A Multiomics Approach to Identify Genes Associated with Childhood Asthma Risk and Morbidity

Erick Forno^{1*}, Ting Wang^{1*}, Qi Yan¹, John Brehm¹, Edna Acosta-Perez², Angel Colon-Semidey³, Maria Alvarez³, Nadia Boutaoui¹, Michelle M. Cloutier⁴, John F. Alcorn¹, Glorisa Canino², Wei Chen¹, and Juan C. Celedón¹

¹Division of Pediatric Pulmonary Medicine, Allergy, and Immunology, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, Pennsylvania; ²Behavioral Sciences Research Institute and ³Department of Pediatrics, University of Puerto Rico, San Juan, Puerto Rico; and ⁴Department of Pediatrics, University of Connecticut Health Center, Connecticut Children's Medical Center, Farmington, Connecticut

ORCID ID: 0000-0001-6497-9885 (E.F.).

Abstract

Childhood asthma is a complex disease. In this study, we aim to identify genes associated with childhood asthma through a multiomics “vertical” approach that integrates multiple analytical steps using linear and logistic regression models. In a case-control study of childhood asthma in Puerto Ricans ($n = 1,127$), we used adjusted linear or logistic regression models to evaluate associations between several analytical steps of omics data, including genome-wide (GW) genotype data, GW methylation, GW expression profiling, cytokine levels, asthma-intermediate phenotypes, and asthma status. At each point, only the top genes/single-nucleotide polymorphisms/probes/cytokines were carried forward for subsequent analysis. In step 1, asthma modified the gene expression-protein level association for 1,645 genes; pathway analysis showed an enrichment of these genes in the cytokine signaling system ($n = 269$ genes). In steps 2–3, expression levels of 40 genes were associated with intermediate phenotypes (asthma onset age, forced expiratory volume in 1 second, exacerbations, eosinophil counts, and skin test reactivity); of those, methylation of seven genes was also associated with asthma. Of these seven candidate genes, *IL5RA* was also significant in analytical steps 4–8. We then

measured plasma IL-5 receptor α levels, which were associated with asthma age of onset and moderate-severe exacerbations. In addition, *in silico* database analysis showed that several of our identified *IL5RA* single-nucleotide polymorphisms are associated with transcription factors related to asthma and atopy. This approach integrates several analytical steps and is able to identify biologically relevant asthma-related genes, such as *IL5RA*. It differs from other methods that rely on complex statistical models with various assumptions.

Keywords: asthma genetics; asthma genomics; childhood asthma; multiomics

Clinical Relevance

This study integrates several pathogenic steps and “omics” data to identify biologically relevant asthma-related genes, such as IL-5 receptor α . Unlike other methods that rely on complex statistical models with various assumptions, each step in our approach uses linear or logistic regression and is simple to interpret.

(Received in original form January 4, 2017; accepted in final form June 2, 2017)

*These authors share first authorship.

This work was supported by National Institutes of Health (NIH) grant HL125666 (E.F.) and Children's Hospital of Pittsburgh of UPMC grant (E.F.), NIH grant HG007358 (W.C.), NIH grants HL079966 and HL117191 (J.C.C.), and by an endowment from the Heinz Foundation.

Author Contributions: E.F., T.W., J.B., G.C., and J.C.C. contributed to study design; E.F., J.B., E.A.-P., A.C.-S., M.A., N.B., M.M.C., J.F.A., G.C., and J.C.C. contributed to data collection; E.F., T.W., Q.Y., W.C., and J.C.C. contributed to data analysis; all coauthors contributed intellectually to data interpretation and manuscript revision, and reviewed and approved the final version for submission.

Correspondence and requests for reprints should be addressed to Juan C. Celedón, M.D., Dr.P.H., Division of Pediatric Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh of UPMC, 4401 Penn Avenue, Pittsburgh, PA 15224. E-mail: juan.celedon@chp.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Cell Mol Biol Vol 57, Iss 4, pp 439–447, Oct 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1165/rcmb.2017-0002OC on June 2, 2017

Internet address: www.atsjournals.org

Asthma is the most common noncommunicable chronic disease of childhood (1) and affects over 6 million children in the United States, leading to over 3.5 million exacerbations, 135,000 hospitalizations, and approximately 14 million missed school days each year (2). It is a multifactorial disease with heritability as high as 80–90% (3, 4). To date, several dozen asthma-related genes have been identified (5), yet single-nucleotide polymorphisms (SNPs) in such genes account for less than half of the disease's heritability.

Our ability to identify novel genetic polymorphisms associated with disease has increased exponentially with the use of microarrays and now high-throughput technologies, such as sequencing. However, this steep increase in the number of markers analyzed has also introduced new problems in terms of statistical analysis and biological interpretation of novel results (6). Strategies used to deal with these issues suffer from their own problems: mathematical corrections for multiple comparisons can be excessively conservative and may discard associations that could be biologically important; different methods rely on various assumptions that are not necessarily generalizable to other circumstances, et cetera (7). This is particularly important for analysis of multiomics data (7). Conversely, attempts at replicating findings in independent cohorts presume that the populations are similar enough or that the specific polymorphisms are universally relevant.

The pathogenesis of complex diseases, such as asthma, involves several steps: genetic variants are subject to epigenetic regulation, which helps determine the level of gene expression; expressed genes are, in turn, translated into proteins, which may have direct effects on disease “intermediate phenotypes”; and these intermediate processes ultimately alter an individual's risk of disease or disease severity. In this study, we propose a novel approach that aims to identify genes and pathways that are relevant to asthma by integrating several analytical steps that are based on the conceptual framework of disease pathogenesis. Furthermore, we aimed to do so through a “vertical” approach that uses existing data in the same population to sequentially narrow our analytical focus, thereby maximizing efficiency, minimizing multiple comparisons, and avoiding complex mathematical models built on numerous assumptions.

