ORIGINAL ARTICLE

Identification of Four Novel Loci in Asthma in European American and African American Populations

Berta Almoguera[1], Lyam Vazquez[1], Frank Mentch[1], John Connolly[1], Jennifer A. Pacheco[2], Agnes S. Sundaresan[3], Peggy L. Peissig[4], James G. Linneman[4], Catherine A. McCarty[5], David Crosslin[6], David S. Carrell[7], Todd Lingren[8], Bahram Namjou-Khales[8], John B. Harley[8,9], Eric Larson[7], Gail P. Jarvik[6], Murray Brilliant[4], Marc S. Williams[3], Iftikhar J. Kullo[10], Erik B. Hysinger[1], Patrick M. A. Sleiman[1,11], and Hakon Hakonarson[1,11]

[1]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [2]Northwestern University Feinberg School of Medicine, Chicago, Illinois; [3]Geisinger Health System, Danville, Pennsylvania; [4]Marshfield Clinic Research Foundation, Marshfield, Wisconsin; [5]Essentia Institute of Rural Health, Duluth, Minnesota; [6]University of Washington Medical Center, Seattle, Washington; [7]Group Health Research Institute, Seattle, Washington; [8]Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [9]U.S. Department of Veterans Affairs Medical Center, Cincinnati, Ohio; [10]Mayo Clinic, Rochester, Minnesota; and [11]Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

Abstract

Rationale: Despite significant advances in knowledge of the genetic architecture of asthma, specific contributors to the variability in the burden between populations remain uncovered.

Objectives: To identify additional genetic susceptibility factors of asthma in European American and African American populations.

Methods: A phenotyping algorithm mining electronic medical records was developed and validated to recruit cases with asthma and control subjects from the Electronic Medical Records and Genomics network. Genome-wide association analyses were performed in pediatric and adult asthma cases and control subjects with European American and African American ancestry followed by metaanalysis. Nominally significant results were reanalyzed conditioning on allergy status.

Measurements and Main Results: The validation of the algorithm yielded an average of 95.8% positive predictive values for both cases and control subjects. The algorithm accrued 21,644 subjects (65.83% European American and 34.17% African American). We identified four novel population-specific associations with asthma after metaanalyses: loci 6p21.31, 9p21.2, and 10q21.3 in the European American population, and the *PTGES* gene in African Americans. *TEK* at 9p21.2, which encodes TIE2, has been shown to be involved in remodeling the airway wall in asthma, and the association remained significant after conditioning by allergy. *PTGES*, which encodes the prostaglandin E synthase, has also been linked to asthma, where deficient prostaglandin E_2 synthesis has been associated with airway remodeling.

Conclusions: This study adds to understanding of the genetic architecture of asthma in European Americans and African Americans and reinforces the need to study populations of diverse ethnic backgrounds to identify shared and unique genetic predictors of asthma.

Keywords: asthma; genetics; genome-wide association study

(Received in original form April 26, 2016; accepted in final form September 8, 2016)

The Electronic Medical Records and Genomics Network was initiated and funded by National Human Genome Research Institute through the following grants: U01HG8657 and U01HG006375 (Group Health Cooperative/University of Washington), U01HG8666 and U01HG006828 (Cincinnati Children's Hospital Medical Center), U01HG6379 and U01HG006379 (Mayo Clinic), U01HG8679 and U01HG006382 (Geisinger Clinic), U01HG8684 and U01HG006830 (The Children's Hospital of Philadelphia), U01HG8673 and U01HG006388 (Northwestern University), U01HG006389 (Essentia Institute of Rural Health, Marshfield Clinic Research Foundation, and Pennsylvania State University), and U01HG8701 and U01HG006385 (Vanderbilt University Medical Center serving as the Coordinating Center). Generation of the Electronic Medical Records and Genomics Imputed merged Phase I and Phase II dataset was supported by the following awards: U01HG004438 (CIDR) and U01HG004424 (the Broad Institute). The study was additionally funded by an Institutional Development Award to the Center for Applied Genomics from The Children's Hospital of Philadelphia.

Author Contributions: H.H., P.M.A.S., E.B.H., B.A., L.V., F.M., and J.C. conceived and designed the study. H.H., P.M.A.S., J.B.H., E.L., G.P.J., M.B., M.S.W., and I.J.K. obtained financial support. A.S.S., P.L.P., J.G.L., J.A.P., C.A.M., D.C., D.S.C., T.L., B.N.-K., and J.B.H. contributed materials and/or analysis tools. B.A., L.V., F.M., E.B.H., P.M.A.S., A.S.S., P.L.P., J.G.L., J.A.P., C.A.M., D.C., T.L., and B.N.-K. performed data analysis. B.A., L.V., F.M., J.C., E.B.H., P.M.A.S., and H.H. wrote the manuscript, and all authors reviewed and revised the paper.

Correspondence and requests for reprints should be addressed to Patrick M. A. Sleiman, Ph.D., Center for Applied Genomics, The Children's Hospital of Philadelphia, Leonard Madlyn Abramson Research Center, 3615 Civic Center Boulevard, Suite 1216E, Philadelphia, PA 19104. E-mail: sleimanp@email.chop.edu

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

Am J Respir Crit Care Med Vol 195, Iss 4, pp 456–463, Feb 15, 2017

Originally Published in Press as DOI: 10.1164/rccm.201604-0861OC on September 9, 2016
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: Despite significant advances in the knowledge of the genetic architecture of asthma, loci identified to date do not fully explain the ancestral disparities observed across populations.

