frontiers in Allergy

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/falgy.2021.738741

Translational Analysis of Moderate to Severe Asthma GWAS Signals Into Candidate Causal Genes and Their Functional, Tissue-Dependent and Disease-Related Associations

Michael A. Portelli[1†], Kamini Rakkar[1†], Sile Hu[2], Yike Guo[2], Ian M. Adcock[3] on behalf of the U-BIOPRED Study Group and Ian Sayers[1]*

[1] Centre for Respiratory Research, Translational Medical Sciences, School of Medicine, National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham University Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom, [2] Data Science Institute, Imperial College London, London, United Kingdom, [3] The National Heart and Lung Institute, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:
Reinoud Gosens,
University of Groningen, Netherlands

Reviewed by:
Lena Uller,
Lund University, Sweden
Jinming Gao,
Peking Union Medical College Hospital (CAMS), China

***Correspondence:**
Kamini Rakkar
kamini.rakkar2@nottingham.ac.uk

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Asthma, a section of the journal Frontiers in Allergy

Received: 09 July 2021
Accepted: 06 September 2021
Published: 18 October 2021

Citation:
Portelli MA, Rakkar K, Hu S, Guo Y, Adcock IM and Sayers I (2021) Translational Analysis of Moderate to Severe Asthma GWAS Signals Into Candidate Causal Genes and Their Functional, Tissue-Dependent and Disease-Related Associations. Front. Allergy 2:738741. doi: 10.3389/falgy.2021.738741

Asthma affects more than 300 million people globally and is both under diagnosed and under treated. The most recent and largest genome-wide association study investigating moderate to severe asthma to date was carried out in 2019 and identified 25 independent signals. However, as new and in-depth downstream databases become available, the translational analysis of these signals into target genes and pathways is timely. In this study, unique (U-BIOPRED) and publicly available datasets (HaploReg, Open Target Genetics and GTEx) were investigated for the 25 GWAS signals to identify 37 candidate causal genes. Additional traits associated with these signals were identified through PheWAS using the UK Biobank resource, with asthma and eosinophilic traits amongst the strongest associated. Gene expression omnibus dataset examination identified 13 candidate genes with altered expression profiles in the airways and blood of asthmatic subjects, including *MUC5AC* and *STAT6*. Gene expression analysis through publicly available datasets highlighted lung tissue cell specific expression, with both *MUC5AC* and *SLC22A4* genes showing enriched expression in ciliated cells. Gene enrichment pathway and interaction analysis highlighted the dominance of the *HLA-DQA1/A2/B1/B2* gene cluster across many immunological diseases including asthma, type I diabetes, and rheumatoid arthritis. Interaction and prediction analyses found *IL33* and *IL18R1* to be key co-localization partners for other genes, predicted that *CD274* forms co-expression relationships with 13 other genes, including the *HLA-DQA1/A2/B1/B2* gene cluster and that *MUC5AC* and *IL37* are co-expressed. Drug interaction analysis revealed that 11 of the candidate genes have an interaction with available therapeutics. This study provides significant insight into these GWAS signals in the context of cell expression, function, and disease relationship with the view of informing future research and drug development efforts for moderate-severe asthma.

Keywords: GWAS, causal genes, eQTL, SNP, moderate to severe asthma

INTRODUCTION

Asthma is one the of most predominant non-communicable diseases throughout the world. It accounts for over 400,000 deaths per year and by World Health Organization estimates, 24.8 million disability adjusted life years (DALYS) were attributable to Asthma in 2016 (1). Although there is no cure for asthma, most symptoms can be managed well with medication. However, patients with severe asthma, which represent ∼4% of all patients, suffer from uncontrolled symptoms and frequent exacerbations despite medication (2).

Over the years, many genome and phenome wide association studies (GWAS, PheWAS) have been completed resulting in a large number of signals mainly *via* single nucleotide polymorphisms (SNPs) associated with asthma relevant traits (3–7), which have been reviewed in detail (8, 9). GWAS and PheWAS (10–12) have greatly advanced asthma research and translating these signals into candidate causal genes is the next step in moving to greater mechanistic understanding of asthma, therapeutic targets and pathways for investigation. However, many single SNPs are largely defined and generally mapped to the closest gene, regardless of whether the SNP has any effect on the function of that gene e.g., expression. This ignores the complexity of the 3D architecture of DNA which may result in SNPs linearly far away from a gene being closer than thought in the 3D structure and having a functional role (13). Indeed, it has been shown that over 90% of disease-linked variants are located in non-coding regions of the genome (14). Therefore, accurately determining causal genes is important for understanding the biology underlying GWAS and PheWAS signals, especially within respiratory relevant tissues and compartments, including the lung. An additional layer of complexity is the possibility of specific SNP-tissue interactions, where SNP-gene regulation may occur differently in different tissues and environments (15, 16).

In this study, we investigated the 25 signals (including four novel signals; rs11603634, rs1090584, rs10905284, and rs61816761) identified in the recent Moderate to Severe Asthma GWAS (7). We aimed to identify candidate causal genes from these signals and understand their association with asthma, highlighting potential targets for downstream investigation utilizing unique [Unbiased biomarkers for the prediction of respiratory disease outcomes (U-BIOPRED)] and publicly available datasets. We then investigated these genes *via* multiple gene enrichment and interaction analyses and gene expression analysis in the lung and in whole blood. This study represents a significant advance in our understanding of the mechanistic underpinnings of these signals and provides candidate genes and pathways for future drug development in moderate-severe asthma.

MATERIALS AND METHODS

Signal to Trait Association Analysis—PheWAS Analysis From GeneATLAS UK Biobank Data

To investigate traits associated with the signals, 25 signals taken from the Moderate to Severe Asthma GWAS were individually searched for in the online GeneATLAS database (17) according to the risk allele. This database of associations between traits and variants uses the UK Biobank cohort (18). It has 778 traits and therefore with Bonferroni correction an adjusted *P*-value of 6.42×10^{-5} was used to determine significance. Proxies were used for the following signals which were not present in the database: rs61816761 (rs61816766, $R^2 = 0.50$), rs367983479 (rs1504215, $R^2 = 0.85$), rs71266076 (rs7824993, $R^2 = 0.81$), rs7305461 (rs1131017, $R^2 = 0.75$), rs112502960 (rs62076439, $R^2 = 1.0$), rs61840192 (rs1031163, $R^2 = 1.0$), rs560026225 (rs72687036, $R^2 = 0.66$), and rs776111176 (rs3997872, $R^2 = 0.82$). Displayed traits were selected and broadly grouped into asthma, blood/immune cell, allergic, other respiratory, other inflammatory, and auto-immune categories. An overview of the project pipeline is provided in **Figure 1**.

Signal to Gene Association Analysis

To determine likely signal to gene associations in lung tissue, T-cells and whole blood, we interrogated publicly available databases (GTEx V8, Open Target Genetics (OTG), HaploReg V4.1, LD Matrix 5.0), a unique dataset presenting expression quantitative trait loci (eQTLs) for bronchial epithelial brushes ($n = 117$), bronchial biopsies ($n = 84$), and nasal epithelial brushes ($n = 75$) (U-BIOPRED), as well as a general literature search for previously published associations. Using these datasets, we noted eQTLs [GTEx, OTG (19), HaploReg (20), U-BIOPRED], sQTLs (GTEx), variant 2 gene scores (OTG), Gene to eQTL localization (OTG), and tagged functional variation [LDLink (LDProxy) (21)] for the sentinel SNP reported for these signals in the original study or, in cases where the sentinel SNP was not covered, a proxy SNP. We used a corrected significance threshold of $P < 0.05$, as presented in each dataset, to identify eQTLs across all datasets, and a H4 > 0.8 threshold for the Gene to eQTL co-localization. Proxy SNPs were selected based on linkage disequilibrium R^2 scores from the publicly available reference haplotypes for the EUR super population (CEU+GBR+TSI+FIN+IBS) from 1000_Phase 3 (Version 5) of the 1,000 Genomes Project accessed through LDLink, where the selected proxy was the SNP with the highest R^2 value (minimum $R^2 = 0.2$) that was present in the interrogated database.

Each signal to gene association identified was given a score point, with greater weighting given to U-BIOPRED eQTLs due to the specificity of the eQTLs to the respiratory related samples. This generated a total score, where genes scoring $s \geq 3$ were selected for further downstream analysis. In cases where no genes achieved $s \geq 3$ for a particular signal, the gene with the highest available score was selected. In the case of signal rs61816761, where all candidate genes scored below 3, *FLG* was selected as the most likely candidate causal gene due to existing literature (22, 23).

Gene Function and Disease Association Analysis

Gene function was determined using the Gene 2 Function (24) and UniProt (25) databases, and gene function and biological processes annotations were listed.

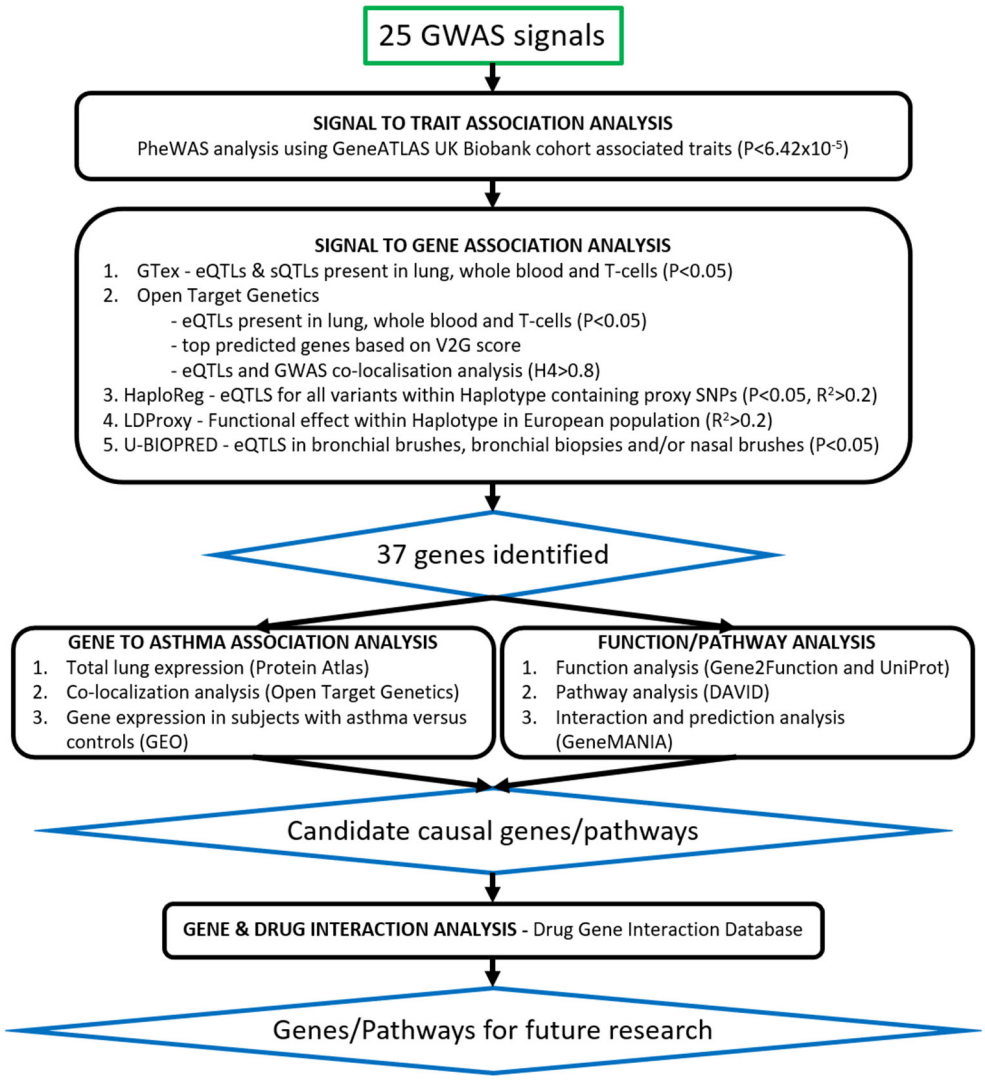


FIGURE 1 | Translational analysis pipeline. Twenty-five signals from the Moderate to Severe Asthma GWAS were taken through signal to trait and signal to gene association analyses. Signals were searched for in publicly available databases such as GeneATLAS, PheWAS, Open Target Genetics, GTex, HaploReg and LDLink as well as our U-BIOPRED (UB) datasets of Bronchial Biopsies, Bronchial Brushes, and Nasal Brushes. Gene association was scored as per **Table 3** and 37 genes were identified. Candidate causal genes were then searched for in gene expression datasets from subjects with asthma vs. controls and single-cell-RNA lung tissue datasets. Asthma related pathways and interactions between genes were identified using the DAVID Functional Annotation Tool and GeneMANIA prediction server and drug gene interactions were identified using the Drug Gene Interaction Database.

Total and Single Cell RNA Expression Analysis in Lung Tissue

To determine the candidate gene expression profile in lung tissue, the Human Protein Atlas (HPA, http://www.proteinatlas.org) resource was used (26). For total lung expression, normalized expression (NX) values were used. The HPA provides NX values through normalizing the average transcript per million (TPM) value of all individual lung samples from three transcriptomics datasets (HPA, GTEx, and FANTOM5). For single cell (sc) lung tissue gene expression, transcripts per million protein coding genes (pTPM) values were used. The HPA provides pTPM values through normalization of read counts from four lung tissue datasets (Single Cell Expression Atlas, the Human Cell Atlas, the Gene Expression Omnibus (GEO) and the European Genome-phenome Archive). Both NX and pTPM values were obtained for 36 candidate genes (*FLG* had no data available). $\log_{10}$(pTPM) values were used for the heatmap. Genes which showed enriched/enhanced expression for lung epithelial (blue bars) or blood/immune cells (red bars) were highlighted in the bar chart. Enriched/enhanced expression was determined by using the same scoring system as the HPA. Enriched/enhanced expression is defined by the HPA as having NX values at least four times of any other tissue/cell type from their full transcriptomics data of 37 tissues and 43 single cell types.

Co-localization Analysis of Tissue Gene Expression and the Reported Trait "Asthma"

To identify evidence linking gene expression across tissue/cell types with the asthma trait, the OTG portal was used. H4 scores for studies where asthma was the exclusive trait were obtained for the 37 candidate genes focusing on lung tissue and/or blood/immune cell types. Median H4 scores of all variants (signals) were plotted for each gene. Genes *AAGAB, CD247, DEXI, FLG, GATA3, HLA-DQA1/A2, B2, IL33, KIF1A, KIAA1109, MSL1, MUC5AC, PGAP3*, and *ZBTB10* had no association data for lung/blood/immune cells/tissue and therefore have not been included in the analysis.

Differential mRNA Expression Analyses in Airway Epithelial Cells Isolated From Subjects With Asthma or Control Subjects

Genes which showed enriched epithelial expression in the HPA database and/or colocalization in lung tissue (median H4 > 0.8) were selected for further investigation to explore mRNA expression in bronchial epithelial cells isolated from subjects with asthma or control subjects. The freely available U-BIOPRED gene expression dataset, GSE43696, was used from the GEO resource (27), as it was the largest adult dataset available with a control group. The GSE43696 dataset comprises of Agilent Human GE 4×44K V2 Gene Expression data of bronchial epithelial cells for 20 control, 50 mild-moderate, and 38 severe asthma subjects (28, 29). In this dataset, mild-moderate asthma was defined as subjects having an FEV_1 of >60% predicted, with/without low-moderate dose inhaled corticosteroids. Severe asthma was defined as subjects having severe refractory asthma, including the continuous use of high-dose inhaled corticosteroids and/or frequent use of oral corticosteroids with continuing symptoms and/or chronic airflow limitation. Control subjects had normal lung function, no history of respiratory disease or recent respiratory infection and no evidence of bronchial hyperresponsiveness. Expression values were plotted for each gene stratified by subject group. Data was statistically analyzed using a Kruskal-Wallis test with a two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli to control the FDR at 5%. $P < 0.05$ was considered significant.

Differential mRNA Expression Analyses in Blood Cells Isolated From Subjects With Asthma or Control Subjects

Candidate genes which showed enriched blood cell expression in the HPA database or colocalization in blood (median H4 > 0.8) were selected for further investigation to explore mRNA expression in blood cells between asthma and control subjects. The freely available gene expression dataset with adult subjects which also included a control group, GSE69683, was used from the GEO resource (27). The GSE69683 dataset comprises of Affymetrix HT HG-U133+ PM Array gene expression data of blood cell samples with 87 control, 77 mild-moderate, and 246 severe asthma subjects from the U-BIOPRED cohort (30). Samples were selected based on their non-smoker status (<5 pack-year smoking history). Mild-moderate asthma was defined as having controlled or partially controlled asthma symptoms, as defined by the Global Initiative for Asthma (GINA), whilst receiving a dose of <500μg fluticasone propionate/day or equivalent. Severe asthma was defined as having uncontrolled symptoms defined according to GINA guidelines and/or frequent exacerbations (more than two per year) despite high-dose inhaled corticosteroids (ICS) (ICS ≥ 1,000 μg fluticasone propionate/day or equivalent dose). Control subjects had no history of asthma or wheeze, had no other chronic respiratory disease, had never smoked and their prebronchodilator FEV1 was ≥ 80% predicted. Expression values were plotted for each gene stratified by subject group. Data was analyzed using the Shapiro-Wilk test to test for normality before using the Kruskal-Wallis test (non-normal distribution) or Welch's ANOVA (normal distribution) to analyze differences between groups. Only genes GSDMB and STAT6 had normally distributed data. Both statistical tests were combined with a two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli to control the FDR at 5%. $P < 0.05$ was considered significant.

Gene Functional Annotation Analysis

To identify gene or gene clusters in functional and disease related associations, the DAVID Bioinformatics Resource (31, 32) was used. The 37 candidate genes underwent enrichment analysis in the Genetic Association Database of complex diseases and disorders (GAD), Gene Ontology (GO) Term and KEGG and REACTOME pathway databases. The 37 genes were used as a set and clusters below 5% FDR were considered significant. The top 10 GAD terms were listed and fold enrichment (FE) scores were plotted against -$\log_{10}$(P-value) with the area of points representing the number of genes in the cluster for GO Terms and pathway analysis.

Gene Interaction Analysis

To identify and predict candidate causal gene interactions, the GeneMANIA prediction server was used which normalizes and weights interaction networks from various sources of data to build an interaction map (33). The 37 candidate genes were inputted as a set to produce a network map depicting physical interactions, co-localization, co-expression and predicted genes and interactions. Scores for predicted genes and the number of gene interactions were used to identify the strongest predicted genes.

Gene Drug Interaction Analysis

We used The Drug Gene Interaction Database (DGIdb) (34) to identify known drug interactions with the 37 genes highlighted by our signal to gene analysis and 5 predicted genes (*CD274, IL37, IRAK4, PDCD1LG2,* and *ZAP70*) which were found to have the strongest interactions through the gene interaction analysis. Drugs showing an interaction score, i.e., the numeric representation of publication count and source count, the ratio of average known gene partners for all drugs to the known partners for the given drug, and the ratio of average known drug partners for all genes to the known partners for the given gene, of >1.0 were selected. In addition we queried each gene at

TABLE 1 | The 25 signals from the moderate to severe asthma GWAS.

SNP	Closest gene reported in paper	Risk allele	Non-risk allele	Risk allele frequency in European population
rs61816761	FLG	A	G	2.37%
rs7523907	CD247	T	C	54.08%
rs12479210	IL1RL1	T	C	38.73%
rs34290285	D2HGDH	G	A	74.26%
rs560026225*	KIAA1109	GATT	G	23.60%
rs1837253	TSLP	C	T	74.16%
rs1438673	WDR36	C	T	50.78%
rs3749833	C5orf56	C	T	26.08%
rs1986009	RAD50	A	C	18.71%
rs9273410	HLA-DQB1	A	C	55.30%
rs776111176*	HLA-DQA1	A	AAT	14.85%
rs367983479*	BACH2	CA	C	61.50%
rs71266076*	MIR5708	C	CT	36.93%
rs144829310	IL33	T	G	16.40%
rs10905284	GATA3	C	A	42.94%
rs11603634	MUC5AC	G	A	50.36%
rs7936312	C11orf30	T	G	47.42%
rs7305461	RPS26	A	C	44.61%
rs703816	STAT6	C	T	43.41%
rs10519068	RORA	G	A	87.25%
rs72743461	SMAD3	A	C	23.60%
rs7203459	CLEC16A	T	C	75.44%
rs2941522	IKZF3	T	C	48.29%
rs112502960	ZNF652	A	G	35.92%
rs61840192	LOC101928272	G	A	57.30%

The 25 signals with the original reported gene (closest), asthma and non-risk allele and asthma risk allele frequency are listed. All analyses conducted in this study are based on the risk allele.
**Insertion and deletions (Indels).*

clinicaltrials.gov in order to identify drugs currently in Phase II, III, and IV trials for our identified genes as targets.

RESULTS

PheWAS Analysis Indicates Signals Are Associated With Blood Cell Counts, Asthma, and Other Inflammatory Disorders

The GeneATLAS database identified traits associated with the risk allele of the 25 signals or proxies from the Moderate to Severe Asthma GWAS (**Tables 1**, **2**). As expected, all signals showed association with asthma with odds ratios (OR) >1 (**Figure 2**). All signals, apart from rs61816761 (*via* proxy rs61816766), showed association with a blood/immune cell traits particularly eosinophil levels but also lymphocyte and neutrophil counts.

Signal rs11603634 was relatively specific in showing only asthma, eosinophil percentage and platelet count associations. All signals apart from rs61816761 and rs9273410 showed association with eosinophil number and/or percentage highlighting the role of this cell type in asthma. All signals apart from rs61816761, rs3749833, rs10905284, rs11603634, rs112502960, and rs776111176 showed an association with allergy. Of these, signals rs3749833, rs10905284, rs11603634, rs112502960, and rs776111176 were associated with eosinophils but not allergy. Conversely rs9273410 was the only signal which showed an association with allergy but not eosinophils. These nuances in associations highlight the different asthma phenotypes.

Signals, rs1986009, rs9273410, rs7936312, rs776111176, and rs2941522 showed broad associations across many trait and trait groups. Signals rs9273410, rs776111176, and rs72743461 were the only ones associated with insulin-dependent diabetes. Signal rs776111176 was the only one associated with rheumatoid arthritis.

Some signals showed protective associations (OR < 1, black dots) for other disease traits such as rs2941522 (inflammatory bowel disease), rs72743461 (diabetes), rs1986009 (psoriasis) rs776111176 (inflammatory bowel diseases, ulcerative colitis, psoriasis, hyperthyroidism, multiple sclerosis, sarcoidosis), and rs9273410 (inflammatory bowel diseases, ulcerative colitis, psoriasis and multiple sclerosis). A full list of traits and ORs and *P*-values are available in **Supplementary Table 1**.

Signal to Gene Analysis Identifies 37 Candidate Causal Genes Associated With the 25 Moderate-Severe Association Signals

Using our scoring system for identifying candidate causal genes (**Table 3**) we identified 33 moderate to strong signal:gene relationships ($s \geq 3$) for 21/25 signals (**Table 4**). The four remaining signals presented with weaker evidence of a relationship ($s \leq 2$).