Materials and Methods

Study Population

As part of a case–control study of childhood asthma in Puerto Ricans (Hartford and Puerto Rico; HPR), we recruited 1,127 children (618 with asthma [cases] and 509 without asthma [control subjects]). The details of subject recruitment have been published previously (8–10). In brief, from September 2003 to July 2008, 449 Puerto Rican children were recruited from elementary/middle schools in Hartford, Connecticut, and from March 2009 to June 2010, 678 Puerto Rican children living in San Juan, Puerto Rico, were recruited using a multistage probability sample design. A household was eligible if one or more resident was a child 6–14 years old. Participants had to have four Puerto Rican grandparents and had to be living in the same household for 1 year or longer. In households with more than one eligible child, one child was randomly selected for screening. There were no significant differences in age, sex, or area of residence between eligible children who did and did not participate at either study site. We selected, as cases, children with physician-diagnosed asthma and at least one episode of wheeze in the prior year, and, as control subjects, children with no physician-diagnosed asthma and no wheeze in the prior year. Written parental consent was obtained for participating children, from whom written assent was also obtained. The study was approved by the Institutional Review Boards of Connecticut Children's Medical Center (Hartford, CT), the University of Puerto Rico (San Juan, PR), Brigham and Women's Hospital (Boston, MA), and the University of Pittsburgh (Pittsburgh, PA).

Study Procedures

Study participants completed a protocol that included questionnaires, spirometry, blood sample collection, and skin test reactivity (STR), as previously described (11). In brief, spirometry was conducted according to American Thoracic Society criteria modified for children (12); spirometry was repeated 15 minutes after administration of 200 µg of albuterol metered-dose inhaler. STR included histamine (positive control), saline solution (negative control), and allergen extracts from house dust mite mix (*Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*), German

cockroach (*Blattella germanica*), cat dander, dog dander, mixed grass pollen, mugwort sage, ragweed, mixed tree pollen, mold mix, *Alternaria tenuis*, and mouse extracts. Peripheral blood eosinophil count was measured by Coulter counter techniques and then log₁₀ transformed for analysis. In addition to asthma status, five intermediate phenotypes were selected *a priori* for analysis: age of asthma onset; moderate–severe asthma exacerbations (defined by EPR-3 criteria [13]); post-bronchodilator forced expiratory volume in 1 second (FEV₁); eosinophil counts; and the total number of STR⁺.

Cytokine Plasma Levels

For a previously published analysis (10), 14 Th17 pathway-related cytokines (IL-1b, IL-4, IL-6, IL-10, IL-17A, IL-17F, IL-21, IL-22, IL-23, IL-25, IL-31, IL-33, IFN-γ, and TNF-α) were measured in plasma using the Bio-Plex Pro Human Th17 cytokine panel (Bio-Rad Laboratories, Hercules, CA). Assays were designed on magnetic beads in a capture-sandwich immunoassay format. Undetectable cytokine levels were assigned a constant (half the lowest limit of detection). Cytokine data were available for 578 participants. All cytokine levels were log₁₀ transformed for data analysis.

Genome-Wide Genotyping

Genome-wide (GW) genotyping (GWG) was performed using the Illumina HumanOmni2.5 BeadChip platform (Illumina, San Diego, CA), which measures approximately 2.5 million SNPs, as previously described (14, 15). Subjects with a call rate less than 95% were removed from the analysis. SNPs were removed if they were not in Hardy–Weinberg equilibrium ($P < 10^{-6}$) in control subjects, had minor allele frequency lower than 1%, or had a failure rate greater than 5%. After all subject and marker quality control steps, approximately 1.83 million SNPs for 951 HPR participants were included in the analysis.

GW White Blood Cell DNA Methylation

High-quality white blood cell (WBC) DNA (750 ng) was available from 340 HPR participants. After bisulfite conversion using EZ-DNA methylation kits (Zymo Research, Irvine, CA), DNA (200 ng) was used for GW WBC DNA methylation (GWM) analysis using the HumanMethylation 450K BeadChip (Illumina), as previously described (16). Methylation data were read

using the *methylumi* R package (version 3.0.1; <https://www.r-project.org/about.html>), and β values for each CpG were calculated as: $\beta = M / (M + U + \alpha)$, where M and U represent methylated and unmethylated signal intensities at the specific site, and α is an arbitrary offset (usually 100) intended to stabilize β values where fluorescent intensities are low (16). Raw methylation data were normalized with the SWAN (subset-quantile within array normalization) method using the *minfi* R package (16, 17).

GW Whole-Blood Gene Expression Profiling

RNA was available from 141 HPR participants. After extraction from whole blood using PAXgene blood miRNA kits (Qiagen, Valencia, CA), globin transcripts were then depleted using GLOBINclear kit (Life Technologies, Carlsbad, CA). RNA quality and concentration were determined using Agilent RNA 6,000 Nano kit (Agilent Technologies, Santa Clara, CA). GW whole-blood gene expression (GWE) profiling was measured in 141 whole-blood globin-cleaned RNA samples using the HumanHT-12 v4 Expression BeadChip (Illumina). Background subtraction and quantile normalization were performed using the *lumi* R package, as previously described (15). Probes with greater than 70% absent points among all 141 samples were excluded from the downstream analysis. A total of approximately 15,000 probes was included in the final analysis. Please refer to Figure E1 in the online supplement for overlap between GWG, GWM, and GWE availability for the cohort.