What This Study Adds to the Field: Using a phenotyping algorithm for asthma based on electronic medical record data mining and independently validated across multiple study sites, we have identified four novel asthma susceptibility loci: 6p21.31, 9p21.2, and 10q21.3 in the European American population; and the prostaglandin synthase E gene (*PTGES*) at 9q34.11 in African Americans. Biologic support exists for the leading candidate genes at the 9p21.2 and 9q34.11 loci *TEK*, encoding the endothelial tyrosine kinase, and *PTGES*, both of which have been linked with asthma in *in vitro* and *in vivo* models. However, this is the first report of genetic variants in these genes contributing to the susceptibility to asthma. The advantage of using a phenotyping algorithm over traditional cross-sectional studies includes a more accurate and objective definition of the asthma diagnosis based on *International Classification of Diseases, 9th Edition* codes and the prescription of asthma medications, with more comprehensive information on the phenotype. This approach has allowed us to uncover novel loci involving pathways previously suspected to contribute to asthma susceptibility and suggests that novel therapies may be developed to more effectively treat the underlying causes of this disease.

Asthma is a common chronic disease of the airways characterized by recurrent episodes of wheezing, coughing, and shortness of breath. In the United States, it is estimated that at least 22.9 million people suffer from the condition, with significant ancestral disparities regarding prevalence, severity, and treatment outcomes across populations (1, 2). Compared with European Americans (EAs), African Americans (AAs) have higher lifetime prevalence of asthma (14.09% vs. 12.41%) (3), higher morbidity, with twice the hospitalization and 1.5 times the intensive care unit admission or intubation (4), and 7.6 times the death rate (5).

Asthma is highly heritable (6). It is well established that genetic risk factors vary within and between diverse populations (7, 8). Over the past decade, several population-specific and shared genetic risk factors have been identified through genome-wide association studies (GWAS). These loci include the *IL1R1*/*IL18R1* region at 2q12.1 (9, 10), the cytokine cluster at 5q31.1 (11, 12), and the innate immune mediator *TSLP* (5q22.1) (10, 13, 14); the HLA region at 6p21.32-33 (11–13), *IL33* at 9p24.4 (9, 10, 15), *SMAD3* at 15q22.33 (9, 15), 17q21-12 (*ORMDL3/GSMDB*) (16, 17), *DENND1B* (1q31) (18), and *CDHR3* (7q22.3) (19). However, despite the significant advances in knowledge of the genetic architecture of asthma, specific contributors to the observed variability in the burden of asthma between populations remain to be discovered.

In this study, we conducted a GWAS in patients with asthma of EA and AA descent recruited from the Electronic Medical Records and Genomics (eMERGE) network, using an algorithm based on electronic medical record (EMR) mining. The eMERGE network is a National Human Genome Research Institute–funded consortium that comprises a diverse cohort of more than 57,000 participants linked to EMRs for phenotype mining from nine participating sites in the United States (20). The large number of study participants and diversity of the eMERGE sites provide an excellent means for investigating the genetic architecture of asthma across populations of different ancestry.

Methods

Study Subjects and Definition of the Asthma Phenotype

Pediatric asthma cases and control subjects from The Children's Hospital of Philadelphia (CHOP) were selected from the biorepository at the Center for Applied Genomics, which has a collection of more than 160,000 samples including more than 90,000 internal pediatric samples genotyped using standard genome-wide arrays and linked to subjects' EMRs. All subjects have consented to analyses and EMR mining from the full longitudinal record, which has a mean duration greater than 5.5 years per subject. Mean age of these subjects is 11 years and 47% are of EA ancestry, 43% are AAs, and 10% from other ancestry groups (20).

The adult asthma cohort was collected from seven sites from the eMERGE Network (*see* Table E1 in the online supplement). Characteristics of the sample collected at each site can be found in detail (20). For all sites except Mount Sinai, EA ancestry represented more than 80% of the subjects and the mean age of cases was greater than 48 years.

The project was approved by the institutional review boards at CHOP and each of the eMERGE participating sites. Written informed consent was obtained from each participant in accordance with institutional requirements and the Declaration of Helsinki Principles.

Asthma cases and control subjects were identified using an EMR-based algorithm developed at CHOP, similar to Pacheco and colleagues (21), and validated and implemented by the seven eMERGE sites.

The inclusion criteria for cases included being greater than or equal to 4 years and either having received greater than or equal to one prescriptions of asthma-related medications (*see* Table E2) and diagnosed with greater than or equal to one of *International Classification of Diseases, 9th Edition* (*ICD9*) codes 493.00–493.92 in greater than or equal to two in-person visits, on separate calendar days; or having greater than or equal to three in-person visits in any 12-month period, on separate calendar days, with affirmative mentions of "wheezing" or "asthma" in EMR notes. The algorithm was also designed to identify subjects with allergic asthma, defined as individuals having either *ICD9* codes for extrinsic asthma (493.00, 493.01, or 493.02), or greater than or equal to two *ICD9* codes related to allergies (*see* Table E3) on separate calendar days. Exclusion criteria used for cases included having any *ICD9* codes related to respiratory conditions or organ transplantation (*see* Table E4).

Subjects were selected as controls if they had no *ICD9* code for asthma, no history of asthma medications, no positive confirmation of "wheezing" or "asthma" in the patient charts, and no exclusionary *ICD9* codes (*see* Table E4).

The validity of the algorithm was initially assessed on a sample of 100 randomly selected EMRs at CHOP revealing some modification requirements: several prescriptions for medications not included

in the initial algorithm were added, and allergy and *ICD9* codes 416.xx and 748.5 were added as exclusions for control subjects (*see* Table E4). Consequently, the algorithm was revised to its current form, rerun, and formally validated to generate positive predictive values (PPVs). For the validation, case records were searched for a positive diagnosis of asthma based on physician diagnosis, coughing or wheezing symptoms, and clinical bronchodilator response where available. Similarly, control subjects were reviewed for the absence of clinical evidence of asthma. Evidence of exclusionary *ICD9* codes was also assessed.

