24E-05	typed	3	158526223	0.0915	3366	Passed
rs10517505	2.82E-05	typed	4	61345972	0.03773	3366	Passed
rs2063058	7.95E-05	typed	4	102076982	0.3972	3366	Passed
rs7678098	9.77E-05	typed	4	139302875	0.1982	3366	Passed
rs10045539	7.36E-05	typed	5	166795002	0.3526	3366	Passed
rs551216	8.27E-05	typed	6	69581991	0.3304	3366	Passed
rs352099	6.88E-05	typed	6	114225811	0.2368	3366	Passed
rs1859732	3.52E-05	typed	7	14276567	0.4147	3366	Passed
rs6968536	5.71E-05	typed	7	146816475	0.418	3366	Passed
rs17170583	6.56E-05	typed	7	147236636	0.2062	3366	Passed
rs10977022	8.52E-05	typed	9	8385838	0.3541	3366	Passed
rs989813	7.58E-05	typed	9	8385969	0.4643	3366	Passed
rs10815844	2.80E-06	imputed	9	8386530	0.4668	3348	Failed
rs1500327	3.52E-05	typed	9	8390467	0.4747	3366	Passed
rs909323	7.17E-05	typed	9	15467175	0.03892	3366	Passed
rs815845	7.09E-05	typed	9	84216090	0.3066	3366	Passed
rs10746763	4.19E-05	typed	9	88131557	0.3743	3366	Passed
rs11101238	9.75E-05	typed	10	51029279	0.4581	3366	Passed
rs7093142	3.63E-05	typed	10	63649483	0.363	3366	Passed
rs1248696	6.99E-05	typed	10	79616605	0.1117	3366	Passed
rs1369752	6.28E-06	typed	10	80248755	0.1696	3366	Passed
rs11002582	2.08E-06	imputed	10	80264319	0.1721	3312	Passed
rs45605435	5.44E-06	imputed	10	97002331	0.07755	3288	Failed
rs10509685	1.92E-06	typed	10	97021389	0.07813	3366	Passed
rs17453855	1.27E-07	imputed	10	97030377	0.07403	3350	Passed
rs45444597	4.25E-07	imputed	10	97031118	0.08453	3348	Failed
rs7479972	6.12E-05	typed	11	15393887	0.05734	3366	Passed
rs12364788	7.83E-05	typed	11	19427092	0.3122	3366	Passed

Continued on next page

SNP	Stage 1 p-value	Type	Chr	Position	MAF	NCHROBS	Sequenom design
rs7301594	5.10E-05	typed	12	46575008	0.432	3366	Passed
rs6582622	7.76E-05	typed	12	46605239	0.1684	3366	Passed
rs1358274	7.52E-05	typed	12	78660092	0.1512	3366	Passed
rs2216144	8.38E-05	typed	12	97666478	0.4736	3366	Passed
rs11836499	7.69E-05	typed	12	130832922	0.156	3366	Passed
rs206070	1.95E-05	typed	13	32896846	0.1735	3366	Passed
rs206319	8.61E-05	typed	13	32986219	0.3164	3366	Passed
rs1324758	7.61E-05	typed	13	35011445	0.2493	3366	Passed
rs9587425	8.30E-05	typed	13	108326074	0.2264	3366	Passed
rs1953447	7.30E-06	typed	14	33442259	0.336	3366	Passed
rs16973971	7.93E-05	typed	15	54304560	0.1447	3366	Failed
rs78869717	6.80E-06	imputed	15	96672831	0.02063	3344	Passed
rs7204751	7.28E-05	typed	16	73079683	0.1396	3366	Passed
rs1285294	3.22E-05	typed	17	77925681	0.3624	3366	Passed
rs6034235	3.05E-07	imputed	20	1614914	0.1469	2736	Failed
rs13037155	1.14E-07	imputed	20	1614915	0.142	2768	Failed
rs13045074	1.36E-06	imputed	20	1615105	0.2439	3202	Failed
rs17770331	1.10E-06	typed	20	1615308	0.2516	3366	Passed
rs2277761	6.10E-07	imputed	20	1617069	0.2523	3310	Passed
rs4813191	6.14E-06	imputed	20	1626479	0.3008	3338	Passed
rs3818168	1.78E-06	imputed	20	1651969	0.2442	3358	Passed
rs6135736	1.05E-05	typed	20	1687450	0.2445	3366	Passed
rs4814511	4.78E-05	typed	20	1710272	0.2706	3366	Passed

MAF, Minor allelic frequency; NCHROBS, Non-missing allele count (posterior probability >0.9)
Chr= chromosome

Table 7.20: Single-nucleotide polymorphisms (SNPs) selected for stage 2 of the genome-wide association study of the persistence of capsular group C meningococci (MenC)-IgG concentrations following childhood immunisation with MenC conjugate vaccine.

SNP	p-value	Type	Chr	Position	MAF	NCHROBS	Sequenom design
rs7523543	1.92E-05	typed	1	25186716	0.2906	3366	Passed
rs2056957	9.59E-06	imputed	1	25192642	0.2879	3310	Passed
rs12058690	6.83E-06	imputed	1	25193278	0.3105	3208	Failed
rs6661520	7.53E-05	typed	1	60666641	0.4756	3366	Failed
rs1048694	4.55E-05	typed	1	151669997	0.3007	3366	Passed
rs1308152	8.10E-05	typed	1	151679441	0.3197	3366	Failed
rs4317815	2.83E-05	typed	1	221985915	0.1913	3366	Passed
rs3003892	3.61E-05	typed	1	221994043	0.2077	3366	Passed
rs4579763	3.55E-05	typed	1	222012950	0.1263	3366	Passed
rs3134628	8.81E-05	typed	2	38977612	0.3197	3366	Passed
rs11896112	1.53E-05	typed	2	146416386	0.1159	3366	Passed
rs11886348	7.58E-05	typed	2	235193436	0.04011	3366	Passed
rs17744749	1.47E-05	typed	3	9291441	0.05645	3366	Passed
rs4597706	3.28E-05	typed	3	107175934	0.3491	3366	Passed
rs17217776	1.93E-05	typed	3	185295704	0.08021	3366	Passed
rs111307985	7.57E-07	imputed	3	185308823	0.07474	3238	Failed
rs113473966	5.26E-06	imputed	3	185308832	0.07117	3316	Failed
rs9878208	6.90E-05	typed	3	185356176	0.05288	3366	Passed
rs3755963	4.31E-05	typed	4	894255	0.3015	3366	Passed
rs4698250	9.98E-05	typed	4	14312034	0.4097	3366	Passed
rs4916794	2.04E-05	typed	5	89573928	0.4257	3366	Passed
rs890785	6.82E-05	typed	5	115270867	0.456	3366	Passed
rs6579757	4.84E-05	typed	5	149161437	0.4409	3366	Passed
rs12524068	8.51E-05	typed	6	16096546	0.08734	3366	Passed
rs946223	8.73E-05	typed	6	46034367	0.4649	3366	Passed
rs2202766	4.09E-05	typed	6	106237505	0.1028	3366	Passed
rs8185884	9.68E-06	typed	8	43626988	0.02644	3366	Passed
rs17087710	8.64E-05	typed	9	87409862	0.06744	3366	Passed
rs7043283	9.85E-05	typed	9	118256674	0.1622	3366	Passed
rs1065209	8.85E-05	typed	10	115543045	0.1916	3366	Passed
rs11016808	2.72E-05	typed	10	129017883	0.2201	3366	Passed
rs10829607	5.71E-05	typed	10	129030343	0.2908	3366	Passed
rs11026723	7.60E-05	typed	11	22689884	0.3925	3366	Passed

Continued on next page

SNP	p-value	Type	Chr	Position	MAF	NCHROBS	Sequenom design
rs11033335	3.82E-05	typed	11	35957322	0.2974	3366	Passed
rs4278546	7.84E-06	imputed	11	79217391	0.4275	3216	Failed
rs1944719	7.72E-05	typed	11	79275128	0.4406	3366	Passed
rs525721	3.41E-05	typed	11	131829753	0.3259	3366	Passed
rs2048607	6.12E-05	typed	12	71211437	0.3476	3366	Passed
rs17052202	1.40E-06	imputed	13	35968822	0.1023	3362	Passed
rs61948743	9.55E-06	imputed	13	35978696	0.09299	3366	Failed
rs9316182	9.83E-05	typed	13	46688364	0.437	3366	Failed
rs4772329	1.61E-05	typed	13	88389149	0.4219	3366	Passed
rs76107923	7.61E-06	imputed	14	42338009	0.02414	3356	Failed
rs8023524	4.65E-05	typed	15	43020983	0.1797	3366	Passed
rs13379920	5.27E-05	typed	15	83756393	0.3069	3366	Passed
rs7171527	7.60E-06	imputed	15	94934598	0.1822	3250	Passed
rs8027043	3.07E-05	typed	15	94949710	0.1705	3366	Passed
rs11648329	7.95E-05	typed	16	24346879	0.4332	3366	Passed
rs1918733	6.94E-05	typed	17	48364231	0.4926	3366	Passed
rs1557400	1.58E-05	typed	18	74947450	0.02436	3366	Passed
rs62105211	5.20E-06	imputed	18	74969284	0.02082	3314	Failed
rs146824711	6.56E-06	imputed	18	74991578	0.02186	3340	Failed
rs2836999	7.49E-05	typed	21	40761545	0.3295	3366	Passed
rs8140505	2.82E-05	typed	22	24616876	0.2142	3366	Passed
rs9608250	5.50E-06	imputed	22	24617271	0.1979	3224	Failed
rs5770911	3.37E-05	typed	22	51011241	0.04813	3366	Passed
rs5770917	2.76E-05	typed	22	51017353	0.04843	3366	Passed

MAF, Minor allelic frequency; NCHROBS, Non-missing allele count (posterior probability >0.9)
Chr= chromosome, SNP= single-nucleotide polymorphism

7.5 Appendix: Genome-wide association study of persistence of serological immunity to *Haemophilus influenza* type b following immunisation - Stage 2

7.5.1 Background

This section describes stage 2 of a two-stage genetic association study to assess the genetic determinants of the persistence of PRP-specific serological immunity following childhood immunisation with Hib conjugate vaccine. The premise behind this tiered study design was described in subsection '7.4.1 Background'.

7.5.2 Materials and methods

7.5.2.1 Stage 2 participants

Participants included in this section were a subset of those described in chapter 3, for whom post-vaccination PRP-specific IgG concentrations, or sera, were available. This subset comprised of children from studies I and III, described in subsection '7.4.2.1 Stage 2 participants',

7.5.2.2 Vaccines

The children described in this section received three infant doses of Hib capsular polysaccharide conjugated to tetanus toxoid: at 2, 3 and 4 months of age, as described in subsection '4.3.3.2 Vaccines'. A subset of these also received a booster Hib conjugate vaccine prior to serological assessment, either at 1 year of age or as toddlers.

7.5.2.3 Immunological measurements

The VEU, Manchester, measured PRP-specific IgG for study I, using the assay described subsection '4.3.3.3 Immunological measurements'. Rijksinstituut voor Volksgezondheid en Milieu (RIVM) measured PRP-specific IgG for study III, by ELISA.

7.5.2.4 SNP selection

SNPs from stage 1 of the GWAS were chosen for assessment in stage 2, using the selection algorithm described in subsection '7.4.2.5 SNP selection'.

7.5.2.5 Genotyping

Sequenom MassARRAY® assays were completed at the Genome Institute of Singapore, as described in subsection '2.2.3.6 Genotyping'.

7.5.2.6 Quantitatively trait loci analysis

Analysis of this two-stage GWAS was described in subsection '7.4.2.7 Quantitatively trait loci analysis'.

7.5.3 Results

7.5.3.1 Stage 2 demographics

Demographics of the two-stage GWAS of the persistence of PRP-specific IgG concentrations following Hib immunisation are displayed in Table 7.21.

Table 7.21: Demographics of the two-stage genome-wide association study of the persistence of polyribosyl ribitol phosphate-specific IgG concentrations following childhood immunisation with *Haemophilus influenzae* type b conjugate vaccine.

Stage	Study	Male/Female	Time since vaccination days* (IQR)	Age at vaccination days* (IQR)	Reference
	Cohort 2	221/197	4551 (4518 - 4607)	524.0 (165.5 - 1055.5)	(214)
1	Study 1	37/24	275.5 (263.3 - 287.8)	123 (118 - 128)	(145)
	Study 3	121/117	256 (248.5 - 264)	122 (118 - 127)	(283)
1/2	Study 2 & Study I	413/375	763.0 (400.5 - 2562.5)	365 (120 - 365)	(282)
2	Study III	29/34	247 (237 - 257)	123 (117 - 131)	(301)

IQR= interquartile range, * = median.

7.5.3.2 SNP selection

Eight hundred and ninety-three SNPs from stage 1 were associated with the persistence of PRP-specific IgG concentrations with a p-value $<1\times10^{-4}$ (Figure 7.13a). Seventy-five variants remained following the SNP filtering algorithm, described in subsection '7.4.2.5 SNP selection' (Figure 7.13b). Six of these SNPs failed Sequenom design, leaving 69 SNPs for Sequenom genotyping in stage 2 (Table 7.22).

7.5.3.3 Association analysis

Genotyping was ongoing at the time this report was written.

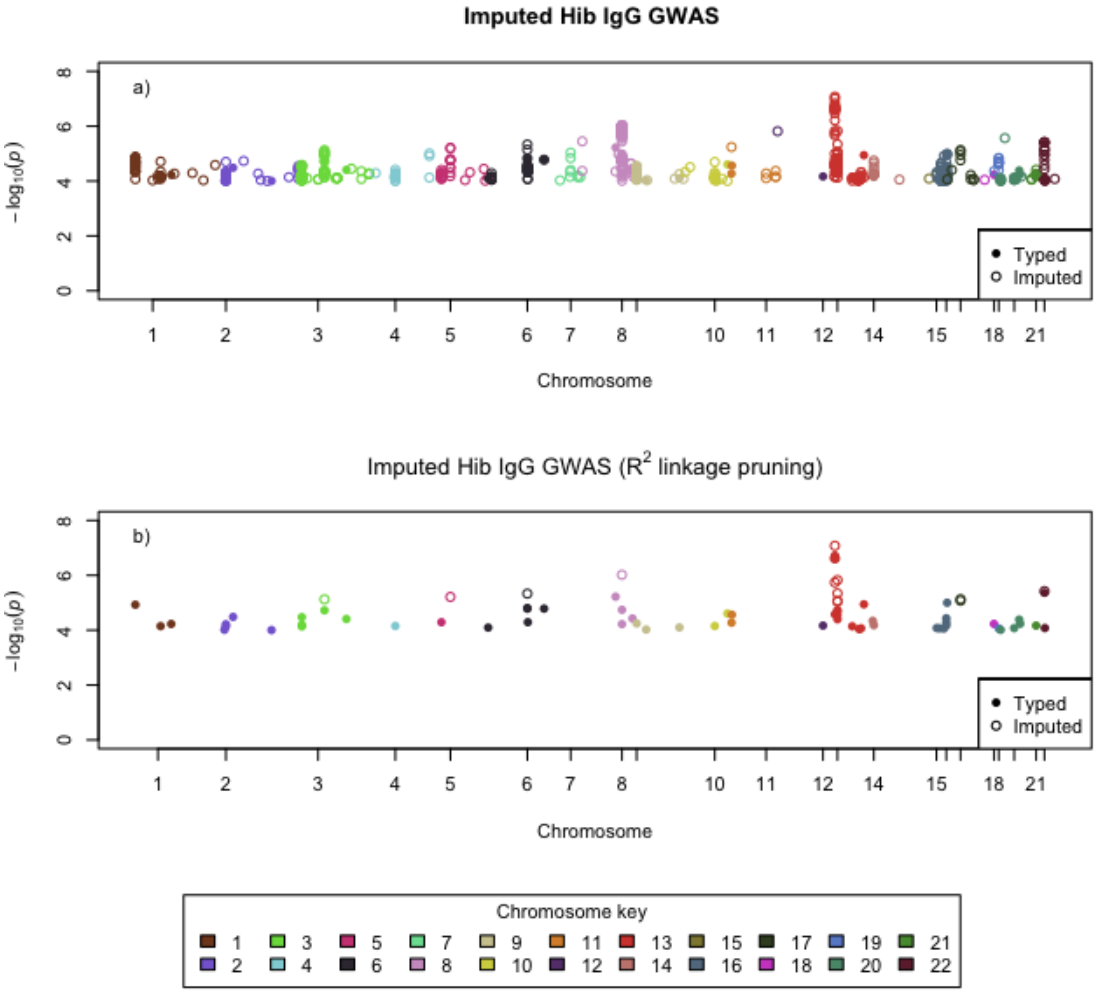


Figure 7.13: Manhattan plots illustrating the selection of stage 1 single-nucleotide polymorphisms (SNPs) for stage 2 genotyping, in the genome-wide association study of the persistence of polyribosyl ribitol phosphate (PRP)-specific IgG concentrations following immunisation with *Haemophilus influenzae* type b vaccine. a) All SNPs in the imputed dataset with p-values $< 1\times10^{-4}$ in stage 1 of the genome-wide association study of the persistence of $\log_{10}$ PRP-specific IgG concentrations following vaccination, b) SNPs selected for stage 2 analysis using linkage disequilibrium pruning, as described in subsection '7.4.2.5 SNP selection'

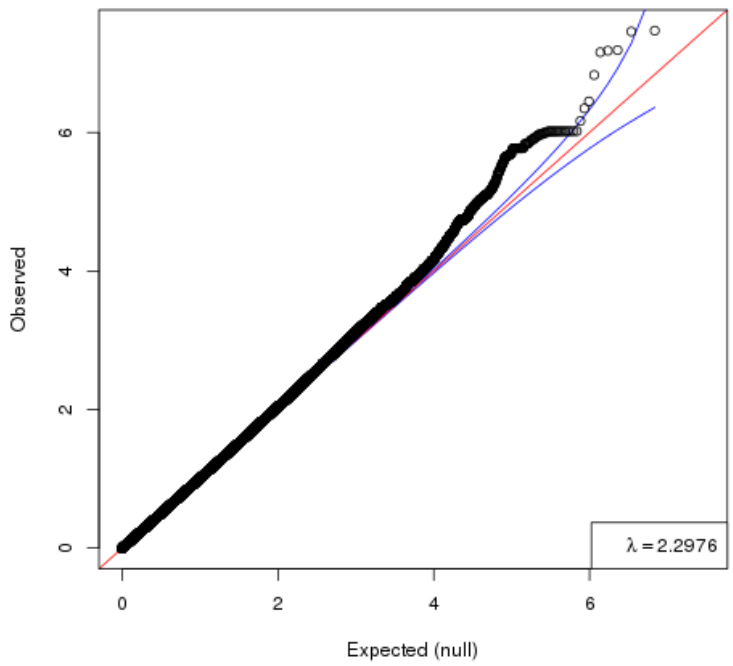


Figure 7.14: Quantile-quantile plot of p-values observed in association analysis between single-nucleotide polymorphisms and the persistence of $\log_{10}$ transformed tetanus toxoid (TT)-specific IgG concentrations following immunisation with TT containing vaccines, when principal components 1-3 were included in regression analysis. The red line represents the null distribution and the blue lines represent the upper and lower 95% confidence intervals. λ= inflation factor.

Table 7.22: Single-nucleotide polymorphisms (SNPs) selected for stage 2 of the genome-wide association study of the persistence of polyribosyl ribitol phosphate-specific IgG concentrations, following childhood immunisation with *Haemophilus influenzae* type b conjugate vaccine.

SNP	p-value	SNP type	Chr	Position	MAF	NCHROBS	Sequenom design
rs17037564	1.18E-05	typed	1	12060869	0.05645	3366	Passed
rs17105542	7.12E-05	typed	1	81336283	0.1197	3366	Passed
rs6537880	5.93E-05	typed	1	110369740	0.2228	3366	Passed
rs6545975	9.68E-05	typed	2	25385485	0.4138	3366	Passed
rs882631	6.07E-05	typed	2	29280791	0.2891	3366	Passed
rs17692899	7.26E-05	typed	2	29285254	0.2002	3366	Passed
rs10185108	3.28E-05	typed	2	49030757	0.4792	3366	Passed
rs13397382	9.91E-05	typed	2	154510972	0.1845	3366	Passed
rs9845420	7.33E-05	typed	3	9502463	0.1548	3366	Passed
rs1060894	6.36E-05	typed	3	9508157	0.02941	3366	Passed
rs2279440	3.35E-05	typed	3	9518511	0.1349	3366	Passed
rs6790644	1.89E-05	typed	3	71592594	0.3916	3366	Passed
rs11720121	7.48E-06	imputed	3	71595258	0.3995	3342	Passed
rs11715147	3.91E-05	typed	3	131725318	0.1777	3366	Passed
rs1453474	7.03E-05	typed	4	72423624	0.03773	3366	Passed
rs7715554	5.13E-05	typed	5	32972458	0.07427	3366	Passed
rs72755778	6.13E-06	imputed	5	57611114	0.2171	3312	Failed
rs12195797	8.05E-05	typed	6	224716	0.03476	3366	Passed
rs11751353	1.56E-05	typed	6	107619620	0.2484	3366	Passed
rs7773295	1.65E-05	typed	6	107637486	0.2023	3366	Passed
rs8180652	4.66E-06	imputed	6	107666030	0.2479	3344	Passed
rs9320269	5.11E-05	typed	6	109032557	0.1482	3366	Passed
rs9384084	1.64E-05	typed	6	153515792	0.03179	3366	Passed

Continued on next page

SNP	p-value	SNP type	Chr	Position	MAF	NCHROBS	Sequenom design
rs920200	6.00E-06	typed	8	96901876	0.3339	3366	Passed
rs117450413	9.59E-07	imputed	8	114268090	0.04516	3366	Passed
rs10505204	6.01E-05	typed	8	114395458	0.04872	3366	Passed
rs6987707	1.81E-05	typed	8	114483641	0.04664	3366	Passed
rs747769	3.74E-05	typed	8	143207696	0.352	3366	Passed
rs10511511	5.66E-05	typed	9	8900067	0.1061	3366	Failed
rs4879847	9.50E-05	typed	9	34983974	0.1699	3366	Passed
rs7851052	7.96E-05	typed	9	126179204	0.1979	3366	Passed
rs12246215	7.05E-05	typed	10	86325884	0.03892	3366	Passed
rs2039488	2.48E-05	typed	10	121244739	0.08021	3366	Passed
rs7121565	5.33E-05	typed	11	11032988	0.4884	3366	Passed
rs1038909	2.74E-05	typed	11	12010327	0.3773	3366	Passed
rs12824981	6.80E-05	typed	12	129637348	0.4147	3366	Passed
rs1001853	8.40E-08	imputed	13	32561254	0.2797	3328	Passed
rs1001854	1.82E-06	imputed	13	32561356	0.3423	3012	Passed
rs9562389	1.91E-07	typed	13	32565369	0.2876	3366	Passed
rs459747	2.65E-05	typed	13	32586749	0.3782	3366	Passed
rs17591215	2.42E-07	typed	13	32587062	0.246	3366	Passed
rs34858589	2.40E-07	imputed	13	32589954	0.246	3366	Passed
rs3751377	2.80E-05	typed	13	39587889	0.1396	3366	Passed
rs7317852	1.86E-05	typed	13	39611148	0.1031	3366	Passed
rs9576725	8.55E-06	imputed	13	39651501	0.1105	3348	Passed
rs55713408	4.56E-06	imputed	13	39678827	0.1118	3266	Passed
rs7987816	1.49E-06	imputed	13	39678987	0.1041	3228	Failed
rs10162229	3.99E-05	typed	13	39682081	0.1619	3366	Passed
rs1333435	2.07E-05	typed	13	39698787	0.1708	3366	Passed
rs9548645	9.15E-06	imputed	13	39721225	0.1747	3246	Passed
rs6563127	7.21E-05	typed	13	80216147	0.0814	3366	Passed
rs9556980	9.15E-05	typed	13	99249554	0.3363	3366	Passed
rs7328957	8.56E-05	typed	13	104557841	0.1171	3366	Passed
rs1151455	1.14E-05	typed	13	112140849	0.3072	3366	Passed
rs17097766	4.58E-05	typed	14	24294206	0.02466	3366	Passed
rs11160030	6.57E-05	typed	14	27305570	0.2249	3366	Passed
rs4589552	8.37E-05	typed	16	767519	0.1673	3366	Passed
rs8058889	8.58E-05	typed	16	9089962	0.104	3366	Passed
rs7498385	8.70E-05	typed	16	20446839	0.3648	3366	Passed
rs17705438	6.73E-05	typed	16	26948830	0.3333	3366	Failed
rs9806952	3.68E-05	typed	16	26949796	0.3547	3366	Passed
rs12051010	5.15E-05	typed	16	26952896	0.2436	3366	Passed
rs2370287	1.00E-05	typed	16	29263902	0.3752	3366	Passed
rs8068625	8.42E-06	imputed	17	36766280	0.4634	2954	Passed
rs9889300	7.39E-06	imputed	17	36770130	0.4001	2722	Passed
rs1551247	5.90E-05	typed	18	55929965	0.2668	3366	Failed
rs10422323	8.66E-05	typed	19	13104027	0.1387	3366	Failed
rs6053378	9.72E-05	typed	20	5395483	0.2311	3366	Passed
rs285198	8.41E-05	typed	20	42364734	0.3503	3366	Passed
rs6099642	4.12E-05	typed	20	56051451	0.4557	3366	Passed
rs6128767	5.84E-05	typed	20	58717961	0.05377	3366	Passed
rs2212606	6.81E-05	typed	21	40044940	0.4899	3366	Passed
rs5758828	4.31E-06	typed	22	22924995	0.07665	3366	Passed
rs5758832	3.77E-06	imputed	22	22925702	0.07616	3348	Passed
rs2330579	8.36E-05	typed	22	24024430	0.3158	3366	Passed

MAF, Minor allelic frequency; NCHROBS, Non-missing allele count (posterior probability >0.9)
Chr= chromosome

7.6 Appendix: Genome-wide association study of persistence of serological immunity to tetanus toxoid following immunisation - Stage 2

7.6.1 Background

This section describes stage 2 of a two-stage genetic association study to assess the genetic determinants of TT-specific serological immunity, following childhood immunisations with TT-containing vaccines. The principle behind this tiered study design was described in subsection '7.4.1 Background'.

7.6.2 Materials and methods

7.6.2.1 Stage 2 participants

Participants included in this section were a subset of those described in chapter 3, for whom post-vaccination TT-specific IgG concentrations, or sera, were available. This subset comprised of children from studies I, III and IV, described in subsection '7.4.2.1 Stage 2 participants'.

7.6.2.2 Vaccines

All participants described in this section received three infant doses of a TT-containing vaccine: at 2, 3 and 4 months of age. In addition, some individuals received further doses of TT-containing vaccines at 1 year old of age or as toddlers, as described in subsection '4.4.3.2 Vaccines'.

7.6.2.3 Immunological measurements

The VEU measured TT-specific IgG for study I and RIVM measured TT-specific IgG for study III, using assays described in subsection '4.4.3.3 Immunological measurements'. Aventis Pasteur, Toronto, measured TT-specific IgG for study IV using an ELISA with an in-house reference standard. The LLQ for this assay was 0.1 IU/mL, with measures below this recorded as half the LLQ.

7.6.2.4 SNP selection

SNPs from stage 1 analysis were chosen for assessment in stage 2, using the selection algorithm described in subsection '7.4.2.5 SNP selection'.

7.6.2.5 Genotyping

Sequenom MassARRAY® assays were completed at the Genome Institute of Singapore, as described in subsection '2.2.3.6 Genotyping'. A number of stage 2 SNPs were within the HLA locus, due to the polymorphic nature of this locus iPLEX primer design is problematic. TaqMan® SNP assays, which are more conducive to allelic discrimination in polymorphic regions of the genome were used to genotype the HLA SNPs selected for stage 2 analysis. TaqMan® SNP genotyping assays are based on the TaqMan® probe and primer chemistry. The incorporation of a minor groove binder enables the use of short probes, which are more suited to polymorphic regions of the genome than long probes. Two allele specific probes, each incorporating a different fluorescent dye and a non-fluorescent quencher, specifically anneal to denatured template DNA. Allele specific dye detection is achieved by the exonuclease cleavage of the allele-specific 5′ dye label during the polymerisation reaction.

The TaqMan® SNP genotyping approach is suited to projects assessing only a few genetic markers. To maximise the efficiency of this methodology, conditional linear regression analysis was used to assess for independent signals among the HLA SNPs associated with TT-specific IgG concentrations. Conditional linear regression analyses were undertaken by 'forward' and 'backwards' procedures, using the statistical package 'stats' in R (153). Imputed SNP genotypes were called based upon posterior probabilities of > 0.7. Only complete cases were included in conditional analysis. Covariates described in subsection '4.4.3.5 Quantitative trait loci analysis' were included in conditional regression analysis.

7.6.2.6 Quantitatively trait loci analysis

Analysis of this two-stage GWAS was described in subsection '7.4.2.7 Quantitatively trait loci analysis'.

7.6.3 Results

7.6.3.1 Stage 2 demographics

Participant demographics of the two-stage GWAS of the persistence of TT-specific IgG concentrations following immunisation with TT-containing vaccines are displayed in Table 7.23.

Table 7.23: Demographics of the two-stage genome-wide association study of the persistence of tetanus toxoid (TT)-specific IgG concentrations following childhood immunisation with TT-containing vaccines.

Stage	Study	Male/Female	Time since vaccination days* (IQR)	Age at vaccination days* (IQR)	Reference
1	Study 1	36/25	275.5 (263.3 - 287.8)	123 (118 - 128)	(145)
	Study 3	121/117	256 (248.5 - 264)	122 (118 - 127)	(283)
1/2	Study 2 & Study I	413/375	616.0 (299.0 - 1497.5)	1185 (365 -1185)	(282)
2	Study III	29/34	247 (237 - 257)	123 (117 - 131)	(301)
	Study IV	45/53	1054.5 (983.8 - 1132.3)	365 (365 - 365)	(300)

IQR= interquartile range, * = median

7.6.3.2 SNP selection

Eight hundred and eighty-six SNPs from stage 1 were associated with the persistence of TT-specific IgG concentrations with a p-value of $<1\times10^{-4}$ (Figure 7.15a). Seventy-seven SNPs from stage 1 were selected for assessment in stage 2, using the selection algorithm described in subsection '7.4.2.5 SNP selection' (Figure 7.15b). Twenty-two of these SNPs failed Sequenom design, leaving 55 SNPs for Sequenom genotyping in stage 2 (Table 7.27).

7.6.3.3 Human leukocyte antigen SNPs

Four hundred and seven participants had called genotypes for all 18 HLA SNPs selected for stage 2 genotyping (subsection '7.6.3.2 SNP selection'), and were included in conditional regression analysis. Backward stepwise analysis resulted in the selection of eight SNPs, whereas forward stepwise analysis resulted in selection of four SNPs, with two SNPs being selected by both methods. The results of these procedures are displayed in Table 7.24 and Table 7.25. The four SNPs selected using the forward method were chosen for TaqMan® genotyping.

