



Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents

Yale Jiang^{1,2,9}, Olena Gruzieva^{3,9}, Ting Wang¹, Erick Forno¹, Nadia Boutaoui¹, Tao Sun⁴, Simon K. Merid³, Edna Acosta-Pérez⁵, Inger Kull³, Glorisa Canino⁵, Josep M. Antó⁶, Jean Bousquet^{7,8}, Erik Melén^{3,10}, Wei Chen^{1,10} and Juan C. Celedón^{1,10}

Affiliations: ¹Division of Pulmonary Medicine, Dept of Pediatrics, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA, USA. ²School of Medicine, Tsinghua University, Beijing, China. ³Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. ⁴Dept of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. ⁵Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, Puerto Rico. ⁶ISGlobal, Barcelona Institute for Global Health, Barcelona, Spain. ⁷CESP, Inserm U1018, Villejuif, France. ⁸University Hospital, Montpellier, France. ⁹These two authors contributed equally to this work. ¹⁰These three authors are joint senior authors.

Correspondence: Juan C. Celedón, Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, 4401 Penn Avenue, Pittsburgh, PA 15224, USA. E-mail: juan.celedon@chp.edu



@ERSpublications

In Puerto Rican and Swedish children, a TWAS identified 59 DEGs by atopy and 40 DEGs by atopic asthma, including *SIGLEC8* and *IL17RB*. Transcriptomic data from WBCs could be used to develop predictive models of atopy/atopic asthma in prospective studies. <http://ow.ly/kF6v30nYnYC>

Cite this article as: Jiang Y, Gruzieva O, Wang T, *et al*. Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents. *Eur Respir J* 2019; 53: 1900102 [<https://doi.org/10.1183/13993003.00102-2019>].

ABSTRACT Early allergic sensitisation (atopy) is the first step in the development of allergic diseases such as atopic asthma later in life. Genes and pathways associated with atopy and atopic asthma in children and adolescents have not been well characterised.

A transcriptome-wide association study (TWAS) of atopy and atopic asthma in white blood cells (WBCs) or whole blood was conducted in a cohort of 460 Puerto Ricans aged 9–20 years (EVA-PR study) and in a cohort of 250 Swedish adolescents (BAMSE study). Pathway enrichment and network analyses were conducted to further assess top findings, and classification models of atopy and atopic asthma were built using expression levels for the top differentially expressed genes (DEGs).

In a meta-analysis of the study cohorts, both previously implicated genes (*e.g.* *IL5RA* and *IL1RL1*) and genes not previously reported in TWASs (novel) were significantly associated with atopy and/or atopic asthma. Top novel genes for atopy included *SIGLEC8* ($p=8.07\times10^{-13}$), *SLC29A1* ($p=7.07\times10^{-12}$) and *SMPD3* ($p=1.48\times10^{-11}$). Expression quantitative trait locus analyses identified multiple asthma-relevant genotype–expression pairs, such as rs2255888/*ALOX15*. Pathway enrichment analysis uncovered 16 significantly enriched pathways at adjusted $p<0.01$, including those relevant to T-helper cell type 1 (Th1) and Th2 immune responses. Classification models built using the top DEGs and a few demographic/parental history variables accurately differentiated subjects with atopic asthma from nonatopic control subjects (area under the curve 0.84).

We have identified genes and pathways for atopy and atopic asthma in children and adolescents, using transcriptome-wide data from WBCs and whole blood samples.

This article has supplementary material available from erj.ersjournals.com

Received: Jan 15 2019 | Accepted after revision: March 02 2019

Copyright ©ERS 2019

Introduction

Over the last few decades, the prevalence of allergic diseases has increased in the USA and worldwide. Sensitisation to allergens (atopy) in early life precedes the development of allergic diseases later in childhood, including food allergies, allergic rhinitis and atopic asthma [1]. Puerto Rican and Swedish children bear a high burden of atopy and asthma [1–3]. In studies of school-aged children, estimates of the prevalence of asthma and allergic rhinitis in Puerto Ricans were 16.1% and 48.4%, respectively [2, 4]. Among Swedish children aged 7–8 years, estimates of the prevalence of current wheeze and allergic rhinitis were 13% and 13.8%, respectively [5].

Recent meta-analyses of genome-wide association studies (GWASs) identified 18 susceptibility loci for asthma [6] and 41 susceptibility loci for allergic rhinitis [7]. However, such loci accounted for a small proportion of the heritability of asthma (~3.5%) and allergic rhinitis (~7.8%), and alternative approaches such as transcriptomics could help identify the “missing heritability” of atopic diseases. Yet another complementary approach is to study intermediate phenotypes such as atopy, as susceptibility genes for atopy often overlap with those for common allergic diseases [7].

In contrast to the many GWASs conducted to date, there have been few transcriptome-wide association studies (TWASs) of atopy and asthma, all limited by modest sample size and no or insufficient replication [8–12]. For example, a study of targeted RNA sequencing of 105 genes in nasal brushings from subjects with (n=50) and without (n=50) asthma showed that nasal transcriptomic profiles could identify subjects with *IL13*-driven asthma and T-helper cell type 2 (Th2)-skewed immune responses [9]. In another study (including mostly adults), RNA sequencing profiles from nasal brushings were used to first develop a panel of 90 genes to differentiate subjects with mild-to-moderate asthma (n=53) from control subjects (n=97), and then test the “classifier panel” in 40 subjects with (n=13) and without (n=27) asthma [10]. Although the panel performed well in training cohorts, external replication was limited and thus no individual genes could confidently be identified as conferring susceptibility to asthma.

While airway epithelial studies have shown promising results, blood samples are easy to collect and often available in epidemiological studies. A genome-wide analysis of transcriptomics in circulating CD19⁺ B-lymphocytes from 41 adults provided suggestive evidence of increased *IL4R* expression in subjects with dust mite allergy and asthma compared with control subjects [11]. In another study in young adolescents with asthma, microarray expression profiles from CD4⁺ T-lymphocytes were differentially expressed by atopic status [12] and five “atopic signature” genes were then nominally replicated in a cohort of 30 subjects. A later integrative study in asthma controls focused on the transcriptomic components in blood that vary with degree of asthma control [13].

Here, we report the findings from the first TWAS of atopy and atopic asthma in white blood cells (WBCs) from children and adolescents in Puerto Rico, with validation studies in a cohort of Swedish children.

Methods

Further details are given in the supplementary material.

Study population

The EVA-PR (Epigenetic Variation and Childhood Asthma in Puerto Ricans) study is a case–control study of asthma in Puerto Ricans aged 9–20 years, recruited using an approach similar to that used in previous studies in the metropolitan area of San Juan (Puerto Rico) [14, 15]. Of the 543 participating children, 460 had measurements of IgE to five common allergens in Puerto Rico (dust mite (Der p 1), cockroach (Bla g 2), cat dander (Fel d 1), dog dander (Can f 1) and mouse urinary protein (Mus m 1)) [16], WBC count and differential, and RNA from WBCs, and were thus included in the current analysis. The study was approved by the institutional review boards of the University of Puerto Rico (San Juan) and the University of Pittsburgh (Pittsburgh, PA, USA). Written parental consent and assent were obtained from all participants.

RNA sequencing and data pre-processing

Sequencing was performed using 350 ng of high-quality RNA extracted from WBCs after removing haemoglobin (RNA integrity number >7). Library preparation was done using the TruSeq Stranded Total RNA Library Prep Kit with the Ribo-Zero Gold High Throughput Kit (Illumina, San Diego, CA, USA), according to the manufacturer’s protocol. Libraries were run on the NextSeq 500 using the NextSeq 500/550 High Output Kit version 2 (Illumina).

Quality control for raw RNA sequencing FASTQ files was performed using FastQC [17]. Low-quality reads and 3′ adapters were trimmed with Trim Galore! [18] and Cutadapt [19]. Trimmed reads were then aligned to reference human genome hg19 with STAR [20] and subsequently annotated in the Illumina

iGenomes database using RSEM [21]. Samples with low alignment percentage and genes with low expression were removed from downstream analyses. After pre-processing, a total of 16 880 genes and 460 samples were retained in the final analysis.

Principal components were calculated from genome-wide genotypic data, which were obtained using the HumanOmni2.5 BeadChip platform (Illumina), as previously described [22]. Imputation of nongenotyped single nucleotide polymorphisms (SNPs) was performed with the Imputation Server [23], using the Haplotype Reference Consortium r1.1 2016 [24] as the reference panel.

The BAMSE study

We also analysed whole blood samples using Affymetrix Human Transcriptome Array 2.0 (Affymetrix, Santa Clara, CA, USA) in 269 16-year-old children selected from BAMSE (Children [Barn], Allergy, Milieu, Stockholm, Epidemiology), a Swedish population-based cohort (with oversampling of asthma and rhinitis cases) [25]. After quality control (RNA integrity number >6) and outlier detection, gene expression data from 250 blood samples were included in the current analysis of atopy and atopic asthma with available WBC count and differential. Allergen-specific IgE levels were measured using the ImmunoCAP System (Thermo Fisher/Phadia, Uppsala, Sweden).

Statistical analyses

In EVA-PR and BAMSE, atopy was defined as at least one positive IgE (≥ 0.35 IU·mL⁻¹) to the allergens tested. Atopic asthma was defined as current asthma (physician-diagnosed asthma and current wheeze) and atopy in EVA-PR, and as current asthma (assessed using a modified version of the GA²LEN questionnaire2) and atopy in BAMSE [26].

Differentially expressed genes (DEGs) were analysed based on the raw count table using DESeq2 [27] for RNA sequencing data in EVA-PR and LIMMA for microarray data in BAMSE. In both EVA-PR and BAMSE, multivariable models of atopic asthma were adjusted for age, sex, batch label and the proportions (percentages) of WBC subtypes (eosinophils, lymphocytes, monocytes and neutrophils). All multivariable models of atopy were additionally adjusted for asthma status. In EVA-PR, all models were additionally adjusted for the first five principal components derived from genotypes, to account for population stratification. A zero-mean normal prior was put on the nonintercept coefficients to ensure that fold changes were independent of the choice of reference level. Next, we conducted a meta-analysis of the cohort-specific results using a fixed effect model with in-house code using the R package meta [28, 29]. The false discovery rate (FDR) approach was applied to adjust for multiple testing.

In EVA-PR, we performed an expression quantitative trait loci (eQTL) analysis (at a distance <1 Mb from gene start and end sites) for the top findings in the meta-analysis of atopy using the R package Matrix eQTL [30].

Pathway and network analysis

Ingenuity Pathway Analysis was performed for the top DEGs (FDR-adjusted $p < 0.1$) in the meta-analysis of atopy and atopic asthma to identify enriched canonical pathways and gene networks [31]. We then calculated the partial correlation between gene expression levels in our discovery cohort, first in all subjects and then separately by asthma status, using the R package FastGGM [32], a powerful tool to detect conditional gene networks under the framework of the Gaussian graphical model.

Classification models of atopy and atopic asthma

Logistic regression with penalised maximum likelihood was applied to build classification models of atopy and atopic asthma in EVA-PR. We first built two models for comparison: one including only demographic and parental history variables (age, sex and parental history of asthma) as predictors, and another including demographic variables, parental history and gene expression levels of randomly chosen genes. Next, we built a model using the top DEGs in our meta-analysis, and another using the top DEGs plus demographic variables and parental history. Models were built separately for atopy and atopic asthma.

Results

The main characteristics of study participants are shown in table 1. In EVA-PR, subjects with atopy were significantly more likely to be male and to have asthma than control subjects. Compared with participants in EVA-PR, those in BAMSE had, as expected for a population-based cohort study, a lower proportion of asthmatic subjects and a narrower age range.

TABLE 1 Characteristics of participants in the EVA-PR and BAMSE studies

	EVA-PR [#]		BAMSE [¶]	
	Atopy	No atopy	Atopy	No atopy
Subjects	300	160	134	116
Age years	15.2 (10–20)	15.7 (10–20)	16.7 (16.0–17.9)	16.7 (15.9–18.1)
Female	131 (43.7)	87 (54.4)	47 (35.1)	59 (50.9)
Asthma	165 (55)	63 (39.4)	42 (31.3)	10 (8.6)
Sensitisation to allergens				
Dust mite (Der p 1)	282 (94)		48 (35.8)	
Cat dander (Fel d 1)	49 (16.3)		72 (53.7)	
Dog dander (Can f 1)	168 (56)		85 (63.4)	
Mouse urinary protein (Mus m 1)	12 (4)		Not measured	
Cockroach (Bla g 2)	174 (58)		Not measured	

Data are presented as n, mean (range) or n (%). [#]: n=460; [¶]: n=250.

