ONTOGENY AND GENETIC CORRELATES OF THE TLR MEDIATED PEDIATRIC INNATE IMMUNE RESPONSE

by

Nathan Patrick Corbett

B.Sc. Biology, Acadia University, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2013

© Nathan Patrick Corbett, 2013
Abstract

In early life, humans are particularly vulnerable to morbidity and mortality due to infectious disease. A key system that is critical to both early response to pathogens, and also to the success or failure of a vaccine to induce protective immunity is the innate immune system. It is our working hypothesis that changes in the developing immune system mediate changes in both vaccine response and infectious morbidity and mortality.

This thesis presents published and unpublished work wherein we analyze the innate immune response of a defined population of newborns from the greater Vancouver area in British Columbia, Canada. In this work, we set out to define the development of early response by the human infant immune system to molecular danger signals known as pathogen-associated molecular patterns (PAMPS) by the well-defined Toll-Like Receptor (TLR) system expressed by peripheral blood mononuclear cells (PBMC). In addition, we have correlated this response with the occurrence of pertinent genetic variance between individuals, in the hope of identifying immune modulating variants in situ. Such variants will provide the basis for later testing of our hypothesis that genetic variance in early life innate immune response contributes to the significant variability in morbidity and mortality.
Preface

Portions of Chapters 2 and 3 have been published previously in: Corbett, N. P., Blimkie, D., Ho, K. C., Cai, B., Sutherland, D. P., Kallos, A., et al. (2010). Ontogeny of Toll-like receptor mediated cytokine responses of human blood mononuclear cells. PloS one, 5(11), e15041. doi: 10.1371/journal.pone.0015041. Nathan conducted the majority of testing and all statistical analysis for this work, and co-authored the manuscript with Dr. Tobias Kollmann. Dr. Edgardo Fortuno provided help revising latter drafts, and Kevin Ho was instrumental in aiding visual analysis and presentation of data.

Publications arising from this work include:


"Prospective Study of Innate Immunity from Birth to Two Years of Age" was approved by the UBC C&W Research Ethics Board (certificate number H06-03339).
Table of Contents

Abstract .................................................................................................................................... ii
Preface ..................................................................................................................................... iii
Table of Contents ................................................................................................................... iv
List of Tables ........................................................................................................................ viii
List of Figures ......................................................................................................................... ix
List of Abbreviations ............................................................................................................ xii
Acknowledgements .............................................................................................................. xvi
Dedication ............................................................................................................................ xvii

Chapter 1: Introduction ........................................................................................................ 1

1.1 Cellular innate immunity: Not a servant, but a collaborator ......................................... 3
1.2 TLR structure, function, and specificity ........................................................................ 5
   1.2.1 TLR signaling ....................................................................................................... 9
1.3 Innate immunity: Critical in early life ......................................................................... 12
   1.3.1 Relevance to allergy and autoimmunity ............................................................. 13
1.3.2 Explanatory paradigms ....................................................................................... 14
   1.3.2.1 Innate skewing of neonatal vaccination ...................................................... 15
1.4 Genetic correlates of innate immune response .............................................................. 17

Chapter 2: Body of Thesis ................................................................................................... 19

2.1 Methods ........................................................................................................................ 19
   2.1.1 TLR stimulation .................................................................................................. 19
   2.1.2 Birth cohort study design ................................................................................... 20
2.1.3 Blood sample processing and in vitro stimulation ....................................................... 21
2.1.4 Assessment of cytokines in culture supernatant ..................................................... 22
2.1.5 Global analysis of secreted cytokine production in response to TLR stimulation of neonatal, 1- and 2-year old, and adult peripheral blood monocytic cells ................... 23
2.1.6 Affymetrix GoldenGate SNP array design ............................................................... 24
2.1.7 Data preparation ..................................................................................................... 25
2.1.8 Pilot cohort data analysis ....................................................................................... 26
2.1.9 Full cohort SNP correlation analysis ...................................................................... 27
  2.1.9.1 Exploratory nonparametric SNP analysis ....................................................... 27
  2.1.9.2 GenABEL based SNP association analysis .................................................. 28
2.2 Results ....................................................................................................................... 32
  2.2.1 Luminex analysis of cytokine expression in pilot longitudinal cohort ............... 32
    2.2.1.1 TLR7/8 and TLR9-driven antiviral IFN-α2 responses reached adult levels by 1 year of life ........................................................................................................... 34
    2.2.1.2 Th-1 supporting IL-12p70 production increased but remained below adult levels even at 2 years of age ................................................................................. 34
    2.2.1.3 Th-1-supporting IFN-γ production always remained below adult levels, and in fact was found to decrease from birth up to age 2 years ................................. 35
    2.2.1.4 Th17-supporting IL-12p40, IL-23 and IL-6 cytokine production declined from birth over the first 2 years of life ................................................................. 35
    2.2.1.5 TNF-α and IL-1β both increased over the first 2 years of life for TLR2/1 as well as TLR4 responses, but not for TLR7/8 responses ............................................. 36
2.2.1.6 Anti-inflammatory IL-10 production declined from birth onwards to at, or below, adult levels ................................................................................................................... 37

2.2.2 Preliminary genetic correlates of innate immune response at birth .................. 37

2.2.2.1 TLR-1 .......................................................................................................... 38

2.2.2.2 IL6, PAM, rs4833095 (TLR-1 coding SNP, nonsynonymous) .................. 40

2.2.3 Genetic correlates of innate immune response at birth .................................. 44

2.2.3.1 SNP rs10751209 near the FADD gene and TNFα response to 3M-003 .... 46

2.2.3.2 SNP rs2069830 (coding, nonsynonymous) in the IL-6 gene and IL-1β response to CpG-A.............................................................................................................. 48

2.2.3.3 SNP rs583911 in the IL-12A gene and IL-6 response to pl:C .................... 50

Chapter 3: Conclusion ......................................................................................................... 52

3.1 Bulk cytokine measurement of longitudinal cohort................................................. 52

3.2 Pediatric immunogenetics ..................................................................................... 59

3.2.1 Introduction ..................................................................................................... 59

3.2.2 Kruskal-Wallis significance testing and rs4833095 ........................................... 60

3.2.3 GenABEL association testing identifies rs10751209 & rs2069830 as significant predictors of neonatal cytokine response to 3M-003 and CpG-A, respectively .... 62

3.3 Suggested future research directions .................................................................... 68

3.4 Strengths and weaknesses of this project ............................................................ 69

3.5 Conclusion ............................................................................................................ 70

References......................................................................................................................... 72

Appendices......................................................................................................................... 88

Appendix A....................................................................................................................... 88
A.1  Statistical testing for significant cytokine response over control. ...................... 88
A.2  R program to perform statistical analysis of NIH01 cohort year 0 time point. .. 89
A.3  R program to plot statistically significant results and perform power

calculations. .................................................................................................................. 100

Appendix B ............................................................................................................... 113
  B.1  Statistical analysis of cytokine response cohort.............................................. 113

Appendix C ............................................................................................................... 115
  C.1  GVS haplotype testing rs10751209............................................................... 115
  C.2  GVS haplotype testing rs2069830 ................................................................. 116
  C.3  GVS haplotype testing rs583911 ................................................................. 116
List of Tables

Table 1. Kruskal-Wallis testing for significant difference between unstimulated and stimulated samples. ................................................................................................................. 88

Table 2. Comparison of the concentrations of secreted cytokines for longitudinal neonatal and adult responses to cell-surface TLR stimulation. Cytokines secreted MC in tissue culture media containing the indicated TLR ligands were determined in pg/ml for 35 neonate-infant subjects and 25 adult subjects. Statistically significant (p<0.01) differences between age groups are indicated in bold type-set. NP = not performed. ......................................................... 113

Table 3. Comparison of the concentrations of secreted cytokines for longitudinal neonatal and adult responses to endosomal TLR stimulation. Cytokines secreted MC in tissue culture media containing the indicated TLR ligands were determined in pg/ml for 35 neonate-infant subjects and 25 adult subjects. Statistically significant (p<0.01) differences between age groups are indicated in bold type-set. NP = not performed. ......................................................... 114

Table 4. Pairwise haplotype LD testing for 10Kb up and downstream of rs10751209, filtered for target SNP. Relevant color code: Orange = UTR SNP ................................................................. 115

Table 5. Pairwise haplotype LD testing for 10Kb up and downstream of rs583911, filtered for target SNP. Relevant color code: Orange = UTR SNP ................................................................. 116
List of Figures

Figure 1. Bulk measurement of secreted cytokine production in response to TLR stimulation of MC from neonatal, 1- and 2-year olds, and adult. .............................................................. 32

Figure 2. rs2069830 Minor allele associates with increased IL-6 response to PAM. Differences between (top strand) common homozygous AA (0), heterozygous AG (1), and rare homozygous GG (2) neonates (cord blood in vitro) in PBMC IL-6 cytokine response to PAM3CSK4 TLR2 ligand, as represented by median/interquartile boxplot with a per-subject beeswarm plot overlayed for SNP rs4833095. ....................................................................... 40

Figure 3. Quartile-Quartile plot of genotype groups as defined by the SNP rs4833095 within the TLR1 gene. Here the normal distribution was compared to the quartiles of cytokine response distributions for each group. The theoretical ‘normal’ approximation of the given data is represented by a line derived from the linear regression, with 95% confidence interval denoted by dash-lines. ............................................................................................................. 41

Figure 4. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs 4833095 within the TLR1 gene. Here the exponential distribution was compared to the quartiles of cytokine response distributions for each group. The theoretical ‘normal’ approximation of the given data is represented by a linear regression line, with 95% confidence interval denoted by dash-lines. ............................................................................................................. 42

Figure 5. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs 4833095 within the TLR1 gene. Here the normal distribution was compared to the quartiles of log10-transformed cytokine response distributions for each group. The theoretical ‘normal’
approximation of the given data is represented by a linear regression line, with 95% confidence interval denoted by dash-lines. 43

Figure 6. Graphical representation of principal component analysis of whole-array (minus AIMs) SNP variance. 44

Figure 7 rs10751209 minor allele associates with decreased TNFα response to 3M-003. 46

Figure 8. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs10751209 near the FADD gene. Here the exponential distribution was compared to the quartiles of log2-transformed cytokine response distributions for each group. The theoretical ‘exponential’ approximation of the given data is represented by a linear regression line, with 95% confidence interval denoted by dash-lines. 47

Figure 9. rs2069830 minor allele associates with nonexistent IL-1B response to CpG-A. Differences between common homozygous GG, heterozygous AG (cord blood in vitro) in PBMC IL-1β cytokine response to CpG-A TLR9 ligand, as represented by median/interquartile boxplot with a per-subject beeswarm plot overlaid for SNP rs2069830 which has been used to flag variation in the IL-6 gene. 48

Figure 10. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs2069830 in the IL-6 gene. Here the exponential distribution was compared to the quartiles of log2-transformed cytokine response distributions for each genotype group. 49

Figure 11. rs583911 minor allele associates with decreased IL-6 response to poly I:C. Differences between common homozygous AA, heterozygous AG, and homozygous GG genotype groups (cord blood in vitro) in PBMC IL-6 cytokine response to poly I:C TLR3 ligand, as represented by median/interquartile boxplot with a per-subject beeswarm plot overlaid for SNP rs583911, which has been used to flag variation in the IL-12A gene. 50
Figure 12. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs583911 in the IL-12A gene. Here the exponential distribution was compared to the quartiles of log$_2$-transformed cytokine response distributions for each genotype group.

Figure 13. General descriptive trends of cytokine response in early life (Corbett et al. 2010).
List of Abbreviations

Alum — Aluminum hydroxide salts
TB — Tuberculosis
HIV — Human Immunodeficiency Virus
T cell — Thymus derived lymphocyte
B cell — Bone marrow derived lymphocyte
TNF — tumor necrosis factor
LPS — lipopolysaccharide
TLR — Toll-Like Receptor
RAG — Recombination Activating Gene
mya — Millions of years ago
PRR — Pattern Recognition Receptor
PAMP — Pathogen Associated Molecular Pattern
TIR — Toll/interleukin-1 receptor (homology domain)
LRR — Leucine-rich repeat
MD2 — Lymphocyte antigen 96
CpG — 2’-deoxyribocytidine-phosphate-guanosine
PamCysPamSK₄ — N-palmitoyl-S-[2-hydroxy-3-(palmitoyloxy)propyl]-(R)-cysteinyln(lysyl)₃-lysine
Pam₃CSK₄ (PAM) — N-palmitoyl-S-[2,3-bis-(palmitoyloxy)-propyl]-(R)-cysteinyl(lysyl)₃-lysine
LDL — Low-Density Lipoprotein
CD36 — Cluster of Differentiation 36
ER — Endoplasmic Reticulum
APC — Antigen Presenting Cell
DC — Dendritic Cell
cDC — classical Dendritic Cell
pDC — plasmacytoid Dendritic Cell
MyD88 — Myeloid differentiation primary response gene (88)
TRIF — TIR-domain-containing adapter-inducing interferon-β
NF-κβ — Nuclear Factor Kappa-light-chain-enhancer of activated B cells
MAPK — Mitogen-Activated Protein Kinase(es)
IRF — Interferon Regulatory Factor
TRIF — TIR domain-containing adaptor protein-inducing IFN-β
TRAM — TRIF-Related Adaptor Molecule
ICS — Intracellular Cytokine Staining
IFN — Interferon
IL — Interleukin
IRAK — interleukin-1 receptor-associated kinase
Ig — Immunoglobulin
TH — T-helper cell (or immune response type)
Cas — Cas-Br-E murine leukemia virus
PFU — Plaque forming units
BFA — Brefeldin A
MFI — Mean fluorescent intensity
pg — picogram
mL — milliliter
SEM — Standard Error of Mean
HWE — Hardy-Weinberg Equilibrium
GB — Gigabyte
RAM — Random Access Memory
I/O — Input/Output (computer science)
AIM — Ancestry Informative Marker
MAF — Minor Allele Frequency
QC — Quality Control
HET — Heterozygosity
FDR — False-Discovery Rate
IBS — Identity By State
ANOVA — Analysis Of Variance
GVS — Genome Variation Server (Washington University)
ND — Not Done
Poly I:C (or pI:C) — Polyinosinic:polycytidylic acid
3M-003 — A synthetic TLR7+8 agonist manufactured by the 3M corporation.
MC — Monocyte
Q-Q plot — Quartile to Quartile comparison plot
KB — Kilobase (genetics)
kb — Kilobyte (computer science)
SD — Standard Deviation
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>G</td>
<td>Guanosine</td>
</tr>
<tr>
<td>C</td>
<td>Cytidine</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage Domain</td>
</tr>
<tr>
<td>HCB</td>
<td>Han-Chinese Hap-Map cohort code</td>
</tr>
<tr>
<td>CEU</td>
<td>Caucasian Hap-Map cohort code</td>
</tr>
<tr>
<td>WB</td>
<td>Whole Blood</td>
</tr>
<tr>
<td>CHILD study</td>
<td>The Canadian Healthy Infant Longitudinal Development study</td>
</tr>
<tr>
<td>GPU</td>
<td>Graphics Processing Unit</td>
</tr>
<tr>
<td>TDT</td>
<td>Transmission Disequilibrium Test</td>
</tr>
</tbody>
</table>
Acknowledgements

I owe a great debt of gratitude to my peers and mentors, who span diverse nationalities and
disciplines of the arts, sciences, and the medical profession. Specifically, Dr. Tobias
Kollmann has been my mentor, friend, and role model. Without his boundless energy,
brilliant mind, and phenomenal compassion, this work would not exist. Dr. Edgardo Fortuno
III has been a daily source of scientific debate and a great deal of laboratory training, as well
as a generous friend. Dr. Scott Tebbutt was instrumental in the creation of the genetic
correlation component of this thesis, and was of tremendous help in navigating what for me
was entirely undiscovered country. Drs. Stuart Turvey and Pascal Lavoie contributed superb
criticism and ideas as members of my supervisory committee. Mr. Darren Blimkie, Ms.
Juliet Crabtree, and Ms. Annie Rein-Weston were instrumental in the preparation of the early
phases of this study, and trained me in many of the pertinent lab techniques. Mr. Kevin Ho
was a tremendous support in rationalizing and indexing the huge quantity of frozen blood
samples that will constitute the basis of many future works on this cohort study. Kevin also
worked closely with me as a superb visual analytics expert, and I am tremendously indebted
to his remarkable skill and tenacity. Bing Cai, Darren P Sutherland, Elaine Fernandez,
Francis Thomai, Kinga Smolen, and Arlene Kallos also provided substantial support in the
smooth operation and acquisition of this project.

Finally, I would like to thank my family, and friends, who have been steadfast supporters of
my quixotic nature, and have tolerated surprisingly well my obsession with medical science.
Dedication

I dedicate this work to the millions of children who die each year from vaccine preventable disease, may we save them all one day in the not so distant future.
Chapter 1: Introduction

While the medical and scientific communities have had great success in effecting public health improvements through trial-by-error empirical vaccine research, we have at this point only managed to deal successfully with the so-called “low hanging fruit” pathogens for which simple live attenuated or killed organism vaccines adjuvanted with aluminum hydroxide (alum) have proven successful. Such approaches do not prevent more difficult targets like malaria, tuberculosis (TB), and human immunodeficiency virus (HIV), which are major drivers of human misery, loss of life, widespread social problems, and social inequality within and between regions (Collison, Dey, Hannah, & Stevenson, 2007; Jimenez, 2001; Medzhitov, Preston-Hurlburt, & Janeway, 1997). There is no question that these diseases cause tremendous mortality and untold suffering (Lawn, Cousens, Darmstadt, Paul, & Martines, 2004; Olusanya, 2005). Whole vaccines—vaccines that include a whole killed or otherwise attenuated pathogen rather than only a constituent part—can fail because the very evolutionary pressures that have crafted these organisms toward pathogenicity have endowed them with characteristics that preclude or prevent either the development of immunological memory, or protective immune response (Horn & McCulloch, 2010; Kwiatkowski, 2005). What is more, the host immune system will be bound by the constraints of its own evolved adaptations (Wlasiuk & Nachman, 2010). In such a heavily co-evolutionary system, both the pathogen and the host’s defenses will have established interactive modes at any given point in time along the developmental path that are defined by the various constraints with which the immune system must contend at that particular developmental point (Reviewed in: Goldszmid & Trinchieri, 2012). In certain cases, this means that a pathogen may exploit an
immunological weakness that in other contexts serves as an advantageous adaptation, or which presages one.

This is to say that there is no “better” immune system; rather, there are certain incumbent realities, which immunologists and vaccinologists must understand in order to intervene in the right way at the right time (Dietert & Piepenbrink, 2008). Given this complicated context, the basic immunology and molecular biology of the immune response is critical for the improvement of global health. Because the early source of vaccines—empirical trial and error—has all but dried up, we must understand both our pathogens and our immunological and ontological (i.e. every developmental stage) selves to develop vaccines that target the right antigens and are tailored for the right developmental time-points. This brave new world of rational vaccinology (Reviewed in: Buonaguro & Pulendran, 2011) will not come easily, and will require the move to a systematic approach. In other words, rational vaccinology requires the integration of fields ranging from bioinformatics to complex systems theory with the field of experimental immunology. This work stands at only the nearest edge of this complexity, and represents an attempt to add descriptive fuel to the analytic revolution. Laboratories are now being built with adjoining space for servers and workstations for this very purpose, indicating that we stand at the cusp of a paradigm shift in-progress for the means and methods of biology in general, and biomedical science in particular.

The primary goal of the largely observational work presented in this thesis is twofold: First, to identify the trends of cytokine response to defined targets of the innate immune response in the very young such that we may contribute to an ongoing discourse about the
fundamental character and mechanisms of the early immune response. Second, to suggest genetic correlates of these cytokine response, which could give some insight into the bimolecular basis of the remarkable variability in individual in vitro cytokine response to defined TLR ligands (Corbett et al., 2010; Kollmann et al., 2009).

