THE UNIVERSITY OF CHICAGO

CHARACTERIZING GENETIC RISKS FOR ASTHMA

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF HUMAN GENETICS

BY

SELENE MARILYN CLAY

CHICAGO, ILLINOIS

MARCH 2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Carole Ober. I am incredibly thankful for all her mentorship, support, patience, and encouragement throughout graduate school. I always appreciated that I could rely on her for help and feedback whenever I needed it, and I could not have asked for a better advisor. I would also like to thank my committee members for providing me with invaluable advice during both one-on-one and committee meetings: Haky Im, Matthew Stephens, Dan Nicolae, and Jim Gern. I am grateful to each of them for all their help.

Next, I want to thank all past, present, and unofficial members of the Ober lab. Specifically, I'd like to thank Nathan Schoettler for his guidance on the HLA project and Matt Dapas for his mentorship on the rare variant project. I'd also like to thank Emma Thompson, William Wentworth-Sheilds, Andreanne Morin, Britney Helling, Marcus Soliai, Kevin Magnaye, Charles Washington, Ethan Zhong, Sahar Mozaffari, Qing Gong, Rebecca Anderson, Christine Billstrand, Adriana Prichina, Chis McKennan, and Britney Snyder for all their help and expertise. I'll always have fun memories chatting with everyone while working on the 1000-piece puzzles in the dry lab. Thanks to Sue Levison for helping me with everything during graduate school, and the Department of Human Genetics for providing a wonderful environment.

I want to thank my family for their encouragement. Thanks to all my fellow graduate students and friends: Kate Farris, Katie Aracena, Esha Bandyopadhyay, Jennifer Blanc, Shreya Ramachandran, Jojo Tang, Maggie Steiner, Qi Zhan, Sammy Keyport, Grace Hansen, Meike Lobb-Rabe, and so many more. I also want to give special thanks to my friends who got together once a week to watch trashy reality TV. Lastly, I want to thank my partner Andrew and dog Mickey for their support throughout these years.

ABSTRACT

Asthma is a disease of the airways with significant clinical heterogeneity regarding age-of-onset, co-occurrence with allergic diseases, lung function measures, and more. Genome-wide association studies (GWAS) have successfully reported over 150 asthma susceptibility loci. However, uncovering the causal genes and mechanisms underlying its pathogenesis has been challenging in part due to extensive linkage disequilibrium (LD), which makes it difficult to uncover the specific causal variants and genes, and in part because most asthma GWAS are often conducted in populations of European ancestries with limited information on asthma subtypes and associated traits. In this thesis, I address these gaps through complementary approaches that identify putatively genes in two major subtypes of asthma and in asthma-associated quantitative traits. First, I examined one of the most highly associated asthma loci on chromosome 6p21, encoding the human leukocyte antigen (HLA) genes in childhood-onset asthma and adult-onset asthma. Using Bayesian approaches for fine mapping both GWAS loci and gene expression in three different asthma-relevant cell types, I identified putatively causal childhood- and adult-onset asthma variants that are both shared and distinct to each type of asthma and highlight roles for both gene expression and protein coding variation in the HLA genes for asthma risk. In the second project I examine the contribution of rare variants to specific asthma-associated quantitative phenotypes in a population of children of diverse ancestries. I examined whole genome sequencing data and detailed clinical information reflecting the major allergic, pulmonary, and immune components of asthma. I performed gene-based variant set tests and followed up associations with independent, predicted gene expression and mouse knockout resources. Overall, I report novel associations between genes and allergic and inflammatory

phenotypes. Together, these studies build on the results of asthma GWASs, identifying novel variants and genes associated with asthma and its associated phenotypes.

CHAPTER 1

INTRODUCTION

1.1 Asthma is a complex, heterogeneous disease

Asthma is a chronic, inflammatory disease of the airways. As one of the most common noncommunicable diseases in the world, asthma has been diagnosed in upwards of 339 million people worldwide and represents a significant global health burden (1). While symptoms generally vary, the diagnosis of asthma is typically based on the presence of wheezing, shortness of breath, cough, and chest tightness. These reflect underlying features of bronchial hyperresponsiveness and inflammation, leading to airflow obstruction and difficulty breathing and even respiratory failure in severe cases (1). Symptoms can be induced by varying triggers such as exercise, exposure to allergens, and respiratory tract infections (2).

Asthma is also a heterogeneous disease that varies between patients with respect to different clinical features. This includes variation in response to treatment, co-occurrence with other allergic diseases, and the presence of eosinophilia, as examples (3,4). In general, these features reflect the allergic, inflammatory, and pulmonary components of asthma that vary across individuals and can define different endotypes of this disease (5). Based on these observations, it has been suggested that “asthma” is in fact multiple, overlapping diseases with perturbations in distinct molecular pathways (5). For example, a subtype based on age is well recognized; childhood-onset asthma and adult-onset asthma have distinct characteristics, such as different prevalence in males and females at different ages, comorbidities, and lung function impairment, among others, suggesting that differences in genetic risk may contribute to this heterogeneity.

There is significant phenotype heterogeneity within childhood-onset asthma as well (6). Two studies performed by the Inner City Asthma Consortium (ICAC), Asthma Phenotypes in the Inner City (APIC) (4) and the URban Environment and Childhood Asthma (URECA) (7), examined this question in children of diverse ancestries living in US urban environments. A study by Zoratti *et al.* applied hierarchical clustering analyses in the APIC cohort and identified five asthma clusters defined by 33 baseline and asthma severity variables, most of which were grouped by degree of allergy and marks of allergic inflammatory, pulmonary physiology, and rhinitis (6). Another study from Bacharier *et al*. similarly performed clustering analyses in the URECA cohort (8), finding that patterns of atopy, wheezing, and lung function defined five phenotype clusters in high-risk children with and without asthma. Similar studies in adults have revealed specific-subtypes of asthma defined by measures of lung function, medication use, allergic sensitization and obesity (9).

Despite this heterogeneity, heritability estimates for asthma estimated from twin studies range from 35-70% (10–13), reflecting significant genetic contribution in disease etiology. Numerous genome-wide association studies (GWAS), which test associations between common variants genome-wide with a trait, have been performed for asthma, identifying over 150 independent loci (14–19). These include highly replicated associations at several loci containing the *ORMDL3*/*GSDMB* genes at 17q12-q21, the *IL1RL1*/*IL18R1* genes on 2q12.1, *IL33* on 17q21.12 locus, and the human leukocyte antigen (HLA) genes on 6p21.32, among others.

Finally, GWAS for different subtypes have been performed. A 2019 study from Pividori *et al*. examined childhood- and adult-onset asthma in individuals from the UK Biobank, finding i) adult-onset asthma risk loci are largely a subset of those associated with childhood-onset asthma but with smaller effect sizes, ii) SNP-based heritability is 0.327 for childhood-onset and

0.098 for adult-onset asthma, and iii) childhood-onset asthma loci were enriched for genes with highest expression in skin, blood cells, and small intestine, whereas adult-onset asthma loci were enriched for genes in lung, blood cells, spleen, and small intestine (15). GWAS have also analyzed other subtypes of asthma, such as severe asthma (20), and for other associated phenotypes of asthma including total serum IgE (21), allergic sensitization (22), pulmonary function (23), and food allergy (24), among others (25,26), providing variable degrees of insight into the development of asthma and the shared genetic architecture between these traits.

However, uncovering the causal variants and genes at these loci has been difficult. Extensive linkage disequilibrium (LD) between the causal variant(s) and nearby SNPs makes pinpointing the causal variants particularly challenging, especially in populations of European ancestries. Since most association signals reside in non-coding sequences, uncovering which genes their effects are mediated through in the relevant cell type is not always clear (27). With the exception of a few loci (28,29), the putatively causal variants and genes are largely unknown in asthma. Approaches like fine-mapping or studying rare variants in genes may be able to identify candidate genes and mechanisms more directly.

1.2 Variation in the HLA region is robustly associated with asthma

One region where variation is robustly associated with asthma risk is the HLA region (14–16,30–33), which contains genes involved in the immune response. Located across a 4Mb stretch on chromosome 6p21, the HLA locus is divided into three regions containing a dense cluster of roughly 260 genes (**Figure 1.1A**) (34). The class I region contains the classical *HLA-A*, *-B*, and *-C* genes, which are present on the surface of nearly all nucleated cells and code for proteins that bind T cell receptors (TCRs) on CD8+ T cells (**Figure 1.1B**) (35,36). The class II region includes genes encoding the HLA-DQ, -DR, and -DP molecules, which are present on the

surface on a subset of immune cells (like B cells, macrophages, and dendritic cells), and present peptides to the TCRs on CD4+ T cells (**Figure 1.1B**) (35–37). The class III region does not contain HLA genes per se, but rather other genes, many of which are involved in the immune function (38).

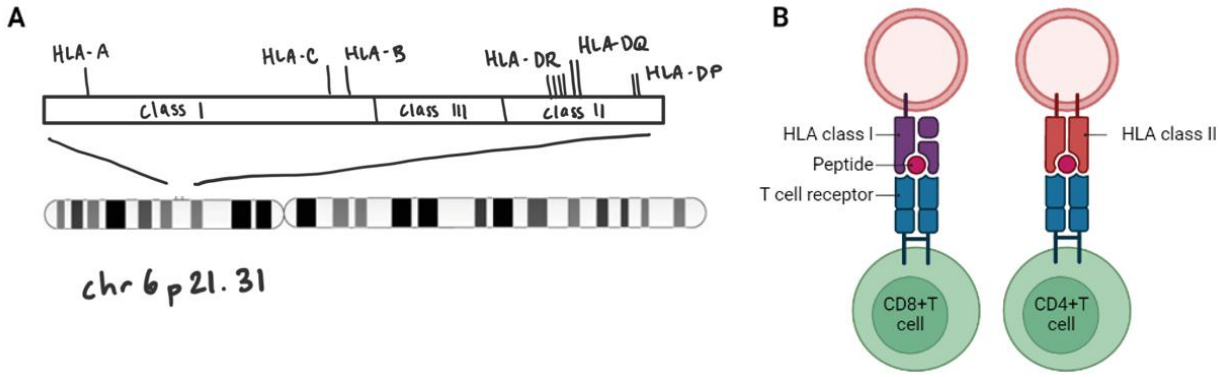


Figure 1.1 HLA Region Overview. **A)** Figure shows an overview of the HLA locus on chromosome 6. Modified from Xie *et al.* (2020) (39). **B)** This illustration shows the interaction between the HLA class I and class II molecules, peptide, and TCR. Adapted from "*Extracellular Vesicles (EV) for Direct Antigen Presentation*" by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

The classical HLA genes are among the most polymorphic genes in the genome with over 15,000 alleles identified to date (40). Most of amino acid substitutions reside in the peptide-binding groove of the HLA molecule or in the regions that interact with the TCRs (36). This specific recognition of HLA, peptide, and TCR is principal in driving the immune system to recognize self from non-self and initiate the adaptive immune response to infection. Both the class I and class II regions include "nonclassical" HLA molecules that have little polymorphism and different (often poorly characterized) functions.

The classical HLA genes also have a complex nomenclature due to the large number of alleles and initial approaches to HLA typing by serology and, more recently, by DNA sequencing (40). The first two digits of each allele are the serotype (e.g. *HLA-DRB1**03) (now

defined by molecular typing). The next two digits refer to the subtypes that differ in amino acid sequence (*HLA-DRB1**03:01), which is the four-digit type or allele. Additional digits refer to synonymous variation in the gene and are not thought to impact antigen binding or function.

The HLA region is associated with more diseases than any other region of the genome, with variation at this extended locus primarily influencing risks for or protection from autoimmune and infectious diseases and transplantation rejection (34,37,41). The HLA genes have also been implicated in the development of allergic diseases and the ability to respond to allergens, and genetic studies have revealed associations at the HLA loci with traits such as allergic rhinitis (hay fever) (25), atopic dermatitis (eczema) (42), total serum IgE (43), and allergic sensitization (22).

The HLA region has been among the most significant associations in asthma GWAS (14–16,30–33). In the Pividori *et al.* GWAS for childhood- and adult-onset asthma in individuals from the UK Biobank, multiple, significant signals were identified in the HLA region for both. In fact, this locus was the most significant association for adult-onset asthma. Independent associations were detected in the class I (*HLA-C*/*B*) and in the class II (*HLA-DR*/*DQ*) regions for both subtypes. The authors used PrediXcan (44) to impute expression of genes at GWAS loci; in the HLA region, expression of over 40 genes was predicted to be associated with childhood- and/or adult-onset asthma in blood, lung, and skin.

Despite its importance in asthma, determining the causal variants and genes in this region has been challenging. In addition to its remarkably high gene density and high levels of polymorphism, there is strong LD across the region making it particularly difficult to determine which disease-associated variants are causal and which genes underlie these associations.

Additionally, GWAS only focus on SNP associations which do not capture the complexity of the HLA region with respect to HLA allelic and amino acid variation.

Finally, standard RNAseq mapping pipelines that map reads to a single reference genome can provide biased expression estimates when it comes to the polymorphic HLA genes due to the potentially high number of sequence differences between the individual's HLA type and the reference (45). However, the recent availability of large samples for GWAS has provided increased statistical power to uncover multiple independent associations at the HLA locus, and imputation of HLA types to four-digit resolution using GWAS SNPs provides a practical approach for genotyping thousands of individuals (46–48).

1.3 Rare variant studies can identify effector genes

Rare variants comprise the bulk of human genetic variation (49) and, unlike common variants, tend to have lower LD with nearby variants, may have stronger effects on traits, and can be more population specific. Overall, rare variants can more directly point to causal genes and mechanisms (50). However, the contribution of rare variants towards complex traits like asthma and allergy have been overall less studied in comparison to common variants through GWAS for several reasons. First, rare variants are not included on commonly used genotyping arrays and cannot be reliably imputed. As a result, to study rare variants, other methods are required such as direct sequencing to identify variation in specific genes, either using a customized SNP array, whole-exome sequencing (WES), or whole-genome sequencing (WGS). Additionally, there is low power to detect associations between complex traits and individual rare variants, requiring large sample sizes for sufficient power. This is especially challenging since sample sizes tend to be lower in studies where individuals have been subjected to more costly sequencing.

Statistical methods have been developed that can increase power in rare variant studies. By grouping variants together according to correlated functional impact (e.g. by gene, window, biological pathway), we can gain power to test for associations by jointly testing associations of a set of variants instead of testing each individual variant. Common methods include the burden tests, which collapse information from gene variants into a single score that is then tested for association with the trait. These tests are more powerful when variants have the same direction of effect (51–54). There are also variance component tests like the sequence kernel association test (SKAT), which test for association by evaluating the distribution of genetic effects for a group of variants and are more powerful in the presence of both trait-increasing and -decreasing variants (55). Omnibus methods can take into account different scenarios that may happen across the genome (56,57). Additionally, applying a variant set test reduces the multiple testing burden from millions of rare variants to however many sets (e.g. 25,000 genes) are examined.

Several studies have explored the contribution of rare variants to asthma and allergic disease (58,59,68,60–67). One notable example is from Smith *et al.* who examined rare loss of function (LOF) variants in the *IL33* and *IL1RL1* genes (59), which have previously been linked to asthma through GWAS (31,69,70). *IL1RL1* encodes ST2, the receptor for IL-33. They discovered a rare, LOF mutation in *IL33* (rs146597587-C) which disrupts a canonical splice acceptor site that was associated with lower eosinophil count, lower asthma risk, and lower *IL33* gene expression in adipose tissue. This finding led to the discovery of a new therapeutic (astegolimab), which inhibits binding of *IL33* to its receptor ST2 (71).

While there are rare variant studies on various inflammatory, allergic, and pulmonary traits in cohorts of diverse ancestries (59,60,72–78), most of these studies did not use WGS and may have missed key variants and genes. Studying the contribution of rare variation genome-

wide to these traits may identify core genes and pathways, which may be also able to inform new potential therapeutic targets.

1.4 Dissertation overview

In this study, I identified putatively causal genes for two subtypes of asthma as well as asthma-associated quantitative traits. In Chapter 2, I examine one of the most highly associated asthma loci (the HLA region) in childhood- and adult-onset asthma in individuals from the UK Biobank and test the hypothesis that causal variants in the HLA region are both shared and distinct between childhood- and adult-onset asthma and that the variants exert their effects through either altering protein properties or gene expression of the HLA genes. I identify putatively causal variants using Bayesian fine-mapping approaches and through analyzing gene expression in asthma-relevant cell types. In Chapter 3, I test the hypothesis that rare variants in genes are associated with asthma-associated quantitative phenotypes. Using WGS from two cohorts of children of diverse ancestries, I performed gene-based variant set analyses and examined associations with phenotypes reflecting the allergic, pulmonary, and immune components of asthma, and validated key associations using independent predicted gene expression and mouse knockout resources. These studies build on and complement the results from asthma GWAS, identify novel variants and genes associated with asthma and its associated phenotypes, and ultimately improve our understanding of the genetic architecture of this complex disease while providing potential therapeutic targets.

CHAPTER 2

Fine-Mapping Studies Distinguish Genetic Risks for Childhood- and Adult-Onset Asthma in the HLA Region

2.1 ABSTRACT[1]

Background: Genome-wide association studies of asthma have revealed robust associations with variation across the human leukocyte antigen (HLA) complex with independent associations in the HLA class I and class II regions for both childhood-onset asthma (COA) and adult-onset asthma (AOA). However, the specific variants and genes contributing to risk are unknown.

Methods: We used Bayesian approaches to perform genetic fine mapping for COA and AOA (n=9,432 and 21,556, respectively; n=318,167 shared controls) in White British individuals from the UK Biobank and to perform expression quantitative trait locus (eQTL) fine mapping in immune (lymphoblastoid cell lines, n=398; peripheral blood mononuclear cells, n=132) and airway cells (nasal epithelial cells, n=188) from ethnically diverse individuals. We also examined putatively causal protein coding variation from protein crystal structures and conducted replication studies in independent multi-ethnic cohorts from the UK Biobank (COA n=1,686; AOA n=3,666; controls n=56,063).

Results: Genetic fine mapping revealed both shared and distinct causal variation between COA and AOA in the class I region but only distinct causal variation in the class II region. Both gene expression levels and amino acid variation contributed to risk. Our results from eQTL fine mapping and amino acid visualization suggested that the *HLA-DQA1**03:01 allele and variation associated with expression of the nonclassical *HLA-DQA2* and *HLA-DQB2* genes accounted entirely for the most significant association with AOA in GWAS. Our studies also suggested a

[1] Citation for chapter: Clay SM *et al*. Fine-mapping studies distinguish genetic risks for childhood- and adult-onset asthma in the HLA region. *Genome Medicine*. 2022.

potentially prominent role for HLA-C protein coding variation in the class I region in COA. We replicated putatively causal variant associations in a multi-ethnic cohort.

Conclusions: We highlight roles for both gene expression and protein coding variation in asthma risk and identified putatively causal variation and genes in the HLA region. A convergence of genomic, transcriptional, and protein coding evidence implicates the *HLA-DQA2* and *HLA-DQB2* genes and *HLA-DQA1**03:01 allele in AOA.

2.2 INTRODUCTION

Asthma is a chronic, inflammatory disease of the airways, affecting over 330 million people worldwide and representing a significant global health burden (79). Genome-wide association studies (GWASs) have reported over 150 independent loci associated with asthma (14–18), including highly replicated, significant associations at the human leukocyte antigen (HLA) region on chromosome 6p21. Recently, we performed a GWAS for childhood-onset asthma (COA) and adult-onset asthma (AOA) (15) in individuals from the UK Biobank (80). Each revealed independent, broad regions of association spanning HLA class I (*HLA-C*/*B*) and class II (*HLA-DR*/*DQ*) genes. While the HLA class II region was the most significant locus for AOA, variants in this region had greater effect sizes for COA compared to AOA (15). In contrast, associations with variants in the class I region were similar and had similar effect sizes in both.

Overall, the HLA region is the most frequently associated locus with asthma and allergic diseases (81). Whereas its central role in adaptive immunity has been extensively characterized (82–84), determining the causal variants and their putative functions has been particularly challenging due to the remarkably high gene density, extraordinary levels of genetic polymorphism, and striking linkage disequilibrium (LD) that characterize this region (34,36). These features make identification of causal disease-associated variants and the genes that

underlie these associations particularly challenging. As a result, the specific HLA region variants and genes contributing to asthma risk are not known.

In this study, we hypothesized that the causal variants at the HLA locus include those that are both shared and distinct to COA and AOA, and that some causal variants exert their effects on asthma risk by modifying the expression of HLA genes while others alter protein properties by changing amino acid sequences in functional domains. Using Bayesian approaches for fine mapping GWAS loci (genetic fine mapping) and for fine mapping expression quantitative trait loci (eQTLs; expression fine mapping) of HLA region genes in cell types relevant to asthma, we identified putatively causal COA and AOA variants and genes in the HLA class I and class II regions and replicated a subset of causal variants in an ethnically-diverse sample.

2.3 METHODS

2.3.1 Study Subjects in the Discovery and Replication Samples

We examined COA and AOA HLA loci in a discovery cohort comprised of the same adult individuals from the UK Biobank data release July 19, 2017 and using the same inclusion/exclusion criteria and phenotype definitions reported in Pividori et al.(15). Briefly, we filtered out individuals with poor-quality genotypes, ambiguous sex assignments, and related individuals (15). Subjects included in the discovery analysis were restricted to White British ancestry (UK Biobank Data-Field 22006). COA cases were subjects with self-reported doctor-diagnosed asthma before 12 years of age (n=9,432), AOA cases were subjects with self-reported doctor-diagnosed asthma between 26-65 years of age (n=21,556), and controls were subjects with no reported asthma up to the latest age of study participation (n=318,167). Individuals with chronic obstructive pulmonary disease, chronic bronchitis and emphysema were excluded from the AOA and control groups.

The replication cohort consisted of individuals from the UK Biobank who were not in the White British ancestry set. We used the same definitions as above for COA (n=1,686), AOA (n=3,666), and non-asthmatic controls in the replication cohort (n=56,063). The self-reported ancestry for the replication cohort is 70.7% White, 12.4% Black or Black British, and 16.8% Asian or Asian British **(Figure S2.1**).

2.3.2 Genotypes and HLA Alleles

Allele dosages were extracted for genotyped and imputed SNPs from UK Biobank v3 within the boundaries of the asthma HLA loci in the COA and AOA GWAS as defined by Pividori et al.(15) using the rbgen 0.1 package in R 3.6.1(85). All SNPs that passed the following genotype quality control filters: call rate > 95% or information score > 80%, Hardy-Weinberg equilibrium test p-value > 1×10^{-10}, and minor allele frequency (MAF) > 0.1% were included, as previously described (15). A total of 8,624 and 10,006 SNPs passed the genotype filters at the HLA class I and class II locus, respectively. Four-digit resolution (also known as two-field resolution) of classical HLA allele dosages were imputed from SNP data by the UK Biobank using HLA*IMP:02 (80,86). We excluded alleles with low-frequencies (<1%, specific to each self-reported race). We translated the imputed HLA allele dosages to their corresponding amino acid dosages using publicly available data from SNP2HLA (46) (http://software.broadinstitute.org/mpg/snp2hla/) that map HLA alleles to their amino acid sequences. All amino acid polymorphisms were encoded as biallelic, as previously described (25,46,87). Amino acid polymorphisms with low-frequencies (<1%) were excluded. Base-pair positions for variants, genes and other genomic features are based on Human Genome Assembly hg19 (88,89). We use the term “HLA allele” when referring to the four-digit nomenclature, “amino acid polymorphism” when the amino acid is the target of analysis, and “SNP” when the

SNP is the target of analysis. We note that SNPs determine amino acid polymorphisms and amino acid polymorphisms determine HLA alleles.

2.3.3 Variant Associations

Associations between variants (SNPs, HLA alleles, and amino acid polymorphisms) and COA and AOA were tested using logistic regression. Sex and the first 10 ancestral principal components (PCs) were included as covariates in the analyses in White British subjects. We used the test for heterogeneity from METAL (90) to determine if variant effects were shared between COA and AOA. We focused on significantly associated variants and amino acid polymorphisms and applied a Bonferroni correction to the results from the heterogeneity test.

To determine whether the associations with variants of interest were driven in part by allergy, we extracted allergy status (any eczema, hay fever, and/or food allergy) and performed regressions where we 1) included allergy status as a covariate in the regression, 2) used an interaction term between allergy status and the variant of interest, and 3) examined associations excluding all individuals with allergy. To determine whether there were any sex differences in HLA-associated risks, we 1) performed associations for the candidate variants separately in males and females, and 2) including an interaction term between sex and the variant in the regression model.

2.3.4 Fine-Mapping the HLA Region

Sum of Single Effects (SuSiE) (91) (susieR R package version 0.9.0) was used to fine map the asthma-associated HLA loci and determine putatively causal variants for COA and AOA, separately in the class I and class II regions. SuSiE is a Bayesian analog of stepwise conditional analysis that improves on previous methods by taking account of the uncertainty of the selection of associated SNPs. Additionally, SuSiE can handle individual-level data (allowing us to

examine SNPs, HLA alleles, and amino acid polymorphisms), detect multiple causal signals, and identify cases when the variable with the lowest p-value is not the causal variable (91). SuSiE reports credible sets, which are sets of variants that include at least one causal variant with high probability. To assess whether SNPs, amino acid polymorphisms, and/or HLA alleles were causal for asthma risk, genotype dosages for these three types of polymorphisms were included in a single genotype matrix: class I SNPs, HLA alleles, and HLA amino acid polymorphisms were examined together for the fine-mapping studies in the class I region; class II SNPs, HLA alleles, and HLA amino acid polymorphisms were examined together for the class II region. See section 2.6.1.1 for more details.

2.3.5 Gene Expression and eQTL Studies

To test whether SNPs identified in the fine-mapping results using SuSiE are eQTLs in asthma-relevant cell types, RNA-seq data from three cell sources from ethnically diverse individuals was evaluated (**Table S2.1**). Lymphoblastoid cell lines (LCLs), peripheral blood mononuclear cells (PBMCs), and upper airway (nasal) epithelial cells (NECs) were considered as surrogates for immune, lung and epithelial tissues that were implicated as relevant cells/tissues in asthma risk in the COA and AOA GWAS (15). The LCLs were derived from blood previously collected from 398 Hutterites (92), a founder population of European descent. Unstimulated PBMCs were derived from blood (n=132) (93) and NECs from nasal swabs (n=188) (94), which were previously collected from children from the URban Environment and Childhood Asthma (URECA) birth cohort (7). In both studies, written informed consent was obtained from the parents of children, and assent from children age 6 or older. Both studies were approved by institutional review boards, and all study subjects were assigned randomly generated ID codes.

RNA-sequencing reads from these studies were remapped. For the polymorphic HLA genes, we aligned RNA-seq reads to reference sequences from the International ImMunoGeneTics (IMGT) database (40) for each individual's known HLA type (45) (see Supplementary Methods for more details). For eQTL mapping in the PBMCs and NECs, we performed linear regressions with QTLtools (95) using a nominal pass and *cis*-window size of ±1 Mb from the transcription start site (TSS). We performed eQTL mapping in the LCLs using Genome-wide Efficient Mixed Model Association (GEMMA) (96), including a kinship matrix as a random effect to account for relatedness between the Hutterites. Relevant covariates were included for all tests (see section 2.6.1.3 for additional details). In the LCLs, some SNPs in the AOA or COA credible sets had high missingness. In these instances, we ran the eQTL mapping in the subset of individuals with high quality genotypes. In the PBMCs and NECs, some SNPs in the AOA or COA credible sets failed QC in genotypes extracted from whole genome sequence data. In those instances, we extracted high quality genotypes from the Illumina MEGA array in the subset of individuals that had previous array-based genotyping and used those genotypes in the eQTL studies.

2.3.6 Fine Mapping eQTLs

We performed eQTL fine mapping in three asthma-relevant cell types, LCLs, PBMCs, and NECs, using SuSiE for the expression of genes in which SNPs in any of the COA or AOA credible sets were significantly associated with their expression at an False Discovery Rate (FDR) (97) threshold of 0.05. The same covariates in each of the eQTL studies were regressed from the dataset (see section 2.6.1.3 for more details).

SuSiE was used to fine map eQTLs using a window of ± 1 Mb around the TSS of each gene. SNPs in any of the eQTL credible sets were then compared to the SNPs in the COA and AOA GWAS credible sets to assess overlap.

2.3.7 Structural Visualization of Amino Acid Variants

Based on results of fine mapping, visualization of amino acid polymorphisms was performed for each associated coding variation (amino acid polymorphism). Amino acid polymorphisms were aligned to their positions on the protein. Crystal X-ray structures were downloaded from the Protein Data Bank (PDB) (98) for HLA molecules containing the risk/protective amino acid variants of interest, if available: 5VGE (HLA-C*07:02) (99), 6DIG (HLA-DQA1*01:02/HLA-DQB1*06:02) (100), and 4D8P (HLA-DQA1*03:01/HLA-DQB1*02:01) (101). The locations of amino acids within the HLA molecules were visualized with PyMOL v2.0.7 (102) (https://pymol.org/2/).

2.3.8 Replication of Fine-Mapping Results

Because variants within a credible set are highly correlated, we selected candidate variants (tag SNPs) from each credible set for replication. These included rs2428494 (the shared class I COA and AOA credible set 1 [CS1] SNP), HLA-C p.11 (class I COA credible set 2 [CS2] amino acid polymorphism), rs28407950 (class II COA CS1 lead SNP), rs35571244 (lead class II COA CS2 SNP), rs9272346, rs1063355, rs3828789, and rs9274660 (class II AOA CS1 SNPs that were also in the eQTL CSs and overlapped with LCL enhancer marks), and HLA-DQA1*03:01 (class II AOA CS2 HLA allele). We performed logistic regressions for COA and AOA separately in self-reported White (excluded from White British Ancestry cohort), Asian British, and Black British individuals from the UK Biobank and then performed meta-analysis as implemented in METAL

(90). Logistic regressions were performed using sex and the top 20 ancestry PCs as covariates in a one-sided test that required the direction of effect to be the same as in the discovery sample.

2.4 RESULTS

2.4.1 HLA Allele and Amino Acid Associations

Similar to our previous study of White British individuals (15), we identified 9,432 COA cases, 21,556 AOA cases, and 318,167 shared non-asthma controls (**Table 2.1**). Because our previous GWASs did not include HLA alleles or amino acid variants (15), we first considered the imputed 4-digit HLA alleles (103) provided by the UK Biobank and identified a total of 78 alleles for the *HLA-A* (n=13), *HLA-B* (n=18), *HLA-C* (n=14), *HLA-DRB1* (n=15), *HLA-DQB1* (n=12), and *HLA-DQA1* (n=7) genes that met the frequency threshold of 0.01. Of the 78 HLA alleles at six loci, 19 were associated with COA and 14 with AOA at a genome-wide significance threshold ($p<5x10^{-8}$; **Table S2.2**). Overall, the effect sizes of associated alleles were larger for COA compared to AOA (**Figure S2.2**). Our results were concordant using either an additive or dominant model (**Table S2.3**). The only alleles that showed significant heterogeneity were *HLA-DQA1**01:02, *HLA-DQA1**03:01, *HLA-DQB1**03:02, *HLA-DQB1**06:02, *HLA-DRB1**15:01 (**Table S2.4**).

Table 2.1. Sample Composition.

	Childhood-onset asthma	Adult-onset asthma	Controls
Sample size	9,432	21,554	318,167
Mean age of asthma onset in years (SD)	6 (3)	44 (10)	NA
Any asthma medication use (%)	41.6	44.8	0.5
Female Sex (%)	40.7	63.6	53.5
Allergic Disease (ever, %)	34.0	24.6	10.6
Allergic rhinitis (%)	26.9	21.6	8.6
Atopic dermatitis (%)	11.9	4.3	2.3
Food allergy (%)	1.5	0.8	0.4

Table 2.1 Sample Composition. (continued)

Mean FEV_1 percent predicted (SD)	87.62 (18.20)	91.80 (17.05)	98.41 (15.62)
Mean FEV_1/FVC (SD)	0.71 (0.08)	0.74 (0.07)	0.76 (0.06)
Mean Eosinophil count (SD)	0.23 (0.18)	0.22 (0.17)	0.17 (0.12)
Mean Tobacco exposure in home (SD)*	0.58 (4.89)	0.62 (4.93)	0.51 (4.35)
Current tobacco smoking (**%** never/occasionally/most days)	91.4/ 6.0/ 2.6	93.5/ 4.7/ 1.8	90.4/ 7.0/ 2.6

*hours/week among non-smokers. SD: standard deviation

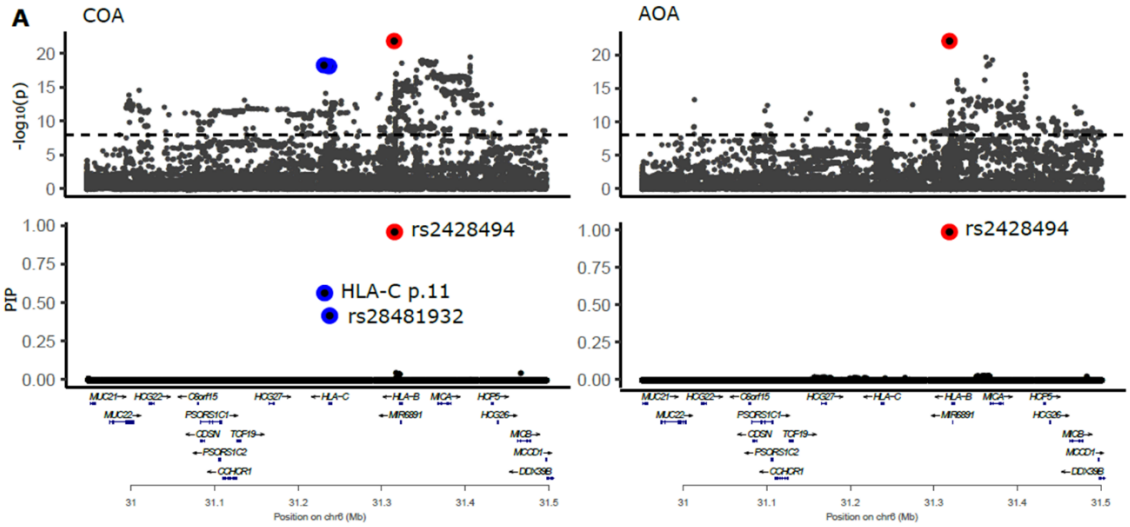
After filtering out low frequency amino acid polymorphisms (<1%), we tested the 741 amino acid polymorphisms at the six HLA loci for associations with COA and AOA. Of these, 188 amino acid polymorphisms were associated with COA and 152 were associated with AOA ($p<5x10^{-8}$, **Table S2.5, File S2.1**). P-values were overall more significant and estimated ORs were larger for class II compared to class I HLA alleles and amino acid polymorphisms (**Figure S2.3**). The magnitude of the ORs were also generally larger for COA compared to AOA. Both of these observations are consistent with the GWAS results (15). Using a test for heterogeneity, eight *HLA-DQA1*, seven *HLA-DQB1*, and 11 *HLA-DRB1* amino acid polymorphisms were significant, suggesting age of onset-specific effects (**Table S2.4**).

2.4.2 Fine-Mapping the HLA Class I Region

To fully capture genetic variation at the HLA region, we combined the genotypes for 19,499 HLA alleles, amino acid polymorphisms and SNPs, and performed genetic fine mapping on the combined set of variants separately in the class I and class II regions, with the goal of identifying causal variation for COA and AOA. We used the Bayesian regression method SuSiE (91) to perform fine mapping of the 9,021 combined variants at the class I locus using the locus boundaries defined previously (15).

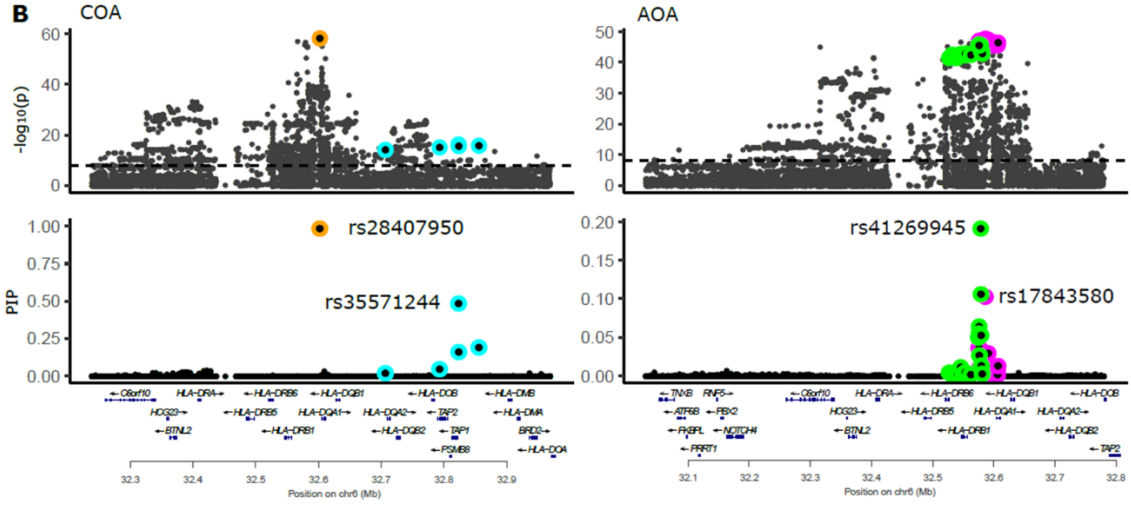
Fine mapping the COA class I locus identified two credible sets, indicating the presence of two independent associations with COA in the class I region. Credible set 1 (CS1) consisted of a single variant and CS2 consisted of two variants. The CS1 SNP (rs2428494) (red point, **Figure 2.1A**) is in an intron of *HLA-B*, and its posterior inclusion probability (PIP) is 0.97. This was the lead SNP in both the COA and AOA GWASs at the HLA class I locus(15). In CS2 (blue points, **Figure 2.1A**), the probability was nearly equally divided between two highly correlated variants (LD $r^2 = 0.99$, calculated from our data), with PIP values of 0.43 and 0.57 for rs28481932 and HLA-C p.11, respectively. SNP rs28481932 is upstream of *HLA-C* and HLA-C p.11 is an amino acid polymorphism (p.11 Ala/Ser). The risk amino acid (alanine) is on *HLA-C*02, *03, *05, *06, *07, *08, *12, *15, *16, *17*, and **18* alleles in this sample (including rare alleles), and the amino acid (serine) that is associated with decreased risk of asthma is on *HLA-C*01:02, *04:01, *04:07*, **14:02*, and **14:03* alleles. The HLA class I locus in AOA included one credible set with rs2428494 having a PIP of 1.00 (**Figure 2.1A**). Therefore, in the Class I region, rs2428494 is a shared causal SNP for COA and AOA and HLA-C p.11 or rs28481932 is a causal variant for COA. Alternatively, one or both may tag untyped or rare causal variation in LD with the candidate variants identified by SuSiE.