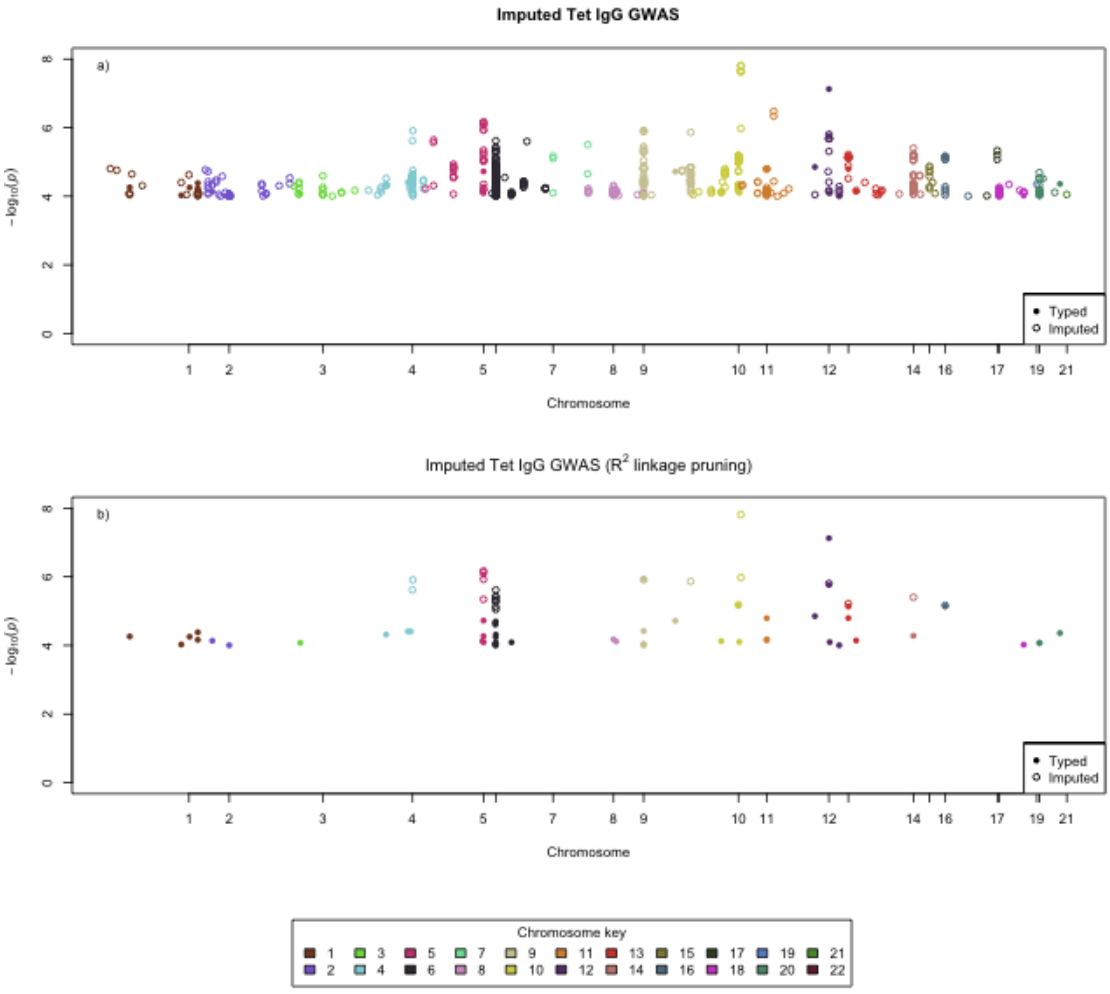


Figure 7.15: Manhattan plots illustrating the selection of stage 1 single-nucleotide polymorphisms for stage 2 genotyping in the genome-wide association study of the persistence of tetanus toxoid (TT)-specific IgG concentrations following immunisation with TT-containing vaccines. a) All SNPs in the imputed dataset with p-values $<1\times10^{-4}$ in stage 1 of the genome-wide association study of the persistence of $\log_{10}$ TT-specific IgG concentrations following vaccination, b) SNPs selected for stage 2 analysis using linkage disequilibrium pruning, as described in subsection '7.4.2.5 SNP selection'

Table 7.24: Human leukocyte antigen (HLA) single nucleotide polymorphisms (SNPs) remaining in the regression model of tetanus toxoid (TT)-specific IgG concentrations following backward stepwise elimination procedures.

	Estimate	SE	t.value	p-value
Intercept	-0.2495	0.0882	-2.83	0.005
rs2523663	-0.0843	0.0466	-1.81	0.071
rs2523650	-0.0714	0.0447	-1.60	0.111
rs2507964	0.1316	0.0744	1.77	0.078
rs9268877	0.1597	0.1010	1.58	0.114
rs2894221	-0.0974	0.0697	-1.40	0.163
rs186229361	-0.1120	0.0673	-1.66	0.097
rs147882356	-0.2288	0.1030	-2.21	0.027
rs9272374	-0.0771	0.0524	-1.47	0.142
Age_at_vaccination	-0.0011	0.0003	-3.59	3.73e-4
Study_2	0.5978	0.0643	9.30	9.74e-19
Study_1	0.2979	0.0795	3.75	2.06e-4
PC6	12.6709	6.1500	2.06	0.040
PC7	-40.0504	19.0000	-2.11	0.036

The 'Estimate' for the explanatory variables is the beta coefficient, SE= standard error .

Table 7.25: Human leukocyte antigen (HLA) single nucleotide polymorphisms (SNPs) incorporated in the regression model of tetanus toxoid (TT)-specific IgG concentration following forward stepwise inclusion procedures.

	Estimate	SE	t.value	p-value
Intercept	-0.0840	0.0784	-1.07	0.284
Study_3	-0.3545	0.0550	-6.44	3.36e-10
rs145763900	-0.1274	0.0410	-3.11	0.002
Time_since_vaccination	0.0001	2.59e-05	3.93	9.89e-05
rs2894221	-0.0600	0.0397	-1.51	0.131
rs9272374	-0.0915	0.0457	-2.00	0.046
rs3094188	-0.0735	0.0403	-1.82	0.070

The 'Estimate' for the explanatory variables is the beta coefficient, SE= standard error.

7.6.3.4 Association analysis

Genotyping was ongoing at the time this report was written.

Table 7.26: Results of association analysis between imputed classical human leukocyte antigen alleles and the persistence of tetanus toxoid (TT)-specific IgG concentrations following childhood immunisations with TT containing vaccines.

HLA allele	Estimate	p-value

Continued on next page

HLA allele	Estimate	p-value
HLAA_0101	-0.0893	0.02009
HLAA_0201	0.0676	0.06146
HLAA_0202	-0.0125	0.77411
HLAA_0205	0.3545	0.01721
HLAA_0206	-0.1786	0.15840
HLAA_0301	0.0585	0.19969
HLAA_0302	-0.0628	0.63119
HLAA_1101	-0.0392	0.55779
HLAA_2301	-0.1434	0.26085
HLAA_2402	0.0966	0.08595
HLAA_2403	-0.1801	0.72580
HLAA_2501	0.0561	0.63133
HLAA_2601	-0.0808	0.68034
HLAA_2608	-0.0524	0.88528
HLAA_2902	-0.1424	0.05355
HLAA_3001	-0.1485	0.36533
HLAA_3002	0.1607	0.16452
HLAA_3004	-0.0125	0.77411
HLAA_3101	-0.0951	0.30894
HLAA_3201	0.0204	0.85746
HLAA_3303	0.0756	0.68046
HLAA_3402	-0.4874	0.33873
HLAA_6601	-0.0852	0.73986
HLAA_6801	-0.0563	0.55075
HLAA_6802	-0.1533	0.55114
HLAA_7406	-0.0801	0.68128
HLAA_7410	-0.0125	0.77411
HLAB_0702	0.1238	0.00837
HLAB_0801	-0.1342	0.00210
HLAB_1302	0.0789	0.52256
HLAB_1401	0.1033	0.54718
HLAB_1402	-0.1436	0.19558
HLAB_1501	-0.0831	0.24666
HLAB_1503	1.0805	0.03534
HLAB_1507	-0.0125	0.77411
HLAB_1510	-0.0125	0.77411
HLAB_1516	-0.0125	0.77411
HLAB_1517	0.2339	0.26285
HLAB_1518	0.2018	0.31390
HLAB_1524	-0.0125	0.77411
HLAB_1801	0.0993	0.23137
HLAB_2701	0.1866	0.19274
HLAB_2702	0.5351	0.29633
HLAB_2705	-0.1961	0.05717
HLAB_3501	-0.0242	0.68036
HLAB_3502	0.1084	0.51845
HLAB_3503	0.0905	0.51692
HLAB_3508	-0.0806	0.75366
HLAB_3701	0.0434	0.70472
HLAB_3801	-0.0238	0.86152
HLAB_3901	0.0845	0.66462
HLAB_3905	-0.0125	0.77411
HLAB_3906	1.8050	0.00046
HLAB_4001	-0.1094	0.13390
HLAB_4002	0.2247	0.11870
HLAB_4101	0.1595	0.59199
HLAB_4102	-0.0562	0.75928
HLAB_4402	0.0205	0.69090
HLAB_4403	-0.1497	0.03017

Continued on next page

HLA allele	Estimate	p-value
HLAB_4404	0.2559	0.48557
HLAB_4405	0.3238	0.27716
HLAB_4501	0.2556	0.16495
HLAB_4701	0.2101	0.47809
HLAB_4901	0.0252	0.81364
HLAB_5001	0.3079	0.11425
HLAB_5101	0.0713	0.32219
HLAB_5108	-0.0125	0.77411
HLAB_5201	-0.2212	0.36553
HLAB_5301	-0.4663	0.37068
HLAB_5501	0.2805	0.05989
HLAB_5601	-0.1137	0.58836
HLAB_5701	-0.0081	0.92392
HLAB_5702	-0.0125	0.77411
HLAB_5801	-0.1942	0.28870
HLAC_0102	0.0887	0.31968
HLAC_0202	0.1087	0.14596
HLAC_0302	-1.1435	0.02677
HLAC_0303	0.0441	0.55699
HLAC_0304	-0.1168	0.07931
HLAC_0401	-0.0232	0.61238
HLAC_0501	-0.0214	0.67644
HLAC_0602	0.0659	0.24095
HLAC_0701	-0.0808	0.03616
HLAC_0702	0.1340	0.00375
HLAC_0704	0.1711	0.11340
HLAC_0801	-0.0125	0.77411
HLAC_0802	-0.0687	0.44519
HLAC_1202	-0.2212	0.36553
HLAC_1203	0.0308	0.69808
HLAC_1402	0.0483	0.79556
HLAC_1502	-0.0828	0.44969
HLAC_1505	-0.0125	0.77411
HLAC_1601	-0.1442	0.05061
HLAC_1602	-0.0125	0.77411
HLAC_1701	0.0765	0.64184
HLADQA_0101	0.0258	0.54797
HLADQA_0102	0.1073	0.00593
HLADQA_0103	-0.0365	0.57391
HLADQA_0201	-0.0201	0.64218
HLADQA_0301	-0.0031	0.93943
HLADQA_0401	0.0176	0.87759
HLADQA_0501	-0.0861	0.01449
HLADQB_0201	-0.1741	0.00010
HLADQB_0202	-0.0720	0.17889
HLADQB_0203	0.1637	0.47707
HLADQB_0301	0.0474	0.23733
HLADQB_0302	-0.0145	0.78743
HLADQB_0303	0.0318	0.64233
HLADQB_0304	0.0635	0.69731
HLADQB_0305	-0.0125	0.77411
HLADQB_0402	0.0141	0.90178
HLADQB_0501	-0.0061	0.89659
HLADQB_0502	0.1561	0.27268
HLADQB_0503	0.1184	0.25033
HLADQB_0504	-0.3076	0.39673
HLADQB_0601	-0.2524	0.36032
HLADQB_0602	0.1324	0.00276
HLADQB_0603	-0.0156	0.81006

Continued on next page

HLA allele	Estimate	p-value
HLADQB_0604	-0.0013	0.98851
HLADQB_0609	-0.0661	0.77394
HLADRB_0101	-0.0152	0.79292
HLADRB_0102	-0.0621	0.66575
HLADRB_0103	0.2215	0.29072
HLADRB_0104	0.0038	0.97156
HLADRB_0301	-0.1656	0.00021
HLADRB_0401	0.1087	0.03877
HLADRB_0402	-0.0217	0.94141
HLADRB_0403	-0.2884	0.13816
HLADRB_0404	-0.1725	0.06288
HLADRB_0405	-0.0727	0.84060
HLADRB_0407	0.0007	0.99657
HLADRB_0701	-0.0193	0.65342
HLADRB_0801	-0.0114	0.92642
HLADRB_0803	0.1880	0.52989
HLADRB_0804	-0.0125	0.77411
HLADRB_0901	-0.1146	0.40774
HLADRB_1001	0.0749	0.68540
HLADRB_1101	0.2917	0.00443
HLADRB_1102	0.5707	0.05372
HLADRB_1103	0.1672	0.51585
HLADRB_1104	-0.0390	0.79297
HLADRB_1201	-0.0896	0.39572
HLADRB_1202	-0.1003	0.73463
HLADRB_1301	-0.0063	0.92548
HLADRB_1302	-0.0180	0.82108
HLADRB_1303	0.0254	0.88926
HLADRB_1305	-0.0125	0.77411
HLADRB_1310	-0.0125	0.77411
HLADRB_1401	0.1269	0.24516
HLADRB_1404	-0.0125	0.77411
HLADRB_1501	0.1454	0.00103
HLADRB_1502	-0.2524	0.36032
HLADRB_1601	0.1263	0.40937
HLADRB_1602	0.3213	0.38854

Estimate = β coefficient

Table 7.27: Single-nucleotide polymorphisms (SNPs) selected for stage 2 of the genome-wide association study of the persistence of tetanus toxoid (TT)-specific IgG concentrations following childhood immunisation with TT containing vaccines.

SNP	p-value	SNP type	Chr	Position	MAF	NCHROBS	Sequenom design
rs1075473	5.48E-05	typed	1	65098136	0.06833	3366	Passed
rs4915236	9.38E-05	typed	1	201537485	0.2802	3366	Passed
rs4233516	5.57E-05	typed	1	223632876	0.1263	3366	Passed
rs6428904	6.89E-05	typed	1	245363423	0.1649	3366	Passed
rs6697031	4.14E-05	typed	1	245367423	0.1034	3366	Passed
rs982353	7.34E-05	typed	2	36573454	0.2166	3366	Passed
rs12617125	9.91E-05	typed	2	81472511	0.2635	3366	Passed
rs2371260	8.41E-05	typed	3	28180249	0.2403	3366	Passed
rs17005475	4.84E-05	typed	4	83067671	0.03951	3366	Passed
rs8192049	3.91E-05	typed	4	140597052	0.04337	3366	Passed
rs7669258	3.94E-05	typed	4	148308126	0.1809	3366	Passed
rs72732116	2.41E-06	imputed	4	152783376	0.1455	3278	Passed
rs76463503	1.23E-06	imputed	4	153626438	0.02227	3188	Passed
rs31223	7.45E-05	typed	5	156625279	0.3791	3366	Passed

Continued on next page

7.6 Appendix: Genome-wide association study of persistence of serological immunity to tetanus toxoid following immunisation - Stage 2

SNP	p-value	SNP type	Chr	Position	MAF	NCHROBS	Sequenom design
rs7720980	7.57E-07	imputed	5	159296102	0.2663	3282	Passed
rs1432675	1.89E-05	typed	5	159298474	0.1904	3366	Passed
rs891902	4.54E-06	imputed	5	159308795	0.1972	3362	Passed
rs13179580	8.82E-07	typed	5	159320525	0.2712	3366	Passed
rs7712049	5.36E-05	typed	5	159325351	0.3431	3366	Passed
rs10072006	6.75E-07	imputed	5	159326237	0.2744	3320	Failed
rs4921240	8.10E-05	typed	5	159327355	0.2445	3366	Passed
rs71603644	1.19E-06	imputed	5	159336790	0.1664	3064	Passed
rs3094188	2.19E-05	typed	6	31142245	0.3446	3366	Failed
rs2523663	9.67E-05	typed	6	31439740	0.3381	3366	Failed
rs2523650	2.47E-05	typed	6	31449022	0.3131	3366	Failed
rs2507964	8.90E-05	typed	6	31458157	0.3158	3366	Failed
rs2894221	4.87E-06	imputed	6	31461771	0.4039	3340	Failed
rs2894254	5.75E-05	typed	6	32345689	0.1227	3366	Failed
rs115365653	4.79E-06	imputed	6	32394913	0.3268	3366	Failed
rs9268877	9.16E-05	typed	6	32431147	0.4043	3366	Failed
rs9268880	2.01E-05	typed	6	32431358	0.2876	3366	Failed
rs186229361	3.74E-06	imputed	6	32435955	0.1083	3020	Failed
rs147882356	3.93E-06	imputed	6	32441679	0.3333	2814	Failed
rs145763900	7.40E-06	imputed	6	32445270	0.2972	3176	Failed
rs7749057	9.94E-05	typed	6	32448904	0.1637	3366	Failed
rs149028172	5.38E-06	imputed	6	32449523	0.3011	2654	Failed
rs114822165	8.92E-06	imputed	6	32583831	0.1101	3196	Failed
rs9272374	2.42E-06	imputed	6	32604669	0.1996	2780	Failed
rs9273327	8.35E-05	typed	6	32623223	0.139	3366	Failed
rs1794282	4.84E-05	typed	6	32666526	0.1209	3366	Failed
rs9442907	8.13E-05	typed	6	74123363	0.183	3366	Passed
rs7820882	6.66E-05	typed	8	68389059	0.3491	3366	Passed
rs11986395	7.64E-05	typed	8	75419366	0.4192	3366	Passed
rs7847373	8.96E-05	typed	9	1845946	0.109	3366	Passed
rs7019385	1.16E-06	typed	9	1849664	0.07903	3366	Passed
rs7860262	1.22E-06	imputed	9	1850836	0.08966	3346	Passed
rs10963648	9.98E-05	typed	9	1867013	0.08972	3366	Passed
rs10963660	3.79E-05	typed	9	1868585	0.1025	3366	Passed
rs6559700	1.92E-05	typed	9	85751913	0.1064	3366	Passed
rs113925093	1.37E-06	imputed	9	126234674	0.09241	3322	Passed
rs11819353	7.49E-05	typed	10	80674984	0.1676	3366	Passed
rs3740542	6.56E-06	imputed	10	126199289	0.06254	3342	Passed
rs11245253	6.70E-06	typed	10	126205537	0.06298	3366	Passed
rs10794142	7.91E-05	typed	10	128607787	0.3654	3366	Passed
rs79697632	1.06E-06	imputed	10	132642718	0.045	3222	Passed
rs75727401	1.54E-08	imputed	10	132643632	0.02879	3300	Passed
rs7128942	7.08E-05	typed	11	68387984	0.1673	3366	Passed
rs11228294	1.61E-05	typed	11	68388567	0.2356	3366	Failed
rs2513281	6.69E-05	typed	11	68417094	0.1536	3366	Passed
rs4913449	1.40E-05	typed	12	68835609	0.478	3366	Passed
rs2694403	1.73E-06	typed	12	105961210	0.2231	3366	Passed
rs1526856	1.56E-06	imputed	12	105967003	0.2237	3352	Passed
rs2711740	7.50E-08	typed	12	105978253	0.2406	3366	Passed
rs7974869	8.03E-05	typed	12	108026082	0.2213	3366	Passed
rs7972363	9.90E-05	typed	12	133179464	0.4649	3366	Passed
rs9507111	7.36E-06	typed	13	24093467	0.1771	3366	Passed
rs9510765	6.05E-06	imputed	13	24104109	0.1768	3360	Failed
rs12430762	1.60E-05	typed	13	24125369	0.1898	3366	Passed
rs7981875	7.20E-05	typed	13	44848338	0.09566	3366	Passed
rs4904054	5.24E-05	typed	14	82782185	0.4335	3366	Passed
rs12897855	3.95E-06	imputed	14	82787076	0.4883	3338	Passed
rs4494544	7.21E-06	typed	16	26986654	0.3835	3366	Passed
rs9746000	6.82E-06	imputed	16	26987784	0.3779	3318	Failed
rs7245374	9.53E-05	typed	18	69999553	0.145	3366	Passed
rs6085838	8.46E-05	typed	20	7028871	0.1179	3366	Passed
rs2026056	8.46E-05	typed	20	7029823	0.1182	3366	Passed
rs3787538	4.37E-05	typed	20	61303004	0.4115	3366	Passed

MAF, Minor allelic frequency; NCHROBS, Non-missing allele count (posterior probability >0.9)
Chr= chromosome

7.6.4 Pleiotropy in immunological responses to routine childhood immunisations

Table 7.28: Genes with single-nucleotide polymorphisms (SNPs) associated with more than one of the vaccine-induced serological measures. SNPs with an association p<1×10^{-4} with one or more of the serological measures are displayed, along with p-value (P) and beta values (β) corresponding to each of serological measures. P-values <1×10^{-4} are shown in bold. Within each gene SNPs are displayed in order of their chromosomal position.

Gene	SNPs	MenC-SBA P	MenC-SBA β	MenC-IgG P	MenC-IgG β	Hib-IgG P	PRP-IgG β	TT-IgG P	TT-IgG β
SRSF7	rs3112183	**2.09E-05**	-0.12	**8.71E-05**	-0.14	2.50E-01	0.04	3.11E-01	-0.06
SRSF7	rs3134630	**2.55E-05**	-0.12	1.17E-04	-0.14	2.54E-01	0.04	3.11E-01	-0.06
SRSF7	rs3134629	**2.52E-05**	-0.12	1.16E-04	-0.14	2.50E-01	0.04	3.10E-01	-0.06
SRSF7	rs3134628	**2.90E-05**	-0.12	**8.81E-05**	-0.14	2.86E-01	0.04	3.17E-01	-0.06
SRSF7	rs12621103	**2.74E-05**	-0.12	**8.84E-05**	-0.14	2.48E-01	0.04	3.23E-01	-0.06
SRSF7	rs12472454	**1.40E-05**	-0.12	3.02E-04	-0.12	2.80E-02	0.08	8.51E-01	0.01
SRSF7	rs13024811	**3.18E-05**	-0.12	1.03E-04	-0.14	2.50E-01	0.04	3.29E-01	-0.06
RP11-115H18.1	rs7627215	**9.69E-05**	-0.11	**3.19E-05**	-0.15	1.13E-01	-0.06	1.73E-01	-0.08
RP11-115H18.1	rs7649780	**9.67E-05**	-0.11	**3.18E-05**	-0.15	1.13E-01	-0.06	1.74E-01	-0.08
RP11-115H18.1	rs9836395	**9.72E-05**	-0.11	**3.19E-05**	-0.15	1.13E-01	-0.06	1.74E-01	-0.08
RP11-115H18.1	rs6785039	**9.76E-05**	-0.11	**3.20E-05**	-0.15	1.13E-01	-0.06	1.74E-01	-0.08
RP11-115H18.1	rs6437734	**9.84E-05**	-0.11	**3.22E-05**	-0.15	1.14E-01	-0.06	1.73E-01	-0.08
RP11-115H18.1	rs6798341	2.24E-04	-0.11	**4.03E-05**	-0.15	1.07E-01	-0.06	1.36E-01	-0.09
RP11-115H18.1	rs885190	1.00E-04	-0.11	**3.23E-05**	-0.15	1.15E-01	-0.06	1.73E-01	-0.08
RP11-115H18.1	rs9845513	1.00E-04	-0.11	**2.53E-05**	-0.16	7.24E-02	-0.07	1.66E-01	-0.08
RP11-115H18.1	rs41517052	**9.53E-05**	-0.11	**3.36E-05**	-0.15	1.07E-01	-0.06	1.60E-01	-0.08
RP11-115H18.1	rs55652157	**9.29E-05**	-0.11	**3.46E-05**	-0.15	1.12E-01	-0.06	1.68E-01	-0.08
RP11-115H18.1	rs4597706	**9.70E-05**	-0.11	**3.28E-05**	-0.15	1.07E-01	-0.06	1.60E-01	-0.08
RP11-115H18.1	rs13092776	1.06E-04	-0.11	**2.41E-05**	-0.16	9.00E-02	-0.06	1.59E-01	-0.08
RP11-115H18.1	rs13092989	**9.82E-05**	-0.11	**3.30E-05**	-0.15	1.07E-01	-0.06	1.58E-01	-0.08
PIWIL1	rs11060834	**8.72E-05**	-0.14	2.69E-04	-0.17	4.71E-01	0.03	5.03E-01	-0.05
PIWIL1	rs80287981	2.15E-03	-0.22	**8.54E-05**	-0.35	6.65E-01	0.04	2.42E-01	-0.18
PIWIL1	rs11060836	**8.71E-05**	-0.14	6.79E-04	-0.16	6.33E-01	0.02	3.99E-01	-0.06
PIWIL1	rs55887533	**9.97E-05**	-0.14	6.09E-04	-0.16	6.44E-01	0.02	3.98E-01	-0.06
PIWIL1	rs55881231	**9.97E-05**	-0.14	6.09E-04	-0.16	6.44E-01	0.02	3.98E-01	-0.06
PIWIL1	rs11836499	**7.69E-05**	-0.14	6.14E-04	-0.16	6.92E-01	0.02	3.94E-01	-0.06
PIWIL1	rs61940168	**6.93E-05**	-0.15	5.13E-04	-0.16	5.35E-01	0.03	4.55E-01	-0.06
PPARGC1B	rs13167468	7.82E-02	0.08	**9.91E-05**	0.23	8.97E-01	0.01	1.04E-01	-0.15
PPARGC1B	rs10051894	1.08E-01	0.04	**4.66E-05**	0.14	4.12E-01	0.03	8.41E-01	0.01
PPARGC1B	rs6579757	1.11E-01	0.04	**4.84E-05**	0.14	3.86E-01	0.03	7.92E-01	0.01
PPARGC1B	rs17653642	3.97E-02	-0.14	4.65E-01	-0.06	**3.60E-05**	-0.38	7.82E-03	-0.39
PTPRD	rs10758969	**9.64E-05**	-0.10	1.73E-02	-0.08	5.66E-01	-0.02	3.32E-01	-0.05
PTPRD	rs10977022	**8.52E-05**	-0.11	5.89E-04	-0.12	3.63E-01	-0.03	9.05E-02	-0.10
PTPRD	rs10758970	**8.43E-05**	-0.10	1.81E-02	-0.08	5.79E-01	-0.02	3.32E-01	-0.05
PTPRD	rs989813	**7.58E-05**	-0.10	1.78E-02	-0.08	5.96E-01	-0.02	3.30E-01	-0.05
PTPRD	rs989814	**4.26E-05**	-0.11	1.86E-02	-0.08	5.62E-01	-0.02	5.29E-01	-0.04
PTPRD	rs989815	**6.63E-05**	-0.11	1.64E-02	-0.08	5.54E-01	-0.02	2.98E-01	-0.06
PTPRD	rs10815844	**2.80E-06**	-0.12	5.20E-03	-0.10	1.34E-01	-0.05	3.73E-01	-0.05
PTPRD	rs1500327	**3.52E-05**	-0.11	7.93E-03	-0.09	7.30E-01	-0.01	7.64E-01	0.02
PTPRD	rs72704370	6.06E-01	0.02	7.27E-01	0.02	**3.37E-05**	0.27	8.63E-01	0.02
PTPRD	rs72704374	7.60E-01	0.01	7.27E-01	0.02	**2.80E-05**	0.26	7.69E-01	0.03
PTPRD	rs12551425	7.39E-01	0.01	7.27E-01	0.02	**3.26E-05**	0.26	8.10E-01	0.02
PTPRD	rs72704378	7.36E-01	0.01	7.28E-01	0.02	**3.40E-05**	0.26	8.22E-01	0.02
PTPRD	rs72704379	7.36E-01	0.01	7.28E-01	0.02	**3.40E-05**	0.26	8.22E-01	0.02
PTPRD	rs72704381	7.35E-01	0.01	7.29E-01	0.02	**3.45E-05**	0.26	8.25E-01	0.02
PTPRD	rs1433542	7.17E-01	0.02	7.17E-01	0.02	**3.81E-05**	0.26	8.82E-01	0.01
PTPRD	rs12551769	4.96E-01	0.03	6.97E-01	0.02	**3.98E-05**	0.26	9.73E-01	0.00
PTPRD	rs12552171	6.89E-01	0.02	6.95E-01	0.02	**5.39E-05**	0.25	9.76E-01	0.00
PTPRD	rs10511511	6.86E-01	0.02	6.93E-01	0.02	**5.66E-05**	0.25	9.86E-01	0.00
PTPRD	rs12554855	6.54E-01	0.02	6.53E-01	0.03	**7.77E-05**	0.25	8.03E-01	-0.02
PTPRD	rs1897671	7.54E-01	0.01	8.29E-01	0.01	**8.09E-05**	0.25	9.49E-01	-0.01
PTPRD	rs976385	7.33E-01	0.02	7.83E-01	0.02	**8.29E-05**	0.25	8.98E-01	-0.01
PTPRD	rs34657422	7.99E-01	0.01	8.02E-01	0.01	**8.61E-05**	0.25	9.49E-01	-0.01
PTPRD	rs35114587	7.50E-01	0.01	8.29E-01	0.01	**7.82E-05**	0.25	9.52E-01	-0.01
PTPRD	rs1019955	7.82E-01	0.01	7.55E-01	0.02	**8.84E-05**	0.25	8.99E-01	-0.01
PTPRD	rs2016766	7.84E-01	0.01	7.42E-01	0.02	**9.04E-05**	0.25	9.02E-01	-0.01
PTPRD	rs59834115	7.36E-01	0.01	7.80E-01	0.02	**8.08E-05**	0.25	9.01E-01	-0.01
PTPRD	rs12552247	7.41E-01	0.01	7.90E-01	0.02	**8.19E-05**	0.25	9.06E-01	-0.01
PTPRD	rs57781824	7.42E-01	0.01	8.03E-01	0.01	**7.79E-05**	0.25	9.15E-01	-0.01
PTPRD	rs72706206	7.46E-01	0.01	7.92E-01	0.02	**8.05E-05**	0.25	9.12E-01	-0.01
PTPRD	rs72706213	8.93E-01	0.01	9.34E-01	-0.00	**4.74E-05**	0.26	7.73E-01	-0.03
PTPRD	rs72706215	8.93E-01	0.01	9.34E-01	-0.00	**4.74E-05**	0.26	7.74E-01	-0.03
PTPRD	rs72706217	9.63E-01	0.00	9.95E-01	-0.00	**3.53E-05**	0.26	7.91E-01	-0.03
PTPRD	rs7027460	9.08E-01	0.01	8.67E-01	-0.01	**5.02E-05**	0.25	8.45E-01	-0.02
PTPRD	rs12551480	9.73E-01	0.00	9.10E-01	-0.01	**4.56E-05**	0.26	8.51E-01	-0.02
PTPRD	rs12552617	8.72E-01	0.01	9.42E-01	-0.00	**5.96E-05**	0.26	7.80E-01	-0.03

Continued on next page

Gene	SNPs	MenC-SBA P	MenC-SBA β	MenC-IgG P	MenC.IgG β	Hib-IgG P	PRP-IgG β	TT-IgG P	TT-IgG β
NTM	rs625676	9.19E-03	0.07	**3.43E-05**	0.15	1.79E-01	0.05	3.41E-01	0.06
NTM	rs548559	1.34E-02	0.07	**3.38E-05**	0.15	1.84E-01	0.05	4.44E-01	0.05
NTM	rs525721	1.03E-02	0.07	**3.41E-05**	0.15	1.38E-01	0.06	2.88E-01	0.06
NTM	rs632968	6.33E-03	0.08	**9.37E-05**	0.15	1.23E-01	0.06	3.47E-01	0.06
NTM	rs142025623	3.84E-01	-0.07	8.88E-01	-0.02	**4.28E-05**	-0.42	2.12E-02	-0.34
RPH3AL	rs11150891	**4.76E-05**	-0.11	2.12E-01	-0.05	1.46E-01	-0.05	8.41E-01	0.01
RPH3AL	rs117849583	**6.36E-05**	0.13	6.18E-02	0.08	2.50E-01	0.05	8.17E-01	0.01
RPH3AL	rs4563095	4.14E-01	0.06	7.47E-01	-0.03	**8.71E-05**	0.39	1.91E-01	0.21
DENND1A	rs7851052	2.11E-01	0.04	1.90E-01	0.06	**7.96E-05**	0.17	9.45E-03	0.17
DENND1A	rs10123099	2.00E-01	0.04	1.82E-01	0.06	**5.61E-05**	0.18	9.84E-03	0.17
DENND1A	rs73664359	6.02E-01	0.02	3.38E-01	0.06	1.97E-03	0.20	**2.22E-05**	0.41
DENND1A	rs76274475	2.04E-01	0.06	2.85E-01	0.06	7.81E-04	0.21	**1.49E-05**	0.43
DENND1A	rs78861713	2.09E-01	0.06	2.87E-01	0.06	7.56E-04	0.21	**1.44E-05**	0.43
DENND1A	rs75339791	2.23E-01	0.06	2.86E-01	0.06	8.86E-04	0.21	**1.42E-05**	0.43
DENND1A	rs73575569	4.89E-01	0.03	3.46E-01	0.05	3.15E-03	0.18	**3.69E-05**	0.40
DENND1A	rs113925093	2.63E-01	0.05	4.92E-01	0.04	6.49E-04	0.21	**1.37E-06**	0.46
DENND1A	rs57067576	5.03E-01	0.03	3.54E-01	0.05	3.29E-03	0.18	**3.44E-05**	0.40
DENND1A	rs79085337	1.10E-01	0.08	2.20E-01	0.07	4.35E-03	0.19	**2.48E-05**	0.44
DENND1A	rs76438231	1.10E-01	0.08	2.19E-01	0.07	4.39E-03	0.19	**2.50E-05**	0.44
DENND1A	rs60485679	1.11E-01	0.08	2.14E-01	0.08	4.87E-03	0.18	**3.09E-05**	0.43
DENND1A	rs7849651	1.12E-01	0.08	2.08E-01	0.08	5.31E-03	0.18	**3.41E-05**	0.43
DENND1A	rs10986082	3.28E-01	0.05	2.00E-01	0.08	2.38E-02	0.14	**1.43E-05**	0.42
DENND1A	rs74524716	2.52E-01	0.06	8.26E-01	0.01	1.38E-02	0.17	**6.54E-05**	0.41
DENND1A	rs113237886	1.62E-01	0.08	2.83E-01	0.08	6.64E-02	0.14	**4.17E-05**	0.50
DENND1A	rs75041694	1.89E-01	0.08	4.71E-01	0.06	4.84E-02	0.15	**2.48E-05**	0.52
DENND1A	rs77587038	2.05E-01	0.07	3.77E-01	0.06	1.27E-02	0.18	**4.60E-05**	0.46
DENND1A	rs112698181	3.19E-01	0.07	2.20E-01	0.11	4.14E-02	0.18	**7.88E-05**	0.55
DENND1A	rs187609854	2.48E-01	0.06	4.51E-01	0.05	1.58E-02	0.17	**2.10E-05**	0.46
DENND1A	rs113702595	2.19E-01	0.06	4.63E-01	0.05	2.36E-02	0.15	**4.35E-05**	0.43
DOCK1	rs10794142	6.64E-01	0.01	2.42E-01	0.04	5.33E-02	-0.07	**7.91E-05**	-0.23
DOCK1	rs61874056	7.64E-03	-0.09	**1.13E-05**	-0.20	4.40E-02	0.09	4.47E-01	0.06
DOCK1	rs11016808	8.09E-03	-0.09	**2.72E-05**	-0.18	1.56E-01	0.06	4.75E-01	0.05
DOCK1	rs11016809	2.98E-03	-0.09	**4.75E-05**	-0.16	4.08E-01	0.03	6.35E-01	-0.03
DOCK1	rs11016812	3.04E-03	-0.09	**5.41E-05**	-0.16	3.94E-01	0.03	6.37E-01	-0.03
DOCK1	rs11016829	8.55E-03	-0.08	**3.60E-05**	-0.17	1.39E-01	0.06	4.01E-01	0.06
DOCK1	rs35483211	5.44E-03	-0.08	**3.68E-05**	-0.16	4.38E-01	0.03	7.32E-01	-0.02
DOCK1	rs34168048	4.20E-03	-0.08	**7.57E-05**	-0.15	3.41E-01	0.04	8.30E-01	-0.01
DOCK1	rs10829607	4.44E-03	-0.08	**5.71E-05**	-0.15	3.08E-01	0.04	8.42E-01	-0.01
DOCK1	rs10829608	3.35E-03	-0.09	**6.42E-05**	-0.15	3.66E-01	0.03	8.06E-01	-0.01
DOCK1	rs7098854	4.73E-03	-0.08	**9.91E-05**	-0.15	3.92E-01	0.03	8.29E-01	-0.01

