Functional characterization of human variants of *NFKBIA*: a key regulator of immune responsiveness implicated in susceptibility to infectious and inflammatory disease

by

Salman Ali

B.Sc., Simon Fraser University, 2007

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

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver)

April 2010

© Salman Ali, 2010

Abstract

IkBα is an important regulator of inflammation. Single nucleotide polymorphisms (SNPs) rs3138053, rs2233406 and rs2233409 in the promoter of the gene NFKBIA, which encodes for $IkB\alpha$, have been shown to be associated with a variety of infectious and inflammatory conditions. In this study, we investigated the functional impact of the promoter variants of NFKBIA on human immune responsiveness. Using a coding SNP that was in strong linkage disequilibrium (LD) with NFKBIA SNPs rs3138053/rs2233406/rs2233409, we designed and validated an allele-specific PCR assay that could detect subtle differences in allele ratios minor (GTT) between the major (ACC) and promoter variants of SNPs rs3138053/rs2233406/rs2233409. Peripheral blood mononuclear cells (PBMCs) of homozygous (ACC/ACC) and heterozygous (ACC/GTT) individuals were stimulated with 100ng/ml LPS and live cultures of Streptococcus pneumoniae (moi 7.8-30) serotype 14 for 3 and 4 hours. PBMCs of neonatal NFKBIA homozygotes and heterozygotes were stimulated with various Toll-likereceptor (TLR) ligands of the innate immunity cascade to assay for differences in the innate immune response.

NFKBIA heterozygotes of European descent displayed 1.21 (1.14-1.27 95% CI)-1.26 (1.18-1.34 95% CI) fold higher expression of the major allele transcript (ACC) relative to the minor allele transcript (GTT). For the same ethnicity, at 3 hours stimulation, *NFKBIA* homozygotes (ACC/ACC) produced higher levels of *NFKBIA* mRNA than heterozygotes following stimulation with LPS (1.4 fold. *p*=0.0095) and *S. pneumoniae* (1.51 fold, *p*=0.024). Higher TNF α secretion was seen from PBMCs of heterozygotes as compared to homozygotes (of European descent) in the presence of LPS (1.57 fold , *p*<0.05 at a dose of 100ng/ml), Pam3CSK4 (2.29 fold, *p*<0.01 at a dose of 100ng/ml and 1.91 fold, *p*<0.05 at a dose of 1000 ng/ml), 3M003 (1.79 fold, *p*<0.001 at a dose 10µM) and 3M002 (3.30 fold, *p*<0.001 at a dose of 10µM). The results presented here provide preliminary functional evidence behind the observed associations of these SNPs with infectious and inflammatory conditions. A global understanding of the functional consequences of regulatory polymorphisms of *NFKBIA* will be provided with subsequent experiments that examine differences between all *NFKBIA* genotypes (including minor allele homozygotes) for IkB α protein expression and NF- κ B translocation in all major ethnic groups.

Abstract	
Chapter 1: Introduction1	
1.1 From genotype to phenotype: Understanding functional consequences of genetic variation.1	
 1.2 The NF-κB pathway: Overview	
 1.3 The role of IκB proteins in regulating NF-κB activity4 1.3.1 Importance of IκBα in regulation of NF-κB4 1.3.2 Functional consequences of altered IκBα function in mammals5 	
1.4 Understanding linkage disequilibrium (LD) 5 1.4.1 Linkage disequilibrium (LD) 6 1.4.2 Using LD for association mapping 6 1.4.3 Using LD for genome wide association studies 6 1.4.4 Patterns of LD in different populations 7 1.4.5 Tag SNP coverage and transferability in different populations 7	
 1.5 Association of observed variation in the gene (<i>NFKBIA</i>) encoding IκBα with infection, autoimmunity/inflammation, and cancer	
1.6 Opposing associations of <i>NFKBIA</i> promoter polymorphisms with infection and autoimmunity: A genetic link to the hygiene hypothesis?12	
1.7 Validating regulatory variants and the genetics of gene expression 12 1.7.1 Cis and trans-variation 13	,
1.8 Approaches for validating <i>cis</i> -acting variants	
1.9 Hypothesis15	
Chapter 2: Material and methods22	
2.1 Analysis of linkage disequilibrium (LD)22	

Table of Contents

2.1.1 Source populations	22 22
2.2 Applysis of NEKRIA mDNA symposium, and ellelis impolence	22
2.2 Analysis of <i>NFKBIA</i> mRNA expression, and allelic impalance	23
2.2.2 Genomic DNA extraction	23
2.2.3 Optimization of bacterial growth	23
2.2.4 PBMC isolation	25
2.2.5 PBMC stimulation	25
2.2.7 Genotyping	25
2.2.8 NFKBIA gene expression	26
2.2.9 Genotyping for allelic imbalance	27
2.3 Neonatal innate immune response	27
2.4 Statistics	28
2.4.1 Statistical analysis for allelic imbalance	28
2.4.2 Statistical analysis for gene expression	28
2.4.3 Statistical analysis for innate immune response	28
2.5 Order of operations	28
Chapter 3: Results	33
3.1 Emerging picture of the NEKBIA promoter haplotype	33
3.1.1 Analysis of variants	33
3.1.2 Linkage disequilibrium(LD)analysis	33
3.1.3 Promoter haplotype analysis	33
3.2 Functional analysis of NFKBIA promoter variants	34
3.2.1 Putative transcription factor binding sites	34
3.2.2 Risk haplotype prediction	34
3.3 Allelic imbalance	35
3.3.1 Validation of allele specific PCR assay	35
3.3.2 Generation of allele ratios	35
3.3.3 Determining allelic imbalance in NEKBIA IS 1050851 heterozygotes	35
3.4 Determining impact of allelic imbalance on <i>NFKBIA</i> mRNA expression	36
3.4.1 Impact of allelic imbalance on NFKBIA mRNA expression in NFKBIA homozy	Jotes
3.4.2 Impact of allelic imbalance on NEKBIA mRNA expression in NEKBIA homozyc	notes
and heterozygotes of European descent	37
3.5 Determining impact of allelic imbalance on the innate immune response	37
3.5.1 Impact of allelic imbalance on the innate immune response in NFKBIA	-
homozygotes and heterozygotes	37
3.5.2 Impact of allelic imbalance on the innate immune response in NFKBIA	20
3.5.3 Impact of allelic imbalance on pathways that do not involve IkBg degradation	38 38

Chapter 4: Discussion
4.1 Overview
4.2 Allele specific PCR514.2.1 Choice of tagging SNP514.2.2 Precision of assay52
4.3 Choice of cell types
4.4 Analysis of variation in gene expression534.4.1 Ligand and pathogen specific differences534.4.2 Inter-individual differences534.4.3 Ethnic differences544.4.4 Artefactual differences55
4.5 Choice of ligand and time course selection
4.6 Choice of CpGA mediated TLR9 secretion of IFNα as a non-IkBα requiring pathway56
4.7 Follow up experiments: Need and feasibility
4.8 Clinical implication of dysregulated TNF α production in inflammatory/autoimmune disease 58
4.9 From bench to bedside: Diagnostic value of <i>NFKBIA</i> expression and <i>NFKBIA</i> polymorphisms in determining response to anti TNFα therapy
4.10 Conclusion
References
Appendix: Haplotypes and LD measure for the entire NFKBIA gene in 4 major ethnic groups70

List of Tables

19
29
39
41

List of Figures

Figure 1.1	Steps needed to verify if variation in a candidate gene	
	influences susceptibility to a complex disease	17
Figure 1.2	Validating impact of cis acting promoter variants on gene	
	expression	20
Figure 1.3	Hypothesis: Impact of minor variants in the NFKBIA promoter	
	on NF-кВ signaling	21
Figure 2.1	Streptococcus pneumoniae serotype 14 shows a reduced lag time for exponentia	
	growth when grown in media BHI media supplemented with 5%	
	FCS	30
Figure 2.2	Determining the optimal growth range for Streptococcus pneumonia	
	serotype 14	31
Figure 2.3	Overview of experimental protocol	32
Figure 3.1	Linkage disequilibrium (LD) plot of the NFKBIA promoter for the 4	
	HapMap populations	40
Figure 3.2	Visual representation of functional binding sites for transcription	
	factors of NFKBIA	42
Figure 3.3	Validating NFKBIA allele specific PCR assay	43
Figure 3.4	Standard curve for detecting allelic imbalance in NFKBIA	
	heterozygotes	44
Figure 3.5	NFKBIA heterozygotes (ACC/GTT) display allelic imbalance	45
Figure 3.6	NFKBIA heterozygotes display a trend of reduced NFKBIA	
	mRNA expression than NFKBIA homozygotes	46
Figure 3.7	NFKBIA mRNA expression is significantly reduced in NFKBIA	
	heterozygotes of European descent	47
Figure 3.8	NFBIA heterozygotes display a higher TNFα response	48
Figure 3.9	NFBIA heterozygotes of European origin also display a higher	
	TNFa response	49
Figure 3.10	Pathways that do not require $I\kappa B\alpha$ signaling are not affected	
	by SNPs in the <i>NFKBIA</i> promoter	50
Figure 4.1	Possible future areas of research in mapping functional role of	
	NFKBIA promoter variants	60
Figure 4.2	Overall summary of research	61

Figure A.1	Haplotyope for the NFKBIA gene in individuals of European	
	Descent	70
Figure A.2	Haplotype for the NFKBIA gene in individuals of African descent	71
Figure A.3	Haplotype for the NFKBIA gene in individuals of Chinese (Han)	
	descent	72
Figure A.4	Haplotype for the NFKBIA gene in individuals of Japanese descent	72
Figure A.5	Strong LD is seen in non-coding regions of NFKBIA	73

List of Abbreviations and Symbols

APOE	Apolipoprotein E type 4	dbSNP	Single nucleotide polymorphism database
ARDS	Acute respiratory distress syndrome	EDA-ID	Ectodermal dysplasia with immunodeficiency
AS	Ankylosing spondylitis	EMSA	Electrophoretic mobility shift assay
BAFFR	B cell –activating factor receptor	FCS	Fetal calf serum
Bcl-3	B-cell CLL/lymphoma 3	5'UTR	5 prime untranslated region
βTRCP	SCFIĸBE3 ubiquitin ligase complex	GVA	Genome variation server
ВНІ	Brain heart infusion	HBV	Hepatitis B virus
CEU	Utah residents of Northern and Western European descent	HBVC	HBV genotype C
CFU	Colony forming units	НСВ	Han Chinese of Beijing
ChIP	Chromation immunoprecipation	НСС	HBV induced carcinoma

IFNα	Interferon alpha	IL-1	Interleukin 1
IKBL	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	IL-1R	Interleukin-1 receptor
ΙΚΚα	l kappa-B kinase, alpha	IL-6	Interleukin 6
ΙΚΚβ	l kappa-B kinase, beta	IPD	Invasive pneumococcal disease
ΙΚΚγ	l kappa-B kinase, gamma	JPT	Japanese of Tokyo
ΙκΒα	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	Kd	Dissociation constant
ΙκΒβ:	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta	LADA	Latent autoimmune diabetes in adults
ΙκΒε	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon	LCL	Lymphoblastoid cell lines
ΙκΒζ	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, zeta	LD	Linkage disequilibrium
IRAK4	Interleukin-1 receptor- associated kinase 4	LPS	Lipopolysaccharide

LTβR	lymphotoxin-β receptor	NIK	NF-κB inducing kinase
MEF	Mouse embryonic fibroblasts	NKC	Natural killer cells
ΜΟΙ	Multiplicity of infection	NLS	Nuclear localization signal
MS	Multiple sclerosis	OD	Optical density
MYD88	myeloid differentiation primary response protein	OR	Odds ratio
NES	Nuclear export signal	РВМС	Peripheral blood mononuclear cells
NFKBIA	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	PBS	Phosphate buffered saline
NFKBIB	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta	PGA	Programs for Genomic Applications
NFKBIE	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon	PID	Primary immune diseases
NF-ĸB	Nuclear factor of kappa light polypeptide gene enhancer in B-cell	RA	Rheumatoid arthritis

RANK	Receptor activator of nuclear factor-kappa B	THP1	Human acute monocytic leukemia cell line
RHD	Rel homology domain	3'UTR	3 prime untranslated region
RIP1	Ral-interacting protein 1	TLR	Toll like receptor
ROR	Retinoid related orphan receptor	TNFα	Tumour necrosis factor, alpha
RSV	Respiratory syncytial virus	TNFR	Tumour necrosis factor receptor
SLE	Systemic lupus erythematosus	TRADD	TNFRSF1A-associated death domain
SNPs	Single nucleotide polymorphisms	TRAF2	TNF receptor-associated factor 2
TCR	T cell receptor	YRI	Yoruba of Ibadan, Nigeria

Acknowledgements

First and foremost I would like to thank God for blessing me with the means, ability and the will to pursue this project. Secondly, I would like to thank Dr. Stuart Turvey for providing me the ideal training desired by most graduate students. This included but is not limited to giving constant input and feedback on my experiments and the status of my project, providing valuable advice on how to write persuasively and how to do high impact presentations and lastly for injecting some enthusiasm in me during the most challenging parts of this project. I would also like to sincerely thank the members of my supervisory committee that includes Dr. Richard Stokes, Dr. Scott Tebbutt and Dr. Rusung Tan who have provided me invaluable advice and support over the last two years. A vote of thanks also goes to Aaron Hirschfeld for his well written standard operating protocols (SOPs) and general friendliness in guiding me through the most trivial of calculations. A thanks also goes to Dr. Tobias Kollmann and his lab for their collaborative support in this project and to the Speert lab for providing me lab space for conducting bacterial experiments. A general thanks goes to the current and former personnel of my lab that includes Chris, Nico, Anthony, Julia and Rachel for their support. The experiments in this project were made possible by the gracious financial support of the National Training Program in Allergy and Asthma Research. Lastly, I want to thank my mother, father and in family their continued everyone else my for support.

Chapter 1: Introduction

1.1 From genotype to phenotype: Understanding functional consequences of genetic variation

Candidate gene studies and genome wide association studies have allowed us to identify variants in genes that associate with complex diseases. The knowledge of such variants could be used to: 1) predict an individual's genetic risk for acquiring a complex disease, 2) design therapeutics to reduce disease and 3) design drugs that specifically target symptoms in an individual based on their genetic makeup, while minimizing adverse drug reactions.

To date only limited progress has been made in translating the results of genetic association studies into clinical predictors of disease onset. A notable exception is that of the well replicated association of the apolipoprotein E type 4 (*APOE4*) allele as a risk factor for developing Alzheimer's disease (1). However, the precise functional effect of the *APOE4* allele at the biological level is yet to be determined (2).

In order to use genetic variation data in assessing risk and designing therapeutics, functional validation is required to prove that: 1) the presence of such variations leads to biological effects in the human host and 2) such biological effects are of clinical importance. Figure 1.1 illustrates the steps of this process.

Genetic variation in the promoter regions of the gene nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (*NFKBIA*) has been shown to be associated with infectious, inflammatory and cancerous conditions. The overexpression of the protein encoded by *NFKBIA*-nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha-(IkBα) has also been recommended as a therapeutic choice for treating chronic inflammatory conditions (*3, 4*). However, to date, no functional studies have been undertaken to verify the biological consequences of genetic variation in the promoter of *NFKBIA*. <u>The central goal of this project was to functionally validate these observations</u> i.e., whether specific variants or haplotypes in the promoter of *NFKBIA* can contribute towards the susceptibility to or exacerbation of infectious or autoimmune/inflammatory conditions.

Our goals were to determine:

- 1) Whether promoter polymorphisms of NFKBIA lie in functionally important regions?
- 2) Whether alleles of such polymorphisms display altered expression levels (allele specific differences)?
- 3) Whether humans of different genotypes for these polymorphisms show altered expression levels for the gene?
- 4) Whether humans of different genotypes show differential and clinically relevant innate immune responses?

This chapter has three major sections. The first section describes the pathways utilized by $I\kappa B\alpha$ in regulation of inflammation, the consequences of altered signaling of these pathways along with the altered functioning of $I \kappa B \alpha$. The second section briefly describes the concept of linkage disequilibrium (LD), followed by an overview of the association studies done with the promoter variants of the IkBa gene-NFKBIA. The third section discusses the techniques frequently utilized in functionally validating promoter variants. Finally we provide a hypothesis to explain the observed associations of NFKBIA promoter with infectious variants and inflammatory/autoimmune conditions.

1.2 The NF-κB pathway: Overview

The nuclear factor of kappa light polypeptide gene enhancer in B-cell (NF- κ B) pathway plays an important role in immune response, inflammatory disease and cell death (4). The NF- κ B family consists of 5 distinct proteins p50, p52, p65(RelA), C-Rel and RelB that either exist as homo or heterodimers (5). The p50-p65 heterodimers are transcriptionally active whereas the p50-p50 homodimers are transcriptionally repressive (6). All NF- κ B proteins share a conserved region known as the Rel homology domain (RHD) which has a nuclear localization signal, as well as dimerization and DNA-binding functions (7). Although NF- κ B can be activated by a variety of receptor systems responding to a variety of stimuli, they can be grouped into 2 major pathways.

1.2.1 Classical pathway of NF-KB signaling

The classical or canonical pathway is critical for the initiation of the inflammatory response as well as the inhibition of apoptosis (4). It regulates the expression of the vast majority of proinflammatory cytokines, chemokines, leukocyte adhesion molecules and prosurvival and antiapoptosis genes (8). It is initiated by the stimulation of tumour necrosis factor receptor (TNFR), Toll-like receptor (TLR), Interleukin-1 receptor (IL-1R) and T cell receptor (TCR) pathways. Although the intermediate proteins involved in each of these cascades differ, all

pathways converge in the activation of the I kappa-B kinase (IKK) complex. The IKK complex of kinases consists of the catalytic subunits IKK α , IKK β and the regulatory subunit IKK γ that phosphorylate downstream IkB inhibitory proteins (*9, 10*). The IkB proteins in an unstimulated state inhibit NF-kB (p50-p65) activity and sequester it into the cytoplasm. Upon stimulation the IkB proteins are phosphorylated, ubiquitinylated, and targeted for degradation in the proteasome. This frees the p50-p65 heterodimer allowing it to translocate to the nucleus where it initiates the transcription of proinflammatory genes (*11*).

