Gene-based association study of rare variants in children of diverse ancestries implicates TNFRSF21 in the development of allergic asthma

Selene Clay  
*University of Chicago*

Katherine Rivera-Spoljaric  
*Washington University School of Medicine in St. Louis*

et al.

Follow this and additional works at: [https://digitalcommons.wustl.edu/oa_4](https://digitalcommons.wustl.edu/oa_4)

Part of the *Medicine and Health Sciences Commons*

Please let us know how this document benefits you.

**Recommended Citation**

Clay, Selene; Rivera-Spoljaric, Katherine; and et al., "Gene-based association study of rare variants in children of diverse ancestries implicates TNFRSF21 in the development of allergic asthma." *Journal of Allergy and Clinical Immunology*. 153, 3. 809 - 820. (2024).

[https://digitalcommons.wustl.edu/oa_4/3526](https://digitalcommons.wustl.edu/oa_4/3526)

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Gene-based association study of rare variants in children of diverse ancestries implicates TNFRSF21 in the development of allergic asthma

Selene Clay, PhD, a Jehan Alladina, MD, b, c, e Neal P. Smith, MS, c, d, e Cynthia M. Visness, PhD, f Robert A. Wood, MD, g George T. O’Connor, MD, MS, h Robyn T. Cohen, MD, h Gurjot K. Khurana Hershey, MD, PhD, j Carolyn M. Kercsmar, MD, l Rebecca S. Gruchalla, MD, PhD, k Michelle A. Gill, MD, l Andrew H. Liu, MD, m Haejin Kim, MD, n Meyer Kattan, MD, o Leonard B. Bacharier, MD, p Deepa Rastogi, MD, q Katherine Rivera-Spoljaric, MD, r Rachel G. Robison, MD, s,t Peter J. Gergen, MD, t William W. Busse, MD, u Alexandra-Chloe Villani, PhD, c, d, e Josalyn L. Cho, MD, v Benjamin D. Medoff, MD, b, e James E. Gern, MD, w Daniel J. Jackson, MD, w Carole Ober, PhD, x and Matthew Dapas, PhD a

Chicago, Ill; Boston and Cambridge, Mass; Durham, NC; Baltimore and Rockville, Md; Cincinnati, Ohio; Dallas, Tex; St Louis, Mo; Aurora, Colo; Detroit, Mich; New York, NY; Nashville, Tenn; Washington, DC; Madison, Wis; and Iowa City, Iowa

Background: Most genetic studies of asthma and allergy have focused on common variation in individuals primarily of European ancestry. Studying the role of rare variation in quantitative phenotypes and in asthma phenotypes in populations of diverse ancestries can provide additional, important insights into the development of these traits. Objectives: We sought to examine the contribution of rare variants to different asthma- or allergy-associated quantitative traits in children with diverse ancestries and explore their role in asthma phenotypes.

Methods: We examined whole-genome sequencing data from children participants in longitudinal studies of asthma (n = 1035; parent-identified as 67% Black and 25% Hispanic) to identify rare variants (minor allele frequency < 0.01). We assigned variants to genes and tested for associations using an omnibus variant-set test between each of 24,902 genes and 8 asthma-associated quantitative traits. On combining our results with external data on predicted gene expression in humans and mouse knockout studies, we identified 3 candidate genes. A burden of rare variants in each gene and in a combined 3-gene score was tested for its associations with clinical phenotypes of asthma. Finally, published single-cell gene expression data in lower airway mucosal cells after allergen challenge were used to assess transcriptional responses to allergen.

Results: Rare variants in USF1 were significantly associated with blood neutrophil count (P = 2.18 × 10−7); rare variants in TNFRSF21 with total IgE (P = 6.47 × 10−6) and PIK3R6 with eosinophil count (P = 4.10 × 10−5) reached suggestive significance. These 3 findings were supported by independent data from human and mouse studies. A burden of rare variants in TNFRSF21 and in a 3-gene score was associated with allergy-related phenotypes in cohorts of children with mild and severe asthma. Furthermore, TNFRSF21 was significantly upregulated in bronchial basal epithelial cells from adults with allergic asthma but not in adults with allergies (but not asthma) after allergen challenge.

Conclusions: We report novel associations between rare variants in genes and allergic and inflammatory phenotypes in children with diverse ancestries, highlighting TNFRSF21 as contributing to the development of allergic asthma. (J Allergy Clin Immunol 2024;153:809-20.)

Key words: Whole-genome sequencing, total IgE, neutrophils, eosinophils

Genome-wide association studies (GWASs) have identified thousands of associations between common variants and hundreds of complex traits. However, rare variants, which comprise the bulk of human genetic variation, have been largely

From the Department of Human Genetics, University of Chicago, Chicago; bthe Division of Pulmonary and Critical Care Medicine and cthe Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital and Harvard Medical School, Boston; dBroad Institute of Massachusetts Institute of Technology and Harvard, Cambridge; eMassachusetts General Hospital Cancer Center, Boston; fRho, Inc, Durham; gPediatric Allergy and Immunology Department, Johns Hopkins Bloomberg School of Public Health, Baltimore; hthe Department of Pediatrics, Boston University School of Medicine, Boston; ithe Division of Asthma Research and jthe Division of Pulmonary Medicine, Cincinnati Children’s Hospital Medical Center, Cincinnati; kInternal Medicine and Pediatrics, University of Texas Southwestern Medical Center, Dallas; lPediatric Infectious Diseases, St Louis Children’s Hospital, St Louis; mthe Breathing Institute, Children’s Hospital Colorado, Aurora; nAllergy and Immunology, Henry Ford Health, Detroit; othe Department of Pediatrics, Columbia University Medical Center, New York; pthe Department of Pediatrics, Monroe Carell Jr Children’s Hospital at Vanderbilt University Medical Center, Nashville; qthe Division of Pulmonology and Sleep Medicine, Children’s National Hospital, Washington; rthe Department of Pediatric Allergy, Immunology, and Pulmonary Medicine, Washington University School of Medicine, St Louis; sAnn & Robert H. Lurie Children’s Hospital, Chicago; tthe National Institute of Allergy and Infectious Diseases, Rockville; uthe Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison; vthe Division of Pulmonary, Critical Care and Occupational Medicine, University of Iowa Carver College of Medicine, Iowa City; and wthe Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison.

