Original article

Genome-wide association study of the age of onset of childhood asthma

Erick Forno, MD, MPH,[a,c,d] Jessica Lasky-Su, ScD,[a,d] Blanca Himes, PhD,[a,d] Judie Howrylak, MD,[a] Clare Ramsey, MD, MPH,[e] John Brehm, MD, MPH,[a,b,d] Barbara Klanderman, PhD,[a] John Ziniti, BA,[a] Erik Melén, MD, PhD,[a,f,g] Goran Pershagen, MD, PhD,[f] Magnus Wickman, MD, PhD,[f,h] Fernando Martinez, MD,[i] Dave Mauger, PhD,[j] Christine Sorkness, PharmD,[k] Kelan Tantisira, MD,[a,b,d] Benjamin A. Raby, MD, MPH,[a,b,d] Scott T. Weiss, MD, MS,[a,d] and Juan C. Celedón, MD, DrPH[l] *Boston, Mass, Miami, Fla, Winnipeg, Manitoba, Canada, Stockholm, Sweden, Tucson, Ariz, Hershey and Pittsburgh, Pa, and Madison, Wis*

Background: Childhood asthma is a complex disease with known heritability and phenotypic diversity. Although an earlier onset has been associated with more severe disease, there has been no genome-wide association study of the age of onset of asthma in children.

Objective: We sought to identify genetic variants associated with earlier onset of childhood asthma.

Methods: We conducted the first genome-wide association study of the age of onset of childhood asthma among participants in the Childhood Asthma Management Program (CAMP) and used 3 independent cohorts from North America, Costa Rica, and Sweden for replication.

Results: Two single nucleotide polymorphisms (SNPs) were associated with earlier onset of asthma in the combined analysis of CAMP and the replication cohorts: rs9815663 (Fisher $P = 2.31 \times 10^{-8}$) and rs7927044 ($P = 6.54 \times 10^{-9}$). Of these 2 SNPs, rs9815663 was also significantly associated with earlier asthma onset in an analysis including only the replication cohorts. Ten SNPs in linkage disequilibrium with rs9815663 were also associated with earlier asthma onset ($2.24 \times 10^{-7} < P < 8.22 \times 10^{-6}$). Having 1 or more risk alleles of the 2 SNPs of interest (rs9815663 and rs7927044) was associated with lower lung function and higher asthma medication use during 4 years of follow-up in CAMP.

Conclusions: We have identified 2 SNPs associated with earlier onset of childhood asthma in 4 independent cohorts. (J Allergy Clin Immunol 2012;■■■:■■■-■■■.)

Key words: *Asthma, pediatrics, age of onset, asthma genetics,* C1orf100*, genome-wide association study, pediatric asthma*

Asthma is a complex disease affecting approximately 7 million children in the United States.[1] Variants in more than 40 genes have been associated with asthma.[2,3] Of these potential asthma susceptibility genes, a handful (eg, *ORMDL3*, *PDE4D*, and *DENND1B*) have been identified by using genome-wide association studies (GWASs).[4-7] Recently, a large-scale GWAS confirmed *ORMDL3* and identified several other variants, including *IL1RL1/IL18R1*, *HLA-DQ*, *IL-33*, and *SMAD3*.[8]

Childhood asthma has significant phenotypic heterogeneity. The age of onset of asthma has important phenotypic and prognostic implications,[9,10] and an earlier age of onset is associated with increased severity of asthma in children with symptoms persisting into school age and adolescence.[11,12]

From [a]the Channing Laboratory and [b]the Division of Pulmonary/Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston; [c]the Division of Pediatric Pulmonology, Department of Pediatrics, University of Miami; [d]Harvard Medical School, Boston; [e]the Department of Medicine, University of Manitoba, Winnipeg; [f]the Institute of Environmental Medicine, Karolinska Institutet, Stockholm; [g]Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm; [h]Sachs Children's Hospital, Stockholm; [i]the Department of Pediatrics, University of Arizona, Tucson; [j]the Penn State Hershey College of Medicine, Hershey; [k]the University of Wisconsin, Madison; and [l]the Division of Pediatric Pulmonary Medicine, Allergy and Immunology, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh School of Medicine.

The Genetics of Asthma in Costa Rica Study is supported by grants HL04370 and HL66289 from the US National Institutes of Health. The Childhood Asthma Management Program is supported by contracts NO1-HR-16044, 16045, 16046, 16047, 16048, 16049, 16050, 16051, and 16052 with the National Heart, Lung, and Blood Institute and General Clinical Research Center grants M01RR00051, M01RR0099718-24, M01RR02719-14, and RR00036 from the National Center for Research Resources. The genome-wide association study in BAMSE (Children, Allergy, Milieu, Stockholm, Epidemiological Survey) is supported by the European Commission as part of GABRIEL (a multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community), contract no. 018996 under the Integrated Program LSH-2004-1.2.5-1, and the Wellcome Trust under WT084703MA: "A second-generation genome-wide association study for asthma." Single-nucleotide Polymorphism Health Association Asthma Resource Project (SHARP) genotyping was performed by Affymetrix, Inc, under US Federal Government contract no. N02-HL-6-4278 from the National Heart, Lung, and Blood Institute. The BAMSE study was supported by the Swedish Research Council, Stockholm County Council, Centre for Allergy Research, Karolinska Institutet and the Swedish Heart Lung Foundation. E. M. is supported by a post doc grant from the Swedish Heart Lung Foundation, the Swedish Fulbright Commission and Riksbankens Jubileumsfond, and Erik Rönnberg's scholarship for research on early childhood diseases. SHARP was funded by the National Heart, Lung, and Blood Institute and was composed of researchers from the Asthma Clinical Research Network (ACRN), the Childhood Asthma Management Program (CAMP), and the Childhood Asthma Research and Education (CARE) network.

Disclosure of potential conflict of interest: F. Martinez is a consultant for MedImmune and has received lecture fees from Abbott. D. Mauger is a consultant for GlaxoSmithKline. K. Tantisira has received research support from the National Institutes of Health. J. C. Celedon is on the advisory board for Genentech; has received research support from the National Institutes of Health; and is author of a chapter in Up To Date. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication June 14, 2011; revised January 9, 2012; accepted for publication March 9, 2012.

Corresponding author: Erick Forno, MD, MPH, Department of Pediatrics, University of Miami, 1580 NW 10th Ave #125, Miami, FL 33136. E-mail: eforno@med.miami.edu.

0091-6749/$36.00

© 2012 American Academy of Allergy, Asthma & Immunology

doi:10.1016/j.jaci.2012.03.020

Abbreviations used

BAMSE: Children, Allergy, Milieu, Stockholm, Epidemiological Survey
CAMP: Childhood Asthma Management Program
CAMPCS/2: CAMP Continuation Study, Part 2
C1orf100: Chromosome 1 open-reading frame 100
ETS: Environmental tobacco smoke
FBAT: Family-based association testing
GACRS: Genetics of Asthma in Costa Rica Study
GWAS: Genome-wide association study
LD: Linkage disequilibrium
PACT: Pediatric Asthma Controller Trial
SNP: Single nucleotide polymorphism

In recent years, 2 studies looking at variants of *ORMDL3* found them to be strongly associated with asthma only among those whose symptoms started before 4 to 5 years of age.[13,14] However, there have been no genome-wide studies directly assessing the genetic determinants of the age of onset of asthma in children. We present the results of a GWAS of the age of onset of asthma in a cohort of North American children enrolled in the Childhood Asthma Management Program (CAMP), followed by replication studies in 3 independent cohorts of asthmatic children from Latin America, North America, and Europe.

METHODS

Population for GWAS

CAMP was a multicenter clinical trial of the effects of anti-inflammatory medications in children with mild-to-moderate asthma aged 5 to 12 years at enrollment. Study protocol and subject recruitment have been described in detail.[15,16] Of the 1024 children in CAMP, we included 573 genotyped non-Hispanic white children (413 index children in nuclear families and 160 singletons) in our analysis. Further details can be found in the Methods section in this article's Online Repository at www.jacionline.org. CAMP was approved by the Institutional Review Boards of Brigham and Women's Hospital and the other participating centers.

Replication cohorts

Genetics of Asthma in Costa Rica Study (GACRS). The protocols for subject recruitment and data collection for the GACRS have been previously described.[17-19] A total of 591 children (aged 6-14 years) were included in the replication analysis.

