Genome-wide association study of asthma, total IgE, and lung function in a cohort of Peruvian children

Ayobami T. Akenroye, MD, MPH,[a]* Tonya Brunetti, PhD,[b]* Karina Romero, MD, MSc,[c,d] Michelle Daya, PhD,[b] Kanika Kanchan, PhD,[e] Gautam Shankar, MS,[e] Sameer Chavan, MS,[b] Meher Preethi Boorgula, MS,[b] Elizabeth A. Ampleford, PhD,[f] Héllen Freitas Fonseca, MSc,[g] Gregory A. Hawkins, PhD,[h] Helena Mariana Pitangueira Teixeira, MSc,[g] Monica Campbell, MS,[b] Nicholas Rafaels, MS,[b] Alexandra Winters, PhD,[e] Eugene R. Bleecker, MD,[i] Alvaro A. Cruz, MD,[j] Mauricio L. Barreto, MD, PhD,[k] Deborah A. Meyers, PhD,[i] Victor E. Ortega, MD, PhD, ATSF,[f] Camila A. Figueiredo, PhD,[g] Kathleen C. Barnes, PhD,[b] William Checkley, MD, PhD,[c,l] Nadia N. Hansel, MD, MPH,[c]‡ and Rasika A. Mathias, ScD[e]‡

Baltimore, Md; Denver, Colo; Lima, Peru; Winston-Salem, NC; Tucson, Ariz; and Salvador, Brazil

Background: Genetic ancestry plays a role in asthma health disparities.

Objective: Our aim was to evaluate the impact of ancestry on and identify genetic variants associated with asthma, total serum IgE level, and lung function.

Methods: A total of 436 Peruvian children (aged 9-19 years) with asthma and 291 without asthma were genotyped by using the Illumina Multi-Ethnic Global Array. Genome-wide proportions of indigenous ancestry populations from continental America (NAT) and European ancestry from the Iberian populations in Spain (IBS) were estimated by using ADMIXTURE. We assessed the relationship between ancestry and the phenotypes and performed a genome-wide association study.

Results: The mean ancestry proportions were 84.7% NAT (case patients, 84.2%; controls, 85.4%) and 15.3% IBS (15.8%; 14.6%). With adjustment for asthma, NAT was associated with higher total serum IgE levels ($P < .001$) and IBS was associated with lower total serum IgE levels ($P < .001$). NAT was associated with higher FEV_1 percent predicted values ($P < .001$), whereas IBS was associated with lower FEV_1 values in the controls but not in the case patients. The HLA-DR/DQ region on chromosome 6 (Chr6) was strongly associated with total serum IgE (rs3135348; $P = 3.438 \times 10^{-10}$) and was independent of an association with the haplotype *HLA-DQA1*∼*HLA-DQB1*:04.01∼04.02 ($P = 1.55 \times 10^{-05}$). For lung function, we identified a locus (rs4410198; $P = 5.536 \times 10^{-11}$) mapping to Chr19, near a cluster of zinc finger interacting genes that colocalizes to the long noncoding RNA *CTD-2537I9.5*. This novel locus was replicated in an independent sample of pediatric case patients with asthma with similar admixture from Brazil ($P = .005$).

Conclusion: This study confirms the role of HLA in atopy, and identifies a novel locus mapping to a long noncoding RNA for lung function that may be specific to children with NAT. (J Allergy Clin Immunol 2021;■■■:■■■-■■■.)

Key words: *Asthma, immunoglobulin E, lung function, admixture, genome wide association analyses, Peru, ancestry, allergy*

Asthma is the most prevalent noncommunicable chronic disease in children and a major cause of emergency department visits, hospitalizations, and school absences.[1,2] African Americans and Hispanics are more likely to have severe asthma and worse asthma-related outcomes.[3] Indeed, asthma is a complex disease to which environmental and socioeconomic factors may contribute. Nonetheless, there is evidence that these factors do not completely explain these disparities and that genetics plays a role in asthma development, atopy, and pulmonary function.[4]

The genome-wide association study (GWAS) approach has been successful in identifying numerous loci/genes associated with asthma.[5] Recently, Pividori et al reported 60 unique loci

From [a]the Division of Pediatric Allergy and Immunology, [c]the Division of Pulmonary and Critical Care Medicine, [e]the Division of Allergy and Clinical Immunology, Johns Hopkins University School of Medicine, Baltimore; [b]the Division of Biomedical Informatics and Personalized Medicine, University of Colorado, Denver; [d]the A.B. PRISMA, Lima; [f]the Department of Internal Medicine, Center for Precision Medicine, Wake Forest School of Medicine, Winston-Salem; [g]Instituto de Ciências da Saúde, Universidade Federal da Bahia (UFBA), Salvador; [h]the Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem; [i]the Department of Medicine, University of Arizona, Tucson; [j]the Fundação ProAR, Salvador; [k]the Centro de Integração de Dados e Conhecimento para Saúde, Fiocruz, Salvador; and [l]the Department of International Health, Program in Global Disease Epidemiology and Control, Johns Hopkins Bloomberg School of Public Health, Baltimore.

*These authors contributed equally to this work.

‡These authors contributed equally to this work.

Supported by National Institute of Environmental Health Sciences, National Institutes of Health (grants R01ES018845 and R01ES018845-S1 and NHLBI R01 HL142992). A.T.A was supported by the National Institute of Allergy and Infectious Diseases T32 Research Training Grant in Pediatric Allergy and Immunology (grant no. 2T32AI007007-41) and the Johns Hopkins University Provost's Postdoctoral Award. K.R. was a Fogarty Global Health Fellow through the consortium comprising the University of North Carolina, Johns Hopkins University, Morehouse School of Medicine, and Tulane University during the conduct of this work (through grant 5R25TW009340) W.C. was supported by Pathway to Independence Award R00HL096955 from the National Heart, Lung, and Blood Institute, National Institutes of Health.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication June 8, 2020; revised February 20, 2021; accepted for publication February 24, 2021.

Corresponding author: Rasika Mathias, ScD, Division of Allergy and Clinical Immunology, Johns Hopkins University School of Medicine, 5501 Hopkins Bayview Circle, Baltimore, MD 21224. E-mail: rmathias@jhmi.edu. Or: Nadia Hansel, MD, MPH, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, 5501 Hopkins Bayview Circle, Baltimore, MD 21224. E-mail: nhansel1@jhmi.edu.

0091-6749/$36.00

https://doi.org/10.1016/j.jaci.2021.02.035

Abbreviations used
BMI: Body mass index
CAAPA: Consortium on Asthma among African Ancestry Populations
Chr: Chromosome
COPD: Chronic obstructive pulmonary disease
eQTL: Expression quantitative trait locus
GASP: Genetic Asthma Susceptibility to Indoor Pollution in Peru
GTEx: Genotype-Tissue Expression
GWAS: Genome-wide association analyses
IBS: Iberian populations in Spain
LD: Linkage disequilibrium
LncRNA: Long noncoding RNA
MAF: Minor allele frequency
MEGA: Multi-Ethnic Genotyping Array
NAT: Indigenous ancestry populations from continental America
PAMPAS: Pampas de San Juan de Miraflores
PC: Principal component
PCA: Principal component analysis
PP: Posterior probability
SNP: Single-nucleotide polymorphism
TGP: Thousand Genomes Project
VILLA: Villa El Salvador
YRI: Yoruban

associated with childhood onset of asthma in a cohort of individuals of European ancestry.[6] Genetic risk loci may however differ by ancestral population.[7] For instance, the Genes-environments and Admixture in Latino Americans (GALA II) study confirmed the existence of a strong association between the *ORMDL3* locus on 17q21 that was previously identified in European populations, but it identified a novel locus (*MUC22*) 6p21 in Latinos with asthma.[5] The Consortium on Asthma among African Ancestry Populations (CAAPA) also showed that the effect of the genetic variants at the *ORMDL3* locus differed according to the ancestral haplotype on which the variant was found.[8] Thus, the study of admixed populations such as those in Peru may provide opportunities to confirm previously identified loci in this specific population and also facilitate the identification of novel loci.

In the International Study of Asthma and Allergies in Childhood (ISAAC), Peru was one of the countries with the highest prevalence of childhood asthma, with a significant proportion of these children having severe disease.[9] Our study goal was to bridge the gaps stemming from a minimal representation of this population in asthma genetics interrogations to date. We have done so by determining the global estimates of ancestry for children with and without asthma from the Genetic Asthma Susceptibility to Indoor Pollution in Peru (GASP) study, and evaluating whether these global ancestry estimates are associated with asthma status, lung function, and total serum IgE level (IgE). Acknowledging our limited sample size, we also performed a GWAS for asthma, total serum IgE level, and lung function in this pediatric sample from Peru.

METHODS

Study participants and setting

We analyzed data from the GASP study, which evaluated the association between genetics, environment, and asthma status among children and adolescents residing in Peru. As described in a previous article,[10] case patients and controls were recruited from 2 adjacent communities in Lima: Pampas de San Juan de Miraflores (PAMPAS) and Villa El Salvador (VILLA) between 2011 and 2014. Subjects were eligible if they were 9 to 19 years of age. Children were considered to have asthma if they reported a physician diagnosis of asthma and asthma symptoms or taking asthma medications within the past year. Patients with other chronic respiratory conditions; pregnancy; current or past history of tuberculosis; history of hospitalization for cardiovascular disease in the preceding 3 months; and/or history of ocular, abdominal, or thoracic surgery in the past 3 months were excluded. Patients who were unwilling or unable to provide a blood sample were also excluded. The controls were children without asthma symptoms or the use of asthma medications in the past year and a normal FEV_1-to–forced vital capacity ratio and an FEV_1 value above 80%.

At baseline, questionnaires were completed by each child or caregiver. The questionnaire included demographic data, comorbidities (including other allergic diseases), and data on asthma control and severity. Baseline anthropometry and lung function were assessed, and predicted values and z scores were calculated by using multiethnic reference values derived by the Global Lung Function Initiative.[11] Total serum IgE level was measured by using the ImmunoCAP 250 system (ThermoFisher Scientific, Waltham, Mass); the levels in all samples were above the detection thresholds. For specific total serum IgE antibody testing, a level higher than 0.10 kUa/L indicated a positive total serum IgE antibody response to mixes of 3 common allergens (animal, mold, and dust mite). The institutional review boards at the Johns Hopkins University School of Medicine (Baltimore, Md) and AB PRISMA (Lima, Peru) approved this study; all subjects or parents provided written consent.