Relationships with the greatest score were observed for the signals tagged by rs12479210 (*IL1RL1*) and rs34290285 (*D2HGDH*) (score = 13) and multiple genes were identified for signals rs12479210, rs34290285, rs1438673, rs3749833, rs1986009, rs7305461, rs72743461, rs7203459, rs776111176, rs2941522, and rs112502960. We identified novel candidate causal signal:gene relationships for SNPs rs3749833, rs1986009, rs71266076, rs7936312, and rs2941522.

The strongest eQTLs, as defined by the most significant combined *P* and B-values for the risk allele, were observed in whole blood for rs2941522 (*GSDMB* & *ORMDL3*) and rs7305461 (*RPS26* & *STAT6*) (**Figure 3**). Weaker associations (*P*-value) presenting with the greatest observed effect (B-value) were observed for rs776111176 in whole blood, lung (*HLA-DQA2*) and for signal rs9273410 in bronchial brushes (*HLA-DQB1*) and in whole blood and lung (*HLA-DQB2*) (**Figure 3**).

Gene Function and Biological Processes Analysis Identifies Distinct Gene Groups

The largest gene function clusters were observed for genes involved in DNA/RNA binding (*BACH2, GATA3, IRF1, RORA, RPS26, SMAD3, STAT6*, and

TABLE 2 | SNPs utilized in pipeline analyses.

Variant	LD block /base pairs	Number of SNPs in LD (R^2 >0.1)	Proxy used in other analyses (MAF)	R^2	UBIOPRED Bronchial Biopsy (MAF)	R^2	UBIOPRED Bronchial Brush (MAF)	R^2	UBIOPRED Nasal Brush (MAF)	R^2
rs61816761	988,882	20	rs61816766 (C, 2%)	0.50	rs61816764 (T, 5%)	0.21	-	-	-	-
rs7523907	41,739	56	rs3108155 (G, 39%)	0.83	rs2056625 (A, 36%)	0.73	-	-	-	-
rs12479210	458,394	578	-	-	rs2270298 (G, 25%)	0.47	rs2241116 (A, 20%)	0.46	-	-
rs34290285	69,955	157	-	-	rs34077392 (C, 37%)	0.43	-	-	-	-
rs560026225	622,215*	150*	rs72687036 (G, 24%)	0.66	-	-	-	-	-	-
rs1837253	71,462	29	-	-	-	-	-	-	-	-
rs1438673	227,233	353	-	-	rs2289277 (G, 46%)	0.73	-	-	rs7524421 (G, 24%)	1.00
rs3749833	526,932	325	-	-	-	-	-	-	rs11748326 (T, 23%)	0.76
rs1986009	365,939	239	rs12652920 (C, 19%)	1.00	rs12652920 (C, 19%)	1.0	rs4705952 (G, 27%)	0.21	-	-
rs9273410	554,975	5,739	-	-	-	-	-	-	-	-
rs776111176	574,227*	1,447*	rs3997872 (A, 16%)	0.82	-	-	-	-	-	-
rs367983479	244,383	191	rs1504215 (A, 36%)	0.85	rs4142967 (T, 46%)	0.56	-	-	-	-
rs71266076^	175,000	206	rs7824993 (A, 35%)	0.81	-	-	-	-	-	-
rs144829310	496,605	234	-	-	rs1929996 (C, 31%)	0.43	-	-	-	-
rs10905284	67,858	96	-	-	-	-	-	-	-	-
rs11603634	146,648	187	-	-	rs11602802 (C, 33%)	0.40	-	-	-	-
rs7936312	316,410	277	rs7936323 (A, 47%)	0.96	-	-	-	-	-	-
rs7305461	315,101	120	rs1131017 (C, 41%)	0.75	-	-	-	-	-	-
rs703816	284,371	165	-	-	rs167769 (T, 33%)	0.82	-	-	-	-
rs10519068	108,343	123	-	-	rs2279292 (C, 15%)	0.91	rs2279296 (G, 45%)	0.13	rs2279296 (G, 45%)	0.13
rs72743461	30,796	45	-	-	-	-	rs10152544 (C, 48%)	0.30	rs10152544 (C, 48%)	0.30
rs7203459	429,719	435	-	-	rs12919828 (G, 25%)	0.98	rs6498135 (A, 40%)	0.30	-	-
rs2941522	805,248	896	-	-	rs907091 (T, 49%)	0.96	-	-	-	-
rs112502960	270,023	370	rs62076439 (T, 36%)	1.00	-	-	-	-	-	-
rs61840192	201,710	249	rs1031163 (T, 42%)	1.00	-	-	-	-	-	-

Each sentinel SNP reported in the original moderate to severe asthma GWAS corresponds to a haplotype consisting of related SNPs (R^2 > 0.2) as identified in column 2. These SNPs provided us with potential proxies in cases where the original sentinel SNP was not covered by the in-silico platform and also gave an idea of the size of the LD pattern of the signal. In each case, the SNP with the highest R^2 value available was selected as proxy. Multiple proxies were used in instances where an originally selected proxy was unavailable in one or more of the U-BIOPRED datasets. Where a proxy has not been stated (-) then the sentinel SNP was used.

**Indicates the sentinel SNP was not found in 1,000 Genomes Project accessed through LDLink (Phase 3) or HaploReg (Phase 1), ergo values are reflective of the proxy SNP used.*

^This sentinel SNP (rs71266076) was not found in the 1,000 Genomes Project Phase 3 data accessed through LDLink therefore Phase 1 data accessed through HaploReg was used instead.

MAF, minor allele frequency in the European population from 1,000 Genomes Project Phase 3.

FIGURE 2 | PheWAS of signals for moderate-severe asthma. Signals (risk allele in brackets) or their proxies were searched for in the GeneATLAS PheWAS database of the UK Biobank cohort. A Bonferroni adjusted *p*-value of 6.42×10^{-5} was used. The interaction between signals (top) and PheWAS traits (right) are represented on a grid and the area of the circle represents the -$\log_{10}$ (*p*-value) of the association. A larger area indicates a lower *p*-value. Only selected PheWAS traits have been displayed and organized into asthma (blue), blood/immune cell (red), allergy (yellow), other respiratory (green), inflammatory (purple), and auto-immune (gray) groups. The trait "basophil percentage" did not meet the Bonferroni corrected *P*-value for any of the signals and therefore was not included in the Figure. A full list of PheWAS terms and Beta and *P*-values are available in **Supplementary Table 1**. For disease traits (i.e., all except blood cell traits) a black dot represents an odds ratio of <1 with respect to the moderate to severe asthma risk allele. Proxies were used for the following signals which were not present in the database: rs61816761 (rs61816766, $R^2 = 0.50$), rs367983479 (rs1504215, $R^2 = 0.85$), rs71266076 (rs7824993, $R^2 = 0.81$), rs7305461 (rs1131017, $R^2 = 0.75$), rs112502960 (rs62076439, $R^2 = 1.0$), rs61840192 (rs1031163, $R^2 = 1.0$), rs560026225 (rs72687036, $R^2 = 0.66$), and rs776111176 (rs3997872, $R^2 = 0.82$).

WDR36) or signal transduction (*CD247, GNGT2, HLA-DQA1/A2/B1/B2, IL18R1, IL1RL1, IL33, LRRC32, PDCD1,* and *TSLP*). Both *FLG* and *MUC5AC* are structural components involved in barrier formation and GSDMB is involved in cell death. Some genes such as *DEXI, KIF1A, KIAA1109, ZBTB10,* and *ZNF652* had very limited or no information and as yet their function or involvement in biological processes is unknown. The gene function of implicated genes is described in **Table 5**.

TABLE 3 | Scoring system for signal to gene analyses.

Evidence	Weighting
GTEX respiratory relevant cell/tissue eQTL ($P < 0.05$)	1
GTEX Blood eQTL ($P < 0.05$)	1
GTEX respiratory relevant cell/tissue sQTL ($P < 0.05$)	1
GTEX blood sQTL ($P < 0.05$)	1
OTG eQTL Resp ($P < 0.05$)	1
OTG eQTL blood ($P < 0.05$)	1
OTG V2G score (presence of)	1
OTG co-localization study (H4 > 0.8)	1
HaploReg eQTL ($P < 0.05$)	1
Functional variant (presence of in an LD $r^2 > 0.1$)	1
UBIOPRED brush eQTL ($P < 0.05$)	2
UBIOPRED biopsy eQTL ($P < 0.05$)	2
UBIOPRED nasal eQTL (P<0.05)	2
Literature association (presence of)	1
Total possible score	**16**

Each potential candidate causal gene was scored based on supporting evidence. eQTL and sQTL associations were reported on presenting a p-value of <0.05, co-localization studies on presenting an H4 > 0.8 and functional variants if present withing a linkage disequilibrium block with an $r^2 > 0.1$. Double weighting was given to eQTL associations observed in U-BIOPRED, due to the increased relevance of lung tissue and airway epithelial tissues to our study.

TABLE 4 | Selection of genes of interest relative to signal of association.

SNP	Gene in paper	Identified genes	Score	UB score	Total score
rs61816761	FLG	FLG	2	0	2
		TUFT1	0	1	2
		SELENBP1	0	1	2
		C1orf68	0	1	2
rs7523907	CD247	CD247	7	0	7
		BRP44	0	1	2
		CREG1	0	1	2
rs12479210	IL1RL1	IL1RL1	11	1	13
		IL1R1	0	1	2
		IL18R1	6	0	6
rs34290285	D2HGDH	D2HGDH	11	1	13
		PDCD1	3	0	3
		GAL3ST2	1	0	1
		KIF1A	1	1	3
rs560026225	KIAA1109	KIAA1109	4	0	4
rs1837253	TSLP	TSLP	2	0	2
		WDR36	1	0	1
rs1438673	WDR36	WDR36	3	0	3
		CAMK4	1	0	1
		TMEM232	0	1	2
		TSLP	1	1	3
		STMN1	0	1	2
rs3749833	C5orf56	SLC22A5	8	0	8
		SLC22A4	2	0	2

(Continued)

TABLE 4 | Continued

SNP	Gene in paper	Identified genes	Score	UB score	Total score
		P4HA2	2	0	2
		C5orf56	1	0	1
		IRF1	3	0	3
		LOC553103	0	1	2
		SEPT8	0	1	2
		ANKRD43	0	1	2
rs1986009	RAD50	SLC22A5	3	0	3
		SLC22A4	3	0	3
		ACSL6	0	1	2
		IL13	0	1	2
		IL4	0	1	2
rs9273410	HLA-DQB1	HLA-DQB1	2	3	8
		HLA-DQB2	2	3	8
		HLA-DQA1	2	0	2
		HLA-DQA2	2	0	2
		ATF6B	1	0	1
rs776111176	HLA-DQA1	HLA-DQA1	6	2	8
		HLA-DQA2	6	0	6
		HLA-DQB1	6	6	12
		HLA-DQB2	6	0	6
		HLA-DRB3	0	2	2
		HLA-DOB	4	0	4
		AGER	0	2	2
		LY6G6E	0	2	2
		PFDN6	0	2	2
		NEU1	0	2	2
		DOM3Z	0	2	2
		LY6G6D	0	2	2
		COL11A2	0	2	2
rs367983479	BACH2	BACH2	3	0	3
		ANKRD6	0	1	2
		MAP3K7	0	1	2
		GABRR2	0	1	2
rs71266076	MIR5708	ZBTB10	1	0	1
rs144829310	IL33	IL33	2	2	6
		ERMP1	1	0	1
		TPD52L3	0	1	2
rs10905284	GATA3	GATA3	2	0	2
rs11603634	MUC5AC	MUC5AC	2	2	6
		TNNT3	0	1	2
rs7936312	C11orf30	LRRC32	4	0	4
		BRCA2	1	0	1
rs7305461	RPS26	RPS26	4	0	4
		SUOX	4	0	4
		RAB5B	2	0	2
		ERBB3	2	0	2
		ESYT1	1	0	1
		GDF11	1	0	1
		RNF41	1	0	1

(Continued)

TABLE 4 | Continued

SNP	Gene in paper	Identified genes	Score	UB score	Total score
rs703816	STAT6	STAT6	5	0	5
		NEMP1	1	0	1
		RBMS2	1	0	1
		SPRYD4	1	0	1
		EEF1AKMT3	1	0	1
		ZBTB39	0	1	2
		CDK4	0	1	2
		ESYT1	0	1	2
rs10519068	RORA	RORA	4	0	4
		ICE2	0	1	2
		ANXA2	0	1	2
		FOXB1	0	1	2
s72743461	SMAD3	SMAD3	4	0	4
		AAGAB	3	0	3
		C15orf61	0	1	2
		BPGM	0	1	2
		MAP2K1	0	1	2
rs7203459	CLEC16A	CLEC16A	2	1	4
		TEKT5	0	1	2
		PRM1	0	1	2
		DEXI	3	0	3
rs2941522	IKZF3	ORMDL3	4	0	4
		GSDMB	5	0	5
		GSDMA	2	0	2
		PGAP3	3	0	3
		MSL1	1	1	3
		IKZF3	2	0	2
		ARL5C	0	1	2
rs112502960	ZNF652	ZNF652	3	0	3
		GNGT2	3	0	3
		PHOSPHO1	2	0	2
		TMEM92	0	1	2
		HOXB4	0	1	2
		NCRNA00253	0	1	2
rs61840192	LOC101928272	GATA3	4	0	4
		GATA3-AS1	2	0	2

Signals were analyzed using the translational pipeline including publicly available datasets such as Open Target Genetics, GTex, HaploReg and LDLink and our UBIOPRED (UB) datasets of Bronchial Biopsies, Bronchial Brushes, and Nasal Brushes. Each gene was scored based on supporting evidence as laid out in ***Table 3****, with double weighting being given to eQTL associations observed in UB, based on the datasets localized data. The genes highlighted in green were taken forward in downstream analyses.*

Single Cell RNA Expression in Lung Tissue Indicates Cell Type Specificity in Healthy Individuals

The HPA resource was used to look at overall and single cell type gene expression profile in lung tissue from healthy donors for the 37 candidate causal genes. Genes which showed enriched/enhanced expression (4-fold higher expression than in other cells/tissues) in epithelial or blood/immune cells were also identified. All genes except *FLG* showed expression in lung tissue and cells (**Figure 4**). *KIF1A* showed lower levels of lung expression and was undetectable in the cell types analyzed potentially due to the sensitivity of this technique. *MUC5AC* showed overall highest expression in lung tissue and along with *IL33* and *SLC22A4* was enriched in club and ciliated cells. *IL33* also showed high expression in endothelial cells. Both *IRF1* and *RPS26* showed high lung cell type expression but no cell type specificity. *CD247, GNGT2, HLA-DQB2,* and *PDCD1* were expressed at low levels in the datasets and were specific for T-cells and macrophages. *HLA-DQA1/A2, HLA-DQB1* were highly expressed overall and specifically in macrophages, whereas *IL18R1* and *IL1RL1* showed greater expression in granulocytes. A full list of enriched tissues and cell types is available in **Supplementary Table 2**.

Co-localization Analysis Reveals Genes With High Asthma Trait Associations in Lung Tissues and Blood/Immune Cells

The OTG platform was used to investigate whether gene expression in lung tissue and/or blood immune/cells was associated with the asthma trait (**Figure 4B**). *BACH2, IRF1, LRRC32, PDCD1, RORA, STAT6, and TSLP* showed a strong link (H4 score>0.8) to blood/immune cells only, whereas *ZNF652* was exclusively linked to lung tissue expression. Genes *GSDMB, IL1RL1, IL18R1, ORMDL3, RPS26*, and *SUOX* on the other hand showed strong association to both lung tissue and blood/immune cells.

Genes Show Differential Gene Expression in Bronchial Epithelial Cells Isolated From Subjects With Severe Asthma

From the above analyses, 10 genes (*GSDMB, IL18R1, IL1RL1, IL33, MUC5AC, ORMDL3, RPS26, SLC22A4, SUOX, and ZNF652)* showed enriched epithelial expression and/or co-localization of the asthma trait in lung tissue. These specific genes were further investigated in the bronchial epithelium of mild-moderate or severe asthma subjects compared to control subjects (**Figure 5**). Six of the 10 studied genes showed differential expression in asthma vs. control subject cells. Expression levels of *IL18RL1* ($P = 0.019$, MA and 0.0008, SA), *IL1RL1* ($P = 0.0034$, MA and 0.010, SA), and *ORMDL3* ($P = 0.040$, MA and 0.0029, SA) were higher in both mild-moderate and severe asthma subjects compared to controls (**Figures 5A,B,D**) whereas expression levels of *MUC5AC* ($P = 0.022$) and *RPS26* ($P = 0.0004$) were higher in severe asthma subjects only compared to controls (**Figures 5C,E**). Expression levels of *SLC22A4* ($P = 0.0051$, MA and <0.0001, SA) was lower in both mild-moderate and severe asthma subjects compared to controls (**Figure 5F**). Expression levels of *SLC22A4* ($P = 0.0028$) was decreased in severe asthma (**Figure 5F**) compared to mild-moderate asthma. Genes *GSDMB, IL33, SUOX*, and *ZNF652* did not show any significant differences in expression across groups (**Supplementary Figure 1**).

Genes Show Differential Gene Expression in Blood Cells Isolated From Subjects With Severe Asthma

From the above analyses, 18 genes (*BACH2, CD247, HLA-DQA1/A2, IL1RL1, ORMDL3, RORA, STAT6, GNGT2, GSDMB, HLA-DQB1/B2, IL18R1, IRF1, LRRC32, PDCD1, RPS26, SUOX,* and *TSLP*) showed enriched blood/immune expression and/or co-localization of the asthma trait in blood. These genes were further investigated in blood cells of mild-moderate or severe asthma subjects compared to control subject cells (**Figure 6**). Seven of the 18 genes studied showed differential expression across control and asthma subject groups. Expression levels of *BACH2* ($P = <0.0001$ for both) and *CD247* ($P = 0.0002$, vs. C and $P = 0.0204$ vs. MA for *CD247*) were significantly lower in severe asthma compared to controls and mild-moderate asthma (**Figures 6A,B**). Expression levels of *HLA-DQA1/A2* ($P = 0.0077$), *ORMDL3* ($P = 0.0021$), and *RORA* ($P = 0.0005$) were significantly lower in severe asthma subjects only compared to controls (**Figures 6C–E**). Expression levels of *IL1RL1* ($P = 0.0002$ vs. MA, $P = <0.0001$ vs. SA) and *STAT6* ($P = 0.0079$ vs. MA, $P = <0.0001$ vs. SA) were significantly higher in mild-moderate and severe asthma compared to controls (**Figures 6F,G**). Genes *GNGT2, GSDMB, HLA-DQB1/B2, IL18R1, IRF1, LRRC32, PDCD1, RPS26, SUOX,* and *TSLP* didn't show any significant differences in expression between groups (**Supplementary Figure 2**).

Genetic Association Database of Complex Diseases and Disorders Identifies Genes Associated With Asthma

The DAVID functional annotation tool was used to perform enrichment analysis for the 37 candidate genes in the GAD. The top 10 gene clusters (5% FDR) are listed in **Table 6**. The full table is available in **Supplementary Table 3**. Asthma was significantly associated with the largest cluster of genes comprising of *FLG, IL33, GSDMB, SLC22A5, SMAD3, TSLP, GATA3, SUOX, MUC5AC, IL1RL1, IRF1, ORMDL3, STAT6, HLA-DQA2, IL18R1, HLA-DQA1,* and *HLA-DQB1* ($P = 1.78 \times 10^{-15}$). Genes were also associated with other, predominantly autoimmune, diseases including Type 1 diabetes, Celiac disease, Crohn's disease, and rheumatoid arthritis which complements the findings of the PheWAS. There is considerable overlap between genes associated with asthma and other diseases with *HLA-DQA1/B1* genes present in ∼74% of significant terms (**Supplementary Table 4**).

Pathway and Gene Ontology Term Enrichment Analysis Highlights Genes Involved in Asthma and the Strong Presence of a HLA-DQA1/A2/B1/B2 Cluster

GO term enrichment analysis (**Figure 7A**) identified a predominant gene cluster of *HLA-DQA1/A2/B1/B2* genes. These genes were enriched mainly in processes involving MHC class II molecules and vesicle-membrane transport interactions. KEGG and REACTOME pathway analyses also identified this gene cluster as being involved in asthma, MHC class II antigen presentation, autoimmune processes, pathogenic infections, and diabetes (**Figure 7B**). PD-1 signaling was the most significant pathway identified which was enriched for *CD247*, *PDCD1*, *HLA-DQA1/A2,* and *HLA-DQB1/2*. Other significant pathways, including phosphorylation, translocation and generation of second messenger molecules, involved the *CD247*, *HLA-DQA1/A2/B1/B2* genes. Full result tables are available in **Supplementary Tables 5, 6**.

Gene Interactions Analysis Highlights Key Co-localization and Co-expression Relationships and Predicts Additional Related Genes

Gene interaction analysis of the 37 candidate genes identified that the strongest physical interactions were observed between *HLA-DQA1* and *B1*, *SMAD3* and *SMAD4* (predicted gene), *CD247* and *ZAP70* (predicted gene), and *FLG* and *CASP14* (predicted gene). *IL33, LRRC32*, *IL18R1* and CD274 (predicted gene) were observed to have the greatest number of co-localized partners. *GATA3* and *IL18R1* were observed to have the greatest number of co-expressional partners. *AAGAB, KIF1A,* and *KIAA1109* showed no interaction between any other gene (**Figure 8**). The predicted genes with the strongest interactions were *CD274, IL37, IRAK4, PDCD1LG2,* and *ZAP70* (**Supplementary Table 7**).

Gene Drug Interaction and Clinical Trial Data Analyses Identifies 14 Genes Targeted by Licensed Therapeutics and 5 Genes Under Clinical Trials for Asthma

Of the 37 candidate causal genes and five predicted genes, DGIdb identified that 14 of the genes have known interactions with currently available therapeutics, with highest scoring genes being, *SLC22A5* (Carnitine, 63.79), *MUC5AC* (Ensituximab, 31.9), and *CD274* (Avelumab, 23.92). Novel inhibitory antibodies under Phase II/III/IV clinical trials have been identified as potential novel therapeutics for asthma using https://clinicaltrials.gov, for 5 of the highlighted genes [*GATA3, IL1RL1, IL33, SMAD3,* and *TSLP* (**Table 7**)].

DISCUSSION

Asthma is a heterogenous disease in which both genetic and environmental factors contribute to susceptibility and progression (35). Severe asthma, characterized by uncontrolled symptoms, a burden of medication and frequent exacerbations, remains inadequately managed in many patients (2). The current study aimed to provide significant translation of 25 genetic signals identified as risk factors for the development of moderate to severe asthma to new biological insight using a broad range of approaches and datasets. The main findings are (i) moderate-severe asthma genetic signals drive a large number of inflammatory cell traits particularly eosinophil levels and are risk factors for related traits such as atopic dermatitis and other inflammatory diseases, (ii) our signal to gene pipeline identified 37 candidate causal genes, (iii) genes show enrichment in lung tissue and inflammatory cells illustrating both the role in

FIGURE 3 | Bubble plots identifying the eQTL patterns for each selected SNP:Gene pairing as identified in **Table 3** and tissue specific eQTLs. Plot identifies each eQTL *P*-value (x-axis) and *B*-value (y-axis) stratified based on the asthma risk allele for each identified compartment as denoted by the corresponding color; Blue = Lung, Red = Whole Blood, Green = T-cells, Yellow = U-BIOPRED Bronchial Biopsy, Orange = U-BIOPRED Bronchial Brush and Purple—U-BIOPRED Nasal Brush. The size of the bubble indicates the generated score in **Table 3**. EQTLs listed are those that achieved a *P*-value of $P < 0.05$ in one of the interrogated datasets.

