Fakultät für Medizin der Technischen Universität München

Genetics of complex lung disorders - new approaches in asthma and COPD

Muralidharan Sargurupremraj

Vollständiger Abdruck der von der promotionsführenden Einrichtung

Fakultät für Medizin

der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors Dr. rer.nat. : Doktor der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Prof. Dr. Juliane Winkelmann

Prüfer der Dissertation:

1. apl. Prof. Dr. Bertram Müller-Myhsok
2. apl. Prof. Dr. Johannes Beckers

Die Dissertation wurde am 29.08.2016 bei der Technischen Universität München eingereicht und durch die promotionsführende Einrichtung Fakultät für Medizin am 15.03.2017 angenommen.

Table of contents

Table of contents

Abbreviations

LD	Linkage disequilibrium
AHR	Airway hyper-responsiveness
BHR	Bronchial hyper-responsiveness
GWLA	Genome-wide linkage analysis
GWAS	Genome-wide association study
eQTL	expression quantitative trait loci
ER	Endoplasmic reticulum
WGS	Whole genome sequencing
WES	Whole exome sequencing
CNV	Copy number variant
MAF	Minor allele frequency
IPA	Ingenuity pathway analysis
PCR	Polymerase chain reaction
cM	Cenitmorgan
IL	Interleukin
SPT	Skin prick test
SNP	Single nucleotide polymorphism
LINE-1/L1	Long interspersed nuclear elements - 1
SV	Structural variants
TNF	Tumour necrosis factor
Mb	Mega base-pairs
GTEx	Genotype-tissue expression
GRASP	Genome-wide repository of associations between SNPs and Phenotypes
GSDMA	Gasdermin A
ORMDL3	Orosomucoid like 3
IFN	Interferon
OR	Odds ratio
MANOVA	Multiple analysis of variance
LOD	Logarithm of odds
H3K4Me1	Histone 3 lysine 4 mono-methylation
H3K27ac	Histone 3 lysine 27 acetylation

Summary

One of the main hallmark features of asthma is airway hyper-responsiveness (AHR), and strong correlation exists between AHR and the inflammatory processes, suggesting the involvement of a network of cytokines. Other salient features include the IgE production, eosinophil infiltration and the shift of the fine balance between the different T - helper cells (Th1, Th2) towards a Th2 bias. Th2 cells occupy a central stage in the regulation of inflammatory cytokines and secrete a panel of them with several overlapping functions, through which it orchestrates the allergen-induced responses. Pro-inflammatory cytokines such as TNFα, IL-1b, IL-6, IL-8, and IL-33 being the characteristic molecular signatures found in the bronchoalveolar lavage and sputum of asthmatics, are receiving increased attention for their role in the maintenance of the inflammatory cascade.

The main aim of the thesis is to investigate such circulating (serum) cytokines as intermediary phenotypes and at the application of genetic approaches to yield insights into the association of underlying constitutional host immunity and genetic component in a family based setting. Serum samples of 923 individuals coming from 218 families were measured for 17 cytokines (Eotaxin, GM-CSF, IFNγ, IL-1b, IL-1RA, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12(p40), IL-13, IL-17, IL-23, IL-33, TSLP, and TNFα). The determination of heritable components in the variation of serum cytokine levels showed pro-inflammatory cytokines: TNFα and IL-8 as the most heritable with tight genetic control. In the genome-wide linkage analysis, one genomic region, 17q22-24, exceeded the genome-wide significance threshold for the linkage with anti-inflammatory cytokine levels of IL-12(p40). Furthermore, common variants (Minor Allele Frequency; MAF > 5%) from a set of loci implicated in published population-based association studies of asthma phenotypes, were tested for their strength of association with cytokine levels using the family-based association tests for quantitative traits (QFAM). Some of the key findings are the significant association of SNPs residing in *IL1RN* , *NR3C2* and *GSDMA* with cytokines involved in the Th2-mediated immune response. Significantly associated variants are followed up and annotated using a custom pipeline consisting of publicly available functional datasets on gene expression, regulatory networks and association catalogue. Interestingly, high-LD SNPs of the intronic variants residing in *IL1RN* and *NR3C2* are

associated with differential expression of the genes in blood cells and immune cells, and significantly enriched for histone modification marks associated with promoter and enhancer activation in multiple cell types including the lung. Lookup of significantly associated variants from the QFAM test, in the meta-analysis summary statistics of a consortium based (GABRIEL) largest Genome-Wide Association Study (GWAS) of asthmatic individuals, shows the *GSDMA* variant to be genome-wide associated with asthma. The present work mainly argues that the intermediary phenotypes such as the serum levels of cytokines can function as an alternative to the classical clinical phenotypes, especially when the clinical criteria are ambiguous.

It is well-known that common variants contribute modestly to the phenotype through their moderate effect sizes and subtle regulatory effects, and explain only a small proportion of the heritability. This forms one of the outlook of this thesis, as low frequency variants (MAF < 1%) in the summary data from the GABRIEL study was explored, which led to the identification of a larger proportion of associated SNPs with larger effect estimates. Given the rare nature of the identified variants, further explorations with greater sample size are warranted. Additionally, factors like structural variants and gene-environment interactions that might explain some of the missingness are explored and discussed in a disease type with strong environmental component - Chronic Obstructive Pulmonary Disease (COPD).

1. Introduction:

The World Health Organisation (WHO) estimates the number of asthmatic individuals to be steadily increasing, from approximately 150 million people a decade back to about 334 million people today [1], making this non-communicable disease the most common and serious chronic lung disease. The estimates are considered to be in the lower end of the spectrum as a result of the stringent criteria applied in asthma diagnosis. All ethnic backgrounds suffer from asthma, with developed countries suffering more than the developing nations. Asthma prevalence in U.S is higher compared to the other countries in the world and varies drastically among different ethnic populations [2]. One common pattern in asthma prevalence is its higher distribution in children than adults, and a possible explanation for this high prevalence in children is atopic sensitization [3]. Additionally, it is also shown that lifestyle changes can contribute as a major risk factor in asthma frequency, as the transition from a rural community to urbanised area is associated with increased prevalence of asthma [4]. With a marked increase in the world population and urbanisation, a further increase in asthmatics is expected over the next two decades. Recent assessments show that asthma accounts for 1 in every 250 deaths worldwide, with 250,000 annual deaths attributed to the disease [5]. Cost effective management strategies that are proven to reduce morbidity and mortality is an effective substitute until there is a greater understanding of causality in asthma.

1.1. Analytical studies in genetic epidemiology (Asthma)

From hallmark discoveries such as Mendel's law of segregation and the proposal of the double helix model of the DNA by Watson and Crick, immense progress has been made in the field of genetics and in understanding their role in disease development. In contrast to monogenic disorders like Cystic Fibrosis (CF) where a single gene explains the major part of the observed phenotype, asthma is rather ambiguous in establishing the relationship between cause and effect. It does not follow the classical Mendelian patterns owing to its heterogeneous nature. Attributable to the large effect size of the variants and their high penetrance, single-gene disorders like CF are less prevalent (1 in 2000–3000 live births) [6] compared

to counterparts like asthma that are more common. The prevalence of asthma and its severity varies significantly as a result of its multifactorial nature that includes both genetic and environmental factors; asthma is thus rightly termed as a common but complex genetic disorder. The genetic complexity of asthma is best evident from characteristic features such as the involvement of multiple genes (gene sets) in the pathogenesis (polygenic), different gene sets contributing to inter-individual variability in susceptibility (genetic heterogeneity), and the possibility of unaffected individuals carrying the same disease genotype (low disease penetrance). Twin studies and family-based studies have already established a strong genetic component about asthma onset [7]. This clustering in families is also evident from segregation studies; one such study by Meyers and co-workers showed that two loci in the 5q region combinatorially explained 87.4% of the variation in total serum IgE levels [8]. However, it has now become clear that no one genetic region or a single genetic approach comprehensively explains complex diseases. Rather, a multitude of approaches augmented by in-depth functional follow-up will be needed. In the following sections, such genetic approaches and their contribution to the field of asthma genetics will be discussed.

1.2. Before the era of genome-wide association studies

One approach (linkage analysis) uses no prior information in identifying regions of the genome that are linked or co-inherited with a particular phenotype of the disease of interest. It uses highly polymorphic markers called the microsatellites that are evenly spaced throughout the genome and comes with the prerequisite of affected families spanning for a minimum of two generations. Evidence of linkage in a typical genome-wide linkage analysis (GWLA) usually spans 10 to 20 Mb with the possibility of multiple genes residing in the region. Thus, follow-up of linkage results is vital in deciphering the actual causal locus. One can either analyse the variations residing in the region systematically by mapping finely spaced genetic markers (positional cloning) or target certain candidate genes in the region based on an apriori hypothesis. Positional cloning is a time-consuming process that involves cloning of the candidate region, followed by the identification of a single causative gene using mutation analysis comparing groups with and without the disease. So far genome-wide screens for asthma and related phenotypes have been carried out in

different populations including the East Asians [9], and owing to their differences in sample size and study design the overall results vary substantially. Nevertheless, asthma/atopy genes successfully identified through linkage and positional cloning approaches were discussed in detail elsewhere [10,11] and Table-1 is an attempt to summarise the genes, linked phenotype and list the regions that are successfully replicated. In genetic studies, it is customary to follow up the hits by replicating them in another independent dataset as an attempt to identify the true disease-susceptibility loci. GWLA in particular struggles to replicate the signals from study to study due to low statistical power, possibly owing to the difficulties involved in ascertaining large pedigrees (large sample size). Two recent meta-analyses of asthma and atopy linkage studies overcome this barrier by combining data from nine Caucasian asthma populations [12] and across different ethnic populations (including Asian, Hispanic, etc.,) [13] conducted for asthma or related phenotypes. They aimed to investigate the consistency of linkage findings across different studies. Altogether, genome-wide evidence for linkage was detected for asthma/atopy phenotypes in 12 chromosomal regions, with few overlapping findings between the two meta-analysis studies. Other than the two regions (2p21-p14 and 6p21.31–p21.1) in European families that reached genome-wide significance levels with asthma [12,13], the remaining regions were linked with lung function phenotypes (6p22.3-21.1, 2q22.1-q23.3, 7q12.11-q31.1, and 5q23.2-q34 for bronchial hyper-responsiveness; BHR) [11,12] and atopy-related phenotypes (2q32-q34 for eosinophil count, 3p25.3-q24 and 17p12-q25 for positive skin prick test; SPT, 5q23-q33 for a quantitative measure of SPT, and 5q11.2-q14.3 and 6pter-p22.3 for IgE) [11-13]. The lesser predominance of asthma-linked regions across these studies might reflect the disease heterogeneity and the need for relatively homogeneous measures, such as quantitative traits (BHR, IgE levels, eosinophil count, cytokines levels, etc.,) instead of a simple affected or not affected status of the disease (dichotomous trait). In addition to identifying new regions and replicating known regions, the main objectives of such meta-analysis are to investigate the shared regions across different phenotypes. Interestingly, overlapping of linked regions between different traits (6p22.3-21.1 for asthma and BHR; 3p14.1-q12.3 for asthma and SPT; 5q23.2 - q34 & 7p21.1 - 14.1 for asthma, BHR and IgE) are observed. This suggests the possibility of pleiotropic effects exerted by the gene residing in these regions leading

to multiple phenotypic expressions and ascertains asthma to be more of a spectrum rather than a single disease entity.

Positionally cloned genes	Chromosomal region	Linked phenotype in the original study	Replicated phenotype in meta-analysis studies
DPP10	2q14	Asthma, IgE	BHR, Denham et al., 2008 [13]
CYF1P2	5q33	Asthma	SPTQ, Bouzigon et al., 2010 [14]
HLAG	6p21	Asthma	BHR, Denham et al., 2008 [13] ; Asthma, Bouzigon et al., 2010 [14]
GPRA	7p14	Atopy, Asthma	—
SFRS8	12q24	Asthma	—
PHF11	13q14	Asthma, IgE	—
ADAM33	20p13	Asthma	—
IRAK3	12q14	Early asthma	—
COL6A5	3q21	Atopy	SPT, Denham et al., 2008 [13] ; SPT, Bouzigon et al., 2010 [14]
CHML	1qter	Atopy,Asthma	—

Table-1: Linked regions that are suggestively replicated in large-scale meta-analysis.

Genes identified by GWLA and positional cloning function as targets for candidate gene studies that rely on such apriori information. It is clearly evident from pathway analysis that a multitude of biological pathways are involved in asthma pathogenesis [14], and thus it is highly likely that there is a larger number of plausible candidate genes. The number of candidate gene studies in asthma could be well over 600, and examining close to 200 genes [10]. The cytokine cluster of genes and β2 adrenergic receptor (*ADRB2*) on 5q, *ICOS*, *CD28* on 2q, Toll-like receptor (*TLR*) family of genes, C-C chemokine receptor genes on 3p, the high-

affinity IgE receptor gene (*FCER1*) and Clara cell secretory protein (*CC16*) on 11q are some of the genes that have been studied more frequently. Overall, the results vary enormously from study to study as a result of population and phenotype heterogeneity, posing a considerable challenge in identifying the true association.

Immunological pathways driving the inflammatory response are a complex system involving multiple genes, systematically regulated under the influence of allergens. Therefore, it is natural for most of the candidate gene studies to have immune-response genes as their study targets. Interestingly, the genes from 5q31 cluster involved in the IL-4/IL-13 pathway exhibit less inter-study variability and are frequently found to be associated with asthma/atopy-related traits [15,16]. *IL-4* and *IL-13* genes are located in close physical proximity (12.5 kb apart) and are involved in isotype switching from IgM or IgG to IgE in activated B lymphocytes [15]. In vivo studies have already demonstrated that the large insert of 5q31 transgenics influences several quantitative phenotypes of asthma by altering gene dosage [17]. Promoter polymorphism in *IL-4* (-589 C/T) is known to be associated with IgE levels in asthmatics [18]. A number of genetic epidemiology studies have successfully replicated this association of *IL-4* promoter polymorphism with asthma in different non-European populations (Chinese, Taiwanese) [19], Russians [20], and in the admixed population of the Madeira island (Caucasoid/Sub-Saharan) [21] where the polymorphism is additionally shown to be an important risk and severity predictor for atopy asthma. Similarly, polymorphisms in the exonic region of *IL-13* (exon 4) are associated with elevated serum IgE levels in children from three different populations [15] and in the German population [22]; moreover, serum IL-13 levels are associated with both allergic and non-allergic forms of asthma in the British and Japanese populations [16]. It is important to remember that the tightly packed nature of this locus involving multiple genes (interleukin genes *IL-4, IL-13, IL-9, IL-5*, as well as *CD14* and *ADRB2*) mandates a careful interpretation of any significant association, as the identified variant could be in high linkage disequilibrium (LD) with another novel/unknown gene. Other possible causes of differences in association studies include population-specific gene-gene or gene-environment interactions.

1.3. GWAS in asthma

Candidate gene studies are, strictly speaking, more of a retrospective approach that starts with the outcome and traces back to investigate exposures. They are advantageous compared to earlier genome-wide screens (GWLA), as it can be applied to independent case-control subjects and are not limited to family-based data. By increasing the case-control ratio and using a larger sample size, the power to detect true associations in a case-control study is enhanced [23]. However, case-control studies using candidate gene approach suffers from the important constraint of focusing only on the genes selected because of an apriori hypothesis, in contrast to the GWAS approach that maps the whole genome. The human genome encodes one SNP for every 300 base pairs [24], and on the genome scale, this comes close to 10 million SNPs at a frequency of > 1% in the world's human population and constitutes 90% of the variation (remaining 10% being rare variants). It is technically and as well statistically challenging to analyse all of these variants, at which point the LD pattern between the SNPs is used. Thereby, tag SNPs that are representative of all high-LD SNPs in a given region (‘haplotype block’) of the genome are genotyped instead of genotyping 10 million SNPs. Comparison of GWAS chips from different populations shows that genotyping chips with 500K-1000K SNPs are a sufficient starting point for a good GWAS [25]. The number of GWAS studies has increased exponentially with the steady drop in the cost of genotyping, and identified true hits satisfying a stringent statistical threshold (10-7 to 10-8 in contrast to the conventional p-value of 0.05) to control the genome-wide type I error rate [26]. So far GWAS studies in asthma have identified close to 45 genes [27], based on the criteria of study size being greater than 300 individuals and replication of the genome-wide hit in at least one additional population (Figure-1).

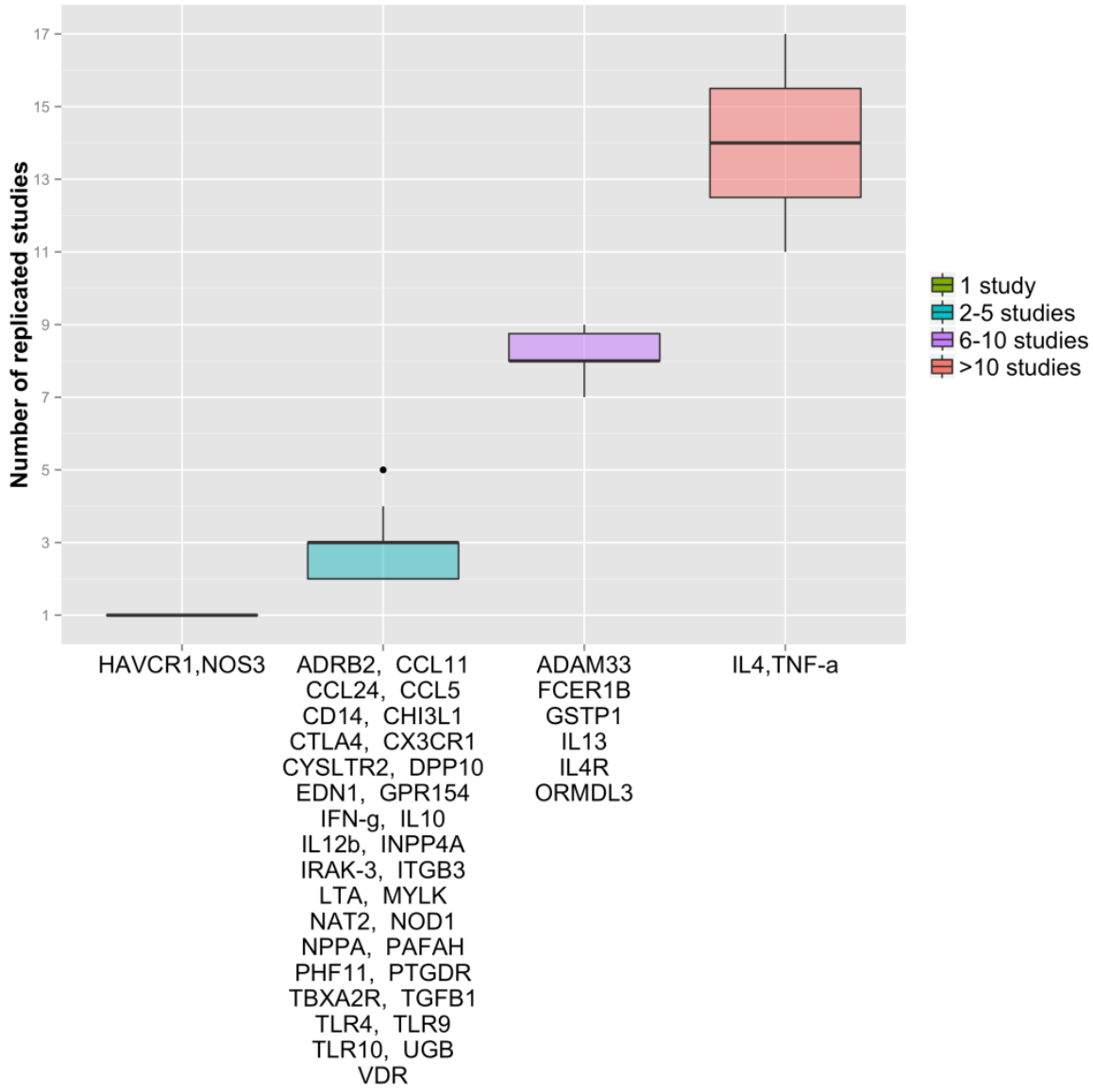


Figure-1: Replication of significant findings in multiple studies. (Data from Weiss et al., 2009 Table 1) [27]

IL-4 and *TNFα* lead the list of genes replicated in multiple studies (> 5 studies) with replication in 11 and 17 studies respectively. *TNFα* shares the locus (6p21) with the well characterised – major histocompatibility complex (HLA) region that has shown strong linkage evidence with asthma [28]. As mentioned in earlier sections, pro-inflammatory cytokines such TNFα are found in high levels in the sputum and bronchoalveolar lavage of asthmatics and are known to play a vital role in asthma severity through the activation and subsequent accumulation of neutrophils [29]. Other important observations include the replication of genes previously known by linkage and positional cloning such as *DPP10*, *FCER1B*, *TLR*,

PHF11, and *ADRB2* [10,11]. *ORMDL3* that is replicated in more than five studies were the first to be discovered by GWAS [30] and owing to their tight LD structure, variants from this locus function as expression quantitative trait locus (eQTL) regulating the expression of *ORMDL3* and other neighbouring genes [11]. It is for this reason of obscurity in the regulatory effects the region is commonly termed as the 17q21 locus. The biological function of the 17q21 locus in mammalian cells is little known, and its mode of action has been recently proposed [31]. *ORMDL3* from this locus encodes a transmembrane protein in the endoplasmic reticulum (ER) that regulates the activity of an epithelial cell-specific membrane protein (SERCA), resulting in epithelial cell remodelling. In addition to studying the association with qualitative asthma traits, intermediate quantitative traits that reflect the severity of asthma were used in one study and identified a novel variant in the promoter region of *CHI3L1* that is genome-wide associated with elevated serum YKL-40 levels [32]. Accounting for the characteristic small effect estimates of variants in complex traits, most of the genome-wide approaches require a larger sample size to achieve statistical power. Consortium based meta-analyses are designed specifically to address this issue, where genotype data from multiple studies with identical phenotype definition are combined. The first GWAS meta-analysis of approximately 10000 asthma cases and 16000 controls integrating close to 500K SNPs from 23 European studies was performed by the GABRIEL consortium [33]. It was soon followed by the meta-analysis of approximately 3000 asthma cases and 3000 controls in ethnically diverse North American populations by the EVE consortium [34], where the analysis was carried out both in ethnically identical groups and in the overall group. In addition to the replication of previously identified loci, both these studies identified new asthma susceptible loci (*PYNHIN1* - in African-descent population by EVE; and *IL-33*, *SMAD3*, *IL1RL1/IL18R1*, *IL2RB1*, *HLA-DQ* - in European population by GABRIEL) [33,34]. Recently, a more gene-centric approach is advocated over SNP-for-SNP replication [35]. This is true if the comparison is made across ethnically different populations, as the underlying assumption of identical allele frequency or haplotype structure does not hold in such cases. In this regard, 17q21 and *IL1RL1* are the only regions that are genome-wide significant in both the studies (Figure-2), followed by *TSLP* and *IL-33* genome-widely significant in one study and suggestive in the other. In addition, there were few associations

(*RORA*, *SMAD3*) significant in the European population (GABRIEL study) but nominally associated with European Americans (EVE study), suggesting the possibility of a difference in risk susceptibility even between people of European descent. Nevertheless, it is important to specify that the variants identified by both these approaches explain little of the overall outcome, as in the GABRIEL study only 35% of the individuals can be classified as asthmatics based on these common variants [33].