TWAS of atopy

In the multivariable analysis of atopy, which was adjusted for WBC subtype and other covariates, we identified 69 DEGs in EVA-PR at FDR-adjusted $p < 0.05$. In BAMSE, there were no DEGs at FDR-adjusted $p < 0.05$, but 17 of the 69 DEGs in EVA-PR were differentially expressed at $p < 0.05$, with similar estimates of effect size across the two studies (supplementary table S1). Next, to increase the statistical power of DEG analysis, we combined the results from the two cohorts in a transcriptome-wide meta-analysis, which identified 59 DEGs by atopy, at an FDR-adjusted $p < 0.05$ (figure 1a and supplementary figure S1). Table 2 lists the top 20 genes (by p-value) in this meta-analysis, for which the direction of the association between gene expression and atopy was consistent across the two studies.

For the top DEGs by atopy, we found 43 enriched pathways at FDR-adjusted $p < 0.05$. Table 3 lists the top 20 enriched pathways, of which the two most significant ones were the Th2 pathway and the Th1/Th2 activation pathway. Other top pathways included those for nerve growth factor (NGF) and adenosine/adenosine salvage I.

TWAS of atopic asthma

We then conducted a transcriptomic analysis of atopic asthma, finding substantial overlap between the results from this analysis and those for atopy (figure 1b and supplementary table S2). We identified 49

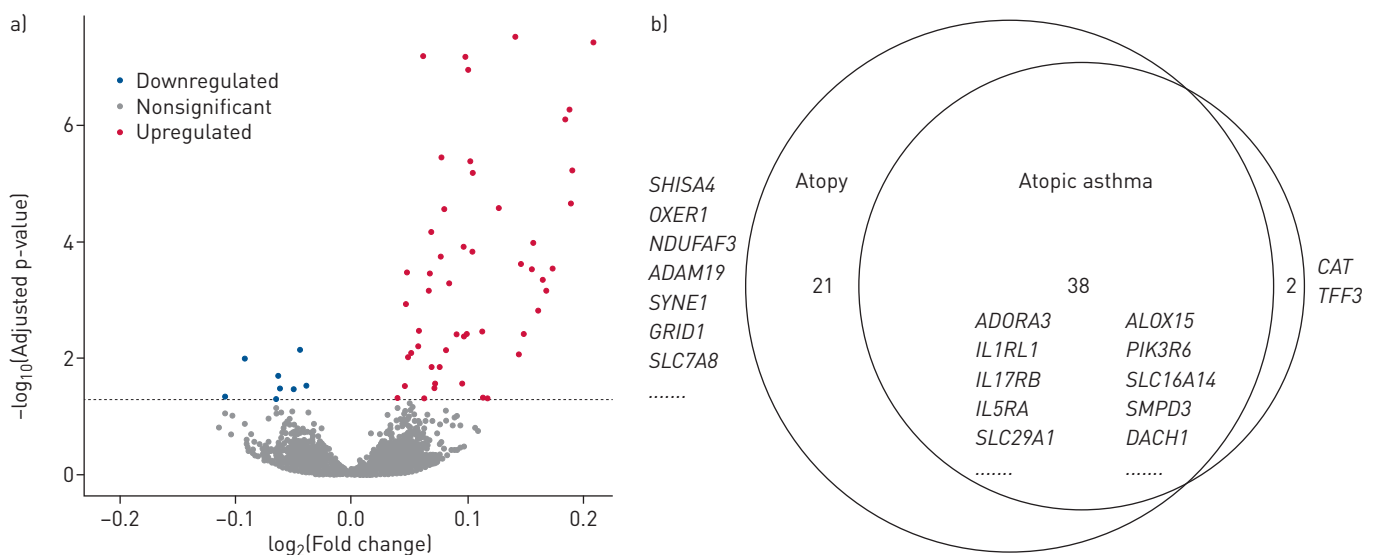


FIGURE 1 Meta-analysis of gene expression demonstrates substantial differences in gene expression between atopic subjects and control subjects. DEG: differentially expressed gene. a) Volcano plot showing the $\log_2$ (fold change) and $-\log_{10}$ (adjusted p-value) of all the genes in the meta-analysis. The statistically significant DEGs between atopic subjects and control subjects are represented by red (upregulated) and blue (downregulated) symbols. The statistical criterion for a gene to be considered differentially expressed was false discovery rate-adjusted $p < 0.05$. b). Venn diagram depicting the distribution of the DEGs in the analysis between atopic subjects and control subjects, and between atopic asthmatic subjects and nonatopic nonasthmatic control subjects.

TABLE 2 Top differentially expressed genes from the meta-analysis of atopy

Gene symbol	Description	EVA-PR			BAMSE			Meta-analysis		
		Fold change	p-value	FDR-adjusted p-value	Fold change	p-value	FDR-adjusted p-value	Fold change	p-value	FDR-adjusted p-value
SIGLEC8	Sialic acid binding Ig-like lectin 8	1.18	2.97×10^{-12}	2.51×10^{-8}	1.05	9.34×10^{-3}	NS	1.10	8.07×10^{-13}	2.86×10^{-8}
IL5RA	Interleukin-5 receptor subunit α	1.15	2.06×10^{-9}	3.48×10^{-6}	1.18	1.97×10^{-4}	NS	1.16	2.03×10^{-12}	3.59×10^{-8}
IL17RB	Interleukin-17 receptor B	1.19	1.29×10^{-14}	2.18×10^{-10}	1.01	2.39×10^{-1}	NS	1.04	5.14×10^{-12}	6.07×10^{-8}
SLC29A1	Solute carrier family 29 member 1 (Augustine blood group)	1.15	1.23×10^{-9}	2.31×10^{-6}	1.05	9.24×10^{-4}	NS	1.07	7.07×10^{-12}	6.26×10^{-8}
SMPD3	Sphingomyelin phosphodiesterase 3	1.15	2.44×10^{-10}	1.03×10^{-6}	1.04	5.31×10^{-3}	NS	1.07	1.48×10^{-11}	1.05×10^{-7}
IL1RL1	Interleukin-1 receptor-like 1	1.14	6.15×10^{-9}	8.94×10^{-6}	1.13	2.24×10^{-3}	NS	1.14	8.46×10^{-11}	4.99×10^{-7}
ALOX15	Arachidonate 15-lipoxygenase	1.15	9.79×10^{-10}	2.11×10^{-6}	1.11	1.22×10^{-2}	NS	1.14	1.48×10^{-10}	7.50×10^{-7}
OLIG2	Oligodendrocyte transcription factor 2	1.17	5.22×10^{-12}	2.94×10^{-8}	1.01	3.18×10^{-1}	NS	1.06	7.96×10^{-10}	3.52×10^{-6}
PIK3R6	Phosphoinositide-3-kinase regulatory subunit 6	1.12	9.07×10^{-8}	7.65×10^{-5}	1.05	2.41×10^{-3}	NS	1.07	1.04×10^{-9}	4.10×10^{-6}
CLC	Charcot-Leyden crystal galectin	1.14	2.54×10^{-8}	2.38×10^{-5}	1.17	9.22×10^{-3}	NS	1.14	1.65×10^{-9}	5.85×10^{-6}
ADORA3	Adenosine A3 receptor	1.14	6.58×10^{-9}	8.94×10^{-6}	1.04	2.47×10^{-2}	NS	1.07	1.95×10^{-9}	6.27×10^{-6}
CEBPE	CCAAT/enhancer binding protein ϵ	1.12	1.04×10^{-6}	6.04×10^{-4}	1.07	2.32×10^{-3}	NS	1.09	9.52×10^{-9}	2.59×10^{-5}
PMP22	Peripheral myelin protein 22	1.15	1.00×10^{-9}	2.11×10^{-6}	1.02	1.75×10^{-1}	NS	1.06	1.05×10^{-8}	2.66×10^{-5}
PRSS33	Protease, serine 33	1.15	7.21×10^{-10}	2.03×10^{-6}	1.01	3.15×10^{-1}	NS	1.05	2.77×10^{-8}	6.55×10^{-5}
PTGDR2	Prostaglandin D2 receptor 2	1.12	3.85×10^{-7}	2.60×10^{-4}	1.04	2.34×10^{-2}	NS	1.07	5.61×10^{-8}	1.17×10^{-4}
RAB44	Member RAS oncogene family	1.08	6.77×10^{-5}	2.11×10^{-2}	1.07	2.39×10^{-4}	NS	1.07	7.15×10^{-8}	1.41×10^{-4}
HRASLS5	HRAS-like suppressor family member 5	1.14	2.11×10^{-8}	2.19×10^{-5}	1.02	1.60×10^{-1}	NS	1.05	9.10×10^{-8}	1.69×10^{-4}
HRH4	Histamine receptor H4	1.11	4.78×10^{-6}	2.18×10^{-3}	1.10	7.45×10^{-3}	NS	1.11	1.37×10^{-7}	2.42×10^{-4}
CCR3	CC motif chemokine receptor 3	1.11	1.20×10^{-5}	4.72×10^{-3}	1.15	4.20×10^{-3}	NS	1.11	1.77×10^{-7}	2.85×10^{-4}
CDK15	Cyclin-dependent kinase 15	1.14	8.22×10^{-9}	9.25×10^{-6}	1.01	3.56×10^{-1}	NS	1.03	2.14×10^{-7}	3.29×10^{-4}

FDR: false discovery rate; NS: nonsignificant [i.e. FDR-adjusted $p > 0.05$]. All models were adjusted for age, sex, asthma status, batch and proportions of white blood cell subtypes (eosinophils, lymphocytes, neutrophils and monocytes). Models in EVA-PR were additionally adjusted for principal components.

DEGs by atopic asthma in EVA-PR at FDR-adjusted $p < 0.05$. In BAMSE, there were no DEGs at FDR-adjusted $p < 0.05$, but five of the 49 DEGs by atopic asthma in EVA-PR were differentially expressed at $p < 0.05$ in BAMSE, with similar estimates of effect size. 40 DEGs were then identified in the meta-analysis at FDR-adjusted $p < 0.05$. For atopic asthma, a pathway enrichment analysis revealed pathways related to lipid metabolism (e.g. docosahexaenoic acid signalling) and arachidonic acid signalling (e.g. eicosanoids signalling) (supplementary table S3).

eQTL analysis

Next, we conducted a *cis*-eQTL analysis in EVA-PR, which identified 3772 SNPs that were significantly associated with our top genes from the expression analysis for atopy at FDR-adjusted $p < 0.05$ (supplementary table S4). Such SNPs included multiple eQTLs in the GTEx database (e.g. rs2255888 in *ALOX15* and rs35092096 in *CCR3*) [33]. Of interest, multiple SNPs were associated with expression of *IL1RL1* (a known asthma-susceptibility gene) [6].

Network analysis

A network analysis was performed to illustrate interactions among our top DEGs by atopy (supplementary figure S2), first in all subjects and then separately in atopic subjects and in control subjects. Partial correlation among these genes remained strong, even after conditioning on global expression profile. This was particularly true for the gene most differentially expressed, *SIGLEC8*. We also found that some strong correlations in control subjects weakened significantly in atopic children, such as those for the gene pair *IL34* and *CYSLTR2*.

We next performed hierarchical clustering with log-transformed transcripts per kilobase million values in EVA-PR, based on the 59 DEGs (FDR-adjusted $p < 0.05$) in the meta-analysis of atopy, identifying four

TABLE 3 Top functional Ingenuity pathways enriched by the top differentially expressed genes in the meta-analysis of atopy

Ingenuity canonical pathway	$-\log_{10}(\text{Adjusted p-value})$	Gene symbol
Th2 pathway	3.83	<i>CCR3, PTGDR2, IL17RB, IL1RL1, PIK3R6</i>
Th1 and Th2 activation pathway	3.41	<i>CCR3, PTGDR2, IL17RB, IL1RL1, PIK3R6</i>
Type II diabetes mellitus signalling	2.75	<i>CACNA1D, PIK3R6, CACNG8, SMPD3</i>
MSP–RON signalling pathway	2.74	<i>CSF1, PIK3R6, RPS6KA2</i>
Fc γ RIIB signalling in B-lymphocytes	2.54	<i>CACNA1D, PIK3R6, CACNG8</i>
Role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis	2.35	<i>IL1RL1, CSF1, PIK3R6, FZD1, CEBPE</i>
Amyotrophic lateral sclerosis signalling	2.21	<i>CACNA1D, GRID1, PIK3R6</i>
CREB signalling in neurons	2.21	<i>CACNA1D, GRID1, PIK3R6, CACNG8</i>
UVA-induced MAPK signalling	2.21	<i>PIK3R6, RPS6KA2, SMPD3</i>
Adenine and adenosine salvage I	2.18	<i>APRT</i>
Role of osteoblasts, osteoclasts and chondrocytes in rheumatoid arthritis	2.11	<i>IL1RL1, CSF1, PIK3R6, FZD1</i>
nNOS signalling in skeletal muscle cells	2.08	<i>CACNA1D, CACNG8</i>
NGF signalling	2.08	<i>PIK3R6, RPS6KA2, SMPD3</i>
Atherosclerosis signalling	2.05	<i>CCR3, ALOX15, CSF1</i>
Role of tissue factor in cancer	2.02	<i>CSF1, PIK3R6, RPS6KA2</i>
Cellular effects of sildenafil (Viagra)	2.01	<i>CACNA1D, MYH9, CACNG8</i>
DHA signalling	1.88	<i>ALOX15, PIK3R6</i>
α -tocopherol degradation	1.88	<i>CYP4F12</i>
Tetrapyrrole biosynthesis II	1.78	<i>ALAS2</i>
PKC θ signalling in T-lymphocytes	1.75	<i>CACNA1D, PIK3R6, CACNG8</i>