1.2.2 Alternative pathway of NF-кВ signaling

The major function of the non-canonical pathway is in mediating the adaptive immune response, the regulation of lymphoid organogenesis and B cell survival and maturation. It is initiated by a subset of receptors from the TNF receptor family that includes lymphotoxin-beta receptor (LT β R), CD40, receptor activator of nuclear factor-kappa B (RANK), and B cell–activating factor receptor (BAFFR) (6). The non-canonical pathway involves the recruitment of an upstream NF- κ B inducing kinase (NIK) that phosphorylates and activates IKK α dimers (*12*). The IKK α dimers ubiquitinate p100 proteins (instead of the I κ B proteins) that in a resting state bind and inhibit the translocation of the p52-Rel B NF- κ B (instead of p50-p65) dimers in the cytoplasm. From a disease perspective, the classical pathway of NF- κ B signaling is more relevant as it has been implicated in a variety of infectious, autoimmune and cancerous syndromes.

1.2.3 The importance of regulating NF-κB signaling: Examining consequences of altered NF-κB signaling:

Loss of NF- κ B signaling is prominent in infectious conditions caused by Mendelian defects in genes encoding proteins of the innate immunity cascade. These primary immune disease (PID) studies have provided valuable insight in the role of the members of the TLR signaling cascade. Patients who have autosomal recessive mutations in genes that encode for key adaptor proteins in the TLR cascade such as myeloid differentiation primary response protein (MYD88), or signaling intermediates such as interleukin-1 receptor-associated kinase 4 (IRAK4) display undue susceptibility to pyogenic bacterial functions (*13, 14*). Similarly, patients with autosomal dominant mutations in the genes that encode for regulatory proteins such as IkB α and IKK γ also display loss of NF- κ B signaling (*15, 16*). Other members of the TLR cascade have also been implicated with loss of NF- κ B signaling. For more details refer to review papers referenced at the end of the chapter (*17, 18*).

Autoimmune conditions are complex syndromes that can be triggered both by genetic predisposition and by environmental factors. A gain or increase in NF-kB signaling has been

observed in certain autoimmune conditions (*19*). Patients with rheumatoid arthritis (RA) display constitutively high levels of proinflammatory cytokines such as TNF α , IL-1 and IL-6 (*3*). Activated forms of p50 and p65 subunits of NF- κ B are also seen in nuclei of synovial lining cells in patients with RA (*20*). Multiple sclerosis (MS) patients also display elevated levels of the p65 subunits of macrophages and oligodendrocytes from active MS lesions (*21*).

1.3 The role of IκB proteins in regulating NF-κB activity

The IkB family of proteins are the most important NF-kB interacting proteins (*11*). They comprise a family of 3 classical members (IkB α , IkB β , IkB ϵ) and 2 novel IkB-like members (IkB ζ and B-cell CLL/lymphoma 3 (Bcl-3)). All members have a characteristic ankyrin rich repeats that interact with nuclear localization signals (NLS) present on the Rel homology domains of NF-kB dimers, and an N-terminal regulatory domain that has a role in their inducible degradation (*22*). All members display unique binding preferences to NF-kB dimers and undergo signal induced proteasomal degradation with different kinetics (*23*).

1.3.1 Importance of IκBα in regulation of NF-κB

 $I\kappa B\alpha$ is the best studied member of the IkB family that displays a strong negative feedback control of NF-κB (5, 23). In an unstimulated state, IκBα masks the NLS of p50, but not the NLS on the p65 subunit of the p50-p65 NF- κ B heterodimer and thus sequesters it in the cytoplasm. In response to an upstream stimulus, the activated IKK complex phosphorylates IkBα on Ser32 and Ser36 residues. This leads to the polyubiquitination of the Lys19 residue by the SCF^{IKB}E3 ubiquitin ligase complex (BTRCP) which catalyzes the formation of degradative Lys48-linked polyubiquitin chains (24). Ubiquitination of $I \kappa B \alpha$ leads to its degradation by the 26s proteasome. The now unhindered NLS on the p50 subunit of the p50-p65 heterodimer allows this complex to migrate to the nucleus and bind to NF-kB response elements on proinflammatory genes and begin their subsequent transcription. The *NFKBIA* gene that encodes IκBα is also a target gene that is expressed early in response to the P50-P65 heterodimer. This results in a strong induction of IkBa mRNA synthesis, followed by a restoration of IkBa protein levels in the cell (25). IkB α then translocates to the nucleus and binds to NF-kB, and the complex is exported back to the cytoplasm (26). This feedback regulation of NF- κ B ensures that activation of NF- κ B is limited and transient. The apparent simplicity of such an autoregulatory loop has been questioned (26). Studies have revealed that under unstimulated conditions, the unbound NLS of the p65 subunit of the p50p65 heterodimer and the nuclear export signal (NES) work in opposing directions, resulting in a constant shuttling of the complex between the nucleus and the cytoplasm (27). Degradation of IkB α tips this dynamic balance in favor of nuclear localization of NF- κ B (5). Although the precise transport mechanism behind the transport of $I\kappa$ B α to the nucleus, and subsequent return to the cytoplasm with the p50p65 dimer is not clear, recent experiments have shown $I\kappa B\alpha$ is remarkably efficient at increasing the dissociation rate (kd) of NF- κ B from DNA (28).

1.3.2 Functional consequences of altered IkBa function in mammals

Mice

Mice that lack *NFKBIA*, the gene encoding IkB α , though normal at birth, display severe runting, skin defects and extensive granulopoiesis, leading to death by 8 days (*29*). Other IkB α knockout experiments in mice reveal increased percentages of monocytes/macrophages in spleen cells of 5, 7 and 9 day old pups. Death is accompanied by severe dermatitis and increased levels of TNF α mRNA in the skin (*30*).

Humans

In humans, to date, individuals with mutations leading to dramatically decreased IkB α levels have not been found. Consequently, hypermorphic mutations, i.e., mutations that enhance the inhibitory capacity of IkB α by preventing its phosphorylation and subsequent degradation have been observed. Such individuals display impaired NF-kB signaling and impaired T-cell function, coupled with ectodermal dysplasia with immunodeficiency (EDA-ID) (*15, 31-33*). Thus, although direct consequences of the loss or gain of IkB α activity has not been observed, functional experiments in mice and primary immunodeficiency cases in humans suggests that the under expression of IkB α can lead to inflammation whereas sustained expression can lead to immunodeficiency.

1.4 Understanding linkage disequilibrium (LD)

Genetic association studies have been pivotal in providing clues for the genetic nature of complex syndromes. Such studies have become the method of choice for clinicians and researchers alike for determining susceptibility to disease. It is a method which utilizes families, cases and controls, or cohort data to statistically relate genetic factors to the phenotype of interest (*34*). The dominant theory behind conducting a genetic association study is the 'common disease-common variant hypothesis.' The hypothesis states that many common variants in the population with a frequency of 1% or greater have a modest effect on an individual's phenotype (*35*). It is the joint effect of these variants that determines disease susceptibility in the population. These variants are called polymorphisms. They can be tandem repeats, insertions, deletions or single nucleotide polymorphisms (SNPs). SNPs, by far, are the most abundant polymorphisms with more than 10 million present in the human genome (*36*). Currently, owing to technological limitations, it is not possible to account for the identity of all

SNPs. However, as it will be shown in the sections below, by relying on the concept of linkage disequilibrium (LD), it is possible to identity the majority of these SNPs in cases and controls (*37*). This section briefly describes the use of linkage disequilibrium (LD) in mapping genetic variation. From a research perspective, it is pivotal that the reader be acquainted with the concept of LD as this concept was used in this project to functionally validate promoter variants of *NFKBIA*.

1.4.1 Linkage disequilibrium (LD)

Mutation occurs at a slow rate in the human genome, at 10^{-8} per site per generation (38). Initially a variant such as a SNP is jointly inherited by the offspring along with all the accompanying sequences on the same chromosome from the parent. However, during meiosis homologous recombination occurs, allowing genetic material to be exchanged from the paternal and maternal chromatids (39). Thus, the chromosome that is passed on to the next offspring contains sequences that are not identical to the chromosome in the parent. Over time, owing to several generations of recombination events occurring during meiosis, the chromosome housing the SNP gradually loses its ancestral sequence homogeneity. Since the frequency of recombination is proportional to the frequency of genetic distance between two nucleotides, sequences in the vicinity of the original mutation tend to be retained This leads to the formation of islands of nucleotide sequences that still retain the ancestral sequence on a chromosome (referred to as the Haplotype), interspersed by 'hot spots' of recombination (40). Nucleotides or variants in a haplotypes are said to be linked and this association is defined as linkage disequilibrium (LD) i.e., non-random association of two alleles at one or many loci.

1.4.2 Using LD for association mapping

Initially, investigators relied on 'linkage analysis' which served as a prelude to the use of LD, to identify candidate loci implicated in disease. This approach relies on the fact that genetic markers closely linked to the locus of interest tend to remain associated with the candidate locus through several rounds of recombination. In an affected family, as a causative gene segregates through a kindred, other nearby markers on the same chromosome tend to segregate together (*34*). Linkage analysis in combination with positional cloning showed remarkable success in identifying genes responsible for single gene Mendelian disorders such as Cystic fibrosis and Duchenne muscular dystrophy (*41, 42*).

1.4.3 Using LD for genome wide association studies

In the case of complex diseases the traditional method of linkage analysis had initially proven to be difficult because previously used markers of linkage analysis such as microsatellites did not

provide sufficient coverage of the genome to map multiple variants that were presumed to be associated with the various phenotypes of such diseases (43). Furthermore, it was believed based on simulations by Kruglyak that patterns of LD would be low in the human genome and would be limited to a maximum of 3 kB segments of chromosomes (44). The poor genomic coverage of previous markers was resolved by SNPs that were more abundantly present in the genome. Secondly, with the availability of larger sequence data sets, it was proven that the extent of LD was greater than what was assumed initially, with results indicating that as much as 65% to 85% of the human genome was comprised of haplotype blocks 10 kb in length or greater (45). The discovery of the relative abundance of haplotype blocks in the human genome led to the postulation that a single variant in a haplotype block-owing to strong LD could be predictive of the identity of other variants in the block and could be used as a 'tag' to distinguish one haplotype block from the other (46). Therefore, one would only need to catalogue a sufficient number of 'tagging' SNPs to scan the entire human genome for any common risk variants that would associate with disease. However, in order to choose the appropriate tag SNPs for conducting genome wide studies of association, genome wide 'LDmaps' were needed for different human ancestral populations. This became the impetus for the international Hap Map project which was undertaken with the purpose of cataloguing such variation in various ethnic groups (47).

1.4.4 Patterns of LD in different populations

The first phase of the HapMap project involved the studying of 1 million SNPs in the 4 different population groups that included the Yoruba of Ibadan (YRI), Nigeria, Utah residents of Northern and Western European descent (CEU), Han Chinese of Beijing (HCB), China , and Japanese individuals of Tokyo (JPT), Japan (*47*). This was followed by the second phase that involved the identification of an additional 3.1 million SNPs. Results indicated that the degree of LD was lower in YRI population but similar across regions with higher LD in the 4 population panels (*48*).

1.4.5 Tag SNP coverage and transferability in different populations

An important statistic of LD is r^2 which measures how well a SNP acts as a proxy for a nearby SNP (An r^2 value of 100 implies 100% concordance of a selected SNP with a nearby SNP). Preliminary results of the HapMap study showed that selecting as few as 447,579 SNPs for CEU, 434,476 SNPs for CHB+JPT and 604,886 SNPs for YRI as tags was sufficient to account for 100% of all the common SNPs in the HapMap panel. The second phase, however, reduced the coverage to at least 80% while increasing the minimum number of tag SNPs to 579,978, 670,407 and 780,336 (for CEU, CHB+JPT and YRI respectively). Owing to a larger number of

SNPs needed for adequately covering SNPs in YRI population, most existing genotyping platforms provide poor coverage for SNPs in this panel when compared to SNPs in CEU or CEU+JPT panels. The highest coverage provided is by the Illumina HumanHap 550k platform, which is 68% coverage for the YRI panel versus 95% for CEU and CHB+JPT (*49*). Another debated issue is how useful are selected tag SNPs for providing coverage in the genomes of other major population groups. The emerging understanding is that transferability of selected tag SNPs is high when selected from a HapMap panel that is geographically adjacent to the target population (*50*).

1.5 Association of observed variation in the gene (*NFKBIA*) encoding IκBα with infection, autoimmunity/inflammation, and cancer

A brief description of the major association studies done with variants in the *NFKBIA* gene is described below. A summary table (table 1.1) is also included at the end of the chapter with relevant information, such as sample size, ethnicity, *P* values, etc.

1.5.1 Association of NFKBIA polymorphisms with infections

Trachoma

The first known association of *NFKBIA* polymorphisms was reported by Chamay *et al.* in Gambian patients with trachoma (*51*). Trachoma is an infectious disease caused by repeated infections of the ocular surface by *Chlamydia trachomatis*. Subsequent inflammation leads to scarring followed by inturning of the eyelashes and blinding corneal opacification (*52*). Three promoter SNPs of *NFKBIA*, rs3138053, rs2233406, and rs2233409 were analyzed. The SNPs rs3138053 and rs2233406 were in complete linkage disequilibrium, and the minor allele haplotype (GT) was shown to be protective against infection (*p*=0.046).

Invasive pneumococcal disease (IPD)

Chapman *et al.* examined associations of 62 SNPs in *NFKBIA*, *NFKBIB*, and *NFKBIE* for susceptibility to invasive pneumoccal disease (IPD) in individuals of European descent (53). The minor alleles of *NFKBIA* promoter SNPs rs3138053 and rs2233406 were shown to be protective against IPD in an initial study group (odds ratio (OR) 0.64(0.48-0.87), 0.57(0.43-0.76)) and against pneumcoccal empyema in a subgroup (OR 0.42 (0.21-0.83), 0.47(0.24-0.93). However, the association for rs3138053 was lost after applying the Bonferroni correction for multiple testing.

Respiratory syncytial virus (RSV)

Siezen *et al.* analyzed variants in the innate immunity cascade for associations with susceptibility to RSV in premature infants (54). The minor allele of the *NFKBIA* promoter SNP rs2233409 was shown to be more protective against RSV in premature infants, born with underdeveloped lungs, than in term children (p=0.0261).

1.5.2 Association of NFKBIA polymorphisms with autoimmune/inflammatory conditions Sarcoidosis

Abdallah *et al.* analyzed Dutch patients with sarcoidosis for associations with *NFKBIA* promoter SNPs rs3138053, rs2233406, and rs2233409 (55). The minor allele of the rs2233409 SNP was more prevalent in patients with sarcoidosis (p=0.008). The minor allele haplotype (GTT) was associated with risk of sarcoidosis (p=0.01).

Multiple sclerosis (MS)

Miterski *et al.* found that the minor allele of an 8 base pair insertion/deletion variant (708ins8) in the promoter of *NFKBIA* was protective against multiple sclerosis (P<0.01) (56). The variant seems to be rare as neither follow up studies have been performed with this SNP nor has a reference SNP accession id (rs number) been submitted to the single nucleotide polymorphism database (dbSNP).

Acute respiratory distress syndrome (ARDS)

Zhai *et al.* examined US patients with acute respiratory distress syndrome for associations with *NFKBIA* promoter SNPs rs3138053, rs2233406 and rs2233409 (57) The GTT haplotype, consisting of minor alleles rs3138053/rs2233406 and the major allele for rs2233409, was associated with the risk of ARDS (OR 1.66 (1.09-2.53), p=0.02).

Graves' disease

Kurlowicz *et al.* examined Polish patients with Graves' disease for associations with SNPs in nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1 (*IKBL*) (rs2071592 (promoter), rs2071592 (intron1), rs3130062 (exon4) and *NFKBIA* (rs2233409 and rs2233406 (promoter), rs696 (3' untranslated region (UTR)) (58). Although SNPs for the *NFKBIA* gene were not associated with the risk of Graves' disease, the minor alleles for the promoter SNPS rs2233406 and rs2233409 were associated with clinically evident ophthalmopathy (OR 1.67 (1.20-2.36), *p*= 0.036 and OR 1.65 (1.18-2.38), *p*=0.07 respectively). The minor allele haplotype (TT) was also significantly associated with the presence of clinically evident ophthalmopathy (*p*=0.003).

Note: The next four studies (Primary Sjögren's syndrome, Systemic lupus erythematosus, Rheumatoid arthritis and Ankylosing spondylitis) were performed by one research group based in Taiwan. A great deal of overlap is seen for associations of SNPs in the *NFKBIA* promoter with each of the four diseases examined. Also, the two promoter SNPS rs3138053 and rs2233406 do not display strong linkage disequilibrium, which is contrary to what is seen in major HapMap populations. An email inquiry was made about the ethnicity of the population. The authors indicated that the Taiwanese nationality assigned for the patients did not imply ethnic homogeneity, but on the other hand included a heterogenous population of Han Chinese, Hakka, Taiwan Aboriginals, and other populations [personal communication July 4th, 2009]. Therefore results from these 4 studies may not be generally applicable at the population level owing to the inclusion of rare ethnicities in the study.

Primary Sjögren's syndrome

Patients had a lower frequency for the minor allele of SNP rs3138053 (G) (OR 0.5 (0.2-0.9), p=0.02) and the major allele of SNP rs2233406 (C) (OR 16.2 (10.0-26.2)) than controls. The haplotype (ATACC) was associated with the risk of Sjögren's syndrome (OR 34.14 (17.77-65.59), p<0.007) while the haplotype (ACACC) was associated with protection (OR 0.05(0.03-0.08), p<0.007) (59).

Systemic lupus erythematosus (SLE)

The minor allele for *NFKBIA* promoter SNP rs2233406 and the haplotype ATACC were associated with the risk of SLE (OR 2.0 (1.2-3.4), *p*=0.01 and OR 3.2 (1.6-6.2), *P*<0.002) (60).