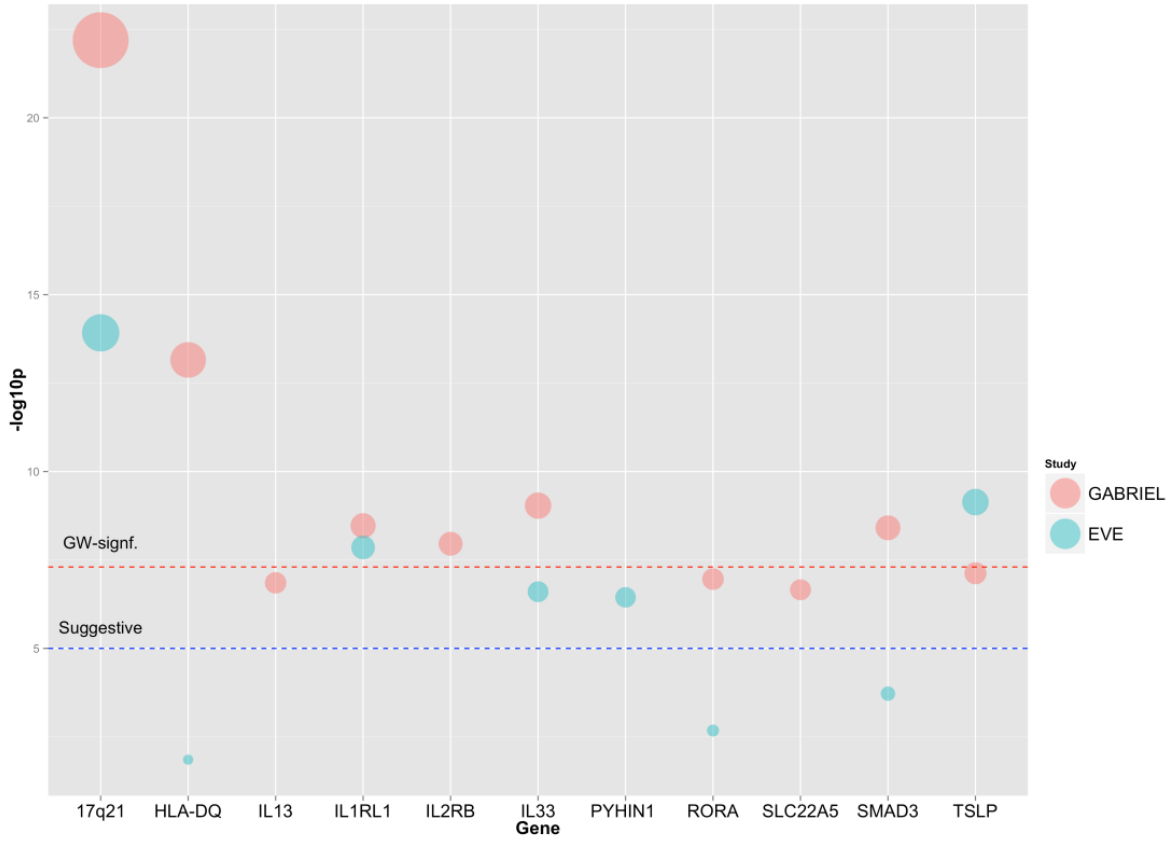


Figure-2: Significant hits from the largest asthma GWAS studies. (Data from Ober et al., 2011 Table-5) [11]

Altogether, the findings from the GWAS approach replicate very little of what is known from earlier candidate gene studies. A plausible explanation for this might be the heterogeneity observed in asthma phenotypes and also to an extent the inadequate representation of previously known regions in commercial genotyping chips. The difference in coverage of the variants in genotyping chips might also explain the less successful efforts in replicating the findings from a European population to any non-European population (Asian or African). The other aspect to

be noted in individual GWAS studies and the meta-analysis approach is that despite having a larger sample size, only a few common variants reach genome-wide significance and with smaller effect sizes. Besides, genetic heterogeneity observed between individuals that can dilute the observed effect sizes [36]. Early life environmental exposure, interacting with the genotype of individuals, might play an important role in conferring the heterogeneity and result in selective association with small effect estimates that is evident only in exposed individuals. This forms the principle of gene-environment (GxE) interaction studies. Prior to the genome-wide approach, initial interaction studies focused on selecting genetic regions and environmental risk factors mainly from experimentally verified data. Dose-dependent relationships between bacterial endotoxin (environmental exposure) and the gene that codes for endotoxins receptor protein (*CD14*), were one of the initial candidate interaction studies [37]. Followed by testing the relationship between the 17q21 locus and tobacco smoke [38], which revealed an enhanced risk of asthma conferred by the 17q21 variant and individuals exposed to tobacco smoke in early life. It opened up a new possibility of expanding the GxE studies to a genomic scale, where estimates of both the main genetic effects (outcome) and their interaction (environmental risk factor) will be determined. The first GxE interaction study at the genome level was carried out to study the relationship of childhood asthma and farm-related exposures in 1,700 children from rural regions of Europe [39]. The study failed to show any significant interaction with common SNPs for which the study was well powered, and instead detected association with rare variants residing in a glutamate gene (*GRM1*) that plays a vital role in T-cell mediated immunity. Other than the initial success of experimental approaches that use prior knowledge on GxE interaction, agnostic approaches in the genome-wide scale are still in their infancy, as new methodological approaches such as next-generation analytic tools being currently proposed as a suitable alternative [40].

1.4. Post-GWAS where next?

Despite the identification of multiple asthma susceptible loci by GWAS studies of different endpoints (childhood asthma, severe asthma, and other asthma-related traits) [41], it still only manages to explain a small fraction of the heritability. Several hypotheses have been proposed to explain the “missing heritability” or "dark matter” [42], and since GWAS ignore low-frequency (MAF < 5%) and rare (MAF < 0.5%) variants it is likely for those variants to hold the answer. In addition, structural variations such as insertion/deletions (INDELS) and copy number variants (CNVs) that can be detected by direct sequencing approaches are largely ignored in a typical GWAS set-up. Next-generation sequencing technologies such as whole genome sequencing (WGS) and whole exome sequencing (WES) are designed specifically to bridge this gap. Previous attempts of large-scale sequencing efforts such as the 1000 genomes project (1,092 individuals / 14 populations) by using a combination of low-coverage WGS and WES [43], have already identified approximately 40 million of common and rare genetic variants including 38 million SNPs, and 1.4 million structural variants. The numbers are expected to rise with future efforts such as the UK10K project that aims to sequence the whole-exomes of 6,000 individuals through high read depth and the whole-genomes of 4,000 European-ancestry individuals at low-coverage (www.UK10K.org). Nevertheless, the validation of rare variants, in particular, comes with the high cost associated with deep sequencing of a large sample size that is required to attain enough statistical power for identifying rare variants. For this reason, it is customary for studies to opt for a combinatorial approach of performing relatively low-coverage WGS and follow up with cost-effective, targeted sequencing for accuracy. In one such recent study, WGS results from a small number of index individuals (16 individuals from asthma enriched and depleted families) were followed up by targeted sequencing to identify structural variants including the CNVs [44]. The authors of this study, however, emphasise the need for a larger sample size and low-cost genotyping technologies to validate their results. With these technologies at hand and the possibility of cost reduction, the next wave of asthma genetic studies might include full genome sequencing analyses with the power to answer the century-long search for asthma genes.

1.5. Family-based study designs in genetic epidemiology and their advantages

Genetic epidemiology differs from its classic counterpart by explicitly taking genetic factors, gene-environment interactions, and familial information into consideration. While the initial studies (familial aggregation, heritability, and linkage analysis) relied mainly on family-based approaches, later LD association mapping (candidate association studies, GWAS) studies that are more powerful in detecting moderate effects relied on the inclusion of relatively less laborious population-based (unrelated) approaches. However, association-based studies failed to replicate their findings or to explain the missing heritability, in spite of reporting larger numbers of true hits. One classic example for this could be the inability of GWAS studies to replicate the breast cancer susceptibility loci (*BRCA1*, *BRCA2*) identified by family-based approaches [45]. A possible explanation for this could be the enrichment of genetic effects more in the affected as compared to unrelated individuals and thus the greater power to detect the association. Similar differences seen in the allele frequency between cases and controls is exploited in association-based approaches to detect disease variants, any introduction of ethnically different allele frequency results in population heterogeneity (or admixture) and thus the greater probability of false positives. Population-based studies that aim to increase sample sizes by including unrelated, ethnically different controls suffer from this stratification, whereas the family-based controls are completely immune and robust to population substructure. Other advantages of family-based designs include their immunity to genotyping errors as they can easily be detected by comparing parental and offspring genotype and check for any Mendelian inconsistencies [46]. In addition, approaches in family-based designs have evolved significantly from the strict parametric model (assumes specific genetic model) to non-parametric methods such as the TDTae (extension of Transmission Disequilibrium Test) that is developed for robustness against genotyping errors and hidden genetic structures [47], and the FBAT (Family-Based Association Test) approach that permits different genetic models, as well as multiple markers and phenotypes [45]. Recent efforts of inferring genotype information from genotyped individuals to impute un-genotyped individuals based on familial information in a cohort-based study (Framingham Heart Study,

FHS) [48] has opened the possibilities of genotyping more individuals at a lesser cost. Such a pseudo sequencing strategy has already been used in increasing the power to detect rare variants in a family-based study [49]. Altogether, it has led to a renewed interest in seeing both family-based designs and population-based designs as complementary instead of competitive.

1.6. Asthma phenotypes - Th2 associated asthma

Asthma is an umbrella term for a group of clinical features including reversible expiratory airflow limitation and airway inflammation. The characteristics of these entities are so diverse with a sporadic occurrence that it is not possible to group them as a single disease entity. As a result, it is quite difficult to unify the phenotypes as one true phenotype representing the disease. Moreover, unifying procedures that are based on a single dominant character like atopic status can lead to possible bias, and thus unbiased approaches like cluster analyses that include additional parameters like personal characteristics, age at onset, treatment, lung function, etc. have been proposed [50-52]. Interestingly, in spite of the differences in study design, all unbiased studies identified two major phenotypes/subgroups: early-onset allergic phenotype and the less atopic late-onset phenotype. While the clusters showed age at disease onset as a key determining factor in defining the subgroups, it is equally important to perform phenotyping based on molecular signatures. The study by Woodruff et al. is one such study that showed it is possible to classify asthmatic subjects into ‘Th2 high’ and ‘Th2 low’ based on the expression profile of IL-13 - inducible genes (*POSTN*, *CLCA1*, *SERPINB2*) [53]. The study also validated the classification by showing marked differences in the cytokine expression profile and markers of inflammation between the two classes, with more atopy seen in Th2 high compared to Th2 low people. It is vital to integrate clinical-clusters with molecular profiles for a better resolution, especially in the overlapping region encompassing the late-onset category. Defining the phenotype is an important prerequisite for genetic studies, and in particular, intermediate phenotypes are proposed to be a better alternative to questionnaire-based qualitative phenotypes. While the initial genetic and epidemiological studies in asthma focused on qualitative traits, the possibility of erroneous estimation of disease prevalence along with the diagnostic bias has encouraged researchers to study quantifiable biological markers [54]. Moreover,

tightly defined quantitative intermediate phenotypes are often regarded as the units of binary traits and are much closer to the underlying biology; and is more informative being closely related to the gene expression. Importantly, the results are easily interpretable when compared to genetic association studies involving qualitative traits.

2. Aim of the study:

Complex gene-environment interactions in the aetiology of asthma make the identification of causal genes arduous by just applying methodologies used in monogenic disorders. Furthermore, difficulties in defining the diagnostic criteria and the heterogeneity observed in the clinical phenotypes compounds the issue.

For the past two decades, genetic studies all over the world have identified myriads of risk loci. Owing to the heterogeneous genetic sources for the phenotype and heterogenous epigenetic backgrounds or environmental influences most of the studies have been inconclusive. Affected family-based study has the advantage of identically phenotyped families with shared genetic and environmental background, which may be a promising alternative to explain the aetiological heterogeneity observed so far.

Herein, intermediate sub-phenotypes in the form of circulating serum cytokines are explored to investigate its correlations with the genetic component. As shown in the introduction, there is evidence from prior own studies and literature, of the direct involvement of key cytokines (Eotaxin, GM-CSF, IFNγ, IL-1b, IL1RA, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12(p40), IL-13, IL-17, IL-23, IL-33, TSLP and TNFα) in asthma pathology. Cytokine measurements from 218 asthmatic families with the affected sib-pair (ASP) structure are used as the phenotype, and its heritability is determined for the first time. Linked chromosomal regions using autosomal microsatellite markers are examined, and SNP marker tested for association at genes implicated earlier with the autoimmune response.

3. Materials and methods:

3.1. Probands

Samples were collected from 218 core asthma families in two stages and probands clinically evaluated for the severity of asthma [55,56]. The sample definition is identical to previous reports with regards to the study participants, clinical characteristics and the general selection process. For the first stage, 103 asthma sib-pair families were selected, five families were excluded due to anomalous segregation (non-mendelian inheritance) of microsatellite markers, and one family was excluded with monozygotic twins, leaving 97 in total for an initial genome-wide linkage scan [55]. In the second study stage, 121 asthma sib pair families were selected over a period of 18 months. Identical phenotyping procedures were carried out by three university hospitals, and six paediatric pulmonary practices in both the stages of sample collection. Families comprising at least two children with clinically confirmed asthma are included and children with a history of low birth weight or prematurity, along with any severe pulmonary disease other than asthma were excluded. Altogether, the final sample consisted of 944 individuals (218 families); and after the exclusion of 21 samples with insufficient clinical details, 923 individuals remained for further analysis (Parents: n=424; Children: n=499). Whole genome linkage scan was performed with approximately 400 microsatellite markers [57]. Notably, clinical diagnosed asthmatic children over the age 3 had a history of recurrent wheezing that lasted for at least 3 years, without any other respiratory diseases. Clinical procedures contained detailed interviews of every family member, skin prick tests (SPT) for frequent allergens and blood sample assays (for IgE and allergen-specific IgE (RAST) measurements and ELISA (Pharmacia Diagnostics, Uppsala, Sweden) were applied to determine the total IgE levels, and allergen-specific IgE (CAP FEIA system). The specificity test enables to differentiate atopic (allergic) asthma from the non-atopic type. The parent/guardians of all participants under 18 gave informed consent. Ethics Commission of ‘Nordrhein-Westfalen’ approved the study methods in 1995 and later again in 2001 by ‘Bayerische Landesärztekammer München’.

3.2. Serum measurements

As reported in Pukelsheim et al. [58] concentration levels of serum eotaxin, GM-CSF, IFNγ, IL-1b, IL1RA, IL-8, IL-10, IL-12(p40), IL-4, IL-5, IL-6 IL-13, IL-17, IL-23, IL-33, TSLP and TNFα were quantified using custom-made MILLIPLEX MAP Human Cytokine/Chemokine Panels (Millipore, Schwalbach/Ts, Germany). The assays were performed as per the manufacturer recommendations and an 8-point standard curve for every cytokine is used for the dilutions. Intra-assay sensitivity has a mean coefficient of variance (CV) of 4.5% to 10.4%. Samples were analysed on a Luminex 100 device (BioRad, München, Germany) and evaluated using the Bio-Plex Manager software (BioRad). Detailed information on the sample measurements is available in the Supplementary Table S1 of the original article by Pukelsheim et al. [58]). Values reaching below the detection limit were set to half of the limits, and if exceeding the maximum limit (3000 pg/ml), it is set to 3000 pg/ml.

3.3. DNA preparation and genotyping

Peripheral white blood cells were used in the DNA isolating procedure, by the salt out method [59], and in few cases with the Qiamp blood kit (Qiagen-Hilden, Germany). Microsatellite genotyping is based on an in-house microsatellite-mapping panel (Genethon reference) and includes 18 markers on the X chromosome in addition to the autosomal markers. Detailed information on the microsatellite panel is available in the methods section of earlier German Asthma Family Study [55].

SNP selection is based on literature mining and included loci/regions implicated in other asthma studies. Information from publicly available disease-specific databases (http://geneticassociationdb.nih.gov - for the known genetic associations; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM - human OMIM; http:// www.bork.embl-heidelberg.de/g2d/ - gene ontologies) were used in the refining process of identified candidate genetic regions. In addition, SNP information from dbSNP (http:// www.ncbi.nlm.nih.gov/SNP/), Ensembl (http://www.ensembl.org) and innate immunity PGA (http://innateimmunity.net) were used to finalise a total of 550 SNPs. SNP-genotyping was done over a 5-year period, and highly polymorphic markers were selected that are equally spaced in the genome with markers also at

telomeric positions. It is performed in-house or at Illumina (San Diego, CA, USA) by use of the Sentrix bead arrays. PCR conditions included initial denaturation for 10 min at 95°C, followed by extension of 45 cycles for 20 sec at 95°C, 30 sec at 72°C and a final extension for 10 min at 72°C and stored at 4°C. Primers were obtained from Metabion GmbH (Planegg, Martinstried, Germany) and MWG Biotech AG (Ebersberg, Germany). For validation of the selected SNPs, allele-specific primer extension products generated from amplified DNA sequences were tested with the mass spectrometry - MALDI-TOF (MassARRAY, SEQUENOM Inc., San Diego, CA, USA). Mass spectra were processed and analysed for peak area calculation, peak identification, and allele frequency estimation (SpectroTYPER RT 2.0, SEQUENOM Inc, San Diego, CA, USA). After amplification procedures, PCR products are sequenced using the ABI 377 automatic sequencer and scored using GENE-SCAN and GENOTYPER (ABI) software.

3.4. Statistical analysis

Statistical analysis was performed using the freely-available statistical software, R 2.12.2 (R studio 0.93) [60]. Descriptive analysis performed using cross-tabulation; distribution of single variables and normalisation. Effect estimation by linear and logistic regression analysis corrected for the covariates age and sex. Merlin 1.1.2 [61] was used in the calculation of QTL, single and multipoint non-parametric linkage analysis - Kong and Cox linear model in determining significance based on a likelihood-based approach (LOD scores). Variance components linkage analysis was performed along with the estimation of narrow sense heritability or additive genetic heritability (h2) adjusted for the covariates. PLINK 1.07 [62] was used to perform the Transmission disequilibrium test (TDT) and the family-based tests of association with quantitative phenotypes by using the QFAM option. Association of cytokine levels with genotypes SNPs was calculated after testing for missing data, correction of paternity errors and calculation of Hardy-Weinberg equilibrium. Instead of fitting a maximum likelihood variance components model as done in the QTDT, QFAM does a simple linear regression of phenotype on genotype and provides pointwise empirical significance. It applies special permutation steps that are robust and by preserving the correlation structure between SNPs, it provides a less stringent correction for multiple testing, importantly accounting for the relatedness between

individuals. Multiple linear regression analysis was performed using variables on the drug intake and serum cytokine levels. Correlation matrix and scatterplot matrix are used to visually ascertain the interrelationships between variables. In addition, the combined effect of different drug intake on cytokine levels was tested using the multiple analysis of variance (MANOVA).

3.5. Lookups and Insilico annotation

Figure-3 shows the prioritisation pipeline of significantly associated variants by using publicly available functional datasets. In the data-driven approach, significant SNPs and its high-LD SNPs were interrogated for its effect on amino acid sequence, gene expression and the possible involvement of regulatory regions from experimentally confirmed datasets. On the other hand, the same set of SNPs are investigated based on a priori knowledge by using association catalogues and biological knowledge base. Variant effect predictor (VEP) online tool was used to catalogue the LD variants (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP). The following section brief on the tools and steps involved under each category,

Protein altering effect: ANNOVAR - [63]. ANNOVAR imports precompiled database on amino acid changes and based on the allele information, and genomic coordinates of the SNPs the resulting biological consequence of the variant is determined. In addition, it provides the protein-altering effect of a SNP predicted by function prediction algorithms like SIFT (Sorting Intolerant From Tolerant), and PolyPhen (Polymorphism Phenotyping).

eQTL lookup: Expression quantitative trait loci (eQTL) represent the genetic markers that are associated with the gene expression. GTEx V2 (web-based portal) [64], holds experimental information on eQTLs from seven different cell/tissue type and can be accessed through http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi. Additionally, the GRASP (Genome-Wide Repository of Associations between SNPs and Phenotypes) database [65] was queried to identify SNPs that are known to be functioning as eQTLs in published datasets.

Epigenetic annotation: Association of variants with epigenetic marks was studied by using HaploReg V3 [66]. It takes the rsID of a variant as the input and annotates the region for the enrichment of specific histone modification marks for the chromatin state, using Chip-Seq data on multiple cell/tissue types from ENCODE

(Encyclopedia of DNA Elements) [67] and Roadmap (National Institutes of Health Roadmap Epigenome projects) [68].

Association catalog: NHGRI-EBI GWAS catalog [69] is a manually curated, literature-derived collection of published GWAS studies assaying at least 100K SNPs and SNP-trait associations with p-values < 1.0 x 10-5.