Th: T-helper cell; MSP: macrophage stimulating protein; RON: RON protein tyrosine kinase/receptor d'origine nantais; CREB: cAMP response element binding protein; UVA: ultraviolet A; MAPK: mitogen-activated protein kinase; nNOS: neuronal nitric oxide synthase; NGF: nerve growth factor; DHA: docosahexaenoic acid; PKC: protein kinase C.

individual clusters (figure 2). In this analysis, the prevalence of atopy differs between the first two clusters, but not between the last two clusters (supplementary figure S3). However, when we compared eosinophil proportions in samples across these four clusters, we found that the samples with higher expression of the top DEGs for atopy had more eosinophils in peripheral blood (figure 3). We then tested for an association between expression levels of the top DEGs for atopy or atopic asthma and the measured proportions of each cell subset in peripheral blood (supplementary table S5). Of the top 59 DEGs for atopy, 51 genes were significantly associated (FDR-adjusted $p < 0.05$) with the proportion of eosinophils in peripheral blood. However, a few of the top DEGs were associated with the proportions of other cell types (e.g. *APRT* and *TMEM160* with lymphocyte proportions, and *PEPD* with neutrophil proportions). All top 40 DEGs for atopic asthma were significantly associated with eosinophils in peripheral blood. Next, we deconvolved the expression profile of WBCs based on known cell proportions and permutations in order to investigate cell type-specific differential expression using CIBERSORT [34]. 38 genes were significantly enriched in eosinophils (FDR-adjusted $p < 0.05$). Of these 38 genes, 24 were among the top DEGs for atopy and 20 were among the top DEGs for atopic asthma.

Classification models for atopy and atopic asthma

Finally, we built classification models in our discovery cohort using only demographic variables and parental history, achieving area under the curve (AUC) values of 0.56 for atopy and 0.74 for atopic asthma. These results were similar after adding expression levels for a random set of genes (data not shown). When using top DEGs alone as predictors, the models reached AUC values of 0.77 and 0.81 for atopy and atopic asthma, respectively. However, the performance of the classification models improved substantially after expression levels of the top DEGs, demographic variables and parental history were included as predictors, achieving AUC values of 0.77 for atopy and 0.84 for atopic asthma (figure 4). For atopy, we used 51 genes (FDR-adjusted $p < 0.05$ in the meta-analysis for atopy and available in both cohorts), while for atopic asthma we included the top 33 genes (FDR-adjusted $p < 0.05$ in the meta-analysis for atopic asthma and available in both cohorts).

Discussion

In a TWAS of WBCs, we identified DEGs by atopy and atopic asthma in 460 Puerto Rican children. In parallel, we conducted a TWAS of atopy and atopic asthma in 250 Swedish children, which yielded nonsignificant results due to the small sample size, but yielded replication of some of our top findings in Puerto Rican children at $p < 0.05$, in the same direction of association and with similar estimates of effect

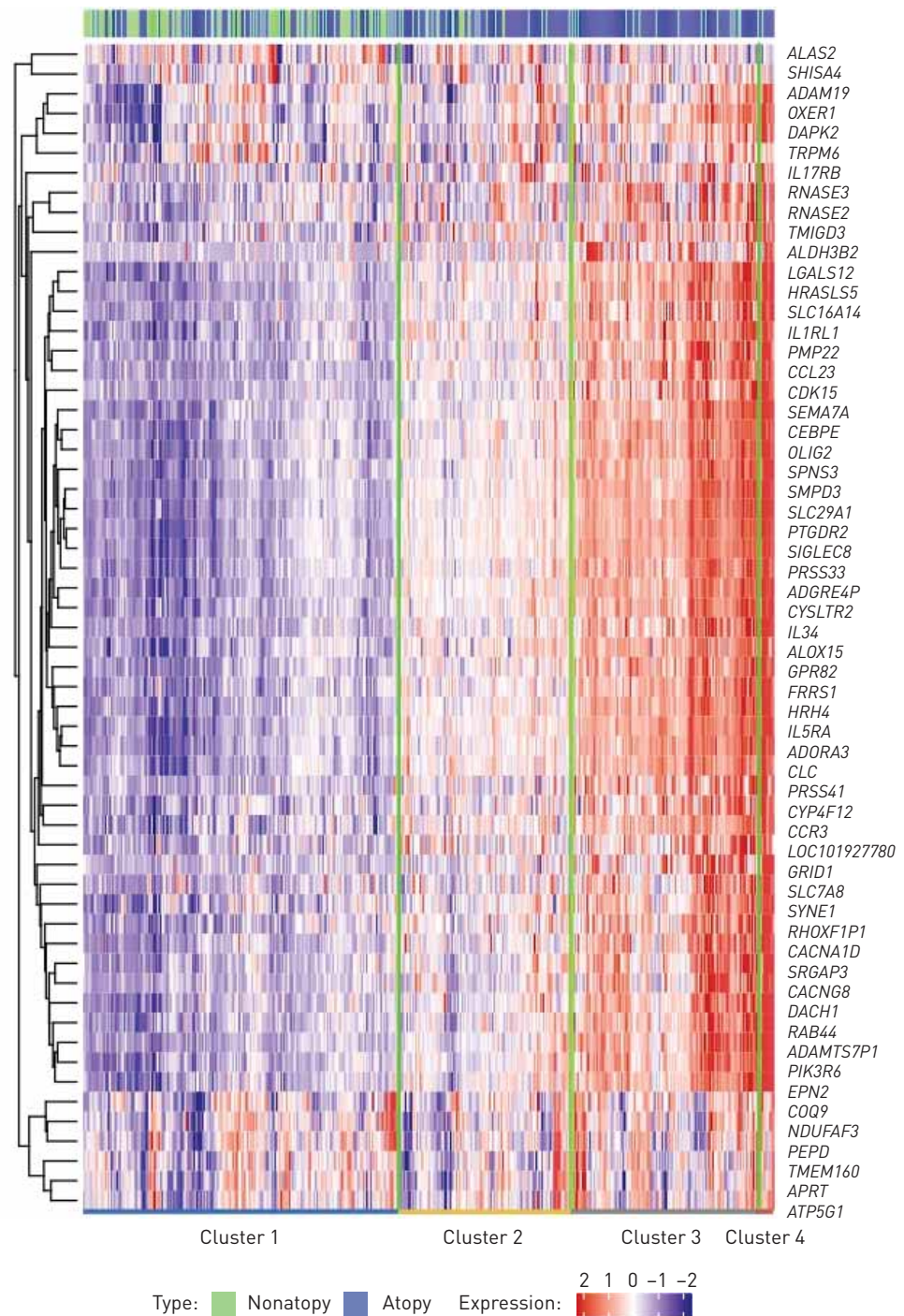


FIGURE 2 Hierarchical clustering based on the top differentially expressed genes in the meta-analysis. Four clusters are identified, with most of the genes upregulated from cluster 1 to cluster 4.

size. Next, we conducted a meta-analysis of TWASs from both study cohorts, identifying 59 DEGs by atopy and 40 DEGs by atopic asthma.

In our meta-analysis, the top DEG by atopy was *SIGLEC8*, which has an extracellular binding domain that induces apoptosis in eosinophils and inhibits FcεRI-dependent mediator release in mast cells [35]. SNPs in *SIGLEC8* were associated with asthma and eosinophilic oesophagitis in candidate gene studies [36, 37], but this gene was not associated with asthma in a recent meta-analysis of GWASs [6]. Of interest, expression of one of the top 20 genes in our meta-analysis of atopy (*OLIG2*, encoding oligodendrocyte transcription factor 2) was previously reported to downregulate RNA and protein levels of *SIGLEC8* in

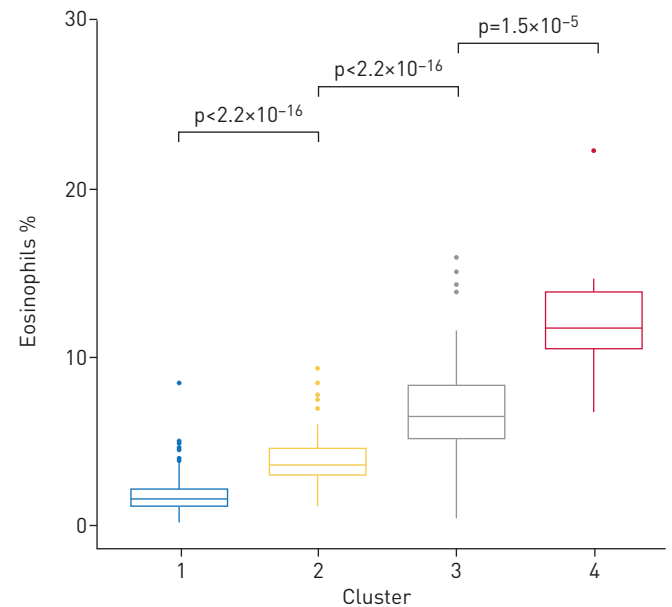


FIGURE 3 Eosinophil proportion comparison for each cluster identified in figure 2, showing significant difference both globally and between clusters individually in the Kruskal-Wallis test. Box-and-whisker plots show median with interquartile range (IQR) (boxes) and upper/lower limit within 1.5 IQR from the box range (whiskers); outliers are indicated by individual data points.

peripheral blood eosinophils [38]. Moreover, expression levels of *OLIG2* and *CCL23* (one of our top DEGs for atopic asthma) in whole blood were significantly associated with suboptimal disease control in a meta-analysis of data from 693 ethnically diverse children and adults [13]. Both *OLIG2* and *CCL23* are TREM-1 (triggering receptor expressed in myeloid cells-1)/lipopolysaccharide signalling genes, and TREM-1 is a key amplifier of inflammatory responses induced by pattern recognition receptors.

Other top study results are supported by experimental evidence or findings from other human studies. *SLC29A1* is involved in adenosine-regulated inhibition of IgE-dependent degranulation of mast cells [39]. *IL5RA*, a gene implicated in asthma in our prior study [40], regulates Th2 cell differentiation and Th2 cell effector functions [41]. Moreover, *IL5RA* and *ALOX15* (also among our top genes) were both significantly associated with eosinophil count in a transcriptomic analysis of blood samples from subjects with asthma [36]. This is consistent with our prior findings for total IgE [42] and, together with the results from our hierarchical cluster analysis and other complementary analyses, further emphasises the key role of eosinophil-dependent mechanisms in the pathogenesis of atopy and atopic asthma.

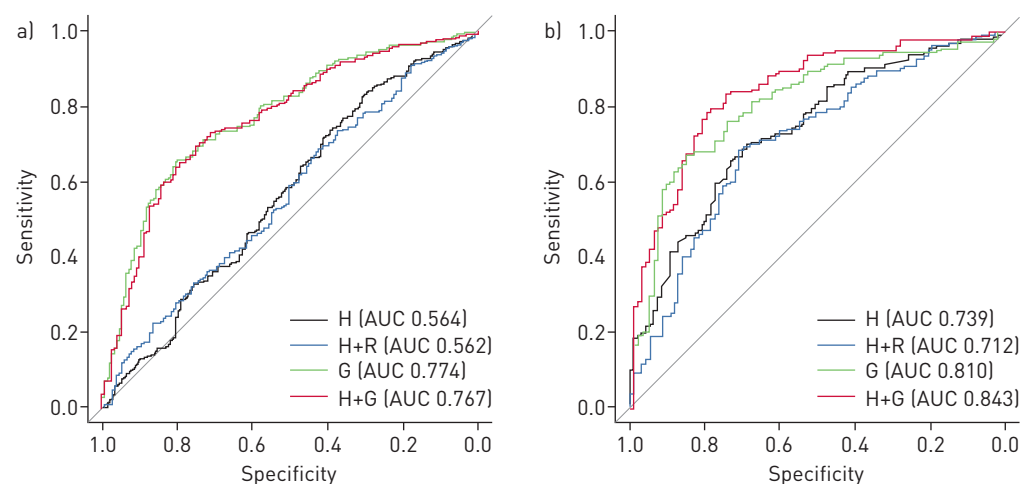


FIGURE 4 Classification models of a) atopy and b) atopic asthma. H: age, sex and asthma history of parents as predictors; R: expression profile of randomly chosen genes (the same number as G) as predictors; G: expression profile of top differentially expressed gene predictors (false discovery rate-adjusted $p < 0.05$ in the meta-analysis); AUC: area under the receiver operating characteristic curve.

