Sequencing the *IL4* locus in African Americans implicates rare noncoding variants in asthma susceptibility

Gabe Haller, BA,* Dara G. Torgerson, PhD, Carole Ober, PhD, and Emma E. Thompson, PhD *Chicago, Ill*

Background: Common genetic variations in the *IL4* gene have been associated with asthma and atopy in European and Asian populations, but not in African Americans.
Objective: Because populations of African descent have increased levels of genetic variation compared with other populations, particularly with respect to low frequency or rare variants, we hypothesized that rare variants in the *IL4* gene contribute to the development of asthma in African Americans.
Methods: To test this hypothesis, we sequenced the *IL4* locus in 72 African Americans with asthma and 70 African American controls without asthma to identify novel and rare polymorphisms in the *IL4* gene that may be contributing to asthma susceptibility.
Results: We report an excess of private noncoding single nucleotide polymorphisms (SNPs) in the subjects with asthma compared with control subjects without asthma ($P = .031$). Tajima's D is significantly more negative in subjects with asthma (–0.375) than controls (–0.073; $P = .04$), reflecting an excess of rare variants in the subjects with asthma.
Conclusion: Our findings indicate that SNPs at the *IL4* locus that are potentially exclusive to African Americans are associated with susceptibility to asthma. Only 3 of the 26 private SNPs (ie, SNPs present only in the subjects with asthma or only in the controls) are tagged by single SNPs on one of the common genotyping platforms used in genome-wide association studies. We also find that most of the private SNPs cannot be reliably imputed, highlighting the importance of sequencing to identify genetic variants contributing to common diseases in African Americans. (J Allergy Clin Immunol 2009;124:1204-9.)

Key words: *Rare variants, private alleles, asthma, IL4, IgE, African Americans*

Abbreviations used
MAF: Minor allele frequency
SNP: Single nucleotide polymorphism

IL-4 is the major T_H2 cytokine that induces the isotype switch to IgE in B lymphocytes[1,2] and is involved in host response to both parasitic infection and allergens.[3] Genetic variations in the *IL4* gene have been associated with expression levels of IL-4[4] as well as susceptibility to atopic diseases, including asthma (see review[5]), and to parasitic infections.[6,7] Even though *IL4* is among the most replicated asthma/atopy susceptibility loci,[5] only 2 studies have included individuals of African descent.[8,9] To date, all or part of the *IL4* gene has been sequenced in 78 European and 50 non-European individuals (38 African American, 6 Asian, 6 Hispanic).[8,10,11] Together, 5 single nucleotide polymorphisms (SNPs) have been identified in the coding region of the gene (3 of which are rare nonsynonymous SNPs).[12] On the other hand, an abundance of noncoding SNPs have been identified at the *IL4* locus, several of which have been implicated in disease risk. For example, 2 common SNPs have been associated with allergic disease (including asthma) in many studies.[5] These polymorphisms, a functional promoter SNP (-589C/T; rs2243250) and an SNP of unknown function in the 5′ untranslated region (-33C/T; rs2070874), are in linkage disequilibrium in European and Asian HapMap samples (Europeans, $r^2 = 1$; Asians, $r^2 = 1$; Africans, $r^2 = 0.18$). Other variations in intron 2 (3017 T/G, rs2227284, and 2 repeat polymorphisms) have also been associated with asthma and IgE in diverse populations. However, none of these common polymorphisms or any of the other tested variants in the *IL4* gene has been associated with disease risk in African Americans when *P* values are corrected for multiple testing.[8,9]

Because populations of African descent have increased levels of genetic variation compared with other populations, particularly with respect to low-frequency or rare variants,[13,14] we hypothesized that rare variants in the *IL4* gene contribute to the development of asthma in African Americans. To test this hypothesis, we sequenced nearly the entire *IL4* gene in 72 African American subjects with asthma and 70 African American control subjects. We report a significant excess of private SNPs (ie, SNPs present only in subjects with asthma or only in controls) and rare noncoding SNPs (minor allele frequency [MAF] < 5%) in subjects with asthma compared with control subjects, supporting the hypothesis that rare variants in the *IL4* gene play an important role in disease susceptibility in African Americans.