Canonical pathways and regulatory network analysis: “Data were analyzed through the use of QIAGEN’s Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity).” “The [networks, functional analyses, etc.] were generated through the use of QIAGEN’s Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity).” Genes defined from a given list of SNPs are mapped to its corresponding gene set in the IPA knowledge base, and the associated canonical pathway is displayed. In addition, the IPA generated network gives insight on the molecular interactions of biological molecules and for this analysis direct physical interactions like protein-protein, chemical modifications were used. Network size was set to the default size of 35 nodes and the relationship between the network are based on the experimental information from the IPA knowledge base. Focus molecules in the network are Gene/Protein/ Chemical identifiers that made the user-defined cutoff and map to the global molecular network, and the significance is evaluated based on the following parameters; (1) P-value: calculated by Fischer’s exact test that is the probability of the observed association by random chance alone; (2) Ratio: represents the number of molecule map to the total number of genes that participate in the putative canonical pathway; and the (3) score value representation of the negative log p-value.

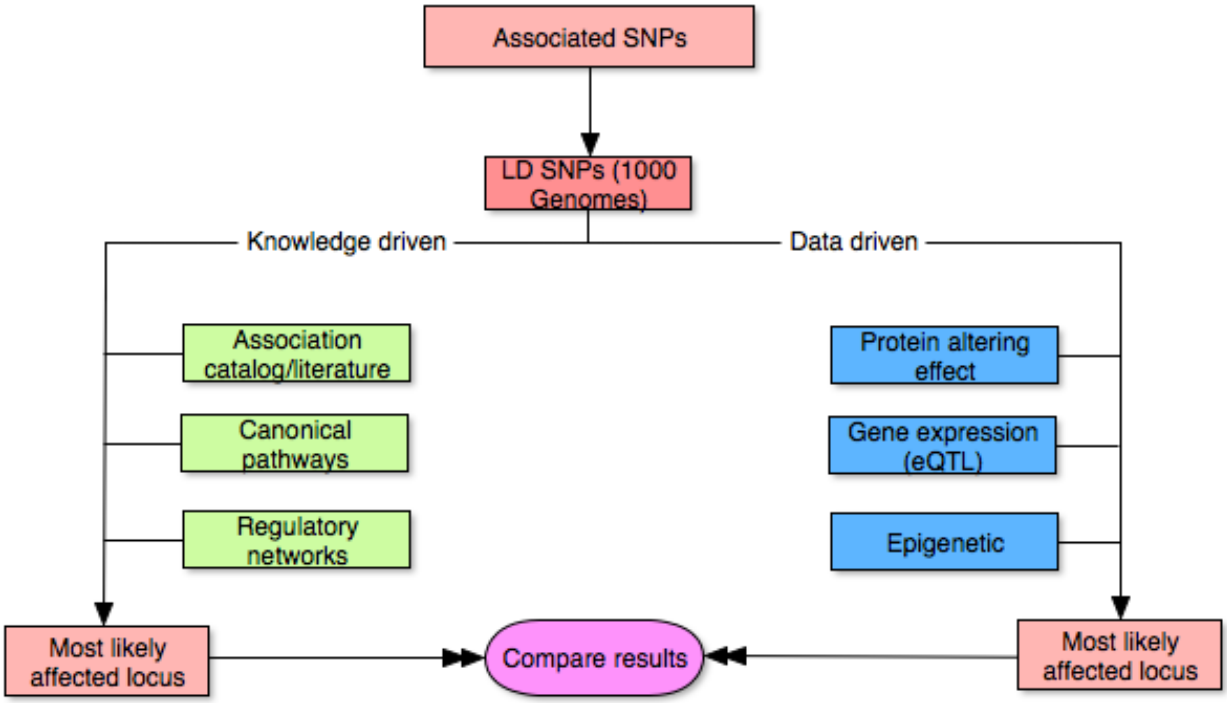


Figure-3: Overview of candidate SNP and gene prioritisation pipeline.

Contributions: Matthias Wjst (MW) conceived, and coordinated the study. Katrin Pukelsheim and David Kutschke performed the cytokine measurements. Muralidharan Sargurupremraj (MS) and MW analyzed the data. MW contributed reagents/materials/ analysis tools. MS wrote the papers.

4. Results:

The main demographic details of the study participants are summarised in Table-2. The samples across the two subsets (parents and children) were equally distributed with a mean age of 40 and 11 years respectively. Male and female distribution is almost equal in both the groups. The percentage of individuals with physician-diagnosed asthma is higher in children (88.7%) than in parents (21.9%), and the youngest asthmatic child was 4 years old while the oldest being 34 years old.

Characteristics	Parents	Children
Number, n (%)	424(44.9)	499(52.8)
Mean age (in Years)	40 ± 5.5	11 ± 3.8
Asthma Diagnosis (%)	93(21.9)	443(88.7)
Sex (%)		
Male	223(52.5)	275(55.1)
Female	201(47.4)	224(44.8)
Skin prick test		
D.pter Skin >= 3mm	129(30.4)	221(44.2)
D.far Skin >= 3mm	111(26.1)	189(37.8)
Serum-specific IgE (CAP) test		
D.pter	101(24.0)	237(48.3)
D.far	94(22.3)	234(47.5)
FEV1(ml)	3.6 ± 0.9	2.6 ± 0.9
In(IgE) (kU/L)	4.1 ± 1.6	5.0 ± 1.9
In(eosinophil) (count/mm3)	0.5 ± 1.7	1.1 ± 1.6

Table-2: Clinical characteristics of the study population (n = 923). Descriptors of categorical variables include n (total sample) followed by percentage. Descriptors of continuous variables include mean values followed by standard deviation. Abbreviations: D.far, *Dermatophagoides farinae*; D.pter, *Dermatophagoides pteronyssinus*; FEV1, forced expiratory volume in 1 second. Note: Percentages may not add up to 100 due to rounding. Standard deviation expressed as ±

4.1. Drug effects on cytokine levels

Significant effects were found in patients treated with glucocorticoid, cromoglycate, and antihistamines ($p < 0.05$) (Table-3). The coefficient is negative for the pro-inflammatory cytokines (GM-CSF, IL-8, TNFα and IL-1b) with antihistamine, cromoglycate, and the leukotriene antagonist treated group. Similarly, the chemokine-Eotaxin exhibits a negative association with glucocorticoid. On the other hand, the treatment of glucocorticoids was associated with higher IL-4 and IL-5. Multivariate analysis of variance that studies the overall drug effects on serum cytokines shows that glucocorticoid contributes to the model more than any other substance (Supplementary 10.1).

Drug	Regression model							
	Eotaxin	GMCSF	IL-4	IL-5	IL-8	TNFα	IL-1b	IL1RA
Glucocorticoid	-0.071 0.0048*	-0.079 0.309	0.2 0.0427*	0.19 0.013*	-0.038 0.56540	-0.004 0.9007	0.010 0.9963	-0.021 0.685
Leukotriene antagonist	-0.014 0.9035	-0.105 0.764	-0.453 0.3079	0.006 0.986	0.018 0.95172	-0.111 0.4089	-0.739 0.0231*	-0.434 0.057
β_2-sympathomimetic	-0.053 0.0582	-0.018 0.837	0.025 0.8212	0.003 0.976	0.157 0.03440*	0.037 0.2674	0.081 0.3179	0.111 0.052
Aminophylline	0.018 0.7315	0.173 0.281	-0.314 0.1202	-0.221 0.156	-0.118 0.38702	-0.004 0.9427	-0.193 0.1925	-0.109 0.295
Cromoglycate	0.023 0.3962	-0.006 0.939	0.259 0.0152*	0.172 0.037*	-0.238 0.00102*	-0.028 0.3794	-0.211 0.0074*	-0.057 0.299
Anticholinergic	0.069 0.0880	0.15 0.237	0.123 0.4407	-0.046 0.709	-0.165 0.12741	0.061 0.2061	0.056 0.6364	0.029 0.726
Antihistamine	-0.03 0.4733	-0.317 0.016*	0.026 0.8774	0.010 0.998	0.009 0.93582	-0.13 0.0099*	-0.197 0.1095	-0.164 0.058

Table-3: Multiple regression analysis - effect of drug intake on serum cytokine levels (Significantly correlated 8 cytokines from the total of 17 cytokines). Depicted are regression coefficients followed by the p-value (* Significant).

4.2. Heritability

Additive genetic variance (combined effect of all loci) was used to determine narrow sense heritability from the serum concentration of 17 cytokines as shown in Figure-4. After adjusting for confounders, IL-8 (h2 = 1.00) had the strongest genetic component followed by TNFα (h2 = 0.55), IFNγ (h2 = 0.51) and IL-1b (h2 = 0.51), suggesting a tight genetic control for the production of these chemokines and pro-inflammatory cytokines. IL-6 (h2 = 0.43) showed moderate heritability with shared environmental component. Heritability for other cytokines (IL-5, IL-10, IL-13, IL-23 and IL1RA) varied from weak genetic influence to no genetic influence (TSLP and IL-33).

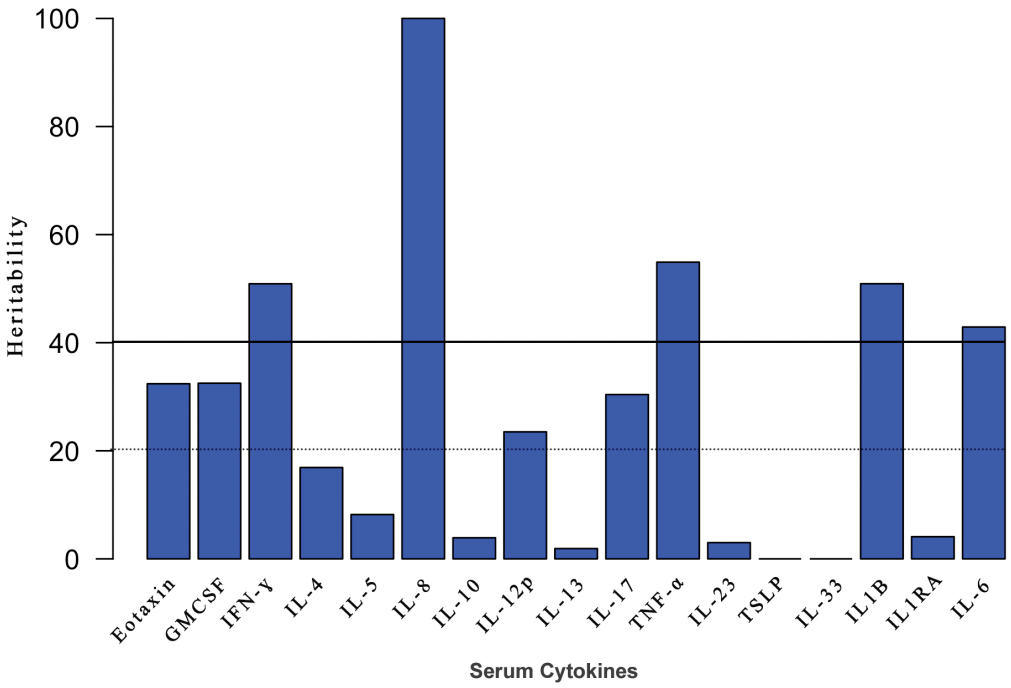


Figure-4: Heritability estimates of serum cytokines. Effect sizes ranging from environmental effect (heritability < 20, dotted line) to genetic effect (heritability > 40, straight line). Heritability expressed as percentage.

4.3. Linkage

Non-parametric genome-wide linkage analysis (NPL) with 334 autosomal markers adjusted for age and sex (linear model) revealed two quantitative trait loci (QTL) (D6S344; D17S949) that are above the thresholds (Figure-5). One peak reached the genome-wide linkage significance level (LOD > 3.6) with a LOD score of 4.05 (crude p-value of 7.80 x 10-6) for IL-12(p40), followed by the suggestive evidence of linkage (LOD > 2.2) with a LOD score of 2.56 (crude p-value of 3.00 x 10-4) for Eotaxin levels (Table-4).

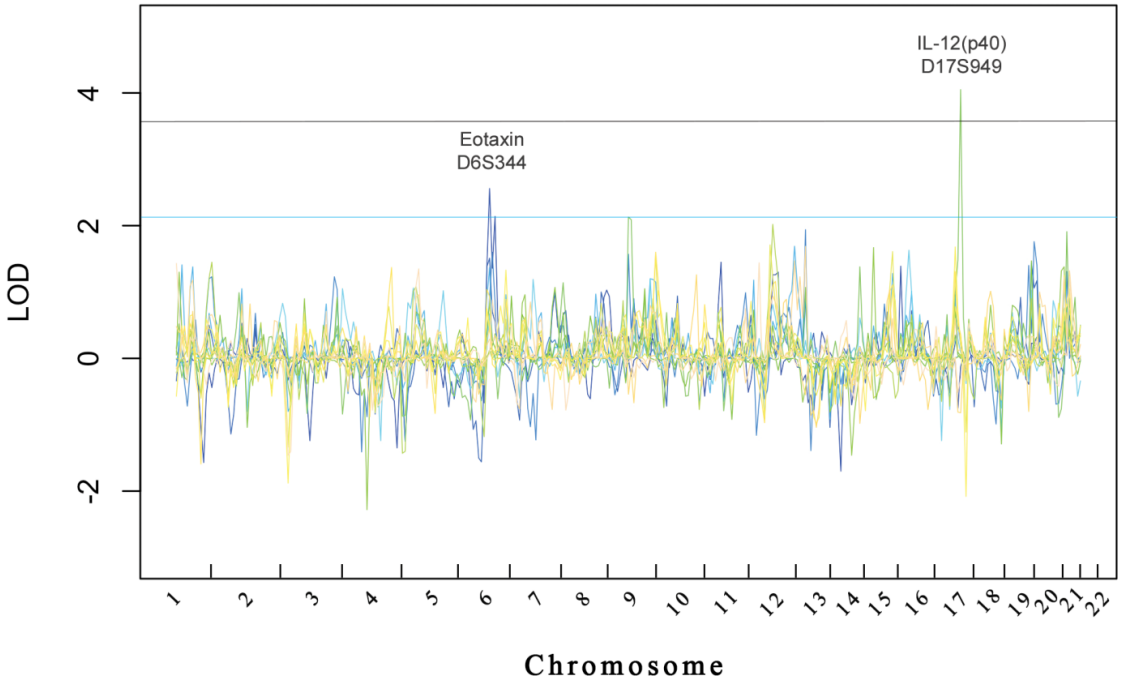


Figure-5: Genome-wide linkage scan with 334 autosomal markers. QTL with a genome-wide significant LOD > 3.6 (Black horizontal line) on 17q24.3, and suggestive LOD > 2.2 on 6p25.3 (Blue horizontal line). LOD scores of the markers used in the linkage panel are plotted in cM relative to their chromosomal location.

Pheno-type	Chr †	Marker	cM	LOD	p-value	Candidate genes (distance from the marker in Mb *)
Eotaxin	6p25.3 6p24.3	D6S344 D6S470	1.40 18.22	2.56 2.14	3.00E-04 8.00E-04	*FOXF2*(-0.2), *SERPINB1*(1.2) *GCNT2*(0.4), *EDN1*(2.2)
IL-12(p40)	9p24.2 9p24.1 17q24.3	D9S288 D9S286 D17S949	9.83 18.06 93.27	2.13 2.09 4.05	8.00E-04 9.00E-04 7.80E-06	*RFX3*(-0.4), *JAK2*(1.0) *IL-33*(-1.7), *PTPRD*(0.2) *MAP2K6*(-0.9), *KCNJ2*(-0.2), *SSTR2*(2.6)
IL-4	6p25.3	D6S344	1.40	1.51	4.00E-03	*FOXF2*(-0.2), *SERPINB1*(1.2)
IL-5	9p24.2	D9S288	9.83	1.57	3.00E-03	*RFX3*(-0.4), *JAK2*(1.0)
TNF-α	12q13.11	D12S85	62.10	2.02	1.00E-03	*WNT1*(2.0), *KCNH3*(2.5), *GPD1*(3.1)

Table-4: Non-parametric linkage analysis (n = 923). † Chromosome position. Abbreviations: *EDN1*, endothelin 1; *FOXF2*, forkhead box F2; *GCNT2*, glucosaminyl (N-acetyl) transferase 2; *GPD1*, glycerol-3-phosphate dehydrogenase 1; *JAK2*, janus kinase 2; *KCNH3*, potassium voltage-gated channel H3; *IL-33*, interleukin-33; *MAP2K6*, mitogen-activated protein kinase kinase 6; *PTPRD*, protein tyrosine phosphatase receptor type *DKCNJ2*, potassium inwardly rectifying channel J2; *RFX3*, regulatory factor X 3; *SERPINB1*, Serpin peptidase inhibitor, clade B1; *SSTR2*, Somatostatin receptor 2; *WNT1*, wingless-type MMTV integration site 1. *Distance separating the marker and the candidate genes are taken from UCSC genome bioinformatics site (http://genome.ucsc.edu/). Negative and positive values represent the position of the gene located downstream or upstream from the marker, respectively.

4.4. Association

Quantitative trait association analysis of 550 genotyped SNPs shows that of the multiple significant associations from TDT (Supplementary-10.2), few were replicated in parents or children only. Association was found to be highly significant with SNP markers rs439154 (2q13), rs2070948 (4q31.23), rs3894194 (17q21.1), rs3135499 (16q12.1), rs3782905 (12q13.11) residing in *IL1RN*, *NR3C2*, *GSDMA*, *NOD2* and *VDR* gene respectively, amongst the children (Table-5). In Interleukin-1 receptor antagonist gene (*IL1RN*), individuals with the homozygous minor allele (AA) of the intronic variant (rs439154) was associated with significantly lower serum levels of IL-12(p40) than in those with the AG genotype or the GG genotype

(Supplementary-10.3.a). IL-23, TSLP, and IL-33 show statistical significance with those individuals carrying the homozygous major allele (GG) and except for the slightly higher levels of IL-23 and TSLP, no major differences in the serum levels were seen possibly as a result of the small subject number (Supplementary-10.3.a). In mineralocorticoid receptor gene (*NR3C2*), serum levels of IL-13 and eotaxin show a statistically significant association with the intronic polymorphism and are higher for the individuals with homozygous major allele (TT) (Supplementary-10.3.b). Furthermore, isolated associations were seen amongst the *GSDMA* and *VDR* variants with the serum levels of IL-33 and IL-13 respectively.

Gene name	SNP	Chr:pos †	MAF	Functional	Allele	Cytokine	p-value	
							AA vs AG	**AG vs GG**
IL1RN	rs439154	2:113884401	0.47	Intronic	A/G	IL-4 IL-5 IL-12(p40) IL-23 IL-33 TSLP	0.032 0.021 0.003 0.042 0.467 0.019	0.529 0.145 0.036 0.0006 0.028 0.004
							AA vs AT	**AT vs TT**
NR3C2	rs2070948	4:149358729	0.33	Intronic	A/T	Eotaxin IFNγ IL-13	0.021 0.043 0.026	0.036 0.136 0.004
							CC vs CT	**CT vs TT**
GSDMA	rs3894194	17:38121993	0.42	Missense	C/T	IL-33	0.015	0.001
							AA vs AC	**AC vs CC**
NOD2	rs3135499	16:50766127	0.37	3' UTR	A/C	GMCSF	0.029	0.031
							CC vs CG	**CG vs GG**
VDR	rs3782905	12:48266167	0.24	Intronic	C/G	IL-13	0.013	0.659

Table-5: SNP association with cytokine expression profile. † Chromosome position with respect to NCBI built 37; MAF= Minor allele frequency.

4.5. Insilico annotation

Locus boundaries: Significantly associated SNP from each locus were used to define their boundaries based on the linkage disequilibrium (LD) between the lead SNP and proxy SNPs (LD value-r2 > 0.5) using the PLINK tool set for pairwise LD calculation, with 1000 genomes (CEU) as the reference population. In total, it resulted in 149 SNPs of which 95 located in the intronic and non-coding transcript regions, 4 intergenic variants and 35 downstream and upstream variants, 7 variants from the UTR location, 1 synonymous variant, 1 splice region variant, 3 non-synonymous variant and 3 other variants (Supplementary-10.4). Variant effect predictor (VEP) tool was used to catalogue all the genes residing in the genomic coordinates of these 149 SNPs. Collectively, there were 10 protein-coding genes (Supplementary-10.5).

As outlined in the Methods section, the primary objective was to identify the most plausible locus and the pipeline consists of two different parts; using experimental data and using prior knowledge.

Protein altering effect: The effect of a SNP on the amino acid (AA) change and the biological consequence was determined using the ANNOVAR tool. Of the 149 SNPs, 3 missense variants from the *GSDMA* locus lead to AA change with deleterious protein-altering effect predicted by the SIFT, PolyPhen algorithm and the integrated Condel score (Table-6).

Lead SNP	Chr †	LD	LD SNP	Conse-quence	Nearby Gene	AA change	SIFT	Poly-Phen	Con-del
rs3894194	17	1.0	rs3894194	missense-variant	*GSDMA*	p.Arg18Gln	Ben (0.14)	Ben (0.05)	Neutral (0.46)
rs3894194	17	0.8	rs7212938	missense-variant	*GSDMA*	p.-Val128Leu	Del (0.04)	Ben (0.01)	Del (0.48)
rs3894194	17	0.6	rs56030650	missense-variant	*GSDMA*	p.Thr314Asn	Ben (0.1)	Del 0.52)	Del (0.72)

Ben-Benign; Del-Deleterious

Table-6: Functional prediction of variants. † Chromosome; LD = Linkage disequilibrium; AA = Aminoacid; SIFT = Sorting Intolerant From Tolerant; PolyPhen = Polymorphism Phenotyping; Condel = Consensus deleteriousness.

Gene expression: In addition, to the published eQTL effect (Supplementary 10.6) of the SNPs from the 17q21 locus using GRASP, lookup using the GTEx dataset showed significant eQTL effect on the *GSDMA* gene in different tissue types including blood cells, immune cells, brain, and liver (Figure-6). No significant eQTLs were found for the SNPs from other loci using GTEx, except for few high-LD SNPs that were known to be functioning as eQTLs in published dataset (using GRASP) (Table-7) altering the gene expression levels in immune cells and blood dendritic cells of *Mycobacterium tuberculosis* (MTB) infected individuals.

