Title

An admixture mapping meta-analysis implicates genetic variation at 18q21 with asthma susceptibility in Latinos

Journal

Journal of Allergy and Clinical Immunology, 143(3)

ISSN

0091-6749

Authors

Gignoux, Christopher R

Torgerson, Dara G

Pino-Yanes, Maria

et al.

Publication Date

2019-03-01

DOI

10.1016/j.jaci.2016.08.057

Peer reviewed

HHS Public Access

Author manuscript

J Allergy Clin Immunol. Author manuscript; available in PMC 2020 March 01.

Published in final edited form as:

J Allergy Clin Immunol. 2019 March ; 143(3): 957–969. doi:10.1016/j.jaci.2016.08.057.

An admixture mapping meta-analysis implicates genetic variation at 18q21 with asthma susceptibility in Latinos

A full list of authors and affiliations appears at the end of the article.

Abstract

Background: Asthma is a common but complex disease with racial/ethnic differences in prevalence, morbidity, and response to therapies.

Objective: We sought to perform an analysis of genetic ancestry to identify new loci that contribute to asthma susceptibility.

Methods: We leveraged the mixed ancestry of 3902 Latinos and performed an admixture mapping meta-analysis for asthma susceptibility. We replicated associations in an independent study of 3774 Latinos, performed targeted sequencing for fine mapping, and tested for disease correlations with gene expression in the whole blood of more than 500 subjects from 3 racial/ethnic groups.

Results: We identified a genome-wide significant admixture mapping peak at 18q21 in Latinos ($P = 6.8 \times 10^{-6}$), where Native American ancestry was associated with increased risk of asthma (odds ratio [OR], 1.20; 95% CI, 1.07–1.34; $P = .002$) and European ancestry was associated with protection (OR, 0.86; 95% CI, 0.77–0.96; $P = .008$). Our findings were replicated in an independent childhood asthma study in Latinos ($P = 5.3 \times 10^{-3}$, combined $P = 2.6 \times 10^{-7}$). Fine mapping of 18q21 in 1978 Latinos identified a significant association with multiple variants 5′ of SMAD family member 2 (*SMAD2*) in Mexicans, whereas a single rare variant in the same window was the top association in Puerto Ricans. Low versus high *SMAD2* blood expression was

Corresponding author: Christopher R. Gignoux, PhD, University of Colorado Denver-Anschutz Medical Campus, Colorado Center for Personalized Medicine and Department of Biostatistics, Colorado Center for Personalized Medicine, Aurora, CO 80045. chris.gignoux@ucdenver.edu.

*These authors contributed equally to this work.

‡Christopher R. Gignoux, PhD, is currently affiliated with the Department of Genetics, Stanford Center for Computational, Evolutionary, and Human Genomics, Stanford University, Stanford, California.

§Lawrence H. Uricchio, PhD, is currently affiliated with the Department of Biology, Stanford Center for Computational, Evolutionary, and Human Genomics, Stanford University, Stanford, California.

Disclosure of potential conflict of interest: C. R. Gignoux receives grant support and travel support from the National Institutes of Health (NIH) and holds stock with 23andMe. M. Pino-Yanes receives payments for lectures from Affymetrix. L. H. Uricchio receives grant support from the NIH. J. Galanter receives grant support from the NIH. R. Kumar receives grant support from the NIH. N. Thakur receives grant support from the National Institute of General Medical Sciences (NIGMS) and National Heart, Lung, and Blood Institute (NHLBI). S. S. Oh receives grant funding from the NIH. M. McGarry receives grant support from the NIH. M. A. Seibold receives research support from Pfizer and MedImmune. H. J. Farber receives grant support from the NIH. P. Avila receives grant and travel support from the NIH. E. Brigino-Buenaventura receives grant support, honorarium, and travel support from the Sandler Foundation. A. M. Levin received research support from the NIH. B. A. Raby receives royalties from UpToDate and holds stock with CureSpark. F. J. Martinez receives grant support from the NIH/NHLBI and Johnson & Johnson and serves as a consultant for Copeval. D. L. Nicolae receives grant support from the NIH. S. Sen has received grants from the NIH. L. Keoki Williams receives grant support from the National Institute of Allergy and Infectious Diseases (NIAID), NHLBI, and NIH. R. D. Hernandez receives grant support from the NIH. The rest of the authors declares that they have no relevant conflicts of interest.

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

correlated with case status (13.4% lower expression; OR, 3.93; 95% CI, 2.12–7.28; $P < .001$). In addition, lower expression of *SMAD2* was associated with more frequent exacerbations among Puerto Ricans with asthma.

Conclusion: Ancestry at 18q21 was significantly associated with asthma in Latinos and implicated multiple ancestry-informative noncoding variants upstream of *SMAD2* with asthma susceptibility. Furthermore, decreased *SMAD2* expression in blood was strongly associated with increased asthma risk and increased exacerbations.

Keywords

Asthma; *asthma exacerbations*; *admixture mapping*; *meta-analysis*; *Latinos*; SMAD2; *gene expression*; *targeted sequencing*; *rare variation*

Asthma prevalence varies dramatically by race/ethnicity. In the United States childhood asthma prevalence is highest among Puerto Ricans (24.8%), intermediate among African Americans(16.3%), and lowest among European Americans (7.8%) and Mexican Americans (7.8%).[1] Substantial evidence supports both environmental and genetic contributions to asthma, with heritability estimates ranging to upward of 75%.[2] Genome-wide association studies (GWASs) have identified more than 25 genetic risk factors for asthma.[3] Nonetheless, known genetic associations account for only a small proportion of the genetic basis of asthma and have provided limited insight into racial/ethnic disparities in its prevalence and severity. This might be due in part to the presence of population-specific genetic risk factors and the limited number of GWASs in non-European populations.[4–6] For example, many asthma-associated variants identified in European American subjects demonstrate significant effect heterogeneity or have failed to replicate in non-European populations.[7,8] Additionally, rare genetic polymorphisms, which are more likely to be population specific,[9] can contribute to the genetic underpinnings of complex diseases.[10] These unexplored genetic factors can contribute to disparities in asthma susceptibility between racial/ethnic groups.

Latinos are primarily admixed descendants of Native American, European, and Sub-Saharan African ancestors,[11,12] and African Americans are primarily admixed descendants of Sub-Saharan African and European ancestors.[13] This variation in genetic ancestry at the locus-specific level can be leveraged to identify novel loci associated with asthma through admixture mapping,[14–16] an approach that is complementary to traditional GWASs. Specifically, admixture mapping compares genetic ancestry between cases and control subjects rather than genotypes at individual single nucleotide polymorphisms (SNPs), as in a GWAS. The variation in ancestry patterns stems from the unique patterns of history encoded in the genes of admixed populations. Admixture mapping identifies disease-associated loci through ancestry linkage disequilibrium (LD) with causal variation rather than SNP LD and thus might better capture risk loci in which multiple rare variants are more strongly linked to genetic ancestry than to a single common variant. Importantly, these longer blocks of ancestry result in a decreased multiple testing burden. Admixture mapping is especially relevant when the disease cause might be different in different ancestral populations, as in asthmatic patients. Recently, we demonstrated that admixture mapping and ancestry deconvolution can identify associations with asthma in African Americans[17] and Latinos,

[18,19] total IgE levels,[20] bronchodilator response,[21] and lung function in Mexicans and Mexican Americans.[22]

Here we extend the transethnic meta-analysis framework used to study asthma in the EVE Asthma Genetics Consortium[23,24] by performing an admixture mapping meta-analysis of 7606 Latino and 3106 African American asthma cases and control subjects. We identified a novel locus in which ancestry is associated with asthma at 18q21.1 in Latinos and implicate multiple noncoding variants 5′ of SMAD family member 2 (*SMAD2*) with asthma susceptibility. Furthermore, we correlate the expression of *SMAD2* with asthma susceptibility and exacerbations, further implicating *SMAD2* as playing an important role in asthma.

METHODS

Our study approach is outlined in Fig 1.

Study subjects

Discovery.—Self-identified Latino and African American subjects from 8 independent studies participating in the EVE Asthma Genetics Consortium were included in our analysis (total of 3902 Latinos and 3106 African Americans; Table I).[23] The EVE Consortium is a large multiethnic assembly of asthma studies with existing genome-wide SNP genotypes from 9 US institutions. Detailed descriptions of each study have been published previously.[23] Briefly, 3 case-control (Genetics of Asthma in Latino Americans [GALA I] Mexican, GALA I Puerto Rican, and the Children's Health Study) and 2 trio (Childhood Asthma Research Network and Mexico City Childhood Asthma Study) studies were included in the discovery analysis. Additional details can be seen in Table I. In the current project we used all genotyped autosomal SNPs passing prior quality control standards (imputed genotypes were not included in ancestry estimation).[23] Local institutional review boards approved the studies.