Genotyping and quality control for GASP samples

GASP subjects were genotyped by using the Illumina Multi-Ethnic Genotyping Array (MEGA), which was specifically designed to capture genetic variation in populations with a significant African and Native American contribution.[12] Genotyping plates were balanced by asthma status and sex. Duplicate sample concordance, HapMap concordance, and mendelian errors of HapMap trios were used as controls for each set of 91 samples plated.[13] Any unresolved sex mismatches and ambiguously imputed sex (defined as samples with an F statistic value between 0.20 and 0.65, as assessed using PLINK1.9) were removed.[14,15] Thereafter, we excluded samples with a genotyping rate less than 98.5%. We then excluded all samples with strong cryptic relatedness (PI_HAT > 0.3) and excess heterozygosity (±3 SDs from the mean).

Genotyped single-nucleotide polymorphisms (SNPs) that did not pass quality control procedures were removed by using 2 criteria: less than a 99% call rate and/or deviations from Hardy-Weinberg equilibrium ($P < 10^{-6}$). Ancestral outliers in the data set were evaluated by using a set of linkage disequilibrium (LD)-pruned SNPs in a combined data set of GASP and Thousand Genomes Project (TGP) samples. Principal component (PC) analysis (PCA) was performed by using King, GENESIS, and PC-AiR.[16] Study samples approximately 6 SD away from the mean of the Peru sample for PCs 1 and 2 were dropped. Approximately 15% of subjects were filtered out, resulting in a total of 743 samples (for the PAMPAS sampling site, n = 408; for the VILLA sampling site, n = 319) with high-quality genotype data (see Fig E1 in this article's Online Repository at www.jacionline.org) Before the analyses, we excluded 16 additional subjects on account of ambiguity in their case patient–control status.

Assessment of genetic ancestry: ADMIXTURE and PCA

In all, 3 ancestral reference populations were used in the ancestry deconvolution for the GASP samples: 107 TGP samples from the Iberian populations in Spain (IBS), 88 TGP Yoruban (YRI) samples, and 43 samples of indigenous ancestry populations from continental America (NAT).[17] The specific samples selected were those reported with no admixture themselves and included Bolivian Aymara (n = 25), Maya (n = 6), Mixtec (n = 5), Nahua (n = 1), Peruvian Quechua (n = 2), and Tlapanec (n = 4). We used IBS as the European population, given the Spanish introgression and gene flow patterns in Peruvians following the Spanish conquest.[18] SNPs were LD-pruned on the

basis of an SNP window size of 50, variant count per step of 5, and variance inflation factor of 2[14,15] on the combined data set of GASP and 3 reference populations (IBS, YRI, and NAT). The number of SNPs after LD pruning was 107,095. PCA was performed by using this set of 107,095 SNPs and including all 3 reference populations by using King, GENESIS, and PC-AiR.[16]

To obtain global estimates of ancestry, we used ADMIXTURE.[19] The first step was an unsupervised analysis conducted for clusters (K = 1-5 with cross-validation) to find the number of putative source populations in GASP alone. We found that a K value of 2 had the smallest cross-validation error. Data from GASP were then merged with the IBS and NAT reference populations on the basis of confirmation from the aforementioned PCA that these were the 2 major contributing reference populations.[17] Global estimates of admixture by using maximum likelihood estimates were then obtained by assuming a K value of 2 ancestral clusters.

Imputation of data from the TGP Pilot

Subsequent to the aforementioned SNP-based quality control, we removed SNPs with a minor allele frequency (MAF) less than .0001 and all ambiguous allele SNPs. Additionally, strand flips were resolved before imputation by using the Michigan Imputation Server against the TGP.[20] A total of 956,459 SNPs were used as input for imputation. Only imputed variants with an imputation quality (R^2) greater than 0.3 were used in the downstream association analysis; these included SNPs and short indels (insertion/deletion variants) returned from the imputation server. Additional filters on allele frequency are described in the following sections of this article.

Analysis of association between global genetic ancestry and asthma susceptibility, total serum IgE levels, and lung function

We tested for an association between the global estimates of ancestry and asthma, lung function, and total serum IgE. For these analyses, the global estimates of ancestry from ADMIXTURE were used for percentage of Utah residents with ancestry from Northern and Western Europe (CEU) and percentage of NAT. For asthma susceptibility, we used logistic regression, and for total serum IgE level and lung function, we used a linear model. All models included age, sex, and socioeconomic status. Asthma status was included as a covariate for analyses of lung function and total serum IgE level that combined case patients and controls. Other potential confounders were also included if they were associated with the phenotype in univariate models. For asthma, body mass index (BMI) was also included as a covariate and site in models combining the PAMPAS and VILLA sampling sites. For lung function, we used prebronchodilator FEV_1 percent predicted and included height, BMI, and site as additional covariates. The results were however similar with use of z scores. Total serum IgE level was log-transformed and additionally adjusted for site.

Statistical models for genetic association analysis

All tests for association were performed in the R package GENESIS, version 2.4.0[21,22] by testing each variant (genotyped and imputed) under an additive model. Primary GWASs were performed for (1) asthma, (2) log-transformed total serum IgE level ($\log_{10}$[IgE] value) in the combined sample of case patients and controls, and also (3) lung function in the combined sample by using covariates specified later in this article. Stratified analyses by case patient group and control group were also performed for any GWAS signals identified for total serum IgE level or lung function. Given our small sample size, a MAF of 5% or higher was applied universally for genotyped and imputed SNPs. As already described, any imputed SNP with an R^2 value less than or equal to 0.3 and any genotyped SNP with less than a 99% call rate, and/or deviations from Hardy-Weinberg equilibrium ($P < 10^{-6}$) was also discarded. Finally, the quality of the resulting data set used for the genome-wide analysis was confirmed by plotting Q-Q plots (see Fig E2 in this article's Online Repository at www.jacionline.org).

A logistic mixed effect model was used for asthma and linear mixed effects models for total serum IgE and lung function. All 3 models included GASP-specific PCs as covariates in addition to age, sex, and socioeconomic status. The PCs used in these tests for association were derived by using a total of 246,361 LD-pruned genotyped SNPs in the GASP Peruvian samples (ie, these PCs were calculated without any reference ancestral samples). This is a larger set of SNPs than that used in the aforementioned analysis including ancestral populations because there we had to limit the starting set of SNPs to those that overlapped between the MEGA array and were used for the reference 43 NAT reference population[17]; here we were able to perform LD pruning on the full set of SNPs from the MEGA array. The first 20 PCs generated by using King, GENESIS, and PC-AiR with reference[16] were visually examined on a scree plot, with the first 4 PCs identified as accounting for the most variance in the data set, and used as covariates in association analysis models. Site was not included as a covariate, as ancestry differences were addressed with the inclusion of the PCs. BMI was also included as a covariate for asthma and lung function. For total serum IgE level and lung function, asthma was included as a covariate to account for any possible association with asthma rather than with the total serum IgE level or lung function phenotype. For any identified associations, we further performed stratified analyses within the case patients and controls separately. All data cleanup, Manhattan, and Q-Q plots were generated by using custom written scripts and the R package qqman, version 0.1.[23]

We implemented standard GWAS thresholds for discovery ($P < 5 \times 10^{-8}$) and suggestive evidence ($P < 1 \times 10^{-5)}$ for each of the 3 phenotypes. Additionally, we sought to replicate the 60 significant childhood asthma loci identified by Pividori et al[6] within extracted flanking regions of plus or minus 0.4 Mb (similar to the scale of Pividori et al[6]). We used 2 significance thresholds for these lookups: the first was a simple correction for number of loci tested, which assumes a single causal variant per locus ($P < .05/60$), and the second was a correction for the number of independent SNPs (at an Rsq $R^2 = 0.7$) tested across the 60 loci ($P < .05/1,599$).

Colocalization analysis of GWAS signals and cis-eQTLs

Genotype-Tissue Expression (GTEx) analysis V7 (dbGap accession no. phs000424.v7.p2) expression quantitative trait locus (eQTL) results were downloaded from the GTEx portal for lung and whole blood tissues; a false discovery rate of 0.05 or less was used to ascertain the significant transcripts. We performed pairwise colocalization analysis of GWAS signals with cis-eQTL data by using the R-package coloc.[24,25] The method usef approximate Bayes factor computations and tests pairwise colocalization of SNPs in GWAS data set with eQTLs. It generates 5 posterior probabilities (PPs), namely, PP0, PP1, PP2, PP3 and PP4, for the locus by using the evidence for competing hypotheses of either no colocalization or colocalization.[24] A PP3 value of approximately 75% indicated evidence against colocalization. In contrast, a PP4 value of approximately 75% supported evidence of colocalization; therefore, the first step was to find all genes with a PP3 value less than 75% and a PP4 value of approximately75% and to then examine the PP for each SNP within the region for the likely causal variant.

Imputation of HLA alleles, haplotypes, and amino acids

HLA alleles for the HLA class I genes *HLA-A*, *HLA-B*, and *HLA-C* and the HLA class II genes *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, and *HLA-DRB1* were imputed by using the R package HIBAG v1.3[26,27] with use of the attribute bagging method to impute HLA alleles by using hg19 genome assembly and Illumina Infinium Multi-Ethnic Global BeadChip prediction model with 2-field (4-digit) resolution. We filtered out all the genes with a call rate lower than 95%, and analysis was limited to those alleles with a frequency approximately 5% for each passing gene. Each allele was then analyzed under an additive model in HIBAG for asthma, log-transformed total serum IgE level, and lung function by incorporating the same covariates as used in the aforementioned GWAS analysis.