METHODS

Study samples

DNA from 142 unrelated African Americans who participate in the Chicago Asthma Genetics study were included in this study. The 72 subjects with asthma met the following criteria: (1) a physician's diagnosis of asthma (with no conflicting diagnosis); (2) the presence of at least 2 self-reported symptoms (cough, wheeze, shortness of breath); (3) current use of asthma medications; (4) either bronchial hyperresponsiveness, defined as a ≥20% decrease in FEV_1 after inhalation of ≤25 mg/mL methacholine, or

From the Department of Human Genetics, University of Chicago.
*G. Haller is currently affiliated with Washington University, Division of Biology and Biomedical Sciences, St Louis, Mo.
Supported by R01 HL72414, U01 HL049596, and M01 RR00055. D.G.T. is supported by R01 HL087665 and U10 HL064313. E.E.T. is supported by T32 HL007605.
Disclosure of potential conflict of interest: The authors have declared that they have no conflict of interest.
Received for publication July 20, 2009; revised September 9, 2009; accepted for publication September 10, 2009.
Available online November 12, 2009.
Reprint requests: Emma E. Thompson, PhD, University of Chicago, Department of Human Genetics, 920 E. 58th Street, Chicago, IL 60637. E-mail: eethomps@uchicago.edu.
0091-6749/$36.00

doi:10.1016/j.jaci.2009.09.013

reversibility to inhaled bronchodilator, defined as a ≥15% increase in baseline FEV_1 after inhalation of a bronchodilator (albuterol) or after treatment; (5) birth weight <4.4 lb; and (6) <3 pack-years of cigarette smoking. The 70 control subjects had no personal or family history of asthma among first-degree relatives and were ≥18 years of age. All subjects with asthma and control subjects reported at least 3 grandparents of African or African American descent. Total serum IgE measurements were available for all but 19 subjects with asthma and 11 controls without asthma; allergen skin prick testing was not performed in the control subjects, and those data therefore were not included in this study.

Sequencing studies

The consensus sequence AF395008 (Genbank) was used to design overlapping primer sets to cover the entire *IL4* gene, including 1743 bp upstream of the ATG start site, and 1724 bp downstream of the last exon (exon 4; see this article's Table E1 in the Online Repository at www.jacionline.org). Nucleotide positions throughout this article are given with respect to the ATG start site unless otherwise noted. PCR was performed in a total volume of 23 μL, with 5 mmol/L deoxyribonucleotide triphosphates (dNTPs), 37.5 mmol/L $MgCl_2$, 12.5 μmol/L each primer, 5 μL 5x Taq Buffer, 0.2 μL GoTaq Flexi DNA polymerase (Promega, Madison, Wis), and 20 ng genomic DNA. Unincorporated nucleotides and excess primers were removed from PCR products by using Exonuclease (New England Biolabs, Ipswich, Mass)/Shrimp Alkaline Phosphatase (USBio, Marblehead, Mass). All amplifications were sequenced in both directions by using BigDye Terminator Sequencing Kits (Applied Biosystems, Foster City, Calif). The chimpanzee consensus sequence (GeneID: 449565) was aligned to the human reference sequence with ClustalW[15] and used to determine ancestral alleles.

SNP identification

The Phred-Phrap-Consed-PolyPhred package was used to assemble the sequences and identify SNPs.[16] All sequences were visually inspected. Because of the sequence overlap, more than 1 call for each genotype was often obtained for each position in a sample.

Genotyping a variable element in intron 3

The genotypes for a variable element in intron 3 (VE6566),[10] a 70-bp copy number variant, were determined by size separation on 3% agarose gels (1-3 copies). DNA was amplified by using PCR primers (Table E1) that flanked the variable element. Genotypes at a TG dinucleotide repeat in the second intron of *IL4*[17,18] were not included in this study because we could not discern genotypes by sequencing or by an electrophoretic gel assay.