Replication.—We replicated a genome-wide significant association with ancestry in the Genes-Environments & Admixture in Latino Americans (GALA II) study,[18] a large, multicenter, case-control study of Latino children between the ages of 8 and 21 years with and without asthma (see the Methods section in this article's Online Repository at www.jacionline.org). Subjects were genotyped on the Affymetrix Axiom Genome-Wide LAT1 Array (World Array 4; Affymetrix, Santa Clara, Calif),[25] as described by Galanter et al[18] (see the Methods section in this article's Online Repository). Local institutional review boards approved the study, and all subjects and legal guardians provided written informed assent/consent.

Targeted sequencing

Sequencing of a 342-Kb region at 18q21.1 was performed on 1978 Latinos from the replication study (GALA II), including 986 Puerto Ricans (495 cases and 491 control subjects) and 992 Mexicans (495 cases and 497 control subjects), through the National Heart, Lung, and Blood Institute's Resequencing and Genotyping Service (chromosome 18:

45321283–45663680, hg19 coordinates). Target enrichment was performed with a custom NimbleGen SeqCap EZ library, for which 87.3% of the target could be uniquely captured. Sequencing was performed to an average of 257× coverage (with 97% of the target having >20× coverage) on a HiSeq device by using paired-end reads of 150 bp in length. Reads were mapped to the human genome build hg19, and variants were called by using Genome Analysis Toolkit best practices.[26]

Gene expression measurements

We measured expression of *SMAD2*, *SMAD3*, zinc finger and BTB domain containing 7C (*ZBTB7C*), and the housekeeping gene β-glucuronidase (*GUSB*) in peripheral blood using TaqMan RT-PCR assays (Thermo Fisher Scientific, Waltham, Mass). Total RNA was isolated from PAXgene Blood RNA tubes (PreAnalytiX GmbH, Hombrechtikon, Switzerland), and RNA integrity was assessed with the Agilent BioAnalyzer (Agilent Technologies, Santa Clara, Calif). Samples with RNA integrity of less than 6 were excluded from further analysis. *SMAD2* expression was measured in GALA II Puerto Ricans (107 cases and 54 control subjects) and Mexicans (122 cases and 56 control subjects) and African Americans from the Study of African Americans, Asthma, Genes and Environment (SAGE; 114 cases and 51 control subjects, see the Methods section in this article's Online Repository).[27,28] Lastly, we downloaded *SMAD2* expression data from isolated $CD4^+$ lymphocytes in the Gene Expression Omnibus[29] from Mobini et al.[30] The study included subjects from Germany with and without asthma, seasonal allergic rhinitis, and eczema, totaling 15 asthmatic patients and 8 asthmatic control subjects.

Statistical analyses

Ancestry estimation.—We estimated local ancestry at every SNP on our genotyping for each study individually using 2 LAMP algorithms: LAMP[31] for case-control studies and LAMP-HAP[32] for family-based studies. We ran LAMP assuming 3 ancestral populations consistent with prior studies and with reference panels designed with each platform in mind. Extensive details on platform choice, rationale, and methods are available in the Methods section in this article's Online Repository.

Admixture mapping.—Briefly, we treat ancestry estimates as an additive count (0, 1, 2) of ancestries per locus. We used a 2-^{df} likelihood ratio test to evaluate the total local effect of the 3 primary ancestries in Latinos on asthma susceptibility (see the Methods section in this article's Online Repository). We performed separate versions of generalized linear models for each study design: logistic regression for case-control studies (GALA I Mexican, GALA I Puerto Rican, and Children's Health Study) and an extended transmission disequilibrium test (TDT) Poisson regression for trio-based studies (Mexico City Childhood Asthma Study and Childhood Asthma Research Network). *P* values were then meta-analyzed across studies to evaluate the joint association by using the Fisher method. We determined a study-specific significance threshold using empiric autoregression to estimate the effective number of tests using the *coda* package in R software (see the Methods section in this article's Online Repository).

To evaluate the effect of a specific ancestry, we coded ancestry at each position as a biallelic state (eg, African vs non-African) and evaluated the additive case of 0, 1, or 2 copies of that ancestry at any locus. Similar to GWASs, we then used logistic regression for case-control studies using R software[33] and the TDT[34] for trio studies by using custom Python[35] scripts available on request. Single-ancestry tests were combined by using fixed-effects models in PLINK[36] to evaluate statistical significance, magnitude of effect, and study heterogeneity (I^2). Sites with an I^2 value of heterogeneity of greater than 50% were inspected to determine whether a random-effects model was warranted (to incorporate between-study heterogeneity).

Imputation and fine mapping.—To identify genetic variants driving the association with local ancestry, we performed genotype imputation using a standard pipeline (see the Methods section in this article's Online Repository) across an approximately 5-Mb region around *SMAD2*. We analyzed each study using a similar framework, as described above (ie, logistic regression and TDT). A fixed-effects meta-analysis was performed in PLINK, and a random-effects meta-analysis was performed with the R *package metafor*.[37]

Targeted sequencing analysis.—Pooled rare variant association tests were performed by using the optimized Sequence Kernel Association Test (SKAT-O, available at http://www.hsph.harvard.edu/skat/download/),38,39 as implemented in R software[40] and on individual variants by using logistic regression. The kernel provides an efficient method for combining evidence across individual variants, where pooling provides additional power compared with single rare variant tests. We performed tests of association with asthma as a case-control test for variants within coding exons of *SMAD2* and *ZBTB7C* separately and agnostically across the joint 342-Kb region using a sliding window of 5 Kb in length and a step size of 500 bp (total of 676 windows) to evaluate all possible loci within the region of ancestry association. We also performed a similar scan of overall expression quantitative trait loci (eQTLs), adjusting for asthma as a covariate. A total of 10,000 permutations were run at an a value of .05 to determine statistical significance, and from this, we could determine a false discovery rate in R software to adjust for multiple comparisons (method = BH). For individual variants at 18q21.1, we performed additional imputation using IMPUTE2 for the remaining 1778 Latinos from the GALA II study who were not directly sequenced with the filters described above. Reference haplotypes included subjects from the GALA II study who were sequenced (n = 1978) and subjects from the 1000 Genomes Project (phase 1). All tests of association were adjusted for local and global ancestry, age, sex, and recruitment area.

Gene expression analysis.—To provide additional lines of evidence for our associations, we performed gene expression at the genes of interest using quantitative PCR on whole blood for a subset of GALA II and SAGE II subjects with available RNA. We rescaled expression to fold change levels ($2^{-\Delta CT}$). We maximized goodness of fit of a cut point of gene expression by using Bayes factors calculated from generalized linear models, adjusting for age, sex, recruitment center, and ancestry. These were used downstream for regression testing with observed exacerbations and as outcome for single variant and pooled (SKAT-O) association tests. Additional details of exacerbation calculations, specific models

used, estimation of cell composition, and other details are available in the Methods section in this article's Online Repository.

RESULTS

Admixture mapping meta-analysis and replication

We performed local ancestry estimation, ancestry interpolation, and admixture mapping in 5 different studies of childhood asthma from the EVE Consortium, comprising 3902 Latino subjects. The multiple testing threshold using autocorrelation was determined to be a *P* value of less than 4×10^{-5}, a value similar to prior analyses leveraging permutations.[18] We identified a genome-wide significant admixture mapping peak in Latinos at 18q21 using the likelihood ratio test ($P = 6.8 \times 10^{-6}$; Fig 2, A and B). This peak was driven by multiple ancestry signals at the locus: Native American ancestry was found to increase risk (odds ratio [OR], 1.20; 95% CI, 1.07–1.34; $P = .002$), whereas European ancestry was found to be protective (OR, 0.86; 95% CI, 0.77–0.96; $P = .008$; Fig 2, C); African ancestry did not appear to play a significant role ($P = .4$; Fig 2, C) but appears protective in the trio-based studies. We note that Native American and European ancestries are negatively correlated, and therefore these effects are not independent. The admixture peak is bounded by 2 recombination hotspots and overlaps 2 protein-coding genes: *SMAD2* and *ZBTB7C* (Fig 2, B).

We replicated the association between asthma and ancestry at 18q21 in an independent sample of 3774 Latinos from the GALA II Study (min $P = 5.3 \times 10^{-3}$, Table II). The direction of effect in the discovery and replication populations was homogeneous for both Native American and European ancestry (OR, 1.09 [$P = 6.3 \times 10^{-3}$] and 0.87 [$P = 5.8 \times 10^{-3}$], respectively; Table II). In African Americans we found no significant association with ancestry at 18q21 (P = .7; see Fig E1 in this article's Online Repository at www.jacionline.org). In European Americans we are unable to perform admixture mapping; however, no SNP within the 18p21 region was previously identified as being significantly associated with asthma in the EVE Consortium meta-analysis.[23]

In addition to the likelihood ratio test, we performed a meta-analysis of single-ancestry admixture mapping across Latino subjects and identified 2 genome-wide significant peaks for European ancestry at 9q22 and 12p12 ($P < 4 \times 10^{-5}$, see Fig E2 and Table E1 in this article's Online Repository at www.jacionline.org). Both of these peaks failed to replicate in GALA II (lowest $P = .17$ and .21, respectively).