7.7 Appendix: Transcriptomic analysis following childhood influenza vaccination

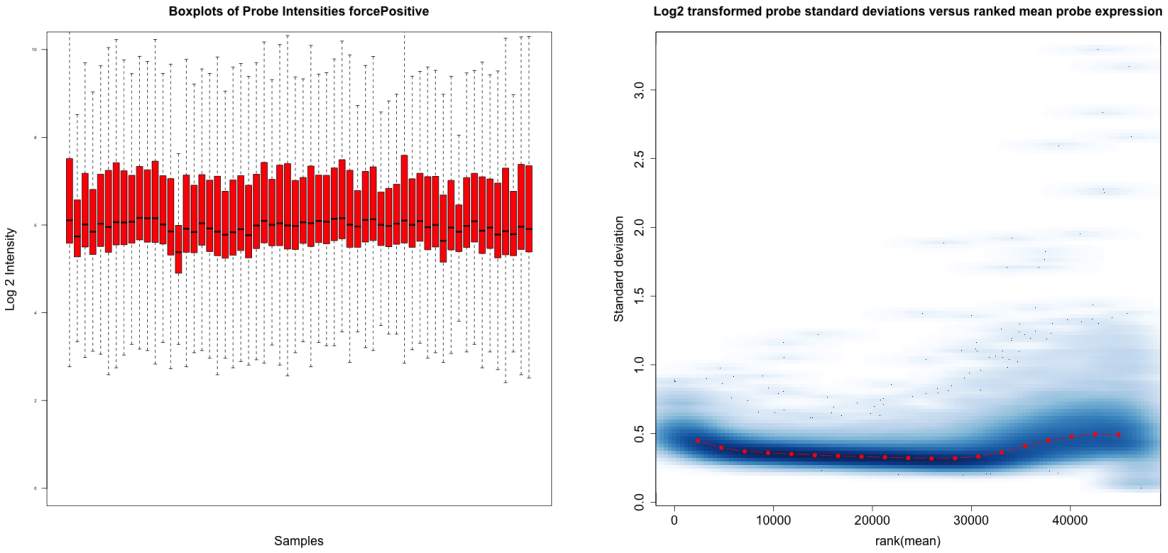
7.7.1 Gene expression following a single dose of seasonal trivalent inactivated influenza vaccine

7.7.1.1 Power calculations

Power calculations incorporated a variance estimate derived from a publicly available childhood gene expression dataset downloaded from the Gene Expression Omnibus `http://www.ncbi.nlm.nih.gov/geo/`. The 75^{th} percentile of the standard deviation of transcripts detected in a set of healthy caucasian child controls ($n = 5$) from a study conducted by Berry et al, was used as the generalisable standard deviation estimate (422).

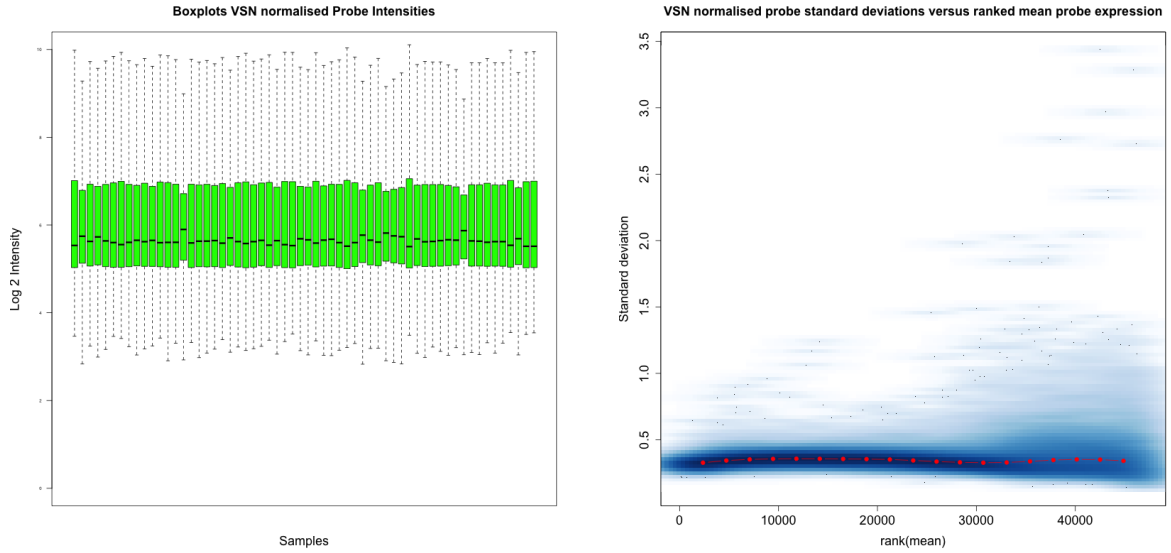
7.7.1.2 Gene expression microarray normalisation

Normalisation is necessary to make different microarrays comparable, given potential differences in samples processing, RNA quantities and labelling efficiency. Conceptually, microarray normalisation is similar to quantitative PCR normalisation, which normalises relative to levels of housekeeping genes assumed to be constant between experimental conditions. However, instead of comparing to particular housekeeping genes, microarray normalisation methods are based on the assumption that the majority of genes will not change with an intervention. Four different normalisation methods were assessed: Variance Stabilisation and Normalisation (VSN), Robust spline normalization (RSN), median normalisation and quantile normalisation. With the exception of the VSN method background subtracted data were forced positive using the 'lumiB' R package, and then to account for the linear relationship between standard deviation and mean spot intensity were $\log_2$ transformed (423). Data were then normalised using the respective normalisation methods.



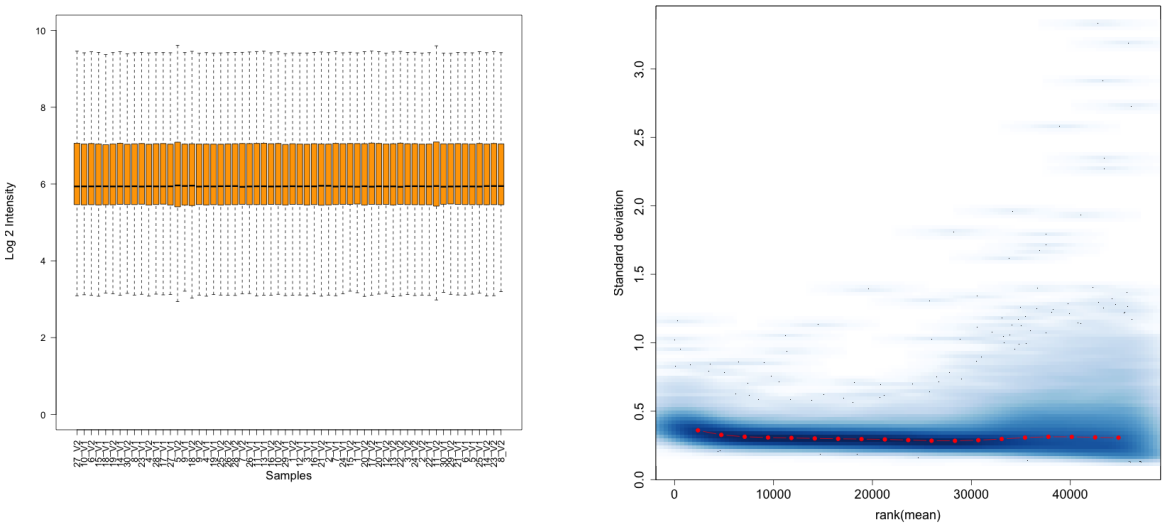
(a) (b)

Figure 7.16: RNA microarray quality control on $\log_2$ transformed background subtracted probe intensity data from the study of peripheral blood gene expression 21 days after immunisation with trivalent inactivated influenza vaccine. a) Box and whisker plots showing the distribution of $\log_2$ transformed RNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for $\log_2$ transformed RNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of standard deviation-mean dependence.



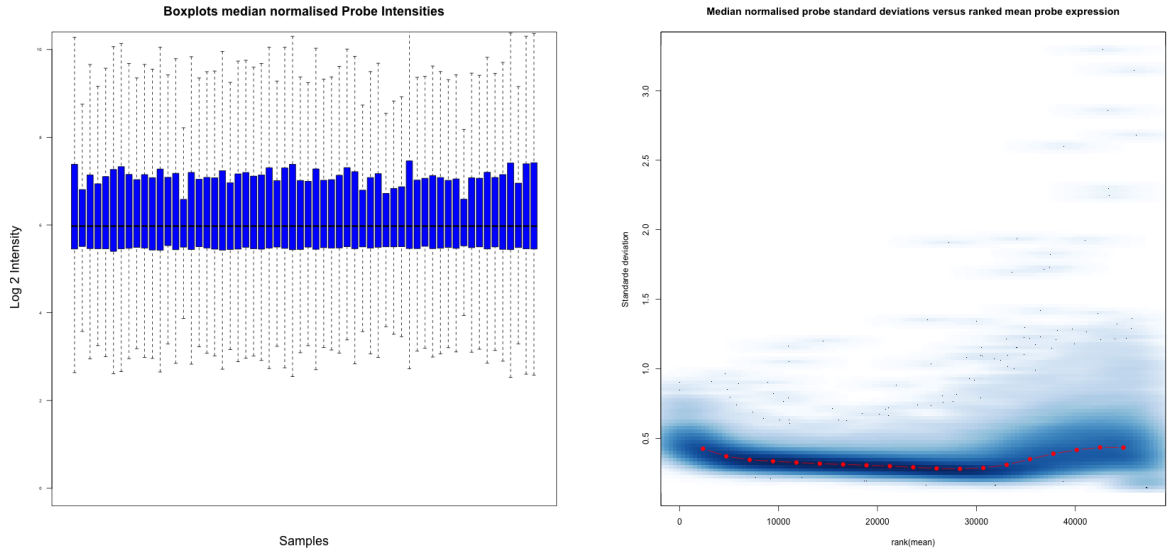
(a) (b)

Figure 7.17: RNA microarray quality control on variance stabilising normalised (VSN) background subtracted probe intensity data from the study of peripheral blood gene expression 21 days after immunisation with trivalent inactivated influenza vaccine. a) Box and whisker plots showing the distribution of VSN normalised RNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for VSN normalised RNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of standard deviation-mean dependence.



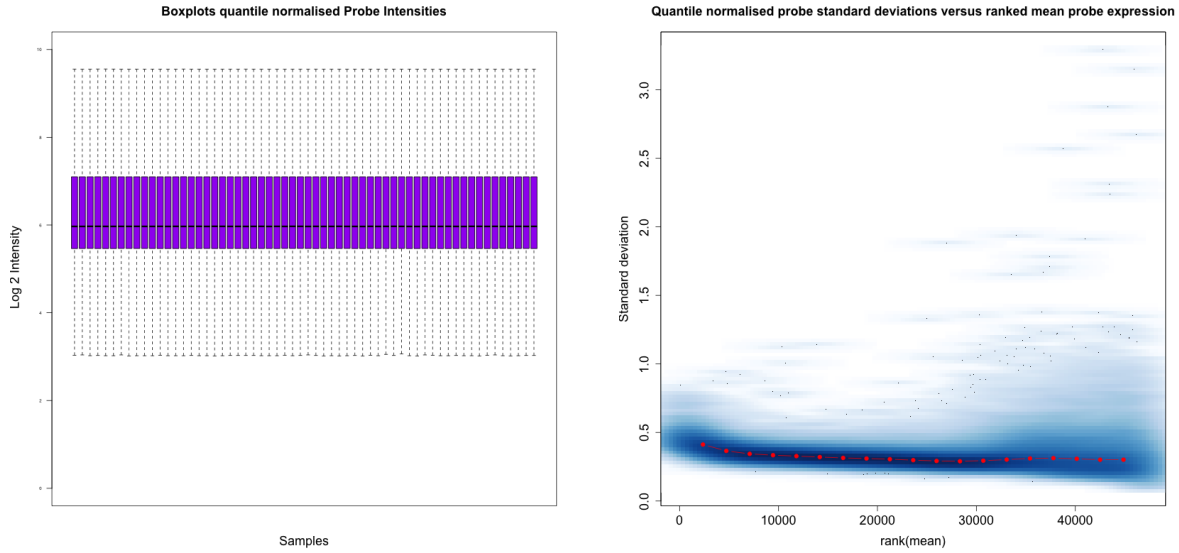
(a) (b)

Figure 7.18: RNA microarray quality control on robust spline normalised (RSN) background subtracted probe intensity data from the study of peripheral blood gene expression 21 days after immunisation with trivalent inactivated influenza vaccine. a) Box and whisker plots showing the distribution of RSN normalised RNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for RSN normalised RNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of standard deviation-mean dependence.



(a) (b)

Figure 7.19: RNA microarray quality control on median normalised background subtracted probe intensity data from the study of peripheral blood gene expression 21 days after immunisation with trivalent inactivated influenza vaccine. a) Box and whisker plots showing the distribution of median normalised RNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for median normalised RNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of dependence of standard deviation-mean dependence.



(a) (b)

Figure 7.20: RNA microarray quality control on quantile normalised background subtracted probe intensity data from the study of peripheral blood gene expression 21 days after immunisation with trivalent inactivated influenza vaccine. a) Box and whisker plots showing the distribution of quantile normalised RNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for quantile normalised RNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of dependence of standard deviation-mean dependence.

7.7.1.3 Sensitivity analysis

Table 7.29: Sensitivity analysis of the top 25 transcripts ranked in order of statistical evidence for differential expression 21 days after immunisation with trivalent inactivated influenza vaccine, with the two possible outlier samples identified in the quality control process excluded (subsection '5.2.5.2 Gene transcript microarray quality control'.)

Gene	Log FC	P Value	FDR	Definition
LOC643932	-0.25	0.000061	0.70	PREDICTED: Homo sapiens similar to 40S ribosomal protein S3a
GAPT	-0.31	0.000085	0.70	GRB2-binding adaptor protein
C14ORF4	0.24	0.000378	0.88	chromosome 14 open reading frame 4
SIT1	0.22	0.000437	0.88	signalling threshold regulating transmembrane adaptor 1
TRIP4	-0.17	0.000462	0.88	thyroid hormone receptor interactor 4
LOC729798	0.37	0.000605	0.88	similar to ribosomal protein L23a
LOC649210	-0.33	0.000776	0.88	similar to Ig lambda chain V region 4A precursor
CHI3L1	0.39	0.000820	0.88	chitinase 3-like 1
SFI1	0.23	0.000974	0.88	Sfi1 homolog
ASMTL	0.22	0.001012	0.88	acetylserotonin O-methyltransferase-like
LOC728484	0.27	0.001022	0.88	miscellaneous RNA
RPL23AP13	0.28	0.001051	0.88	ribosomal protein L23a pseudogene 13
LOC390183	0.27	0.001070	0.88	miscellaneous RNA
TOR2A	0.19	0.001076	0.88	torsin family 2, member A
GAPT	-0.23	0.001129	0.88	GRB2-binding adaptor protein
PGRMC1	-0.25	0.001131	0.88	progesterone receptor membrane component 1
LOC100132317	0.31	0.001137	0.88	similar to C20orf69 protein
LOC650254	-0.19	0.001173	0.88	hypothetical protein FLJ40722, transcript variant 1
SLC45A4	0.20	0.001240	0.88	solute carrier family 45, member 4
PLCG1	0.29	0.001298	0.88	phospholipase C, gamma 1
TYSND1	0.21	0.001300	0.88	trypsin domain containing 1
LOC649946	0.25	0.001408	0.88	ribosomal protein L23a pseudogene
VEGFB	0.18	0.001662	0.88	vascular endothelial growth factor B
C16ORF13	-0.21	0.001666	0.88	chromosome 16 open reading frame 13
CYSLTR1	-0.28	0.001690	0.88	cysteinyl leukotriene receptor 1

FC= Fold change ; FDR= False discovery rate (Benjamini-Hochberg); ORF= open reading frame

7.7.1.4 Gene expression study participant information

Table 7.30: Demographics of children in the study of peripheral blood gene expression profiles 21 days following trivalent inactivated influenza vaccine (TIV).

Randomised number	Sex	Previous influenza vaccine	Ethnicity	Age at vaccination (months)	Days since TIV
1	M	BAXTER	Caucasian	83	21
2	M	GSK	Caucasian	129	21
3	M	BAXTER	Caucasian	140	21
4	F	BAXTER	Caucasian	169	21
5	F	GSK	Caucasian	108	21
6	M	BAXTER	Caucasian	121	21
7	M	BAXTER	Caucasian	115	21
8	M	GSK	Caucasian	152	21
9	F	GSK	Caucasian	103	21
10	M	GSK	Caucasian	82	20
11	F	GSK	Caucasian	140	21
12	F	GSK	Caucasian	64	21
13	F	GSK	Caucasian	141	20
14	F	BAXTER	Caucasian	156	20
15	M	BAXTER	Caucasian	126	20
16	F	BAXTER	Caucasian	167	21
17	F	BAXTER	Caucasian	122	17
18	M	GSK	Caucasian	99	20
19	F	GSK	Caucasian	151	21
20	F	GSK	Caucasian	152	20
21	M	BAXTER	Caucasian	158	23
22	F	GSK	Caucasian	139	23
23	F	BAXTER	Caucasian	98	21
24	F	GSK	Caucasian	81	21
25	M	GSK	Caucasian	69	20
26	M	GSK	Caucasian	94	20
27	F	GSK	Caucasian	51	21
28	M	BAXTER	Caucasian	55	21
29	M	GSK	Caucasian	153	21
30	F	GSK	Caucasian	89	21

M= male, F= female

7.7.1.5 Gene Set Enrichment Analysis

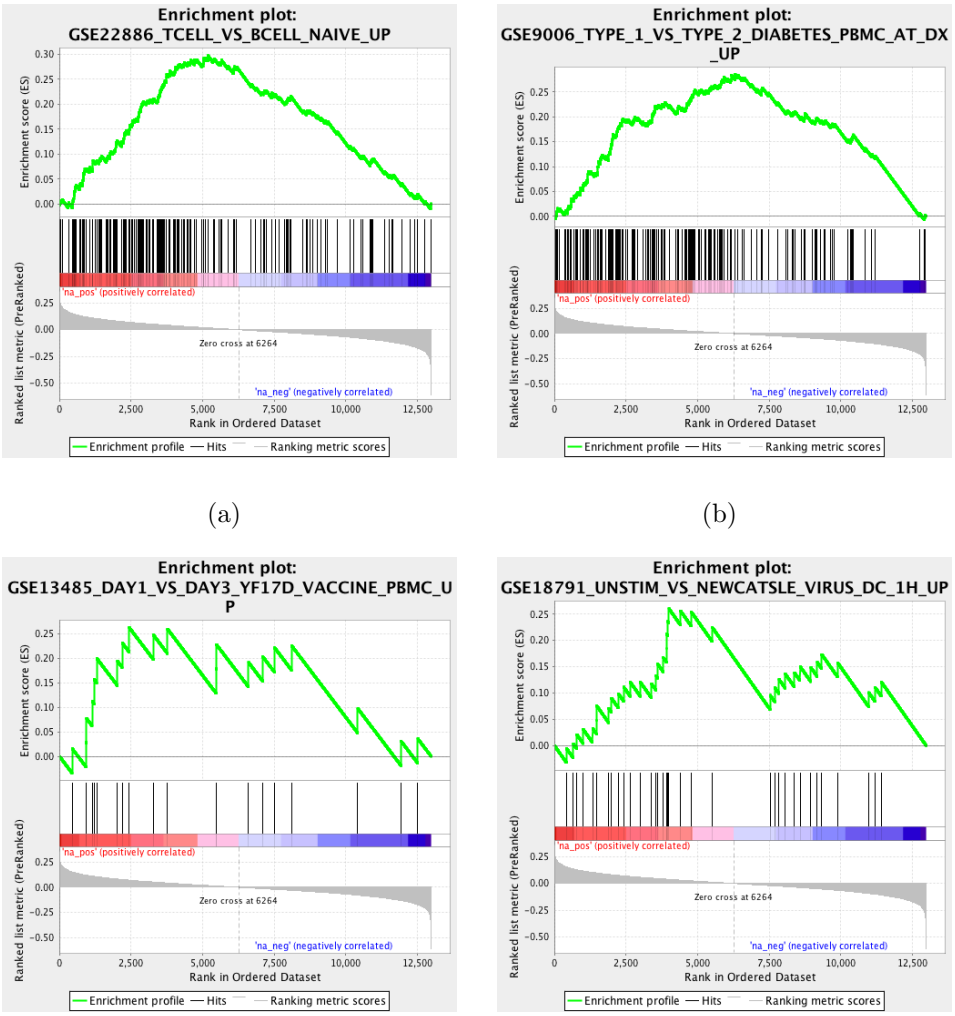
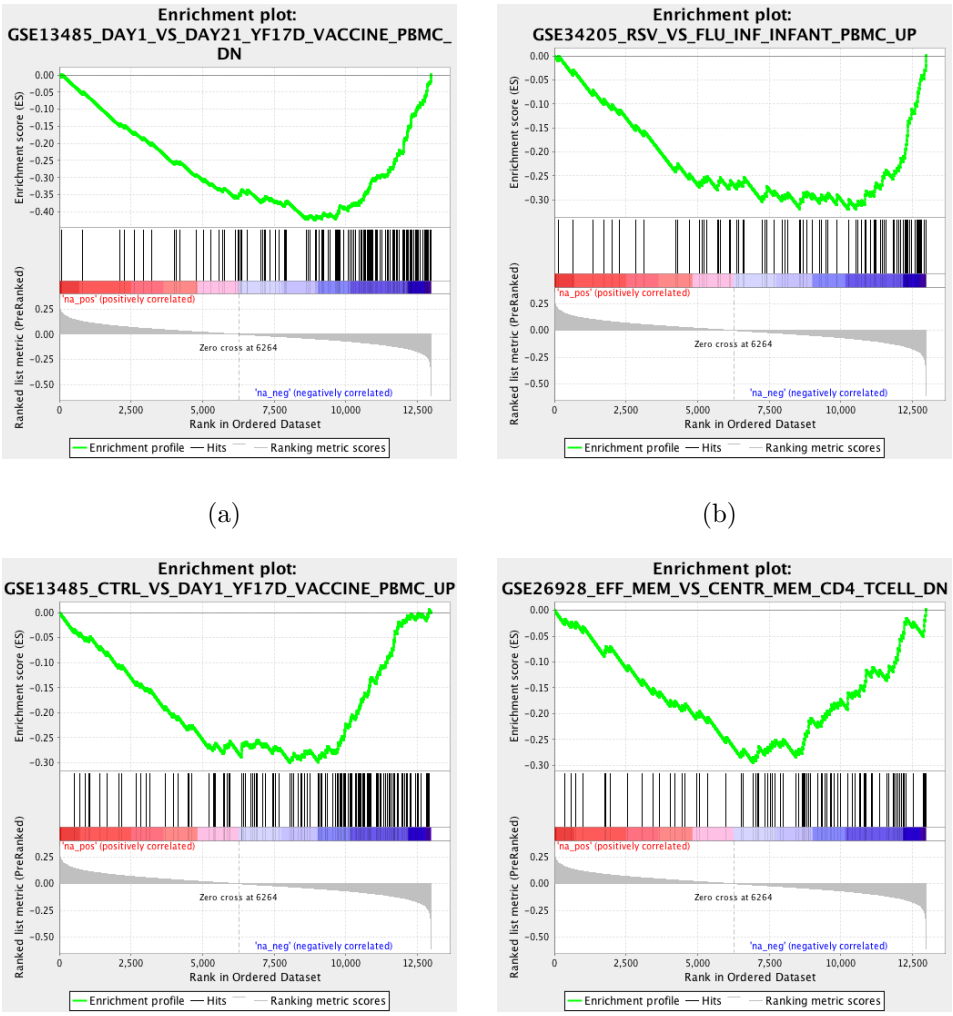


Figure 7.21: The top four upregulated gene sets from gene set enrichment analysis conducted on peripheral blood gene expression profiles 21 days after immunisation with trivalent inactivated influenza vaccine.



(a) (b)

(c) (d)

Figure 7.22: The top four downregulated gene sets from gene set enrichment analysis conducted on peripheral blood gene expression profiles 21 days after immunisation with trivalent inactivated influenza vaccine.

7.7.1.6 Ingenuity Pathway Analysis

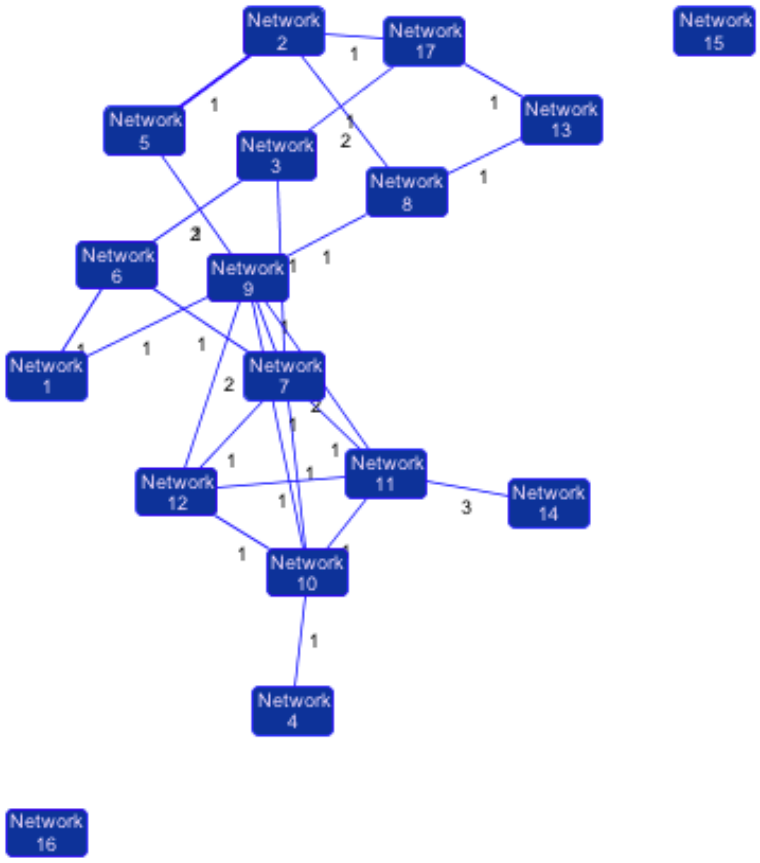


Figure 7.23: The relatedness of networks derived from Ingenuity® Pathway analysis on peripheral blood gene expression profiles 21 days after immunisation with trivalent inactivated influenza vaccine. Networks that share constituents are connected, with the degree of overlap corresponding to the denoted number.

7.7.1.7 Relationship between gene expression and immunological responses to TIV

Table 7.31: The networks produced by Ingenuity® Pathway Analysis of genes with differential expression in peripheral blood 21 days following immunisation with trivalent inactivated influenza vaccine associated ($p<0.05$) with post-vaccination haemagglutination inhibition assay titres.

Score	Focus molecules	Diseases and Functions
34	17	Developmental Disorder, Hereditary Disorder, Skeletal and Muscular Disorders
31	16	RNA Damage and Repair, Cancer, Gastrointestinal Disease
29	15	Cell Signalling, Cell Death and Survival, Renal Necrosis/Cell Death
25	14	Cell-To-Cell Signalling and Interaction, Cellular Movement, Haematological System Development and Function
23	13	Cancer, Cell Death and Survival, Tumour Morphology
19	11	Lymphoid Tissue Structure and Development, Cardiovascular Disease, Cardiovascular System Development and Function
9	6	Cell-To-Cell Signalling and Interaction, Haematological System Development and Function, Immune Cell Trafficking
2	1	Cancer, Cell Death and Survival, Cellular Development

Table 7.32: The top 25 genes with fold-changes 21 days following immunisation with trivalent inactivated influenza vaccine associated with post-vaccination T-cell immunity to 2009 H1N1 nucleoprotein (NP).

Gene	t-value	p-value
MCRS1	4.00	3.97E-04
LOC653545	-3.95	4.57E-04
LBH	3.93	4.84E-04
HS.162734	-3.89	5.41E-04
LOC647691	-3.74	8.10E-04
ACTG1	3.64	1.05E-03
MAX	3.57	1.27E-03
ZFP36L2	3.54	1.38E-03
TNFRSF25	3.52	1.44E-03
RALGPS2	-3.51	1.48E-03
SLC25A20	3.48	1.60E-03
LOC642817	3.43	1.82E-03
MGAT2	3.40	1.98E-03
ABCA7	-3.38	2.09E-03
CCT7	3.37	2.14E-03
UTP23	-3.33	2.35E-03
LOC100133760	-3.33	2.36E-03
PSMC4	3.32	2.46E-03
LOC286016	-3.31	2.48E-03
LOC642161	3.28	2.71E-03
ERP29	3.24	2.95E-03
DNAJB6	3.21	3.21E-03
SFRS2B	3.18	3.49E-03
LETMD1	3.16	3.69E-03
CENPV	-3.16	3.71E-03

Table 7.33: The networks produced by Ingenuity Pathway Analysis conducted on genes with fold-changes associated (p <0.015) with 2009 H1N1 nucleoprotein (NP) ELISpots following trivalent inactivated influenza vaccine (TIV)..

Score	Focus molecules	Diseases and functions
33	17	Organismal Development, Cellular Growth and Proliferation, Nervous System Development and Function
32	18	Cell-To-Cell Signalling and Interaction, Haematological System Development and Function, Immunological Disease
31	16	Cell Morphology, Organ Morphology, Reproductive System Development and Function
31	16	Developmental Disorder, Hereditary Disorder, Metabolic Disease
28	15	Cellular Assembly and Organisation, Developmental Disorder, Haematological Disease
27	15	Cell Morphology, Cellular Assembly and Organisation, Cellular Function and Maintenance
2	1	Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder
2	1	Developmental Disorder, Hereditary Disorder, Skeletal and Muscular Disorders

Table 7.34: The top 25 genes with fold-changes 21 days following immunisation with trivalent inactivated influenza vaccine associated with post-vaccination T-cell immunity to 2009 H1N1 matrix protein 1 (M1).