BAMSE (Children, Allergy, Milieu, Stockholm, Epidemiological Survey). BAMSE is a study of Swedish children recruited at birth between 1994 and 1996 and followed prospectively. BAMSE protocols have been previously described.[20,21] A total of 107 genotyped children with physician-diagnosed asthma and persistent wheezing at age 8 years were included in the analysis.

Pediatric Asthma Controller Trial (PACT). The PACT compared the effectiveness of different controller regimens for asthmatic children aged 6 to 14 years with mild-to-moderate persistent asthma and documented bronchial reversibility, methacholine sensitivity, or both.[22] A total of 233 genotyped children were included in the analysis. Data from PACT were available through the Single-nucleotide Polymorphism Health Association Asthma Resource Project (see the acknowledgments section). All studies (GACRS, BAMSE, and PACT) were approved by the respective institutional review boards, ethics committees, or both of the participating institutions.

Phenotyping

The age of onset of asthma was obtained in CAMP, GACRS, and PACT by means of parental report through a questionnaire at the beginning of each study. The age of onset was analyzed as a continuous variable; any age of onset reported to be less than 6 months was considered 0.5 years. In BAMSE questionnaires were mailed to parents when participating children were approximately 1, 2, 4, and 8 years of age. For our analysis, the age of onset was assigned to be the midpoint of each time interval: 0.5, 1.5, 3, 5, or 8 years. Details on phenotyping for other variables can be found in the Methods section in this article's Online Repository.

Gene expression

Gene expression profiling was performed on $CD4^+$ lymphocytes collected from 299 subjects participating in the CAMP Continuation Study, Part 2 (CAMPCS/2). Details can be found in the Methods section in this article's Online Repository.

Genotyping and quality control

Genome-wide single nucleotide polymorphism (SNP) genotyping was performed by Illumina, Inc (San Diego, Calif), on the HumanHap550v3 BeadChip for CAMP subjects and their parents. After stringent quality control (see the Methods section, Table E1, and Fig E1 in this article's Online Repository at www.jacionline.org for details), 512,296 SNPs remained for analysis. Genotyping details for replication cohorts can be found in the Methods section in this article's Online Repository. When SNPs selected for replication from CAMP were not available in PACT (genotyping done with a different platform), we performed imputation for the original SNPs using data from HapMap and the 1000 Genomes Project (see the Methods section in this article's Online Repository for details).

Statistical methods

The population-based GWAS of the age of onset of asthma in CAMP was performed by using survival analysis in R software with PLINK.[23] We used an additive model adjusted for age at enrollment, sex, and environmental tobacco smoke (ETS) exposure during infancy. The main eigenvectors describing the population substructure, as identified by using EIGENSTRAT,[24] were included as covariates to correct for population stratification. We also performed family-based association testing (FBAT; a generalization of the transmission disequilibrium test to test association with any phenotype, sampling structure, and pattern of missing marker information) in the 403 index children in nuclear families.[25,26] SNPs with the lowest *P* values in the population-based survival analysis that also had FBAT *P* values of less than .20 were considered for replication.

SNPs were tested in GACRS by using the same adjusted survival and FBAT analyses and in BAMSE and PACT by using adjusted survival analysis. Replication of the original finding was defined as a nominal 1-sided *P* value of less than .05 with an effect in the same direction as in the GWAS. Fisher combined *P* values for all cohorts and for the replication cohorts only were calculated by using the population-based *P* values from CAMP, BAMSE, and PACT and the FBAT *P* value from GACRS (only the *P* value from the FABT is presented to adjust for potential population stratification because genome-wide genotypic data to estimate eigenvectors were not available for this cohort). Bonferroni correction was used as a reference to control for multiple tests; the significance threshold was a *P* value of less than 9.8×10^{-8} (0.05/512,296) for the GWAS and the analysis of all cohorts and a *P* value of less than .0036 (0.05/14) for the analysis including only the replication cohorts.

RESULTS

Table I summarizes the baseline characteristics of all cohorts. Compared with children in CAMP, those in GACRS had earlier onset of asthma, higher lung function and eosinophil counts, and lower total IgE levels and were less likely to have been exposed to ETS; children in BAMSE had similar age of onset, lower frequency of exposure to ETS, and higher lung function; and

TABLE I. Main characteristics of the study participants

	CAMP	GACRS	BAMSE†	PACT
Country of origin	United States and Canada	Costa Rica	Sweden	United States
No.	573‡	591	107	233
Male sex, no. (%)	338 (59.5)	351 (59.3)	69 (64.5)	147 (63.1)
ETS, no. (%)	203 (35.9)	181 (30.7)*	27 (26.2%)*	103 (44.6%)*
Age at enrollment (y), mean (SD)	8.9 (2.1)	9.0 (1.8)	Birth*	9.9 (2.2)*
Age of onset, mean (SD)	3.1 (2.5)	2.5 (2.3)*	3.1 (2.4)	3.3 (2.7)
FEV_1 (% predicted), mean (SD)	93.9 (14.1)	99.7 (17.1)*	104.8 (11.8)*	99.1 (12.0)*
FEV_1/FVC ratio (%), mean (SD)	79.7 (8.4)	82.5 (7.4)*	84.1 (6.6)*	80.2 (7.3)
Total IgE (IU/mL),§ median (IQR)	433 (173-639)	391 (114-932)*	NA	173 (60-390)*
Eosinophils (cells/mm^3),§ median (IQR)	400 (200-639)	520 (270-790)*	NA	336 (200-632)

FVC, Forced vital capacity; *IQR*, interquartile range; *NA*, data not available.
*$P < .05$ for the comparison between each cohort and CAMP.
†In BAMSE pulmonary function data were available for 90 (84%) of 107 children; total IgE levels and eosinophil numbers were not available.
‡One hundred seventy in the budesonide treatment arm and 403 in the nedocromil/placebo arm.
§Analyzed as $\log_{10}$.

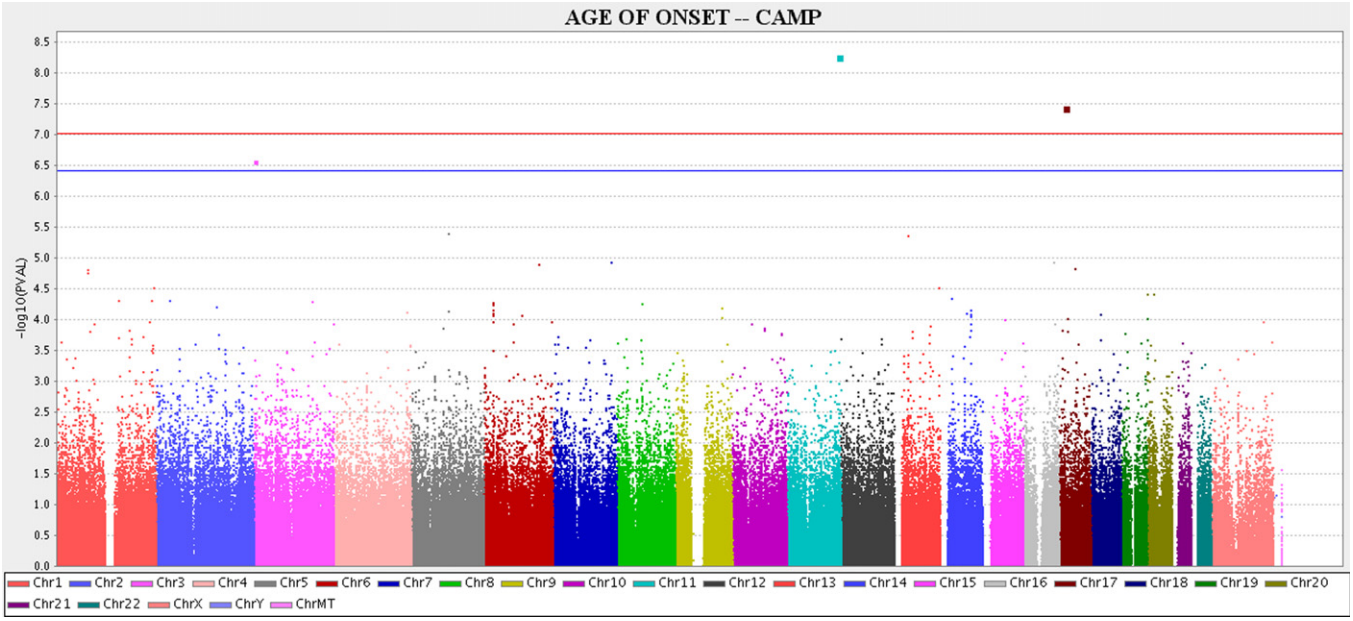

FIG 1. Plots for CAMP GWAS. Manhattan plot shows the $-\log_{10}$ *P* value by the chromosomal position for the GWAS with age of onset as the phenotype in the population-based adjusted survival analysis in CAMP. *Red line,* Threshold for a *P* value of less than .05 after Bonferroni correction ($P < 9.8 \times 10^{-8}$). *Blue line,* Threshold for a *P* value of less than .10 after Bonferroni correction ($P < 2.0 \times 10^{-7}$). *Chr,* Chromosome.

children in PACT were older and had higher lung function, higher frequency of exposure to ETS, and lower total IgE levels.