We also note that although genome-wide admixture mapping peaks remain rare, we were able to provide replication to a peak found on 6p21 for asthma.[18] A comprehensive set of suggestive peaks, the relevant underlying ancestry test, and the genes within each peak are available in Table E1.

Fine mapping of genetic associations

We imputed SNP genotypes from the 1000 Genomes Project (phase I) within a 5-Mb region centered on the admixture peak at 18q21 in all of the discovery populations; however, we

found no significant associations at SNPs within 18q21 (see Fig E3 in this article's Online Repository at www.jacionline.org).

In our replication study we performed targeted sequencing of 352 Kb of the admixture peak to augment genotype imputation and to better enable tests of association with pooled variants in our populations of interest. We observed a significant association with pooled variants in Mexicans within a 5-Kb window upstream of SMAD2 at chromosome 18 (45495471–45500349 [hg19]; P = 2.0×10^{-4}; false discovery rate, 0.045; Fig 3).[41,42] When we stratified based on local ancestry, we found the association at pooled variants to be strongest on haplotypes of non–Native American ancestry and subjects who were homozygous for African ancestry among Mexican subjects (see Figs E4 and E5 in this article's Online Repository at www.jacionline.org). Consistent with these findings, the top SKAT-O window harbored 7 individual variants at a frequency of less than 1% and with a P value of less than .05 were protective and African in origin (Table III). Similarly, an additional variant with a P value of less than .05 within the significant SKAT-O window was at a frequency of 6% in Mexicans, was more common on Native haplotypes, and was a risk factor for asthma (rs59002988; Table III). Although the function of these variants is unknown, 4 of these variants are within ENCODE transcription factor–binding sites and/or DNAse peaks (Fig 3 and see Table E2 in this article's Online Repository at www.jacionline.org).[41]

We found no significant association with pooled variants in Puerto Ricans only; however, the top associated individual variants were either directly within the window implicated in Mexicans (4 variants, min $P = 4.6 \times 10^{-4}$ at rs59002988) or nearby (3 variants, min $P = 3.8 \times 10^{-4}$ at rs7238092 [137 bp from window start], Fig 3). Furthermore, rs59002988 was locus-wide significantly associated with asthma in all of GALA II (OR, 1.67; $P = 1 \times 10^{-5}$) and contributes to the SKAT-O signal overlapping that same window.

Within chromosome 18 (45495283–45500283), we identified 56 variants in Mexicans and 61 variants in Puerto Ricans through direct sequencing, of which 30 were shared between populations. A stratified analysis partitioning shared versus private variants (ie, variants only found in 1 population) identified a marginal association in both classes of variation in Mexicans (see Fig E6 in this article's Online Repository at www.jacionline.org), suggesting that both private and shared variants contribute to the association. No significant associations were observed at individual variants in Mexicans, and no significant associations were observed with pooled variants in the coding exons of *SMAD2* or *ZBTB7C* in either Mexicans or Puerto Ricans (see Table E3 in this article's Online Repository at www.jacionline.org).

eQTL analysis and gene expression associations

We measured expression of *SMAD2*, *SMAD3* (the cytosolic heterodimeric partner of *SMAD2*), and *ZBTB7C* using RT-PCR from whole blood RNA in a subset of 161 Puerto Ricans from the replication study (GALA II). We found no significant association between genotypes at individual or pooled variants and gene expression in whole blood of 177 Mexicans (SMAD2) and 144 Puerto Ricans (*SMAD2* and *ZBTB7C*) from the GALA II study that were sequenced in this study (see Fig E7 in this article's Online Repository at

www.jacionline.org). However, asthma cases had significantly lower mean levels of *SMAD2* gene expression ($P < .001$). In contrast, neither *SMAD3* nor *ZBTB7C* demonstrated any difference in expression between cases and control subjects ($P = .8$ and .9, respectively).

We then investigated *SMAD2* gene expression in whole blood in all 503 African and Mexican Americans alongside the Puerto Ricans previously measured. We determined the best-fit cut point at 127% of mean expression in control subjects across all populations when partitioning by low versus high *SMAD2* expression (see Fig E8 in this article's Online Repository at www.jacionline.org). Each population has a different cut point, and here the transethnic value is provided to maximize the lower bounds on the OR across the entire sample. *SMAD2* expression of less than this cut point was associated with a 4-fold increase in the odds of asthma (OR, 3.93; 95% CI, 2.12–7.28; $P < .001$), with no ethnic heterogeneity ($P = .34$, Cochran Q). In addition, the $CD4^+$ lymphocyte data set replicated the association between asthma and low *SMAD2* expression in subjects from Germany (OR, 1.9; 95% CI, 1.3–2.8; $P = .02$; see Fig E9 and Table E4 in this article's Online Repository at www.jacionline.org). Although the sample size of this additional source of replication is small, we observe an overlapping effect size with our larger set of samples from SAGE II and GALA II.

In the subset of subjects with estimated blood cell counts, the proportion of cell types was not significantly associated with *SMAD2* expression (min marginal $P = .35$ for the proportion of $CD4^+$ T cells) nor was the joint effect of all types (likelihood ratio test: $P = .93$). Thus *SMAD2* expression was not associated with or confounded by peripheral blood cell proportions, which is consistent with previous findings.[43] The association between asthma status and *SMAD2* expression was strongest in those with an age of asthma onset of less than 5 years (see Fig E10 in this article's Online Repository at www.jacionline.org), which is consistent with prior discoveries of a stronger genetic effect in early-onset asthma.[44]

Lower SMAD2 gene expression was associated with higher exacerbation scores across the 342 cases (OR per 10% decrease in expression, 1.11; 95% CI, 1.02–1.20; P = .01; see Table IV and Fig E11 in this article's Online Repository at www.jacionline.org for the distribution of exacerbation scores). There was no significant heterogeneity (P = .08, Cochran Q), and the P value remained significant (P = .04) even when modeling the overall association through random-effects meta-analysis in META-SOFT.[45] As with the case-control association, incorporating peripheral blood cell counts did not attenuate the significance of the findings. Oral steroids could affect gene expression; however, only 5 of the participants report taking oral steroids. Including a variable for steroid use was not in itself significant nor did it affect the association between gene expression and asthma status. Although we found a highly significant association with exacerbations, we did not observe associations with other measures of severity, including asthma control or spirometric values.

We then built logistic regression models to test the ability of *SMAD2* gene expression to explain asthma exacerbations in Latinos and African Americans. We incorporated age, population, medication use, and spirometry in the full models (see Fig E12 and Table E5 in this article's Online Repository at www.jacionline.org). Subjects were dichotomized into

those with low and high exacerbation scores. We applied 3 prediction models: use of controller medication (any long-term asthma medication), response to albuterol, and *SMAD2* gene expression. Incorporating all 3 predictors (along with age and ethnicity) minimized the Aikake Information Criterion (AIC), demonstrating the additional power of *SMAD2* expression to predict high-end health care users beyond standard clinical measurements. Increasing the cut point of high versus low exacerbation resulted in improved specificity and prediction accuracy (Fig 4, B).

DISCUSSION

Our novel investigation of admixture mapping and asthma among 3902 Latinos from the EVE Asthma Genetics Consortium identified a genome-wide significant association between asthma and ancestry at 18q21 centered on the *SMAD2-ZBTB7C* locus. We replicated this finding among 3774 subjects in the GALA II study and identified a significant association with multiple noncoding rare variants 5′ upstream of *SMAD2* in Mexicans. Although we stratified by population for the sake of analysis, the strong fixed-effects analysis is consistent with a Hispanic/Latino-wide association rather than one that is geographically restricted. Although we do note that the original forest plots might look somewhat heterogeneous, the results at our top hit are overall consistent and reflect the need for large sample sizes to uncover accurate effects. In addition, we demonstrated that low *SMAD2* expression is associated with increased risk of asthma and more frequent asthma exacerbations in whole blood. Genetic variation at 18q21 was not identified as a risk factor for asthma in prior meta-analyses using traditional GWAS methods, including in these same studies.[23,46] This demonstrates the power of admixture mapping to detect genetic associations that are different from those identified by using traditional GWASs and for identifying loci in which rare genetic variation contributes to asthma susceptibility.

An important and unique contribution offered by admixture mapping is its potential for increased coverage of untyped variants caused by ancestry LD compared with genotype LD.[47] Our admixture mapping approach identified Native American ancestry at 18q21 as a risk factor for asthma in Latinos. Indeed, within the 5-Kb window showing a significant association with multiple noncoding variants in Mexicans, variants showing a marginal association on their own includes one that is risky for asthma and more common on Native American haplotypes and 7 that are protective for asthma and more common on African haplotypes. These observations are consistent with the admixture signal and might explain why no association was identified with individual variants in the original EVE Consortium meta-analysis because the effect appears to be driven predominantly by rare and low frequency variation. Therefore admixture mapping with follow-up sequencing provides the capability to discover additional genetic variation important for disease susceptibility by using existing GWAS data.