Principal Component Analysis

Principal component (PC) analysis was performed using genotyping data, methylation data, and expression data separately. PCs can capture different sources of variation due to global ancestry, population stratification, unmeasured environmental factors, or technical factors. The first two PCs derived from GWG data, the first 10 PCs from GWE data, and the first two PCs from GWM data were included as covariates in the respective analytic steps.

Analytical Approach

Please see the online supplement for further details. We performed our analysis in several steps, working under a “vertical” approach that follows the conceptual

framework of pathogenic pathway from genotype to disease (Figure 1)—including genetics (GWG), epigenetics (GWM), genomics (GWE), proteomics (cytokine levels), phenomics (asthma-related intermediate phenotypes), and disease status (asthma versus control):

1. Modification of gene expression versus protein level by asthma status: our initial hypothesis was that gene expression and protein level would be most closely associated; furthermore, we hypothesized that genes and pathways implicated in asthma risk or morbidity would be differentially activated between cases and control subjects. Therefore, in this first step, we evaluated the association between GWE and cytokine levels by asthma status using the general model: $P_i \sim A + E_j + A \cdot E_j + \text{covariates}$, where P_i is the plasma level of cytokine, i , A is the participant's asthma status, E_j is the expression level of gene, j , and $A \cdot E_j$ is the interaction term. Next, we evaluated genes the association of which with protein level differed by asthma status (interaction $P < 0.01$) in a gene-set enrichment analysis using Fisher's exact test (Reactome and KEGG databases), and selected the top enriched pathway(s) as candidate(s) for subsequent steps.
2. Gene expression versus intermediate phenotypes: we assessed the expression levels of the genes carried forward from step 1 versus five asthma-related intermediate phenotypes (age of asthma onset, asthma moderate/severe exacerbations, baseline FEV₁, peripheral blood eosinophil counts, and positive allergy skin testing) using linear or logistic regression as appropriate. From this analysis, we selected the genes with the 20 lowest P values plus those with P less than 0.01 in at least three of five intermediate phenotypes. In addition, we included one gene (*IL18*) that was nominally significant in step 1, but nonsignificant in step 2 as a “negative control,” hypothesizing that it would not be consistently significant in subsequent analytical steps.
3. Methylation versus asthma status: next, we evaluated the association between DNA methylation levels at the CpG sites of the genes carried forward from step 2 and asthma status using regression modeling. We carried forward genes with false discovery rate-adjusted P less than 0.05.

4. Genotype versus asthma status: we then evaluated the association between all SNPs (from GWG) in the top genes from step 3 and asthma status using regression modeling.
5. Expression quantitative trait loci (eQTL) analysis: simultaneously, we evaluated SNPs that were associated with the expression levels of the top genes from step 3, based on result from our recent eQTL analysis (15).
6. Genotype versus protein level: we included all SNPs that were significant in steps 4 or 5 and evaluated their association with the plasma levels of the corresponding cytokines from step 1.
7. Genotype versus intermediate phenotypes: we included all SNPs that were nominally significant in step 6 and evaluated the association between the SNPs and the corresponding intermediate phenotypes from step 2.
8. Protein level versus intermediate phenotypes: in the final step, we analyzed the association between the protein levels that showed nominally significant associations in step 1 and the intermediate phenotypes from steps 2 and 7.

IL-5 receptor α and in silico transcription factor analysis. We measured IL-5 receptor α (IL-5R α) plasma levels (the protein for which gene *IL5RA* codes) in a subset of participants ($n = 130$) and evaluated the association between IL-5R α levels and asthma phenotypes. In addition, we assessed the effects of *IL5RA* SNPs on IL-5R α levels and their modification by asthma status. All models were adjusted for age, sex, parental asthma, and household income. We also evaluated the regulatory potential of *IL5RA* SNPs using is-rSNP (18), software designed to predict whether an SNP has a regulatory effect on transcription factor binding sites *in silico*.

Results

Modification of Gene Expression versus Protein Level by Asthma Status

Characteristics of study participants are shown in Table 1. Asthma modified the gene–protein association in a total of 1,645 genes (interaction $P < 0.01$); of the 14 cytokines tested, the main associations were with IL-10, IL-17A, IL-17F, IL-22, and IL-23. Results from gene-set enrichment analysis for

Figure 1. Pathogenic and analytical steps. Pathogenic steps from genotype to disease (left) and analytical steps (right) performed in the current study. Shown are the numbers of genes, probes, or single-nucleotide polymorphisms (SNPs) included in each sequential step, as well as the generic analytical model used. All models were adjusted for pertinent covariates (not shown in the figure for simplicity).

those 1,645 genes showed that the Reactome cytokine signaling system ($n = 269$ genes; see Figure 2) was significantly enriched for those genes and cytokines (see Table E1). Thus, these 269 genes were carried forward.

Gene Expression versus Intermediate Phenotypes

Age of onset showed nominally significant associations with $n = 60$ of 269 genes (22%); exacerbations with $n = 42$ genes (16%); FEV₁ with $n = 40$ (15%), eosinophils with $n = 32$

(12%); and SPT⁺ with $n = 66$ (25%)—all significantly more than expected by chance. From this step, we selected 40 genes that were among the top P values and those consistently associated with more than two intermediate phenotypes. In addition, *IL18* was included as a negative control. Table E2 shows the detailed results from this step.