The odds ratios (ORs) remained largely unchanged when accounting for allergy in the associations, and no sex differences in risk were observed for any of these variants, suggesting that the associations were not reflecting confounding due to the presence of allergic diseases in either COA or AOA (**Table S2.6, S2.7**).



Class I Fine-Mapping Results

	CS	Var	PIP	GWAS p-val	GWAS OR	95% CI	Allele (Non-risk/Risk)	RAF
COA	CS1 (red)	rs2428494	0.97	8.77×10^{-23}	1.16	1.12-1.19	T/A	0.47
	CS2 (blue)	HLA-C p.11	0.57	3.12×10^{-19}	1.24	1.18-1.30	Ser/Ala	0.87
		rs28481932	0.43	4.36×10^{-19}	1.24	1.18-1.30	G/A	0.87
AOA	CS1 (red)	rs2428494	1.00	4.52×10^{-23}	1.10	1.08-1.13	T/A	0.47



Class II Fine-Mapping Results

	CS	Var	PIP	GWAS p-val	GWAS OR	95% CI	Allele (Non-risk/Risk)	RAF
COA	CS1 (orange)	rs28407950	1.00	1.37×10^{-59}	1.35	1.31-1.40	T/C	0.75
	CS2 (cyan)	rs35571244	0.50	1.07×10^{-17}	1.25	1.19-1.32	T/C	0.07
		rs35599935	0.20	3.39×10^{-17}	1.25	1.19-1.32	T/C	0.07
		rs34975158	0.17	4.73×10^{-17}	1.25	1.19-1.32	G/A	0.07
		rs4148878	0.06	1.60×10^{-16}	1.24	1.18-1.31	T/G	0.07
		rs33998906	0.03	1.42×10^{-15}	1.24	1.18-1.31	C/T	0.07
AOA	CS1 (magenta)	rs17843580	0.08	4.01×10^{-49}	1.17	1.14-1.19	A/G	0.60
		+ 32 SNPs, 19 HLA-DQA1 amino acids, 8 HLA-DQB1 amino acids						
	CS2 (green)	rs41269945	0.19	5.08×10^{-47}	1.20	1.17-1.22	A/T	0.19
		*+ 21 SNPs, HLA-DQA1*03:01, 5 HLA-DQA1 amino acids, 5 HLA-DRB1 amino acids*						

Figure 2.1. Fine-Mapping Results for Childhood- and Adult-Onset Asthma. In the **A)** HLA class I and **B)** HLA class II region, the upper panels show the -$\log_{10}$(p-values) from the GWAS (15); the dashed line indicates genome-wide significance. The lower panels show the **Figure 2.1**

Figure 2.1. Fine-Mapping Results for Childhood- and Adult-Onset Asthma. (continued) fine-mapping probabilities (PIPs) for the same variants. The colors represent the level-95% credible sets (HLA class I: CS1=red, CS2=blue; HLA class II CS1=orange and CS2=cyan for COA, and CS1=magenta, CS2=green for AOA). The tables show summaries of the variants in each credible set. A total of 9,021 variants (SNPs, HLA alleles, amino acid polymorphisms) were examined in the class I region and 10,428 variants in the class II region.

2.4.3 Fine-Mapping the HLA Class II Region

We used 10,428 combined variants at the class II region for fine mapping. Two credible sets were identified for COA (**Figure 2.1B**). CS1 (orange point) included one SNP (rs28407950; PIP 1.00) located downstream of *HLA-DQB1*. This SNP was the lead GWAS SNP in the HLA class II region for COA(15). CS2 (cyan points) contained five SNPs spanning 152 kb. The SNP (rs35571244) with the highest PIP (0.50) was located at the proximal end of the class II region upstream of *TAP1*. The minimum r^2 between all variants in CS2 was 0.79 (median r^2=0.99). No HLA alleles or amino acid polymorphisms were included in either of the COA credible sets.

Two credible sets were also identified for AOA, but neither included variants in the COA credible sets (**Figure 2.1B**, **Table S2.8**). CS1 (magenta points) contained 60 variants: 33 SNPs, 19 *HLA-DQA1* amino acid polymorphisms, and eight *HLA-DQB1* amino acid polymorphisms. The minimum r^2 between all variants in CS1 was 0.94 (median r^2=0.99), spanning 32.1 kb across the *HLA-DQA1* and *HLA-DQB1* genes. The lead SNP in the AOA GWAS (15) (rs17843580), which is located downstream of *HLA-DQA1*, was also the lead SNP in CS1 (PIP 0.08), but because there were so many variants in this credible set all individual PIPs were small. CS2 (green) spanned 54.6 kb from *HLA-DRB5* to *HLA-DQA1*, and included 33 variants: 22 SNPs, five *HLA-DQA1* amino acid polymorphisms, one *HLA-DQA1* allele (*HLA-DQA1**03:01), and five *HLA-DRB1* amino acid polymorphisms. The minimum r^2 between all CS2 variants was 0.88 (median r^2=0.96). The variant with the highest PIP was a SNP (rs41269945) located between *HLA-DQA1* and *HLA-DQB1*. Five perfectly correlated *HLA-DQA1* amino acids had the highest

PIPs (each 0.07) among the amino acids: Thr26, Gln47, Arg56, Val76, and Thr187, which together define *HLA-DQA1**03 alleles (*03:01, *03:02, and *03:03 in our data).

The association of these putatively causal SNPs with AOA were not due to confounding by inclusion of allergic diseases, as the results were similar when excluding individuals with these comorbidities (**Table S2.6**). However, an interaction between the COA class II CS1 SNP (rs28407950) and sex was nominally significant (p=0.02) (**Table S2.7**), reflecting a nominally larger effect in females (OR=1.43 95% CI [1.35-1.51]) compared to males (OR=1.31 95% CI [1.25-1.37]) for this variant.

2.4.4 Fine-Mapping Simulations in the HLA Region

Most previous fine-mapping studies excluded the HLA region due to its genomic complexities (104–106). To validate that SuSiE accurately detects multiple independent causal signals in this region, we conducted simulations in each of the HLA class I and class II regions for both binary (e.g. case-control status) and quantitative (e.g. gene expression) outcomes (see Supplementary Methods for additional details). In the null simulations (with zero causal variants), SuSiE accurately reported no credible sets (**Figure S2.4**). In the simulations ranging between one to three causal variants, SuSiE correctly detected the correct number of causal signals, each containing the true causal variant in the class I and class II regions for both the binary and quantitative outcomes. The true causal variant had the highest PIP 61% of the time. These simulations demonstrated that fine-mapping studies in the HLA region with SuSiE can accurately identify credible sets containing the true causal variants despite the complexities of this region.

2.4.5 Fine-Mapping eQTLs and Functional Annotations in the HLA Region

Genetic variation can influence disease risk by altering protein function or by influencing expression levels of disease-associated genes. Earlier studies demonstrated different functional properties of HLA alleles defined by amino acid polymorphisms (82,107), while more recent studies have shown regulatory effects of SNPs on HLA gene expression (108–112). Our results suggested that both types of mechanisms may mediate the effects of HLA genes on asthma risk. We first assessed the potential regulatory effects of asthma-associated SNPs using gene expression data from three asthma-relevant cell types (**Table S2.1**) from individuals with HLA types that were known or could be determined. This allowed us to map the RNA-sequencing (RNA-seq) data against sequences corresponding to each person's known HLA type (**Table S2.9**) and avoid mapping biases that arise from the large number of sequence differences between HLA alleles and reference genome (45).

We first performed eQTL studies for all genes whose transcription start site (TSS) was ± 1 Mb of the SNPs in each credible set. This identified 245 significant eQTLs (False Discovery Rate, FDR<0.05) involving 46 SNPs and 17 genes (*CCHCR1*, *AL662844.4*, *MIR6891*, *PSBM9*, *TAP1*, *TAP2*, *PPT2*, *HLA-B*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DQB1-AS1*, *HLA-DQA2*, *HLA-DQB2*, *HLA-DRB5*, *HLA-DRB6*, *HLA-DRB9*, and *HLA-DPB2*) (**Tables S2.10, S2.11, File S2.2**). We then used SuSiE to perform fine mapping for each gene with at least one eQTL in that cell type, resulting in expression fine mapping of 15 genes in LCLs, three in PBMCs, and six in NECs (**Table S2.12**). Five of the eQTL fine-mapping studies identified credible sets with SNPs that were also in the class II AOA CS1 (**Table S2.13, File S2.3**). These included credible sets with eQTLs for *HLA-DQB2* in LCLs, PBMCs, and NECs (**Figure 2.2A**) and *HLA-DQA2* in LCLs and PBMCs (**Figure 2.2B**). The AOA CS1 risk alleles were associated with increased

expression of both *HLA-DQA2* and *HLA-DQB2* in these cells (**Figure 2.2C, Figure S2.5**). None of the other SNPs in eQTL fine-mapping credible sets overlapped with SNPs in the class II AOA CS2, the class II COA CS1 or CS2, or the AOA or COA class I credible sets.

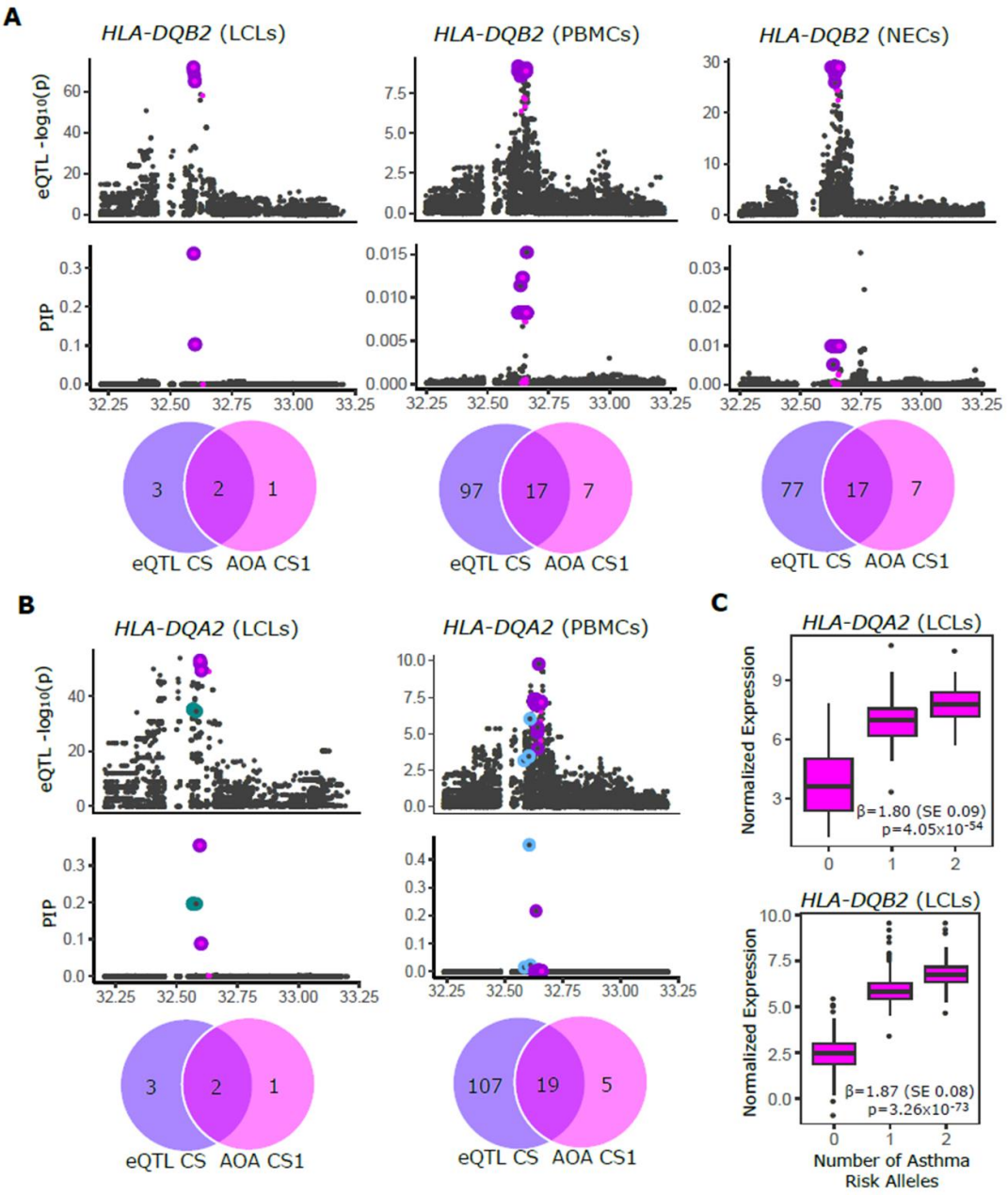


Figure 2.2. eQTL Fine Mapping and Functional Annotations. For **A)** *HLA-DQB2* and **B)** *HLA-DQA2*, upper panels show eQTL -$\log_{10}$(p-values) and PIPs for each variant tested within ± 0.5 Mb of the gene TSS. The colored outlines show the eQTL credible sets; the magenta points are SNPs that were also in the class II AOA CS1. A purple outline around a magenta point shows the variants shared between the eQTL credible set and the class II AOA CS1. The green and blue

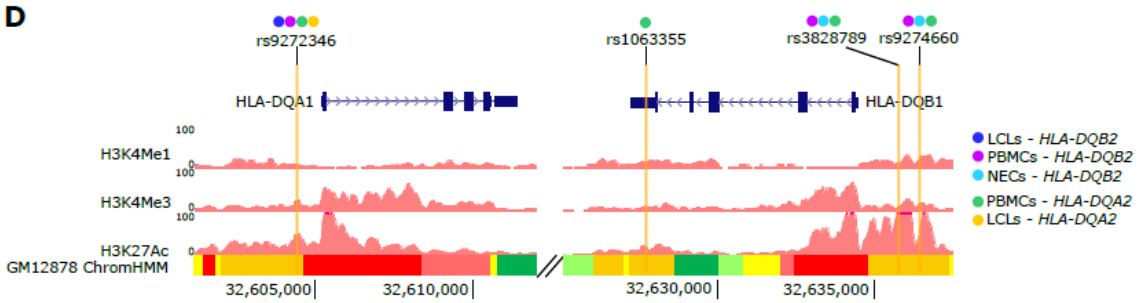


Figure 2.2. eQTL Fine Mapping and Functional Annotations. (continued) outlines indicate other credible sets. X-axis position is in Mb (hg19). The Venn diagrams show the number of variants in the eQTL credible set and the number that overlaps with SNPs in the class II AOA CS1. **C)** Normalized expression of *HLA-DQA2* (upper panel) and *HLA-DQB2* (lower-panel) in LCLs by the number of asthma-risk alleles (A) for rs9272346, a representative class II AOA CS1 SNP. **D)** Four SNPs in the class II AOA CS1 and in one or more eQTL credible set overlapped with functional annotations. The colors above each rsID indicate which eQTL credible set they were in. All four SNPs overlapped with strong enhancers in GM12878 cells (LCLs; bottom track). The GM12878 ChromHMM results correspond to active promoters (red), weak promoters (light red), strong enhancers (orange), weak/poised enhancers (yellow), transcriptional transition/elongation (dark green), and weak transcription (light green). Modified from http://genome.ucsc.edu(113).

To prioritize among the 20 SNPs that were in both the class II AOA CS1 and eQTL credible sets for *HLA-DQA2* or *HLA-DQB2*, we overlapped these SNPs with functional annotations from nine cell lines (GM12878 (LCLs), H1-hESC, K562, HepG2, HUVEC, HMEC, HSMM, NHEK, NHLF) available from ENCODE (114). Four of the 20 SNPs overlapped an enhancer region in LCLs (**Figure 2.2D**) and also resided in or near (approximately 70bp to 700bp) weak enhancer marks in three epithelial cell-derived lines (NHEK, keratinocytes; HMEC, mammary epithelial cells; HEPG2; liver hepatocellular cancer) (**Figure S2.6**). None of the SNPs overlapped with or were near enhancer marks in the other ENCODE cell lines. rs9272346, located upstream of *HLA-DQA1*, was in four of five of the eQTL credible sets. The other three SNPs were located near or within *HLA-DQB1*, with rs1063355, rs3828789, and rs9274660 in one, three, and three of the five eQTL credible sets, respectively. Surprisingly, the AOA GWAS lead SNP (rs17843580), which had the highest PIP in the AOA CS1, and the

remaining 15 SNPs in CS1 did not overlap an enhancer region in any of the cell types (**Figure S2.6**).

2.4.6 Structural Visualization of Amino Acid Variants

HLA class I and class II molecules bind and present peptides to T cell receptors (TCRs). The polymorphic features of HLA class I and class II molecules, particularly in the peptide binding domain, serve the crucial functions of diversifying antigen presentation and restricting pathogen-evasion of immune recognition. Therefore, we next explored the possibility that the amino acid variants with the highest PIPs in some of the credible sets affect peptide presentation or interactions with the TCR (**Table S2.14**). The amino acid in the class I region with the highest PIP in the COA CS2, HLA-C p.11, is located within the peptide-binding pocket of the HLA-C protein (**Figure 2.3A**). The serine allele was associated with protection from COA and is polar and uncharged; the alanine was associated with risk and is aliphatic and hydrophobic. Thus, both the location and structure of this polymorphic site suggest that peptide presentation by or other functional properties of HLA-C may be different between these alleles.

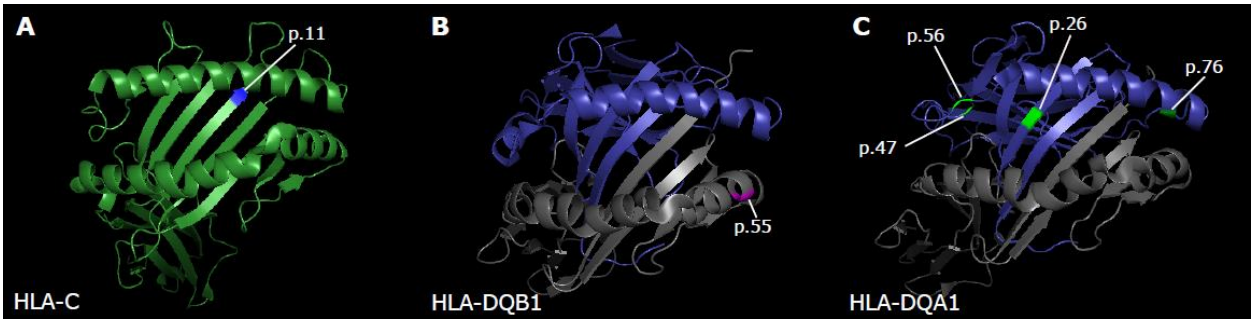


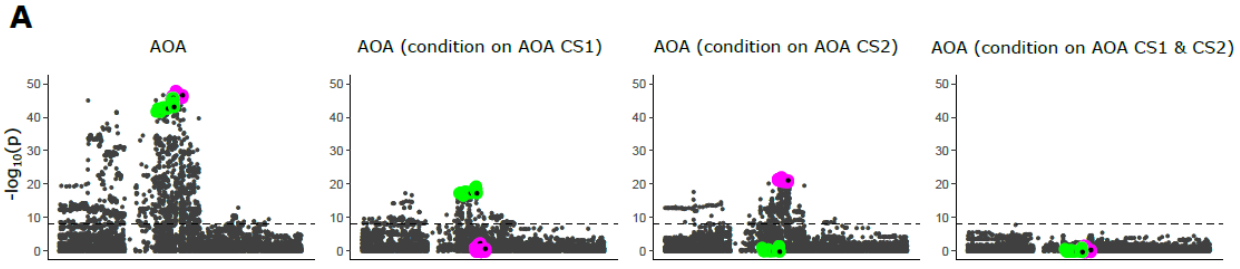
Figure 2.3. Localization of Asthma-Associated Amino Acid Variants.
Ribbon-figure representations of the peptide-binding pocket are shown for each HLA protein, and the amino-acid variant in focus is highlighted. **A)** HLA-C p.11, shown in blue, is located within the peptide-binding pocket of the HLA-C molecule(99) (forest green). **B)** HLA-DQB1 p.55, shown in magenta, lies in the region that may interact with the T-cell receptor on the HLA-DQB1(100) protein (grey) in complex with HLA-DQA1 (blue). **C)** HLA-DQA1 p.26, p.47, p.56, and p.76 (green) shown on the HLA-DQA1 protein(101) (blue). p.26 lies in the peptide-binding pocket, p.76 in the region that may interact with the TCR, and p.56 and p.47 in regions outside of the peptide-binding pocket.

The amino acid with the highest PIP in the AOA class II CS1 was at position 55 in the HLA-DQB1 protein (**Figure 2.3B**). Arginine contains a positively charged side chain and was associated with protection from AOA ($p=4.50x10^{-49}$; OR=0.86, CI 0.84-0.88). This is a multiallelic position, with leucine and proline as alternate alleles; both were associated with asthma risk (Leucine: $p=3.36x10^{-6}$; OR=1.06, CI 1.03-1.08; Proline: $p=1.83x10^{-28}$; OR=1.12, CI 1.10-1.15). The risk variants were both hydrophobic whereas the protective variant was positively charged and polar. This location may be in a region that interacts with the TCR (115). Because this variant is in strong LD with SNPs that are eQTLs in the class II AOA CS1 (median $r^2=0.99$), it is unclear if this variant, the eQTLs, or both are causal for AOA.

Among the 10 amino acids in class II AOA CS2, HLA-DQA1 Ser26, Gln47, Arg56, and Val76 were in perfect LD with each other, had the highest PIPs of the amino acid polymorphisms in CS2, and were associated with AOA risk. These amino acids are present exclusively on the *HLA-DQA1**03:01, 03:02, and 03:03 alleles (**Table S2.15**). The *HLA-DQA1**03:02 and *HLA-DQA1**03:03 alleles occurred at low frequency (<1%) in this sample and were not included in the GWAS or fine-mapping studies. Of these four amino acids, Ser26 is in the peptide-binding pocket and Val76 is in a region that may interact with the TCR (**Figure 2.3C**). The other amino acid polymorphisms were not in regions with obvious functional effects. At position 26, both the risk-associated serine and the protection-associated threonine have polar uncharged side chains, although the serine sidechain is smaller. Position 76 was multiallelic, with all three amino acids (valine, leucine, and methionine) having hydrophobic side chains, and valine having the smallest molecular weight. Position 187 was also perfectly correlated with these amino acids and captured in CS2 but was not part of the crystal structure. Our data support *HLA-DQA1**03 alleles as risk alleles for AOA.

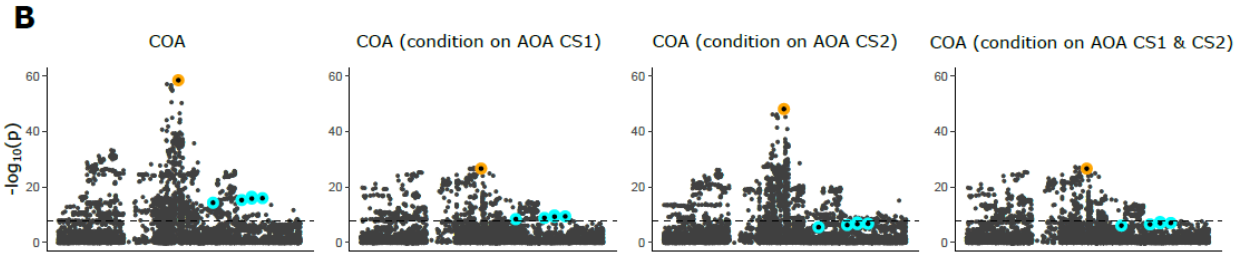
2.4.7 Conditional Analyses to Assess Independent Effects

To determine whether the candidate eQTL SNP for *HLA-DQA2* and *HLA-DQB2* (rs9272346) in CS1 and the *HLA-DQA1**03:01 allele in CS2 accounts for all the association signal at the class II AOA locus, we performed three conditional analyses in which we included the number of alleles for rs9272346, for *HLA-DQA1**03:01, and for both as covariates in association tests of variants at this locus with AOA (**Figure 2.4A**). In the first conditional analysis including number of rs9272346 (CS1) alleles as a covariate, the AOA association signal for *HLA-DQA1**03:01 remained genome-wide significant; in the second conditional analysis including the number of *HLA-DQA1**03:01 alleles (CS2) as a covariate, the AOA association signal for rs9272346 also remained genome-wide significant. When both rs9272346 and *HLA-DQA1**03:01 were included as covariates, the significance across the locus was reduced. These analyses indicate that the most significant asthma locus in AOA is due to variation in two credible sets, whose effects are likely attributable to their impact on expression of the *HLA-DQA2* and *HLA-DQB2* genes and of the *HLA-DQA1**03:01 allele.



		Associations		Condition on AOA CS1		Condition on AOA CS2	
CS	variant	p-value	OR [95% CI]	p-value	OR [95% CI]	p-value	OR [95% CI]
AOA CS1	rs9272346	1.98×10^{-47}	1.16 [1.14-1.19]	NA	NA	5.48×10^{-22}	1.12 [1.10-1.14]
AOA CS2	DQA1*03:01	6.91×10^{-47}	1.19 [1.16-1.22]	6.21×10^{-20}	1.13 [1.10-1.16]	NA	NA

Figure 2.4. Class II Conditional Analyses. For **A)** adult-onset asthma (AOA) and **B)** childhood-onset asthma (COA), results show the -$\log_{10}$(p-value) for each variant after conditioning on rs9272346 (representative AOA CS1 SNP), and/or *HLA-DQA1**03:01 (representative AOA CS2 variant). Colored outlines correspond to different credible sets; AOA



		Associations		Condition on AOA CS1		Condition on AOA CS2		Condition on AOA CS1 & AOA CS2	
CS	SNP	p-value	OR [95% CI]	p-value	OR [95% CI]	p-value	OR [95% CI]	p-value	OR [95% CI]
COA CS1	rs28407950	$1.37x10^{-59}$	1.35 [1.31-1.40]	$7.80x10^{-28}$	1.32 [1.26-1.39]	$2.92x10^{-49}$	1.33 [1.28-1.38]	$8.04x10^{-28}$	1.32 [1.26-1.39]
COA CS2	rs35571244	$1.07x10^{-17}$	1.25 [1.19-1.32]	$5.62x10^{-11}$	1.19 [1.13-1.26]	$2.04x10^{-08}$	1.18 [1.11-1.25]	$1.30x10^{-08}$	1.18 [1.11-1.25]

Figure 2.4. Class II Conditional Analyses. (continued) CS1: magenta, AOA CS2: green, COA CS1: orange, and COA CS2: cyan. See Fig. 1 for comparison.

To confirm that the two class II AOA putatively causal variants do not contribute to class II COA risk, we repeated the analysis above for SNPs at the COA class II GWAS locus (**Figure 2.4B**). When conditioning on either the AOA CS1 SNP rs9272346, the AOA CS2 *HLA-DQA1**03:01 allele, or both, the COA associations remain genome-wide significant, although the magnitudes of the associations are reduced, likely due to including additional covariates in the model and the LD in the region (**Table S2.16**). These results further support the argument that risk for COA and AOA are due to different causal variants in the HLA class II region.

Finally, we performed conditional analyses to assess the independence between the class I and class II signals. For each of the COA class I variants, we tested for their association with COA after conditioning on the tag class II SNPs and vice versa for the class II region. We similarly did this for AOA. For all results, the odds ratios (ORs) are largely similar and the 95% confidence intervals (CIs) overlap between the marginal and conditional associations, suggesting that the class I and class II signals are indeed independent (**Table S2.17**).

2.4.8 Replication of Fine-Mapping Results

To replicate the COA and AOA putatively causal SNPs identified in the UK Biobank White British ancestry individuals, we examined a replication cohort of UK Biobank multi-ethnic individuals who were initially excluded from our studies. Asthma and allergy phenotypes were defined using the same criteria as in the discovery sample. The prevalence of both were similar in the discovery and replication samples (**Table S2.18, S2.19**). To allow for allele frequency and effect size heterogeneity between the replication cohorts (n=43,449 White British, n=10,327 Asian or Asian British, n=7,637 Black or Black British), we tested each variant for association with COA and AOA within each cohort and then performed a meta-analysis of the results. We required that the same allele is associated with asthma with the same direction of effect as in the discovery cohort. All of the variants, except the class II COA CS2 and the class I AOA CS1 SNPs, replicated at a significance threshold adjusted for multiple testing of $5.0x10^{-3}$ (**Table 2.2, Figure S2.7**, **Table S2.20**). Additionally, the *HLA-DQA1**03:01 allele had the most significant association for AOA compared to the other HLA alleles tested.

Table 2.2 Results of Replication Meta-analysis. P-values, odds ratios (ORs), and 95% confidence intervals (95% CI) in the replication cohorts are reported for each of the candidate variants from the discovery COA and AOA credible sets.

Group	**CS**	**Variant**	**Allele (non-risk/risk)**	**p-value**	**OR**	**95% CI**
COA	Class I CS1	rs2428494	T/A	$9.72x10^{-04}$	1.13	1.05-1.21
	Class I CS2	HLA-C p.11	Ser/Ala	$2.66x10^{-03}$	1.16	1.05-1.29
	Class II CS1	rs28407950	T/C	$1.30x10^{-09}$	1.29	1.19-1.40
	Class II CS2	rs35571244	T/C	$2.74x10^{-01}$	1.08	0.94-1.25
AOA	Class I CS1	rs2428494	T/A	$1.27x10^{-02}$	1.06	1.01-1.12

Table 2.2 Results of Replication Meta-analysis. (continued)

AOA	Class II CS1	rs9274660	A/G	$6.10x10^{-08}$	1.14	1.09-1.20
		rs3828789	G/T	$4.59x10^{-08}$	1.15	1.09-1.20
		rs1063355	T/G	$4.31x10^{-08}$	1.15	1.09-1.20
		rs9272346	G/A	$7.52x10^{-08}$	1.14	1.09-1.20
	Class II CS2	HLA-DQA1	*03:01	$1.27x10^{-04}$	1.13	1.06-1.21

Overall, all but two of the candidate variants (AOA rs2428494 and COA rs35571244) from the discovery cohort were significantly associated with COA or AOA in the replication cohort with the same direction of effect. Both variants were nominally associated with COA or AOA, with the same directions of effect, but were not significant after multiple test correction.

2.5 DISCUSSION

The HLA region is associated with more diseases than any other region of the genome (116), and variation in this region has been consistently associated with asthma risk in GWASs (14–17,31–33). However, the causal variants and genes have been unknown. Most previous large studies focused only on SNPs, which do not fully capture the extensive protein polymorphism at this locus. A recent study reported colocalizations between eQTLs for *HLA-B*, *HLA-DQB1*, *HLA-DQA1*, *HLA-DRA*, *TAP1*, and *RNF5* in induced pluripotent stem cells with asthma GWAS SNPs (111). However, they did not separate COA and AOA, examine associations with HLA alleles or amino acids, or study gene expression in asthma-relevant cell types. Our study addressed these limitations and used relevant cell types and eQTL fine mapping to identify putatively causal eQTLs. These studies extended our earlier observations of COA and AOA having both shared and distinct genetic risk to the HLA region. We also prioritized putatively causal variation by examining the location of associated amino acid variants within the functional domains of HLA

proteins and of associated SNPs to functional annotations of gene regulation. In the class I region, we identified both a COA-specific association that may be mediated by HLA-C protein coding variation and a shared causal variant with unknown function. We did not find any shared causal variation between COA and AOA in the class II region, which was supported by replication studies. Nevertheless, our data strongly suggest that HLA class II-associated risk for AOA is mediated by both protein coding variation associated with the *HLA-DQA1**03 alleles and differential expression of the nonclassical *HLA-DQA2* and *HLA-DQB2* genes.

Our fine-mapping studies in White British individuals from the UK Biobank revealed a lead GWAS SNP in an intron of *HLA-B* in the class I region that was putatively causal for both COA and AOA (rs2428494). However, rs2428494 was not in any eQTL credible sets. This SNP was also not an eQTL for any genes in GTEx tissues (117) or in immune cells in the Database of Immune Cell Expression, eQTLs, and Epigenomics (DICE) (118), and did not reside in ENCODE (114) *cis* regulatory elements (**Figure S2.6**). However, rs2428494 was predicted to be in flanking active TSSs for several T cell subsets in Roadmap (119). Additionally, this association was replicated in a multi-ethnic cohort and was the lead class I region SNP in a meta-analysis of GWASs for allergic rhinitis, a common co-morbidity with asthma, with the same allele associated with risk (25). Otherwise, little is known about this variant.

A second credible set in the class I region included a SNP (rs28481932) and an HLA-C amino acid polymorphism at position 11 and was specific to COA. We did not accrue any functional evidence for the SNP to mediate its effects through gene expression based on our eQTL studies and co-localizations with ENCODE annotations (**Figure S2.6**). However, the HLA-C amino acid polymorphism had a higher PIP and its location in the HLA-C protein makes it a promising functional candidate. This amino acid polymorphism was also associated with

COA in the replication cohort. Along with other class I classical HLA genes, *HLA-C* is expressed on the cell surface of nearly all nucleated cells, where it presents intracellular peptides to the TCR of CD8 T cells. TCR recognition of foreign (e.g. viral) peptides presented by class I molecules activates these T cells, leading to the cytotoxic killing of infected cells by CD8 T cells and promoting inflammation. Position 11 lies in a β-sheet in the peptide-binding pocket and the amino acid substitution may change the peptide binding properties and alter antigen presentation and recognition by CD8 T cells. HLA-C is also a ligand for killer immunoglobulin receptors (KIRs) expressed on natural killer (NK) cells (120), which survey MHC class I levels on cell surfaces and can induce cell death when levels decrease during cell stress or viral infection (121). Modulation of NK cell activity may be another potential role for this variant in COA (122,123).

In the class II region, two credible sets were identified for COA, neither of which overlapped with the class II AOA credible sets. One of the credible sets (CS1, rs28407950) was replicated in the multi-ethnic cohort providing strong evidence that this SNP, or an untyped or rare variant in LD with it, is indeed a causal variant for COA. Additionally, we observed a nominally significant interaction between sex and rs28407950, with the risk variant having a stronger effect in females compared to males. Similarly, none of the class II COA CS2 SNPs were in eQTL credible sets and therefore not likely causal for resting gene expression in these cells. The fact that no HLA alleles or amino acids were included in either COA credible set largely rules out protein variation mediating risk at this locus for COA.

In contrast, fine-mapping studies identified the same causal variation underlying both AOA risk and expression of *HLA-DQA2* and *HLA-DQB2* genes at the class II region; these associations with AOA were replicated by fine-mapping studies in the multi-population cohort. While previous asthma GWAS have implicated the classical and highly polymorphic *HLA-*

DQA1 and *HLA-DQB1* class II genes(31,33,124), our study revealed associations with *HLA-DQA2* and *HLA-DQB2* genes in risk for AOA. The *HLA-DQA2* and *HLA-DQB2* genes are highly conserved paralogues of the *HLA-DQA1* and *HLA-DQB1* genes (125,126), respectively, but are virtually devoid of amino acid polymorphisms in the peptide binding pocket (125). Little is known about the functions of the HLA-DQα2/HLA-DQβ2 protein, although it has been shown to form heterodimers in Langerhans cells and can present antigens and activate T cells (127). It is notable that SNPs in the AOA CS1 were associated with increased *HLA-DQB2* and *HLA-DQA2* expression in different cell types, pointing to their potentially broad effects in both immune and airway epithelial tissues. The finding that asthma-associated SNPs in the class II region were associated with increased *HLA-DQA2* and *HLA-DQB2* expression is consistent with our earlier results showing that predicted increased expression of *HLA-DQA2* and *HLA-DQB2* (44) was among the most significant gene-based associations with asthma risk (15). Our current study confirms this prediction and further implicates increased expression of these highly conserved and poorly characterized genes as potential mediators of risk for AOA. Moreover, a recent GWAS of asthma hospitalizations in White British adults from the UK Biobank also reported that the HLA class II region was the most significant association with hospitalizations (128). Some of the associated SNPs were also eQTLs for HLA genes, including *HLA-DQA2*. This raises the possibility that the AOA class II CS1 variants in our study may also be associated with asthma hospitalizations and severity due to their effects on *HLA-DQA2* expression.

The *HLA-DQA1**03 alleles may also play an important role in AOA risk. The class II AOA CS2 contained the *HLA-DQA1**03:01 allele and five amino acids that define the *HLA-DQA1**03 alleles; some of these polymorphisms are in regions with potential impact on peptide presentation and TCR interactions (**Figure 2.3C**). We suggest therefore that *HLA-DQA1**03:01

may be the causal variant for AOA in CS2. T cell activation and proliferation is driven by both differential expression levels and protein coding variation in the HLA genes which may affect binding affinities to different peptides(129,130). The *HLA-DQA1**03:01 allele association was replicated, and it was the strongest HLA allele association in the replication cohort. Taken together, our studies indicate that increased expression of the *HLA-DQA2* and *HLA-DQB2* genes and coding variation in the HLA-DQA1*03 protein mediate the risks conferred by variation for AOA at the most significant GWAS locus (15).