Our admixture mapping peak at 18q21 was fairly broad. However, fine mapping of this peak implicated noncoding variants directly between *SMAD2* and *ZBTB7C*. *ZBTB7C* has limited functional characterization and no known role in asthma pathophysiology. Furthermore, we found no significant difference in the expression of *ZBTB7C* in whole blood of asthmatic patients versus control subjects. In contrast, *SMAD2* was found to exhibit decreased

expression in asthmatic patients versus control subjects and is a well-characterized cofactor involved in the *TGFB* signaling pathway often studied in patients with asthma and other inflammatory diseases. The *TGFB* pathway has been implicated in multiple components of asthma biology: negative regulation of allergic airway inflammation,[48] airway remodeling,[49] and drug response.[50] Ligation of *TGFB* receptors activates the proximal transcription factors *SMAD2* and *SMAD3*; these translocate to the nucleus to regulate transcription of several hundred target genes along with a complex of DNA binding cofactors. Lower levels of *SMAD2* would then be expected to correlate with lower levels of *TGFB*-mediated signaling effect,[51] which is consistent with our observation that lower *SMAD2* expression is associated with asthma. We note that effects are not consistent in overall inflammation levels and locally in the airway, but rather, these levels would broadly represent inflammatory response. Among cases, we would expect that limited *TGFB* signaling in the blood would result in more airway inflammation, thereby leading to increased asthma exacerbations.

Although *TGFB* pathway genes are known to play a functional role in asthmatic patients, they have rarely been identified through GWASs, and to our knowledge, variation within or near *SMAD2* has not been previously associated with asthma. Interestingly, 2 previous meta-analyses[23,46] identified an association between *SMAD3*, the cytosolic heterodimeric partner of *SMAD2*, and asthma in Europeans and European Americans. Here variation in *SMAD3* was not significantly associated with asthma through admixture mapping in either Latinos or African Americans, nor did we find that *SMAD3* was differentially expressed in the whole blood of asthma cases versus control subjects among Puerto Ricans.

Fine mapping implicated multiple noncoding variants 5′ of *SMAD2* in asthma susceptibility, including several variants in the proximity of regulatory elements. However, none of the variants were correlated with expression of either *SMAD2* or *ZBTB7C* in peripheral blood, and thus we find no evidence that these variants affect the transcription of either of these genes in *cis*. This might indicate that the variants contribute to asthma susceptibility through a mechanism independent of *SMAD2* or *ZBTB7C* expression or a lack of statistical power to identify eQTLs in the subset of subjects examined or that the variants identified regulate the expression of *SMAD2*, *ZBTB7C*, or both in a tissue other than peripheral blood.

Despite our inability to identify *cis*-eQTLs, our findings support the role of differential regulation of *SMAD2* gene expression in the blood in asthmatic patients, with a 4-fold increase in the odds of asthma for low versus high *SMAD2* expression. Thus the disconnect between our genetic and gene expression associations at 18q21 might reflect a more complex and multilayered contribution of *SMAD2* on asthma susceptibility. The effects of *SMAD2* expression are seen jointly when pooled across 3 ethnic groups, with the largest effect observed in Puerto Ricans and those with early asthma onset (≤5 years of age).

The potential for using *SMAD2* expression as a biomarker for asthma and asthma severity in peripheral blood is attractive because of its ease of collection compared with other tissues. Including peripheral blood *SMAD2* expression levels improved the explanatory power of statistical models of asthma exacerbations beyond the use of traditional variables collected in the clinic. By estimating white blood cell counts for a subset of our data, we ruled out

potential confounding by cell type, another attractive feature of a biomarker measured in peripheral blood. Furthermore, we found that *SMAD2* expression was also significantly lower in atopic asthma cases versus control subjects in isolated $CD4^+$ lymphocytes in German subjects. This provides additional support that the difference we observe is not due to differences in white blood cell counts and that *SMAD2* expression might be a useful biomarker in subjects of European ancestry.

Our results uncover a novel asthma locus in Hispanics/Latinos that replicates, is the heterodimeric partner of another well-replicated asthma locus, and provides additional insight through gene expression in whole blood. However, this study is not without shortcomings.

First, we deployed our admixture mapping framework in a highly diverse collection of studies with unique ancestries and environmental exposures. We cannot adequately measure the full scope of environmental exposures that can affect expression patterns.

In addition, because our gene expression measurements came from studies of childhood asthma, we only have access to 1 relevant tissue type reflecting the inflammatory nature of asthma (whole blood). Although we did not identify a consensus set of eQTL loci, the intrinsic heterogeneity of exposures across subjects can conceal associations. We note that the large Genotype-Tissue Expression Consortium[52] was also unable to identify eQTLs for *SMAD2* in whole blood. The small subset of subjects might not reflect the entire spectrum of expression patterns. However, our fine mapping does result in a suggestion of transcription factor–binding sites and/or DNAse peaks underlying the signal, suggesting a regulatory component reflecting the noncoding nature of our top hits. The novel methods, large-scale computational approach, numerous independent study designs and platforms, fine mapping, and biological relevance to disease provide a new window into the pathophysiology of asthma in Hispanic/Latino subjects.

In conclusion, an admixture mapping meta-analysis across 3902 Latinos identified a novel genome-wide significant association with ancestry at chromosome 18q21 centered on *SMAD2-ZBTB7C*, which replicated in an additional 3774 Latinos. We further implicated multiple noncoding variants 5′ of *SMAD2* and asthma susceptibility, which vary in frequency between the ancestral haplotypes of Latinos. Low peripheral blood *SMAD2* expression was associated with a 4-fold increase in the odds of asthma. In asthmatic patients low *SMAD2* expression in peripheral blood is associated with increased asthma exacerbations in Latinos.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Christopher R. Gignoux, PhD[a,b,‡,*], Dara G. Torgerson, PhD[c,*], Maria Pino-Yanes, PhD[c,d], Lawrence H. Uricchio, PhD[b,§], Joshua Galanter, MD, MAS[b,c], Lindsey A. Roth, MA[c], Celeste Eng, BS[c], Donglei Hu, PhD[c], Elizabeth A. Nguyen, BS[c], Scott

Huntsman, MS[c], Rasika A. Mathias, ScD[e], Rajesh Kumar, MD, MSPH[f], Jose Rodriguez-Santana, MD[g], Neeta Thakur, MD, MPH[c], Sam S. Oh, PhD[c], Meghan McGarry, MD[h], Andres Moreno-Estrada, MD, PhD[i], Karla Sandoval, PhD[i], Cheryl A. Winkler, PhD[j], Max A. Seibold, PhD[k], Badri Padhukasahasram, PhD[l], David V. Conti, PhD[m], Harold J. Farber, MD, MSPH[o], Pedro Avila, MD[p], Emerita Brigino-Buenaventura, MD[q], Michael Lenoir, MD[r], Kelley Meade, MD[n], Denise Serebrisky, MD[s], Luisa N. Borrell, DDS, PhD[t], William Rodriguez-Cintron, MD[u], Shannon Thyne, MD[c], Bonnie R. Joubert, PhD[v], Isabelle Romieu, MD[w], Albert M. Levin, PhD[l], Juan-Jose Sienra-Monge, MD[x], Blanca Estela del Rio-Navarro, MD[x], Weiniu Gan, PhD[y], Benjamin A. Raby, MD, MPH[z], Scott T. Weiss, MD[z], Eugene Bleecker, MD[aa], Deborah A. Meyers, PhD[aa], Fernando J. Martinez, MD[bb], W. James Gauderman, PhD[m], Frank Gilliland, MD[m], Stephanie J. London, MD, PhD[v], Carlos D. Bustamante, PhD[i], Dan L. Nicolae, PhD[cc], Carole Ober, PhD[dd], Saunak Sen, PhD[ee], Kathleen Barnes, PhD[e], L. Keoki Williams, MD, MAS[l,ff], Ryan D. Hernandez, PhD[a,b], Esteban G. Burchard, MD, MPH[a,b,c]

Affiliations

[a]Program in Pharmaceutical Sciences and Pharmacogenomics, University of California, San Francisco; [b]Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco; [c]Department of Medicine, University of California, San Francisco; [d]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid; [e]Department of Medicine, Johns Hopkins University, Baltimore; [f]Ann and Robert H. Lurie Children's Hospital of Chicago, Feinberg School of Medicine, Northwestern University, Chicago; [g]Centro de Neumologia Pediatrica, San Juan; [h]Department of Pediatrics, University of California, San Francisco; [i]Department of Genetics, Stanford University, Palo Alto; [j]Molecular Genetics Epidemiology Section, Frederick National Laboratory for Cancer Research; [k]Integrated Center for Genes, Environment, and Health, Department of Pediatrics, Division of Pulmonary and Critical Care Medicine, National Jewish Health, Denver; [l]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit; [m]Department of Preventative Medicine, University of Southern California, Los Angeles; [n]Children's Hospital and Research Center Oakland; [o]Department of Pediatrics, Section of Pulmonology, Baylor College of Medicine and Texas Children's Hospital, Houston; [p]Division of Allergy-Immunology, Feinberg School of Medicine, Northwestern University, Chicago; [q]Department of Allergy & Immunology, Kaiser Permanente-Vallejo Medical Center, Vallejo; [r]Bay Area Pediatrics, Oakland; [s]Pediatric Pulmonary Division, Jacobi Medical Center, Bronx; [t]Department of Health Sciences, Graduate Program in Public Health, Lehman College, City University of New York, Bronx; [u]Veterans Caribbean Health Care System, San Juan; [v]National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park; [w]Nutritional Epidemiology Group, International Agency for Research on Cancer, Lyon; [x]Departmento de Alergia e Inmunologia, Clinica Hospital Infantil de Mexico Federico Gomez, Mexico City; [y]Division of Lung Diseases, National Heart, Lung, and Blood Institute, Bethesda; [z]Department of Medicine, Harvard Medical School, Boston; [aa]Center for

Genomics & Personalized Medicine Research, Wake Forest University, Winston-Salem; [bb]BIO5 Institute, University of Arizona, Tucson; [cc]Physical Sciences Division, Department of Statistics, University of Chicago [dd]Department of Human Genetics, University of Chicago; [ee]Department of Preventive Medicine, University of Tennessee Health Sciences Center, Memphis; [ff]Department of Internal Medicine, Henry Ford Health System, Detroit.