Methylation versus Asthma Status

A total of 7 of 41 genes (17%) had DNA methylation levels that were associated with

asthma status: Conserved helix-loop-helix ubiquitous kinase (*CHUK*), IL-1 receptor-2 (*IL1R2*), IL-1 receptor antagonist (*IL1RN*), IL-5Rα (*IL5RA*), IFN regulatory factor 7 (*IRF7*), tyrosine-protein kinase Lyn (*LYN*), and suppressor of cytokine signaling 2 (*SOCS2*). *IL5RA* had the most significant P value and the most differentially methylated CpG sites. Table 2 shows the summary findings for these seven genes (for all analytical steps).

SNP versus eQTL

We evaluated all SNPs associated with expression of the seven selected genes. Table 2 shows the false discovery rate-adjusted P values as well as the corresponding SNPs. SNP rs17878498 was associated with expression of *IL5RA*; it is located approximately 870 bases upstream from *IL5RA* in chromosome 3p26.6, and is part of the binding site motif for transcription factor *SPI1*, which is also involved in IL-4 signaling and glucocorticoid receptor regulation.

SNP versus Asthma Status

Table 2 shows the top P value associated with asthma status for the seven selected genes, as well as the total number of SNPs with P less than 0.05 for each gene. Three genes (*IL5RA*, *LYN*, and *SOCS2*) had the

Table 1. Characteristics of Study Participants

	Children with GWG	Children with GWM	Children with GWE
No. of children	948	309	138
Age, yr	10.1 (2.7)	8.6 (1.9)*	10.4 (2.7)
Male sex, %	51.6	57.3	56.5
FEV ₁ , L	1.95 (0.71)	1.60 (0.51)	1.98 (0.71)
Eosinophils, cells/mm ³	239 (133–426)	389 (202–599)*	340 (181–554)*
STR ⁺ , %	60.9	74.1	76.9
Asthma, %	54.9	56.6	50.7
Age of onset, yr [†]	2.4 (2.5)	2.2 (2.0)	2.6 (2.4)
Exacerbation, % [†]	42.2	58.9	55.7

Definition of abbreviations: GWE, genome-wide expression; GWG, genome-wide genotypic; GWM, genome-wide methylation; STR⁺, positive skin test reactivity.

Shown are mean (SD) or median (interquartile range) for continuous variables, and percent for dichotomous variables.

* $P < 0.05$ compared to data from children with GWG, but without GWM or GWE.

[†]Among children with asthma only.

Figure 2. Venn diagram showing number of genes in top pathways from gene-set enrichment analysis. Reactome pathway “cytokine signaling in the immune system” ($n = 269$ genes) and the contained pathways were highly enriched for genes significant in analytical step 1 (modification of gene expression versus cytokine level by asthma status). The next encompassing pathway (“immune system”) was deemed too broad for further analysis.

top P values and/or the most SNP associations (17, 25, and 56, respectively).

SNP versus Cytokine Level

We included 134 SNPs in the 7 selected genes. Three genes (*IL5RA*, *IRF7*, and *LYN*) had SNPs that were significantly associated with the cytokine with which there was differential expression–protein association by asthma status (step 1). *IL5RA* SNP rs6808454 was associated with IL-17F plasma levels ($P = 0.024$), *IRF7* SNP rs143197172 with IL-17A ($P = 0.022$), and five SNPs in high linkage disequilibrium in *LYN* SNP with IL-17A ($P = 0.025$). In addition, each gene had other SNPs with P less than 0.05 (five total SNPs in *IL5RA*, two in *IRF7*, and five in *LYN*).

SNP versus Intermediate Phenotype

Among SNPs that were significant in step 6, *IL5RA* SNP rs6773701 was associated with eosinophil counts ($P = 0.05$), *IRF7* SNP rs142972371 with exacerbations ($P = 0.016$), and *LYN* SNP rs112735328 was associated with the total number of STR⁺ ($P = 0.009$).

Cytokine Level versus Intermediate Phenotype

Finally, in step 8, IL-17F (associated with *IL5RA*) was significantly associated with

higher eosinophil counts ($\beta = 0.30$; 95% confidence interval [CI] = 0.06–0.54; $P = 0.013$), and also showed borderline association with lower FEV₁ ($\beta = -248$ ml; 95% CI = -511 ml to $+16$ ml; $P = 0.067$).

Of the seven selected genes, *IL5RA* was significant in all eight analytical steps. *IRF7* was significant in six steps, and the other five genes were significant in four to five steps. Contrarily, *IL18*, which was selected as a “negative control” because it did not pass step 2, was only significant in three out of eight steps.

IL-5R α plasma levels. We measured IL-5R α plasma levels in a subset of participants: median IL-5R α level was 0.08 pg/ml (minimum, 0.058 pg/ml; maximum, 3.61 pg/ml). In our adjusted analysis, children who had moderate–severe asthma exacerbations had lower IL-5R α levels (-1.98 pg/ml [95% CI = -3.81 to -1.03]; $P = 0.044$) and children with earlier age of onset had higher IL-5R α ($+1.17$ pg/ml [95% CI = 1.01–1.36]; $P = 0.033$). In addition, the effects of five *IL5RA* SNPs on IL-5R α levels were modified by asthma status (Figure E1). Table 3 shows the summary of all SNPs in *IL5RA* that showed nominally significant associations in our stepwise analysis.