Received for publication April 3, 2023; revised September 25, 2023; accepted for publication October 12, 2023.

Available online November 7, 2023.

Corresponding author: Selene Clay, PhD, The University of Chicago, Human Genetics, 925 E. 58th St, Chicago, IL 60637. E-mail: seleneclay@uchicago.edu.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections.

0901-6749 © 2023 The Authors. Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.1016/j.jaci.2023.10.023
**Abbreviations used**

AA: Allergic asthma  
AC: Allergic (nonasthma) control  
APIC: Asthma Phenotypes in the Inner City  
GWAS: Genome-wide association study  
LW-HA: Low wheeze/high atopy  
MAF: Minor allele frequency  
PIK3R6: Phosphoinositide-3-kinase regulatory subunit  
ppb: Parts per billion  
STAAR: Variant-Set Test for Association using Annotation  
TNFRSF21: Tumor necrosis factor receptor superfamily member 21  
TSS: Transcription start site  
URECA: Urban Environment and Childhood Asthma  
USF1: Upstream transcription factor 1  
VAMP3: Vesicle-associated membrane protein 3

overlooked because they have not been well represented by the variants included on commonly used genotyping arrays and because most studies are underpowered to detect individual rare variant associations. However, the increased availability of whole-genome sequencing and the development of variant-set tests that jointly test the association of multiple variants in a gene allow for a more comprehensive study of the role of genome-wide rare variation in complex traits. Although these variants do not necessarily explain a significant proportion of heritability, rare variant studies can more directly identify causal genes and mechanisms, as well as novel therapeutic targets.

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

Other studies have explored the role of rare variation in asthma-associated traits, including allergic, inflammatory, and pulmonary traits, some in individuals with diverse ancestries. However, these studies most often examined just 1 or a limited number of traits, and most used targeted sequencing or exonic variants, or performed GWASs of individual rare variants. Capturing all coding and noncoding variation across the entire genome through whole-genome sequencing and using a gene-based test may be a more effective approach for studying the genetic architecture of asthma-associated quantitative phenotypes and clinical phenotypes of asthma.

In this study, we explored the contributions of rare variants (minor allele frequency [MAF] < 0.01) to 8 asthma- and allergy-associated quantitative traits in 1035 children with diverse ancestries from 2 longitudinal studies: the Asthma Phenotypes in the Inner City (APIC) study and the URBan Environment and Childhood Asthma (URECA) birth cohort study. Three of the top gene associations that we identified were further supported by external data, and single-cell sequencing data from lower airway mucosal cells before and after allergen challenge highlighted a potentially important role of TNFRSF21 in the development of allergic asthma (AA), and as a possible therapeutic target for AA.

**METHODS**

**Study populations**

These studies included children from 2 cohorts that were component studies of the Inner-City Asthma Consortium (for study overview, see Fig E1 in this article’s Online Repository at www.jacionline.org). APIC was a 1-year longitudinal study of children and adolescents (ages 6-17 years) with asthma living in low-income areas of 9 US cities (Baltimore, Md; Boston, Mass; Chicago, Ill; Cincinnati, Ohio; Dallas, Tex; Denver, Colo; Detroit, Mich; New York, NY; and Washington, DC). All APIC participants had a physician’s diagnosis for asthma and at least 2 asthma episodes requiring bronchodilator administration in the previous year. URECA is a prospective birth cohort study of children living in low-income areas of 4 US cities (Baltimore, Md; Boston, Mass; New York, NY; and St Louis, Mo). Mothers were enrolled into this study during pregnancy; at least 1 parent had a history of asthma or an allergic disease; asthma in the child was defined by physician’s diagnosis at age 7 or 10 years, lung function criteria, and/or reported symptoms.

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

**Asthma-associated quantitative phenotypes**

Eight quantitative traits that reflect component features of asthma were available for both cohorts and were the focus of this study. We included 2 quantitative measures of allergic sensitization: total serum IgE (IU) concentration and the number of sensitizations to 15 common inhaled and food allergens (mouse, dog, house dust mite [×2], cat, roach, mold [×2], ragweed, maple, oak, timothy grass, peanut, egg, and milk), according to serum allergen-specific IgE concentration (positive is ≥0.35 kU/L). Immune and inflammatory phenotypes included blood cell counts (eosinophils and neutrophils, both in cells/mm³) and fractional exhaled nitric oxide in parts per billion, which was measured following American Thoracic Society guidelines. Lung function measures included percent predicted FEV₁, FEV₁/forced vital capacity (using the normalized z score), and bronchodilator response (percentage change from baseline in FEV₁ following 4 inhalations of albuterol). The Mann-Whitney U test was used to assess differences in traits between individuals with and without asthma, and the chi-square test was used to assess differences in self-reported race and sex ratios (% female).
Quantitative trait normalization

We performed trait normalization for the 8 quantitative traits. We adjusted for potential confounding factors by fitting a linear mixed model accounting for asthma status, age at which the trait was measured, sex, the first 10 principal components of ancestry (from common variants), and genetic relatedness between individuals modeled as a random effect. Asthma was included as a covariate to account for disease-induced phenotypic differences; analyses for the most significant results were repeated without including asthma as a covariate in individuals with asthma only to exclude the possibility of collider bias effects. We applied a quasi-Poisson linear mixed model for the count data (blood neutrophil count, blood eosinophil count, and allergen sensitization) and a gaussian linear mixed model for all other traits. Total IgE and blood neutrophil count were log transformed, and blood eosinophil count was square-root transformed before regression fitting. For neutrophil count, we additionally corrected for the common variant within Duffy blood group gene, ACKR1/DARC (rs2814778; 1 for CC, 0 for CT, TT), because the homozygous CC genotype is associated with decreased neutrophil counts in individuals with African ancestry. We also repeated analyses without including rs2814778 as a covariate for comparison. For bronchodilator response, we additionally corrected for FEV1 % predicted and FEV1/VC. Residuals from all regressions were rank-based, inverse-normal transformed, and used as the outcome variables in further analyses.