Acknowledgments

Supported in part by the National Institutes of Health (NIH; AI061774, AI077439, AI079139, CA113710, DK064695, ES015794, HL078885, HL079055, HL087699, HL088133, HL104608, M01-RR00188, and MD006902); ARRA grant RC2 HL101651; the Flight Attendant Medical Research Institute (FAMRI); UCSF Chancellor's Research Fellowship, Dissertation Year Fellowship, and in part by NIH Training Grant T32GM007175 and T32HG000044 (to C.R.G.); an RWJF Amos Medical Faculty Development Award (to E.G.B.); the Sandler Foundation; the American Asthma Foundation (to E.G.B. and L.K.W.); and NHLBI K23 (K23HL111636) and NCATS KL2 (KL2TR000143; to J.M.G.). M.P.-Y. was funded by a Postdoctoral Fellowship from Fundación Ramón Areces. K.B. was supported in part by the Mary Beryl Patch Turnbull Scholar Program. R.A.M. was supported in part by the MOSAIC initiative of Johns Hopkins University. This publication was supported by the National Center for Advancing Translational Sciences, NIH, through UCSF-CTSI grant no. KL2TR000143. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN26120080001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

We thank the families and patients for their participation and the numerous health care providers and community clinics for their support and participation in all the EVE Consortium and replication studies. In particular, we thank GALA I, GALA II, and SAGE study coordinator Sandra Salazar and the recruiters who obtained the data: Duanny Alva, MD; Gaby Ayala-Rodriguez; Lisa Caine; Elizabeth Castellanos; Jaime Colon; Denise DeJesus; Blanca Lopez; Brenda Lopez, MD; Louis Martos; Vivian Medina; Juana Olivo; Mario Peralta; Esther Pomares, MD; Jihan Quraishi; Johanna Rodriguez; Shahdad Saeedi; Dean Soto; Emmanuel Viera; and Ana Taveras. Resequencing services were kindly provided through the Resequencing and Genotyping Service by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under US Federal Government contract no. HHSN268201100037C from the National Heart, Lung, and Blood Institute. Some computations were performed with the UCSF Biostatistics High Performance Computing System. Finally, the authors would like to thank Dean Sheppard, MD; David Erle, MD; and Amy J. Markowitz for helpful edits, comments and advice on early versions of the manuscript.

Abbreviations used

eQTL	Expression quantitative trait locus
GALA I	Genetics of Asthma in Latino Americans
GALA II	Genes-Environments & Admixture in Latino Americans
GWAS	Genome-wide association study
LD	Linkage disequilibrium
OR	Odds ratio
SAGE	Study of African Americans, Asthma, Genes and Environment
SKAT-O	Optimized Sequence Kernel Association Test

SMAD2	SMAD family member 2
SNP	Single nucleotide polymorphism
TDT	Transmission disequilibrium test
ZBTB7C	Zinc finger and BTB domain containing 7C

REFERENCES

1. National Health Interview Survey (NHIS). Data. 2011 Available at: http://www.cdc.gov/asthma/nhis/2011/table4-1.htm; 2011.
2. Willemsen G, van Beijsterveldt TCEM, van Baal CGCM, Postma D, Boomsma DI. Heritability of self-reported asthma and allergy: a study in adult Dutch twins, siblings and parents. Twin Res Hum Genet 2008;11:132–42. [PubMed: 18361713]
3. A catalog of published genome-wide association studies. 2013 Available at: http://www.genome.gov/gwastudies; 2013Accessed March 2013.
4. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. Nature 2011;475:163–5. [PubMed: 21753830]
5. Ober C, Hoffjan S. Asthma genetics 2006: the long and winding road to gene discovery. Genes Immun 2006;7:95–100. [PubMed: 16395390]
6. Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. Trends Genet 2009;25:489–94. [PubMed: 19836853]
7. Galanter JM, Torgerson D, Gignoux CR, Sen S, Roth LA, Via M, et al. Cosmopolitan and ethnic-specific replication of genetic risk factors for asthma in 2 Latino populations. J Allergy Clin Immunol 2011;128:37–43.e12. [PubMed: 21621256]
8. Wu H, Romieu I, Shi M, Hancock DB, Li H, Sienra-Monge JJ, et al. Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. J Allergy Clin Immunol 2010;125:321–7.e13. [PubMed: 19910030]
9. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci U S A 2011;108:11983–8. [PubMed: 21730125]
10. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264–72. [PubMed: 16554528]
11. Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, et al. Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc Natl Acad Sci U S A 2010;107(suppl 2): 8954–61. [PubMed: 20445096]
12. Choudhry S, Coyle NE, Tang H, Salari K, Lind D, Clark SL, et al. Population stratification confounds genetic association studies among Latinos. Hum Genet 2006;118:652–64. [PubMed: 16283388]
13. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc Natl Acad Sci U S A 2010;107:786–91. [PubMed: 20080753]
14. Choudhry S, Taub M, Mei R, Rodriguez-Santana J, Rodriguez-Cintron W, Shriver MD, et al. Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region. Hum Genet 2008;123:455–68. [PubMed: 18401594]
15. Cheng C-Y, Kao WHL, Patterson N, Tandon A, Haiman CA, Harris TB, et al. Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. PLoS Genet 2009;5:e1000490. [PubMed: 19461885]
16. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A 2006;103:14068–73. [PubMed: 16945910]
17. Torgerson DG, Capurso D, Ampleford EJ, Li X, Moore WC, Gignoux CR, et al. Genome-wide ancestry association testing identifies a common European variant on 6q14.1 as a risk factor for

asthma in African American subjects. J Allergy Clin Immunol 2012;130:622–9.e9. [PubMed: 22607992]

18. 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. [PubMed: 24406073]
19. Torgerson DG, Gignoux CR, Galanter JM, Drake KA, Roth LA, Eng C, et al. Case-control admixture mapping in Latino populations enriches for known asthma-associated genes. J Allergy Clin Immunol 2012;130:76–82.e12. [PubMed: 22502797]
20. 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. [PubMed: 25488688]
21. Drake KA, Torgerson DG, Gignoux CR, Galanter JM, Roth LA, Huntsman S, et al. A genome-wide association study of bronchodilator response in Latinos implicates rare variants. J Allergy Clin Immunol 2014;133:370–8. [PubMed: 23992748]
22. Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, Zakharia F, Sikora M, Contreras AV, et al. Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science 2014;344: 1280–5. [PubMed: 24926019]
23. 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. [PubMed: 21804549]
24. Myers RA, Himes BE, Gignoux CR, Yang JJ, Gauderman WJ, Rebordosa C, et al. Further replication studies of the EVE Consortium meta-analysis identifies 2 asthma risk loci in European Americans. J Allergy Clin Immunol 2012;130:1294–301. [PubMed: 23040885]
25. Hoffmann TJ, Zhan Y, Kvale MN, Hesselson SE, Gollub J, Iribarren C, et al. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. Genomics 2011;98: 422–30. [PubMed: 21903159]
26. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43,1110. [PubMed: 25431634]
27. Nishimura KK, Galanter JM, Roth LA, Oh SS, Thakur N, Nguyen EA, et al. Early-life air pollution and asthma risk in minority children. The GALA II and SAGE II studies. Am J Respir Crit Care Med 2013;188:309–18. [PubMed: 23750510]
28. Thakur N, Oh SS, Nguyen EA, Martin M, Roth LA, Galanter J, et al. Socioeconomic status and childhood asthma in urban minority youths. The GALA II and SAGE II studies. Am J Respir Crit Care Med 2013;188:1202–9. [PubMed: 24050698]
29. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30: 207–10. [PubMed: 11752295]
30. Mobini R, Andersson BA, Erjefalt J, Hahn-Zoric M, Langston MA, Perkins AD, et al. A module-based analytical strategy to identify novel disease-associated genes shows an inhibitory role for interleukin 7 receptor in allergic inflammation. BMC Syst Biol 2009;3:19. [PubMed: 19216740]
31. Pasaniuc B, Sankararaman S, Kimmel G, Halperin E. Inference of locus-specific ancestry in closely related populations. Bioinformatics 2009;25:i213–21. [PubMed: 19477991]
32. Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, et al. Fast and accurate inference of local ancestry in Latino populations. Bioinformatics 2012;28:1359–67. [PubMed: 22495753]
33. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2012.
34. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–16. [PubMed: 8447318]
35. Python Software Foundation. Available at: http://www.python.org/.