In silico transcription factor analysis. Finally, we evaluated the regulatory

potential of *IL5RA* SNPs using is-rSNP (18). Table E3 shows all results with adjusted P less than 0.05. Among the top transcription factors associated with these *IL5RA* SNPs, several have been associated with asthma or atopy, or other lung processes and diseases, including hypoxia-inducible factor 1 α and 1 β (19–21), serum response factor (22, 23), upstream transcription factor 1 (24), CCAAT-enhancer binding protein β (25), peroxisome proliferator-activated receptor α (26), signal transducer and activator of transcription 3 (27), Runt-related transcription factor 1 (28), *et cetera*.

Alternative/confirmatory analysis. To evaluate whether our results were solely dependent on the order of the steps, we performed a secondary, confirmatory analysis, designed so it would be “order agnostic.” We looked at the overlap in results from analyzing SNP data for the genes significant in step 2 versus four outcomes: (1) asthma; (2) intermediate phenotypes; (3) gene expression (eQTLs); and (4) cytokine levels. We additionally adjusted this analysis by WBC differential (percent of neutrophils, lymphocytes, monocytes, eosinophils, basophils). Rather than reducing the number of SNPs at each step, we looked at the overlap in results for the four outcomes. There were 148 SNPs from 40 genes (Figure 3 and Table E4) associated with at least three of these four outcomes. There was only one SNP, rs71058675 in *IL5RA*, associated with all four outcomes. There was also only one eQTL, rs1153462 (also in *IL5RA*), associated with asthma and intermediate phenotypes. Thus, this analysis confirms that our results are not entirely dependent on the selected order of the analytical steps. Of the other top genes, *SOC2* was associated with only two of the four outcomes, and *IL1RN* with one.

Discussion

In the present study, we demonstrate a multiomics approach able to identify *IL5RA* as an asthma-related gene. To our knowledge, this is the first study of childhood asthma that integrates several analytical stages—based on the concept of pathogenic steps—to assess the significance of a gene or group of genes.

Table 2. Stepwise Analytical Approach Summary for Seven Final Genes

Step	<i>CHUK</i>	<i>IL1R2</i>	<i>IL1RN</i>	<i>IL5RA</i>	<i>IRF7</i>	<i>LYN</i>	<i>SOCS2</i>	<i>IL18</i>	Comments
1	0.03	0.007	0.01	0.007	0.038	0.093	0.10	0.04	Top interaction <i>P</i> value
2	IL-23	IL-17A	IL-17A	IL-17F	IL-17A	IL-17A	IL-23	IL-23	Associated cytokine(s)
	n/s	0.002	0.007	n/s	0.009	0.027	0.022	n/s	<i>P</i> value for exacerbations
	0.041	0.048	0.006	n/s	n/s	0.009	0.005	n/s	Age of onset of asthma
	0.042	n/s	n/s	n/s	n/s	n/s	0.023	n/s	FEV ₁
	0.031	0.03	n/s	0.046	0.042	0.015	0.011	n/s	STR ⁺
	n/s	n/s	0.014	6.3 × 10⁻¹²	n/s	n/s	0.042	n/s	Eosinophils
3	0.042	0.039	0.031	0.012	0.049	0.042	0.042	0.02	Top FDR <i>P</i> value
	1	1	1	4	1	1	1	2	Number of FDR <i>P</i> < 0.05 sites
4	n/s	n/s	n/s	0.019	n/s	n/s	n/s	n/s	Top FDR <i>P</i> value
	—	—	—	rs17878498	—	—	—	—	Associated SNP(s)
5	0.024	0.052	0.026	0.0055	0.015	0.014	0.0016	0.009	Top FDR <i>P</i> value
	4	0	5	17	4	25	56	10	Number of FDR <i>P</i> < 0.05 SNPs
6	n/s	n/s	n/s	0.024	0.022	0.025	n/s	n/s	Top <i>P</i> value
	—	—	—	—	—	6	—	—	Number of <i>P</i> < 0.05 SNPs
7	n/s	n/s	n/s	0.0507	0.01	0.049	0.005	n/s	Top <i>P</i> value
	—	—	—	Eosinophils	Exacerbations	Total STR ⁺	Onset age	—	Associated phenotype(s)
8	n/s	n/s	n/s	0.013	n/s	n/s	n/s	n/s	Top <i>P</i> value
	—	—	—	Eosinophils	—	—	—	—	Associated phenotype(s)

Definition of abbreviations: FDR, false discovery rate; n/s, not statistically significant; SNP, single-nucleotide polymorphism; STR⁺, positive skin test reactivity. Final candidate genes: conserved helix-loop-helix ubiquitous kinase (*CHUK*); IFN regulatory factor 7 (*IRF7*); IL-1 receptor-2 (*IL1R2*); IL-1 receptor antagonist (*IL1RN*); IL-5 receptor α (*IL5RA*); tyrosine-protein kinase Lyn (*LYN*); and suppressor of cytokine signaling 2 (*SOCS2*). Bold typeface denotes *P* < 0.05.