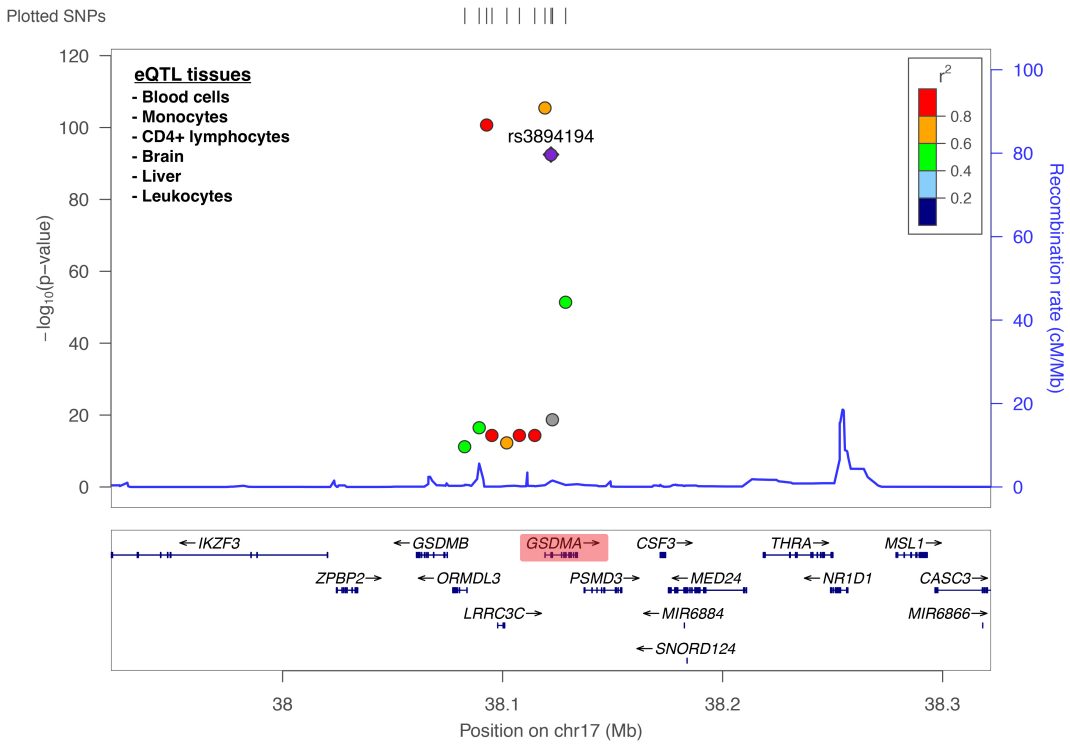


Figure-6: Tissue-specific eQTL effect of *GSDMA* locus (using GTEx). Regional association plot of eQTL association p-value of the lead SNP rs3894194 and its LD SNPs. (Associated gene is highlighted in red).

Lead SNP	Chr †	LD	LD SNP	eQTL association p-value	Gene	Cell/tissue Type	PMID
rs439154	2	0.9	rs2637988	5.40E-07	*IL1RN*	Blood cells	21829388
rs2070948	4	0.8	rs4635799	2.09E-02	*NR3C2*	Blood dendritic cells (MTB)	22233810
rs3135499	16	0.9	rs11647841	4.27E-04	*NOD2*	CD4+ lympho-cytes	20833654

Table-7: Tissue-specific eQTL effects of other loci (using GRASP) and the corresponding publication (ID) PMID. † Chromosome; LD = Linkage disequilibrium; eQTL = expression quantitative trait loci.

Epigenetic: annotation using Haploreg V3 identified putative regulatory SNPs based on the enrichment for chromatin states estimated by histone modification marks in different cell/tissue types from ENCODE and ROADMAP projects. SNPs from the *NR3C2*, *NOD2* and *GSDMA* locus are enriched for histone modification marks for promoter (H3K4Me1) and enhancer (H3K27ac) activation in multiple cell/tissue types including the lung (Figure-7a,7b).

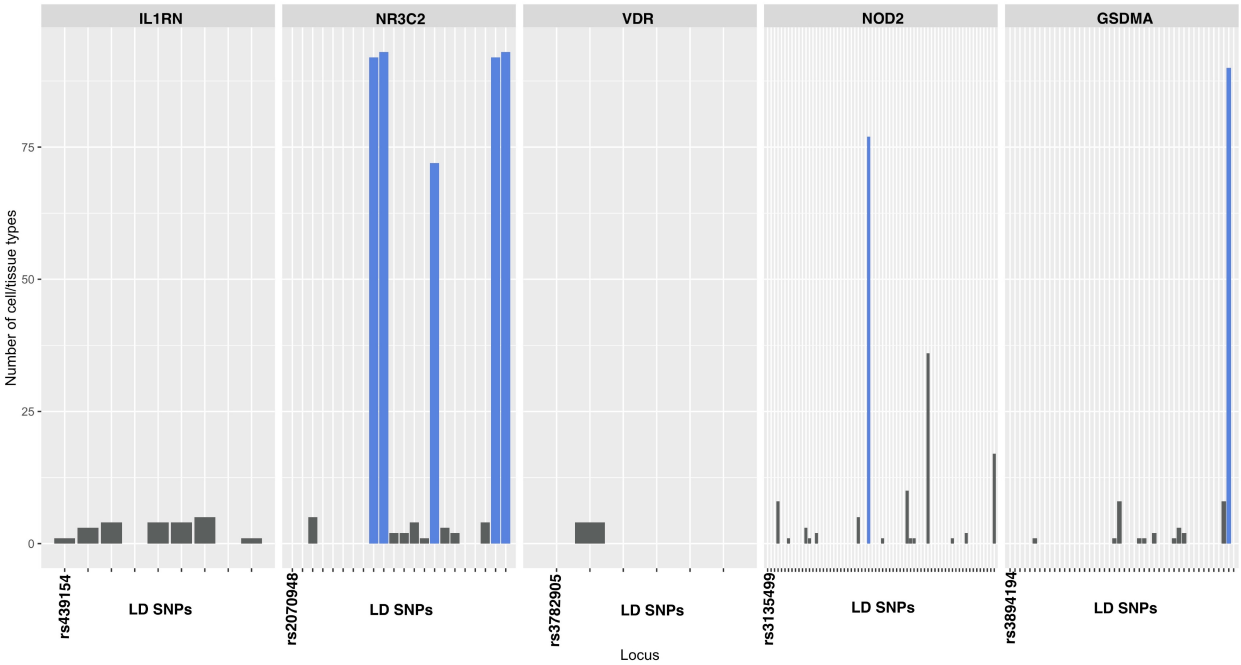


Figure-7a: Enrichment of promoter associated histone modification mark (H3K4Me1) in multiple cell/tissue types. Blue highlighted bar represents the SNPs that are enriched in multiple cell/tissue types including the lung. Prime SNP indicated followed by its LD SNPs. Note: Number of high-LD (r2 > 0.5) SNPs varies for each SNP.

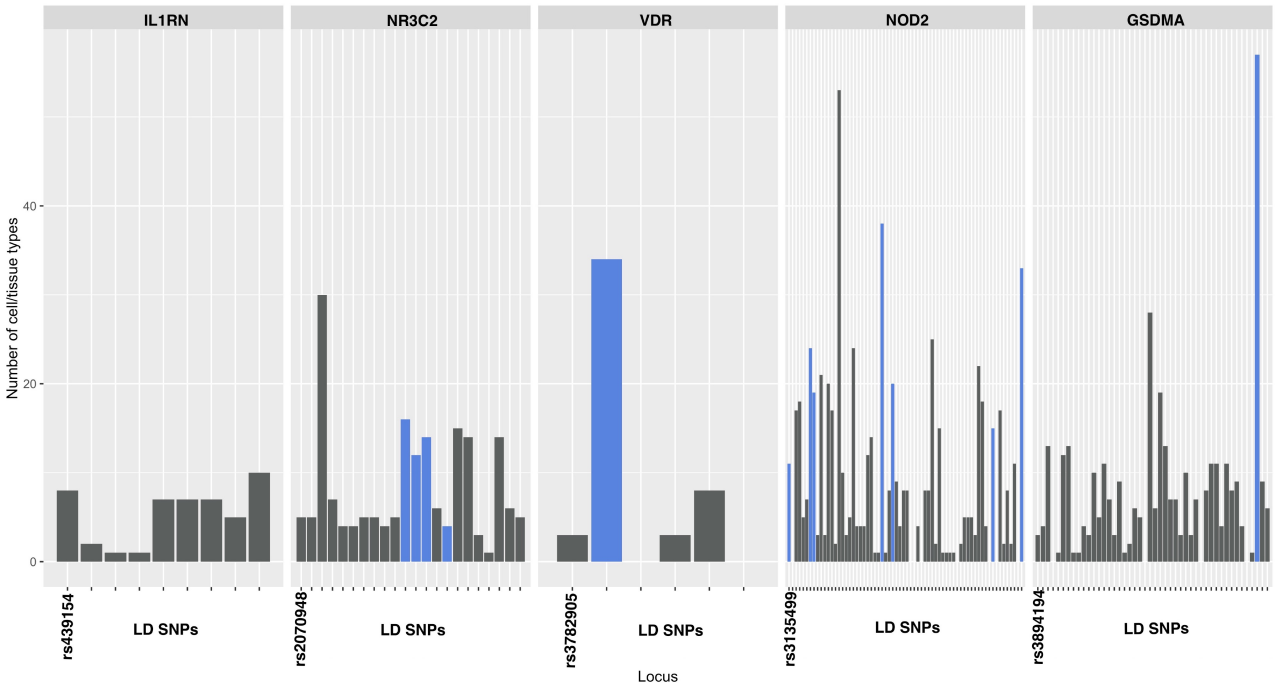


Figure-7b: Enrichment of enhancer associated histone modification mark (H3K27ac) in multiple cell/tissue types. Blue highlighted bar represents the SNPs that are enriched in multiple cell/tissue types including the lung. Prime SNP indicated followed by its LD SNPs. Note: Number of high-LD (r2 > 0.5) SNPs varies for each SNP.

Association catalog: Literature-based mining using the recent NHGRI-EBI GWAS catalog showed that of the 149 SNPs queried, SNPs from the 17q21 locus were known to be genome-wide associated with the highest odds for childhood-onset asthma, along with other traits such as asthma with hay-fever, white blood cell count and asthma status (Table-8).

Lead SNP	Chr ↑	LD	LD SNP	DISEASE/TRAIT	p-value	OR	PMID
rs3894194	17	0.8	rs7212938	Asthma and hay fever	4.00E-10	1.16	24388013
rs3894194	17	1.0	rs3894194	Asthma (childhood onset)	3.00E-21	1.59	24241537
rs3894194	17	0.5	rs4794820	Asthma	1.00E-08	1.33	22561531
rs3894194	17	0.6	rs3859192	White blood cell count	2.00E-12	0.14 ¥	22037903

Lead SNP	Chr †	LD	LD SNP	DISEASE/TRAIT	p-value	OR	PMID
rs3894194	17	1.0	rs3894194	Asthma	5.00E-09	1.17	20860503

Table-8: NHGRI-EBI GWAS catalog of genome-wide genotype-phenotype associations. † Chromosome; LD = Linkage disequilibrium; publication (ID) PMID; ψ Beta value.

Canonical pathways and regulatory networks: Table-9 shows the most significant canonical pathways across the entire dataset from the IPA knowledge base. Glucocorticoid receptor signalling was the top pathway with two molecules contributing (Ratio) significantly to the overall pathway network estimated by Fischer's exact test.

Pathway	p-value	Ratio	Overlapping molecule
Glucocorticoid receptor signalling	2.95E-03	2/272	*NR3C2, IL1RN*
Role of hypercytokinemia in Influenza pathogenesis	1.29E-02	1/41	*IL1RN*
Graft Vs Host Disease signalling	1.38E-02	1/44	*IL1RN*
Role of cytokines in mediating communications between immune cells	1.63E-02	1/52	*IL1RN*
IL-10 signalling	2.13E-02	1/68	*IL1RN*

Table-9: List of the genes in the top canonical pathways.

5. Lookup of supplemental data from GABRIEL study

Different factors are held responsible for the unexplained or missing heritability in asthma genetics like imprecise phenotype definition, sample heterogeneity, etc. Of these shortcomings, the bias in SNP selection based on the allele frequency (MAF > 5%) and omitting low frequency and rare variants has the greatest impact as they are much more pronounced in the genome than previously anticipated. This section aims to stress their importance by using supplemental data from the largest GWAS (10,365) on asthmatic individuals (GABRIEL study) and study such rare variants, which are omitted in a typical GWAS setting as part of the quality control procedure. However, considering the low error rate of genotyping chips (< 0.01%) it is now considered as being inadequate. Upon reanalysis of the supplemental data, rare variants were seven-fold enriched in terms of the effect estimates and makes up to 45% of the total genome-wide significant SNPs (n=109) that satisfies the significance threshold (P <= 7.20 x 10-8) (Figure-8). They are distributed amongst more than 50 unique genes with odds ratio scaling up to 50. In addition to the loci reported in the original study, new variants with stronger effect sizes were identified such as the *DDR1*, a tyrosine kinase receptor that is phosphorylated by collagen; *PERP*, a component of intercellular desmosome junctions; *FOXP2*, a regulator of lung development; and *CYP2A6*, the primary enzyme responsible for the oxidation of nicotine.

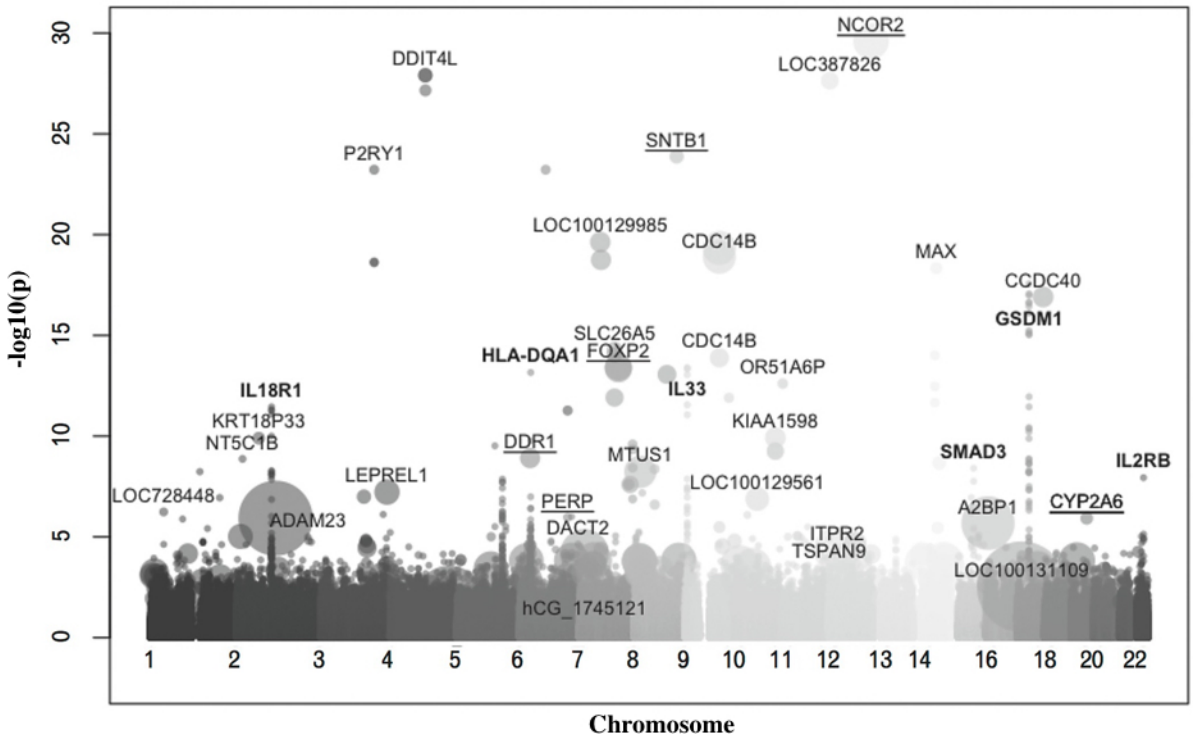


Figure-8: Asthma risk by rare variants from the Gabriel study. Manhattan plot showing the significance of SNP association from the fixed effects model. Size of the symbol correlates with the odds ratio (OR) magnitude (if OR greater than 1 or 1/OR if OR less than 1). Genes that are reported in the original publication are depicted in bold and locus with strong effect sizes SNPs are depicted by underline.

6. Discussion:

In this thesis, I demonstrated that a better definition of the phenotype for a complex disease like asthma is necessary and showed the plausibility of an intermediary phenotype like the circulating levels of serum cytokines as an alternative for a better understanding of the underlying pathology. ASP analysis is an effective study design in examining allele sharing and in identifying disease-associated genetic markers, additionally by studying common variants (n=550) that are shown to have a moderate effect on asthma phenotypes it gives enough statistical power in detecting true associations even under small sample size. By this approach, I show the genetic association of key cytokines that are involved in inflammatory pathways as well as functional evidence from publicly available datasets. With additional functional validation, the identified variants can be listed as potential biomarkers in the diagnosis, considering the ambiguity in the clinical diagnosis of asthma and its myriad causes. Three different analyses were performed, as the outcome of single part analysis may be misleading. The approach followed a stepwise approach to elucidate the inheritance of quantitative traits. In spite of the differences, each analysis corroborates each other and allows conclusions on the genetic influences on serum cytokines.

The samples for this study were collected in two stages and higher serum levels of the chemoattractant cytokine IL-8 were observed in the second part of the study (details in Materials and Methods section). IL-8 levels are strongly correlated with elevated neutrophil levels in airway secretions of acute asthma patients [70]. Hull et al. [71] first reported IL-8 in virus-inflicted respiratory illness while later studies associated IL-8 with asthma and concluded its role in genetic predisposition to asthma [72]. Noticeably, the heritability estimates of IL-8 (h^2 =1.00) is higher than with any other cytokine, possibly reflecting unapparent infections running in the families. Nevertheless, reported high heritability of other cytokines like TNFα, IFNγ and IL-1b support the previous arguments that SNPs in these regions affect allergen-induced cytokine production [73].

6.1. Implications from linkage analysis

Though several genome-wide linkage studies have been completed for asthma and atopy-related phenotypes, to the best of my knowledge this is the first report of genome-wide screening in relation to the serum cytokine levels. Except for a few that used population-based semiparametric [74] and parametric methods [75], most of the studies utilised non-parametric sib-pair linkage strategies. In particular, few studies have used the ASP structure with stringent selection criteria for affected subjects to minimise the type I (false positive) and type II (false negative) error rate [55,76]. ASP being a model-free method requires no assumption on the mode of inheritance and is more informative in the presence of incomplete penetrance. By performing ASP linkage analysis of serum cytokine levels and following the recommended criteria for 'significance' by Lander and Kruglyak (1995) [77], the present work found genome-wide significant evidence for linkage in one region (17q24.3, LOD=4.05; crude p-value = 7.80E-06), along with suggestive significance (6p25.3, LOD=2.56; crude p-value = 3.00E-04) and few other potentially interesting regions (6p24.3-25.3, 9p24.2-24.1, 12q13.11). Previous studies have reported suggestive evidence of linkage in the 17q24.3 region for SPT [78] and eosinophil count [79] in French and Dutch families respectively. The later meta-analysis of genome-wide linkage data for asthma and related phenotypes [12,13] using different bin widths (bins of fixed centimorgan window sizes) confirmed the significance across a broad region of 17q for atopy-related phenotypes (atopic asthma, SPT). In addition, the 17q region is known to be linked with other atopic diseases like dermatitis [80] and includes the 17q21 asthma susceptibility locus (*ORMDL3/ GSDMA*) that is identified by genome-wide association study of single nucleotide variants [30]. Interestingly, in the present study, the evidence for linkage in the 17q region was genome-wide significant for IL-12p40 cytokine levels and showed nominal linkage signals for the 17q21 region. Sharing of identical regions across multiple phenotypes might be the result of pleiotropic effects, exerted by genes residing in the region, which is common amongst studies investigating intermediate traits. Few other regions showed suggestive (6p25.3) and marginal (6p24.3) evidence of linkage for serum eotaxin levels. Multiple genome-wide screens have already reported the 6p region to be linked with asthma and related phenotypes [81],

making it the most replicated region. Refinement of this region showed 6p24.3 to be linked with asthma in Danish population [82], and 6p21 with asthma and bronchial hyper-responsiveness in a collaborative study involving multiple independent samples (Chicago, Hutterite, Dutch families) [28]. The other interesting region is 9p24, which showed marginal evidence of linkage for IL-12(p40) levels. Recent large-scale GWAS showed SNPs in this region flanking the IL-33 gene to be genome-wide associated with asthma [33,34].

The most convincing finding of our study is the evidence of linkage to IL-12(p40) levels at D17S949 on chromosome 17q24.3, and is located about 6cM from a locus (*BP14*; blood pressure QTL 14) that is known to function as quantitative trait loci (QTL) for blood pressure changes in murine models [83]. Interestingly D17S949 is known to exhibit significant allele sharing in affected sib pairs for essential hypertension [84]. Apart from the long established relationship between asthma and hypertension [85], both these entities involves spasms of smooth muscles. This similarity proposes the possibility of one disease type predisposing the other under specific stimuli, as acute severe asthmatics exhibit elevated blood pressure [86] and experimental evidence exists to show the relation between pro-inflammatory cytokines and hypertension [87]. However, it is equally important to remember the effect of β2 – Agonists (Bronchodilators) on inducing blood pressure changes [88]; this, along with other environmental confounders, makes it difficult to attain a definitive conclusion. The other gene that is in proximity with D17S949 is the *SSTR2* - a G protein-coupled receptor (GPCR) subtype for somatostatins (SST), that is vital in the initiation of physiological actions of SSTs. Somatostatins are known to function as an integral part of the immunoregulatory circuit by limiting the production of IFNγ in the granuloma cells of chronic inflammatory mice models [89]. Furthermore, from the expression studies on granuloma cells, it is known that of all the somatostatin receptor subtypes, mRNA for the *SSTR2* is expressed more, blocking its activity with anti-*SSTR2* antiserum prevents the IFN-inhibitory effect of SSTs particularly in the T-cells [90]. Based on in vitro cell-specific localization studies (using immunohistochemistry) it is known that *SSTR2* localises more in the human immune cells and lung fibroblasts [91], and that binding of somatostatin or its analogues on such receptors attenuated inflammatory airway hyper-responsiveness [92], bleomycin-induced pulmonary fibrosis [93]. Altogether, it suggests the central

role of GPCR receptors like *SSTR2* in orchestrating airway inflammation and remodelling airway architecture.