Despite our delineation of specific HLA class I and class II variants and genes associated with risk for COA and AOA, our study had limitations. First, the fine-mapping method we used (SuSiE) has not yet been extended for logistic models or previously used in the HLA region. However, the use of linear methods to binary data can be justified (see section 2.6.1.1 for Supplementary Methods) (91,131–133), and our simulations indicate that SuSiE can accurately detect causal signals for binary traits (e.g. case-control status) as well as for quantitative traits (e.g. gene expression) in the genetically complex HLA region, at least in large sample sizes. Second, we did not identify potential mechanisms for all putatively causal signals for COA and AOA. In those cases, these variants may be eQTLs in other cell types not profiled in this study (134), in response to specific stimuli (135), or at specific developmental stages (136). Further experimental evidence is needed to elucidate the relationships between the putatively causal variants identified in this study and their impacts on gene function and/or expression, and ultimately on risk for asthma. Third, not all of the variants in the credible sets replicated, which may be due to smaller sample size or may reflect true ethnic differences in risk alleles. The extensive diversity of HLA genes and haplotypes between worldwide populations and the known effects of local selection pressures on these genes makes multi-ancestry replication particularly

challenging (137,138), and large cohorts are needed to replicate the fine-mapping results. However, the single variant association tests replicated nearly all representative variants in COA and AOA for class I and class II credible sets and the direction of effect was the same across discovery and replication cohorts. Given the importance of the HLA locus in asthma, additional studies need to be performed in larger cohorts from other racial and ethnic groups to gain a full picture of the roles of HLA genes in asthma. Finally, we discovered that HLA-DQA1*03:01 and a set of amino acids that are co-inherited on the HLA-DQA1*03 alleles were putatively causal for AOA. However, because the allele frequencies of the other HLA-DQA1*03 alleles (*03:02 and *03:03) were too infrequent in this sample to examine individually, we cannot determine if the effect is due solely to the HLA-DQA1*03:01 allele or a general effect of all *03 alleles.

In addition to age of asthma onset, many other known epidemiological factors distinguish asthma with onset in childhood compared to onset after puberty, including sex ratios, the importance of respiratory viral infections, and comorbidities with allergic diseases or obesity, as examples (139). Future studies can determine whether these well-established differences in COA and AOA may also be explained, at least in part, by differences in causal variation in the HLA region identified in this study.

2.5.1 Conclusions

Overall, our study highlights roles for both expression and protein coding variation in asthma risk. We suggest a prominent role for HLA-C protein coding variation in COA and both gene expression levels and protein coding changes in the HLA-DQ genes in AOA. We further propose that the *HLA-DQA1**03:01 allele and SNPs that regulate the expression of the under-characterized *HLA-DQA2* and *HLA-DQB2* genes explain the class II HLA risk at the most significant AOA locus. Our study identified potential therapeutic targets for asthma and utilized

a strategy that can serve as a model for fine mapping other HLA-associated diseases by integrating approaches to narrow putatively causal variants and genes in the HLA region.

2.5.2 Author Contributions

Conceptualization and design of study: Selene Marilyn Clay., Hae Kyung Im, Dan L. Nicolae, Carole Ober; Patient samples/data: James E. Gern, Daniel J. Jackson, Mario G. Rosasco, Matthew C. Altman, Matthew Dapas; Statistical Analyses: Selene Marilyn Clay, Andrew M. Goldstein, Peter Carbonetto, Matthew Stephens, Hae Kyung Im, Dan L. Nicolae; Interpretation of data: Selene Marilyn Clay, Carole Ober, and Nathan Schoettler; Writing: Selene Marilyn Clay and Carole Ober wrote the paper, all authors read and commented on drafts. All authors read and approved the final manuscript.

2.6 SUPPLEMENTARY INFORMATION

2.6.1 Supplementary Methods

2.6.1.1 Fine Mapping the HLA Region

We used Sum of Single Effects (SuSiE (91)) (susieR R package version 0.9.0) to fine map the HLA loci for childhood-onset asthma (COA) and adult-onset asthma (AOA). The susieR R package does not currently allow for the inclusion of covariates, so sex and the first 10 ancestral principal components (PCs) were regressed out of the genotype matrix and phenotype vector using linear regression; we used the residuals of the genotype matrix and phenotype vector as inputs to SuSIE. The SuSiE method is based on a linear regression, and so when applied to binary data, it will estimate and test for effects in terms of risk differences, rather than the more conventional odds ratio (OR). Applying linear methods to binary data is justified here because the estimated ORs were all small (<1.3), the allele frequencies were not too extreme, and the sample size (here, limited by the smaller number of cases) was large (131–133). See Pirinen *et al.* Section 3 for detailed discussion of applying linear methods to binary data and the relationship between estimated risk differences and ORs. We assumed at most $L = 10$ causal variants and set susieR to estimate the residual and prior variances. We retained only level-95% credible sets (coverage = 0.95). We took the additional step of discarding credible sets in which the "purity" (smallest absolute correlation among all pairs of variants within the credible set) was less than 0.50. We only considered credible sets that contained at least one variant reaching genome-wide significance to avoid any possible artifacts.

2.6.1.2. HLA Fine-Mapping Simulations

Because the HLA region is extraordinarily complex, we assessed the performance of SuSiE in this region by simulation. Existing genotype and covariate data were used to leverage the true

correlation structure in the class I and class II regions to simulate both binary (e.g. case/control status) and quantitative (e.g. gene expression) outcomes.

To simulate binary (e.g. case/control status) outcomes, we used the genotype matrix *X* (HLA class I or class II loci defined by Pividori *et al.*(15)) and covariate matrix *Z* from the UK Biobank and set the individual-level log-odds of asthma to be

$$\ln \frac{p_i}{1-p_i} = \sum_j \beta_j X_{ij} + \sum_k \delta_k Z_{ik} + \alpha$$

for individual *i*, SNPs *j*, covariates *k*, fixed effect vectors β and δ, and a fixed intercept α. We used the true matrix of covariates and covariate effects δ estimated from a logistic regression, separately for COA and AOA simulations. β_j was set to 0 for all non-causal variants. For causal variants, $\beta_j$was set using effect sizes similar to what was found in the Pividori COA and AOA GWASs (15). We randomly selected 0-3 variants from a random uniform distribution to be causal (with non-zero effects) for both the class I and class II regions using both COA and AOA effect sizes. We simulated case/control status for each individual as $Y_i \sim Bernoulli(p_i)$ independently and regressed out the covariates in Z from *X* and *Y*.

For quantitative outcomes (e.g. gene expression), we used *X* and *Z* in the HLA class I and HLA class II regions from the nasal epithelial cell (NEC) dataset from URECA described below. We set the individual-level mean to be

$$\mu_i = \sum_j \beta_j X_{ij} + \sum_k \delta_k Z_{ik} + \alpha$$

and we used the true matrix of covariates and effects δ estimated from a linear regression. β_j was set to 0 for all non-causal variants, and causal $\beta_j$were set using effect sizes similar to what was found from the NEC eQTL studies (described below). We similarly randomly selected 0-3 causal signals in both the class I and class II regions and set the individual level response to be $Y_i \sim$

$N(\mu_i, \sigma^2)$. We similarly regressed out the covariates in Z from X and Y. These simulations were used to test how well SuSiE recovers the causal effects over each simulation.

2.6.1.3. Gene Expression and eQTL Studies

Lymphoblastoid Cell Lines (LCLs): We examined RNA-seq data previously collected from LCLs from 398 Hutterites (92). The Hutterites are a founder population of European descent with well characterized HLA types for the polymorphic *HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-G*, *HLA-DPB1*, *HLA-DRB1*, *HLA-DQB1*, and *HLA-DQA1* genes(140). The sample was composed of 191 males and 207 females who were between the ages of 10 and 60 at the time of sample collection. Informed consent was obtained from all participants under University of Chicago IRB-approved protocols.

Standard RNA-seq pipelines that map reads to a reference genome can provide biased expression estimates at the highly polymorphic HLA loci due to the potentially large number of differences between the sequence of an individual's HLA type and the reference sequence used for mapping (45,111). Expression estimates can be improved by mapping RNA-seq reads to the sequences for each individual's known HLA type (45). For the polymorphic HLA genes, we aligned RNA-seq reads to reference sequences from the IMGT database (40) for each individual's known HLA type, removing duplicate reads with WASP (141). Sequencing reads were mapped and quantified using STAR/2.6.1 (142) for other genes. Samples with >7M uniquely mapped reads underwent trimmed means of M-value (TMM) normalization and voom transformation (143). We corrected for extraction date and sequencing batch with limma (144).

To perform eQTL mapping, associations between SNPs and expression of genes in the HLA class I and class II regions were performed with Genome-wide Efficient Mixed Model Association (GEMMA) (96) using a kinship matrix to correct for relatedness between Hutterite

individuals. We used a linear mixed model (LMM) including age and sex as covariates and considered all variants within 1 Mb of the transcription start site (TSS) of each expressed gene.

Peripheral Blood Mononuclear Cells (PBMCs): We examined unstimulated PBMC RNA-seq data from 132 (78 males, 54 females) African-American children from the URban Environment and Childhood Asthma (URECA) birth cohort who were 2 years old at the time of sample collection (7,93). Whole genome sequencing (WGS) was performed using the Illumina NovaSEQ6000 with 150 bp paired-end reads. Reads were aligned to the GRCh38 human reference genome (including alternate loci and decoy contigs) using BWA-MEM (145) (Burrows-Wheeler Aligner; v0.7.17). Aligned reads underwent duplicate removal (Picard MarkDuplicates v2.8.1) and base quality score recalibration (GATK BaseRecalibrator; v3.8) against known sites (dbSNP138, known indels, and Mills and 1 KG gold standard indels) provided in the GATK resource bundle (146). Reads that mapped to the primary HLA region (chr6:28510120-33480577), reads that mapped to the GRCh38 HLA contigs, and unmapped reads were used for WGS HLA typing. We used HLA-LA (147) to infer HLA types from WGS for *HLA-A, HLA-B, HLA-C, HLA-E, HLA-F*, *HLA-DRB1*, *HLA-DQB1*, *HLA-DQA1*, *HLA-DPB1*, and *HLA-DPA1*. Reads were mapped and normalized as previously described. To perform eQTL mapping, we examined linear regressions with QTLtools (95), using a nominal pass and *cis*-window size of 1 Mb around the TSS. We included sex, collection site, the first three ancestral PCs, and 19 latent factors (148) to account for unwanted variation as covariates in the analysis.

Nasal Epithelial Cells (NECs): We examined NEC RNA-seq data from 188 (92 females, 96 males) African-American children (age 11 at time of sample collection) from the URECA cohort(149). As described above for the PBMCs, we used HLA-LA to infer HLA types from whole-genome sequences, mapped RNA-seq reads as described above, and used QTLtools to

perform eQTL mapping, using sex, the first three ancestral PCs, collection site, epithelial cell proportion, sequencing batch, and seven latent factors (148) as covariates in the analysis.

2.6.2 Supplementary Tables

Table S2.1. RNA-seq Sample Composition. Sample composition from RNA-seq data collected from LCLs (Hutterites (92)) and the PBMCs (7,93) and NECs (149) from URECA (URban Environment and Childhood Asthma).

Cells	Study	Ancestry	Sample Size	Age (yrs)	Sex
LCLs (EBV-transformed B cells)	Hutterites	European American	398	10-60	191 M, 207 F
Peripheral blood mononuclear cells (PBMCs)	URECA	African American	132	2	78 M, 54 F
Nasal epithelial cells (NECs)	URECA	African American	188	11	96 M, 92 F

Table S2.2. HLA Allele Associations. Results are shown for each HLA allele with frequency > 1%. Significant association p-values (<5x10-8) with are shown in bold font.

	Childhood-Onset Asthma					Adult-Onset Asthma				
HLA allele	**p-value**	**OR**	**95% CI**	**beta**	**SE**	**p-value**	**OR**	**95% CI**	**beta**	**SE**
A*01:01	5.23E-02	1.037	0.999-1.075	0.036	0.019	8.62E-03	1.033	1.008-1.059	0.033	0.013
A*02:01	6.05E-03	0.955	0.923-0.986	-0.046	0.017	8.83E-01	0.998	0.976-1.020	-0.002	0.011
A*03:01	1.86E-01	0.972	0.932-1.013	-0.028	0.021	1.17E-03	0.954	0.927-0.981	-0.047	0.014
A*11:01	8.79E-01	1.005	0.945-1.067	0.005	0.031	3.23E-02	0.955	0.916-0.996	-0.046	0.021
A*23:01	7.39E-01	1.019	0.911-1.139	0.019	0.057	9.43E-01	1.003	0.929-1.081	0.003	0.039
A*24:02	9.36E-01	1.002	0.947-1.060	0.002	0.029	1.88E-01	1.025	0.987-1.064	0.025	0.019
A*25:01	3.98E-02	1.125	1.005-1.258	0.118	0.057	7.72E-01	1.012	0.934-1.095	0.012	0.040
A*26:01	9.75E-01	1.002	0.899-1.115	0.002	0.055	9.45E-01	1.003	0.932-1.078	0.003	0.037
A*29:02	5.80E-04	1.128	1.053-1.208	0.120	0.035	3.81E-01	1.022	0.973-1.072	0.022	0.025
A*30:01	1.17E-01	1.116	0.972-1.281	0.110	0.070	9.47E-01	1.003	0.910-1.106	0.003	0.050
A*31:01	5.71E-01	1.026	0.938-1.121	0.026	0.045	2.58E-04	1.114	1.051-1.180	0.108	0.030
A*32:01	7.38E-01	1.014	0.936-1.096	0.013	0.040	8.10E-01	1.007	0.954-1.061	0.007	0.027
A*68:01	3.83E-02	0.911	0.834-0.995	-0.093	0.045	5.26E-01	0.982	0.926-1.039	-0.019	0.029
B*07:02	1.34E-01	0.969	0.929-1.009	-0.032	0.021	7.98E-07	0.932	0.905-0.958	-0.071	0.014
B*08:01	3.21E-06	1.100	1.056-1.144	0.095	0.020	**9.19E-10**	1.088	1.059-1.118	0.085	0.014
B*13:02	9.41E-01	1.004	0.903-1.115	0.004	0.054	1.90E-01	1.048	0.977-1.123	0.047	0.036
B*14:01	5.26E-01	0.956	0.831-1.099	-0.045	0.071	2.25E-01	0.943	0.858-1.036	-0.058	0.048
B*14:02	1.98E-01	0.940	0.854-1.033	-0.062	0.048	5.82E-03	0.913	0.856-0.974	-0.091	0.033
B*15:01	1.13E-03	1.101	1.039-1.167	0.097	0.030	3.87E-07	1.107	1.064-1.151	0.102	0.020
B*18:01	6.83E-01	1.016	0.940-1.097	0.016	0.039	1.58E-01	0.962	0.912-1.015	-0.038	0.027
B*27:05	5.55E-02	0.927	0.857-1.001	-0.076	0.040	9.12E-01	0.997	0.947-1.048	-0.003	0.026
B*35:01	**5.20E-13**	0.749	0.692-0.810	-0.289	0.040	2.22E-06	0.889	0.846-0.933	-0.118	0.025
B*37:01	5.27E-01	0.960	0.844-1.090	-0.041	0.065	2.47E-01	0.950	0.871-1.035	-0.051	0.044
B*40:01	2.63E-01	0.964	0.904-1.027	-0.037	0.033	6.73E-01	0.991	0.949-1.034	-0.009	0.022
B*44:02	**4.06E-10**	1.150	1.100-1.201	0.140	0.022	2.89E-03	1.048	1.016-1.080	0.046	0.016
B*44:03	4.73E-05	1.131	1.065-1.199	0.123	0.030	5.68E-02	1.041	0.998-1.084	0.040	0.021
B*49:01	9.84E-01	0.999	0.870-1.145	-0.001	0.070	7.44E-01	0.985	0.897-1.080	-0.015	0.047

Table S2.2. HLA Allele Associations. (continued)

B*51:01	1.50E-01	0.943	0.870-1.021	-0.059	0.041	3.20E-01	1.027	0.974-1.081	0.026	0.027
B*55:01	1.81E-01	0.927	0.829-1.035	-0.076	0.057	2.12E-01	0.954	0.885-1.027	-0.047	0.038
B*57:01	4.97E-04	0.869	0.802-0.940	-0.141	0.040	1.31E-05	0.890	0.843-0.937	-0.117	0.027
C*01:02	3.51E-06	0.812	0.743-0.886	-0.208	0.045	9.63E-02	0.955	0.903-1.008	-0.047	0.028
C*02:02	6.53E-01	1.018	0.942-1.099	0.018	0.039	1.39E-01	1.040	0.987-1.094	0.039	0.026
C*03:03	6.90E-01	1.013	0.951-1.078	0.013	0.032	9.66E-03	1.057	1.013-1.101	0.055	0.021
C*03:04	2.31E-01	1.033	0.979-1.088	0.032	0.027	6.03E-02	1.035	0.998-1.072	0.034	0.018
C*04:01	**6.16E-12**	0.819	0.774-0.867	-0.199	0.029	5.39E-06	0.919	0.885-0.952	-0.085	0.019
C*05:01	**6.35E-09**	1.138	1.089-1.189	0.129	0.022	8.18E-02	1.028	0.996-1.059	0.027	0.016
C*06:02	2.73E-02	0.943	0.895-0.993	-0.058	0.026	9.28E-03	0.955	0.922-0.988	-0.046	0.018
C*07:01	4.83E-04	1.069	1.029-1.109	0.067	0.019	1.51E-06	1.064	1.037-1.091	0.062	0.013
C*07:02	1.42E-01	0.970	0.932-1.010	-0.030	0.021	4.81E-06	0.938	0.912-0.964	-0.064	0.014
C*07:04	6.31E-03	1.156	1.041-1.283	0.145	0.053	1.43E-01	1.056	0.981-1.136	0.055	0.037
C*08:02	1.21E-01	0.939	0.866-1.016	-0.063	0.041	5.35E-03	0.926	0.877-0.977	-0.077	0.028
C*12:03	9.59E-02	0.926	0.846-1.013	-0.077	0.046	5.33E-01	0.981	0.924-1.041	-0.019	0.030
C*15:02	7.09E-01	0.979	0.873-1.095	-0.022	0.058	2.21E-01	1.047	0.972-1.128	0.046	0.038
C*16:01	8.91E-06	1.163	1.087-1.242	0.151	0.034	1.53E-02	1.059	1.011-1.109	0.058	0.024
DQA1*01:01	**1.98E-35**	0.747	0.713-0.782	-0.292	0.023	**1.89E-16**	0.885	0.859-0.911	-0.122	0.015
DQA1*01:02	1.84E-01	0.975	0.939-1.012	-0.025	0.019	**1.23E-13**	0.908	0.885-0.931	-0.097	0.013
DQA1*01:03	**5.08E-10**	0.802	0.748-0.859	-0.221	0.035	**1.45E-08**	0.878	0.839-0.918	-0.130	0.023
DQA1*02:01	9.52E-01	0.999	0.958-1.040	-0.001	0.021	9.54E-01	1.001	0.973-1.029	0.001	0.014
DQA1*03:01	**7.26E-17**	1.160	1.120-1.200	0.148	0.018	**6.91E-47**	1.187	1.159-1.215	0.172	0.012
DQA1*04:01	4.79E-02	0.896	0.802-0.998	-0.110	0.056	4.46E-01	0.973	0.905-1.044	-0.028	0.036
DQA1*05:01	**2.01E-14**	1.140	1.102-1.179	0.131	0.017	2.92E-03	1.036	1.012-1.060	0.035	0.012
DQB1*02:01	**2.93E-09**	1.126	1.082-1.171	0.119	0.020	8.52E-05	1.056	1.027-1.084	0.054	0.014
DQB1*02:02	2.84E-02	1.060	1.006-1.117	0.059	0.027	4.94E-02	1.036	1.000-1.073	0.036	0.018
DQB1*03:01	**4.15E-18**	1.175	1.132-1.218	0.161	0.019	**1.82E-09**	1.080	1.053-1.108	0.077	0.013
DQB1*03:02	4.65E-04	1.086	1.037-1.137	0.083	0.024	**5.68E-25**	1.174	1.138-1.210	0.161	0.016
DQB1*03:03	1.22E-03	0.894	0.834-0.956	-0.113	0.035	2.06E-01	0.972	0.929-1.015	-0.029	0.023
DQB1*04:02	4.19E-02	0.894	0.802-0.995	-0.112	0.055	2.05E-01	0.955	0.889-1.025	-0.046	0.036

Table S2.2. HLA Allele Associations. (continued)

DQB1*05:01	**4.18E-25**	0.772	0.735-0.810	-0.259	0.025	**1.73E-13**	0.890	0.862-0.917	-0.117	0.016
DQB1*05:03	**1.35E-11**	0.663	0.588-0.746	-0.411	0.061	5.07E-03	0.905	0.844-0.970	-0.099	0.036
DQB1*06:02	2.33E-02	1.049	1.006-1.092	0.048	0.021	**1.21E-08**	0.920	0.893-0.946	-0.084	0.015
DQB1*06:03	**9.88E-09**	0.811	0.755-0.871	-0.209	0.037	**5.10E-10**	0.862	0.821-0.902	-0.149	0.024
DQB1*06:04	2.47E-07	0.771	0.697-0.850	-0.261	0.051	4.26E-05	0.877	0.823-0.934	-0.131	0.032
DQB1*06:09	7.20E-02	0.868	0.743-1.012	-0.142	0.079	3.22E-01	0.951	0.860-1.050	-0.050	0.051
DRB1*01:01	**3.65E-23**	0.754	0.712-0.797	-0.283	0.029	**2.41E-11**	0.887	0.856-0.918	-0.120	0.018
DRB1*01:03	6.56E-04	0.806	0.712-0.912	-0.215	0.063	3.55E-01	0.965	0.893-1.041	-0.036	0.039
DRB1*03:01	**3.84E-10**	1.133	1.089-1.178	0.125	0.020	1.33E-04	1.054	1.026-1.083	0.053	0.014
DRB1*04:01	**8.56E-22**	1.236	1.183-1.290	0.212	0.022	**1.03E-34**	1.203	1.168-1.239	0.185	0.015
DRB1*04:04	1.51E-01	1.055	0.980-1.133	0.053	0.037	1.01E-05	1.114	1.061-1.168	0.108	0.024
DRB1*07:01	9.29E-01	0.998	0.957-1.040	-0.002	0.021	9.37E-01	0.999	0.971-1.026	-0.001	0.014
DRB1*08:01	5.66E-02	0.895	0.798-1.003	-0.111	0.058	6.42E-01	0.983	0.912-1.058	-0.018	0.038
DRB1*09:01	8.66E-01	1.011	0.890-1.148	0.011	0.065	3.00E-04	1.161	1.070-1.258	0.149	0.041
DRB1*11:01	**4.29E-10**	1.270	1.178-1.368	0.239	0.038	2.58E-02	1.064	1.007-1.123	0.062	0.028
DRB1*11:04	2.96E-01	0.926	0.801-1.069	-0.077	0.074	6.64E-01	0.979	0.890-1.077	-0.021	0.049
DRB1*12:01	5.07E-01	0.959	0.847-1.085	-0.042	0.063	1.14E-04	0.840	0.768-0.917	-0.175	0.045
DRB1*13:01	**1.96E-08**	0.814	0.757-0.874	-0.206	0.037	**1.69E-09**	0.865	0.824-0.906	-0.145	0.024
DRB1*13:02	7.27E-08	0.795	0.731-0.864	-0.230	0.043	2.23E-05	0.891	0.844-0.939	-0.115	0.027
DRB1*14:01	**4.54E-10**	0.671	0.591-0.760	-0.400	0.064	1.22E-02	0.910	0.845-0.979	-0.094	0.038
DRB1*15:01	5.51E-02	1.041	0.999-1.084	0.040	0.021	**3.56E-09**	0.917	0.891-0.944	-0.086	0.015

Table S2.3. Allele Associations: Additive vs. Dominant Model. For each HLA allele, the p-value, odds ratio (OR), and number of homozygotes are shown for childhood- and adult-onset asthma.

HLA allele	p-value	OR	p_dom	OR_dom	N_hom
Childhood-Onset Asthma					
A*01:01	5.23E-02	1.04	1.51E-02	1.05	12278
A*02:01	6.05E-03	0.95	1.45E-02	0.95	25258
A*03:01	1.86E-01	0.97	1.28E-01	0.96	7016
A*11:01	8.79E-01	1.00	9.28E-01	1.00	1207
A*23:01	7.39E-01	1.02	7.04E-01	1.02	93
A*24:02	9.36E-01	1.00	9.76E-01	1.00	1780
A*25:01	3.98E-02	1.12	1.77E-02	1.13	84
A*26:01	9.75E-01	1.00	9.65E-01	1.00	107
A*29:02	5.80E-04	1.13	3.14E-04	1.14	575
A*30:01	1.17E-01	1.12	1.08E-01	1.12	48
A*31:01	5.71E-01	1.03	4.71E-01	1.03	215
A*32:01	7.38E-01	1.01	7.16E-01	1.02	376
A*68:01	3.83E-02	0.91	3.14E-02	0.91	323
B*07:02	1.34E-01	0.97	1.09E-01	0.96	7337
B*08:01	3.21E-06	1.10	8.52E-08	1.13	6926
B*13:02	9.41E-01	1.00	8.09E-01	1.01	143
B*14:01	5.26E-01	0.96	5.80E-01	0.96	43
B*14:02	1.98E-01	0.94	1.97E-01	0.94	224
B*15:01	1.13E-03	1.10	1.62E-04	1.12	1408
B*18:01	6.83E-01	1.02	8.18E-01	1.01	453
B*27:05	5.55E-02	0.93	6.24E-02	0.93	501
B*35:01	5.20E-13	0.75	7.27E-12	0.76	767
B*37:01	5.27E-01	0.96	4.64E-01	0.95	70
B*40:01	2.63E-01	0.96	4.34E-01	0.97	1034
B*44:02	4.06E-10	1.15	3.25E-10	1.17	4170
B*44:03	4.73E-05	1.13	1.15E-04	1.13	1098
B*49:01	9.84E-01	1.00	9.94E-01	1.00	46
B*51:01	1.50E-01	0.94	2.61E-01	0.96	467
B*55:01	1.81E-01	0.93	2.05E-01	0.93	128
B*57:01	4.97E-04	0.87	4.98E-04	0.87	534
C*01:02	3.51E-06	0.81	4.67E-06	0.81	372
C*02:02	6.53E-01	1.02	7.13E-01	1.01	414
C*03:03	6.90E-01	1.01	6.54E-01	1.02	1090
C*03:04	2.31E-01	1.03	1.28E-01	1.04	2184
C*04:01	6.16E-12	0.82	2.95E-11	0.82	2318
C*05:01	6.35E-09	1.14	1.53E-09	1.16	4270
C*06:02	2.73E-02	0.94	4.66E-02	0.95	2725
C*07:01	4.83E-04	1.07	6.18E-05	1.09	10218

Table S2.3. Allele Associations: Additive vs. Dominant Model. (continued)

C*07:02	1.42E-01	0.97	1.41E-01	0.97	8299
C*07:04	6.31E-03	1.16	4.73E-03	1.16	116
C*08:02	1.21E-01	0.94	1.14E-01	0.94	411
C*12:03	9.59E-02	0.93	7.67E-02	0.92	269
C*15:02	7.09E-01	0.98	7.78E-01	0.98	108
C*16:01	8.91E-06	1.16	3.25E-06	1.18	624
DQA1*01:01	1.98E-35	0.75	5.96E-33	0.74	6799
DQA1*01:02	1.84E-01	0.98	5.96E-01	0.99	12290
DQA1*01:03	5.08E-10	0.80	3.39E-10	0.79	1021
DQA1*02:01	9.52E-01	1.00	5.09E-01	1.02	7018
DQA1*03:01	7.26E-17	1.16	1.01E-18	1.21	13460
DQA1*04:01	4.79E-02	0.90	4.37E-02	0.89	139
DQA1*05:01	2.01E-14	1.14	4.38E-15	1.18	17673
DQB1*02:01	2.93E-09	1.13	1.38E-12	1.17	7498
DQB1*02:02	2.84E-02	1.06	6.25E-03	1.07	3502
DQB1*03:01	4.15E-18	1.18	2.14E-19	1.22	10184
DQB1*03:02	4.65E-04	1.09	3.51E-04	1.10	3415
DQB1*03:03	1.22E-03	0.89	2.88E-03	0.90	928
DQB1*04:02	4.19E-02	0.89	4.19E-02	0.89	143
DQB1*05:01	4.18E-25	0.77	1.15E-22	0.77	4916
DQB1*05:03	1.35E-11	0.66	1.07E-11	0.66	150
DQB1*06:02	2.33E-02	1.05	1.42E-03	1.08	7011
DQB1*06:03	9.88E-09	0.81	9.71E-09	0.81	902
DQB1*06:04	2.47E-07	0.77	4.96E-07	0.77	249
DQB1*06:09	7.20E-02	0.87	8.32E-02	0.87	41
DRB1*01:01	3.65E-23	0.75	4.21E-21	0.76	3066
DRB1*01:03	6.56E-04	0.81	4.65E-04	0.80	122
DRB1*03:01	3.84E-10	1.13	8.77E-14	1.18	7380
DRB1*04:01	8.56E-22	1.24	2.11E-24	1.28	4557
DRB1*04:04	1.51E-01	1.05	1.62E-01	1.05	548
DRB1*07:01	9.29E-01	1.00	5.91E-01	1.01	7026
DRB1*08:01	5.66E-02	0.90	5.29E-02	0.89	125
DRB1*09:01	8.66E-01	1.01	8.23E-01	1.01	70
DRB1*11:01	4.29E-10	1.27	5.94E-10	1.25	489
DRB1*11:04	2.96E-01	0.93	3.41E-01	0.94	46
DRB1*12:01	5.07E-01	0.96	4.86E-01	0.96	74
DRB1*13:01	1.96E-08	0.81	2.77E-08	0.81	865
DRB1*13:02	7.27E-08	0.79	1.84E-07	0.80	467
DRB1*14:01	4.54E-10	0.67	5.27E-10	0.68	132
DRB1*15:01	5.51E-02	1.04	2.71E-03	1.07	7136
Adult-Onset Asthma					
A*01:01	8.62E-03	1.03	9.82E-04	1.05	12730
A*02:01	8.83E-01	1.00	6.92E-01	0.99	26275

Table S2.3. Allele Associations: Additive vs. Dominant Model. (continued)

A*03:01	1.17E-03	0.95	4.06E-03	0.95	7222
A*11:01	3.23E-02	0.96	2.10E-02	0.95	1254
A*23:01	9.43E-01	1.00	9.56E-01	1.00	98
A*24:02	1.88E-01	1.03	2.18E-01	1.02	1851
A*25:01	7.72E-01	1.01	9.48E-01	1.00	92
A*26:01	9.45E-01	1.00	8.95E-01	1.00	110
A*29:02	3.81E-01	1.02	3.39E-01	1.02	596
A*30:01	9.47E-01	1.00	9.83E-01	1.00	52
A*31:01	2.58E-04	1.11	7.53E-04	1.11	240
A*32:01	8.10E-01	1.01	8.28E-01	1.01	391
A*68:01	5.26E-01	0.98	6.57E-01	0.99	329
B*07:02	7.98E-07	0.93	4.27E-06	0.93	7557
B*08:01	9.19E-10	1.09	2.26E-12	1.12	7179
B*13:02	1.90E-01	1.05	1.66E-01	1.05	152
B*14:01	2.25E-01	0.94	1.84E-01	0.94	48
B*14:02	5.82E-03	0.91	6.98E-03	0.91	230
B*15:01	3.87E-07	1.11	6.56E-08	1.12	1470
B*18:01	1.58E-01	0.96	1.80E-01	0.96	461
B*27:05	9.12E-01	1.00	8.51E-01	1.00	527
B*35:01	2.22E-06	0.89	4.08E-06	0.89	805
B*37:01	2.47E-01	0.95	2.01E-01	0.94	74
B*40:01	6.73E-01	0.99	7.56E-01	0.99	1078
B*44:02	2.89E-03	1.05	2.22E-03	1.05	4309
B*44:03	5.68E-02	1.04	9.97E-02	1.04	1140
B*49:01	7.44E-01	0.98	7.18E-01	0.98	47
B*51:01	3.20E-01	1.03	2.80E-01	1.03	488
B*55:01	2.12E-01	0.95	1.74E-01	0.95	139
B*57:01	1.31E-05	0.89	3.32E-05	0.89	546
C*01:02	9.63E-02	0.95	8.80E-02	0.95	391
C*02:02	1.39E-01	1.04	1.25E-01	1.04	425
C*03:03	9.66E-03	1.06	1.50E-02	1.06	1139
C*03:04	6.03E-02	1.03	4.32E-02	1.04	2271
C*04:01	5.39E-06	0.92	6.64E-06	0.91	2409
C*05:01	8.18E-02	1.03	6.45E-02	1.03	4419
C*06:02	9.28E-03	0.96	5.35E-03	0.95	2840
C*07:01	1.51E-06	1.06	4.57E-08	1.08	10602
C*07:02	4.81E-06	0.94	1.20E-05	0.93	8566
C*07:04	1.43E-01	1.06	1.81E-01	1.05	125
C*08:02	5.35E-03	0.93	2.41E-03	0.92	433
C*12:03	5.33E-01	0.98	4.60E-01	0.98	281
C*15:02	2.21E-01	1.05	2.51E-01	1.04	114
C*16:01	1.53E-02	1.06	1.85E-02	1.06	656
DQA1*01:01	1.89E-16	0.89	1.39E-16	0.87	7080

Table S2.3. Allele Associations: Additive vs. Dominant Model. (continued)

DQA1*01:02	1.23E-13	0.91	5.27E-12	0.90	12671
DQA1*01:03	1.45E-08	0.88	5.63E-09	0.87	1063
DQA1*02:01	9.54E-01	1.00	7.79E-01	1.00	7299
DQA1*03:01	6.91E-47	1.19	1.65E-44	1.22	14122
DQA1*04:01	4.46E-01	0.97	4.31E-01	0.97	143
DQA1*05:01	2.92E-03	1.04	2.76E-04	1.05	18256
DQB1*02:01	8.52E-05	1.06	2.53E-07	1.08	7750
DQB1*02:02	4.94E-02	1.04	1.46E-02	1.04	3633
DQB1*03:01	1.82E-09	1.08	1.29E-09	1.09	10575
DQB1*03:02	5.68E-25	1.17	6.18E-23	1.18	3606
DQB1*03:03	2.06E-01	0.97	3.02E-01	0.98	960
DQB1*04:02	2.05E-01	0.96	2.19E-01	0.96	147
DQB1*05:01	1.73E-13	0.89	1.81E-14	0.88	5140
DQB1*05:03	5.07E-03	0.91	5.75E-03	0.91	155
DQB1*06:02	1.21E-08	0.92	4.68E-08	0.92	7232
DQB1*06:03	5.10E-10	0.86	1.09E-10	0.85	936
DQB1*06:04	4.26E-05	0.88	7.07E-05	0.88	254
DQB1*06:09	3.22E-01	0.95	3.59E-01	0.95	42
DRB1*01:01	2.41E-11	0.89	2.91E-12	0.88	3212
DRB1*01:03	3.55E-01	0.96	3.53E-01	0.96	123
DRB1*03:01	1.33E-04	1.05	4.99E-07	1.08	7622
DRB1*04:01	1.03E-34	1.20	6.85E-38	1.23	4757
DRB1*04:04	1.01E-05	1.11	4.80E-05	1.11	588
DRB1*07:01	9.37E-01	1.00	9.03E-01	1.00	7301
DRB1*08:01	6.42E-01	0.98	6.28E-01	0.98	128
DRB1*09:01	3.00E-04	1.16	1.17E-04	1.17	71
DRB1*11:01	2.58E-02	1.06	1.51E-02	1.07	497
DRB1*11:04	6.64E-01	0.98	6.53E-01	0.98	49
DRB1*12:01	1.14E-04	0.84	2.81E-04	0.85	72
DRB1*13:01	1.69E-09	0.86	9.40E-10	0.86	895
DRB1*13:02	2.23E-05	0.89	3.51E-05	0.89	482
DRB1*14:01	1.22E-02	0.91	1.25E-02	0.91	137
DRB1*15:01	3.56E-09	0.92	5.10E-08	0.92	7358

Table S2.4. HLA Heterogeneity Test. The HLA alleles and amino acid polymorphisms that are heterogeneous are shown. For each variant, the childhood-onset (COA) and adult-onset asthma (AOA) p-value, odds ratio (OR) and 95% confidence interval (CI) are shown, along with the direction of the two variants and heterogeneity test results.

gene	variant	COA p-val	COA OR	COA 95% CI	AOA p-val	AOA OR	AOA 95% CI	dir	HetPVal
HLA ALLELE									
DQB1	*0602	2.33E-02	1.049	1.006-1.092	1.21E-08	0.920	0.893-0.946	+-	1.96E-08
DRB1	*1501	5.51E-02	1.041	0.999-1.084	3.56E-09	0.917	0.891-0.944	+-	3.62E-08
DQB1	*0302	4.65E-04	1.086	1.037-1.137	5.68E-25	1.174	1.138-1.210	++	2.19E-06
DQA1	*0102	1.84E-01	0.975	0.939-1.012	1.23E-13	0.908	0.885-0.931	--	2.11E-05
DQA1	*0301	7.26E-17	1.160	1.120-1.200	6.91E-47	1.187	1.159-1.215	++	3.70E-05
HLA AMINO ACID POLYMORPHISM									
DRB1	47_F	5.92E-03	1.042	1.011-1.073	8.92E-10	0.940	0.921-0.958	+-	3.86E-10
DQA1	34_E	2.96E-02	0.968	0.940-0.996	3.69E-09	1.061	1.040-1.081	-+	1.27E-08
DQB1	9_F	1.75E-01	1.027	0.988-1.068	3.44E-09	0.921	0.896-0.946	+-	3.15E-07
DRB1	-1_S	1.72E-01	1.029	0.987-1.071	4.99E-09	0.919	0.893-0.945	+-	3.87E-07
DRB1	71_A	1.72E-01	1.029	0.987-1.071	4.99E-09	0.919	0.893-0.945	+-	3.87E-07
DQB1	9_Y	1.82E-01	0.974	0.937-1.012	1.27E-08	1.081	1.052-1.110	-+	7.76E-07
DQB1	-5_L	9.79E-02	0.970	0.934-1.005	6.33E-18	0.895	0.873-0.918	--	1.11E-06
DRB1	96_Q	4.49E-01	1.015	0.976-1.056	8.87E-10	0.918	0.893-0.943	+-	1.31E-06
DQB1	185_I	8.92E-01	1.003	0.965-1.041	2.17E-12	1.094	1.066-1.121	++	1.43E-06
DRB1	11_P	3.08E-01	1.021	0.981-1.062	6.49E-09	0.921	0.895-0.947	+-	1.61E-06
DRB1	13_R	3.08E-01	1.021	0.981-1.062	6.49E-09	0.921	0.895-0.947	+-	1.61E-06
DRB1	133_L	3.08E-01	1.021	0.981-1.062	6.49E-09	0.921	0.895-0.947	+-	1.61E-06
DRB1	142_M	3.08E-01	1.021	0.981-1.062	6.49E-09	0.921	0.895-0.947	+-	1.61E-06
DQB1	97_F	9.26E-02	0.969	0.933-1.005	2.65E-17	0.896	0.873-0.919	--	2.19E-06
DQB1	185_T	5.99E-01	0.990	0.952-1.028	8.91E-13	0.912	0.889-0.935	--	3.64E-06
DQA1	187_T	2.32E-16	1.157	1.117-1.197	3.63E-47	1.187	1.160-1.215	++	2.09E-05
DQA1	26_S	2.32E-16	1.157	1.117-1.197	3.63E-47	1.187	1.160-1.215	++	2.09E-05
DQA1	47_Q	2.32E-16	1.157	1.117-1.197	3.63E-47	1.187	1.160-1.215	++	2.09E-05
DQA1	56_R	2.32E-16	1.157	1.117-1.197	3.63E-47	1.187	1.160-1.215	++	2.09E-05

Table S2.4. HLA Heterogeneity Test. (continued)

DQA1	76_V	2.32E-16	1.157	1.117-1.197	3.63E-47	1.187	1.160-1.215	++	2.09E-05
DQA1	207_M	1.84E-01	0.975	0.939-1.012	1.23E-13	0.908	0.885-0.931	--	2.11E-05
DQB1	125_G	5.72E-05	0.931	0.898-0.963	1.44E-22	0.888	0.867-0.909	--	6.71E-05
DQA1	175_E	3.93E-10	1.100	1.067-1.133	2.86E-33	1.131	1.108-1.153	++	7.47E-05
DRB1	180_V	2.25E-14	0.874	0.844-0.904	4.30E-41	0.853	0.833-0.872	--	7.51E-05
DRB1	-24_L	9.46E-14	0.879	0.849-0.909	4.48E-39	0.859	0.839-0.878	--	1.16E-04
DRB1	104_S	1.47E-12	0.896	0.869-0.923	7.50E-36	0.878	0.860-0.895	--	2.04E-04

Table S2.5. HLA Amino Acid Polymorphism Associations. Results for COA and AOA are shown for amino acids (freq > 1%); associations that are genome-wide significance are shown in bold font. For each amino acid, the locus is specified (A, B, C, DQA1, DQB1, DRB1), then the position along the protein. For biallelic positions, the reference then the alternative AA is listed. For multiallelic positions, only the amino acid at that position is noted. See File S2.1 for full table.