Rheumatoid arthritis (RA)

The minor allele of rs2233406 (T) and the major allele of rs2233407(A) were associated with the risk of RA (OR 1.6 (1.1-2.4), P=0.027) and (5.1 (1.4-18.2), P=0.007). The TAC haplotype was also associated with the risk of RA (OR 1.8 (1.1-2.8), P=0.01) (61).

Ankylosing spondylitis (AS)

The minor allele of the SNP rs2233406 was associated with the increased risk of AS (4.1(2.7-6.0), P<0.001). The haplotypes ATACC and ACATC were associated with the risk of AS (OR 8.0 (4.26-15.02), P<0.007 and OR 8.54 (1.99-36.69), P=0.007 respectively) (62).

Ulcerative colitis

Szamosi *et al* found that the minor allele genotype (GG) for the 3'UTR variant rs696 of *NFKBIA* was associated with the risk of ulcerative colitis in Hungarian individuals with inflammatory bowel disease. (OR 0.003 2.97(1.45-6.08) (*63*).

Crohn's disease

Klein *et al.* found that the major allele genotype and (AA) and the major allele (A) of the 3'UTR variant rs696 were more prevalent in patients with Crohn's disease than controls (P<0.003, relative risk 1.77 and P<0.02, relative risk 1.22) (64).

Type 2 diabetes

Romzova *et al.* found that the AA genotype for the 3'UTR (A to G) variant was associated with the risk of diabetic nephropathy (OR 3.59, *P*=0.0015) (*65*). The relative position of this variant and the rs number was not mentioned. However, based on a search of SNPs in dbSNP, the only A to G common variant is rs696. The only other 3' UTR variant A to G variant is rs1131420, but there is no population frequency data for this variant in dbSNP.

Latent autoimmune diabetes in adults (LADA)

Latent autoimmune diabetes in adults occurs in subset of individuals with type 2 diabetes and is characterized by autoimmune and immune mediated β -cell dysfunction (66). Katarina *et al.* found that the AA genotype of a 3'UTR SNP rs696 was associated with the risk of LADA (OR 2.68, *P*<0.0001) in a Czech population (67).

1.5.3 Association of NFKBIA polymorphisms with cancer

Myeloma

Spink *et al.* found that risk haplotype (GCCTATCA) corresponding to SNPs rs3138053 (A/G), rs22233409 (C/T), rs3138054 (G/A), rs2233419 (C/T), rs1957106 (C/T), rs10782383 (C/T), rs8904 (C/T) and +2921 (A/G) (not validated) was associated with the increased risk of multiple myeloma (OR 2.29 (2.10-2.49), P= 0.006) (*68*).

Colorectal cancer

Gao *et al.* found that the AG genotype for rs696 in the 3'UTR was associated with the risk of colorectal cancer in Chinese patients \geq 50 (OR 3.06 (1.55-6.02, *P*=0.001)). The GG genotype of the same SNP was associated with poor survival rate in Swedish patients (OR 3.10 (1.28-7.60), *P*=0.01) (69).

Hepatocelullar induced carcinoma

Yongchao *et al.* examined patients with hepatitis B virus (HBV) infections for genetic susceptibility to HBV induced carcinoma (HCC) (70). Two HBV genotypes were examined: patients with HBV genotype B (HBVB) and patients with HBV genotype C (HBVC) infections. The CT genotype for the rs2233406 and the AG genotype for rs3138053 SNPs were more prevalent in patients with HBVC infections and HCC as compared to patients with HBVC infections and no HCC (1.71 (1.07-2.73), P=0.024 and 4.02 (2.14-7.59), P=0.000). The haplotype GTC (3138053(A/G), 2233406 (C/T), 2233408 (C/T)) was associated with the risk of HCC in patients with HBVC infections (3.142 (1.443-6.838, P=0.002)).

1.6 Opposing associations of *NFKBIA* **promoter polymorphisms with infection and autoimmunity: A genetic link to the hygiene hypothesis?**

The vast majority of promoter polymorphisms examined to date for *NFKBIA* associate with either infectious or autoimmune/inflammatory conditions. After excluding studies that analyzed rare ethnic groups such as the ones present in the Taiwanese study group, a general pattern emerges-the minor variants for SNPs rs3138053(G) and rs2233406(T) that associate with protection from infectious conditions are associated with the risk of acquiring autoimmune/inflammatory conditions (see table 1.2). Epidemiological studies in Europe have shown that the decline of certain infections such as rheumatic fever, hepatitis A, tuberculosis have met with a concomitant rise of autoimmune diseases such as multiple sclerosis, type 1 diabetes, and Crohn's disease in these regions (71). The hygiene hypothesis states that the decline of infections in the West is the causal link for the increase of immune disorders (72). Therefore, the next logical step would be to validate these associations to see if there is a functional explanation for these observations. In subsequent sections an overview is provided on the various approaches used in validating regulatory SNPs and specifically SNPs that lie in the promoter.

1.7 Validating regulatory variants and the genetics of gene expression

Even though SNPs that occur in coding regions of genes are predicted to be more deleterious because they affect protein function (73), the vast majority of variants that associate with complex syndromes lie in regulatory regions of the human genome (74). Therefore, there is a need to validate such variants for functional relevance to disease. In the absence of known functional effects of such variants at the biochemical level, the analysis of gene expression can serve as a intermediate phenotype for evaluating the functional effects of such variants (75). The measure of gene expression has been recommended as a surrogate for linking clinical features of a patient to a variant associated with the syndrome (76). Genetics of gene

expression studies therefore typically involve measuring steady state mRNA levels of many genes using microarray analysis or RNA sequencing, followed by genotyping of individuals for polymorphic markers (SNPs) and a statistical analysis to identify regulatory variants that associate with expression levels for genes of interest (77).

1.7.1 Cis and trans-variation

mRNA expression from the alleles of a given gene is under the control of *cis* or *trans* acting variants. *Cis* variants lie on the same chromosome (allele) of the gene that they regulate and they reside on the regulatory regions of the genes such as promoters, enhancers and splice sites (78). Thus, the impact of *cis* acting variants on gene expression is allele specific. *Trans*-acting factors such as transcription factors may reside on different chromosomes and thus variants of such factors affect the expression of both alleles equally (79). In comparison to *cis* acting variants, *trans*-acting variants are difficult to detect as they can be present anywhere in the genome relative to the locus and their effects on gene expression tend to be of a smaller magnitude than *cis* variants (*80*). *Cis*-activating variants have been validated by *in vitro* studies involving the use of functional assays in cell lines and by *in vivo* approaches by the measure of differential allelic expression (DAE) in human tissues. The mapping of polymorphisms in *trans*-acting factors has been difficult and has relied on traditional linkage analysis methods.

1.8 Approaches for validating *cis*-acting variants

1.8.1 Direct methods: Functional assays

In vitro promoter assays

A direct way to functionally validate the effects of a *cis*-acting polymorphism is to clone the promoter of the gene in an expression vector with a reporter that lacks endogenous promoter activity. The construct is transiently transfected into a cell line and the activity of a reporter gene is measured (73). When examining multiple variants, several constructs with individual variants, can be generated and transfected into cell lines. A control construct without the gene promoter is also transfected to measure any baseline activity of the reporter gene. A difference in reporter activity relative to control (baseline activity) between the variants of a polymorphism would indicate the functional importance of such a variant (73). However, a reporter gene assay is more suitable for measuring large differences in promoter activity, such as might be observed in the context of a mutation or deletion of functionally important transcription elements. The variation in promoter activity owing to the presence of regulatory variants can be of a small magnitude (*81*). Such low level differences are hard to distinguish from differences seen from a variety of non-genetic sources. Differences in promoter activity can result from variation in

transfection efficiencies of the promoter constructs. Such studies do not take into the account the effect of trans-acting variants that can impact promoter activity (*82, 83*).

Electrophoretic mobility shift assays (EMSA)

An EMSA is used to asses the strength of binding of a protein to a target DNA sequence that is known or predicted to be of functional importance. In its most basic form the assay involves generating fluorescently or radioactively labeled oligonucleotides corresponding to individual alleles of a SNP (*84*). The oligonucleotides are incubated with nuclear or cell extracts from tissues, cells or organelles. The lysate-oligonucelotide mix and oligonucleotides are run in parallel on a non-denaturing gel. If there is a binding interaction between the target site and proteins in the lysate, this leads to the retardation of the mixture on the gel relative to the lane that contains pure lysates or probe. This 'band shift' then corresponds or represents a functional interaction of the target site with the protein (*73*). This approach is limited by the size of the oligonucleotides (approximately 20-30 nucleotides), and is unable to provide full resolution for oligonucleotides that bind multiple target sites (*85, 86*).

Chromation immunoprecipation (ChIP) assays

A ChIP assay is useful for studying interactions between transcription factors and their target DNA sites *in vivo*. It involves the cross-linking of transcription factors to target sites by the use of cross-linking agents such as formaldehyde followed by sonication to fragment the DNA and immunoprecipiation by a transcription-factor specific antibody. The amount of DNA that is present in the immunoprecipated protein-DNA complex is determined by PCR for each allele. A difference in the amount of DNA bound to the complex illustrates the strength of binding of a transcription factor to its target. This elucidates the functional impact of any variant that lies in the binding site (*84*).

1.8.2 Indirect methods: Differential allelic expression (DAE) or allelic imbalance (AI) The problem of quantitating promoter variants:

When it comes to quantitating the impact of a promoter variant on the transcription efficiency of the transcript encoded, it is difficult to discriminate between the two allelic transcripts. Figure 1.2a illustrates this conundrum. Since the mRNA transcript lacks the promoter, it is difficult to discriminate between the transcript that arises from the major allele versus the transcript that arises from the minor allele. The solution is to quantify a coding SNP in the mRNA that is linked with the promoter variant. By pre-screening an individual for the linkage of both variants i.e., the major variant in the promoter co-segregates with a major variant in the coding region while the minor variant of the promoter co-segregates with the minor variant in the coding region, it is

possible to discriminate between the major and minor alleles of the region (Fig. 1.2b). The amount of each transcript can be quantified and expressed as a ratio of major transcript to minor transcript or vice versa. This ratio is compared to the ratio of both alleles in (base line condition) genomic DNA to determine if there is allelic imbalance in the gene.

There are several advantages for utilizing differential allelic expression assays as opposed to other methods of validation. Both alleles are expressed in their natural environment. They are exposed to the same trans-acting factors. Comparisons of alleles are made within an individual as opposed to between individuals. This minimizes inter-individual variation (*87*).

Methods for detecting allelic imbalance

Most indirect methods of quantitating *cis* allelic variation work on the basic principle of quantitating the ratio of the alleles of the coding SNP that are in strong linkage with the promoter. The difference arises in what is used as the indicator or a 'surrogate' for determining allelic ratio. Methods such as SNaPshot, sequencing, pyrosequencing, rely on the 'endpoint' readings of PCR. Probe based assays such as TaqMan real-time PCR allow greater precision, as they rely on measuring differences in the threshold of amplification, referred to as ΔCt of each allele (*87*). This method relies on comparing differences of allelic amplification cycles for each allele in a heterozygous individual for a SNP in the coding region. Under genomic condition i.e., in the genomic DNA of an individual, the allelic ratio in a gene is 1:1. In a probe based assay such as the TaqMan assay, a 1:1 expression in theory should correspond to no net difference between the thresholds of amplification for either alleles of a gene. In practical terms there is always a background difference in the genomic DNA resulting from differing probe binding efficiencies for each allele. This difference is unique for each individual and is accounted for when determining allele ratios. Any difference in ΔCt in heterozygous individual cDNA versus the difference seen in the genomic DNA indicates allelic imbalance (*88*).

1.9 Hypothesis

Based on the role of *NFKBIA* that has been inferred from examining: 1) targeted *NFKBIA* gene knockouts in mice, 2) humans who display abnormalities in IκBα function, and 3) the association of SNPs in the *NFKBIA* regulatory regions with disease, we hypothesize that the presence of minor variants of SNPs rs3138053, rs2233406 and rs2233409 in the promoter of *NFKBIA* will have functional consequences. Specifically, this will lead to a reduced expression of this *NFKBIA* allelic transcript, resulting in allele specific differences in expression, which will translate to a lower overall *NFKBIA* mRNA expression in individuals who have the minor allele,

ultimately resulting in an increase in proinflammatory cytokine secretion. Figures 1.4a and 1.4b illustrate this hypothesis.



Fig. 1.1. Steps needed to verify if variation in a candidate gene influences susceptibility to a complex disease. Genetic association studies provide hints for the association of variants in a gene with diseases. Further validation at the biological level is needed along with clinical validation to show that variants within the candidate gene of interest have causal roles in the onset of complex disease. The next step would involve developing therapeutic measures to optimize the activity of the gene in order to treat the disease.

		Pro	omoter			Intro	nic	Exe	onic	3'l	JTR								Ethnically		Functional
Position	-881	-826	-297	-519	-550	+1678	+2025	+212	+1059	+2787	+420	+2921	n/a	Association	Р	Odds Rato	Cases	Controls	Matched?	Ethnicity	Data?
	138053	233406	233409	233408	233407	138054	233419	957106	1782383	304	96	a.	a.								
rs id	3	57	52	R	S	, e	N N		Ę	8	⁶	2	2								
Allelic Variant	(A/G)	(C/T)	C/T)	(C/T)	(A/G)	(G/A)	(C/T)	(C/T)	(C/T)	(C/T)	(G/A)		(A/G)		0.040		100	101	Vee	Combion	Nia
Invasive pneumococcal disease (53)	G	÷												protection	0.046	0.57(0.43-0.76)	762	260	Yes	Gampian	NO
invasive priedinococcar discuse (00)	0	'												protection	0.028	0.47(0.24-0.93)	370	40	Yes	Caucasians	No
Respiratory syncytial virus (54) Sarcoidosis (55)	G	т	T T											protection risk	0.0261	0.61	470 205	1008 201	No Yes	Undeclared Caucasians	No No
Multiple Sclerosis (56) Acute respiratory distress syndrome													G	risk	<0.025	1.36 (1.06-1.75)	185	438	Yes	Caucasian	Yes
(57)	G	т	С											risk	0.02	1.66 (1.09-2.53)	382	828	Yes	Caucasians	No
Graves' disease (58)		т	т											risk	0.03	1.73 (1.02-2.94)	481	455	Yes	Caucasians	No
Primary Sjögren's Syndrome (59)	А	т	С	С	Α									risk	< 0.001	34.14(17.77-65.59)	98	110	No	Taiwanese	No
Systemic lupus erythematosus (60)	Α	Т	С	С	Α									risk	<0.002	3.2 (1.6-6.2)	110	120	No	Taiwanese	No
Rheumatoid arthritis (61)		Т		С	Α									risk	0.01	1.8 (1.1-2.8)	140	115	No	Taiwanese	No
Ankylosing spondylitis (62)	A	Т	С	С	A									risk	0.001	8.54 (1.99-36.69)	154	112	No	Taiwanese	No
Ulcerative colitis (63)											GG			risk	0.003	2.97(1.45-6.08)	266	149	No	Hungarian	No
Crohn's disease (64)											AA			risk	< 0.003	1.77 (relative risk)	145	441	No	Undeclared	No
Turne 2 dishetes (65)											A			risk	< 0.02	1.22 (relative risk)	145	441	No	Undeclared	No
Type 2 diabetes (65)													AA	risk	0.0008	2.59	246	159	res	Caucasian	NO
Latent autoimmune diabetes in adults (67)											AA			risk	<0.0001	2.68	31	153	Yes	Caucasian	No
Myeloma (68)	G		С			А	Т	С	т	С		А		risk	0.006	2.29(2.10-2.49)	157	196	No	Undeclared	No
Colorectal cancer (69)											AG GG			risk Survival	0.001	3.20(1.61-6.38)	199 155	577 438	Yes	Chinese	No No
Hepatocelullar carcinoma (70)	G	т		С							00			risk	0.002	3.142(1.443-6.838)	202	482	Yes	Chinese	No

Table 1.1. Genetic association of studies of variants of NFKBIA

The table summarizes all the genetic association studies involving variants of *NFKBIA* done to date. For the variants located at position +2921, no rs number is available. For multiple sclerosis and type 2 diabetes, the identity (relative location from the start site or the rs number) of the variant was not mentioned in the publication nor could it be traced on dbSNP.

Position of promoter SNP	-881	-826	-297	Association	Reference
rs id	3138053	2233406	2233409		1
Change	(A/G)	(C/T)	(C/T)		
Infectious disorders					
Trachoma	G	Т		Protection	(51)
Respiratory syncytial virus			Т	Protection	(53)
Invasive pneumococcal disease	G	Т		Protection	(54)
Inflammatory/Autoimmune disorders					
Sarcoidosis	G	Т	Т	Risk	(55)
Acute respiratory distress syndrome	G	Т	С	Risk	(57)
Graves' disease		Т	Т	Risk	(58)

Table 1.2. The GT haplotype is associated with protection against infections and risk for inflammatory conditions

SNPs rs3138053 and rs2233406 show complete (100%) linkage in individuals of European, African, Chinese (Han) descent and 91% linkage in individuals of Japanese descent. The 223409 SNP shows strong linkage with rs3138053 in individuals of European, Chinese (Han), and Japanese descent (83%, 83%, 89% respectively), but display only 56% linkage in individuals of African descent.