Whole-genome sequencing and variant calling

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

To isolate rare variants, we first selected the 21,073,226 variants with MAF less than 0.01 in the combined APIC and URECA children (see [Fig E2 in this article’s Online Repository at www.jacionline.org]). We then excluded variants that were common (MAF \( \geq 0.01 \)) in any of the 1000 Genomes Project super-populations (African, American, East Asian, European, and South Asian) because variants with deleterious effects are unlikely to be found at common frequencies in any population. This removed 4.56% of variants in our data set, leaving a final set of 20,093,812 variants for downstream analysis.

Association tests

To group variants for gene-based association testing, we binned all coding and noncoding rare variants located between the 3'UTR and 5kb upstream of the 5' UTR for each gene, including both protein-coding genes and ncRNAs. We used STAAR (variant-Set Test for Association using Annotation information) for gene-based association testing with the 8 quantitative phenotypes described above. STAAR incorporates 3 variant-set tests: the burden test, the sequence kernel association test, and the aggregated Cauchy association test. We did not weigh the variants by their frequency due to the narrow range of MAFs (0.00097-0.099) in this sample. We excluded genes with fewer than 5 rare variants, resulting in 24,902 genes for analysis. We used a stringent Bonferroni correction of 0.05/(24,902 \( \times 8 \)) phenotypes less than 2.51 \( \times 10^{-7} \) and considered a P-value threshold of less than 1 \( \times 10^{-5} \) for suggestive significance. For the top gene associations with each trait, we further examined associations stratified by coding and noncoding variants in the gene set. When a subset of variants assigned to one gene set resided in the exon of another nearby gene, we repeated analyses after excluding those variants from the gene set. For all genes with evidence of association, we also examined associations for each variant within the gene set through linear regression using the same covariates used in the gene-based tests.

External gene and variant validation

To further evaluate the associations with the most significant gene for each of the 8 traits, we first assessed whether predicted gene expression was associated with the same or related trait in the phenomeXcan database, which reports associations between predicted gene expression based on genetic variation and specific phenotypes (primarily from the UK Biobank). We then evaluated publicly available mouse knockout studies of the associated genes to see whether the resulting phenotypes included traits related to those in our study. We next examined which variants resided in active enhancer and transcriptional start sites in blood and epithelial cells from ROADMAP.

Burden of rare variants and asthma phenotypes

For each of the 3 candidate genes prioritized using external evidence, we defined the number of rare variants in that gene as its “burden score” for each individual. The total of rare variants across all 3 genes for each individual was used in a “3-gene score.” These burden scores were tested for association with asthma and allergy phenotypes that were previously defined within the APIC and URECA cohorts, as described.

We performed Mann-Whitney U tests to determine whether the individual or 3-gene scores were associated with clinical phenotypes. The APIC cohort was composed of children with a doctor’s diagnosis of asthma. Using predefined phenotypes in APIC, we compared phenotype A (minimally symptomatic asthma and rhinitis, low allergy/inflammation, and normal pulmonary physiology during the study) to each of the more severe phenotypes B to E (B: lower allergy and inflammation, intermediate rhinitis symptoms, and mildly impaired pulmonary physiology; C: minimally symptomatic asthma and rhinitis with an intermediate degree of allergy and allergic inflammation; D: symptomatic rhinitis and higher levels of allergy and allergic inflammation, with intermediate impairment of pulmonary physiology; E: highest levels of symptomatic asthma and rhinitis and highest degree of allergy and allergic inflammation, with most impaired pulmonary physiology). The URECA cohort includes children with and without asthma, with the former comprised largely of children with mild asthma. Using predefined phenotypes in URECA, we compared children with little to no allergy or asthma (phenotype low wheeze/low atopy) to each of 5 other phenotypes: low wheeze/high atopy (LW-HA), transient wheeze/low atopy, high wheeze/low atopy, high wheeze/high atopy, and high wheeze/low atopy. The comparisons within each
TABLE I. Sample composition

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All</th>
<th>Asthma</th>
<th>Nonasthma</th>
<th>P value (asthma vs nonasthma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>1035</td>
<td>681</td>
<td>226</td>
<td>—</td>
</tr>
<tr>
<td>Mean age (y)</td>
<td>9.99</td>
<td>10.56</td>
<td>9.37</td>
<td>1.51 ( \times 10^{-7} )</td>
</tr>
<tr>
<td>% Female</td>
<td>46.18</td>
<td>43.86</td>
<td>52.65</td>
<td>0.025</td>
</tr>
<tr>
<td>% Self-reported race and ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>67.25</td>
<td>75.22</td>
<td>72.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Hispanic</td>
<td>24.93</td>
<td>27.50</td>
<td>16.81</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.35</td>
<td>1.03</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>Other/mixed/unknown</td>
<td>6.47</td>
<td>6.27</td>
<td>6.64</td>
<td></td>
</tr>
</tbody>
</table>

Traits, median (IQR)

- Allergen sensitization (positive test results) 3.00 (7) 4.67 (7.67) 1.00 (3.33) 7.65 \( \times 10^{-21} \)
- Total IgE concentration (IU/mL) 158 (463) 243 (655) 70 (123) 7.67 \( \times 10^{-21} \)
- Blood eosinophil count (cells/mm\(^3\)) 200 (300) 300 (340) 200 (200) 3.13 \( \times 10^{-12} \)
- Total IgE concentration (IU/mL) 158 (463) 243 (655) 70 (123) 7.67 \( \times 10^{-21} \)
- Blood eosinophil count (cells/mm\(^3\)) 200 (300) 300 (340) 200 (200) 3.13 \( \times 10^{-12} \)
- Bronchodilator response (% change from baseline) 8.40 (10.82) 9.80 (11.84) 5.60 (8.06) 7.83 \( \times 10^{-14} \)

Results are reported separately by asthma status for the combined APIC and URECA samples. The ages correspond to the year at which total IgE was measured (all other ages listed in Table E1). The medians and IQRs (in parentheses) are listed for all quantitative traits. Missing data were not included in calculations.