Internal validation at CHOP consisted of manual chart review of two independent samples: a first sample of 250 randomly selected subjects (76 cases and 174 control subjects) and a second sample of 301 asthma cases. External validation of the phenotyping algorithm included 170 subjects: 100 from Cincinnati Children's Hospital Medical Center (50 cases, 50 control subjects), 20 from Geisinger Clinic (10 cases, 10 control subjects), and 50 from Marshfield (25 cases, 25 control subjects). PPVs with 95% confidence interval were calculated for cases and control subjects at each site.

Genome-Wide Genotyping, Imputation, and Population Stratification Assessment

Genotyping was performed at each eMERGE site using Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA) platforms (*see* Table E5). Imputation of the CHOP cohort was conducted at the Center for Applied Genomics. The eMERGE Coordinating Center (Pennsylvania State University) led the imputation of the adult dataset.

A standard preimputation quality control of the genotyping data was implemented at the individual and single-nucleotide polymorphism (SNP) level using pairwise identity-by-descent values (22). Cryptic relatedness and duplicated samples were also assessed by pairwise identity-by-descent values and removed. Haplotypes were phased using SHAPEIT version 2 (O. Delaneau, Department of Statistics, University of Oxford, Oxford, UK) (23). Genome-wide imputation of the pediatric samples was performed with the IMPUTE2 package (B. N. Howie, Department of Statistics, University of Oxford, Oxford, UK) (24) using the cosmopolitan reference panel from the 1,000 Genomes Project, which included 1,092 samples from multiple ancestry groups.

Genotype imputation in the adult cohort was performed as described (25), using SHAPEIT (23) and IMPUTE2 (24) with the cosmopolitan reference panel from the 1,000 Genomes Project.

Population ancestry was determined by principal component analysis using Eigenstrat 3.0 (A. Price, Department of Genetics, Harvard Medical School, Boston, MA) (26) on the genotyping data in the pediatric samples and the imputed data in the adults (using 35,324 SNPs common to all Illumina and Affymetrix genotyping platforms). Samples were separated into AAs and EAs based on the principal components, which were regenerated for each cohort and included as covariates to control for population stratification in each individual analysis.

Association Analysis

We performed four sets of GWAS analyses in the pediatric and adult datasets, including each ancestry group separately, and we then reanalyzed nominally associated SNPs conditioning on allergy status. For the GWAS, we used SNPTEST (J. Marchini, Department of Statistics, University of Oxford, Oxford, UK) (27) to take into account the genotype uncertainty introduced by the imputation. An additive model was applied to the genotype dosages generated by IMPUTE2 including the proportion of missing data as a covariate. The first 10 principal components and sex were included as covariates in the analysis of the pediatric and the adult cohort and because the imputation at the eMERGE Coordinating Center was conducted for each genotyping platform and site separately, chip type and site were also used as covariates in the adults. Poorly imputed variants ("info" score $\leqslant 0.7$), variants with a minor allele frequency less than or equal to 1%, and variants not meeting Hardy-Weinberg equilibrium in control subjects ($P < 5 \times 10^{-8}$) were removed from the analysis. Then, METAL (C. J. Willer, Department of Biostatistics, University of Michigan, Ann Arbor, MI) was used for the metaanalyses of the EAs, AAs, and the combined sample, and also in the pediatric and adult data sets, using an inverse variance fixed-effects method, with control for genomic inflation. Additionally, genomic inflation factors for each individual GWAS and metaanalyses were calculated.

A cis-expression quantitative trait locus (eQTL) effect was investigated for all significant variants and variants in linkage disequilibrium (LD; $r^2 > 0.6$) by mining HaploReg version 4 (L. D. Ward, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology and The Broad Institute of MIT and Harvard, Cambridge, MA) (28) and the NCBI Genotype-Tissue Expression version 6 (J. Lonsdale, National Disease Research Interchange, Philadelphia, PA) (29) in all available tissues.

Results

Validation and Implementation of the EMR-based Asthma Algorithm at CHOP and the eMERGE Sites

The internal validation of the algorithm by manual chart review yielded PPVs (95% confidence interval) of 97.07% (94.66–98.45%) and 99.4% (96.8–100%) at CHOP, and the external PPVs amounted to 96% (85.14–99.30%) and 82% (68.08–90.95%) at Cincinnati, 100% (65.54–100%) for both at Geisinger, and 96% (77.67–99.79%) for both at Marshfield Clinic, for cases and control subjects, respectively.

The implementation of the algorithm in the Center for Applied Genomics' biorepository identified 11,784 asthma cases and control subjects. After quality control filtering, 215 samples were removed because of low genotyping call rate and 446 because of cryptic relatedness. Principal component analysis on the remaining 11,569 samples classified 5,680 subjects as AA (49.10%), 5,414 as EA (46.80%), and 474 belonging to other ancestry groups (4.10%). Only individuals in the EA and AA clusters were included in the analysis. The mean age of the final sample ($\pm$SD) was 14.02 $\pm$ 7.7 years, 50.95% of whom were male.

The number of subjects accrued across the eMERGE sites was 11,929, of whom 40.31% were males and with a mean age ($\pm$SD) of 64.52 $\pm$ 17.52 years. Principal component analysis identified 8,970 subjects as EAs (75.19%), 2,025 as AAs (17.31%), and 934 were classified as "other" (7.49%). Only EA and AA clusters were included in the analysis (Table 1).

The algorithm was also used to capture the allergy status of the asthma cases. Among the 5,304 pediatric and adult cases, 69.42% had allergic asthma: 79.1% and 51.2% in AA children and adults, respectively, and 84.6% and 42.7% in the EA pediatric and adult cohorts, respectively (Table 1).