Our findings for *IL17RB* are particularly interesting. In mouse models, allergen challenge triggers the interleukin (IL)-4 pathway, which promotes allergic airway inflammation and airway responsiveness. In turn, activation of IL-4 pathways induces expression of IL-25 and IL-17RB in serum and the lungs [43, 44]. In a prior study of human subjects with asthma, *IL17RB* expression in peripheral blood CD4⁺ T-cells was significantly associated with serum total IgE [45], albeit only in males. Consistent with a key role for *IL17RB* in atopy in humans, genetic deletion of *IL17RB* significantly lessens allergic sensitisation in mice [44].

We identified other plausible candidate genes for atopy and atopic asthma. *SMPD3*, encoding sphingomyelin phosphodiesterase 3 [46], is hypermethylated and downregulated in the small airways of subjects with chronic obstructive pulmonary disease [47]. *PIK3R6*, an IL-17A gene target in airway smooth muscle cells [48], has been shown to interact with second-hand smoke exposure on forced expiratory volume in 1 s. Other genes, such as *ADAMTS7P1*, *EPN2* and *RAB44*, have not been previously reported in TWASs and merit further investigation.

DEGs by atopy often overlapped with those for atopic asthma. Most of our top findings were correlated with genotypic data, emphasising a genetic predisposition to atopy. Moreover, a network analysis based on partial correlation revealed close interactions between significant DEGs.

Consistent with known mechanisms of atopy, we found that the top two most enriched pathways were the Th2 and Th1/Th2 activation pathways. Other top pathways in our analysis were that for NGF (a critical factor in the pathogenesis of allergic airway inflammation and proliferation of airway smooth muscle cells) [49] and that for adenine/adenosine salvage I (adenosine has been contraindicated in allergy or severe asthma) [50].

We compared our findings for gene expression and atopic asthma in WBCs with those of a prior meta-analysis of gene expression in airway epithelium and asthma [51]. Of the top 40 DEGs by atopic asthma in our study, seven were differentially expressed by asthma in airway epithelium (supplementary table S6). Some of these differences may be due to inclusion of cases of both atopic and nonatopic asthma in the prior analysis in airway epithelium, while others may indicate true tissue-specific differences in gene expression in atopic asthma. We also compared our results with those of a prior European study of sputum transcriptomics in 104 subjects with moderate-to-severe asthma and 16 healthy volunteers [52]. In that study, three transcriptome-associated clusters were identified, with one of them enriched for the IL-13/Th2 pathway. Of the 20 genes comprising this cluster, seven were also associated with atopy or atopic asthma (FDR-adjusted $p < 0.05$) in our meta-analysis (*PRSS33*, *CLC*, *LGALS12*, *ALOX15*, *OLIG2*, *HRH4* and *CCR3*).

We mapped the top DEGs in our meta-analysis to the top CpG sites discovered in epigenome-wide association studies (EWASs) of atopy and atopic asthma in blood and airway epithelial samples. None of the genes containing or near the top 30 CpG sites recently associated with atopy in an EWAS of nasal epithelium [53] were DEGs by atopy in our study, but two of the top DEGs by atopic asthma in our study (*IL5RA* and *SIGLEC8*) were part of differentially methylated regions by asthma in an EWAS of whole blood [54, 55].

Results from our classification models further support the validity of our top findings, as well as the plausibility of developing predictive transcriptomic panels of atopy and atopic asthma, using WBCs, in future longitudinal studies. Although classification models of atopic asthma using transcriptomic data from airway epithelium in adults have shown greater accuracy (AUC 0.99) [10] than that shown in the current study (AUC 0.84), WBCs are easier to collect than nasal or bronchial epithelium [10].

There were differences in both the type of sample (WBCs *versus* whole blood) and the platform used to assess gene expression (RNA sequencing *versus* microarray) between EVA-PR and BAMSE. This, together with differences in ethnicity and sample size between the two studies, likely explains lack of significant replication of our findings in EVA-PR in BAMSE. However, a previous study of the transcriptomics of atopic dermatitis found generally good correlation between the results of RNA sequencing and microarrays, except for genes with low expression levels [56]. Moreover, we found nominal replication at $p < 0.05$, with consistent direction and estimated effect size, for some of our top findings in EVA-PR in BAMSE. Although there were no significant DEGs for atopy or atopic asthma in BAMSE after FDR correction for multiple testing, the top genes associated with atopy or atopic asthma in BAMSE (shown in supplementary table S7) were enriched in pathways for energy-coupled proton and ion transmembrane transport.

In summary, our combined transcriptomic analysis in Puerto Rican and Swedish children identified 59 DEGs by atopy and 40 DEGs by atopic asthma, including *SIGLEC8* and *IL17RB*. Transcriptomic data from WBCs could be used to develop predictive models of atopy and atopic asthma in future prospective studies.

Acknowledgements: We thank all the participating children and families. The BAMSE team would like to thank Charles Auffray (European Institute for Systems Biology and Medicine, Lyon, France) and his team for generating the transcriptomics data.

Author contributions: W. Chen, E. Melén and J.C. Celedón conceived and designed the study. Y. Jiang, O. Gruziova and S.K. Merid conducted the primary analysis and interpreted data. E. Forno, T. Wang, E. Acosta-Pérez, G. Canino, I. Kull, J.M. Antó and J. Bousquet participated in data collection and data analysis. Y. Jiang, W. Chen and J.C. Celedón prepared the first draft of the manuscript. All authors reviewed the draft for intellectual content and approved submission of the final version of the manuscript.

Conflict of interest: Y. Jiang has nothing to disclose. O. Gruziova has nothing to disclose. T. Wang has nothing to disclose. E. Forno has nothing to disclose. N. Boutaoui has nothing to disclose. T. Sun has nothing to disclose. S.K. Merid has nothing to disclose. E. Acosta-Pérez has nothing to disclose. I. Kull has nothing to disclose. G. Canino has nothing to disclose. J.M. Antó has nothing to disclose. J. Bousquet reports personal fees for advisory board work, consultancy and lectures from Chiesi, Cipla, Hikma, Menarini, Mundipharma, Mylan, Novartis, Sanofi-Aventis, Takeda, Teva and Uriach, and owns shares in Kyomed, outside the submitted work. E. Melén has nothing to disclose. W. Chen has nothing to disclose. J.C. Celedón has received research materials from GSK, Merck and Pharmavite, outside the submitted work.

Support statement: This study was supported by grants HL079966, HL117191 and MD011764 (Principal Investigator: J.C. Celedón) from the US National Institutes of Health (NIH). E. Forno's contribution was supported by NIH grant HL125666. Y. Jiang was supported by the China Scholarship Council. The BAMSE cohort was supported by grants from the Swedish Heart-Lung Foundation, the Swedish Research Council, Stockholm County Council (ALF), the Strategic Research Program (SFO) in Epidemiology at Karolinska Institutet and MeDALL (Mechanisms of the Development of ALLergy; European Union grant agreement number 261357). E. Melén is supported by a grant from the European Research Council (grant agreement number 757919, TRIBAL). Funding information for this article has been deposited with the Crossref Funder Registry.

References

- Pearce N, Ait-Khaled N, Beasley R, *et al.* Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax* 2007; 62: 758–766.
- Jacobs TS, Forno E, Brehm JM, *et al.* Underdiagnosis of allergic rhinitis in underserved children. *J Allergy Clin Immunol* 2014; 134: 737–739.
- Rosser FJ, Forno E, Cooper PJ, *et al.* Asthma in Hispanics. An 8-year update. *Am J Respir Crit Care Med* 2014; 189: 1316–1327.
- Akinbami LJ, Moorman JE, Bailey C, *et al.* Trends in asthma prevalence, health care use, and mortality in the United States, 2001–2010. *NCHS Data Brief* 2012; (94): 1–8.
- Bjerg A, Sandstrom T, Lundback B, *et al.* Time trends in asthma and wheeze in Swedish children 1996–2006: prevalence and risk factors by sex. *Allergy* 2010; 65: 48–55.
- Demais F, Marguerite-Jeannin P, Barnes KC, *et al.* Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet* 2018; 50: 42–53.
- Waage J, Standl M, Curtin JA, *et al.* Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nat Genet* 2018; 50: 1072–1080.
- Yick CY, Zwiderman AH, Kunst PW, *et al.* Transcriptome sequencing (RNA-Seq) of human endobronchial biopsies: asthma versus controls. *Eur Respir J* 2013; 42: 662–670.
- Poole A, Urbanek C, Eng C, *et al.* Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. *J Allergy Clin Immunol* 2014; 133: 670–678.
- Pandey G, Pandey OP, Rogers AJ, *et al.* A nasal brush-based classifier of asthma identified by machine learning analysis of nasal RNA sequence data. *Sci Rep* 2018; 8: 8826.
- Pascual M, Roa S, Garcia-Sanchez A, *et al.* Genome-wide expression profiling of B lymphocytes reveals IL4R increase in allergic asthma. *J Allergy Clin Immunol* 2014; 134: 972–975.
- Howrylak JA, Moll M, Weiss ST, *et al.* Gene expression profiling of asthma phenotypes demonstrates molecular signatures of atopy and asthma control. *J Allergy Clin Immunol* 2016; 137: 1390–1397.
- Croteau-Chonka DC, Qiu W, Martinez FD, *et al.* Gene expression profiling in blood provides reproducible molecular insights into asthma control. *Am J Respir Crit Care Med* 2017; 195: 179–188.
- Bird HR, Davies M, Duarte CS, *et al.* A study of disruptive behavior disorders in Puerto Rican youth: II. Baseline prevalence, comorbidity, and correlates in two sites. *J Am Acad Child Adolesc Psychiatry* 2006; 45: 1042–1053.
- Brehm JM, Acosta-Pérez E, Klei L, *et al.* Vitamin D insufficiency and severe asthma exacerbations in Puerto Rican children. *Am J Respir Crit Care Med* 2012; 186: 140–146.
- Forno E, Cloutier MM, Datta S, *et al.* Mouse allergen, lung function, and atopy in Puerto Rican children. *PLoS One* 2012; 7: e40383.
- Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. www.bioinformatics.babraham.ac.uk/projects/fastqc Date last accessed: March 8, 2019.
- Krueger F. Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 0.4. 2015. www.bioinformatics.babraham.ac.uk/projects/trim_galore Date last accessed: March 8, 2019.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011; 17: 10–12.
- Dobin A, Davis CA, Schlesinger F, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15–21.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011; 12: 323.
- Brehm JM, Acosta-Pérez E, Klei L, *et al.* African ancestry and lung function in Puerto Rican children. *J Allergy Clin Immunol* 2012; 129: 1484–1490.
- Das S, Forer L, Schönherr S, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016; 48: 1284–1287.
- McCarthy S, Das S, Kretschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; 48: 1279–1283.