		Childhood-Onset Asthma				Adult-Onset Asthma			
HLA	**Amino Acid**	**p-value**	**OR**	**beta**	**SE**	**p-value**	**OR**	**beta**	**SE**
HLA-A	-22_V	9.08E-01	0.986	-0.014	0.125	7.62E-01	1.026	0.026	0.086
HLA-A	-20_A	8.48E-01	0.977	-0.024	0.124	7.98E-01	1.022	0.022	0.085
HLA-A	-15_L	2.77E-03	1.046	0.045	0.015	8.28E-01	0.998	-0.002	0.010
HLA-A	-15_V	2.34E-03	0.955	-0.046	0.015	7.95E-01	1.003	0.003	0.010
HLA-A	-11_L	2.64E-02	1.053	0.051	0.023	2.74E-02	1.035	0.035	0.016
HLA-A	-11_S	2.61E-02	0.950	-0.051	0.023	3.35E-02	0.967	-0.033	0.016
HLA-A	-2_W	9.54E-01	0.993	-0.007	0.120	8.66E-01	1.014	0.014	0.082
HLA-A	9_F	1.44E-01	0.977	-0.023	0.016	6.14E-01	0.995	-0.005	0.011
HLA-A	9_S	2.24E-01	1.029	0.028	0.023	5.67E-01	1.009	0.009	0.016
HLA-A	9_T	1.61E-02	1.067	0.065	0.027	1.63E-02	1.045	0.044	0.018
HLA-A	9_Y	3.73E-01	0.981	-0.019	0.021	9.16E-02	0.976	-0.024	0.014
HLA-A	17_R_S	1.13E-02	1.133	0.125	0.049	2.84E-01	0.963	-0.038	0.036
HLA-A	44_R_K	4.94E-02	1.037	0.036	0.019	9.33E-03	1.033	0.033	0.013
HLA-A	56_G	5.07E-02	0.937	-0.065	0.033	3.17E-02	0.953	-0.048	0.023
HLA-A	56_R	3.86E-02	1.073	0.070	0.034	3.31E-02	1.050	0.049	0.023
HLA-A	62_E	9.32E-01	1.002	0.002	0.026	2.39E-01	1.021	0.020	0.017
HLA-A	62_G	6.31E-03	0.956	-0.045	0.017	6.51E-01	0.995	-0.005	0.011
HLA-A	62_L	8.73E-04	1.122	0.116	0.035	4.63E-01	1.018	0.018	0.024
HLA-A	62_Q	6.07E-02	1.028	0.028	0.015	7.35E-01	0.997	-0.003	0.010
HLA-A	62_R	1.56E-01	0.962	-0.039	0.027	6.54E-01	0.992	-0.008	0.018
HLA-A	63_E	4.24E-01	0.983	-0.017	0.022	9.14E-01	0.998	-0.002	0.015
HLA-A	63_N	1.56E-01	0.962	-0.039	0.027	6.54E-01	0.992	-0.008	0.018

Table S2.6. Putatively Causal Variants and Allergy. Each row is a result for each variant reflecting each credible set (CS). The original p-value and odds ratio (OR) for the risk allele are shown, then the p-value and OR when allergy was included as a covariate in the regression ("Allergy Cov"), when all individuals with allergy were excluded ("No allergy"), and the interaction between the variant and allergy status are shown.

CS	Var	Original p-value	Original OR	Allergy Cov p-value	Allergy Cov OR	No allergy p-value	No allergy OR	Interaction p-value
Childhood-Onset Asthma								
Class I CS1	rs2428494	8.77×10^{-23}	1.157	4.97×10^{-21}	1.151	6.11×10^{-16}	1.160	0.494
Class I CS2	HLA-C p.11	3.12x10-^{19}	1.241	2.12×10^{-18}	1.236	7.84×10^{-12}	1.224	0.562
Class II CS1	rs28407950	1.37x10-^{59}	1.355	1.31×10^{-54}	1.339	7.74×10^{-41}	1.361	0.219
Class II CS2	rs35571244	1.07×10^{-17}	1.253	3.21×10^{-16}	1.242	2.78×10^{-11}	1.243	0.9995
Adult-Onset Asthma								
Class I CS1	rs2428494	4.52×10^{-23}	1.104	3.13×10^{-21}	1.100	1.66×10^{-16}	1.099	0.939
Class II CS1	rs9272346	1.98×10^{-47}	1.163	1.42×10^{-47}	1.164	1.82×10^{-37}	1.166	0.77
Class II CS2	DQA1*0301	6.91×10^{-47}	1.187	1.72×10^{-48}	1.202	1.62×10^{-40}	1.211	0.352

Table S2.7. Putatively Causal Variants and Sex. For each variant, the p-value and ORs for the risk allele are shown for the original analysis ("Original"), in just female participants, and in just males. The p-value for the interaction between sex and the variant is also shown.

CS	Var	Original p-value	Original OR	Females p-value	Females OR	Males p-value	Males OR	Interaction p-value
Childhood-Onset Asthma								
Class I CS1	rs2428494	$8.77x10^{-23}$	1.157	$3.46x10^{-09}$	1.146	$3.71x10^{-15}$	1.164	0.613
Class I CS2	HLA-C p.11	$3.12x10^{-19}$	1.241	$4.59x10^{-12}$	1.302	$3.68x10^{-09}$	1.200	0.098
Class II CS1	rs28407950	$1.37x10^{-59}$	1.355	$2.87x10^{-33}$	1.426	$6.08x10^{-29}$	1.308	0.024
Class II CS2	rs35571244	1.07x10-^{17}	1.253	$3.85x10^{-08}$	1.253	$4.94x10^{-11}$	1.253	0.996
Adult-Onset Asthma								
Class I CS1	rs2428494	$4.52x10^{-23}$	1.104	$6.63x10^{-14}$	1.099	$9.69x10^{-11}$	1.112	0.575
Class II CS1	rs9272346	$1.98x10^{-47}$	1.163	$3.02x10^{-31}$	1.165	$7.18x10^{-18}$	1.159	0.812
Class II CS2	DQA1*0301	$6.91x10^{-47}$	1.187	$1.45x10^{-34}$	1.202	$2.32x10^{-14}$	1.162	0.174

Table S2.8. SuSiE Credible Set Results. Median r2 reported between all variants within the CS. Posterior Inclusion Probability (PIP) reported for each variant within the CS. P-value, odds ratio (OR) and confidence interval (CI) reported for the variant's association for that specific group (childhood-onset asthma or adult-onset asthma).

CS	Median r2	Var	PIP	p-val	OR	CI	Freq	Ref	Alt
				Class I Child					
CS1	1	rs2428494	0.969	8.77E-23	1.157	1.123-1.191	0.474	T	A
CS2	1.00	HLA-C p.11	0.573	3.12E-19	0.806	0.768-0.844	0.127	A	S
		rs28481932	0.426	4.36E-19	0.807	0.769-0.845	0.127	A	G
				Class I Adult					
CS1	1	rs2428494	0.999	4.52E-23	1.104	1.082-1.125	0.475	T	A
				Class II Child					
CS1	1	rs28407950	0.997	1.37E-59	0.738	0.712-0.765	0.246	C	T
CS2	0.99	rs35571244	0.495	1.07E-17	1.253	1.189-1.319	0.072	T	C
		rs35599935	0.201	3.39E-17	1.251	1.187-1.318	0.071	T	C
		rs34975158	0.171	4.73E-17	1.249	1.185-1.315	0.071	G	A
		rs4148878	0.056	1.60E-16	1.243	1.180-1.308	0.071	T	G
		rs33998906	0.029	1.42E-15	1.242	1.177-1.309	0.067	C	T
				Class II Adult					
CS1	0.99	rs17843580	0.078	4.01E-49	1.167	1.143-1.190	0.601	A	G
		HLA-DQB1 p.55	0.039	4.50E-49	0.858	0.841-0.876	0.409	L,P	R
		HLA-DQA1 p.11	0.030	3.23E-48	0.859	0.841-0.876	0.390	Y	C
		HLA-DQA1 p.18	0.030	3.23E-48	0.859	0.841-0.876	0.390	S	F
		HLA-DQA1 p.45	0.030	3.23E-48	0.859	0.841-0.876	0.390	V	A
		HLA-DQA1 p.47	0.030	3.23E-48	0.859	0.841-0.876	0.390	C,K,Q	R
		HLA-DQA1 p.48	0.030	3.23E-48	0.859	0.841-0.876	0.390	L	W
		HLA-DQA1 p.50	0.030	3.23E-48	0.859	0.841-0.876	0.390	L,V	E
		HLA-DQA1 p.52	0.030	3.23E-48	0.859	0.841-0.876	0.390	H,R	S
		HLA-DQA1 p.53	0.030	3.23E-48	0.859	0.841-0.876	0.390	Q,R	K

Table S2.8. SuSiE Credible Set Results. (continued)

HLA-DQA1 p.55	0.030	3.23E-48	0.859	0.841-0.876	0.390	R	G
HLA-DQA1 p.56	0.030	3.23E-48	0.859	0.841-0.876	0.390	R,x	G
HLA-DQA1 p.61	0.030	3.23E-48	0.859	0.841-0.876	0.390	F	G
HLA-DQA1 p.64	0.030	3.23E-48	0.859	0.841-0.876	0.390	T	R
HLA-DQA1 p.66	0.030	3.23E-48	0.859	0.841-0.876	0.390	I	M
HLA-DQA1 p.69	0.030	3.23E-48	0.859	0.841-0.876	0.390	L,T	A
HLA-DQA1 p.76	0.030	3.23E-48	0.859	0.841-0.876	0.390	L,V	M
HLA-DQA1 p.80	0.030	3.23E-48	0.859	0.841-0.876	0.390	S	Y
HLA-DQA1 p.175	0.030	3.23E-48	0.859	0.841-0.876	0.390	E,K	Q
HLA-DQA1 p.218	0.030	3.23E-48	0.859	0.841-0.876	0.390	R	Q
rs17612788	0.023	9.67E-49	1.165	1.141-1.189	0.605	G	A
rs9273084	0.015	5.25E-48	1.164	1.140-1.187	0.607	T	C
HLA-DQB1 p.203	0.014	2.30E-47	0.859	0.841-0.876	0.383	I,x	V
rs9274660	0.011	7.88E-48	1.163	1.139-1.187	0.608	A	G
rs34843907	0.010	1.51E-47	1.163	1.139-1.187	0.604	T	G
rs17843619	0.010	1.11E-47	1.163	1.139-1.186	0.607	C	A
rs9272629	0.009	2.34E-46	1.161	1.137-1.185	0.600	G	T
rs1130034	0.009	1.27E-47	1.163	1.139-1.187	0.608	T	C
rs9273497	0.009	6.01E-47	1.163	1.139-1.187	0.601	T	C
rs17612576	0.009	1.16E-47	1.163	1.139-1.186	0.608	G	A
HLA-DQB1 p.84	0.008	3.66E-47	0.860	0.842-0.877	0.387	Q,x	E
HLA-DQB1 p.85	0.008	3.66E-47	0.860	0.842-0.877	0.387	L,x	V
HLA-DQB1 p.89	0.008	3.66E-47	0.860	0.842-0.877	0.387	T,x	G
HLA-DQB1 p.90	0.008	3.66E-47	0.860	0.842-0.877	0.387	T,x	I
HLA-DQB1 p.220	0.008	3.75E-47	0.860	0.842-0.877	0.387	H,x	R
HLA-DQB1 p.221	0.008	3.75E-47	0.860	0.842-0.877	0.387	H,x	Q
rs9273088	0.008	1.59E-47	1.163	1.139-1.186	0.608	A	C
rs3828789	0.008	1.36E-47	1.163	1.139-1.186	0.608	G	T

Table S2.8. SuSiE Credible Set Results. (continued)

		rs1140343	0.007	6.88E-48	1.164	1.140-1.188	0.603	T	G
		rs1063349	0.007	1.49E-47	1.163	1.139-1.186	0.608	T	C
		rs17843577	0.007	3.38E-46	1.160	1.136-1.184	0.600	C	T
		rs9272346	0.007	1.98E-47	1.163	1.139-1.186	0.608	G	A
		rs9273215	0.006	1.67E-47	1.163	1.139-1.186	0.608	G	A
		rs9273329	0.006	1.66E-47	1.163	1.139-1.186	0.608	G	A
		rs9273326	0.006	1.70E-47	1.163	1.139-1.186	0.608	T	C
		rs17612781	0.006	2.06E-47	1.163	1.139-1.186	0.608	T	C
		rs9272625	0.006	3.49E-46	1.161	1.137-1.185	0.599	T	C
		rs17612625	0.006	1.77E-47	1.163	1.139-1.186	0.608	A	C
		rs17612633	0.006	1.80E-47	1.163	1.139-1.186	0.608	G	C
		rs17843573	0.006	1.89E-47	1.163	1.139-1.186	0.608	C	T
		rs9273493	0.006	1.85E-47	1.162	1.139-1.186	0.608	T	C
		rs1063355	0.006	1.90E-47	1.162	1.139-1.186	0.608	T	G
		rs9273330	0.006	3.83E-47	1.162	1.138-1.186	0.605	C	T
		rs17612858	0.006	1.89E-47	1.163	1.139-1.186	0.608	A	T
		rs9273524	0.005	2.14E-47	1.162	1.138-1.186	0.608	T	C
		rs9273339	0.005	2.64E-47	1.162	1.138-1.186	0.608	G	A
		rs17612928	0.004	9.47E-47	1.161	1.137-1.185	0.602	A	G
		rs1063348	0.003	2.42E-47	1.162	1.138-1.186	0.605	A	G
		rs17612802	0.003	8.36E-48	1.164	1.140-1.188	0.602	C	T
		HLA-DQA1 p.-16	0.003	6.05E-48	0.861	0.843-0.878	0.412	M	L
CS2	0.96	rs41269945	0.193	5.08E-47	1.197	1.168-1.227	0.188	A	T
		rs41269955	0.112	1.93E-45	1.207	1.176-1.239	0.169	G	A
		HLA-DQA1 p.26	0.066	3.63E-47	1.187	1.160-1.215	0.204	T	S
		HLA-DQA1 p.47	0.066	3.63E-47	1.187	1.160-1.215	0.204	C,K,R	Q
		HLA-DQA1 p.56	0.066	3.63E-47	1.187	1.160-1.215	0.204	G,x	R
		HLA-DQA1 p.76	0.066	3.63E-47	1.187	1.160-1.215	0.204	L,M	V

Table S2.8. SuSiE Credible Set Results. (continued)

HLA-DQA1 p.187	0.066	3.63E-47	1.187	1.160-1.215	0.204	A	T
HLA-DQA1*0301	0.057	6.91E-47	1.187	1.159-1.215	0.202	-	-
rs34141382	0.056	1.48E-45	1.200	1.170-1.231	0.179	T	C
rs1391371	0.051	5.75E-47	1.187	1.159-1.214	0.204	A	T
rs9272461	0.028	1.84E-46	1.187	1.159-1.214	0.203	G	A
rs34763586	0.016	6.01E-45	1.194	1.165-1.224	0.185	T	C
rs7760841	0.014	3.69E-44	1.196	1.166-1.226	0.176	C	T
rs17426593	0.013	2.67E-45	1.191	1.162-1.220	0.193	T	C
rs35117964	0.008	8.13E-43	1.187	1.158-1.216	0.190	A	G
rs3104413	0.007	3.07E-44	1.184	1.156-1.212	0.192	C	G
rs34578704	0.007	7.07E-43	1.204	1.172-1.236	0.161	C	A
rs34415150	0.006	9.06E-44	1.188	1.159-1.216	0.184	A	G
rs9272785	0.005	2.92E-44	1.189	1.160-1.218	0.192	G	A
rs35265698	0.005	8.55E-44	1.184	1.155-1.212	0.192	C	G
rs35371668	0.005	2.51E-43	1.190	1.160-1.219	0.180	C	T
rs34039593	0.005	8.50E-44	1.184	1.155-1.212	0.191	T	G
rs35294087	0.004	9.99E-44	1.183	1.155-1.211	0.192	A	G
rs504594	0.004	8.78E-44	1.184	1.155-1.212	0.191	C	A
rs9271608	0.004	8.79E-44	1.184	1.155-1.212	0.191	A	G
rs3129751	0.004	9.40E-44	1.184	1.155-1.212	0.191	A	C
rs35118762	0.004	1.29E-43	1.183	1.155-1.211	0.192	C	T
HLA-DRB1 p.-24	0.004	2.98E-43	1.194	1.164-1.224	0.178	L,x	F
HLA-DRB1 p.33	0.004	2.98E-43	1.194	1.164-1.224	0.178	N	H
HLA-DRB1 p.96	0.004	2.98E-43	1.194	1.164-1.224	0.178	E,H,Q,x	Y
HLA-DRB1 p.180	0.004	2.98E-43	1.194	1.164-1.224	0.178	V,x	L
HLA-DRB1 p.13	0.004	3.00E-43	1.194	1.164-1.224	0.178	F,G	H
rs3997872	0.004	1.11E-43	1.183	1.155-1.211	0.191	T	A

Table S2.9. HLA Allele Frequencies by Study. Sample composition from RNAseq data collected from LCLs (Hutterites) and the PBMCs and NECs from URECA (URban Environment and Childhood Asthma). NA, not available.

Locus	Allele	Hutterites	URECA
HLA-A	01:01	0.179	0.045
HLA-A	02:01	0.256	0.130
HLA-A	02:02	NA	0.030
HLA-A	02:04	NA	0.003
HLA-A	02:05	NA	0.024
HLA-A	02:11	NA	0.003
HLA-A	02:17	NA	0.003
HLA-A	02:31	NA	0.003
HLA-A	02:60	NA	0.003
HLA-A	03:01	0.138	0.099
HLA-A	11:01	0.013	0.009
HLA-A	23:01	NA	0.139
HLA-A	24:02	0.207	0.024
HLA-A	25:01	NA	0.003
HLA-A	26:01	0.073	0.006
HLA-A	29:01	NA	0.003
HLA-A	29:02	NA	0.045
HLA-A	30:01	0.022	0.042
HLA-A	30:02	NA	0.066
HLA-A	30:04	NA	0.003
HLA-A	31:01	0.028	0.009
HLA-A	32:01	0.069	0.009
HLA-A	33:01	0.003	0.030
HLA-A	33:03	NA	0.039
HLA-A	34:01	0.012	NA
HLA-A	34:02	NA	0.030
HLA-A	36:01	NA	0.015
HLA-A	66:01	NA	0.003
HLA-A	66:02	NA	0.006
HLA-A	66:03	NA	0.003
HLA-A	68:01	NA	0.039
HLA-A	68:02	NA	0.069
HLA-A	68:03	NA	0.006
HLA-A	68:15	NA	0.003
HLA-A	74:01	NA	0.051
HLA-A	80:01	NA	0.003
HLA-B	07:02	0.023	0.087
HLA-B	07:05	NA	0.009
HLA-B	08:01	0.087	0.054
HLA-B	13:02	0.075	0.015

Table S2.9. HLA Allele Frequencies by Study. (continued)

HLA-B	14:01	NA	0.015
HLA-B	14:02	NA	0.018
HLA-B	14:03	NA	0.003
HLA-B	15:01	0.034	NA
HLA-B	15:03	NA	0.078
HLA-B	15:10	NA	0.039
HLA-B	15:16	NA	0.003
HLA-B	15:17	NA	0.003
HLA-B	15:18	NA	0.003
HLA-B	15:39	0.024	NA
HLA-B	15:54	NA	0.003
HLA-B	18:01	0.024	0.045
HLA-B	27:03	NA	0.003
HLA-B	27:05	0.002	0.009
HLA-B	27:08	0.074	NA
HLA-B	35:01	0.083	0.057
HLA-B	35:03	0.008	NA
HLA-B	35:05	NA	0.003
HLA-B	35:08	0.067	NA
HLA-B	35:12	NA	0.003
HLA-B	37:01	NA	0.015
HLA-B	38:01	0.042	0.003
HLA-B	39:08	NA	0.003
HLA-B	39:10	NA	0.006
HLA-B	40:01	0.078	0.015
HLA-B	40:02	NA	0.009
HLA-B	40:16	NA	0.003
HLA-B	41:01	NA	0.003
HLA-B	41:02	NA	0.003
HLA-B	41:03	NA	0.003
HLA-B	42:01	NA	0.060
HLA-B	42:02	NA	0.009
HLA-B	44:02	0.013	0.015
HLA-B	44:03	NA	0.054
HLA-B	45:01	NA	0.033
HLA-B	48:02	NA	0.003
HLA-B	49:01	NA	0.033
HLA-B	50:01	0.003	NA
HLA-B	51:01	0.183	0.024
HLA-B	52:01	NA	0.015
HLA-B	53:01	NA	0.099
HLA-B	56:01	0.016	NA
HLA-B	57:01	0.087	0.003

Table S2.9. HLA Allele Frequencies by Study. (continued)

HLA-B	57:02	NA	0.006
HLA-B	57:03	NA	0.042
HLA-B	57:04	NA	0.003
HLA-B	58:01	NA	0.021
HLA-B	58:02	NA	0.033
HLA-B	78:01	NA	0.018
HLA-B	81:01	NA	0.012
HLA-C	01:02	0.071	0.012
HLA-C	02:02	0.053	0.009
HLA-C	02:10	NA	0.081
HLA-C	03:02	0.003	0.012
HLA-C	03:03	0.017	0.003
HLA-C	03:04	0.110	0.057
HLA-C	03:05	NA	0.006
HLA-C	04:01	0.127	0.157
HLA-C	05:01	0.057	0.030
HLA-C	06:02	0.164	0.075
HLA-C	07:01	0.073	0.160
HLA-C	07:02	0.039	0.072
HLA-C	07:04	NA	0.006
HLA-C	08:02	NA	0.039
HLA-C	08:04	NA	0.012
HLA-C	12:03	0.147	0.015
HLA-C	14:02	0.069	0.006
HLA-C	14:03	NA	0.003
HLA-C	15:02	NA	0.012
HLA-C	15:05	NA	0.030
HLA-C	16:01	0.031	0.084
HLA-C	16:02	NA	0.003
HLA-C	17:01	NA	0.075
HLA-C	18:01	NA	0.036
HLA-C	18:04	NA	0.003
HLA-DQA1	01:01	0.220	0.108
HLA-DQA1	01:02	0.109	0.259
HLA-DQA1	01:03	0.061	0.036
HLA-DQA1	02:01	0.128	0.133
HLA-DQA1	03:01	0.197	0.136
HLA-DQA1	04:01	0.056	0.093
HLA-DQA1	05:01	0.229	0.235
HLA-DQB1	02:01	0.176	0.253
HLA-DQB1	03:01	0.190	0.178
HLA-DQB1	03:02	0.063	0.066
HLA-DQB1	03:03	0.089	0.024

Table S2.9. HLA Allele Frequencies by Study. (continued)

HLA-DQB1	03:132	NA	0.003
HLA-DQB1	04:02	0.083	0.069
HLA-DQB1	05:01	0.218	0.111
HLA-DQB1	05:02	0.013	0.030
HLA-DQB1	05:03	NA	0.018
HLA-DQB1	06:02	0.096	0.160
HLA-DQB1	06:03	0.061	0.018
HLA-DQB1	06:04	NA	0.030
HLA-DQB1	06:08	NA	0.003
HLA-DQB1	06:09	NA	0.030
HLA-DRB1	01:01	0.145	0.018
HLA-DRB1	01:02	NA	0.027
HLA-DRB1	01:03	NA	0.003
HLA-DRB1	03:01	0.117	0.087
HLA-DRB1	03:02	NA	0.054
HLA-DRB1	04:01	0.115	0.015
HLA-DRB1	04:02	NA	0.003
HLA-DRB1	04:03	0.002	0.003
HLA-DRB1	04:04	0.044	0.012
HLA-DRB1	04:05	0.019	0.021
HLA-DRB1	04:07	NA	0.012
HLA-DRB1	04:08	NA	0.003
HLA-DRB1	04:10	NA	0.003
HLA-DRB1	04:11	NA	0.009
HLA-DRB1	07:01	0.128	0.117
HLA-DRB1	07:11	NA	0.003
HLA-DRB1	08:01	NA	0.003
HLA-DRB1	08:02	NA	0.003
HLA-DRB1	08:03	0.052	NA
HLA-DRB1	08:04	NA	0.054
HLA-DRB1	08:06	NA	0.006
HLA-DRB1	09:01	NA	0.036
HLA-DRB1	10:01	0.066	0.018
HLA-DRB1	11:01	0.111	0.075
HLA-DRB1	11:02	NA	0.024
HLA-DRB1	11:04	NA	0.006
HLA-DRB1	11:08	0.001	NA
HLA-DRB1	12:01	NA	0.030
HLA-DRB1	13:01	0.061	0.048
HLA-DRB1	13:02	NA	0.063
HLA-DRB1	13:03	NA	0.045
HLA-DRB1	13:04	NA	0.012
HLA-DRB1	13:31	NA	0.003

Table S2.9. HLA Allele Frequencies by Study. (continued)

HLA-DRB1	14:01	NA	0.021
HLA-DRB1	15:01	0.094	0.021
HLA-DRB1	15:03	NA	0.120
HLA-DRB1	15:13	NA	0.003
HLA-DRB1	16:01	0.013	NA
HLA-DRB1	16:02	NA	0.012
HLA-DRB1	16:34	NA	0.003

Table S2.10. List of SNPs in the Credible Sets Excluded from eQTL Analyses. SNPs noted if there was high missingness or if it failed QC for other reasons (e.g. failed HWE).

Group	CS	rsid	Hutterites (LCL)	URECA (PBMC and NEC)
Class I Child	CS2	rs28481932	high missingness	
Class II Child	CS1	rs28407950	high missingness	
	CS2	rs35571244		failed QC
Class II Adult	CS1	rs17843580	high missingness	
		rs17612788	high missingness	
		rs9273084	high missingness	
		rs34843907		failed QC
		rs17843619	high missingness	
		rs9272629	high missingness	failed QC
		rs1130034	failed QC	failed QC
		rs9273497	high missingness	
		rs17612576	high missingness	
		rs9273088	high missingness	
		rs3828789	high missingness	
		rs1140343	high missingness	failed QC
		rs1063349	high missingness	failed QC
		rs17843577	high missingness	
		rs9273215	failed QC	failed QC
		rs9273329	high missingness	
		rs9273326	high missingness	
		rs17612781	high missingness	
		rs9272625	high missingness	failed QC
		rs17612625	high missingness	
		rs17612633	high missingness	
		rs17843573	high missingness	
		rs9273493	high missingness	
		rs1063355	high missingness	
		rs9273330	high missingness	
		rs17612858	high missingness	
		rs9273524	high missingness	
		rs9273339	high missingness	
		rs17612928	failed QC	failed QC
		rs1063348	high missingness	failed QC
		rs17612802	high missingness	
	CS2	rs41269945	high missingness	failed QC
		rs41269955	failed QC	failed QC
		rs34141382	high missingness	failed QC

Table S2.10. List of SNPs in the Credible Sets Excluded from eQTL Analyses. (continued)

rs9272461	high missingness	failed QC
rs34763586	high missingness	failed QC
rs7760841	failed QC	
rs17426593	high missingness	failed QC
rs35117964	high missingness	failed QC
rs3104413	high missingness	
rs34578704	failed QC	failed QC
rs34415150	high missingness	
rs9272785	failed QC	failed QC
rs35265698	high missingness	
rs34039593	high missingness	
rs35294087	high missingness	
rs504594	failed QC	
rs9271608	high missingness	
rs35118762	high missingness	
rs3997872	high missingness	

Table S2.11. eQTL Results for All Credible Set SNPs. P-values are shown for each SNP/gene that have an FDR<0.05. ɣgenotypes in subset of individuals from MEGA array. *genotypes from subset of individuals (see Methods in section 2.3). See File S2.2 for full table.

CS	SNP	Cell	gene	p
Class I COA AOA CS1	rs2428494	LCL	*HLA-B*	1.10E-03
Class I COA AOA CS1	rs2428494	LCL	*MIR6891*	1.66E-08
Class I COA CS2	rs28481932*	LCL	*CCHCR1*	1.83E-06
Class I COA CS2	rs28481932*	LCL	*AL662844.4*	3.06E-05
Class II COA CS1	rs28407950*	LCL	*HLA-DQA2*	2.16E-17
Class II COA CS1	rs28407950*	LCL	*HLA-DQB2*	5.15E-17
Class II COA CS1	rs28407950*	LCL	*HLA-DRB5*	6.75E-06
Class II COA CS1	rs28407950*	LCL	*HLA-DQB1-AS1*	5.19E-05
Class II AOA CS1	rs34843907	LCL	*HLA-DQA2*	1.51E-50
Class II AOA CS1	rs9272346	LCL	*HLA-DQA2*	4.05E-54
Class II AOA CS1	rs9274660	LCL	*HLA-DQA2*	7.86E-50
Class II AOA CS1	rs34843907	LCL	*HLA-DQB2*	2.31E-66
Class II AOA CS1	rs9272346	LCL	*HLA-DQB2*	3.26E-73
Class II AOA CS1	rs9274660	LCL	*HLA-DQB2*	5.88E-59
Class II AOA CS1	rs34843907	LCL	*HLA-DRB6*	7.54E-20
Class II AOA CS1	rs9272346	LCL	*HLA-DRB6*	1.23E-21
Class II AOA CS1	rs9274660	LCL	*HLA-DRB6*	3.11E-19
Class II AOA CS1	rs34843907	LCL	*HLA-DRB9*	4.57E-06
Class II AOA CS1	rs9272346	LCL	*HLA-DRB9*	1.67E-06
Class II AOA CS1	rs9274660	LCL	*HLA-DRB9*	1.41E-06
Class II AOA CS1	rs34843907	LCL	*TAP2*	4.57E-05
Class II AOA CS1	rs9272346	LCL	*TAP2*	1.98E-04
Class II AOA CS1	rs9274660	LCL	*TAP2*	1.35E-05
Class II AOA CS2	rs1391371	LCL	*HLA-DQA2*	6.16E-26
Class II AOA CS2	rs3129751	LCL	*HLA-DQA2*	9.92E-20
Class II AOA CS2	rs35371668	LCL	*HLA-DQA2*	6.43E-25
Class II AOA CS2	rs1391371	LCL	*HLA-DQB2*	1.68E-12
Class II AOA CS2	rs3129751	LCL	*HLA-DQB2*	2.35E-10
Class II AOA CS2	rs35371668	LCL	*HLA-DQB2*	4.05E-12
Class II AOA CS2	rs1391371	LCL	*HLA-DRB5*	1.16E-10
Class II AOA CS2	rs3129751	LCL	*HLA-DRB5*	2.09E-11
Class II AOA CS2	rs35371668	LCL	*HLA-DRB5*	2.13E-11
Class II AOA CS2	rs1391371	LCL	*HLA-DRB6*	7.81E-10
Class II AOA CS2	rs3129751	LCL	*HLA-DRB6*	5.60E-08
Class II AOA CS2	rs35371668	LCL	*HLA-DRB6*	1.05E-09
Class II AOA CS2	rs1391371	LCL	*PPT2*	1.08E-05

Table S2.12. HLA Region eQTLs. For all SNPs in each credible set (CS) and each gene within 1 Mb of the transcription start site (TSS) examined across the three datasets (lymphoblastoid cell lines [LCLs], peripheral blood mononuclear cells [PBMCs], and nasal epithelial cells [NECs], the genes with eQTLs at FDR < 0.05 are shown. n.i., no information for the SNP (Table S6). These genes were then included in the eQTL fine-mapping studies. * analyses performed in a subset of individuals who had genotypes for all SNPs in the CS. ¥ analyses performed in a subset of individuals with genotypes from MEGA array

CS	LCLs (n=398, European American)	PBMCs (n=133, African American)	NECs (n=189, African American)
Class I COA CS1/AOA CS1	*HLA-B* *MIR6891*	No eQTLs	*HLA-B* ¥
Class I COA CS2	*CCHCR1* * *AL662844.4* *	No eQTLs	No eQTLs
Class II COA CS1	*HLA-DQA2* * *HLA-DQB2* * *HLA-DRB5* * *HLA-DQB1-AS1* *	*HLA-DQA2* *HLA-DQB2*	HLA-DQB2
Class II COA CS2	*HLA-DQA2* *HLA-DPB2* *HLA-DRB9* *PSMB9* *TAP2*	No eQTLs	No eQTLs
Class II AOA CS1	*HLA-DQA2* *HLA-DQB2* *HLA-DRB6* *HLA-DRB9* *TAP2*	*HLA-DQA2* *HLA-DQB2* *HLA-DRB6*	*HLA-DQB1* *HLA-DQA2* *HLA-DQB2* *HLA-DRB6*
Class II AOA CS2	*HLA-DQA2* *HLA-DQB2* *HLA-DRB5* *HLA-DRB6* *PPT2* *TAP1*	*HLA-DQA2*	*HLA-DQA1* *HLA-DQA2*

Table S2.13. eQTL Fine-Mapping Results. SNPs in adult-onset asthma (AOA) class II CS1 that overlap with class II eQTL credible sets are shown with their PIPs. SNPs are noted if they are shared with the class II AOA CS1. See File S2.3 for full table.

rsid	PIP	in AOA CS
LCL: HLA-DQB2		
rs9272346	0.3411	yes
rs3104367	0.3411	
rs34843907	0.1064	yes
rs9272634	0.1064	
rs9272756	0.1064	
LCL: HLA-DQA2		
rs9272346	0.359057	yes
rs3104367	0.359057	
rs34843907	0.092248	yes
rs9272634	0.092248	
rs9272756	0.092248	
PBMC: HLA-DQB2		
rs3134993	0.0154	
rs17612858	0.0125	yes
rs9273046	0.0116	
rs9272346	0.0084	yes
rs9273084	0.0084	yes
rs17843573	0.0084	yes
rs17612576	0.0084	yes
rs17843577	0.0084	yes
rs17843580	0.0084	yes
rs17612625	0.0084	yes
rs17612633	0.0084	yes
rs17612781	0.0084	yes
rs17612788	0.0084	yes
rs17612802	0.0084	yes
rs9273326	0.0084	yes
rs9273329	0.0084	yes
rs9273330	0.0084	yes
rs3828789	0.0084	yes
rs9274660	0.0084	yes

Table S2.14. Amino Acid Associations with the Highest PIPs in Each Credible Set. For each of the putatively causal amino acid polymorphisms in the credible sets (CSs), the position (Pos), reference (Ref) or alternative (Alt) amino acid, frequency, p-value, odds ratio (OR), 95% confidence interval (CI), posterior inclusion probability (PIP), and secondary structure are described. A dash ("-") indicates that the variant is unlikely to reside in a functional domain.

CS	HLA Locus	Pos	Ref	Alt	Frequency	p-value	OR	95% CI	PIP	Secondary Structure
Class I COA CS2	HLA-C	11	Ala	Ser	0.13	3.12×10^{-19}	0.806	0.768-0.844	0.573	Beta Strand
Class II AOA CS1	HLA-DQB1	55	Lys, Pro	Arg	0.41	4.50×10^{-49}	0.858	0.841-0.876	0.039	Alpha Helix
Class II AOA CS2	HLA-DQA1	26	Thr	Ser	0.20	3.63×10^{-47}	1.187	1.160-1.215	0.066	Beta Strand
		47	Cys	Gln	0.20	3.63×10^{-47}	1.187	1.160-1.215	0.066	-
		56	Gly,x	Arg	0.20	3.63×10^{-47}	1.187	1.160-1.215	0.066	-
		76	Ile	Val	0.20	3.63×10^{-47}	1.187	1.160-1.215	0.066	Alpha Helix
		187	Ala	Thr	0.20	3.63×10^{-47}	1.187	1.160-1.215	0.066	-

Table S2.15. Amino Acids in the Credible Sets and Their Corresponding HLA Alleles. Alleles include both common and rare HLA Alleles.