Figure 1.3. Hypothesis: Impact of minor variants in the *NFKBIA* promoter on NF-κB **signaling**. Individuals who are homozygous (ACC/ACC) for the *NFKBIA* promoter SNPs rs3138053, rs2233406, and rs2233409, are expected to display normal negative regulation of NF-κB where IκBα, upon phosphorylation by the IKK complex, would be ubiquitinated and subsequently proteasomically degraded. This would allow the p50p65 NF-κB heterodimer to bind response elements of proinflammatory genes and begin their transcription. NF-κB (p50p65) would also bind response elements on *NFKBIA*, the gene that encodes for IκBα, and initiate its expression. Newly synthesized IκBα would bind NF-κB (p50p65) and translocate it back to the cytoplasm (A). In individuals who are heterozygous (ACC/GTT) for these SNPs, there would be a reduction (dotted arrow) in the amount of mRNA transcript that is produced. This would lead to a reduction (dotted arrow) in the amount of IkBα that is available to bind NF-κB (p50p65). A reduced inhibition on NF-κB (p50p65) translocation to the nucleus would in turn lead to an increase in the expression of proinflammatory genes which would subsequently lead to an increase in inflammation (large text) (B)

Chapter 2: Material and methods

2.1 Analysis of linkage disequilibrium (LD)

2.1.1 Source populations

LD was examined in 4 ethnic groups which included individuals of European, African, Chinese (Han) and Japanese descent. To incorporate the maximum amount of variants for each ethnic group, genotype data was taken from 2 source populations- one genotyped by the HapMap consortium (www.hapmap.org) and the second genotyped by Programs for Genomic Applications (PGA) (http://pga.jgi-psf.org/). Even though there was considerable overlap between individuals used by HapMap and PGA, yet PGA assayed for more variants of *NFKBIA*. (as shown in table 2.1). The only ethnicity for which there was no overlap between individuals genotyped by PGA or HapMap was for individuals of African descent. PGA genotyped African Americans of South-Western USA whereas HapMap genotyped Yoruba of Ibadan, Nigeria. To eliminate any redundant genotypes, parameters in the Genome Variation Server (GVA) were set to 'merge' duplicate genotypes from the same individual found in two source populations. More detail is provided on the use of GVA in the next section. The location assigned to a *NFKBIA* SNP i.e. whether it was an intronic, exonic, 3' UTR, 5'UTR or promoter SNP was based on the designation given by the dbSNP (http://www.ncbi.nlm.nih.gov/SNP).

2.1.2 Software

Haplotype analyses

The Genome Variation Server (GVA) (<u>http://gvs.gs.washington.edu/GVS/</u>) was used to download genotype information for each ethnic group. In the search engine, the entire *NFKBIA* gene sequence plus 1000 bp of upstream sequence, was initially screened for variants. The 'genotype' and 'marker haplotype' files, obtained from GVA, were uploaded in Haploview 3.1. Common variants i.e., variants with a minor allele frequency $\geq 1\%$ were only included in the analysis. In our case we chose to define LD between two SNPs as 'strong' if the r^2 values were greater than or exceeded 80%, as this has been used traditionally by the HapMap consortium as an accepted level of tag SNP coverage (*48*).

Functional site analyses

The "SNP@Promoter" search engine (<u>http://variome.kobic.re.kr/SNPatPromoter/</u>) was used to analyze the *NFKBIA* promoter for any variants that were present in putative transcription factor binding sites.

2.2 Analysis of *NFKBIA* mRNA expression, and allelic imbalance

2.2.1 Subject recruitment

Recruitment for NFKBIA allele specific and total mRNA expression measure

Ninety seven healthy individuals were recruited from the research department at the BC Children's Hospital. Initially saliva samples were obtained for genotyping and determining *NFKBIA* haplotype for SNPs rs3138053/2233406/2233409/1050851(A/G, C/T, C/T, C/T). Each individual had consented for the extraction and use of their DNA for this study which was approved by the board of ethics at the University of British Columbia. Individuals with HIV infection, spleen abnormailities, significant heart and lung disease, some kidney and liver diseases, diabetes, cancer, cerebrospinal fluid leaks, sickle disease and other primary immunodeficiences were excluded from the study. 32 individuals who had consented for the use of their blood in this study were approached for blood samples.

2.2.2 Genomic DNA extraction

Genomic DNA was extracted from controls and case blood samples using QIAamp ® DNA Blood Mini Kit (Qiagen, MD, USA). Genomic DNA was extracted from case saliva samples using ORAgene® DNA extraction kit (DNA genotek, ON, Canada). DNA was eluted in 25µl of T.E. buffer and was quantified by a NANODROP 1000[™] UV/VIS spectrophotometer (Thermo Scientific, USA). Between 224.75 to 2538.75 nanograms of DNA was eluted in 25µl of Tris-EDTA (T.E.) buffer.

2.2.3 Optimization of bacterial growth

Optimizing growth medium

In spite of its pathogenicity in blood, *Streptococcus pneumoniae* grows relatively poorly on media. Many isolates require prolonged preincubation prior to entering the exponential phase of growth. In a recent publication, 47 clinical isolates of *S. pneumoniae* representing 15 serotypes were grown in brain heart infusion (BHI) media with and without 5% fetal calf serum (FCS). Supplementation with 5% FCS led to a significant shortening of the lag phase in serotypes associated with invasive disease (*89*)(*91*). The clinical isolate used in this project, serotype 14, is one of the 7 commonly invasive serotypes (4, 6B, 9V, 14, 18C, 19F, and 23F) known to cause the majority of pediatric invasive pneumococcal disease cases in Canada (*90*). We monitored growth of serotype 14 for 10 hours using a spectrophotometer adjusted at 600nm wavelength in BHI medium supplemented with and without 5% FCS. Comparisons were made for cultures started from overnight cultures (which were colony inoculations grown in BHI medium to an optical density (OD) of 0.1) or single colonies. All cultures were grown under anaerobic conditions at 37°C with 5% CO₂. With the inclusion of 5% FCS, cultures inoculated from over
night cultures (Fig. 2.1a) or single colonies (Fig. 2.1b) had a shorter lag time to exponential growth compared to cultures that were grown in BHI medium only. In the case of cultures inoculated from overnight cultures the lag time for growth in BHI medium supplemented with 5% FCS, was shorter by 90 minutes when compared to cultures that were grown in media lacking 5% FCS (Fig. 2.1a). For cultures started from single colonies, quantitative comparisons for time required to reach exponential phase could not be made as cultures grown in the (control) BHI medium failed to reach exponential phase after 10 hours of incubation (Fig. 2.1b). However, single colony cultures took a longer time to reach the exponential phase than cultures that were started from overnight cultures (440 minutes versus 380 minutes).

Determining the mid-log phase of exponential growth

In order to determine the optical density at which *S. pneumoniae* had reached the mid-log stage of exponential growth, 2 additional growth curve experiments were conducted. A clinical isolate of *S. pneumoniae*, serotype 14, was sub-cultured twice on Columbia agar plates with 5% sheep blood. A single colony was inoculated overnight in 50mL BHI medium. The next day (after 12 hours), once the culture had reached an O.D. of 0.5, 300µl of the culture was inoculated in media consisting of 50mL BHI + 5% FCS to an O.D. of 0.1. Growth was monitored for 10 hours. Plate counts were taken periodically, throughout the growth curve, whenever an O.D increase of 0.1 was observed. The next day, colonies were counted and colony count per ml for each O.D. was determined. Results from two growth curves indicated that the mid-log phase occurred at an approximate O.D. of 0.5 (Fig. 2.2a). Colony counts were plotted against colony forming units (CFU)/mL for each O.D. to yield a standard curve for predicting colony counts for a given O.D. of *S. pneumoniae* in culture (Fig. 2.2b).

Bacterial growth on the day of the experiment

A clinical isolate of *Steptococcus pneumoniae*, serotype 14, was sub-cultured twice on Columbia agar plates with 5% sheep blood. On the day of the experiment, a single colony from the second subculture was inoculated in 25 mL of BHI medium supplemented with 5% FCS. The culture was grown under anaerobic conditions at 37°C with 5% CO2 for 6 hours. At an optical density reading of 0.5, the bacteria were pelleted by centrifugation at 5000X g for 10 minutes. The bacterial pellet was serially diluted in appropriate volumes of PBS to generate multiplicity of infections (MOI) of 20 and 10. Prior to stimulation of cells, the MOIs were plated on Columbia blood agar plates, with 5% V/W sheep blood, in order to determine the number of live bacteria present in each dilution

2.2.4 PBMC isolation

Heparinized whole blood was mixed in a 1:1 ratio with R10 medium consisting of RPMI 1640 medium with 10% (v/v) fetal calf serum and 2 mM l-glutamine and 1 mM sodium pyruvate (Gibco). Human peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation (Ficoll-Paque[™] Plus, Amersham Pharmacia Biotech). PBMCs were washed twice with R10 and viable cells were quantified by trypan blue exclusion.

2.2.5 PBMC stimulation

PBMCs were suspended in R10 and were dispensed into a flat-bottom 24-well plate (6.6 x 10^5 cells in 666µl R10). 66.7µl of LPS (1000ng/mL) (*Escherichia coli* 0111:B4) and 66.7µl *S. pneumoniae* (serotype 14), at MOIs of 20 and 10, were added separately to duplicate wells The cells were incubated at 37°C in a 5% CO₂ atmosphere for 3 and 4 hours post stimulus. Supernatants were harvested and stored at -20°C. 333.33 µl of buffer RLT (Qiagen) was added to the cells and samples were stored at -80°C.

2.2.6 RNA isolation and cDNA reverse transcription

RNA was isolated from lysates stored at -80°C using the RNeasy[®] Plus Mini Kit (Qiagen, Hilden, USA). An additional DNAse treatment step was performed prior to adding of buffer RW1. RNA samples were eluted in 17µl of DNAse/RNAse free H₂O. 10µl of sample (with a minimum concentration of 10ng/µl RNA) was used as template for reverse transcription into cDNA using the SuperScript®VILO[™] cDNA synthesis kit (Invitrogen). In order to increase the yield of cDNA, the incubation time at 42°C was increased to 120 minutes. Following reverse transcription, each sample was quantified and diluted to a concentration of 800ng/µl, which was determined to be the optimal cDNA template concentration for quantitative PCR.

2.2.7 Genotyping

The *NFKBIA* SNPs rs2233406, rs3138053 and rs2233409 were ordered from Applied Biosystems as part of the TaqMan® Pre-Designed SNP Genotyping Assay. The context sequences of the SNPs are as follows:

Rs3138053	5'-ATTCGTTTATGCTATCTGACCTACA[C/T]TGTGCTCCCGCAGAA
	AAAGGATCGT-3'
Rs2233406	5'-TGGTGGTTGTGGATACCTTGCAATA[A/G]CAGAGTAGCTATTG
	TGTTCATAAGT-3'
Rs2233409	5'-TGTAATCCTGTCCCTCTGCAAGTGA[A/G]CCTTCTTTCCCTGG
	GGTTTCCCACG-3'

The primers for the coding SNP rs1050851 were custom designed. The primer sequences are as follows:

Forward	5'-AAGTGATCCGCCAGGTGAAG-3'
Reverse	5'-GCTGCAGGTTGTTCTGGAAGT-3'

The probe sequences to detect the rs1050851 major allele (C) and the minor allele (T) were as follows (the position at which the nucleotide change occurs is underlined):

Reporter1 (VIC) (C)	5'-ACCTGGC <u>C</u> TTCCTCA-3'
Reporter2 (FAM) (T)	5'-ACCTGGC <u>T</u> TTCCTCA-3'

The genotyping reagents were added in the following format:

Component	Volume (µl)
2X PerfeCta master mix	10 µl
Nuclease-free H ₂ O	9 µl
40X SNP assay mix	0.5 µl
gDNA (1-25 ng)	0.5 µl
Total volume	20 µl

Allele calling

The automated allele calling feature was used in accompanying sequence detection software, included with the Applied Biosystems 7300 Real Time PCR Machine.

2.2.8 NFKBIA gene expression

The forward and reverse primer sequences for *NFKBIA* were as follows

Forward	5'-TCAACAGAGTTACCTACCAGGGCT-3'
Reverse	5'-TCCTCTGTGAACTCCGTGAACTCT-3'

The forward and reverse primer sequences for ACTB (β -actin) were as follows

Forward	5'-GTTGCGTTACACCCTTTCTT-3'
Reverse	5'-ACCTTCACCGTTCCAGTTT-3'

Quantitative PCR was performed on cDNA using SYBR green reagent. The PCR reagents were added in the following format:

Component	Volume (μl)
2X SYBR green	12.5 µl
RNAse and DNAse free H ₂ O	10.5 µl
Forward and reverse primers	0.5 μl(5μM concentration) for each primer
Template (cDNA) concentration	1µl (800ng/mL)
Total volume	25 µl

The reaction setup for *NFKBIA* genotyping and gene expression in the Applied Biosystems 7300 Real Time PCR Machine was as follows:

	Cycle 1	Cycle 2	Cycle 3		
	HOLD	HOLD	CYCLE (40 cycles)		
			Denature	Anneal/Extend	
Temperature	50°C	95°C	95°C	60°C	
Time	2 min	10 min	15 sec	1 min	

The fluorescence was measured at cycle 3, step 2 of the reaction.

2.2.9 Genotyping for allelic imbalance

cDNA from PBMCs of individuals heterozygous or homozygous (major allele) for the promoter SNPs rs3138053/rs2233406/rs2233409 and the coding SNP rs1050851 was genotyped for all conditions (unstimulated, stimulated with LPS for 3 hours, stimulated with LPS for 4 hours, stimulated with *S. pneumoniae* for 3 hours, and stimulated with *S. pneumoniae* for 4 hours). Genotyping for each condition was performed in triplicate. Genomic DNA for each donor was also genotyped in triplicate. Detailed descriptions of assay design, validation as well as calculation of allelic imbalance is described in chapter 3.

2.3 Neonatal innate immune response

The neonatal immune response analysis was done on neonatal cord blood by the Kollmann lab. For information regarding subject recruitment, blood isolation, processing, and cytokine analysis refer to the publication by Kollmann *et al (91)*. Briefly, cord blood from healthy, full-term elective Caesarian sections without labour was collected directly into sodium heparin-containing vacutainers (BD Biosciences). PBMCs were isolated by density gradient centrifugation. 96 well plates containing 1.3 µl of TLR ligands (PAM₃CSK₄ (TLR2/1); poly(I:C) (TLR3); 0111:B4 LPS (TLR4), 3M-002 (TLR8), 3M-003 (TLR7/8); 3M-013 (TLR7); CpG (A type); RFSL (TLR 2/6)) were prepared. 180 µl of cell suspension (monocyte mixed with 1:1 RPMI 1640) was added to each TLR plate and the plates were incubated for 6 hours at 37°C. Supernatants were thawed at room temperature, and filtered into a clean 96 well plate. The Luminex assay was performed using the Upstate/Millipore "Flex Kit." Genomic DNA was extracted from cord samples of 20 neonates using the QIAamp ® DNA Blood Mini Kit.

2.4 Statistics

2.4.1 Statistical analysis for allelic imbalance

A 95% confidence interval of the mean allelic ratio was taken for each of the 5 conditions (unstimulated, stimulated with LPS for 3 hours, stimulated with LPS for 4 hours, stimulated with *S. pneumoniae* for 3 hours and *S. pneumoniae* for 4 hours).

2.4.2 Statistical analysis for gene expression

A non-parametric Mann-Whitney-test was used to compare the mean gene expression values of *NFKBIA* relative to the *ACTB* for 3 and 4 hour stimulations with LPS and *S. pneumoniae*.

2.4.3 Statistical analysis for innate immune response

A two-way ANOVA was used to analyze mean cytokine response for each ligand concentration with a Bonferroni post-test to account for multiple testing.

2.5 Order of operations

The order in which all the major steps in the experimental protocol were performed is outlined in Fig. 2.3.

Table 2.1. Data on the number of individuals and variants in different population ethnicgroups examined for LD analysis of *NFKBIA* variants

		No of	No.of	
Population source	Geographical origin	individuals	variants	Genotyped by
Northern and Western European	Utah (USA)	30	6	НарМар
Northern and Western European	Utah (USA)& France	23	29	PGA
African	Ibadan, Nigeria	30	7	НарМар
African	South-Western USA	24	29	PGA
Asian (Japanese)	Tokyo, Japan	45	6	НарМар
Asian (Japanese)	Tokyo, Japan	45	3	PGA
Asian (Han chinese)	Beijing, china	45	6	НарМар
Asian (Han chinese)	Beijing, china	45	3	PGA

Thes	e data were o	downlo	aded	on January	2 nd , 2010 a	and were dow	nloa	ided aga	ain on I	March 21 st ,
2010	to record an	y new	SNP	genotypes	for the ab	ove population	ons.	No nev	v SNP	genotypes
were	observed	in	the	second	download	relative	to	the	first	download.



Figure 2.1. Streptococcus pneumoniae serotype 14 shows a reduced lag time for exponential growth when grown in media BHI media supplemented with 5% FCS. Growth of *S. pneumoniae* was monitored in cultures started from overnight cultures (A) and single colonies (B) for 10 hours under anaerobic conditions at 37° C with 5% CO₂



Figure 2.2 Determining the optimal growth range for *Streptococcus pneumonia* **serotype 14.** A single colony of *S. pneumoniae* serotype 14, was subcultured twice and grown in BHI medium supplemented with 5% FCS for 10 hours. Plate counts were taken for an OD increase of 0.1. Data from two growth curves is shown (A). CFU/ml were plotted against OD readings to yield a standard curve for determining colony counts for a given OD (B).



Figure 2.3. Overview of experimental protocol. Experimental procedures were performed as outlined in the Figure. All assays were done in healthy adults, with the exception of the innate immune response assay which was done in healthy neonates.

Chapter 3: Results

3.1 Emerging picture of the NFKBIA promoter haplotype

3.1.1 Analysis of variants

As both the type of variation and the location of the variant in a gene determine the subsequent impact on gene function, it was important to get an understanding of the nature and location of variants in the *NFKBIA* gene. Table 3.1 shows the location of variants of *NFKBIA* in different ethnic groups. In the European panel, of the 28 variants genotyped by HapMap and PGA, there were only 2 synonymous-coding variants (rs 1050851 and rs1957106). The remainder of the variants consisted of 13 variants in the intron, 2 in the 5' UTR, 5 in the 3' UTR and 6 variants in the promoter region. The African panel consisted of 30 variants (3 exonic, 15 intronic, 0 in the 5' UTR, 5 in the 3' UTR, and 7 promoter variants). Both the CHB and JPT panels consisted of 9 variants (2 exonic, 1 intronic, 0 in the 5' UTR, 3 in the 3' UTR, and 3 promoter variants).