- 25 Ballardini N, Bergström A, Wahlgren C-F, *et al.* IgE antibodies in relation to prevalence and multimorbidity of eczema, asthma, and rhinitis from birth to adolescence. *Allergy* 2016; 71: 342–349.
- 26 Hallberg J, Ballardini N, Almqvist C, *et al.* Impact of IgE sensitization and rhinitis on inflammatory biomarkers and lung function in adolescents with and without asthma. *Pediatr Allergy Immunol* 2019; 30: 74–80.
- 27 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15: 550.
- 28 Lee CH, Cook S, Lee JS, *et al.* Comparison of two meta-analysis methods: inverse-variance-weighted average and weighted sum of z-scores. *Genomics Inform* 2016; 14: 173–180.
- 29 Schwarzer G. meta: an R package for meta-analysis. *R News* 2007; 7: 40–45.
- 30 Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012; 28: 1353–1358.
- 31 Krämer A, Green J, Pollard J Jr, *et al.* Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2013; 30: 523–530.
- 32 Wang T, Ren Z, Ding Y, *et al.* FastGGM: an efficient algorithm for the inference of Gaussian graphical model in biological networks. *PLoS Comput Biol* 2016; 12: e1004755.
- 33 Lonsdale J, Thomas J, Salvatore M, *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; 45: 580–585.
- 34 Newman AM, Liu CL, Green MR, *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015; 12: 453–457.
- 35 Yokoi H, Choi OH, Hubbard W, *et al.* Inhibition of FcεRI-dependent mediator release and calcium flux from human mast cells by sialic acid-binding immunoglobulin-like lectin 8 engagement. *J Allergy Clin Immunol* 2008; 121: 499–505.
- 36 Choy DF, Jia G, Abbas AR, *et al.* Peripheral blood gene expression predicts clinical benefit from anti-IL-13 in asthma. *J Allergy Clin Immunol* 2016; 138: 1230–1233.
- 37 Gao P-S, Shimizu K, Grant AV, *et al.* Polymorphisms in the sialic acid-binding immunoglobulin-like lectin-8 (Siglec-8) gene are associated with susceptibility to asthma. *Eur J Hum Genet* 2010; 18: 713–719.
- 38 Hwang SM, Uhm TG, Lee SK, *et al.* Olig2 is expressed late in human eosinophil development and controls Siglec-8 expression. *J Leukoc Biol* 2016; 100: 711–723.
- 39 Gomez G, Nardone V, Lotfi-Emran S, *et al.* Intracellular adenosine inhibits IgE-dependent degranulation of human skin mast cells. *J Clin Immunol* 2013; 33: 1349–1359.
- 40 Forno E, Wang T, Yan Q, *et al.* A multiomics approach to identify genes associated with childhood asthma risk and morbidity. *Am J Respir Cell Mol Biol* 2017; 57: 439–447.
- 41 Vercelli D. Discovering susceptibility genes for asthma and allergy. *Nat Rev Immunol* 2008; 8: 169–182.
- 42 Chen W, Wang T, Pino-Yanes M, *et al.* An epigenome-wide association study of total serum IgE in Hispanic children. *J Allergy Clin Immunol* 2017; 140: 571–577.
- 43 Weathington NM, Kanth SM, Gong Q, *et al.* IL-4 induces *IL17Rb* gene transcription in monocytic cells with coordinate autocrine IL-25 signaling. *Am J Respir Cell Mol Biol* 2017; 57: 346–354.
- 44 Petersen BC, Budelsky AL, Baptist AP, *et al.* Interleukin-25 induces type 2 cytokine production in a steroid-resistant interleukin-17RB⁺ myeloid population that exacerbates asthmatic pathology. *Nat Med* 2012; 18: 751–758.
- 45 Hunninghake GM, Chu JH, Sharma SS, *et al.* The CD4⁺ T-cell transcriptome and serum IgE in asthma: IL17RB and the role of sex. *BMC Pulm Med* 2011; 11: 17.
- 46 Filosto S, Castillo S, Danielson A, *et al.* Neutral sphingomyelinase 2: a novel target in cigarette smoke-induced apoptosis and lung injury. *Am J Respir Cell Mol Biol* 2011; 44: 350–360.
- 47 Vucic EA, Chari R, Thu KL, *et al.* DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol* 2014; 50: 912–922.
- 48 Dragon S, Hirst SJ, Lee TH, *et al.* IL-17A mediates a selective gene expression profile in asthmatic human airway smooth muscle cells. *Am J Respir Cell Mol Biol* 2014; 50: 1053–1063.
- 49 Yang Y-G, Tian W-M, Zhang H, *et al.* Nerve growth factor exacerbates allergic lung inflammation and airway remodeling in a rat model of chronic asthma. *Exp Ther Med* 2013; 6: 1251–1258.
- 50 Layland J, Carrick D, Lee M, *et al.* Adenosine: physiology, pharmacology, and clinical applications. *JACC Cardiovasc Interv* 2014; 7: 581–591.
- 51 Tsai Y-H, Parker JS, Yang IV, *et al.* Meta-analysis of airway epithelia gene expression in asthma. *Eur Respir J* 2018; 51: 1701962.
- 52 Kuo C-HS, Pavlidis S, Loza M, *et al.* T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J* 2017; 49: 1602135.
- 53 Forno E, Wang T, Qi C, *et al.* DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. *Lancet Respir Med* 2019; 7: 336–346.
- 54 Reese SE, Xu CJ, den Dekker HT, *et al.* Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *J Allergy Clin Immunol* 2018; in press [https://doi.org/10.1016/j.jaci.2018.11.043].
- 55 Xu CJ, Soderhall C, Bustamante M, *et al.* DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med* 2018; 6: 379–388.
- 56 Suárez-Fariñas M, Ungar B, da Rosa JC, *et al.* RNA sequencing atopic dermatitis transcriptome profiling provides insights into novel disease mechanisms with potential therapeutic implications. *J Allergy Clin Immunol* 2015; 135: 1218–1227.

Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents

Yale Jiang, MDc^{*}, Olena Gruzieva, PhD^{*}, Ting Wang, PhD, Erick Forno, MD, MPH, Nadia Boutaoui, PhD, Tao Sun, MS, Simon K. Merid, MSc, Edna Acosta-Pérez, PhD, Inger Kull, PhD, Glorisa Canino, PhD, Josep M. Antó, MD, PhD, Jean Bousquet, MD, PhD, Erik Melén, MD, PhD^{**}, Wei Chen, PhD^{**}, Juan C. Celedón, MD, DrPH, ATSF^{**}

^{*}Equal contribution ^{**}Shared senior authors

Online Data Supplement

Table of contents	
Online Methods	- Study population - RNA sequencing and data preprocessing in EVE-PR - Gene expression data processing in BAMSE - Disease definition - Network analysis - Classification model
Table S1	Top differentially expressed genes (DEGs) for atopy in the two studies (FDR-adjusted $P < 0.05$)
Table S2.	Top 20 DEGs in TWAS of atopic asthma in meta-analysis.
Table S3	Functional ingenuity pathways enriched by top DEGs in the meta-analysis of atopic asthma.
Table S4	The most significant Cis-eQTL(1Mb) pairs for each significant DEG (FDR < 0.05)
Table S5	Associations between top DEGs by atopy or atopic asthma and cell type proportions
Table S6	Top DEGs of atopic asthma get validated in TWAS of asthma in airway epithelium
Table S7	Top differentially expressed genes in BAMSE
Fig S1	(a) Heatmap of DEGs between atopic subjects and controls. (b) Expression of <i>SIGLEC8</i> and (c) <i>IL17RB</i> in atopic and control subjects.
Fig S2	Network analysis of significant genes in EVA-PR.
Fig S3	Bar plot of atopy proportion in 4 distinct clusters identified from DEGs

Online Methods

Study population

Participants in the *Epigenetic Variation and Childhood Asthma in Puerto Ricans* study (EVA-PR) were recruited from households in the metropolitan area of San Juan and Caguas (Puerto Rico) from October 2013 through August 2017. Of 1,111 households selected using multi-stage probabilistic sampling in a previous study [1], 638 had ≥ 1 eligible child (i.e. aged 9 to 20 years, with four Puerto Rican grandparents), and 543 (85.1%) children (one per household) agreed to participate. There were no significant differences in age, sex, or area of residence between eligible children who did and did not participate. We selected as cases children who had physician-diagnosed asthma and wheeze in the previous year. We selected as controls children who had neither physician-diagnosed asthma nor wheeze in the prior year. All study participants had to have four Puerto Rican grandparents to ensure their Puerto Rican descent.

The study protocol included questionnaires on respiratory and general health, and collection of blood samples for measurement of allergen-specific IgEs and RNA extraction from WBCs. Plasma levels of IgE specific to five common allergens (dust mite [Der p 1], cockroach [Bla g 2], cat dander [Fel d 1], dog dander [Can f 1], and mouse urinary protein [Mus m 1]) were determined by using the UniCAP 100 system (Pharmacia & Upjohn, Kalamazoo, Mich). For each allergen, an IgE level of 0.35 IU/mL or greater was considered positive. WBC count, including percentages of WBC subtypes, was conducted in peripheral blood using a Coulter-Counter technique.

BAMSE (Swedish: *Barn, Allergi, Miljö, Stockholm, Epidemiologi*) is a population-based birth cohort in which 4,089 infants born between 1994-1996 were recruited at birth and prospectively followed during childhood and adolescence. Information on demographic characteristics, parental allergies, and child's health has been collected from repeated questionnaires up to the age of 16 years[2]. Response rates were high, with as many as 78% of the original cohort remaining at 16-year-follow-up. Blood samples were collected at ages 4, 8, and 16 years. Sera were screened with Phadiatop® (a mix of typical airborne allergens: birch, timothy, mugwort, cat, dog, horse, mold [*Cladosporium herbarum*], house dust mite [*Dermatophagoides pteronyssinus*], followed by specific allergen IgE analyses if positive (ImmunoCAP System; Thermo Fisher/Phadia AB, Uppsala, Sweden). Automated cell counts were obtained by flow cytometry performed at the Karolinska University Laboratory in Stockholm, Sweden.

RNA sequencing and data preprocessing in EVA-PR

Sequencing was performed at the Genomics and Proteomics Research Core of the University of Pittsburgh, using 350 ng of high-quality RNA extracted from WBCs after removing hemoglobin (RNA Integrity Number [RIN] > 7). Libraries were run on the NextSeq 500 using NextSeq® 500/550 High Output Kit v2, using paired-end reads at 75 cycles and with 80 million reads per sample. After quality control, alignment and trimming, an additional filtering of low-expressed genes (mean count < 2) was performed. After preprocessing, a total of 16,880 genes and 460 samples were retained in the final analysis.

Table 1. Summary of RNA sequencing mapping quality in EVA-PR

	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
Total_Reads	29,615,100	51,563,242	57,626,763	57,193,283	62,509,597	93,838,498
Remaining_Reads	29,179,226	51,223,011	56,791,896	56,625,302	61,966,716	92,796,617
Uniquely-mapped_Rate(%)	66.3	83.22	85.28	84.92	87.28	92.22
Multi-mapped_Rate(%)	3.39	5.425	6.93	7.44	8.855	16.51
Unmapped_Rate(%)	3.03	5.22	7.285	7.642	9.162	29.09
Duplicate_Rate(%)	7.35	13.52	16.21	17.51	20.36	46.26
#Gene(count>0)	16,021	17,099	17,308	17,292	17,468	20,813
Total_count	7,471,063	15,423,764	19,524,916	19,575,059	23,259,733	41,361,210

Gene expression data processing in BAMSE

RNA gene expression in BAMSE was available from pilot phase 45 samples and extended phase 224 samples at 16 years old children. Whole blood was collected in PAXGene tubes, and RNA was extracted using the PAXgene Blood RNA kit (QIAGEN). Assessments of extracted RNA yield and quality were performed with Dropsense96 (Trinean) and Tapestation (Agilent) instruments, respectively. RNA of highest quality was selected for amplification, labeling, and hybridization on Affymetrix HTA 2.0 Genechips using the Affymetrix WT PLUS kit (Affymetrix Inc.). The quality assessment discarded six samples in total [3]. Data analysis was based on the space signal transformation robust multiarray average (SST-RMA) algorithm [4] which combines Guanine Cytosine Count Normalization (GCCN) and Signal Space Transformation (SST) approaches, first normalizing intensities based on the probe affinity difference associated with GC content, then ‘stretching’ the intensity distribution by decompressing the Fold Change ratios with a power law mapping, prior to applying RMA algorithm with quantile normalization [5]. The Combat method [6] was used to adjust for batch effect between pilot and extended phases. Data production (RNA extraction, QC, preparation, hybridization, CEL files) and analysis (pre-processing) was performed

by the European Institute for Systems Biology and Medicine in Lyon. Automated cell counts were obtained by flow cytometry performed at the Karolinska University Laboratory in Stockholm, Sweden.

Disease definition

Current asthma was defined as doctor diagnosed asthma plus wheezing or medications on asthma in the past year. Current rhinitis was defined as the presence of hay fever or a runny or stuffy nose accompanied by sneezing and itching at a time when he/she did not have a cold or flu in the last 12 months. Current dermatitis was defined as the presence of itchy rash which was coming and going for at least six months or at any time in the last 12 months.

Network analysis

Gaussian graphical model (GGM) is characterized for conditional dependence among large sets of random variables and embedded in the R package FastGGM [7]. We applied the algorithm to calculate the partial correlation between gene expression levels in our discovery cohort, in all subjects and separately by asthma status respectively. The gene sets are limited to the 59 top DEGs identified in meta-analysis of atopy ($FDR < 0.05$) to relatively accurately estimate the correlation between the genes on condition of the whole-transcriptome expression profile.

Classification model

Logistic regression with penalized maximum likelihood was applied to build classification models using R package caret [8]. The inclusion criteria of genes as predictors is that the gene should be among the top DEGs in the meta-analysis and available in both datasets, for atopy (51 genes) and atopic asthma (33 genes) respectively. For each phenotype, 4 models altogether were build for comparison. First, 1) one including only demographic and parental history (age, gender, parental history of asthma) as predictors was made, and 2) another including demographic and parental history, and gene expression levels of randomly chosen genes was built as negative control. Next, we built 3) one model only using the top DEGs in our meta-analysis and 4) another one using the top DEGs plus demographic and parental history to see whether our identified genes can significantly improve the performance of prediction model.

References

1. Brehm, J.M., et al., *African ancestry and lung function in Puerto Rican children*. J Allergy Clin Immunol, 2012. **129**(6): p. 1484-1490 e6.

2. Ballardini, N., et al., *IgE antibodies in relation to prevalence and multimorbidity of eczema, asthma, and rhinitis from birth to adolescence*. Allergy, 2016. **71**(3): p. 342-9.
3. Xu, C.J., et al., *DNA methylation in childhood asthma: an epigenome-wide meta-analysis*. Lancet Respir Med, 2018. **6**(5): p. 379-388.
4. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data*. Biostatistics, 2003. **4**(2): p. 249-64.
5. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias*. Bioinformatics, 2003. **19**(2): p. 185-93.
6. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods*. Biostatistics, 2007. **8**(1): p. 118-27.
7. Wang, T., et al., *FastGGM: An Efficient Algorithm for the Inference of Gaussian Graphical Model in Biological Networks*. PLoS Comput Biol, 2016. **12**(2): p. e1004755.
8. Kuhn, M., *Building predictive models in R using the caret package*. Journal of statistical software, 2008. **28**(5): p. 1-26.