Data analysis

Polymorphisms were tested for Hardy-Weinberg equilibrium in the subjects with asthma, the controls, and the combined sample by using Haploview.[19] Single SNPs were tested for association with asthma status using a χ^2 test as implemented in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/).[20] The proportion of private SNPs in subjects with asthma compared with controls was evaluated by permutation test, in which case/control status was permuted 100,000 times in the subjects with asthma and controls (combined), holding the genotypes constant and preserving the pattern of missing data, to build an empirical distribution of the differences in the proportion of private SNPs. As a second test, a weighted-sum statistic was calculated to test for association of the locus as a whole with case status, using the method of Madsen and Browning.[21] In this method, each variant was weighted by its MAF in unaffected individuals, and then individuals were given a score consisting of the sum of their weighted alleles (with rarer alleles given greater weight). Individuals were ranked on the basis of their score, and a sum of ranks for affected individuals was calculated. Significance was determined by permuting case/control status 100,000 times to produce an empirical distribution of summed ranks, and to preserve the pattern of missing data. For all permutation tests, a threshold for statistical significance was set at P <.05 (ie, less than 5000 of the 100,000 permutations were greater than the observed sum of ranks for affected individuals).

Two measures of nucleotide diversity are commonly used to compare SNP frequencies among samples of varying size and DNA fragment length: the average number of pairwise differences in a given set of chromosomes (π),[22] and nucleotide diversity estimated from the allele frequency of the polymorphic sites (θ_w).[23] The difference between these 2 estimates (relative to their SE) is expressed as Tajima's D, which is expected to be 0 under neutrality. A positive Tajima's D reflects a larger π than θ_w and indicates an excess of intermediate-frequency variants, whereas a negative value (larger θ_w than π) indicates an excess of rare variants. Significant deviations from 0 in either direction indicate a skew in the allele frequency spectrum, which can be a sign of nonneutral evolution (ie, selection) or demographic events (ie, population history). Permutation tests were used to assess whether Tajima's D was significantly more negative in subjects with asthma compared with controls. A total of 100,000 permutations were conducted by randomly sampling 72 individuals without replacement to represent subjects with asthma from the pooled set of cases and controls. The remaining sample of 70 individuals was then taken to represent controls. For each permutation, Tajima's D was calculated for the sampled subjects with asthma and controls, and the difference between their values was calculated to assess the probability that we have observed a larger difference between the values of Tajima's D for subjects with asthma and controls than expected by chance.

Admixture estimates

European admixture in 112 African American subjects (54 cases, 58 controls) was estimated at both the genomic and local scales by using genotypes from more than 1 million SNPs on the Illumina Human 1 M array. Genomic European admixture was estimated from the first principal component in a principal component analysis including the HapMap CEU and YRI as reference populations in EIGENSTRAT.[24] Local European admixture at the *IL4* locus was estimated using 63,598 SNPs on chromosome 5 with the program Local Ancestry in adMixed Populations (LAMP)[25] by assuming 20 generations of European admixture, a constant recombination rate of 10^{-7}, and a population admixture rate of 81% obtained from genomic admixture estimates. Wilcoxon rank sum tests (WRSTs) with continuity corrections were used to compare percent European ancestry in subjects with asthma and controls, and between individuals who harbored private SNPs and those who did not. Genomic European admixture did not differ between subjects with asthma and controls (P = .64; see this article's Fig E1 in the Online Repository at www.jacionline.org) or between individuals who carried private polymorphisms and those who did not (Wilcoxon rank-sum test; P = .83). At the local scale, there was also no significant difference in European admixture between subjects with asthma and controls (P = .28; see this article's Fig E2 in the Online Repository at www.jacionline.org), or between individuals who carried private polymorphisms and those who did not (P = .49).

Estimates of linkage disequilibrium and SNP imputation

Linkage disequilibrium between SNPs identified from the sequencing of *IL4* and SNPs genotyped on the Illumina 1 M genotyping platform in 112 African American individuals (54 cases, 58 controls), was estimated by using Haploview 4.1.[19] The *IL4* SNPs from sequencing were considered to be tagged by SNPs on the Illumina 1 M if they demonstrated an r^2 value >0.5 and were within 500 kb of the *IL4* locus. To evaluate the potential for imputing the *IL4* SNPs identified by sequencing, we generated 2 datasets (a query and reference panel) including the same 112 African American individuals who were typed on the Illumina 1 M platform. The query dataset included only the 588 SNPs typed on the Illumina 1 M platform that were within 500 kb of the *IL4* transcription start and stop site. The reference panel included both the 588 SNPs from the Illumina 1 M and the *IL4* SNPs from the sequencing of the same individuals. The *IL4* SNPs in the query dataset were imputed with Mach 1.0[26] by using 50 iterations of the Markov sampler and considering 200 haplotype states when updating each individual. The accuracy in imputing

the minor allele for each SNP was defined as the number of times the minor allele was correctly imputed/the number of minor alleles observed in the data.