Table 1. Participants of the Study by Site and Ancestry Group

Cohort	African Americans			European Americans			Total
	Control Subjects	Cases (Allergy)	Total	Control Subjects	Cases (Allergy)	Total	
CHOP	2,993	2,379 (1,881)	5,372	4,097	1,180 (998)	5,277	10,649
Adults eMERGE	1,367	658 (337)	2,025	7,878	1,092 (466)	8,970	10,995
Total	4,360	3,037 (2,218)	7,397	11,975	2,272 (1,464)	14,247	21,644

Definition of abbreviations: CHOP = The Children's Hospital of Philadelphia; eMERGE = Electronic Medical Records and Genomics.
The number of cases with allergic asthma is indicated in parentheses.

Metaanalyses Reveal Three Novel Population-Specific Susceptibility Loci and Confirm 5q22.1 as a Shared Risk Locus for Asthma across Populations

We performed GWAS in each of the four cohorts separately and then conducted metaanalyses in the combined sample of 21,644 subjects, in each of the ancestry groups and in the pediatric and adult samples separately (top results with $P \leqslant 10^{-6}$ in Tables E6–E10).

Genomic inflation factors were 1.015 and 1.056 for the pediatric EAs and AAs, respectively; 1.021 and 1.036 for the adult EAs and AAs, respectively; and 1.015, 1.007, 0.996, 1.013, and 1.000 for the metaanalysis of the combined, EA, AA, pediatric, and adult cohorts, respectively, indicating no population stratification (*see* Figure E1).

Two genome-wide significant association signals were identified in the EA-only metaanalysis, consisting of 12 SNPs mapping to two separate loci, including 9p21.2 and 6p21.31 (Table 2, Figure 1; *see* Figure E2 and Table E11). The 9p21.2 locus, which included four low-frequency SNPs in LD, spanned three genes: *TEK* (TEK tyrosine kinase, endothelial), *EQTN* (equatorin, sperm acrosome associated), and *MOB3B* (MOB kinase activator 3B). Notably, the significant variants are in LD with a missense variant in *TEK* (rs682632; Q346P; $P = 4.38 \times 10^{-6}$). Fine mapping of the region evidenced no eQTL effects in any of the tissues relevant to asthma but effects in other tissues were observed ($P < 0.05$) (*see* Table E12).

The second significant locus in EAs mapped to 6p21.31 and included eight SNPs in LD (Table 2; *see* Figure E2) in the intergenic region between *GRM4* (glutamate receptor, metabotropic 4) and *HMGA1* (high mobility group AT-hook 1), all variants with slight eQTL effects on *HMGA1* in lung (lowest $P = 4.7 \times 10^{-3}$) (*see* Table E12).

The AA-only metaanalysis led to the identification of one genome-wide significant SNP in the intronic region of the prostaglandin E synthase gene (*PTGES*), on chromosome 9 (rs11788591; $P = 4.45 \times 10^{-8}$; P value_pediatric $= 2.22 \times 10^{-5}$; P value_adult $= 2.05 \times 10^{-4}$) (Figure 1; *see* Figure E2 and Table E11).

We identified the previously reported SNP rs1837253 in 5q22.1 reaching genome-wide significance in the metaanalysis of the combined sample ($P = 4.22 \times 10^{-8}$; P value_EA $= 1.86 \times 10^{-4}$; P value_AA $= 4.80 \times 10^{-4}$) (*see* Table E11).

Besides 5q22.1, we also aimed to assess the association of other known asthma SNPs/loci in our sample (*see* Table E13). rs3771166 at 2q12.1, rs2244012 at 5q31.1, rs2381416 at 9p24.1, and all reported SNPs at 17q12–21.1 showed nominal significance in the metaanalysis of the combined sample ($P < 5 \times 10^{-3}$ in all cases). Other loci, such as 11q13.5 ($P = 4.73 \times 10^{-5}$) and *CLEC16A* ($P = 7.22 \times 10^{-4}$), also showed association with asthma in our study.

Because asthma is a disease typically occurring during childhood and the pediatric cohort constitutes a more homogeneous population in terms of genetic influence, we also ran metaanalyses in the pediatric and the adult population separately. No significant results were observed in the analysis of the pediatric sample (*see* Table E9). However, a haplotype in the intronic region of *GSDMB* was among the top

Table 2. Genome-Wide Significant Associations with Asthma after Metaanalysis of the Pediatric and Adult European American Populations

Marker	Genomic Position (*bp*)	Locus	Closest Gene (Distance) (*kb*)	r^2	Frequency of Risk Allele	β (SE)	*P* Value		
							All	Pediatrics	Adults
rs72721168	27308288	9p21.2	*EQTN* (11kb)	1	0.959	−0.602 (0.098)	7.02×10^{-10}	2.62×10^{-4}	3.01×10^{-7}
rs72721166	27304548	9p21.2	*EQTN* (7.4kb)	0.96	0.038	0.599 (0.102)	3.83×10^{-9}	1.58×10^{-4}	3.60×10^{-6}
rs72721164	27300439	9p21.2	*EQTN* (33kb)	0.96	0.038	0.597 (0.101)	4.06×10^{-9}	1.55×10^{-4}	3.92×10^{-6}
rs72721158	27275906	9p21.2	*EQTN* (21kb)	0.84	0.036	0.551 (0.100)	4.04×10^{-8}	4.28×10^{-4}	1.62×10^{-5}
rs1776883	34156444	6p21.31	*GRM4* (33kb)	1	0.472	−0.222 (0.038)	5.29×10^{-9}	6.50×10^{-6}	1.02×10^{-4}
rs1776888	34158331	6p21.31	*GRM4* (35kb)	0.96	0.531	0.219 (0.038)	7.63×10^{-9}	7.35×10^{-6}	1.27×10^{-4}
rs1776881	34156323	6p21.31	*GRM4* (33kb)	0.97	0.53	0.218 (0.038)	7.96×10^{-9}	7.14×10^{-6}	1.39×10^{-4}
chr6:34158205:I (rs34754950)	34158205	6p21.31	*GRM4* (35kb)	0.99	0.467	−0.217 (0.038)	1.09×10^{-8}	9.45×10^{-6}	1.47×10^{-4}
rs1776886	34157113	6p21.31	*GRM4* (34kb)	0.85	0.512	0.218 (0.039)	1.90×10^{-8}	7.21×10^{-6}	3.12×10^{-4}
rs1776885	34156978	6p21.31	*GRM4* (34kb)	0.98	0.477	−0.212 (0.038)	2.56×10^{-8}	1.78×10^{-5}	2.01×10^{-4}
rs1776889	34159166	6p21.31	*GRM4* (36kb)	0.84	0.498	−0.212 (0.038)	2.64×10^{-8}	5.28×10^{-5}	8.11×10^{-5}
rs1776884	34156970	6p21.31	*GRM4* (34kb)	0.93	0.557	0.209 (0.038)	3.78×10^{-8}	2.41×10^{-4}	2.72×10^{-5}