Gene	t-value	p-value
JAK2	5.39	8.00E-06
LOC729513	-4.77	4.80E-05
TMOD2	4.69	5.80E-05
C1ORF19	4.52	9.40E-05
ZNF124	4.44	1.20E-04
COG6	4.39	1.35E-04
ATF2	4.29	1.80E-04
NPCDR1	-4.26	1.94E-04
ANKRD49	4.23	2.14E-04
EIF5	-4.21	2.26E-04
CHML	-4.15	2.59E-04
NOTCH2NL	-4.10	3.00E-04
FNTA	4.05	3.42E-04
TOMM20	4.05	3.42E-04
HS.443853	-4.05	3.46E-04
STXBP3	4.05	3.47E-04
TRIM33	4.05	3.49E-04
OGFRL1	4.02	3.74E-04
PPM1A	4.02	3.77E-04
ZNF800	3.98	4.15E-04
HS.555252	-3.98	4.21E-04
HSPA14	3.97	4.29E-04
CCDC117	3.97	4.32E-04
RHEB	3.97	4.34E-04
ZDHHC13	3.96	4.42E-04

Table 7.35: The networks produced by Ingenuity Pathway Analysis conducted on genes with fold-changes that associated ($p < 0.015$) with 2009 H1N1 matrix protein 1 (M1) ELISpots following trivalent inactivated influenza vaccine.

Score	Focus molecules	Diseases and functions
48	33	RNA Post-Transcriptional Modification, Infectious Disease, Cellular Assembly and Organisation
45	32	Molecular Transport, Protein Trafficking, Cellular Assembly and Organisation
38	29	Post-Translational Modification, Cell Morphology, Hair and Skin Development and Function
34	28	Protein Synthesis, Cell-To-Cell Signalling and Interaction, Cell Signalling
34	27	Cellular Development, Haematological System Development and Function, Haematopoiesis
34	27	Cellular Assembly and Organisation, Gene Expression, Cancer
33	27	Molecular Transport, Lymphoid Tissue Structure and Development, Organ Morphology
32	26	Cancer, Haematological Disease, Small Molecule Biochemistry
32	26	DNA Replication, Recombination, and Repair, Energy Production, Nucleic Acid Metabolism
31	26	RNA Damage and Repair, Cellular Assembly and Organisation, Cellular Compromise
30	25	Cancer, Cell Morphology, Developmental Disorder
27	23	Post-Translational Modification, Cell Death and Survival, Infectious Disease
25	22	Post-Translational Modification, Cell Cycle, DNA Replication, Recombination, and Repair
23	16	Developmental Disorder, Hereditary Disorder, Neurological Disease
23	21	Carbohydrate Metabolism, Lipid Metabolism, Post-Translational Modification
23	24	Protein Synthesis, Cellular Development, Respiratory System Development and Function
21	20	Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organisation, Tissue Development
21	20	Auditory Disease, Hereditary Disorder, Immunological Disease
21	20	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry
20	19	Cell Cycle, Cellular Assembly and Organisation, Cell Morphology
20	19	Cell Signalling, Cellular Movement, Haematological System Development and Function
20	19	Cell Cycle, Dermatological Diseases and Conditions, Hereditary Disorder
20	19	Cell Cycle, Cell Morphology, Cell-mediated Immune Response
19	19	Cell Cycle, Developmental Disorder, Gastrointestinal Disease
18	18	Increased Levels of AST, Dermatological Diseases and Conditions, Infectious Disease

Table 7.36: The top 25 genes with fold-changes 21 days following immunisation with trivalent inactivated influenza vaccine associated with post-vaccination T-cell immunity to 2009 H1N1 nonstructural protein 1 (NS1).

Gene ID	t-value	p-value
ZNF330	4.42	1.26E-04
HS.162734	-4.14	2.73E-04
MOSPD2	-3.89	5.40E-04
ABCA7	-3.85	5.94E-04
B3GNT8	-3.63	1.09E-03
ARL5B	-3.60	1.16E-03
RBM41	3.50	1.51E-03
MARS2	3.49	1.55E-03
FBXL17	3.46	1.69E-03
ZRANB2	-3.45	1.75E-03
ACTG1	3.39	2.05E-03
G3BP2	3.36	2.19E-03
ZCCHC6	-3.33	2.35E-03
LOC728188	3.31	2.46E-03
ACOT2	3.30	2.53E-03
C1ORF55	3.30	2.59E-03
BRWD1	-3.28	2.68E-03
HS.573541	-3.27	2.75E-03
DKFZP761P0423	3.23	3.10E-03
OPN3	3.19	3.43E-03
LOC648695	3.15	3.73E-03
VPS13C	-3.15	3.73E-03
ARID4B	-3.14	3.80E-03
B3GAT3	-3.14	3.85E-03
MAP2K4	-3.13	3.98E-03

Table 7.37: The networks produced by Ingenuity Pathway Analysis conducted on genes with fold-changes associated ($p<0.015$) with 2009 H1N1 nonstructural protein 1 (NS1) ELISpots following trivalent inactivated influenza vaccine.

Score	Focus.Molecules	Diseases and functions
40	20	Cell Cycle, Cellular Assembly and Organisation, Cellular Function and Maintenance
38	19	Cell Morphology, Cell Death and Survival, Embryonic Development
30	16	DNA Replication, Recombination, and Repair, Gene Expression, Cell Morphology
30	16	Cell Signalling, Post-Translational Modification, Protein Synthesis
29	16	Embryonic Development, Organ Development, Organ Morphology
23	13	Cell Cycle, Cell-To-Cell Signalling and Interaction, Haematological System Development and Function
11	7	Cancer, Organismal Injury and Abnormalities, Reproductive System Disease
2	1	Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder

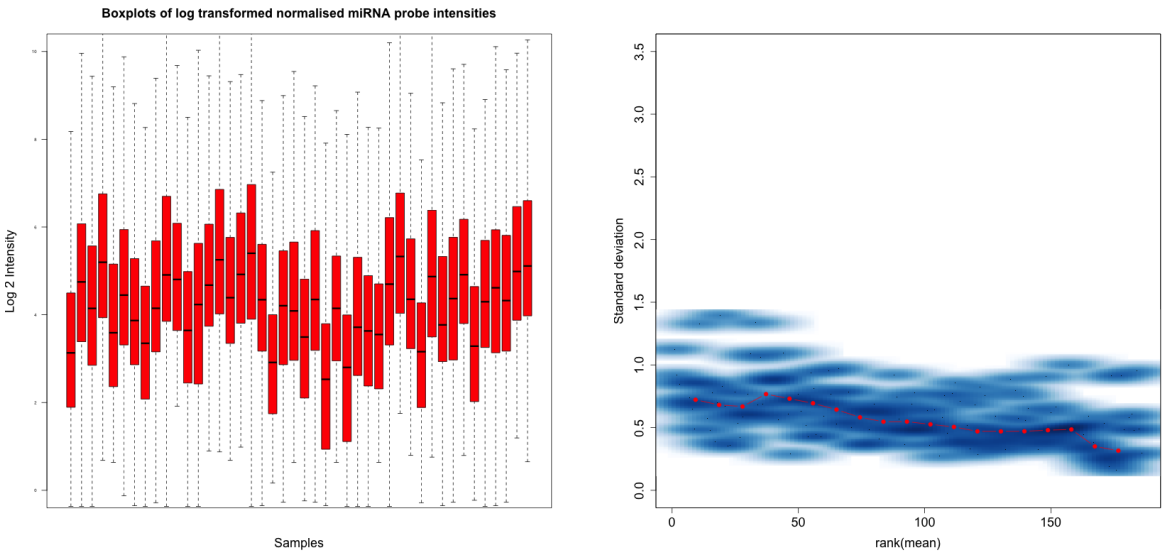
7.7.1.8 Comparison of gene changes following trivalent inactivated influenza vaccine between individuals who previously received adjuvanted or non-adjuvanted 2009 pandemic influenza vaccination

Table 7.38: The networks produced by Ingenuity Pathway Analysis conducted using genes with p-values less than <0.015 in the linear regression model between gene fold-change following trivalent inactivated influenza vaccine and which of the pandemic vaccines was administered the previous year.

Score	Focus molecules	Top diseases and functions
26	12	DNA Replication, Recombination, and Repair, Cell Morphology, Cellular Function and Maintenance
24	11	Cancer, Dermatological Diseases and Conditions, Hereditary Disorder
24	11	Cellular Movement, Connective Tissue Development and Function, Cell-To-Cell Signalling and Interaction
13	7	Biliary Hyperplasia, Cancer, Developmental Disorder

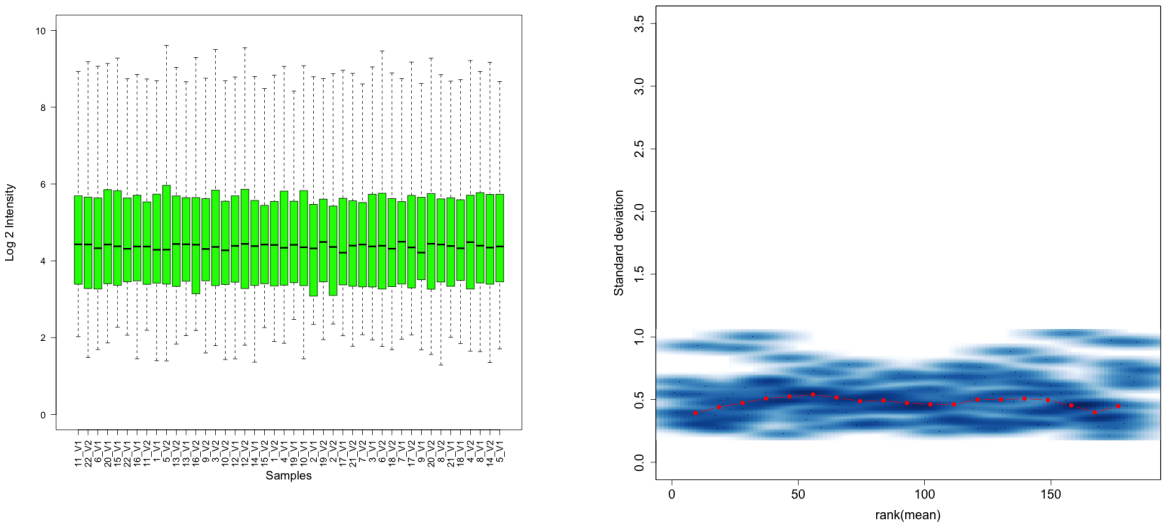
7.7.2 Assessment of serum microRNA following childhood immunisation with pandemic influenza vaccine

7.7.2.1 MicroRNA microarray normalisation



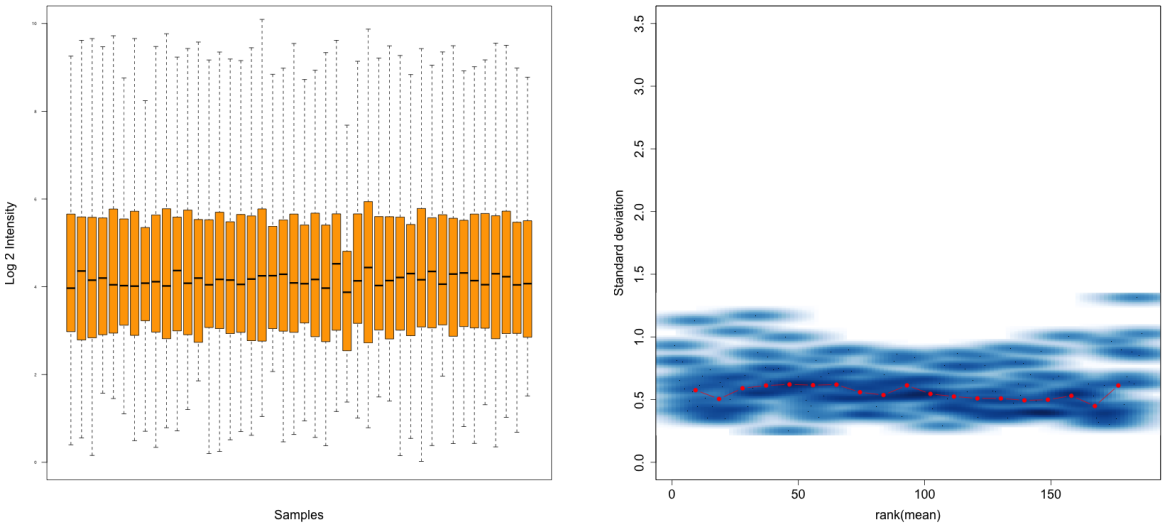
(a) (b)

Figure 7.24: MicroRNA microarray quality control on $\log_2$ transformed patched probe intensity data from the study of serum microRNA following immunisation with pandemic influenza vaccine. a) Box and whisker plots showing the distribution of $\log_2$ transformed MicroRNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for $\log_2$ transformed microRNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of standard deviation-mean dependence.



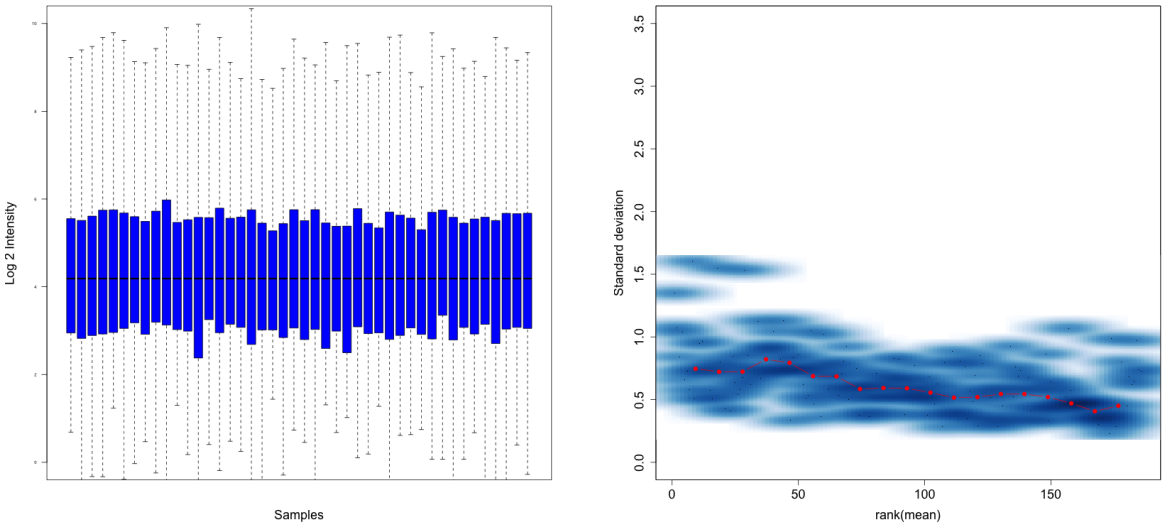
(a) (b)

Figure 7.25: MicroRNA microarray quality control on variance stabilising normalised (VSN) patched probe intensity data from the study of serum microRNA following immunisation with pandemic influenza vaccine. a) Box and whisker plots showing the distribution of VSN normalised MicroRNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for VSN normalised microRNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of dependence of of standard deviation-mean dependence.



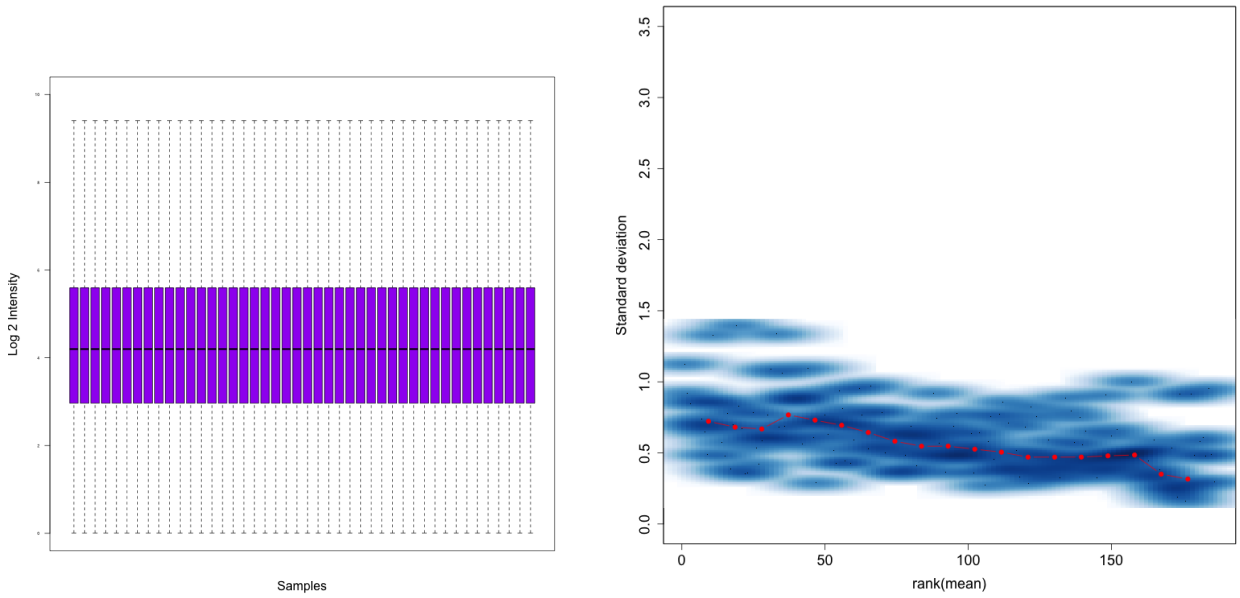
(a) (b)

Figure 7.26: MicroRNA microarray quality control on robust spline normalised (RSN) background patched intensity data from the study of serum microRNA following immunisation with pandemic influenza vaccine. a) Box and whisker plots showing the distribution of RSN normalised MicroRNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for RSN normalised microRNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of dependence of standard deviation-mean dependence.



(a) (b)

Figure 7.27: MicroRNA microarray quality control on median normalised patched probe intensity data from the study of serum microRNA following immunisation with pandemic influenza vaccine. a) Box and whisker plots showing the distribution of median normalised MicroRNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for median normalised microRNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of dependence of the standard deviation mean.



(a) (b)

Figure 7.28: MicroRNA microarray quality control on quantile normalised patched probe intensity data from the study of serum microRNA following immunisation with pandemic influenza vaccine. a) Box and whisker plots showing the distribution of quantile normalised MicroRNA feature intensities. The boxes represent the first, second (median) and third quartiles. The whiskers represent the rest of the distribution, terminating at the maximum and minimum intensity values. b) Standard deviations against the ranked mean probe intensities for quantile normalised microRNA data. The red dots show the running median estimator, which is approximately horizontal as there is little evident of standard deviation-mean dependence.

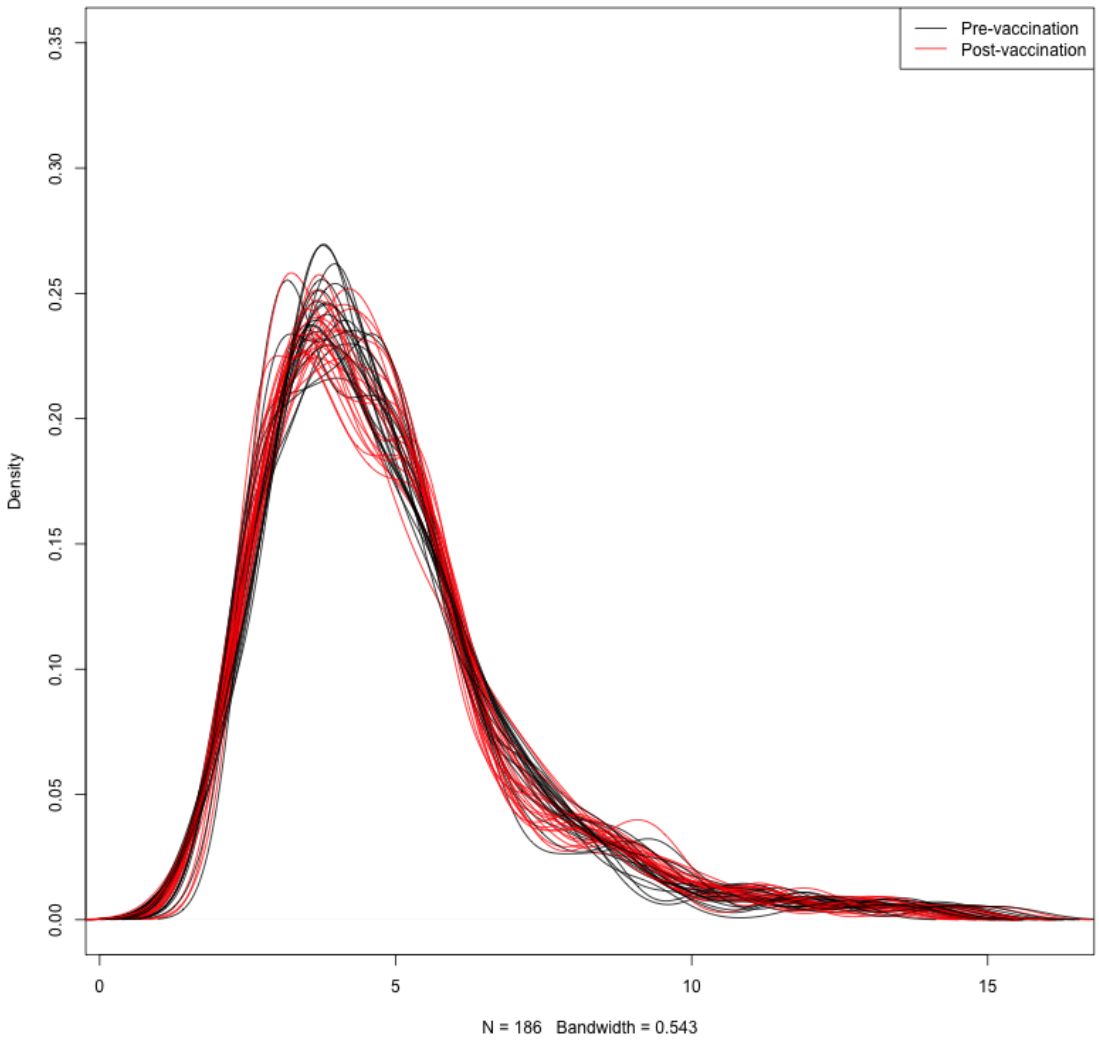


Figure 7.29: Density of microRNA probes with varying variance stabilising normalised (VSN) patched expression values for each sample from the study of serum microRNA levels following immunisation with pandemic influenza vaccine

7.7.2.2 MicroRNA study participant demographics

Table 7.39: Participant demographics in the study of serum microRNA levels following immunisation with pandemic influenza vaccine.

Random No.	Visit	Time to spin*	Sex	Vaccine	Age at vaccination#	Ethnicity	Days post-vaccination
1	1	01:07	M	Baxter	109	Caucasian	-
	3	02:24					25
2	1	02:25	F	Baxter	142	Caucasian	-
	3	01:22	F				21
3	1	03:15	F	Baxter	15	Caucasian	-
	2	18:55	F				28
4	1	01:55	F	GSK	58	Mixed	
	3	01:25	F			(Asian & Caucasian)	24
5	1	02:05	M	Baxter	38	Caucasian	-
	3	02:30	M				25
6	1	02:20	F	Baxter	107	Caucasian	-
	3	01:00	F				21
7	1	02:40	M	Baxter	112	Caucasian	-
	3	01:30	M				21
8	1	01:48	F	GSK	89	Caucasian	-
	3	02:55	F				26
9	1	02:33	M	Baxter	69	Caucasian	-
	3	01:25	M				25
10	1	02:30	M	Baxter	41	Caucasian	-
	3	02:05	M				21
11	1	02:00	M	Baxter	145	Caucasian	-
	3	16:30	M				25
12	1	01:37	M	GSK	45	Caucasian	-
	3	03:00	M				26
13	1	02:50	F	Baxter	71	Caucasian	-
	3	00:45	F				21
14	1	01:29	M	GSK	115	Caucasian	-
	3	02:50	M				26
15	1	01:45	F	GSK	125	Caucasian	-
	3	16:35	F				25
16	1	02:10	M	GSK	18	Caucasian	-
	3	02:00	M				29
17	1	01:30	F	GSK	11	Caucasian	-
	3	01:05	F				25
18	1	02:05	M	Baxter	124	Caucasian	-
	3	01:00	M				21
19	1	23:10	F	GSK	67	Caucasian	-
	3	00:40	F				28
20	1	02:05	F	Baxter	21	Caucasian	-
	3	01:22	F				18
21	1	00:59	F	GSK	51	Caucasian	-
	3	02:05	F				25
22	1	00:50	F	Baxter	15	Mixed	-
	3	18:05	F			(Asian & Caucasian)	25

Random No. = Randomisation number , *=hh:mm, V1= Visit 1(vaccination), #= months)

7.7.2.3 Table of detected microRNA in serum

Table 7.40: MicroRNAs detected in at least 65% of serum samples in the study assessing microRNA levels following immunisation with pandemic influenza vaccine.

hsa.miR.575	hsa.miR.4298	hsv1.miR.H6.3p	hsa.miR.486.5p	hsa.miR.19a
hsa-miR-1207-5p	hsa-miR-1268	hsa-miR-766	hsa-miR-21	hsa-miR-3188
hsa-miR-1202	hsa-miR-129*	hsa-miR-4281	hsa-miR-199a-3p	ebv-miR-BART12
hsa-miR-4270	hsa-miR-3180-5p	hsa-miR-92a	hsa-miR-15a	hsa-miR-1249
hsa-miR-638	hsa-miR-572	hsa-miR-4259	hsa-miR-3663-3p	hsa-miR-29c
hsa-miR-483-5p	hsv1-miR-H8	hsa-miR-142-5p	hbv-miR-B2RC	hsa-miR-25
hsa-miR-2861	hsa-miR-933	hsa-miR-1260b	hsa-miR-27a	hsa-miR-1274b_v16.0
hsa-miR-3679-5p	hsa-miR-940	hsa-miR-151-5p	hsa-miR-1224-5p	hsa-miR-24
kshv-miR-K12-3	hsa-miR-1228	hsa-miR-4323	hsa-miR-762	hsa-miR-20a
hsa-miR-3141	hsa-miR-1471	hsa-miR-4327	ebv-miR-BART16	hsa-miR-30e
hsa-miR-3196	hsa-miR-26a	hsa-miR-32*	hsa-miR-130b	hsa-miR-1246
hsv1-miR-H17	hsv1-miR-H6-5p	hsa-miR-320c	hsa-miR-144	hsa-miR-4271
hsa-miR-30b	kshv-miR-K12-8*	hsa-miR-146a	hsa-miR-320a	hsa-miR-3676
hsa-miR-3656	hsa-miR-494	hsa-miR-4286	hsa-miR-3149	hsa-miR-140-3p
hsa-miR-1915	hsa-miR-423-5p	hsa-miR-150*	hsa-miR-23a	hsa-miR-20b
hcmv-miR-UL70-3p	hsa-miR-3665	hsa-miR-140-5p	hsa-miR-877*	hsa-miR-3162-5p
hsv2-miR-H25	hsa-miR-1238	hsa-miR-466	hsa-miR-107	hsa-miR-424
hsa-miR-642b	hsa-miR-30d	hsa-miR-29b	hsa-miR-342-3p	hsa-miR-93
hsa-miR-142-3p	hsa-miR-1225-3p	hsa-miR-185	hsa-miR-101	hsv2-miR-H6
hsv2-miR-H10	hsa-miR-103a	hsa-miR-151-3p	hsa-miR-296-5p	hsa-miR-192
hsa-miR-671-5p	hsa-miR-22	hsa-miR-4306	hsa-miR-17	hsa-miR-150
hsa-miR-129-3p	hsa-miR-199a-5p	hsa-miR-15b	hsa-miR-451	hsa-miR-595
hsa-miR-634	hsa-miR-130a	hsa-miR-221	hsa-miR-630	hsa-miR-498
hsa-miR-1225-5p	hsa-miR-1825	hsv1-miR-H7*	hsa-miR-361-5p	hsa-miR-590-5p
hsa-miR-3648	hsa-miR-134	hsa-miR-718	hsv2-miR-H24	hsa-miR-1290
hsa-miR-1280	hsa-miR-1275	hsa-miR-320b	hsa-miR-106b	hsa-miR-1914*
hsa-miR-3138	hsa-miR-720	hsa-miR-3195	hsa-miR-19b	hsa-miR-135a*
hsa-miR-23b	hsa-miR-188-5p	hsa-miR-191*	hsa-miR-148b	hsa-miR-328
hsa-miR-939	hsa-miR-320d	hsa-miR-320e	hsa-miR-3610	hsa-miR-197
hsa-miR-550a	kshv-miR-K12-12*	hsa-miR-126	hsa-miR-16	hsa-miR-574-3p
hsa-miR-765	hsa-miR-1281	hsa-miR-27b	hsa-miR-324-3p	hsv1-miR-H1*
hsa-miR-4257	hsa-miR-483-3p	hsa-miR-29a	hsa-miR-26b	
hsa-miR-1234	hsa-miR-148a	hsa-miR-425	hsa-miR-574-5p	
hsa-miR-4284	hsa-miR-223	hsa-miR-1260	hsa-miR-4313	

hcmv =human cytomegalovirus, hsv= herpes simplex virus, kshv= karposi sarcoma-associated herpesvirus
hsa= homo sapiens, miR= mature microRNA sequence, *= minor product of the miRNA hairpin, 5p= 5′, 3p= 3′

Table 7.41: Results of the Ingenuity Pathway Analysis microRNA target filter on the 22 differentially expressed (false discovery rate <0.05) serum microRNAs following immunisation with pandemic influenza vaccine. Ingenuity Knowledge Base microRNA target data were filtered for only experimental results and mRNA involved in cellular and/or humoral immune response

MicroRNA identification	Log_2 fold-change	Database source	mRNA target
hsa-miR-3679-5p	0.43	Ingenuity Expert Findings	MAPK1
hsa-miR-142-3p	-0.37	TargetScan Human,miRecords	ADCY9
hsa-miR-142-3p	-0.37	Ingenuity Expert Findings	BCL2L1
hsa-miR-30b	-0.32	TarBase,TargetScan Human	AP2A1
hsa-miR-30b	-0.32	TarBase,TargetScan Human	ATP2A2
hsa-miR-30b	-0.32	Ingenuity Expert Findings,TargetScan Human	BCL6
hsa-miR-30b	-0.32	TarBase	F2
hsa-miR-30b	-0.32	TarBase,TargetScan Human,miRecords	GNAI2
hsa-miR-30b	-0.32	TarBase	ITGA2
hsa-miR-30b	-0.32	TarBase	JUN
hsa-miR-30b	-0.32	TarBase	LMNB2
hsa-miR-30b	-0.32	TarBase	MET
hsa-miR-30b	-0.32	TarBase	PPP2R4
hsa-miR-30b	-0.32	TarBase,TargetScan Human	PPP3CA
hsa-miR-30b	-0.32	miRecords	TP53
hsa-miR-642b	0.26	Ingenuity Expert Findings	Sos

References

[1] WHO, UNICEF, and World Bank. *State of the world's vaccines and immunization*. World Health Organization, Geneva, World Health Organization, 3rd edition, 2009.

[2] A. J. Roome, S. J. Walsh, M. L. Cartter, and J. L. Hadler. **Hepatitis B vaccine responsiveness in Connecticut public safety personnel.** *JAMA*, **270**(24):2931–4, 1993.

[3] KP Perrett, AP Winter, E Kibwana, C Jin, TM John, LM Yu, R Borrow, N Curtis, and AJ Pollard. **Antibody persistence after serogroup C meningococcal conjugate immunization of United kingdom primary-school children in 1999-2000 and response to a booster: a phase 4 clinical trial.** *Clin Infect Dis*, **50**(12):1601–10, 2010.

[4] H Van Loveren, JG Van Amsterdam, RJ Vandebriel, TG Kimman, HC Rümke, PS Steerenberg, and JG Vos. **Vaccine-induced antibody responses as parameters of the influence of endogenous and environmental factors.** *Environ Health Perspect*, **109**(8):757–64, 2001.

[5] D. O'Connor and A. J. Pollard. **Characterizing vaccine responses using host genomic and transcriptomic analysis**. *Clin Infect Dis*, **57**(6):860–9, 2013.

[6] T Höhler, E Reuss, N Evers, E Dietrich, C Rittner, CM Freitag, J Vollmar, PM Schneider, and R Fimmers. **Differential genetic determination of immune responsiveness to hepatitis B surface antigen and to hepatitis A virus: a vaccination study in twins.** *Lancet*, **360**(9338):991–5, 2002.