36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75. [PubMed: 17701901]
37. Schwartz AG, Wenzlaff AS, Bock CH, Ruterbusch JJ, Chen W, Cote ML, et al. Admixture mapping of lung cancer in 1812 African-Americans. Carcinogenesis 2011;32:312–7. [PubMed: 21115650]
38. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012;91:224–37. [PubMed: 22863193]
39. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics 2012;13:762–75. [PubMed: 22699862]
40. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2013.
41. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74. [PubMed: 22955616]
42. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002;12:996–1006. [PubMed: 12045153]
43. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics 2006;7:115. [PubMed: 16704732]
44. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 2007;448:470–3. [PubMed: 17611496]
45. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet 2011;88: 586–98. [PubMed: 21565292]
46. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363:1211–21. [PubMed: 20860503]
47. Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, et al. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum Biol 2012;84:343–64. [PubMed: 23249312]
48. Hansen G, McIntire JJ, Yeung VP, Berry G, Thorbecke GJ, Chen L, et al. CD4(+) T helper cells engineered to produce latent TGF-beta1 reverse allergen-induced airway hyperreactivity and inflammation. J Clin Invest 2000;105:61–70. [PubMed: 10619862]
49. Sagara H, Okada T, Okumura K, Ogawa H, Ra C, Fukuda T, et al. Activation of TGF-beta/Smad2 signaling is associated with airway remodeling in asthma. J Allergy Clin Immunol 2002;110:249–54. [PubMed: 12170265]
50. 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. [PubMed: 14617512]
51. Hough C, Radu M, Dore JJ. Tgf-beta induced Erk phosphorylation of smad linker region regulates smad signaling. PLoS One 2012;7:e42513. [PubMed: 22880011]
52. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648–60. [PubMed: 25954001]

Key messages

- Genetic ancestry at 18q21 is associated with asthma susceptibility in Latinos.
- Targeted sequencing of the admixture peak at 18q21 identified multiple noncoding variants associated with asthma. These variants are at different frequencies in the ancestral populations of Latinos.
- Low *SMAD2* expression correlates with increased susceptibility to asthma and more frequent asthma exacerbations.

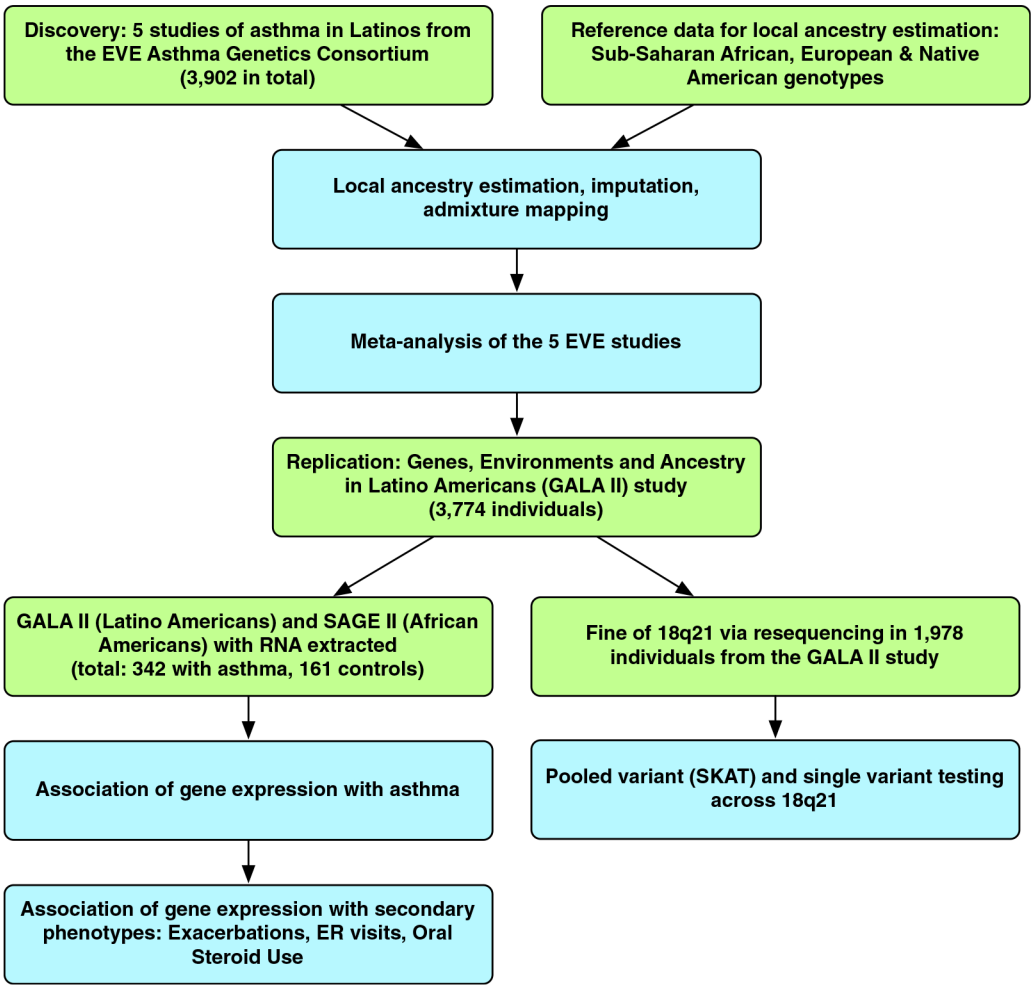

FIG 1.

Outline of the study approach. Data generation steps are in green, and analysis steps are in blue. Numbers of subjects available for analyses can be found in Table I.

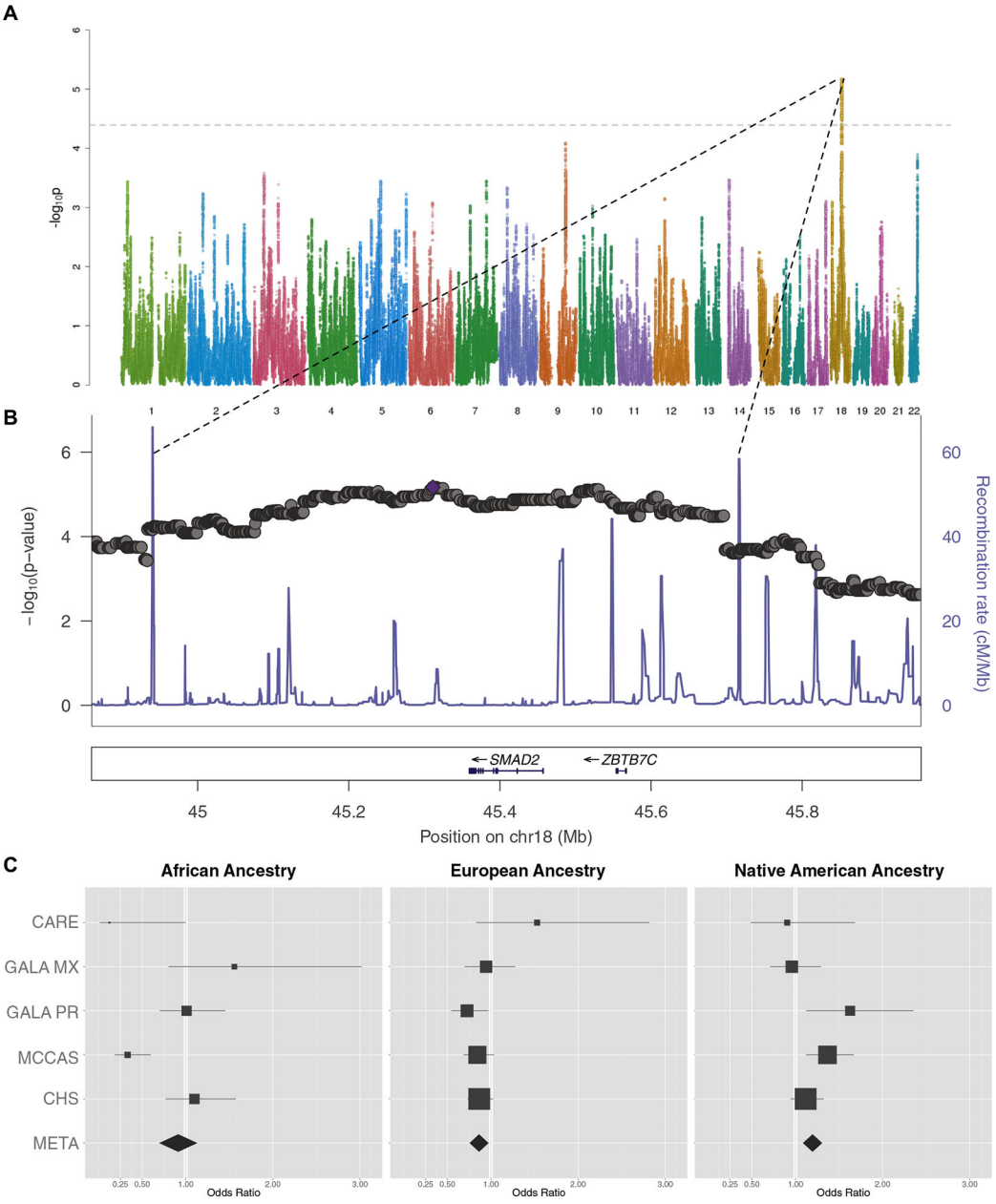

FIG 2.