TABLE 5 | Gene function of candidate causal genes.

Gene	Full name	Molecular function/biological process
Barrier formation/defense		
FLG	Filaggrin	(1) protein binding (2) structural constituent of skin epidermis/establishment of skin barrier (3) peptide cross-linking
MUC5AC	Mucin 5AC, oligomeric mucus/gel-forming	(1) phosphatidylinositol-mediated signaling (2) gel-forming secretory mucin
Cell death		
GSDMB	Gasdermin B	(1) wide pore channel activity (2) granzyme-mediated programmed cell death signaling pathway (3) pyroptosis
DNA/RNA binding		
BACH2	BTB domain and CNC homolog 2	(1) sequence-specific double-stranded DNA binding
GATA3	GATA binding protein 3	(1) transcription regulatory region sequence-specific DNA binding (2) core promoter proximal region sequence-specific DNA binding (3) E-box binding (4) negative regulation of mammary gland epithelial cell proliferation (5) positive regulation of transcription, DNA-templated (6) positive regulation of transcription from RNA polymerase II promoter (7) mesenchymal to epithelial transition (8) lymphocyte migration (9) positive regulation of interleukin-5 and−13 secretion
IRF1	Interferon regulatory factor 1	(1) protein binding (2) RNA polymerase II core promoter proximal region sequence-specific DNA binding (3) apoptotic process (4) cell cycle arrest (5) cellular response to interferon-beta (6) defense response to virus
RORA	RAR related orphan receptor A	(1) DNA-binding transcription factor activity (2) ligand-activated transcription factor activity (3) oxysterol binding (4) RNA polymerase II cis-regulatory region sequence-specific DNA (5) cellular response to sterol (6) intracellular receptor signaling pathway (7) positive regulation of transcription, DNA-templated (8) positive regulation of transcription by RNA polymerase II
RPS26	Ribosomal protein S26	(1) negative regulation of RNA splicing (2) cadherin binding involved in cell-cell adhesion (3) mRNA binding (4) protein binding (5) structural constituent of ribosome (6) cytoplasmic translation (7) negative regulation of RNA splicing
SMAD3	SMAD family member 3	(1) RNA polymerase II core promoter proximal region sequence-specific DNA binding (2) transforming growth factor beta receptor binding and signaling pathway (3) zinc ion binding (4) phosphatase binding (5) protein kinase binding (6) ubiquitin protein ligase binding (7) protein homodimerization activity (8) negative regulation of transcription from RNA polymerase II promoter (9) positive regulation of epithelial to mesenchymal transition (10) evasion or tolerance of host defenses by virus (11) negative regulation of cell growth (12) positive regulation of transcription factor import into nucleus (13) positive regulation of nitric oxide biosynthetic process (14) negative regulation of fat cell differentiation (15) negative regulation of cytosolic calcium ion concentration (16) positive regulation of extracellular matrix assembly

(Continued)

TABLE 5 | Continued

Gene	Full name	Molecular function/biological process
STAT6	Signal transducer and activator of transcription 6	(1) protein binding (2) protein phosphatase binding (3) DNA-binding transcription activator activity, RNA polymerase II-specific
WDR36	WD repeat domain 36	(1) poly(A) RNA binding
Membrane binding/transport		
AAGAB	Alpha and gamma adaptin binding protein	(1) protein binding (2) protein transport (3) may be involved in endocytic recycling of growth factor receptors such as EGFR
CLEC16A	C-type lectin domain containing 16A	(1) possible involvement in autophagy and endosomal transport
SLC22A4	Solute carrier family 22 member 4	(1) carnitine transport (2) quaternary ammonium group transport (3) carnitine transmembrane transporter activity (4) cation:cation antiporter activity (5) PDZ domain binding (6) amino acid import across plasma membrane
SLC22A5	Solute carrier family 22 member 5	(1) carnitine transport (2) quaternary ammonium group transport (3) carnitine transmembrane transporter activity (4) PDZ domain binding (5) sodium-dependent organic cation transport
ORMDL3	ORMDL sphingolipid biosynthesis regulator 3	(1) Protein binding (2) positive regulation of autophagy (3) positive regulation of protein localization to nucleus (4) may indirectly regulate ER-mediated calcium signaling
Metabolic processes		
D2HGDH	D-2-hydroxyglutarate dehydrogenase	(1) (R)-2-hydroxyglutarate dehydrogenase activity
PGAP3	Post-GPI attachment to proteins phospholipase 3	(1) Possible involvement in GPI anchor metabolic process and hydrolase activity, acting on ester bonds
MSL1	MSL complex subunit 1	(1) Histone H4-K16 acetylation
SUOX	Sulfite oxidase	(1) Possible involvement in sulfur compound metabolic processing
Signal transduction		
CD247	CD247 molecule	(1) Protein binding (2) identical protein binding (3) protein tyrosine kinase binding
GNGT2	G protein subunit gamma transducin 2	(1) GTPase activity
HLA-DQA1/A2/B1/B2	Major histocompatibility complex, class II, DQ beta 1	(1) Humoral immune response mediated by circulating immunoglobulin (2) immunoglobulin production involved in immunoglobulin mediated immune response (3) antigen processing and presentation of exogenous peptide antigen *via* MHC class II (4) T cell receptor signaling pathway
IL18R1	Interleukin 18 receptor 1	(1) Protein binding (2) interleukin-18-mediated signaling pathway (3) positive regulation of interferon-gamma production (4) positive regulation of NF-kappaB transcription factor activity (5) positive regulation of NIK/NF-kappaB signaling (6) positive regulation of T-helper 1 cell cytokine production
IL1RL1	Interleukin 1 receptor like 1	(1) protein binding (2) IL33 receptor
IL33	Interleukin 33	(1) positive regulation of chemokine secretion (2) positive regulation of macrophage activation (3) cytokine activity (4) protein binding
LRRC32	Leucine rich repeat containing 32	(1) TGF-beta binding and signaling pathway
PDCD1	Programmed cell death 1	(1) protein binding (2) cell surface receptor signaling pathway (3) negative regulation of CD4-positive, alpha-beta T cell proliferation (4) positive regulation of activated CD8-positive, alpha-beta T cell apoptotic process (5) response to cytokine (6) signal transduction (7) T cell co-stimulation

(Continued)

TABLE 5 | Continued

Gene	Full name	Molecular function/biological process
TSLP	Thymic stromal lymphopoietin	(1) cytokine activity (2) positive regulation of tyrosine phosphorylation of STAT protein (3) promotes T helper type 2 cell responses (4) positive regulation of chemokine production (5) positive regulation of cytokine-mediated signaling pathway
Unknown		
DEXI	DEXI homolog	Unknown
KIAA1109	KIAA1109	(1) Protein binding
KIF1A	Kinesin family member 1A	(1) Identical protein binding
ZNF652	Zinc finger protein 652	(1) Protein binding
ZBTB10	Zinc finger and BTB domain containing 10	(1) Protein binding

Gene function and biological process for the candidate genes as determined using the Gene 2 Function and UniProt databases, listing associated disease states and gene function annotations.

inflammation and structural changes in the airways, (iv) 32 of the 37 genes had additional support for a role in asthma including some with differential expression in the airways and blood cells of severe patients, (v) genes show enrichment for pathways relevant to T cell function, interferon signaling, endoplasmic reticulum (ER) biology and others, (vi) gene interaction analyses identified predicted genes already known to be involved in asthma and finally, (vii) these genes and pathways show potential as targets for novel drug development or repurposing.

PheWAS Identified Features of Asthma and Other Traits Driven by Moderate-Severe Asthma Signals

Our initial analysis focused on understanding the uniqueness of the moderate-severe asthma signals by testing for association with a large number other traits *via* PheWAS. This approach can identify novel associations with quantitative traits e.g., lung function and provide significant inference regarding the mechanisms underlying the original association. All signals showed associations with the asthma trait including the *MUC5AC* signal that was identified as a potential moderate-severe asthma signal originally (7) and blood/immune cell trait associations were very prominent. Recently, this *MUC5AC* signal has been identified as potentially more relevant to adult onset asthma (36), a phenotype that shares less overlap with atopic comorbidities than childhood onset asthma. This may explain, at least in part, the differential profile compared to many of the other signals which showed associations with e.g., atopic dermatitis.

Interestingly, some signals showed trait specificity whilst others had broader trait associations. Of particular note are signals rs9273410 and rs776111176 both corresponding to *HLA-DQA1/A2/B1/B2* candidate genes, which had the widest disease associations including Hay fever, type 1 diabetes, inflammatory bowel disease, ulcerative colitis and psoriasis. The gene association for these results were confirmed in DAVID pathway and gene enrichment analysis where the *HLA-DQA1/A2/B1/B2* gene cluster was present in ∼74% of diseases and pathways. Interestingly, PheWAS analysis for these signals showed a protective effect for some auto-immune diseases. Previous studies of signals in this region have confirmed inverse disease dependence risk between asthma and autoimmune disease, for example signal rs2395185, the asthma risk allele T was shown to be protective for ulcerative colitis (37, 38). These differences in trait profiles across diseases may offer insight into disease pathophysiology for this gene cluster.

Another notable association with the majority of signals was the association with blood eosinophils levels, with eosinophils known to be an effector cell in asthma and linked to severe asthma pathology (39). Indeed, current treatments such as Mepolizumab, Benralizumab, and Reslizumab reduce exacerbations through blocking IL5/IL5 signaling and reducing eosinophil numbers (40–43).

Interestingly, signal rs2941522 corresponding to candidate genes *ORMDL3, GSDMB, PGAP3,* and *MSL1* had the strongest association with neutrophil counts. This observation may point to a particular pathology for carriers of this variant and a role in non-eosinophilic or neutrophilic asthma (44, 45). Multiple GWAS and functional studies relate asthma to *ORMDL3* and *GSDMB* (46, 47) however to our knowledge no studies have been published investigating this variant or linked genes with regards to neutrophils in asthma.

Signals, rs12479210, rs144829310, and rs1986009 had the strongest eosinophilic associations, corresponding to *IL1RL1, IL18R1, SLC22A4/5,* and *IL33* candidate causal genes. *IL18R1* and *IL1RL1* have been shown in many studies to have variants associated with asthma as well as having altered expression profiles in asthmatic subjects (3, 28). Predictive interaction analysis revealed strong physical interactions between *IL1RL1* and *TMED1* which is unsurprising as both are related to *IL33* signaling (48) which has been extensively researched in the context of asthma (49).

Signal to Gene Analyses Identifies 37 Plausible Candidate Causal Genes

The complexity of signals identified through GWAS highlight the importance of considering relevant tissue compartments when determining signal to gene associations. For many of the signals, we identified multiple plausible candidate causal genes to a single signal, with different genes acting as eQTLs in different cells and tissue types. This is potentially as anticipated as it is feasible that a signal may influence multiple genes and pathways. For example,

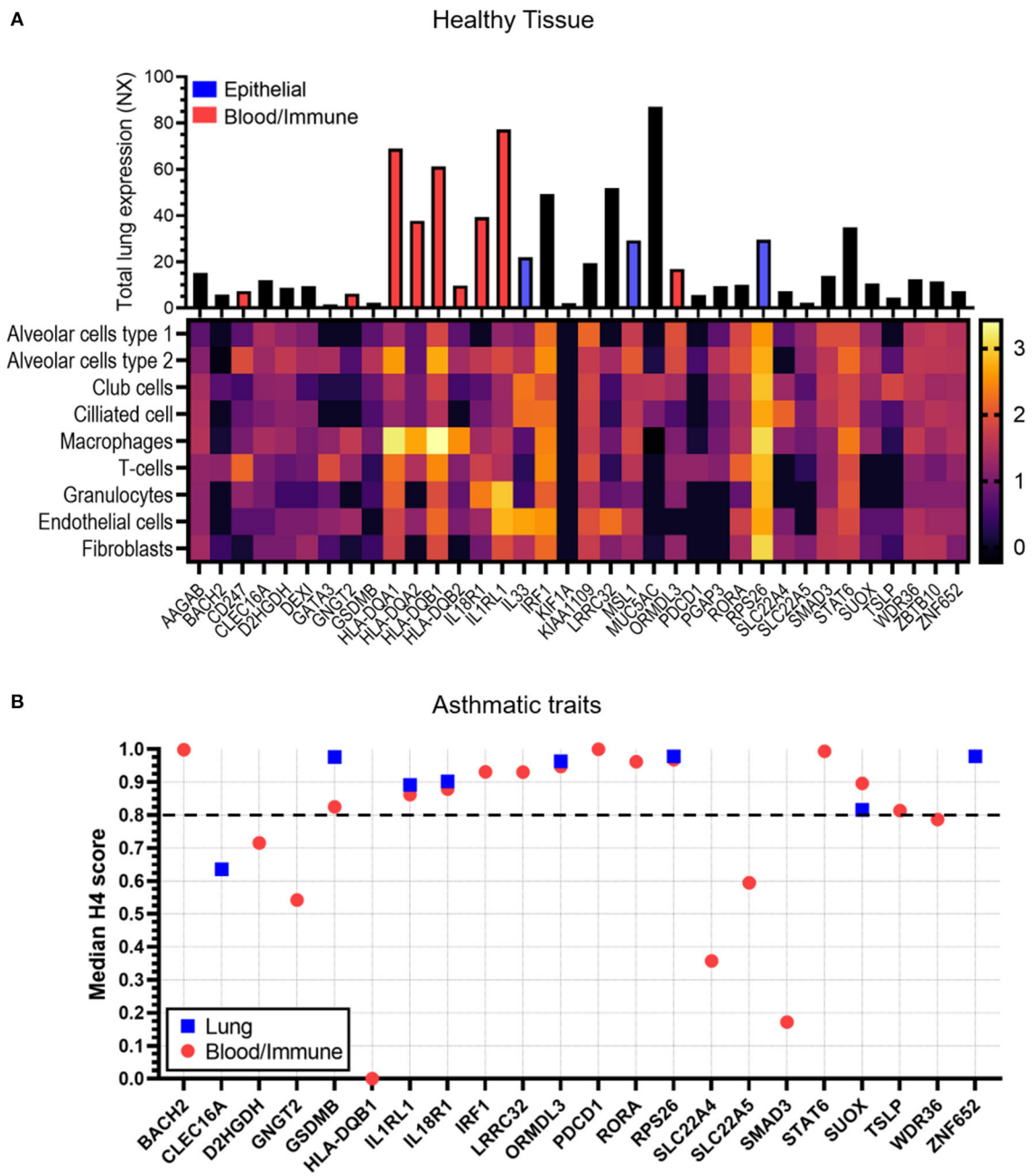


FIGURE 4 | Tissue expression and colocalization analysis of candidate genes. **(A)** Data was taken from the Human Protein Atlas for both total lung RNA expression scores (bar chart, NX) and single cell RNA expression [heatmap, $\log_{10}$ (median transcripts per million protein coding genes)] in healthy tissue. Bar directly corresponds to gene in heatmap. The Human Protein Atlas was also analyzed for genes which showed enriched/enhanced gene expression in epithelial cells (blue bar) and blood/immune cells (red bar) compared to other cell types. A full list of enriched/enhanced tissue and cell gene expression data per gene is available in **Supplementary Table 2**. No data was available for FLG and therefore has not been shown. **(B)** Co-localization analysis (H4 scores) data was taken from the Open Target Genetics (OTG) platform for lung (blue squares) or blood/immune cells (red circles) in studies with the asthma trait (exclusive). Scores represent evidence of association between candidate gene, specific tissue and asthma trait. A higher score indicates greater association. Dotted line shows cut off value of 0.8. Candidate genes *AAGAB, CD247, DEXI, FLG, GATA3, HLA-DQA1/A2, B2, IL33, KIF1A, KIAA1109, MSL1, MUC5AC, PGAP3*, *and ZBTB10* had no association data for lung/blood/immune cells/tissue in OTG and therefore have not been shown.

considering signal rs12479210, we observe an eQTL for *IL1RL1* in lung tissue but not in a T-cell population, where *IL18R1* is the observed eQTL. This strongly suggests that a single signal may drive the differential expression of multiple genes and contribute to asthma mechanisms in multiple ways, the *IL1RL1* signal remains one of the most reproducible asthma signals potentially for this reason (50). Additionally, we highlight that when eQTLs are common across different tissue types, these may present

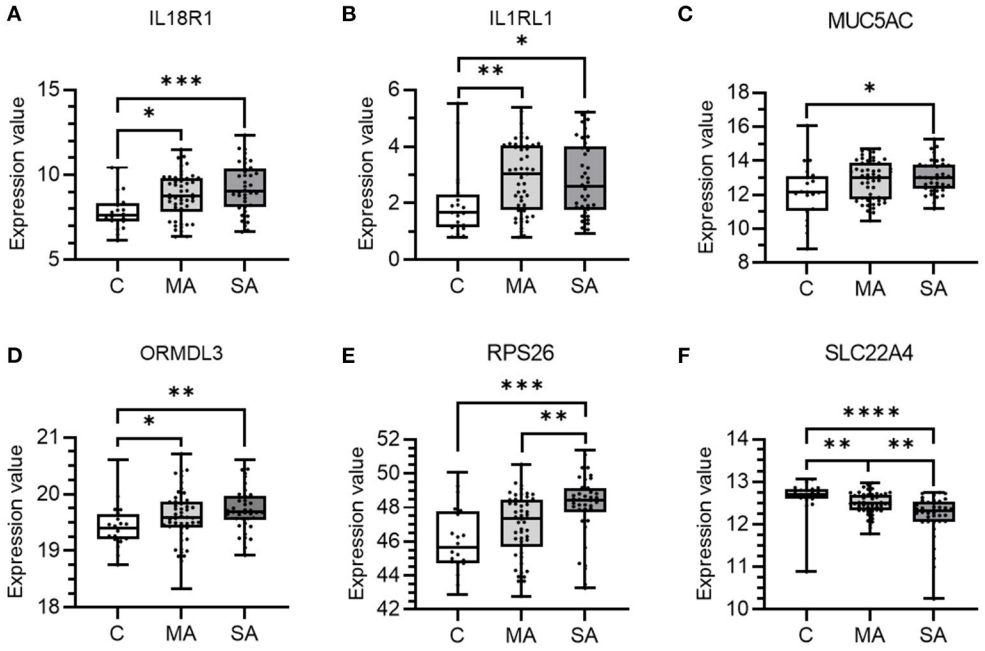


FIGURE 5 | mRNA expression of candidate causal genes in bronchial epithelial cells taken from patients with asthma and controls. Boxes shows the median and IQR and the whiskers show the minimum and maximum data. Bronchial epithelial brush samples were from controls (C, $n = 20$) and patients with mild-moderate (MA, $n = 50$) and severe (SA, $n = 38$) asthma. Data is shown for **(A)** IL18R1, **(B)** IL1RL1, **(C)** MUC5AC, **(D)** ORMDL3, **(E)** RPS26, and **(F)** SLC22A4. Expression values were taken from the dataset and a Kruskal-Wallis test with a two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli used to control the FDR at 5% was performed. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

dramatically different effects on gene expression. For example, when considering the above locus of association, we observe a positive eQTL for *IL1RL1* in whole blood (B = 0.1), but a negative eQTL effect in whole lung (B = −0.372). This emphasizes the need for further analysis on these identified eQTLs in order to unpick the relationship between the eQTLs and asthma relevant phenotypes (51). The 37 identified candidate causal genes broadly mapped to; barrier formation/defense, cell death, DNA/RNA binding, membrane binding and transport, metabolic processes, and signal transduction. Importantly, independent biological studies of asthma have identified a potential role for several of these pathways e.g., epithelial barrier/defense (52) and DNA/RNA binding (53), however more novel mechanisms identified require further investigation.

Gene to Asthma Analyses Provide Significant Support for a Role of Candidate Causal Genes in the Structural Changes in the Airways and Inflammation in Asthma

Structural remodeling is a major component of asthma pathophysiology resulting in the narrowing of the airways. Changes comprise of loss of ciliated cells, goblet cell metaplasia, increased sputum production, basement membrane thickening and smooth muscle cell hypertrophy, and hyperplasia leading to an overall decline in lung function (54–56). Airway remodeling is thought to occur in response predominantly to chronic inflammation, which is supported by studies showing that steroid treatment in asthmatic patients both reduces airway inflammation and has a positive effect on airway remodeling (57, 58). The strong association of these signals with inflammatory traits and a candidate causal genes list with genes predominantly involved in signal transduction indicates these signals may contribute to airway remodeling.

Immune cell infiltration of the lung epithelium is a magnifier of asthmatic inflammation (59). *STAT6* expression was increased in the blood cells of asthmatic patients compared to controls in the dataset analyzed in this study. It is a transcription activator and has been linked to goblet cell metaplasia and an increase in MUC5AC production (60). The MUC5AC protein is a secreted gel forming mucin and its protein levels have also been shown to be increased in asthmatic airway mucus (61). Furthermore, in the airway epithelial dataset analyzed, *MUC5AC* showed a significant increase in gene expression in patients with severe asthma. Indeed, IL-13, an inflammatory cytokine highly associated with asthma induced activation of STAT6 has been shown to increase mucin expression and mucus metaplasia in both airway epithelial cells and submucosal glands in mice and has been linked to goblet cell hyper/metaplasia in humans (62). Both BACH2 and RORA are also DNA binding proteins, however in blood cells, their expression values were significantly lower in asthmatics. RORA has been shown to play a role in lung

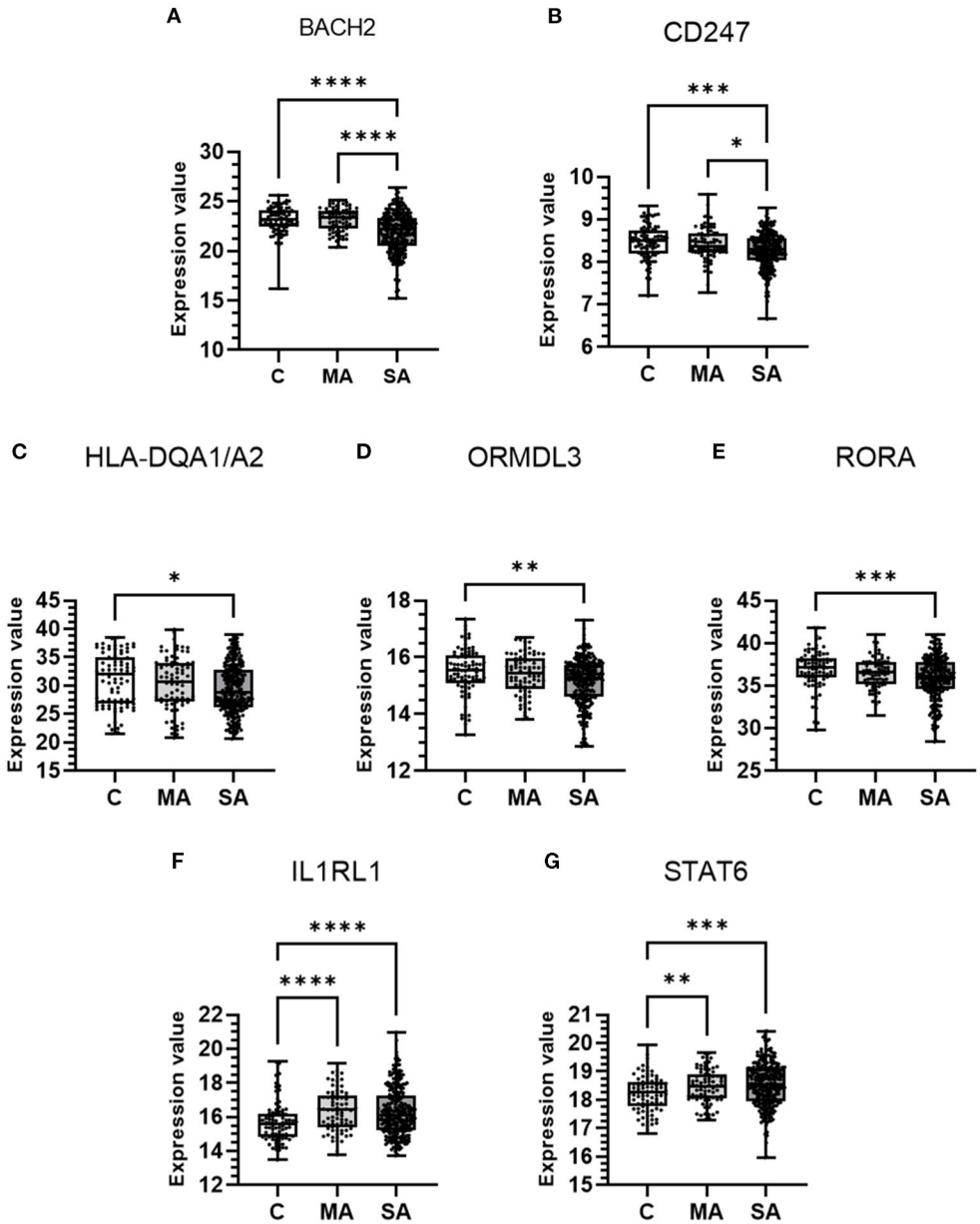


FIGURE 6 | mRNA expression of candidate causal genes in blood cells taken from patients with asthma and controls. Boxes shows the median and IQR and the whiskers show the minimum and maximum data. Blood samples were from controls (C, $n = 87$) and patients with mild-moderate (MA, $n = 77$) and severe (SA, $n = 246$) asthma. Data is shown for **(A)** BACH2, **(B)** CD247, **(C)** HLA-DQA1/A2, **(D)** ORMDL3, **(E)** RORA, **(F)** IL1RL1, and **(G)** STAT6. Expression values were taken from the dataset and either a Kruskal-Wallis test or Welch's ANOVA (STAT6), both with a two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli used to control the FDR at 5%, was performed. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

development (63) and childhood asthma (64) and studies in mice have shown BACH2 to repress T-cell cytokine production (65, 66) indicating lower levels of this protein in asthma may contribute to unregulated inflammation. Together, these findings suggest that imbalances in transcriptional regulators in inflammatory cells may, in part, contribute to asthma pathology possibly *via* downstream processes which affect multiple cells types including the airway epithelium. However, some caution is required as these expression datasets were from mixed cell populations, so we cannot exclude the influence of different cell populations between groups driving some mRNA level differences. Furthermore, the signal for the *MUC5AC* candidate gene, rs11603634, associated with the asthma, eosinophils and platelets traits only, indicating a role specific to severe/eosinophilic asthma that is potentially not driven by allergic mechanisms due to the lack of association with related traits in PheWAS.