RESULTS

Sequence variation

We identified 92 genetic polymorphisms in the 12.4 kb of DNA encompassing the *IL4* gene in 142 African Americans (see this article's Table E2 in the Online Repository at www.jacionline.org). Of these, 86 were SNPs and 6 were insertion/deletion polymorphisms. More than half of the SNPs (N = 57) had MAFs <0.05. One SNP was in the coding region of exon 1 (44A/G; Leu33Leu; rs2070874), 17 were in the 3′ downstream region, 13 were in the 5′ upstream region, and 61 were intronic. All SNPs were in Hardy-Weinberg equilibrium in the combined sample, in subjects with asthma, and in controls without asthma after applying a Bonferroni correction for multiple testing. Only a single SNP (4400A/G; rs2243270) had a Hardy-Weinberg uncorrected *P* value <.05 in both the control ($P = .02$) and combined ($P = .008$) samples. The MAFs for each of the 92 polymorphisms in subjects with asthma and controls and in atopic cases and atopic controls are shown in Table E2. Consistent with our previous report, which included the subjects in this study,[9] the -589C/T (rs2243250), -33C/T (rs2070874), and 3017 T/G (rs2227284) SNPs had similar minor allele frequencies in subjects with asthma and controls ($P > .15$).

Private SNPs and the allele frequency spectrum in subjects with asthma and controls

We identified 26 private SNPs: 18 occurred only in subjects with asthma and 8 occurred only in controls, all with MAF ≤0.02. Only 2 of these SNPs were present in dbSNP (4614/rs2243272 and 5342/rs2243277), each of which had been identified by sequencing a single heterozygous individual in an African American sample (N = 46 chromosomes).[11] Two of the private SNPs (-996 and -593) in our study occurred together in 2 subjects with asthma (CA4748 and CA9994) and presumably reside on the same haplotype in these individuals; 2 additional private SNPs (612, 8982) were each present in 2 subjects with asthma, 1 private SNP (4614) was present in 3 subjects with asthma, and 1 private SNP (-1621) was present in 2 controls (Fig 1).

The proportion of private SNPs (among all SNPs detected) was higher in subjects with asthma than controls ($P = .031$; permutation test; Table I). The sex ratios among individuals carrying private SNPs (16 females, 10 males) and among individuals not carrying private SNPs (73 females, 30 males) were not different ($P = .358$). Private SNPs were not associated with European admixture in either subjects with asthma (Fisher exact test; $P = 1.0$) or controls ($P = .710$; see this article's Table E3 in the Online Repository at www.jacionline.org), indicating that an excess of private SNPs in the subjects with asthma is not a result of differences in the number of admixed individuals or the proportion of European admixture between the groups. In addition, the weighted sum method[21] revealed an association between variation at the *IL4* locus and asthma status ($P = .017$; permutation test). Last, we investigated whether individuals with private SNPs had higher IgE levels than individuals without private SNPs. Seventeen of 23 (77%) subjects with private SNPs had IgE levels above the median value (72 IU/mL), whereas only 39 of 80 (49%) subjects without private SNPs had IgE levels above the median ($P = .004$), indicating that increased IgE levels are associated with rare, private SNPs and suggesting that at least some of these SNPs may be regulatory in function.

We next calculated Tajima's D to identify differences in the allele frequency spectrum between subjects with asthma and controls. Tajima's D[22] considers the difference between 2 measures of nucleotide diversity, θ_W and π, and is positive when there is an excess of intermediate frequency variants, negative when there is an excess of rare variants (at either high-frequency or low-frequency derived SNPs), and 0 under neutral expectations (where $\theta_W = \pi$). The value for Tajima's D in the control sample is –0.073, which is in the 78th percentile of values from an empirical distribution based on 327 autosomal loci in African Americans (Seattle SNPs). However, Tajima's D is more negative in subjects with asthma (–0.375) than in controls (–0.073; $P = .041$; permutation test; Table I), reflecting the increased number of low-frequency variants in the subjects with asthma.