1.1 Cellular innate immunity: Not a servant, but a collaborator

Elie Metchnikoff’s seminal presentation to the Pasteur institute in 1891 provided the first conclusive evidence that phagocytes were part of the body’s protective mechanisms and not pathology (Metchnikoff, 1893). This work began inquiry into the most fundamental aspects of basic immunology research that we continue here. This advent came more than a hundred years following the development of the smallpox vaccine by Dr. Edward Jenner that kicked-off the now well-established and broadly acknowledged paradigm of the adaptive immune system. This work, in turn, led to the discovery that immunity to disease is mediated by pathogen targeting antibodies (Reviewed in: Llewelyn, Hawkins, & Russell, 1992), which are the proteinaecious products of a cellular system comprised of T and B cells (Gowans, McGregor, Cowen, & Ford, 1962; Miller & Mitchell, 1968; Mitchell & Miller, 1968).

For much of the history of immunology, phagocytic cells were thought to be simple servants of the adaptive humoral immunity, which served by gobbling up antibody bound (or “opsinized”, from the Greek opsono for ‘I prepare victuals for’) pathogens (Wright & Douglas, 1903). However, Metchnikoff’s phagocytes may be having their own day in the sun now, as evidenced by the recent Nobel prize in physiology & medicine awarded to Ralph
Steinmann, Jules Hoffmann, and Bruce Beutler (Wagner, 2012) for their foundational work in innate immunity research. In this work, Steinmann, Hoffmann and Beutler discovered the dendritic cell (Steinman & Cohn, 1973; 1974; Steinman, Lustig, & Cohn, 1974), the pathological role of tumor necrosis factor (TNF) in septic response to lipopolysaccharide (LPS), as well as how to mitigate it (Beutler, Greenwald, Hulmes, & Chang, 1985a; Beutler, Milsark, & Cerami, 1985b; Peppel, Crawford, & Beutler, 1991). They also discovered the protective role of the genetic locus of the Toll homolog in drosophila for protection against fungal infection (Lemaitre, Nicolas, Michaut, Reichhart, & Hoffmann, 1996). Because the Nobel Prize can not be awarded posthumously, Charles Janeway is conspicuously absent from this narrative, though he cloned the first human toll-like receptor (TLR-4) along with Ruslan Medzhitov (Medzhitov et al., 1997), and was a powerfully productive and inspirational figure in innate as well as adaptive immunology.

It is from within the understanding that the adaptive immune response evolved out of that of innate immunity (J. A. Hoffmann, Kafatos, & Janeway, 1999; Litman & Cooper, 2007), that one may easily comprehend that adaptive immune response takes its cue and instruction from that of the innate immune system (Jones, 2005; Nemazee, Gavin, Hoebe, & Beutler, 2006) (Reviewed in: Iwasaki & Medzhitov, 2004). This understanding levels the hierarchy of the immune system, and establishes command and control-like roles for the humble phagocyte, especially the dendritic cell, as we will explore.

The innate immune response has a character, as we will briefly explain in section 1.3.2, and the phagocyte defines this character. Therefore, we theorize that the phagocytic cell forms a
major component of differences in immune response, and thus, vaccination response, that we see between individuals and between groups of subjects as different stages of human development. Given that the phagocyte delivers more information than simply antigen to the adaptive immune system (Reviewed in: Iwasaki & Medzhitov, 2010), and that signal is primarily through secreted peptides known as cytokines, we have chosen to study secreted cytokine response to defined stimuli as a way to understand the development of TLR response in early life (Corbett et al., 2010; Kollmann et al., 2009). We take this approach because the TLR system is a primary way by which phagocytic cells collect information about the environment, which has been shown to play a key role in immune development.

1.2 TLR structure, function, and specificity

Much of the published literature examining innate immunity has focused on pattern recognition receptors (PRR), such as the TLR, which are ancient evolutionarily-conserved cell surface and endosomal receptors involved in the recognition of conserved microbial pathogen-associated molecular patterns (PAMP) (Turvey & Hawn, 2006). Of course, there are other PAMP recognition systems active in human immunology and they interact with the TLR system via both direct and signaling pathway-mediated crosstalk (Reviewed in: Kawai & Akira, 2011). However, we have chosen to focus exclusively on the TLR system here, as the TLR system plays a critical role in early-life survival, as has been evidenced by remarkably high infant mortality suffered by individuals lacking TLR signaling (Ku et al., 2007; Picard et al., 2010). Recognition of a specific PAMP by a specific TLR triggers events
that result in the initiation of antigen uptake, processing, and presentation, as well as in the expression of co-stimulatory molecules and secretion of cytokine mediators (Reviewed in: Kumar, Kawai, & Akira, 2011), as we will explain in detail.

The TOLL-receptor was first discovered via forward genetics screening in drosophila due to its role in dorsoventral patterning during drosophila embryonic development (Anderson, Jürgens, & Nüsslein-Volhard, 1985; Hashimoto, Hudson, & Anderson, 1988). Later this gene was found to have immunological function through loss of protective effects versus fungal infection in flies with specific mutations in TOLL (Lemaitre et al., 1996). The first human homolog, hTLR, later TLR4, was discovered shortly thereafter (Medzhitov et al., 1997), and the TLR field has since become a major research area in immunology. TLR receptors are type 1 membrane glycoproteins composed of a relatively highly conserved intracellular cytoplasmic Toll/interleukin-1 receptor (TIR) homology domain (Medzhitov, 2007), and a more variable extracellular leucine-rich repeat (LRR) domain ( Reviewed in: Mikami, Miyashita, Takatsuka, Kuroki, & Matsushima, 2012). The LRR domain has a curved horseshoe-like structure, with parallel beta-pleated sheets forming the inside curve, and a varied structure including alpha helices and varied sequences that form pockets or grooves that conform to all or part of the target PAMP (Botos, Segal, & Davies, 2011; Yoon et al., 2012). Depending on the particular binding site in the LRR ectodomain (of which there can be more than one), TLR receptors can homodimerize, as in the case of TLR3 binding to dsRNA (Leonard et al., 2008; L. Liu et al., 2008), or heterodimerize, as in the case of TLR1&2 binding PAM₃CSK₄ (Jin et al., 2007). The case of TLR4 is an interesting one, in that two unattached molecules of co-receptor MD2 associate with two TLR4 LRR regions.
and two LPS molecules, and the whole complex associates “back to back” to mediate signaling through the TIR domains (Park et al., 2009).

The TLR system includes 10 known receptors in humans (12 in mice), 9 of which have known ligands, which include: lipoproteins, which are detected by TLR2 associated with TLR1 (triacylated lipopeptides), or TLR6 (diacylated lipopeptides) (Aliprantis, Yang, Mark, Suggett, & Devaux, 1999; Brightbill, Libraty, Krutzik, & Yang, 1999; Takeuchi et al., 2001); peptidoglycan (TLR2) (Schwandner, Dziarski, Wesche, & Rothe, 1999; Takeuchi, Hoshino, Kawai, Sanjo, & Takada, 1999; Yoshimura et al., 1999); lipopolysaccharide (TLR4 with the secreted cofactor MD2) (Ozinsky et al., 2000; Shimazu, Akashi, Ogata, & Nagai, 1999; R. B. Yang et al., 1998); conserved sites on bacterial flagellin (TLR5) (Hayashi et al., 2001; Letran et al., 2011; K. D. Smith et al., 2003; Yoon et al., 2012); ssRNA or the chemical agonists imiquimod (TLR7) and resiquimod (TLR7&8); and dsDNA via the unmethylated 2’-deoxyribocytidine-phosphate-guanosine (CpG) motif, which is often methylated in animals, but not in bacteria or viruses (TLR9). It has been predicted via computational modeling that TLR10 binds the di-acylated peptide PamCysPamSK₄ as a homodimer, or Pam₃CSK₄ as a heterodimer with TLR2 (Govindaraj, Manavalan, Lee, & Choi, 2010). Interestingly, there are also emerging roles for the TLR system in so-called sterile inflammation caused by endogenous factors, such as oxidized low-density lipoprotein (LDL) and amyloid-β activation of a CD36/TLR4/TLR6 complex, which can be associated with pathology (Stewart et al., 2010). It is highly likely that other sensory heterocomplexes will be discovered, and will further enrich our understanding of this nuanced system.
TLRs are spatially organized by which cell compartments on which they are displayed, and by which cell types carry them within the body. TLRs are active at the cell surface, or within the endocytic compartment. They are produced within the endoplasmic reticulum (ER), and are either trafficked to the plasma membrane, or to endocytic vesicles that then fuse with active lysosomes. The plasma membrane tropic TLRs are TLR1,2,4,5,6&10, while the endosomal tropic TLRs are TLR3,7,8,&9 (Reviewed in: Kawai & Akira, 2010). The surface TLRs are the least conserved, perhaps suggesting that there are alternative mechanisms for some of their functions, while the endocytic TLRs are more highly conserved (Barreiro et al., 2009), and have been shown to be critical for antiviral immune function (Gowen et al., 2007; Mancuso et al., 2009; S.-Y. Zhang et al., 2007a).

Of course, TLRs are not acting in a vacuum, they are presented on immune cells, many of which are antigen-presenting cells (APC) (Guermonprez, Valladeau, Zitvogel, Théry, & Amigorena, 2002). The major human APCs are dendritic cells (DCs), both classical (cDC) and plasmacytoid (pDC) (Barchet, Cella, & Colonna, 2005), as well as neutrophils, macrophages, and B cells (reviewed in: Geissmann et al., 2010). Interestingly, each of these APC expresses a somewhat unique set of PRRs, which suggests that each cell type may have specific TLR mediated functions (Pulendran & Ahmed, 2006). cDCs and monocytes express TLRs 1,2,3,4,5,6&8 at various levels of expression (Seya, Funami, & Taniguchi, 2005)(Hornung et al., 2002; Seya, Funami, & Taniguchi, 2005) while pDCs express TLRs 1,6,7,9,&10(Hornung et al., 2002). Cell-type specific effects are important in neonatal immunology because it has been hypothesized that specific cellular subsets do not mature in time to mediate protective immunity (H. H. Lee et al., 2008; Siegrist, 2009; Velilla, Rugeles,
& Chougnet, 2006). Whether or not differences in neonatal innate immunity are mediated at the cell type or general level, signaling pathways will be critical in understanding the roots of phenotypic difference.

### 1.2.1 TLR signaling

On dimerization and association of their intracellular TIR domains, the TLR receptors signal through two key intermediaries via the phosphorylation of MyD88 and TRIF. MyD88 mediates signals from all TLRs, except TLR3, via activation of the transcription factor complex NF-κβ as well as the mitogen-activated protein kinases (MAPKs), leading to the induction of inflammatory cytokines (Akira & Uematsu, 2006). On the other hand, TIR domain-containing adaptor protein-inducing IFN-β (TRIF) mediates signals from TLRs 3&4 through MyD88 independent pathways, leading to the activation of the interferon regulatory factor-3 (IRF-3) and NF-κβ transcription factor complexes, and the production of type 1 interferon as well as inflammatory cytokines. Because of the centrality of MyD88 to most TLR signaling, these pathways are generally divided into MyD88 dependent (TLR1,2,5,6,7,8,9), and myD88 independent (TLR 3). The TIR family proteins TRIF-related adaptor molecule (TRAM) and TIR-associated protein (TIRAP) recruit TRIF to TLR4 as well as MyD88 to the TLR2-4 heterodimer TIR regions. This is interesting, as TLR4 must signal through MyD88 to attain full function, though it can also signal through TRIF.

Together, the sum of TLR-mediated responses support immediate innate effector functions (Barton, 2008), such as phagocytosis (Underhill et al., 1999), and they also direct the ensuing
adaptive immune response (Takeshita et al., 2006). It is important to note that innate PRRs are essential in initiating and orchestrating the immune response not just to infection, but—through recognition of adjuvants—they also direct the quantity, quality, and longevity of the adaptive immune response to vaccination (Reviewed in: Pulendran & Ahmed, 2006; 2011).

Our previous studies and those of other groups have identified significant differences between cord and adult peripheral blood responses to TLR stimulation (Levy, 2007; Philbin & Levy, 2009). Our comprehensive comparison of neonatal to adult innate response to TLR stimulation consisted of simultaneous analysis of both global (through cytokine quantification in the culture supernatant) as well as single-cell analysis (via intracellular cytokine staining (ICS) of the response of all four major APCs to a broad range of well-defined TLR stimuli (Blimkie et al., 2010; Corbett et al., 2010; Jansen et al., 2008; Kollmann et al., 2009). Using this platform, we described that monocytes, cDC and pDC from neonates display a consistently and significantly reduced capacity, compared to adults, to produce interleukin-12p70 (IL-12p70), interferon-α2 (IFN-α2) and IFN-γ. This is combined with a more modest and less consistently reduced capacity to produce TNF-α, a similar or even greater capacity to produce IL-6, IL-23 and IL-1β, and a much greater capacity to produce IL-10 in response to TLR stimulation. These differences may relate to the observed increase in neonatal infectious morbidity and mortality as well as in suboptimal immune responses to vaccines administered around birth (Barrios et al., 1996; Levy, 2007; Philbin & Levy, 2009).

However, it is well recognized that while the incidence of infections is highest in the neonatal period, infants continue to remain at increased risk for several years (Klein, Barker,
Remington, & Wilson, 2006; Levy et al., 1999). Furthermore, most vaccines are not administered around birth, but in infancy (Siegrist, 2001). Therefore, a more complete understanding of the ontogeny of the innate immune system over the first few years of life (i.e. beyond the neonatal period) is needed. In the current study we set out to apply our comprehensive approach to innate immune profiling following TLR stimulation over the first 2 years of life. We were able to do so by following a cohort large enough to allow robust statistical analyses. Our findings contradict the notion of a linear progression from an ‘immature’ neonatal to a ‘mature’ adult pattern (Yerkovich et al., 2007), but instead, indicate the existence of qualitative and quantitative age-specific patterns in innate immune reactivity in response to TLR stimulation.
1.3 Innate immunity: Critical in early life

Our working hypothesis in the Kollmann lab is that age–dependent changes in the developing innate immune system are central to differences in vaccine efficacy and survival in the face of infection and injury. We believe that interventions can be developed that modulate suboptimal (Dietert & Piepenbrink, 2008) innate immune response in such a way that vaccination can be effective early enough in life to save millions of lives per year worldwide (Philbin & Levy, 2009; Reviewed in: Siegrist, 2007). Alternatively, innate modulation could save lives directly (Philbin et al., 2012; Wynn & Levy, 2010; Wynn, Neu, Moldawer, & Levy, 2008). The TLR research field is particularly relevant to early life, as studies of individuals with interleukin-1 receptor-associated kinase 4 (IRAK4) mutations, who lack many TLR responses, suffer high infectious mortality in early life (~50%), but do not suffer significant mortality after reaching 8 years of age, and do not suffer from abnormal infectious morbidity beyond age 14 (Ku et al., 2007). This finding has since been replicated, and extended to MyD88 deficiency (Picard et al., 2010).

Interventional methodology for the innate immune response could, by nature, be difficult to implement, due to: (a) the genetic diversity present in the innate immune receptors of humans (Wlasiuk & Nachman, 2010), (b) difficulties acquiring sufficient study sample sizes in neonatal populations (Tarnow-Mordi, Isaacs, & Dutta, 2010), and (c) the complexity of cross-talk inherent in the innate immune sensory system (Ghosh et al., 2007). Differences between individual innate immune systems at the genetic level could explain heterogeneity in both disease susceptibility and vaccine efficacy within populations (Buonaguro & Pulendran,
It is important to note that this methodology differs from the more traditional one pursued by Casanova et al. in New York (Casanova & Abel, 2004; Casanova, Abel, & Quintana-Murci, 2011; Casanova, Holland, & Notarangelo, 2012; Chandesris et al., 2012), where they look for significant immune pathology and search for a genetic cause. In contrast, we are interested in studying the functioning immune system, as subclinical variations in response are far more common, and are more likely to yield insights into broadly applicable correlates of disease and protection when taken together. While we do learn from reductionist case-study approaches, the emerging individualized-medicine and bioinformatics paradigms demand characterization of populations as a whole, not only in sickness.

1.3.1 Relevance to allergy and autoimmunity

Studying the nature of the innate immune response in infants could also provide insights into the pathogenesis of allergy and autoimmunity. It has recently been found that the innate immune system plays a major role in keeping the young adaptive response in check (Reviewed in: Barton, 2008). This regulation is mediated by the presence of TLR ligands within environmental stimuli, and it appears that specific patterns of TLR activation and antigenic stimulation during early development, including peptidoglycan and LPS in household dust, and subclinical infections deriving from breast milk, are tolerogenic and improve the ability of the infant immune system to mount T-helper type-1 (T$_{H1}$) responses in situations where they would previously have mounted ineffectual and potentially allergy inducing T$_{H2}$ or T$_{H17}$ type responses (Reviewed in: Levy, 2007). While an understanding of
the mechanisms of this tolerance induction are beyond the scope of this work, our findings concerning cytokine responses are relevant to this line of research.

1.3.2 **Explanatory paradigms**

The immune response to a given pathogen or insult exhibits a particular character that is defined by the nature of the stimulus and the given homeostatic context. The T helper cell (T\(_{H}\)) is a subclass of T lymphocyte that does not exert significant effector function in the body, but rather, it controls the actions of other surrounding cells, and is required for effective response (Reviewed in: Murphy, 2011). The T helper cell takes its cues from the innate immune system through direct interaction with the APC (Barton, 2008), as well as through indirect signaling by surrounding cells of the surrounding tissue or lymph node (Reviewed in: Murray & Smale, 2012). DCs serve as the primary APCs, and it has been hypothesized that the cDC subset functions as the principal activator of naïve T cells. Accordingly, it is thought that the pDC acts as a kind of “helper” dendritic cell by expressing CD40 on activation, which enhances the ability of the cDC to produce IL-12 (Reviewed in: Murphy, 2011, pp. 342-353). IL-12 production by the activating DC is crucial for the creation of T\(_{H1}\) polarized T\(_{H}\) cells, and has shown to be a major component of the inherent differences between adult and neonatal immune response (H. H. Lee et al., 2008). Indeed, the potent induction of pDC subsets via the TLR7/8 ligand R848 has been shown to vastly augment the neonatal IL-12 response, perhaps indicating a role for the pDC-cDC interaction in the relative lack of IL-12 response to stimuli in this group (Philbin et al., 2012).
modulation of T cell activation is a field currently under intense research (Reviewed in: Germain, 2012), and deserves further reading than could be given here, as this work pertains directly to that rapidly expanding body of research.

1.3.2.1 Innate skewing of neonatal vaccination

Infants are more susceptible to infection, and these early infections can become recalcitrant, as if the immune system learns the ‘wrong’ response, and can not un-learn it (Prendergast, Klenerman, & Goulder, 2012). We also know that this has something to do with the type of immune response mounted, specifically, that infants mount unhelpful TH2 or TH17 responses over TH1 in response to antigens for which the latter would be desirable (Levy, Goriely, & Kollmann, 2012). To better coach this understanding within a research framework, however, we must look at the original research.

Neonatal mice can develop effective TH1 type IFN-γ/CD8+ T cell-mediated responses to Cas-Br-E murine leukemia virus (Cas) if the infection is with a relatively small quantity of infectious virus (0.3 PFU) (Sarzotti, Robbins, & Hoffman, 1996). However, when challenged with increased quantities of the same virus, a TH2 response was the result. This effect persisted when those same mice were re-vaccinated as adults, regardless of the re-challenge concentration, thus suggesting the skewing of the typical adult TH1 response into a TH2 response through prior exposure (Barrios et al., 1996). This work could suggest that early life immune exposures define immune responses to those stimuli for life.
The fundamental difference between these two different neonatal vaccination outcomes appears to be that attenuated or purely antigenic—extracellular—stimuli generate $T_{H2}$ type responses, while those elicited by live infectious—and thus intracellular-available—virus generate effective and protective cell mediated immunity (CMI), which is typically associated with effective Th1 priming and responds to challenge with a Th1 type response in the neonatally vaccinated adult (Fadel, Ozaki, & Sarzotti, 2002). Indeed, work by our own group (Kollmann et al., 2007) has shown that an attenuated Th1 inducing–intracellular vector (a genetically modified ActA deficient strain of *Listeria Monocytogenes*) can induce potent, long-lasting, protective Th1 responses in the neonatal mouse (Kollmann et al., 2007). While the mechanisms of this alternate path to CMI response in the neonate are as yet unclear, they present a tantalizing possibility for future neonatal vaccination strategies.