GWAS *P* values are shown in Fig 1. Two SNPs (rs7927044 [chromosome 11q24] and rs10521233 [on 17p12]) were significantly associated with the age of onset of asthma after the Bonferroni correction. These 2 SNPs, as well as the 12 SNPs with the next-lowest *P* values, were carried forward for replication in GACRS, BAMSE, and PACT.

Three SNPs were significantly associated with asthma onset in at least 1 of the replication cohorts in the same direction (earlier onset) as in the GWAS (Table II). SNP rs9815663 (3p26) showed significant association in CAMP BAMSE and PACT and a borderline significant association in GACRS ($P = .07$). SNP rs7927044 (11q24) showed significant association in CAMP and BAMSE. SNP rs4658627 (1q44, within 5 kb of the gene chromosome 1 open-reading frame 100 *[C1orf100]*) was significantly associated with an earlier asthma onset in CAMP and GACRS. After Bonferroni correction, SNPs rs9815663 and rs7927044 were significantly associated (at a genome-wide level) with earlier asthma onset in the combined analysis of all cohorts; rs9815663 was also significant in the combined analysis of the replication cohorts (excluding the GWAS). For rs4658627, the association with earlier asthma onset approached but did not achieve genome-wide statistical significance for all cohorts and for the replication-only combined analyses.

Fig 2 depicts the Kaplan-Meier curves for age of asthma onset in CAMP for these SNPs. For rs9815663, the mean ages of asthma onset in children with 0 (AA), 1 (TA), and 2 (TT) copies of the risk allele were 3.4 (SD, 2.6), 2.6 (SD, 2.2), and 1.7 (SD, 1.8) years,

TABLE II. Results for selected SNPs in CAMP, GACRS, BAMSE, and PACT

			CAMP				GACRS			BAMSE			PACT			Combined *P* values	
SNP	Chromosome	Allele	MAF	β	*P* value	FBAT	MAF	β	FBAT	MAF	β	*P* value	MAF	β	*P* value	All cohorts	Replication only
rs17100817	1	A	0.0295	−0.89	1.6×10^{-5}	0.128	0.0158	−0.08	0.41	0.0157	0.76	.15	0.3402	NA	.45	2.85×10^{-4}	.31
rs4658627	1	A	0.2730	−0.28	3.1×10^{-5}	**0.001**	0.2418	−0.13	**0.006**	0.2849	−0.16	.19	0.2160	−0.61	.12*	5.92×10^{-6}	.0068
rs9815663	3	T	0.2050	−0.42	2.8×10^{-7}	**0.029**	0.1820	−0.17	**0.075**	0.2011	−0.33	**.035**	0.1977	−0.49	**.0099**	$\mathbf{2.31 \times 10^{-8}}$	**.0018**
rs6908183	6	G	0.3155	0.26	5.9×10^{-5}	**0.006**	0.2196	0.01	**0.055**	0.3210	−0.02	.44	0.2665	NA	.20	2.00×10^{-4}	.10
rs384112	6	T	0.0111	−1.62	1.3×10^{-5}	**0.095**	0.0605	−0.06	0.38	0.0085	0.65	.29	0.1017	NA	.47	4.00×10^{-4}	.43
rs2699467	7	T	0.0581	−0.59	1.2×10^{-5}	0.155	0.1262	0.15	0.47	0.0668	0.07	.41	0.1195	−0.18	.25	3.54×10^{-4}	.42
rs960957	9	T	0.3357	0.26	6.7×10^{-5}	**0.008**	0.4733	0.1	0.45	0.3343	0.11	.24	0.3433	NA	.32	1.07×10^{-3}	.35
rs7927044	11	A	0.0104	−2.03	$\mathbf{5.5 \times 10^{-8}}$	0.108	0.0134	−0.16	0.36	0.0156	−2.15	**.0036**	0.0249	0.51	.25*	$\mathbf{6.54 \times 10^{-9}}$	.013
rs7328278	13	C	0.0183	−1.24	4.4×10^{-6}	0.157	0.0285	0.13	0.27	0.0142	2.51	**.0084**	0.0189	NA	.19	2.98×10^{-6}	.017
rs17753596	14	T	0.1239	0.38	8.0×10^{-5}	0.18	0.0680	−0.11	0.11	0.1686	−0.03	.44	0.0973	0.23	.18	4.14×10^{-4}	.15
rs9940339	16	T	0.0262	−0.84	1.2×10^{-5}	**0.059**	0.0135	−0.04	0.22	0.0341	−0.24	.29	0.0419	−0.09	.39	2.07×10^{-4}	.29
rs10521233	17	G	0.0724	−0.63	$\mathbf{3.6 \times 10^{-8}}$	**0.032**	0.0735	−0.14	0.19	0.0554	−0.01	.49	0.0582	−0.02	.48	2.69×10^{-6}	.40
rs2074289	17	G	0.0183	−0.99	1.5×10^{-5}	**0.04**	0.1156	−0.04	0.24	0.0312	−0.09	.42	NA	NA	NA	1.58×10^{-4}	.33
rs6110858	20	A	0.4342	0.25	3.9×10^{-5}	0.137	0.4620	0.07	**0.096**	0.4632	0.04	.4	0.3655	−0.28	**.048**	6.43×10^{-5}	.05

P and β values are from population-based adjusted survival analysis for age of onset, except FBAT (*P* value from family-based association test [CAMP and GACR]). Combined *P* values were calculated by using the Fisher approach for all cohorts and for replication cohorts only (GACRS, BAMSE, and PACT) by using the population-based approach *P* values for CAMP, BAMSE, and PACT and the FBAT *P* value for GACRS.

MAF, Minor allele frequency; *NA*, data not available.

*PACT genotyping was performed on an Affymetrix platform, which does not include rs4658627 or rs7927044; analysis was performed with imputed genotypic data.