FVC, Forced vital capacity; FENO, fractional exhaled nitric oxide; IQR, interquartile range; IU, international units; ppb, parts per billion.

Patterns of gene expression and responses to allergen challenge

To gain further insight into the potential functional or clinical effects of the prioritized genes, we used publicly available single-cell RNA-sequencing data from lower airway mucosal cells from 4 individuals with AA and 4 individuals with allergies alone (no asthma) at baseline and 24 hours after segmental allergen challenge from Alladina et al.\(^5\) Using data on differentially expressed genes (Data File S6 in the published manuscript) and their interactive website (https://villani.mgh.harvard.edu/allergy-asthma/), we examined the expression patterns and differential expression between individuals with AA and individuals with allergy only at baseline and after allergen challenge for the 3 prioritized genes.

Results are reported separately by asthma status for the combined APIC and URECA samples. The ages correspond to the year at which total IgE was measured (all other ages listed in Table E1). The medians and IQRs (in parentheses) are listed for all quantitative traits. Missing data were not included in calculations.

Overview of associations

We performed gene-based variant-set tests for 8 quantitative traits in participants from the APIC and URECA cohorts. We assigned rare variants to 25,605 genes, and required gene sets to have at least 5 variants, resulting in 24,902 genes examined. The mean number of variants in each variant set was 508 (median number, 207; range, 5-17,174; see Fig E5 in this article’s Online Repository available at www.jacionline.org). The gene-based association test results are presented in Table E2 in the Online Repository available at www.jacionline.org; the most significant gene for each of the 8 traits is listed in Table II.

All associations remained nominally significant \((P < 0.05)\) in analyses that did not include asthma as a covariate in the trait normalization, and all associations except the association of VAMP3 (vesicle-associated membrane protein 3) with allergic sensitization remained nominally significant in the analyses including only children with asthma (see Table E3 in this article’s Online Repository at www.jacionline.org). For 2 genes, VAMP3 and USF1 (upstream transcription factor 1), the variants that were within 5kb upstream of their 5’UTR included some that were in the exons of neighboring genes and were therefore designated as coding variants. Nine single nucleotide variants in the VAMP3 set were coding variants for CAMTA1 (calmodulin binding transcription activator 1); see Fig E6 in this article’s Online Repository at www.jacionline.org; when these single nucleotide

phenotype were not completely independent (all compared against the same reference group), so Bonferroni corrections would have been overly conservative.

RESULTS

Sample composition

We examined rare variation and allergy- and asthma-associated phenotypes from 1035 children in APIC \((n = 508)\) and URECA \((n = 527)\). The parent-reported racial and ethnic composition of their children was 67% Black non-Hispanic, 25% Hispanic, and 8% other (White, mixed, unknown) in APIC and 72% Black non-Hispanic, 20% Hispanic, and 8% other (White, mixed, unknown) in URECA. The demographic and clinical characteristics of these children are presented in Table 1 and Table E1; a principal-component analysis plot of the genetic ancestries of these children is shown in Fig E3 (in the Online Repository available at www.jacionline.org). In this sample, there were fewer females and fewer parent-reported Black and more parent-reported Hispanic individuals among children with asthma compared with children without asthma, but neither proportion of self-reported race and ethnicity nor % female was significantly different after correcting for 11 comparisons \((P_{\text{corrected}} < 0.045)\). Except for blood neutrophil count, measurements of all the clinical phenotypes significantly differed between children with and without asthma after multiple testing correction (Table I). The correlations between phenotypes are shown in Fig E4 (in the Online Repository available at www.jacionline.org).
variants were excluded, the $P$ values for VAMP3 were largely unchanged ($5.13 \times 10^{-3}$ for the set and $P = .85$ for coding variation). The USF1 results are further discussed in the next section.

We next used publicly available resources to investigate additional evidence supporting the most significant gene associations for each of the 8 phenotypes (see Tables E4 and E5 in this article’s Online Repository at www.jacionline.org). Three associations were supported by orthogonal data: USF1 with blood neutrophil count, TNFRSF21 (tumor necrosis factor receptor superfamily member 21) with total IgE, and PIK3R6 (phosphoinositide-3-kinase regulatory subunit) with blood eosinophil count. These are discussed below and presented in bold font in Table II. The remaining associations did not have corroborating support from these external resources and were not further investigated.

**USF1**

**Blood neutrophil count**

The most significant gene-trait pair was USF1 with blood neutrophil count (Table II and Fig 1, A). This association was also significant ($P = 1.97 \times 10^{-8}$) in a secondary analysis that

![Image](FIG 1. USF1 and neutrophil count. A, Manhattan plot for gene set associations with blood neutrophil count. Each point represents a gene. The red line is the Bonferroni significance threshold. B, Marginal associations and locations of variants in the USF1 variant set. Variants are color coded green if they are noncoding and orange if coding in either USF1 or its neighboring gene, ARHGAP30. C, USF1 association with other traits. FENO, Fractional exhaled nitric oxide; FVC, forced vital capacity.)