Given the tendency for Th1 responses to induce maternal rejection during pregnancy (Dubanchet, Martal, & Frydman, 2003; Vitoratos et al., 2006), and the necessity of generating a protective humoral immunity before the termination of breast feeding and the associated supply of maternal antibodies (Chirico, Marzollo, & Cortinovis, 2008), it makes good evolutionary sense that CMI in the neonate should arise only in response to intracellular danger signals. This is especially the case in stringent developmental environments where erroneous responses could cause much harm (Reviewed in: Koga, Aldo, & Mor, 2009). Indeed, within neonatal responses, very different characteristics emerge depending on the extremity of an intracellular pathogen exposure:
“[In neonatal mice] High doses of antigen preferentially stimulate type 2 cytokine responses, eliciting interleukin-4 (IL-4), IL-5, and IL-10 and promoting humoral immunity, whereas low doses of antigen induce type 1 cytokine responses eliciting IL-2, IFN-γ, and tumor necrosis factor-a (TNF-a) and promoting cell-mediated immunity…” (M Sarzotti et al., 1996).

In summary, the neonate is capable of responding to infection with a protective TH1 response, but further work is required to determine how this can be encouraged without causing harm..

1.4 Genetic correlates of innate immune response

While various studies have characterized ex vivo innate immune cytokine response in the very young (Belderbos et al., 2009a; Belderbos, Levy, & Bont, 2009b; Corbett et al., 2010; Karlsson, Hessle, & Rudin, 2002; Kollmann et al., 2009; Nguyen et al., 2010; Upham et al., 2002; Yerkovich et al., 2007), none have attempted to correlate these responses with genetic variation. Given that the TLR receptors are thought to have been subject to heavy selection pressure in human history (Barreiro et al., 2009), and that genetic variation in this system has direct relevance for immunological disease (Netea, van de Veerdomk, van Deuren, & van der Meer, 2011; Schwartz, Netea, & van der Meer, 2011), it follows that association of TLR response with genetic variation could provide useful insight into a critical aspect of immune (dys)function. This nuanced approach was succinctly described as follows:
‘Human genetics attempts to decipher the natural history and pathogenesis of infectious diseases—the most ancient and widespread life-taking experiments of Nature—at the molecular level. In doing so, it reconciles the holistic and reductionist approaches of naturalists and chemists or, to quote Hans Selye, ‘problem-finders’ and ‘problem-solvers’ (Quintana-Murci, Alcaïs, Abel, & Casanova, 2007)

In this tradition, we seek to add genetic information to the search for correlates of healthy immune response, to which correlates of immune protection will later be added in subsequent works via the ascertainment of antibody titers to vaccines for the current study population. To date, correlates of protection are largely undefined, though a major effort is under way to develop tools and methodologies for the task. Of note, Gregory Poland at the Mayo Clinic, and Bali Pulendran at Emory University are making significant progress (Buonaguro & Pulendran, 2011; Haralambieva et al., 2011; Poland, Ovsyannikova, & Jacobson, 2009).
Chapter 2: Body of Thesis

2.1 Methods

2.1.1 TLR stimulation

TLR stimulation plates were prepared as described previously (Kollmann et al. 2009). Briefly, deep-96-well (VWR) source plates containing 1.3 ml of various TLR ligands at 10 times the desired concentration were prepared using sterile procedures under a laminar airflow hood. The following TLR ligands were used at the concentrations noted in the figure or table legends: PAM₃CSK₄ (PAM; TLR2/1; EMC Microcollections); poly(I:C) (pI:C; TLR3; Amersham Biosciences); 0111:B4 LPS (LPS; TLR4, InvivoGen); 3M-003 (3M-003; TLR7/8, 3M); ODN 2336 (CpG-A; TLR9; Coley). For the 6-h ICS plates, brefeldin A (BFA; Sigma-Aldrich) was added at a concentration of 100 µg/ml (10 times the desired final concentration of 10 µg/ml) to all wells, except those wells containing TLR3 or 9 ligands. BFA was not added to the 10× source plates for the plates that were used to obtain 18-h supernatants for Luminex and ELISA cytokine quantification. Source plates were sealed with sterile aluminum plate sealers (USA Scientific), frozen at -80°C, and thawed before use. Twenty microliters from each well of the source plate was dispensed into each well of recipient 96-well round-bottom polystyrene plates (Corning) using the Evolution P3 Precision Pipetting Platform (PerkinElmer) under a laminar airflow hood using sterile procedures. Recipient plates were sealed with sterile aluminum plate sealers and frozen at -80°C until use. As we have shown (Jansen et al., 2008, Blimkie et al., 2011) our approach is
stable over time and between technicians. To further ensure stability of our experimental platform, we also periodically (3 month intervals) ran the same adult controls samples, to ensure all reagents and procedures would fall within < 1 SDEV of the original observed run of these adult controls (Jansen et al. 2008).

2.1.2 Birth cohort study design

All studies were approved by the Institutional Ethics Review Board at both the University of Washington and the University of British Columbia. Our longitudinal study was set-up specifically for the purpose of analyzing the ontogeny of the human innate immune system over the first 2 years of life. Parents were contacted prior to birth at the BC Children’s hospital, and written informed consent was obtained from both parents. Only healthy, full-term newborns after elective Caesarean sections without labor were enrolled in this study between 2005 and 2008.

Prior to each postnatal blood draw, written informed consent was obtained again from the parent or legal guardian. The infant’s health was assessed via a history-driven health assessment (HDHA), focusing on past or current chronic or acute abnormalities or medical conditions. Infants were excluded from the entire study if any the following exclusion criteria were met at any of time of the study: (1) The diagnosis of a significant chronic medical condition including: HIV infection; immune deficiency; immunosuppression by disease or medication; cancer; bone marrow or organ transplantation; blood product administration within the last 3 mo; bleeding disorder; known congenital malformation or genetic disorder;
If the parent or legal guardian were unable to read and/or comprehend English; (3) if the family moved outside of the Greater Vancouver Area during the study period (i.e., would be unavailable for follow-up). Additionally, any febrile illness within the last 24 h, recent immunizations (within 4 weeks of live vaccine, or 7 days after non-live vaccines), or brief (<1 mo) immunosuppressive medication use within one month, would result in a deferral of the blood draw to a later date. Subjects in the birth cohort were 55% female and 45% male; 54% percent of our birth cohort subjects were identified by their parents as having Caucasian ancestry; 32% as having Asian ancestry; 8% as having African American ancestry; 2% as having Native American/First Peoples ancestry; 2% as having Native Hawaiian/Pacific Islander ancestry; 2% as having Hispanic ancestry. This distribution is representative of the American Northwest/Canadian Southwest coasts from which it was randomly sampled. Healthy adults, unrelated to the infants, aged 23 to 48, of equal male-female ratio were recruited from the same area, and had a similar ethnic background as our infant study subjects. All blood draws were performed in the hospital by a trained pediatric phlebotomist. Peripheral blood (7-10 ml) was drawn via sterile venipuncture into Vacutainers containing 143 USP units of sodium-heparin (Becton Dickinson (BD) Biosciences, catalog no. 8019839) using batches we had previously confirmed to be free of innate immune activating substance in assays performed as described elsewhere (Kollmann et al., 2009).

2.1.3 Blood sample processing and in vitro stimulation

Blood samples were processed as described previously (Kollmann et al., 2009). Neonatal cord blood or adult peripheral blood mononuclear cells (MC) were isolated by Ficoll-Paque
density gradient centrifugation. MC were cultured in RPMI 1640 supplemented with 100 U penicillin/ml, 100 mg streptomycin/ml (Invitrogen), and 10% human AB serum (Gemini Bio-Products). Two hundred microliters of cell suspension (2.5 x 10^6 MC/ml) was added to each well of the premade plates containing the specific TLR ligands. For the ICS assays, cells were incubated for 6 h at 37°C in 5% CO_2. For the TLR3 and TLR9 ligands, BFA was added 3 h later at a final concentration of 10 µg/ml, which provides optimal detection of intracellular cytokine production in response to these ligands (Kollmann et al., 2009). After culture, cells were treated with a final concentration of 2 mM EDTA for 15 min at 37°C, then spun down and resuspended in 100 µl of 1x BD FACS Lysing Solution, sealed, and stored frozen at -80°C until staining. An identical set of plates was incubated in parallel for 18 h without BFA; at 18 h, these plates were spun and 100 µl of supernatant was removed and frozen at -80°C for later Luminex analysis.

2.1.4 Assessment of cytokines in culture supernatant

Supernatants were thawed at room temperature, and then filtered through a 1.2-µm filter plate (Millipore) into a clean 96-well plate to remove any remaining cellular debris using a multi-screen HTS vacuum manifold (Millipore). The Luminex assay was performed using the Upstate/Millipore "Flex Kit" system using the high-biotin protocol and overnight incubation at 4°C. Cytokines measured were IL-6, IL-10, IL-12p40, IL-12p70, TNF-α, IFN-α, and IFN-γ. Samples were diluted 1-to-1 (and, if needed to fall within the standard curve, 10- or 20-fold) with RPMI 1640 supplemented 10% human AB serum. Beadlytes, biotin, and streptavidin-phycoerythrin, were used at half the manufacturer’s recommended
concentrations. Assays were read using Luminex 200 Total System (Luminex) running either the Bio-plex (Bio-Rad) or the MasterPlex (MiraiBio) softwares, and the downstream analysis was performed using Excel (Microsoft) and an in-house database. To determine the IL-23 concentration, filtered supernatants were diluted 1:4 in diluent contained in the human IL-23 (p19/p40) ELISA kit (eBioscience), and assays were performed according to the manufacturer’s specifications. Plates were read at 450 nm with 570-nm subtraction. A sigmoid logistic curve was used to generate the standard curve. To allow assessment of the level of cytokine production detected in culture supernatants of unstimulated samples data was not subtracted from the stimulated samples but shown side by side, as this data is possibly biologically relevant (Romero et al., 2006)

2.1.5 Global analysis of secreted cytokine production in response to TLR stimulation of neonatal, 1- and 2-year old, and adult peripheral blood monocytic cells

Our experiments were set up to provide a global view of differences between neonatal and adult cells of cytokines produced after 18 h of stimulation by all cells present in the culture (Figure 1 and Tables 2&3). Although less specific than flow-cytometric analysis (i.e. the source of the cell type producing a given cytokine is not directly identifiable), this multiplexed bead array based approach offers the advantage of permitting quantitative analysis of cytokine responses for which there are no sufficiently sensitive or specific tools available for flow cytometric analysis (e.g. IL-12p70, IL-23). It also allows for cytokine networks to functionally interact in vitro, as cytokine secretion is not inhibited through brefeldin A (BFA). As we had previously shown, multiplexed bead array-based analysis and
ICS do not always produce congruent results, but instead support and complement each other, and in combination, provide what is currently the most comprehensive approach to innate immune profiling (Shooshtari et al., 2010).

2.1.6 Affymetrix GoldenGate SNP array design

Our SNP array was selected based on the Seattle SNPs genome variation suite (GVS) tagging feature that selects tag-SNPs based on known human LD for major populations as assayed in Seattle (Washington, USA), an area with a similar population structure to that of Vancouver (Canada). In this manner, tag-SNPs that mark regions of DNA that are commonly inherited together could be chosen. As well, SNPs that cause peptide changes in key immune related proteins were included to give capacity for candidate gene functional variant discovery. The Illumina GoldenGate technology (Fan et al., 2003) is a form of SNP microarray assay (Reviewed in: Fan et al., 2006) that can be read by an Illumina sequencer to yield high throughput genotyping when sequencing is not feasible. As Fan et al. state in their original 2003 article, “the assay performs allelic discrimination directly on genomic DNA (gDNA), generates a synthetic allele-specific PCR template afterward, then performs PCR on the artificial template. In contrast, conventional SNP genotyping assays typically use PCR to amplify a SNP of interest. Allelic discrimination is then carried out on the PCR product…”. The PCR product itself remains attached to an allele specific fluorophore-linked probe (that identifies which allele is present in the source gDNA), and also a unique pre-selected identifying “tag” sequence in the proximal sequence that can be used to localize this DNA-fluorophore conjugate to a spot on microarray or a microbead surface (Fan et al., 2003).
When this PCR product is annealed to an array and examined with fluorescence microscopy, high levels of one fluorophore at a site indicate a homozygous condition for that variant. Intermediate levels indicate a heterozygote, or (potentially) a genotyping error due to problems with probe specificity or non-specific replication, errors which must be accounted for when analyzing GoldenGate data (Lynch et al., 2009). In this study, GoldenGate genotyping was performed by genomic DNA extraction from cord blood samples, PCR amplification with SNP array specific primers, ligation to the SNP genotyping microarray, and reading of the target SNP base by an Illumina sequencer. Data was then converted from raw image form into a genotyping report via the use of the Illumina Bead Studio software (Illumina Inc.).

2.1.7 Data preparation

Luminex data were processed from raw machine output by the Masterplex CT software package from Mirai Biosystems (San Francisco, California), and mean fluorescent intensity (MFI) values were fitted to 5-paramater logistic curves (5-PL) versus a known 5 point (3 fold) standard curve (10,000 pg/mL to 13.7pg/mL) prepared by seria dilution using the Masterplex QT software package. Curve fits were checked manually and assigned an appropriate weighting (1/X, 1/X^2, 1/Y, 1/Y^2) as per best practices for 5-PL curve fitting (A. Liu, n.d.). Given that data were derived from asymptotic cases at both low and high concentrations, extremely high values (those greater than 10,000 * loading dilution factor * 2) were considered to be unreliable, and were capped at 10,000 * loading dilution factor to facilitate data visualization and analysis. Values that occurred below the minimum
asymptote or defaulted to the minimum value of the standard curve were set to 0 to reduce their effect on relative ranking between treatment groups. Due to the aforementioned effects and transformations, only nonparametric statistics were deemed appropriate for the current dataset. Though higher dilutions could have theoretically prevented saturation of the luminex assay, pilot experiments showed that very high dilution factors in luminex experiments amplify variance significantly. Dilutions less than twofold could not be used due to the limitation of culture supernatant volume, which was required for multiple assays. Accordingly, 2, 10, and 20 fold dilutions of culture supernatant were used, with each cytokine being assigned a dilution appropriate to its experimental concentration range. After extensive testing and review of the data, I decided that variance between dilutions was not a directly linear relationship. Therefore, each cytokine was assigned to a single relevant dilution, in order to avoid cross-comparison between dilutions, where asymptotic effects would occur at different real cytokine concentrations, and where linearity between dilutions could not be assumed. Such considerations were based on careful piloting and analysis of luminex experiments, as well as previous experience (Kollmann et al., 2009).

2.1.8 Pilot cohort data analysis

Data for the 35 subject cohort study of cytokine response to TLR ligands were processed using a combination of Excel (Microsoft, Seattle Washington), the statistical programming language R (R Development Core Team, 2011), and Prism 5 (GraphPad, La Jolla California). Results for each time point were plotted with error bars representing the standard error of mean (SEM).
2.1.9 **Full cohort SNP correlation analysis**

Data analysis presented here constitutes two separate attempts. The first highly exploratory association methodology used as much data as possible and involved no pre-screening for Hardy-Weinberg equilibrium (HWE) or low frequency variants, while the second is derived from established genomics theory and included HWE testing for genotyping errors, as will be explained in more detail below.

2.1.9.1 **Exploratory nonparametric SNP analysis**

In order to parse the dataset into the 48 treatment/cytokine groups required by the study design and perform association testing for all of the measured SNPs (~740), I used the same methodology as for the pilot cohort, but required considerably more computational power and efficiency.

Computational power was acquired through the use of an 8 core Macintosh Pro with 32GB of RAM, and use of the parallel computing R package ‘snow’—which stands for “simple network of workstations”—(Tierney, Rossini, & Li, 2008), which allowed what would have been a week long computation to be performed overnight. The addition of the package ‘snowfall’ (Knaus, Porzelius, & Binder, 2009) allowed for easy distribution of working dataset files to worker processes affiliated with each processor core, and simple aggregation of results in fast RAM; as opposed to hard drive-based file storage, which proved
problematic due to I/O errors in parallel file access. Furthermore, conversion of my R code away from using a for-loop structure—which is known to be computationally inefficient in R—to one using nested functions allowed for a further significant speed enhancement, which allowed for computations to be run, checked, and repeated during working hours. In addition to simple association testing, it was necessary to generate quartile-quartile plots to analyze the distribution of data and p-values using the ‘car’ package (Weisberg, 2010), and to present the most statistically significant genotype sorted cytokine response to defined stimulus results using the ‘beeswarm’ package by Aron Eklund, a selection of which is shown in the results section.

2.1.9.2 GenABEL based SNP association analysis

Data for the full 98-subject cohort was analyzed in the R statistical programming language & environment. The ‘reshape’ package (Wickham, 2007) was used to reformat the data from input-friendly justified table (one row per entry) to the various formats required to work with the dataset, such as the wide (one column per subject) format often required by analytics packages. Data were then screened for errors or absence of critical descriptors before being loaded into the GenABEL statistical genetics package (Aulchenko, Ripke, Isaacs, & van Duijn, 2007), which is an adaptation and expansion of the well known PLINK (Purcell et al., 2007) package developed by Shaun Purcell at MIT. 16 ancestry informative markers (AIMs) (N. Yang et al., 2005) were removed before analysis. AIMs had been included in the genotyping array to indicate population ancestry, and could confound the unbiased population stratification correction employed here. Once loaded into gwaa.class raw data
format, the GenABEL package allowed for quality control of the genotyping by checking for violations of HWE (p \leq 0.05 rejected) and removing SNPs for which variants were too rare (MAF < 5%) to be reliably assayed. Summary of QC was as follows:

738 markers and 95 people in total

113 (15.31165%) markers excluded as having low (<5%) minor allele frequency

Low frequency alleles cause statistical difficulties in small group

50 (6.775068%) markers excluded because of low (<95%) call rate

Low call rates indicate an unreliable marker that may not genotype consistently.

43 (5.826558%) markers excluded because they are out of HWE (P <0.05)

Abnormal deviation from HWE is an indication of systemic genotyping error (Hosking et al., 2004).

1 (1.052632%) people excluded because of low (<95%) call rate

Mean autosomal HET is 0.3656041 (s.e. 0.03711875)

0 people excluded because too high autosomal heterozygosity (FDR <1%)

High or low autosomal heterozygosity can flag systemic genotyping errors.

Mean IBS was 0.6982521 (s.e. 0.02436779), as based on 509 autosomal markers

2 (2.105263%) people excluded because of too high IBS (\geq 0.98)

This was expected, as two study subjects were reportedly twins.

In total, 548 (74.25474%) markers passed all criteria

In total, 92 (96.84211%) people passed all criteria
In addition, the population was tested for population stratification via principal component analysis, as described previously by Paschou et al. (2007), and also in the GenABEL manual (“ABEL tutorial,” 2011). Briefly, the SNP variance was measured via distance matrix calculations for frequencies of all SNPs excluding a set of AIMs, as those markers would bias grouping towards the traditional race-definitions for which they were selected. The resultant distance matrix was then reduced down to a two-factor analysis, from which groups were identified via the k-means clustering algorithm to ensure an approach that was unbiased by prior knowledge of racial groupings (Hartigan & Wong, 1979). Association between a given SNP allele and a given phenotype were deemed significant when controlled for array-wide significance by the GenABEL algorithm QTscore after 200 bootstrap-resamplings to control for sampling effects, and controlling for the effects of gender (via regression compensation), which were shown to describe a significant amount of variance of log2 transformed phenotypic data in the majority of test conditions by ANOVA (p<0.05), whereas birth weight and birth date were not deemed to be sufficiently powerful predictors of variance to include in the model (data not shown). The QTscore algorithm also controls for population stratification by a multi-step process of variance testing and comparison (Aulchenko, 2007). Briefly, once population data is identified for the algorithm, each population may be assessed independently to determine if either population has a significantly different variance, indicating population stratification has a significant effect in the given condition. If the two populations are different with respect to variance, they are analyzed independently, with the top ranked hits being reported for both with correspondingly reduced N values reported. If variance is not significantly different, the population is analyzed as a whole. This methodology allows for populations to be treated as
separate when they vary phenotypically, and to be analyzed together when they do not. In doing so, we strike a balance between sensitivity to population heterogeneity, where sub-populations could give significant results on their own, and population uniformity, where the population as a whole contributes to the significance of a result. In the current study, sample sizes were sufficiently small that no sub-population produced a significant result, however, this was not an *a priori* given. Test conditions for which a significantly associated SNP was found were plotted by the genotypes of that SNP using the beeswarm method mentioned previously. The Seattle SNP Genome Variation Server (GVS) was used to determine what other known SNPs are in linkage with associated SNPs in a similar population composition.
2.2 Results

2.2.1 Luminex analysis of cytokine expression in pilot longitudinal cohort

Figure 1. Bulk measurement of secreted cytokine production in response to TLR stimulation of MC from neonatal, 1- and 2-year olds, and adult.
Cytokines secreted by MC in tissue culture containing the indicated TLR ligands were measured by ELISA (IL-23) or by Luminex’s xMAP cytokine assays. The MC were stimulated for 18 h at 37°C, 5% CO2, after which the supernatants were harvested and frozen at -80°C until time of assay. The IL-23 sandwich ELISA was done on samples from 25 neonate-infant subjects (0, 1, & 2 years of age) and 25 adult subjects (A). The other cytokines were assayed via Luminex multiplex analysis for 35 neonate-infant subjects and 25 adult subjects. The y-axes represent the mean cytokine concentration in pg/ml; error bars indicate SEM. ND = not done.