HLA haplotypes were generated across *HLA-DQA1* and *HLA-DRB1* by using the BIGDAWG R package.[28] Only samples with complete haplotype data (ie, those with alleles at both genes) were retained for analysis, and the analysis was limited to those alleles with a frequency of approximately 5%. Each haplotype was then analyzed under an additive model in Plink by using the exact analyses models for asthma, $\log_{10}$(IgE) value, and lung function stated earlier in this article for the SNP association analysis; BIGDAWG could not accommodate covariates and quantitative traits. Additional analyses were performed within Plink to facilitate a conditional analysis to evaluate independence of the identified SNP from the HLA alleles/haplotypes.

Replication of Chr19 locus for lung function

Replication for the genome-wide signification association on Chr19 with lung function (FEV_1 percent predicted) in GASP was assessed in 4 studies: (1) ProAR (Program for Control of Asthma and Allergic Rhinitis), (2) SCAALA (Social Changes, Asthma and Allergies in Latin America) Program, (3) Hispanics participants randomized to a long-acting β-agonist and inhaled corticosteroids with or without rescue long-acting β-agonist and inhaled corticosteroid combination therapy for 6 months in the AstraZeneca-sponsored COMPASS trial,[29] and (4) a large published GWAS of lung function in 400,102 individuals of European ancestry from the UKBiobank and the SpiroMeta Consortium.[30] A total of 12 SNPs with a discovery *P* value less than 5E–08 were examined for replication in each of these data sets. No correction was made, as these 12 SNPs represent a single association signal in GASP.

The ProAR severe asthma case-control study[31] was carried out in the city of Salvador, Bahia, Brazil, in 2013. Adult individuals (aged >18 years) with no other lung diseases were recruited, and a total of 1065 individuals with FEV_1 percent predicted data available were included in the analysis. ProAR DNA samples were genotyped on the Illumina MEGA array and imputed to the CAAPA reference panel on the Michigan imputation server. Association tests were performed by using PLINK and included age, sex, and the first PC as covariates. The SCAALA cohort was used first to assess risk factors for asthma and allergies in children and adolescents aged 4 to 11 years who were living in the city of Salvador and later to study the genetics determinants of such conditions. For this replication, we included 947 unrelated children with available FEV_1 percent predicted data.

SCAALA DNA samples were genotyped by using the commercial panel 2.5 HumanOmni Beadchip, which is available from Illumina (www.illumina.com), and imputed to the 1000 Genomes Project (TGP) reference panel on the Michigan imputation server. Association tests were performed by using PLINK and included age, sex, and the first 4 PCs as covariates.

The COMPASS trial participants were genotyped on the Illumina Human OmniExpress-12v1 chip and imputed to the TOPmed freeze 5 reference panel on the Michigan imputation server. Pulmonary function was assessed from data obtained during the run-in period of the COMPASS trial in 249 individuals with asthma from Argentina and 312 individuals from Mexico, of whom 471 were adults (>18 years of age). Compass samples were genotyped on the Illumina Human OmniExpress-12v1 chip and imputed to the TOPMed freeze 5 reference panel on the Michigan imputation server reference panel on the Michigan imputation server. Association tests were performed by using PLINK and included age, sex, and the first 4 PCs as covariates (n = 947).

Summary statistics from the UK Biobank and the SpiroMeta Consortium GWAS were obtained from the GWAS catalog (ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/ShrineN_30804560_GCST007432/Shrine_30804560_FEV1_meta-analysis.txt.gz), and the Chr19 variants associated with lung function in GASP were extracted and assessed for replication.

RESULTS

Study subject characteristics

The characteristics of participants at enrollment are shown in Table I. The cohort members from both sites were similar in age, sex, total serum IgE level, FEV_1 percent predicted value, and distribution of the 2 ancestral populations. Compared with the controls, the case patients included a higher proportion of males (56% vs 47%) and they had higher total serum IgE levels (1233 vs 652 kU/L); case patients from both sites had a significantly lower FEV_1 percent predicted value (14% vs 118% [$P = .003$]). Fig 1, *A* shows the ADMIXTURE global genetic ancestry proportions among the study cohort by case patient–control status and sampling site, allowing for 2-way admixture. The 2-way admixture is also evident from the PCA analysis in Fig 1, *B*, which shows the distribution of the GASP samples relative to 3 ancestral reference groups with variability due largely to NAT and IBS contribution. On average, the global genetic ancestry in the study subjects was 15% European (IBS) and 85% indigenous ancestry. Differences were observed between the 2 sites (Table I and Fig 1, *C*), with higher indigenous and lower European ancestry in the PAMPAS sampling site (86% and 14%, respectively) than in the VILLA sampling site (82% and 18%, respectively) ($P \leq .001$ and $P < .001$, respectively).

Association of genetic ancestry with asthma susceptibility, total serum IgE level, and lung function

We did not find any significant association between indigenous ancestry or European ancestry and asthma susceptibility (see Table E1 in this article's Online Repository at www.jacionline.org). European ancestry was associated with a lower FEV_1 value (percent predicted and *z* score [FEV_1 percent predicted, $\beta = –0.192$; $P = .001$]; given 2-way admixture, the exact opposite was noted for indigenous ancestry ($\beta = 0.192$; $P = .001$) (see Table E2 in this article's Online Repository at www.jacionline.org). In analyses stratified by asthma status, these associations between ancestry and lung function remained statistically significant in the controls but not in the case patients; although the direction of effect was the same between the 2 groups, the magnitude of the effect was larger in the controls (for NAT, $\beta = 0.272$; $P = .001$) than in the case patients (for NAT, $\beta = 0.111$; $P =$ not significant). As seen in Table E2, European ancestry was associated with lower total serum IgE levels ($\beta = –0.010$; $P < .001$), and this effect was highly consistent between the case patients with asthma ($\beta = –0.009$; $P = .004$) and the controls ($\beta = –.025011$; $P = .02$). Once again, given 2-way admixture, the exact opposite was observed for indigenous ancestry. Given the observation (in Fig 1, *B*) of some African contribution to a small set of samples, a sensitivity analysis was performed by using a K value of 3 in ADMIXTURE. We were able to confirm the robustness of our results with a K value of 2: there was virtually no change in the NAT estimation (ie, any YRI ancestry was absorbed into the IBS ancestry), and the described patterns with NAT remain unchanged.

Genome-wide association analyses

Manhattan plots and Q-Q plots of the primary GWAS for asthma, total serum IgE level, and lung function are shown in Fig 2 and Fig E2 (in this article's Online Repository at www.jacionline.org) ($\Lambda = \sim 1$), respectively.

Asthma status. We did not find any variants that met GWAS significance ($P < 5 \times 10^{-8}$) in the asthma GWA analysis (Fig 2, *A*). However, there were 13 variants spanning 9 loci that met the suggestive GWAS threshold ($P < 1 \times 10^{-5}$); 12 of 13

TABLE I. Baseline participant characteristics by site and case-control status

Characteristic	PAMPAS sampling site			VILLA sampling site		
	Case patients (n = 256)	Controls (n = 152)	*P* value	Case patients (n = 180)	Controls (n = 139)	*P* value
Demographic characteristic						
Age (y), mean (SD)	13.5 (2.7)	13.8 (2.6)	.28	13.3 (2.6)	12.8 (2.7)	.07
Male, no. (%)	151 (59)	73 (48)	.03	94 (52)	65 (47)	.33
SES (no.), mean (SD)*	-0.35 (1.6)	-0.81 (1.6)	.005	0.92 (1.4)	0.38 (1.5)	.001
Total serum IgE level, kU/L (%)†	1202 (1523)	599 (758)	<.001	1257 (1498)	658 (1092)	<.001
Atopic disease, no. (%)‡						
Eczema	36 (14.6)	19 (13.2)	.69	17 (10.3)	3 (2.2)	.005
Rhinitis	21 (15.2)	2 (3.7)	.03	21 (20.4)	1 (3.1)	.02
Ancestry (no.), mean (SD)						
Indigeneous (NAT)	85.7 (9.2)	86.8 (9.7)	.22	82.0 (10.3)	83.9 (9.6)	.10
European (IBS)	14.3 (9.2)	13.2 (9.7)	.22	18.0 (10.3)	16.1 (9.6)	.10
Clinical characteristic						
BMI (ng/m^2), no. (%)						
Underweight/normoweight§	146 (57.0)	102 (67.1)	.10	77 (42.8)	83 (59.7)	.01
Overweight	77 (30.1)	32 (21.1)		66 (36.7)	37 (26.6)	
Obese	33 (12.9)	18 (11.8)		37 (20.6)	19 (13.7)	
Baseline FEV_1 (% predicted), mean (SD)	114.1 (14.5)	118.4 (13.8)	.003	113.6 (15.6)	119.2 (13.9)	<.001
Baseline FEV_1 (raw in liters), mean (SD)	2.86 (0.82)	3.03 (0.80)	.04	2.86 (0.78)	2.80 (0.77)	.48
Baseline FEV_1 (z score), mean (SD)	1.22 (1.25)	1.58 (1.20)	.004	1.18 (1.34)	1.66 (1.22)	<.001
Asthma severity, no. (%)						
Mild intermittent	40 (20.4)	N/A	N/A	31 (22.6)	N/A	N/A
Mild persistent	82 (41.8)	N/A		64 (46.7)	N/A	
Moderate persistent	41 (20.9)	N/A		23 (16.8)	N/A	
Severe persistent	33 (16.8)	N/A		19 (13.9)	N/A	
Sensitivity to environmental allergens, no. (%)						
House dust mix (*Der pter, Der far, Bla g*)	195 (76)	88 (58)	.001	132 (73)	76 (55)	<.001
Mold and yeast mix (*Pen, Clad, Asp, Candida, Alternaria, Setomelanomma*)	154 (62)	49 (35)	<.001	76 (42)	35 (25)	.001
Animal and epidermal mix (*Can f, Fel d, Mus m1, rat, guinea pig*)	118 (46)	30 (20)	<.001	79 (44)	28 (20)	<.001
Exhaled nitric oxide (ppb), median (IQR)	37.1 (34.9)	21.7 (24.6)	<.001	44.2 (41.0)	24.3 (31.1)	<.001

Asp, *Aspergillus*; *Bla g*, *Blattella germanica*; *Can f*, Canus familiaris; *Clad*, Cladosporium; *Der far*, *Dermatophagoides farinae*; *Der pter*, *Dermatophagoides pteronyssinus*; *Fel d*, *Felinus domesticus*; *IQR,* Interquartile range; *N/A*, not available; *SES*, socioeconomic status.