**TABLE II. Results of rare-variant, gene-based association studies**

<table>
<thead>
<tr>
<th>Trait</th>
<th>Gene*</th>
<th>Location</th>
<th># Var</th>
<th>Coding $P$ value</th>
<th>Noncoding $P$ value</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood neutrophil count</td>
<td>USF1</td>
<td>1q23.3</td>
<td>79</td>
<td>.98</td>
<td>$1.15 \times 10^{-9}$</td>
<td>$2.18 \times 10^{-7}$</td>
</tr>
<tr>
<td>Total IgE</td>
<td>TNFRSF21</td>
<td>6p12.3</td>
<td>794</td>
<td>.24</td>
<td>$5.90 \times 10^{-6}$</td>
<td>$6.47 \times 10^{-6}$</td>
</tr>
<tr>
<td>FENO</td>
<td>CTBP1-AS</td>
<td>4p16.3</td>
<td>151</td>
<td>.72</td>
<td>$3.26 \times 10^{-5}$</td>
<td>$3.43 \times 10^{-5}$</td>
</tr>
<tr>
<td>FEV1/FVC</td>
<td>VRK3</td>
<td>19q13.33</td>
<td>547</td>
<td>.04</td>
<td>$3.59 \times 10^{-5}$</td>
<td>$3.15 \times 10^{-5}$</td>
</tr>
<tr>
<td>Bronchodilator response</td>
<td>MRPL44</td>
<td>2q36.1</td>
<td>129</td>
<td>$1.68 \times 10^{-3}$</td>
<td>$3.68 \times 10^{-5}$</td>
<td>$3.44 \times 10^{-5}$</td>
</tr>
<tr>
<td>Blood eosinophil count</td>
<td>PIK3R6</td>
<td>17p13.1</td>
<td>608</td>
<td>$8.93 \times 10^{-3}$</td>
<td>$5.12 \times 10^{-5}$</td>
<td>$4.10 \times 10^{-5}$</td>
</tr>
<tr>
<td>FEV1% predicted</td>
<td>TEX36-AS1</td>
<td>10q26.13</td>
<td>93</td>
<td>NA</td>
<td>$8.61 \times 10^{-3}$</td>
<td>$6.08 \times 10^{-5}$</td>
</tr>
<tr>
<td>Allergen sensitization</td>
<td>VAMP3</td>
<td>1p36.23</td>
<td>124</td>
<td>.75</td>
<td>$3.96 \times 10^{-5}$</td>
<td>$6.42 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

The most significant gene association for each trait is given. For each quantitative trait, the gene, location, number of variants in the gene set (# Var), $P$ value in analyses considering just coding or just noncoding variants in the variant set, and the overall $P$ value are given. One association (USF1 with blood neutrophil count, $P = 2.18 \times 10^{-7}$) met the Bonferroni-adjusted significance threshold of $2.51 \times 10^{-7}$. The full results are presented in Table E2. The 3 bolded associations are the results with support from external sources and are described in further detail.

*CTBP1-AS, C-terminal binding protein 1 antisense RNA; VRK3, vaccinia-related kinase 3; MRPL44, mitochondrial ribosomal protein L44; TEX36-AS1, testis expressed protein 36 antisense RNA 1.
did not correct for the genotype at the Duffy blood group gene (see Methods). The association was more significant when only the 63 noncoding variants were included in the set \((P = 1.15 \times 10^{-6})\); the set comprised only of the 16 coding variants showed no evidence of associations \((P = .98)\) (Table II). No single variant contributed disproportionately to the association (Fig 1, B). The variant set included 13 SNVs in the coding region of the nearby gene \((ARHGAP30)\). When we removed these variants, the association for \(USF1\) was more significant \((P = 1.55 \times 10^{-9} \text{ vs } 2.18 \times 10^{-7})\). This variant set was not associated with any of the other 7 traits (Fig 1, C).

We evaluated the \(USF1\) gene association further using publicly available mouse knockout data and phenomeXcan, a resource that reports associations between predicted gene expression from multiple tissues and GWAS traits.\[^{46}\] PhenoXcan reported that predicted \(USF1\) expression across 33 tissues (see Table E6 in this article’s Online Repository at www.jacionline.org) was significantly associated with both neutrophil percentage \((P = 4.66 \times 10^{-4})\) and neutrophil count \((P = 8.02 \times 10^{-3})\), with lower predicted \(USF1\) expression associated with higher values of both measures. Furthermore, mice with \(Usf1\) knocked out in bone marrow cells had increased blood neutrophil counts,\[^{52}\] which is consistent with the direction of effect from phenomeXcan. According to epithelial cell annotations in ROADMAP,\[^{49}\] 28% of rare variants we identified for \(USF1\) were located in the transcription start site (TSS) and none were in enhancers. Using the blood cell annotations, 19% were located in the TSS and 34% were in enhancers (see Fig E7 in this article’s Online Repository at www.jacionline.org). These orthogonal data based on gene expression and mouse knockout studies support our results, validating the association between \(USF1\) and neutrophil counts, and further suggesting that rare variants in \(USF1\) impact neutrophilia via their effects on gene expression.

**TNFRSF21 is associated with total IgE**

The second most significant association, which reached suggestive significance \((P = 6.47 \times 10^{-6})\), was between \(TNFRSF21\) and total IgE levels (Fig 2, A). Similar to \(USF1\), no single variant contributed disproportionately to the association (Fig 2, B). \(TNFRSF21\) was also nominally associated with both allergen sensitization and blood eosinophil count (Fig 2, C). However, these 3 traits were highly correlated (Fig E4), and when we repeated the associations including total IgE levels as a covariate in the model, the associations with allergen sensitization and blood eosinophil count were no longer significant (Fig 2, C). This conditional analysis suggests that this gene may be a marker of more generalized atopy or type 2 inflammation. In addition, the association with \(TNFRSF21\) was slightly more significant when considering the 776 noncoding variants \((P = 5.90 \times 10^{-6} \text{ vs } 6.47 \times 10^{-6})\), whereas the association with the 18 coding variants was not significant \((P = .24; \text{ Table II})\).

Total IgE was not available in the phenomeXcan data set, and predicted \(TNFRSF21\) expression was not associated with the available allergic traits in the UK Biobank (food allergy, allergic rhinitis, and eczema). However, the Mouse Genome Informatics resource reported that \(Tnfrsf21\) mouse knockouts had significantly increased IgE levels in response to protein challenge,\[^{28,53}\] supporting a relationship between \(TNFRSF21\) and total IgE. ROADMAP epithelial cell annotations indicated that 3% of rare variants were located in the TSS and 4% were in enhancers. According to blood cell annotations, 0.1% were located in the TSS and 4% were in enhancers (Fig E7). Although more limited, these results support the association between \(TNFRSF21\) and total IgE and suggest that rare variants in this gene may influence IgE levels or type 2 inflammation via their impact on gene expression.