Graphs were prepared using Excel (Microsoft). Statistical differences between neonate-infant pairs were analyzed using the Wilcoxon matched-pair signed rank test, while the Mann-Whitney unpaired test was used to compare the neonates/infants to the adults. To correct for multiple comparisons, we computed the Bonferroni corrected acceptable Type I error rate (alpha) as 0.05/6 = 0.01 (0.0083), as these would yield false discoveries at a rate of < 0.05 comparing cord blood to 1- and 2- year old infant samples as well as adult, 1 year infant samples to 2 year old infant and adult samples, and 2 year old infant samples to adult (i.e. a total of 6 comparisons for each data set). Accordingly, we consider p-values of less than 0.01 to indicate a significant difference. Significant differences are described here and discussed in the main body. Full statistical results can be found in Appendix B.1.
2.2.1.1 TLR7/8 and TLR9-driven antiviral IFN-α2 responses reached adult levels by 1 year of life

As we have previously shown, following TLR7/8 (3M-003) or TLR9 (CpG-A) stimulation, production of IFN-α2 (the only IFN-α detected by our Luminex assay) was much reduced at birth in comparison to adults. The IFN-α2 response to polyinosinic:polycytidylic acid (pI:C), although overall much lower as compared to TLR7/8 or TLR9 stimulation, was found to be similar between neonatal cord blood and adult peripheral blood. However, in response to either TLR7/8 or TLR9 stimulation, IFN-α2 appeared to be produced at levels similar to adult MCs by 1 year of age. This high-level response was maintained at least up to 2 years of age. In contrast, the IFN-α2 response to TLR3 stimulation actually decreased from birth to age 1, decreasing further again by 2 years of age. Neither PAM nor LPS induced significant IFN-α2 production at any age.

2.2.1.2 Th-1 supporting IL-12p70 production increased but remained below adult levels even at 2 years of age

The production of IL-12p70 was strongest in all age groups in response to pI:C and 3M-003, with only marginal responses induced by LPS, and no significant response to PAM or CpG-A. The capacity to produce IL-12p70 displayed a steady increase from birth onwards in that the same subjects produced more IL-12p70 at 1 year of age than at birth, and again more at 2 years of age than at 1 year of age. But even at 2 years of age, the cohort we followed
produced less than half of the amount of IL-12p70 compared to our adult control group. This trend was the same for TLR3, TLR4, or TLR7/8 stimulation.

2.2.1.3 Th-1-supporting IFN-\(\gamma\) production always remained below adult levels, and in fact was found to decrease from birth up to age 2 years

IFN-\(\gamma\) production was only detectable after TLR7/8 stimulation. In contrast to the trend for IL-12p70, IFN-\(\gamma\) was observed to decrease after TLR7/8 stimulation between birth and 1 year of age, with no significant change thereafter. Importantly, at all early-life time points tested, IFN-\(\gamma\) production in response to TLR7/8 stimulation was found to be far below that of adult MC.

2.2.1.4 Th17-supporting IL-12p40, IL-23 and IL-6 cytokine production declined from birth over the first 2 years of life

We previously described a higher than adult-level of IL-23 production in cord blood MC. In response to LPS and 3M-003, we found production of IL-23 to drop from the cord blood-high to adult-levels by 1-year of age, where it remained up until at least 2 years of age. This trend, a decline from birth to lower levels by 1 year of age, was also observed for the IL-12/23p40 subunit that together with the p19 subunit make up IL-23. However, contrary to IL-23, IL-12p40 in response to TLR4 or TLR7/8 stimulation was detected at much higher levels in the adult. It is important to remember that IL-12p40 also functions as a homodimer (Goriely & Goldman, 2008). PAM did not induce production of IL-23 or IL-12p40 above
background; however, PAM induced significant IL-6 production in all age groups, as did LPS and 3M-003. Similar to IL-12p40, IL-6 dropped from a high cord blood-level that was either at (for 3M-003) or just below (for PAM, LPS) the adult-level of production, to a lower level at 1-year of age and lower again at 2-years of age. As we previously described (J. A. Hoffmann et al., 1999; Jansen et al., 2008; Litman & Cooper, 2007), CpG-A and pI:C produced artifactual high reads for IL-23, and were thus not included in the IL-23 analysis. Neither pI:C nor CpG-A induced levels above unstimulated samples for either IL-12p40 or IL-6.

2.2.1.5 TNF-α and IL-1β both increased over the first 2 years of life for TLR2/1 as well as TLR4 responses, but not for TLR7/8 responses

We previously described an adult-level production of TNF-α in cord blood in response to TLR7/8 stimulation, and we observe the same effect here. The TNF-α response to 3M-003, however, dropped from a high in neonates to significantly lower levels by 1-year of age where they were maintained up to at least age 2 years. IL-1β production in response to TLR7/8 stimulation followed the same exact pattern. However, contrary to the trend following TLR7/8 stimulation, both TNF-α and IL-1β increased from a level below the adult at birth to an adult-level production following TLR2/1 or TLR4 stimulation by 1 year of age, where it was maintained up until at least 2 years of age. Neither pI:C nor CpG-A-induced production of either TNF-α or IL-1β above background in any of the age groups tested.
2.2.1.6 Anti-inflammatory IL-10 production declined from birth onwards to at, or below, adult levels

Cord blood production of the anti-inflammatory and immune-regulatory cytokine IL-10, after TLR stimulation has previously been described by us (Kollmann et al., 2009) and by others (Nguyen et al., 2010) as at or above adult levels. This was indeed the case for this cohort as well, in that TLR2/1 and TLR4 induced production of IL-10 in cord blood at levels similar to adult, and TLR7/8 and TLR9 stimulation induced production of IL-10 in cord blood at levels far above those found in adult MC. However, we found in this longitudinal study that while the ability to produce IL-10 following TLR2/1 stimulation was maintained at adult levels up until the age of 1 year, it dropped to below neonatal and adult levels by 2 years of age. The IL-10 response following TLR4 stimulation declined from birth to two years of age. Finally, IL-10 production following both TLR7/8 and TLR9 stimulation dropped from the high at birth by 1 year, and continued to drop to reach the low adult levels only by age 2 years. pI:C was not found to induce IL-10 production above background in any of the age groups tested.

2.2.2 Preliminary genetic correlates of innate immune response at birth

Given that very early life is the most clinically significant period with regard to infectious mortality in humans (Corbett et al., 2010; Kollmann et al., 2009; Lawn et al., 2004), we have decided to focus our attention the cytokine responses of cells that were collected at birth from cord blood. In order to have some hope of identifying associations, it was necessary to add significantly more individuals to the dataset beyond the 35 included in the longitudinal cohort
study (Corbett et al., 2010). In addition to the initial 35 subjects, we assayed a further 100 subjects, which yielded a viable dataset of N=99 after the removal of subjects with incomplete files. Given that this analysis is exploratory, we did not try to expand this cohort any further. Leaving individuals “in the bank” will allow for the confirmation of any hits within in the same cohort, a methodology, which has been shown to provide more power than simple one-shot analysis. Within cohort validation will minimize confounding variables, and produce robust conclusions. We project that an N of ~100-150 can be achieved using our current biobank. Accordingly, any suggested correlations must be discernable with 95% power at N=100-150 after correction for multiple testing and false discovery rate (FDR). For SNP correlation protocol, refer to section 2.3.3 (pg. 27). While a great many results returned “significant” in nonparametric testing, only one locus appeared many times in the results with very low p-values: TLR1. Given this interesting finding, we further examined this particular locus as it was represented through nonparametric cytokine association.

2.2.2.1 TLR-1

The TLR-1 gene correlated with cytokine response for 4 SNP/Treatment/Cytokine conditions of the top 50 such conditions tested non-parametrically (as per 2.3.3 on pg. 27):

- Uncorrected p-value=0.00018
  - IL-6 PAM rs4833095
- Uncorrected p-value=0.00020
  - IL-12p70 PAM rs5743617
- Uncorrected p-value=0.00032
  - IL-6 PAM rs5743614
Later analysis with the GenABEL suite showed that rs5743617 failed HWE testing (p>0.05), which would render any associations for that SNP within this genotyping dataset unviable due to the probability that such a result was due to genotyping error at that site (Hosking et al., 2004). Accordingly, only the most significant SNP with no HWE errors is included for discussion here, though in a larger cohort—or one with parental genetic information—such variants could be checked by other methods (e.g. TDT analysis) and included or excluded (Reviewed in: Whittemore, 2006). While rs5743617 is likely worthy of follow-up, as will be discussed (pg. 60), at this time we have insufficient data to explore it further.
2.2.2.2 IL6, PAM, rs4833095 (TLR-1 coding SNP, nonsynonymous)

Figure 2. rs2069830 Minor allele associates with increased IL-6 response to PAM. Differences between (top strand) common homozygous AA (0), heterozygous AG (1), and rare homozygous GG (2) neonates (cord blood in vitro) in PBMC IL-6 cytokine response to PAM3CSK4 TLR2 ligand, as represented by median/interquartile boxplot with a per-subject beeswarm plot overlayed for SNP rs4833095.

While this technique was superseded by more advanced analysis, there is something to be learned from the comparison of the two methods. Note the very small p-value despite the presence of outliers that would reduce the strength of a parametric association. In each statistical technique, there are different strengths and weaknesses.

Given that the TLR1 gene is encompassed by a 41kb linkage block in European populations (CEU), and a 28kb linkage block in Han Chinese and Japanese populations (CHB+JPT), this SNP could be tracking any number of variants in TLR1 that have been correlated with immune protection, or vulnerability (Alexopoulou et al., 2002; Georgel, Macquin, & Bahram, 2009; Randhawa et al., 2011; Uciechowski, Imhoff, & Lange, 2011). According to HapMap
ss76891328, this SNP has opposite prevalence in the Caucasian and Asian populations (Gibbs et al., 2003). As the Kruskal-Wallis test that was used to identify this difference in cytokine production relies upon homogeneity of the tested population distribution, a quartile-quartile (Q-Q) plot was used to check the homogeneity of the distributions of each genotype-defined population (Figure 3). Based on Q-Q plot analysis, the genotype populations for rs4833095 deviate from normality, and do so in a relatively uniform way. This would indicate that normal statistics could be applied for power calculations after a mathematical transformation that yields approximate normality (Genser, Cooper, Yazdanbaksh, Barreto, & Rodrigues, 2007).

![Figure 3. Quartile-Quartile plot of genotype groups as defined by the SNP rs4833095 within the TLR1 gene. Here the normal distribution was compared to the quartiles of cytokine response distributions for each group. The theoretical ‘normal’ approximation of the given data is represented by a line derived from the linear regression, with 95% confidence interval denoted by dash-lines.](image-url)
Further analysis of these distributions revealed that they can best be described by an exponential distribution, to which they all conform well (Figure 4).

![Figure 4](image1)

**Figure 4.** Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs 4833095 within the TLRI gene. Here the exponential distribution was compared to the quartiles of cytokine response distributions for each group. The theoretical ‘normal’ approximation of the given data is represented by a linear regression line, with 95% confidence interval denoted by dash-lines.

Given that the data conform to an exponential distribution, it follows that a log-transform of the dataset yield an approximately normal distribution, as shown in Figure 5 below.
Figure 5. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs 4833095 within the TLRI gene. Here the normal distribution was compared to the quartiles of log_{10}-transformed cytokine response distributions for each group. The theoretical ‘normal’ approximation of the given data is represented by a linear regression line, with 95% confidence interval denoted by dash-lines.

The log-transformed dataset is thus, suitable for power calculations for future experiments performed given similar conditions and equipment, which makes for greatly simplified analysis and improved estimation of true statistical power.

Given this information, it was apparent that further statistical work was needed in order to best account for population variability and stratification using advanced modeling, and also for the calculation of sample sizes required to confirm any significant findings. The establishment of methods for analyzing this dataset also allowed for the use of highly computationally efficient programs that are well established within the genetics community.
2.2.3 Genetic correlates of innate immune response at birth

As outlined in the methods section, the R package GenABEL was used as the primary analytics package, as integrated into an R program written to translate the raw study data files into a quality-controlled and standardized format (A.1), and plotted with power calculations (A.3). In accounting for population stratification, the QTscore algorithm needed to be fed a variable defining populations within the genotyped group. In order to avoid the issues inherent in using self-reported race as a population descriptor, we opted for an unbiased approach based on principal component analysis of all-SNP variance (Figure 6). Two significant SNPs were correlated with cytokine response to defined stimulus, and they are presented below. As well, one SNP was close to statistical significance, and was of sufficient interest—given its relevance to the crucial IL-12 gene—to bear further scrutiny.

Figure 6. Graphical representation of principal component analysis of whole-array (minus AIMs) SNP variance.
This analysis gave at least two clearly defined populations, which were then subjected to kmeans clustering, which confirmed the presence of two discernible populations by the genetic similarity that could be ascertained by our SNP array. These two groups were then coded 1 or 0 for the purpose of controlling for population stratification effects within the QTscore algorithm.
2.2.3.1 SNP rs10751209 near the FADD gene and TNFα response to 3M-003

TNFα production in response to 3M-003 varied significantly via allelic correlation analysis by the GenABEL suite QTscore algorithm with stratification for population heterogeneity as detected by PC analysis and kmeans clustering, and correction for effects of gender on variance (p=0.04, N=74). pg/mL TNFα were plotted vs. genotype (Figure 7).

![Graph showing TNFα production in response to 3M-003](image)

**Figure 7 rs10751209 minor allele associates with decreased TNFα response to 3M-003.**

Graph shows common homozygous AA, heterozygous AG, and rare homozygous GG neonates (cord blood *in vitro*) in PBMC TNFα cytokine response to 3M-003 TLR7/8 agonist, as represented by median/interquartile boxplot with a per-subject beeswarm plot overlaid for SNP rs10751209, which has been used to flag variation in the FADD gene.

With the exception of two heterozygous outliers, which are quite obviously more than the standard 2.5SD from the mean, there is an obvious downward trend in production of IL-6 with increasing dosage of the G allele at SNP rs1071209. Power calculations show that the suggested significant difference between AA and AG conditions would require 52
individuals per group to attain a power of 0.8 at an alpha of 0.05. Comparison of AA vs.
AG+GG would require fewer subjects, however, 104 is attainable with our current biobank,
and thus confirmation is feasible. An LD analysis of 10KB on either side of rs10751209 for
the combined HCB and CEU HapMap populations on the Seattle SNP GVS server gives one
SNP in linkage: rs11235564 (Appendix A ). Distribution analysis by QQ plot shows that all
genotypic groups conformed to the exponential distribution through the large majority of the
cytokine pg/mL range, with some deviation at the high and low concentration extremes
(Figure 8). FADD has previously been associated with induction of antibacterial genes,
including IFNs (Balachandran, Thomas, & Barber, 2004).

Figure 8. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs10751209
near the FADD gene. Here the exponential distribution was compared to the quartiles of log2-
transformed cytokine response distributions for each group. The theoretical ‘exponential’
approximation of the given data is represented by a linear regression line, with 95% confidence interval
denoted by dash-lines.
2.2.3.2 SNP rs2069830 (coding, nonsynonymous) in the IL-6 gene and IL-1β response to CpG-A

IL-1β response to CpG-A varied significantly in allelic association testing by the GenABEL suite QTscore algorithm with bootstrap correction, stratification for population heterogeneity—as detected by PC analysis and kmeans clustering—and correction for effects of gender on variance (array wide corrected p=0.005, N=82). pg/mL IL-1β in culture supernatant was plotted versus genotype (Figure 9).

Figure 9. rs2069830 minor allele associates with nonexistent IL-1β response to CpG-A. Differences between common homozygous GG, heterozygous AG (cord blood in vitro) in PBMC IL-1β cytokine response to CpG-A TLR9 ligand, as represented by median/interquartile boxplot with a per-subject beeswarm plot overlaid for SNP rs2069830 which has been used to flag variation in the IL-6 gene.

From the above plot, it is quite obvious that this variant interacts significantly with other unknown factors, given the tremendous heterogeneity in response seen in the common
homozygous condition. Given that rs2069830 is a missense mutation (proline to serine) at position 36 in the protein, it is possible that it is a functional variant. The heterozygous condition shows a severe deficit in IL-1β production in 6 of 7 individuals tested. However, a number of individuals in the homozygous condition also show no significant response. Both genotypic groups show deviation from the expected exponential distribution at high cytokine concentrations (Figure 10).

![Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs2069830 in the IL-6 gene. Here the exponential distribution was compared to the quartiles of log2-transformed cytokine response distributions for each genotype group.](image)

The calculated size of a confirmatory cohort is a mere 7 subjects per condition, or N=14 total to yield a power of 0.8 at an alpha of 0.05. However, given the deviations from the exponential distribution, this should be treated with caution, especially given that deviations occur at higher concentrations, which have a significant effect on parametric statistics.
2.2.3.3 SNP rs583911 in the IL-12A gene and IL-6 response to pI:C

While it did not register as being statistically significant when corrected for array-wide significance, the next most significant SNP by that measure was quite intriguing, and is included here although it is not recommended for follow up with the same strength as the previous significant associations. Grouping by genotype shows a marked negative effect of the haplotype marked by the less common G allele at rs583911 on expression of IL-6 in response to the TLR3 ligand poly I:C (Figure 11).

![Figure 11. rs583911 minor allele associates with decreased IL-6 response to poly I:C. Differences between common homozygous AA, heterozygous AG, and homozygous GG genotype groups (cord blood in vitro) in PBMC IL-6 cytokine response to poly I:C TLR3 ligand, as represented by median/interquartile boxplot with a per-subject beeswarm plot overlaid for SNP rs583911, which has been used to flag variation in the IL-12A gene.](image-url)
There was no marked deviation from the expected exponential distribution in this case (Figure 12).

Figure 12. Quartile-Quartile plot of cytokine response per genotype as defined by the SNP rs583911 in the IL-12A gene. Here the exponential distribution was compared to the quartiles of log_{2}-transformed cytokine response distributions for each genotype group.

Power calculations reveal that confirming a significant difference between AA and AG genotypes would require 43 subjects per condition, for an N of 86 to attain a power of 0.8 at an alpha of 0.05. Comparison of the AA and GG phenotypes would require a smaller sample size, but may not be possible given the relatively small proportion of individuals in that condition and the potentially problematic nature of pre-selecting a cohort by genotype.
Chapter 3: Conclusion

3.1 Bulk cytokine measurement of longitudinal cohort

Our study aimed to profile the human innate immune response to a wide range of well-defined TLR ligands over the first 2 years of life. Using a well-established, robust, high-throughput profiling platform, we were able to follow a large number of children from birth to two years of age. We detected a pattern that suggests the existence of age-specific responses rather than a global, linear progression from a neonatal to an adult pattern (Figure 13).