HLA region	CS	HLA Locus	Position	OR	Ref	Alt	HLA alleles with the alternate amino acid polymorphism
Class I	COA CS2	HLA-C	p.11	0.806	A	S	0102, 0401, 0407, 1402, 1403
Class II	AOA CS1	HLA-DQB1	p.55	0.858	L,P	R	0501, 0502, 0503, 0504, 0601, 0602, 0603, 0604, 0605, 0609, 0401, 0402
		HLA-DQA1	p.11	0.859	Y	C	0101, 0102, 0103, 0104
		HLA-DQA1	p.18	0.859	S	F	0101, 0102, 0103, 0104
		HLA-DQA1	p.45	0.859	V	A	0101, 0102, 0103, 0104
		HLA-DQA1	p.47	0.859	C,K,Q	R	0101, 0102, 0103, 0104
		HLA-DQA1	p.48	0.859	L	W	0101, 0102, 0103, 0104
		HLA-DQA1	p.50	0.859	L,V	E	0101, 0102, 0103, 0104
		HLA-DQA1	p.52	0.859	H,R	S	0101, 0102, 0103, 0104
		HLA-DQA1	p.53	0.859	Q,R	K	0101, 0102, 0103, 0104
		HLA-DQA1	p.55	0.859	R	G	0101, 0102, 0103, 0104
		HLA-DQA1	p.56	0.859	R,x	G	0101, 0102, 0103, 0104
		HLA-DQA1	p.61	0.859	F	G	0101, 0102, 0103, 0104
		HLA-DQA1	p.64	0.859	T	R	0101, 0102, 0103, 0104
		HLA-DQA1	p.66	0.859	I	M	0101, 0102, 0103, 0104
		HLA-DQA1	p.69	0.859	L,T	A	0101, 0102, 0103, 0104
		HLA-DQA1	p.76	0.859	L,V	M	0101, 0102, 0103, 0104
		HLA-DQA1	p.80	0.859	S	Y	0101, 0102, 0103, 0104
		HLA-DQA1	p.175	0.859	E,K	Q	0101, 0102, 0103, 0104
		HLA-DQA1	p.218	0.859	R	Q	0101, 0102, 0103, 0104
		HLA-DQB1	p.203	0.859	I,x	V	0501, 0502, 0503, 0602, 0603, 0604, 0609
		HLA-DQB1	p.84	0.860	Q,x	E	0501, 0502, 0503, 0601, 0602, 0603, 0604, 0605, 0609
		HLA-DQB1	p.85	0.860	L,x	V	0501, 0502, 0503, 0601, 0602, 0603, 0604, 0605, 0609
		HLA-DQB1	p.89	0.860	T,x	G	0501, 0502, 0503, 0601, 0602, 0603, 0604, 0605, 0609
		HLA-DQB1	p.90	0.860	T,x	I	0501, 0502, 0503, 0601, 0602, 0603, 0604, 0605, 0609
		HLA-DQB1	p.220	0.860	H,x	R	0501, 0502, 0503, 0601, 0602, 0603, 0604, 0609
		HLA-DQB1	p.221	0.860	H,x	Q	0501, 0502, 0503, 0601, 0602, 0603, 0604, 0609
		HLA-DQA1	p.-16	0.861	M	L	0101, 0102, 0103, 0104, 0401, 0601

Table S2.15. Amino Acids in the Credible Sets and Their Corresponding HLA Alleles. (continued)

Class II	AOA CS2	HLA-DQA1	p.26	1.187	T	S	0301, 0302, 0303
		HLA-DQA1	p.47	1.187	C,K,R	Q	0301, 0302, 0303
		HLA-DQA1	p.56	1.187	G,x	R	0301, 0302, 0303
		HLA-DQA1	p.76	1.187	L,M	V	0301, 0302, 0303
		HLA-DQA1	p.187	1.187	A	T	0301, 0302, 0303
		HLA-DRB1	p.-24	1.194	L,x	F	0401, 0402, 0403, 0404, 0405, 0406, 0407, 0408, 0410, 0411
		HLA-DRB1	p.33	1.194	N	H	0401, 0402, 0403, 0404, 0405, 0406, 0407, 0408, 0410, 0411
		HLA-DRB1	p.96	1.194	E,H,Q,x	Y	0401, 0402, 0403, 0404, 0405, 0406, 0407, 0408, 0410, 0411
		HLA-DRB1	p.180	1.194	V,x	L	0401, 0402, 0403, 0404, 0405, 0406, 0407, 0408, 0410, 0411
		HLA-DRB1	p.13	1.194	F,G	H	0401, 0402, 0403, 0404, 0405, 0406, 0407, 0408, 0410, 0411, 1410

Table S2.16. Average r^2 Between Childhood-Onset and Adult-Onset Asthma SNPs

Variant	rs2428494 (Class I COA AOA CS1)	HLA-C p.11 (Class I COA CS2)	rs28407950 (Class II COA CS1)	rs35571244 (Class II COA CS2)	rs9272346 (Class II AOA CS1)	HLA-DQA1*0301 (Class II AOA CS2)
rs2428494	1.00	0.0085	0.0154	0.0112	0.0739	0.0088
HLA-C p.11		1.00	0.0522	0.0017	0.0162	0.0055
rs28407950			1.00	0.0097	0.5013	0.0814
rs35571244				1.00	0.0240	0.1585
rs9272346					1.00	0.1622
HLA-DQA1*0301						1.00

Table S2.17. Marginal vs. Conditional Association Results. P-value, odds ratio (OR), and 95% confidence interval (CI) shown for the original, marginal association and for the association when conditioning on either the class I or class II signals.

Variant	Marginal Association		Conditional Association		
	p-value	**OR [95% CI]**	**p-value**	**OR 95% CI]**	**Conditioned on**
Childhood-Onset Asthma					
rs2428292 (class I CS1)	8.77×10^{-23}	1.16 [1.12-1.19]	5.05×10^{-13}	1.11 [1.08-1.15]	rs28407950 + rs35571244
HLA-C p.11 (class I CS2)	3.12×10^{-19}	0.81 [0.77-0.84]	4.89×10^{-08}	0.87 [0.83-0.92]	rs28407950 + rs35571244
rs28407950 (class II CS1)	1.37×10^{-59}	0.74 [0.71-0.77]	4.22×10^{-44}	0.77 [0.74-0.80]	rs2428494 + HLA-C p.11
rs35571244 (class II CS2)	1.07×10^{-17}	1.25 [1.19-1.32]	5.59×10^{-13}	1.21 [1.15-1.28]	rs2428494 + HLA-C p.11
Adult-Onset Asthma					
rs2428292 (class I CS1)	4.52×10^{-23}	1.10 [1.08-1.13]	2.03×10^{-10}	1.07 [1.05-1.09]	rs9272346 + HLA-DQA1*0301
rs9272346 (class II CS1)	1.98×10^{-47}	1.16 [1.14-1.19]	1.09×10^{-34}	1.14 [1.12-1.17]	rs2428494
*HLA-DQA1**0301 (class II CS2)	6.91×10^{-47}	1.19 [1.16-1.21]	1.99×10^{-41}	1.18 [1.15-1.20]	rs2428494

Table S2.18. Sample Composition of the Replication Cohort.

	Childhood-onset asthma	**Adult-onset asthma**	**Controls**
Sample size	1,686	3666	56,063
Age of asthma onset in years	Range: 0-11 Mean (SD): 6 (3)	Range: 26-65 Mean (SD): 43 (10)	NA
Female Sex	46.1%	64.4%	54.6%
Allergic Disease (ever)	30.5%	26.4%	10.3%

Table S2.19. Self-Reported Ethnic Composition of the Replication Cohort. Number of individuals in each self-reported ethnic group are shown for childhood-onset asthma (COA), adult-onset asthma (AOA), and the non-asthmatic controls (Ctl).

Group	COA	AOA	Ctl
White	14	26	391
British	537	1063	16533
Irish	314	689	9651
Any other white background	395	830	13008
White and Black Caribbean	18	28	459
White and Black African	11	18	308
Black or Black British	0	1	22
Caribbean	114	266	3294
African	60	115	2826
Any other Black background	6	7	84
White and Asian	35	45	617
Asian or Asian British	0	5	30
Chinese	29	53	1311
Indian	91	284	4542
Pakistani	28	127	1352
Bangladeshi	3	10	180
Any other Asian background	31	99	1455

Table S2.20. Replication Meta-Analysis Results. For the candidate variants from each credible set (CS), the SNP, Risk allele, and p-value, odds ratio (OR), 95% confidence interval (CI), and risk allele frequency (AF) are shown for the meta-analysis and for associations with self-reported White British, Asian British, and Black British individuals.

		Meta-Analysis		White British			Asian British			Black British		
snp	Risk Allele	p-value	OR [95% CI]	p-value	OR [95% CI]	AF	p-value	OR [95% CI]	AF	p-value	OR [95% CI]	AF
Childhood-Onset, Class I CS1												
rs2428494	A	9.72E-04	1.13 [1.05-1.21]	6.58E-05	1.17 [1.08-1.27]	0.4	7.58E-01	0.93 [0.76-1.14]	0.3	2.83E-01	1.07 [0.86-1.33]	0.3
HLA-C p.11	Ala	2.66E-03	1.16 [1.05-1.29]	2.97E-04	1.24 [1.10-1.40]	0.2	4.53E-01	1.01 [0.80-1.28]	0.2	3.40E-01	1.05 [0.82-1.35]	0.3
Childhood-Onset, Class I CS2												
rs28407950	C	1.30E-09	1.29 [1.19-1.40]	1.01E-09	1.35 [1.22-1.48]	0.3	4.27E-02	1.19 [0.98-1.45]	0.4	1.90E-01	1.11 [0.88-1.41]	0.3
rs35571244	C	2.74E-01	1.08 [0.94-1.25]	1.50E-01	1.09 [0.93-1.27]	0.1	3.96E-01	1.08 [0.61-1.92]	0.0	4.03E-01	1.06 [0.66-1.70]	0.1
Adult-Onset, Class I CS1												
rs2428494	A	1.27E-02	1.06 [1.01-1.12]	5.29E-02	1.05 [0.99-1.11]	0.4	6.63E-02	1.09 [0.97-1.23]	0.3	6.44E-02	1.12 [0.97-1.31]	0.4
Adult-Onset, Class II CS1												
rs9274660	G	6.10E-08	1.14 [1.09-1.20]	3.06E-06	1.14 [1.08-1.21]	0.6	1.06E-03	1.20 [1.07-1.35]	0.5	1.79E-01	1.07 [0.93-1.24]	0.5
rs3828789	T	4.59E-08	1.15 [1.09-1.20]	3.19E-06	1.14 [1.08-1.21]	0.6	6.57E-04	1.21 [1.08-1.36]	0.5	1.79E-01	1.07 [0.93-1.24]	0.5
rs1063355	G	4.31E-08	1.15 [1.09-1.20]	3.08E-06	1.14 [1.08-1.21]	0.6	6.34E-04	1.21 [1.08-1.36]	0.5	1.79E-01	1.07 [0.93-1.24]	0.5
rs9272346	A	7.52E-08	1.14 [1.09-1.20]	3.64E-06	1.14 [1.08-1.21]	0.6	1.19E-03	1.20 [1.07-1.35]	0.5	1.71E-01	1.07 [0.93-1.24]	0.5
Adult-Onset, Class II CS2												
HLA-DQA1	*03:01	1.27E-04	1.13 [1.06-1.21]	1.50E-03	1.12 [1.04-1.20]	0.2	1.18E-02	1.21 [1.03-1.42]	0.2	1.02E-01	1.18 [0.91-1.53]	0.1

2.6.3 Supplementary Figures

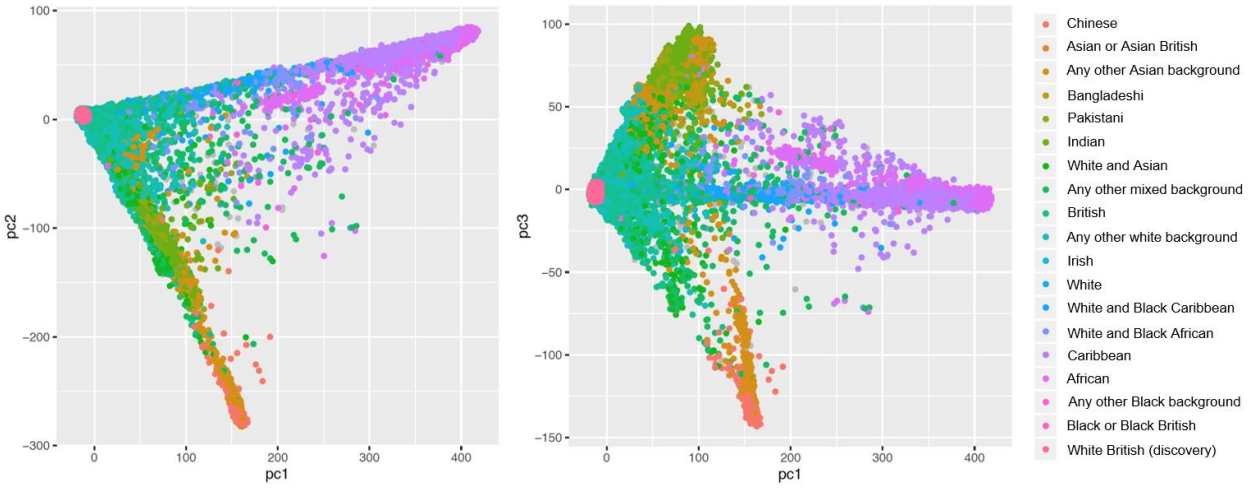


Figure S2.1. Ancestry PCs for the Replication and Discovery Cohorts. Ancestry PC1, PC2, and PC3 are shown for the discovery cohort ("White British (discovery)") and the replication cohort, with the colors corresponding to self-reported ancestry.

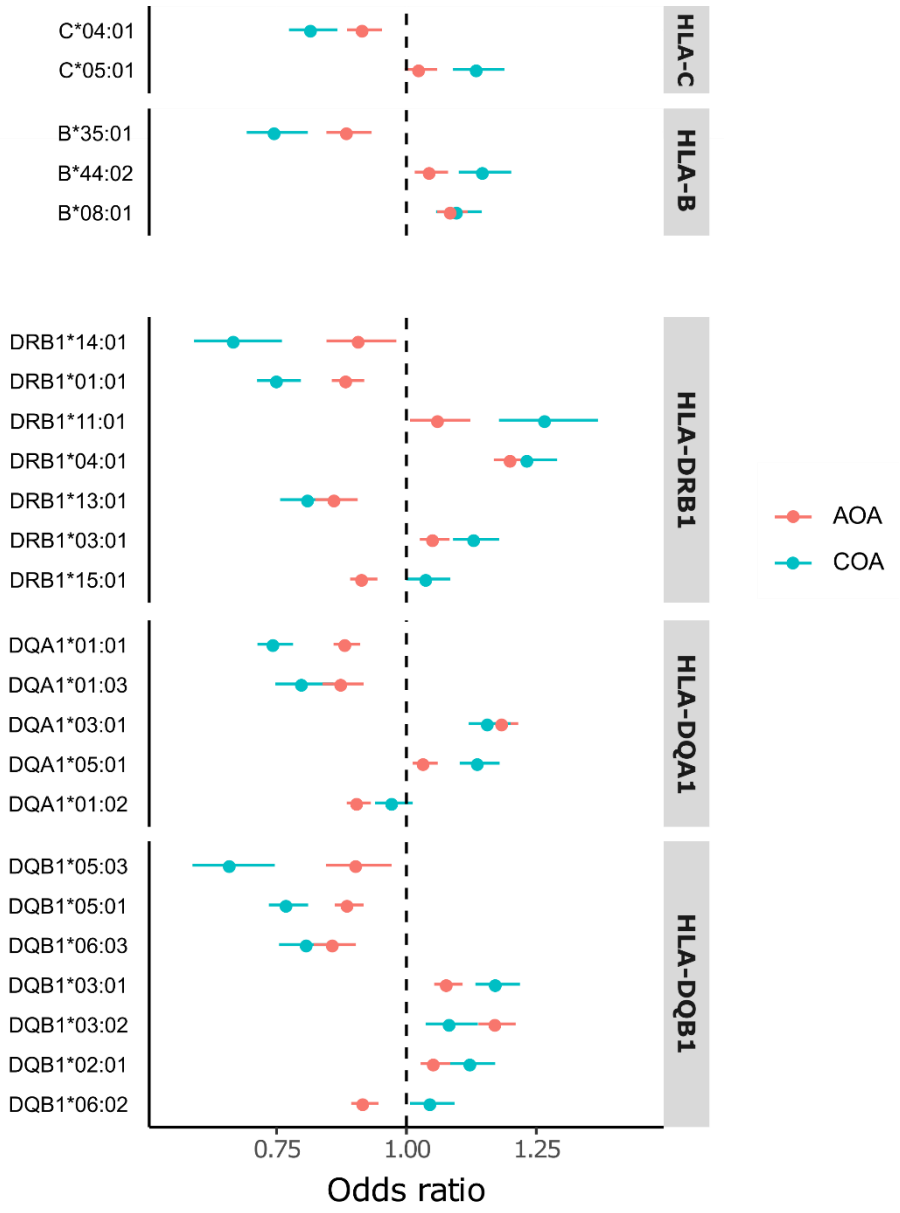


Figure S2.2. HLA Allele Associations. Odds ratios and 95% confidence intervals are shown for the HLA alleles that were significantly associated ($p<5.0x10^{-8}$) with either childhood-onset asthma (COA, blue) and/or adult-onset asthma (AOA, red). The results for all alleles for the six HLA loci are shown in Table S2.2.

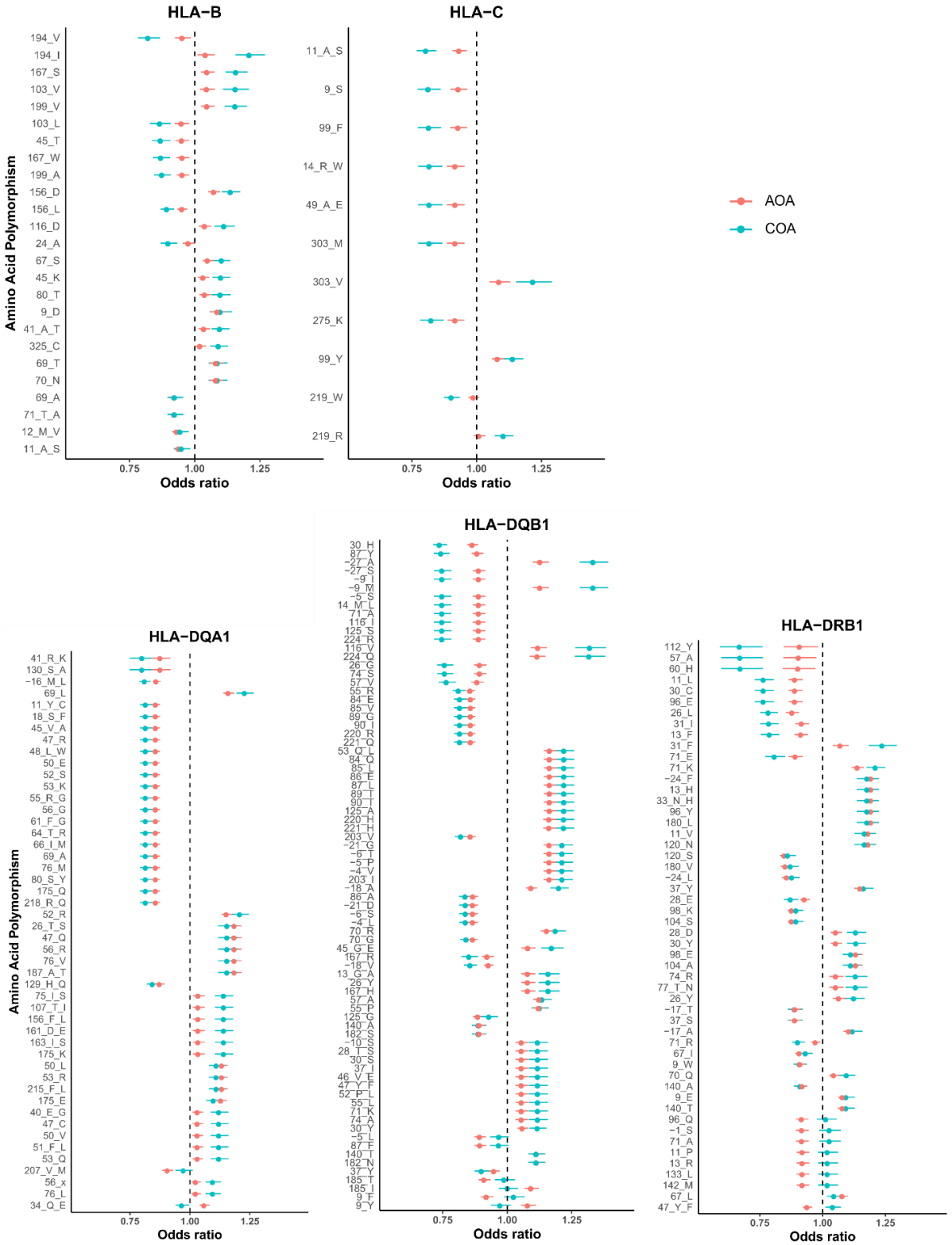


Figure S2.3. Amino Acid Associations. Odds ratios and 95% confidence intervals are shown for HLA amino acid polymorphisms that were significantly associated ($p<5x10^{-8}$) with either childhood-onset asthma (COA, blue) and/or adult-onset asthma (AOA, red).

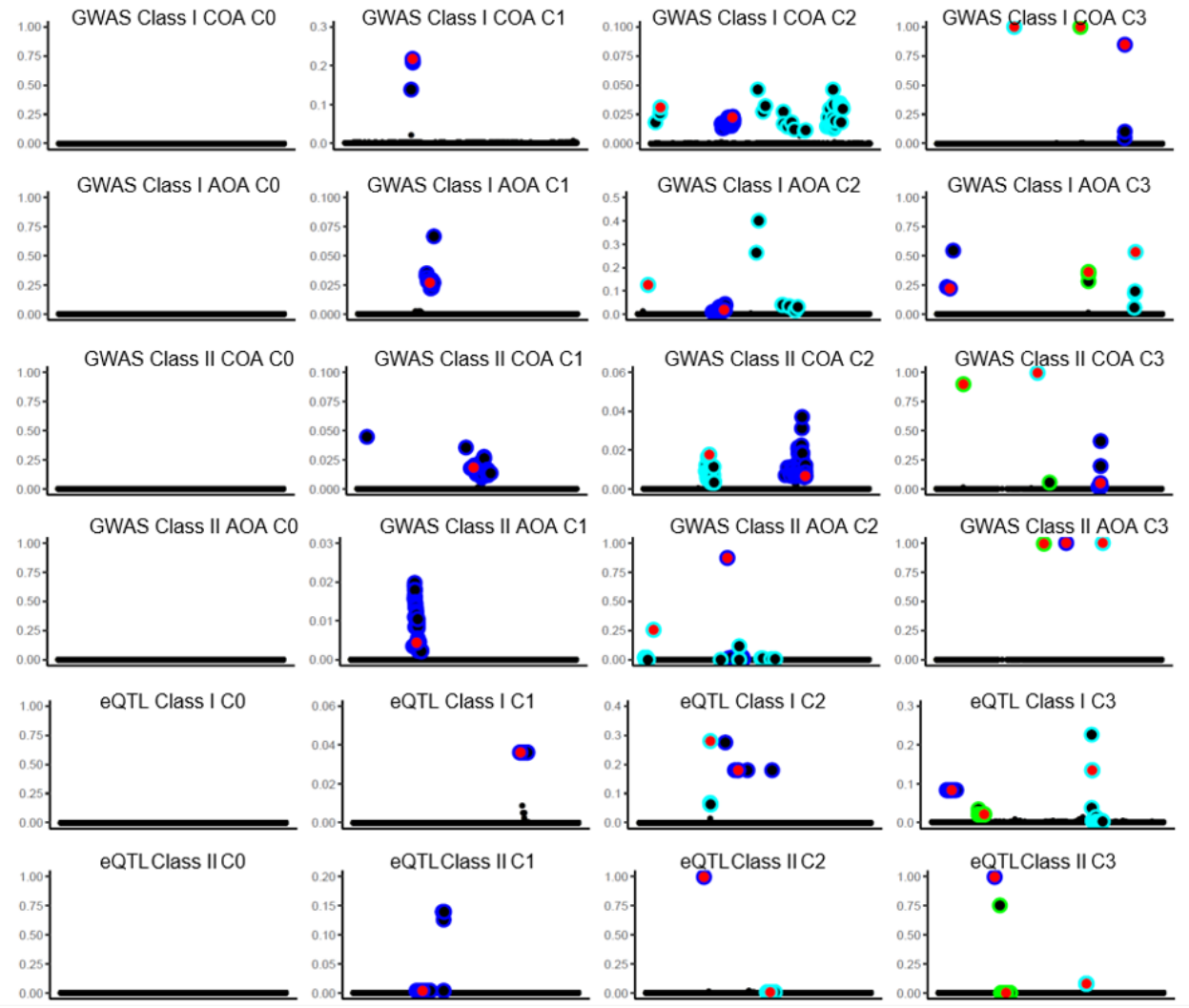


Figure S2.4. Fine-Mapping Simulations in the HLA Region. Each panel is a simulation. The top four rows are the simulated binary ("GWAS") traits and the bottom two rows are the simulated quantitative ("eQTL") traits. Simulations were performed for both the class I and class II regions separately. The binary outcomes were also simulated using covariate effects estimated for either childhood-onset asthma (COA) or adult-onset asthma (AOA) (see Supplementary Methods section 2.6 for more details). C refers to the number of causal variants (0-3). The colors represent the credible sets detected by SuSiE, with the designated causal effect variant(s) in red. SuSiE correctly identified the accurate number of causal signals and reported a true causal signal in each credible set in all the simulations.

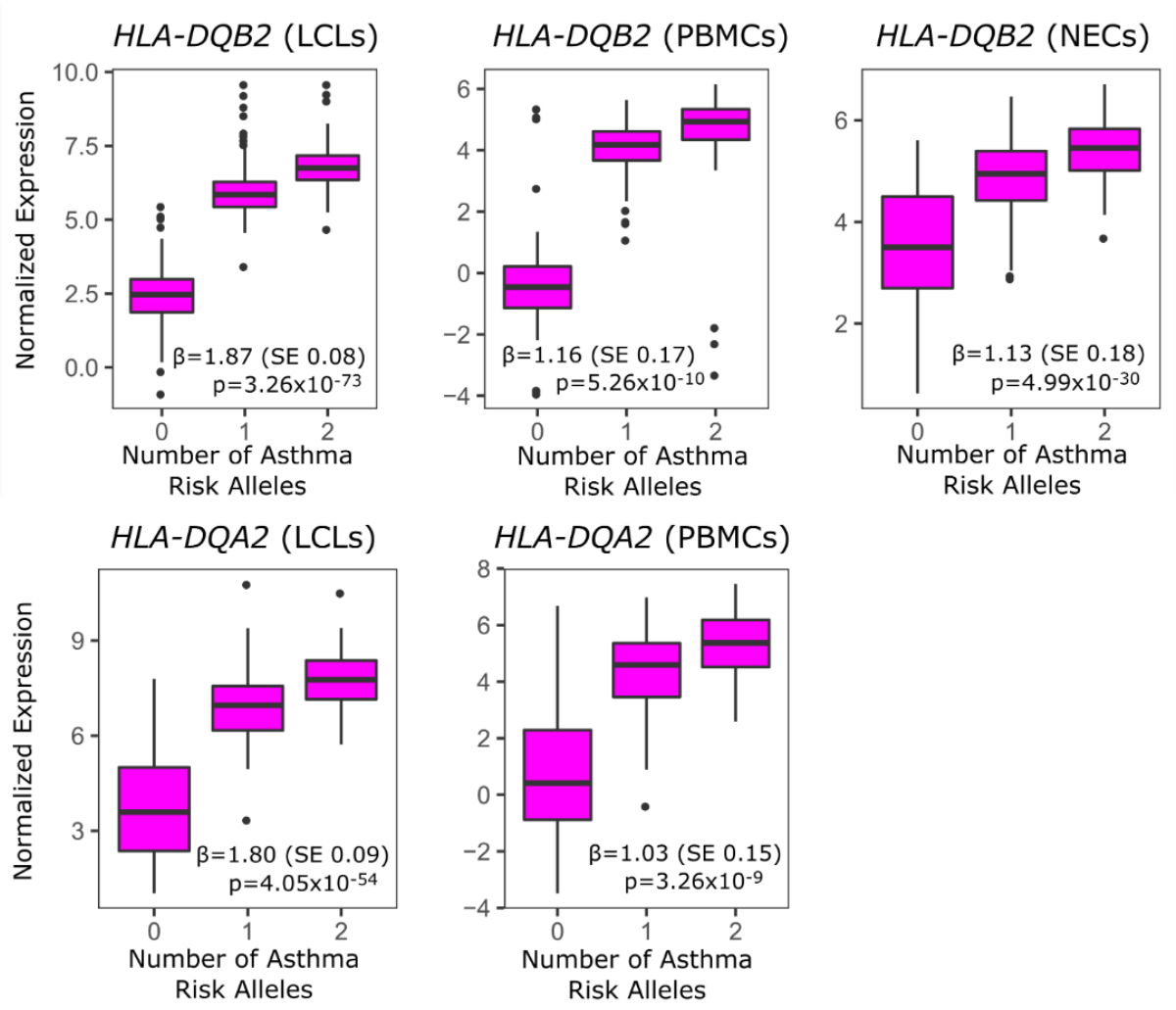


Figure S2.5. Expression of *HLA-DQB2* and *HLA-DQA2*. Normalized expression of each gene by the number of asthma-risk alleles for rs9272346 (A) for lymphoblastoid cell lines (LCLs) and peripheral blood mononuclear cells (PBMCs) and rs9274660 (G) for nasal epithelial cells (NECs), which were representative class II AOA CS1 SNPs.

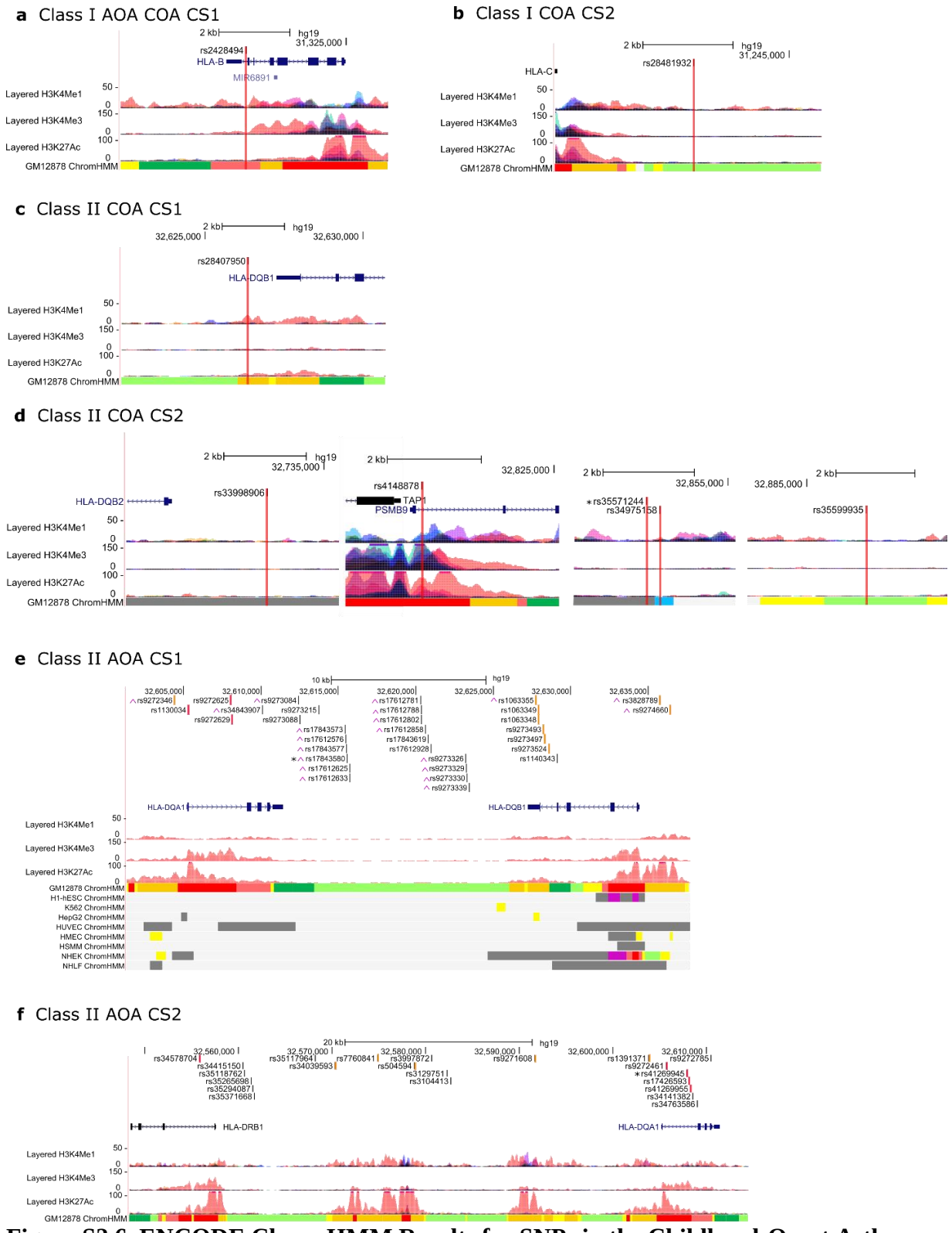


Figure S2.6. ENCODE ChromHMM Results for SNPs in the Childhood-Onset Asthma and Adult-Onset Asthma Credible Sets. Vertical red line indicates the location of each SNP. Layered H3K4Me1, H3K4ME3, and H3K27Ac marks and ChromHMM states are shown for the GM12878 cells. Red: active promoter, light red: weak promoter, purple: inactive/poised

Figure S2.6. ENCODE ChromHMM Results for SNPs in the Childhood-Onset Asthma and Adult-Onset Asthma Credible Sets.(continued) promoter, orange: strong enhancer, yellow: weak/poised enhancer, blue: insulator, dark green: transcriptional transition/elongation, light green: weak transcribed, gray: polycomb-repressed, light gray: heterochromatin/low signal. Asterisk denotes rsid with the highest PIP. **a)** rs2428494 (shared class I CS) was predicted to reside in a weak promoter, **b)** rs28481932 (class I COA CS2) in a weakly transcribed region, and **c)** rs28407950 (class II COA CS1) in an strong enhancer. **d)** Class II COA CS2 SNPs were predicted to reside in polycomb-repressed, active promoter, polycomb-repressed, insulator, and weakly transcribed regions (from left to right). **e)** Class II AOA CS1 SNPs (see main manuscript for a discussion of these results). The red or orange mark next to the rsid indicates it is predicted to reside in an active promoter or strong enhancer, respectively. Magenta ^ indicates if it was an eQTL in our study. **f)** Class II AOA CS2 SNPs. Figures created from http://genome.uscs.edu (113).

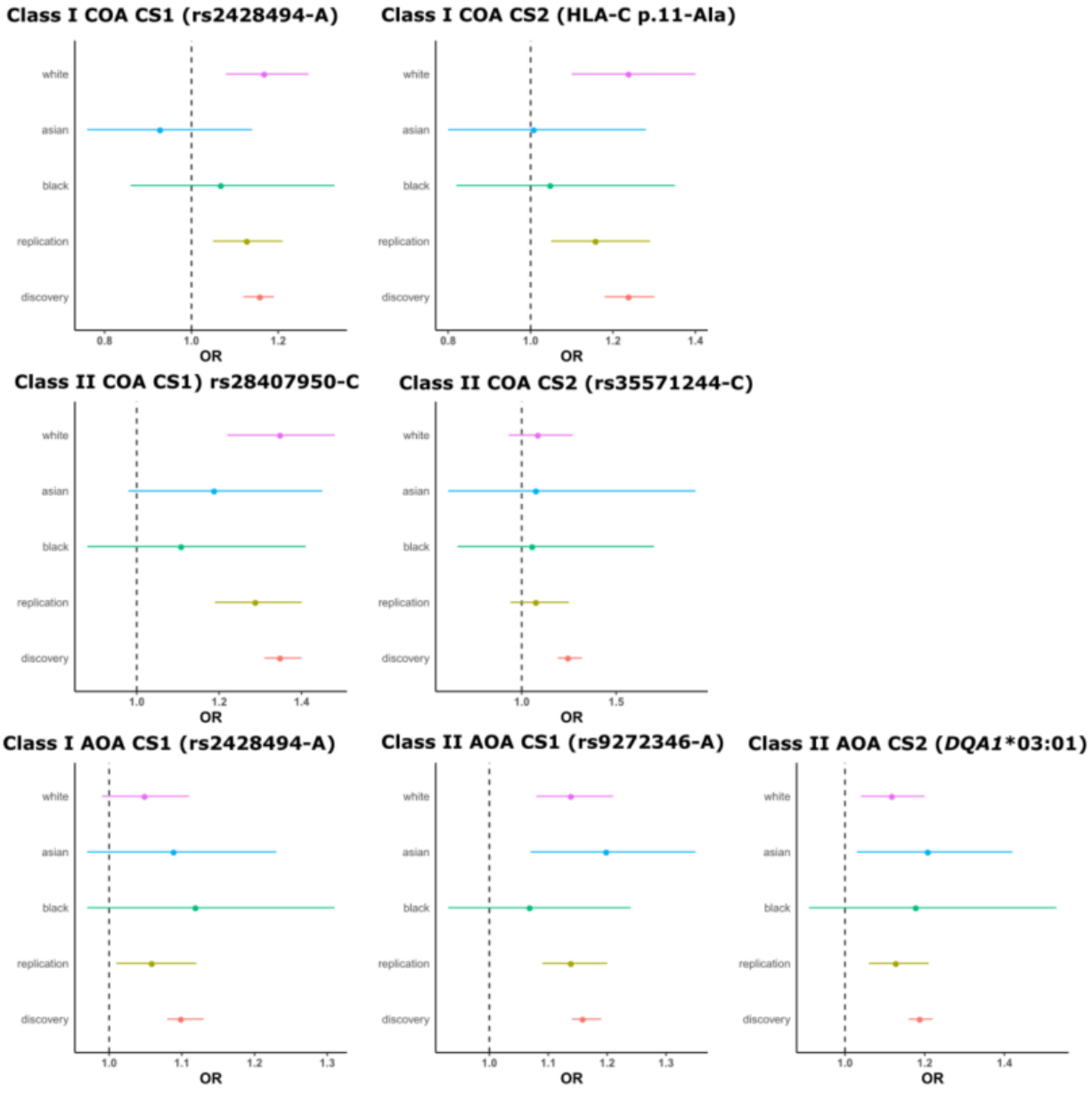


Figure S2.7. Replication Results. Odds ratios and 95% confidence intervals are shown for the candidate variant allele or amino acid polymorphism in the discovery COA and AOA CSs for self-reported White (British, Irish, White, Any other White background), self-reported Black or Black British, self-reported Asian or Asian British, entire replication cohort (consisting of White, Black or Black British, Asian or Asian British), and the White British discovery cohort.