Table S1. Top differentially expressed genes (DEGs) for atopy in the two studies (FDR-adjusted P value < 0.05)

Gene	EVA-PR			BAMSE			
	Effect size	P value	FDR	Effect size	P value	FDR*	Significance**
<i>IL17RB</i>	1.19	1.29E-14	2.18E-10	1.01	2.39E-01	NS	
<i>SIGLEC8</i>	1.18	2.97E-12	2.51E-08	1.05	9.34E-03	NS	+
<i>OLIG2</i>	1.17	5.22E-12	2.94E-08	1.01	3.18E-01	NS	
<i>SMPD3</i>	1.15	2.44E-10	1.03E-06	1.04	5.31E-03	NS	+
<i>CCL23</i>	1.14	3.13E-10	1.06E-06	0.99	5.03E-01	NS	
<i>PRSS33</i>	1.15	7.21E-10	2.03E-06	1.01	3.15E-01	NS	
<i>ALOX15</i>	1.15	9.79E-10	2.11E-06	1.11	1.22E-02	NS	+
<i>PMP22</i>	1.15	1.00E-09	2.11E-06	1.02	1.75E-01	NS	
<i>SLC29A1</i>	1.15	1.23E-09	2.31E-06	1.05	9.24E-04	NS	+
<i>IL5RA</i>	1.15	2.06E-09	3.48E-06	1.18	1.97E-04	NS	+
<i>IL1RL1</i>	1.14	6.15E-09	8.94E-06	1.13	2.24E-03	NS	+
<i>ADORA3</i>	1.14	6.58E-09	8.94E-06	1.04	2.47E-02	NS	+
<i>IL34</i>	1.14	7.33E-09	8.94E-06	1.01	5.46E-01	NS	
<i>ADAMTS7P1</i>	1.14	7.41E-09	8.94E-06	NA	NA	NA	
<i>CDK15</i>	1.14	8.22E-09	9.25E-06	1.01	3.56E-01	NS	
<i>HRASLS5</i>	1.14	2.11E-08	2.19E-05	1.02	1.60E-01	NS	
<i>TFF3</i>	1.12	2.21E-08	2.19E-05	0.97	5.15E-02	NS	
<i>CLC</i>	1.14	2.54E-08	2.38E-05	1.17	9.22E-03	NS	+
<i>EPN2</i>	1.11	4.59E-08	4.07E-05	NA	NA	NA	
<i>PIK3R6</i>	1.12	9.07E-08	7.65E-05	1.05	2.41E-03	NS	+
<i>DACH1</i>	1.11	1.26E-07	1.01E-04	1.02	1.20E-01	NS	
<i>RHOXF1P1</i>	1.13	1.72E-07	1.32E-04	NA	NA	NA	
<i>ADGRE4P</i>	1.12	3.07E-07	2.17E-04	NA	NA	NA	
<i>PRSS41</i>	1.11	3.09E-07	2.17E-04	1.01	6.01E-01	NS	
<i>PTGDR2</i>	1.12	3.85E-07	2.60E-04	1.04	2.34E-02	NS	+
<i>ALDH3B2</i>	1.08	4.43E-07	2.88E-04	1.01	5.84E-01	NS	
<i>LOC101927780</i>	1.12	5.24E-07	3.27E-04	NA	NA	NA	
<i>KHDRBS3</i>	1.12	7.07E-07	4.26E-04	0.99	2.54E-01	NS	
<i>CEBPE</i>	1.12	1.04E-06	6.04E-04	1.07	2.32E-03	NS	+
<i>RNASE2</i>	1.11	1.07E-06	6.04E-04	1.07	2.25E-01	NS	
<i>CACNG8</i>	1.12	1.24E-06	6.64E-04	NA	NA	NA	
<i>CACNA1D</i>	1.11	1.26E-06	6.64E-04	1.02	9.27E-02	NS	
<i>DUOX2</i>	1.09	2.42E-06	1.24E-03	0.99	6.33E-01	NS	
<i>SLC7A8</i>	1.11	3.10E-06	1.54E-03	1.01	5.36E-01	NS	
<i>SLC16A14</i>	1.11	3.90E-06	1.88E-03	1.00	6.91E-01	NS	
<i>GPR82</i>	1.11	4.03E-06	1.89E-03	1.03	1.23E-01	NS	
<i>HRH4</i>	1.11	4.78E-06	2.18E-03	1.10	7.45E-03	NS	+
<i>LONRF2</i>	1.10	5.56E-06	2.47E-03	0.98	1.04E-01	NS	
<i>GRID1</i>	1.10	6.69E-06	2.90E-03	1.01	4.27E-01	NS	
<i>RNASE3</i>	1.10	6.89E-06	2.91E-03	1.02	5.68E-01	NS	
<i>TMIGD3</i>	1.11	9.20E-06	3.79E-03	NA	NA	NA	

<i>AJAP1</i>	1.11	1.17E-05	4.69E-03	1.00	9.14E-01	NS	
<i>CCR3</i>	1.11	1.20E-05	4.72E-03	1.15	4.20E-03	NS	+
<i>NDUFAF3</i>	0.95	1.32E-05	5.05E-03	0.98	1.04E-01	NS	
<i>THBS4</i>	1.10	2.29E-05	8.58E-03	1.01	6.19E-01	NS	
<i>ABO</i>	1.08	2.49E-05	9.12E-03	1.00	9.30E-01	NS	
<i>RIMKLA</i>	1.10	2.58E-05	9.26E-03	0.98	3.39E-01	NS	
<i>CPLX2</i>	1.08	2.86E-05	1.01E-02	0.99	5.66E-01	NS	
<i>LYPD1</i>	1.08	3.31E-05	1.14E-02	1.00	7.06E-01	NS	
<i>PDZK1IP1</i>	0.91	3.75E-05	1.25E-02	0.98	5.18E-01	NS	
<i>RNF214</i>	1.05	3.77E-05	1.25E-02	1.01	5.71E-01	NS	
<i>PRPF38B</i>	1.04	4.71E-05	1.53E-02	1.01	5.40E-01	NS	
<i>ACOT11</i>	1.09	6.49E-05	2.07E-02	1.00	7.91E-01	NS	
<i>RAB44</i>	1.08	6.77E-05	2.11E-02	1.07	2.39E-04	NS	+
<i>SRGAP3</i>	1.08	8.31E-05	2.55E-02	NA	NA	NA	
<i>SHISA4</i>	0.92	9.20E-05	2.77E-02	0.96	3.83E-02	NS	+
<i>RAB27A</i>	1.06	1.00E-04	2.94E-02	1.02	3.24E-01	NS	
<i>DNAJC6</i>	0.91	1.01E-04	2.94E-02	0.99	6.02E-01	NS	
<i>PPP1R1B</i>	1.06	1.05E-04	2.99E-02	1.00	9.53E-01	NS	
<i>ADAM19</i>	1.07	1.10E-04	3.10E-02	1.05	8.80E-04	NS	+
<i>ALAS2</i>	0.93	1.12E-04	3.10E-02	0.91	1.41E-01	NS	
<i>OLIG1</i>	1.09	1.20E-04	3.26E-02	0.98	1.56E-01	NS	
<i>VLDLR</i>	1.09	1.22E-04	3.26E-02	0.99	5.60E-01	NS	
<i>RBM25</i>	1.04	1.28E-04	3.38E-02	1.02	3.11E-01	NS	
<i>LRRC17</i>	1.09	1.44E-04	3.75E-02	0.98	1.21E-01	NS	
<i>RGS1</i>	1.09	1.49E-04	3.81E-02	0.96	2.03E-02	NS	+
<i>NAGPA</i>	0.95	1.52E-04	3.82E-02	1.00	8.19E-01	NS	
<i>AP3B2</i>	0.96	1.55E-04	3.84E-02	0.99	5.75E-01	NS	
<i>TRPM6</i>	1.08	1.82E-04	4.46E-02	1.05	6.03E-02	NS	

All results were adjusted for asthma status, age, gender, batch, cell proportions of eosinophils, neutrophils, monocytes, and lymphocytes. 5 PCs of genotype were additionally adjusted in EVA-PR study.

NA means that the expression of the gene is not available for analysis.

*NS means non-significant, i.e., the FDR-adjusted P value > 0.05 in BAMSE.

**A plus sign indicates raw P value < 0.05 in BAMSE.

Table S2. Top 20 DEGs in TWAS of atopic asthma in meta-analysis

Gene	Description	EVA-PR			BAMSE			Meta-analysis		
		Fold Change	P Value	FDR	Fold Change	P Value	FDR*	Fold Change	P Value	FDR
<i>SLC29A1</i>	Solute carrier family 29 member 1 (Augustine blood group)	1.20	2.95E-10	1.25E-06	1.06	6.86E-03	NS	1.10	2.71E-11	6.97E-07
<i>OLIG2</i>	Oligodendrocyte transcription factor 2	1.22	4.24E-12	7.15E-08	1.04	7.57E-02	NS	1.09	3.94E-11	6.97E-07
<i>CLC</i>	Charcot-Leyden crystal galectin	1.18	2.09E-08	3.10E-05	1.32	7.28E-04	NS	1.19	7.52E-11	8.87E-07
<i>SIGLEC8</i>	Sialic acid binding Ig like lectin 8	1.21	5.94E-11	3.34E-07	1.05	5.91E-02	NS	1.13	1.94E-10	1.56E-06
<i>CEBPE</i>	CCAAT/enhancer binding protein epsilon	1.20	5.07E-10	1.55E-06	1.08	2.19E-02	NS	1.14	2.20E-10	1.56E-06
<i>IL17RB</i>	Interleukin 17 receptor B	1.19	6.41E-10	1.55E-06	1.03	2.96E-02	NS	1.06	4.17E-10	2.46E-06
<i>ALOX15</i>	Arachidonate 15-lipoxygenase	1.17	2.97E-08	3.86E-05	1.17	3.68E-03	NS	1.17	6.58E-10	3.32E-06
<i>IL5RA</i>	Interleukin 5 receptor subunit alpha	1.18	2.21E-08	3.10E-05	1.18	9.85E-03	NS	1.18	1.71E-09	7.19E-06
<i>SMPD3</i>	Sphingomyelin phosphodiesterase 3	1.19	2.65E-09	5.60E-06	1.05	3.68E-02	NS	1.09	1.83E-09	7.19E-06
<i>ADORA3</i>	Adenosine A3 receptor	1.20	6.28E-10	1.55E-06	1.05	9.76E-02	NS	1.11	2.88E-09	1.02E-05
<i>IL1RL1</i>	Interleukin 1 receptor like 1	1.17	1.10E-07	1.03E-04	1.16	6.07E-03	NS	1.17	3.79E-09	1.22E-05
<i>HRH4</i>	Histamine receptor H4	1.14	8.57E-06	4.52E-03	1.17	1.26E-03	NS	1.15	3.90E-08	1.14E-04
<i>HRASLS5</i>	HRAS like suppressor family member 5	1.19	6.55E-09	1.23E-05	1.03	1.60E-01	NS	1.08	4.19E-08	1.14E-04
<i>PRSS33</i>	Protease, serine 33	1.16	2.00E-07	1.69E-04	1.04	3.13E-02	NS	1.08	5.05E-08	1.28E-04
<i>CCL23</i>	C-C Motif Chemokine Ligand 23	1.19	1.11E-11	9.36E-08	1.00	8.66E-01	NS	1.07	9.94E-08	2.19E-04
<i>PTGDR2</i>	Prostaglandin D2 receptor 2	1.17	3.57E-08	4.01E-05	1.04	1.26E-01	NS	1.09	1.01E-07	2.19E-04
<i>IL34</i>	Interleukin 34	1.18	1.26E-08	2.12E-05	1.03	2.01E-01	NS	1.08	1.05E-07	2.19E-04
<i>RAB44</i>	Member RAS oncogene family	1.12	4.10E-05	1.78E-02	1.09	1.13E-03	NS	1.10	1.64E-07	3.06E-04
<i>PIK3R6</i>	Phosphoinositide-3-kinase regulatory subunit 6	1.12	2.25E-05	1.09E-02	1.07	2.31E-03	NS	1.09	1.80E-07	3.19E-04
<i>PMP22</i>	Peripheral myelin protein 22	1.17	3.27E-08	3.95E-05	1.02	4.18E-01	NS	1.07	9.41E-07	1.39E-03

All results were adjusted for age, gender, batch in both studies. 5 PCs of genotype were additionally adjusted in EVA-PR study.

*NS means non-significant, i.e., the FDR-adjusted P value > 0.05 in BAMSE.