17q23.2 also harbours potassium ion (K+) channel genes (*KCNJ2*) and the mitogen-activated protein kinase kinase (*MAP2K6*) gene. K+ channel genes have been shown to be involved in human airway smooth muscle cell (HASM) hyperplasia (increased smooth muscle bulk) and contribute to airway remodelling by promoting the proliferation of HASM [94]. Although the initial ion channel studies in asthma primarily focused on voltage-gated Ca2+ channels, there is a recent shift of focus towards K+ channels owing to its modulatory effect on membrane potential by regulating the activation of other ion channels that are involved in smooth muscle contraction. *MAP2K6* regulates the transcription and activation of pro-inflammatory genes by phosphorylating and activating protein kinase (p38) in response to inflammatory cytokines and other environmental stress. Features that are central to the pathophysiology of asthma such as eosinophil apoptosis, Th2 cell-mediated IL5 and IL-13 synthesis, and the expression of IgE receptor are all known to be associated with p38 *MAP2K6* activation [95]. Its importance is further demonstrated by the reversal of allergen-induced complications [96] and corticosteroid insensitivity [97] during asthma treatment by inhibiting p38 MAPK. Potential therapeutic applications targeting potassium channel openers/modulators and inhibiting the p38 protein kinase pathway have already been discussed in the treatment of complex lung disorders [98].

The other noteworthy linkage signal is that of the microsatellite marker D6S344 at chromosome 6p25.3 with eotaxin levels. D6S344 are located closely to the serine protease inhibitor genes (serpins) and are known to inhibit chymotrypsin-like proteases that are involved in regulating the release of reactive oxidants [99]. Dysfunctioning serpins can thus result in the excessive release of reactive oxygen species, bringing oxidative stress along with the depletion of antioxidant defences. In vivo studies already show the outcome of oxidative stress with increasing degrees of airway hyper-responsiveness in guinea pigs [100].

6.2. Implications from association analysis

In the family-based association test for quantitative traits (QFAM), intronic variants of the mineralocorticoid receptor (*NR3C2*) and interleukin 1 receptor antagonist (*IL1RN*) gene satisfied the significance threshold ($p < 0.05$) for association with serum cytokine levels. The significance is based on the permutation procedure to control for the dependence of individuals within the same family. By preserving the correlational structure between SNPs, the permutation scheme provides a less stringent correction for multiple testing compared to the bonferroni that assumes all tests as independent [62].

Stress is a known risk factor for asthma, as stress-induced cortisols secreted by the neuroendocrine (HPA: hypothalamic–pituitary–adrenal) organs controls over the Th1/Th2 balance, favouring predominantly type 2 response [101]. Cortisols exerts its effect by binding with the mineralocorticoid (MR) (Eg. NR3C2) and glucocorticoid receptors (GR) (Eg. NR3C1). *NR3C2* is a highly polymorphic gene and functions as receptors for both the mineralocorticoids and glucocorticoids response elements and were chosen for investigation due to its implication in psychological stress. In vitro assays have already shown that SNPs residing in these genes regulate cortisol levels [102] by modulating the HPA (hypothalamic–pituitary–adrenal) axis. Equally, it is important to remember that stress may not modify the immune response by its own, rather it accentuates the inflammatory response to environmental triggers. Therefore, the hypothesis is that variants in this genetic region might be associated with individual asthma susceptibility modulated by psychological stress. So far two studies have established the link between glucocorticoid receptor-gene (*NR3C1*) variants and asthma development [103,104], while yet to show the association of the mineralocorticoid receptor gene (*NR3C2*). In this study, an intronic variant (rs2070948) of *NR3C2* was found to be significantly associated with quantitative asthma phenotypes (serum levels of eotaxin, IFNγ, and IL-13); highlighting the importance of corticosteroid pathway in asthma pathology.

In particular, the association of *NR3C2* SNP (rs2070948) with IFNγ is interesting, as it is long known that stressed subjects have a significantly higher production of IFNγ compared to subjects with low-stress perception [105]. Furthermore, in vivo studies demonstrated increased IFNγ production along with enhanced airway hyperreactivity in ovalbumin-sensitized and stressed mice [106],

strengthening the hypothesis of psychological distress being perceived by the immune system through the secretion of pro-inflammatory cytokines. This association of mineralocorticoid system with pro-inflammatory cytokines complements the in vivo molecular evidence of arterial leukocyte infiltration and increased expression of pro-inflammatory markers [107] as a result of the infusion of mineralocorticoids such as aldosterone. The exact mechanism by which aldosterone affects inflammatory processes is best illustrated by the in vitro knockdown of MR in human aortic endothelial cells resulting in decreased aldosterone-mediated endothelial exocytosis - the initial response in leukocyte trafficking to tissue sites of inflammation [108].

Additionally, this study demonstrates that the possession of *IL1RN* rs439154 polymorphism major allele (GG) is significantly associated with IL-23 levels along with the nominal associations of IL-33 and TSLP levels. More importantly, haplotypes estimated from this SNP region were known to be associated with a higher prevalence of asthma [109] and high total IgE levels [110] in our earlier studies on German family samples and German adult populations respectively. The fact that previous studies have analysed the effect of SNPs in this region with atopy-related traits ensure that these are likely associations and not false positives. *IL1RN* belongs to the IL-1 family of cytokines involved in inflammatory processes and encodes for the IL-1 receptor antagonist (IL1RA) polypeptide. IL1RA protein exhibits anti-inflammatory properties by binding competitively with the pro-inflammatory IL-1 receptor sites and inhibiting the inflammatory cascade (downstream). Thus, depending on the *IL1RN* expression range, two possibilities exist, lower level of IL1RA activity as a result of reduced *IL1RN* expression promoting an excess inflammation or increased expression resulting in higher levels of IL1RA activity and reduced inflammatory response. The protective role of IL1RA is confirmed by experimental data, where IL1RA treatment of antigen-sensitized animals exhibited the reversal of histamine-induced bronchial hyperreactivity [111]. Maintenance of this balance between pro- and anti-inflammatory systems have already been explored in in vivo models [112,113].

Of the multiple cytokines associated with *IL1RN*, the association with the levels of IL-33 is interesting as it plays an agnostic role in inflammatory cascade by signalling towards Th2-driven inflammation. IL-33 belongs to the IL-1 family of

cytokines, released mainly by damaged and dying (necrosis) cells. Released IL-33 are often in their active state, and by binding to their respective receptors stimulate early inflammatory responses. Additional properties of IL-33 include co-stimulating Th2 cytokines with TSLP and regulating maturation of mast cell precursors to enhance the production of other pro-inflammatory agonists [114].

6.3. Implications from insilico annotations

The main challenge in genetic association studies is establishing the causal link between the SNP and the gene and is particularly challenging for the non-coding variants. In such cases, the LD structure of the lead variant can be helpful in identifying cell type-specific eQTLs or regulatory regions of the gene, which potentially links the variants to the genes involved in disease pathogenesis. In this study, publicly available databases on gene expression, regulatory elements, and biological networks were queried to predict the functional effects of loci. As expected, missense variants in the *GSDMA* locus in addition to functioning as significant eQTL for the same gene, lead to amino-acid changes with deleterious protein-altering effect. Interestingly, high-LD SNPs of the intronic variants residing in *IL1RN* and *NR3C2* are associated with its differential expression in blood cells and immune cells. The annotation relied mainly on the eQTL data, as a large proportion of the common disease variants identified so far seems regulatory [115]. However, there is a considerable variation in the sample sizes in eQTL studies thereby limiting the power to detect true eQTLs, more importantly, limited by the availability of disease-relevant tissues. Ongoing efforts of profiling eQTLs from different tissues by the Genotype-Tissue Expression (GTeX) consortium should contribute in the identification of eQTLs in asthma-relevant tissues [64]. Furthermore, *NR3C2* SNPs are significantly enriched for the histone modification marks that are associated with promoter and enhancer activation in multiple cell types (mainly the lung), emphasising the regulatory effect of these SNPs on the gene expression. Recent estimates by the ENCODE consortium shows that only approximately 27% of the regulatory elements (histone marks) interacts with the nearest promoter [116], suggesting the closest gene need not be the target for a given SNP association. This brings to one of the limitations of the current pipeline as it does not include results from chromatin-capture experiments that assess long-range interactions. Both

experimental data and prior knowledge were used in the present functional annotation pipeline. Noticeably, the *GSDMA* locus exhibits functional relevance in almost all the annotation steps, except for the network-based analysis. The use of previous biological knowledge in the pathway analysis could contradict the hypothesis-free nature of genome-wide association signals (Table-8), possibly explaining the failure to detect the GWAS-identified *GSDMA* locus in the network analysis. On the other hand, glucocorticoid receptor signalling pathway is the top biological pathway contributed by *NR3C2* and *IL1RN*, further stressing the importance of corticosteroid pathway in asthma pathology. The annotation efforts exhibit the complexity in the downstream effects of a locus, as it does not fit clearly into a function or pathway and identifying the mechanistic relevance of these loci to the trait of interest requires additional experimental assays.

6.4. Implications from the lookup of major meta-analysis data

The main focus of the lookup is to address the missing heritability seen in genetic association studies of asthma and discuss the need for identifying variants that explain the main pathology. Apart from multiple factors attributed to the unexplained heritability, importance is given to the rare variants which are omitted in a typical GWAS setting and that provides the highest risk estimates. To study such variants, supplemental data from the largest asthma GWAS to date (GABRIEL study) was used [33]. Rationale of including low frequency (MAF < 1%) variants in this lookup comes from the fact that error rate of genotyping chips is reported to be less than 0.01%. Being said that calling of these rare variants and any future implications need to be verified by targeted deep sequencing strategies. Along with hits already reported in the original publication (*HLA-DQA*, *IL-33*, *SMAD3*, *GSDMA*, *IL18R1*, *ILR2B*), there were new hits identified with much larger effect estimates (*DDR1*, *PERP*, *FOXP2*, *CYP2A6*). Discoidin domain receptor 1 (*DDR1*) is a receptor tyrosine kinase that lies in proximity with several HLA class I genes and are activated by the binding of collagen components (ligand) of the extracellular matrix. Experimental evidence shows that they are strongly over-expressed in the basolateral surface of the bronchial epithelium [117], colocalized with its ligand. Furthermore, invivo studies on *DDR1* knockout mice have demonstrated resistance to lung inflammation and fibrosis that are induced by bleomycin administration [118], emphasising *DDR1* role

in cytokine production and the downstream inflammatory processes. It is well known that consistent inflammation brings remodelling of the airway epithelium, leading to clinical symptoms of asthma, and thus it is expected to see significant differences in *DDR1* expression between asthmatics and non-asthmatics. However, given the fact that *DDR1* is strongly expressed in the epithelium of bronchial biopsies from individuals with and without asthma [119], suggests that *DDR1* might play a regulatory role in the inflammation through means other than the epithelial damage repair process.

The other interesting hit from the lookup is the gene involved in the drug metabolism (xenobiotics) (*CYP2A6*). Cytochrome P450 is the primary enzyme responsible for the oxidation of nicotine and in the metabolism of several pharmaceuticals and carcinogens [120]. Genetic polymorphisms residing in this gene has already been implicated in multiple outcomes like smoking habits [121] and pulmonary emphysema [122]. Though initial reports in high ranking journals have associated the *CYP2A6* polymorphism with nicotine metabolism and tobacco dependancy [121], recent studies with novel and advanced genotyping methods show additional loci (*EGLN2*) [123] affecting the smoking behaviour. Marked inter-individual variation in the activities of *CYP2A6* [124] and possible suppressed expression by inflammatory stimuli place them as an interesting candidate for studying its association with asthma.

As mentioned earlier, the primary technology for the detection of rare variants with high confidence is by targeted deep sequencing of the region of interest. Additional cost-effective approach such as inferring (imputation) based on a comprehensive genome catalogue like the 1,000 Genomes Project (http://www.1000genomes.org/page.php) [125], can facilitate the extension of already-genotyped SNPs to additional common and rare variants. Nevertheless, when it comes to the interpretation, it is imperative to remember that such events are rare possibly involving other mediators that modulate the disease process. Thus it is highly unlikely that few functional variants can explain the phenotype under study. Complicating matters further, the sample size that is required to detect low-frequency variants with high-risk association scales roughly linearly (1/MAF) requiring larger N as the variants become rarer [126]. Inter-individual differences also pose a significant challenge in finding appropriate subjects, as individuals with rare variants at one

locus may have ancestral differences in allele frequency. In addition to these problems, statistical approaches used in evaluating common variants are not applicable for rare variants. On the other hand, structural variations like copy number variants (CNVs), inversions, repeat genetic elements and translocations could explain some of the missingness. Similar to trait-associated SNPs, structural variants associated with the disease can either be rare with larger effect sizes or common with modest effects, and various approaches have been developed to identify and integrate them in GWAS such as exploiting the LD relationship between SNP and common CNVs [127]. Of note, common CNVs and de novo CNVs, seem highly unlikely to contribute to family resemblance and heritability, but could emphasise the involvement of an active environmental component. Specifically, it is worthwhile to mention that the current estimates of heritability for common traits can be significantly inflated due to the underlying assumption of no gene-gene or gene-environment interactions. In conclusion, the next wave of asthma genetic studies will probably not so much on improving genotyping chips but use full-genome sequencing analysis supported by family-based studies, and learn from phenotypically well-defined data by exploring in the context of regulatory annotations.

7. Potential role of structural variants like transposable elements in COPD onset:

7.1. Key regulators in COPD

COPD is characterised by the co-occurrence of bronchial inflammation (bronchitis) and alveolar wall destruction (emphysema). The main clinical feature associated with COPD is the reduction of forced expiratory volume in one second (FEV1) along with the decrease in the ratio of FEV1 to forced vital capacity (FVC) [128], resulting in obstruction of the airways due to reduced lung volume and lesser contractile response of the airway walls. Though environmental influence such as air pollution plays a role in their onset, smoking is regarded as the most common cause of COPD. Pathophysiologically, it involves inflammatory cell infiltration (neutrophils, macrophages) and hyperplasia of the airway smooth muscle cells (ASMC), releasing proteolytic enzymes, which generate oxidants and thereby imposing oxidative stress [129]. ASMC and the mucociliary system (goblet cells and ciliated cells) are the primary target of action for these inflammatory mediators. As biopsies of bronchial airways and small airways from COPD patients show abnormal goblet cell differentiation (metaplasia), histopathologic changes of the small airways and mucus hypersecretion with the loss of mucociliary clearance [130]. These changes manifest as what is known to be the hallmarks of COPD, namely emphysema (airway remodelling and obstruction) and bronchitis (narrowing of bronchial walls – inflammation and mucus hypersecretion). Fibroblasts are also involved in the airway remodelling as they proliferate under the stimulus of transforming growth factor-beta (TGF-β) from epithelial cells resulting in fibrosis in the small airways [131]. The airway wall thickens as a result of this remodelling process which narrows the airway lumen causing irreversible airflow limitation and increased airway hyperreactivity.

7.2. Developmental control of transposable elements

In complex gene-environment interaction, the genome is continuously challenged by deleterious genotoxic events, which are efficiently evaded by the host surveillance networks. Their failure triggers genomic instability, and many such events were reported in complex disorders such as COPD [132]. Global distribution of transposable elements (TEs) and their ability to translocate (retrotranspose) to one position from another poses a greater challenge to maintain this integrity [133]. TEs such as Long Interspersed Nuclear Elements (LINEs or L1s) introduce somatic mutations during translocation and lead to post-insertional genomic instability events [134], but their role in complex disorders like COPD is unclear. The possibility of retrotransposition events during embryogenesis and affecting the integrity of developmental genes in the early growth phase is an interesting proposal. Any genetic damage occurring during the early growth stage of the lung could result in susceptibility to damage either by the host machinery (protease/anti-protease imbalance) or environmental factors like cigarette smoking and pollutants. The theory is further supported by L1 insertion mediated mosaicism seen in somatic and germline cells [135], and the occurrence of L1 RNA in embryonic cells [136]. Moreover, quantitative studies have already shown the frequency of retrotransposition to be higher in reproductive cells [137]. For instance, epigenetic changes during lung development are known to play a vital role in the development of bronchopulmonary dysplasia (BPD) [138]; this can pose as a risk factor altering the lung morphology in early childhood and any associated lower lung functions can ultimately result in the development of COPD.

7.3. Epigenetics of transposable elements

DNA methylation is a heritable non-genetic change that in addition to regulating the gene expression has evolved to keep a check on the repeat elements and silence them. Almost one-third of the DNA methylation occurs in TEs like Alu elements and L1s [133], and since the TEs account for nearly half of the genome, this can function as a representative marker for global methylation profile. Nevertheless, such sites are highly prone to environmental influences and can be hypomethylated, leading to genome instability and altered gene expression. For

instance, one study established such an association by showing altered gene expression as a result of hypomethylation in lung tissue samples from idiopathic pulmonary fibrosis patients [139]. As a further support to the hypothesis hypomethylation of L1 elements is associated with a rapid decline in lung function measures (FEV1 and FVC) [140]. Lung function tests being a major determining factor in diagnosing and measuring the severity of lung disorders, the impact of hypomethylation on lung function is intriguing.

7.4. Oxidative stress and hypomethylation

Oxidative stress as a result of oxidant/antioxidant imbalance plays a central role in the pathogenesis of COPD (Figure-9). Under stress conditions, oxidation susceptible GC-rich sites oxidize to guanyl neutral radical [141] that react with superoxides from the cigarette smoke to form a stable oxidation product (8-OHdG) [142]. 8-OHdG, a marker of oxidative stress, plays a vital role in altering the methylation profile by inhibiting the binding capacity of DNA methyltransferase to CpG islands, resulting in the hypomethylation of guanine and cytosine residues. The hypomethylation property of 8-OHdG is further strengthened by its ability to cause transversions (G > T) and reduce the methylation hotspots (CpG dinucleotides) [143]. Interestingly, 5'-UTR region of the L1 mRNA is GC rich (approximately 60%), which makes it highly susceptible to oxidative stress mediated hypomethylation. Of the various environmental factors, smoking is considered to play a decisive role, as prenatal exposure to tobacco smoke is associated with global hypomethylation in adulthood [144]. Similar, evidence show that cigarette smoking and other environmental factors like particulate air pollution can synergistically determine the overall methylation status [145].

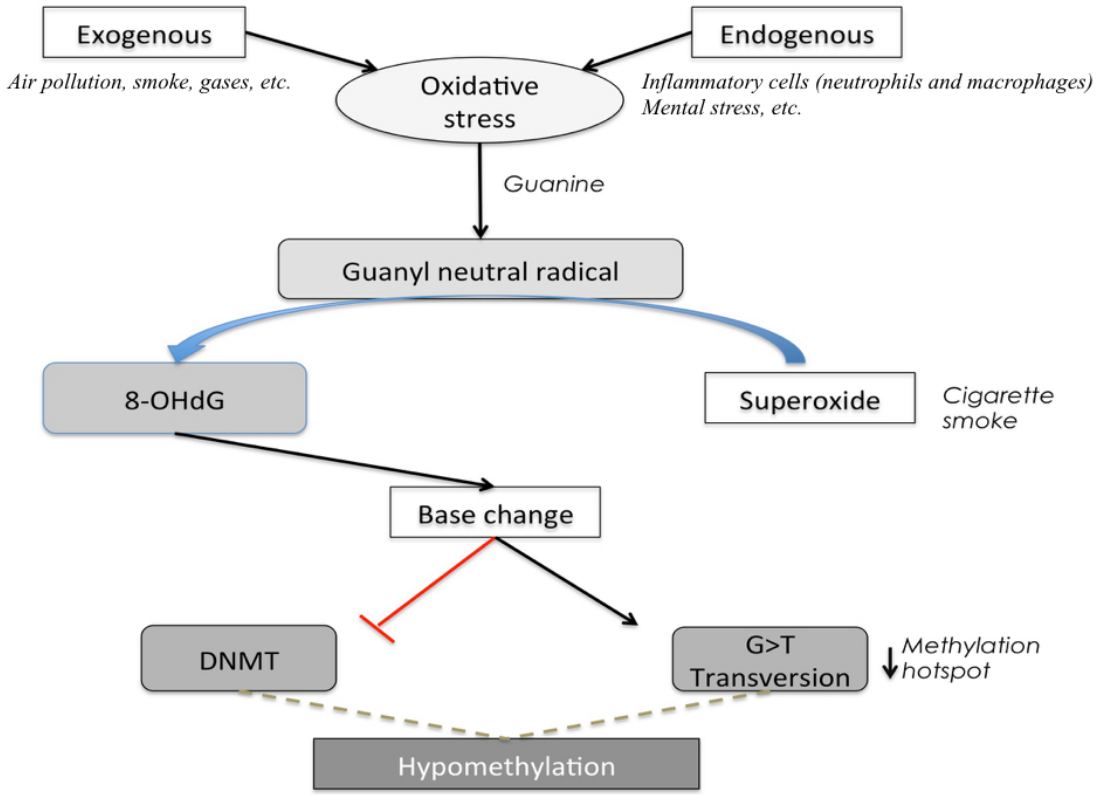


Figure-9: Oxidative stress mediated by exogenous and endogenous sources.

Figure-10a discusses the effect of oxidative stress mediated hypomethylation in activating and transposing L1s that leads to deleterious structural alterations in the genome followed by a cascade of signalling events (Figure-10b). These events can result in the death of cells or inflammatory response or both leading to a continued decline in lung function. All these evidence strongly suggest that these are not isolated events in the COPD onset and that epigenetic changes mediated by oxidative stress play a central role.