3.1.2 Linkage disequilibrium(LD)analysis

As LD has been critical for mapping complex disease variants in the human genome, it was important to study the *NFKBIA* LD pattern in different ethnic groups. In our analysis of the entire *NFKBIA* gene, we discovered LD was not continuous but limited to mostly non-coding variants in the gene (appendix: fig. A1-4). We categorized variants that showed strong LD and discovered that 75-95% of variants that displayed strong LD resided in non-coding regions of *NFKBIA* which included the promoter, introns, and the 3' UTR (appendix: fig. A5a). Next, we determined the percentage of variants-within these three sites- that showed strong local-linkage i.e., we <u>excluded</u> variants in these sites that didn't show strong LD with a variant within the same region but showed strong LD with a variant outside the region. In the 4 populations examined, 52-100% of the variants in the promoter showed strong local LD whereas 28-65% of the variants in the 3' UTR showed strong LD (appendix: fig. A5b).

3.1.3 Promoter haplotype analysis

Since the promoter variants were in strong LD, we examined this region exclusively within each of the 4 populations (fig 3.1a-d). We determined that 2 promoter variants (rs2233406, and 3138053) comprised a small, strongly-linked 2-variant haplotype in all 4 populations. The European panel (fig. 3.1a) was comprised of a 4-variant haplotype consisting of rs2233409, rs2233406, and rs3138053 in strong LD, and a fourth variant rs11569591-an 8bp insertional/deletion (structural) variant- that displayed 100% LD with rs2233406 and rs3138053.

The African panel (fig. 3.1b) had a second 2-variant haplotype consisting of SNPs rs11569593 and rs11569590 in 100% LD. The CHB panel (fig. 3.1c) had a 3 variant haplotype where the third variant, rs2233409, displayed 83% LD with rs2233406 and rs3138053.

3.2 Functional analysis of NFKBIA promoter variants

3.2.1 Putative transcription factor binding sites

In order to see if any of the *NFKBIA* promoter variants resided in functional sites for the binding of transcription factors we searched the entire *NFKBIA* gene sequence in the SNP@Promoter search engine (<u>http://variome.kobic.re.kr/SNPatPromoter/</u>) for any putative transcription factor binding sites. The search revealed that the variant rs3138053 that was found in all of the 4 populations, resided at the binding site for retinoic related orphan receptor 2 (RORa2). Previous have identified it as a binding site for RORa1 and RORa2 (*92*). Functional experiments using human cell lines have shown that RORa2 binds to this response element and transcriptionally up regulates IkBa expression, and mutagenesis in this region leads to a consequent decline in IkBa expression (*93*). There is a putative binding site AP-2 at an intronic variant (rs2233411) that was in 100% LD with the variant rs11569591 resides in a functional binding site for 2 transcription factors- KROX and Egr. Table 3.2, shows all the variants of *NFKBIA* that reside in putative transcription factor binding sites.

3.2.2 Risk haplotype prediction

Based on the strong LD observed in the promoter region of *NFKBIA* in all 4 populations, and the presence of putative transcription factor binding sites in this region (Table 3.2), we came up with a risk haplotype for the population groups examined (Fig. 3.2). The first risk haplotype was common in all populations and consisted of the risk variant rs3138053 which resided in the RORα1 and RORα2 binding site and was in strong LD with SNP rs2233406. The second haplotype was unique to individuals of European descent and consisted of 2 additional risk variants rs11569591 and rs2233411 in strong LD with SNPs rs3138053, rs2233406 and rs2233409, both of which resided in putative transcriptional factor binding sites. Therefore, we predicted that owing to the presence of the rs3138053 risk variant in all 4 populations, we would see a subsequent impact on allelic and mRNA expression of *NFKBIA*, but this effect would be more pronounced in individuals of European descent who have 2 additional risk variants in their haplotype. It is to be noted that the actual genotyping and functional validation i.e., analysis of *NFKBIA* allelic imbalance, *NFKBIA* gene expression and innate immune response was done for SNPs rs3138053, rs2233406 and rs2233409 as these SNPs had been shown to be associated with disease.

3.3 Allelic imbalance

3.3.1 Validation of allele specific PCR assay

In order to measure allelic imbalance we investigated whether: 1) we could detect differences in allele ratios using TaqMan probes and 2) what range of allele ratios were detectable using this technique. To answer these questions we selected a synonymous coding SNP rs1050851 which was in strong LD with promoter SNPs rs3138053, rs2233406 and rs2233409 and thus acted as a tag for differentiating between the major (ACC) and minor (GTT) allele transcripts of *NFKBIA*. Genomic DNA was extracted from a healthy individual who is homozygous for the *NFKBIA* rs1050851 coding SNP wild type allele "C/C" and from an individual who is homozygous for the minor allele "T/T." The concentration of genomic DNA was quantified and diluted to ensure that both samples had similar starting concentrations (187.82 \pm 2.03ng/µL for "C/C" and 182.79 \pm 3.68ng/µL). The genomic DNA from the samples was mixed in different concentrations to mimic allelic imbalance seen in cDNA samples during gene expression. The allelic ratios were 1:0.5, 1:1, 1:1.5, 1:2, 1:2.5, 1:3, 1:3.5, 1:4, 1:8 CC:TT and 8:1, 4:1, 3.5:1, 3:1, 2.5:1, 2:1, 1.5:1, 0.5:1 TT:CC. Each genomic mixture was genotyped for the rs1050851 SNP using the Applied Biosystems 7300 Real Time PCR Machine. Figure 3.3a illustrates this procedure.

The cycle at which amplification of the DNA template enters the exponential phase is referred to as the threshold of activation or "ct." We observed that as the ratio of the major allele to the minor allele was decreased from 8:1 (C/C:T/T) to 1:8(C/C:T/T), a corresponding increase in Δ ct was detected in the real time PCR read out for each ratio (Fig. 3.3b). We also observed a Δ ct value in a 1:1 ratio, which theoretically should be zero. This difference is a result of different probe binding efficiencies of the Vic and Fam probes for their respective alleles and has been reported in other allele specific PCR assays involving TaqMan probes (94). This background difference was subtracted from the observed " Δ ct" for all the mixtures of genomic DNA to obtain a corrected " Δ ct" reading for each sample.

3.3.2 Generation of allele ratios

The corrected " Δ ct" reading for each mixture was plotted against the logarithm of the known ratio of CC:TT for each sample to obtain a standard curve (Fig. 3.4). This standard curve was used to determine the allelic ratios in cDNA samples for which the allelic ratio was not known.

3.3.3 Determining allelic imbalance in NFKBIA rs1050851 heterozygotes

To determine if *NFKBIA* heterozygotes (ACC/GTT) for promoter SNPs rs3138053, rs2233406 and rs2233409 displayed allelic imbalance, we stimulated PBMCs from *NFKBIA* heterozygotes (ACC/GTT) who were also heterozygous for the coding SNP rs1050851 (C/T) with LPS (100ng/ml) and *S. pneumoniae* (moi 7.8-30) for 3 and 4 hours. The cDNA along with the

genomic DNA for each individual was genotyped in tandem for the rs1050851 SNP which was used as a 'tag' for genotyping the 3 promoter SNPs that could not be assayed for in cDNA (for more details on this, refer to section entitled 'the problem of quantitating promoter variants' in chapter 1). The Δ Ct read out for the genomic samples was subtracted from the Δ Ct readout each of the three conditions (unstimulated, stimulated for 3 hours, and stimulated for 4 hours) to obtain a 'corrected Δ Ct' (Δ Ct') read out. Using the standard curve for calculating unknown allele ratios we determined the allele ratio for each condition.

In the unstimulated condition, the mean allele transcript ratio of the major allele (ACC) to the minor allele (GTT) was 1.21 (1.15-1.27 95% CI). In the presence of LPS (100 ng/ml), the mean allelic ratio for 3 and 4 hour stimulation was 1.22 (1.18-1.25 95% CI) and 1.21 (1.15-1.27 95% CI) respectively (Fig.3.5a). The ratios for each condition were not significantly different from each other, but were significantly higher than the base line 1:1 ratio seen in the genomic sample. Similarly, in the presence of *S. pneumoniae* the ratios for the 3 hour and 4 hour stimulation were 1.23 (1.16-1.31 95% CI) and 1.25 (1.17-1.33 95% CI) respectively (Fig. 3.5b). Since 11 out of the 12 individuals analyzed were of European descent, we also analyzed the mean allele ratios for these 11 individuals. In the one individual of Chinese (Han) descent, the allele ratios for unstimulated, LPS 3 and 4 hour stimulated, *S. pneumoniae* 3 and 4 hour stimulated conditions were 1.11, 1.22, 1.26, 1.12, 1.13 respectively. With the exclusion of this sample, the mean allele ratios were 1.22 (1.15-1.29), 1.22 (1.18-1.26), 1.21 (1.14-1.27), 1.24(1.17-1.32), 1.26 (1.18-1.34) respectively. Therefore, by the use of a validated allele specific PCR assay, we were able to detect allele specific differences in individuals who were heterozygous for SNPs rs3138053, rs2233406, rs2233409 and rs1050851.

3.4 Determining impact of allelic imbalance on NFKBIA mRNA expression

3.4.1 Impact of allelic imbalance on NFKBIA mRNA expression in NFKBIA homozygotes and heterozygotes

To investigate whether allelic imbalance had a subsequent downstream impact on *NFKBIA* expression, we measured *NFKBIA* expression relative to the housekeeping gene *ACTB* (β -*actin*) at 3 and 4 hours post stimulus in the presence of LPS and *S.pneumoniae*. Initially, we examined *NFKBIA* expression in all individuals who were heterozygous or homozygous for the promoter SNPs rs3138053/rs2233406/rs2233409. This included individuals of all ethnicities (Fig. 3.6a-d). Although no significant differences were seen in *NFKBIA* expression, a pattern was seen where *NFKBIA* homozygotes (ACC/ACC) displayed a higher fold expression of *NFKBIA* than *NFKBIA* heterozygotes (ACC/GTT); this pattern was more evident at 3 hours (*p*=

0.057and p=0.073 for stimulation with LPS and *S.pneumoniae* respectively) (Fig.3.6a and 3.6c) than 4 hours (p= 0.14 and p=0.10 respectively) (Fig.3.6b and 3.6d).

3.4.2 Impact of allelic imbalance on NFKBIA mRNA expression in NFKBIA homozygotes and heterozygotes of European descent

We had hypothesized earlier that owing to the presence of an 8 insertion/deletion polymorphism variant (rs11569591), and an intronic variant (2233411) in strong LD with the promoter SNPs rs3138053/rs2233406 in individuals of European descent, we would see more profound differences in mRNA expression of *NFKBIA* in heterozygotes or homozygotes (for the minor allele) of this group than homozygotes for the major allele. We observed that when only examining individuals of European descent at 3 hours a 1.4 fold greater expression of *NFKBIA* was seen in homozygotes (ACC/ACC) than heterozygotes (ACC/GTT) for LPS (*p*=0.0095, Mann Whitney-U-test) and a 1.51 greater expression was seen in samples stimulated with *S.pneumoniae* (*p*=0.024, Mann-Whitney-U-test) (Fig 3.7a and 3.7c). A similar pattern was seen at the 4 hour time point. However the differences were not statistically significant (Fig 3.7b and 3.7d).

3.5 Determining impact of allelic imbalance on the innate immune response

NFKBIA encodes for IκBα which negatively regulates NF-κB translocation. NF-κB positively regulates the expression of proinflammatory cytokines genes *TNFα* and *IL6*. A readout of proinflammatory cytokines is frequently used in a clinical immunology setting to determine the level of inflammation in an individual. However, when trying to specifically understand the genetic contribution to an individual's phenotype, it is important to minimize any confounding effects to the cytokine readout that could result from environmental effects such as an infection or an autoimmune condition or the interaction of the adaptive immune system with innate immune system, as is the case in adults. To circumvent this problem, the innate immune analysis was conducted on PBMCs from cord blood belonging to healthy infants born at BC Children's Hospital. Healthy infants were chosen over healthy adults as infants display poor immunological memory and predominantly rely on their innate immune system for protection against pathogens (*95*).

3.5.1 Impact of allelic imbalance on the innate immune response in NFKBIA homozygotes and heterozygotes

To determine if there was a difference in the innate immune response between *NFKBIA* homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) PBMCs of these genotypes were stimulated with a variety of Toll Like Receptor (TLR) ligands for 6 hours and the innate immune

response was measured. A significantly higher response was seen for TNF α in PBMCs of heterozygotes (ACC/GTT) than homozygotes (ACC/ACC) when stimulated with TLR ligands LPS (1.80 fold, *p*<0.05 at a dose of 100ng/ml), Pam3CSK4 (3.20 fold, *p*<0.01 at a dose of 100 ng/ml, 2.76 fold , *p*<0.01 at a dose of 100ng/ml), 3M003 (1.63 fold, *p*<0.01 at a dose of 10 µM), 3M002 (2.39 fold, *p*<0.001 at a dose of 10.0 uM), and RFSL (3.3 fold, *p*<0.05 at a dose of 1000ng/ml, 2.36 fold *p*<0.01 at a dose of 10000ng/ml) (Fig.3.8 (a-e)).

3.5.2 Impact of allelic imbalance on the innate immune response in NFKBIA homozygotes and heterozygotes of European descent

To determine if reduced mRNA expression of *NFKBIA* in heterozygotes (ACC/GTT) of European descent led to an exacerbation of the innate immune response, we analyzed TNF α production between homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) of European origin. Significantly higher fold production of TNF α was seen in PBMCs of heterozygotes (ACC/GTT) than homozygotes (ACC/ACC) in the presence of LPS (1.57 fold , *p*<0.05 at a dose of 100ng/ml), Pam3CSK4 (2.29 fold, *p*<0.01 at a dose of 100ng/ml and 1.91 fold, *p*<0.05 at a dose of 1000 ng/ml), 3M003 (1.79 fold, *p*<0.001 at a dose 10µM) and 3M002 (3.30 fold, *p*<0.001 at a dose of 10µMI) (Fig. 3.9 (a-d)). However, these differences did not significantly differ in magnitude from the differences seen between homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) of mixed descent. Analysis for stimulation with RFSL could not be done as sufficient samples of European descent were not available (3 homozygotes (ACC/ACC) versus 5 heterozygotes (ACC/GTT)) for stimulation with this ligand in order to make statistically meaningful comparisons.

3.5.3 Impact of allelic imbalance on pathways that do not involve IkBa degradation

In order to see if the observed differences in cytokine production between homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) were not chance occurrences, we compared differences for pathways that do not require IkBa degradation. TLR9, in addition to activating the classical signaling pathway uses a non IkBa pathway to produce interferon alpha (IFNa) (96). In the presence of the TLR9 ligand CpGA, both homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) did not show any differences in IFNa production (Fig. 3.10a). Similarly, homozygotes (ACC/ACC) and heterozygotes (ACC/ACC) and heterozygotes (ACC/GTT) of European descent also did not show any differences in IFNa production (Fig. 3.10b).

Region	Europe	Africa	CHB	JPT
promoter	6	7	3	3
5'utr	2	0	0	0
exon	2	3	2	2
intron	13	15	1	1
3'utr	5	5	3	3
total	28	30	9	9

Table 3.1. Region specific location of variants in NFKBIA



Figure 3.1. Linkage disequilibrium (LD) plot of the *NFKBIA* promoter for the 4 major populations. LD plots correspond to European, African, Chinese (Han) of Beijing (CHB), and Japanese individuals of Tokyo(JPT). Variant information was downloaded from the genome variation server database and uploaded into Haploview 3.1. The r^2 are represented by a black to white gradient where black corresponds to 100% and white corresponds to 0-4% linkage between two SNPs.

Ethnicity	Variant rs	Туре	Position	Transcription Factor Binding Site	Linkage with rs2233409	Linkage with rs2233406	Linkage with rs3138053
JPT	3138053	A/G	-894	ROR alpha2	88	91	n/a
CHB	3138053	A/G	-894	ROR alpha2	83	100	n/a
CEU	2233411	A/T	+394	AP-2	100	82	82
CEU	11569591	-/GGAGGGGG	-599	KROX/Egr	79	100	100
CEU	3138053	A/G	-894	ROR alpha2	86	100	n/a
YRI	2233411	A/T	+394	AP-2	100	45	46
YRI	3138053	A/G	-894	ROR alpha2	36	100	n/a

 Table 3.2. Putative transcription factor binding sites occupied by NFKBIA variants

GT haplotype

European/African/CHB/JPT



Figure 3.2. Visual representation of functional binding sites for transcription factors of *NFKBIA.* Variants that show 100% linkage with rs3138053 are shown. <u>GGAGGGGGG</u> corresponds to the major (insertion) allele for the structural variant rs11569591.



Figure 3.3. Validating *NFKBIA* **allele specific PCR assay.** Genomic DNA from *NFKBIA* coding SNP (rs1050851) major (CC) and minor (TT) allele homozygotes was mixed in different proportions of the C allele to the T allele (A). The change in allele ratios corresponded to a change in the Δ Ct values when each mixture was genotyped for the rs1050851 SNP (B). The major allele is "CC" is represented by the blue curve and the minor allele "TT" is represented by the red curve.



Figure 3.4. Standard curve for detecting allelic imbalance in *NFKBIA* heterozygotes. The Δ Ct value for 1:1 ratio of a *NFKBIA* 1050851 heterozygote was subtracted from the Δ Ct value for each allelic ratio mixture genotyped to obtain a corrected Δ Ct' value. The log of each allele ratio was plotted against Δ Ct to yield a standard curve.