Table S3. Functional Ingenuity Pathways enriched by top DEGs in the meta-analysis of atopic asthma

Ingenuity Canonical Pathways	-log(P-value)	Molecules
Th2 Pathway	5.06	<i>CCR3,PTGDR2,IL17RB,IL1RL1,PIK3R6</i>
Th1 and Th2 Activation Pathway	4.62	<i>CCR3,PTGDR2,IL17RB,IL1RL1,PIK3R6</i>
Type II Diabetes Mellitus Signaling	3.71	<i>CACNA1D,PIK3R6,CACNG8,SMPD3</i>
FcγRIIB Signaling in B Lymphocytes	3.26	<i>CACNA1D,PIK3R6,CACNG8</i>
Amyotrophic Lateral Sclerosis Signaling	2.93	<i>CACNA1D,CAT,PIK3R6</i>
Atherosclerosis Signaling	2.76	<i>CCR3,ALOX15,CSF1</i>
Sorbitol Degradation I	2.73	<i>SORD</i>
nNOS Signaling in Skeletal Muscle Cells	2.57	<i>CACNA1D,CACNG8</i>
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2.52	<i>IL1RL1,CSF1,PIK3R6,CEBPE</i>
PKCθ Signaling in T Lymphocytes	2.44	<i>CACNA1D,PIK3R6,CACNG8</i>
Docosahexaenoic Acid (DHA) Signaling	2.37	<i>ALOX15,PIK3R6</i>
Granulocyte Adhesion and Diapedesis	2.32	<i>IL1RL1,CCL23,HRH4</i>
Netrin Signaling	2.18	<i>CACNA1D,CACNG8</i>
Eicosanoid Signaling	2.15	<i>ALOX15,CYSLTR2</i>
CREB Signaling in Neurons	2.10	<i>CACNA1D,PIK3R6,CACNG8</i>
MSP-RON Signaling Pathway	2.09	<i>CSF1,PIK3R6</i>
LPS/IL-1 Mediated Inhibition of RXR Function	2.07	<i>IL1RL1,CAT,ALDH3B2</i>
Role of NFAT in Cardiac Hypertrophy	2.06	<i>CACNA1D,PIK3R6,CACNG8</i>
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	2.02	<i>IL1RL1,CSF1,PIK3R6</i>

Table S4. The most significant Cis-eQTL(1Mb, FDR < 0.01) pairs for each significant DEG (FDR < 0.05)

SNP	Chromosome	Position	Gene	Beta	P value	FDR
rs1045280	17	4622638	<i>ALOX15</i>	-0.40	2.82E-12	3.26E-09
rs4982729	14	23576611	<i>SLC7A8</i>	0.11	1.25E-08	6.87E-06
rs112983975	1	201086427	<i>SHISA4</i>	0.38	1.77E-06	5.64E-04
rs6670361	1	100200184	<i>FRRS1</i>	0.16	3.11E-06	9.11E-04
rs2756117	6	153066908	<i>SYNE1</i>	0.27	6.54E-06	1.65E-03
rs58546689	5	157565904	<i>ADAM19</i>	-0.28	6.73E-06	1.68E-03
rs2215044	19	14936194	<i>CYP4F12</i>	0.14	6.79E-06	1.70E-03
rs187984959	16	68940698	<i>SMPD3</i>	0.53	7.55E-06	1.87E-03
rs117153386	16	69634727	<i>IL34</i>	0.17	8.38E-06	2.04E-03
rs1243659	14	21081507	<i>RNASE3</i>	0.35	9.24E-06	2.21E-03
rs10152239	15	82448410	<i>ADAMTS7P1</i>	0.16	9.60E-06	2.27E-03
rs1390153	17	46969711	<i>ATP5G1</i>	0.38	1.02E-05	2.38E-03
rs17030975	3	53768205	<i>IL17RB</i>	-0.09	1.10E-05	2.54E-03
rs76162241	14	21042591	<i>RNASE2</i>	1.03	1.14E-05	2.62E-03
rs854200	3	45460460	<i>CCR3</i>	0.18	1.15E-05	2.64E-03
rs149916667	3	2609485	<i>IL5RA</i>	-0.75	1.26E-05	2.85E-03
rs9841726	3	53515794	<i>CACNA1D</i>	-0.09	1.48E-05	3.27E-03
rs59607495	2	102004515	<i>IL1RL1</i>	-0.33	1.72E-05	3.71E-03
rs1968359	19	40890938	<i>CLC</i>	-0.34	1.74E-05	3.73E-03
rs115897028	17	33584084	<i>CCL23</i>	-0.75	1.86E-05	3.96E-03
rs185768534	2	201845329	<i>CDK15</i>	0.37	1.93E-05	4.09E-03
rs7018994	9	77436641	<i>TRPM6</i>	-0.15	2.07E-05	4.36E-03
rs13333044	16	2223000	<i>PRSS33</i>	-0.29	2.74E-05	5.47E-03
rs10157580	1	112714950	<i>ADORA3</i>	0.14	3.19E-05	6.14E-03
rs10424222	19	51815704	<i>SIGLEC8</i>	-0.60	3.37E-05	6.43E-03
rs2025525	14	61903156	<i>LOC101927780</i>	-0.11	3.41E-05	6.50E-03
rs71394505	15	63597072	<i>DAPK2</i>	-0.21	3.49E-05	6.65E-03
rs56099328	21	34427671	<i>OLIG2</i>	0.16	4.16E-05	7.68E-03
rs74035835	16	88872631	<i>APRT</i>	-0.40	4.28E-05	7.85E-03
rs78272784	17	9281305	<i>PIK3R6</i>	0.20	4.40E-05	8.02E-03
rs9568025	13	48801876	<i>CYSLTR2</i>	0.15	4.57E-05	8.29E-03
rs192394913	17	19166383	<i>EPN2</i>	0.43	4.86E-05	8.65E-03
rs78488736	6	36660105	<i>RAB44</i>	0.18	4.93E-05	8.69E-03
rs4792578	17	15064170	<i>PMP22</i>	-0.13	5.70E-05	9.78E-03
rs2352178	10	87845206	<i>GRID1</i>	-0.02	5.72E-05	9.81E-03

Table S5. Associations between top differentially expressed genes and cell type proportions, by statistical significance at FDR-adjusted $P < 0.05$ (Y= Yes, N=No)

Gene	Atopy.DEG	Atopic Asthma.DEG	Neutrophils.	Lymphocytes	Monocytes	Eosinophils	Basophils
<i>HRH4</i>	Y	Y	N	N	N	Y	Y
<i>CCR3</i>	Y	Y	N	N	N	Y	N
<i>IL1RL1</i>	Y	Y	N	N	N	Y	N
<i>PEPD</i>	Y	N	Y	N	Y	N	N
<i>COQ9</i>	Y	N	N	N	N	N	N
<i>CLC</i>	Y	Y	N	N	N	Y	N
<i>EPN2</i>	Y	Y	N	N	N	Y	N
<i>CACNG8</i>	Y	Y	N	N	N	Y	N
<i>SPNS3</i>	Y	N	N	N	N	Y	N
<i>HRASLS5</i>	Y	Y	N	N	N	Y	N
<i>IL5RA</i>	Y	Y	N	N	N	Y	N
<i>SRGAP3</i>	Y	N	N	N	N	Y	N
<i>ALDH3B2</i>	Y	Y	N	N	N	Y	N
<i>PTGDR2</i>	Y	Y	N	N	N	Y	N
<i>PMP22</i>	Y	Y	N	N	N	Y	N
<i>DAPK2</i>	Y	N	Y	Y	N	Y	N
<i>ADGRE4P</i>	Y	Y	N	N	N	Y	N
<i>ATP5G1</i>	Y	N	N	N	N	N	N
<i>ADAM19</i>	Y	N	N	N	N	Y	N
<i>RAB44</i>	Y	Y	N	N	Y	Y	N
<i>SEMA7A</i>	Y	Y	N	N	N	Y	N
<i>NDUFAF3</i>	Y	N	N	N	N	N	N
<i>IL34</i>	Y	Y	N	N	N	Y	N
<i>OXER1</i>	Y	N	Y	N	N	Y	N
<i>ADAMTS7P1</i>	Y	Y	N	N	N	Y	N
<i>RHOXF1P1</i>	Y	Y	N	N	N	Y	N
<i>LGALS12</i>	Y	Y	N	N	N	Y	N

<i>ALAS2</i>	Y	N	N	N	N	N	N
<i>ADORA3</i>	Y	Y	N	N	N	Y	N
<i>PIK3R6</i>	Y	Y	N	N	N	Y	N
<i>PRSS33</i>	Y	Y	N	N	N	Y	N
<i>SMPD3</i>	Y	Y	N	N	N	Y	N
<i>CACNA1D</i>	Y	Y	N	N	N	Y	N
<i>CYP4F12</i>	Y	N	N	N	N	Y	N
<i>APRT</i>	Y	N	N	Y	N	N	N
<i>SHISA4</i>	Y	N	N	N	N	N	N
<i>CDK15</i>	Y	Y	N	N	N	Y	N
<i>SLC7A8</i>	Y	N	N	N	N	Y	N
<i>OLIG2</i>	Y	Y	N	N	N	Y	N
<i>PRSS41</i>	Y	Y	N	Y	N	Y	N
<i>CCL23</i>	Y	Y	N	N	N	Y	N
<i>SIGLEC8</i>	Y	Y	N	N	N	Y	N
<i>TMIGD3</i>	Y	Y	N	N	N	Y	N
<i>LOC101927780</i>	Y	Y	N	Y	N	Y	N
<i>TMEM160</i>	Y	N	N	Y	N	N	N
<i>RNASE3</i>	Y	N	N	N	N	Y	N
<i>GRID1</i>	Y	N	N	N	N	Y	N
<i>SYNE1</i>	Y	N	N	Y	N	Y	N
<i>RNASE2</i>	Y	Y	N	Y	N	Y	N
<i>ALOX15</i>	Y	Y	N	N	N	Y	N
<i>DACH1</i>	Y	Y	N	N	N	Y	N
<i>CEBPE</i>	Y	Y	N	N	N	Y	N
<i>TRPM6</i>	Y	N	Y	N	N	Y	N
<i>CYSLTR2</i>	Y	N	N	N	N	Y	N
<i>SLC29A1</i>	Y	Y	N	N	N	Y	N
<i>GPR82</i>	Y	N	N	N	N	Y	N
<i>SLC16A14</i>	Y	Y	N	N	N	Y	N
<i>TFF3</i>	N	Y	N	N	N	Y	N
<i>IL17RB</i>	Y	Y	N	N	N	Y	N
<i>FRRS1</i>	Y	Y	N	Y	N	Y	N
<i>CAT</i>	N	Y	N	N	N	Y	N

The first two columns denote whether the gene is among the DEGs in the meta-analysis of atopy or atopic asthma, at FDR-adjusted P value < 0.05. The other five columns denote whether expression of the gene is associated with cell type proportions (i.e. neutrophils, lymphocytes, monocytes, eosinophils, and basophils).

Table S6. Top DEGs by atopic asthma in white blood cells or whole blood samples vs. top DEGs by asthma in a prior study of airway epithelium

Gene	EVA-PR (white blood cells)			Meta-analysis (white blood cells and whole blood samples)			Previous transcriptomic analysis of asthma in airway epithelium*		
	Effect size	P value	FDR	Effect size	P value	FDR	Effect size	P value	FDR
<i>CLC</i>	1.18	2.09E-08	3.10E-05	1.19	7.52E-11	8.87E-07	4.60	4.15E-06	0.0005
<i>SIGLEC8</i>	1.21	5.94E-11	3.34E-07	1.13	1.94E-10	1.56E-06	4.61	4.07E-06	0.0005
<i>ALOX15</i>	1.17	2.97E-08	3.86E-05	1.17	6.58E-10	3.32E-06	4.63	3.59E-06	0.0004
<i>IL17RB</i>	1.19	6.41E-10	1.55E-06	1.06	4.17E-10	2.46E-06	3.40	0.0007	0.0219
<i>IL1RL1</i>	1.17	1.10E-07	1.03E-04	1.17	3.79E-09	1.22E-05	5.14	2.70E-07	6.00E-05
<i>PIK3R6</i>	1.12	2.25E-05	1.09E-02	1.09	1.80E-07	3.19E-04	3.82	0.0001	0.0066
<i>TFF3</i>	1.15	7.01E-08	7.39E-05	1.06	6.03E-06	6.68E-03	4.51	6.56E-06	0.0007

*Tsai Y-H et al. Meta-analysis of airway epithelia gene expression in asthma. Eur Respir J 2018; p. 1701962.