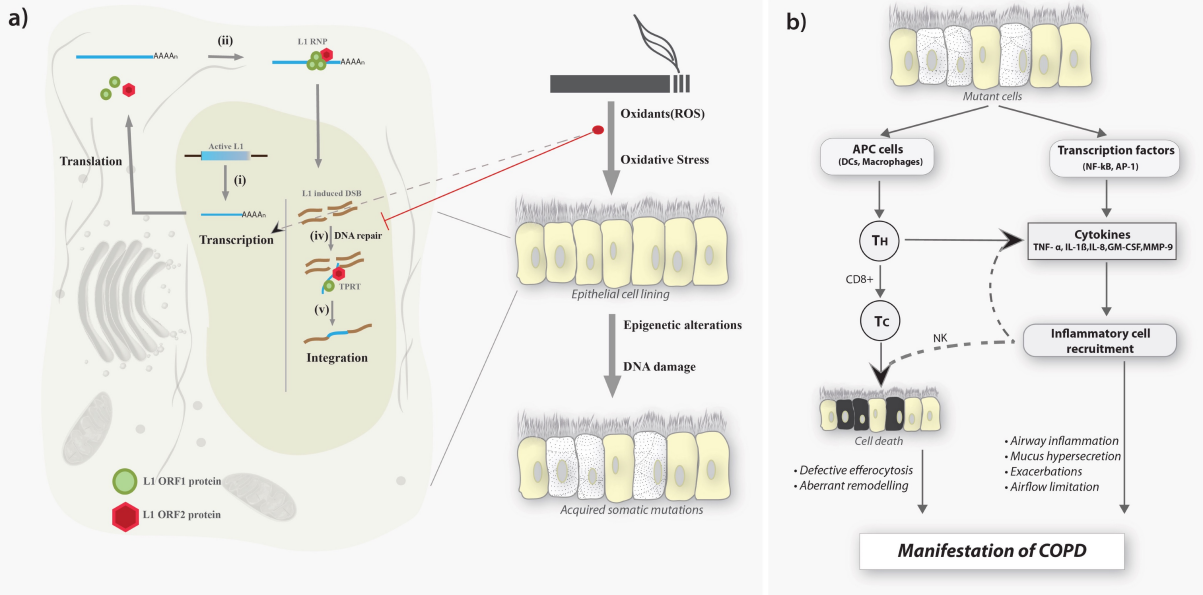


Figure-10a: L1 retrotransposon life cycle. L1 life cycle begins with the transcription of active L1, followed by polyadenylation to create L1 RNA. Spliced L1 RNA is later nuclear exported for translation. In the cytoplasm, L1 RNA coded proteins (ORF1 and ORF2) binds with other L1 RNA that are retrotranscription competent L1 (RC-L1) to form the ribonucleoprotein (L1 RNP) complex. L1 RNP is nuclear imported for the retrotransposition mechanism resulting in double-strand break (DSB) by the endonuclease activity of L1 ORF2. Lesions created by the digesting activity is repaired and integrated into the genome by reverse transcription (TPRT) method. Smoke particles and heavy metals can interact with L1 lifecycle either at the early stages by altering the methylation profile or at the late stages by impairing the repair pathway resulting in the accumulation of somatic mutations (Granulated cells). **10b: Somatic mutation on disease onset and exacerbation.** Mutated somatic cells are recognized as foreign cells by the host system and presented by the antigen-presenting cells (APCs), triggering the T helper (Th) and cytotoxic T cells (Tc) that migrate to the infected sites inducing cell death. Failure in removing the dying/dead cells (efferocytosis) results in aberrant remodeling of the cells. On the other hand, mutant cells can increase the release of cytokines by interacting with transcription factors and recruit inflammatory cells to destabilize the immune balance, thereby manifesting the features of COPD.

7.5. Proposal for future studies of COPD and transposable elements

Distribution of TEs varies across individuals probably attributed to their activity in somatic tissues and the low selection pressure encountered by these elements, enabling them to evolve rapidly and making their identification arduous. However, with the advent of next generation sequencing technologies, different algorithms for detecting the inserts were proposed [146]. Earlier methods rely on prior information and identify bonafide TEs based on sequence (homology-based) and structural (structure-based) similarity [147]. However, they are prone to bias in identifying TEs belonging to the previously identified families. De-novo approaches, on the other hand, attempts to discover new TEs using their unique repetitive nature [148]. Figure-11 discusses the main theme of this approach. Research interest in structural variants has increased exponentially over the past decade, and so far approximately 5,000 insertions have been reported. Database of Retrotransposition Insertion Polymorphism (DbRIP) represents a comprehensive non-redundant list of such variants (SINE, Alu and LINE) [149].

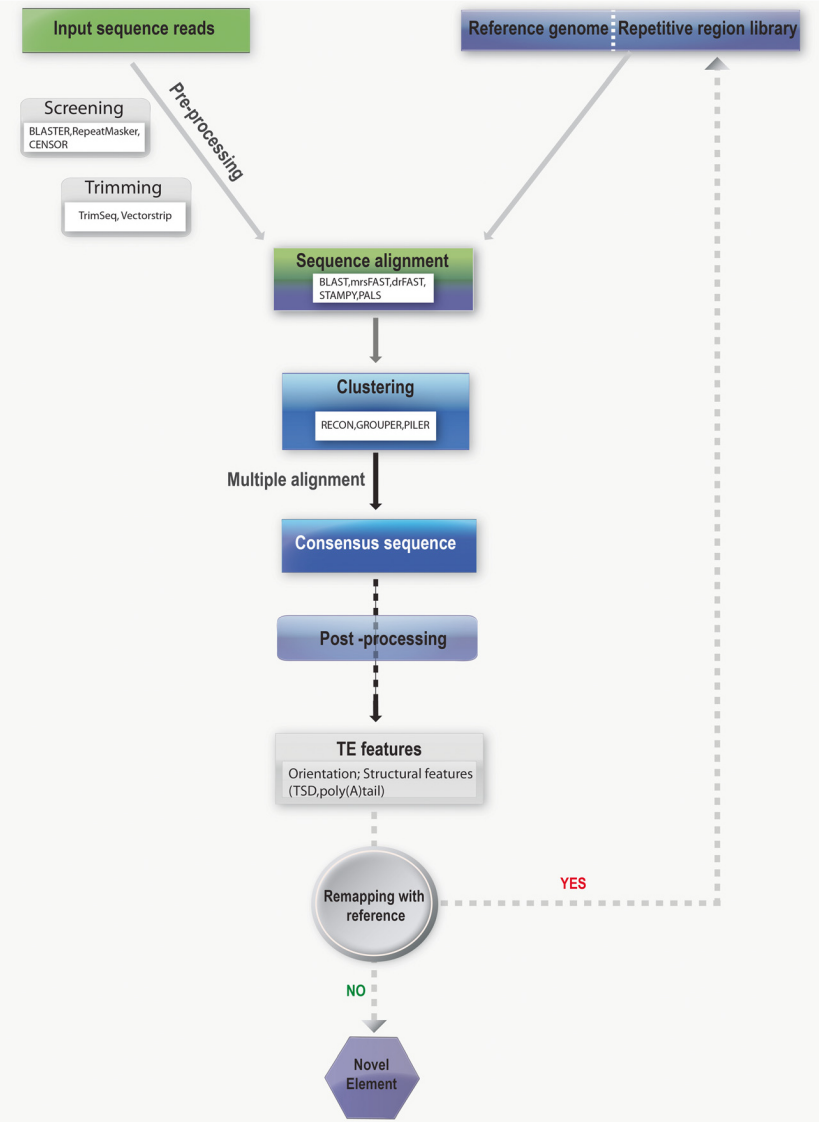


Figure-11: General scheme of pipeline in identifying repeat elements. (White inset boxes - Examples of available tools). Sequence reads are pre-screened for the presence of TEs, and cryptic structures (poly (A) tail) are trimmed from the sequence to avoid excessive mismatches. Followed by mapping against the reference genome and repeat library to form clusters of consensus sequences. In the post-processing step, consensus sequences are remapped with the reference based on the characteristic features of TEs. This yields either concordant (YES) or discordant combinations (NO). Concordant combinations are the elements that are already in the library while the discordant combinations represent the putative novel elements.

8. Synthesis:

The major challenge of addressing asthma is the ambiguity observed in the clinical definition and the overlapping phenotype with other pulmonary disease. The approach used in this study is more of a multi-dimensional approach, where an alternate phenotype in the form of serum cytokine levels is proposed, and their genetic relations are explored. By using a family-based sample of 923 individuals from 218 asthma families, the study has enough statistical power to detect common variants associations. Primary outcomes include the confirmation of the immune-regulatory role of the 17q24.3 region by identifying its strong linkage with IL-12(p40) levels. Furthermore, IL-12(p40) along with IL-33 and IL-23 are shown to be associated (novel) with intronic variants of *IL1RN* and *NR3C2* gene, which is not otherwise detected using clinical diagnosis alone. After performing annotation analysis using a custom pipeline, functional and regulatory role of the identified variants were confirmed. Further functional studies are needed for a better understanding of the molecular mechanism that drives the diseases phenotype. It is to be mentioned that in this study three cytokines (IL-4, IL-5, and IL-8) were positively correlated with the intake of anti-asthmatic drugs, but considering the overall effect of drug intake, cytokine levels implicated in the association and linkage analysis remains unaffected by the treatment history. Though the study does not show replication of results in an independent case-control dataset, the huge difference in the proportion of asthmatics between parents and children (21.9%, and 88.7% respectively) can be seen as case-control sets.

The main strength of the study is the affected sib-pair structure that controls the confounding effect of environmental factors and provides the statistical power to study variants that are common in the population. The present work mainly argues that the intermediary phenotypes such as the serum levels of cytokines can function as an alternative to the classical clinical phenotypes, especially when the clinical criteria are ambiguous. As common variants contribute modestly to the phenotype through their moderate effect size, association data from another asthma study are explored, leading to the identification of many more associated SNPs, all being rare variants with effect estimates scaling up to an OR of 50. Although this may be due to

some population stratification, intermediary phenotypes and rare variants might further explain some of the missing genetic heritability. This is explored and further discussed in another pulmonary disease type with strong environmental component, Chronic Obstructive Pulmonary Disease (COPD).

9. References:

1. Asher, M.I. and P. Ellwood, *"Global Burden of Disease due to Asthma"* in *The Global Asthma Report.* Global Asthma Network, 2014. **12**: p. 16.
2. Akinbami, O.J., et al., *Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010*. NCHS Data Brief, 2012. **94**(1): p. 1-7.
3. Von Mutius, E., *Environmental factors influencing the development and progression of pediatric asthma.* J Allergy Clin Immunol, 2002. **109**(6): p. S525-S532.
4. Jie, Y., et al., *Urban vs. rural factors that affect adult asthma.* Rev Environ Contam Toxicol, 2013. **226**: p. 33-63.
5. Cruz A.A., J. Bousquet and N.G Khaltaev, *"Chronic respiratory diseases: a major global health problem" in Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach.* World Health Organization, 2007. **1**: p. 5.
6. Barrett, P.M., A. Alagely and E.J. Topol, *Cystic fibrosis in an era of genomically guided therapy.* Hum Mol Genet, 2012. **21**(R1): p. R66-R71.
7. Cookson, W.O., *Asthma genetics.* Chest, 2002. **121**(3 Suppl): p. 7S-13S.
8. Meyers, D.A., et al., *Two locus segregation and linkage analysis for total serum IgE levels.* Clin Exp Allergy, 1995. **25**: p. 113-15.
9. Kimura, K., et al., *Linkage and association of atopic asthma to markers on chromosome 13 in the Japanese population.* Hum Mol Genet, 1999. **8**(8): p. 1487-90.
10. Postma, D.S. and G.H. Koppelman, *Genetics of asthma: where are we and where do we go?.* Proc Am Thorac Soc, 2009. **6**(3): p. 283-87.
11. Ober, C. and T.C. Yao, *The genetics of asthma and allergic disease: a 21st century perspective.* Immunol Rev, 2011. **242**(1): p. 10-30.
12. Denham, S., et al., *Meta-analysis of genome-wide linkage studies of asthma and related traits.* Respir Res, 2008. **9**(38): p. 1-12.
13. Bouzigon, E., et al., *Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy.* Eur J Hum Genet, 2010. **18**(6): p. 700-06.
14. Hansel, N.N. and G.B. Diette, *Gene expression profiling in human asthma.* Proc Am Thorac Soc, 2007. **4**(1): p. 32-36.
15. Graves, P.E., et al., *A cluster of seven tightly linked polymorphisms in the IL-13 gene is associated with total serum IgE levels in three populations of white children.* J Allergy Clin Immunol, 2000. **105**(3): p. 506 - 13.

16. Heinzmann, A., et al., *Genetic variants of IL-13 signalling and human asthma and atopy.* Hum Mol Genet, 2000. **9**(4): p. 549-59.
17. Symula, D.J., et al., *Functional screening of an asthma QTL in YAC transgenic mice.* Nat Genet, 1999. **23**(2): p. 241-244.
18. Rosenwasser, L.J., et al., *Promoter polymorphisms in the chromosome 5 gene cluster in asthma and atopy.* Clin Exp Allergy, 1995. **25**: p. 74-78.
19. Wang, W., et al., *A study on the relationship between interleukin-4 promoter polymorphism and asthma in a Xinjiang Uyger population.* Zhonghua Jie He He Hu Xi Za Zhi, 2004. **27**(7): p. 460-64.
20. Gervaziev, Y.V., V.A. Kaznacheev and V.B. Gervazieva, *Allelic polymorphisms in the interleukin-4 promoter regions and their association with bronchial asthma among the Russian population.* Int Arch Allergy Immunol, 2006. **141**(3): p. 257-64.
21. Berenguer, A.G., et al., *Genetic polymorphisms and asthma: findings from a case-control study in the Madeira island population.* Biol Res, 2014. **47**(40): p. 1-8.
22. Liu, X., et al., *An IL13 coding region variant is associated with a high total serum IgE level and atopic dermatitis in the German Multicenter Atopy Study (MAS-90).* J Allergy Clin Immunol, 2000. **106**(1): p. 167-70.
23. Hong, E.P. and J.W. Park, *Sample size and statistical power calculation in genetic association studies.* Genomics Inform, 2012. **10**(2): p. 117-22.
24. International HapMap Consortium., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
25. Ha, N.T., S. Freytag and H. Bickeboeller, *Coverage and efficiency in current SNP chips.* Eur J Hum Genet, 2014. **22**(9): p. 1124-30.
26. Visscher, P.M., et al., *Five years of GWAS discovery.* Am J Hum Genet, 2012. **90**(1): p. 7-24.
27. Weiss, S.T., B.A. Raby and A. Rogers, *Asthma genetics and genomics 2009.* Curr Opin Genet Dev, 2009. **19**(3): p. 279-82.
28. Nicolae, D., et al., *Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21.* Am J Hum Genet, 2005. **76**(2): p. 349-57.
29. Kikuchi, S., et al., *Association of tumor necrosis factor-α and neutrophilic inflammation in severe asthma.* Allergol Int, 2005. **54**(4): p. 621-25.
30. Moffatt, M.F., et al., *Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma.* Nature, 2007. **448**(7152): p. 470-73.
31. Cantero-Recasens, G., et al., *The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress.* Hum Mol Genet, 2010. **19**(1): p. 111-21.
32. Ober, C., et al., *Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function.* N Engl J Med, 2008. **358**(16): p. 1682-91.

33. Moffatt, M.F., et al., *A Large-Scale, Consortium-Based Genomewide Association Study of Asthma.* N Engl J Med, 2010. **363**(13): p. 1211-21.
34. Torgerson, D.G., et al., *Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations.* Nat Genet, 2011. **43**(9): p. 887-92.
35. Barnes, K.C., *Genetic studies of the etiology of asthma.* Proc Am Thorac Soc, 2011. **8**(2): p. 143-48.
36. Manchia, M., et al., *The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases.* PLoS One, 2013. **8**(10): e76295.
37. Simpson, A., et al., *Endotoxin exposure, CD14, and allergic disease: an interaction between genes and the environment.* Am J Respir Crit Care Med, 2006. **174**(4): p. 386-92.
38. Bouzigon, E., et al., *Effect of 17q21 variants and smoking exposure in early-onset asthma.* N Engl J Med, 2008. **359**(19): p. 1985-94.
39. Ege, M.J., et al., *Gene-environment interaction for childhood asthma and exposure to farming in Central Europe.* J Allergy Clin Immunol, 2011. **127**(1): p. 138-44.
40. Manuck, S.B. and J.M. McCaffery, *Gene-environment interaction.* Annu Rev Psychol, 2014. **65**: p. 41-70.
41. Weiss, S.T. and E.K. Silverman, *Pro: genome-wide association studies (GWAS) in asthma.* Am J Respir Crit Care Med, 2011. **184**(6): p. 631-33.
42. Eichler, E.E., et al., *Missing heritability and strategies for finding the underlying causes of complex disease.* Nat Rev Genet, 2010. **11**(6): p. 446-50.
43. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
44. Campbell, C.D., et al., *Whole-genome sequencing of individuals from a founder population identifies candidate genes for asthma.* PLoS One, 2014. **9**(8): e104396.
45. Ott, J., Y. Kamatani and M. Lathrop, *Family-based designs for genome-wide association studies.* Nat Rev Genet, 2011. **12**(7): p. 465-74.
46. Gordon, D., S.C. Heath and J. Ott, *True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms.* Hum Hered, 1999. **49**(2): p. 65-70.
47. Cheng, K.F. and J.H. Chen, *A simple and robust TDT-type test against genotyping error with error rates varying across families.* Hum Hered, 2007. **64**(2): p. 114-22.
48. Chen, M.H., et al., *Using family-based imputation in genome-wide association studies with large complex pedigrees: the Framingham Heart Study.* PLoS One, 2012. **7**(12): e51589.

49. Saad, M. and E.M. Wijsman, *Power of family-based association designs to detect rare variants in large pedigrees using imputed genotypes.* Genet Epidemiol, 2014. **38**(1): p. 1-9.
50. Moore, W.C., et al., *Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program.* Am J Respir Crit Care Med, 2010. **181**(4): p. 315-23.
51. Haldar, P., et al., *Cluster analysis and clinical asthma phenotypes.* Am J Respir Crit Care Med, 2008. **178**(3): p. 218-24.
52. Siroux, V., et al., *Identifying adult asthma phenotypes using a clustering approach.* Eur Respir J, 2011. **38**(2): p. 310-17.
53. Woodruff, P.G., et al., *T-helper type 2-driven inflammation defines major subphenotypes of asthma.* Am J Respir Crit Care Med, 2009. **180**(5): p. 388-95.
54. Los, H., P.E. Postmus and D.I. Boomsma, *Asthma genetics and intermediate phenotypes: a review from twin studies.* Twin Research, 2001. **4**(02): p. 81-93.
55. Wjst, M., et al., *A Genome-wide Search for Linkage to Asthma.* Genomics, 1999. **58**(1): p. 1 - 8.
56. Wjst, M., et al., *Asthma families show transmission disequilibrium of gene variants in the vitamin D metabolism and signalling pathway.* Respir Res, 2006. **7**(1): p. 1-11.
57. Altmüller, J., et al., *Phenotypic and genetic heterogeneity in a genome-wide linkage study of asthma families.* BMC Pulm Med, 2005. **5**(1): p. 1-10.
58. Pukelsheim, K.A.S., et al., *Cytokine Profiles in Asthma Families Depend on Age and Phenotype.* PLoS One, 2010. **5**(12): e14299.
59. Miller, S.A., D.D. Dykes and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells.* Nucleic Acids Res, 1988. **16**(3): p. 1215.
60. R Core Team., *R: a language and environment for statistical computing*. *Vienna (Austria): R Foundation for statistical computing*. Team 2011, 2011. **3**.
61. Abecasis, G.R. et al., *Merlin—rapid analysis of dense genetic maps using sparse gene flow trees.* Nature Genetics, 2001. **30**(1): p.97-101.
62. Purcell, S. et al., *PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.* The American Journal of Human Genetics, 2007. **81**(3): p.559-75.
63. Wang, K., M. Li and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): e164.
64. GTEx Consortium., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-85.

65. Leslie, R., C.J. ODonnell and A.D. Johnson, *GRASP: analysis of genotype--phenotype results from 1390 genome-wide association studies and corresponding open access database.* Bioinformatics, 2014. **30**(12): p. i185-i194.
66. Ward, L.D. and M. Kellis, *HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.* Nucleic Acids Res, 2012. **40**: p. 930-34.
67. ENCODE Project Consortium., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
68. Chadwick, L.H., *The NIH Roadmap Epigenomics Program data resource.* Epigenomics, 2012. **4**(3): p. 317-24.
69. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.* Nucleic Acids Res, 2014. **42**: p. 1001-06.
70. Ordoñez, C.L., et al., *Increased neutrophil numbers and IL-8 levels in airway secretions in acute severe asthma: Clinical and biologic significance.* Am J Respir Crit Care Med, 2000. **161**(4): p. 1185-90.
71. Hull, J., A. Thomson and D. Kwiatkowski, *Association of respiratory syncytial virus bronchiolitis with the interleukin 8 gene region in UK families.* Thorax, 2000. **55**(12): p. 1023-27.
72. Heinzmann, A., et al., *Association study suggests opposite effects of polymorphisms within IL8 on bronchial asthma and respiratory syncytial virus bronchiolitis.* J Allergy Clin Immunol, 2004. **114**(3): p. 671.
73. Anderson, G.G. and J.F. Morrison, *Molecular biology and genetics of allergy and asthma.* Arch Dis Child, 1998. **78**(5): p. 488-96.
74. Ober, C., et al., *Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma.* Hum Mol Genet, 1998. **7**(9): p. 1393-98.
75. Bleecker, E.R., et al., *Genome screen for susceptibility loci for bronchial hyperresponsiveness in a genetically homogeneous Dutch population.* Am J Respir Crit Care Med, 1999. **159**(3): p. 645-45.
76. Brasch-Andersen, C., et al., *Significant linkage to chromosome 12q24. 32--q24. 33 and identification of SFRS8 as a possible asthma susceptibility gene.* Thorax, 2006. **61**(10): p. 874-79.
77. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.* Nat Genet, 1995. **11**(3): p. 241-47.
78. Bouzigon, E., et al., *Clustering patterns of LOD scores for asthma-related phenotypes revealed by a genome-wide screen in 295 French EGEA families.* Hum Mol Genet, 2004. **13**(24): p. 3103-13.
79. Koppelman, G.H., et al., *Genome-wide search for atopy susceptibility genes in Dutch families with asthma.* J Allergy Clin Immunol, 2002. **109**(3): p. 498-06.