In our analysis, *IL5RA* showed significant associations in all analytical steps. *IL5RA* SNPs were associated with eosinophil counts, IL-17F (which in turn is associated with asthma and intermediate phenotypes), and asthma, as well as with *IL5RA* expression and IL-5R α plasma levels. *IL5RA* methylation was associated with asthma. *IL5RA* expression was associated with eosinophils, STR, and its association with IL-17F was modified by asthma. In addition, IL-5R α levels were

Table 3. *IL5RA* Single-Nucleotide Polymorphisms with Significant Associations in the Stepwise Analysis

Step 4 Gene Expression ~ SNP		Step 5 Asthma Status ~ SNP		Step 6 Cytokine (IL17F) ~ SNP		Step 7 Phenotypes ~ SNP		IL5-R α Level ~ SNP · Asthma	
SNP	<i>P</i> Value	SNP	<i>P</i> Value	SNP	<i>P</i> Value	SNP	<i>P</i> Value	SNP	<i>P</i> Value
rs17878498	0.018	rs3804790	0.005	rs6808454	0.024	rs6773701	0.0507	rs340832	0.043
		rs11294168	0.007	rs6773701	0.034			rs340833	0.034 [†]
		rs1153461	0.016	rs71058675	0.045			rs340807	0.10
		rs35439082	0.017	rs2600027	0.046			rs340809	0.009 [†]
		rs3856847	0.025	rs3804790	0.049			rs3792424	0.18
		rs1153462	0.027						0.059 [†]
		rs7644398	0.029						0.051
		rs71058675	0.029						0.010 [†]
		rs35418165	0.031						0.073
		rs62230291	0.035						0.029 [†]
		rs17887062	0.036						
		rs17879465	0.041						
		rs6808454	0.042						
		rs142128948	0.043						
		rs150254181	0.045						
		rs2600027	0.049						
		rs6773701	0.049						

Definition of abbreviation: SNP, single-nucleotide polymorphism.

All SNPs sorted by *P* value. SNPs significant in more than one step are shown bold typeface.

[†]Interaction term (IL-5 receptor α ~ SNP + SNP · asthma + covariates).

Figure 3. Results from confirmatory analysis. Shown is the number of SNPs associated with each of four outcomes: asthma; intermediate phenotypes; cytokine levels; and gene expression. All results are from adjusted regression models adjusted for relevant covariates and principal components. Gene expression analysis additionally adjusted for white blood cell differential cell types. A total of 148 SNPs from 40 genes (see Table E4) were associated with at least three of four outcomes. Only one SNP, rs71058675, in *IL-5 receptor α* (*IL5RA*), was associated with all four outcomes. There was also only one expression quantitative trait loci, rs1153462 (also in *IL5RA*), associated with asthma and intermediate phenotypes. CellAdj, adjusted for CBC cell differential; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; MedPhn, asthma intermediate phenotype.

associated with asthma age of onset and moderate-severe exacerbations. Of note, save for the association between *IL5RA* expression and eosinophils, most of these associations would have been missed in unidimensional analyses that customarily use statistical methods to adjust for multiple markers (e.g., for *IL5RA* SNPs versus asthma, the top SNP [rs17878498] had a value of $P=0.019$). Although these steps are certainly not independent from each other, we propose that this “vertical validation” method, which evaluates significance at multiple pathogenic levels, is preferable to identify genes that are of greater biological plausibility and relevance. Furthermore, we observed associations between specific SNPs and several transcription factors, which show that these SNPs may have a direct regulatory role within known IL-5R α signaling pathways.

The IL-5R, expressed on eosinophils, mast cells, and basophils, is a heterodimer composed of subunits IL-5R α (specific to the IL-5R) and IL-5R β (common to the IL-3, IL-5, and granulocyte-macrophage colony-stimulating factor receptors) (29). We previously reported that an SNP adjacent to *IL5RA* is associated to the age of onset of asthma (30). The soluble form of IL-5R α can occur by cleavage of existing

cell surface receptors or by alternate splicing of *IL5RA* mRNA (31), and is a high-affinity competitive antagonist of IL-5 that prevents IL-5 from binding to surface IL-5R, and thus decreases activation and survival of eosinophils *in vitro* (32). Clinical studies have shown changes in soluble IL-5R α levels in association with eczema in children (33), chronic obstructive pulmonary disease (34), and nasal polyposis (35). The balance between surface and soluble IL-5R α has been shown to play a part in eosinophil regulatory pathways (31), and it could have implications for the use of anti-IL-5 therapeutic agents in asthma (36, 37). *IL5RA* was among the top results in a recent epigenome-wide association study of immune IgE (38). Anti-IL-5R α monoclonal antibody, benralizumab, has been recently reported to reduce exacerbations and improve lung function in subjects with severe uncontrolled asthma (39, 40).

Our study has several limitations. Although we used the concept of disease pathogenesis to design our analytical approach, it is critical to underscore that our cross-sectional analysis evaluates associations and cannot adjudicate causality. GWE and GWM were not available for all study participants, and thus we had a

relatively small sample size for some of the analyses, particularly those involving GWE, and thus limited statistical power to detect modest effects. We subjectively decided to use the gene expression-protein level as our first analytical step, hypothesizing that those two would be most closely related. However, it is important to clarify that we are not proposing that this is the only way or specific order in which these analyses must be undertaken; we simply propose that a “vertical validation” (e.g., significance at multiple pathogenic levels using multiomics integration) may render more relevant results than “horizontal replication” (e.g., attempting to replicate results from one analysis in a separate sample). This vertical validation approach does not expand the spectrum of findings across platforms, but rather uses these platforms to sequentially validate our findings. Pathway analysis may have biased our results toward better-characterized genes. We did not correct for multiple comparisons at all analytical steps; because the steps are not fully independent from each other, multiple testing correction at each level would have likely been overstringent. Future studies may be necessary to explore the optimal inferential approach to account for this problem. Finally, our analysis might have yielded different results had we proceeded in a different analytical order. However, we performed a secondary analysis that was agnostic to the order of the analytical steps, and *IL5RA* remained the top result. Furthermore, we propose that those results from analyzing the data in yet another order would likely be biologically relevant as well.