**FIG 2.** \(TNFRSF21\) and total IgE. A, Manhattan plot for total IgE levels. Red line is Bonferroni significance and blue is suggestive significance \((1.00 \times 10^{-8})\). B, Marginal associations of variants in the \(TNFRSF21\) variant set. C, \(TNFRSF21\) association with the other traits before and after conditioning on total IgE. FENO, Fractional exhaled nitric oxide; FVC, forced vital capacity.
**PIK3R6 is associated with blood eosinophil count**

The most significant association for blood eosinophil counts was with \( \text{PIK3R6} \) with \( P = 4.10 \times 10^{-5} \) (Fig 3, A). In contrast to the previous 2 gene-trait pairs, the \( P \) value for the associations between \( \text{PIK3R6} \) and eosinophil count was slightly less significant when analyzing just the 581 noncoding variants \( (P = 5.19 \times 10^{-5}) \) and retained nominal significance \( (P = 3.90 \times 10^{-3}) \) when considering only the 27 coding variants. This indicates that both sets of variants contributed to the association. No single variant was responsible for the entire signal (Fig 3, B), and \( \text{PIK3R6} \) was not associated with any other trait (Fig 3, C).

The association was validated by phenomeXcan: the most significant association for predicted \( \text{PIK3R6} \) expression was with eosinophil percentage \( (P = 3.96 \times 10^{-6}) \) across 26 tissues (Table E6). It was also associated with eosinophil count \( (P = 1.85 \times 10^{-3}) \). Furthermore, \( \text{Pik3r6} \) knockout mice reported decreased granulocyte numbers.\(^{48,54}\) Although the type of granulocyte was not specified, these measures would have included eosinophils. According to epithelial cell annotations in ROADMAP, 5% of the rare variants were located in enhancers and none were in the TSS. Similarly, using the blood cell annotations, 16% were located in enhancers and none were in the TSS (Fig E7). Taken together, these results support a possible association between \( \text{PIK3R6} \) variants and eosinophil counts.

**TNFRSF21 and a 3-gene score is associated with respiratory phenotypes**

We next asked whether rare variant gene scores, reflecting the burden of rare variants for each individual, for each of the 3 candidate genes (\( \text{USF1, PIK3R6, TNFRSF21} \)) or a composite 3-gene score was associated with respiratory phenotypes in the URECA and APIC cohorts. In these analyses, we assigned a score to each individual that reflected the number of rare variants in each gene carried by that individual, or the number of rare variants across all 3 genes carried by that individual. We used phenotype groupings previously defined in these children.\(^{27,50}\) In URECA, composed mostly of children without asthma or with mild asthma, the \( \text{TNFRSF21} \) score was nominally associated with the LW-HA phenotype compared with the low wheeze, low atopy group \( (P = 2.25 \times 10^{-2}) \) (Fig 4, A; see Fig E8 and Table E7 in the Online Repository available at www.jacionline.org). The gene scores for \( \text{USF1} \) and \( \text{PIK3R6} \) were not associated with any phenotypes in the URECA children. However, the 3-gene score was also associated with the LW-HA phenotype and was more significant than the association with \( \text{TNFRSF21} \) alone \( (P = 5.69 \times 10^{-5}) \). These results in URECA children with allergic sensitization indicated a positive association between more rare variants and more atopy in the absence of wheeze.

In APIC, composed of children at all levels of asthma severity, the only nominally significant association with a single gene score was also between \( \text{TNFRSF21} \) and phenotype E compared with the mildest group (phenotype A) \( (P = 4.2 \times 10^{-2}) \). Phenotype E describes children with symptomatic asthma and the highest IgE levels, eosinophil counts, and allergen sensitizations compared with the other phenotype groups, consistent with this gene showing at least modest associations with these 3 phenotypes in the gene-based association tests. The 3-gene score also had a stronger association \( (P = 2.69 \times 10^{-2}) \) with phenotype E than did the \( \text{TNFRSF21} \) score alone, possibly reflecting the contribution of rare variants in all 3 genes in this phenotype of severe asthma. There were no other nominally significant associations between the other gene scores and phenotypes in the APIC.
children (Fig 4, B, Fig E8, and Table E7). These results in APIC further indicated a positive association between more rare variants in these genes and more severe disease.

**TNFRSF21 is highly expressed in basal epithelial cells and responsive to allergen challenge**

Finally, to gain further insight into the expression patterns and transcriptional regulation of the prioritized genes in airway cells, we used publicly available single-cell RNA-sequencing data in lower airway mucosal cells from individuals with AA and allergic (nonasthma) controls (AC) measured at baseline and after allergen challenge. 51

**DISCUSSION**

In this study, we examined the contribution of rare variants to allergy- and asthma-associated quantitative traits using a gene-based approach. We included a broad panel of allergic, immune/inflammatory, and pulmonary traits measured in children with diverse ancestries. Overall, our studies suggested that rare non-coding variation within or just upstream of genes contribute more...
to variation in these traits than do rare coding variation (Table II), further supporting other findings that complex traits are primarily driven by both common and rare noncoding variation that have effects on gene expression rather than on protein function.\textsuperscript{55-57} Indeed, 3 of the associations were supported by orthogonal evidence derived from independent data sets based on predicted gene expression and from mouse knockout studies. All 3 of these genes represent novel associations with immune, atopic, and inflammatory phenotypes.