Results of the admixture mapping meta-analysis. A, Admixture mapping Manhattan plot. B, Summary of admixture mapping findings across the 18q21 locus. C, Forest plots for each ancestry, including ORs and CIs, with the size of the square inversely proportional to the SE. Meta-analysis estimates through fixed-effects models are shown as diamonds, with values in Table II.

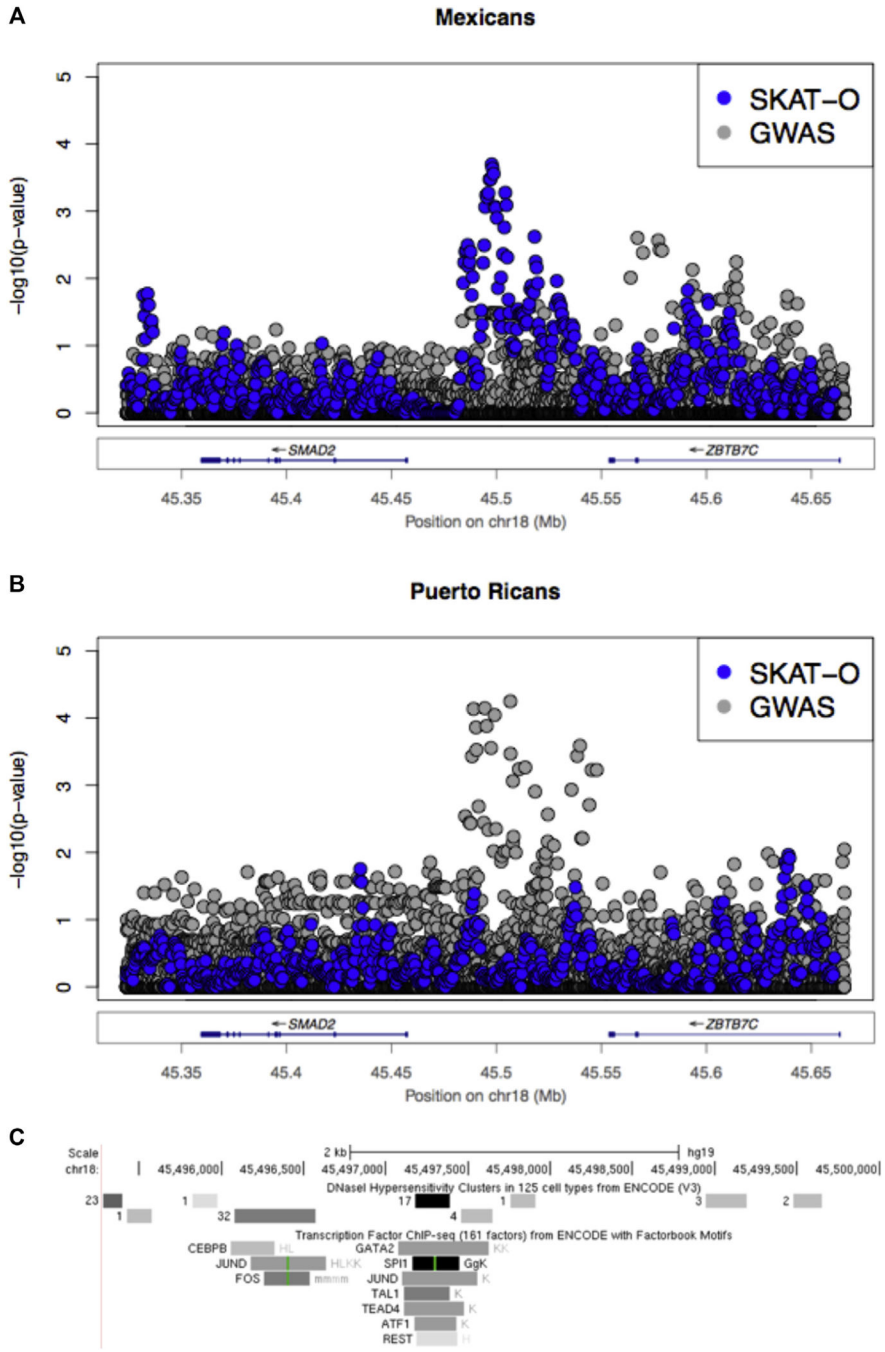

FIG 3.

Fine mapping of 18q21. **A** and **B**, pooled (SKAT-O) and individual (GWAS) variants in Mexicans (Fig 3, *A*) and Puerto Ricans (Fig 3, *B*). **C**, Regulatory elements from ENCODE,[41] with darker boxes having stronger signals and numbers indicating hypersensitive cell lines. Also, transcription factor bindings site are shown through ChIP-seq in ENCODE.[41] *Green lines* indicate the highest-scoring binding motif for the corresponding factor and most associated cell type from the UCSC Genome Browser.[42]

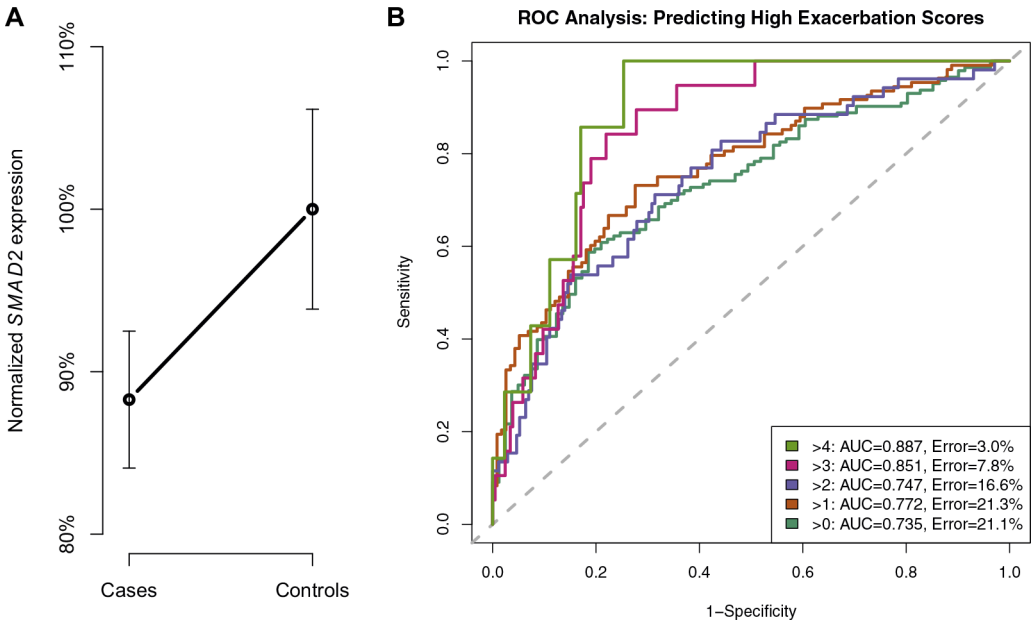

FIG 4.

Analysis of *SMAD2* expression in whole blood. **A**, Means and 95% CIs of whole blood *SMAD2* expression, which was lower in cases ($P = .002$). Neither *ZBTB7C* nor *SMAD3* showed any differential expression. **B**, Receiver operating characteristic (*ROC*) curves of high exacerbation scores in patients with asthma by using clinical variables, demographic data, and *SMAD2* gene expression accounting for age, population, spirometry, and controller medication use. AUCs and 10-fold cross-validated prediction errors are shown in legends.

TABLE I.