CHAPTER 3

Gene-based association study of rare variants identifies novel genes for asthma-associated quantitative traits in children of diverse ancestries

3.1 ABSTRACT

Background: Most genetic studies of asthma and allergy have focused on common variation in individuals primarily of European ancestry. Studying the role of rare variation in phenotypes measured during youth and in populations of diverse ancestries can provide additional, important insights into the development of these traits.

Objective: The goal of this study was to examine the contribution of rare variants to different asthma- or allergy-associated quantitative traits in children with diverse ancestries.

Methods: We examined whole-genome sequencing (WGS) data from a cohort of children (n=1,035; 67% Black, 25% Hispanic) to identify rare variants (minor allele frequency < 0.01). We assigned variants to genes and tested for associations using an omnibus variant-set test between each of 24,902 genes and eight asthma-associated quantitative traits. We further supported our findings with external data on predicted gene expression and mouse knockout studies.

Results: Rare variants in *USF1* were significantly associated with blood neutrophil count ($p=2.18x10^{-7}$). This finding as well as suggestive significant associations with two other gene-trait pairs (*TNFRSF21* with total IgE [$p=6.47x10^{-6}$] and *PIK3R6* with eosinophil count [$p=4.10x10^{-5}$]) were supported by independent data sources.

3.2 INTRODUCTION

Genome-wide association studies (GWAS) have successfully identified thousands of associations between common variants and hundreds of complex traits (116,150,151). However, rare variants, which comprise the bulk of human genetic variation (49), have been largely overlooked because they are not well represented by the variants included on commonly used genotyping arrays and because most studies are underpowered to detect individual rare variant associations. However, the increased availability of whole-genome sequencing (WGS) and the development of variant-set tests that jointly test the association of multiple variants in a set (e.g., a gene) allow for a more comprehensive study of the role of genome-wide rare variation in complex traits. Although these variants do not necessarily explain a significant proportion of heritability (64), rare variant studies can more directly identify causal genes and mechanisms (50).

Previous studies have explored the contribution of rare variants in asthma and allergic diseases and have implicated genes harboring rare variants in asthma using various study designs across diverse ancestries (58,59,68,60–67). Among the more notable findings are associations between asthma and rare variants in the interleukin 33 (*IL33*) (59,67) and filaggrin (*FLG*) (61,152) genes. Indeed, the discovery of a rare loss-of-function variant in *IL33* that conferred protection from asthma by disrupting binding to its receptor, ST2, led to the identification of astegolimab, an ST2 inhibitor, as an effective therapy for reducing exacerbations in individuals with hard-to-treat asthma (71).

Other studies have explored the role of rare variation in asthma-associated traits, including allergic, inflammatory, and pulmonary traits in cohorts of diverse ancestries (59,60,72–78). However, these studies most often examined just one or a limited number of traits, and the

majority have used only targeted sequencing (59,60,77) or exonic variants (67,72,75,76), or performed GWAS of individual rare variants (73,74). Capturing all coding and non-coding variation across the entire genome through WGS and using a gene-based approach may be better able to study the genetic architecture of asthma-associated quantitative phenotypes.

In this study, we explored the contributions of rare variants (minor allele frequency [MAF] < 0.01) to asthma- and allergy-associated quantitative traits in 1,035 children with diverse ancestries from two longitudinal studies: the Asthma Phenotypes in the Inner City (APIC) study and the URban Environment and Childhood Asthma (URECA) birth cohort study. We used WGS data to perform gene-based association tests, examining associations with traits reflecting various phenotypic components of asthma related to allergy (total immunoglobulin E [IgE], sensitization to specific allergens), inflammatory/immune responses (fraction of exhaled nitric oxide [FeNO], blood eosinophil counts, blood neutrophil counts) and lung function (forced expiratory volume in one second [FEV_1] % predicted, FEV_1/forced vital capacity [FVC], bronchodilator responses). We supported detected associations using external resources of predicted gene expression associations and mouse knockout studies. Our study revealed three novel gene associations with allergic and inflammatory phenotypes, demonstrating the influence of rare variants on these traits and offering potential therapeutic targets.

3.3 METHODS

3.3.1 Study Populations

These studies included children from two cohorts that were component studies of the Inner-City Asthma Consortium (ICAC) (153). APIC (4,6) was a one-year longitudinal study of children and adolescents (ages 6-17) with asthma living in low-income areas of nine U.S. cities (Baltimore,

MD; Boston, MA; Chicago, IL; Cincinnati, OH; Dallas, TX; Denver, CO; Detroit, MI; New York, NY; Washington DC). All APIC participants had a physician's diagnosis for asthma and at least two asthma episodes requiring bronchodilator administration in the previous year (6). URECA is a prospective birth cohort study of children living in low-income areas of four U.S. cities (Baltimore, MD; Boston, MA; New York, NY, St. Louis, MO) (7). Mothers were enrolled into this study during pregnancy; at least one parent had a history of asthma or an allergic disease (7), and asthma in the child was defined by physician's diagnosis at age 7 or 10 years, lung function criteria, and/or reported symptoms (154).

Data for both cohorts were obtained following written informed consent from a parent and assent from the children. The clinical studies in URECA were approved by a Central IRB at the University of Wisconsin and WIRB IRB # 20142570. The clinical studies in APIC were approved at each recruiting site: Johns Hopkins IRB #5; Boston University Medical Campus (Blue Panel IRB); Children's Memorial Medical Center IRB #2011-14581; Cincinnati Children's Hospital IRB; University of Texas Southwestern Medical School IRB #8843; National Jewish Health IRB; Henry Ford Health System IRB #6782; Columbia University Medical Center IRB #1; Children's National Medical Center IRB. The studies described in this manuscript were approved by the University of Chicago IRB 19-0046).

3.3.2 Asthma-Associated Quantitative Phenotypes

Eight quantitative traits that reflect component features of asthma were available for both cohorts. We included two quantitative measures of allergic sensitization: total serum IgE (IU) concentration and the number of sensitizations to 15 common inhaled and food allergens (mouse, dog, house dust mite (x2), cat, roach, mold (x2), ragweed, maple, oak, Timothy grass, peanut, egg, and milk), according to serum allergen-specific IgE concentration (positive is >=0.35

kU/L). Immune and inflammatory phenotypes included blood cell counts (eosinophils and neutrophils, both in cells/mm^3) and FeNO in parts per billion (ppb), which was measured following American Thoracic Society guidelines (155). Lung function measures included percent predicted FEV_1, FEV_1/FVC (using the normalized z-score), and bronchodilator response (percentage change from baseline in FEV_1 following four inhalations of albuterol) (4,7). The Mann-Whitney U Test was used to assess differences in traits between individuals with and without asthma, and the chi-square test was used to assess differences in self-reported race and sex ratios (% female).

3.3.3 Quantitative Trait Normalization

We performed trait normalization for the eight quantitative traits. We adjusted for potential confounding factors by fitting a linear mixed model (LMM) accounting for asthma status, sex, the first 10 principal components (PCs) of ancestry (from common variants(156)), and genetic relatedness between individuals modeled as a random effect. Asthma was included as a covariate to account for disease-induced phenotypic differences; however, we repeated analyses for the most significant results without including asthma as a covariate and in individuals with asthma only to exclude the possibility of collider bias effects (157). We applied a quasi-poisson LMM for the count data (blood neutrophil count, blood eosinophil count, and allergen sensitization) and a gaussian LMM for all other traits. Total IgE and blood neutrophil count were log transformed, and blood eosinophil count was square-root transformed prior to regression fitting. For neutrophil count, we additionally corrected for the SNP within Duffy blood group gene, *ACKR1/DARC* (rs2814778; 1 for CC, 0 for CT, TT) because the homozygous CC genotype is associated with decreased neutrophil counts in individuals of African-ancestries (158). We also repeated analyses without including rs2814778 as a covariate for comparison. For bronchodilator

response we additionally corrected for FEV_1 % predicted and FEV_1^2. Residuals from all regressions were rank-based, inverse-normal transformed (159) and used as the outcome variables in further analyses.

3.3.4 Whole Genome Sequencing and Variant Calling

Whole-genome sequencing (WGS) was performed and variant calls were generated as described in Dapas *et al.* (156) Briefly, WGS was performed using the Illumina NovaSEQ6000, generating 150bp paired-end reads. Sequences were processed according to Genome Analysis Toolkit (GATK) best practices, and reads were aligned to the GRCh38 human reference genome (145). Aligned reads underwent duplicate removal and base quality score recalibration against known sites in the GATK resource bundle (146). Sample swaps were tested with VerifyBamID (160). See Dapas *et al.* 2022 (156) for more details.

To isolate rare variants, we first selected all variants with minor allele frequency (MAF) < 0.01 in the combined APIC and URECA children (**Figure S3.1**). This yielded 21,073,226 rare variants. We further excluded variants that were common (MAF ≥ 0.01) in any of the 1000 Genomes Project super populations (African, American, East Asian, European, and South Asian) because variants with deleterious effects are unlikely to be found at common frequencies in any population (49). This removed 4.56% of variants in our dataset, leaving a final set of 20,093,812 variants for downstream analysis.

3.3.5 Association Tests

To group variants for gene-based association testing, we binned all coding and non-coding rare variants between the 3'UTR and 5kb upstream of the 5'UTR for each gene, including both protein-coding genes and ncRNAs. We used STAAR (variant-Set Test for Association using

Annotation infoRmation) (56) for gene-based association testing with the eight quantitative phenotypes described above. STAAR incorporates three variant set tests: the burden test (51–54), sequence kernel association test (SKAT) (55), and aggregated Cauchy association test (ACAT-V) (161). STAAR can apply a MAF weight to variants; however, we did not weight the variants by their frequency due to the narrow range of MAFs (0.00097 – 0.099) in this sample. We did not evaluate genes that had fewer than five rare variants detected across the cohort. We used a stringent Bonferroni correction of $0.05/(24{,}902 \times 8 \text{ phenotypes}) < 2.51 \times 10^{-7}$ and considered a threshold of $p < 1 \times 10^{-5}$ for suggestive significance. For the top gene associations with each trait, we further examined associations stratified by coding and non-coding variants in the gene set. When we observed cases in which a subset of variants in one gene set resided in the exon of another nearby gene, we repeated analyses after excluding those variants from the gene set. For all genes with evidence of association, we also examined associations for each variant within the gene set through linear regression using the same covariates.

3.3.6 External Association Validation

To evaluate the associations with the most significant gene for each of the eight traits, we first assessed if predicted gene expression was associated with the same or related trait in the phenomeXcan database (162), which reports associations between predicted gene expression based on genetic variation and specific phenotypes (primarily from the UK Biobank (80)). We also evaluated publicly available mouse knockout studies of the associated genes to see if the resulting phenotypes included traits related to those in our study (163). We next examined which variants resided in active enhancer and transcriptional start sites in blood and epithelial cells from ROADMAP (119).

3.4 RESULTS

3.4.1 Sample Composition

We examined rare variation and allergy- and asthma-associated phenotypes from 1,035 children in APIC (n=508) and URECA (n=527). The parent-reported racial composition of their children was 67% Black non-Hispanic, 25% Hispanic, and 8% other (White, mixed, unknown) in APIC and 72% Black non-Hispanic, 20% Hispanic, and 8% other (White, mixed, unknown) in URECA. The demographic and clinical characteristics of these children are shown in **Table 3.1**; a principal component analysis (PCA) plot of the genetic ancestries of these children is shown in **Figure S3.2**. In this sample, there were fewer self-reported Black and more Hispanic individuals among children with asthma compared to children without asthma, but neither proportion of self-reported race nor % female was significantly different after correcting for 11 comparisons ($p_{corrected}<0.0045$). Except for blood neutrophil count, measurements of all the clinical phenotypes significantly differed between children with and without asthma after multiple testing correction (**Table 3.1**). The correlations between phenotypes are shown in **Figure S3.3.**

Table 3.1. Sample Composition. Results are reported separately by asthma status (all, asthma only, and non-asthma only). The ages correspond to the year at which total IgE was measured (all other ages shown in **Table S3.1**). The medians and interquartile ranges (IQR; in parentheses) are shown for all quantitative traits. Missing data were not included in calculations. IU: International units. Ppb: parts per billion. FeNO: fractional exhaled nitric oxide. FEV_1: forced expiratory volume in one second. FVC: forced vital capacity.

Characteristic	All	Asthma	Non-asthma	P-value (Asthma vs Non-asthma)
Sample Size	1035	681	226	-
Mean Age (yrs)	9.99	10.56	9.37	$p=1.51x10^{-7}$
% Female	46.18	43.86	52.65	p=0.025
% Self-Reported Race				
Black	67.25	65.20	75.22	p=0.015
Hispanic	24.93	27.50	16.81	
White	1.35	1.03	1.33	

Table 3.1. Sample Composition. (continued)

% Self-Reported Race Other/Mixed/Unknown	6.47	6.27	6.64	
Traits, median (IQR)				
Allergen sensitization (positive tests)	3.00 (7)	4.67 (7.67)	1.00 (3.33)	$p=7.65x10^{-21}$
Total IgE concentration (IU)	158 (463)	243 (655)	70 (123)	$p=7.67x10^{-21}$
Blood eosinophil count (cells/mm^3)	200 (300)	300 (340)	200 (200)	$p=3.13x10^{-12}$
FeNO (ppb)	17.00 (25.54)	19.50 (29)	8.00 (11)	$p=2.77x10^{-16}$
Blood neutrophil count (cells/mm^3)	2700 (1800)	2700 (1900)	2500 (1700)	p=0.01
FEV_1 % predicted	95.50 (20.37)	93.70 (21.54)	100.75 (16.61)	$p=1.20x10^{-8}$
FEV_1/FVC	-1.06 (1.60)	-1.27 (1.67)	-0.52 (1.20)	$p=1.49x10^{-17}$
Bronchodilator Response (% change from baseline)	8.40 (10.82)	9.80 (11.84)	5.60 (8.06)	$p=7.83x10^{-14}$

3.4.2 Overview of Associations

We performed gene-based variant set tests for eight quantitative traits in participants from the APIC and URECA cohorts. We assigned rare variants to 25,605 genes, and required gene sets to have at least 5 variants, resulting in 24,902 genes examined. The mean number of variants in each variant set was 508 (median number: 207, range 5 - 17,174, **Figure S3.4**). The full gene-based association test results are shown in **Table S3.2** and **File S3.1**; the most significant gene for each of the eight traits is shown in **Table 3.2**.

Table 3.2. Results of rare variant, gene-based association studies. The most significant gene association for each trait is shown. For each quantitative trait, the gene, location, number of variants in the gene set (# Var), p-value in analyses considering just coding or just non-coding variants in the variant set, and the overall p-value are shown. One association (*USF1* with blood neutrophil count, $p=2.18\times10^{-7}$) met the Bonferroni-adjusted significance threshold of 2.51×10^{-7}. The three bolded associations are described in further detail.

Trait	Gene*	Location	# Var	Coding p-value	Noncoding p-value	p-value
Blood neutrophil count	***USF1***	**1q23.3**	**79**	**0.98**	**1.15×10^{-9}**	**2.18×10^{-7}**
Total IgE	***TNFRSF21***	**6p12.3**	**794**	**0.24**	**5.90×10^{-6}**	**6.47×10^{-6}**
FeNO	*CTBP1-AS*	4p16.3	151	0.72	3.26×10^{-5}	2.34×10^{-5}
FEV_1/FVC	*VRK3*	19q13.33	547	0.04	3.59×10^{-5}	3.15×10^{-5}
Bronchodilator Response	*MRPL44*	2q36.1	129	1.68×10^{-3}	3.68×10^{-5}	3.44×10^{-5}
Blood eosinophil count	***PIK3R6***	**17p13.1**	**608**	**8.93×10^{-3}**	**5.12×10^{-5}**	**4.10×10^{-5}**
FEV_1 % predicted	*TEX36-AS1*	10q26.13	93	NA	8.61×10^{-5}	6.08×10^{-5}
Allergen sensitization	*VAMP3*	1p36.23	124	0.75	3.96×10^{-5}	6.42×10^{-5}

* *USF1*: upstream transcription factor 1, *TNFRSF21*: tumor necrosis factor receptor superfamily member 21, *CTBP1-AS*: C-terminal binding protein 1 antisense RNA, *VRK3*: vaccinia-related kinase 3, *MRPL44*: mitochondrial ribosomal protein L44, *PIK3R6*: phosphoinositide-3-kinase-regulatory subunit 6, *TEX36-AS1*: testis expressed protein 36 Antisense RNA 1, *VAMP3*: vesicle associated membrane protein 3.

All associations remained nominally significant in analyses that did not include asthma as a covariate in the trait normalization, and all associations except *VAMP3* with allergic sensitization remained nominally significant in the analyses including only children with asthma (**Table S3.3**). For two genes, *VAMP3* and *USF1*, the variants that were within 5kb upstream of their 5'UTR included some that were in the exons of neighboring genes and were therefore designated as coding variants. Nine SNPs in the *VAMP3* set were coding variants for *CAMTA1*

(calmodulin binding transcription activator 1, **Figure S3.5**); when these SNPs were excluded the p-values for *VAMP3* were largely unchanged ($5.13x10^{-5}$ for the set and p=0.85 for coding variation). The *USF1* results are further discussed in the next section.

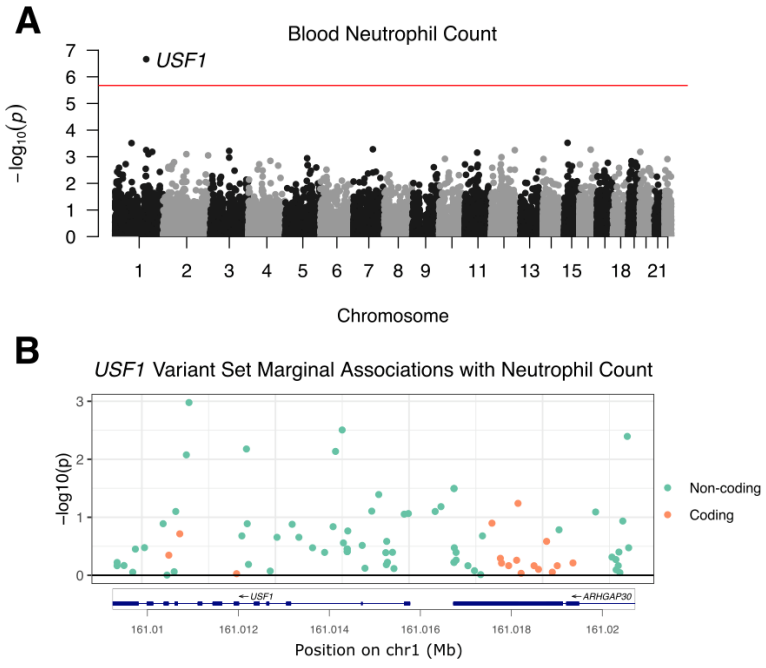
We next used publicly available resources to investigate additional evidence supporting the top gene associations for each of the eight phenotypes (**Table S3.4, S3.5**). Three associations were supported by orthogonal data: *USF1* with blood neutrophil count, *TNFRSF21* with total IgE, and *PIK3R6* with blood eosinophil count. These are discussed below and shown in bold font in **Table 3.2**. The remaining associations did not have corroborating support from these external resources.

3.4.3 USF1 is Associated with Blood Neutrophil Count

The most significant gene-trait pair was *USF1* with blood neutrophil count (**Table 3.2, Figure 3.1A**). This association was also significant ($p=1.97x10^{-8}$) in a secondary analysis that did not correct for genotype at the Duffy blood group gene (see Methods in section 3.3.3). The association was more significant when only the 63 non-coding variants were included in the set ($p=1.15x10^{-9}$); the set comprised only of the 16 coding variants showed no evidence of associations (p=0.98) (**Table 3.2**). No single variant contributed disproportionately to the association (**Figure 3.1B**). The variant set included 13 SNVs in the coding region of the nearby gene (*ARHGAP30*). When we removed those variants, the association for *USF1* was more significant ($p=1.55x10^{-9}$ vs $2.18x10^{-7}$). This variant set was not associated with any of the other seven traits (**Figure 3.1C**).

We next used phenomeXcan, a resource which reports associations between predicted gene expression from multiple tissues and GWAS traits (162), and publicly available mouse

knockout data as complementary approaches to validate the gene association. PhenomeXcan reported that predicted *USF1* expression across 33 tissues (**Table S3.6**) was significantly associated with both neutrophil percentage ($p=4.66x10^{-4}$) and neutrophil count ($p=8.02x10^{-3}$), with lower predicted *USF1* expression associated with higher values of both measures. Furthermore, mice with *Usf1* knocked out in bone-marrow cells had increased blood neutrophil counts (164), which is consistent with the direction of effect from phenomeXcan. According to epithelial cell annotations in ROADMAP (119), 28% of rare variants were located in the transcription start site (TSS) and none were in enhancers. Using the blood cell annotations, 19% were located in the TSS and 34% were in enhancers (**Figure S3.6**). These orthogonal data based on gene expression and mouse knock-out studies support our results and validate the association between *USF1* and neutrophil counts.



C

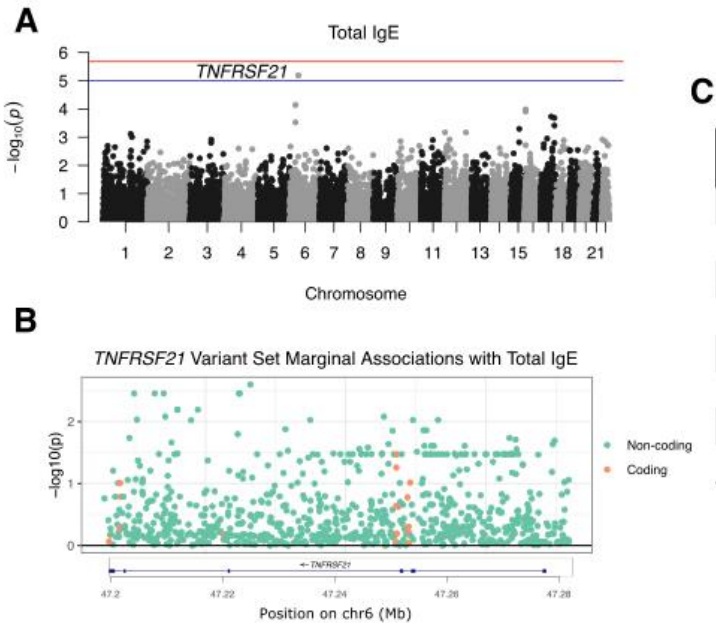
Trait	*P*-Value
Allergen sensitization	0.43
Total IgE	0.64
Blood Eosinophil Count	0.67
Blood Neutrophil Count	**$2.18x10^{-7}$**
FeNO	0.59
FEV_1 % predicted	0.40
FEV_1/FVC	0.57
Bronchodilator Response	0.74

Figure 3.1. *USF1* and Neutrophil Count. A) Manhattan plot for gene set associations with blood neutrophil count. Each point represents a gene. The red line is the Bonferroni significance threshold. **B)** Marginal associations and locations of variants in the *USF1* variant set. Variants

Figure 3.1. *USF1* and Neutrophil Count. (continued) are color coded green if they are non-coding and orange if coding in either *USF1* or its neighboring gene, *ARHGAP30*. **C)** *USF1* association with other traits.

3.4.4 TNFRSF21 is Associated with Total IgE

The second most significant association, which reached suggestive significance ($p=6.47x10^{-6}$), was between *TNFRSF21* and total IgE levels (**Figure 3.2A**). Similar to *USF1*, no single variant contributed disproportionately to the association (**Figure 3.2B**). *TNFRSF21* was also nominally associated with both allergen sensitization and blood eosinophil count (**Figure 3.3C**). However, these three traits were highly correlated (**Figure S3.3**), and when we repeated the associations including total IgE levels as a covariate in the model the associations with allergen sensitization and blood eosinophil count were no longer significant (**Figure 3.3C**). This conditional analysis suggests that the associations with allergic sensitization and eosinophil count were due to the correlations with total IgE levels. Additionally, the association between *TNFRSF21* was slightly more significant when considering the 776 non-coding variants ($p=5.90x10^{-6}$ vs $6.47 x10^{-6}$), whereas the association was not significant when examining only the 18 coding variants (p=0.24, **Table 3.2**).



C

Trait	*P*-Value	Conditional *P*-Value
Allergen sensitization	$3.78x10^{-4}$	1.00
Total IgE	**$6.47x10^{-6}$**	**1.00**
Blood Eosinophil Count	$7.16x10^{-3}$	0.97
Blood Neutrophil Count	0.91	0.96
FeNO	0.25	0.85
FEV_1 % predicted	0.17	0.38
FEV_1/FVC	0.43	0.84
Bronchodilator Response	0.25	0.37

Figure 3.2. *TNFRSF21* and Total IgE. A) Manhattan plot for total IgE levels. Red line is Bonferroni significance and blue is suggestive significance ($1.00x10^{-5}$) **B)** Marginal associations of variants in the *TNFRSF21* variant set. **C)** *TNFRSF21* association with the other traits before and after conditioning on total IgE.

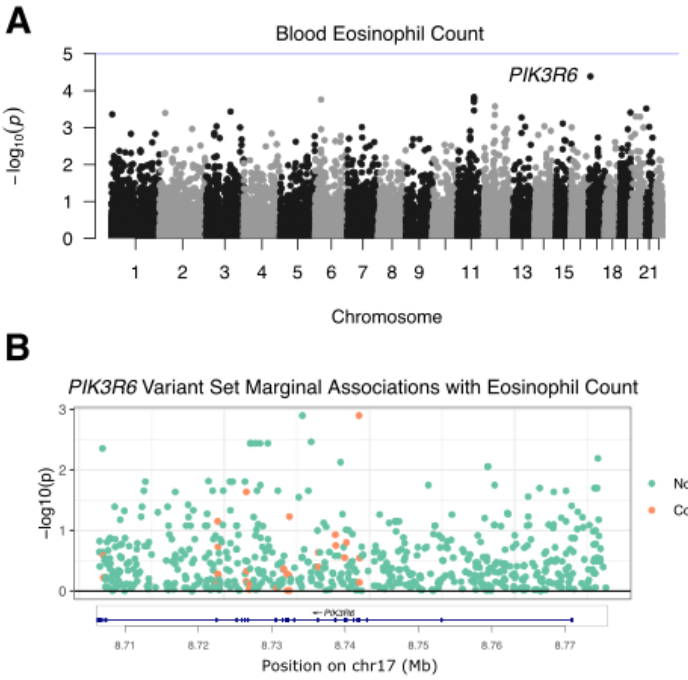
Total IgE was not available in the phenomeXcan dataset and predicted *TNFRSF21* expression was not associated with the available allergic traits (food allergy, allergic rhinitis, and eczema). However, the Mouse Genome Informatics resource reported that *Tnfrsf21* mouse knockouts had increased IgE levels (163,165), supporting the relationship between *TNFRSF21* and total IgE. ROADMAP epithelial cell annotations indicated 3% of rare variants were located in the TSS and 4% were in enhancers. According to blood cell annotations, 0.1% were located in the TSS and 4% were in enhancers (**Figure S3.6**).

3.4.5 PIK3R6 is Associated with Blood Eosinophil Count

The most significant association for blood eosinophil count was with *PIK3R6* with $p=4.10x10^{-5}$ (**Figure 3.3A**). In contrast to the previous two gene-trait pairs, the p-value for the associations

between *PIK3R6* and eosinophil count was slightly less significant when analyzing just the 581 non-coding variants ($p=5.19x10^{-5}$) and retained nominal significance ($p=3.90x10^{-3}$) when considering only the 27 coding variants. This indicates that both sets of variants contributed to the association. No single variant was responsible for the entire signal (**Figure 3.3B**), and *PIK3R6* was not associated with any other traits (**Figure 3.3C**).

The association was validated by phenomeXcan: the most significant association for predicted *PIK3R6* expression was with eosinophil percentage ($p=3.96x10^{-6}$) across 26 tissues (**Table S3.6**). It was also associated with eosinophil count ($p=1.85x10^{-4}$). Furthermore, *Pik3r6* knockout mice reported decreased granulocyte numbers (163,166). Although the type of granulocyte was not specified, these measures would have included eosinophils. According to epithelial cell annotations in ROADMAP, 5% of the rare variants were located in enhancers and none were in the TSS. Similarly, using the blood cell annotations, 16% were located in enhancers and none were in the TSS (**Figure S3.6**). Taken together, these results further support a possible association between *PIK3R6* variants and eosinophil counts.



C

Trait	*P*-Value
Allergen sensitization	0.42
Total IgE	0.86
Blood Eosinophil Count	**4.10×10^{-5}**
Blood Neutrophil Count	0.26
FeNO	0.50
FEV_1 % predicted	0.18
FEV_1/FVC	0.76
Bronchodilator Response	0.33

Figure 3.3. *PIK3R6* and Eosinophil Count. A) Manhattan plot for blood eosinophil count. **B)** Marginal associations of variants in *PIK3R6* variant set. **C)** *PIK3R6* association with the other traits.

3.4.6 Individual Test Performance

The STAAR test used in our study provides p-values that represent the combined results from three different gene-based association tests: the burden, SKAT, and ACAT-V tests (51–55,161). The burden test is more powerful when a high proportion of variants are causal and with the same direction of effect on the outcome. SKAT is more powerful when the effects of different variants are in different directions. ACAT-V is more powerful when the contributing variants are very rare or when only a few in the variant set are causal (56,161,167). Among the three tests, the most significant associations between the three gene-phenotype pairs discussed above were with the burden test: *USF1* and neutrophil count ($p_{Burden}=7.28 \times 10^{-8}$, $p_{SKAT}=1.84 \times 10^{-3}$, $p_{ACAT\text{-}V}=2.21 \times 10^{-5}$), *TNFRSF21* and total IgE ($p_{Burden}=2.25 \times 10^{-6}$, $p_{SKAT}=0.014$, $p_{ACAT\text{-}V}=5.13 \times 10^{-5}$), and

PIK3R6 and eosinophil count ($p_{Burden}=1.39x10^{-5}$, $p_{SKAT}=2.59x10^{-3}$, $p_{ACAT\text{-}V}=1.82x10^{-3}$). For all three gene-trait pairs, the p-values from the Burden test alone were more significant than the p-value from combined STAAR test.

We performed associations for each variant in the gene set and examined the direction of effect for the marginal associations of variants that were nominally significant ($p<0.05$) (**Figure S3.7**). For *USF1* and neutrophil count, all nominally significant variants had $\beta < 0$. For *TNFRSF21* and total IgE, 81.4% of nominally significant variants had $\beta < 0$; and for *PIK3R6* and eosinophil count, 86.7% of variants had $\beta < 0$. The fact that >80% of associated variants in each of the three genes had the same direction of effect is consistent with the Burden test providing the most significant p-values for each gene-trait pair.

3.5 DISCUSSION

In this study, we examined the contribution of rare variants to allergy- and asthma-associated quantitative traits using a gene-based approach. We included a broad panel of allergic, immune/ inflammatory, and pulmonary traits measured in children with diverse ancestries. Overall, our studies suggested that rare noncoding variation within or upstream of genes contribute more to variation in these traits than rare coding variation (**Table 3.2**), further supporting other findings that complex traits are primarily driven by both common and rare noncoding variation that likely have effects on gene expression (168–170). Indeed, three of the associations were supported by orthogonal, independent data based on predicted gene expression and mouse knockout studies. All three of these findings represent novel gene associations with immune, atopic, and inflammatory phenotypes.

One association was significant after multiple test correction for 24,902 genes. The gene set for *USF1* was associated with blood neutrophil count in the APIC and URECA children, and predicted *USF1* expression was associated with neutrophil counts in adults in the UK Biobank. Moreover, a mouse model with *Usf1* knocked out had increased neutrophil counts (164). Both the phenomeXcan and mouse knockout results indicated that decreased expression of this gene is associated with elevated levels of circulating neutrophils. *USF1* encodes a transcription factor belonging to the basic helix-loop-helix leucine zipper family of proteins that can regulate expression through E-box motifs (171). This gene is also located within 2Mb of the Duffy blood group gene (*ACKR1*/*DARC* (158)), but because this association remained significant when conditioning on the Duffy SNP, we were able to rule out that the association between *USF1* and neutrophil counts is not due to LD with the Duffy null allele (see Methods). Several studies have reported a relationship between *USF1* and immune-related traits (164,172–175), and the locus containing this gene has been significantly associated with both white blood cell count (176) and granulocyte percentage of myeloid white cells (177) in GWAS. As a transcription factor, USF1 may be regulating expression of genes important in inflammatory processes and possibly even asthma: several studies have reported asthma risk alleles at the loci encoding the *MUC5AC* and *ORMDL3/GSDMB* genes affecting binding of this transcription factor (20,178,179).

We also report an association between *PIK3R6* and eosinophil count. This was further supported by both phenomeXcan and *Pik3r6* knockout mice (166). *PIK3R6* encodes a lipid kinase which acts as a regulatory subunit for the PI3K gamma complex and is primarily activated by G protein coupled receptors (180). Notably, several studies in human cohorts have implicated *PIK3R6* expression in other eosinophilic or allergic diseases. For example, *PIK3R6* was among the most significantly differentially expressed genes in peripheral blood cells between

eosinophilic and non-eosinophilic chronic obstructive pulmonary disease (COPD) (181), between patients with allergic asthma and non-asthma/non-allergic controls (182), and in a meta-analysis of atopy in white blood cells (183). Altogether, these studies point to a potentially important role of this gene in eosinophilic-mediated allergic traits.

The second most significant association was between *TNFRSF21* and total IgE. Measures of IgE were not available in phenomeXcan, but *Tnfrsf21* -/- mice had increased IgE levels (165). *TNFRSF21* encodes the Death Receptor 6 (DR6) protein and is a member of the tumor necrosis factor superfamily and signals through the NF-kB pathway (184,185). Several mouse studies have also implicated this gene in the Th2 response (165,186,187), specifically through activating the Jun amino terminal kinase (JNK) pathway in regulating Th cell differentiation (165). Th2 inflammation is characterized by an increase in Th2 cytokines, including IL-4, IL-5, IL-9, and IL-13 (188), which play central roles in allergic disease and, particularly IL-13, in asthma.

Some of the findings for other traits were also potentially interesting but did not have corroborating evidence from other studies (**Table S3.4, S3.5**). For example, the most significant association with bronchodilator response was with the mitochondrial ribosomal protein L44 (*MRPL44*), (189). In another study of rare variants with asthma and allergy-related phenotypes, Morin *et al.* (2019) reported an association between *MRPL44* and eosinophil count (60). It is therefore possible that this gene plays a role in asthma severity. *VAMP3* was associated with allergic sensitization in our study. While mouse knockout studies reported a "normal immune system phenotype" and other non-allergic traits (e.g. "abnormal neurotransmitter secretion" and "abnormal vibrissa morphology" (163)), phenomeXcan reported associations between variants that predict expression of *VAMP3* and allergic traits ("hayfever, allergic rhinitis, or eczema") ($p=2.24x10^{-4}$) and asthma ($p=2.83x10^{-3}$). *VAMP3* regulates expression of FcεRI, the high affinity

receptor for IgE that has previously implicated in genetic studies of asthma and allergic diseases and in allergic responses (21,190–193).

There are limitations to this study. First, we had low power due to the modest sample size (n=1,035) to detect associations with individual rare variants. We were also not well powered to detect associations with asthma. Thus, associations between these genes and asthma will have to be explored in future, larger studies. Second, the rare variant method we used (STAAR) was not able to report effect estimates or standard errors, which is a limitation of this study. Third, not all gene candidates identified in our studies (*CTPB1-AS*, *TEX36-AS1*) were available in phenomeXcan or the Mouse Genome Informatics resource, and not all phenotypes (total IgE, bronchodilator response, FEV_1/FVC) were available in phenomeXcan, so it was not possible to validate all top associations. Fourth, the findings of this study have not been replicated in an independent cohort with WGS and similar phenotypes. Nonetheless, we were able to validate several of the most significant associations using phenomeXcan and/or results from mouse knockout studies. These combined data underscore the robustness of the associations across biological contexts and provide convergent support for expression levels of these genes impacting asthma-associated quantitative phenotypes.

In summary, we identified rare variants in genes associated with allergic and inflammatory phenotypes using WGS data from a cohort of children with diverse ancestries. We further validated three associations through external gene expression and mouse knockout resources. Overall, our studies in a well-characterized cohort of children highlight the importance of rare variation in the development of asthma-associated quantitative phenotypes and propose novel candidate genes that may contribute to these important traits and serve as therapeutic targets.

3.5.1 Author Contributions

Conceptualization and design of study: Selene Marilyn Clay, Carole Ober, Matt Dapas; Patient samples/data: Cynthia M. Visness, Robert A. Wood, George T. O'Connor, Robyn T. Cohen, Gurjit K. Khurana Hershey, Carolyn M. Kercsmar, Rebecca S. Gruchalla, Michelle A. Gill, Andrew H. Liu, Haejin Kim, Meyer Kattan, Leonard B. Bacharier, Deepa Rastogi, Katherine Rivera-Spoljaric, Rachel G. Robison, Peter J. Gergen, , William W. Busse, James E. Gern, Daniel J. Jackson; Statistical Analyses: Selene Marilyn Clay, Matt Dapas; Interpretation of Data: Selene Marilyn Clay, Matt Dapas, Carole Ober; Writing: Selene Marilyn Clay, Matt Dapas, and Carole Ober wrote the paper, all authors read and commented on drafts.