In summary, we present a multiomics approach that integrates several pathogenic steps and is able to identify biologically relevant asthma-related genes, such as *IL5RA*. Our approach differs from other methods for integrative analysis that rely on complicated statistical models with various assumptions; each step in our method uses linear or logistic regression, and is simple to interpret. Further studies will need to evaluate and validate this approach in diverse settings, including other complex diseases with multiomics data. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- World Health Organization. Asthma. Geneva: WHO; Fact Sheet No 307. 2014 [updated 2017 Apr; accessed 2017 Jul 5]. Available from: <http://www.who.int/mediacentre/factsheets/fs307/en/>
- Centers for Disease Control and Prevention. Asthma data, statistics, and surveillance: most recent asthma data. 2015 [updated 2017 Feb; accessed 2017 Jul 5]. Available from: http://www.cdc.gov/asthma/most_recent_data.htm
- Palmer LJ, Burton PR, James AL, Musk AW, Cookson WO. Familial aggregation and heritability of asthma-associated quantitative traits in a population-based sample of nuclear families. *Eur J Hum Genet* 2000;8:853–860.
- Thomsen SF, van der Sluis S, Kyvik KO, Skytthe A, Skadhauge LR, Backer V. Increase in the heritability of asthma from 1994 to 2003 among adolescent twins. *Respir Med* 2011;105:1147–1152.
- Ortiz RA, Barnes KC. Genetics of allergic diseases. *Immunol Allergy Clin North Am* 2015;35:19–44.
- Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, Milanese L. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics* 2016;17:15.
- Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat Rev Genet* 2015;16:85–97.
- Brehm JM, Ramratnam SK, Tse SM, Croteau-Chonka DC, Pino-Yanes M, Rosas-Salazar C, Litonjua AA, Raby BA, Boutaoui N, Han YY, et al. Stress and bronchodilator response in children with asthma. *Am J Respir Crit Care Med* 2015;192:47–56.
- Jacobs TS, Forno E, Brehm JM, Acosta-Pérez E, Han YY, Blatter J, Colón-Semidey A, Alvarez M, Canino G, Celedón JC. Underdiagnosis of allergic rhinitis in underserved children. *J Allergy Clin Immunol* 2014;134:737–739.e6.
- Han YY, Forno E, Brehm JM, Acosta-Pérez E, Alvarez M, Colón-Semidey A, Rivera-Soto W, Campos H, Litonjua AA, Alcorn JF, et al. Diet, interleukin-17, and childhood asthma in Puerto Ricans. *Ann Allergy Asthma Immunol* 2015;115:288–293.e1.
- Forno E, Acosta-Pérez E, Brehm JM, Han YY, Alvarez M, Colón-Semidey A, Canino G, Celedón JC. Obesity and adiposity indicators, asthma, and atopy in Puerto Rican children. *J Allergy Clin Immunol* 2014;133:1308–1314, 1314.e1–5.
- American Thoracic Society. 1994 update. *Am J Respir Crit Care Med* 1995;152:1107–1136.
- National Asthma Education and Prevention Program. Expert panel report 3 (EPR-3): guidelines for the diagnosis and management of asthma—summary report 2007. *J Allergy Clin Immunol* 2007;120(5 suppl):S94–S138.
- Brehm JM, Acosta-Pérez E, Klei L, Roeder K, Barnada MM, Boutaoui N, Forno E, Cloutier MM, Datta S, Kelly R, et al. African ancestry and lung function in Puerto Rican children. *J Allergy Clin Immunol* 2012;129:1484–1490.e6.
- Chen W, Brehm JM, Lin J, Wang T, Forno E, Acosta-Pérez E, Boutaoui N, Canino G, Celedón JC. Expression quantitative trait loci (eQTL) mapping in Puerto Rican children. *PLoS One* 2015;10: e0122464.
- Wang T, Guan W, Lin J, Boutaoui N, Canino G, Luo J, Celedón JC, Chen W. A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. *Epigenetics* 2015;10:662–669.
- Maksimovic J, Gordon L, Oshlack A. SWAN: subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol* 2012;13:R44.
- Macintyre G, Bailey J, Haviv I, Kowalczyk A. is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics* 2010;26: i524–i530.
- Ørvevik J, Låg M, Lecureur V, Gilot D, Lagadic-Gossman D, Refsnes M, Schwarze PE, Skuland T, Becher R, Holme JA. AhR and Arnt differentially regulate NF- κ B signaling and chemokine responses in human bronchial epithelial cells. *Cell Commun Signal* 2014;12:48.
- Baay-Guzman GJ, Bebenek IG, Zeidler M, Hernandez-Pando R, Vega MI, Garcia-Zepeda EA, Antonio-Andres G, Bonavida B, Riedl M, Kleerup E, et al. HIF-1 expression is associated with CCL2 chemokine expression in airway inflammatory cells: implications in allergic airway inflammation. *Respir Res* 2012;13:60.
- Crotty Alexander LE, Akong-Moore K, Feldstein S, Johansson P, Nguyen A, McEachern EK, Nicatia S, Cowburn AS, Olson J, Cho JY, et al. Myeloid cell HIF-1 α regulates asthma airway resistance and eosinophil function. *J Mol Med (Berl)* 2013;91:637–644.
- Soulez M, Rouviere CG, Chafey P, Hentzen D, Vandromme M, Lautredou N, Lamb N, Kahn A, Tuil D. Growth and differentiation of C2 myogenic cells are dependent on serum response factor. *Mol Cell Biol* 1996;16:6065–6074.
- Sisson TH, Ajayi IO, Subbotina N, Dodi AE, Rodansky ES, Chibucos LN, Kim KK, Keshamouni VG, White ES, Zhou Y, et al. Inhibition of myocardin-related transcription factor/serum response factor signaling decreases lung fibrosis and promotes mesenchymal cell apoptosis. *Am J Pathol* 2015;185:969–986.
- Gao E, Wang Y, Alcorn JL, Mendelson CR. Transcription factor USF2 is developmentally regulated in fetal lung and acts together with USF1 to induce SP-A gene expression. *Am J Physiol Lung Cell Mol Physiol* 2003;284:L1027–L1036.
- Ruddy MJ, Wong GC, Liu XK, Yamamoto H, Kasayama S, Kirkwood KL, Gaffen SL. Functional cooperation between interleukin-17 and tumor necrosis factor- α is mediated by CCAAT/enhancer-binding protein family members. *J Biol Chem* 2004;279: 2559–2567.
- Hecker M, Behnk A, Morty RE, Sommer N, Vadász I, Herold S, Seeger W, Mayer K. PPAR- α activation reduced LPS-induced inflammation in alveolar epithelial cells. *Exp Lung Res* 2015;41:393–403.
- Lim H, Cho M, Choi G, Na H, Chung Y. Dynamic control of Th2 cell responses by STAT3 during allergic lung inflammation in mice. *Int Immunopharmacol* 2015;28:846–853.
- Haley KJ, Lasky-Su J, Manoli SE, Smith LA, Shahsafaei A, Weiss ST, Tantisira K. RUNX transcription factors: association with pediatric asthma and modulated by maternal smoking. *Am J Physiol Lung Cell Mol Physiol* 2011;301:L693–L701.
- Tavernier J, Devos R, Cornelis S, Tuytens T, Van der Heyden J, Fiers W, Plaetinck G. A human high affinity interleukin-5 receptor (IL5R) is composed of an IL5-specific α chain and a β chain shared with the receptor for GM-CSF. *Cell* 1991;66: 1175–1184.
- Forno E, Lasky-Su J, Himes B, Howrylak J, Ramsey C, Brehm J, Klanderman B, Ziniti J, Melén E, Pershagen G, et al. Genomewide association study of the age of onset of childhood asthma. *J Allergy Clin Immunol* 2012;130:83–90.e4.
- Liu LY, Sedgwick JB, Bates ME, Vrtis RF, Gern JE, Kita H, Jarjour NN, Busse WW, Kelly EA. Decreased expression of membrane IL-5 receptor α on human eosinophils: IL-5 down-modulates its receptor via a proteinase-mediated process. *J Immunol* 2002;169: 6459–6466.
- Monahan J, Siegel N, Keith R, Caparon M, Christine L, Compton R, Cusik S, Hirsch J, Huynh M, Devine C, et al. Attenuation of IL-5-mediated signal transduction, eosinophil survival, and inflammatory mediator release by a soluble human IL-5 receptor. *J Immunol* 1997;159:4024–4034.
- Semic-Jusufagic A, Gevaert P, Bachert C, Murray C, Simpson A, Custovic A. Increased serum-soluble interleukin-5 receptor α level precedes the development of eczema in children. *Pediatr Allergy Immunol* 2010;21:1052–1058.
- Rohde G, Gevaert P, Holtappels G, Fransen L, Borg I, Wiethage A, Arinir U, Tavernier J, Schultze-Werninghaus G, Bachert C. Soluble interleukin-5 receptor α is increased in acute exacerbation of chronic obstructive pulmonary disease. *Int Arch Allergy Immunol* 2004;135: 54–61.
- Gevaert P, Bachert C, Holtappels G, Novo CP, Van der Heyden J, Fransen L, Depraetere S, Walter H, van Cauwenberge P, Tavernier J. Enhanced soluble interleukin-5 receptor α expression in nasal polyposis. *Allergy* 2003;58:371–379.