r^2 of all associated single-nucleotide polymorphisms with the top single-nucleotide polymorphism included. Effect sizes (β) and SE for each individual study along with the *P* value for the Cochran test of heterogeneity is included in Table E11.

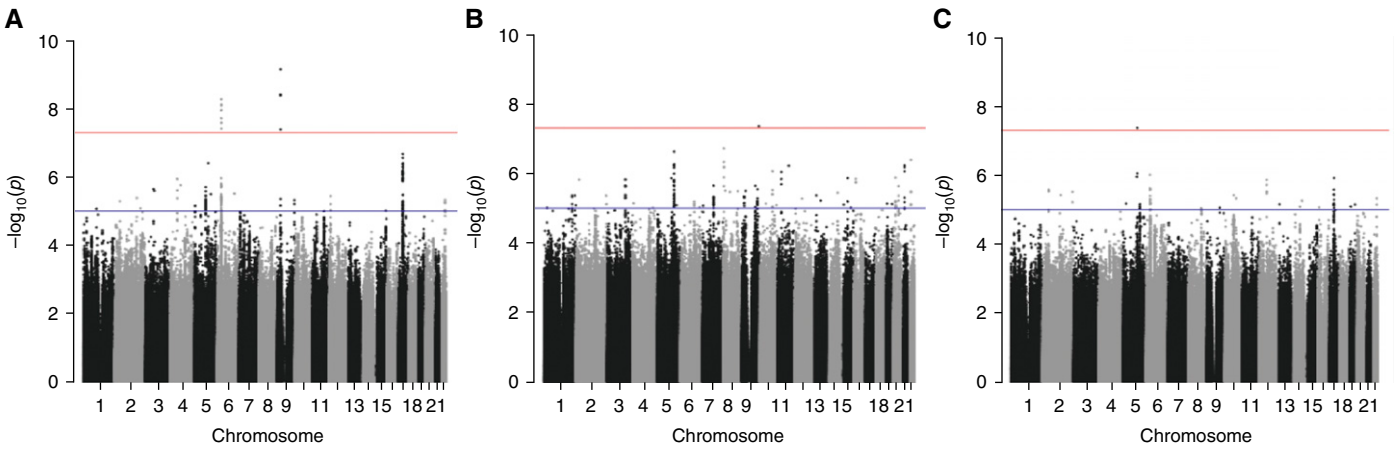

Figure 1. Manhattan plots of the genome-wide association study conducted in European Americans (*A*), African Americans (*B*), and the combined sample (*C*), showing the four loci significantly associated with asthma: 6p21.31 and 9p21.2 in European Americans, *PTGES* in locus 9q34.11 in African Americans, and 5q22.1 in the combined analysis. The threshold for genome-wide significance, $P \leqslant 5.0 \times 10^{-8}$, is shown as a *red line*. A suggestive threshold of 1.0×10^{-5} is shown as a *blue line*.

associations ($P = 3.44 \times 10^{-6}$). In the adult-only analysis, a common intronic variant in *PSORS1C1* (psoriasis susceptibility 1 candidate 1) reached genome-wide significance (rs3095318; $P = 1.61 \times 10^{-11}$) (*see* Figure E2 and Table E10). The variant is associated with a reduced expression of *PSORS1C1* in lung tissue ($P = 9.4 \times 10^{-6}$).

Conditioning on Allergy Status Indicates rs72721168 at 9p21.2 Is a Risk Factor for Asthma in EAs Independent of Allergy and Identifies Two Additional Loci

To determine if the observed associations were driven by asthma specifically or more broadly associated with an allergic phenotype we reanalyzed all SNPs that reached nominal significance ($P < 5 \times 10^{-3}$) adjusting for allergy status in both EAs (n = 41,835 variants) and AAs (n = 67,774 variants) followed by metaanalyses in the two populations and in the combined sample.

Adjusting the analysis by allergy, the 9p21.2 locus and three low-frequency SNPs on chromosome 10 reached genome-wide significance in EAs (Table 3; *see* Figure E2). The region includes the genes jumonji domain containing 1C (*JMJD1C*) and receptor accessory protein 3 (*REEP3*) at 10q21.3. All three variants are associated with decreased expression of *JMJD1C* in lung tissue ($P = 0.029$) in the Genotype-Tissue Expression data.

Discussion

In this study, we present the results from metaanalyses of asthma in individuals of EA and AA ancestry. We replicated the known association of 5q22.1 with asthma and we identified four novel loci at 6p21.31, 9p21.2, and 10q21.3 in EAs and *PTGES* in AAs.