Figure 3.5. NFKBIA heterozygotes (ACC/GTT) display allelic imbalance. Peripheral blood mononuclear cells of individuals heterozygous (ACC/GTT) for the NFKBIA promoter SNPs (rs3138053/rs2233406/rs2233409) and the synonymous coding SNP (rs1050851) were stimulated with 100ng/ml of LPS (A) and Streptococcus pneumoniae (serotype 14) (B) for 3 and 4 hours. cDNA from unstimulated PBMCs and PBMCs stimulated for 3 and 4 hours was genotyped for the NFKBIA coding SNP rs1050851. Values represent means with 95% confidence intervals for 12 individuals. 11 out of 12 individuals assay were of European descent. А ratio of 1:1 represents allele ratio observed DNA. in genomic



Β

Figure 3.6. *NFKBIA* heterozygotes display a trend of reduced *NFKBIA* mRNA expression than *NFKBIA* homozygotes. Peripheral blood mononuclear cells of individuals who were either homozygous (ACC/ACC) or heterozygous (ACC/GTT) for *NFKBIA* promoter SNPs rs3138053/2233406/2233409 were stimulated with 100ng/mL of LPS (A and B) or *Streptococcus pneumonia* serotype 14 (C and D), for 3 and 4 hours. *NFKBIA* expression was determined by quantitative PCR. Relative expression was analyzed by $2^{-\Delta \Delta Ct}$ using β -actin as a reference gene. Values represent means ±SEM of 16 homozygotes (dark bars) and 16 heterozygotes (grey bars). Statistical analysis was performed by the non-parameteric Mann-Whitney-test.



Figure 3.7. *NFKBIA* mRNA expression is significantly reduced in *NFKBIA* heterozygotes of European descent. Peripheral blood mononuclear cells of individuals of European origin who were either homozygous (ACC/ACC) or heterozygous (ACC/GTT) for *NFKBIA* promoter SNPs rs3138053/2233406/2233409 were stimulated with 100ng/mL of LPS (A and B) or *Streptococcus pneumonia* serotype 14 (C and D), for 3 and 4 hours. *NFKBIA* expression was determined by quantitative PCR. Relative expression was analyzed by $2^{-\Delta \Delta Ct}$ using β -actin as a reference gene Values represent means ±SEM of 11 homozygotes (dark bars) and 13 heterozygotes (grey bars). Statistical analysis was performed by the non-parameteric Mann-Whitney-test.



Figure 3.8. *NFBIA* heterozygotes display a higher TNF α response. Peripheral blood mononuclear cells of healthy neonates were stimulated with various Toll like receptor (TLR) ligands for 6 hours and the TNF α response was measured using the luminex platform. Responses were compared between homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) for *NFKBIA* promoter SNPs rs3138053/rs2233406/rs2233409. Responses for TLR ligands LPS, Pam3CSK4, 3M003, 3M002 and RFSL (A, B, C, D, E) are shown. Values represent means ±SEM of 5-8 heterozygotes (open circles) and 12 homozygotes (black squares). A two-way ANOVA was performed with a post-test.*, p<0.05, **,p<0.01, ***p<0.001.



Figure 3.9. *NFBIA* heterozygotes of European origin also display a higher TNF α response. Peripheral blood mononuclear cells of healthy neonates were stimulated with various Toll like receptor (TLR) ligands for 6 hours and the TNF α response was measured using the luminex platform. Responses were compared between homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) for *NFKBIA* promoter SNPs rs3138053/rs2233406/rs2233409. Responses for TLR ligands LPS, Pam3CSK4, 3M003, and 3M002 (A, B, C, D) are shown. . Values represent means ±SEM of 5-7 heterozygotes (open circles) and 5-10 homozygotes (black squares). A two-way ANOVA was performed with a post-test.*, p<0.05, **,p<0.01, ***p<0.001.



Figure 3.10. Pathways that do not require $I\kappa B\alpha$ signaling are not affected by SNPs in the *NFKBIA* promoter. Homozygotes (black squares, n=12) and heterozygotes (open circles, n=8) of mixed descent (A) as well as homozygotes (black squares, n=10) and heterozygotes (open circles, n=7) of European descent (B) did not show any differences in IFN α production.

Chapter 4: Discussion

4.1 Overview

In this project we analyzed the functional impact of minor variants of SNPs rs2233406, rs2233409 and rs3138053 in the *NFKBIA* promoter. Using LD and functional analysis we identified a 2 variant risk haplotype (consisting of SNPs rs3138053 and rs223406) which was common in the 4 major ethnic groups examined and a 5 variant risk haplotype (consisting of rs3138053, rs2233406, rs2233409, rs11569591 and rs2233411) which was unique in individuals of European descent. Subsequently using a tag SNP (rs1050851) that was in strong LD with SNPs rs2233406, rs2233409 and rs3138053 we were able to detect allelic imbalance in PBMCs of heterozygotes. This was followed by a measure of *NFKBIA* expression wherein heterozygous individuals of European descent showed reduced expression of *NFKBIA* in comparison to homozygotes. Analysis of neonatal innate immune responses revealed that *NFKBIA* heterozygotes displayed greater secretion of TNF α than homozygotes in response to various TLR ligands. In this chapter I will be reviewing the techniques utilized in this study, the feasibility of future follow up experiments and how the results from this project may be applicable to a clinical setting.

4.2 Allele specific PCR

4.2.1 Choice of tagging SNP

We validated and applied an allele specific PCR assay to investigate allelic imbalance in the *NFKBIA* promoter. The advantages of working with a candidate gene such as *NFKBIA* is that it only has 3 synonymous SNPs in the coding region: rs1050851, rs2233411, and rs1957106 that are present at a minor allele frequency of greater than 1 percent. rs2233411 has a minor allele frequency in individuals of African descent and no known minor allele frequency in individuals of other ethnicities (dbSNP). It shows poor linkage (less than 5%) with promoter variants in all the 4 major population groups examined. rs1957106 has minor allele frequencies of 30.4%, 22.5%, 18.2%, and 25.0% in individuals of European, African, Chinese (Han) and Japanese descents respectively. However, rs1957106 shows poor linkage (less than 35%) with promoter SNPs in *NFKBIA* for all the ethnic groups examined. rs1050851 has a minor allele frequencies of 24.1% in individuals of European descent and minor allele frequencies of 0.9%, 2.3%, 4.7% for individuals of African, Chinese (Han) and Japanese descents, within the *NFKBIA* gene in individuals of European descent. It shows strong linkage with 5 variants, within the *NFKBIA* gene in individuals of European descent, three of which are promoter SNPs (rs2233406, rs2233409, rs3138053), one is a promoter structural

variant (rs1156951) and one is an intronic variant (rs 2233411). rs1156951 and rs2233411 are in strong LD with the promoter SNPs rs2233406 and rs3138953, and were shown to occupy putative transcription factor binding sites (refer to results). Strong linkage of rs1050851 with promoter variants is not seen in other ethnic groups (African, CHB, JPT) examined in this study. This was also evident from our genotyping results where 11 out of 12 individuals who were heterozygous for SNPs rs3138053, rs2233406, rs2233409 and rs1050851 were of European descent. Since it is possible that rs1050851 might be in strong linkage with variants outside the *NFKBIA* gene, using the Genome Variation Server, we searched 20,000 base pairs upstream and 20,000 bp down stream of rs1050851 for any variants that were in strong linkage with rs1050851. The greatest linkage (55.8%) was seen for variants rs10148482 and rs3138045. No data on Hap Map is available about the location of these SNPs relative to the *NFKBIA* gene nor have these SNPs been identified to occupy any putative binding sites. Therefore, rs1050851 is a valid choice for a tag in measuring allelic imbalance of *NFKBIA* transcripts in *NFKBIA* promoter heterozygotes.

4.2.2 Precision of assay

Our standard curve gave us the precision to measure allelic imbalance from 1:0.5 (CC:TT or TT:CC) to 1:8 (CC:TT or TT:CC) for rs1050851. Our measured allelic imbalance values ranged between 0.94 to 1.49 (CC:TT) with the median value ranging from 1.21 to 1.24 for each of the 5 conditions (Unstimulated, stimulated with LPS for 3 hours, stimulated with LPS for 4 hours, stimulated with *S. pneumoniae* for 3 hours and stimulated with *S. pneumoniae* for 4 hours). Therefore our measured values fall within the known range of the standard curve. The only other measure of allelic imbalance in an innate immunity gene was done by Cambino *et al.* for an exonic SNP (rs352140) in TLR9 where heterozygous individuals showed allelic imbalance values ranging from 0.5-1.87 of the major allele relative to the minor allele (97). The assay was done using an Allelotype platform, which utilized mass spectrometry based nucleotide acid analysis for genotyping in lymphoblastoid cell lines (LCLs). The next section discusses the potential issues associated with selecting cell types for allelic imbalance studies.

4.3 Choice of cell types

In this study we performed *NFKBIA* allele specific and gene expression experiments on primary cells. Most allele specific PCR assays have been done using lymphoblastoid cell lines (LCLs) immortalized using Epstein Barr Virus (*98, 99*). The advantages of working with LCLs is that they provide a fixed environment for investigating differences in gene expression (*80*). However, such a system is not representative of *in vivo* biology and does not take into account epigenetic factors that might influence gene expression (*100*). The allelic imbalance (AI) observed in clonal

cell lines might in part result from a greater preponderance of monoclonal allelic expression resulting from epigenetic imprinting (101). Therefore, in a LCL population of cells the greater prevalence of paternal or maternal allele expressing clones may result in a false positive result of allelic imbalance. A recent study has shown that as much as 20% of an LCL derived population might be affected by monoclonal allelic expression (102). Primary cells offer the advantage of being experimentally unmanipulated and the results obtained from performing experiments are more reflective of *in vivo* biology.

4.4 Analysis of variation in gene expression

4.4.1 Ligand and pathogen specific differences

NFKBIA allele specific and mRNA expression was examined using LPS and Streptococcus pneumoniae (serotype 14). We noted that mean allelic imbalance values and standard deviations for PBMCs that were stimulated with S. pneumoniae (1.23±.0.11(SD), 1.25±0.12(SD) for 3 and 4 hours respectively) were nominally higher than PBMCs that were stimulated with LPS (1.22±0.055(SD), 1.22±0.093(SD) for 3 and 4 hours respectively). However there were more pronounced differences in mean values and standard deviations when examining NFKBIA mRNA expression. Homozygote (ACC/ACC) and heterozygote (ACC/GTT) PBMCs stimulated with LPS for 3 hours had NFKBIA expression values of 2.106±0.497 (SD) and 2.954±0.936 (SD), P= 0.0094 with LPS while with S. pneumoniae the values were 2.402±0.967 (SD) and 3.629±2.182 (SD), P= 0.0244 respectively. Similarly for the 4 hour stimulation homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) stimulated with LPS had NFKBIA expression values of 2.499±.0882 (SD) and 2.748±0.598 (SD), P=0.115 while stimulation with S.pneumoniae yielded values of 2.865±1.095 (SD) and 3.622±1.235 (SD), P= 0.0524). Similar differences were seen when examining expression between mixed ethnicities. The greater mean expression of NFKBIA and standard deviation for samples stimulated with S. pneumoniae can probably be attributed to the varying level of MOIs of S. pneumoniae given to PBMCs. Since we were working with a live pathogen, our goal was to limit the range MOI stimulation between MOIs 10-15. But in practice, judging from plate counts, we noted the MOIs varied between 7.8 to 30.

4.4.2 Inter-individual differences

Inter-individual differences in gene expression from peripheral mononuclear cells (PBMCs) resulting from non-genetic factors such as diet, age and gender do exist (*103, 104*). In our study the mean allelic imbalance readouts with the standard deviations (SD) for all 12 individuals examined were 1.21±0.1 (unstimulated), 1.22±0.055 (LPS stimulated for 3 hours), 1.22±0.093 (LPS stimulated for 4 hours), 1.23±.0.11 (*S. pneumoniae* stimulated for 3 hours), and 1.25±0.12 (*S. pneumoniae* stimulated for 4 hours). We were not able to detect any differences in allelic

imbalance between unstimulated and stimulated samples. However, we did show that allelic imbalance is prevalent in all samples. This implies that the AI seen for *NFKBIA* promoter variants exists at base line levels and is not stimulus driven. It is to be noted that our purpose was to *detect* allelic imbalance in the *NFKBIA* promoter. In order to get an absolute quantification of allelic imbalance in stimulated and unstimulated conditions a larger set of individuals might be needed to detect such subtle differences, if they exist.

4.4.3 Ethnic differences

We also noted the NFKBIA mRNA levels in individuals of European descent were significantly lower in NFKBIA heterozygotes than homozygotes at the three hour time point (2.106±0.497 (SD) vs 2.954±0.936 (SD), P= 0.0094 with LPS and 2.402±0.967 (SD) vs 3.629±2.182 (SD) S. pneumoniae, P= 0.0244) than the four hour time point (2.499±.0882 (SD) vs with 2.748±0.598 (SD), P=0.115 with LPS and 2.865±1.095 (SD) vs 3.622±1.235 (SD), P= 0.0524). We had originally hypothesized this effect based on the presence of an insertional/deletion risk variant rs11569591 and an intronic variant rs2233411 both of which were in strong LD with the rs3138053 risk variant that. No previous association studies of NFKBIA involving these variants have been done. However, a 4 base pair insertion/deletion -94(ATTG) polymorphism (rs28362491) that is located between two putative key promoter regulatory elements of the NFKB1 gene-that encodes for the NF-kB1 protein-sub unit of the NFKB1/p50 and NFKB1/p105 heterodimers has been associated with the risk of increased risk of ulcerative colitis, psoriasis, nasopharyngeal carcinoma, oral squamous cell carcinoma, cervical squamous cell carcinoma, superficial bladder cancer and prostate cancer (105-111). NFKB1 promoter-luciferase reporter plasmid constructs containing the -94delATTG, which were transfected in Hela or HT-29 cell lines, showed reduced promoter activity when compared to constructs containing the

-94insATTG allele (*105*). The rs11569591 -599(-/GGAGGGGG) polymorphism has a minor allele frequency of 28.7% in individuals of European descent and 16.7% in individuals of African descent where the minor allele is the insertion allele and the major allele is the deletion allele (dbSNP). Based on the results obtained from association and functional studies done with the 4 base -94delATTG polymorphism in *NFKB1*, it is highly plausible that the associated 8 base rs11569591 -599(-/GGAGGGGG) polymorphism in the *NFKBIA* promoter in combination with rs3138053 can potentially further impact promoter activity, leading to a more reduced *NFKBIA* expression level as was seen in populations where this polymorphism was commonly prevalent.

No studies to date have compared expression levels between individuals homozygous and heterozygous for either of the 4 SNPs (rs3138053, rs2233406, rs2233409, and rs1050851) examined in this study. However, it has been previously shown that the minor allele (GTT)

haplotype has been associated with the risk of sarcoidosis; in a separate study, untreated and treated patients with sarcoidosis show levels of NF-κB protein, in the nuclear extract of PBMCs, that are twice as high than healthy controls (*55, 112*).

4.4.4 Artefactual differences

Artefactual differences in mRNA expression can occur if a *cis* acting polymorphism is present on the primer target sequence of the gene (*113*). The primer pair used for measuring *NFKBIA* mRNA expression amplifies a 173bp sequence in exon 3 of the gene. All exonic polymorphisms map to exons 1,2, and 6 but not exon 3 in *NFKBIA* (http://www.ncbi.nlm.nih.gov/SNP). Similarly, for the house keeping gene *ACTB* (beta actin), the primer sets used amplify a 147 bp sequence in exon 6. There is a synonymous polymorphism in this region, rs13447409, but it occurs at a low minor allele frequency (0.6% in a mixed population).

4.5 Choice of ligand and time course selection

LPS was chosen as a prototypical proinflammatory stimuli for studying NFKBIA expression. Even though from the perspective of NF- κ B signaling, any ligand for a receptor system that culminates in the activation of the classical NF- κ B pathway would be suitable for stimulation. differences exist between the specific downstream effects of ligands on the NF-kB/lkBa autoregulatory loop. Studies using mice embryonic fibroblasts (MEFs) have shown that TNFa mediated activation of NF-kB involves a negative feedback loop mediated by IkBa, whereas LPS and pathogenic mediated activation does not involve a negative feedback loop (114). In a study done by Covert et al., when MEFs were stimulated with LPS over a 3 hour time course, the IkBa protein level decreased and remained consistently low, while IkBa mRNA transcript levels increased and remained consistently high (115). This finding is partially consistent with our trial experiments with LPS stimulation of human PBMCs, where cells were stimulated with 100ng/ml of LPS for 1,2,3 and 4 hours. We noticed that the NFKBIA expression peaked at the 3 hour time point and remained consistent at the 4 hour time point. In the presence of TNF α , NFκB shows damped oscillations in activity which has been shown to be a consequence of the negative feedback inhibition by IkBa. In contrast NF-kB activation is stable in the presence of LPS (116). Taken together this implies that in response to an LPS stimulation, there is proteasomal degradation of IκBα that allows subsequent binding of NF-κB to its response elements on NFKBIA leading to an up-regulation of NFKBIA transcription. However, this stimulus does not lead to the restoration of the inhibiton of $I\kappa B\alpha$ on NF- κB . From the viewpoint of examining the functional differences of promoter variants in the promoter of NFKBIA, a LPS stimulus is more useful, as it has allowed us to see differences in NFKBIA mRNA expression at time points were NFKBIA expression is consistent as opposed to a TNF α induction model where *NFKBIA* expression shows an oscillatory behaviour (*116*). Therefore when studying the immediate impact of regulatory variation on *NFKBIA* expression the LPS stimulus is more useful, whereas for studying the impact of regulatory variation on the NF- κ B/I κ B α autoregulatory loop a TNF α stimulus would be more relevant.

