FUNCTIONAL ANALYSIS OF POLYMORPHISMS IN ASTHMA GENES: TSLP AND IL1RL1

by

Loubna Akhabir

B.Sc., Faculté Des Sciences et Techniques de Mohammedia, 1999

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2014

Abstract

Thymic stromal lymphopoietin (TSLP) and Interleukin 1 receptor-like 1 (IL1RL1) are important for the initiation and maintenance of a Th2 inflammatory environment in the asthmatic lung. TSLP and IL-33, the exclusive IL1RL1 ligand, are secreted by epithelial cells and other immune cells and play essential roles in Th2 polarization, activation and proliferation of immune cells and participate in many asthma cardinal features such as chronic inflammation, airway remodeling and mucus production. *TSLP* and *IL1RL1* are two of the most consistently associated genes in genome-wide association studies of asthma. rs1837253 in *TSLP* was identified as a putative causal SNP based on consistent association data both from candidate gene and genome-wide association studies; as well as the absence of significant linkage disequilibrium with other single nucleotide polymorphisms (SNPs) in the region. In contrast, there are several asthma-associated *IL1RL1* variants and due to the extensive linkage disequilibrium in the region encompassing *IL1RL1* and many other genes, potential causal SNP(s) are unknown.

In this thesis, I describe the functional analysis of the *TSLP* SNP rs1837253, the study of SNPs in the *IL1RL1* region to identify potentially functional variants by *in silico* analysis using a lung expression Quantitative Trait Locus dataset and published association data, the functional analysis of the *IL1RL1* SNPs: rs1420101 and rs3771180 and finally an analysis of gene expression of TSLP short and long isoforms as well as IL1RL1 receptor and soluble isoforms in relation to clinical phenotypes.

Preface

Parts of chapter 1 were published as the following two peer-reviewed review articles:

[Akhabir L], Sandford AJ. Genome-wide association studies for discovery of genes involved in asthma. *Respirology*. 2011 Apr; 16(3): 396-406.

[Akhabir L], Sandford AJ. Genetics of Interleukin 1 receptor-like 1 in immune and inflammatory diseases. *Current Genomics*. 2010 Dec; 11(8): 591-606.

A version of chapter 3 was published in the Journal of Allergy and Clinical Immunology. [Akhabir L], Bérubé JC, Bossé Y, Laviolette M, Hao K, Nickle DC, Timens W, Sin DD, Paré PD, Postma DS, Sandford AJ. Lung eQTL dataset identifies important functional polymorphisms in the asthma-associated IL1RL1 region. J Allergy Clin Immunol. 2014 Apr 17. Epub ahead of print. PMID: 24746754.

I designed and conducted all the analysis in the study and wrote the manuscript, JCB assisted with some of the expression analysis, YCB, YB and KH supplied specific datasets and helped with data extraction, YB and ML were responsible for the Laval University arm of the eQTL study, DCN and KH participated in the Merck arm of the lung eQTL study, WT and DSP were responsible for the Groningen arm of the eQTL study, DDS and PDP were responsible for the UBC arm of the eQTL study, AS supervised me throughout the study, provided guidance and reviewed the analysis and manuscript writing; all the co-authors reviewed and edited the manuscript.

The work outlined in this thesis was conducted with approval from University of British Columbia Office of Research and in accordance with the University of British

Columbia Policies and procedures, Biosafety Practices and Public Health Agency of Canada Guidelines. The work using radioactive isotopes was in accordance with the University of British Columbia radiation safety office policies and was approved by the committee on radioisotopes and radiation hazards; license number: MEDI-3242-16.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Description
°C	Degrees Celsius
3'UTR	3' untranslated region
5'UTR	5' untranslated region
ABI	Applied Biosystems
ACEI	Angiotensin-converting-enzyme inhibitor
ADAM33	ADAM metallopeptidase domain 33
ADRA1B	Adrenoceptor alpha 1B
AIC	Akaike Information Criterion
ALT	Alanine aminotransferase
ANOVA	Analysis of variance
ANTXR1	Anthrax toxin receptor 1
ARE	Adenylate-uridylate-rich elements
ATCC	American Type culture collection
ATF6	Activating transcription factor 6
ATP	Adenosine triphosphate
AUF-1	AU-rich element RNA-binding protein 1
AU-rich	Adenylate-uridylate-rich
BMI	Body mass index
bp	Base pair
BTNL2	Butyrophilin-like 2 (MHC class II associated)
C/EBP	CCAAT-enhancer-binding protein
C6orf10	Chromosome 6 open reading frame 10
cAMP	Cyclic AMP
CAPPS	Canadian Asthma Primary Prevention Study
CCL11	Chemokine (C-C motif) ligand 11
CCL12	Chemokine (C-C motif) ligand 12
CCL17	Chemokine (C-C motif) ligand 17
CD4	Cluster of differentiation 4
CD40	CD40 molecule, TNF receptor superfamily member 5

Abbreviation	Description
CD8	Cluster of differentiation 80
CD86	Cluster of differentiation 86
CDHR3	Cadherin-related family member 3
cDNA	Complementary DNA
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
ChIP-Seq	Chromatin immunoprecipitation-sequencing
CHRNA2	Cholinergic receptor, nicotinic, alpha 2
CLP	Cecal ligation and puncture
COL18A1	Collagen, type XVIII, alpha 1
COL5A3	Collagen, type V, alpha 3
COL6A5	Collagen, type VI, alpha 5
COPD	Chronic obstructive pulmonary disease
CRB1	Crumbs homolog 1 (Drosophila)
CRLF2	Cytokine receptor-like factor 2
CSE	Cigarette smoke extract
CSF2	Colony stimulating factor 2 (granulocyte-macrophage)
Ct	threshold cycle
CTCF	CCCTC-binding factor
CTNNA3	Catenin (cadherin-associated protein), alpha 3
CXCR2	Chemokine (C-X-C motif) receptor 2 (or IL8RB)
DC	dendritic cells
DENND1B	DENN/MADD domain containing 1B
DMEM	Dulbecco’s Minimum Essential Medium
DPP10	Dipeptidyl-peptidase 10
EMSA	Electrophoretic mobility shift assay
ENCODE	The encyclopedia of DNA elements
eQTL	Expression quantitative trait locus
ERK1	Mitogen-activated protein kinase 3
ERK2	Mitogen-activated protein kinase 1
ESPERR	Evolutionary and sequence pattern extraction through reduced representations

Abbreviation	Description
EVE	The EVE consortium to identify asthma susceptibility genes in ethnically diverse populations
FCER1A	Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide
FEV_1	Forced expiratory volume in one second
FEV_1%VC	Forced expiratory volume in one second expressed as a percentage of the vital capacity
FOXF2	Forkhead Box F2
FOXP3	Forkhead box P3
FVC	Forced vital capacity
GATA2	GATA binding protein 2
gDNA	Genomic DNA
GFP	Green fluorescent protein
GINA	Global Initiative for Asthma
Glmnet	Lasso and elastic-net regularized generalized linear models
GM-CSF	Colony stimulating factor 2 (granulocyte-macrophage)
GR	Glucocorticoid receptor
GRE	Glucocorticoid response element
GRK2	Beta adrenergic receptor kinase
GSDMA	Gasdermin A
GSDMB	Gasdermin B
GWAS	Genome-wide association study
HAC	Histone acetylases
HapMap	Haplotype Map Project
HBB	Hemoglobin subunit beta
HeLa	Cancerous cell line derived from Henrietta Lacks
HepG2	Hepatocellular carcinoma human cell line
HLA	Human leukocyte antigen
HLA-DQ	Major histocompatibility complex, class II, DQ
HLA-DQB1	Major histocompatibility complex, class II, DQ beta 1
HLA-DRA	Major histocompatibility complex, class II, DR alpha
HLA-DRB1	Major histocompatibility complex, class II, DR beta 1
HMEC	Human mammary epithelial cell

Abbreviation	Description
HPLC	High-performance liquid chromatography
HuR	Hu antigen R (aka ELAVL1, embryonic lethal, abnormal vision, Drosophila)-like 1
HUVEC	Human umbilical vein endothelial cell
HWE	Hardy-Weinberg equilibrium
IBD	Inflammatory bowel disease
ICOS	Inducible T-cell co-stimulator
IgE	Immunoglobulin type E
IIAM	The International Institute for Advancement of Medicine
IKZF2	IKAROS family zing finger 2 (Helios)
IL10	Interleukin 10
IL13	Interleukin 13
IL18R1	interleukin 18 receptor 1
IL18RAP	Interleukin 18 receptor accessory protein
IL1R2	Interleukin 1 receptor, type 2 (a.k.a. CD121b)
IL1RL1	interleukin 1 receptor-like 1
IL1RL2	interleukin 1 receptor-like 2
IL1α	Interleukin 1 alpha
IL1β	Interleukin 1 beta
IL25	Interleukin 25
IL25R	Interleukin 25 receptor
IL2RB	Interleukin 2 receptor, beta
IL33	interleukin 33
IL4	Interleukin 4
IL5	Interleukin 5
IL6	Interleukin 6
IL7	Interleukin 7
IL8	Interleukin 8
ILC2	Innate lymphoid cells, type 2
IRE	Iron-responsive elements
IRES	Internal ribosome entry sites

Abbreviation	Description
Jak	Janus kinase
K562	Myelogenous leukemia cell line
KHSRP	KH-Type Splicing Regulatory Protein
LABA	Long-acting beta2-agonist
LB	Lysogeny broth
LD	Linkage disequilibrium
lincRNA	Long intergenic non-coding RNA
LOD	Logarithm (base 10) of odds
LPS	Lipopolysaccharide
luc	Luciferase
MAF	Minor allele frequency
MAP4K4	Mitogen-activated protein kinase kinase kinase kinase 4
MFSD9	Major facilitator superfamily, member 9
MGB	Minor groove binding
MHC	Major histocompatibility complex
MI	Myocardial infarction
minP	Minimal promoter
miRNA	MicroRNA
MKLN1	Muskelin 1, intracellular mediator containing kelch motifs
mRNA	Messenger RNA
MYC	c-Myc
MYD88	Myeloid differentiation primary response gene 88
NAPS	Nucleic acid protein service unit in UBC
NCBI	National Center for Biotechnology Information
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NIH	National Institutes of Health
NKT	Natural killer cells
NPSR1	Neuropeptide S receptor 1
NRG1	Neuregulin 1
NRMSE	Normalized root mean square error
NSAIDS	Non-steroidal anti-inflammatory drugs

Abbreviation	Description
OCA2	Oculocutaneous albinism II
ORMDL3	ORM1-like 3 (S. cerevisiae)
OVA	Ovalbumin
OX40	Tumor necrosis factor receptor superfamily, member 4
OX40L	Tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally activated glycoprotein 1, 34kDa)
PAGE	Polyacrylamide gel electrophoresis
PC20	Provocative concentration causing a 20% drop of FEV_1
PCR	Polymerase chain reaction
PDE4D	Phosphodiesterase 4D, cAMP-specific
PGK	Phosphoglycerate kinase
PHF11	PHD finger protein 11
PKA	cAMP-dependent protein kinase
Poly(I:C)	Polyinosinic:polycytidylic acid, sodium salt
pQTL	Protein quantitative trait locus
PRNP	Prion protein
PYH1N1	Pyrin and HIN domain family member 1
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RA	Rheumatoid arthritis
RAD50	RAD50 homolog (S. cerevisiae)
RE	Restriction enzyme
RMSE	Root-mean-square error
RPLP0	Ribosomal protein, large, P0
RT	Room temperature
SH2B3 (aka LNK)	SH2B adaptor protein 3
siRNA	Small interfering RNA
SLC9A2	Solute carrier family 9, subfamily A (NHE2, cation proton antiporter 2), member 2
SLC9A4	Solute carrier family 9, subfamily A (NHE2, cation proton antiporter 2), member 4
SMAD3	SMAD family member 3
SNP	Single nucleotide polymorphism

Abbreviation	Description
Sp1	Specificity protein 1
Src	Proto-oncogene tyrosine protein kinase Src
sST2	Soluble suppression of tumorigenicity 2 (soluble IL1RL1)
ST2	Suppression of tumorigenicity 2 (IL1RL1)
ST2L	Receptor suppression of tumorigenicity 2 (receptor form of IL1RL1)
STAT6	Signal transducer and activator of transcription 6
T2D	Type 2 diabetes
TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)
TCR	T-cell receptor
TE	Tris-EDTA buffer
TFBS	Transcription factor binding site
TFIID	Transcription factor II D
TGFβ	Transforming growth factor beta
Th1	T helper 1
Th17	T helper 17
Th2	T helper 2
TLE4	Transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)
TLR1	Toll-like receptor 1
TLR4	Toll-like receptor 4
TMEM182	Transmembrane protein 182
TNFα	Tumor necrosis factor alpha
TNXB	Tenascin XB
Treg	Regulatory T cells
TSLP	Thymic stromal lymphopoietin
TSS	Transcription start site
TTP	Tristetraprolin
UBC	University of British Columbia
UPR	Unfolded protein response
USF1	Upstream stimulatory factor 1
USF2	Upstream stimulatory factor 2

Abbreviation	Description
UTR	Untranslated region
UV	Ultraviolet
vST2	Variant form of suppression of tumorigenicity 2
WDR36	WD repeat domain 36
ZPBP2	Zone pellucida binding protein 2

Acknowledgements

Firstly, I would like to acknowledge and whole-heartedly thank my PhD supervisor Dr. Andrew Sandford. Throughout my graduate training, Dr. Sandford has been an inspiring mentor and an incredible example of scientific integrity and ethics. Many thanks Andy for helping me drastically improve my writing, thank you for being so thoughtful, helpful and ever so generous with mentorship, resources and training opportunities. This work would have been absolutely impossible without your unwavering support.

I would like to extend my warmest thanks to Dr. Peter Paré, who trusted me with my first UBC technician job which introduced me to the world of asthma research. Dr. Paré was always very supportive and kind towards me and I am grateful to him. My thanks also go to him for being a member of my supervisory committee, along with Dr. Keith Walley and Dr. Del Dorscheid. Drs. Walley and Dorscheid were generous with their support, resources and time and I sincerely thank them. I would also like to acknowledge the support and kind mentorship of Dr. Denise Daley. Dr. Daley inspires me to want to be better than myself and to move forward despite any hurdles one might encounter along the way.

My heart-felt thanks also go to Amritpal Singh, who has been a great friend and an amazing supporter and introduced me to the joys (and frustrations) of the R programming language. For that and a million more things, I'll forever be thankful to him. Big thanks as well to my very dear friend Dr. Katherine Thain; I will never forget our endless functional assays troubleshooting conversations, while drinking Starbucks coffee in the park.

I also thank all the kind, friendly, and collaborative trainees, technicians, research associates, principal investigators and administrative personnel of the UBC James Hogg research centre.

I would also like to sincerely acknowledge the support of the lung eQTL consortium as represented by Dr. Paré and Dr. Sin at the University of British Columbia, Dr. Bossé and Dr. Laviolette at Université de Laval, Dr. Nickle and Dr. Hao at Merck and Dr. Postma and Dr. Timmens at the University of Groningen.

My acknowledgements to funding from AllerGen Networks of Centres of Excellence, UBC Faculty of Medicine graduate awards and UBC four year fellowship program.

Last but certainly not least, I would like to acknowledge the contribution of my parents, Naima Elachqar and Ahmed Akhabir, who gave me the gifts of love, compassion and ethics and I thank them for always having faith in me, and for their encouragements and prayers. I dearly thank my siblings Kamal and Nadia for their unconditional love and support. My loving thanks go to my amazing daughter Kenza for calling me the best scientist in the world, for loving me, gracing my life day in and day out and smiling at me every morning. My uncles, aunts and cousins have strongly encouraged and supported me from Morocco, Europe and Canada; each and every one of them deserves my sincerest thanks and gratitude.

To my daughter Kenza, the treasure of my life

Chapter 1: Introduction

1.1 The fundamentals of asthma

Asthma is a common chronic inflammatory disease, characterized by shortness of breath, reversible airflow obstruction, airway hyperreactivity and chronic lung inflammation. In 2012, 2.4 million Canadians (8.1% of the population) were afflicted with the disease, according to government statistics (http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/health50a-eng.htm). It is estimated that 300 million people worldwide have asthma and that the prevalence of the disease is increasing (1).

1.1.1 Clinical presentation and management

1.1.1.1 Diagnosis

According to the Global Initiative for Asthma (GINA) (2), asthma should be diagnosed based on the patient's history and symptoms as follows. The presence of any of the following signs increases the likelihood of presence of asthma: wheezing i.e. a high-pitched whistling sound when breathing out, history of cough, worsening at night, recurrent wheeze, recurrent difficulty breathing or chest tightness, nocturnal worsening or seasonal pattern of symptoms, concomitant allergic disease such as eczema, family history of asthma and atopy, occurrence or worsening of symptoms in the presence of triggers such as animals, chemicals, smoke, respiratory infections, etc.

When asthma is suspected based on symptoms and history, lung function testing confirms the diagnosis and provides a quantitative objective assessment of disease severity. Spirometry is a lung function test for measuring airflow limitation and reversibility thereof.

An increase in Forced Expiratory Volume in one second (FEV_1) of ≥ 12% and ≥ 200 ml after administration of a bronchodilator is an indication of reversible airflow obstruction consistent with an asthma diagnosis.

1.1.1.2 Treatment

There are two types of asthma medication: relievers and controllers. 1) Relievers are medications which alleviate symptoms promptly such as short-acting β_2-agonists (e.g. Albuterol) which act by dilating the airways rapidly and rescue the patient from an asthma attack or reduce acute symptoms such as coughing or shortness of breath. Bronchodilators only treat airway tightness and do not act on the lung inflammation. 2) Controller medications act by reducing airway inflammation and are generally taken daily. Examples of controllers are inhaled corticosteroids (e.g. fluticasone), leukotriene receptor antagonists (e.g. Montelukast), long-acting β_2-agonists (e.g. Salmeterol), and theophylline. Inhaled corticosteroids function by repressing gene transcription of inflammatory genes by binding glucocorticoid receptors in the cytoplasm and translocating to the nucleus where they recruit histone deacetylases (gene transcription repressors) as well as inhibit the recruitment of histone acetylases to the promoters of inflammatory genes (3). Different kinds of inhaled steroids have the potential to cause local or systemic side-effects (4).

Leukotriene receptor antagonists bind, as their name indicates, to leukotriene receptors; leukotrienes are products of arachidonic acid metabolism and participate in the pathophysiology of asthma by binding to their receptors on the surface of mast cells, eosinophils, endothelial and smooth muscle cells. Leukotriene receptor antagonists block the pro-inflammatory and other harmful effects of leukotriene signaling (5).

Long-acting β_2-agonists (LABAs) are bronchodilators that provide long-term relief from airway obstruction by binding to β_2-adrenergic receptor on the surface of smooth muscle cells, increasing the activity of adenylyl cyclase, leading to the accumulation of cAMP and the subsequent activation of cAMP-dependent protein kinase (PKA) leading to the functional effect of relaxation on the airways for 12 hours per dose (6). Adding LABA therapy to inhaled corticosteroids is generally a more effective asthma treatment; indeed there is evidence that β_2-adrenergic receptor agonists enhance the glucocorticoid response element-dependent effect of inhaled corticosteroids via the cAMP-PKA signaling pathway, resulting in more efficient anti-inflammatory effect of the corticosteroids (7).

The treatment of asthmatic patients is step-wise and depends on the severity of disease and the level of symptom control. Patients are always given short-acting bronchodilators which are used on an as-needed basis to prevent severe exacerbations and relieve recurrent symptoms. With increasing levels of severity, the patients are prescribed one or more controller medications usually starting with inhaled corticosteroids of increasing dosages. Combination of LABA and inhaled corticosteroids is prescribed if control is not attained, then leukotriene modifiers or theophylline are added to the treatment plan for difficult to treat asthma. Oral corticosteroids are prescribed as a last resort in the case of severe asthma. When control of asthma symptoms is reached and sustained for at least 3 months, treatment can be gradually reduced to decrease the risk of side-effects and maintain only necessary medication for control.

Patient education and regular medical monitoring are key elements to successful symptom control and increase in the quality of life of asthmatic patients.

1.1.2 Genetic associations

1.1.2.1 Overview

Asthma is the result of a complex interaction between environmental factors and genetic variants that confer susceptibility; the heritability of asthma is estimated at 48% to 79% (8). The genetic basis of allergic sensitization, including asthma, has been long recognized with the Human Leukocyte Antigen (*HLA)* locus being the first specific chromosomal region implicated (9). The genetic contribution to asthma can be demonstrated by twin studies where monozygotic twins are more concordant for asthma and other allergic traits than dizygotic twins (10). However, monozygotic twins are not completely concordant for these phenotypes, clearly demonstrating the importance of environmental factors.

Studies of the genetics of asthma have been conducted using family-based designs that detect the co-inheritance (linkage) of genetic variants with the phenotype. The power of this approach is that the genes are detected by virtue of their chromosomal location alone and not on the basis of prior knowledge of their function. Therefore, the entire human genome can be screened in an unbiased fashion. Several novel genes such as *ADAM33*, *DPP10, NPSR1* (*GPRA*), *PHF11* and *HLA-G* were identified as putative susceptibility loci using this approach (11-16). On the other hand, the results of linkage studies have shown poor reproducibility. In a recent meta-analysis of 20 genome-wide linkage studies there were only two chromosomal regions (2p21–p14 and 6p21) that showed significant evidence for linkage in European families, after adjustment for multiple comparisons (17).

An alternative approach is to use hypothesis-driven association studies of specific asthma candidate genes (18, 19). This is most commonly performed using a case-control design although cohort and family-based approaches are possible. However, there has also

been a general lack of reproducibility in these types of studies (20). There are many possible reasons for such inconsistent results and these include small sample sizes, differences in phenotype definition and lack of adequate matching of study subjects for ethnic background (21).

A complicating factor in association studies is the phenomenon of linkage disequilibrium (LD) which is the association between alleles at different sites on a chromosome. LD tends to be lower for polymorphisms that are further apart (due to the effect of recombination), although there is not a simple relationship between LD and genetic distance. The pattern of LD across the genome is not uniform and differs between populations of different ancestry. As a consequence of LD, an association of a genetic polymorphism with a disease outcome does not necessarily imply causality. The polymorphism may be in LD with several nearby variants, any one of which could be the causal locus. On the other hand, the presence of LD reduces the number of polymorphisms that have to be assayed in a given chromosomal region as one variant can act as a surrogate for many others.

Most recently, the association study design has been extended from the examination of a specific candidate gene to an unbiased, genome-wide approach (22). In order to survey the entire genome, a large number of genetic polymorphisms (~0.5-2 million) need to be assayed in each study subject. These studies utilize single nucleotide polymorphisms (SNPs) as they are the most efficiently assayed and frequent type of genetic variant. In the most commonly used approach to genome-wide association studies (GWAS) the genotype frequencies at each SNP are compared between cases and controls. However, such a large number of comparisons require the use of extremely stringent statistical correction to avoid an

overwhelmingly large number of false positive results. For example, in a GWAS consisting of 500,000 SNPs, to survive a Bonferroni correction requires a p value $<1\times10^{-7}$ (0.05/500,000). Therefore, very large sample sizes are needed in such studies if genome-wide statistical significance is to be achieved e.g. several thousand cases and controls. On the other hand, the large amount of genetic data generated in these studies allows efficient correction for differences in ethnic background.

Advances in genotyping technology have made GWAS feasible, although still expensive. As large sample sizes are essential the studies are generally carried out by consortia of investigators who pool their resources.

1.1.2.2 Asthma GWAS studies

Thus far, there have been approximately twenty GWAS to look for susceptibility loci for asthma and related traits. Seventeen of these studies utilized asthma or childhood asthma as the phenotype. The first GWAS for asthma was published in 2007 by Moffatt *et al.*(23) and it demonstrated that hypothesis-driven approaches have limited power (16) to identify susceptibility genes because that study uncovered a novel associated locus on chromosome 17q21 encompassing the genes *ORMDL3, GSDMB* and *ZPBP2.* None of these genes would have been selected in a candidate association study based on current knowledge of the functions of these genes. Nevertheless, this finding has been consistently replicated in independent populations of European ancestry and in other ethnic groups (20, 24-39) and thus chromosome 17q21 seems to be a true asthma susceptibility locus. Only one study found a lack of association and that was in an African American population (24). In general, these replication studies have demonstrated that the associations were stronger for early-onset

asthma. Two of the studies (26, 32) showed an interaction of the polymorphisms in the chromosome 17q21 region with cigarette smoke exposure.

It is currently unclear which gene(s) in this chromosomal region are responsible for the association with asthma or through which mechanisms; efforts are underway to elucidate that question. Moffatt *et al*. (23) demonstrated that the SNPs associated with asthma were also associated with *ORMDL3* expression in lymphoblastoid cell lines (derived from B cells). Halapi and coworkers (33) determined gene expression in peripheral blood leukocytes and found that the polymorphism most strongly associated with asthma in the study by Moffatt *et al*.(23), rs7216389, was significantly correlated with the expression of both the *GSDMB* and *ORMDL3* genes. Expression of these two genes was highly correlated suggesting that they are coordinately regulated. It was later also shown that SNPs within the locus were associated with *ORMDL3* and *GSDMA* gene expression in cord blood mononuclear cells (40).

In order to determine the cause of the chromosome 17q21 genetic association, Verlaan *et al*. (41) performed an extensive, elegant functional study which showed alteration of regulation of genes in the region was due to chromatin remodeling occurring differentially by alleles. These data suggested that more than one causal polymorphism exists in this region and that these SNPs have effects on the regulation of several genes. A similar study showed that allelic variation in the region was a result of the interaction between epigenetic changes and SNPs and reported another SNP, rs4795397 in the promoter region of *ZPBP2*, which associated with allele-specific nucleosome occupancy (42).

The chromosome 17q21 locus posed a challenge to researchers (43) as the functions of the genes composing it are unclear and efforts have since been made to uncover the

mechanism(s) by which this region could be involved in the pathogenesis of asthma. The current understanding of the function of the *ORMDL3* (ORM1-like 3) gene is that it regulates endoplasmic reticulum-mediated calcium signaling resulting in the unfolded protein response (UPR), which is thought to trigger an inflammatory response (44). In an ovalbumin (OVA)-induced asthma mouse model, it was shown that *ORMDL3*'s expression was allergen-inducible in the lung epithelium and that this induction is dependent on STAT6 and not NF-κB; the *in vitro* part of the study demonstrated *ORMDL3* transfection led to the activation of ATF6, a signaling protein of the ER-localized UPR (45). However, more recent *in vitro* studies illustrated that in airway epithelial cell lines, upon siRNA knockdown of *ORMDL3*, no changes were observed in IL-6 or IL-8 production or UPR activation upon stimulation or to gene expression of a number of immune-related genes (46).

Other data implicated ORMDL3 in the regulation of sphingolipid metabolism (47). Sphingolipids are a major component of cell membranes and are involved in numerous functions such as cell proliferation, signal transduction and apoptosis (48). A newly-emerged potential mechanism of the involvement of ORMDL3 in asthma is through the regulation of eosinophil trafficking, recruitment and activation, which was found *in vivo* in a mouse model of allergic airway inflammation and *in vitro* in murine eosinophils (49).

Whether these functions provide an explanation for the association with asthma remains to be determined. ORMDL3 was suggested as the most plausible gene causing the asthma association of chromosome 17q21 locus (23) however there is increasing evidence making the case for the other genes in the region more compelling. A SNP in *GSDMA* was the strongest expression quantitative trait locus (eQTL) in the lung among all the SNPs in the region and human lung epithelial cells expressed abundant messenger RNA (mRNA) and

protein levels of this gene (50); Hao *et al.* presented results of a canonical pathways enrichment analysis that showed GSDMA might be the most likely candidate for causing the asthma association in the region.

The *GSDMB* (gasdermin B) gene is expressed in the epithelial cells of the skin and intestinal tract and appears to be involved in tumorigenesis (51). However, *GSDMB* is also highly expressed in T cells and at low levels in the fetal lung and bronchial epithelium (52). *ZPBP2* (zona pellucida binding protein 2) encodes a protein involved in fertilization and possibly has a structural role in the biogenesis of the acrosome during spermatogenesis (53). More recently, two SNPs within *ZPBP2* were shown to associate with gene expression of the neighboring genes; the study's main finding was sex-specific methylation patterns in the *ZPBP2* promoter but not in *ORMDL3* or the *GSDM* genes proximal promoters; *ZPBP2* differential promoter methylation potentially explained the sex-specific asthma association in the French-Canadian cohort studied (54). Thus, more studies are warranted to elucidate the causal mechanisms behind the asthma association of the *ZPBP2*/*ORMDL3/GSDMB* locus.

The GWAS of childhood asthma conducted by Sleiman *et al*. (55) confirmed the chromosome 17q21 association at a genome-wide significance level. These authors also found several SNPs associated with childhood asthma that were in a large region with high levels of LD on chromosome 1q31 containing *DENND1B* and the 3'end of *CRB1*. In this study, the 1q31 association was originally observed in a population of European descent and validated in an independent European population. Sleiman *et al*. replicated the association of this region in African Americans although interestingly the alternative allele at each polymorphism was associated in this ethnic group. The association of *DENND1BI* with childhood asthma in this GWAS was entirely novel and is indicative of the usefulness of

GWAS in detecting genes that would not have been the subject of a candidate gene approach. Although the *DENND1B1* association was convincing and statistically significant, it has not been replicated in GWAS since. While *CRB1* does not seem to be an attractive asthma candidate, as its reported roles lie in photoreception, *DENND1BI* is of great interest. *DENND1BI* encodes for the DENN/MADD domain containing 1B protein. It is thought to interact with the tumor necrosis factor receptor type 1 to block its signaling and is expressed by dendritic cells (DCs), which play a pivotal role in the inflammatory state characteristic of asthma, and by natural killer T (NKT) lymphocytes. No functional studies have been performed to elucidate the biology behind this gene's putative role in asthma.

In 2009, two additional GWAS using the childhood asthma phenotype were published (56, 57). In the study performed by Himes *et al*. two regions of the genome (on chromosomes 5 and 8) showed associations that approached genome-wide significance. The chromosome 8p21 polymorphism is in the *CHRNA2* (nicotinic cholinergic receptor α2) gene. Nicotinic cholinergic receptor genes have been implicated in lung disease as polymorphisms in them were associated with lung cancer (58, 59) and chronic obstructive pulmonary disease (COPD) (60, 61) and have been reported to influence smoking behavior in GWAS meta-analyses for the number of cigarettes smoked per day and smoking initiation (62, 63). The other noteworthy result of the Himes *et al*. study was the association of *PDE4D* SNPs on chromosome 5p12. The *PDE4D* (cAMP-specific phosphodiesterase 4D) gene encodes an enzyme that hydrolyses cyclic AMP, and thus is upstream of many important signaling pathways. Furthermore, PDE4D plays an important role in airway smooth muscle contractility (64). Variants in the *PDE4D* gene were previously associated with stroke (although the association is controversial (65)) and other disorders (66-68). Several

compounds designed to inhibit PDE4D are in clinical drug development as therapeutics for asthma and COPD (69, 70). A comprehensive review of these efforts has recently been published (71). Himes *et al*. (57) found that chromosome 17q21 polymorphisms were associated with asthma and in the same direction as reported by Moffatt *et al*. (23) but the p values fell short of genome-wide significance (p=0.0007 to 0.02).

In the second 2009 study, Hancock *et al*. (56) studied childhood asthma in 492 trios of Hispanic asthma cases and their parents. As expected from the sample size, no SNP reached genome-wide statistical significance. However, variants on chromosome 9q21 upstream of *TLE4* showed some of the smallest p values for association with asthma in the initial cohort and one of the associations was replicated in an independent population, although it was not significant after Bonferroni correction. Therefore, these data provide suggestive evidence of involvement of this locus in susceptibility to childhood asthma. TLE4 (Transducin-like enhancer of split 4 (E (sp1) homolog, *Drosophila*) is an inhibitor of transcription with no previous association with asthma or related phenotypes. As in the study by Himes and coworkers (57), chromosome 17q21 SNPs in the *ORMDL3* region were associated with asthma but only at modest significance levels (p=0.01 to 0.04). Two polymorphisms in the region of the *PDE4D* gene were also associated with asthma in this population (p=0.02 and 0.04).

Increased immunoglobulin E (IgE) levels and eosinophilia are two of the main characteristics of asthma and are highly correlated with severity of disease (72-75). Results for the first GWAS of total serum IgE were reported in 2008 by Weidinger *et al*. (76). The GWAS population was from Southern Germany and consisted of 1530 individuals. No single SNP showed genome-wide statistical significance. Efforts to follow up the variants that

showed the lowest p value regardless of genome-wide significance were undertaken using a total of 9769 additional samples from four independent populations. Interestingly, the top findings were for SNPs in *FCER1A*, the gene encoding the α chain of the high affinity receptor for IgE on chromosome 1q23. The high affinity receptor for IgE is composed of three subunits α, β and γ; the α subunit is responsible for IgE binding whereas the β and γ subunits mediate the signaling. The gene encoding the β subunit has also been implicated in atopy and asthma in some studies but not others (18). Recently, it has been shown *in vitro* that the α subunit could function as a single-chain receptor and was able to bind IgE and subsequently internalizes and shuttles the antigen into lysosomes (77). These new data suggest that the α chain could be sufficient for IgE signaling. Although this GWAS result is not surprising based on the function of the gene and previous candidate studies of *FCER1A* (78, 79), it is reassuring that this gene was the top hit of an IgE GWAS and can be regarded as a validation of the GWAS approach.

Another interesting finding of the Weidinger *et al*. (76) GWAS was the association with SNP rs12368672. This SNP is located approximately 7 kb from the signal transducer and activator of transcription 6 (*STAT6*) promoter and is in high LD with an intronic SNP in *STAT*6 (rs324011) that has been shown to have allele-specific effects on *STAT6* promoter activity *in vitro* as well as STAT6 mRNA levels *in vivo* (80). STAT6 is a transcription factor that plays an important role in activating genes involved in IgE synthesis (81) and polymorphisms in these genes have previously been associated with serum IgE levels (82, 83).

Both *FCER1A* and *STAT6* are genes with solid biological plausibility to influence IgE levels. However, Weidinger *et al*. (76) also identified a novel gene that had no previous data

to show involvement in atopy or asthma. *RAD50* is located on chromosome 5q31 near the important cytokine gene cluster containing *IL13*, *IL4*, *IL5*, *IL3* and *CSF2*: all of which are involved in the inflammatory state present in the asthmatic lung. *RAD50* encodes a protein responsible for DNA double-strand break repair. However, the 3' end of *RAD50* contains several conserved non-coding sequences and enhancer elements which act as a locus control region for regulation of the neighbouring *IL4* and *IL13* genes (84).

Interestingly, there was a second report associating *RAD50* with asthma in the GWAS performed by Li *et al.* in a Caucasian population with severe or difficult to treat asthma (85). Most probably due to the small sample size (n=473 cases), no single SNP was significantly associated with the phenotype at genome-wide threshold. The most significant association was seen for an intronic SNP in *RAD50*; additional SNPs spanning the region comprising *RAD50* and *IL13* were associated as well. It is of note that SNPs in the locus control region, regulating *IL13* and present in the *RAD50* gene, were found associated with asthma. It is uncertain whether the asthma and total IgE levels associations are truly caused by variants in the *RAD50* gene, in which case efforts should be put towards uncovering yet unknown functions of this enzyme, are surrogates for variants in *IL13*, or are merely affecting *IL13* via its locus control region in *RAD50* (86). Functional follow-up of the region is warranted to fill this knowledge gap.

The other major finding of the study by Li *et al.* (85) was the association of multiple SNPs in the *HLA-DQB1* and *HLA-DRB1* genes, part of the HLA class II region. The role of HLA in antigen presentation and thus in the development of allergy is well-established and there are published HLA genetic association studies of atopy and asthma from as early as 1973 (9). The *HLA-DQB1* and *HLA-DRB1* regions in particular have been previously

associated with asthma (87, 88) and atopy (89). In total, five asthma GWAS reported genome-wide significant associations of HLA loci with childhood and adult asthma (28, 35, 90-92).

Li *et al.* (85) also reported marginal associations of *GSDMB* and *FCER1A* SNPs that had been reported in previous GWAS (23, 76).

Choudhry *et al*. performed an asthma GWAS (93) with a small sample (96 cases with moderate to severe asthma and 88 controls) of individuals from Puerto Rico. As would be expected, no single polymorphism survived correction for multiple comparisons and thus no genome-wide significance was observed. The most significant associations were observed for SNPs on chromosomes 5q23 and 13q13.

In 2009, Gudbjartsson *et al*. performed a GWAS with blood eosinophil count as the phenotype using an initial cohort of 9392 Icelandic individuals and followed-up the top SNPs in different populations (94). The most significant association with eosinophil levels in the initial cohort was with a SNP in *SH2B3* (a.k.a. *LNK*) on chromosome 12q, which codes for an adaptor protein thought to be involved in T cell signaling. *In vivo* studies suggest that the SH2B3 adaptor protein is implicated in cytokine signaling and hematopoietic homeostasis (95). The associated SNP (rs3184504) was previously associated with susceptibility to celiac disease (96) and type-1 diabetes (97). This suggests that SH2B3 protein might be important in general immune system processes potentially common to all immune diseases. This conclusion is supported by the observation that rs3184504 was associated not just with eosinophils count but with the levels of several blood cell types (94). Another SNP that was associated at a genome-wide significance level was in the *GATA2* gene on chromosome 3q,

which encodes a transcription factor involved in hematopoietic cell development and proliferation and seems particularly critical for eosinophil development (98).

The results of Gudbjartsson *et al*. (94) also highlighted genes in an essential pathway for T helper type 2 (Th2) immunity, namely *IL1RL1* and the gene encoding its ligand *IL33*. SNPs in the *IL1RL1-IL18R1* region were significantly associated with eosinophil count at the genome-wide level and were also very strongly associated with asthma in different populations. This region has been previously implicated in asthma, inflammation and a number of immune disorders (99-101). The association of an *IL33* SNP with eosinophil level was strong ($p<10^{-5}$) but not genome-wide significant. However, the same SNP was also associated with asthma at a similar level of significance.

Gudbjartsson *et al*. found two other loci that were significantly associated with eosinophil count using the genome-wide cutoff. The first was the region containing the *IL5* gene in the cytokine gene cluster on chromosome 5q. IL-5 is a cytokine that plays important roles in the Th2 type immune response that characterizes asthma; it stimulates B cells to increase IgE production and is a major regulatory protein for eosinophils (102). The second was a region near the *IKZF2* gene (a.k.a. Helios) a member of the zinc finger family that is expressed in the thymus and is thought to have a role in T cell development (103). Another interesting locus, though one that did not survive the genome-wide statistical significance cutoff, is on chromosome 5q and contains the *WDR36* and *TSLP* genes. Interestingly, the same region was identified in a GWAS for eosinophilic esophagitis (104). WDR36 is thought to participate in the regulation of T cell activation (105). TSLP (thymic stromal lymphopoietin) is expressed in the airway epithelium and stimulates dendritic cells to

promote the differentiation of naïve $CD4^+$ T cells to Th2 cells (106). Previous candidate gene studies have shown association of *TSLP* variants with asthma (107-110) and IgE levels (108).

A GWAS conducted for asthma by Mathias *et al*. (111) was performed in two populations of African-descent with a total sample size of 1864 individuals. Although the study yielded a number of nearly genome-wide significant loci, there were inherent limitations due to the differences in SNP allele frequencies and probably LD patterns between the two populations studied (Barbados founders and African-Americans), although both are of African descent. In addition, the major findings were not successfully replicated in European cohorts. However, a meta-analysis using less stringent statistical thresholds highlighted three putative loci for asthma susceptibility in the combined African-descent populations: *DPP10, ADRA1B,* and *PRNP*. The *DPP10* gene on chromosome 2q14 encodes the inactive dipeptidyl peptidase 10 and although this protein possesses no protease activity it is involved in regulation of potassium channels via direct binding (112). It is noteworthy that *DPP10* was one of the genes implicated in the pathogenesis of asthma by positional cloning (12). This association has been relatively well replicated in different populations (113, 114). *ADRA1B* codes for the α_{1b}-adrenergic receptor, which is a member of the G-protein-coupled receptor family. It is expressed in the lung and been shown to induce proliferation of vascular smooth muscle cells (115).

The *PRNP* gene on chromosome 20p encodes the prion protein; the function of this protein is not well characterized but speculative roles include ion transport and signal transduction. It has been linked to inflammation by regulating phagocytosis (116) and apoptosis (117) and was shown to be under-expressed in alveolar macrophages of allergic asthmatics versus controls (118).

The largest GWAS for asthma to date included ~10,000 cases composed of samples from different studies and ~16,000 controls (35). The analysis of all subjects yielded *HLA-DQ* as the strongest associated locus for the phenotype of doctor-diagnosed asthma. SNPs in the *IL18R1*, *GSDMA*/*GSDMB*, *IL33*, *SMAD3* and *IL2RB* genes were the most significant findings after *HLA-DQ*. The *IL18R1* association may be due to its neighboring gene *IL1RL1* as the SNP is in an LD block that includes a group of amino acid-changing polymorphisms in *IL1RL1*. Interestingly, *IL33* was also among the most significant findings of the same study and this cytokine is the ligand for *IL1RL1*. The analysis of childhood asthma samples resulted in the replication of the association of the chromosome 17 locus that includes *ORDML3*.

It is interesting that this study highlighted the inflammatory signaling axis IL1RL1/IL-33. IL-33 and its receptor IL1RL1 have been consistently implicated in a wide range of inflammatory and immune disorders including asthma (119). Most importantly, variants in the *IL33* and *IL1RL1* genes were among the top findings of the eosinophil count GWAS (94). Although the phenotypes of the two studies are different, it is safe to conclude that the IL1RL1/IL-33 pathway is of importance in asthma susceptibility. Moffatt *et al*. also reported a suggestive association of the *TSLP* variant rs1837253 ($P= 3\times10^{-6}$), previously associated with asthma in candidate gene studies.

The chromosome 17q21, *IL1RL1*/*IL18R1*, *TSLP* and *IL33* loci were replicated in an EVE consortium-led meta-analysis of asthma GWAS in North-American multi-ethnic groups (37) at genome-wide statistical significance levels. In the Torgerson *et al*. study, SNPs in those genes were associated with asthma in all three North-American ethnic groups studied (European, African and Latino), and were among 75 SNPs in 15 chromosomal locations

found to be associated with a p value smaller than the cut-off of 10^{-6} in the combined meta-analysis of three populations. Another important result of this study was the asthma association of *PYH1N1* (Pyrin and HIN domain family member 1) which was unique to the study subjects of African descent; an unprecedented finding.

Further replication of the association of *TSLP* rs1837253 with asthma came from a large Japanese GWAS of adult asthma (90). Hirota *et al*. replicated the association at the MHC locus as well as the association with rs1837253 on chromosome 5q22 upstream of *TSLP*; and reported three new adult asthma loci in the Japanese population studied: rs7686660 on chromosome 4q31, a SNP close to a lung function-associated SNP although only weak LD exists between the SNPs; rs10508372 on chromosome 10p14 lies in a gene desert, the closest gene is *GATA3* which is an important regulator of Th2 cell differentiation and rs1701704 on chromosome 12q13 in a LD block containing 13 genes and 2 kb upstream from *IKZF4*, a gene involved in T regulatory cell regulation.

Previously, the chromosome 17 locus was more consistently associated with childhood asthma probably explaining its absence from the Hirota *et al*. Japanese GWAS significant hits. It was however soon after replicated again in a European GWAS for severe asthma (38) and constituted the most significant association in their meta-analysis.

Together, these GWAS findings make the chromosome 17q21 locus, the locus on Chr 2 containing the *IL1RL1* gene, the locus on Chr 9 containing the *IL33* gene, and the Chr 5 *TSLP* SNP rs1837253 the most consistently replicated in GWAS along with the HLA genes.

The most recently reported asthma GWAS was undertaken to study a more specific phenotype, namely early childhood asthma with severe exacerbations (120). DNA from Danish children with recurrent asthma-related acute hospitalizations between 2 and 6 years of

age, was used for the discovery GWAS; replication was performed using the publically available Gabriel childhood onset data (35) as well as samples from European birth cohorts. Five loci were associated with the phenotype at the genome-wide significance level; four of them were the now-established asthma loci: *GSDMB/ORMDL3* in the chromosome 17 locus, *IL1RL1*, *IL33* and *RAD50*. SNP rs6967330 in cadherin-related family member 3 (*CDHR3*) is a new locus reported in this study significantly associated with early childhood asthma with severe exacerbations with an odds ratio of 1.45 ($p=1.4\times10^{-8}$).

Even though the specific role of CDHR3 is unknown, it belongs to the cadherin family of calcium-dependent membrane proteins involved in cell-cell adhesion; Bonnelykke *et al* and others (121) showed that it is expressed in the airway epithelium in an allele-specific manner; the asthma risk allele was associated with greater cell-surface expression of the CDHR3 protein.

In addition to CDHR3, a number of other genes encoding structural proteins have been identified in the GWAS published thus far. In the eosinophil count GWAS (94), *TNXB* (tenascin XB) on chromosome 6p21 was among the genes that approached genome-wide significance level. Tenascin XB is believed to regulate collagen fibril deposition in the extracellular matrix (122). The gene might be involved in systemic lupus erythematosus via its association with complement 4 genes (123). In the Moffatt *et al*. study (23), *ANTXR1* was among the genome-wide significantly associated genes; it codes for anthrax toxin receptor 1 and is thought to play a role in cell adhesion and migration and binds directly to actin (124). In the same study (23), *NRG1* was identified; this gene encodes for neuregulin 1 which been reported to play a role in cell adhesion (125). In the Li *et al*. study (85), variants in *COL5A3* and *MKLN1* were associated with asthma though not at the genome-wide significance level.

These genes encode a member of the collagen family and muskmelon 1, respectively. Collagen is one of the major components of the extracellular matrix and is implicated in the airway remodeling observed in asthma (126) as an important component causing thickening of the epithelial basement membrane. Muskelin 1 is a cytoplasmic protein involved in regulation of cytoskeleton arrangement; it also participates in cell spreading by facilitating communication between the cytoplasm and the nucleus (127). Additional genes encoding members of the Collagen family were reported at genome-wide significant levels: *COL18A1* and *COL6A5* (128). Multiple SNPs in the *CTNNA3* gene were the top findings of a Korean GWAS for toluene-diisocyanate-induced asthma (129), albeit not reaching genome-wide statistical significance (no SNP did in that study, probably due to the small sample size: 84 cases and 263 controls). The *CTNNA3* gene on chromosome 10q22 encodes for catenin (cadherin-associated protein) α3. This protein is involved in cell-cell adhesion and has been linked to Alzheimer disease in some studies (130, 131). A SNP in *CTNNA3* was among the top hits in a GWAS of nicotine dependence (132).

In conclusion, it is clear from the collective GWAS for asthma that the airway epithelium plays a central role in asthma, both in a structural capacity (adhesion proteins etc) and in an effector immune capacity (secretion of immune proteins such as TSLP and IL-33).

1.2 Single nucleotide polymorphisms and disease susceptibility

SNPs are the markers of choice for genetic association studies for the ease of genotyping with recently-developed technology. The inherent LD properties of genomic loci make association studies possible with a limited number of SNPs that can account for the majority of genetic variability. As of the latest reference genome build (37.4), the number of human

SNPs in the NIH-funded dbSNP public database was 53,567,890 (38,072,522 of them validated).

1.2.1 Potential function of SNPs

Between any two human individuals, DNA sequence varies by an average of 0.1% to 0.4% (133, 134) but this nucleotide variation, along with larger scale variation, greatly contributes to phenotypic variation between human beings including disease susceptibility. SNPs occur every 500 to 1000 nucleotides, at a variable density throughout the genome. Association studies are used to identify those SNPs responsible for different traits in populations. Not all SNPs have the potential to induce a functional consequence, so in order to identify functional SNPs a good place to start is to examine the location of the SNPs. Changes in gene function due to SNPs can occur in two ways. First, SNPs can cause changes in the function of the protein by alteration of the primary structure, either as a result of exonic SNPs or intronic SNPs in splice sites. Second, SNPs located in any other part of the gene can be functional by resulting in expression change; these SNPs are generally called regulatory SNPs.

1.2.1.1 Promoter SNPs

The promoter of a protein-coding gene is the genomic locus where the transcriptional machinery is orchestrated, initiated by the binding of transcription factor II D (TFIID) to the TATA box typically located 25bp from transcription start sites (TSS). Gene promoters can have one or multiple TSS (135). Generally, promoters contain several different types of regulatory elements for the binding of transcription factors, the Mediator complex, chromatin remodeling complexes and histone acetylases for transcription activation and aiding

communication with the RNA polymerase. There are also reports of promoters containing MicroRNA (miRNA)-binding sequences that infer enhancer activity by miRNA binding (136).

Regulatory SNPs present in the promoter can disrupt binding sites for a variety of regulatory proteins, possibly miRNA and of course transcription machinery proteins. One example of a regulatory SNP in a promoter is rs4073 in the *IL8* promoter which confers increased susceptibility to idiopathic pulmonary fibrosis by resulting in allele-specific up-regulation of *IL8* expression (137).

1.2.1.2 5′ Untranslated region (UTR) SNPs

The 5′ UTR is the regulatory locus at the 5′ end of protein-coding genes, which starts at the TSS and ends immediately before the start codon of the region that translates into protein. The majority of translation regulation occurs in the 5′UTR (138) where scanning of the translational machinery happens in order to find the initiation codon AUG. Consequently, SNPs occurring in the 5′UTR can be functional and affect gene regulation by disrupting any of the cis-regulatory features in the 5′UTR such as binding sites for activators or repressor proteins, internal ribosome entry sites (IRES), iron-responsive elements (IRE) (139) and/or the GC richness or the upstream open reading frames (140, 141).

1.2.1.3 Exonic SNPs

Exons are the only gene component that directly affects protein structure, by translation of codons into amino acids. Exonic SNPs are translatable in the protein gene product and can be non-synonymous or synonymous. They can be synonymous, i.e. the new codon generated by

the SNP allele gives rise to the same amino acid or non-synonymous, i.e. the codon change results in a different amino acid in the protein. Non-synonymous SNPs can be:

- nonsense SNPs when the affected codon changes into a termination codon which generally results in a non-functional truncated protein.
- missense SNPs when the codon change results in an amino acid change in the protein.

Missense SNPs can cause a structural change in the protein; potentially altering its conformation, ligand affinity or ability to interact with other proteins. Missense SNPs can have from no or minimal effect to disastrous effect on the resulting protein. The majority of deleterious non-synonymous SNPs are rare with frequencies below 1% (142). The mutation that causes most cases of Sickle cell anemia is SNP rs334 in the *HBB* gene, which encodes the beta subunit of hemoglobin. This SNP is a missense SNP that causes A to T change resulting in the replacement of a glutamic acid residue with a valine in the protein.

1.2.1.4 Intronic SNPs

Introns are regions of a gene that are transcribed into the precursor mRNA molecule then spliced out to generate mature mRNA. Introns are very important for the regulation of gene expression (143) and indeed host a variety of regulatory binding sites such as enhancers, may contain alternative promoters, encode non-coding RNA (144) and have an essential role in regulation of splicing (145).

Consequently, intronic SNPs can interfere with any of the above-mentioned regulatory roles and have functional consequences such as altering the binding affinity of an enhancer or disrupting the splicing process and hence resulting in an alteration of gene expression levels.

A haplotype comprised of three intronic SNPs in the *OCA2* gene accounts for most of the variability in human eye color in Caucasians (146).

1.2.1.5 3′UTR SNPs

The 3′UTR of a gene is the trailer genomic locus that is translated in the mRNA molecule and immediately follows the translation termination codon and precedes the poly-A tail that indicates the end of transcription. The 3′UTR is an important regulatory region controlling processes such as mRNA stability, nuclear transport, and translation. This region contains binding loci for miRNA molecules (147), Adenylate-uridylate-rich elements (AU-rich elements a.k.a. AREs) (148), repressor proteins (149, 150) and regulators of polyadenylation (151).

miRNAs are short single-stranded non-coding RNA molecules that physically interact by partial base pairing to miRNA response elements in the 3′UTR of a target mRNA resulting in mRNA degradation or down-regulation of translation according to the classical view of miRNA's role in post-transcriptional regulation; however an increasing number of published reports show this binding results in promoting translation (152). More recent research show expanded roles for miRNA as well as locations of binding other than 3′UTR, namely 5′UTR (153), promoters (136) and coding regions (154).

miRNA have been reported to play important roles in important biological process such as cell fate (155) and lipid homeostasis (156) as well as in pathological states including lung inflammation (147, 157, 158).

AU-rich elements host proteins that stabilize mRNA such as HuR (159) and AUF-1 (160) or alternatively destabilize the mRNA and accelerate its decay such as TTP (161) and KHSRP (162).

As a result, SNPs located in the 3′UTR of a gene can play an important regulatory role by disrupting a binding site for regulatory proteins for any of the processes mentioned above and can thus have considerable functional consequences. For instance, a SNP which is located in the binding site of a miRNA can potentially decrease the affinity of the miRNA to the messenger RNA molecule, thus hindering the down-regulation resulting from miRNA binding and resulting in increased translation of mRNA into protein. SNPs can also potentially alter the secondary structure of the mRNA and result in a perturbed localization and translation (163).

1.2.1.6 Intergenic SNPs

As the name indicates, intergenic regions are DNA regions not containing protein-coding genes. GWAS for a multitude of phenotypes showed that intergenic SNPs constituted about 40% of phenotype-associated loci (22) even though genotyping microarrays are biased towards containing protein-coding SNPs (164). More recently, the Encyclopedia of Coding Elements (ENCODE) researchers showed that regulatory DNA was an enriched location for common phenotype-associated variants (165). Hemberg *et al* shed light on the important fact that the majority of conserved non-coding DNA serves as promoter-distal regulatory protein binding sites (166).

SNPs occurring in intergenic regions can be functional through a variety of ways. They could disrupt long-range regulatory interactions (167) such as at enhancer binding sites

(168); enhancer elements are genomic sites that recruit proteins serving to unravel repressed chromatin and facilitate the building of transcriptional complexes. Intergenic regions also contain insulator elements, which function as a boundary to a single gene's regulatory interactions between enhancer/repressor elements and transcriptional complexes. Insulators can function to block transcription (169) and can also prevent gene silencing by stopping the spread of chromatin (170). Clear mechanisms of insulator functioning are still lacking but there is a good deal of evidence that CCCTC-binding factor (CTCF) is responsible for at least some of these insulating functions (171, 172, 172, 173) although new ENCODE data show long-range regulation can bypass this known mechanism of insulation, suggesting other means of insulation yet undiscovered (167).

In intergenic regions, SNPs can also occur in large intergenic non-coding RNAs (lincRNAs). LincRNAs have been shown to associate with disease and to participate in a variety of biological processes (174, 175).

An example of an intergenic SNP is rs6983267 in the chromosome 8q24 locus, a gene desert which is the most common region for somatic amplification in all cancers and was robustly associated with a variety of cancers in GWAS. rs6983267 was determined to be a functional SNP that confers cancer risk by differential binding of Transcription factor 7-like 2 (TCF7L2) to the enhancer element the SNP is located in and this binding in turn regulates the expression of MYC oncogene (176, 177).

1.3 Interleukin 1 Receptor Ligand 1

Interleukin 1 receptor-like 1 (IL1RL1), also called T1, ST2, DER4 and FIT-1, is a member of the interleukin 1 super-family (178) but does not bind interleukin 1 (IL1) (179). IL1RL1 was

an orphan receptor until the description of its ligand, interleukin-33 (IL-33) in 2005 (180). Since then, IL-33 binding to IL1RL1 has been associated with a variety of disease states and in particular to inflammatory processes as outlined in recent reviews (100, 119).

1.3.1 *IL1RL1* gene and proteins

The *IL1RL1* gene is located in chromosome 2q12 and is composed of 11 exons (181). A number of IL1 family members reside in the immediate vicinity of the *IL1RL1* gene namely *IL1R2, IL1R1, IL1RL2,* IL-18 receptor 1 (*IL18R1)* and IL-18 receptor accessory protein (*IL18RAP*) (Figure **1-1**).

Figure 1-1 *IL1RL1* and neighboring genes

The region spans about 300 kb and is in high LD. There is evidence for the involvement of the genes surrounding *IL1RL1* in human and experimental disease, and therefore the causal locus responsible for genetic association signals from this region is difficult to determine.

The *IL18R1* and *IL18RAP* genes code for components of the heterodimeric IL-18 receptor (the α and β chains, respectively). The cytokine IL-18 is a modulator of innate and adaptive immune responses that acts by inducing T helper type 1 (Th1) cell differentiation and T and NK cell maturation or by activating IgE production and Th2 cell differentiation in specific cytokine milieus (182-184). High levels of *IL18* mRNA and protein were observed

in the lungs of smokers and COPD patients (185) and expression of an alternatively-spliced variant of *IL18R1* was associated with atopy (186). *IL18R1* expression was also higher in human primary keratinocytes derived from skin lesions of psoriasis and atopic dermatitis patients compared with healthy controls (187). IL-18 signaling has been implicated in host defense (188) and rheumatoid arthritis (189). Additionally, genetic association data have implicated the IL-18 receptor genes in asthmatic and allergic phenotypes (190, 191).

1.3.1.1 Expression of IL1RL1

IL1RL1 gene transcription is initiated at two separate promoters: a proximal promoter and a distal promoter. The alternative usage of these two promoters leads to differential 3' processing of the mRNA isoforms (192, 193). Three known isoforms are produced: isoform 1 which codes for IL1RL1-b (aka ST2L), a long membrane-bound protein; isoform 2 which codes for IL1RL1-a (aka sST2), a short soluble protein; a third isoform which codes for IL1RL1-c (aka vST2) (192), a variant membrane-anchored form of the protein. The soluble form of IL1RL1 corresponds to the extra-cellular domain of IL1RL1-b except for nine amino-acids in the C-terminal region.

IL1RL1-b is mainly expressed on cells of hematopoietic provenance, mainly Th2 cells (194). It has been shown that binding of IL1RL1-b with its ligand on the surface of basophils, eosinophils and mast cells promotes their activation (195), increases adhesion and survival (196) and degranulation (197), respectively. However, a chronic model of allergic disease showed a protective effect of IL-33/IL1RL1-b on mast cell activation in response to antigens (198) where a prolonged IL-33 treatment reduced antigen-mediated degranulation.

IL-33/IL1RL1-b has also been shown to play a role in activating macrophages (199-201) and stimulating DCs to produce chemokines CCL17 and CCL12 (202) as well as increased expression of MHC class II and costimulatory molecules CD86, CD40 and CD80 (203). It was also shown *in vivo* that IL-33 was a potent stimulant of type 2 innate lymphoid cells, causing their expansion and heightened cytokine production (204).

The short form, IL1RL1-a, is expressed by various cells including epithelial cells, endothelial cells, fibroblasts and smooth muscle cells. This expression is augmented upon stimulation with IL-1α, IL-1β, TNFα, LPS and other factors inducing cell stress such as cardiac infarction and hypoxia (205). The tissue distribution of IL1RL1-a seems to be relatively ubiquitous, with the highest levels found in the lung followed by the heart and the brain (206).

Several studies show that the membrane-bound IL1RL1 protein acts as a specific marker for Th2 cells (194). *In vitro* blockade of IL1RL1 signaling with recombinant IL1RL1 protein to compete with the endogenous receptor resulted in the abrogation of differentiation to and activation of Th2, but not Th1, effector cells (207). Interestingly, IL1RL1 has been found to play a considerable role in the newly discovered immune type2 effector leukocytes, known as nuocytes or group 2 innate lymphoid cells (ILC2) (204, 208, 209). An IL13-GFP mouse model was utilized to define these as cells not corresponding to a previously known leukocyte lineage that express ICOS, IL1RL1 and IL25R. The ILC2's function includes the innate immune response to helminth infection with *Nippostrongylus brasiliensis* by secretion of high levels of IL13 in response to IL-25 and IL-33 (208).

The ligand for IL1RL1 is IL-33, a member of the interleukin 1 family (180). The signaling of IL1RL1-b binding to IL-33 results in the activation of the Mitogen-Activated

Protein kinases ERK1, ERK2 and p38 and the subsequent activation of NF-κB (180, 196). IL1RL1-a corresponds to the extra-cellular domain of IL1RL1-b and *in vitro* studies have shown that it can also bind IL-33 and act as a decoy receptor inhibiting the activation of NF-κB (210) and the subsequent inflammatory response. This was confirmed in an animal model where introduction of soluble IL1RL1 decreased pro-inflammatory cytokine (IL-4, IL-5 and IL-13) production in a murine asthma model after treatment with IL-33. It was shown that this protective effect of the soluble IL1RL1 seems to be IL-10 dependent in an animal model of ischemia reperfusion injury (211).

1.3.1.2 IL1RL1 in asthma and other respiratory diseases

Increased eosinophil count is a phenotype associated with the majority of asthma cases and correlates with severity of the disease as well as response to glucocorticoid treatment (212). *In vivo* studies demonstrated that eosinophilic inflammation following allergic stimuli is significantly decreased following treatments with recombinant soluble IL1RL1 or antibodies directed against the membrane-anchored IL1RL1 (213, 214). This anti-inflammatory effect persists in the absence of eosinophils, such as in mast cells purified from eosinophil-deficient mice (215).

Soluble IL1RL1 has been clearly implicated in experimental asthma; it was shown to be sufficient to reduce experimental allergic airway inflammation using an intravenous IL1RL1 gene transfer mouse model (214) and $ST2^{-/-}$ knockdown mice showed decreased airway inflammation (216). IL1RL1-positive Th2 cells were shown to be responsible for the enduring airway hyperresponsiveness (AHR) after cessation of allergen challenge in mice; this inflammatory feature was vastly abrogated by an IL1RL1 receptor antibody (217).

Additionally, anti-IL-33 antibody or soluble IL1RL1 treatment significantly reduced airway inflammation in an OVA-mediated experimental allergic asthma model (218). However studies of experimental airway allergic inflammation demonstrated a clear increase of both receptor and soluble IL1RL1 proteins in response to allergen challenge (216, 219). IL1RL1 expression has also been shown to be higher in human asthmatic lungs (99); soluble IL1RL1 levels has been shown to be greater in the serum and induced sputum of asthmatic patients compared with controls, and these higher levels correlated with disease activity, lung function decrease as well as an increase in the serum levels of inflammatory cytokines (220).

The IL-33/IL1RL1-b axis has also been implicated in asthma-associated airway fibrosis; increased IL1RL1-b was observed on the surface of fibrocytes from allergen-exacerbated asthmatic patients and recombinant IL-33 treatment resulted in the promotion of fibrocyte migration and proliferation (221); *in vitro* studies using lung fibroblasts showed IL-33 induced eotaxin/CCL11 production (219). IL-33/IL1RL1-b was also shown to promote airway remodeling features namely collagen synthesis by human and murine lung fibroblasts; in addition IL-33 levels (but not IL-13) were associated with reticular basement membrane thickness in endobronchial biopsies from severe steroid-resistant pediatric asthma patients (222).

Since the late 1990s, genetic studies have shown linkage of chromosome 2q with asthma, lung function (as assessed by the FEV_1 expressed as percentage of the vital capacity (FEV_1%VC), a common clinically-useful index for airflow limitation), eosinophilia and IgE levels (223-225).

Polymorphisms in *IL18R1*, a gene in tight LD with *IL1RL1*, were associated with asthma, atopic asthma and airway hyperresponsiveness using a candidate gene approach in a

Danish population and the association consistently replicated in two other European populations (226). In the same year, another candidate gene study documented significant association of the gene cluster containing *IL1RL1, IL18R1* and *IL18RAP* with asthma and atopy in a Dutch population (190). Additional association evidence was reported by the same group using pathway analysis to detect gene-gene interactions in the Toll Like Receptor (TLR)-related pathway. IL1RL1-a has been shown to down-regulate gene expression of TLR4 and TLR1 *in vitro* after treatment with LPS and *in vivo* in a LPS-induced shock mouse model (227).

Twenty-nine genes implicated in TLR regulation were selected for a pathway analysis in Dutch populations (228). *IL1RL1* SNPs were associated with allergy and asthma phenotypes as single SNPs although the significance did not survive multiple testing correction. In addition, when gene × gene interactions were tested using the multifactor dimensionality reduction approach, *IL1RL1* SNPs were identified as interacting factors in analyses of IgE phenotypes (228).

In a study performed by our group in collaboration with others, we investigated three Canadian and one Australian population but failed to detect any significant association with *IL1RL1* that survived correction for multiple comparisons (229). The same cohorts, in addition to one American population, were used in an association study of genes in the vitamin D pathway with asthma and atopy phenotypes. *IL1RL1* SNPs were selected for this study based on the fact that *IL1RL1* was shown to be transcriptionally regulated by vitamin D (230). The genotyping covered more variants of *IL1RL1* than the initial study and the number of candidate genes was substantially less (11 versus 120 genes). Significant associations of these variants were observed with asthma and atopy phenotypes (231).

Given the role of eosinophils in the pathogenesis of asthma, alleles that associate with increased eosinophil count could be detrimental in terms of asthma risk and severity. In a GWAS of eosinophil count in an Icelandic population, a SNP in *IL1RL1* (rs1420101) showed the most significant association. The A allele of rs1420101 associated with increased eosinophil count and in further analyses with increased serum IgE as well as with three asthma phenotypes (asthma, atopic asthma, non-atopic asthma) in nine European populations and one east Asian population (94). rs1420101 is an intronic SNP which is in high LD (r^2 greater than 80%) with a large number of other variants in *IL1RL1*, *IL18R1* and *IL18RAP*; this group of SNPs contains mostly intronic SNPs in addition to a coding-synonymous and a few 3′ and 5′ UTR SNPs.

It is of note that an association of a SNP in *IL33* (rs3939286) with eosinophil count, asthma and atopic asthma was reported in the same study, although the *IL33* association with eosinophil count did not reach genome-wide significance (1.8×10^{-5}). The same *IL33* SNP was associated with nasal polyposis in a Belgian population in a candidate gene study (232). Wu *et al*. used GWAS data of childhood asthma in a Mexican population (233) to perform a candidate gene analysis. In this study, 237 genes were selected from published studies of human and animal model of asthma which had at least one SNP associated with an asthma phenotype. They reported *IL1RL1* among the most significant associations. Furthermore, their results were subjected to multi-marker analysis, which confirmed *IL1RL1* as a significant finding as well as *IL18R1*.

In asthma GWAS, the *IL1RL1* locus is one of the most replicated associated loci at genome-wide statistical significance levels and in different populations (35, 37, 38, 234). This genetic association is certainly very well supported by the biology of IL1RL1 and

related proteins. IL-33 is secreted by the airway epithelium in response to stress such as allergens or viruses, and binds to IL1RL1-b on the surface of immune cells. There are excellent reviews about the central role of the epithelium in initiating and sustaining immune responses (235); IL-33/ IL1RL1-b plays a crucial role in that process.

Although the binding of IL1RL1-b and its ligand IL-33 is known to trigger the NF-κB signaling pathway leading to the transcription of Th2-type cytokines, the role of the soluble IL1RL1-a remains unclear. Several animal models and *in vitro* studies show that IL1RL1-a prevents IL-33/ IL1RL1-b signaling and consequently attenuates inflammation, indicating its role as a negative regulator of the pro-inflammatory IL-33/ IL1RL1-b axis. Human data on the other hand clearly demonstrate a consistent increase of IL1RL1-a in an array of pathological conditions as well as the correlation of this increase with severity. Additionally, there was a report of an animal study showing that mice deficient in IL1RL1 (receptor and soluble) showed attenuated airway inflammation after challenge with an allergen (236); together these animal and human studies suggest that IL1RL1-a might be participating in the excessive inflammation observed in asthma. However, the model used for this study was the transgenic TCR-mouse model; these animals are pre-disposed to autoimmune disorders because they carry rearranged TCR α and β genes from a diabetogenic T cell clone.

The above studies do not seem to be consistent with the antagonist role of IL1RL1-a but rather indicate a possible involvement in promoting the pathology of asthma. An alternative explanation would be that the increase of IL1RL1-a is a means of preventing an exaggerated immune response but either occurs too late or is insufficient to remedy the pathological state.

Evidently, soluble IL1RL1 plays a role in the regulation of the immune response, notably in severe disease. Exactly what that role is and the mechanisms underlying it need to be clarified in order to develop efficient strategies for developing therapeutics using the IL1RL1 proteins.

Human data in COPD seem to indicate an involvement of soluble IL1RL1 in the early stages of COPD (237). This study however involved a small number of patients and needs replication.

1.3.1.3 IL1RL1 in allergy and immune disorders

A SNP in *IL1RL1*, rs3771175 was associated at genome-wide statistical significance level in a GWAS for allergic sensitization, defined by the presence of allergen-specific IgE against common environmental antigens (238).

A SNP in the distal promoter region of *IL1RL1*, rs6543116 (-26999G/A), was associated with increased risk for atopic dermatitis (239). This study suggested a functional effect of rs6543116 as the A allele correlated with an up-regulation of the gene transcription as well as serum levels. An atopic dermatitis GWAS in a Japanese population confirmed *IL1RL1* as a susceptibility gene (rs13015714 with a P value of 8.36×10^{-18}) (240).

Castano *et al*. found a significant association of *IL1RL1* SNPs with chronic rhinosinusitis using a cohort of surgery-unresponsive chronic rhinitis patients; this association was stronger in more severe disease (241). An association of serum levels of IL-33 and SNPs in the *IL33* gene with Japanese cedar pollinosis was reported; Japanese cedar pollinosis is the most common form of allergic rhinitis in Japan (242); and IL-33 and IL1RL1 levels were shown to be elevated in allergic rhinitis patients (243).

IL1RL1 and closely linked genes have been implicated in an array of autoimmune diseases. Levels of IL1RL1-a have been shown to be increased in various conditions such as Systemic Lupus Erythematosus and rheumatoid arthritis (RA) (244). Mok *et al*. found that elevated serum IL1RL1-a levels in Systemic Lupus Erythematosus patients correlated with disease activity (245). To date, GWAS performed in Chinese and European populations have not found association of *IL1RL1* SNPs with RA (246, 247).

Studies in animal models demonstrated that recombinant IL1RL1-a protein, or anti-IL1RL1 antibody could significantly attenuate the severity of experimental arthritis (248, 249) and *Il1rl1* knock-out mice were shown to develop a less severe form of disease and had reduced pro-inflammatory cytokine production. Additionally, human studies have shown increased levels of IL-33 and IL1RL1 in RA synovium paralleling increased inflammation (250, 251). Studies in animal models strongly suggest that the involvement of IL-33/IL1RL1 in RA is through triggering mast cell degranulation in the RA synovium (252). Although there is good evidence for a role of IL-33/IL1RL1 in human and experimental arthritis, no SNPs in these genes were found associated with susceptibility to RA in GWAS data (253, 254).

The IL1RL1/IL-33 axis was implicated in inflammatory bowel disease (IBD) for the first time in 2010 in a study characterizing IL1RL1 and IL-33 protein and mRNA expression in IBD patients (253). There was an increase in soluble IL1RL1 levels in the gut, which was mainly associated with the active state of ulcerative colitis, indicating a possible negative regulation of the IL1RL1/IL-33 pathway in order to dampen the inflammation. Pastorelli *et al*. confirmed the observation of elevated levels of IL1RL1 and IL-33 in the serum and mucosa of IBD patients; they also showed that anti-TNF decreased IL1RL1-b levels and

increased the soluble isoform making more decoy receptor available in order to sequester IL-33 and reduce the inflammation (255).

A SNP 1.5 kb downstream of *IL18RAP* (rs917997) was associated with susceptibility to IBD in a Dutch population; the same SNP was associated with celiac disease in three European populations (256) and in a GWAS of celiac disease in a UK population along with another SNP in the intergenic region between *IL1RL1* and *IL18R1* (rs13015714) (257). The same SNP downstream of *IL18RAP* (rs917997) was associated with Crohn's disease in a GWAS (258). SNPs in *IL1RL1* (rs13015714 and rs2058660) and in *IL33* (rs3939286) were found associated with IBD in an Italian population (259).

These genetic and mechanistic data suggest that IL1RL1/IL-33 plays a role in the gut mucosa similar to the airway epithelium i.e. IL1RL1 isoform A/IL-33 eliciting a Th2 immune response and IL1RL1-a serving as a negative regulator.

A genetic linkage study implicated chromosome 2q14 with type 2 diabetes with a logarithm of odds (LOD) score of 4.5 (260). A LOD score greater than 3 indicates evidence that the locus linkage did not occur by chance with odds >1000 to 1. There is evidence that soluble IL1RL1 directly acts on macrophages to suppress their ability to produce pro-inflammatory cytokines (227). Macrophages are instrumental in diabetes pathogenesis. In an animal model of diabetes (multiple low-dose streptozotocin-induced diabetes), Mensah-Brown *et al*. (261) showed that specific disruption of the *Il1rl1* gene significantly enhanced inflammation in their mouse model as estimated by an increase in cellular infiltration in pancreatic islets and a reduction in cells immuno-positive for insulin. Human data show that IL1RL1-a levels are increased in type 2 diabetes patients (262) and correlate with diabetes markers such as glucose levels, circulating triglycerides (263).

In summary, the available data on the involvement of IL1RL1 and its ligand IL-33 in immune and autoimmune disorders are reasonably consistent; a clearer understanding of the balance between IL1RL1-b/IL-33, IL1RL1-a/IL-33 and its regulation is needed in order to make that axis an attractive target for therapeutic intervention.

1.3.1.4 IL1RL1 in other disorders

In order to understand more fully the biology of IL1RL1, knowledge of its involvement in other disorders is necessary. Understanding the similarities and differences between the roles of IL1RL1 in different tissues and disorders would further the quest to finding general and disease-specific IL1RL1 pathways and mechanisms.

1.3.1.4.1 Cardiovascular disease

IL1RL1 has been extensively studied in the context of cardiovascular disease and it was generally found that increased level of IL1RL1 is correlated with poor prognosis in different instances of cardiovascular disease, pointing perhaps to a role of soluble IL1RL1 as marker for the severity of the immune response. Increased IL1RL1-a is indicative of an overwhelming immune response that is hard to control and thus leads to unfavorable outcome in cardiovascular disease patients, such as after a myocardial infarction (MI). A summary of representative studies is presented below.

In vitro and animal model studies have demonstrated that IL-33/IL1RL1-b signaling protects cardiomyocytes from apoptosis by suppressing Caspase 3 activity and promoting the expression of anti-apoptotic proteins *in vitro* and improves survival in experimental MI animals (264); *in vivo* data also showed that IL-33 induces eosinophilic pericarditis whereas IL1RL1-a prevents eosinophilia and improves systolic function (265). Human studies have

shown an increase of soluble IL1RL1 after myocardial stress or injury, and MI (205, 266); IL1RL1-a levels were associated with diastolic load (267), cardiac abnormalities on electrocardiogram, poor prognosis in dyspneic and MI patients (268, 269) and increased all-cause and cardiovascular mortality (270). In a study following 150 patients admitted to hospital with acutely destabilized heart failure, multiple serum samples were collected between admission and discharge and soluble IL1RL1 levels were measured. The results showed that IL1RL1-a levels were a powerful predictor of 90-day mortality. Indeed, IL1RL1-a serum levels are considered a reliable biomarker for heart failure (271-273) and are a promising predictor of left ventricular and infarct recovery after acute MI (274).

The company Critical Diagnostics in collaboration with the Brigham and Women's Hospital in Boston has developed a diagnosis kit called Presage that uses soluble IL1RL1 levels for diagnosis and prognosis of cardiovascular disease. This kit is approved by the U.S. Food and Drug Administration for clinical use to contribute to prognosis of chronic heart failure patients (275).

1.3.1.4.2 Infections

Studies of IL1RL1 show it confers protection from infection, which is consistent with its involvement in the Th2 immune response. The increase in IL1RL1 signaling skews T cells to Th2 and prevents a harmful parasite-specific Th1 polarized response. A summary of representative studies is presented below.

Using the cecal ligation and puncture (CLP) model in BALB/c mice, a widely used model for experimental sepsis, it was demonstrated that IL-33 treatment was protective from peritonitis and enhanced bacterial clearance (276). The study also showed that the protective

effect of IL-33 treatment was achieved via the inhibition of a TLR-signaling-induced protein, GRK2 in human neutrophils. GRK2 plays a prominent role in sepsis as it down-regulates CXCR2 (a receptor for IL-8, a chemokine that attracts neutrophils to infection sites) thus leading to inefficient clearance of bacteria. Other *in vivo* studies confirmed IL1RL1 involvement by showing the reversal of sepsis-related immune-compromised state in CLP ST2 knockout mice (277) and heightened susceptibility of ST2 knockout mice to polymicrobial sepsis due to an impaired bacteria killing by IL1RL1-lacking macrophages (278).

In agreement with the role of IL1RL1 proteins in the promotion of Th2 responses, mRNA levels of both receptor and soluble forms of the IL1RL1 transcript were shown to be up-regulated in an animal model of *Toxoplasma gondii* parasitic infection and this up-regulation correlated with protection from the infection (279). In addition, $ST2^{-/-}$ knockout mice demonstrated increased susceptibility and more severe disease compared to wild type mice as assessed by weight loss, increased parasite transcript levels and typical disease pathology.

Despite *in vivo* and *in vitro* data, not all reports on human subjects support the correlation of IL1RL1-a levels with sepsis severity and/or outcome in human subjects (276, 280, 281).

In 2008, a small study of a Somali population reported an association of a SNP in the 3′UTR of *IL18R1* (rs3213733) with variability in Rubella vaccine-induced humoral immunity (282). Variants in *IL18R1* are in high LD with SNPs in *IL1RL1*. It is interesting that the same SNP was recently shown to be associated with asthma in two different studies, in Mexican

and Japanese populations (94, 283). As LD patterns differ between populations, this suggests a potential functional role of this SNP in regulating gene expression/function.

Additional evidence for a role of the IL-33/IL1RL1 axis in host immune defense comes from *in vivo* studies showing the protective role of IL-33 in intestinal infection with nematodes (284, 285).

A recent study highlighted a new IL-33/IL1RL1 disease involvement in coxsackievirus B5-induced experimental pancreatitis; mice lacking IL-33 or IL1RL1 demonstrated more severe pancreatic disease, increased weight loss and heightened virus load than wild type mice (286).

1.3.1.4.3 Liver and kidney disorders

In a candidate gene association study of the course of hepatitis C in a Japanese population, 103 genes including *IL1RL1* and *IL18R1* were investigated (287). SNPs in both these genes as well as other genes involved in immune responses were significantly associated with serum levels of alanine aminotransferase (ALT) indicating an involvement in liver inflammation. ALT levels are routinely used as a diagnostic test of liver function and elevated levels are an indicator of infections and other disorders. Nevertheless, this group's data were not corrected for multiple testing and thus need to be replicated in other populations. It has been shown by others that IL1RL1-a serum levels along with IL-33 levels were indeed increased in chronic hepatitis C and correlated with ALT levels in Chinese patients (288). Preliminary data also show increased serum levels of the two proteins in liver failure patients (289). Animal studies have also shown increased hepatitis severity and inflammatory cells influx to the liver of ST2-deficient mice (290) as well as an over-

expression of *Il1rl1* and *Il33* mRNA in mouse fibrotic liver which was replicated in human tissue sections (291).

IL-33/IL1RL1-a was implicated in kidney disease in a mouse model of acute kidney injury where treatment with a soluble IL1RL1-a fusion protein showed lesser disease as assessed by $CD4^+$ T infiltration to the kidneys, lower serum creatinine levels and decreased acute tubular necrosis and apoptosis; the inverse effect (worsened disease) was observed upon treatment with recombinant IL-33 (292). In human subjects, elevated serum levels of IL1RL1-a were observed in chronic kidney disease patients (293) but in another study of biomarkers for identification of incident chronic kidney disease, IL1RL1-a levels association failed to reach statistical significance after correction for multiple testing (294).

1.3.2 Concluding remarks

The *IL1RL1* gene has been clearly shown to be an asthma gene in numerous GWAS. Several polymorphisms were reported from the chromosome 2 region encompassing *IL1RL1*, indicating independent signals as the polymorphisms are not all in LD with each other. This is interesting and warrants functional analysis in order to identify the putative causal polymorphisms and their mechanism of action. Animal and human studies show that IL1RL1 proteins are important in the inflammatory cascade leading to several features of asthma pathogenesis. The receptor isoform signals after binding of IL-33 on the surface of Th2 cells and innate cells leading to pro-inflammatory gene up-regulation. The shorter soluble isoform of IL1RL1 also binds IL-33, preventing the inflammatory signaling through the receptor. IL1RL1 proteins and the ligand IL-33 play a role in a number of human disorders including allergic diseases, infections and cardiovascular conditions.

My thesis work regarding the *IL1RL1* gene was aimed at identifying the putative causal polymorphisms to explain the asthma genetic association and performing functional analysis to investigate the mechanistic roles of the polymorphisms.

1.4 Thymic Stromal Lymphopoietin

1.4.1 TSLP gene and proteins

Thymic stromal lymphopoietin (TSLP) is an IL2-family cytokine, originally identified as a growth factor in the conditioned medium of a murine stromal cell line Z210R.1 (295). TSLP stands as a pivotal protein in the interface between the environment and cellular responses (296) as it is secreted by epithelial cells in the lung, skin and gut (297) in response to a diversity of stimuli such as viruses (298, 299), TLR ligands (300), cigarette smoke (301), diesel exhaust (302), allergens (303-305) and some chemical compounds (306, 307) and leads to Th2 polarization; other cells such as fibroblasts, smooth muscle cells (308) and mast cells (309) also express TSLP.

The TSLP receptor (TSLPR) is a heterodimer composed of the IL-7 receptor alpha chain and a chain specific to the TSLP receptor called cytokine receptor-like factor 2 (CRLF2) (310); it is expressed in the lung on the surface of dendritic cells (311), B cells (312), T cells (313), innate immune cells (314, 315), smooth muscle cells (316), mast cells (317) and can also be expressed on airway epithelial cells (299) and its production seems to be controlled by the NF-κB pathway. The binding of TSLP to its receptor activates the STAT3 and/or STAT5 pathways by phosphorylation of Jak and Src kinases; (318-321) resulting in the activation of various cytokine and chemokine gene expression (Figure 1-2).

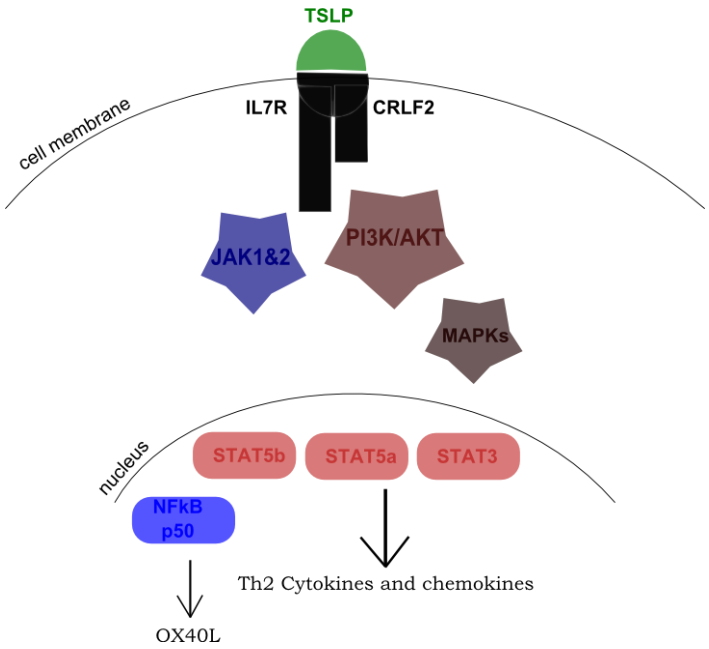

Figure 1-2 TSLP signaling

On lung dendritic cells, TSLP induces NF-κB activation which leads to OX40L gene up-regulation by way of the p50 subunit.

The *TSLP* gene is located on chromosome 5q22, centromeric to the Th2-gene cluster on 5q31 (*IL5*, *IL13*, *IL4*). *TSLP* is comprised of 5 exons and encodes 2 transcript variants: TSLP isoform 1, which is the longest isoform and TSLP isoform 2 which differs in the 5′UTR starting from a different start codon and has a shorter N-terminal domain.

1.4.2 TSLP in the homeostatic state

TSLP is constitutively expressed in the gut by intestinal epithelial cells and plays a role in the immune tolerance of commensal organisms (322, 323). In the mouse gut, TSLP has been shown to promote Regulatory T cell (Treg) development and differentiation of tolerogenic DCs (324); this process has also been demonstrated in the context of maternal fetal

interaction in order to ensure a successful pregnancy by the induction of Treg cells by TSLP-induced DCs (325).

Using a TSLPR-knockout mouse model, it has been shown that TSLP plays a role in lymphoid development by promoting the expansion and survival of $CD4^+$ thymocytes and peripheral T cells (326). In the absence of exogenous antigens or cytokines, TSLP -activated DCs were shown to induce a sustained expansion of autologous naïve $CD4^+$ T cells thus providing evidence for a role for TSLP in homeostatic proliferation of naive T cells mimicking the homeostatic expansion induced by self-peptide-MHC complexes (327); TSLP-activated DCs are also capable of promoting the generation and proliferation of $CD4^+$ $CD25^+$ $FOXP3^+$ regulatory T cells in the human thymus. *In vivo* data showed a direct action of TSLP on $CD8^+$ T cells resulting in their increased survival (328). Hence, these data together support that TSLP is important for T cell homeostasis.

1.4.3 TSLP in allergic disease

Animal studies have strongly implicated TSLP in allergic disease as it has been shown to be sufficient to induce allergic airway inflammation in a lung-specific TSLP transgene mouse model (329) as well as spontaneous inflammation with the hallmarks of atopic dermatitis in skin-specific inducible TSLP transgene mouse model (330). Additionally, mice lacking the TSLP receptor ($TSLPR^{-/-}$) failed to develop an inflammatory lung response to inhaled antigen unless supplemented with wild-type $CD4^+$ T cells (331), and the $TSLPR^{-/-}$ mice also do not develop allergic skin inflammation (106).

Treating OVA-induced allergic rhinitis mice with a neutralizing anti-TSLP antibody inhibited disease development (332). Some animal studies advance the idea of the

involvement of TSLP in the phenomenon of the atopic march (333, 334), which stipulates that atopic dermatitis in early life eventually leads to the development of allergic rhinitis and asthma later in life (335).

Although the murine and human TSLP proteins only share 43% homology, they seem to play similar roles. In humans, expression of TSLP is increased in the asthmatic lung and correlates with disease severity (336, 337). The classical inflammatory pathway to explain the role of TSLP in airway inflammation is the activation of DCs to engender an inflammatory Th2 phenotype (lacking IL10 production) (321). TSLP/TSLPR interaction on DCs results in the rapid up-regulation of OX40 ligand and co-stimulatory molecules CD80 and CD86 on the DC surface as well as the production of chemokines for immune cell recruitment (311); the DCs then migrate to the lymph node when they interact with naïve T helper cells to polarize them into Th2 cells.

A study containing human and *in vitro* data showed another disease pathway by which TSLP promotes an inflammatory environment in the lung: TSLP interacts directly with Treg cells dampening their tolerogenic capacities (338); indeed Tregs in asthma have decreased IL-10 production and lessened suppressive activity (339). TSLP has also been reported to be up-regulated in chronic rhinosinusitis (340).

The first report of genetic linkage of *TSLP* with allergic disease was of the SNP rs2289276 with IgE phenotypes in Costa-Rican girls (108). There was no linkage for male subjects in the same region. Sex-specific effects of TSLP were previously observed in lung disease in mice (341).

In candidate gene association studies in Japanese populations, rs3806933 and rs2289276 were associated with lower FEV_1% predicted and lower FEV_1/FVC ratio but not

lung function decline (342); additionally, the same two SNPs were associated with childhood and adult asthma and rs2289278 was associated with pulmonary function (109). Harada *et al*. also showed in their study that corticosteroids inhibited the TSLP expression stimulated by poly (I:C) treatment.

He *et al*. from our laboratory first reported the association of the T allele of rs1837253 with protection from asthma and AHR in one Australian and three Canadian asthma cohorts; the SNP was the most significant finding of the study (107). The association of rs1837253-T allele with decreased asthma risk was replicated, by the same group who first reported the linkage of *TSLP* with allergic phenotype, in a candidate study in only white males (110). The same SNP (in the same direction) was significantly associated with protection against allergic rhinitis in asthmatic children in a candidate gene study (343).

rs1837253 is presently one of the most consistently reported SNPs in asthma GWAS in different ethnic groups (35, 37, 90, 344).

1.4.4 TSLP in other disorders

Intestinal DCs display a tolerogenic phenotype which is driven by their contact with intestinal epithelial cells (345). Intestinal epithelial cells constitutively express TSLP, which regulates the local tolerogenic DCs, hence participating in immune tolerance in the gut (346, 347).

In addition to its participation in the homeostatic tolerance in the gut, TSLP has been implicated in IBD; the two main IBD disorders are Crohn's disease and ulcerative colitis.

TSLP protects against helminthic infections by promoting Th2-type immunity; additionally, in a Dextran sodium sulfate-induced colitis mouse model, the lack of the TSLP

receptor leads to exaggerated disease in the form of increased weight loss and heightened pro-inflammatory cytokine production (322). However, in another study of an experimental colitis mouse model, TSLP expression was decreased (324); the apparent contradiction is due to the fact that ulcerative colitis is classically thought of as a Th2 disorder when in fact it does not fit into the Th1/Th2 paradigm (348). TSLP was found to be down-regulated in non-inflamed Crohn's disease patients compared with controls in a genome-wide gene expression study in endoscopic biopsies from Crohn's disease patients (349), consistent with the Th1 nature of Crohn's disease. TSLP-receptor-deficient mice were more susceptible to autoimmune gastritis (350).

In a collagen-induced arthritis mouse model, TSLP treatment worsened arthritis and joint destruction and TSLPR knock-out mice demonstrated considerably lessened disease (351). In humans, TSLP was found to be elevated in skin biopsies from patients with diffuse cutaneous systemic sclerosis and interacted with the pro-fibrotic cytokines IL-13 and TGFβ (352). Elevated TSLP levels were also observed in systemic sclerosis patients; it was reported in the same study that bleomycin-induced fibrosis was reduced in TSLPR-deficient mice (353).

In addition to the established implication of TSLP in the promotion of Th2-type inflammation, there is increasing evidence for its involvement in Th17-type inflammation (354, 355).

TSLP was shown to be up-regulated in human atherosclerotic lesions whilst normal human vessels do not express TSLP; the over-expression of TSLP was concomitant with Th17 inflammation activation (356). In the hepatitis C virus-infected liver, TSLP was shown to induce DCs to polarize T helper cells into Th17 cells (355). In addition, SNPs in *TSLP*

were found associated in a candidate gene study of Graves' disease in a Taiwanese population, TSLP serum levels were also elevated in Graves' disease patients compared to controls and this was associated with increased Th17 cell differentiation in those patients (357).

TSLP has also been implicated in the biology of cancers. Knockdown of TSLP was shown to result in decreased growth and metastasis in breast cancer cells; *in vivo* data also support the effect of lack of TSLP in decreasing tumor progression in breast cancer animal models (358).

It was also found that pancreatic cancer-associated fibroblasts release TSLP, thereby decreasing the Th1/Th2 cells ratio by priming DCs present in the tumor stroma and tumor-draining lymph nodes; a decreased Th1/Th2 ratio is associated with a reduced survival of pancreatic cancer patients (359).

1.4.5 Concluding remarks

The involvement of the *TSLP* gene in asthma has been clearly established in GWAS, with the solid replication of the rs1837253 SNP's association in several studies of different populations. The association is thus unique in the sense that the multiple replications were not only at the gene level but also at the SNP level. The biology of TSLP indicates its importance in the upstream stages of the inflammatory cascade leading to several features of asthma pathogenesis. The lung epithelium secretes TSLP in response to a variety of stimuli and in turn TSLP acts upon many effector cells causing an amplification of inflammatory signaling. The fact that TSLP also plays a role in non-pathological states such as gut homeostasis and protection against helminth infections intimates the importance of

investigating the asthma-specific molecular mechanisms of TSLP action. Therefore, my thesis work was aimed at investigating the role of rs1837253 in terms of the mechanism potentially causing the asthma association.

1.5 Research question and hypotheses

TSLP and *IL1RL1* genes are clearly important in asthma susceptibility according to the numerous GWAS significant findings in diverse populations. There is high biological plausibility for these genetic findings and Figure 1-3 summarizes the expression and roles of TSLP and IL1RL1/IL33 in asthma pathophysiology.

The purpose of this thesis is to firstly identify putative functional polymorphisms that are causing the asthma associations and secondly uncover the molecular mechanisms of those associated polymorphisms. I hypothesized that functional alleles of SNPs associated with asthma risk in the *IL1RL1* gene lead to decreased expression of the soluble isoform of IL1RL1 and/or increased expression of the receptor form. Additionally, I hypothesized that the *TSLP* SNP rs1837253 is the functional polymorphism causing the asthma association and that the protective T allele leads to decreased expression of TSLP.

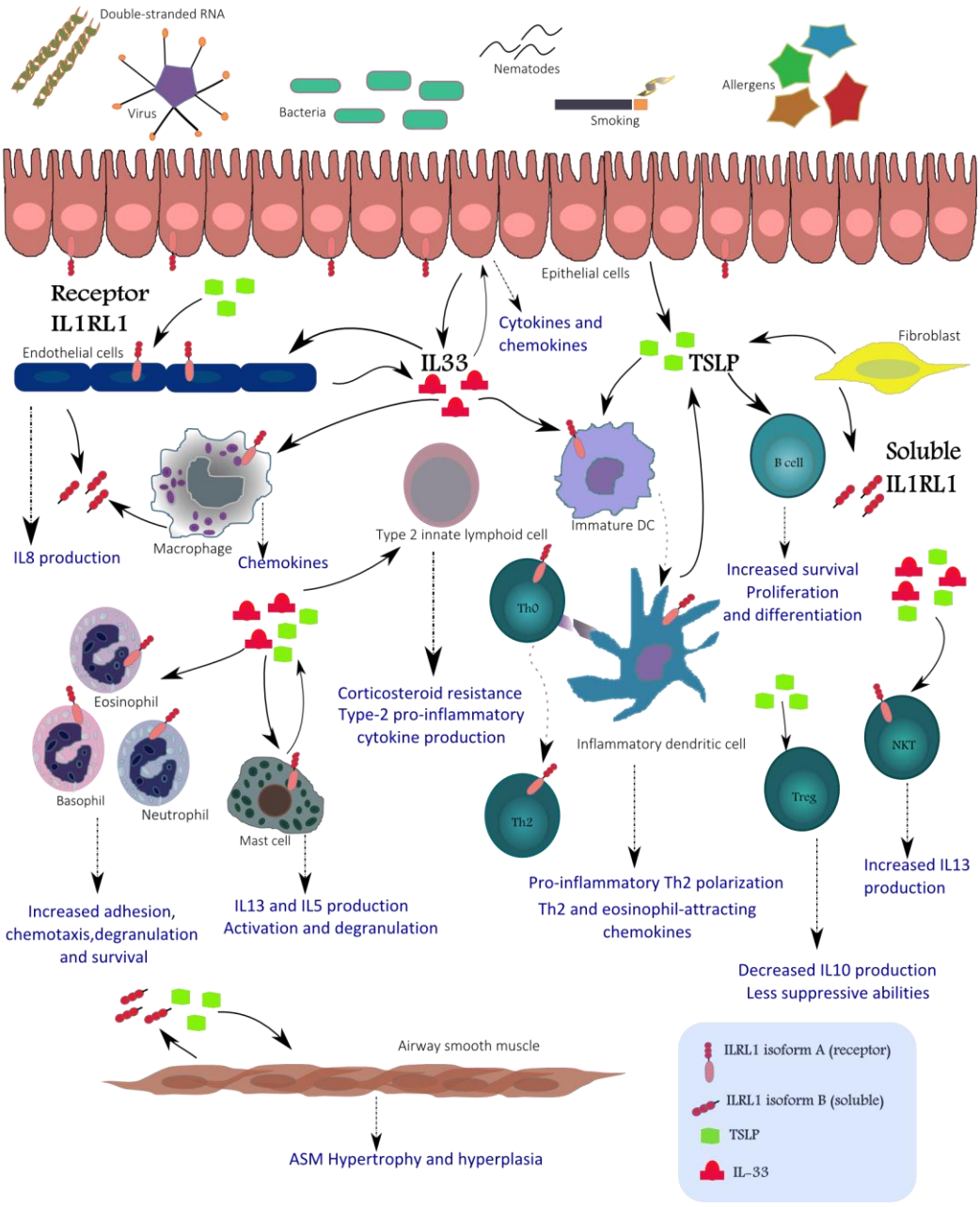

Figure 1-3 TSLP and IL1RL1 expression and cellular pathways in the lung

Chapter 2: Methodology

2.1 Samples

2.1.1 IIAM lung samples

Lungs were obtained from the International Institute for the Advancement of Medicine (IIAM), which provides non-transplantable organs and tissues to biomedical researchers for the purpose of research and education. The IIAM samples were stored at -80°C in the James Hogg Research Centre biobank. DNA was extracted from 18 formalin-fixed, paraffin-embedded IIAM lungs using the QIAamp DNA mini kit (Qiagen catalog # 51304) and total RNA was extracted from the same samples using High Pure RNA Paraffin kit (Roche catalog # 03270289001) following the manufacturer's procedures.

2.1.1 Blood samples

DNA and RNA samples from 208 subjects were graciously offered by Drs. Edith Chen and Greg Miller from the Department of Psychology, University of British Columbia. These samples were from blood drawn from asthmatics and medically healthy children aged 9 to 18 from Vancouver, British Columbia. The children were recruited for a study approved by the Research Ethics Board of the University of British Columbia (360).

2.1.2 CAPPS samples

The Canadian Asthma Primary Prevention Study (CAPPS) (361-363) was established by Drs. Allan Becker and Moira Chan-Yeung in the early 1990s from the two Canadian cities Vancouver and Winnipeg. 545 mothers of high asthma risk infants were recruited during the

third trimester of pregnancy; high risk was defined as having at least one first-degree asthmatic relative or two first-degree relatives with IgE-mediated allergic disease. The primary purpose of the study was to assess the effectiveness of an intervention on asthma incidence. Thus the families were randomized to a control arm or an intervention arm. The intervention consisted of active avoidance of common allergens (house dust mite, pets, and environmental tobacco smoke), encouraging breast-feeding and delaying of introduction of solid foods. Blood samples were collected and questionnaires filled at the time of birth and when the children were 1, 7 and 15 years of age.

DNA was extracted using QIAamp DNA kits (Qiagen) and quantified using PicoGreen® assays; PicoGreen® is a reagent that fluoresces upon binding double-stranded DNA and exhibits higher sensitivity and specificity than quantification by ultraviolet (UV) absorption. CAPPS DNA samples were processed in our laboratory and plated aliquots were sent to the Centre National de Genotypage (CNG, Paris, France) for whole-genome genotyping as part of the GABRIEL consortium large-scale GWAS of asthma.

During the CAPPS 15-year follow-up (which occurred during the year 2010), blood samples were available from 195 children. Freshly-collected blood in red-top tubes was centrifuged and serum collected and stored at -80°C. I used the CAPPS serum samples for IL1RL1 protein quantification.

2.2 TaqMan® genotyping

DNA samples were genotyped using TaqMan® SNP genotyping assays (Applied Biosystems Inc.). Each TaqMan® SNP genotyping assay contains a pair of sequence-specific forward and reverse primers as well as two minor groove binder (MGB) probes around the SNP of

interest with identical sequences except in the SNP site; each allele-specific probe is composed of a 5′ reporter VIC® or FAM™ dye to differentiate between the two SNP alleles and a 3′ non-fluorescent quencher. As long as a TaqMan® probe is intact, the reporter dye does not emit fluorescence. The probe binds to the SNP flanking sequence and if a sample contains the allele corresponding to the probe, annealing occurs with no mismatch and the probe is degraded by 5′ → 3′ exonuclease cleavage during extension of the PCR product. Upon cleavage of the probe, the dye emits fluorescence as it is no longer quenched by the close proximity of the quencher; a probe located on a sequence that does not contain its specific allele binds with one nucleotide mismatch in the middle, and is merely displaced during extension and hence keeps its integrity and its reporter dye does not emit fluorescence.

The assays were subjected to quality control by first genotyping purchased Coriell DNA samples (Coriell Institute for Medical Research) which are very good quality DNA samples that have been genotyped as part of the International HapMap Project. 5ng of DNA or no template negative control (water or TE buffer instead of DNA) per well were plated in 384-well plates and dried-down by evaporation at room temperature in a biosafety cabinet overnight. The subsequent day, each well received 5µl of PCR mastermix containing 10X TaqMan® universal PCR master mix and 40X TaqMan® SNP genotyping assay. PCR was performed in the ABI 7900HT Prism® sequence detection instrument using the following protocol: a *Taq* DNA polymerase activation stage of 95°C for 10 min followed by 40 cycles of [denaturation at 92°C for 15 s – annealing and extension at 60°C for 1 min]. At the end of the PCR run, amplification plots were checked and threshold cycle (Ct) values recorded as QC measures of the assay. Negative controls should not amplify and thus are expected to

have no Ct values i.e. their amplification curve should not intersect with the threshold line (set in the linear phase of the amplification plot that corresponds to the exponential phase of the PCR) to ensure no contamination occurred on the plate. Subsequently, an allelic discrimination run was performed on the same instrument in order to assign genotypes to the samples; fluorescence was measured for each reporter dye and a resulting allelic discrimination plot was produced. Up to four clusters are observed, one cluster for negative controls (no amplification) with no or background-level fluorescence from either probe, two homozygote clusters with fluorescence from only one reporter dye and a heterozygote with equal fluorescence from both probes. The resulting genotypes were compared with publically-available genotypes for Coriell samples. An assay was deemed to have passed my QC measures if Coriell samples amplified with reasonable Ct values (less than 30 cycles) and all the genotypes obtained matched the HapMap genotypes. In case of disagreement between my genotyping and the public genotypes, I sent off PCR products from the discrepant samples for sequencing at the NAPS unit sequencing facility at UBC to ascertain the genotype from a third party.

Unknown genotype DNA samples were genotyped following the same procedure as the Coriell QC plate; 20 positive controls (Coriell samples) and 20 negative controls (no-template) were used in each 384-well plate.

2.3 Gene expression studies

2.3.1 House-keeping gene selection

Before performing gene expression experiments for my genes of interest in lung tissue-derived RNA samples, and in the absence of relevant literature to use, I proceeded to select

the best house-keeping gene to utilize for standardization; a good housekeeping gene should have high stable expression that is unaltered by experimental conditions or treatments or by the disease state of the individuals the samples were derived from. I selected twelve common house-keeping genes to test for stability of gene expression in our tissue of choice. Table 2-1 indicates the names and description of our chosen common reference genes.

Gene name	Abbreviation	Gene function	Assay ID
18S rRNA	18SRNA	Ribosomal RNA	Hs99999901_s1
ribosomal protein, large, P0	RPLP0	Catalysis of protein synthesis	HS99999902_m1
ß-actin	ACTB	Cytoskeletal structural protein	Hs99999903_m1
Cyclophilin A	PPIA	Serine-threonine phosphatase inhibitor	Hs99999904_m1
Glyceraldehyde- 3 -phosphate dehydrogenase	GAPDH	Glycolysis enzyme	Hs99999905_m1
Phosphoglycerokinase	PGK1	Glycolysis enzyme	Hs99999906_m1
ß2-Microglobulin	B2M	Cytoskeletal protein involved in cell locomotion	Hs99999907_m1
ß-Glucuronidase	GUSB	Exoglycosidase in lysosomes	Hs99999908_m1
Hypoxanthine ribosyl transferase	HPRT1	Metabolic salvage of purines	Hs99999909_m1
TATA-binding protein	TBP	Transcription by RNA polymerases	Hs99999910_m1
Transferrin receptor	TFRC	Cellular iron uptake	Hs99999911_m1
Guanine nucleotide-binding protein, ß-peptide 2-like 1	GNB2L1	Binding and anchorage of protein kinase C	Hs00272002_m1

Table 2-1 House-keeping genes candidates

RNA was extracted from IIAM lung tissues and reverse transcription was performed to convert the RNA to cDNA.

Gene expression assays of the twelve genes were performed on 384-well plates in triplicate using TaqMan gene expression assays (Applied Biosystems) comprised of unlabelled PCR primers and a 6-FAM™ dye-labeled TaqMan® MGB probe.

During the PCR, the probe anneals to a specific sequence between the primer sites. Similar to the chemistry used for the TaqMan® genotyping assay, the fluorescence is produced when the quencher dye and the reporter dye are separated from the DNA polymerase cleaving the probe when the latter is hybridized to its specific sequence on the cDNA molecules.

The assays were ran on an ABI Prism® 7900HT sequence detection system (Applied Biosystems) following the standard manufacturer's protocol using 20 ng of starting cDNA

template per well. Standard curves were obtained by serial dilutions of reverse-transcribed commercially-available universal RNA (Stratagene).

Gene expression results were analyzed manually in Excel worksheets using the standard curve method.

The stability of the twelve tested reference genes (Table 2-1) was assessed using three publically available software packages: BestKeeper, which runs pair-wise correlation analysis of all gene pairs and calculates the geometric mean of the most stably-expressed genes (364), GeNorm, which calculates two values: gene expression stability measure M and mean pair-wise variation V between each gene and the rest of the genes (365) and NormFinder, which determines the most stable gene as determined by the gene with the least intra- and inter-group expression variation (366). We selected the most stable reference gene as the most replicated highly ranked by all three algorithms.

2.3.2 Quantitative PCR (qPCR) experiments

For both *TSLP* and *IL1RL1* genes, I used commercially available TaqMan® gene expression assays (Applied Biosystems Inc.) to determine transcript expression in lung and blood samples. Table 2-2 indicates the assays used for each gene. The assays were run in triplicate in 384-well plate format and performed on an ABI Prism® 7900HT sequence detection system (Applied Biosystems Inc.).

Gene	Isoform targeted	Taqman assay ID
TSLP	Both long and short isoforms	Hs00263639_m1
TSLP	Long isoform	Hs00263633_m1
IL1RL1	Short isoform	Hs01073297_m1
IL1RL1	Long isoform	Hs01073295_m1

Table 2-2 TaqMan gene expression assay ID for qPCR of *TSLP* and *IL1RL1*

2.3.3 Allele-specific expression assay

In order to further investigate whether rs1837253 had an effect on *TSLP* gene expression, I conducted allele-specific expression assays using TaqMan® SNP genotyping probes. The advantage of an allele-specific approach as opposed to genotype-specific approach is the fact that in the former, each allele's expression is assessed in a single heterozygote individual, hence controlling for any other genetic variability not related to the SNP in question as well as for any extrinsic influences on gene expression such as environmental effects.

For allele-specific assays, heterozygote samples for a coding SNP are needed. The principle of the assay is to compare cDNA probe fluorescence between alleles; the cDNA signal is standardized to the genomic DNA (gDNA) signal, which should have equal signal from both probes or otherwise would be used to correct for any technology-related probe signal inequality due to preferential binding of one probe over the other.

Since rs1837253 is not in the *TSLP* mRNA, I selected rs2289276, a SNP in the 5′UTR of *TSLP* for expression measurement. If rs1837253 does indeed affect *TSLP* expression by allele, one chromosome (allele) from a heterozygote sample will engender a greater number of mRNA molecules than the other chromosome (alternative allele). By measuring the cDNA synthesized from this RNA, an increased signal from the probe detecting the allele located on the over-expressed chromosome will be observed. As a consequence, the ratio of the signals from VIC® and FAM™ probes will depart from 1, which would be observed in the case of no allele-specific differential expression. The extent of LD between the SNP of interest and the marker coding SNP is of no relevance for this assay, as the direction of the departure from 1 is inconsequential, only the presence or absence of departure from 1 is informative as it pertains to an allele-specific difference.

2.4 Luciferase reporter gene assays

2.4.1 Overview

Reporter gene assays are very valuable tools to assess the regulatory impact of a DNA region. To investigate a suspected functional SNP, reporter gene assays are very useful as each allele could be tested under the exact same conditions and surrounded by the same vector and insert DNA so that any difference observed could only be explained by the base change caused by the SNP studied. Many reporter genes are available; for our work we used the firefly luciferase gene as an experimental reporter and *Renilla* luciferase as a control reporter to control for transfection efficiency as well as to differentiate between specific and non-specific cellular responses.

2.4.2 Experimental procedure

2.4.2.1 Cloning

DNA was amplified around the SNP of interest using primers designed to contain restriction enzyme (RE) sites specific to the following enzymes: *Sac*I for the forward primers and *Xho*I for the reverse primers.

SNP flanking sequences used for primer design were retrieved from dbSNP database build 137 and primer design was performed using the publically available online PCR primer design tool, Primer3 (367, 368).

I optimized each primer pair's annealing temperature in order to obtain a unique PCR product and no non-specific primer binding.

PCR was performed using the optimized annealing temperature for each primer pair. PCR products were verified using electrophoresis on a 2% agarose gel, and specific bands

were excised under 320 nm UV light. DNA was extracted from the gel bands using a commercially available gel extraction kit (Qiagen catalog# 28704), restriction-digested at 37°C for one hour and the DNA was purified using clean-up spin columns (Qiagen). DNA quantification at this step and subsequent steps was performed using a NanoDrop spectrophotometer.

DNA inserts for each allele were prepared by performing two PCR reactions in parallel using homozygote DNA samples, one PCR reaction using heterozygote DNA samples or by site-directed mutagenesis using one allele's prepared DNA insert.

In preparation for cloning, minimum promoter vectors were subjected to double restriction digest using the appropriate enzymes, de-phosphorylated using calf intestinal alkaline phosphatase (New England Biolabs) and purified by agarose electrophoresis.
The DNA insert and vectors were ligated at room temperature using T4 ligase, an appropriate DNA ligase for cohesive ends of DNA, then stored at -20°C until transformation.

2.4.2.1.1 TSLP

Primers to amplify a 656 bp DNA region around *TSLP* rs1837253 were as follows:

Forward primer: 5'- GTACGAGCTCAAGACAGTCCTCAGGCCAAA -3'

Reverse primer: 5'- GCATCTCGAGGTCTGCCTTGTGGCATCATA -3'

My SNP of interest is located in the 5' end of the *TSLP* gene therefore I used the pGL4.23[*luc2*/minP] vector for plasmid construction as it contains a 5' cloning site and a minimal promoter that drives luciferase expression.

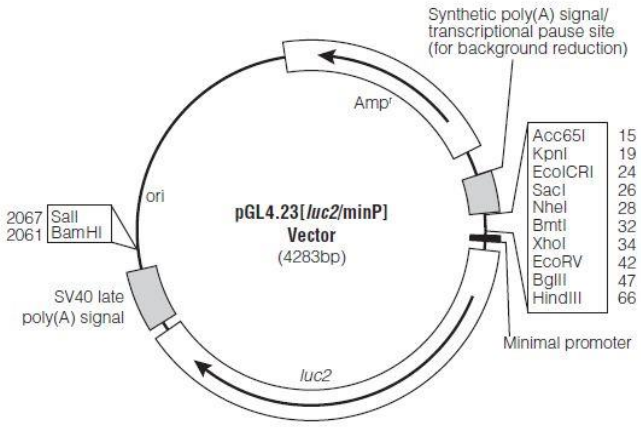

Figure 2-1 Map of vector pGL 4.23[*luc2*/minP]

2.4.2.1.2 IL1RL1

Primers to amplify an 809 bp region surrounding the *IL1RL1* SNP rs3771180 were as follows:

Forward primer: 5'- GTACGAGCTCAACCCCAAACCCTCAAACTT -3'

Reverse primer: 5'- GCATCTCGAGGCTGCTTACTTCGCATTTCC -3'

My SNP of interest is located in the 5' end of the *IL1RL1* gene; therefore the vector used for plasmid construction was pGL4.23[*luc2*/minP] (Figure 2-1).

2.4.2.1.3 Site-directed mutagenesis

In order to produce the C allele for rs1837253 C/T *TSLP* SNP, I performed site-directed mutagenesis using the Stratagene Quickchange site-directed mutagenesis kit (Agilent technologies Inc., Santa Clara, CA). Two complementary primers were designed using db SNP (build 137) containing the desired allele (C) in the middle and at least 15 nucleotides on each side as follows:

TSLP-C 5'- GATCCTTTTATACATAAACAACGTGTCTAAACTATGAAGC -3'

TSLP-Ccomp 5'- GCTTCATAGTTTAGACACGTTGTTTATGTATAAAAGGATC -3'

The primers were ordered to be made and PAGE-purified from Integrated DNA technologies (IDT, San Diego, CA). Using the T allele plasmid as template, I ran a primer extension reaction for 18 cycles using *Pfu* Turbo DNA polymerase. *Pfu* is a proof reading polymerase which does not displace the primers resulting in a nicked double-stranded circular DNA. After cycling, digestion with the endonuclease *Dpn*I removes the parental DNA template thus selecting for the newly synthesized DNA. Transformation into XL-1 Blue supercompetent *E. coli* cells repairs the nick in the new plasmid. Sequencing was used to confirm the identity of the new C allele plasmid.

2.4.2.2 Transformation

MAX Efficiency® DH5α™ Competent *E. coli* cells (Life Technologies) were used for transformation. pUC19 vector was used as a transformation control. Colonies were grown overnight in LB-Agar plates with Ampicillin in a 37°C incubator and counted the next morning. Individual colonies were picked and grown in culture tubes containing LB medium in a 37°C shaking incubator. DNA from the cultured bacteria from each colony was purified using Purelink™ Quick Plasmid Miniprep kit (Life Technologies) and plasmids were checked using restriction digest, with the two enzyme sites used for cloning and gel electrophoresis was used to check for the presence of DNA inserts of the expected size (Figure 2-2). Verification of insert-containing plasmids was performed by sequencing.

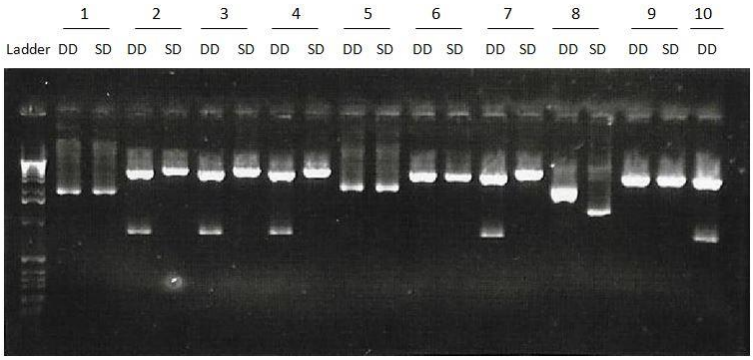

Figure 2-2 Example of a gel of restriction digest for selecting plasmid with insert

The first lane contains 1kb DNA ladder; DD: double restriction digest (both restriction enzymes used for cloning); SD: Single restriction digest (one of the restriction enzymes used for cloning); Numbers 1 to 10 indicate plasmid DNA isolated from bacterial colonies 1 to 10.
This gel shows DNA isolated from colonies 2, 3, 4, 7 and 10 contain an insert of the desired size. The sequence was verified by sequencing.

The correct plasmid-carrying bacterial colonies were grown and used to extract DNA using Qiagen plasmid maxiprep kit (Qiagen) following the manufacturer's protocol.

2.4.2.3 Transfection

A549 cells were grown in Dulbecco's Minimum Essential Medium (DMEM) in a 24-well cell culture plate until they reached about 70% confluence. 1 µg plasmid DNA for each SNP allele was mixed in DMEM without serum supplemented with Plus™ reagent. Co-transfection with *Renilla* plasmid DNA was performed as an internal transfection control at a 3 to 1 ratio (firefly to *Renilla*).

Lipofectamine® LTX reagent was added after a 15 minute incubation time and the mixtures were left at room temperature to allow the formation of DNA/lipid complexes for 30 minutes. Media were removed from the cells and replaced with media with serum and 100 µl of DNA/lipid complex mixture to each well. Plates were incubated at 5% CO_2 and 37°C for 24 hours for transfection. Treatments were added for the TSLP experiments 24-hour post-

transfection and the cells were incubated for a subsequent 24 hours. The treatments used were: Dexamethasone (Sigma-Aldrich) at six concentrations from 0.01 to 100 nM, Polyinosinic-polycytidylic acid (Poly(I:C)) at 10, 20 and 30 µg/ml and cigarette smoke extract (CSE) at 100, 50 and 5 fold dilutions.

2.4.2.3.1 Cigarette smoke extract

CSE was generated using 3R4F standardized cigarettes (University of Kentucky). Smoke was bubbled using a vacuum pump from three cigarettes to 5 ml phosphate buffered saline to produce a 100% CSE solution. The solution was filtered through a 0.2 µm filter and diluted 5, 50 and 100 fold in DMEM culture medium. Fresh CSE solution was made prior to each experiment.

2.4.2.4 Luciferase reporter gene assay

Dual-Glo® Luciferase assay system (Promega) reagents were used to measure Luciferase reporter gene activity for both firefly and *Renilla* vectors. Firefly and *Renilla* luciferases are distinct enzymes with different substrates, which allows for the usage of these two enzymes for successive measurement in the same sample. Firefly luciferase requires beetle luciferin, ATP, magnesium and molecular oxygen whilst *Renilla* luciferase requires coelenterate luciferin (coelenterazine) and molecular oxygen.

After the transfection incubation time, media were removed from each well and replaced with Dual-Glo® luciferase assay reagent, which is composed of the luciferase buffer and firefly luciferase substrate; after a 10 minute incubation necessary for complete cell lysis, the contents of each well was mixed by pipetting then transferred to 2 wells of a Corning

Costar® white 96-well plate suitable for luminescence measurement and the firefly luminescence was measured using a GENios multifunction microplate reader (Tecan). After the firefly luminescence measurement, an equal volume of Stop & Glo® reagent (composed of Dual-Glo® Stop & Glo® buffer and substrate) was added to each well. The *Renilla* luminescence was measured after a 10 minute incubation at room temperature.

The ratio of luminescence from the experimental reporter (firefly) to the control reporter (*Renilla*) was calculated after background-subtraction using control wells containing empty vector that were treated using the same procedure.

2.5 Electrophoretic mobility shift assays

2.5.1 Overview

In order to test the hypothesis that a SNP plays a functional role in gene regulation, an experimental approach is to study possible DNA/protein interactions using electrophoretic mobility shift assays (EMSA). Functional SNPs can participate in gene regulation by binding regulatory proteins such as transcription factors.

The principle of EMSA is the usage of non-denaturing polyacrylamide gels to separate protein-bound DNA from unbound DNA. The DNA used is labeled double-stranded DNA probes of 20 to 40 nucleotides in length, designed around the SNP of interest. The potential binding proteins exist in a crude nuclear extract isolated from appropriate cells.

2.5.2 Experimental procedure

The adenocarcinoma human alveolar basal epithelial cell line A549 was selected for nuclear proteins extractions for the purpose of EMSAs for both *TSLP* and *IL1RL1*. A549 cells were

purchased from ATCC, and cultured in DMEM in 6-well culture plates. Nuclear extracts were obtained using nuclear extraction kits for use with transcription factor assays (Panomics, catalog# AY2002) according to the manufacturer's protocol and were quantified using optical densitometry at 280 nm on a NanoDrop 8000 spectrophotometer.

Pairs of oligonucleotides for each SNP allele were designed based on the SNP flanking sequence in the NCBI SNP database (http://www.ncbi.nlm.nih.gov) to be approximately 25 bases around the SNP site and ordered to be synthesized and purified using high-performance liquid chromatography from Integrated DNA Technologies. Table 2-3 outlines the sequences of the oligonucleotides used.

Gene	SNP	Allele	Oligonucleotide	Complementary oligonucleotide
TSLP	rs1837253	T	5'- TAT ACA TAA ACA ATG TGT CTA AAC T - 3'	5' - AGT TTA GAC ACA TTG TTT ATG TAT A - 3'
		C	5' - TAT ACA TAA ACA ACG TGT CTA AAC T - 3'	5' - AGT TTA GAC ACG TTG TTT ATG TAT A - 3'
IL1RL1	rs3771180	A	5' - AAA TGC TGG TGA ATC ATG TAC TA- 3'	5' - TAG TAC ATG ATT CAC CAG CAT TT- 3'
		C	5' - AAA TGC TGG TGC ATC ATG TAC TA- 3'	5' - TAG TAC ATG ATG CAC CAG CAT TT- 3'

Table 2-3 EMSA oligonucleotides sequences

Annealed complementary oligonucleotides for each SNP allele were 5' end-radio-labeled with $^{[\gamma\text{-}32]}$P-ATP (Perkin Elmer) and the percentage of incorporation was measured using the ratio of labeled oligonucleotide to unlabeled oligonucleotide after counting in a scintillation counter (Beckman Coulter LS6500 Multi-Purpose Scintillation counter). An incorporation of greater than 30% was necessary for the use of the radio-labeled oligonucleotides.

A commercially available kit from Promega, Gel shift assay system (Catalog# E3050) was used for the EMSA experiments. The double-stranded oligonucleotides were incubated with A549 nuclear extracts in the presence of DNA binding buffer on ice for one hour after a

pre-incubation of nuclear extract and binding buffer for 10 minutes at room temperature; the cold environment is beneficial for preventing non-specific binding. Negative and positive controls were included in the form of Sp1 oligonucleotide in the absence and presence of HeLa cell nuclear extract, respectively.

In order to predict proteins that would differentially bind alleles, I used the following publically available algorithms which use pre-defined transcription factor binding sites (TFBS) (such as in the TRANSFAC database (369)) to construct binding site weight matrices for TFBS prediction: PROMO (370), JASPAR (371), TFsitescan (372), Consite (373) and TFSearch (374).

For competition reactions, the unlabeled oligonucleotides were incubated with the nuclear extracts and binding buffer for one hour on ice before adding the radio-labeled oligonucleotides. The reactions were run on 2-8% gradient CosmoPAGE native gels (Nacalai USA, Inc) for about 30 to 45 minutes until the loading dye reached ¾ down the gel. The gels were dried for 45 minutes to one hour at 57ºC on a slab gel dryer (Drygel Sr. Model SE1160, Hoefer Scientific Instruments). The gels were subsequently exposed to a Super resolution storage phosphor screen (Perkin Elmer Life and Analytical Sciences) overnight and visualized on a Cyclone® Plus storage phosphor system from Perkin Elmer using OptiQuant Image Analysis software (Perkin Elmer).

2.6 Enzyme-linked immunosorbent assays (ELISAs)

2.6.1 Overview

ELISA is a valuable tool utilized to quantify protein levels from biological fluids and generally involves 3 essential elements:

- A solid phase (immunosorbent): Usually a 96-well plate, to which the antigens to measure or antibodies will be immobilized.
- A conjugate: Either the antigen (protein to be measured) or an enzyme-labeled antibody.
- A substrate: the substrate is used for detection of the bound enzyme and consists of hydrogen peroxide and a chromogen.

Three main types of ELISA can be performed: "sandwich" ELISA, antigen-capture (direct) ELISA and competitive (blocking) ELISA. I performed "sandwich" ELISA; Antigens of interest are coated on the solid phase through a coated specific capture antibody; primary detection antibodies are then added and bind the antigen. Washing the wells of the plate ensures unbound antigens and antibodies are eliminated and bound antigen can be quantified by the substrate of an enzyme conjugated to a secondary antibody that binds the primary.

ELISA assays were performed to measure soluble IL1RL1 (hereafter referred to as sST2) levels in serum samples.

2.6.2 ST2 ELISA

I used reagents from R&D Systems (Catalog# DY523) to build a sandwich ELISA system to detect sST2 in serum samples of CAPPS samples. According to the manufacturer, the capture antibody is specific to natural and recombinant human ST2 as well as free ST2 and IL-33-complexed ST2.

96-well plates were coated overnight with 500 ng/ml capture antibody; the wells were washed using an automatic plate washer, blocked with bovine serum albumin then washed.

Consequently, undiluted serum samples were applied to the wells; serial dilutions of recombinant ST2 were used for a standard in each plate and samples were plated in triplicate for the first plate then in duplicate for the remainder of the plates. After the 2-hour incubation of samples and standard protein, biotinylated detection antibody was applied, the wells washed then detection was achieved with the addition of streptavidin-horseradish peroxidase, followed by 3,3′,5,5′-Tetramethylbenzidine (TMB) substrate; the signal was stopped by sulfuric acid then read at 450 nm and 570 nm reference.

2.7 PCR-based splicing assay

In order to assess whether the *IL1RL1* intronic SNP rs1420101 had an effect on splicing of *IL1RL1* mRNA, four primers were designed to form three primer pairs to amplify different length fragments of cDNA. The primers were: a reverse primer targeting exon 6, a reverse primer targeting exon 7, a reverse primer targeting exon 8 and finally a forward primer targeting exon 5 to use for all the reverse primers. The primer pairs composed of the forward primer and each of the reverse primers will be hereafter referred to as R6, R7 and R8 respectively. Figure 2-3 shows the primers' alignment and the expected sizes of PCR products using the primer pairs.

The cDNA samples were cycled in a ThermoHybaid thermal cycler for 35 cycles (denaturation step: 92°C for 30 seconds, an annealing step: 61°C for 1 minute then elongation at 72°C for 1 minute) followed by additional elongation for 10 minutes at 72°C.

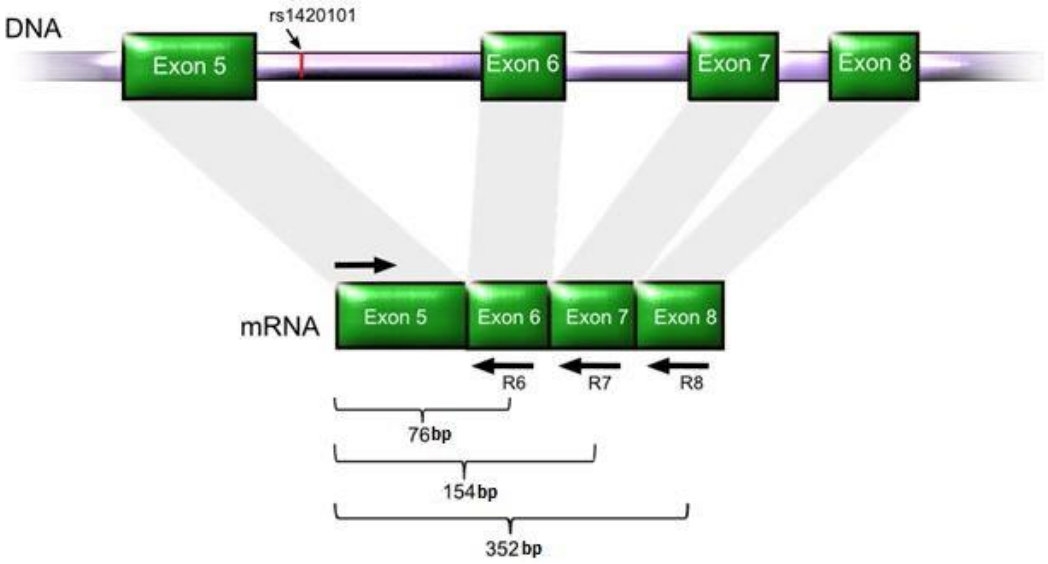

Figure 2-3 rs1420101 splicing assay primer alignment

2.8 Lung eQTL study dataset

A previously generated lung expression eQTL dataset (50, 375, 376) was used to investigate the SNPs of interest. The data were generated from non-tumor lung tissue samples collected from 1,424 individuals undergoing lung resectional surgery at three academic sites: Laval University, Québec, Canada (500 patients); the University of British Columbia, Canada (437 subjects); the University Medical Center Groningen, The Netherlands (487 subjects). The Québec and Vancouver subjects gave written consent and the studies were approved by the appropriate ethics boards. The Groningen study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines ("Code of conduct; Dutch Federation of Biomedical Scientific Societies"; http://www.federa.org).

1,237 DNA samples were successfully genotyped using the Illumina Human1M-Duo BeadChip array. For gene expression, total RNA was hybridized to a custom Affymetrix

HU133 microarray, comprised of 751 control probe-sets and 51,627 non-control probe-sets resulting in 1,329 successfully assayed samples. Quality controls on genotyping and gene expression data as well as expression traits processing have been described previously (50, 375, 376).

2.9 Data analysis

2.9.1 TaqMan® genotyping data

TaqMan® genotyping results were verified for Hardy-Weinberg equilibrium (HWE) as ascertained by the equations below for each genotyping dataset.

$$p^2 + 2pq + q^2 = 1$$
$$p + q = 1$$

with p and q being the allele frequencies. Expected and observed values were compared using Chi-square tests with one degree of freedom in Excel.

Association of SNP genotypes (obtained with TaqMan® genotyping) with phenotypes such as asthma adjusting for gender and sex were performed in JMP software using Chi-square tests. Statistical power calculations were performed using the Quanto software (377).

2.9.2 Gene expression data

2.9.2.1 qPCR data

Differences in *TSLP* and *IL1RL1* expression levels, generated by qPCR, between genotypic groups and phenotypic groups, were assessed by one-way ANOVA tests in Excel statistics package.

To test for allele-specific expression, the cDNA ratios of FAM™ over VIC® probes (corresponding to C to T alleles) were compared with the ratio of the gDNA using a t-test in Excel initially, then replicated in R studio using the Stats R package.

2.9.2.2 Lung eQTL data

eQTL analysis was run in the lung expression dataset using linear models in R for each of the *TSLP* and *IL1RL1* probes in order to investigate the association of gene expression with genotypes and assess the portion of variability of gene expression explained by the SNPs. In an effort to uncover additional explanation for the variability of the gene expression, I investigated the potential relationship of gene expression with clinical and demographic variables collected for the lung eQTL study as well as interactions between the variables and SNPs on gene expression levels. For that purpose, multiple linear regression models (one model for each *TSLP* and *IL1RL1* isoform) were used in RStudio using the R packages 'Stats', 'FactoMiner', 'glmnet' and 'cvTools' for exploratory data analysis and 'Lattice', 'BiplotGUI', 'ggplots' and 'gridExtra' for graphical representations.

For this analysis, I utilized 45 of the 97 variables available from the lung expression dataset. Variable selection for inclusion in the multiple linear models was performed using the following approach:

1. Irrelevant variables were removed, such as dates, unknown and unspecified medication intake; this resulted in 89 variables.

2. Categorical variables with one factor present in less than 10 subjects were removed; for example only 5 subjects were taking leukotriene antagonist medication therefore this variable was removed. This resulted in 85 variables.

3. Missing values were ascertained for each variable and the variables with complete data were plotted against the number of subjects available. The plot is shown below (Figure 2-4) and indicates the number of subjects available for each number of variables without missing data. The blue points correspond to the variables and the red line (at 45 variables) corresponds to the cutoff I utilized in order to maximize the number of subjects as well as the number of variables to include in the statistical models. The linear models do not allow for missing data.

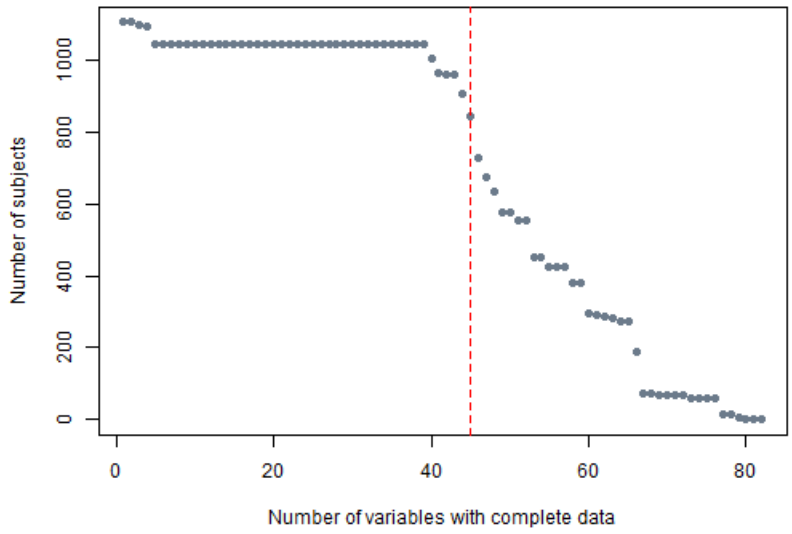

Figure 2-4 Variable selection using number of missing data for each variable

This resulted in 845 subjects with complete data for 45 variables; the characteristics of these subjects for all variables are outlined in appendix B.1 (page 238). I ran multiple linear models for the *TSLP* and *IL1RL1* isoforms gene expression using the following probes: X100122252_TGI_at for the long *TSLP* isoform, X100142361_TGI_at for the short *TSLP*

isoform, X100312840_TGI_at for the soluble *IL1RL1* isoform and X100302151_TGI_at for the receptor *IL1RL1* isoform.

Subsequently, I used the 'Step' function in R which implements stepwise model selection based on the Akaike Information Criterion (AIC) (378). The AIC is a method of model selection among several possible models; the selected model maximizes the fit with the least number of variables.

2.9.2.2.1 TSLP analysis

For analysis of the *TSLP* gene expression, the resulting 45 variables were included in the two multiple linear models along with interaction terms of the rs1837253 SNP with all the 45 variables. There were 279 and 123 possible models for the long and short isoforms expression respectively. Applying the 'step' function to the result of the two models resulted in the best fitted model, which results are presented in Chapter 5.

2.9.2.2.2 IL1RL1 analysis

There were 21 SNPs identified as important predictors of *IL1RL1* expression and potential functional SNPs (details explained in Chapter 3): 10 SNPs from the manual annotation (Table 3-4) and the 11 SNPs from the Glmnet approach (Table 3-5). Because 4 of the SNPs were selected in the manual annotation approach based on their high LD with eQTL SNPs, genotypes were unavailable for them in the lung dataset. Subsequently, using 1000 Genomes data for LD determination, I removed SNPs in high LD ($r^2 > 0.8$) with other SNPs in the list, as the statistical modeling I used did not allow for highly correlated variables. Finally, I had a list of 9 independent SNPs for the interaction analysis: rs10496354, rs10496357, rs4850994,

rs7607548, rs1558622, rs3771180, rs1420101, rs4988956 and rs1420106. The inclusion of all 45 variables, 9 SNPs and interaction terms of each variable with each SNP would result in 2372 possible models; as this number is higher than the number of subjects (845), it is not possible to apply a linear model.

Therefore, I proceeded to shrink the number of possible models by first testing separate multiple linear models without interaction terms (including the 9 SNPs and 45 clinical and demographic variables); applying the 'step' function, I obtained the best fitted model. I then used this model and added interaction terms between the variables and the SNPs. There were 516 and 67 possible models for the receptor and soluble isoforms expression, respectively. Applying the 'step' function to the result of this model resulted in the final results presented in Chapter 4.

Graphs depicting main effects of associations between gene expression and clinical variables were produced using the function 'termplot' from the 'stats' R package. This function allows for accurate plotting of regression terms against their predictors i.e. it produces plots of association of variables and the response (i.e. gene expression) as adjusted for the effects of all the other variables in the same model.

Graphs depicting interactions of SNPs with clinical variables are based on pairwise relationships and are not adjusted for the rest of the variables in the model, because the function 'termplot' does not allow for interaction plotting. Nevertheless, these graphs are useful in showing the general trend of the association although they need to be considered together with the coefficient estimate and p value.

2.9.3 Soluble IL1RL1 protein data

IL1RL1 protein data were handled and analyzed for association with SNPs and phenotypes using linear regression models in R Studio using the R packages 'Stats', 'FactoMiner', 'Lattice', 'BiplotGUI', 'ggplots' and 'gridExtra'.

2.9.4 Reporter gene data

Reporter gene data were analyzed in R studio using the t-test function in the 'Stats' R package and plotted using the 'Lattice' R package.

2.10 Methodology for IL1RL1 computational analysis (chapter 3)

2.10.1 Lung eQTL study data

From the aforementioned lung eQTL study dataset, 1,111 subjects had complete genotyping, gene expression and clinical data and were hence used further for analysis. eQTL determination was performed in RStudio by using a linear regression model on all transcript-SNP pairs within the chromosome 2 region encompassing *IL1RL1* (Chr2: 102239363-103583057 build 37.3).

2.10.2 Manual SNP prioritization workflow

The following steps were used to prioritize SNPs as causal for the asthma associations:

1. All SNPs were selected that showed association with expression of any gene in the lung in the region spanning 1 Mb around the *IL1RL1* gene.
2. eQTLs in the region associated with the expression of genes other than *IL1RL1* and/or *IL18R1* were removed. Only SNPs associated with expression of both *IL1RL1* and

IL18R1 using a p value cut-off of less than 1×10^{-6} were retained (all eQTLs for *IL18R1* were also eQTLs for *IL1RL1*). The rationale for this step was that none of the aforementioned genes were associated with asthma and related phenotypes in GWAS data. In contrast, both *IL1RL1* and *IL18R1* have been identified in numerous studies and one of the goals of our search for putative causal SNPs was to determine which of these two genes is responsible for the association signal.

3. The SNPs from the eQTL dataset as well as asthma-associated SNPs in the same chromosome region from the literature were organized into LD bins with a cut-off of 0.7 for r^2.
4. *In silico* analysis was performed on all SNPs to determine functional potential.
5. The relationship between the associated SNP effect (protection/risk) and the direction of expression effect (which allele was responsible for greater or lesser expression levels) was investigated for all eQTLs.
6. Final selection of putative causal polymorphism was performed by manual annotation using the ENCODE data (379).

2.10.3 Lasso and elastic-net regularized generalized linear models (Glmnet)

The "glmnet" R library was used to select SNPs that best predict expression levels of *IL1RL1* probe-sets. This method fits a generalized linear model to the data; which is suited for high dimensional data (N samples << N predictors) and allows for correlated predictors. The elastic net penalty allows for the selection of k predictors from a total of p predictors (k<<p). The number of predictors selected is defined by a tuning parameter (α) defined by the user.

In addition, I incorporated a cross-validation step in order to replicate the resulting SNPs in randomly split folds of the dataset; the idea being that true predictor SNPs will stay predictors in any subset of the dataset. For that purpose, I used the R package "cvTools" which is a package composed of tools for cross-validation in regression models and is used to assist with model selection. The data were randomly split into 5 folds, one test set and four training sets (5-fold cross-validation). This procedure was repeated ten times (10x5 fold cross-validation); for each iteration, I obtained a list of SNPs whose coefficients were non-zeros across all 5 folds, called active coefficients, i.e. SNPs that explain the variability of the expression data.

I counted the number of times each resulting SNP had an active coefficient in each of the 10 iterations, and created a novel scoring system by adding the number of iterations for which each SNP was active; I utilized this score as a reflection of the degree of replication and thus validity of the SNP as an active SNP. The maximum score for a SNP for any given probe-set is 10 points (1 point for each iteration).

I considered five scores: a score out of a maximum of 60 for all six probe-sets (5 *IL1RL1* and 1 *IL18R1* probe-sets), a score out of a maximum of 50 for all five *IL1RL1* probe-sets, a score out of a maximum of 40 for the four soluble *IL1RL1* probe-sets, a score out of a maximum of 10 for the one receptor *IL1RL1* probe-set and a score out of a maximum of 10 for the one *IL18R1* probe-set. ENCODE publically available data were utilized to annotate the active SNPs.

In addition to my novel scoring system, and in order to test the efficiency of the resulting SNPs in predicting the corresponding probe's expression, I calculated the normalized root-mean square error (NRMSE), by calculating RMSE then dividing it by the

data range to obtain the normalized RMSE. NRMSE is more informative than RMSE because it is expressed as a percentage as opposed to the scale of the expression data as is the case for RMSE. The formula used is given below:

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{Obs}_i - \mathrm{Pred}_i)^2}}{Obs_{\max} - Obs_{\min}}$$

The NRMSE indicates in percentage the difference between the predicted and the observed value, in this case of gene expression; and is hence a valuable way to judge whether the SNP selection is valid, does indeed predict the gene expression and determines how well the prediction model fits the real expression data.

2.10.4 Web resources

LD information was obtained from publically available data in the following websites: HapMap (http://hapmap.ncbi.nlm.nih.gov/), SeattleSNPs Genome variation Server 138 (http://pga.gs.washington.edu) and 1000 Genomes (http://browser.1000genomes.org/).

The LD data were downloaded and probed for our SNPs of interest in Excel worksheets. eQTL data were manipulated and analyzed using the RStudio open source software version 0.96.331 (http://cran.r-project.org/).

Chapter 3: *IL1RL1* SNP prioritization using Generalized Linear Model regression with Elastic net and *in silico* manual analysis

3.1 Overview

IL1RL1 (a.k.a. ST2) and IL-33 play an important role in the eosinophilic inflammation characteristic of asthma. The *IL1RL1* gene is located on chromosome 2 in the midst of a cytokine gene cluster with *IL1R1*, *IL1RL2*, *IL18R1* and *IL18RAP*: all encoding proteins involved in inflammation and immunity. This gene cluster showed significant association in candidate-gene studies as well as multiple GWAS for asthma, childhood asthma, severe asthma, and eosinophil count. The region is in relatively high LD, making it challenging to narrow down the asthma association signal to one or more causal SNPs.

My aim was to identify functional SNPs that may be causal for the asthma associations and thus elucidate the regulation of the *IL1RL1* gene. I hypothesized that SNPs associated with risk of asthma would be associated with a decrease in expression of the soluble isoform of *IL1RL1* and/or an increase in expression of the *IL1RL1* receptor isoform in the lung.

Using a lung eQTL dataset in combination with published genome-wide association data, I was able to manually narrow down the search for putative causal SNPs from 644 SNPs to 10 excellent functional candidates and determined that the SNPs associated with expression of *IL1RL1* in the lung are mainly regulating the short soluble isoform.

Using a generalized linear model regression with elastic net penalty, I uncovered a list of SNPs in and around *IL1RL1* that best predict soluble *IL1RL1* expression as well as four

SNPs, about 260 kb from the *IL1RL1* gene that best predict the expression of the receptor *IL1RL1*.

Hence, I provide a strong proof-of-concept for the usage of generalized linear model regression with elastic net penalty to select predictors of gene expression as well as propose the presence of a distal regulatory region controlling the expression of the membrane-bound isoform of IL1RL1 but not the soluble isoform.

3.2 Background

GWAS and most candidate gene association studies utilize the LD structure of the genome (380) by interrogating TagSNPs (381) in order to reduce the number of redundant SNPs assayed while still conserving as much information as possible about the genetic variation in a region. When a significant association is observed with a phenotype, the TagSNP or any one of the SNPs in the same LD bin may be causing the association signal. Thus, knowledge of the associated SNPs alone does not necessarily give information about causality or the mechanism underlying the association of the particular gene or locus with the phenotype studied.

The *IL1RL1* locus is located on chromosome 2q12 in a cytokine gene cluster with *IL18R1, IL18RAP, IL1R1* and *IL1RL2;* this region is characterized by long stretches of high LD (r^2 greater than 0.8), which also encompass neighboring genes such as *SLC9A4* and *MAP4K4. IL1RL1* exists mainly in two isoforms: a ubiquitously expressed short soluble isoform and a long receptor isoform (382) expressed exclusively on the surface of hematopoietic cells, although a third isoform has also been described (192). The receptor

isoform binds IL-33 resulting in a pro-inflammatory Th2 type immune response. The soluble isoform also binds IL-33 but acts as a decoy receptor, preventing the inflammatory signaling. SNPs in the *IL1RL1* region have been associated with asthma in candidate gene studies and GWAS (101). rs1420101 in intron 5 of *IL1RL1* was significantly associated with eosinophil count in a GWAS (94). Subsequently, an investigation of 237 candidate genes in asthmatic children showed that rs13431828 and rs1041973 in the second and third exons of the receptor isoform of *IL1RL1* were among the most strongly associated SNPs; the authors also implicated *IL18R1* after multimarker analysis (114). The GABRIEL study GWAS reported that rs3771166 in *IL18R1* was associated with asthma in a meta-analysis of all the study populations (35). Furthermore, 11 SNPs in the region including *IL1RL1*, *IL18R1* and *IL18RAP* were associated with asthma at a genome-wide significance level in at least one population (35). In addition, the locus was significantly associated with asthma in a meta-analysis of GWAS in North American populations (37) and rs3213733 in *IL18R1* was associated with asthma in a GWAS in a Japanese population (283). In a recent GWAS, multiple SNPs (either directly genotyped or imputed using 1000 Genomes data) were associated with severe asthma: rs3771166, rs9807989 in *IL18R1* and rs13035227 in *IL1RL1* (38).

Thus, chromosome 2q12 constitutes one of the most robustly replicated asthma loci but the causal gene(s) and SNP(s) involved remain to be determined. Investigating the regulation of the *IL1RL1* gene is important because of the diverse biological roles of IL1RL1 and IL-33 (100, 101). In tissues affected by asthmatic inflammation, the receptor isoform of IL1RL1 is found in a dimer with IL1 receptor accessory protein on the surface of various cell types (Th2 cells, mast cells, macrophages and dendritic cells) and binds IL-33 to give rise to a MYD88-

mediated inflammatory signaling cascade resulting in NF-κB activation and subsequent pro-inflammatory cytokine and chemokine gene expression (180). The soluble isoform of IL1RL1 has been shown to inhibit this inflammatory cascade, by binding to IL-33 thus acting as a decoy receptor (210). In mice models, blockade of the IL1RL1/IL-33 axis, e.g. by administering soluble IL1RL1 or antibodies against IL1RL1, showed beneficial inhibition of inflammation (211, 217).

eQTL are loci in the genome that regulate the mRNA expression of a gene. The change in mRNA level of a gene is an important aspect of regulation and most SNPs associated with complex traits are located in non-coding regions (164) suggesting that they result in an alteration of expression rather than function of proteins involved in the complex trait. SNPs which are associated with complex genetic disease are significantly enriched for eQTLs (15). In this study, I used two complementary approaches to assess which polymorphisms in the region encompassing the IL1 cytokine gene cluster are most likely to be functional and therefore causal for the asthma associations. Firstly, I used a lung eQTL dataset (50, 375, 376) combined with published asthma genetic association studies to manually curate the evidence for the functional importance of SNPs in the region. This analysis was focused on *IL1RL1* and *IL18R1*, which are the genes most consistently associated with asthma and associated phenotypes. Secondly, I used an unbiased algorithm to determine the best predictors of the soluble and/or receptor isoforms of *IL1RL1* using the same lung eQTL dataset. Lung eQTLs are of interest and relevance to asthma because the genes in my region of interest are expressed in the airway epithelium. My *in silico* analyses also generated important insights into the regulation of the *IL1RL1* gene. SNPs that are eQTLs in the lung

are excellent candidates for causing the associations with asthma and thus interesting targets for further functional characterization; such as the experiments described in Chapter 4.

3.3 Results

3.3.1 Manual SNP prioritization

3.3.1.1 Initial SNP selection in the region encompassing IL1RL1 gene

Data points from the entire chromosome 2 eQTL peak using 10% false discovery rate were extracted from the eQTL dataset. A probe-set is a group of 25-mer probes which hybridize specifically to one target transcript; a gene may have one or several probe-sets targeting it in order to cover known and possible splice variants.

A data point is comprised of a SNP - expression probe-set pair and represents the association between the two as quantified by a p value, that is, whether the expression measured by the probe-set is different when stratified by SNP genotypes.

There were no significant trans-eQTLs either in the *IL1RL1* region or targeting it, i.e., there were no SNPs in the region that targeted transcripts in other regions (> 1 Mb away) or SNPs in other regions that targeted transcripts in the *IL1RL1* locus. Therefore, the analysis described in this study is concerned only with cis-eQTLs.

A visual inspection of a HapMap LD heat map of the region showed extensive LD ($r^2 \sim 0.8$) over about 1 Mb (Figure 3-1) therefore I focused on a 1.3 Mb chromosomal region to account for this LD and included all probe-sets for the genes present in the region. This region spans eleven annotated genes (Figure 1-1). The numbers of probe-sets for each gene on the microarray used in this study were: *MAPK4* – three probe-sets, *IL1R2* - one probe-set, *IL1RL2* - one probe-set, *IL1R1* - two probe-sets, *IL1RL1* - five probe-sets, *IL18R1* - one

probe-set, *IL18RAP* - one probe-set, *SLC9A4* - one probe-set, *SLC9A2* - two probe-sets, *MFSD9* - three probe-sets and *TMEM182* - three probe-sets.

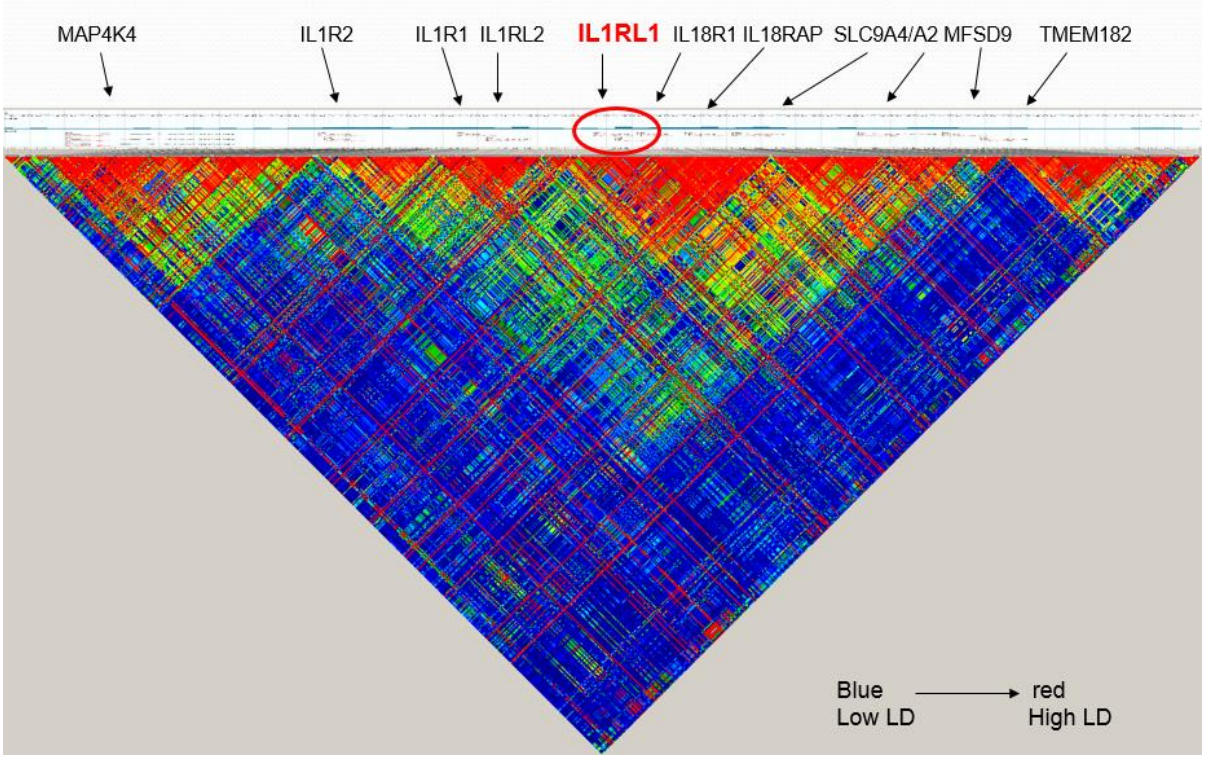

Figure 3-1 Linkage Disequilibrium map of the chromosome 2 region encompassing *IL1RL1* and neighboring genes

There were 641 cis-eQTLs in the region of interest. Table **3-1** contains details of the number of SNPs associated with different numbers of probe-sets and genes.

Number of probes	1	2	3	4	5	6	7
Number of SNPs	314	4	15	36	94	177	1
Percentage of SNPs	49.0%	0.6%	2.3%	5.6%	14.7%	27.6%	0.2%

Number of genes	1	2	3
Number of SNPs	366	234	41
Percentage of SNPs	57.1%	36.5%	6.4%

Table 3-1 Distribution of number of SNPs associated with probesets and genes from the list of 641 SNPs

3.3.1.2 Selection of eQTLs specific to expression of IL1RL1 and IL18R1

To narrow down the region of interest, 217 SNPs were removed which were eQTLs for *MAPK4*, *IL1R2*, *IL1RL2*, *MFSD9* or *TMEM182*. Interestingly, 83 of the removed SNPs were associated with expression of *SLC9A2* and were located ~500 kb away in *IL1R2, ILR1* and *IL1RL2*. This step resulted in a list of 424 SNPs.

Applying an arbitrary p value cutoff ($p<1\times10^{-6}$) for SNP association with expression, 1460 SNP-probe-set pairs remained in the analysis; they were comprised of 365 unique SNPs and 8 expression probe-sets corresponding to 4 genes. 304 of the eQTLs (82%) targeted the expression of *IL1RL1*, 186 SNPs (51%) targeted *IL18R1*, 63 (17%) targeted *SLC9A2* and 8 (2%) targeted *MFSD9*. At this statistical threshold, no eQTL was associated with the expression of *IL18RAP*. Only SNPs targeting the expression of *IL1RL1* and *IL18R1* were retained for further consideration to be functional candidates.

All of the resulting 304 SNPs were associated with the expression of *IL1RL1*. 91% of them associated with 4, 5 or 6 expression probe-sets (6 is the maximum number of probe-sets for *IL1RL1* and *IL18R1*); 186 of them were also eQTLs for *IL18R1*. All of the SNPs associated with *IL18R1* were also associated with one or more probe-sets for *IL1RL1;* i.e. there were no *IL18R1*-specific eQTLs but there were 118 *IL1RL1*-specific associations. The 186 SNPs that were common eQTLs for *IL1RL1* and *IL18R1* were examined further as potential causal SNPs. The list of these SNPs is presented in Appendix 6.3A.1.

3.3.1.3 Linkage disequilibrium analysis

An LD analysis of the 186 SNPs using CEU HapMap LD data resulted in 9 LD bins with an r^2 cutoff of 0.7. After a survey of the genetic association literature (35, 94, 99, 101, 114, 190,

228, 383, 384) for the *IL1RL1* region, I identified a list of 31 SNPs associated with asthma belonging to 8 separate LD bins. The two groups of LD bins (derived from the eQTL SNPs and from the published associations) overlap, sharing 7 LD bins; thus the total number of LD bins of interest increases to 12 bins for a total number of 205 SNPs. Combining the LD bins, rather than considering only the LD bins in common, permitted consideration of SNPs involved in gene regulation that do not involve expression changes. SNPs in high LD ($r^2 > 0.8$) with variants from the literature were added to the list bringing the total number to 232 SNPs in 12 LD bins.

207 SNPs of the 232 (89%) belong to 6 HapMap LD bins, three remaining LD bins comprise 9, 6 and 6 SNPs respectively and 3 SNPs are singletons. The LD analysis was repeated for the 232 SNP list using 1000 Genomes Project data and this resulted in the 232 SNPs belonging to 24 different LD bins.

Of those 24 bins, 4 LD bins contained 56% of the SNPs and 7 bins were composed of singleton SNPs. This was expected as more subjects were used to calculate LD in the 1000 Genomes Project; this however did not affect the conclusions of the study. The list of the 232 SNPs is shown in Appendix 6.3A.2.

3.3.1.4 Assessment of the functional potential of the SNPs

The list of 232 SNPs was subjected to functional impact analysis using the web-based NIH funcpred tool (385). The funcpred tool analyzes potential functionality of SNPs based on 1) predictions of the SNPs disrupting various regulatory sites such as splicing, transcription factor binding sites, enhancer sites, silencer sites or miRNA binding sites, 2) conservation as measured by the UCSC vertebrate multiz alignment and conservation score using 17 species

(386) 3) Polyphen program prediction of the effect of non-synonymous SNPs (387) and finally 4) a UCSC bioinformatics metric called the Regulatory Potential score (386, 388) generated by the ESPERR (evolutionary and sequence pattern extraction through reduced representations) algorithm; the Regulatory Potential score relies on alignment between 7 species to classify genomic regions as more closely related to neutral DNA or to regulatory DNA; a SNP not being assigned a score is due to a lack of alignment in at least 3 non-human species. As a result, the 232 SNPs were shortlisted to a total of 77 SNPs belonging to 12 HapMap LD bins leaving only SNPs predicted to have at least one of the functional potentials investigated. The complete list of the 77 SNPs is in Appendix 6.3A.3.

3.3.1.5 Association and eQTL direction relationship

20 SNPs associated with asthma were identified in 8 published studies (35, 37, 94, 99, 101, 114, 190, 226, 228, 231, 233, 283, 383, 384) that were also associated with *IL1RL1* expression in the lung eQTL study. The direction of association with asthma (risk or protection) was compared with the eQTL effect direction (increase or decrease of expression). The *IL1RL1* probe-sets used in the lung eQTL dataset enabled me to differentiate between the two main isoforms and thus to test the hypothesis that a risk allele for asthma is associated with a reduction in the soluble IL1RL1 isoform expression.

Fifteen of the SNPs (75%) compared showed concordance with this hypothesis (Table 3-2), that is, the eQTL data for shortlisted SNPs consistently showed that the protective eQTL alleles resulted in increased expression levels of the *IL1RL1* soluble isoform.

Hapmap LD bin	Encode LD bin	Gene	Isoform1 Location	Isoform2 Location	Chromosomal location (Hg19)	SNP ID	RefSNP alleles	MAF	Assoc. allele	Assoc. direction	assoc. allele eQTL effect on sST2	sST2 p value	assoc. allele eQTL effect on ST2L	ST2L p value	References
2	-	IL1RL1	Intron5	Intron5	102957716	rs1420101	T/C	T=0.37	T	risk	Decrease	< 2.2E-16	decrease	0.00063	Zhu *et al* Eur J Hum Genet 2008
-	-	IL1RL1	Intron10	0	102966067	rs1921622	A/G	A=0.411	A	risk	Decrease	4.02E-12	none	0.1192	Reijmerink *et al* JACI 2008
5	3	IL18R1	0	0	102979028	rs2287037	A/G (rev)	T=0.339	A	risk	Decrease	< 2.2E-16	decrease	0.000686	Zhu *et al* Eur J Hum Genet 2008
3	2	IL18R1	0	0	102985424	rs2058622	C/T	A=0.278	C	risk	Decrease	< 2.2E-16	decrease	0.000117	Zhu *et a* Eur J Hum Genet 2008
2	-	IL18R1	0	0	103001402	rs1035130	A/G	T=0.257	A	risk	Decrease	< 2.2E-16	decrease	0.000499	Reijmerink *et al* JACI 2008
3	2	IL18RAP	0	0	103035044	rs1420106	C/T (rev)	A=0.280	G	risk	Decrease	< 2.2E-16	decrease	0.000176	Reijmerink *et al* JACI 2008
-	-	IL1RL1	Intron1	0	102938389	rs1420089	A/G (rev)	C=0.102	C	protection	Increase	2.6E-05	none	0.85	Bossé *et al* Respir Res 2009
-	-	IL1RL1	Intron8	3' UTR	102961929	rs1946131	A/G (rev)	T=0.100	C	protection	Increase	0.002668	increase	0.025	Bossé *et al* Respir Res 2009
4	4	IL1RL1	Exon11	0	102968212	rs10204137	A/G	G=0.321	G	protection	Increase	0.001339	increasing trend	0.14	Moffatt *et al* NEJM 2010
4	4	IL1RL1	Exon11	0	102968356	rs10192157	C/T	T=0.403	T	protection	Increase	0.001355	increasing trend	0.1029	Moffatt *et al* NEJM 2010
4	4	IL1RL1	Exon11	0	102968362	rs10206753	C/T	C=0.403	C	protection	Increase	0.00246	increasing trend	0.1169	Moffatt *et al* NEJM 2010
4	4	IL18R1	0	0	102986222	rs3771166	C/T (rev)	A=0.382	A	protection	Increase	0.002366	increasing trend	0.1162	Moffatt *et al* NEJM 2010
4	5	IL18R1	0	0	103009537	rs4851004	C/T	T=0.464	T	protection	increasing trend	0.5852	decrease	0.03189	Moffatt *et al* NEJM 2010; Wu *et al* JACI 2010
4	5	IL18R1	0	0	103011237	rs2287033	A/G (rev)	C=0.464	C	protection	increasing trend	0.3994	decrease	0.01253	Moffatt *et al* NEJM 2010; Wu *et al* JACI 2010
4	5	IL18R1	0	0	103015687	rs1420094	A/G (rev)	T=0.465	T	protection	none	0.3794	Decrease	0.01436	Wu *et al* JACI 2010; Zhu *et al* Eur J Hum Genet 2008; Moffatt *et al* NEJM 2010

Table 3-2 eQTLs with direction of association concordant with the direction of *IL1RL1*-a gene expression

Isoform1: the receptor isoform; isoform2: soluble IL1RL1; MAF: minor allele frequency in CEU population; Assoc. allele: allele associated with asthma in the literature; Assoc.direction: Direction of the asthma association; sST2: soluble IL1RL1; ST2L: receptor isoform of IL1RL1. SNPs colored the same are in high LD with each other ($r^2 > 0.8$).

Figure 3-2 illustrates a representative pair of eQTL plots for one of the SNPs concordant with my hypothesis: rs1420089 which C allele leads to increased levels of soluble IL1RL1 (probe X100312840_TGI_at) but not membrane-bound IL1RL1 (probe X100302151_TGI_at); rs1420089-C allele was associated with asthma risk (231).

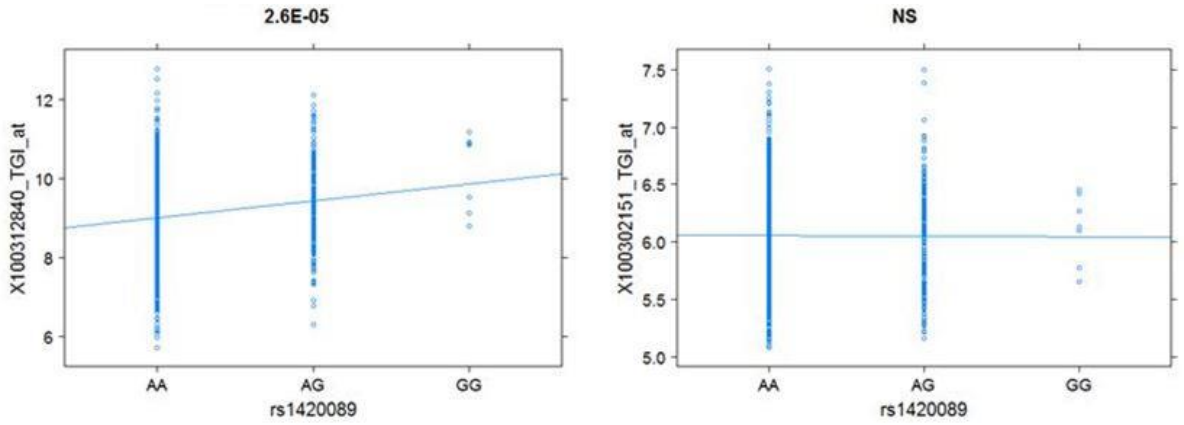

Figure 3-2 Lung eQTL plots for rs1420089

X100312840_TGI_at: probe for soluble IL1RL1; X100302151_TGI_at: probe for receptor IL1RL1; NS: Not statistically significant.

4 SNPs (20%) were associated with a decrease of the soluble isoform but were protective alleles (Table 3-3); which is not consistent with my hypothesis.

Hapmap LD bin	Encode LD bin	Gene	Isoform1 Location	Isoform2 Location	Chromosomal location (Hg19)	SNP ID	RefSNP alleles	MAF	Assoc. allele	Assoc. direction	assoc. allele eQTL effect on sST2	sST2 p value	assoc. allele eQTL effect on ST2L	ST2L p value	References
1	-	IL1RL2/IL1RL1	0	0	102887128	rs11674302	C/T	C=0.136	C	protection	Decrease	6.641E-06	none	0.1519	Moffatt *et al* NEJM 2010
1	-	IL1RL1	5' UTR	5' UTR	102954653	rs13431828	C/T	T=0.161	T	protection	Decrease	3.824E-06	none	0.2143	Moffatt *et al* NEJM 2010
1	-	IL1RL1	Exon3	Exon3	102955468	rs1041973	A/C	A=0.288	A	protection	Decrease	1.265E-11	Decreasing trend	0.002555	Wu *et al* JACI 2010
1	-	IL18R1	0	0	102997884	rs3213733	G/T (rev)	A=0.192	T	protection	Decrease	2.389E-06	decreasing trend	0.1396	Reijmerink *et al* Allergy 2010; Imada *et al* BMC Res Notes 2009; Wu *et al* JACI2010

Table 3-3 eQTLs with direction of association in disagreement with direction of *IL1RL1*-a gene expression

Isoform1: *IL1RL1*-b (the receptor isoform); isoform2: *IL1RL1*-a (soluble isoform); MAF: minor allele frequency in CEU population; Assoc. allele: allele associated with asthma in the literature; Assoc.direction: Direction of the asthma association; sST2: soluble *IL1RL1*; ST2L: receptor isoform of *IL1RL1*.

The 4 SNPs are rs11674302 (intergenic between *IL1RL2* and *IL1RL1*), rs13431828 (5′ end of *IL1RL1*), rs1041973 (non-synonymous SNP in *IL1RL1* exon 3) and rs3213733 (intronic SNP in *IL18R1*). These 4 SNPs are in low to moderate LD with each other according to 1000 Genomes LD data (r^2 ranging from 0.63 to 0.29).

Figure 3-3 is a group of eQTL plots for the 4 SNPs showing X100148162_TGI_at expression (one of the probes for the soluble isoform) by genotype.

For one of the eQTLs associated with asthma in the literature, rs3771180 (37), neither the direction of association nor the associated allele could be identified.

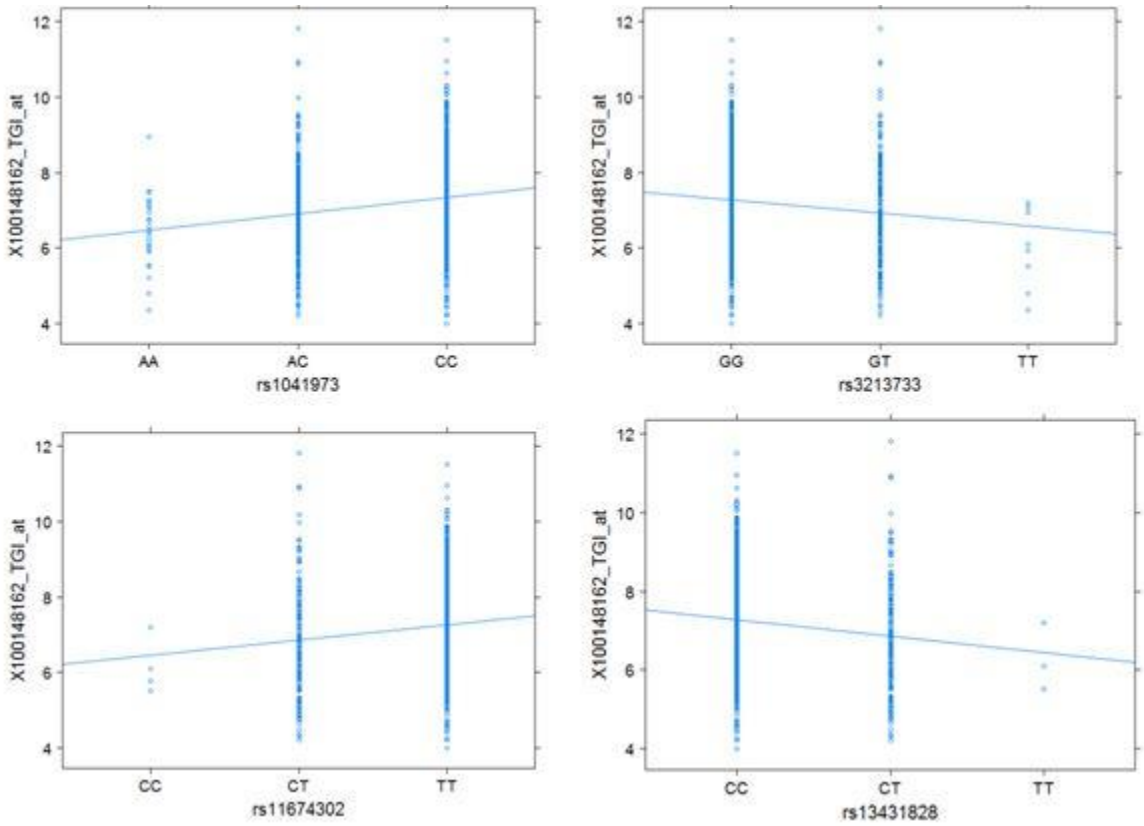

Figure 3-3 SNPs associated with lower of *IL1RL1-a* (soluble) gene expression and with asthma protection in the literature

X100148162_TGI_at: probe for soluble IL1RL1.

3.3.1.6 Final selection of putative causal SNPs by manual annotation

The SNP selection from each LD bin was performed manually by assessing the entire prediction information available, the SNP location, as well as the eQTL data whenever

available. The polymorphism location is important information to consider when judging functional impact (Figure 3-4).

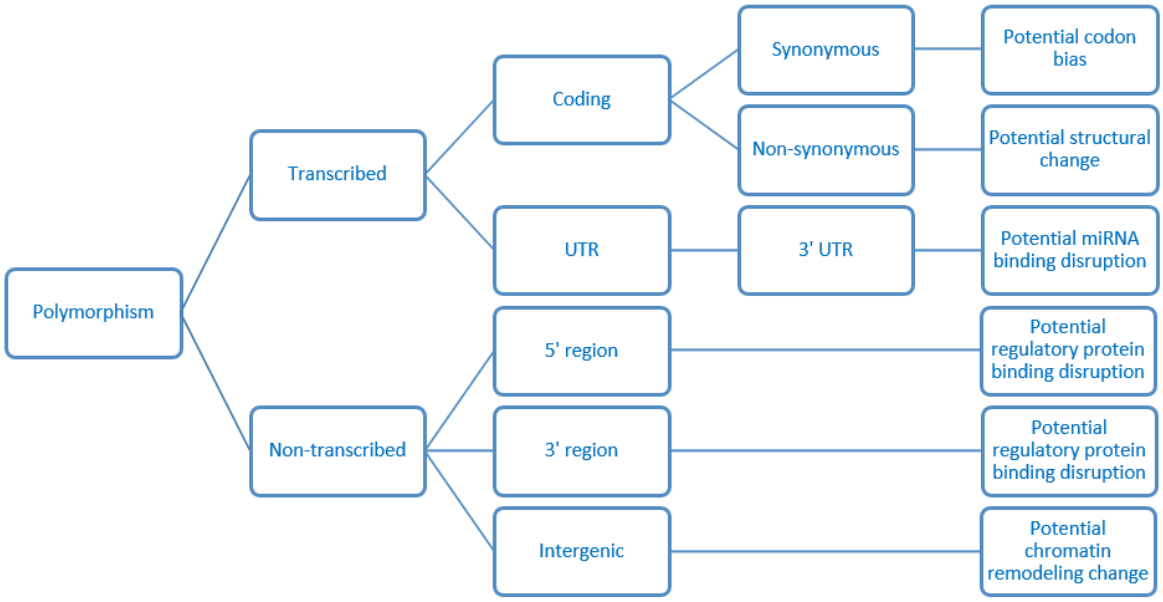

Figure 3-4 SNP functional impact by location

As a final step to select the best candidates, I subjected the list of 77 SNPs (Appendix A section 3) to a web-based tool: HaploReg (389), populated by recently completed ENCODE-generated data. My list was 3.2 and 7 fold enriched in enhancers in K562 and HUVEC cell lines, respectively. I used the experimentally-derived ENCODE database in order to further prioritize the SNPs. This resulted in the final list of candidates shown in Table 3-4: 10 SNPs belonging to 5 different LD bins (according to both HapMap and 1000 Genomes data).

Hapmap LD bins	Encode LD bin	pos (hg19)	variant	RefSeq genes	dbSNP func annot
1	2	102953617	rs3771180	*IL1RL1*	intronic
1	2	102954653	rs13431828	*IL1RL1*	5'-UTR
2	8	102957716	rs1420101	*IL1RL1*	intronic
3	8	102963072	rs6543119	*IL1RL1*	intronic
4	20	102968007	rs4988956	*IL1RL1*	missense
4	20	102968075	rs4988957	*IL1RL1*	synonymous
5	18	102972807	rs4851567	4.3 kb 5' of *IL1RL1*	intergenic
6	7	103035044	rs1420106	208 bp 5' of *IL18RAP*	5' near-gene
6	7	103039584	rs3817465	*IL18RAP*	intronic
6	7	103052995	rs6746271	*IL18RAP*	intronic

Table 3-4 Final candidate functional SNPs using the manual annotation approach
pos: chromosomal position; dbSNP func annot: dbSNP database functional annotation

3.3.2 Lasso and elastic-net regularized generalized linear models results

Out of 1111 lung samples, 1031 had 100% genotype call rate and were retained for the Glmnet analysis. The analysis was performed for each of six probe-sets: 5 for *IL1RL1* (among them one capturing expression of only the long isoform and four capturing both isoforms but more reflective of the soluble isoform expression) and one for *IL18R1*.

Out of the 737 SNPs that went into the analysis, Glmnet yielded 43 SNPs with active coefficients.

I calculated the pair-wise correlation between the probe-sets expression datasets. The four *IL1RL1* probe-sets targeting the expression of both isoforms were very highly correlated with correlation coefficients ranging from 0.927 to 0.975. Due the high correlation, I expected real active SNPs to replicate between the 4 probe-sets. Thus, three SNPs were removed because they had active coefficients only for one of the *IL1RL1* probe-sets (rs4467294, rs13018263 and rs1523198).

The probe-set targeting only the long isoform of *IL1RL1* was correlated to the other 4 probe-sets of *IL1RL1* with correlation coefficients ranging from 0.273 to 0.389. Interestingly, expression of the receptor *IL1RL1* isoform was more correlated with *IL18R1* expression with

a correlation coefficient of 0.49. Soluble *IL1RL1* expression was also more correlated with *IL18R1* expression than with *IL1RL1* receptor expression (correlation coefficients ranging from 0.631 to 0.708). This indicates co-regulation of the expression of soluble *IL1RL1* with *IL18R1* but much less with the expression the receptor *IL1RL1*. As the focus of this study was the regulation of *IL1RL1*, I removed the one SNP which was only active for the *IL18R1* probe-set: rs13011360.

The filtering of the remainder of the active SNPs was based on the scoring system I designed: 11 SNPs had one or more scores that were at least 75% of the maximum score indicating good replication between iterations (numbers in bold in the table) and were retained with high with confidence (Table **3-5**).

	SNPs	N probe 1	N Probe 2	N probe 3	N probe 4	N probe 5	N probe 6	Score all probes (/60)	Score IL1RL1 (/50)	both IL1RL1 isoforms score (/40)	ST2L score (/10)	IL18R1 score (/10)
1	rs4399750	0	**10**	**10**	**10**	**10**	**8**	**48**	**40**	**40**	0	**8**
2	rs11685424	0	**10**	**10**	**10**	**10**	**8**	**48**	**40**	**40**	0	**8**
3	rs1558622	0	**10**	**10**	**10**	**10**	**8**	**48**	**40**	**40**	0	**8**
4	rs7568913	0	**10**	**10**	**10**	**10**	1	**41**	**40**	**40**	0	1
5	rs12712141	**10**	**10**	**10**	**8**	**8**	**10**	**56**	**46**	**36**	**10**	**10**
6	rs974389	0	**10**	**9**	6	**10**	5	**40**	35	**35**	0	5
7	rs10178436	0	6	6	1	**10**	0	23	23	**23**	0	0
8	rs10496357	**7**	6	5	2	**7**	0	27	27	20	**7**	0
9	rs7607548	**10**	0	0	0	0	0	10	10	0	**10**	0
10	rs4850994	**10**	0	0	0	0	0	10	10	0	**10**	0
11	rs10496354	**8**	0	0	0	0	0	8	8	0	**8**	0

Table 3-5 Final candidate functional SNPs using the Glmnet approach
Probeset1 is for the receptor isoform of *IL1RL1* (X100302151_TGI_at); probesets 2 to 5 are for the soluble isoform of *IL1RL1* (X100302783_TGI_at, X100312840_TGI_at, X100148210_TGI_at, and X100148162_TGI_at, respectively) and probeset 6 is for *IL18R1* (X100136977_TGI_at).

Among these active SNPs, 1 SNP was a predictor for all expression probe-sets, 4 were predictors of receptor *IL1RL1* expression solely, 3 SNPs were predictors for expression of

both soluble *IL1RL1* and *IL18R1*, and finally 3 SNPs were predictors for soluble *IL1RL1* expression solely (Figure 3-5).

Interestingly, the SNP predictors of the receptor form of *IL1RL1* were located 270-670 kb from the *IL1RL1* TSS whereas all the SNP predictors of the soluble form of *IL1RL1* were located in and immediately around the *IL1RL1* gene itself. This is suggestive of long-range regulation being involved in the gene expression of the long membrane-bound isoform but not the soluble isoform.

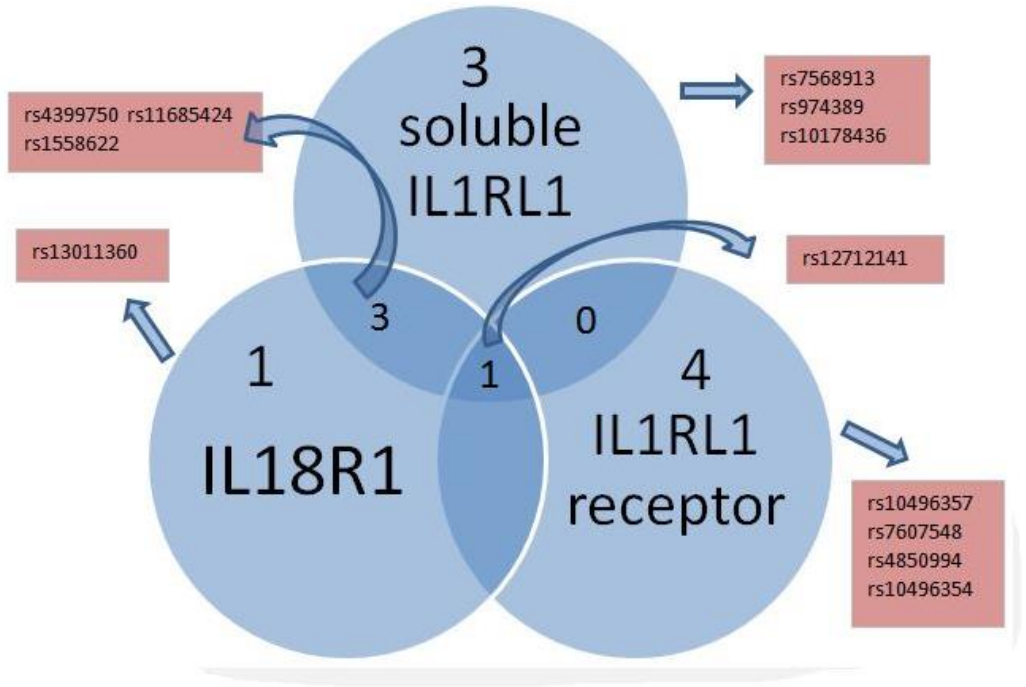

Figure 3-5 Glmnet candidate SNPs overlap between genes and isoforms

The normalized root mean square errors of the group of predictor SNPs for each probe (on each of the 10 iterations, 5 folds per iteration) were calculated to determine the fit of the prediction to the real expression data, i.e. how much variation does the group of SNPs identified explain from the expression variability. Results are shown in Table **3-6**. The

average NRMSE calculated for all six probes was 18.10%; the predictor SNPs do thus explain on average around 82% of the gene expression.

	Average RMSE	Average %	stdev RMSE	stdev %
X100302151_TGI_at	0.2059	20.60%	0.0194	1.90%
X100302783_TGI_at	0.1786	17.90%	0.012	1.20%
X100312840_TGI_at	0.1647	16.50%	0.0114	1.10%
X100148210_TGI_at	0.1884	18.80%	0.016	1.60%
X100148162_TGI_at	0.1645	16.40%	0.0132	1.30%
X100136977_TGI_at	0.1842	18.40%	0.0139	1.40%

Table 3-6 Root mean square errors of the gene expression prediction by the SNPs
stdev: standard deviation

3.4 Discussion

The need to integrate data from different sources, namely genetic association data and expression microarray data arises from a desire by researchers to make more sense of genetic data and move closer to clinical applications. Many groups have taken up the task to identify causal loci from large genetic association datasets; sometimes by using elaborate bioinformatic tools and statistical analysis such as the Differential Allelic Co-Expression test used to estimate the correlation between a SNP genotype with multiple expression levels in a biological pathway assuming that a SNP associated with multiple mRNA transcripts is more likely to be causing the genetic association (390, 391). Fridley *et al*. applied a Bayesian Latent Variable Model to rank SNPs associated with ovarian cancer in GWAS as opposed to ranking the associated SNPs based on their statistical p value; the model incorporates multiple features of the polymorphisms including their allele frequency, functional annotation and association with gene expression (391). Saccone *et al*. successfully applied a systematic prioritization method, namely Genomic Information Networks, to highlight biologically relevant associated loci using genetic association studies of nicotine dependence (392). However, it is of note that in this study, I was interested in dissecting the genetic

association signal from one locus rather than confirming its association with our phenotype of interest: asthma.

Hsu *et al*. prioritized genes associated with osteoporosis phenotypes in a GWAS by performing subsequent genome-wide expression profiling in human and mouse tissue as well as bioinformatic analysis, namely gene enrichment tests (393). Our study achieved a greater resolution, which is to the SNP level, as it was focused on a small region of chromosome 2 encompassing the *IL1RL1* gene and specifically aimed to decipher the biology behind the asthma association of this high LD locus.

Ho *et al*. conducted a genome-wide study of soluble IL1RL1-a (sST2) levels in the serum of 2,991 participants in the Framingham heart offspring study; they found 388 SNPs associated with circulating sST2 levels, all in and around *IL1RL1* in a 400 kb region which were summarized by 11 independent SNPs (with an LD r^2 cutoff of 0.2). 7 out of the 11 SNPs were cis-eQTLs for IL1RL1 probes in our lung study. After conditional analysis, 4 SNPs stayed significantly associated with sST2; all eQTLs for IL1RL1 in our lung dataset.

In this study, I analyzed genetic association data from the *IL1RL1* locus, a high LD region and a corresponding expression microarray dataset in the lung. I used both a manual annotation and an unbiased algorithmic method to select the SNPs with the most potential for causality. In the manual annotation, I included SNPs that were common between *IL1RL1* and *IL18R1* because all GWAS data show association with SNPs in both genes. Both genes could be involved in disease pathogenesis e.g. by means of global chromatin regulation occurring in the area between the two neighboring genes. However, the fact that all the eQTL SNPs for *IL18R1* are also associated with expression of *IL1RL1* but not *vice versa* suggests that

IL1RL1 is the main asthma gene in the region and that *IL18R1* is involved by way of shared regulation. Using Glmnet, I included all available genotypes in the analysis.

It is important to note that using either the manual or Glmnet approach, I found that SNPs in separate LD bins differentially control expression of the soluble and receptor *IL1RL1* isoforms. I identified SNPs associated with expression of soluble *IL1RL1*, but not receptor *IL1RL1*, located in the *IL1RL1* proximal promoter. This is consistent with evidence in the literature that receptor *IL1RL1* distal and not the proximal promoter is responsible for expression of Th2 genes (394). Despite elegant evidence in the mouse that deletion of the proximal promoter obliterates IL1RL1 expression in a cell-type-specific manner (395), it is still unclear in what exact conditions and/or cell type the differential promoter usage is executed. The Glmnet analysis suggests that there is more complex regulation of the full-length receptor form in the shape of long-range regulation from a region more than 260 kb from the *IL1RL1* TSS.

I aimed to adopt a systematic approach to the region and leverage a lung eQTL dataset in order to increase the likelihood of identifying the most promising SNPs responsible for the asthma association in the region. The fact that most SNPs were associated with more than one gene may reflect the presence of SNPs in loci involved with chromatin remodeling in intergenic regions that affect more than single genes (high-order chromatin organization).

Using manual annotation, I found that several LD groups were significantly associated with lung gene expression as well as with asthma in the literature; this indicates possible independent disease mechanisms and/or separate regulation "switches" in the region. Hence, functional studies of one or more SNPs in each of the LD bins may uncover different mechanisms involved in asthma and related inflammatory phenotypes.

Interestingly, using the Glmnet analysis, the SNPs found to best explain soluble *IL1RL1* expression levels belong to only one LD bin. Those SNPs were only in and around *IL1RL1*. I cannot exclude the possibility that SNPs from other LD bins are involved in the regulation of soluble *IL1RL1* expression but I suggest the identified LD bin is the most important in terms of the effect on gene expression. On the other hand, SNPs that best predicted receptor *IL1RL1* expression were further from the target gene itself and ranged from 262 kb to as far as 679 kb from *IL1RL1* TSS.

Long-range gene regulation has been consistently reported in the literature (396) such as in the insulin locus where the regulatory element is 360 kb far from the target gene (397, 398) and the long-range interaction in the glucocorticoid receptor gene leading to transcription repression (399). Long-range regulation is possible due to the inherent flexibility of DNA and the 3D structure of chromatin; research is ongoing to develop more methods to uncover the higher-order organization of the chromatin and the dynamics of protein-DNA interactions (400, 401). Strong evidence exists for an important role of proteins such as Cohesin and CTCF in facilitating long-range regulation and interaction permitting the physical proximity of gene promoter with their distal regulatory elements (402). A further avenue to pursue could be to search for CTCF binding elements within the region identified here as potentially important for receptor *IL1RL1* long-range regulation.

The *IL1RL1* locus is known to have two distinct 5′ non-coding regulatory regions: proximal and distal promoters (382) however the reported location of the distal element is in the non-coding first exon which is of varied length for each isoform, around 600 bp from the transcription start site. I advance the idea that investigating a further *IL1RL1* distal regulatory element around the *MAP4K4* and *IL1R2* genes is warranted.

Regulation of the *IL1RL1* gene is important because of the diverse biological roles of IL1RL1 and IL-33 (100, 101). In tissues affected by asthmatic inflammation, the receptor isoform of IL1RL1 is found in a dimer with IL1 receptor accessory protein on the surface of various cell types (Th2 cells, mast cells, macrophages and DCs) and binds IL-33 to give rise to a MYD88-mediated inflammatory signaling cascade resulting in NF-κB activation and subsequent pro-inflammatory cytokine and chemokine gene expression (180). The soluble isoform of IL1RL1 has been shown to inhibit this inflammatory cascade, by binding to IL-33 thus acting as a decoy receptor (210).

SNP selection can be a time-consuming task depending on the SNP list and can necessitate a certain degree of subjectivity when the SNPs have similar features and cannot be clearly prioritized in terms of functional potential. For instance, in LD bin#1, rs7586983 was a clear choice because among the 16 SNPs in the list, it was the only highly conserved SNP (conservation score=0.89), its regulatory potential score was very high (0.347. On the other hand, although only comprised of 3 SNPs, LD bin # 5 was more difficult to narrow down as the SNPs are all located in the promoter region, 3 of them are predicted to disrupt transcription factor binding, none is evolutionarily-conserved and the closest pair is 3.9 kb apart.

In the case of LD bin #9, there are 8 SNPs, 5 of which are in very close proximity (within 287 bp) and are located in the last exon (exon 11) of the receptor isoform. These SNPs include coding SNPs that were highly significantly associated with asthma in the GABRIEL study GWAS (35). These SNPs should be followed up for functional impact as a group, using reporter gene assays.

Out of the 4 SNPs identified as predictors for the expression of receptor *IL1RL1* (Table **3-5**), rs10496354 is located 679 kb from *IL1RL1* TSS and according to the ENCODE data is located at an enhancer histone mark in the HMEC cell line (mammary epithelial cells). This SNP is in high LD (r^2 greater than 0.8) with near-by SNPs found to bind the transcription factor CCAAT/enhancer-binding protein (C/EBP) beta by Chromatin Immunoprecipitation Sequencing (ChIP-Seq). C/EBPb is known to be involved in expression regulation of inflammatory cytokines genes such as *IL5* and was reported to be important in lung epithelial inflammation (403). rs10496357 is located 672 kb from the *IL1RL1* TSS and is coincident with a promoter histone mark in HepG2 cell line; it is in a DNAse hypersensitive site in 5 cell lines including liver, muscle and blood tissues as well as in fibroblasts; this SNP is in high LD (r^2 greater than 0.8) with SNPs found to be bound to the inflammatory transcription factor NF-κB and RNA polymerase II by ChIP-Seq. rs4850994, situated 273 kb from *IL1RL1* TSS, corresponds to an enhancer histone mark in 3 cell lines including lung fibroblasts and is in a DNAse hypersensitive site in 13 different cell lines including aortic smooth muscle and two pulmonary fibroblast cell lines. This SNP is also in high LD (r^2 greater than 0.8) with SNPs in *IL1R2* found to be bound to NF-κB by ChIP-Seq. The last SNP, rs7607548, is located 262 kb from *IL1RL1* TSS and coincides with an enhancer histone mark in 4 cell lines including lung fibroblasts; it is also in complete LD (r^2 =1) with a near-by SNP found to be bound to the transcription factors USF1 and USF2 by ChIP-Seq.

Concerning the soluble isoform of IL1RL1, Glmnet analysis yielded 6 SNPs that best predict expression levels (Figure **3-5**). The 6 SNPs are all in high LD with a minimum pair-wise r^2 of 0.86 and are located from 10 kb 5′ of *IL1RL1* to the first intron. According to the ENCODE functional data, one of these SNPs, rs7568913 was found to bind GATA2 by

ChIP-Seq and rs1558622 is located in an enhancer histone mark in 4 cell lines including lung fibroblasts, keratinocytes and mammary epithelial cells.

Finally, Glmnet identified one SNP that predicted the expression of all the probes tested (five for *IL1RL1* and one for *IL18R1)*: rs12712141. This SNP binds STAT3 as assayed by ChIP-Seq and is located in the distal promoter of *IL1RL1*. STAT3 is an important transcription factor activated in response to a variety of cytokines and growth factors and plays an important role in lung inflammation (404). The fact that there is one SNP (out of 11) which is shared between the soluble and membrane-bound isoforms indicates that the isoforms do share some degree of regulation and indeed in these data, I see a small correlation between their expression (average $r^2 = 0.33$).

Ultimately, my objective was met as I narrowed down a large number of associated SNPs (by virtue of gene expression, genetic association or LD with the first two) to ten SNPs utilizing manual annotation, and to 11 SNPs using Glmnet analysis with elastic net: very manageable numbers of polymorphisms which I believe to have excellent functional potential. Investigating these SNPs may uncover important biological mechanisms to explain, at the molecular level, the involvement of *IL1RL1* in asthma as well as provide insight into the differential regulation of its soluble and membrane-bound isoforms.

3.5 Conclusion

In silico analysis has become necessary as a follow up to GWAS studies due to the large number of polymorphisms significantly associated with disease states as well as gene expression data that are generated from genome-wide microarrays. A plethora of bioinformatics methods for manually prioritizing candidate disease-causing genes and SNPs

exists (405) but it is not a straight forward process (406) due to the lack of centralized databases, the inconsistencies between certain online resources and the limitation of predictive software based solely on sequence alignments. I was however able to narrow down our search for putative causal SNPs to 10 SNPs that were deemed to be excellent functional candidates from 641 SNPs. Using an unbiased approach based on algorithms in the R programming environment, I was able to identify one important LD bin comprising SNPs in and around the *IL1RL1* gene which best predict the expression of the soluble isoform of IL1RL1 as well as a list of 4 SNPs that best predict the membrane-bound isoform; all four SNPs are located more than 260 kb away from the *IL1RL1* TSS.

This study offers a strong proof-of-concept of the utilization of generalized linear models to select SNPs that best predict gene expression data. To the best of my knowledge, this is the first report of the usage of Glmnet for SNP prioritization as well as the usage of this type of analysis for a continuous variable (such as gene expression in this case).

Three separate important regulatory regions are proposed, one shared between the two *IL1RL1* isoforms and one for each of the isoforms. I also suggest the presence of an important regulatory region controlling the gene expression of the receptor form of IL1RL1 at a distal location as far as 260 kb from the TSS of its target gene *IL1RL1* but not controlling the soluble form. Finally, differential regulation of the two *IL1RL1* isoforms was confirmed and specific SNPs with strong potential for causing the asthma association in the region were identified. The next chapter outlines the functional follow-up of two of the SNPs that I determined (as explained in this chapter) to be good functional candidates.

Chapter 4: Functional follow-up of *IL1RL1* polymorphisms and gene expression analysis

4.1 Overview

The *IL1RL1* gene and its resulting soluble and receptor proteins have emerged as key regulators of the inflammatory process implicated in a large variety of human pathologies. IL1RL1 is important for both innate and adaptive immunity as IL1RL1-b binding with its ligand IL-33 leads to polarization of T helper cells into Th2 and also activates innate immune cells including type 2 innate lymphoid cells. The resulting inflammation is down-regulated by the soluble form of IL1RL1; levels of IL1RL1 are recognized as biomarkers for the severity of various conditions. Thus far, there have been two studies of functional SNPs in the *IL1RL1* gene, one was a candidate SNP study to investigate promoter SNPs associated with atopic dermatitis (239) and the second is a recently published report of the functional analysis of five coding non-synonymous SNPs in the *IL1RL1* receptor (407); it is noteworthy that SNPs located outside of exons and promoters may also have sizable functional effects on gene regulation and thus functional characterization of associated variants is necessary to determine the causal pathways leading to expression and/or function changes in the proteins.

In this chapter, I first describe expression studies of *IL1RL1* soluble and receptor isoforms. The two transcript's expression levels were investigated by qPCR and soluble protein levels by ELISA; I subsequently describe the results of association analysis of the measured serum IL1RL1 levels with asthma as well as with SNPs.

Secondly, I report the results of my association analysis of *IL1RL1* transcript levels with clinical phenotypes as well as SNP effects on modifying these relationships using the lung expression dataset.

Thirdly, I report functional analysis of two *IL1RL1* SNPs: 1) rs1420101 for which I describe the effect of genotype on *IL1RL1* expression, asthma association analysis and the results of a PCR-based splicing assay and 2) rs3771180, for which I describe the effect of genotype on expression, the results of EMSA assays and finally the results of reporter gene assays.

The two SNPs I report functional analysis for, were identified in the top ten candidate SNPs by the manual *in silico* annotation described in Chapter 3. Figure 4-1 outlines the location of all candidate SNPs from the manual annotation. While all these SNPs are viable functional candidates, it was beyond the scope of this thesis to empirically characterize all ten in terms of their functional effects. Therefore, I selected rs1420101 in intron 5 of *IL1RL1* (same location in both the receptor and soluble isoforms' transcripts) and rs3771180 in the proximal promoter. rs1420101 was significantly associated with asthma and eosinophil count in genome-wide and candidate gene association studies (94, 226), rs3771180 was also significantly associated with asthma in GWAS (37), and was shown in ENCODE data to be in a DNAse sensitive region in 30 cell lines, to bind 3 proteins by ChIP-Seq and to be near an enhancer epigenetic mark in 4 cell lines, including lung fibroblasts.

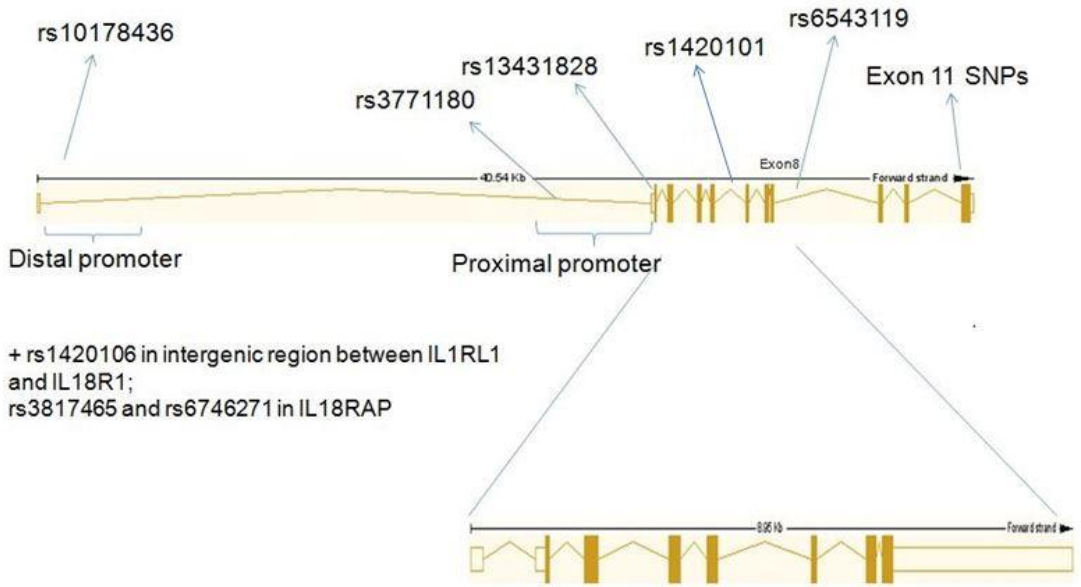

Figure 4-1 Genomic location of functional candidate SNPs selected by manual annotation

4.2 Results

4.2.1 Gene expression studies

4.2.1.1 qPCR

IL1RL1 qPCR experiments were performed in two tissues: blood (samples described in section 2.1.1) and lung (IIAM samples described in section 2.1.1). As expected, the receptor isoform *IL1RL1*-b was the predominantly expressed isoform in blood; the mean Ct for the soluble isoform was 33 and that of the receptor isoform was 28. The soluble isoform IL1RL1-a was the predominant isoform in the lung; the mean Ct was 28 and 33 respectively for the short and the receptor isoform.

4.2.1.2 Protein expression in CAPPS serum samples

Soluble IL1RL1 levels were shown to be elevated in asthmatics (99); in Chapter 3, I reported on a large number of SNPs associated with *IL1RL1* transcript expression. In this section, I

aimed to replicate the association of serum levels with asthma in the CAPPS children and hypothesized that 1) serum IL1RL1 levels will be higher in asthmatic children compared with non-asthmatics; and 2) that protein levels of IL1RL1 will be associated with SNPs in directions consistent with my previous mRNA analysis (Chapter 3).

Serum samples from CAPPS children (section 2.1.3) collected during the 15-year follow-up visits were used for soluble IL1RL1 protein quantification using ELISA. Samples from 195 subjects were quantified and 187 subjects had complete data for protein, phenotypes and SNP genotypes. The characteristics of these subjects are shown in Table 4-1.

	N	Sex (F)	FEV1% predicted	PC20 (mg/ml)	D.pteronyssinus SPT	D.Farinae SPT	Cat SPT	Dog SPT	Trees Midwest SPT	Grass Mix SPT	Weed Mix SPT	Peanut SPT
Asthmatics	21 (11.2%)	11 (52.4%)	96.7 (±8.8)	15.3 (±3.2)	7 (33.3%)	7 (33.3%)	12 (57.1%)	3 (14.3%)	5 (23.8%)	10 (47.6%)	4 (19.0%)	5 (23.8%)
Controls	166 (88.8%)	65 (39.1%)	97.5 (±1.1)	29.3 (±1.9)	48 (28.9%)	44 26.5%)	45 (27.1%)	6 (3.6%)	41 (24.7%)	59 (35.5%)	34 (20.5%)	13 (7.8%)

Table 4-1 Characteristics of CAPPS children used for IL1RL1 protein levels analysis
N: Number of subjects in each group; F: female; PC20: methacholine provocative concentration causing 20% drop in FEV_1; SPT: skin prick test; *Dermatophagoides pteronyssinus* and *D. farinae* are the two major house douse mite species.

Figure 4-2 shows the distribution of serum IL1RL1 levels. IL1RL1 protein levels had a median concentration of 293.6 pg/ml (shown by the blue vertical line in the graph) and an interquartile range of 161.9 to 512.0 pg/ml. The protein levels were not normally-distributed so the values were log2-transformed before proceeding with parametric statistical testing.

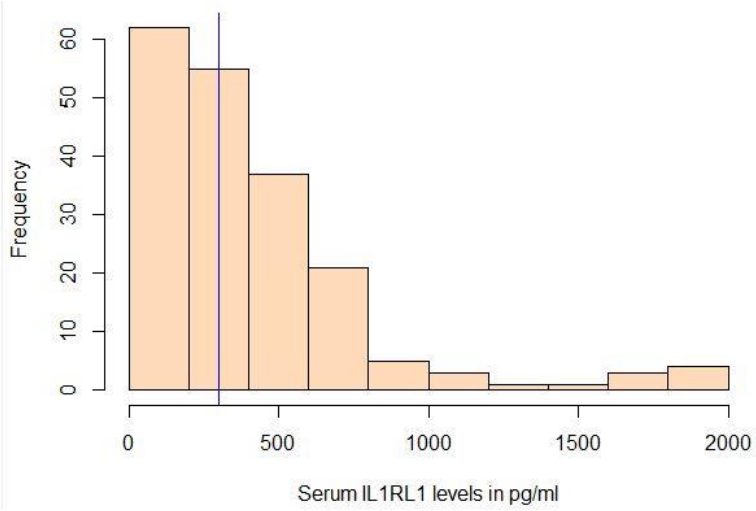

Figure 4-2 Distribution of serum IL1RL1 protein levels as measured by ELISA

Of the 187 CAPPS teenagers, 21 had been diagnosed with asthma by a physician. Using linear regression, a significant association was found between serum IL1RL1-a levels and doctor-diagnosed asthma (p_{adj}=0.015); CAPPS asthmatic children had lower soluble IL1RL1 protein compared to the controls (Figure 4-3). This association of protein levels with asthma was adjusted for sex, as sex was the only confounding variable approaching statistical significance (p=0.06) with regard to protein levels.

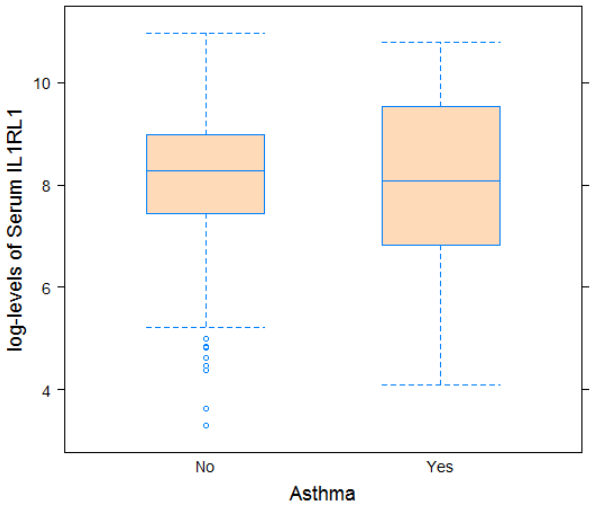

Figure 4-3 Serum IL1RL1 levels by asthma status in the CAPPS teenagers

Linear regression models were used to assess the relationship of serum IL1RL1 levels with nine genotyped *IL1RL1* variants. Two SNPs, rs4090473 and rs950880 were significantly associated with serum levels after Bonferroni correction for multiple testing (Figure **4-4).**

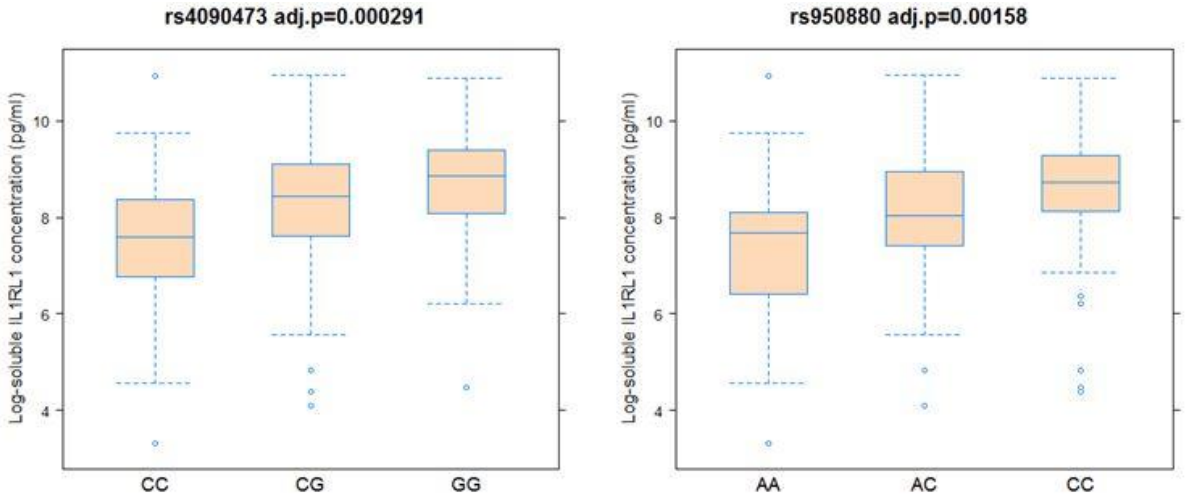

Figure 4-4 Serum IL1RL1 levels by genotypes of significant SNPs

Three SNPs were nominally associated but did not survive correction. I compared the soluble IL1RL1 serum level results with the eQTL data in the lung; Table **4-2** summarizes the analysis results as well as genotype frequencies by asthma status and states the eQTL p values in the lung dataset for each SNP.

The most significant protein quantitative trait locus (pQTL), rs4090473, was the second most significant eQTL for soluble *IL1RL1* in the entire chromosome 2q region with $p=9.9\times10^{-65}$ and in high LD with the most significant eQTL (rs1362349, $p=9.4\times10^{-65}$) with r^2 of 0.89. The eQTL direction was the same as for the protein, i.e. the minor allele, C, was associated with lower mRNA and protein expression.

The second most significant pQTL measured by ELISA was rs950880; this SNP was in an LD group of soluble *IL1RL1* eQTLs with the third lowest p values at $p=1.8\times10^{-34}$. The direction of the association with protein expression was in agreement with the mRNA expression: the A allele was associated with lower soluble IL1RL1 expression.

SNP	Genotypes	N	Asthma N	Asthma %	Control N	Control %	P value	Corr P value	Lung eQTL P value
rs4090473	CC	53	4	7.5%	49	92.5%	**0.00003234**	**0.00029106**	9.9×10^{-65}
	GC	88	16	18.2%	72	81.8%			
	GG	36	5	13.9%	31	86.1%			
rs950880	CC	57	10	17.5%	47	82.5%	**0.000176**	**0.001584**	1.8×10^{-34}
	AC	88	11	12.5%	77	87.5%			
	AA	32	4	12.5%	28	87.5%			
rs1420103	CC	104	16	15.4%	88	84.6%	**0.008096**	NS	9.2×10^{-20}
	AC	58	7	12.1%	51	87.9%			
	AA	15	2	13.3%	13	86.7%			
rs6719130	CC	145	19	13.1%	126	86.9%	**0.01918**	NS	3.0×10^{-6}
	CT	30	6	20.0%	24	80.0%			
	TT	2	0	0.0%	2	100.0%			
rs1921622	TT	42	4	2.3%	37	21.0%	**0.02815**	NS	1.1×10^{-13}
	CT	96	16	9.1%	73	41.5%			
	CC	35	5	2.8%	29	16.5%			
rs1946131	GG	148	22	14.9%	126	85.1%	0.07925	NS	NS
	AG	24	2	8.3%	22	91.7%			
	AA	5	1	20.0%	4	80.0%			
rs3771175	TT	138	20	14.5%	118	85.5%	0.6781	NS	4.1×10^{-8}
	AT	37	5	13.5%	32	86.5%			
	AA	2	0	0.0%	2	100.0%			
rs1420089	TT	143	16	11.2%	127	88.8%	0.2824	NS	3.6×10^{-11}
	CT	32	9	28.1%	23	71.9%			
	CC	2	0	0.0%	2	100.0%			

Table 4-2 Serum IL1RL1 levels by *IL1RL1* SNP genotypes

N: Sample size; Corr P value: Bonferroni-corrected P value (P value multiplied by 9). P values in bold are significant at a threshold of $p < 0.05$. Values in green in the last column denote lung eQTL p value ranking consistent with pQTL ranking.

Subsequently to the work reported above, Ho *et al*. published the first GWAS of circulating soluble IL1RL1 levels (407); they reported 388 significant SNPs ($p<5\times10^{-8}$) in the *IL1RL1* region, which they shortlisted to 11 independent SNPs using an LD r^2 threshold of 0.2 and further decreased to 4 SNPs after statistical conditional analysis.

As explained in Chapter 3, I identified a similar number of independent SNPs in my manual *in silico* annotation. I compared my resulting SNPs and the significant pQTL

(ELISA) with the final 4 SNPs identified by Ho *et al.* (Table 4-3). All 4 SNPs in the GWAS of circulating soluble IL1RL1 levels were also eQTLs in the lung expression dataset, and were concordant in direction of association to that in our eQTL data as shown in the 7th and 8th column of the table (except for one SNP where no comparison was possible due to missing data).

Method	Encode LD bin	Chromosomal position (hg19)	SNP	RefSeq genes location	Lung eQTL?	Protein association direction (pQTL)	mRNA association direction (eQTL)	Notes
Glmnet	-	102248081	rs10496354	65kb 5' of MAP4K4	Yes			
Glmnet	-	102255114	rs10496357	58kb 5' of MAP4K4	Yes			
Glmnet	-	102654228	rs4850994	9.2kb 3' of IL1R2	Yes			
Glmnet	-	102665552	rs7607548	15kb 5' of IL1R1	Yes			
Glmnet	1	102917788	rs4399750	10kb 5' of IL1RL1	Yes			
Glmnet	1	102920037	rs7568913	7.9kb 5' of IL1RL1	Yes			
ELISA	1	102922987	rs4090473	11.9kb 5' of IL1RL1	Yes			
Glmnet	1	102926511	rs10178436	1.5kb 5' of IL1RL1	Yes			
Glmnet	1	102926981	rs11685424	980bp 5' of IL1RL1	Yes			
Glmnet	1	102930147	rs1558622	IL1RL1	Yes			
soluble IL1RL1 GWAS + my ELISA data	2	102932562	rs950880	IL1RL1	Yes	A --> lower Exp.	A --> lower Exp.	p<E-16 6 probes IL1RL1+ IL18R1
Glmnet	1	102936981	rs974389	IL1RL1	Yes			
ELISA data	3	102938389	rs1420089	IL1RL1	Yes			
Soluble IL1RL1 GWAS	4	102948632	rs1420103	IL1RL1	Yes	A --> higher Exp.	A --> higher Exp.	p<E-16 6 probes IL1RL1+ IL18R1
ELISA data	4	102948632	rs1420103	IL1RL1	Yes			
Glmnet	1	102953067	rs12712141	IL1RL1	Yes			
Soluble IL1RL1 GWAS	5	102953444	rs17639215	IL1RL1	Yes	A --> higher Exp.	A --> higher Exp.	I removed this SNP from manual annotation because eQTL p>E-06
manual	6	102953617	rs3771180	IL1RL1	Yes			
manual	6	102954653	rs13431828	5'UTR IL1RL1	Yes			
Soluble IL1RL1 GWAS	-	102957291	rs13029918	IL1RL1	Yes	G --> lower Exp.	NA	I removed this SNP from manual annotation because eQTL p>E-06
manual	7	102957716	rs1420101	IL1RL1	Yes			
ELISA data	5	102958236	rs6719130	IL1RL1	Yes			
ELISA data	6	102960210	rs3771175	IL1RL1	Yes			
ELISA data		102961929	rs1946131	IL1RL1	No			
manual	7	102963072	rs6543119	IL1RL1	Yes			
ELISA data		102966067	rs1921622	IL1RL1	Yes			
manual	8	102968007	rs4988956	IL1RL1	No			
manual	8	102968075	rs4988957	IL1RL1	No			
manual	3	102972807	rs4851567	IL18R1	Yes			
manual	9	103035044	rs1420106	104bp 5' of IL18RAP	Yes			
manual	9	103039584	rs3817465	IL18RAP	Yes			
manual	9	103052995	rs6746271	IL18RAP	Yes			

Table 4-3 Comparison of my identified candidate SNPs with Ho *et al.* soluble *IL1RL1* GWAS SNPs

Method: How the SNPs were identified; exp: expression; SNPs of the same color are in high LD with each other ($r^2 > 0.8$) except gray-colored SNPs which are not in LD ($r^2 < 0.2$) with any other SNP in this list; green color in 7th and 8th columns signifies concordance between direction of association between pQTL and eQTL association with expression.

4.2.1.3 Lung eQTL data analysis

IL1RL1 soluble and receptor isoforms play important roles in the asthmatic lung and participate in processes leading to several features of asthma pathophysiology. Polymorphisms in the *IL1RL1* gene are associated with asthma and affect gene expression levels at the mRNA and protein levels. However, although I have shown highly significant associations of SNPs with *IL1RL1* gene expression, there remains considerable variability in the expression levels that is unexplained. For example, rs1420101 explains only 12.2% and 1.2% of the variance in soluble and receptor *IL1RL1* mRNA levels, respectively. The lung expression dataset provided a good opportunity for me to explore the causes of the remaining variability by examining the effect of relevant phenotypes, as well as their interaction with *IL1RL1* SNPs, on *IL1RL1* transcript expression. In this section, I will describe the results of these analyses.

Details of variable selection for this analysis are described in section 2.9.2.2. Briefly, the lung dataset contained 1111 subjects with 97 clinical and demographic variables. There was considerable missingness in most of the variables but only subjects with complete data (expression, genotypes and phenotypic and demographic variables) are allowed in the linear models. First, irrelevant variables such as dates and unspecified medication use were removed. Secondly, categorical variables with less than 10 subjects in one category were removed. Lastly, plotting of the number of subjects by the number of variables with complete data for those subjects (Figure 2-4) was performed; I used this plot to ascertain an arbitrary cutoff (45 variables) in order to simultaneously maximize the number of variables and subjects to include in the analysis.

The final number of subjects with complete data for the 45 variables included in the statistical models was 845 subjects. The characteristics of these subjects are delineated in Table 4-4; a summary of the full list of variables is included as appendix 6.3B.1.

Variables	Mean (± SE) or N (%)
Age	59.1 ± 0.48
BMI	25.22 ± 0.18
Pack year	35.95 ± 1.02
Predicted FEV1	2.95 ± 0.02
Predicted FVC	3.85 ± 0.03
Sex	M: 452 (54.1%)
Smoking status	Current: 170 (20.3%), Ex: 527 (63.0%), No: 139 (16.6%)
Weight	70.89 ± 0.57

Table 4-4 Characteristics of lung dataset subjects used for *IL1RL1* expression analysis

The *IL1RL1* gene is targeted by five probes in the lung expression dataset. One probe targeting the receptor *IL1RL1*, one probe targeting both isoforms, two probes targeting the soluble *IL1RL1* isoform transcript and one probe with unknown annealing sites; the 4 probes not targeting the receptor *IL1RL1* exclusively were highly correlated (Table 4-5).

	X100302151_TGI_at (Long)	X100302783_TGI_at (short)	X100312840_TGI_at (short)	X100148210_TGI_at (both isoforms)	X100148162_TGI_at (short)	Location in soluble transcript	Location in receptor transcript
X100302151_TGI_at	1	0.3598763	0.3718859	0.2604865	0.3762355	NA	Exon 10
X100302783_TGI_at	0.3598763	1	0.9531528	0.952927	0.9401503	unknown	unknown
X100312840_TGI_at	0.3718859	0.9531528	1	0.9335779	0.9738763	3'UTR	Intron 8
X100148210_TGI_at	0.2604865	0.952927	0.9335779	1	0.9252433	Exon 8	Exon 8
X100148162_TGI_at	0.3762355	0.9401503	0.9738763	0.9252433	1	3' UTR	Intron8

Table 4-5 Location of the five probes on the *IL1RL1* transcripts and their pairwise correlation

The two most correlated probes anneal exclusively to the short isoform transcript (they anneal to an intron of the receptor isoform). Therefore, I arbitrarily selected one of these two probes, X100312840_TGI_at, as a representative to investigate the relationship of soluble *IL1RL1* with clinical phenotypes. The probe X100302151_TGI_at anneals to the 10th exon of *IL1RL1*, only present in the transcript encoding the receptor isoform and was therefore used

to investigate potential association of clinical phenotypes to *IL1RL1*-b. I did not have the sequence of the X100302783_TGI_at probe but due to the high correlation of the expression levels with the expression of the other three probes, it is likely that it aligns to the short isoform. Figure 4-5 depicts the distribution of the expression levels by the selected probes.

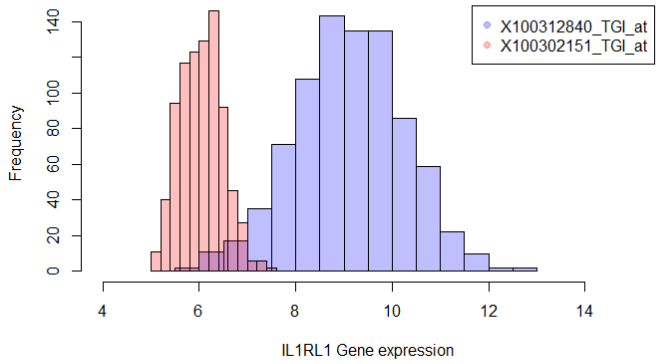

Figure 4-5 *IL1RL1* probe expression distribution in lung tissue

X100312840_TGI_at (blue) is one of the short *IL1RL1* isoforms' probes. X100302151_TGI_at (pink) is the probe for the receptor *IL1RL1* isoform.

One of the aims of this section was to analyze the potential effect of these SNPs on the association of *ILRL1* expression with clinical variables. As explained in Chapter3, I identified 21 SNPs as important predictors of *IL1RL1* expression and potential functional SNPs: 10 SNPs from the manual annotation (Table 3-4) and the 11 SNPs from the Glmnet approach (Table 3-5). From these 21, I used the following list of 9 independent SNPs with available genotypes in the lung expression dataset for the analysis: rs10496354, rs10496357, rs4850994, rs7607548, rs1558622, rs3771180, rs1420101, rs4988956 and rs1420106.

For both of the *IL1RL1* isoforms, I investigated the following two questions: 1) Are *IL1RL1* expression levels associated with one or more clinical variables? And 2) Do any of the 9 candidate SNPs described above significantly interact with those clinical variables, i.e. do genotypes modify the effect of the variable on expression?

For this purpose, I ran multiple linear regression models (one for each *IL1RL1* probe) that included main effects and interaction terms (phenotype*genotype) with all the variables (see section 2.9.2.2 for details of the modeling procedure). Batch was included as a main effect only, as it represented a potential confounding variable.

Graphs illustrating the main effects of the associations of gene expression with the variables were generated from the multiple regression models and reflect the true effects adjusted to all other effects in the models.

Graphs illustrating the interactions of SNPs with clinical variables were generated from a pairwise correlation, and not from the multiple regression models; they are intended to illustrate the general trend of the association but do not exactly depict the results discussed in this chapter unless put in the context of the coefficient estimates obtained from the model.

I multiplied the p values by 4 to account for the two models tested for each of the two probes (X100312840_TGI_a for soluble *IL1RL1* and X100302151_TGI_at for receptor *IL1RL1*).

The statistical model for the soluble *IL1RL1* isoform yielded an adjusted R^2 coefficient of determination of 0.337; that means 33.7% of the variability of the soluble isoform expression is explained by the variables and interactions in the model, the results of which are presented in Table 4-6. The receptor IL1RL1 model yielded an adjusted R^2 of 0.642.

All results with significant uncorrected p values (α threshold 0.05) in the final models are delineated in Table 4-6.

	Receptor *IL1RL1*			Soluble *IL1RL1*		
Term	Estimate	P	Pcorr	Estimate	P	Pcorr
Antibiotics:rs1420101 GG	0.515	0.024	0.096			
Antibiotics:rs1420106 TT.CT	-0.337	0.024	0.094			
Antibiotics:rs3771180 CC	-0.426	3.3E-03	**0.013**			
Antibiotics:rs4988956 GG	0.636	7.6E-03	**0.031**			
Antidepression	0.159	0.037	0.149			
α-1 antitrypsin deficiency:rs3771180 CC	0.737	9.7E-03	**0.039**			
Cystic fibrosis:rs3771180 CC	0.535	3.0E-04	**0.001**			
Idiopathic pulmonary fibrosis:rs3771180 CC	0.313	0.046	0.186			
BatchLAVAL	-0.518	<2.0E-16	**<8.0E-16**	-0.174	0.047	0.188
BatchUBC	0.185	1.4E-06	**5.7E-06**			
Bisphosphonates:rs10496357 TT	-0.316	6.5E-06	**2.6E-05**			
Bisphosphonates:rs1420101 GG	0.249	0.015	0.058			
Bronchiectasis	0.452	0.035	0.138			
Bronchiolitis obliterans	-0.624	0.025	0.099			
Diuretics	0.104	5.2E-03	**0.021**	0.486	0.045	0.178
Diuretics:rs1420101 GG				-0.775	0.007	**0.027**
Large cell carcinoma:rs3771180 CC	0.216	0.041	0.165			
Long acting beta agonists:rs1420106 TT.CT	0.184	0.026	0.106			
No medication:rs1420101 GG	0.412	2.6E-04	**0.001**			
No medication:rs1420106 TT.CT	-0.230	1.9E-03	**0.008**			
Oral steroids				1.555	0.001	**0.003**
Oral steroids:rs10496357 TT				-0.428	0.037	0.147
Oral steroids:rs1420101 GG				-0.762	0.021	0.086
rs1420106 TT.CT	0.112	7.2E-03	**0.029**			
rs4988956 AG	-0.115	0.012	**0.046**			
Sequestration/fungus	-0.678	0.031	0.124			
Smoking status (No)				0.497	0.019	0.076
Solitary granuloma	0.620	0.026	0.103			
Squamous cell carcinoma	0.135	0.013	0.052			
Weight				0.005	0.009	**0.035**

Table 4-6 Analysis of main effects and interactions of *IL1RL1* isoform gene expression, clinical variables and SNP genotypes

P: P value; P_{corr}: corrected p value; No medication: a categorical variable where "Yes" signifies a subject is not taking any of the medications specified in the study, and "No" signifies a subject is taking at least one of the medications specified in the study.

The receptor *IL1RL1* expression had a range of 2.62 arbitrary expression units (4.98 – 7.60). Soluble *IL1RL1* expression was 2.3 fold greater than receptor *IL1RL1*, with a range of 6.16 (5.73 – 11.89).

Although there was no statistically significant main effect of antibiotic use on *IL1RL1* transcript expression, there was a significant interaction between four SNPs and antibiotic

use on the expression of receptor *IL1RL1* (Figure 4-6), two of theses associations survived correction (rs3771180 and rs4988956).

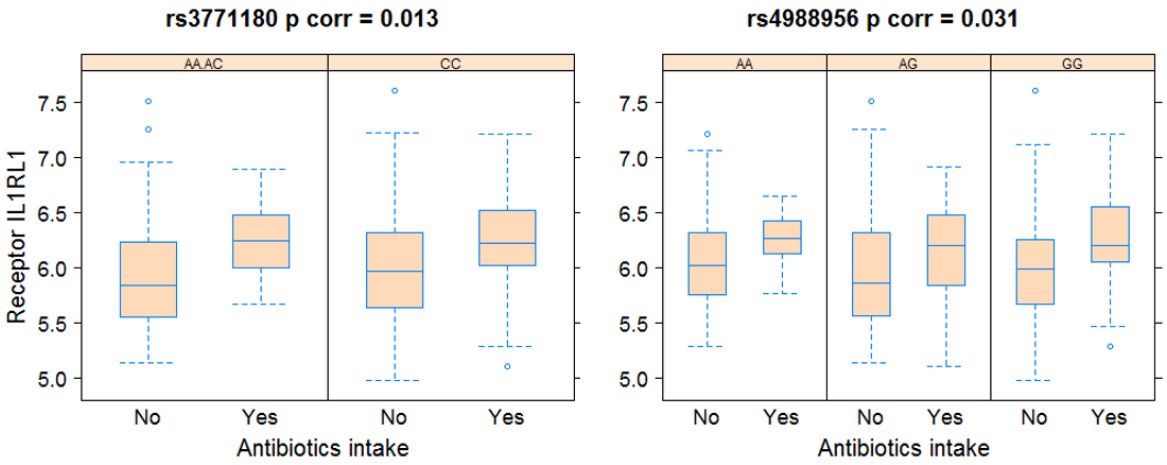

Figure 4-6 Interaction of rs3771180 and rs4988956 with antibiotic use on receptor *IL1RL1* expression

The blue circles in the graphs correspond to samples which are higher than the 75th quartile by greater than 1.5 x the interquartile range of the data.

rs3771180 significantly interacted with diagnoses of α-1 antirypsin deficiency and cystic fibrosis and nominally with idiopathic pulmonary fibrosis to influence the gene expression of receptor *IL1RL1* with corrected p values of 0.039, 0.001 and 0.186 respectively. The expression of receptor *IL1RL1* was significantly greater in subjects carrying the CC genotype (Figure 4-7); there was only one subject in the minor genotype group in α-1 antirypsin deficiency subjects. However, 10 CF subjects had the minor allele (A); 32 had the CC genotype and exhibited a 20% greater receptor expression of the total expression range as compared with the minor allele subjects.

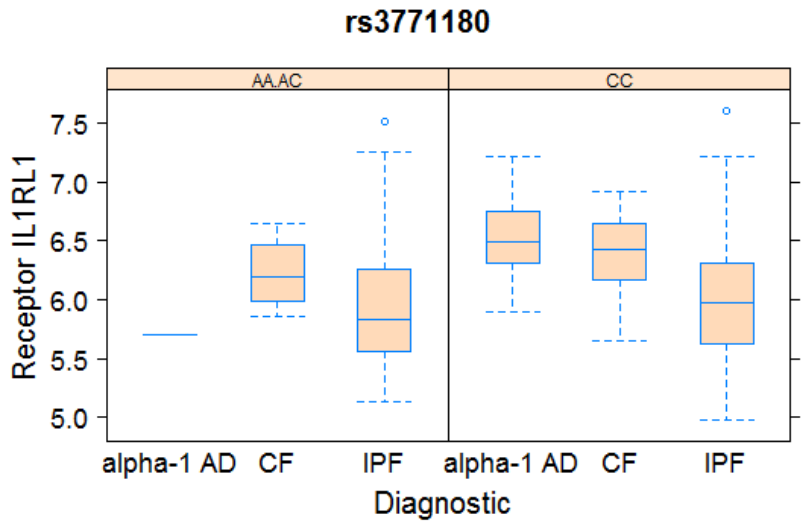

Figure 4-7 Interaction of rs3771180 alleles with three diagnoses on the receptor *IL1RL1* expression

There was a very significant batch effect on both *IL1RL1* transcripts' expression. Figure 4-8 shows that effect in the entire cohort. The following are the p values for the receptor *IL1RL1*: Laval vs. Groningen: $p<2\times10^{-16}$, UBC vs. Groningen: $p=1.4\times10^{-5}$; for the soluble *IL1RL1* (measured by the probe X100312840_TGI_at), Laval vs. Groningen: $p=2.5\times10^{-5}$, UBC vs. Groningen: p=0.69. The batch effect was identical for the other soluble *IL1RL1* probes; there was no batch effect for the X100148210_TGI_at probe (shared between the isoforms).

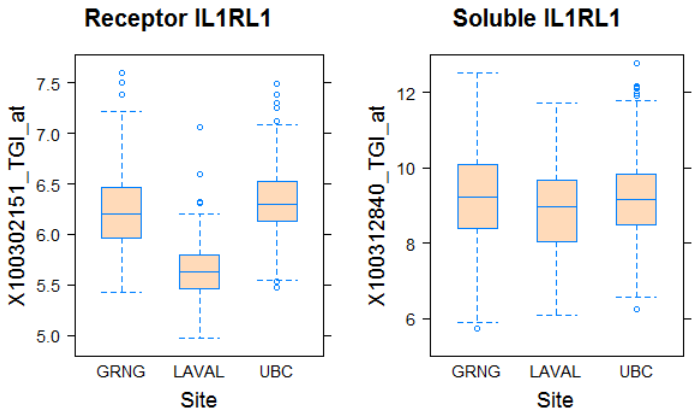

Figure 4-8 Batch effect on *IL1RL1* isoform expression

Although there was no statistically significant main effect of bisphosphonate use on *IL1RL1* gene expression, there was a significant interaction between the SNP rs10496357 and bisphosphonate use on the expression of receptor *IL1RL1* with p_{corr}= 2.6 10^{-5} (Figure 4-9). Only 3 subjects with the CC genotypes were taking the medication, therefore the CC and CT genotypes were grouped and compared to the TT genotype. In individuals taking bisphosphonates, there was significantly lower receptor *IL1RL1* in the TT genotype (65 subjects) compared with the CC and CT genotypes (30 subjects).

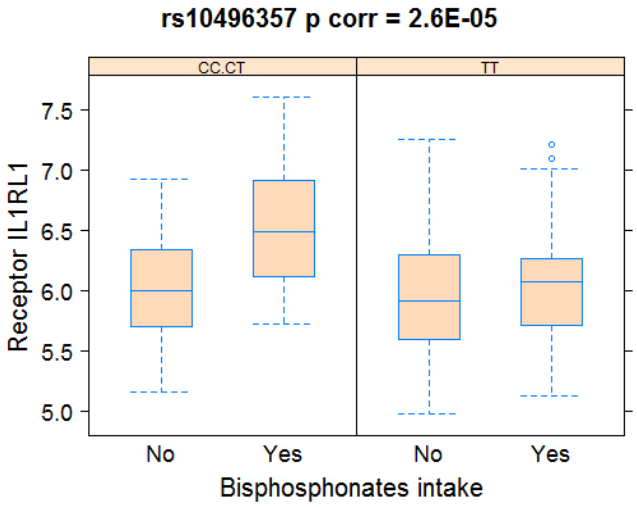

Figure 4-9 Interaction of rs10496357 with bisphosphonate use on receptor *IL1RL1* expression

There was a significant association between diuretic use and receptor *IL1RL1* gene expression (p_{corr}= 0.021) but not the soluble isoform expression. However, the SNP rs1420101 significantly interacted with this medication intake to modify soluble *IL1RL1* gene expression with p_{corr}= 0.027 (Figure 4-10). Whereas the association of soluble *IL1RL1* (albeit only nominally significant) was positive, this relationship was reversed in subjects with the GG genotypes. The rs1420101 genotype distribution by diuretic intake was as follows: AA: Yes: 18/ No: 110, AG: Yes: 65/ No: 338 and GG: Yes: 43/ No: 263.

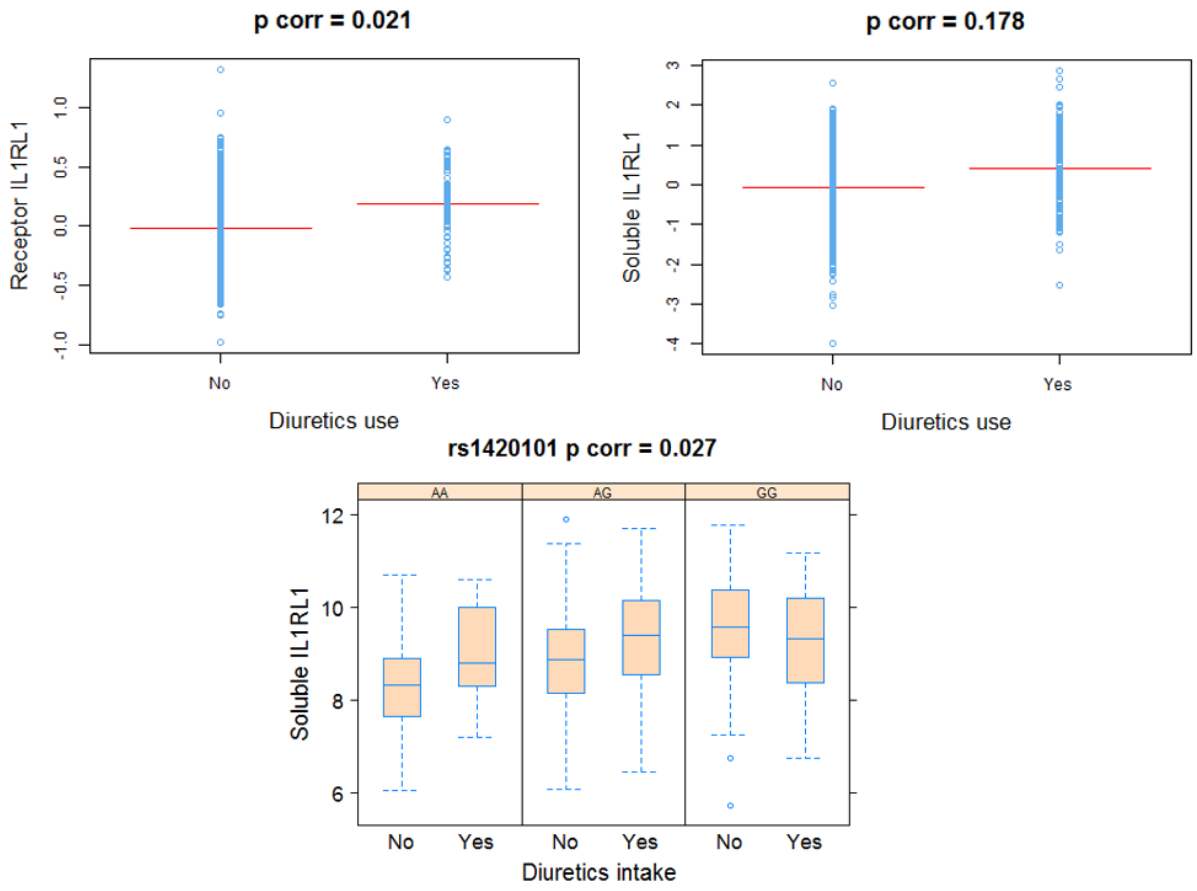

Figure 4-10 Main effect of diuretic intake and interaction with rs1420101 on *IL1RL1* gene expression

Although no statistically significant main effect was observed for the no medication intake variable on *IL1RL1* gene expression, rs1420101 and rs1420106 genotypes were significantly associated with receptor *IL1RL1* expression stratified by this variable (p_{corr} = 0.001 and 0.008 respectively**)** (Figure 4-11).

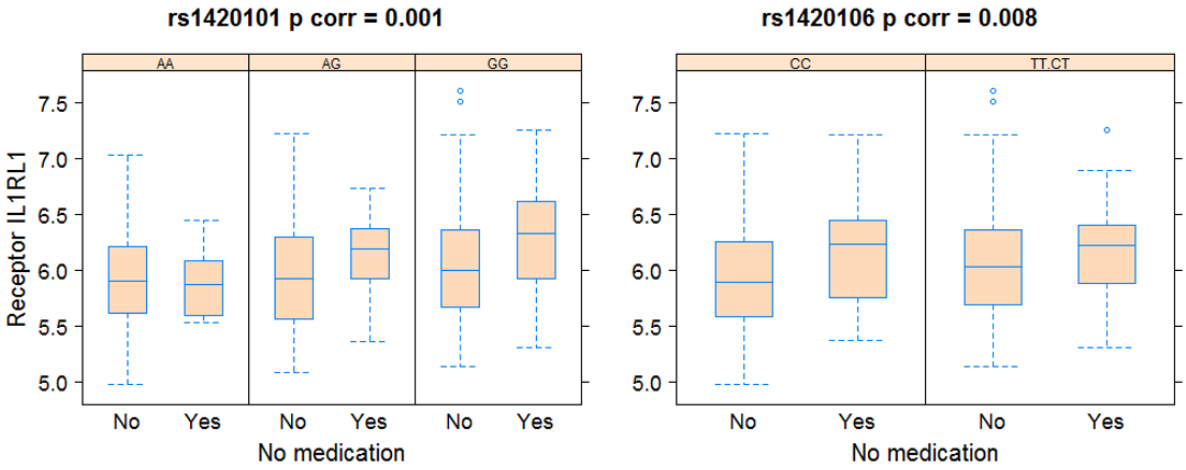

Figure 4-11 Interaction of rs1420101 and rs1420106 with no medication intake with *IL1RL1* receptor expressiosson

There was a statistically significant association between oral steroid intake and soluble *IL1RL1* (p_{corr}=0.003) (Figure 4-12); rs10496357 and rs1420101 modified this association but this did not survive multiple testing correction. Subjects taking oral steroids (112 subjects) had significantly higher soluble *IL1RL1* expression (25% of the expression range).

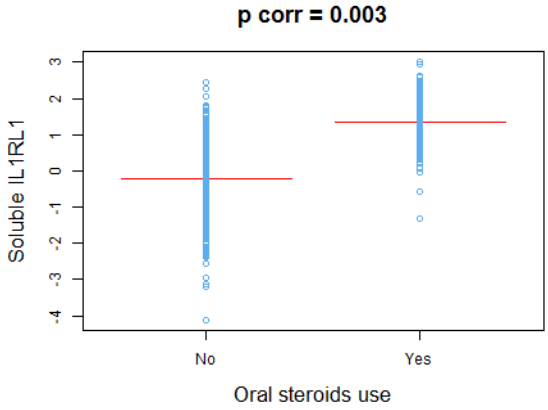

Figure 4-12 Association of oral steroid use with soluble *IL1RL1* expression

In addition, there was a nominal association of smoking status with soluble *IL1RL1* expression where non-smokers have greater expression as compared with former or current

smokers (Figure 4-13); however this association did not survive multiple testing correction. Finally, there was a positive association between weight and soluble IL1RL1 expression with $p_{corr} = 0.035$.

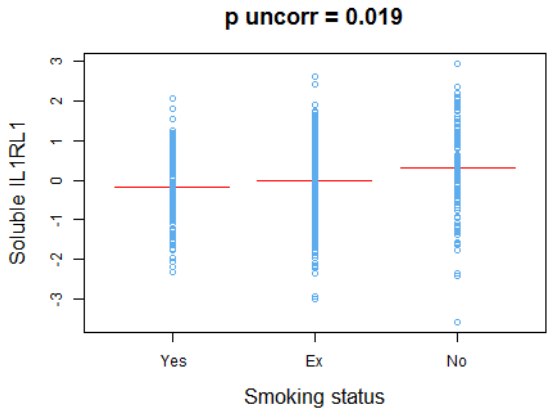

Figure 4-13 Association of smoking status with soluble *IL1RL1* expression

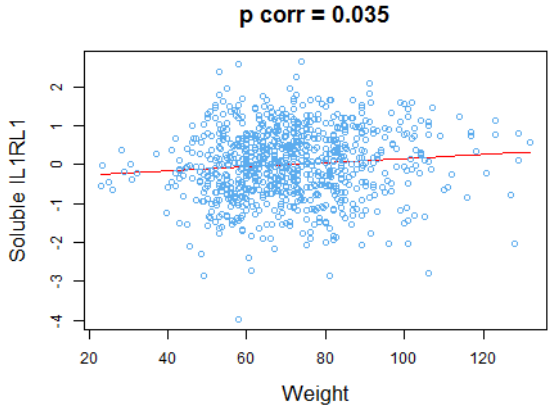

Figure 4-14 Association of weight with soluble *IL1RL1* expression

4.2.2 Functional analysis of SNP rs1420101

4.2.2.1 Analysis of rs1420101 relationship to IL1RL1 gene expression

eQTL analysis of the lung dataset showed that rs1420101 was significantly associated with all five *IL1RL1* probes albeit at a much lesser significance with the receptor isoform expression probe. The A allele leads to a markedly lower expression of the soluble isoform

($p=2.2\times10^{-16}$) and a trend toward increasing expression of the receptor isoform (p=0.00079) (Figure 4-15).

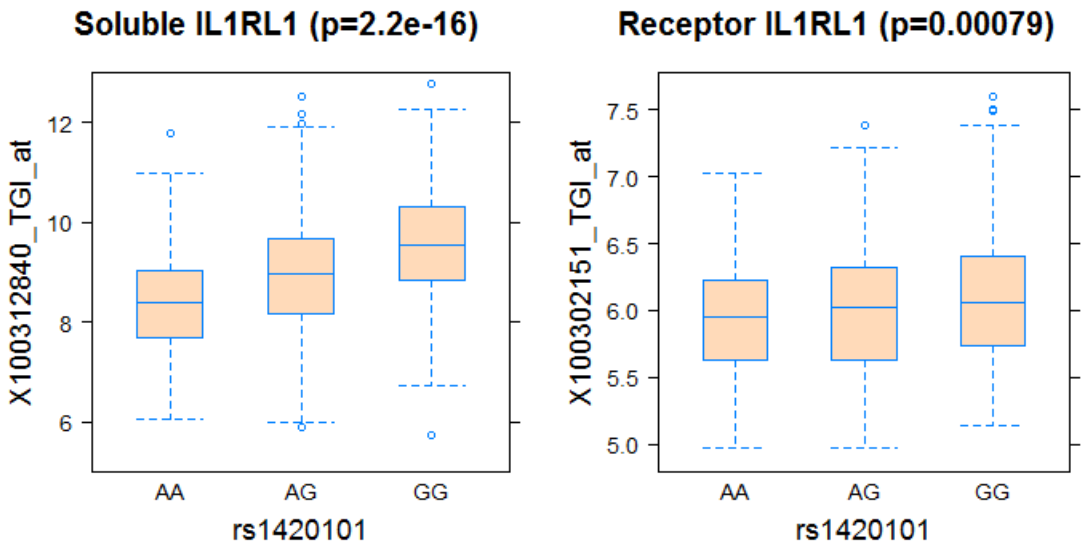

Figure 4-15 *IL1RL1* isoform expression by rs1420101 genotypes

Soluble *IL1RL1* expression was measured using the probe X100148210_TGI_at and the receptor expression was measured using the probe; X100302151_TGI_at. A/G alleles are on the reverse strand.

4.2.2.2 Asthma association analysis

I hypothesized that rs1420101-T is associated with asthma risk in the cases and controls from the blood and lung samples. In order to test that, rs1420101 was genotyped in the blood samples. The genotypic data were in Hardy-Weinberg equilibrium. Table 4-7 shows the genotype distributions in cases and controls.

	N of controls	%	N asthmatics	%
CC	19	19.39	14	14.29
CT	21	21.43	22	22.45
TT	12	12.24	10	10.2

Table 4-7 rs1420101 genotyping data (blood samples)

There was no significant association between genotypes and asthma status in the blood samples. The lung expression study samples had genotypes available for rs1420101; but there was no significant association between asthma and rs1420101; of note is that there were only 30 asthmatic samples for this test.

4.2.2.3 PCR-based splicing assay

In virtue of its intronic location, I hypothesized that rs1420101 plays a role in *IL1RL1* mRNA splicing and I designed a PCR-based assay to investigate this question (see section 2.7 for details). cDNA samples derived from blood RNA, IIAM lung RNA or B-lymphoblastoid cell line RNA were amplified using the R6, R7 and R8 primer pairs.

Lung (Figure 4-16) and blood-derived (Figure 4-17) cDNA produced PCR products of the expected sizes for R6 and R7 primer pairs (76 and 154 bp, respectively) regardless of genotype.

In the blood samples, the R7 primer pair produced very faint bands for AA and GG genotypes; consequently, I verified the presence of a product by sequencing using the R7 primer pairs of DNA cut-out from the gel in the expected location. Sequencing was performed in both directions and the AA and AG genotypes were both confirmed to produce a PCR product for R7 of the expected size.

However, only samples with the GG genotype, for both blood and lung, gave the expected product using the R8 primer pair (352 bp); the resulting bands observed were considerably less intense than the other bands. In contrast, B-lymphoblastoid cDNA samples amplified using all primer pairs and produced several band sizes in addition to the expected ones.

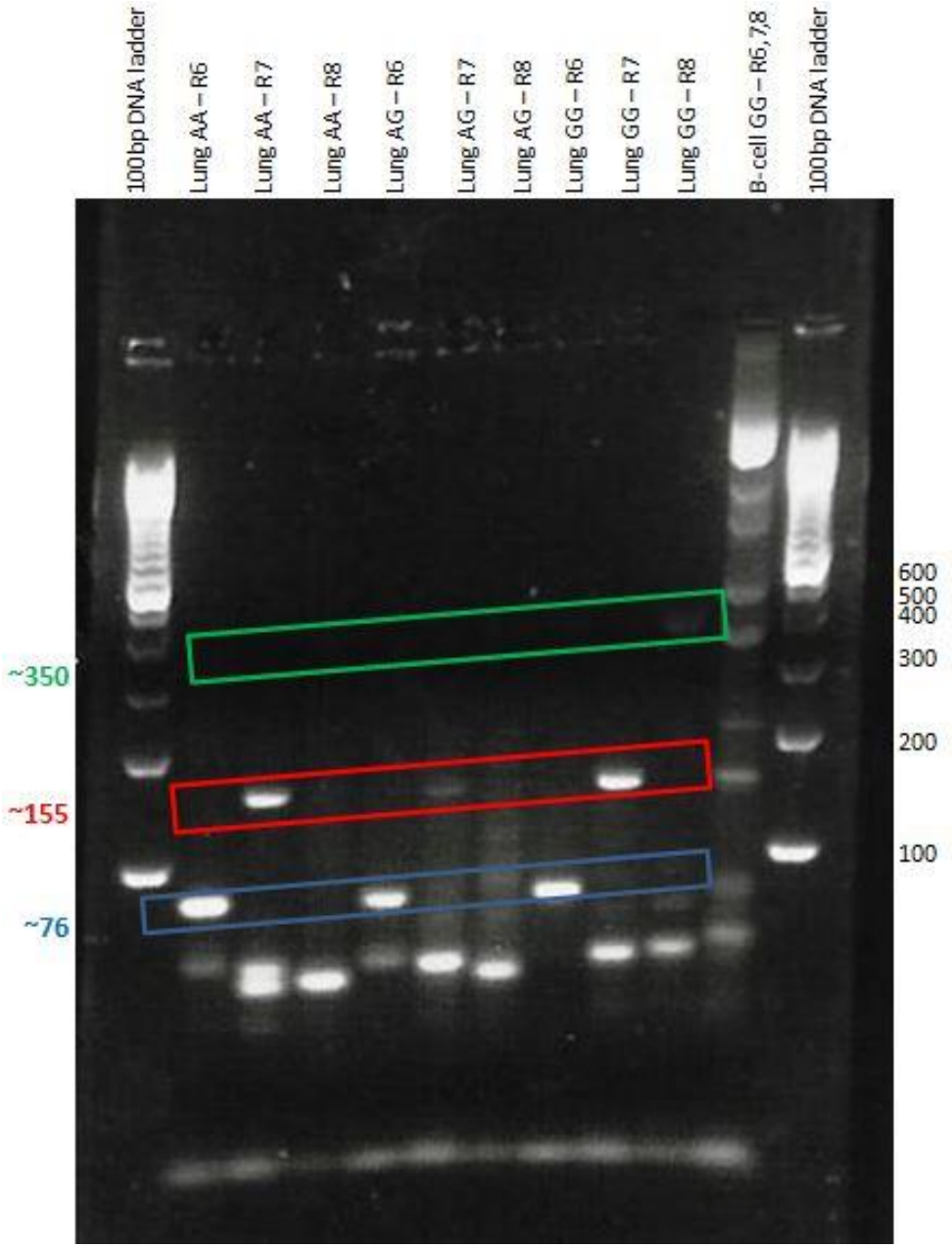

Figure 4-16 rs1420101 splicing assay gel for lung samples

The lanes in order are: lane1: 100bp ladder; lane2, 3 and 4: lung sample of AA genotype amplified with R6, R7 and R8 primer pairs respectively; lane 5, 6 and 7: lung sample of AG genotype amplified with R6, R7 and R8 primer pairs respectively; lane 8, 9 and 10: lung sample of GG genotype amplified with R6, R7 and R8 primer pairs respectively; lane 11: B cell multiplex lane contains PCR product using all B cell cDNA and all 3 primer pairs simultaneously.
The blue box corresponds to 76bp band size; the red box corresponds to 155bp band size; the green box corresponds to 350bp band size.

The PCR band pattern did not differ between asthmatic and control cDNA samples of the same genotypes (Figure 4-17).

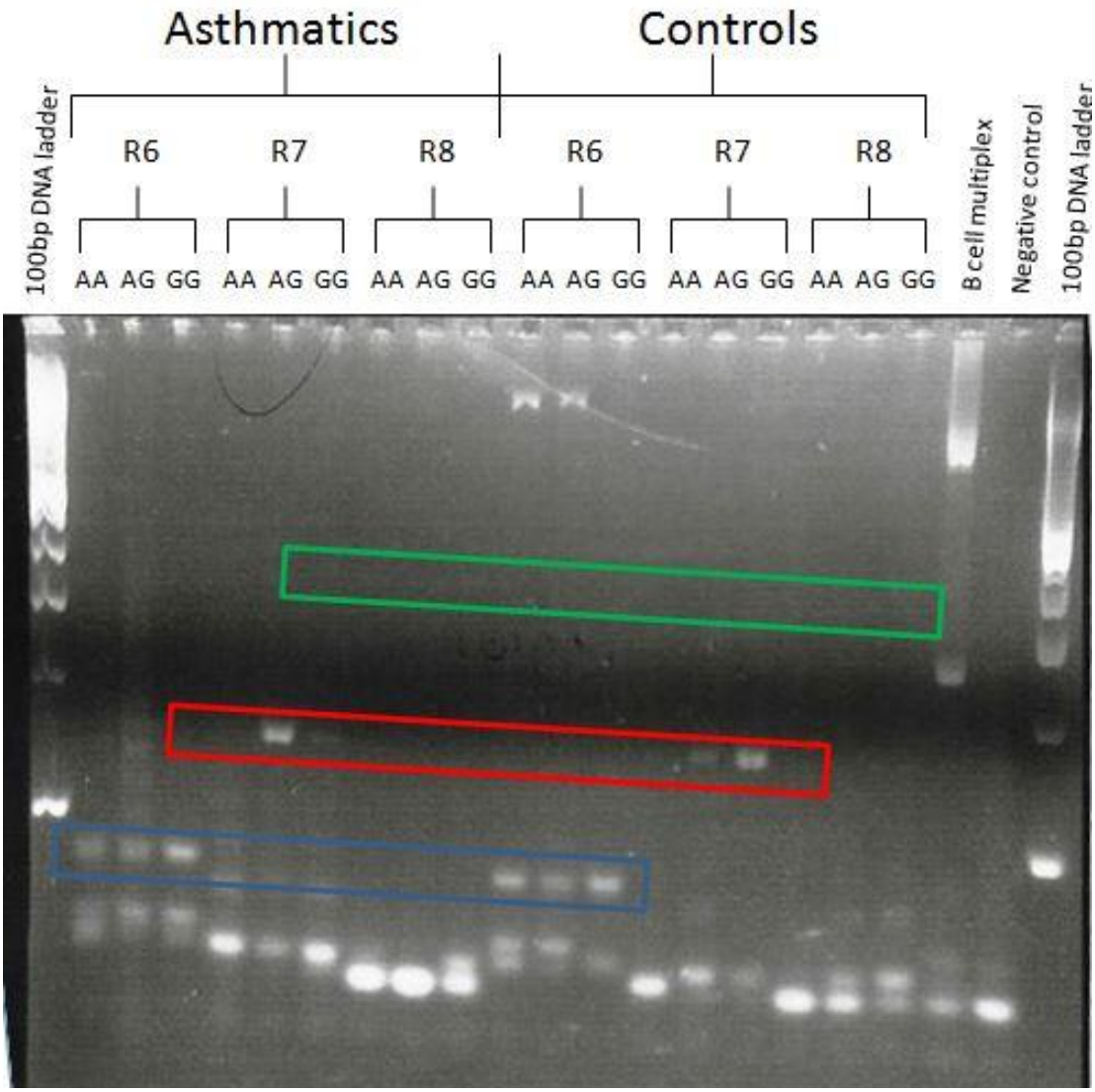

Figure 4-17 rs1420101 splicing assay gel for blood samples by asthma status

R6, R7 and R8 refer to the primer pairs using the reverse primers R6, R7 and R8 respectively; B cell multiplex lane contains PCR product using all B cell cDNA and all 3 primer pairs simultaneously.
The blue box corresponds to 76bp band size; the red box corresponds to 155bp band size; the green box corresponds to 350bp band size.

4.2.3 Functional analysis of SNP rs3771180

4.2.3.1 Analysis of rs3771180 in relation to IL1RL1 gene expression

eQTL analysis of the lung dataset showed that rs3771180 was significantly associated with the 4 *IL1RL1*-a probes with $p < 10^{-7}$ but not associated with the *IL1RL1* receptor isoform (p=0.3) (Figure 4-18).

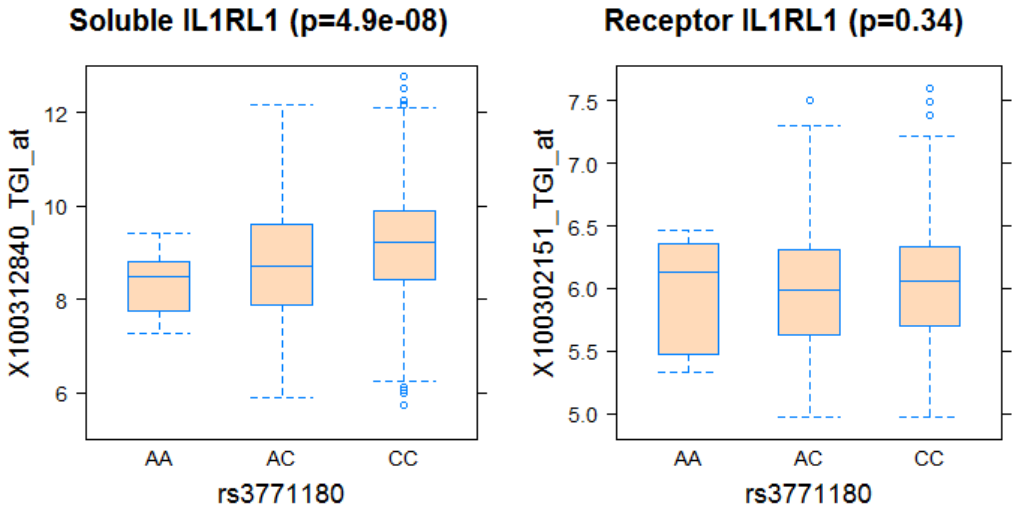

Figure 4-18 *IL1RL1* probe expression by rs3771180 genotypes.

Soluble *IL1RL1* isoform was measured by the probe X100148162_TGI_at; receptor *IL1RL1* isoform was measured by the probe X100302151_TGI_at.

4.2.3.2 Electrophoretic mobility shift assay

I hypothesized that rs3771180 was a functional SNP and that its asthma protection allele would lead to higher soluble *IL1RL1* expression and/or lower receptor *IL1RL1* expression.

Due to its location 5' of the gene, I aimed to determine whether rs3771180 results in an allele-specific differential binding of a nuclear protein.

I selected the A549 cell line as a model for human lung epithelial cells; this cell line has been shown to offer a good representation of the immune response in the lung epithelium

(408-410) and is frequently used for *in vitro* studies relevant to asthma. I also used Human Umbilical Vein Cell (HUVEC) line as a control because ENCODE ChIP-Seq data showed rs3771180 bound nuclear protein in this cell line.

Independent EMSA assays showed DNA/protein interactions between A549 nuclear extract and rs3771180 oligonucleotides irrespective of allele. Using HUVEC nuclear extracts also led to the formation of DNA/protein complexes on EMSA gels.

EMSA assays showed no differential protein/DNA interactions by alleles of rs3771180 using nuclear extracts from A549 cell lines (blue ovals in Figure 4-19). There seemed to be allele-specific differential binding in HUVEC but it was not replicated (blue ovals in Figure 4-19). Sp1 transcription factor oligonucleotides and HeLa cell nuclear extract were used as positive controls in each EMSA experiment and a band of DNA/protein complex was observed consistently.

Considerable attention and time went into optimizing the EMSA experiments in order to obtain at least 3 independent experiments showing this result; including optimizing the nuclear extraction to obtain more concentrated extracts, incubation times for radioactivity labeling and for binding of oligonucleotides and nuclear extracts, temperature for the incubations, efficient gel electrophoresis and drying, as well as visualization.

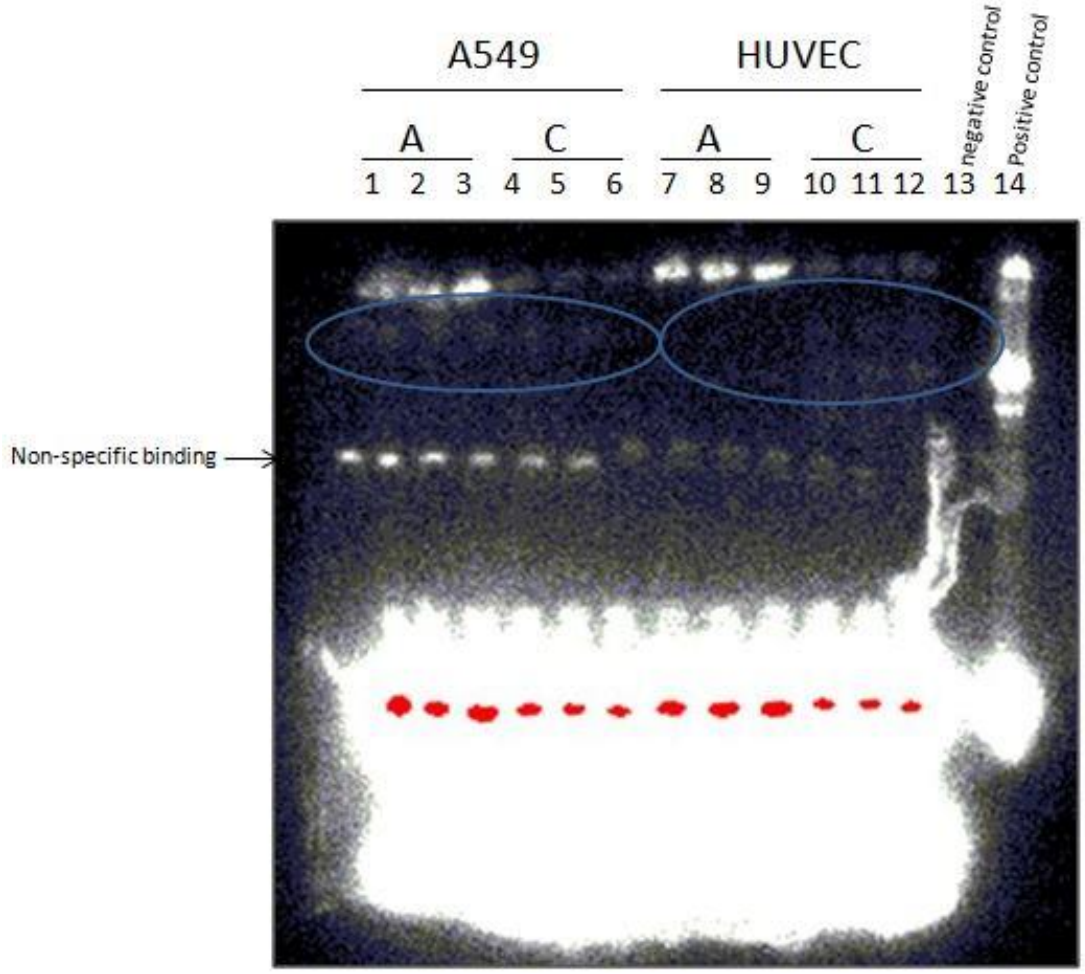

Figure 4-19 rs3771180 EMSA gel

Details of the lanes: 1: A hot only; 2: A + non-specific cold oligo; 3: A + A cold; 4: C hot only; 5: C + non-specific cold oligo; 6: C + C cold; 7: A hot only; 8: A + non-specific cold oligo; 9: A + A cold; 10: C hot only; 11: C + non-specific cold oligo; 12: C + C cold; 13: Negative control; 14: Sp1 and HeLa nuclear extract positive control. Lanes 1 to 6: A549 nuclear extract, lanes 7 to 12: HUVEC nuclear extract.
The red lines are because of high intensity due to unused (unbound) radioactively-labeled oligonucleotides.

4.2.3.3 Luciferase reporter gene assay

In addition to EMSA, another method to determine whether a SNP of interest is regulatory is Luciferase reporter gene assays.

I hypothesized that a plasmid containing the asthma-protective A allele of rs3771180 will lead to altered luciferase reporter activity.

Luciferase activity ratio was ascertained by the Firefly luciferase over the *Renilla* luciferase activities ratio and the signal was background-subtracted using wells of non-transfected cells containing all reagents. The A allele and C allele plasmids had a mean activity ratio of -0.07 and 2.54 respectively (Figure 4-20).

Although there was a trend towards decreased expression for the A allele plasmid Luciferase activity compared to the C allele, this difference did not reach statistical significance (p=0.12). This result was obtained from two independent experiments, ten biological replicates per allele in each experiment and ten technical duplicates for each biological replicate.

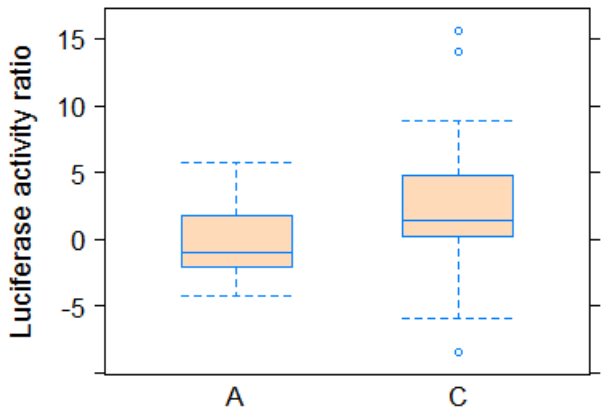

Figure 4-20 rs3771180 Luciferase activity ratio by allele

4.3 Discussion

IL1RL1 has been shown to be involved in asthma in several GWAS and is likely a true asthma gene given the consistent replication as well as its biological plausibility. *IL1RL1* encodes two main transcripts: the long isoform which produces IL1RL1-b, a receptor bound to membranes of hematopoietic cells, mainly Th2 cells and mast cells, and the short isoform

which produces IL1RL1-a, a smaller soluble protein that also binds IL-33, preventing the receptor signaling.

I investigated transcript expression levels of both *IL1RL1* isoforms in blood and lung samples; as expected, blood *IL1RL1* was predominantly the receptor isoform and lung *IL1RL1* was mostly the soluble isoform.

I also performed ELISA assays to investigate protein soluble IL1RL1 levels. I found a modest but significant association between serum IL1RL1-a levels and doctor-diagnosed asthma in the CAPPS teenagers' samples (p=0.015 adjusted for sex).

The direction of the asthma association I found was opposite to published studies, i.e. I found that asthmatics had lower IL1RL1-a protein levels. The CAPPS children dataset comprised mainly mild asthmatics among the 22 asthmatics diagnosed by a doctor. In the entire dataset, there were 41 subjects with airway hyperresponsiveness as defined by PC20 lower than 7.8 mg/ml; among them only 6 (15%) were diagnosed by a doctor to have asthma. Furthermore, a proportion of the non-asthmatics were allergic subjects as assessed by skin prick testing (Table 4-1). Hence, the comparison of serum levels was made between the asthmatics and other children, some of whom had AHR and/or were allergic. Elevated serum IL1RL1-a levels have been shown in allergic subjects (411) and subjects with allergic diseases such as allergic rhinitis (243, 412). The CAPPS cohort was initially designed to study children who would be at high risk of developing asthma and allergic disease and the asthma status was observed prospectively.

My data showed decreased soluble IL1RL1 protein levels in the CAPPS asthmatic children, and this was concordant with the eQTL data showing lower soluble *IL1RL1* mRNA in asthmatics; decreased levels agree with my hypothesis that soluble IL1RL1 is protective

from asthma. Together, these observations suggest elevated serum IL1RL1 might be found in more severe disease or during exacerbations, however in mild disease, as is the case for the CAPPS cohort, the level of inflammation might not stimulate increased soluble IL1RL1.

Although serum IL1RL1 levels were reported to be significantly elevated in asthmatics, this effect was significantly stronger in moderate asthmatics compared to mild asthmatics (220, 413) or was present only when the levels were measured in an acute asthmatic episode (220). Ali *et al*. also reported increased soluble IL1RL1-a (a.k.a. sST2) levels in the serum of asthmatic children during acute episodes compared with after the acute episodes; however they also found a negative correlation between soluble IL1RL1 and eosinophil counts measured during an asthma exacerbation as well as a stronger negative correlation between eosinophil counts and the soluble IL1RL1 levels during the attack and during convalescence (99). The same type of observation was made in allergic rhinitis by Baumann *et al*. who ascertained that, while being significantly increased in seasonally pollen exposed allergic subjects compared to non-allergic controls, nasal soluble IL1RL1 was inversely correlated with nasal symptom scores of severity (412). These data indicate soluble IL1RL1 is produced in acute inflammatory situations, likely as a result of a regulatory loop to lessen the IL-33/IL1RL1-b signaling. There is evidence for such a regulatory loop whereby coding SNPs in the *IL1RL1* receptor (not present in the soluble isoform transcript) led to increased responsiveness of IL-33 to its receptor IL1RL1-b, which in turn led to greater release of soluble IL1RL1 (407).

Indeed, soluble IL1RL1 has been clearly shown in asthma animal models to inhibit allergic inflammation by dampening the signaling cascade precipitated by IL-33 binding to the receptor IL1RL1 isoform (217, 218). An inflammatory role of soluble IL1RL1 in asthma

does not align with those *in vivo* findings therefore the elevated levels of soluble IL1RL1 observed in published studies of pathological states such as asthma or cardiovascular disease most probably indicate a response of the immune system to the inflammatory cascade that may prevent further lung damage. In fact, the ST2 ELISA kit from R&D, the most used kit for soluble IL1RL1 level quantification, detects both free and IL33-bound soluble IL1RL1. High levels detected using this kit indicate the total level of soluble IL1RL1 but leave unanswered the question of the proportion of free to IL-33-bound soluble IL1RL1. High levels of IL-33 have been observed in asthma (413, 414); that would require an increase of soluble IL1RL1 to neutralize it.

Ho *et al*. published the first GWAS of circulating soluble IL1RL1 levels; they found that missense SNPs located in the intracellular domain of the receptor IL1RL1 participated in a potential auto-regulatory loop whereby the alleles modify the receptor signaling and result in an increased IL-33 production which in turn leads to heightened soluble IL1RL1 secreted levels (407). Ho *et al*. reported that SNPs accounted for 45% of the soluble IL1RL1 levels adjusted for age and sex with 388 SNPs associated with levels at $p < 5 \times 10^{-8}$, all in and around *IL1RL1*; this is well aligned with my lung eQTL data. Ho *et al*. distilled the associated SNPs down to 11 independent SNPs as defined by an LD $r^2 < 0.2$ and further decreased to 4 SNPs after statistical conditional analysis. In my *in silico* manual annotation of the lung eQTL dataset described in Chapter 3, I identified a similar number of independent SNPs. I compared my resulting SNPs with the Ho *et al.* soluble IL1RL1 GWAS data SNPs (Table 4-3). Of the 4 SNPs Ho *et al*. obtained after conditional analysis, rs950880 is in high LD with rs1420101, one of my candidate SNPs ($r^2 = 0.91$). rs1420103 was in moderate LD with another of my candidate SNPs, rs1420106 ($r^2 = 0.59$) and the other two soluble IL1RL1

GWAS SNPs: rs13029918, and rs17639215, were not in LD with any of my candidates (r^2 threshold of 0.2).

Two SNPs survived the stringent Bonferroni correction. The most significant pQTL, rs4090473, was in almost complete LD (r^2 between 1 and 0.89) with a group of SNPs (rs4399750 and others) I identified as important predictors of soluble *IL1RL1* expression using the Glmnet analysis detailed in Chapter 3. It was the second most significant eQTL for soluble *IL1RL1* in the entire chromosome 2q region with $p=9.9\times10^{-65}$ and was in high LD with the most significant eQTL. These SNPs were part of a group of 25 tightly correlated SNPs, all eQTLs for the soluble *IL1RL1* at a maximum p value of 8×10^{-65}; this group of SNPs span a 37 kb region encompassing the *IL1RL1* promoters and first intron. The protein data were concordant with the eQTL data as the minor allele, C, was associated with lower mRNA and protein expression.

The second most significant pQTL, rs950880, was one of the 4 SNPs identified in the soluble IL1RL1 GWAS (407) after conditional analysis as highly associated with serum levels; it was in very high LD with rs1420101 ($r^2 = 0.91$), one of the candidate SNPs I identified using the manual *in silico* analysis and investigated for potential splicing effect in this thesis.

Subsequently, I attempted to uncover any relationship of the *IL1RL1* gene expression with clinical phenotypes as well as investigate any potential role of the candidate functional SNPs, identified in Chapter 3, in modifying those associations.

Out of the 5 *IL1RL1* probes, the receptor probe (X100302151_TGI_at) had the lowest expression levels and the smallest range of expression values (5.08 – 7.51); the other 4

probes were all tightly correlated (92.5% – 97.4%), and had greater expression levels and variation than the receptor isoform with ranges from 7.04 to 9.89 expression units.

A model containing all the nine independent SNPs (among the SNPs identified in Chapter 3) explained 26.3% of the soluble *IL1RL1* expression but only 6.7% of the receptor transcript expression. I ascertained that there was a considerable batch effect possibly confounding the lung expression data, and therefore added this covariate to the SNPs models; this resulted in a modest increase in the adjusted R^2 of the soluble transcript expression (to 28.4%) but a considerable increase in that of the receptor transcript (from 6.7% to 58%). This indicates that the genetic variation conferred by these SNPs plays a major role in determining the level of soluble *IL1RL1* mRNA but not as much in the receptor transcript's regulation. In order to attempt to explain additional variability of both isoforms' expression, I added clinical and demographical variables to the SNPs as well as interaction terms between them. An unbiased best fitted model selection was applied which results in final models with the highest adjusted R^2 for each of the isoforms.

The receptor isoform expression was significantly associated with conditions such as α-1 antitrypsin deficiency and CF and with medication intake such as bisphosphonates and antibiotics. However, the amount of variability explained by the addition of these variables was minimal. The adjusted R^2 for the receptor isoform expression increased by only 9%.

Soluble *IL1RL1* transcript expression was positively associated with diuretic use except for subjects carrying the rs1420101-GG genotype (asthma-protective). Diuretic drugs are prescribed for treatment of many conditions, mainly heart failure and hypertension; soluble IL1RL1 is a well-recognized marker of heart disease. Soluble *IL1RL1* expression was also associated with oral steroids intake where subjects taking oral steroids exhibited greater

expression. Oral steroids are prescribed in order to dampen the body's immune response in many inflammatory and immune diseases such as asthma, IBD, severe allergies, autoimmune diseases and joint diseases; since soluble IL1RL1 acts as decoy to the receptor to IL-33, a greater expression would indeed benefit patients of any of these disorders.

Similarly to the receptor isoform, the soluble IL1RL1 transcript model with the clinical and demographical variables added yielded an adjusted R^2 increase of about 5%, suggesting that in this dataset, the genetic variation accounted for the majority of the variability observed. This is consistent with the soluble IL1RL1 GWAS study, where the authors found clinical variables accounted for 14% of the protein levels (407). There remains a considerable amount of soluble IL1RL1 variability (~66%) unexplained by my model.

rs1420101 is a T to C polymorphism in the fifth intron of *IL1RL1* with a minor allele frequency of 37%. The T allele was associated with asthma risk and increased eosinophil count (94, 226, 234). I found this SNP to be highly associated with decreased *IL1RL1* soluble isoform gene expression in the lung eQTL dataset (Figure 4-15) and this SNP was in high LD (r^2=0.91) with a SNP I found significantly associated with protein levels of soluble IL1RL1 in CAPPS children samples (rs1420101 was not tested in that analysis as it was covered by a Tag-SNP).

I could not replicate the asthma association of rs1420101in blood samples. This is most likely due to the small sample size (46 asthmatics and 50 controls); I had a 66% power to detect a genetic effect of 2 fold increase in risk and 28% power to detect a 1.5 fold increase in risk. Using the lung dataset, I could not replicate the asthma association either because of the low number of asthmatics in the cohort (24 with complete genotypes). That meant that the analysis had 64% power to detect a 2 fold risk increase and 27% power to detect a more

modest effect of 1.5 fold risk increase, which would be in line with published genetic effect sizes in complex disease. Indeed, this SNP, rs1420101, was associated with asthma by Gudbjartsson *et al.* with an OR of 1.16 (94).

Due to its intronic location, I hypothesized that the rs1420101-T allele leads to down-regulation of expression by playing a role in mRNA splicing. It is indeed predicted to abolish a splicing site using ESE Finder (415) and FAS-ESS (416) web-based bioinformatics tools.

I used a PCR-based assay to attempt to uncover any splicing pattern difference between rs1420101 genotypes in lung and blood sample cDNA. The R6 and R7 primer pairs produced PCR products of the expected size but no amplification was observed for the R8 primer pair in subjects of AA or AG genotypes and very faint PCR bands were obtained for the GG genotypes; although the G allele should result in lower transcript levels than the A allele as shown in the lung dataset (Figure 4-15), I do not believe this to be the result of lower levels since the same GG samples produced normal PCR bands using the other two primer pairs. It might be a technical artifact of binding of the R8 reverse primer because I optimized all three primer pairs to use the same annealing temperature in order to amplify them together and thus minimize variability.

In silico analysis of splicing in *IL1RL1* using the ASG web-based tool (417) showed exon 6 to be a potential cassette exon; I did not find any evidence for that, as shown by the absence of PCR bands of sizes consistent with skipping of exon 6 in the transcripts (there would have been no product from R6 primer pairs due the fact that the R6 reverse primer anneals to exon 6, for example).

Overall, I found no evidence of a differential splicing pattern by rs1420101 genotypes or by asthma status as demonstrated by the presence of PCR product of expected sizes based

on normal splicing and reference sequences. Despite the classical view that intronic SNPs can be functional by affecting splicing, it is now known that intronic SNPs may influence gene expression by disrupting regulatory protein binding which could lead to enhancing or down-regulating gene expression; rs1420101 could in fact be functional by other means which could be investigated by DNA binding assays.

To determine whether rs3771180 is in fact a functional SNP as identified in the manual *in silico* analysis discussed in Chapter 3, I investigated its relationship to *IL1RL1* gene expression and performed functional assays: EMSA and reporter gene to investigate a potential allele-specific effect on binding to a nuclear protein and on expression.

rs3771180 is a C to A polymorphism in the non-coding first intron of *IL1RL1* with a minor allele frequency of 16% in European populations. I identified rs3771180 as a strong eQTL for *IL1RL1* short isoform transcripts in the lung (Figure 4-18). This SNP was reported by Torgerson *et al.* to be significantly associated with asthma in three populations from different ethnic groups (37) and is in complete or high LD with several SNPs also associated with asthma in the literature, such as rs11674302, rs13431828 (35), rs1041973 (233) and rs13408661 (234).

As 1) Torgerson *et al.* did not indicate the allele or the direction of the rs3771180 association; and 2) the minor alleles of rs13431828 and rs10173081, both SNPs in complete LD with rs3771180, were associated with protection from asthma (35, 37), I used the surrogate SNPs information to assume that the association of rs3771180 with asthma is a protective one conferred by the A allele. I hypothesized that this SNP is functional by leading to transcription alteration by binding a regulatory protein in an allele-specific manner.

EMSA experiments did not support this hypothesis; although there were bands indicating the formation of DNA/protein complexes they seemed to be non-specific as they were not competed out by cold oligonucleotides. I used nuclear extract from the HUVEC cell line as a control as it was shown in the ENCODE data (using the HaploReg web resource) that this SNP binds nuclear proteins in HUVEC by Chip-Seq. My result does not necessarily mean that the rs3771180 SNP site does not participate in binding of a regulatory protein or complex; if this SNP is indeed regulatory, it is possible its function requires some further DNA region not included in the 23 nucleotides of the EMSA oligonucleotide. ENCODE ChIP-Seq data indicate rs3771180 binds STAT3, a transcription activator as well as cFos and cJun; these proteins are part of the AP-1 complex which usually binds enhancer elements resulting in upregulation of gene expression. Additionally, ENCODE data show that rs3771180 is located in a DNAse hypersensitive region in 30 cell types including A549 and other lung cell lines, and exhibits enhancer histone marks in 4 of them (endothelial cells, mammary epithelial cells, erythroleukemia cells and epidermal keratinocytes).

Luciferase reporter gene assays showed a trend of decreased transcription by the plasmid containing the A allele compared with the C allele plasmid although it did not reach statistical significance (Figure 4-20). That is in agreement with the mRNA data from the lung dataset analysis, where I showed that A allele expression was significantly lower than the C allele for the probes of the soluble isoform but not the receptor isoform (Figure 4-18). It is also consistent with asthma protection associated with the C allele, as higher soluble IL1RL1 would be available to neutralize IL-33 and prevent its signaling through the IL1RL1 receptor.

In summary, I found that the receptor IL1RL1 expression was lower and less variable than that of the soluble isoform. Most of its variation in the lung expression dataset was due a

batch effect. My candidate SNPs and other variables explained only a small amount of the variability. I found several significant associations of clinical variables with the soluble isoform but their cumulative effects did not improve greatly on the amount of expression variability explained by my nine candidate SNPs. However, 66% of the variability has not been accounted for in my statistical modeling.

Protein levels of soluble IL1RL1 were inversely associated with asthma status, which is in agreement with my hypothesis. The two significant pQTLs were also eQTLs in the same direction in the lung and represented the two highest signals (lowest p values) in the lung eQTL analysis.

I performed functional analysis of two *IL1RL1* SNPs. rs1420101 was significantly associated with soluble *IL1RL1* expression. My functional analysis of this SNP failed to show allele-specific differential splicing of *IL1RL1* mRNA. rs3771180 was significantly associated with soluble *IL1RL1* expression. Using EMSA and reporter gene assays, I found no significant allele-specific functional effect of this SNP on expression.

Chapter 5: Functional follow-up of *TSLP* polymorphism rs1837253 and gene expression analysis

5.1 Overview

Thymic stromal lymphopoietin is an important protein in asthma pathophysiology and has been called a master switch of allergic inflammation (418). TSLP is expressed by epithelial cells upon external stimuli such as injury, virus infection, certain microbial products, allergen or cigarette smoke exposure (419, 420) and plays an important role in promoting and establishing a Th2-inflammatory environment in the lung (296, 329, 421).

The *TSLP* gene is located on chromosome 5q22.1, is composed of 5 exons and encodes 2 transcript variants: *TSLP* isoform 1 (long) and isoform 2 (short). rs1837253 is a SNP located 3.9 kb 5′ of *TSLP*. First reported associated with asthma and AHR, a related phenotype (107), it was then replicated at genome-wide statistical significance levels in major GWAS (35, 37, 283) and constitutes one of the most consistent asthma-associated genes. rs1837253 is not in LD with any other SNP in European or African populations (r^2 >0.2). It is in moderate LD with a group of 3 SNPs in *TSLP* with r^2=0.44 in the 1000 Genomes Asian superpopulation and it is also in weak LD (r^2=0.24) with a different group of 3 SNPs in *TSLP* in the ad-mixed American superpopulation.

In this chapter, I discuss the following: 1) the results of the most stable housekeeping gene for lung gene expression qPCR studies, 2) gene expression data by genotype and by allele performed by qPCR, 3) gene expression analysis in relation to clinical phenotypes and rs1837253 genotypes using the lung expression dataset and finally 4) functional analysis of *TSLP* SNP rs1837253 (EMSA and Luciferase reporter gene assays).

The objective of this chapter is to test my hypothesis that rs1837253 is a functional SNP, which regulates *TSLP* gene expression; and that rs1837253-T allele will lead to the differential binding of a regulatory protein and subsequently to decreased gene expression.

5.2 Results

5.2.1 Gene expression studies

5.2.1.1 House-Keeping gene study

I used three software packages in order to identify the most stable expression amongst a list of twelve candidates for house-keeping genes. I ranked the twelve genes using the outputs from each software package: M score for geNorm, stability value for NormFinder and coefficient of correlation for BestKeeper. The rankings from geNorm and NormFinder were in agreement for all but the 11^{th} and 12^{th} positions, which were reversed; BestKeeper however produced a ranking which was only partially concordant with the other two. Table 5-1 contains the rankings' details; numbers in red are discordant rankings between software packages. There are two columns for the BestKeeper results because the program handles a maximum of 10 genes to compare stability at one time; therefore I performed the analysis of all 12 genes twice, each time removing two low-ranking genes. Results from both analyses are shown as the columns labeled BestKeeper 1 and 2 in the table.

	geNorm		NormFinder		BestKeeper1		BestKeeper2	
Gene	M score	Rank	Stability	Rank	r	Rank	r	Rank
PPIA	1.105	1	0.115	1	0.728	8	0.748	8
RPLP0	1.196	2	0.353	2	0.941	2	0.957	1
GNB2L1	1.224	3	0.447	3	0.898	6	0.932	3
PGK1	1.265	4	0.478	4	0.899	5	0.927	4
ACTB	1.349	5	0.58	5	0.955	1	0.956	2
TFRC	1.379	6	0.646	6	0.906	4	0.925	5
RNR1	1.463	7	0.715	7	0.922	3	0.925	5
B2M	1.608	8	0.838	8	0.801	7	0.735	9
GUSB	1.657	9	0.895	9	0.658	9	-	-
TBP	1.779	10	1.034	11	-	-	0.869	7
HPRT1	1.808	11	1.032	10	-	-	0.565	10
GAPDH	2.081	12	1.293	12	0.627	10	-	-

Table 5-1 House-keeping stability ranking results using three algorithms

I decided to select the gene *RPLP0* as the best house-keeping to use for qPCR assays in the lung samples because it was ranked second by all three algorithms.

5.2.1.2 qPCR assays

I hypothesized that the *TSLP* SNP rs1837253 is a functional SNP. Due to its location in an intergenic region, I aimed to test the hypothesis that it is a regulatory SNP by investigating its potential effect on the TSLP gene expression.

I conducted qPCR experiments in 18 cDNA samples obtained from IIAM lung-derived RNA and 103 cDNA samples obtained from blood-derived RNA (section 2.1.1). I utilized two gene expression TaqMan® assays: one detecting the *TSLP* long isoform transcript and the second detecting both long and short *TSLP* transcripts. As determined in the house-keeping gene optimization, I used *RPLP0* as the reference gene.

No detectable gene expression of the long isoform was observed in any of the blood or lung samples; I observed reasonable expression levels of the short isoform with Ct values averaging 34.6 cycles (range: 29.3 to 40) in the lung samples only. *TSLP* levels in the blood samples were too low to analyze.

Figure 5-1 shows the amplification plot of the blood samples; amplification curves outlined in red were obtained using the assay which anneals to both isoforms; samples did not start amplifying until about the 35th cycle, indicating negligible *TSLP* transcripts present in these samples, the rest of the amplification curves were obtained using the gene expression assay which anneals exclusively to the long isoform; these amplification lines resemble no template control samples, indicating absolutely no detectable long *TSLP* transcripts in these samples.

Nonetheless, the blood-derived DNA samples were used for rs1837253 genotyping and asthma association analysis.

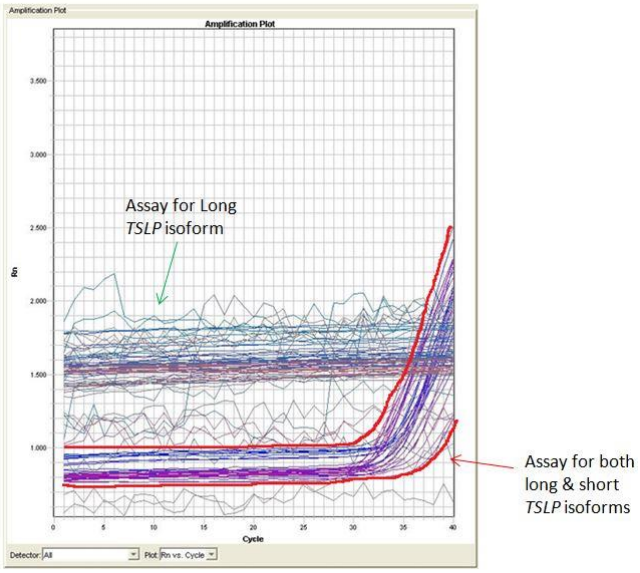

Figure 5-1 Amplification plot of blood cDNA using *TSLP* gene expression assays

Lung and blood DNA samples were genotyped using TaqMan® SNP genotyping assay. Genotyping data were in Hardy-Weinberg equilibrium. A summary is provided in Table 5-2.

	PATH (blood samples)		IIAM (lung samples)	
	N (%) asthmatics	N (%) controls	N asthmatics	N controls
CC	21 (45.7%)	16 (32.0%	4 (57.1%)	9 (69.2%)
CT	18 (39.1%)	28 (56.0%)	3 (42.9%)	4 (30.8%)
TT	7 (15.2%)	6 (12.0%)	0 (0.0%)	0 (0.0%)

Table 5-2 TSLP rs1838253 genotyping summary

Using lung sample qPCR data, there was no significant difference in the expression of the short isoform by asthma status (p=0.24) or by rs1837253 genotype (p=0.89). The genotype comparison was made between major homozygotes and heterozygotes, as none of the 18 samples were homozygotes for the minor allele. There was however a significant difference in expression in females with p=0.009 albeit with a small sample size (3 CT/ 5 CC); with decreased expression in female subjects with the CT genotype compared to the CC major allele homozygote.

In the absence of qPCR data and demographics about the blood samples (section 2.1.1), the only test I could perform for these samples was for association between rs1837253 genotypes and asthma status. I found no significant evidence for such association in these samples (p=0.25).

5.2.1.3 Allele-specific expression

Since rs1837253-T allele is protective from asthma, and given the SNP location in the intergenic region 5' upstream of *TSLP*, I hypothesized that it is a regulatory SNP and the T allele leads to gene expression down-regulation.

As the gene expression analysis by genotype failed to show a significant association, I endeavored to analyze the expression by allele in heterozygous samples. Comparing transcripts carrying different alleles from the same heterozygous sample offers some

advantages over the analysis of samples carrying different genotypes; chiefly, it removes the analysis from any confounding due to environmental causes and the approach is inherently protected from population stratification.

In order to determine if *TSLP* expression was different by alleles, TaqMan® genotyping assays for a *TSLP* marker SNP rs2289276 were performed in order to measure the levels of each allele in cDNA samples heterozygous for both my SNP of interest, rs1837253 and the marker SNP rs2289276.

C/T ratios were calculated for gDNA in order to establish the ratio of 1. cDNA C/T ratios did not depart significantly from 1 (p=0.79) (Figure 5-2).

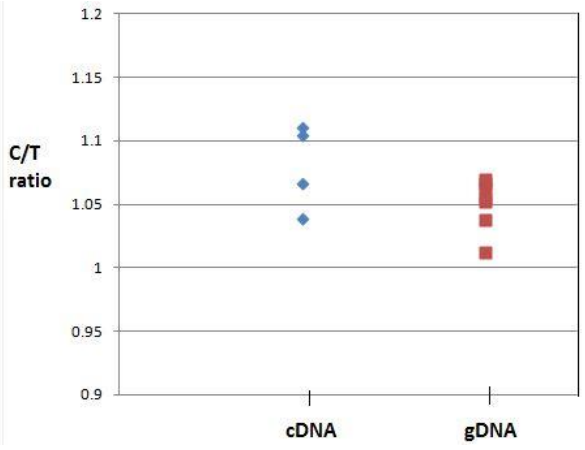

Figure 5-2 Allele-specific expression of *TSLP* cDNA using the rs2289276 coding SNP

5.2.1.4 Association with clinical phenotypes

The *TSLP* gene has been associated with asthma in multiple GWAS; TSLP plays an established role in initiating asthmatic inflammatory cascades by interacting with both innate and adaptive immune cells. However, there is still a knowledge gap in the involvement of the two TSLP short and long isoforms; the specific roles of which are unclear. Thus far, there

have been only two *in vitro* studies from one research group about the inducible nature of the long isoform and the constitutive expression of the short isoform. Their respective roles in experimental and human asthma still remain to be investigated.

In this section, I aimed to investigate the expression of the two isoforms' transcripts using the lung expression dataset. I hypothesized that 1) the expression levels of the two *TSLP* isoforms will be different from each other, 2) the expression of one or both isoforms will be associated with traits relevant to asthma pathogenesis, and 3) that the asthma-associated SNP rs1837253 will modify these associations.

Similarly to the analysis of the *IL1RL1* probes' expression, I used the lung expression dataset. Details of variable selection for this analysis are described in section 2.9.2.2. Briefly, the lung dataset contained 1111 subjects with 97 clinical and demographic variables. There was considerable missingness in most of the variables but only subjects with complete data (expression, genotypes and phenotypic and demographic variables) are allowed in the linear models. First, irrelevant variables such as dates and unspecified medication use were removed. Secondly, categorical variables with less than 10 subjects in one category were removed. Lastly, plotting of the number of subjects by the number of variables with complete data for those subjects (Figure 2-4) was performed; I used this plot to ascertain an arbitrary cutoff (45 variables) in order to simultaneously maximize the number of variables and subjects to include in the analysis.

The final number of subjects with complete data for the 45 variables included in the statistical models was 845 subjects. The characteristics of these subjects are delineated in Table 5-3; a summary of the full list of variables is included as appendix 6.3B.1.

Variables	Mean (± SE) or N (%)
Age	59.1 ± 0.48
BMI	25.22 ± 0.18
Pack year	35.95 ± 1.02
Predicted FEV1	2.95 ± 0.02
Predicted FVC	3.85 ± 0.03
Sex	M: 452 (54.1%)
Smoking status	Current: 170 (20.3%), Ex: 527 (63.0%), No: 139 (16.6%)
Weight	70.89 ± 0.57

Table 5-3 Characteristics of subjects included in *TSLP* expression analysis

In the lung expression dataset, *TSLP* expression was assessed using two probes: X100122252_TGI_at for the long *TSLP* and X100142361_TGI_at for the short *TSLP* isoform. The level of the short *TSLP* expression was considerably greater than that of the long isoform and their expression levels were poorly correlated with r^2 = 15% (Figure 5-3).

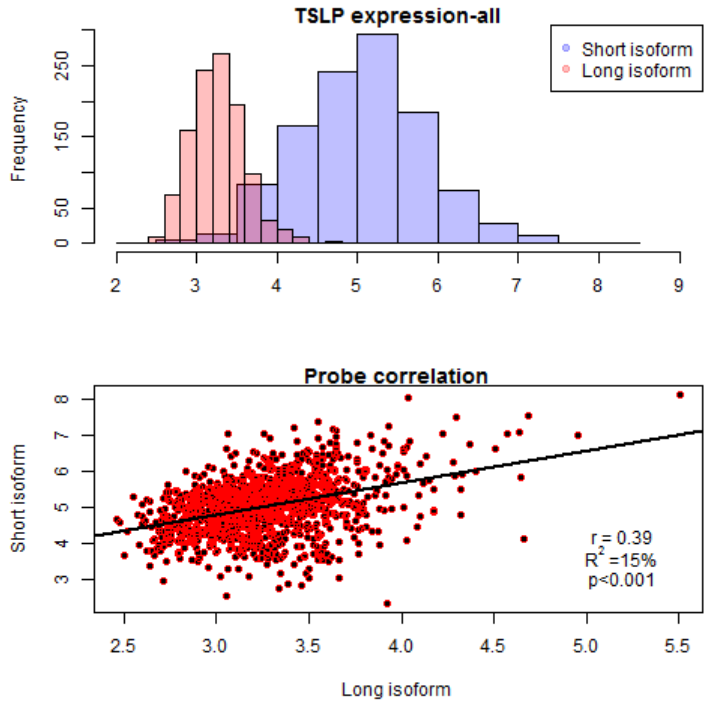

Figure 5-3 *TSLP* probes expression and correlation in the lung dataset.

The probe for long *TSLP* isoform is X100122252_TGI_at; the probe for short *TSLP* isoform is X100142361_TGI_at.

The lung samples were collected at three university sites: UBC, Laval and Groningen. Of note is that the centre variable was very highly confounding in the analysis; Figure **5-4** shows the difference of *TSLP* levels by study centre.

For each probe, the differences between the gene expression in each site were highly statistically significant. The following are the p values for the long *TSLP* probe: Laval vs. Groningen: $p=2.1\times10^{-14}$, UBC vs. Groningen: p=0.004; for the short *TSLP* isoform, Laval vs. Groningen: p=0.003, UBC vs. Groningen: $p<2\times10^{-16}$.

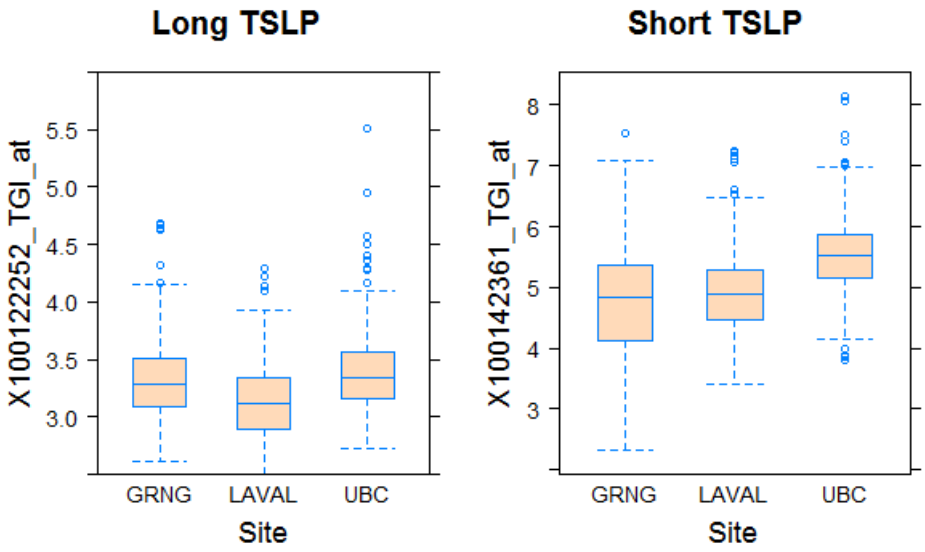

Figure 5-4 Batch effect in *TSLP* probes expression data

The X axis is the study centre: UBC: University of British Columbia, LAVAL: Université Laval, GRNG: University of Groningen.

For each gene expression probe (short and long), I investigated the relationship of the *TSLP* SNP rs1837253 with expression (Figure 5-5). The SNP was modestly associated with expression but accounted for little variability (0.5%).

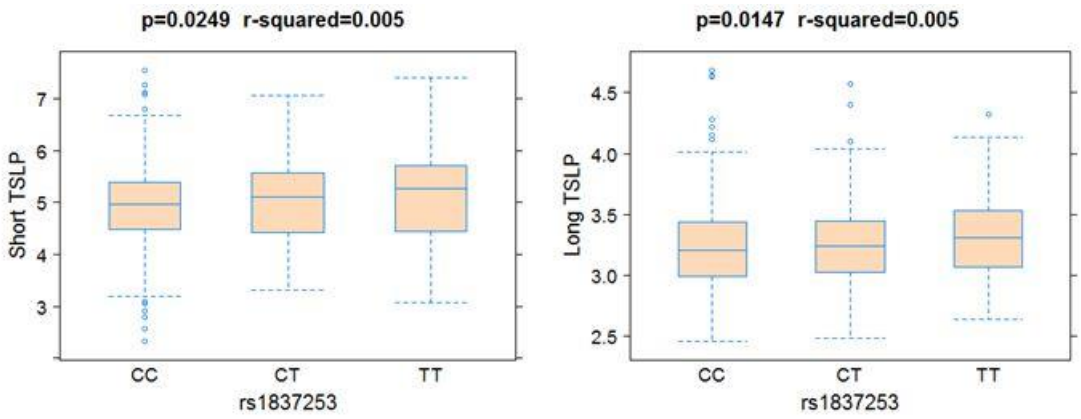

Figure 5-5 eQTL plot of rs1837253

Adding the batch, age and sex covariates to account for their potentially confounding effect, the SNP was still significantly associated with expression and explained 14% and 9% of the variability of short and long isoform expression, respectively (Figure 5-6).

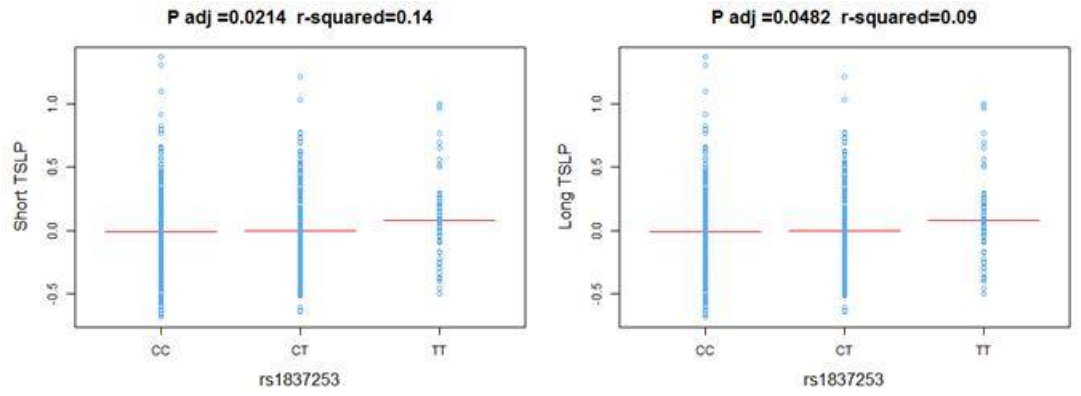

Figure 5-6 eQTL plot of rs1837253 including covariates

For each gene expression probe (short and long), I ran a multiple linear regression model where the response was gene expression and the predictors were 1) rs1837253 genotype, 2) the 45 clinical and demographical variables, 3) batch to account for its

confounding effect, and 4) interaction terms for each clinical and demographical variable with the SNP genotypes

Details of the statistical modeling are provided in section 2.9.2.2. P values obtained were corrected by multiplying by 2 for the two tests (one per probe). All nominally significant results (α threshold = 0.05) of these two analyses are shown in Table 5-4.

	Long *TSLP* isoform			Short *TSLP* isoform		
Term	Estimate	P	P_{corr}	Estimate	P	P_{corr}
Antibiotics	-0.1328	0.0178	**0.0356**	-0.2980	9.0E-04	**0.0018**
Antiplatelet : rs1837253 TT.CT				-0.3107	0.0243	**0.0486**
Bisphosphonates : rs1837253 TT.CT				0.3932	0.0242	**0.0484**
Insulin	0.1658	0.0068	**0.0136**			
NSAIDS				-0.2774	0.0251	0.0503
Oral steroids				0.2908	0.0205	**0.0410**
Oral steroids : rs1837253 TT.CT	-0.1530	0.0330	0.06606	-0.4423	0.0079	**0.0159**
Predicted FEV1	-0.8968	0.0263	0.0526			
Predicted FVC	0.6935	0.0306	0.0612			
Sex	-0.1090	0.0446	0.0893			
Short acting inhaled beta agonists				-0.1557	0.0225	**0.0449**
Smoking status : rs1837253 TT.CT				0.3440	0.0137	**0.0273**
Weight	0.0019	0.0254	0.0507	0.0058	1.6E-04	**3.1E-04**
BatchLAVAL	-0.1763	7.02E-07	**1.40E-06**			
BatchUBC				0.6328	1.5E-15	**3.0E-15**

Table 5-4 Results of multiple linear regression models of *TSLP* expression with phenotypes and genotypes
NSAIDS: Non-steroidal anti-inflammatory drugs.

Antibiotic use was significantly associated with *TSLP* gene expression with p_{corr} = 0.0356 and 0.0018 for long and short *TSLP* expression, respectively. Subjects taking antibiotics (90 subjects) had significantly lower short and long *TSLP* transcripts (Figure 5-7).

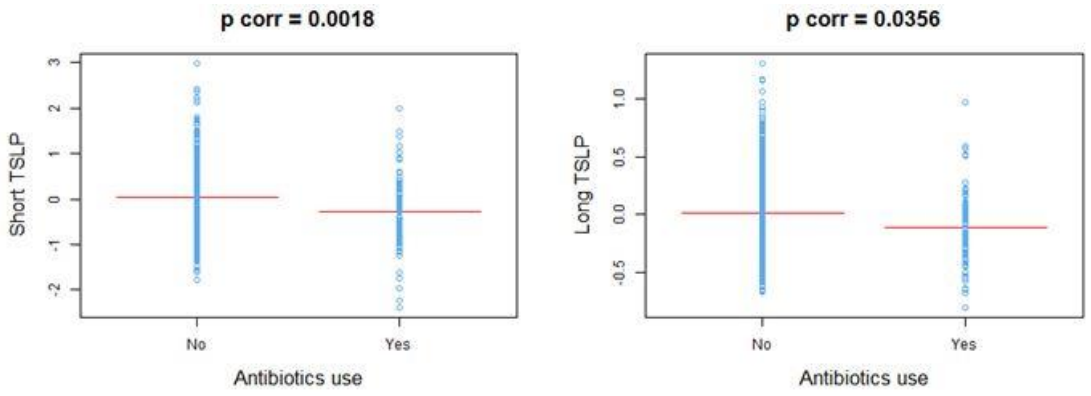

Figure 5-7 Association of antibiotics use with *TSLP* gene expression

Although antiplatelet medication did not associate with *TSLP* gene expression, there was a significant interaction of medication*SNP for short *TSLP* expression with p_{corr} = 0.0486. In subjects taking antiplatelet medication, carrying one or two T alleles (58 subjects) was associated with significantly lower short *TSLP* expression than the CC genotype. rs1837253 also significantly interacted with bisphosphonate use, whereby in subjects taking this medication, carrying TT or CT genotypes (45 subjects) was associated with greater short *TSLP* expression with p_{corr} = 0.0484 (Figure 5-8).

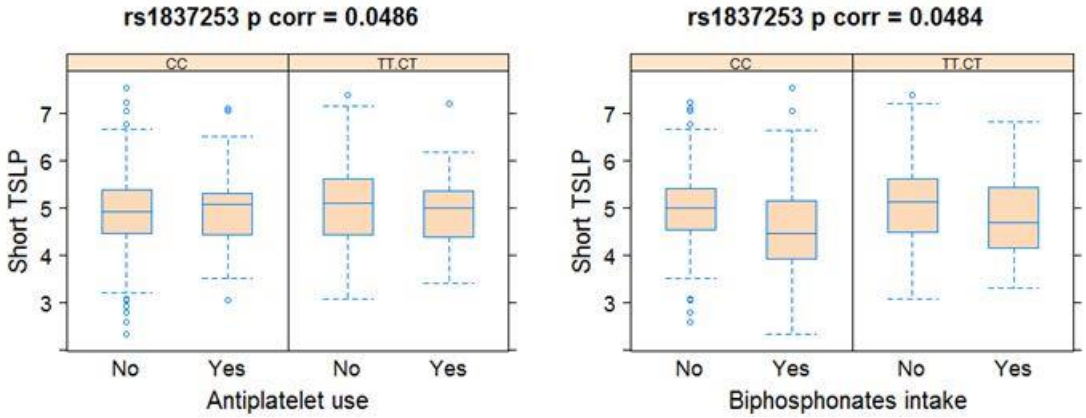

Figure 5-8 Interaction of rs1837253 with antiplatelet and bisphosphonate use on short *TSLP* expression

Insulin use was significantly associated with long *TSLP* expression with p_{corr}=0.0136 (Figure 5-9). In the 31 subjects using insulin, there was greater long *TSLP* expression.

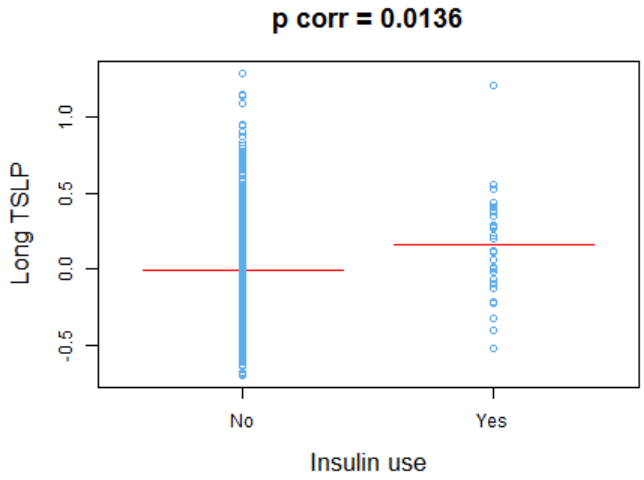

Figure 5-9 Association of insulin use with long *TSLP* expression

Intake of non-steroidal anti-inflammatory drugs (NSAIDs) was marginally associated with lower short *TSLP* expression with p_{uncorr}=0.0251 (Figure 5-10). 58 subjects in the dataset were taking this medication.

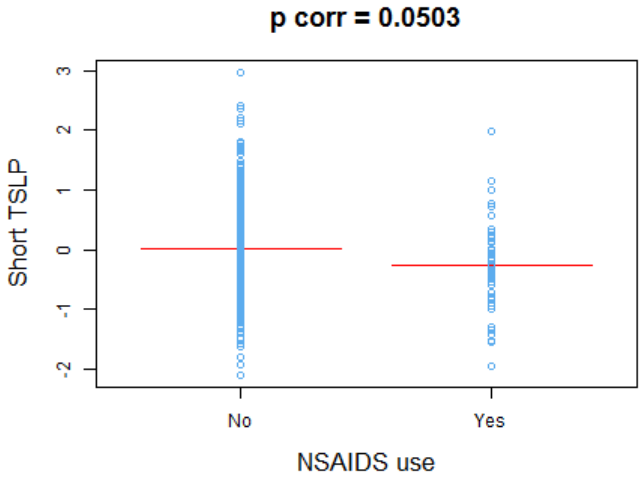

Figure 5-10 Association of NSAIDs use with short *TSLP* expression

NSAIDS: Non-steroidal anti-inflammatory drugs.

Oral steroid intake was significantly associated with short *TSLP* expression with p p_{corr}=0.0410; rs1837253-T allele significantly modified this association with p_{corr}=0.0159 (Figure 5-11). This effect was present in the long isoform but did not reach statistical significance after correction.

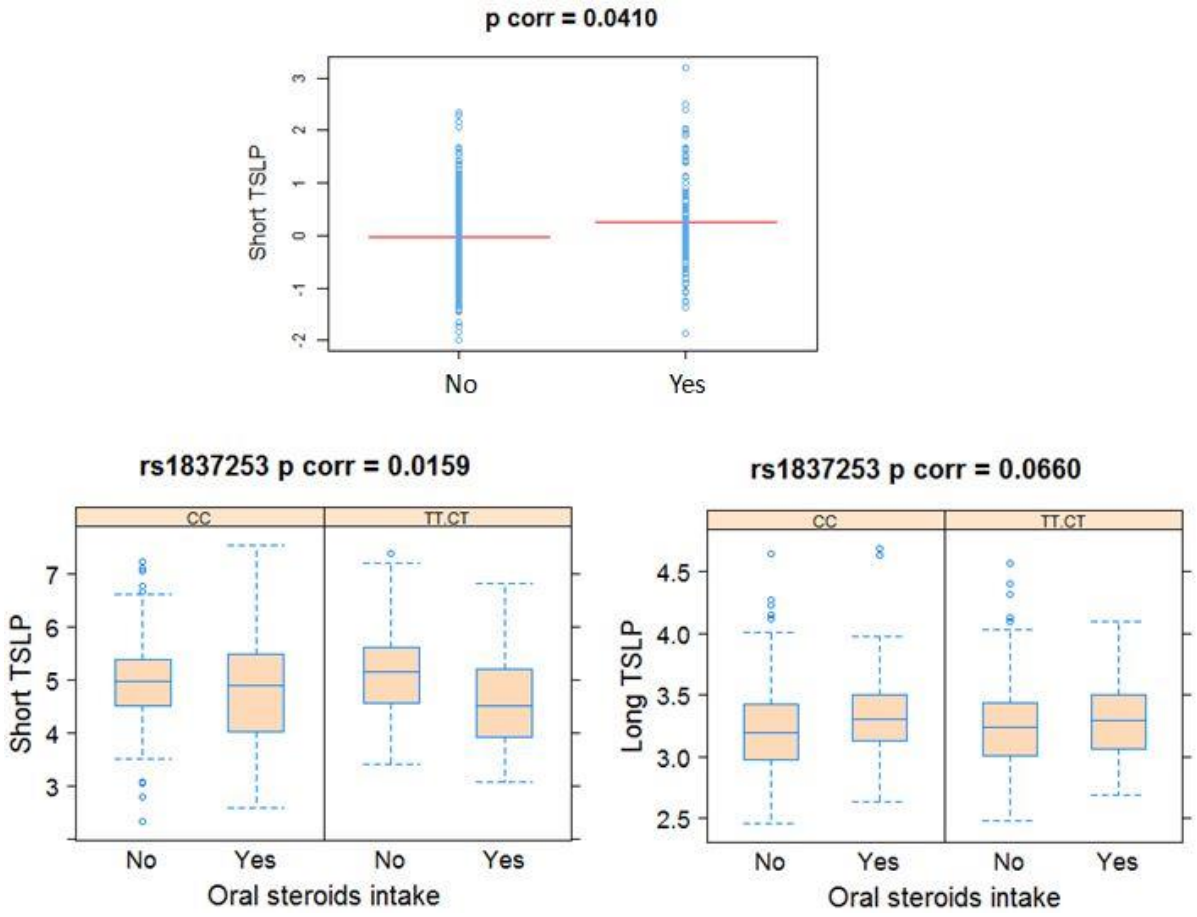

Figure 5-11 Main effect of oral steroid intake and interaction with rs1837253 on *TSLP* expression

There was a marginal association between sex and long *TSLP* with p_{uncorr}=0.0446 (Figure 5-12) whereby males exhibited lower long *TSLP* expression than females.

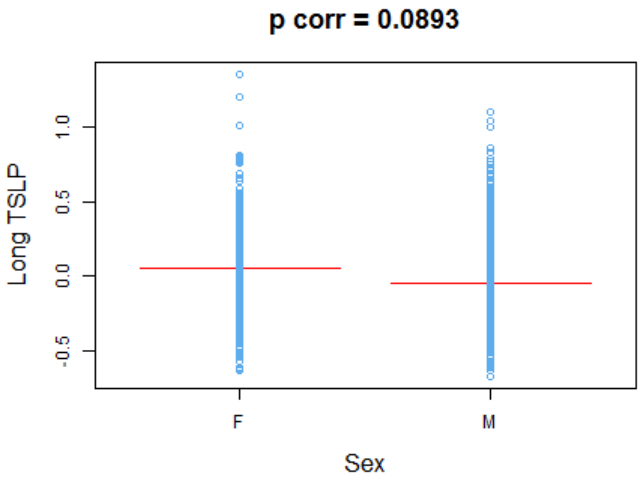

Figure 5-12 Association of sex with long *TSLP* expression

There was also a significant association between short acting inhaled beta agonists and short *TSLP* with p_{corr}=0.0449 (Figure 5-13); 137 subjects in the dataset were taking this medication and exhibited lower short *TSLP* expression.

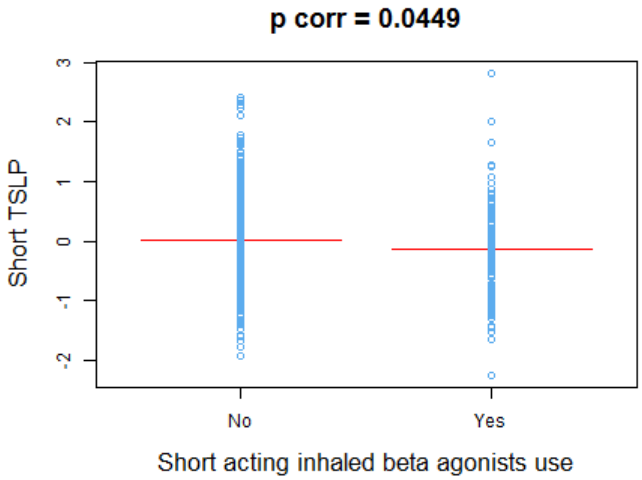

Figure 5-13 Association of short acting beta agonist use with short *TSLP* expression

rs1837253 also significantly interacted with the smoking status with p_{corr}=0.0273. In former smokers, the T allele (238 subjects) was associated with higher short *TSLP* levels than subjects with the CC genotype (Figure 5-14).

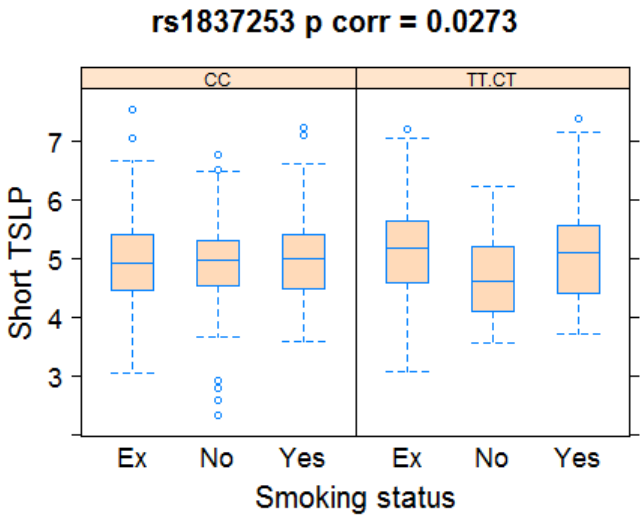

Figure 5-14 Interaction of rs1837253 with smoking on short *TSLP* expression

Ex: former smokers; No: non-smokers; Yes: current smokers.

Finally, short *TSLP* expression was significantly associated with weight with p_{corr}=0.0003; the same trend was observed for long TSLP but it did not reach statistical significance after correction (p_{uncorr}=0.0254) (Figure 5-15).

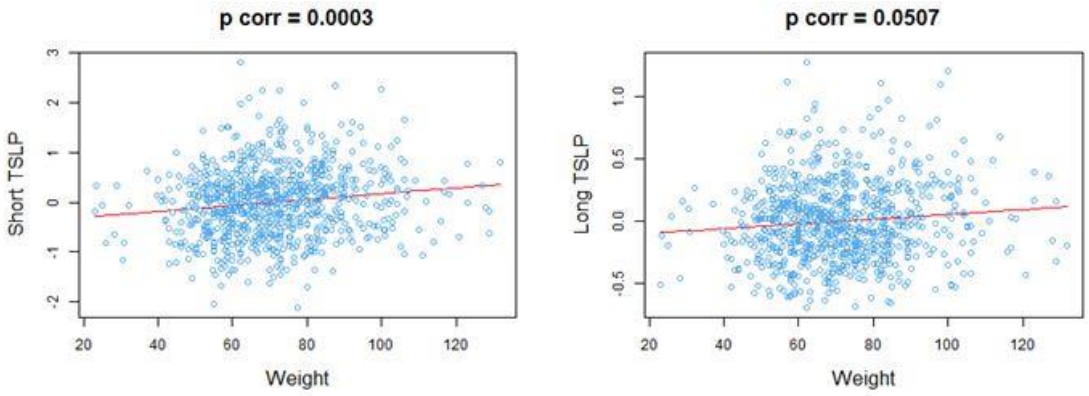

Figure 5-15 Association of weight with *TSLP* isoform expression

5.2.2 Electrophoretic mobility shift assays

The objective of this chapter was to test the hypothesis that rs1837253 is a functional SNP, which regulates *TSLP* gene expression; and that rs1837253-T allele will lead to the differential binding of a regulatory protein and subsequently to altered gene expression.

In order to investigate potential differential regulatory protein(s) binding on or around the rs1837253 SNP site, EMSA was performed using A549 nuclear extract.

The A549 type II alveolar epithelial cell line has been shown to be a convenient and relevant model to study immunotoxicity of the lung epithelium (408-410) and was thus established as a good model for *in vitro* studies of asthma inflammation.

Nuclear protein binding to the rs1837253-T allele oligonucleotide was considerably greater than to the C-allele oligonucleotide. A representative gel is shown in Figure 5-16.

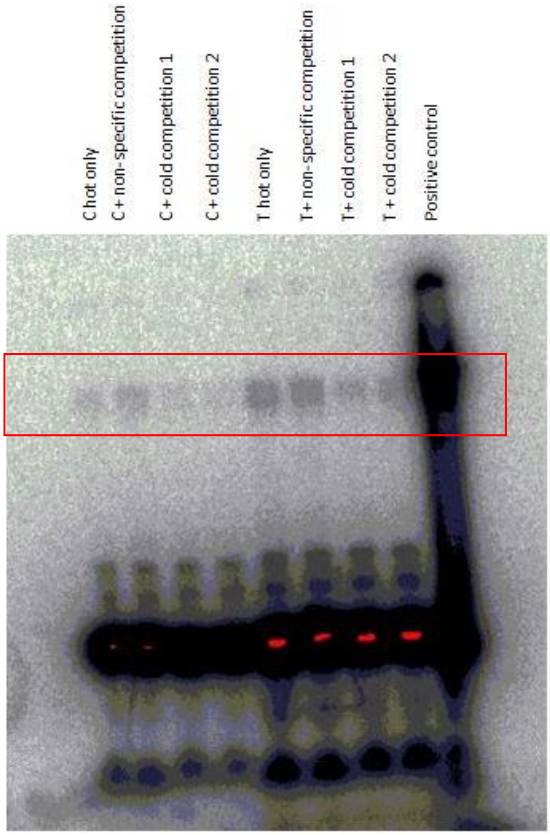

Figure 5-16 rs1837253 EMSA with cold competition

The first four lanes are for the C allele oligonucleotide. The second set of four lanes is for the T allele oligonucleotide. For each allele, the first lane: nuclear extract + hot oligonucleotide; the second lane: nuclear extract + hot oligonucleotide + cold oligonucleotide at 50 times; the third lane: nuclear extract + hot oligonucleotide + cold oligonucleotide at 100 times.

In silico analysis predicted allelic differential binding of rs1837253-T to the glucocorticoid receptor (GR) and to Forkhead box F2 (FOXF2). In EMSA experiments, The T allele oligonucleotide complex seemed be obliterated by incubation with consensus sequence oligonucleotides of the glucocorticoid response element (GRE) and FoxF2 (Figure 5-17) but it was not specific to these candidate competitor oligonucleotides since one of two non-specific cold competitors also resulted in the same effect. I was unable to show specific competition of the SNP complex with candidate response elements oligonucleotides.

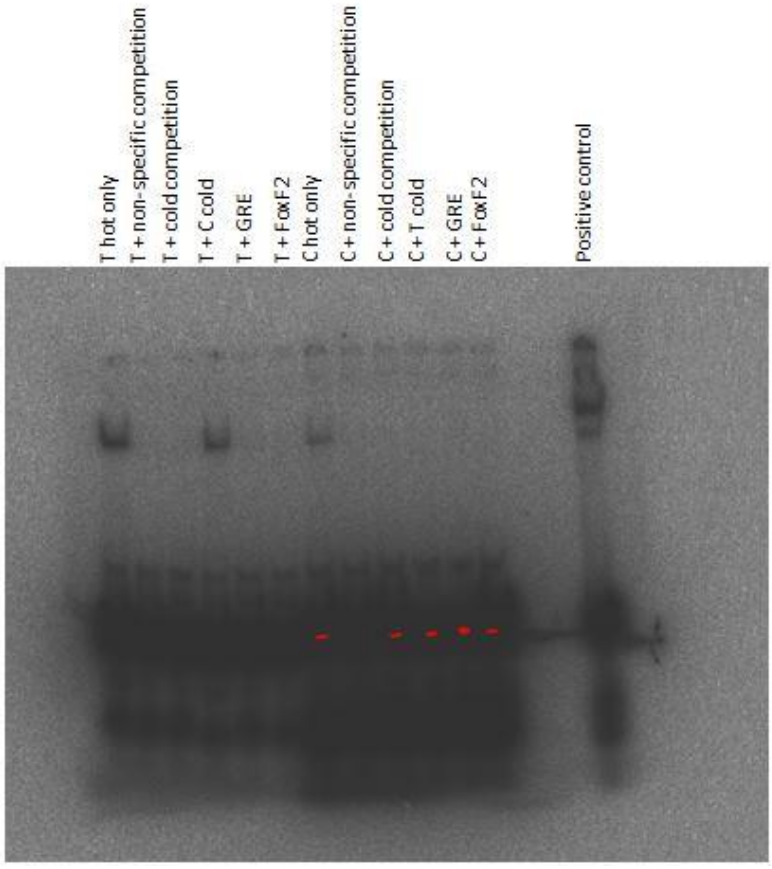

Figure 5-17 rs1837253 EMSA with unsuccessful cold competition

5.2.1 Luciferase reporter gene assays

To further accumulate evidence that the asthma associated rs1837253 is a regulatory SNP, I tested the following hypotheses: 1) a plasmid containing the rs1837253-T allele will lead to decreased luciferase reporter activity as compared with the C allele plasmid, 2) treatments of the cells with poly(I:C) and cigarette smoke extract (CSE) will increase luciferase activity, and treatment with Dexamethasone will decrease luciferase activity as compared with untreated cells, and finally that 3) the effect of the treatments will differ by rs1837253 alleles.

I performed Luciferase reporter gene assays using A549 cells transfected with plasmids of each allele. Luciferase activity ratio was ascertained by the Firefly luciferase over the *Renilla* luciferase activities ratio and the signal was background-subtracted using wells of non-transfected cells containing all reagents. The C allele and T allele plasmids had a mean activity ratio of -0.47 and 0.21 respectively.

Luciferase activity from untreated cells was significantly higher ($p=3.6\times10^{-4}$) in the plasmid containing the rs1837253-T allele (N=34 in duplicate, 6 independent experiments) (Figure 5-18).

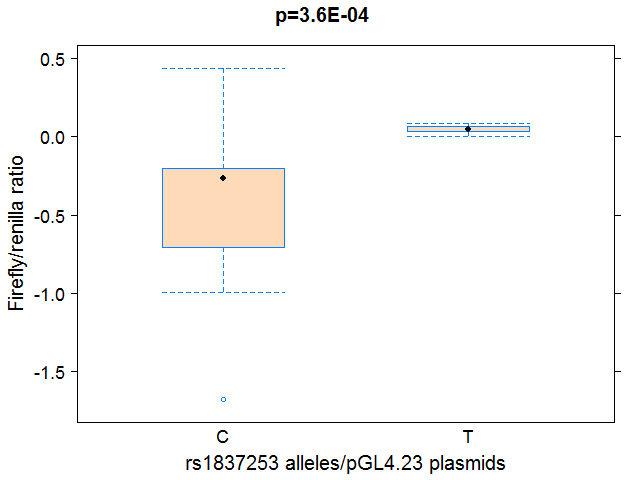

Figure 5-18 Luciferase ratio difference between rs1837253 alleles

I tested the effect of three treatments on A549-transfected cells with plasmids containing allelic inserts: Dexamethasone, Poly(I:C) and cigarette smoke extract. Treatments with increasing concentrations of Dexamethasone (0.01nM to 100nM) did not affect the C-plasmid luciferase activity. There was a modest decrease in luciferase activity after treatment with 10nM Dexamethasone for the T-plasmid activity with p=0.034 from a mean of 0.210 for untreated cells to a mean of 0.004 for treated cells (N=8 in duplicate) (Figure 5-19). However, no other concentration of Dexamethasone showed a significant effect on the T allele.

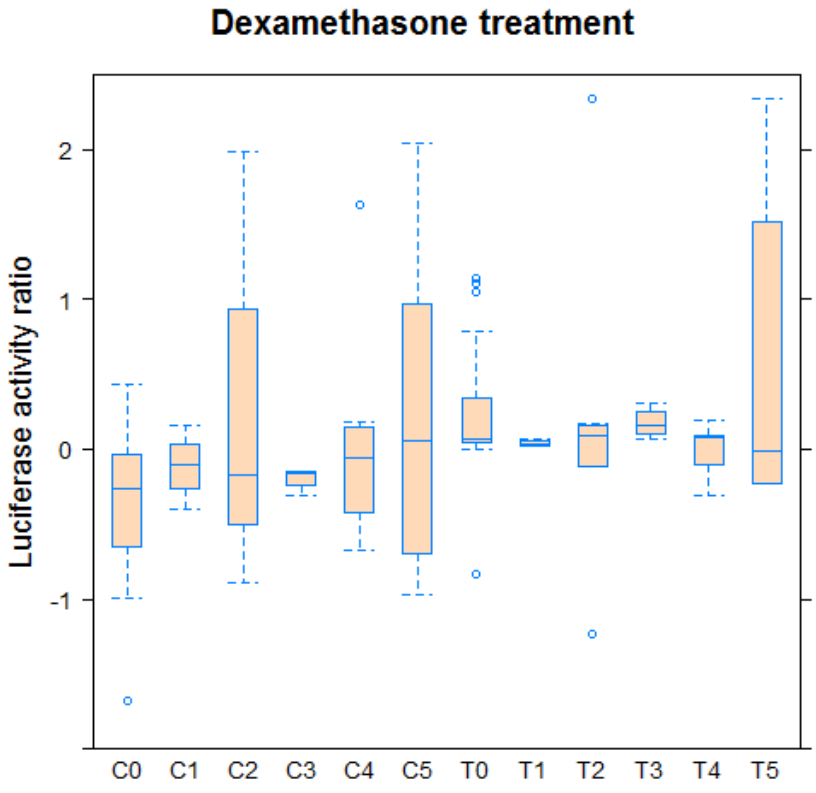

Figure 5-19 Effect of Dexamethasone on Luciferase activity

The x axis represents the treatments by allele: C and T are the rs1837253 allele-specific plasmids; the numbers from 0 to 5 are increasing concentrations of Dexamethasone; 0: no treatment, 1: 0.01 nM, 2: 0.1 nM, 3: 1 nM, 4: 10 nM and 5: 100 nM. Example: C2 is the C allele plasmid treated with 0.1 nM Dexamethasone.

Poly (I:C) treatment significantly increased luciferase activity in cells transfected with the C allele plasmid but not the T allele plasmid. The increase was in a dose-dependent manner (Figure 5-20).

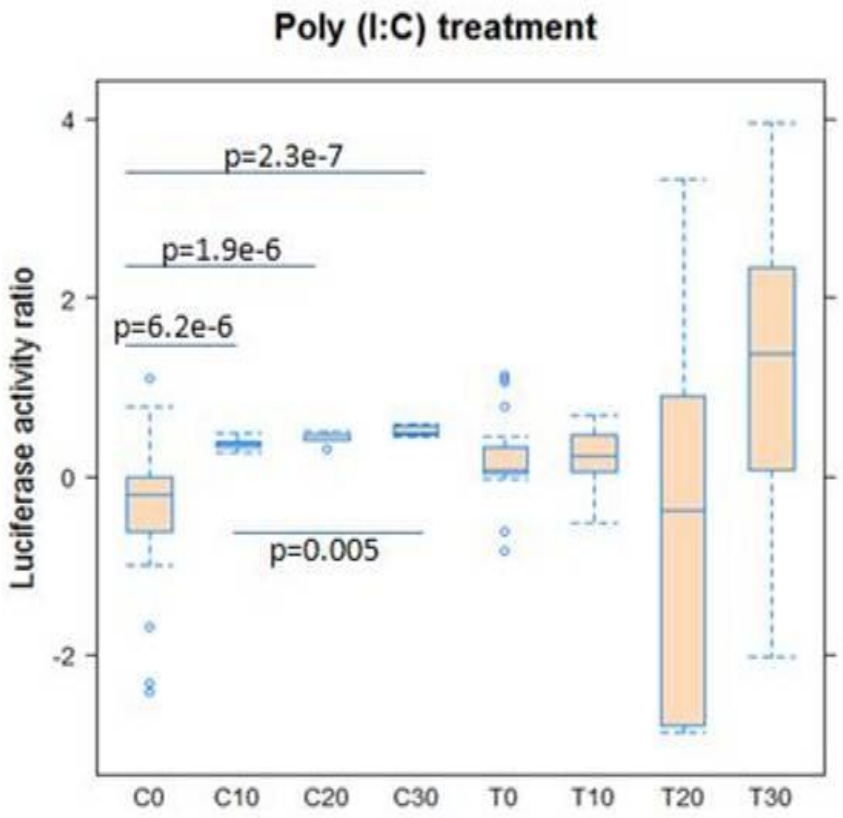

Figure 5-20 Effect of Poly (I:C) on Luciferase activity

The x axis represents the treatments by allele: C and T are the rs1837253 allele-specific plasmids.
Poly (I:C) treatment figure: 0: no treatment; 10: poly (I:C) at 10 μg/ml; 20: poly (I:C) at 20 μg/ml, 30: poly (I:C) at 30 μg/ml.

CSE treatment at 100 fold and 50 fold dilutions also led to an increased luciferase activity in the C allele plasmid transfected cells compared with the untreated cells (Figure 5-21). However, this effect did not seem dose-dependent. Cells transfected with the T allele did not exhibit significant differences in luciferase activity for any of the three concentrations used.

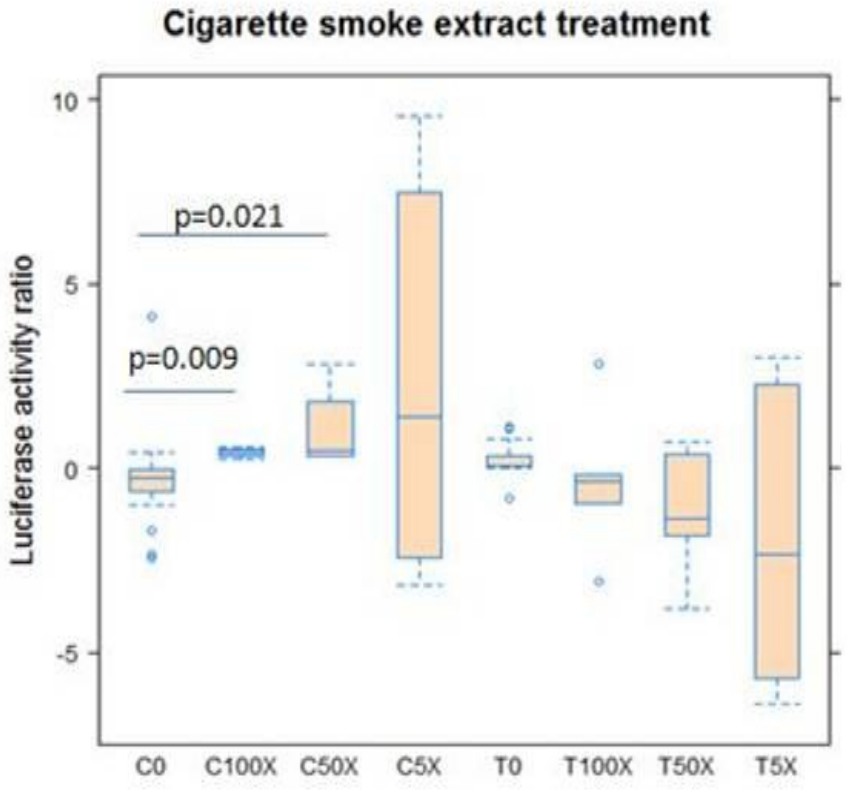

Figure 5-21 Effect of CSE on Luciferase activity

The x axis represents the treatments by allele: C and T are the rs1837253 allele-specific plasmids.
CSE treatment figure: 0: no treatment; 100X: CSE at 100 fold dilution; 50X: CSE at 50 fold dilution; 5X: CSE at 5 fold dilution.

5.3 Discussion

TSLP is a unique asthma gene in that each of its GWAS replications were at the SNP level. rs1837253 is located upstream of *TSLP* and is not in LD (r^2 >0.2) with any other SNP in European or African populations. It is in moderate LD with a group of 3 SNPs in *TSLP* with r^2=0.44 in the 1000 Genome Asian superpopulation (which includes Chinese, Japanese and Vietnamese populations) and it is also in weak LD (r^2=0.24) with a group of 3 SNPs in *TSLP* in the ad-mixed American superpopulation which includes Mexican, Puerto-Rican, Columbian and Peruvian populations. This same SNP was identified in a Japanese population as well as European and African-descent populations.

With the absence of LD in Europeans and Africans and the persistent replication in GWAs at genome-wide significance level, rs1837253 is certainly the most probable cause of the *TSLP* asthma association.

Before proceeding with qPCR assays for *TSLP* gene expression, I performed an analysis of 12 potential house-keeping genes searching for the most stable to utilize in my lung samples. It is interesting that *GAPDH*, a gene very often used as a house-keeping gene in the literature, ranked at the bottom of my stability list using three software packages. I selected *RPLP0*, as the stability of this gene was the second highest ranked and the most concordant using all three methods. These data will be a valuable resource for future studies of lung gene expression, as validated housekeeping genes are an essential prerequisite for accurate quantification of transcript level (422).

Due to the location of rs1837253 in the intergenic region 3.9kb 5' of *TSLP*, I hypothesized it was a functional regulatory SNP. In order to test this hypothesis, I first investigated the SNP effect on gene expression by qPCR.

TSLP encodes two transcripts: the long isoform 1, and the short isoform 2. I used two TaqMan gene expression assays in order to differentiate between the two: one assay containing a specific probe which anneals to the long isoform only and the second assay containing a probe which anneals to a region shared between the two transcripts. In the lung samples, I only observed detectable gene expression of the short isoform and no expression of the long isoform; this is consistent with published *in vitro* data showing very low or undetectable levels of the long isoform in the lung while the short isoform was readily expressed in the lung and all other tissues tested; the long isoform however was strongly inducible by poly(I:C) as compared to the scarce induction of the short isoform (423).

I did not observe any difference in TSLP expression by rs1837253 genotypes in my samples using qPCR but the sample size for this analysis was small.

Allele-specific qPCR showed no significant departure from a ratio of one for the cDNA levels of C over T alleles. Hence, I could not reject the null hypothesis that *TSLP* rs1837253 is not associated with allele-specific differential expression based on my allele-specific qPCR data. It is important to note that the assumption is that the marker SNP rs2289276 used for this experiment does not itself lead to allele-specific differential expression. rs2289276 was previously associated with asthma in candidate gene studies in Asian (109, 424), Turkish (425) and ad-mixed American subjects (110); in Asians, the LD r^2 of this SNP with rs1837253 is 0.44; which is different from European populations where rs1837253 is not in LD with any other SNP (r^2 threshold of 0.2). Harada *et al*. showed that the rs2289276-T allele binds a potential repressor transcription factor, Ap-2α less than the C allele and they showed the T allele is protective from asthma in a Japanese population (109, 423). The reason I selected this SNP as a marker for allele-specific PCR was for technical reasons and the selection preceded the Harada *et al.* publication mentioned above.

My qPCR assays did not show differential *TSLP* expression by genotype; I stratified the genotypic groups by asthma status, and still did not find a significant relationship between the SNP and gene expression; there was however a suggestive significant difference of expression between females and males (regardless of genotype); males exhibited higher short *TSLP* expression.

I investigated this relationship in the lung eQTL data and short *TSLP* expression was not significantly higher in males; in contrast, males showed lower long *TSLP* than females in

this analysis (Figure 5-12). My analysis comprised 836 subjects with gene expression, phenotypes and rs1837253 genotypes.

In the first study reporting the involvement of TSLP in allergic inflammation, Hunninghake *et al*. found an association with lower cockroach-specific IgE levels in Costa Rican girls, and not in boys (108). The same group later reported sex-specific association of rs1837253 and rs2289276 minor alleles with protection from asthma; intriguingly the rs1837253 association was in males whereas that of rs2289276 was in females (110).

In other allergic disease, the rs1837253-T allele was also associated with protection from allergic rhinitis in asthmatic boys only (343) and three *TSLP* SNPs in a Chinese population were associated with chronic rhinosinusitis in a sex-specific manner (426).

In my analysis, *TSLP* expression was indeed significantly associated with sex; males had a lower short transcript expression and higher long transcript expression. This is the first report of differential *TSLP* gene expression by sex and this phenomenon may contribute to the sex-specific effects noted in the previous genetic association studies.

In the lung expression dataset, I observed a marked difference between the expression of the two *TSLP* isoforms and a low correlation suggesting separate roles or regulation; this supports the findings of a published study reporting the inducibility of the long isoform and the constitutive expression of the short isoform regardless of treatments (423).

In fact, in a very recent high impact publication, the authors propose that the reason the long TSLP isoform increases efficiently in response to stimulation by poly (I:C) and inflammatory cytokines and short TSLP only increases modestly in comparison, is because the isoforms have separate roles and regulation. Indeed, they show that the short TSLP

protein does not activate STAT5 in oral keratinocytes, nor does it interfere with the long TSLP signaling and that the short isoform has antimicrobial properties (427).

In this thesis, I described the poly (I:C) induction performed in my reporter gene assays; I had hypothesized that rs1837253 is important for this viral-induced stimulation reported by Bjerkan *et al*. and others. I treated A549 cells transfected with rs1837253 allele-specific plasmids with increasing concentrations of poly (I:C) and hypothesized that this treatment will lead to higher luciferase activity and that this increase will be different by allele. Poly (I:C) treatment did indeed lead to increased luciferase activity in a dose-dependent manner but only for the C allele. This novel result is consistent with my hypothesis. The T allele is the asthma protective allele, and here I have shown that a plasmid containing this allele is not responsive to the poly (I:C) induction seen in an identical plasmid differing only in the one allele. This might have important implications for the role viral infections play in asthma development as well as exacerbations.

An important novel result from my studies is that the rs1837253-T allele results in higher expression of the short *TSLP* isoform in the lung (corrected p=0.0082). This is the first report of a putative functional effect of the rs1837253 polymorphism: a variant that has been consistently associated with asthma and related traits. Although my data do not directly implicate rs1837253, it is highly likely to be the causal SNP for this differential expression because of the lack of LD with any other polymorphism. The direction of the association was surprising, as this allele is protective from asthma and I hypothesized that it would result in diminished *TSLP* production. Nevertheless, in the light of the new data about the potential opposite roles of the TSLP proteins, my data indicate short TSLP might be important for asthma protection and that rs1837253 is important for the regulation of this isoform.

rs1837253 was a modest eQTL in the lung expression data and when adjusted for age, sex and batch confounding factors, accounted for 14% and 9% of the variability of the *TSLP* short and long isoforms expression respectively as ascertained by the models' adjusted R^2.

In order to determine whether greater explanation for the *TSLP* gene expression variability can be obtained by clinical and demographical variables and whether my SNP of interest modifies the relationship of these variables with *TSLP* gene expression, I performed multiple linear regression models using clinical and demographical variables available in the lung expression dataset.

I showed significant negative correlation between short *TSLP* expression and use of antibiotics and short acting inhaled β agonists. The short TSLP protein has been recently shown to have an antimicrobial role (428) and this negative association with antibiotic use might be explained by the medication rendering this role redundant. Similarly, short acting inhaled β agonists are used to relieve asthma symptoms and their negative correlation with short *TSLP* might indicate a lessening of a need of an inflammatory role for this isoform.

Smoking is known to act as an inflammatory stimulus in the lung; here I showed a significant interaction between the *TSLP* SNP rs1837253 and smoking status whereby there was higher short *TSLP* expression in former smokers carrying the T allele compared with subjects with the CC genotype. I also found a significant positive association between oral steroid intake and short *TSLP* expression; this relationship was significantly modified by the presence of the rs1837253-T allele. Under the assumption that oral steroid use is a surrogate for lung inflammation in those subjects, this association suggests the asthma-protective T allele (but not the C allele) leads to higher short TSLP in order to participate in anti-inflammatory efforts. Significant interactions of rs1837253 were also observed for use of

antiplatelet and bisphosphonate drugs, influencing the short *TSLP* transcript expression. No data exist thus far to elucidate the potential cause for these two associations; TSLP has not been associated with osteoporosis or coagulation pathways. Therefore, these data need to be replicated elsewhere and investigated.

The short *TSLP* isoform expression was also significantly correlated with higher weight. This is the first report of this finding in the lung. Differential expression of *TSLP* was only ever reported by one research group in adipose tissue of obese subjects with metabolic syndrome (a condition constituted of metabolic and inflammatory phenotypes observed in obese individuals) (429). Low-grade chronic inflammation has been associated with obesity in many studies (430, 431) and there is some evidence of a possible link between asthma and obesity (432, 433), although a study aiming to detect common associated genes between asthma and obesity using GWAS data failed to find significant association; however, gene-based analysis of GWAS aiming to detect common associated genes between asthma and obesity, led to a suggestive association of the *IL1RL1* SNP rs13431828 with obesity (434). My finding builds on the evidence towards a link between asthmatic inflammation and obesity as well as provides a possible mechanism to the epidemiologic observation of obesity's association with later onset severe asthma.

Despite the numerous associations of *TSLP* gene expression with the variables, the adjusted R^2 only increased by 9% and 7% for the short and long isoforms, respectively. There remains considerable unexplained variability not accounted for by the rs1837253 and the clinical and demographical variables in the models (77% and 84% for short and long isoforms, respectively).

Generally, there were less significant results for the *TSLP* long isoform compared with the short isoform; this might be due to either its low levels in the lung samples or it might be a true negative result given the emergent role of the short TSLP protein as having a different role than the long isoform; the samples I used were lungs of predominantly non-asthmatic subjects, and the *TSLP* long isoform has been shown to be very low or undetectable in lung cells or keratinocytes and only increases under stimulating conditions (423, 435). In a recent publication, Takai *et al.* suggested that the long *TSLP* isoform is the predominant transcript responsible for the production of the *TSLP* protein in primary human keratinocytes (428); using the same TaqMan® gene expression assays I have used: one assay for the long isoform and one assay for both isoforms, they showed that under unstimulated conditions, the total *TSLP* transcript levels constituted 50 to 100 times the level of the long isoform and there was no detectable TSLP protein. After stimulation with TLR ligands or with pro-inflammatory cytokines, the long *TSLP* isoform was greatly up-regulated and so was the total *TSLP* transcript with no significant difference between the two levels; there was a marked release of TSLP protein. Bjerkan *et al.* also stimulated keratinocytes with poly (I:C) and pro-inflammatory cytokines and found the same result, i.e. that the expression of the long isoform was increased to a much greater degree than that of the short isoform. Taken together, these results suggest separate roles of the two isoforms. Indeed, Bjerkan et al. showed that signaling of the short isoform was different, as it does not activate STAT5 in DCs whereas the long isoform does.

Although a similar study in airway epithelial cells is needed to answer the same question in the lung, it is likely the expression pattern described in keratinocytes would be relevant to lung cells; it is established that TSLP is mainly produced by the epithelium in the

airway, the skin cells and the gut; TSLP production by keratinocytes is thought to predispose to and certainly predates asthma. This participates in a phenomenon referred to as the atopic march. In atopic march mouse models, keratinocyte-derived TSLP was shown to correlate with skin sensitization and asthma severity (436-438) but there was no distinction between isoforms in those studies.

Using the lung expression dataset, my analysis was underpowered to detect an association of this SNP with asthma. Out of 1104 total subjects, there were 29 asthmatics, among them 3, 12 and 14 subjects carrying the TT, CT and CC genotypes, respectively. I had 32% power to detect a genetic effect of 1.5 and 10.2% power to detect an effect of 1.2; which is in line with genetic effect sizes for SNPs in complex diseases. The odds ratio for the *TSLP* SNP association in GWAS indeed ranged from 1.17 (37) to 1.21 (90).

EMSA assays using C and T oligonucleotides and A549 cell nuclear extract showed a clear difference between the oligonucleotides binding to nuclear protein(s) indicating an allele-specific effect of rs1837253 in the formation of a DNA/protein complex in those lung cells.

However, competition experiments using cold T and C oligonucleotides were inconclusive even after switching from room temperature incubation of probes and nuclear extracts to incubation on ice in order to limit free protein movement and discourage non-specific binding. Additionally, the T allele oligonucleotide complex band obliteration by a GRE consensus sequence oligonucleotide was not consistently replicated.

ENCODE data indicate the presence of two histone enhancer marks in epithelial cells and keratinocytes in the vicinity of rs1837253; H3K27ac histone enhancer mark was present inside a 10.5 kb region starting 285 bp from the SNP site and ending inside the *TSLP* gene;

and H3K4me1, present inside a 15 kb region encompassing the *TSLP* gene and starting 150 bp 3' of the SNP site. H3K4me1 is the monomethylation of the 4th residue (Lysine) from the N-terminus of H3 Histone and is an established epigenetic modification to serve as a marker of enhancers of actively transcribed genes. H3K27ac is the acetylation of the 27th residue (Lysine) of the H3 histone, and together with H3K4me1, marks active enhancers as opposed to inactive ("poised") enhancers which lack the latter epigenetic mark (439, 440). The ENCODE data also show DNAse sensitivity in lung fibroblasts and aortic smooth muscle cells indicating active chromatin in those cells around the SNP site.

Reporter gene assay data strengthened the evidence for the involvement of the T allele in *TSLP* short isoform up-regulation by showing that this allele leads to higher luciferase activity than the C allele plasmid. This result supports my novel finding of a functional role of the rs1837253 polymorphism and provides the first insight into the mechanism underlying the association with *TSLP* gene expression and with asthma.

Dexamethasone is a glucocorticosteroid drug with anti-inflammatory properties, used to treat a number of conditions including asthma. Dexamethasone and other glucocorticoids have very broad immunosuppressive effects (441), which occur after the compounds cross the cell membrane and bind to glucocorticoid receptors, forming a complex which is then able to bind specific DNA response elements on several genes and alter gene transcription. Dexamethasone has been shown to inhibit TSLP production (442, 443).

I hypothesized that the *TSLP* SNP rs1837253 is important for that repression and tested it using luciferase assays. I did not observe any effect on the C allele plasmid-transfected cells but did find a statistically significant decrease in luciferase activity for the T-plasmid activity after treatment with 10nM Dexamethasone. However, no other concentration of

Dexamethasone showed a significant effect on the T allele. This decrease was modest and the higher concentration (100nM) did not show a significant difference compared with the untreated cells; although the mean is lower, there was much variability in the luciferase activity for the 100 nM concentration. Therefore I cannot conclude the effect is true with great confidence.

Viral infections in early life are associated with persistent wheezing and later development of asthma (444) and the majority of asthma exacerbations are preceded by viral infections in children and adults (445, 446). Asthmatic epithelial cells secrete excessive amounts of TSLP after viral infection in comparison to cells from non-asthmatic subjects (298); this is accompanied by lower interferon beta production which could participate in the association of viral infections with increased asthma risk in genetically-predisposed individuals. This viral-induced production of TSLP in epithelial cells has been shown in human and experimental asthma models (447, 448). Poly (I:C) is a synthetic double-stranded RNA designed to mimic viral genetic material and was shown to be a rodent and human TLR3 ligand (449). My data showed an induction of luciferase expression by the C allele, and not by the asthma-protective T allele.

Environmental cigarette smoke *in utero* and in early life has been associated with development of asthma later in life (450). TSLP expression has also been shown to increase in response to stimulation using cigarette smoke *in vivo* (301) and *in vitro* in lung smooth muscle (451). In addition, nasal epithelial cells (co-cultured with dendritic cells) from smokers exhibited greater TSLP production than cells from non-smokers after influenza infection (452).

I hypothesized that my SNP of interest is involved in this cigarette smoke induction of *TSLP* expression; and therefore cells transfected with the allele-specific plasmids were treated with fresh cigarette smoke extract diluted to 3 concentrations. I found that similar to the poly (I:C) treatment, CSE treatment led to an increase in luciferase activity in the C allele plasmid transfected cells compared with the untreated cells but seemingly not in a dose-dependent manner. This effect was not present in the cells transfected with the T allele, although there seemed to be a decreasing trend in that allele. This is in line with my hypothesis

In summary, the T allele led to higher short *TSLP* expression and no significant change in long *TSLP* expression, this allele exhibited differential binding to nuclear protein in EMSA, and led to increased expression compared to the C allele in luciferase reporter gene assays; cells transfected with the C allele plasmid and not T allele were responsive to poly (I:C) and CSE treatments resulting in an expression increase..

My data, together with the new protein data suggesting opposite roles for the two TSLP isoforms (427), indicate a role for the rs1837253 T allele in protection from asthma by regulating the expression of the short TSLP isoform and potentially preventing its stimulation by pro-inflammatory stimuli.

Chapter 6: Conclusion

The objective of this thesis was to identify causal SNPs in two genes, *TSLP* and *IL1RL1*, with very solid genetic and biological evidence of involvement in asthma pathogenesis. Both genes have been identified in several asthma GWAS in multiple populations.

Finding an association of a phenotype with SNPs in a specific gene might be a valid endpoint for a study as this can identify genes involved in disease-causing pathways. However, there is a compelling case to be made for following-up the association studies in order to identify the specific phenotype-causing SNP. Identifying causal SNPs can help uncover the molecular mechanism underlying the involvement of the gene in disease pathogenesis and help decipher association signals from multi-gene loci with extensive LD.

This thesis outlines the first functional analysis of the sole asthma-associated SNP in *TSLP*, rs1837253, which has been replicated in multiple populations; I demonstrated that this SNP is indeed functional, the T allele was associated with increased short *TSLP* expression, and the T allele differentially bound a nuclear protein or complex and led to increased expression in a reporter gene system.

I also report the first usage of generalized linear models to prioritize SNPs for functional follow-up and also its first use to deal with gene expression data. I have identified a viable list of candidate functional SNPs from the asthma-associated *IL1RL1*/*IL18R1* chromosome 2q region; and uncovered a potential novel distal regulatory region ranging from 680 kb to 260 kb from the *IL1RL1* transcription start site, controlling the expression of the *IL1RL1* receptor isoform transcript.

6.1 IL1RL1

6.1.1 Summary of findings

The purpose of the *IL1RL1* work in this thesis was to firstly identify putative asthma-causing functional SNPs, secondly to investigate the gene expression of *IL1RL1* soluble and receptor transcripts isoforms, and protein expression of the soluble isoform as well as the relationship of expression to clinical phenotypes and thirdly to uncover the molecular mechanisms of identified putative functional SNPs.

Using two approaches, a manual annotation and a statistical learning method (Glmnet), I narrowed down 604 associated SNPs (by virtue of gene expression, genetic association or LD with associated SNPs) to ten and eleven SNPs respectively. I uncovered three separate potentially important regulatory regions: one shared between the soluble and receptor *IL1RL1* isoforms and one for each of the isoforms. I also suggest the presence of an important regulatory region controlling the gene expression of the receptor *IL1RL1* at a distal location as far as 260 kb from the TSS of its target gene *IL1RL1* but not controlling the soluble form.

I hypothesized that functional alleles of SNPs associated with asthma risk in the *IL1RL1* gene lead to decreased expression of the soluble isoform of *IL1RL1* and/or increased expression of the receptor form. I found no statistically significant associations of the receptor *IL1RL1* expression of the receptor in a subset of the lung expression dataset of about 300 subjects with complete phenotypic data. In the entire lung expression dataset, receptor *IL1RL1* expression was lower and less variable than that of the soluble isoform. The soluble *IL1RL1* expression was negatively correlated with asthma and antihistamine intake; and there was a positive correlation between smoking and soluble *IL1RL1* expression; this correlation

was significantly modified by four SNPs. Soluble *IL1RL1* was also significantly associated with BMI.

I performed functional analysis of the rs1420101 and rs3771180 polymorphisms. rs1420101 was significantly associated with soluble *IL1RL1* expression; although a strong functional candidate, it did not show differential splicing of *IL1RL1* mRNA as assessed by a PCR-based method. rs3771180 was significantly associated with soluble *IL1RL1* expression. The SNP showed DNA/protein complex formation in EMSA experiments but no clear difference between alleles, and luciferase reporter gene assays corroborated this negative finding.

6.1.2 Implications of the work

This thesis is the first report of the usage of generalized linear models using gene expression data with SNPs as predictors. Exploring the *IL1RL1* region using gene expression data and intersecting it with asthma association findings in the literature allowed me to identify potentially functional SNPs in this region of extensive LD. eQTL data showed the soluble isoform is regulated by a multitude of independent groups of SNPs in the lung.

Using multiple linear models, I found that the receptor *IL1RL1* expression did not vary considerably as compared with the soluble isoform's expression. Subjects diagnosed with conditions with inflammatory symptoms such as α-1 antitrypsin deficiency and CF showed significantly higher receptor expression and soluble *IL1RL1* was significantly greater in subjects taking oral steroids. Despite these associations, the variability of the transcripts' expression was mostly accounted for by genetic variation.

The SNPs I selected for functional analysis were not analyzed elsewhere. Therefore, efforts towards identifying asthma-causing polymorphisms will include different SNPs than the ones for which I am reporting negative findings.

I aimed to select good putative functional SNPs that could be causing the asthma association as well as perform *in vitro* studies to verify that my candidate SNPs were functional and uncover their mechanism of regulating the *IL1RL1* gene. I hypothesized that alleles associated with asthma risk would lead to decreased gene expression of soluble *IL1RL1* accompanied with no change or increase of the receptor *IL1RL1* expression and that alleles associated with protection from asthma would lead to the opposite effect. I utilized the results of my manual SNP prioritization for selection of SNPs for functional analysis. I selected rs1420101 and rs3771180 as both were excellent candidates: eQTLs, associated with asthma and predicted to have functional potential. I hypothesized that rs1420101 leads to defects in *IL1RL1* splicing as it was present in an intron shared between the two *IL1RL1* isoforms. I designed a PCR-based assay to detect size differences between PCR products from cDNA amplified with a forward primer and different reverse primers annealing to successive exons. However, rs1420101 did not show evidence supporting my hypothesis.

rs3771180 is a proximal promoter SNP with strong evidence of functional potential according to my manual annotation; it was highly associated with soluble IL1RL1 expression in the lung but failed to show differential binding to nuclear proteins by EMSA and similarly did not show a significant difference in luciferase expression by allele. Therefore, it is unlikely that either rs1420101 or rs3771180 is a causal polymorphism for the associations with asthma and these results narrow down the search for the functional SNPs in this important asthma locus.

6.1.3 Limitations and future directions

A limitation to the expression studies in this thesis is that the lung gene expression dataset was derived from subjects undergoing lung resection following cancer diagnosis; although the region of the lung sampled for the study were distant from the tumor site, this was not an asthma cohort. I used this dataset under the assumption that SNP effects on gene expression would be the same in asthma patients although allele frequencies may be different in the two diseases. A similar dataset from asthmatics and controls would have been ideal but was not available. Additionally, there was much sparsity in the phenotypic data, which decreased the number of samples I could use for gene expression/phenotype association from 1104 to 845 subjects. Variable selection for the linear models was performed in an unbiased manner, which resulted in the removal of several relevant variables such as asthma. In addition, several variables had too few subjects in the different categories and this rendered some comparisons underpowered because of the stratification by genotype that is required for studying interactions.

Another limitation in the IL1RL1 work is that, although I found no evidence of a differential splicing pattern by rs1420101 genotypes or by asthma status using the PCR-based splicing assay, the PCR-based technique has relatively low resolution to observe smaller changes in size. A more robust approach would have to be employed in order to definitively state that this SNP has no effect on splicing. For example, the rs1420101 SNP splicing data in this thesis should be replicated using mini-gene systems in order to verify the negative finding and provide a more definitive answer.

Moreover, the SNP selection for functional analysis was performed based on the manual annotation because the Glmnet analysis required additional training in statistical learning methods which occurred later during my thesis work.

As future direction, putative functional SNPs from the glmnet analysis should be investigated for functional potential using molecular biology techniques like those described in this thesis; in particular, the identified potential distal regulatory region for receptor *IL1RL1* should be investigated. This could be done by analyzing the individual SNPs in the region using EMSA and reporter gene assays and also by chromosome capture techniques to uncover potential physical interaction of this regulatory region with *IL1RL1* promoters.

Blood and lung samples from a large cohort (500 or more) of asthmatics and controls with genotypes, phenotypes and gene expression data would be instrumental in further investigating the differential expression of the *IL1RL1* isoforms and distinguishing between their respective regulatory SNPs. This would also allow comparison of tissue-specific differences and may shed more light on the differential *IL1RL1* promoters' usage. Thus far, there have been only a limited number of published *in vitro* studies investigating the topic; in rats cardiac myocytes, the proximal promoter is used for the two isoforms (205); whereas in mice mast cells, both isoforms are produced from the distal promoter (453). Human *in vitro* studies also indicate complex regulation. For example, data suggest that in Th2 cells, the receptor transcription is controlled by the distal promoter (394) and in fibroblasts the soluble isoform is controlled by the proximal promoter (395). There are also data from a leukemia cell line UT-7, where both promoters are used for the transcription of the two isoforms (382).

6.2 TSLP

6.2.1 Summary of findings

My purpose was to investigate the *TSLP* SNP rs1837253, as the T allele of this polymorphism was associated with protection from asthma in several GWAS in different populations. I hypothesized that this allele leads to decreased *TSLP* expression.

qPCR assays in a small number of lung samples showed no detectable long *TSLP*, no differential expression of the short *TSLP* by genotypes or by alleles; although for the allele-specific expression assay, I utilized a coding marker SNP which might result in differential allelic expression so this result was inconclusive.

In the larger lung expression dataset, long *TSLP* was detectable but its levels were lower and of smaller range than the short isoform. Male subjects had lower long *TSLP* expression in the lung; but it did not survive correction. I showed rs1837253 was an eQTL in the lung albeit a modest one, and that it significantly modified the relationship of smoking status to the short *TSLP* expression. Former smokers with the T allele exhibited higher gene expression than the subjects carrying the CC genotype. I showed for the first time a significantly higher long *TSLP* expression in CF subjects as well in subjects with higher weight. Antibiotic use was significantly associated with lower short *TSLP* consistent with its emerging role as an antimicrobial peptide.

Using EMSA, I showed that the T allele of rs1837253 differentially bound a nuclear protein or complex in A549 nuclear extract, but could not show consistent competition of the resulting DNA/protein complex. Reporter gene assays however confirmed a differential effect of the T allele, as a plasmid containing that allele showed higher luciferase expression than the C allele plasmid. Dexamethasone treatment did not significantly affect the C allele

but showed some effect on decreasing the T allele plasmid luciferase activity. Finally, the C allele plasmid-transfected cells had significantly increased luciferase expression after Poly (I:C) and CSE treatments but not the protective T allele.

6.2.2 Implications of the work

To the best of my knowledge, this is the first investigation and demonstration of the functional potential of the asthma-associated *TSLP* SNP rs1837253. In the last few years, the important role of TSLP in initiating, amplifying and maintaining the asthmatic airway inflammation has become clear. TSLP, however, is also involved in homeostasis rendering a direct therapeutic approach challenging. It is therefore crucial to understand the role of the asthma-associated SNP in order to decipher the particular asthma-specific pathway and consequently have the basis to develop specific therapeutics which would hopefully have the least harmful secondary effects on patients. AMG 157, a human TSLP monoclonal antibody reduced allergen-induced asthmatic early and late responses in stable allergic asthmatics in a recent 12-weeks clinical trial (454).

The *TSLP* gene expression analysis detailed in this thesis showed a clear difference between the levels of the two isoforms in the lung as well as a low correlation between them; this certainly warrants future studies into TSLP expression to be designed to differentiate between the isoforms. At present, it is still unclear which isoform is important for asthma as most reports use gene expression assays that do not discriminate between isoforms which could confound the results. A very recent study thoroughly investigated the two TSLP isoforms in oral keratinocytes (427) and the authors elucidate the issue of the TSLP isoforms (at least in keratinocytes) by finding the short TSLP does not activate STAT5 in DCs as the

long TSLP does. Furthermore, the short TSLP isoform does not interfere with the signaling either and thus they were able to differentiate between an inflammatory role for the long TSLP and a role as an antimicrobial peptide for the short TSLP. This study potentially explains my finding that the asthma protective T allele of rs1837253 binds a nuclear protein or complex in a lung cell line and leads to higher expression in a luciferase reporter gene system and is associated with greater short *TSLP* transcript levels in human lung samples. I showed that cells transfected with the C allele plasmid increased luciferase production in response to the pro-inflammatory stimuli used (Poly(I:C) and CSE) but not the T allele. Together with the EMSA finding, this could imply the nuclear protein(s) that bind to the T allele inhibit the signaling necessary for an inflammatory response to the stimuli.

This thesis reports the first interaction of the *TSLP* genome-wide associated SNP genotype with smoking status and oral steroids use on gene expression. Previously published significant interactions of *TSLP* SNPs were: 1) rs2289276 was associated with FEV_1 in asthmatic Turkish children without allergic rhinitis compared to with allergic rhinitis (425); 2) rs3806933 significantly modified the effect of breast-feeding on food sensitization (455); 3) rs1898671 significantly interacted with smoking to influence asthma risk in an American admixed population (456) and finally 4) rs2289276 was significantly associated with protection from asthma in females only in a stratified analysis (110).

6.2.3 Limitations and future directions

The *TSLP* gene follow-up of the asthma association consisted of the functional analysis of a single SNP, rs1837253. The reason for that was the consistency of its replication in several GWAS; to date, no other *TSLP* SNP was identified in a GWAS. However, other *TSLP* SNPs

have been associated with asthma and related phenotypes in candidate gene association studies (109, 110, 424, 425, 456, 457) although the majority of these studies did not consider rs1837253 for association analysis. These studies are summarized in 6.3B.2 .These *TSLP* SNPs might also be important for gene regulation and asthma; but the fact that these SNPs were not identified in GWAS, even in the GWAS of the Japanese population (90), where some of these candidate genetic association studies were performed, indicates that their involvement in asthma is minor compared to rs1837253 or that they served as a surrogate for it, as there is some degree of LD in the Asian population.

A clear limitation of my TSLP EMSA data is the lack of consistent successful cold competition; many strategies were used to attempt to remedy that problem and the assays were repeated multiple times (~18 times). Nevertheless, the differential binding to the T allele was always replicated.

A limitation of my studies is that I did not perform any methylation assays to verify any potential effect of my SNP of interest; there was a report of hypomethylation leading to heightened *TSLP* expression in atopic dermatitis lesions (458) as well as a significant difference of methylation status of the *TSLP* promoter in children with and without prenatal smoke exposure (459); but the scope of my thesis work was to investigate the asthma-associated SNP rs1837253 and there were no reported CpG sites around it, as verified using the UCSC genome browser.

In future work, the identification of the protein or complex which binds to the T allele and the successful cold competition are very desirable. Mass spectrometry after stable isotope labeling of the DNA/protein complex could be used to identify the proteins bound

(460) followed by verification using consensus sequence oligonucleotides in EMSA competition.

6.3 Concluding remarks

More extensive research needs to take place in order to clearly understand the specific roles of both *TSLP* and *IL1RL1* transcript isoforms in the lung. Understanding the different regulation and expression patterns of each isoform would greatly enhance our understanding of asthma pathogenesis by answering some of the unresolved questions in the mechanisms of asthmatic lung inflammation. Given the upstream production of IL-33 and TSLP from the epithelium, these proteins and their receptors present a great opportunity for therapeutic intervention before chronic inflammation becomes established in the lung. An excellent way of investigating those mechanisms is to follow up SNP associations with functional analysis while paying the utmost attention to the selection of the SNPs most likely to be worth the time and financial commitment of this follow-up. In a metaphorical sense, functional SNPs pinpoint to the exact "location" in the signaling cascades where homeostatic processes and healthy immune responses break-down and should be used to identify those proteins or signaling hubs' breaking points or imbalances in the normal immune responses. Asthma therapies have not changed in a long time and are still focused on relieving excessive existing inflammation and acute bronchospasms. Prenatal screening for asthma susceptibility and eventually a cure are long overdue, and efforts should be directed towards the opportunity that epithelial genes, such as TSLP and IL-33, offer to understand the disease onset mechanisms.

Bibliography

(1) Le Galès-Camus C. World Health Organization.Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach. : World Health Organization; 2007.

(2) GINA Report, Global Strategy for Asthma Management and Prevention. Internet. Cited Oct 26, 2013 http://www.ginasthma.org/documents/4; 2013.

(3) Barnes PJ, Adcock IM. How do corticosteroids work in asthma? Ann Intern Med 2003 Sep 2;139(5 Pt 1):359-370.

(4) Barnes NC. The properties of inhaled corticosteroids: similarities and differences. Prim Care Respir J 2007 Jun;16(3):149-154.

(5) Montuschi P, Peters-Golden ML. Leukotriene modifiers for asthma treatment. Clin Exp Allergy 2010 Dec;40(12):1732-1741.

(6) Giembycz MA, Newton R. Beyond the dogma: novel beta2-adrenoceptor signalling in the airways. Eur Respir J 2006 Jun;27(6):1286-1306.

(7) Kaur M, Chivers JE, Giembycz MA, Newton R. Long-acting beta2-adrenoceptor agonists synergistically enhance glucocorticoid-dependent transcription in human airway epithelial and smooth muscle cells. Mol Pharmacol 2008 Jan;73(1):203-214.

(8) Pinto LA, Stein RT, Kabesch M. Impact of genetics in childhood asthma. J Pediatr (Rio J) 2008 Aug;84(4 Suppl):S68-75.

(9) Marsh DG, Bias WB, Hsu SH, Goodfriend L. Association of the HL-A7 cross-reacting group with a specific reaginic antibody response in allergic man. Science 1973 Feb 16;179(4074):691-693.

(10) Sandford A, Weir T, Pare P. The genetics of asthma. Am J Respir Crit Care Med 1996 Jun;153(6 Pt 1):1749-1765.

(11) Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002 Jul 25;418(6896):426-430.

(12) Allen M, Heinzmann A, Noguchi E, Abecasis G, Broxholme J, Ponting CP, et al. Positional cloning of a novel gene influencing asthma from chromosome 2q14. Nat Genet 2003 Nov;35(3):258-263.

(13) Laitinen T, Polvi A, Rydman P, Vendelin J, Pulkkinen V, Salmikangas P, et al. Characterization of a common susceptibility locus for asthma-related traits. Science 2004 Apr 9;304(5668):300-304.

(14) Zhang Y, Leaves NI, Anderson GG, Ponting CP, Broxholme J, Holt R, et al. Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. Nat Genet 2003 Jun;34(2):181-186.

(15) Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010 Apr 1;6(4):e1000888.

(16) Nicolae D, Cox NJ, Lester LA, Schneider D, Tan Z, Billstrand C, et al. Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21. Am J Hum Genet 2005 Feb;76(2):349-357.

(17) Bouzigon E, Forabosco P, Koppelman GH, Cookson WO, Dizier MH, Duffy DL, et al. Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy. Eur J Hum Genet 2010 Jun;18(6):700-706.

(18) Hoffjan S, Nicolae D, Ober C. Association studies for asthma and atopic diseases: a comprehensive review of the literature. Respir Res 2003 Dec 4;4:14.

(19) Vercelli D. Discovering susceptibility genes for asthma and allergy. Nat Rev Immunol 2008 Mar;8(3):169-182.

(20) Rogers AJ, Raby BA, Lasky-Su JA, Murphy A, Lazarus R, Klanderman BJ, et al. Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. Am J Respir Crit Care Med 2009 Jun 15;179(12):1084-1090.

(21) Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001 Feb;2(2):91-99.

(22) Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med 2010 Jul 8;363(2):166-176.

(23) Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 2007 Jul 26;448(7152):470-473.

(24) Sleiman PM, Annaiah K, Imielinski M, Bradfield JP, Kim CE, Frackelton EC, et al. ORMDL3 variants associated with asthma susceptibility in North Americans of European ancestry. J Allergy Clin Immunol 2008 Dec;122(6):1225-1227.

(25) Tavendale R, Macgregor DF, Mukhopadhyay S, Palmer CN. A polymorphism controlling ORMDL3 expression is associated with asthma that is poorly controlled by current medications. J Allergy Clin Immunol 2008 Apr;121(4):860-863.

(26) Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. N Engl J Med 2008 Nov 6;359(19):1985-1994.

(27) Galanter J, Choudhry S, Eng C, Nazario S, Rodriguez-Santana JR, Casal J, et al. ORMDL3 gene is associated with asthma in three ethnically diverse populations. Am J Respir Crit Care Med 2008 Jun 1;177(11):1194-1200.

(28) Hirota T, Harada M, Sakashita M, Doi S, Miyatake A, Fujita K, et al. Genetic polymorphism regulating ORM1-like 3 (Saccharomyces cerevisiae) expression is associated with childhood atopic asthma in a Japanese population. J Allergy Clin Immunol 2008 Mar;121(3):769-770.

(29) Wu H, Romieu I, Sienra-Monge JJ, Li H, del Rio-Navarro BE, London SJ. Genetic variation in ORM1-like 3 (ORMDL3) and gasdermin-like (GSDML) and childhood asthma. Allergy 2009 Apr;64(4):629-635.

(30) Bisgaard H, Bonnelykke K, Sleiman PM, Brasholt M, Chawes B, Kreiner-Moller E, et al. Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. Am J Respir Crit Care Med 2009 Feb 1;179(3):179-185.

(31) Leung TF, Sy HY, Ng MC, Chan IH, Wong GW, Tang NL, et al. Asthma and atopy are associated with chromosome 17q21 markers in Chinese children. Allergy 2009 Apr;64(4):621-628.

(32) Flory JH, Sleiman PM, Christie JD, Annaiah K, Bradfield J, Kim CE, et al. 17q12-21 variants interact with smoke exposure as a risk factor for pediatric asthma but are equally associated with early-onset versus late-onset asthma in North Americans of European ancestry. J Allergy Clin Immunol 2009 Sep;124(3):605-607.

(33) Halapi E, Gudbjartsson DF, Jonsdottir GM, Bjornsdottir US, Thorleifsson G, Helgadottir H, et al. A sequence variant on 17q21 is associated with age at onset and severity of asthma. Eur J Hum Genet 2010 Aug;18(8):902-908.

(34) Madore AM, Tremblay K, Hudson TJ, Laprise C. Replication of an association between 17q21 SNPs and asthma in a French-Canadian familial collection. Hum Genet 2008 Feb;123(1):93-95.

(35) Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010 Sep 23;363(13):1211-1221.

(36) Ferreira MA, McRae AF, Medland SE, Nyholt DR, Gordon SD, Wright MJ, et al. Association between ORMDL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia. Eur J Hum Genet 2011 Apr;19(4):458-464.

(37) Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet 2011 Jul 31;43(9):887-892.

(38) Wan YI, Shrine NR, Soler Artigas M, Wain LV, Blakey JD, Moffatt MF, et al. Genome-wide association study to identify genetic determinants of severe asthma. Thorax 2012 May 5;67(9):762--8.

(39) Kang MJ, Yu HS, Seo JH, Kim HY, Jung YH, Kim YJ, et al. GSDMB/ORMDL3 variants contribute to asthma susceptibility and eosinophil-mediated bronchial hyperresponsiveness. Hum Immunol 2012 Sep;73(9):954-959.

(40) Lluis A, Schedel M, Liu J, Illi S, Depner M, von Mutius E, et al. Asthma-associated polymorphisms in 17q21 influence cord blood ORMDL3 and GSDMA gene expression and IL-17 secretion. J Allergy Clin Immunol 2011 Jun;127(6):1587-94.e6.

(41) Verlaan DJ, Berlivet S, Hunninghake GM, Madore AM, Lariviere M, Moussette S, et al. Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am J Hum Genet 2009 Sep;85(3):377-393.

(42) Berlivet S, Moussette S, Ouimet M, Verlaan DJ, Koka V, Al Tuwaijri A, et al. Interaction between genetic and epigenetic variation defines gene expression patterns at the asthma-associated locus 17q12-q21 in lymphoblastoid cell lines. Hum Genet 2012 Jul;131(7):1161-1171.

(43) Holloway JW, Koppelman GH. 17q21 Variants and Asthma - Questions and Answers. N Engl J Med 2008 Nov 6;359(19):2043-2045.

(44) Cantero-Recasens G, Fandos C, Rubio-Moscardo F, Valverde MA, Vicente R. The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress. Hum Mol Genet 2010 Jan 1;19(1):111-121.

(45) Miller M, Tam AB, Cho JY, Doherty TA, Pham A, Khorram N, et al. ORMDL3 is an inducible lung epithelial gene regulating metalloproteases, chemokines, OAS, and ATF6. Proc Natl Acad Sci U S A 2012 Oct 9;109(41):16648-16653.

(46) Hsu KJ, Turvey SE. Functional analysis of the impact of ORMDL3 expression on inflammation and activation of the unfolded protein response in human airway epithelial cells. Allergy Asthma Clin Immunol 2013 Feb 1;9(1):4-1492-9-4.

(47) Breslow DK, Collins SR, Bodenmiller B, Aebersold R, Simons K, Shevchenko A, et al. Orm family proteins mediate sphingolipid homeostasis. Nature 2010 Feb 25;463(7284):1048-1053.

(48) Zheng W, Kollmeyer J, Symolon H, Momin A, Munter E, Wang E, et al. Ceramides and other bioactive sphingolipid backbones in health and disease: lipidomic analysis, metabolism and roles in membrane structure, dynamics, signaling and autophagy. Biochim Biophys Acta 2006 Dec;1758(12):1864-1884.

(49) Ha SG, Ge XN, Bahaie NS, Kang BN, Rao A, Rao SP, et al. ORMDL3 promotes eosinophil trafficking and activation via regulation of integrins and CD48. Nat Commun 2013;4:2479.

(50) Hao K, Bossé Y, Nickle D, Paré PD, Postma D, Laviolette M, et al. Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. PLoS Genet 2012;8(11):e1003029.

(51) Komiyama H, Aoki A, Tanaka S, Maekawa H, Kato Y, Wada R, et al. Alu-derived cis-element regulates tumorigenesis-dependent gastric expression of GASDERMIN B (GSDMB). Genes Genet Syst 2010 Feb;85(1):75-83.

(52) Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 2004 Apr 20;101(16):6062-6067.

(53) Lin YN, Roy A, Yan W, Burns KH, Matzuk MM. Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. Mol Cell Biol 2007 Oct;27(19):6794-6805.

(54) Naumova AK, Al Tuwaijri A, Morin A, Vaillancourt VT, Madore AM, Berlivet S, et al. Sex- and age-dependent DNA methylation at the 17q12-q21 locus associated with childhood asthma. Hum Genet 2013 Jul;132(7):811-822.

(55) Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, et al. Variants of DENND1B associated with asthma in children. N Engl J Med 2010 Jan 7;362(1):36-44.

(56) Hancock DB, Romieu I, Shi M, Sienra-Monge JJ, Wu H, Chiu GY, et al. Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in mexican children. PLoS Genet 2009 Aug;5(8):e1000623.

(57) Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. Am J Hum Genet 2009 May;84(5):581-593.

(58) Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008 Apr 3;452(7187):638-642.

(59) Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. PLoS Genet 2010 Aug 5;6(8):10.1371/journal.pgen.1001053.

(60) Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 2009 Mar;5(3):e1000421.

(61) Pillai SG, Kong X, Edwards LD, Cho MH, Anderson WH, Coxson HO, et al. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2010 Dec 15;182(12):1498-1505.

(62) Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 2010 May;42(5):448-453.

(63) Stephens SH, Hartz SM, Hoft NR, Saccone NL, Corley RC, Hewitt JK, et al. Distinct Loci in the CHRNA5/CHRNA3/CHRNB4 Gene Cluster Are Associated With Onset of Regular Smoking. Genet Epidemiol 2013 Dec;37(8):846-859.

(64) Hansen G, Jin S, Umetsu DT, Conti M. Absence of muscarinic cholinergic airway responses in mice deficient in the cyclic nucleotide phosphodiesterase PDE4D. Proc Natl Acad Sci U S A 2000 Jun 6;97(12):6751-6756.

(65) Rosand J, Bayley N, Rost N, de Bakker PI. Many hypotheses but no replication for the association between PDE4D and stroke. Nat Genet 2006 Oct;38(10):1091-2; author reply 1092-3.

(66) Reneland RH, Mah S, Kammerer S, Hoyal CR, Marnellos G, Wilson SG, et al. Association between a variation in the phosphodiesterase 4D gene and bone mineral density. BMC Med Genet 2005 Mar 7;6:9.

(67) Kuhlenbaumer G, Berger K, Huge A, Lange E, Kessler C, John U, et al. Evaluation of single nucleotide polymorphisms in the phosphodiesterase 4D gene (PDE4D) and their association with ischaemic stroke in a large German cohort. J Neurol Neurosurg Psychiatry 2006 Apr;77(4):521-524.

(68) Calboli FC, Tozzi F, Galwey NW, Antoniades A, Mooser V, Preisig M, et al. A genome-wide association study of neuroticism in a population-based sample. PLoS One 2010 Jul 9;5(7):e11504.

(69) Gross NJ, Giembycz MA, Rennard SI. Treatment of chronic obstructive pulmonary disease with roflumilast, a new phosphodiesterase 4 inhibitor. COPD 2010 Apr;7(2):141-153.

(70) Kodimuthali A, Jabaris SS, Pal M. Recent advances on phosphodiesterase 4 inhibitors for the treatment of asthma and chronic obstructive pulmonary disease. J Med Chem 2008 Sep 25;51(18):5471-5489.

(71) Higgs G. Is PDE4 too difficult a drug target? Curr Opin Investig Drugs 2010 May;11(5):495-498.

(72) Burrows B, Martinez FD, Halonen M, Barbee RA, Cline MG. Association of asthma with serum IgE levels and skin-test reactivity to allergens. N Engl J Med 1989 Feb 2;320(5):271-277.

(73) Wardlaw AJ, Brightling C, Green R, Woltmann G, Pavord I. Eosinophils in asthma and other allergic diseases. Br Med Bull 2000;56(4):985-1003.

(74) Burrows B, Knudson RJ, Cline MG, Lebowitz MD. A reexamination of risk factors for ventilatory impairment. Am Rev Respir Dis 1988 Oct;138(4):829-836.

(75) Rosenberg HF, Phipps S, Foster PS. Eosinophil trafficking in allergy and asthma. J Allergy Clin Immunol 2007 Jun;119(6):1303-10; quiz 1311-2.

(76) Weidinger S, Gieger C, Rodriguez E, Baurecht H, Mempel M, Klopp N, et al. Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. PLoS Genet 2008 Aug;4(8):e1000166.

(77) Platzer B, Fiebiger E. The signal peptide of the IgE receptor alpha-chain prevents surface expression of an immunoreceptor tyrosine-based activation motif-free receptor pool. J Biol Chem 2010 May 14;285(20):15314-15323.

(78) Bae JS, Kim SH, Ye YM, Yoon HJ, Suh CH, Nahm DH, et al. Significant association of FcepsilonRIalpha promoter polymorphisms with aspirin-intolerant chronic urticaria. J Allergy Clin Immunol 2007 Feb;119(2):449-456.

(79) Potaczek DP, Sanak M, Szczeklik A. Additive association between FCER1A and FCER1B genetic polymorphisms and total serum IgE levels. Allergy 2007 Sep;62(9):1095-1096.

(80) Schedel M, Frei R, Bieli C, Cameron L, Adamski J, Lauener R, et al. An IgE-associated polymorphism in STAT6 alters NF-kappaB binding, STAT6 promoter activity, and mRNA expression. J Allergy Clin Immunol 2009 Sep;124(3):583-9, 589.e1-6.

(81) Kaplan MH, Wurster AL, Smiley ST, Grusby MJ. Stat6-dependent and -independent pathways for IL-4 production. J Immunol 1999 Dec 15;163(12):6536-6540.

(82) Schedel M, Carr D, Klopp N, Woitsch B, Illig T, Stachel D, et al. A signal transducer and activator of transcription 6 haplotype influences the regulation of serum IgE levels. J Allergy Clin Immunol 2004 Nov;114(5):1100-1105.

(83) Weidinger S, Klopp N, Wagenpfeil S, Rummler L, Schedel M, Kabesch M, et al. Association of a STAT 6 haplotype with elevated serum IgE levels in a population based cohort of white adults. J Med Genet 2004 Sep;41(9):658-663.

(84) Lee GR, Fields PE, Griffin TJ, Flavell RA. Regulation of the Th2 cytokine locus by a locus control region. Immunity 2003 Jul;19(1):145-153.

(85) Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, et al. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. J Allergy Clin Immunol 2010 Feb;125(2):328-335.e11.

(86) Lee DU, Rao A. Molecular analysis of a locus control region in the T helper 2 cytokine gene cluster: a target for STAT6 but not GATA3. Proc Natl Acad Sci U S A 2004 Nov 9;101(45):16010-16015.

(87) Munthe-Kaas MC, Carlsen KL, Carlsen KH, Egeland T, Haland G, Devulapalli CS, et al. HLA Dr-Dq haplotypes and the TNFA-308 polymorphism: associations with asthma and allergy. Allergy 2007 Sep;62(9):991-998.

(88) Martyn MB, Molis W, Jacobson RM, Poland GA, Weaver AL, Juhn YJ. Human leukocyte antigen type and progression from onset of symptoms to development of asthma. Allergy Asthma Proc 2010 Mar-Apr;31(2):120-125.

(89) Aron Y, Desmazes-Dufeu N, Matran R, Polla BS, Dusser D, Lockhart A, et al. Evidence of a strong, positive association between atopy and the HLA class II alleles DR4 and DR7. Clin Exp Allergy 1996 Jul;26(7):821-828.

(90) Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. Nat Genet 2011 Jul 31;43(9):893-896.

(91) Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M, et al. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. PLoS Genet 2011 Jul;7(7):e1002170.

(92) Lasky-Su J, Himes BE, Raby BA, Klanderman BJ, Sylvia JS, Lange C, et al. HLA-DQ strikes again: genome-wide association study further confirms HLA-DQ in the diagnosis of asthma among adults. Clin Exp Allergy 2012 Dec;42(12):1724-1733.

(93) Choudhry S, Taub M, Mei R, Rodriguez-Santana J, Rodriguez-Cintron W, Shriver MD, et al. Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region. Hum Genet 2008 Jun;123(5):455-468.

(94) Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat Genet 2009 Mar;41(3):342-347.

(95) Velazquez L, Cheng AM, Fleming HE, Furlonger C, Vesely S, Bernstein A, et al. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. J Exp Med 2002 Jun 17;195(12):1599-1611.

(96) Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 2008 Apr;40(4):395-402.

(97) Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007 Jul;39(7):857-864.

(98) Hirasawa R, Shimizu R, Takahashi S, Osawa M, Takayanagi S, Kato Y, et al. Essential and instructive roles of GATA factors in eosinophil development. J Exp Med 2002 Jun 3;195(11):1379-1386.

(99) Ali M, Zhang G, Thomas WR, McLean CJ, Bizzintino JA, Laing IA, et al. Investigations into the role of ST2 in acute asthma in children. Tissue Antigens 2009 Mar;73(3):206-212.

(100) Liew FY, Pitman NI, McInnes IB. Disease-associated functions of IL-33: the new kid in the IL-1 family. Nat Rev Immunol 2010 Feb;10(2):103-110.

(101) Akhabir L, Sandford A. Genetics of interleukin 1 receptor-like 1 in immune and inflammatory diseases. Curr Genomics 2010 Dec;11(8):591-606.

(102) Takatsu K, Nakajima H. IL-5 and eosinophilia. Curr Opin Immunol 2008 Jun;20(3):288-294.

(103) Getnet D, Grosso JF, Goldberg MV, Harris TJ, Yen HR, Bruno TC, et al. A role for the transcription factor Helios in human CD4(+)CD25(+) regulatory T cells. Mol Immunol 2010 Apr;47(7-8):1595-1600.

(104) Rothenberg ME, Spergel JM, Sherrill JD, Annaiah K, Martin LJ, Cianferoni A, et al. Common variants at 5q22 associate with pediatric eosinophilic esophagitis. Nat Genet 2010 Apr;42(4):289-291.

(105) Mao M, Biery MC, Kobayashi SV, Ward T, Schimmack G, Burchard J, et al. T lymphocyte activation gene identification by coregulated expression on DNA microarrays. Genomics 2004 Jun;83(6):989-999.

(106) He R, Geha RS. Thymic stromal lymphopoietin. Ann N Y Acad Sci 2010 Jan;1183:13-24.

(107) He JQ, Hallstrand TS, Knight D, Chan-Yeung M, Sandford A, Tripp B, et al. A thymic stromal lymphopoietin gene variant is associated with asthma and airway hyperresponsiveness. J Allergy Clin Immunol 2009 Aug;124(2):222-229.

(108) Hunninghake GM, Lasky-Su J, Soto-Quiros ME, Avila L, Liang C, Lake SL, et al. Sex-stratified linkage analysis identifies a female-specific locus for IgE to cockroach in Costa Ricans. Am J Respir Crit Care Med 2008 Apr 15;177(8):830-836.

(109) Harada M, Hirota T, Jodo AI, Hitomi Y, Sakashita M, Tsunoda T, et al. Thymic stromal lymphopoietin gene promoter polymorphisms are associated with susceptibility to bronchial asthma. Am J Respir Cell Mol Biol 2011 Jun;44(6):787-793.

(110) Hunninghake GM, Soto-Quiros ME, Avila L, Kim HP, Lasky-Su J, Rafaels N, et al. TSLP polymorphisms are associated with asthma in a sex-specific fashion. Allergy 2010 Dec;65(12):1566-1575.

(111) Mathias RA, Grant AV, Rafaels N, Hand T, Gao L, Vergara C, et al. A genome-wide association study on African-ancestry populations for asthma. J Allergy Clin Immunol 2010 Feb;125(2):336-346.e4.

(112) Zagha E, Ozaita A, Chang SY, Nadal MS, Lin U, Saganich MJ, et al. DPP10 modulates Kv4-mediated A-type potassium channels. J Biol Chem 2005 May 13;280(19):18853-18861.

(113) Gao J, Li W, Willis-Owen SA, Jiang L, Ma Y, Tian X, et al. Polymorphisms of PHF11 and DPP10 are associated with asthma and related traits in a Chinese population. Respiration 2010;79(1):17-24.

(114) Wu H, Romieu I, Shi M, Hancock DB, Li H, Sienra-Monge J, et al. Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. J Allergy Clin Immunol 2010 2;125(2):321-327.e13.

(115) Hu ZW, Shi XY, Lin RZ, Hoffman BB. Alpha1 adrenergic receptors activate phosphatidylinositol 3-kinase in human vascular smooth muscle cells. Role in mitogenesis. J Biol Chem 1996 Apr 12;271(15):8977-8982.

(116) de Almeida CJ, Chiarini LB, da Silva JP, E Silva PM, Martins MA, Linden R. The cellular prion protein modulates phagocytosis and inflammatory response. J Leukoc Biol 2005 Feb;77(2):238-246.

(117) Sunyach C, Checler F. Combined pharmacological, mutational and cell biology approaches indicate that p53-dependent caspase 3 activation triggered by cellular prion is dependent on its endocytosis. J Neurochem 2005 Mar;92(6):1399-1407.

(118) Madore AM, Perron S, Turmel V, Laviolette M, Bissonnette EY, Laprise C. Alveolar macrophages in allergic asthma: an expression signature characterized by heat shock protein pathways. Hum Immunol 2010 Feb;71(2):144-150.

(119) Murphy GE, Xu D, Liew FY, McInnes IB. Role of interleukin 33 in human immunopathology. Ann Rheum Dis 2010 Jan;69 Suppl 1:i43-47.

(120) Bonnelykke K, Sleiman P, Nielsen K, Kreiner-Moller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet 2014 Jan;46(1):51-55.

(121) Ross AJ, Dailey LA, Brighton LE, Devlin RB. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. Am J Respir Cell Mol Biol 2007 Aug;37(2):169-185.

(122) Mao JR, Taylor G, Dean WB, Wagner DR, Afzal V, Lotz JC, et al. Tenascin-X deficiency mimics Ehlers-Danlos syndrome in mice through alteration of collagen deposition. Nat Genet 2002 Apr;30(4):421-425.

(123) Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am J Hum Genet 2007 Jun;80(6):1037-1054.

(124) Garlick KM, Mogridge J. Direct interaction between anthrax toxin receptor 1 and the actin cytoskeleton. Biochemistry 2009 Nov 10;48(44):10577-10581.

(125) Kanakry CG, Li Z, Nakai Y, Sei Y, Weinberger DR. Neuregulin-1 regulates cell adhesion via an ErbB2/phosphoinositide-3 kinase/Akt-dependent pathway: potential implications for schizophrenia and cancer. PLoS One 2007 Dec 26;2(12):e1369.

(126) Tang ML, Wilson JW, Stewart AG, Royce SG. Airway remodelling in asthma: current understanding and implications for future therapies. Pharmacol Ther 2006 Nov;112(2):474-488.

(127) Valiyaveettil M, Bentley AA, Gursahaney P, Hussien R, Chakravarti R, Kureishy N, et al. Novel role of the muskelin-RanBP9 complex as a nucleocytoplasmic mediator of cell morphology regulation. J Cell Biol 2008 Aug 25;182(4):727-739.

(128) Castro-Giner F, Bustamante M, Ramon Gonzalez J, Kogevinas M, Jarvis D, Heinrich J, et al. A pooling-based genome-wide analysis identifies new potential candidate genes for atopy in the European Community Respiratory Health Survey (ECRHS). BMC Med Genet 2009 Dec 6;10:128-2350-10-128.

(129) Kim SH, Cho BY, Park CS, Shin ES, Cho EY, Yang EM, et al. Alpha-T-catenin (CTNNA3) gene was identified as a risk variant for toluene diisocyanate-induced asthma by genome-wide association analysis. Clin Exp Allergy 2009 Feb;39(2):203-212.

(130) Morgan AR, Hamilton G, Turic D, Jehu L, Harold D, Abraham R, et al. Association analysis of 528 intra-genic SNPs in a region of chromosome 10 linked to late onset Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 2008 Sep 5;147B(6):727-731.

(131) Busby V, Goossens S, Nowotny P, Hamilton G, Smemo S, Harold D, et al. Alpha-T-catenin is expressed in human brain and interacts with the Wnt signaling pathway but is not responsible for linkage to chromosome 10 in Alzheimer's disease. Neuromolecular Med 2004;5(2):133-146.

(132) Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet 2007 Jan 1;16(1):24-35.

(133) Jorder L, Wooding S. Genetic variation, classifiction and 'race'. Nat Genet 2004;36(11s):S28-33.

(134) Tishkoff SA, Kidd KK. Implications of biogeography of human populations for 'race' and medicine. Nat Genet 2004 Nov;36(11 Suppl):S21-7.

(135) Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 2006 Jun;38(6):626-635.

(136) Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R. MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc Natl Acad Sci U S A 2008 Feb 5;105(5):1608-1613.

(137) Ahn MH, Park BL, Lee SH, Park SW, Park JS, Kim DJ, et al. A promoter SNP rs4073T>A in the common allele of the interleukin 8 gene is associated with the development of idiopathic pulmonary fibrosis via the IL-8 protein enhancing mode. Respir Res 2011 Jun 8;12:73-9921-12-73.

(138) Pickering BM, Willis AE. The implications of structured 5' untranslated regions on translation and disease. Semin Cell Dev Biol 2005 Feb;16(1):39-47.

(139) Gray NK, Hentze MW. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. EMBO J 1994 Aug 15;13(16):3882-3891.

(140) Iacono M, Mignone F, Pesole G. uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene 2005 Apr 11;349:97-105.

(141) Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol Cell 2009 May;101(5):251-262.

(142) Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet 2007 Apr;80(4):727-739.

(143) Rose AB. Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol 2008;326:277-290.

(144) St Laurent G, Shtokalo D, Tackett MR, Yang Z, Eremina T, Wahlestedt C, et al. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. BMC Genomics 2012 Sep 24;13:504-2164-13-504.

(145) Solis AS, Shariat N, Patton JG. Splicing fidelity, enhancers, and disease. Front Biosci 2008 Jan 1;13:1926-1942.

(146) Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, et al. A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am J Hum Genet 2007 Feb;80(2):241-252.

(147) Sayed D, Abdellatif M. MicroRNAs in development and disease. Physiol Rev 2011 Jul;91(3):827-887.

(148) Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 2008 Jun 18;582(14):1977-1986.

(149) Tan-Wong SM, French JD, Proudfoot NJ, Brown MA. Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. Proc Natl Acad Sci U S A 2008 Apr 1;105(13):5160-5165.

(150) Fu L, Ma W, Benchimol S. A translation repressor element resides in the 3' untranslated region of human p53 mRNA. Oncogene 1999 Nov 11;18(47):6419-6424.

(151) Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. Mol Cell 2011 Sep 16;43(6):853-866.

(152) Lee S, Vasudevan S. Post-transcriptional stimulation of gene expression by microRNAs. Adv Exp Med Biol 2013;768:97-126.

(153) Lee I, Ajay SS, Yook JI, Kim HS, Hong SH, Kim NH, et al. New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. Genome Res 2009 Jul;19(7):1175-1183.

(154) Forman JJ, Legesse-Miller A, Coller HA. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. Proc Natl Acad Sci U S A 2008 Sep 30;105(39):14879-14884.

(155) Choi E, Choi E, Hwang KC. MicroRNAs as novel regulators of stem cell fate. World J Stem Cells 2013 Oct 26;5(4):172-187.

(156) Vickers KC, Sethupathy P, Baran-Gale J, Remaley AT. Complexity of microRNA function and the role of isomiRs in lipid homeostasis. J Lipid Res 2013 May;54(5):1182-1191.

(157) Sittka A, Schmeck B. MicroRNAs in the lung. Adv Exp Med Biol 2013;774:121-134.

(158) Rebane A, Akdis CA. MicroRNAs: Essential players in the regulation of inflammation. J Allergy Clin Immunol 2013 Jul;132(1):15-26.

(159) Peng SS, Chen CY, Xu N, Shyu AB. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. EMBO J 1998 Jun 15;17(12):3461-3470.

(160) Kiledjian M, DeMaria CT, Brewer G, Novick K. Identification of AUF1 (heterogeneous nuclear ribonucleoprotein D) as a component of the alpha-globin mRNA stability complex. Mol Cell Biol 1997 Aug;17(8):4870-4876.

(161) Lai WS, Carballo E, Strum JR, Kennington EA, Phillips RS, Blackshear PJ. Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA. Mol Cell Biol 1999 Jun;19(6):4311-4323.

(162) Gherzi R, Lee KY, Briata P, Wegmuller D, Moroni C, Karin M, et al. A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. Mol Cell 2004 Jun 4;14(5):571-583.

(163) Chen JM, Ferec C, Cooper DN. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. Hum Genet 2006 Oct;120(3):301-333.

(164) Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009 Jun 9;106(23):9362-9367.

(165) Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012 Sep 7;337(6099):1190-1195.

(166) Hemberg M, Gray JM, Cloonan N, Kuersten S, Grimmond S, Greenberg ME, et al. Integrated genome analysis suggests that most conserved non-coding sequences are regulatory factor binding sites. Nucleic Acids Res 2012 Sep;40(16):7858-7869.

(167) Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature 2012 Sep 6;489(7414):109-113.

(168) Glinskii AB, Ma S, Ma J, Grant D, Lim CU, Guest I, et al. Networks of intergenic long-range enhancers and snpRNAs drive castration-resistant phenotype of prostate cancer and contribute to pathogenesis of multiple common human disorders. Cell Cycle 2011 Oct 15;10(20):3571-3597.

(169) Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet 2006 Sep;7(9):703-713.

(170) Raab JR, Kamakaka RT. Insulators and promoters: closer than we think. Nat Rev Genet 2010 Jun;11(6):439-446.

(171) Shrimali S, Srivastava S, Varma G, Grinberg A, Pfeifer K, Srivastava M. An ectopic CTCF-dependent transcriptional insulator influences the choice of Vbeta gene segments for VDJ recombination at TCRbeta locus. Nucleic Acids Res 2012 Sep;40(16):7753-7765.

(172) Gillen AE, Harris A. The role of CTCF in coordinating the expression of single gene loci. Biochem Cell Biol 2011 Oct;89(5):489-494.

(173) Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, et al. CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 2008 Jun 17;105(24):8309-8314.

(174) Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 2009 Jul 14;106(28):11667-11672.

(175) Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 2011 Aug 28;477(7364):295-300.

(176) Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res 2010 Sep;20(9):1191-1197.

(177) Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet 2009 Aug;41(8):882-884.

(178) Mitcham JL, Parnet P, Bonnert TP, Garka KE, Gerhart MJ, Slack JL, et al. T1/ST2 signaling establishes it as a member of an expanding interleukin-1 receptor family. J Biol Chem 1996 Mar 8;271(10):5777-5783.

(179) Kumar S, Minnich MD, Young PR. ST2/T1 protein functionally binds to two secreted proteins from Balb/c 3T3 and human umbilical vein endothelial cells but does not bind interleukin 1. J Biol Chem 1995 Nov 17;270(46):27905-27913.

(180) Schmitz J, Owyang A, Oldham E, Song Y, Murphy E, McClanahan TK, et al. IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. Immunity 2005 Nov;23(5):479-490.

(181) Li H, Tago K, Io K, Kuroiwa K, Arai T, Iwahana H, et al. The cloning and nucleotide sequence of human ST2L cDNA. Genomics 2000 Aug 1;67(3):284-290.

(182) Gracie JA, Robertson SE, McInnes IB. Interleukin-18. J Leukoc Biol 2003 Feb;73(2):213-224.

(183) McInnes IB, Liew FY, Gracie JA. Interleukin-18: a therapeutic target in rheumatoid arthritis? Arthritis Res Ther 2005;7(1):38-41.

(184) Nakanishi K, Yoshimoto T, Tsutsui H, Okamura H. Interleukin-18 is a unique cytokine that stimulates both Th1 and Th2 responses depending on its cytokine milieu. Cytokine Growth Factor Rev 2001 Mar;12(1):53-72.

(185) Kang MJ, Homer RJ, Gallo A, Lee CG, Crothers KA, Cho SJ, et al. IL-18 is induced and IL-18 receptor alpha plays a critical role in the pathogenesis of cigarette smoke-induced pulmonary emphysema and inflammation. J Immunol 2007 Feb 1;178(3):1948-1959.

(186) Watanabe M, Kaneko H, Shikano H, Aoki M, Sakaguchi H, Matsui E, et al. Predominant expression of 950delCAG of IL-18R alpha chain cDNA is associated with reduced IFN-gamma production and high serum IgE levels in atopic Japanese children. J Allergy Clin Immunol 2002 Apr;109(4):669-675.

(187) Wittmann M, Purwar R, Hartmann C, Gutzmer R, Werfel T. Human keratinocytes respond to interleukin-18: implication for the course of chronic inflammatory skin diseases. J Invest Dermatol 2005 Jun;124(6):1225-1233.

(188) Kawakami K, Qureshi MH, Zhang T, Okamura H, Kurimoto M, Saito A. IL-18 protects mice against pulmonary and disseminated infection with Cryptococcus neoformans by inducing IFN-gamma production. J Immunol 1997 Dec 1;159(11):5528-5534.

(189) Dayer JM. Interleukin-18, rheumatoid arthritis, and tissue destruction. J Clin Invest 1999 Nov;104(10):1337-1339.

(190) Reijmerink NE, Postma DS, Bruinenberg M, Nolte IM, Meyers DA, Bleecker ER, et al. Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy. J Allergy Clin Immunol 2008 9;122(3):651-654.e8.

(191) Zhu J, Hastie T. Classification of gene microarrays by penalized logistic regression. Biostatistics 2004 Jul;5(3):427-443.

(192) Tago K, Noda T, Hayakawa M, Iwahana H, Yanagisawa K, Yashiro T, et al. Tissue distribution and subcellular localization of a variant form of the human ST2 gene product, ST2V. Biochem Biophys Res Commun 2001 Aug 3;285(5):1377-1383.

(193) Bergers G, Reikerstorfer A, Braselmann S, Graninger P, Busslinger M. Alternative promoter usage of the Fos-responsive gene Fit-1 generates mRNA isoforms coding for either secreted or membrane-bound proteins related to the IL-1 receptor. EMBO J 1994 Mar 1;13(5):1176-1188.

(194) Yanagisawa K, Naito Y, Kuroiwa K, Arai T, Furukawa Y, Tomizuka H, et al. The expression of ST2 gene in helper T cells and the binding of ST2 protein to myeloma-derived RPMI8226 cells. J Biochem 1997 Jan;121(1):95-103.

(195) Suzukawa M, Iikura M, Koketsu R, Nagase H, Tamura C, Komiya A, et al. An IL-1 cytokine member, IL-33, induces human basophil activation via its ST2 receptor. J Immunol 2008 Nov 1;181(9):5981-5989.

(196) Suzukawa M, Koketsu R, Iikura M, Nakae S, Matsumoto K, Nagase H, et al. Interleukin-33 enhances adhesion, CD11b expression and survival in human eosinophils. Lab Invest 2008 Nov;88(11):1245-1253.

(197) Pushparaj PN, Tay HK, H'ng SC, Pitman N, Xu D, McKenzie A, et al. The cytokine interleukin-33 mediates anaphylactic shock. Proc Natl Acad Sci U S A 2009 Jun 16;106(24):9773-9778.

(198) Jung MY, Smrz D, Desai A, Bandara G, Ito T, Iwaki S, et al. IL-33 induces a hyporesponsive phenotype in human and mouse mast cells. J Immunol 2013 Jan 15;190(2):531-538.

(199) Kurowska-Stolarska M, Stolarski B, Kewin P, Murphy G, Corrigan CJ, Ying S, et al. IL-33 amplifies the polarization of alternatively activated macrophages that contribute to airway inflammation. J Immunol 2009 Nov 15;183(10):6469-6477.

(200) Brint EK, Xu D, Liu H, Dunne A, McKenzie AN, O'Neill LA, et al. ST2 is an inhibitor of interleukin 1 receptor and Toll-like receptor 4 signaling and maintains endotoxin tolerance. Nat Immunol 2004 Apr;5(4):373-379.

(201) Ohto-Ozaki H, Kuroiwa K, Mato N, Matsuyama Y, Hayakawa M, Tamemoto H, et al. Characterization of ST2 transgenic mice with resistance to IL-33. Eur J Immunol 2010 Sep;40(9):2632-2642.

(202) Kurokawa M, Matsukura S, Kawaguchi M, Ieki K, Suzuki S, Watanabe S, et al. Interleukin-33-activated dendritic cells induce the production of thymus and activation-regulated chemokine and macrophage-derived chemokine. Int Arch Allergy Immunol 2013;161 Suppl 2:52-57.

(203) Rank MA, Kobayashi T, Kozaki H, Bartemes KR, Squillace DL, Kita H. IL-33-activated dendritic cells induce an atypical TH2-type response. J Allergy Clin Immunol 2009 May;123(5):1047-1054.

(204) Barlow JL, Peel S, Fox J, Panova V, Hardman CS, Camelo A, et al. IL-33 is more potent than IL-25 in provoking IL-13-producing nuocytes (type 2 innate lymphoid cells) and airway contraction. J Allergy Clin Immunol 2013 Oct;132(4):933-941.

(205) Weinberg EO, Shimpo M, De Keulenaer GW, MacGillivray C, Tominaga S, Solomon SD, et al. Expression and regulation of ST2, an interleukin-1 receptor family member, in cardiomyocytes and myocardial infarction. Circulation 2002 Dec 3;106(23):2961-2966.

(206) Mildner M, Storka A, Lichtenauer M, Mlitz V, Ghannadan M, Hoetzenecker K, et al. Primary sources and immunological prerequisites for sST2 secretion in humans. Cardiovasc Res 2010 Sep 1;87(4):769-777.

(207) Coyle AJ, Lloyd C, Tian J, Nguyen T, Erikkson C, Wang L, et al. Crucial role of the interleukin 1 receptor family member T1/ST2 in T helper cell type 2-mediated lung mucosal immune responses. J Exp Med 1999 Oct 4;190(7):895-902.

(208) Neill DR, Wong SH, Bellosi A, Flynn RJ, Daly M, Langford TK, et al. Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. Nature 2010 Apr 29;464(7293):1367-1370.

(209) Spooner CJ, Lesch J, Yan D, Khan AA, Abbas A, Ramirez-Carrozzi V, et al. Specification of type 2 innate lymphocytes by the transcriptional determinant Gfi1. Nat Immunol 2013 Dec;14(12):1229-1236.

(210) Hayakawa H, Hayakawa M, Kume A, Tominaga S. Soluble ST2 blocks interleukin-33 signaling in allergic airway inflammation. J Biol Chem 2007 Sep 7;282(36):26369-26380.

(211) Fagundes CT, Amaral FA, Souza AL, Vieira AT, Xu D, Liew FY, et al. ST2, an IL-1R family member, attenuates inflammation and lethality after intestinal ischemia and reperfusion. J Leukoc Biol 2007 Feb;81(2):492-499.

(212) Spahn JD. Asthma biomarkers in sputum. Immunol Allergy Clin North Am 2007 Nov;27(4):607-22; vi.

(213) Lambrecht BN, De Veerman M, Coyle AJ, Gutierrez-Ramos JC, Thielemans K, Pauwels RA. Myeloid dendritic cells induce Th2 responses to inhaled antigen, leading to eosinophilic airway inflammation. J Clin Invest 2000 Aug;106(4):551-559.

(214) Lohning M, Stroehmann A, Coyle AJ, Grogan JL, Lin S, Gutierrez-Ramos JC, et al. T1/ST2 is preferentially expressed on murine Th2 cells, independent of interleukin 4, interleukin 5, and interleukin 10, and important for Th2 effector function. Proc Natl Acad Sci U S A 1998 Jun 9;95(12):6930-6935.

(215) Cho KA, Suh JW, Sohn JH, Park JW, Lee H, Kang JL, et al. IL-33 induces Th17-mediated airway inflammation via mast cells in ovalbumin-challenged mice. Am J Physiol Lung Cell Mol Physiol 2012 Feb 15;302(4):L429-40.

(216) Oshikawa K, Yanagisawa K, Tominaga S, Sugiyama Y. Expression and function of the ST2 gene in a murine model of allergic airway inflammation. Clin Exp Allergy 2002 Oct;32(10):1520-1526.

(217) Kearley J, Buckland KF, Mathie SA, Lloyd CM. Resolution of allergic inflammation and airway hyperreactivity is dependent upon disruption of the T1/ST2-IL-33 pathway. Am J Respir Crit Care Med 2009 May 1;179(9):772-781.

(218) Lee HY, Rhee CK, Kang JY, Byun JH, Choi JY, Kim SJ, et al. Blockade of IL-33/ST2 ameliorates airway inflammation in a murine model of allergic asthma. Exp Lung Res 2014 Jan 21.

(219) Kurokawa M, Matsukura S, Kawaguchi M, Ieki K, Suzuki S, Odaka M, et al. Expression and effects of IL-33 and ST2 in allergic bronchial asthma: IL-33 induces eotaxin production in lung fibroblasts. Int Arch Allergy Immunol 2011;155 Suppl 1:12-20.

(220) Oshikawa K, Kuroiwa K, Tago K, Iwahana H, Yanagisawa K, Ohno S, et al. Elevated soluble ST2 protein levels in sera of patients with asthma with an acute exacerbation. Am J Respir Crit Care Med 2001 Jul 15;164(2):277-281.

(221) Bianchetti L, Marini MA, Isgro M, Bellini A, Schmidt M, Mattoli S. IL-33 promotes the migration and proliferation of circulating fibrocytes from patients with allergen-exacerbated asthma. Biochem Biophys Res Commun 2012 Sep 14;426(1):116-121.

(222) Saglani S, Lui S, Ullmann N, Campbell GA, Sherburn RT, Mathie SA, et al. IL-33 promotes airway remodeling in pediatric patients with severe steroid-resistant asthma. J Allergy Clin Immunol 2013 Sep;132(3):676-685.e13.

(223) Ober C, Cox NJ, Abney M, Di Rienzo A, Lander ES, Changyaleket B, et al. Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma. Hum Mol Genet 1998 Sep;7(9):1393-1398.

(224) Postma DS, Meyers DA, Jongepier H, Howard TD, Koppelman GH, Bleecker ER. Genomewide screen for pulmonary function in 200 families ascertained for asthma. Am J Respir Crit Care Med 2005 Aug 15;172(4):446-452.

(225) Wjst M, Fischer G, Immervoll T, Jung M, Saar K, Rueschendorf F, et al. A genome-wide search for linkage to asthma. German Asthma Genetics Group. Genomics 1999 May 15;58(1):1-8.

(226) Zhu G, Whyte MK, Vestbo J, Carlsen K, Carlsen KH, Lenney W, et al. Interleukin 18 receptor 1 gene polymorphisms are associated with asthma. Eur J Hum Genet 2008 Sep;16(9):1083-1090.

(227) Sweet MJ, Leung BP, Kang D, Sogaard M, Schulz K, Trajkovic V, et al. A novel pathway regulating lipopolysaccharide-induced shock by ST2/T1 via inhibition of Toll-like receptor 4 expression. J Immunol 2001 Jun 1;166(11):6633-6639.

(228) Reijmerink NE, Bottema RW, Kerkhof M, Gerritsen J, Stelma FF, Thijs C, et al. TLR-related pathway analysis: novel gene-gene interactions in the development of asthma and atopy. Allergy 2010 Feb;65(2):199-207.

(229) Daley D, Lemire M, Akhabir L, Chan-Yeung M, He JQ, McDonald T, et al. Analyses of associations with asthma in four asthma population samples from Canada and Australia. Hum Genet 2009 May;125(4):445-459.

(230) Wang TT, Tavera-Mendoza LE, Laperriere D, Libby E, MacLeod NB, Nagai Y, et al. Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin D3 target genes. Mol Endocrinol 2005 Nov;19(11):2685-2695.

(231) Bosse Y, Lemire M, Poon AH, Daley D, He JQ, Sandford A, et al. Asthma and genes encoding components of the vitamin D pathway. Respir Res 2009 Oct 24;10:98.

(232) Buysschaert ID, Grulois V, Eloy P, Jorissen M, Rombaux P, Bertrand B, et al. Genetic evidence for a role of IL33 in nasal polyposis. Allergy 2010 May;65(5):616-622.

(233) Wu H, Romieu I, Shi M, Hancock DB, Li H, Sienra-Monge JJ, et al. Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. J Allergy Clin Immunol 2010 Feb;125(2):321-327.e13.

(234) Ramasamy A, Kuokkanen M, Vedantam S, Gajdos ZK, Couto Alves A, Lyon HN, et al. Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association near HLA. PLoS One 2012;7(9):e44008.

(235) Kato A, Schleimer RP. Beyond inflammation: airway epithelial cells are at the interface of innate and adaptive immunity. Curr Opin Immunol 2007 Dec;19(6):711-720.

(236) Mangan NE, Dasvarma A, McKenzie AN, Fallon PG. T1/ST2 expression on Th2 cells negatively regulates allergic pulmonary inflammation. Eur J Immunol 2007 May;37(5):1302-1312.

(237) Hacker S, Lambers C, Pollreisz A, Hoetzenecker K, Lichtenauer M, Mangold A, et al. Increased soluble serum markers caspase-cleaved cytokeratin-18, histones, and ST2 indicate apoptotic turnover and chronic immune response in COPD. J Clin Lab Anal 2009;23(6):372-379.

(238) Bonnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. Nat Genet 2013 Jun 30;Epub ahead of print.

(239) Shimizu M, Matsuda A, Yanagisawa K, Hirota T, Akahoshi M, Inomata N, et al. Functional SNPs in the distal promoter of the ST2 gene are associated with atopic dermatitis. Hum Mol Genet 2005 Oct 1;14(19):2919-2927.

(240) Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Sakashita M, et al. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. Nat Genet 2012 Nov;44(11):1222-1226.

(241) Castano R, Bosse Y, Endam LM, Desrosiers M. Evidence of association of interleukin-1 receptor-like 1 gene polymorphisms with chronic rhinosinusitis. Am J Rhinol Allergy 2009 Jul-Aug;23(4):377-384.

(242) Sakashita M, Yoshimoto T, Hirota T, Harada M, Okubo K, Osawa Y, et al. Association of serum interleukin-33 level and the interleukin-33 genetic variant with Japanese cedar pollinosis. Clin Exp Allergy 2008 Dec;38(12):1875-1881.

(243) Kamekura R, Kojima T, Takano K, Go M, Sawada N, Himi T. The role of IL-33 and its receptor ST2 in human nasal epithelium with allergic rhinitis. Clin Exp Allergy 2012 Feb;42(2):218-228.

(244) Kuroiwa K, Arai T, Okazaki H, Minota S, Tominaga S. Identification of human ST2 protein in the sera of patients with autoimmune diseases. Biochem Biophys Res Commun 2001 Jun 29;284(5):1104-1108.

(245) Mok MY, Huang FP, Ip WK, Lo Y, Wong FY, Chan EY, et al. Serum levels of IL-33 and soluble ST2 and their association with disease activity in systemic lupus erythematosus. Rheumatology (Oxford) 2010 Mar;49(3):520-527.

(246) Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 2009 Nov;41(11):1234-1237.

(247) Graham RR, Hom G, Ortmann W, Behrens TW. Review of recent genome-wide association scans in lupus. J Intern Med 2009 Jun;265(6):680-688.

(248) Leung BP, Xu D, Culshaw S, McInnes IB, Liew FY. A novel therapy of murine collagen-induced arthritis with soluble T1/ST2. J Immunol 2004 Jul 1;173(1):145-150.

(249) Palmer G, Talabot-Ayer D, Lamacchia C, Toy D, Seemayer CA, Viatte S, et al. Inhibition of interleukin-33 signaling attenuates the severity of experimental arthritis. Arthritis Rheum 2009 Mar;60(3):738-749.

(250) Matsuyama Y, Okazaki H, Tamemoto H, Kimura H, Kamata Y, Nagatani K, et al. Increased levels of interleukin 33 in sera and synovial fluid from patients with active rheumatoid arthritis. J Rheumatol 2010 Jan;37(1):18-25.

(251) Hong YS, Moon SJ, Joo YB, Jeon CH, Cho ML, Ju JH, et al. Measurement of interleukin-33 (IL-33) and IL-33 receptors (sST2 and ST2L) in patients with rheumatoid arthritis. J Korean Med Sci 2011 Sep;26(9):1132-1139.

(252) Xu D, Jiang HR, Li Y, Pushparaj PN, Kurowska-Stolarska M, Leung BP, et al. IL-33 exacerbates autoantibody-induced arthritis. J Immunol 2010 Mar 1;184(5):2620-2626.

(253) Beltran CJ, Nunez LE, Diaz-Jimenez D, Farfan N, Candia E, Heine C, et al. Characterization of the novel ST2/IL-33 system in patients with inflammatory bowel disease. Inflamm Bowel Dis 2010 Jul;16(7):1097-1107.

(254) Orozco G, Viatte S, Bowes J, Martin P, Wilson AG, Morgan AW, et al. Novel RA susceptibility locus at 22q12 identified in an extended UK genome wide association study. Arthritis Rheum 2013 Sep 24.

(255) Pastorelli L, Garg RR, Hoang SB, Spina L, Mattioli B, Scarpa M, et al. Epithelial-derived IL-33 and its receptor ST2 are dysregulated in ulcerative colitis and in experimental Th1/Th2 driven enteritis. Proc Natl Acad Sci U S A 2010 Apr 27;107(17):8017-8022.

(256) Zhernakova A, Festen EM, Franke L, Trynka G, van Diemen CC, Monsuur AJ, et al. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. Am J Hum Genet 2008 May;82(5):1202-1210.

(257) van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2007 Jul;39(7):827-829.

(258) Imielinski M, Baldassano RN, Griffiths A, Russell RK, Annese V, Dubinsky M, et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. Nat Genet 2009 Dec;41(12):1335-1340.

(259) Latiano A, Palmieri O, Pastorelli L, Vecchi M, Pizarro TT, Bossa F, et al. Associations between genetic polymorphisms in IL-33, IL1R1 and risk for inflammatory bowel disease. PLoS One 2013 Apr 25;8(4):e62144.

(260) Elbein SC, Das SK, Hallman DM, Hanis CL, Hasstedt SJ. Genome-wide linkage and admixture mapping of type 2 diabetes in African American families from the American Diabetes Association GENNID (Genetics of NIDDM) Study Cohort. Diabetes 2009 Jan;58(1):268-274.

(261) Mensah-Brown E, Shahin A, Parekh K, Hakim AA, Shamisi MA, Hsu DK, et al. Functional capacity of macrophages determines the induction of type 1 diabetes. Ann N Y Acad Sci 2006 Nov;1084:49-57.

(262) Fousteris E, Melidonis A, Panoutsopoulos G, Tzirogiannis K, Foussas S, Theodosis-Georgilas A, et al. Toll/interleukin-1 receptor member ST2 exhibits higher soluble levels in type 2 diabetes, especially when accompanied with left ventricular diastolic dysfunction. Cardiovasc Diabetol 2011 Nov 21;10:101-2840-10-101.

(263) Miller AM, Purves D, McConnachie A, Asquith DL, Batty GD, Burns H, et al. Soluble ST2 associates with diabetes but not established cardiovascular risk factors: a new inflammatory pathway of relevance to diabetes? PLoS One 2012;7(10):e47830.

(264) Seki K, Sanada S, Kudinova AY, Steinhauser ML, Handa V, Gannon J, et al. Interleukin-33 prevents apoptosis and improves survival after experimental myocardial infarction through ST2 signaling. Circ Heart Fail 2009 Nov;2(6):684-691.

(265) Abston ED, Barin JG, Cihakova D, Bucek A, Coronado MJ, Brandt JE, et al. IL-33 independently induces eosinophilic pericarditis and cardiac dilation: ST2 improves cardiac function. Circ Heart Fail 2012 May 1;5(3):366-375.

(266) Weinberg EO, Shimpo M, Hurwitz S, Tominaga S, Rouleau JL, Lee RT. Identification of serum soluble ST2 receptor as a novel heart failure biomarker. Circulation 2003 Feb 11;107(5):721-726.

(267) Bartunek J, Delrue L, Van Durme F, Muller O, Casselman F, De Wiest B, et al. Nonmyocardial production of ST2 protein in human hypertrophy and failure is related to diastolic load. J Am Coll Cardiol 2008 Dec 16;52(25):2166-2174.

(268) Shah RV, Chen-Tournoux AA, Picard MH, van Kimmenade RR, Januzzi JL. Serum levels of the interleukin-1 receptor family member ST2, cardiac structure and function, and long-term mortality in patients with acute dyspnea. Circ Heart Fail 2009 Jul;2(4):311-319.

(269) Dieplinger B, Gegenhuber A, Kaar G, Poelz W, Haltmayer M, Mueller T. Prognostic value of established and novel biomarkers in patients with shortness of breath attending an emergency department. Clin Biochem 2010 Jun;43(9):714-719.

(270) Chen LQ, de Lemos JA, Das SR, Ayers CR, Rohatgi A. Soluble ST2 is associated with all-cause and cardiovascular mortality in a population-based cohort: the Dallas Heart Study. Clin Chem 2013 Mar;59(3):536-546.

(271) Pascual-Figal DA, Ordonez-Llanos J, Tornel PL, Vazquez R, Puig T, Valdes M, et al. Soluble ST2 for predicting sudden cardiac death in patients with chronic heart failure and left ventricular systolic dysfunction. J Am Coll Cardiol 2009 Dec 1;54(23):2174-2179.

(272) Shimpo M, Morrow DA, Weinberg EO, Sabatine MS, Murphy SA, Antman EM, et al. Serum levels of the interleukin-1 receptor family member ST2 predict mortality and clinical outcome in acute myocardial infarction. Circulation 2004 May 11;109(18):2186-2190.

(273) Moore SA, Januzzi JL,Jr. Found in translation soluble ST2 and heart disease. J Am Coll Cardiol 2010 Jan 19;55(3):251-253.

(274) Weir RA, Miller AM, Murphy GE, Clements S, Steedman T, Connell JM, et al. Serum soluble ST2: a potential novel mediator in left ventricular and infarct remodeling after acute myocardial infarction. J Am Coll Cardiol 2010 Jan 19;55(3):243-250.

(275) Mueller T, Dieplinger B. The Presage((R)) ST2 Assay: analytical considerations and clinical applications for a high-sensitivity assay for measurement of soluble ST2. Expert Rev Mol Diagn 2013 Jan;13(1):13-30.

(276) Alves-Filho JC, Sonego F, Souto FO, Freitas A, Verri WA,Jr, Auxiliadora-Martins M, et al. Interleukin-33 attenuates sepsis by enhancing neutrophil influx to the site of infection. Nat Med 2010 Jun;16(6):708-712.

(277) Hoogerwerf JJ, Leendertse M, Wieland CW, de Vos AF, de Boer JD, Florquin S, et al. Loss of suppression of tumorigenicity 2 (ST2) gene reverses sepsis-induced inhibition of lung host defense in mice. Am J Respir Crit Care Med 2011 Apr 1;183(7):932-940.

(278) Buckley JM, Liu JH, Li CH, Blankson S, Wu QD, Jiang Y, et al. Increased susceptibility of ST2-deficient mice to polymicrobial sepsis is associated with an impaired bactericidal function. J Immunol 2011 Oct 15;187(8):4293-4299.

(279) Jones LA, Roberts F, Nickdel MB, Brombacher F, McKenzie AN, Henriquez FL, et al. IL-33 receptor (T1/ST2) signalling is necessary to prevent the development of encephalitis in mice infected with Toxoplasma gondii. Eur J Immunol 2010 Feb;40(2):426-436.

(280) Hoogerwerf JJ, Tanck MW, van Zoelen MA, Wittebole X, Laterre PF, van der Poll T. Soluble ST2 plasma concentrations predict mortality in severe sepsis. Intensive Care Med 2010 Apr;36(4):630-637.

(281) Parenica J, Malaska J, Jarkovsky J, Lipkova J, Dastych M, Helanova K, et al. Soluble ST2 levels in patients with cardiogenic and septic shock are not predictors of mortality. Exp Clin Cardiol 2012 Winter;17(4):205-209.

(282) Dhiman N, Ovsyannikova IG, Vierkant RA, Pankratz VS, Jacobson RM, Poland GA. Associations between cytokine/cytokine receptor single nucleotide polymorphisms and humoral immunity to measles, mumps and rubella in a Somali population. Tissue Antigens 2008 Sep;72(3):211-220.

(283) Imada Y, Fujimoto M, Hirata K, Hirota T, Suzuki Y, Saito H, et al. Large scale genotyping study for asthma in the Japanese population. BMC Res Notes 2009 Mar 31;2:54.

(284) Humphreys NE, Xu D, Hepworth MR, Liew FY, Grencis RK. IL-33, a potent inducer of adaptive immunity to intestinal nematodes. J Immunol 2008 Feb 15;180(4):2443-2449.

(285) Yasuda K, Muto T, Kawagoe T, Matsumoto M, Sasaki Y, Matsushita K, et al. Contribution of IL-33-activated type II innate lymphoid cells to pulmonary eosinophilia in intestinal nematode-infected mice. Proc Natl Acad Sci U S A 2012 Feb 28;109(9):3451-3456.

(286) Sesti-Costa R, Silva GK, Proenca-Modena JL, Carlos D, Silva ML, Alves-Filho JC, et al. The IL-33/ST2 pathway controls coxsackievirus B5-induced experimental pancreatitis. J Immunol 2013 Jul 1;191(1):283-292.

(287) Saito T, Ji G, Shinzawa H, Okumoto K, Hattori E, Adachi T, et al. Genetic variations in humans associated with differences in the course of hepatitis C. Biochem Biophys Res Commun 2004 Apr 30;317(2):335-341.

(288) Wang J, Zhao P, Guo H, Sun X, Jiang Z, Xu L, et al. Serum IL-33 levels are associated with liver damage in patients with chronic hepatitis C. Mediators Inflamm 2012;2012:819636.

(289) Roth GA, Zimmermann M, Lubsczyk BA, Pilz J, Faybik P, Hetz H, et al. Up-regulation of interleukin 33 and soluble ST2 serum levels in liver failure. J Surg Res 2010 Oct;163(2):e79-83.

(290) Milovanovic M, Volarevic V, Radosavljevic G, Jovanovic I, Pejnovic N, Arsenijevic N, et al. IL-33/ST2 axis in inflammation and immunopathology. Immunol Res 2012 Apr;52(1-2):89-99.

(291) Marvie P, Lisbonne M, L'helgoualc'h A, Rauch M, Turlin B, Preisser L, et al. Interleukin-33 overexpression is associated with liver fibrosis in mice and humans. J Cell Mol Med 2010 Jun;14(6B):1726-1739.

(292) Akcay A, Nguyen Q, He Z, Turkmen K, Won Lee D, Hernando AA, et al. IL-33 exacerbates acute kidney injury. J Am Soc Nephrol 2011 Nov;22(11):2057-2067.

(293) Bao YS, Na SP, Zhang P, Jia XB, Liu RC, Yu CY, et al. Characterization of interleukin-33 and soluble ST2 in serum and their association with disease severity in patients with chronic kidney disease. J Clin Immunol 2012 Jun;32(3):587-594.

(294) Ho JE, Hwang SJ, Wollert KC, Larson MG, Cheng S, Kempf T, et al. Biomarkers of cardiovascular stress and incident chronic kidney disease. Clin Chem 2013 Nov;59(11):1613-1620.

(295) Friend SL, Hosier S, Nelson A, Foxworthe D, Williams DE, Farr A. A thymic stromal cell line supports in vitro development of surface IgM+ B cells and produces a novel growth factor affecting B and T lineage cells. Exp Hematol 1994 Mar;22(3):321-328.

(296) Headley MB, Zhou B, Shih WX, Aye T, Comeau MR, Ziegler SF. TSLP conditions the lung immune environment for the generation of pathogenic innate and antigen-specific adaptive immune responses. J Immunol 2009 Feb 1;182(3):1641-1647.

(297) Reche PA, Soumelis V, Gorman DM, Clifford T, Liu M, Travis M, et al. Human thymic stromal lymphopoietin preferentially stimulates myeloid cells. J Immunol 2001 Jul 1;167(1):336-343.

(298) Uller L, Leino M, Bedke N, Sammut D, Green B, Lau L, et al. Double-stranded RNA induces disproportionate expression of thymic stromal lymphopoietin versus interferon-beta in bronchial epithelial cells from donors with asthma. Thorax 2010 Jul;65(7):626-632.

(299) Miazgowicz MM, Elliott MS, Debley JS, Ziegler SF. Respiratory syncytial virus induces functional thymic stromal lymphopoietin receptor in airway epithelial cells. J Inflamm Res 2013;6:53-61.

(300) Kato A, Favoreto S,Jr, Avila PC, Schleimer RP. TLR3- and Th2 cytokine-dependent production of thymic stromal lymphopoietin in human airway epithelial cells. J Immunol 2007 Jul 15;179(2):1080-1087.

(301) Nakamura Y, Miyata M, Ohba T, Ando T, Hatsushika K, Suenaga F, et al. Cigarette smoke extract induces thymic stromal lymphopoietin expression, leading to T(H)2-type immune responses and airway inflammation. J Allergy Clin Immunol 2008 Dec;122(6):1208-1214.

(302) Bleck B, Tse DB, Gordon T, Ahsan MR, Reibman J. Diesel exhaust particle-treated human bronchial epithelial cells upregulate Jagged-1 and OX40 ligand in myeloid dendritic cells via thymic stromal lymphopoietin. J Immunol 2010 Dec 1;185(11):6636-6645.

(303) Hammad H, Chieppa M, Perros F, Willart MA, Germain RN, Lambrecht BN. House dust mite allergen induces asthma via Toll-like receptor 4 triggering of airway structural cells. Nat Med 2009 Apr;15(4):410-416.

(304) Kouzaki H, O'Grady SM, Lawrence CB, Kita H. Proteases induce production of thymic stromal lymphopoietin by airway epithelial cells through protease-activated receptor-2. J Immunol 2009 Jul 15;183(2):1427-1434.

(305) Li DQ, Zhang L, Pflugfelder SC, De Paiva CS, Zhang X, Zhao G, et al. Short ragweed pollen triggers allergic inflammation through Toll-like receptor 4-dependent thymic stromal lymphopoietin/OX40 ligand/OX40 signaling pathways. J Allergy Clin Immunol 2011 Dec;128(6):1318-1325.e2.

(306) Yamashita S, Segawa R, Satou N, Hiratsuka M, Leonard WJ, Hirasawa N. Induction of Thymic Stromal Lymphopoietin Production by Nonanoic Acid and Exacerbation of Allergic Inflammation in Mice. Allergol Int 2013 Sep 25.

(307) Satou N, Ishihara K, Hiratsuka M, Tanaka H, Endo Y, Saito S, et al. Induction of thymic stromal lymphopoietin production by xylene and exacerbation of picryl chloride-induced allergic inflammation in mice. Int Arch Allergy Immunol 2012;157(2):194-201.

(308) Redhu NS, Saleh A, Halayko AJ, Ali AS, Gounni AS. Essential role of NF-kappaB and AP-1 transcription factors in TNF-alpha-induced TSLP expression in human airway smooth muscle cells. Am J Physiol Lung Cell Mol Physiol 2011 Mar;300(3):L479-85.

(309) Moon PD, Kim HM. Thymic stromal lymphopoietin is expressed and produced by caspase-1/NF-kappaB pathway in mast cells. Cytokine 2011 Jun;54(3):239-243.

(310) Pandey A, Ozaki K, Baumann H, Levin SD, Puel A, Farr AG, et al. Cloning of a receptor subunit required for signaling by thymic stromal lymphopoietin. Nat Immunol 2000 Jul;1(1):59-64.

(311) Ito T, Wang YH, Duramad O, Hori T, Delespesse GJ, Watanabe N, et al. TSLP-activated dendritic cells induce an inflammatory T helper type 2 cell response through OX40 ligand. J Exp Med 2005 Nov 7;202(9):1213-1223.

(312) Scheeren FA, van Lent AU, Nagasawa M, Weijer K, Spits H, Legrand N, et al. Thymic stromal lymphopoietin induces early human B-cell proliferation and differentiation. Eur J Immunol 2010 Apr;40(4):955-965.

(313) Omori M, Ziegler S. Induction of IL-4 expression in CD4(+) T cells by thymic stromal lymphopoietin. J Immunol 2007 Feb 1;178(3):1396-1404.

(314) Wong CK, Hu S, Cheung PF, Lam CW. Thymic stromal lymphopoietin induces chemotactic and prosurvival effects in eosinophils: implications in allergic inflammation. Am J Respir Cell Mol Biol 2010 Sep;43(3):305-315.

(315) Han H, Headley MB, Xu W, Comeau MR, Zhou B, Ziegler SF. Thymic stromal lymphopoietin amplifies the differentiation of alternatively activated macrophages. J Immunol 2013 Feb 1;190(3):904-912.

(316) Redhu NS, Shan L, Movassagh H, Gounni AS. Thymic stromal lymphopoietin induces migration in human airway smooth muscle cells. Sci Rep 2013;3:2301.

(317) Allakhverdi Z, Comeau MR, Jessup HK, Delespesse G. Thymic stromal lymphopoietin as a mediator of crosstalk between bronchial smooth muscles and mast cells. J Allergy Clin Immunol 2009 Apr;123(4):958-60.e2.

(318) Isaksen DE, Baumann H, Trobridge PA, Farr AG, Levin SD, Ziegler SF. Requirement for stat5 in thymic stromal lymphopoietin-mediated signal transduction. J Immunol 1999 Dec 1;163(11):5971-5977.

(319) Wu J, Liu F, Zhao J, Wei Y, Lv J, Dong F, et al. Thymic stromal lymphopoietin promotes asthmatic airway remodelling in human lung fibroblast cells through STAT3 signalling pathway. Cell Biochem Funct 2013 Aug;31(6):496-503.

(320) Zhong J, Kim MS, Chaerkady R, Wu X, Huang TC, Getnet D, et al. TSLP signaling network revealed by SILAC-based phosphoproteomics. Mol Cell Proteomics 2012 Jun;11(6):M112.017764.

(321) Arima K, Watanabe N, Hanabuchi S, Chang M, Sun SC, Liu YJ. Distinct signal codes generate dendritic cell functional plasticity. Sci Signal 2010 Jan 19;3(105):ra4.

(322) Taylor BC, Zaph C, Troy AE, Du Y, Guild KJ, Comeau MR, et al. TSLP regulates intestinal immunity and inflammation in mouse models of helminth infection and colitis. J Exp Med 2009 Mar 16;206(3):655-667.

(323) Zeuthen LH, Fink LN, Frokiaer H. Epithelial cells prime the immune response to an array of gut-derived commensals towards a tolerogenic phenotype through distinct actions of thymic stromal lymphopoietin and transforming growth factor-beta. Immunology 2008 Feb;123(2):197-208.

(324) Spadoni I, Iliev ID, Rossi G, Rescigno M. Dendritic cells produce TSLP that limits the differentiation of Th17 cells, fosters Treg development, and protects against colitis. Mucosal Immunol 2012 Mar;5(2):184-193.

(325) Du MR, Guo PF, Piao HL, Wang SC, Sun C, Jin LP, et al. Embryonic Trophoblasts Induce Decidual Regulatory T Cell Differentiation and Maternal-Fetal Tolerance through Thymic Stromal Lymphopoietin Instructing Dendritic Cells. J Immunol 2014 Jan 22.

(326) Al-Shami A, Spolski R, Kelly J, Fry T, Schwartzberg PL, Pandey A, et al. A role for thymic stromal lymphopoietin in CD4(+) T cell development. J Exp Med 2004 Jul 19;200(2):159-168.

(327) Hanabuchi S, Watanabe N, Liu YJ. TSLP and immune homeostasis. Allergol Int 2012 Mar;61(1):19-25.

(328) Rochman Y, Leonard WJ. The role of thymic stromal lymphopoietin in CD8+ T cell homeostasis. J Immunol 2008 Dec 1;181(11):7699-7705.

(329) Zhou B, Comeau MR, De Smedt T, Liggitt HD, Dahl ME, Lewis DB, et al. Thymic stromal lymphopoietin as a key initiator of allergic airway inflammation in mice. Nat Immunol 2005 Oct;6(10):1047-1053.

(330) Yoo J, Omori M, Gyarmati D, Zhou B, Aye T, Brewer A, et al. Spontaneous atopic dermatitis in mice expressing an inducible thymic stromal lymphopoietin transgene specifically in the skin. J Exp Med 2005 Aug 15;202(4):541-549.

(331) Al-Shami A, Spolski R, Kelly J, Keane-Myers A, Leonard WJ. A role for TSLP in the development of inflammation in an asthma model. J Exp Med 2005 Sep 19;202(6):829-839.

(332) Miyata M, Hatsushika K, Ando T, Shimokawa N, Ohnuma Y, Katoh R, et al. Mast cell regulation of epithelial TSLP expression plays an important role in the development of allergic rhinitis. Eur J Immunol 2008 Jun;38(6):1487-1492.

(333) Demehri S, Morimoto M, Holtzman MJ, Kopan R. Skin-derived TSLP triggers progression from epidermal-barrier defects to asthma. PLoS Biol 2009 May 19;7(5):e1000067.

(334) Han H, Xu W, Headley MB, Jessup HK, Lee KS, Omori M, et al. Thymic stromal lymphopoietin (TSLP)-mediated dermal inflammation aggravates experimental asthma. Mucosal Immunol 2012 May;5(3):342-351.

(335) Suzuki Y, Kodama M, Asano K. Skin barrier-related molecules and pathophysiology of asthma. Allergol Int 2011 Mar;60(1):11-15.

(336) Ying S, O'Connor B, Ratoff J, Meng Q, Mallett K, Cousins D, et al. Thymic stromal lymphopoietin expression is increased in asthmatic airways and correlates with expression of Th2-attracting chemokines and disease severity. J Immunol 2005 Jun 15;174(12):8183-8190.

(337) Shikotra A, Choy DF, Ohri CM, Doran E, Butler C, Hargadon B, et al. Increased expression of immunoreactive thymic stromal lymphopoietin in patients with severe asthma. J Allergy Clin Immunol 2012 Jan;129(1):104-11.e1-9.

(338) Nguyen KD, Vanichsarn C, Nadeau KC. TSLP directly impairs pulmonary Treg function: association with aberrant tolerogenic immunity in asthmatic airway. Allergy Asthma Clin Immunol 2010 Mar 15;6(1):4-1492-6-4.

(339) Xystrakis E, Kusumakar S, Boswell S, Peek E, Urry Z, Richards DF, et al. Reversing the defective induction of IL-10-secreting regulatory T cells in glucocorticoid-resistant asthma patients. J Clin Invest 2006 Jan;116(1):146-155.

(340) Nagarkar DR, Poposki JA, Tan BK, Comeau MR, Peters AT, Hulse KE, et al. Thymic stromal lymphopoietin activity is increased in nasal polyps of patients with chronic rhinosinusitis. J Allergy Clin Immunol 2013 May 17.

(341) Taneda S, Segerer S, Hudkins KL, Cui Y, Wen M, Segerer M, et al. Cryoglobulinemic glomerulonephritis in thymic stromal lymphopoietin transgenic mice. Am J Pathol 2001 Dec;159(6):2355-2369.

(342) Masuko H, Sakamoto T, Kaneko Y, Iijima H, Naito T, Noguchi E, et al. Lower FEV1 in non-COPD, nonasthmatic subjects: association with smoking, annual decline in FEV1, total IgE levels, and TSLP genotypes. Int J Chron Obstruct Pulmon Dis 2011;6:181-189.

(343) Bunyavanich S, Melen E, Wilk JB, Granada M, Soto-Quiros ME, Avila L, et al. Thymic stromal lymphopoietin (TSLP) is associated with allergic rhinitis in children with asthma. Clin Mol Allergy 2011 Jan 18;9:1-7961-9-1.

(344) Ferreira MA, Matheson MC, Tang CS, Granell R, Ang W, Hui J, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol 2013 Dec 30.

(345) Iliev ID, Mileti E, Matteoli G, Chieppa M, Rescigno M. Intestinal epithelial cells promote colitis-protective regulatory T-cell differentiation through dendritic cell conditioning. Mucosal Immunol 2009 Jul;2(4):340-350.

(346) Hanabuchi S, Ito T, Park WR, Watanabe N, Shaw JL, Roman E, et al. Thymic stromal lymphopoietin-activated plasmacytoid dendritic cells induce the generation of FOXP3+ regulatory T cells in human thymus. J Immunol 2010 Mar 15;184(6):2999-3007.

(347) Iliev ID, Spadoni I, Mileti E, Matteoli G, Sonzogni A, Sampietro GM, et al. Human intestinal epithelial cells promote the differentiation of tolerogenic dendritic cells. Gut 2009 Nov;58(11):1481-1489.

(348) Fuss IJ. Is the Th1/Th2 paradigm of immune regulation applicable to IBD? Inflamm Bowel Dis 2008 Oct;14 Suppl 2:S110-2.

(349) Noble CL, Abbas AR, Lees CW, Cornelius J, Toy K, Modrusan Z, et al. Characterization of intestinal gene expression profiles in Crohn's disease by genome-wide microarray analysis. Inflamm Bowel Dis 2010 Oct;16(10):1717-1728.

(350) Nishiura H, Kido M, Aoki N, Iwamoto S, Maruoka R, Ikeda A, et al. Increased susceptibility to autoimmune gastritis in thymic stromal lymphopoietin receptor-deficient mice. J Immunol 2012 Jan 1;188(1):190-197.

(351) Hartgring SA, Willis CR, Dean CE,Jr, Broere F, van Eden W, Bijlsma JW, et al. Critical proinflammatory role of thymic stromal lymphopoietin and its receptor in experimental autoimmune arthritis. Arthritis Rheum 2011 Jul;63(7):1878-1887.

(352) Christmann RB, Mathes A, Affandi AJ, Padilla C, Nazari B, Bujor AM, et al. Thymic stromal lymphopoietin is up-regulated in the skin of patients with systemic sclerosis and induces profibrotic genes and intracellular signaling that overlap with those induced by

interleukin-13 and transforming growth factor beta. Arthritis Rheum 2013 May;65(5):1335-1346.

(353) Usategui A, Criado G, Izquierdo E, Del Rey MJ, Carreira PE, Ortiz P, et al. A profibrotic role for thymic stromal lymphopoietin in systemic sclerosis. Ann Rheum Dis 2013 Feb 14.

(354) Tanaka J, Watanabe N, Kido M, Saga K, Akamatsu T, Nishio A, et al. Human TSLP and TLR3 ligands promote differentiation of Th17 cells with a central memory phenotype under Th2-polarizing conditions. Clin Exp Allergy 2009 Jan;39(1):89-100.

(355) Lee HC, Sung SS, Krueger PD, Jo YA, Rosen HR, Ziegler SF, et al. Hepatitis C virus promotes T-helper (Th)17 responses through thymic stromal lymphopoietin production by infected hepatocytes. Hepatology 2013 Apr;57(4):1314-1324.

(356) Lin J, Chang W, Dong J, Zhang F, Mohabeer N, Kushwaha KK, et al. Thymic stromal lymphopoietin over-expressed in human atherosclerosis: potential role in Th17 differentiation. Cell Physiol Biochem 2013;31(2-3):305-318.

(357) Tsai KH, Tsai FJ, Lin HJ, Lin HJ, Liu YH, Liao WL, et al. Thymic stromal lymphopoietin gene promoter polymorphisms and expression levels in Graves' disease and Graves' ophthalmopathy. BMC Med Genet 2012 Nov 30;13:116-2350-13-116.

(358) Olkhanud PB, Rochman Y, Bodogai M, Malchinkhuu E, Wejksza K, Xu M, et al. Thymic stromal lymphopoietin is a key mediator of breast cancer progression. J Immunol 2011 May 15;186(10):5656-5662.

(359) De Monte L, Reni M, Tassi E, Clavenna D, Papa I, Recalde H, et al. Intratumor T helper type 2 cell infiltrate correlates with cancer-associated fibroblast thymic stromal lymphopoietin production and reduced survival in pancreatic cancer. J Exp Med 2011 Mar 14;208(3):469-478.

(360) Chen E, Chim LS, Strunk RC, Miller GE. The role of the social environment in children and adolescents with asthma. Am J Respir Crit Care Med 2007 Oct 1;176(7):644-649.

(361) Becker AB, Manfreda J, Ferguson AC, Dimich-Ward H, Watson WT, Chan-Yeung M. Breast-feeding and environmental tobacco smoke exposure. Arch Pediatr Adolesc Med 1999 Jul;153(7):689-691.

(362) Chan-Yeung M, Manfreda J, Dimich-Ward H, Ferguson A, Watson W, Becker A. A randomized controlled study on the effectiveness of a multifaceted intervention program in the primary prevention of asthma in high-risk infants. Arch Pediatr Adolesc Med 2000 Jul;154(7):657-663.

(363) Chan-Yeung M, Ferguson A, Watson W, Dimich-Ward H, Rousseau R, Lilley M, et al. The Canadian Childhood Asthma Primary Prevention Study: outcomes at 7 years of age. J Allergy Clin Immunol 2005 Jul;116(1):49-55.

(364) Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol Lett 2004 Mar;26(6):509-515.

(365) Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002 Jun 18;3(7):RESEARCH0034.

(366) Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 2004 Aug 1;64(15):5245-5250.

(367) Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics 2007 May 15;23(10):1289-1291.

(368) Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. Nucleic Acids Res 2012 Aug;40(15):e115.

(369) Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, et al. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res 1998 Jan 1;26(1):362-367.

(370) Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 2002 Feb;18(2):333-334.

(371) Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 2004 Jan 1;32(Database issue):D91-4.

(372) Ghosh D. Object-oriented transcription factors database (ooTFD). Nucleic Acids Res 2000 Jan 1;28(1):308-310.

(373) Sandelin A, Wasserman WW, Lenhard B. ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res 2004 Jul 1;32(Web Server issue):W249-52.

(374) Akiyama Y. TFSEARCH: Searching Transcription Factor Binding Sites. Available at: http://www.rwcp.or.jp/papia/.

(375) Bosse Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, et al. Molecular signature of smoking in human lung tissues. Cancer Res 2012 Aug 1;72(15):3753-3763.

(376) Bosse Y, Sin D, Laviolette M, Sandford AJ, Hogg J, Daley D, et al. Hypothesis-driven Research on Genomic Data Derived from a Large Scale Lung EQTL Mapping Study WebmedCentral LUNG 2010;1(9):WMC00724.

(377) Gauderman WJ, Morrison JM. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006; Available at: http://biostats.usc.edu/software.

(378) Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theroretic approach. : Springer; 2002.

(379) ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012 Sep 6;489(7414):57-74.

(380) Gu CC, Yu K, Rao DC. Characterization of LD Structures and the Utility of HapMap in Genetic Association Studies. Advances in Genetics: Academic Press; 2008. p. 407-435.

(381) Pettersson FH, Anderson CA, Clarke GM, Barrett JC, Cardon LR, Morris AP, et al. Marker selection for genetic case-control association studies. Nat Protoc 2009;4(5):743-752.

(382) Iwahana H, Yanagisawa K, Ito-Kosaka A, Kuroiwa K, Tago K, Komatsu N, et al. Different promoter usage and multiple transcription initiation sites of the interleukin-1 receptor-related human ST2 gene in UT-7 and TM12 cells. Eur J Biochem 1999 Sep;264(2):397-406.

(383) Faber TE, Schuurhof A, Vonk A, Koppelman GH, Hennus MP, Kimpen JL, et al. IL1RL1 Gene Variants and Nasopharyngeal IL1RL-a Levels Are Associated with Severe RSV Bronchiolitis: A Multicenter Cohort Study. PLoS One 2012;7(5):e34364.

(384) Savenije OE, Kerkhof M, Reijmerink NE, Brunekreef B, de Jongste JC, Smit HA, et al. Interleukin-1 receptor-like 1 polymorphisms are associated with serum IL1RL1-a, eosinophils, and asthma in childhood. J Allergy Clin Immunol 2011 Mar;127(3):750-6.e1-5.

(385) Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009 Jul;37(Web Server issue):W600-5.

(386) King DC, Taylor J, Elnitski L, Chiaromonte F, Miller W, Hardison RC. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res 2005 Aug;15(8):1051-1060.

(387) Sunyaev S, Ramensky V, Koch I, Lathe W,3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet 2001 Mar 15;10(6):591-597.

(388) Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, Chiaromonte F. ESPERR: Learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Research 2006 December 01;16(12):1596-1604.

(389) Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012 Jan;40(Database issue):D930-4.

(390) Woo JH, Cho SB, Lee E, Kim JH. Identifying regulatory relationships among genomic loci, biological pathways, and disease. Artif Intell Med 2010 7;49(3):161-165.

(391) Fridley BL, Iversen E, Tsai YY, Jenkins GD, Goode EL, Sellers TA. A latent model for prioritization of SNPs for functional studies. PLoS One 2011;6(6):e20764.

(392) Saccone SF, Saccone NL, Swan GE, Madden PAF, Goate AM, Rice JP, et al. Systematic biological prioritization after a genome-wide association study: an application to nicotine dependence. Bioinformatics 2008 August 15;24(16):1805-1811.

(393) Hsu YH, Zillikens MC, Wilson SG, Farber CR, Demissie S, Soranzo N, et al. An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. PLoS Genet 2010 Jun 10;6(6):e1000977.

(394) Carter RW, Sweet MJ, Xu D, Klemenz R, Liew FY, Chan WL. Regulation of ST2L expression on T helper (Th) type 2 cells. Eur J Immunol 2001 Oct;31(10):2979-2985.

(395) Lipsky BP, Toy DY, Swart DA, Smithgall MD, Smith D. Deletion of the ST2 proximal promoter disrupts fibroblast-specific expression but does not reduce the amount of soluble ST2 in circulation. Eur J Immunol 2012 Jul;42(7):1863-1869.

(396) Hakim O, Sung MH, Hager GL. 3D shortcuts to gene regulation. Curr Opin Cell Biol 2010 Jun;22(3):305-313.

(397) Xu Z, Lefevre GM, Felsenfeld G. Chromatin structure, epigenetic mechanisms and long-range interactions in the human insulin locus. Diabetes Obes Metab 2012 Oct;14 Suppl 3:1-11.

(398) Palstra RJ, de Laat W, Grosveld F. Beta-globin regulation and long-range interactions. Adv Genet 2008;61:107-142.

(399) Ramamoorthy S, Cidlowski JA. Ligand-induced repression of the glucocorticoid receptor gene is mediated by an NCoR1 repression complex formed by long-range chromatin

interactions with intragenic glucocorticoid response elements. Mol Cell Biol 2013 May;33(9):1711-1722.

(400) Sajan SA, Hawkins RD. Methods for identifying higher-order chromatin structure. Annu Rev Genomics Hum Genet 2012;13:59-82.

(401) Hou C, Corces VG. Throwing transcription for a loop: expression of the genome in the 3D nucleus. Chromosoma 2012 Apr;121(2):107-116.

(402) Merkenschlager M, Odom DT. CTCF and cohesin: linking gene regulatory elements with their targets. Cell 2013 Mar 14;152(6):1285-1297.

(403) Roos AB, Barton JL, Miller-Larsson A, Dahlberg B, Berg T, Didon L, et al. Lung epithelial-C/EBPbeta contributes to LPS-induced inflammation and its suppression by formoterol. Biochem Biophys Res Commun 2012 Jun 22;423(1):134-139.

(404) Gao H, Ward PA. STAT3 and suppressor of cytokine signaling 3: potential targets in lung inflammatory responses. Expert Opin Ther Targets 2007 Jul;11(7):869-880.

(405) Mooney SD, Krishnan VG, Evani US. Bioinformatic tools for identifying disease gene and SNP candidates. Methods Mol Biol 2010;628:307-319.

(406) Coassin S, Brandstatter A, Kronenberg F. Lost in the space of bioinformatic tools: a constantly updated survival guide for genetic epidemiology. The GenEpi Toolbox. Atherosclerosis 2010 Apr;209(2):321-335.

(407) Ho JE, Chen WY, Chen MH, Larson MG, McCabe EL, Cheng S, et al. Common genetic variation at the IL1RL1 locus regulates IL-33/ST2 signaling. J Clin Invest 2013 Oct 1;123(10):4208-4218.

(408) Carterson AJ, Honer zu Bentrup K, Ott CM, Clarke MS, Pierson DL, Vanderburg CR, et al. A549 lung epithelial cells grown as three-dimensional aggregates: alternative tissue culture model for Pseudomonas aeruginosa pathogenesis. Infect Immun 2005 Feb;73(2):1129-1140.

(409) Berger JT, Voynow JA, Peters KW, Rose MC. Respiratory carcinoma cell lines. MUC genes and glycoconjugates. Am J Respir Cell Mol Biol 1999 Mar;20(3):500-510.

(410) Winton HL, Wan H, Cannell MB, Gruenert DC, Thompson PJ, Garrod DR, et al. Cell lines of pulmonary and non-pulmonary origin as tools to study the effects of house dust mite proteinases on the regulation of epithelial permeability. Clin Exp Allergy 1998 Oct;28(10):1273-1285.

(411) Sahlander K, Larsson K, Palmberg L. Increased serum levels of soluble ST2 in birch pollen atopics and individuals working in laboratory animal facilities. J Occup Environ Med 2010 Feb;52(2):214-218.

(412) Baumann R, Rabaszowski M, Stenin I, Tilgner L, Gaertner-Akerboom M, Scheckenbach K, et al. Nasal levels of soluble IL-33R ST2 and IL-16 in allergic rhinitis: inverse correlation trends with disease severity. Clin Exp Allergy 2013 Oct;43(10):1134-1143.

(413) Hamzaoui A, Berraies A, Kaabachi W, Haifa M, Ammar J, Kamel H. Induced sputum levels of IL-33 and soluble ST2 in young asthmatic children. J Asthma 2013 Oct;50(8):803-809.

(414) Guo Z, Wu J, Zhao J, Liu F, Chen Y, Bi L, et al. IL-33 promotes airway remodeling and is a marker of asthma disease severity. J Asthma 2014 May 6.

(415) Zeyda M, Wernly B, Demyanets S, Kaun C, Hammerle M, Hantusch B, et al. Severe obesity increases adipose tissue expression of interleukin-33 and its receptor ST2, both predominantly detectable in endothelial cells of human adipose tissue. Int J Obes (Lond) 2013 May;37(5):658-665.

(416) Cancello R, Henegar C, Viguerie N, Taleb S, Poitou C, Rouault C, et al. Reduction of macrophage infiltration and chemoattractant gene expression changes in white adipose tissue of morbidly obese subjects after surgery-induced weight loss. Diabetes 2005 Aug;54(8):2277-2286.

(417) Oliveira AG, Carvalho BM, Tobar N, Ropelle ER, Pauli JR, Bagarolli RA, et al. Physical exercise reduces circulating lipopolysaccharide and TLR4 activation and improves insulin signaling in tissues of DIO rats. Diabetes 2011 Mar;60(3):784-796.

(418) Ali Z, Ulrik CS. Obesity and asthma: a coincidence or a causal relationship? A systematic review. Respir Med 2013 Sep;107(9):1287-1300.

(419) Ford ES. The epidemiology of obesity and asthma. J Allergy Clin Immunol 2005 May;115(5):897-909; quiz 910.

(420) Melen E, Himes BE, Brehm JM, Boutaoui N, Klanderman BJ, Sylvia JS, et al. Analyses of shared genetic factors between asthma and obesity in children. J Allergy Clin Immunol 2010 Sep;126(3):631-7.e1-8.

(421) Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet 2006 Aug 15;15(16):2490-2508.

(422) Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. Cell 2004 Dec 17;119(6):831-845.

(423) Leipzig J, Pevzner P, Heber S. The Alternative Splicing Gallery (ASG): bridging the gap between genome and transcriptome. Nucleic Acids Res 2004 Aug 3;32(13):3977-3983.

(424) Liu YJ. Thymic stromal lymphopoietin: master switch for allergic inflammation. J Exp Med 2006 Feb 20;203(2):269-273.

(425) Allakhverdi Z, Comeau MR, Jessup HK, Yoon BR, Brewer A, Chartier S, et al. Thymic stromal lymphopoietin is released by human epithelial cells in response to microbes, trauma, or inflammation and potently activates mast cells. J Exp Med 2007 Feb 19;204(2):253-258.

(426) Takai T. TSLP expression: cellular sources, triggers, and regulatory mechanisms. Allergol Int 2012 Mar;61(1):3-17.

(427) Jang S, Morris S, Lukacs NW. TSLP promotes induction of Th2 differentiation but is not necessary during established allergen-induced pulmonary disease. PLoS One 2013;8(2):e56433.

(428) Bustin SA. Why the need for qPCR publication guidelines?--The case for MIQE. Methods 2010 Apr;50(4):217-226.

(429) Harada M, Hirota T, Jodo AI, Doi S, Kameda M, Fujita K, et al. Functional analysis of the thymic stromal lymphopoietin variants in human bronchial epithelial cells. Am J Respir Cell Mol Biol 2009 Mar;40(3):368-374.

(430) Liu W, Xu LS, Liu QJ, Dong FZ, Qiu RF, Wen MC, et al. Two single nucleotide polymorphisms in TSLP gene are associated with asthma susceptibility in Chinese Han population. Exp Lung Res 2012 Oct;38(8):375-382.

(431) Birben E, Sahiner UM, Karaaslan C, Yavuz TS, Cosgun E, Kalayci O, et al. The genetic variants of thymic stromal lymphopoietin protein in children with asthma and allergic rhinitis. Int Arch Allergy Immunol 2014;163(3):185-192.

(432) Zhang Y, Wang X, Zhang W, Han D, Zhang L, Bachert C. Polymorphisms in thymic stromal lymphopoietin gene demonstrate a gender and nasal polyposis-dependent association with chronic rhinosinusitis. Hum Immunol 2013 Feb;74(2):241-248.

(433) Bjerkan L, Schreurs O, Engen SA, Jahnsen FL, Baekkevold ES, Blix IJ, et al. The short form of TSLP is constitutively translated in human keratinocytes and has characteristics of an antimicrobial peptide. Mucosal Immunol 2014 May 21.

(434) Xie Y, Takai T, Chen X, Okumura K, Ogawa H. Long TSLP transcript expression and release of TSLP induced by TLR ligands and cytokines in human keratinocytes. J Dermatol Sci 2012 Jun;66(3):233-237.

(435) Takai T, Chen X, Xie Y, Vu AT, Le TA, Kinoshita H, et al. TSLP expression induced via Toll-like receptor pathways in human keratinocytes. Methods Enzymol 2014;535:371-387.

(436) Leyva-Castillo JM, Hener P, Jiang H, Li M. TSLP produced by keratinocytes promotes allergen sensitization through skin and thereby triggers atopic march in mice. J Invest Dermatol 2013 Jan;133(1):154-163.

(437) Zhang Z, Hener P, Frossard N, Kato S, Metzger D, Li M, et al. Thymic stromal lymphopoietin overproduced by keratinocytes in mouse skin aggravates experimental asthma. Proc Natl Acad Sci U S A 2009 Feb 3;106(5):1536-1541.

(438) Jiang H, Hener P, Li J, Li M. Skin thymic stromal lymphopoietin promotes airway sensitization to inhalant house dust mites leading to allergic asthma in mice. Allergy 2012 Aug;67(8):1078-1082.

(439) Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 2010 Dec 14;107(50):21931-21936.

(440) Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res 2011 Aug;21(8):1273-1283.

(441) Baschant U, Tuckermann J. The role of the glucocorticoid receptor in inflammation and immunity. J Steroid Biochem Mol Biol 2010 May 31;120(2-3):69-75.

(442) Segawa R, Yamashita S, Mizuno N, Shiraki M, Hatayama T, Satou N, et al. Identification of a cell line producing high levels of TSLP: advantages for screening of anti-allergic drugs. J Immunol Methods 2014 Jan 15;402(1-2):9-14.

(443) Ozawa T, Koyama K, Ando T, Ohnuma Y, Hatsushika K, Ohba T, et al. Thymic stromal lymphopoietin secretion of synovial fibroblasts is positively and negatively regulated by Toll-like receptors/nuclear factor-kappaB pathway and interferon-gamma/dexamethasone. Mod Rheumatol 2007;17(6):459-463.

(444) Sigurs N, Gustafsson PM, Bjarnason R, Lundberg F, Schmidt S, Sigurbergsson F, et al. Severe respiratory syncytial virus bronchiolitis in infancy and asthma and allergy at age 13. Am J Respir Crit Care Med 2005 Jan 15;171(2):137-141.

(445) Johnston SL, Pattemore PK, Sanderson G, Smith S, Lampe F, Josephs L, et al. Community study of role of viral infections in exacerbations of asthma in 9-11 year old children. BMJ 1995 May 13;310(6989):1225-1229.

(446) Wark PA, Johnston SL, Moric I, Simpson JL, Hensley MJ, Gibson PG. Neutrophil degranulation and cell lysis is associated with clinical severity in virus-induced asthma. Eur Respir J 2002 Jan;19(1):68-75.

(447) Lee HC, Headley MB, Loo YM, Berlin A, Gale M,Jr, Debley JS, et al. Thymic stromal lymphopoietin is induced by respiratory syncytial virus-infected airway epithelial cells and

promotes a type 2 response to infection. J Allergy Clin Immunol 2012 Nov;130(5):1187-1196.e5.

(448) Qiao J, Li A, Jin X. TSLP from RSV-stimulated rat airway epithelial cells activates myeloid dendritic cells. Immunol Cell Biol 2011 Feb;89(2):231-238.

(449) Matsumoto M, Seya T. TLR3: interferon induction by double-stranded RNA including poly(I:C). Adv Drug Deliv Rev 2008 Apr 29;60(7):805-812.

(450) Kumar R. Prenatal factors and the development of asthma. Curr Opin Pediatr 2008 Dec;20(6):682-687.

(451) Smelter DF, Sathish V, Thompson MA, Pabelick CM, Vassallo R, Prakash YS. Thymic stromal lymphopoietin in cigarette smoke-exposed human airway smooth muscle. J Immunol 2010 Sep 1;185(5):3035-3040.

(452) Horvath KM, Brighton LE, Zhang W, Carson JL, Jaspers I. Epithelial cells from smokers modify dendritic cell responses in the context of influenza infection. Am J Respir Cell Mol Biol 2011 Aug;45(2):237-245.

(453) Gachter T, Werenskiold AK, Klemenz R. Transcription of the interleukin-1 receptor-related T1 gene is initiated at different promoters in mast cells and fibroblasts. J Biol Chem 1996 Jan 5;271(1):124-129.

(454) Gauvreau GM, O'Byrne PM, Boulet LP, Wang Y, Cockcroft D, Bigler J, et al. Effects of an Anti-TSLP Antibody on Allergen-Induced Asthmatic Responses. N Engl J Med 2014 May 20.

(455) Hong X, Wang G, Liu X, Kumar R, Tsai HJ, Arguelles L, et al. Gene polymorphisms, breast-feeding, and development of food sensitization in early childhood. J Allergy Clin Immunol 2011 Aug;128(2):374-81.e2.

(456) Liu M, Rogers L, Cheng Q, Shao Y, Fernandez-Beros ME, Hirschhorn JN, et al. Genetic variants of TSLP and asthma in an admixed urban population. PLoS One 2011;6(9):e25099.

(457) Cheong KA, Chae SC, Kim YS, Kwon HB, Chung HT, Lee AY. Association of thymic stromal lymphopoietin gene -847C>T polymorphism in generalized vitiligo. Exp Dermatol 2009 Dec;18(12):1073-1075.

(458) Luo Y, Zhou B, Zhao M, Tang J, Lu Q. Promoter demethylation contributes to TSLP overexpression in skin lesions of patients with atopic dermatitis. Clin Exp Dermatol 2014 Jan;39(1):48-53.

(459) Wang IJ, Chen SL, Lu TP, Chuang EY, Chen PC. Prenatal smoke exposure, DNA methylation, and childhood atopic dermatitis. Clin Exp Allergy 2013 May;43(5):535-543.

(460) Mittler G, Butter F, Mann M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res 2009 Feb;19(2):284-293.

Appendices

Appendix A - Supplementary tables for Chapter 3

A.1 List of the 186 SNPs in the *IL1RL1* region

	SNP rs	Chromosomal position build 37.3	Nearby Genes	Alleles
1	rs10178214	102858921	IL1RL2||LOC100129822	G/T
2	rs13027294	102860074	IL1RL2||LOC100129822	C/G
3	rs11677452	102865236	IL1RL2||LOC100129822	A/T
4	rs9646944	102865875	IL1RL2||LOC100129822	C/G
5	rs10189629	102879464	LOC100129822||IL1RL1	A/C
6	rs1882510	102883618	LOC100129822||IL1RL1	C/T
7	rs11692065	102883975	LOC100129822||IL1RL1	C/T
8	rs11674302	102887128	LOC100129822||IL1RL1	C/T
9	rs12469892	102895784	LOC100129822||IL1RL1	A/G
10	rs951774	102912664	LOC100129822||IL1RL1	A/C
11	rs11690644	102914214	LOC100129822||IL1RL1	A/G
12	rs11693697	102915662	LOC100129822||IL1RL1	C/T
13	rs4399750	102917788	LOC100129822||IL1RL1	C/T
14	rs7568913	102920037	LOC100129822||IL1RL1	C/T
15	rs4090473	102922987	LOC100129822||IL1RL1	G/C
16	rs10178436	102926511	LOC100129822||IL1RL1	C/T
17	rs11685424	102926981	LOC100129822||IL1RL1	A/G
18	rs1558622	102930147	IL1RL1	G/A
19	rs10189711	102930881	IL1RL1	A/G
20	rs12712135	102930948	IL1RL1	A/G
21	rs953934	102932293	IL1RL1	T/C
22	rs950880	102932562	IL1RL1	C/A
23	rs11123918	102935237	IL1RL1	C/T
24	rs10182639	102935805	IL1RL1	A/C
25	rs11690443	102936131	IL1RL1	A/T
26	rs974389	102936981	IL1RL1	G/A
27	rs4142132	102937482	IL1RL1	G/A
28	rs13001325	102939036	IL1RL1	C/T
29	rs1420088	102939434	IL1RL1	T/C
30	rs11123920	102939833	IL1RL1	C/T
31	rs6706844	102940412	IL1RL1	C/T
32	rs12996772	102947201	IL1RL1	A/T
33	rs1420103	102948632	IL1RL1	C/A
34	rs1420102	102948819	IL1RL1	T/C
35	rs12466380	102948939	IL1RL1	A/G

	SNP rs	Chromosomal position build 37.3	Nearby Genes	Alleles
36	rs12479210	102949161	IL1RL1	C/T
37	rs13019081	102950822	IL1RL1	A/C
38	rs1997467	102951073	IL1RL1	G/A
39	rs1997466	102951467	IL1RL1	G/C
40	rs1362350	102951798	IL1RL1	G/C
41	rs2310220	102951851	IL1RL1	A/G
42	rs1362349	102951972	IL1RL1	G/C
43	rs17026974	102952360	IL1RL1	A/G
44	rs12712141	102953067	IL1RL1	C/T
45	rs3771180	102953617	IL1RL1	T/G
46	rs13431828	102954653	IL1RL1	C/T
47	rs13408569	102955056	IL1RL1	C/G
48	rs13408661	102955082	IL1RL1	A/G
49	rs1041973	102955468	IL1RL1	A/C
50	rs873022	102955683	IL1RL1	G/T
51	rs3771177	102955860	IL1RL1	T/G
52	rs10173081	102957348	IL1RL1	C/T
53	rs3732129	102957532	IL1RL1	T/C
54	rs1420101	102957716	IL1RL1	T/C
55	rs12905	102960007	IL1RL1	A/G
56	rs3771175	102960210	IL1RL1	A/T
57	rs12712142	102960584	IL1RL1	A/C
58	rs6543119	102963072	IL1RL1	A/T
59	rs13017455	102964742	IL1RL1	C/T
60	rs1921622	102966067	IL1RL1	A/G
61	rs10197862	102966549	IL1RL1	A/G
62	rs11123923	102967844	IL1RL1	A/C
63	rs13015714	102971865	IL1RL1||IL18R1	G/T
64	rs12998521	102974417	IL1RL1||IL18R1	G/T
65	rs2287037	102979028	IL1RL1||IL18R1	T/C
66	rs2058622	102985424	IL18R1	G/A
67	rs3771170	102985980	IL18R1	T/A
68	rs2058623	102986170	IL18R1	T/C
69	rs1465321	102986618	IL18R1	T/C
70	rs17027029	102990648	IL18R1	C/G
71	rs3771164	102991786	IL18R1	T/A
72	rs2270297	102992675	IL18R1	T/C
73	rs6753717	102993161	IL18R1	A/C
74	rs6750020	102994714	IL18R1	A/G
75	rs3213733	102997884	IL18R1	C/A
76	rs3771161	103003961	IL18R1	C/A
77	rs3771158	103009894	IL18R1	G/A
78	rs3732127	103013750	IL18R1	G/C
79	rs12991737	103018128	IL18R1||LOC442035	A/T
80	rs10202813	103019740	IL18R1||LOC442035	G/T
81	rs1035127	103019919	IL18R1||LOC442035	G/A
82	rs10197310	103020030	IL18R1||LOC442035	A/T
83	rs4851007	103024813	IL18R1||LOC442035	G/T
84	rs4851575	103025203	IL18R1||LOC442035	A/G
85	rs10181785	103025274	IL18R1||LOC442035	C/T
86	rs12712148	103025547	IL18R1||LOC442035	A/G
87	rs11687768	103025738	IL18R1||LOC442035	A/G
88	rs7586983	103028066	LOC442035	C/T
89	rs1807782	103033147	LOC442035	T/C
90	rs1420106	103035044	LOC442035	G/A
91	rs2293223	103035468	LOC442035/IL18RAP	C/T

	SNP rs	Chromosomal position build 37.3	Nearby Genes	Alleles
92	rs3755268	103038527	LOC442035/IL18RAP	G/C
93	rs3755267	103038587	LOC442035/IL18RAP	T/G
94	rs3817465	103039584	IL18RAP	A/T
95	rs2272127	103039873	IL18RAP	C/G
96	rs2272128	103039929	IL18RAP	A/G
97	rs11694658	103045020	IL18RAP	A/G
98	rs2160232	103046880	IL18RAP	G/A
99	rs6716784	103048467	IL18RAP	G/T
100	rs2041756	103049910	IL18RAP	G/A
101	rs10166330	103050390	IL18RAP	C/T
102	rs6543134	103050458	IL18RAP	G/T
103	rs2110735	103050925	IL18RAP	G/A
104	rs2110734	103052206	IL18RAP	T/C
105	rs6746271	103052995	IL18RAP	C/G
106	rs10176820	103054420	IL18RAP	C/T
107	rs2058660	103054449	IL18RAP	G/A
108	rs2058658	103054803	IL18RAP	T/C
109	rs4851009	103055644	IL18RAP	A/G
110	rs2075186	103057251	IL18RAP	T/G
111	rs1558650	103060024	IL18RAP	T/A
112	rs11694360	103061147	IL18RAP	A/G
113	rs11123928	103061286	IL18RAP	A/G
114	rs7597017	103062116	IL18RAP	A/G
115	rs6734736	103062880	IL18RAP	C/T
116	rs6708413	103063369	IL18RAP	A/G
117	rs11465730	103066858	IL18RAP	A/G
118	rs7559479	103068787	IL18RAP	A/G
119	rs917997	103070568	IL18RAP||SLC9A4	T/C
120	rs4070554	103074493	IL18RAP||SLC9A4	A/G
121	rs6761825	103075561	IL18RAP||SLC9A4	C/T
122	rs6705001	103076210	IL18RAP||SLC9A4	A/G
123	rs6543141	103076351	IL18RAP||SLC9A4	A/G
124	rs6705385	103076569	IL18RAP||SLC9A4	A/C
125	rs6705498	103076670	IL18RAP||SLC9A4	A/G
126	rs6719196	103076888	IL18RAP||SLC9A4	G/T
127	rs4241210	103078740	IL18RAP||SLC9A4	A/G
128	rs6720564	103079297	IL18RAP||SLC9A4	C/T
129	rs10210176	103079516	IL18RAP||SLC9A4	A/C
130	rs6717915	103079619	IL18RAP||SLC9A4	A/C
131	rs6718157	103079814	IL18RAP||SLC9A4	A/T
132	rs917996	103082273	IL18RAP||SLC9A4	C/A
133	rs741284	103083324	IL18RAP||SLC9A4	G/C
134	rs12463588	103085257	IL18RAP||SLC9A4	C/G
135	rs2310302	103086049	IL18RAP||SLC9A4	C/G
136	rs12469887	103086758	IL18RAP||SLC9A4	C/T
137	rs990171	103086770	IL18RAP||SLC9A4	C/A
138	rs10172116	103087573	IL18RAP||SLC9A4	C/T
139	rs4140786	103088176	IL18RAP||SLC9A4	T/G
140	rs1403552	103088777	IL18RAP||SLC9A4	T/C
141	rs10201184	103089078	IL18RAP||SLC9A4	C/G
142	rs13030642	103091585	SLC9A4	A/C
143	rs1468791	103092021	SLC9A4	G/A
144	rs13018263	103092270	SLC9A4	C/T
145	rs7597819	103092906	SLC9A4	A/G
146	rs6737668	103093081	SLC9A4	C/T
147	rs10469840	103093243	SLC9A4	C/T

	SNP rs	Chromosomal position (build 37.3)	Nearby Genes	Alleles
148	rs759382	103094213	SLC9A4	T/G
149	rs759381	103094323	SLC9A4	T/A
150	rs13019784	103123301	SLC9A4	A/G
151	rs9989842	103123633	SLC9A4	C/G
152	rs9989749	103123642	SLC9A4	A/G
153	rs6708949	103123965	SLC9A4	C/G
154	rs6724109	103125018	SLC9A4	C/G
155	rs6751949	103125138	SLC9A4	A/G
156	rs6724322	103125182	SLC9A4	C/T
157	rs4851607	103125632	SLC9A4	C/T
158	rs10195948	103125736	SLC9A4	C/T
159	rs12712155	103127963	SLC9A4	A/T
160	rs4851609	103128866	SLC9A4	C/T
161	rs11676371	103129692	SLC9A4	C/G
162	rs2192758	103132269	SLC9A4	G/C
163	rs2192757	103132378	SLC9A4	T/C
164	rs6714379	103133310	SLC9A4	A/G
165	rs1523203	103135759	SLC9A4	G/A
166	rs4851611	103135938	SLC9A4	A/T
167	rs1403550	103136309	SLC9A4	C/T
168	rs1403551	103136446	SLC9A4	G/T
169	rs4851613	103137990	SLC9A4	C/T
170	rs6750851	103138761	SLC9A4	A/G
171	rs6750971	103138825	SLC9A4	A/G
172	rs10193407	103139298	SLC9A4	C/T
173	rs11123935	103139751	SLC9A4	A/G
174	rs1024798	103141651	SLC9A4	G/C
175	rs3849365	103144391	SLC9A4	G/A
176	rs6724213	103151219	SLC9A4||SLC9A2	A/C
177	rs2871474	103151441	SLC9A4||SLC9A2	A/G
178	rs4851616	103151862	SLC9A4||SLC9A2	C/T
179	rs11685483	103159093	SLC9A4||SLC9A2	A/C
180	rs6739426	103160443	SLC9A4||SLC9A2	A/G
181	rs11899041	103161053	SLC9A4||SLC9A2	A/T
182	rs12712157	103165129	SLC9A4||SLC9A2	C/T
183	rs1303960	103165832	SLC9A4||SLC9A2	G/A
184	rs2005881	103173059	SLC9A4||SLC9A2	G/A
185	rs4851619	103176411	SLC9A4||SLC9A2	C/T
186	rs6761871	103176797	SLC9A4||SLC9A2	A/C

A.2 List of the 232 SNPs in the *IL1RL1* region

	HapMap LD bin	ENCODE LD bin	SNP rs	Chromosomal position build 37.3	Nearby Genes	SNP source
1	1	1	rs10178214	102858921	IL1RL2||LOC100129822	eQTL
2	1	1	rs13027294	102860074	IL1RL2||LOC100129822	eQTL
3	1	1	rs11677452	102865236	IL1RL2||LOC100129822	eQTL
4	1	1	rs9646944	102865875	IL1RL2||LOC100129822	eQTL
5	1	5	rs10189629	102879464	LOC100129822||IL1RL1	eQTL
6	1	5	rs11692065	102883975	LOC100129822||IL1RL1	eQTL
7	1	5	rs11674302	102887128	LOC100129822||IL1RL1	eQTL + Association
8	1	3	rs11690644	102914214	LOC100129822||IL1RL1	eQTL
9	1	2	rs3771180	102953617	IL1RL1	eQTL + Association
10	1	2	rs13431828	102954653	IL1RL1	eQTL + Association
11	1	2	rs13408569	102955056	IL1RL1	eQTL + LD
12	1	3	rs13408661	102955082	IL1RL1	eQTL + LD
13	1	17	rs1041973	102955468	IL1RL1	eQTL + Association
14	1	3	rs10173081	102957348	IL1RL1	eQTL + Association
15	1	2	rs3771175	102960210	IL1RL1	eQTL
16	1	3	rs10197862	102966549	IL1RL1	eQTL + Association
17	1	2	rs17027029	102990648	IL18R1	eQTL
18	1	2	rs3771164	102991786	IL18R1	eQTL
19	1	2	rs3213733	102997884	IL18R1	eQTL + Association
20	1	2	rs3771161	103003961	IL18R1	eQTL
21	1	2	rs3771158	103009894	IL18R1	eQTL
22	1	2	rs3732127	103013750	IL18R1	eQTL
23	1	2	rs12991737	103018128	IL18R1||LOC442035	eQTL
24	1	2	rs10202813	103019740	IL18R1||LOC442035	eQTL
25	1	2	rs10197310	103020030	IL18R1||LOC442035	eQTL
26	1	2	rs10181785	103025274	IL18R1||LOC442035	eQTL
27	1	2	rs12712148	103025547	IL18R1||LOC442035	eQTL
28	1	2	rs11687768	103025738	IL18R1||LOC442035	eQTL
29	1	2	rs7586983	103028066	LOC442035	eQTL
30	1	2	rs2293223	103035468	LOC442035/IL18RAP	eQTL
31	1	20	rs2272127	103039873	IL18RAP	eQTL
32	1	2	rs10166330	103050390	IL18RAP	eQTL
33	1	2	rs10176820	103054420	IL18RAP	eQTL
34	1	2	rs2075186	103057251	IL18RAP	eQTL
35	1	2	rs10210176	103079516	IL18RAP||SLC9A4	eQTL
36	1	2	rs741284	103083324	IL18RAP||SLC9A4	eQTL
37	1	2	rs10172116	103087573	IL18RAP||SLC9A4	eQTL
38	1	2	rs1403552	103088777	IL18RAP||SLC9A4	eQTL
39	1	2	rs13030642	103091585	SLC9A4	eQTL
40	1	12	rs13018263	103092270	SLC9A4	eQTL
41	2	10	rs1882510	102883618	LOC100129822||IL1RL1	eQTL
42	2	10	rs12469892	102895784	LOC100129822||IL1RL1	eQTL
43	2	10	rs951774	102912664	LOC100129822||IL1RL1	eQTL
44	2	21	rs7600901	102915571	LOC100129822||IL1RL1	LD
45	2	13	rs1420103	102948632	IL1RL1	eQTL
46	2	13	rs2310220	102951851	IL1RL1	eQTL
47	2	7	rs13015714	102971865	IL1RL1||IL18R1	eQTL
48	2	7	rs2058622	102985424	IL18R1	eQTL + Association
49	2	7	rs3771170	102985980	IL18R1	eQTL
50	2	7	rs2058623	102986170	IL18R1	eQTL
51	2	7	rs1465321	102986618	IL18R1	eQTL
52	2	7	rs2270297	102992675	IL18R1	eQTL + Association
53	2	7	rs6753717	102993161	IL18R1	eQTL
54	2	7	rs6750020	102994714	IL18R1	eQTL
55	2	7	rs1035127	103019919	IL18R1||LOC442035	eQTL
56	2	7	rs4851007	103024813	IL18R1||LOC442035	eQTL

	HapMap LD bin	ENCODE LD bin	SNP rs	Chromosomal position build 37.3	Nearby Genes	SNP source
57	2	7	rs4851575	103025203	IL18R1||LOC442035	eQTL
58	2	7	rs1807782	103033147	LOC442035	eQTL
59	2	7	rs1420106	103035044	IL18RAP	eQTL + Association
60	2	7	rs3755268	103038527	LOC442035/IL18RAP	eQTL
61	2	7	rs3755267	103038587	LOC442035/IL18RAP	eQTL
62	2	7	rs3817465	103039584	IL18RAP	eQTL
63	2	7	rs2272128	103039929	IL18RAP	eQTL
64	2	7	rs11694658	103045020	IL18RAP	eQTL
65	2	7	rs2160232	103046880	IL18RAP	eQTL
66	2	7	rs6716784	103048467	IL18RAP	eQTL
67	2	7	rs2041756	103049910	IL18RAP	eQTL
68	2	7	rs6543134	103050458	IL18RAP	eQTL
69	2	7	rs2110735	103050925	IL18RAP	eQTL
70	2	7	rs2110734	103052206	IL18RAP	eQTL
71	2	7	rs6746271	103052995	IL18RAP	eQTL
72	2	7	rs2058660	103054449	IL18RAP	eQTL
73	2	7	rs2058658	103054803	IL18RAP	eQTL
74	2	7	rs4851009	103055644	IL18RAP	eQTL
75	2	7	rs1558650	103060024	IL18RAP	eQTL
76	2	7	rs6734736	103062880	IL18RAP	eQTL
77	2	7	rs6708413	103063369	IL18RAP	eQTL
78	2	7	rs7559479	103068787	IL18RAP	eQTL
79	2	7	rs917997	103070568	IL18RAP||SLC9A4	eQTL
80	2	7	rs4070554	103074493	IL18RAP||SLC9A4	eQTL
81	2	7	rs6761825	103075561	IL18RAP||SLC9A4	eQTL
82	2	7	rs6705001	103076210	IL18RAP||SLC9A4	eQTL
83	2	7	rs6543141	103076351	IL18RAP||SLC9A4	eQTL
84	2	7	rs4241210	103078740	IL18RAP||SLC9A4	eQTL
85	2	7	rs6720564	103079297	IL18RAP||SLC9A4	eQTL
86	2	7	rs6717915	103079619	IL18RAP||SLC9A4	eQTL
87	2	7	rs6718157	103079814	IL18RAP||SLC9A4	eQTL
88	2	7	rs917996	103082273	IL18RAP||SLC9A4	eQTL
89	2	7	rs990171	103086770	IL18RAP||SLC9A4	eQTL
90	2	7	rs1468791	103092021	SLC9A4	eQTL
91	2	7	rs7597819	103092906	SLC9A4	eQTL
92	2	7	rs6737668	103093081	SLC9A4	eQTL
93	2	7	rs10469840	103093243	SLC9A4	eQTL
94	2	4	rs11685483	103159093	SLC9A4||SLC9A2	eQTL
95	2	4	rs6739426	103160443	SLC9A4||SLC9A2	eQTL
96	2	4	rs11899041	103161053	SLC9A4||SLC9A2	eQTL
97	2	4	rs12712157	103165129	SLC9A4||SLC9A2	eQTL
98	2	4	rs1303960	103165832	SLC9A4||SLC9A2	eQTL
99	2	4	rs2005881	103173059	SLC9A4||SLC9A2	eQTL
100	2	4	rs4851619	103176411	SLC9A4||SLC9A2	eQTL
101	2	4	rs6761871	103176797	SLC9A4||SLC9A2	eQTL
102	3	22	rs4851563	102885535	LOC100129822||IL1RL1	LD
103	3	24	rs1476984	102912269	LOC100129822||IL1RL1	LD
104	3	24	rs1420089	102938389	IL1RL1	Association
105	3	24	rs17696376	102965153	IL1R1	LD
106	3	22	rs4851567	102972807	IL1RL1||IL18R1	LD
107	3	24	rs12105808	102974222	IL1RL1 / IL18R1	LD
108	3	23	rs13425475	103025181	IL18R1||LOC442035	LD
109	3	23	rs4851581	103034749	LOC442035	LD
110	3	23	rs11465722	103063489		LD
111	4	18	rs11694697	102915662	LOC100129822||IL1RL1	eQTL
112	5	4	rs4399750	102917788	LOC100129822||IL1RL1	eQTL

	HapMap LD bin	ENCODE LD bin	SNP rs	Chromosomal position build 37.3	Nearby Genes	SNP source
113	5	4	rs7568913	102920037	LOC100129822||IL1RL1	eQTL
114	5	4	rs4090473	102922987	LOC100129822||IL1RL1	eQTL
115	5	4	rs10178436	102926511	LOC100129822||IL1RL1	eQTL
116	5	4	rs11685424	102926981	LOC100129822||IL1RL1	eQTL
117	5	4	rs1558622	102930147	IL1RL1	eQTL
118	5	4	rs10189711	102930881	IL1RL1	eQTL
119	5	4	rs12712135	102930948	IL1RL1	eQTL
120	5	4	rs953934	102932293	IL1RL1	eQTL
121	5	4	rs11123918	102935237	IL1RL1	eQTL
122	5	4	rs10182639	102935805	IL1RL1	eQTL
123	5	4	rs11690443	102936131	IL1RL1	eQTL
124	5	4	rs974389	102936981	IL1RL1	eQTL
125	5	4	rs4142132	102937482	IL1RL1	eQTL
126	5	4	rs1420088	102939434	IL1RL1	eQTL
127	5	4	rs11123920	102939833	IL1RL1	eQTL
128	5	4	rs6706844	102940412	IL1RL1	eQTL
129	5	4	rs12996772	102947201	IL1RL1	eQTL
130	5	4	rs1420102	102948819	IL1RL1	eQTL
131	5	4	rs12466380	102948939	IL1RL1	eQTL
132	5	4	rs1997467	102951073	IL1RL1	eQTL
133	5	4	rs1997466	102951467	IL1RL1	eQTL
134	5	4	rs1362350	102951798	IL1RL1	eQTL
135	5	4	rs1362349	102951972	IL1RL1	eQTL
136	5	4	rs12712141	102953067	IL1RL1	eQTL
137	6	11	rs950880	102932562	IL1RL1	eQTL
138	6	11	rs13001325	102939036	IL1RL1	eQTL
139	6	11	rs12479210	102949161	IL1RL1	eQTL
140	6	11	rs13019081	102950822	IL1RL1	eQTL
141	6	9	rs17026974	102952360	IL1RL1	eQTL
142	6	9	rs873022	102955683	IL1RL1	eQTL
143	6	9	rs3771177	102955860	IL1RL1	eQTL
144	6	9	rs3732129	102957532	IL1RL1	eQTL
145	6	11	rs1420101	102957716	IL1RL1	eQTL + Association
146	6	9	rs12905	102960007	IL1RL1	eQTL
147	6	9	rs1035130	103001402	IL18R1	eQTL + Association
148	7	27	rs12999517	102959260	IL1R1	LD
149	7	27	rs1946131	102961929	IL1RL1	Association
150	7	27	rs12989197	102962739	IL1R1	LD
151	7	27	rs12996097	102963628	IL1R1	LD
152	7	27	rs13028993	102963949	IL1R1	LD
153	7	27	rs12999542	102965329	IL1R1	LD
154	8	19	rs1921622	102966067	IL1RL1	eQTL + Association
155	9	25	rs13424006	102967236	IL1RL1	LD
156	9	25	rs6751967	102967413	IL1RL1	LD
157	9	25	rs6749114	102967587	IL1RL1	LD
158	9	25	rs4988955	102967928	IL1RL1	LD
159	9	25	rs4988956	102968007	IL1RL1	LD
160	9	25	rs4988957	102968075	IL1RL1	LD
161	9	25	rs10204137	102968212	IL1RL1	Association
162	9	25	rs4988958	102968285	IL1RL1	LD
163	9	25	rs10192157	102968356	IL1RL1	Association
164	9	25	rs10206753	102968362	IL1RL1	Association
165	9	25	rs1362348	102984624	IL18R1	Association
166	9	25	rs3771166	102986222	IL18R1	Association
167	9	25	rs1974675	102986375	IL18R1	Association
168	9	26	rs10439410	102990788	IL18R1	LD

	HapMap LD bin	ENCODE LD bin	SNP rs	Chromosomal position build 37.3	Nearby Genes	SNP source
169	9	26	rs6758936	102991369	IL18R1	LD
170	9	26	rs2041739	102994333	IL18R1	LD
171	9	26	rs10208196	102996345	IL18R1	LD
172	9	26	rs3213732	102998279	IL18R1	LD
173	9	26	rs6760621	102999952	IL18R1	LD
174	9	26	rs6706002	103006104	IL18R1	LD
175	9	26	rs6749014	103006448	IL18R1	LD
176	9	26	rs4851004	103009537	IL18R1	Association
177	9	26	rs2287033	103011237	IL18R1	Association
178	9	26	rs1420094	103015687	IL18R1	Association
179	9	26	rs3732124	103018052	IL18R1||LOC442035	LD
180	9	26	rs4851571	103019000	IL18R1||LOC442035	LD
181	9	26	rs4851572	103019031	IL18R1||LOC442035	LD
182	9	26	rs2110662	103020139	IL18R1||LOC442035	LD
183	9	28	rs1861245	102966906	IL1RL1	Association
184	10	9	rs11694360	103061147	IL18RAP	eQTL
185	10	9	rs11123928	103061286	IL18RAP	eQTL
186	10	9	rs7597017	103062116	IL18RAP	eQTL
187	10	6	rs11465730	103066858	IL18RAP	eQTL
188	10	6	rs6705385	103076569	IL18RAP||SLC9A4	eQTL
189	10	6	rs6705498	103076670	IL18RAP||SLC9A4	eQTL
190	10	6	rs6719196	103076888	IL18RAP||SLC9A4	eQTL
191	10	6	rs12463588	103085257	IL18RAP||SLC9A4	eQTL
192	10	6	rs2310302	103086049	IL18RAP||SLC9A4	eQTL
193	10	6	rs12469887	103086758	IL18RAP||SLC9A4	eQTL
194	10	6	rs4140786	103088176	IL18RAP||SLC9A4	eQTL
195	10	6	rs10201184	103089078	IL18RAP||SLC9A4	eQTL
196	10	15	rs759382	103094213	SLC9A4	eQTL
197	10	15	rs759381	103094323	SLC9A4	eQTL
198	10	4	rs13019784	103123301	SLC9A4	eQTL
199	10	4	rs9989842	103123633	SLC9A4	eQTL
200	10	4	rs9989749	103123642	SLC9A4	eQTL
201	10	4	rs6708949	103123965	SLC9A4	eQTL
202	10	4	rs6724109	103125018	SLC9A4	eQTL
203	10	4	rs6751949	103125138	SLC9A4	eQTL
204	10	4	rs6724322	103125182	SLC9A4	eQTL
205	10	4	rs4851607	103125632	SLC9A4	eQTL
206	10	16	rs10195948	103125736	SLC9A4	eQTL
207	10	4	rs12712155	103127963	SLC9A4	eQTL
208	10	4	rs4851609	103128866	SLC9A4	eQTL
209	10	4	rs11676371	103129692	SLC9A4	eQTL
210	10	4	rs2192758	103132269	SLC9A4	eQTL
211	10	4	rs2192757	103132378	SLC9A4	eQTL
212	10	4	rs6714379	103133310	SLC9A4	eQTL
213	10	4	rs1523203	103135759	SLC9A4	eQTL
214	10	4	rs4851611	103135938	SLC9A4	eQTL
215	10	4	rs1403550	103136309	SLC9A4	eQTL
216	10	4	rs1403551	103136446	SLC9A4	eQTL
217	10	4	rs4851613	103137990	SLC9A4	eQTL
218	10	4	rs6750851	103138761	SLC9A4	eQTL
219	10	4	rs6750971	103138825	SLC9A4	eQTL
220	10	4	rs10193407	103139298	SLC9A4	eQTL
221	10	4	rs11123935	103139751	SLC9A4	eQTL
222	10	4	rs1024798	103141651	SLC9A4	eQTL
223	10	4	rs6724213	103151219	SLC9A4||SLC9A2	eQTL
224	10	4	rs2871474	103151441	SLC9A4||SLC9A2	eQTL

	HapMap LD bin	ENCODE LD bin	SNP rs	Chromosomal position build 37.3	Nearby Genes	SNP source
225	10	4	rs4851616	103151862	SLC9A4||SLC9A2	eQTL
226	11	14	rs3849365	103144391	SLC9A4	eQTL
227	12	8	rs12712142	102960584	IL1RL1	eQTL
228	12	8	rs6543119	102963072	IL1RL1	eQTL
229	12	8	rs13017455	102964742	IL1RL1	eQTL
230	12	8	rs11123923	102967844	IL1RL1	eQTL
231	12	8	rs12998521	102974417	IL1RL1||IL18R1	eQTL
232	12	8	rs2287037	102979028	IL1RL1||IL18R1	eQTL + Association

SNP source legend:
eQTL: SNP identified from the lung eQTL dataset
LD: SNP identified to be in high LD with an eQTL or asthma-associated SNP
Association: SNP identified from a published genetic association study of asthma

A.3 List of the 77 SNPs in the *IL1RL1* region

	pos (hg19)	SNP	HapMap LD bin	Encode LD bin	Ref	Alt	CEU	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP func annot
1	103087573	rs10172116	1	3	C	T	0.18						Ets	2.2kb 5' of SLC9A4	
2	103025274	rs10181785	1	3	C	T	0.18						RXRA,Roaz	9.9kb 5' of IL18RAP	
3	102955468	rs1041973	1	13	C	A	0.23			Huvec, K562			AP-4, AscI2, LBP-1, Maf, Myf, NRSF	IL1RL1	missense
4	103025738	rs11687768	1	3	A	G	0.18						Hbp1	9.4kb 5' of IL18RAP	
5	103025547	rs12712148	1	3	G	A	0.18						Dlx5, En-1, Pax-4, Pou1f1, RXRA	9.6kb 5' of IL18RAP	
6	102954653	rs13431828	1	2	C	T	0.14			Huvec, NHEK, HSMM, K562	K562, HCM, HEEpiC, HMVEC-Dbl-Ad, HPAEC, HRGEC, PANC-1, SAEC, WI-38	CFOS, CJUN, POL2, FOSL1, JUNB, JUND	LXR	IL1RL1	5'-UTR
7	103088777	rs1403552	1	3	C	T	0.18						DMRT1,RREB-1	984bp 5' of SLC9A4	
8	102990648	rs17027029	1	3	G	C	0.18						AFP1, FoxQ1, HNF1, Myf, SIX5, SRF	IL18R1	intronic
9	103057251	rs2075186	1	3	G	T	0.18						Tgif1	IL18RAP	intronic
10	103035468	rs2293223	1	3	C	T	0.18		K562	GM12878			5 altered motifs	IL18RAP	intronic
11	102997884	rs3213733	1	3	C	A	0.18						GATA,Pax-8,TAL1	IL18R1	intronic
12	103013750	rs3732127	1	3	G	C	0.17			K562			EBF	IL18R1	3'-UTR
13	102960210	rs3771175	1	2	T	A	0.15			K562	HCM,NB4		BRCA1, Ets, Irf, Myb, tcf12	IL1RL1	3'-UTR
14	102953617	rs3771180	1	2	G	T	0.14			Huvec, HMEC, K562, NHEK	(1)	CFOS,CJUN,STAT3	AP-1, Bch2, GR, PRDM1, STAT	IL1RL1	intronic
15	103028066	rs7586983	1	3	C	T	0.18						PPAR	7.1kb 5' of IL18RAP	
16	102865875	rs9646944	1	1	G	C	0.2				Chorion,Hepatocytes		GATA,HDAC2,HMGN3	9.4kb 3' of IL1RL2	
17	103035044	rs1420106	2	7	A	G	0.79			GM12878, K562, Huvec	(2)	CCNT2, CJUN, CMYC, JUND, PU1, STAT1, TAL1	Evi-1	104bp 5' of IL18RAP	
18	103033147	rs1807782	2	7	C	T	0.79				Th1			2kb 5' of IL18RAP	
19	102985424	rs2058622	2	7	A	G	0.79			GM12878			Evi-1, Osf2, PEBP, SIX5	IL18R1	intronic

	pos (hg19)	SNP	HapMap LD bin	Encode LD bin	Ref	Alt	CEU	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP func annot
20	103054449	rs2058660	2	7	G	A	0.79			K562, GM12878			HNF4, Ik-1, Ik-2, Nanog, RXRA, Sox, TCF4	IL18RAP	intronic
21	102992675	rs2270297	2	7	T	C	0.79				AG04450,HPF			IL18R1	intronic
22	103039929	rs2272128	2	7	G	A	0.79			GM12878, K562			AIRE,Nkx6-1	IL18RAP	intronic
23	102951851	rs2310220	2	7	G	A	0.74			K562, Huvec			GCNF,Zfp740	IL1RL1	intronic
24	103039584	rs3817465	2	7	A	T	0.79			GM12878, K562	Th1,Monocytes-CD14+_RO01746,NB4		GR,VDR	IL18RAP	intronic
25	103024813	rs4851007	2	7	T	G	0.79				Melano		Evi-1, GR, HDAC2, Osf2, PEBP, Sin3Ak-20	9.6kb 3' of IL18R1	
26	103025203	rs4851575	2	7	G	A	0.79						PPAR	9.9kb 5' of IL18RAP	
27	103052995	rs6746271	2	7	G	C	0.79			HSMM, K5562, GM12878, Huvec	(3)	NFKB, EBF1, ELF1, IRF4, MEF2A, PAX5C20	Maf	IL18RAP	intronic
28	102993161	rs6753717	2	7	A	C	0.79						EWSR1-FLI1,TCF12	IL18R1	intronic
29	103068787	rs7559479	2	7	G	A	0.79						Egr-1, SETDB1, Sox, Zfp410	IL18RAP	3'-UTR
30	103070568	rs917997	2	7	T	C	0.79						TAL1	1.5kb 3' of IL18RAP	
31	103086770	rs990171	2	7	A	C	0.79						CDP,GATA	3kb 5' of SLC9A4	
32	103063489	rs11465722	3	19	C	T	0.12						Myc	IL18RAP	intronic
33	102974222	rs12105808	3	18	A	T	0.13		GM12878	Huvec		EBF1	MZF1::1-4,ZBTB7A,Zfp740	IL18R1	
34	103025181	rs13425475	3	19	G	A	0.12						RAR,RXRA,SMC3	10kb 3' of IL18R1	
35	102938389	rs1420089	3	18	T	C	0.13					GATA2	Evi-1,Lhx4	IL1RL1	intronic
36	102972807	rs4851567	3	18	G	A	0.13		NHLF, Huvec, GM12878, NHEK, H1, HMEC, K562, HepG2	HSMM	GM12878, GM18507, GM19238, Huh-7, Urothelia, CD20+, CD34+_mobilized, H7-hESC, PANC-1, SAEC, Th2	EGR1,POL2	ATF3, ATF6, AhR, RXRA, SP1	IL18R1	

	pos (hg19)	SNP	HapMap LD bin	Encode LD bin	Ref	Alt	CEU	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP func annot
37	103034749	rs4851581	3	19	A	G	0.12			GM12878, K562	Th1		Foxa,Foxp1,HDAC2	399bp 5' of IL18RAP	
38	102915662	rs11693697	4	14	T	C	0.18						Hoxa5,Hoxb13	12kb 5' of IL1RL1	
39	102926981	rs11685424	5	5	G	A	0.49						Ik-2	980bp 5' of IL1RL1	
40	102951798	rs1362350	5	5	G	C	0.49			K562, Huvec			Pax-4	IL1RL1	intronic
41	102922987	rs4090473	5	5	C	G	0.49			K562			CIZ, HNF4, RAR, RXRA	5kb 5' of IL1RL1	
42	103001402	rs1035130	6	9	C	T	0.25						AIRE,RREB-1,STAT	IL18R1	synonymous
43	102960007	rs12905	6	9	G	A	0.25			K562			HDAC2	IL1RL1	3'-UTR
44	102957716	rs1420101	6	8	C	T	0.34						HNF4,Pdx1,RXRA	IL1RL1	intronic
45	102955860	rs3771177	6	9	G	T	0.25			Huvec			AP-2rep,ZBTB7A	IL1RL1	intronic
46	102963628	rs12996097	7	23	G	A	0.13						CDP,Roaz	IL1RL1	intronic
47	102965392	rs12999542	7	23	A	C	0.13			K562			ATF3, ATF6, E2F, HEY1	IL1RL1	intronic
48	102963949	rs13028993	7	23	T	C	0.13						BCL, Egr-1, FXR, GATA, HNF1, Irf, Hkx2, PU.1, Pax-5, RXRA	IL1RL1	intronic
49	102966067	rs1921622	8	15	G	A	0.52						7 altered motifs	IL1RL1	intronic
50	102968356	rs10192157	9	20	C	T	0.41			Huvec			PRDM1,Pax-5,Roaz	IL1RL1	missense
51	102968212	rs10204137	9	21	A	G	0.32			Huvec			WT1	IL1RL1	missense
52	102968362	rs10206753	9	20	T	C	0.41			Huvec			Pax-5,Pou2f2,Roaz	IL1RL1	missense
53	102968007	rs4988956	9	20	G	A	0.41			Huvec	HMVEC-dBl-Ad,HMVEC-dLy-Neo,Th2	GATA2	Zbtb3	IL1RL1	missense
54	102968075	rs4988957	9	20	T	C	0.41			Huvec	HMVEC-dLy-Neo,Th2	GATA2	AP-1,FAC1,p300	IL1RL1	synonymous
55	102968285	rs4988958	9	20	T	C	0.41			Huvec			AP-3,GATA,Pou3f1	IL1RL1	synonymous
56	103006104	rs6706002	9	22	A	G	0.54						AP-4, BCL, HEN1, LBP-1	IL18R1	intronic
57	102991369	rs6758936	9	22	G	A	0.54						Nkx2	IL18R1	intronic
58	103089078	rs10201184	10	6	G	C	0.43						GR	683bp 5' of SLC9A4	
59	103129692	rs11676371	10	4	G	C	0.81			HSMM, Huvec			NRSF	SLC9A4	intronic
60	103085257	rs12463588	10	6	C	G	0.43						E2F	4.5kb 5' of SLC9A4	

	pos (hg19)	SNP	HapMap LD bin	Encode LD bin	Ref	Alt	CEU	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP func annot
61	103086758	rs12469887	10	6	T	C	0.43						AP-1, HDAC2, PU.1, Pax-5, STAT	3kb 5' of SLC9A4	
62	103127963	rs12712155	10	4	A	T	0.81					CEBPB	Arid5a,Mef2,Sox	SLC9A4	intronic
63	103123301	rs13019784	10	4	A	G	0.81			HSMM	HSMMtube		Pax-5,ZBTB33	SLC9A4	intronic
64	103136309	rs1403550	10	4	T	C	0.81						CDP,Dbx2	SLC9A4	intronic
65	103136446	rs1403551	10	4	T	G	0.81						Smad3,Sox,YY1	SLC9A4	intronic
66	103132269	rs2192758	10	4	C	G	0.81						Hic1	SLC9A4	intronic
67	103086049	rs2310302	10	6	G	C	0.43						Hsf	3.7kb 5' of SLC9A4	
68	103088176	rs4140786	10	6	G	T	0.43						Ets,GATA,VDR	1.6kb 5' of SLC9A4	
69	103128866	rs4851609	10	4	T	C	0.81			NHEK, Huvec	Osteobl,HAEpiC		BCL, Egr-1, Ets, FEV, GATA, NERF1a, Nrf-2, PU.1, Pax-5, Tel2	SLC9A4	intronic
70	103137990	rs4851613	10	4	T	C	0.81						Pou2f2	SLC9A4	intronic
71	103123965	rs6708949	10	4	G	C	0.81			HSMM			Pbx3	SLC9A4	intronic
72	103133310	rs6714379	10	4	A	G	0.81						Arid3a, Dbx1, Dlx1, Hoxa5, Hoxa7, HoxB4, HoxC6, Lhx3, Mef2, Nkx6, Pax7, Pitx2, Pou2f2	SLC9A4	intronic
73	103144391	rs3849365	11	11	G	A	0.68						CDP, CEBPB, Fox, Foxa, Foxp1, HMG-IY, HNF1	SLC9A4	intronic
74	102960584	rs12712142	12	8	C	A	0.38			K562			AhR::Arnt,Arnt	IL1RL1	3'-UTR
75	102974417	rs12998521	12	8	G	T	0.38		GM12878		Th1,GM06990			IL18R1	
76	102979028	rs2287037	12	8	C	T	0.38							IL18R1	
77	102963072	rs6543119	12	8	A	T	0.38				HCM,HRCEpiC,RPTEC		Mef2,NF-kappaB	IL1RL1	intronic

Legend:
Ref: Reference allele
Alt: Alternative allele
CEU: Minor allele frequency in CEU HapMap population
GERP cons: Genomic evolutionary rate profiling conservation score
SiPhy cons: Site-specific phylogenetic analysis conservation score

(1) A549, HSMM, HSMMtube, HUVEC, Hela-S3, HNEK, HPDE6-E6E7, Myometr, RWPE1, AG10803, AoAf, HAEpiC, HCF, HCFaa, HCPEpiC, HEEpiC, HIPEpiC, HMF, LBI, HMVEC-Dbi-Neo, HPAEC, HRCEpiC, NH-A, NHDF-Ad, PANC-1, PrEC, RPTEC, SAEC, SKMC

(2) GM12878, K562, Th1, AoSMC, GM18507, GM19238, Adult_CD4_Th0, CD34+_Mobilized_sp, HAEpiC, HBMEC, HCM, HCPEpiC, HL-60, HNPCEpiC, HPAF, HRCEpiC, HRE, Jurjat, CD14+_RO01746, NB4, NH-A, NHLF

(3) GM12878, HSMM, HSMMtube, HUVEC, AoSMC, PanIslets, AG04449, AG04450, Adult_CD4_Th0, HA-sp, HAEpiC, Hac, HBMEC, HCF, HCFaa, HCM, HCPEpiC, HFF-Myc, HIPEpiC-LBI, HMVEC-dNeo, HNPCEpiC, HPAEC, HPAF, HRCEpiC, HRE, HRGEC

Refer to http://genome.ucsc.edu/ENCODE/cellTypes.html for a description of the ENCODE cell lines

Appendix B - Supplementary tables for Chapters 4 and 5

B.1 Characteristics of subjects used for gene expression multiple linear models for *TSLP* and *IL1RL1* isoforms gene expression

Variables	Mean (± SE) or N (%)
Age	59.1 ± 0.48
BMI	25.22 ± 0.18
Height	1.67 ± 0.003
Pack year	35.95 ± 1.02
Predicted FEV1	2.95 ± 0.02
Predicted FVC	3.85 ± 0.03
Weight	70.89 ± 0.57
ACEI	Yes: 105 (12.5%)
Anti_depression	Yes: 84 (10.0%)
Antibiotics	Yes: 90 (10.8%)
Anticholinergic long acting	Yes: 33 (3.9%)
Anticholinergic short acting	Yes: 79 (9.4%)
Anticoagulants	Yes: 44 (5.3%)
Antihistamine	Yes: 27 (3.2%)
Antiplatelet	Yes: 138 (16.5%)
ARB	Yes: 72 (8.6%)
Beta blockers	Yes: 113 (13.5%)
Bisphosphonates	Yes: 95 (11.4%)
Calcium channel blockers vasodilators	Yes: 122 (14.6%)
Cardiac disease	Yes: 162 (19.4%)
Combined beta agonists & anticholinergic	Yes: 54 (6.5%)
Combined beta agonists & steroid	Yes: 58 (6.9%)
Digoxin	Yes: 19 (2.3%)
Diuretics	Yes: 126 (15.1%)
Estrogen replacement	Yes: 33 (3.9%)
Gabapentine dilantin valproic acid phenobarbital	Yes: 25 (3.0%)
Hypnotic anxiolytic antipsychotics	Yes: 191 (22.8%)
Inhaled steroids	Yes: 132 (15.8%)
Insulin	Yes: 32 (3.8%)
Long acting beta agonists	Yes: 83 (9.9%)
Narcotic	Yes: 60 (7.2%)
Nasal steroid	Yes: 25 (3.0%)
No medication	Yes: 92 (11.0%)
NSAIDS	Yes: 58 (6.9%)
Oral hypoglycemics	Yes: 42 (5.0%)

Variables	Mean (± SE) or N (%)
Oral steroids	Yes: 112 (13.4%)
Progestins	Yes: 14 (1.7%)
Protein pump inhibitors	Yes: 159 (19.0%)
Sex	M: 452 (54.1%)
Short acting inhaled beta agonists	Yes: 137 (16.4%)
Smoking status	Current: 170 (20.3%), ex: 527 (63.0%), no: 139 (16.6%)
Statins	Yes: 154 (18.4%)
Theophyline	Yes: 48 (5.7%)
Thyroid replacement	Yes: 64 (7.6%)
Vitamins & minerals	Yes: 220 (26.3%)

ACEI: Angiotensin-converting-enzyme inhibitor; ARB: angiotensin II receptor blocker.

B.2 Summary of *TSLP* candidate gene association results in the literature

SNP	Chr5 location	LD r2 with rs1837253 in CEU	LD r2 with rs1837253 in study population	Phenotype	Population	N (cases / controls)	Functional analysis	Notes	References
rs11466749	110412585	<0.2	<0.2	Atopic asthma	Turkish	506 / 157	No	-	Birben E *et al*. Int Arch Allergy Immunol 2014; 163(3):185-92
rs1898671	110408002	<0.2	<0.2	Asthma	admixed american	discovery: 387 / 212; replication: 1716 / 16888	No	Association stronger in ex-smokers and non-existant in never-smokers	Liu M *et al*. PloS One 2011; 6(9):e25099
rs2289276	110407507	<0.2	<0.2	FEV1 in asthmatics without allergic rhinitis + Eosinophils count in girls	Turkish	506 / 157	No	-	Birben E *et al*. Int Arch Allergy Immunol 2014; 163(3):185-92
rs2289276	110407507	<0.2	0.44	Asthma	Han Chinese	531 / 540	No	study of the 2 SNPs only	Liu W *et al*. Exp Lung Res 2012 Oct; 38(8):375-82
rs2289276	110407507	<0.2	0.44	childhood atopic & adult asthma	Japanese	Children: 639 / 838; adults: 641 / 376	protective T allele binds Ap-2a less than C allele	No replication in European samples	Harada M *et al*. Am J Resir Cell Mol Biol 2011 Jun; 44(6): 787-93
rs2289276	110407507	<0.2	<0.2	childhood atopic asthma	Costa Rican+African-Americans + hispanics	lots	No	No replication in European samples	Hunninghake GM *et al*. Allergy 2010 Dec; 65(12): 1566-75
rs2289278	110409148	<0.2	<0.2	Asthma + lower FEV1/FVC	Han Chinese	531 / 540	No	study of the 2 SNPs only	Liu W *et al*. Exp Lung Res 2012 Oct; 38(8):375-82
rs2289278	110409148	<0.2	<0.2	lower FEV1/FVC in adult asthma	Japanese	639 / 838	No	No replication in European samples	Harada M *et al*. Am J Resir Cell Mol Biol 2011 Jun; 44(6): 787-93
rs3806933	110406742	<0.2	0.29	Vitiligo	South Korean	160 / 568	the risk allele C decreased promoter activity compared with T allele construct	-	Cheong KA *et al*- Exp dermatol 2009 Dec; 18(12): 1073-5
rs3806933	110406742	<0.2	0.29	childhood atopic & adult asthma	Japanese	639 / 838	T allele was associated with higher expression of long TSLP isoform	No replication in European samples	Harada M *et al*. Am J Resir Cell Mol Biol 2011 Jun; 44(6): 787-93