The strongest association found in this study was with variants at 9p21.1 and asthma in the EAs. The most plausible candidate gene in the region is *TEK*, which encodes the endothelial tyrosine kinase, commonly referred to as TIE2 and that plays a central role in the regulation of the development and remodeling of the vascular system (30). TIE2 has been involved in remodeling the airway wall in patients with asthma and models of chronic asthma in mice (31, 32), evidenced by the correlation between the degree of airway obstruction and expression of TIE2 receptors in the respiratory epithelium in mice; and increased levels of TIE2 ligands, angiopoietins 1 and 2, in the sputum of patients with asthma has been reported (31). Despite the importance of TIE2 signaling in asthma, to our knowledge this is the first report of genetic variants in *TEK* contributing to the susceptibility to asthma. Furthermore, this was the only locus where the association remained genome-wide significant after adjusting for allergy status of the asthma cases, indicating that *TEK*

Table 3. Genome-Wide Significant Associations in the Metaanalysis of Pediatric and Adult European American Subjects Adjusted by Allergy Status of the Cases

Marker	Genomic Position (*bp*)	Locus	Closest Gene (Distance) (*kb*)	r^2	Frequency of Risk Allele	β (SE)	*P* Value		
							All	Pediatrics	Adults
rs72721168	27308288	9p21.2	*EQTN* (11kb)	—	0.957	−0.927 (0.165)	1.97×10^{-8}	2.44×10^{-3}	1.72×10^{-6}
rs75446656	65100016	10q21.3	*JMJD1C* (intron 2)	1	0.036	0.970 (0.176)	3.60×10^{-8}	4.94×10^{-7}	3.64×10^{-4}
rs35632171	65109502	10q21.3	*JMJD1C* (intron 2)	1	0.964	−0.970 (0.176)	3.65×10^{-8}	4.98×10^{-7}	3.67×10^{-4}
rs36080042	65426785	10q21.3	*REEP3* (68kb)	0.6	0.036	0.964 (0.176)	4.70×10^{-8}	1.13×10^{-7}	6.99×10^{-4}

r^2 of all associated single-nucleotide polymorphisms with the top single-nucleotide polymorphism included. Table E14 shows the *P* values of the variants significantly associated with asthma in the nonadjusted analysis and the analysis adjusting by the allergy status.

contributes to the risk of asthma independent of allergy status.

The second associated locus in the EAs mapped to a 20-kb LD block at 6p21.31 flanked by *GRM4* and *HGMA1*, the latter being expressed in lung and subject to eQTL effects from the associated variants. Notably, a 280-kb LD block in 6p21.32 also showed suggestive association with asthma, independent of 6p21.31, in the EAs and the combined sample. This locus includes the major histocompatibility complex region and has been linked to asthma and related phenotypes in populations of European (11, 14, 16), Asian (13), and Latino descent (33). Interestingly, *PSORS1C1* at locus 6p21.33 was associated with asthma in adults. This gene has been previously associated with asthma in Latino population (33), or chronic obstructive pulmonary disease biomarkers (34).

Metaanalysis in AAs revealed a novel association of asthma with rs11788591, a common variant in the prostaglandin E synthase gene. Similar to TIE2, the biologic role of prostaglandin E_2 synthase and its product, the inflammatory mediator prostaglandin E_2, in asthma and other related phenotypes has been extensively investigated (35–39). Stumm and colleagues (36) reported a correlation between deficient prostaglandin E_2 synthesis by lung fibroblasts and airway remodeling in mouse models of asthma. Similarly, Reeves and colleagues (38) recently described reduced expression of prostaglandin E_2 synthase by airway epithelial cells of patients with asthma compared with control subjects. Although further confirmation of *PTGES* as a genetic susceptibility factor for asthma is needed, our findings are supported by the evidence of prostaglandin E_2–mediated airway remodeling in asthma.

We replicated the previously reported association at *TSLP* on 5q22.1 across EAs and AAs. *TSLP* is a TH2 response mediator (40) and has been associated with various atopic traits including asthma in Asian populations (13), as well as AAs (10), EAs (10, 14), and Latinos (10). Besides *TSLP*, two of the most extensively replicated loci across ethnically diverse populations, 17q21-12 and 5q31.1 (16, 41–43), also showed suggestive association among the EAs and AAs, respectively. The association at 17q21-12 consisted on a LD block of over 200 kb spanning from *PGAP3* to *ORMDL3/GSDMB*. Interestingly, the metaanalysis in the pediatric sample also yielded a signal in this locus, different from that found in EAs, associated with asthma in both populations and mapping to the intronic region of *GSDMB*. Together these results suggest that different haplotypes may be associated with asthma in different populations, or that the algorithm used captures different subtypes of asthma depending on the population/age. These population-specific patterns of association have been previously reported by our group (17) and others (10).

We also aimed to explore whether other SNPs reported in asthma GWAS or metaanalyses replicated in our sample. Most of the well-established asthma susceptibility loci, such as *IL1RL1* at 2q12.1, *RAD50* at 5q31.1, the HLA region, *IL33*, or the *ORMDL3/GSDMB/GSDMA* locus (17q21-12) and others, such as *CLEC16A* or locus 11q13.5, showed a trend toward significance in our study.

The replication of well-established asthma susceptibility loci along with the high PPVs obtained supports the validity of our algorithm in capturing individuals with asthma using EMR data for investigation of genetic risk factors, as also extensively demonstrated by the eMERGE network (44). Even though the use of a phenotyping algorithm may not necessarily be more accurate than a prospective study, one of the advantages of using the algorithm for the ascertainment of cases compared with the traditional cross-sectional phenotyping is the accurate and objective definition of the asthma diagnosis based on at least two *ICD9* codes and the prescription of specific asthma medications with the possibility of a lengthy follow-up of each individual. However, given the nature of EMR data, there are also limitations, such as the actual need of developing and validating algorithmic approaches to identify study subjects. Data in EMRs can be either structured (*ICD9* codes or medications) or unstructured (notes, free text, and so forth), and the latter, which may constitute up to 80% of the value in EMR data (45, 46) often requires the application of natural language processing tools to derive relevant phenotyping elements and thus dedicated expertise. Algorithm sharing is limited by the ability of participating sites to implement these practices and there will always be limits to the capacity to fully capture the richness of clinical notes.