![Figure 13. General descriptive trends of cytokine response in early life (Corbett et al. 2010).](image)

Specifically, we found an increase in the production of IFN-α2 that reached adult response levels after stimulation with TLR7/8 or TLR9 ligands by 1-year of age. This stood in contrast to the infants’ capacity to produce IL-12p70, which in response to TLR3 and TLR7/8, increased from birth to 1 year of age, and again to 2 years of age; but even then,
response never reached adult levels (Figure 1). While the ability to produce TNF-α or IL-1β reached adult levels after TLR2/1 and TLR4 stimulation, it in fact decreased following stimulation with TLR7/8 to below adult levels by 1 year of age. IL-6, IL-12p40, IL-23 and IL-10 following TLR2/1, TLR4, and TLR7/8 (and TLR9 for IL-10) stimulation decreased by 1 year of age from the neonatal high level, and continued to decrease to either reach or even drop below adult levels of response. These findings suggest potentially important time periods during which vaccination may result in biased immune responses, or particular windows of vulnerability to specific pathogens. Deciphering the mechanisms underlying our observations is likely to yield important insights; this constitutes what we consider a necessary next step investigating innate immune ontogeny.

Most studies examining early life innate immune function have focused on cord blood (Levy, 2007; Philbin & Levy, 2009). Only a few previous studies have analyzed postnatal innate immune development (Angelone et al., 2006; Belderbos et al., 2009a; Nguyen et al., 2010; Upham et al., 2002; Yerkovich et al., 2007). Our study design offered several unique advantages over these previous studies. For example, ours was the only study that examined the ontogeny of innate immune responses in a truly longitudinal fashion—i.e. following the same subjects over time which will likely reduce variability due to genetic heterogeneity (Leonard et al., 2008; L. Liu et al., 2008; Turvey & Hawn, 2006). Furthermore, we were able to enroll, retain and analyze high percentages of our subjects at each time point for each assay, allowing for a well-powered statistical analysis. Lastly, our cohort was followed over the same time period in the life cycle that elevated risk for infectious diseases related morbidity and mortality has repeatedly been observed—the first 2 years of life (Klein et al.,
This same time period also covers the span in which most childhood vaccines are given (Siegrist, 2001).

Our analysis of the cellular composition of the human APC compartment in MC revealed several significant changes over the first 2 years of life, namely the juxtaposed increase of B cells, and a decrease of monocytes from birth to 1- and 2-years of age (Kato, 1935; Schwandner et al., 1999; Shearer, Rosenblatt, & Gelman, 2003; Takeuchi et al., 1999; Yoshimura et al., 1999). Nguyen et al. and Belderbos et al. had looked at WB (Belderbos et al., 2009a; Nguyen et al., 2010), while we analyzed the innate cell content in MC. WB contains neutrophils, the predominant white blood cell in peripheral blood. Neutrophil numbers are known to drop from a high in early life (Kato, 1935) leading to a potential relative increase in monocytes, cDC and pDC, precisely as described by Nguyen et al. and Belderbos et al. (Belderbos et al., 2009a; Nguyen et al., 2010). The changes in the composition of APC in MC over the first few years of life we describe here may impact the cytokine quantification in bulk culture supernatant as shown in Figure 5.

Nguyen et al. followed children from birth up to only 1-year of age by analyzing the impact of TLR4 and TLR9 stimulation on APC surface maturation and cytokine secretion in WB cultures (Nguyen et al., 2010; Ozinsky et al., 2000; Shimazu et al., 1999; R. B. Yang et al., 1998). With respect to the 2 TLRs examined by Nguyen et al., our findings are in full agreement with theirs, but importantly, our studies extend them to 2 years of age and include several additional qualitatively and quantitatively important insights. For example, similar to our results, Nguyen et al. describe that the TNF-α response to LPS reaches adult levels by 1-
year of age, and that the LPS induced IL-6 production drops from a high at birth to lower levels at 1-year of age. In our study, we extended this observation to show that this drop in IL-6 production in response to LPS continued up until at least 2-years of age, in fact to a level below that of the adult. Similarly to our results, Nguyen et al. also found no striking change in the very high early-life LPS induced IL-10 production by 1-year of age. Again, our findings extended this through our discovery of a pronounced drop in LPS induced IL-10 production to occur between 1- and 2-years of age to a level significantly below that of adults. Similar to our findings, Nguyen et al. detect a significant drop in CpG-A induced IL-6 and IL-10 production between birth and 1-year of age. Again, we observed this drop to continue further between years 1 and 2 of life, and to also occur after TLR7/8 stimulation. Both our results and those of Nguyen et al. indicate that the early-life IL-12p70 response to LPS reaches adult-levels by 1-year of age. However, we found that even then the amount of IL-12p70 produced in response to LPS was still markedly lower as compared to, e.g., pI:C or 3M-003. Finally, for the high IL-12p70-inducing TLRs (TLR3 and TLR7/8), while production of IL-12p70 increased after birth, it remained at a significantly lower level than adults even up to 2-years of age.

The fact that our data and that of Nguyen et al. support each other up to at least 1 year post birth, is not only reassuring, it also produces a significant insight that neither study alone had addressed: a comparison of MC to WB responses. For cord blood vs. adult peripheral blood we had previously completed the direct MC vs. WB comparison and found there to be striking and significant qualitative and quantitative differences in TLR responses between neonatal vs. adult MC and WB (Kollmann et al., 2009). The fact that our current
observations using MC and those of Nguyen et al. who used WB (Nguyen et al., 2010) uncovered the same trends over the first year of life suggests that the impact of soluble factors, such as adenosine (Levy, 2007; Levy, Coughlin, & Cronstein, 2006) or other cells contained in WB on the TLR response, while important around birth (Levy, 2007), may be less important by 1-year of age. Another study investigating the innate responses to TLR stimulation from newborn up to 1-mo old infants confirmed, in a direct comparison of MC to WB, that the impact of soluble factors or other cells contained in WB (as compared to MC) on TLR responses disappears by a month after birth (Belderbos et al., 2009a; Hayashi et al., 2001; Letran et al., 2011; K. D. Smith et al., 2003; Yoon et al., 2012).

In work that is similar to ours and that of Nguyen, Belderbos et al., also describe a rapid postnatal increase in IFN-α2 production in response to TLR7/8 stimulation, and an LPS-induced increase in IL-12p70 production by 1 month, but only to levels still below those of the adult (Belderbos et al., 2009a). However, they analyzed the cytokine response in WB culture supernatant after stimulation with LPS following IFN-γ priming, and used concentrations of pI:C and CpG-A much higher than ours. The latter aspect, the need for higher pI:C and CpG-A concentrations when stimulating WB vs. MC is consistent with our previous observation of a lack of response of APC in WB at concentrations that maximally stimulate APC in MC (Jansen et al., 2008; Kollmann et al., 2009; Pulendran & Ahmed, 2006; 2011). Using this much higher concentration of pI:C in WB cultures, Belderbos et al. describe a rapid increase to adult levels of IL-12p70 production by 1-mo of age (Belderbos et al., 2009a). While we also detected an increase in pI:C-induced IL-12p70 production postnatally, we did not see this level reach those of adults even by 2-years of age. This
difference may relate to the higher concentration of pI:C, or may indicate changes between 1-mo and 1-2-years of life. The observation by Belderbos et al. regarding TLR9-induced IL-10 production is in line with our observation and those of Nguyen, in that it appears to rapidly drop off from a neonatal high (Belderbos et al., 2009a; Nguyen et al., 2010). Similar to Belderbos et al., we did not detect a drop in IL-10 production in response to LPS stimulation from the high cord levels by 1 year of age. We did, however, observe a striking drop of IL-10 produced in response to LPS to below adult levels between 1- and 2-years of age.

Two other studies investigating the postnatal ontogeny of innate immune responses in humans come from the same research group (Upham et al., 2002; Yerkovich et al., 2007). Similar to us, this group used MC; but their MC were cryopreserved and thawed prior to TLR stimulation, while ours were stimulated fresh within 4 hours of collecting the blood. Furthermore, this group stimulated with LPS only after IFN-γ priming, while our stimulations occurred without prior priming. Despite these differences in experimental setup, Upham et al. describe an IL-12p70 production pattern consistent with our findings, in that even by 12 years of age, IL-12p70 appears to still not be produced at levels comparable to adults, while early life elevated IL-10 production reaches the low adult level by at least 5 years of age (Upham et al., 2002). Yerkovich et al. observe a level of IL-6 and IL-10 production at birth that is similar to adult levels but which drops off by 1- and even further by 2-years of age (Blimkie et al., 2010; Corbett et al., 2010; Jansen et al., 2008; Kollmann et al., 2009; Yerkovich et al., 2007). This agrees with our findings. Neither Yerkovich et al. nor Upham et al. used any of our other TLR stimuli, limiting the extent to which these studies can be compared to ours (Upham et al., 2002).
Our *in vitro* findings with innate cells from newborns and infants found in blood and after exposure to adult serum may not accurately reflect the response to microbes or vaccine adjuvants found in tissues of an infant. However, several observations suggest that our findings described in this study are at least consistent with clinical presentations of the human newborn and infant. For example, infection with HSV results in the most severe morbidity and mortality if it occurs before 1-month of life (Keller, 2009). As it is IFN-α2 that is most relevant to protection from HSV infection, the rapid increase in IFN-α2 production early in life may explain this clinical observation (Keller, 2009). The delay in IFN-γ and IL-12p70 production beyond 2 years of life is consistent with the known heightened susceptibility beyond 2 years of age to infection with microbes such as TB, where protection is known to depend on robust IFN-γ and IL-12p70 production (Barrios et al., 1996; Donald, Ben J Marais, & Barry, 2010; Levy, 2007; Philbin & Levy, 2009). The change of the age-specific predominant innate immune response we observed may also impact vaccine immune response preferences. For example, BCG immunization at birth has been shown to support a stronger T_{H17}-adaptive T cell response as compared to BCG given at 4 months of age (Burl et al., 2010), while a delay in BCG vaccination for several months post-birth results in a somewhat stronger T_{H1} type T cell response (Kagina et al., 2009). This observation is entirely consistent with the preferential production of T_{H17} supporting innate cytokines around birth we detected.

Overall, our bulk cytokine measurement findings contradict the notion that the ontogeny of the innate immune response to TLR stimulation progresses in a linear fashion from birth to adulthood. Instead, it appears that age-specific regulatory mechanisms are in place
governing the TLR response by the major human APC subsets. With the precise knowledge of the most pertinent ontogenic changes in TLR responses in hand, we are now well positioned to initiate a targeted interrogation of the underlying molecular mechanisms governing early life innate immune ontogeny. This, in turn, will form the rational basis on which to attempt age-specific and age-appropriate interventions aimed at improving the immune-mediated protection of this highly vulnerable population. Accordingly, we now proceed to discuss preliminary immunogenetic analysis of the present cohort.

3.2 Pediatric immunogenetics

3.2.1 Introduction

Engaging in the difficult work of finding meaningful information within a sea of data has become a core aspect of contemporary biomedical research (Petsko, 2007). The dawning of the genomic age has created a need for skills of data quality-control, manipulation, and analysis that were previously entirely outside the scope of the average researcher. In addition to designing methodologies to improve the reliability and usability of the large luminex dataset presented here, interfacing this data with genetic and clinical correlates and covariates presented a task that seemed impossible with simple tools like Microsoft Excel. While there were off-the-shelf analytic programs available for purchase, there was no budget for such expensive virtual equipment. Accordingly, the R scripts and analysis presented here represent only one manifestation of what was a lengthy and labor-intensive effort to learn a new language and a more mathematical/logical view of science and statistical inference.
3.2.2 Kruskal-Wallis significance testing and rs4833095

Initially, our thinking in approaching the correlation of SNP variants with cytokine levels was straightforward. Given that our data did not fit a normal distribution, standard ANOVA type statistics could not be used to determine which individuals carrying variant genotypes were significantly different from others. Without a theoretical population distribution on which to base the modeling required to make inferences about the significance (i.e. the probability of an observation reflecting a reasonably reliable reality), one must resort to rank-based measures of difference that do not take into account the scale of the observations, or their distribution with regard to the measure, but rather, their relative position to each other. Specifically, the Kruskal-Wallis (KW) test was developed to determine if there is significant difference between two or more groups based on their relative rank distributions. While the KW test allows for scale-free quantification of difference, it relies on some fundamental assumptions. Primarily, it depends on the distribution of each group to be compared being similar. This arises from the logic that if group distributions are similar, then the presence of a proportion of data points above those of another group should indicate that one is larger than the other. However, split distributions, as can be the case in epistatic phenomena (where more factors than the one being dissected influence the dependent variable), violate this logic. In looking through hundreds of top-ranked associations, we found that there was a certain logic to these associations, but that they were rarely convincing to the eye of a biologist looking for something that influences objective biological reality. Some cogent examples presented themselves, especially surrounding the TLR1 gene (Section 2.2.2.2), which is crucial in the activity of the TLR1/2 and TLR1/6 complexes. The association
presented here shows an effect on IL-6 production in response to the TLR-2 ligand PAM3CSK, with the rare variant being more responsive (Figure 2). Given the number of “hits” in this region, it is possible that any of a number of variants are causative (Alexopoulou et al., 2002; Georgel et al., 2009; Randhawa et al., 2011; Uciechowski et al., 2011). In addition the SNP rs4833095 is a non-synonymous variant causing an Asparagine to Serine, which would eliminate a positively charged -NH2 group, replacing it with a smaller structure terminating in a negatively charged –OH group. This SNP has been associated with poor maternal and fetal outcomes in women infected with *P. falciparum*, including a doubling of rates of malarial anemia (Hamann, Bedu-Addo, Eggelte, Schumann, & Mockenhaupt, 2010). Given that the variant of this SNP common in Caucasian populations appears to decrease production of IL-6, this raises interesting questions as to the mechanism of reduced resistance to *P. falciparum* infection. It is interesting to note that prevalence of this SNP is opposite in Caucasian (CEU) and all other (HCB,JPT,YRI) HapMap populations (Frazer et al., 2007), perhaps reflecting selection pressure caused by *P. falciparum* and other diseases endemic to equatorial regions. An application of this research finding could be to test for low IL-6 production by PBMC cells of individuals who have chronic recurrent or severe *P. falciparum* infection. Alternatively, IL-6 could be administered to individuals with *P. falciparum* infection to determine if this improves outcomes, though such an intervention would be difficult to carry out given the systemic effects of IL-6 administration and the cost of the recombinant protein.

One association that dominated the nonparametric association testing, rs5743617, did not have allele frequencies remotely close to populations in the HapMap datasets. In testing for
HWE, it became clear that this was a genotyping error that happened to segregate multiple cytokines in a remarkably enticing way. It is puzzling that a genotyping error would be able to do this, and thus, follow-up for this SNP is warranted, as something may have been causing the genotyping error that associates closely with immune response. Given that genotyping was performed on pellets of the same cells that produced the cytokines, it is conceivable that some kind of recalcitrant protein interacting with the rs5743617 site interfered with genotyping in an immune-indicative way. Via the use of a stringent HWE cutoff of p=0.05 rs5743617 no longer appeared as an association in subsequent analysis. This was an important lesson in genetic quality control, but also, perhaps, a lead worthy of a side-project.

3.2.3 GenABEL association testing identifies rs10751209 & rs2069830 as significant predictors of neonatal cytokine response to 3M-003 and CpG-A, respectively

Scale is important in biology. A very high cytokine concentration does not occur for no particular reason. Rather, the multicellular system that we study in vitro in culturing freshly stimulated human PBMCs is a computational system that processes the given input and yields a variety of graded responses. A very high response means something different than does one that is merely the highest of a group that vary stochastically around a mean response. Despite the fact that our machinery and methods are not always capable of reflecting the naturally occurring distributions, due to measurement bias at the high and low ends of a given range, the majority of responses conform to an exponential distribution (Figure 8,10,&12). Through capping of the maximum value at a reasonable point (see
methods), we can ensure that limit-effects inherent in the 5-PL logistic model used to derive cytokine concentrations do not unduly influence parametric statistical approaches. Log$_2$ conversion is common in phenotype-SNP association (Balding, 2006; Genser et al., 2007), and it is especially pertinent in this case, given the trend towards exponential response distributions. Accordingly, data were log$_2$ transformed before being error checked and imported into the GenABEL genomics package in R for advanced association testing. The QTscore algorithm was selected for its ability to perform adjustment of the linear model for covariates (of which subject gender was deemed to be most broadly applicable across cytokines and treatments via exploratory ANOVA analysis as per the GenABEL manual) population stratification adjustment using unbiased whole-array SNP variance derived populations, also as per the GenABEL manual, and quality control of the genomic data, including HWE testing, among others. Though the current cohort did include one pair of twins (IBS>0.98), they were excluded, given that the methodology of association used here is for unrelated individuals with no family genotyping information, and the impossibility of gaining significance from only two twins. Correction for population distribution effects was performed by bootstrap resampling 200 times, also as per the GenABEL manual (“ABEL tutorial,” 2011; Aulchenko et al., 2007). P values were also corrected for multiple testing bias with 1 degree of freedom (Pc1df). In effect, this method accounts for and attempts to mitigate all significant confounds of the relationship between a given allelic variant and a given cytokine response to a defined treatment.

We have identified here a significant (corrected p=0.04) negative effect of the G allele of rs10751209 on production of the canonical inflammatory cytokine TNFα at birth, which has
well established roles in innate immunity, sepsis, Rheumatoid Arthritis, and induction of apoptosis in cells expressing the receptor TNFR1 (Apostolaki, Armaka, Victoratos, & Kollias, 2010; Peppel et al., 1991).

The association between variation at a SNP near the FADD gene and TNFα production levels could be mediated in two different ways, relating to two different functions of FADD in innate immune cells (Reviewed in: Tourneur & Chiocchia, 2010). The first, more traditional, function is the induction of apoptosis after the activation of the TNFR1 by TNFα binding. In cases where FADD is not present, the activated TNFR1 complex does not induce apoptosis. Accordingly, the expression levels and functionality of the FADD protein in the cytoplasm is crucial in determining cellular responses to TNFα (Micheau & Tschopp, 2003). In PBMCs, this could serve as a self-regulatory system whereby cells producing TNFα also express the TNFR1 receptor, instituting an apoptotic negative-feedback loop, whereby expression is contained to prevent tissue damage or sepsis. In such a scenario, a gain of function mutation in FADD would result in the reduced TNFα expression seen here. Conversely, it has been hypothesized that FADD plays a key role in the induction of IFNβ in response to intracellular dsRNA in a TLR3 independent fashion. It is possible that this “innateosome” or RIFT pathway proposed by Balachandran et al. (2004) plays a role in the response to 3M-003 in human PBMC cells, and that a hypofunctional variant in FADD exerts an inhibitory effect on that pathway, leading to lowered activation of IRF7 (Balachandran, Venkataraman, Fisher, & Barber, 2007). Type I IFNs have been shown to mediate TNFα expression in activated Macrophages (Mancuso et al., 2007), and also in neonatal survival of acute inflammation (X. Zhang et al., 2007b). FADD deficiency has been associated with a suite of disorders,
including susceptibility to viral infections, in humans (Bolze et al., 2010). It is possible that FADD plays a direct role in endosomal TLR signaling, or that the RIG-I pathways, with which it is known to be associated, play a role in response to 3M-003. Accordingly, further study of the role of FADD in neonatal innate immune TLR response is warranted based on the results shown here. Power calculations show that the suggested significant difference between AA and AG conditions would require 52 individuals per group to attain a power of 0.8 at an alpha of 0.05. Comparison of AA vs AG+GG would require fewer subjects, however, 104 is attainable with our current biobank, and thus confirmation is feasible. An LD analysis of 10KB on either side of rs10751209 for the combined HCB and CEU HapMap populations on the Seattle SNP GVS server gives one SNP in linkage: rs11235564 (Appendix B), suggesting that this variant and surroundings should also be included in any detailed search for the causative variant.