One association was significant after multiple test correction for 24,902 genes. The gene set for USF1 was associated with blood neutrophil count in the APIC and URECA children, and predicted USF1 expression was associated with neutrophil counts in adults in the UK Biobank. Moreover, a mouse model with Usp1 knockout had increased neutrophil counts.\textsuperscript{52} Both the phenomeXcan and mouse knockout results indicated that decreased expression of this gene is associated with elevated levels of circulating neutrophils. USF1 encodes a transcription factor belonging to the basic helix-loop-helix leucine zipper family of proteins that can regulate expression through E-box motifs.\textsuperscript{38} This gene is also located within 2Mb of the Duffy blood group gene (ACKR1/DARC),\textsuperscript{54} but because this association remained significant when conditioning on the Duffy variant, the association between USF1 and neutrophil counts was not due to linkage disequilibrium with the Duffy null allele (see Methods). Several studies have reported a relationship between USF1 and immune-related traits,\textsuperscript{52,59-62} and the locus containing this gene has been significantly associated with both white blood cell count\textsuperscript{63} and granulocyte percentage of myeloid white cells\textsuperscript{64} in GWASs. As a transcription factor, USF1 may be regulating expression of genes important in inflammatory processes and possibly even asthma: several studies have reported that asthma risk alleles at the loci encoding the MUC5AC and ORMDL3/GSDMB genes affect binding of USF1.\textsuperscript{65-67} Although our studies did not support a role for rare variants in this gene contributing to asthma or atopy susceptibility in children, these combined data strongly support a role for this gene in determining circulating neutrophil levels and possibly in other inflammatory conditions.\textsuperscript{59-62}

We also report an association between PIK3R6 and eosinophil count. This was further supported by both phenomeXcan and Pik3r6 knockout mice.\textsuperscript{54} PIK3R6 encodes a lipid kinase that acts as a regulatory subunit for the PI3K gamma complex and is primarily activated by G protein–coupled receptors.\textsuperscript{60} Notably, several studies in human cohorts have implicated PIK3R6 expression in other eosinophilic or allergic diseases. For example, PIK3R6 was among the most significantly differentially expressed genes in peripheral blood cells between eosinophilic and noneosinophilic chronic obstructive pulmonary disease,\textsuperscript{69} between patients with AA and nonasthma/nonallergic controls,\textsuperscript{70} and in a meta-analysis of atopy in white blood cells.\textsuperscript{71} Together, these studies point to a potentially important role of this gene in eosinophilic-related traits.

The second most significant association was between TNFRSF21 and total IgE. Measures of IgE were not available in phenomeXcan, but Tnfrsf21-deficient mice had increased IgE levels in response to protein challenge.\textsuperscript{53} TNFRSF21 encodes the Death Receptor 6 protein and is a member of the tumor necrosis factor superfamily and signals through the NF-κB pathway.\textsuperscript{72,73} Studies in mice have also implicated this gene in the Tgβ2 response, specifically through activating the Jun amino terminal kinase (JNK) pathway in regulating Tgβ2-cell differentiation.\textsuperscript{53} Tgβ2 inflammation is a hallmark of allergic disease, but evidence supporting a role of TNFRSF21 in this pathway has come largely from mouse studies. Our results implicate this gene more directly in the development of AA in humans.

Studies of quantitative traits are more powerful for identifying genetic associations than are studies of binary traits, particularly in small samples,\textsuperscript{76} but the goal of these studies is generally to identify variants that ultimately affect disease risk or outcomes. Therefore, we used quantitative trait mapping to identify 3 candidate genes that may impact risk for asthma or allergic disease. Analyses of these 3 genes with clinical phenotypes of allergy and asthma in the APIC and URECA cohorts converged on a role of rare variants in TNFRSF21 and the 3-gene score with AA phenotypes. In the URECA cohort, both scores were associated with the LW-HA phenotype. In the APIC cohort, TNFRSF21 and the 3-gene score were associated with the most severe asthma phenotype (symptomatic asthma, high IgE, eosinophilia, and sensitization to multiple allergens). The different associations between the 2 cohorts likely reflect the different ascertainment of children in each. URECA is a birth cohort in which a family member had asthma or allergic disease.
As a result, this cohort includes children without asthma and with generally mild asthma among affected children.26,50 In contrast, APIC is a 1-year longitudinal study of children with asthma, of whom nearly half were classified as “difficult to treat.”26,27 Despite these differences, both studies of urban children highlight TNFRSF21 and the 3-gene score in the development of allergy and asthma phenotypes.

A role for TNFRSF21 in AA was additionally supported by the observation of greater increase in TNFRSF21 expression in basal epithelial cells in adults with AA after allergen challenge compared with adults with allergy but not asthma and establishes a role for this gene in T2 inflammatory responses in human airways. Our genetic studies further suggest that rare variants in this gene contribute to this response and to the development of allergic disease and asthma in children.

Strengths of the study include the use of 2 study cohorts of children with whole-genome sequencing and comprehensive longitudinal phenotyping. Each study defined distinct groups of children on the basis of multiple measurements conducted over 1 (APIC) or over 10 (URECA) years, thereby adding precision beyond that achieved in cross-sectional observations. In addition, the study populations include a high percentage of Black and Hispanic children with high rates of allergy and asthma, populations that have been underrepresented in genetic studies.77 Despite these strengths, there are limitations. First, because of the modest sample size (n = 1035), we had low power to detect associations with individual rare variants and we were not well powered to detect genome-wide associations with asthma per se or with clinical subtypes of asthma. The lack of true (general population) controls also reduced power to detect associations with asthma or atopy. As a result, we had to limit our studies of asthma and atopy to 3 candidate genes that were put forward from our genome-wide studies of quantitative phenotypes. Second, the rare variant method we used (STAAr) does not report effect estimates or SEs, which limits interpretation of our results. Third, not all gene candidates identified in our studies (CTPB1-AS, TEX36-AS1) were available in phenomeXcan or the Mouse Genome Informatics resource, and not all phenotypes (total IgE, bronchodilator response, FEV1/forced vital capacity) were available in phenomeXcan, so it was not possible to validate all associations using these databases. Fourth, the genetic findings of this study have not been replicated in an independent cohort with whole-genome sequencing and similar quantitative phenotypes. Nonetheless, we were able to validate 3 of the most significant associations using orthogonal data from phenomeXcan and/or mouse knockout studies. Finally, the associations with AA phenotypes in APIC and URECA were only nominally significant. However, despite the small sample sizes, results in 2 independent cohorts highlighted the TNFRSF21 gene. A role for this gene in AA was further supported as an allergen-responsive gene in bronchial epithelial cells among adults with AA compared with adults with allergy but not asthma.26,51 These combined data underscore the robustness of the associations across biological contexts and provide convergent support for expression levels of TNFRSF21 impacting asthma-associated quantitative phenotypes and AA.