[7] MJ Newport, T Goetghebuer, HA Weiss, H Whittle, CA Siegrist, and A and Marchant. **Genetic regulation of immune responses to vaccines in early life.** *Genes Immun*, **5**(2):122–9, 2004.

[8] YC Lee, MJ Newport, T Goetghebuer, CA Siegrist, HA Weiss, AJ Pollard, and A and Marchant. **Influence of genetic and environmental factors on the immunogenicity of Hib vaccine in Gambian twins.** *Vaccine*, **24**(25):5335–40, 2006.

[9] HB Konradsen, J Henrichsen, H Wachmann, and N Holm. **The influence of genetic factors on the immune response as judged by pneumococcal vaccination of mono- and dizygotic Caucasian twins.** *Clin Exp Immunol*, **92**(3):532–6, 1993.

[10] K. Yan, W. Cai, F. Cao, H. Sun, S. Chen, R. Xu, X. Wei, X. Shi, and W. Yan. **Genetic effects have a dominant role on poor responses to infant vaccination to hepatitis B virus**. *J Hum Genet*, **58**(5):293–7, 2013.

[11] Michael C. Neale and Lon R. Cardon. *Methodology for genetic studies of twins and families*. Kluwer Academic published in cooperation with NATO Scientific Affairs Division, 1992. GB9242329 bnb 2212 by Michael C. Neale and Lon R. Cardon. Based on workshops taught at Leuven, Belgium in 1987, 1988 and 1991 and at Boulder, Colo., in 1990. North Atlantic Treaty Organization. Scientific Affairs Division.

[12] M. C. Neale and M. B. Miller. **The use of likelihood-based confidence intervals in genetic models.** *Behav Genet*, **27**(2):113–20, 1997.

[13] TA Manolio, FS Collins, NJ Cox, DB Goldstein, LA Hindorff, DJ Hunter, MI McCarthy, EM Ramos, LR Cardon, A Chakravarti, JH Cho, AE Guttmacher, A Kong, L Kruglyak, E Mardis, CN Rotimi, M Slatkin, D Valle, AS Whittemore, M Boehnke, AG Clark, EE Eichler, G Gibson, JL Haines, TF Mackay, SA McCarroll, and PM Visscher. **Finding the missing heritability of complex diseases.** *Nature*, **461**(7265):747–53, 2009.

[14] P. M. Visscher, W. G. Hill, and N. R. Wray. **Heritability in the genomics era–concepts and misconceptions**. *Nat Rev Genet*, **9**(4):255–66, 2008.

[15] O. Zuk, E. Hechter, S. R. Sunyaev, and E. S. Lander. **The mystery of missing heritability: Genetic interactions create phantom heritability**. *Proc Natl Acad Sci U S A*, **109**(4):1193–8, 2012.

[16] E. Kostem and E. Eskin. **Improving the accuracy and efficiency of partitioning heritability into the contributions of genomic regions**. *Am J Hum Genet*, **92**(4):558–64, 2013.

[17] S. S. Huang, J. A. Finkelstein, S. L. Rifas-Shiman, K. Kleinman, and R. Platt. **Community-level predictors of pneumococcal carriage and resistance in young children**. *Am J Epidemiol*, **159**(7):645–54, 2004.

[18] A. Roche, P. T. Heath, M. Sharland, D. Strachan, A. Breathnach, J. Haigh, and Y. Young. **Prevalence of nasopharyngeal carriage of pneumococcus in preschool children attending day care in London**. *Arch Dis Child*, **92**(12):1073–6, 2007.

[19] PL Tan, RM Jacobson, GA Poland, SJ Jacobsen, and VS Pankratz. **Twin studies of immunogenicity–determining the genetic contribution to vaccine failure.** *Vaccine*, **19**(17-19):2434–9, 2001.

[20] Stanley A. Plotkin, Walter A. Orenstein, and Paul A. Offit. *Vaccines*. Saunders/Elsevier, Philadelphia, Pa., 5th edition, 2008.

[21] S. A. Plotkin. **Vaccines: correlates of vaccine-induced immunity**. *Clin Infect Dis*, **47**(3):401–9, 2008.

[22] E. Png, A. Thalamuthu, R. T. Ong, H. Snippe, G. J. Boland, and M. Seielstad. **A genome-wide association study of hepatitis B vaccine response in an Indonesian population reveals multiple independent risk variants in the HLA region**. *Hum Mol Genet*, **20**(19):3893–8, 2011.

[23] Y. Kamatani, S. Wattanapokayakit, H. Ochi, T. Kawaguchi, A. Takahashi, N. Hosono, M. Kubo, T. Tsunoda, N. Kamatani, H. Kumada, A. Puseenam, T. Sura, Y. Daigo, K. Chayama, W. Chantratita, Y. Nakamura, and K. Matsuda. **A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians.** *Nat Genet*, **41**(5):591–5, 2009.

[24] Z. Hu, Y. Liu, X. Zhai, J. Dai, G. Jin, L. Wang, L. Zhu, Y. Yang, J. Liu, M. Chu, J. Wen, K. Xie, G. Du, Q. Wang, Y. Zhou, M. Cao, L. Liu, Y. He, Y. Wang, G. Zhou, W. Jia, J. Lu, S. Li, H. Yang, Y. Shi, W. Zhou, and H. Shen. **New loci associated with chronic hepatitis B virus infection in Han Chinese**. *Nat Genet*, **45**(12):1499–503, 2013.

[25] L. Pan, L. Zhang, W. Zhang, X. Wu, Y. Li, B. Yan, X. Zhu, X. Liu, C. Yang, J. Xu, G. Zhou, A. Xu, H. Li, and Y. Liu. **A genome-wide association study identifies polymorphisms in the HLA-DR region associated with non-response to hepatitis B vaccination in Chinese Han populations**. *Hum Mol Genet*, **23**(8):2210–9, 2014.

[26] I. G. Ovsyannikova, R. B. Kennedy, M. O'Byrne, R. M. Jacobson, V. S. Pankratz, and G. A. Poland. **Genome-wide association study of antibody response to smallpox vaccine**. *Vaccine*, **30**(28):4182–9, 2012.

[27] R. B. Kennedy, I. G. Ovsyannikova, V. S. Pankratz, I. H. Haralambieva, R. A. Vierkant, R. M. Jacobson, and G. A. Poland. **Genome-wide genetic associations with IFN-gamma response to smallpox vaccine**. *Hum Genet*, **131**(9):1433–51, 2012.

[28] R. B. Kennedy, I. G. Ovsyannikova, I. H. Haralambieva, N. D. Lambert, V. S. Pankratz, and G. A. Poland. **Genome-wide SNP associations with rubella-specific cytokine responses in measles-mumps-rubella vaccine recipients**. *Immunogenetics*, **66**(7-8):493–9, 2014.

[29] F. Dudbridge and A. Gusnanto. **Estimation of significance thresholds for genomewide association scans**. *Genet Epidemiol*, **32**(3):227–34, 2008.

[30] N. M. Pajewski, S. D. Parker, G. A. Poland, I. G. Ovsyannikova, W. Song, K. Zhang, B. A. McKinney, V. S. Pankratz, J. C. Edberg, R. P. Kimberly, R. M. Jacobson, J. Tang, and R. A. Kaslow. **The role of HLA-DR-DQ haplotypes in variable antibody responses to anthrax vaccine adsorbed**. *Genes Immun*, **12**(6):457–65, 2011.

[31] I. G. Ovsyannikova, R. M. Jacobson, R. A. Vierkant, M. M. O'Byrne, and G. A. Poland. **Replication of rubella vaccine population genetic studies: validation of HLA genotype and humoral response associations**. *Vaccine*, **27**(49):6926–31, 2009.

[32] I. G. Ovsyannikova, V. S. Pankratz, R. A. Vierkant, R. M. Jacobson, and G. A. Poland. **Consistency of HLA associations between two independent measles vaccine cohorts: a replication study**. *Vaccine*, **30**(12):2146–52, 2012.

[33] S. Davila, F. E. Froeling, A. Tan, C. Bonnard, G. J. Boland, H. Snippe, M. L. Hibberd, and M. Seielstad. **New genetic associations detected in a host response study to hepatitis B vaccine**. *Genes Immun*, **11**(3):232–8, 2010.

[34] CM Gelder, R Lambkin, KW Hart, D Fleming, OM Williams, M Bunce, KI Welsh, SE Marshall, and J Oxford. **Associations between human leukocyte antigens and nonresponsiveness to influenza vaccine.** *J Infect Dis*, **185**(1):114–7, 2002.

[35] B. Posteraro, R. Pastorino, P. Di Giannantonio, C. Ianuale, R. Amore, W. Ricciardi, and S. Boccia. **The link between genetic variation and variability in vaccine responses: systematic review and meta-analyses**. *Vaccine*, **32**(15):1661–9, 2014.

[36] Z. K. Li, J. J. Nie, J. Li, and H. Zhuang. **The effect of HLA on immunological response to hepatitis B vaccine in healthy people: a meta-analysis**. *Vaccine*, **31**(40):4355–61, 2013.

[37] N Dhiman, IG Ovsyannikova, RA Vierkant, JE Ryan, VS Pankratz, RM Jacobson, and GA Poland. **Associations between SNPs in toll-like receptors and related intracellular signaling molecules and immune responses to measles vaccine: preliminary results.** *Vaccine*, **26**(14):1731–6, 2008.

[38] C. E. Moore, B. J. Hennig, K. P. Perrett, J. C. Hoe, S. J. Lee, H. Fletcher, D. Brocklebank, D. O'Connor, M. D. Snape, A. J. Hall, S. Segal, A. V. Hill, and A. J. Pollard. **Single nucleotide polymorphisms in the Toll-like receptor 3 and CD44 genes are associated with persistence of vaccine-induced immunity to the serogroup C meningococcal conjugate vaccine**. *Clin Vaccine Immunol*, **19**(3):295–303, 2012.

[39] TG Kimman, S Banus, N Reijmerink, J Reimerink, FF Stelma, GH Koppelman, C Thijs, DS Postma, and M Kerkhof. **Association of interacting genes in the toll-like receptor signaling pathway and the antibody response to pertussis vaccination.** *PLoS ONE*, **3**(11):e3665, 2008.

[40] IG Ovsyannikova, N Dhiman, IH Haralambieva, RA Vierkant, MM O'Byrne, RM Jacobson, and GA Poland. **Rubella vaccine-induced cellular immunity: evidence of associations with polymorphisms in the Toll-like, vitamin A and D receptors, and innate immune response genes.** *Hum Genet*, **127**(2):207–21, 2010.

[41] N Dhiman, IG Ovsyannikova, JM Cunningham, RA Vierkant, RB Kennedy, VS Pankratz, GA Poland, and RM Jacobson. **Associations between measles vaccine immunity and single-nucleotide polymorphisms in cytokine and cytokine receptor genes.** *J Infect Dis*, **195**(1):21–9, 2007.

[42] IH Haralambieva, N Dhiman, IG Ovsyannikova, RA Vierkant, VS Pankratz, RM Jacobson, and GA Poland. **2'-5'-Oligoadenylate synthetase single-nucleotide polymorphisms and haplotypes are associated with variations in immune responses to rubella vaccine.** *Hum Immunol*, **71**(4):383–91, 2010.

[43] S. Banus, R. W. Bottema, C. L. Siezen, R. J. Vandebriel, J. Reimerink, M. Mommers, G. H. Koppelman, B. Hoebee, C. Thijs, D. S. Postma, T. G. Kimman, and F. F. Stelma. **Toll-like receptor 4 polymorphism associated with the response to whole-cell pertussis vaccination in children from the KOALA study**. *Clin Vaccine Immunol*, **14**(10):1377–80, 2007.

[44] K. Gröndahl-Yli-Hannuksela, J. Vuononvirta, A. M. Barkoff, M. Viander, O. Van Der Meeren, J. Mertsola, and Q. He. **Gene polymorphism in toll-like receptor 4: effect on antibody production and persistence after acellular pertussis vaccination during adolescence**. *J Infect Dis*, **205**(8):1214–9, 2012.

[45] P. P. Majumder, H. F. Staats, N. Sarkar-Roy, B. Varma, T. Ghosh, S. Maiti, K. Narayanasamy, C. C. Whisnant, J. L. Stephenson, and D. K. Wagener. **Genetic determinants of immune-response to a polysaccharide vaccine for typhoid**. *Hugo J*, **3**(1-4):17–30, 2009.

[46] H. D. Clifford, S. T. Yerkovich, S. K. Khoo, G. Zhang, J. Upham, P. N. Le Souëf, P. Richmond, and C. M. Hayden. **TLR3 and RIG-I gene variants: associations with functional effects on receptor expression and responses to measles virus and vaccine in vaccinated infants**. *Hum Immunol*, **73**(6):677–85, 2012.

[47] S. N. Ladhani, S. Davila, M. L. Hibberd, P. T. Heath, M. E. Ramsay, M. P. Slack, A. J. Pollard, and R. Booy. **Association between single-nucleotide polymorphisms in Mal/TIRAP and interleukin-10 genes and susceptibility to invasive *Haemophilus influenzae* serotype b infection in immunized children.** *Clin Infect Dis*, **51**(7):761–7, 2010.

[48] B Yucesoy, VJ Johnson, K Fluharty, ML Kashon, JE Slaven, NW Wilson, DN Weissman, RE Biagini, DR Germolec, and MI Luster. **Influence of cytokine gene variations on immunization to childhood vaccines.** *Vaccine*, **27**(50):6991–7, 2009.

[49] T Höhler, E Reuss, CM Freitag, and PM Schneider. **A functional polymorphism in the IL-10 promoter influences the response after vaccination with HBsAg and hepatitis A.** *Hepatology*, **42**(1):72–6, 2005.

[50] C Wang, J Tang, W Song, E Lobashevsky, CM Wilson, and RA Kaslow. **HLA and cytokine gene polymorphisms are independently associated with responses to hepatitis B vaccination.** *Hepatology*, **39**(4):978–88, 2004.

[51] N Dhiman, IG Ovsyannikova, RA Vierkant, VS Pankratz, RM Jacobson, and GA Poland. **Associations between cytokine/cytokine receptor single nucleotide polymorphisms and humoral immunity to measles, mumps and rubella in a Somali population.** *Tissue Antigens*, **72**(3):211–20, 2008.

[52] IG Ovsyannikova, JE Ryan, RM Jacobson, RA Vierkant, VS Pankratz, and GA Poland. **Human leukocyte antigen and interleukin 2, 10 and 12p40 cytokine responses to measles: is there evidence of the HLA effect?** *Cytokine*, **36**(3-4):173–9, 2006.

[53] L. Pan, W. Zhang, Z. Liang, X. Wu, X. Zhu, J. Li, T. Li, L. Wang, H. Li, and Y. Liu. **Association between polymorphisms of the cytokine and cytokine receptor genes and immune response to hepatitis B vaccination in a Chinese Han population**. *J Med Virol*, **84**(1):26–33, 2012.

[54] IG Ovsyannikova, RM Jacobson, N Dhiman, RA Vierkant, VS Pankratz, and GA Poland. **Human leukocyte antigen and cytokine receptor gene polymorphisms associated with heterogeneous immune responses to mumps viral vaccine.** *Pediatrics*, **121**(5):e1091–9, 2008.

[55] C Janeway. *Immunobiology : the immune system in health and disease*. Garland Science ; Churchill Livingstone, New York ; London Oxford, 6th edition, 2005.

[56] C. T. Watson and F. Breden. **The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease**. *Genes Immun*, 2012.

[57] K. Kolibab, S. L. Smithson, B. Rabquer, S. Khuder, and M. A. Westerink. **Immune response to pneumococcal polysaccharides 4 and 14 in elderly and young adults: analysis of the variable heavy chain repertoire**. *Infect Immun*, **73**(11):7465–76, 2005.

[58] A. H. Lucas and D. C. Reason. **Polysaccharide vaccines as probes of antibody repertoires in man**. *Immunol Rev*, **171**:89–104, 1999.

[59] A. H. Lucas, G. R. McLean, D. C. Reason, A. P. O'Connor, M. C. Felton, and K. D. Moulton. **Molecular ontogeny of the human antibody repertoire to the *Haemophilus influenzae* type B polysaccharide: expression of canonical variable regions and their variants in vaccinated infants**. *Clin Immunol*, **108**(2):119–27, 2003.

[60] A. Hamrouni, A. Aublin, P. Guillaume, and J. L. Maryanski. **T cell receptor gene rearrangement lineage analysis reveals clues for the origin of highly restricted antigen-specific repertoires**. *J Exp Med*, **197**(5):601–14, 2003.

[61] S. Gras, Z. Chen, J. J. Miles, Y. C. Liu, M. J. Bell, L. C. Sullivan, L. Kjer-Nielsen, R. M. Brennan, J. M. Burrows, M. A. Neller, R. Khanna, A. W. Purcell, A. G. Brooks, J. McCluskey, J. Rossjohn, and S. R. Burrows. **Allelic polymorphism in the T cell receptor and its impact on immune responses**. *J Exp Med*, **207**(7):1555–67, 2010.

[62] T. Romo-González and E. Vargas-Madrazo. **Structural analysis of substitution patterns in alleles of human immunoglobulin VH genes**. *Mol Immunol*, **42**(9):1085–97, 2005.

[63] L. Liu and A. H. Lucas. **IGH V3-23*01 and its allele V3-23*03 differ in their capacity to form the canonical human antibody combining site specific for the capsular polysaccharide of *Haemophilus influenzae* type b**. *Immunogenetics*, **55**(5):336–8, 2003.

[64] E. H. Sasso, J. H. Buckner, and L. A. Suzuki. **Ethnic differences in VH gene polymorphism**. *Ann N Y Acad Sci*, **764**:72–3, 1995.

[65] A. J. Feeney, M. J. Atkinson, M. J. Cowan, G. Escuro, and G. Lugo. **A defective Vkappa A2 allele in Navajos which may play a role in increased susceptibility to *Haemophilus influenzae* type b disease**. *J Clin Invest*, **97**(10):2277–82, 1996.

[66] B. Nadel, A. Tang, G. Lugo, V. Love, G. Escuro, and A. J. Feeney. **Decreased frequency of rearrangement due to the synergistic effect of nucleotide changes in the heptamer and nonamer of the recombination signal sequence of the V kappa gene A2b, which is associated with increased susceptibility of Navajos to *Haemophilus influenzae* type b disease**. *J Immunol*, **161**(11):6068–73, 1998.

[67] M. G. Scott, D. L. Crimmins, D. W. McCourt, I. Zocher, R. Thiebe, H. G. Zachau, and M. H. Nahm. **Clonal characterization of the human IgG antibody repertoire to *Haemophilus influenzae* type b polysaccharide. III. A single VKII gene and one of several JK genes are joined by an invariant arginine to form the most common L chain V region**. *J Immunol*, **143**(12):4110–6, 1989.

[68] S. D. Boyd, B. A. Gaëta, K. J. Jackson, A. Z. Fire, E. L. Marshall, J. D. Merker, J. M. Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D. Nguyen, K. C. Nadeau, M. Egholm, D. B. Miklos, J. L. Zehnder, and A. M. Collins. **Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements**. *J Immunol*, **184**(12):6986–92, 2010.

[69] E. H. Sasso, T. Johnson, and T. J. Kipps. **Expression of the immunoglobulin VH gene 51p1 is proportional to its germline gene copy number**. *J Clin Invest*, **97**(9):2074–80, 1996.

[70] C. T. Watson, K. M. Steinberg, J. Huddleston, R. L. Warren, M. Malig, J. Schein, A. J. Willsey, J. B. Joy, J. K. Scott, T. A. Graves, R. K. Wilson, R. A. Holt, E. E. Eichler, and F. Breden. **Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation**. *Am J Hum Genet*, **92**(4):530–46, 2013.

[71] T. R. O'Brien, I. Kohaar, R. M. Pfeiffer, D. Maeder, M. Yeager, E. E. Schadt, and L. Prokunina-Olsson. **Risk alleles for chronic hepatitis B are associated with decreased mRNA expression of HLA-DPA1 and HLA-DPB1 in normal human liver**. *Genes Immun*, **12**(6):428–33, 2011.

[72] E. Valkov, A. Stamp, F. Dimaio, D. Baker, B. Verstak, P. Roversi, S. Kellie, M. J. Sweet, A. Mansell, N. J. Gay, J. L. Martin, and B. Kobe. **Crystal structure of Toll-like receptor adaptor MAL/TIRAP reveals the molecular basis for signal transduction and disease protection**. *Proc Natl Acad Sci U S A*, **108**(36):14879–84, 2011.

[73] A. D. Skol, L. J. Scott, G. R. Abecasis, and M. Boehnke. **Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies**. *Nat Genet*, **38**(2):209–13, 2006.

[74] L Serpell and J Green. **Parental decision-making in childhood vaccination.** *Vaccine*, **24**(19):4041–6, 2006.

[75] FL Black, W Hierholzer, JP Woodall, and F Pinhiero. **Intensified reactions to measles vaccine in unexposed populations of american Indians.** *J Infect Dis*, **124**(3):306–17, 1971.

[76] DM Reif, BA McKinney, AA Motsinger, SJ Chanock, KM Edwards, MT Rock, JH Moore, and JE Crowe. **Genetic basis for adverse events after smallpox vaccination.** *J Infect Dis*, **198**(1):16–22, 2008.

[77] S. L. Stanley, S. E. Frey, P. Taillon-Miller, J. Guo, R. D. Miller, D. C. Koboldt, M. Elashoff, R. Christensen, N. L. Saccone, and R. B. Belshe. **The immunogenetics of smallpox vaccination**. *J Infect Dis*, **196**(2):212–9, 2007.

[78] C. A. Dinarello. **Infection, fever, and exogenous and endogenous pyrogens: some concepts have changed**. *J Endotoxin Res*, **10**(4):201–22, 2004.

[79] Y. N. Martin, O. E. Salavaggione, B. W. Eckloff, E. D. Wieben, D. J. Schaid, and R. M. Weinshilboum. **Human methylenetetrahydrofolate reductase pharmacogenomics: gene resequencing and functional genomics.** *Pharmacogenet Genomics*, **16**(4):265–77, 2006.

[80] W. Urano, A. Taniguchi, H. Yamanaka, E. Tanaka, H. Nakajima, Y. Matsuda, H. Akama, Y. Kitamura, and N. Kamatani. **Polymorphisms in the methylenetetrahydrofolate reductase gene were associated with both the efficacy and the toxicity of methotrexate used for the treatment of rheumatoid arthritis, as evidenced by single locus and haplotype analyses**. *Pharmacogenetics*, **12**(3):183–90, 2002.

[81] C. G. Casey, J. K. Iskander, M. H. Roper, E. E. Mast, X. J. Wen, T. J. Török, L. E. Chapman, D. L. Swerdlow, J. Morgan, J. D. Heffelfinger, C. Vitek, S. E. Reef, L. M. Hasbrouck, I. Damon, L. Neff, C. Vellozzi, M. McCauley, R. A. Strikas, and G. Mootrey. **Adverse events associated**

with smallpox vaccination in the United States, January-October 2003. *JAMA*, **294**(21):2734–43, 2005.

[82] David Salisbury, Mary Ramsay, and Karen Noakes. *Immunisation Against Infectious Disease (The Green Book)*. Department of Health, 2013.

[83] N. Gossger, M. D. Snape, L. M. Yu, A. Finn, G. Bona, S. Esposito, N. Principi, J. Diez-Domingo, E. Sokal, B. Becker, D. Kieninger, R. Prymula, P. Dull, E. Ypma, D. Toneatto, A. Kimura, A. J. Pollard, and European MenB Vaccine Study Group. **Immunogenicity and tolerability of recombinant serogroup B meningococcal vaccine administered with or without routine infant vaccinations according to different immunization schedules: a randomized controlled trial**. *JAMA*, **307**(6):573–82, 2012.

[84] TD Querec, RS Akondy, EK Lee, W Cao, HI Nakaya, D Teuwen, A Pirani, K Gernert, J Deng, B Marzolf, K Kennedy, H Wu, S Bennouna, H Oluoch, J Miller, RZ Vencio, M Mulligan, A Aderem, R Ahmed, and B Pulendran. **Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans.** *Nat Immunol*, **10**(1):116–25, 2009.

[85] H. I. Nakaya, J. Wrammert, E. K. Lee, L. Racioppi, S. Marie-Kunze, W. N. Haining, A. R. Means, S. P. Kasturi, N. Khan, G. M. Li, M. McCausland, V. Kanchan, K. E. Kokko, S. Li, R. Elbein, A. K. Mehta, A. Aderem, K. Subbarao, R. Ahmed, and B. Pulendran. **Systems biology of vaccination for seasonal influenza in humans.** *Nat Immunol*, **12**(8):786–95, 2011.

[86] S. Li, N. Rouphael, S. Duraisingham, S. Romero-Steiner, S. Presnell, C. Davis, D. S. Schmidt, S. E. Johnson, A. Milton, G. Rajam, S. Kasturi, G. M. Carlone, C. Quinn, D. Chaussabel, A. K. Palucka, M. J. Mulligan, R. Ahmed, D. S. Stephens, H. I. Nakaya, and B. Pulendran. **Molecular signatures of antibody responses derived from a systems biology study of five human vaccines**. *Nat Immunol*, **15**(2):195–204, 2014.

[87] G. Obermoser, S. Presnell, K. Domico, H. Xu, Y. Wang, E. Anguiano, L. Thompson-Snipes, R. Ranganathan, B. Zeitner, A. Bjork, D. Anderson, C. Speake, E. Ruchaud, J. Skinner, L. Alsina, M. Sharma, H. Dutartre, A. Cepika, E. Israelsson, P. Nguyen, Q. A. Nguyen, A. C. Harrod, S. M. Zurawski, V. Pascual, H. Ueno, G. T. Nepom, C. Quinn, D. Blankenship, K. Palucka, J. Banchereau, and D. Chaussabel. **Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines**. *Immunity*, **38**(4):831–44, 2013.

[88] K. L. Bucasas, L. M. Franco, C. A. Shaw, M. S. Bray, J. M. Wells, D. NiŰo, N. Arden, J. M. Quarles, R. B. Couch, and J. W. Belmont. **Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans**. *J Infect Dis*, **203**(7):921–9, 2011.

[89] C. A. Scherer, C. L. Magness, K. V. Steiger, N. D. Poitinger, C. M. Caputo, D. G. Miner, P. L. Winokur, D. Klinzman, J. McKee, C. Pilar, P. A. Ward,

M. H. Gillham, N. J. Haulman, J. T. Stapleton, and S. P. Iadonato. **Distinct gene expression profiles in peripheral blood mononuclear cells from patients infected with vaccinia virus, yellow fever 17D virus, or upper respiratory infections**. *Vaccine*, **25**(35):6458–73, 2007.

[90] M. T. Vahey, Z. Wang, K. E. Kester, J. Cummings, D. G. Heppner, M. E. Nau, O. Ofori-Anyinam, J. Cohen, T. Coche, W. R. Ballou, and C. F. Ockenhouse. **Expression of genes associated with immunoproteasome processing of major histocompatibility complex peptides is indicative of protection with adjuvanted RTS,S malaria vaccine.** *J Infect Dis*, **201**(4):580–9, 2010.

[91] C. L. Fuller, K. C. Brittingham, M. W. Porter, M. J. Hepburn, P. L. Petitt, P. R. Pittman, and S. Bavari. **Transcriptome analysis of human immune responses following live vaccine strain (LVS) Francisella tularensis vaccination**. *Mol Immunol*, **44**(12):3173–84, 2007.

[92] F. Mosca, E. Tritto, A. Muzzi, E. Monaci, F. Bagnoli, C. Iavarone, D. O'Hagan, R. Rappuoli, and E. De Gregorio. **Molecular and cellular signatures of human vaccine adjuvants**. *Proc Natl Acad Sci U S A*, **105**(30):10501–6, 2008.

[93] E. Caproni, E. Tritto, M. Cortese, A. Muzzi, F. Mosca, E. Monaci, B. Baudner, A. Seubert, and E. De Gregorio. **MF59 and Pam3CSK4 boost adaptive responses to influenza subunit vaccine through an IFN type I-independent mechanism of action**. *J Immunol*, **188**(7):3088–98, 2012.

[94] K. S. Korsholm, R. V. Petersen, E. M. Agger, and P. Andersen. **T-helper 1 and T-helper 2 adjuvants induce distinct differences in the magnitude, quality and kinetics of the early inflammatory response at the site of injection**. *Immunology*, **129**(1):75–86, 2010.

[95] D. Gaucher, R. Therrien, N. Kettaf, B. R. Angermann, G. Boucher, A. Filali-Mouhim, J. M. Moser, R. S. Mehta, D. R. Drake, E. Castro, R. Akondy, A. Rinfret, B. Yassine-Diab, E. A. Said, Y. Chouikh, M. J. Cameron, R. Clum, D. Kelvin, R. Somogyi, L. D. Greller, R. S. Balderas, P. Wilkinson, G. Pantaleo, J. Tartaglia, E. K. Haddad, and R. P. Sékaly. **Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses**. *J Exp Med*, **205**(13):3119–31, 2008.

[96] R. B. Belshe, K. M. Edwards, T. Vesikari, S. V. Black, R. E. Walker, M. Hultquist, G. Kemble, E. M. Connor, and CAIV-T Comparative Efficacy Study Group. **Live attenuated versus inactivated influenza vaccine in infants and young children**. *N Engl J Med*, **356**(7):685–96, 2007.

[97] W. Zhu, B. W. Higgs, C. Morehouse, K. Streicher, C. S. Ambrose, J. Woo, G. W. Kemble, B. Jallal, and Y. Yao. **A whole genome transcriptional analysis of the early immune response induced by live attenuated and inactivated influenza vaccines in young children.** *Vaccine*, **28**(16):2865–76, 2010.

[98] R. G. Cao, N. M. Suarez, G. Obermoser, S. M. Lopez, E. Flano, S. E. Mertz, R. A. Albrecht, A. Garcia-Sastre, A. Mejias, H. Xu, H. Qin, D. Blankenship, K. Palucka, V. Pascual, and O. Ramilo. **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**. *J Infect Dis*, **210**(2):224–33, 2014.

[99] A. D. Henn, S. Wu, X. Qiu, M. Ruda, M. Stover, H. Yang, Z. Liu, S. L. Welle, J. Holden-Wiltse, H. Wu, and M. S. Zand. **High-resolution temporal response patterns to influenza vaccine reveal a distinct human plasma cell gene signature**. *Sci Rep*, **3**:2327, 2013.

[100] D. Furman, V. Jojic, B. Kidd, S. Shen-Orr, J. Price, J. Jarrell, T. Tse, H. Huang, P. Lund, H. T. Maecker, P. J. Utz, C. L. Dekker, D. Koller, and M. M. Davis. **Apoptosis and other immune biomarkers predict influenza vaccine responsiveness**. *Mol Syst Biol*, **9**:659, 2013.

[101] D. E. Zak, E. Andersen-Nissen, E. R. Peterson, A. Sato, M. K. Hamilton, J. Borgerding, A. T. Krishnamurty, J. T. Chang, D. J. Adams, T. R. Hensley, A. I. Salter, C. A. Morgan, A. C. Duerr, S. C. De Rosa, A. Aderem, and M. J. McElrath. **Merck Ad5/HIV induces broad innate immune activation that predicts CD8? T-cell responses but is attenuated by preexisting Ad5 immunity**. *Proc Natl Acad Sci U S A*, **109**(50):E3503–12, 2012.

[102] M. J. McElrath, S. C. De Rosa, Z. Moodie, S. Dubey, L. Kierstead, H. Janes, O. D. Defawe, D. K. Carter, J. Hural, R. Akondy, S. P. Buchbinder, M. N. Robertson, D. V. Mehrotra, S. G. Self, L. Corey, J. W. Shiver, D. R. Casimiro, and Step Study Protocol Team. **HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis**. *Lancet*, **372**(9653):1894–905, 2008.

[103] S. P. Buchbinder, D. V. Mehrotra, A. Duerr, D. W. Fitzgerald, R. Mogg, D. Li, P. B. Gilbert, J. R. Lama, M. Marmor, C. Del Rio, M. J. McElrath, D. R. Casimiro, K. M. Gottesdiener, J. A. Chodakewitz, L. Corey, M. N. Robertson, and Step Study Protocol Team. **Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study) a double-blind, randomised, placebo-controlled, test-of-concept trial**. *Lancet*, **372**(9653):1881–93, 2008.

[104] I. H. Haralambieva, A. L. Oberg, N. Dhiman, I. G. Ovsyannikova, R. B. Kennedy, D. E. Grill, R. M. Jacobson, and G. A. Poland. **High-dimensional gene expression profiling studies in high and low responders to primary smallpox vaccination**. *J Infect Dis*, **206**(10):1512–20, 2012.