*SES score is based on a PCA that includes 12 household assets, parental level of education, and number of persons in the household. Higher number denotes higher SES.

†All total serum IgE levels in the samples were above the level of assay detection.

‡Eczema (390 subjects in the PAMPAS sampling site and 300 in the VILLA sampling site [690 of 727 subjects]) and rhinitis (192 in the PAMPAS sampling site and 135 in the VILLA sampling site [327 of 727 subjects]); data collected on only a subset of patients.

§Between both sites, only 3 individuals fell into the underweight category.

mapped to intronic regions, whereas 1 SNP was located in an intergenic region, with the nearest gene being *IP6K2*. (see Table E3 in this article's Online Repository at www.jacionline.org).

Total serum IgE level. Association analysis for total serum IgE resulted in 30 variants that met the GWAS threshold; all of them were located on Chr6 in the HLA region closest to *HLA-DQA1* (Table II and Fig 2, *B*). The most significant SNP in the region was rs3135348 ($P = 3.438 \times 10^{-10}$). An additional 414 variants in the same HLA region also met the suggestive GWAS threshold (see Table E3). The region spans a broader range of HLA genes containing *HLA-DR/DQ* genes. There were an additional 68 variants that met the suggestive GWAS threshold outside of the HLA region, primarily spanning Chr1, Chr2, Chr16, and Chr17 (see Table E3).

Lung function: FEV_1. We identified 12 variants that met the GWAS threshold of significance ($P < 5 \times 10^{-8}$). All of these SNPS were located on Chr19, near zinc finger interacting genes (Table II and Fig 2, *C*). The SNP with the most significant *P* value ($P = 5.536 \times 10^{-11}$) was rs4410198, which was located in *LOC107985322*, a noncoding transcript near the zinc finger interacting genes. There were an additional 118 variants that met the suggestive GWAS threshold of significance ($P < 1 \times 10^{-5}$) outlined in Table E3.

Overview of the Chr6 and Chr19 loci showing significant association with total serum IgE and lung function, respectively. To further investigate the 2 GWAS loci identified for total serum IgE level and lung function, we performed separate stratified analyses within the groups of case patients with asthma and controls and plotted the variant in each region by using Locus Zoom.[32] The peak SNP for total serum IgE level in the analysis including all subjects (with adjustment for asthma status), rs3135348, was located at position Chr6:32,394,098 (Fig 3, *A* [*left*]). We observed evidence for association at rs3135348 in the control analysis (n = 291 [Fig 3, *C* (*left*)]; β = –0.2588; $P = 5.448 \times 10^{-5}$) and a slightly stronger signal in the asthma-only analysis (n = 436 [Fig 3, *B* (left)]; β = –0.2020; $P = 4.257 \times 10^{-6}$). rs3135348 was a significant eQTL for a large number of HLA genes across a variety of tissues in the GTEx data (*HLA-DQA1, HLA-DQA2,*

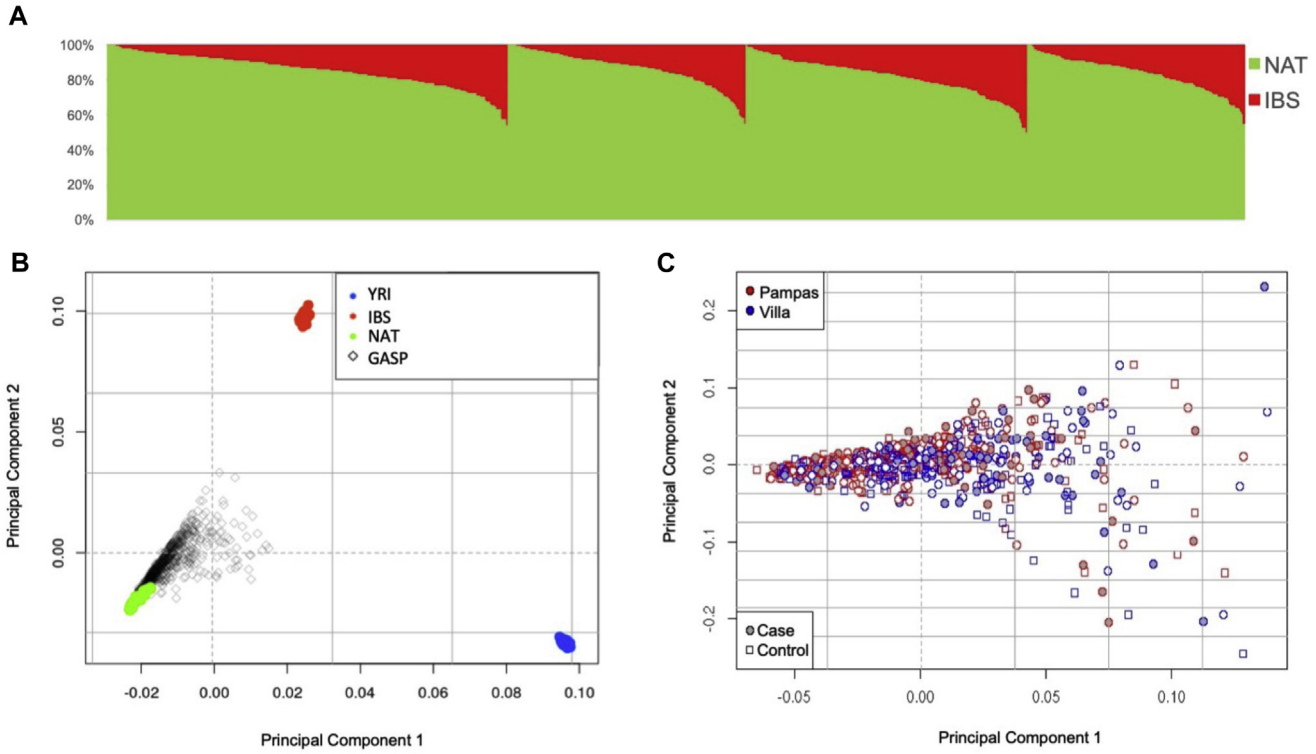

FIG 1. Summary of ancestry de-convolution in the GASP study subjects. **A,** Global admixture estimates by asthma case–control status and by PAMPAS and PAMPAS sampling sites, as estimated by using ADMIXTURE. Ancestry was estimates by using 107 TGP Iberian samples from Spain (IBS [*red*]) and 43 NAT samples (*green*) as reference populations. **B**, PCs for the Peruvian GASP samples, including 88 African samples (YRI [*green*]), 107 TGP Iberian samples from Spain (IBS [*red*]), and 43 NAT samples (*green*) as reference populations. **C,** PC analysis for the GASP study subjects alone showing the first 2 components by sampling site and by asthma case patient–control status.

HLA-DQB1, HLA-DQB1-AS1, HLA-DQB2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-DRB6, and *HLA-DRB9*).[33]

The peak SNP for lung function, rs4410198, was located at position Chr19:56,122,538 (Fig 3, *A* [*right*]). At this locus, the strength of the association was high in the asthma-only analysis (n = 436 [Fig 3, *B* (*right*)]; β = 6.870; $P = 1.547 \times 10^{-8}$) and was found to a lesser degree in the nonasthma analysis (n = 291 [Fig 3, *C* (*right*)]; β = 4.514; *P* = .00057). Formal colocalization analysis was performed for this novel locus to identify a potential target gene.

Replication of prior asthma GWAS

We compared the results from our GWAS to what is to date the most comprehensive asthma GWAS focused on childhood asthma (ie, that by Pividori et al[6]). The authors of that study identified 61 significant asthma loci by using UK Biobank data from both children and adults. Because GASP is a pediatric cohort, we focused on the 60 loci identified exclusively from childhood-onset asthma or shared between children and adults. We selected variants from our asthma association analyses that overlapped with each loci and highlighted these on a Manhattan plot (Fig 2, *A*), and we used 2 significance thresholds for these lookups: the first was a simple correction for number of loci tested, and the second was a correction for the number of independent SNPs tested across the 60 loci. Although none of the variants within the 60 known loci reached GWAS significance, we were able to replicate 2 loci at the locus-corrected threshold (Fig 2, *A* [*green line*]). Both loci identified, namely, 12q13.2 and 17q12, are exclusively childhood related (Fig 2, *A* [*red dots*]). The peak SNP on Chr12 was rs12578859 (*P* = .0001), and on Chr17, it was rs12450091 (*P* = .0002); neither of these cross the threshold correcting for number of independent SNPs tested.

Next, we investigated the overlap of the 60 loci with our results for total serum IgE level association analysis. One shared locus (ie, between adult and childhood asthma), namely, 6p21.32, overlapped the same *HLA-DR/DQ* gene region that we reported as meeting GWAS level of significance (Fig 2, *B*) for total serum IgE level. There were an additional 4 loci (ie, 2p25.1, 6p21.33, 7p15.1, and 18q21.33) that were replicated when the locus-corrected threshold was used. The 6p21.33 locus spans a broader HLA region that includes both *HLA-B* and *HLA-C* genes. The peak SNP on Chr2 was rs57838855 (*P* = .0005), on Chr7 the peak SNP was rs6962289 (*P* = .0004), at 6p21.33 it was rs9378247 (*P* = .0002), and on Chr18 it was rs56173102 (*P* = 1.82E-05). Only rs56173102 met the threshold correcting for number of independent SNPs tested, and it is interesting to note that this was the only childhood-specific locus from the prior published GWAS; the other 3 were shared with adult and childhood asthma loci.