80. Bradley, M., et al., *Susceptibility loci for atopic dermatitis on chromosomes 3, 13, 15, 17 and 18 in a Swedish population.* Hum Mol Genet, 2002. **11**(13): p. 1539-48.
81. Hoffjan, S. and C. Ober, *Present status on the genetic studies of asthma.* Curr Opin Immunol, 2002. **14**(6): p. 709-17.
82. Haagerup, A., et al., *Asthma and atopy - a total genome scan for susceptibility genes.* Allergy, 2002. **57**(8): p. 680-86.
83. Feng, M., et al., *Genetic analysis of blood pressure in 8 mouse intercross populations.* Hypertension, 2009. **54**(4): p. 802-09.
84. Rutherford, S., et al., *Chromosome 17 and the inducible nitric oxide synthase gene in human essential hypertension.* Hum Genet, 2001. **109**(4): p. 408-15.
85. Ford, F.M., et al., *Hypertension and asthma: psychological aspects.* Soc Sci Med, 1989. **29**(1): p. 79-84.
86. Salako, B.L. and S.O. Ajayi, *Bronchial asthma: a risk factor for hypertension?.* Afr J Med Med Sci, 2000. **29**(1): p. 47-50.
87. Granger, J.P., *An emerging role for inflammatory cytokines in hypertension.* Am J Physiol Heart Circ Physiol, 2006. **290**(3): p. 923-24.
88. Newnham, D.M., *Asthma medications and their potential adverse effects in the elderly: recommendations for prescribing.* Drug Saf, 2001. **24**(14): p. 1065-80.
89. Weinstock, J.V. and D. Elliott, *The somatostatin immunoregulatory circuit present at sites of chronic inflammation.* Eur J Endocrinol, 2000. **143**: p. S15-S19.
90. Elliott, D.E., et al., *SSTR2A is the dominant somatostatin receptor subtype expressed by inflammatory cells, is widely expressed and directly regulates T cell IFN-gamma release.* Eur J Immunol, 1999. **29**(8): p. 2454-63.
91. Taniyama, Y., et al., *Systemic distribution of somatostatin receptor subtypes in human: an immunohistochemical study.* Endocr J, 2005. **52**(5): p. 605-11.
92. Elekes, K., et al., *Inhibitory effects of synthetic somatostatin receptor subtype 4 agonists on acute and chronic airway inflammation and hyperreactivity in the mouse.* Eur J Pharmacol, 2008. **578**(3): p. 313-22.
93. Borie, R., et al., *Activation of somatostatin receptors attenuates pulmonary fibrosis.* Thorax, 2008. **63**(3): p. 251-58.
94. Shepherd, M.C., et al., *KCa3.1 Ca2+ activated K+ channels regulate human airway smooth muscle proliferation.* Am J Respir Cell Mol Biol, 2007. **37**(5): p. 525-31.
95. Nath, P., et al., *Importance of p38 mitogen-activated protein kinase pathway in allergic airway remodelling and bronchial hyperresponsiveness.* Eur J Pharmacol, 2006. **544**(3): p. 160-67.

96. Chung, K.F., *p38 mitogen-activated protein kinase pathways in asthma and COPD.* CHEST Journal, 2011. **139**(6): p. 1470-79.
97. Bhavsar, P., et al., *Effect of p38 MAPK inhibition on corticosteroid suppression of cytokine release in severe asthma.* Eur Respir J, 2010. **35**(4): p. 750-56.
98. Pelaia, G., et al., *Potential role of potassium channel openers in the treatment of asthma and chronic obstructive pulmonary disease.* Life sciences, 2002. **70**(9): p. 977-90.
99. Janciauskiene, S., *Conformational properties of serine proteinase inhibitors (serpins) confer multiple pathophysiological roles.* Biochimica et Biophysica Acta (BBA), 2001. **1535**(3): p. 221-35.
100. de Boer, J., et al., *Role of nitric oxide and superoxide in allergen-induced airway hyperreactivity after the late asthmatic reaction in guinea-pigs.* Br J Pharmacol, 2001. **133**(8): p. 1235-42.
101. Elenkov, I.J., *Glucocorticoids and the Th1/Th2 balance.* Ann N Y Acad Sci, 2004. **1024**: p. 138-46.
102. Derijk, R.H., et al., *Corticosteroid receptor-gene variants: modulators of the stress-response and implications for mental health.* Eur J Pharmacol, 2008. **585**(3): p. 492-01.
103. Pietras, T., et al., *The Bcl I single nucleotide polymorphism of the human glucocorticoid receptor gene h-GR/NR3C1 promoter in patients with bronchial asthma: pilot study.* Mol Biol Rep, 2011. **38**(6): p. 3953-58.
104. Panek, M., et al., *The role of functional single nucleotide polymorphisms of the human glucocorticoid receptor gene NR3C1 in Polish patients with bronchial asthma.* Mol Biol Rep, 2012. **39**(4): p. 4749-57.
105. Maes, M., et al., *The effects of psychological stress on humans: increased production of pro-inflammatory cytokines and a Th1-like response in stress-induced anxiety.* Cytokine, 1998. **10**(4): p. 313-18.
106. Joachim, R.A., et al., *Stress enhances airway reactivity and airway inflammation in an animal model of allergic bronchial asthma.* Psychosom Med, 2003. **65**(5): p. 811-15.
107. Neves, M.F., et al., *Role of aldosterone in angiotensin II-induced cardiac and aortic inflammation, fibrosis, and hypertrophy.* Can J Physiol Pharmacol, 2005. **83**(11): p. 999-1006.
108. Jeong, Y., et al., *Aldosterone activates endothelial exocytosis.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3782-87.
109. Gohlke, H., et al., *Association of the interleukin-1 receptor antagonist gene with asthma.* Am J Respir Crit Care Med, 2004. **169**(11): p. 1217-23.
110. Pattaro, C., et al., *Association between interleukin-1 receptor antagonist gene and asthma-related traits in a German adult population.* Allergy, 2006. **61**(2): p. 239-44.

111. Mao, X.Q., et al., *Imbalance production between interleukin-1beta (IL-1beta) and IL-1 receptor antagonist (IL-1Ra) in bronchial asthma.* Biochem Biophys Res Commun, 2000. **276**(2): p. 607-12.
112. Selig, W. and J. Tocker, *Effect of interleukin-1 receptor antagonist on antigen-induced pulmonary responses in guinea pigs.* Eur J Pharmacol, 1992. **213**(3): p. 331-36.
113. Nakae, S., et al., *IL-1 is required for allergen-specific Th2 cell activation and the development of airway hypersensitivity response.* International Immunology, 2003. **15**(4): p. 483-90.
114. Allakhverdi, Z., et al., *Cutting edge: The ST2 ligand IL-33 potently activates and drives maturation of human mast cells.* J Immunol, 2007. **179**(4): p. 2051-54.
115. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-95.
116. Amartya Sanyal, B.L.G.J.J.D., *The long-range interaction landscape of gene promoters.* Nature, 2012. **489**(7414): p. 109.
117. Sakamoto, O., et al., *Expression of discoidin domain receptor 1 tyrosine kinase on the human bronchial epithelium.* Eur Respir J, 2001. **17**(5): p. 969-74.
118. Avivi-Green, C., M. Singal and W.F. Vogel, *Discoidin domain receptor 1--deficient mice are resistant to bleomycin-induced lung fibrosis.* Am J Respir Crit Care Med, 2006. **174**(4): p. 420-27.
119. Roberts, M.E., et al., *Discoidin domain receptor 1 regulates bronchial epithelial repair and matrix metalloproteinase production.* Eur Respir J, 2011. **37**(6): p. 1482-93.
120. Raunio, H., A. Rautio and O. Pelkonen, *The CYP2A subfamily: function, expression and genetic polymorphism.* IARC scientific publications, 1998. **148**: p. 197-07.
121. Pianezza, M.L., E.M. Sellers and R.F. Tyndale, *Nicotine metabolism defect reduces smoking.* Nature, 1998. **393**(6687): p. 750-50.
122. Minematsu, N., et al., *Association of CYP2A6 deletion polymorphism with smoking habit and development of pulmonary emphysema.* Thorax, 2003. **58**(7): p. 623-28.
123. Bloom, A.J., et al., *Variants in two adjacent genes, EGLN2 and CYP2A6, influence smoking behavior related to disease risk via different mechanisms.* Hum Mol Genet, 2014. **23**(2): p. 555-61.
124. Godoy, W., et al., *CYP2A6/2A7 and CYP2E1 expression in human oesophageal mucosa: regional and inter-individual variation in expression and relevance to nitrosamine metabolism.* Carcinogenesis, 2002. **23**(4): p. 611-16.

125. Consortium, 1.G.P., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
126. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
127. Li, B., D.J. Liu and S.M. Leal, *Identifying rare variants associated with complex traits via sequencing.* Curr Protoc Hum Genet, 2013. **1**: p.1-26.
128. Fragoso, C.A.V., et al., *The ratio of FEV1 to FVC as a basis for establishing chronic obstructive pulmonary disease.* Am J Respir Crit Care Med, 2010. **181**(5): p. 446.
129. Rahman, I., *Oxidative stress and gene transcription in asthma and chronic obstructive pulmonary disease: antioxidant therapeutic targets.* Curr Drug Targets Inflamm Allergy, 2002. **1**(3): p. 291-15.
130. Jeffery, P.K., *Structural and inflammatory changes in COPD: a comparison with asthma.* Thorax, 1998. **53**(2): p. 129-36.
131. Minshall, E.M., et al., *Eosinophil-associated TGF-β1 mRNA expression and airways fibrosis in bronchial asthma.* Am J Respir Cell Mol Biol, 1997. **17**(3): p. 326-33.
132. Tzortzaki, E.G., et al., *Oxidative DNA damage and somatic mutations: a link to the molecular pathogenesis of chronic inflammatory airway diseases.* Chest, 2012. **141**(5): p. 1243-50.
133. Kochanek, S., D. Renz and W. Doerfler, *DNA methylation in the Alu sequences of diploid and haploid primary human cells.* The EMBO journal, 1993. **12**(3): p. 1141-51.
134. Miki, Y., et al., *Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer.* Cancer Res, 1992. **52**(3): p. 643-45.
135. van den Hurk, J.A., et al., *L1 retrotransposition can occur early in human embryonic development.* Hum Mol Genet, 2007. **16**(13): p. 1587-92.
136. Kano, H. et al., *L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism.* Genes Dev, 2009. **23**(11): p. 1303-12.
137. Belancio, V.P., et al., *Somatic expression of LINE-1 elements in human tissues.* Nucleic Acids Res, 2010. **38**(12): p. 3909-22.
138. Merritt, T.A., et al., *Epigenetic influences in the development of bronchopulmonary dysplasia.* Arch of Peri Med, 2011. **17**(1): p. 17-22.
139. Rabinovich, E.I., et al., *Global methylation patterns in idiopathic pulmonary fibrosis.* PLoS One, 2012. **7**(4): e33770.
140. Lange, N.E., et al., *Alu and LINE-1 methylation and lung function in the normative ageing study.* BMJ open, 2012. **2**(5): p. 1-7.
141. Devasagayam, T.P., et al., *Formation of 8-hydroxy(deoxy)guanosine and generation of strand breaks at guanine residues in DNA by singlet oxygen.* Biochemistry, 1991. **30**(25): p. 6283-89.

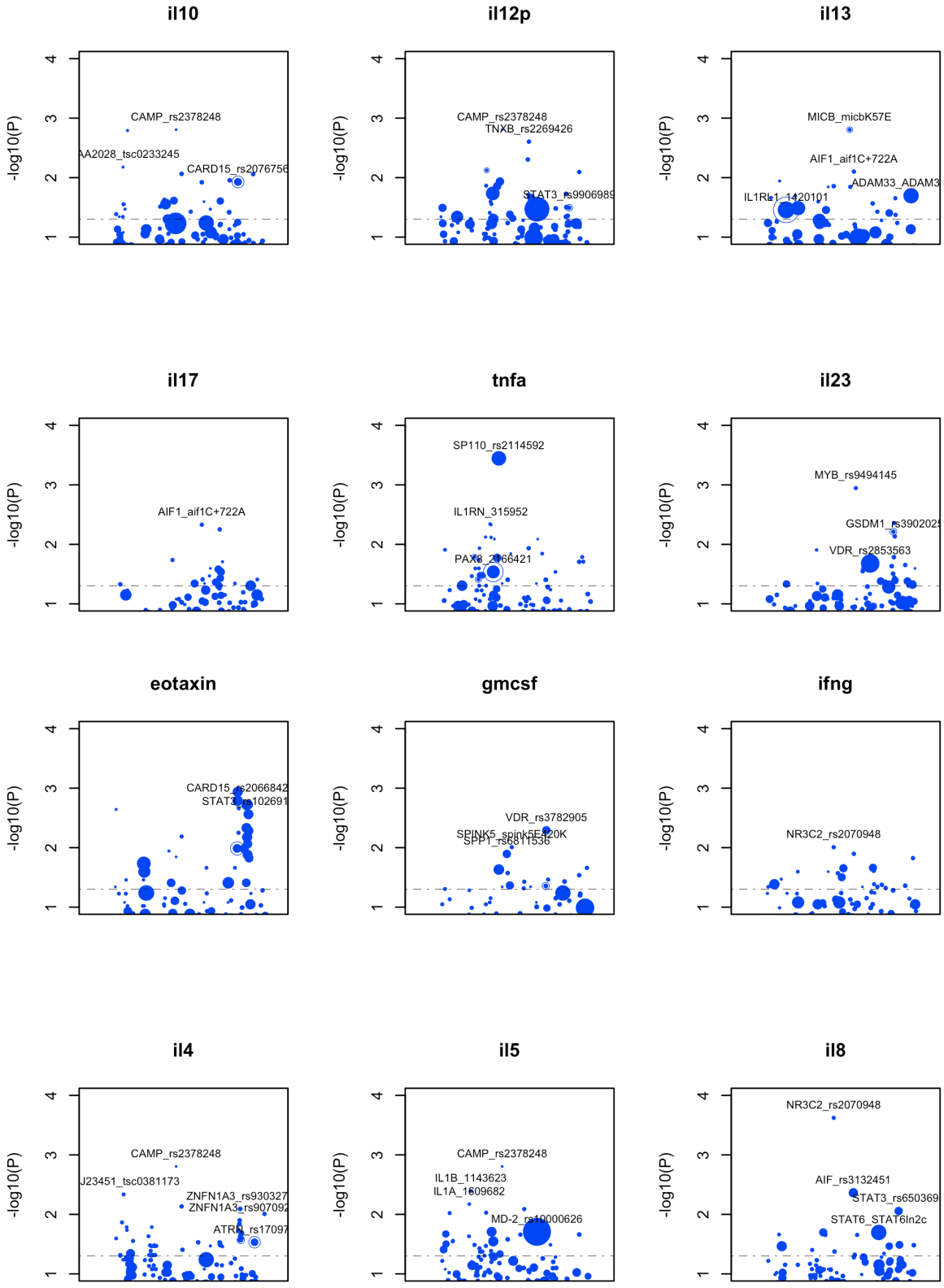
142. Misiaszek, R., et al., *Oxidative DNA damage associated with combination of guanine and superoxide radicals and repair mechanisms via radical trapping.* J Biol Chem, 2004. **279**(31): p. 2106-15.
143. Kuchino, Y., et al., *Misreading of DNA templates containing 8-hydroxydeoxyguanosine at the modified base and at adjacent residues.* Nature, 1987. **327**(6117): p. 77-79.
144. Flom, J.D., et al., *Prenatal smoke exposure and genomic DNA methylation in a multiethnic birth cohort.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(12): p. 2518-23.
145. Baccarelli, A., et al., *Rapid DNA methylation changes after exposure to traffic particles.* Am J Respir Crit Care Med, 2009. **179**(7): p. 572-78.
146. Bergman, C.M. and H. Quesneville, *Discovering and detecting transposable elements in genome sequences.* Brief Bioinform, 2007. **8**(6): p. 382-92.
147. Quesneville, H., D. Nouaud and D. Anxolabéhère, *Detection of new transposable element families in Drosophila melanogaster and Anopheles gambiae genomes.* J Mol Evol, 2003. **57**: p. S50-S59.
148. Hormozdiari, F., et al., *Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes.* Genome Res, 2009. **19**(7): p. 1270-78.
149. Liang, P. and W. Tang, *Database documentation of retrotransposon insertion polymorphisms.* Front Biosci (Elite Ed), 2012. **4**: p. 1542-55.

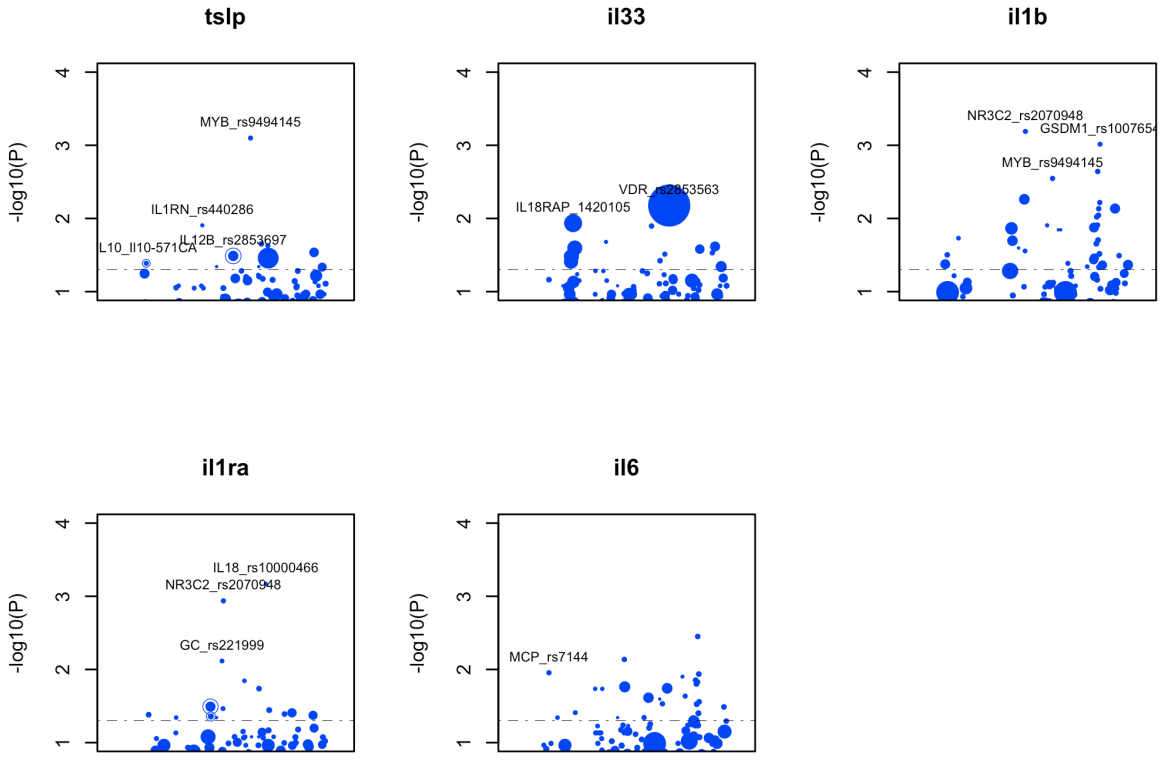
10. Supplementary information

10.1 Multivariate analysis of variance (MANOVA) - overall effect of drugs on serum cytokine levels.

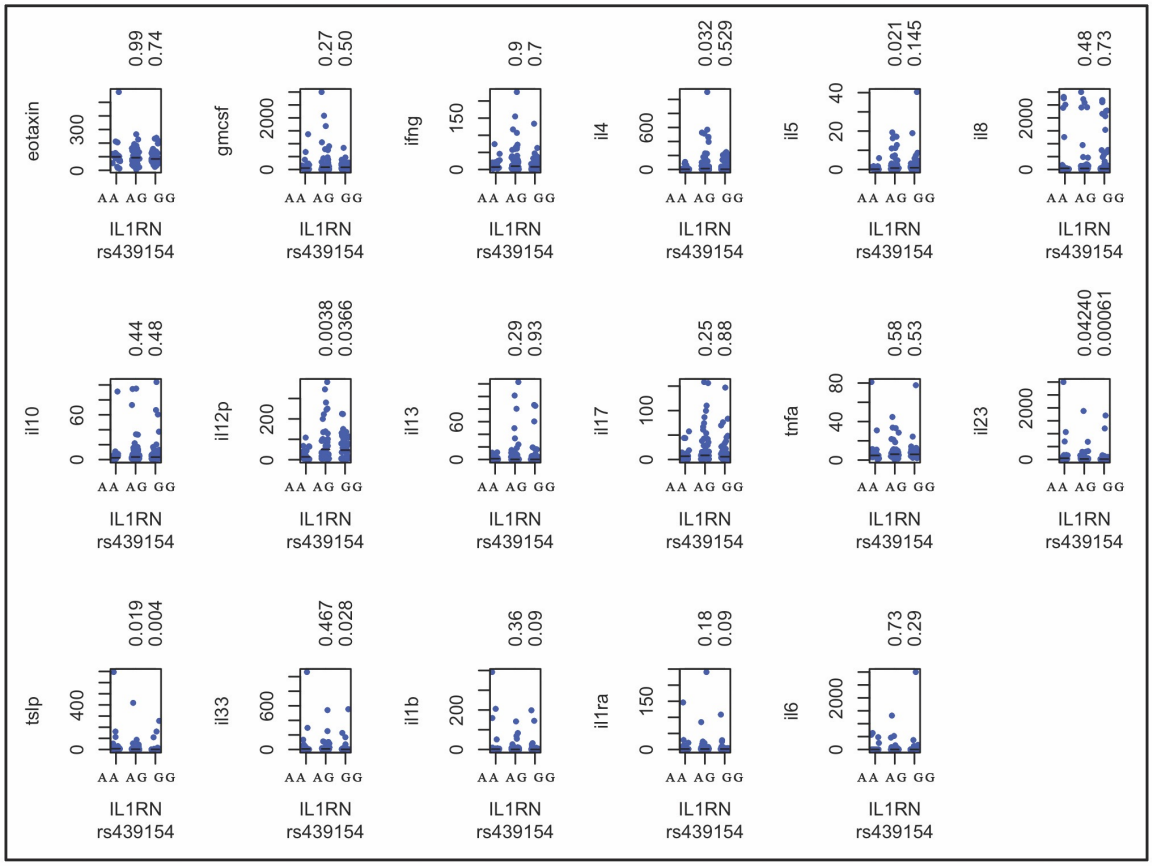
Drug	MANOVA		
	Pillai's trace	approx F	p-value Bonferroni adjusted
Glucocorticoid	0.041	3.48	6.00E-04
Leukotriene antagonist	0.005	0.42	1
β_2-sympathomimetic	0.025	2.08	1.54E-01
Aminophylline	0.012	1.01	1
Cromoglycate	0.027	2.31	7.02E-02
Anticholinergic	0.012	1.01	1
Antihistamine	0.007	0.55	1