In summary, we identified rare variants in genes associated with allergic and inflammatory phenotypes using whole-genome sequencing data from children with diverse ancestries. We further validated 3 associations through external data sources, which converged on a role for rare variants in these 3 genes and, particularly in TNFRSF21, with asthma and atopy. Our study in well-characterized cohorts of children highlights the importance of rare variation in the development of asthma-associated quantitative phenotypes and AA phenotypes and identifies novel candidate genes that may serve as therapeutic targets.

DISCLOSURE STATEMENT

This work was supported by the National Institutes of Health (NIH) (grant nos. U19 AI62310, R01 HL104608, UG3/UL1 OD023282, HHSN272200900052C, HHSN272201000052I, UM1 AI114271, UM1 AI160040, DP2 CA247831, and UH2 AI4434); the Department of Defense (grant no. PR150903/W81XWH-16-1-0493), and funding from Sanofi and Regeneron. Site data collection was supported by the following NIH grants: RR00052 and UL1 TR001079 (Baltimore); M01 RR00533, UL1 RR025771, and 1 UL1 TR001430 (Boston); UL1 TR000150 (Chicago); UL1 TR000451 and UL1 TR001105 (Dalslas); UI1 RR025780 (Denver); UL1 TR000040, M01 RR00071, and UL1 TR001873 (New York); UL1 TR000075 (DC); and UL1 TR000077 (Cincinnati). S.M.C. was supported by NIH grant number T32 GM007197. J.A. was supported by NIH grant number KL2 TR002542. M.D. was supported by NIH grant numbers TL1 TR002388 and T32 HL07605. Single-cell sequencing studies were supported by the NIH (grant no. KL2 TR002452 to J.A., grant no. DP2 CA247831 to A.-C.V., and grant no. UH2 AI44434 to J.L.C.), and a DOD grant PR150903/W81XWH-16-1-0493 and Sanofi iAward (B.D.M.).

Disclosure of potential conflict of interest: All authors, with the exception of P. J. Gergen, report grants from the National Institutes of Health (NIH)/National Institute of Allergy and Infectious Diseases during the conduct of study. S. Clay, J. Alladina, N. P. Smith, C. M. Visness, M. Dasapas, M. Kattan, P. J. Gergen, and J. L. Cho have nothing to disclose outside the submitted work. W. W. Busse reports consulting fees from Novartis, GlaxoSmithKline, Genentech, Sanofi, AstraZeneca, and Regeneron and royalties from Elsevier outside the submitted work. M. A. Gill reports an honorarium for and support for travel to the 2017 AAAAI meeting during the conduct of study and monetary compensation from the American Academy of Pediatrics for her work teaching the biannual Pediatrics board review course, PREP The Course. G. K. K. Hershey reports grants from Adare during the conduct of the study. D. J. Jackson reports personal fees from Novartis, Pfizer, Regeneron, AstraZeneca, Sanofi, and Vifor Pharma, grants and personal fees from GlaxoSmithKline, and grants from NIH/National Heart, Lung, and Blood Institute, outside the submitted work. R. S. Gruchalla reports government employment from the Center for Biologics Evaluation and Research as well as personal fees from Consulting Massachusetts Medical Society, outside the submitted work. A. H. Liu reports personal fees from Phadia ThermoFisher as consulting honoraria, grants and nonfinancial support from ResMed/Propeller Health, nonfinancial support from Revenio, grants and personal fees from Avillion, and personal fees from Labcorp, outside the submitted work. L. B. Bacharier reports book royalties from Elsevier, consulting fees from Sanofi, Regeneron, Genentech, GlaxoSmithKline, DBV Technologies, Teva, Medscape, Kinaset, OM Pharma, and AstraZeneca, honoraria from Sanofi, Regeneron, and GlaxoSmithKline, participation in advisory board for DBV Technologies, AstraZeneca, and Vertex, leadership role in the American Academy of Allergy, Asthma & Immunology and the American Board of Allergy and
Immunology, and medical writing services for Sanofi/Regeneron. J. E. Gern reports consulting fees from AstraZeneca and Meissa Vaccines and 2 patents related to the methods to enhance the production of rhinoviruses and stock options with Meissa Vaccines. C. M. Kercsmar reports royalties from UpToDate. R. A. Wood reports grants from DBV, Aimmune, Regeneron, Genentech, Novartis, Food Allergy Research & Education (FARE), and Genentech and royalties from UpToDate. R. G. Robison received grant support from DBV Technologies and Aimmune Therapeutics. A.-C. Villani has a financial interest in 10X Genomics; the company designs and manufactures gene sequencing technology for use in research, and such technology is being used in this research. Dr Villani’s interests were reviewed by The Massachusetts General Hospital and Mass General Brigham in accordance with their institutional policies. B. D. Medoff receives research funding from and served as a consultant for Sanofi and Regeneron. C. Ober reports personal fees from the American Association of Asthma, Allergy and Immunology.

We acknowledge Dr Hae Kyung (Haky) Im for her advice and helpful discussions, the University of Chicago Center for Research Informatics for its cluster resources, and the coordinators and participants in all the included studies.

REFERENCES

34. Reich D, Nalls MA, Kaoo WHL, Akylbekova EL, Tandon A, Patterson N, et al. Reduced neutrophil count in people of African descent is due to a regulatory SNP. Nature 2021;599:628-34.