Given that rs2069830 is a missense mutation (proline to serine) at position 36 in the IL-6 protein, it is possible that it is a functional variant. Previous work has shown that the first 28 bases at the N terminus of IL-6 are expendable, while positively charged C terminal amino acids are crucial for function (Krüttgen et al., 1990; Lütticken, Krüttgen, Möller, Heinrich, & Rose-John, 1991). The question of the role of the remaining N-proximal amino acids has been partially addressed by haplotype based studies of Hepatitis B infection, where it is clear that there is an association between this C proximal region, or associated regulatory elements, and antiviral immune response in adults (Yee, Im, Borg, Yang, & Liang, 2009). The heterozygous condition shows a severe deficit in IL-1β production in 6 of 7 individuals tested (Figure 9). However, a number of individuals in the homozygous condition also show no
significant response also, suggesting that there is a potential epistasis effect or unrelated environmental cause. While IL-6 has been shown to induce significant levels of IL-1 receptor antagonist *in vivo*, this has not been shown *in vitro* (Tilg, Trehu, Atkins, Dinarello, & Mier, 1994). This as-yet unobserved phenomenon could represent a novel IL-6 gain of function mutation wherein the usual function of IL-6 in transitioning the immune response away from early mediators (i.e. TNFα and IL-1β) is significantly exaggerated. If this is the case, then individuals carrying this relatively rare variant will be particularly resilient to inflammatory tissue damage early in life (Zhao et al., 2008). This theory is given further credence by the fact that treatment with anikinra, an IL-1β receptor antagonist has been shown to significantly reduce pathologies associated with neonatal-onset multisystem inflammatory disease (Goldbach-Mansky et al., 2006). If this hypothesis is correct, this variant will be one to watch. The fact that this effect was seen for CpG-A stimulation specifically suggests that there may be a role for IL-6 production by the dendritic cell in inducing IL-1β production by other cells within the PBMC that has been compromised by this genetic variant. Rs2069830 was not shown to be in LD with any nearby variants in the GVS database. Power analysis shows that a relatively small confirmatory cohort of 14 would achieve the desired power to confirm the significance of this effect. Given deviation from the expected distribution, confirmation is especially important here, and a cohort several fold larger than calculated will likely be necessary to proceed with confidence.

While the association was not statistically significant (Pc1df=0.08), rs583911 represents a variant with known immune function in a gene that is key to the T_{H1} immune response, IL-12A. Accordingly, it is worth a brief discussion. rs583911 has been associated with
childhood ALL (J. S. Chang et al., 2010), which suggests a role for variants in this genomic region in the development of the adaptive and innate immune systems. If this variant really does induce decreased IL-6 production in early life, this could have major implications for the health of individuals carrying this variant. High IL-6 production and insufficient IL-12p70 has been correlated with inflammatory conditions and poor response to BCG vaccination in neonates (H. H. Lee et al., 2008; Siegrist, 2007). An N of 86 will be required to confirm the suspected difference between AA and AG genotypes in IL-6 production with a power of 0.8 at an alpha of 0.05.

The fact that the top associations were both with intracellular—and thus what are often described as antiviral—TLRs (TLR7/8 and TLR9) is interesting, especially given the growing awareness in pediatric immunology of the importance of developmental stages in the quality and character of antiviral immune response (Reviewed in: Prendergast et al., 2012). Because these variants modulate the cytokine responses (IL-1β and TNFα) that play key roles in modulating the advancement to adaptive response or retreat into tolerance (Reviewed in: Germain, 2012), it could be possible that such genetic differences could act as part of a genotypic array for use in predicting outcomes for infants with viral infection. Indeed, the previously discussed SNP rs4833095 predicted a tremendous difference in the incidence of malaria anemia. Such a tool could be instrumental in determining the level of medical intervention that is warranted. Medical intervention always comes with risks and costs, and thus the development of a suite of tools with prognostic value for the course of an infection can give physicians invaluable insight into the costs versus benefits of interventions.
3.3 Suggested future research directions

SNP-phenotype associations suggested here may be confirmed within the same cohort using biobanked samples. For future studies, incorporation of more environmental variables will allow for more robust multiple-regression strategies that can determine the contribution of a wider range of variables to cytokine responses. The CHILD study aims to do exactly that, and is under way at sites across Canada.

Given that significant variability was seen within most SNPs plotted by genotype, it stands to reason that SNPxSNP effects are taking place, as multiple variants interact in the intimate signaling networks that mediate cytokine responses. Previous work on immune networks has shown that multiple variants can make significant contributions to disease progression (Giha, Nasr, & Iriemenam, 2011). While it was previously computationally prohibitive, epistasis detection has recently been made quite feasible by the advent of massively parallel processing techniques (Reviewed in: Wang, Liu, Feng, & Wong, 2011) that can be carried out on consumer-grade graphics processing units (GPUs), which contain hundreds of processor cores linked to fast memory. While the associations detected here were interesting and worthy of follow-up, it is clear that any single genetic variant can not explain most of the variability in a given defined response to a defined stimulus, excepting where that variant compromises a critical component. Accordingly, epistasis testing and environmental factors will be needed to understand variance in cytokine production. Beyond this, subsequent health outcomes will need to be correlated with variants of interest to determine their utility as descriptors of the character of a given immune system as it develops and responds to
infections and vaccines, or as it produces autoimmunity and other defects of immunoregulation.

3.4 Strengths and weaknesses of this project

The work presented here represents the first longitudinal cohort analysis of infant innate immune response to incorporate candidate-gene genetic correlates, and provide for further follow-up study to confirm preliminary conclusions. While candidate gene studies have looked for associations with vaccine response, they often neglect the fundamental innate-immune relevant genes and phenotypes that were tested here (Haralambieva et al., 2011). In looking for associations known adaptive-immunity related genes, studies such as these turn up little that is novel. By contrast, we struck out to identify variants that influence innate immune response and (in later phases) vaccine response. Cohort methodology reduces variance due to genetic and environmental confounding factors, as the same individuals make up each time point. This stands in contrast to other studies, where each time point is made up of a different cohort of infants. In seeking to understand the fundamental mechanisms of immune response, cohorts present a wealth of information, and the work presented here represents only the beginning of what will be a deeply informative body of research on this population.

An inherent weakness in study designs of this type is the difficulty of determining true associations in diverse population with obvious genetic population stratification effects on immune response. A solution to this problem would be to collect blood and genetic consent from the parents as well as infants, which would allow for use of the Transmission
Disequilibrium Test (TDT), which is not susceptible to population stratification confounding, and is designed to quantify SNP-phenotypic associations (Reviewed in: Whittemore, 2006). While compensating for population structure by accounting for differences in population variance as we have done here does improve the reliability of results, it does not remove the problem entirely, as the TDT would do. Given the difficulty of finding subjects for this kind of study, it will be advisable to attempt to recruit parents as well in future studies. In addition, familial information would allow for better assessment of the LD “landscape” around each SNP (Reviewed in: Balding, 2006).

3.5 Conclusion

The integrative conclusion from this body of work is that the innate immune response changes considerably from birth to two years of age, and variability within that response can be at least partially explained by genetic variation. We show here that infant PBMCs produce significantly less Th1 stimulating cytokine in response to TLR ligands, and that they produce high levels of Th17 inducing cytokines. We also show that variants in the FADD, IL-12A and IL-6 genes can have significant effects on innate immune response to TLR signals at birth. Accordingly, we present here a synthesis of two core modalities in the study of immune development: the genetic basis of immune variability, and the ontogeny of the human immune system. The next logical step will be to extend this analysis to determine if there are genetic variants that influence the ontogeny itself, for which an expanded cohort of year 1 and year 2 cytokine quantitation is currently under way. While the modulation of immune response by genetics is something that is readily comprehensible, and an area of
intense study (Poland, Ovsyannikova, Jacobson, & Smith, 2007), the genetic modulation of immune ontogeny is a largely undiscovered country that leads us to ask deeper questions about the fundamental network structure of the immune system (G. W. Hoffmann, n.d.), and how this structure changes as it is formed during development. This is true of developmental biology in general, that it forces us to ask hard questions about the creation of complex structures (Philbin & Levy, 2009). It is within these difficult questions that we will find a deeper understanding of this core life-sustaining system.
References


Burl, S., Adetifa, U. J., Cox, M., Touray, E., Ota, M. O., Marchant, A., Whittle, H., et al. (2010). Delaying Bacillus Calmette-Guerin Vaccination from Birth to 4 1/2 Months of Age Reduces Postvaccination Th1 and IL-17 Responses but Leads to Comparable Mycobacterial Responses at 9 Months of Age. The Journal of Immunology, 185(4), 2620–2628. doi:10.4049/jimmunol.1000552


85


Appendices

Appendix A

R based statistical analysis of pediatric immunophenotyping cohort.

A.1 Statistical testing for significant cytokine response over control.

Table 1. Kruskal-Wallis testing for significant difference between unstimulated and stimulated samples.

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>Treatment</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-10</td>
<td>siC</td>
<td>1.62E-11</td>
</tr>
<tr>
<td>IL-10</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IL-10</td>
<td>3M-003</td>
<td>8.44E-59</td>
</tr>
<tr>
<td>IL-10</td>
<td>CpG-A</td>
<td>4.22E-42</td>
</tr>
<tr>
<td>IL-10</td>
<td>JPS</td>
<td>7.85E-50</td>
</tr>
<tr>
<td>IL-10</td>
<td>PAM</td>
<td>1.03E-36</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>siC</td>
<td>1.83E-44</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>3M-003</td>
<td>4.18E-79</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>CpG-A</td>
<td>1.33E-08</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>JPS</td>
<td>1.63E-28</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>PAM</td>
<td>1.42E-25</td>
</tr>
<tr>
<td>IL-12p70</td>
<td>siC</td>
<td>1.32E-33</td>
</tr>
<tr>
<td>IL-12p70</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IL-12p70</td>
<td>3M-003</td>
<td>2.04E-31</td>
</tr>
<tr>
<td>IL-12p70</td>
<td>CpG-A</td>
<td>0.053054975</td>
</tr>
<tr>
<td>IL-12p70</td>
<td>JPS</td>
<td>0.00059797</td>
</tr>
<tr>
<td>IL-12p70</td>
<td>PAM</td>
<td>0.7724118</td>
</tr>
<tr>
<td>IFN-α</td>
<td>siC</td>
<td>0.008299</td>
</tr>
<tr>
<td>IFN-α</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IFN-α</td>
<td>3M-003</td>
<td>1.01E-10</td>
</tr>
<tr>
<td>IFN-α</td>
<td>CpG-A</td>
<td>0.01583152</td>
</tr>
<tr>
<td>IFN-α</td>
<td>JPS</td>
<td>9.11E-12</td>
</tr>
<tr>
<td>IFN-α</td>
<td>PAM</td>
<td>9.13E-12</td>
</tr>
<tr>
<td>TNFα</td>
<td>siC</td>
<td>1.35E-11</td>
</tr>
<tr>
<td>TNFα</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>TNFα</td>
<td>3M-003</td>
<td>0.64E-18</td>
</tr>
<tr>
<td>TNFα</td>
<td>CpG-A</td>
<td>2.42E-12</td>
</tr>
<tr>
<td>TNFα</td>
<td>JPS</td>
<td>1.18E-15</td>
</tr>
<tr>
<td>TNFα</td>
<td>PAM</td>
<td>3.30E-15</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>siC</td>
<td>6.72E-55</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>3M-003</td>
<td>3.42E-114</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>CpG-A</td>
<td>7.04E-12</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>JPS</td>
<td>2.34E-30</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>PAM</td>
<td>1.37E-25</td>
</tr>
<tr>
<td>IFN-α2</td>
<td>siC</td>
<td>3.12E-16</td>
</tr>
<tr>
<td>IFN-α2</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IFN-α2</td>
<td>3M-003</td>
<td>6.11E-09</td>
</tr>
<tr>
<td>IFN-α2</td>
<td>CpG-A</td>
<td>1.65E-30</td>
</tr>
<tr>
<td>IFN-α2</td>
<td>JPS</td>
<td>0.4585247</td>
</tr>
<tr>
<td>IFN-α2</td>
<td>PAM</td>
<td>0.13297817</td>
</tr>
<tr>
<td>IL-23</td>
<td>siC</td>
<td>5.12E-19</td>
</tr>
<tr>
<td>IL-23</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IL-23</td>
<td>3M-003</td>
<td>1.13E-07</td>
</tr>
<tr>
<td>IL-23</td>
<td>CpG-A</td>
<td>5.60E-19</td>
</tr>
<tr>
<td>IL-23</td>
<td>JPS</td>
<td>1.11E-08</td>
</tr>
<tr>
<td>IL-23</td>
<td>PAM</td>
<td>0.017892772</td>
</tr>
<tr>
<td>IL-1b</td>
<td>siC</td>
<td>2.12E-09</td>
</tr>
<tr>
<td>IL-1b</td>
<td>Jstim</td>
<td>1</td>
</tr>
<tr>
<td>IL-1b</td>
<td>3M-003</td>
<td>8.89E-22</td>
</tr>
<tr>
<td>IL-1b</td>
<td>CpG-A</td>
<td>2.42E-07</td>
</tr>
<tr>
<td>IL-1b</td>
<td>JPS</td>
<td>1.33E-17</td>
</tr>
<tr>
<td>IL-1b</td>
<td>PAM</td>
<td>6.88E-16</td>
</tr>
</tbody>
</table>
A.2 R program to perform statistical analysis of NIH01 cohort year 0 time point.

# Set year to analyze:  
Yr <- 0

# Repository: http://cran.stat.sfu.ca/  
# Install necessary packages if absent:  
if(require(date)==FALSE)  
{  
install.package('date',repos="http://cran.stat.sfu.ca")  
}
library(date)

if(require(mail)==FALSE)  
{  
install.packages('mail',repos="http://cran.stat.sfu.ca")  
}
library(mail)

# GenABEL  
if(require(GenABEL)==FALSE)  
{  
install.packages('GenABEL',repos="http://cran.stat.sfu.ca")  
}
library(GenABEL)

# reshape  
if(require(reshape)==FALSE)  
{  
install.packages('reshape',repos="http://cran.stat.sfu.ca")  
}
library(reshape)

SNPs <- read.table("Source Data/NIH01_SNP.csv", header=T, sep="","as.is=T")
SNPs<- SNPs[!duplicated(SNPs$SNP_key),]

# Now, to get the SNP information in illumina like format, we need genotypes that follow the  
# C/- . Here, "-" stands of a deletion. Missing data can be coded as "-" or "00". Make sure that  
# the coding for missing is "00" if you use one of the codings A/-, T/-, G/-, C/- !  
SNPs$Genotype<-sub(" ","",SNPs$Genotype)  
SNPs[,Genotype'] <- sub("-","00",SNPs[,Genotype'])
#Remove AIMs SNPs from analysis, as they could confound the population stratification compensation.
AIMS <- read.table('AIMS.csv',header=F,sep=',',as.is=T)
SNP <- SNPs[!SNPs$SNP_Name %in% AIMS$V1,]
AIMSrem <- SNPs[SNPs$SNP_Name %in% AIMS$V1,]

#Cleanup
SNPs <- SNP
rm(SNP)
molten<-melt.data.frame(SNPs,measure.vars='Genotype')
casted<-cast(molten,formula=SNP_Name~Subject,add.missing=T,fill="00")
load("Source Data/SNPinfo.RData")
NIH01_SNP_wide<- merge(pheno,casted,by.x='name',by.y='SNP_Name')
#I now have a data.frame that is quite similar to the input format given in the ABEL tutorial.
NIH01_SNP_wide_nospace<-NIH01_SNP_wide

print(paste('Prepared genotype data for ',length(NIH01_SNP_wide_nospace[1,])-4," subjects.

#Now we prepare the phenotype information
luminex <- read.csv("Source Data/Luminex2.csv", header=TRUE, row.names='Key',
fill=TRUE, strip.white=TRUE, comment.char= "*", na.strings= c("#N/A",""), as.is=T)
luminex$Accepted_Value=gsub","","",luminex$Accepted_Value)
luminex <- luminex[!is.na(as.numeric(luminex$Accepted_Value)),]
print('Luminex loaded')
enrol.dat<- read.csv('Source Data/Enrolment data.csv', header=TRUE, fill=TRUE,
strip.white=TRUE, comment.char= "*", na.strings= c("#N/A",""),0, as.is=T)
enrol.dat$Weight<-sub("NA",NA,enrol.dat$Weight)
enrol.dat$Weight <-as.numeric(enrol.dat$Weight)
print('Enrolment data loaded')

cohort <- enrol.dat[!is.na(enrol.dat$Subject),]
print('Enrolment data subsetted to Study.ID with valid Subject# for current study')

NIH01 <- merge(cohort,luminex,by='Subject')
NIH01 <- NIH01[,c(1,6:11,13:17)]
colnames(NIH01)<-
c('id','dob_mm.dd.yyyy','sex','m.ethnic','f.ethnic','multiplicity','weight','year','cytokine','stim','dilution','value')
save(NIH01,file="Data Backup/NIH01_cohort_and_luminex.RData'
print('NIH01 cohort and luminex file completed and saved for reference.'

90
NIH01 <- NIH01[which(NIH01$year==Yr),]

# Fix typographic error in dob (checked samples before and after to infer correct year).
NIH01$dob_mm.dd.yyyy <- sub("6-3-2005","6-3-2008",NIH01$dob_mm.dd.yyyy)

# Convert dob to numeric
NIH01$dob <- as.numeric(as.date(NIH01$dob_mm.dd.yyyy,"mdy"))

# Set first birthday as day zero.
NIH01$dob <- NIH01$dob-min(NIH01$dob)

# Set unknown weights to the population mean
is.na(NIH01$weight) <- mean(na.omit(NIH01$weight))

# Transform dataset to something like normal distribution.
NIH01$value <- log2(as.numeric(NIH01$value)+0.1)

# Test for significant effect of treatment vs. unstim, and remove all conditions that did not induce a significant effect.
cytokines <- unique(NIH01$cytokine)
treatments <- unique(NIH01$stim)

# Parsing the dataset into unit of analysis groups.
result <- data.frame()
for(j in 1:length(cytokines))
{
  cyt <- cytokines[j]
  for(i in 1:length(treatments))
  {
    trt <- treatments[i]
    temp <- subset(NIH01, cytokine==cyt & stim==trt)
    control <- subset(NIH01, cytokine==cyt & stim="Unstim")
    test <- t.test(as.numeric(temp$value), y=as.numeric(control$value), alternative= "two.sided")
    current <- cbind(cyt, trt, test$p.value)
    result <- rbind(result, current)
  }
}
result$V3 <- as.numeric(levels(result$V3)[result$V3])
write.csv(result, file= "results/TestConditionVsUnstImp.values.csv")

drop.list <- result[result$V3>0.05,]

# Format the drop list such that it matches the GenABEL format.
drop.list <- paste(drop.list$cyt, drop.list$trt, sep= "_")
drop.list <- sub("-", ", ", drop.list)

# Reshape the dataset for import to GenABEL
molten <- melt.data.frame(NIH01, measure.vars= "value")
casted <- cast(NIH01, formula= id+ sex+ dob+ weight~ cytokine+ stim)
# Convert M/F sex code to the GenABEL 1/0 standard.
casted[, 'sex'] <- sub("F", 0, casted[, 'sex'])
casted[, 'sex'] <- sub("M", 1, casted[, 'sex'])

# set unknown sex individuals to female. If they are male they will not pass the QC check.
casted[which(is.na(casted$sex)),]$sex <- 0

# remove any duplicated individuals caused by birthweight and sex assignment.
casted <- casted[!duplicated(casted$id),]

NIH01pheno <- casted

print(paste('Prepared phenotype data for ', length(NIH01pheno[,1]), ' subjects.'))