Imputing rare and private SNPs

None of the private SNPs and only a single rare variant identified in this study were present on the Illumina 1 M genotyping array. Furthermore, only 3 of the private SNPs in *IL4* were tagged by SNPs on that platform ($r^2 > 0.5$, 2 private to controls and 1 private to a subject with asthma; see this article's Table E4 in the Online Repository at www.jacionline.org). As expected, a greater proportion of private SNPs failed to be correctly imputed at even a single minor allele compared with SNPs that were shared between subjects with asthma and controls (57% vs 1.7%; Fisher exact test; $P < 10^{-6}$; see this article's Table E5 and Fig E3 in the Online Repository at www.jacionline.org). This is similarly true for rare SNPs (57% vs 3.1%; $P = 9 \times 10^{-6}$).

DISCUSSION

The prevalence of asthma in the United States has risen in recent decades,[27] and the disease burden disproportionately affects African Americans.[28] In addition, the clinical presentation of asthma with respect to associations with and severity of atopy and bronchial hyperresponsiveness differs between African Americans and European Americans.[28] Yet relatively few studies have investigated the genetics of asthma in African American populations, in whom associations are less often replicated,[29,30] as is the case at the *IL4* locus.[8,9] One explanation for this finding could be differences in haplotype structure and patterns of linkage disequilibrium between individuals of European or Asian ancestry and those of African ancestry (see review[14]), as suggested in a recent genome-wide association study for asthma genes.[29] An alternative explanation is that rare alleles present in African-derived populations, but not in European or Asian populations, contribute to disease susceptibility, possibly even masking the effects of other variants. Because of their unique demographic history, African-derived populations harbor more nucleotide diversity in general, and more rare alleles in particular, than populations of European or Asian descent.[13,14,31] It is possible, therefore, that rare alleles account for more of the genetic risk for common diseases in African-derived populations. Our sequencing study revealed an excess of private and rare SNPs in African Americans with asthma compared with African American control subjects without asthma, suggesting that this may indeed be the case at some loci.