TABLE 6 | DAVID functional annotation tool analysis of the genetic association database of complex diseases and disorders (GAD).

Term	Genes	Count (%)	FE*	*p*-value	FDR
Asthma	FLG, IL33, GSDMB, SLC22A5, SMAD3, TSLP, GATA3, SUOX, MUC5AC, IL1RL1, IRF1, ORMDL3, STAT6, HLA-DQA2, IL18R1, HLA-DQA1, HLA-DQB1	17 (45.95)	14.53	1.78×10^{-15}	1.07×10^{-12}
Diabetes mellitus, type 1	KIAA1109, GSDMB, CLEC16A, PDCD1, SUOX, HLA-DQA2, BACH2, HLA-DQA1, HLA-DQB1	9 (24.32)	32.76	1.58×10^{-10}	4.76×10^{-08}
Celiac disease	IL1RL1, KIAA1109, IRF1, CLEC16A, CD247, BACH2, IL18R1, HLA-DQA1, HLA-DQB1	9 (24.32)	22.97	2.72×10^{-09}	5.44×10^{-07}
Nasal polyposis	IL33, IL1RL1, WDR36, HLA-DQA1, HLA-DQB1	5 (13.51)	140.38	3.00×10^{-08}	4.30×10^{-06}
Obesity|asthma	IL33, IL1RL1, GSDMB, TSLP, ORMDL3, IL18R1	6 (16.22)	60.47	3.57×10^{-08}	4.30×10^{-06}
Ulcerative colitis	SLC22A4, GSDMB, SLC22A5, ORMDL3, STAT6, HLA-DQA1, HLA-DQB1	7 (18.92)	31.27	6.45×10^{-08}	6.46×10^{-06}
Arthritis, rheumatoid	SLC22A4, IL1RL1, KIAA1109, CD247, HLA-DQA2, HLA-DQA1, HLA-DQB1	7 (18.92)	19.79	9.79×10^{-07}	8.41×10^{-05}
Crohn's disease	GSDMB, SLC22A5, SMAD3, ORMDL3, BACH2, HLA-DQA1	6 (16.22)	23.12	4.63×10^{-06}	3.48×10^{-04}
Diabetes, type 1	SLC22A4, SLC22A5, IRF1, CLEC16A, PDCD1, HLA-DQA1, HLA-DQB1	7 (18.92)	14.48	6.02×10^{-06}	3.62×10^{-04}
Rheumatoid arthritis	SLC22A4, SLC22A5, CLEC16A, PDCD1, HLA-DQB2, HLA-DQA1, HLA-DQB1	7 (18.92)	14.48	6.02×10^{-06}	3.62×10^{-04}

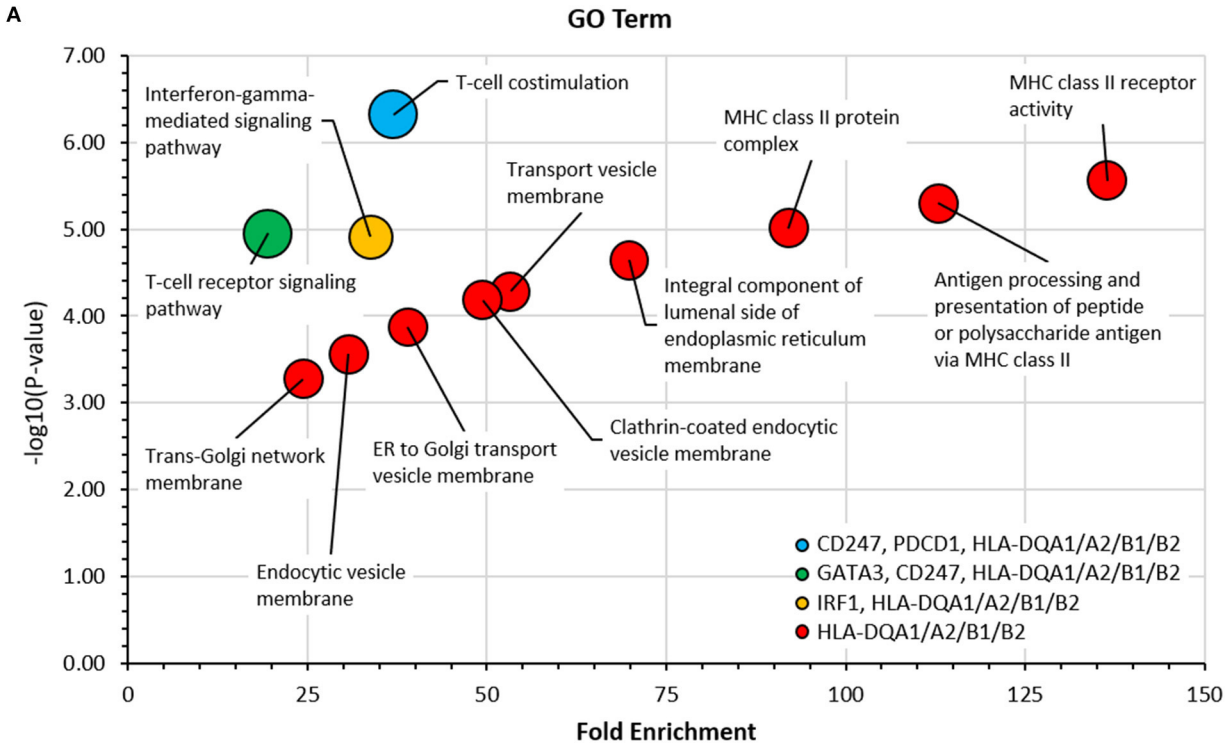
The top 10 genes clusters and disease associations are shown, the full table is available in ***Supplementary Table 3****. All results are below 5% FDR.*
**Fold enrichment.*

Eosinophils, mast cells and T-helper (Th) cells are the key producers of inflammatory cytokines, IL-4, IL-5, and IL-13, associated with Th2 asthma (59, 67). From the signal association with eosinophils in the PheWAS analysis, it suggests inflammation plays a key role in asthma pathogenesis. Indeed, more than half of the 37 candidate causal genes were found, through gene function and pathway analyses, to be involved in signaling pathways related to inflammatory cytokines. Some genes such as *TSLP, IL-33* and its receptor *IL1RL1* have well-established asthma inflammatory roles (51, 68). *IL1RL1*, showed increased expression in both airway epithelium and blood of asthmatic patients compared to controls. Furthermore, inflammatory pathology is supported by the presence of *GATA3* and *STAT6* in asthma and inflammatory gene clusters in the pathway analysis. Both of these transcription factors have been shown to be vital for Th2 cell activation in asthma (69). Expression of *IL18R1*, a receptor for IL-18, was also increased in asthmatic airway epithelium. IL-18 has been shown to regulate both Th1 and Th2 immune responses (70, 71) through INFγ and IL-12 (72) and induce IgE production *via* IL-4 and STAT6 (73). Additionally, for *IL1RL1* and *IL18R1* colocalization analysis revealed strong association with both lung tissue and blood/immune cells indicating infiltration into the lung.

Gene Functional Annotation, Pathway, and Gene Interaction Analysis Provides Insight Into Asthma Mechanisms Including Prediction of Other Asthma Related Genes

Functional and pathway analysis has revealed that the 37 candidate genes are involved in predominantly transcriptional regulation, membrane interaction and cytokine signaling, indicating that these processes may be disrupted or imbalanced in patients with asthma. A strong *HLA-DQA1/A2/B1/B2* gene cluster has emerged involving in a broad range of processes including auto-immune disease, viral and bacterial infection, antigen presentation and ER/Golgi/vesicle membrane interaction. If the gene cluster is broadened to include *CD247*, then combined these five genes are present in over 75% pathway and GO terms and processes. The *HLA* genes are part of a class of major histocompatibility complex (MHC) molecules which present exogenous allergen on airway dendritic cells and interact with CD247, part of the T-cell antigen receptor complex, to achieve T-cell activation (74, 75). From the blood dataset utilized in this study, both *CD247* and *HLA-DQA1/A2* gene expression were found to be decreased in asthmatic patients. Although studies in asthma are lacking, type 2 (non-insulin dependent) diabetics have shown chronic inflammation resulted in immunosuppression and CD247 downregulation (76). Seeing as chronic inflammation is also present in severe asthmatics, it is possible that a similar process is involved. Indeed, the signals, rs9273410 and rs776111176, for the *HLA-DQ* genes also had a strong risk association with diabetes albeit, type 1 (insulin-dependent) diabetes. Genetic variants in the *HLA* region have long been associated with asthma (77), however deconvolution of this region has been difficult.

Another feature which has appeared is the interaction/involvement of the candidate genes with the ER apparatus. The role of the ER, ER stress and the unfolded protein response (UPR) in asthma remains unclear. An increase in ER stress has been reported in asthma, and ER stress has also been shown in turn to increase mucin production, including MUC5AC, and airway remodeling (78, 79). In some studies, ORMDL3, an ER transmembrane protein, has been heavily implicated in asthma related ER stress (78, 79). *ORMDL3* gene expression was shown to be increased in asthmatic airway epithelium but decreased in blood compared to controls in the datasets used in this study. In



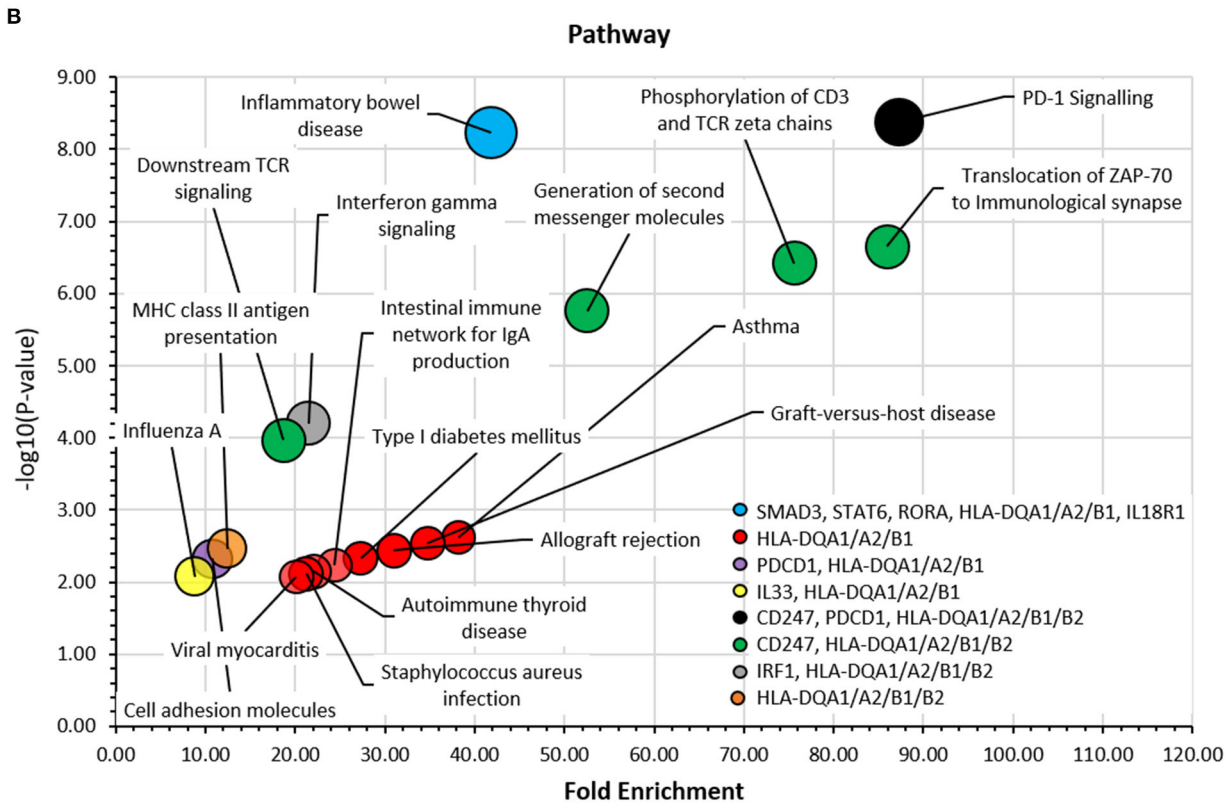


FIGURE 7 | Gene ontology (GO) term and pathway analyses. The DAVID bioinformatic tools was used to assess candidate genes in **(A)** GO Term and **(B)** KEGG and REACTOME Pathway analyses. -$\log_{10}$ (*p*-value) was plotted against fold enrichment for gene groups and the area of the circle is proportional to number of genes in the group. 5% FDR. Full analyses can be viewed in **Supplementary Tables 5, 6**.

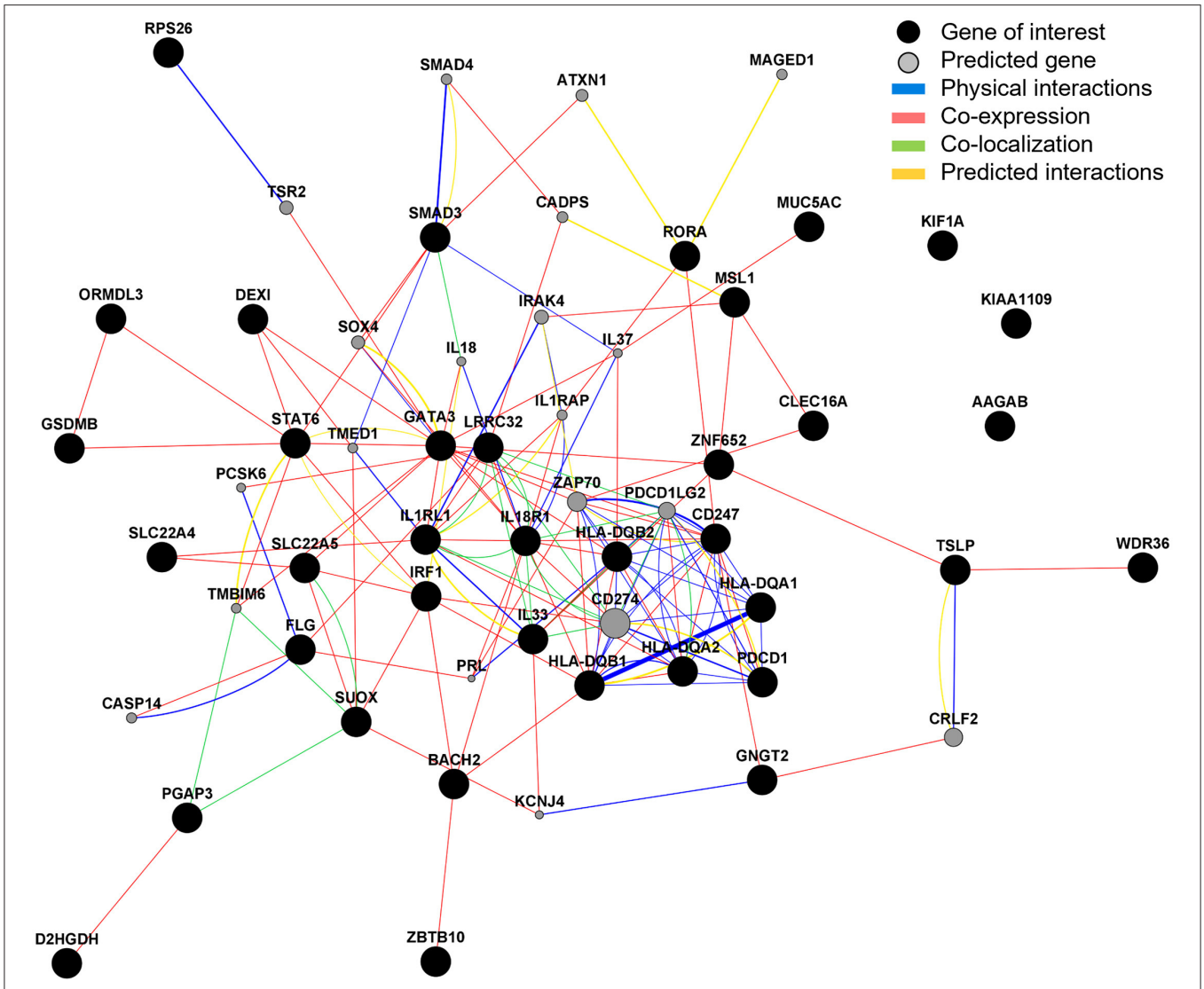


FIGURE 8 | Gene interaction analysis. GeneMANIA was used to explore and predict gene interactions. Black circles represent inputted candidate genes and gray circles are predicted genes (selected to show top 20 with strongest predicted association). The area of the gray circle represents prediction scores and thickness of line represent interaction scores, with a larger area/thicker line representing stronger prediction/interaction.

mice, *ORMDL3* expression has been shown to be induced by allergen, IL-4 and IL-13 *via* STAT6 and in bronchial epithelial cells, overexpression of *ORMDL3* has been shown to trigger activating transcription factor 6 (ATF6), which has also been implicated in airway remodeling in asthma (80). However, the role of *ORMDL3* in ER driven asthma pathology remains to be resolved (81, 82).

Analyses of 42 Genes Identified in the Current Study Provide Significant Opportunities for Drug Development and Or/Repurposing

Of the 37 candidate genes and five strongest predicted genes, drugs for *GATA3, IL33, IL1RL1, SMAD3,* and *TSLP* are either already in use or in clinical trials for asthma, clearly validating these genes from our pipeline as therapeutic targets. Drugs supported by genetic evidence are twice as likely to go through the drug development pipeline, be successful in Phases II and III and ultimately go into clinic (83). Drugs that target *MUC5AC* (ENSITUXIMAB) and *STAT6* (CHEMBL363332 and CHEMBL1374370) have emerged as potential avenues of future research *via* repurposing for severe asthma and show these two proteins are viable therapeutic targets. The *PDCD1* and *CD274* gene had the largest number of drug associations, albeit predominantly chemotherapy drugs. Although not a prominent gene to come out of this analysis, *PDCD1* has been shown in a small study to be increased in asthma patients after whole lung allergen challenge (84). However, caution must be taken when interpreting drug-gene interactions. For

TABLE 7 | Known drug interactions and molecules currently in clinical trials of 37 candidate causal genes and five predicted genes.