4.6 Choice of CpGA mediated TLR9 secretion of IFNα as a non-IκBα requiring pathway

Type 1 interferons have key roles in anti viral and anti proliferative responses. Specifically, IFN α and IFN β bind to the interferon receptor and initiate the Janus kinase/signal transducer of transcription (JAK-STAT) signaling cascade. This leads to the transcription of host genes involved in the inhibition of viral replication (*117*). Activation of TLRS 3, 4, 7, 8, and 9 results in the production of type 1 interferons. TLRs 3 and 4 activate the gene encoding IFN β via the TIR-domain-containing adapter-inducing interferon- β -Interferon regulatory factor 3 (TRIF-IRF3) complex while TLRs 7, 8, 9 activate the gene encoding IFN α via the Myd88-IRF7 complex (*118*). The expression of the IFN β gene occurs upon the binding of the 1) jun oncogene/activating transcription factor (c-jun/ATF-2) heterodimer, 2) IRFs 1, 7 and 9, and 3) NF-κB to their respective binding sites in the IFN β promoter. However, the binding of c-jun/ATF-2 and NF-κB to the IFN β promoter is not required for the induction of IFN β mRNA expression (*119*). Owing to the fact that IFN β expression can occur in the presence of NF-κB, we chose not to use IFN β as a readout cytokine for pathways that do not require IkB α degradation.

The IFN α promoter lacks NF- κ B binding sites but has binding sites for IRF transcription factors (*119*). Kawai *et al*, showed that in response to the TLR9 ligand CpG, Myd88 formed a complex with IRF7 and not IRF3, followed by recruitment of TRAF6 which lead to the activation of IFN α promoters (*118*). In its inactive form, IRF-7 resides in the cytoplasm. However unlike the activation of NF- κ B which requires the phosphorylation of I κ B α by the IKK complex, activation of IRF-7 requires its direct phosphorylation by IKK α (*120*). Our preference for studying the CpGA mediated activation of TLR9 (as opposed to the activation of TLRs 7 and 8) was primarily owing to the fact that CpGA stimulation of plasmocytoid dendritic cells has been shown to lead to a greater upregulation of IFN α mRNA expression over IFN β mRNA expression and a substantial increase in IFN α cytokine secretion (*121, 122*). Thus, as IFN α expression is not induced by NF- κ B, and the primary transcription factor (IRF-7) involved in its induction is not regulated by I κ B α , we selected IFN α as a cytokine whose production would not require I κ B α degradation. However, to rule out any possible secretion of IFN α by an uncharacterized receptor system or pathway that might be regulated by I κ B α , confirmatory experiments would be

required. This would include analyzing IFNα mRNA and protein expression in the context of selective knockdown of *NFKBIA* mRNA expression and IkBα inhibition.

4.7 Follow up experiments: Need and feasibility

Our hypothesis was that the presence of minor variants in the promoter of NFKBIA would lead to a reduced expression of its minor allele transcripts, leading to an overall reduction of NFKBIA mRNA levels. This would result in a reduction of IkBa molecules (which are needed to bind and sequester NF-kB in the cytosol) allowing more NF-kB molecules to bind and initiate proinflammatory cytokine expression. Our results have provided supporting evidence for reduced expression of minor allele containing transcripts and reduced mRNA levels of NFKBIA in individuals with the minor allele as well as increased inflammatory cytokine production. However, reduced IkBa production and increased NF-kB translocation are yet to be validated. Figure 4.1 outlines the key aspects of this study. An NF- κ B EMSA and an I κ B α western blot would be suitable for examining variations in $I\kappa B\alpha$ expression and NF- κB translocation between homozygotes and heterozygotes. These approaches have been used in validating promoter variants in other genes (87, 105). The differences observed in this project at the gene level are subtle; a 20-25% difference was seen in the expression of the major and minor allele transcripts, 28.5-33.8% differences in NFKBIA mRNA expression. Western blots and EMSAs are good qualitative measures of differences and their feasibility for measuring quantitative differences would be better served when examining larger differences such as those seen in the context of mutations or non-synonymous polymorphisms as opposed to regulatory polymorphisms. We were not able to recruit sufficient numbers of individuals who were homozygous for the minor alleles of rs3138053, rs2233406, and rs2233409 (GTT/GTT). Of the 97 individuals screened we only came across one individual who was homozygous for the minor allele of all 3 SNPs. Similarly, we came across a single neonate that was homozygous for all 3 SNPs. It would be informative to screen a larger subset of individuals to recruit minor allele homozygotes to see if the impact of the minor variants leads to more dramatic differences in NFKBIA gene expression and innate immune response in such individuals. A larger recruitment would also be needed to see if the magnitude of differences seen in TNF α output between homozygotes and heterozygotes of European descent differs in comparison to the magnitude of differences of TNFa output seen in other ethnicities. Another informative future experiment would be to analyze NFKBIA allelic imbalance and expression in homozygotes (ACC/ACC) and heterozygotes (ACC/GTT) using commonly invasive and rarely invasive serotypes of S. pneumoniae. Studies done in our lab have shown that human acute monocytic leukemia cell line (THP1) cells stimulated with commonly invasive serotypes of S. pneumoniae display a reduced TNF α secretion, and thus a reduced immune response, than rarely invasive ones

(123). It would be interesting to see if these differences are less pronounced or more pronounced in different *NFKBIA* genotypes of the host PBMCs when stimulated with different *S. pneumonia* serotypes.

4.8 Clinical implication of dysregulated TNF α production in inflammatory/autoimmune disease

The minor allele haplotype GTT has been shown to be associated with the risk of auto immune/inflammatory conditions. In this study it was shown that individuals heterozygous for these promoter SNPs displayed higher TNF α secretion than homozygotes in response to ligands for TLRs 1,2,4,7, and 8 which are involved in recognition of pathogen associated molecular patterns. TNF α was originally identified as an endotoxin induced glycoprotein, which caused haemorrhagic necrosis of sarcomas (*124*). Subsequent studies have revealed that TNF α is a key mediator of the inflammatory response to infection (*125*). TNF α is produced mainly by activated macrophages but is also produced by other immune cells such as mast cells, T and B lymphocytes, natural killer (NK) cells and neutrophils as well as smooth muscle cells, adipocytes and fibroblasts (*124, 126*). TNF α binds to its target receptors TNFR1 and TNFR2 which are expressed on most cell types with the exception of erythrocytes (*126*). TNFR1 mediates signaling through proinflammatory and programmed cell death pathway via the activation of NF- κ B and TNFR2, which has roles in signaling proliferation of thymocytes,

cytotoxic T cells, and human mononuclear cells (127). Upon binding of TNF α to TNFR1 at the plasma membrane, the TNFR1 death domain binds the TNFRSF1A-associated death domain (TRADD) adapter protein. TRADD recruits TNF receptor-associated factor 2 (TRAF2) protein and Ral-interacting protein 1 (RIP1) protein which, through a series of signaling intermediates, signal the activation of NF- κ B via the classical NF- κ B signaling pathway (128). NF- κ B in turn binds to consensus sites in the promoter of the TNF α gene and thereby initiates TNF α transcription. Subsequent synthesis of the TNFa protein results in a positive feedback that leads to further production of TNF α (129, 130). Elevated TNF α levels are not observed in healthy individuals but are prominent in response to pathogenic stimuli. Elevated serum and tissue levels of TNF α routinely associate with inflammatory and infectious conditions, and serum levels of TNF α correlate with the severity of infections (131, 132). Whereas elevated TNF α levels are needed in responding to infections, excessively high TNF α levels are associated with chronic inflammatory and autoimmune conditions which include rheumatoid arthritis, inflammatory bowel disease, psoriatic arthritis, vasculitis, and ankylosing spondilitis (133). Consequently, therapies that target TNF α production have shown promising results in reducing the severity of such conditions (134).

4.9 From bench to bedside: Diagnostic value of *NFKBIA* expression and *NFKBIA* polymorphisms in determining response to anti TNFα therapy

Anti-TNF α therapy is not applicable to all patients of chronic inflammatory conditions. One third of rheumatoid arthritis patients that are prescribed anti-TNF α therapy are non-responders (135). A subset of patients also display adverse side effects. Specifically, association and functional studies have shown that anti-TNF α therapy increases the risk of bacterial infections. For a recent review, refer to review articles cited at the end of this chapter (136, 137). Our results can potentially add more depth in the prognosis of non-response and adverse response to anti-TNF α therapy. Expression levels of soluble TNFR2 which acts a surrogate for TNF α levels in blood was shown to be elevated 12 years prior to the development of rheumatoid arthritis and were positively associated with the incidence of rheumatoid arthritis in women (138). Consequently, *NFKBIA* expression has been shown to positively associate with nonrespondence to anti TNF α therapy in rheumatoid arthritis patients and has been recommended as a prognostic marker for anti-TNF α therapy (139). The importance of knowing an individual's genetic response to anti-TNF α therapy will become more relevant once we enter the age of personalized medicine.

4.10 Conclusion

Genetic association studies have been critical in identifying genetic variants that associate with disease. Yet, the information obtained from such studies has functioned as the proverbial "tip of the iceberg" when it comes to understanding the precise disease pathology resulting from such variants and in determining therapeutic approaches applicable for reducing such pathologies. Our journey began with the observation that minor variants of SNPs rs3138053, rs2233406 and rs2233409 in the promoter of NFKBIA were associated with the risk of autoimmune and inflammatory conditions while at the same time they were protective against infections (Fig. 4.2). Our study has shown individuals heterozygous for NFKBIA promoter variants rs3138053, rs2233406, rs2233409 display allelic imbalance, reduced expression of NFKBIA mRNA than homozygotes while showing elevated levels of TNFa. The results presented here provide further evidence for the importance of the IkB family of proteins, specifically for IkB α in regulating the inflammatory response. A global understanding of the functional consequences of regulatory polymorphisms of NFKBIA will be provided with subsequent experiments that examine differences between all NFKBIA promoter genotypes (including minor allele homozygotes) for ΙκΒα protein expression and NF-κB translocation in all major ethnic groups.






Figure 4.2. Overall summary of research. Our goal was to gain a functional understanding of the opposing associations of the minor allele haplotype GTT (corresponding to SNPs rs3138053, rs2233406, rs2233409) with infectious and auto immune conditions. Through a series of functional experiments, we were able to demonstrate the impact of the minor allele variants on *NFKBIA* allele specific expression, *NFKBIA* gene expression and the innate immune response.

References

- 1. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993; 261 (5123): 921.
- 2. Zhong N, Weisgraber KH. Understanding the Basis for the Association of Apoe4 with Alzheimer's Disease: Opening the Door for Therapeutic Approaches. Curr Alzheimer Res 2009.
- 3. Kurylowicz A, Nauman J. The role of nuclear factor-kappaB in the development of autoimmune diseases: a link between genes and environment. Acta Biochim Pol 2008; 55 (4): 629.
- 4. Niederberger E, Geisslinger G. The IKK-NF-kappaB pathway: a source for novel molecular drug targets in pain therapy? Faseb J 2008; 22 (10): 3432.
- 5. Hayden MS, Ghosh S. Shared principles in NF-kappaB signaling. Cell 2008; 132 (3): 344.
- 6. Solt LA, May MJ. The IkappaB kinase complex: master regulator of NF-kappaB signaling. Immunol Res 2008; 42 (1-3): 3.
- 7. Hay RT, Vuillard L, Desterro JM, Rodriguez MS. Control of NF-kappa B transcriptional activation by signal induced proteolysis of I kappa B alpha. Philos Trans R Soc Lond B Biol Sci 1999; 354 (1389): 1601.
- 8. Bonizzi G, Karin M. The two NF-kappaB activation pathways and their role in innate and adaptive immunity. Trends Immunol 2004; 25 (6): 280.
- 9. Zandi E, Chen Y, Karin M. Direct phosphorylation of IkappaB by IKKalpha and IKKbeta: discrimination between free and NF-kappaB-bound substrate. Science 1998; 281 (5381): 1360.
- 10. Rothwarf DM, Zandi E, Natoli G, Karin M. IKK-gamma is an essential regulatory subunit of the IkappaB kinase complex. Nature 1998; 395 (6699): 297.
- 11. Vallabhapurapu S, Karin M. Regulation and function of NF-kappaB transcription factors in the immune system. Annu Rev Immunol 2009; 27: 693.
- 12. Tas SW, Vervoordeldonk MJ, Tak PP. Gene therapy targeting nuclear factor-kappaB: towards clinical application in inflammatory diseases and cancer. Curr Gene Ther 2009; 9 (3): 160.
- McDonald DR, Brown D, Bonilla FA, Geha RS. Interleukin receptor-associated kinase-4 deficiency impairs Toll-like receptor-dependent innate antiviral immune responses. J Allergy Clin Immunol 2006; 118 (6): 1357.
- 14. von Bernuth H, Picard C, Jin Z, et al. Pyogenic bacterial infections in humans with MyD88 deficiency. Science 2008; 321 (5889): 691.
- 15. Courtois G, Smahi A, Reichenbach J, et al. A hypermorphic IkappaBalpha mutation is associated with autosomal dominant anhidrotic ectodermal dysplasia and T cell immunodeficiency. J Clin Invest 2003; 112 (7): 1108.
- 16. Smahi A, Courtois G, Vabres P, et al. Genomic rearrangement in NEMO impairs NFkappaB activation and is a cause of incontinentia pigmenti. The International Incontinentia Pigmenti (IP) Consortium. Nature 2000; 405 (6785): 466.
- 17. van de Vosse E, van Dissel JT, Ottenhoff TH. Genetic deficiencies of innate immune signalling in human infectious disease. Lancet Infect Dis 2009; 9 (11): 688.

- 18. Courtois G, Gilmore TD. Mutations in the NF-kappaB signaling pathway: implications for human disease. Oncogene 2006; 25 (51): 6831.
- 19. Pai S, Thomas R. Immune deficiency or hyperactivity-Nf-kappab illuminates autoimmunity. J Autoimmun 2008; 31 (3): 245.
- 20. Handel ML, McMorrow LB, Gravallese EM. Nuclear factor-kappa B in rheumatoid synovium. Localization of p50 and p65. Arthritis Rheum 1995; 38 (12): 1762.
- 21. Gveric D, Kaltschmidt C, Cuzner ML, Newcombe J. Transcription factor NF-kappaB and inhibitor I kappaBalpha are localized in macrophages in active multiple sclerosis lesions. J Neuropathol Exp Neurol 1998; 57 (2): 168.
- 22. Dale E, Davis M, Faustman DL. A role for transcription factor NF-kappaB in autoimmunity: possible interactions of genes, sex, and the immune response. Adv Physiol Educ 2006; 30 (4): 152.
- 23. Whiteside ST, Israel A. I kappa B proteins: structure, function and regulation. Semin Cancer Biol 1997; 8 (2): 75.
- 24. Karin M, Ben-Neriah Y. Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity. Annu Rev Immunol 2000; 18: 621.
- 25. Brown K, Park S, Kanno T, Franzoso G, Siebenlist U. Mutual regulation of the transcriptional activator NF-kappa B and its inhibitor, I kappa B-alpha. Proc Natl Acad Sci U S A 1993; 90 (6): 2532.
- 26. Ferreiro DU, Komives EA. Molecular Mechanisms of System Control of NF-kappaB Signaling by IkappaBalpha. Biochemistry.
- 27. Ghosh S, Karin M. Missing pieces in the NF-kappaB puzzle. Cell 2002; 109 Suppl: S81.
- 28. Bergqvist S, Alverdi V, Mengel B, Hoffmann A, Ghosh G, Komives EA. Kinetic enhancement of NF-kappaBxDNA dissociation by IkappaBalpha. Proc Natl Acad Sci U S A 2009; 106 (46): 19328.
- 29. Beg AA, Sha WC, Bronson RT, Baltimore D. Constitutive NF-kappa B activation, enhanced granulopoiesis, and neonatal lethality in I kappa B alpha-deficient mice. Genes Dev 1995; 9 (22): 2736.
- Klement JF, Rice NR, Car BD, et al. IkappaBalpha deficiency results in a sustained NFkappaB response and severe widespread dermatitis in mice. Mol Cell Biol 1996; 16 (5): 2341.
- 31. Janssen R, van Wengen A, Hoeve MA, et al. The same IkappaBalpha mutation in two related individuals leads to completely different clinical syndromes. J Exp Med 2004; 200 (5): 559.
- 32. McDonald DR, Mooster JL, Reddy M, Bawle E, Secord E, Geha RS. Heterozygous Nterminal deletion of IkappaBalpha results in functional nuclear factor kappaB haploinsufficiency, ectodermal dysplasia, and immune deficiency. J Allergy Clin Immunol 2007; 120 (4): 900.
- 33. Lopez-Granados E, Keenan JE, Kinney MC, et al. A novel mutation in NFKBIA/IKBA results in a degradation-resistant N-truncated protein and is associated with ectodermal dysplasia with immunodeficiency. Hum Mutat 2008; 29 (6): 861.
- 34. Motsinger AA, Haas DW, Hulgan T, Ritchie MD. Human genomic association studies: a primer for the infectious diseases specialist. J Infect Dis 2007; 195 (12): 1737.
- 35. Harris H. The principals of human biochemical genetics: North Holland Publishing Company, 1975.
- Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 2003; 33 Suppl: 228.