[105] H. A. Fletcher, A. Keyser, M. Bowmaker, P. C. Sayles, G. Kaplan, G. Hussey, A. V. Hill, and W. A. Hanekom. **Transcriptional profiling of mycobacterial antigen-induced responses in infants vaccinated with BCG at birth**. *BMC Med Genomics*, **2**:10, 2009.

[106] I. H. Haralambieva, A. L. Oberg, I. G. Ovsyannikova, R. B. Kennedy, D. E. Grill, S. Middha, B. M. Bot, V. W. Wang, D. I. Smith, R. M. Jacobson, and G. A. Poland. **Genome-wide characterization of transcriptional patterns in high and low antibody responders to rubella vaccination**. *PLoS One*, **8**(5):e62149, 2013.

[107] D. Li, J. P. Lewinger, W. J. Gauderman, C. E. Murcray, and D. Conti. **Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies**. *Genet Epidemiol*, **35**(8):790–9, 2011.

[108] G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, G. A. McVean, and 1000 Genomes Project Consortium. **An integrated map of genetic variation from 1,092 human genomes**. *Nature*, **491**(7422):56–65, 2012.

[109] A. Kong, V. Steinthorsdottir, G. Masson, G. Thorleifsson, P. Sulem, S. Besenbacher, A. Jonasdottir, A. Sigurdsson, K. T. Kristinsson, M. L. Frigge, A. Gylfason, P. I. Olason, S. A. Gudjonsson, S. Sverrisson, S. N. Stacey, B. Sigurgeirsson, K. R. Benediktsdottir, H. Sigurdsson, T. Jonsson, R. Benediktsson, J. H. Olafsson, O. T. Johannsson, A. B. Hreidarsson, G. Sigurdsson, A. C. Ferguson-Smith, D. F. Gudbjartsson, U. Thorsteinsdottir, K. Stefansson, and DIAGRAM Consortium. **Parental origin of sequence variants associated with complex diseases**. *Nature*, **462**(7275):868–74, 2009.

[110] S. Debey, U. Schoenbeck, M. Hellmich, B. S. Gathof, R. Pillai, T. Zander, and J. L. Schultze. **Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types**. *Pharmacogenomics J*, **4**(3):193–207, 2004.

[111] I. Dinu, J. D. Potter, T. Mueller, Q. Liu, A. J. Adewale, G. S. Jhangri, G. Einecke, K. S. Famulski, P. Halloran, and Y. Yasui. **Gene-set analysis and reduction**. *Brief Bioinform*, **10**(1):24–34, 2009.

[112] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods*, **5**(7):621–8, 2008.

[113] P. J. Hurd and C. J. Nelson. **Advantages of next-generation sequencing versus the microarray in epigenetic research**. *Brief Funct Genomic Proteomic*, **8**(3):174–83, 2009.

[114] A. Sîrbu, G. Kerr, M. Crane, and H. J. Ruskin. **RNA-Seq vs Dual- and Single-Channel Microarray Data: Sensitivity Analysis for Differential Expression and Clustering**. *PLoS One*, **7**(12):e50986, 2012.

[115] Z Wang, M Gerstein, and M Snyder. **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet*, **10**(1):57–63, 2009.

[116] F. Ozsolak and P. M. Milos. **RNA sequencing: advances, challenges and opportunities.** *Nat Rev Genet*, **12**(2):87–98, 2011.

[117] PI de Bakker, R Yelensky, I Pe'er, SB Gabriel, MJ Daly, and D Altshuler. **Efficiency and power in genetic association studies.** *Nat Genet*, **37**(11):1217–23, 2005.

[118] J. P. Ioannidis, E. E. Ntzani, T. A. Trikalinos, and D. G. Contopoulos-Ioannidis. **Replication validity of genetic association studies**. *Nat Genet*, **29**(3):306–9, 2001.

[119] D. O'Connor, C. E. Moore, M. D. Snape, T. John, A. V. Hill, and A. J. Pollard. **Exonic single nucleotide polymorphisms within TLR3 associated with infant responses to serogroup C meningococcal conjugate vaccine**. *Vaccine*, **32**(27):3424–30, 2014.

[120] WHO. **Control of epidemic meningococcal disease**, 1998.

[121] C. Trotter, S. Samuelsson, A. Perrocheau, S. de Greeff, H. de Melker, S. Heuberger, and M. Ramsay. **Ascertainment of meningococcal disease in Europe**. *Euro Surveill*, **10**(12):247–50, 2005.

[122] K. A. Cartwright, J. M. Stuart, D. M. Jones, and N. D. Noah. **The Stonehouse survey: nasopharyngeal carriage of meningococci and Neisseria lactamica**. *Epidemiol Infect*, **99**(3):591–601, 1987.

[123] R. Gold, I. Goldschneider, M. L. Lepow, T. F. Draper, and M. Randolph. **Carriage of Neisseria meningitidis and Neisseria lactamica in infants and children**. *J Infect Dis*, **137**(2):112–21, 1978.

[124] S. J. Gray, C. L. Trotter, M. E. Ramsay, M. Guiver, A. J. Fox, R. Borrow, R. H. Mallard, E. B. Kaczmarski, and Meningococcal Reference Unit. **Epidemiology of meningococcal disease in England and Wales 1993/94 to 2003/04: contribution and experiences of the Meningococcal Reference Unit**. *J Med Microbiol*, **55**(Pt 7):887–96, 2006.

[125] NE Rosenstein, BA Perkins, DS Stephens, T Popovic, and JM Hughes. **Meningococcal disease.** *N Engl J Med*, **344**(18):1378–88, 2001.

[126] I Goldschneider, EC Gotschlich, and MS Artenstein. **Human immunity to the meningococcus. I. The role of humoral antibodies.** *J Exp Med*, **129**(6):1307–26, 1969.

[127] R Borrow, P Balmer, and E Miller. **Meningococcal surrogates of protection–serum bactericidal antibody activity.** *Vaccine*, **23**(17-18):2222–7, 2005.

[128] D. Riding and N. L. Corkill. **Prophylactic Vaccination in Epidemic Meningococcal Meningitis**. *J Hyg (Lond)*, **32**(2):258–67, 1932.

[129] EC Gotschlich, I Goldschneider, and MS Artenstein. **Human immunity to the meningococcus. IV. Immunogenicity of group A and group C meningococcal polysaccharides in human volunteers.** *J Exp Med*, **129**(6):1367–84, 1969.

[130] R. Gold and M. S. Artenstein. **Meningococcal infections. 2. Field trial of group C meningococcal polysaccharide vaccine in 1969-70**. *Bull World Health Organ*, **45**(3):279–82, 1971.

[131] M. S. Artenstein, R. Gold, J. G. Zimmerly, F. A. Wyle, H. Schneider, and C. Harkins. **Prevention of meningococcal disease by group C polysaccharide vaccine**. *N Engl J Med*, **282**(8):417–20, 1970.

[132] M. L. Lepow, I. Goldschneider, R. Gold, M. Randolph, and E. C. Gotschlich. **Persistence of antibody following immunization of children with groups A and C meningococcal polysaccharide vaccines**. *Pediatrics*, **60**(5):673–80, 1977.

[133] R. Gold, M. L. Lepow, I. Goldschneider, T. L. Draper, and E. C. Gotschlich. **Clinical evaluation of group A and group C meningococcal polysaccharide vaccines in infants**. *J Clin Invest*, **56**(6):1536–47, 1975.

[134] A. Leach, P. A. Twumasi, S. Kumah, W. S. Banya, S. Jaffar, B. D. Forrest, D. M. Granoff, D. E. LiButti, G. M. Carlone, L. B. Pais, C. V. Broome, and B. M. Greenwood. **Induction of immunologic memory in Gambian children by vaccination in infancy with a group A plus group C meningococcal polysaccharide-protein conjugate vaccine**. *J Infect Dis*, **175**(1):200–4, 1997.

[135] Augusto E. de Taunay, Paulo A. Galvao, Jose S. de Morais, Emil C Gotschlich, and Roger A. Feldman. **Disease prevention by meningococcal serogroup C polysaccharide vaccine inpreschool children: results after eleven months in Sao Paulo, Brazil**. *Pediatr Res*, **8**(4):429–429, 1974.

[136] P. Richmond, R. Borrow, E. Miller, S. Clark, F. Sadler, A. Fox, N. Begg, R. Morris, and K. Cartwright. **Meningococcal serogroup C conjugate vaccine is immunogenic in infancy and primes for memory**. *J Infect Dis*, **179**(6):1569–72, 1999.

[137] E. Miller, D. Salisbury, and M. Ramsay. **Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story**. *Vaccine*, **20 Suppl 1**:S58–67, 2001.

[138] L. Salleras, A. Dominguez, and N. Cardenosa. **Dramatic decline of serogroup C meningococcal disease in Catalonia (Spain) after a mass vaccination campaign with meningococcal C conjugated vaccine**. *Vaccine*, **21**(7-8):729–33, 2003.

[139] P. De Wals, G. Deceuninck, N. Boulianne, and G. De Serres. **Effectiveness of a mass immunization campaign using serogroup C meningococcal conjugate vaccine**. *JAMA*, **292**(20):2491–4, 2004.

[140] J. M. MacLennan, F. Shackley, P. T. Heath, J. J. Deeks, C. Flamank, M. Herbert, H. Griffiths, E. Hatzmann, C. Goilav, and E. R. Moxon. **Safety, immunogenicity, and induction of immunologic memory by a serogroup C meningococcal conjugate vaccine in infants: A randomized controlled trial**. *JAMA*, **283**(21):2795–801, 2000.

[141] MD Snape, KP Perrett, KJ Ford, TM John, D Pace, LM Yu, JM Langley, S McNeil, PM Dull, F Ceddia, A Anemona, SA Halperin, S Dobson, and AJ Pollard. **Immunogenicity of a tetravalent meningococcal glycoconjugate vaccine in infants: a randomized controlled trial.** *JAMA*, **299**(2):173–84, 2008.

[142] F. A. Wyle, M. S. Artenstein, B. L. Brandt, E. C. Tramont, D. L. Kasper, P. L. Altieri, S. L. Berman, and J. P. Lowenthal. **Immunologic response of man to group B meningococcal polysaccharide vaccines**. *J Infect Dis*, **126**(5):514–21, 1972.

[143] G. Blanchard-Rohner, M. D. Snape, D. F. Kelly, D. O'Connor, T. John, E. Kibwana, H. Parks, K. Ford, P. M. Dull, and A. J. Pollard. **Seroprevalence and placental transmission of maternal antibodies specific for Neisseria meningitidis Serogroups A, C, Y and W135 and influence of maternal antibodies on the immune response to a primary course of MenACWY-CRM vaccine in the United Kingdom**. *Pediatr Infect Dis J*, **32**(7):768–76, 2013.

[144] KP Perrett, MD Snape, KJ Ford, TM John, LM Yu, JM Langley, S McNeil, PM Dull, F Ceddia, A Anemona, SA Halperin, S Dobson, and AJ Pollard. **Immunogenicity and immune memory of a nonadjuvanted quadrivalent meningococcal glycoconjugate vaccine in infants.** *Pediatr Infect Dis J*, **28**(3):186–93, 2009.

[145] D. Pace, M. Snape, S. Westcar, M. Hamaluba, L. M. Yu, N. Begg, J. Wysocki, H. Czajka, G. Maechler, D. Boutriau, and A. J. Pollard. **A new combination *Haemophilus influenzae* type B and Neisseria meningitidis serogroup C-tetanus toxoid conjugate vaccine for primary immunization of infants**. *Pediatr Infect Dis J*, **26**(11):1057–9, 2007.

[146] LL Gheesling, GM Carlone, LB Pais, PF Holder, SE Maslanka, BD Plikaytis, M Achtman, P Densen, CE Frasch, and H KŁyhty. **Multicenter comparison of Neisseria meningitidis serogroup C anti-capsular polysaccharide antibody levels measured by a standardized enzyme-linked immunosorbent assay.** *J Clin Microbiol*, **32**(6):1475–82, 1994.

[147] CM Elie, PK Holder, S Romero-Steiner, and GM Carlone. **Assignment of additional anticapsular antibody concentrations to the Neisseria meningitidis group A, C, Y, and W-135 meningococcal standard reference serum CDC1992.** *Clin Diagn Lab Immunol*, **9**(3):725–6, 2002.

[148] S. Little. **Amplification-refractory mutation system (ARMS) analysis of point mutations**. *Curr Protoc Hum Genet*, **Chapter 9**:Unit 9.8, 2001.

[149] S. Gabriel, L. Ziaugra, and D. Tabbaa. **SNP genotyping using the Sequenom MassARRAY iPLEX platform**. *Curr Protoc Hum Genet*, **Chapter 2**:Unit 2.12, 2009.

[150] F. Sanger, S. Nicklen, and A. R. Coulson. **DNA sequencing with chain-terminating inhibitors**. *Proc Natl Acad Sci U S A*, **74**(12):5463–7, 1977.

[151] J. K. Bonfield, C. Rada, and R. Staden. **Automated detection of point mutations using fluorescent sequence trace subtraction.** *Nucleic Acids Res*, **26**(14):3404–9, 1998.

[152] Gregory Warnes, with contributions from Gregor Gorjanc, Friedrich Leisch, and Michael Man. *genetics: Population Genetics*, 2013. R package version 1.3.8.1.

[153] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[154] B. Servin and M. Stephens. **Imputation-based analysis of association studies: candidate regions and quantitative traits**. *PLoS Genet*, **3**(7):e114, 2007.

[155] P. Scheet and M. Stephens. **A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase**. *Am J Hum Genet*, **78**(4):629–44, 2006.

[156] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. de Bakker, Mark J. Daly, and Pak C. Sham. **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *American journal of human genetics*, **81**(3):559–575, September 2007.

[157] P. Richmond, R. Borrow, D. Goldblatt, J. Findlow, S. Martin, R. Morris, K. Cartwright, and E. Miller. **Ability of 3 different meningococcal C conjugate vaccines to induce immunologic memory after a single dose in UK toddlers.** *J Infect Dis*, **183**(1):160–3, 2001.

[158] R. Borrow, P. Richmond, E. B. Kaczmarski, A. Iverson, S. L. Martin, J. Findlow, M. Acuna, E. Longworth, R. O'Connor, J. Paul, and E. Miller. **Meningococcal serogroup C-specific IgG antibody responses and serum bactericidal titres in children following vaccination with a meningococcal A/C polysaccharide vaccine**. *FEMS Immunol Med Microbiol*, **28**(1):79–85, 2000.

[159] L Alexopoulou, AC Holt, R Medzhitov, and RA Flavell. **Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3.** *Nature*, **413**(6857):732–8, 2001.

[160] K. A. Cavassani, M. Ishii, H. Wen, M. A. Schaller, P. M. Lincoln, N. W. Lukacs, C. M. Hogaboam, and S. L. Kunkel. **TLR3 is an endogenous sensor of tissue necrosis during acute inflammatory events.** *J Exp Med*, **205**(11):2609–21, 2008.

[161] J. M. Weiss, J. Sleeman, A. C. Renkl, H. Dittmar, C. C. Termeer, S. Taxis, N. Howells, M. Hofmann, G. Köhler, E. Schöpf, H. Ponta, P. Herrlich, and J. C. Simon. **An essential role for CD44 variant isoforms in epidermal Langerhans cell and blood dendritic cell function**. *J Cell Biol*, **137**(5):1137–47, 1997.

[162] S. Goodison, V. Urquidi, and D. Tarin. **CD44 cell adhesion molecules**. *Mol Pathol*, **52**(4):189–96, 1999.

[163] K. K. Ryckman, K. Fielding, A. V. Hill, M. Mendy, P. Rayco-Solon, G. Sirugo, M. A. van der Sande, P. Waight, H. C. Whittle, A. J. Hall, S. M. Williams, and B. J. Hennig. **Host genetic factors and vaccine-induced immunity to HBV infection: haplotype analysis.** *PLoS One*, **5**(8):e12273, 2010.

[164] Y. Guo, M. Audry, M. Ciancanelli, L. Alsina, J. Azevedo, M. Herman, E. Anguiano, V. Sancho-Shimizu, L. Lorenzo, E. Pauwels, P. B. Philippe, R. Pérez de Diego, A. Cardon, G. Vogt, C. Picard, Z. Z. Andrianirina, F. Rozenberg, P. Lebon, S. Plancoulaine, M. Tardieu, Valérie Doireau, E. Jouanguy, D. Chaussabel, F. Geissmann, L. Abel, J. L. Casanova, and S. Y. Zhang. **Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity.** *J Exp Med*, **208**(10):2083–98, 2011.

[165] B. Kobe and A. V. Kajava. **The leucine-rich repeat as a protein recognition motif.** *Curr Opin Struct Biol*, **11**(6):725–32, 2001.

[166] M. Yamamoto, S. Sato, H. Hemmi, K. Hoshino, T. Kaisho, H. Sanjo, O. Takeuchi, M. Sugiyama, M. Okabe, K. Takeda, and S. Akira. **Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway**. *Science*, **301**(5633):640–3, 2003.

[167] M. Yoneyama, M. Kikuchi, T. Natsukawa, N. Shinobu, T. Imaizumi, M. Miyagishi, K. Taira, S. Akira, and T. Fujita. **The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses**. *Nat Immunol*, **5**(7):730–7, 2004.

[168] L. Gitlin, W. Barchet, S. Gilfillan, M. Cella, B. Beutler, R. A. Flavell, M. S. Diamond, and M. Colonna. **Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus**. *Proc Natl Acad Sci U S A*, **103**(22):8459–64, 2006.

[169] N. Court, V. Vasseur, R. Vacher, C. Frémond, Y. Shebzukhov, V. V. Yeremeev, I. Maillet, S. A. Nedospasov, S. Gordon, P. G. Fallon, H. Suzuki, B. Ryffel, and V. F. Quesniaux. **Partial redundancy of the pattern recognition receptors, scavenger receptors, and C-type lectins for the long-term control of Mycobacterium tuberculosis infection**. *J Immunol*, **184**(12):7057–70, 2010.

[170] P. R. Taylor, L. Martinez-Pomares, M. Stacey, H. H. Lin, G. D. Brown, and S. Gordon. **Macrophage receptors and immune recognition**. *Annu Rev Immunol*, **23**:901–44, 2005.

[171] M. G. Netea, C. Wijmenga, and L. A. O'Neill. **Genetic variation in Toll-like receptors and disease susceptibility**. *Nat Immunol*, **13**(6):535–42, 2012.

[172] SY Zhang, E Jouanguy, S Ugolini, A Smahi, G Elain, P Romero, D Segal, V Sancho-Shimizu, L Lorenzo, A Puel, C Picard, A Chapgier, S Plancoulaine, M Titeux, C Cognet, H von Bernuth, CL Ku, A Casrouge, XX Zhang, L Barreiro, J Leonard, C Hamilton, P Lebon, B Héron, L Vallée, L Quintana-Murci, A Hovnanian, F Rozenberg, E Vivier, F Geissmann, M Tardieu, L Abel, and JL Casanova. **TLR3 deficiency in patients with herpes simplex encephalitis.** *Science*, **317**(5844):1522–7, 2007.

[173] L Quintana-Murci, A AlcaŢs, L Abel, and JL Casanova. **Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases.** *Nat Immunol*, **8**(11):1165–71, 2007.

[174] C. L. Ku, H. von Bernuth, C. Picard, S. Y. Zhang, H. H. Chang, K. Yang, M. Chrabieh, A. C. Issekutz, C. K. Cunningham, J. Gallin, S. M. Holland, C. Roifman, S. Ehl, J. Smart, M. Tang, F. J. Barrat, O. Levy, D. McDonald, N. K. Day-Good, R. Miller, H. Takada, T. Hara, S. Al-Hajjar, A. Al-Ghonaium, D. Speert, D. Sanlaville, X. Li, F. Geissmann, E. Vivier, L. Maródi, B. Z. Garty, H. Chapel, C. Rodriguez-Gallego, X. Bossuyt, L. Abel, A. Puel, and J. L. Casanova. **Selective predisposition to bacterial infections in IRAK-4-deficient children: IRAK-4-dependent TLRs are otherwise redundant in protective immunity**. *J Exp Med*, **204**(10):2407–22, 2007.

[175] H. von Bernuth, C. Picard, Z. Jin, R. Pankla, H. Xiao, C. L. Ku, M. Chrabieh, I. B. Mustapha, P. Ghandil, Y. Camcioglu, J. Vasconcelos, N. Sirvent, M. Guedes, A. B. Vitor, M. J. Herrero-Mata, J. I. Aróstegui, C. Rodrigo, L. Alsina, E. Ruiz-Ortiz, M. Juan, C. Fortuny, J. Yagüe, J. Antón, M. Pascal, H. H. Chang, L. Janniere, Y. Rose, B. Z. Garty, H. Chapel, A. Issekutz, L. Maródi, C. Rodriguez-Gallego, J. Banchereau, L. Abel, X. Li, D. Chaussabel, A. Puel, and J. L. Casanova. **Pyogenic bacterial infections in humans with MyD88 deficiency**. *Science*, **321**(5889):691–6, 2008.

[176] M. Sironi, M. Biasin, R. Cagliani, D. Forni, M. De Luca, I. Saulle, S. Lo Caputo, F. Mazzotta, J. Macías, J. A. Pineda, A. Caruz, and M. Clerici. **A common polymorphism in TLR3 confers natural resistance to HIV-1 infection.** *J Immunol*, **188**(2):818–23, 2012.

[177] E. Kindberg, S. Vene, A. Mickiene, Ą Lundkvist, L. Lindquist, and L. Svensson. **A functional Toll-like receptor 3 gene (TLR3) may be a risk factor for tick-borne encephalitis virus (TBEV) infection.** *J Infect Dis*, **203**(4):523–8, 2011.

[178] D. N. O'Dwyer, M. E. Armstrong, G. Trujillo, G. Cooke, M. P. Keane, P. G. Fallon, A. J. Simpson, A. B. Millar, E. E. McGrath, M. K. Whyte,

N. Hirani, C. M. Hogaboam, and S. C. Donnelly. **The Toll-like receptor 3 L412F polymorphism and disease progression in idiopathic pulmonary fibrosis**. *Am J Respir Crit Care Med*, **188**(12):1442–50, 2013.

[179] CT Ranjith-Kumar, W Miller, J Sun, J Xiong, J Santos, I Yarbrough, RJ Lamb, J Mills, KE Duffy, S Hoose, M Cunningham, A Holzenburg, ML Mbow, RT Sarisky, and CC Kao. **Effects of single nucleotide polymorphisms on Toll-like receptor 3 activity and expression in cultured cells.** *J Biol Chem*, **282**(24):17696–705, 2007.

[180] C. Gorbea, K. A. Makar, M. Pauschinger, G. Pratt, J. L. Bersola, J. Varela, R. M. David, L. Banks, C. H. Huang, H. Li, H. P. Schultheiss, J. A. Towbin, J. G. Vallejo, and N. E. Bowles. **A role for Toll-like receptor 3 variants in host susceptibility to enteroviral myocarditis and dilated cardiomyopathy**. *J Biol Chem*, **285**(30):23208–23, 2010.

[181] P. Cho, L. Gelinas, N. P. Corbett, S. J. Tebbutt, S. E. Turvey, E. S. Fortuno, and T. R. Kollmann. **Association of common single-nucleotide polymorphisms in innate immune genes with differences in TLR-induced cytokine production in neonates**. *Genes Immun*, **14**(4):199–211, 2013.

[182] A. Svensson, P. Tunbäck, I. Nordström, L. Padyukov, and K. Eriksson. **Polymorphisms in Toll-like receptor 3 confer natural resistance to human herpes simplex virus type 2 infection**. *J Gen Virol*, **93**(Pt 8):1717–24, 2012.

[183] N. S. Crowcroft, R. Booy, T. Harrison, L. Spicer, J. Britto, Q. Mok, P. Heath, I. Murdoch, M. Zambon, R. George, and E. Miller. **Severe and unrecognised: pertussis in UK infants**. *Arch Dis Child*, **88**(9):802–6, 2003.

[184] Public Health Englandhind. **Health Protection Agency general information, http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/**, accessed January 2014.

[185] A. J. Pollard. **New combination vaccines still need a boost**. *Arch Dis Child*, **92**(1):1–2, 2007.

[186] N. Le Saux, N. J. Barrowman, D. L. Moore, S. Whiting, D. Scheifele, S. Halperin, and Canadian Paediatric Society/ Health Canada Immunization Monitoring Program-Active (IMPACT). **Decrease in hospital admissions for febrile seizures and reports of hypotonic-hyporesponsive episodes presenting to hospital emergency departments since switching to acellular pertussis vaccine in Canada: a report from IMPACT**. *Pediatrics*, **112**(5):e348, 2003.

[187] I. M. Onorato, S. G. Wassilak, and B. Meade. **Efficacy of whole-cell pertussis vaccine in preschool children in the United States**. *JAMA*, **267**(20):2745–9, 1992.

[188] P. E. Fine and J. A. Clarkson. **Reflections on the efficacy of pertussis vaccines**. *Rev Infect Dis*, **9**(5):866–83, 1987.

[189] M. E. Pichichero and J. R. Casey. **Acellular pertussis vaccines for adolescents**. *Pediatr Infect Dis J*, **24**(6 Suppl):S117–26, 2005.

[190] O. N. Tinnion and M. Hanlon. **Acellular vaccines for preventing whooping cough in children**. *Cochrane Database Syst Rev*, page CD001478, 2000.

[191] J. Storsaeter, H. O. Hallander, L. Gustafsson, and P. Olin. **Levels of anti-pertussis antibodies related to protection after household exposure to Bordetella pertussis**. *Vaccine*, **16**(20):1907–16, 1998.

[192] J. Taranger, B. Trollfors, T. Lagergård, V. Sundh, D. A. Bryla, R. Schneerson, and J. B. Robbins. **Correlation between pertussis toxin IgG antibodies in postvaccination sera and subsequent protection against pertussis**. *J Infect Dis*, **181**(3):1010–3, 2000.

[193] E. Miller. **Overview of recent clinical trials of acellular pertussis vaccines**. *Biologicals*, **27**(2):79–86, 1999.

[194] N. Kitchin, J. Southern, R. Morris, F. Hemme, K. Cartwright, M. Watson, and E. Miller. **A randomised controlled study of the reactogenicity of an acellular pertussis-containing pentavalent infant vaccine compared to a quadrivalent whole cell pertussis-containing vaccine and oral poliomyelitis vaccine, when given concurrently with meningococcal group C conjugate vaccine to healthy UK infants at 2, 3 and 4 months of age**. *Vaccine*, **24**(18):3964–70, 2006.

[195] P. Olin, F. Rasmussen, L. Gustafsson, H. O. Hallander, and H. Heijbel. **Randomised controlled trial of two-component, three-component, and five-component acellular pertussis vaccines compared with whole-cell pertussis vaccine. Ad Hoc Group for the Study of Pertussis Vaccines**. *Lancet*, **350**(9091):1569–77, 1997.

[196] A. M. Wendelboe, A. Van Rie, S. Salmaso, and J. A. Englund. **Duration of immunity against pertussis after natural infection or vaccination.** *Pediatr Infect Dis J*, **24**(5 Suppl):S58–61, 2005.

[197] S. S. Long, H. W. Lischner, A. Deforest, and J. L. Clark. **Serologic evidence of subclinical pertussis in immunized children**. *Pediatr Infect Dis J*, **9**(10):700–5, 1990.

[198] F. G. Versteegh, J. F. Schellekens, A. F. Nagelkerke, and J. J. Roord. **Laboratory-confirmed reinfections with Bordetella pertussis**. *Acta Paediatr*, **91**(1):95–7, 2002.

[199] H. W. Hethcote. **An age-structured model for pertussis transmission**. *Math Biosci*, **145**(2):89–136, 1997.

[200] B. A. Halperin, A. Morris, D. Mackinnon-Cameron, J. Mutch, J. M. Langley, S. A. McNeil, D. Macdougall, and S. A. Halperin. **Kinetics of the antibody response to tetanus-diphtheria-acellular pertussis vaccine in women of childbearing age and postpartum women**. *Clin Infect Dis*, **53**(9):885–92, 2011.

[201] N. R. Kitchin, J. Southern, R. Morris, F. Hemme, S. Thomas, M. W. Watson, K. Cartwright, and E. Miller. **Evaluation of a diphtheria-tetanus-acellular pertussis-inactivated poliovirus-*Haemophilus influenzae* type b vaccine given concurrently with meningococcal group C conjugate vaccine at 2, 3 and 4 months of age**. *Arch Dis Child*, **92**(1):11–6, 2007.

[202] S. M. Kerfoot, E. M. Long, M. J. Hickey, G. Andonegui, B. M. Lapointe, R. C. Zanardo, C. Bonder, W. G. James, S. M. Robbins, and P. Kubes. **TLR4 contributes to disease-inducing mechanisms resulting in central nervous system autoimmune disease**. *J Immunol*, **173**(11):7070–7, 2004.

[203] F. Fransen, R. M. Stenger, M. C. Poelen, H. H. van Dijken, B. Kuipers, C. J. Boog, J. P. van Putten, C. A. van Els, and P. van der Ley. **Differential effect of TLR2 and TLR4 on the immune response after immunization with a vaccine against Neisseria meningitidis or Bordetella pertussis.** *PLoS One*, **5**(12):e15692, 2010.

[204] Z. Vadasz, E. Bamberger, K. Ben-Tikva, M. Feterman, I. Srugo, and A. Kessel. **Toll-like receptor-4 expression in infants with pertussis infection**. *Infection*, **41**(1):195–8, 2013.

[205] C. L. Collins, P. Salt, N. McCarthy, T. Chantler, L. Lane, F. Hemme, L. Diggle, J. Buttery, N. R. Kitchin, E. R. Moxon, and A. J. Pollard. **Immunogenicity and safety of a low-dose diphtheria, tetanus and acellular pertussis combination vaccine with either inactivated or oral polio vaccine as a pre-school booster in UK children**. *Vaccine*, **22**(31-32):4262–9, 2004.

[206] F. Faul, E. Erdfelder, A. G. Lang, and A. Buchner. **G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences**. *Behav Res Methods*, **39**(2):175–91, 2007.

[207] C. A. Janeway and R. Medzhitov. **Innate immune recognition**. *Annu Rev Immunol*, **20**:197–216, 2002.

[208] T. H. Mogensen. **Pathogen recognition and inflammatory signaling in innate immune defenses**. *Clin Microbiol Rev*, **22**(2):240–73, Table of Contents, 2009.

[209] S. Banus, R. M. Stenger, E. R. Gremmer, J. A. Dormans, F. R. Mooi, T. G. Kimman, and R. J. Vandebriel. **The role of Toll-like receptor-4 in pertussis vaccine-induced immunity**. *BMC Immunol*, **9**:21, 2008.

[210] D. M. Agnese, J. E. Calvano, S. J. Hahm, S. M. Coyle, S. A. Corbett, S. E. Calvano, and S. F. Lowry. **Human toll-like receptor 4 mutations but not CD14 polymorphisms are associated with an increased risk of gram-negative infections**. *J Infect Dis*, **186**(10):1522–5, 2002.

[211] N. C. Arbour, E. Lorenz, B. C. Schutte, J. Zabner, J. N. Kline, M. Jones, K. Frees, J. L. Watt, and D. A. Schwartz. **TLR4 mutations are associated with endotoxin hyporesponsiveness in humans**. *Nat Genet*, **25**(2):187–91, 2000.

[212] C. L. Trotter, N. J. Andrews, E. B. Kaczmarski, E. Miller, and M. E. Ramsay. **Effectiveness of meningococcal serogroup C conjugate vaccine 4 years after introduction**. *Lancet*, **364**(9431):365–7, 2004.

[213] MD Snape, DF Kelly, P Salt, S Green, C Snowden, L Diggle, A Borkowski, LM Yu, ER Moxon, and AJ Pollard. **Serogroup C meningococcal glycoconjugate vaccine in adolescents: persistence of bactericidal antibodies and kinetics of the immune response to a booster vaccine more than 3 years after immunization.** *Clin Infect Dis*, **43**(11):1387–94, 2006.

[214] MD Snape, DF Kelly, S Lewis, C Banner, L Kibwana, CE Moore, L Diggle, T John, LM Yu, R Borrow, A Borkowski, C Nau, and AJ Pollard. **Seroprotection against serogroup C meningococcal disease in adolescents in the United Kingdom: observational study.** *BMJ*, **336**(7659):1487–91, 2008.

[215] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al. **The sequence of the human genome**. *Science*, **291**(5507):1304–51, 2001.

[216] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner,

A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. **Initial sequencing and analysis of the human genome**. *Nature*, **409**(6822):860–921, 2001.