Gene	Drug	Interaction type/company/disorder	PMIDs	Interaction score	clinicaltrials.gov
AAGAB	-	-	-	-	-
BACH2	-	-	-	-	-
CD247	MUROMO-B-CD3	-	2503348, 8861551	15.95	-
CD247	BLI-TUMOMAB	-	-	3.04	-
CD247	AZACITIDINE	-	15795105	1.25	-
CD274	AVELUMAB	Antibody (inhibitory) Clinical Trials	28472902, 26918451, 23724846	23.92	*
CD274	ATEZOLIZUMAB	Antibody (inhibitory) Clinical Trials	24403232, 24829850, 28472902, 26970723, 26918451, 26952546, 27412122	10.63	*
CD274	DURVALUMAB	Antibody (inhibitory) Clinical Trials	28472902, 29416316, 28717238, 28214651	7.97	*
CD274	LODAPOLIMAB	Antibody (inhibitory)	-	5.32	-
CD274	CX-072	Clinical Trials/ CytomX Therapeutics/Solid tumor	-	5.32	NCT03993379, NCT03013491, NCT04596150
CD274	BMS-936559	Antibody (inhibitory) Clinical trials/Bristol-Myers Squibb/cancer and NIAID/HIV	-	5.32	NCT01452334, NCT01455103, NCT02576457, NCT02028403
CD274	ENVAFOLIMAB	Antibody (inhibitory) Clinical Trial/ Tracon Pharmaceuticals Inc./ Undifferentiated Pleomorphic Sarcoma|Myxofibrosarcoma	-	5.32	NCT04480502
CD274	PIDILIZUMAB	Antibody (inhibitory) Clinical trial/ Northwestern University|Gateway for Cancer Research|National Cancer Institute (NCI)/ Lymphoma	-	2.66	NCT02530125
CD274	NIVOLUMAB	Antibody (inhibitory) Clinical Trials	26086854, 28472902, 26918451	2.48	*
CD274	PEMBROLIZUMAB	Antibody (inhibitory) Clinical Trials	25891174, 28472902, 27718847, 26918451, 26086854	2.17	*
CD274	M-7824	Clinical Trials	-	1.77	*
CLEC16A	-	-	-	-	-
D2HGDH	-	-	-	-	-
DEXI	-	-	-	-	-
FLG	-	-	-	-	-
GATA3	PEGASPARGASE	-	24141364	2.45	-
GATA3	SB010	Clinical Trial/ Sterna Biologicals GmbH & Co. KG/Asthma	-	-	NCT01743768
GATA3	SB011	Clinical Trial/ Sterna Biologicals GmbH & Co. KG/ Mild to Moderate Atopic Dermatitis	-	-	NCT02079688
GATA3	SB012	Clinical Trial/ Sterna Biologicals GmbH & Co. KG/Ulcerative Colitis	-	-	NCT02129439
GNGT2	-	-	-	-	-

(Continued)

TABLE 7 | Continued

Gene	Drug	Interaction type/company/disorder	PMIDs	Interaction score	clinicaltrials.gov
GSDMB	-	-	-	-	-
HLA-DQA1	LUMIRACOXIB	-	20639878	7.97	-
HLA-DQA1	LAPATINIB	-	24687830, 21245432	2.28	-
HLA-DQA1	AZATHIOPRINE	-	25217962	1.28	-
HLA-DQA2	-	-	-	-	-
HLA-DQB1	LUMIRACOXIB	-	20639878	2.66	-
HLA-DQB1	BUCILLAMINE	-	-	2.66	-
HLA-DQB1	CLAVULANIC ACID	-	10535882, 30664875	2.28	-
HLA-DQB1	FLOXACILLIN	-	30664875	1.52	-
HLA-DQB2	-	-	-	-	-
IL18R1	-	-	-	-	-
IL1RL1	MSTT1041A	Clinical Trial/University of Leicester/COPD	-	-	NCT03615040
IL1RL1	GSK3772847	Clinical Trial/GSK/Moderately Severe Asthma			NCT03207243
IL33	ITEPEKIMAB	Clinical Trial/Regerenron/COPD	-	-	NCT04701983
IL33	MSTT1041A	Clinical Trial/ Hoffmann-La Roche/Severe Asthma	-	-	NCT02918019
IL33	MEDI3506	Clinical Trial/AstraZenica/Uncontrolled Moderate-Severe Asthma	-	-	NCT04570657
IL33	ANB020	Clinical Trial/University of Leicester/AnaptysBio/Asthma	-	-	NCT04256044
IL37	-	-	-	-	-
IRAK4	PF-06650833	Inhibitory Clinical Trial/Yale University/COVID-19 Clinical Trial/Pfizer/rheumatoid arthritis	-	3.36	NCT04575610, NCT02996500
IRAK4	KT-474	Small molecule degrader Clinical Trial/ Kymera Therapeutics Inc./Atopic Dermatitis, Hidradenitis Suppurativa	-	-	NCT04772885
IRAK4	CA-4948	Inhibitory Clinical Trial/Curis Inc./Acute Myelogenous Leukemia, Myelodysplastic Syndrome/ Relapsed Hematologic Malignancy, Refractory Hematologic Malignancy	-	-	NCT04278768, NCT03328078
IRF1	-	-	-	-	-
KIF1A	-	-	-	-	-
KIAA1109	-	-	-	-	-
LRRC32	-	-	-	-	-
MSL1	-	-	-	-	-
MUC5AC	ENSITUXIMAB	-	-	31.9	-
ORMDL3	-	-	-	-	-
PDCD1	CEMIPLIMAB	Antibody (inhibitory), inhibitor (inhibitory)	29863979, 29089720	9.11	-
PDCD1	SPARTALIZUMAB	Antibody (inhibitory)	-	9.11	-
PDCD1	TISLELIZUMAB	Antibody (inhibitory)	-	9.11	-
PDCD1	PIDILIZUMAB	Antibody (inhibitory)	-	6.83	-
PDCD1	AMP-224	Antibody (inhibitory)	-	4.56	-
PDCD1	MGA-012	-	-	4.56	-
PDCD1	BALSTILIMAB	-	-	4.56	-
PDCD1	SYM-021	-	-	4.56	-
PDCD1	SASANLIMAB	-	-	4.56	-
PDCD1	SINTILIMAB	-	-	4.56	-

(Continued)

TABLE 7 | Continued

Gene	Drug	Interaction type/company/disorder	PMIDs	Interaction score	clinicaltrials.gov
PDCD1	DOSTARLIMAB	-	-	4.56	-
PDCD1	NIVOLUMAB	Inhibitor (inhibitory), antibody (inhibitory)	23289116	2.13	-
PDCD1	M-7824	-	-	1.52	-
PDCD1	PEMBROLIZUMAB	Antibody (inhibitory), antagonist (inhibitory), inhibitor (inhibitory)	25685857	1.04	-
PDCD1	Pembrolizumab/ Vibostolimab Coformulation (MK-7684A)	Clinical Trial/Merck Sharp & Dohme Corp./ Metastatic Non-Small Cell Lung Cancer	-	-	NCT04725188
PDCD1LG2	AMP-224	Antibody	-	15.95	-
PDCD1LG2	PEMBROLIZUMAB	-	28619999	2.9	-
PGAP3	-	-	-	-	-
RORA	CHOLESTEROL	Agonist (activating)	10592235, 17139284, 17016423	31.9	-
RORA	T091317	Agonist (activating)	-	2.66	-
RORA	MELATONIN	-	8702678, 12595746	1.77	-
RPS26	-	-	-	-	-
SLC22A4	IMATINIB	-	23127916, 22875622	2.81	-
SLC22A5	CARNITINE	-	21422191	63.79	-
SLC22A5	IMATINIB	-	28762371, 23127916	1.41	-
SMAD3	VACTOSERTIB	Clinical Trial/ MedPacto, Inc./Solid state tumors	-	-	NCT02160106
STAT6	CHEMBL1374370	-	-	5.32	-
STAT6	CHEMBL363332	-	-	2.66	-
SUOX	-	-	-	-	-
TSLP	MEDI9929	Clinical Trial/ MedImmune LLC/Severe Asthma, Atopic Dermatitis	-	-	NCT02698501, NCT02054130, NCT02512900
TSLP	MRx-4DP0004	Clinical Trial/ 4D pharma plc/Asthma & COVID-19	-	-	NCT03851250, NCT04363372
WDR36	-	-	-	-	-
ZAP70	TRIDOLGOSIR	-	17897956	21.26	
ZAP70	ALOISINE	Inhibitory		10.63	
ZBTB10	-	-	-	-	-
ZNF652	-	-	-	-	-

Interaction scores for known drug interactions with genes highlighted by our signal to gene analysis from the Drug Gene Interaction Database (DGIdb) are listed. The score is a numeric representation of publication count and source count, the ratio of average known gene partners for all drugs to the known partners for the given drug, and the ratio of average known drug partners for all genes to the known partners for the given gene. In interaction score cut off value of >1.0 were selected. Genes with drugs in clinical trials are also listed, CD274 had over 2,000 entries in clinicaltrials.gov and therefore drugs which appeared on the clinical trials database but not the DGIdb have been shown in ***Supplementary Material****.*
Too many clinical trials to list, see* ***Supplementary Material**.*

example, there are many drugs available, which may offer a starting point for targeting of the *HLA-DQA1/B1* genes such as azathioprine, which is an immunosuppressant already in use to treat rheumatoid arthritis, Crohn's disease and ulcerative colitis. However, in the blood cell dataset analysis, *HLA-DQA1* was decreased in asthmatics indicating a lack of gene expression is linked to risk and therefore inhibitory drugs at best would be ineffective and at worst intensify the asthma phenotype.

A Move to Personalized Medicine

For any therapy to be truly effective in a disease such as asthma which is heterogenous, medicines need to be stratified based on evidence that that drug target and/or pathway is driving disease and genetic variants may help identify these individuals. The combination of variants may lead to a particular type of asthma such as IL33 high, which then may be amenable to therapy targeting IL33. Therefore, asthma studies which measure gene expression and also determine genotype offer the most

comprehensive basis for understanding the effects of these signals and genes, as then risk allele, gene expression profile, and asthma phenotype can be bridged. For example, the rs11603634 signal, with candidate gene *MUC5AC*, is inherited in ∼50% of the European population. The airway epithelium datasets analyzed in this study showed increased expression of MUC5AC and a brief analysis in the original GWAS paper has shown the individuals homozygous for the risk allele, A, show increased *MUC5AC* expression compared to homozygous non-risk allele, C, individuals (7). Together these results indicate *MUC5AC* is a strong candidate for inhibitor therapy particularly for individuals that carry the risk allele.

On the other hand, signal rs61816761 (candidate gene *FLG*) is inherited in ∼2% of the European population and was highly specific in only associating with eczema/dermatitis in addition to asthma. This signal includes a loss of function variant and results from this study are consistent with the literature which indicate loss of function variants in the gene are associated with asthma in children with eczema (23). Therefore, therapy for these individuals would require restoring the loss of either FLG itself or of another molecule in its pathway but would probably only be effective in individuals with this variant. Prediction analysis in this study revealed novel physical interactions with *CASP14* and *PCSK6* and co-expressional interactions with *PRL, CASP14,* and *LRRC32*. These interactions may be compromised in subjects with this variant and are options for future exploration.

Therefore, any therapy, including those being investigated for repurposing in asthma, should be stratified and targeted to take into account the individual's genotype. There is previous proof of concept for this e.g., the prevention of asthma exacerbation by the IL4R antagonist, pitrakinra was shown to be effective in carriers of specific IL4R genotypes, although the specific mechanism is not known at this time (85).

CONCLUSIONS

The results from this study replicate and importantly extend previous GWAS translational study findings. A study by Dong et al. which integrated eQTL data and GWAS summary data, using Sherlock Bayesian analysis, identified 11 candidate genes implicated in severe asthma of which four genes (*HLA-DQA1, GNGT2, STAT6,* and *SLC22A5*) overlap with the genes identified through our pipeline (86). Furthermore, a study by El-Husseini et al. which focused on druggable candidate genes identified through GWAS and eQTL analysis also showed overlap with the genes identified in this study notably *SLC22A4* and *SUOX* (87). Collectively, these results provide direct support for the role of 37 candidate causal genes in moderate to severe asthma pathological mechanisms and indirect evidence through interaction for an additional five genes. Overall, the findings demonstrate the contribution of altered airway structural cell and inflammatory cell mechanisms underlying asthma including novel inferences regarding genes not previously identified as potentially contributing to asthma. These genes and pathways can form the basis of novel drug development and/or drug repurposing with the aim to treat moderate to severe asthma where there is an unmet clinical need potentially in a stratified approach using genetics to guide therapy to the individual patient.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: GEO GSE69683; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69683, GEO GSE43696; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43696.

AUTHOR CONTRIBUTIONS

IS conceived the study, contributed to the design, and contributed to writing the manuscript. SH, YG, and IA provided U-BIOPRED datasets and contributed to drafting the manuscript. MP and KR carried out the analysis and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The IS laboratory receives grant support from Asthma UK, British Lung Foundation, NIHR, MRC, GSK, NC3Rs, BBSRC, and Boehringer Ingelheim. The U-BIOPRED initiative was funded by Innovative Medicines Initiative (EU-IMI 115010). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 16/06/2020.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/falgy.2021.738741/full#supplementary-material

REFERENCES

1. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet.* (2017) 390:1211–59. doi: 10.1016/S0140-6736(17)32154-2
2. Hekking PW, Wener RR, Amelink M, Zwinderman AH, Bouvy ML, Bel EH. The prevalence of severe refractory asthma. *J Allergy Clin Immunol.* (2015) 135:896–902. doi: 10.1016/j.jaci.2014.08.042
3. 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) 363:1211–21. doi: 10.1056/NEJMoa0906312

4. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet.* (2011) 43:887–92. doi: 10.1038/ng.888
5. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet.* (2018) 50:42–53. doi: 10.1038/s41588-017-0014-7
6. Akhabir L, Sandford AJ. Genome-wide association studies for discovery of genes involved in asthma. *Respirology.* (2011) 16:396–406. doi: 10.1111/j.1440-1843.2011.01939.x
7. 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) 7:20–34. doi: 10.1016/S2213-2600(18)30389-8
8. Kim KW, Ober C. Lessons learned from GWAS of asthma. *Allergy Asthma Immunol Res.* (2019) 11:170–87. doi: 10.4168/aair.2019.11.2.170
9. Vicente CT, Revez JA, Ferreira MAR. Lessons from 10 years of genome-wide association studies of asthma. *Clin Transl Immunol.* (2017) 6:e165. doi: 10.1038/cti.2017.54
10. Claar DD, Larkin EK, Bastarache L, Blackwell TS, Loyd JE, Hartert TV, et al. A phenome-wide association study identifies a novel asthma risk locus near TERC. *Am J Respir Crit Care Med.* (2016) 193:98–100. doi: 10.1164/rccm.201507-1267LE
11. Roden DM. Phenome-wide association studies: a new method for functional genomics in humans. *J Physiol.* (2017) 595:4109–15. doi: 10.1113/JP273122
12. Diogo D, Tian C, Franklin CS, Alanne-Kinnunen M, March M, Spencer CCA, et al. Phenome-wide association studies across large population cohorts support drug target validation. *Nat Commun.* (2018) 9:4285. doi: 10.1038/s41467-018-06540-3
13. Bhatia S, Kleinjan DA. Disruption of long-range gene regulation in human genetic disease: a kaleidoscope of general principles, diverse mechanisms and unique phenotypic consequences. *Hum Genet.* (2014) 133:815–45. doi: 10.1007/s00439-014-1424-6
14. 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) 337:1190–5. doi: 10.1126/science.1222794
15. Li X, Hastie AT, Hawkins GA, Moore WC, Ampleford EJ, Milosevic J, et al. eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. *Allergy.* (2015) 70:1309–18. doi: 10.1111/all.12683
16. Ghosh D, Ding L, Bernstein JA, Mersha TB. The utility of resolving asthma molecular signatures using tissue-specific transcriptome data. *G3.* (2020) 10:4049–62. doi: 10.1534/g3.120.401718
17. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet.* (2018) 50:1593–9. doi: 10.1038/s41588-018-0248-z
18. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* (2015) 12:e1001779. doi: 10.1371/journal.pmed.1001779
19. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, et al. Open Targets Platform: new developments and updates 2 years on. *Nucleic Acids Res.* (2019) 47:D1056–D65. doi: 10.1093/nar/gky1133
20. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* (2012) 40:D930–4. doi: 10.1093/nar/gkr917
21. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* (2015) 31:3555–7. doi: 10.1093/bioinformatics/btv402
22. Brown SJ, McLean WH. One remarkable molecule: filaggrin. *J Invest Dermatol.* (2012) 132:751–62. doi: 10.1038/jid.2011.393
23. Chan A, Terry W, Zhang H, Karmaus W, Ewart S, Holloway JW, et al. Filaggrin mutations increase allergic airway disease in childhood and adolescence through interactions with eczema and aeroallergen sensitization. *Clin Exp Allergy.* (2018) 48:147–55. doi: 10.1111/cea.13077
24. Hu Y, Comjean A, Mohr SE, FlyBase C, Perrimon N. Gene2 function: an integrated online resource for gene function discovery. *G3.* (2017) 7:2855–8. doi: 10.1534/g3.117.043885
25. UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* (2021) 49:D480–D9.
26. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* (2015) 347:1260419. doi: 10.1126/science.1260419
27. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* (2002) 30:207–10. doi: 10.1093/nar/30.1.207
28. Li X, Hawkins GA, Moore WC, Hastie AT, Ampleford EJ, Milosevic J, et al. Expression of asthma susceptibility genes in bronchial epithelial cells and bronchial alveolar lavage in the Severe Asthma Research Program (SARP) cohort. *J Asthma.* (2016) 53:775–82. doi: 10.3109/02770903.2016.1158268
29. Voraphani N, Gladwin MT, Contreras AU, Kaminski N, Tedrow JR, Milosevic J, et al. An airway epithelial iNOS-DUOX2-thyroid peroxidase metabolome drives Th1/Th2 nitrative stress in human severe asthma. *Mucosal Immunol.* (2014) 7:1175–85. doi: 10.1038/mi.2014.6
30. Bigler J, Boedigheimer M, Schofield JPR, Skipp PJ, Corfield J, Rowe A, et al. A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts. *Am J Respir Crit Care Med.* (2017) 195:1311–20. doi: 10.1164/rccm.201604-0866OC
31. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* (2009) 4:44–57. doi: 10.1038/nprot.2008.211
32. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* (2009) 37:1–13. doi: 10.1093/nar/gkn923
33. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* (2010) 38:W214–20. doi: 10.1093/nar/gkq537
34. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, et al. DGIdb: mining the druggable genome. *Nat Methods.* (2013) 10:1209–10. doi: 10.1038/nmeth.2689
35. Portelli MA, Hodge E, Sayers I. Genetic risk factors for the development of allergic disease identified by genome-wide association. *Clin Exp Allergy.* (2015) 45:21–31. doi: 10.1111/cea.12327
36. 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) 104:665–84. doi: 10.1016/j.ajhg.2019.02.022
37. 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) 130:861–8 e7. doi: 10.1016/j.jaci.2012.04.041
38. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet.* (2009) 41:216–20. doi: 10.1038/ng.275
39. Bakakos A, Loukides S, Bakakos P. Severe eosinophilic asthma. *J Clin Med.* (2019) 8:91375. doi: 10.3390/jcm8091375
40. Leckie MJ, ten Brinke A, Khan J, Diamant Z, O'Connor BJ, Walls CM, et al. Effects of an interleukin-5 blocking monoclonal antibody on eosinophils, airway hyper-responsiveness, and the late asthmatic response. *Lancet.* (2000) 356:2144–8. doi: 10.1016/S0140-6736(00)03496-6
41. Khorasanizadeh M, Eskian M, Rezaei N. Reductions in eosinophil biomarkers by benralizumab in patients with asthma. *Acta Med Iran.* (2017) 55:352–3. doi: 10.1016/j.rmed.2016.01.003
42. Padilla-Galo A, Levy-Abitbol R, Olveira C, Valencia Azcona B, Perez Morales M, Rivas-Ruiz F, et al. Real-life experience with benralizumab during 6 months. *BMC Pulm Med.* (2020) 20:184. doi: 10.1186/s12890-020-01220-9
43. Virchow JC, McDonald M, Garin M, Korn S. Reslizumab as add-on therapy in patients with refractory asthma. *BMJ Open Respir Res.* (2020) 7:494. doi: 10.1136/bmjresp-2019-000494

44. Carr TF, Zeki AA, Kraft M. Eosinophilic and non-eosinophilic asthma. *Am J Respir Crit Care Med.* (2018) 197:22–37. doi: 10.1164/rccm.201611-2232PP
45. Crisford H, Sapey E, Rogers GB, Taylor S, Nagakumar P, Lokwani R, et al. Neutrophils in asthma: the good, the bad and the bacteria. *Thorax.* (2021). doi: 10.1136/thoraxjnl-2020-215986
46. Das S, Miller M, Broide DH. Chromosome 17q21 genes ORMDL3 and GSDMB in asthma and immune diseases. *Adv Immunol.* (2017) 135:1–52. doi: 10.1016/bs.ai.2017.06.001
47. Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, et al. A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle. *J Allergy Clin Immunol.* (2018) 142:749–64 e3. doi: 10.1016/j.jaci.2017.12.974
48. Connolly DJ, O'Neill LA, McGettrick AF. The GOLD domain-containing protein TMED1 is involved in interleukin-33 signaling. *J Biol Chem.* (2013) 288:5616–23. doi: 10.1074/jbc.M112.403899
49. Makrinioti H, Toussaint M, Jackson DJ, Walton RP, Johnston SL. Role of interleukin 33 in respiratory allergy and asthma. *Lancet Respir Med.* (2014) 2:226–37. doi: 10.1016/S2213-2600(13)70261-3
50. Gordon ED, Palandra J, Wesolowska-Andersen A, Ringel L, Rios CL, Lachowicz-Scroggins ME, et al. IL1RL1 asthma risk variants regulate airway type 2 inflammation. *JCI Insight.* (2016) 1:e87871. doi: 10.1172/jci.insight.87871
51. Portelli MA, Dijk FN, Ketelaar ME, Shrine N, Hankinson J, Bhaker S, et al. Phenotypic and functional translation of IL1RL1 locus polymorphisms in lung tissue and asthmatic airway epithelium. *JCI Insight.* (2020) 5:132446. doi: 10.1172/jci.insight.132446
52. Heijink IH, Kuchibhotla VNS, Roffel MP, Maes T, Knight DA, Sayers I, et al. Epithelial cell dysfunction, a major driver of asthma development. *Allergy.* (2020) 75:1902–17. doi: 10.1111/all.14421
53. Holgate ST, Wenzel S, Postma DS, Weiss ST, Renz H, Sly PD. Asthma. *Nat Rev Dis Primers.* (2015) 1:15025. doi: 10.1038/nrdp.2015.25
54. Bergeron C, Tulic MK, Hamid Q. Airway remodelling in asthma: from benchside to clinical practice. *Can Respir J.* (2010) 17:e85–93. doi: 10.1155/2010/318029
55. Pascual RM, Peters SP. Airway remodeling contributes to the progressive loss of lung function in asthma: an overview. *J Allergy Clin Immunol.* (2005) 116:477–86. doi: 10.1016/j.jaci.2005.07.011
56. Fehrenbach H, Wagner C, Wegmann M. Airway remodeling in asthma: what really matters. *Cell Tissue Res.* (2017) 367:551–69. doi: 10.1007/s00441-016-2566-8
57. Chetta A, Zanini A, Foresi A, Del Donno M, Castagnaro A, D'Ippolito R, et al. Vascular component of airway remodeling in asthma is reduced by high dose of fluticasone. *Am J Respir Crit Care Med.* (2003) 167:751–7. doi: 10.1164/rccm.200207-710OC
58. Hoshino M, Nakamura Y, Sim JJ, Yamashiro Y, Uchida K, Hosaka K, et al. Inhaled corticosteroid reduced lamina reticularis of the basement membrane by modulation of insulin-like growth factor (IGF)-I expression in bronchial asthma. *Clin Exp Allergy.* (1998) 28:568–77. doi: 10.1046/j.1365-2222.1998.00277.x
59. Murdoch JR, Lloyd CM. Chronic inflammation and asthma. *Mutat Res.* (2010) 690:24–39. doi: 10.1016/j.mrfmmm.2009.09.005
60. Young HW, Williams OW, Chandra D, Bellinghausen LK, Perez G, Suarez A, et al. Central role of Muc5ac expression in mucous metaplasia and its regulation by conserved 5' elements. *Am J Respir Cell Mol Biol.* (2007) 37:273–90. doi: 10.1165/rcmb.2005-0460OC
61. Lachowicz-Scroggins ME, Yuan S, Kerr SC, Dunican EM, Yu M, Carrington SD, et al. Abnormalities in MUC5AC and MUC5B protein in airway mucus in asthma. *Am J Respir Crit Care Med.* (2016) 194:1296–9. doi: 10.1164/rccm.201603-0526LE
62. Atherton HC, Jones G, Danahay H. IL-13-induced changes in the goblet cell density of human bronchial epithelial cell cultures: MAP kinase and phosphatidylinositol 3-kinase regulation. *Am J Physiol Lung Cell Mol Physiol.* (2003) 285:L730–9. doi: 10.1152/ajplung.00089.2003
63. Melen E, Kho AT, Sharma S, Gaedigk R, Leeder JS, Mariani TJ, et al. Expression analysis of asthma candidate genes during human and murine lung development. *Respir Res.* (2011) 12:86. doi: 10.1186/1465-9921-12-86
64. Acevedo N, Saaf A, Soderhall C, Melen E, Mandelin J, Pietras CO, et al. Interaction between retinoid acid receptor-related orphan receptor alpha (RORA) and neuropeptide S receptor 1 (NPSR1) in asthma. *PLoS ONE.* (2013) 8:e60111. doi: 10.1371/journal.pone.0060111
65. Kuwahara M, Ise W, Ochi M, Suzuki J, Kometani K, Maruyama S, et al. Bach2-Batf interactions control Th2-type immune response by regulating the IL-4 amplification loop. *Nat Commun.* (2016) 7:12596. doi: 10.1038/ncomms12596
66. Yang L, Chen S, Zhao Q, Sun Y, Nie H. The critical role of Bach2 in shaping the balance between CD4(+) T cell subsets in immune-mediated diseases. *Mediators Inflamm.* (2019) 2019:2609737. doi: 10.1155/2019/2609737
67. Pelaia G, Vatrella A, Busceti MT, Gallelli L, Calabrese C, Terracciano R, et al. Cellular mechanisms underlying eosinophilic and neutrophilic airway inflammation in asthma. *Mediators Inflamm.* (2015) 2015:879783. doi: 10.1155/2015/879783
68. Marone G, Spadaro G, Braile M, Poto R, Criscuolo G, Pahima H, et al. Tezepelumab: a novel biological therapy for the treatment of severe uncontrolled asthma. *Expert Opin Investig Drugs.* (2019) 28:931–40. doi: 10.1080/13543784.2019.1672657
69. Lambrecht BN, Hammad H. The immunology of asthma. *Nat Immunol.* (2015) 16:45–56. doi: 10.1038/ni.3049
70. Zhu G, Whyte MK, Vestbo J, Carlsen K, Carlsen KH, Lenney W, et al. Interleukin 18 receptor 1 gene polymorphisms are associated with asthma. *Eur J Hum Genet.* (2008) 16:1083–90. doi: 10.1038/ejhg.2008.67
71. Nakanishi K, Yoshimoto T, Tsutsui H, Okamura H. Interleukin-18 regulates both Th1 and Th2 responses. *Annu Rev Immunol.* (2001) 19:423–74. doi: 10.1146/annurev.immunol.19.1.423
72. Dinarello CA. IL-18: A TH1-inducing, pro-inflammatory cytokine and new member of the IL-1 family. *J Allergy Clin Immunol.* (1999) 103:11–24. doi: 10.1016/S0091-6749(99)70518-X
73. Yoshimoto T, Mizutani H, Tsutsui H, Noben-Trauth N, Yamanaka K, Tanaka M, et al. IL-18 induction of IgE: dependence on CD4+ T cells, IL-4 and STAT6. *Nat Immunol.* (2000) 1:132–7. doi: 10.1038/77811
74. Smith-Garvin JE, Koretzky GA, Jordan MS. T cell activation. *Annu Rev Immunol.* (2009) 27:591–619. doi: 10.1146/annurev.immunol.021908.132706
75. Gelfand EW, Joetham A, Wang M, Takeda K, Schedel M. Spectrum of T-lymphocyte activities regulating allergic lung inflammation. *Immunol Rev.* (2017) 278:63–86. doi: 10.1111/imr.12561
76. Eldor R, Klieger Y, Sade-Feldman M, Vaknin I, Varfolomeev I, Fuchs C, et al. CD247, a novel T cell-derived diagnostic and prognostic biomarker for detecting disease progression and severity in patients with type 2 diabetes. *Diabetes Care.* (2015) 38:113–8. doi: 10.2337/dc14-1544
77. Kontakioti E, Domvri K, Papakosta D, Daniilidis M. HLA and asthma phenotypes/endotypes: a review. *Hum Immunol.* (2014) 75:930–9. doi: 10.1016/j.humimm.2014.06.022
78. Pathinayake PS, Hsu AC, Waters DW, Hansbro PM, Wood LG, Wark PAB. Understanding the unfolded protein response in the pathogenesis of asthma. *Front Immunol.* (2018) 9:175. doi: 10.3389/fimmu.2018.00175
79. Miao K, Zhang L, Pan T, Wang Y. Update on the role of endoplasmic reticulum stress in asthma. *Am J Transl Res.* (2020) 12:1168–83.
80. Miller M, Tam AB, Cho JY, Doherty TA, Pham A, Khorram N, et al. ORMDL3 is an inducible lung epithelial gene regulating metalloproteases, chemokines, OAS, and ATF6. *Proc Natl Acad Sci USA.* (2012) 109:16648–53. doi: 10.1073/pnas.1204151109
81. Hsu KJ, Turvey SE. Functional analysis of the impact of ORMDL3 expression on inflammation and activation of the unfolded protein response in human airway epithelial cells. *Allergy Asthma Clin Immunol.* (2013) 9:4. doi: 10.1186/1710-1492-9-4
82. Debeuf N, Zhakupova A, Steiner R, Van Gassen S, Deswarte K, Fayazpour F, et al. The ORMDL3 asthma susceptibility gene regulates systemic ceramide levels without altering key asthma features in mice. *J Allergy Clin Immunol.* (2019) 144:1648–59 e9. doi: 10.1016/j.jaci.2019.06.041
83. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support