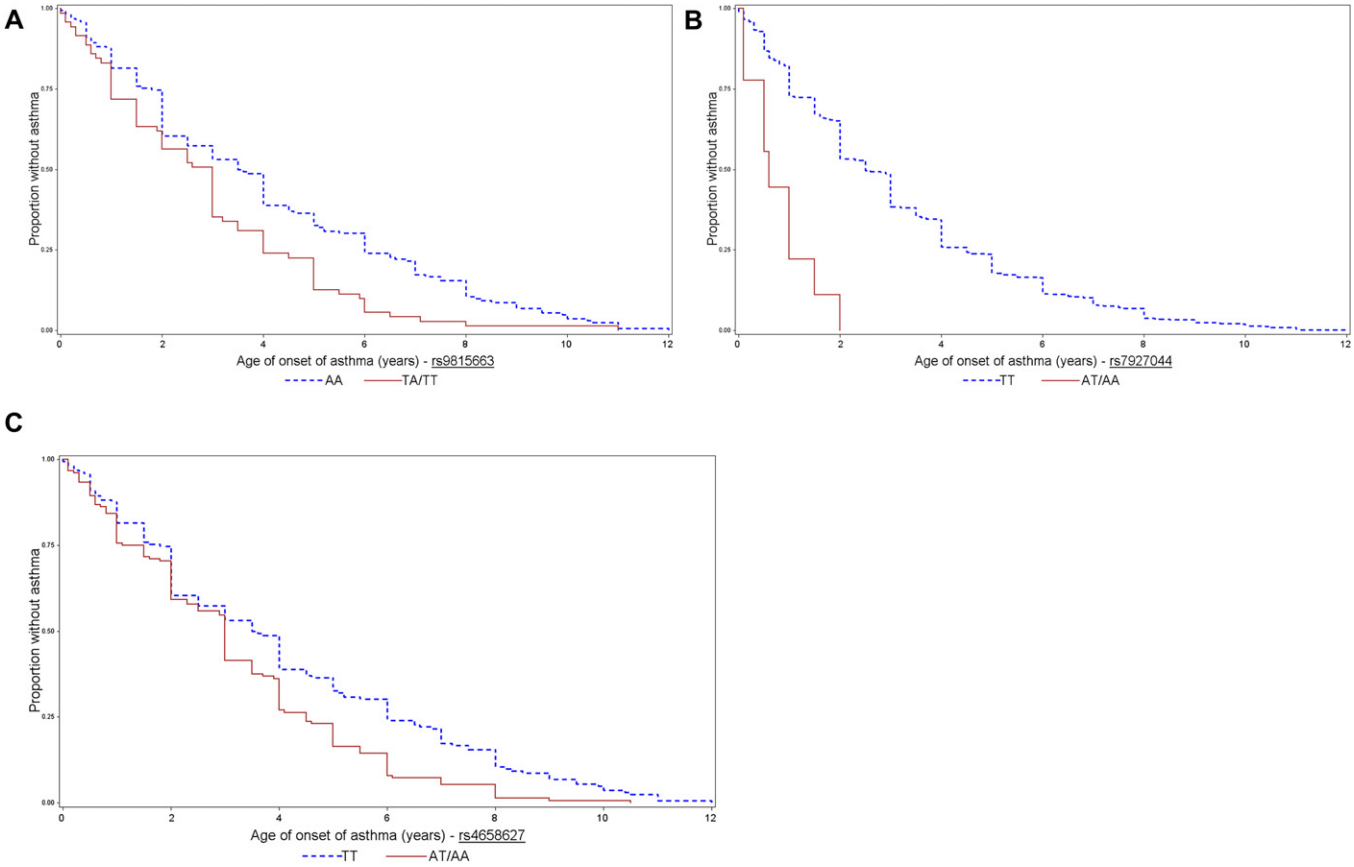

FIG 2. Kaplan-Meier curves for age of onset of asthma in CAMP by genotype of replicated SNPs. Time to onset of asthma in CAMP for rs9815663 **(A)**, rs7927044 **(B)**, and rs4658627 **(C)** is shown. *Blue dashed line*, 0 risk alleles; *red continuous line*, 1 to 2 copies of risk allele.

respectively. For rs7927044, children with 0 (TT) and 1 (TA) copies of the risk allele had mean ages of onset of 3.1 (SD, 2.5) and 0.8 (SD, 0.6) years, respectively; no subjects were homozygous AA. For rs4658627, children with 0 (TT), 1 (TA), and 2 (AA) copies of the risk allele had mean ages of onset of 3.5 (SD, 2.8), 2.8 (SD, 2.2), and 2.6 (SD, 2.1) years, respectively.

TABLE III. Age of onset of asthma and asthma phenotypes in CAMP by genotype

No. of (+) SNPs	No. (%)	Mean age of onset (y)	Mean FEV_1 (% predicted)†	Mean albuterol use score‡
0	201 (35.1)	3.4 (2.57)	94.5% (14.3)	1.86 (1.33)
≥1	372 (64.9)	2.5 (2.13)*	93.7% (13.9)*	2.11 (1.27)*

Age of onset, lung function, and medication use in genotyped children from CAMP (n = 573) for children with 0 versus any risk alleles of the 2 earlier-onset SNPs identified (rs9815663 and rs7927044). Numbers in parentheses are SDs, unless otherwise stated.
*$P < .01$ (0 vs ≥1 SNPs) from adjusted longitudinal model.
†FEV_1 at the end of CAMP (month 48 of follow-up).
‡Albuterol use score at the end of CAMP. Score: 0, none; 1, less than once a week; 2, at least once a week; 3, at least twice a week; and 4, daily.

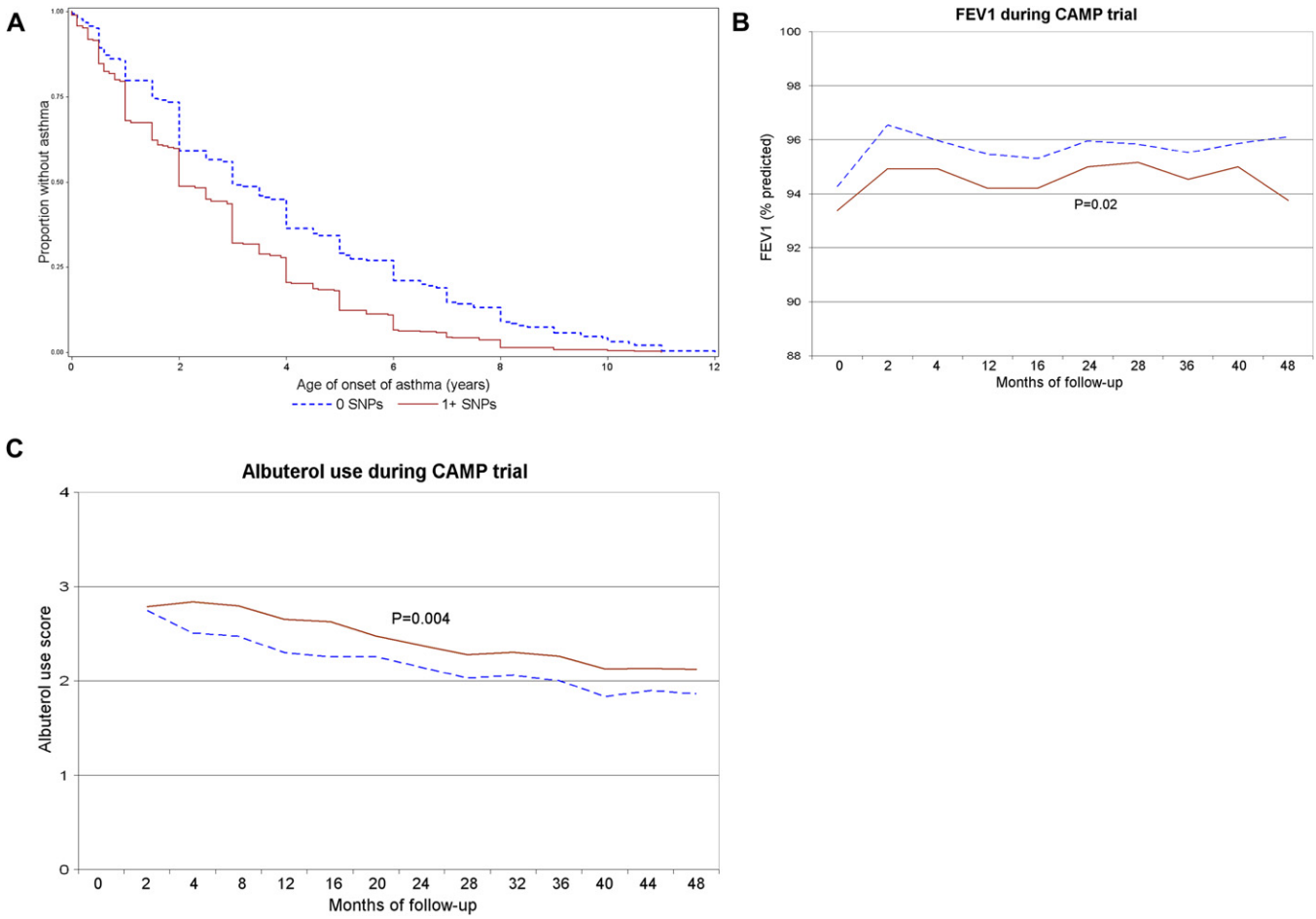

FIG 3. Age of onset and asthma phenotypes by the presence of any the 2 earlier-onset SNPs (rs9815663 and rs7927044). **A**, Time to onset of asthma in CAMP. *Blue dashed line*, Children with none of the earlier-onset SNPs; *Red continuous line*, children with 1 or more of the SNPs. **B**, FEV_1 during 4 years of follow-up in CAMP. **C**, Albuterol use score during 4 years of follow-up in CAMP.