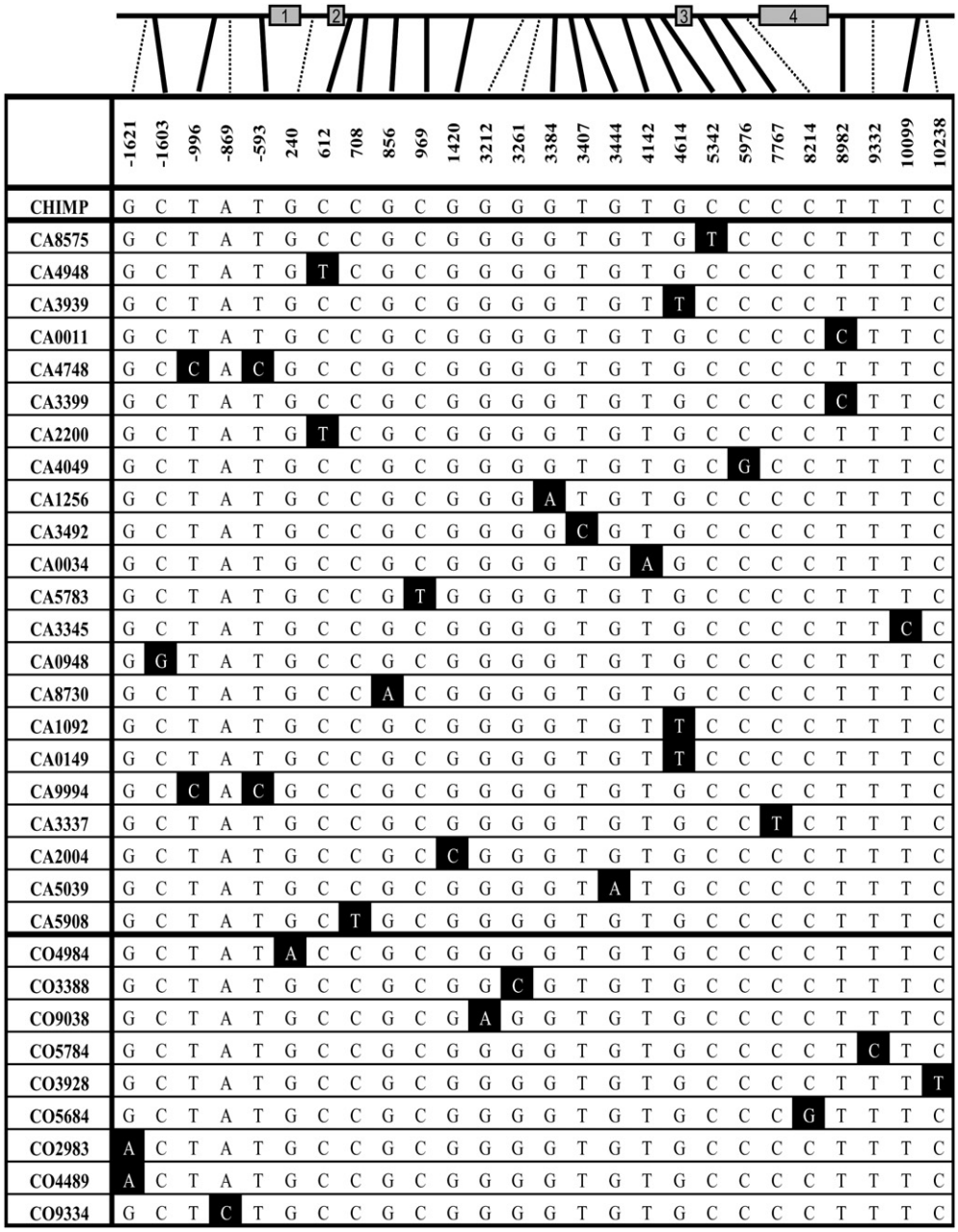

FIG 1. Private SNP haplotypes in subjects with asthma and controls. A schematic of the *IL4* gene is shown at the *top* of the figure, with relative SNP positions (with respect to ATG start site) marked. SNPs present only in asthma cases *(CA)* or only in asthma controls (*CO*; ie, private SNPs) are shown. SNPs present in asthma cases are connected to the gene figure by *solid lines*, and SNPs present in controls are connected by *hatched lines*. Individual haplotypes appear in rows, with nucleotides against a dark background representing the derived (private) allele.

The idea that rare SNPs contribute to common diseases is not new,[32-34] and a number of recent studies have reported excesses of rare, highly penetrant coding SNPs in individuals with cardiovascular disease phenotypes, tuberculosis, colorectal cancer, pancreatitis, folate response, type 1 diabetes, trichotillomania, and obsessive compulsive disorder.[35-48] In contrast, we present here for the first time an association between a common disease and rare, noncoding variants. Because relatively few sequencing studies have been conducted in disease susceptibility genes, and because most of those studies focus on coding regions, it is not known whether this is an unusual finding or reflects a shared genetic architecture among asthma or other common disease susceptibility genes in African Americans. Regardless, the finding of an excess of private and rare variants in African American subjects with asthma compared with controls is consistent with the hypothesis that multiple rare variants at disease susceptibility loci contribute to complex disease risk and, if generalizable to other loci, could explain the relatively poor replicability of disease susceptibility loci in African Americans.[29,30,49-51] In that case, characterizing the genetic architecture of common

TABLE I. Descriptive statistics for SNPs at the *IL4* locus

	Proportion of private SNPs	Permutation *P* value	θ_W*	π†	Tajima D	Permutation *P* value
Combined sample (n = 142)	—	—	12.53×10^{-4}	11.09×10^{-4}	–0.356	—
Controls without asthma (n = 70)	0.076	.031	11.26×10^{-4}	11.00×10^{-4}	–0.073	.041
Subjects with asthma (n = 72)	0.211	—	12.71×10^{-4}	11.20×10^{-4}	–0.375	—

*The Watterson[23] estimate of the population mutation rate ($\theta = 4N_e\mu$, where N_e is the effective population size and μ is the per generation mutation rate).
†Average number of pairwise differences between samples.[22]

diseases in populations of African descent may be particularly challenging.