Using the same approach for lung function association analysis, we found a single locus, 8q24.21, replicated at the locus-corrected threshold (Fig 2, *C*). The peak SNP was

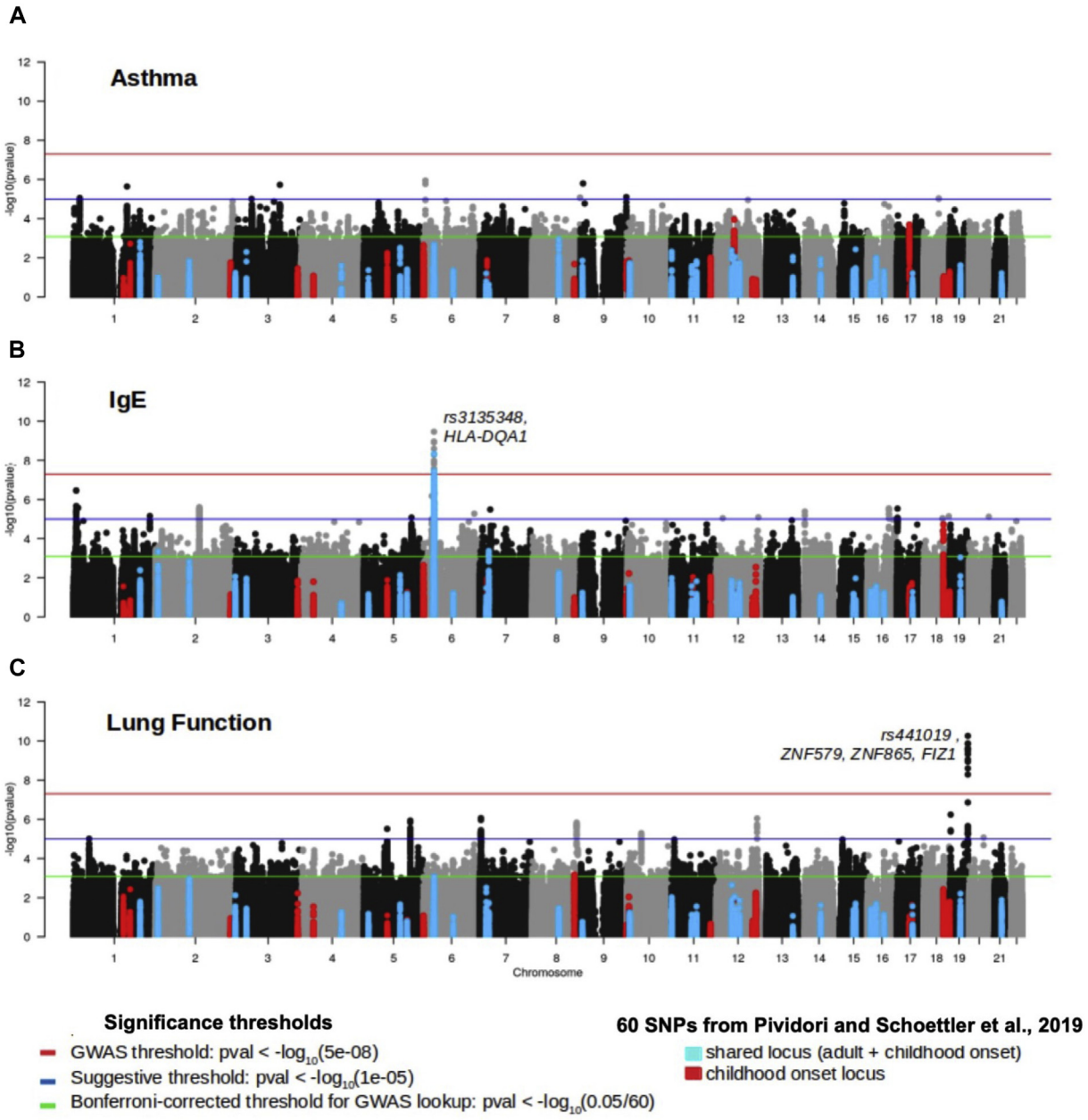

FIG 2. Genome-wide Manhattan plots for asthma (**A**), IgE level (**B**), and lung function (**C**). Plots are annotated on the basis of prior signals for childhood-onset asthma (unique or shared with adult-onset asthma) from 60 different loci (defined as ±0.4 Mb from the reference SNP values in Table 2 of Pividori et al[6]); 3 levels of statistical significance are shown.

rs4645958 (P = .0007), and it did not meet the threshold correcting for number of independent SNPs tested.

Finally, because the Pividori et al[6] GWAS is a study of subjects who are largely European in ancestry and may not be representative of genetic contributions to asthma and its related phenotypes in this Peruvian cohort of predominantly indigenous ancestry, we looked up a set of additional variants identified from the non-European samples within the EVE[34] and CAAPA[8] consortia. We were able to replicate only 1 variant, namely, rs335016 (P = .003), for asthma; however, the effect size was not in the same direction (see Table E4 in this article's Online Repository at www.jacionline.org).

Chr19 region colocalization with cis eQTLs of CTD-2537I9.5 in lung tissue

We jointly analyzed the GWAS data with cis-eQTLs in lung and whole blood tissues from GTEx. We identified 37 significant transcripts in lung tissue within an approximately 1-Mb region of the peak SNP for lung function, rs4410198. Among these transcripts, only the long noncoding RNA (LncRNA) *CTD-2537I9.5* showed strong evidence of colocalization (PP3 = 0.02 and PP4 = 98% [see Table E5 in this article's Online Repository at www.jacionline.org]**).** The top colocalizing SNP, namely, rs34164618, had a GWAS association P value (P = 2.44E–710), eQTL association P value (P = 3.47E–06),

TABLE II. Genetic variants passing the GWAS threshold ($P < 5 \times 10^{-8}$) for any of the 3 phenotypes of asthma, total serum IgE level, and lung function

rsID	Chr	hg19 position	Ref/Alt	MAF	Genotyped	R^2	Nearby gene(s)	Asthma		$\log_{10}$ (total serum IgE)		Lung function	
								OR	*P* value	Effect size	*P* value	Effect size	*P* value
rs6913471	6	32339925	T/A	0.273	Imputed	0.955		0.965605416	.7811	0.215	**4.93E–08**	–1.099	.2016
rs4496841	6	32389997	C/T	0.719	Imputed	0.988		0.854704059	.1852	–0.209	**1.39E–08**	1.363	.08916
rs3135359	6	32390578	T/C	0.736	Imputed	0.995		0.878973966	.2832	–0.218	**5.33E–09**	1.147	.1599
rs9296027	6	32393062	C/G	0.241	Imputed	0.995		1.093080656	.4805	0.218	**2.74E–08**	–1.303	.1278
rs9469109	6	32393161	A/T	0.241	Imputed	0.995		1.094174284	.4793	0.218	**2.73E–08**	–1.304	.1275
rs3135348	6	32394098	A/G	0.687	Imputed	0.966		0.922193691	.4957	–0.229	**3.44E–10**	1.464	.06675
rs9501400	6	32394184	G/A	0.247	Imputed	0.996		1.112934254	.398	0.221	**1.50E–08**	–1.433	.09281
rs9501622	6	32394251	T/A	0.241	Imputed	0.998		1.095269005	.4695	0.218	**2.64E–08**	–1.311	.1254
rs732163	6	32394911	G/A	0.24	Imputed	1.000		1.11516235	.3907	0.218	**2.70E–08**	–1.213	.1563
rs4959100	6	32397813	C/T	0.237	Imputed	1.000		1.122995872	.3612	0.216	**4.63E–08**	–1.358	.1152
rs9469110	6	32398525	G/T	0.237	Genotyped	1.000		1.122995872	.3613	0.216	**4.64E–08**	–1.358	.1151
rs3129854	6	32398781	G/C	0.719	Imputed	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
NA	6	32398853	T/TA	0.719	Imputed	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
rs984778	6	32400088	C/T	0.719	Genotyped	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
rs9501626	6	32400344	C/A	0.237	Genotyped	1.000		1.122995872	.3613	0.216	**4.64E–08**	–1.358	.1151
rs3135338	6	32401217	C/T	0.719	Genotyped	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
rs3135336	6	32401829	G/A	0.72	Imputed	0.995		0.860707976	.2091	–0.225	**1.16E–09**	1.238	.1254
rs3135335	6	32401845	C/G	0.72	Imputed	0.995		0.860707976	.2091	–0.225	**1.16E–09**	1.238	.1254
rs2027856	6	32402705	G/A	0.237	Genotyped	1.000		1.122995872	.3613	0.216	**4.64E–08**	–1.358	.1151
rs3135397	6	32403941	A/T	0.719	Imputed	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
rs3129866	6	32404065	G/C	0.719	Imputed	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
rs3129867	6	32404220	G/C	0.723	Imputed	0.999		0.884263663	.3042	–0.22	**2.57E–09**	1.243	.1233
rs2395173	6	32404859	A/G	0.719	Genotyped	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
rs3135395	6	32405192	T/G	0.719	Genotyped	1.000		0.862431115	.2136	–0.224	**1.16E–09**	1.231	.1263
rs2395178	6	32405362	G/C	0.719	Imputed	1.000		0.862431115	.2138	–0.224	**1.16E–09**	1.231	.1263
rs3129869	6	32405671	A/C	0.72	Genotyped	0.999		0.863293977	.2186	–0.224	**1.09E–09**	1.237	.1244
rs3129871	6	32406342	A/C	0.709	Imputed	0.999		0.832768156	.1245	–0.211	**1.05E–08**	1.153	.1516
rs9271364	6	32586787	A/G	0.689	Imputed	1.000		0.923116346	.5037	–0.201	**4.97E–08**	0.497	.5375
rs9272518	6	32606446	G/T	0.495	Imputed	0.701	HLA-DQA1	1.044982355	.7238	–0.228	**4.78E–09**	1.725	.04204
rs9273395	6	32627094	C/T	0.634	Imputed	0.936		0.88603396	.3118	–0.206	**3.47E–08**	1.365	.09296
rs4335869	19	56085656	T/A	0.297	Imputed	0.872		1.065026839	.6251	–0.022	.5873	5.489	**1.33E–10**
NA	19	56086758	GA/GAA	0.328	Imputed	0.848		1.030454534	.8187	–0.023	.5706	5.433	**1.43E–10**
rs28699417	19	56087272	T/C	0.323	Imputed	0.887		1.034584607	.7907	–0.021	.6048	5.279	**2.54E–10**
rs28379489	19	56087281	T/A	0.323	Imputed	0.887		1.035619709	.7818	–0.021	.6044	5.276	**2.65E–10**
NA	19	56088110	C/CA	0.249	Imputed	0.593		1.09089668	.6007	–0.025	.6331	6.372	**5.19E–09**
rs12972695	19	56088487	A/G	0.33	Imputed	0.892		1.010050167	.9379	–0.026	.5089	5.144	**4.73E–10**
rs10403008	19	56089947	C/G	0.307	Imputed	0.903	ZNF579	1.057597684	.6543	–0.009	.8208	5.211	**3.39E–10**
rs34164618	19	56090076	G/T	0.306	Imputed	0.903	ZNF579	1.071436209	.5838	–0.009	.8128	5.271	**2.44E–10**
rs12609355	19	56105932	G/A	0.3	Imputed	0.904	FIZ1	1.108491409	.4255	–0.014	.7245	5.195	**1.14E–09**
rs3803890	19	56110700	G/A	0.288	Imputed	0.879	FIZ1	1.133148453	.3485	–0.008	.8501	5.369	**8.98E–10**
rs4410198	19	56122538	G/A	0.296	Imputed	0.832	ZNF865	1.087628894	.5364	0.001	.9873	5.83	**5.54E–11**
rs146619376	19	56127441	C/G	0.277		0.854	ZNF865	1.068226717	.626	–0.006	.8831	5.351	**2.51E–09**