To assess the potential combined effects of the SNPs of interest (rs9815663 and rs7927044) on asthma onset, we conducted an exploratory analysis grouping children from CAMP (n = 573) in 2 strata: those with 0 risk alleles for any of the 2 SNPs (n = 201) and those who had 1 or more risk alleles (n = 372, Table III and Fig 3). Children with 0 risk alleles had a mean age of onset of 3.4 years (SD, 2.57), whereas those who had 1 or more risk alleles had a mean onset of 2.5 years (SD, 2.13; $P < .0001$). Given the known association between earlier asthma onset and increased severity, we tested for association between the SNPs of interest and FEV_1, a measure of lung function that is inversely correlated with disease severity. FEV_1 was approximately 2.8% lower (95% CI, 0.5-5.2; $P = .018$) in children with 1 or more risk alleles than in children without risk alleles. Children with 1 or more risk alleles also had a higher mean score for albuterol use than those without risk alleles ($P = .004$); by trial's end, the difference was approximately 0.3 points (95% CI, 0.1-0.5; $P = .016$). Results were similar when we excluded children who received budesonide during the CAMP trial (data not shown).

Using imputed genotypic data for CAMP (from the 1000 Genomes Project), we identified 10 nearby SNPs in high linkage disequilibrium (LD; $r^2 > 0.75$) with SNP rs9815663. Fig 4 shows the LD plot and the $-\log_{10} P$ values for the associated SNPs. All 10 SNPs were associated with age of onset of asthma in CAMP, with P values ranging from 8.3×10^{-6} to 2.2×10^{-7}. Three additional SNPs in moderate LD ($r^2 = 0.5$-0.75) had P values

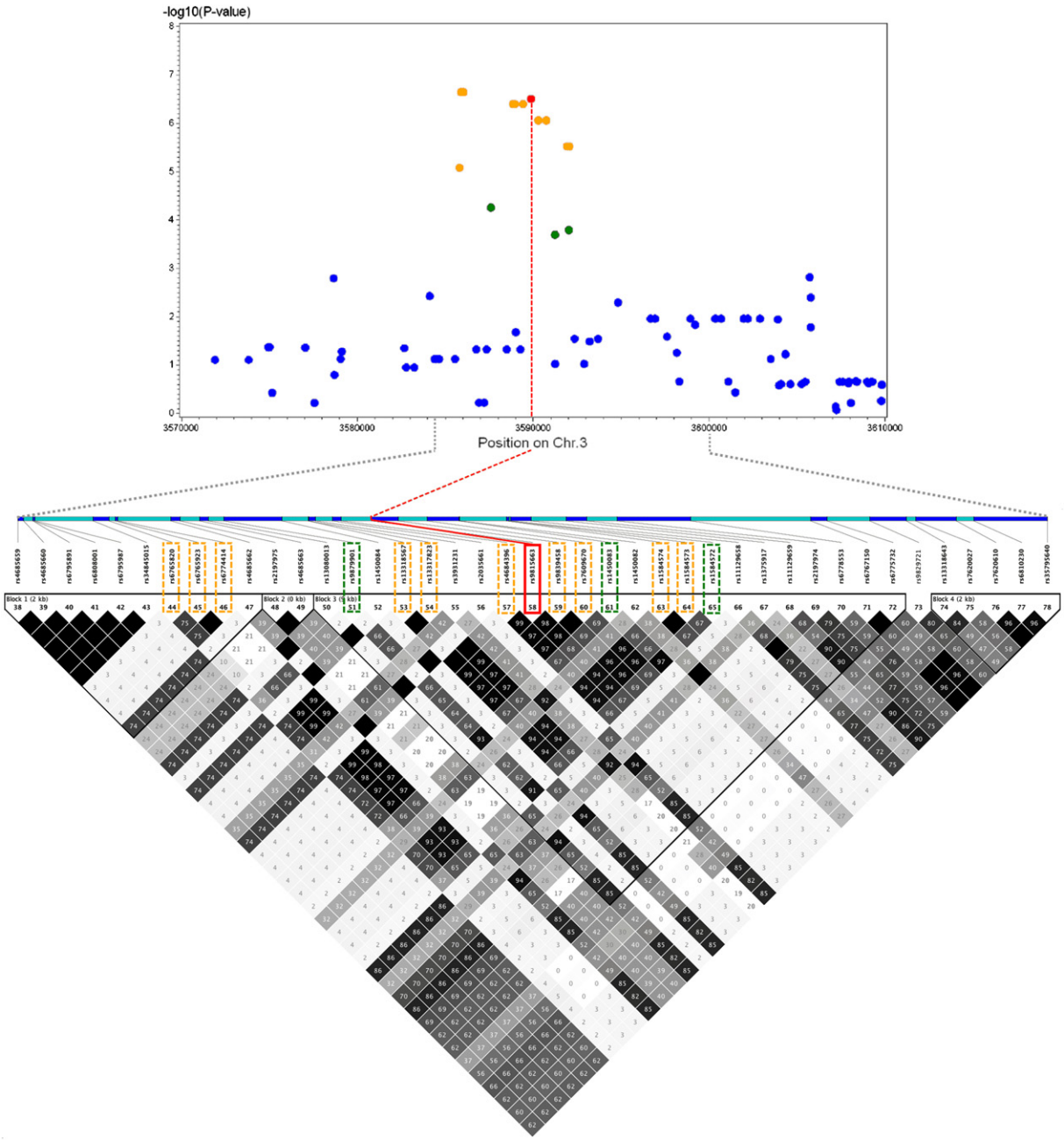

FIG 4. LD plot from CAMP for SNP rs9815663 and $-\log_{10}$ *P* values for associated SNPs. LD plot for SNP rs9815663 (±30 kb) is shown *(bottom)* with $-\log_{10}$ *P* values for association with age of onset of asthma in CAMP *(top)*. *Red*, rs9815663; *orange*, 10 SNPs in high LD ($r^2 > 0.75$); *green*, 3 SNPs in moderate LD ($r^2 > 0.50$).

between 2.0×10^{-4} and 5.5×10^{-5}. There were no SNPs in significant LD with rs4658627 or rs7927044, whether using genotyped or imputed CAMP data (see Fig E2 in this article's Online Repository at www.jacionlone.org).

We then focused on *C1orf100*, which lies approximately 5 kb from SNP rs4658627, to assess its potential function. In CAMP participants with expression data available from $CD4^+$ T lymphocytes (n = 227), adjusted longitudinal analysis showed that the expression level of *C1orf100*, although not associated with age of onset of asthma, was associated with lower percent predicted FEV_1 ($P < .0001$) and FEV_1/forced vital capacity ratio ($P < .0001$) throughout the 4 years of the trial. Exploratory analysis including the effects of *C1orf100* levels, treatment with budesonide, and an interaction term (*C1orf100**budesonide) showed a significant interaction between *C1orf100* expression and budesonide on FEV_1 (Table IV), whereas the interaction was not significant for FEV_1/forced vital capacity ratio ($P = .43$). Expression levels of *C1orf100* were also associated with an increased risk of severe exacerbations (emergency department visits, hospitalizations, or prednisone courses for asthma), increased nighttime symptom scores, and more missed schooldays for asthma in CAMP (data not shown).

TABLE IV. Longitudinal analysis for FEV_1 (percent predicted) and *C1orf100* expression levels in CAMP

	Model 1	Model 2
Intercept‡	96.1% (2.9)	96.2% (2.9)
C1orf100§	−0.36% (0.08)†	−0.44% (0.09)†
Budesonide‖	3.63% (1.6)*	3.23% (1.6)*
*C1orf100**budesonide¶	—	0.32% (0.16)*

Models are from adjusted longitudinal analysis for percent predicted FEV_1. Model 1 shows the main effects of *C1orf100* expression level and treatment with inhaled budesonide. Model 2 also includes the interaction term. Numbers represent β coefficients from mixed-effects regression models with SEs in parentheses. In summary, higher expression levels of *C1orf100* were associated with lower FEV_1; budesonide improved FEV_1 and also partially reversed the reduction in FEV_1 associated with higher *C1orf100* levels.
*$P < .05$.
†$P < .001$.
‡Children in CAMP with the lowest levels of *C1orf100* and no budesonide treatment had FEV_1 of approximately 96.2%.
§*C1orf100* gene expression levels as log-intensity values. There was a decrease of approximately 0.4% in FEV_1 for each log increase in *C1orf100* level.
‖Children receiving budesonide had an FEV_1 of approximately 3.2% to 3.6% higher than those not receiving budesonide.
¶Children with higher *C1orf100* levels had a greater response to budesonide: for each log increase in the level, treatment improved their FEV_1 by approximately 0.3% *more* than those with lower expression levels of the gene.