[217] International Human Genome Sequencing Consortium. **Finishing the euchromatic sequence of the human genome**. *Nature*, **431**(7011):931–45, 2004.

[218] K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R. C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, W. Sun, H. Wang, Y. Wang, X. Xiong, L. Xu, M. M. Waye, S. K. Tsui, H. Xue, J. T. Wong, L. M. Galver, J. B. Fan, K. Gunderson, S. S. Murray, A. R. Oliphant, M. S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J. F. Olivier, M. S. Phillips, S. Roumy, C. Sallée, A. Verner, T. J. Hudson, P. Y. Kwok, D. Cai, D. C. Koboldt, R. D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L. C. Tsui, W. Mak, Y. Q. Song, P. K. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, et al. **A second generation human haplotype map of over 3.1 million SNPs**. *Nature*, **449**(7164):851–61, 2007.

[219] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio. **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits**. *Proc Natl Acad Sci U S A*, **106**(23):9362–7, 2009.

[220] LA. Hindorff, J. MacArthur, J. Morales, HA. Junkins, PN. Hall, AK. Klemm, and TA. Manolio. **A Catalog of Published Genome-Wide Association Studies, www.genome.gov/gwastudies**, Accessed [December 2013].

[221] O. Delaneau, J. F. Zagury, and J. Marchini. **Improved whole-chromosome phasing for disease and population genetic studies**. *Nat Methods*, **10**(1):5–6, 2013.

[222] B. E. Huang, C. I. Amos, and D. Y. Lin. **Detecting haplotype effects in genomewide association studies**. *Genet Epidemiol*, **31**(8):803–12, 2007.

[223] B. N. Howie, P. Donnelly, and J. Marchini. **A flexible and accurate genotype imputation method for the next generation of genome-wide association studies**. *PLoS Genet*, **5**(6):e1000529, 2009.

[224] Y. Li, C. J. Willer, J. Ding, P. Scheet, and G. R. Abecasis. **MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes**. *Genet Epidemiol*, **34**(8):816–34, 2010.

[225] S. R. Browning and B. L. Browning. **Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering**. *Am J Hum Genet*, **81**(5):1084–97, 2007.

[226] O. Delaneau, J. Marchini, and J. F. Zagury. **A linear complexity phasing method for thousands of genomes**. *Nat Methods*, **9**(2):179–81, 2012.

[227] J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly. **A new multipoint method for genome-wide association studies by imputation of genotypes**. *Nat Genet*, **39**(7):906–13, 2007.

[228] CC Spencer, Z Su, P Donnelly, and J Marchini. **Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip.** *PLoS Genet*, **5**(5):e1000477, 2009.

[229] B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, and G. R. Abecasis. **Fast and accurate genotype imputation in genome-wide association studies through pre-phasing**. *Nat Genet*, **44**(8):955–9, 2012.

[230] J. Marchini and B. Howie. **Genotype imputation for genome-wide association studies**. *Nat Rev Genet*, **11**(7):499–511, 2010.

[231] M.C.J. Maiden, A.B. Ibarz-Pavón, R. Urwin, S.J. Gray, N.J. Andrews, S.C. Clarke, A.M. Walker, M.R. Evans, J.S. Kroll, K.R. Neal, D.A.A. Ala'Aldeen, D.W. Crook, K. Cann, S. Harrison, R. Cunningham, D. Baxter, E. Kaczmarski, J. MacLennan, J.C. Cameron, and J.M. Stuart. **Impact of meningococcal serogroup C conjugate vaccines on carriage and herd immunity**. *Journal of Infectious Diseases*, **197**(5):737–743, 2008.

[232] M. E. Ramsay, N. J. Andrews, C. L. Trotter, E. B. Kaczmarski, and E. Miller. **Herd immunity from meningococcal serogroup C conjugate vaccination in England: database analysis**. *BMJ*, **326**(7385):365–6, 2003.

[233] R. A. Manz, A. E. Hauser, F. Hiepe, and A. Radbruch. **Maintenance of serum antibody levels**. *Annu Rev Immunol*, **23**:367–86, 2005.

[234] M.C.J. Maiden and J.M. Stuart. **Carriage of serogroup C meningococci 1 year after meningococcal C conjugate polysaccharide vaccination**. *Lancet*, **359**(9320):1829–1830, 2002.

[235] C. A. Anderson, F. H. Pettersson, G. M. Clarke, L. R. Cardon, A. P. Morris, and K. T. Zondervan. **Data quality control in genetic case-control association studies**. *Nat Protoc*, **5**(9):1564–73, 2010.

[236] GenABEL project developers. *GenABEL: genome-wide SNP association analysis*, 2013. R package version 1.8-0.

[237] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. **Principal components analysis corrects for stratification in genome-wide association studies**. *Nat Genet*, **38**(8):904–9, 2006.

[238] Stephane Champely. *pwr: Basic functions for power analysis*, 2012. R package version 1.1.1.

[239] GF Santos, RR Deck, J Donnelly, W Blackwelder, and DM Granoff. **Importance of complement source in measuring meningococcal bactericidal titers.** *Clin Diagn Lab Immunol*, **8**(3):616–23, 2001.

[240] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

[241] Scott Melville. *NCBI2R: NCBI2R-An R package to navigate and annotate genes and SNPs*, 2014. R package version 1.4.5.

[242] J. P. Higgins, S. G. Thompson, J. J. Deeks, and D. G. Altman. **Measuring inconsistency in meta-analyses**. *BMJ*, **327**(7414):557–60, 2003.

[243] P. I. de Bakker, M. A. Ferreira, X. Jia, B. M. Neale, S. Raychaudhuri, and B. F. Voight. **Practical aspects of imputation-driven meta-analysis of genome-wide association studies**. *Hum Mol Genet*, **17**(R2):R122–8, 2008.

[244] R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, and C. J. Willer. **LocusZoom: regional visualization of genome-wide association scan results**. *Bioinformatics*, **26**(18):2336–7, 2010.

[245] A. Zuko, K. T. Kleijer, A. Oguro-Ando, M. J. Kas, E. van Daalen, B. van der Zwaag, and J. P. Burbach. **Contactins in the neurobiology of autism**. *Eur J Pharmacol*, **719**(1-3):63–74, 2013.

[246] M. J. Barber, L. M. Mangravite, C. L. Hyde, D. I. Chasman, J. D. Smith, C. A. McCarty, X. Li, R. A. Wilke, M. J. Rieder, P. T. Williams, P. M. Ridker, A. Chatterjee, J. I. Rotter, D. A. Nickerson, M. Stephens, and R. M. Krauss. **Genome-wide association of lipid-lowering response to statins in combined study populations**. *PLoS One*, **5**(3):e9763, 2010.

[247] C. S. Fox, Y. Liu, C. C. White, M. Feitosa, A. V. Smith, N. Heard-Costa, K. Lohman, A. D. Johnson, M. C. Foster, D. M. Greenawalt, P. Griffin, J. Ding, A. B. Newman, F. Tylavsky, I. Miljkovic, S. B. Kritchevsky, L. Launer, M. Garcia, G. Eiriksdottir, J. J. Carr, V. Gudnason, T. B. Harris, L. A. Cupples, I. B. Borecki, GIANT Consortium, MAGIC Consortium, and GLGC Consortium. **Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women**. *PLoS Genet*, **8**(5), 2012.

[248] K. Wang, H. Zhang, C. S. Bloss, V. Duvvuri, W. Kaye, N. J. Schork, W. Berrettini, H. Hakonarson, and Price Foundation Collaborative Group. **A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa**. *Mol Psychiatry*, **16**(9):949–59, 2011.

[249] E. J. Benjamin, J. Dupuis, M. G. Larson, K. L. Lunetta, S. L. Booth, D. R. Govindaraju, S. Kathiresan, J. F. Keaney, M. J. Keyes, J. P. Lin, J. B. Meigs, S. J. Robins, J. Rong, R. Schnabel, J. A. Vita, T. J. Wang, P. W. Wilson, P. A. Wolf, and R. S. Vasan. **Genome-wide association with select biomarker traits in the Framingham Heart Study**. *BMC Med Genet*, **8 Suppl 1**:S11, 2007.

[250] T. T. Lu and J. G. Cyster. **Integrin-mediated long-term B cell retention in the splenic marginal zone**. *Science*, **297**(5580):409–12, 2002.

[251] X. Wang, L. B. Rodda, O. Bannard, and J. G. Cyster. **Integrin-mediated interactions between B cells and follicular dendritic cells influence germinal center B cell fitness**. *J Immunol*, **192**(10):4601–9, 2014.

[252] K. A. Sutton, M. K. Jungnickel, Y. Wang, K. Cullen, S. Lambert, and H. M. Florman. **Enkurin is a novel calmodulin and TRPC channel binding protein in sperm**. *Dev Biol*, **274**(2):426–35, 2004.

[253] A. D. Paterson, D. Waggott, A. P. Boright, S. M. Hosseini, E. Shen, M. P. Sylvestre, I. Wong, B. Bharaj, P. A. Cleary, J. M. Lachin, J. E. Below, D. Nicolae, N. J. Cox, A. J. Canty, L. Sun, S. B. Bull, MAGIC (Meta-Analyses of Glucose Consortium), Insulin related traits, Diabetes Control Group, Complications Trial/Epidemiology of Diabetes Interventions, and Complications Research. **A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose**. *Diabetes*, **59**(2):539–49, 2010.

[254] L. Jostins, S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. P. Schumm, Y. Sharma, C. A. Anderson, J. Essers, M. Mitrovic, K. Ning, I. Cleynen, E. Theatre, S. L. Spain, S. Raychaudhuri, P. Goyette, Z. Wei, C. Abraham, J. P. Achkar, T. Ahmad, L. Amininejad, A. N. Ananthakrishnan, V. Andersen, J. M. Andrews, L. Baidoo, T. Balschun, P. A. Bampton, A. Bitton, G. Boucher, S. Brand, C. B§ning, A. Cohain, S. Cichon, M. D'Amato, D. De Jong, K. L. Devaney, M. Dubinsky, C. Edwards, D. Ellinghaus, L. R. Ferguson, D. Franchimont, K. Fransen, R. Gearry, M. Georges, C. Gieger, J. Glas, T. Haritunians, A. Hart, C. Hawkey, M. Hedl, X. Hu, T. H. Karlsen, L. Kupcinskas, S. Kugathasan, A. Latiano, D. Laukens, I. C. Lawrance, C. W. Lees, E. Louis, G. Mahy, J. Mansfield, A. R. Morgan, C. Mowat, W. Newman, O. Palmieri, C. Y. Ponsioen, U. Potocnik, N. J. Prescott, M. Regueiro, J. I. Rotter, R. K. Russell, J. D. Sanderson, M. Sans, J. Satsangi, S. Schreiber, L. A. Simms, J. Sventoraityte, S. R. Targan, K. D. Taylor, M. Tremelling, H. W. Verspaget, M. De Vos, C. Wijmenga, D. C. Wilson, J. Winkelmann, R. J. Xavier, S. Zeissig, B. Zhang, C. K. Zhang, H. Zhao, M. S. Silverberg, V. Annese, H. Hakonarson, S. R. Brant, G. Radford-Smith, C. G. Mathew, J. D. Rioux, E. E. Schadt, et al. **Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease**. *Nature*, **491**(7422):119–24, 2012.

[255] J. B. Maller, G. McVean, J. Byrnes, D. Vukcevic, K. Palin, Z. Su, J. M. Howson, A. Auton, S. Myers, A. Morris, M. Pirinen, M. A. Brown, P. R. Burton, M. J. Caulfield, A. Compston, M. Farrall, A. S. Hall, A. T. Hattersley, A. V. Hill, C. G. Mathew, M. Pembrey, J. Satsangi, M. R. Stratton, J. Worthington, N. Craddock, M. Hurles, W. Ouwehand, M. Parkes, N. Rahman, A. Duncanson, J. A. Todd, D. P. Kwiatkowski, N. J. Samani, S. C. Gough, M. I. McCarthy, P. Deloukas, P. Donnelly, and Wellcome Trust Case Control Consortium. **Bayesian refinement of association signals for 14 loci in 3 common diseases**. *Nat Genet*, **44**(12):1294–301, 2012.

[256] A. P. Boyle, E. L. Hong, M. Hariharan, Y. Cheng, M. A. Schaub, M. Kasowski, K. J. Karczewski, J. Park, B. C. Hitz, S. Weng, J. M. Cherry, and M. Snyder. **Annotation of functional variation in personal genomes using RegulomeDB**. *Genome Res*, **22**(9):1790–7, 2012.

[257] E. T. Cirulli and D. B. Goldstein. **Uncovering the roles of rare variants in common disease through whole-genome sequencing.** *Nat Rev Genet*, **11**(6):415–25, 2010.

[258] M. Parkes, A. Cortes, D. A. van Heel, and M. A. Brown. **Genetic insights into common pathways and complex relationships among immune-mediated diseases**. *Nat Rev Genet*, **14**(9):661–73, 2013.

[259] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, and H. Parkinson. **The NHGRI GWAS Catalog, a curated resource of SNP-trait associations**. *Nucleic Acids Res*, **42**(Database issue):D1001–6, 2014.

[260] I. S. Lossos, A. A. Alizadeh, R. Rajapaksa, R. Tibshirani, and R. Levy. **HGAL is a novel interleukin-4-inducible gene that strongly predicts survival in diffuse large B-cell lymphoma**. *Blood*, **101**(2):433–40, 2003.

[261] Y. Natkunam, E. D. Hsi, P. Aoun, S. Zhao, P. Elson, B. Pohlman, H. Naushad, M. Bast, R. Levy, and I. S. Lossos. **Expression of the human germinal center-associated lymphoma (HGAL) protein identifies a subset of classic Hodgkin lymphoma of germinal center derivation and improved survival**. *Blood*, **109**(1):298–305, 2007.

[262] I. Romero-Camarero, X. Jiang, Y. Natkunam, X. Lu, C. Vicente-Dueñas, I. Gonzalez-Herrero, T. Flores, J. L. Garcia, G. McNamara, C. Kunder, S. Zhao, V. Segura, L. Fontan, J. A. Martinez-Climent, F. J. Garcia-Criado, J. D. Theis, A. Dogan, E. Campos-Sánchez, M. R. Green, A. A. Alizadeh, C. Cobaleda, I. Sánchez-García, and I. S. Lossos. **Germinal centre protein HGAL promotes lymphoid hyperplasia and amyloidosis via BCR-mediated Syk activation**. *Nat Commun*, **4**:1338, 2013.

[263] J. Choe, H. S. Kim, R. J. Armitage, and Y. S. Choi. **The functional role of B cell antigen receptor stimulation and IL-4 in the generation of human memory B cells from germinal center B cells**. *J Immunol*, **159**(8):3757–66, 1997.

[264] H. J. Westra, M. J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, M. W. Christiansen, B. P. Fairfax, K. Schramm, J. E. Powell, A. Zhernakova, D. V. Zhernakova, J. H. Veldink, L. H. Van den Berg, J. Karjalainen, S. Withoff, A. G. Uitterlinden, A. Hofman, F. Rivadeneira, P. A. 't Hoen, E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez, M. A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Völker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dicey, S. A. Gharib, D. A. Enquobahrie, T. Lumley, G. W. Montgomery, S. Makino, H. Prokisch, C. Herder, M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R. C. Jansen, P. M. Visscher, J. C. Knight, B. M. Psaty, S. Ripatti, A. Teumer, T. M. Frayling, A. Metspalu, J. B. van Meurs, and L. Franke. **Systematic identification of trans eQTLs as putative drivers of known disease associations**. *Nat Genet*, **45**(10):1238–43, 2013.

[265] B. P. Fairfax, P. Humburg, S. Makino, V. Naranbhai, D. Wong, E. Lau, L. Jostins, K. Plant, R. Andrews, C. McGee, and J. C. Knight. **Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression**. *Science*, **343**(6175):1246949, 2014.

[266] S. A. Spector, M. Qin, J. Lujan-Zilbermann, K. K. Singh, M. G. Warshaw, P. L. Williams, P. Jean-Philippe, T. Fenton, G. K. Siberry, and IMPAACT P1065 Protocol Team. **Genetic variants in toll-like receptor 2 (TLR2), TLR4, TLR9, and FCγ receptor II are associated with antibody response to quadrivalent meningococcal conjugate vaccine in HIV-infected youth**. *Clin Vaccine Immunol*, **20**(6):900–6, 2013.

[267] G. K. Siberry, P. L. Williams, J. Lujan-Zilbermann, M. G. Warshaw, S. A. Spector, M. D. Decker, B. E. Heckman, E. F. Demske, J. S. Read, P. Jean-Philippe, W. Kabat, S. Nachman, and IMPAACT P1065 Protocol Team. **Phase I/II, open-label trial of safety and immunogenicity of meningococcal (groups A, C, Y, and W-135) polysaccharide diphtheria toxoid conjugate vaccine in human immunodeficiency virus-infected adolescents**. *Pediatr Infect Dis J*, **29**(5):391–6, 2010.

[268] L. R. Cardon and L. J. Palmer. **Population stratification and spurious allelic association**. *Lancet*, **361**(9357):598–604, 2003.

[269] R. Borrow, D. Goldblatt, N. Andrews, J. Southern, L. Ashton, S. Deane, R. Morris, K. Cartwright, and E. Miller. **Antibody persistence and immunological memory at age 4 years after meningococcal group C conjugate vaccination in children in the United kingdom**. *J Infect Dis*, **186**(9):1353–7, 2002.

[270] M. E. Ramsay, J. McVernon, N. J. Andrews, P. T. Heath, and M. P. Slack. **Estimating *Haemophilus influenzae* type b vaccine effectiveness in England and Wales by use of the screening method**. *J Infect Dis*, **188**(4):481–5, 2003.

[271] D. F. Kelly, A. J. Pollard, and E. R. Moxon. **Immunological memory: the role of B cells in long-term protection against invasive bacterial pathogens**. *JAMA*, **294**(23):3019–23, 2005.

[272] P. B. Beeson. **Age and sex associations of 40 autoimmune diseases**. *Am J Med*, **96**(5):457–62, 1994.

[273] G. Gavazzi and K. H. Krause. **Ageing and infection**. *Lancet Infect Dis*, **2**(11):659–66, 2002.

[274] J. Kovarik and C. A. Siegrist. **Immunity in early life**. *Immunol Today*, **19**(4):150–2, 1998.

[275] J. Rowe, C. Macaubas, T. Monger, B. J. Holt, J. Harvey, J. T. Poolman, R. Loh, P. D. Sly, and P. G. Holt. **Heterogeneity in diphtheria-tetanus-acellular pertussis vaccine-specific cellular immunity during infancy: relationship to variations in the kinetics of postnatal maturation of systemic th1 function**. *J Infect Dis*, **184**(1):80–8, 2001.

[276] C. A. Koch and J. L. Platt. **Natural mechanisms for evading graft rejection: the fetus as an allograft**. *Springer Semin Immunopathol*, **25**(2):95–117, 2003.

[277] P. G. Holt, J. B. Clough, B. J. Holt, M. J. Baron-Hay, A. H. Rose, B. W. Robinson, and W. R. Thomas. **Genetic 'risk' for atopy is associated with delayed postnatal maturation of T-cell competence**. *Clin Exp Allergy*, **22**(12):1093–9, 1992.

[278] R. Lamason, P. Zhao, R. Rawat, A. Davis, J. C. Hall, J. J. Chae, R. Agarwal, P. Cohen, A. Rosen, E. P. Hoffman, and K. Nagaraju. **Sexual dimorphism in immune response genes as a function of puberty**. *BMC Immunol*, **7**:2, 2006.

[279] L. Dumitrescu, C. L. Carty, N. Franceschini, L. A. Hindorff, S. A. Cole, P. Bužková, F. R. Schumacher, C. B. Eaton, R. J. Goodloe, D. J. Duggan, J. Haessler, B. Cochran, B. E. Henderson, I. Cheng, K. C. Johnson, C. S. Carlson, S. A. Love, K. Brown-Gentry, A. Q. Nato, M. Quibrera, G. Anderson, R. V. Shohet, J. L. Ambite, L. R. Wilkens, L. L. Marchand, C. A. Haiman, S. Buyske, C. Kooperberg, K. E. North, M. Fornage, and D. C. Crawford. **Post-genome-wide association study challenges for lipid traits: describing age as a modifier of gene-lipid associations in the Population Architecture using Genomics and Epidemiology (PAGE) study**. *Ann Hum Genet*, **77**(5):416–25, 2013.

[280] M. F. Moffatt, I. G. Gut, F. Demenais, D. P. Strachan, E. Bouzigon, S. Heath, E. von Mutius, M. Farrall, M. Lathrop, W. O. Cookson, and GABRIEL Consortium. **A large-scale, consortium-based genomewide association study of asthma**. *N Engl J Med*, **363**(13):1211–21, 2010.

[281] J. Graham, W. A. Hagopian, I. Kockum, L. S. Li, C. B. Sanjeevi, R. M. Lowe, J. B. Schaefer, M. Zarghami, H. L. Day, M. Landin-Olsson, J. P. Palmer, M. Janer-Villanueva, L. Hood, G. Sundkvist, A. Lernmark, N. Breslow, G. Dahlquist, G. Blohmé, Diabetes Incidence in Sweden Study Group,

and Swedish Childhood Diabetes Study Group. **Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes**. *Diabetes*, **51**(5):1346–55, 2002.

[282] C. S. Waddington, W. T. Walker, C. Oeser, A. Reiner, T. John, S. Wilkins, M. Casey, P. E. Eccleston, R. J. Allen, I. Okike, S. Ladhani, E. Sheasby, K. Hoschler, N. Andrews, P. Waight, A. C. Collinson, P. T. Heath, A. Finn, S. N. Faust, M. D. Snape, E. Miller, and A. J. Pollard. **Safety and immunogenicity of AS03B adjuvanted split virion versus non-adjuvanted whole virion H1N1 influenza vaccine in UK children aged 6 months-12 years: open label, randomised, parallel group, multicentre study**. *BMJ*, **340**:c2649, 2010.

[283] A. Khatami, E. A. Clutterbuck, A. J. Thompson, J. A. McKenna, D. Pace, J. Birks, M. D. Snape, and A. J. Pollard. **Evaluation of the Induction of Immune Memory following Infant Immunisation with Serogroup C Neisseria meningitidis Conjugate Vaccines - Exploratory Analyses within a Randomised Controlled Trial**. *PLoS One*, **9**(7):e101672, 2014.

[284] D. M. Church, V. A. Schneider, T. Graves, K. Auger, F. Cunningham, N. Bouk, H. C. Chen, R. Agarwala, W. M. McLaren, G. R. Ritchie, D. Albracht, M. Kremitzki, S. Rock, H. Kotkiewicz, C. Kremitzki, A. Wollam, L. Trani, L. Fulton, R. Fulton, L. Matthews, S. Whitehead, W. Chow, J. Torrance, M. Dunn, G. Harden, G. Threadgold, J. Wood, J. Collins, P. Heath, G. Griffiths, S. Pelan, D. Grafham, E. E. Eichler, G. Weinstock, E. R. Mardis, R. K. Wilson, K. Howe, P. Flicek, and T. Hubbard. **Modernizing reference genome assemblies**. *PLoS Biol*, **9**(7):e1001091, 2011.

[285] J.P Steibel and G Abecasis. **QpowR: Interactive power calculator for two-stage genetic association studies of quantitative traits**, 2008.

[286] D. M. Evans, C. C. Spencer, J. J. Pointon, Z. Su, D. Harvey, G. Kochan, U. Oppermann, U. Opperman, A. Dilthey, M. Pirinen, M. A. Stone, L. Appleton, L. Moutsianas, L. Moutsianis, S. Leslie, T. Wordsworth, T. J. Kenna, T. Karaderi, G. P. Thomas, M. M. Ward, M. H. Weisman, C. Farrar, L. A. Bradbury, P. Danoy, R. D. Inman, W. Maksymowych, D. Gladman, P. Rahman, A. Morgan, H. Marzo-Ortega, P. Bowness, K. Gaffney, J. S. Gaston, M. Smith, J. Bruges-Armas, A. R. Couto, R. Sorrentino, F. Paladini, M. A. Ferreira, H. Xu, Y. Liu, L. Jiang, C. Lopez-Larrea, R. Diaz-Peña, A. López-Vázquez, T. Zayats, G. Band, C. Bellenguez, H. Blackburn, J. M. Blackwell, E. Bramon, S. J. Bumpstead, J. P. Casas, A. Corvin, N. Craddock, P. Deloukas, S. Dronov, A. Duncanson, S. Edkins, C. Freeman, M. Gillman, E. Gray, R. Gwilliam, N. Hammond, S. E. Hunt, J. Jankowski, A. Jayakumar, C. Langford, J. Liddle, H. S. Markus, C. G. Mathew, O. T. McCann, M. I. McCarthy, C. N. Palmer, L. Peltonen, R. Plomin, S. C. Potter, A. Rautanen, R. Ravindrarajah, M. Ricketts, N. Samani, S. J. Sawcer, A. Strange, R. C. Trembath, A. C. Viswanathan, M. Waller, P. Weston, P. Whittaker, S. Widaa, N. W. Wood, G. McVean,

J. D. Reveille, B. P. Wordsworth, M. A. Brown, P. Donnelly, Spondyloarthritis Research Consortium of Canada (SPARCC), Australo-Anglo-American Spondyloarthritis Consortium (TASC), and Wellcome Trust Case Control Consortium 2 (WTCCC2). **Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility**. *Nat Genet*, **43**(8):761–7, 2011.

[287] M. Stefanidakis, G. Newton, W. Y. Lee, C. A. Parkos, and F. W. Luscinskas. **Endothelial CD47 interaction with SIRPgamma is required for human T-cell transendothelial migration under shear flow conditions in vitro**. *Blood*, **112**(4):1280–9, 2008.

[288] JPT Higgins and S Green. **Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]**, 2011.

[289] Z. Z. Shi, B. Han, G. M. Habib, M. M. Matzuk, and M. W. Lieberman. **Disruption of gamma-glutamyl leukotrienase results in disruption of leukotriene D(4) synthesis in vivo and attenuation of the acute inflammatory response**. *Mol Cell Biol*, **21**(16):5389–95, 2001.

[290] J. C. Barrett, D. G. Clayton, P. Concannon, B. Akolkar, J. D. Cooper, H. A. Erlich, C. Julier, G. Morahan, J. Nerup, C. Nierras, V. Plagnol, F. Pociot, H. Schuilenburg, D. J. Smyth, H. Stevens, J. A. Todd, N. M. Walker, S. S. Rich, and Type 1 Diabetes Genetics Consortium. **Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes**. *Nat Genet*, **41**(6):703–7, 2009.

[291] P. Anderson, M. Pichichero, R. Insel, P. Farsad, and M. Santosham. **Capsular antigens noncovalently or covalently associated with protein as vaccines to *Haemophilus influenzae* type b: comparison in two-year-old children**. *J Infect Dis*, **152**(3):634–6, 1985.

[292] H. Jakobsen, S. Hannesdottir, S. P. Bjarnarson, D. Schulz, E. Trannoy, C. A. Siegrist, and I. Jonsdottir. **Early life T cell responses to pneumococcal conjugates increase with age and determine the polysaccharide-specific antibody response and protective efficacy**. *Eur J Immunol*, **36**(2):287–95, 2006.

[293] K. K. Kamboj, C. L. King, N. S. Greenspan, H. L. Kirchner, and J. R. Schreiber. **Immunization with *Haemophilus influenzae* type b-CRM(197) conjugate vaccine elicits a mixed Th1 and Th2 CD(4+) T cell cytokine response that correlates with the isotype of antipolysaccharide antibody**. *J Infect Dis*, **184**(7):931–5, 2001.

[294] Z Lai and JR Schreiber. **Antigen processing of glycoconjugate vaccines; the polysaccharide portion of the pneumococcal CRM(197) conjugate vaccine co-localizes with MHC II on the antigen processing cell surface.** *Vaccine*, **27**(24):3137–44, 2009.

[295] F. Y. Avci, X. Li, M. Tsuji, and D. L. Kasper. **A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design**. *Nat Med*, **17**(12):1602–9, 2011.

[296] M. E. Tabangin, J. G. Woo, and L. J. Martin. **The effect of minor allele frequency on the likelihood of obtaining false positives**. *BMC Proc*, **3 Suppl 7**:S41, 2009.

[297] I. G. Ovsyannikova, R. A. Vierkant, V. S. Pankratz, R. M. Jacobson, and G. A. Poland. **Human leukocyte antigen genotypes in the genetic control of adaptive immune responses to smallpox vaccine**. *J Infect Dis*, **203**(11):1546–55, 2011.

[298] G. Blanchard-Rohner, M. D. Snape, D. F. Kelly, D. O'Connor, T. John, E. Kibwana, H. Parks, K. Ford, P. M. Dull, and A. J. Pollard. **Seroprevalence and placental transmission of maternal antibodies specific for Neisseria meningitidis Serogroups A, C, Y and W135 and influence of maternal antibodies on the immune response to a primary course of MenACWY-CRM vaccine in the United Kingdom**. *Pediatr Infect Dis J*, **32**(7):768–76, 2013.

[299] M. D. Snape, T. Dawson, P. Oster, A. Evans, T. M. John, B. Ohene-Kena, J. Findlow, L. M. Yu, R. Borrow, E. Ypma, D. Toneatto, and A. J. Pollard. **Immunogenicity of two investigational serogroup B meningococcal vaccines in the first year of life: a randomized comparative trial.** *Pediatr Infect Dis J*, **29**(11):e71–9, 2010.

[300] A. Khatami, A. Peters, H. Robinson, N. Williams, A. Thompson, H. Findlow, A. J. Pollard, and M. D. Snape. **Maintenance of immune response throughout childhood following serogroup C meningococcal conjugate vaccination in early childhood**. *Clin Vaccine Immunol*, **18**(12):2038–42, 2011.

[301] M. D. Snape, C. L. Klinger, E. D. Daniels, T. M. John, H. Layton, L. Rollinson, S. Pestridge, S. Dymond, E. Galiza, S. Tansey, D. A. Scott, S. A. Baker, T. R. Jones, L. M. Yu, W. C. Gruber, E. A. Emini, S. N. Faust, A. Finn, P. T. Heath, and A. J. Pollard. **Immunogenicity and reactogenicity of a 13-valent-pneumococcal conjugate vaccine administered at 2, 4, and 12 months of age: a double-blind randomized active-controlled trial**. *Pediatr Infect Dis J*, **29**(12):e80–90, 2010.

[302] C. L. Trotter, M. E. Ramsay, and M. P. Slack. **Rising incidence of *Haemophilus influenzae* type b disease in England and Wales indicates a need for a second catch-up vaccination campaign**. *Commun Dis Public Health*, **6**(1):55–8, 2003.

[303] M. Pittman. **Variation and type specificity in the bacterial species *Haemophilus influenzae***. *J Exp Med*, **53**(4):471–92, 1931.

[304] E. C. Anderson, N. T. Begg, S. C. Crawshaw, R. M. Hargreaves, A. J. Howard, and M. P. Slack. **Epidemiology of invasive *Haemophilus influenzae***

infections in England and Wales in the pre-vaccination era (1990-2). *Epidemiol Infect*, **115**(1):89–100, 1995.

[305] G. Tudor-Williams, J. Frankland, D. Isaacs, R. T. Mayon-White, J. A. MacFarlane, M. P. Slack, E. Anderson, D. G. Rees, and E. R. Moxon. ***Haemophilus influenzae*** **type b disease in the Oxford region**. *Arch Dis Child*, **64**(4):517–9, 1989.

[306] A. J. Howard, K. T. Dunkin, J. M. Musser, and S. R. Palmer. **Epidemiology of *Haemophilus influenzae* type b invasive disease in Wales**. *BMJ*, **303**(6800):441–5, 1991.

[307] R. Booy, S. A. Hodgson, M. P. Slack, E. C. Anderson, R. T. Mayon-White, and E. R. Moxon. **Invasive *Haemophilus influenzae* type b disease in the Oxford region (1985-91)**. *Arch Dis Child*, **69**(2):225–8, 1993.

[308] LeRoy Fothergill and Joyce Wright. **Influenzal meningitis: The relation of age incidence to the bactericidal power of blood against the causal organism**. *J Immunol*, **24**(4):273–284, 1933.

[309] R. Booy, S. Hodgson, L. Carpenter, R. T. Mayon-White, M. P. Slack, J. A. Macfarlane, E. A. Haworth, M. Kiddle, S. Shribman, and J. S. Roberts. **Efficacy of *Haemophilus influenzae* type b conjugate vaccine PRP-T**. *Lancet*, **344**(8919):362–6, 1994.