for drug mechanisms on the probability of drug approval. *PLoS Genet*. (2019) 15:e1008489. doi: 10.1371/journal.pgen.1008489

84. Esnault S, Kelly EA, Schwantes EA, Liu LY, DeLain LP, Hauer JA, et al. Identification of genes expressed by human airway eosinophils after an *in vivo* allergen challenge. *PLoS ONE*. (2013) 8:e67560. doi: 10.1371/journal.pone.0067560
85. Slager RE, Otulana BA, Hawkins GA, Yen YP, Peters SP, Wenzel SE, et al. IL-4 receptor polymorphisms predict reduction in asthma exacerbations during response to an anti-IL-4 receptor alpha antagonist. *J Allergy Clin Immunol*. (2012) 130:516–22 e4. doi: 10.1016/j.jaci.2012.03.030
86. Dong Z, Ma Y, Zhou H, Shi L, Ye G, Yang L, et al. Integrated genomics analysis highlights important SNPs and genes implicated in moderate-to-severe asthma based on GWAS and eQTL datasets. *BMC Pulm Med*. (2020) 20:270. doi: 10.1186/s12890-020-01303-7
87. El-Husseini ZW, Gosens R, Dekker F, Koppelman GH. The genetics of asthma and the promise of genomics-guided drug target discovery. *Lancet Respir Med*. (2020) 8:1045–56. doi: 10.1016/S2213-2600(20)30363-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Portelli, Rakkar, Hu, Guo, Adcock and Sayers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

SUPPLEMENTARY MATERIAL

Translational analysis of moderate to severe asthma GWAS signals into candidate causal genes and their functional, tissue-dependent and disease-related associations

Michael A. Portelli[†1], Kamini Rakkar[†1*], Sile Hu [2], Yike Guo [2], Ian M. Adcock [3] on behalf of the U-BIOPRED Study Group & Ian Sayers[1]

[†]These authors have contributed equally to this work and share first authorship

[1]Centre for Respiratory Research, Translational Medical Sciences, School of Medicine, National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham University Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom

[2]Data Science Institute, Imperial College London, London, United Kingdom

[3]The National Heart and Lung Institute, Imperial College, London, United Kingdom

***Correspondence:** Kamini Rakkar, kamini.rakkar2@nottingham.ac.uk

Keywords: GWAS, causal genes, eQTL, SNP, moderate to severe asthma

Running Title: Asthma GWAS and causal genes

INDEX

Table/Figure	Title	Page No.
Table S1.	GeneATLAS PheWAS for each signal.	3
Table S2.	Protein Atlas data for tissue/cell enriched or enhanced gene and protein expression.	24
Table S3.	DAVID Functional Annotation Tool analysis of the Genetic Association Database of complex diseases and disorders (GAD).	26
Table S4.	Frequency of genes in GAD Terms.	27
Table S5.	DAVID Functional Annotation Tool analysis of GO Terms.	28
Table S6.	DAVID Functional Annotation Tool analysis of the KEGG/REACTOME Pathway	29
Table S7.	GeneMANIA predicted gene scores	30
Figure S1.	mRNA expression of candidate causal genes in bronchial epithelial cells taken from patients with asthma and controls	31
Figure S2.	mRNA expression of candidate causal genes in blood taken from patients with asthma and control subjects.	32

Supplementary Table S1. GeneATLAS PheWAS for each signal.

Proxies were used for the following signals which were not present in the database: rs61816761 (rs61816766, R^2 = 0.50), rs367983479 (rs1504215, R^2 = 0.85), rs71266076 (rs7824993, R^2 = 0.81), rs7305461 (rs1131017, R^2 = 0.75), rs112502960 (rs62076439, R^2 = 1.0) and rs61840192 (rs1031163, R^2 = 1.0). Signals rs776111176 and rs560026225 (no proxies available) were not found and have not been displayed.

rs7523907				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	0.0240	2.34E-14	-	T
asthma	0.0047	9.28E-14	1.05	T
allergy/hypersensitivity/anaphylaxis	0.0033	8.50E-10	1.05	T
hayfever/allergic rhinitis	0.0027	9.07E-09	1.05	T
Eosinophill count	0.0013	1.67E-08	-	T
J40-J47 Chronic lower respiratory diseases	0.0024	1.62E-05	1.03	T
J45 Asthma	0.0021	1.95E-05	1.04	T
rs12479210				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	0.0962	4.05E-197	-	T
Eosinophill count	0.0066	8.19E-180	-	T
asthma	0.0104	4.97E-58	1.11	T
J45 Asthma	0.0052	3.99E-25	1.09	T
J40-J47 Chronic lower respiratory diseases	0.0057	9.06E-24	1.08	T
hayfever/allergic rhinitis	0.0044	4.89E-20	1.09	T
allergy/hypersensitivity/anaphylaxis	0.0045	2.16E-16	1.07	T
Neutrophill percentage	-0.1128	4.91E-13	-	T
J33 Nasal polyp	0.0013	1.88E-11	1.18	T
Mean sphered cell volume	0.0402	4.54E-06	-	T
Ease of skin tanning	-0.0078	6.54E-06	-	T
Mean corpuscular volume	0.0296	5.15E-05	-	T

emphysema/chronic bronchitis	0.0012	5.57E-05	1.06	T
rs34290285				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	-0.0767	1.37E-101	-	A
Eosinophill count	-0.0053	1.53E-95	-	A
asthma	-0.0107	2.06E-49	0.90	A
J45 Asthma	-0.0059	1.56E-25	0.91	A
J40-J47 Chronic lower respiratory diseases	-0.0064	3.97E-24	0.92	A
hayfever/allergic rhinitis	-0.0047	1.08E-18	0.92	A
allergy/hypersensitivity/anaphylaxis	-0.0049	1.32E-15	0.93	A
Lymphocyte percentage	0.0733	9.26E-07	-	A
eczema/dermatitis	-0.0017	5.32E-06	0.93	A
rs1837253				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	0.0592	3.04E-62	-	C
Eosinophill count	0.0043	4.69E-62	-	C
asthma	0.0110	6.85E-53	1.11	C
J33 Nasal polyp	0.0023	4.55E-28	1.34	C
J45 Asthma	0.0058	2.82E-25	1.10	C
J40-J47 Chronic lower respiratory diseases	0.0062	1.81E-23	1.09	C
hayfever/allergic rhinitis	0.0049	3.81E-20	1.10	C
allergy/hypersensitivity/anaphylaxis	0.0054	1.68E-18	1.08	C
J30-J39 Other diseases of upper respiratory tract	0.0024	4.18E-08	1.07	C
Number of self-reported non-cancer illnesses	0.0216	7.43E-08	-	C
nasal/sinus disorder	0.0014	1.49E-06	1.10	C
rs1438673				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill count	-0.0038	3.60E-65	-	T
Eosinophill percentage	-0.0526	5.94E-64	-	T
asthma	-0.0074	6.25E-32	0.93	T

hayfever/allergic rhinitis	-0.0043	4.31E-20	0.92	T
J45 Asthma	-0.0044	2.66E-19	0.93	T
allergy/hypersensitivity/anaphylaxis	-0.0043	4.27E-16	0.94	T
J40-J47 Chronic lower respiratory diseases	-0.0039	7.12E-13	0.95	T
J33 Nasal polyp	-0.0012	7.20E-11	0.86	T
Monocyte percentage	0.0184	9.85E-07	-	T
rs3749833				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Mean platelet (thrombocyte) volume	-0.0285	7.72E-57	-	C
Eosinophill count	0.0036	5.38E-46	-	C
Eosinophill percentage	0.0436	8.07E-35	-	C
asthma	0.0084	8.54E-32	1.09	C
Neutrophill count	0.0267	5.87E-21	-	C
Lymphocyte percentage	-0.1356	4.37E-20	-	C
Standing height	-0.0860	1.25E-19	-	C
White blood cell (leukocyte) count	0.0319	1.28E-19	-	C
Platelet count	0.8731	2.52E-17	-	C
Comparative height size at age 10	-0.0089	5.39E-12	-	C
J40-J47 Chronic lower respiratory diseases	0.0041	3.58E-11	1.06	C
J45 Asthma	0.0036	1.86E-10	1.06	C
Leg predicted mass (right)	-0.0125	3.48E-09	-	C
Leg fat-free mass (right)	-0.0133	3.83E-09	-	C
Leg fat-free mass (left)	-0.0130	1.08E-08	-	C
Leg predicted mass (left)	-0.0121	1.21E-08	-	C
Whole body water mass	-0.0467	2.34E-08	-	C
M72 Fibroblastic disorders	0.0013	2.42E-08	1.14	C
Neutrophill percentage	0.0943	4.46E-08	-	C
Whole body fat-free mass	-0.0618	5.07E-08	-	C
Basal metabolic rate	-8.2019	1.05E-07	-	C
Sitting height	-0.0337	1.29E-07	-	C

Number of self-reported non-cancer illnesses	0.0209	1.64E-07	-	C
J33 Nasal polyp	0.0011	1.87E-07	1.15	C
Trunk fat-free mass	-0.0300	2.44E-07	-	C
Trunk predicted mass	-0.0286	2.84E-07	-	C
Weight	-0.1211	5.03E-06	-	C
Weight	-0.1207	6.26E-06	-	C
Monocyte count	0.0014	7.38E-06	-	C
Mean corpuscular haemoglobin	-0.0141	1.55E-05	-	C
Basophill count	0.0004	1.61E-05	-	C
Arm fat-free mass (left)	-0.0037	1.99E-05	-	C
Arm predicted mass (left)	-0.0035	2.05E-05	-	C
I10 Essential (primary) hypertension	0.0035	2.58E-05	1.02	C
Hand grip strength (right)	-0.0659	3.44E-05	-	C
I10-I15 Hypertensive diseases	0.0034	3.90E-05	1.02	C
Arm fat-free mass (right)	-0.0033	5.69E-05	-	C
rs1986009				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	0.1132	1.06E-176	-	A
Eosinophill count	0.0080	6.13E-169	-	A
asthma	0.0094	6.50E-31	1.10	A
J45 Asthma	0.0051	1.14E-15	1.09	A
Neutrophill percentage	-0.1557	1.25E-15	-	A
J40-J47 Chronic lower respiratory diseases	0.0052	2.16E-13	1.07	A
Lymphocyte count	0.0089	2.37E-10	-	A
Standing height	-0.0642	1.99E-09	-	A
Mean platelet (thrombocyte) volume	-0.0111	4.22E-08	-	A
psoriasis	-0.0014	2.96E-07	0.88	A
eczema/dermatitis	0.0021	9.63E-07	1.09	A
White blood cell (leukocyte) count	0.0183	4.37E-06	-	A
Immature reticulocyte fraction	-0.0006	3.51E-05	-	A

thyroid problem (not cancer)	-0.0024	4.81E-05	0.96	A
rs9273410				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
malabsorption/coeliac disease	0.0061	< 1E-300	4.13	A
K90 Intestinal malabsorption	0.0059	< 1E-300	3.24	A
E10 Insulin-dependent diabetes mellitus	0.0034	7.69E-77	1.65	A
Asthma	0.0124	7.55E-76	1.13	A
thyroid problem (not cancer)	0.0074	6.26E-52	1.14	A
K90-K93 Other diseases of the digestive system	0.0057	1.44E-49	1.21	A
hypothyroidism/myxoedema	0.0057	4.87E-37	1.13	A
J45 Asthma	0.0065	3.55E-35	1.12	A
E00-E07 Disorders of thyroid gland	0.0052	6.56E-34	1.14	A
E03 Other hypothyroidism	0.0046	4.41E-32	1.15	A
Mean reticulocyte volume	0.1659	4.41E-32	-	A
hyperthyroidism/thyrotoxicosis	0.0022	2.42E-31	1.34	A
J40-J47 Chronic lower respiratory diseases	0.0066	1.59E-29	1.09	A
Mean sphered cell volume	0.0990	8.55E-27	-	A
Red blood cell (erythrocyte) count	-0.0066	3.35E-25	-	A
Diabetes	0.0046	7.88E-25	1.11	A
Lymphocyte count	-0.0117	1.33E-23	-	A
Haemoglobin concentration	-0.0183	4.59E-23	-	A
E10-E14 Diabetes mellitus	0.0044	9.35E-23	1.10	A
E16 Other disorders of pancreatic internal secretion	0.0010	2.97E-20	1.55	A
E15-E16 Other disorders of glucose regulation and pancreatic internal secretion	0.0010	6.97E-20	1.54	A
H36 Retinal disorders in diseases classified elsewhere	0.0011	2.82E-19	1.42	A
Mean platelet (thrombocyte) volume	0.0146	2.55E-18	-	A
Reticulocyte count	-0.0004	1.12E-17	-	A
E05 Thyrotoxicosis [hyperthyroidism]	0.0012	2.54E-15	1.30	A
Number of self-reported non-cancer illnesses	0.0285	2.34E-14	-	A

Psoriasis	-0.0017	1.58E-13	0.86	A
bowel problem	0.0039	1.46E-12	1.06	A
E14 Unspecified diabetes mellitus	0.0012	1.93E-12	1.23	A
Haematocrit percentage	-0.0379	3.54E-12	-	A
White blood cell (leukocyte) count	-0.0229	4.66E-12	-	A
High light scatter reticulocyte count	-0.0001	7.57E-12	-	A
Reticulocyte percentage	-0.0070	9.00E-12	-	A
Number of treatments/medications taken	0.0357	1.28E-11	-	A
Monocyte count	-0.0020	1.94E-11	-	A
Mean corpuscular haemoglobin concentration	-0.0127	3.13E-11	-	A
allergy/hypersensitivity/anaphylaxis	0.0037	9.99E-11	1.06	A
D64 Other anaemias	0.0024	1.00E-10	1.09	A
D60-D64 Aplastic and other anaemias	0.0024	1.68E-10	1.09	A
Mean corpuscular volume	0.0492	3.92E-10	-	A
Dermatology	-0.0025	8.88E-09	0.94	A
E11 Non-insulin-dependent diabetes mellitus	0.0025	1.03E-08	1.06	A
High light scatter reticulocyte percentage	-0.0021	3.69E-08	-	A
hayfever/allergic rhinitis	0.0027	4.07E-08	1.05	A
K51 Ulcerative colitis	-0.0011	4.51E-08	0.87	A
Lymphocyte percentage	-0.0755	4.85E-08	-	A
Red blood cell (erythrocyte) distribution width	0.0093	1.46E-07	-	A
Neutrophill percentage	0.0844	1.69E-07	-	A
H30-H36 Disorders of choroid and retina	0.0015	5.95E-07	1.09	A
connective tissue disorder	0.0009	6.16E-07	1.15	A
inflammatory bowel disease	-0.0010	1.14E-06	0.89	A
Vasculitis	0.0006	1.55E-06	1.22	A
diabetic eye disease	0.0005	2.83E-06	1.27	A
D50-D53 Nutritional anaemias	0.0015	2.99E-06	1.08	A
bone disorder	0.0015	3.30E-06	1.07	A
L40 Psoriasis	-0.0008	3.73E-06	0.87	A

L97 Ulcer of lower limb, not elsewhere classified	0.0005	3.79E-06	1.23	A
chronic/degenerative neurological problem	-0.0008	4.73E-06	0.88	A
Anaemia	0.0010	5.60E-06	1.11	A
C44 Other malignant neoplasms of skin	-0.0018	6.36E-06	0.95	A
C43-C44 Melanoma and other malignant neoplasms of skin	-0.0019	8.37E-06	0.95	A
I77 Other disorders of arteries and arterioles	0.0006	1.39E-05	1.18	A
G35 Multiple sclerosis	-0.0005	1.85E-05	0.84	A
D50 Iron deficiency anaemia	0.0013	2.76E-05	1.07	A
G35-G37 Demyelinating diseases of the central nervous system	-0.0006	3.73E-05	0.86	A
J33 Nasal polyp	0.0008	3.77E-05	1.11	A
Platelet count	-0.4016	3.81E-05	-	A
Haematology	0.0011	5.01E-05	1.08	A
M81 Osteoporosis without pathological fracture	0.0011	5.11E-05	1.08	A
rs144829310				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	0.1473	3.51E-268	-	T
Eosinophill count	0.0104	5.67E-257	-	T
Asthma	0.0144	2.41E-63	1.15	T
J33 Nasal polyp	0.0026	1.02E-26	1.41	T
J45 Asthma	0.0068	2.07E-24	1.12	T
J40-J47 Chronic lower respiratory diseases	0.0075	9.26E-24	1.11	T
hayfever/allergic rhinitis	0.0048	3.85E-14	1.09	T
allergy/hypersensitivity/anaphylaxis	0.0048	4.54E-11	1.07	T
J30-J39 Other diseases of upper respiratory tract	0.0032	7.18E-10	1.10	T
Neutrophill percentage	-0.1201	4.75E-09	-	T
nasal/sinus disorder	0.0017	2.43E-07	1.13	T
Ease of skin tanning	-0.0104	5.63E-06	-	T
J32 Chronic sinusitis	0.0009	2.44E-05	1.17	T

rs10905284				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	-0.0348	3.19E-28	-	A
Eosinophill count	-0.0024	3.84E-25	-	A
Lymphocyte percentage	0.0857	8.46E-11	-	A
asthma	-0.0037	1.16E-08	0.97	A
Lymphocyte count	0.0055	8.62E-07	-	A
J45 Asthma	-0.0024	1.56E-06	0.96	A
J40-J47 Chronic lower respiratory diseases	-0.0026	4.15E-06	0.97	A
J33 Nasal polyp	-0.0008	1.46E-05	0.90	A
rs11603634				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
asthma	0.0032	5.10E-07	1.03	G
Eosinophill percentage	0.0150	1.45E-06	-	G
Eosinophill count	0.0010	1.08E-05	-	G
K80-K87 Disorders of gallbladder, biliary tract and pancreas	-0.0017	4.29E-05	0.96	G
Platelet count	-0.3654	6.26E-05	-	G
K80 Cholelithiasis	-0.0015	6.32E-05	0.96	G
rs7936312				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	0.0640	1.63E-94	-	T
Eosinophill count	0.0045	5.44E-89	-	T
asthma	0.0104	1.86E-61	1.11	T
hayfever/allergic rhinitis	0.0049	6.70E-26	1.10	T
allergy/hypersensitivity/anaphylaxis	0.0053	6.21E-23	1.08	T
J45 Asthma	0.0047	1.63E-21	1.08	T
J40-J47 Chronic lower respiratory diseases	0.0048	3.48E-18	1.07	T
dermatology	0.0036	4.16E-18	1.09	T
eczema/dermatitis	0.0028	4.69E-18	1.12	T