# Now we can convert the dataset into the GenABEL raw format:
write.table(NIH01_SNP_wide_nospace, file = "GenABEL/NIH01_SNP_wide_nospace.illu", sep = "\t", row.names = F, quote = F, eol = "\n")
convert.snp.illumina(inf = "GenABEL/NIH01_SNP_wide_nospace.illu", out = "NIH01geno.raw", strand = "file")

write.table(NIH01pheno, file = "GenABEL/NIH01_phenotypes.dat", sep = "\t", row.names = F, quote = F, eol = "\n")
print('Phenotype file formatted and saved as .dat file')

df <- load.gwaa.data(phe = "GenABEL/NIH01_phenotypes.dat", gen = "GenABEL/NIH01geno.raw", force = T, id = "id")

qc1 <- check.marker(df, maf = 0.05, p.lev = 0.05, ibs.threshold = 0.98)

summary(qc1)
data1 <- df[qc1$idok, qc1$snpok]
data1 <- Xfix(data1)
attach(phdata(data1))
descriptives.marker(data1)[2]
data1.gkin <- ibs(data1[, autosomal(data1)], weight = "freq")
data1.dist <- as.dist(0.5 - data1.gkin)
data1.mds <- cmdscale(data1.dist)
jpeg("results/PPplot.jpeg", width = 300)
plot(data1.mds, main = "")
dev.off()
km <- kmeans(data1.mds, centers = 2, nstart = 1000)
cl1 <- names(which(km$cluster == 1))
cl2 <- names(which(km$cluster == 2))
if (length(cl1) > length(cl2)) {
  x <- cl2
  cl2 <- cl1
  cl1 <- x
}
# Define what the relatedness groupings are.
c11
cohort[match(c11, cohort$Subject), c(2, 7, 8, 9)]
table(cohort[match(c11, cohort$Subject), c(8, 9)])
c12
cohort[match(c12, cohort$Subject), c(2, 7, 8, 9)]
table(cohort[match(c12, cohort$Subject), c(8, 9)])

# Select only the main genetically similar group.
data2 <- data1[c12, ]
qc2 <- check.marker(data2, fdr=0.2)
data2 <- data2[qc2$idok, qc2$snpok]

# This quite nicely creates a dummy code for the two groups by autosomal principal
# component analysis.
pop <- as.numeric(idnames(data1) %in% c11)

dobinfo <- as.numeric(phdata(data1)$dob)
weightinfo <- as.numeric(phdata(data1)$weight)

jpeg("popinfo.jpeg", width=500)
par(mfcol=c(2,1))
hist(dobinfo/365, col="lightblue", breaks=40, xlab="Years from first sample at
collection.", main="Histogram of date of DOB distribution in NIH01 cohort SNP correlation
subset.")

hist(weightinfo, col="lightblue", breaks=40, xlab="Birth weight in grams.", main="Histogram
of birth weight NIH01 cohort SNP correlation subset")
dev.off()

result <- qtscore(
  IFN.g_3M.003
  ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result, sort="Pc1df", top=25), file='results/IFNg_3M003.csv', sep=" ",
  eol="\n", append=T, col.names=F)

result <- qtscore(
  IFN.g_CpG.A
  ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
result <- qtscore(
  IFN.g_LPS
  ~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IFNg_LPS.csv',sep="","eol="\n",append=T, col.names=F)
write.table(cbind("IFN.g_LPS",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
  IFN.g_PAM
  ~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IFNg_PAM.csv',sep="","eol="\n")
write.table(cbind("IFN.g_PAM",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
  IFN.g_pIC
  ~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IFNg_pIC.csv',sep="","eol="\n")
write.table(cbind("IFN.g_pIC",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
  IFNa2_3M.003
  ~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IFNa2_3M003.csv',sep="","eol="\n")
write.table(cbind("IFNa2_3M.003",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
  IFNa2_CpG.A
  ~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IFNa2_CpGA.csv',sep="","eol="\n")
write.table(cbind("IFNa2_CpG.A",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)
IFNa2_pIC
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IFNa2_pIC.csv',sep="",',
,col="\n")
write.table(cbind("IFNa2_pIC",results(result)),file='results/Aggregate.csv',sep="",',
,col="\n",append=T, col.names=F)

result <- qtscore(
IL.10_3M.003
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL.10_3M003.csv',sep="",',
,col="\n")
write.table(cbind("IL.10_3M.003",results(result)),file='results/Aggregate.csv',sep="",',
,col="\n",append=T, col.names=F)

result <- qtscore(
IL.10_CpG.A
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL10_CpGA.csv',sep="",',
,col="\n")
write.table(cbind("IL.10_CpG.A",results(result)),file='results/Aggregate.csv',sep="",',
,col="\n",append=T, col.names=F)

result <- qtscore(
IL.10_LPS
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL10_LPS.csv',sep="",',
,col="\n")
write.table(cbind("IL.10_LPS",results(result)),file='results/Aggregate.csv',sep="",',
,col="\n",append=T, col.names=F)

result <- qtscore(
IL.10_PAM
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL10_PAM.csv',sep="",',
,col="\n")
write.table(cbind("IL.10_PAM",results(result)),file='results/Aggregate.csv',sep="",',
,col="\n",append=T, col.names=F)

result <- qtscore(
IL.10_pIC
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL10_pIC.csv',sep="",',
,col="\n")
write.table(cbind("IL.10_pIC",results(result)),file='results/Aggregate.csv',sep="",',
,col="\n",append=T, col.names=F)
result <- qtscore(
IL.12p40_3M.003
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL12p40_3M.003.csv',
sep="",eol="\n")
write.table(cbind("IL.12p40_3M.003",results(result)),file='results/Aggregate.csv',sep="",eol=
"\n",append=T, col.names=F)

result <- qtscore(
IL.12p40_CpG.A
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL12p40_CpG.A.csv',sep="",eol="\n")
write.table(cbind("IL.12p40_CpG.A",results(result)),file='results/Aggregate.csv',sep="",eol="\n",append=T, col.names=F)

result <- qtscore(
IL.12p40_LPS
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL12p40_LPS.csv',sep="",eol="\n")
write.table(cbind("IL.12p40_LPS",results(result)),file='results/Aggregate.csv',sep="",eol="\n",append=T, col.names=F)

result <- qtscore(
IL.12p40_PAM
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL12p40_PAM.csv',sep="",eol="\n")
write.table(cbind("IL.12p40_PAM",results(result)),file='results/Aggregate.csv',sep="",eol="\n",append=T, col.names=F)

result <- qtscore(
IL.12p40_pIC
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL.12p40_pIC.csv',sep="",eol="\n")
write.table(cbind("IL.12p40_pIC",results(result)),file='results/Aggregate.csv',sep="",eol="\n",append=T, col.names=F)

result <- qtscore(
IL.12p70_3M.003
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL.12p70_3M003.csv',
sep="",eol="\n")
write.table(cbind("IL.12p70_3M.003", results(result)), file='results/Aggregate.csv', sep='", eol="\n", append=T, col.names=F)

result <- qtscore(
    IL.12p70_CpG.A ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result, sort="Pc1df", top=25), file='results/IL12p70_CpG.A.csv', sep="","eol="\n")
write.table(cbind("IL.12p70_CpG.A", results(result)), file='results/Aggregate.csv', sep="","eol="\n", append=T, col.names=F)

result <- qtscore(
    IL.12p70_LPS ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result, sort="Pc1df", top=25), file='results/IL12p70_LPS.csv', sep="","eol="\n")
write.table(cbind("IL.12p70_LPS", results(result)), file='results/Aggregate.csv', sep="","eol="\n", append=T, col.names=F)

result <- qtscore(
    IL.12p70_pIC ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result, sort="Pc1df", top=25), file='results/IL12p70_pIC.csv', sep="","eol="\n")
write.table(cbind("IL.12p70_pIC", results(result)), file='results/Aggregate.csv', sep="","eol="\n", append=T, col.names=F)

result <- qtscore(
    IL.1b_3M.003 ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result, sort="Pc1df", top=25), file='results/IL1b_3M003.csv', sep="","eol="\n")
write.table(cbind("IL.1b_3M.003", results(result)), file='results/Aggregate.csv', sep="","eol="\n", append=T, col.names=F)

result <- qtscore(
    IL.1b_CpG.A ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result, sort="Pc1df", top=25), file='results/IL1b_CpGA.csv', sep="","eol="\n")
write.table(cbind("IL.1b_CpG.A", results(result)), file='results/Aggregate.csv', sep="","eol="\n", append=T, col.names=F)

result <- qtscore(
    IL.1b_LPS ~ sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL1b_LPS.csv',sep="\n",eol="\n")
write.table(cbind("IL.1b_LPS",results(result)),file='results/Aggregate.csv',sep="\n",eol="\n",append=T, col.names=F)

result <- qtscore( IL.1b_PAM
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL1b_PAM.csv',sep="\n",eol="\n")
write.table(cbind("IL.1b_PAM",results(result)),file='results/Aggregate.csv',sep="\n",eol="\n",append=T, col.names=F)

result <- qtscore( IL.1b_pIC
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL1b_pIC.csv',sep="\n",eol="\n")
write.table(cbind("IL.1b_pIC",results(result)),file='results/Aggregate.csv',sep="\n",eol="\n",append=T, col.names=F)

result <- qtscore( IL.23_3M.003
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL23_3M003.csv',sep="\n",eol="\n")
write.table(cbind("IL.23_3M.003",results(result)),file='results/Aggregate.csv',sep="\n",eol="\n",append=T, col.names=F)

result <- qtscore( IL.23_PAM
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL23_PAM.csv',sep="\n",eol="\n")
write.table(cbind("IL.23_PAM",results(result)),file='results/Aggregate.csv',sep="\n",eol="\n",append=T, col.names=F)

result <- qtscore( IL.6_3M.003
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL6_3M003.csv',sep="\n",eol="\n")
write.table(cbind("IL.6_3M.003",results(result)),file='results/Aggregate.csv',sep="\n",eol="\n",append=T, col.names=F)

result <- qtscore(
IL.6_CpG.A
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL6_CpGA.csv',sep="","eol="\n")
write.table(cbind("IL.6_CpG.A",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
IL.6_LPS
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL6_LPS.csv',sep="","eol="\n")
write.table(cbind("IL.6_LPS",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
IL.6_PAM
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL6_PAM.csv',sep="","eol="\n")
write.table(cbind("IL.6_PAM",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
IL.6_pIC
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/IL6_pIC.csv',sep="","eol="\n")
write.table(cbind("IL.6_pIC",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
TNFa_3M.003
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/TNFa_3M003.csv',sep="","eol="\n")
write.table(cbind("TNFa_3M.003",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
TNFa_CpG.A
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/TNFa_CpGA.csv',sep="","eol="\n")
write.table(cbind("TNFa_CpG.A",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)
result <- qtscore(
TNFa_LPS
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/TNFa_LPS.csv',sep="","eol="\n")
write.table(cbind("TNFa_LPS",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
TNFa_PAM
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/TNFa_PAM.csv',sep="","eol="\n")
write.table(cbind("TNFa_PAM",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

result <- qtscore(
TNFa_pIC
~sex, data1, times = 200, quiet = F, trait = "gaussian", strata=pop)
write.table(descriptives.scan(result,sort="Pc1df",top=25),file='results/TNFa_pIC.csv',sep="","eol="\n")
write.table(cbind("TNFa_pIC",results(result)),file='results/Aggregate.csv',sep="","eol="\n",append=T, col.names=F)

A.3 R program to plot statistically significant results and perform power calculations.

#Plotting program written for NIH01 project by Nathan Corbett

if(!require('car')) install.packages('car')
if(!require('beeswarm')) install.packages('beeswarm')
if(!require('VGAM')) install.packages('VGAM')
if(!require('SNPassoc')) install.packages('SNPassoc')
if(!require('plyr')) install.packages('plyr')
if(!require('wvioplot')) install.packages('wvioplot')

require('car')
#http://cran.r-project.org/web/packages/car/car.pdf
require('beeswarm')
require('VGAM')
require('SNPassoc')
require('plyr')
#require('wvioplot')
#VARS TO SET: current working location and the number of SNPS you want to look at.

datafile<- 'NIH01_complete_justified.RData'
snpfile <- 'results/Aggregate.csv'

if(exists('NIH01'))
{
  print('Dataset previously loaded')
} else {
  datadir <- getwd()
  load(datafile)
  NIH01 <- lumigene
  print(paste('Dataset ',datafile,' loaded.',sep=''))
  #Excluding un-stim values, as they are too low to consider significant, which the non-parametric test could show.
  NIH01 <- subset(NIH01, Year !='Adult' & Treatment !='Unstim')
}

snp.by.sig <- read.table(snpfile,header=T,sep=',', as.is=T)
snp.by.sig$Test_p_value <- as.numeric(snp.by.sig$Pc1df)
snp.by.sig <- snp.by.sig[which(!is.na(snp.by.sig$Test_p_value)),,]
.snp.by.sig <- snp.by.sig[order(snp.by.sig$Test_p_value),,]
snp.by.sig$Test_p_value <- round(snp.by.sig$Test_p_value, digits = 4)
snp.by.sig <- subset(snp.by.sig, snp.by.sig$Test_p_value <= 0.1)
sig.snp <- snp.by.sig
colnames(sig.snp)[1]<-'SNP_Name'
# sig.snp$Cytokine <- gsub('[.][','-',sig.snp$Cytokine)
# sig.snp$Treatment <- gsub('[.][','-',sig.snp$Treatment)
x<- length(sig.snp$Test_p_value)

NIH01$Gene <- sub(' ','',NIH01$Gene)
NIH01 <- NIH01[!is.na(NIH01$Accepted_Value),,]
NIH01 <- NIH01[!is.na(NIH01$Genotype),,]
NIH02 <- NIH01
NIH01 <- NIH01[which(NIH01$Year==0 | NIH01$Year=='0'),,]

print('Dataset subsetted to significant SNPs only.')

if(length(as.numeric(NIH01$Accepted_Value))<2)
date <- Sys.Date()
system(paste('mkdir Plots/', date, sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots', sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots/Normal', sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots/Log-trans-Normal', sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots/Chisq', sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots/Exponential', sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots/Logistic', sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots/Laplace', sep='', wait=T))
system(paste('mkdir Plots/', date, '/QQplots/Versus', sep='', wait=T))
system(paste('mkdir Plots/', date, '/Boxplots', sep='', wait=T))
system(paste('mkdir Plots/', date, '/Eth-Scatter-Boxplots', sep='', wait=T))
system(paste('mkdir Plots/', date, '/Beeplots', sep='', wait=T))
system(paste('mkdir Plots/', date, '/Stripchart', sep='', wait=T))
system(paste('mkdir results/Effect_stats/', sep='', wait=T))

Statplots <- function(j, NIH01, sig.snp, x) {
  require('SNPassoc')
  require('plyr')
  attach(sig.snp)

  #Parsing the dataset into unit of analysis groups.
  cyt    <- Cytokine[j]
  trt    <- Treatment[j]
  snp    <- SNP_Name[j]
  p.val  <- Test_p_value[j]
  chr    <- Chromosome[j]
  pos    <- Position[j]
  strand <- Strand[j]
  allele1 <- A1[j]
  allele2 <- A2[j]

  temp    <- subset(NIH01, Cytokine==cyt & Treatment==trt & SNP_Name==snp)

  #subset to relevant year for subsequent plots
gene <- paste(unique(temp$Gene), collapse = '')
sig.snp.temp <- subset(sig.snp, Cytokine==cyt & Treatment==trt & SNP_Name==snp)
temp <- merge(temp, sig.snp.temp, all.x=T)
#converting unordered genotypes into ordered factor.
#remove duplications of individual subjects.
    temp <- temp[!duplicated(temp$Subject),]
#calculate genotype frequencies for each SNP within the population for which data is available.
    tempSNP <- snp(temp$Genotype, sep=' ',
                   allow.partial.missing=F, remove.spaces=F)
    snpinfo <- summary(tempSNP)
    MAF <- min(snpinfo$allele.freq[,2])/100
    HWEp <- snpinfo$HWE
    tempSNP <- additive(tempSNP)

    #add a new column to temp that includes numerical codes for the genotypes
    where 0 is the most common homozygous case, 1 is the heterozygous case, and 2 is the least
    common homozygous case
    temp$Geno.code <- tempSNP
    N = length(temp$Subject)

    code0 <- subset(temp, Geno.code == 0)
    codenum0 <- code0$Geno.code[1]
    geno0 <- code0$Genotype[1]
    medcode0 <- median(code0$Accepted_Value)
    meancode0 <- mean(code0$Accepted_Value)
    stdev0 <- sd(code0$Accepted_Value)

    code1 <- subset(temp, Geno.code == 1)
    codenum1 <- code1$Geno.code[1]
    geno1 <- code1$Genotype[1]
    medcode1 <- median(code1$Accepted_Value)
    meancode1 <- mean(code1$Accepted_Value)
    stdev1 <- sd(code1$Accepted_Value)

    code2 <- subset(temp, Geno.code == 2)
    codenum2 <- code2$Geno.code[1]
    geno2 <- code2$Genotype[1]
    medcode2 <- median(code2$Accepted_Value)
    meancode2 <- mean(code2$Accepted_Value)
    stdev2 <- sd(code2$Accepted_Value)

    codes <- list(code0, code1, code2)

    n0 <- length(code0[,6])
    n1 <- length(code1[,6])

103
n2 <- length(code2[,6])

if(n1<1)
{
  temp <- subset(temp, Geno.code == 0 | Geno.code==2)
genotype <-c(geno0,geno2)
genocode<-c(0,2)
} else{
  if(n2<1)
  {
    temp <- subset(temp, Geno.code ==0 |
    Geno.code==1)
genotype <- c(geno0,geno1)
genocode <- c(0,1)
  } else{
    genotype <-
    c(geno0,geno1,geno2)
genocode<-c(0,1,2)
  }
}

#Effect size calculation
require(pwr)
E.data <- subset(temp, Geno.code == genocode[2])
E.data <- E.data$Accepted_Value
C.data <- subset(temp, Geno.code == genocode[1])
C.data <- C.data$Accepted_Value

se.d.raw1<-function(E.data,C.data){
n1<-length(C.data)
n2<-length(E.data)
d<-(mean(E.data)-mean(C.data))/
  sqrt(((n2-1)*var(E.data)+(n1-1)*var(C.data))/(n1+n2-2))
  se<-sqrt(((n1+n2-1)/(n1+n2-3))*((4/(n1+n2))*(1+(d^2/8))))
  names(se)<-'se.d'
  names(d)<-'effect'
  return(c(se,d))
}

sesame <- se.d.raw1(E.data,C.data)

test <- pwr.t.test(d = sesame[2], sig.level = 0.05, power = 0.8, type = c('two.sample'))

effect.results<-
cbind(cyt, trt, snp, genocode[1], length(E.data), mean(E.data), sd(E.data), genocode[2], length(C.data), mean(C.data), sd(C.data), sesame[2], sesame[1], test$n, MAF, HWEp)
colnames(effect.results) <- c('Cytokine','Treatment','SNP','Common allele','N common','Mean common','Stdev common','Het or rare','N het/rare','Mean het/rare','Stdev het/rare','Effect size d','se of d','1/2 n required for confirmation','Minor allele frequency','Hardy-Weinberg equilibrium p-value')

#Create table of effect size results
write.csv(effect.results, file=paste('results/Effect_stats/NIH01_top- 'x,'_SNPs_Effect_Size.csv', sep=''))

#Create jpeg device to write to.
jpeg(file=paste('Plots/',date,'/Beeplots/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NIH01_top- 'x,'_SNPs.png', sep=''),height=20,width=40,res=300,units='cm')

#Create space for legend.
plot.new()
par(oma=c(2,0,0,2))

jpeg(file=paste('Plots/',date,'/Beeplots/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NIH01_top- 'x,'_SNPs.png', sep=''),height=20,width=40,res=300,units='cm')

#Thanks to: http://www.r-statistics.com/2011/03/beeswarm-boxplot-and-plotting-it-with-r/ for this idea:
beeswarm(Accepted_Value~Geno.code, data=temp,
 method = 'swarm',
 pch = 16,
 cex=1.5,
 xlab = 'Genotype (top strand)', ylab = paste('pg/mL',cyt, 'in supernatant.'),
 main=paste(cyt,trt,snp,gene,'N=',N,'p=',p.val,sep=''),
 labels = genotype)
if(n2 == 0)
{
 mtext(paste('N of',geno0,' =',n0,'  N of ',geno1,' =',n1,sep=' '),side=1, line=4)
} else{
if(n1 == 0)
{
 mtext(paste('N of',geno0,' =',n0,'  N of ',geno2,' =',n2,sep=' '),side=1, line=4)
} else{
 mtext(paste('N of',geno0,' =',n0,'  N of ',geno1,' =',n1,'  N of ',geno2,' =',n2,sep=' '),side=1, line=4)
}

boxplot(Accepted_Value ~ Geno.code, data = temp, add = T, names = genotype,
 col='#0000ff22')
# my thanks goes to Greg Snow for the tip on the transparency colour (from 2007):
dev.off()

jpeg(file=paste('Plots/',date,'/Eth-Scatter-Boxplots/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NIH01_top-',x,'_SNPs.png', sep=''),height=20,width=40,res=300,units='cm')

plot.new()
par(oma=c(2,0,0,2))
layout(matrix(c(1,2), nrow = 1), widths = c(0.7, 0.3))

par(xpd=F)

par(mar=c(4,6,4,0))
opar <- par(las=1)