DISCUSSION

We report the first GWAS of the age of onset of asthma in children and report 2 SNPs (rs9815663 and rs7927044) significantly associated with an earlier age of onset of asthma in a combined analysis of 4 cohorts.

Childhood asthma is a complex disease, and our efforts to understand its causes and determinants have been hindered by its phenotypic heterogeneity. It is now recognized that asthma consists of diverse phenotypes, but the definitions of these phenotypes are also variable, depending on the methodologies used and the predictors included in the analysis.[27] Nonetheless, the age of onset of symptoms has been consistently identified as an important determinant of the severity of asthma in childhood.[9,11,12,28]

The first of the 2 SNPs we report, rs9815663, which is located in chromosome 3p26.2, is not in any known gene. However, it was consistently associated with earlier asthma onset in all the cohorts tested (same direction of association and similar effect size), and it met the criteria for significance after Bonferroni correction in the CAMP GWAS and in the combined analysis of all cohorts ($P < 9.8 \times 10^{-8}$), as well as in the analysis including only the replication cohorts ($P < .0036$). Using imputed data in CAMP, we identified several other SNPs in high and moderate LD with rs9815663 that were also consistently associated with earlier age of asthma onset. The gene closest to this SNP codes for IL-5 receptor α (*IL5RA).* IL-5 plays a role in eosinophil homeostasis and activation[29] and is a potential target for future asthma therapies[30,31]; *IL5RA* is selectively expressed in the bronchial muscle and has been shown to play an eosinophil-independent role in airway hyperresponsiveness.[32]

The second SNP, rs7927044, had the lowest *P* value in CAMP, replicated very strongly in BAMSE, and was significantly associated with earlier asthma onset in the combined analysis of all cohorts. However, this SNP did not replicate in the analysis including only the replication cohorts (no significant replication in either GACRS or PACT by using the imputed data or an available marker in strong LD with rs7927044 [rs1364780, $r^2 = 1$]), likely because of insufficient power (this SNP had low minor allele frequency [1% to 2%] in all the cohorts). Given this and the fact that there were no SNPs in significant LD with rs7927044, these results must be cautiously interpreted, awaiting follow-up in additional cohorts.

A third SNP (rs4658267) was not significantly associated with an earlier asthma onset in the combined analysis of all cohorts or in the analysis including only the replication cohorts. However, this SNP had the fourth-lowest *P* value in the analysis of all cohorts and the second-lowest *P* value in the analysis that included only the replication cohorts. Of interest, SNP rs4658267 is located within 5 kb of the gene *C1orf100* (RefSeq: NM_001012970) on chromosome 1q44 (assembly NCBI36/Hg18, dbSNP build 130). The function of *C1orf100* is still unknown, but the analysis of $CD4^+$ T-lymphocyte gene expression data showed that higher levels of expression of this gene were associated with lower lung function and other measures of asthma severity in CAMP. Although *C1orf100* cannot be confidently identified as a candidate gene for susceptibility to earlier onset of asthma, our genetic expression data for *C1orf100* lend significant biological plausibility to SNP rs4658267 and merits follow-up. $CD4^+$ T lymphocytes participate in airway inflammation, hyperreactivity, and remodeling and might be involved in steroid resistance in asthmatic patients.[33,34]

In this study children with at least 1 copy of the risk alleles for an earlier asthma onset had a lower lung function (measured based on FEV_1) and higher reported use of albuterol than those who had no risk alleles. This is consistent with prior results from CAMP, which showed that duration of asthma symptoms was associated with lower lung function, higher symptom scores, and increased use of albuterol,[12] as well as with those of epidemiologic studies reporting an association between early asthma onset and increased disease severity in adults.[35,36] Conversely, children with a later age of onset of asthma tend to have a preserved level of lung function at age 6 years when compared with children who have never wheezed.[11]

As is likely the case for other complex traits, genetic determinants might interact with environmental factors, pharmacologic factors, or both. Although the results of our exploratory analysis need to be interpreted with caution, we report an interaction between $CD4^+$ T-lymphocyte expression levels of a gene near one of our SNPs of interest *(C1orf100)* and treatment with inhaled budesonide on a measure of lung function (FEV_1) among children in CAMP. To our knowledge, this is the first report of a potential interaction between genetic determinants associated with age of onset of asthma, lung function later in childhood, and pharmacologic treatment of asthma.

Our study has several limitations. First, we have inadequate power to identify modest genetic effects on asthma onset because of sample size. Second, the age of onset of asthma was ascertained retrospectively through parental report, which might introduce recall bias. Such bias would be nondifferential with regard to genotypic data and thus likely skew our results toward the null hypothesis of no association. Third, because of budgetary constraints, only 14 SNPs among the top-ranked results in our initial cohort were carried forward for replication. However, 2 of these 14 SNPs were also significant in a combined analysis of all cohorts. Clearly a broader search with a larger set of markers is warranted. Fourth, both genotypic data imputation and high-LD surrogate markers have limitations.

In summary, we have identified 2 SNPs associated with earlier age of onset of childhood asthma in a combined analysis of 4 independent cohorts. We report that having at least 1 risk allele in

any of the 2 earlier-onset SNPs is associated with lower lung function and higher medication use during 4 years of follow-up in CAMP. These SNPs might be associated with different disease pathogenesis, prognosis, or both. Further studies are needed in this area.

We thank all the families for their invaluable participation in the GACRS and CAMP studies. We acknowledge the CAMP investigators and research team for their help in data collection. All work on data collected for GACRS was conducted at the Channing Laboratory of the Brigham and Women's Hospital under appropriate policies and human subject protections.

The BAMSE Study was supported by the Swedish Research Council, Stockholm County Council, Centre for Allergy Research, Karolinska Institutet and the Swedish Heart Lung Foundation. E.M. is supported by a post doc grant from the Swedish Heart Lung Foundation, the Swedish Fulbright Commission and Riksbankens Jubileumsfond, Erik Rönnberg's scholarship for research on early childhood diseases.

The single-nucleotide polymorphism health association-asthma resource project (SHARP) was funded by the National Heart Lung and Blood Institute, and was composed of researchers from the Asthma Clinical Research Network (ACRN), the Childhood Asthma Management Program (CAMP), and the Childhood Asthma Research and Education (CARE) network.

Clinical implications: Differences in the age of onset in childhood asthma might be associated with different asthma phenotypes, and identifying genes associated with age of onset might have implications for management and prognosis.