Currently, the relative contribution of rare genetic variants to asthma risk is unknown. Future resequencing studies of *IL4* and additional genes in larger samples of well characterized subjects with asthma and controls are required to address this important question. Our study was limited with respect to sample size and the corresponding low power to detect clinical differences between individuals with and without private or rare SNPs. As a result, we were not able to address the possibility that rare variants in the *IL4* gene are markers for clinical subtypes of asthma, such as atopic asthma (based on allergen skin prick tests), steroid-resistant asthma, or childhood versus adult onset. However, resquencing studies in large samples are becoming more feasible with the availability of high-throughput Next Generation sequencing technologies,[52] which will enable more comprehensive surveys of rare variations in asthma and other common diseases. A final caveat is that sequencing artifacts caused by the misincorporation of bases during PCR can bias studies of rare variations. We have attempted to minimize this possibility by sequencing each amplicon in both directions and by designing overlapping PCR fragments that would allow the identification of the same variant in independent amplicons. However, even if some were a result of PCR artifacts, we would not expect this to occur more commonly in the subjects with asthma than in the controls. Therefore, we are confident that the excess of rare or private SNPs in subjects with asthma compared with controls in our study is not a result of PCR or sequencing artifacts.

Last, we note that only 2 of the 26 private variants present in our sample (Fig 1) have been reported previously (rs2243272 and rs2243277), and none of the private variants are present on the Affymetrix 6.0 or Illumina 1 M genotyping platforms. Furthermore, only 3 of the 26 private variants are tagged by SNPs on the Illumina 1 M genotyping platform (Table E4). We also find that a greater proportion of private SNPs failed to be correctly imputed at even a single minor allele compared with SNPs that are shared between subjects with asthma and controls, which reflects the challenges of imputing rare alleles in general (Fig E3). Given that our reference and query panels for imputation are perfectly matched (ie, identical samples used in both), our findings represent the best case scenario for a reference panel similar in size to the HapMap CEU and YRI samples. Thus, if an excess of rare or private variation reflects a common mechanism of disease susceptibility in African Americans, genome-wide association studies with 1 million or more SNP genotypes may still miss the majority of susceptibility alleles for common diseases in populations of African descent.

Clinical implications: Rare, noncoding SNPs in *IL4* may play an important role in asthma susceptibility in African Americans. This study highlights the importance of resequencing for discovering risk variants in African populations.

REFERENCES

1. Finkelman FD, Holmes J, Katona IM, Urban JF Jr, Beckmann MP, Park LS, et al. Lymphokine control of in vivo immunoglobulin isotype selection. Annu Rev Immunol 1990;8:303-33.
2. Murphy KM, Reiner SL. The lineage decisions of helper T cells. Nat Rev Immunol 2002;2:933-44.
3. Lynch NR, Hagel IA, Palenque ME, Di Prisco MC, Escudero JE, Corao LA, et al. Relationship between helminthic infection and IgE response in atopic and nonatopic children in a tropical environment. J Allergy Clin Immunol 1998;101:217-21.
4. Rosenwasser LJ, Klemm DJ, Dresback JK, Inamura H, Mascali JJ, Klinnert M, et al. Promoter polymorphisms in the chromosome-5 gene-cluster in asthma and atopy. Clin Exp Allergy 1995;25:74-8.
5. Ober C, Hoffjan S. Asthma genetics 2006: the long and winding road to gene discovery. Genes Immun 2006;7:95-100.
6. Blackwell JM, Mohamed HS, Ibrahim ME. Genetics and visceral leishmaniasis in the Sudan: seeking a link. Trends Parasitol 2004;20:268-74.
7. Gatlin MR, Black CL, Mwinzi PN, Secor WE, Karanja DM, Colley DG. Association of the gene polymorphisms IFN-gamma +874, IL-13-1055 and IL-4 -590 with patterns of reinfection with Schistosoma mansoni. PLoS Negl Trop Dis 2009;3:e375.
8. Basehore MJ, Howard TD, Lange LA, Moore WC, Hawkins GA, Marshik PL, et al. A comprehensive evaluation of IL4 variants in ethnically diverse populations: association of total serum IgE levels and asthma in white subjects. J Allergy Clin Immunol 2004;114:80-7.
9. Donfack J, Schneider DH, Tan Z, Kurz T, Dubchak I, Frazer KA, et al. Variation in conserved non-coding sequences on chromosome 5q and susceptibility to asthma and atopy. Respir Res 2005;6:145.
10. Kabesch M, Tzotcheva I, Carr D, