Genetic Epidemiology of Atopy and Asthma

By

Yize I Wan, BMedsci. (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

May 2011

To my family,

with all my love

Contents

Abstract

The evidence for genetic contributions to the development of asthma and atopy has been well established. Refining the major genetic factors underlying these contributions will lead to a greater understanding of the pathophysiology of these conditions and potentially identify novel therapeutic targets.

This thesis describes a series of studies designed primarily using genome-wide association (GWA) approaches to examine common single nucleotide polymorphisms (SNPs) contributing to these phenotypes in the Caucasian population. Susceptibility to atopy was assessed using both non-parametric association tests of SNPs across the genome and focused analysis of two regions on chromosomes 3p22.1-q22.1 and 17p12-q24.3 previously identified through a meta-analysis of genome-wide linkage studies (GSMA). The discovery cohort consisted of 1,083 cases and 2,770 controls, replication analyses were undertaken in four independent population cohorts. A GWA study of severe asthma was carried out in 933 cases and 3,346 controls with replication in a further 231 cases and 1,345 controls. The contribution of SNPs within all previously reported asthma susceptibility loci identified using a comprehensive literature search was also evaluated.

Overall, there is evidence for a large number of loci influencing both atopy and severe asthma, each harbouring modest effects. A number of potentially novel loci meeting nominal significance in both GWA studies have been identified requiring further work. Strong evidence was found to support the IL1RL1-IL33 signalling pathway in asthma pathogenesis. Molecular characterisation of the 5' untranslated regions of *IL1RL1* and *IL33* suggest a complex regulatory role of identified common variants involving multiple promoters for *IL1RL1*. A number of asthma specific variants within the chromosome 2q12 and 9p24.1 regions were detected using next generation re-sequencing in homogenised pools of cases and controls warranting further analyses.

This work has identified a potentially important pathway in which to focus the development of effective asthma therapies. Future directions will include functional analysis of replicated variants and tests of interactions between the multiple genetic and environmental factors likely to be involved in disease.

Acknowledgements

I would like to thank Prof. Ian Hall for giving me the opportunity to undertake this PhD under his continual encouragement, guidance, and support. This has allowed me to build a solid foundation in scientific research as well as broaden my horizons and expand my interests before I embark upon the rest of my career. Thank you for providing invaluable advice, reassurance, and perspective as well as always being generous with your time and good nature. I am very grateful to Dr. Ian Sayers for his consistent supervision and practical direction throughout my PhD. Thank you also for teaching me to be methodical, structured, and thorough in my work as well as always challenging me to improve and pushing me to reach my highest potential.

I would like to stress that none of this work would have been possible without the contribution of thousands of individuals, from study participants to administrative staff, whether they are aware of it or not. Thank you also to all of our collaborators for their help and efforts in making progress with this work in order to achieve our shared goals.

I am extremely grateful to the University of Nottingham for all of the education and experiences I have gained over the past six years. This has been an intense time of personal development and learning to deal with some of life's trials and tribulations; this would not have been possible without the support of those around me. I would also like to thank the friends I have made both in Therapeutics and at the University of Leicester for all of their help and advice as well as providing a healthy work-life balance with lots of fun times and trips out.

Finally, I dedicate this thesis to my family in particular my parents, who have shaped the person I am. Thank you for your unwavering support in all that I do and doing everything that was in your power to help me realise my dreams. Thank you for teaching me humility, integrity, and perserverance; for instilling the confidence to keep challenging myself; and for cultivating the belief that anything is possible. I hope I have made you proud.

Abbreviations

AAGC	Australian Asthma Genetics Cohort
ALSPAC	Avon Longitudinal Study of Parents and Children
AUGOSA	Asthma UK Genetics of Severe Asthma
ATCC	American Type Culture Collection
ATS	American Thoracic Society
β_2-agonist	β_2-adrenergic receptor agonist
B58C	British 1958 Birth Cohort
BEAS-2B	Bronchial epithelial cell line
BHR	Bronchial hyper-responsiveness
CAP-RAST	Allergosorbent test
CEU	European ancestry in Utah
Chr	Chromosome
CI	Confidence interval
CNV	Copy number variation
ELISA	Enzyme linked immunosorbent assay
ERS	European Respiratory Society
Fc Epsilon R1	High affinity IgE receptor
FEV_1	Forced expiratory volume in the 1st second
GABRIEL	Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community
GINA	Global Initiative for Asthma
GWA	Genome-wide association
GWL	Genome-wide linkage
HASM	Human airway smooth muscle
HBEC	Human bronchial epithelial cell
HLA	Human leukocyte antigen
HMC-1	Human mast cell line
HWE	Hardy-Weinberg equilibrium
ICS	Inhaled gluco-corticosteroid
IgE	Immunoglobulin E
IL1RL1	Interleukin 1 receptor-like 1
IL18R1	Interleukin 18 receptor 1
IL33	Interleukin 33
Indel	Insertion/deletion polymorphism
LD	Linkage disequilibrium
LOD	Log_{10} ratio of the odds of two loci being linked compared to the odds of observing the data by chance
MAF	Minor allele frequency
MHC	Major histocompatibility complex
NGS	Next-generation sequencing
OR	Odds ratio
ORMDL3	ORM1-like protein 3
PBMC	Peripheral blood mononuclear cell
PCA	Principal components analysis
PCR	Polymerase chain reaction
PEF	Peak expiratory flow
Q-Q plot	Quantile-Quantile plot
RACE	Rapid amplification of cDNA ends
RAST	Radioallergosorbent test
SD	Standard deviation
SNP	Single nucleotide polymorphism
Soton	Southampton Cohort
SPT	Skin prick test
T1DGC	Type 1 Diabetes Genetics Consortium
TENOR	The Epidemiology and Natural History of Asthma: Outcomes and Treatment Regimes
Th	T helper
THP-1	Human acute monocytic leukemia cell line
WTCCC2	Wellcome Trust Case Control Consortium Phase 2

Publications Relevant to Thesis

Full Papers

Wan YI, Strachan DP, Evans DM, Henderson J, McKeever T, Holloway JW, Hall IP, Sayers I. **A genome-wide association study to identify genetic determinants of atopy in subjects from the United Kingdom.** J Allergy Clin Immunol 2011; 127(1):223-31, 231.e1-3. Epub 2010 Nov 20.

Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson WO; GABRIEL Consortium. **A large-scale, consortium-based genomewide association study of asthma.** N Engl J Med 2010; 363(13):1211-1221.

Wan YI, Shrine NRG, Soler-Artigas M, Wain LV, Blakey JD, Moffatt MF, Bush A, Cookson WOCM, Strachan DP, Heaney L, Al-Momani BAH, Mansur AH, Manney S, Thomson NC, Chaudhuri R, Brightling CE, Bafadhel M, Singapuri A, Niven R, Simpson A, Holloway JW, Howarth PH, Hui J, Musk AW, James AL, Ferreira MA, Baltic S, Martin NG, Thompson PJ, Tobin MD, Sayers I, Hall IP. **Genome-Wide Association Study To Identify Genetic Determinants Of Severe Asthma (AUGOSA).** (Under submission).

Oral Communications

Wan YI, Strachan DP, Stefansson K, Halapi E, McKeever T, Holloway JW, Hall IP, Sayers I. **Genome-Wide Association Study to Identify Genetic Determinants of Atopy.** J Allergy Clin Immun 2010; 125; 2(1): AB193. AAAAI International Conference 2010.

Wan YI, Shrine NRG, Soler-Artigas M, Wain LV, Moffatt MF, Bush A, Cookson WOCM, Strachan DP, Heaney L, Al-Momani BAH, Mansur AH, Manney S, Thomson NC, Chaudhuri R, Brightling CE, Bafadhel M, Singapuri A, Niven R, Simpson A, Holloway JW, Howarth PH, Hui J, Musk AW, James AL, Tobin MD, Sayers I, Hall IP. **Comprehensive Evaluation Of**

Mild-To-Moderate Asthma Genes In Severe Asthma (AUGOSA). Am J Respir Crit Care Med 2011; 183: A6374. ATS International Conference 2011.

Abstracts (Poster Discussion)

Wan YI, Strachan DP, Hall IP, Sayers I. **Fine mapping of Atopy Loci on Chromosomes 3 & 17.** Am J Respir Crit Care Med 2009; 179: A5410. ATS International Conference 2009.

Wan YI, Shrine NRG, Soler-Artigas M, Wain LV, Moffatt MF, Bush A, Cookson WOCM, Strachan DP, Heaney L, Al-Momani BAH, Mansur AH, Manney S, Thomson NC, Chaudhuri R, Brightling CE, Bafadhel M, Singapuri A, Niven R, Simpson A, Holloway JW, Howarth PH, Hui J, Musk AW, James AL, Tobin MD, Sayers I, Hall IP. **Genome-Wide Association Study To Identify Genetic Determinants Of Severe Asthma (AUGOSA).** Am J Respir Crit Care Med 2011; 183: A6170. ATS International Conference 2011.

Abstract (Thematic Presentation)

Shrine NRG, Wain LV, Soler-Artigas M, Wan YI, Sayers I, Hall IP, Tobin MD. **The effect on GWAS association tests of phasing cases and controls together or separately before imputation.** European Mathematical Genetics Meeting 2011.

Chapter 1

General Introduction

1.1 Allergic Disease

1.1.1 Definition

Allergy is a hypersensitivity reaction initiated by immunological mechanisms resulting from antigen-induced changes in reactivity (Von Pirquet 1946). Allergic diseases such as asthma, allergic rhinitis, and atopic dermatitis are characterised by inflammatory responses mediated by immunoglobulin E (IgE) to environmental allergens resulting in both acute and chronic responses in a range of end organs and tissues (Johansson *et al.* 2004).

1.1.2 Epidemiology

Over recent decades, the global prevalence of allergic disorders has increased substantially (Bajekal and Prescott 2003). This has been associated with increasing allergic sensitisation in the developed world (Masoli *et al.* 2004). Although this trend appears now to be stabilising, allergic diseases remain extremely common and continue to represent an important source of patient morbidity and healthcare utilisation (Gupta *et al.* 2007). In the UK, 39% of children and 30% of adults have been diagnosed with one or more atopic disorder. Furthermore, approximately 10% of primary care prescribing costs is utilised on treatment for these conditions and direct NHS costs for managing allergic problems are estimated at over £1 billion per year (Gupta *et al.* 2004).

1.1.3 Pathogenesis

Allergic diseases result from an augmented response of the immune system to external substances. An acute reaction in response to allergen exposure is known as a type I hypersensitivity reaction and is characterised by the excessive activation of leukocytes by IgE antibodies resulting in the induction of an extreme inflammatory response. Type 1 immediate

hypersensitivity results from two temporally distinct processes when an allergen is inhaled, ingested or injected into a susceptible individual [Figure 1.1].

The primary sensitisation stage occurs when the allergen is presented to the immune system inducing IgE production. Allergen uptake and presentation by dendritic cells activate T-helper (Th) lymphocytes leading to the release of cytokines such as interleukin 4 (IL4), interleukin (IL5), and interleukin 13 (IL13). This activates B lymphocytes resulting in the secretion of specific IgE antibodies. These antibodies then bind to mast cell membranes via the high affinity IgE receptor (Fc epsilon R1). During subsequent provocation upon re-exposure to the allergen and internalisation at a later date, the allergen binds to membrane bound IgE molecules on the surface of immune cells such as eosinophils, mast cells, and basophils. This induces degranulation of pre-formed mediators and new vasoactive mediator synthesis, triggering the release of various inflammatory mediators including histamine, prostaglandins, and leukotrienes leading to the clinical symptoms of allergy ranging from rash to anaphylaxis [Table 1.1] (Holgate and Church 1993).

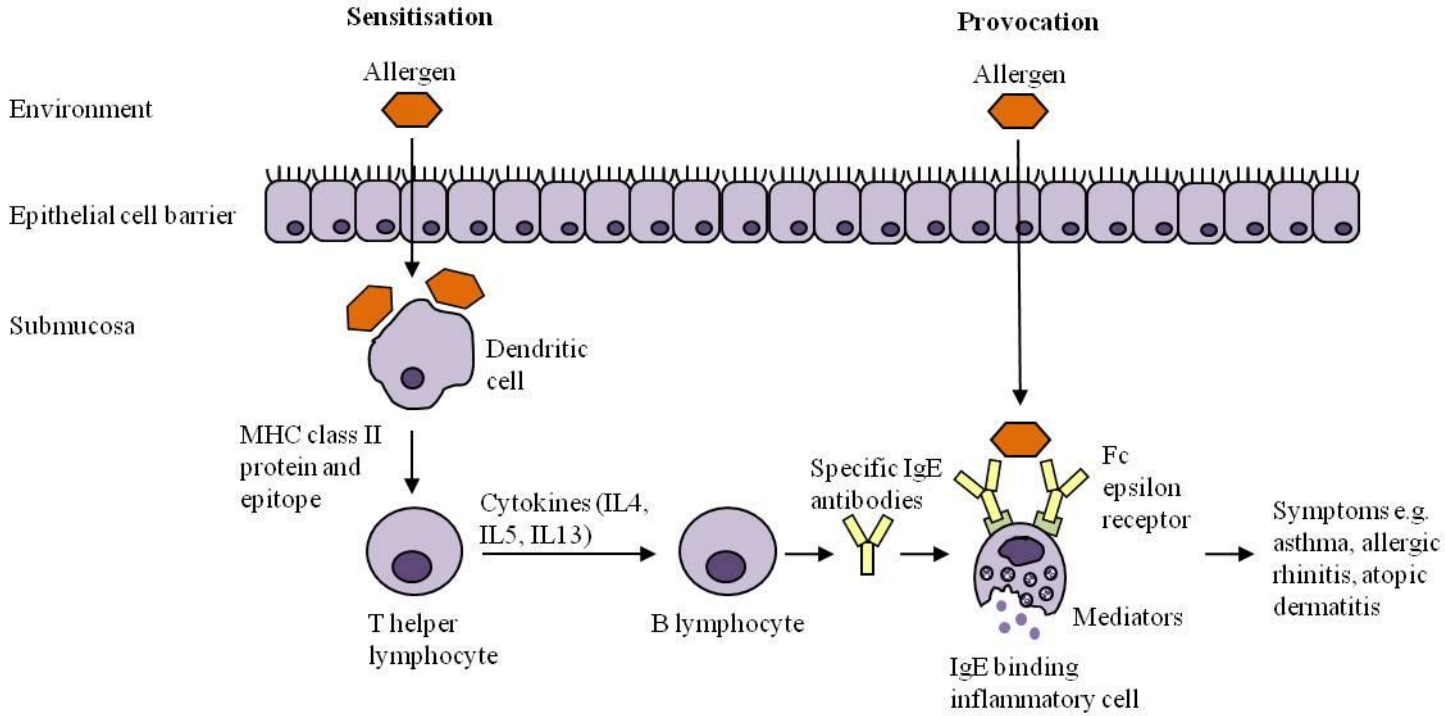


Figure 1.1 - Immunological mechanisms involved in allergic disease. Primary sensitisation following antigen exposure induces secretion of specific IgE antibodies. Subsequent provocation by the same allergen triggers inflammatory mediator release. Adapted from (Holgate and Church 1993).

Table 1.1 - Products of mast cell degranulation.

Mediator	Biological effect
Preformed and stored within granules	
Histamine	Vasodilatation, chemotaxis, broncho-constriction, mucus secretion, increased capillary permeability
Tryptase	Activates complement factor 3 in innate immunity
Eosinophil Chemotactic factor	Eosinophil chemotaxis
Neutrophil Chemotactic factor	Neutrophil chemotaxis
Newly synthesised	
Leukotrienes e.g. LTA_4, LTB_4, LTC_4, LTD_4, LTE_4, and LTF_4	Increased vascular permeability, chemokines and mucus secretion, smooth muscle contraction
Prostaglandins e.g. PGI_2, PGE_2, and $PGF_{2\alpha}$	Broncho-constriction, platelet aggregation, vasodilatation
Thromboxanes	Broncho-constriction
Platalet-activating factor	Broncho-constriction, eosinophil and neutrophil chemotaxis

The type 1 or early-phase reaction is often followed by a late-phase reaction in many individuals developing 2 to 6 hours post allergen exposure and peaking at 6 to 9 hours. Although this typically resolves within 2 days, symptoms in the skin involve pain, oedema, and erythema, and in the lungs are characterised by airway narrowing and mucus hypersecretion. The late-phase reaction results from local recruitment and activation of leukocytes such as Th2 cells, eosinophils and basophils as well as the sustained production of mediators from IgE activated mast cells and allergen primed T cells (Kay 2001).

Prolonged or repetitive allergen exposure to specific allergens can result in chronic inflammation with significant leukocyte recruitment, extracellular matrix alteration, and structural cell modification at the affected site (Galli *et al.* 2008).

1.1.4 Allergens

A range of allergens are recognised as being able to induce hypersensitivity reactions in susceptible individuals. Properties which make a substance an allergen may facilitate access to the immune system in a susceptible host dependent on a number of factors such as its molecular complexity, biochemical characteristics, as well as the genetic predisposition of the host (Aas 1978).

1.1.4.1 Sources of Allergens

Environmental allergens can be classified into two main types, the first involving the induction of IgE production by sensitising the subject and subsequently causing an allergic reaction upon re-exposure. Common such allergens include house dust mite faecal particles, grass pollen, animal dander, and certain medicines. Most allergens are proteins or haptens and many inhalant allergens such house dust mite serine proteases, *Dermatophagoides pteronyssinus* 1 (Der p 1) and 9 (Der p 9) exhibit enzymatic activity in hypersensitivity (Robinson *et al.* 1997). The second type of allergen is classified as inducing an adaptive immune response to exogenous antigens such as bacteria or toxins, and endogenous antigens produced by viruses replicating within a host cell. Responses to these allergens cause local inflammation independent of IgE production (Janeway *et al.* 2001).

1.1.4.2 Allergen Sensitisation Testing

A number of techniques have been developed to assess IgE-mediated hypersensitivity to allergens. The most widely used allergy test is the skin prick test (SPT) using aqueous solutions of a variety of allergens. The skin test lancet is injected through drops of solution to prick the superficial epidermal layer of the skin. Introduction of the antigen and a positive test results in raised red lesions. The size and flare of these wheals are measured 15 to 20 minutes post-application and the number of positive reactions, commonly defined as wheals ≥3mm in diameter, summated as an index of skin test positivity. Both negative antigen-free and positive histamine controls are required, for example, 50% glycerol and 1mg/ml histamine hydrochloride respectively (Ten *et al.* 1995).

Serum or plasma levels of IgE production, either total levels or specific to different allergens after allergen exposure can be measured *in vitro*. The radioallergosorbent test (RAST) uses an immobilised test antigen bound to an insoluble matrix such as paper or agarose beads. When patient serum is added, allergen specific IgE antibodies bind to the allergen complex. Subsequent addition of radioactively labelled anti-IgE allows the estimation of specific IgE production in an individual (Ortolani *et al.* 1989). A positive allergen result is typically defined by levels of specific IgE ≥0.35KU_A/L (Boyano Martinez *et al.* 2001). Non-radioactive versions

of this technique include the allergosorbent test (CAP-RAST) and enzyme linked immunosorbent assay (ELISA) using fluorescence and enzyme substrates (Luczynska *et al.* 1989; Darsow *et al.* 1995). These methods can be more specific and less likely to be influenced by concurrent drug treatment but tend to be more expensive and time consuming.

The above techniques used in allergen sensitisation testing have been shown to be significantly associated although the quantitative correlation remains weak, particularly in individuals with low levels of sensitivity (Stenius *et al.* 1971; Norman *et al.* 1973; Bryant *et al.* 1975). Reasons for this may include the use of non-standardised allergenic extracts or patient-dependent factors such as differing abilities in mast cell release (Witteman *et al.* 1996). Furthermore, previous studies have also shown that total IgE levels raised due to helminth parasite infections can cause mast cell Fc epsilon R1 saturation resulting in lower detected specific IgE antibody responses against given allergens (Godfrey and Gradidge 1976).

1.1.5 Atopy

Atopy is the predisposition to an IgE isotype production and the development of immediate hypersensitivity in response to exposure to common environmental allergens (Jackola *et al.* 2002). The presence of atopy can be defined by elevated levels of serum total IgE, specific IgE, positive SPT to common allergens or a combination of these markers.

The heritable tendency to produce allergen-specific IgE is one of the strongest risk factors for the development of allergic diseases such as asthma, allergic rhinitis, and atopic dermatitis (Coca and Cooke 1923; Burrows *et al.* 1989). The prevalence of asthma increases with increasing serum IgE concentrations (Sunyer *et al.* 1995). In addition, 60% of asthma subjects show immediate hypersensitivity to defined environmental agents such as nitrogen dioxide and ozone as well as commonly suffer from other atopic diseases such as allergic rhinitis and atopic dermatitis (Valacer 2000).

1.1.6 Clinical Manifestations of Allergic Disease

Allergic hypersensitivity to environmental allergens can affect many tissue types in the human body such as the lungs, the nasal tract or the skin resulting in a wide range of end-organ manifestations of disease.

1.1.6.1 Asthma

Asthma is a chronic inflammatory condition of the airways characterised by recurrent episodes of reversible airway obstruction and increased bronchial hyper-responsiveness (BHR) (Dantzker *et al.* 1987). According to WHO estimates, approximately 300 million people suffer from asthma worldwide with an additional 100 million predicted to be diagnosed by 2025, as a result of increasing urbanisation (Masoli *et al.* 2004). Symptoms of asthma include cough, wheeze, chest tightness, and dyspnoea particularly at night and during the early morning; these may be severe and potentially life-threatening during acute exacerbations which can occur episodically and vary in onset. Many well characterised sub-types of asthma have been identified such as atopic, aspirin-induced, and occupational asthma. The immediate asthmatic response is mediated by IgE dependent mast cell release of mediators such as histamine and leukotrienes. The late reaction is a manifestation of eosinophilic airway inflammation associated with the development of airway hyper-responsiveness.

In asthma, repeated allergen exposure and inflammatory responses can cause epithelial damage exposing deeper airway structures. Inflammatory and structural cell production of growth factors and pro-fibronogenic cytokines lead to angiogenesis, smooth muscle proliferation, and basement membrane thickening. Along with cellular infiltration of eosinophils and mononuclear cells, this pathological remodelling leads to fibrosis and chronic BHR [Figure 1.2] (Hogg 1997).

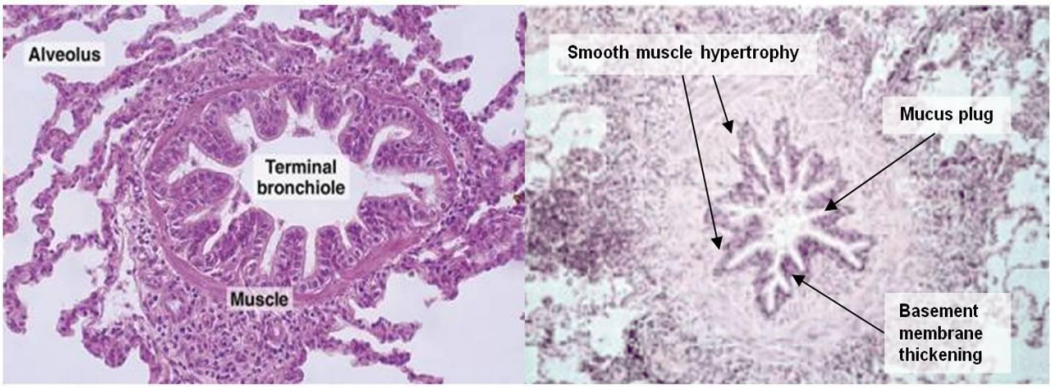


Figure 1.2 – Inflammatory and structural changes in asthma. Haematoxylin and Eosin stained micrographs showing terminal bronchioles from a normal subject (left) and a subject who died during an asthma attack (right), taken from (Saetta and Turato 2001; Junqueira and Carneiro 2009). In comparison, the asthma subject sample displays smooth muscle hypertrophy, thickening of the basement membrane, and a marked inflammatory process in the airway wall mainly characterised by eosinophils (Roche 1998; Holgate 2000).

Clinical care is complicated due to heterogeneity in disease expression and severity resulting in variability of clinical symptoms, as well as a range of physiological and pathological abnormalities in individuals (Wenzel *et al.* 1999). Currently utilised guidelines for treatment are based on four to six steps describing clinical characteristics including frequency of symptoms, lung function, as well as requirement for preventer and relief medication (National Asthma Education and Prevention Program 2007; Global Initiative for Asthma 2008) (described in more detail in Section 1.1.7). Patients with severe or difficult-to-treat asthma constitute approximately 5% to 8% of the total asthma population (Holgate *et al.* 2011). This sub-group of patients disproportionately consume healthcare resources related to asthma and contribute the largest proportion of morbidity and mortality (The ENFUMOSA Study Group 2003). Multiple distinct subgroups of asthma phenotypes have been identified using cluster analysis (Haldar *et al.* 2008; Moore *et al.* 2010). Within the overall asthma population, patients were found to fall mainly into two groups consisting of early-onset atopic individuals and obese individuals with non-eosinophilic airway inflammation (Haldar *et al.* 2008). Multiple subgroups were identified within severe asthmatics; these clusters differed mainly in lung function, age of onset, and atopic status [Figure 1.3] (Haldar *et al.* 2008; Moore *et al.* 2010).

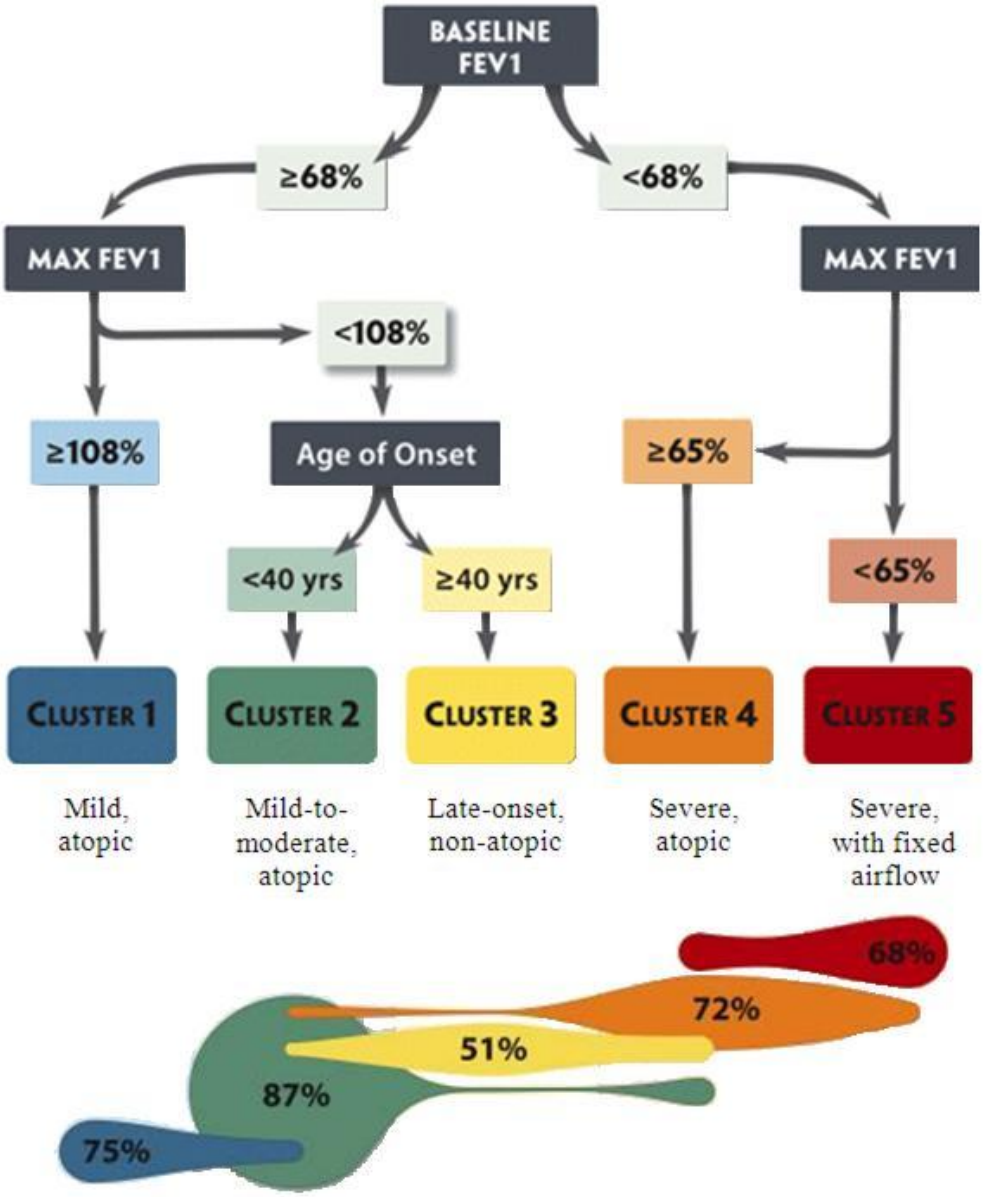


Figure 1.3 – Cluster analysis showing 5 distinct, severe asthma phenotypes ranging from milder asthma (Cluster 1) to more severe disease (Clusters 4 to 5). 80% of 726 subjects from the Severe Asthma Research Program (SARP) were successfully clustered, the size of coloured figures are proportional to frequency of individuals. Adapted from (Moore *et al.* 2010).

1.1.6.2 Allergic Rhinitis

Allergic rhinitis is an inflammatory disease of the nasal mucosa characterised by pruritis, irritation, and excessive mucus production in response to specific environmental allergens such as seasonal plants, airborne chemicals, and dust particles (Skoner 2001). Although frequently under diagnosed, Allergic rhinitis is the most common manifestation of allergic disease with a

high prevalence ranging from 21% to 26% in Europe (Bauchau and Durham 2004). It is a major healthcare burden as well as having an important impact in determining an individual's quality of life. Allergic rhinitis is an important co-morbidity in asthma with 20% to 30% of clinically diagnosed patients also having asthma (Bousquet *et al.* 2008). Conversely, up to 80% of asthma patients have also been shown to exhibit symptoms of allergic rhinitis (Shaaban *et al.* 2008).

1.1.6.3 Atopic Dermatitis

Atopic dermatitis is a chronically relapsing, inflammatory skin disease resulting in symptoms such as pruritis and eczematous lesions typically presenting in early childhood (Leung *et al.* 2003). The skin of an affected individual reacts easily and abnormally to irritants, food, and environmental allergens increasing vulnerability to bacterial infections. The prevalence of atopic dermatitis is approximately 15% in industrialised countries but has been difficult to establish since no standardised diagnostic criteria exist (Williams *et al.* 1999). There is increasing evidence for a genetic component in determining atopic dermatitis through breakdown of the dermal barrier as well as support for its role in the 'atopic march' (Hanifin 2009). This hypothesis describes the phenomenon whereby some children with atopic dermatitis also go on to develop asthma and/or allergic rhinitis in later life (Spergel 2010).

1.1.7 Management

Management for allergic diseases is primarily based on control of symptoms. A number of non-pharmacological methods have been proposed such as minimising allergen exposure in sensitised individuals in order to reduce both the frequency and severity of immediate hypersensitivity responses (Ownby *et al.* 2002; Simpson and Custovic 2004; Lucas *et al.* 2005). However, this strategy can be ineffective particularly for air-borne allergens such as grass pollens. Furthermore, more studies are required in this area as there is currently insufficient evidence to support the effectiveness of these strategies.

A range of specific pharmacotherapy agents targeting the immunologic response are currently available with differing value dependent on the type and severity of the allergic disorder. In

mild disease, allergic rhinitis is predominantly managed using anti-histamines targeting the histamine H1 receptor such as diphenhydramine (Scadding *et al.* 2008). Emollients form the basis for atopic dermatitis management. Additionally, topical steroids are commonly used in both diseases with increasing severity of symptoms. Asthma management requires the use of a β_2-adrenergic receptor agonist (β_2-agonist) such as salbutamol to induce bronchodilation and an inhaled gluco-corticosteroid (ICS) such as beclamethasone to reduce inflammation (Barnes and Barnes 2006; Holgate *et al.* 2008). Some patients are prescribed cysteinyl leukotriene modifiers either in the form of receptor antagonists or synthesis inhibitors and others with more severe forms of asthma can require the use of IgE targeting monoclonal antibodies such as omalizumab (Holgate *et al.* 2005).

As mentioned in Section 1.1.6.1, asthma severity can be classified on the basis of the intensity of treatment required to achieve good asthma control. Mild asthma can be well-controlled with low-intensity treatment such as low-dose ICS, leukotriene modifiers or cromones. Partly-controlled asthma is based on having any of the following characteristics in a week: daytime symptoms, limited activity, nocturnal symptoms, need for reliever medication, <80% predicted lung function based on forced expiratory volume in the 1st second (FEV_1) or peak expiratory flow (PEF), and exacerbations occurring one or more times per year. Symptoms are uncontrolled if three or more features of partly-controlled asthma are present in a given week or if exacerbations occur weekly. According to the Global Initiative for Asthma (GINA) guidelines for the management of asthma, five treatment steps have been defined with increasing disease severity [Table 1.2] (Bateman *et al.* 2008).

Table 1.2 – GINA defined management approach based on control. Adapted from (Bateman *et al.* 2008).

Treatment steps				
Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**
Asthma education, environmental control				
As needed rapid-acting β_2-agonist	As needed rapid-acting β_2-adrenergic receptor agonists			
Controller options (ICS, long-acting β_2-agonist)	*Select one*	*Select one*	*To Step 3, select one or more*	*To step 4, add either*
	Low-dose ICS	Low-dose ICS plus long-acting β_2-agonist	Medium or high-dose ICS plus long-acting β_2-agonist	Oral gluco-corticosteroid (lowest dose)
	Leukotriene modifier	Medium or high-dose ICS; Low-dose ICS plus leukotriene modifier	Leukotriene modifier; Sustained release theophylline	Anti-IgE treatment
		Low-dose ICS plus sustained release theophylline		

Variability in asthma control is to an extent due to heterogeneity in drug treatment responses in allergic diseases. Widespread differences even in patients with seemingly identical clinical phenotypes have been investigated using pharmacogenetic approaches (Hall and Sayers 2007). Part of this variability could be due to differences in the genetics of individual patients, for example, polymorphisms in the genes encoding the β_2-adrenergic receptor and 5-lipoxygenase enzyme family have been shown to influence response to β_2-agonist and anti-leukotriene therapies respectively (Drazen *et al.* 2000).

1.1.8 Susceptibility to Disease

Atopic disorders are complex multi-factorial diseases and do not follow classic Mendelian inheritance patterns. A large number of interacting genetic and environmental factors are involved in determining disease expression (Sibbald and Turner-Warwick 1979; Barnes 1999).

1.1.8.1 Environmental Factors

Environmental factors such as exposure to house dust mite, tobacco smoke, and viral infections are important determinants of asthma and atopic disease such as allergic rhinitis (Carlsen *et al.* 1984; Skoner 2001). Although debated, one theory termed the hygiene hypothesis suggests that childhood exposure to lower levels of bacterial or parasite antigens may in turn result in a

tendency for T lymphocytes to polarise towards a Th2 rather than a Th1 cytokine profile and so increase risk of allergic disease in later life (Wills-Karp *et al.* 2004). Common precipitants to asthma exacerbations include exercise particularly in cold weather, exposure to airborne allergens or pollutants and viral upper respiratory tract infections (Lemanske *et al.* 2003). For example, 60% of asthma subjects exhibit immediate hypersensitivity to defined environmental allergens such as nitrogen dioxide and ozone (Carlsen *et al.* 1984; Valacer 2000).

In addition to complex interactions of genes within multiple pathways, environmental factors are also important in the expression of genetic predisposition. Most current hypotheses propose that changes in environmental factors favour the manifestation of allergic disease in those who are genetically susceptible and it is the interaction between these that leads to clinical disease (Ober and Vercelli 2011).

1.1.8.2 Disease Heritability

A genetic predisposition to the development of allergic diseases such as asthma, allergic rhinitis, and atopic dermatitis has been recognised since the early 1900s. In 1916, Cooke and Van der Veer reported that almost 50% of patients with allergic rhinitis and asthma had a positive family history of similar illness (Cooke and Veer 1916). Since then, the genetic component of susceptibility to atopy has been confirmed by numerous epidemiological studies showing familial clustering of disease and concordance rates amongst twins (Meyers *et al.* 1995).

The heritability of asthma has been reported at as high as 75% and with each additional atopic first degree family member, the 3% risk of developing atopic dermatitis in the general population increases by two-fold (Duffy *et al.* 1990; Kuster *et al.* 1990). Correlation rates between twins for intradermal skin testing were significantly lower in dizygotic pairs (r^2=0.46) compared to in monozygotic twin pairs (r^2=0.82) (Hopp *et al.* 1984). This was similarly demonstrated for allergic rhinitis where a 45% to 60% concordance rate reduces to 25% and for atopic dermatitis, a rate of 72% to 86% falls to 21% to 23% in monozygotic compared to in dizygotic twins (Schultz Larsen and Holm 1985; Larsen *et al.* 1986; Feijen *et al.* 2000).

Although a genetic contribution to atopic traits has been well established, heritability estimates of various atopic phenotypes vary, ranging from 34% to 74% (Koppelman *et al.* 2002). Numerous studies have been carried out to identify the inheritance pattern of these allergic diseases including using segregation analyses resulting in many different genetic models being proposed (Los *et al.* 1999).

1.2 Genetics of Complex Disease

1.2.1 Statistical Approaches

Identifying the genetic determinants of complex, multifactorial diseases has been more difficult than that of single gene traits. Generally, three main approaches have been used to detect loci and genes that are implicated in disease transmission: linkage and association in populations presumed to be ancestrally homogenous, as well as admixture mapping using multiple genetically diverse populations [Figure 1.4]. Admixture mapping involves combining known population history with information from subjects' measured genotypes using known ancestry-informative markers within admixed populations. This thesis will focus on linkage and association approaches.

Linkage analysis tests for co-segregation of genetic markers with the disease phenotype using family-based cohorts such as multiplex extended pedigrees and nuclear families with affected sib-pairs (Risch 1990). This approach tests genetic variation across the genome using linkage markers based on a positional map of genetic markers relative to each other in terms of recombination frequency. The greater the frequency of segregation during crossover of homologous chromosomes occurring between two genetic markers, the further apart they are assumed to be.

Association studies examine the co-occurrence of a marker and disease between cases and controls to determine if the marker occurs more frequently in people with the disease compared to those without. This approach can be carried out using both family-based and unrelated subjects commonly using single nucleotide polymorphism (SNP) markers based on a map of their physical distances along each chromosome. Over 12 million SNPs have been deposited in NCBI GeneBank, approximately one million of which have been validated with identified SNPs across the genome occuring approximately every 500 base pairs (bp) (Sachidanandam *et al.* 2001; 1000 Genomes Project Consortium *et al.* 2010).

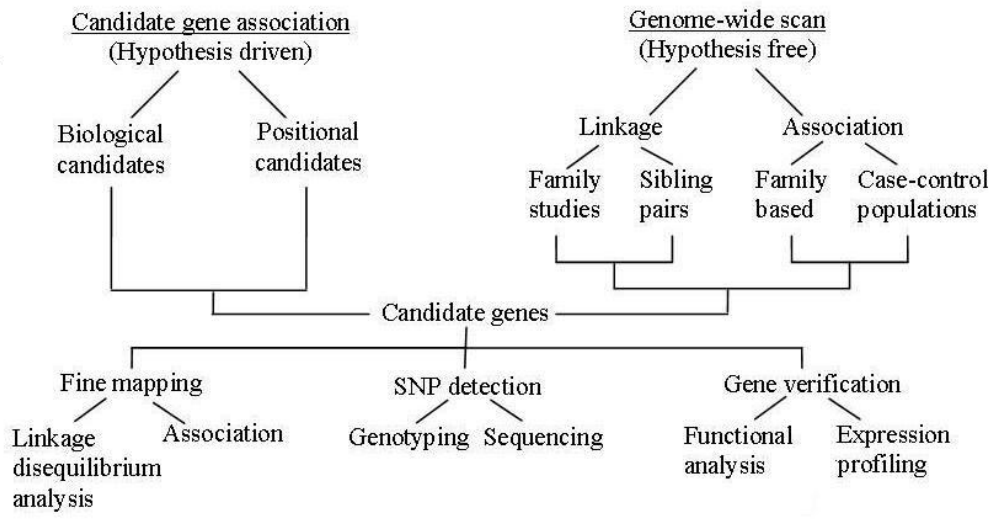


Figure 1.4 - Approaches to complex disease gene discovery.

1.2.2 Linkage Analysis

Genetic linkage is the deviation of loci from independent assortment during meiosis resulting in a tendency for certain loci or alleles located close in chromosomal position to be inherited together. Linkage analysis uses the assumption that a genetic marker and the disease causing variant are in physical proximity and therefore results in transmission from parent to offspring more often than expected under independent inheritance (Mendel's second law). The basis for using this approach comes from the evidence for genetic heritability of a trait found using family studies. Maximum likelihood of disease ratio methods are used to assess the likelihood that a trait co-segregates with a genetic marker, expressed as an LOD score measure of the $\log_{10}$ ratio of the odds of two loci being linked compared to the odds of observing the data by chance (Morton 1955). In genome-wide linkage studies using more than 300 markers, a value of +3.6 is taken as significance evidence for linkage (Feijen *et al.* 2000). Once loci have been located, fine mapping can be carried out to locate candidate disease susceptibility genes and subsequent association analyses termed positional cloning are often used to provide additional evidence for genes in determining disease [Figure 1.4].

Studies based on linkage approaches are hypothesis-free thereby offering the potential to identify novel susceptibility genes, particularly those in pathways not obviously implicated in allergic disease. Linkage is able to identify multiple regions simultaneously and is particularly

useful in detecting genes of modest to large effects usually with an odds ratios (OR) greater than 5 (Ober and Hoffjan 2006). The use of this approach has successfully identified genes such as breast cancer 1, early onset (*BRCA1*) in determining breast cancer through mapping the linkage signal (LOD 5.98) on chromosome 17q21 by positional cloning (Hall *et al.* 1990). However, the genetic component for the majority of complex diseases is likely to be composed of multiple genes with only modest effects. Evidence provided by linkage studies can have limited statistical power and methods in following up linked regions such as high-throughput sequencing can be costly and labour intensive, in particular when large numbers of broad linkage regions are identified (Wills-Karp *et al.* 2004).

1.2.2 Association Analysis

The aim of an association study is to determine whether affected individuals are more likely to have a particular allele of a marker or gene than the unaffected individuals in the same population. This is based on the common disease, common variant hypothesis which states that genetic variants in the population affecting risk of common diseases may be polymorphisms that are common in the whole population with a minor allele frequency (MAF) greater than 5% (Risch and Merikangas 1996). It is assumed that these variants have not been eliminated through natural selection due to low effect sizes, genetic drift given small effective human population sizes, and different selection pressures in modern society. Furthermore, adaptation to environmental changes such as levels of pathogen exposure may have resulted in the selection for alleles that confer increased protection against infections but predispose to autoimmunity, for example, in interleukin 18 receptor accessory protein (*IL18RAP*) (Sironi and Clerici 2010). Each of these common genetic variants will contribute a small, additive or multiplicative effect resulting in part, in the evolutionarily neutral state of these common diseases.

The candidate gene approach utilises association analysis to assess the contribution of polymorphisms in genes with hypothesised biological functions in determining disease [Figure 1.3]. Association is determined when differences in allele frequencies are found to be statistically significant between a disease affected case population and an unaffected control

population (Cardon and Bell 2001). This approach has been used to identify common disease variants such as the apolipoprotein E (*APOE*) ε4 allele in Alzheimer's disease, present in the general population at approximately 30% (Corder *et al.* 1994). Association between markers and disease can implicate candidate genes and be used to determine true causation suggesting that the markers or genes are involved in disease risk. However, association can also result from the marker being in linkage disequilibrium (LD) with another disease causing allele or gene. The subsequent non-random association of these alleles mean that they are not transmitted independently thus increasing the rate of false-positive association results (Baron 2001). False-positive results in association studies can also be attributed to the potential for population stratification and the value of results depends on the selection of candidate genes with the greatest involvement in disease pathophysiology (Hirschhorn *et al.* 2005).

1.2.3 Genome-Wide Association (GWA) Studies

A genome-wide association (GWA) study is a powerful, recent approach that rapidly scans thousands of SNP markers across complete sets of deoxyribonucleic acid (DNA) or genomes of case-control cohorts in order to find genetic variations associated with a particular disease (Hirschhorn *et al.* 2005). This approach was made possible following the completion of the Human Genome Project in 2003 and the International HapMap Project in 2005 through providing computerised databases that contain the reference human genome sequence, a map of human genetic variation and new technologies that can quickly and accurately analyse whole-genome samples (Sachidanandam *et al.* 2001; Venter *et al.* 2001; The International HapMap 2003). Of the over 10 million common SNPs believed to be present in the human genome, these variants have been demonstrated to cluster into haplotypes groups. This allows for a large proportion of SNPs across the genome to be accurately sampled using fewer, carefully chosen SNPs (Gabriel *et al.* 2002). GWA studies have the ability to identify novel disease genes and pathways simultaneously including those with small effect sizes (Hirschhorn *et al.* 2005). As a result, they have been used extensively to study common diseases including coronary artery disease, Crohn's disease, rheumatoid arthritis, as well as type 1 and type 2 diabetes (Smyth *et al.* 2006; Petretto *et al.* 2007; Wellcome Trust Case Control Consortium 2007).

Since almost the entire genome is analysed for the genetic associations of a particular disease, this technique allows the genetic factors of a disease to be investigated in a non-hypothesis driven manner and represents a promising way to study complex, common diseases in which many genetic variations contribute to disease risk. Despite the advantages of using GWA to study complex traits, this approach has a strong potential for false-positives due to the large number of permutations generated requiring stringent statistical criteria in analysis such as the Bonferroni correction (Miller 1981). Similarly for all association studies, GWA is subject to influences from factors such as population stratification and phenotype misclassification between cases and controls (Baron 2001). Large sample sizes are required to detect genetic variants of modest effects with power calculations indicating that 2,000 individuals are needed to detect an OR ≤ 2 (Risch and Merikangas 1996). In addition, genome coverage from SNP chips used in GWA has recently been demonstrated to be limited at about 60% rather than previously reported levels of 80% to 90%, particularly for low-frequency and rare variants (Hao *et al.* 2008).

1.3 Genetics of Allergic Disease

1.3.1 Genetic Studies of Allergic Disease

Different approaches have been used to study genetic determinants in the development of allergic disease with increasing numbers of candidate loci and genes being implicated for disease susceptibility. Genetic studies indicate that many genes may be involved in the pathogenesis of allergic disease in addition to other disease specific genes which regulate the development of different types of allergic conditions (Barnes 2000). For example in atopy development, the production of total IgE and specific IgE have been shown to have influences from different genetic loci in additional to common loci controlling both traits (Dold *et al.* 1992; Marsh 1996). However, the majority of studies have been limited in both statistical power and experimental design resulting in problems encountered such as type I error. This has in turn led to difficulty in reproducing findings across studies and the inability to reliably confirm levels of contribution from multiple genetic factors (discussed in detail in Section 1.4.1).

1.3.2 Genome–Wide Linkage (GWL) Studies

Several genome-wide screens have linked atopy and related atopic diseases to loci across the genome containing genes coding for proteins involved in host immune responses. The first locus linked to atopy defined by IgE responses was identified on chromosome 11q13 in a study of seven families to the marker D11S97 with a maximum LOD score of 5.58 (Cookson *et al.* 1989). Subsequently, a gene coding for the β subunit of Fc epsilon R1: membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide) (*MS4A2*) was identified in this region (Dizier *et al.* 1999). However, linkage of atopy to the 11q13 region has not been reproduced consistently across several studies using a variety of genetic markers and phenotypes in populations with differing ethnic backgrounds (Moffatt and Cookson 1997). Other linked loci that have been replicated for statistical significance in more than one genome-wide screen for asthma include regions 2q33, 5q23-31, 6p24-21, 11q21-13, 12q24-12, and 13q14-12 (Wills-Karp *et al.* 2004).

Genome-wide searches for allergic rhinitis have been carried out in six populations identifying linkage regions on chromosomes 3q and 4q as common to atopy (serum total and specific IgE), atopic asthma, allergic rhinitis, and atopic dermatitis (Haagerup *et al*. 2004). Other regions may contain trait-specific genetic factors such as 3q13.31 for allergic rhinitis. A study testing dense markers in this region using a total of 236 allergic Danish families found that the signal intensified using 125 sib-pairs with rhinitis alone (LOD 5.55 compared to 3.71 for increased specific IgE) (Brasch-Andersen *et al*. 2006). To date, GWL scans have been conducted for atopic dermatitis in four different populations. Regions on chromosomes 3p, 3q, 17q, and 18q have been linked in at least two independent studies (Hoffjan *et al*. 2005).

Linkage results for atopy and related traits with regions replicated in at least two studies are summarised in Table 1.3. The independent identification with which these regions have been detected indicates that they might indeed contain disease susceptibility genes.

Table 1.3 – References to studies showing loci replicated for linkage ≥2 independent studies.

LINKAGE				
Loci	**Atopy (total IgE, specific IgE, SPT)**	**Asthma**	**Allergic rhinitis**	**Atopic dermatitis**
2q12-q33			Dizier *et al.* 1999; Haagerup *et al.* 2001	
3p6-24				Bradley *et al.* 2002; Haagerup *et al.* 2004
3q13			Haagerup *et al.* 2001; Brasch-Andersen *et al.* 2006; Bu *et al.* 2006	
3q14-21				Lee *et al.* 2000; Bradley *et al.* 2002
5q23-q33	Marsh *et al.* 1994; Hizawa *et al.* 1995; Meyers *et al.* 1995; Noguchi *et al.* 1997; Hizawa *et al.* 1998; Palmer *et al.* 1998	Postma *et al.* 1995; CSGA 1997; Noguchi *et al.* 1997; Ober *et al.* 1998		
6p21.1-p23	Caraballo and Hernandez 1990; Daniels *et al.* 1996; Hizawa *et al.* 1998; Wjst *et al.* 1999	CSGA 1997; Wjst *et al.*1999		
6p22-24			Haagerup *et al.* 2001; Bu *et al.* 2006	
11q13	Cookson *et al.* 1989; Young *et al.* 1992; Shirakawa *et al.* 1994; Shirakawa *et al.* 1994; Hizawa *et al.* 1995; Daniels *et al.* 1996; Doull *et al.* 1996; Palmer *et al.* 1998	Van Herwerden *et al.* 1995; Daniels *et al.* 1996; Doull *et al.* 1996		
12q14-q24.33	Barnes *et al.* 1996; Nickel *et al.* 1997; Wjst *et al.* 1999	Barnes *et al.* 1996; Ottman 1996; Ober *et al.* 1998; Barnes *et al.* 1999; Wjst *et al.* 1999		
13q14.3-qter	Daniels *et al.* 1996; Hizawa *et al.* 1998	CSGA 1997; Kimura *et al.* 1999		
14q11.2-q13	Moffatt *et al.* 1994; Mansur *et al.* 1999			
17q21- 25				Cookson et al. 2001; Bradley *et al.* 2002
18q11-21				Bradley *et al.* 2002; Haagerup *et al.* 2004
19q13		CSGA 1997; Kimura *et al.*1999		
21q21		CSGA 1997; Kimura *et al.*1999		

Positional cloned genes for asthma and related traits are listed in Table 1.4, six genes have been identified and positively replicated using this approach for asthma including at the level of a SNP. Currently, no genes have been identified veritably for both allergic rhinitis and atopic dermatitis by positional cloning. Studies identifying linkage to regions containing susceptibility genes for another inflammatory skin condition, psoriasis including to chromosome 17q containing the psoriasis susceptibility 2 (*PSORS2*) and psoriasis susceptibility 5 (*PSORS5*) genes suggest an important role for dermal expressed genes in atopic dermatitis in addition to general atopic genes (Cookson *et al.* 2001).

Table 1.4 – Genes identified by positional cloning in determining susceptibility to atopy and related traits. * Genes replicated in ≥2 independent studies.

POSITIONAL CLONING					
Loci	**Genes**	**Phenotype**	**Putative function**	**1st positional cloning study**	
				Population	**Reference**
2q14	*DPP10*	Asthma* IgE	Controls bronchial reactivity via voltage-gated potassium channels	239 UK and Australian families	Allen *et al.* 2003
6p21	*HLA-G*	Asthma* Allergy	Alters the Th1/Th2 cell balance	129 US Caucasian families	Nicolae *et al.* 2005
7p14	*NPSR1*	Asthma* IgE Atopy	Controls smooth muscle contractibility	1,972 German children	Kormann *et al.* 2005
12q24	*SFSWAP*	Asthma*	Involved in T cell activation via CD45 antigen	236 Danish families	Brasch-Andersen *et al.* 2006
13q14	*PHF11*	Asthma* IgE Atopic dermatitis	Promotes Th1-type cytokine gene expression through interaction with NFkB protein	202 Australian families	Ferreira *et al.* 2005
20p13	*ADAM33*	Asthma* IgE	Involvement in epithelial repair	460 US Caucasian families	Van Eerdewegh *et al.* 2002

DPP10: dipeptidyl-peptidase 10 (non-functional); *HLA-G*: major histocompatibility complex, class I, G; *NPSR1*: neuropeptide S receptor 1; *SFSWAP*: splicing factor, suppressor of white-apricot homolog (Drosophila); *PHF11*: PHD finger protein 11; *ADAM33*: ADAM metallopeptidase domain 33.

More than 20 distinct chromosomal regions have been identified for linkage with atopy and related phenotypes such as serum total IgE, specific IgE, SPT, and asthma in more than 17 distinct populations (Ferreira *et al.* 2005). A large amount of heterogeneity remains with few replicated findings and reported linkages meeting the accepted criteria for significant genome-wide linkage i.e. a LOD score of 3.6. The results of these linkage analyses suggest that multiple genomic regions contain genes influencing atopic phenotypes [Figure 1.5].

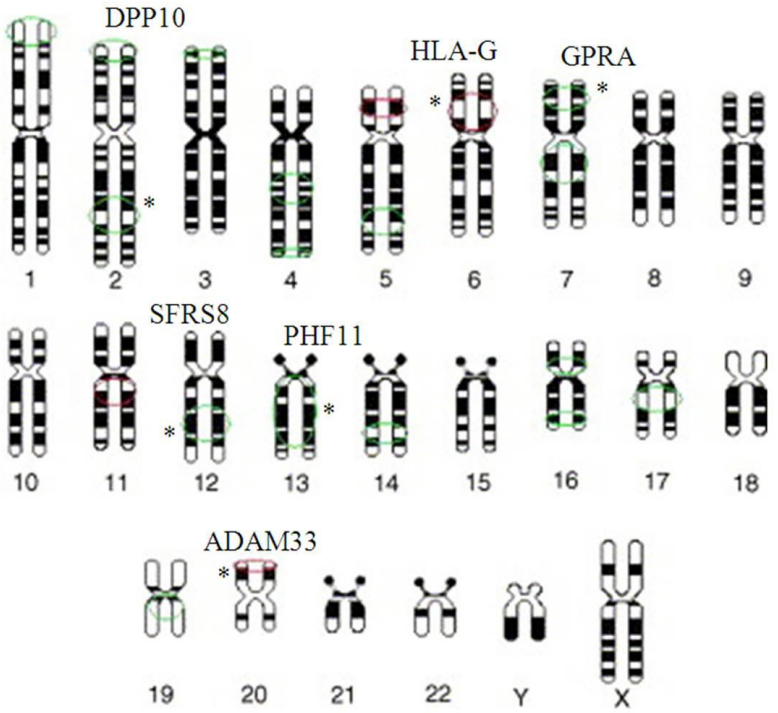


Figure 1.5 – Results of 16 genome-wide scans for asthma or asthma-related intermediate phenotypes. Specific chromosomal regions have been highlighted for importance (green) and particular importance (red) based on consistency of results. Locations of key positionally cloned genes studied detailed in Table 1.4 are shown. Adapted from (Carroll 2005).

1.3.3 Meta-Analysis of GWL Studies

Previous work performed in Nottingham combined data from 11 GWL studies for asthma and related traits in order to investigate the consistency of linkage results across studies in Caucasian populations using asthma enriched families (1,093 pedigrees, n=4,746) (Denham *et al.* 2008). The genome-wide scan meta-analysis (GSMA) study yielded supporting evidence for overlap of susceptibility loci for atopy, BHR, and total IgE including chromosome 5q23.2-q34 containing candidate genes *IL4*, *IL13*, *IL5*, *IL12B*, and the major histocompatibility complex (MHC) or human leukocyte antigen (HLA) locus on chromosome 6p22.3-p21.1 implicating potentially pleiotropic effects for these genes. Significant evidence for linkage was shown for chromosomal regions 3p22.1-q22.1 and 17p12-q24.3 as containing susceptibility loci for positive SPT (1,093 pedigrees, n=4,746). For asthma (1,267 pedigrees, n=5,832), no single chromosomal region at genome-wide statistical significance could be found to be clearly linked. This may have been due to the heterogeneity in disease diagnosis across the multiple studies used (Denham *et al.* 2008).

1.3.4 Candidate Gene Studies

Findings of candidate genes for atopy and related traits are summarised in Table 1.5. Most commonly, the candidate gene approach has been used to examine the cytokine gene cluster on chromosome 5q23-31 and HLA group of genes on chromosome 6p21-22. A review carried out in 2006 of nearly 500 disease association papers showed that 118 asthma candidate genes have been identified by candidate gene association, of which 79 have been associated with an asthma or atopy-related phenotype in two or more independent studies (Ober and Hoffjan 2006). As of 2011, there are now over 1,000 published studies examining association of SNPs in asthma and allergy candidate genes (Postma *et al.* 2011). Several of the most investigated genes include the *HLA-DR* genes, interleukin 4 receptor (*IL4R*), and interleukin 13 (*IL13*), which have been replicated over 10 times in various studies although this may have been subject to publication bias.

HLA-DR genes, for example, were chosen due to the involvement of these molecules in the immune response by binding to antigen derived peptides prior to T lymphocyte presentation. As a result, a number of association studies have been carried out looking at their potential roles in determining genetic susceptibility to atopy (Marsh *et al.* 1982; Aron *et al.* 1996; Nicolae *et al.* 2005). Although a number of association studies produced varying consistency of results, evidence for this region was supported by a linkage analysis study identifying a number of allele polymorphisms in *HLA-DR* genes associated with specific IgE responses to a number of allergens (Young *et al.* 1994; Holloway *et al.* 1996).

In some studies, candidate genes have been selected based on positional properties within loci identified for linkage. The gene encoding filaggrin, a protein involved in terminal differentiation of the epidermis and formation of the skin barrier was identified and replicated on chromosome 1q21, a region showing strong linkage to atopic dermatitis (Palmer *et al.* 2006; Sandilands *et al.* 2007). Following multiple replication studies, two losses of function genetic variants within this gene have shown significant association with the expression of atopic dermatitis (Weidinger *et al.* 2006).

Table 1.5 – Summary of candidate genes replicated ≥5 independent studies for atopy and asthma, ≥4 for atopic dermatitis (AD) and ≥2 for allergic rhinitis (AR).

CANDIDATE GENE				
Loci	**Genes**	**Phenotype**	**Putative function**	**Selected references**
1q21	*Filaggrin*	AD	Involvement in protective dermal barrier function by encoding the Filaggrin protein which bind to keratin fibers in epithelial cells.	Seguchi *et al.* 1996; Palmer *et al.* 2006; Weidinger *et al.* 2006; Sandilands *et al.* 2007
5q31	*IL4, IL13, CD14*	Atopy	Th2 cytokines: inducing differentiation of naive Th cells into Th2 cells (IL4); IgE isotype switching following release from Th2 cells (IL13). Pattern recognition receptor in the innate immune response expressed predominantly by macrophages (CD14).	Baldini *et al.* 1999; Gao *et al.* 1999; Ober *et al.* 2000; Cardaba *et al.* 2001; Koppelman *et al.* 2001; Bourgain *et al.* 2003
		Asthma		Ober *et al.* 2000; Takabayashi *et al.* 2000; Zhu *et al.* 2000; Noguchi *et al.* 2001
		AD		Heinzmann *et al.* 2000; Liu *et al.* 2000; Tsunemi *et al.* 2002; Hummelshoj *et al.* 2003
5q31-33	*SPINK5*	AD	Role in skin and hair morphogenesis, anti-inflammatory and/or antimicrobial protection of mucous epithelia. Mutations in the SPINK5 protein may result in Netherton's syndrome (elevated IgE, with atopic manifestations).	Walley *et al.* 2001; Tsunemi *et al.* 2002; Kato *et al.* 2003; Nishio *et al.* 2003; Kabesch *et al.* 2004; Kusunoki *et al.* 2005
5q32-34	*ADRB2*	Asthma	Targeted site for β_2-agonists as a membrane-bound G-protein-coupled receptor involved in smooth muscle relaxation and vasodilatation.	Reihsaus *et al.* 1993; Turki *et al.* 1995; Hopes *et al.* 1998; Ober *et al.* 2000; Hakonarson *et al.* 2001; Bourgain *et al.* 2003
6p21	*TNF*	Asthma	Mediates the inflammatory response in the acute phase response by promoting neutrophil migration through expression of endothelial adhesion molecules (TNFα); vascular endothelial and phagocytic cell binding (LTA).	Moffatt and Cookson 1997; Tan *et al.* 1997; Chagani *et al.* 1999; Moffatt *et al.* 1999; Trabetti *et al.* 1999; Lin *et al.* 2002; Witte *et al.* 2002; Di Somma *et al.* 2003
	LTA	Asthma		Moffatt and Cookson 1997; Tan *et al.* 1997; Moffatt *et al.* 1999; Cardaba *et al.* 2001; Immervoll *et al.* 2001; Lin *et al.* 2002; Witte *et al.* 2002
6p21.3	*HLA-DR* genes	Atopy	Antigen presentation to T lymphocytpes to stimulate Th cell production.	Marsh *et al.* 1982; Freidhoff *et al.* 1986; Marsh *et al.* 1987; Ansari *et al.* 1989; Young *et al.* 1994
11q13	*MS4A2*	Atopy	Expressed on antigen-presenting cells to form antigen-IgE complexes leading to degranulation of basophils and mast cells to release inflammatory mediators.	Shirakawa *et al.* 1994; Hill and Cookson 1996; Nagata *et al.* 2001
		Asthma		Li and Hopkin 1997; Cox *et al.* 1998; Green *et al.* 1998; Hijazi *et al.* 1998; Palmer *et al.* 1999; van Hage-Hamsten *et al.* 2002
		AD		Cox *et al.* 1998
11q22	*IL18*	AR	Pro-inflammatory cytokine working in combination with IL12.	Kruse *et al.* 2003; Lee *et al.* 2006; Sebelova *et al.* 2007
14q11	*CMA1*	AD	Mast cell serine protease conferring pro-inflammatory effects.	Mao *et al.* 1996; Mao *et al.* 1998; Soderhall *et al.* 2001; Weidinger *et al.* 2005
16p12	*IL4R*	AD	Encodes an alpha-chain for both the IL4 and IL13 cytokine receptors.	Hershey *et al.* 1997; Oiso *et al.* 2000; Callard *et al.* 2002; Novak *et al.* 2002; Hosomi *et al.* 2004
17q23	*EPO*	AR	Eosinophil activation in the Th2 response.	Nakamura *et al.* 2003; Nakamura *et al.* 2004

IL4: interleukin 4; *IL13*: interleukin 13; *CD14*: CD14 molecule; *SPINK5*: serine peptidase inhibitor, Kazal type 5; *ADRB2*: adrenergic, beta-2-, receptor, surface; *TNF*: tumor necrosis factor; *LTA*: lymphotoxin alpha (TNF superfamily, member 1); *MS4A2*: membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide); *IL18*: interleukin 18; *CMA1*: chymase 1, mast cell; *IL4R*: interleukin 4 receptor; *EPO*: erythropoietin.

1.3.5 GWA Studies

Rapid advances in genotyping technologies have resulted in increasing numbers of GWA studies in allergic disease identifying for example in asthma, a widely replicated locus on chromosome 17q12-21 containing genes ORM1-like protein 3 (*ORMDL3*), chemokine (C-C motif) ligand 11 (*CCL11*) and gasdermin B (*GSDML*) as well as additional genes including chitinase 3-like 1 (cartilage glycoprotein-39) (CHI3L1), interleukin 1 receptor-like 1 (*IL1RL1*), and WD repeat domain 36 (*WDR36*) on chromosomes 1q31, 2q12, and 5q22 respectively [Table 1.6] (Moffatt *et al.* 2007; Ober *et al.* 2008; Gudbjartsson *et al.* 2009). Recently, the biggest collaborative effort thus far investigating the genetic determinants of asthma was published by the Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community (GABRIEL) Consortium. This study consisted of 10,365 case subjects and 16,110 controls and observed genome-wide significance between asthma and SNPs within previously reported loci and genes including interleukin 18 receptor 1 (*IL18R1*) ($p=3x10^{(-9)}$), the *HLA-DQ* group of genes ($p=7x10^{(-14)}$), and interleukin 33 (*IL33*) ($p=9x10^{(-10)}$) as well as the chromosome 17q12-21 locus ($p=6x10^{(-23)}$) specific to childhood-onset disease (Moffatt *et al.* 2010).

Importantly, association studies to date have mostly focused on mild-to-moderate disease and genetic risk factors for severe asthma remain unclear. In 2009, a relatively small GWA study was conducted on a population of patients with severe or difficult to treat asthma (473 cases, 1,892 controls) carried out by The Epidemiology and Natural History of Asthma: Outcomes and Treatment Regimes (TENOR) study (Li *et al.* 2010). This study identified association with multiple SNPs in the DNA repair protein RAD50-interleukin 13 (*RAD50-IL13*) ($p=3x10^{(-7)}$) and *HLA-DR/DQ* ($p=9x10^{(-6)}$) regions although no loci met conventional genome-wide significance criteria ($p<5x10^{(-8)}$) (McCarthy *et al.* 2008).

Additional potential disease susceptibility genes for atopy, asthma, and atopic dermatitis identified by GWA studies are listed in Table 1.6. Many of these findings have been supported by functional analyses. No study has yet been conducted for allergic rhinitis using this approach.

Table 1.6 – Susceptibility genes identified using GWAS.

GWA STUDIES								
Loci	**Genes**	**Phenotype**	**Putative function**	**Population**	**Genotyping platform**	**p value**	**Replication**	**Reference**
1p31	*DENND1B*	Asthma (children)	Cytokine release and regulation	US Caucasian (793 cases, 1,988 controls)	Illumina 300K	$9x10^{(-9)}$	Yes	Sleiman *et al.* 2010
1q23	*FceR1α*	Atopy (total IgE)	Codes for the α chain of high affinity IgE receptor	European (1,530 individuals)	Affymetrix 500K	$2×10^{(-20)}$	Yes	Weidinger *et al* 2008
2q12	*IL18R1*	Asthma	Promotes Th1 response	European (10,365 cases, 16,110 controls)	Illumina 610K	$3x10^{(-9)}$	No	Moffatt *et al.* 2010
2q12.3-14.2	*DPP10*	Asthma	Controls bronchial reactivity	African-American (464 cases, 471 controls)	Illumina 650K	$<10^{(-5)}$ in all subjects	Yes	Mathias *et al.* 2010
5q12.1	*PDE4D*	Asthma	Modulates airway relaxation	US Caucasian (259 cases, 846 controls)	Illumina 650K	$9x10^{(-7)}$	Yes	Himes *et al.* 2009
5q31.1	*RAD50*	Asthma (severe)	Role in apoptosis	US Caucasian (453 cases, 1,892 controls	Illumina 500K	$3x10^{(-7)}$	No	Li *et al.* 2010
5q33	*ADRB1*	Asthma	Codes for a α1-adrenergic receptor subtype	African-American (464 cases, 471 controls)	Illumina 650K	$<10^{(-5)}$ in all subjects	Yes	Mathias *et al.* 2010
6p21.3	*HLA-DR/DQ* genes	Asthma (severe)	Antigen presentation	US Caucasian (453 cases, 1,892 controls	Illumina 300K	$1x10^{(-5)}$	No	Li *et al.* 2010
		Asthma		European (10,365 cases, 16,110 controls)	Illumina 610K	$7x10^{(-14)}$	No	Moffatt *et al.* 2010
9q21.31	*TLE4*	Asthma (childhood-onset)	Promote inflammatory processes	Mexican (482 trios)	Illumina 550K	$7x10^{(-4)}$	Yes	Hancock *et al.* 2009
9q24	*IL33*	Asthma	Drives production of Th2-associated cytokines	European (10,365 cases, 16,110 controls)	Illumina 610k	$9x10^{(-10)}$	No	Moffatt *et al.* 2010
10q21.3	*CTNNA3*	Asthma (Occupational TDI)	Recruitment of E-cadherin and formation of cell-cell adhesion	Korean (84 cases, 265 controls)	Affymetrix 500K	$6x10^{(-6)}$	No	Kim *et al.* 2009
11q13	*LRRC32, C11orf30*	Atopic dermatitis	Unknown	European (939 cases, 975 controls)	Affymetrix 500K	$7.64x10^{(-10)}$ in all subjects	Yes	Esparza-Gordillo 2009
15q22.33	*SMAD3*	Asthma	Transcriptional modulator of regulatory T cells	European (10,365 cases, 16,110 controls)	Illumina 610K	$4x10^{(-9)}$	No	Moffatt *et al.* 2010
17q12-21	*GSDMB /GSDMA*	Asthma (childhood-onset)	Unknown	European (10,365 cases, 16,110 controls)	Illumina 610K	$6x10^{(-23)}$	No	Moffatt *et al.* 2010
17q12-21	*ORMDL3*	Asthma (childhood-onset)	Unknown	European (944 cases, 1,243 controls)	Illumina 300K	$9x10^{(-11)}$	Yes	Moffatt *et al.* 2010
17q24.3	*GNA13*	Asthma	G-protein coupled receptor molecule	African-American (464 cases, 471 controls)	Illumina 650K	$<10^{(-5)}$ in all subjects	Yes	Mathias *et al.* 2010
20pter-p12	*PRNP*	Asthma	Immune cell activation, signal transduction, cell adhesion, antioxidant activity	African-American (464 cases, 471 controls)	Illumina 650K	$<10^{(-5)}$ in all subjects	Yes	Mathias *et al.* 2010
22q13	*IL2RB*	Asthma	Regulate homeostasis of effector T-cell subgroups	European (10,365 cases, 16,110 controls)	Illumina 610K	$1x10^{(-8)}$	No	Moffatt *et al.* 2010

DENND1B: DENN/MADD domain containing 1B; *IL18R1*: interleukin 18 receptor 1; *DPP10*: dipeptidyl-peptidase 10 (non-functional); *PDE4D*: phosphodiesterase 4D, cAMP-specific; *RAD50*: DNA repair protein RAD50; *ADRB1*: adrenergic, beta-1-, receptor; *TLE4*: transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila); *IL33*: interleukin 33; *CTNNA3*: catenin (cadherin-associated protein), alpha 3; *LRRC32*: leucine rich repeat containing 32; *C11orf30*: chromosome 11 open reading frame 30; *SMAD3*: SMAD family member 3; *GSDMB/A*: gasdermin B/A; *ORMDL3*: ORM1-like protein 3; *GNA13*: guanine nucleotide binding protein (G protein); *PRNP*: prion protein; *IL2RB*: interleukin 2 receptor, beta.

1.3.6 Gene-Gene Interactions

Genes identified using the above methods have tended to be analysed individually. Interactions of multiple genes and pathways are likely to be involved in the genetic basis of complex disease highlighting the need to examine the effects of gene-gene interaction or epistasis in determining the genetic susceptibility of atopy and related traits (Martinez 2007). For asthma, carrying the risk genotypes for both *IL4R* and *IL13* has been shown to increase risk of developing disease five-fold relative to carrying risk genotypes for only one or equally none of these genes (Howard *et al.* 2002). Although not extensively investigated, the importance of epistasis for other atopic traits has also been demonstrated. For example, interaction between genes coding for *IL4R* and the thromboxane A2 receptor (*TBXA2R*) on determining serum IgE has been reported (Tanaka *et al.* 2002). To date, few studies have adequately addressed epistatic gene-gene interactions in allergic disease.

1.3.7 Gene-Environment Interactions

Similarly, gene-environment interactions remain to be determined. As the environmental components in these complex diseases are diverse, the interaction between gene expression and environmental influences need also to be taken into consideration (Martinez 2005). Environmental factors such as exposure to house dust mite, viral infection, and cigarette smoke may influence sensitisation following allergen exposure through alteration of normal homeostatic defence mechanisms by increasing allergen access via epithelial permeability (D'Amato *et al.* 2005).

A number of studies have shown differential phenotypic effects of the same genetic polymorphisms as a result of environmental factors, in particular, between variants in *CD14*, toll-like receptor 2 (*TLR2*), and chromosome 5q with farming exposure (Eder *et al.* 2004; Eder *et al.* 2005; Bieli *et al.* 2007; Leynaert *et al.* 2008), pet microbes (Williams *et al.* 2008), household dust mite (Zambelli-Weiner *et al.* 2005), and tobacco smoke (Meyers *et al.* 2005) respectively. Linkage of hyper-responsiveness to the 5q locus was stronger in families with environmental tobacco smoke exposure during pregnancy or early childhood (Meyers *et al.* 2005). The risk allele for the development of asthma and elevated IgE of the *CD14* -159 SNP,

rs2569190 has been shown to differ in communities with varying endotoxin exposures (Simpson *et al.* 2006). In a number of studies, homozygotes for the T allele polymorphisms in *CD14* have increased likelihood of developing asthma and elevated IgE when exposed to high levels of endotoxin but the same genotype was shown to be protective against atopy and asthma when exposures were low (Eder *et al.* 2005; Zambelli-Weiner *et al.* 2005). Interaction of *ORMDL3* variants with smoke exposure in asthma has also been shown in both early and late-onset individuals (Flory *et al.* 2009). The differential association of flexural atopic dermatitis with *IL4R* dependent upon previous history of bacterial infection has also been reported (Gern *et al.* 2004).

1.4 The 'Missing Heritability' Component

1.4.1 Deficiencies in Previous Studies

Current GWA studies are only sufficiently powered to detect common casual variants. Assuming that some identified polymorphisms are truly associated with disease, lack of replicated and inconsistent findings may have resulted from a number of different factors. The main difficulties arise from phenotype definition and population ascertainment (Holloway *et al.* 1999). Studies have been carried out using a range of population cohorts with varying ancestry and differences in defining affected individuals with disease. This not only has detrimental impact on the ability to detect common variants but also reduces accuracy in estimating disease heritability. Identified markers linked to or associated with disease may have variable linkage correlation across populations reducing the effectiveness in tagging candidate genes or other disease causative polymorphisms. There may be individual and population specific genetic effects which may be independent of other genetic variants and environmental exposures, or result from epistasis and gene-environment interactions.

The proportions of phenotypic variation attributed to additive genetic factors, described by heritability estimates, have tended to be calculated from family studies (Visscher *et al.* 2008). Although theoretical models and data have proposed that the genetic effects are most likely to be additive in complex diseases, the accuracy of subsequently estimations has been challenged by not accounting for additional non-additive genetic factors, shared environmental exposures as well as interactions (Stunkard *et al.* 1990; Hill *et al.* 2008; Manolio *et al.* 2009). Furthermore, larger sample sizes and improved phenotyping are required to yield adequate statistical power and reduce error in detecting polymorphisms with lower allele frequencies and modest genetic effects.

Given the low effect sizes seen (typically with OR 1.2 to 1.5), even in results from large-scale studies such as that carried out by the GABRIEL Consortium, it is unlikely that identified common variants will explain all of the disease variability identified in allergic disease (Park *et al.* 2010). Currently, reported common variants have only been able to determine less than 10%

of this variability. The source of this 'missing heritability' has been widely debated (Maher 2008). Other forms of structural variation warranting further study include insertion/deletion polymorphisms (indels) and copy number variants (CNVs), for example, a deletion up to 50% of glutathione S-transferase mu 1 (*GSTM1*) in determining asthma with an OR of 1.59 to 1.89 has been identified (Ivaschenko *et al.* 2002; Brasch-Andersen *et al.* 2004). A deletion of two genes in the late cornified envelope (*LCE*) gene cluster, *LCE3C* and *LCE3B* has also been shown to confer susceptibility to psoriasis and has recently been replicated (de Cid *et al.* 2009; Hüffmeier *et al.* 2010).

Recent interest in the contribution of rare polymorphisms has focused on identifying variants not previously studied using GWA approaches: low-frequency SNPs with MAF of 0.3% to 5% and rare SNPs with MAFs of less than 0.3%. The common disease, rare variant hypothesis is based on the assumption that low frequency causative SNPs are likely to be non-synonymous and increase disease risk by having a deleterious effect on the protein function (Goldstein 2009). The effects of these SNPs may not be strong enough to be selected out of the population and may persist due to genetic drift or founder effects. Re-sequencing of two genes, potassium inwardly-rectifying channel, subfamily J, member 11 (*KCNJ110*) and hematopoietically expressed homeobox (*HHEX*) associated with diabetes in 13,715 individuals discovered an abundance of rare variants (Coventry *et al.* 2010). This study suggested that short, recent gene genealogical branches may have arisen as consequence of recent accelerating population growth. Therefore, selection would only have had sufficient time to influence the frequencies of only the most severely harmful variants.

Low-frequency and rare SNPs are thought to occur frequently in the genome with approximately 138,700 validated SNPs currently deposited online (dbSNP 134), most recently by the 1000 Genomes Project (Gorlov *et al.* 2008; 1000 Genomes Project Consortium *et al.* 2010). Rare variants are likely to have larger effect sizes with an OR between 2 and 5 filling the effect gap of detected heritable variants between linkage and association studies (Bodmer and Bonilla 2008). Recent evidence suggests that rare variants may be responsible for previously seen association signals attributed to common SNPs through synthetic association

demonstrating that 30% of synthetic associations were caused by the presence of rare causative variants and furthermore, the strength of the association with common variants increased with increasing numbers of rare causative variants (Dickson *et al.* 2010). However, there is evidence to challenge this hypothesis demonstrated by results from pathway analyses implying functional roles of GWA identified variants. Furthermore, less overlap between GWA signals and well-powered linkage studies, in addition to lower frequencies of rare variants in re-sequenced regions identified through GWA has been shown, than would be expected if synthetic associations were widespread (Anderson *et al.* 2011).

1.4.2 Next-Generation Sequencing (NGS)

Over the past few years, developments in next-generation sequencing (NGS) platforms have enabled the sequencing of several individual genomes (Venter *et al.* 2001; Levy *et al.* 2007; Bentley *et al.* 2008; Wang *et al.* 2008; Wheeler *et al.* 2008). These findings have emphasised the highly polymorphic nature of the human genome and the fact that compared to common variants, the catalogue of rare variant information remains highly incomplete. Sequencing of large numbers of individuals is required to identify rare polymorphisms across the genome. However, this is currently unfeasible in hundreds of individuals. Recent advances in NGS have made re-sequencing candidate genes to a greater depth of coverage more affordable to identify rare variants within hypothesised or previously identified loci (Stratton 2008). Methods to multiplex larger sample sizes have also been incorporated into NGS platforms and can now be carried out requiring fewer runs to sequence mupltiple individuals in pools (Dahl *et al.* 1998; Craig *et al.* 2008; Harismendy and Frazer 2009). Use of this approach has generated support for the contribution of rare variants in common, complex diseases and phenotypes such as colorectal adenomas, low plasma levels of high density lipoprotein (HDL) cholesterol hypertriglyceridemia, and blood pressure (Cohen *et al.* 2004; Fearnhead *et al.* 2004; Ji *et al.* 2008). Re-sequencing technology has been used in combination with GWA giving evidence of overlap in rare and common variant signals, patients with high plasma triglyceride levels were found to carry twice as many rare, coding genetic variants within four candidate genes identified through GWA (Johansen *et al.* 2010).

A number of different strategies in identifying rare variants have been proposed varying in the extent of the genome covered, whether to target loci identified through GWA, the whole exome or the whole genome (Shendure and Ji 2008). Selecting populations for sequencing has also been debated, whether or not to focus on extreme phenotypes and to specifically incorporate family based designs. The current challenges faced by NGS approaches include both technical limitations such as low accuracy genotyping in regions with larger numbers of rare variants and in interpretation such as whether effects detected are disease causative or due to population stratification or cryptic relatedness within study samples (Day-Williams and Zeggini 2011).

The 1000 Genomes Project has undertaken a combination of low-coverage whole-genome sequencing of 179 individuals from four populations with an average of 3.7x coverage, high-coverage sequencing of two parent-child trios and exon-targeted sequencing of 697 individuals from seven populations (1000 Genomes Project Consortium *et al.* 2010). Data released from the August 2010 pilot has generated a catalogue of 8 million previously unknown SNPs and around one million structural variants due to small indels. This project has shown the feasibility of the use of pooled DNA in multiplexing for the identification of rare variants in candidate genes and provided further support for the common disease, rare variant hypothesis. Further efforts in high coverage sequencing are currently been carried out across the scientific community to catalogue reference databases of low-frequency and rare variants including the UK10K Project (http://www.uk10k.org/) aimed at sequencing the exomes of 4,000 individuals to 6x sequencing depth.

1.5 Summary

Numerous chromosomal regions such as 2q, 5q, and 11q have shown strong evidence of linkage to atopy and allergic diseases including asthma, allergic rhinitis, and atopic dermatitis and have been predicted to contain disease susceptibility genes. The use of positional cloning has identified novel susceptibility genes including *DPP10*, *HLA-DR/DQ* genes, and *ADAM33* for asthma. Polymorphisms within several candidate genes including *IL4*, *IL4R*, *IL13*, *HLA-DRB1*, and *MS4A2* have shown reproducible association with atopy, asthma, and related phenotypes. GWA approaches have also been used, identifying confident association with genes such as *ORMDL3* with childhood-onset asthma and *FCER1A* with serum total IgE.

Due to the heterogeneity of these traits, disease expression and severity are influenced by numerous genetic pathways interacting with environmental components such as epistasis between *IL4R* and *TBXA2R* in serum IgE as well as influences of house dust mite endotoxin on *CD14* polymorphisms in asthma and elevated IgE. The total numbers of genetic loci implicated are extensive and the proportion of effect from individual genes remains unclear. Furthermore, it seems unlikely that all of the genes underlying these complex disorders have been identified.

Lack of replicated results and inconsistent findings result from a number of different factors in study design, implementation, and interpretation. Other sources of genetic variation may be responsible for the missing component in disease heritability including phenocopies, CNVs, and differential gene expression (Moffatt *et al.* 2007; Emilsson *et al.* 2008). Furthermore, the interaction between these factors and with the environment needs also to be addressed, for example, the importance of epigenetic causes such as DNA methylation. Genetic variability resulting from CpG islands, for example, has been described in influencing *ADAM33* expression in epithelial cells of asthma patients (Yang *et al.* 1999).

As focus on treatments for allergic diseases turn towards total prevention and better therapeutic interventions, information about specific genetic susceptibility in individuals will become increasingly valuable. Determining the genetic basis in the development and outcome of these

diseases could help to identify prior risk and prevent disease onset as well as predict variations in clinical responses to treatment. Genetic factors have also been shown to be responsible for up to 80% of the variability in treatment responses in Caucasian asthmatic patients (Drazen *et al.* 2000). Studies in genetics can also help to question the underlying biological mechanisms behind these conditions by identifying new pathways not previously indicated in disease pathogenesis, which can in turn form the basis for novel therapies.

Although genetic determinants of allergic disease have been widely studied, two areas which remain least understood are the specific IgE and severe asthma phenotypes. No study has yet used GWA approaches to identify genetic factors contributing to the development of atopy defined by elevated specific IgE and only a single small study has been published for severe asthma in 473 cases and 1,892 controls (Li *et al.* 2010).

1.6 Project Aims

The overall objective of this project was to obtain a greater insight into the genetic determinants of two of the least understood aspects of allergic disease. Two main phenotypes were investigated, firstly, type I hypersensitivity in response to common environmental allergens i.e. atopy defined by elevated specific IgE $\geq 0.35 KU_A/L$ or positive SPT, and secondly, severe asthma based on GINA criteria for poor control with steps 3 or above.

The aims of individual results chapters within this thesis were to:

Chapter 3: Conduct a GWA study to attempt to identify novel atopy susceptibility genes.

Chapter 4: Further define atopy susceptibility loci within chromosomal loci 3p22.1-q22.1 and 17p12-q24.3 showing significance linkage for SPT response identified by the GSMA.

Chapter 5: Use GWA to attempt to identify common variants determining susceptibility to severe asthma.

Chapter 6: Evaluate the contribution within severe asthma of previously identified candidate genes for mild-to-moderate disease.

Chapter 7: Characterise the 5' untranslated region (UTR), establish transcription start sites and identify putative promoter locations for the *IL1RL1* and *IL33* genes.

Chapter 8: Use re-sequencing to identify if the relative abundance of variation within the GWA identified loci on chromosome 2q12 and 9p24.1 determine susceptibility to severe asthma.

Chapter 2

Materials and Methods

2.1 Study Populations and Phenotyping

2.1.1 Atopy Analyses

Studies of atopy in Chapters 3 and 4 involved conducting a primary genome-wide association (GWA) study in the British 1958 Birth Cohort. Replication of loci identified for association was performed using both *in silico* replication (utilising directly genotyped or imputed SNPs) in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort and by prospective genotyping in the Nottingham Gedling and Southampton cohorts. In Chapter 4, *in silico* replication was also carried out in the deCODE Icelandic cohort. Demographic data for these population cohorts are summarised in Table 2.1 and key details summarised in the following text.

Table 2.1 Study populations used in atopy analyses.

Discovery									
Study cohort	**N total**	**N male**	**N female**	**Mean age, years (SD)**	**Median total IgE, kU_A/L (Range)**	**Atopy marker**	**N case**	**N control**	**SNP Array**
British 1958 Birth Cohort D1	1,428	Data unavailable				Elevated specific IgE $\geq 0.35 kU_A/L$	388	1,040	Illumina HumanHap500, Affymetrix Human Mapping 500K
British 1958 Birth Cohort D2	3,853	1,889	1,964	45.2 (0.4)	27 (0-1281)	Elevated specific IgE $\geq 0.35 kU_A/L$	1,083	2,770	Illumina HumanHap500
Replication *in silico*									
Study cohort	**N total**	**N male**	**N female**	**Mean age, years**	**Median total IgE, kU_A/L (Range)**	**Atopy marker**	**N case**	**N control**	**SNP Array**
ALSPAC	1,324	642	682	7.5	22 (0.7-99.8)	Positive SPT ≥2mm	282	1,042	Illumina HumanHap300
deCODE	28,058	Data unavailable				Positive SPT ≥3mm or ≥50% of histamine-positive control	908	27,150	Illumina HumanHap300
Replication genotyping									
Study cohort	**N total**	**N male**	**N female**	**Mean age, years, (SD)**	**Median total IgE, kU_A/L (Range)**	**Atopy marker**	**N case**	**N control**	**SNP Array**
Nottingham Gedling	1,268	635	633	56.2 (12.3)	22.9 (0-2000)	Positive SPT ≥3mm	394	831	KBiosciences
Southampton	682	339	343	42.3	9.7 (0-39893.5)	Elevated specific IgE $>0.35 kU_A/L$	295	381	KBiosciences
						Positive SPT >3mm	382	294	
Total n available for pooled analyses									
Phenotype	**N total**						**N case**	**N control**	
Elevated specific IgE	4,529						1,378	3,151	
Positive SPT	31,283						1,966	29,317	
Serum total IgE levels	7,276								

2.1.1.1 The British 1958 Birth Cohort

The British 1958 Birth Cohort is a longitudinal study of 18,558 individuals born during one week in March 1958 across the UK. Follow up at ages 7, 11, 16, 23, 33, and 42 years were conducted using parental interviews, school medical examinations, and individual interviews (Power *et al.* 2006). Protocols for the 2002 to 2004 biomedical examination were approved by the South East MultiCentre Research Ethics Committee. Written consent to participate in genetic association studies was given by all participants. DNA was obtained from 7,703 singletons of Caucasian ethnicity during 2002 to 2004. Phenotype data at age 44 to 45 years were collected for specific immunoglobulin E (IgE) levels in response to house dust mite, mixed grass or cat fur measured with Hytec enzyme immunoassay as well as serum total IgE levels (Hycor Biomedical, Irvine, CA, USA). Analyses included 1,083 cases defined by elevated specific IgE $\geq 0.35 kU_A/L$ to one or more allergens and 2,770 controls. Serum total IgE level ($\log_e$ transformed) analyses were carried out in 3,873 individuals.

2.1.1.2 The ALSPAC Cohort

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prospective birth cohort study of 14,000 infants born between 1st April 1991 and 31st December 1992 from mothers residing within the three Bristol-based health districts to study determinants of health, behaviour and development. Ethics approval was granted from the ALSPAC Law and Ethics Committee and Local Research Ethics Committees. Informed consent was obtained from all study participants (Roberts *et al.* 2005). DNA was extracted from approximately 70% of children using a number of commercial extraction methods. Skin prick testing at the age of 7.5 years was carried out using a panel of six allergens: house dust mite, mixed grass, cat fur, mixed nuts, peanut, and milk. Atopy was defined by a positive skin prick test (SPT) with a mean wheal diameter of ≥2mm to one or more of house dust mite, mixed grass, and cat fur. A total of 282 cases and 1,042 controls with genotype data were included in analyses. Serum IgE levels were measured by ELISA in 1,423 individuals.

2.1.1.3 The deCODE Icelandic Population

The deCODE Icelandic cohort was compiled by cross-matching patient records between 1977 to 2001 from the Department of Allergy and Pulmonary Medicine, National University Hospital of Iceland with a genealogy database of the entire Icelandic nation. Informed consent was obtained from all patients, and ethics approval granted by the Icelandic Data protection Commision and the National Bioethics Committee of Iceland (Hakonarson *et al.* 2001). A biobank of 100,000 genetic samples was created for participation in GWA studies of various disease projects at deCODE Genetics. Atopic individuals related to at least one other patient within and including six meiotic events were selected as cases. Atopy was defined by positive SPT measured by a wheal diameter ≥3mm or ≥50% of the histamine-positive control response to 12 aeroallergens: birch, grass, *Rumex crispus*, cat, dog, horse, Cladosporium, Mucor, Alternaria, house dust mite (Der p 1), house dust mite (Der f 1), and *Lepidoglyphus destructor* (Greer Laboratories, Lenoir, NC, USA) identifying 908 cases. A sub-group of 239 non-asthmatic, atopic case individuals were identified. The 27,150 general population controls used in this study consisted of individuals from other ongoing GWA studies at deCODE.

2.1.1.4 The Nottingham Gedling Cohort

The Nottingham Gedling cohort was collected in January 1991 from the general population, through a systematic sample of 2,633 adults (age 18 to 70 years) living in Nottingham, UK for participation in a cross-sectional study of the relationship between diet, asthma, and chronic obstructive pulmonary disease. Ethical approval was granted by the Nottingham City Hospital and Nottingham University Ethics Committee boards, and written consent was obtained from all subjects (McKeever *et al.* 2002). DNA was extracted using a Genomic DNA Maxi-Prep kit (Qiagen, Crawley, UK). Skin prick testing to the three common allergens: house dust mite, grass pollen, and cat fur was performed, with concomitant saline and histamine controls (Bencard, Brentford, UK). The response to each allergen was measured as the mean of two perpendicular diameters of the wheal, excluding flares and pseudopods. Analyses were carried out in 1,286 individuals for whom blood samples for DNA extraction were available. Atopy was defined as a mean wheal diameter ≥3mm than the saline control resulting in 385 cases and

822 controls. Serum total IgE levels were measured by enzyme linked immunosorbent assay (ELISA) in 1,293 individuals.

2.1.1.5 The Southampton Cohort

The Southampton cohort consists of a subset of non-related adult individuals (parents only) from the Southampton Asthma Families cohort. The Southampton Asthma Families cohort comprises 341 Caucasian families recruited from the Southampton area with at least two biological siblings (ages 5 to 21 years) with a current physician diagnosis of asthma taking regular asthma medication. Ethics approval was obtained from the Southampton and South West Hampshire joint ethics committee, written consent was obtained from all subjects (Sayers *et al.* 2003). DNA was extracted from whole blood using a Genomic DNA Maxi-Prep kit (Qiagen, Crawley, UK). Serum specific IgE level and total IgE level data were determined by radio allergosorbent test (RAST) to response to house dust mite, grass, cat, dog, *Alternaria*, and tree allergens. Skin prick testing was also completed for the same common allergens. Analyses were carried out using data from 757 non-related adult individuals (parents only) classified into 295 cases and 381 controls using atopy defined by elevated specific IgE levels >0.35kU_A/L or 382 cases and 294 controls using atopy defined by positive SPT with >3mm wheal diameter to one or more allergens. Serum total IgE levels were measured by ELISA in 675 individuals.

2.1.2 Asthma Analyses

Studies of severe asthma in Chapters 5, 6, and 8 involved conducting a primary GWA study as part of the Asthma UK Genetics of Severe Asthma (AUGOSA) study. In Chapter 5, subsequent *in silico* replication was carried out for SNPs of interest in the Australian Asthma Genetics Consortium (AAGC) study. Demographic data for these population cohorts are summarised in Table 2.2. A subset of individuals from the Nottingham Gedling cohort was also used for targeted re-sequencing in Chapter 8, described previously in Section 2.1.1.4.

Table 2.2 – Study populations used for severe asthma analyses. * Sex was unspecified for one individual within AUGOSA and six within GABRIEL Severe.

Discovery						
Study cohort		**N total**	**N male**	**N female**	**Mean age, years, (SD)**	**SNP Array**
Cases						
AUGOSA		682*	226	455	46 (13.9)	Illumina HumanHap610-Quad, Illumina HumanHap660w-Quad
GABRIEL Severe		344*	133	205	43 (17.8)	Illumina HumanHap660w-Quad
Controls						
Busselton		576	245	331	58 (15.7)	Illumina HumanHap550
WTCCC2		1,483	753	730	45 (0.4)	Illumina Human1M-duo v3
T1DGC		1,313	640	673	45 (0.3)	Illumina HumanHap660w-Quad
Total n available for analysis						
Cases		1,026				
Controls		3,372				
Replication *in silico*						
Study		**N total**	**N male**	**N female**	**Mean age, years, (SD)**	**SNP Array**
AAGC	Cases	231	90	141	51 (16.3)	Illumina HumanHap610-Quad
	Controls	1,345	530	815	32 (15.0)	Illumina HumanHap610-Quad

2.1.2.1 The Asthma UK Genetics of Severe Asthma (AUGOSA) Study

Severe Asthma Cases

The Asthma UK Genetics of Severe Asthma (AUGOSA) study consisted of 682 cases with severe asthma collected from eight centres across the UK by the AUGOSA Consortium and an additional 344 cases collected by the Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community (GABRIEL) Consortium.

Cases recruited by the AUGOSA Consortium were 682 individuals of European ancestry with severe asthma classified as steps 3 or above based on the Global Initiative for Asthma (GINA) criteria, recruited across eight UK-based centres: Nottingham QMC, Nottingham City Hospital, Belfast, Birmingham, Glasgow, Leicester, Manchester, and Southampton. Written consent was obtained and ethics approval granted by the Nottingham QMC Hospital, City Hospital and Nottingham University Ethics Committee boards. DNA extraction was carried out by phenol extraction at Centre National de Génotypage (Paris, France). Using GINA criteria, individuals were classified with step 3 severe asthma if they suffered from daily asthma attacks affecting activity or nocturnal symptoms more than once a week. Individuals were classified with step 4 severe asthma if they suffered from continuous symptoms limiting physical activity or frequent symptoms at night (Bateman *et al.* 2008). FEV_1 or PEF is 60% to 80% predicted for patients in step 3 and less than 60% in step 4, FEV_1 or PEF variability is greater than 30% for both groups. Patients in GINA step 5 were prescribed maintenance oral gluco-corticosteroids.

The GABRIEL Consortium is a collection of over 40,000 European cases and population-matched controls obtained from 23 different studies aimed to identify determinants of asthma. All participants or their parents provided signed informed consent to participate in the study, and ethics approval obtained by corresponding local committees (Moffatt *et al.* 2010). DNA extraction was carried out by each individual centre. Asthma was determined using physician diagnosis and sub-group analyses carried out for (i) age-of-onset at 16 years or older, (ii) cases with unknown age of onset, (iii) occupational asthma or (iv) severe asthma. Individuals with severe asthma were recruited from three specialist asthma clinics; adult and childhood clinics

based at the Royal Brompton Hospital, London and an adult clinic at the Glenfield Hospital, Leicester. A total of 344 subjects were included in analyses defined using combinations of ATS and ERS criteria, as well as specialised protocols involving parameters of airway inflammation, airway physiology, quality of life, and control of symptoms (Bacharier *et al.* 2008; Stanojevic *et al.* 2008). ATS definitions are based on reduced peak flows and responses to corticosteroids, chronic airflow limitation, and rapidly progressive loss of lung function (ATS 2000). ERS definitions are based on poor symptomatic control, episodic exacerbations, persistent and variable airways obstruction in addition to requirement for β_2-agonists and oral gluco-corticosteroids (Chung *et al.* 1999). For clarity, these subjects will be referred to as GABRIEL Severe cases in order to distinguish them from the complete case population used in the published GABRIEL study totalling 10,365 individuals (Moffatt *et al.* 2010).

General Population Controls

A total of 3,353 control subjects without history of asthma or wheeze were collected from the UK and Western Australia, all of whom were of European ancestry. This consisted of 564 individuals collected by the Busselton Health Study, 1,483 individuals collected by the Wellcome Trust Case-Control Consortium Phase 2 (WTCCC2), and 1,299 individuals collected by the Type 1 Diabetes Genetics Consortium (T1DGC).

The Busselton Health Study consists of residents from the town of Busselton in the south-west of Western Australia who have been involved in a series of cross-sectional, whole-population health surveys since 1966 for studies of common diseases, primarily cardiovascular and respiratory and their relationships to risk factors, lifestyle, and environmental variables. Informed written consent for participation and subsequent use of data was obtained from participants at the time of the survey. Ethics approval was obtained from the University of Western Australia Human Research Ethics Committee (James *et al.* 2010). DNA from each subject was extracted using phenol/chloroform extraction and ethanol precipitation. Controls were defined as those who have consistently answered ‘No’ to ‘Has your doctor ever told you that you had asthma?’ at all previous surveys that they have attended from 1996 to 1994, 576 non-asthmatic, non-wheeze individuals were included in analyses.

The WTCCC2 involved 15 WTCCC collaborative studies and 12 independent studies totalling approximately 120,000 samples established in April 2008 with the aim of performing GWA studies in 13 disease conditions. Both patients and family members gave written consent for the use of their DNA for research within the laboratory. Ethics approval was granted by the Multi-Centre Research Ethics Committee (MREC) (Wellcome Trust Case Control Consortium 2007). DNA was extracted from EDTA-venous blood samples using conventional salt lysis methodologies. Control subjects consisted of 6,000 common controls from WTCCC, an additional 3,000 from the British 1958 Birth Cohort, and 3,000 from the UK Blood Service Collection, of which, 1,483 non-asthmatic, non-wheeze control individuals were included in analyses.

The T1DGC is an international, multicenter program organised to promote research to identify genes and alleles determining risk of susceptibility to Type 1 Diabetes. Written informed consent was obtained for all individuals involved in the study and ethics approval granted by each institution's Internal Review Board (IRB) or Ethics Committee at each clinic (Barrett *et al.* 2009). DNA extraction was carried out using variations of the salting out and chloroform methods. Cases and trio families (one child with diabetes and both biological parents) were collected across multiple ethnicities. Controls were defined as individuals without type 1 diabetes and recruited from matched populations. A total of 1,313 non-asthmatic, non-wheeze control individuals of Caucasian ethnicity were identified for inclusion in analyses.

2.1.2.2 The Australian Asthma Genetics Consortium (AAGC) Study

The Australian Asthma Genetics Consortium (AAGC) was established in 2009 with the aim of identifying genetic causes underlying asthma through analysis of whole-genome genetic data from 7,197 unrelated individuals of European ancestry. Study subjects were recruited from eight independent cohorts by six participating institutions across Australia. Ethics approval and informed consent were obtained by individual centres. DNA extraction was carried out using variations of the salting out and chloroform methods (Ferreira *et al.* 2011). Case individuals with severe asthma were recruited by the Lung Institute of Western Australia (LIWA) and defined as those with a physician diagnosis of asthma receiving medication inclusive of inhaled

gluco-corticosteroids (ICS) ≥400μg in combination with a long-acting β_2-agonist. Non-asthmatic control subjects were recruited through the LIWA (n=35) and the Queensland Institute of Medical Research (QIMR) studies (n=1,310). In the replication study, a total of 231 cases and 1,345 controls were included in analyses.

2.2 Genotyping and Sequencing

2.2.1 Genome-Wide SNP Genotyping

Analysis of genotype data from individuals included in study cohorts were generated primarily using DNA microarray technology through genome-wide genotyping. Single nucleotide polymorphisms (SNPs) were tested across the whole genome from each DNA sample using fluorescence based allele calling and probes covalently bound to a chemical matrix (Gruber *et al.* 2002). Assays with varying density of SNP coverage used in analyses were designed by the Illumina (San Diego, USA) and Affymetrix (Santa Clara, USA) genotyping laboratories.

Illumina utilise single-base extension with labelled bases on 50-mer oligonucleotide probes (one per SNP) designed almost entirely on the International HapMap Project tagging SNPs on coded microscopic beads (BeadArrary). Affymetrix employ differential hybridisation with 25-mer oligonucleotide probes (4 to 5 replicates per allele) designed using 50% HapMap tagging SNPs in combination with unbiased SNPs covering the whole genome on spatially arranged solid glass or silicon surfaces (GeneChip). Features of genome-wide SNP platforms used in genotyping of cohorts used in this thesis are detailed in Table 2.3. Genotype calls were made using the Illuminus and Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) calling algorithms for the Illumina and Affymetrix platforms respectively.

Table 2.3 – SNP Genotyping Arrays designed by Illumina and Affymetrix used in analyses. CEU denotes European ancestry in Utah, MHC denotes major histocompatibility complex, CNVs denotes copy number variants.

Genotyping laboratory	Array name	Number of SNPs tested	Average coverage
Illumina	HumanHap300	>317,000 HapMap Phase 1 tag SNPs 7,300 non-synonymous SNPs Additional tag SNPs in the MHC region, regions with CNVs, mitochondrial SNPs	1 SNP per 9kb across the genome (CEU)
Illumina	HumanHap550	~550,000 HapMap release 20 tag SNPs 7,800 non-synonymous SNPs Additional tag SNPs in the MHC region, regions with CNVs, mitochondrial SNPs	1 SNP per 5.5kb across the genome (CEU)
Illumina	HumanHap610-Quad	620,901 SNPs HapMap release 23 tag SNPs 7,577 non-synonymous SNPs Additional tag SNPs in the MHC region, regions with CNVs, mitochondrial SNPs	1 SNP per 4.7kb across the genome
Illumina	HumanHap660w-Quad	657,366 SNPs HapMap release 24 tag SNPs 10,051 non-synonymous SNPs Additional tag SNPs in the MHC region, regions with CNVs, mitochondrial SNPs	1 SNP per 4.4kb across the genome (CEU)
Illumina	Human1M-duo v3	1,199,187 SNPs HapMap release 23 tag SNPs 21,877 non-synonymous SNPs Additional tag SNPs in the MHC region, regions with CNVs, mitochondrial SNPs	1 SNP per 2.4kb across the genome (CEU)
Affymetrix	Human Mapping 500K	~262,200 (Nsp I restriction enzyme) ~238,000 (Sty I restriction enzyme)	1 SNP per 5.8kb across the genome

2.2.1.1 Atopy Analyses

Genome-wide genotyping of the British 1958 Birth Cohort, the ALSPAC cohort, and the deCODE Icelandic population cohort used in atopy analyses were undertaken independently by collaborating study centres. Genotyping procedures and quality control measures carried out prior to association testing are summarised and described below.

The British 1958 Birth Cohort

Individuals from the British 1958 Birth Cohort were genotyped using the Illumina Infinium HumanHap550 SNP BeadChip. SNPs with minor allele frequency (MAF) <1% or $p<10^{(-7)}$ following tests for Hardy-Weinberg equilibrium (HWE) as well as individuals of non-European ancestry or displaying high levels of missing genotypes or heterozygosity were excluded from further analyses. A total of 459,334 autosomal SNPs passed all quality control measures. Analyses included 3,853 individuals with results of association testing with specific IgE data and 3,873 individuals with results of association testing with serum total IgE level data. In addition, the Affymetrix GeneChip Human Mapping 500K Array was also used to genotype a subset of 1,468 individuals with the same quality control measures applied as above.

The ALSPAC Cohort

Individuals from the ALSPAC Cohort were genotyped using the Illumina HumanHap300 SNP chip. SNPs with MAF <1%, >5% missing genotypes or failed an exact test of HWE with $p<5x10^{(-7)}$ were excluded from further analyses. After quality control, 315,807 autosomal SNPs were left in analyses. Individuals of non-European ancestry or who displayed high levels of missing genotypes or heterozygosity were also excluded. Analyses were carried out in 1,324 individuals with results of association testing with SPT data and 1,423 individuals with results of association testing with serum total IgE level data.

The deCODE Icelandic Population

Individuals from the deCODE Icelandic Population were genotyped using the Illumina HumanHap300 SNP chip. Exclusion of SNPs with MAF <1% or deviated from HWE with $p<10^{(-10)}$ and individuals with non-European ancestry or who displayed high levels of missing genotypes or heterozygosity was carried out. A total of 312,179 autosomal SNPs were tested following quality control. Analyses included 28,140 individuals with results of association testing using SPT data.

2.2.1.2 Asthma Analyses

Genotyping of severe asthma cases recruited by both the AUGOSA study and the GABRIEL Consortium were carried out by Centre National de Génotypage (Paris, France). Controls had previously been genotyped independently by the relevant centres. Genotype data for all individuals in the AUGOSA study were combined prior to applying quality control measures. This work was carried out as part of this thesis; details of methods used for quality control are described in detail in Chapter 5. Procedures for genome-wide genotyping undertaken by the AAGC study and quality control measures carried out prior to association testing are described below.

The AUGOSA Study

Case subjects recruited by the AUGOSA Consortium were genotyped using the Illumina HumanHap610-Quad SNP BeadChip for 121 individuals and the Illumina HumanHap660w SNP BeadChip for 561 individuals. All 344 subjects recruited by the GABRIEL Consortium were genotyped using the Illumina HumanHap660w SNP BeadChip. Controls subjects were genotyped using the Illumina HumanHap610-Quad SNP BeadChip for all 576 individuals from the Busselton Health Study, the Illumina Human1M-duo SNP Beadship for all 1,483 individuals from the WTCCC2, and the Illumina HumanHap550 SNP BeadChip for all 1,313 individuals from the T1DGC. This resulted in six groups of individuals for genotyping. Individuals with <90% of SNPs called were excluded and SNPs were excluded if they had low call rates (proportion of genotypes called <90%), were not in HWE ($p<10^{(-4)}$), had a low MAF <1% or with differential missingness between cases and controls ($p<10^{-6)}$). Any SNPs not

present in all genotyping platforms used were excluded resulting in a total of 490,303 autosomal SNPs. This led to 9,414 SNPs being excluded leaving 480,889 SNPs in the analysis carried out in 1,026 cases and 3,372 controls remaining after genotyping quality control. An additional 119 individuals were excluded following principal components analysis (PCA) for underlying population structure leaving 933 cases and 3,346 controls (see Chapter 5 for details).

The AGGC Study

Individuals from the AGGC study cohort were genotyped using the Illumina HumanHap610-Quad SNP BeadChip. Quality control measures were carried out to remove SNPs with MAF <1%, >5% missing genotypes or failed an exact test of HWE with $p<10^{(-6)}$. Replication analyses were carried out in 231 severe asthma cases and 1,345 controls.

2.2.2 Direct SNP Genotyping (Atopy Analyses)

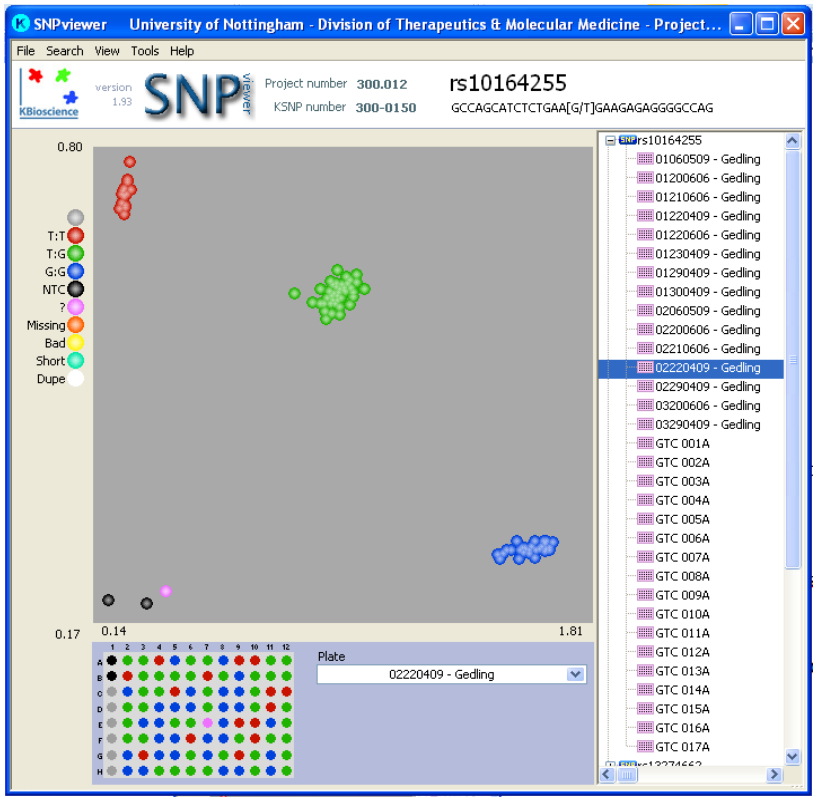
Prospective genotyping of selected SNPs in Chapters 3 and 4 atopy analyses was carried out in the Nottingham Gedling (1,286 individuals) and Southampton cohorts (757 individuals) by KBioscience (Hoddesdon, UK). The KBioscience Competitive Allele-Specific Polymerase Chain Reaction (KASP) genotyping system uses two allele-specific primers with interacting oligos, labelled at the 5' end with FAM or CAL Fluor Orange 560 for alternate alleles. Complementary oligos with 3' quenchers prevent fluorescence signal generation until the corresponding fluor-labelled oligo corresponding to the amplified allele gets incorporated into the template. This occurs during polymerase chain reaction (PCR) when the allele-specific primer becomes incorporated into the template.

Direct SNP Genotyping Quality Control

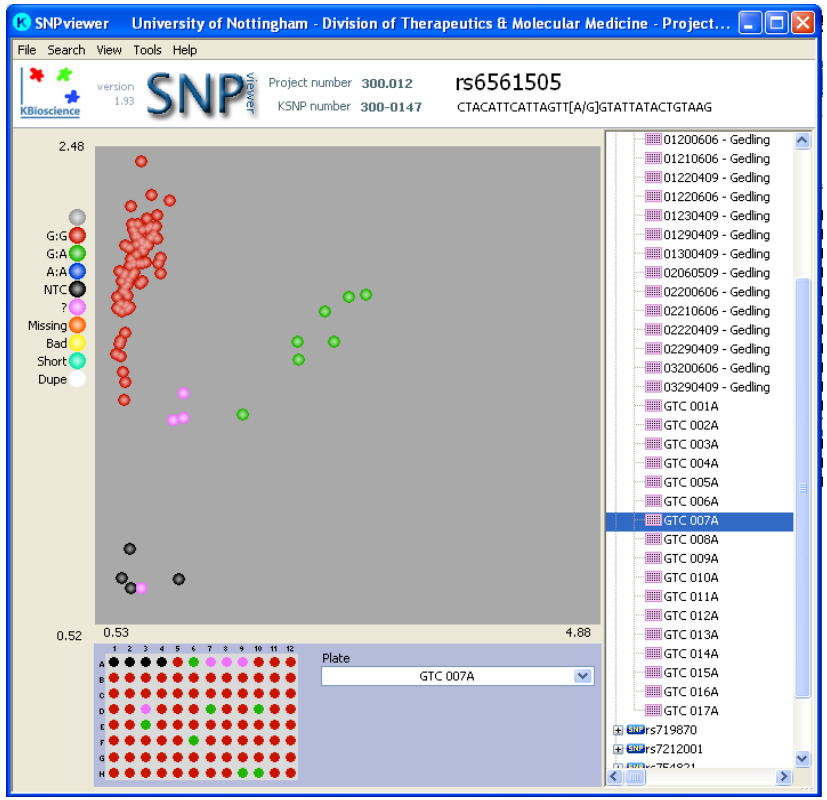
Genotyping quality control measures were carried out including positive and negative control samples. Individuals with <90% of SNPs called were excluded as well as SNPs based on low call rates (<90% genotypes called), low MAF (≤5%) or not in HWE ($p\leq0.05$). Genotype calling was carried out using software SNPviewer2 [Figure 2.1].

Figure 2.1 - Genotyping calling using SNPviewer2 for genotyped SNPs in atopy analyses. The x and y axis represent scales of the intensity of fluorescence detected. Example screenshots: A) rs10164255 showing clear clustering into TT (red), TG (green) and GG (blue). The non-template control (black) is also identified; one anomaly (pink) can be seen. B) rs6561505 show greater variation of data within each cluster and four anomalies. Only genotypes classed as homozygotes (red or green) and heterozygotes (blue) were included in analyses.

A



B



2.2.3 Targeted Re-sequencing (Asthma Analyses)

Two identified loci in severe asthma analyses were re-sequenced in a subset of cases from the AUGOSA study (200 individuals) and a subset of controls from the Nottingham Gedling cohort (200 non-asthmatic, non-wheeze, non-atopic individuals) by SourceBioscience (Nottingham, UK). Individual DNA samples were sheared and target sequence enrichment carried out using 120bp oligos. Cases and controls were then each split into three indexed pools consisting of 66, 66, and 68 individuals. Paired-end sequencing using 100bp reads for cases and controls was run on two separate lanes using the Illumina HiSeq 2000 Systems by fluorescence-based reversible terminator technology. Principles and methodology used are described in detail in Chapter 8.

2.2.4 Dye Terminator Sequencing

Dye terminator sequencing was carried out to verify products of candidate gene expression in Chapters 3 and 4 as well as to sequence insert regions of DNA plasmids produced by 5' rapid amplification of cDNA ends (5' RACE) in Chapter 7. The Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Cat. No. 4337456) was used containing enzymes and dNTPs labelled with fluorescent markers allowing fragments ending at every base pair of the sequence to be produced by the PCR reaction. DNA sequencing is a form of PCR with DNA synthesis in the presence of dNTPs and dideoxynucleoside triphosphates (ddNTPs) in a ratio of 4:1. Extension of the DNA polymerase during DNA synthesis requires a 3' hydroxyl. Using the Sanger chain termination method, fluorophore labelled ddNTPs lacking hydroxyl groups were added to terminate DNA synthesis during the reaction. At the end of the reaction, polyacrylamide gel electrophoresis was used to separate each terminated strand and the DNA sequence to be read by the position of bands in the gel. Automated sequencing was carried out in the School of Biomedical Sciences, University of Nottingham.

Thermal Cycle Sequencing

Samples were prepared by setting up reactions as follows and amplified using a heated lid PCR machine (25 cycles: 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes; 4°C pause):

DNA samples	3µl
Primer (10pM/µl)	1µl
2x Sequencing Buffer	2µl
Big Dye version 3.1	2µl
DEPC Water	to 10µl

Ethanol Precipitation

Samples were precipitated at room temperature for 20 minutes by the addition of 50µl 100% Ethanol, 2µl Sodium Acetate, and 10µl water before centrifugation at 13,000rpm for 45 seconds. The supernatant was removed and resulting pellet washed with 250µl 70% Ethanol before centrifugation for 10 minutes. The ethanol was then removed and pellet dried at 72°C by placing the tubes with lids open in a PCR machine.

Sequencing Reaction and Analysis

Sequence data produced by laser radiation were recorded using an ABI PRISM® 310 Genetic Analyser (Applied Biosystems). Samples were run in the sequencing machines on a sequencing gel or column. Isolated DNA fragments were separated according to size by capillary gel electrophoresis. Each fluorescent peak detected corresponded to a particular base in the sequence. Data in the form of sequence chromatogram files and sequence data files were analysed with Chromas Lite 2.01 software (www.technelysium.com/auy/chromas_lite.html) (see Chapters 3, 4, and 7 for details).

2.3 Bioinformatics

A range of bioinformatics techniques have been used to carry out *in silico* work using online sequence databases, search engines, and genetic analysis software. URLs for online resources are listed in Appendix 1.

2.3.1 Online Resources

Genome sequence and polymorphism information were obtained from NCBI (dbSNP, Nucleotide, Gene), the UCSC Genome Browser, Ensembl, and the 1000 Genomes Project. Data from both genome releases NCBI Build 36.1 (release date March 2006) and Genome Reference Consortium GRCh37 (release date February 2009) were used.

Haplotype data for the CEU (Utah residents with ancestry from northern and western Europe) population panel was obtained from the International HapMap Project using HapMap3 Genome Browser release #2 (Phase 3 - genotypes, frequencies & LD). Searches based on the genomic regions of interest were carried out to generate phased haplotype display images (see Chapter 4 for details). The SNAP pairwise LD program (Broad Institute) was used to process HapMap3 data for individual SNPs throughout this thesis; search options were based on distance limits of 500bp and no limits for the r^2 threshold.

PCR suitability of primers designed for gene expression and 5' RACE were assessed using the web-based tool PCR Primer Stats by detecting the presence of undesirable internal secondary structures such as palindromes, self-annealing bases, and hair-pins. Pass/fail/warning values were generated based on single base runs, dinucleotide base runs, % GC, melting temperature, GC clamp, self-annealing, and hairpin formation. PCR products were compared to reference sequences using search tool BLAST (NCBI) (see Chapters 3, 4, and 7 for details).

Transcription factor analyses were carried out using four search tools: WWW Signal Scan TFD, WWW Signal Scan TRANSFAC, Transcription Element Search System (TESS), and TFSEARCH. Protein information for IL1RL1 and IL33 was obtained using protein sequence

database UniProtKB/Swiss-Prot. Protein prediction was carried out using analysis tool EMBOSS Transeq (EMBL-EBI) by entering nucleotide sequences of interest detected using 5' rapid amplification of cDNA ends (5' RACE) (Rice *et al.* 2000). Procedures for transcription factor and protein sequence analyses are described in detail in Chapter 7.

2.3.2 Genetics Analysis Software

A number of open-source command line programs were downloaded and run using the UNIX operating system. Principal components analysis (PCA) was carried out using EIGENSOFT 3.0 (Patterson *et al.* 2006). Haplotype phasing was carried out using MACH 1.0 (Li and Abecasis 2006) and subsequent imputation using minimac (http://genome.sph.umich.edu/wiki/Minimac). Genome-wide association analysis quality control and association testing for genotype data was carried out using PLINK 1.07 (Purcell *et al.* 2007). Association testing for imputed data was carried out using ProbABEL 0.1-3 (Aulchenko *et al.* 2010). Detailed methods for PCA, imputation, and association testing are described in Chapter 5. MeffLi variables were calculated using SNPSpD (http://gump.qimr.edu.au/general/daleN/SNPSpD/) (see Chapter 6 for details).

Power calculations were estimated using CaTS 0.0.2; estimation parameters used were based on the sample sizes and disease prevalence of tested population cohorts under an allele additive model, detailed in the relevant results chapters (Skol *et al.* 2006). Region plots were generated using the online graphical tool LocusZoom.1.1 (Pruim *et al.*). General statistical analyses were carried out using R version 2.12.1, Microsoft Excel 2007, SPSS 15.0 for Windows, and STATA/SE 10.1 for Windows. Manhattan and Quantile-quantile (Q-Q) plots were also generated using scripts written in R version 2.12.1.

2.4 Statistical Analyses

A range of statistical methods have been used to analyse genotype data in the stages of quality control, association testing, and analysis of results. The principles behind these analyses are explained below. Procedural details are given in the relevant results chapters.

2.4.1 Principal Components Analysis (PCA)

Principal components analysis (PCA) of study populations was carried out to correct for population substructure in severe asthma analyses based on the covariance of effect allele loadings (0, 1, 2) using EIGENSOFT 3.0 (Price *et al.* 2006). PCA converts the set of observations of potentially correlated variables into a set of values of uncorrelated variables called principal components, less than or equal to the number of the original variables. The first principal component is defined to maximise variance between individuals, subsequent components in turn maximise remaining variance under the constraint of being orthogonal i.e. uncorrelated with the preceding components. Individuals with very different locations on a principal component may have very different ancestry proportions; outliers were defined with variances $>6\sigma$.

2.4.2 Association Testing

SNP association testing was carried out using an allele additive model to carry out per-allele tests given no presumptions about the underlying disease mechanisms. A number of statistical tests based on a range of probability distributions were used in analyses.

2.4.2.1 Degrees of Freedom

The degrees of freedom of each test describe the number of components that need to be known before the vector is fully determined. In general, this is equal to the number of independent scores included in the estimate of a parameter minus the number of parameters estimated as intermediate steps.

2.4.2.2 Pearson's Chi-squared (χ^2) Test

The Pearson's Chi-squared (χ^2) test carried out in MS Excel 2007 was used to compare elevated specific IgE and positive SPT case control status between alleles by comparing observed versus expected frequencies computed under the null hypothesis of independence. Quality control tests of SNP MAF and HWE in atopy analyses were also carried out using this test. The test statistic $X^2=\Sigma((O-E)^2/E)$ was calculated by the sum of the squared differences between the observed and expected frequencies [Table 2.5].

Table 2.5 – Cross-tabulation of allele frequencies in case control analyses

Frequencies	**Observed (O)**		**Expected (E)**	
	Case	Control	Case	Control
Minor allele	a	b	(a+b)(a+c)/(a+b+c+d)	(a+b)(b+d)/(a+b+c+d)
Major allele	c	d	(c+d)(a+c)/(a+b+c+d)	(c+d)(b+d)/(a+b+c+d)

The p value and standard deviation (SD) were calculated by comparing X^2 to a χ^2 distribution with one degree of freedom. The χ^2 distribution is based on the sum of the squared values of the N(0,1) distribution observations. It has a positive skew with a mean equal to the number of degrees of freedom; as this increases, the χ^2 curve approaches a normal distribution. Odds ratios (OR) with 95% confidence intervals (CI) were calculated for case control analyses of elevated specific IgE and positive SPT using OR=(a/b)/(c/d) and 95% CI=OR±1.96SD.

2.4.2.3 Fisher's Exact Test

Fisher's exact test using PLINK 1.07 was carried out for quality control tests of SNP MAF, HWE, and differential missingness in severe asthma analyses by assuming fixed row and column totals and using the hypergeometric distribution to compute probabilities of possible tables conditional on these values. P values were determined by p=[(a+b)!(c+d)!(a+c)!(b+d)!]/(a!b!c!d!n!). This test does not rely on using approximate theoretical distributions and is therefore more accurate in evaluating the significance level of the test, especially with small numbers of observations or imbalances between cases and controls.

2.4.2.4 Linear Regression

Linear regression using STATA/SE 10.1 was used to test association of serum total IgE with alleles by calculating the best-fitting regression line for the observed data through minimising the sum of the squares of the vertical deviations from each data point to the line. Serum total IgE data were $\log_e$ transformed to result in a normal distribution. The equation of the straight line given by $y=\alpha+\beta x$ with constant (α) and beta coefficient (β) is used to estimate the effect size for trend increases per effect allele with standard error (SE). The p value was calculated by assessing the test statistic $t=\beta/SE$ with n-2 degrees of freedom using the t distribution. The t distribution is similar to the N(0,1) distribution but altered by the number of degrees of freedom. The t distribution has a mean equal to 0; with infinite numbers of degrees of freedom, the distribution becomes normal.

2.4.2.5 Logistic Regression

Logistic regression using PLINK 1.07 was used to test association of severe asthma case control status with alleles using genotype data. ProbABEL 0.1-3 was used to test association using imputed data with effect allele dose on a continuous scale between 0 and 2 to reflect imputation uncertainty. This model is a variation of linear regression for dichotomous variables which fits observed data to a logit function logistic curve. The logistic function is given by $f(z)=1/(1+e^{-z})$ with output $f(z)$ and input z ranging from negative to positive infinity. The logit $z=\alpha+\beta x$ is the measure of the contribution of the independent variable. The likelihood ratio test was used to compare the fit of the null and the alternative models by assessing how many times more likely the data are using one model than the other. The test statistic D=-2ln(likelihood for null model/likelihood for alternative model) was approximated using the χ^2 distribution with one degree of freedom to calculate the p value. OR with 95% CI were calculated using $OR=EXP(\beta)$ and 95% CI=OR±1.96SE.

2.4.3 Test Inflation

Test inflation statistics (λ) were calculated in R version 2.12.1 using $\lambda=[median(\beta/SE)^2]/[median(qchisq(0.5,1)]$ where qchisq(0.5,1) is the χ^2 distribution with one degree of freedom. This is equal to one under the null hypothesis.

2.4.4 Statistical Power

Retrospective power analyses for GWA analyses were carried out using CaTS 0.0.2 at the 5% significance level to give estimates of the minimum effect sizes of polymorphisms likely to be detected in analyses using the available sample sizes. The power of a statistical test is given by one minus the false negative rate or type II error. Power increases with increasing sample size and decreasing study variance.

2.4.5 Meta-Analysis

Meta-analysis was used to combine results from individual studies in order to identify a common measure of effect size. This aimed to provide greater power in estimating the true effect size for this phenotype as opposed to those derived from a single study under a single set of assumptions and conditions. A fixed effects model with weights inversely proportionate to individual study variances was used based on the assumption that the true effect size θ is normally distributed with $N(\theta,\sigma_\theta)$, where σ_θ^2 is the within study variance of the effect size.

In atopy analyses, this was carried out using the metan function in STATA/SE 10.1. As all studies have the same expected fixed effect size using this function, between study differences are not considered. Therefore, tests for heterogeneity in effects across studies were carried out using test statistics which measured the weighted sum of the differences between k individual study effects and the overall estimate under a χ^2 distribution with k-1 degrees of freedom.

In severe asthma analyses, weights for each study ($W_{1..n}$ for primary AUGOSA, $W_{2..n}$ for replication AAGC) were calculated in MS Excel using the effect size estimates ($\beta_{1..n}$) and variances ($SE_{1..n}$) per $SNP_{1..n}$ given by $1/(SE_{1..n}^2)$. The results from meta-analysis were calculated for example for SNP_1 using $\beta_1(meta)=(\beta_1W_1+\beta_2W_2)/(W_1+W_2)$ and $SE_1(meta)=\sqrt{[1/(W_1+W_2)]}$.

2.4.6 Linkage Disequilibrium (LD)

Linkage disequilibrium (LD) occurs when some combinations of alleles occur more frequently in a population than would be expected from a random formation of haplotypes based on the

frequencies of each relevant allele i.e. the non-random association of alleles at two or more loci. Levels of LD are influenced by factors including linkage, selection, recombination rate, genetic drift, non-random mating, and population structure. For a two locus, bi-allele haplotype model A_1B_1, A_1B_2, A_2B_1, A_2B_2 with relative frequencies x_{11}, x_{12}, x_{21}, x_{22} respectively, allele frequencies are calculated using $p_1=x_{11}+x_{12}$ (A_1), $p_2=x_{21}+x_{22}$ (A_2), $q_1=x_{11}+x_{21}$ (B_1), $q_2=x_{12}+x_{22}$ (B_2). Hence the frequency of A_1B_1 is given by $x_{11}=p_1q_1$. Parameter $D=x_{11}-p_1q_1$ constrained by allele frequencies measures the difference between observed and expected haplotype frequencies. A value of 0 denotes independence under linkage equilibrium (LE) (Lewontin 1964). Two pairwise measures of LD are generally used based on the D parameter. D' is given by D/D_{max}, r^2 is given by $D^2/(p_1p_2q_1q_2)$ and adjusted for population size. Both D' and r^2 give an indication of recombination history and range from 0 to 1 i.e. complete LE to LD, r^2 can also be used to predict the frequency of one polymorphism given the other.

2.4.7 Imputation

Genotypes for untyped SNPs in the ALSPAC cohort, the AUGOSA study, and the AAGC study were inferred using imputation to allow comparison of GWA data across the multiple population cohorts genotyped using different platforms. This approach relies on the assumption that apparently unrelated individuals can share haplotypes inherited from distant common ancestors. Reference sample haplotypes were downloaded from the International HapMap Project and the 1000 Genomes Project. Selected SNPs were imputed in atopy analyses to allow comparison of genotypes across the platforms used. Imputation of genotyped SNPs to over 6 million SNPs were carried out in severe asthma analyses. Imputation carried out using MACH 1.0 Markov Chain based haplotyper for phasing and minimac for imputation. The implemented method uses genotyped SNPs as proxies to randomly generate a pair of haplotypes compatible with observed genotypes for each sample individual. These initial estimates are then refined by carrying out a series of 20 to 100 iterations, during which a Hidden Markov Model (HMM) generates a new haplotype pair for each individual and describes the haplotype pair as an imperfect mosaic of the other haplotypes. The probability of change in the mosaic pattern between every pair of consecutive markers and the probability of observing an imperfection in the mosaic at each specific point is measured in order to construct a consensus haplotype

sampled in each round. Quality measures based on r^2 were calculated by estimating the squared correlation between imputed and true genotypes. A cut-off of $r^2<0.3$ was used to identify the majority of poorly imputed SNPs whilst only excluding <1% of well imputed SNPs.

2.4.8 Significance Testing

Incorrect rejection of the null hypothesis is more likely to occur when a set of statistical inferences are considered simultaneously such as when testing a group of SNPs. Corrections for multiple testing generally require a stronger level of evidence to be observed in order for an individual comparison to be deemed significant, so as to compensate for the number of inferences being made. In addition to the use of nominal significance and evaluation for SNP clustering to assess statistical significance, two additional approaches were used to correct for multiple testing at the 5% significance level described below.

2.4.8.1 Bonferroni Correction

The Bonferroni correction is derived by observing Boole's inequality (Miller 1981). If n tests are performed, each significant with probability β (where β is unknown), then the probability that at least one of them comes out significant is $\leq n\beta$. If the required significance level for the entire series of tests is equal to α; by solving for β, $\beta=\alpha/n$. This correction does not require tests to be independent.

2.4.8.2 MeffLi Correction

A large number of SNPs tested exhibit linkage disequilibrium (LD) not accounted for by the Bonferroni correction. SNP spectral decomposition (SNP SpD) was used to calculate the effective number of independent variables, given by $M_{eff}=1+(M-1)(1-Var_{\lambda obs}/M)$, to allow for pairwise SNP correlation (Nyholt 2004). The variance of the Eigenvalues of the correlation matrix of SNP alleles ($Var_{\lambda obs}$) estimated using reference genotypes obtained from Busselton controls will vary from M (the total number of SNPs) when they are perfectly correlated to 0 when there is no correlation. The resulting significance threshold was calculated using $p=\alpha/M_{eff}$ given the required significance level for the entire series of tests is equal to α.

2.5 Cellular and Molecular Techniques

A range of techniques were used to analyse RNA and DNA for testing expression and sequencing of candidate genes in Chapters 3, 4, and 7.

2.5.1 Cell Culture

Sub-culture of cells sourced from consented individuals without history of allergic or respiratory disease was carried out and detailed in the following text. These cells were cryopreserved in liquid nitrogen (-179°C) using 10% dimethyl sulphoxide (DSMO, diluted with FCS) and routinely maintained in the Division of Therapeutics & Molecular Medicine. All cultured cells were incubated at 37°C in 5% CO_2. Reagents for cell culture were obtained from Sigma-Aldrich (Dorset, UK) unless otherwise stated. All cell culture plastics were obtained from Corning life sciences from Fisher Scientific (Loughborough, UK).

2.5.1.1 Human Airway Smooth Muscle (HASM) Cells

Adherent, primary human airway smooth muscle (HASM) cells were obtained by colleagues from thoracotomy surgery and bronchial biopsies from three donors (D0501, D0401, D0502) (Daykin *et al.* 1993). Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) containing 2mM glutamine from Sigma-Aldrich (Dorset, UK), supplemented with 10% FCS. Cells were utilised between passages 3 and 5.

2.5.1.2 Human Bronchial Epithelial Cells (HBECs)

Adherent, undifferentiated human bronchial epithelial cells (HBECs) sourced from 3 donors (5F0750, 4F0507, 7F3158) were purchased from Clonetics-Biowhittaker™ (MD, USA). Cells were cultured in bronchial epithelial cell basal medium (BEGM) from Lonza (Berkshire, UK) supplemented with 52µg/ml bovine pituitary extract (BPE), 0.5µg/ml hydrocortisone, 0.5µg/ml human recombinant EGF (hEGF), 0.5ng/ml epinephrine, 5µg/ml insulin, 6.5ng/ml triiodothyronine, 10µg/ml transferrin, 50µg/ml gentamicin, and 50µg/ml amphotericin-B. Cells were utilised at passage 3 until 80% to 90% confluency.

2.5.1.3 Dermal Fibroblast Cells

Flasks of confluent dermal fibroblast cells from two donors (6C193 at passage 8, C240209 at passage 5) were provided by Dr. Christine Pullar, University of Leicester. RNA extraction was carried out without further cell culture.

2.5.1.4 Bronchial Epithelial Cell Line (BEAS-2B)

Adherent cells, obtained from the American Type Culture Collection (ATCC) via LGC Promochem (Middlesex, UK), were cultured from an immortalized human bronchial epithelial cell line (BEAS-2B) originally isolated from the normal human bronchial epithelium of a non-cancerous donor for lung transplantation (Reddel *et al.* 1988). Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) containing 2mM glutamine from Sigma-Aldrich (Dorset, UK), supplemented with 10% FCS. Cells were utilised up to passage 13.

2.5.1.5 Human Mast Cell Line (HMC-1)

Non-adherent cells, obtained from the ATCC via LGC Promochem (Middlesex, UK), were sub-cultured from a human mast cell line originally derived from a patient with mast cell leukaemia (Butterfield *et al.* 1988). Cells were cultured in GIBCO™ Iscove's Modified Dulbecco's Medium (IMDM) containing GlutaMAX I medium and 25mM HEPES buffer from Invitrogen-GIBCO (Paisley, UK), supplemented with 10% FCS. Cells were utilised at passage 3.

2.5.1.6 Human Acute Monocytic Leukaemia Cell Line (THP-1)

Non-adherent cells, obtained from the ATCC via LGC Promochem (Middlesex, UK), were sub-cultured from a human acute monocytic leukaemia cell line (THP-1) originally isolated from the peripheral blood of a 1 year old male with acute monocytic leukaemia (Tsuchiya *et al.* 1980). These were cultured in RPMI-1640 medium from LGC Promochem (Middlesex, UK). Cells were utilised up to passage 14.

2.5.2 Molecular Biology Methods

2.5.2.1 Total RNA Extraction

A panel of total RNA samples were collected for use in gene expression profiling and 5' RACE. Commercial total RNA was obtained for peripheral blood mononuclear cells (PBMCs) from two donors (RNA051206, RNA051206) from 3H Biomedical (Uppsala, Sweden); lung tissue from two donors (07020236, 1002004), brain tissue from two donors (073P010203017A, 030503013), kidney tissues from two donors (08010041, 0811001), and placenta tissue from two donors (114P020303015A, 03070032) all from Applied Biosystems/Ambion (Warrington, UK). Total RNA was isolated from cultured HASM, HBEC, dermal fibroblast, BEAS-2B, and THP-1 cells using the RNeasy mini Kit (Qiagen. Cat. No. 74104) and the RNAase-free DNase set (Qiagen. Cat No. 79254). This method of spin-column technology allows isolation of all RNA molecules >200 nucleotides in length. Prior to all RNA preparations, procedures were undertaken to minimise contamination using RNase Away (Invitrogen).

Cell density following sub-culture was ascertained using a haemocytometer to estimate cell counts using 10µl aliquots. Total RNA was extracted from confluent cells. Confluent cells were trypsinised, washed with PBS, and pelleted before re-suspension in 600µl of Buffer RLT with added β-Mercaptoethanol (10µl β-ME/ml Buffer RLT). The lysate was pipetted onto a Qiashredder column for centrifugation at maximum speed for 2 minutes to homogenise the mixture. One volume (600µl) of 70% ethanol was added and mixed well by pipetting. Up to 700µl of the sample with resulting precipitate was loaded onto an RNeasy column with collection tube centrifuged at 10,000rpm for 15 seconds discarding the flow-through. Surplus volumes were aliquoted onto the column successively.

Subsequent on-column DNase treatment was carried out. Solid DNaseI supplied with the RNase free DNase set (QIAGEN) was gently dissolved in 550µl DNase free water and stored at -20˚C. 350µl of Buffer RW1 was pipetted onto the RNeasy column and centrifuged for 15 seconds at 10,000rpm discarding the flow through. 10µl of the DNaseI stock solution was added to 70µl Buffer RDD and mixed by gentle inversion. This DNaseI mix was then added to

the spin column membrane and incubated at room temperature for 15 minutes to degrade the DNA. Buffer RW1 (350µl) was added to wash the column and centrifuged for 15 seconds at 10,000rpm discarding the flow-through. This procedure was repeated twice using 500µl RPE Buffer with added ethanol (RNeasy mini kit) centrifuging for 15 seconds and 2 minutes respectively. A clean collection tube was added to the column and centrifuged at maximum speed for 1 minute to remove all residual wash solution before transferring to a 1.5ml RNase free microfuge tube. 30µl RNAse free water was added directly onto the membrane and centrifuged at 10,000rpm for 1 minute to elute the RNA. Samples were stored at -80˚C.

Quality control for RNA integrity was checked by analysing 1µl of each sample by 1% agarose gel electrophoresis to check the 28s and 18S rRNA bands in total RNA seen at 4.5kb and 1.9kb respectively (see Section 2.5.2.5). Using 1µl of the sample, the Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies) was used to measure the RNA concentration and assess RNA purity indicated by a 260/280 10mm absorbance ratio between 1.9 and 2.1.

2.5.2.2 cDNA Synthesis

Reverse transcription of single stranded RNA to synthesise complementary cDNA was carried out using the SuperScript™ First-Strand Synthesis System for RT-PCR (Invitrogen. Cat. No. 11904-018). The first-strand cDNA synthesis reaction is catalysed by SuperScript™ II Reverse Transcriptase (RT) and primed using random hexamers to produce cDNA of different lengths representing the whole RNA sequence. PCR sensitivity was increased by removal of the RNA template from the cDNA:RNA hybrid molecule by digestion with RNase H after first-strand synthesis. RT negative samples were prepared alongside each cDNA synthesis reaction to act as negative controls for genomic DNA contamination. The final cDNA samples were diluted using DEPC water in the ratios of 1:5 for use in PCR and 1:25 for use in Taqman® analysis.

The protocol used for first strand cDNA synthesis is detailed as follows. RNA/primer mixtures were prepared in sterile 0.5ml tubes and incubated at 65°C for 5 minutes followed by standing on ice for at least 1 minute:

Total RNA	600ng

Random hexamers (50ng/µl)	1µl
10mM dNTP mix	1µl
DEPC-treated water	to 10µl

Subsequently, 2µl 10x RT Buffer, 25mM $MgCl_2$, 0.1M DTT, and 1µl RNaseOUT Recombinant Ribonuclease Inhibitor were added to each reaction. The reaction was mixed gently and collected by brief centrifugation before incubation at 25°C for 2 minutes. To test reactions, 1µl (50 units) of SuperScript™ II RT was added and incubated at 25°C for 10 minutes, 42°C for 50 minutes and 70°C for 15 minutes before chilling on ice. The reactions were collected by brief centrifugation and 1µl of RNase H added to each tube prior to incubation for 20 minutes at 37°C.

2.5.2.3 Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a sensitive technique designed to indirectly detect levels of mRNA expression using cDNA for PCR amplification using gene-specific primer pairs and a *Taq* DNA polymerase enzyme. The PCR reaction involves three consecutive steps which can be repeated up to 30 to 35 cycles, with exponential amplification of the DNA product. Denaturation separates the double stranded DNA molecule followed by annealing of primers to their target sites and then extension to amplify the primer defined sequence. Typically 25 to 30 cycles of reaction occurs, yielding more than 10^7 copies of the amplified sequence. Efficiency of the reaction reduces after 40 cycles.

Primer Design for Gene Expression

Forward and reverse primers for annealing to opposing strands of the DNA duplex chain were designed by hand, crossing intron and exon boundaries where possible to ensure cDNA specificity. Suitability for primer use was examined using the following rules:

Primer length 18bp to 25bp
Primers with ≤3 contiguous bases
Avoiding repetitive sequences
40% to 60% G/C content
Amplified sequence length 300bp to 1,000bp
Matched primer Tm (<5˚C difference) i.e. at which 50% of the sequence is annealed

G/C at the 3' end of the primer but ≤3

Primer integrity was assessed using the web-based tool PCR Primer Stats (see Section 2.3.1). All primers used in analyses in expression of different genes are listed in Appendix 2.

PCR Assay Conditions for Gene Expression

Samples were prepared by setting up reactions as follows and amplified using a heated lid PCR machine (40 cycles: 94°C for 45 seconds, variable annealing temperatures for 30 seconds, 72°C for 1.5 minutes; 72°C for 10 minutes; 4°C pause):

10x PCR Buffer	2.5µl
Magnesium Chloride (50mM)	3mM
dNTP mixture (10mM dATP/dTTP/dGTP/dCTP)	200µM
Taq DNA polymerase (2.5U/µl)	0.025U/µl
Forward primer (10µM)	0.2µM
Reverse primer (10µM)	0.2µM
DNA template	2-10ng
Sterile distilled water	to 25µl

Optimisation was carried out by varying annealing temperatures and Magnesium ion (Mg^{2+}) concentrations between 1.5mM and 3mM. Positive control housekeeping gene, Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) and negative controls were included in each experiment. Specific conditions following assay optimisation for testing expression of different genes are detailed in Appendix 2.

Real-Time Quantitative PCR (Taqman®)

Real-time quantitative PCR based on Taqman® methodology was used to measure levels of *IL1RL1* and *IL33* relative mRNA expression. This procedure follows the general principle of PCR but detects amplified DNA as the reaction progresses in real-time. Principles behind this technique and the protocols followed are detailed in Chapter 7.

2.5.2.4 Rapid Amplification of cDNA Ends (RACE)

Rapid Amplification of cDNA Ends (RACE) was used to obtain the RNA transcript sequences of the *IL1RL1* and *IL33* genes found within tested mRNA templates from a small known

sequence within each transcript to the 5' end (5' RACE). This technique results in the production of a cDNA copy of the RNA sequence of interest using reverse transcription, followed by PCR amplification of the cDNA copies, which are then sequenced. Detailed principles and protocols used are detailed in Chapter 7.

2.5.2.5 Agarose Gel Electrophoresis

Nucleic acids such as PCR products can be analysed using agarose gel electrophoresis. Applying a current through the electrophoresis chamber causes negative DNA fragments to migrate from the negative electrode at the top of the gel to the positive electrode at the bottom with fragments moving at a slower rate with increasing size. The DNA fragments are stained by ethidium bromide which intercalates between the bases and fluoresces under ultraviolet (UV) light. Accurate visualisation of separated PCR products can then be carried out under an UV transilluminator (Anachem, UK) using a DNA ladder to estimate the size and concentration of fragments. Different percentage agarose gels were used to run PCR products depending on the resolution of separation required. Smaller PCR products migrate at a faster rate across the electrophoresis gel requiring higher percentage agarose gels.

Agarose (Invitrogen, electrophoresis grade. Cat. No. 540-5510UB) at the required concentration was added to 1x TAE Buffer (0.4M Tris Base, 0.2M acetic acid and 0.01M ethylendiaminetetraacetic acid), e.g. for 1% gels, a mixture of 1g agarose per 100ml 1x TAE Buffer was used. The mixture was microwave heated to dissolve and 5µl 10mg/ml ethidium bromide (Sigma, UK catalogue number: E-1510) was added per 100ml of solution. When cooled to 50˚C, the mixture was poured into a gel tray with well combs and left to set at room temperature. The set gel was placed into the electrophoresis chamber (Bio-Rad) and 1x TAE buffer was added to cover the gel to a depth of 3mm above the gel surface. Loading dye (6x orange G loading dye: 0.25% orange G dye, 30% glycerol) was added to each PCR product (2µl loading dye per 10µl PCR product) to increase density. A 1kb or 100bp DNA ladder (10% DNA ladder, 20% BlueJuice™; Invitrogen) was also run. Gels were run at approximately 80mA to 100mA for 35 minutes for small SciPlas/Fisher Scientific (Leicestershire, UK) gel tanks and 125-155mA for 1 hour for large Biometra Agagel Maxi (Glasgow, UK) gel tanks

until the dye front had migrated sufficiently. DNA bands were subsequently visualised under UV light in a GeneGenius Gel Doc system and photographed using Syngene software.

2.5.2.5 Gel Extraction and Purification

Extraction of gel-fractionated DNA from agarose gels was carried out using the StrataPrep™ DNA Gel extraction kits (Stratagene, Cat. No. 400766, 400768). DNA extraction eliminates the need for manipulation of resins, toxic phenol-chloroform extractions, and time consuming ethanol precipitations. Buffers were used to wash away unwanted constituents and DNA eluted to result in a purified product for restriction digestion, ligation, and probe labelling. An 80% recovery rate is expected from DNA that is 250bp to 9kb in length; a 50% recovery is expected for longer DNA up to 23kb.

300µl of Buffer QG was added to the gel slices in an eppendorf tube and incubated at 50˚C to allow the gel slice to dissolve. The sample was then applied to the spin column in 2ml collection tubes and centrifuged for 1 minute at 13,000rpm allowing for DNA to bind to the columns discarding any flow-through. 0.75ml of Buffer PE was added and unwanted contaminants washed off by further centrifugation for 1 minute. This step was repeated to ensure purification of the DNA. Then, 50µl of Buffer EB was added and tubes centrifuged for 1 minute to elute the DNA. The purified DNA was checked using agarose gel electrophoresis and stored at 5˚C until required.

2.6 Individual Contributions and Collaborations

The work described in this thesis has resulted through a combination of individual and collaborative efforts. A number of established population cohorts have been used in analyses, in which genotyping using genome-wide SNP platforms was previously carried out by individual study centres.

Studies of atopy within multiple populations was possible due to provision of summary results after quality control and association testing of genotype data through collaboration with Prof. David P Strachan (St Georges' University, UK), Dr. John Henderson (University of Bristol, UK), Dr. Tricia McKeever (University of Nottingham, UK), Dr. John W Holloway (University of Southampton, UK), and Dr. Eva Halapi (deCODE Genetics, Iceland).

In studies of severe asthma, processing of genotyped and sequencing data including quality control and association testing was carried out in collaboration with Dr. Nick R G Shrine (University of Leicester, UK). Summary results for SNPs studied in the follow-up replication cohort were provided by Dr. Manuel A R Ferreira (Queensland Institute of Medical Research, Australia).

Descriptions of all population cohorts used including details on the set up of each study, participant recruitment and phenotyping are given in Section 2.1. Genotyping procedures, quality control measures, and association testing undertaken, were either summarised by collaborators where stated or carried out by myself as described in Section 2.2.

Chapter 3

Genome-Wide Association Study to Identify Novel Atopy Susceptibility Loci

3.1 Introduction

As discussed in Chapter 1, there is extensive evidence supporting a genetic contribution to the risk of developing atopy. This was first established through familial clustering of disease and concordance rates amongst both monozygotic and dizygotic twins. Efforts to identify causative genetic loci and genes have been made through studies using linkage, candidate gene association, and more recently, genome-wide association (GWA). In recent years, GWA approaches have been used extensively to study a range of allergic diseases and associated traits. To date, a total of eight GWA studies have been published for asthma, one for atopic dermatitis, and one for serum total IgE levels by studying these phenotypes as the primary outcome (see Chapter 1: Section 1.3).

Results presented from these studies have supported association of previously reported candidate genes such as *IL13* and *HLA-DRB1* as well as identifying novel loci including 1q31 and 9q21.31 for asthma and 11q13 for atopic dermatitis (Ober *et al.* 2008; Moffatt *et al.* 2010). This methodology has allowed identification of genes including *ORMDL3* for childhood asthma susceptibility and *FCER1A* for determining serum total IgE levels (Moffatt *et al.* 2007; Weidinger *et al.* 2008). However, no study utilising a GWA approach has yet examined predisposition to atopy using measures of specific IgE.

The majority of novel loci identified by GWA studies in allergic diseases appear to be trait dependent; for example, different genotype and allele distributions in *TGFB1* polymorphisms have been shown in atopic compared to non-atopic asthmatics as well as in patients with more severe airflow obstruction (Mak *et al.* 2006). An exception to this has been the Th2 cytokine

cluster on chromosome 5q31-33 where association with genetic variants across this region is evident across multiple atopy phenotypes (Palmer *et al.* 2000). Published literature suggests that IgE is a distinct sub-phenotype involved in atopic disease development and therefore may be controlled by independent genetic factors (Weinmayr *et al.* 2007; Moffatt *et al.* 2010).

Considering the extent of phenotypic variability in defining complex diseases such as asthma, using related traits for genetic studies such as serum IgE levels and skin prick test (SPT) responses to common allergens may help to determine disease severity and characterise different forms of disease, for example, atopic and non-atopic asthma (Sandford *et al.* 1998). Furthermore, the use of more quantifiable and objective measures such as specific IgE responses, serum eosinophil counts, and bronchial hyperresponsiveness (BHR), which are closely associated with atopy and asthma could prove to be useful surrogate markers for atopic diseases in genetic studies. These measures can increase study power by improving phenotyping as they are less likely to underestimate disease prevalence compared to physician diagnosis or questionnaire-based self-reporting, particularly in asymptomatic individuals (Kauffmann *et al.* 2002).

3.2 Aim

The aim of this chapter was to identify susceptibility loci for atopy defined by elevated specific IgE or positive SPT using a non-parametric based GWA study.

3.3 Methods

3.3.1 Study Populations

The primary GWA study was carried out using the British 1958 Birth Cohort consisting of 1,083 cases and 2,770 controls genotyped using the Illumina HumanHap550 SNP platform. Statistically significant SNPs were assessed for replication *in silico* in 1,324 individuals from the ALSPAC cohort genotyped on the Illumina HumanHap300 SNP platform as well as being prospectively genotyped in 1,286 individuals from the Nottingham Gedling cohort and 757 individuals from the Southampton cohort. Details of study recruitment and phenotyping are described in Chapter 2: Section 2.1.1. Genotyping and quality control procedures undertaken are summarised in Chapter 2: Section 2.2.1.1.

Atopy was defined using elevated specific IgE $\geq 0.35 kU_A/L$ in the British 1958 Birth Cohort, positive skin prick test (SPT) with a mean wheal diameter ≥2mm in the ALSPAC cohort, positive SPT ≥3mm in the Nottingham Gedling cohort, and measures of both specific IgE $\geq 0.35 kU_A/L$ and SPT >3mm were taken in the Southampton cohort. Analysis of SNP association with serum total IgE levels was also undertaken in 3,873 individuals from the British 1958 Birth Cohort, 1,423 individuals from the ALSPAC, 1,293 individuals from the Nottingham Gedling cohort, and 675 individuals from the Southampton cohort. Adjustments for serum total IgE level data in the British 1958 Birth Cohort had been taken for potential confounders: sex, date & time of collection, postal delay, laboratory batch and nurse.

Power estimates calculated using CaTS 0.02 for each study dataset used in elevated specific IgE and positive SPT analyses are listed in Table 3.1. Each individual cohort had low power to detect modest effects, which was improved through combined analyses.

Table 3.1 – Study power for datasets used in atopy analyses. Values for the minimum detectable OR for a SNP with a given MAF are presented at 80% power.

Study		**Minimum detectable OR with MAF**		
		10%	**25%**	**40%**
Discovery				
British 1958 Birth Cohort		1.19	1.13	1.11
Replication *in silico*				
ALSPAC		1.39	1.27	1.24
Replication genotyping				
Nottingham Gedling		1.30	1.21	1.19
Southampton	Elevated specific IgE	1.30	1.21	1.19
	Positive SPT	1.22	1.16	1.15
Combined				
Elevated specific IgE		1.16	1.11	1.10
Positive SPT		1.18	1.12	1.11

3.3.2 Association Testing

Summary data in the form of allele counts following GWA testing of 459,334 autosomal SNPs genotyped in the British 1958 Birth Cohort passing quality control was provided for analysis. Allele counts for SNPs undergoing replication analyses were provided from a previous GWA study in the ALSPAC cohort. All tests of association for elevated specific IgE and positive SPT in these populations were carried out using Pearson's Chi-squared (χ^2) per-allele tests with one degree of freedom (df) comparing allele frequencies between cases and controls. Linear regression was used to test association of serum total IgE in the British 1958 Birth Cohort (see Chapter 2: Section 2.4.2).

During analyses, odds ratios (ORs) with 95% confidence intervals (CI) were calculated to assess minor allele effect sizes using provided allele counts. Following prospective genotyping of SNPs in the Nottingham Gedling cohort and the Southampton cohort, Pearson's Chi-squared (χ^2) per-allele tests with 1df were carried out to test association with elevated specific IgE and positive SPT. All serum total IgE level data were $\log_e$ transformed and SNP association tested using linear regression and beta trend (β) coefficients with 95% CI were calculated for the minor allele. All association testing of data from prospective genotyping and calculations of effect sizes were carried out in STATA/SE 10.1.

Pooled analyses for each phenotype used to define atopy were performed using all available count data. This included the British 1958 Birth Cohort and the Southampton cohort for

elevated specific IgE; the ALSPAC, Nottingham Gedling and Southampton cohorts for positive SPT; the Nottingham Gedling and Southampton cohorts for serum total IgE levels. Forest plots were generated using Microsoft Excel 2007 by plotting the minor allele effect size estimates (OR and β with 95% CI) per study cohort for each SNP. Weighted meta-analysis using a fixed effects model of individual per-allele effect sizes for serum total IgE levels was carried out in order to include data across all four study populations carried out in STATA/SE 10.1.

3.3.3 Statistical Significance Testing

Genome-wide statistical significance in the British 1958 Birth Cohort GWA study was defined using $p \leq 1.09 \times 10^{(-7)}$ as determined by the Bonferroni correction at the 5% significance level for a total of 459,334 SNPs. In consideration of the fact that this may be over-conservative particularly in regions with poor SNP coverage, two additional less stringent thresholds were used: nominally defined $p \leq 10^{(-5)}$ for any single SNP and clustering of ≥3 SNPs with $p \leq 10^{(-4)}$ within a 100kb region. Loci were selected for replication analyses by setting limits of 300kb flanking all identified SNPs. Distinct loci with no overlap were taken forward for replication analyses.

3.3.4 SNP Selection for Replication Studies

Replication of SNPs showing the highest statistical significance within each locus passing quality control with minor allele frequency (MAF) >5% and in Hardy-Weinberg equilibrium (HWE $p>0.05$) was tested. Positive replication was assessed at the 5% significance level. Non-genotyped SNPs in the ALSPAC cohort were imputed.

3.3.5 Candidate Gene Expression

Expression profiling of possible candidate genes identified by statistically significant SNP associations was carried out using polymerase chain reaction (PCR) in two independent cDNA panels with two repeats each. Assays were designed to detect all published gene transcripts documented online using NCBI. PCR products were sequence verified (see Chapter 2: Section 2.5).

3.3.6 Evaluation of Previously Reported Atopy Genes

Association of SNPs within atopy candidate genes previously reported in a smaller dataset of the British 1958 Birth Cohort was also analysed (Maier *et al.* 2006).

3.4 Results

3.4.1 Discovery GWA Study

The Quantile-Quantile (Q-Q) plot shows observed versus expected -$\log_{10}$ p values for all tested SNPs in the British 1958 Birth Cohort showing a small number of SNPs with extreme p values and additional SNPs with more modest effects [Figure 3.1]. The Manhattan plot shows -$\log_{10}$ p values for all tested SNPs ordered by chromosomal position showing the highest significance SNPs lying on chromosomes 8 and 13 [Figure 3.2].

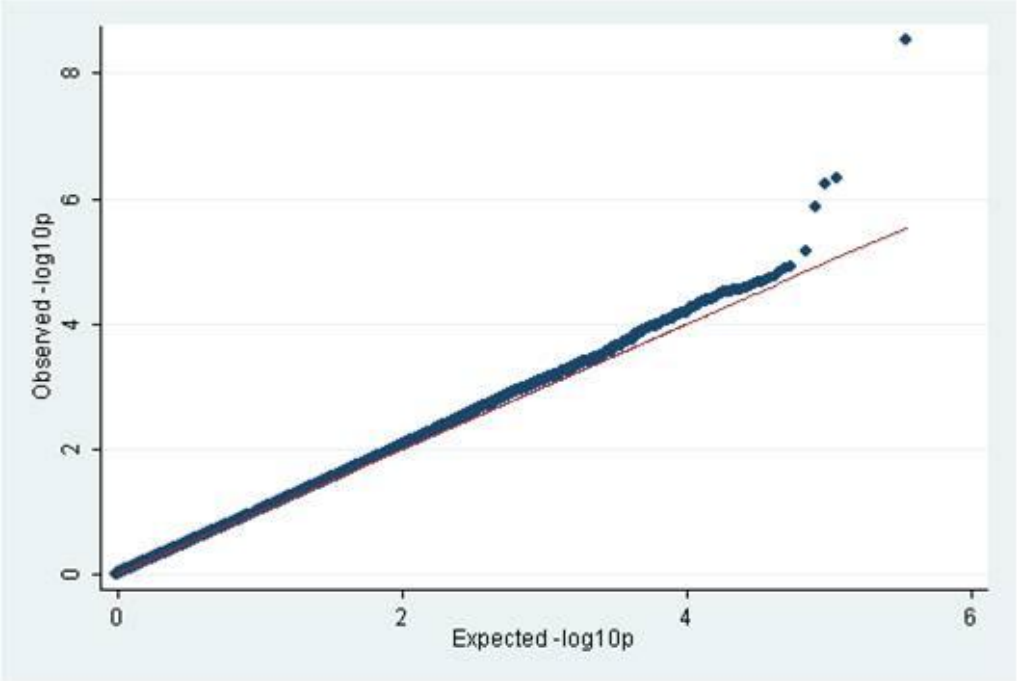


Figure 3.2 – Q-Q plot showing results of the discovery GWA study. The straight line shows the distribution of 459,334 SNPs analysed in 1,083 cases and 2,770 controls under the null hypothesis.

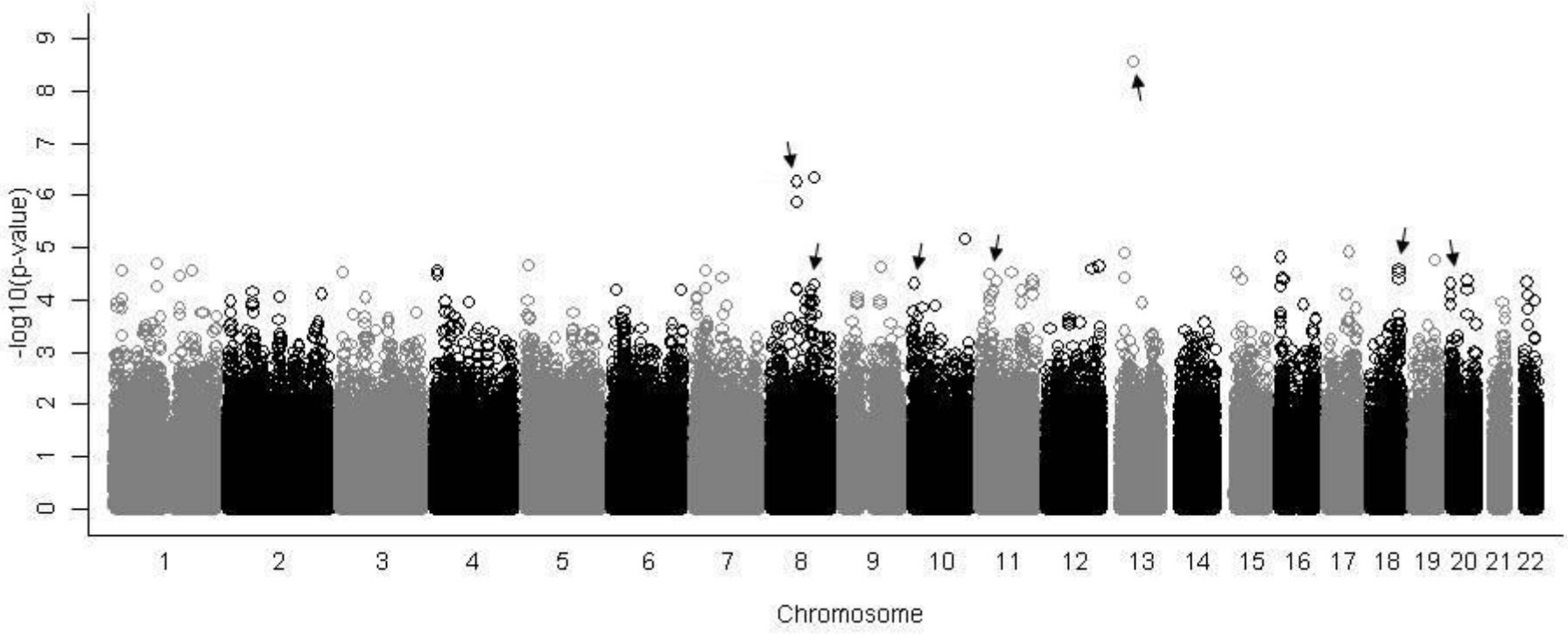


Figure 3.3 – Manhattan plot showing results of the discovery GWA study for 459,334 SNPs analysed in 1,083 cases and 2,770. The arrows indicate the position of the seven loci taken forward for replication studies based on both genome-wide significance ($p \leq 1.09 \times 10^{(-7)}$), nominal significance $p \leq 0.05$, and clustering of ≥3 SNPs with $p \leq 10^{(-4)}$ within a 100kb region.

A single SNP, rs6561505 on chromosome 13q14 reached genome-wide significance ($p=2.15x10^{(-9)}$). A further four SNPs were identified using the lower threshold of $p\leq10^{(-5)}$: rs7835153 ($p=4.50x10^{(-7)}$), rs7835920 ($p=5.54x10^{(-7)}$), rs17202769 ($p=1.39x10^{(-6)}$), and rs11244450 ($p=6.50x10^{(-6)}$), on chromosomes 8q22, 8q12 (2 SNPs), and 10q26 respectively. An additional 11 SNPs were identified based on the pre-defined criteria for clustering with $p\leq10^{(-4)}$ (≥3 SNPs within 100kb) on chromosomes 8q12 (2 SNPs), 11q24 (3 SNPs), 18q22 (3 SNPs), and 20p13 (3 SNPs).

The total of 16 SNPs was grouped into seven distinct loci. The SNPs with the highest statistical significance in each locus were rs7835920 (8q12), rs7835153 (8q22), rs11244450 (10q26), rs4936003 (11q24), rs6561505 (13q14), rs10514054 (18q22), and rs6082335 (20p13). All seven SNPs were taken into replication studies following quality control checks. MAFs of identified SNPs were similar between cohorts for all SNPs apart from rs10514054 and rs6082335 [Table 3.2]. All SNPs apart from rs7835153 showed a risk effect for the minor allele [Table 3.3].

Table 3.2 – Minor allele frequencies for seven identified SNPs taken into replication studies. * r^2 quality metrics for imputed SNPs using MACH 1.0 were 0.86 (rs11244450), 0.96 (rs4936003), 0.97 (rs6561505), 0.99 (rs6082335).

SNP	Allele (Major/Minor)	British 1958 Birth Cohort (n=3,853)	ALSPAC (n=1,324)	Nottingham Gedling (n=1,286)	Southampton (n=757)
rs7835920	A/C	0.25	0.27	0.25	0.25
rs7835153	A/G	0.42	0.42	0.43	0.43
rs11244450	G/A	0.09	0.08*	0.10	0.09
rs4936003	A/G	0.24	0.24*	0.23	0.26
rs6561505	G/A	0.08	0.08*	0.12	0.12
rs10514054	C/A	0.12	0.12	0.39	0.36
rs6082335	G/A	0.38	0.37*	0.25	0.25

Table 3.3 - Association with atopy phenotypes for studied SNPs: elevated specific IgE, positive SPT, and log_e serum total IgE levels. Results shown for studied SNPs identified in the discovery GWA study cohort: British 1958 Birth Cohort (B58C) and replication cohorts: ALSPAC, Nottingham Gedling, and Southampton (Soton). Pooled p values and effect sizes were calculated using available count data. Significant p values p≤0.05 are shown in bold. $^{+}$ Risk minor allele. $^{-}$ Protective minor allele.

Chr	SNP (minor allele), Gene	Elevated specific IgE		Positive SPT			Log_e serum total IgE levels				Combined		
		British 1958 Birth Cohort	Soton	ALSPAC	Nottingham Gedling	Soton	British 1958 Birth Cohort	ALSPAC	Nottingham Gedling	Soton	Elevated specific IgE	Positive SPT	Log_e serum total IgE levels
8q12	rs7835920 (C), intergenic	**$5.54x10^{(-7)}$ +**	0.30	0.35	0.06 +	0.49	**$5.79x10^{(-4)}$ +**	0.55	0.13	0.12	**$8.20x10^{(-7)}$ +**	0.11	0.24
8q22	rs7835153 (G), *ANKRD46*	**$4.50x10^{(-7)}$ -**	0.17	0.13	0.82	0.17	**$3.47x10^{(-5)}$ -**	0.45	0.82	0.38	**$7.00x10^{(-5)}$ -**	0.82	0.72
10q26	rs11244450 (A), intergenic	**$6.50x10^{(-6)}$ +**	0.33	**0.045 -**	0.73	0.12	**$2.14x10^{(-4)}$ +**	0.89	0.13	0.87	**$2.08x10^{(-4)}$ +**	0.14	0.49
11q24	rs4936003 (G), *KIRREL3*	**$4.20x10^{(-5)}$ +**	0.33	**$7.33x10^{(-4)}$ +**	0.32	0.69	0.054 -	0.48	0.86	0.63	**$4.36x10^{(-4)}$ +**	0.052 +	0.23
13q14	rs6561505 (A), *FNDC3A*	**$2.15x10^{(-9)}$ +**	0.44	0.73	0.16	0.27	**$7.75x10^{(-5)}$ +**	0.87	0.84	0.86	**$5.45x10^{(-7)}$ +**	0.31	0.34
18q22	rs10514054 (A), *NETO1*	**$2.54x10^{(-5)}$ +**	0.17	0.06 +	0.38	0.60	**$6.52x10^{(-3)}$ +**	0.13	0.19	**0.03 -**	**$1.32x10^{(-3)}$ +**	0.90	0.48
20p13	rs6082335 (A), intergenic	**$5.00x10^{(-5)}$ +**	0.10	0.80	0.97	0.31	**0.04 +**	0.96	0.34	0.43	**$2.78x10^{(-5)}$ +**	0.71	0.83

ANKRD46: ankyrin repeat domain 46; *KIRREL3*: kin of irregular chiasm-like 3; *FNDC3A*: fibronectin type III domain containing protein; *NETO1*: neuropilin tolloid-like 1.

3.4.3 *In silico* Replication

All seven identified SNPs were studied in the ALSPAC cohort, only three of which were genotyped in ALSPAC (rs7835920, rs7835153, rs10514054). None of these SNPs showed replicated association with SPT positivity [Table 3.3]. Imputed data were provided for the remaining four non-genotyped SNPs with r^2 quality metrics 0.86 (rs11244450), 0.96 (rs4936003), 0.97 (rs6561505), 0.99 (rs6082335). Association was seen with rs4936003 ($p=9.93x10^{(-4)}$) showing the same direction of effect for the minor allele and marginally with rs11244450 (p=0.045) where the opposing allele conferred risk in ALSPAC [Table 3.3].

3.4.4 Replication Genotyping

All seven identified SNPs were prospectively genotyped in the Nottingham Gedling and Southampton cohorts. None of the SNPs showed evidence of association with elevated specific IgE or with positive SPT in either cohort [Table 3.3].

3.4.5 Total IgE Analyses

Six studied SNPs also showed association with serum total IgE levels in the British 1958 Birth Cohort: rs7835920 ($p=5.79x10^{(-4)}$), rs7835153 ($p=3.47x10^{(-5)}$), rs11244450 ($p=2.14x10^{(-4)}$), rs6561505 ($p=7.75x10^{(-5)}$), rs10514054 ($p=6.52x10^{(-3)}$), and rs6082335 ($p=4.37x10^{(-2)}$) [Table 3.3]. A single SNP, rs10514054 replicated association with serum total IgE levels in the Southampton cohort (p=0.03) but showed opposing directions of effect for the minor allele: the British 1958 Birth Cohort (OR 1.14, 95% CI 1.04-1.26), the Southampton cohort (OR 0.53, 95% CI 0.30-0.94).

3.4.6 Combined Analyses

Combined analyses of count data for elevated specific IgE and positive SPT across cohorts were conducted to estimate pooled p values and effect sizes [Table 3.3]. Forest plots of the per-allele effect sizes for each of the SNPs are presented [Figures 3.3 and 3.4]. Results tables for all SNPs are listed in Appendix 3.

Pooled data showed significant associations with elevated specific IgE for all seven SNPs although these effects were driven by the initial British 1958 Birth Cohort dataset. No SNP showed significant association with positive SPT [Table 3.3]. Pooled analysis of serum total IgE levels was carried out for the Nottingham Gedling and Southampton cohorts. No SNP met statistical significance in this analysis [Table 3.3].

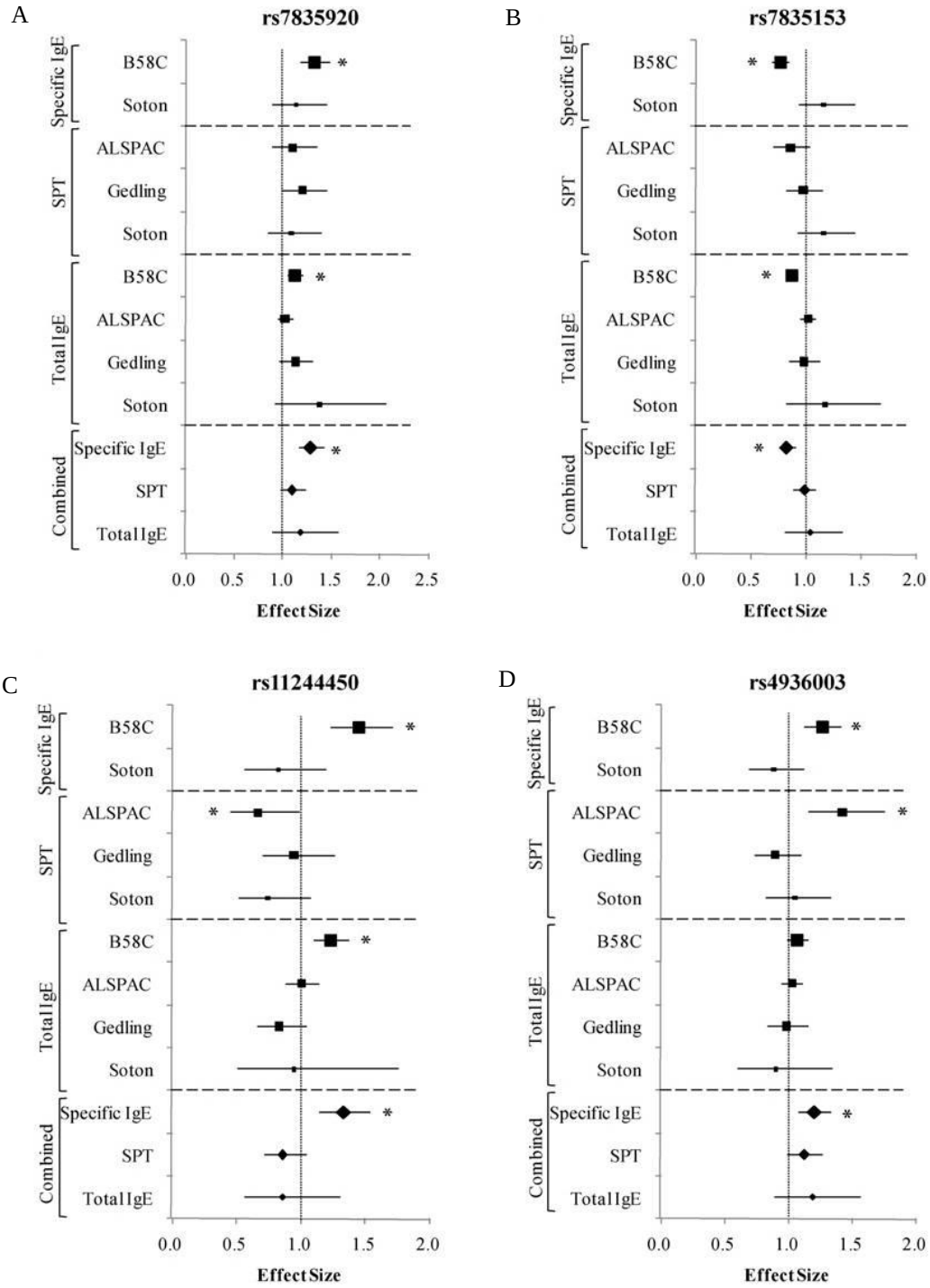


Figure 3.4 - Forest plots of studied SNPs 1 to 4: (A) rs7835920 (8q12), (B) rs7835153 (8q22), (C) rs11244450 (10q26), (D) rs4936003 (11q24). Per-allele summary effect sizes for the minor allele are shown for atopy phenotypes: elevated specific IgE (OR), positive SPT (OR), and serum total IgE levels (β). Each row represents one study, relative sizes of data points are proportional to the cohort sample size, horizontal lines represent 95% confidence intervals. Diamonds represent the meta-analyses. * denotes statistical significance ($p \leq 0.05$).

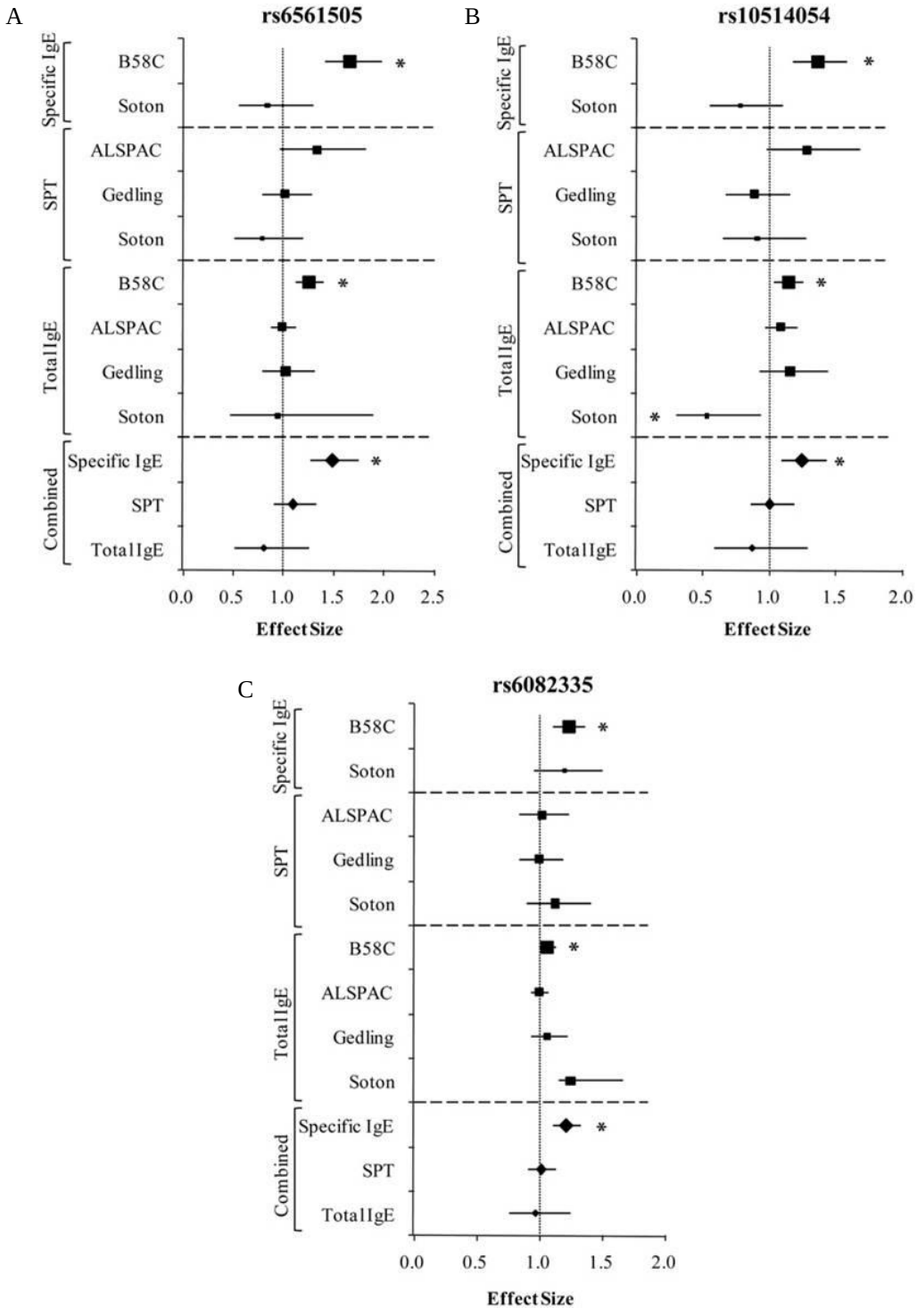


Figure 3.4 - Forest plots of studied SNPs 5 to 7: (A) rs6561505 (13q14), (B) rs10514054 (18q22), (C) rs6082335 (20p13). Per-allele summary effect sizes for the minor allele are shown for atopy phenotypes: elevated specific IgE (OR), positive SPT (OR), and serum total IgE levels (β). Each row represents one study, relative sizes of data points are proportional to the cohort sample size, horizontal lines represent 95% confidence intervals. Diamonds represent the meta-analyses. * denotes statistical significance ($p \leq 0.05$).

3.4.7 Meta-Analysis

Four datasets were available with serum total IgE levels data in addition to combined analysis on count data available only from the Nottingham Gedling and Southampton cohorts. Therefore, a meta-analysis of the per-allele effect size estimates for each individual study was carried out. Three SNPs met statistical significance as determined by the Bonferroni correction at the 5% significance level for seven SNPs ($p \leq 0.007$), but these p values were driven by the British 1958 Birth Cohort, which formed the largest contribution [Table 3.4]. This analysis also confirmed high levels of between-study heterogeneity of effect sizes.

Table 3.4 – Meta-analysis of individual per-allele study effect sizes of studied SNPs for serum total IgE weighted inversely proportionally to study variance. A meta-analysis of individual per-allele effect sizes across all studies for total IgE was carried out using STATA 10.1 with weights inversely proportional to study variance. Significant p values ($p \leq 0.007$) are shown in bold. Effect sizes presented for the minor allele.

SNP	Allele (Major/ Minor)	Study heterogeneity (p value)	Effect (95% CI)	p value
rs7835920	A/C	0.166	1.09 (1.04-1.15)	**$6.67 \times 10^{(-4)}$**
rs7835153	A/G	**0.005**	0.95 (0.91-0.99)	0.017
rs11244450	G/A	**0.010**	1.09 (1.01-1.18)	0.028
rs4936003	A/G	0.638	1.05 (0.99-1.10)	0.092
rs6561505	G/A	**0.036**	1.11 (1.03-1.21)	**0.007**
rs10514054	C/A	0.071	1.11 (1.04-1.19)	**0.003**
rs6082335	G/A	0.144	1.05 (1.01-1.10)	0.023

3.4.8 Candidate Gene Expression

Four identified SNPs are positioned within intronic regions of the following genes: rs7835153 within ankyrin repeat domain 46 (*ANKRD46*) on 8q22, rs4936003 within kin of irregular chiasm-like 3 (*KIRREL3*) on 11q24; rs6561505 within fibronectin type III domain containing protein (*FNDC3A*) on 13q14; and rs10514054 within neuropilin tolloid-like 1 (*NETO1*) on 18q22. Using NCBI to assess published transcript variants, three transcripts were published for *ANKRD46* (NM_198401), a single transcript for *KIRREL3* (NM_032531), four for *FNDC3A* (NM_001079673), and five for *NETO1* (NM_138999).

Expression profiling of these genes was carried out in two independent panels consisting of cDNA samples from human airway smooth muscle (HASM), human bronchial epithelial cells (HBEC), lung, dermal fibroblasts, human mast cell line (HMC-1), peripheral blood mononuclear cells (PBMC), brain, kidney, and placenta. Primers were designed to detect all published transcripts and are listed in Appendix 2. In the tested panels, *ANKRD46* and *FNDC3A* were expressed ubiquitously; *KIRREL3* was expressed in brain only; *NETO1* was not shown to be expressed [Figure 3.4]. Expression of *NETO1* was tested using two primer sets to ensure validity of expression assays. PCR products were sequenced and verified against published gene transcripts using NCBI blast and an example for *FNDC3A* is shown in Figure 3.9.

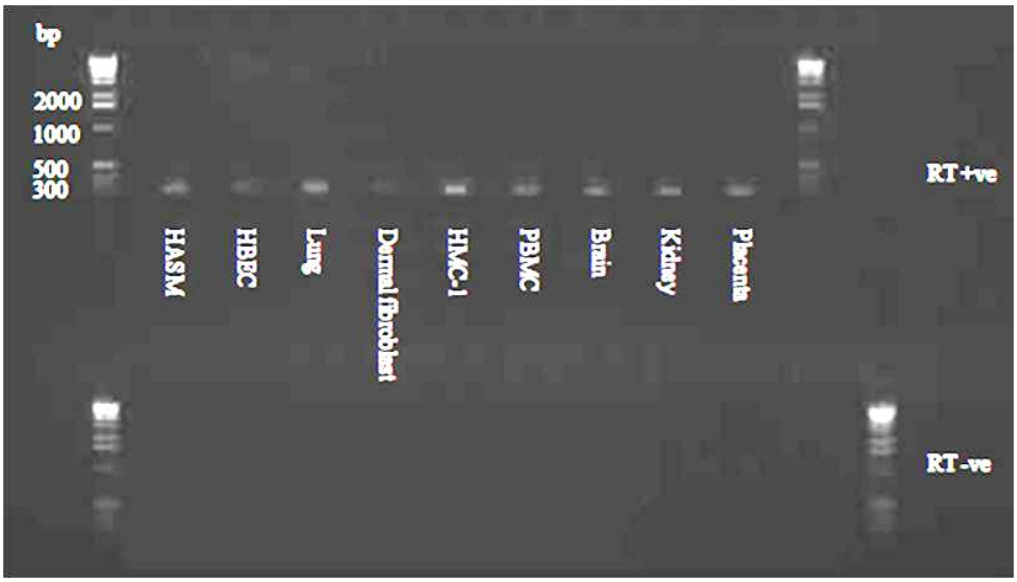


Figure 3.5 – Gel electrophoresis using 1.5% agarose under UV illumination showing ubiquitous expression of *ANKRD46* (8q22). Primers were designed to detect all three published transcripts (NM_198401). This is a representative gel from two experiments showing similar results. A product of 302bp was seen in all cell types examined. HASM denotes human airway smooth muscle, HBEC denotes human bronchial epithelial cell, HMC-1 denotes human mast cell line, and PBMC denotes peripheral blood mononuclear cells. 1kb DNA ladders are shown on the left and right extremes, RT -ve and water controls are shown in the second row.

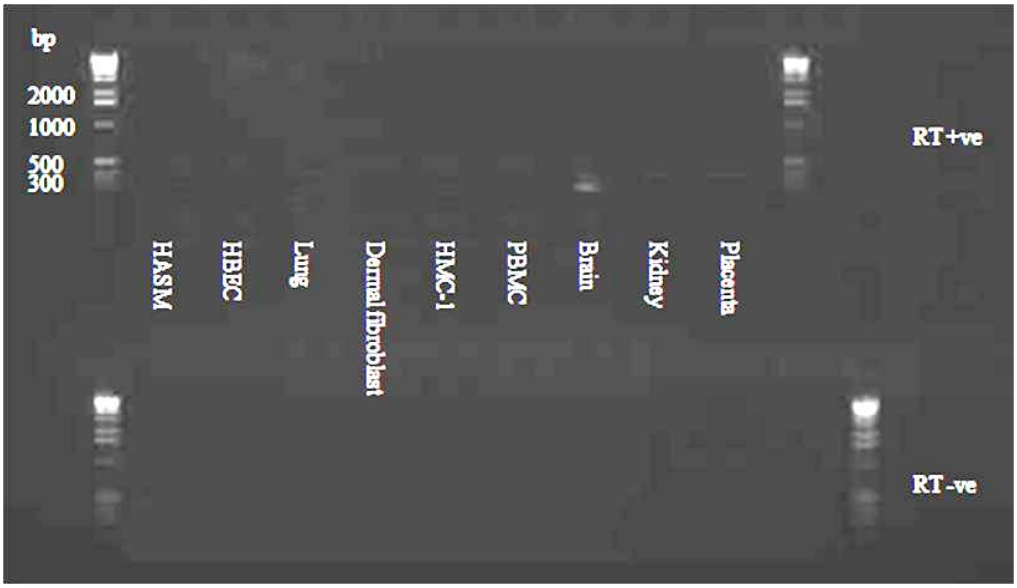


Figure 3.6 – Gel electrophoresis using 1.5% agarose under UV illumination showing expression in brain of *KIRREL3* (11q24). Primers were designed to detect the single published transcript (NM_032531). This is a representative gel from two experiments showing similar results. A product of 300bp was seen in brain. HASM denotes human airway smooth muscle, HBEC denotes human bronchial epithelial cell, HMC-1 denotes human mast cell line, and PBMC denotes peripheral blood mononuclear cells. 1kb DNA ladders are shown on the left and right extremes, RT -ve and water controls are shown in the second row.

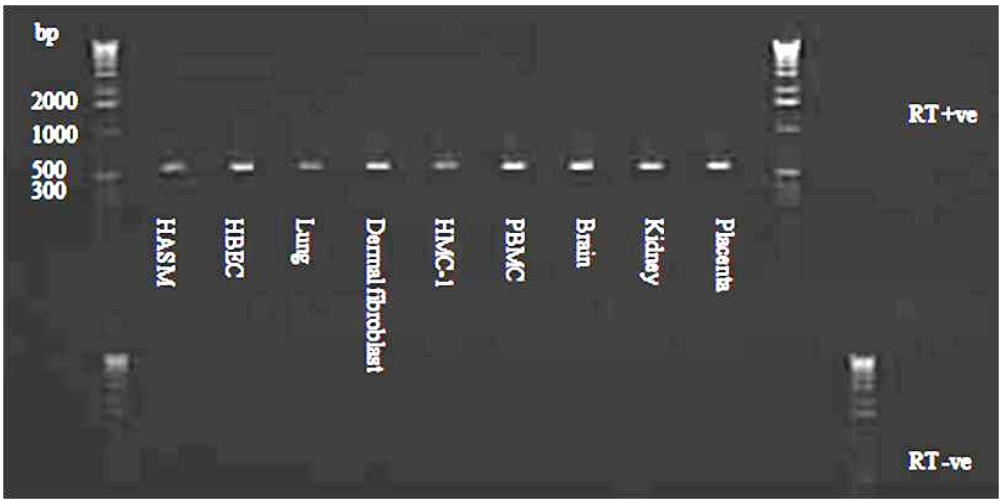


Figure 3.7 – Gel electrophoresis using 1.5% agarose under UV illumination showing ubiquitous expression of *FNDC3A* (13q14). Primers were designed to detect all four published transcripts (NM_001079673). This is a representative gel from two experiments showing similar results. A product of 563bp was seen in all cell types examined. HASM denotes human airway smooth muscle, HBEC denotes human bronchial epithelial cell, HMC-1 denotes human mast cell line, and PBMC denotes peripheral blood mononuclear cells. 1kb DNA ladders are shown on the left and right extremes, RT -ve and water controls are shown in the second row.

Figure 3.8 – Gel electrophoresis using 1.5% agarose under UV illumination showing absence of expression of *NETO1* (18q22). Primers were designed to detect all five published transcripts (NM_001079673). Expression of *NETO1* was tested using two primer sets to ensure validity of expression assays. This is a representative gel from two experiments showing similar results. Products of 366bp (assay 1) and 203bp (assay 2) were not seen in any cell types examined. HASM denotes human airway smooth muscle, HBEC denotes human bronchial epithelial cell, HMC-1 denotes human mast cell line, and PBMC denotes peripheral blood mononuclear cells. 1kb DNA ladders are shown on the left and right extremes, RT -ve and water controls are shown in the second row.

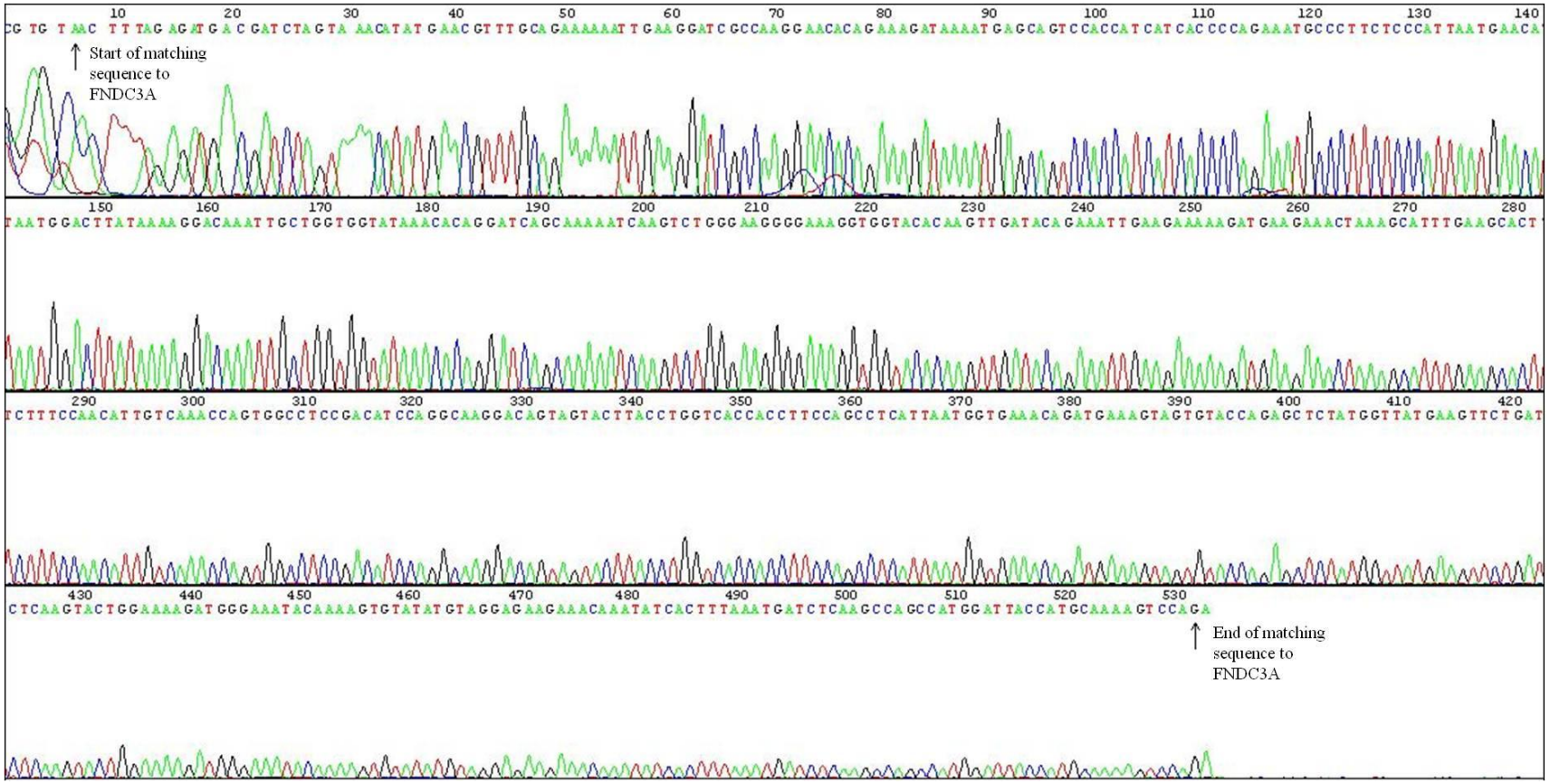


Figure 3.9 – Example of sequencing results for a 563bp standard PCR product for *FNDC3A* in human airway smooth muscle (HASM) obtained from excision from the gel shown in Figure 3.7. All products seen in cell types examined were excised and purified for sequencing. Matching sequence to *FNDC3A* (NM_001079673) using BLAST (NCBI).

3.4.9 Candidate Gene Look-up

Candidate genes previously identified in a smaller dataset of the British 1958 Birth Cohort were identified and association results for SNPs within a list of widely replicated candidate genes from the literature were evaluated in the studied British 1958 Birth Cohort [Table 3.5]. Association with elevated specific IgE was seen with SNPs within previously identified atopy genes including *FCER1A* (5 SNPs, lowest p=0.002), *IL13* (2 SNPs, lowest p=0.02), and signal transducer and activator of transcription 6, interleukin-4 induced (*STAT6*) (6 SNPs, lowest p=0.003). No SNPs within *CD14* (5q31.1) and *HLA-DRB1* (6p21) were genotyped in the current studied British 1958 Birth Cohort dataset. Therefore, these genes could not be formally evaluated.

Table 3.5 – Evaluation within the British 1958 Birth Cohort of previously reported atopy candidate genes. *Note no SNPs were able to be tested within *CD14* and *HLA-DRB1*.

Chromosome	Gene	Number of SNPs tested within gene	Highest significance SNP					
			SNP	Function	Elevated specific IgE		Serum total IgE levels	
					p value	OR (95% CI)	p value	OR (95% CI)
1q23	*FCER1A*	5	rs2251746	Intronic	**$2.14x10^{(-3)}$**	0.84 (0.75-0.94)	**$1.04x10^{(-8)}$**	0.82 (0.76-0.88)
1q31-q32	*NOS1*	45	rs9658535	Intronic	0.19	1.77 (0.76-4.16)	0.31	1.36 (0.75-2.44)
1q31-q32	*IL10*	8	rs1800896	5' UTR	0.27	0.95 (0.86-1.04)	0.46	0.98 (0.92-1.04)
2q33	*CD28*	9	rs10490573	Intronic	0.058	1.13 (1.00-1.28)	0.24	1.05 (0.97-1.14)
2q33	*CTLA4*	1	rs733618	5' UTR	0.61	1.05 (0.84-1.26)	0.99	1.00 (0.89-1.13)
5q31	*IL13*	2	rs20541	Q [Gln] ⇒ R [Arg]	**0.020**	1.16 (1.02-1.32)	**$5.97x10^{(-6)}$**	1.21 (1.11-1.31)
5q31.1	*IL4*	3	rs2243248	5' UTR	**0.020**	0.78 (0.63-0.96)	0.49	0.96 (0.84-1.08)
5q31.1	*CD14*	*0	-	-	-	-	-	-
5q32	*SPINK5*	16	rs2303064	D [Asp] ⇒ N [Asn]	0.23	1.10 (0.94-1.28)	0.29	1.05 (0.95-1.17)
6p21	*HLA-DQB1*	1	rs1063355	3' UTR	0.57	0.97 (0.88-1.08)	0.63	0.98 (0.92-1.05)
6p21	*HLA-DRB1*	*0	-	-	-	-	-	-
11q13	*MS4A6E*	4	rs2165832	Intronic	**0.029**	0.89 (0.80-0.99)	0.81	0.99 (0.93-1.06)
11q13	*MS4A2*	1	rs10750935	5' UTR	0.59	1.03 (0.93-1.14)	0.68	0.99 (0.92-1.05)
12q13	*STAT6*	6	rs1059513	3' UTR	**0.003**	0.78 (0.66-0.92)	**$1.95x10^{(-6)}$**	0.79 (0.71-0.87)
16p12.1-p12.2	*IL4R*	18	rs1805011	E [Glu] ⇒ A [Ala]	**0.035**	0.84 (0.72-0.99)	**$7.63x10^{(-3)}$**	0.88 (0.80-0.97)

FCER1A: Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide; *NOS1*: nitric oxide synthase 1 (neuronal); *IL10*: interleukin 10; *CD28*: CD28 molecule; *CTLA4*: cytotoxic T-lymphocyte-associated protein 4; *IL13*: interleukin 13; *IL4*: interleukin 4; *CD14*: CD14 molecule; *SPINK5*: serine peptidase inhibitor, Kazal type 5; *MS4A6E*: membrane-spanning 4-domains, subfamily A, member 6E; *MS4A2*: membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide); *STAT6*: signal transducer and activator of transcription 6, interleukin-4 induced; *IL4R*: interleukin 4 receptor.

3.5 Discussion

Results presented in this Chapter describe a large-scale population based GWA study for elevated specific IgE to three common allergens in 3,853 subjects from the British 1958 Birth Cohort with follow-up investigations in 7,423 individuals from three additional cohorts using atopy defined by elevated specific IgE and/or positive SPT. Overall, no replicated evidence for any of the atopy susceptibility loci identified in the discovery cohort using elevated specific IgE to define atopy was found.

The Q-Q plot for all SNPs in the British 1958 Birth Cohort specific IgE dataset gives an indication of the likelihood of identifying susceptibility loci. Deviation of plotted results from the predicted distribution of p values in the absence of genetic effects is seen for a small number of SNPs predominantly at the extreme p values. However, in addition there was also a large number of SNPs with small effect sizes above the line, indicating probable existence of numerous regions containing SNPs with small effects. This suggests that genetic influences on atopy defined using elevated specific IgE do not come from a single or even a few major genes with large effect sizes but rather a larger number of loci with modest effects.

Seven potential loci were identified in the initial GWA scan for susceptibility to atopy, all of which are in regions that have previously reported linkage to IgE levels and the development of various atopic traits: 8q12 (Gusareva *et al.* 2009), 8q22 (Dizier *et al.* 2000; Ciullo *et al.* 2006), 10q26 (Tremblay *et al.* 2008), 11q24 (Anderson *et al.* 2002), 13q14 (Chen *et al.* 2006), 18q22 (Van Eerdewegh *et al.* 2002), and 20p13 (Jang *et al.* 2005; Lesueur *et al.* 2007). Chromosome 13q14 is a region supported by both previous linkage and association results in which a number of studies have been conducted to examine candidate genes including plant homeodomain finger protein 11 (*PHF11*) and cysteinyl leukotriene type 2 receptor (*CYSLTR2*) with potentially functional roles in both atopy and asthma (Fukai *et al.* 2004; Sanak *et al.* 2007).

Identified SNPs on 8q22, 11q24, 13q14, and 18q22 are intronic in *ANKRD46*, *KIRREL3*, *FNDC3A*, and *NETO1* respectively. Few studies have been published regarding these genes

and their functions remain unclear, but some suggestions have been proposed for their functional roles. *ANKRD46* has been proposed to participate in phosphorylation during glycolysis (Shin *et al.* 2007). *KIRREL3* has been shown to have a functional role in PDZ domain binding, a domain found in diverse signalling proteins (Bhalla *et al.* 2008). *FNDC3A* has been proposed to be involved in extracellular matrix (ECM)-receptor interactions (Obholz *et al.* 2006). *NETO1* encodes a predicted transmembrane protein which contains a specific intracellular motif, shown in other proteins to be a strong promoter of endocytosis (Stöhr *et al.* 2002).

The pivotal SNP on 13q14 (rs6561505, $p=2.15x10^{(-9)}$) reached genome-wide significance for increased risk of atopy (OR 1.67, 95% CI 1.41-1.98) and is located in *FNDC3A*, a gene not previously implicated in atopy or related traits, approximately 300kb in distance from both *CYSLTR2* upstream and *PHF11* downstream. In order to replicate these findings from the GWA study, a combination of *in silico* replication, prospective genotyping, and combined analyses was used. This SNP was not associated with atopy defined by elevated specific IgE or positive SPT in the additional three cohorts (1,007 cases and 2,153 controls in total). The minor allele for this SNP also increased the risk of raised serum total IgE in the British 1958 Birth Cohort but not in the additional cohorts. Pooled results for elevated specific IgE showed a significant combined effect of OR 1.49 (95% CI 1.28-1.75) for the risk allele.

Whilst this study focused on the identification of signals reaching conventional GWA significance, a more inclusive region-based strategy was also implemented involving prospective genotyping of an additional six SNPs within loci showing at least some evidence of association. For the majority of SNPs genotyped, there was no evidence for association in the additional cohorts. Despite significant associations for elevated specific IgE for rs7835153 (8q22), rs11244450 (10q26), and rs10514054 (18q22) in the British 1958 Birth Cohort, there was little evidence of consistent association in the additional cohorts for these SNPs. Overall it seems unlikely that these three loci confer markedly increased susceptibility to atopy in these populations.

A single SNP, rs4936003 (11q24) showed replicated association with positive SPT in the ALSPAC cohort ($p=7.33x10^{(-4)}$) with the same direction of effect for the minor allele (OR 1.43, 95% CI 1.16-1.76) as in the British 1958 Birth Cohort (OR 1.27, 95% CI 1.13-1.43). This is the only SNP which showed association with atopy across two independent cohorts, using subjects of different ages and definitions of atopy. As this SNP, intronic of the *KIRREL3* gene, did not show evidence of association with total IgE, this indicates that it may play a role in sensitisation to allergens different from influencing IgE production *per se*. In the limited profiling panel used, expression of this gene was only detected in the brain template. Interestingly, one study has shown the involvement of the gene product mKirre in the support of hematopoietic stem cells by stromal cells (Ueno *et al.* 2003).

In addition, some suggestive evidence for association has been identified for two intergenic SNPs, rs7835920 (8q12) and rs6082335 (20p13). For these SNPs, although not all replication results are statistically significant, a consistent effect can be seen for the risk allele throughout. The results for these two SNPs suggest that there may be genes underlying atopy susceptibility in these regions. Combined analyses for elevated specific IgE showed effect sizes of OR 1.17, 95% CI 1.29-1.43 ($p=8.20x10^{(-7)}$) and OR 1.22, 95% CI 1.11-1.34 ($p=2.78x10^{(-5)}$) respectively although these were likely to have been driven by the initial signal in the British 1958 Birth Cohort.

Since four datasets were available with total IgE measures, a meta-analysis of the individual study per-allele effect estimates was conducted in addition to combined analysis on available raw data from the Nottingham Gedling and Southampton cohorts only. Three SNPs met nominal significance ($p\leq0.01$): rs7835920 (8q12, $p=6.67x10^{(-4)}$), rs6561505 (13q14, p=0.007), and rs10514054 (18q22, p=0.003) although as anticipated, these p values were driven by the British 1958 Birth Cohort which formed the largest contribution. Significant levels of study heterogeneity of effect sizes were identified for three SNPs ($0.005\leq p\leq0.036$). This indicates that difficulty in replicating significant results for these SNPs may have been influenced by differences in the way each study was conducted.

Potential association for SNPs in candidate genes previously identified in a smaller dataset of the British 1958 Birth Cohort were also examined (Maier *et al.* 2006). In the British 1958 Birth cohort, association with elevated specific IgE was seen with SNPs within previously identified atopy genes including *FCER1A* (five SNPs: lowest p=0.002, OR 0.84), *IL13* (two SNPs: lowest p=0.02, OR 1.16), and *STAT6* (six SNPs: lowest p=0.003, OR 0.78) (Heinzmann *et al.* 2000; Schedel *et al.* 2004; Sanak *et al.* 2007). Whilst there was some evidence of association for both specific IgE and total IgE in the current study, none of these signals met conventional GWAS significance criteria in the larger dataset. Genes around the *HLA* locus was unable to be formally examined due to limitations in SNP coverage on the Illumina HumanHap550 Array. This cohort has demonstrated replication of previously identified atopy genes with more modest p values and effect sizes.

A number of factors may have contributed to difficulty in replication. Firstly, 'winner's curse bias' may have resulted in inflated effect size estimates obtained from the discovery cohort, requiring much larger replication studies in order to have sufficient power to detect unbiased effect sizes. Numbers of study subjects used in this study may have been unlikely to yield enough statistical power to detect variants with modest effects particularly in the presence of substantial allelic heterogeneity (Wright *et al.* 1999). Despite efforts to increase total sample sizes through combined analyses, there was still inadequate power overall, which was further reduced as a result of study heterogeneity. Using a two-stage replication design, a total of more than 16,000 individuals was required in replication cohorts in order to have 80% power to detect risk allele odds ratios of greater than 1.2.

Secondly, heterogeneity across study populations in both phenotype definitions and variable coverage of SNPs across the genome in genotyping arrays used may have complicated analyses. Not all of the key SNPs studied were genotyped in the ALSPAC cohort requiring the use of imputed data. Although the r^2 quality metrics for these SNPs were relatively high ranging from 0.86 (rs11244450) up to 0.99 (rs6082335), this was a further limitation during replication.

Many SNPs showing association with elevated specific IgE in the discovery GWA study were also associated with serum total IgE levels in the British 1958 Birth cohort. In addition, measured elevated specific IgE and positive SPT results for the same subjects in the Southampton cohort correlated well (Pearson's r^2=0.73, p<0.001). However, the relationship between these phenotypes is unclear and it is likely that the use of different atopy markers across cohorts may have contributed to difficulty in replication due to different genetic contributions to these related phenotypes. Despite some overlap in SNP association, loci identified for atopy defined using specific IgE and positive SPT in this study appear to be unique compared to those for increased serum total IgE levels suggesting that the processes controlling sensitisation may differ from those controlling level of IgE production. Carrying out a GWA study of total IgE within the British 1958 Birth Cohort using this dataset may have helped to provide further insight into this question, but this was not possible as related analyses were already included as part of the GABRIEL study (Moffatt *et al.* 2010).

Previous research using subjects from the ALSPAC cohort has shown that 70% of all atopic children were identified from their reactions to the three common allergens alone i.e. house dust mite (Der p 1), mixed grass pollen and cat fur (Perkin *et al.* 2006). However, there is still some risk of phenotype misclassification in which atopic individuals will be considered as controls in the absence of a more extensive panel. In addition, variations in the antigen panels used across cohorts may have complicated replication as well as differences in population demographics. The British 1958 Birth Cohort, Nottingham Gedling, and Southampton cohorts are adult populations whereas subjects in the ALSPAC cohort had a mean age of 7.5 years at the time of atopy data collection.

The main conclusion that can be drawn from this GWA study is that overall there is evidence for a large number of genetic factors influencing atopy risk, each with a small contribution. Therefore, the detection of further susceptibility loci will require GWA studies with more complete genome SNP coverage and using larger sample sizes together with the combination of results across multiple studies. These results have significant implications for the design of reliable studies to identify genes regulating elevated specific IgE and contributing to atopy.

The results from this Chapter have highlighted the importance for well designed and large GWA studies to address these issues.

Chapter 4

Assessing Atopy Susceptibility within Previously Identified Linkage Loci on Chromosomes 3 and 17

4.1 Introduction

Prior to the discovery and widespread use of genome-wide association (GWA) approaches, considerable efforts had been made to map the chromosomal locations of susceptibility genes for asthma and atopy through the use of linkage in families. Over the past decade, numerous genome-wide linkage (GWL) scans in multiple populations have identified more than 20 distinct susceptibility regions including chromosomes 2q, 5q, 6p, 11q, and 13q as well as positionally cloned genes such as *ADAM33*, *GRPA*, *PHF11*, *DPP10*, and *HLA-G* (Ober 05, Bouzigon 10) (see Chapter 1, Section 1.3.2).

However, inconsistency in study designs including phenotype definitions, family ascertainment schemes, and analytical methods have resulted in a lack of replication and consistent evidence for individual loci. Previous work performed by colleagues in Nottingham completed a meta-analysis of data from GWL studies of asthma and related traits: bronchial hyper-responsiveness (BHR), allergen positive skin prick test (SPT), and serum total Immunoglobulin E (IgE) (Denham *et al.* 2008) (see Chapter 1: Section 1.3.3). This study aimed to increase statistical power and reduce the potential for type 1 and type 2 errors in identifying novel candidate genes as well as infer reasons for the inconsistency and lack of reproducibility of findings.

The rank based genome-scan meta-analysis (GSMA) method was used to combine linkage data from nine Caucasian asthma proband populations. The close inter-relation between asthma and atopy suggested by co-occurence of these phenotypes has allowed the study of atopic phenotypes in families ascertained through a proband with asthma. Significant evidence for linkage was shown for chromosomal regions 3p22.1-q22.1 and 17p12-q24.3 suggesting that

they may contain susceptibility loci for positive SPT at least in asthma enriched families (1,093 pedigrees, n=4,746) [Figure 4.1]. Results from this study supported the contribution of regions that contain previously identified asthma susceptibility genes to the risk of developing allergen SPT responses and by inference atopy.

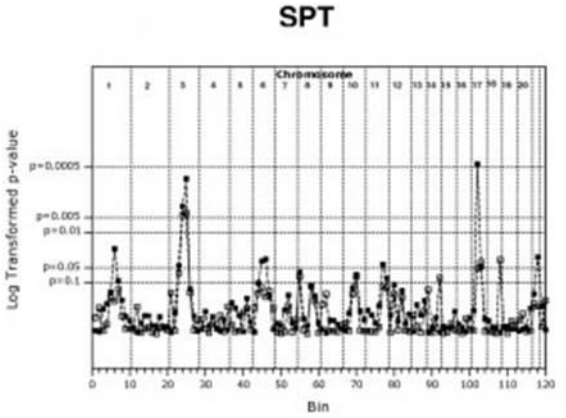


Figure 4.1 - Weighted GSMA results for positive SPT response. This plot taken from the GSMA shows significant evidence for linkage to positive SPT response on chromosomes 3p22.1-q22.1 and 17p12-q24.3. Taken from (Denham *et al.* 2008).

The identified regions were large, spanning 70.3Mb and 28.5Mb on chromosomes 3p22.1-q22.1 and 17p12-q24.3 respectively, each harbouring a large number of potential candidate genes. The genotype data generated from the GWA study described in Chapter 3 were able to be used to test association of SNPs within these regions in order to carry out fine mapping to try and localise causative loci and variants within genes responsible for the possible linkage signals seen.

Although these regions were not identified for significant association and did not feature in the work derived in Chapter 3, this provides an alternative, complimentary approach to identifying susceptibility genes. The results from the GWA study indicate that there are likely to be multiple loci with modest effects in determining susceptibility to asthma. This focused approach may yield more statistical power as it limits the regions studied and thus the level of stringency needed when correcting for multiple testing.

4.2 Aim

The aim of this Chapter was to further define previously identified atopy susceptibility loci on chromosomes 3 and 17 using data from a GWA study in the British 1958 Birth Cohort.

4.3 Methods

4.3.1 Study Populations

As in Chapter 3, primary analyses in Chapter 4 were carried out in the British 1958 Birth Cohort (1,083 cases and 2,770 controls), identified SNPs of interests were assessed using *in silico* replication in the ALSPAC cohort (1,324 individuals) as well as prospective genotyping in the Nottingham Gedling (1,286 individuals) and Southampton cohorts (757 individuals) (see Chapter 3: Section 3.1.1 for details).

The work presented in Chapter 4 was commenced using data from a subset of 388 cases and 1,040 controls from the British 1958 Birth Cohort prior to the complete dataset becoming available. This subset of individuals was also genotyped using the Affymetrix Human Mapping 500K SNP platform. In addition, the deCODE Icelandic cohort was used in Chapter 4 during *in silico* replication. This cohort consisted of 908 cases and 27,150 general population controls genotyped using the Illumina HumanHap300 SNP platform. Details of study recruitment and phenotyping in all cohorts are described in Chapter 2: Section 2.1.1. Genotyping and quality control procedures undertaken are summarised in Chapter 2: Section 2.2.1.1.

Atopy was defined using positive SPT ≥3mm or ≥50% of the histamine control in the deCODE Icelandic cohort. A sub-group of 239 non-asthmatic, atopic cases were also identified. No serum total IgE level data was available for analysis in this cohort.

4.3.2 Association Testing

Details of summary data generation in the British 1958 Birth Cohort and the ALSPAC cohort, methods for association testing following prospective genotyping in the Nottingham Gedling and Southampton cohorts as well as in pooled analyses are described in Chapter 3: Section 3.3.2. In Chapter 4, the subset dataset analysed in the British 1958 Birth Cohort (1,468 individuals) will be referred to as 'B58C specific IgE D1'; the complete dataset (3,853 individuals) will be referred to as 'B58C specific IgE D2'. Allele counts for SNPs undergoing

replication analyses were provided from a previous GWA study in the deCODE Icelandic cohort. Association tests for positive SPT in total cases and the sub-group of non-asthmatic cases against general population controls were carried out using Pearson's Chi-squared (χ^2) per-allele tests with one degree of freedom (df) comparing allele frequencies (see Chapter 2: Section 2.4.2). The flow of analyses carried out is summarised in Figure 4.2.

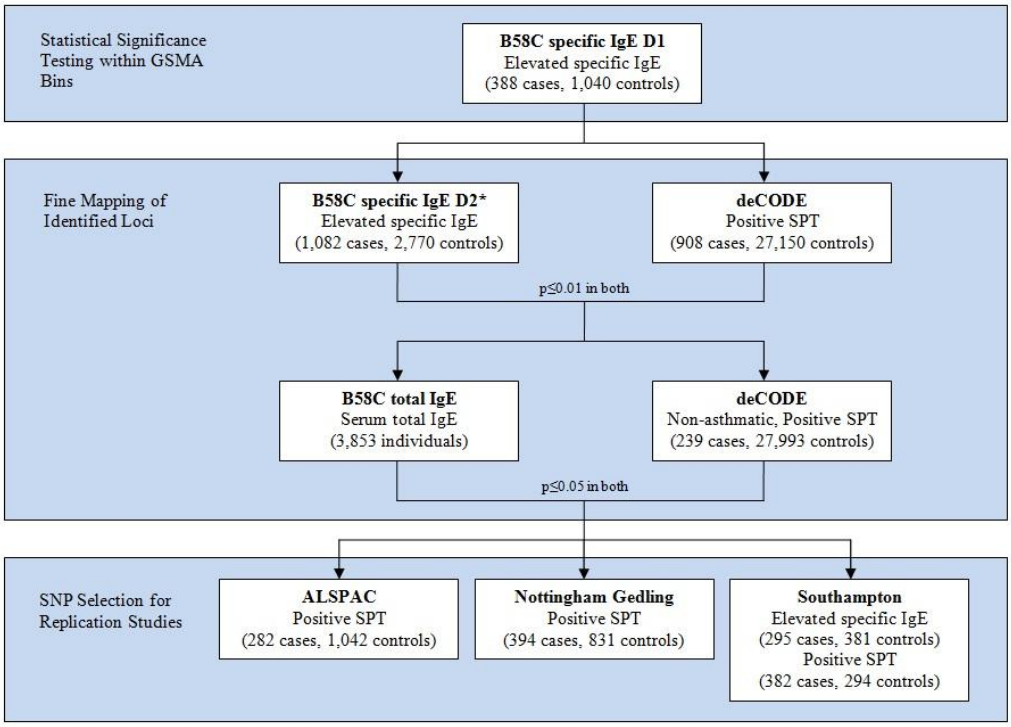


Figure 4.2 – Summary flow diagram of analyses to assess atopy susceptibility within previously identified linkage loci on Chromosomes 3 and 17. * Note: individuals included in B58C specific IgE D2 encompasses those in B58C specific IgE D1.

4.3.3 Statistical Significance Testing within GSMA Bins

Results of GWA testing using B58C specific IgE D1 (1,468 individuals) were provided for SNPs within GSMA reported susceptibility bins: 3p22.1-q22.1 (38,845kb to 134,474kb), 17p12-q24 (11,325kb to 66,600kb). These were grouped according to the genotyping platform used (Illumina HumanHap550 or Affymetrix Human Mapping 500K). Statistical significance was assessed using nominally defined $p \leq 0.001$ as well as Bonferroni corrected p values at the 5% significance level: $p \leq 3.20 \times 10^{(-6)}$ for 15,640 SNPs on 3p22.1-q22.1 and $p \leq 5.58 \times 10^{(-6)}$ for 8,543 SNPs on 17p12-q24 tested on the Illumina platform; $p \leq 3.55 \times 10^{(-6)}$ for 14,083 SNPs on

3p22.1-q22.1 and $p \leq 7.00 \times 10^{(-6)}$ for 7,146 SNPs on 17p12-q24 tested on the Affymetrix platform.

Susceptibility loci were identified based on clustering of statistically significant SNPs, consistency of results across genotyping platforms, the surrounding linkage disequilibrium (LD), and relationship with potential candidate genes in the same region. Intervals shown to be ≤500kb apart were combined and distinct loci taken forward for fine mapping.

4.3.4 Fine Mapping of Identified Loci

Subsequently, summary data from the complete GWA dataset in the British 1958 Birth Cohort i.e. B58C specific IgE D2 (3,853 individuals) and the deCODE Icelandic cohort became available. Distinct loci identified using the method described above using B58C specific IgE D1 were no longer pursued if they did not contain any associated SNPs ($p \leq 0.01$) within both the B58C specific IgE D2 and deCODE Icelandic cohort datasets. In order to identify potential atopy-specific loci, SNPs within remaining loci were tested for association ($p \leq 0.05$) with serum total IgE levels in the British 1958 Birth Cohort within the complete dataset (3,853 individuals) and non-asthmatic, atopic cases (239 individuals) in the deCODE Icelandic cohort. All SNPs meeting p value thresholds were assessed to ensure that the same direction of effect was shown across cohorts. Atopy-specific loci were taken forward for replication analyses.

4.3.5 SNP Selection for Replication Studies

Replication was tested for SNPs showing the highest statistical significance within each atopy specific locus as well as consistent association in both datasets from the British 1958 Birth Cohort. Quality control was assessed for minor allele frequency (MAF) >5% and in Hardy-Weinberg equilibrium (HWE $p>0.05$). Positive replication was assessed at the 5% significance level.

4.3.6 Candidate Gene Expression

Expression profiling of possible candidate genes identified by statistically significant SNP associations was carried out using polymerase chain reaction (PCR) in two independent cDNA

panels with two repeats each. Assays were designed to detect all published gene transcripts documented online using NCBI. PCR products were sequence verified (see Chapter 2: Section 2.5).

4.4 Results

4.4.1 Association Testing of GSMA Bins in the British 1958 Birth Cohort

No SNPs met Bonferroni corrected threshold for significance [Figure 4.3]. A total of 40 SNPs on 3p22.1-q22.1 and 22 SNPs on 17p12-q24 met criteria for nominal significance ($p \leq 0.001$). All SNPs were plotted on positional maps showing LD blocks and all candidate genes in the region (Appendix 4). A total of 14 groups of SNPs were identified on chromosome 3 and 9 on chromosome 17, combining interval regions ≤500kb apart resulted in 11 distinct loci on chromosome 3 and 8 distinct loci on chromosome 17 [Table 4.1].

Table 4.1 – Loci identified for fine mapping by results of all SNPs tested for association in the B58C specific IgE D1 within GSMA identified susceptibility bins on chromosomes 3p22.1-q22.1 and 17p12-q24.3. Loci identified for fine mapping are labelled A to K (11 loci and 8 candidate genes) on chromosome 3 and A to H (8 loci and 6 candidate genes) on chromosome 17 [Figure 4.3].

Chromosome	Loci (kb)		Size (kb)	Gene(s)
3	A	45405-45580	175	*LARS2*
	B	54560-54600	40	*CACNA2D3*
	C	56930-57050	120	-
	D	61415-61510	95	*PTPRG*
	E	64115-64710	595	-
	F	66390-66910	520	*SLC25A26, LRIG1*
	G	87070-87355	285	-
	H	97725-98160	435	*EPHA6*
	I	114230-114550	320	*WDR52*
	J	118625-118725	100	-
	K	120090-120210	120	*IGSF11*
17	A	13080-13105	25	-
	B	13913-14060	147	*COX10*
	C	39025-39080	55	-
	D	39990-40415	425	*ADAM11, EFTUD2, LOC146909*
	E	47965-48050	85	-
	F	50100-50145	45	-
	G	53200-53220	20	-
	H	54650-55280	630	*GDPD1, VMP1*

LARS2: leucyl-tRNA synthetase 2, mitochondrial; *CACNA2D3*: calcium channel, voltage-dependent, alpha 2/delta subunit 3; *PTPRG*: protein tyrosine phosphatase, receptor type, G; *SLC25A26*: solute carrier family 25, member 26; *LRIG1*: leucine-rich repeats and immunoglobulin-like domains 1; *EPHA6*: EPH receptor A6; *WDR52*: WD repeat domain 52; *IGSF11*: immunoglobulin superfamily, member 11; *COX10*: COX10 homolog, cytochrome c oxidase assembly protein, heme A: farnesyltransferase (yeast); *ADAM11*: ADAM metallopeptidase domain 11; EFTUD2: elongation factor Tu GTP binding domain containing 2; *GDPD1*: glycerophosphodiester phosphodiesterase domain containing 1; *VMP1*: vacuole membrane protein 1.

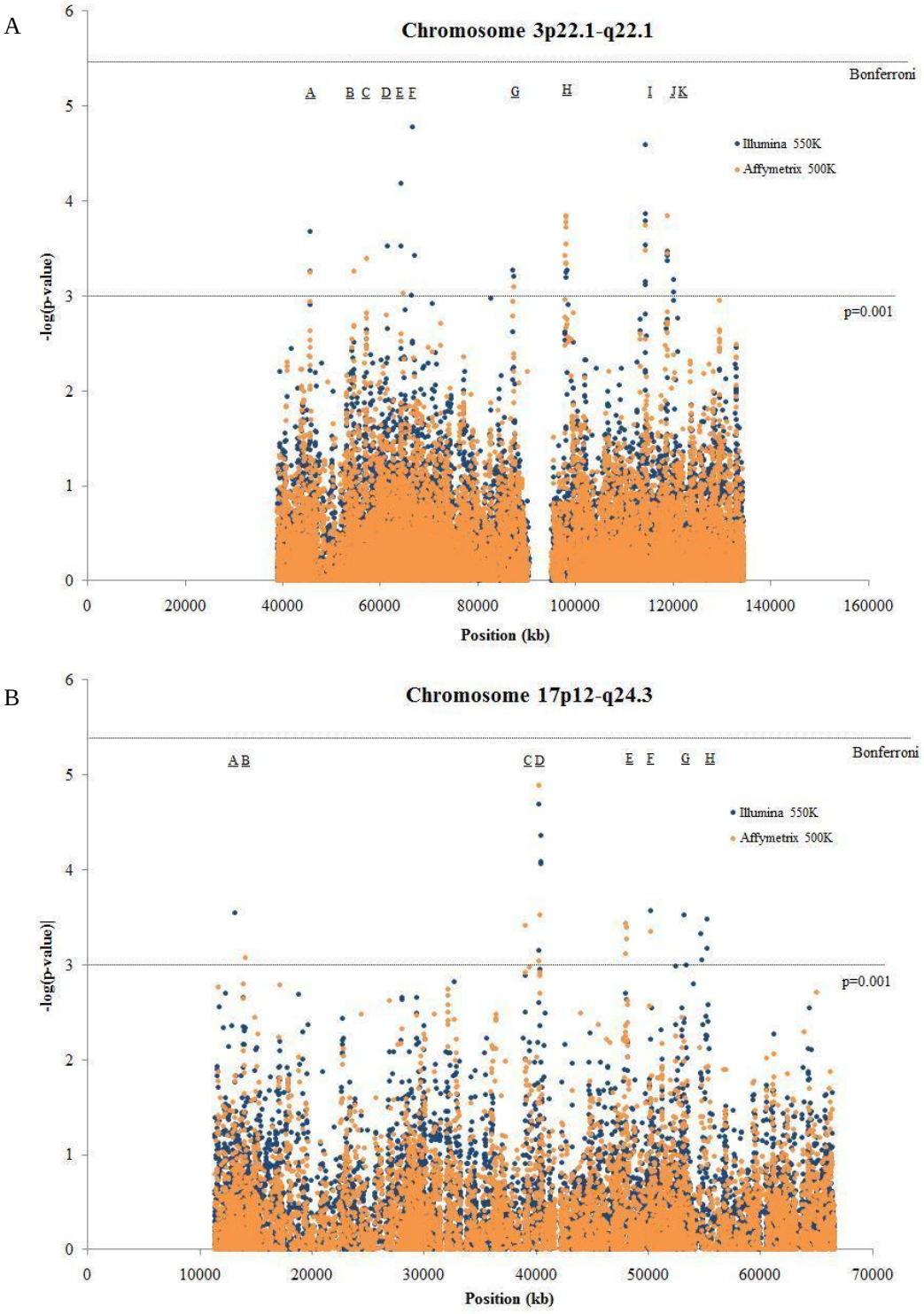


Figure 4.3 – Results of all SNPs tested for association in the B58C specific IgE D1 within GSMA identified susceptibility bins on chromosomes 3p22.1-q22.1 and 17p12-q24.3. SNPs genotyped using the Illumina 550K and Affymetrix 500K platforms are shown in blue and orange respectively. The Bonferroni (chromosome and platform specific) and nominal significance thresholds (p=0.001) are marked. Loci identified for fine mapping are labelled A) A to K on chromosome 3 and B) A to H on chromosome 17 [Table 4.1].

4.4.2 Fine Mapping of Identified Loci

The complete dataset from the British 1958 Birth Cohort became available. Therefore, SNP results within the original GSMA set susceptibility bins were assessed using B58C specific IgE D2. Results indicate that the physical positions of identified susceptibility loci had shifted following the inclusion of more individuals in analyses [Figure 4.4]. Four loci (D, E, F, G) on chromosome 3 and five loci on chromosome 17 (B, C, D, G, H) identified using the subset B58C specific IgE D1 dataset remained significant within the complete B58C specific IgE D2 dataset. Summary data from association tests with atopy in the deCODE Icelandic cohort were provided for SNPs within these loci. Only three loci (D, F, G) on chromosome 3 and a single locus (D) on chromosome 17 contained SNPs associated with positive SPT in the deCODE Icelandic cohort. On chromosome 3, these were rs2886524 ($p=6.60x10^{(-4)}$) within locus D, rs7628307 (p=0.008) and rs9845175 (p=0.002) within locus F, rs1002765 (p=0.002) and rs6768771 (p=0.003) within locus G. On chromosome 17, a single SNP was identified within locus D: rs1126642 (p=0.007).

Only locus D on chromosome 17 contained SNPs showing association with both serum total IgE levels in the British 1958 Birth Cohort (eight out of 78 tested SNPs: $0.010 \leq p \leq 0.045$) and non-asthmatic atopy in the deCODE Icelandic Population cohort (one out of three tested SNPs: p=0.018 for rs1002765). The highest significance SNP, rs7212001 within this atopy specific region showing consistent association in both British 1958 Birth Cohort datasets ($p=4.31x10^{(-5)}$ for specific IgE in D1, p=0.016 for specific IgE in D2, p=0.024 for total IgE) was taken forward for replication analyses.

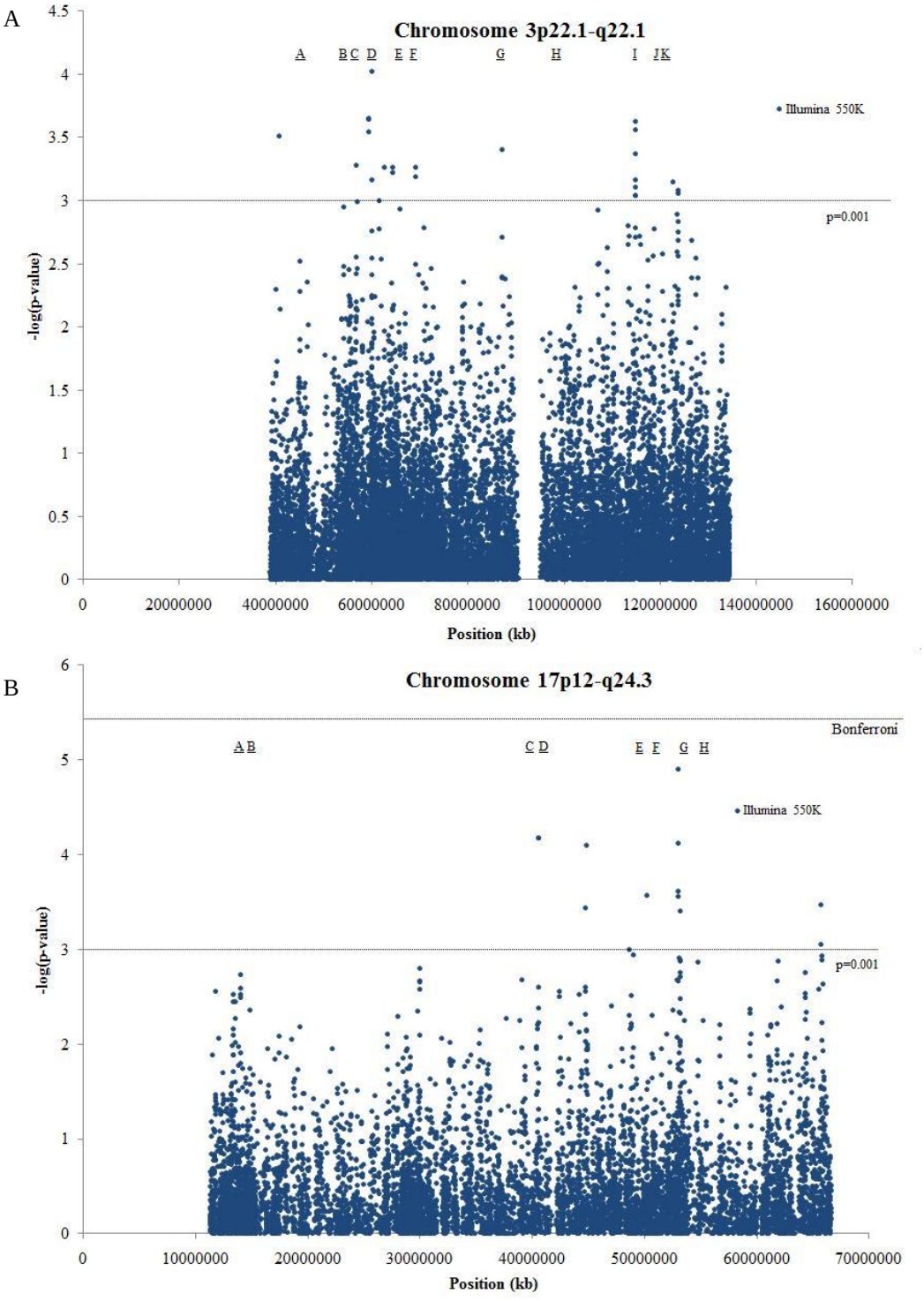


Figure 4.4 – Results of all SNPs tested for association in the B58C specific IgE D2 within GSMA identified susceptibility bins on chromosomes 3p22.1-q22.1 and 17p12-q24.3. SNPs were genotyped using the Illumina 550K platform only, shown in blue. The Bonferroni (chromosome and platform specific) and nominal significance thresholds (p=0.001) are marked. Loci identified for fine mapping are labelled A) A to K on chromosome 3 and B) A to H on chromosome 17 [Table 4.1].

4.4.3 *In silico* Replication

The single locus (D) identified on 17q21.3 was studied in the ALSPAC cohort by analysis of rs7212001, but this SNP did not show replicated association with SPT positivity, p=0.316 [Table 4.2].

4.4.4 Replication Genotyping

The identified SNP, rs7212001 was genotyped in the Nottingham Gedling and Southampton cohorts and did not show statistically significant evidence of association for either specific IgE or positive SPT. However, all results showed the same direction of effect for the minor allele [Table 4.2].

4.4.5 Total IgE Analyses

Association with serum total IgE levels in the British 1958 Birth Cohort was seen for rs7212001 (p=0.024) showing the same direction of effect for specific IgE (OR 0.88, 95% CI 0.79-0.97) and serum total IgE (OR 0.93, 95% CI 0.87-0.99). This was not replicated in any of the other cohorts.

Table 4.2 - Association results for rs7212001 (17q21.31) in individual cohorts and combined analyses. P values and effect sizes for the minor (2^{nd}) allele measured using OR for elevated specific IgE and positive SPT; β for serum total IgE.

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs7212001 'T/C'	Elevated specific IgE	B58C	1,083	2,768	0.88 (0.79-0.97)	**0.011**
		Southampton	281	370	0.91 (0.73-1.13)	0.394
	Positive SPT	ALSPAC	282	1,040	0.92 (0.76-1.11)	0.384
		deCODE	966	27,129	1.04 (0.95-1.14)	0.316
		Gedling	395	847	0.95 (0.80-1.13)	0.585
		Southampton	367	283	0.90 (0.72-1.12)	0.341
	Serum total IgE	B58C	3,873		0.93 (0.87-0.99)	**0.024**
		ALSPAC	1,423		1.01 (0.95-1.08)	0.753
		Gedling	1,167		1.00 (0.88-1.15)	0.951
		Southampton	654		0.84 (0.59-1.19)	0.319
	Combined	Specific IgE	1,364	3,138	0.88 (0.81-0.97)	**$7.87x10^{(-3)}$**
		Positive SPT	2,010	29,299	0.96 (0.89-1.02)	0.167
		Total IgE	1,821		0.93 (0.67-1.17)	0.547

4.4.6 Combined Analyses

Combined analyses of allele count data for elevated specific IgE and positive SPT across cohorts were conducted to estimate pooled p values and effect sizes [Table 4.2 and Figure 4.5]. A significant association with elevated specific IgE was seen for this SNP although these effects were driven by the British 1958 Birth Cohort dataset. No SNP showed significant association with positive SPT. Pooled analysis of serum total IgE levels was carried out for the Nottingham Gedling and Southampton cohorts and no SNP met statistical significance.

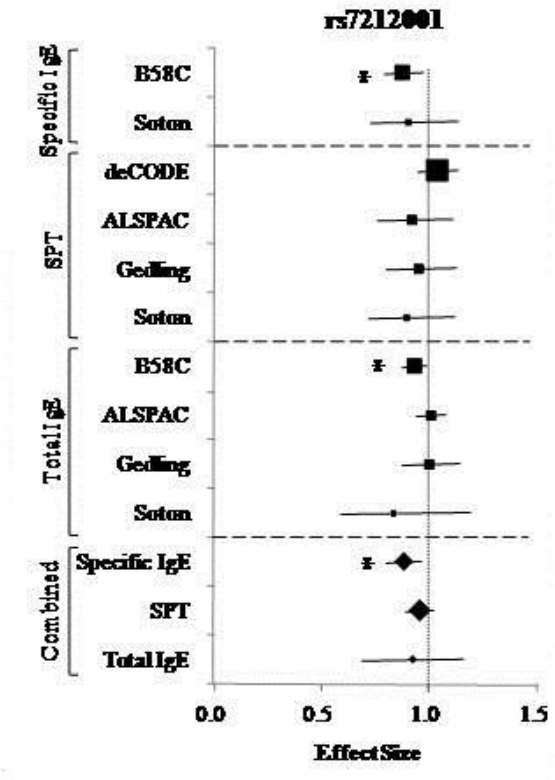


Figure 4.5 - Forest plot for rs7212001 (17q21.31). Per-allele summary effect sizes for the minor allele are shown for atopy phenotypes: elevated specific IgE (OR), positive SPT (OR), and serum total IgE levels (β). Each row represents one study, relative sizes of data points are proportional to the cohort sample size, horizontal lines represent 95% confidence intervals. Diamonds represent the meta-analyses. * denotes statistical significance ($p \leq 0.05$).

4.4.7 Candidate Gene Expression

The identified SNP rs7212001 is within an intronic region of Kinesin family member 18B (*KIF18B*). Two published transcript variants were documented on NCBI for *KIF18B* (NM_001080443). Expression profiling of this gene was carried out in two independent panels consisting of cDNA samples from human airway smooth muscle (HASM), human bronchial epithelial cells (HBEC), lung, dermal fibroblasts, human mast cell line (HMC-1), peripheral blood mononuclear cells (PBMC), brain, kidney, and placenta. Primers were designed to detect all published transcripts and are listed in Appendix 2. In the tested panels, *KIF18B* expression was observed in dermal fibroblast only [Figure 4.6]. PCR products were sequenced and verified against published gene transcripts using NCBI blast.

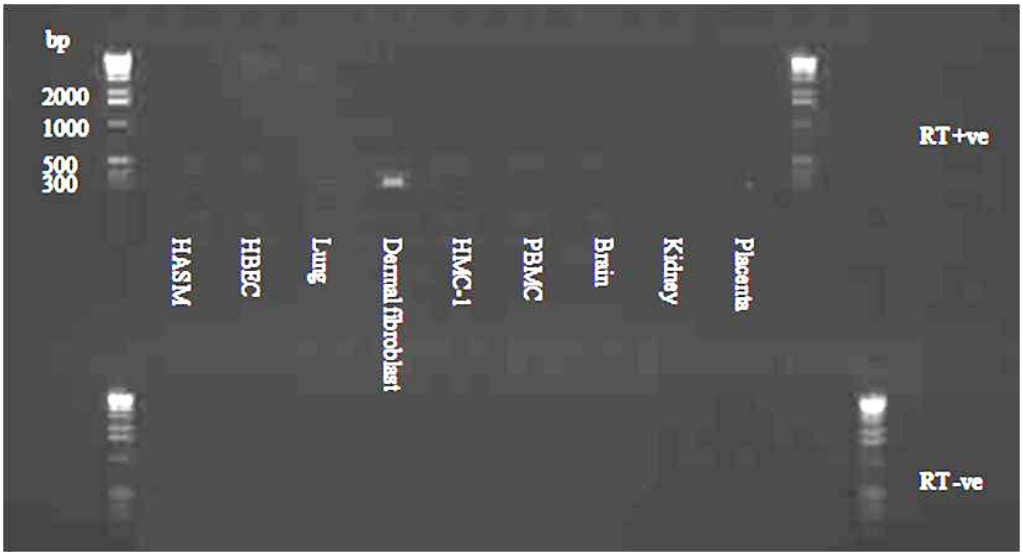


Figure 4.6 – Gel electrophoresis under UV illumination showing expression in dermal fibroblast of *KIF18B*. Primers were designed to detect all two published transcripts (NM_001080443). This is a representative gel from two experiments showing similar results. A product of 331bp was seen in all cell types examined. HASM denotes human airway smooth muscle, HBEC denotes human bronchial epithelial cell, HMC-1 denotes human mast cell line, PBMC denotes peripheral blood mononuclear cells. 1kb DNA ladders are shown on the left and right extremes, RT -ve and water controls are shown in the second row.

4.5 Discussion

This Chapter describes results from assessment of previously reported linkage loci for atopy on chromosomes 3 and 17 to identify potential susceptibility polymorphisms; these loci were identified through a meta-analysis of nine GWL scans carried out in Caucasian asthma families. Three adjacent bins implicating chromosomal region 3p22.1-q22.1 and two adjacent bins implicating 17p12-q24.3 showed significant evidence for linkage to atopy defined by positive SPT responses in the meta-analysis. Testing SNPs within these regions for association with elevated specific IgE in 1,428 subjects from the British 1958 Birth Cohort was carried out. Fine mapping was subsequently undertaken using a more complete dataset from the same cohort (inclusive of an additional 2,425 subjects) as well as 28,140 individuals from the deCODE Icelandic cohort. Follow-up investigations in 7,423 individuals from three additional cohorts were carried out using atopy defined by elevated specific IgE and/or positive SPT. Overall, no causative loci could be determined.

Within the British 1958 Birth Cohort specific IgE D1 dataset, consistency of genotyping results was demonstrated between the Illumina and Affymetrix SNP platforms. However, there were 11.1% and 19.5% increased SNP coverage on chromosomes 3 and 17 respectively using the Illumina platform for genotyping. No SNPs on either platform met Bonferroni corrected threshold for significance. A second less stringent nominal significance level ($p \leq 0.001$) was used identifying a further 40 SNPs on chromosome 3 and 22 SNPs on chromosome 17.

Following analysis for clustering of significant SNPs, extent of LD, and positions of candidate genes in the region, a total of 11 independent susceptibility loci were identified on chromosome 3 and eight loci on chromosome 17. Four loci on chromosome 3 and five loci on chromosome 17 identified using the British 1958 Birth Cohort specific IgE D1 remained significant within the larger dataset, the British 1958 Birth Cohort specific IgE D2 which became available during analyses, in which there was a nearly three-fold increase in the number of individuals genotyped.

Three loci on chromosome 3 and one on chromosome 17 replicated association with positive SPT in the deCODE Icelandic cohort. Further analysis for association with both total IgE in the British 1958 Birth Cohort and non-asthmatic atopy in the deCODE Icelandic Population cohort ($p \leq 0.05$) suggested that the single locus on chromosome 17q21.3 contained SNPs showing atopy specific association. This identified a single candidate gene, *KIF18B* which encodes a motor protein which may be involved in adenosine triphosphate (ATP) binding and microtubule-based movement (Lee *et al.* 2010). The top hit SNP in this region, rs7212001 was studied for replication *in silico* and by prospective genotyping. There was little evidence of consistent association in the additional cohorts for this SNP and therefore, it seems unlikely to confer increased susceptibility to atopy in these populations.

The GSMA assessed GWL scans carried out in family cohorts ascertained using asthma probands. Analyses detailed in this Chapter were carried out in general population cohorts, therefore identified loci in the GSMA may be more important in determining susceptibility to asthma rather than atopy. Efforts had been made to standardise asthma diagnosis and criteria for positive SPT in order to reduce between-study heterogeneity. However, varying phenotype definitions were used and allergens tested differed across studies. This is likely to have influenced the production of allergen-specific IgE and therefore confounded the use of elevated specific IgE as a marker of atopy for association testing in this study. Furthermore, although asthma and atopy are intrinsically linked, increasing evidence in particular from genetic studies to support the existence of trait specific loci suggest that the methodology used in this study was perhaps unlikely to identify atopy susceptibility but rather more pleiotropic genes (Palmer *et al.* 2000).

Results from previous Chapter 3 describing the GWA study from which SNP data for fine mapping was generated suggest that a large number of genes are likely to be involved in the development of atopy, each conferring a modest effect. Given this, the power to find association produced by a single or few gene(s) in specific chromosomal regions may be low. Estimations of study power for individual cohorts used in analyses (detailed in Chapter 3) show that there is overall low statistical power to detect small effects. In addition, the

significant signals seen may be due to several genes in combination in close proximity within the entire region and therefore fine mapping for individual genes may not be appropriate.

In the past year, a second GSMA has been completed in 20 GWL scans in multiple ethnicities showing consistency of findings (n=3,024 families). This study identified novel loci for asthma and atopy susceptibility as well as provided further genome-wide significant support for 3p25.3-q24 and 17q12-q24 for SPT response in individuals of European ancestry (Bouzigon *et al.* 2010). Convincing evidence for these two regions has been reported by numerous studies demonstrating both linkage and association for asthma and atopy phenotypes (Ober *et al.* 1999; Bradley *et al.* 2002). These results suggest that 3p22.1-q22.1 and 17p12-q24.3 may harbour several susceptibility genes, including some pleiotropic effects on correlated phenotypes including toll-like receptor 9 (*TLR9*), chemokine (C-C motif) receptor 3 (*CCR3*), chemokine (C-C motif) receptor 5 (*CCR5*) on chromosome 3, and interestingly, *ORMDL3* on chromosome 17 (Bradley *et al.* 2002; Moffatt *et al.* 2007).

Shifts in the positions of significant peaks seen between the two British 1958 Birth Cohort datasets indicate that differences in sample sizes used in genetic association studies have a considerable influence on the accuracy and ability in localising causative variants. These results indicate that the initial findings may have been modest and less robust to type II errors. Using the smaller cohort to select distinct loci for further analysis may have prematurely eliminated regions harbouring the true causative variants. Due to practical limitations, sub-group genotyping in two phases was carried out in the British 1958 Birth Cohort, but this is likely to have introduced 'batch effects' or technical bias leading to non-biological experimental variation in the two datasets used.

Similarly to Chapter 3, there are problems in heterogeneity across the fine mapping and replication study populations in demographics, phenotype definitions, and variable SNP coverage across genotyping platforms. For example, within locus D on chromosome 17, a total of 78 genotyped SNPs were tested in the British 1958 Birth Cohort compared to three SNPs in the deCODE Icelandic cohort. In addition, less robust phenotyping in the deCODE Icelandic

cohort through defining control subjects as all individuals in the general population may have resulted in misclassified subjects who could have been considered cases using elevated specific IgE as the marker of atopy or alternatively have asthma symptoms. For this reason, study power estimations were not carried out for this cohort. Furthermore, variation in environmental exposures in particular in the deCODE Icelandic cohort are likely to have influenced the ability to detect consistent signals.

Subtle, underlying population stratification due to varying ancestry may have led to potential differences between the deCODE Icelandic cohort and other British cohorts used in this study. This variation in linkage disequilibrium patterns in combination with a poor overlap of genotyped SNPs within these chromosomal regions may have precipitated the lack of replicated loci to carry forward for fine mapping.

The results presented in this Chapter have not been able to identify atopy susceptibility loci through fine mapping of two susceptibility regions for SPT positivity on chromosomes 3p22.1-q22.1 and 17p12-q24.3 identified through a previous meta-analysis of nine GWL scans. This further supports conclusions drawn from the GWA study described in Chapter 3 indicating evidence for multiple loci harbouring modest effects.

This fine mapping approach reduces the number of regions studies and therefore increases statistical power and previous research suggest that signals generated by linkage are likely to be driven by variants with larger effect sizes. However, due to the large number of individuals used in the discovery cohort compared to much smaller replication samples, effect sizes were most likely overestimated. In combination to the other phenotyping and genotyping problems elucidated earlier, this would limit an ability to find variants determining disease susceptibility with small effects. Current methodologies to reliably identify causative genes require greater refinement in both study power and design.

Chapter 5

Genome-Wide Association Study to Identify Genetic Determinants of Severe Asthma

5.1 Introduction

From the studies performed to date, it is clear that the total number of genetic loci implicated in asthma susceptibility is large and the proportion of effect from individual genetic variants both independently and in combination remains unclear. Importantly, association studies to date have predominantly focused on mild-to-moderate disease and genetic risk factors for severe asthma remain especially unclear (see Chapter 1: Sections 1.3.4 and 1.3.5).

Previous studies of severe asthma have mostly used related traits such as measures of lung function defined using FEV_1 or medication requirements such as steroid use. These have reported association between polymorphisms in previously reported asthma genes with increased declines in lung function and near fatal events including *IL4* and *IL4R* (Burchard *et al.* 1999; Rosa-Rosa *et al.* 1999; Sandford *et al.* 2000), *TGFB1* and chemokine (C-C motif) ligand 2 (*CCL2*) (Pulleyn *et al.* 2001; Szalai *et al.* 2001), and *ADAM33* (Jongepier *et al.* 2004) as well as *ADRB2* with worsened treatment response (Israel *et al.* 2001). Furthermore, a SNP near the *ORMDL3* locus was reported to be associated with asthma severity in a group of early-onset asthma cases (p=0.0012) (Halapi *et al.* 2010). In 2009, a relatively small genome-wide association (GWA) study was conducted on a population of patients with severe or difficult to treat asthma from the TENOR study (473 cases, 1,892 asthma-free controls) and identified association with multiple SNPs in the *RAD50-IL13* and *HLA-DR/DQ* regions although no signal met conventional genome-wide significance (Li *et al.* 2010).

The Asthma UK Genetics of Severe Asthma (AUGOSA) study has established the largest genotype database of severe asthma patients to date consisting of 1,026 individuals of

European ancestry recruited across UK-based centres. Participants from this study contributed to the group of cases used in the recent GABRIEL consortium GWA study. Furthermore, some overlap in controls also occurred with individuals collected by Busselton and the WTCCC2. In addition to identifying determinants of asthma *per se*, GABRIEL also reported analyses for severe asthma (927 individuals after tests of quality control) of identified loci for all asthma subjects but did not identify any significant associations (Moffatt *et al.* 2010).

The AUGOSA study presented the opportunity to use the GWA approach with the potential to identify novel loci to address two key questions: firstly to determine whether the genetic susceptibility to severe asthma differs from that of asthma *per se*, and secondly to evaluate the contribution of polymorphisms in recently reported asthma susceptibility genes identified through GWA in a more extreme phenotype.

5.2 Aim

The aim of this Chapter was to use GWA to identify common genetic variants influencing susceptibility to the development of severe asthma.

5.3 Methods

5.3.1 Study Populations

The primary GWA study was carried out using the Asthma UK Genetics of Severe Asthma (AUGOSA) study consisting of 1,026 cases and 3,372 individuals genotyped using the Illumina HumanHap550, HumanHap610-Quad, HumanHap660w-Quad, and Human1M-duo v3 SNP platforms. Power estimations calculated using CaTS 0.0.2 indicate that this dataset had 80% power (α=0.05) to detect an odds ratio (OR) of approximately 1.19 for a SNP with a minor allele frequency (MAF) of 10%; 1.14 for a SNP with a MAF of 25%; and 1.12 for a SNP with a MAF of 40%. Replication of SNPs of interest was undertaken in the Australian Asthma Genetics Consortium (AAGC) study consisting of 231 cases and 1,345 controls genotyped using the Illumina HumanHap610-Quad SNP platform. This replication cohort dataset had 80% power to detect an OR of 1.46 (MAF=10%), 1.31 (MAF=25%), and 1.28 (MAF=40%). Details of study recruitment and phenotyping for these population cohorts are described in Chapter 2: Section 2.1.2.

Genotyping quality control measures undertaken in the AAGC study are described in Chapter 2: Section 2.2.1.2. Remaining autosomal SNPs were then imputed up to 7.8 million SNPs against reference panel data from the March 2010 release of the 1000 Genomes Project and the February 2009 release of the HapMap3 Genome Browser release #2 using Impute2 (Howie *et al.* 2009). Post-imputation exclusion of SNPs with $r^2_{imp}<0.3$, MAF<1%, and Hardy-Weinberg equilibrium (HWE) $p<10^{(-6)}$ were carried out. Genotype data for a subsequent total of 5.7 million SNPs were tested for association using a standard case-control allelic test with a genomic inflation factor of 1.02.

5.3.2 Genotyping Quality Control

Results of genotyping in all case subjects from the AUGOSA study were returned from Centre National de Génotypage (Paris, France) in .txt files detailing genotyping success (report.txt), summary of results per SNP (success_mark.txt), summary of results per individual (success_sample.txt), and the raw genotype calls for all SNPs genotyped in each individual

(data.txt). Results files containing raw genotype calls for all control subjects were provided by collaborators.

Initial Filtering

Initial filtering of supplied genotyped data was carried out in PLINK 1.07. Only autosomal SNPs were retained for analyses. Relevant data columns were extracted using R version 2.12.1 to create .ped and .map PLINK input files. The .ped file contained the results per individual created using the data.txt file; the first 6 columns of this file contained the following information: Family ID (0=unknown), Individual ID, Paternal ID (0=unknown), Maternal ID (0=unknown), Sex (1=male, 2=female, 0=unknown), Affection status (1=unaffected, 2=affected). Each line of the .map file described a single SNP containing three columns extracted from the success_mark.txt file: Chromosome (1 to 22, X, Y or 0 if unplaced), rs# or snp identifier, Base-pair position (bp units).

The following command line was run in UNIX to remove individuals with <90% SNPs genotyped (--mind 0.1), SNPs with <90% of individuals genotyped (--geno 0.1), SNPs with MAF <1% (--maf 0.01), SNPs not meeting HWE $p<10^{(-4)}$ (--hwe 0.0001):

plink --file augosa --map3 --maf 0.01 --mind 0.1 --geno 0.1 --hwe 0.0001 --recode

No individuals were removed during this filtering; the numbers of SNPs excluded are listed in Table 5.1. A subset of 490,303 SNPs present in all six cohorts was then taken.

Table 5.1 – Number of samples remaining and number of SNPs excluded after initial filtering. AUG610 and AUG660 denote results of cases recruited by the AUGOSA Consortium genotyped on the Illumina HumanHap610-Quad and 660w-Quad platforms respectively, BHS denotes the Busselton Health Study.

Filter	**AUG610**	**AUG660**	**GABRIEL Severe**	**BHS**	**WTCCC2**	**T1DGC**
Sex Chromosomes	15303	13392	15303	15303	42558	13845
MAF <1%	29485	23970	26832	31118	666	23615
<90% samples genotyped	3115	1770	1994	2169	0	1260
HWE $p<10^{(-4)}$	259	948	902	1035	2109	3238

Tests of Differential Missingness

The proportion of missing genotypes between cases and controls for each SNP was compared using Fisher's exact test using the following command line in UNIX:

plink --file augosa --test-missing

A total of 9,414 SNPs were excluded with $p<10^{(-4)}$ under the null hypothesis of equal proportions of missing data in cases and controls, leaving 480,889 SNPs for testing.

Principal components analysis (PCA)

Correction for underlying population structure in the data was undertaken by Dr. Nick R G Shrine using principal components analysis (PCA) (see Chapter 2: Section 2.4.1) based on the covariance of effect allele loadings (0, 1, 2) between SNPs and samples as implemented in EIGENSOFT 3.0. PCA can show components that are influenced by the linkage disequilibrium (LD) structure if SNPs in strong LD are used. Therefore, a subset of 57,213 uncorrelated SNPs for the PCA was identified by removing SNPs correlated by $r^2>0.1$ within a window of 50 adjacent SNPs in PLINK 1.07. The PCA results are shown in Figure 5.1 indicating generally homogenous populations. A total of 119 outliers were identified with variances $>6\sigma$ for exclusion leaving 933 cases and 3,346 controls in the analysis [Table 5.2]. The first 10 principal components was stored in a PCA.txt file with columns: Family ID, Individual ID, $PCA_{1..10}$.

Table 5.2 – Number of samples remaining after Principal components analysis (PCA). AUG610 and AUG660 denote results of cases recruited by the AUGOSA Consortium genotyped on the Illumina HumanHap610-Quad and 660w-Quad platforms respectively, BHS denotes the Busselton Health Study.

Filter	**AUG610**	**AUG660**	**GABRIEL Severe**	**BHS**	**WTCCC2**	**T1DGC**
Total	121	561	344	576	1483	1313
PCA outliers	8	31	54	11	2	13
Remaining	113	530	290	565	1481	1300

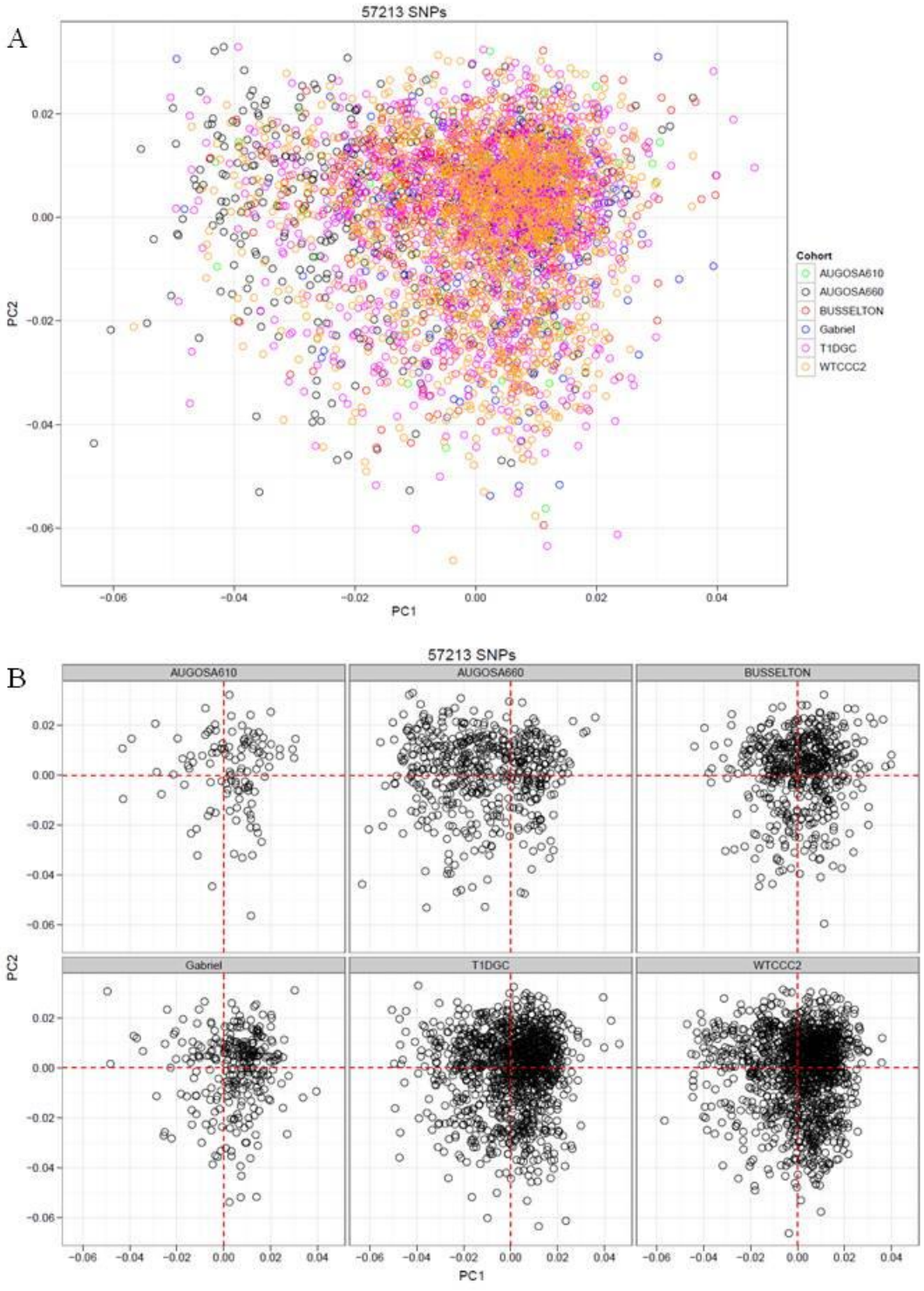


Figure 5.5 - Example plots showing the first two principal components. AUG610 and AUG660 denote results of cases recruited by the AUGOSA Consortium genotyped on the Illumina HumanHap610-Quad and 660w-Quad platforms respectively. A) No clear separation of clusters for each cohort (three case cohorts: AUG610, AUG660, GAB and three control cohorts: BUSSELTON, WTCCC2, T1DGC) suggests homogeneous populations. B) The PCA analysis identified 119 outliers with variances >6σ, these were excluded from subsequent analyses.

5.3.3 Genome-wide Association (GWA) Testing

Association testing was carried out using logistic regression models with one degree of freedom (df) (see Chapter 2: Section 2.4.2) comparing allele frequencies between cases and controls using PLINK 1.07. Identified principal components were used as covariates in the association analysis (--covar). Modified results files after quality control measures were analysed using the following command line in UNIX:

```
plink --bfile augosa_QC --logistic --beta --ci 0.95 --covar PCA.txt
```

The .assoc.logistic file produced contained regression coefficients (--beta) with 95% confidence intervals (--ci 0.95) and p values for association for the default minor allele per SNP.

5.3.4 Post-Association Quality Control

Control-control comparisons using the three control cohorts were run by appointing the status of each cohort in turn as cases in an association test with the others. This identified 21 SNPs with $p<10^{(-6)}$ in a given test for exclusion, leaving 480,868 SNPs in the analysis. The test inflation statistic (λ) (see Chapter 2: Section 2.4.3) was calculated using the command line in R version 2.12.1 for data frame (results) with regression coefficients (BETA) and standard errors (SE):

```
lambda<-median((results$BETA/results$SE)^2,na.r=T)/qchisq(0.5,1)
```

Manhattan and Quantile-Quantile (Q-Q) plots were generated to visualise SNP results (see Chapter 2: Section 2.3.2).

5.3.5 Statistical Significance Testing

Genome-wide significance was defined using $p\leq1.02x10^{(-7)}$ as determined by the Bonferroni correction at the 5% significance level (see Chapter 2: Section 2.4.8). Suggestive regions of interest were identified as those with a sentinel SNP showing association with asthma (at a threshold of $p<5x10^{(-5)}$) with at least one additional SNP within 500kb also reaching a threshold of $p<5x10^{(-5)}$.

5.3.6 Imputation Using the 1000 Genomes Project

Imputation using the 1000 Genomes Project (1000G) reference panel (see Chapter 2: Section 2.4.7) was used to improve the resolution of regions identified for association from genotyped data carried out by Dr. Nick R G Shrine in two stages. Genotyped data were imputed to 6,858,242 SNPs and GWA testing of imputed data was carried out. However, analyses were limited to regions identified from genotyped data. Genotypes were compared across the quality controlled 4,279 cases and controls to phase haplotypes, and then compared to haplotype blocks in the August 2010 release of the 1000G data. The European (CEU) panel was used consisting of 120 individuals genotyped at 6,858,242 SNPs downloaded from the MaCH website. The phasing step requires all alleles to be defined on the positive strand. Strand assignments of our samples were checked against UCSC and dbSNP. The phasing was performed using MaCH 1.0 and the imputation using minimac. Post imputation quality control was carried out to exclude SNPs with MAF <1% and imputation quality threshold $r^2_{imp}<0.3$. Association testing of remaining SNPs were performed using ProbABEL 0.1-3 using a logistic model with the dose of the effect allele (on a continuous scale between 0 and 2 to reflect imputation uncertainty) as the independent variable and 10 ancestry principal components derived from genotyped SNPs as covariates. Post association filters were applied to remove SNPs showing significant association in control-control comparisons ($p<10^{(-6)}$) leaving 6,103,628 SNPs.

5.3.7 Replication and Meta-Analysis

Replication analyses were carried out for a selected number of SNPs using the AAGC study. This was based on the following four criteria: genotyped SNPs identified through the primary GWA study, imputed SNPs with lower p values than genotyped SNPs in regions containing the genotyped SNPs of interest, imputed SNPs with $p>10^{(-5)}$ responsible for secondary peaks in regions with known asthma genes, as well as SNPs in new regions identified through analysis using imputed data with $p<10^{(-5)}$ and $r^2_{imp}>0.7$. Statistical significance for replication was assessed using a 5% significance threshold and results of inverse-variance weighted meta-analysis assessed using conventional criteria for genome-wide significance ($p=5x10^{(-8)}$) (McCarthy *et al.* 2008).

5.3.8 Evaluation of GABRIEL loci

The contributions of risk polymorphisms identified for mild-to-moderate asthma by the GABRIEL Consortium were assessed in the AUGOSA severe asthma population. Regions ±500kb of SNPs reported to be associated with asthma in the GABRIEL study were examined, including both regions reported to have reached genome-wide significance i.e. $p \leq 7.2 \times 10^{(-8)}$ and those providing suggestive evidence of association i.e. $p \leq 5 \times 10^{(-7)}$ in GABRIEL.

5.4 Results

5.4.1 Loci Identified by GWA Testing

The primary analysis involved 480,868 SNPs in 933 cases and 3,346 controls, resulting in a modest test inflation statistic (λ=1.04). The Q-Q plot shows observed versus expected -$\log_{10}$ p values for all tested SNPs showing a small number of SNPs with extreme p values and additional SNPs with more modest effects [Figure 5.2]. The Manhattan plot shows -$\log_{10}$ p values for all tested SNPs ordered by chromosomal position showing the highest significance SNPs lying on chromosomes 2, 5, 6, 7, 13, and 17 [Figure 5.3].

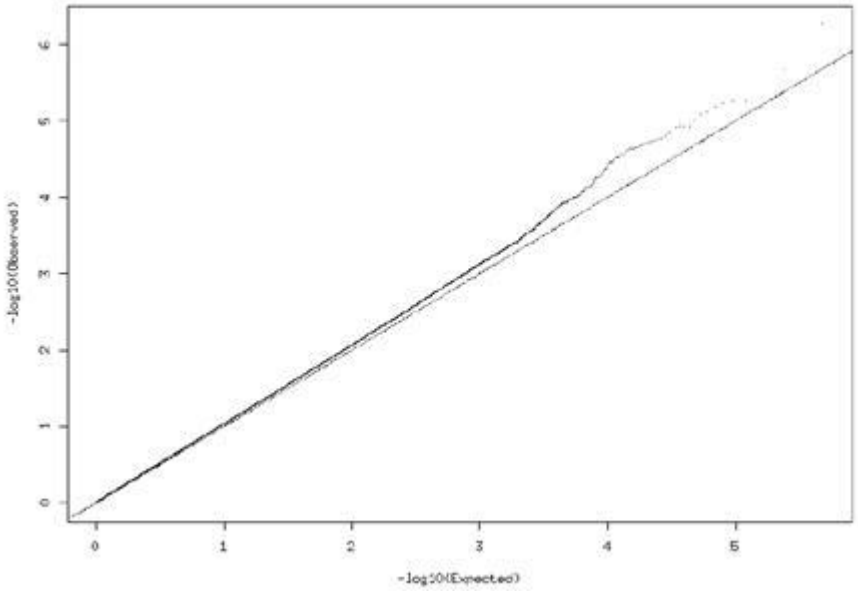


Figure 5.2 – Quantile-Quantile (Q-Q) plot showing GWA results for genotyped SNPs. The straight line shows the distribution of 480,868 SNPs analysed in 933 cases and 3,346 controls under the null hypothesis.

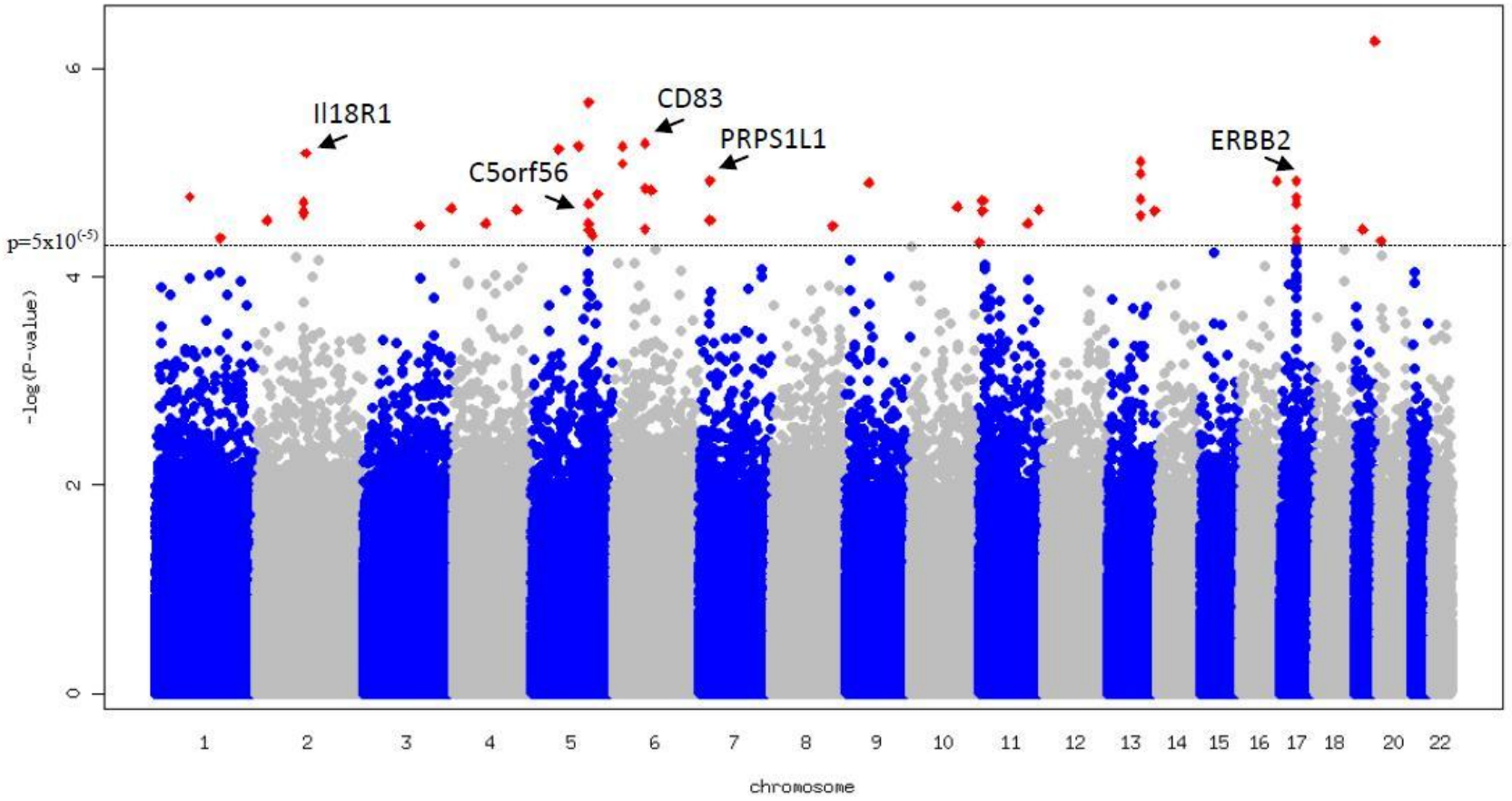


Figure 5.3 - Manhattan plot showing GWA results for 480,868 genotyped SNPs analysed in 933 cases and 3,346 controls under analysis. * Note: the signal on chromosome 13q31.1 was intergenic.

No SNPs met genome-wide significance for association with severe asthma using a cut-off defined by the Bonferroni correction: $p=1.02x10^{(-7)}$. Using a second criterion to evaluate other potential loci below this threshold, eight SNPs were identified with $p<5x10^{(-5)}$, with at least one other SNP within 500kb with $p<5x10^{(-5)}$ [Table 5.3]. Further assessment of supporting evidence within the region for these SNPs suggests that six of these loci may contain susceptibility genes for severe asthma [Figures 5.4 to 5.6]. Region plots showing low supporting evidence for the two eliminated SNPs are shown in Appendix 5. Sentinal SNPs for the six suggestive loci were rs11745587 within chromosome 5 open reading frame 56 (*C5orf56*) on 5q31.1 ($p=2.09x10^{(-6)}$), rs9382963 upstream of CD83 molecule (*CD83*) on 6p23 ($p=5.61x10^{(-6)}$), rs2496764 within an intergenic region on 13q31.1 ($p=7.86x10^{(-6)}$), rs12699949 within phosphoribosyl pyrophosphate synthetase 1-like 1 (*PRPS1L1*) on 7p21.1 ($p=1.19x10^{(-5)}$), rs1810132 within v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (*ERBB2*) on 17q12-21 ($p=1.73x10^{(-5)}$), and rs3771166 within interleukin 18 receptor 1 (*IL18R1*) on 2q12.1 ($p=1.93x10^{(-5)}$).

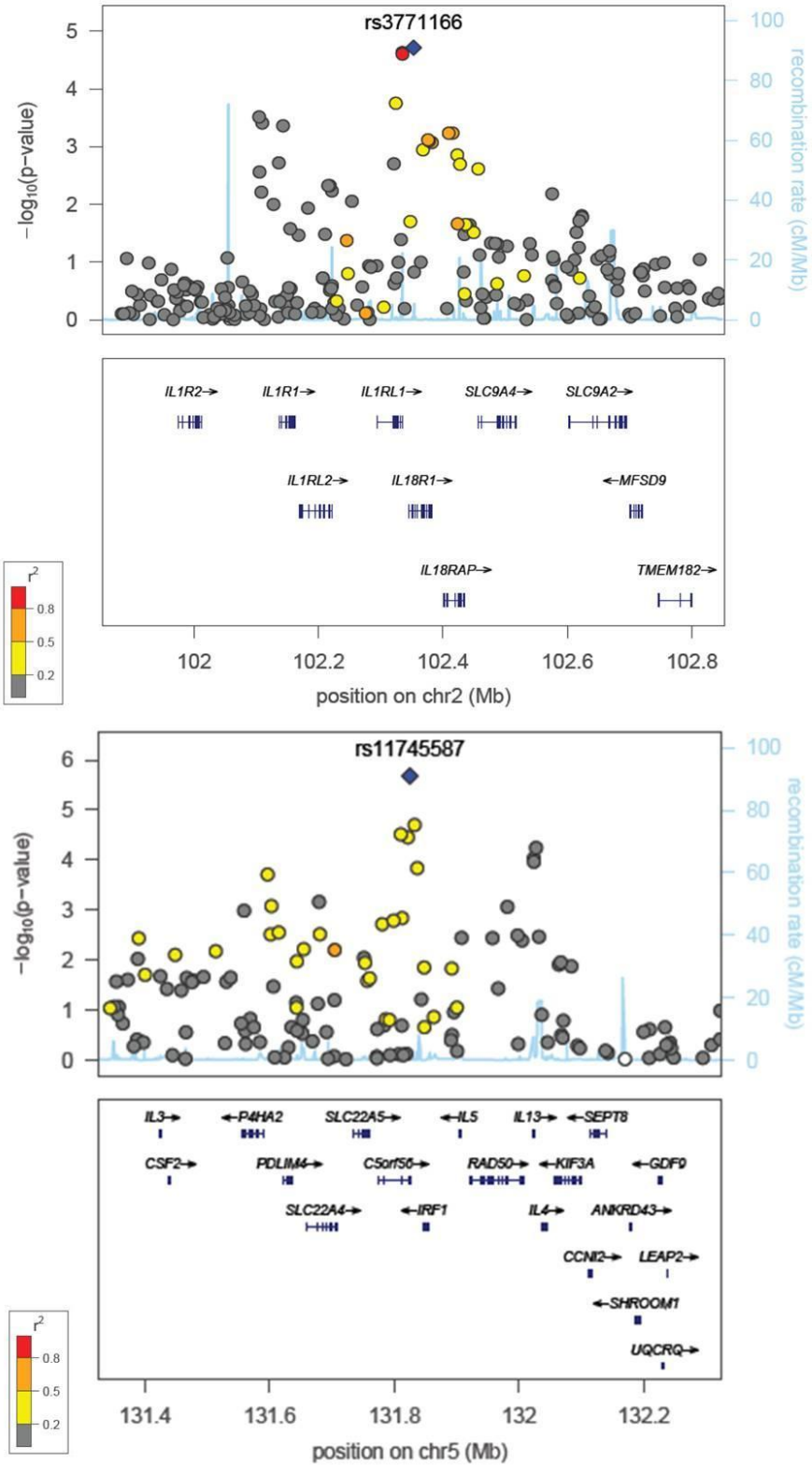


Figure 5.4 - Region plots for suggestive loci 1 to 2 (rs3771166 and rs11745587) using genotyped data identified by GWA testing. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

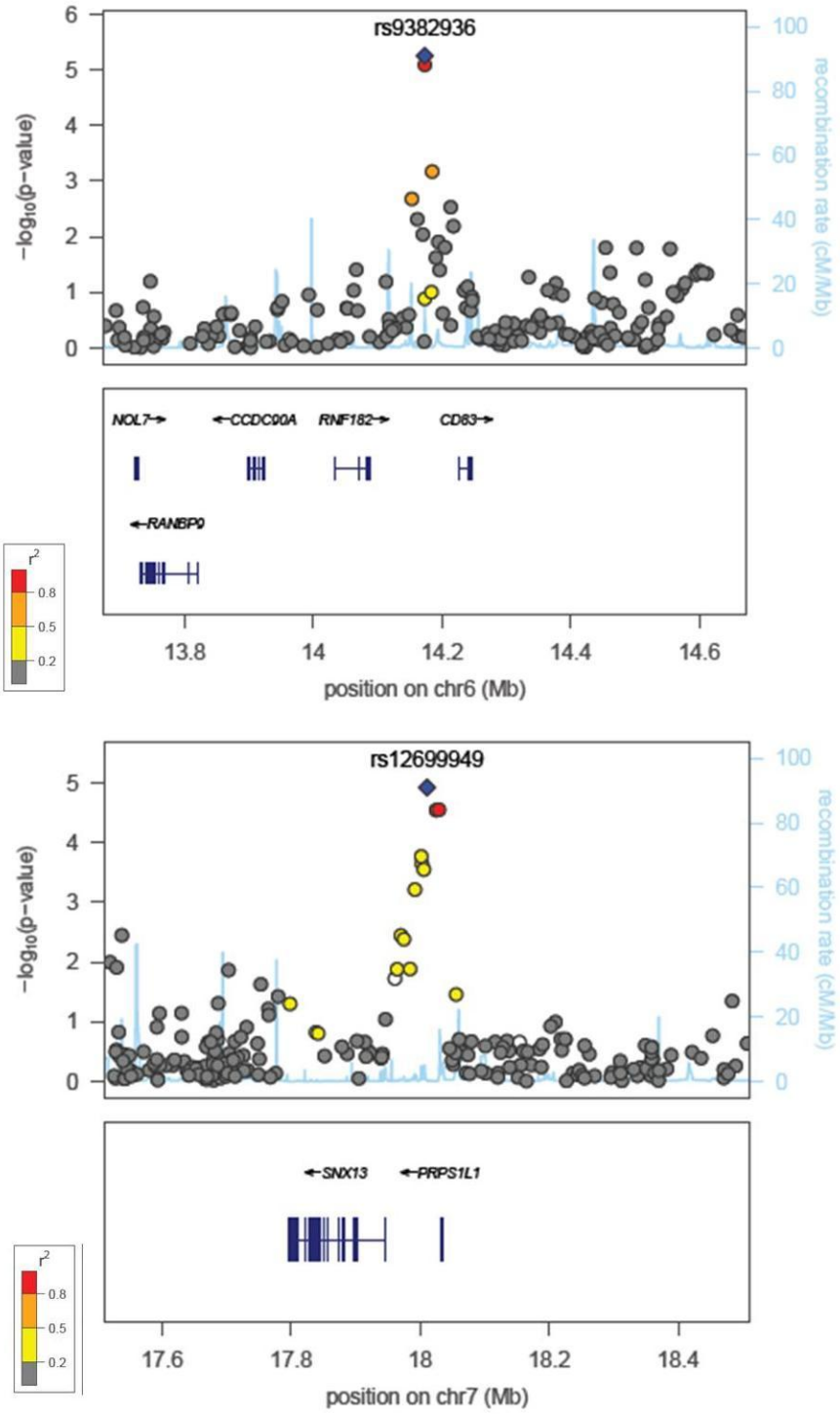


Figure 5.5 - Region plots for suggestive loci 3 to 4 (rs9382936 and rs12699949) using genotyped data identified by GWA testing. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

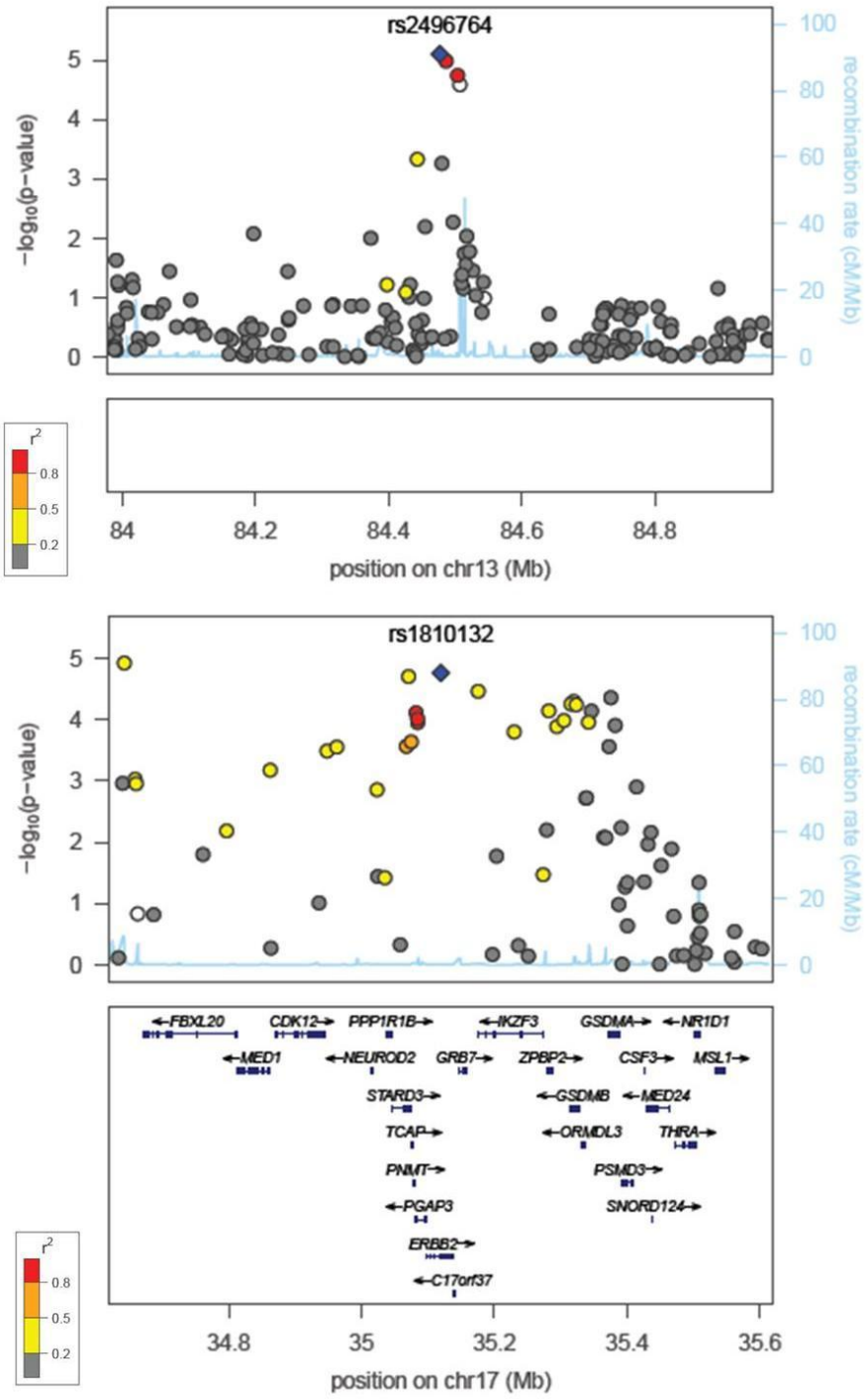


Figure 5.6 - Region plots for suggestive loci 5 to 6 (rs2496764 and rs1810132) using genotyped data identified by GWA testing. Note rs2496764 is intergenic. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

Table 5.3 – Genotyped SNPs meeting significance criteria for association with severe asthma. Ref denotes reference allele; effect refers to the minor allele. SNPs with low supporting evidence in the region were not taken into additional analyses.

Chromosome	Gene	SNP	Position	Ref	Effect	MAF	OR (95% CI)	p value
Suggestive loci meeting significance criteria								
2q12.1	*IL18R1*	rs3771166	102352654	G	A	0.38	0.79 (0.71-0.88)	$1.93x10^{(-5)}$
5q31.1	*C5orf56*	rs11745587	131824821	G	A	0.36	1.30 (1.17-1.45)	$2.09x10^{(-6)}$
6p23	*CD83*	rs9382936	14173097	G	A	0.25	1.31 (1.17-1.48)	$5.61x10^{(-6)}$
7p21.1	*PRPS1L1*	rs12699949	18010787	G	A	0.32	0.77 (0.69-0.87)	$1.19x10^{(-5)}$
13q31.1	Intergenic	rs2496764	84477159	G	A	0.20	1.34 (1.18-1.52)	$7.86x10^{(-6)}$
17q12-21	*ERBB2*	rs1810132	35119531	T	C	0.31	1.28 (1.14-1.43)	$1.73x10^{(-5)}$
Eliminated loci due to low region support								
6p12.1	*AL591034.2*	rs9395865	53415653	G	A	0.29	0.76 (0.67-0.85)	$5.27x10^{(-6)}$
11p15.4	*TRIM5*	rs10500656	5879658	G	A	0.15	1.36 (1.18-1.56)	$1.87x10^{(-5)}$

IL18R1: interleukin 18 receptor 1; *C5orf56*: chromosome 5 open reading frame 56; *CD83*: CD83 molecule; *PRPS1L1*: phosphoribosyl pyrophosphate synthetase 1-like 1; *ERBB2*: v-erb-b2 erythroblastic leukemia viral oncogene homolog 2; *TRIM5*: tripartite motif containing 5.

Table 5.4 – Top imputed SNP in suggestive regions identified by GWA testing using genotyped data. Ref denotes reference allele, effect refers to the minor allele. r^2_{imp}: Imputation quality. r^2_{LD}: Linkage disequilibrium with corresponding genotyped SNPs in Table 5.3 based on 1000G June 2010.

Chromosome	Gene	SNP	Position	Ref	Effect	MAF	r^2_{imp}	r^2_{LD}	OR (95% CI)	p value
2q12.1	*IL18R1*	rs9807989	102337632	T	C	0.33	0.91	0.84	0.76 (0.67-0.85)	$5.20x10^{(-6)}$
7p21.1	*PRPS1L1*	rs12699948	18044210	C	G	0.28	0.92	0.75	0.75 (0.66-0.85)	$4.84x10^{(-6)}$
13q31.1	Intergenic	rs9547037	84476839	G	T	0.19	0.90	0.83	1.38 (1.20-1.58)	$6.60x10^{(-6)}$
17q12-21	*STARD3*	rs9972882	35061224	C	A	0.25	0.98	0.52	1.32 (1.17-1.49)	$5.17x10^{(-6)}$

IL18R1: interleukin 18 receptor 1; *PRPS1L1*: phosphoribosyl pyrophosphate synthetase 1-like 1; *STARD3*: StAR-related lipid transfer (START) domain containing 3.

5.4.3 Imputation Using the 1000 Genomes Project

Imputation was used to refine association signals within the six suggestive loci [Figures 5.7 to 5.9]. The test inflation statistic for GWA testing was unchanged following imputation (λ=1.04). This analysis identified a further four SNPs with a lower p value than the original genotyped SNP in the same region: rs9807989 within *IL18R1* on 2q12.1 ($p=5.20x10^{(-6)}$), rs12699948 within *PRPS1L1* on 7p21.1 ($p=4.84x10^{(-6)}$), rs9547037 intergenic on 13q31.1 ($p=6.60x10^{(-6)}$), and rs9972882 within StAR-related lipid transfer (START) domain containing 3 (*STARD3*) on 17q12-21 ($p=5.17x10^{(-6)}$) [Table 5.4].

Two imputed SNPs produced secondary peaks in other regions with known asthma associated genes: rs13035227 within interleukin 1 receptor-like 1 (*IL1RL1*) on 2q12.1 ($p=8.91x10^{(-5)}$), and rs847 within *IL13* on 5q31 ($p=4.05x10^{(-5)}$) [Figure 5.6]. An additional 12 SNPs with $p<10^{(-5)}$ and $r^2_{imp}=0.7$ produced signals in new regions ($1.19x10^{(-5)} \leq p \leq 2.82x10^{(-7)}$) following analyses of imputed data. Two of these SNPs, rs9897185 and rs4794820 produced secondary peaks in the previously identified 17q12-21 locus. All SNP results are detailed in Table 5.5, region plots for the secondary peaks produced by rs13035227 (2q12.1), rs847 (5q31), and rs9897185 (17q12-21) are shown in Figures 5.10 and 5.11. Region plots for the secondary peak produced by rs4794820 (17q12-21) and all other new regions identified through imputation are shown in Appendix 5.

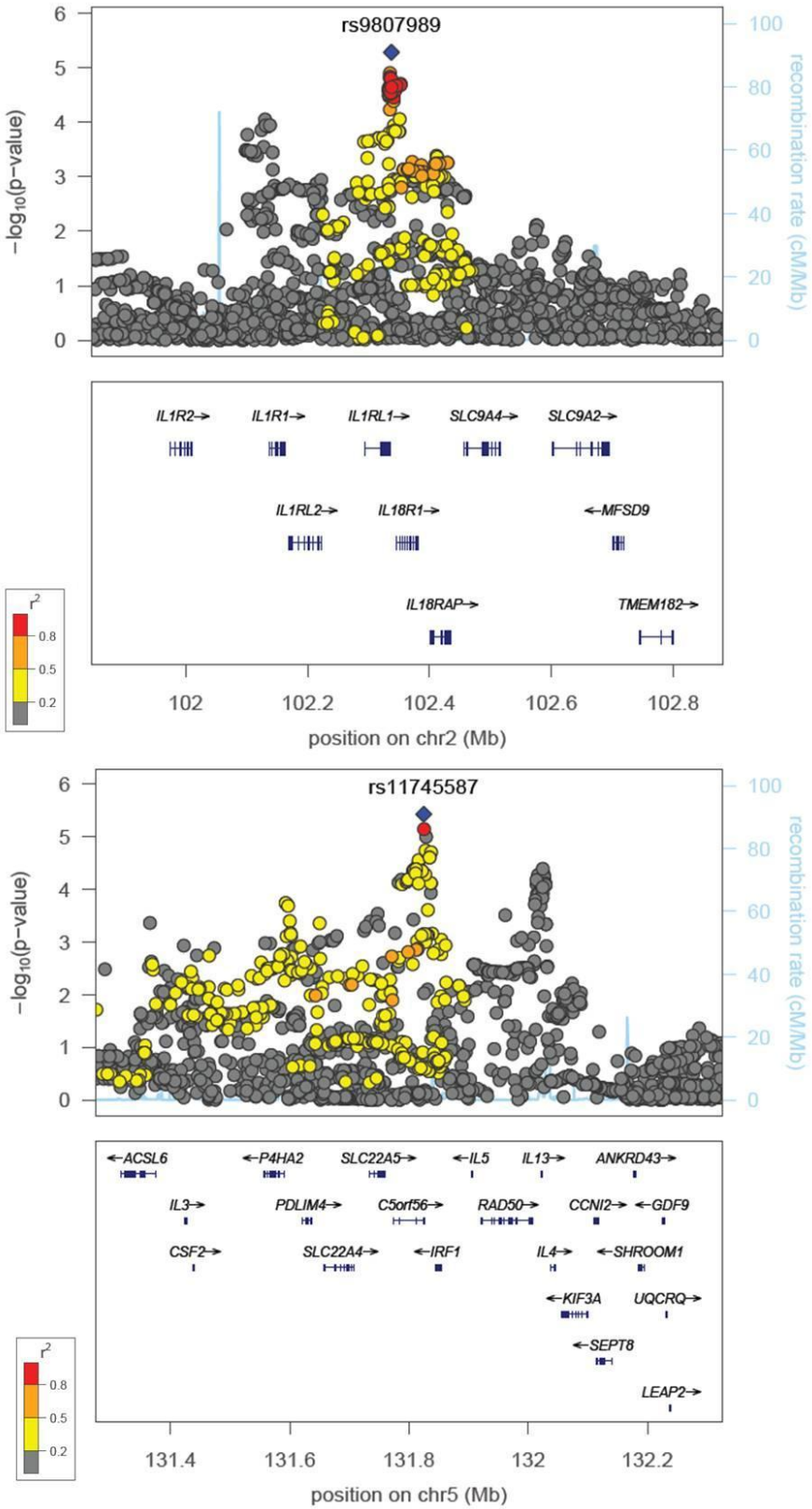


Figure 5.7 - Region plots using imputed data for suggestive loci 1 and 2 (rs3771166 and rs11745587) identified by GWA testing. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

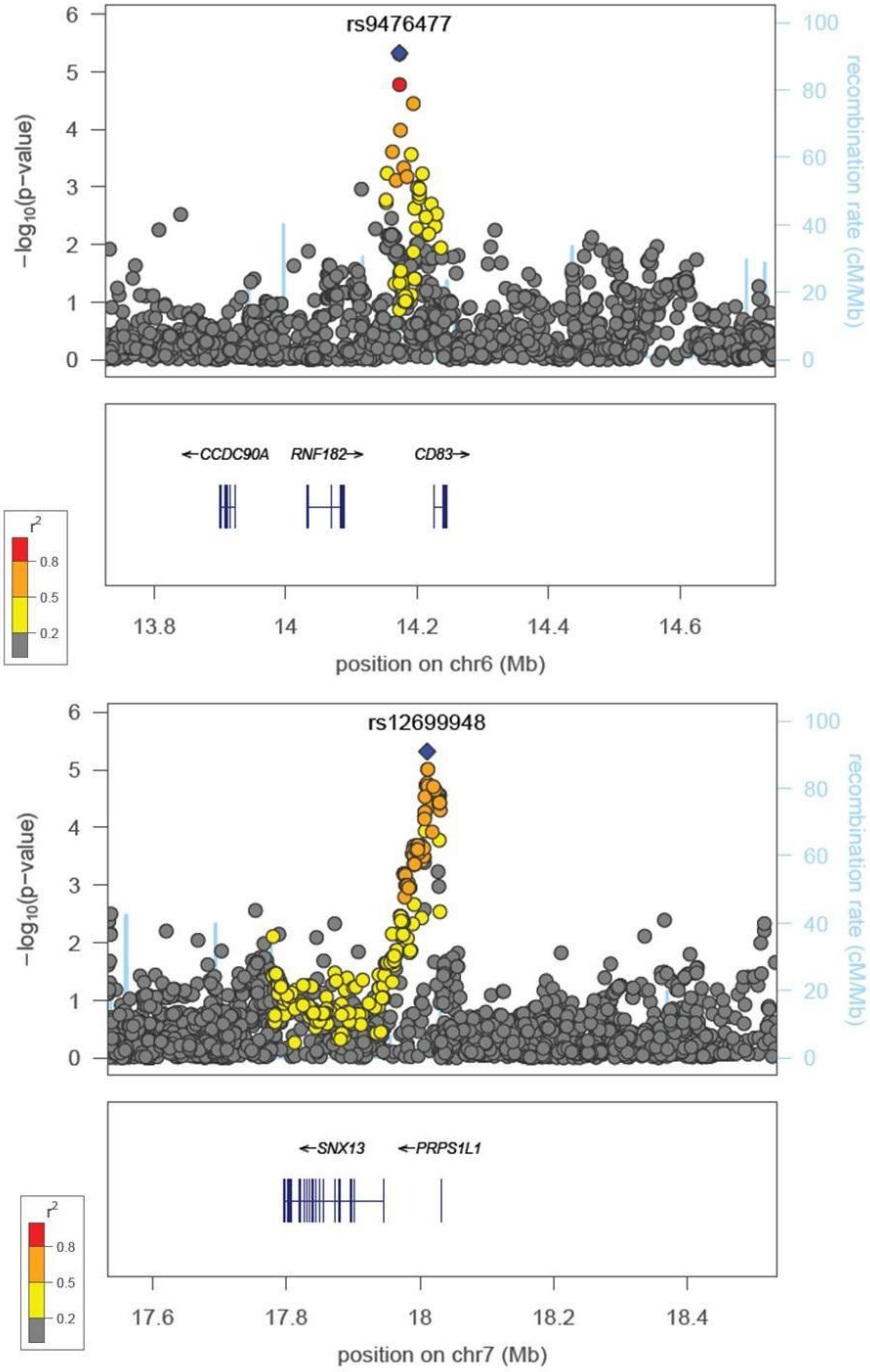


Figure 5.8 - Region plots using imputed data for suggestive loci 3 and 4 (rs9382936 and rs12699949) identified by GWA testing. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

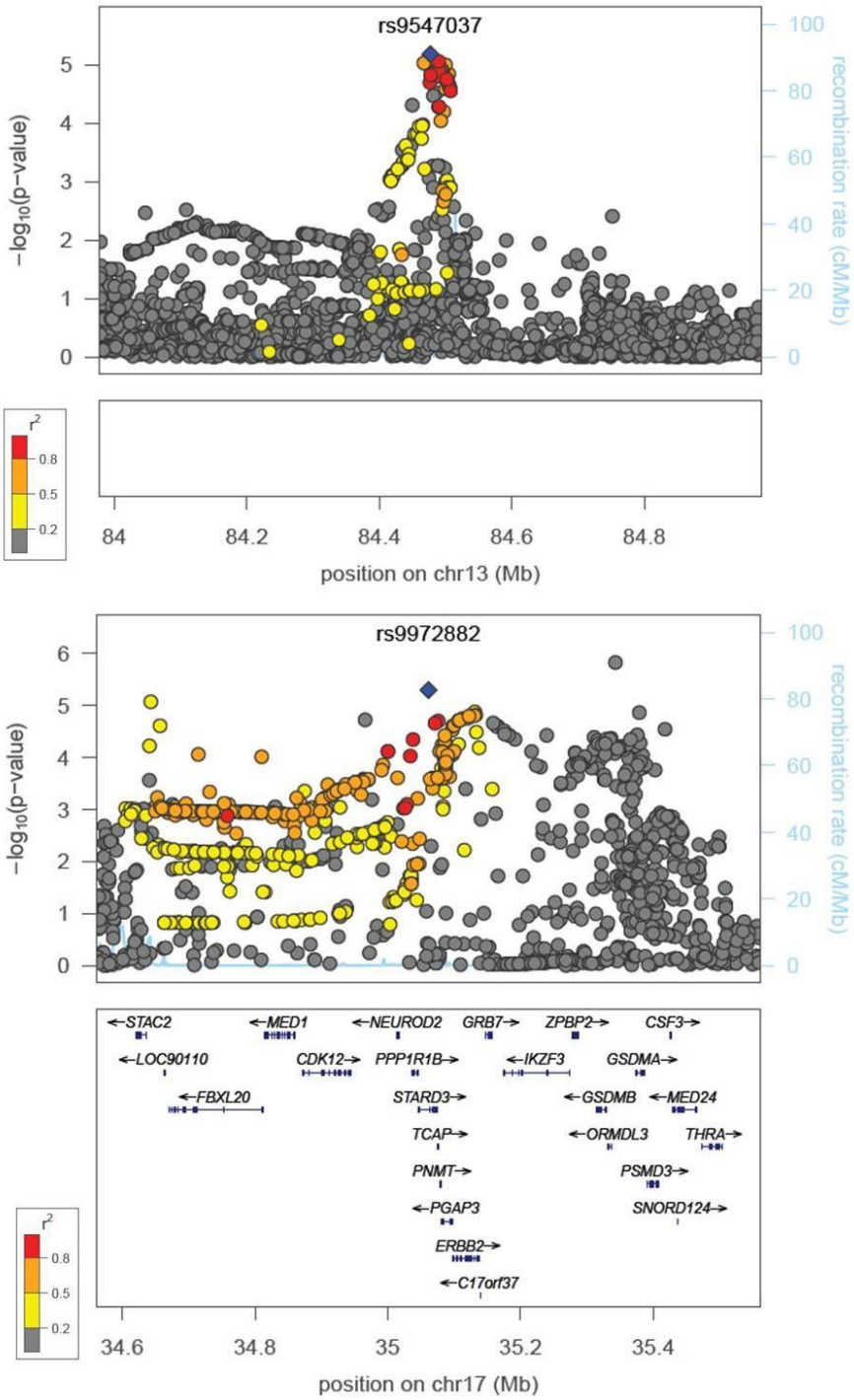


Figure 5.9 - Region plots using imputed data for suggestive loci 5 and 6 (rs2496764 and rs1810132) identified by GWA testing. Note rs9547037 is intergenic. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

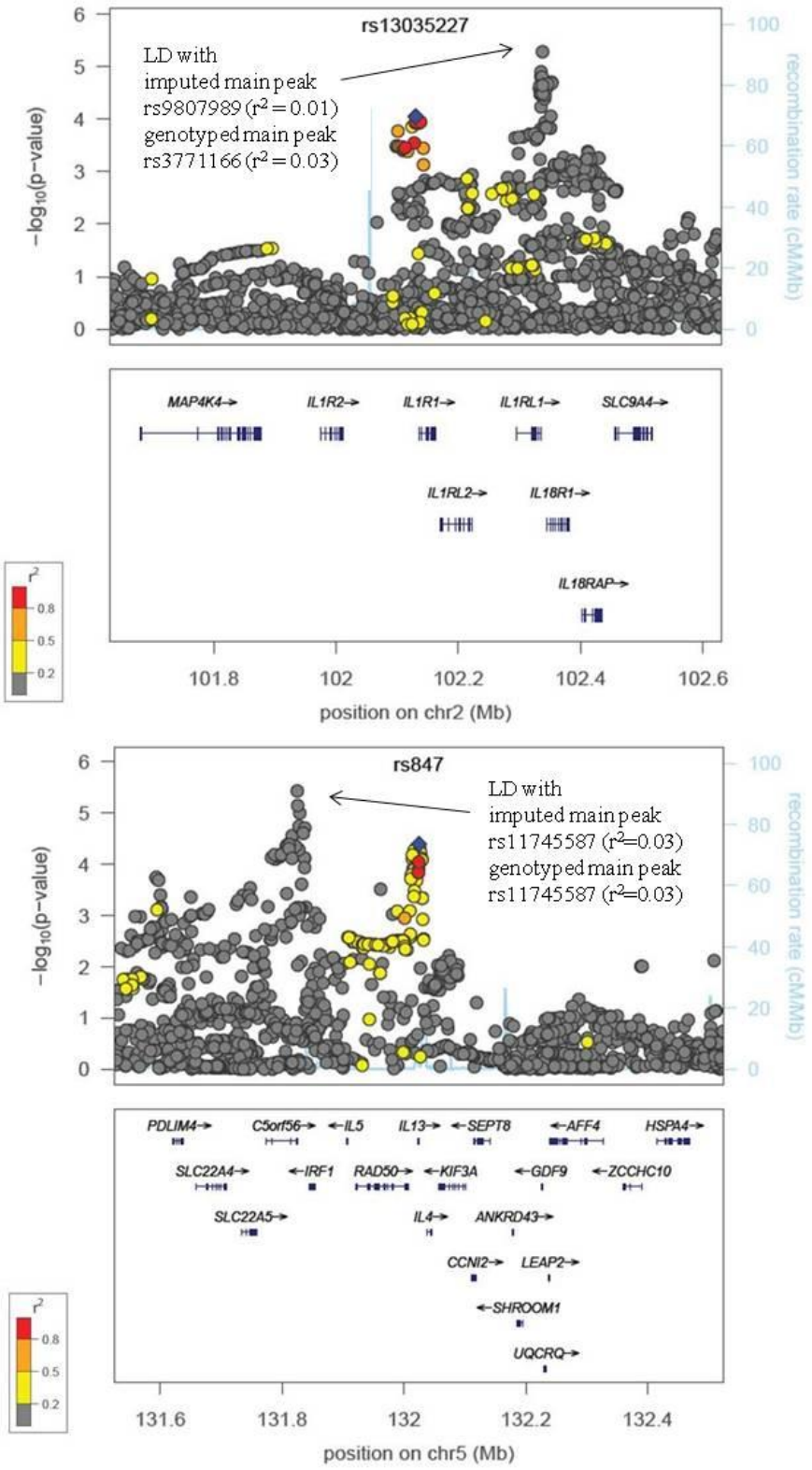


Figure 5.10 - Region plots for identified regions with secondary peaks following imputation. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

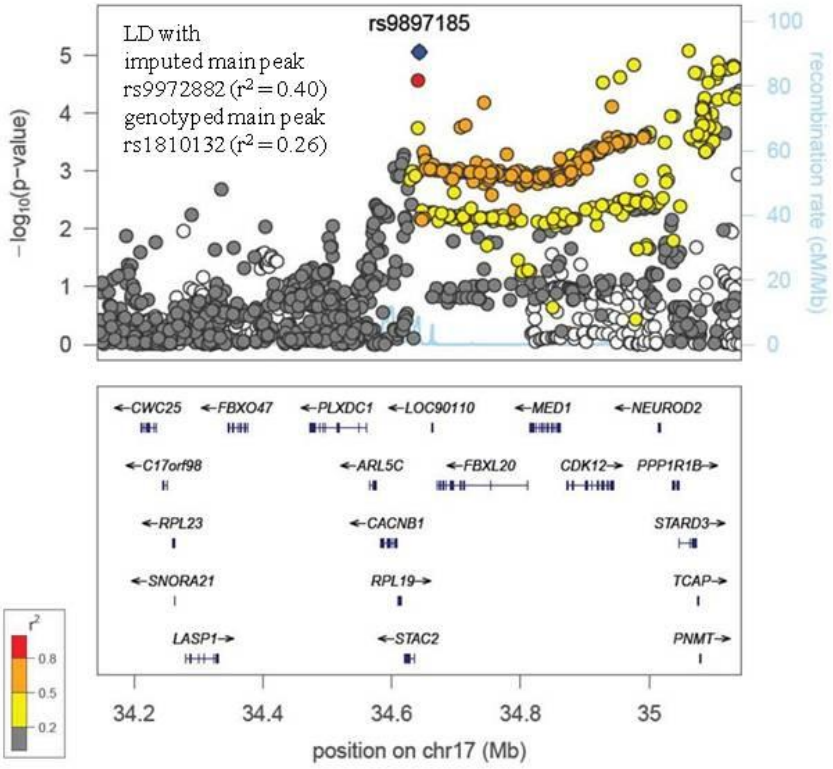


Figure 5.11 - Region plot for an identified region with a secondary peak following imputation. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs shown by their colour (see key). Fine scale recombination rate is plotted in blue.

5.4.4 Replication *In Silico*

A total of 24 SNPs were followed up using *in silico* replication in the AAGC study [Table 5.5]. This consisted of the six genotyped SNPs identified by the primary GWA study, four imputed SNPs with lower p values than genotyped SNPs in the same six regions of interest, two imputed SNPs responsible for secondary peaks in regions with known asthma associated genes, and 12 SNPs in new regions identified through imputation.

Two loci were replicated in the AAGC study cohort. The first of these was on 17q12-21 by rs4797820 within *ORMDL3* (p=0.002). The second was on 2q12.1 by a cluster of 3 SNPs: rs3771166 (p=0.001), rs9807989 (p=0.003) both within *IL18R1*, and rs13035227 within *IL1R1* (p=0.002). In non-replicated regions, a consistent direction of effect for the minor allele was seen across studies for 12 out of the 20 remaining SNPs. Following meta-analysis, the signal on *ORMDL3* met conventional genome-wide significance ($p=1.03x10^{(-8)}$) and the signal on *IL18R1* approached this threshold ($p=5.59x10^{(-8)}$) [Table 5.5].

Table 5.5 – Results for follow-up SNPs of *in silico* replication and meta-analysis. GENO denotes genotyped SNP. Significant p values p≤0.05 in replication, $p<5x10^{(-8)}$ in meta-analysis are shown in bold.

				AUGOSA (933 cases, 3,346 controls)			AAGC (231 cases, 1,345 controls)			Meta-analysis	
6 Genotyped SNPs from regions identified in genotype analysis											
Chromosome	**Gene**	**SNP**	**Position**	**r^2_{imp}**	**OR (95% CI)**	**p value**	**r^2_{imp}**	**OR (95% CI)**	**p value**	**OR (95% CI)**	**p value**
2	*IL18R1*	rs3771166	102352654	GENO	0.79 (0.71-0.88)	**$1.93x10^{(-5)}$**	GENO	0.71 (0.57-0.87)	**0.001**	0.77 (0.70-0.85)	$1.24x10^{(-7)}$
5	*C5orf56*	rs11745587	131824821	GENO	1.30 (1.17-1.45)	**$2.09x10^{(-6)}$**	GENO	1.13 (0.92-1.39)	0.25	1.26 (1.15-1.39)	$2.13x10^{(-6)}$
6	*CD83*	rs9382936	14173097	GENO	1.31 (1.17-1.48)	**$5.61x10^{(-6)}$**	GENO	1.00 (0.79-1.25)	0.98	1.24 (1.12-1.38)	$5.68x10^{(-5)}$
7	*PRPS1L1*	rs12699949	18010787	GENO	0.77 (0.69-0.87)	**$1.19x10^{(-5)}$**	GENO	0.89 (0.72-1.10)	0.27	0.80 (0.72-0.88)	$1.19x10^{(-5)}$
13	Intergenic	rs2496764	84477159	GENO	1.34 (1.18-1.52)	**$7.86x10^{(-6)}$**	GENO	0.99 (0.78-1.27)	0.96	1.26 (1.12-1.41)	$8.03x10^{(-5)}$
17	*ERBB2*	rs1810132	35119531	GENO	1.28 (1.14-1.43)	**$1.73x10^{(-5)}$**	GENO	1.07 (0.86-1.32)	0.56	1.23 (1.11-1.36)	$4.54x10^{(-5)}$
5 Imputed SNPs in regions above with a lower p value than the genotyped SNP											
2	*IL18R1*	rs9807989	102337632	0.91	0.76 (0.67-0.85)	**$5.20x10^{(-6)}$**	0.91	0.72 (0.58-0.89)	**0.003**	0.75 (0.68-0.83)	$5.59x10^{(-8)}$
7	*PRPS1L1*	rs12699948	18010735	0.92	0.75 (0.66-0.85)	**$4.84x10^{(-6)}$**	0.93	0.92 (0.74-1.15)	0.46	0.79 (0.71-0.88)	$1.44x10^{(-5)}$
13	Intergenic	rs9547037	84476839	0.90	1.38 (1.20-1.58)	**$6.60x10^{(-6)}$**	0.90	1.00 (0.78-1.29)	0.97	1.28 (1.13-1.44)	$8.01x10^{(-5)}$
17	*STARD3*	rs9972882	35061224	0.98	1.32 (1.17-1.49)	**$5.17x10^{(-6)}$**	0.99	1.07 (0.85-1.34)	0.55	1.26 (1.13-1.40)	$1.64x10^{(-5)}$
2 SNPs with $p>10^{(-5)}$ responsible for secondary peaks in regions with known asthma genes											
2	*IL1R1*	rs13035227	102130269	0.96	1.36 (1.16-1.58)	**$8.91x10^{(-5)}$**	GENO	1.53 (1.17-1.99)	**0.002**	1.40 (1.22-1.59)	$6.69x10^{(-7)}$
5	*IL13*	rs847	132024568	0.86	1.35 (1.17-1.55)	**$4.05x10^{(-5)}$**	0.94	1.12 (0.88-1.43)	0.37	1.29 (1.14-1.45)	$6.43x10^{(-5)}$
12 SNPs from new regions identified by imputation with $p<10^{(-5)}$ and r^2_{imp}=0.7 (includes 2 secondary peaks in previously identified 17q12-21 locus)											
2	Intergenic	chr2:211694960	211694960	0.70	1.76 (1.40-2.21)	**$1.27x10^{(-6)}$**	0.78	1.11 (0.73-1.70)	0.62	1.59 (1.30-1.94)	$6.80x10^{(-6)}$
5	*FLJ37543*	rs7715669	60972053	GENO	0.74 (0.65-0.84)	**$5.92x10^{(-6)}$**	GENO	0.93 (0.74-1.18)	0.56	0.78 (0.69-0.87)	$2.29x10^{(-5)}$
5	*NDFIP1*	rs6867913	141426164	0.99	0.72 (0.63-0.82)	**$1.74x10^{(-6)}$**	0.97	0.90 (0.69-1.15)	0.39	0.75 (0.67-0.85)	$3.82x10^{(-6)}$
6	*GCLC*	rs9395865	53415653	GENO	0.76 (0.67-0.85)	**$5.27x10^{(-6)}$**	GENO	1.15 (0.93-1.42)	0.19	0.84 (0.76-0.93)	$9.27x10^{(-4)}$
6	Intergenic	rs6922932	73384242	0.78	0.70 (0.60-0.82)	**$7.36x10^{(-6)}$**	0.83	1.02 (0.79-1.31)	0.88	0.78 (0.68-0.89)	$1.99x10^{(-4)}$
8	*DUSP4*	rs650230	29322727	0.78	1.37 (1.19-1.57)	**$6.40x10^{(-6)}$**	0.76	1.04 (0.83-1.30)	0.73	1.27 (1.13-1.42)	$5.74x10^{(-5)}$
9	*ACO1*	rs10970976	32423526	0.98	0.76 (0.67-0.86)	**$9.14x10^{(-6)}$**	0.96	0.84 (0.67-1.05)	0.13	0.78 (0.70-0.86)	$3.69x10^{(-6)}$
11	*OR52E4*	rs4453217	5883859	0.87	1.43 (1.25-1.64)	**$2.82x10^{(-7)}$**	0.88	0.94 (0.72-1.22)	0.63	1.31 (1.16-1.48)	$1.50x10^{(-5)}$
11	*ETS1*	rs7125574	127866404	0.93	1.34 (1.18-1.51)	**$5.28x10^{(-6)}$**	0.93	0.92 (0.73-1.16)	0.49	1.23 (1.10-1.37)	$2.13x10^{(-4)}$
17	*STAC2*	rs9897185	34642935	GENO	1.31 (1.16-1.47)	**$1.19x10^{(-5)}$**	GENO	0.95 (0.75-1.19)	0.64	1.22 (1.10-1.35)	$2.50x10^{(-4)}$
17	*ORMDL3*	rs4794820	35342870	0.94	0.76 (0.68-0.85)	**$1.52x10^{(-6)}$**	0.93	0.72 (0.59-0.89)	**0.002**	0.75 (0.69-0.83)	**$1.03x10^{(-8)}$**
19	*ZNF665*	rs16984547	58373854	GENO	1.53 (1.30-1.82)	**$5.45x10^{(-7)}$**	GENO	1.06 (0.76-1.49)	0.74	1.43 (1.23-1.66)	$3.54x10^{(-6)}$

IL18R1: interleukin 18 receptor 1; *C5orf56*: chromosome 5 open reading frame 56; *CD83*: CD83 molecule; *PRPS1L1*: phosphoribosyl pyrophosphate synthetase 1-like 1; *ERBB2*: v-erb-b2 erythroblastic leukemia viral oncogene homolog 2; *STARD3*: StAR-related lipid transfer (START) domain containing 3; *IL1R1*: interleukin 1 receptor-like 1; *IL13*: interleukin 13; *NDFIP1*: Nedd4 family interacting protein 1; *GCLC*: glutamate-cysteine ligase, catalytic subunit; *DUSP4*: dual specificity phosphatase 4; *ACO1*: aconitase 1, soluble; *OR52E4*: olfactory receptor, family 52, subfamily E, member 4; *ETS1*: v-ets erythroblastosis virus E26 oncogene homolog 1 (avian) ; *STAC2*: SH3 and cysteine rich domain 2; *ORMDL3*: ORM1-like protein 3: *ZNF665*: zinc finger protein 665.

5.4.6 Evaluation of GABRIEL Loci

All SNPs reported in the GABRIEL study for genome-wide significance and suggestive evidence for association with mild-to-moderate asthma were tested to assess the degree of association with severe asthma in the AUGOSA study. As expected, at least some evidence for association ($p \leq 0.05$) with the majority of these loci was seen, apart from rs11071559 on chromosome 15q22 (p=0.159) and rs2284033 on chromosome 22q12 (p=0.105) [Table 5.6]. Due to reduced coverage on the Illumina genotyping platform, no proxy SNP in LD was identified within the *HLA-DQ* locus. The highest significance SNPs in the AUGOSA study within GABRIEL identified loci (reported SNP ±500kb) were also identified [Table 5.7]. Region plots based on these SNPs are shown in Figures 5.12 to 5.14 (genome-wide significance) and Figures 5.15 to 5.16 (suggestive evidence).

Table 5.6 - Association results using genotyped data in AUGOSA of SNPs identified with significant or suggestive evidence of effects on the risk of asthma by the GABRIEL Consortium. Odds ratios (ORs) were calculated by designating alternative alleles (Alt) as effect alleles. Ref denotes reference allele, Alt alternative allele and CI confidence interval. We were unable to test for association with the HLA-DQ locus due to reduced coverage on the Illumina genotyping platform and no proxy SNP in LD was identified. The SNP in closest proximity was rs2187668 (32,713,862bp), p=0.26. * Opposite coding allele.

				AUGOSA (1026 cases, 3,345 controls)				GABRIEL (10,365 cases 16,110 controls)			
Chromosome	**Gene**	**SNP**	**Position**	**Ref**	**Alt**	**OR (95% CI)**	**P value**	**Ref**	**Alt**	**OR (95% CI)**	**p value**
Genome-wide significant loci ($p \leq 7.2x10^{(-8)}$) in the GABRIEL study											
2	*IL18R1*	rs3771166	102352654	G	A	0.79 (0.71-0.88)	$1.93x10^{(-5)}$	G	A	0.87 (0.83-0.91)	$3.40x10^{(-9)}$
6	*HLA-DQ*	rs9273349	32733847	-	-	-	-	T	C	1.18 (1.13-1.24)	$7.00x10^{(-14)}$
9	*IL33*	rs1342326	6180076	T	C	1.18 (1.03-1.35)	0.018	T	C	1.20 (1.13-1.28)	$9.20x10^{(-10)}$
15	*SMAD3*	rs744910	65233839	A	G	1.21 (1.09-1.35)	$2.88x10^{(-4)}$	G	A	0.89 (0.86-0.92)*	$3.90x10^{(-9)}$
17	*GSDMB*	rs2305480	35315722	G	A	0.80 (0.72-0.89)	$5.56x10^{(-5)}$	G	A	0.85 (0.81-0.90)	$9.60x10^{(-8)}$
17	*GSDMA*	rs3894194	35375519	G	A	1.25 (1.12-1.39)	$4.39x10^{(-5)}$	G	A	1.17 (1.11-1.23)	$4.60x10^{(-9)}$
22	*IL2RB*	rs2284033	35863980	G	A	0.92 (0.82-1.02)	0.105	G	A	0.89 (0.86-0.93)	$1.20x10^{(-8)}$
Suggestive loci ($p \leq 5x10^{(-7)}$) in the GABRIEL study											
5	*SLC22A5*	rs2073643	131751187	C	T	1.15 (1.04-1.28)	0.009	T	C	0.90 (0.87-0.94)*	$2.20x10^{(-7)}$
5	*IL13*	rs1295686	132023742	C	T	1.29 (1.14-1.47)	$9.30x10^{(-5)}$	T	C	0.87 (0.83-0.92)*	$1.40x10^{(-7)}$
15	*RORA*	rs11071559	58857280	C	T	0.89 (0.76-1.05)	0.159	C	T	0.85 (0.80-0.90)	$1.10x10^{(-7)}$

IL18R1: interleukin 18 receptor 1; *IL33:*interleukin 33; *SMAD3*: SMAD family member 3; *GSDMB*: gasdermin B; *GSDMA*: gasdermin A; *IL2RB*: inerleukin 2 receptor, beta; *SLC22A5*: solute carrier family 22 (organic cation/carnitine transporter), member 5; *IL13*: interkeuin 13; *RORA*: RAR-related orphan receptor A.

Table 5.7 - Highest significance SNPs in AUGOSA within GABRIEL identified loci. Genotyped SNPs are shown in bold, imputed SNPs are shown in non-bold.

Chromosome	Gene	GABRIEL SNP	AUGOSA top SNP in region (±500kb)	OR (95% CI)	p value
Genome-wide significant loci ($p \leq 7.2x10^{(-8)}$) in the GABRIEL study					
2	*IL18R1*	rs3771166	rs9807989	0.76 (0.67-0.85)	$5.20x10^{(-6)}$
6	*HLA-DQ*	rs9273349	rs9281776	1.72 (1.34-2.22)	$2.18x10^{(-5)}$
9	*IL33*	rs1342326	rs394879	0.74 (0.64-0.86)	$5.01x10^{(-5)}$
15	*SMAD3*	rs744910	chr15:65161322	1.66 (1.30-2.11)	$4.61x10^{(-5)}$
17	*GSDMB*	rs2305480	rs4794820	0.76 (0.68-0.85)	$1.52x10^{(-6)}$
17	*GSDMA*	rs3894194	rs4794820	0.76 (0.68-0.85)	$1.52x10^{(-6)}$
22	*IL2RB*	rs2284033	**rs8140025**	**1.17 (1.06-1.30)**	**0.003**
Suggestive loci ($p \leq 5x10^{(-7)}$) in the GABRIEL study					
5	*SLC22A5*	rs2073643	**rs11745587**	**1.30 (1.17-1.45)**	**$2.09x10^{(-6)}$**
5	*IL13*	rs1295686	rs847	1.35 (1.17-1.55)	$4.05x10^{(-5)}$
15	*RORA*	rs11071559	rs7173451	0.85 (0.76-0.95)	0.004

IL18R1: interleukin 18 receptor 1; *IL33*:interleukin 33; *SMAD3*: SMAD family member 3; *GSDMB*: gasdermin B; *GSDMA*: gasdermin A; *IL2RB*: inerleukin 2 receptor, beta; *SLC22A5*: solute carrier family 22 (organic cation/carnitine transporter), member 5; *IL13*: interkeuin 13; *RORA*: RAR-related orphan receptor A.

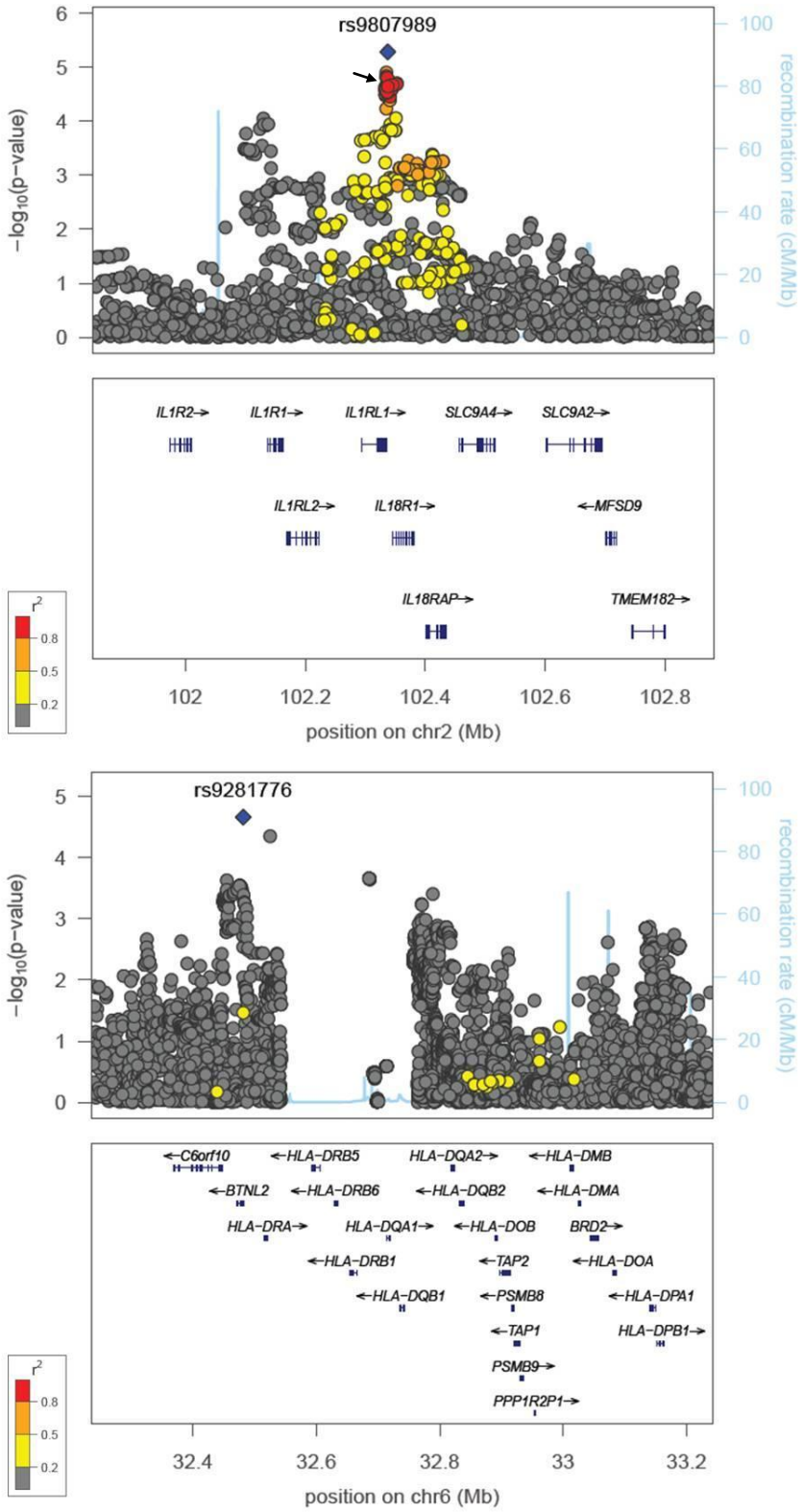


Figure 5.12- Region plots in the AUGOSA study for 9807989 and rs9281776 in genome-wide significant loci ($p \leq 7.2 \times 10^{(-8)}$) identified by the GABRIEL study. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The reported SNP in GABRIEL is shown by the arrow. The correlation (r^2) of each of the surrounding SNPs to the sentinel SNP in AUGOSA is shown by their colour (see key). Fine scale recombination rate is plotted in blue.

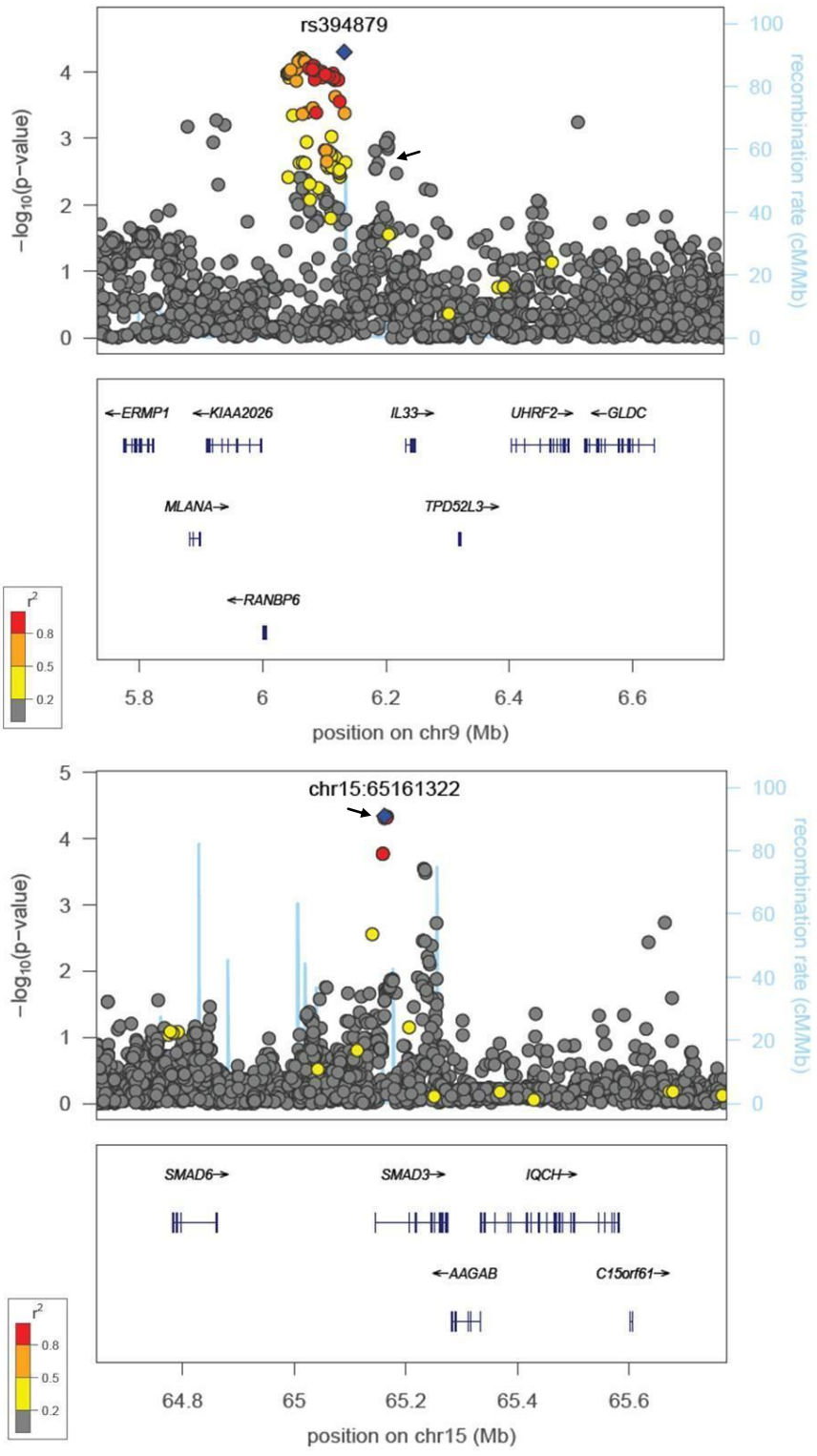


Figure 5.13 - Region plots in the AUGOSA study for rs394879 and chr15:65161322 in genome-wide significant loci ($p \leq 7.2 \times 10^{(-8)}$) identified by the GABRIEL study. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The reported SNP in GABRIEL is shown by the arrow. The correlation (r^2) of each of the surrounding SNPs to the sentinel SNP in AUGOSA is shown by their colour (see key). Fine scale recombination rate is plotted in blue.

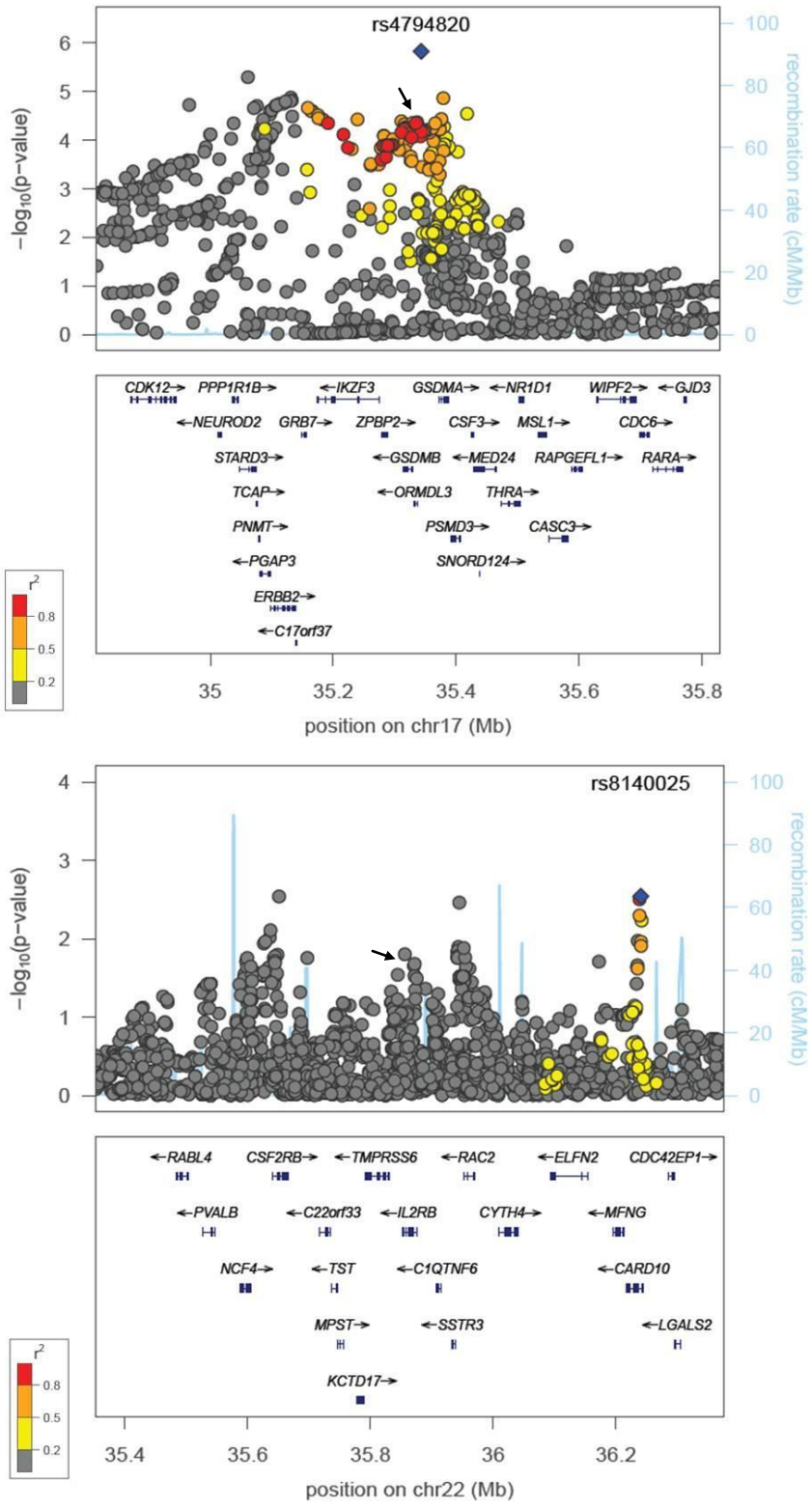


Figure 5.14 - Region plots in the AUGOSA study for rs4794820 and rs8140025 in genome-wide significant loci ($p \leq 7.2 \times 10^{(-8)}$) identified by the GABRIEL study. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The reported SNP in GABRIEL is shown by the arrow. The correlation (r^2) of each of the surrounding SNPs to the sentinel SNP in AUGOSA is shown by their colour (see key). Fine scale recombination rate is plotted in blue.

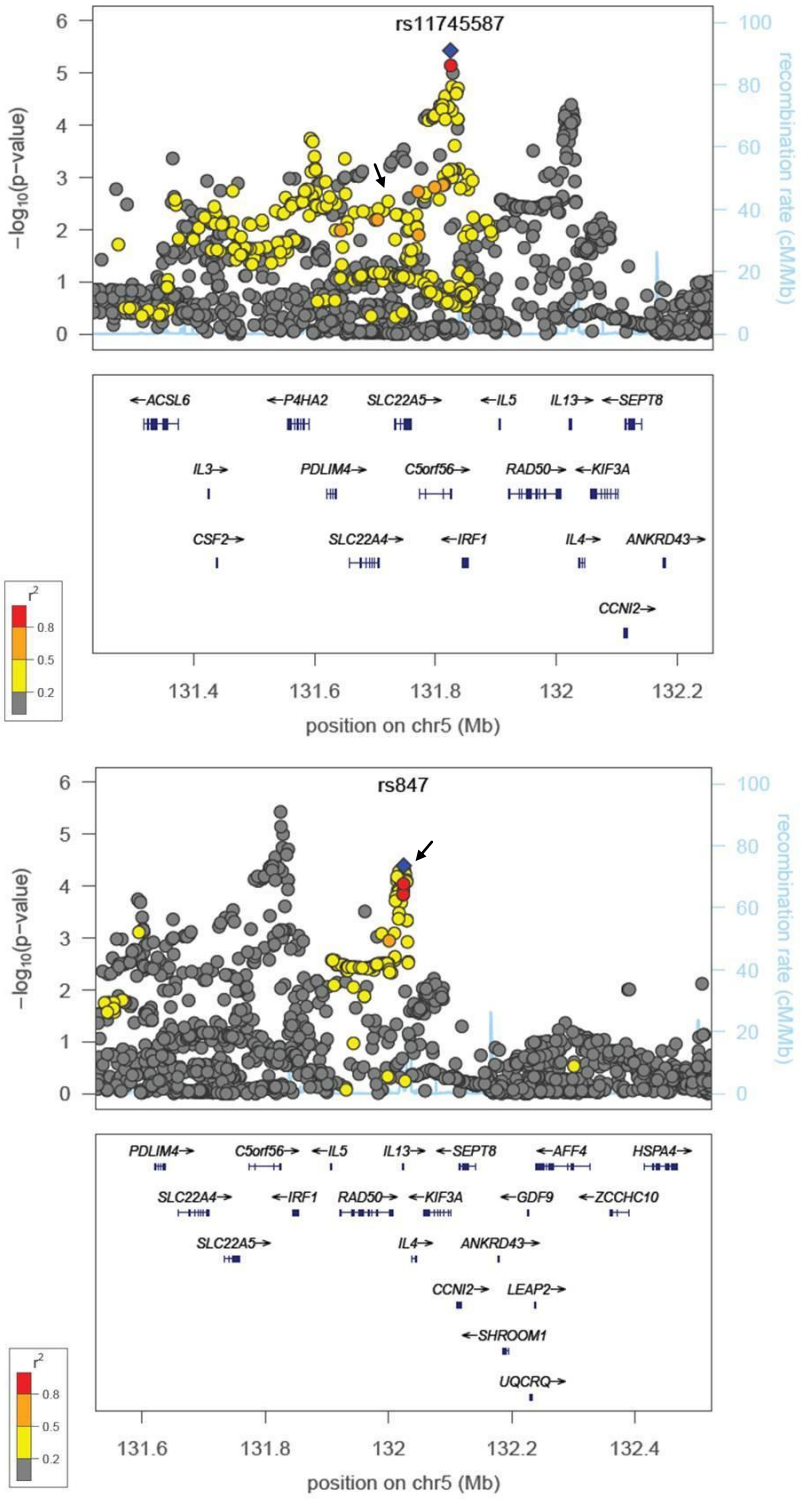


Figure 5.15 - Region plots in the AUGOSA study for rs11745587 and rs847 in suggestive loci ($p \leq 5x10^{(-7)}$) identified by the GABRIEL study. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The reported SNP in GABRIEL is shown by the arrow. The correlation (r^2) of each of the surrounding SNPs to the sentinel SNP in AUGOSA is shown by their colour (see key). Fine scale recombination rate is plotted in blue.

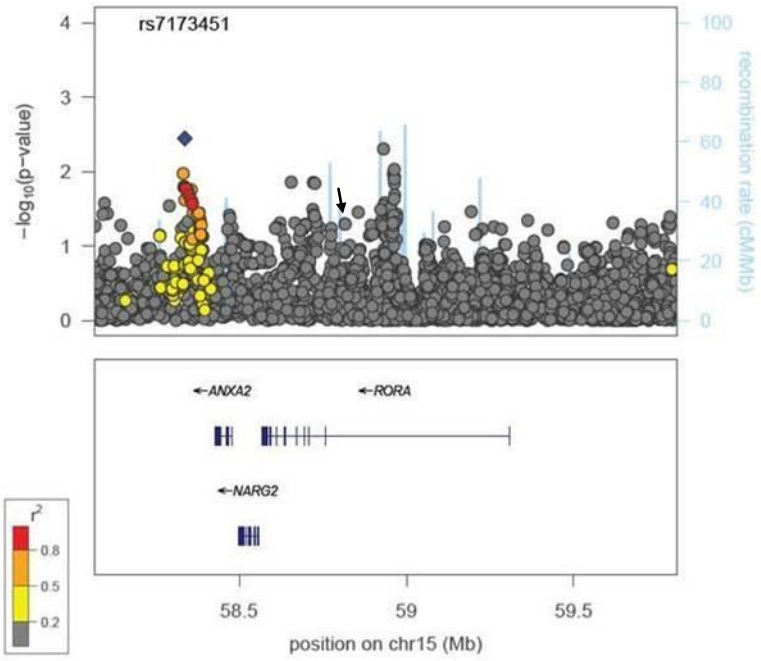


Figure 5.16 - Region plots in the AUGOSA study for rs7173451 in suggestive loci ($p \leq 5x10^{(-7)}$) identified by the GABRIEL study. Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The reported SNP in GABRIEL is shown by the arrow. The correlation (r^2) of each of the surrounding SNPs to the sentinel SNP in AUGOSA is shown by their colour (see key). Fine scale recombination rate is plotted in blue.

5.5 Discussion

This Chapter presents results from the largest severe asthma GWA study conducted to date in a cohort of 933 cases defined by GINA step 3 or above for severity and 3,346 controls without history of asthma or wheeze in order to determine if there are common genetic polymorphisms contributing to susceptibility to severe asthma.

Overall, no novel SNPs meeting genome-wide significance were identified. Further analysis of the results for polymorphisms just below this threshold was carried out to look for regions which did not meet standard genome-wide significance but had supporting evidence with at least one additional SNP with $p<5x10^{(-5)}$ within 500kb. Using this criterion, we identified six loci with suggestive evidence for association. Two of these loci, chromosomes 2q12 ($p=1.93x10^{(-6)}$) and 17q12-21 ($p=1.73x10^{(-5)}$) implicating the *IL1RL1/IL18R1* and *ORMDL3/GSDMB* loci respectively were convincingly replicated in the AAGC cohort (p=0.001 on 2q12, p=0.002 on 17q12-21). These loci have been previously reported by GWA studies for association with mild-to-moderate asthma (Moffatt *et al.* 2007; Gudbjartsson *et al.* 2009; Moffatt *et al.* 2010). The same SNP, rs3771166 on chromosome 2q12 locus was reported by the GABRIEL Consortium showing the same direction of effect for association with the minor allele as that seen in AUGOSA severe asthma subjects. The strongest signal on 17q12-21 in this study was with rs1810132: this SNP is in strong linkage disequilibrium ($r^2=0.97$) with rs2305480 reported by GABRIEL. Recently, an evaluation of four previously associated genes (*ORMDL3, PDE4D, DENND1B,* and *IL1RL1*) was carried out by the AAGC study in their total asthma cohort (986 cases, 1,846 controls) and also confirmed association of variants within *ORMDL3* and *IL1RL1* (Ferreira *et al.* 2011).

A previous GWA study in which the main phenotype was blood eosinophil counts identified additional evidence for an association with asthma for SNPs in *IL1RL1* and *IL33* (Gudbjartsson *et al.* 2009). The same SNP, rs3771166 on chromosome 2q12 locus as reported by the GABRIEL consortium showing the same direction of effect for association with the minor allele was identified within AUGOSA. A stronger protective effect size was observed for this

polymorphism within the AUGOSA severe asthma cohort (OR=0.79, 95% CI 0.71-0.88) compared to that shown in the GABRIEL study (OR=0.87, 95% CI 0.83-0.91) (Moffatt *et al.* 2010).

The linkage disequilibrium (LD) block on 2q12 contains three additional genes: *IL18R1*, interleukin 18 receptor accessory protein (*IL18RAP*), and solute carrier family 9 (sodium/hydrogen exchanger), member 4 (*SLC9A4*). Additional SNPs within these genes have been previously associated with asthma, atopic dermatitis, Crohn's disease, and Coeliac's disease (Shimizu *et al.* 2005; Hunt and al. 2008; Reijmerink *et al.* 2008; Zhernakova *et al.* 2008; Zhu *et al.* 2008). The rs3771166 SNP reported by the GABRIEL study is in high LD with other *IL1RL1* SNPs showing significant association with asthma but lies within the *IL18R1* gene. These findings, along with associated tagging SNPs within the region in LD with rs1420101: rs1974675 (r^2=0.27, p=0.025) and rs10206753 (r^2=0.96, p=0.00057) in the deCODE study, show that it is unclear which of these are the causative genes driving the observed association signals (Gudbjartsson *et al.* 2009).

With current data and the two distinct peaks seen in the region, it has not been possible to dissect the association as to whether the signal in this study lies within the *IL1RL1* or the *IL18R1* gene. However, both genes are plausible biological candidates which may play important roles in the inflammatory cascade in the pathway to asthma pathogenesis (Fukao *et al.* 2000; Schmitz *et al.* 2005). *IL1RL1* and *IL18R1* code for receptors expressed within the respiratory epithelium targeted by the IL33 and IL18 cytokines respectively. IL33 drives the production of Th2 cytokines such as IL4, IL5, and IL13 following epithelial damage (Schmitz *et al.* 2005; Carriere *et al.* 2007). IL18 acts in synergy with IL12 to induce IFNγ production and promote Th1 responses (Fukao *et al.* 2000).

Associations with asthma and SNPs located on chromosome 17q12-21 have been reported and replicated across multiple study populations (Van Den Berg *et al.* 2001; Moffatt *et al.* 2007; Galanter *et al.* 2008; Moffatt *et al.* 2010). Despite this, the region on 17q12-21 harbours a number of genes of currently poorly understood function. In combination with the difficult

linkage disequilibrium structure in this locus, it is difficult to determine which genetic variants are causal. Furthermore, this signal may be specific to childhood-onset disease (Moffatt *et al.* 2007; Moffatt *et al.* 2010). However, the culmination of findings suggests that the strongest candidates within this locus are *ORMDL3* and *GSDMB*. The role of these genes in severe asthma has also been previously reported (Halapi *et al.* 2010). The transmembrane ORMDL3 protein may have functional roles in regulation of endoplasmic reticulum-mediated calcium signalling and protein folding, in turn causing inflammation (Cantero-Recasens *et al.* 2010). It has also been suggested to be involved in cell proliferation, signal transduction and apoptosis through regulation of a component of cell membranes termed sphingolipids and their metabolism (Zheng *et al.* 2006; Breslow *et al.* 2010). *GSDMB* has been shown to be highly expressed in T cells and at low levels in the foetal lung and bronchial epithelium (Su *et al.* 2004).

An additional four loci demonstrating suggestive evidence of association with severe asthma are reported in this Chapter. Reported SNPs lie within intronic and regulatory regions of candidate genes suggesting potential functions via alternative splicing and transcriptional regulation. The most significant result on chromosome 6p23 implicates *CD83*, the gene encoding the CD83 antigen expressed on dendritic cells and which may play a role in immune modulation in the airways (Van Den Berg *et al.* 2001). The identified SNP is approximately 53kb upstream from the coding region of *CD83*. The SNP on chromosome 5q31 lies within the 3' UTR of *C5orf56*, but it is approximately 100kb downstream of the *IL4* and *IL13* genes as well as *RAD50*, a region previously reported for suggestive association with severe asthma by the TENOR study (Li *et al.* 2010). However, the region plots and results from imputation suggest that these may be two independent signals (r^2=0.03). The remaining two regions downstream of *PRPS1L1* on 7p21 and the intergenic region on 13q31 are potentially novel with unknown function and warrant further study. A single previous study reported a rare variant (MAF=0.044, $p=7.3x10^{(-5)}$) associated with rheumatoid arthritis within *PRPS1L1* (Morris *et al.* 2009). Although none of these loci were replicated in the AAGC cohort, all SNPs apart from that within the intergenic region on 13q31 showed the same direction of effect for the minor allele as found within AUGOSA.

No published study has yet analysed imputed genotypes using data released from the 1000 Genomes Project. Results from this analysis have provided strong support for all six association signals identified using genotyped data. A total of five imputed SNPs within identified regions had lower p values than genotyped SNPs on 2q12.1, 6p23, 7p21.1, 13q31.1, and 17q12-21 as well as three SNPs causing secondary peaks within 2q12.1, 5q31.1, and 17q12-21, regions known to contain asthma associated genes. Although, these shadow signals may have been synthesised through imputation, these findings may also suggest that further fine mapping is required within each locus to identify the causative variants. Again, results in the AAGC cohort did not replicate these findings but showed consistent SNP effects. Results from imputation also identified 12 SNPs meeting statistical significance in new regions with reasonable supporting evidence ($p<10^{(-5)}$, $r^2_{imp}>0.7$) on 2q34, 5q12.1, 5q31.3, 6p12.1, 6q13, 8p21.1, 9p21.1, 11p15.4, 11q24.3, and 19q13.41. Similarly, the functional implications of these SNPs are unclear as they do not lie in coding regions resulting in amino acid changes. However, the 19q13 locus has previously been shown to be associated with atopy (Holloway *et al.* 2010).

Replication results were more variable for these regions showing inconsistent effects suggesting that there are overall, low supporting evidence for these regions. However, power calculations demonstrated that this smaller replication cohort of 231 cases and 1,345 controls would have had relatively low power to detect modest effects of SNPs within these regions. Meta-analysis of SNP results between the AUGOSA and AAGC cohorts found nominal statistical significance for all SNPs followed up ($p\leq0.05$). P values were similar to that in the AUGOSA cohort and were most likely to have been driven by the initial signals from the GWA study.

In addition, the contribution of additional previously identified asthma susceptibility loci was assessed in patients with severe disease reported by the recent GABRIEL study (Moffatt *et al.* 2010). Since this GWA study is currently the largest published association study investigating the genetic determinants of asthma, with sufficient statistical power to identify common genetic variants contributing to asthma susceptibility, this second analysis aimed to determine

if these signals also contribute to disease susceptibility in this severe asthma cohort. It is important to note that although individuals used in the AUGOSA analyses are a subset of the total case subjects used in GABRIEL, they only constitute a relatively small proportion (8.94%). Therefore, an association signal specific for mild-to-moderate asthma would be unlikely to originate only from and be confounded by the group of severe asthmatics alone, thereby validating this approach.

For genome-wide significant results when compared to mild-to-moderate asthma from GABRIEL, effect sizes for the alternative allele are greater in this cohort of severe asthmatics for rs3771166 on chromosome 2q12: the observed OR for the protective allele was 0.79 compared to 0.87 in GABRIEL, for rs2305480 on chromosome 17q12-21: the observed OR for the protective allele was 0.80 for compared to 0.85, for rs3894194 on chromosome 17q12-21: the OR for the risk allele was 1.25 compared to 1.17. Further to this, this study has confirmed weaker association with *IL33* and *SMAD3* chromosomes 9p24 ($p=1.80x10^{(-2)}$) and 15q22 ($p=2.88x10^{(-4)}$). The more distal signal seen in AUGOSA in the 9q24 region may indicate different regulatory control of the *IL33* gene in severe compared to mild-to-moderate asthma.

The explanations behind the increased effect size of risk alleles in severe asthma are potentially two-fold. The most likely explanation is that the contribution of genetic effects driven by variation in these genes is greater in more severe patients. However, it is possible that in populations with milder subjects, misclassification of cases and controls may result in an underestimate of true effect sizes. Furthermore, differences in population sizes used in analyses may also reflect our varying ability to determine robust effect sizes.

Although the current study has been the largest effort so far to determine genetic determinants of severe asthma, the numbers of subjects used is still a limitation in being able to generate enough statistical power to detect all variants with modest effects. Whilst these results suggest that it is probable to exclude major effects being driven by a single gene as a risk for severe asthma, this data based primarily on genotyped SNPs suggests that there may be a number of loci which may be specific for severe asthma but with relatively small overall contributions to

the risk of developing severe disease. The obvious solution to resolving this issue is to undertake additional replication studies in severe asthma populations, but these populations by their very definition are hard to recruit: the current study included subjects recruited from eight major centres in the UK and hence obtaining a suitable replication population to take this work forward will require additional international efforts to establish suitably large populations with severe disease. The only other existing cohort of severe asthmatic subjects is the TENOR study with relatively few individuals (n=473) (Li *et al.* 2010).

In summary, results presented in this Chapter provide evidence to support a potentially enhanced role for known genetic risk factors for asthma, for example, polymorphisms in the *IL1RL1*/*IL18R1* and *ORMDL3*/*GSDMB* loci in the development of severe disease. These have highlighted the importance of candidate genes both in pathways initiating Th2 inflammation in response to epithelial damage as well as in down-regulation of airway inflammation and remodelling in determining asthma susceptibility irrespective of severity. A number of novel loci have been identified which may be specific to the development of severe as opposed to milder forms of asthma, but these require further analyses in addition to testing in severe asthmatics compared to individuals with mild-to-moderate disease as opposed to the controls used in the analyses reported herein.

Chapter 6

Evaluation of the Relative Contribution of Previously Reported Candidate Loci in Severe Asthma

6.1 Introduction

As summarised in Chapter 1, a large number of candidate genes have been previously reported in the literature to determine susceptibility to asthma defined using multiple approaches. These have been identified through a number of methods including linkage and positional cloning, candidate gene association and recently, genome-wide association. Replicated findings have been seen for a number of candidate genes for example, toll-like receptor (*TLR2*) (4q32) (Qian *et al.* ; Eder *et al.* 2004; Bjornvold *et al.* 2009; Smit *et al.* 2009), *IL18* (11q22.2-q22.3) (Higa *et al.* 2003; Imboden *et al.* 2006; Lee *et al.* 2006; Lachheb *et al.* 2008; Zhu *et al.* 2008), and *IFNG* (12q14) (Nakao *et al.* 2001; Nagarkatti *et al.* 2002; Wang *et al.* 2006; Kumar and Ghosh 2008) which are involved in the inflammatory pathways and immune regulation.

The results shown for many of these candidate genes suggest that they may play a real role in asthma susceptibility and may explain some of the missing heritability component not identified by GWA studies. However, the growing number of genetic association studies carried out in addition to limited consistency mean that it is still difficult to interpret these findings. Reasons for this include publication and study bias towards positive associations, inadequate power from individual studies due to sample size or phenotyping, genetic heterogeneity between independent study populations, and inadequate correction for multiple testing resulting in both type I and II errors (see Chapter 1: Section 1.4.1). Using data from larger, genome-wide association studies may increase power to detect marginal effects of previously identified candidate genes. To date, a select number of studies have used this approach in complex diseases including asthma (Gabriel *et al.* 2002; Feulner *et al.* 2009; Rogers *et al.* 2009; Wu *et al.* 2010).

Given the results of the GWA study described in Chapter 5, it seems that there is unlikely to be a select small number of novel susceptibility genes with major effects that are responsible for determining severe asthma specifically. However, it is pertinent to note that the current genotyping platforms used do not have complete, uniform coverage across the genome. This means that there may be additional loci that have not been adequately tested in poorly covered intervals as well as rare SNPs and structural variations such as copy number variants (CNVs) (Barrett and Cardon 2006; Barrett *et al.* 2008). The GABRIEL Consortium study observed genome-wide significance between asthma and SNPs within previously reported loci and genes including *IL18R1*, *HLA-DQ*, *IL33*, and chromosome 17q12-21, the latter specific to childhood-onset disease (Moffatt *et al.* 2010). The contribution of these loci to severe disease were assessed in AUGOSA, but a large number of widely replicated genes and loci were not evaluated directly using the approach outlined in Chapter 5, such as *IL4R* (16p12.1-p11.2), *TGFB* (19q13.1), and *ADAM33* (20p13) described in detail in Chapter 1: Section 1.3 (Ober and Hoffjan 2006).

6.2 Aim

The aim of this Chapter was to evaluate within the AUGOSA severe asthma cohort the contribution of candidate genes previously identified in the literature for association with mild-to-moderate asthma.

6.3 Methods

6.3.1 Literature Search

A comprehensive, systematic review of previously published literature in asthma genetics was carried out updated to the 1st January 2011. An inclusive criterion was used to evaluate results from all studies in asthma (atopic, non-atopic, aspirin-induced, occupational, severe and unspecified) regardless of age-of-onset (childhood or adult) and population ancestry (all populations included). Reported loci (genes, SNPs, and microsatellite markers) were recorded for showing positive association in candidate gene studies (nominally defined) and only conventional genome-wide significance in GWA studies ($p<5x10^{(-8)}$) or GWL studies (LOD≥3.6).

Using PubMed, a search was conducted within the following search string limited to humans and publications in English:

"asthma" AND "association" AND "SNP" OR "polymorphism" OR "variant" OR "gene" OR "genetic" OR "genom*"

Results were summarised in a table in MS Excel using the following variables:

Gene, SNP or microsatellite marker, location, effect & baseline alleles, effect size, significance, study design, reference, associated phenotypes, study population (ethnicity, age, sample size), any other notes.

In order to accurately summarise supporting evidence for individual genes and avoid duplication of identified candidate genes reported under different names or symbols in the literature, all gene names were standardised using the HUGO Gene Symbol online database (www.genenames.org). Genes identified on sex chromosomes were excluded.

6.3.2 Defining Loci Co-ordinates

Gene co-ordinates corresponding to the transcriptional start and end sites for each gene were ascertained using the USCS Gene tracks dataset on the USCS Gene tables genome browser (http://genome.ucsc.edu/cgi-bin/hgTables) and selecting genome: Human, assembly: Mar 2006

(NCBI36/hg18). For most Gene and Gene Prediction tracks, any individual gene will have one or more associated transcripts. Therefore, the canonical transcript was selected for each gene to include the longest or most 5' reaching transcript.

Identified microsatellite names were searched using the NCBI UniSTS search engine (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unists) to identify the reference sequence co-ordinates.

Regions were defined by ±10kb of identified gene co-ordinates, SNP positions and microsatellite co-ordinates to allow a margin for error in gene co-ordinate positions and in order not to miss out any SNPs with significant associations just upstream or downstream of each gene. This also helped to include any additional SNPs in LD from which the true signal may be originating.

Three lists of regions were compiled to carry out three different analyses: 1) to evaluate all identified regions, 2) to exclude results from GWA studies, and 3) to evaluate loci reported by candidate gene studies and GWL studies supported by three or more publications only.

6.3.3 Region Evaluation

Each list of regions was evaluated by extracting results from GWA testing for all genotyped SNPs within the region in the AUGOSA study (see Chapter 5 for details). These were identified using a script written in R version 2.12.1.

Co-ordinates in a number of identified regions overlapped, therefore duplicate SNPs within each extracted results files were removed. For the second and third analyses involving exclusion of identified regions, both SNPs within defined region co-ordinates and additional SNPs in LD with r^2>0.3 lying outside the ±10kb criteria were excluded. The test inflation statistic (λ) was calculated for each list of regions (see Chapter 2: Section 2.4.3). Manhattan and Quantile-Quantile (Q-Q) plots were then generated (see Chapter 2: Section 2.3.2).

6.3.4 Region and SNP Analyses

The threshold significance level was determined for each analysis using both the Bonferroni corrected p value at the 5% significance level assuming that each SNP remains an independent test and a MeffLi corrected p value taking into account the extent of LD between tested SNPs (see Chapter 2: Section 2.4.8). MeffLi variables indicating the number of effective independent variables were calculated using SNPSpD. Relevant data columns were extracted using R version 2.12.1 to create .pre and .map SNPSpD input files for SNPs extracted from all three sets of analyses. The .pre file contained the results per individual created using the data.txt file; the columns of this file contained the following information: Family ID (0=unknown), Individual ID, Paternal ID (0=unknown), Maternal ID (0=unknown), Sex (1=male, 2=female, 0=unknown), Genotype_1, … Genotype_n. SNPSpD was run using the following command line in UNIX:

```
./SNPSpD.sh region_lookup_results
```

6.4 Results

6.4.1 Systematic Literature Review

As of 1st January 2011, the systematic literature search identified a total of 1,830 papers. A total of 303 papers met criteria for inclusion in this study [Table 6.1]. Studies in over 19 populations were recorded, shown in Figure 6.1. Results from 1,527 papers were not recorded for the following reasons:

227 Negative reported results only

24 Severity analyses within asthmatic cases only

36 Haplotype-based analyses only

10 Meta-analyses of previous publications

1 Candidate gene look-up

1,229 Irrelevant or reviews

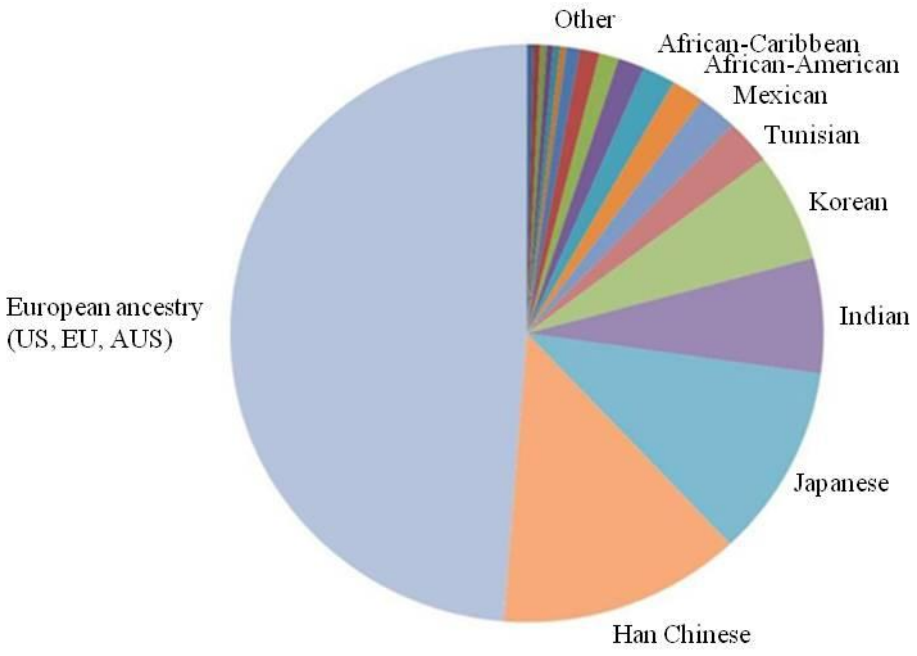


Figure 6.1 - Recorded papers categorised by study population. Other includes Colombian (0.36%), Greek (0.36%), Malay (0.36%), Puerto-Rican (0.36%), Thai (0.36%), Venezuelan (0.36%), Hutterite (0.73%), Egyptian (1.09%), and Iranian (1.09%).

Table 6.1 - Recorded papers categorised by study design, phenotype, European or non-European ancestry, and age of participants. Results were recorded from 303 papers identified for reporting positive association in candidate gene studies (nominally defined) and only conventional genome-wide significance in GWA studies ($p<5x10^{(-8)}$) or GWL studies (LOD≥3.6).

		All populations	European ancestry	Non-European ancestry
Study design	Case-control association	293	142	151
	GWA	6	3	3
	GWL	4	1	3
Cohort	Adult	150	66	84
	Children	79	36	43
	Both	40	30	10
	Unspecified	34	14	20
Asthma phenotype	Atopic	57	19	38
	Non-atopic	5	3	2
	Aspirin-induced	10	0	10
	Occupational	4	3	1
	Severe	4	3	1
	Asthma unspecified	223	118	105

6.4.2 Identified Genes

From all recorded papers, a total of 170 genes were identified and stratified by the number of supporting publications. Stratifying by the type of variation studied, 269 papers reported SNP associations only, 32 reported insertion/deletion polymorphisms (indels) and two reported copy number variations (CNVs). Three genes on sex chromosomes reported by two papers were excluded; these were *TLR8*, *TIMP1*, and *CYSLTR1*. In addition, two SNPs and six microsatellites were also reported resulting in a total of 175 loci (±10kb) for the look-up. A table showing the full list of identified loci, transcriptional start and end co-ordinates, defined regions co-ordinates, and the number of supporting references for the locus is provided in Appendix 6.

6.4.3 Loci Evaluation Results

Three sets of analyses were carried out to evaluate 1) all 175 identified loci, 2) 159 loci excluding results from GWA studies, and 3) 33 loci reported by candidate gene studies and GWL studies supported by three or more publications. Results for each set of analyses are listed below detailing the numbers of tested SNPs across identified loci, defined significances, and test inflation statistics (λ). Manhattan and Q-Q plots comparing all results are shown in Figure 6.1.

6.4.3.1 All Identified Loci

A total of 3,738 unique genotyped SNPs were tested within 175 identified loci (167 genes, two SNPs, six microsatellite regions). The test inflation statistic (λ) for this analysis was 1.17. The defined levels of statistical significance described above were $p=1.34x10^{(-5)}$ as determined by the Bonferroni correction and $p=2.37x10^{(-5)}$ as determined by the MeffLi correction.

Two SNPs met Bonferroni corrected significance: rs1837253 within thymic stromal lymhopoietin (*TSLP*) on 5q22.1 ($p=5.52x10^{(-6)}$) and rs3771166 within interleukin 18 receptor 1 (*IL18R1*) on 2p12.1 ($p=1.93x10^{(-5)}$). An additional four SNPs met MeffLi defined significance: rs10192157 ($p=2.34x10^{(-5)}$), rs10204137 ($p=2.37x10^{(-5)}$), rs10206753 ($p=2.49x10^{(-5)}$) all within *IL1RL1/IL18R1* on 2p12.1, and rs2243204 within *IL13* on 5q31.1 ($p=5.71x10^{(-5)}$) [Table 6.2]. SNP rs3771166 (*IL18R1*) is the same SNP as reported by the GABRIEL study and examined in Chapter 5 tagged by the additional 3 SNPs ($r^2=0.966$). SNP rs2243204 is in weak LD with the *IL13* SNP rs1295686 reported by the GABRIEL study ($r^2=0.464$). The next 10 genes containing SNPs just below defined statistical significance are listed in Table 6.3.

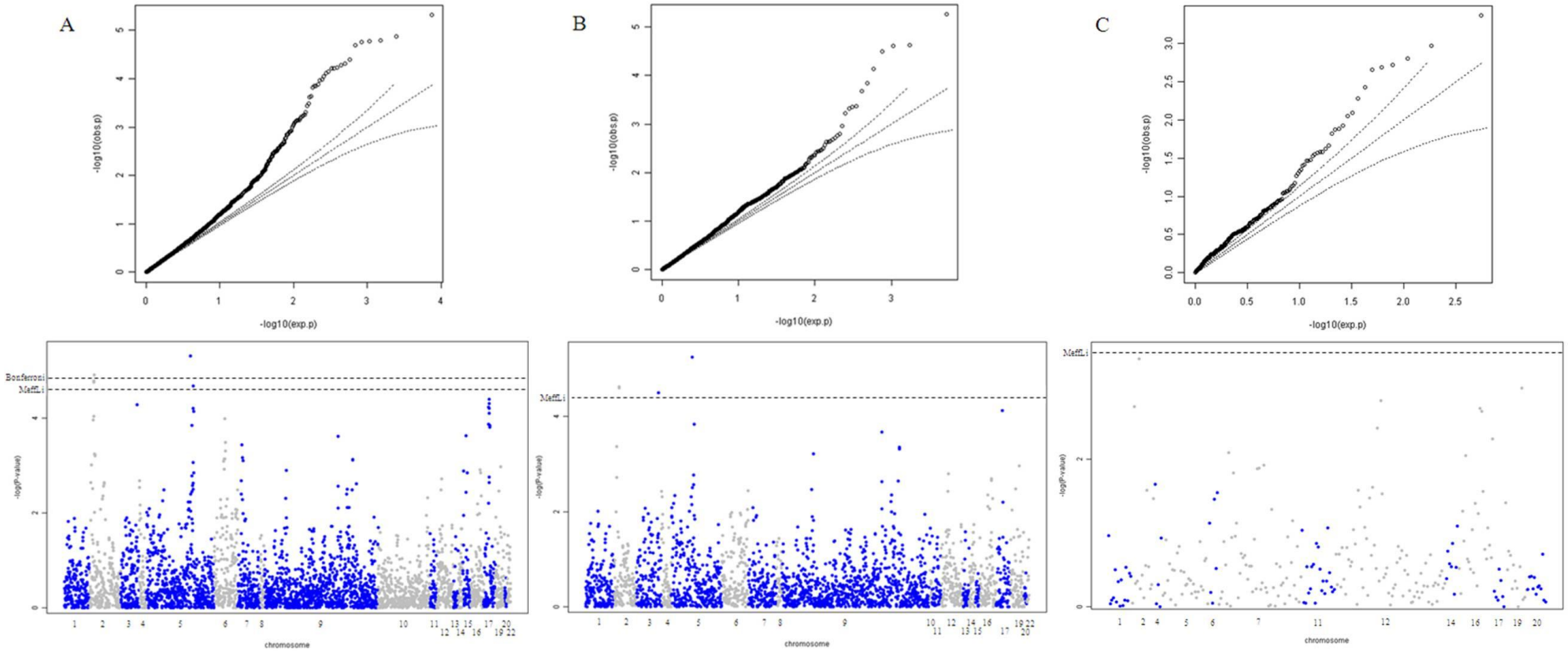


Figure 6.1 – Results of look-up analyses for A) all 175 identified loci, B) 159 loci excluding results from GWA studies, and C) 33 loci reported by candidate gene studies and GWL studies supported by three or more publications only. Q-Q plots of observed versus expected $-\log_{10}$ p values for all tested SNPs. Manhattan plots of $-\log_{10}$ p values for tested SNPs plotted against genomic position. * No SNPs on chromosomes 18 and 21 were tested for A) and B), only SNPs on chromosomes 1, 2, 4-7, 11, 12, 14, 16, 17, 19, and 20 were tested for C).

6.4.3.2 Loci Excluding Results from GWA Studies

A total of 2,800 unique genotyped SNPs identified within 159 loci (152 genes, one SNP, six microsatellites) remained after eliminating results from GWA studies. A further 159 SNPs in remaining loci tagging SNPs (r^2>0.3) in eliminated regions were removed leaving 2,641 SNPs for analysis. The test inflation statistic (λ) for this analysis was 1.19. The defined levels of statistical significance described above were $p=1.89x10^{(-5)}$ as determined by the Bonferroni correction and $p=3.17x10^{(-5)}$ as determined by the MeffLi correction.

No SNPs met Bonferroni corrected significance. However, rs1837253 ($p=5.52x10^{(-6)}$) within *TSLP* and two SNPs within *IL1RL1*: rs10204137 ($p=2.37x10^{(-5)}$), rs10206753 ($p=2.49x10^{(-5)}$) met MeffLi defined significance. All SNPs were also identified by the first set of analyses in all loci [Table 6.2].

6.4.3.3 Replicated Loci Excluding Results from GWA Studies (≥3 Candidate Gene and GWL studies)

A total of 323 unique genotyped SNPs were identified within 33 loci (33 genes) with three or more references from the list of regions tested in the second set of analyses. A further 29 SNPs in the 33 loci tagging SNPs (r^2>0.3) in eliminated loci were removed leaving 294 SNPs for analysis. The test inflation statistic (λ) for this analysis was 1.57. The defined levels of statistical significance described above were $p=1.70x10^{(-4)}$ as determined by the Bonferroni correction and $p=2.78x10^{(-4)}$ as determined by the MeffLi correction.

No SNPs met Bonferroni corrected significance or MeffLi defined significance following this analysis. The most statistically significant SNP was rs3917254 ($p=4.35x10^{(-4)}$), also contained within *IL1RL1* but not in LD with any of the above identified SNPs in the same locus (r^2=0.003) [Table 6.2].

Table 6.2 –Results of statistically significant SNPs from look-up analyses for 1) all 175 identified loci and 2) 159 loci excluding results from GWA studies. Ref denotes reference allele; effect refers to the minor allele.

Chromosome	Gene	SNP	Position	Ref	Effect	OR (95% CI)	p value
1) All identified loci							
2	*IL1RL1*	rs10192157	102334788	G	A	0.79 (0.71-0.88)	$2.34x10^{(-5)}$
2	*IL1RL1*	rs10204137	102334644	A	G	0.79 (0.71-0.88)	$2.37x10^{(-5)}$
2	*IL1RL1*	rs10206753	102334794	A	G	0.79 (0.71-0.88)	$2.49x10^{(-5)}$
2	*IL18R1*	rs3771166	102352654	G	A	0.79 (0.71-0.88)	$1.93x10^{(-5)}$
5	*TSLP*	rs1837253	110429771	G	A	0.75 (0.66-0.85)	$5.52x10^{(-6)}$
5	*IL13*	rs2243204	132027393	G	A	1.48 (1.24-1.78)	$5.71x10^{(-5)}$
2) Loci excluding results from GWA studies							
2	*IL1RL1*	rs10204137	102334644	A	G	0.79 (0.71-0.88)	$2.37x10^{(-5)}$
2	*IL1RL1*	rs10206753	102334794	A	G	0.79 (0.71-0.88)	$2.49x10^{(-5)}$
5	*TSLP*	rs1837253	110429771	G	A	0.75 (0.66-0.85)	$5.52x10^{(-6)}$

IL1RL1: interleukin 1 receptor-like 1; *IL18R1*: interleukin 18 receptor 1; *TSLP*: thymic stromal lymphopoietin; *IL13*: interleukin 13.

Table 6.3 – Pivotal SNPs in top 10 genes just below statistical significance in analysis of all identified loci. Ref denotes reference allele; effect refers to the minor allele.

Chromosome	Gene	SNP	Position	Ref	Effect	OR (95% CI)	p value
3	*MYLK*	rs1343700	125054444	A	G	1.25 (1.12-1.39)	$3.20x10^{(-5)}$
5	*IRF1*	rs2548993	131836768	A	G	1.25 (1.11-1.40)	$1.46x10^{(-4)}$
5	*RAD50*	rs17772583	131981409	A	G	0.81 (0.71-0.92)	$8.67x10^{(-4)}$
6	*BTNL2*	rs3817963	32476065	A	G	1.25 (1.12-1.39)	$2.86x10^{(-4)}$
7	*NPSR1*	rs425990	34668324	A	G	1.21 (1.09-1.34)	$5.79x10^{(-4)}$
9	*PTPRD*	rs7873669	9876403	A	C	1.75 (0.64-0.87)	$2.13x10^{(-4)}$
15	*SMAD3*	rs744910	65233839	A	G	1.22 (1.10-1.35)	$2.88x10^{(-4)}$
17	*ZPBP2*	rs11557467	35282160	A	C	1.24 (1.12-1.38)	$7.18x10^{(-5)}$
17	*GSDMB*	rs2290400	35319766	G	A	1.25 (1.12-1.39)	$5.05x10^{(-5)}$
17	*ORMDL3*	rs7216389	35323475	G	A	1.24 (1.12-1.38)	$5.73x10^{(-5)}$

MYLK: myosin light chain kinase; *IRF1*: interferon regulatory factor 1; *RAD50*: DNA repair protein RAD50; *BTNL2*: butyrophilin-like 2 (MHC class II associated); *NPSR1*: neuropeptide S receptor 1; *PTPRD*: protein tyrosine phosphatase, receptor type, D; *SMAD3*: SMAD family member 3; *ZPBP2*: zona pellucida binding protein 2; *GSDMB*: gardermin B; *ORMDL3*: interleukin 13.

6.4.4 Examination for Region Support

Supporting evidence within the region for identified loci containing significant SNPs was assessed. Both the chromosome 2p12.1 and 5q31.1 loci were reported by the GABRIEL study and evaluated previously in the AUGOSA study (see Chapter 5 for details). Low supporting evidence is seen for the *TSLP* locus and this was confirmed by no other SNPs within this locus reaching $p<5x10^{(-3)}$ within results for all regions tested [Figure 6.3].

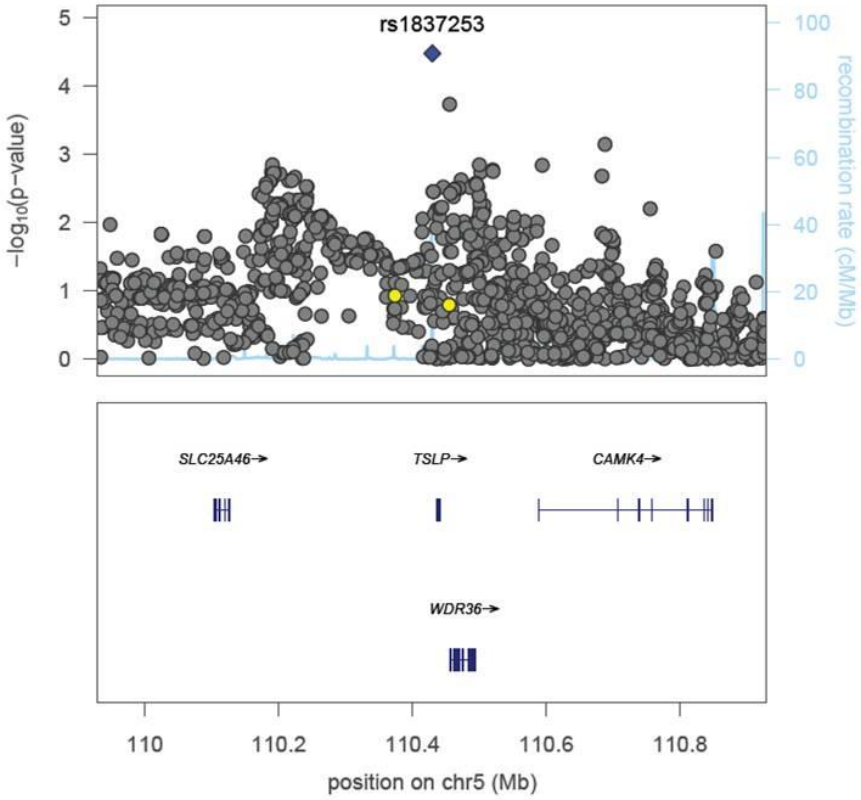


Figure 6.3 - Region plots for rs1837253 (*TSLP*, 5q22.1). Statistical significance of each SNP on the $-\log_{10}$ scale is shown as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and the correlation (r^2) of each of the surrounding SNPs to the sentinel SNP is shown by their colour (see key). Fine scale recombination rate is plotted in blue.

6.4.5 Previously Reported SNPs within the 2q12 Locus

The strongest candidate gene which has not been previously identified by GWA studies was *IL1RL1* on 2q12. All six papers in the literature reporting positive association to the 2q12 locus were reviewed to evaluate consistency of findings across studies and further analysis was carried out to test association of reported SNPs in the AUGOSA study [Table 6.4]. The variable number tandem repeat (VNTR) polymorphism reported by Settin *et al.* and Zeyrek *et al.* was not tested; both rs447713 and rs3087271 reported by Pattaro *et al.* were not typed on the AUGOSA genotyping platforms. All reported SNPs were also significantly associated with severe asthma within AUGOSA ($1.93x10^{(-5)}<p<1.98x10^{(-3)}$) apart from rs1041973 reported by Wu *et al* (p=0.236) showing the same direction of effect for the tested alllele.

Table 6.4 - Association results in AUGOSA of positively associated SNPs reported on chromosome 2q12. Odds ratios (ORs) were calculated by designating alternative alleles (Alt) as effect alleles. Ref denotes reference allele. All six papers in the previous literature reporting positive association to the chromosome 2q12 locus were reviewed to evaluate consistency of findings across studies and results for published SNPs within the AUGOSA cohort.

Study				**Polymorphism**						
Reference	**Sample size**	**Ethnicity**	**Age**	**Gene**	**SNP**	**Position**	**Ref**	**Alt**	**OR (95% CI)**	**p value**
Gudbjartsson *et al.* 2009	7,996 cases, 44,890 controls	9 European, 1 East Asian sets	Adult	*IL1RL1*	rs1420101	102324148	G	A	1.16 (1.11-1.21)	$5.50x10^{(-12)}$
AUGOSA							G	A	1.22 (1.10-1.36)	$1.77x10^{(-4)}$
Settin e*t al*. 2008	69 cases, 98 controls	Egyptian	Children	*IL1R1*	VNTR	Intron 2	240bp	410bp	7.5 (3.4-14)	<0.05
AUGOSA							Not tested			
Zeyrek *et al*. 2008	328 cases, 246 controls	Turkish	Children	*IL1R1*	VNTR	Intron 2	410bp	595bp	0.44 (0.22-0.85)	0.008
AUGOSA							Not tested			
Pattaro *et al*. 2006	25 cases, 526 controls	European	Adult	*IL1R1*	rs447713	1135067560	A	G	Unknown	0.0013
AUGOSA							Not typed			
Pattaro *et al*. 2006	25 cases, 526 controls	European	Adult	*IL1R1*	rs3087271	115075088	A	C	Unknown	0.0227
AUGOSA							Not typed			
Wu *et al.* 2010	492 case trios	Mexican	Children	*IL1RL1*	rs13431828	102321085	C	T	0.45 (0.29-0.70)	$2.00x10^{(-4)}$
AUGOSA							G	A	0.77 (0.65-0.91)	$1.98x10^{(-3)}$
Wu *et al.* 2010	492 case trios	Mexican	Children	*IL1RL1*	rs1041973	102321900	C	A	0.58 (0.43-0.78)	$3.50x10^{(-4)}$
AUGOSA							C	A	0.93 (0.81-1.05)	0.236
Moffatt *et al.* 2010	10,365 cases, 16,110 controls	European	Adult	*IL18R1*	rs3771166	102352654	G	A	0.87 (0.83-0.91)	$3.40x10^{(-9)}$
AUGOSA							G	A	0.79 (0.71-0.88)	$1.93x10^{(-5)}$

IL1RL1: interleukin 1 receptor-like 1; *IL1R1* : interleukin 1 receptor 1 ; *IL18R1* : interleukin 18 receptor 1.

6.4 Discussion

The aim of this chapter was to assess the contribution of genetic variants previously reported to be associated with asthma irrespective of diagnosis criteria in the AUGOSA severe asthma cohort. A comprehensive literature review was undertaken in order to identify relevant gene regions for analysis.

Assessments for significant loci were conducted using an appropriate significance threshold taking into account the true independency of the genomic regions examined. This is the first study to comprehensively evaluate the role of previously associated asthma genes in a severe asthma cohort. Firstly, genes forming the largest contribution to the severe asthma phenotype from all reported results were determined. A total of six SNPs met defined significance threshold tagging three distinct loci: *IL1RL1/IL18R1* on chromosome 2p12.1 ($p=1.93x10^{(-5)}$), *TSLP* on chromosome 5q22.1 ($p=5.52x10^{(-6)}$), and *IL13* on chromosome 5q31.1 ($p=5.71x10^{(-5)}$).

Assuming that genes identified using the GWA approach are accepted as true determinants of susceptibility to severe asthma within this cohort, the next aim was to ascertain how much of the remaining variation seen can be explained by other identified genes with much more modest effects. Excluding all genes identified through GWA approaches and all tagging SNPs, the *IL1RL1* ($p=2.37x10^{(-5)}$), and *TSLP* ($p=5.52x10^{(-6)}$) loci remained significant following the second set of analyses.

Finally, analyses were undertaken to reduce the number of regions tested by only focusing on positively replicated genes supported by three or more independent papers. This aimed to evaluate the impact of publication bias in previous literature. No SNPs reached statistical significance; however, the *IL1RL1* loci showed highest significance just below the defined threshold ($p=4.35x10^{(-4)}$). This suggests that although results of candidate genes reported may be influenced by bias towards reporting positive associations, there is still substantial evidence to support identified loci by results from these analyses. Limiting analyses to widely replicated

results may be liable to miss a number of potentially important loci. The inflation test statistics resulting from these analyses were high: in all regions λ=1.17, excluding GWA identified loci λ=1.19, replicated remaining loci λ=1.57. However, the overall test statistic for the GWA study described in Chapter 5 is low, λ=1.04. This provides strong evidence to suggest contributions from multiple genes in addition to those identified using GWA approaches.

The chromosome 2p12.1 and 5q31.1 loci have already been discussed earlier in this thesis in Chapter 5, results from this Chapter emphasise the potential importance of these loci in determining genetic susceptibility to severe asthma. Taken together, the findings from the GWA studies in GABRIEL and AUGOSA as well as this candidate gene look-up, the signal for the *IL1RL1* locus appears to show the strongest evidence for association with both mild-to-moderate and severe asthma. All previously published SNPs which showed positive association on chromosome 2q12 were reviewed to evaluate consistency of findings across studies and within the AUGOSA cohort. All SNPs reported showed the same direction of effect and except for rs1041973 reported by Wu *et al* were also significantly associated with severe asthma within AUGOSA (Pattaro *et al.* 2006; Settin *et al.* 2008; Zeyrek *et al.* 2008; Gudbjartsson *et al.* 2009; Moffatt *et al.* 2010; Wu *et al.* 2010).

Low supporting evidence was suggested for the *TSLP* locus. The *TSLP* gene has only been reported by one candidate gene association paper involving 5,565 individuals from four independent asthma studies (He *et al.* 2009). The GABRIEL study also reported suggestive association with severe asthma for the same SNP, rs1837253 within *TSLP* and showed the same direction of effect for the minor allele as identified in this Chapter ($p=4.85x10^{(-6)}$, OR=0.75 in the AUGOSA analyses; $p=3x10^{(-6)}$, OR=0.56 reported in the GABRIEL Study). However, the association result in the GABRIEL Study was only present in the 290 severe asthma subjects recruited by the GABRIEL Consortium and not replicated in the 637 core AUGOSA severe asthma subjects.

The *TSLP* gene encodes an IL7-like cytokine and plays an important role in the maturation of T cell populations through activation of antigen presenting cells (Park *et al.* 2000; Ziegler and

Liu 2006). This gene was originally studied due to its function as a biologically plausible candidate in immunomodulation of the Th2 phenotype and allergic airway inflammation (Headley *et al.* 2009). High expression of *TSLP* has been shown in airway epithelial cells with increased Th2 attracting chemokines and disease severity (Ying *et al.* 2005). This hypothesis has also been supported by work carried out in murine models testing both asthma and allergic phenotypes (Yoo *et al.* 2005; Zhou *et al.* 2005). Furthermore, TSLP has been shown to act in synergy with IL33 in upregulation of immune-mediated allergic sensitisation (Allakhverdi *et al.* 2007; Guo *et al.* 2009; Matsuda *et al.* 2010).

Although not reaching genome-wide significance, another 5q22 SNP, rs2416257 ($p=1.2x10^{(-4)}$) within the WD repeat domain 36 (*WDR36*) gene in the same LD block as *TSLP* was studied in the deCODE GWA study of eosinophil numbers (Gudbjartsson *et al.* 2009). *WDR36* encodes a T-cell activation protein which is co-regulated with the T-cell growth factor, IL2. This SNP was shown to be highly associated with atopic asthma ($p=4.2x10^{(-6)}$) but not with non-atopic asthma (p=0.78). These results as well as sub-phenotype analyses carried out for atopic asthma, airway hyper-responsiveness and atopy defined by SPT in the original candidate gene paper reporting this gene suggest that it may predominantly play a role in allergic disease and thus be a determinant of atopy rather than asthma. There has been evidence from functional studies to support the role of gene as being specific to both traits (Mao *et al.* 2004).

Evidence to suggest that over-expression of *TSLP* may be sex-specific and vary across individual cohorts has also been published (Taneda *et al.* 2001; Kowalewska *et al.* 2007). Therefore, the fact that results were not replicated between the two severe asthma cohorts studied in the GABRIEL study should not be evidence to dismiss a potential role for this locus. Severe asthma has been shown to be more prevalent in adult females compared to males (Wenzel 2005). Furthermore, criteria for selection of subjects by GABRIEL were based on the ATS and ERS definitions for severe asthma and in addition, partially recruited from a childhood clinic. Therefore, these individuals may have more of an atopic phenotype compared to those recruited by the core AUGOSA cohort and may be a reason for the inconsistent result. The reason to explain why this locus was not identified by the main GABRIEL GWA study for

mild-to-moderate asthma but has shown significant association with severe disease may be the increased inflammatory environment in severe asthmatic airways.

A recent meta-analysis of North-American asthma GWA studies in European American, African-American or African-Caribbean, and Latino ancestry populations (n=5,416 with replication in an additional 12,649 individuals) confirmed association of 17p21, near *IL1RL1*, *TSLP*, and *IL33* across the three ethnic groups (Torgerson *et al.* 2011). In addition, a new locus at *PYHIN1* was identified specific to individuals of African descent ($p=3.9x10^{(-9)}$). This study suggests contributions to asthma susceptibility from loci which are robust to differences in ancestry detectable using sufficiently large sample sizes as well as ancestry-specific associations.

This study population consisted of individuals with severe disease of European ancestry only. Many of the genes identified from previous studies may have effects specific to different forms of asthma and in different ethnic populations. Furthermore, underlying differences in allele frequencies may mean that although a variant is important in disease susceptibility across a number of populations, a study carried out in one, in which the variant is rare, will require a much greater sample size in order to generate enough statistical power to observe an effect. Under the assumption that many of the genes responsible for mild-to-moderate asthma are also involved in determining severe disease suggested by results from Chapter 5, a lack of replication for many candidate regions is potentially a reflection that much of the previously reported association came from studies with small sample sizes and inadequate gene coverage across the different genotyping arrays used. The Q-Q plots suggest that genes identified by GWA approaches have been able to explain some of the more extreme deviations from the line i.e. a small number of loci with large effects such as 2q12, 5q31, and 17q12-21. Positive results from previously reported candidate genes studies may actually be tagging the GWA signals seen at a later date.

However despite this, there is still evidence to suggest the role of multiple additional loci with smaller effects seen by the deviation of a large number of SNPs lower down even after

removing tests for GWA reported hits. For example, the most suggestive gene identified just below statistical significance encodes for the myosin light chain kinase (*MYLK*) on 3q21.1. Several linkage studies have previously reported this region (Dizier *et al.* 2000; Lee *et al.* 2000; Blumenthal *et al.* 2004; Kurz *et al.* 2005). Functional studies have shown increased levels of both smooth muscle and non-muscle myosin light chain kinase in asthma, which may result from functional alterations in airway smooth muscle (Benayoun *et al.* 2003). This gene was listed as a candidate for evaluation because a single study had reported association with severe asthma and a variant within *MYLK* in the African-American population (Flores *et al.* 2007). The SNP identified had not been genotyped in European individuals in the same study. The association signal seen in AUGOSA ($p=5.21x10^{(-3)}$) may provide some supporting evidence for this gene and help to explain some of the more severe symptoms observed in asthma patients of African-American ancestry (Barnes *et al.* 2007).

A potential weakness in the approach used is that publications reporting negative associations only were not included in these analyses. Using only studies with positive results weights the literature gathered towards unconfirmed associations. Comparing the relative numbers of positive and negative studies would help to evaluate the empirical evidence for identified loci to better put p value thresholds into context. Assessing studies reporting negative association could also help to determine the conditions in which a locus may confer susceptibility as contributing genetic loci are likely to vary between different sub types of asthma, levels of severity, and populations differing in age and ancestry.

Currently, few other published studies have used this approach to apply GWA to evaluate previously reported candidate genes to such a detailed extent (Rogers *et al.* 2009; Moffatt *et al.* 2010). The most extensive so far is a study in the Mexican population (492 families), which tested association of 237 candidate genes with childhood-onset asthma showing the most significant results for tumour growth factor, beta 1 (*TGFB1*) ($p=3x10^{(-4)}$), *IL1RL1* ($p=2x10^{(-4)}$), *IL18R1* ($p=9\times10^{(-3)}$), and *DPP10* ($p=8x10^{(-4)}$) (Wu *et al.* 2010). In other diseases, a study of Alzheimer's Disease (491 cases, 479 controls) confirmed association with GWA signals as

well as identified evidence based on nominal significance for contributions from candidate genes reported by case-control studies (Feulner *et al*. 2009).

There is limited power within the AUGOSA study to detect variants with small effects especially those with a low minor allele frequency (MAF). In determining statistical significance, the corrections taken for multiple testing are likely to have been over-conservative. Under the hypothesis that there is a single gene or locus contributing to susceptibility to disease, SNPs would have met statistical significance defined using the Bonferroni correction. If degrees of linkage disequilibrium between SNPs are taken into consideration, then sufficient SNPs to explain a larger proportion of heritability would have met the MeffLi corrected significance threshold. However, current statistical corrections for multiple testing need to be re-evaluated to adequately take into consideration contributions from multiple genes. For example, if there were 30 loci expected to be associated with disease, a matched 30-fold correction would need to be taken.

In summary, these results provide evidence to support contributions from multiple loci with small effects in the development of severe asthma in combination to a select number of loci meeting or approaching genome-wide significance. Further functional work is needed to help to distinguish pathways involved in the onset of this complex disease. In addition, fine mapping and re-sequencing of these loci will provide value to try and identify rare and/or intermediate variants which may also be contributing to the disease phenotype.

Chapter 7

Molecular Characterisation of the Gene Promoter Regions of *IL1RL1* and *IL33*

7.1 Introduction

Two recent large scale genome-wide association (GWA) studies have been carried out identifying common polymorphisms in the *IL1RL1* and *IL33* genes determining asthma (Gudbjartsson *et al.* 2009; Moffatt *et al.* 2010). Firstly, the deCODE Icelandic GWA study identified chromosomes 2q12 and 9p24 for genome-wide significant association with blood eosinophil counts and also asthma. The reported SNPs, rs1420101 is located within *IL1RL1* (2q12) and rs393928 is 32kb upstream of *IL33* (9p24) (Gudbjartsson *et al.* 2009). Secondly, the GABRIEL study also reported asthma association of SNPs within these loci: rs3771166 within the *IL18R1* gene in high linkage disequilibrium (LD) with other SNPs showing significant association within *IL1RL1* as well as rs1342326 flanking *IL33* approximately 52kb upstream (Moffatt *et al.* 2010).

A number of candidate gene studies have identified polymorphisms in *IL1RL1* and *IL33* to be associated with atopic diseases including asthma: a three SNP haplotype within *IL1RL1* (OR 1.77) (Reijmerink *et al.* 2008) and atopic dermatitis: distal promoter -26,999 G/A (OR 1.86) (Shimizu *et al.* 2005). The distal promoter SNP -26,999 G/A was also shown to be associated with asthma severity with an increased severity score 10.1 compared to 9.3 based on a score ranging from 5 to 15 (5 to 7 mild, 8 to 11 moderate, 12 to 15 severe) (Castano *et al.* 2009).

In addition to genetic links made between the *IL1RL1* and *IL33* genes with asthma and related traits, there has been accumulating evidence to suggest important roles for these genes in inflammatory disease pathogenesis by biological studies. The IL1RL1 (also known as ST2) receptor exists both as a transmembrane receptor and in a soluble form measurable in serum

(Kakkar and Lee 2008). IL1RL1 is targeted by its natural ligand IL33, a cytokine within the IL1 superfamily (Schmitz *et al.* 2005). The contribution of IL1RL1-IL33 signalling in eosinophil-mediated inflammation has been shown through both direct effects on eosinophil maturation, survival, and activation as well as indirect recruitment and regulation of Th2 cytokines including IL4, IL5, and IL13 in inflamed tissues (Carriere *et al.* 2007; Cherry *et al.* 2008). Higher serum IL33 levels were detected in patients with Japanese cedar pollinosis, a disease of allergic rhinitis and allergic conjunctivitis resulting from exposure to Japanese cedar pollen (Sakashita *et al.* 2008). Furthmore, there is evidence to suggest increased levels of expression of both the soluble form of the IL1RL1 receptor and IL33 in more severe asthma. Serum IL1RL1 levels during asthma exacerbations were shown to be statistically correlated with percent predicted peak expiratory flow (r^2=0.634, p=0.004) (Oshikawa *et al.* 2001). Using bronchial biopsies from asthma patients, *IL33* mRNA expression was found to be significantly higher compared to controls, particularly in subjects with severe disease (p=0.024 in moderate asthma, p=0.002 in severe asthma) (Préfontaine *et al.* 2009).

Effects of the distal promoter region SNP, -26,999 G/A in increasing transcriptional activity of *IL1RL1* has been shown to induce preferential activation of the Th2 response using reporter gene assays to measure different levels of the IL1RL1 protein and total IgE concentrations (Shimizu *et al.* 2005). However, given the low effect sizes detected (OR <2) and common frequency in the population of associated polymophisms, for example, for rs1921622 (MAF 0.42), rs1861246 (MAF 0.24), and rs1921622 (MAF 0.42) reported within *IL1RL1* (Reijmerink *et al.* 2008), the functionalities of reported SNPs are still unclear. Using findings from these studies, it seems likely that SNPs within both *IL1RL1* and *IL33* may influence transcription of these genes and subsequent levels of protein expression. A recent study in a Dutch childhood cohort found multiple SNPs in *IL1RL1* associated with increased serum levels of soluble IL1RL1 receptor as well as blood eosinophil counts and asthma (Savenije *et al.* 2011).

A previous study investigating alternative splicing of the *IL1RL1* gene demonstrated two variants with distinct non-coding primary exons and a common exon 2 containing the transcriptional start site (TSS). This was carried out in a human leukaemic cell line (UT7) and

a human fibroblastic cell line (TM12) in which promoter usage varied (Iwahana *et al.* 1999). Expression of *IL33* mRNA in epithelial cells has been shown to be positively regulated by inflammatory stimuli such as TNF, which may be provoked through interaction with environmental allergens (Kato and Schleimer 2007; Smith 2010). Additional information on transcriptional regulatory control in airway cells will also help to understand factors which may alter gene expression in asthma. Therefore, molecular characterisation of these genes will help to infer the potential function of SNPs both identified in the previous literature and by work described in the earlier Chapters of this thesis.

7.2 Aim

The aim of this Chapter was to investigate sequence upstream of the *IL1RL1* and *IL33* genes, by characterising the 5' UTR, establish transcription start sites, and identify putative promoter locations.

7.3 Methods

7.3.1 Real-Time Quantitative PCR

Taqman® real-time quantitative PCR was used to quantify levels of *IL1RL1* and *IL33* relative mRNA expression in a representative cell and tissue panel consisting of lung tissue, human airway smooth muscle (HASM), human bronchial epithelial cells (HBEC), peripheral blood mononuclear cells (PBMC), human bronchial epithelial cell line (BEAS-2B), human mast cell line (HMC-1) and human acute monocytic leukaemia cell line (THP-1) (see Chapter 2: Section 2.5).

Taqman® is based on the fluorescence emitted by a target gene specific fluorogenic probe. The probe is designed to be flanked by two PCR primers and is labelled with a 5' reporter and a 3' quencher dye. As the two dyes are in close proximity with each other, the quencher dye is able to limit the fluorescence emitted by the reporter dye. During the PCR extension phase, the 5' nuclease activity of the *Taq* DNA polymerase cleaves the probe resulting in separation of the 5' and 3' dyes, therefore emitting fluorescence. The fluorescence emission detected using a laser light distributed on thin-walled reaction tubes with optical fibres are then directed to a spectrograph with a charge-coupled device camera.

An assay previously designed and optimised to test expression of *IL1RL1* and *IL33* by Jane Fox, Division of Therapeutics and Molecular Medicine was used. Primers (Invitrogen) and probes dual labelled with 5' FAM and 3' TAMRA dyes (Applied Biosystems) had been designed by Beacon designer 7 (Stratagene) (Appendix 2). Universal PCR Master Mix (Applied Biosystems. Cat. No. 1275) and Pre-Developed Assay Reagents were used in assays.

Each sample was run in triplicate using 900nM forward and reverse primers and 250nM probe using a MX3005 qPCR system (Stratagene). RT negative and template negative controls were analysed in parallel. A standard protocol was used to detect amplification (50°C for 2 minutes; 95°C for 10 minutes; 40 cycles: 95°C for 15 seconds, 60°C for 60 seconds; 4°C pause). Quantification of the target gene expression was monitored using MxPro QPCR software and

calculated using the Comparative method with measures of the number of threshold cycles (Ct) at which amplification first becomes exponential: ΔCt=Ct of target gene – Ct of housekeeping gene, obtaining a final value via $2^{-\Delta Ct}$. Hypoxanthine ribosyltransferase (*HPRT*) was used as the endogenous housekeeping gene for data normalisation as this had been shown by Jane Fox to have the most uniform expression between cDNA samples compared to other tested housekeeping genes *18S*, and *GAPDH*. Two repeats were performed per experiment and the results averaged.

7.3.2 Rapid Amplification of 5' cDNA Ends (5'RACE)

Heterogeneous transcriptional start sites and the promoter regions of the *IL1RL1* and *IL33* genes were characterised using 5' Rapid Amplification of cDNA ends (5' RACE) [Figure 7.1]. 5'RACE was performed using the GeneRacer™ Kit (Invitrogen. Cat. No. L1502-01). Resulting PCR products were cloned into pCR® 4-TOPO® vectors using the TOPO TA Cloning® Kit for Sequencing (Invitrogen. Cat. No. K4575-01). Plasmid DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen. Cat. No. 27104).

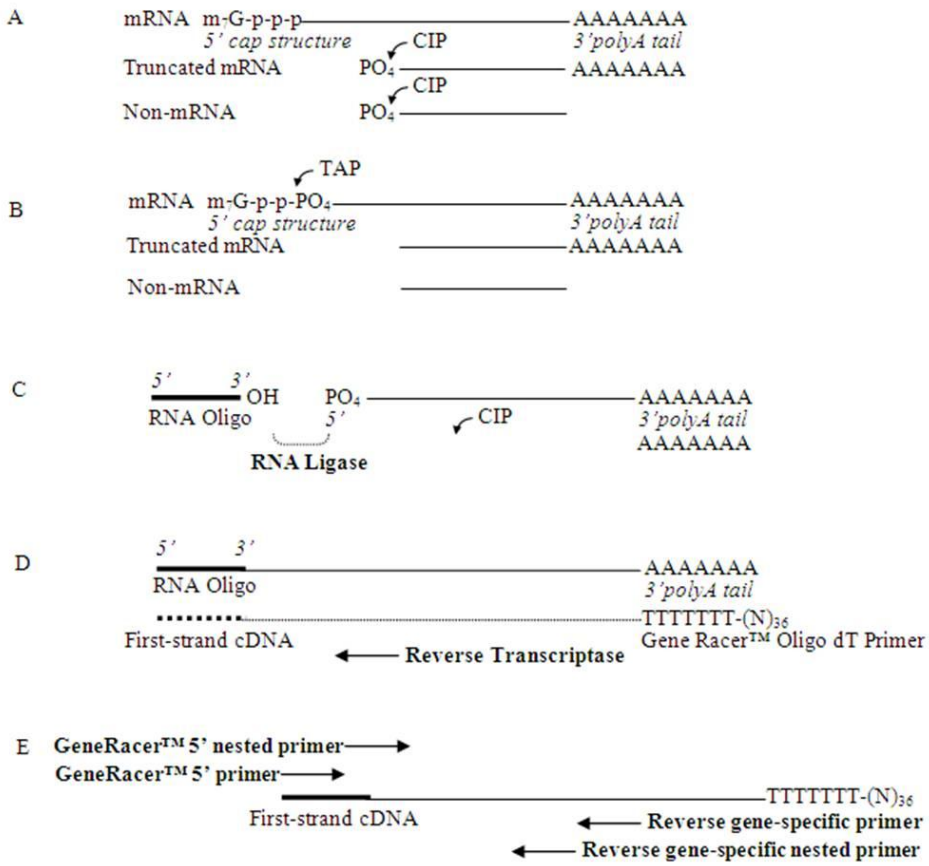


Figure 7.1- Principles of 5' RACE. A) RNA dephosphorylation. B) Removing the mRNA cap structure. C) Ligating the RNA Oligo to decapped mRNA. D) Reverse transcribing mRNA. E) Amplification of cDNA ends.

7.3.2.1 Synthesis of RACE-Ready cDNA

Total RNA was treated with calf intestinal phosphatase (CIP) to remove 5' phosphates. CIP has no effect on full-length, capped mRNA and therefore eliminated truncated mRNA and non-mRNA from subsequent ligation with the GeneRacer™ RNA Oligo [Figure 7.1 A]. Desphosphorylated RNA was treated with tobacco acid pyrophosphatase (TAP) to remove the 5' cap structure from intact, full-length mRNA leaving a 5' phosphate required for ligation to the GeneRacer™ RNA Oligo using T4 RNA ligase [Figure 7.1 B, C]. This provided a known priming site for the GeneRacer™ PCR primers after transcription of mRNA into cDNA. The ligated mRNA was reverse transcribed using SuperScript™ III RT and the GeneRacer™ Oligo dT Primer to create RACE-ready first-strand cDNA with known priming sites at the 5' end [Figure 7.1 D].

RNA Dephosphorylation

In addition to test RNA, a HeLa RNA positive control was provided in the kit. Each reaction was set up on ice in a 1.5ml sterile microcentrifuge tube using:

RNA	2.5µg
10x CIP Buffer	1µl
RNaseOut™ (40U/µl)	1µl
DEPC water	to 10µl

Reagents were mixed gently by pipetting and vortexed briefly before being centrifuged and incubated at 50°C for 1 hour. Each reaction was then centrifuged briefly and placed on ice.

RNA Precipitation

90µl of DEPC water and 100µl phenol:chloroform was added to each reaction and vortexed vigorously for 30 seconds before centrifugation at maximum speed in a microcentrifuge for 5 minutes at room temperature. The top, aqueous phase was transferred to a new microcentrifuge tube. Subsequently, 2µl 10mg/ml mussel glycogen (to aid precipitation), 10µl 3M sodium acetate, pH 5.2 and 220µl 95% ethanol were added, vortexed, and frozen on dry ice for 10 minutes. Centrifugation at maximum speed in a microcentrifuge was carried out for 20 minutes at 4°C to pellet the RNA. A pipette was used to carefully remove the supernatant. Then, 500µl 70% ethanol was added, the tube inverted several times and vortexed briefly before repeated centrifugation for 20 minutes and removal of the resulting supernatant. The pellet was air-dried for 1 to 2 minutes at room temperature before resuspension in 7µl DEPC water.

Removing the mRNA Cap Structure

Each reaction was set up on ice in a 1.5ml sterile microcentrifuge tube using:

Dephosphorylated RNA	7µl
10x TAP Buffer	1µl
RNaseOut™ (40U/µl)	1µl
TAP (0.5U/µl)	1µl

Reagents were mixed gently by pipetting and vortexed briefly before being centrifuged and incubated at 37°C for 1 hour. Each reaction was then centrifuged briefly and placed on ice. The samples were then precipitated and resuspended in 7µl DEPC water as before.

Ligating the RNA Oligo to Decapped mRNA

7µl of dephosphorylated, decapped RNA was added to a tube containing lyophilised GeneRacer™ RNA Oligo (0.25µg), mixed gently to resuspend the Oligo and centrifuged before incubation at 65°C for 5 minutes to relax the RNA secondary structure. The reaction was placed on ice, centrifuged briefly, and the following reagents added and mixed gently before brief centrifugation and incubation at 37°C for 1 hour:

10x Ligase Buffer	7µl
10mM ATP	1µl
RNaseOut™ (40U/µl)	1µl
T4 RNA ligase (5U/µl)	1µl

The tubes were returned to ice and precipitated as before with resuspension in 10µl of DEPC water.

Reverse Transcribing mRNA

The following reagents were added to 10µl of ligated RNA:

5µM Oligo dT primers	1µl
dNTP mix (10µM each)	1µl
DEPC water	1µl

The reaction was incubated at 65°C for 5 minutes to remove any RNA secondary structure, placed on ice for at least 1 minute and centrifuged briefly before adding:

5x First Strand Buffer	4µl
RNaseOut™ (40U/µl)	1µl
Superscript™ III RT (200U/µl)	1µl

This was mixed well, centrifuged briefly, and incubated at 50°C for 50 minutes. Subsequent incubation at 70°C for 15 minutes was carried out to inactivate the reaction prior to chilling on ice for 2 minutes and brief centrifugation. 1µl of RNase H (2U) was added and incubated at 37°C for 20 minutes to digest away the RNA template before repeated centrifugation. Synthesised RACE-ready cDNA was stored at -20°C for use as templates in RACE PCR reactions.

7.3.2.2 PCR Primer Design for RACE

Gene-specific primers were designed using NCBI sequences taking the following parameters into consideration: primer length 23bp to 25bp, 50% to 70% GC, annealing temperature >70°C given by 4(G+C)+2(A+T), <2 G/C residues in the last 5 bases at the 3' end, no self-complementary sequences within the primer or complementary to kit primers, especially at the 3' end. Gene-specific nested primers were designed to have similar annealing temperatures to the GeneRacer™ nested primer and distant enough from the original gene-specific primer in order to distinguish the products of the original and nested PCRs by size. Known SNPs in coding sequences were identified and avoided during primer design. Primers for 5' RACE were designed in exon 3 for *IL1RL1* common to both published transcripts. Designed primers for *IL33* were in exon 4 common to published transcripts 1 and 2 only. This was due to transcript 3 of *IL33* being documented online on 10th December 2010, after RACE experiments had already been carried out. *IL1RL1* and *IL33* primers used for 5' RACE are listed in Appendix 2.

7.3.2.3 Amplification of cDNA Ends

The first-strand cDNA was amplified using a reverse gene-specific primer and the GeneRacer™ 5' Primer, homologous to the GeneRacer™ RNA Oligo. Only completely reverse transcribed mRNA with the GeneRacer™ RNA Oligo ligated to the 5' end was amplified using PCR and subsequent nested PCR. Platinum® *Taq* DNA Polymerase High Fidelity results in single A-overhangs to the 3' ends of PCR products to allow subsequent ligation with linearised vectors containing single overhanging 3' T residues [Figure 7.1 E].

Two PCR reactions were performed to increase product specificity, the second nested using 1µl of the first reaction as the template and with primers nested inside those used for the first amplification. Negative controls using the gene-specific primers were set up to check for background binding and amplification in test samples as well as a template negative to check for purity of reagents and procedural contamination. Positive control primers for β-actin were tested on the HeLa control sample.

20µl reactions were set up for each reaction (cycling parameters):

First PCR (94^0C for 2mins; 5 cycles at 94^0C for 1min, 72^0C for 2mins; 5 cycles at 94^0C for 1min, 70^0C for 2mins; 20 cycles at 94^0C for 1min, 68^0C for 2.5mins; 68^0C for 10mins; 4^0C hold)

GeneRacer™ 5' Primer	3µl
Reverse gene-specific primer	1µl
RACE-ready cDNA	1µl
10x High Fidelity PCR Buffer	5µl
dNTPs (10mM each)	1µl
Platinum Taq DNA polymerase High Fidelity (5U/µl)	0.5µl
$MgSO_4$ 50mM	2µl
Sterile water	36.5µl

Nested PCR (94^0C for 2mins; 25 cycles at 94^0C for 1min, 65^0C for 30secs, 68^0C for 2mins; 68^0C for 10mins; 4^0C hold).

7.3.2.4 Ligation and Transformation of 5' RACE PCR Products

5' RACE PCR products were purified and cloned into pCR® 4-TOPO® vectors containing flanking sequences recognised by sequencing primers M13R [Figure 7.2]. Topoisomerase I enzyme from *Vaccinia* virus covalently bound to the vector ligates T-overhangs present on the cut vector to A-overhangs on the PCR product. The constructs produced by the cloning reaction were transformed into One Shot® Chemically Competent TOP10 *E. coli*. Ligation of a PCR product disrupts expression of the lethal *E. coli* gene, *ccdB* permitting growth of only positive recombinants upon transformation.

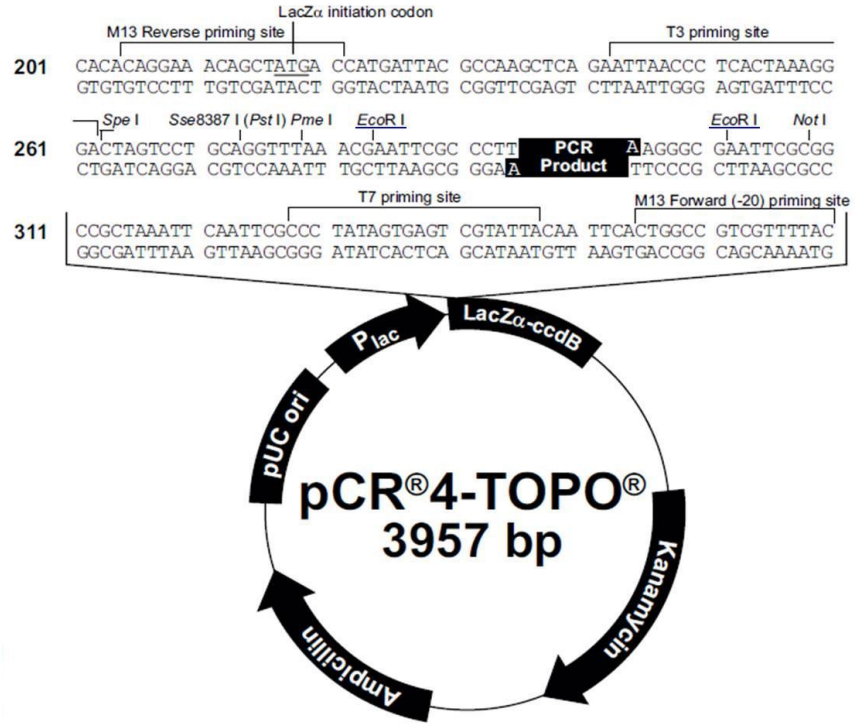


Figure 7.2 – pCR®4-TOPO vector map taken from the TOPO TA Cloning® Kit for Sequencing manual showing the sequence surrounding multiple cloning sites. M13 reverse primers were used for sequencing the insert and *Eco*RI sites (underlined) used to excise the PCR product inserts from the vector.

4µl of the PCR product was added to 1µl Salt solution and 1µl TOPO® vector, mixed gently before incubation for 5 minutes at room temperature, and placed on ice. 2µl of the reaction was added into a vial of thawed TOP10 cells and gently mixed with the pipette tip. Cells were left on ice for 5 minutes prior to heat-shock in a hot block for 30 seconds at 42°C and then immediately transferred to ice. SOC medium (250µl) was added to ensure even spreading and the tube placed horizontally at 37^0C to shake at 200rpm for 1 hour. The broth and cells were then plated onto pre-warmed 100µg/ml ampicillin selective agar plates at volumes of 100µl and 200µl to obtain a range of colony densities and incubated at 37^0C overnight. Single colonies obtained were picked into 4ml LB broths containing ampicillin at 100µg/ml and again incubated at 37^0C overnight with shaking at 200rpm.

7.3.2.4 Purification of Plasmid DNA from LB Broth

Plasmid DNA was isolated using the QIAprep Spin Miniprep kit (Qiagen. Cat. No. 27104). This procedure is based on alkaline lysis of bacterial cells followed by adsorption of DNA onto silica in the presence of high salt before washing and eluting the resulting plasmid DNA. A restriction enzyme recognises a specific sequence of DNA (approximately 4bp to 6bp) and cleaves the DNA at this point. Digestion using *Eco*RI restriction enzyme was performed on the DNA prepared to check the presence of inserts.

DNA Plasmid Preparation

1.5ml of prepared broths was transferred to 1.5ml eppendorf tubes and centrifuged at 13,000rpm for 5 minutes to pellet the cells. The supernatant was discarded and tubes inverted onto paper towel to drain. Cells were vortexed and resuspended in 250µl Buffer P1 (with RNase A 100mg/ml added to give a final concentration of 100µg/ml). 250µl Buffer P2 was added and tubes gently inverted 4 to 6 times to lyse cells. 350µl Buffer N3 was added within 5 minutes to avoid DNA denaturing which would result in a restriction enzyme-resistant open circular form of the plasmid. The tubes were gently inverted 4 to 6 times to neutralise the mixture and precipitate proteins, chromosomal DNA, cellular debris, and SDS before centrifugation at 13,000rpm for 10 minutes. The resulting supernatant was transferred to a QIAprep column placed in a collection tube and spun for 30 seconds. The flow-through was discarded after centrifugation for 45 seconds with consecutive washes of 0.5ml Buffer PB and 0.75ml Buffer PE. The flow-through was again discarded and columns centrifuged for 2 minutes to dry and transferred to a clean 1.5ml eppendorf tube. 50µl of sterile water was carefully added directly onto the column membrane, left to rest for 1 minute before centrifugation for 1 minute to elute the plasmid DNA before storing at -20°C.

Restriction Endonuclease Digestion of DNA to Identify PCR Inserts

The following reaction per plasmid sample was set up, gently vortexed, and centrifuged before incubation at 37°C for 2 hours:

*Eco*RI enzyme	0.5µl
10x *Eco*RI Buffer	1µl

Plasmid product from miniprep 3.5µl

3µl was run with 6x loading dye for agarose gel electrophoresis using 1.5% agarose gel.

7.3.2.5 Sequencing and Analysis of Inserts

DNA obtained from 5' RACE clone minipreps was sequenced using vector primer M13R (see Chapter 2: Section 2.2.4). Sequence data was analysed by comparison with reference contigs to identify individual exons, splice borders, and primers. Only sequences bearing the GeneRacer 5' primers were accepted as true transcripts.

7.3.3 Sequence Data Analysis

Transcription start sites were identified, putative promoter regions were analysed for transcription factor binding sites, and protein prediction carried out for novel transcripts detected (see Chapter 2: Section 2.3.1). Transcription factors known to be important in asthma including Nuclear factor-κB (NF-κB), activator protein-1 (AP-1), CAAT/Enhancer binding proteins (C/EBP), signal transduction-activated transcription factors (STAT), cyclic AMP response element binding protein (CREB), CREB binding protein (CBP), Glucocorticoid receptors (GRs), and nuclear factor of activated T-cells (NF-AT) were mapped (Barnes and Adcock 1998). Four online databases were used to search for promoter factor binding sites, those reported by two or more databases were identified (see Chapter 2: Section 2.3.1). Locations of SNPs reported for genome-wide association were determined in relation to characterised gene structures. All transcript information was based on Genome Build Ch37.1 and data from earlier builds converted accordingly using the Convert utility on the UCSC Genome Browser annotation tracks page.

7.4 Results

7.4.1 Defining Genomic Arrangements

Previously reported transcripts for *IL1RL1* and *IL33* documented online in NCBI databases were used to build maps of expected gene structures. Two published transcripts have been reported for *IL1RL1* with six exons in common. Variant 1 (NM_016232.4) encodes the longest isoform 1 composed of 11 exons. Variant 2 (NM_003856.2) differs in the 5' and 3' UTRs compared to variant 1 resulting in a distinct isoform 2 with a shorter C-terminus [Figure 7.3]. A third variant (NM_173459.1) has previously been reported for *IL1RL1* containing a new exon compared to variant 1, which leads to a protein with a different C-terminus localising on the plasma membrane of cells (Tominaga *et al.* 1999; Tago *et al.* 2001). This transcript was however permanently supressed on NCBI because it was weakly supported by other studies. Three known transcripts have been reported for *IL33* with the same N- and C- termini but differing in length. Variant 1 (NM_033439.3) encodes the longest isoform composed of 8 exons. Variant 2 (NM_001199640.1) encodes a shorter isoform which does not contain exon 5 of variant 1. Variant 3 further lacks exons 3 and 4 resulting in the shortest isoform [Figure 7.4].

Asthma associated SNPs responsible for the signals reported in *IL1RL1* and *IL33* genes by the deCODE and GABRIEL studies were mapped in relation to published structures (Gudbjartsson *et al.* 2009; Moffatt *et al.* 2010). All GWA tested SNPs genotyped in the AUGOSA severe asthma study for each gene were plotted and r^2 identified with reported SNPs: rs1420101 (*IL1RL1*) and rs3939286 (*IL33*) by the deCODE study, rs1342326 (*IL33*) by the GABRIEL study [Figures 7.3 and 7.4].

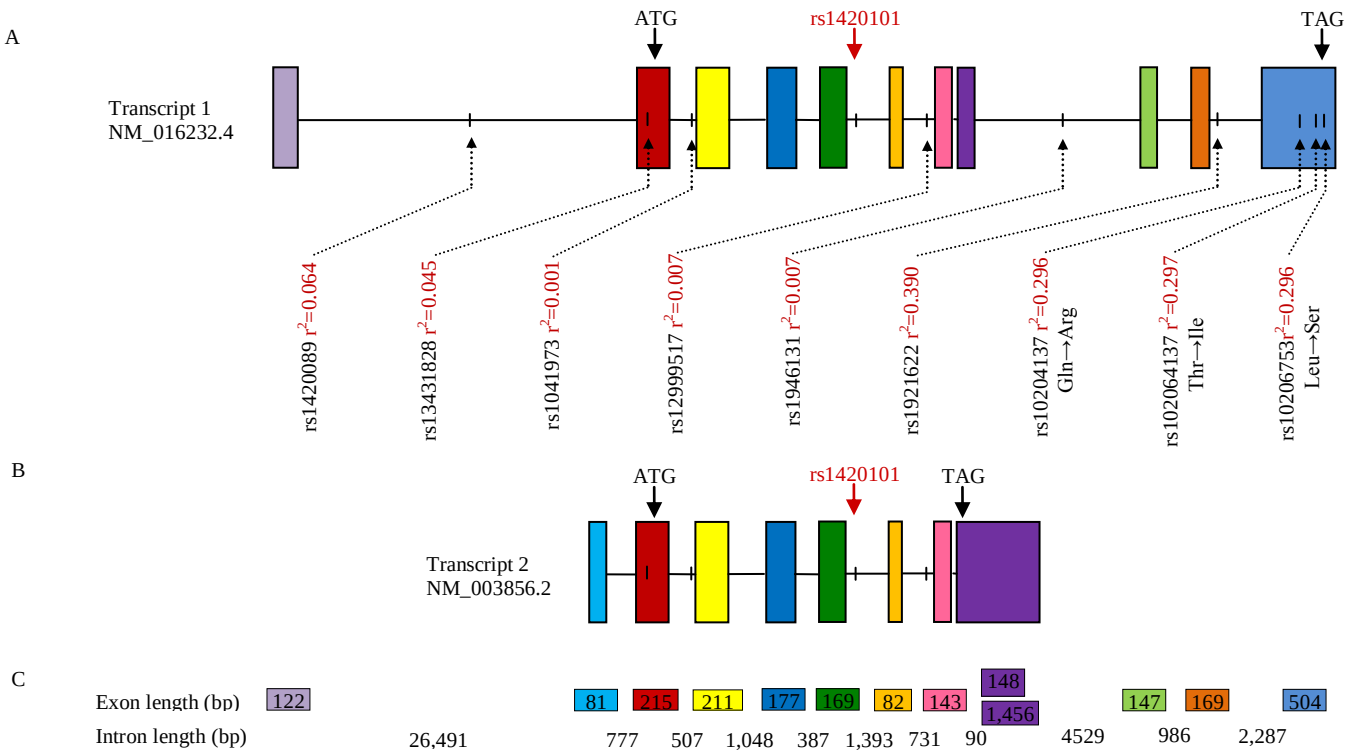


Figure 7.3 - Schematic representation of the transcript arrangements obtained from published sequences for *IL1RL1*. A) Published Transcript 1 (NM_016232.4). B) Published Transcript 2 (NM_003856.2). C) Exon and intron sizes (bp). Established exons are colour coded, translational start (ATG) and stop codon (TAG) sites are shown. Asthma associated SNP, rs1420101 reported by deCODE and r^2 values with genotyped SNPs (in black) in the AUGOSA study are marked. Amino acid changes resulting from SNPs in coding regions are detailed. Primers for 5' RACE were designed in common exon 3 (yellow).

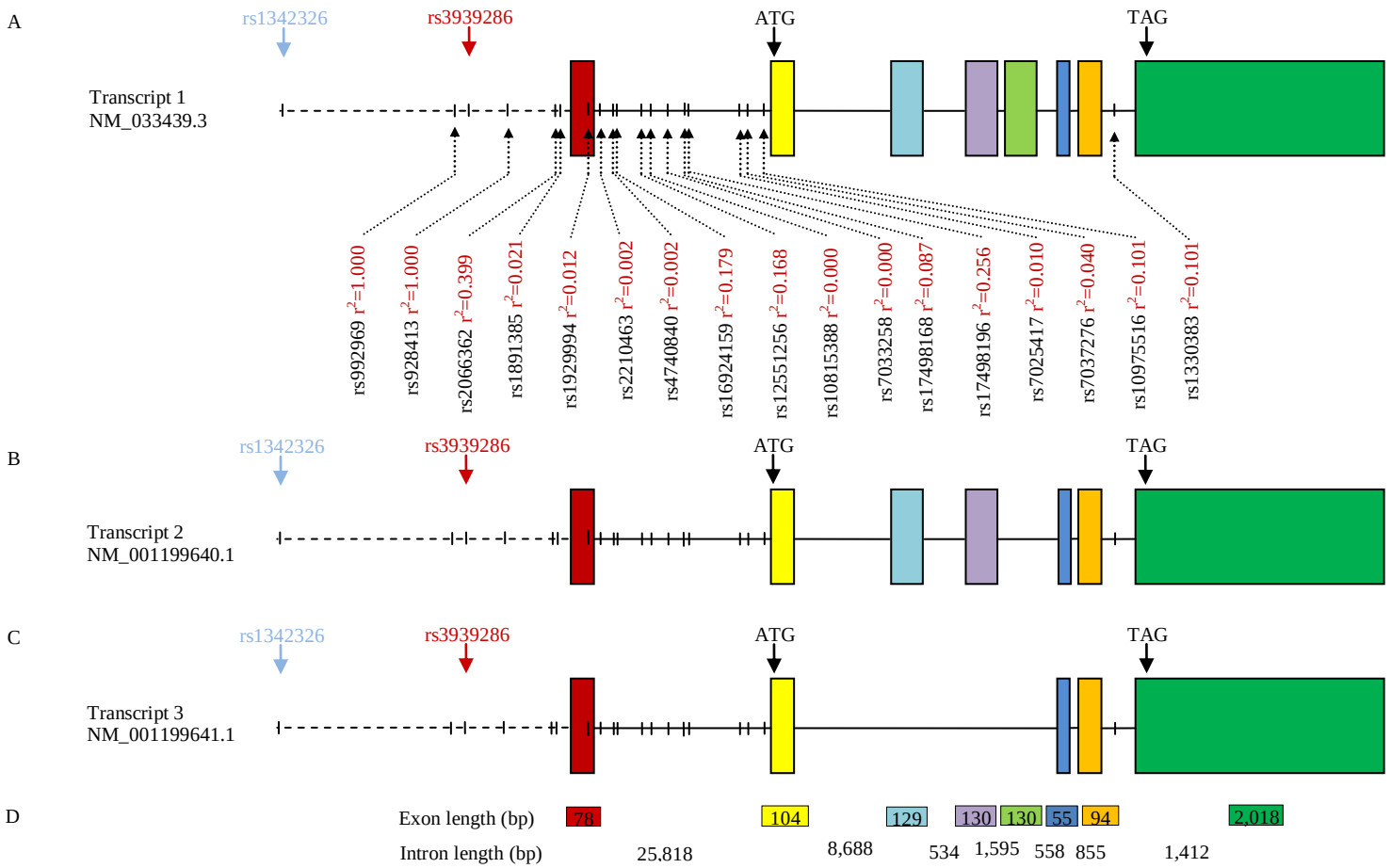


Figure 7.4 - Schematic representation of the transcript arrangements obtained from published sequences for *IL33*. A) Published Transcript 1 (NM_033439.3). B) Published Transcript 2 (NM_001199640.1). C) Published Transcript 3 (NM_001199641.1). D) Exon and intron sizes (bp). Established exons are colour coded, translational start (ATG) and stop codon (TAG) sites are shown. No LD was found between genotyped SNPs (in black) in the AUGOSA study and rs3939286 reported by deCODE, r^2 values with rs1342326 reported by GABRIEL are shown. Primers for 5' RACE were designed in exon 4 common to transcripts 1 and 2 (purple).

7.4.2 Quantitative Gene Expression

Relative levels of mRNA expression of *IL1RL1* and *IL33* were averaged from two repeat experiments using Taqman® quantitative PCR [Figure 7.5]. All RT and template negative controls were free of contamination. Both mRNA transcripts were expressed in lung tissue; levels of *IL1RL1* mRNA were most abundant in HMC-1 and of *IL33* in HASM cell templates. Low quantities of *IL1RL1* and *IL33* were detected in HBEC templates (*IL1RL1* in one donor, *IL33* in two donors). Lung tissue (donor 1) and HASM cell (donor 3) samples were selected as templates for carrying out 5' RACE for *IL1RL1*. Lung tissue (donor 1) and HMC-1 samples were chosen for *IL33* [Figure 7.3].

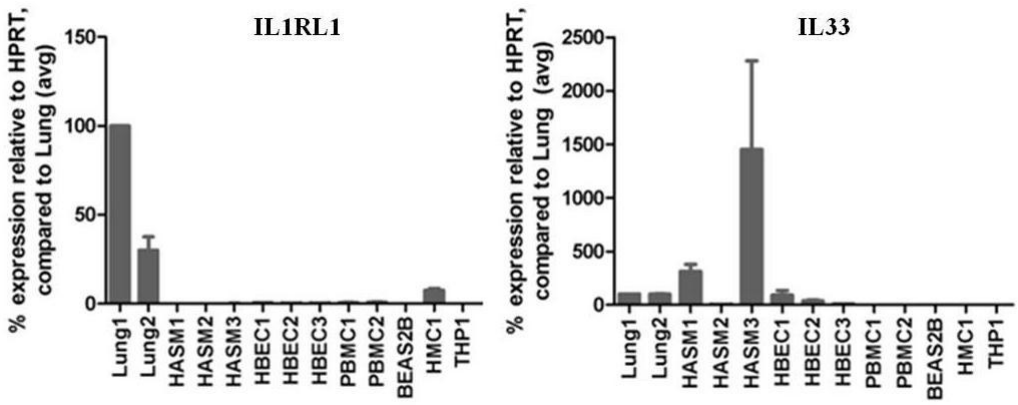


Figure 7.5 – Relative mRNA expression of *IL1RL1* and *IL33* using Taqman® quantitative PCR expression. Mean fold difference (%) and standard error of the mean (SEM) of two separate experiments carried out in triplicate, data was normalised to Hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) housekeeping gene, compared to lung.

7.4.3 *IL1RL1* and *IL33* Transcripts Detected Using 5' RACE

7.4.3.1 Primary and Nested PCR Reactions

5' RACE was performed on cDNA templates synthesised using oligo DT primer reactions on total RNA (*IL1RL1* in lung tissue and HMC-1, *IL33* in lung tissue and HASM). Electrophoresis gel pictures showing the products of the first and second nested PCR reactions are shown in Figures 7.6 and 7.7 for *IL1RL1* and *IL33* respectively.

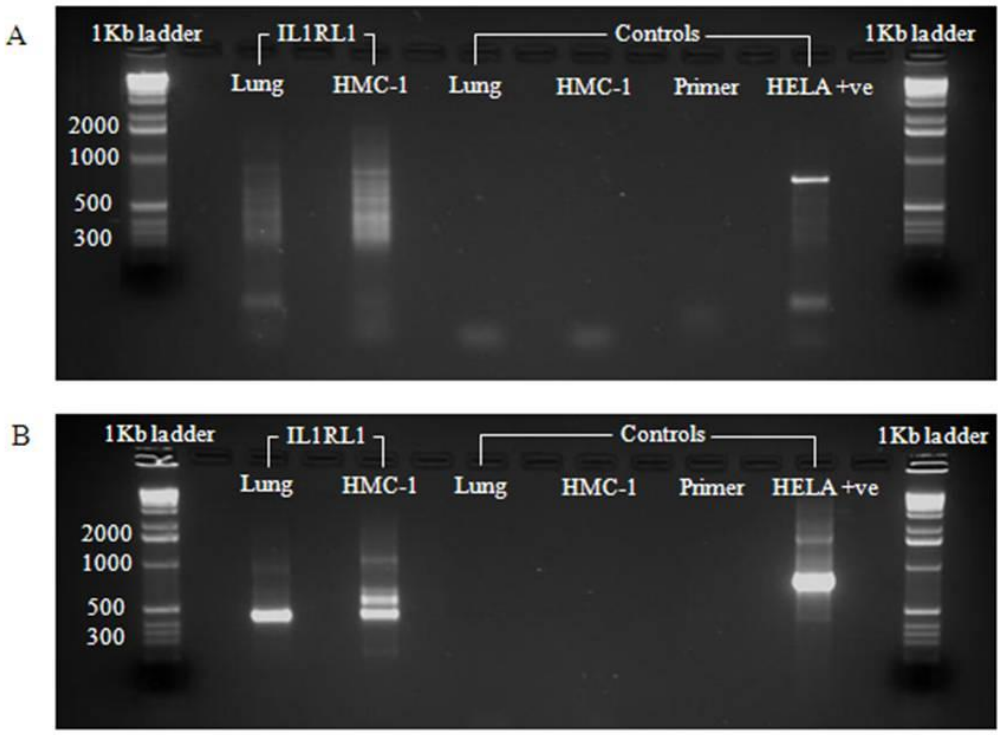


Figure 7.6 - 5' RACE PCR reactions for *IL1RL1* in lung and HMC-1 visualised using 2% agarose gel electrophoresis. A) PCR reactions performed on 1st strand RACE cDNA from lung and HMC-1 total RNA. B) Nested PCR reactions performed on products shown in gel picture A. Products seen indicate a single main transcript in lung, two main with multiple minor transcripts in human mast cell line (HMC-1). Primer negative (in lung and HMC-1) as well as template negative controls were free from contamination. An immortalised HeLa cell line (HELA) was used as the positive control.

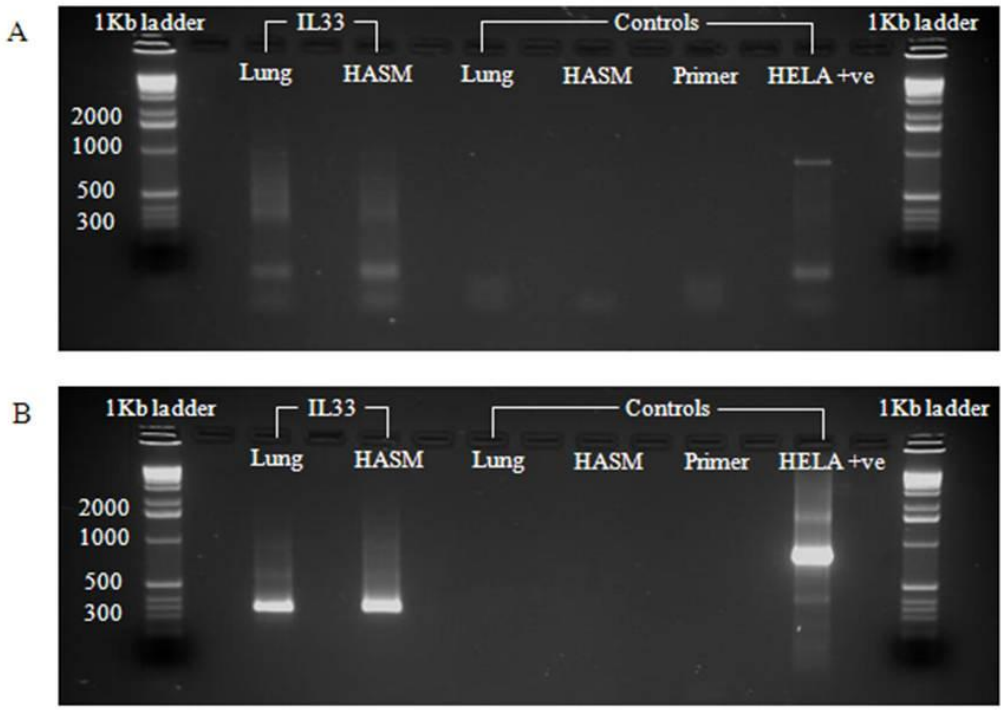


Figure 7.7 - 5' RACE PCR reactions for *IL33* in lung and HASM visualised using 2% agarose gel electrophoresis. A) PCR reactions performed on 1st strand RACE cDNA from lung and HASM total RNA. B) Nested PCR reactions performed on products shown in gel picture A. Products seen indicate a single main transcript in lung and in human airway smooth muscle (HASM). Primer negative (in lung and HASM) as well as template negative controls were free from contamination. An immortalised HeLa cell line (HELA) was used as the positive control.

7.4.3.2 Detected Transcripts

A total of 48 clones were obtained for each gene in each test template. Gel pictures following *Eco*RI plasmid digestion to check for the presence of inserts prior to sequencing are shown in Appendix 7. Following sequencing, the presence of the GeneRacer 5' primer sequences and the gene specific primers were verified prior to studying the transcript sequences.

For *IL1RL1*, a total of 60 clones were successfully sequenced: lung (n=21) and HMC-1 (n=39). The vast majority (95%) of clones in lung were reported transcript variant 2 (NM_003856.2), a single clone confirmed the presence of transcript variant 1 (NM_016232.4). Transcript 1 was confirmed by 49% of clones in HMC-1 and transcript 2 was also detected as well as three novel transcripts in this mast cell line, each identified through multiple clones [Figure 7.8]. For *IL33*, a total of 46 clones were successfully sequenced: lung (n=23) and HASM (n=23). All

clones matched the common 5' sequence of reported transcripts 1 and 2, which can be detected using the primers designed [Figure 7.9].

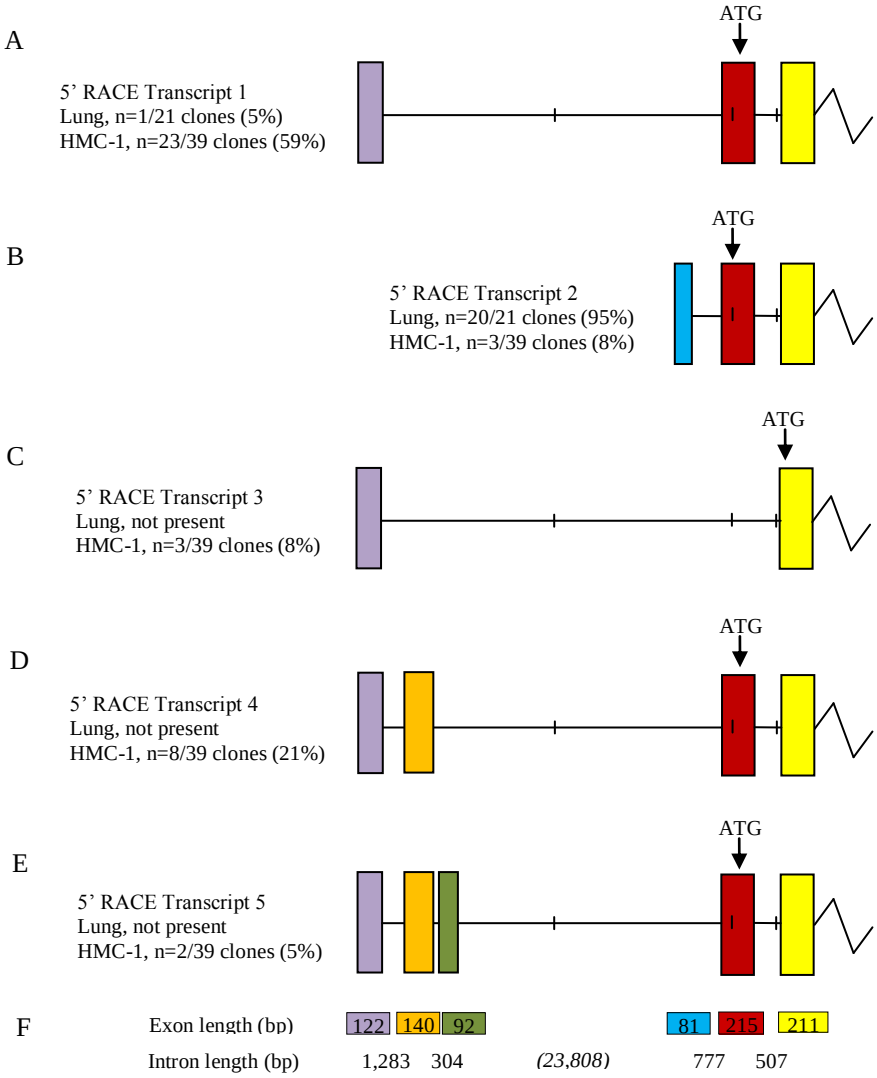


Figure 7.8 - Schematic representation of the *IL1RL1* transcript arrangements obtained from 5'RACE. A) Transcript 1 refers to published transcript 1 (NM_016232.4). B) Transcript 2 to published transcript 2 (NM_003856.2). C) Novel transcript 3. D) Novel transcript 4. E) Novel transcript 5. F) Exon and intron length (bp). ATG sites are marked.

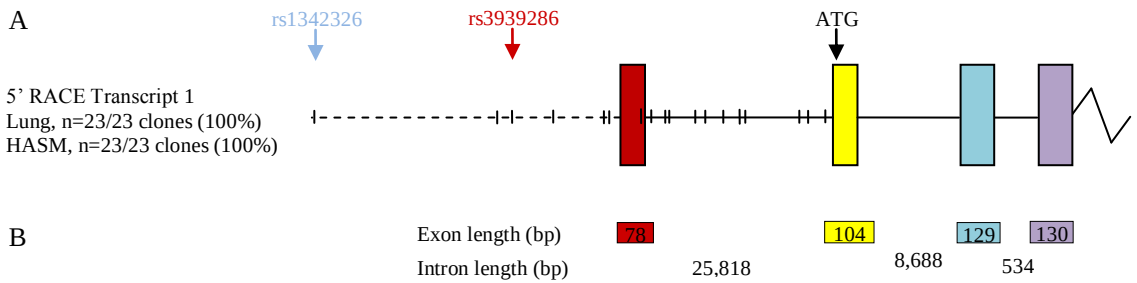
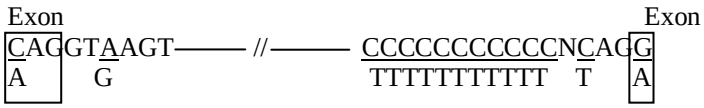


Figure 7.9 - Schematic representation of the *IL33* transcript arrangements obtained from 5'RACE. A) Single transcript detected matching published transcript 1 (NM_033439.3), 2 (NM_001199640.1), and 3 (NM_001199641.1). ATG sites, asthma associated SNPs, rs3939286 reported by deCODE and rs1342326 reported by GABRIEL are marked.

7.4.3.3 Consensus Splice Sites for Novel Exons

Novel exons were identified in *IL1RL1* transcripts 3, 4, and 5 detected by 5' RACE. Sequences surrounding each splice site obeyed the GT-AG rule and showed high levels of concordance with the consensus splice site sequences for the donor site and acceptor sites (Breathnach *et al.* 1978; Zhang 1998):

In genomic DNA



In cDNA

This provided further confidence that the exon arrangements determined by 5'RACE were not artefacts [Table 7.1].

Table 7.1 - Sequences around splice borders of novel exons in *IL1RL1* RACE transcripts show agreement with consensus splice sequences. Intronic sequence is shown in subscript. Green = mismatch to consensus, red = matches consensus, black = surrounding sequence.

	Donor site	**Branch point**	**Acceptor site**
Transcript 3			
Exon 1-2	AATGAGG$_{GTAAGTG}$	AATGAGG|GTAAACA	$_{TTTGCAG}$GTAAACA
Transcript 4			
Exon 1-2	AATGAGG$_{GTAAGTG}$	AATGAGG|GCATAAG	$_{CTTCCAG}$GCATAAG
Exon 2-3	TGCAGAG$_{GTTTGGA}$	TGCAGAG|ATATAGG	$_{AGTTGAG}$ATATAGG
Transcript 5			
Exon 1-2	AATGAGG$_{GTAAGTG}$	AATGAGG|GCATAAG	$_{CTTCCAG}$GCATAAG
Exon 2-3	TGCAGAG$_{GTTTGGA}$	TGCAGAG|ATATAGG	$_{AGTTGAG}$ATATAGG
Exon 3-4	TGTTTAG$_{GTGAGGA}$	TGTTTAG|ATATAGG	$_{AGTTGAG}$ATATAGG

7.4.3.3 Analysis of Transcriptional Start Sites

Multiple transcriptional start sites (TSSs) were identfied for both genes. For *IL1RL1*, the TSSs for exons were spread over regions of 105bp on transcript 1 and 99bp on transcript 2; a potential TATA box site was found in transcript 1 approximately 22bp to 27bp upstream of a cluster of TSS sites [Figure 7.10]. Detected novel transcripts 3, 4, and 5 share the same exon 1 as transcript 1. Combining clones for lung and HMC-1, the TSSs for transcript 1 clustered at one main site, accounting for 41% of clones and resulting in a 98bp length exon 1. Three main sites were seen for transcript 2 resulting in exon 1 with lengths 38bp, 80bp, and 81bp collectively accounting for 52% of clones, the TSS coding for the 81bp exon 1 is reported in the literature.

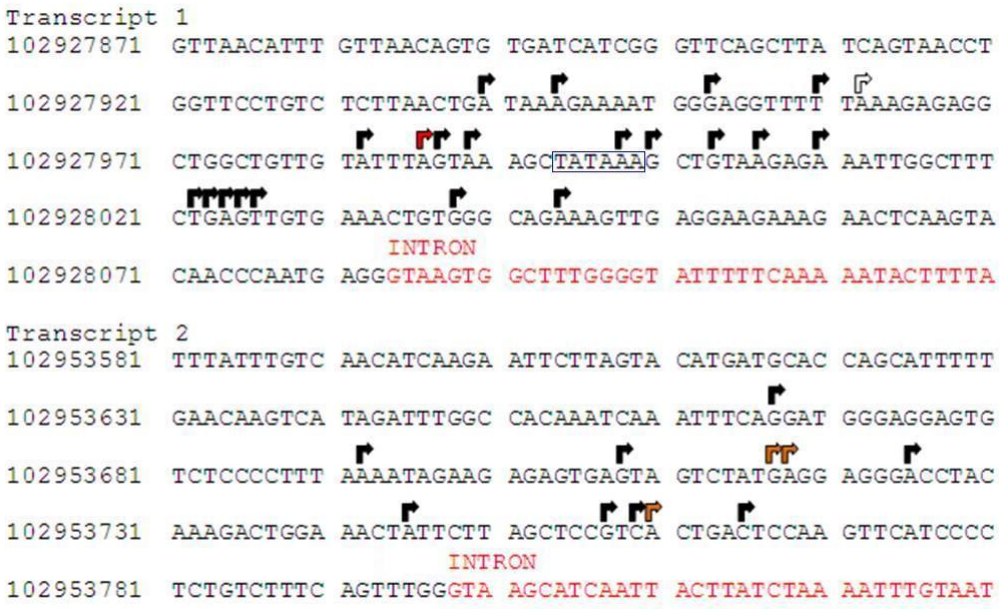


Figure 7.10 – Transcription start sites (TSSs) for *IL1RL1* transcripts obtained from 5' RACE. Combining clones for lung and HMC-1, the TSSs for transcript 1 clustered at one main site (red), no clones with the same TSS as reported in the literature were seen (white). Three main sites were seen for transcript 2 (orange), the first being the reported TSS. A potential TATA box site is shown in the blue box.

Combining *IL33* clones for lung and HASM, the TSSs clustered at one main site, accounting for 58% of clones and resulting in a 42bp length exon 1. No clones with exactly the same TSS as reported in the literature were seen. The TSSs for exons were spread over a 67bp region [Figure 7.11].

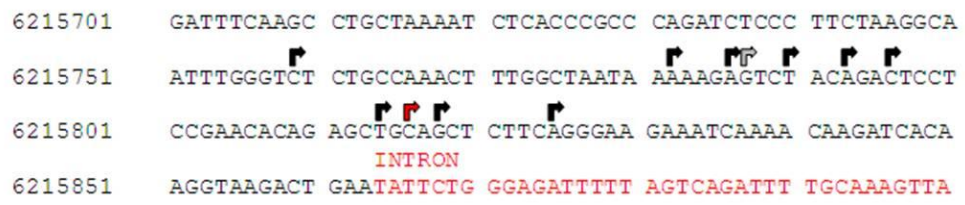


Figure 7.11 – Transcription start sites (TSSs) for *IL33* transcripts obtained from 5' RACE. Combining clones for lung and HASM, the TSSs clustered at one main site (red), no clones with the same TSS as reported in the literature were seen (white).

7.4.4 IL1RL1 Open Reading Frame Analysis

The 5' UTRs created by the novel *IL1RL1* transcript variants detected by 5' RACE were examined to identify any open reading frames (ORFs) present which may potentially result in different protein coding or have regulatory effects on downstream gene expression. Information from the UniProtKB/Swiss-Prot online database show that the longer transcript 1 codes for the transmembrane (NP_057316.3) and the shorter transcript 2 codes for the soluble (NP_057316.3) forms of the IL1RL1 receptor protein respectively.

Protein prediction for the 3 novel transcripts detected in HMC-1 was carried out using EMBOSS Transeq (EMBL-EBI), in-frame amino acid sequences for each transcript are detailed in Appendix 8. Transcript 4 lacks exon 2 which contains the published ATG site, the resulting ATG for this transcript may lie within exon 3 coding for a truncated protein with a loss of 117 amino acids. Transcripts 3 and 5 contain additional exons but are likely to have the same ATG site and subsequently code for the same protein as transcript 1.

7.4.5 Promoter Binding Sites

The sequences 2kb upstream of the most 3' TSS for *IL1RL1* and *IL33* were used to identify putative promoter binding sites [Figure 7.12]. Multiple promoter binding sites were identified across all *IL1RL1* and *IL33* transcripts with consensus from ≥2 online databases for Nuclear factor-κB (NF-κB), activator protein-1 (AP-1), CAAT/Enhancer binding proteins (C/EBP), signal transduction-activated transcription factors (STAT), cyclic AMP response element binding protein (CREB), CREB binding protein (CBP), and nuclear factor of activated T-cells (NF-AT) were analysed (Barnes and Adcock 1998). C/EBP and CREB binding sites were not identified for *IL1RL1* transcript 2.

Figure 7.12 Promoter binding site analyses (2kb). The sequences upstream of the most 3' TSS were used to identify putative promoter binding sites for all transcripts of *IL1RL1* and *IL33*. Transcription factors (AP-1 ◇, C/EBP ▮, CREB ▮, GR ▪, CBP ▮, GATA △) are positioned on the scale relative to the distance from the first exon.

7.5 Discussion

This Chapter describes work carried out using a 5' RACE approach to amplify the extreme 5' ends of mRNA transcripts of *IL1RL1* from lung tissue and a human mast cell line (HMC-1), as well as *IL33* from lung tissue and human airway smooth muscle (HASM) cells. *IL1RL1* (IL1RL1) and *IL33* (IL33) have been recognised to exert an important role in the development of inflammation and airway hyperresponsiveness in asthma by both genetic association and functional studies (Schmitz *et al.* 2005; Reijmerink *et al.* 2008; Ali *et al.* 2009; Gudbjartsson *et al.* 2009; Moffatt *et al.* 2010). However, the transcriptional regulatory mechanisms operating on their expression in human airway cells and tissues are poorly understood. Furthermore, the effects of identified polymorphic variation within these genes are still to be determined. The presence of previously published transcripts were confirmed in tested templates as well as three novel *IL1RL1* transcripts with a potentially alternative open reading frame. Results were analysed to establish the positions of the transcription start sites (TSSs) and multiple locations for transcription factor binding were found in the putative promoter regions of these genes.

Both the *IL1RL1* and *IL33* genes were found to be expressed across templates in a representative airway cells and tissues panel with higher levels of *IL1RL1* expression particularly in HMC-1 inflammatory cells and *IL33* preferentially expressed in structural cells such as HASM and HBECs. This supports findings from previous studies showing abundant *IL33* mRNA expression in multiple tissue related cells types including endothelial cells, smooth muscle, and mucosal epithelial cells (Schmitz *et al.* 2005; Moussion *et al.* 2008). Expression of *IL1RL1* was reported previously to be predominantly by cells involved in allergic inflammation such as mature mast cells (Gächter *et al.* 1998; Moritz *et al.* 1998; Allakhverdi *et al.* 2007). Injury to epithelial cells may increase the product of IL33 cytokines, which in turn result in recruitment of IL1RL1 to the affected site. Results from 5' RACE have confirmed the presence of the two published transcripts for *IL1RL1* and the three published transcripts for *IL33*. Two variant forms of the IL1RL1 receptor protein are known to exist, a transmembrane form and a soluble form coded by the longer transcript 1 and shorter trancript 2 respectively. A third variant encoding a truncated membrane-bound form of the receptor

protein has also previously been published but weakly supported (Tominaga *et al.* 1999; Tago *et al.* 2001). Findings from this Chapter have also detected the presence of three novel *IL1RL1* transcripts closely related to published transcript 1, one of which potentially codes for a shorter 5' end of the transmembrane form of the receptor.

The 5' UTRs of detected transcripts for each gene differ because of the variation in the 5' terminal exons. It is therefore possible that regulation at the transcript level differs between the splice variants observed by potentially introducing or removing regulatory elements or by causing the gene to be expressed from an alternative promoter. All TSSs are more widespread for *IL1RL1* over an 105bp region on transcript 1 and a 99bp region on transcript 2 compared to *IL33* over a 67bp region for all transcripts. A potential TATA box was identified for *IL1RL1* transcript 1. A range of TSSs over a large area could be due to fewer clones produced from the experiment reducing the likelihood of identifying definite clusters or due to artefacts of degredation. Alternatively, multiple locations for TSSs could also suggest that gene expression is differentially regulated by the use of alternate promoters and has cell-specific expression patterns (Smale 1997). This is indicative of the ability of RNA polymerase II to initiate mRNA synthesis over a region of sequences. The TATA box motif is important in establishing the position of the transcriptional initiation of RNA polymerase II. Genes without TATA boxes are reported to show greater variation in TSS intiation positions (Smale 1997).

These results provides evidence to suggest that *IL33* transcriptional regulation is dependent on fewer controlling mechanisms than *IL1RL1* and therefore, potentially under tighter regulation. This may help to explain the functional roles of these genes in asthma. Structural changes in the airway following tissue injury may increase production of IL33 cytokines and transmembrane forms of the IL1RL1 receptor (Oshikawa *et al.* 2001; Smith 2010). During exacerbations and the induction of the inflammatory cascade, soluble IL1RL1 receptors may be upregulated via the recruitment of inflammatory cells such as mast cells to act as a decoy receptor (Allakhverdi *et al.* 2007).

Multiple promoter binding sites for transcription factors reported to be important in asthma were identified in the 2kb 5' region upstream of each gene including AP-1, C/EBP, CREB, GR, CBP, and GATA (Barnes and Adcock 1998). Altered gene function may occur not only due to coding region variability but also due to polymorphic variation within the promoter. Promoter polymorphisms might result in altered expression levels of an otherwise normal protein potentially resulting in a modified behaviour of gene-product dependent pathways.

Previous studies have proposed ways in which differing promoter usage and alternative splicing in *IL1RL1* and *IL33* may be involved in determining asthma susceptibility (Iwahana *et al.* 1999; Kato and Schleimer 2007; Smith 2010). The use of both proximal and distal promoters with multiple TSSs in other cells types has also been shown (Iwahana *et al.* 1999). The SNPs reported for association with asthma identified by the deCODE Icelandic and the GABRIEL GWA studies were mapped relative to identified transcripts for *IL1RL1* and *IL33*. The signal generated by the SNP reported by deCODE, rs1420101 within intron 5 of *IL1RL1* could, through tagging of more 5' causative polymorphisms, be driven by differential expression of the different receptor splice variants resulting in production of different levels of the *IL1RL1* receptor protein in cases compared to controls (Schmitz *et al.* 2005). Intronic SNPs may also influence either the transcriptional activity or the splicing efficiency of the genes, or alter the expression of alternative transcripts (Manolio *et al.* 2009). Association with *IL33* SNPs may be due to effects on promoter binding sites therefore determining levels of transcription. Both SNPs flanking *IL33* reported by the deCODE Icelandic and GABRIEL studies are 5' of the coding regions: rs13423626 being 32kb upstream (deCODE) and rs3939286 being 52kb upstream (GABRIEL) of the identified TSSs; these could be tagging polymorphisms within distal promoters positioned more closely to the gene coding regions.

Difficulty in determining the functionality of these SNPs is also in part due to the strong LD structure of the 2q12 locus making it uncertain where the true causative origin of the association signal lies (see Chapter 5 Discussion). Mapping of genotyped SNPs tested for association in the AUGOSA GWA study described in Chapter 5 across the *IL1RL1* and *IL33* transcript regions has shown that coverage using current genotyping platforms is still relatively

low. Furthermore, it is likely that these detected polymorphisms are tagging rare variants within the identified locus through synthetic associations that have not been able to be detected by current GWA approaches (Dickson *et al.* 2010). These questions will only be able to be addressed by advances in next-generation sequencing, which have made re-sequencing candidate genes more affordable to identify rare variants within hypothesised or previously identified loci. This approach was used in Chapter 8 to further study the *IL1RL1* and *IL33* susceptibility loci on chromosome 2q12 and 9p24.1 respectively.

Chapter 8

Re-sequencing to Identify Rare Variants within the Chromosome 2q12 and 9p24.1 Loci

8.1 Introduction

Substantial evidence for the role of the chromosome 2q12 and 9p24.1 regions in asthma has been found. However, effect sizes for identified polymorphisms remain small and it has been difficult to determine which genes the association signals are originating from. Although associations of polymorphisms within 2q12 in particular and asthma have been widely replicated, few direct SNP associations have been reproduced (Reijmerink *et al.* 2008; Ali *et al.* 2009; Gudbjartsson *et al.* 2009; Reijmerink *et al.* 2010; Wu *et al.* 2010). This loose region-based replication as well as the work described in this thesis and the wider literature suggests that there may be multiple signals in the relevant regions containing *IL1RL1*/*IL18R1*, and *IL33* (Savenije *et al.* 2011). In addition, identified SNPs may be tagging other polymorphisms in nearby genes or rare variants. Furthermore, an individual may be predisposed to disease as a result of individual variants with functional effects or multiple variants within a gene having a deleterious effect in synergy.

As explained in Chapter 1 and shown by results from Chapter 5 to 7, due to small effect sizes, it is unlikely that common variants assayed using current SNP genotyping platforms will explain all the heritability still to be identified in asthma and there is increasing interest in the contribution of rare variants (Maher 2008; Moffatt *et al.* 2010). The common disease, rare variant hypothesis proposes that multiple rare genetic polymorphisms, each with relatively high penetrance, form major contributions to common disease susceptibility (Bodmer and Bonilla 2008). Population-level processes such as mutation, selection, and random genetic drift are likely to have acted against common variants favouring the existence of multiple rare variations contributing to disease arising from recent population growth (Pritchard 2001;

Coventry *et al.* 2010). Rare and potentially functionally significant polymorphisms have also been shown to occur naturally in many genes affecting complex, multi-factorial traits (Crawford *et al.* 2004; Schork *et al.* 2008).

Data released from the August 2010 pilot of the 1000 Genomes Project has made available an online database of eight million previously unknown SNPs illustrating the fact that potentially a large proportion of the effects of specific polymorphisms on disease risk still remain to be identified. These less frequent variants have also been proposed as more likely to be functional than predominantly common variations tested using most genome-wide SNP genotyping platforms (Gorlov *et al.* 2008). Findings from some previous studies in the literature have shown that rare variants are likely to have larger effect sizes with an OR between 2 and 5 filling the effect gap of detected heritable variants between linkage and association studies (Bodmer and Bonilla 2008) (see Chapter 1, Section 1.4).

Recent advances in next-generation sequencing (NGS) have allowed affordable re-sequencing of candidate genes to test for association of previously unidentified rare variants with disease. This approach will help to determine the relationship of common and rare polymorphisms and identify the location of causative polymorphisms; this approach has recently being used to identify low frequency variants determining high plasma triglyceride levels (Johansen *et al.* 2010). Studies in other complex diseases have observed frequency differences between cases and controls in rare SNPs identified through sequencing of coding regions in candidate genes and loci (Cohen *et al.* 2004; Fearnhead *et al.* 2004; Ji *et al.* 2008). Although this has been to a certain extent subject to publication bias, these findings indicate that multiple rare variants may be associated with susceptibility to complex disease.

8.2 Aim

The aim of this Chapter was to carry out re-sequencing of the chromosome 2q12 and 9q24.1 loci in 200 severe asthma cases and 200 general population controls to identify whether an excess of variants contribute to the determination of developing severe asthma as well as establish if unique variants exist in the asthma and control populations.

8.3 Methods

8.3.1 Study Populations

A total of 200 severe asthma cases and 200 controls were used in re-sequencing. Case subjects consisted of 200 individuals from the AUGOSA study. A subset of general population controls was selected from the Nottingham Gedling cohort to be non-asthmatic, non-wheeze, and non-atopic controls. Details of study recruitment and phenotyping are described in Chapter 2: Sections 2.1 and 2.2. Selection of individual samples were made based on the highest yields of DNA available in order to maximise the success of the sequencing reaction (minimum required concentration 50ng/ul) whilst attempting to approximately sex-match the case and control groups. Control individuals were chosen to be slightly older in order to provide further assurance of their non-asthmatic status. Using these criteria, the resulting samples were made up of cases in which 69.4% were female with an average age of 48.1 years (SD 14.9) and controls in which 73.0% were female with an average age of 56.7 years (SD 13.0).

8.3.2 Region Selection

For both the chromosome 2q12 and 9p24.1 loci, regions were defined to include the entire association signal from genotyped data by either the most distal SNP with $-\log_{10}$ p value >2.5 on either side of the sentinel SNP, or by the co-ordinates of the *IL1RL1/IL18R1* and *IL33* genes (±10kb), whichever was furthest from the sentinel SNP. All co-ordinates were converted from NCBI Build 36.1 to Genome Reference Consortium GRCh37 using the UCSC Batch Convertor tool and the Lift Genome Annotations tool.

8.3.3 The 1000 Genomes Project

Sequencing and genotyping efforts carried out by the 1000 Genomes projects are currently underway and data is being updated and deposited online continuously. The latest data releases were analysed using the Variant Call Format (VCF) version 4.0 files uploaded online (http://www.1000genomes.org/) for genotype and indel data as well as the browser for sequence data. The full project genotype release (16 December 2010) and indel release (16

February 2011) was available at the time of analysis. This resource was used to evaluate new variants currently identified for the *IL1RL1* and *IL33* regions of interested intervals, defined by chromosome 2: 102,738,158-103,208,945bp and chromosome 9: 5,924,967-6,267,982bp respectively based on NCBI Build 37.1 co-ordinates. Sequence data was assessed using the online browser (December 2010). Genotype and indel data containing calls based on 629 individuals was extracted from the EBI FTP site (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/) using the following commands using tabix in UNIX:

tabix -h
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20100804/ALL.2of4intersectoin.20100804.genotypes.vcf.gz
2:102738158-103208945
9:5924967-6267982

8.3.4 Re-sequencing Design

All technical stages of re-sequencing in terms of library preparation and target enrichment, cluster generation and sequencing, base calling and data processing were carried out by SourceBioscience (Nottingham, UK). Paired-end sequencing using 100bp reads for cases and controls were run on two separate lanes in a single flow cell using the Illumina HiSeq2000™ Systems pipeline designed by Illumina (San Diego, USA).

Due to the cost implications of generating individual level data, a degree of multiplexing was carried out by splitting cases and controls into three randomly assigned pools of 66, 66, and 68 individuals. Concentrations of each sample were determined using an Agilent BioAnalyzer prior to pooling in order to ensure that each sample contributed the same quantity of DNA for sequencing. DNA samples within each pool were sheared in order to carry out target sequence enrichment using 120bp oligos (baits). These were designed using web-based design tool eArray (http://earray.chem.agilent.com) for use in the Agilent SureSelect Custom Capture kit (x10 reactions) with capture size range of 500kb to 1.5Mb. Biotinylated RNA baits hybridise to regions of interest on fragmented DNA, which allow these to be selected out using streptavidin coated magnetic beads for focused amplification prior to cluster generation and sequencing (Mamanova *et al.* 2010). Indexing was then carried out by bar-coding each pool of samples

using unique adaptor sequences prior to library preparation and subsequent sequencing (Harismendy and Frazer 2009).

Illumina sequencing uses automated cluster generation using cBot followed by sequencing-by-synthesis reversible terminator technology (Shendure and Ji 2008; Metzker 2010). Genomic DNA is randomly fragmented and bound to two unique adaptor oligo-nucleotides at either ends with resulting fragments of approximately size 150bp to 200bp. This allows hybridisation to a lawn of oligo-nucleotides corresponding to the sequences of ligated adapters immobilized on the surface of flow cell channels. Following 3' extension from hybridised adaptors, the original templates are denatured leaving copies immobilised onto the flow cell surface. The free, distal ends of these fragments bind to complementary oligos nearby. Solid phase bridge amplification is then carried out to synthesis dense clusters of double stranded fragments (approximately 2,000) in order to produce stronger fluorescence signals [Figure 8.1]. The reverse strand is then removed to increase the efficiency of sequencing, carried out using labelled reversible nucleotide terminators, primers, and DNA polymerase. Emitted fluorescence from each cluster is captured following laser excitation and bases identified sequentially.

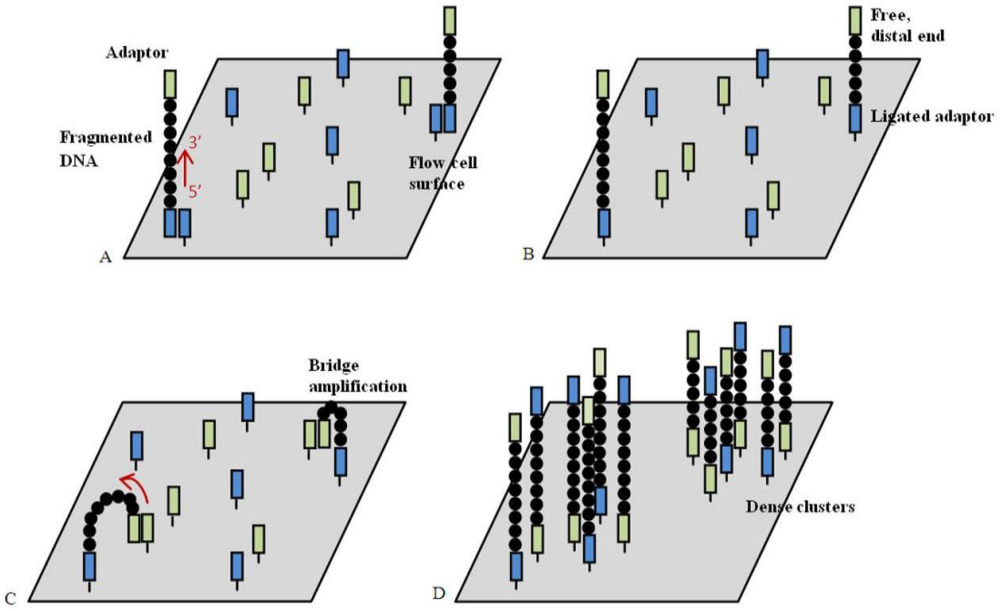


Figure 8.1 – Solid phase bridge amplification and cluster formation prior to Illumina sequencing. A) Adaptor bound DNA fragments hybridise to ligated adaptors on the flow cell surface. B) Original templates are denatured to leave immobilised copies following 3' extension. C) Free, distal ends bind to complimentary copies for bridge amplification. D) Dense template clusters are generated by repeated extension and denaturation.

As mentioned above, the regions defined for re-sequencing were chromosome 2: 102,738,158-103,208,945bp (343,015bp size) and chromosome 9: 5,924,967-6,267,982bp (470,787bp size) based on Genome Reference Consortium GRCh37 co-ordinates. The total region size sequenced in a single lane was therefore 53.7Mb (813,802bp x66). According to Illumina product specifications, approximately 10GB of data for each lane of sequencing with 60% on-target efficiency is expected using the HiSeq2000 Systems pipeline. Therefore, the estimated average coverage per individual was 37-fold calculated (10GB data/3 pools/54MB region x 0.60 efficiency).

8.4 Results

8.4.1 Region Selection

A total region of 814kb covering association signals in the *IL1RL1/IL18R1* and *IL33* regions was identified for re-sequencing [Table 8.1]. For *IL33*, which did not show significant regions of association in AUGOSA, the region was extended to include the complete *IL33* gene (+10kb) which was intersected by the cut off defined by the association region [Figures 8.2 and 8.3].

Table 8.1 – Re-sequencing regions on 2q12 and 9p24.1 based on Genome Reference Consortium GRCh37.

Target		**Region**		
Chromosome	**Gene**	**Start (bp)**	**End (bp)**	**Length**
2q12	*IL1RL1/IL18R1*	102,738,158	103,208,945	470,787bp
9p24.1	*IL33*	5,924,967	6,267,982	343,015bp
			TOTAL	**813.802kb**

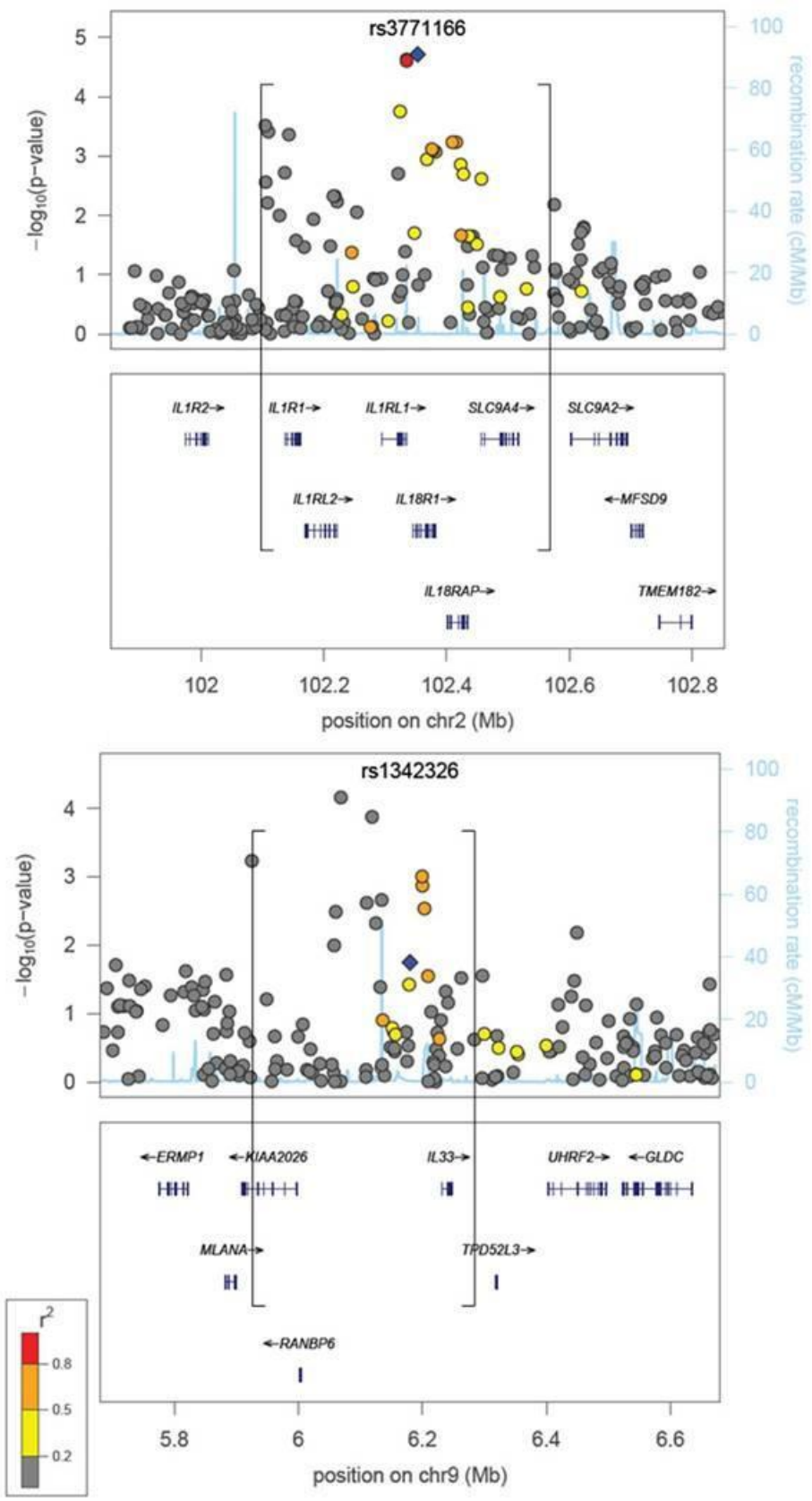


Figure 8.2 – Annotated regional plots for the sentinel SNPs on 2q12 (rs3771166) and 9p24.1 (rs1342326) used to define regions for re-sequencing. Defined regions are shown by the square brackets.

8.4.2 Results from the 1000 Genomes Project

Data from the latest releases from the 1000 Genomes Project were analysed for chromosome 2q12 and 9p24.1 regions defined for re-sequencing. Polymorphisms not previously published and not tested in the AUGOSA GWA study within *IL1RL1*, *IL18R1*, and *IL33* plus 2kb upstream of all three genes identified in the CEU population (n=60 at low coverage x2 to x6, n=3 from a trio family) are summaried in Table 8.2. Results from the 1000 Genomes Project from sequencing at low coverage have identified 21 novel SNPs (MAF <1%: 17 SNPs, <2%: 2 SNPs, <3%: 2 SNPs) in *IL1RL1*, 12 novel SNPs (all MAF <1%) in *IL18R1*, and three novel SNPs (all MAF <1%) in *IL33*. A single synonymous SNP in *IL1RL1* was identified; all remaining SNPs were in the 5' UTR, 3' UTR, and intronic regions. Sequencing of a trio family has identified four indels in *IL1RL1*, nine indels in *IL18R1*, and a SNP and three indels in *IL33* [Table 8.2].

Table 8.2 – New variants in *IL1RL1, IL18R1,* and *IL33* identified by the 1000 Genomes Project from sequencing CEU populations at low coverage and a trio family not present on the genotyping platforms used in the AUGOSA study. Ref and alt denotes reference and alternate alleles respectively. Positions are based on Genome Reference Consortium GRCh37. * Called in the trio family only.

Chr	Gene	Position (bp)	Variation	Ref/alt	Function	MAF
2q12	*IL1RL1*	102929183	SNP	G/A	5' UTR	0.008
		102930603	SNP	C/A	5' UTR	0.017
		102930605	Indel	T/TAGC	5' UTR	Tri o *
		102931322	SNP	G/A	5' UTR	0.025
		102931527	SNP	A/G	5' UTR	0.008
		102931622	SNP	G/T	5' UTR	0.008
		102931763	Indel	A/AC	5' UTR	Tri o *
		102935208	SNP	G/A	5' UTR	0.008
		102935998	SNP	T/G	5' UTR	0.008
		102939614	SNP	A/G	5' UTR	0.008
		102940755	SNP	T/C	5' UTR	0.017
		102949305	Indel	C/CA	5' UTR	Tri o *
		102949645	SNP	A/T	5' UTR	0.025
		102949652	SNP	T/C	5' UTR	0.008
		102949848	SNP	G/A	5' UTR	0.008
		102951053	Indel	G/GA	5' UTR	Tri o *
		102952845	SNP	A/G	5' UTR	0.009
		102952902	SNP	T/C	5' UTR	0.007
		102954495	SNP	A/G	Intronic	0.008
		102954751	SNP	C/T	Synonymous: Leu [L]	0.002
		102957731	SNP	G/T	Intronic	0.008
		102958648	SNP	C/T	Intronic	0.004
		102960262	SNP	A/G	Intronic	0.008
		102960632	SNP	A/G	Intronic	0.004
		102964745	SNP	T/A	Intronic	0.002
2q12	*IL18R1*	102980480	SNP	T/C	Intronic	0.008
		102980590	SNP	A/C	Intronic	0.005
		102985646	Indel	G/GC	Intronic	Tri o *
		102986971	SNP	A/G	Intronic	0.009
		102987503	Indel	C/CT	Intronic	Tri o *
		102988409	SNP	A/G	Intronic	0.008
		102989060	Indel	C/T	Intronic	Tri o *
		102989580	SNP	C/T	Intronic	0.008
		102990071	Indel	T/TA	Intronic	Tri o *
		102991422	SNP	T/C	Intronic	0.004
		102993385	SNP	C/A	Intronic	0.008
		102996624	Indel	C/CACTG	Intronic	Tri o *
		102997129	Indel	C/CACTG	Intronic	Tri o *
		103001628	SNP	G/A	Intronic	0.008
		103003973	SNP	T/G	Intronic	0.007
		103004251	SNP	A/G	Intronic	0.008
		103006399	SNP	G/A	Intronic	0.001
		103007692	Indel	G/GC	Intronic	Tri o *
		103009730	Indel	G/GA	Intronic	Tri o *
		103010547	SNP	T/C	Intronic	0.008
		103013388	Indel	T/TG	Intronic	Tri o *
9q24.1	*IL33*	6242273	SNP	T/G	Intronic	0.002
		6242785	SNP	T/G	Intronic	0.001
		6244171	SNP	C/A	Intronic	0.008
		6249037	SNP	C/T	Intronic	0.004
		6252040	SNP	A/C	Intronic	Tri o *
		6252056	Indel	C/CA	Intronic	Tri o *
		6252330	Indel	T/TAA	Intronic	Tri o *
		6257936	Indel	T/TCTA	3' UTR	Tri o *

8.4.3 Bait Design for Target Enrichment

Bait groups for each region were designed using paired-end oligos with 120bp length centred to give even distribution across the target region using eArray. The allowed overlap outside of the defined region was restricted to 20bp and a 5x tiling frequency was used. Standard repeat masked regions based on the UCSC RepeatMasker track were avoided. The resulting baits for each region were downloaded as a .zip file containing a .summary file describing the design and results, a .fate file listing the number of baits generated per target region, a .bed file mapping the bait positions, and a .tdt file detailing each bait sequence and targeted position.

A total of 8,886 oligos (8,705 in normal regions, 181 in regions with high GC content) were designed for chromosome 2q12. A total of 4,503 oligos (4,448 in normal regions, 55 in regions with high GC content) were designed for chromosome 9q24.1. The .bed files were uploaded using the UCSC Genome Browser for visualisation and analysed with the .fate file showing that successful bait coverage across each region [Figure 8.4]. High GC baits (>60%) were identified using Microsoft Excel in order to enrich for GC rich regions through inclusion of these baits to normal GC baits in a ratio of 4:1 in the final library.

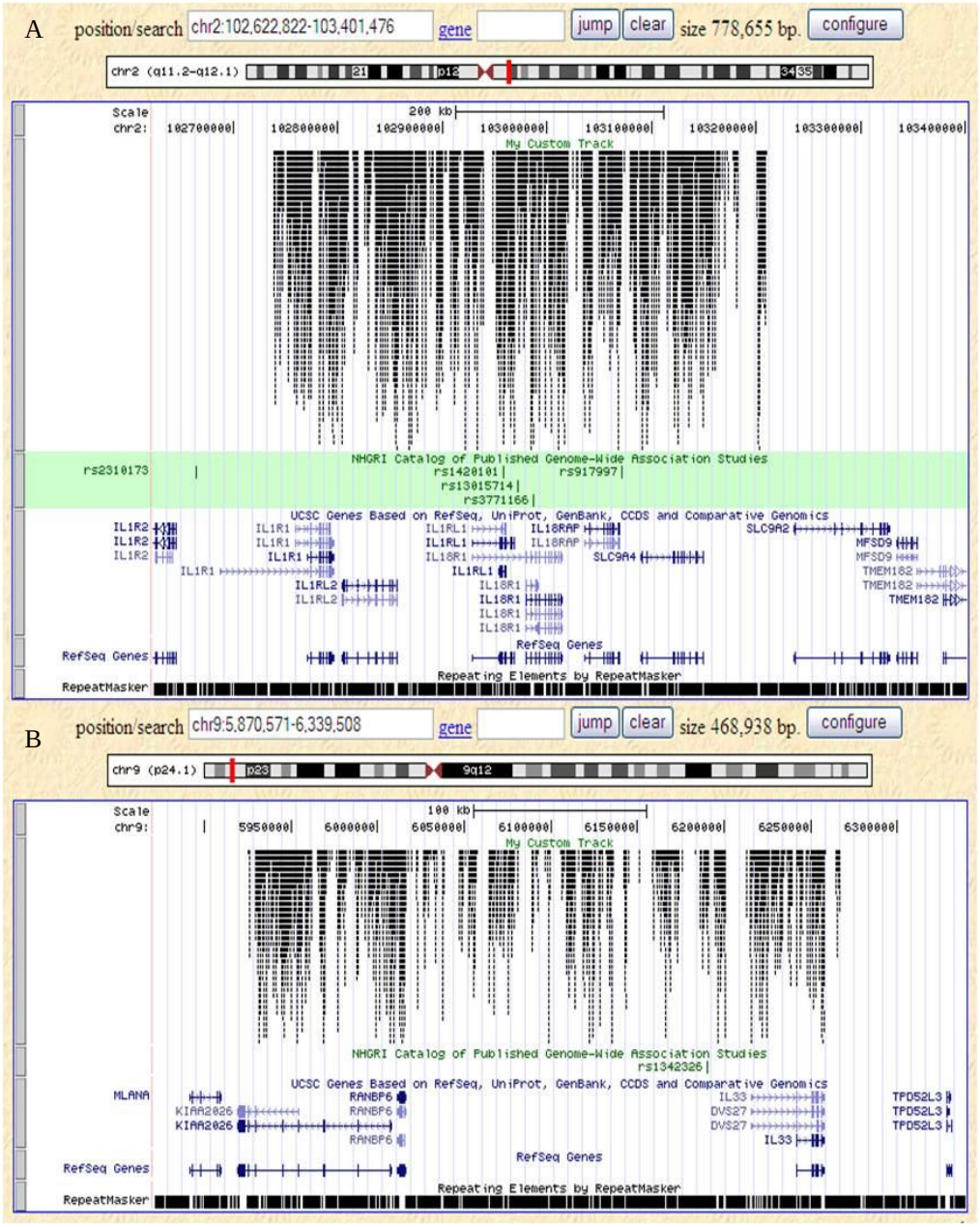


Figure 8.3 – Baits (RNA oligo probes) designed to cover A) the *IL1RL1/IL18R1* region on 2p12 and B) the *IL33* region on 9p24.1 visualised using UCSC Genome Browser.

8.4.4 Variants Identified by Re-sequencing

Results of re-sequencing were provided by SourceBioscience following base calling. In collaboration with Dr. N G Shrine, reads from sequencing were aligned and compared to the reference genome. SNP and indel variant detection was then carried out using Syzygy software (http://www.broadinstitute.org/software/syzygy/) by scanning each genomic position to identify positions at which differences between read and reference occur in 'significant' amounts.

Indications of quality were given for variants called, this is a stringent threshold based on whether or not there is concordance between the number of alleles counted on the plus strand and the number counted on the negative strand. A Fisher's exact test was used to compare if the proportion of non-reference to reference alleles is the same between the plus and negative strands and a variant was classed as high quality if $p>0.1$ and low if $p<0.1$. This was not used to further filter the data as some low quality variants may have been real as evidenced by mapping to an rs code from dbSNP. A total of 9,560 variants were called across all pools, of which 4,851 were high quality. In the total regions sequenced, 30.5% of known SNPs documented on dbSNP and the 1000 Genomes project was identified overall (1,596 out of 4,488 SNPs on 2q12, 1,760 out of 6,504 SNPs on 9p24.1). This was expected as the average MAF of SNPs which were not identified was 0.02. In addition, the list of known SNPs was compiled across multiple populations and therefore numerous SNPs may not be in Caucasian populations.

Excluding called SNPs mapped to known variants documented by dbSNP and the 1000 Genomes Project, using a filter of low quality scores corresponding to $p<0.1$ error, 48.9% of all SNPs called across all pools would have been removed [Table 8.3]. However, all of these 4,666 low quality SNPs were observed in all six pools indicating that these were likely to have been real. Therefore, putative variants for comparison of cases and controls were not filtered using quality thresholds in individual pools. This was chosen in order to not to remove potentially real variants through quality threshold bias especially those which may have been

very rare whilst ensuring some confidence that they were not artefact as demonstrated by reproducibility across independent pools.

Multiple SNPs and indels were detected in both the chromosome 2 and 9 regions re-sequenced. The distribution of these variants between the cases and controls are listed in Table 8.3 and grouped by level of confidence using the quality threshold in Table 8.4. For all variants called, 71.9% (4,080 out of 5,676) on 2q12 and 77.3% (3,004 out of 3,884) on 9p24.1 were novel to our samples. Novel SNPs within these regions were not present on the genotyping platforms used in the AUGOSA study and therefore had not been tested for association in Chapter 5. All replicated variants were compared and both cases and controls were found to carry multiple variants unique to each affection status group.

Table 8.3 – Numbers of variants detected by re-sequencing on 2q12 and 9p24.1. Novel SNPs did not match to a rs ID in dbSNP or variant documented by the 1000 Genomes Project (all populations). High quality denotes $p<0.1$ and low quality denotes $p>0.1$ using a Fisher's exact test to compare concordance of base calling between strands.

Region		**2q12**	**9p24.1**
Total variants detected			
Controls	Pool 1	5,466 (2,658 High quality)	3,690 (1,789 High quality)
	Pool 2	5,506 (2,698 High quality)	3,759 (1,858 High quality)
	Pool 3	5,477 (2,669 High quality)	3,718 (1,817 High quality)
Cases	Pool 4	5,360 (2,552 High quality)	3,620 (1,719 High quality)
	Pool 5	5,399 (2,591 High quality)	3,648 (1,747 High quality)
	Pool 6	5,416 (2,608 High quality)	3,706 (1,805 High quality)
Variants replicated across pools			
Common to all pools (controls and cases)		5,580 (2,772 High quality)	3,806 (1,905 High quality)
Unique to controls (pools 1, 2, and 3)		68 (68 High quality)	58 (58 High quality)
Unique to cases (pools 4, 5, and 6)		28 (28 High quality)	20 (20 High quality)
Total		5,676 (2,868 High quality)	3,884 (1,983 High quality)

Table 8.4 – Numbers of novel variants detected by re-sequencing on 2q12 and 9p24.1. Multiple variants unique to each pool of controls and cases were detected.

Region	**2q12**	**9p24.1**
Novel variants replicated across pools		
Common to all pools (controls and cases)	4,019 (1,235 High quality)	2,952 (1,070 High quality)
Unique to controls (pools 1, 2, and 3)	42 (42 High quality)	39 (39 High quality)
Unique to cases (pools 4, 5, and 6)	19 (19 High quality)	13 (13 High quality)
Total	4,080 (1,296 High quality)	3,004 (1,122 High quality)

Prediction of functional changes implicated by identified SNPs was carried out using FANS version 1.7 (Liu *et al.* 2008). The majority of novel variants did not lie within the *IL1RL1*,

IL18R1, and *IL33* genes. However, *IL1RL1* was found to carry three exonic SNPs with potential functional changes within both controls and cases in additional to multiple SNPs unique to controls and to cases. A single indel which may cause a change in the open reading frame of *IL18R1* was also detected. All remaining variants within these candidate genes were intronic [Table 8.5].

Table 8.5 – High quality novel variants detected by re-sequencing within candidate genes: *IL1RL1, IL18R1,* and *IL33.* Ref and alt denotes reference and alternate alleles respectively. Positions are based on Genome Reference Consortium GRCh37.

Gene	Position	Function	Ref	Alt
Common to all pools (controls and cases)				
IL1RL1	102955425	Non-synonymous: mis-sense (conservative change)	C	T
	102965608	Non-synonymous: mis-sense (protein domain abolished)	G	A
	102968231	Non-synonymous: mis-sense (conservative change)	G	T
	A further 75 intronic variants (and 6 variants within 2kb upstream) are not listed			
IL18R1	103013389	Open reading frame change	G	Deletion
	A further 80 variants (and 1 SNP within 2kb upstream) are not listed			
IL33	71 intronic variants are not listed			
Unique to controls (pools 1, 2, and 3)				
IL1RL1	102939726	Intronic	T	Insertion
	102939727	Intronic	A	G
	102939729	Intronic	A	G
	102939731	Intronic	A	G
	102939754	Intronic	G	A
	102939801	Intronic	T	Deletion
	102943140	Intronic	A	Deletion
	102945241	Intronic	G	A
	102946684	Intronic	G	A
	102946867	Intronic	G	T
	102946954	Intronic	A	Deletion
IL18R1	None detected			
IL33	None detected			
Unique to cases (pools 4, 5, and 6)				
IL1RL1	102945127	Intronic	G	A
	102947242	Intronic	G	A
IL18R1	None detected			
IL33	None detected			

8.5 Discussion

This Chapter describes work carried out to test whether the relative abundance of variants within the group of cases compared to the group of controls determines susceptibility to severe asthma within two regions identified by GWA studies on chromosomes 2q12 and 9q24.1. These were re-sequenced in two separate pools consisting of 200 severe asthma cases and 200 general population controls to identify the variants present and establish whether different variants exist between cases and controls.

The high depth of coverage from re-sequencing (estimate average of 37-fold) was powered to detect rare variants. Using pooled data, multiple SNP and indel variants were found in both the group of cases and the group of controls. Variants replicated across independent pools unique to each affection status group were identified: 42 in controls and 19 in cases on 2q12, 39 in controls and 13 in cases on 9p24.1. The increased allelic heterogeneity observed in the loci of interest seen provides further evidence to support the common disease, rare variant hypothesis in asthma (Pritchard 2001). The functional implications of these variants remains unclear as the majority of did not lie within the coding regions of the candidate genes *IL1RL1*, *IL18R1*, and *IL33*. This supports the observation that it is difficult to identify the causal variants responsible for the association signals detected by GWA studies. All SNPs within gene regions were found to be intronic. This challenges the hypothesis that rare and low frequency variants are more likely to have significantly functional effects such as those resulting in protein changes (Gorlov *et al.* 2008).

A relatively larger number of unique SNPs were called in controls compared to cases on both 2q12 and 9q24.1. Differences in the collective risk of harbouring rare SNPs detected between the controls and the severe asthma cases would support the familial cluster of disease seen in asthma. Theoretical calculations suggest that under the assumption of the existence of only a few predisposing variations with low penetrance as opposed to multiple rare variants with additive effects, complex diseases influenced by rare variants are unlikely to show familial clustering (Bodmer and Bonilla 2008). Variants detected by high depth re-sequencing may be

individually rare but sufficiently common in aggregate to contribute to the variation seen in common diseases.

This study has only been able to identify the collective abundance of variants within each pool in order to compare the total group of cases and controls. The hypothesis under test is based on the assumption that all the SNPs within these loci have similar effects in determining disease. Due to the limitations in study design and implementation, additional, in-depth inferences about the implications of SNPs identified cannot be drawn from these preliminary data. Cost and logistical implications meant that individual level sequencing could not be undertaken in the current study. Pooled sequencing is a more cost-effective approach to study rare variants in large populations by reducing the costs of both sample preparation and sequencing (Sham *et al.* 2002). However, as costs come down, individual level re-sequencing will become more affordable and may provide additional insight.

Allele frequency generation using pooled sequencing has been a commonly used and validated approach, particularly in GWA studies (Kwok *et al.* 1994; Taillon-Miller *et al.*). Careful quantification of of individual DNA samples and accurate combining was carried out to provide accurate allele calling. However, a limitiation to the pooling approach is the difficulty in ensuring absolute equimolar representation of samples. Therefore, due to uncertainties in the pooling accuracy and sequencing read quality, rare alleles are often difficult to detect in pooled sequencing and allele frequency estimates may be incorrect by around 5% (Marth *et al.* 2001). As a result, allele frequency data generated using this approach may not be accurate and so were not assessed for identified SNPs in the current study. Therefore, it has not been possible to study association of any individual SNPs to generate allele frequency data or determine effects in disease. Furthermore, in pooled samples, very rare variants such as a single heterozygote SNP in one individual may have an allele frequency which falls below sequencing error rates. As a result, some true variants may not have been detected.

Any single SNP may be present in individuals in isolation or in distinct combinations with other SNPs in different people. Sequencing studies are currently underpowered to assess SNPs

in rare variant association analyses using single-point tests; as a result, methods have been developed to combine information across multiple variant sites, described for example in a review by Asimit and Zeggini (Asimit and Zeggini 2010). However, these require SNP frequency data and therefore have not been used in analyses described in this Chapter. Several locus-based statistical methods exist including the collapsing method which compares the load of rare variants between cases and controls, as well as the allele-matching method which takes into account the direction of effect of alleles in the tested gene by comparing the number of shared alleles between cases and controls. Recently, a novel approach to test for the presence of mixed effects across a set of rare variants has also been proposed (Neale *et al.* 2011). This could allow detection of both risk and protective variants in the same genes or pathways in determining disease.

The findings presented here add weight to the hypothesis that rare variation explains some of the genetic heritability still to be identified in asthma. The functions of identified variants are still unclear and inferences on how they relate to common variants detected by GWA approaches in the sequenced regions cannot be made from results presented in this Chapter. Further work will need to be conducted to determine the location of casual polymorphisms, resulting functional effects, and the relationship between common and rare variants. As the gold standard of GWA studies has been validation through replication, this provides a further challenge for the future in selecting appropriate individuals and populations for follow-up studies (Chanock *et al.* 2007). Validation of identified variants through genotyping for example will be required as well as individual level data to be able to assess the frequency of detected variants. Given the nature of looking for infrequent variation, measures to minimise bias and confounding need to be taken into account by considering factors such as adequate phenotyping and population stratification (Schork *et al.* 2009). Furthermore, appropriate methods for statistical analyses are still in development for testing these associations (Hoggart *et al.* 2008; Li and Leal 2008; Schork *et al.* 2008).

Further advances to increase throughput and decrease costs in sequencing technologies are required in order to identify sufficient numbers of rare variants in large numbers of individuals

prior to testing associations with disease. These include developments in methods to improve target enrichment of selected regions, including enriching for DNA samples in multiplex pools at the same time (Johansson *et al.* 2011; Kenny *et al.* 2011; Singh *et al.* 2011). Current efforts from a number of large-scale sequencing projects may provide further insight into the impact of rare variants as well as provide usable reference data through assembling entire genomes of individuals at higher depth (1000 Genomes Project Consortium *et al.* 2010) (http://www.uk10k.org/). Furthermore, there may be additional low-frequency and rare polymorphisms in other regions identified across the genome such as the *IL13* and *TSLP* loci as well as the chromosome 17q12-21 locus which may be important in the development of asthma (see Chapters 5 and 6). This calls for designs such as whole-genome or at a slightly reduced cost, whole-exome sequencing.

Chapter 9

General Discussion

A summary of positive associations identified using primary GWA studies in atopy and severe asthma is given in Table 9.1. The strength of association in the original discovery cohort and subsequent replication is detailed.

Table 9.1 – Summary of positive associations from primary GWA studies. Results shown for studied SNPs identified in the discovery GWA study cohort and replicated SNPs in subsequent follow-up cohorts. + Risk minor allele. ¯ Protective minor allele.

Phenotype	**Chr**	**Gene**	**SNP**	**Discovery**		**Replication**	
				Cohort	**p value**	**Cohort**	**p value**
Specific IgE	8q12	Intergenic	rs7835920	British 1958 Birth Cohort	$5.54x10^{(-7)+}$		
	8q22	*ANKRD46*	rs7835153		$4.50x10^{(-7)-}$		
	10q26	Intergenic	rs11244450		$6.50x10^{(-6)+}$		
	11q24	*KIRREL3*	rs4936003		$4.20x10^{(-5)+}$	ALSPAC	$7.33x10^{(-4)+}$
	13q14	*FNDC3A*	rs6561505		$2.15x10^{(-9)+}$		
	18q22	*NETO1*	rs10514054		$2.54x10^{(-5)+}$		
	20p13	Intergenic	rs6082335		$5.00x10^{(-5)+}$		
Severe asthma	2q12.1	*IL18R1*	rs3771166	AUGOSA	$1.93x10^{(-5)-}$	AAGC	0.001 ¯
	5q31.1	*C5orf56*	rs11745587		$2.09x10^{(-6)+}$		
	6p23	*CD83*	rs9382936		$5.61x10^{(-6)+}$		
	7p21.1	*PRPS1L1*	rs12699949		$1.19x10^{(-5)-}$		
	13q31.1	Intergenic	rs2496764		$7.86x10^{(-6)+}$		
	17q12-21	*ERBB2*	rs1810132		$1.73x10^{(-5)+}$		

9.1 Genetic Determinants of Atopy

Genetic heritability to allergic disease has been long established (see Chapter 1: Section 1.3). However, it has been difficult to determine the contribution of individual genetic variants to disease. The first part of this thesis aimed to study atopy by using a genome-wide association (GWA) approach to identify novel atopy susceptibility loci and fine mapping of previously reported linkage signals through candidate region association.

Overall, results from the GWA scan in the British 1958 Birth Cohort suggest that there is evidence for a large number of loci influencing risk of atopy defined using elevated specific IgE and/or positive skin prick test (SPT), each conferring a modest effect. A single SNP on chromosome 13q14 met genome-wide significance and a further six loci were identified using SNPs meeting nominal significance on chromosomes 8q12, 8q22, 10q26, 11q24, 18q22, and 20p13. A combination of *in silico* replication, prospective genotyping, and combined analyses were used to replicate these findings.

Subsequently, a more focused approach was used to test association of SNPs within chromosomes 3p22.1-q22.1 and 17p12-q24.3. These loci were identified through a meta-analysis of genome-wide linkage studies (GSMA) as having significant linkage ($p<4.17x10^{(-4)}$) with atopy in asthma-enriched families (Denham *et al.* 2008). Fine mapping was carried out to try and localise the causative signal before attempting replication of a single SNP on chromosome 17q21.31 using the same methodology as for the GWA scan.

For the majority of assessed SNPs from the GWA scan, inconsistent replication was seen across the cohorts for both statistical significance and direction of effect of the minor allele. Although significant effect sizes were seen for some SNPs following combined analyses, they were most likely driven by the initial signals from the GWA scan. Results from replication of the SNP identified through fine mapping of reported linkage signals were inconsistent within tested populations in spite of findings from a second, recent GSMA also providing supporting evidence for chromosome 17q12-q24 (Bouzigon *et al.* 2010).

While it is not possible to confidently confirm any of the identified loci as being determinants of disease, some suggestive evidence for association can be concluded. For example in the GWA scan, although not all replication results were statistically significant, SNPs on chromosomes 8q12 and 20p13 showed a consistent effect for the minor allele across study cohorts. In addition, the SNP on chromosome 11q24 showed statistically significant replication of association with atopy across two independent cohorts using subjects of different ages and different definitions of atopy, both elevated specific IgE and positive SPT. It is possible that these loci contain susceptibility genes which influence risk of atopy. However, the function of genes implicated by significant SNPs identified (*ANKRD46*, *KIRREL3*, *FNDC3A*, and *NETO1* using GWA; *KIF18B* using fine mapping) are unknown and variable mRNA expression is seen across a range of cells and tissues involved in allergic disease including in the lung and skin (see Chapter 3: Section 3.4.8 and Chapter 4: Section 4.4.7).

Association with SNPs within previously reported candidate genes such as *FCER1A*, *IL13*, and *STAT6* has demonstrated the ability of the British 1958 Birth Cohort dataset used in being able to detect atopy susceptibility polymorphisms (Heinzmann *et al.* 2000). Despite some overlap in SNP association across markers of atopy analysed, genetic factors controlling allergen specific atopic responses via elevated specific IgE or positive SPT may be different to that of serum total IgE levels. With the exception of the *IL13* and *HLA* regions, distinct variants may determine these traits specifically (Moffatt *et al.* 2010). This highlights the potential existence of both pleitropic genes determining the generalised allergic response as well as those unique to sensitisation.

9.2 Genetic Determinants of Severe Asthma

The genetic architecture of asthma has been widely investigated with consistent replication of a number of loci and genes reported providing relatively strong evidence of their contributions to disease. These include *ORMDL3* identified through GWA as well as *IL4R* (16p12.1-p11.2), *TGFB* (19q13.1), and *ADAM33* (20p13) through positional cloning and candidate gene association (Ober and Hoffjan 2006; Moffatt *et al.* 2007). However, the majority of studies conducted to date have focused on mild-to-moderate disease. The second part of this thesis aimed to use GWA to establish if there are any novel variants determining severe asthma and to evaluate the contribution within severe forms of disease of previously identified genes for association with mild-to-moderate asthma.

Results from GWA testing did not identify any SNPs meeting genome-wide statistical significance. However, six loci with suggestive evidence for association were identified just below this threshold; previously known loci on chromosomes 2q12.1 and 17q12-21 as well as potentially novel loci on chromosomes 5q31, 6p23, 7p21 containing *C5orf56*, *CD83*, *PRPS1L1* repectively, and an intergenic region on 13q31. Identified SNPs of interest were tested in a smaller cohort showing convincing replication for the 2q12.1 and 17q12-21 regions and consistent directions of effects for all loci apart from 13q31. These findings mean that there are unlikely to be major effects driven by common polymorphisms within a single or a few gene(s) on the risk of developing severe asthma. There may be a small number of additional loci which may be specific to this phenotype but with overall relatively small contributions.

Association of susceptibility loci reported by the GABRIEL Consortium study for mild-to-moderate asthma were confirmed for severe asthmatics with SNPs within *IL18R1* (2q12.1), *SLC22A5* and *IL13* (5q31.1), *HLA-DQ* (6p21.32), *IL33* (9p24.1), *SMAD3* (15q22.33), and *GSDMB* and *GSDMA* (17q12-21). This reinforces the hypothesis that determinants of asthma are likely to come from the same genes regardless of severity of disease. Effect sizes for the alternative allele were greater within severe asthmatics compared to individuals with milder disease for the chromosomes 2q12.1 and 17q12-21 SNPs. This could be explained by the fact

that contribution of genetic effects driven by these polymorphisms may be greater in more severe disease.

A large number of widely replicated findings from candidate gene studies had not been evaluated directly within the GWA scan such as *TLR2* (4q32) (Qian *et al.* ; Eder *et al.* 2004; Bjornvold *et al.* 2009; Smit *et al.* 2009), *IL18* (11q22.2-q22.3) (Higa *et al.* 2003; Imboden *et al.* 2006; Lee *et al.* 2006; Lachheb *et al.* 2008; Zhu *et al.* 2008), and *IFNG* (12q14) (Nakao *et al.* 2001; Nagarkatti *et al.* 2002; Wang *et al.* 2006; Kumar and Ghosh 2008). Therefore, a comprehensive literature review was undertaken in order to identify all previously reported genes and loci in order to evaluate whether polymorphisms within these regions contributed to disease risk in severe asthma.

In this analysis, the extent of linkage disequilibrium between tested SNPs was taken into account and significant association was found for SNPs within 3 distinct loci implicating *IL1RL1/IL18R1* (2p12.1), *TSLP* (5q22.1), and *IL13* (5q31.1). The *IL1RL1* and *TSLP* genes remained significant even when all reported genes identified through GWA approaches for asthma were excluded. High test inflation statistics from assessing only loci identified through candidate gene associations (λ=1.19) were seen in spite of an overall low test inflation statistic from the initial severe asthma GWA study (λ=1.04). This provides strong evidence to suggest that there are contributions from multiple genes with more modest effects in determining disease in addition to those identified using GWA approaches.

9.3 Chromosomes 2q12 and 9p24 as Asthma Susceptibility Loci

Substantial evidence for polymorphisms within chromosomes 2q12 and 9p24 as contributing to asthma have been reported by both GWA and candidate gene association studies (Gabriel *et al.* 2002; Shimizu *et al.* 2005; Reijmerink *et al.* 2008; Castano *et al.* 2009; Gudbjartsson *et al.* 2009). Work carried out in this thesis for severe asthma has yielded strong support for the 2q12 locus and provided further suggestive evidence for 9p24. Identified genetic variants for association are common SNPs (MAF >5%) with unknown function. Therefore, these loci were further examined through molecular characterisation of the 5' gene regions of *IL1RL1* and *IL33* as well as next-generation re-sequencing of the region bearing the GWA signals seen.

The IL1RL1 receptor and its natural ligand, the IL33 cytokine have been recognised to exert important roles in the development of airway hyperresponsiveness in asthma. IL1RL1-IL33 signalling is thought to result in eosinophil-mediated inflammation via direct effects on eosinophils and indirect control of Th2 cytokines including IL4, IL5, and IL13 in inflamed tissues (Carriere *et al.* 2007; Cherry *et al.* 2008). Airway and immune cells and tissues likely to be important in asthma exhibit mRNA expression of both the *IL1RL1* and *IL33* genes, with increased levels of *IL1RL1* particularly in HMC-1 inflammatory cells and *IL33* preferentially in structural cells such as HASM and HBEC (see Chapter 7: Section 7.4.2).

Results from 5' RACE have identified multiple splice variants in both genes and a potentially novel open reading frame (ORF) in *IL1RL1* coding for a shorter form of the membrane-bound receptor. Multiple promoter binding sites for transcription factors with important roles in airway disease including AP-1, C/EBP, CREB, GR, CBP, and GATA were identified in the 2kb region upstream of the gene transcription start sites (TSSs) (Barnes and Adcock 1998).

The reported SNP implicating *IL1RL1* by the deCODE Icelandic GWA study is contained within an intronic region (Gudbjartsson *et al.* 2009). This SNP could be in LD with more 5' causative polymorphisms and therefore be driving the association signal through differential expression between *IL1RL1* splice variants, and by inference production of different levels of

the IL1RL1 receptor protein in cases compared to controls. Asthma associated SNPs in *IL33* are upstream of the established TSSs suggesting that these signals may be due to effects on transcription factor binding sites therefore determining levels of transcription varying between cases and controls.

These findings suggest that repeated airway injury in asthma resulting in structural changes may increase the production of IL33 cytokines and membrane-bound forms of the IL1RL1 receptor. During exacerbations and the induction of the inflammatory cascade, decoy soluble IL1RL1 receptors may be upregulated via the recruitment of inflammatory cells such as mast cells. This supports previously published functional work showing increased levels of expression of both the soluble form of the IL1RL1 receptor and IL33 in more severe asthma (Oshikawa *et al.* 2001; Préfontaine *et al.* 2009). The role of IL33 as an endogenous 'danger' signal or 'alarmin' released from dead or dying cells following injury due to trauma or infection has also been proposed (Moussion *et al.* 2008). IL33 either acting alone or with other mediators such as TSLP may be an important activator of initiator and effector cells responsible for innate immunity such as mast cells (Allakhverdi *et al.* 2007).

Results from re-sequencing of the chromosome 2q12 and 9p24.1 regions identified multiple variants detected by re-sequencing in homogenised pools of cases compared to controls. Both cases and controls were found to carry unique variants. Carrying severe asthma specific variants may have implications in determining disease susceptibility. The majority of these variants did not lie with the coding regions of *IL1RL1, IL18R1*, and *IL33* genes and were not common SNPs detected by GWA signals. These findings provide support for the common disease, rare variant hypothesis and warrant further work before the function of these variants and relation to previously identified SNPs can be determined.

9.4 Strengths and Weaknesses in Study Design

This thesis presents the first use of GWA approaches to identify genetic determinants of both the atopy *per se* and severe asthma phenotypes. Study cohorts used in analyses were the largest accessible at the given time. The AUGOSA cohort is the largest collection of individuals with severe asthma established to date; the only other existing cohort to study this phenotype is the TENOR study with relatively low numbers of recruited subjects (Li *et al.* 2010).

The total numbers of individuals in the initial GWA scans are relatively low for both atopy (1,083 cases, 2,070 controls) and severe asthma analyses (933 cases, 3,346 controls) compared to recently published GWA studies, for example on lung function (20,288 individuals) (Repapi *et al.* 2010) and mild-to-moderate asthma (10,365 cases, 16,110 controls) (Moffatt *et al.* 2010). However, attempts to increase study power were made by increasing the number of controls due to the difficulty encountered in recruiting defined cases. Severe asthma is an extreme phenotype therefore increasing statistical power due to expected augmentations in effect sizes of identified polymorphisms during case-control comparisons. Furthermore, combined analyses were carried out using multiple cohorts to increase the total n numbers for association testing following efforts to replicate identified signals. However, this method may not have been optimal in the absence of large effect sizes as pooled calculations when combining all individuals with potentially large amounts of heterogeneity may have prevented accurate signal identification.

Despite limiting the regions studied for fine mapping of linkage signals for atopy, the power to find linkage produced by a single or even a few gene(s) in specific chromosomal regions may have been low given that it is likely that a large number of genes are likely to be involved in the development of atopy, each conferring a modest effect. As explained in Chapter 4, the approach taken to localise the susceptibility loci was probably unsuccessful due to the problems in effect size replication, limitations in study power, and analysis of individual cohorts in stages. Additionally, fine mapping for individual genes may have been inappropriate

if the significant signals seen were due to several genes in combination in close proximity within the entire region.

The over-conservative nature of commonly used corrections for multiple testing to determine genome-wide significance such as the Bonferroni correction was taken into account through the use of nominal significance and the MeffLi correction. An inclusive approach was taken to identify significantly associated loci in the consideration that each SNP does not undergo a truly independent test due to linkage disequilibrium (LD), joint effects through epistasis, and strong evidence to suggest the existence of multiple polymorphisms with modest effects. Nevertheless, this thesis has highlighted the need for more appropriate methods to assess statistical significance, particularly when the *a priori* hypothesis is based on a large number of contributing polymorphisms to disease.

Difficulty in identifying robust signals with confidence due to limited consistency in replication in terms of both statistical significance and effect size may have resulted due to a number of different factors. Attempts at replicating initially significant signals in subsequent cohorts with much smaller sample sizes inevitably leads to a type of ascertainment bias known as 'winner's curse bias'. The overestimation of the genetic effect in the discovery cohort will require much larger follow-up studies in order to generate enough power to adequately replicate signals with unbiased effect sizes. This is likely to have occurred in both the atopy and severe asthma GWA scans, particularly given the modest effects of expected variants.

Confounding through varying degrees of phenotypic heterogeneity between studies can also be used to explain these difficulties encountered as demonstrated by the meta-analysis of serum total IgE levels in atopy analyses. This would have occurred through recruitment of individuals with differences such as in age, population demographics, subtle population stratification, and accuracy in case-control definition. In atopy analyses, high correlation of results between specific IgE, positive SPT, and serum total IgE levels were seen. Furthermore, good representation of an atopic state was expected from a core test panel of three allergens in the initial study i.e. house dust mite (Der p 1), mixed grass pollen, and cat fur (Perkin *et al.* 2006).

However, this is based on the assumption that the genetic factors underlying sensitivity to, for example, house dust mite is the same as that to cat fur. Therefore, it may not have been appropriate to group these allergens together in phenotype defintion. The use of different antigen panels and measures to define atopy between studies may have caused complications as the relationship between these specific traits is still uncertain and there is still some risk of some cases being classed as controls in the absence of a more extensive panel. In severe asthma analyses, some patients with milder forms of disease may have been misclassified as severe cases leading to an underestimation of effects. Furthermore, currently no uniform definition of asthma severity has yet been established, differences in methods used across studies such as the ATS, ERS, and GINA criteria are also likely to have contributed to the difficulty in replicating findings.

Variable SNP coverage across the genome in genotyping arrays in different study cohorts may have been another contributing factor, particularly when fine mapping signals. Imputation is generally accepted as a good surrogate as shown by both imputation against a selected group genotyped SNPs in the ALSPAC cohort and against genome-wide data generated by the 1000 Genomes Project. Although study cohorts were genotyped reasonably densely using the highest coverage platforms at the time, this is still limiting in many regions, exemplified by mapping of genotyped SNPs in AUGOSA surrounding the *IL1RL1* and *IL33* genes. This has made it difficult to determine the location of the causative polymorphisms and is currently inadequate in identifying rare variants. In the absence of denser genotyping, imputation has helped to inform interpretation of association signal strength, maximise replication data, and guide pinpointing of causative variants through detection of new peaks in identified regions. However, this statistical approach is driven by LD with pre-existing signals and is thus automatically biased to providing further supporting evidence for these regions.

9.5 Future Work

Hundreds of GWA studies carried out in common diseases have led to the identification of many statistically compelling and biologically plausible variations (Manolio *et al*. 2008). However in general, greater than 90% of the expected heritable component of a disease remains unexplained after extensive GWAS interrogation (Maher 2008). Current methodologies to reliably identify causative genetic factors contributing to disease require greater refinement in both study power and design. Atopy and asthma are both collections of symptoms with varying levels of inflammation, reversibility, exacerbations, severity, and degrees of treatment response. Overall, there is a large amount of evidence to suggest that genetic variation determining the loosely defined phenotypes of atopy and severe asthma are likely to come from multiple polymorphisms with modest effects. These polymorphisms are probable to be a combination of common, intermediate, and rare SNPs as well as other forms of structural variation such as indels and CNVs. Effects on disease are also expected to be exerted through multiple interacting gene pathways with a number of environmental influences that may differ between populations and epigenetic influences.

Improvements in individual recruitment, phenotyping, and genotyping will help to reduce confounding through heterogeneity across studies. Large-scale, international collaborations will drive efforts to generate cohorts with sufficient statistical power. Advances in genotyping technology will help to increase SNP coverage and address the role of rare and intermediate variants both in focused regions as well as genome-wide through whole exome sequencing. More detailed fine mapping techniques will then be needed to localise identified loci and confirm causative variants. Further work to elucidate the function of variants found will also be required to confirm their roles in disease pathogenesis. These could include functional genomic approaches such as targeting transcription and mapping expression by using microarrays for specific genes or expression quantitative trait loci (eQTL) mapping studies, for example. This will help to distinguish biological pathways whilst testing for interactions in gene pathways and with other environmental risk factors predisposing to disease using new analytical methods

such as genome-wide interaction studies (GWIS) (Murcray *et al.* 2009; Manning *et al.* 2011; Ober and Vercelli 2011).

The low predictive power of risk loci identified by GWAS, with a typical genotype relative risk of only 1.1 to 1.5 has been somewhat disappointing for the genetic community. However, what has been identified as a problem in estimating heritability using these variants may simply be more a reflection of the difficulty in being able to fully describe the architecture of phenotypic variation. Heritability does not equate to genetic contributions but rather the ratio of the genetic to the total phenotypic variance in a population (Visscher *et al.* 2008). Findings from current studies may reflect the complex interactions between different genetic and environmental factors (Manolio *et al.* 2008). These factors will be different in individual patients and hence be determined by different patterns of genetic contributions. Identification of specific variants may not be effective in predicting individual susceptibility risks but may substantially explain the burden of disease in the general population (Moffatt *et al.* 2010).

Furthermore, genetic approaches to these complex diseases has provided the opportunity to identify previously unrecognised biological pathways in disease pathophysiology (Hirschhorn 2009). GWA studies indicated that asthma and chronic obstructive pulmonary disease (COPD) may share common genetic factors influencing lung development (Postma *et al.* 2011). In other diseases, examples include the complement pathway in patients with age-related macular degeneration and the *IL23* associated pathways in Crohn's disease (Klein *et al.* 2005; Lettre and Rioux 2008). The value of this research is to question our understanding of disease mechanisms in parallel with identifying pathways important in causing certain symptoms and disease outcomes in order to lead us to better targeted therapies in the future. Combining information on genetic variations from individuals with clinical factors may increase the sensitivity and specificitiy of predictions in disease aetiology, in particular in identifying sub-phenotypes and response to treatment (Wu *et al.* 2010). Further incorporation of other influences such as environmental interactions, for example, with levels of allergen exposure could also help to provide further opportunities for preventative medicine.

Appendices

Appendix 1 URLs for Online Resources

NCBI (The National Centre for Biotechnology Information) public database

http://www.ncbi.nlm.nih.gov/

dbSNP: SNP database provided by NCBI

http://www.ncbi.nlm.nih.gov/projects/SNP

Nucleotide: sequence database provided by NCBI

http://www.ncbi.nlm.nih.gov/nuccore

Entrez Gene: gene database provided by NCBI

http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene

BLAST sequence alignment tool provided by NCBI

http://blast.ncbi.nlm.nih.gov/BLAST.cgi

The International HapMap Project provided by NCBI

http://hapmap.ncbi.nlm.nih.gov/

UCSC Genome Browser

http://genome.ucsc.edu/

Ensembl Genome Browser public database

http://www.ensembl.org/index/html

The 1000 Genomes Project Browser

http://browser.1000genomes.org/index.html

SNAP pairwise LD program (Broad Institute)

http://www.broadinstitute.org/mpg/snap/ldsearch.php

EMBOSS Transeq Nucleotide to protein sequence converter (EMBL-EBI)

http://www.ebi.ac.uk/Tools/emboss/transeq/

UniProtKB/Swiss-Prot Protein database

http://www.uniprot.org/

TOPO2 Transmembrane protein display program (UCSF)

http://www.sacs.ucsf.edu/TOPO-run/wtopo.pl

PCR Primer Stats sequence analysis tool

http://www.ualberta.ca/~stothard/javascript/pcr_primer_stats.html

WWW Signal Scan TFD, WWW Signal Scan TRANSFAC

http://www-bimas.cit.nih.gov/molbio/signal/

Transcription Element Search System (TESS)

http://www.cbil.upenn.edu/cgi-bin/tess/tess

TFSEARCH

http://www.cbrc.jp/research/db/TFSEARCH.html

Appendix 2 PCR Primers and Conditions

Standard PCR							
Target gene	**No. of bases**	**Primer name**	**Primer sequence**	**Position**	**Product length**	**Annealing temp.**	**No. of cycles**
ANKRD46	20	ANKRD46-EP-F753	5'-GACCATGCAAACAGCTGAGA-3'	Exon 5	302bp	59.7˚C	40
	20	ANKRD46-EP-R1054	5'-AACAAACATTGGAAGCCAGG-3'	Exon 6			
KIRREL3	20	KIRREL3-EP-F200	5'-GAGGCACAAAGAGACCGAAG-3'	Exon 1	300bp	59.7˚C	40
	20	KIRREL3-EP-R499	5'-TTGGCCTTCATTCATTCTCC-3'	Exon 2/Exon 3			
FNDC3A	20	FNDC3A-EP-F786	5'-GGAGATGTAGATGCTCACTC-3'	Exon 5/Exon 6	563bp	61.6˚C	40
	21	FNDC3A-EP-R1349	5'-CTGGACTTTTGCATGGTAATC-3'	Exon 9/Exon 10			
NETO1	20	NETO1-EP-F94	5'-CATTCAGAGATGCTGGCAGA-3'	Exon 1	366bp	60.0˚C	40
	21	NETO1-EP-R459	5'-TTGTCCACAGAAACGTCCAA-3'	Exon 4			
NETO1	20	NETO1-EP-F121	5'-CCTAAGGTTGTAGCAAGTTT-3'	Exon 1/Exon 2	203bp	61.0˚C	40
	19	NETO1-EP-R323	5'-CGGCTTCTATGATGTAGAT-3'	Exon 3/Exon 4			
KIF18B	20	KIF18B- EP-F1163	5'-GCTGAAGAGCAATGTGACCA-3'	Exon 7/Exon 8	331bp	59.7˚C	40
	20	KIF18B-EP-R1493	5'-GGACTGCTCCTGGTCTGAAG-3'	Exon 10			
GAPDH	24	GAPDH-control-F152	5'CCACCCATGGCAAATTCCATGGCA-3'	Exon 2	600bp	50.0˚C	40
	24	GAPDH-control-R749	5'-TCTAGACGGCAGGTCAGGTCCACC-3'	Exon 6			

Taqman® PCR						
Target gene	**Forward primer**	**Reverse primer**	**Probe**	**Product length**	**Exons targeted**	**Cat. No./scale**
IL1RL1	5'-CAAACAAAAGTATTCCCACTCAGG-3'	5'-ACATTCGCATATCCAGTCCTATTG-3'	5'-TCGTGTGTTTGCCTCAGGCCAACT-3'	148bp	3/4 at 492bp to 376bp	PB-DD101-004 (40nmol)
IL33	5'-TAGGAGAGAAACCACCAAAAGG-3'	5'-ATATACCAAAGGCAAAGCACTC-3'	5'-CATCTGGTACTCGCTGCCTGTCAACA-3'	117bp	2/3 at 197bp to 313bp	PB-DD101-004 (40nmol)

5' RACE PCR					
Target gene	**No. of bases**	**Primer name**	**Primer sequence**	**Position**	**PCR**
IL1RL1	27	IL1RL1-5'RACEa-R531	5'-CAGGTATAAATACCAGAATCAGCAACT-3'	Exon 3	Primary
	26	ST2-5'RACEb-R481	5'- TGGCCTGAGGCAAACACACGATTTCT-3'	Exon 3	Nested
IL33	25	IL33-5'RACEa-R295	5'-CACTCCACAGTAGACTGCTGTTGAC-3'	Exon 4	Primary
	26	IL33-5'RACEb-R269	5'-GGCAGCGAGTACCAGATGTCTTTTGT-3'	Exon 4	Nested

Appendix 3 Association Results for Studied Atopy SNPs Identified through GWA

Association results between studied SNPs detailing per-allele p values and effect sizes measured using OR for dichotomised traits: elevated specific IgE and positive SPT; Beta Trend coefficient for continuous trait: serum total IgE. Data for serum total IgE corrected for sex, sample collection differences at age 44-45yrs in B58C; uncorrected in ALSPAC, Nottingham Gedling, and Southampton. Effect sizes presented for the minor (2nd) allele.

(i) Chr 8 (59,471,286 bp)

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs7835920 'A/C'	Elevated specific IgE	B58C	1,083	2,770	1.33 (1.19-1.49)	**$5.54x10^{(-7)}$**
		Southampton	292	376	1.14 (0.89-1.46)	0.302
	Positive SPT	ALSPAC	281	1,042	1.10 (0.90-1.36)	0.349
		Gedling	385	822	1.21 (0.99-1.47)	0.060
		Southampton	377	291	1.09 (0.85-1.40)	0.495
	Serum total IgE	B58C	3,873		1.13 (1.06-1.22)	**$5.79x10^{(-4)}$**
		ALSPAC	1,423		1.02 (0.95-1.11)	0.552
		Gedling	1,117		1.13 (0.97-1.32)	0.129
		Southampton	663		1.38 (0.92-2.07)	0.119
	Combined	Specific IgE	1,375	3,146	1.29 (1.17-1.43)	**$8.20x10^{(-7)}$**
		Positive SPT	1,043	2,155	1.10 (0.98-1.24)	0.114
		Total IgE	1,780		1.19 (0.90-1.57)	0.236

(ii) Chr 8 (101,642,996 bp) – *ANKRD46*

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs7835153 'A/G'	Elevated specific IgE	B58C	1,082	2,759	0.77 (0.70-0.85)	**$4.50x10^{(-7)}$**
		Southampton	290	380	1.17 (0.94-1.45)	0.171
	Positive SPT	ALSPAC	282	1,036	0.86 (0.72-1.05)	0.132
		Gedling	390	828	0.98 (0.83-1.16)	0.819
		Southampton	377	293	1.17 (0.94-1.45)	0.172
	Serum total IgE	B58C	3,873		0.88 (0.82-0.93)	**$3.47x10^{(-5)}$**
		ALSPAC	1,423		1.03 (0.96-1.10)	0.451
		Gedling	1,129		0.98 (0.86-1.13)	0.823
		Southampton	665		1.18 (0.82-1.69)	0.376
	Combined	Specific IgE	1,372	3,139	0.83 (0.76-0.91)	**$7.00x10^{(-5)}$**
		Positive SPT	1,049	2,157	0.99 (0.89-1.10)	0.817
		Total IgE	1,794		1.05 (0.82-1.34)	0.718

(iii) Chr 10 (126,026,768 bp)

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs11244450 'G/A'	Elevated specific IgE	B58C	1,079	2,763	1.46 (1.24-1.72)	**$6.50x10^{(-6)}$**
		Southampton	291	374	0.83 (0.57-1.21)	0.332
	Positive SPT	ALSPAC	242	1,042	0.67 (0.45-1.00)	**0.045**
		Gedling	382	826	0.95 (0.71-1.27)	0.733
		Southampton	376	289	0.75 (0.52-1.08)	0.120
	Serum total IgE	B58C	3,873		1.23 (1.10-1.37)	**$2.14x10^{(-4)}$**
		ALSPAC	1,423		1.01 (0.89-1.15)	0.889
		Gedling	1,120		0.84 (0.67-1.05)	0.131
		Southampton	660		0.95 (0.51-1.77)	0.872
	Combined	Specific IgE	1,370	3,137	1.33 (1.14-1.55)	**$2.08x10^{(-4)}$**
		Positive SPT	1,040	2,157	0.87 (0.72-1.05)	0.142
		Total IgE	1,780		0.86 (0.57-1.31)	0.486

(iv) Chr 11 (126,258,841 bp) – *KIRREL3*

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs4936003 'A/G'	Elevated specific IgE	B58C	1,073	2,754	1.27 (1.13-1.43)	**$4.20x10^{(-5)}$**
		Southampton	295	381	0.89 (0.69-1.13)	0.332
	Positive SPT	ALSPAC	282	1,042	1.43 (1.16-1.76)	**$7.33x10^{(-4)}$**
		Gedling	386	826	0.90 (0.74-1.11)	0.321
		Southampton	382	294	1.05 (0.82-1.34)	0.685
	Serum total IgE	B58C	3,873		1.08 (1.00-1.16)	0.054
		ALSPAC	1,423		1.03 (0.95-1.12)	0.477
		Gedling	1,125		0.99 (0.84-1.16)	0.864
		Southampton	671		0.91 (0.61-1.35)	0.628
	Combined	Specific IgE	1,368	3,135	1.20 (1.09-1.34)	**$4.36x10^{(-4)}$**
		Positive SPT	1,050	2,162	1.13 (1.00-1.27)	0.052
		Total IgE	1,769		1.19 (0.90-1.58)	0.229

(v) Chr 13 (48,651,730 bp) – *FNDC3A*

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs6561505 'G/A'	Elevated specific IgE	B58C	1,081	2,768	1.67 (1.41-1.98)	**$2.15x10^{(-9)}$**
		Southampton	288	370	0.85 (0.55-1.30)	0.442
	Positive SPT	ALSPAC	242	1,042	1.33 (0.97-1.83)	0.078
		Gedling	394	825	1.24 (0.92-1.67)	0.158
		Southampton	371	286	0.79 (0.52-1.20)	0.273
	Serum total IgE	B58C	3,873		1.26 (1.12-1.41)	**$7.75x10^{(-5)}$**
		ALSPAC	1,423		0.99 (0.88-1.12)	0.870
		Gedling	1,167		1.03 (0.80-1.32)	0.835
		Southampton	675		0.94 (0.47-1.90)	0.863
	Combined	Specific IgE	1,369	3,138	1.49 (1.28-1.75)	**$5.45x10^{(-7)}$**
		Positive SPT	1,007	2.,153	1.10 (0.91-1.33)	0.305
		Total IgE	1,842		0.81 (0.52-1.26)	0.344

(vi) Chr 18 (68,532,581 bp)

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs10514054 'C/A'	Elevated specific IgE	B58C	1,082	2,770	1.37 (1.18-1.59)	**$2.54x10^{(-5)}$**
		Southampton	290	377	0.79 (0.56-1.11)	0.169
	Positive SPT	ALSPAC	282	1,041	1.29 (0.98-1.69)	0.069
		Gedling	385	831	0.89 (0.68-1.16)	0.381
		Southampton	377	290	0.91 (0.65-1.28)	0.599
	Serum total IgE	B58C	3,873		1.14 (1.04-1.26)	**$6.52x10^{(-3)}$**
		ALSPAC	1,423		1.09 (0.98-1.21)	0.128
		Gedling	1,126		1.16 (0.93-1.44)	0.190
		Southampton	662		0.53 (0.30-0.94)	**0.030**
	Combined	Specific IgE	1,372	3,147	1.25 (1.09-1.43)	**$1.32x10^{(-3)}$**
		Positive SPT	1,044	2,162	1.01 (0.86-1.19)	0.901
		Total IgE	1,782		0.87 (0.59-1.29)	0.484

(vii) Chr 20 (2,114,427 bp)

SNP	Phenotype	Cohort	N		Effect (95% CI)	p value
			Case	Control		
rs6082335 'G/A'	Elevated specific IgE	B58C	1,076	2,741	1.24 (1.12-1.37)	**$5.00x10^{(-5)}$**
		Southampton	290	376	1.22 (0.96-1.51)	0.104
	Positive SPT	ALSPAC	242	1,042	1.03 (0.85-1.24)	0.795
		Gedling	385	826	1.00 (0.84-1.20)	0.973
		Southampton	375	291	1.12 (0.90-1.41)	0.309
	Serum total IgE	B58C	3,873		1.07 (1.00-1.14)	**0.044**
		ALSPAC	1,423		1.00 (0.93-1.08)	0.961
		Gedling	1,123		1.07 (0.93-1.23)	0.344
		Southampton	662		1.25 (1.16-1.67)	0.434
	Combined	Specific IgE	1,366	3,117	1.22 (1.11-1.34)	**$2.78x10^{(-5)}$**
		Positive SPT	1,003	2,159	1.02 (0.92-1.14)	0.714
		Total IgE	1,785		0.97 (0.76-1.25)	0.828

Appendix 4 LD Structure on Chromosomes 3p22.1-q22 and 17p12-q24.3

The following LD plots show the positions of SNPs meeting statistical significance ($p \leq 0.001$) for association with atopy in the British 1958 Birth Cohort. This analysis tested association of SNPs genotyped on both the Illumina HumanHap550 and the Affymetrix Human Mapping 500K SNP platforms in a subset of 1,468 individuals with elevated specific IgE $\geq 0.35 kU_A/L$. This informed selection of regions for fine mapping prior to identifying SNPs to take forward for *in silico* replication and prospective genotyping.

SNPs identified on chromosome 3p22.1-q22 are shown in Figures A4.1 and A4.2. SNPs identified on chromosome 17p12-q24.3 are shown in Figures A4.3 and A4.4.

SNPs genotyped on each platform are colour coded as below:

* Ilumina HumanHap550
* Affymetrix Human Mapping 500K
* Both SNP platforms

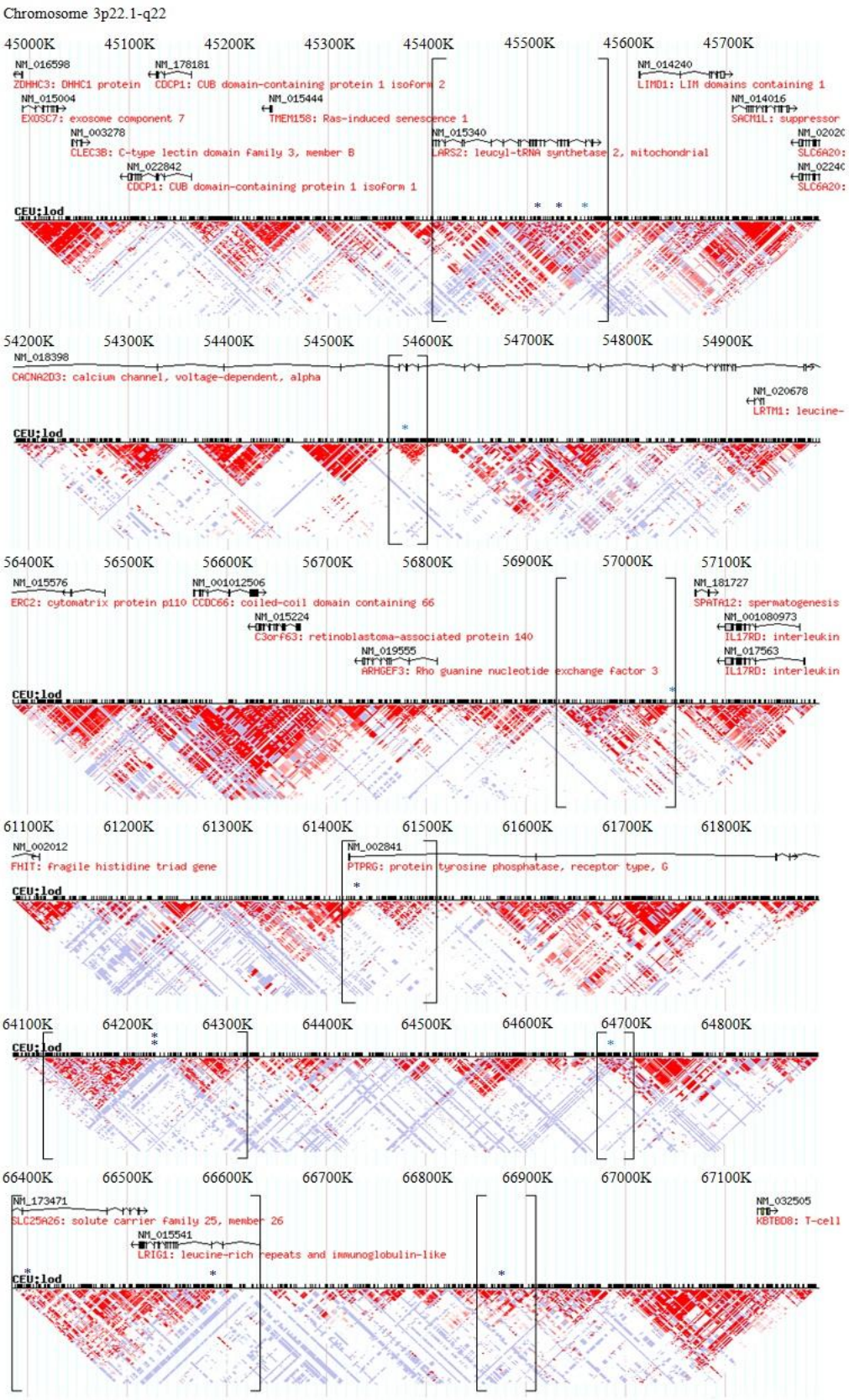


Figure A4.1 - SNPs identified on chromosome 3p22.1-q22 meeting statistical significance (p≤0.001) for association with atopy in a subset of the British 1958 Birth Cohort (45,000kb to 67,200kb).

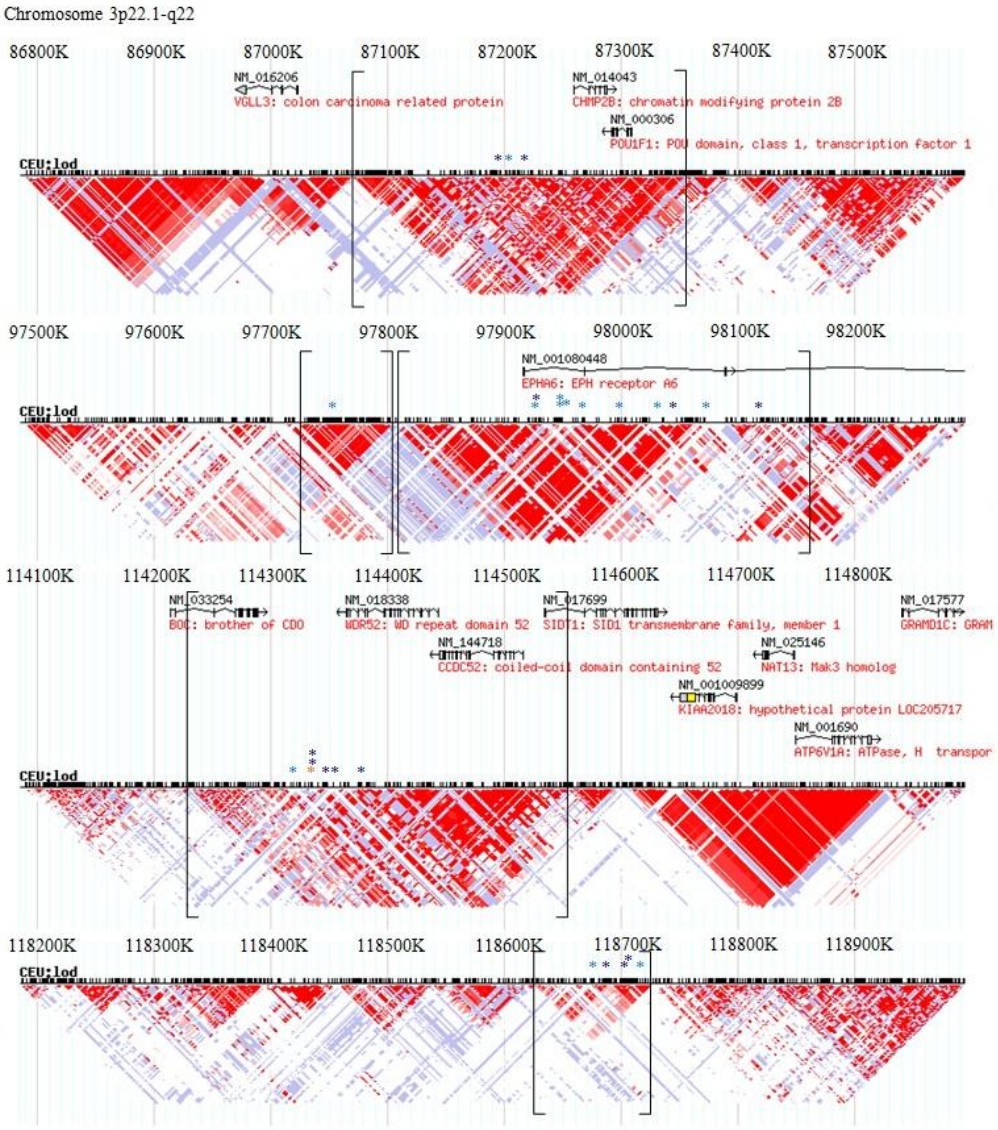


Figure A4.2 - SNPs identified on chromosome 3p22.1-q22 meeting statistical significance (p≤0.001) for association with atopy in a subset of the British 1958 Birth Cohort (86,800kb to 119,000kb).

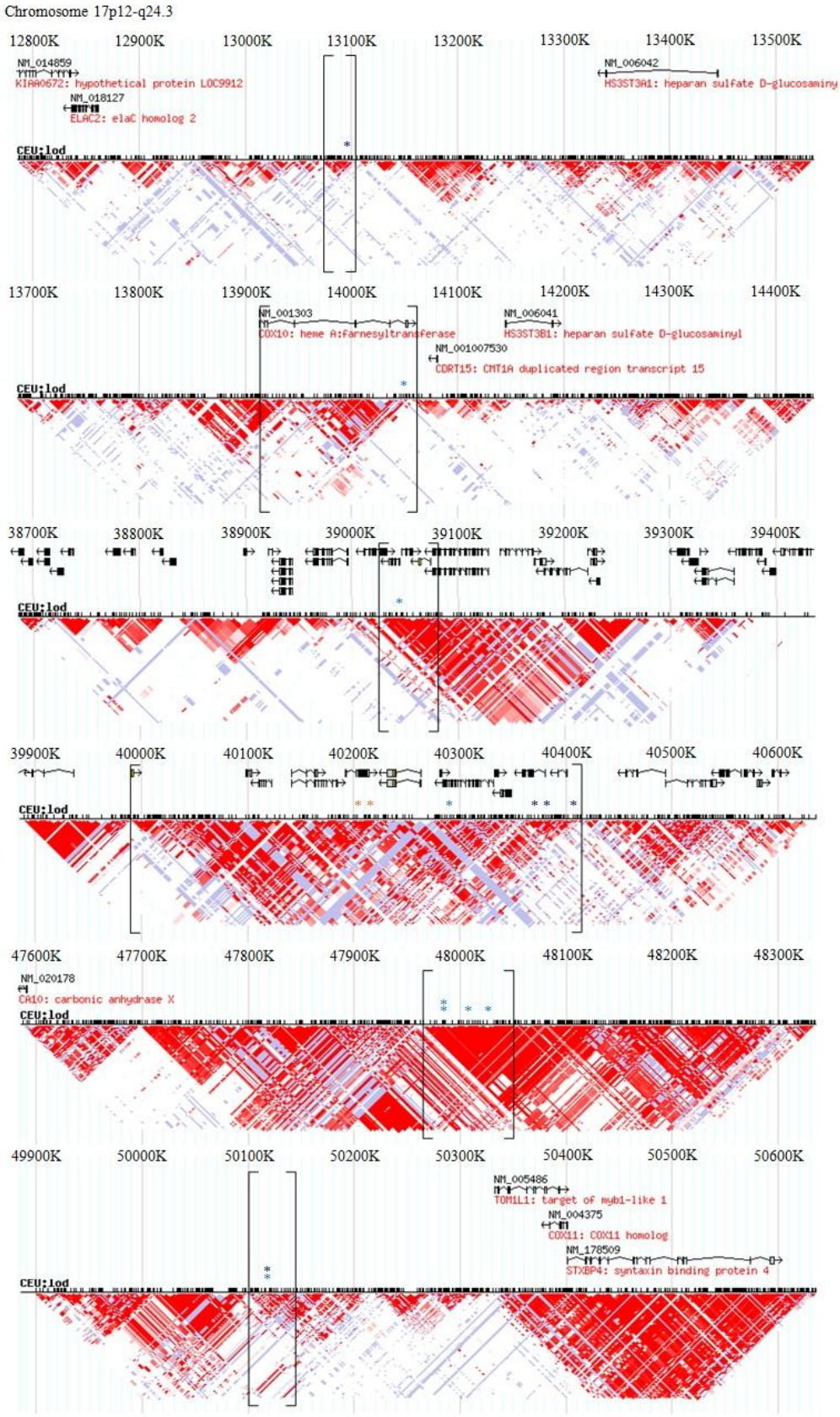


Figure A4.3 - SNPs identified on chromosome 17p12-q24.3 meeting statistical significance (p≤0.001) for association with atopy in a subset of the British 1958 Birth Cohort (12,800kb to 50,600kb).

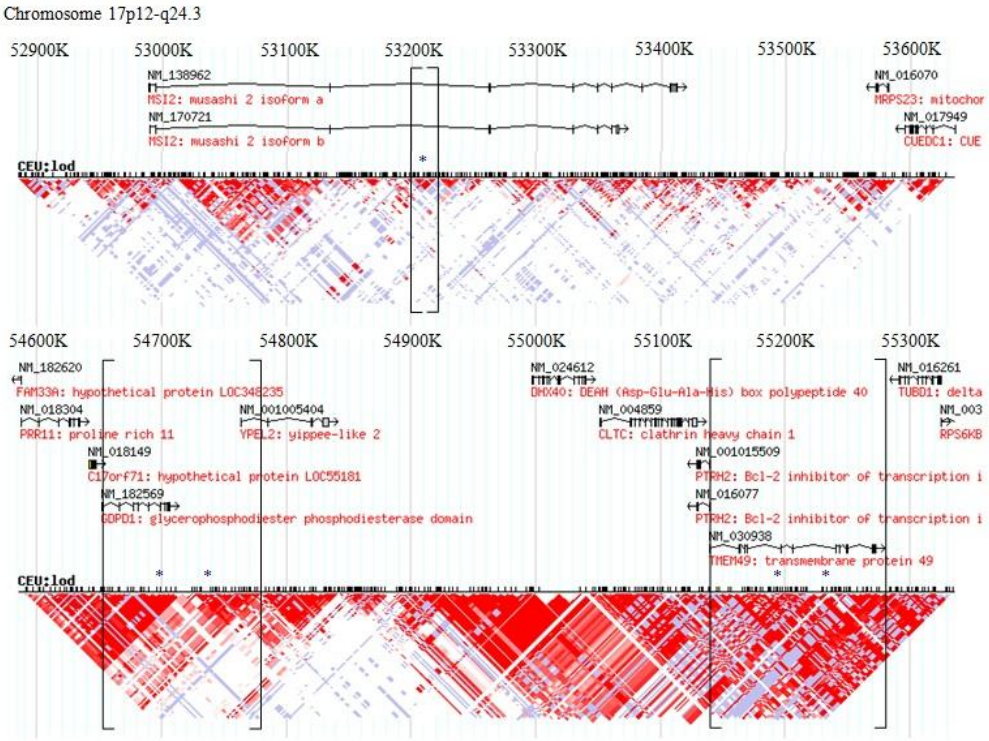


Figure A4.4 - SNPs identified on chromosome 17p12-q24.3 meeting statistical significance (p≤0.001) for association with atopy in a subset of the British 1958 Birth Cohort (52,900kb to 554,400kb).

Appendix 5 Region Plots for Eliminated Loci following Genotype Analyses and New Regions Identified through Imputation Analyses

The following region plots show the statistical significance of each SNP on the $-\log_{10}$ scale as a function of chromosome position (NCBI build 36). The sentinel SNP is shown in blue and in Figure 5.1A also flagged by the arrow. The correlation (r^2) of each of the surrounding SNPs to the sentinel SNP is shown by their colour (see key). Fine scale recombination rate is plotted in blue.

Two SNPs meeting statistical significance following GWA testing of genotyped SNPs: rs9395865 on 6p12.1 ($p=5.27x10^{(-6)}$) and rs10500656 on 11p15.4 ($p=1.87x10^{(-5)}$) were found to have low supporting evidence in the region. Based on the region plots shown in Figure A5.1, these loci were not taken forward for replication analyses.

Results from imputation identified 12 SNPs meeting statistical significance in new regions with reasonable supporting evidence ($p<10^{(-5)}$, $r^2_{imp}>0.7$). Two of these SNPs, rs4794820 and rs9897185 produced secondary peaks in the previously identified 17q12-21 region. The region plot for rs9897185 is shown in Chapter 5: Figure 5.6. Region plots showing supporting evidence from surrounding SNPs for the remaining 11 new regions are shown in Figures A5.2 to A5.7.

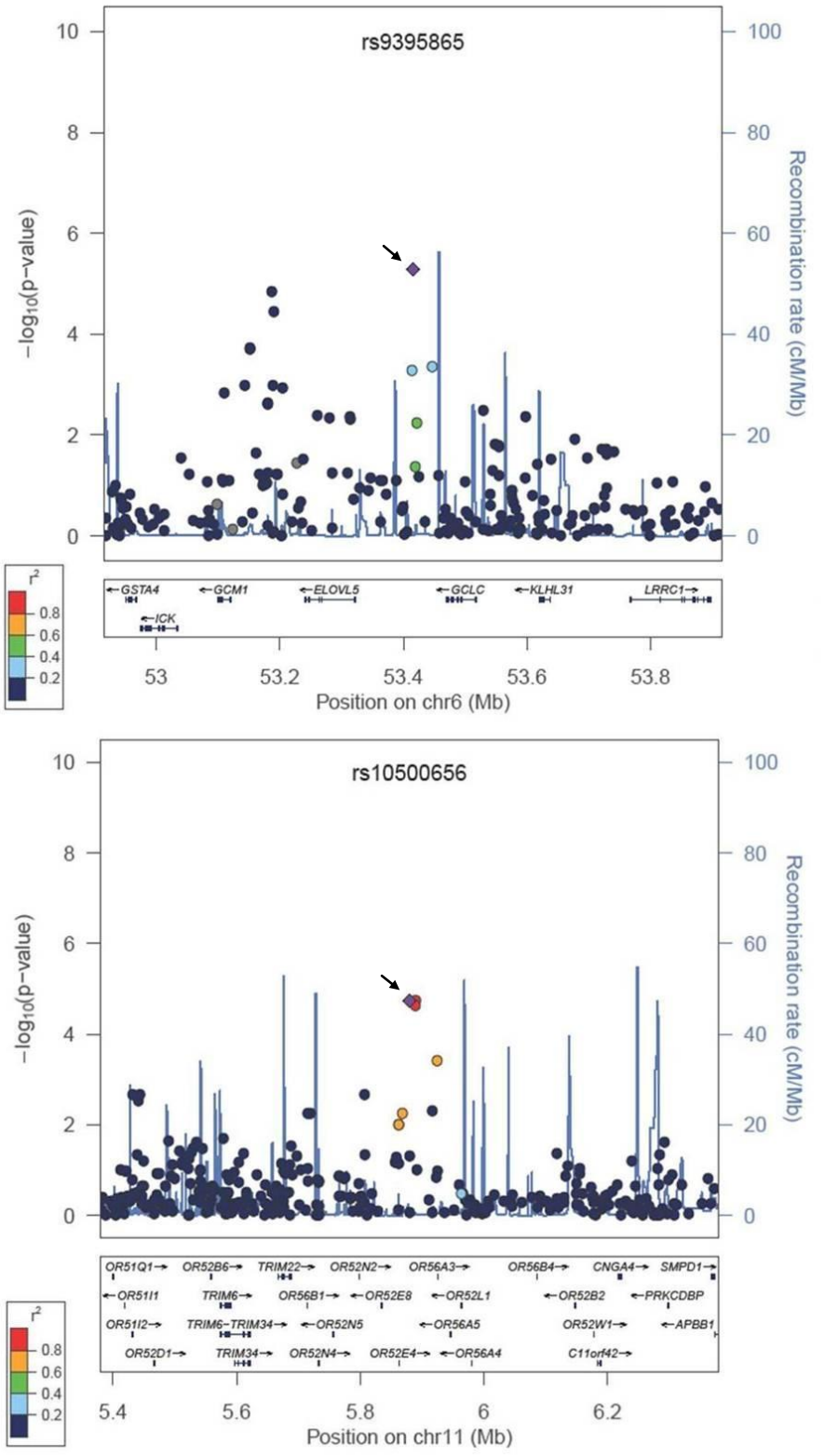


Figure A5.1 - Region plots for rs9395865 and rs10500656 with low supporting evidence from surrounding SNPs.

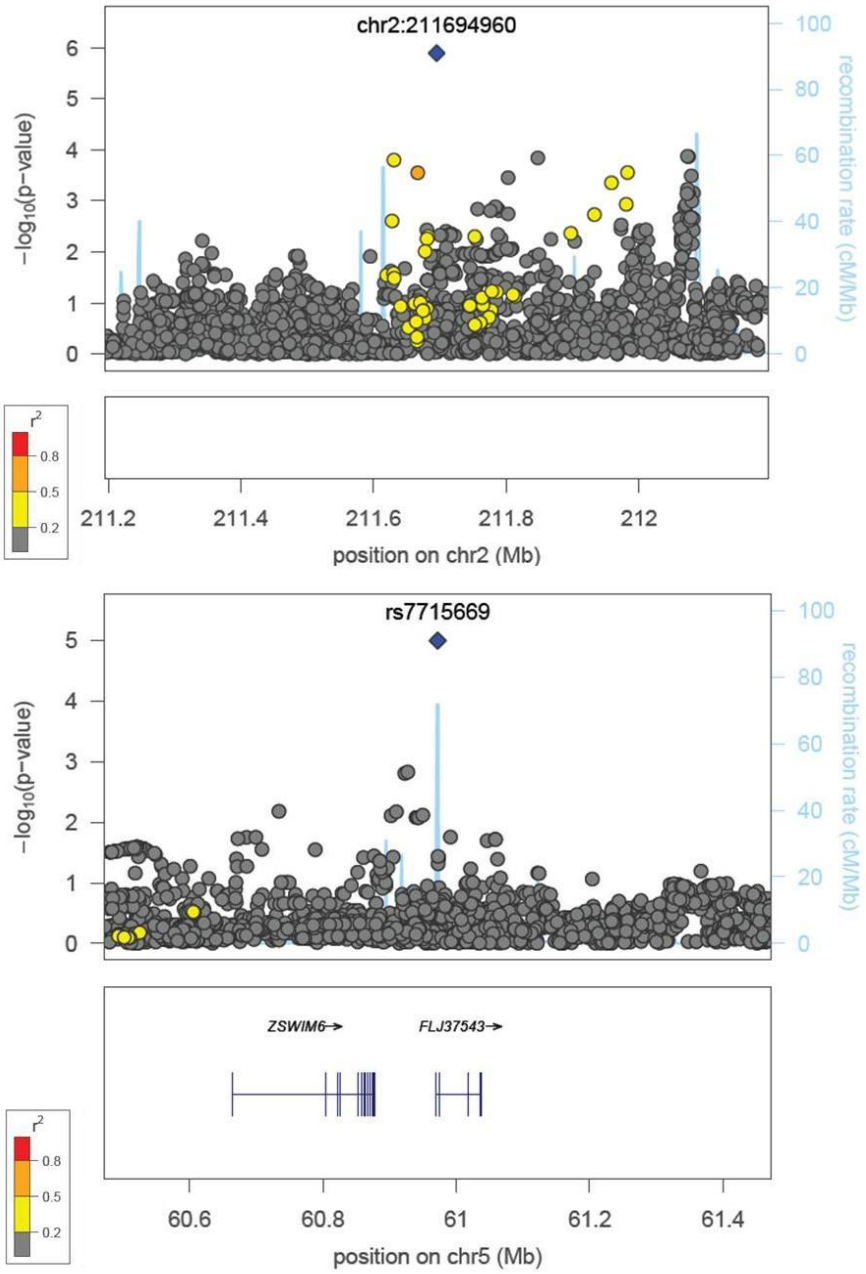


Figure A5.2 - Region plots for chr2:211694960 and rs7715669 in new regions identified by imputation with $p<10^{(-5)}$ and $r^2_{imp}=0.7$. Note chr2:211694960 is intergenic.

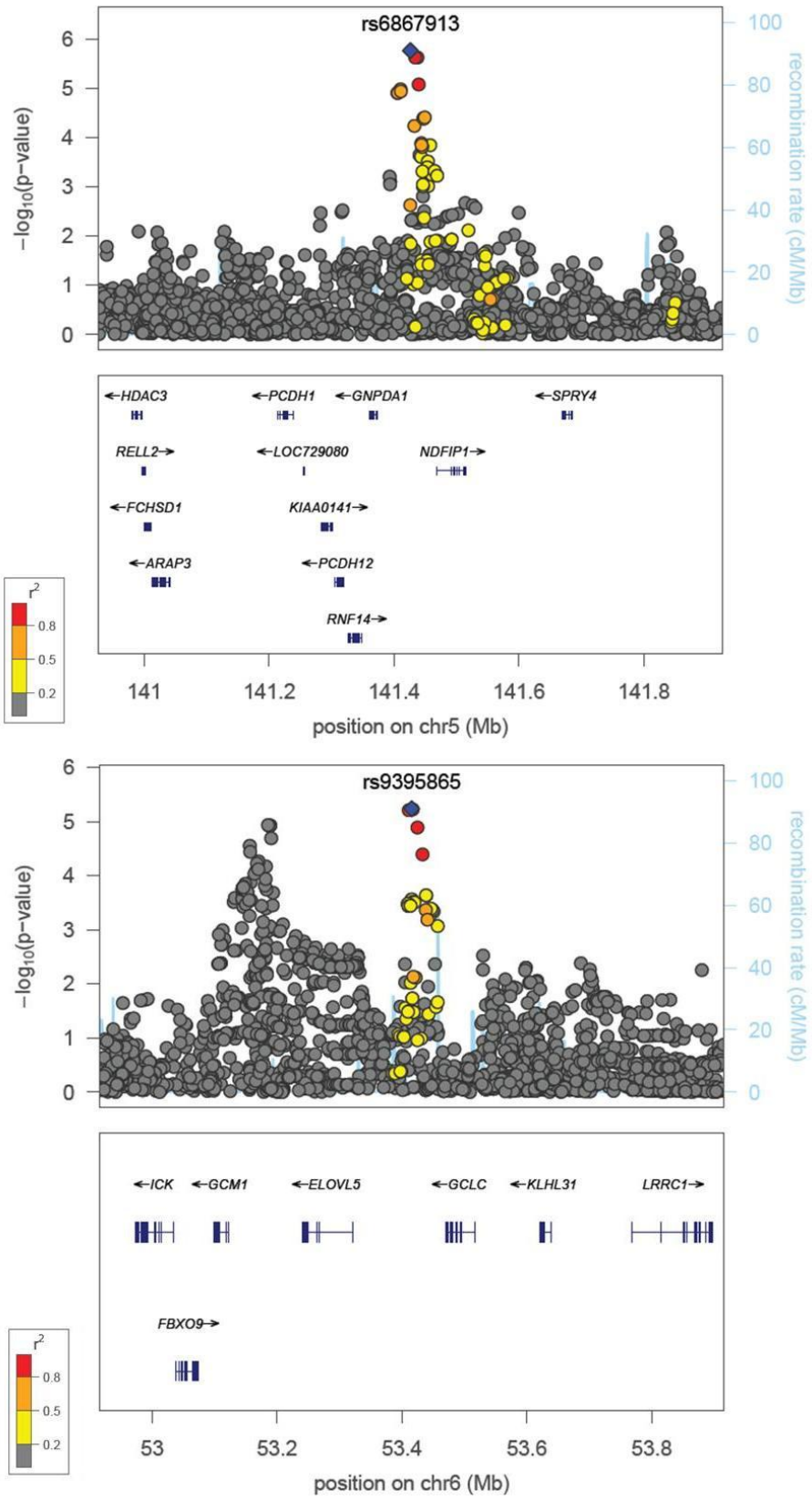


Figure A5.3 - Region plots for rs6867913 and rs9395865 in new regions identified by imputation with $p<10^{(-5)}$ and $r^2_{imp}=0.7$.

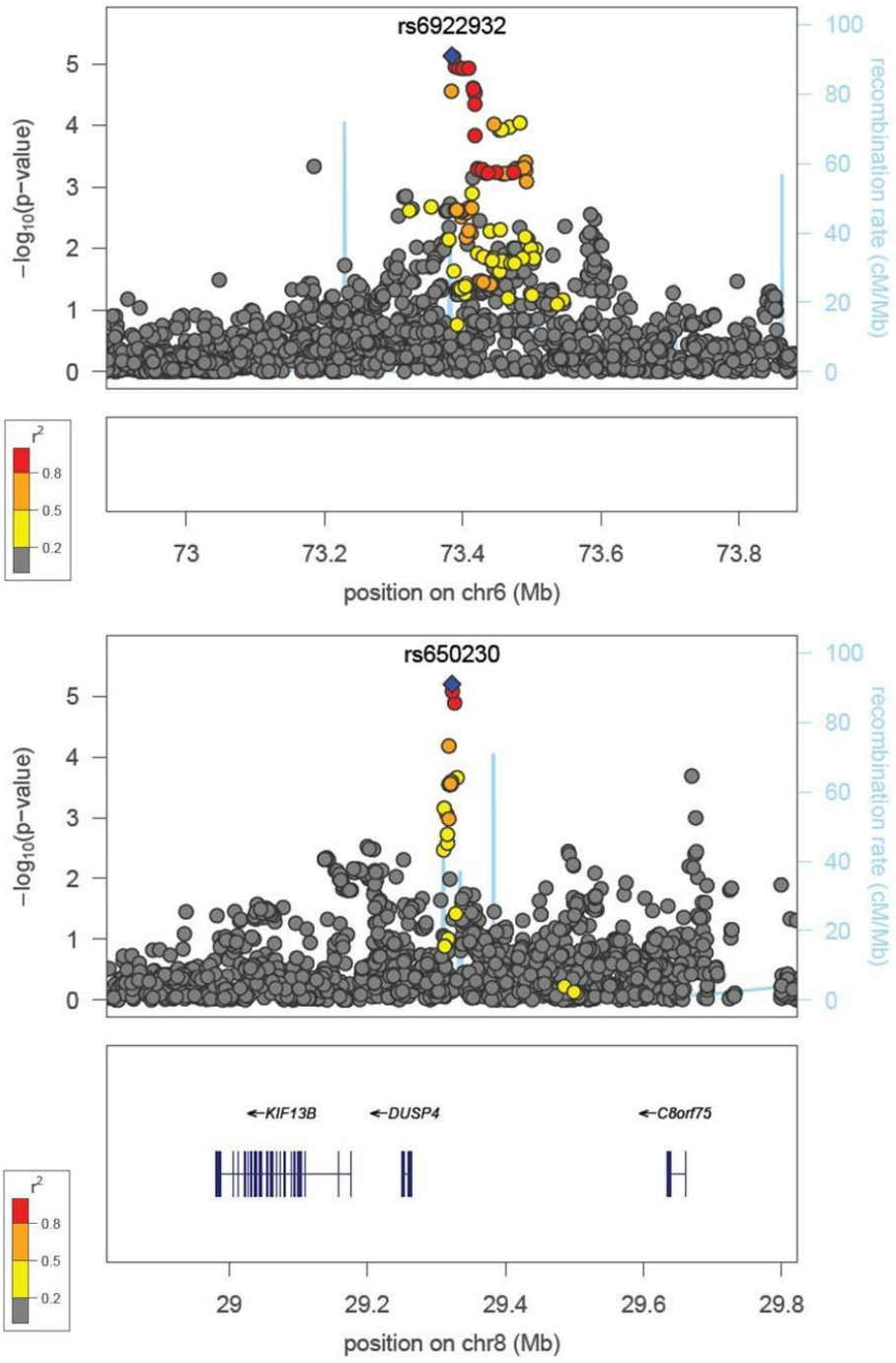


Figure A5.4 - Region plots for rs6922932 andrs650230 in new regions identified by imputation with $p<10^{(-5)}$ and $r^2_{imp}=0.7$. Note rs6922932 is intergenic.

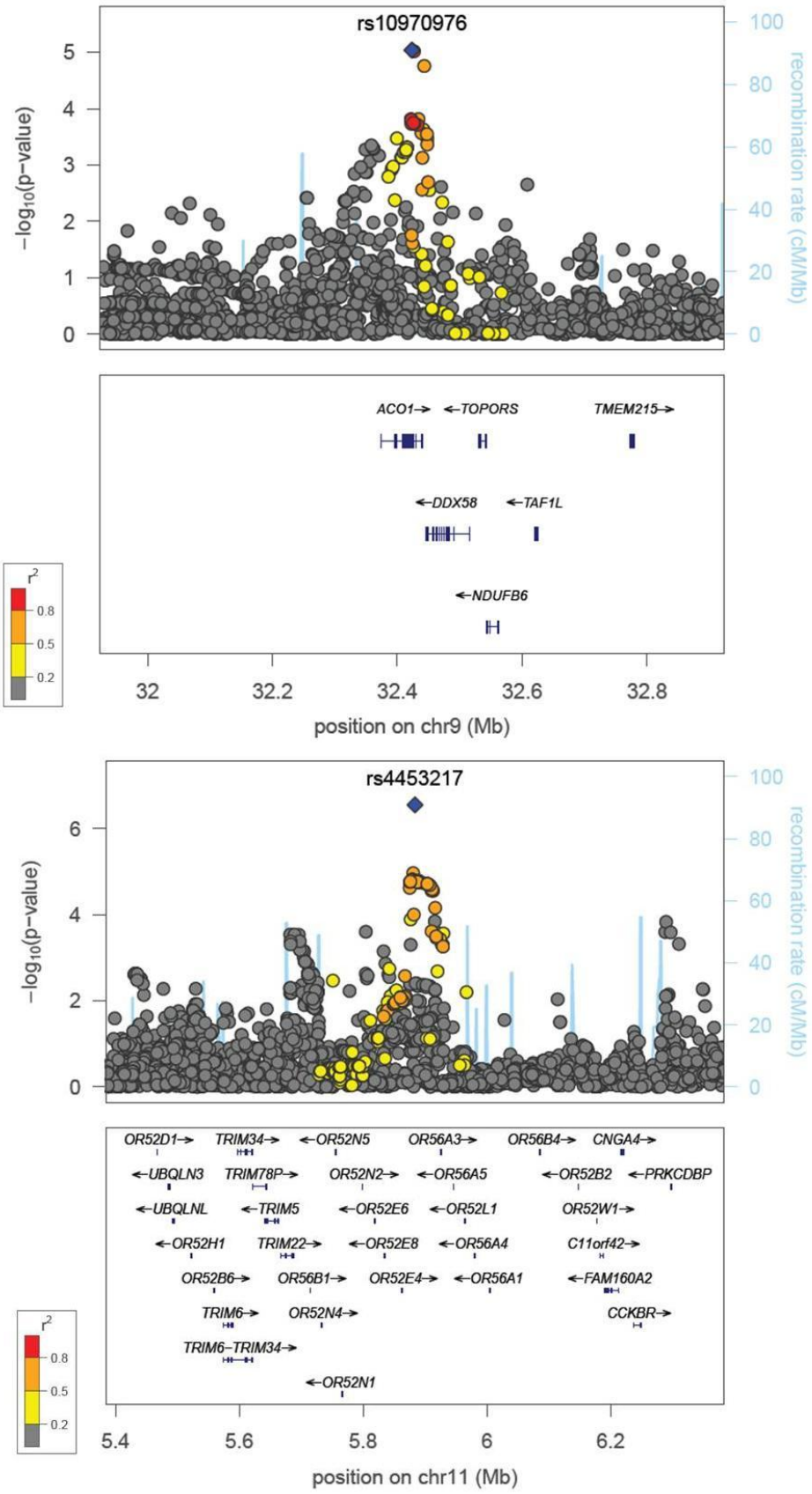


Figure A5.5 - Region plots for rs10970976 and rs4453217 in new regions identified by imputation with $p<10^{(-5)}$ and $r^2_{imp}=0.7$.

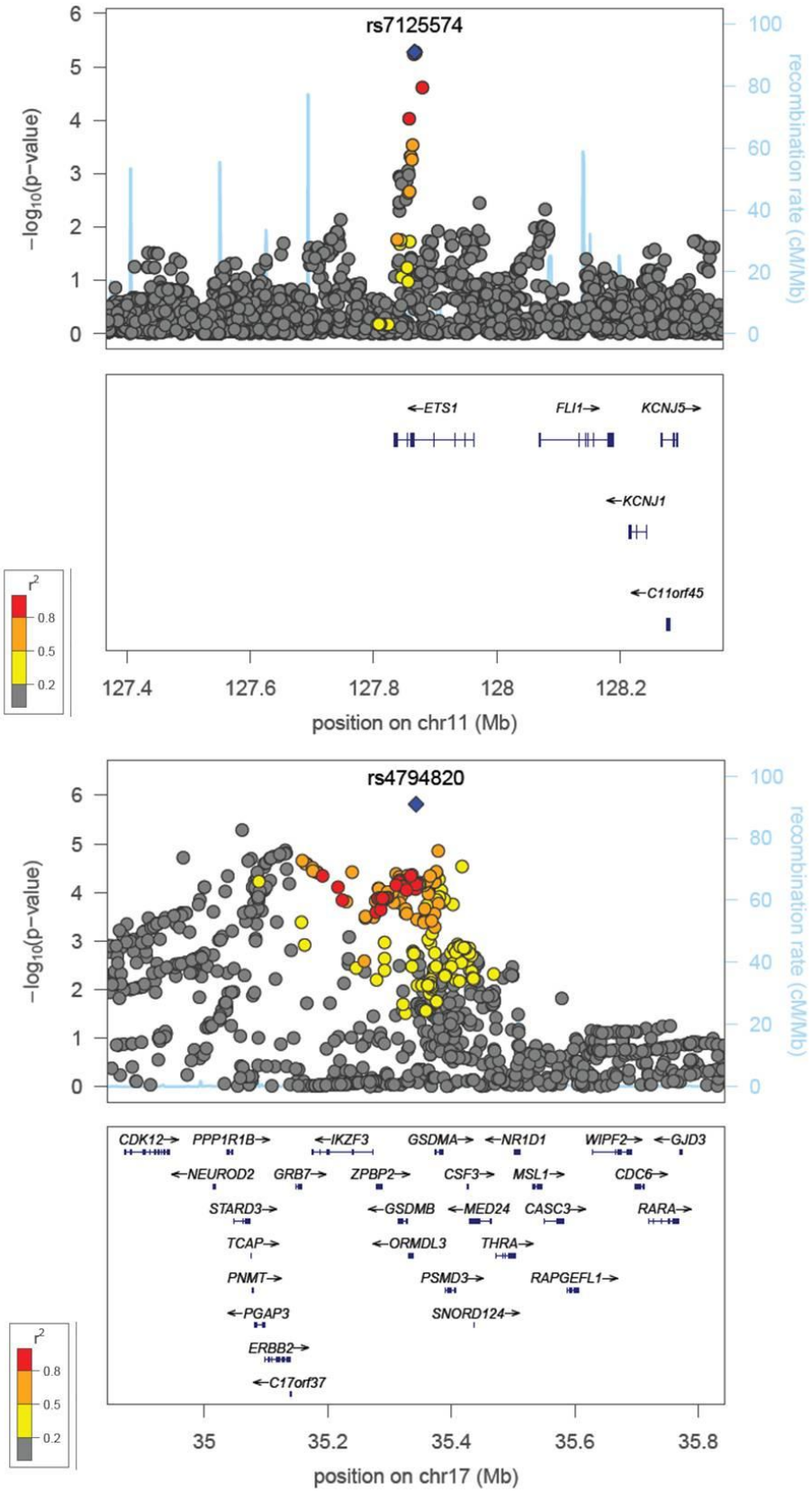


Figure A5.6 - Region plots for rs7125574 and rs4794820 in new regions identified by imputation with $p<10^{(-5)}$ and $r^2_{imp}=0.7$.

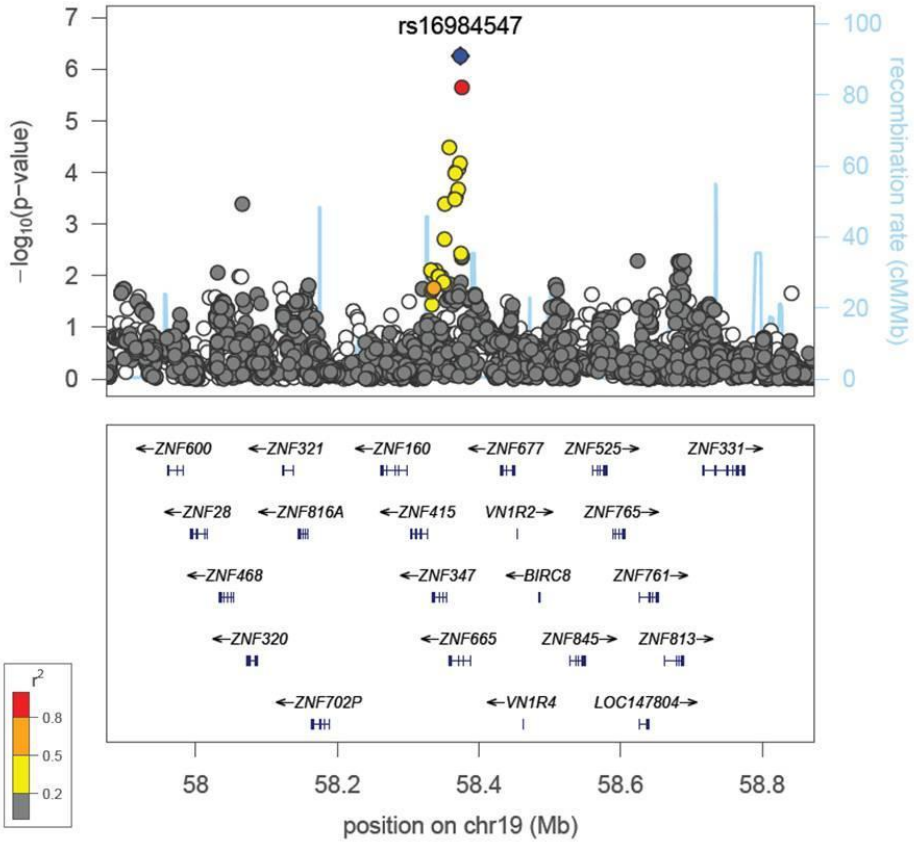


Figure A5.6 - Region plots rs16984547 in a new region identified by imputation with $p<10^{(-5)}$ and $r^2_{imp}=0.7$.

Appendix 6 Identified Candidate Loci from Previous Literature and Corresponding Region Co-ordinates

The full list of identified loci previously reported in the literature and corresponding region co-ordinates used to extract SNP results for the evaluation of candidate loci are listed in Table A6. A total of 1,830 papers were reviewed identifying 171 genes, two SNPs and six microsatellites from 303 recorded papers. Region co-ordinates (±10kb) were based on the position of transcriptional start and end sites of genes, location of SNPs and sequence-tagged sites (STS) start and end positions of microsatellites. The relevant references specific to this table are listed in the following pages.

Genes							
Chromosome	**Gene**	**Transcriptional**		**References**	**Region**		**Positive references**
		Start	**End**		**Start**	**End**	
1p31.2	*PTGER3*	71190702	71286079	Kim *et al.* 2007; Park *et al.* 2010	71180702	71296079	2
1p33-p32	*TGFBR3*	91920574	92124243	Kim *et al.* 2010	91910574	92134243	1
1p21	*S1PR1*	101474892	101479664	Sun *et al.* 2010	101464892	101489664	1
1p13.3	*GSTM1*	110031964	110037890	Ivaschenko *et al.* 2002; Brasch-Andersen *et al.* 2004; Hanene *et al.* 2007; Mak *et al.* 2007; Islam *et al.* 2009	110021964	110047890	5
1p13.2	*CHIA*	111635006	111664708	Bierbaum *et al.* 2005; Chatterjee *et al.* 2008	111625006	111674708	2
1q21.3	*FLG*	150541274	150564303	Muller *et al.* 2009	150531274	150574303	1
1q22	*DAP3*	153925497	153974943	Hirota *et al.* 2004	153915497	153984943	1
1q21-q22	*DARC*	157441824	157442914	Vergara *et al.* 2008	157431824	157452914	1
1q25.2-q25.3	*PTGS2*	184907591	184916179	Szczeklik *et al.* 2004; Chan *et al.* 2007; Kormann *et al.* 2008	184897591	184926179	3
1q31.3	*DENND1B*	195788007	196011179	Sleiman *et al.* 2010	195778007	196021179	1
1q32.1	*CHI3L1*	201414681	201422545	Ober *et al.* 2008; Rathcke *et al.* 2009	201404681	201432545	2
1q31-q32	*IL10*	205007570	205012462	Hang *et al.* 2003; Chatterjee *et al.* 2005; Wjst 2005; Zedan *et al.* 2008	204997570	205022462	4
1q41-q42	*PARP1*	224615014	224662424	Tezcan *et al.* 2009	224605014	224672424	1
1q44	*NLRP3*	245646097	245679029	Hitomi *et al.* 2009	245636097	245689029	1
2q11.2	*INPP4A*	98427844	98564716	Sharma *et al.* 2008	98417844	98574716	1
2q12	*IL1R1*	102136832	102162766	Pattaro *et al.* 2006; Settin *et al.* 2008; Zeyrek *et al.* 2008	102126832	102172766	3
2q12	*IL1RL1*	102294393	102334929	Gudbjartsson *et al.* 2009; Wu *et al.* 2009; Moffatt *et al.* 2010	102284393	102344929	3
2q12	*IL18R1*	102345524	102381649	Wu *et al.* 2009; Moffatt *et al.* 2010	102335524	102391649	2
2q14	*IL1B*	113303807	113310827	Zeyrek *et al.* 2008	113293807	113320827	1
2q14.2	*IL1RN*	113601608	113608064	Mao *et al.* 2000; Gohlke *et al.* 2004	113591608	113618064	2
2q14.1	*DPP10*	115636201	116318406	Wu *et al.* 2009; Zhou *et al.* 2009	115626201	116328406	2
2q22.1	*HNMT*	138438277	138490404	Yan *et al.* 2000	138428277	138500404	1
2q31.3	*PDE11A*	178196222	178645728	DeWan *et al.* 2010	178186222	178655728	1
2q33	*CTLA4*	204440753	204446928	Howard *et al.* 2002; Schubert *et al.* 2006	204430753	204456928	2
3p26-p24	*IL5RA*	3086420	3127031	Cheong *et al.* 2005	3076420	3137031	1
3p25	*PPARG*	12368000	12450855	Oh *et al.* 2009	12358000	12460855	1
3p21.3	*CX3CR1*	39279988	39296531	Tremblay *et al.* 2006	39269988	39306531	1
3p21.3	*CCR3*	46258691	46283166	Fukunaga *et al.* 2001; Al-Abdulhadi and Al-Rabia 2009	46248691	46293166	2
3p21	*CCR2*	46370238	46377429	Mak *et al.* 2007	46360238	46387429	1
3p21	*CCR5*	46386636	46392701	Hall *et al.* 1999; McGinnis *et al.* 2002; Srivastava *et al.* 2003; Berce *et al.* 2008	46376636	46402701	4
3p21.3	*TLR9*	52230137	52235219	Lazarus *et al.* 2003; Lachheb *et al.* 2008	52220137	52245219	2

3p21.1	*IL17RB*	53855616	53874867	Jung *et al.* 2009	53845616	53884867	1
3q21	*CD86*	123256910	123322673	Wu *et al.* 2007	123246910	123332673	1
3q21	*MYLK*	124813832	125085839	Flores *et al.* 2007	124803832	125095839	1
3q21	*TF*	134947489	134980325	Isada *et al.* 2010	134937489	134990325	1
4p14	*TLR10*	38450646	38460984	Lazarus *et al.* 2004	38440646	38470984	1
4p14	*TLR1*	38474270	38482807	Kormann *et al.* 2008	38464270	38492807	1
4p16.1	*TLR6*	38504802	38507553	Tantisira *et al.* 2004; Hoffjan *et al.* 2005; Kormann *et al.* 2008	38494802	38517553	3
4q12	*PDGFRA*	54790020	54859169	Wu *et al.* 2006	54780020	54869169	1
4q13.3	*MUC7*	71372524	71383303	Kirkbride *et al.* 2001	71362524	71393303	1
4q13-q21	*IL8*	74825138	74828297	Heinzmann *et al.* 2004	74815138	74838297	1
4q23	*ADH5*	100211152	100228954	Corydon *et al.* 2007; Wu *et al.* 2007	100201152	100238954	2
4q26-q27	*IL2*	123592075	123597100	Christensen *et al.* 2006; Trajkov *et al.* 2008; Sekigawa *et al.* 2009	123582075	123607100	3
4q31	*IL15*	142777203	142874062	Bierbaum *et al.* 2006	142767203	142884062	1
4q32	*TLR2*	154824890	154846692	Eder *et al.* 2004; Bjornvold *et al.* 2009; Smit *et al.* 2009; Qian *et al.* 2010	154814890	154856692	4
5q35	*ADAMTS12*	33563043	33927881	Kurz *et al.* 2006	33553043	33937881	1
5p13	*IL7R*	35892747	35912680	Kurz *et al.* 2006	35882747	35922680	1
5p13.1	*PTGER4*	40715788	40729594	Kurz *et al.* 2006; Kim *et al.* 2007; Park *et al.* 2007	40705788	40739594	3
5q12	*PDE4D*	58300622	59225378	Himes *et al.* 2009	58290622	59235378	1
5q22.1	*TSLP*	110435288	110441621	He *et al.* 2009	110425288	110451621	1
5q22.2	*WDR36*	110455768	110494099	Gudbjartsson *et al.* 2009	110445768	110504099	1
5q23-q31	*IRF1*	131846683	131854326	Nakao *et al.* 2001; Wang *et al.* 2006	131836683	131864326	2
5q23-q31	*RAD50*	131920528	132007494	Li *et al.* 2010	131910528	132017494	1
5q31	*IL13*	132021763	132024700	Heinzmann *et al.* 2000; Howard *et al.* 2001; Heinzmann *et al.* 2003; Park *et al.* 2004; Xi *et al.* 2004; Hosseini-Farahabadi *et al.* 2007; Black *et al.* 2009; Bottema *et al.* 2010; Li *et al.* 2010; Undarmaa *et al.* 2010; Wu *et al.* 2010	132011763	132034700	11
5q23-q31	*IL4*	132037271	132046267	Noguchi *et al.* 1998; Chouchane *et al.* 1999; Beghe *et al.* 2003; Child *et al.* 2003; Nagarkatti *et al.* 2004; Wilson *et al.* 2004; Gervaziev *et al.* 2006; Schubert *et al.* 2006; Hosseini-Farahabadi *et al.* 2007; Kamali-Sarvestani *et al.* 2007; Trajkov *et al.* 2008; Amirzargar *et al.* 2009	132027271	132056267	13
5q22-q32	*CD14*	139991500	139993439	Sharma *et al.* 2004; Zambelli-Weiner *et al.* 2005; Lachheb *et al.* 2008; Wang *et al.* 2009	139981500	140003439	4
5q32	*SCGB3A2*	147238466	147241946	Niimi *et al.* 2002	147228466	147251946	1
5q31-q32	*SPINK5*	147423727	147497118	Kabesch *et al.* 2004; Liu *et al.* 2009	147413727	147507118	2
5q31-q32	*ADRB2*	148186348	148188381	Hopes *et al.* 1998; Santillan *et al.* 2003; Thakkinstian *et al.* 2005; Szczepankiewicz *et al.* 2009	148176348	148198381	4
5q32	*CSF1R*	149413050	149473128	Shin *et al.* 2010	149403050	149483128	1
5q31-q32	*SLC6A7*	149549712	149570828	Kim *et al.* 2010	149539712	149580828	1

5q33.3	*TIMD4*	156278947	156322844	Cai *et al.* 2009	156268947	156332844	1
5q33.2	*HAVCR1*	156389108	156418065	Chae *et al.* 2003; Chae *et al.* 2003; Chae *et al.* 2004; Gao *et al.* 2005	156379108	156428065	4
5q34	*CYFIP2*	156625668	156755184	Noguchi *et al.* 2005	156615668	156765184	1
5q31.1-q33.1	*IL12B*	158674368	158690059	Randolph *et al.* 2004; Wjst *et al.* 2006	158664368	158700059	2
6p24.3	*LY86*	6533932	6600215	Lee *et al.* 2008	6523932	6610215	1
6p24.1	*EDN1*	12398514	12405413	Immervoll *et al.* 2001; Zhu *et al.* 2008	12388514	12415413	2
6p21.3	*MICA*	31479349	31491069	Hui *et al.* 2006	31469349	31501069	1
6p21.3	*LTA*	31648071	31650077	Moffatt and Cookson 1997; Albuquerque *et al.* 1998; Mak *et al.* 2007; Huang *et al.* 2008	31638071	31660077	4
6p21.3	*TNF*	31651328	31654091	Moffatt and Cookson 1997; Albuquerque *et al.* 1998; Chagani *et al.* 1999; Winchester *et al.* 2000; Fan and Jung 2002; Di Somma *et al.* 2003; Shin *et al.* 2004; Gupta *et al.* 2005; Li *et al.* 2006; Schubert *et al.* 2006; Sharma *et al.* 2006; Munthe-Kaas *et al.* 2007; Castro-Giner *et al.* 2008; Kumar *et al.* 2008; Kumar *et al.* 2008; Trajkov *et al.* 2008; Zedan *et al.* 2008; Jimenez-Morales *et al.* 2009; Puthothu *et al.* 2009; Undarmaa *et al.* 2010	31641328	31664091	20
6p21.3	*BTNL2*	32470490	32482878	Li *et al.* 2010	32460490	32492878	1
6p21.3	*HLA-DRA*	32515624	32520802	Li *et al.* 2010	32505624	32530802	1
6p21.3	*HLA-DRB1*	32628467	32635757	Soriano *et al.* 1997; Lara-Marquez *et al.* 1999; Cho *et al.* 2000; Horne *et al.* 2000; Di Somma *et al.* 2003; Torio *et al.* 2003; Parapanissiou *et al.* 2005; Juhn *et al.* 2007; Munthe-Kaas *et al.* 2007; Li *et al.* 2010	32618467	32645757	10
6p21.3	*HLA-DQA1*	32713160	32719407	Mapp *et al.* 2000; Gao *et al.* 2003; Parapanissiou *et al.* 2005; Li *et al.* 2010; Moffatt *et al.* 2010	32703160	32729407	5
6p21.3	*HLA-DQB1*	32735634	32742444	Balboni *et al.* 1996; Horne *et al.* 2000; Mapp *et al.* 2000; Gao *et al.* 2003; Torio *et al.* 2003; Movahedi *et al.* 2008; Li *et al.* 2010; Moffatt *et al.* 2010	32725634	32752444	8
6p21.3	*TAP1*	32920963	32929726	Ismail *et al.* 1997	32910963	32939726	1
6p21.3	*HLA-DPB1*	33151737	33162954	Caraballo *et al.* 1991; Choi *et al.* 2004	33141737	33172954	2
6p12	*VEGFA*	43845930	43862201	Sharma *et al.* 2009	43835930	43872201	1
6p12	*IL17F*	52209442	52217257	Hizawa *et al.* 2006, 602	52199442	52227257	2
6p12.2	*GSTA1*	52764346	52776616	Polimanti *et al.* 2010	52754346	52786616	1
6q21	*FYN*	112089177	112147958	Szczepankiewicz *et al.* 2008	112079177	112157958	1
6q25	*SOD2*	160020138	160034343	Mak *et al.* 2006	160010138	160044343	1
7p21-p15	*IL6*	22733322	22738145	Settin *et al.* 2008; Trajkov *et al.* 2008	22723322	22748145	2
7p15-p14	*NOD1*	30430674	30484790	Weidinger *et al.* 2005	30420674	30494790	1
7p14.3	*NPSR1*	34664421	34884469	Kormann *et al.* 2005; Hersh *et al.* 2007; Castro-Giner *et al.* 2009; Vergara *et al.* 2009	34654421	34894469	4
7p14-p12	*AOAH*	36519132	36730678	Barnes *et al.* 2006	36509132	36740678	1
7p14	*TCRG*	38245705	38262515	Walsh *et al.* 2010	38235705	38272515	1
7q11.2	*CCL26*	75236777	75257000	Chae *et al.* 2004	75226777	75267000	1
7q11.23	*CCL24*	75279049	75281028	Chae *et al.* 2004	75269049	75291028	1

7q21.3-q22	*SERPINE1*	100557098	100569262	Buckova *et al.* 2002	100547098	100579262	1
7q36	*NOS3*	150319079	150342609	Lee *et al.* 2000; Yanamandra *et al.* 2005; Holla *et al.* 2006	150309079	150352609	3
8p23.1	*DEFB1*	6715508	6722939	Levy *et al.* 2005	6705508	6732939	1
8q22	*NAT2*	18293034	18303003	Nacak *et al.* 2002; Batra *et al.* 2006	18283034	18313003	2
8p12-p11	*IDO1*	39890484	39905104	Sekigawa *et al.* 2009	39880484	39915104	1
9p24.1	*IL33*	6231683	6247982	Gudbjartsson *et al.* 2009; Moffatt *et al.* 2010	6221683	6257982	2
9p24.1-p23	*PTPRD*	8304245	10602505	Shyur *et al.* 2008	8294245	10612505	1
9q31	*IKBKAP*	110669620	110736429	Takeoka *et al.* 2001	110659620	110746429	1
9q32-q34	*TNC*	116822625	116920307	Matsuda *et al.* 2005	116812625	116930307	1
9q34.2-q34.3	*PTGDS*	138990821	138996015	Noguchi *et al.* 2002	138980821	139006015	1
10p13	*MRC1*	17891367	17993184	Hattori *et al.* 2009	17881367	18003184	1
10q21	*CTNNA3*	67349730	69125955	Kim *et al.* 2009	67339730	69135955	1
10q24	*PLAU*	75340895	75347261	Begin *et al.* 2007	75330895	75357261	1
10q25.1	*GSTO1*	106004667	106021203	Polimanti *et al.* 2010	105994667	106031203	1
11q12-q13.1	*MS4A3*	59580676	59595164	Adra *et al.* 1999	59570676	59605164	1
11q12-q13	*MS4A2*	59612712	59622592	Shirakawa *et al.* 1994; Adra *et al.* 1999; Palmer *et al.* 1999; Cui *et al.* 2003; Sharma *et al.* 2004; Zhang *et al.* 2004; Sharma and Ghosh 2009; Wang *et al.* 2009	59602712	59632592	8
11q12-q13.3	*GPR44*	60374973	60380020	Huang *et al.* 2004; Wang *et al.* 2009	60364973	60390020	2
11q11-qter	*SCGB1A1*	61943098	61947243	Laing *et al.* 2000; Sharma and Ghosh 2004; Candelaria *et al.* 2005	61933098	61957243	3
11q12-q13	*CHRM1*	62432726	62445588	Maeda *et al.* 2006	62422726	62455588	1
11q12.3	*SLC22A10*	62814005	62835822	Adra *et al.* 1999	62804005	62845822	1
11q13-qter	*GSTP1*	67107861	67110699	Fryer *et al.* 2000; Child *et al.* 2003; Aynacioglu *et al.* 2004; Lee *et al.* 2005; Hanene *et al.* 2007; Kamada *et al.* 2007; Li *et al.* 2008; Islam *et al.* 2009	67097861	67120699	8
11q13	*CTTN*	69922259	69960338	Ma *et al.* 2008	69912259	69970338	1
11q22.2-q22.3	*IL18*	111519185	111540050	Higa *et al.* 2003; Imboden *et al.* 2006; Lachheb *et al.* 2008; Lee *et al.* 2008; Zhu *et al.* 2008	111509185	111550050	5
12p13	*SLC6A12*	169510	193632	Pasaje *et al.* 2010	159510	203632	1
12p13.31	*LPCAT3*	6955607	6996103	Srivastava *et al.* 1985	6945607	7006103	1
12q12-q14	*VDR*	46521586	46585081	Poon *et al.* 2004; Raby *et al.* 2004; Wjst *et al.* 2006; Saadi *et al.* 2009	46511586	46595081	4
12q13.2	*STAT2*	55021648	55040176	Hsieh *et al.* 2009	55011648	55050176	1
12q13	*STAT6*	55775459	55791428	Chae *et al.* 2003	55765459	55801428	1
12q14	*IFNG*	66834816	66839788	Nakao *et al.* 2001; Nagarkatti *et al.* 2002; Wang *et al.* 2006; Kumar and Ghosh 2008	66824816	66849788	4
12q22	*LTA4H*	94918741	94953496	Holloway *et al.* 2008; Via *et al.* 2010	94908741	94963496	2
12q14-qter	*NOS1*	116135361	116283965	Grasemann *et al.* 1999; Gao *et al.* 2000; Grasemann *et al.* 2000; Shao *et al.* 2004; Martinez *et al.* 2007	116125361	116293965	5
12q24.33	*SFRS8*	130761587	130850235	Eder *et al.* 2006	130751587	130860235	1

13q12	*ALOX5AP*	30207668	30236556	Holloway *et al.* 2008; Via *et al.* 2010	30197668	30246556	2
13q14.2	*CYSLTR2*	48178953	48181499	Pillai *et al.* 2004	48168953	48191499	1
13q14.11	*PHF11*	48967801	49001118	Zhang *et al.* 2003; Gao *et al.* 2009	48957801	49011118	2
14q24-q31	*RNASE3*	20429401	20430347	Munthe-Kaas *et al.* 2007	20419401	20440347	1
14	*CMA1*	24044551	24047311	Sharma *et al.* 2005; Hossny *et al.* 2008	24034551	24057311	2
14q22.1	*PTGDR*	51804180	51813192	Oguma *et al.* 2004; Sanz *et al.* 2006; Zhu *et al.* 2007	51794180	51823192	3
14q22	*PTGER2*	51850862	51865072	Kim *et al.* 2007; Park *et al.* 2007	51840862	51875072	2
14q24.1	*ARG2*	67156331	67188189	Salam *et al.* 2009; Vonk *et al.* 2010	67146331	67198189	2
15q21-q22	*SMAD3*	65145248	65274587	Moffatt *et al.* 2010	65135248	65284587	1
15q26.3	*IL16*	79276273	79392157	Burkart *et al.* 2006; Hosseini-Farahabadi *et al.* 2007	79266273	79402157	2
16p13.13	*SOCS1*	11255774	11257540	Harada *et al.* 2007	11245774	11267540	1
16p13.1	*ABCC1*	15950934	16144431	Kedda *et al.* 2004	15940934	16154431	1
16p12.1-p11.2	*IL4R*	27232751	27283600	Ober *et al.* 2000; Takabayashi *et al.* 2000; Risma *et al.* 2002; Beghe *et al.* 2003; Zhang *et al.* 2007; Zhang *et al.* 2007; Lv *et al.* 2009; Berce and Potocnik 2010; Undarmaa *et al.* 2010	27222751	27293600	9
16p11	*IL27*	28418183	28425656	Chae *et al.* 2007	28408183	28435656	1
16q12	*NOD2*	49288550	49324488	Weidinger *et al.* 2005; Wjst 2005	49278550	49334488	2
16q12-q22	*NQO1*	68300804	68318034	Li *et al.* 2009	68290804	68328034	1
16q24	*CYBA*	87237197	87244958	Izakovicova Holla *et al.* 2009	87227197	87254958	1
17p13.3	*PAFAH1B1*	2443672	2535659	Kruse *et al.* 2000; Ito *et al.* 2002	2433672	2545659	2
17p13.3	*TRPV1*	3415489	3447085	Zhang *et al.* 2010	3405489	3457085	1
17q11.2-q21.1	*CCL2*	29606408	29608333	Chelbi *et al.* 2008	29596408	29618333	1
17q21.1-q21.2	*CCL11*	29636799	29639312	Batra *et al.* 2007	29626799	29649312	1
17q11.2-q12	*CCL5*	31222608	31231490	Fryer *et al.* 2000; Yao *et al.* 2003; Al-Abdulhadi *et al.* 2005; Lachheb *et al.* 2007	31212608	31241490	4
17q21.1	*ZPBP2*	35277980	35287675	Madore *et al.* 2008	35267980	35297675	1
17q21.2	*GSDMB*	35314373	35328429	Moffatt *et al.* 2010	35304373	35338429	1
17q21.2	*GSDML*	35314373	35328429	Madore *et al.* 2008; Wu *et al.* 2009	35304373	35338429	2
17q21	*ORMDL3*	35330821	35337380	Moffatt *et al.* 2007; Brown *et al.* 2008; Galanter *et al.* 2008; Madore *et al.* 2008; Li *et al.* 2009; Wu *et al.* 2009, Halapi *et al.* 2010; Moffatt *et al.* 2010	35320821	35347380	8
17q21.32	*ITGB3*	42686206	42745076	Weiss *et al.* 2005	42676206	42755076	1
17q21.2	*TBX21*	43165608	43178484	Raby *et al.* 2006; Suttner *et al.* 2009; Undarmaa *et al.* 2010	43155608	43188484	3
17q21.33	*TAC4*	45270669	45280378	Klassert *et al.* 2010	45260669	45290378	1
17q21.3-q23	*MPO*	53702215	53713295	Polonikov *et al.* 2009	53692215	53723295	1
17q23	*ACE*	58908165	58928711	Benessiano *et al.* 1997; Urhan *et al.* 2004	58898165	58938711	2
19p13.3	*TBXA2R*	3545503	3557658	Unoki *et al.* 2000; Leung *et al.* 2002; Kim *et al.* 2007; Undarmaa *et al.* 2010	3535503	3567658	4

19p13.3-p13.2	*C3*	6628845	6671662	Hasegawa *et al.* 2004; Inoue *et al.* 2008	6618845	6681662	2
19p13.3-p13.2	*ICAM1*	10242516	10258291	Puthothu *et al.* 2006	10232516	10268291	1
19p13.1	*PTGER1*	14444277	14447174	Park *et al.* 2010	14434277	14457174	1
19p13.1	*IL12RB1*	18031370	18058742	Wjst 2005	18021370	18068742	1
19q13.1	*TGFB1*	46528490	46551656	Pulleyn *et al.* 2001; Cakebread *et al.* 2004; Nagpal *et al.* 2005; Mak *et al.* 2006; Hatsushika *et al.* 2007; Salam *et al.* 2007; Lv *et al.* 2009; Wu *et al.* 2009; Zhang *et al.* 2010	46518490	46561656	9
19q13	*PLAUR*	48844571	48866342	Barton *et al.* 2009	48834571	48876342	1
19q13.3	*PTGIR*	51815564	51820194	Kim *et al.* 2007	51805564	51830194	1
19q13.33-q13.41	*SIGLEC8*	56646062	56653520	Gao *et al.* 2010	56636062	56663520	1
19q13.4	*CACNG6*	59187353	59207732	Lee *et al.* 2010	59177353	59217732	1
20p13	*ADAM33*	3596619	3610738	Van Eerdewegh *et al.* 2002; Howard *et al.* 2003; Werner *et al.* 2004; Hirota *et al.* 2006; Noguchi *et al.* 2006; Sakagami *et al.* 2007; Su *et al.* 2008; Thongngarm *et al.* 2008; Zhang *et al.* 2009; Undarmaa *et al.* 2010	3586619	3620738	10
20q12-q13	*MMP9*	44070953	44078607	Nakashima *et al.* 2006	44060953	44088607	1
22q11.23	*MIF*	22566564	22567409	Wu *et al.* 2009	22556564	22577409	1
22q11.23	*GSTT1*	22706138	22714284	Fan and Jung 2002; Brasch-Andersen *et al.* 2004; Babusikova *et al.* 2009	22696138	22724284	3
22q13	*IL2RB*	35851825	35875908	Moffatt *et al.* 2010	35841825	35885908	1

SNPs						
Chromosome	**Marker**	**Position**	**References**	**Region Start**	**Region End**	**Evidence**
9q21.31	rs2378383	81229182	(Hancock *et al.* 2009)	81219182	81239182	1
2q14	D2S308 SNP	261056	(Allen *et al.* 2003)	251056	271056	1

Microsatellites							
Chromosome	**Marker**	**STS start**	**STS end**	**References**	**Region Start**	**Region End**	**Evidence**
5q13	D5S2498	87363565	87363812	Teerlink *et al.* 2009	87353565	87373812	1
11q13	D11S1337	67888234	67888517	Huang *et al.* 2003	67878234	67898517	1
14q24	D14S588	69290039	69290158	Hakonarson *et al.* 2002	69280039	69300158	1
14q24	D14S603	69593706	69593828	Hakonarson *et al.* 2002	69583706	69603828	1
5q31-33	D5S820	156054978	156055188	Yokouchi *et al.* 2000	156044978	156065188	1
5q31-33	D5S1471	166809050	166809216	Yokouchi *et al.* 2000	166799050	166819216	1

References to Candidate Loci Tables

Adra, C. N., *et al.* (1999). Clin Genet **55**(6): 431-7.
Al-Abdulhadi, S. A., *et al.* (2005). Genes Immun **6**(1): 24-30.
Al-Abdulhadi, S. A. and Al-Rabia, M. W. (2009). Egypt J Immunol **16**(1): 83-93.
Albuquerque, R. V., *et al.* (1998). Clin Exp Allergy**28**(5): 578-84.
Allen, M., *et al.* (2003). Nat Genet **35**(3): 258-63.
Amirzargar, A. A., *et al.* (2009). J Investig Allergol Clin Immunol **19**(6): 433-8.
Aynacioglu, A. S., *et al.* (2004). Br J Clin Pharmacol **57**(2): 213-7.
Babusikova, E., *et al.* (2009). J Physiol Pharmacol **60 Suppl 5**: 27-30.
Balboni, A., *et al.* (1996). Eur Respir J **9**(2): 207-10.
Barnes, K. C., *et al.* (2006). J Allergy Clin Immunol **118**(1): 70-7.
Barton, S. J., *et al.* (2009). J Allergy Clin Immunol **123**(6): 1391-400 e17.
Batra, J., *et al.* (2006). Pharmacogenomics **7**(5): 673-82.
Batra, J., *et al.* (2007). J Med Genet **44**(6): 397-403.
Beghe, B., *et al.* (2003). Clin Exp Allergy **33**(8): 1111-7.
Begin, P., *et al.* (2007). Am J Respir Crit Care Med **175**(11): 1109-16.
Benessiano, J., *et al.* (1997). J Allergy Clin Immunol **99**(1 Pt 1): 53-7.
Berce, V. and Potocnik, U. (2010). Wien Klin Wochenschr **122 Suppl 2**: 11-8.
Berce, V., *et al.* (2008). J Asthma **45**(9): 780-4.
Bierbaum, S., *et al.* (2005). Am J Respir Crit Care Med **172**(12): 1505-9.
Bierbaum, S., *et al.* (2006). Allergy **61**(5): 576-80.
Bjornvold, M., *et al.* (2009). Genes Immun **10**(2): 181-7.
Black, S., *et al.* (2009). Allergy **64**(8): 1172-8.
Bottema, R. W., *et al.* (2010). Int Arch Allergy Immunol **153**(3): 259-67.
Brasch-Andersen, C., *et al.* (2004). Hum Mutat **24**(3): 208-14.
Brown, S. J., *et al.* (2008). J Allergy Clin Immunol **121**(4): 940-46 e3.
Buckova, D., *et al.* (2002). Allergy **57**(5): 446-8.
Burkart, K. M., *et al.* (2006). J Allergy Clin Immunol **117**(1): 86-91.
Cai, P. C., *et al.* (2009). Tissue Antigens **74**(1): 11-6.
Cakebread, J. A., *et al.* (2004). Springer Semin Immunopathol **25**(3-4): 361-75.
Candelaria, P. V., *et al.* (2005). Immunogenetics **57**(1-2): 25-32.
Caraballo, L., *et al.* (1991). Hum Immunol **32**(3): 157-61.
Castro-Giner, F., *et al.* (2008). Eur Respir J **32**(2): 350-61.
Castro-Giner, F., *et al.* (2009). Thorax **65**(2): 124-31.
Chae, S. C., *et al.* (2003). Biochem Biophys Res Commun **312**(2): 346-50.
Chae, S. C., *et al.* (2003). Hum Immunol **64**(12): 1177-82.
Chae, S. C., *et al.* (2004). Biochem Biophys Res Commun **320**(1): 131-7.
Chae, S. C., *et al.* (2004). Hum Immunol **65**(12): 1427-31.
Chae, S. C., *et al.* (2007). J Hum Genet **52**(4): 355-61.
Chagani, T., *et al.* (1999). Am J Respir Crit Care Med **160**(1): 278-82.
Chan, I. H., *et al.* (2007). Allergy **62**(7): 802-9.
Chatterjee, R., *et al.* (2005). Clin Exp Allergy **35**(7): 914-9.
Chatterjee, R., *et al.* (2008). J Allergy Clin Immunol **122**(1): 202-8, 208 e1-7.
Chelbi, H., *et al.* (2008). Genes Immun **9**(7): 575-81.
Cheong, H. S., *et al.* (2005). J Hum Genet **50**(12): 628-34.
Child, F., *et al.* (2003). Respir Med **97**(12): 1247-56.
Cho, S. H., *et al.* (2000). Clin Exp Allergy**30**(11): 1568-75.
Choi, J. H., *et al.* (2004). J Allergy Clin Immunol **113**(3): 562-4.
Chouchane, L., *et al.* (1999). Int Arch Allergy Immunol **120**(1): 50-5.
Christensen, U., *et al.* (2006). Eur J Hum Genet **14**(2): 227-35.
Corydon, T. J., *et al.* (2007). J Med Genet **44**(8): 509-15.
Cui, T., *et al.* (2003). Chin Med J (Engl) **116**(12): 1875-8.
DeWan, A. T., *et al.* (2010). J Allergy Clin Immunol **126**(4): 871-873 e9.
Di Somma, C., *et al.* (2003). Hum Immunol **64**(3): 359-65.
Eder, W., *et al.* (2004). J Allergy Clin Immunol **113**(3): 482-8.
Eder, W., *et al.* (2006). Allergy **61**(9): 1117-24.
Fan, R. and Jung, J. (2002). Hum Hered **54**(3): 132-50.
Flores, C., *et al.* (2007). Genet Epidemiol **31**(4): 296-305.
Fryer, A. A., *et al.* (2000). Am J Respir Crit Care Med **161**(5): 1437-42.
Fryer, A. A., *et al.* (2000). Genes & Immunity **1**(8): 509-14.
Fukunaga, K., *et al.* (2001). Eur Respir J **17**(1): 59-63.
Galanter, J., *et al.* (2008). Am J Respir Crit Care Med **177**(11): 1194-200.

Gao, P. S., *et al.* (2000). Biochem Biophys Res Commun **267**(3): 761-3.
Gao, J., *et al.* (2003). Chin Med J (Engl) **116**(7): 1078-82.
Gao, P. S., *et al.* (2005). J Allergy Clin Immunol **115**(5): 982-8.
Gao, J., *et al.* (2009). Respiration **79**(1): 17-24.
Gao, P. S., *et al.* (2010). Eur J Hum Genet **18**(6): 713-9.
Gervaziev, Y. V., *et al.* (2006). Int Arch Allergy Immunol **141**(3): 257-64.
Gohlke, H., *et al.* (2004). Am J Respir Crit Care Med **169**(11): 1217-23.
Grasemann, H., *et al.* (1999). Clin Exp Allergy **29 Suppl 4**: 39-41.
Grasemann, H., *et al.* (2000). Biochem Biophys Res Commun **272**(2): 391-4.
Gudbjartsson, D. F., *et al.* (2009). Nat Genet **41**(3): 342-7.
Gupta, V., *et al.* (2005). J Asthma **42**(10): 839-41.
Hakonarson, H., *et al.* (2002). Am J Med Genet **71**(3): 483-91.
Halapi, E., *et al.* (2010). Eur J Hum Genet **18**(8): 902-8.
Hall, I. P., *et al.* (1999). Lancet **354**(9186): 1264-5.
Hancock, D. B., *et al.* (2009). PLoS Genet **5**(8): e1000623.
Hanene, C., *et al.* (2007). Mediators Inflamm **2007**: 19564.
Hang, L. W., *et al.* (2003). J Clin Lab Anal **17**(5): 168-73.
Harada, M., *et al.* (2007). Am J Respir Cell Mol Biol **36**(4): 491-6.
Hasegawa, K., *et al.* (2004). Hum Genet **115**(4): 295-301.
Hatsushika, K., *et al.* (2007). Clin Exp Allergy **37**(8): 1165-74.
Hattori, T., *et al.* (2009). Immunogenetics **61**(11-12): 731-8.
He, J.-Q., *et al.* (2009). J Allergy Clin Immunol **124**(2): 222-9.
Heinzmann, A., *et al.* (2000). Hum Mol Genet **9**(4): 549-59.
Heinzmann, A., *et al.* (2003). J Allergy Clin Immunol **112**(4): 735-9.
Heinzmann, A., *et al.* (2004). J Allergy Clin Immunol **114**(3): 671-6.
Hersh, C. P., *et al.* (2007). Am J Respir Crit Care Med **176**(9): 849-57.
Higa, S., *et al.* (2003). Clin Exp Allergy **33**(8): 1097-102.
Himes, B. E., *et al.* (2009). Am J Hum Genet **84**: 581-593.
Hirota, T., *et al.* (2004). J Hum Genet **49**(7): 370-5.
Hirota, T., *et al.* (2006). Clin Exp Allergy **36**(7): 884-91.
Hitomi, Y., *et al.* (2009). J Allergy Clin Immunol **124**(4): 779-85 e6.
Hizawa, N., *et al.* (2006). Clin Exp Allergy **36**(9): 1109-14.
Hoffjan, S., *et al.* (2005). BMC Med Genet **6**: 34.
Holla, L. I., *et al.* (2006). Int Arch Allergy Immunol **141**(2): 103-9.
Holloway, J. W., *et al.* (2008). Allergy **63**(8): 1046-53.
Hopes, E., *et al.* (1998). BMJ **316**(7132): 664.
Horne, C., *et al.* (2000). Eur Respir J **15**(5): 911-4.
Hosseini-Farahabadi, S., *et al.* (2007). Iran J Allergy Asthma Immunol **6**(1): 9-14.
Hossny, E. M., *et al.* (2008). J Investig Allergol Clin Immunol **18**(5): 376-81.
Howard, T. D., *et al.* (2001). Am J Respir Cell Mol Biol **25**(3): 377-84.
Howard, T. D., *et al.* (2002). J Allergy Clin Immunol **110**(5): 743-51.
Howard, T. D., *et al.* (2003). J Allergy Clin Immunol **112**(4): 717-22.
Hsieh, Y. Y., *et al.* (2009). Int J Biol Sci **5**(1): 74-81.
Huang, J. L., *et al.* (2004). Hum Mol Genet **13**(21): 2691-7.
Huang, S. C., *et al.* (2008). Pediatr Neonatol **49**(2): 30-4.
Huang, S. K., *et al.* (2003). Hum Genet **113**(1): 71-5.
Hui, J., *et al.* (2006). Clin Exp Allergy **36**(6): 728-34.
Imboden, M., *et al.* (2006). Clin Exp Allergy **36**(2): 211-8.
Immervoll, T., *et al.* (2001). Hum Mutat **18**(4): 327-36.
Inoue, H., *et al.* (2008). J Hum Genet **53**(8): 728-38.
Isada, A., *et al.* (2010). J Hum Genet **55**(3): 167-74.
Islam, T., *et al.* (2009). Thorax **64**(3): 197-202.
Ismail, A., *et al.* (1997). J Allergy Clin Immunol **99**(2): 216-23.
Ito, S., *et al.* (2002). J Hum Genet **47**(2): 99-101.
Ivaschenko, T. E., *et al.* (2002). J Mol Med **80**(1): 39-43.
Izakovicova Holla, L., *et al.* (2009). Int Arch Allergy Immunol **148**(1): 73-80.
Jimenez-Morales, S., *et al.* (2009). Hum Immunol **70**(4): 251-6.
Juhn, Y. J., *et al.* (2007). Tissue Antigens **69**(1): 38-46.
Jung, J. S., *et al.* (2009). Chest **135**(5): 1173-80.
Kabesch, M., *et al.* (2004). Clin Exp Allergy **34**(3): 340-5.
Kamada, F., *et al.* (2007). Int Arch Allergy Immunol **144**(4): 275-86.
Kamali-Sarvestani, E., *et al.* (2007). J Investig Allergol Clin Immunol **17**(6): 361-6.
Kedda, M. A., *et al.* (2004). J Allergy Clin Immunol **113**(5): 889-95.

Kim, H. K., *et al.* (2010). Exp Mol Med **42**(6): 420-7.
Kim, J. H., *et al.* (2010). J Hum Genet **55**(6): 358-65.
Kim, S. H., *et al.* (2007). Pharmacogenet Genomics **17**(4): 295-304.
Kim, S. H., *et al.* (2009). Clin Exp Allergy **39**(2): 203-12.
Kirkbride, H. J., *et al.* (2001). Eur J Hum Genet **9**(5): 347-54.
Klassert, T. E., *et al.* (2010). J Neuroimmunol **227**(1-2): 202-7.
Kormann, M. S., *et al.* (2005). Am J Respir Crit Care Med **171**(12): 1358-62.
Kormann, M. S., *et al.* (2008). J Allergy Clin Immunol **122**(1): 86-92, 92 e1-8.
Kruse, S., *et al.* (2000). Am J Med Genet **66**(5): 1522-30.
Kumar, A., *et al.* (2008). Indian J Med Sci **62**(8): 323-30.
Kumar, A. and Ghosh, B. (2008). Genes Immun **9**(4): 294-301.
Kumar, V., *et al.* (2008). Natl Med J India **21**(3): 120-2.
Kurz, T., *et al.* (2006). J Allergy Clin Immunol **118**(2): 396-402.
Lachheb, J., *et al.* (2007). Human Immunology **68**(8): 675-80.
Lachheb, J., *et al.* (2008). Int J Immunogenet **35**(1): 63-8.
Lachheb, J., *et al.* (2008). Tissue Antigens **71**(5): 417-25.
Laing, I. A., *et al.* (2000). Am J Respir Crit Care Med **161**(1): 124-7.
Lara-Marquez, M. L., *et al.* (1999). Clin Exp Allergy**29**(1): 60-71.
Lazarus, R., *et al.* (2003). Genomics **81**(1): 85-91.
Lazarus, R., *et al.* (2004). Am J Respir Crit Care Med **170**(6): 594-600.
Lee, C. C., *et al.* (2008). J Clin Lab Anal **22**(1): 39-44.
Lee, J. S., *et al.* (2010). BMC Med Genet **11**: 138.
Lee, S. W., *et al.* (2008). J Microbiol Immunol Infect **41**(6): 445-9.
Lee, Y. C., *et al.* (2000). Allergy **55**(10): 959-63.
Lee, Y. L., *et al.* (2005). Chest **128**(3): 1156-62.
Leung, T. F., *et al.* (2002). Pediatr Allergy Immunol **13**(1): 10-7.
Levy, H., *et al.* (2005). J Allergy Clin Immunol **115**(2): 252-8.
Li, X., *et al.* (2010). J Allergy Clin Immunol **125**(2): 328-335 e11.
Li, Y. F., *et al.* (2006). Am J Respir Crit Care Med **173**(9): 970-6.
Li, Y. F., *et al.* (2008). Environ Health Perspect **116**(3): 409-15.
Li, Y. F., *et al.* (2009). Mutat Res **678**(1): 53-8.
Liu, Q., *et al.* (2009). BMC Med Genet **10**: 59.
Lv, J., *et al.* (2009). J Asthma **46**(6): 582-5.
Ma, S. F., *et al.* (2008). Genet Epidemiol **32**(8): 757-66.
Madore, A. M., *et al.* (2008). Hum Genet **123**(1): 93-5.
Maeda, Y., *et al.* (2006). Am J Respir Crit Care Med **174**(10): 1119-24.
Mak, J. C., *et al.* (2006). Clin Exp Allergy **36**(4): 440-7.
Mak, J. C., *et al.* (2006). J Allergy Clin Immunol **117**(1): 92-6.
Mak, J. C., *et al.* (2007). Clin Exp Allergy **37**(8): 1150-7.
Mak, J. C., *et al.* (2007). Int Arch Allergy Immunol **144**(2): 114-22.
Mao, X. Q., *et al.* (2000). Biochem Biophys Res Commun **276**(2): 607-12.
Mapp, C. E., *et al.* (2000). Clin Exp Allergy **30**(5): 651-6.
Martinez, B., *et al.* (2007). Int Arch Allergy Immunol **144**(2): 105-13.
Matsuda, A., *et al.* (2005). Hum Mol Genet **14**(19): 2779-86.
McGinnis, R., *et al.* (2002). Eur J Immunogenet **29**(6): 525-8.
Moffatt, M. F. and Cookson, W. O. (1997). Hum Mol Genet **6**(4): 551-4.
Moffatt, M. F., *et al.* (2007). Nature **448**(7152): 470-3.
Moffatt, M. F., *et al.* (2010). N Engl J Med **363**(13): 1211-21.
Movahedi, M., *et al.* (2008). Iran J Allergy Asthma Immunol **7**(4): 215-20.
Muller, S., *et al.* (2009). Pediatr Allergy Immunol **20**(4): 358-61.
Munthe-Kaas, M. C., *et al.* (2007). Allergy **62**(4): 429-36.
Munthe-Kaas, M. C., *et al.* (2007). Allergy **62**(9): 991-8.
Nacak, M., *et al.* (2002). Br J Clin Pharmacol **54**(6): 671-4.
Nagarkatti, R., *et al.* (2002). J Allergy Clin Immunol **110**(3): 410-2.
Nagarkatti, R., *et al.* (2004). Int Arch Allergy Immunol **134**(3): 206-12.
Nagpal, K., *et al.* (2005). J Allergy Clin Immunol **115**(3): 527-33.
Nakao, F., *et al.* (2001). J Allergy Clin Immunol **107**(3): 499-504.
Nakashima, K., *et al.* (2006). Biochem Biophys Res Commun **344**(1): 300-7.
Niimi, T., *et al.* (2002). Am J Hum Genet **70**(3): 718-25.
Noguchi, E., *et al.* (1998). Clin Exp Allergy**28**(4): 449-53.
Noguchi, E., *et al.* (2002). Clin Exp Allergy **32**(1): 93-6.
Noguchi, E., *et al.* (2005). Am J Respir Crit Care Med **172**(2): 183-8.
Noguchi, E., *et al.* (2006). Clin Exp Allergy **36**(5): 602-8.

Ober, C., *et al.* (2000). Am J Med Genet **66**(2): 517-26.
Ober, C., *et al.* (2008). N Engl J Med **358**(16): 1682-91.
Oguma, T., *et al.* (2004). N Engl J Med **351**(17): 1752-63.
Oh, S. H., *et al.* (2009). Respir Med **103**(7): 1020-4.
Palmer, L. J., *et al.* (1999). Clin Exp Allergy**29**(11): 1555-62.
Parapanissiou, E., *et al.* (2005). Tissue Antigens **65**(5): 481-4.
Park, B. L., *et al.* (2004). J Hum Genet **49**(10): 517-27.
Park, B. L., *et al.* (2010). BMB Rep **43**(6): 445-9.
Park, H. W., *et al.* (2007). Clin Exp Allergy **37**(11): 1609-15.
Pasaje, C. F., *et al.* (2010). Ann Hum Genet **74**(4): 326-34.
Pattaro, C., *et al.* (2006). Allergy **61**(2): 239-44.
Pillai, S. G., *et al.* (2004). Pharmacogenetics **14**(9): 627-33.
Polimanti, R., *et al.* (2010). Clin Exp Pharmacol Physiol **37**(8): 870-2.
Polonikov, A. V., *et al.* (2009). J Asthma **46**(5): 523-8.
Poon, A. H., *et al.* (2004). Am J Respir Crit Care Med **170**(9): 967-73.
Pulleyn, L. J., *et al.* (2001). Hum Genet **109**(6): 623-7.
Puthothu, B., *et al.* (2006). Genes Immun **7**(4): 322-6.
Puthothu, B., *et al.* (2009). Pediatr Allergy Immunol **20**(2): 157-63.
Qian, F. H., *et al.* (2010). J Investig Allergol Clin Immunol **20**(4): 340-6.
Raby, B. A., *et al.* (2004). Am J Respir Crit Care Med **170**(10): 1057-65.
Raby, B. A., *et al.* (2006). Am J Respir Crit Care Med **173**(1): 64-70.
Randolph, A. G., *et al.* (2004). Am J Hum Genet **75**(4): 709-15.
Rathcke, C. N., *et al.* (2009). PLoS One **4**(7): e6106.
Risma, K. A., *et al.* (2002). J Immunoly **169**(3): 1604-10.
Saadi, A., *et al.* (2009). BMC Med Genet **10**: 71.
Sakagami, T., *et al.* (2007). J Hum Genet **52**(1): 66-72.
Salam, M. T., *et al.* (2007). Am J Respir Crit Care Med **176**(12): 1192-9.
Salam, M. T., *et al.* (2009). J Allergy Clin Immunol **123**(3): 596-602, 602 e1-8.
Santillan, A. A., *et al.* (2003). J Allergy Clin Immunol **112**(6): 1095-100.
Sanz, C., *et al.* (2006). Allergy **61**(5): 543-8.
Schubert, K., *et al.* (2006). Dis Markers **22**(3): 127-32.
Sekigawa, T., *et al.* (2009). Clin Exp Allergy **39**(7): 972-81.
Settin, A., *et al.* (2008). Indian J Pediatr **75**(10): 1019-23.
Shao, C., *et al.* (2004). J Hum Genet **49**(3): 115-22.
Sharma, M., *et al.* (2004). Immunogenetics **56**(7): 544-7.
Sharma, M., *et al.* (2008). Am J Respir Crit Care Med **177**(7): 712-9.
Sharma, S., *et al.* (2004). Clin Genet **66**(5): 417-25.
Sharma, S. and Ghosh, B. (2004). J Hum Genet **49**(12): 677-83.
Sharma, S., *et al.* (2005). J Hum Genet **50**(6): 276-82.
Sharma, S., *et al.* (2006). Am J Respir Cell Mol Biol **35**(4): 488-95.
Sharma, S., *et al.* (2009). Eur Respir J **33**(6): 1287-94.
Sharma, S. and Ghosh, B. (2009). Int Arch Allergy Immunol **149**(3): 208-18.
Shin, E. K., *et al.* (2010). Hum Genet **128**(3): 293-302.
Shin, H. D., *et al.* (2004). Hum Mol Genet **13**(4): 397-403.
Shirakawa, T., *et al.* (1994). Nat Genet **7**(2): 125-9.
Shyur, S. D., *et al.* (2008). Eur J Hum Genet **16**(10): 1283-8.
Sleiman, P. M., *et al.* (2010). N Engl J Med **362**(1): 36-44.
Smit, L. A., *et al.* (2009). Am J Respir Crit Care Med **179**(5): 363-8.
Soriano, J. B., *et al.* (1997). Am J Respir Crit Care Med **156**(5): 1394-8.
Srivastava, N., *et al.* (1985). Hum Hered **35**(4): 263-4.
Srivastava, P., *et al.* (2003). Thorax **58**(3): 222-6.
Su, D., *et al.* (2008). BMC Med Genet **9**: 82.
Sun, X., *et al.* (2010). J Allergy Clin Immunol **126**(2): 241-9, 249 e1-3.
Suttner, K., *et al.* (2009). J Allergy Clin Immunol **123**(5): 1062-8, 1068 e1-8.
Szczeklik, W., *et al.* (2004). J Allergy Clin Immunol **114**(2): 248-53.
Szczepankiewicz, A., *et al.* (2008). Int Arch Allergy Immunol **145**(1): 43-7.
Szczepankiewicz, A., *et al.* (2009). J Appl Genet **50**(3): 275-81.
Takabayashi, A., *et al.* (2000). Exp Clin Immunogenet**17**(2): 63-70.
Takeoka, S., *et al.* (2001). J Hum Genet **46**(2): 57-63.
Tantisira, K., *et al.* (2004). Genes Immun **5**(5): 343-6.
Teerlink, C. C., *et al.* (2009). Eur J Hum Genet **17**(5): 636-43.
Tezcan, G., *et al.* (2009). J Asthma **46**(4): 371-4.
Thakkinstian, A., *et al.* (2005). Am J Epidemiol **162**(3): 201-11.

Thongngarm, T., *et al.* (2008). Asian Pac J Allergy Immunol **26**(4): 205-11.
Torio, A., *et al.* (2003). Hum Immunol **64**(8): 811-5.
Trajkov, D., *et al.* (2008). Iran J Allergy Asthma Immunol **7**(3): 143-56.
Tremblay, K., *et al.* (2006). Genes Immun **7**(8): 632-9.
Undarmaa, S., *et al.* (2010). J Hum Genet **55**(6): 342-9.
Unoki, M., *et al.* (2000). Hum Genet **106**(4): 440-6.
Urhan, M., *et al.* (2004). Allergy Asthma Proc **25**(4): 243-7.
Van Eerdewegh, P., *et al.* (2002). Nature **418**(6896): 426-30.
Vergara, C., *et al.* (2008). Am J Respir Crit Care Med **178**(10): 1017-22.
Vergara, C., *et al.* (2009). Clin Exp Allergy **39**(10): 1558-68.
Via, M., *et al.* (2010). Clin Exp Allergy **40**(4): 582-9.
Vonk, J. M., *et al.* (2010). Pharmacogenet Genomics **20**(3): 179-86.
Walsh, K. M., *et al.* (2010). Mutat Res **690**(1-2): 89-94.
Wang, J., *et al.* (2009). Mol Biol Rep **36**(6): 1549-53.
Wang, J. Y., *et al.* (2009). J Clin Immunol **29**(2): 205-9.
Wang, T. N., *et al.* (2006). Clin Exp Allergy **36**(9): 1147-52.
Weidinger, S., *et al.* (2005). Clin Exp Allergy **35**(7): 866-72.
Weidinger, S., *et al.* (2005). J Allergy Clin Immunol **116**(1): 177-84.
Weiss, L. A., *et al.* (2005). Am J Respir Crit Care Med **172**(1): 67-73.
Werner, M., *et al.* (2004). Clin Exp Allergy **34**(1): 26-31.
Wilson, N. M., *et al.* (2004). Pediatr Pulmonol **38**(1): 75-81.
Winchester, E. C., *et al.* (2000). Hum Genet **107**(6): 591-6.
Wjst, M. (2005). BMC Genet **6**: 2.
Wjst, M., *et al.* (2006). Respir Res **7**: 60.
Wu, H., *et al.* (2007). J Allergy Clin Immunol **120**(2): 322-8.
Wu, H., *et al.* (2009). Allergy **64**(4): 629-35.
Wu, H., *et al.* (2009). J Allergy Clin Immunol **125**(2): 321-327 e13.
Wu, J., *et al.* (2009). Tissue Antigens **73**(4): 302-6.
Wu, L. S., *et al.* (2006). Int Arch Allergy Immunol **141**(1): 37-46.
Wu, X., *et al.* (2010). J Asthma **47**(3): 238-44.
Xi, D., *et al.* (2004). J Huazhong Univ Sci Technolog Med Sci **24**(3): 219-22.
Yan, L., *et al.* (2000). Pharmacogenetics **10**(3): 261-6.
Yanamandra, K., *et al.* (2005). Biochem Biophys Res Commun **335**(2): 545-9.
Yao, T. C., *et al.* (2003). J Allergy Clin Immunol **111**(6): 1285-92.
Yokouchi, Y., *et al.* (2000). Genomics **66**(2): 152-60.
Zambelli-Weiner, A., *et al.* (2005). J Allergy Clin Immunol **115**(6): 1203-9.
Zedan, M., *et al.* (2008). Allergy Asthma Proc **29**(3): 268-73.
Zeyrek, D., *et al.* (2008). Allergy Asthma Proc **29**(5): 468-74.
Zhang, H., *et al.* (2007). Hum Mutat **28**(10): 1046.
Zhang, W., *et al.* (2007). Respir Med **101**(1): 186-90.
Zhang, X., *et al.* (2004). Ann Allergy Asthma Immunol **93**(5): 499-503.
Zhang, X., *et al.* (2009). Mol Biol Rep **36**(6): 1505-9.
Zhang, Y., *et al.* (2003). Nat Genet **34**(2): 181-6.
Zhang, Y., *et al.* (2010). Respirology **15**(4): 643-50.
Zhou, H., *et al.* (2009). BMC Med Genet **10**: 123.
Zhu, G., *et al.* (2007). Genes Immun **8**(5): 398-403.
Zhu, G., *et al.* (2008). Eur J Hum Genet **16**(9): 1083-90.
Zhu, G., *et al.* (2008). Genes Immun **9**(1): 23-9.

Appendix 7 Gel Eletrctrophoresis of *IL1RL1* and *IL33* Clones

Gel electrophoresis images showing clone inserts following EcoR1 plasmid digestion for *IL1RL1* [Figure A7.1] and *IL33* [Figure A7.2] PCR inserts.

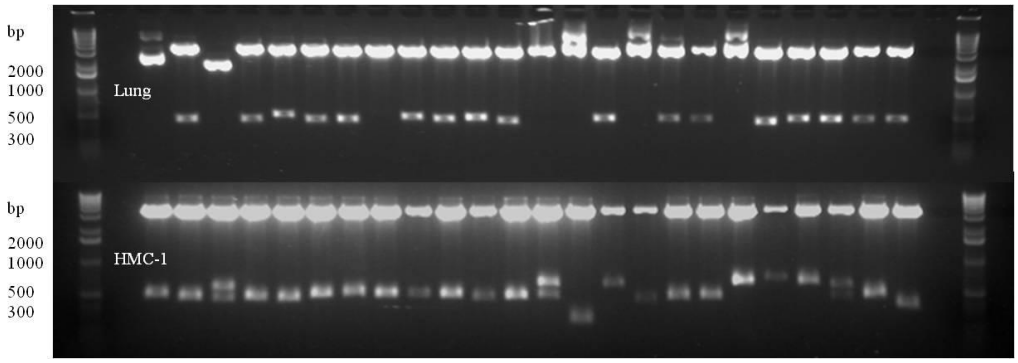


Figure A7.1 – 1% gel electrophoresis showing clone inserts following plasmid digestion using EcoR1 for *IL1RL1*. Inserts in lung and human mast cell line (HMC-1) are shown in the top and bottom rows respectively. 1kb DNA ladders are shown on the left and right extremes. Inserts were roughly the same size in lung, multiple insert sizes can be seen in HMC-1.

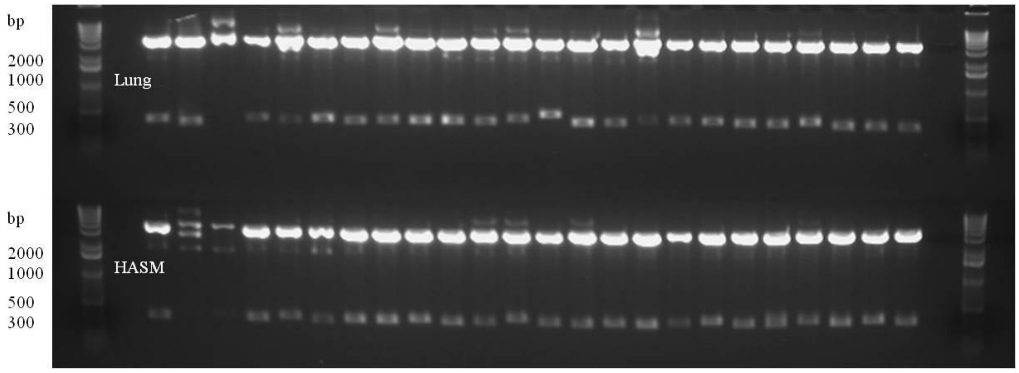


Figure A7.2 – 1% gel electrophoresis showing clone inserts following plasmid digestion using EcoR1 for *IL33*. Inserts in lung and human airway smooth muscle (HASM) are shown in the top and bottom rows respectively. All inserts were roughly the same size in both templates.

Appendix 8 In-Frame Amino Acid Sequences for *IL1RL1*

In frame amino acid sequences were delineated for the five transcripts identified for *IL1RL1* from 5' RACE assuming conservation of the remaining published 3' sequences across all detected transcripts. Transcripts 1 and 2 correspond to published transcripts NM_016232.4 and NM_003856.2 respectively. Transcripts 3, 4 and 5 are novel. The amino acid sequence coded by the open reading frame is in green, potential methyanine site are highlighted in red with the most likely site in yellow.

Transcript 1 (codes for the transmembrane form of the IL1RL1 receptor NP_057316.3)

```
5'3'
KRGWLLYLVKL-SCKRNWLSEL-NCGQKVEEERTQVQPNEVEI-ATLPNSVLKSITNCLM
CGDLHCRMPVTHLE-SQQRVTNTCS-LINRMGFWILAILTILMYSTAAKFSKQSWGLENE
ALIVRCPRQGKPSYTVDWYYSQTNKSIPTQERNRVFASGQLLKFLPAAVADSGIYTCIVR
SPTFNRTGYANVTIYKKQSDCNVPDYLMYSTVSGSEKNSKIYCPTIDLYNWTAPLEWFKN
CQALQGSRYRAHKSFLVIDNVMTEDAGDYTCKFIHNENGANYSVTATRSFTVKDEQGFSL
FPVIGAPAQNEIKEVEIGKNANLTCSACFGKGTQFLAAVLWQLNGTKITDFGEPRIQQEE
GQNQSFSNGLACLDMVLRIADVKEEDLLLQYDCLALNLHGLRRHTVRLSRKNPIDHHSIY
CIIAVCSVFLMLINVLVIILKMFWIEATLLWRDIAKPYKTRNDGKLYDAYVVYPRNYKSS
TDGASRVEHFVHQILPDVLENKCGYTLCIYGRDMLPGEDVVTAVETNIRKSRRHIFILTP
QITHNKEFAYEQEVALHCALIQNDAKVILIEMEALSELDMLQAEALQDSLQHLMKVQGTI
KWREDHIANKRSLNSKFWKHVRYQMPVPSKIPRKASSLTPLAAQKQ-CLL-CAKASEFEA
FLTSPSWLMPLH-SVRSRNIKGIQA
```

Transcript 2 (codes for the soluble form of the IL1RL1 receptor NP_003847.2)

```
5'3' Frame 3
GGTYKDWKLFLAPSLTPSSSPLSFSLVEI-ATLPNSVLKSITNCLMCGDLHCRMPVTHLE
-SQQRVTNTCS-LINRMGFWILAILTILMYSTAAKFSKQSWGLENEALIVRCPRQGKPSY
TVDWYYSQTNKSIPTQERNRVFASGQLLKFLPAAVADSGIYTCIVRSPTFNRTGYANVTI
YKKQSDCNVPDYLMYSTVSGSEKNSKIYCPTIDLYNWTAPLEWFKNCQALQGSRYRAHKS
FLVIDNVMTEDAGDYTCKFIHNENGANYSVTATRSFTVKDEQGFSLFPVIGAPAQNEIKE
VEIGKNANLTCSACFGKGTQFLAAVLWQLNGTKITDFGEPRIQQEEGQNQSFSNGLACLD
MVLRIADVKEEDLLLQYDCLALNLHGLRRHTVRLSRKNPSKECF-DFDHLNFL-QV-AEW
SVVPRDPSRQWEWPVP-NVLLFFGMLFAV-SL-TVPVCWELLCCLNCSSSPTPSYRWFV-
NTQLLLWSSLFSNFMNSLCVTVCERKCTNNRKLNVFFCALL-LALHVVSMVRSIPFSGHN
EHSFC-RGW-SEQKGEVSNYCHFSEKILGATL--DIC-AILALI-RNT-DWVIYMKRGLI
GSQFCRLYGKHGGICFWGHLRSFTHGRRQSKGRHFTQ-KQERERGATLKQPDLMRSHSLL
QGQHQRDGAKPFMMNSPP-SNHLPPGSTSNTGDYHSA-DLGRNTDPNHTTHIIIVKLCKV
FKVHGTHGKSGSSAHFFIASVIHHVIQVPRIPGSLSWPSVCSIHTFQVLL-HPVCIIALV
TLPYLNLFDSLLNTTASSMRAGTSSLPSLGP-CNTWQLASAQLNIC-LNK-MHNQKKKKK
KKKKKK
```

Transcript 3 (results in a shorter protein truncated at the 5’ end with a loss of 117 amino acids)

5'3'
KRGWLLYLVKL-SCKRNWLSEL-NCGQKVEEERTQVQPNEGKQSWGLENEALIVRCPRQG
KPSYTVDWYYSQTNKSIPTQERNRVFASGQLLKFLPAAVADSGIYTCIVRSPTFNRTGYA
NVTIYKKQSDCNVPDYLMYSTVSGSEKNSKIYCPTIDLYNWTAPLEWFKNCQALQGSRYR
AHKSFLVIDNVMTEDAGDYTCKFIHNENGANYSVTATRSFTVKDEQGFSLFPVIGAPAQN
EIKEVEIGKNANLTCSACFGKGTQFLAAVLWQLNGTKITDFGEPRIQQEEGQNQSFSNGL
ACLDMVLRIADVKEEDLLLQYDCLALNLHGLRRHTVRLSRKNPIDHHSIYCIIAVCSVFL
MLINVLVIILKMFWIEATLLWRDIAKPYKTRNDGKLYDAYVVYPRNYKSSTDGASRVEHF
VHQILPDVLENKCGYTLCIYGRDMLPGEDVVTAVETNIRKSRRHIFILTPQITHNKEFAY
EQEVALHCALIQNDAKVILIEMEALSELDMLQAEALQDSLQHLMKVQGTIKWREDHIANK
RSLNSKFWKHVRYQMPVPSKIPRKASSLTPLAAQKQ-CLL-CAKASEFEAFLTSPSWLMP
LH-SVRSRNIKGIQA

Transcript 4 (results in the transmembrane form of the IL1RL1 receptor NP_057316.3)

5'3' Frame 3
REAGCCI--SYKAVREIGFLSCETVGRKLRKKELKYNPMRA-GNRRHVSVWL-QFSY-IL
RPNYSHAIVPEYHCGHNHLILAIPCRVEI-ATLPNSVLKSITNCLMCGDLHCRMPVTHLE
-SQQRVTNTCS-LINRMGFWILAILTILMYSTAAKFSKQSWGLENEALIVRCPRQGKPSY
TVDWYYSQTNKSIPTQERNRVFASGQLLKFLPAAVADSGIYTCIVRSPTFNRTGYANVTI
YKKQSDCNVPDYLMYSTVSGSEKNSKIYCPTIDLYNWTAPLEWFKNCQALQGSRYRAHKS
FLVIDNVMTEDAGDYTCKFIHNENGANYSVTATRSFTVKDEQGFSLFPVIGAPAQNEIKE
VEIGKNANLTCSACFGKGTQFLAAVLWQLNGTKITDFGEPRIQQEEGQNQSFSNGLACLD
MVLRIADVKEEDLLLQYDCLALNLHGLRRHTVRLSRKNPIDHHSIYCIIAVCSVFLMLIN
VLVIILKMFWIEATLLWRDIAKPYKTRNDGKLYDAYVVYPRNYKSSTDGASRVEHFVHQI
LPDVLENKCGYTLCIYGRDMLPGEDVVTAVETNIRKSRRHIFILTPQITHNKEFAYEQEV
ALHCALIQNDAKVILIEMEALSELDMLQAEALQDSLQHLMKVQGTIKWREDHIANKRSLN
SKFWKHVRYQMPVPSKIPRKASSLTPLAAQKQ-CLL-CAKASEFEAFLTSPSWLMPLH-S
VRSRNIKGIQA

Transcript 5 (results in the transmembrane form of the IL1RL1 receptor NP_057316.3)

5'3' Frame 1
KERLAVVFSKAIKL-EKLAF-VVKLWAES-GRKNSSTTQ-GHKEIGGMFLSGYNSSPTRS
-GQTIHMPLFLNTIVGIII-FWLFHAEICCCDC-LECYGIPSRIWDSLCS-MRCLVEI-A
TLPNSVLKSITNCLMCGDLHCRMPVTHLE-SQQRVTNTCS-LINRMGFWILAILTILMYS
TAAKFSKQSWGLENEALIVRCPRQGKPSYTVDWYYSQTNKSIPTQERNRVFASGQLLKFL
PAAVADSGIYTCIVRSPTFNRTGYANVTIYKKQSDCNVPDYLMYSTVSGSEKNSKIYCPT
IDLYNWTAPLEWFKNCQALQGSRYRAHKSFLVIDNVMTEDAGDYTCKFIHNENGANYSVT
ATRSFTVKDEQGFSLFPVIGAPAQNEIKEVEIGKNANLTCSACFGKGTQFLAAVLWQLNG
TKITDFGEPRIQQEEGQNQSFSNGLACLDMVLRIADVKEEDLLLQYDCLALNLHGLRRHT
VRLSRKNPIDHHSIYCIIAVCSVFLMLINVLVIILKMFWIEATLLWRDIAKPYKTRNDGK
LYDAYVVYPRNYKSSTDGASRVEHFVHQILPDVLENKCGYTLCIYGRDMLPGEDVVTAVE
TNIRKSRRHIFILTPQITHNKEFAYEQEVALHCALIQNDAKVILIEMEALSELDMLQAEA
LQDSLQHLMKVQGTIKWREDHIANKRSLNSKFWKHVRYQMPVPSKIPRKASSLTPLAAQK
Q-CLL-CAKASEFEAFLTSPSWLMPLH-SVRSRNIKGIQA

Bibliography

1000 Genomes Project Consortium, *et al.* (2010). "A map of human genome variation from population-scale sequencing." Nat Genet **467**: 1061-1073

Aas, K. (1978). "What makes an allergen an allergen." Allergy **33**(1): 3-14.

Ali, M., *et al.* (2009). "Investigations into the role of ST2 in acute asthma in children." Tissue Antigens **73**(3): 206-212.

Allakhverdi, Z., *et al.* (2007). "Cutting edge: The ST2 ligand IL-33 potently activates and drives maturation of human mast cells." J Immunol **179**(4): 2051-2054.

Anderson, C. A., *et al.* (2011). "Synthetic associations are unlikely to account for many common disease genome-wide association signals." PLoS Biol **9**(1): e1000580.

Anderson, G. G., *et al.* (2002). "Positive association to IgE levels and a physical map of the 13q14 atopy locus." Eur J Hum Genet **10**(4): 266-270.

Aron, Y., *et al.* (1996). "Evidence of a strong, positive association between atopy and the HLA class II alleles DR4 and DR7." Clin Exp Allergy **26**(7): 821-828.

Asimit, J. and Zeggini, E. (2010). "Rare variant association analysis methods for complex traits. Review." Annu Rev Genet **44**: 293-308.

ATS (2000). "Proceedings of the ATS Workshop on Refractory Asthma: Current Understanding, Recommendations, and Unanswered Questions." Am J Respir Crit Care Med **162**(6): 2341-2351.

Aulchenko, Y., *et al.* (2010). "ProbABEL package for genome-wide association analysis of imputed data." BMC Bioinformatics **11**(134).

Bacharier, L. B., *et al.* (2008). "Episodic use of an inhaled corticosteroid or leukotriene receptor antagonist in preschool children with moderate-to-severe intermittent wheezing." J Allergy Clin Immunol **122**(6): 1127-1135 e1128.

Bajekal, M. and Prescott, A. (2003). Respiratory symptoms, atopic conditions and lung function. Health Survey for England 2001, The Stationery Office, London.

Barnes, K. C. (1999). "Gene-environment and gene-gene interaction studies in the molecular genetic analysis of asthma and atopy." Clin Exp Allergy **29**(Suppl 4): 47-51.

Barnes, K. C. (2000). "Evidence for common genetic elements in allergic disease." J Allergy Clin Immunol **106**(Suppl 5): S192-200.

Barnes, K. C., *et al.* (2007). "African Americans with asthma: genetic insights. Review." Proc Am Thorac Soc **4**(1): 58-68.

Barnes, P. J. and Adcock, I. M. (1998). "Transcription factors and asthma." Eur Respir J **12**(1): 221-234.

Barnes, P. J. and Barnes, P. J. (2006). "Drugs for asthma." Br J Pharmacol **147** (Suppl 1): S297-303.

Baron, M. (2001). "The search for complex disease genes: fault by linkage or fault by association?" Mol Psychiatry **6**(2): 143-149.

Barrett, J. C. and Cardon, L. R. (2006). "Evaluating coverage of genome-wide association studies." Nat Genet **38**(6): 659-662.

Barrett, J. C., *et al.* (2009). "Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes." Nat Genet **41**(6): 703-707.

Barrett, J. C., *et al.* (2008). "Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease." Nat Genet **40**(8): 955-962.

Bateman, E. D., *et al.* (2008). "Global strategy for asthma management and prevention: GINA executive summary." Eur Respir J **31**(1): 143-178.

Bauchau, V. and Durham, S. R. (2004). "Prevalence and rate of diagnosis of allergic rhinitis in Europe." Eur Respir J **24**(5): 758-764.

Benayoun, L., *et al.* (2003). "Airway structural alterations selectively associated with severe asthma." Am J Respir Crit Care Med **167**(10): 1360-1368.

Bentley, D. R., *et al.* (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." Nature **456**(7218): 53-59.

Bhalla, K., *et al.* (2008). "Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability." Am J Hum Genet **83**(6): 703-713.

Bieli, C., *et al.* (2007). "A polymorphism in CD14 modifies the effect of farm milk consumption on allergic diseases and CD14 gene expression." J Allergy Clin Immunol **120**(6): 1308-1315.

Bjornvold, M., *et al.* (2009). "A TLR2 polymorphism is associated with type 1 diabetes and allergic asthma." Genes Immun **10**(2): 181-187.
Blumenthal, M. N., *et al.* (2004). "A genome-wide search for allergic response (atopy) genes in three ethnic groups: Collaborative Study on the Genetics of Asthma." Hum Genet **114**(2): 157-164.
Bodmer, W. and Bonilla, C. (2008). "Common and rare variants in multifactorial susceptibility to common diseases." Nat Genet **40**(6): 695-701.
Bousquet, J., *et al.* (2008). "Allergic Rhinitis and its Impact on Asthma (ARIA) 2008 update (in collaboration with the World Health Organization, GA(2)LEN and AllerGen)." Allergy **63**(Suppl 86): 8-160.
Bouzigon, E., *et al.* (2010). "Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy." Eur J Hum Genet **18**(6): 700-706.
Boyano Martinez, T., *et al.* (2001). "Validity of specific IgE antibodies in children with egg allergy." Clin Exp Allergy **31**(9): 1464-1469.
Bradley, M., *et al.* (2002). "Susceptibility loci for atopic dermatitis on chromosomes 3, 13, 15, 17 and 18 in a Swedish population." Hum Mol Genet **11**(13): 1539-1548.
Brasch-Andersen, C., *et al.* (2004). "Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers." Hum Mutat **24**(3): 208-214.
Brasch-Andersen, C., *et al.* (2006). "Highly significant linkage to chromosome 3q13.31 for rhinitis and related allergic diseases." J Med Genet **43**(3): e10.
Breathnach, R., *et al.* (1978). "Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries." Proc Natl Acad Sci U S A **75**(10): 4853-4857.
Breslow, D. K., *et al.* (2010). "Orm family proteins mediate sphingolipid homeostasis." Nature **463**(7284): 1048-1053.
Bryant, D. H., *et al.* (1975). "The correlation between skin tests, bronchial provocation tests and the serum level of IgE specific for common allergens in patients with asthma." Clin Allergy **5**(2): 145-157.
Burchard, E. G., *et al.* (1999). "Association between a sequence variant in the IL-4 gene promoter and FEV(1) in asthma." Am J Respir Crit Care Med **160**(3): 919-922.
Burrows, B., *et al.* (1989). "Association of asthma with serum IgE levels and skin-test reactivity to allergens." N Engl J Med **320**(5): 271-277.
Butterfield, J. H., *et al.* (1988). "Establishment of an immature mast cell line from a patient with mast cell leukemia." Leuk Res **12**(4): 345-355.
Cantero-Recasens, G., *et al.* (2010). "The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress." Hum Mol Genet **19**(1): 111-121.
Cardon, L. R. and Bell, J. I. (2001). "Association study designs for complex diseases." Nat Rev Genet **2**(2): 91-99.
Carlsen, K. H., *et al.* (1984). "Respiratory virus infections and aeroallergens in acute bronchial asthma." Arch Dis Child **59**(4): 310-315.
Carriere, V., *et al.* (2007). "IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo." Proc Natl Acad Sci U S A **104**: 282-287.
Carroll, W. (2005). "Asthma genetics: pitfalls and triumphs." Paediatr Respir Rev **6**(1): 68-74.
Castano, R., *et al.* (2009). "Evidence of association of interleukin-1 receptor-like 1 gene polymorphisms with chronic rhinosinusitis." Am J Rhinol Allergy **23**(4): 377-384.
Chanock, S. J., *et al.* (2007). "Replicating genotype-phenotype associations." Nature **447**(7145): 655-660.
Chen, J. J., *et al.* (2006). "The gene for a rare autosomal dominant form of pompholyx maps to chromosome 18q22.1-18q22.3." J Invest Dermatol **126**(2): 300-304.
Cherry, W. B., *et al.* (2008). "A novel IL-1 family cytokine, IL-33, potently activates human eosinophils." J Allergy Clin Immunol **121**: 1484-1490.
Chung, K. F., *et al.* (1999). "Difficult/therapy-resistant asthma: the need for an integrated approach to define clinical phenotypes, evaluate risk factors, understand pathophysiology and find novel therapies. ERS Task Force on Difficult/Therapy-Resistant Asthma. European Respiratory Society." Eur Respir J **13**(5): 1198-1208.
Ciullo, M., *et al.* (2006). "New susceptibility locus for hypertension on chromosome 8q by efficient pedigree-breaking in an Italian isolate." Hum Mol Genet **15**(10): 1735-1743.

Coca, A. F. and Cooke, R. A. (1923). "On the classification of the phenomena of hypersensitiveness." J Immunol (8): 163-182.
Cohen, J. C., *et al.* (2004). "Multiple rare alleles contribute to low plasma levels of HDL cholesterol." Science **305**(5685): 869-872.
Cooke, R. A. and Veer, A. V. J. (1916). "Human sensitization." J Immunol(1): 201-305.
Cookson, W. O., *et al.* (1989). "Linkage between immunoglobulin E responses underlying asthma and rhinitis and chromosome 11q." Lancet **1**(8650): 1292-1295.
Cookson, W. O., *et al.* (2001). "Genetic linkage of childhood atopic dermatitis to psoriasis susceptibility loci." Nat Genet **27**(4): 372-373.
Corder, E. H., *et al.* (1994). "Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease." Nat Genet **7**(2): 180-184.
Coventry, A., *et al.* (2010). "Deep resequencing reveals excess rare recent variants consistent with explosive population growth." Nat Commun **1**(131): 1-6.
Craig, D. W., *et al.* (2008). "Identification of genetic variants using bar-coded multiplexed sequencing." Nat Methods **5**(10): 887-893.
Crawford, D. C., *et al.* (2004). "Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations." Am J Hum Genet **74**(4): 610-622.
D'Amato, G., *et al.* (2005). "Environmental risk factors and allergic bronchial asthma." Clin Exp Allergy **35**(9): 1113-1124.
Dahl, M., *et al.* (1998). "DeltaF508 heterozygosity in cystic fibrosis and susceptibility to asthma." Lancet **351**(9120): 1911-1913.
Dantzker, D. R., *et al.* (1987). "Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, November 1986." Am Rev Respir Dis **136**(1): 225-244.
Darsow, U., *et al.* (1995). "Atopy patch test with different vehicles and allergen concentrations: an approach to standardization." J Allergy Clin Immunol **95**(3): 677-684.
Day-Williams, A. G. and Zeggini, E. (2011). "The effect of next-generation sequencing technology on complex trait research." Eur J Clin Invest **41**(5): 561-567.
Daykin, K., *et al.* (1993). "Control of histamine induced inositol phospholipid hydrolysis in cultured human tracheal smooth muscle cells." Eur J Pharmacol **246**(2): 135-140.
de Cid, R., *et al.* (2009). "Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis." Nat Genet **41**: 211-215
Denham, S., *et al.* (2008). "Meta-analysis of genome-wide linkage studies of asthma and related traits." Respir Res **9**: 38.
Dickson, S. P., *et al.* (2010). "Rare variants create synthetic genome-wide associations." PLoS Biol **8**(1): e1000294.
Dizier, M. H., *et al.* (2000). "Genome screen for asthma and related phenotypes in the French EGEA study." Am J Respir Crit Care Med **162**(5): 1812-1818.
Dizier, M. H., *et al.* (1999). "Indication of linkage of serum IgE levels to the interleukin-4 gene and exclusion of the contribution of the (-590 C to T) interleukin-4 promoter polymorphism to IgE variation." Genet Epidemiol **16**(1): 84-94.
Dold, S., *et al.* (1992). "Genetic risk for asthma, allergic rhinitis, and atopic dermatitis." Arch Dis Child **67**(8): 1018-1022.
Drazen, J. M., *et al.* (2000). "Heterogeneity of therapeutic responses in asthma." Br Med Bull **56**(4): 1054-1070.
Duffy, D. L., *et al.* (1990). "Genetics of asthma and hay fever in Australian twins." Am Rev Respir Dis **142**(6 Pt 1): 1351-1358.
Eder, W., *et al.* (2005). "Opposite effects of CD 14/-260 on serum IgE levels in children raised in different environments." J Allergy Clin Immunol **116**(3): 601-607.
Eder, W., *et al.* (2004). "Toll-like receptor 2 as a major gene for asthma in children of European farmers." J Allergy Clin Immunol **113**(3): 482-488.
Emilsson, V., *et al.* (2008). "Genetics of gene expression and its effect on disease." Nature **452**(7186): 423-428.
Esparza-Gordillo J, W. S., Fölster-Holst R Anja Bauerfeind, Franz Ruschendorf, Giannino Patone, Klaus Rohde, Ingo Marenholz, Florian Schulz, Tamara Kerscher, Norbert Hubner, Ulrich Wahn, Stefan Schreiber, Andre Franke, Rainer Vogler, Simon Heath, Hansjörg Baurecht, Natalija Novak, Elke Rodriguez, Thomas Illig, Min-Ae Lee-Kirsch, Andrzej Ciechanowicz, Michael Kurek, Tereza Piskackova, Milan Macek,

Young-Ae Lee & Andreas Ruether (2009). "A common variant on chromosome 11q13 is associated with atopic dermatitis." Nature Genetics (41): 596 - 601.
Fearnhead, N. S., *et al.* (2004). "Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas." Proc Natl Acad Sci U S A **101**(45): 15992-15997.
Feijen, M., *et al.* (2000). "Genetics of allergic disease." Br Med Bull **56**(4): 894-907.
Ferreira, M. A., *et al.* (2011). "Association between ORMDL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia." Eur J Hum Genet **19**(4): 458-464.
Ferreira, M. A., *et al.* (2005). "Robust estimation of experimentwise P values applied to a genome scan of multiple asthma traits identifies a new region of significant linkage on chromosome 20q13." Am J Respir Crit Care Med **77**(6): 1075-1085.
Feulner, T. M., *et al.* (2009). "Examination of the current top candidate genes for AD in a genome-wide association study." Mol Psychiatry **15**(7): 756-766.
Flores, C., *et al.* (2007). "A variant of the myosin light chain kinase gene is associated with severe asthma in African Americans." Genet Epidemiol **31**(4): 296-305.
Flory, J. H., *et al.* (2009). "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 **124**(3): 605-607.
Fukai, H., *et al.* (2004). "Association between a polymorphism in cysteinyl leukotriene receptor 2 on chromosome 13q14 and atopic asthma." Pharmacogenetics **14**(10): 683-690.
Fukao, T., *et al.* (2000). "Synergistic effects of IL-4 and IL-18 on IL-12-dependent IFN-gamma production by dendritic cells." J Immunol **164**(1): 64-71.
Gabriel, S. B., *et al.* (2002). "The structure of haplotype blocks in the human genome." Science **296**(5576): 2225-2229.
Gächter, T., *et al.* (1998). "GATA-Dependent expression of the interleukin-1 receptor-related T1 gene in mast cells." Mol Cell Biol **18**(9): 5320-5331.
Galanter, J., *et al.* (2008). "ORMDL3 gene is associated with asthma in three ethnically diverse populations." Am J Respir Crit Care Med **177**(11): 1194-1200.
Galli, S. J., *et al.* (2008). "The development of allergic inflammation." Nature **454**(7203): 445-454.
Gern, J. E., *et al.* (2004). "Effects of dog ownership and genotype on immune development and atopy in infancy." J Allergy Clin Immunol **113**(2): 307-314.
Global Initiative for Asthma (2008). "Global Strategy for Asthma Management and Prevention (GINA)." NIH Consens Statement; NHLBI: [Available at http://www.ginasthma.org].
Godfrey, R. C. and Gradidge, C. F. (1976). "Allergic sensitisation of human lung fragments prevented by saturation of IgE binding sites." Nat Genet **259**: 484-486.
Goldstein, D. B. (2009). "Common genetic variation and human traits." N Engl J Med **360**(17): 1696-1698.
Gorlov, I. P., *et al.* (2008). "Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms." Am J Hum Genet **82**(1): 100-112.
Gruber, J. D., *et al.* (2002). "Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing." Hum Genet **110**(5): 395-401.
Gudbjartsson, D. F., *et al.* (2009). "Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction." Nat Genet **41**(3): 342-347.
Guo, L., *et al.* (2009). "IL-1 family members and STAT activators induce cytokine production by Th2, Th17, and Th1 cells." Proc Natl Acad Sci U S A **106**(32): 13463-13468.
Gupta, R., *et al.* (2004). "Burden of allergic disease in the UK: secondary analyses of national databases." Clin Exp Allergy **34**(4): 520-526.
Gupta, R., *et al.* (2007). "Time trends in allergic disorders in the UK." Thorax **62**(1): 91-96.
Gusareva, E. S., *et al.* (2009). "Mouse to human comparative genetics reveals a novel immunoglobulin E-controlling locus on Hsa8q12." Immunogenetics **61**(1): 15-25.
Haagerup, A., *et al.* (2004). "Fine-scale mapping of type I allergy candidate loci suggests central susceptibility genes on chromosomes 3q, 4q and Xp." Allergy **59**(1): 88-94.
Hakonarson, H., *et al.* (2001). "Allelic frequencies and patterns of single-nucleotide polymorphisms in candidate genes for asthma and atopy in Iceland.[see comment]." Am J Respir Crit Care Med **164**(11): 2036-2044.
Halapi, E., *et al.* (2010). "A sequence variant on 17q21 is associated with age at onset and severity of asthma." Eur J Hum Genet **18**: 902-908.

Haldar, P., *et al.* (2008). "Cluster analysis and clinical asthma phenotypes." Am J Respir Crit Care Med **178**(3): 218-224.
Hall, I. P. and Sayers, I. (2007). "Pharmacogenetics and asthma: false hope or new dawn?" Eur Respir J **29**(6): 1239-1245.
Hall, J. M., *et al.* (1990). "Linkage of early-onset familial breast cancer to chromosome 17q21." Science **250**(4988): 1684-1689.
Hanifin, J. M. (2009). "Evolving concepts of pathogenesis in atopic dermatitis and other eczemas." J Invest Dermatol **129**(2): 320-322.
Hao, K., *et al.* (2008). "Calibrating the performance of SNP arrays for whole-genome association studies." PLoS Genet **4**(6): e1000109.
Harismendy, O. and Frazer, K. (2009). "Method for improving sequence coverage uniformity of targeted genomic intervals amplified by LR-PCR using Illumina GA sequencing-by-synthesis technology." Biotechniques **46**(3): 229-231.
He, J.-Q., *et al.* (2009). "A thymic stromal lymphopoietin gene variant is associated with asthma and airway hyperresponsiveness." J Allergy Clin Immunol **124**(2): 222-229.
Headley, M. B., *et al.* (2009). "TSLP conditions the lung immune environment for the generation of pathogenic innate and antigen-specific adaptive immune responses." J Immunol **182**(3): 1641-1647.
Heinzmann, A., *et al.* (2000). "Genetic variants of IL-13 signalling and human asthma and atopy." Hum Mol Genet **9**(4): 549-559.
Higa, S., *et al.* (2003). "Association between interleukin-18 gene polymorphism 105A/C and asthma." Clin Exp Allergy **33**(8): 1097-1102.
Hill, W. G., *et al.* (2008). "Data and theory point to mainly additive genetic variance for complex traits. Review." PLoS Genet **4**(2): e1000008.
Hirschhorn, J. N. (2009). "Genomewide association studies--illuminating biologic pathways." N Engl J Med **360**(17): 1699-1701.
Hirschhorn, J. N., *et al.* (2005). "Genome-wide association studies for common diseases and complex traits." Nat Rev Genet **6**(2): 95-108.
Hoffjan, S., *et al.* (2005). "The genetics of atopic dermatitis: recent findings and future options." J Mol Med **83**(9): 682-692.
Hogg, J. C. (1997). "The pathology of asthma." APMIS **105**(10): 735-745.
Hoggart, C. J., *et al.* (2008). "Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies." PLoS Genet **4**(7): e1000130.
Holgate, A. T. and Church, M. K. (1993). Allergic disease. Allergy, Gower Medical Publishing, Brookfield.
Holgate, S. T. (2000). "Inflammatory and structural changes in the airways of patients with asthma." Respir Med **94**(Suppl D): S3-6.
Holgate, S. T., *et al.* (2011). Asthma. WAO White Book on Allergy, World Allergy Organisation, Wisconsin.
Holgate, S. T., *et al.* (2005). "Anti-immunoglobulin E treatment with omalizumab in allergic diseases: an update on anti-inflammatory activity and clinical efficacy." Clin Exp Allergy **35**(4): 408-416.
Holgate, S. T., *et al.* (2008). "Treatment strategies for allergy and asthma." Nat Rev Immunol **8**(3): 218-230.
Holloway, J. W., *et al.* (2010). "Atopy susceptibility and chromosome 19q13." Allergy **65**(6): 802-804.
Holloway, J. W., *et al.* (1999). "The genetic basis of atopic asthma." Clin Exp Allergy **29**(8): 1023-1032.
Holloway, J. W., *et al.* (1996). "Lack of evidence of a significant association between HLA-DR, DQ and DP genotypes and atopy in families with HDM allergy." Clin Exp Allergy **26**(10): 1142-1149.
Hopp, R. J., *et al.* (1984). "Specificity and sensitivity of methacholine inhalation challenge in normal and asthmatic children." J Allergy Clin Immunol **74**(2): 154-158.
Howard, T. D., *et al.* (2002). "Gene-gene interaction in asthma: IL4RA and IL13 in a Dutch population with asthma." Am J Hum Genet **70**(1): 230-236.
Howie, B. N., *et al.* (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies." PLoS Genet **6**(e1000529).
Hüffmeier, U., *et al.* (2010). "Replication of LCE3C-LCE3B CNV as a risk factor for psoriasis and analysis of interaction with other genetic risk factors." J Invest Dermatol **130**(4): 979-984.

Hunt, K. A. and al., e. (2008). "Newly identified genetic risk variants for celiac disease related to the immune response." Nat Genet **40**: 395-402.
Imboden, M., *et al.* (2006). "The common G-allele of interleukin-18 single-nucleotide polymorphism is a genetic risk factor for atopic asthma. The SAPALDIA Cohort Study." Clin Exp Allergy **36**(2): 211-218.
Israel, E., *et al.* (2001). "Effect of polymorphism of the beta(2)-adrenergic receptor on response to regular use of albuterol in asthma." Int Arch Allergy Immunol **124**(1-3): 183-186.
Ivaschenko, T. E., *et al.* (2002). "Glutathione- S-transferase micro and theta gene polymorphisms as new risk factors of atopic bronchial asthma." J Mol Med **80**(1): 39-43.
Iwahana, H., *et al.* (1999). "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 **264**(2): 397-406.
Jackola, D. R., *et al.* (2002). "Variable binding affinities for allergen suggest a 'selective competition' among immunoglobulins in atopic and non-atopic humans." Mol Immunol **39**(5-6): 367-377.
James, A. L., *et al.* (2010). "Changes in the prevalence of asthma in adults since 1966: the Busselton health study." Eur Respir J **35**(2): 273-278.
Janeway, C. A., *et al.* (2001). Innate Immunity. Immunobiology, Garland Science, New York and London.
Jang, N., *et al.* (2005). "Polymorphisms within the PHF11 gene at chromosome 13q14 are associated with childhood atopic dermatitis." Genes Immun **6**(3): 262-264.
Ji, W., *et al.* (2008). "Rare independent mutations in renal salt handling genes contribute to blood pressure variation." Nat Genet **40**(5): 592-599.
Johansen, C. T., *et al.* (2010). "Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia." Nat Genet **42**(8): 684-687.
Johansson, H., *et al.* (2011). "Targeted resequencing of candidate genes using selector probes." Nucleic Acids Res **39**(2): e8.
Johansson, S. G. O., *et al.* (2004). "Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization." J Allergy Clin Immunol **113**(5): 832-836.
Jongepier, H., *et al.* (2004). "Polymorphisms of the ADAM33 gene are associated with accelerated lung function decline in asthma." Clin Exp Allergy **34**(5): 757-760.
Junqueira, L. C. and Carneiro, J. (2009). Asthma. Basic Histology, a text and atlas, McGraw-Hill Medical, New York**:** p. 358, Figure 317-313.
Kakkar, R. and Lee, R. T. (2008). "The IL-33/ST2 pathway: therapeutic target and novel biomarker." Nat Rev Drug Discov **7**(10): 827-840.
Kato, A. and Schleimer, R. P. (2007). "Beyond inflammation: airway epithelial cells are at the interface of innate and adaptive immunity." Curr Opin Immunol **19**(6): 711-720.
Kauffmann, F., *et al.* (2002). "Epidemiologic study of the genetics and environment of asthma, bronchial hyperresponsiveness, and atopy." Chest **121**(Suppl 3): 27S.
Kay, A. B. (2001). "Allergy and allergic diseases (part 1)." N Engl J Med **344**(1): 30-37.
Kenny, E. M., *et al.* (2011). "Multiplex target enrichment using DNA indexing for ultra-high throughput SNP detection." DNA Res **18**(1): 31-38.
Klein, R. J., *et al.* (2005). "Complement factor H polymorphism in age-related macular degeneration." Science **308**(5720): 385-389.
Koppelman, G. H., *et al.* (2002). "Genome-wide search for atopy susceptibility genes in Dutch families with asthma." J Allergy Clin Immunol **109**(3): 498-506.
Kowalewska, J., *et al.* (2007). "Thymic stromal lymphopoietin transgenic mice develop cryoglobulinemia and hepatitis with similarities to human hepatitis C liver disease." Am J Pathol **170**(3): 981-989.
Kumar, A. and Ghosh, B. (2008). "A single nucleotide polymorphism (A --> G) in intron 3 of IFNgamma gene is associated with asthma." Genes Immun **9**(4): 294-301.
Kurz, T., *et al.* (2005). "A genome-wide screen on the genetics of atopy in a multiethnic European population reveals a major atopy locus on chromosome 3q21.3." Allergy **60**(2): 192-199.
Kuster, W., *et al.* (1990). "A family study of atopic dermatitis. Clinical and genetic characteristics of 188 patients and 2,151 family members." Arch Dermatol Res **282**(2): 98-102.

Kwok, P. Y., *et al.* (1994). "Comparative Analysis of Human DNA Variations by Fluorescence-Based Sequencing of PCR Products." Genomics **23**(1): 138-144.
Lachheb, J., *et al.* (2008). "Promoter polymorphism of the IL-18 gene is associated with atopic asthma in Tunisian children." Int J Immunogenet **35**(1): 63-68.
Larsen, F. S., *et al.* (1986). "Atopic dermatitis. A genetic-epidemiologic study in a population-based twin sample." J Am Acad Dermatol **15**(3): 487-494.
Lee, H. M., *et al.* (2006). "Interleukin-18/-607 gene polymorphism in allergic rhinitis." Int J Pediatr Otorhinolaryngol **70**(6): 1085-1088.
Lee, Y. A., *et al.* (2000). "A major susceptibility locus for atopic dermatitis maps to chromosome 3q21." Nat Genet **26**(4): 470-473.
Lee, Y. M., *et al.* (2010). "Cell cycle-regulated expression and subcellular localization of a kinesin-8 member human KIF18B." Gene **466**(1-2): 16-25.
Lemanske, R. F., Jr., *et al.* (2003). "6. Asthma." J Allergy Clin Immunol **111**(2 Suppl): S502-519.
Lesueur, F., *et al.* (2007). "Confirmation of psoriasis susceptibility loci on chromosome 6p21 and 20p13 in French families." J Invest Dermatol **127**(6): 1403-1409.
Lettre, G. and Rioux, J. D. (2008). "Autoimmune diseases: insights from genome-wide association studies. Review." Hum Mol Genet **17**(R2): R116-121.
Leung, D. Y., *et al.* (2003). "Atopic dermatitis." Lancet **361**(9352): 151-160.
Levy, S., *et al.* (2007). "The diploid genome sequence of an individual human." Plos Biology **5**(10): e254.
Lewontin, R. C. (1964). "The Interaction of Selection and Linkage. Ii. Optimum Models." Genetics **50**: 757-782.
Leynaert, B., *et al.* (2008). "Association between farm exposure and atopy, according to the CD14 C-159T polymorphism." J Allergy Clin Immunol **118**(3): 658-665.
Li, B. and Leal, S. M. (2008). "Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data." Am J Hum Genet **83**(3): 311-321.
Li, X., *et al.* (2010). "Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions." J Allergy Clin Immunol **125**(2): 328-335.e311.
Li, Y. and Abecasis, G. R. (2006). "Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference." Am J Hum Genet **S79**(2290).
Liu, C. K., *et al.* (2008). "Functional analysis of novel SNPs and mutations in human and mouse genomes." BMC Bioinformatics **9**(Suppl 12): S10.
Los, H., *et al.* (1999). "The importance of genetic influences in asthma." Eur Respir J **14**(5): 1210-1227.
Lucas, S. R., *et al.* (2005). "Physical activity and exercise in asthma: relevance to etiology and treatment." J Allergy Clin Immunol **115**(5): 928-934.
Luczynska, C. M., *et al.* (1989). "A two-site monoclonal antibody ELISA for the quantification of the major Dermatophagoides spp. allergens, Der p I and Der f I." J Immunol Methods **118**(2): 227-235.
Maher, B. (2008). "Personal genomes: The case of the missing heritability." Nature **456**(7218): 18-21.
Maier, L. M., *et al.* (2006). "Association of IL13 with total IgE: evidence against an inverse association of atopy and diabetes." J Allergy Clin Immunol **117**(6): 1306-1313.
Mak, J. C., *et al.* (2006). "Analysis of TGF-beta(1) gene polymorphisms in Hong Kong Chinese patients with asthma." J Allergy Clin Immunol **117** 92-96.
Mamanova, L., *et al.* (2010). "Target-enrichment strategies for next-generation sequencing. Review." Nat Methods **7**(2): 111-118.
Manning, A. K., *et al.* (2011). "Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients." Genet Epidemiol **35**(1): 11-18.
Manolio, T. A., *et al.* (2008). "A HapMap harvest of insights into the genetics of common disease. Review." J Clin Invest **118**(5): 1590-1605.
Manolio, T. A., *et al.* (2009). "Finding the missing heritability of complex diseases." Nat Genet **461**(7265): 747-753.
Mao, M., *et al.* (2004). "T lymphocyte activation gene identification by coregulated expression on DNA microarrays." Genomics **83**(6): 989-999.
Marsh, D. G. (1996). "Mapping the genes for IgE production and allergy." Adv Exp Med Biol **409**: 43-53.

Marsh, D. G., *et al.* (1982). "HLA-Dw2: a genetic marker for human immune response to short ragweed pollen allergen Ra5. I. Response resulting primarily from natural antigenic exposure." J Exp Med **155**(5): 1439-1451.
Marth, G., *et al.* (2001). "Single-nucleotide polymorphisms in the public domain: how useful are they?" Nat Genet **27**(4): 371-372.
Martinez, F. D. (2005). "Gene-environment interactions in asthma and allergies: a new paradigm to understand disease causation." Immunol Allergy Clin North Am **25**(4): 709-721.
Martinez, F. D. (2007). "Genes, environments, development and asthma: a reappraisal." Eur Respir J **29**(1): 179-184.
Masoli, M., *et al.* (2004). "The global burden of asthma: executive summary of the GINA Dissemination Committee report." Allergy **59**(5): 469-478.
Matsuda, A., *et al.* (2010). "Functional role of thymic stromal lymphopoietin in chronic allergic keratoconjunctivitis." Invest Ophthalmol Vis Sci **51**(1): 151-155.
McCarthy, M. I., *et al.* (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." Nat Rev Genet **9**(5): 356-369.
McKeever, T. M., *et al.* (2002). "Prospective study of diet and decline in lung function in a general population." Am J Respir Crit Care Med **165**(9): 1299-1303.
Metzker, M. L. (2010). "Sequencing technologies - the next generation. Review." Nat Rev Genet **11**(1): 31-46.
Meyers, D. A., *et al.* (2005). "Genome screen for asthma and bronchial hyperresponsiveness: interactions with passive smoke exposure." J Allergy Clin Immunol **115**(6): 1169-1175.
Meyers, D. A., *et al.* (1995). "Two locus segregation and linkage analysis for total serum IgE levels." Clin Exp Allergy **25**(Suppl 2): 113-115.
Miller, R. G. (1981). Simultaneous statistical inference. Simultaneous statistical inference, Springer Verlag, London**:** 6-8.
Moffatt, M. F. and Cookson, W. O. C. M. (1997). "Linkage and candidate gene studies in asthma." Am J Respir Crit Care Med **150**: S110-112.
Moffatt, M. F., *et al.* (2010). "A large-scale, consortium-based genomewide association study of asthma." N Engl J Med **363**(13): 1211-1221.
Moffatt, M. F., *et al.* (2007). "Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma." Nature **448**(7152): 470-473.
Moore, W. C., *et al.* (2010). "Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program." Am J Respir Crit Care Med **181**(4): 315-323.
Moritz, D. R., *et al.* (1998). "The IL-1 receptor-related T1 antigen is expressed on immature and mature mast cells and on fetal blood mast cell progenitors." J Immunol **161**(9): 4866-4874.
Morris, A. P., *et al.* (2009). "Identification of novel putative rheumatoid arthritis susceptibility genes via analysis of rare variants." BMC Proc **3**(Suppl 7): S131.
Morton, N. E. (1955). "Sequential tests for the detection of linkage." Am J Hum Genet **7**(3): 277-318.
Moussion, C., *et al.* (2008). "The IL-1-like cytokine IL-33 is constitutively expressed in the nucleus of endothelial cells and epithelial cells in vivo: a novel 'alarmin'?" PLoS One **3**(10): e3331.
Murcray, C. E., *et al.* (2009). "Gene-environment interaction in genome-wide association studies." Am J Epidemiol. **169**(2): 219-226.
Nagarkatti, R., *et al.* (2002). "Association of IFNG gene polymorphism with asthma in the Indian population." J Allergy Clin Immunol **110**(3): 410-412.
Nakao, F., *et al.* (2001). "Association of IFN-gamma and IFN regulatory factor 1 polymorphisms with childhood atopic asthma." J Allergy Clin Immunol **107**(3): 499-504.
National Asthma Education and Prevention Program (2007). "Expert panel report 3: guidelines for the diagnosis and management of asthma." NIH Consens Statement; NHLBI: [Available from http://www.nhlbi.nih.gov/guidelines/asthma].
Neale, B. M., *et al.* (2011). "Testing for an unusual distribution of rare variants." PLoS Genet **7**(3): e1001322.
Nicolae, D., *et al.* (2005). "Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21." Am J Hum Genet **76**(2): 349-357.

Norman, P. S., *et al.* (1973). "Diagnostic tests in ragweed hay fever. A comparison of direct skin tests, IgE antibody measurements, and basophil histamine release." J Allergy Clin Immunol **52**(4): 210-224.
Nyholt, D. R. (2004). "A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other." Am J Hum Genet **74**(4): 765-769.
Ober, C. and Hoffjan, S. (2006). "Asthma genetics 2006: the long and winding road to gene discovery." Genes Immun **7**(2): 95-100.
Ober, C., *et al.* (2008). "Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function." N Engl J Med **358**(16): 1682-1691.
Ober, C., *et al.* (1999). "Genome-wide screen for atopy susceptibility alleles in the Hutterites." Clin Exp Allergy **29**(Suppl 4): 11-15.
Ober, C. and Vercelli, D. (2011). "Gene-environment interactions in human disease: nuisance or opportunity?" Trends Genet **27**(3): 107-115.
Obholz, K. L., *et al.* (2006). "FNDC3A is required for adhesion between spermatids and Sertoli cells." Dev Biol **298**(2): 498-513.
Ortolani, C., *et al.* (1989). "Comparison of results of skin prick tests (with fresh foods and commercial food extracts) and RAST in 100 patients with oral allergy syndrome." J Allergy Clin Immunol **83**(3): 683-690.
Oshikawa, K., *et al.* (2001). "Elevated Soluble ST2 Protein Levels in Sera of Patients with Asthma with an Acute Exacerbation." Am J Respir Crit Care Med **164**(2): 277-281.
Ownby, D. R., *et al.* (2002). "Exposure to dogs and cats in the first year of life and risk of allergic sensitization at 6 to 7 years of age." JAMA **288**(8): 963-972.
Palmer, C. N., *et al.* (2006). "Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis." Nat Genet **38**(4): 441-446.
Palmer, L. J., *et al.* (2000). "Independent inheritance of serum immunoglobulin E concentrations and airway responsiveness." Am J Respir Crit Care Med **161**(6): 1836-1843.
Park, J. H., *et al.* (2010). "Estimation of effect size distribution from genome-wide association studies and implications for future discoveries." Nat Genet **42**(7): 570-575.
Park, L. S., *et al.* (2000). "Cloning of the murine thymic stromal lymphopoietin (TSLP) receptor: Formation of a functional heteromeric complex requires interleukin 7 receptor." J Exp Med **192**(5): 659-670.
Pattaro, C., *et al.* (2006). "Association between interleukin-1 receptor antagonist gene and asthma-related traits in a German adult population." Allergy **61**(2): 239-244.
Patterson, N., *et al.* (2006). "Population structure and eigenanalysis." PLoS Genet **2**(12): e190.
Perkin, M. R., *et al.* (2006). "The predictive value of early life total immunoglobulin E measurement in identifying atopic children in a population-based birth cohort study." Pediatr Allergy Immunol **17**(2): 118-124.
Petretto, E., *et al.* (2007). "A gene harvest revealing the archeology and complexity of human disease." Nat Genet **39**(11): 1299-1301.
Postma, D. S., *et al.* (2011). "Asthma and COPD: Common Genes, Common Environments?" Am J Respir Crit Care Med **183**(12): 1588-1594.
Power, C., *et al.* (2006). "Cohort profile: 1958 British birth cohort (National Child Development Study)." Int J Epidemiol **35**(1): 34-41.
Préfontaine, D., *et al.* (2009). "Increased expression of IL-33 in severe asthma: evidence of expression by airway smooth muscle cells." J Immunol **183**(8): 5094-5103.
Price, A. L., *et al.* (2006). "Principal components analysis corrects for stratification in genome-wide association studies." Nat Genet **38**(8): 904-909.
Pritchard, J. K. (2001). "Are rare variants responsible for susceptibility to complex diseases?" Am J Hum Genet **69**(1): 124-137.
Pruim, R. J., *et al.* "LocusZoom: regional visualization of genome-wide association scan results." Bioinformatics **26**(18): 2336-2337.
Pulleyn, L. J., *et al.* (2001). "TGFbeta1 allele association with asthma severity." Hum Genet **109**(6): 623-627.
Purcell, S., *et al.* (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Qian, F. H., *et al.* "Polymorphisms in the toll-like receptor 2 subfamily and risk of asthma: a case-control analysis in a Chinese population." J Investig Allergol Clin Immunol **20**(4): 340-346.

Reddel, R. R., *et al.* (1988). "Transformation of human bronchial epithelial cells by infection with SV40 or adenovirus-12 SV40 hybrid virus, or transfection via strontium phosphate coprecipitation with a plasmid containing SV40 early region genes." Cancer Res **48**(7): 1904-1909.

Reijmerink, N. E., *et al.* (2010). "TLR-related pathway analysis: novel gene-gene interactions in the development of asthma and atopy." Allergy **65**(2): 199-207.

Reijmerink, N. E., *et al.* (2008). "Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy." J Allergy Clin Immunol **122**(3): 651-654.

Repapi, E., *et al.* (2010). "Genome-wide association study identifies five loci associated with lung function Source." Nat Genet **42**(36-44): 1061-4036.

Rice, P., *et al.* (2000). "EMBOSS: the European Molecular Biology Open Software Suite." Trends Genet **16**(6): 276-277.

Risch, N. (1990). "Linkage strategies for genetically complex traits. II. The power of affected relative pairs." Am J Hum Genet **46**(2): 229-241.

Risch, N. and Merikangas, K. (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-1517.

Roberts, G., *et al.* (2005). "Relationship between aeroallergen and food allergen sensitization in childhood." Clin Exp Allergy **35**(7): 933-940.

Robinson, C., *et al.* (1997). "On the potential significance of the enzymatic activity of mite allergens to immunogenicity. Clues to structure and function revealed by molecular characterization." Clin Exp Allergy **27**(1): 10-21.

Roche, W. R. (1998). "Inflammatory and structural changes in the small airways in bronchial asthma." Am J Respir Crit Care Med **157**(5 Pt 2): S191-194.

Rogers, A. J., *et al.* (2009). "Assessing the reproducibility of asthma candidate gene associations, using genome-wide data." Am J Respir Crit Care Med **179**(12): 1084-1090.

Rosa-Rosa, L., *et al.* (1999). "The R576 IL-4 receptor alpha allele correlates with asthma severity." J Allergy Clin Immunol **104**(5): 1008-1014.

Sachidanandam, R., *et al.* (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature **409**(6822): 928-933.

Saetta, M. and Turato, G. (2001). "Airway pathology in asthma." Eur Respir J **34**: S18-23.

Sakashita, M., *et al.* (2008). "Association of serum interleukin-33 level and the interleukin-33 genetic variant with Japanese cedar pollinosis." Clin Exp Allergy **38**(12): 1875-1881.

Sanak, M., *et al.* (2007). "Genetic variability of the high-affinity IgE receptor alpha subunit (Fc epsilon RI alpha) is related to total serum IgE levels in allergic subjects." Allergol Int **56**(4): 397-401.

Sandford, A. J., *et al.* (2000). "Polymorphisms in the IL4, IL4RA, and FCERIB genes and asthma severity." J Allergy Clin Immunol **106**(1 Pt 1): 135-140.

Sandford, A. J., *et al.* (1998). "Phenotypic heterogeneity in asthma and allergy." Clin Exp Allergy **Suppl 1**(26-31): 32-36.

Sandilands, A., *et al.* (2007). "Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema." Nat Genet **39**(5): 650-654.

Savenije, O. E., *et al.* (2011). "Interleukin-1 receptor-like 1 polymorphisms are associated with serum IL1RL1-a, eosinophils, and asthma in childhood." J Allergy Clin Immunol **127**(3): 750-756 e755.

Sayers, I., *et al.* (2003). "Allelic association and functional studies of promoter polymorphism in the leukotriene C4 synthase gene (LTC4S) in asthma." Thorax **58**(5): 417-424.

Scadding, G. K., *et al.* (2008). "BSACI guidelines for the management of allergic and non-allergic rhinitis." Clin Exp Allergy **38**(1): 19-42.

Schedel, M., *et al.* (2004). "A signal transducer and activator of transcription 6 haplotype influences the regulation of serum IgE levels." J Allergy Clin Immunol **114**(5): 1100-1105.

Schmitz, J., *et al.* (2005). "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 **23**(5): 479-490.

Schork, N. J., *et al.* (2009). "Common vs. rare allele hypotheses for complex diseases." Curr Opin Genet Dev **19**(3): 212-219.
Schork, N. J., *et al.* (2008). "DNA sequence-based phenotypic association analysis." Adv Genet **60**: 195-217.
Schultz Larsen, F. V. and Holm, N. V. (1985). "Atopic dermatitis in a population based twin series. Concordance rates and heritability estimation." Acta Derm Venereol Suppl (Stockh) **114**: 159.
Settin, A., *et al.* (2008). "Gene polymorphisms of IL-6(-174) G/C and IL-1Ra VNTR in asthmatic children." Indian J Pediatr **75**(10): 1019-1023.
Shaaban, R., *et al.* (2008). "Rhinitis and onset of asthma: a longitudinal population-based study." Lancet **372**(9643): 1049-1057.
Sham, P., *et al.* (2002). "DNA Pooling: a tool for large-scale association studies." Nat Rev Genet **3**(11): 862-871.
Shendure, J. and Ji, H. (2008). "Next-generation DNA sequencing." Nat Biotechnol **10**: 1135-1145.
Shimizu, M., *et al.* (2005). "Functional SNPs in the distal promoter of the ST2 gene are associated with atopic dermatitis." Hum Mol Genet **14**: 2919–2927.
Shin, J., *et al.* (2007). "Identification of phosphotyrosine binding domain-containing proteins as novel downstream targets of the EphA8 signaling function." Mol Cell Biol **27**(23): 8113-8126.
Sibbald, B. and Turner-Warwick, M. (1979). "Factors influencing the prevalence of asthma among first degree relatives of extrinsic and intrinsic asthmatics." Thorax **34**(3): 332-337.
Simpson, A. and Custovic, A. (2004). "The role of allergen avoidance in primary and secondary prevention." Pediatr Pulmonol Suppl **26**: 225-228.
Simpson, A., *et al.* (2006). "Endotoxin exposure, CD14, and allergic disease: an interaction between genes and the environment." Am J Respir Crit Care Med **174**: 386-392.
Singh, P., *et al.* (2011). "Target-Enrichment Through Amplification of Hairpin-Ligated Universal Targets for Next-Generation Sequencing Analysis." Methods Mol Biol **733**: 267-278.
Sironi, M. and Clerici, M. (2010). "The hygiene hypothesis: an evolutionary perspective. Review." Microbes Infect **12**(6): 421-427.
Skol, A. D., *et al.* (2006). "Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies." Nat Genet **38**: 209-213.
Skoner, D. P. (2001). "Allergic rhinitis: definition, epidemiology, pathophysiology, detection, and diagnosis." J Allergy Clin Immunol **108**(Suppl 1): S2-8.
Smale, S. T. (1997). "Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes." Biochim Biophys Acta **1351**(1-2): 73-88.
Smit, L. A. M., *et al.* (2009). "CD14 and toll-like receptor gene polymorphisms, country living, and asthma in adults." Am J Respir Crit Care Med **179**(5): 363-368.
Smith, D. E. (2010). "IL-33: a tissue derived cytokine pathway involved in allergic inflammation and asthma." Clin Exp Allergy **40**(2): 200-208.
Smyth, D. J., *et al.* (2006). "A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region." Nat Genet **38**(6): 617-619.
Spergel, J. M. (2010). "From atopic dermatitis to asthma: the atopic march." Ann Allergy Asthma Immunol **105**(2): 99-106; quiz 107-109.
Stanojevic, S., *et al.* (2008). "Reference ranges for spirometry across all ages: a new approach." Am J Respir Crit Care Med **177**(3): 253-260.
Stenius, B., *et al.* (1971). "Clinical significance of specific IgE to common allergens. I. Relationship of specific IgE against Dermatophagoides spp. and grass pollen to skin and nasal tests and history." Clin Allergy **1**(1): 37-55.
Stöhr, H., *et al.* (2002). "A novel gene encoding a putative transmembrane protein with two extracellular CUB domains and a low-density lipoprotein class A module: isolation of alternatively spliced isoforms in retina and brain." Genes Immun **286**(2): 223-231.
Stratton, M. (2008). "Genome resequencing and genetic variation. Review. ." Nat Biotechnol **26**(1): 65-66.
Stunkard, A. J., *et al.* (1990). "The body-mass index of twins who have been reared apart." N Engl J Med **322**(21): 1483-1487.

Su, A. I., *et al.* (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." Proc Natl Acad Sci U S A **101**(16): 6062-6067.
Sunyer, J., *et al.* (1995). "Relationship between serum IgE and airway responsiveness in adults with asthma." J Allergy Clin Immunol **95**(3): 699-706.
Szalai, C., *et al.* (2001). "Polymorphism in the gene regulatory region of MCP-1 is associated with asthma susceptibility and severity." J Allergy Clin Immunol **108**: 375-381.
Tago, K., *et al.* (2001). "Tissue distribution and subcellular localization of a variant form of the human ST2 gene product, ST2V." Biochem Biophys Res Commun **285**(5): 1377-1383.
Taillon-Miller, P., *et al.* (2000). "Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28." Nat Genet **25**(3): 324-328.
Tanaka, K., *et al.* (2002). "Genetic variants of the receptors for thromboxane A2 and IL-4 in atopic dermatitis." Biochem Biophys Res Commun **292**(3): 776-780.
Taneda, S., *et al.* (2001). "Cryoglobulinemic glomerulonephritis in thymic stromal lymphopoietin transgenic mice." Am J Pathol **159**(6): 2355-2369.
Ten, R. M., *et al.* (1995). "Allergy skin testing." Mayo Clin Proc **70**(8): 783-784.
The ENFUMOSA Study Group (2003). "The ENFUMOSA cross-sectional European multicentre study of the clinical phenotype of chronic severe asthma. European Network for Understanding Mechanisms of Severe Asthma." Eur Respir J **22**(3): 470-477.
The International HapMap, C. (2003). "The International HapMap Project." Nature **426**(6968): 789-796.
Tominaga, S., *et al.* (1999). "Presence and expression of a novel variant form of ST2 gene product in human leukemic cell line UT-7/GM." Biochem Biophys Res Commun **264**(1): 14-18.
Torgerson, D. G., *et al.* (2011). "Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations." Nat Genet **43**(9): 887-892.
Tremblay, K., *et al.* (2008). "Genes to diseases (G2D) computational method to identify asthma candidate genes." PLoS One **3**(8): e2907.
Tsuchiya, S., *et al.* (1980). "Establishment and characterization of a human acute monocytic leukemia cell line (THP-1)." Int J Cancer **26**(2): 171-176.
Ueno, H., *et al.* (2003). "A stromal cell-derived membrane protein that supports hematopoietic stem cells." Nature Immunol **4**(5): 457-463.
Valacer, D. J. (2000). "Childhood asthma: causes, epidemiological factors and complications." Drugs **59**(Suppl 1): 1-8; discussion 43-45.
Van Den Berg, T. K., *et al.* (2001). "Cutting edge: CD43 functions as a T cell counterreceptor for the macrophage adhesion receptor sialoadhesin (Siglec-1)." J Immunol **166**(6): 3637-3640.
Van Eerdewegh, P., *et al.* (2002). "Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness." Nature **418**(6896): 426-430.
Venter, J. C., *et al.* (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.
Visscher, P. M., *et al.* (2008). "Heritability in the genomics era-concepts and misconceptions." Nat Rev Genet **9**: 255-266.
Von Pirquet, C. (1946). "Allergie." Ann Allergy **4**(5): 388-390.
Wang, J., *et al.* (2008). "The diploid genome sequence of an Asian individual." Nature **456**(7218): 60-65.
Wang, T. N., *et al.* (2006). "Association of interferon-gamma and interferon regulatory factor 1 polymorphisms with asthma in a family-based association study in Taiwan." Clin Exp Allergy **36**(9): 1147-1152.
Weidinger, S., *et al.* (2008). "Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus." PLoS Genet **4**(8): e1000166.
Weidinger, S., *et al.* (2006). "Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations." J Allergy Clin Immunol **118**(1): 214-219.
Weinmayr, G., *et al.* (2007). "Atopic sensitization and the international variation of asthma symptom prevalence in children." Am J Respir Crit Care Med **176**(6): 565-574.
Wellcome Trust Case Control Consortium (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.

Wenzel, S. E. (2005). "Severe asthma in adults." Exp Lung Res **31 Suppl 1**: 22.
Wenzel, S. E., *et al.* (1999). "Evidence that severe asthma can be divided pathologically into two inflammatory subtypes with distinct physiologic and clinical characteristics." Am J Respir Crit Care Med **160**(3): 1001-1008.
Wheeler, D. A., *et al.* (2008). "The complete genome of an individual by massively parallel DNA sequencing." Nature **452**(7189): 872-876.
Williams, H., *et al.* (1999). "Worldwide variations in the prevalence of symptoms of atopic eczema in the International Study of Asthma and Allergies in Childhood." J Allergy Clin Immunol **103**(1 Pt 1): 125-138.
Williams, L. K., *et al.* (2008). "Gene-environment interactions between CD14 C-260T and endotoxin exposure on Foxp3+ and Foxp3- CD4+ lymphocyte numbers and total serum IgE levels in early childhood." Ann Allergy Asthma Immunol **100**(2): 128-136.
Wills-Karp, M., *et al.* (2004). "Time to draw breath: asthma-susceptibility genes are identified." Nat Rev Genet **5**(5): 376-387.
Witteman, A. M., *et al.* (1996). "The relationship between RAST and skin test results in patients with asthma or rhinitis: a quantitative study with purified major allergens." J Allergy Clin Immunol **97**(1 Pt 1): 16-25.
Wright, A. F., *et al.* (1999). "Population choice in mapping genes for complex diseases." Nat Genet **23**(4): 397-404.
Wu, A. C., *et al.* (2010). "Development of a Pharmacogenetic Predictive Test in asthma: proof of concept." Pharmacogenet Genomics **20**(2): 86-93.
Wu, H., *et al.* (2010). "Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans." J Allergy Clin Immunol **125**(2): 321-327.e313.
Yang, Y., *et al.* (1999). "Epigenetic mechanisms silence a disintegrin and metalloprotease 33 expression in bronchial epithelial cells." J Allergy Clin Immunol **121**(6): 1393-1399.
Ying, S., *et al.* (2005). "Thymic stromal lymphopoietin expression is increased in asthmatic airways and correlates with expression of Th2-attracting chemokines and disease severity." J Immunol **174**(12): 8183-8190.
Yoo, J., *et al.* (2005). "Spontaneous atopic dermatitis in mice expressing an inducible thymic stromal lymphopoietin transgene specifically in the skin." J Exp Med **202**(4): 541-549.
Young, R. P., *et al.* (1994). "HLA-DR and HLA-DP genotypes and immunoglobulin E responses to common major allergens." Clin Exp Allergy **24**(5): 431-439.
Zambelli-Weiner, A., *et al.* (2005). "Evaluation of the CD14/-260 polymorphism and house dust endotoxin exposure in the Barbados Asthma Genetics Study." J Allergy Clin Immunol **115**(6): 1203-1209.
Zeyrek, D., *et al.* (2008). "Association of interleukin-1beta and interleukin-1 receptor antagonist gene polymorphisms in Turkish children with atopic asthma." Allergy Asthma Proc **29**(5): 468-474.
Zhang, M. Q. (1998). "Statistical features of human exons and their flanking regions." Hum Mol Genet **7**(5): 919-932.
Zheng, W., *et al.* (2006). "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 **1758**(12): 1864-1884.
Zhernakova, A., *et al.* (2008). "Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP." Am J Hum Genet **82**(5): 1202-1210.
Zhou, B., *et al.* (2005). "Thymic stromal lymphopoietin as a key initiator of allergic airway inflammation in mice." Nature Immunol **6**(10): 1047-1053.
Zhu, G., *et al.* (2008). "Interleukin 18 receptor 1 gene polymorphisms are associated with asthma." Eur J Hum Genet **16**(9): 1083-1090.
Ziegler, S. F. and Liu, Y.-J. (2006). "Thymic stromal lymphopoietin in normal and pathogenic T cell development and function." Nature Immunol **7**(7): 709-714.