REFERENCES

1. Moorman JE, Rudd RA, Johnson CA, King M, Minor P, Bailey C, et al. National Surveillance for Asthma—United States, 1980-2004. Morb Mortal Wkly Rep Surveill Summ 2007;56:18-54.
2. Ober C, Hoffjan S. Asthma genetics 2006: the long and winding road to gene discovery. Genes Immun 2006;7:95-100.
3. Weiss ST, Raby BA, Rogers A. Asthma genetics and genomics 2009. Curr Opin Genet Dev 2009;19:279-82.
4. 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.
5. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. Am J Hum Genet 2009;84:581-93.
6. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, et al. Variants of DENND1B associated with asthma in children. N Engl J Med 2010;362:36-44.
7. Akhabir L, Sandford AJ. Genome-wide association studies for discovery of genes involved in asthma. Respirology 2011;16:396-406.
8. 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.
9. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. N Engl J Med 1995;332:133-8.
10. Castro-Rodriguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. Am J Respir Crit Care Med 2000;162:1403-6.
11. Taussig LM, Wright AL, Holberg CJ, Halonen M, Morgan WJ, Martinez FD. Tucson Children's Respiratory Study: 1980 to present. J Allergy Clin Immunol 2003;111:661-76.
12. Zeiger RS, Dawson C, Weiss S. Relationships between duration of asthma and asthma severity among children in the Childhood Asthma Management Program (CAMP). J Allergy Clin Immunol 1999;103:376-87.
13. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. N Engl J Med 2008;359:1985-94.
14. Halapi E, Gudbjartsson DF, Jonsdottir GM, Bjornsdottir US, Thorleifsson G, Helgadottir H, et al. A sequence variant on 17q21 is associated with age at onset and severity of asthma. Eur J Hum Genet 2010;18:902-8.
15. The Childhood Asthma Management Program (CAMP): design, rationale, and methods. Childhood Asthma Management Program Research Group. Control Clin Trials 1999;20:91-120.
16. Long-term effects of budesonide or nedocromil in children with asthma. The Childhood Asthma Management Program Research Group. N Engl J Med 2000;343:1054-63.
17. Hunninghake GM, Soto-Quiros ME, Avila L, Ly NP, Liang C, Sylvia JS, et al. Sensitization to *Ascaris lumbricoides* and severity of childhood asthma in Costa Rica. J Allergy Clin Immunol 2007;119:654-61.
18. Ly NP, Soto-Quiros ME, Avila L, Hunninghake GM, Raby BA, Laskey D, et al. Paternal asthma, mold exposure, and increased airway responsiveness among children with asthma in Costa Rica. Chest 2008;133:107-14.
19. Brehm JM, Celedon JC, Soto-Quiros ME, Avila L, Hunninghake GM, Forno E, et al. Serum vitamin D levels and markers of severity of childhood asthma in Costa Rica. Am J Respir Crit Care Med 2009;179:765-71.
20. Kull I, Melen E, Alm J, Hallberg J, Svartengren M, van Hage M, et al. Breast-feeding in relation to asthma, lung function, and sensitization in young schoolchildren. J Allergy Clin Immunol 2010;125:1013-9.
21. Melen E, Nyberg F, Lindgren CM, Berglind N, Zucchelli M, Nordling E, et al. Interactions between glutathione S-transferase P1, tumor necrosis factor, and traffic-related air pollution for development of childhood allergic disease. Environ Health Perspect 2008;116:1077-84.
22. Sorkness CA, Lemanske RF Jr, Mauger DT, Boehmer SJ, Chinchilli VM, Martinez FD, et al. Long-term comparison of 3 controller regimens for mild-moderate persistent childhood asthma: the Pediatric Asthma Controller Trial. J Allergy Clin Immunol 2007;119:64-72.
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, 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.
24. 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;38:904-9.
25. Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 2000;50:211-23.
26. Van Steen K, Lange C. PBAT: a comprehensive software package for genome-wide association analysis of complex family-based studies. Hum Genomics 2005;2:67-9.
27. Henderson J, Granell R, Sterne J. The search for new asthma phenotypes. Arch Dis Child 2009;94:333-6.
28. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. Thorax 2008;63:974-80.
29. Takatsu K, Kouro T, Nagai Y. Interleukin 5 in the link between the innate and acquired immune response. Adv Immunol 2009;101:191-236.
30. Flood-Page P, Swenson C, Faiferman I, Matthews J, Williams M, Brannick L, et al. A study to evaluate safety and efficacy of mepolizumab in patients with moderate persistent asthma. Am J Respir Crit Care Med 2007;176:1062-71.
31. Haldar P, Brightling CE, Hargadon B, Gupta S, Monteiro W, Sousa A, et al. Mepolizumab and exacerbations of refractory eosinophilic asthma. N Engl J Med 2009;360:973-84.
32. Rizzo CA, Yang R, Greenfeder S, Egan RW, Pauwels RA, Hey JA. The IL-5 receptor on human bronchus selectively primes for hyperresponsiveness. J Allergy Clin Immunol 2002;109:404-9.
33. Syed F, Bingham B, Johnson M, Markham AF, Morrison JF. The CD4+ T lymphocyte is a site of steroid resistance in asthma. QJM 1998;91:567-72.
34. Foster PS, Yang M, Herbert C, Kumar RK. CD4(+) T-lymphocytes regulate airway remodeling and hyper-reactivity in a mouse model of chronic asthma. Lab Invest 2002;82:455-62.
35. Strachan DP, Griffiths JM, Johnston ID, Anderson HR. Ventilatory function in British adults after asthma or wheezing illness at ages 0-35. Am J Respir Crit Care Med 1996;154:1629-35.
36. Sears MR, Greene JM, Willan AR, Wiecek EM, Taylor DR, Flannery EM, et al. A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. N Engl J Med 2003;349:1414-22.

METHODS

Population for the GWAS (CAMP)

For the CAMP trial, asthma was defined by symptoms greater than 2 times per week, the use of an inhaled bronchodilator at least twice weekly or the use of daily medication for asthma, and airway responsiveness to less than 12.5 mg/mL of methacholine. Children with severe asthma or other clinically significant conditions were excluded. In CAMP 1041 asthmatic children were followed up for 4 to 6 years. Of these, 968 children and 1518 of their parents contributed DNA samples. For this study, we restricted our analysis to 573 non-Hispanic white children (413 index children in nuclear families and 160 singletons).

Phenotypic assessment

Spirometry was conducted at baseline in CAMP, GACRS, and PACT according to American Thoracic Society recommendations.[E1] In CAMP subsequent measurements were obtained at 4-month intervals throughout the 48 months of the study; the completion rate was approximately 4%. Albuterol use was assessed at the same intervals in CAMP by means of parental report (range of the score was 0-4, where 0 = none, 1 = less than once a week, 2 = at least once a week, 3 = at least twice a week, and 4 = daily). Total serum IgE levels and peripheral eosinophil counts were assessed at the beginning of each study and were $\log_{10}$ transformed for analysis.

Genotyping and quality control in CAMP

Stringent quality control was conducted for the genome-wide genotypic data (see Table E1): 6257 markers were removed because of low clustering scores, and 1329 markers were removed because their flanking sequences did not map to a unique position on the hg17 reference genome sequence. Further quality control was performed with PLINK version 1.03.[E2] The average completion rate for each marker was greater than 99%. Monomorphic markers (n = 3790) and those with 5 or more Mendelian errors (n = 2445) were removed. We assessed genotype reproducibility by plating 4 subjects once on each of 14 plates. All of these replicates had greater than 99.8% concordance. The average genotyping completion rate for each subject was 99.75%. No filtering was done based on Hardy-Weinberg equilibrium because of ascertainment of the cohort through affected probands. Thus 547,645 (97.5%) of the 561,466 SNPs in the BeadChip passed quality control. Furthermore, SNPs with low minor allele frequencies of less than 1% and SNPs located in sex chromosomes were excluded, leaving 512,296 SNPs for analysis.

Genotyping in replication cohorts

SNPs selected for replication were subsequently genotyped in GACRS by using Sequenom MassArray genotyping with iPlex chemistry (12 SNPs) and TaqMan (Applied Biosystems, Foster City, Calif) assays (2 SNPs). The completion rate was 96%, and concordance was 99.9%. In BAMSE genotyping was performed by using the Illumina Human610 quad array (Illumina). All samples had a genotyping success rate of greater than 95%. Genotyping in PACT was performed with the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, Calif). The completion rate was greater than 96.5% for all included subjects, with the average completion rate being greater than 99%. When markers from Illumina were not available in Affymetrix, we used imputed data (see below) and also performed an exploratory analysis by using selected available SNPs with the highest LD possible (highest r^2 for CEU trios from the HapMap).

Genotypic data imputation in PACT

Two of the selected SNPs were not genotyped in PACT (rs4658627 and rs7927044). For rs4658726, imputation on the June 2010 release of the 1000 Genomes Project data was performed with the Markov Chain Haplotyping software (MaCH).[E3] For rs7927044, imputation was performed on data from the HapMap project (Phase 2 Release 22) with the same software. The ratio of the empirically observed dosage variance to the expected (binomial) dosage variance for imputed SNPs used was greater than 0.5 for both imputed SNPs, indicating good quality of imputation. For imputed SNPs, dosage data were used to compute association statistics.

Genotypic data imputation in CAMP

Imputation of all SNPs available in the June 2010 release of the 1000 Genomes Project data that were not genotyped or failed quality control was performed with the Markov Chain Haplotyping software (MaCH).[E3] The ratio of the empirically observed dosage variance to the expected (binomial) dosage variance for imputed SNPs used was greater than 0.5, indicating good quality of imputation.