Nevertheless, a benefit of the use of this algorithm is the extensive information collected on secondary phenotypes, such as allergy, which allowed us to investigate whether the observed associations were specific to asthma or more broadly with an allergic phenotype by adjusting for allergy status. When the analysis was adjusted by the allergic status of the cases, additional novel associations were identified. A novel locus in the intergenic region between *JMJD1C* and *REEP3* in 10q21.3 became significant in EAs. *JMJD1C* encodes a hormone-dependent transcription factor, whose expression has been reported in lymphocytes and lung and has been associated with juvenile idiopathic arthritis (47). It is noteworthy that except for the significant finding of 9p21.2 in EAs, all associations found in the nonadjusted analysis disappeared when allergy was considered. This is not surprising given that associations identified in this study were enriched with genes involved in inflammation, autoimmunity, or allergic processes, such as the inflammatory gene *PTGES*, the HLA locus, the known atopic loci 5q22.1 and 5q31.1 (11, 12, 16, 42, 48), or the established autoimmune susceptibility gene *CLEC16A*, which has been linked to asthma with hay fever in adults (15).

This study adds to the understanding of the genetic architecture of asthma in EA and AA populations with four novel asthma risk loci identified, including one associated with nonallergic asthma, and reinforces the need to study populations of diverse ethnic background to identify both shared and unique genetic predictors that contribute to the observed heterogeneity across ethnic groups. Although further studies are needed to functionally validate the culprit genes at these loci, the biologic relevance of the candidate genes of interest suggests that novel therapies may be developed to more effectively treat the underlying causes of this common disease. ■

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

Acknowledgment: The authors thank all patients and control subjects for their participation in the study. They also thank the U.S. Department of Veterans Affairs and the Electronic Medical Records and Genomics network and co-directors, Teri Manolio and Rongling Li, for their support.

References

1. Weiss KB, Sullivan SD, Lyttle CS. Trends in the cost of illness for asthma in the United States, 1985-1994. *J Allergy Clin Immunol* 2000;106: 493–499.
2. Navarro RP, Schaecher KL, Rice GK. Asthma management guidelines: updates, advances, and new options. *J Manag Care Pharm* 2007;13: S3–11; quiz S12–13.
3. Bhan N, Kawachi I, Glymour MM, Subramanian SV. Time trends in racial and ethnic disparities in asthma prevalence in the united states from the behavioral risk factor surveillance system (BRFSS) study (1999-2011). *Am J Public Health* 2015;105:1269–1275.
4. Mitchell SJ, Bilderback AL, Okelo SO. Racial disparities in asthma morbidity among pediatric patients seeking asthma specialist care. *Acad Pediatr* 2016;16:64–67.
5. Akinbami LJ, Moorman JE, Garbe PL, Sondik EJ. Status of childhood asthma in the United States, 1980-2007. *Pediatrics* 2009;123: S131–S145.
6. van Beijsterveldt CE, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J* 2007;29:516–521.
7. March ME, Sleiman PM, Hakonarson H. The genetics of asthma and allergic disorders. *Discov Med* 2011;11:35–45.
8. March ME, Sleiman PM, Hakonarson H. Genetic polymorphisms and associated susceptibility to asthma. *Int J Gen Med* 2013;6:253–265.
9. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson WO; GABRIEL Consortium. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010;363:1211–1221.
10. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, Himes BE, Levin AM, Mathias RA, Hancock DB, *et al*.; Mexico City Childhood Asthma Study (MCAAS); Children's Health Study (CHS) and HARBORS study; Genetics of Asthma in Latino Americans (GALA) Study, Study of Genes-Environment and Admixture in Latino Americans (GALA2) and Study of African Americans, Asthma, Genes & Environments (SAGE); Childhood Asthma Research and Education (CARE) Network; Childhood Asthma Management Program (CAMP); Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity (SAPPHIRE); Genetic Research on Asthma in African Diaspora (GRAAD) Study. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 2011;43:887–892.
11. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, Bleecker ER. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol* 2010; 125:328–335.e11.
12. Weidinger S, Willis-Owen SA, Kamatani Y, Baurecht H, Morar N, Liang L, Edser P, Street T, Rodriguez E, O'Regan GM, *et al*. A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. *Hum Mol Genet* 2013;22:4841–4856.
13. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, Fujita K, Miyatake A, Enomoto T, Miyagawa T, *et al*. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet* 2011;43:893–896.
14. Bønnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, Linneberg A, Curtin JA, Warrington NM, Standl M, *et al*.; Australian Asthma Genetics Consortium (AAGC); EArly Genetics and Lifecourse Epidemiology (EAGLE) Consortium. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat Genet* 2013;45:902–906.
15. Ferreira MA, Matheson MC, Tang CS, Granell R, Ang W, Hui J, Kiefer AK, Duffy DL, Baltic S, Danoy P, *et al*.; Australian Asthma Genetics Consortium Collaborators. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J Allergy Clin Immunol* 2014;133:1564–1571.
16. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, *et al*. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007;448:470–473.
17. Sleiman PM, Annaiah K, Imielinski M, Bradfield JP, Kim CE, Frackelton EC, Glessner JT, Eckert AW, Otieno FG, Santa E, *et al*. ORMDL3 variants associated with asthma susceptibility in North Americans of European ancestry. *J Allergy Clin Immunol* 2008;122: 1225–1227.
18. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, Wang K, Rafaels NM, Michel S, Bonnelykke K, *et al*. Variants of DENND1B associated with asthma in children. *N Engl J Med* 2010; 362:36–44.
19. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, den Dekker HT, Husby A, Sevelsted A, Faura-Tellez G, *et al*. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* 2014;46:51–55.
20. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, Sanderson SC, Kannry J, Zinberg R, Basford MA, *et al*. The Electronic Medical Records and Genomics (eMERGE) network: past, present, and future. *Genet Med* 2013;15:761–771.
21. Pacheco JA, Avila PC, Thompson JA, Law M, Quraishi JA, Greiman AK, Just EM, Kho A. A highly specific algorithm for identifying asthma cases and controls for genome-wide association studies. *AMIA Annu Symp Proc* 2009;2009:497–501.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
23. Delaneau O, Marchini J; 1000 Genomes Project Consortium; 1000 Genomes Project Consortium. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* 2014;5:3934.
24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
25. Verma SS, de Andrade M, Tromp G, Kuivaniemi H, Pugh E, Namjou-Khales B, Mukherjee S, Jarvik GP, Kottyan LC, Burt A, *et al*. Imputation and quality control steps for combining multiple genome-wide datasets. *Front Genet* 2014;5:370.
26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38: 904–909.
27. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;11:499–511.
28. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–D934.
29. Consortium GT; GTEx Consortium. The genotype-tissue expression (gtex) project. *Nat Genet* 2013;45:580–585.
30. Moss A. The angiopoietin:Tie 2 interaction: a potential target for future therapies in human vascular disease. *Cytokine Growth Factor Rev* 2013;24:579–592.
31. Kanazawa H, Nomura S, Asai K. Roles of angiopoietin-1 and angiopoietin-2 on airway microvascular permeability in asthmatic patients. *Chest* 2007;131:1035–1041.
32. Makinde TO, Agrawal DK. Increased expression of angiopoietins and Tie2 in the lungs of chronic asthmatic mice. *Am J Respir Cell Mol Biol* 2011;44:384–393.
33. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, Nguyen EA, Drake KA, Huntsman S, Hu D, *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.
34. Kim DK, Cho MH, Hersh CP, Lomas DA, Miller BE, Kong X, Bakke P, Gulsvik A, Agustí A, Wouters E, *et al*.; ECLIPSE, ICGN, and COPDGene Investigators. Genome-wide association analysis of blood biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;186:1238–1247.
35. Gauvreau GM, Watson RM, O'Byrne PM. Protective effects of inhaled PGE2 on allergen-induced airway responses and airway inflammation. *Am J Respir Crit Care Med* 1999;159:31–36.