- 37. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008; 118 (5): 1590.
- 38. Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet 2001; 27 (3): 234.
- 39. Collins A. Allelic association: linkage disequilibrium structure and gene mapping. Mol Biotechnol 2009; 41 (1): 83.
- 40. Wahls WP. Meiotic recombination hotspots: shaping the genome and insights into hypervariable minisatellite DNA change. Curr Top Dev Biol 1998; 37: 37.
- 41. Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989; 245 (4922): 1066.
- 42. Kunkel LM, Hoffman EP. Duchenne/Becker muscular dystrophy: a short overview of the gene, the protein, and current diagnostics. Br Med Bull 1989; 45 (3): 630.
- 43. Panoutsopoulou K, Zeggini E. Finding common susceptibility variants for complex disease: past, present and future. Brief Funct Genomic Proteomic 2009; 8 (5): 345.
- 44. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999; 22 (2): 139.
- 45. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002; 296 (5576): 2225.
- 46. Johnson GC, Esposito L, Barratt BJ, et al. Haplotype tagging for the identification of common disease genes. Nat Genet 2001; 29 (2): 233.
- 47. The International HapMap Project. Nature 2003; 426 (6968): 789.
- 48. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007; 449 (7164): 851.
- 49. Gu CC, Yu K, Ketkar S, Templeton AR, Rao DC. On transferability of genome-wide tagSNPs. Genet Epidemiol 2008; 32 (2): 89.
- 50. Xing J, Witherspoon DJ, Watkins WS, Zhang Y, Tolpinrud W, Jorde LB. HapMap tagSNP transferability in multiple populations: general guidelines. Genomics 2008; 92 (1): 41.
- 51. Mozzato-Chamay N, Corbett EL, Bailey RL, Mabey DC, Raynes J, Conway DJ. Polymorphisms in the IkappaB-alpha promoter region and risk of diseases involving inflammation and fibrosis. Genes Immun 2001; 2 (3): 153.
- 52. Burton MJ, Mabey DC. The global burden of trachoma: a review. PLoS Negl Trop Dis 2009; 3 (10): e460.
- 53. Chapman SJ, Khor CC, Vannberg FO, et al. IkappaB genetic polymorphisms and invasive pneumococcal disease. Am J Respir Crit Care Med 2007; 176 (2): 181.
- 54. Siezen CL, Bont L, Hodemaekers HM, et al. Genetic susceptibility to respiratory syncytial virus bronchiolitis in preterm children is associated with airway remodeling genes and innate immune genes. Pediatr Infect Dis J 2009; 28 (4): 333.
- 55. Abdallah A, Sato H, Grutters JC, et al. Inhibitor kappa B-alpha (IkappaB-alpha) promoter polymorphisms in UK and Dutch sarcoidosis. Genes Immun 2003; 4 (6): 450.
- 56. Miterski B, Bohringer S, Klein W, et al. Inhibitors in the NFkappaB cascade comprise prime candidate genes predisposing to multiple sclerosis, especially in selected combinations. Genes Immun 2002; 3 (4): 211.
- 57. Zhai R, Zhou W, Gong MN, et al. Inhibitor kappaB-alpha haplotype GTC is associated with susceptibility to acute respiratory distress syndrome in Caucasians. Crit Care Med 2007; 35 (3): 893.
- 58. Kurylowicz A, Hiromatsu Y, Jurecka-Lubieniecka B, et al. Association of NFKB1 94ins/del ATTG promoter polymorphism with susceptibility to and phenotype of Graves' disease. Genes Immun 2007; 8 (7): 532.

59. Tsan-Teng Ou C-HL, Yu-Chih Lin, Ruei-Nian Li, Wen-Chan Tsai, Hong-Wen Liu, Jeng-Hsien Yen. IkBα Promoter Polymorphisms in Patients with Primary

Sjögren's Syndrome. Journal of Clinical Immunology 2008 (28): 440.

- 60. Lin CH, Wang SC, Ou TT, et al. I kappa B alpha promoter polymorphisms in patients with systemic lupus erythematosus. J Clin Immunol 2008; 28 (3): 207.
- 61. Lin CH, Ou TT, Wu CC, Tsai WC, Liu HW, Yen JH. IkappaBalpha promoter polymorphisms in patients with rheumatoid arthritis. Int J Immunogenet 2007; 34 (1): 51.
- 62. Hung YH, Ou TT, Lin CH, et al. IkBalpha promoter polymorphisms in patients with ankylosing spondylitis. Rheumatol Int 2009.
- 63. Szamosi T, Lakatos PL, Szilvasi A, et al. The 3'UTR NFKBIA variant is associated with extensive colitis in Hungarian IBD patients. Dig Dis Sci 2009; 54 (2): 351.
- 64. Klein W, Tromm A, Folwaczny C, et al. A polymorphism of the NFKBIA gene is associated with Crohn's disease patients lacking a predisposing allele of the CARD15 gene. Int J Colorectal Dis 2004; 19 (2): 153.
- 65. Romzova M, Hohenadel D, Kolostova K, et al. NFkappaB and its inhibitor IkappaB in relation to type 2 diabetes and its microvascular and atherosclerotic complications. Hum Immunol 2006; 67 (9): 706.
- 66. Naik RG, Brooks-Worrell BM, Palmer JP. Latent autoimmune diabetes in adults. J Clin Endocrinol Metab 2009; 94 (12): 4635.
- 67. Katarina K, Daniela P, Peter N, et al. HLA, NFKB1 and NFKBIA gene polymorphism profile in autoimmune diabetes mellitus patients. Exp Clin Endocrinol Diabetes 2007; 115 (2): 124.
- 68. Spink CF, Gray LC, Davies FE, Morgan GJ, Bidwell JL. Haplotypic structure across the I kappa B alpha gene (NFKBIA) and association with multiple myeloma. Cancer Lett 2007; 246 (1-2): 92.
- 69. Lewander A, Butchi AK, Gao J, et al. Polymorphism in the promoter region of the NFKB1 gene increases the risk of sporadic colorectal cancer in Swedish but not in Chinese populations. Scand J Gastroenterol 2007; 42 (11): 1332.
- 70. He Y, Zhang H, Yin J, et al. IkappaBalpha gene promoter polymorphisms are associated with hepatocarcinogenesis in patients infected with hepatitis B virus genotype C. Carcinogenesis 2009; 30 (11): 1916.
- 71. Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. N Engl J Med 2002; 347 (12): 911.
- 72. Bach JF. Infections and autoimmune diseases. J Autoimmun 2005; 25 Suppl: 74.
- 73. Chorley BN, Wang X, Campbell MR, Pittman GS, Noureddine MA, Bell DA. Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. Mutat Res 2008; 659 (1-2): 147.
- 74. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007; 447 (7145): 661.
- 75. Montgomery SB, Dermitzakis ET. The resolution of the genetics of gene expression. Hum Mol Genet 2009; 18 (R2): R211.
- 76. Kraft P, Horvath S. The genetics of gene expression and gene mapping. Trends Biotechnol 2003; 21 (9): 377.
- 77. Williams RB, Chan EK, Cowley MJ, Little PF. The influence of genetic variation on gene expression. Genome Res 2007; 17 (12): 1707.
- 78. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet 2008; 24 (8): 408.
- 79. Wittkopp PJ. Genomic sources of regulatory variation in cis and in trans. Cell Mol Life Sci 2005; 62 (16): 1779.

- 80. Cheung VG, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat Rev Genet 2009; 10 (9): 595.
- 81. Karimi M, Goldie LC, Cruickshank MN, Moses EK, Abraham LJ. A critical assessment of the factors affecting reporter gene assays for promoter SNP function: a reassessment of -308 TNF polymorphism function using a novel integrated reporter system. Eur J Hum Genet 2009; 17 (11): 1454.
- 82. Hollon T, Yoshimura FK. Variation in enzymatic transient gene expression assays. Anal Biochem 1989; 182 (2): 411.
- 83. Pastinen T, Hudson TJ. Cis-acting regulatory variation in the human genome. Science 2004; 306 (5696): 647.
- 84. Prokunina L, Alarcon-Riquelme ME. Regulatory SNPs in complex diseases: their identification and functional validation. Expert Rev Mol Med 2004; 6 (10): 1.
- 85. Knight JC. Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. Clin Sci (Lond) 2003; 104 (5): 493.
- 86. Gurevich I, Zhang C, Aneskievich BJ. Scanning for transcription factor binding by a variant EMSA. Methods Mol Biol; 585: 147.
- 87. Sun C, Southard C, Witonsky DB, Olopade OI, Di Rienzo A. Allelic imbalance (AI) identifies novel tissue-specific cis-regulatory variation for human UGT2B15. Hum Mutat; 31 (1): 99.
- 88. Buckland PR. Allele-specific gene expression differences in humans. Hum Mol Genet 2004; 13 Spec No 2: R255.
- 89. Battig P, Hathaway LJ, Hofer S, Muhlemann K. Serotype-specific invasiveness and colonization prevalence in Streptococcus pneumoniae correlate with the lag phase during in vitro growth. Microbes Infect 2006; 8 (11): 2612.
- 90. Bettinger JA, Scheifele DW, Halperin SA, Kellner JD, Tyrrell G. Invasive pneumococcal infections in Canadian children, 1998-2003: implications for new vaccination programs. Can J Public Health 2007; 98 (2): 111.
- 91. Kollmann TR, Crabtree J, Rein-Weston A, et al. Neonatal innate TLR-mediated responses are distinct from those of adults. J Immunol 2009; 183 (11): 7150.
- 92. Ou TT, Lin CH, Lin YC, et al. IkappaBalpha promoter polymorphisms in patients with primary Sjogren's syndrome. J Clin Immunol 2008; 28 (5): 440.
- 93. Delerive P, Monte D, Dubois G, et al. The orphan nuclear receptor ROR alpha is a negative regulator of the inflammatory response. EMBO Rep 2001; 2 (1): 42.
- 94. Zhu G, Lipsky RH, Xu K, et al. Differential expression of human COMT alleles in brain and lymphoblasts detected by RT-coupled 5' nuclease assay. Psychopharmacology (Berl) 2004; 177 (1-2): 178.
- 95. Newburg DS, Walker WA. Protection of the neonate by the innate immune system of developing gut and of human milk. Pediatr Res 2007; 61 (1): 2.
- 96. Perry AK, Chen G, Zheng D, Tang H, Cheng G. The host type I interferon response to viral and bacterial infections. Cell Res 2005; 15 (6): 407.
- 97. Campino S, Forton J, Auburn S, et al. TLR9 polymorphisms in African populations: no association with severe malaria, but evidence of cis-variants acting on gene expression. Malar J 2009; 8: 44.
- 98. Morley M, Molony CM, Weber TM, et al. Genetic analysis of genome-wide variation in human gene expression. Nature 2004; 430 (7001): 743.
- 99. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. Nat Genet 2007; 39 (10): 1202.

- 100. Heap GA, Yang JH, Downes K, et al. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. Hum Mol Genet; 19 (1): 122.
- 101. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. Science 2007; 318 (5853): 1136.
- 102. Plagnol V, Uz E, Wallace C, et al. Extreme clonality in lymphoblastoid cell lines with implications for allele specific expression analyses. PLoS One 2008; 3 (8): e2966.
- 103. Whitney AR, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U S A 2003; 100 (4): 1896.
- 104. Karlovich C, Duchateau-Nguyen G, Johnson A, et al. A longitudinal study of gene expression in healthy individuals. BMC Med Genomics 2009; 2: 33.
- 105. Karban AS, Okazaki T, Panhuysen CI, et al. Functional annotation of a novel NFKB1 promoter polymorphism that increases risk for ulcerative colitis. Hum Mol Genet 2004; 13 (1): 35.
- 106. Li H, Gao L, Shen Z, et al. Association study of NFKB1 and SUMO4 polymorphisms in Chinese patients with psoriasis vulgaris. Arch Dermatol Res 2008; 300 (8): 425.
- 107. Zhou B, Rao L, Li Y, et al. A functional insertion/deletion polymorphism in the promoter region of NFKB1 gene increases susceptibility for nasopharyngeal carcinoma. Cancer Lett 2009; 275 (1): 72.
- 108. Lin SC, Liu CJ, Yeh WI, Lui MT, Chang KW, Chang CS. Functional polymorphism in NFKB1 promoter is related to the risks of oral squamous cell carcinoma occurring on older male areca (betel) chewers. Cancer Lett 2006; 243 (1): 47.
- 109. Zhou B, Qie M, Wang Y, et al. Relationship between NFKB1 -94 insertion/deletion ATTG polymorphism and susceptibility of cervical squamous cell carcinoma risk. Ann Oncol 2009.
- Tang T, Cui S, Deng X, et al. Insertion/deletion polymorphism in the promoter region of NFKB1 gene increases susceptibility for superficial bladder cancer in Chinese. DNA Cell Biol; 29 (1): 9.
- 111. Zhang P, Wei Q, Li X, et al. A functional insertion/deletion polymorphism in the promoter region of the NFKB1 gene increases susceptibility for prostate cancer. Cancer Genet Cytogenet 2009; 191 (2): 73.
- 112. Drent M, van den Berg R, Haenen GR, van den Berg H, Wouters EF, Bast A. NF-kappaB activation in sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis 2001; 18 (1): 50.
- Benovoy D, Kwan T, Majewski J. Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments. Nucleic Acids Res 2008; 36 (13): 4417.
- 114. Shih VF, Kearns JD, Basak S, Savinova OV, Ghosh G, Hoffmann A. Kinetic control of negative feedback regulators of NF-kappaB/RelA determines their pathogen- and cytokine-receptor signaling specificity. Proc Natl Acad Sci U S A 2009; 106 (24): 9619.
- 115. Covert MW, Leung TH, Gaston JE, Baltimore D. Achieving stability of lipopolysaccharide-induced NF-kappaB activation. Science 2005; 309 (5742): 1854.
- 116. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science 2002; 298 (5596): 1241.
- 117. Charles A. Janeway Jr. PT, Mark Walport, Mark J. Shlomchik. Chapter 2: Innate Immunity. In: Lawrence E, ed. Immuno Biology. New York: Garland Science, 2005: 87.
- 118. Kawai T, Sato S, Ishii KJ, et al. Interferon-alpha induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. Nat Immunol 2004; 5 (10): 1061.

- 119. Randall RE, Goodbourn S. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. J Gen Virol 2008; 89 (Pt 1): 1.
- 120. Hoshino K, Sugiyama T, Matsumoto M, et al. IkappaB kinase-alpha is critical for interferon-alpha production induced by Toll-like receptors 7 and 9. Nature 2006; 440 (7086): 949.
- 121. Abel K, Wang Y, Fritts L, et al. Deoxycytidyl-deoxyguanosine oligonucleotide classes A, B, and C induce distinct cytokine gene expression patterns in rhesus monkey peripheral blood mononuclear cells and distinct alpha interferon responses in TLR9-expressing rhesus monkey plasmacytoid dendritic cells. Clin Diagn Lab Immunol 2005; 12 (5): 606.
- 122. Krieg AM. CpG motifs in bacterial DNA and their immune effects. Annu Rev Immunol 2002; 20: 709.
- 123. Burgess TS, Hirschfeld AF, Tyrrell GJ, Bettinger JA, Turvey SE. Commonly invasive serotypes of Streptococcus pneumoniae trigger a reduced innate immune response compared with serotypes rarely responsible for invasive infection. FEMS Immunol Med Microbiol 2008; 53 (1): 136.
- 124. Bradley JR. TNF-mediated inflammatory disease. J Pathol 2008; 214 (2): 149.
- 125. Clark IA. How TNF was recognized as a key mechanism of disease. Cytokine Growth Factor Rev 2007; 18 (3-4): 335.
- 126. Popa C, Netea MG, van Riel PL, van der Meer JW, Stalenhoef AF. The role of TNFalpha in chronic inflammatory conditions, intermediary metabolism, and cardiovascular risk. J Lipid Res 2007; 48 (4): 751.
- 127. Choi SJ, Lee KH, Park HS, Kim SK, Koh CM, Park JY. Differential expression, shedding, cytokine regulation and function of TNFR1 and TNFR2 in human fetal astrocytes. Yonsei Med J 2005; 46 (6): 818.
- 128. Verstrepen L, Bekaert T, Chau TL, Tavernier J, Chariot A, Beyaert R. TLR-4, IL-1R and TNF-R signaling to NF-kappaB: variations on a common theme. Cell Mol Life Sci 2008; 65 (19): 2964.
- 129. Trede NS, Tsytsykova AV, Chatila T, Goldfeld AE, Geha RS. Transcriptional activation of the human TNF-alpha promoter by superantigen in human monocytic cells: role of NF-kappa B. J Immunol 1995; 155 (2): 902.
- 130. Bashir MM, Sharma MR, Werth VP. TNF-alpha production in the skin. Arch Dermatol Res 2009; 301 (1): 87.
- 131. Robak T, Gladalska A, Stepien H. The tumour necrosis factor family of receptors/ligands in the serum of patients with rheumatoid arthritis. Eur Cytokine Netw 1998; 9 (2): 145.
- 132. Waage A, Halstensen A, Espevik T. Association between tumour necrosis factor in serum and fatal outcome in patients with meningococcal disease. Lancet 1987; 1 (8529): 355.
- 133. Wong M, Ziring D, Korin Y, et al. TNFalpha blockade in human diseases: mechanisms and future directions. Clin Immunol 2008; 126 (2): 121.
- 134. Taylor PC, Feldmann M. Anti-TNF biologic agents: still the therapy of choice for rheumatoid arthritis. Nat Rev Rheumatol 2009; 5 (10): 578.
- 135. Verweij CL. Predicting the future of anti-tumor necrosis factor therapy. Arthritis Res Ther 2009; 11 (3): 115.
- 136. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. Jama 2006; 295 (19): 2275.
- 137. Leombruno JP, Einarson TR, Keystone EC. The safety of anti-tumour necrosis factor treatments in rheumatoid arthritis: meta and exposure-adjusted pooled analyses of serious adverse events. Ann Rheum Dis 2009; 68 (7): 1136.

- 138. Karlson EW, Chibnik LB, Tworoger SS, et al. Biomarkers of inflammation and development of rheumatoid arthritis in women from two prospective cohort studies. Arthritis Rheum 2009; 60 (3): 641.
- 139. Koczan D, Drynda S, Hecker M, et al. Molecular discrimination of responders and nonresponders to anti-TNF alpha therapy in rheumatoid arthritis by etanercept. Arthritis Res Ther 2008; 10 (3): R50.

Appendix: Haplotypes and LD measure for the entire *NFKBIA* gene in 4 major ethnic groups



Figure A.1. Haplotyope for the *NFKBIA* **gene in individuals of European descent.** Variants in the promoter region are shown in a black triangular grid.



Figure A.2. Haplotype for the *NFKBIA* gene in individuals of African descent. Variants in the promoter region are shown in a black triangular grid.



Figure A.3. Haplotype for the *NFKBIA* **gene in individuals of Chinese (Han) descent.** Variants in the promoter region are shown in a black triangular grid.



Figure A.4. Haplotype for the *NFKBIA* **gene in individuals of Japanese descent.** Variants in the promoter region are shown in a black triangular grid.



Figure A.5. Strong LD is seen in non-coding regions of *NFKBIA***.** The total number of variants that displayed strong LD (LD≥80%) for each gene region were divided by the absolute number of variants in each of the 4 populations examined to yield a percentage of LD (A). Region-specific LD was also determined by dividing the number of variants that displayed strong LD within the region by the total number of variants (B).