#Write a boxplot for each condition.

boxplot(Accepted_Value~Geno.code, data=temp, names=genotype, xlab='Genotype (top strand)', ylab=' ', varwidth=F, outline=F, vertical=T, main=paste(cyt,trt,snp,gene,'N=',N,'p=',p.val,sep=' '), add=F)

par(opar)
title( ylab='pg/mL cytokine in supernatant.', line=4)

if(length(genotype)>2)
{
  jpeg(file=paste('Plots/',date,'/Boxplots/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NIH01_top-',x,'_SNPs.png', sep=''),height=20,width=40,res=300,units='cm')
} else {
  jpeg(file=paste('Plots/',date,'/Boxplots/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NIH01_top-',x,'_SNPs.png', sep=''),height=20,width=20,res=300,units='cm')
}

#Create space for legend.

plot.new()
par(oma=c(2,0,0,2))

par(xpd=F)
par(mar=c(4,6,4,0))
opar <- par(las=1)

#Write a boxplot for each condition.
boxplot(temp$Accepted_Value~temp$Geno.code,
names=genotype,
xlab='Genotype (top strand)',
ylab='',
varwidth=F,
method='jitter',
jitter=0.2,
outline=F,
vertical=T,
main=paste(cyt,trt,snp,gene,'N=',N,'p=',p.val,sep=' '),
add=F)

par(opar)
title( ylab='pg/mL cytokine in supernatant.', line=4)

if(n2 == 0)
  {mtext(paste('N of',geno0,' =',n0,'  N of ',geno1,'=','n1,sep='''),side=1, line=4)}
else{
  if(n1 == 0)
    {mtext(paste('N of ',geno0,' =',n0,'  N of ',geno2,'=','n2,sep='''),side=1, line=4)}
  else{
    mtext(paste('N of ',geno0,' =',n0,'  N of ',geno1,'=','n1,' N of ',geno2,'=','n2, sep=' '),side=1, line=4)
  }
}
dev.off()

#Thanks to: http://www.r-bloggers.com/the-many-uses-of-q-q-plots/ for nicely describing the various methods available.
#Thanks to http://blog.revolutionanalytics.com/2010/08/distributions-in-r.html for describing the base distributions in R.

jpeg(file=paste('Plots/',date,'/QQplots/Normal/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NI H01_top-',x,'_SNPs.png', sep=''),height=20,width=40,res=300,units='cm')

if(length(genocode) > 2)
  {
    par(mfrow=c(1,3))
  }
for (v in 1:length(genocode))
{
  geno <- genocode[v]
  geno.real <- geno+1
  geno.group <- codes[[geno.real]]
  this.genotype <- genotype[[geno.real]]
  qqPlot(geno.group$Accepted_Value,
         distribution='norm',
         ylab=paste(this.genotype, 'population', cyt, 'response to',trt,'quantiles',sep=')',
         xlab='Normal distribution quantiles',
         main=paste(gene,snp,this.genotype,sep=')'))

  dev.off()

  jpeg(file=paste('Plots/',date,'/QQplots/Log-trans-Normal/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NIH01_top-',x,'_SNPs.png',
                  sep=''),height=20,width=40,res=300,units='cm')

  if(length(genocode) > 2)
  {
    par(mfrow=c(1,3))
  } else {
    par(mfrow=c(1,2))
  }

  for (v in 1:length(genocode))
  {
    geno <- genocode[v]
    geno.real <- geno+1
    geno.group <- codes[[geno.real]]
    this.genotype <- genotype[[geno.real]]
    qqPlot(log10(geno.group$Accepted_Value+1),
           distribution='norm',
           ylab=paste(this.genotype, 'population', cyt, 'response to',trt,'quantiles',sep=')',
           xlab='Normal distribution quantiles',
           main=paste(gene,snp,this.genotype,sep=')'))

    dev.off()

    jpeg(file=paste('Plots/',date,'/QQplots/Chisq/',gene,'_',p.val,'_',cyt,'_',trt,'_',snp,'_NIH01_top-',x,'_SNPs.png',
                    sep=''),height=20,width=40,res=300,units='cm')
if(length(genocode) > 2)
{
par(mfrow=c(1,3))
} else {
par(mfrow=c(1,2))
}

for (v in 1:length(genocode))
{
    geno <- genocode[v]
geno.real <- geno+1
geno.group <- codes[[geno.real]]
this.genotype <- genotype[[geno.real]]
qqPlot(geno.group$Accepted_Value, distribution='chisq', df=length(genocode)-1,
ylab=paste(this.genotype, 'population', cyt, 'response to', trt, 'quantiles', sep=' '), xlab=paste('Chi-squared distribution quantiles df=', length(genocode)-1), main=paste(gene, snp, this.genotype, sep=' '))
}

dev.off()

jpeg(file=paste('Plots/', date, '/QQplots/Exponential/', gene, '_', p.val, '_', cyt, '_', trt, '_', snp, '_NIH01_top-', x, '_SNPs.png', sep=''), height=20, width=40, res=300, units='cm')

if(length(genocode) > 2)
{
par(mfrow=c(1,3))
} else {
par(mfrow=c(1,2))
}

for (v in 1:length(genocode))
{
    geno <- genocode[v]
geno.real <- geno+1
geno.group <- codes[[geno.real]]
this.genotype <- genotype[[geno.real]]
qqPlot(geno.group$Accepted_Value, distribution='exp', df=length(genocode)-1,
ylab=paste(this.genotype, 'population', cyt, 'response to', trt, 'quantiles', sep=' '), xlab=paste('Chi-squared distribution quantiles df=', length(genocode)-1), main=paste(gene, snp, this.genotype, sep=' '))
}
dev.off()

jpeg(file=paste('Plots/',date,'/QQplots/Logistic/',gene,' .p.val',' .',cyt,' .',trt,' .',snp,'_NIH01_to p- ',x,'_SNPs.png', sep=''),height=20,width=40,res=300,units='cm')

if(length(genocode) > 2)
{
  par(mfrow=c(1,3))
} else {
  par(mfrow=c(1,2))
}
for (v in 1:length(genocode))
{
  geno <- genocode[v]
  geno.real <- geno+1
  geno.group <- codes[[geno.real]]
  this.genotype <- genotype[[geno.real]]
  qqPlot(geno.group$Accepted_Value,
        ylab=paste(this.genotype, 'population', cyt, 'response to',trt,'quantiles',sep=' '),
        xlab='Exponential distribution quantiles',
        main=paste(gene,snp,this.genotype,sep=' '))
}

dev.off()

#When you can define the p arg of the exponential distribution:
#When p = 2 the exponential power distribution becomes the Normal Distribution, when p = 1 the exponential power distribution becomes the Laplace Distribution, when p → ∞ the exponential power distribution becomes the Uniform Distribution.

jpeg(file=paste('Plots/',date,'/QQplots/Laplace/',gene,' .p.val',' .',cyt,' .',trt,' .',snp,'_NIH01_top - ',x,'_SNPs.png', sep=''),height=20,width=40,res=300,units='cm')

if(length(genocode) > 2)
{
  par(mfrow=c(1,3))
} else {
  par(mfrow=c(1,2))
}
for (v in 1:length(genocode))
{
  geno <- genocode[v]
geno.real <- geno+1
geno.group <- codes[[geno.real]]
this.genotype <- genotype[[geno.real]]
qqPlot(geno.group$Accepted_Value,
distribution='laplace',
ylab=paste(this.genotype, 'population', cyt, 'response to', trt, 'quantiles', sep=' '), xlab='Laplace distribution quantiles',
main=paste(gene, snp, this.genotype, sep=' '))
dev.off()

if(length(genocode) > 2)
{
    jpeg(file=paste('Plots/', date, '/QQplots/Versus/', gene, '_', p.val, '_', cyt, '_', trt, '_', snp, '_NIH01_top-', x, '_SNPs.png', sep=''), height=20, width=40, res=300, units='cm')
    par(mfrow=c(1,3))
} else {
    jpeg(file=paste('Plots/', date, '/QQplots/Versus/', gene, '_', p.val, '_', cyt, '_', trt, '_', snp, '_NIH01_top-', x, '_SNPs.png', sep=''), height=20, width=20, res=300, units='cm')
    par(mfrow=c(1,1))
}

for (v in 1:length(genocode))
{
    for(w in 1:length(genocode))
    {
        if(v==w)
        {next}
        if(v==2 & w==1)
        {next}
        if(v==3)
        {next}

        genofoo1 <- genocode[v]
        genofoo2 <- genocode[w]
        geno1.real <- genofoo1+1
        geno2.real <- genofoo2+1
        genofoo1.group <- codes[[geno1.real]]
        genofoo2.group <- codes[[geno2.real]]
        this.genotype1 <- genotype[[geno1.real]]
        this.genotype2 <- genotype[[geno2.real]]

        qqplot(genofoo1.group$Accepted_Value, genofoo2.group$Accepted_Value, plot.it=T,
ylab=paste(this.genotype1, 'population', cyt, 'response to', trt, 'quantiles', sep=' '),
xlab=paste(this.genotype2, 'population', cyt, 'response to', trt, 'quantiles', sep=' '),
main=paste(gene, snp, this.genotype, sep=' '))
abline(0,1)
}
}

dev.off()

print(paste(x-j, ' snp plots remain.'))
return(effect.results)
detach(sig.snp)
}
effect.results <- sapply(1:x, Statplots, NIH01, sig.snp, x)
print("statistics results plotting complete."
)
write.table(effect.results, file=paste('results/effects.csv'), col.names=F, row.names=F, append=T, sep=';', quote=F)

setwd(datadir)
print('Done. Have a nice day.')
## Appendix B

### B.1 Statistical analysis of cytokine response cohort

Table 2. Comparison of the concentrations of secreted cytokines for longitudinal neonatal and adult responses to cell-surface TLR stimulation.

Cytokines secreted MC in tissue culture media containing the indicated TLR ligands were determined in pg/ml for 35 neonate-infant subjects and 25 adult subjects. Statistically significant (p<0.01) differences between age groups are indicated in bold type-set. NP = not performed.

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>year</th>
<th>Interleukin</th>
<th>pAM</th>
<th>pIC</th>
<th>LPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNα2</td>
<td>2</td>
<td>A</td>
<td>66.9 ± 164.3</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>9.8 ± 13.7</td>
<td>58.1 ± 74.2</td>
<td>2</td>
<td>9.7 ± 14</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>11.2 ± 22.6</td>
<td>106.7 ± 97.3</td>
<td>1</td>
<td>9.1 ± 15</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>19.4 ± 16.4</td>
<td>50.0 ± 300</td>
<td>2</td>
<td>14 ± 23</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>1</td>
<td>111.2 ± 229.5</td>
<td>0.0032</td>
<td>0.0152</td>
<td>0.00088</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>39.8 ± 87.3</td>
<td>177.6 ± 285</td>
<td>2</td>
<td>152.3 ± 260.8</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>59.7 ± 169.2</td>
<td>57.3 ± 92.6</td>
<td>1</td>
<td>0.0036 &lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>38.4 ± 33.3</td>
<td>0.3720</td>
<td>0.7624</td>
<td>0.00141</td>
</tr>
<tr>
<td>IL-12p70</td>
<td>2</td>
<td>9.9 ± 1.11</td>
<td>&lt; 0.0001</td>
<td>0.0072</td>
<td>0.0872</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3.7 ± 1.3</td>
<td>0.0003</td>
<td>0.7145</td>
<td>33.1 ± 6.3</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.3 ± 0.3</td>
<td>0.0055</td>
<td>0.0001</td>
<td>8.1 ± 1.4</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>2</td>
<td>22.3 ± 23.6</td>
<td>&lt; 0.0001</td>
<td>0.0640</td>
<td>0.00218</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>54.1 ± 58.2</td>
<td>35.7 ± 56.9</td>
<td>1</td>
<td>0.0001 &lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>241.9 ± 369.6</td>
<td>0.00019</td>
<td>0.1042</td>
<td>0.00099</td>
</tr>
<tr>
<td>IL-23</td>
<td>2</td>
<td>20.3 ± 50.8</td>
<td>&lt; 0.0001</td>
<td>NP</td>
<td>156.1 ± 147.9</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8.6 ± 23.3</td>
<td>0.0001</td>
<td>0.0100</td>
<td>0.00009</td>
</tr>
<tr>
<td>IL-6</td>
<td>2</td>
<td>1704.1 ± 4217.5</td>
<td>0.0005</td>
<td>497.3 ± 1454.3</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3064.7 ± 2991.4</td>
<td>0.0001</td>
<td>506.8 ± 306.3</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>5962.1 ± 6532.8</td>
<td>0.0001</td>
<td>0.00071</td>
<td>0.001</td>
</tr>
<tr>
<td>TNFα</td>
<td>2</td>
<td>1447.5 ± 1652.2</td>
<td>0.261</td>
<td>513.1</td>
<td>2611.6 ± 145.1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2215.4 ± 2479.5</td>
<td>0.0286</td>
<td>325.7 ± 322.2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1749.8 ± 1091.5</td>
<td>0.0052</td>
<td>0.0472</td>
<td>0.00064</td>
</tr>
<tr>
<td>IL-10</td>
<td>2</td>
<td>2775.5 ± 1066.6</td>
<td>375.4 ± 255.9</td>
<td>2</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3765.5 ± 833.5</td>
<td>0.0220</td>
<td>0.0196</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2591.7 ± 330.4</td>
<td>0.0003</td>
<td>0.1095</td>
<td>0.0001</td>
</tr>
<tr>
<td>IL-10</td>
<td>2</td>
<td>46.4 ± 10.7</td>
<td>0.0167</td>
<td>9.8 ± 1.9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>97 ± 14.7</td>
<td>0.0041</td>
<td>0.0316</td>
<td>0.3320</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>107.5 ± 15.2</td>
<td>0.05174</td>
<td>0.00012</td>
<td>0.00037</td>
</tr>
</tbody>
</table>

113
Table 3. Comparison of the concentrations of secreted cytokines for longitudinal neonatal and adult responses to endosomal TLR stimulation.

Cytokines secreted MC in tissue culture media containing the indicated TLR ligands were determined in pg/ml for 35 neonate-infant subjects and 25 adult subjects. Statistically significant (p<0.01) differences between age groups are indicated in bold type-set. NP = not performed.

<table>
<thead>
<tr>
<th>Cytokine</th>
<th>3M-003</th>
<th>A</th>
<th>CpG-A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNα2</td>
<td>516 ± 468</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1964 ± 107</td>
<td>1</td>
<td>0.0180</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1874 ± 1078</td>
<td>1</td>
<td>0.00001</td>
<td>1</td>
</tr>
<tr>
<td>IFNγ</td>
<td>4043.4 ± 3155.5</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1645.6 ± 1748.1</td>
<td>1</td>
<td>0.00001</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1236.3 ± 233.3</td>
<td>1</td>
<td>0.00001</td>
<td>1</td>
</tr>
<tr>
<td>IL-12b70</td>
<td>103.1 ± 28.1</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1236.3 ± 233.3</td>
<td>1</td>
<td>0.00001</td>
<td>1</td>
</tr>
<tr>
<td>IL-12p40</td>
<td>5886.6 ± 368.4</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td>IL-23</td>
<td>74.9 ± 64.1</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td>IL-6</td>
<td>43930 ± 9880.5</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td>TNFα</td>
<td>3578.4 ± 947.4</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1645.6 ± 1748.1</td>
<td>1</td>
<td>0.00001</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1236.3 ± 233.3</td>
<td>1</td>
<td>0.00001</td>
<td>1</td>
</tr>
<tr>
<td>IL-1b</td>
<td>19210.9 ± 4112.2</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td>IL-10</td>
<td>88.4 ± 21</td>
<td>2</td>
<td>0.00001</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1964 ± 107</td>
<td>1</td>
<td>0.0180</td>
<td>0</td>
</tr>
</tbody>
</table>

114
Appendix C

Genotyping supplemental information.

C.1 GVS haplotype testing rs10751209

Chromosome 11: 70059105 - 70059105

Total chromosome span: 70049105 - 70069105

Genes in this region: FADD

Allele Frequency Cutoff (%): 0, monomorphic sites excluded

Data Merging: combined samples with combined variations

Population: HapMap-CEU(unrelated only), Submitter: CSHL-HAPMAP

Population: HapMap-HCB, Submitter: CSHL-HAPMAP

Table 4. Pairwise haplotype LD testing for 10Kb up and downstream of rs10751209, filtered for target SNP. Relevant color code: Orange = UTR SNP

<table>
<thead>
<tr>
<th>SNP-1</th>
<th>SNP-2</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs615363</td>
<td>rs10751209</td>
<td>0.245</td>
</tr>
<tr>
<td>rs1131715</td>
<td>rs10751209</td>
<td>0.146</td>
</tr>
<tr>
<td>rs3740720</td>
<td>rs10751209</td>
<td>0.375</td>
</tr>
<tr>
<td>rs7105001</td>
<td>rs10751209</td>
<td>0.136</td>
</tr>
<tr>
<td>rs7939734</td>
<td>rs10751209</td>
<td>0.144</td>
</tr>
<tr>
<td>rs10751209</td>
<td>rs10793035</td>
<td>0.130</td>
</tr>
<tr>
<td>rs10751209</td>
<td>rs11235564</td>
<td>0.898</td>
</tr>
<tr>
<td>rs10751209</td>
<td>rs12419801</td>
<td>0.069</td>
</tr>
<tr>
<td>rs10751209</td>
<td>rs12786868</td>
<td>0.009</td>
</tr>
</tbody>
</table>
C.2  GVS haplotype testing rs2069830

Chromosome 7: 22767137 - 22767137

Total chromosome span: 22757137 - 22777137

Genes in this region: IL6

Allele Frequency Cutoff (%): 0 , monomorphic sites excluded

Data Merging: combined samples with combined variations

Population: HapMap-CEU(unrelated only), Submitter: CSHL-HAPMAP

Population: HapMap-HCB, Submitter: CSHL-HAPMAP

Rs2069830 did not appear in the resultant output. For full table, please contact the author.

C.3  GVS haplotype testing rs583911

Chromosome 3: 159710390 - 159710390

Total chromosome span: 159700390 - 159720390

Genes in this region: IL12A

Allele Frequency Cutoff (%): 0 , monomorphic sites excluded

Data Merging: combined samples with combined variations

Population: HapMap-HCB, Submitter: CSHL-HAPMAP

Population: HapMap-CEU(unrelated only), Submitter: CSHL-HAPMAP

Table 5. Pairwise haplotype LD testing for 10Kb up and downstream of rs583911, filtered for target SNP.

<table>
<thead>
<tr>
<th>SNP-1</th>
<th>SNP-2</th>
<th>r^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs583911</td>
<td>rs640039</td>
<td>0.189</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs662959</td>
<td>0.149</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs668998</td>
<td>1</td>
</tr>
</tbody>
</table>

Relevant color code: Orange = UTR SNP
<table>
<thead>
<tr>
<th>rs583911</th>
<th>rs890914</th>
<th>0.315</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs583911</td>
<td>rs2115176</td>
<td>0.692</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2133310</td>
<td>0.706</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243115</td>
<td>0.315</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243123</td>
<td>0.315</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243131</td>
<td>0.189</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243136</td>
<td>0.245</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243140</td>
<td>0.227</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243143</td>
<td>0.696</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243147</td>
<td>0.706</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243148</td>
<td>0.315</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2243151</td>
<td>0.706</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs2561284</td>
<td>0.189</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs4608735</td>
<td>0.315</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs4680536</td>
<td>0.008</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs6441283</td>
<td>0.588</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs6441284</td>
<td>0.706</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs11927521</td>
<td>0.692</td>
</tr>
<tr>
<td>rs583911</td>
<td>rs17826053</td>
<td>0.189</td>
</tr>
<tr>
<td>rs485497</td>
<td>rs583911</td>
<td>0.368</td>
</tr>
<tr>
<td><strong>rs568408</strong></td>
<td><strong>rs583911</strong></td>
<td><strong>0.26</strong></td>
</tr>
</tbody>
</table>