36. Stumm CL, Wettlaufer SH, Jancar S, Peters-Golden M. Airway remodeling in murine asthma correlates with a defect in PGE2 synthesis by lung fibroblasts. *Am J Physiol Lung Cell Mol Physiol* 2011;301:L636–L644.
37. Liu T, Laidlaw TM, Feng C, Xing W, Shen S, Milne GL, Boyce JA. Prostaglandin E2 deficiency uncovers a dominant role for thromboxane A2 in house dust mite-induced allergic pulmonary inflammation. *Proc Natl Acad Sci USA* 2012;109: 12692–12697.
38. Reeves SR, Kolstad T, Lien TY, Elliott M, Ziegler SF, Wight TN, Debley JS. Asthmatic airway epithelial cells differentially regulate fibroblast expression of extracellular matrix components. *J Allergy Clin Immunol* 2014;134:663–670.e1.
39. Säfholm J, Manson ML, Bood J, Delin I, Orre AC, Bergman P, Al-Ameri M, Dahlén SE, Adner M. Prostaglandin E2 inhibits mast cell-dependent bronchoconstriction in human small airways through the E prostanoid subtype 2 receptor. *J Allergy Clin Immunol* 2015;136:1232–1239.e1.
40. Ziegler SF, Artis D. Sensing the outside world: TSLP regulates barrier immunity. *Nat Immunol* 2010;11:289–293.
41. Madore AM, Tremblay K, Hudson TJ, Laprise C. Replication of an association between 17q21 SNPs and asthma in a French-Canadian familial collection. *Hum Genet* 2008;123:93–95.
42. Galanter J, Choudhry S, Eng C, Nazario S, Rodríguez-Santana JR, Casal J, Torres-Palacios A, Salas J, Chapela R, Watson HG, *et al*. ORMDL3 gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 2008;177:1194–1200.
43. Leung TF, Sy HY, Ng MC, Chan IH, Wong GW, Tang NL, Waye MM, Lam CW. Asthma and atopy are associated with chromosome 17q21 markers in Chinese children. *Allergy* 2009;64:621–628.
44. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, Basford M, Chute CG, Kullo IJ, Li R, *et al*. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc* 2013; 20:e147–e154.
45. Voorham J, Denig P. Computerized extraction of information on the quality of diabetes care from free text in electronic patient records of general practitioners. *J Am Med Inform Assoc* 2007; 14:349–354.
46. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;309:1351–1352.
47. Thompson SD, Marion MC, Sudman M, Ryan M, Tsoras M, Howard TD, Barnes MG, Ramos PS, Thomson W, Hinks A, *et al*. Genome-wide association analysis of juvenile idiopathic arthritis identifies a new susceptibility locus at chromosomal region 3q13. *Arthritis Rheum* 2012;64:2781–2791.
48. Bisgaard H, Bønnelykke K, Sleiman PM, Brasholt M, Chawes B, Kreiner-Møller E, Stage M, Kim C, Tavendale R, Baty F, *et al*. Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. *Am J Respir Crit Care Med* 2009;179:179–185.