Baseline characteristics of studies

Study name*	Genotyping platform	Study type	Cases/ control subjects/ trios	Analytic method	Recruitment location	Male sex (%)	Age at recruitment (y)	Age of onset (y)	FEV_1†	FVC†	ΔFEV_1
EVE Consortium Hispanic/Latino studies											
CARE	Affymetrix 6.0	Trios	42	TDT/Poisson GLM	US-wide	78.8	4.8 ± 3.2	1.4 ± 1.6	90.0 ± 5.5	101.3 ± 4.5	NA
CHS	Illumina 550K, 610K	Case-control	606/792	Logistic GLM	Southern California	55.9	7.4 ± 1.7	6.8 ± 4.8	101.0 ± 12.6	100.5 ± 12.7	NA
GALA I Mexicans	Affymetrix 6.0	Case-control	252/151	Logistic GLM	Mexico City/Bay Area	44.3	15.9 ± 8.3	8.3 ± 7.7	89.7 ± 18.8	98.0 ± 18.3	19.1 ± 18.9
GALA I Puerto Ricans	Affymetrix 6.0	Case-control	277/191	Logistic GLM	Puerto Rico/New York	42.1	13.8 ± 7.0	3.4 ± 4.8	85.6 ± 15.9	98.6 ± 19.3	12.2 ± 0.2
MCCAS	Illumina 550K	Trios	492	TDT/Poisson GLM	Mexico City	58.7	9.0 ± 2.4	3.3 ± 1.4	90.5 ± 16.8	93.1 ± 16.1	NA
EVE Consortium African American studies											
Barbados	Illumina 650K	Pedigrees	382/461	Pedigree-aware case-control	Barbados	48.2	20.8 ± 12.9	8.2 ± 10.6	NA	NA	NA
GRAAD	Illumina 650K	Case-control	606/792	Logistic GLM	Maryland	44.4	24.4 ± 17.6	11.9 ± 13.2	NA	NA	NA
CAG/CSGA/ SARP	Illumina 1M	Case-control	541/451	Logistic GLM	US-wide	38.9	26.4 ± 15.8	9.8 ± 12.4	64.5 ± 20.5	98.0 ± 18.3	NA
SAPPHIRE	Affymetrix 6.0	Case-control	150/131	Logistic GLM	Detroit, Michigan	21.4	32.3 ± 13.3	10.5 ± 11.6	NA	NA	NA
GALA II Hispanic/Latino replication											
GALA II Mexicans	Affymetrix Axiom World Array 4	Case-control	586/659	Logistic GLM/SKAT	US-wide	48.4	12.8 ± 3.3	5.3 ± 3.7	96.9 ± 13.3	101.7 ± 12.8	6.3 ± 6.9
GALA II Puerto Ricans	Affymetrix Axiom World Array 4	Case-control	896/890	Logistic GLM/SKAT	US-wide	50.7	12.4 ± 3.2	2.6 ± 2.9	86.0 ± 16.4	89.7 ± 16.9	10.8 ± 8.9
GALA II Mixed/Other	Affymetrix Axiom World Array 4	Case-control	402/324	Logistic GLM/SKAT	US-wide	48.2	12.5 ± 3.2	4.4 ± 3.9	92.4 ± 15.0	97.7 ± 14.7	7.6 ± 8.7
Study populations with gene expression											

Study name*	Genotyping platform	Study type	Cases/ control subjects/ trios	Analytic method	Recruitment location	Male sex (%)	Age at recruitment (y)	Age of onset (y)	FEV_1†	FVC†	ΔFEV_1
SAGE II African Americans	TaqMan	Case-control	114/51	Logistic GLM	Bay Area	47.3	13.4 ± 3.7	3.8 ± 3.9	83.3 ± 10.4	87.8 ± 9.3	8.1 ± 5.3
GALA II Mexicans	TaqMan	Case-control	122/56	Logistic GLM	US-wide	51.7	12.5 ± 3.4	5.7 ± 4.3	95.9 ± 13.5	100.5 ± 12.2	9.1 ± 8.4
GALA II Puerto Ricans	TaqMan	Case-control	107/54	Logistic GLM	US-wide	58.1	12.6 ± 2.9	1.9 ± 2.4	89.2 ± 13.6	92.8 ± 14.8	12.4 ± 7.8

Latinos from the EVE Asthma Genetics Consortium were used for discovery, and GALA II was used for replication. GALA II and SAGE II were used for gene expression associations. Characteristics include origin of study and study type. Other descriptions from the EVE Consortium studies can be found in detail in Torgerson et al.[23]

*_CARE_, Childhood Asthma Research Network (University of Arizona); _CHS_, Children's Health Study (University of Southern California); _GALA I_, Genetics of Asthma in Latino Americans (University of California, San Francisco); _GALA II_, Genes-Environments and Admixture in Latino Americans (University of California, San Francisco);

MCCAS, Mexico City Childhood Asthma Study (National Institute of Environmental Health Sciences); _SAGE II_, Study of African Americans, Asthma, Genes and Environment (University of California, San Francisco).

†Values are presented as percent predicted based on age, sex, height, and ethnicity. For studies with bronchodilator response, reported values are before albuterol administration; for other studies, spirometric values are baseline from the time of recruitment. Spirometric values are only reported for studies in the EVE Consortium in which the majority of subjects had reliable tests.

J Allergy Clin Immunol. Author manuscript; available in PMC 2020 March 01.

TABLE II.

Ancestry associations at the 18q21 peak centered on *SMAD2*

18q21	EVE Consortium	GALA II (average)	Combined
African *P* value	.42	.16	.27
African OR (95% CI)	0.91 (0.73–1.14)	1.05 (0.98–1.13)	1.04 (0.97–1.11)
European *P* value	8.35×10^{-3}	5.83×10^{-3}	1.36×10^{-4}
European OR (95% CI)	0.86 (0.77–0.96)	0.87 (0.78–0.96)	0.86 (0.80–0.93)
Native American *P* value	1.63×10^{-3}	6.26×10^{-3}	9.15×10^{-5}
Native American OR (95% CI)	1.20 (1.07–1.34)	1.09 (1.02–1.16)	1.11 (1.05–1.17)
Overall *P* value	6.80×10^{-6}	.017 (min = 5.3×10^{-3})	2.6×10^{-7}

Summary characteristics of admixture mapping findings at the top hit in the chromosome 18q21 region. Meta-analysis of discovery and replication studies was performed assuming fixed effects of effect size estimates, and the Fisher method was used for overall likelihood ratio test meta-analysis.

J Allergy Clin Immunol. Author manuscript; available in PMC 2020 March 01.

TABLE III.

Individual allelic association testing with *P* values of less than .05 in Mexicans (and Puerto Ricans) within the top SKAT-O window

				Mexicans				Puerto Ricans				Ancestral allele frequency (A1)		
SNP	**Position**	**A1**	**A2**	**Frequency A1**	**OR**	**SE**	**P value**	**Frequency A1**	**OR**	**SE**	**P value**	**Native American**	**African (YRI)**	**European (CEU)**
rs76402589	45497831	A	T	0.00429	0.076	1.08	.0164	0.0255	0.71	0.223	.125	0	0.15	0
rs113382450	45499287	G	T	0.00429	0.076	1.08	.0164	0.0255	0.71	0.223	.125	0	0.15	0
rs79405582	45499892	T	G	0.00429	0.076	1.08	.0164	0.0255	0.71	0.223	.125	0	0.15	0
rs112042427	45496169	A	G	0.00429	0.076	1.08	.0164	0.0257	0.72	0.222	.147	0	0.15	0
rs75478387	45496266	T	C	0.00429	0.076	1.08	.0164	0.0257	0.72	0.222	.147	0	0.15	0
rs8099232	45496784	T	G	0.00429	0.076	1.08	.0164	0.0257	0.72	0.222	.147	0	0.15	0
rs7236490	45498260	G	T	0.00468	0.082	1.07	.0194	0.0255	0.71	0.223	.125	0	0.15	0
rs59002988	45496882	T	G	0.0616	1.4	0.172	.0488	0.0282	2.2	0.227	4.6×10^{-4}	0.102	0.07	0.01

The ancestral frequency of A1 for Native American ancestry was inferred from GALA II subjects who were homozygous for Native American ancestry at 18q21 and for subjects of African and European ancestry from YRI and CEU from the 1000 Genomes Project, respectively.

TABLE IV.

Associations of *SMAD2* gene expression with asthma and morbidity outcomes

Case-control phenotypes: <127% of mean control expression*	**No, (cases vs control subjects)**		**OR (95% CI)**	***P* value**
All 3 populations	343 vs 161		**3.93 (2.12–7.28)**	**<.001**
All populations, heterogeneity	343 vs 161		—	.34
Subset, adjusting for blood cell counts	74 vs 32		**22.4 (1.7–294)**	**.018**
$CD4^+$ lymphocytes†	15 vs 8		**1.9 (1.3–2.8)**	**.02**
Exacerbation score per 10% decrease in expression compared with control subjects‡		**No.**	**OR (95% CI)**	***P* value**
All 3 populations		333	**1.06 (1.01–1.12)**	**.011§**
All populations, heterogeneity		333	—	.08
Subset, adjusting for blood cell counts		74	**1.16 (1.01–1.34)**	**.036**

Dichotomous outcomes use a best-fit cut point of 127% of healthy control *SMAD2* gene expression to estimate the OR. Associations with exacerbations are per 10% increase in *SMAD2* expression compared with control subjects. Values in boldface indicate statistical significance.

*Adjusting for age, sex, recruitment site, and ancestry.

†*P* values are given based on the Fisher exact test, but ORs and CIs are given based on logistic regression (*P* = .01) to be more robust. Both tests used the same data.

‡Levels 0 to 9, ordered logistic regression, and *P* value from likelihood ratio test.

§*P* value of less than .05 for random-effects meta-analysis as well.