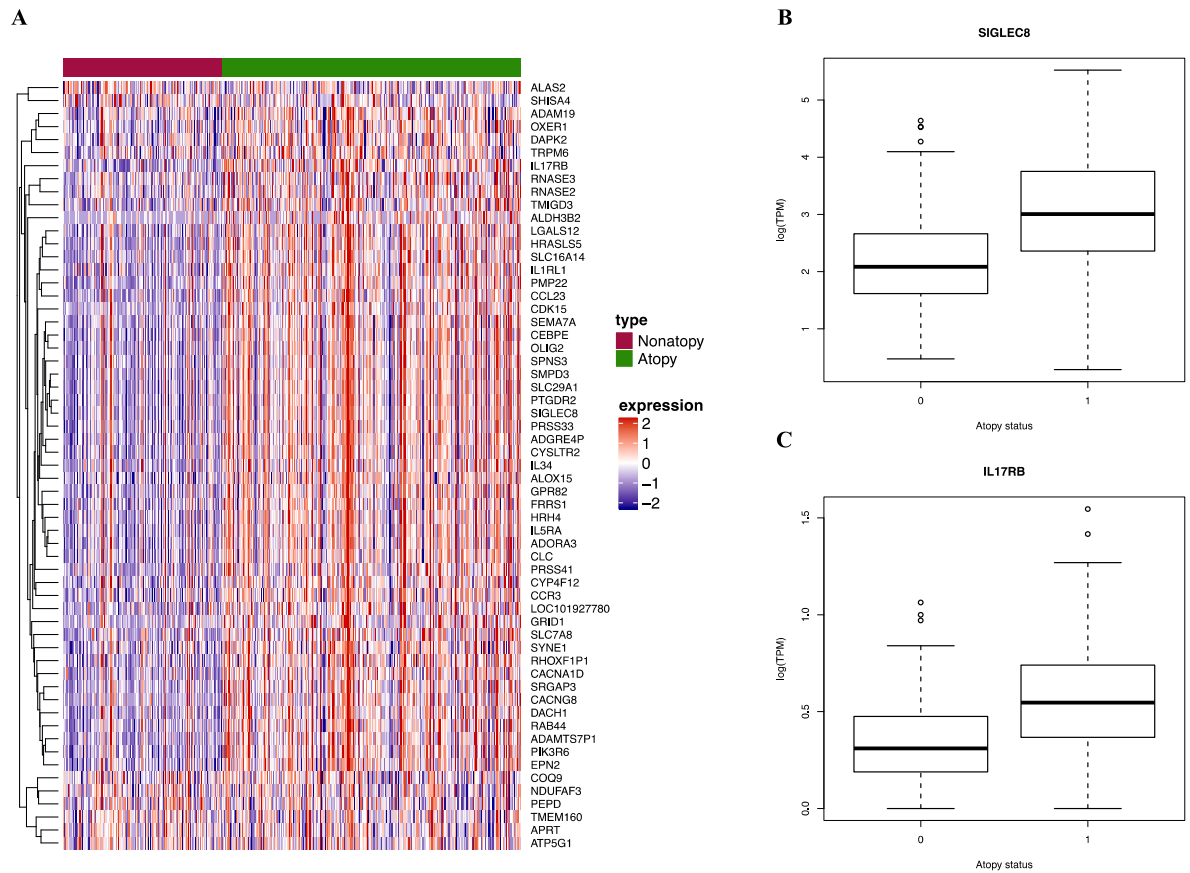
Table S7. Top differentially expressed genes in BAMSE*

Index **	Gene	Description	EVA-PR		BAMSE		Meta-analysis		
			Effect size	P-value	Effect size	P-value	Effect size	P-value	FDR
1	<i>IGFLR1</i>	IGF Like Family Receptor 1	0.99	6.00E-01	0.94	8.41E-06	0.96	2.18E-03	2.91E-01
2	<i>ATP5G1</i>	ATP Synthase Membrane Subunit C Locus 1	0.96	2.26E-02	0.96	6.39E-05	0.96	2.57E-05	1.98E-02
3	<i>RPP21</i>	Ribonuclease P/MRP Subunit P21	0.98	2.14E-01	0.96	2.10E-04	0.97	1.38E-03	2.41E-01
4	<i>MRPL24</i>	Mitochondrial Ribosomal Protein L24	0.98	1.59E-01	0.96	2.23E-04	0.97	8.82E-04	1.96E-01
5	<i>FRRS1</i>	Ferric Chelate Reductase 1	1.07	2.41E-03	1.07	2.41E-04	1.07	3.81E-06	3.86E-03
6	<i>KRTCAP2</i>	Keratinocyte Associated Protein 2	0.99	4.64E-01	0.97	2.99E-04	0.97	6.22E-03	3.85E-01
7	<i>APRT</i>	Adenine Phosphoribosyltransferase	0.96	1.70E-02	0.95	3.10E-04	0.96	4.87E-05	3.32E-02
8	<i>ATP6V1F</i>	ATPase H+ Transporting V1 Subunit F	0.98	6.16E-02	0.95	3.11E-04	0.96	2.68E-04	1.10E-01
9	<i>EIF3G</i>	Eukaryotic Translation Initiation Factor 3 Subunit G	0.98	1.33E-01	0.95	3.11E-04	0.96	8.06E-04	1.88E-01
10	<i>IGF2R</i>	Insulin Like Growth Factor 2 Receptor	1.02	1.84E-01	1.10	4.13E-04	1.05	1.55E-03	2.55E-01
11	<i>SF3A2</i>	Splicing Factor 3a Subunit 2	1.01	5.23E-01	0.96	4.32E-04	0.98	1.15E-01	8.00E-01
12	<i>MZT1</i>	Mitotic Spindle Organizing Protein 1	0.98	1.31E-01	0.95	4.93E-04	0.96	1.02E-03	2.04E-01
13	<i>ZNF747</i>	Zinc Finger Protein 747	0.94	6.92E-05	1.00	8.11E-01	0.97	9.83E-03	1.00E+00
14	<i>ACCS</i>	ACC Synthase-Like Protein 1	1.08	1.85E-04	0.99	7.26E-01	1.05	4.93E-02	1.00E+00

* Genes significantly associated with atopy or atopic asthma in BAMSE only, at an empirical cutoff as P-value < 5×10⁻⁴.

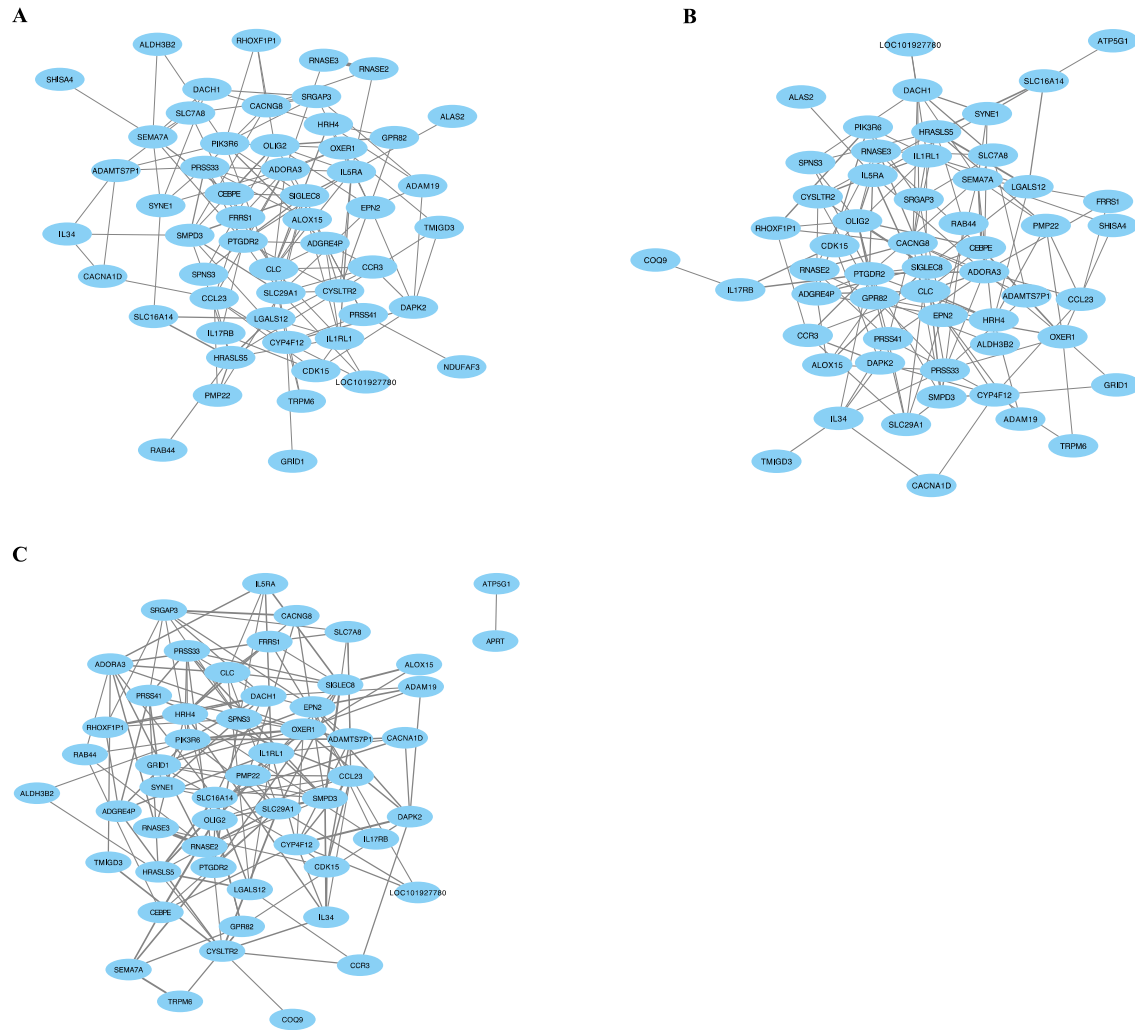
**Genes indexed from 1 to 12 are associated with atopy, and genes 13 and 14 are associated with atopic asthma in BAMSE.

Figure S1. Heatmap of differentially expressed genes (DEGs) and box plot of two top genes in the meta-analysis of atopy



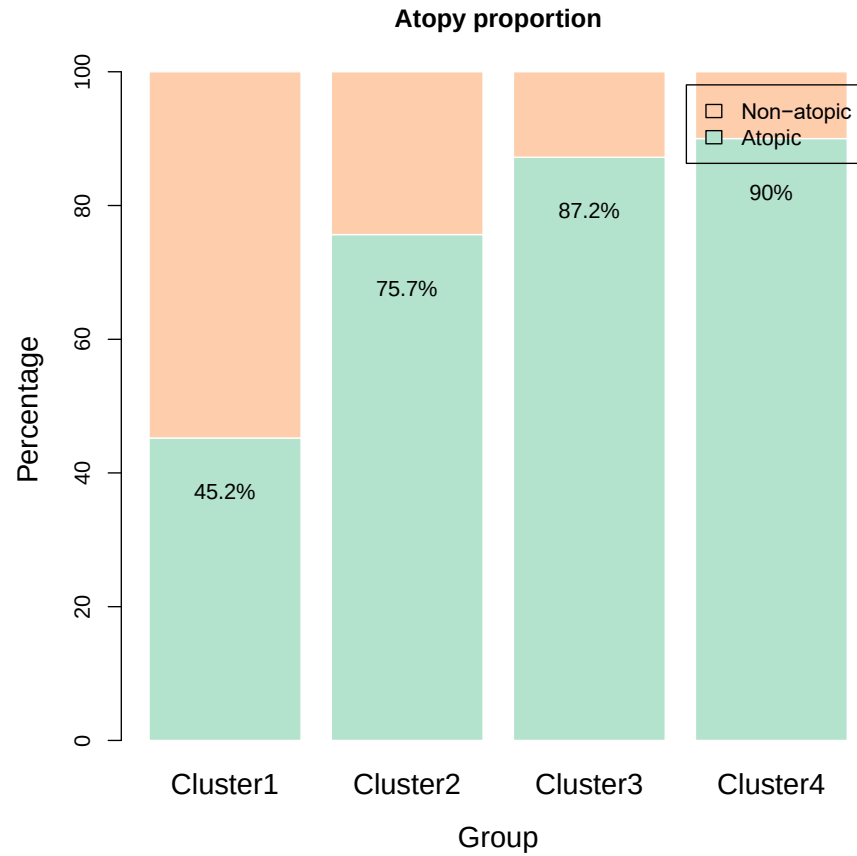
A. The heat map illustrates the expression level of the 59 DEGs (FDR < 0.05 in meta-analysis) between groups in EVA-PR. Each column is a sample, and each row is a gene. For any given gene, the red color represents gene expression that is greater than the overall mean, and the blue color represents gene expression that is less than the overall mean. Hierarchical clustering of genes is represented by the dendrograms on the left of the heat map. B. Expression of *SIGLEC8* and C. *IL17RB* in atopic and control subjects in EVA-PR.

Figure S2. Network analysis of significant 59 genes in EVA-PR.



Network relationship between the significant 59 genes (FDR < 0.05 in meta-analysis) in EVA-PR (P value < 0.01). A. Gene-gene interactions in all subjects. B. Gene-gene interactions in atopic subjects, and C. in non-atopic subjects. The width of edges between genes indicates the level partial correlations.

Figure S3. Bar plot of atopy proportion



Bar plot illustrates proportion of atopy subjects in each of the 4 clusters identified by gene expression in EVA-PR.