10.2 Transmission Disequilibrium Test (TDT) associations of 17 serum cytokines.

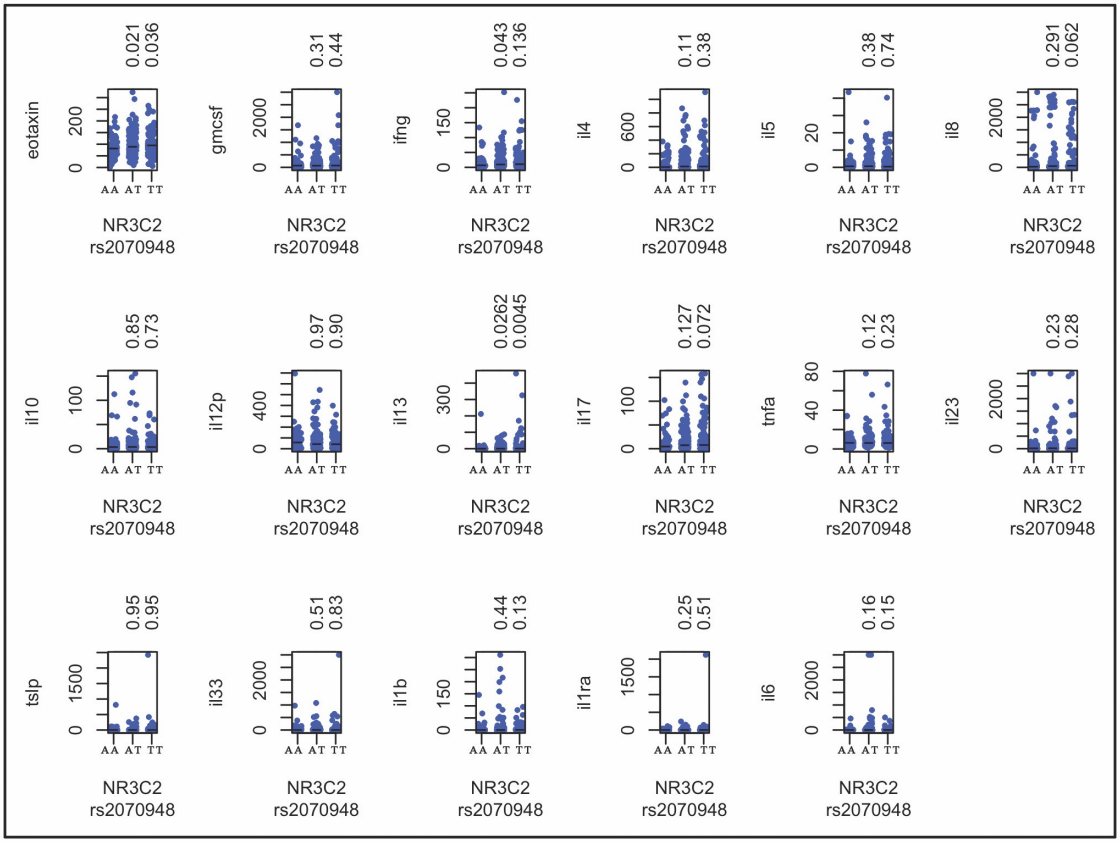



tslp
-log10(P)
MYB_rs9494145
IL1RN_rs440286
IL10_Il10-571CA
IL12B_rs2853697
il33
IL18RAP_1420105
VDR_rs2853563
il1b
NR3C2_rs2070948
GSDM1_rs100765
MYB_rs9494145
il1ra
IL18_rs10000466
NR3C2_rs2070948
GC_rs221999
il6
MCP_rs7144

10.3.a Significant QFAM test association of the *IL1RN* SNPs with serum cytokine levels in children.

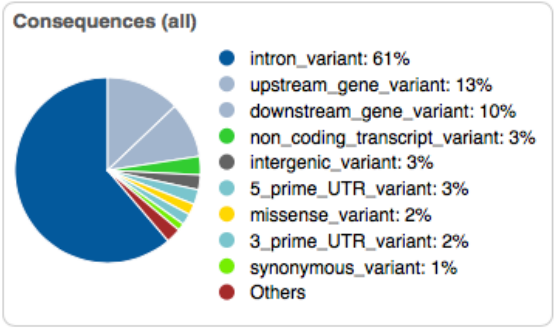


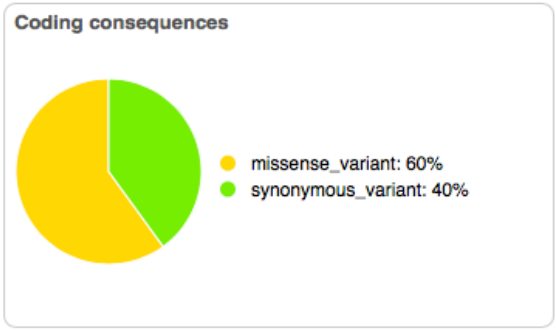
10.3.b Significant QFAM test association of the *NR3C2* SNPs with serum cytokine levels in children.



10.4 Distribution of LD extended SNPs from Variant Effect Predictor (VEP).

Category	Count
Variants processed	149
Variants remaining after filtering	149
Novel / existing variants	2 (1.3%) / 147 (98.7%)
Overlapped genes	13
Overlapped transcripts	15
Overlapped regulatory features	2





10.5 LD SNP information.

Prime_SNP	LD	LD SNP	Gene	BIOTYPE
rs2070948	0.5	rs10029199	-	open_chromatin_region
rs2070948	0.8	rs10031992	*NR3C2*	protein_coding
rs3135499	1.0	rs10521209	*NOD2*	protein_coding
rs3135499	0.8	rs1054987	*RP11-327F22.5*	lincRNA
rs3782905	0.8	rs10875693	*VDR*	protein_coding
rs2070948	0.7	rs11099695	*NR3C2*	protein_coding
rs2070948	0.8	rs11099696	*NR3C2*	protein_coding
rs2070948	0.5	rs112737912	*NR3C2*	protein_coding
rs3135499	0.9	rs11642482	*NOD2*	protein_coding
rs3135499	0.9	rs11642646	*NOD2*	protein_coding
rs3135499	0.6	rs11646242	*CYLD*	protein_coding
rs3135499	0.9	rs11647841	*NOD2*	protein_coding
rs2070948	0.7	rs11929719	*NR3C2*	protein_coding
rs3894194	0.9	rs12451084	*GSDMA*	protein_coding
rs3894194	0.9	rs12451100	*GSDMA*	protein_coding
rs3894194	0.6	rs12601749	*LRRC3C*	protein_coding
rs3894194	0.5	rs12603332	*ORMDL3*	protein_coding
rs3894194	0.9	rs12603481	*LRRC3C*	protein_coding
rs3135499	0.8	rs13332720	*RP11-327F22.5*	lincRNA
rs3135499	0.8	rs13337656	*CYLD*	protein_coding
rs3135499	0.8	rs1420872	*CYLD*	protein_coding
rs2070948	0.8	rs1512344	*NR3C2*	protein_coding
rs3135499	0.5	rs1548990	*CYLD*	protein_coding
rs439154	1.0	rs1665190	*IL1RN*	protein_coding
rs3135499	0.8	rs17223195	*CYLD*	protein_coding
rs3135499	0.9	rs17312836	*NOD2*	protein_coding
rs3135499	0.9	rs17314341	*CYLD*	protein_coding
rs3135499	0.8	rs17314544	*CYLD*	protein_coding
rs2070948	0.7	rs17582206	*NR3C2*	protein_coding

Prime_SNP	LD	LD SNP	Gene	BIOTYPE
rs439154	1.0	rs1794066	*IL1RN*	protein_coding
rs3135499	0.9	rs1861757	*NOD2*	protein_coding
rs3135499	0.9	rs1861758	*NOD2*	protein_coding
rs3135499	0.9	rs1861759	*NOD2*	protein_coding
rs3135499	0.6	rs1861761	*RP11-327F22.5*	lincRNA
rs3135499	0.6	rs1861762	*CYLD*	protein_coding
rs439154	0.6	rs1867830	-	-
rs3135499	0.7	rs2066848	*SNX20*	protein_coding
rs3135499	0.9	rs2066851	*CYLD*	protein_coding
rs3135499	0.9	rs2067085	*NOD2*	protein_coding
rs2070948	1.0	rs2070948	*NR3C2*	protein_coding
rs2070948	1.0	rs2070949	*NR3C2*	protein_coding
rs2070948	0.8	rs2070950	*NR3C2*	protein_coding
rs2070948	0.8	rs2070951	*NR3C2*	protein_coding
rs3135499	0.7	rs2111435	*CYLD*	protein_coding
rs3135499	0.7	rs2160683	*CYLD*	protein_coding
rs3782905	0.9	rs2238138	*VDR*	protein_coding
rs3135499	0.9	rs2357791	*NOD2*	protein_coding
rs3135499	0.5	rs2357792	*NOD2*	protein_coding
rs439154	0.9	rs2592346	*IL1RN*	protein_coding
rs439154	0.9	rs2637988	*IL1RN*	protein_coding
rs2070948	0.5	rs28477241	-	-
rs3135499	1.0	rs3135499	*NOD2*	protein_coding
rs3135499	0.9	rs3135500	*NOD2*	protein_coding
rs3135499	0.6	rs3135501	*CYLD*	protein_coding
rs3135499	0.8	rs3135503	*CYLD*	protein_coding
rs439154	0.9	rs315919	*IL1RN*	protein_coding
rs439154	0.5	rs315933	*IL1RN*	protein_coding
rs3135499	0.9	rs34133110	*NOD2*	protein_coding
rs3135499	0.6	rs34464167	*CYLD*	protein_coding
rs3135499	0.8	rs34552113	*CYLD*	protein_coding
rs3894194	0.9	rs35123741	*LRRC3C*	protein_coding

Prime_SNP	LD	LD SNP	Gene	BIOTYPE
rs3782905	1.0	rs3782905	*VDR*	protein_coding
rs3135499	0.7	rs3785140	*CYLD*	protein_coding
rs3894194	0.6	rs3859191	*GSDMA*	protein_coding
rs3894194	0.6	rs3859192	*GSDMA*	protein_coding
rs3894194	0.7	rs3859193	*GSDMA*	protein_coding
rs3894194	0.6	rs3893044	*LRRC3C*	protein_coding
rs3894194	0.8	rs3894193	*GSDMA*	protein_coding
rs3894194	1.0	rs3894194	*GSDMA*	protein_coding
rs3894194	0.6	rs3902025	*GSDMA*	protein_coding
rs3894194	0.6	rs3907022	*PSMD3*	protein_coding
rs3894194	0.7	rs3931960	*GSDMA*	protein_coding
rs3135499	0.7	rs4027241	*CYLD*	protein_coding
rs3894194	0.5	rs4065275	*ORMDL3*	protein_coding
rs3894194	0.6	rs4065876	*GSDMA*	protein_coding
rs3894194	0.6	rs4065985	*LRRC3C*	protein_coding
rs3894194	0.9	rs4065986	*LRRC3C*	protein_coding
rs3894194	0.6	rs4239225	*GSDMA*	protein_coding
rs439154	1.0	rs439154	*IL1RN*	protein_coding
rs439154	1.0	rs452204	*IL1RN*	protein_coding
rs3894194	0.7	rs4580194	*GSDMA*	protein_coding
rs2070948	0.8	rs4635799	*NR3C2*	protein_coding
rs3135499	0.9	rs4785224	*NOD2*	protein_coding
rs3135499	0.6	rs4785226	*CYLD*	protein_coding
rs3135499	0.6	rs4785227	*RP11-327F22.4*	antisense
rs3135499	0.7	rs4785450	*CYLD*	protein_coding
rs3135499	0.5	rs4785451	*CYLD*	protein_coding
rs3135499	0.6	rs4785452	*RP11-327F22.4*	antisense
rs3135499	0.7	rs4785453	*RP11-327F22.4*	antisense
rs3135499	0.8	rs4785454	*RP11-327F22.5*	lincRNA
rs3135499	0.8	rs4785455	*RP11-327F22.5*	lincRNA
rs3894194	0.5	rs4794820	*RP11-387H17.4*	lincRNA
rs3894194	0.8	rs4794821	*GSDMA*	protein_coding

Prime_SNP	LD	LD SNP	Gene	BIOTYPE
rs3894194	0.9	rs4795406	*LRRC3C*	protein_coding
rs3894194	0.9	rs4795408	-	promoter_flanking_region
rs3894194	0.8	rs4795409	*GSDMA*	protein_coding
rs2070948	0.5	rs4835138	*NR3C2*	protein_coding
rs2070948	0.7	rs4835519	*NR3C2*	protein_coding
rs3894194	0.6	rs55739615	*GSDMA*	protein_coding
rs3894194	0.6	rs56030650	*GSDMA*	protein_coding
rs3894194	0.6	rs56199421	*RP11-387H17.4*	lincRNA
rs3894194	0.6	rs56326707	*GSDMA*	protein_coding
rs2070948	0.5	rs58869535	*NR3C2*	protein_coding
rs3894194	0.6	rs60134943	*GSDMA*	protein_coding
rs3894194	0.6	rs60137005	*GSDMA*	protein_coding
rs3894194	0.8	rs60725845	*GSDMA*	protein_coding
rs3894194	0.6	rs62068170	*LRRC3C*	protein_coding
rs3894194	0.6	rs62068171	*LRRC3C*	protein_coding
rs2070948	0.7	rs62332388	*NR3C2*	protein_coding
rs2070948	0.5	rs62332389	*NR3C2*	protein_coding
rs3135499	0.9	rs6500328	*NOD2*	protein_coding
rs3135499	0.9	rs6500329	*CYLD*	protein_coding
rs3135499	0.6	rs6500331	*CYLD*	protein_coding
rs3894194	0.9	rs6503525	*LRRC3C*	protein_coding
rs3894194	0.9	rs6503526	*GSDMA*	protein_coding
rs2070948	0.8	rs6814934	*NR3C2*	protein_coding
rs3135499	0.9	rs718226	*NOD2*	protein_coding
rs3135499	0.7	rs7194886	*NOD2*	protein_coding
rs3135499	0.9	rs7203691	*NOD2*	protein_coding
rs3135499	0.6	rs7205423	*NOD2*	protein_coding
rs3894194	0.8	rs7212938	*GSDMA*	protein_coding
rs3894194	0.9	rs7216564	*LRRC3C*	protein_coding
rs3894194	0.5	rs7221814	*RP11-387H17.4*	lincRNA
rs3894194	0.7	rs7223717	*GSDMA*	protein_coding

Prime_SNP	LD	LD SNP	Gene	BIOTYPE
rs3782905	0.9	rs7311713	*VDR*	protein_coding
rs3135499	0.6	rs7342715	*CYLD*	protein_coding
rs3135499	0.9	rs748855	*NOD2*	protein_coding
rs3135499	0.8	rs751919	*CYLD*	protein_coding
rs2070948	0.7	rs7658048	*NR3C2*	protein_coding
rs3782905	0.5	rs7968852	*VDR*	protein_coding
rs3782905	0.7	rs7974708	*VDR*	protein_coding
rs3135499	0.9	rs8045009	*NOD2*	protein_coding
rs3135499	0.7	rs8047910	*CYLD*	protein_coding
rs3135499	0.6	rs8053457	*CYLD*	protein_coding
rs3135499	0.5	rs8056611	*NOD2*	protein_coding
rs3135499	0.8	rs8060598	*CYLD*	protein_coding
rs3135499	0.8	rs8061821	*CYLD*	protein_coding
rs3135499	1.0	rs8061960	*NOD2*	protein_coding
rs3135499	0.8	rs8062540	*CYLD*	protein_coding
rs3894194	1.0	rs8069202	*GSDMA*	protein_coding
rs3894194	0.7	rs8071050	-	-
rs3894194	0.8	rs8079416	*RP11-387H17.4*	lincRNA
rs3894194	0.6	rs8080734	*LRRC3C*	protein_coding
rs3894194	0.9	rs8081462	-	-
rs3135499	0.8	rs9635531	*RP11-327F22.4*	antisense
rs3135499	0.8	rs9925070	*CYLD*	protein_coding
rs3135499	0.8	rs9938976	*CYLD*	protein_coding
rs3135499	0.7	rs9940175	*CYLD*	protein_coding
rs2070948	0.8	rs9992256	*NR3C2*	protein_coding

10.6 Significant eQTLs of the GSDMA locus using GRASP.

Lead SNP	LD	LD SNP	eQTL association p-value	Gene	Cell/tissue type	PMID
rs3894194	1.0	rs3894194	1.70E-05	*GSDMA*	blood cells	21829388
rs3894194	1.0	rs3894194	3.30E-93	*GSDMB*	blood cells	21829388
rs3894194	1.0	rs3894194	2.09E-04	*GSDML*	CD4+ lymphocytes	20833654
rs3894194	1.0	rs3894194	7.46E-11	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	1.08E-16	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	1.50E-07	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	1.31E-03	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	1.66E-10	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	6.71E-16	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	4.80E-07	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	1.23E-03	*MED24*	brain	22685416
rs3894194	1.0	rs3894194	1.70E-37	*ORMDL3*	blood cells	21829388
rs3894194	1.0	rs3894194	3.31E-07	*ORMDL3*	CD4+ lymphocytes	20833654
rs3894194	0.9	rs6503526	3.89E-15	*MED24*	brain	22685416
rs3894194	0.9	rs6503526	5.40E-14	*MED24*	brain	22685416
rs3894194	0.9	rs4795408	4.30E-15	*MED24*	brain	22685416
rs3894194	0.9	rs4795408	2.90E-14	*MED24*	brain	22685416
rs3894194	0.9	rs6503525	2.19E-05	*IKZF3*	brain	22685416
rs3894194	0.9	rs6503525	2.52E-12	*IKZF3*	brain	22685416
rs3894194	0.9	rs6503525	7.94E-09	*IKZF3*	brain	22685416
rs3894194	0.9	rs6503525	9.18E-08	*IKZF3*	brain	22685416
rs3894194	0.9	rs6503525	1.90E-12	*IKZF3*	brain	22685416
rs3894194	0.9	rs6503525	1.18E-07	*IKZF3*	brain	22685416
rs3894194	0.9	rs6503525	3.51E-10	*MED24*	brain	22685416
rs3894194	0.9	rs6503525	4.35E-15	*MED24*	brain	22685416
rs3894194	0.9	rs6503525	1.31E-06	*MED24*	brain	22685416
rs3894194	0.9	rs6503525	2.66E-09	*MED24*	brain	22685416
rs3894194	0.9	rs6503525	3.40E-14	*MED24*	brain	22685416
rs3894194	0.9	rs6503525	2.80E-06	*MED24*	brain	22685416

rs3894194	0.9	rs6503525	3.35E-07	*ORMDL3*	blood cells	22692066
rs3894194	0.8	rs8079416	5.80E-05	*GSDMA*	blood cells	21829388
rs3894194	0.8	rs8079416	1.90E-101	*GSDMB*	blood cells	21829388
rs3894194	0.8	rs8079416	5.62E-05	*GSDMB*	intestine	23474282
rs3894194	0.8	rs8079416	5.13E-04	*GSDML*	CD4+ lymphocytes	20833654
rs3894194	0.8	rs8079416	4.96E-10	*MED24*	brain	22685416
rs3894194	0.8	rs8079416	4.35E-15	*MED24*	brain	22685416
rs3894194	0.8	rs8079416	1.56E-06	*MED24*	brain	22685416
rs3894194	0.8	rs8079416	5.20E-03	*MED24*	brain	22685416
rs3894194	0.8	rs8079416	2.71E-09	*MED24*	brain	22685416
rs3894194	0.8	rs8079416	3.39E-14	*MED24*	brain	22685416
rs3894194	0.8	rs8079416	5.72E-03	*MED24*	brain	22685416
rs3894194	0.8	rs8079416	2.50E-39	*ORMDL3*	blood cells	21829388
rs3894194	0.8	rs8079416	7.94E-08	*ORMDL3*	CD4+ lymphocytes	20833654
rs3894194	0.8	rs7212938	1.67E-19	*MED24*	brain	22685416
rs3894194	0.8	rs7212938	8.09E-19	*MED24*	brain	22685416

11. Publications

Sargurupremraj M, Pukelsheim K, Hofer T, Wjst M
“Intermediary quantitative traits - an alternative in the identification of disease genes in asthma?”
Genes and Immunity 2013: 1-7.

Sargurupremraj M, Wjst M
“Transposable elements and their potential role in complex lung disorder.”
Respiratory Research 2013;14: 99.

Wjst M, Sargurupremraj M, Arnold M
“Genome-wide association studies in asthma: what they really told us about pathogenesis”.
Current opinion in allergy and clinical immunology 2013;13: 112-118.

12. Acknowledgement

I take this opportunity to thank Prof. Dr. Matthias Wjst, Prof. Dr. Bertram Mueller-Myhsok and Prof. Dr. Johannes Beckers for their time and valuable suggestions.

It is my greatest gift and pleasure to have such a supportive family (Sargurupremraj, Maheswari, Shobana, Chandru, Sasi Rekha, and Bairav). Lastly, the new addition to my family and the lucky charm, Aadhira. I am indebted for my life. **Vielen Dank.**