3.6 SUPPLEMENTARY INFORMATION

3.6.1 Supplementary Tables

Table S3.1. Mean Age the Data for Each Phenotype Was Collected. Results shown separately for all, asthmatic, and non-asthmatic individuals.

Trait	**All**	**Asthmatics**	**Non-Asthmatics**
Allergen sensitization	9.99	10.55	9.37
Total IgE	9.99	10.56	9.37
Blood eosinophil count	10.00	10.55	9.31
FeNO	10.37	10.65	9.03
Blood neutrophil count	10.05	10.54	9.32
FEV_1 % predicted	10.34	10.64	9.68
FEV_1/FVC	10.31	10.62	9.59
Bronchodilator Response	10.01	10.46	8.83

Table S3.2. Full Gene-Based Association Test Results. Results for associations between all genes and all 8 phenotypes (sens=allergen sensitization, eos=eosinophil count, neut=neutrophil count, tige=total ige, eno=feno, fprev=FEV_1% predicted, frev=bronchodilator response, ffvc=FEV_1/FVC). The overall STAAR p-value is shown, along with the p-values from the SKAT, burden, and ACAT-V variant set tests. See File S3.1 for full results.

gene	chr	nvar	trait	STAAR_p	SKAT_p	Burden_p	ACAT_V_p
A3GALT2	1	143	sens	0.970611	0.533794	0.990058	0.589468
A3GALT2	1	143	eos	0.462919	0.43719	0.466711	0.485311
A3GALT2	1	143	neut	0.274556	0.644225	0.113679	0.38542
A3GALT2	1	142	tige	0.402772	0.302698	0.457292	0.469126
A3GALT2	1	116	eno	0.698906	0.367897	0.791552	0.791552
A3GALT2	1	142	fpred	0.21444	0.428623	0.084373	0.548572
A3GALT2	1	141	frev	0.589703	0.346623	0.443171	0.819654
A3GALT2	1	142	ffvc	0.693839	0.779847	0.796842	0.361835
AADACL3	1	193	sens	0.819601	0.844419	0.804373	0.804373
AADACL3	1	192	eos	0.246032	0.363073	0.207717	0.207717
AADACL3	1	192	neut	0.701226	0.755487	0.667566	0.667566
AADACL3	1	194	tige	0.906042	0.92323	0.893718	0.894908
AADACL3	1	161	eno	0.35814	0.206892	0.481298	0.481298
AADACL3	1	194	fpred	0.276554	0.131244	0.3639	0.564355
AADACL3	1	190	frev	0.023373	0.061767	0.010386	0.062155
AADACL3	1	195	ffvc	0.610375	0.695099	0.542308	0.577105
AADACL4	1	326	sens	0.177448	0.09471	0.243942	0.349337
AADACL4	1	325	eos	0.193504	0.097351	0.26841	0.420868
AADACL4	1	325	neut	0.420348	0.559204	0.267256	0.481595
AADACL4	1	325	tige	0.461807	0.310973	0.61674	0.477538
AADACL4	1	262	eno	0.297132	0.124187	0.533389	0.533389
AADACL4	1	324	fpred	0.809821	0.387931	0.905072	0.815087
AADACL4	1	313	frev	0.965789	0.507139	0.725718	0.988183
AADACL4	1	323	ffvc	0.438068	0.290827	0.603951	0.450012
ABCA4	1	1238	sens	0.905569	0.948153	0.760991	0.885438
ABCA4	1	1234	eos	0.248719	0.363588	0.26387	0.173355
ABCA4	1	1234	neut	0.964753	0.620454	0.14058	0.988913
ABCA4	1	1237	tige	0.94703	0.920266	0.951309	0.95762
ABCA4	1	1034	eno	0.349149	0.310699	0.370426	0.370426
ABCA4	1	1238	fpred	0.490765	0.453333	0.626964	0.389766
ABCA4	1	1218	frev	0.209918	0.38097	0.120789	0.254898
ABCA4	1	1233	ffvc	0.572805	0.560378	0.536338	0.618848
ABCB10	1	364	sens	0.145676	0.19196	0.578631	0.064196
ABCB10	1	362	eos	0.469347	0.474146	0.410378	0.525678
ABCB10	1	362	neut	0.96581	0.984011	0.89557	0.937097
ABCB10	1	363	tige	0.054007	0.245338	0.688814	0.018533

Table S3.3. Most Significant Gene Associations Correcting for Asthma. For each quantitative trait, the gene, location, number of variants in the gene set (# Var), raw p-value for the trait when including and not including asthma as a covariate in the trait normalization and in just asthmatics.

Trait	Gene	Location	# Var	P-value (asthma as a covariate)	P-value (asthma not as a covariate)	P-value (just asthmatics)
Allergen sensitization	*VAMP3*	1p36.23	124	6.42×10^{-5}	7.83×10^{-4}	0.69
Total IgE	*TNFRSF21*	6p12.3	794	6.47×10^{-6}	3.81×10^{-5}	5.55×10^{-5}
Blood eosinophil count	*PIK3R6*	17p13.1	608	4.10×10^{-5}	1.76×10^{-4}	4.36×10^{-6}
FeNO	*CTBP1-AS*	4p16.3	151	2.34×10^{-5}	6.28×10^{-5}	3.72×10^{-4}
Blood neutrophil count	*USF1*	1q23.3	79	2.18×10^{-7}	1.67×10^{-7}	3.72×10^{-5}
FEV_1 % predicted	*TEX36-AS1*	10q26.13	93	6.08×10^{-5}	1.46×10^{-4}	8.42×10^{-3}
FEV_1/FVC	*VRK3*	19q13.33	547	3.15×10^{-5}	1.30×10^{-3}	5.11×10^{-4}
Bronchodilator Response	*MRPL44*	2q36.1	129	3.44×10^{-5}	4.31×10^{-5}	1.15×10^{-3}

Table S3.4. Overview of PhenomeXcan and Mouse KO Supportive Evidence. For each gene and the asthma-associated quantitative phenotype it is associated with in our study, a "yes" indicates that PhenomeXcan and/or independent mouse KO studies support that association, and a "no" indicates conflicting results. We also note which genes and/or traits were unavailable to study. See Table S3.5 for more details.

Trait	Gene	PhenomeXcan	Mouse KO
Blood neutrophil count	*USF1*	yes	yes
Total IgE	*TNFRSF21*	IgE not available	yes
FeNO	*CTBP1-AS*	*CTBP1-AS* not available	*CTBP1-AS* not available
FEV_1/FVC	*VRK3*	FEV_1/FVC not available	FEV_1/FVC not available
Bronchodilator Response	*MRPL44*	Bronchodilator Response not available	Bronchodilator Response not available
Blood eosinophil count	*PIK3R6*	yes	yes
FEV_1 % predicted	*TEX36-AS1*	*TEX36-AS1* not available	*TEX36-AS1* not available
Allergen sensitization	*VAMP3*	yes	no

Table S3.5. Supportive Evidence from PhenomeXcan and Mouse KO Studies. For each gene and the asthma-associated quantitative phenotype it is associated with in our study, the top five results and any additional relevant phenotypes from phenomeXcan (162) are reported, along links to the results of mouse KO studies (163) and accompanying papers describing the relevant phenotype, if available. PhenomeXcan traits are from the UK Biobank (https://biobank.ctsu.ox.ac.uk/crystal/search.cgi, (80)), unless otherwise noted.

Gene	Phenotype	Top 5 PhenomeXcan Results	Relevant PhenomeXcan phenotype(s)	Mouse KO studies
VAMP3	Allergen Sensitization	1. Mean platelet (thrombocyte) volume (p=2.62e-19) 2. Morning/evening person (chronotype) (7.32e-18) 3. Platelet distribution width (1.81e-13) 4. Platelet count (4.30e-10) 5. Getting up in morning (2.45e-9)	1. Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor: Hayfever, allergic rhinitis or eczema (p=2.24e-4) 2. Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor: Asthma (p=2.83e-3)	http://www.informatics.jax.org/marker/MGI:1321389
TNFRSF21	Total IgE	1. Mean platelet (thrombocyte) volume (p=1.98e-9) 2. Treatment/medication mode: rabeprazole sodium (20003_1141168584) (p=2.84e-4) 3. Platelet distribution width (5.39e-4) 4. Job SOC coding: Construction operatives n.e.c. (1.114e-3) 5. Hearing difficulty/problems: No (1.20e-3)	Total IgE not available	http://www.informatics.jax.org/marker/MGI:2151075 https://pubmed.ncbi.nlm.nih.gov/11714751/

Table S3.5. Supportive Evidence from PhenomeXcan and Mouse KO Studies. (continued)

PIK3R6	Blood Eosinophil Count	1. Eosinophil percentage (p=3.96e-6) 2. Ischaemic Stroke, excluding all haemorrhages (2.20e-5), 3. Non-cancer illness code, self-reported: hypothyroidism/myxoedema (2.48e-5) 4. Stroke, excluding SAH (1.62e-4) 5. Diagnoses - main ICD10: 163 Cerebral infarction (1.77e-4)	1. Eosinophil percentage (p=3.96e-6) 2. Eosinophil count (p=1.85e-4)	http://www.informatics.jax.org/marker/MGI:2144613 https://pubmed.ncbi.nlm.nih.gov/25605974/
CTBP1-AS	FeNO	*CTBP1-AS* not available	*CTBP1-AS* not available	*CTBP1-AS* not available
USF1	Blood Neutrophil count	1. Monocyte percentage (p=7.11e-11) 2. Monocyte Count (4.50e-8, from Astle et al 2016) 3. Monocyte count (3.75e-7, from UKB) 4. Platelet distribution width (1.17e-5) 5. Mean reticulocyte volume (2.51e-5)	1. Neutrophil percentage (p=4.04e-4) 2. White blood cell (leukocyte) count (7.44e-4) 3. Neutrophil count (8.02e-3)	http://www.informatics.jax.org/marker/MGI:99542 https://pubmed.ncbi.nlm.nih.gov/34385562/
TEX36-AS1	FEV_1 % Predicted	*TEX36-AS1* not available	*TEX36-AS1* not available	*TEX36-AS1* not available

Table S3.5. Supportive Evidence from PhenomeXcan and Mouse KO Studies. (continued)

VRK3	FEV_1/FVC	1. Treatment/medication code: epanutin 25mg capsule (20003_1140872112) (p=1.67e-5) 2. Number of times self-harmed (1.41e-4) 3. Delivery. Methods: Other than those specified above (1.93e-4) 4. Other specified/unspecified disorders of tympanic membrane (4.18e-4) 5. Diagnoses - main ICD10: H73 Other disorders of tympanic membrane (8.33e-4)	FEV_1/FVC not available	http://www.informatics.jax.org/marker/MGI:2182465
MRPL44	Bronchodilator Response	1. Pain type(s) experienced in last month: Pain all over the body (p=2.22e-4) 2. Sitting height (2.83e-4) 3. Mean reticulocyte volume (3.23e-04) 4. Mean platelet (thrombocyte) volume (3.69e-4) 5. Mean sphered cell volume (5.78e-4)	Bronchodilator Response not available	http://www.informatics.jax.org/marker/MGI:1916413

Table S3.6. PhenomeXcan Single Tissue Results. For i) *USF1* and neutrophil percentage and ii) *PIK3R6* and eosinophil percentage, the tissue and p-value are shown for all tissues with a p-value < 0.05.

Gene trait pair	**Tissue (p-value)**
USF1 and Neutrophil percentage	Thyroid (4.81E-06), Muscle_Skeletal (5.76E-06), Colon_Sigmoid (5.89E-06), Skin_Sun_Exposed_Lower_leg (6.85E-06), Artery_Tibial (7.27E-06), Esophagus_Muscularis (8.49E-06), Esophagus_Mucosa (1.04E-05), Nerve_Tibial (1.16E-05), Adipose_Subcutaneous (1.69E-05), Skin_Not_Sun_Exposed_Suprapubic (2.67E-04), Testis (3.61E-04), Colon_Transverse (4.49E-04), Brain_Hypothalamus (7.26E-04), Lung (7.26E-04), Brain_Nucleus_accumbens_basal_ganglia (7.26E-04), Artery_Aorta (7.95E-04), Whole_Blood (8.59E-04), Brain_Caudate_basal_ganglia (1.15E-03), Cells_Cultured_fibroblasts (1.25E-03), Spleen (2.17E-03), Brain_Spinal_cord_cervical_c-1 (3.33E-03), Ovary (3.68E-03), Brain_Cortex (4.23E-03), Adrenal_Gland (5.70E-03), Brain_Putamen_basal_ganglia (5.70E-03), Pancreas (6.68E-03), Esophagus_Gastroesophageal_Junction (8.82E-03), Brain_Frontal_Cortex_BA9 (9.35E-03), Adipose_Visceral_Omentum (9.35E-03), Prostate (9.59E-03), Artery_Coronary (2.01E-02), Heart_Left_Ventricle (2.37E-02), Liver (4.75E-02)
PIK3R6 and Eosinophil percentage	Thyroid (1.05E-06), Skin_Not_Sun_Exposed_Suprapubic (1.05E-06), Brain_Amygdala (1.05E-06), Lung (1.05E-06), Brain_Cerebellar_Hemisphere (1.05E-06), Cells_EBV-transformed_lymphocytes (1.05E-06), Spleen (1.05E-06), Muscle_Skeletal (1.05E-06), Pancreas (1.05E-06), Brain_Anterior_cingulate_cortex_BA24 (1.05E-06), Whole_Blood (1.05E-06), Brain_Nucleus_accumbens_basal_ganglia (1.05E-06), Brain_Substantia_nigra (1.05E-06), Prostate (1.05E-06), Brain_Caudate_basal_ganglia (1.05E-06), Nerve_Tibial (1.05E-06), Brain_Cerebellum (1.05E-06), Heart_Left_Ventricle (1.05E-06), Small_Intestine_Terminal_Ileum (1.05E-06), Artery_Aorta (1.05E-06), Liver (1.05E-06), Heart_Atrial_Appendage (1.05E-06), Brain_Putamen_basal_ganglia (1.05E-06), Testis (1.17E-06), Adipose_Visceral_Omentum (1.67E-05), Brain_Hypothalamus (3.72E-03)

3.6.2 Supplementary Figures

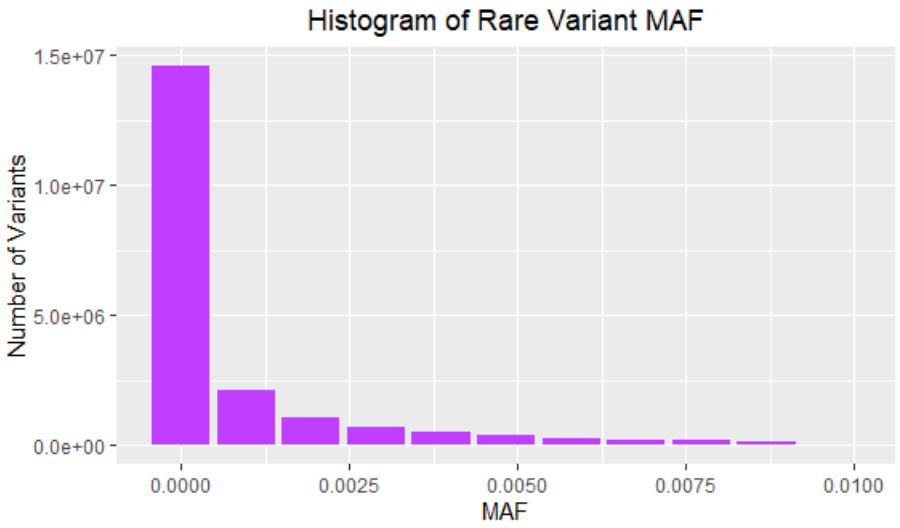


Figure S3.1. Rare Variant Distributions. Histogram of minor allele frequencies of rare variants (MAF < 0.01) in the study.

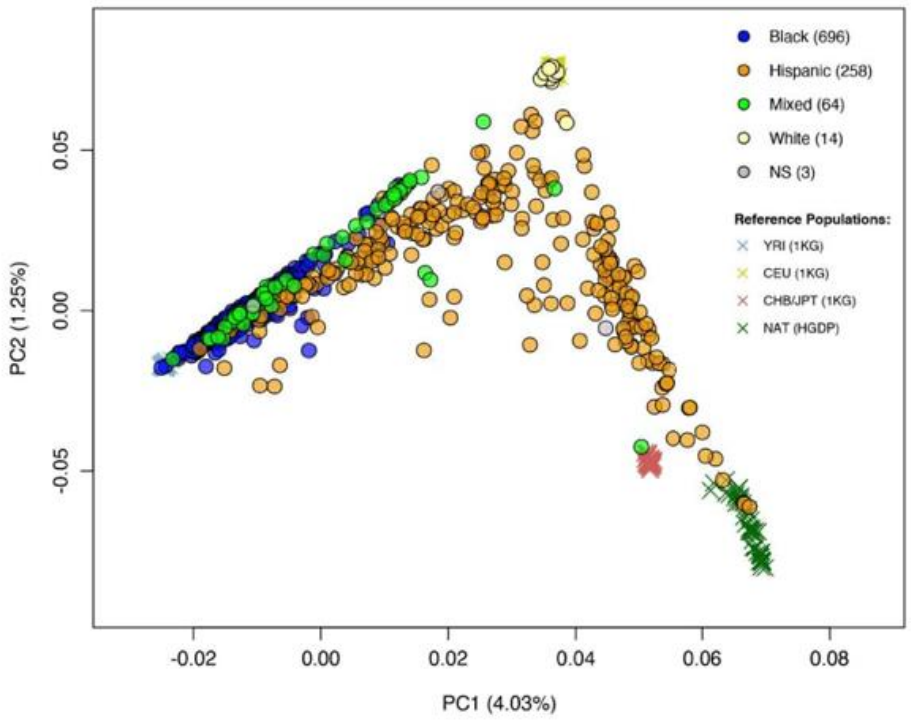
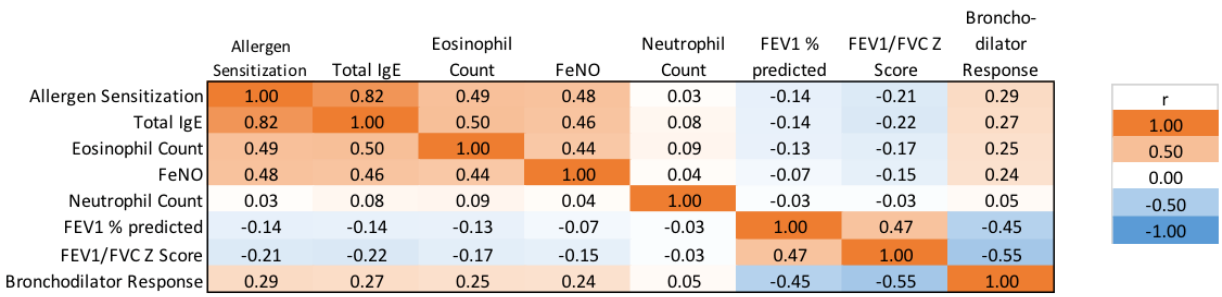


Figure S3.2. Ancestry Composition of Participants. The top two ancestry principal components (PCs) derived from common variants are shown for each individual (circles, colored by self-reported ancestry) or from the four ancestral reference populations (shown by X's). NS: not specified. Figure is taken with permission from Dapas *et al.* 2022 Figure 1A (156).



	Allergen Sensitization	Total IgE	Eosinophil Count	FeNO	Neutrophil Count	FEV1 % predicted	FEV1/FVC Z Score	Broncho-dilator Response
Allergen Sensitization	1.00	0.82	0.49	0.48	0.03	-0.14	-0.21	0.29
Total IgE	0.82	1.00	0.50	0.46	0.08	-0.14	-0.22	0.27
Eosinophil Count	0.49	0.50	1.00	0.44	0.09	-0.13	-0.17	0.25
FeNO	0.48	0.46	0.44	1.00	0.04	-0.07	-0.15	0.24
Neutrophil Count	0.03	0.08	0.09	0.04	1.00	-0.03	-0.03	0.05
FEV1 % predicted	-0.14	-0.14	-0.13	-0.07	-0.03	1.00	0.47	-0.45
FEV1/FVC Z Score	-0.21	-0.22	-0.17	-0.15	-0.03	0.47	1.00	-0.55
Bronchodilator Response	0.29	0.27	0.25	0.24	0.05	-0.45	-0.55	1.00

Figure S3.3. Correlation Between Traits. Spearman correlations are shown between each pair of traits. Allergen sensitization is a count of detectable specific IgEs to 15 common inhaled and food allergens, total IgE is in IU, eosinophil and neutrophil counts are in cells/mm^3, FeNO is in ppb, and bronchodilator response is % change from baseline. A darker orange color corresponds to a more positive correlation and a darker blue color corresponds to a more negative correlation.

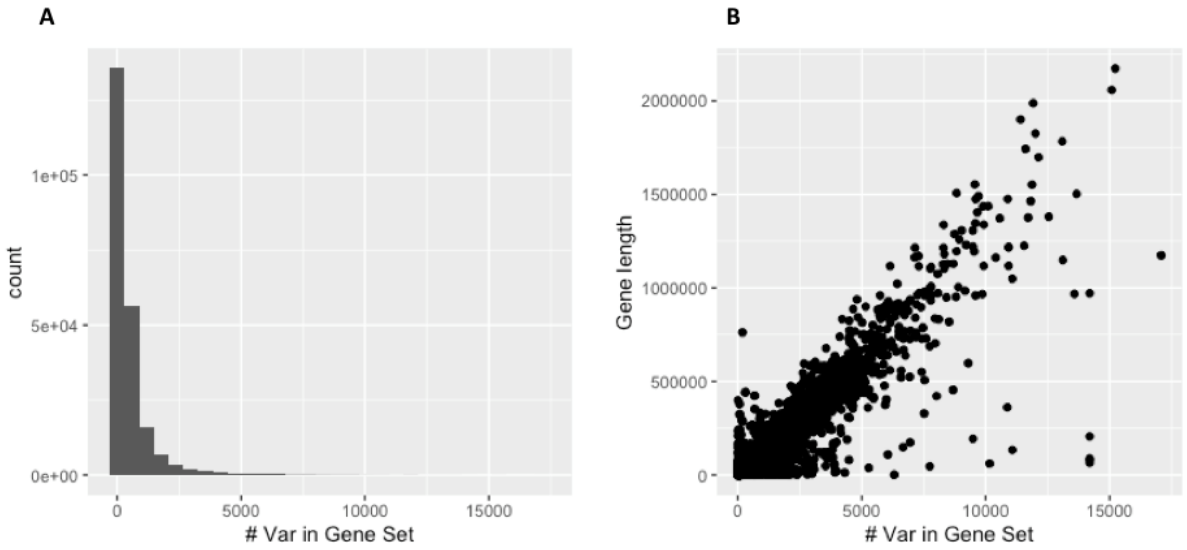
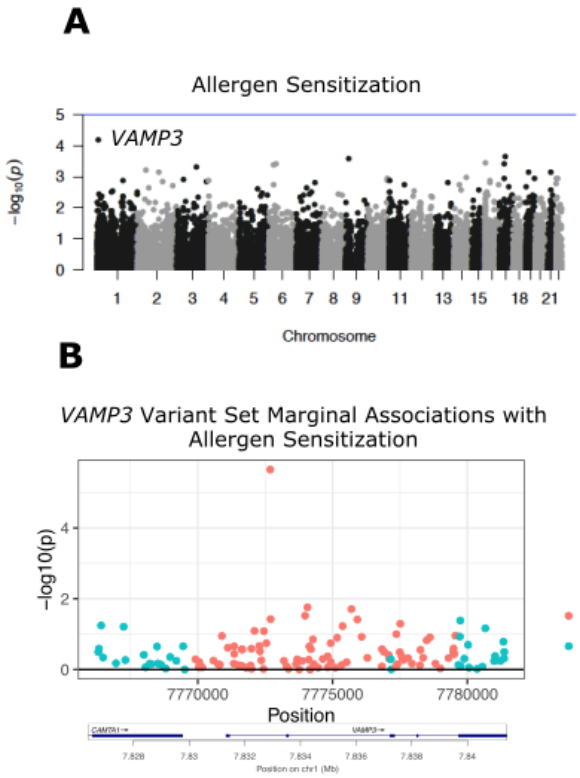


Figure S3.4. Distribution of Variants in Gene Sets. A) Histogram of number of variants in the gene set. **B)** Length of gene plotted against number of variants in the gene set.



C

Trait	*P*-Value
Allergen sensitization	**6.42×10^{-5}**
Total IgE	8.85×10^{-3}
Blood Eosinophil Count	0.19
Blood Neutrophil Count	0.24
FeNO	0.25
FEV_1 % predicted	1.71×10^{-2}
FEV_1/FVC	0.12
Bronchodilator Response	0.81

Figure S3.5. *VAMP3* and Allergen Sensitization. A) Manhattan plot for gene set associations with allergen sensitization. Each point represents a gene and the blue line is suggestive significance (1.00×10^{-5}). **B)** Marginal associations and locations of variants in the *VAMP3* variant set. Variants are color coded green if they are non-coding and orange if coding in either *USF1* or its neighboring gene, *CAMTA1*. **C)** *VAMP3* association with the other traits.

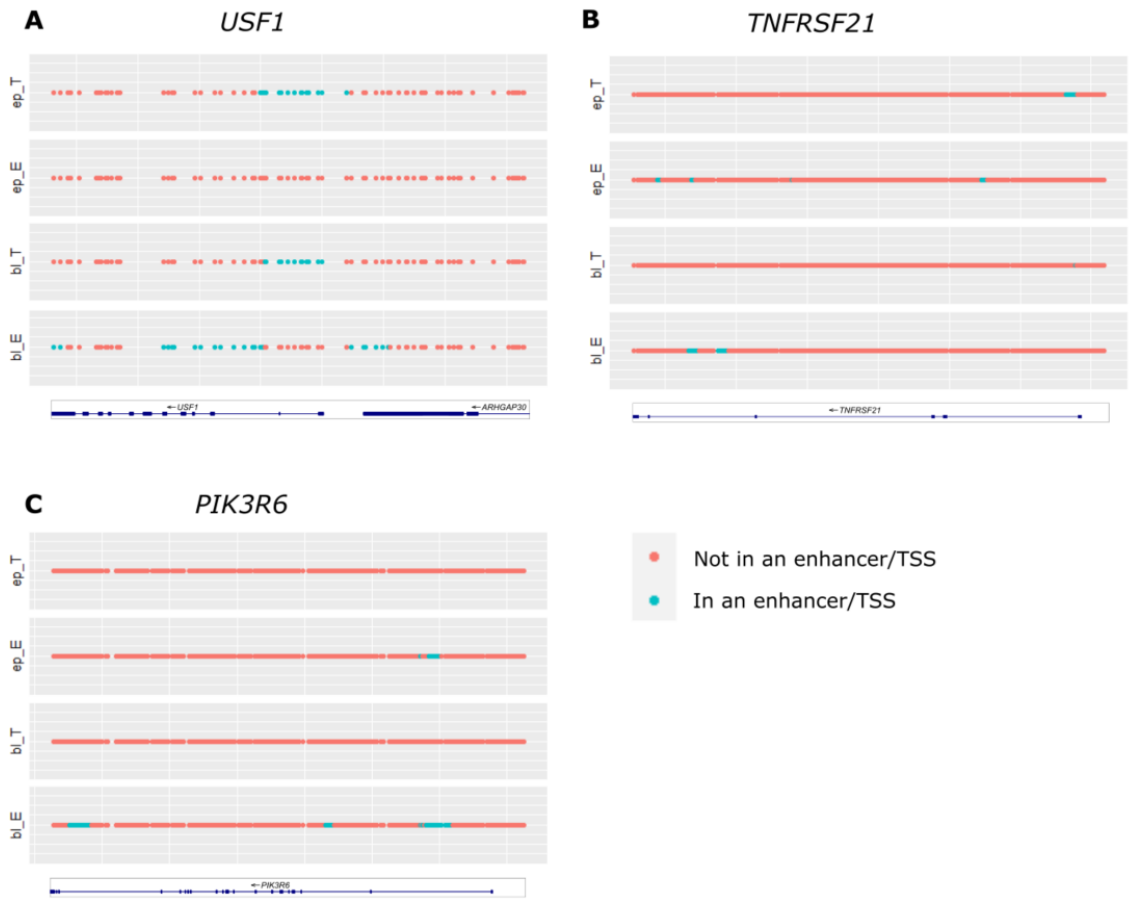


Figure S3.6. ROADMAP Annotations. For **A)** *USF1*, **B)** *TNFRSF21*,and **C)** *PIK3R6*, the rare variants in that gene's variant set are color coded if they do or do not reside in an enhancer or transcription start site (TSS) for epithelial cell TSS (ep_T), epithelial cell enhancer (ep_E), blood cell TSS (bl_T), and blood cell enhancer regions (bl_E).

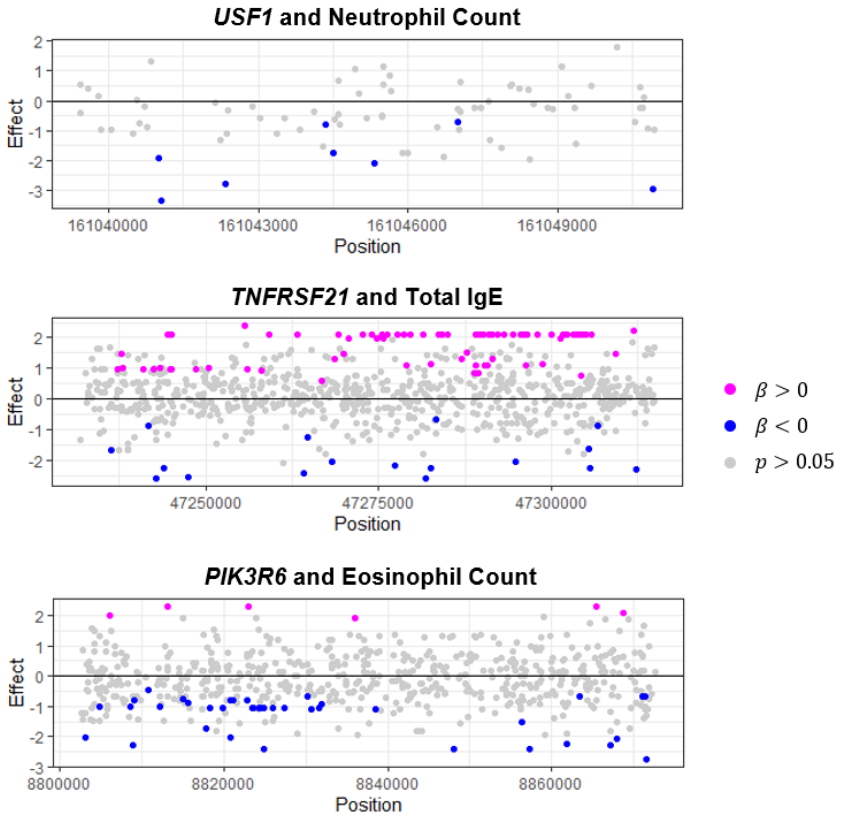


Figure S3.7. Marginal Association Direction of Effect. Effect sizes for the marginal association of variants. Variants that were not marginally significant ($p > 0.05$) are colored gray. Variants with $\beta > 0$ are shown in magenta and variants with $\beta < 0$ are shown in blue.

CHAPTER 4

CONCLUSION

GWAS have successfully identified thousands of statistical associations between common variants and complex traits (116), including over 150 susceptibility loci for asthma (14–19). Despite this, the causal variants, the genes their effects are mediated through, and the relevant biological contexts are not always straightforward to uncover and are therefore unknown for the majority of GWAS loci (194). Here, I build on the results of previous asthma GWAS using complementary approaches and multi-omic datasets. Through genetic fine mapping of GWAS loci and eQTLs to study common variants and applying a gene-based variant set approach to examine rare variants, I demonstrated the utility of these various approaches in identifying novel genes i) at a highly important asthma locus for two subtypes of asthma and ii) for asthma-associated quantitative traits in children of diverse ancestries.

4.1 HLA gene expression and protein coding variation are implicated in childhood- and adult-onset asthma

In chapter 2, I examined one of the most highly associated asthma loci on chromosome 6, which encodes the HLA genes (14–16,30–33). I jointly studied SNPs, HLA amino acids, and HLA alleles to fully capture the genetic variation within this region. Using a Bayesian fine-mapping approach, my study revealed a shared risk variant for both childhood- and adult-onset asthma, as well as a signal unique to childhood-onset asthma in the class I region. However, putatively causal variation in the class II region was not shared. Simulations using existing genotype data in the HLA region also validated that the fine-mapping method SuSiE was able to correctly identify credible sets containing the true causal variants despite the unique complexities of this region.

We assessed the potential regulatory effects of the putatively causal SNPs using gene expression data from asthma-relevant cell types. Two of these datasets were collected from parent-reported Black children (7), and the reduction of LD in individuals of African-ancestries helped with narrowing down the causal variants for gene expression and, by extension, for adult-onset asthma. These findings implicated expression of the highly conserved, yet poorly characterized, *HLA-DQA2* and *HLA-DQB2* genes in the etiology of adult-onset asthma. Protein coding visualization on available crystal structures of the HLA molecules also suggested that the HLA-C p.11 amino acid polymorphism may be specific to COA, whereas the HLA-DQA1*03:01 allele may play a role in adult-onset asthma. T-cell activation and proliferation can be driven by both differential expression levels and protein coding variation in the HLA genes, which can affect binding affinities to different peptides (129,130), and this is a potential mechanism through which these causal signals may impact asthma risk.

Conditional analyses in the class II region showed that adult-onset asthma risk is likely attributable to two causal signals that include the *HLA-DQA1**03:01 allele and variants that impact expression of the *HLA-DQA2* and *HLA-DQB2* genes in immune and epithelial cell types. These conditional studies also supported the finding that there are distinct mechanisms between childhood- and adult-onset asthma with respect in the class II region. Finally, we performed replication analyses to confirm associations of representative variants from each credible set in a meta-analysis of the non-White British cohorts from the UK Biobank. While not all variants were significant at a multiple-testing adjusted significance threshold, they all had the same direction of effect as in the discovery cohort.

Overall, these data suggest that HLA expression and protein coding variation both contribute to asthma risk and extend earlier observations that childhood- and adult-onset asthma

have both shared and distinct genetic risk (15). Many factors distinguish these two subtypes of asthma, including sex ratios, comorbidities of obesity and allergic disease, and the importance of respiratory viral infections in their etiology (139). The study presented here raises the possibility that these dissimilarities may be due, in part, to the differences in causal variation at this locus. This project also used a strategy that can serve as a model for fine mapping other HLA-associated diseases. The results of this study are now published (Clay et al., Fine-mapping studies distinguish genetic risks for childhood- and adult-onset asthma in the HLA region. *Genome Medicine* 14, 55 (2022)) (195).

4.2 Rare variants in genes are associated with asthma-associated atopic and inflammatory phenotypes

In chapter 3, I assessed the contribution of rare variation to asthma-associated quantitative phenotypes in a cohort of children with diverse ancestries. In the U.S., self-reported Black and Hispanic children have a higher prevalence of asthma and more severe symptoms than self-reported White individuals, and increasing representation of relevant populations in genetic studies is crucial in addressing these disparities (196,197). I used a gene-based variant set tests to examine associations with traits reflecting the main features of asthma. Only one association (*USF1* and blood neutrophil count) was significant after multiple-test correction with over 24,000 genes and eight traits. The most significant associations for each of the traits were all nominally significant when just considering just the non-coding variation in the gene set, whereas just three were nominally significant when just considering coding variation. This further supports the many previous findings showing that complex traits are primarily driven by noncoding variation, both common and rare (168–170).

PhenomeXcan and publicly available mouse knockout data served as complementary approaches to validate the most significant associations for each trait, and they supported three gene associations (*USF1* and blood neutrophil count, *PIK3R6* and blood eosinophil count, and *TNFRSF21* and total IgE). In each of these cases, the burden test had the smallest p-values in comparison to the SKAT and ACAT-V variant set tests, suggesting that most variants had the same direction of effect on their trait. Overall, we identified novel genes associated with atopic and inflammatory asthma-associated traits, and the validation results point to the robustness of these findings across biological context and species.

4.3 Future Directions

In this thesis, I identified putatively causal variants and genes for childhood- and adult-onset asthma while highlighting the importance of rare variation in discovering novel genes associated with asthma-relevant phenotypes. However, there are some natural follow-up questions that accompany each of these studies. I propose the following directions.

First, chapters 2 and 3 both demonstrate the importance in including diverse populations in genetic studies. In chapter 2, we performed eQTL mapping and fine mapping in self-reported Black individuals, allowing us to further narrow down putatively causal variants for *HLA-DQA2* and *HLA-DQB2* expression due to the lower LD. We also found associations with quantitative phenotypes in children of diverse ancestries in chapter 3. Despite asthma being a global disease, the majority of genetic studies of asthma have been done in populations of European-ancestries, and increasing representation of all affected populations in genetic studies is crucial to fully understanding its genetic architecture. Current global, large-scale efforts to increase the diversity of cohorts in genetic studies are underway (18,198–202). Future studies of asthma should continue to include increased numbers of diverse cohorts, as uncovering shared and population-

specific causal variation will lead to a more comprehensive understanding of asthma and have more potential impact on global health in general and on global health inequities in particular.

Second, what is noticeably absent from this thesis is a more direct examination of the interactions between genetics and environmental factors in asthma. Epidemiological data have demonstrated the importance of the environment in asthma, including effects of allergens, pollution, early respiratory infections, and smoking, among others (203). Asthma is therefore an excellent model for studying gene-environment interactions. While the APIC and URECA cohorts had several environmental exposure variables available at the time of study, the method we used to study variant sets has not been extended to allow for interaction effects. Future steps could be to apply other methods that allow for the examination of environmental interactions in variant set test association studies (204). Additionally, if available, environmental factors measured prior to the onset of asthma should also be studied to have a better sense of their causality and effect on the course of the disease.