36. Wilson TM, Maric I, Shukla J, Brown M, Santos C, Simakova O, Khoury P, Fay MP, Kozhich A, Kolbeck R, *et al.* IL-5 receptor α levels in patients with marked eosinophilia or mastocytosis. *J Allergy Clin Immunol* 2011;128:1086–1092.e1–3.
37. Castro M, Wenzel SE, Bleecker ER, Pizzichini E, Kuna P, Busse WW, Gossage DL, Ward CK, Wu Y, Wang B, *et al.* Benralizumab, an anti-interleukin 5 receptor α monoclonal antibody, versus placebo for uncontrolled eosinophilic asthma: a phase 2b randomised dose-ranging study. *Lancet Respir Med* 2014;2:879–890.
38. Liang L, Willis-Owen SA, Laprise C, Wong KC, Davies GA, Hudson TJ, Binia A, Hopkin JM, Yang IV, Grundberg E, *et al.* An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* 2015;520:670–674.
39. FitzGerald JM, Bleecker ER, Nair P, Korn S, Ohta K, Lommatzsch M, Ferguson GT, Busse WW, Barker P, Sproule S, *et al.*; CALIMA study investigators. Benralizumab, an anti-interleukin-5 receptor α monoclonal antibody, as add-on treatment for patients with severe, uncontrolled, eosinophilic asthma (CALIMA): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet* 2016;388:2128–2141.
40. Bleecker ER, FitzGerald JM, Chanez P, Papi A, Weinstein SF, Barker P, Sproule S, Gilmartin G, Aurivillius M, Werkström V, *et al.*; SIROCCO study investigators. Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting β 2-agonists (SIROCCO): a randomised, multicentre, placebo-controlled phase 3 trial. *Lancet* 2016;388:2115–2127.