[310] H. Peltola, H. Käyhty, M. Virtanen, and P. H. Mäkelä. **Prevention of *Haemophilus influenzae* type b bacteremic infections with the capsular polysaccharide vaccine**. *N Engl J Med*, **310**(24):1561–6, 1984.

[311] S. B. Black, H. R. Shinefield, B. Fireman, R. Hiatt, M. Polen, and E. Vittinghoff. **Efficacy in infancy of oligosaccharide conjugate *Haemophilus influenzae* type b (HbOC) vaccine in a United States population of 61,080 children. The Northern California Kaiser Permanente Vaccine Study Center Pediatrics Group**. *Pediatr Infect Dis J*, **10**(2):97–104, 1991.

[312] J. Eskola, J. Ward, R. Dagan, D. Goldblatt, F. Zepp, and C. A. Siegrist. **Combined vaccination of *Haemophilus influenzae* type b conjugate and diphtheria-tetanus-pertussis containing acellular pertussis**. *Lancet*, **354**(9195):2063–8, 1999.

[313] J McVernon, N Andrews, MP Slack, and ME Ramsay. **Risk of vaccine failure after *Haemophilus influenzae* type b (Hib) combination vaccines with acellular pertussis.** *Lancet*, **361**(9368):1521–3, 2003.

[314] C. L. Trotter, J. McVernon, N. J. Andrews, M. Burrage, and M. E. Ramsay. **Antibody to *Haemophilus influenzae* type b after routine and catch-up vaccination**. *Lancet*, **361**(9368):1523–4, 2003.

[315] J. McVernon, A. J. Howard, M. P. Slack, and M. E. Ramsay. **Long-term impact of vaccination on *Haemophilus influenzae* type b (Hib) carriage in the United Kingdom**. *Epidemiol Infect*, **132**(4):765–7, 2004.

[316] G. R. Siber, M. Santosham, G. R. Reid, C. Thompson, J. Almeido-Hill, A. Morell, G. deLange, J. K. Ketcham, and E. H. Callahan. **Impaired antibody response to *Haemophilus influenzae* type b polysaccharide and low IgG2 and IgG4 concentrations in Apache children.** *N Engl J Med*, **323**(20):1387–92, 1990.

[317] J. Ward, G. Brenneman, G. W. Letson, and W. L. Heyward. **Limited efficacy of a *Haemophilus influenzae* type b conjugate vaccine in Alaska Native infants. The Alaska H. influenzae Vaccine Study Group**. *N Engl J Med*, **323**(20):1393–401, 1990.

[318] J. P. Pandey, H. H. Fudenberg, G. Virella, C. U. Kyong, C. B. Loadholt, R. M. Galbraith, E. C. Gotschlich, and J. C. Parke. **Association between immunoglobulin allotypes and immune responses to *Haemophilus influenzae* and Meningococcus polysaccharides**. *Lancet*, **1**(8109):190–2, 1979.

[319] D. M. Ambrosino, G. Schiffman, E. C. Gotschlich, P. H. Schur, G. A. Rosenberg, G. G. DeLange, E. van Loghem, and G. R. Siber. **Correlation between G2m(n) immunoglobulin allotype and human antibody response and susceptibility to polysaccharide encapsulated bacteria**. *J Clin Invest*, **75**(6):1935–42, 1985.

[320] D. M. Granoff, B. K. Suarez, J. P. Pandey, and P. G. Shackelford. **Genes associated with the G2m(23) immunoglobulin allotype regulate the IgG subclass responses to *Haemophilus influenzae* type b polysaccharide vaccine**. *J Infect Dis*, **157**(6):1142–9, 1988.

[321] S. V. Hetherington and M. L. Lepow. **Correlation between antibody affinity and serum bactericidal activity in infants**. *J Infect Dis*, **165**(4):753–6, 1992.

[322] J. V. Ravetch and J. P. Kinet. **Fc receptors**. *Annu Rev Immunol*, **9**:457–92, 1991.

[323] M. Torres and A. Casadevall. **The immunoglobulin constant region contributes to affinity and specificity**. *Trends Immunol*, **29**(2):91–7, 2008.

[324] R. Booy, S. A. Taylor, S. R. Dobson, D. Isaacs, G. Sleight, S. Aitken, H. Griffiths, H. Chapel, R. T. Mayon-White, and J. A. Macfarlane. **Immunogenicity and safety of PRP-T conjugate vaccine given according to the British accelerated immunisation schedule**. *Arch Dis Child*, **67**(4):475–8, 1992.

[325] H. Hu, B. Wang, M. Borde, J. Nardone, S. Maika, L. Allred, P. W. Tucker, and A. Rao. **Foxp1 is an essential transcriptional regulator of B cell development**. *Nat Immunol*, **7**(8):819–26, 2006.

[326] C. Shi, M. Sakuma, T. Mooroka, A. Liscoe, H. Gao, K. J. Croce, A. Sharma, D. Kaplan, D. R. Greaves, Y. Wang, and D. I. Simon. **Down-regulation of the forkhead transcription factor Foxp1 is required for monocyte differentiation and macrophage function**. *Blood*, **112**(12):4699–711, 2008.

[327] X. Feng, G. C. Ippolito, L. Tian, K. Wiehagen, S. Oh, A. Sambandam, J. Willen, R. M. Bunte, S. D. Maika, J. V. Harriss, A. J. Caton, A. Bhandoola, P. W. Tucker, and H. Hu. **Foxp1 is an essential transcriptional regulator for the generation of quiescent naive T cells during thymocyte development**. *Blood*, **115**(3):510–8, 2010.

[328] C. E. Busse, I. Czogiel, P. Braun, P. F. Arndt, and H. Wardemann. **Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes**. *Eur J Immunol*, **44**(2):597–603, 2014.

[329] A. Sagardoy, J. I. Martinez-Ferrandis, S. Roa, K. L. Bunting, M. A. Aznar, O. Elemento, R. Shaknovich, L. Fontán, V. Fresquet, I. Perez-Roger, E. F. Robles, L. De Smedt, X. Sagaert, A. Melnick, and J. A. Martinez-Climent. **Downregulation of FOXP1 is required during germinal center B-cell function**. *Blood*, **121**(21):4311–20, 2013.

[330] Y. Jin, S. A. Birlea, P. R. Fain, C. M. Mailloux, S. L. Riccardi, K. Gowan, P. J. Holland, D. C. Bennett, M. R. Wallace, W. T. McCormack, E. H. Kemp, D. J. Gawkrodger, A. P. Weetman, M. Picardo, G. Leone, A. TaŢeb, T. Jouary, K. Ezzedine, N. van Geel, J. Lambert, A. Overbeck, and R. A. Spritz. **Common variants in FOXP1 are associated with generalized vitiligo**. *Nat Genet*, **42**(7):576–8, 2010.

[331] B. Streubel, U. Vinatzer, A. Lamprecht, M. Raderer, and A. Chott. **T(3;14)(p14.1;q32) involving IGH and FOXP1 is a novel recurrent chromosomal aberration in MALT lymphoma**. *Leukemia*, **19**(4):652–8, 2005.

[332] E. E. Adderson, P. G. Shackelford, R. A. Insel, A. Quinn, P. M. Wilson, and W. L. Carroll. **Immunoglobulin light chain variable region gene sequences for human antibodies to *Haemophilus influenzae* type b capsular polysaccharide are dominated by a limited number of V kappa and V lambda segments and VJ combinations**. *J Clin Invest*, **89**(3):729–38, 1992.

[333] S. Ladhani, P. T. Heath, M. E. Ramsay, M. P. Slack, E. Kibwana, A. J. Pollard, and R. Booy. **Long-term immunological follow-up of children with *Haemophilus influenzae* serotype b vaccine failure in the United Kingdom**. *Clin Infect Dis*, **49**(3):372–80, 2009.

[334] E. E. Adderson, P. G. Shackelford, A. Quinn, P. M. Wilson, M. W. Cunningham, R. A. Insel, and W. L. Carroll. **Restricted immunoglobulin VH usage and VDJ combinations in the human response to *Haemophilus influenzae* type b capsular polysaccharide. Nucleotide sequences of monospecific anti-Haemophilus antibodies and polyspecific antibodies cross-reacting with self antigens**. *J Clin Invest*, **91**(6):2734–43, 1993.

[335] A. H. Lucas, J. W. Larrick, and D. C. Reason. **Variable region sequences of a protective human monoclonal antibody specific for the *Haemophilus influenzae* type b capsular polysaccharide**. *Infect Immun*, **62**(9):3873–80, 1994.

[336] G. V. Pinchuk, C. Nottenburg, and E. C. Milner. **Predominant V-region gene configurations in the human antibody response to *Haemophilus influenzae* capsule polysaccharide**. *Scand J Immunol*, **41**(4):324–30, 1995.

[337] N. S. Galbraith, P. Forbes, and H. Tillett. **National surveillance of tetanus in England and Wales 1930-79**. *J Infect*, **3**(2):181–91, 1981.

[338] A. A. Rushdy, J. M. White, M. E. Ramsay, and N. S. Crowcroft. **Tetanus in England and Wales, 1984-2000**. *Epidemiol Infect*, **130**(1):71–7, 2003.

[339] World Health Organisation WHO. **World Health Organisation: Tetanus factsheet, www.wpro.who.int/mediacentre/factsheets/** , 2014.

[340] B. Yucesoy, Y. Talzhanov, V. J. Johnson, N. W. Wilson, R. E. Biagini, W. Wang, B. Frye, D. N. Weissman, D. R. Germolec, M. I. Luster, and M. M. Barmada. **Genetic variants within the MHC region are associated with immune responsiveness to childhood vaccinations**. *Vaccine*, **31**(46):5381–91, 2013.

[341] S. Leslie, P. Donnelly, and G. McVean. **A statistical method for predicting classical HLA alleles from SNP data**. *Am J Hum Genet*, **82**(1):48–56, 2008.

[342] A. T. Dilthey, L. Moutsianas, S. Leslie, and G. McVean. **HLA*IMP–an integrated framework for imputing classical HLA alleles from SNP genotypes**. *Bioinformatics*, **27**(7):968–72, 2011.

[343] T. R. Mercer, D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell, J. A. Jeddeloh, J. S. Mattick, and J. L. Rinn. **Targeted RNA sequencing reveals the deep complexity of the human transcriptome**. *Nat Biotechnol*, **30**(1):99–104, 2012.

[344] S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, and T. R. Gingeras. **Landscape of transcription in human cells**. *Nature*, **489**(7414):101–8, 2012.

[345] M. J. Hangauer, I. W. Vaughn, and M. T. McManus. **Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs**. *PLoS Genet*, **9**(6):e1003569, 2013.

[346] V. Groh, A. Steinle, S. Bauer, and T. Spies. **Recognition of stress-induced MHC molecules by intestinal epithelial gammadelta T cells**. *Science*, **279**(5357):1737–40, 1998.

[347] E. F. Remmers, F. Cosan, Y. Kirino, M. J. Ombrello, N. Abaci, C. Satorius, J. M. Le, B. Yang, B. D. Korman, A. Cakiris, O. Aglar, Z. Emrence, H. Azakli, D. Ustek, I. Tugal-Tutkun, G. Akman-Demir, W. Chen, C. I. Amos, M. B. Dizon, A. A. Kose, G. Azizlerli, B. Erer, O. J. Brand, V. G. Kaklamani, P. Kaklamanis, E. Ben-Chetrit, M. Stanford, F. Fortune, M. Ghabra, W. E. Ollier, Y. H. Cho, D. Bang, J. O'Shea, G. R. Wallace, M. Gadina, D. L. Kastner, and A. Gül. **Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease**. *Nat Genet*, **42**(8):698–702, 2010.

[348] N. A. Patsopoulos, L. F. Barcellos, R. Q. Hintzen, C. Schaefer, C. M. van Duijn, J. A. Noble, T. Raj, P. A. Gourraud, B. E. Stranger, J. Oksenberg, T. Olsson, B. V. Taylor, S. Sawcer, D. A. Hafler, M. Carrington, P. L. De Jager, P. I. de Bakker, IMSGC, and ANZgene. **Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects**. *PLoS Genet*, **9**(11):e1003926, 2013.

[349] D. L. Morris, K. E. Taylor, M. M. Fernando, J. Nititham, M. E. Alarcón-Riquelme, L. F. Barcellos, T. W. Behrens, C. Cotsapas, P. M. Gaffney, R. R. Graham, B. A. Pons-Estel, P. K. Gregersen, J. B. Harley, S. L. Hauser, G. Hom, C. D. Langefeld, J. A. Noble, J. D. Rioux, M. F. Seldin, L. A. Criswell, T. J. Vyse, International MHC Network, Autoimmunity Genetics, and Systemic Lupus Erythematosus Genetics Consortium. **Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans**. *Am J Hum Genet*, **91**(5):778–93, 2012.

[350] J. Sidney, C. Oseroff, S. Southwood, M. Wall, G. Ishioka, F. Koning, and A. Sette. **DRB1*0301 molecules recognize a structural motif distinct from the one recognized by most DR beta 1 alleles**. *J Immunol*, **149**(8):2634–40, 1992.

[351] E. A. James, J. Bui, D. Berger, L. Huston, M. Roti, and W. W. Kwok. **Tetramer-guided epitope mapping reveals broad, individualized repertoires of tetanus toxin-specific CD4+ T cells and suggests HLA-based differences in epitope recognition**. *Int Immunol*, **19**(11):1291–301, 2007.

[352] CA Alper, MS Kruskall, D Marcus-Bagley, DE Craven, AJ Katz, SJ Brink, JL Dienstag, Z Awdeh, and EJ Yunis. **Genetic prediction of non-response to hepatitis B vaccine.** *N Engl J Med*, **321**(11):708–12, 1989.

[353] Y Li, R Ni, W Song, W Shao, S Shrestha, S Ahmad, CK Cunningham, PM Flynn, BG Kapogiannis, CM Wilson, and J Tang. **Clear and independent associations of several HLA-DRB1 alleles with differential antibody responses to hepatitis B vaccination in youth.** *Hum Genet*, **126**(5):685–96, 2009.

[354] L. Handunnetthi, S. V. Ramagopalan, G. C. Ebers, and J. C. Knight. **Regulation of major histocompatibility complex class II gene expression, genetic variation and disease**. *Genes Immun*, **11**(2):99–112, 2010.

[355] A. Finn, C. Blondeau, and F. Bell. ***Haemophilus influenzae* type B (Hib) antibody responses in children given diphtheria-tetanus-acellular pertussis-Hib combination vaccines**. *J Infect Dis*, **181**(6):2117–8, 2000.

[356] D. Goldblatt, P. Richmond, E. Millard, C. Thornton, and E. Miller. **The induction of immunologic memory after vaccination with *Haemophilus influenzae* type b conjugate and acellular pertussis-containing diphtheria, tetanus, and pertussis vaccine combination**. *J Infect Dis*, **180**(2):538–41, 1999.

[357] J. B. Wing, L. Smart, R. Borrow, J. Findlow, H. Findlow, A. W. Heath, and R. C. Read. **Kinetics of immune responses to nasal challenge with meningococcal polysaccharide one year after serogroup-C glycoconjugate vaccination.** *Clin Infect Dis*, **52**(11):1317–23, 2011.

[358] K. Mizuno, K. Hasegawa, T. Katagiri, M. Ogimoto, T. Ichikawa, and H. Yakura. **MPTP delta, a putative murine homolog of HPTP delta, is expressed in specialized regions of the brain and in the B-cell lineage**. *Mol Cell Biol*, **13**(9):5513–23, 1993.

[359] R. Dagan, J. Eskola, C. Leclerc, and O. Leroy. **Reduced response to multiple vaccines sharing common protein epitopes that are administered simultaneously to infants**. *Infect Immun*, **66**(5):2093–8, 1998.

[360] R. Borrow, R. Dagan, F. Zepp, H. Hallander, and J. Poolman. **Glycoconjugate vaccines and immune interactions, and implications for vaccination schedules**. *Expert Rev Vaccines*, **10**(11):1621–31, 2011.

[361] J. C. Tejedor, M. Moro, J. Ruiz-Contreras, J. Castro, J. A. Gómez-Campderá, M. L. Navarro, J. M. Merino, A. Martín-Ancel, J. Roca, M. García-del Río, A. Jurado, F. J. Díez-Delgado, F. Omeñaca, J. García-Sicilia, R. Boceta, P. García-Corbeira, J. M. Jacquet, A. Collard, L. Schuerman, and Spanish DTaP-HBV-IPV-097 Study Group. **Immunogenicity and reactogenicity of primary immunization with a hexavalent diphtheria-tetanus-acellular pertussis-hepatitis B-inactivated polio-*Haemophilus influenzae* type B vaccine coadministered with two doses of a meningococcal C-tetanus toxoid conjugate vaccine**. *Pediatr Infect Dis J*, **25**(8):713–20, 2006.

[362] M. T. Osterholm, N. S. Kelley, A. Sommer, and E. A. Belongia. **Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis**. *Lancet Infect Dis*, **12**(1):36–44, 2012.

[363] T. Jefferson, A. Rivetti, C. Di Pietrantonj, V. Demicheli, and E. Ferroni. **Vaccines for preventing influenza in healthy children**. *Cochrane Database Syst Rev*, **8**:CD004879, 2012.

[364] E. Miller, K. Hoschler, P. Hardelid, E. Stanford, N. Andrews, and M. Zambon. **Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study**. *Lancet*, **375**(9720):1100–8, 2010.

[365] PHE Public Health England. **Pandemic (H1N1) 2009 in England: an overview of initial epidemiological findings and implications for the second wave, www.hpa.org.uk/webc/hpawebfile/** , 2010.

[366] E. Bautista, T. Chotpitayasunondh, Z. Gao, S. A. Harper, M. Shaw, T. M. Uyeki, S. R. Zaki, F. G. Hayden, D. S. Hui, J. D. Kettner, A. Kumar, M. Lim, N. Shindo, C. Penn, K. G. Nicholson, and Writing Committee of the WHO Consultation on Clinical Aspects of Pandemic (H1N1) 2009 Influenza. **Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection**. *N Engl J Med*, **362**(18):1708–19, 2010.

[367] S. Cauchemez, C. A. Donnelly, C. Reed, A. C. Ghani, C. Fraser, C. K. Kent, L. Finelli, and N. M. Ferguson. **Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States**. *N Engl J Med*, **361**(27):2619–27, 2009.

[368] Finnish National Institute for Health and Welfare. **National Institute for Health and Welfare recommends discontinuation of Pandemrix vaccinations, www.thl.fi**, 2010.

[369] H. Nohynek, J. Jokinen, M. Partinen, O. Vaarala, T. Kirjavainen, J. Sundman, S. L. Himanen, C. Hublin, I. Julkunen, P. Olsén, O. Saarenpää-Heikkilä, and T. Kilpi. **AS03 adjuvanted AH1N1 vaccine associated with an abrupt increase in the incidence of childhood narcolepsy in Finland**. *PLoS One*, **7**(3):e33536, 2012.

[370] E. Miller, N. Andrews, L. Stellitano, J. Stowe, A. M. Winstone, J. Shneerson, and C. Verity. **Risk of narcolepsy in children and young people receiving AS03 adjuvanted pandemic A/H1N1 2009 influenza vaccine: retrospective analysis**. *BMJ*, **346**:f794, 2013.

[371] M. Partinen, O. Saarenpää-Heikkilä, I. Ilveskoski, C. Hublin, M. Linna, P. Olsén, P. Nokelainen, R. Alén, T. Wallden, M. Espo, H. Rusanen, J. Olme, H. Sätilä, H. Arikka, P. Kaipainen, I. Julkunen, and T. Kirjavainen. **Increased incidence and clinical picture of childhood narcolepsy following the 2009 H1N1 pandemic vaccination campaign in Finland**. *PLoS One*, **7**(3):e33723, 2012.

[372] P. de Whalley, W. Walker, M. D. Snape, C. Oeser, M. Casey, P. Moulsdale, C. Harrill, N. Andrews, K. Hoschler, B. Thompson, C. Jones, J. Chalk, S. Kerridge, R. Tomlinson, P. T. Heath, A. Finn, S. Faust, E. Miller, and A. J. Pollard. **A 1-year follow-on study from a randomised, head-to-head, multicentre, open-label study of two pandemic influenza vaccines in children**. *Health Technol Assess*, **15**(45):v–vi, xi–xiii, 1–128, 2011.

[373] T. W. Clark, M. Pareek, K. Hoschler, H. Dillon, K. G. Nicholson, N. Groth, and I. Stephenson. **Trial of 2009 influenza A (H1N1) monovalent MF59-adjuvanted vaccine**. *N Engl J Med*, **361**(25):2424–35, 2009.

[374] T. Rowe, R. A. Abernathy, J. Hu-Primmer, W. W. Thompson, X. Lu, W. Lim, K. Fukuda, N. J. Cox, and J. M. Katz. **Detection of antibody to avian influenza A (H5N1) virus in human serum by using a combination of serologic assays**. *J Clin Microbiol*, **37**(4):937–43, 1999.

[375] T. Lambe, A. J. Spencer, C. E. Mullarkey, R. D. Antrobus, L. M. Yu, P. de Whalley, B. A. Thompson, C. Jones, J. Chalk, S. Kerridge, A. V. Hill, M. D. Snape, A. J. Pollard, and S. C. Gilbert. **T-Cell Responses in Children to Internal Influenza Antigens, 1 Year after Immunization with Pandemic H1N1 Influenza Vaccine, and Response to Revaccination with Seasonal Trivalent Inactivated Influenza Vaccine.** *Pediatr Infect Dis J*, 2012.

[376] R. Schmid, P. Baum, C. Ittrich, K. Fundel-Clemens, W. Huber, B. Brors, R. Eils, A. Weith, D. Mennerich, and K. Quast. **Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3.** *BMC Genomics*, **11**:349, 2010.

[377] J. D. Storey and R. Tibshirani. **Statistical significance for genomewide studies**. *Proc Natl Acad Sci U S A*, **100**(16):9440–5, 2003.

[378] P. Liu and J. T. Hwang. **Quick calculation for sample size while controlling false discovery rate with application to microarray analysis**. *Bioinformatics*, **23**(6):739–46, 2007.

[379] Megan Orr and Peng Liu. *ssize.fdr: Sample Size Calculations for Microarray Experiments*, 2012. R package version 1.1.

[380] Megan Orr and Peng Liu. **Sample Size Estimation while Controlling False Discovery Rate for Microarray Experiments Using the ssize.fdr Package**, 2009.

[381] C. Wei, J. Li, and R. E. Bumgarner. **Sample size for detecting differentially expressed genes in microarray experiments**. *BMC Genomics*, **5**:87, 2004.

[382] Gordon K Smyth. **Limma: linear models for microarray data**. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pages 397–420. Springer, New York, 2005.

[383] V. G. Tusher, R. Tibshirani, and G. Chu. **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc Natl Acad Sci U S A*, **98**(9):5116–21, 2001.

[384] Holger Schwender. *siggenes: Multiple testing using SAM and Efron's empirical Bayes approaches*, 2012. R package version 1.36.0.

[385] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A*, **102**(43):15545–50, 2005.

[386] V. K. Mootha, C. M. Lindgren, K. F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, and L. C. Groop. **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes**. *Nat Genet*, **34**(3):267–73, 2003.

[387] A. Krämer, J. Green, J. Pollard, and S. Tugendreich. **Causal analysis approaches in Ingenuity Pathway Analysis**. *Bioinformatics*, **30**(4):523–30, 2014.

[388] W. N. Venables, Brian D. Ripley, and W. N. Modern applied statistics with S. Plus Venables. *Modern applied statistics with S*. Springer, New York, 4th ed. edition, 2002.

[389] Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, and Bill Venables. *gplots: Various R programming tools for plotting data*, 2013. R package version 2.12.1.

[390] S. Sasaki, M. C. Jaimes, T. H. Holmes, C. L. Dekker, K. Mahmood, G. W. Kemble, A. M. Arvin, and H. B. Greenberg. **Comparison of the influenza virus-specific effector and memory B-cell responses to immunization of children and adults with live attenuated or inactivated influenza virus vaccines**. *J Virol*, **81**(1):215–28, 2007.

[391] R. J. Feezor, H. V. Baker, M. Mindrinos, D. Hayden, C. L. Tannahill, B. H. Brownstein, A. Fay, S. MacMillan, J. Laramie, W. Xiao, L. L. Moldawer, J. P. Cobb, K. Laudanski, C. L. Miller-Graziano, R. V. Maier, D. Schoenfeld, R. W. Davis, R. G. Tompkins, Inflammation, and L.rge-Scale Collaborative Research Program Host Response to Injury. **Whole blood and leukocyte RNA isolation for gene expression analyses**. *Physiol Genomics*, **19**(3):247–54, 2004.

[392] J. Liu, E. Walter, D. Stenger, and D. Thach. **Effects of globin mRNA reduction methods on gene expression profiles from whole blood**. *J Mol Diagn*, **8**(5):551–8, 2006.

[393] K. Vartanian, R. Slottke, T. Johnstone, A. Casale, S. R. Planck, D. Choi, J. R. Smith, J. T. Rosenbaum, and C. A. Harrington. **Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis**. *BMC Genomics*, **10**:2, 2009.

[394] N. Dhiman, I. G. Ovsyannikova, A. L. Oberg, D. E. Grill, R. M. Jacobson, and G. A. Poland. **Immune activation at effector and gene expression levels after measles vaccination in healthy individuals: a pilot study**. *Hum Immunol*, **66**(11):1125–36, 2005.

[395] Y. Liu and W. Zhang. **Identification of a new transmembrane adaptor protein that constitutively binds Grb2 in B cells**. *J Leukoc Biol*, **84**(3):842–51, 2008.

[396] E. J. Lowenstein, R. J. Daly, A. G. Batzer, W. Li, B. Margolis, R. Lammers, A. Ullrich, E. Y. Skolnik, D. Bar-Sagi, and J. Schlessinger. **The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling**. *Cell*, **70**(3):431–42, 1992.

[397] W. Zhang, J. Sloan-Lancaster, J. Kitchen, R. P. Trible, and L. E. Samelson. **LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation**. *Cell*, **92**(1):83–92, 1998.

[398] L. He and G. J. Hannon. **MicroRNAs: small RNAs with a big role in gene regulation.** *Nat Rev Genet*, **5**(7):522–31, 2004.

[399] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, and T. Tuschl. **Identification of novel genes coding for small expressed RNAs**. *Science*, **294**(5543):853–8, 2001.

[400] M. Ballarino, F. Pagano, E. Girardi, M. Morlando, D. Cacchiarelli, M. Marchioni, N. J. Proudfoot, and I. Bozzoni. **Coupled RNA processing and transcription of intergenic primary microRNAs**. *Mol Cell Biol*, **29**(20):5632–8, 2009.

[401] R. M. O'Connell, D. S. Rao, A. A. Chaudhuri, and D. Baltimore. **Physiological and pathological roles for microRNAs in the immune system**. *Nat Rev Immunol*, **10**(2):111–22, 2010.

[402] L. A. O'Neill, F. J. Sheedy, and C. E. McCoy. **MicroRNAs: the fine-tuners of Toll-like receptor signalling**. *Nat Rev Immunol*, **11**(3):163–75, 2011.

[403] R. C. Friedman, K. K. Farh, C. B. Burge, and D. P. Bartel. **Most mammalian mRNAs are conserved targets of microRNAs**. *Genome Res*, **19**(1):92–105, 2009.

[404] R. M. O'Connell, K. D. Taganov, M. P. Boldin, G. Cheng, and D. Baltimore. **MicroRNA-155 is induced during the macrophage inflammatory response**. *Proc Natl Acad Sci U S A*, **104**(5):1604–9, 2007.

[405] K. D. Taganov, M. P. Boldin, K. J. Chang, and D. Baltimore. **NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses**. *Proc Natl Acad Sci U S A*, **103**(33):12481–6, 2006.

[406] S. Gilad, E. Meiri, Y. Yogev, S. Benjamin, D. Lebanony, N. Yerushalmi, H. Benjamin, M. Kushnir, H. Cholakh, N. Melamed, Z. Bentwich, M. Hod,

Y. Goren, and A. Chajut. **Serum microRNAs are promising novel biomarkers**. *PLoS One*, **3**(9):e3148, 2008.

[407] X. Chen, H. Liang, J. Zhang, K. Zen, and C. Y. Zhang. **Secreted microRNAs: a new form of intercellular communication**. *Trends Cell Biol*, **22**(3):125–32, 2012.

[408] H. Valadi, K. Ekström, A. Bossios, M. Sjöstrand, J. J. Lee, and J. O. Lötvall. **Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells**. *Nat Cell Biol*, **9**(6):654–9, 2007.

[409] N. Kosaka, H. Iguchi, Y. Yoshioka, F. Takeshita, Y. Matsuki, and T. Ochiya. **Secretory mechanisms and intercellular transfer of microRNAs in living cells**. *J Biol Chem*, **285**(23):17442–52, 2010.

[410] K. C. Vickers, B. T. Palmisano, B. M. Shoucri, R. D. Shamburek, and A. T. Remaley. **MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins**. *Nat Cell Biol*, **13**(4):423–33, 2011.

[411] Y. Xiong, S. Chen, L. Liu, Y. Zhao, W. Lin, and J. Ni. **Increased serum microRNA-155 level associated with nonresponsiveness to hepatitis B vaccine**. *Clin Vaccine Immunol*, **20**(7):1089–91, 2013.

[412] D. Sarkar, R. Parkin, S. Wyman, A. Bendoraite, C. Sather, J. Delrow, A. K. Godwin, C. Drescher, W. Huber, R. Gentleman, and M. Tewari. **Quality assessment and data analysis for microRNA expression arrays**. *Nucleic Acids Res*, **37**(2):e17, 2009.

[413] G. N. Brock, P. Mukhopadhyay, V. Pihur, C. Webb, R. M. Greene, and M. M. Pisano. **MmPalateMiRNA, an R package compendium illustrating analysis of miRNA microarray data**. *Source Code Biol Med*, **8**(1):1, 2013.

[414] K. Wang, Y. Yuan, J. H. Cho, S. McClarty, D. Baxter, and D. J. Galas. **Comparing the MicroRNA spectrum between serum and plasma**. *PLoS One*, **7**(7):e41561, 2012.

[415] P. S. Mitchell, R. K. Parkin, E. M. Kroh, B. R. Fritz, S. K. Wyman, E. L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K. C. O'Briant, A. Allen, D. W. Lin, N. Urban, C. W. Drescher, B. S. Knudsen, D. L. Stirewalt, R. Gentleman, R. L. Vessella, P. S. Nelson, D. B. Martin, and M. Tewari. **Circulating microRNAs as stable blood-based markers for cancer detection**. *Proc Natl Acad Sci U S A*, **105**(30):10513–8, 2008.

[416] Y. Dai, W. Sui, H. Lan, Q. Yan, H. Huang, and Y. Huang. **Comprehensive analysis of microRNA expression patterns in renal biopsies of lupus nephritis patients**. *Rheumatol Int*, **29**(7):749–54, 2009.

[417] M. Jernås, I. Nookaew, H. Wadenvik, and B. Olsson. **MicroRNA regulate immunological pathways in T-cells in immune thrombocytopenia (ITP)**. *Blood*, **121**(11):2095–8, 2013.

[418] A. Diercks and A. Aderem. **Systems approaches to dissecting immunity**. *Curr Top Microbiol Immunol*, **363**:1–19, 2013.

[419] B. Pulendran. **Systems vaccinology: probing humanity's diverse immune systems with vaccines**. *Proc Natl Acad Sci U S A*, **111**(34):12300–6, 2014.

[420] M. Nishimura and S. Naito. **Tissue-specific mRNA expression profiles of human toll-like receptors and related genes.** *Biol Pharm Bull*, **28**(5):886–92, 2005.

[421] A. S. Lihn, B. Richelsen, S. B. Pedersen, S. B. Haugaard, G. S. Rathje, S. Madsbad, and O. Andersen. **Increased expression of TNF-alpha, IL-6, and IL-8 in HALS: implications for reduced adiponectin expression and plasma levels.** *Am J Physiol Endocrinol Metab*, **285**(5):E1072–80, 2003.

[422] M. P. Berry, C. M. Graham, F. W. McNab, Z. Xu, S. A. Bloch, T. Oni, K. A. Wilkinson, R. Banchereau, J. Skinner, R. J. Wilkinson, C. Quinn, D. Blankenship, R. Dhawan, J. J. Cush, A. Mejias, O. Ramilo, O. M. Kon, V. Pascual, J. Banchereau, D. Chaussabel, and A. O'Garra. **An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis**. *Nature*, **466**(7309):973–7, 2010.

[423] Y. Chen, E. R. Dougherty, and M. L. Bittner. **Ratio-based decisions and the quantitative analysis of cDNA microarray images**. *J Biomed Opt*, **2**(4):364–74, 1997.