Third, we demonstrate the benefit of studying specific asthma subtypes. We found that while the risk loci identified in GWAS at the HLA class I and class II loci overlapped between childhood- and adult-onset asthma, the causal variants and genes differed, allowing us to find subtype-specific risk variants and genes. Future genetic studies should continue moving away using a broad phenotype (e.g. a general "asthma" diagnosis) and shift towards finding causal variants and genes for specific subtypes. The GWAS in Pividori *et al.* demonstrated that more loci can be identified in childhood-onset asthma despite the significantly smaller sample size than adult-onset asthma (15). This suggests that focusing on more homogeneous subtypes increases power beyond that afforded by larger and larger sample sizes (81). As previously described in chapter 1.1, childhood-onset asthma itself is an umbellar term for many different

subtypes that capture differential occurrences of atopy, wheeze, and lung function, among other clinical factors (6,8). Improving stratification of individuals in genetic analyses can paint a clearer picture of the shared and specific risk variants and genes. Identification of subtype-specific risks can ultimately improve personalized care for a wider range of individuals.

Additionally, current efforts have largely focused on searching and interpreting lead association signals, such as the loci encoding the *GSDMB*/*ORMDL3* (28,205,206), *IL33* (29), and now the HLA genes. However, as with most complex traits, heritability is not driven by a single gene but through a highly polygenic background (169), and future directions can continue to take on a more genome-wide approach, such as using models that account for the polygenic nature of this disease. For example, others have used polygenic risk scores (PRS) and have linked low vs. high polygenic risk groups to various cellular programs (207) or correlated them with gene expression to identify "core genes" for the trait of interest (208). These types of studies may be able to be extended to asthma.

Finally, it is crucial that we continue to elucidate the biological mechanisms through which the genetic variants and genes impact disease risk. While we identified putatively causal genes in chapters 2 and 3, more efforts are needed to directly link them with disease-relevant cellular programs and the biological networks that are perturbed to understand their causal effect on asthma. The growing availability of genomic, transcriptomic, epigenomic, metabolomic and proteomic annotations from a wide range of disease-relevant cell types, in addition to increased ease in gene editing in cells and animal models can move towards better understanding the causal relationships between genetic variation with the altered biological mechanisms and resulting progression towards asthma (207,209,210). Uncovering all of these layers can move towards ultimately understanding the core processes in asthma.

4.4 Concluding remarks

In this thesis, I describe research studies that bridge the gap between GWAS and gene discovery. By focusing on the HLA region, I show the utility of dissecting a robustly associated asthma locus and integrating multiple ‘omic datatypes towards uncovering causal variants and genes for two major asthma subtypes. This thesis also illustrates the benefits of studying rare variation and using orthogonal data types in finding novel genes for asthma-associated quantitative traits. Overall, the discoveries made in these combined studies, which apply cutting-edge statistical methods and diverse ‘omic datasets, improve our understanding of the genetic architecture of this complex trait and can point to potential therapeutic targets in the treatment of asthma.

CHAPTER 5

Bibliography

1. Network GA. The Global Asthma Report [Internet]. Auckland, New Zealand; 2018. Available from: www.globalasthmanetwork.org
2. Lizzo JM, Cortes S, Doerr C. Pediatric Asthma (Nursing). StatPearls. 2022 May 8;
3. Morosco G, Kiley J. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. J Allergy Clin Immunol. 2007 Nov;120(5):S93.
4. Pongracic JA, Krouse RZ, Babineau DC, Zoratti EM, Cohen RT, Wood RA, et al. Distinguishing characteristics of difficult-to-control asthma in inner-city children and adolescents. J Allergy Clin Immunol. 2016 Oct;138(4):1030–41.
5. Lötvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: A new approach to classification of disease entities within the asthma syndrome. J Allergy Clin Immunol. 2011 Feb 1;127(2):355–60.
6. Zoratti EM, Krouse RZ, Babineau DC, Pongracic JA, O'Connor GT, Wood RA, et al. Asthma phenotypes in inner-city children. J Allergy Clin Immunol. 2016 Oct 1;138(4):1016–29.
7. Gern JE, Visness CM, Gergen PJ, Wood RA, Bloomberg GR, O'Connor GT, et al. The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. BMC Pulm Med. 2009 Dec 8;9(1):17.
8. Bacharier LB, Beigelman A, Calatroni A, Jackson DJ, Gergen PJ, O'Connor GT, et al. Longitudinal Phenotypes of Respiratory Health in a High-Risk Urban Birth Cohort. Am J Respir Crit Care Med. 2019 Jan;199(1):71–82.
9. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of Asthma Phenotypes Using Cluster Analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med. 2010 Feb 2;181(4):315.
10. Thomsen SF, Van Der Sluis S, Kyvik KO, Skytthe A, Backer V. Estimates of asthma heritability in a large twin sample. Clin Exp Allergy. 2010 Mar 12;40(7):1054–61.
11. Thomsen SF, Duffy DL, Kyvik KO, Backer V. Genetic influence on the age at onset of asthma: a twin study. J Allergy Clin Immunol. 2010 Sep 1;126(3):626–30.
12. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of Asthma and Hay Fever in Australian Twins. Am Rev Respir Dis. 1990 Dec 17;142(6_pt_1):1351–8.
13. Nieminen MM, Kaprio J, Koskenvuo M. A Population-based Study of Bronchial Asthma in Adult Twin Pairs. Chest. 1991 Jul 1;100(1):70–5.
14. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet. 2018 Jan 22;50(1):42–53.

15. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med. 2019 Sep 26;7(6):509–22.

16. Olafsdottir TA, Theodors F, Bjarnadottir K, Bjornsdottir US, Agustsdottir AB, Stefansson OA, et al. Eighty-eight variants highlight the role of T cell regulation and airway remodeling in asthma pathogenesis. Nat Commun. 2020;11(1):393.

17. Ferreira MAR, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. Am J Hum Genet. 2019 Apr 4;104(4):665–84.

18. Daya M, Rafaels N, Brunetti TM, Chavan S, Levin AM, Shetty A, et al. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. Nat Commun. 2019 Dec 1;10(1):1–13.

19. Tsuo K, Zhou W, Wang Y, Kanai M, Namba S, Gupta R, et al. Multi-ancestry meta-analysis of asthma identifies novel associations and highlights the value of increased power and diversity. medRxiv. 2022 Sep 23;2021.11.30.21267108.

20. Shrine N, Portelli MA, John C, Soler Artigas M, Bennett N, Hall R, et al. Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. Lancet Respir Med. 2019 Jan 1;7(1):20–34.

21. Granada M, Wilk JB, Tuzova M, Strachan DP, Weidinger S, Albrecht E, et al. A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study. J Allergy Clin Immunol. 2012 Mar;129(3):840-845.e21.

22. Bønnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. Nat Genet. 2013 Aug 30;45(8):902–6.

23. Joost O, Wilk JB, Adrienne Cupples L, Harmon M, Shearman AM, Baldwin CT, et al. Genetic Loci Influencing Lung Function. Am J Respir Crit Care Med. 2002 Mar 15;165(6):795–9.

24. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai H-J, Liu X, et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. Nat Commun. 2015 Feb 24;6:6304.

25. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. Nat Genet. 2018 Aug 16;50(8):1072–80.

26. Paternoster L, Standl M, Chen C-M, Ramasamy A, Bønnelykke K, Duijts L, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2012 Feb 25;44(2):187–92.

27. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012 Sep;337(6099):1190–5.

28. Ober C, Mckennan CG, Magnaye KM, Altman MC, Washington C, Stanhope C, et al. Expression quantitative trait locus fine mapping of the 17q12 – 21 asthma locus in African American children : a genetic association and gene expression study. Lancet Respir Med. 2020;8(20):1–11.

29. Aneas I, Decker DC, Howard CL, Sobreira DR, Sakabe NJ, Blaine KM, et al. Asthma-associated genetic variants induce IL33 differential expression through an enhancer-blocking regulatory region. Nat Commun 2021 121. 2021 Oct 21;12(1):1–12.

30. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017 Dec;49(12):1752–7.

31. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010 Sep 23;363(13):1211–21.

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

33. Lasky-Su J, Himes BE, Raby BA, Klanderman BJ, Sylvia JS, Lange C, et al. HLA-DQ strikes again: Genome-wide association study further confirms HLA-DQ in the diagnosis of asthma among adults. Clin Exp Allergy. 2012 Dec;42(12):1724–33.

34. Trowsdale J, Knight JC. Major Histocompatibility Complex Genomics and Human Disease. Annu Rev Genomics Hum Genet. 2013 Aug 31;14(1):301–23.

35. NIH. Histocompatibility complex [Internet]. Genetics Home Reference. 2019 [cited 2019 Mar 19]. Available from: https://ghr.nlm.nih.gov/primer/genefamily/hla

36. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. Nat Rev Immunol. 2018 Jan 2;18(5):325–39.

37. Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens. 2004 Dec 1;64(6):631–49.

38. Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. Yonsei Med J. 2007 Feb 28;48(1):11–23.

39. Xie M, Li J, Jiang T. Accurate HLA type inference using a weighted similarity graph. BMC Bioinformatics. 2010 Dec 14;11(SUPPL. 11):1–10.

40. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015 Jan;43(Database issue):D423-31.

41. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol. 2017;18(1):76.

42. Margolis DJ, Mitra N, Kim B, Gupta J, Hoffstad OJ, Papadopoulos M, et al. Association of HLA-DRB1 genetic variants with the persistence of atopic dermatitis. Hum Immunol.

2015 Aug;76(8):571–7.

43. Movahedi M, Moin M, Gharagozlou M, Aghamohammadi A, Dianat S, Moradi B, et al. Association of HLA class II alleles with childhood asthma and Total IgE levels. Iran J Allergy Asthma Immunol. 2008 Dec;7(4):215–20.

44. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V., Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015 Aug 27;47(9):1091–8.

45. Aguiar VRC, César J, Delaneau O, Dermitzakis ET, Meyer D. Expression estimation and eQTL mapping for HLA genes with a personalized pipeline. Absher DM, editor. PLOS Genet. 2019 Apr 22;15(4):e1008091.

46. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. Tang J, editor. PLoS One. 2013 Jun 6;8(6):e64683.

47. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG—HLA genotype imputation with attribute bagging. Pharmacogenomics J. 2014 Apr 28;14(2):192–200.

48. Dilthey AT, Moutsianas L, Leslie S, McVean G. HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. Bioinformatics. 2011 Apr 1;27(7):968–72.

49. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature. 2015 Sep 30;526(7571):68–74.

50. Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans. J Hum Genet. 2021 Jan 1;66(1):11.

51. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutat Res. 2007 Feb 3;615(1–2):28–56.

52. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet. 2009 Feb;5(2).

53. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2010 Feb;34(2):188–93.

54. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet. 2008 Sep 12;83(3):311–21.

55. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011 Jul 15;89(1):82–93.

56. Li X, Li Z, Zhou H, Gaynor SM, Liu Y, Chen H, et al. Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-

genome sequencing studies at scale. Nat Genet. 2020 Aug 24;35(9):57.

57. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012 Aug 10;91(2):224–37.

58. DeWan AT, Egan KB, Hellenbrand K, Sorrentino K, Pizzoferrato N, Walsh KM, et al. Whole-exome sequencing of a pedigree segregating asthma. BMC Med Genet. 2012 Oct 9;13.

59. Smith D, Helgason H, Sulem P, Bjornsdottir US, Lim AC, Sveinbjornsson G, et al. A rare IL33 loss-of-function mutation reduces blood eosinophil counts and protects from asthma. PLoS Genet. 2017 Mar 1;13(3).

60. Morin A, Madore AM, Kwan T, Ban M, Partanen J, Rönnblom L, et al. Exploring rare and low-frequency variants in the Saguenay–Lac-Saint-Jean population identified genes associated with asthma and allergy traits. Eur J Hum Genet. 2019 Jan 1;27(1):90.

61. Cameron-Christie S, Mackay A, Wang Q, Olsson H, Angermann B, Lassi G, et al. A broad exome study of the genetic architecture of asthma reveals novel patient subgroups. bioRxiv. 2020 Dec 11;2020.12.10.419663.

62. Chang D, Hunkapiller J, Bhangale T, Reeder J, Mukhyala K, Tom J, et al. A whole genome sequencing study of moderate to severe asthma identifies a lung function locus associated with asthma risk. Sci Rep. 2022 Dec 1;12(1).

63. Campbell CD, Mohajeri K, Malig M, Hormozdiari F, Nelson B, Du G, et al. Whole-genome sequencing of individuals from a founder population identifies candidate genes for asthma. PLoS One. 2014 Aug 12;9(8).

64. Igartua C, Myers RA, Mathias RA, Pino-Yanes M, Eng C, Graves PE, et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. Nat Commun. 2015;6.

65. Chakraborty S, Dakle P, Sinha A, Vishweswaraiah S, Nagori A, Salimath S, et al. Genetic variations in olfactory receptor gene OR2AG2 in a large multigenerational family with asthma. Sci Rep. 2019 Dec 1;9(1).

66. Bogari NM, Amin AA, Rayes HH, Abdelmotelb A, Taher MM, Al-Allaf FA, et al. Next Generation Exome Sequencing of Pediatric Asthma Identifies Rare and Novel Variants in Candidate Genes. Dis Markers. 2021;2021.

67. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. Nat 2021 5997886. 2021 Oct 18;599(7886):628–34.

68. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nat 2021 5907845. 2021 Feb 10;590(7845):290–9.

69. Torgerson DG, EJ A, GY C, WJ G, CR G, PE G, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat

Genet. 2011 Sep;43(9):887–92.

70. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet. 2014 Jan 17;46(1):51–5.

71. Kelsen SG, Agache IO, Soong W, Israel E, Chupp GL, Cheung DS, et al. Astegolimab (anti-ST2) efficacy and safety in adults with severe asthma: A randomized clinical trial. J Allergy Clin Immunol. 2021 Sep;148(3):790–8.

72. Donkel SJ, Portilla Fernández E, Ahmad S, Rivadeneira F, van Rooij FJA, Ikram MA, et al. Common and Rare Variants Genetic Association Analysis of Circulating Neutrophil Extracellular Traps. Front Immunol. 2021 Feb 24;12.

73. Yamada M, Motoike IN, Kojima K, Fuse N, Hozawa A, Kuriyama S, et al. Genetic loci for lung function in Japanese adults with adjustment for exhaled nitric oxide levels as airway inflammation indicator. Commun Biol. 2021 Dec 1;4(1).

74. Drake KA, Torgerson DG, Gignoux CR, Galanter JM, Roth LA, Huntsman S, et al. A genome-wide association study of bronchodilator response in Latinos implicates rare variants. J Allergy Clin Immunol. 2014;133(2).

75. Mousas A, Ntritsos G, Chen MH, Song C, Huffman JE, Tzoulaki I, et al. Rare coding variants pinpoint genes that control human hematological traits. PLoS Genet. 2017 Aug 1;13(8).

76. Yang T, Jackson VE, Smith A V., Chen H, Bartz TM, Sitlani CM, et al. Rare and low-frequency exonic variants and gene-by-smoking interactions in pulmonary function. Sci Rep. 2021 Dec 1;11(1).

77. Ortega VE, Li X, O'Neal WK, Lackey L, Ampleford E, Hawkins GA, et al. The Effects of Rare SERPINA1 Variants on Lung Function and Emphysema in SPIROMICS. Am J Respir Crit Care Med. 2020 Mar 1;201(5):540–54.

78. Lee EY, Mak ACY, Hu D, Sajuthi S, White MJ, Keys KL, et al. Whole-Genome Sequencing Identifies Novel Functional Loci Associated with Lung Function in Puerto Rican Youth. Am J Respir Crit Care Med. 2020 Oct 1;202(7):962–72.

79. Soriano JB, Kendrick PJ, Paulson KR, Gupta V, Abrams EM, Adedoyin RA, et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Respir Med. 2020 Jun 1;8(6):585–96.

80. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct 10;562(7726):203–9.

81. Schoettler N, Rodríguez E, Weidinger S, Ober C. Advances in asthma and allergic disease genetics: Is bigger always better? J Allergy Clin Immunol. 2019 Dec 1;144(6):1495–506.

82. Simmonds M, Gough S. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. Curr Genomics. 2009 Feb 14;8(7):453–65.

83. Mosaad YM. Clinical Role of Human Leukocyte Antigen in Health and Disease. Scand J Immunol. 2015 Oct 1;82(4):283–306.

84. Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. Vol. 22, Clinical Microbiology Reviews. American Society for Microbiology Journals; 2009. p. 370–85.

85. Band G, Marchini J. BGEN: A binary file format for imputed genotype and haplotype data. bioRxiv. bioRxiv; 2018. p. 308296.

86. Motyer A, Vukcevic D, Cortes A, McVean G, Leslie S. Imputation of classical HLA types from UK Biobank genotype data. 2016.

87. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine- mapping studies identify susceptibility loci for multiple common infections. Nat Commun. 2017;8.

88. Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, et al. Modernizing reference genome assemblies. PLoS Biol. 2011 Jul;9(7).

89. Genome Reference Consortium [Internet]. Available from: https://www.ncbi.nlm.nih.gov/grc

90. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Jul 8;26(17):2190.

91. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Ser B (Statistical Methodol. 2020 Jul 10;82(5):1273–300.

92. Cusanovich DA, Caliskan M, Billstrand C, Michelini K, Chavarria C, De Leon S, et al. Integrated analyses of gene expression and genetic association studies in a founder population. Hum Mol Genet. 2016 May 15;25(10):2104–12.

93. Altman MC, Whalen E, Togias A, O'Connor GT, Bacharier LB, Bloomberg GR, et al. Allergen-induced activation of natural killer cells represents an early-life immune response in the development of allergic asthma. J Allergy Clin Immunol. 2018 Dec;142(6):1856–66.

94. Altman MC, Calatroni A, Ramratnam S, Jackson DJ, Presnell S, Rosasco MG, et al. Endotype of allergic asthma with airway obstruction in urban children. J Allergy Clin Immunol. 2021;

95. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. A complete tool set for molecular QTL discovery and analysis. Nat Commun. 2017 May 18;8(1):1–7.

96. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012 Jun 17;44(7):821–4.

97. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc. 1995;57:289–300.

98. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–42.

99. Kaur G, Gras S, Mobbs JI, Vivian JP, Cortes A, Barber T, et al. Structural and regulatory diversity shape HLA-C protein expression levels. Nat Commun. 2017 Jun 26;8.

100. Jiang W, Birtley JR, Hung SC, Wang W, Chiou SH, Macaubas C, et al. In vivo clonal expansion and phenotypes of hypocretin-specific CD4+ T cells in narcolepsy patients and controls. Nat Commun. 2019 Dec 1;10(1):1–17.

101. Tollefsen S, Hotta K, Chen X, Simonsen B, Swaminathan K, Mathews II, et al. Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. J Biol Chem. 2012 Apr 20;287(17):13611–9.

102. Schrödinger L, DeLano W. The PyMOL Molecular Graphics System [Internet]. Schrodinger, LLC; 2020. Available from: http://www.pymol.org/pymol

103. HLA Nomenclature [Internet]. 2019. Available from: http://hla.alleles.org/nomenclature/naming.html

104. Benner C, Havulinna AS, Järvelin MR, Salomaa V, Ripatti S, Pirinen M. Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. Am J Hum Genet. 2017 Oct 5;101(4):539–51.

105. Westra H-J, Martínez-Bonet M, Onengut-Gumuscu S, Lee A, Luo Y, Teslovich N, et al. Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. Nat Genet. 2018;50(10):1366–74.

106. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018;50(11):1505–13.

107. Cruz-Tapias P, Castiblanco J, Anaya J-M. HLA Association with Autoimmune Diseases. In: Anaya J-M, Shoenfeld Y, Rojas-Villarraga A, Roger A. Levy, Cervera R, Editors., editors. Autoimmunity: From Bench to Bedside. Bogota (Colombia): El Rosario University Press; 2013.

108. Jin Y, Roberts GHL, Ferrara TM, Ben S, Van Geel N, Wolkerstorfer A, et al. Early-onset autoimmune vitiligo associated with an enhancer variant haplotype that upregulates class II HLA expression. Nat Commun. 2019;10(1):391.

109. Raj P, Rai E, Song R, Khan S, Wakeland BE, Viswanathan K, et al. Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. Elife. 2016 Feb 15;5(e12089).

110. Apps R, Qi Y, Carlson JM, Chen H, Gao X, Thomas R, et al. Influence of HLA-C expression level on HIV control. Science (80-). 2013 Apr 5;340(6128):87–91.

111. D'Antonio M, Reyna J, Jakubosky D, Donovan MKR, Bonder MJ, Matsui H, et al. Systematic genetic analysis of the MHC region reveals mechanistic underpinnings of HLA type associations with disease. Elife. 2019 Nov 1;8(e48476).

112. Gutierrez-Arcelus M, Baglaenko Y, Arora J, Hannes S, Luo Y, Amariuta T, et al. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. Vol. 52, Nature Genetics. Nature Research; 2020. p. 247–53.

113. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002 May 16;12(6):996–1006.

114. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57–74.

115. Petersen J, Kooy-Winkelaar Y, Loh KL, Tran M, van Bergen J, Koning F, et al. Diverse T Cell Receptor Gene Usage in HLA-DQ8-Associated Celiac Disease Converges into a Consensus Binding Solution. Structure. 2016 Oct 4;24(10):1643–57.

116. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017 Jan 4;45(D1):D896–901.

117. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017 Oct 11;550(7675):204–13.

118. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell. 2018 Nov 29;175(6):1701-1715.e16.

119. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–29.

120. Blais ME, Dong T, Rowland-Jones S. HLA-C as a mediator of natural killer and T-cell activation: Spectator or key player? Vol. 133, Immunology. Wiley-Blackwell; 2011. p. 1–7.

121. Fadda L, Borhis G, Ahmed P, Cheent K, Pageon S V., Cazaly A, et al. Peptide antagonism as a mechanism for NK cell activation. Proc Natl Acad Sci U S A. 2010 Jun 1;107(22):10160–5.

122. Karimi K, Forsythe P. Natural killer cells in asthma. Vol. 4, Frontiers in Immunology. Frontiers Media SA; 2013. p. 159.

123. Kim JH, Jang YJ. Role of natural killer cells in airway inflammation. Vol. 10, Allergy, Asthma and Immunology Research. Korean Academy of Asthma, Allergy and Clinical Immunology; 2018. p. 448–56.

124. Li X, Ampleford EJ, Howard TD, Moore WC, Torgerson DG, Li H, et al. Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. J Allergy Clin Immunol. 2012 Oct;130(4):861-8.e7.

125. Berdoz J, Tiercy J-M, Rollini P, Mach B, Gorski J. Remarkable sequence conservation of the HLA-DQB2 locus (DX beta) within the highly polymorphicDQ subregion of the human MHC. Immunogenetics. 1989;29(4):241–8.

126. Gaur LK, Heise ER, Thurtle PS, Nepom GT. Conservation of the HLA-DQB2 locus in nonhuman primates. J Immunol. 1992 Feb 1;148(3):943–8.

127. Lenormand C, Bausinger H, Gross F, Signorino-Gelo F, Koch S, Peressin M, et al. HLA-DQA2 and HLA-DQB2 Genes Are Specifically Expressed in Human Langerhans Cells and Encode a New HLA Class II Molecule. J Immunol. 2012 Apr 15;188(8):3903–11.

128. Yan Q, Forno E, Herrera-Luis E, Pino-Yanes M, Yang G, Oh S, et al. A genome-wide association study of asthma hospitalizations in adults. J Allergy Clin Immunol. 2021;147(3):933–40.

129. Mine KL, Tedesco-Silva H, Mourão TB, Campos EF, Salzedas LA, Aguiar B, et al. Heightened expression of HLA-DQB1 and HLA-DQB2 in pre-implantation biopsies predicts poor late kidney graft function. Hum Immunol. 2018 Aug 1;79(8):594–601.

130. Farina, Picascia, Pisapia, Barba, Vitale, Franzese, et al. HLA-DQA1 and HLA-DQB1 Alleles, Conferring Susceptibility to Celiac Disease and Type 1 Diabetes, are More Expressed Than Non-Predisposing Alleles and are Coordinately Regulated. Cells. 2019 Jul 19;8(7):751.

131. Pirinen M, Donnelly P, Spencer CCA. Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. Ann Appl Stat. 2013;7(1):369–90.

132. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016 May 15;32(10):1493–501.

133. Banerjee S, Zeng L, Schunkert H, Söding J. Bayesian multiple logistic regression for case-control GWAS. PLoS Genet. 2018 Dec 1;14(12):e1007856.

134. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020 Sep 1;369(6509):1318.

135. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat Genet. 2019 Oct 1;51(10):1494–505.

136. Zhang S, Zhang H, Zhou Y, Qiao M, Zhao S, Kozlova A, et al. Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. Science (80-). 2020 Jul 31;369(6503):561–5.

137. Gourraud P-A, Khankhanian P, Cereb N, Yang SY, Feolo M, Maiers M, et al. HLA Diversity in the 1000 Genomes Dataset. Colombo GI, editor. PLoS One. 2014 Jul 2;9(7):e97282.

138. Luo Y, Kanai M, Choi W, Li X, Yamamoto K, Ogawa K, et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ethnic fine-mapping in HIV host response. medRxiv. 2020;

139. Trivedi M, Denton E. Asthma in children and adults—what are the differences and what can they tell us about asthma? Vol. 7, Frontiers in Pediatrics. Frontiers Media S.A.; 2019. p. 256.

140. Weitkamp L, Ober C. Ancestral and recombinant 16-locus HLA haplotypes in the Hutterites. Immunogenetics. 1999;49(6):491–7.

141. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Methods. 2015 Nov;12(11):1061–3.

142. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15–21.

143. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014 Feb 3;15(2):R29.

144. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47–e47.

145. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul;25(14):1754–60.

146. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43(SUPL.43).

147. Dilthey AT, Mentzer AJ, Carapito R, Cutland C, Cereb N, Madhi SA, et al. HLA*LA-HLA typing from linearly projected graph alignments. Bioinformatics. 2019;35(21):4394–6.

148. McKennan C, Nicolae D. Estimating and Accounting for Unobserved Covariates in High-Dimensional Correlated Data. J Am Stat Assoc. 2020;

149. Altman M, Ramratnam S, Jackson D, Presnell S, Gergen P, Bacharier L, et al. Airway Epithelium Gene Expression Endotyping of Asthma and Airway Obstruction in Urban Children [abstract]. J Allergy Clin Immunol. 2020;145(2):AB176.

150. Kim KW, Ober C. Lessons Learned From GWAS of Asthma. Allergy Asthma Immunol Res. 2019 Mar 1;11(2):170.

151. Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. Clin Transl Immunol. 2017 Dec;6(12):e165.

152. Liu J, Deng Y, Yu B, Mo B, Luo L, Yang J, et al. Targeted resequencing showing novel common and rare genetic variants increases the risk of asthma in the Chinese Han population. J Clin Lab Anal. 2021 Jun 1;35(6).

153. Gergen PJ, Teach SJ, Togias A, Busse WW. Reducing Exacerbations in the Inner City: Lessons from the Inner-City Asthma Consortium (ICAC). J Allergy Clin Immunol Pract. 2016 Jan 1;4(1):22–6.

154. O'Connor GT, Lynch S V., Bloomberg GR, Kattan M, Wood RA, Gergen PJ, et al. Early-life home environment and risk of asthma among inner-city children. J Allergy Clin Immunol. 2018 Apr 1;141(4):1468–75.

155. Recommendations for standardized procedures for the on-line and off-line measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide in adults and children-1999. This official statement of the American Thoracic Society was adopted by the . Am J Respir Crit Care Med. 1999;160(6):2104–17.

156. Dapas M, Thompson EE, Wentworth-Sheilds W, Clay S, Visness CM, Calatroni A, et al. Multi-omic association study implicates PPP1R13B in DNA methylation-mediated genotype and smoking exposure effects on decreased lung function in urban children. medRxiv. 2022 Jun 27;2022.06.24.22276830.

157. Munafò MR, Tilling K, Taylor AE, Evans DM, Smith GD. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol. 2018 Feb 1;47(1):226.

158. Reich D, Nalls MA, Kao WHL, Akylbekova EL, Tandon A, Patterson N, et al. Reduced Neutrophil Count in People of African Descent Is Due To a Regulatory Variant in the Duffy Antigen Receptor for Chemokines Gene. PLOS Genet. 2009 Jan;5(1):e1000360.

159. Auer PL, Reiner AP, Leal SM. The effect of phenotypic outliers and non-normality on rare-variant association testing. Eur J Hum Genet. 2016;24(8).

160. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet. 2012 Nov 2;91(5):839–48.

161. Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. Am J Hum Genet. 2019 Mar 7;104(3):410–21.

162. Pividori M, Rajagopal PS, Barbeira A, Liang Y, Melia O, Bastarache L, et al. PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci Adv. 2020 Sep 1;6(37).

163. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Anagnostopoulos A, et al. Mouse Genome Database (MGD) 2019. Nucleic Acids Res [Internet]. 2019 Jan 8;47(D1):D801–6. Available from: https://pubmed.ncbi.nlm.nih.gov/30407599/

164. Hoekstra M, Ren B, Laurila PP, Hildebrand RB, Soronen J, Frodermann V, et al. Hematopoietic upstream stimulating factor 1 deficiency is associated with increased atherosclerosis susceptibility in LDL receptor knockout mice. Sci Reports 2021 111. 2021 Aug 12;11(1):1–11.

165. Zhao H, Yan M, Wang H, Erickson S, Grewal IS, Dixit VM. Impaired c-Jun amino terminal kinase activity and T cell differentiation in death receptor 6-deficient mice. J Exp Med. 2001 Nov 19;194(10):1441–8.

166. Deladeriere A, Gambardella L, Pan D, Anderson KE, Hawkins PT, Stephens LR. The regulatory subunits of PI3Kγ control distinct neutrophil responses. Sci Signal. 2015 Jan 20;8(360).

167. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. Am J Hum Genet. 2014 Jul 7;95(1):5.

168. Ma M, Ru Y, Chuang LS, Hsu NY, Shi LS, Hakenberg J, et al. Disease-associated variants in different categories of disease located in distinct regulatory elements. BMC Genomics. 2015 Jun 18;16(Suppl 8):S3.

169. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177–86.

170. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet. 2015 Oct 15;24(R1):R102–10.

171. USF1 upstream transcription factor 1 [Homo sapiens (human)] - Gene - NCBI [Internet]. Available from: https://www.ncbi.nlm.nih.gov/gene/7391

172. Ruuth M, Soronen J, Kaiharju E, Merikanto K, Perttilä J, Metso J, et al. USF1 deficiency alleviates inflammation, enhances cholesterol efflux and prevents cholesterol accumulation in macrophages. Lipids Health Dis. 2018 Dec 13;17(1).

173. Fan YM, Hernesniemi J, Oksala N, Levula M, Raitoharju E, Collings A, et al. Upstream Transcription Factor 1 (USF1) allelic variants regulate lipoprotein metabolism in women and USF1 expression in atherosclerotic plaque. Sci Reports 2014 41. 2014 Apr 11;4(1):1–8.

174. Kristiansson K, Ilveskoski E, Lehtimäki T, Peltonen L, Perola M, Karhunen PJ. Association Analysis of Allelic Variants of USF1 in Coronary Atherosclerosis. Arterioscler Thromb Vasc Biol. 2008 May;28(5):983.

175. Reiner AP, Carlson CS, Jenny NS, Durda JP, Siscovick DS, Nickerson DA, et al. USF1 gene variants, cardiovascular risk, and mortality in European Americans: Analysis of two US cohort studies. Arterioscler Thromb Vasc Biol. 2007 Dec 1;27(12):2736–42.

176. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019 Jun 27;570(7762):514.

177. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell. 2016 Nov 17;167(5):1415-1429.e19.

178. Popa SC, Shin JA. The Intrinsically Disordered Loop in the USF1 bHLHZ Domain Modulates Its DNA-Binding Sequence Specificity in Hereditary Asthma. J Phys Chem B. 2019 Nov;123(46):9862–71.

179. Schedel M, Michel S, Gaertner VD, Toncheva AA, Depner M, Binia A, et al. Polymorphisms related to ORMDL3 are associated with asthma susceptibility, alterations in transcriptional regulation of ORMDL3, and changes in TH2 cytokine levels. J Allergy Clin Immunol. 2015 Oct 1;136(4):893-903.e14.

180. PIK3R6 phosphoinositide-3-kinase regulatory subunit 6 [Homo sapiens (human)] - Gene - NCBI [Internet]. Available from: https://www.ncbi.nlm.nih.gov/gene/146850

181. Yun JH, Chase R, Parker MM, Saferali A, Castaldi PJ, Silverman EK, et al. Peripheral blood gene expression signatures of eosinophilic chronic obstructive pulmonary disease.

Am J Respir Cell Mol Biol. 2019 Aug 30;61(3):398–401.

182. Elena-Pérez S, Heredero-Jung DH, García-Sánchez A, Estravís M, Martin MJ, Ramos-González J, et al. Molecular Analysis of IL-5 Receptor Subunit Alpha as a Possible Pharmacogenetic Biomarker in Asthma. Front Med. 2021 Feb 11;7:1148.

183. Jiang Y, Gruzieva O, Wang T, Forno E, Boutaoui N, Sun T, et al. Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents. Eur Respir J. 2019 May 1;53(5):1900102.

184. TNFRSF21 TNF receptor superfamily member 21 [Homo sapiens (human)] - Gene - NCBI [Internet]. Available from: https://www.ncbi.nlm.nih.gov/gene/27242

185. Pan G, Bauer JH, Haridas V, Wang S, Liu D, Yu G, et al. Identification and functional characterization of DR6, a novel death domain-containing TNF receptor. FEBS Lett. 1998 Jul 24;431(3):351–6.

186. Venkataraman C, Justen K, Zhao J, Galbreath E, Na S. Death receptor-6 regulates the development of pulmonary eosinophilia and airway inflammation in a mouse model of asthma. Immunol Lett. 2006 Jul 15;106(1):42–7.

187. Liu J, Na S, Glasebrook A, Fox N, Solenberg PJ, Zhang Q, et al. Enhanced CD4+ T Cell Proliferation and Th2 Cytokine Production in DR6-Deficient Mice. Immunity. 2001 Jul 1;15(1):23–34.

188. Caminati M, Pham D Le, Bagnasco D, Canonica GW. Type 2 immunity in asthma. World Allergy Organ J. 2018 Jan 1;11(1):13.

189. MRPL44 mitochondrial ribosomal protein L44 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2022 Sep 2]. Available from: https://www.ncbi.nlm.nih.gov/gene/65080

190. Mishima S, Sakamoto M, Kioka H, Nagata Y, Suzuki R. Multifunctional regulation of VAMP3 in exocytic and endocytic pathways of RBL-2H3 cells. Front Immunol. 2022 Aug 5;13.

191. Sandford AJ, Shirakawa T, Moffat MF, Daniels SE, Faux JA, Young RP, et al. Localisation of atopy and beta subunit of high-affinity IgE receptor (Fc epsilon RI) on chromosome 11q. Lancet (London, England). 1993 Feb 6;341(8841):332–4.

192. Weidinger S, Gieger C, Rodriguez E, Baurecht H, Mempel M, Klopp N, et al. Genome-Wide Scan on Total Serum IgE Levels Identifies FCER1A as Novel Susceptibility Locus. PLOS Genet. 2008 Aug;4(8):e1000166.

193. Dar SA, Rai G, Ansari MA, Akhter N, Gupta N, Sharma S, et al. FcεR1α gene polymorphism shows association with high IgE and anti-FcεR1α in Chronic Rhinosinusitis with Nasal Polyposis. J Cell Biochem. 2018 May 1;119(5):4142–9.

194. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. Am J Hum Genet. 2018 May;102(5):717–30.

195. Clay SM, Schoettler N, Goldstein AM, Carbonetto P, Dapas M, Altman MC, et al. Fine-mapping studies distinguish genetic risks for childhood- and adult-onset asthma in the

HLA region. Genome Med. 2022;14(1):55.

196. Asthma Facts | AAFA.org [Internet]. [cited 2022 Nov 17]. Available from: https://www.aafa.org/asthma-facts/

197. González Burchard E, Borrell LN. Need for Racial and Ethnic Diversity in Asthma Precision Medicine. N Engl J Med. 2021 Dec 9;385(24):2297–8.

198. Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, et al. Overview of the BioBank Japan Project: Study design and profile. J Epidemiol. 2017 Mar;27(3S):S2–8.

199. Lin J-C, Fan C-T, Liao C-C, Chen Y-S. Taiwan Biobank: making cross-database convergence possible in the Big Data era. Gigascience. 2018 Jan;7(1):1–4.

200. Al Kuwari H, Al Thani A, Al Marri A, Al Kaabi A, Abderrahim H, Afifi N, et al. The Qatar Biobank: background and methods. BMC Public Health. 2015;15(1):1208.

201. Sohail M, Chong A, Quinto-Cortes C, Palma-Martínez M, Ragsdale A, Medina-Muñoz S, et al. Nationwide genomic biobank in Mexico unravels demographic history and complex trait architecture from 6,057 individuals. bioRxiv; 2022.

202. Mulder N, Abimiku A, Adebamowo SN, de Vries J, Matimba A, Olowoyo P, et al. H3Africa: current perspectives. Pharmgenomics Pers Med. 2018;11:59–66.

203. Potaczek DP, Harb H, Michel S, Alhamwe BA, Renz H, Tost J. Epigenetics and allergy: From basic mechanisms to clinical applications. Epigenomics. 2017 Apr 1;9(4):539–71.

204. Lin X, Lee S, Wu MC, Wang C, Chen H, Li Z, et al. Test for Rare Variants by Environment Interactions in Sequencing Association Studies on JSTOR. Biometrics. 2016;72(1):156–64.

205. Washington C, Dapas M, Biddanda A, Magnaye KM, Aneas I, Helling BA, et al. African-specific alleles modify risk for asthma at the 17q12-q21 locus in African Americans. Genome Med. 2022;14(1):112.

206. Raby BA, Weiss ST. Diversity and the splice of life: Mapping the 17q12–21.1 locus for variants associated with early-onset Asthma in African American individuals. Am J Respir Crit Care Med. 2021 Feb 15;203(4):401–3.

207. Laber S, Strobel S, Mercader J-M, Dashti H, Ainbinder A, Honecker J, et al. Discovering cellular programs of intrinsic and extrinsic drivers of metabolic traits using LipocyteProfiler. bioRxiv. 2021 Jan 1;2021.07.17.452050.

208. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet 2021 539. 2021 Sep 2;53(9):1300–10.

209. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. 2016 Dec 15;167(7):1853-1866.e17.

210. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. Science (80-). 2021 Sep 24;373(6562):1464–8.