Gene expression profiling

$CD4^+$ lymphocytes were isolated from peripheral blood samples collected from 299 subjects from 4 clinical centers (Baltimore, Boston, Denver, and St Louis) participating in CAMPCS/2; CAMPCS/2 was the second of two 4-year observational follow-up studies of CAMP participants carried out on completion of the original CAMP study. Blood samples for this analysis were obtained during a routine CAMPCS/2 clinical visit between May 2004 and July 2007. $CD4^+$ T cells were isolated from the collected mononuclear cell layer by using anti-$CD4^+$ microbeads with column separation (Miltenyi Biotec, Auburn, Calif).[E4] Total RNA was extracted with the RNeasy Mini Protocol (Qiagen, Valencia, Calif).[E5] Expression profiles were generated with the Illumina HumanRef8 version2 BeadChip arrays, and arrays were read with the Illumina BeadArray scanner and analyzed with BeadStudio (version 3.1.7) without background correction. Raw expression intensities were processed with the lumi package[E6] of Bioconductor with background adjustment with RMA convolution[E7] and $\log_2$ transformation of each array. The combined samples were quantile normalized. The normalized microarray data are available through the GeneExpression Omnibus of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/, accession no. GSE22324). The expression data for *C1orf100* (assayed on the HumanRef8 v2 array by probe ID ILMN_8320 with the sequence TAGCCACAGTTT CGCTGAATCCTCGACCGCTTAATTCACTGCCAGAGCTC) was assessed for association with FEV_1 (percent predicted), FEV_1/forced vital capacity ratio, and albuterol use by using repeated-measures analysis in SAS version 9.2 (SAS Institute, Inc, Cary, NC) with the MIXED procedure assuming a fixed-effects covariance structure and adjusted for all covariates in the main analyses.

Statistical methods for phenotypic variables

To assess whether SNPs associated with asthma onset were also associated with indicators of asthma severity, we conducted longitudinal analyses of FEV_1 and albuterol use scores in CAMP adjusted for age and height at randomization, sex, race, ETS exposure, study center, total IgE level, and peripheral eosinophil count. Analyses were done by using mixed-effects regression models. Children in the budesonide arm were excluded to avoid confounding by treatment with inhaled corticosteroids, which had an effect on lung function in CAMP. Residual maximum likelihood estimation with spatial-exponential covariance structure was used. Fixed-effects test statistics were adjusted by using the "sandwich" error estimator. Longitudinal *P* values reported are from χ^2 tests with *n-1* degrees of freedom, where *n* is the number of measurements for each outcome. Analysis was done with SAS version 9.2 software.

REFERENCES

E1. American Thoracic Society. Standardization of spirometry, 1994 update. Am J Respir Crit Care Med 1995;152:1107-36.
E2. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, 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.
E3. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet 2008;40:161-9.
E4. Jonuleit H, Schmitt E, Schuler G, Knop J, Enk AH. Induction of interleukin 10-producing, nonproliferating CD4(+) T cells with regulatory properties by repetitive stimulation with allogeneic immature human dendritic cells. J Exp Med 2000;192:1213-22.
E5. Gonzalez P, Zigler JS Jr, Epstein DL, Borras T. Identification and isolation of differentially expressed genes from very small tissue samples. Biotechniques 1999; 26:884-6, 888-92.
E6. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547-8.
E7. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249-64.

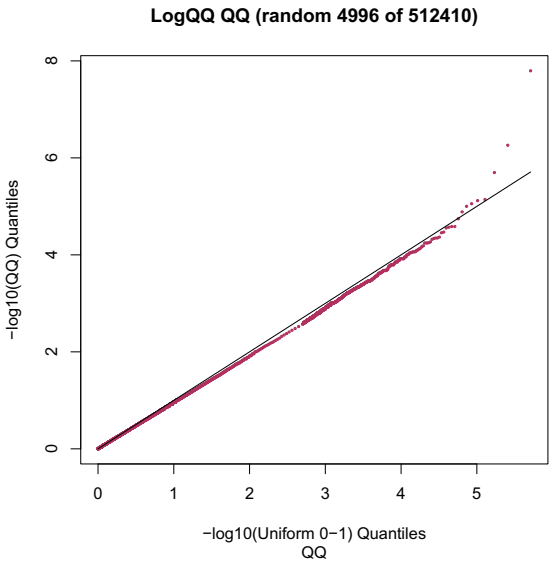

FIG E1. Probability quantile-quantile *(QQ)* plot for the CAMP GWAS.

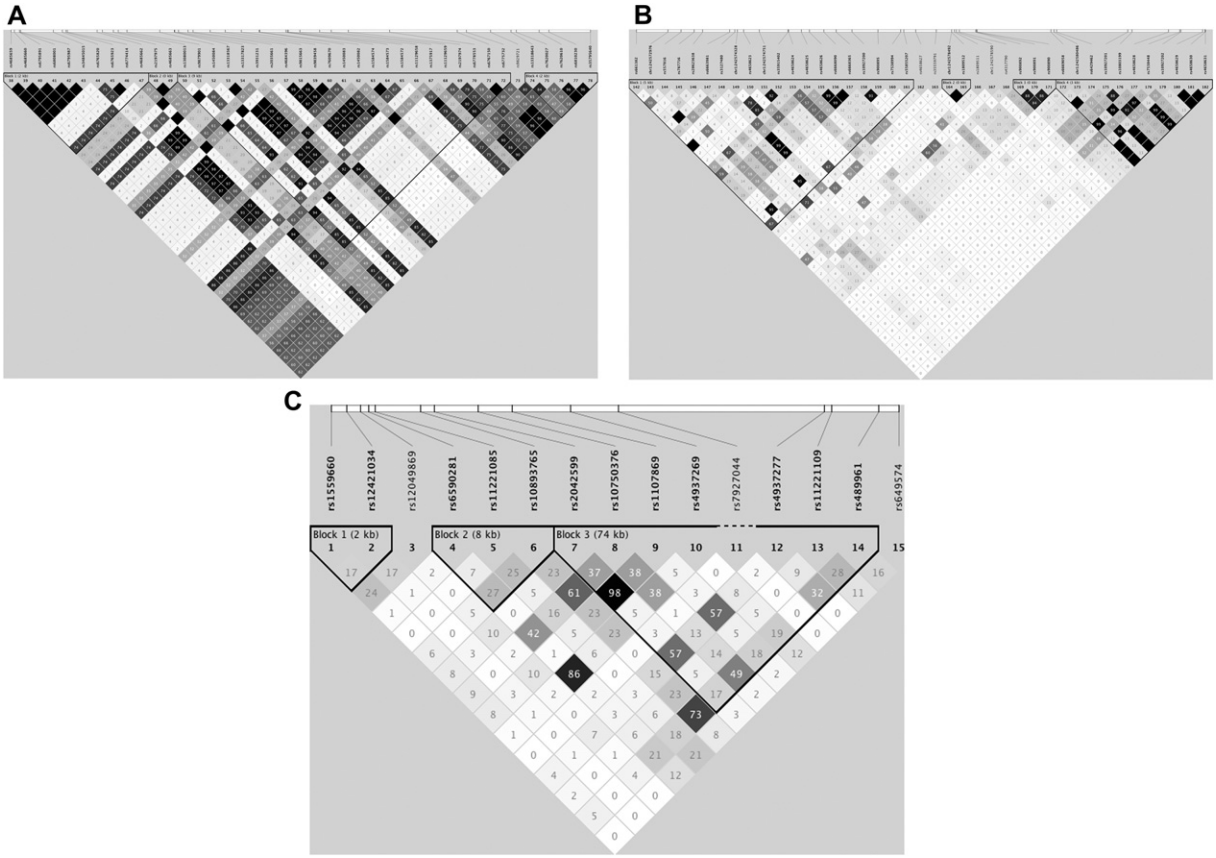

FIG E2. LD plots for SNPs rs9815663 **(A)**, rs7927044 **(B)**, and rs4658627 **(C)**. Figures show the LD plots for the 3 top SNPs from the CAMP GWAS. Plots were obtained by using imputed genotypic data in CAMP generated from the 1000 Genomes Project.

TABLE E1. Summary of the QC and cleaning procedures in the CAMP GWAS

Attribute	Count (% of 561,466 markers on array)
Low Illumina QC score	6,257 (1.1)
Flank sequences do not map to hg17	1,329 (0.2)
Monomorphic	3,790 (0.7)
Parent-offspring inconsistencies >4	2,445 (0.4)
Total no. of failed markers	13,821 (2.5)
Total no. of passed markers	547,645 (97.5)
Autosomes	534,290 (95.2)
Sex linked	13,229 (2.4)
Mitochondrial genome	126 (0.02)
Autosomal markers with MAF <1%	21,994 (3.9)
Total autosomal markers used in the analysis	512,296 (91.2)

MAF, Minor allele frequency; *QC*, quality control.