Alt, Alternative; *NA*, not available; *OR*, odds ratio; *Ref*, reference; *rsID*, rs identifier.
A strong association with the previously implicated HLA-DR/DQ region on Chr6 with total serum IgE level was identified, as was a novel locus mapping to chromosome 19, near zinc finger interacting genes for lung function. Boldface indicates significance at the $P < 5 \times 10^{-8}$ threshold.

and a 50% PP (see Fig E3 in this article's Online Repository at www.jacionline.org). A total of 30 significant transcripts were identified in whole blood tissue; however, none of them showed evidence of colocalization (see Table E5). Similar analysis was done for the peak SNP for total serum IgE levels, namely, rs3135348, in both tissues, but no evidence of colocalization was observed for any significant transcript.

Testing for association with HLA alleles imputed for the Chr6 region

Given that the HLA region is one of the strongest genetic loci across atopic diseases,[35] we imputed HLA alleles for class I and II genes. Table E6 (in this article's Online Repository at www.jacionline.org) shows the genes and alleles that were imputed in these data. There were 14 alleles with a frequency of 5% or higher, which were tested individually for the 3 primary phenotypes as shown in Table E7 (in this article's Online Repository at www.jacionline.org). Similar to the associations at the SNP level, the strongest associations were observed for IgE level. With use of a multiple-testing threshold of .05/(14*3), 3 alleles had significant associations with $\log_{10}$(IgE) value: *HLA-DQA1**04:01 ($P = 4.47 \times 10^{-5}$), *HLA-DQB1**03:02 ($P = 2.13 \times 10^{-4}$), and *HLA-DQB1**04:02 ($P = 4.34 \times 10^{-5}$). These allelic associations were largely represented by the haplotypes across the *HLA-DQA1* and

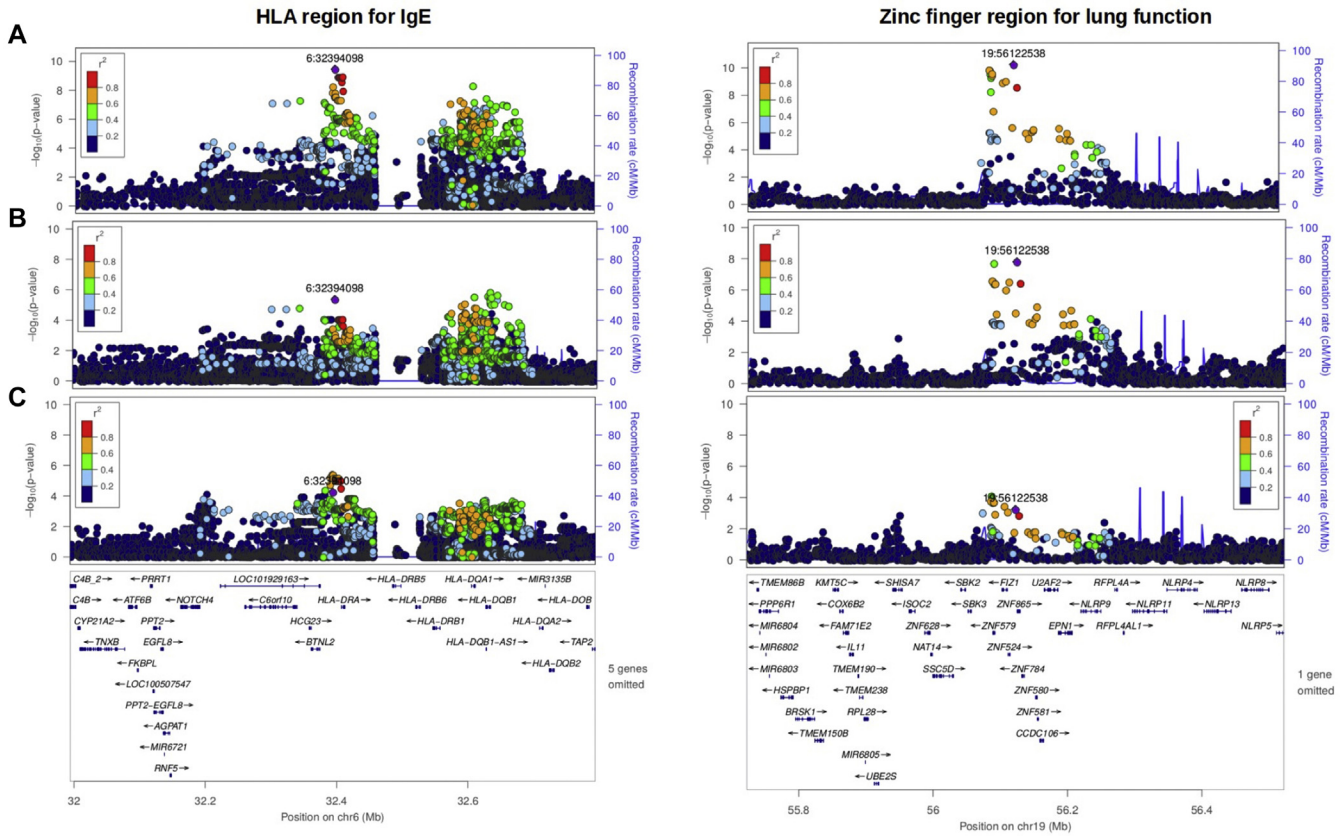

FIG 3. Locus zoom plots of the HLA region (*left*) and zinc finger region (*right*) that passed GWAS significance for total serum IgE level and lung function, respectively. LD between the peak SNP and other variants within the region is color-coded and calculated on the basis of 727 specific GASP samples used in the analysis. Variants are filtered in a minor allele count greater than 10 and a MAF of 0.05 or higher. To facilitate examination of the association signal by case patient–control status, we used full analysis based on all 727 adjusted for asthma (**A**), analysis limited to case patients with asthma (**B**), and analysis based on the asthma-free controls (**C**). The purple diamond represents the reference SNP and is defined on the basis of the peak SNP in the combined analysis from (**A**).

HLA-DQB1 genes (*HLA-DQA1*∼*HLA-DQB1*):04.01∼04.02 ($P = 1.55 \times 10^{-5}$). Given the strong LD that has been well documented in this HLA region, we were unable to separate the independence of the association between the peak SNP in this region for $\log_{10}$(IgE) value and the final haplotype *HLA-DQA1*∼*HLA-DQB1*:04.01∼04.02. Even after adjustment for the effect of the haplotype, the peak variant from the GWAS, namely, rs3135348, retained its significance, although it was somewhat reduced (see Table E7)

Replication of the novel Chr19 locus for lung function

Replication evidence for the Chr19 locus for lung function and the 4 independent data sets is presented in Table E8 (in this article's Online Repository at www.jacionline.org). Significant replication is noted at 7 of the SNPs within the pediatric SCAALA study, with the strongest evidence at rs12609355 ($P = .005$) and rs10403008 ($P = .008$) and with consistent directions of effect between the discovery ($\beta = 5.195$) and replication ($\beta = 2.028$) data. No evidence for replication was noted in the adult PROAR cohort, the COMPASS study (with only a limited number of pediatric case patients), or the study by Shrine et al,[30] which was based on the largely European UK Biobank sample.

DISCUSSION

Given the evidence that Peru is one of the countries with the highest prevalence of childhood asthma and high disease severity,[9] our study goal was to understand whether ancestry was associated with asthma status, lung function, and total serum IgE levels in a cohort of Peruvian children. Recognizing the limitations of a relatively modest sample size, we not only performed a GWAS but specifically sought to replicate prior risk loci associated with childhood-onset asthma and lung function.

We found that participants had a high proportion of indigenous ancestry (85%) but lower proportion of European ancestry (15%), similar to the proportions in prior studies showing indigenous American ancestry as the predominant ancestry in Peruvians.[18] Latino populations worldwide represent some of the most diverse and admixed populations, which has implications for health and disease.[36,37] The source of indigenous American ancestry could also differ between Latino populations, with differing frequencies of risk alleles between these types,[38] and this could account for the difference in the effect of the ancestral population on development of asthma.[39] The differences in the relationship between European ancestry and asthma in these admixed populations are complex, and these differing effects (ie, risk vs

protection) could be due to significant contributions from African ancestry (where there is 3-way admixture), and/or the variability in origin of indigenous ancestry.