Neutrophill percentage	-0.0914	1.57E-09	-	T
Number of self-reported non-cancer illnesses	0.0210	2.27E-09	-	T
Impedance of arm (left)	0.3882	2.38E-09	-	T
Impedance of arm (right)	0.3543	1.52E-08	-	T
inflammatory bowel disease	0.0010	4.20E-08	1.13	T
Impedance of whole body	0.5418	7.21E-07	-	T
K51 Ulcerative colitis	0.0009	1.36E-06	1.12	T
L30 Other dermatitis	0.0006	5.23E-06	1.17	T
Number of treatments/medications taken	0.0212	1.77E-05	-	T
rs10519068				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
asthma	-0.0087	1.47E-20	0.92	A
Eosinophill percentage	-0.0400	3.34E-18	-	A
Eosinophill count	-0.0027	2.92E-16	-	A
hayfever/allergic rhinitis	-0.0043	7.77E-10	0.92	A
allergy/hypersensitivity/anaphylaxis	-0.0042	1.11E-07	0.94	A
J45 Asthma	-0.0033	6.11E-06	0.95	A
rs72743461				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
asthma	0.0110	3.53E-50	1.11	A
Eosinophill percentage	0.0492	1.27E-41	-	A
Eosinophill count	0.0033	2.80E-36	-	A
J45 Asthma	0.0054	1.26E-20	1.10	A
J40-J47 Chronic lower respiratory diseases	0.0058	2.54E-19	1.08	A
hayfever/allergic rhinitis	0.0041	1.35E-13	1.08	A
allergy/hypersensitivity/anaphylaxis	0.0042	2.26E-11	1.06	A
Lymphocyte percentage	-0.0831	4.69E-08	-	A
I25 Chronic ischaemic heart disease	-0.0029	2.05E-07	0.95	A
heart attack/myocardial infarction	-0.0016	4.19E-06	0.93	A
Impedance of whole body	0.5826	5.72E-06	-	A

I20-I25 Ischaemic heart diseases	-0.0027	5.85E-06	0.96	A
E11 Non-insulin-dependent diabetes mellitus	-0.0021	9.53E-06	0.95	A
Lymphocyte count	-0.0056	1.15E-05	-	A
Impedance of arm (right)	0.3224	1.17E-05	-	A
other respiratory problems	0.0011	2.20E-05	1.10	A
Impedance of arm (left)	0.3200	2.79E-05	-	A
E10-E14 Diabetes mellitus	-0.0020	5.08E-05	0.96	A
rs7203459				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill count	-0.0051	9.79E-86	-	C
Eosinophill percentage	-0.0701	3.52E-84	-	C
asthma	-0.0088	2.83E-33	0.92	C
J45 Asthma	-0.0046	1.22E-15	0.93	C
J40-J47 Chronic lower respiratory diseases	-0.0048	8.55E-14	0.94	C
allergy/hypersensitivity/anaphylaxis	-0.0044	2.34E-12	0.94	C
Sitting height	0.0424	5.91E-11	-	C
hayfever/allergic rhinitis	-0.0036	7.10E-11	0.94	C
Lymphocyte count	-0.0077	1.01E-09	-	C
Number of self-reported non-cancer illnesses	-0.0229	2.03E-08	-	C
Hand grip strength (right)	0.0889	4.26E-08	-	C
Neutrophill percentage	0.0936	1.01E-07	-	C
Hand grip strength (left)	0.0858	1.22E-07	-	C
eczema/dermatitis	-0.0020	1.40E-07	0.92	C
Number of operations, self-reported	-0.0164	2.08E-06	-	C
Number of treatments/medications taken	-0.0253	1.08E-05	-	C
White blood cell (leukocyte) count	-0.0155	1.57E-05	-	C
rs2941522				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
White blood cell (leukocyte) count	0.0763	3.59E-133	-	T
Neutrophill count	0.0609	1.13E-129	-	T

Asthma	0.0101	9.93E-58	1.10	T
Monocyte percentage	-0.0597	8.71E-57	-	T
Neutrophill percentage	0.1939	2.59E-37	-	T
Lymphocyte count	0.0125	1.92E-30	-	T
Lymphocyte percentage	-0.1194	4.42E-20	-	T
Basophill count	0.0007	2.16E-19	-	T
J45 Asthma	0.0040	5.06E-16	1.07	T
Mean reticulocyte volume	-0.0966	1.26E-13	-	T
J40-J47 Chronic lower respiratory diseases	0.0040	1.92E-13	1.06	T
hayfever/allergic rhinitis	0.0029	8.03E-10	1.06	T
allergy/hypersensitivity/anaphylaxis	0.0030	2.51E-08	1.04	T
emphysema/chronic bronchitis	0.0016	1.55E-07	1.08	T
Oily fish intake	-0.0092	2.76E-07	-	T
Platelet distribution width	0.0042	1.13E-06	-	T
Eosinophill percentage	-0.0145	3.02E-06	-	T
bronchitis	0.0008	3.18E-06	1.12	T
Eosinophill count	0.0010	3.18E-06	-	T
Number of self-reported non-cancer illnesses	0.0156	9.16E-06	-	T
Platelet count	-0.3829	2.36E-05	-	T
Fresh fruit intake	-0.0121	3.28E-05	-	T
Mean sphered cell volume	-0.0354	3.29E-05	-	T
Mean platelet (thrombocyte) volume	0.0064	4.94E-05	-	T
inflammatory bowel disease	-0.0008	5.98E-05	0.91	T
Salt added to food	0.0066	6.09E-05	-	T
rs703816				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
asthma	0.0072	2.65E-29	1.07	C
Mean platelet (thrombocyte) volume	0.0136	1.93E-16	-	C
J45 Asthma	0.0036	2.73E-13	1.06	C
Eosinophill count	0.0016	1.80E-12	-	C

Sitting height	-0.0400	2.29E-12	-	C
migraine	-0.0025	2.44E-12	0.92	C
Eosinophill percentage	0.0211	2.09E-11	-	C
Standing height	-0.0545	1.04E-10	-	C
J40-J47 Chronic lower respiratory diseases	0.0036	1.19E-10	1.05	C
Platelet distribution width	0.0053	1.06E-09	-	C
neurology	-0.0030	1.01E-08	0.95	C
High light scatter reticulocyte percentage	0.0020	2.54E-08	-	C
High light scatter reticulocyte count	0.0001	7.42E-08	-	C
hayfever/allergic rhinitis	0.0023	1.21E-06	1.04	C
allergy/hypersensitivity/anaphylaxis	0.0026	2.13E-06	1.04	C
Platelet count	-0.4346	2.91E-06	-	C
Salt added to food	0.0075	8.17E-06	-	C
Trunk predicted mass	-0.0215	1.50E-05	-	C
Trunk fat-free mass	-0.0221	1.96E-05	-	C
Alcohol intake frequency.	-0.0117	4.42E-05	-	C
Reticulocyte percentage	0.0040	4.51E-05	-	C
Arm fat percentage (left)	0.0525	5.98E-05	-	C
Arm fat percentage (right)	0.0520	6.38E-05	-	C
rs61816761 (rs61816766 proxy, R^2 = 0.50)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
eczema/dermatitis	0.0175	8.49E-77	2.02	C
dermatology	0.0173	5.81E-47	1.53	C
asthma	0.0163	3.86E-19	1.17	C
Impedance of arm (right)	1.5923	3.45E-18	-	C
Impedance of arm (left)	1.5021	2.86E-15	-	C
Impedance of whole body	1.7950	2.02E-08	-	C
L30 Other dermatitis	0.0019	2.58E-07	1.68	C
J45 Asthma	0.0072	4.36E-07	1.13	C
Hand grip strength (right)	0.1957	1.42E-06	-	C

Hand grip strength (left)	0.1851	5.05E-06	-	C
J40-J47 Chronic lower respiratory diseases	0.0071	7.27E-06	1.10	C
L72 Follicular cysts of skin and subcutaneous tissue	-0.0035	4.15E-05	0.84	C
rs367983479 (rs1504215 proxy, R^2 = 0.85)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill count	-0.0043	7.64E-75	-	A
Eosinophill percentage	-0.0528	1.70E-58	-	A
Mean reticulocyte volume	0.1673	2.27E-33	-	A
Monocyte percentage	0.0458	4.32E-31	-	A
asthma	-0.0075	1.53E-29	0.93	A
Mean sphered cell volume	0.1008	2.37E-28	-	A
thyroid problem (not cancer)	0.0049	6.84E-24	1.09	A
Mean corpuscular volume	0.0744	9.55E-22	-	A
hypothyroidism/myxoedema	0.0042	5.79E-21	1.09	A
White blood cell (leukocyte) count	-0.0289	1.14E-18	-	A
Lymphocyte count	-0.0098	3.02E-17	-	A
Mean corpuscular haemoglobin	0.0255	6.18E-16	-	A
J45 Asthma	-0.0041	4.31E-15	0.93	A
E03 Other hypothyroidism	0.0030	6.81E-15	1.09	A
J40-J47 Chronic lower respiratory diseases	-0.0044	1.89E-14	0.94	A
Red blood cell (erythrocyte) count	-0.0047	1.04E-13	-	A
E00-E07 Disorders of thyroid gland	0.0030	7.11E-13	1.08	A
J33 Nasal polyp	-0.0013	1.31E-11	0.85	A
Platelet crit	-0.0005	7.45E-11	-	A
Neutrophill count	-0.0159	1.79E-09	-	A
C44 Other malignant neoplasms of skin	-0.0023	6.01E-09	0.93	A
C43-C44 Melanoma and other malignant neoplasms of skin	-0.0023	1.68E-08	0.94	A
Platelet count	-0.4646	1.42E-06	-	A
hayfever/allergic rhinitis	-0.0022	6.75E-06	0.96	A

chronic obstructive airways disease/copd	-0.0005	2.32E-05	0.85	A
rs71266076 (rs7824993 proxy, R^2 = 0.81)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
asthma	-0.0068	1.83E-24	0.94	G
Eosinophill percentage	-0.0250	1.81E-14	-	G
Eosinophill count	-0.0017	2.03E-13	-	G
J45 Asthma	-0.0035	6.84E-12	0.94	G
J40-J47 Chronic lower respiratory diseases	-0.0036	5.96E-10	0.95	G
Number of self-reported non-cancer illnesses	-0.0180	1.14E-06	-	G
Number of operations, self-reported	-0.0133	2.17E-05	-	G
Haemoglobin concentration	-0.0074	4.24E-05	-	G
hayfever/allergic rhinitis	-0.0020	5.67E-05	0.96	G
Haematocrit percentage	-0.0214	5.72E-05	-	G
rs7305461 (rs1131017 proxy, R^2 = 0.75)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	-0.0427	3.30E-42	-	G
Eosinophill count	-0.0028	3.92E-34	-	G
Asthma	-0.0061	1.18E-21	0.94	G
Impedance of arm (left)	-0.5935	1.74E-19	-	G
Arm predicted mass (left)	0.0065	4.54E-19	-	G
Impedance of whole body	-0.9859	4.70E-19	-	G
Impedance of arm (right)	-0.5604	8.67E-19	-	G
Arm fat-free mass (left)	0.0068	2.23E-18	-	G
Whole body water mass	0.0627	3.19E-17	-	G
Trunk fat-free mass	0.0435	3.41E-17	-	G
Trunk predicted mass	0.0416	4.16E-17	-	G
Whole body fat-free mass	0.0840	7.30E-17	-	G
Arm fat-free mass (right)	0.0060	2.92E-16	-	G
Arm predicted mass (right)	0.0056	3.95E-16	-	G
Basal metabolic rate	11.0910	5.32E-16	-	G

hypothyroidism/myxoedema	-0.0032	4.23E-14	0.93	G
Leg fat-free mass (right)	0.0148	1.30E-13	-	G
Leg predicted mass (right)	0.0139	1.34E-13	-	G
Impedance of leg (left)	-0.3979	2.19E-13	-	G
Leg fat-free mass (left)	0.0147	2.82E-13	-	G
Leg predicted mass (left)	0.0137	3.61E-13	-	G
Impedance of leg (right)	-0.3912	5.35E-13	-	G
Lymphocyte count	0.0079	7.93E-13	-	G
J45 Asthma	-0.0033	3.01E-11	0.95	G
Weight	0.1574	3.04E-11	-	G
J40-J47 Chronic lower respiratory diseases	-0.0036	1.22E-10	0.95	G
Body mass index (BMI)	0.0517	1.88E-10	-	G
Lymphocyte percentage	0.0827	2.82E-10	-	G
Weight	0.1480	3.16E-10	-	G
thyroid problem (not cancer)	-0.0029	5.72E-10	0.95	G
E03 Other hypothyroidism	-0.0022	2.09E-09	0.94	G
Body mass index (BMI)	0.0482	2.29E-09	-	G
hayfever/allergic rhinitis	-0.0027	1.34E-08	0.95	G
Smoking status	0.0068	1.72E-07	-	G
J33 Nasal polyp	-0.0010	1.74E-07	0.88	G
Time spent watching television (TV)	0.0155	4.27E-07	-	G
Waist circumference	0.1033	8.75E-07	-	G
allergy/hypersensitivity/anaphylaxis	-0.0026	1.42E-06	0.96	G
Processed meat intake	0.0098	1.43E-06	-	G
E00-E07 Disorders of thyroid gland	-0.0018	4.91E-06	0.95	G
Red blood cell (erythrocyte) count	0.0027	6.64E-06	-	G
Arm fat mass (right)	0.0047	8.14E-06	-	G
Hypertension	0.0036	1.08E-05	1.02	G
Arm fat mass (left)	0.0051	1.73E-05	-	G
Leg fat mass (right)	0.0115	2.20E-05	-	G

Hip circumference	0.0646	5.72E-05	-	G
rs112502960 (rs62076439, R^2 = 1.0)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Impedance of leg (left)	-0.7662	4.66E-43	-	T
Standing height	-0.1166	8.19E-42	-	T
Impedance of whole body	-1.5375	8.50E-42	-	T
Impedance of leg (right)	-0.7251	8.82E-39	-	T
Eosinophill percentage	0.0402	1.42E-35	-	T
Eosinophill count	0.0029	1.74E-35	-	T
Impedance of arm (left)	-0.7696	4.36E-30	-	T
Impedance of arm (right)	-0.7018	3.62E-27	-	T
hypertension	0.0083	2.45E-23	1.04	T
asthma	0.0060	1.01E-19	1.06	T
Monocyte count	0.0024	7.77E-17	-	T
I10-I15 Hypertensive diseases	0.0062	3.25E-16	1.04	T
I10 Essential (primary) hypertension	0.0062	3.48E-16	1.04	T
Sitting height	-0.0459	2.88E-15	-	T
Comparative body size at age 10	0.0094	3.37E-13	-	T
Number of treatments/medications taken	0.0343	2.61E-11	-	T
Mean reticulocyte volume	0.0901	2.90E-11	-	T
J40-J47 Chronic lower respiratory diseases	0.0038	4.01E-11	1.05	T
I25 Chronic ischaemic heart disease	0.0032	7.37E-11	1.06	T
I20-I25 Ischaemic heart diseases	0.0034	1.57E-10	1.05	T
J45 Asthma	0.0032	2.53E-10	1.06	T
Monocyte percentage	0.0239	9.11E-10	-	T
angina	0.0022	1.81E-09	1.07	T
Trunk fat percentage	-0.0745	6.34E-09	-	T
Platelet distribution width	-0.0051	1.26E-08	-	T
Mean platelet (thrombocyte) volume	-0.0091	2.51E-08	-	T
E78 Disorders of lipoprotein metabolism and other	0.0031	9.05E-08	1.04	T

lipidaemias				
Haematocrit percentage	0.0279	1.21E-07	-	T
I20 Angina pectoris	0.0022	2.15E-07	1.05	T
Number of self-reported non-cancer illnesses	0.0179	1.01E-06	-	T
heart/cardiac problem	0.0026	1.54E-06	1.04	T
Smoking status	0.0062	3.54E-06	-	T
Immature reticulocyte fraction	0.0005	6.06E-06	-	T
E70-E90 Metabolic disorders	0.0029	6.11E-06	1.03	T
Body fat percentage	-0.0495	8.77E-06	-	T
Red blood cell (erythrocyte) count	0.0027	1.23E-05	-	T
Body mass index (BMI)	0.0361	1.34E-05	-	T
Body mass index (BMI)	0.0363	1.39E-05	-	T
Trunk fat mass	-0.0390	1.46E-05	-	T
heart attack/myocardial infarction	0.0014	1.57E-05	1.06	T
High light scatter reticulocyte count	0.0007	1.64E-05	-	T
Mean time to correctly identify matches	0.9348	2.09E-05	-	T
Haemoglobin concentration	0.0074	3.45E-05	-	T
Time spent watching television (TV)	0.0130	3.98E-05	-	T
rs61840192 (rs1031163, R^2 = 1.0)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	-0.0650	5.66E-95	-	T
Eosinophill count	-0.0046	3.74E-92	-	T
asthma	-0.0096	6.85E-51	0.91	T
J45 Asthma	-0.0061	4.40E-35	0.90	T
J40-J47 Chronic lower respiratory diseases	-0.0062	5.17E-29	0.92	T
J33 Nasal polyp	-0.0013	7.04E-12	0.85	T
hayfever/allergic rhinitis	-0.0032	1.77E-11	0.94	T
allergy/hypersensitivity/anaphylaxis	-0.0033	6.36E-10	0.95	T
Number of self-reported non-cancer illnesses	-0.0165	3.55E-06	-	T
J30-J39 Other diseases of upper respiratory tract	-0.0017	6.19E-06	0.95	T

nasal/sinus disorder	-0.0011	2.28E-05	0.93	T
rs560026225 (rs72687036, R^2 = 0.66)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
Eosinophill percentage	0.0346	1.42E-19	-	G
asthma	0.0066	2.50E-17	1.07	G
Mean platelet (thrombocyte) volume	-0.0160	9.50E-17	-	G
Eosinophill count	0.0022	3.97E-15	-	G
hayfever/allergic rhinitis	0.0033	9.18E-09	1.06	G
J45 Asthma	0.0033	4.65E-08	1.06	G
J40-J47 Chronic lower respiratory diseases	0.0035	2.64E-07	1.05	G
Mean reticulocyte volume	-0.0826	2.72E-07	-	G
Lymphocyte count	-0.0069	3.03E-07	-	G
Sitting height	0.0343	7.20E-07	-	G
Mean sphered cell volume	-0.0495	2.44E-06	-	G
Lymphocyte percentage	-0.0713	9.08E-06	-	G
Standing height	0.0446	1.37E-05	-	G
allergy/hypersensitivity/anaphylaxis	0.0028	1.53E-05	1.04	G
Number of treatments/medications taken	0.0255	3.00E-05	-	G
Mean corpuscular volume	-0.0353	5.45E-05	-	G
K51 Ulcerative colitis	0.0009	6.24E-05	1.12	G
rs776111176 (rs3997872, R^2 = 0.82)				
Trait	**Beta**	**p-value**	**Odds Ratio**	**Effect Allele**
rheumatoid arthritis	0.0077	4.64E-166	2.00	A
M06 Other rheumatoid arthritis	0.0068	3.38E-137	1.92	A
White blood cell (leukocyte) count	0.0870	1.25E-106	-	A
Neutrophill count	0.0680	3.37E-100	-	A
M05 Seropositive rheumatoid arthritis	0.0019	3.12E-84	4.01	A
asthma	0.0154	1.26E-81	1.16	A
E10 Insulin-dependent diabetes mellitus	0.0040	1.07E-76	1.83	A
malabsorption/coeliac disease	-0.0031	5.90E-69	0.49	A

hypothyroidism/myxoedema	0.0093	6.62E-66	1.22	A
Number of treatments/medications taken	0.1040	6.38E-61	-	A
K90 Intestinal malabsorption	-0.0031	7.77E-60	0.54	A
Eosinophill count	0.0044	2.77E-53	-	A
thyroid problem (not cancer)	0.0084	1.19E-46	1.17	A
J45 Asthma	0.0090	3.16E-46	1.16	A
J40-J47 Chronic lower respiratory diseases	0.0093	1.34E-40	1.13	A
diabetes	0.0070	5.03E-39	1.17	A
E03 Other hypothyroidism	0.0059	4.29E-36	1.19	A
vasculitis	0.0018	6.64E-36	1.87	A
M05-M14 Inflammatory polyarthropathies	0.0066	1.35E-33	1.16	A
J33 Nasal polyp	0.0027	1.95E-31	1.42	A
E00-E07 Disorders of thyroid gland	0.0059	4.97E-31	1.16	A
E10-E14 Diabetes mellitus	0.0062	5.88E-31	1.15	A
Lymphocyte percentage	-0.1773	1.16E-26	-	A
Number of self-reported non-cancer illnesses	0.0476	2.61E-26	-	A
H36 Retinal disorders in diseases classified elsewhere	0.0015	1.88E-24	1.62	A
Neutrophill percentage	0.1964	3.45E-24	-	A
Monocyte percentage	-0.0473	5.36E-23	-	A
Mean sphered cell volume	-0.1066	6.22E-22	-	A
M35 Other systemic involvement of connective tissue	0.0016	8.50E-20	1.43	A
joint disorder	0.0078	1.21E-19	1.07	A
Hand grip strength (right)	-0.1525	1.24E-17	-	A
Waist circumference / Hip circumference	0.0013	1.32E-16	-	A
Lymphocyte count	0.0114	4.95E-16	-	A
G35 Multiple sclerosis	-0.0012	6.02E-16	0.68	A
E11 Non-insulin-dependent diabetes mellitus	0.0042	9.19E-16	1.10	A
Reticulocyte count	0.0004	1.02E-15	-	A
chronic/degenerative neurological problem	-0.0017	1.49E-15	0.76	A
Monocyte count	0.0029	1.89E-15	-	A

Reticulocyte percentage	0.0097	2.87E-15	-	A
Hand grip strength (left)	-0.1388	7.06E-15	-	A
Eosinophill percentage	0.0308	8.39E-15	-	A
High light scatter reticulocyte count	0.0002	2.08E-14	-	A
G35-G37 Demyelinating diseases of the central nervous system	-0.0012	2.87E-14	0.71	A
nasal/sinus disorder	0.0023	1.14E-13	1.18	A
High light scatter reticulocyte percentage	0.0033	7.97E-13	-	A
Mean corpuscular haemoglobin	-0.0265	4.04E-12	-	A
Standing height	-0.0746	4.15E-12	-	A
bowel problem	-0.0044	3.44E-11	0.93	A
Sitting height	-0.0476	4.81E-11	-	A
J30-J39 Other diseases of upper respiratory tract	0.0032	5.81E-11	1.10	A
connective tissue disorder	0.0014	8.00E-11	1.24	A
K51 Ulcerative colitis	-0.0015	8.84E-11	0.82	A
ent disorder/not cancer	0.0032	9.43E-11	1.10	A
Platelet count	0.7471	1.64E-10	-	A
K90-K93 Other diseases of the digestive system	-0.0029	4.65E-10	0.91	A
diabetic eye disease	0.0007	1.86E-09	1.45	A
Mean corpuscular volume	-0.0563	2.34E-09	-	A
E16 Other disorders of pancreatic internal secretion	0.0008	2.71E-09	1.40	A
E15-E16 Other disorders of glucose regulation and pancreatic internal secretion	0.0008	2.99E-09	1.40	A
D06 Carcinoma in situ of cervix uteri	0.0019	3.52E-09	1.27	A
Trunk fat-free mass	-0.0384	4.33E-09	-	A
Trunk predicted mass	-0.0364	6.57E-09	-	A
psoriasis	-0.0016	2.04E-08	0.87	A
inflammatory bowel disease	-0.0013	3.32E-08	0.85	A
E14 Unspecified diabetes mellitus	0.0011	3.50E-08	1.22	A
Basophill count	0.0005	7.59E-08	-	A

M31 Other necrotising vasculopathies	0.0005	8.07E-08	1.47	A
sarcoidosis	-0.0006	1.01E-07	0.72	A
K57 Diverticular disease of intestine	0.0034	1.16E-07	1.06	A
Whole body water mass	-0.0494	1.52E-07	-	A
Platelet distribution width	0.0057	1.85E-07	-	A
eczema/dermatitis	0.0022	1.88E-07	1.09	A
Whole body fat-free mass	-0.0650	3.58E-07	-	A
Platelet crit	0.0005	3.70E-07	-	A
H30-H36 Disorders of choroid and retina	0.0017	6.88E-07	1.10	A
Time spent watching television (TV)	0.0190	9.06E-07	-	A
hypertension	0.0050	9.86E-07	1.03	A
hyperthyroidism/thyrotoxicosis	-0.0011	1.05E-06	0.86	A
Leg fat percentage (right)	0.0549	4.54E-06	-	A
Impedance of leg (left)	0.3075	7.19E-06	-	A
Impedance of whole body	0.6220	8.52E-06	-	A
Impedance of leg (right)	0.3041	8.99E-06	-	A
J20-J22 Other acute lower respiratory infections	0.0016	1.06E-05	1.09	A
Leg fat-free mass (right)	-0.0111	1.25E-05	-	A
Leg predicted mass (right)	-0.0103	1.56E-05	-	A
H00 Hordeolum and chalazion	0.0008	1.69E-05	1.19	A
J32 Chronic sinusitis	0.0009	2.02E-05	1.16	A
Leg fat percentage (left)	0.0485	2.06E-05	-	A
H00-H06 Disorders of eyelid, lacrimal system and orbit	0.0017	2.20E-05	1.08	A
Mean reticulocyte volume	-0.0710	2.55E-05	-	A
Basal metabolic rate	-7.2895	2.64E-05	-	A
J22 Unspecified acute lower respiratory infection	0.0015	2.99E-05	1.09	A
Body fat percentage	0.0569	3.37E-05	-	A
I10 Essential (primary) hypertension	0.0038	3.98E-05	1.03	A
Number of operations, self-reported	0.0155	4.27E-05	-	A
I10-I15 Hypertensive diseases	0.0038	4.71E-05	1.03	A

Arm fat-free mass (right)	-0.0037	5.36E-05	-	A

Supplementary Table S2. Protein Atlas data for tissue/cell enriched or enhanced gene and protein expression.

Gene	Tissue	Cell
AAGAB	None	None
BACH2	Lymphoid	horizontal cells
CD247	blood, lymphoid	T-cells
CLEC16A	None	early spermatids, late spermatids
D2HGDH	None	None
DEXI	skeletal muscle	None
FLG	oesophagus, skin, tongue	suprabasal keratinocytes
GATA3	seminal vesicle, skin	syncythiotrophoblasts, extravillous trophoblasts, cytotrophoblasts
GNGT2	blood, lymphoid, retina	cone photoreceptor cells, kupffer cells, macrophages
GSDMB	Intestine	enterocytes, pnaeth cells, mucus-secreting cells
HLA-DQA1	blood, lymphoid	monocytes, macrophages, B-cells, Hofbauer cells
HLA-DQA2	lung, lymphoid	macrophages,Hofbauer cells, B-cells, kupffer cells
HLA-DQB1	Lymphoid	monocytes, macrophages, B-cells, Hofbauer cells
HLA-DQB2	lymphoid, skin	macrophages, melanocytes, basal keratinocytes, suprabasal
IL18R1	Lung	Granulocytes
IL1RL1	kidney, lung, placenta	Granulocytes
IL33	None	lto cells, endothelial cells, club cells, cilliated cells
IRF1	Blood	none
KIF1A	Brain	Horizontal cells, Cone photoreceptor cells, Bipolar cells, Rod photoreceptor cells
KIAA1109	Low	Rod photoreceptor cells
LRRC32	None	extravillous trophoblasts, endothelial cells, smooth muscle cells, lto cells, fibroblasts, sertoli cells
MSL1	None	none
MUC5AC	gallbladder, lung, stomach	club cells, ciliated cells, pancreatic endocrine cells
ORMDL3	None	late spermatids, early spermatids
PDCD1	blood, lymphoid tissue	T-cells

PGAP3	None	none
RORA	Skin	cone photoreceptor cells, suprabasal keratinocytes
RPS26	None	none
SLC22A4	None	ciliated cells, erythroid cells
SLC22A5	kidney, skeletal muscle	enterocytes
SMAD3	Low	urothelial cells, glandular cells
STAT6	None	None
SUOX	None	hepatocytes
TSLP	None	basal keratinocytes, club cells, early spermatids, basal glandular cells, suprabasal keratinocytes, late spermatids
WDR36	None	basal keratinocytes
ZBTB10	None	None
ZNF652	None	None

Supplementary Table S3. DAVID Functional Annotation Tool analysis of the Genetic Association Database of complex diseases and disorders (GAD). Candidate causal gene clusters and disease associations are listed from the GAD database. FE = fold enrichment.

Term	Genes	Count	%	FE	p-value	FDR
Asthma	FLG, IL33, GSDMB, SLC22A5, SMAD3, TSLP, GATA3, SUOX, MUC5AC, IL1RL1, IRF1, ORMDL3, STAT6, HLA-DQA2, IL18R1, HLA-DQA1, HLA-DQB1	17	45.95	14.53	1.78E-15	1.07E-12
Diabetes Mellitus, Type 1	KIAA1109, GSDMB, CLEC16A, PDCD1, SUOX, HLA-DQA2, BACH2, HLA-DQA1, HLA-DQB1	9	24.32	32.76	1.58E-10	4.76E-08
Celiac disease	IL1RL1, KIAA1109, IRF1, CLEC16A, CD247, BACH2, IL18R1, HLA-DQA1, HLA-DQB1	9	24.32	22.97	2.72E-09	5.44E-07
nasal polyposis	IL33, IL1RL1, WDR36, HLA-DQA1, HLA-DQB1	5	13.51	140.38	3.00E-08	4.30E-06
obesity|asthma	IL33, IL1RL1, GSDMB, TSLP, ORMDL3, IL18R1	6	16.22	60.47	3.57E-08	4.30E-06
ulcerative colitis	SLC22A4, GSDMB, SLC22A5, ORMDL3, STAT6, HLA-DQA1, HLA-DQB1	7	18.92	31.27	6.45E-08	6.46E-06
Arthritis, Rheumatoid|	SLC22A4, IL1RL1, KIAA1109, CD247, HLA-DQA2, HLA-DQA1, HLA-DQB1	7	18.92	19.79	9.79E-07	8.41E-05
Crohn Disease|Crohn's disease	GSDMB, SLC22A5, SMAD3, ORMDL3, BACH2, HLA-DQA1	6	16.22	23.12	4.63E-06	3.48E-04
diabetes, type 1	SLC22A4, SLC22A5, IRF1, CLEC16A, PDCD1, HLA-DQA1, HLA-DQB1	7	18.92	14.48	6.02E-06	3.62E-04
rheumatoid arthritis	SLC22A4, SLC22A5, CLEC16A, PDCD1, HLA-DQB2, HLA-DQA1, HLA-DQB1	7	18.92	14.48	6.02E-06	3.62E-04
Bronchiolitis, Viral|Respiratory Syncytial Virus Infections	SLC22A4, SLC22A5, IRF1, STAT6, GATA3, IL18R1, MUC5AC	7	18.92	13.04	1.10E-05	5.50E-04
Cholangitis, Sclerosing	SLC22A4, SLC22A5, HLA-DQA1, HLA-DQB1	4	10.81	68.36	2.34E-05	0.00108
Crohn's disease	SLC22A4, SLC22A5, IRF1, ORMDL3, HLA-DQB1	5	13.51	25.52	3.64E-05	0.001562
hepatitis C	IL1RL1, IRF1, IL18R1, HLA-DQA1, HLA-DQB1	5	13.51	23.68	4.89E-05	0.001959
Asthma|Myocardial Infarction	IL33, IL1RL1, WDR36	3	8.11	235.84	5.87E-05	0.002205
Asthma|Bronchial Hyperreactivity|Hypersensitivity, Immediate	FLG, IL1RL1, IRF1, STAT6, GATA3	5	13.51	21.60	7.01E-05	0.00248
Crohn Disease	SLC22A4, SLC22A5, SMAD3, BACH2, HLA-DQB1	5	13.51	20.69	8.30E-05	0.002676
Graves' disease	IRF1, PDCD1, HLA-DQA1, HLA-DQB1	4	10.81	44.92	8.46E-05	0.002676
diabetes, type 1	SMAD3, CLEC16A, SUOX, BACH2, HLA-DQA1, HLA-DQB1	6	16.22	12.48	9.11E-05	0.002738
Tuberculosis	SLC22A4, SLC22A5, IRF1, HLA-DQA1, HLA-DQB1	5	13.51	19.65	1.01E-04	0.0029
plasma eosinophil count	IL1RL1, WDR36, TSLP	3	8.11	168.45	1.23E-04	0.003337
Lupus Erythematosus, Systemic	CLEC16A, CD247, PDCD1, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	7	18.92	8.41	1.28E-04	0.003337

Asthma|	IL1RL1, ORMDL3, IL18R1, HLA-DQA1, HLA-DQB1	5	13.51	16.66	1.92E-04	0.004813

Supplementary Table S4. Frequency of genes in GAD Terms. A total of 24 GAD terms were under 5% FDR. The frequency of genes were counted and a percentage calculated.

GENE	Number of Terms (total 23)	%
HLA-DQB1	17	73.91
HLA-DQA1	16	69.57
SLC22A5	10	43.48
IL1RL1	10	43.48
IRF1	9	39.13
SLC22A4	9	39.13
ORMDL3	6	26.09
IL18R1	6	26.09
CLEC16A	6	26.09
GSDMB	5	21.74
PDCD1	5	21.74
BACH2	5	21.74
IL33	4	17.39
SMAD3	4	17.39
STAT6	4	17.39
TSLP	3	13.04
GATA3	3	13.04
SUOX	3	13.04
KIAA1109	3	13.04
CD247	3	13.04
WDR36	3	13.04
FLG	2	8.70
MUC5AC	2	8.70
HLA-DQA2	2	8.70
HLA-DQB2	2	8.70

Supplementary Table S5. DAVID Functional Annotation Tool analysis of GO Terms. Candidate causal gene clusters and GO Term associations are listed. FE = fold enrichment.

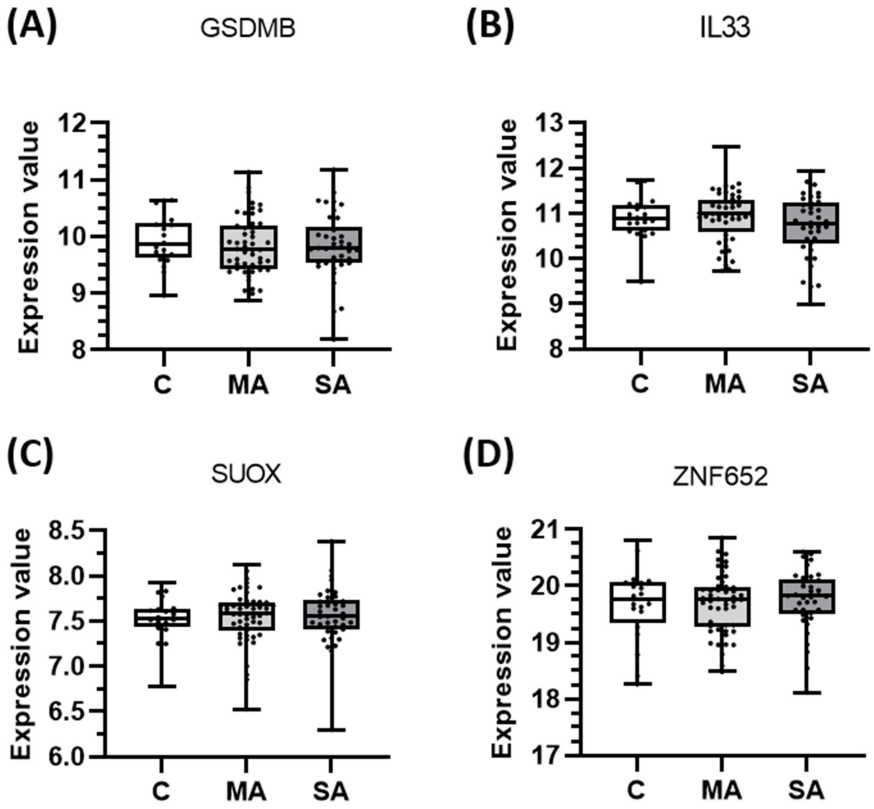
Term	Genes	Count	%	FE	p-value	FDR
T cell costimulation	CD247, PDCD1, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	6	16.22	36.901	4.76E-07	1.50E-04
MHC class II receptor activity	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	136.41	2.77E-06	2.86E-04
MHC class II protein complex	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	92.04	9.75E-06	5.75E-04
integral component of lumenal side of endoplasmic reticulum membrane	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	69.824	2.29E-05	6.76E-04
antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	112.89	5.06E-06	7.99E-04
transport vesicle membrane	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	53.29	5.23E-05	9.71E-04
clathrin-coated endocytic vesicle membrane	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	49.39	6.58E-05	9.71E-04
T cell receptor signaling pathway	GATA3, CD247, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	6	16.22	19.45	1.13E-05	9.77E-04
interferon-gamma-mediated signaling pathway	IRF1, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	5	13.51	33.79	1.24E-05	9.77E-04
ER to Golgi transport vesicle membrane	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	38.94	1.34E-04	0.001587
endocytic vesicle membrane	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	30.68	2.73E-04	0.002688
trans-Golgi network membrane	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	24.40	5.37E-04	0.004523

Supplementary Table S6. DAVID Functional Annotation Tool analysis of the KEGG/REACTOME Pathway. Candidate causal gene clusters and pathway associations are listed from the GAD database. FE = fold enrichment.

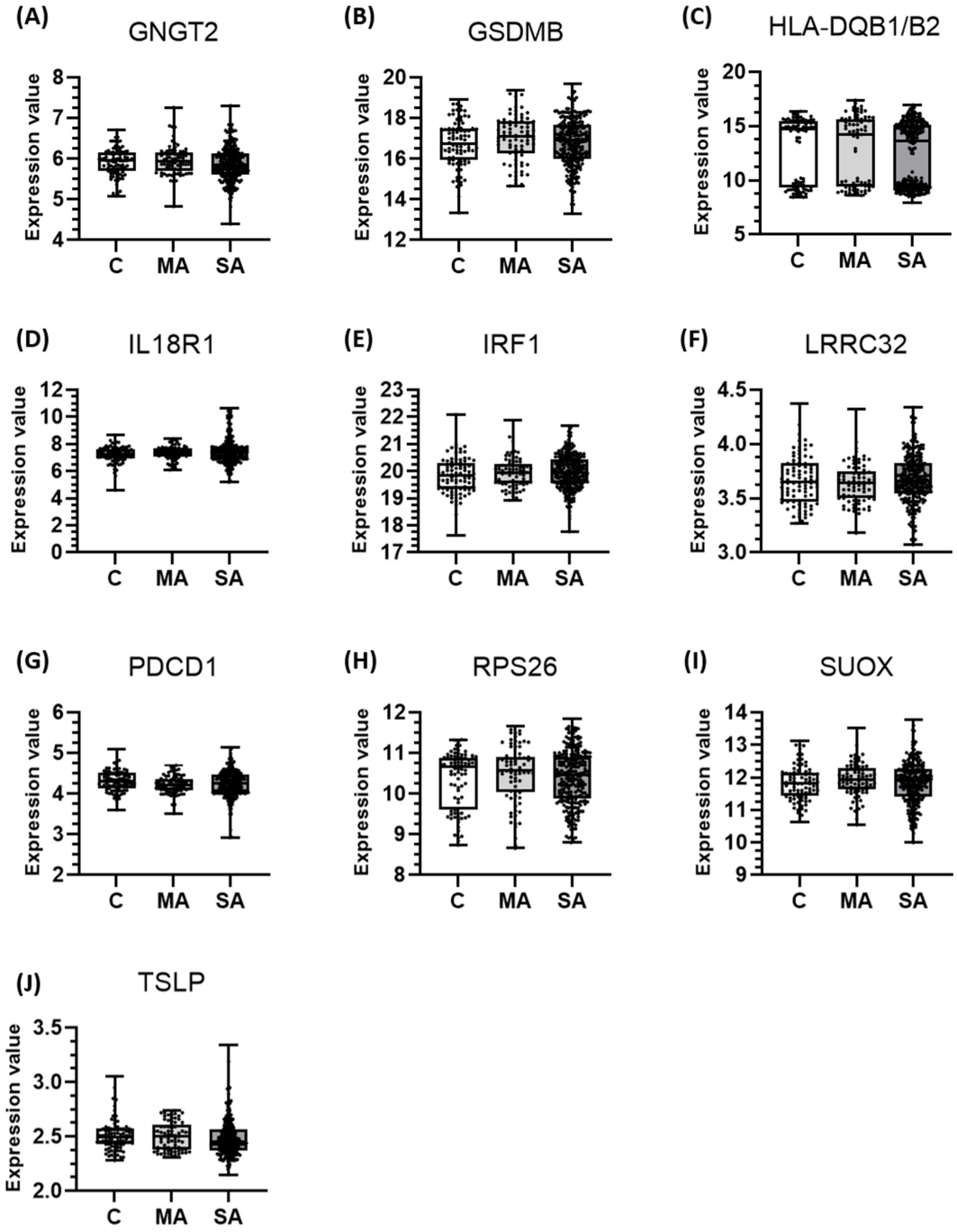
Category	Term	Genes	Count	%	P-value	FE	FDR
KEGG	Inflammatory bowel disease	SMAD3, STAT6, RORA, HLA-DQA2, IL18R1, HLA-DQA1, HLA-DQB1	7	18.92	5.83E-09	41.80	2.86E-07
REACTOME	PD-1 signalling	CD247, PDCD1, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	6	16.22	4.17E-09	87.26	2.92E-07
REACTOME	Translocation of ZAP-70 to Immunological synapse	CD247, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	5	13.51	2.23E-07	85.94	7.79E-06
REACTOME	Phosphorylation of CD3 and TCR zeta chains	CD247, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	5	13.51	3.83E-07	75.63	8.93E-06
REACTOME	Generation of second messenger molecules	CD247, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	5	13.51	1.75E-06	52.52	3.06E-05
REACTOME	Interferon gamma signaling	IRF1, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	5	13.51	6.35E-05	21.48	8.89E-04
REACTOME	Downstream TCR signaling	CD247, HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	5	13.51	1.09E-04	18.72	0.001269
REACTOME	MHC class II antigen presentation	HLA-DQA2, HLA-DQB2, HLA-DQA1, HLA-DQB1	4	10.81	0.003449	12.40	0.034495
KEGG	Asthma	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.002401	38.22	0.037734
KEGG	Graft-versus-host disease	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.002902	34.74	0.037734
KEGG	Allograft rejection	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.003639	30.99	0.037734
KEGG	Type I diabetes mellitus	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.00467	27.30	0.037734
KEGG	Cell adhesion molecules (CAMs)	PDCD1, HLA-DQA2, HLA-DQA1, HLA-DQB1	4	10.81	0.004736	10.77	0.037734
KEGG	Intestinal immune network for IgA production	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.005821	24.39	0.037734
KEGG	Autoimmune thyroid disease	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.007089	22.05	0.037734
KEGG	Staphylococcus aureus infection	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.007628	21.23	0.037734
KEGG	Influenza A	IL33, HLA-DQA2, HLA-DQA1, HLA-DQB1	4	10.81	0.00833	8.79	0.037734
KEGG	Viral myocarditis	HLA-DQA2, HLA-DQA1, HLA-DQB1	3	8.11	0.008471	20.11	0.037734

Supplementary Table S7. GeneMANIA scores for predicted genes. The GeneMANIA tool was used to identify interactions between the 37 candidate genes and predict additional genes. Predicted gene scores were multiplied by the number of genes the predicted gene interacted with to obtain a total score and identify the predicted genes with the strongest interactions.

Predicted gene	score	No. of gene interactions	Total score
CD274	6.06	13	78.78
ZAP70	4.03	9	36.27
PDCD1LG2	3.56	6	21.36
IRAK4	2.99	5	14.95
IL37	2.03	6	12.18
SOX4	2.82	3	8.46
IL18	2.04	4	8.16
CRLF2	3.84	2	7.68
CADPS	2.41	3	7.23
IL1RAP	2.27	3	6.81
TMED1	2.19	3	6.57
PRL	1.64	4	6.56
TMBIM6	2.18	3	6.54
TSR2	2.89	2	5.78
KCNJ4	1.87	3	5.61
SMAD4	2.37	2	4.74
PCSK6	2.03	2	4.06
ATXN1	2.63	1	2.63
MAGED1	2.32	1	2.32
CASP14	2.26	1	2.26



Supplementary Figure S1. mRNA expression of candidate causal genes in bronchial epithelial cells taken from patients with asthma and controls. Boxes shows the median and IQR and the whiskers show the minimum and maximum data. Bronchial epithelial brush samples were from controls (C, n=20) and patients with mild-moderate (MA, n=50) and severe (SA, n=38) asthma. Data is shown for **(A)** GSDMB, **(B)** IL33, **(C)** SUOX and **(D)** ZNF652. Expression values were taken from the dataset and a Kruskal-Wallis test with a two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli used to control the FDR at 5% was performed.


(A)
GNGT2
Expression value
8
7
6
5
4
C
MA
SA
(B)
GSDMB
Expression value
20
18
16
14
12
C
MA
SA
(C)
HLA-DQB1/B2
Expression value
20
15
10
5
C
MA
SA
(D)
IL18R1
Expression value
12
10
8
6
4
2
0
C
MA
SA
(E)
IRF1
Expression value
23
22
21
20
19
18
17
C
MA
SA
(F)
LRRC32
Expression value
4.5
4.0
3.5
3.0
C
MA
SA
(G)
PDCD1
Expression value
6
5
4
3
2
C
MA
SA
(H)
RPS26
Expression value
12
11
10
9
8
C
MA
SA
(I)
SUOX
Expression value
14
13
12
11
10
9
C
MA
SA
(J)
TSLP
Expression value
3.5
3.0
2.5
2.0
C
MA
SA

Supplementary Figure S2. mRNA expression of candidate causal genes in blood taken from patients with asthma and control subjects. Boxes shows the median and IQR and the whiskers show the minimum and maximum data. Blood samples were from controls (C, n=87) and patients with mild-moderate (MA, n=77) and severe (SA, n=246) asthma. Data is shown for **(A)** GNGT2, **(B)** GSDMB, **(C)** HLA-DQB1/B2, **(D)** IL18R1, **(E)** IRF1, **(F)** LRRC32, **(G)** PDCD1, **(H)** RPS26, **(I)** SUOX and **(J)** TSLP. Expression values were taken from the dataset and either a Kruskal-Wallis test or Welch's ANOVA (GSDMB), both with a two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli used to control the FDR at 5%, was performed.