With respect to total serum IgE level, we found that European ancestry is associated with lower total serum IgE levels and indigenous ancestry is associated with higher total serum IgE levels. These findings are consistent with those of prior studies showing that populations of Amer-indigenous ancestry descent have higher total serum IgE levels than do populations with higher European ancestry.[40-42]

Finally, with respect to lung function, consistent with multiple prior studies,[5,43,44] we found indigenous American ancestry to be associated with higher lung function and European ancestry to be associated with lower lung function. When stratified by asthma status, this effect remained statistically significant in the controls but not in the case patients. There is remarkable diversity in the genes associated with lung function, and these genes could differ between patients with asthma and those without asthma.[45] It is also possible that the higher effect seen in controls is a better reflection of the overall effect of genetic ancestry on lung function because the presence of asthma in itself could mediate the relationship between ancestry and lung function. Following the development of asthma, ancestry might lead to smaller differences in lung function when a child with asthma is compared with another, whereas in those without asthma, ancestry may play a larger role in lung function.

Our GWAS was unable to identify genome-wide associations of asthma, and this is likely due to the small sample size; a power calculation of the extremes of effect sizes (odds ratio = 1.013-1.97) noted in Pividori et al[6] shows that our power even for asthma replications was less than 60%. However, we were able to identify associations for total serum IgE level and FEV_1 value. We found a strong association between the *HLA-DQA1* region and total serum IgE levels. Although the most significant SNP in the region, namely, rs3135348, has not been previously reported in the GWAS literature or the GWAS catalog, the MHC region on Chr6, which harbors HLA genes, has consistently been documented to be associated with asthma[46-48] and total serum IgE levels[40,49-53] in diverse populations. In the GABRIEL consortium which looked at genetic signals for asthma, the SNP with the strongest association, rs9271300, was located in the MHC region between *HLA-DRB1* and *HLA-DQA1*.[46] Similarly, in the EVE Asthma Genetics Consortium, which had better representation of Latino and African American participants, *HLA-DRB1* had the strongest association with total serum IgE levels.[40] The MHC region contains immune-regulating genes, including *HLA-DR-B1* variants, which modulate antigen presentation to T_H cells, leading to T_H2 skewing.[50] This is an important step in the development of allergies and in total serum IgE production.[50,54] Multiple other genes in the HLA region, including *HLA-G, HLA-A* and *HLA-DQA2*, have also been shown to be associated with total serum IgE levels.[52]

Beyond associations with specific SNPs, the association with specific alleles and haplotypes for MHC class I and II genes has also been demonstrated in prior studies.[35,55-58] In these populations from Peru, our extension to looking at alleles at the HLA class I and II genes that were imputed with a MAF of approximately 5% mirrors the patterns of association from the GWAS; associations are identified for *HLA-DQA1* (04:01) and *HLA-DQB1* (03:02 and 04:02) alleles only for total serum IgE level, with evidence coming from both the case patients and the controls. The extensive LD within this region is recapitulated in haplotypes across the 2 genes, which includes the alleles driving the association with total serum IgE level. The strongest association is with *HLA-DQA*~*HLA-DQB1**04.01~04.02, with a haplotype frequency of 19% in this Peru sample. Although the haplotype has an increased risk for asthma (odds ratio = 1.18), this is not statistically significant ($P = .296$), and it supports the interpretation that this association is driving atopy and not asthma in these samples. A conditional model with the peak SNP from the GWAS, along with the haplotypes, reveals that there is an independent effect at the SNP beyond the HLA haplotype.

We found a GWAS association between 12 SNPs mapping to a locus on Chr19 and FEV_1 value among the case patients and controls. Notably, we were able to replicate this novel finding at 7 SNPS in similarly admixed pediatric samples from SCAALA, but we were unable to replicate it in either similarly admixed adult samples or the largely European UK Biobank samples. This region on Chr19 houses multiple zinc finger genes, including *ZNF579, ZNF865*, and *FIZ1*. Prior studies have shown that zinc finger genes are associated with bronchodilator response and airway remodeling. The Childhood Asthma Management Program study showed that variants in the *ZNF432* gene were involved in the bronchodilator response among children with asthma and that inhaled steroids modified this response,[59] whereas prior studies in patients with chronic obstructive pulmonary disease (COPD) have suggested that transcription factors of zinc finger family of proteins are involved in airway remodeling and COPD pathogenesis.[60,61] Colocalization analysis was unable to identify a regulatory overlap between our peak association signal and these zinc finger genes. We did find strong evidence for colocalization with the LncRNA *CTD-2537I9.5*. Although not much is known specifically for this LncRNA, there is well-documented evidence for the role of LncRNAs in lung biology in general. LncRNAs are a diverse class of transcribed but not translated RNAs that are approximately 200 nucleotides in length. Although they do not encode proteins, they can interact with both RNA and DNA in the cell and have been shown to lead to transcriptional activation of other proteins such as *HIF1a* and *Myc*, for example.[62] In addition, they have been implicated in altering methylation of DNA, presumably by binding to DNA fragments and preventing methylation and thereby influencing epigenetic regulation.[62-64] Although the majority of studies have implicated LncRNAs in cancer, and specifically in lung cancer,[63] they have also been implicated in almost all types of lung disease, including the following[65]: acute lung injury, in which they have been proposed to function as decoys to some miRs; COPD, in which their regulation is altered with cigarette smoke; idiopathic pulmonary fibrosis, in which they have been implicated in fibroblast proliferation; and pulmonary arterial hypertension with smooth muscle proliferation.

Several loci from prior GWASs replicated in our Peruvian cohort. For asthma, we found the replication of 17q12 and 12q13.2 shown by Pividori et al[6] as being associated with childhood-onset asthma. 17q12 was the most significant locus associated with childhood-onset asthma, with each copy of the risk allele conferring onset approximately 2.5 years earlier than in individuals without the risk.[6] *GSDMB,* a major protein-encoding gene in this region, maps to a broad region of high LD with multiple other genes associated with asthma in Latino and non-Latino populations, including *ORMDL3* and

IKZF3.[66,67] Chr17q12–q21 is associated with asthma risk, and it demonstrates significant ancestral heterogeneity, with the risk increasing as the proportion of European ancestry increases.[8,68] For total serum IgE level, 6p21.32, a shared locus for adult and childhood asthma, directly overlaps the *HLA-DR* region, which met the GWAS level of significance for total serum IgE level in our analyses. The 6p21.33 locus spans a broader region that includes both *HLA-B* and *HLA-C* genes. This validates prior evidence that the MHC region on Chr6 plays an important role in allergy.[40,47,48,51,52] For lung function, 8q24.21 reached the Bonferroni-corrected threshold. This was one of 2 loci identified as being associated with asthma in the genome-wide meta-analyses from populations of African ancestry in CAAPA.[8] It is also the only locus that passes our more stringent correction for multiple testing based on the number of independent SNPs evaluated. This replication in our sample of individuals with predominantly indigenous ancestry supports the idea that this locus may be particularly relevant in admixed populations. The *TATDN1* gene in the region of Chr8q24 has increased expression in airway smooth muscle cells of patients with asthma, and the adjacent binding site for CCAAT/enhancer-binding protein β, is a transcription factor involved in the IL-17 signaling pathway that modulates the effect of house dust mite on lung function.[8,69]

One of the biggest limitations of our work is its limited sample size. Nonetheless, we were able to identify statistically significant associations for total serum IgE level and FEV_1 value We also replicated prior observations regarding the relationship between indigenous ancestry and lung function and total serum IgE level. Although the association between HLA and total serum IgE is not novel, it recapitulates the importance of the *HLA* region in allergy with a strong association for total serum IgE level in this allergy asthmatic pediatric sample from Peru. We have shown that there is an HLA haplotype, *DQA1:04*01~DQB1:04*02*, that is strongly associated with total serum IgE. Even after this haplotype is taken into account, the peak GWAS SNP retains some significance, confirming the inability to fully disentangle quantitative (ie, regulatory eQTL effects) and qualitative (ie, alleles representing antigen-specific binding) effects.[70] Despite our study's limitation regarding sample size, we identified a novel locus for lung function that we were able to replicate in an independent pediatric population also with indigenous central American admixture. We acknowledge that although our colocalization analysis does hone in on a potential LncRNA (*CTD-2537I9.5*) as a target for our novel Chr19 signal, little is known about its specific role in lung physiology and further work is needed to gain mechanistic insights. The peak variants mapping to our GWAS signal are common in both European and American ancestry populations in the 1000 Genomes European and American ancestry populations. A notable exception is rs146619376, which has a MAF of 27% in GASP but a MAF of 4% and 13% in 1000 Genomes European and American populations, respectively. The higher MAF of rs146619376 in GASP suggests that this association may be population specific, and formal analyses utilizing local ancestry in the future would be valuable to resolve the potential for ancestry-specific effects at this local; however, our current data are hampered by sparse genotyping in the reference data. In conclusion, our replication of a novel finding in only a similarly admixed pediatric population and not in adults or Europeans validates the importance of including underrepresented samples in additional explorations of the genetics of allergy and asthma. This is particularly important as we continue to work toward implementing precision medicine initiatives and eliminating health disparities in genetics research.

Key messages

- **Genetic ancestry is associated with asthma-related phenotypes of lung function and total serum IgE.**
- **A novel locus mapping to a LncRNA for lung function may be specific to children with continental American admixture.**

REFERENCES

1. Centers for Disease Control and Prevention. Asthma's impact on the nation 2012. Available at: https://www.cdc.gov/asthma/impacts_nation/asthmafactsheet.pdf. Accessed May 1, 2018.
2. World Health Organization. Asthma fact sheet 2017. Available at: https://www.who.int/news-room/fact-sheets/detail/asthma. Accessed May 1, 2018.
3. Mersha TB. Mapping asthma-associated variants in admixed populations. Front Genet 2015;6:292.
4. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African diaspora genome. Nat Commun 2016;7:12522.
5. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, et al. Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study. J Allergy Clin Immunol 2014;134:295-305.
6. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med 2019;7:509-22.
7. Torgerson DG, Capurso D, Mathias RA, Graves PE, Hernandez RD, Beaty TH, et al. Resequencing candidate genes implicates rare variants in asthma susceptibility. Am J Hum Genet 2012;90:273-81.
8. Daya M, Rafaels N, Brunetti TM, Chavan S, Levin AM, Shetty A, et al. Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. Nat Commun 2019;10:880.
9. Worldwide variations in the prevalence of asthma symptoms: the International Study of Asthma and Allergies in Childhood (ISAAC). Eur Respir J 1998;12:315-35.
10. Hansel NN, Romero KM, Pollard SL, Bose S, Psoter KJ, JU L, et al. Ambient air pollution and variation in multiple domains of asthma morbidity among Peruvian children. Ann Am Thorac Soc 2019;16:348-55.
11. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J 2012;40:1324-43.
12. Bien SA, Wojcik GL, Zubair N, Gignoux CR, Martin AR, Kocarnik JM, et al. Strategies for enriching variant coverage in candidate disease loci on a multiethnic genotyping array. PLoS One 2016;11:e0167758.
13. A haplotype map of the human genome. Nature 2005;437:1299-320.
14. PLINK 1.9. Purcell S, Chang C. Available at: www.cog-genomics.org/plink/2.0. Accessed, February 8, 2018.
15. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.
16. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26:2867-73.
17. Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, et al. A genomewide admixture mapping panel for Hispanic/Latino populations. Am J Hum Genet 2007;80:1171-8.
18. Harris DN, Song W, Shetty AC, Levano KS, Caceres O, Padilla C, et al. Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proc Natl Acad Sci U S A 2018;115:e6526-35.
19. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655-64.
20. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-7.

21. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol 2015;39:276-93.
22. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2019. Available at: http://www.r-project.org/. Accessed February 2, 2018.
23. Turner S. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. Journal of Open Source Software 2018;3:731.
24. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383.
25. Wallace C. Statistical testing of shared genetic control for potentially related traits. Genet Epidemiol 2013;37:802-13.
26. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG–HLA genotype imputation with attribute bagging. Pharmacogenomics J 2014;14:192-200.
27. Zheng X. Imputation-Based HLA Typing with SNPs in GWAS studies. Methods Mol Biol 2018;1802:163-76.
28. Pappas DJ, Marin W, Hollenbach JA, Mack SJ. Bridging ImmunoGenomic data analysis workflow gaps (BIGDAWG): an integrated case-control analysis pipeline. Hum Immunol 2016;77:283-7.
29. Kuna P, Peters MJ, Manjra AI, Jorup C, Naya IP, Martínez-Jimenez NE, et al. Effect of budesonide/formoterol maintenance and reliever therapy on asthma exacerbations. Int J Clin Pract 2007;61:725-36.
30. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet 2019;51:481-93.
31. Cruz AA, Riley JH, Bansal AT, Ponte EV, Souza-Machado A, Almeida PCA, et al. Asthma similarities across ProAR (Brazil) and U-BIOPRED (Europe) adult cohorts of contrasting locations, ethnicity and socioeconomic status. Respir Med 2020;161:105817.
32. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26(18):2336-7.
33. Yizhak K, Aguet F, Kim J, Hess JM, Kubler K, Grimsby J, et al. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. Science 2019;364:eaaw0726.
34. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet 2011;43:887-92.
35. Schoettler N, Rodríguez E, Weidinger S, Ober C. Advances in asthma and allergic disease genetics: is bigger always better? J Allergy Clin Immunol 2019;144:1495-506.
36. Vergara C, Murray T, Rafaels N, Lewis R, Campbell M, Foster C, et al. African ancestry is a risk factor for asthma and high total IgE levels in African admixed populations. Genet Epidemiol 2013;37:393-401.
37. Burchard EG, Avila PC, Nazario S, Casal J, Torres A, Rodriguez-Santana JR, et al. Lower bronchodilator responsiveness in Puerto Rican than in Mexican subjects with asthma. Am J Respir Crit Care Med 2004;169:386-92.
38. Lorenzo Bermejo J, Boekstegers F, Gonzalez Silos R, Marcelain K, Baez Benavides P, Barahona Ponce C, et al. Subtypes of Native American ancestry and leading causes of death: Mapuche ancestry-specific associations with gallbladder cancer risk in Chile. PLoS Genet 2017;13:e1006756.
39. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. Science 2002;298:2381-5.
40. Levin AM, Mathias RA, Huang L, Roth LA, Daley D, Myers RA, et al. A meta-analysis of genome-wide association studies for serum total IgE in diverse study populations. J Allergy Clin Immunol 2013;131:1176-84.
41. Yang JJ, Burchard EG, Choudhry S, Johnson CC, Ownby DR, Favro D, et al. Differences in allergic sensitization by self-reported race and genetic ancestry. J Allergy Clin Immunol 2008;122:820-7.e9.
42. Litonjua AA, Celedon JC, Hausmann J, Nikolov M, Sredl D, Ryan L, et al. Variation in total and specific IgE: effects of ethnicity and socioeconomic status. J Allergy Clin Immunol 2005;115:751-7.
43. Salari K, Choudhry S, Tang H, Naqvi M, Lind D, Avila PC, et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. Genet Epidemiol 2005;29:76-86.
44. Pino-Yanes M, Thakur N, Gignoux CR, Galanter JM, Roth LA, Eng C, et al. Genetic ancestry influences asthma susceptibility and lung function among Latinos. J Allergy Clin Immunol 2015;135:228-35.
45. Brehm JM, Man Tse S, Croteau-Chonka DC, Forno E, Litonjua AA, Raby BA, et al. A Genome-wide association study of post-bronchodilator lung function in children with asthma. Am J Respir Crit Care Med 2015;192:634-7.
46. Moffatt MF, Schou C, Faux JA, Abecasis GR, James A, Musk AW, et al. Association between quantitative traits underlying asthma and the HLA-DRB1 locus in a family-based population sample. Eur J Hum Genet 2001;9:341-6.
47. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, et al. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. J Allergy Clin Immunol 2010;125:328-35.e11.
48. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363(13):1211-21.
49. Kim JH, Lee SY, Kang MJ, Yoon J, Jung S, Cho HJ, et al. Association of genetic polymorphisms with atopic dermatitis, clinical severity and total IgE: a replication and extended study. Allergy Asthma Immunol Res 2018;10:397-405.
50. Potaczek DP, Kabesch M. Current concepts of IgE regulation and impact of genetic determinants. Clin Exp Allergy 2012;42:852-71.
51. Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, et al. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. J Allergy Clin Immunol 2015;135:1502-10.
52. Granada M, Wilk JB, Tuzova M, Strachan DP, Weidinger S, Albrecht E, et al. A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study. J Allergy Clin Immunol 2012;129:840-5.e21.
53. Upham JW, Holt PG, Taylor A, Thornton CA, Prescott SL. HLA-DR expression on neonatal monocytes is associated with allergen-specific immune responses. J Allergy Clin Immunol 2004;114:1202-8.
54. Georas SN, Guo J, De Fanis U, Casolaro V. T-helper cell type-2 regulation in allergic disease. Eur Respir J 2005;26:1119-37.
55. Movahedi M, Moin M, Gharagozlou M, Aghamohammadi A, Dianat S, Moradi B, et al. Association of HLA class II alleles with childhood asthma and Total IgE levels. Iran J Allergy Asthma Immunol 2008;7:215-20.
56. Kontakioti E, Domvri K, Papakosta D, Daniilidis M. HLA and asthma phenotypes/endotypes: a review. Hum Immunol 2014;75:930-9.
57. Takejima P, Agondi RC, Rodrigues H, Aun MV, Kalil J, Giavina-Bianchi P. Allergic and nonallergic asthma have distinct phenotypic and genotypic features. Int Arch Allergy Immunol 2017;172:150-60.
58. Mubarak B, Afzal N, Javaid K, Talib R, Aslam R, Latif W, et al. Frequency of HLA DQβ1*0201 and DQβ1*0301 alleles and total serum IgE in patients with bronchial asthma: a pilot study from Pakistan. Iran J Allergy Asthma Immunol 2017;16:313-20.
59. Wu AC, Himes BE, Lasky-Su J, Litonjua A, Peters SP, Lima J, et al. Inhaled corticosteroid treatment modulates ZNF432 gene variant's effect on bronchodilator response in asthmatics. J Allergy Clin Immunol 2014;133:723-8.e3.
60. Koczulla AR, Jonigk D, Wolf T, Herr C, Noeske S, Klepetko W, et al. Kruppel-like zinc finger proteins in end-stage COPD lungs with and without severe alpha1-antitrypsin deficiency. Orphanet J Rare Dis 2012;7:29.
61. Abe K, Sugiura H, Hashimoto Y, Ichikawa T, Koarai A, Yamada M, et al. Possible role of Kruppel-like factor 5 in the remodeling of small airways and pulmonary vessels in chronic obstructive pulmonary disease. Respir Res 2016;17:7.
62. Shih JW, Kung HJ. Long non-coding RNA and tumor hypoxia: new players ushered toward an old arena. J Biomed Sci 2017;24:53.
63. Jafarzadeh M, Tavallaie M, Soltani BM, Hajipoor S, Hosseini SM. LncRNA HSPC324 plays role in lung development and tumorigenesis. Genomics 2020;112:2615-22.
64. Wu X, Tudoran OM, Calin GA, Ivan M. The many faces of long noncoding RNAs in cancer. Antioxid Redox Signal 2018;29:922-35.
65. Groot M, Zhang D, Jin Y. Long non-coding RNA review and implications in lung diseases. JSM Bioinform Genom Proteom 2018;3:119044.
66. Yan Q, Brehm J, Pino-Yanes M, Forno E, Lin J, Oh SS, et al. A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. Eur Respir J 2017;49:1601505.
67. Galanter J, Choudhry S, Eng C, Nazario S, Rodriguez-Santana JR, Casal J, et al. ORMDL3 gene is associated with asthma in three ethnically diverse populations. Am J Respir Crit Care Med 2008;177:1194-200.
68. Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, et al. A decade of research on the 17q12-21 asthma locus: piecing together the puzzle. J Allergy Clin Immunol 2018;142:749-64.e3.
69. Forno E, Sordillo J, Brehm J, Chen W, Benos T, Yan Q, et al. Genome-wide interaction study of dust mite allergen on lung function in children with asthma. J Allergy Clin Immunol 2017;140:996-1003.e7.
70. Kanchan K, Irizar CS, Bunyavanich S, Mathias RA. Current insights into the genetics of food allergy. J Allergy Clin Immunol 2021;147:15-28.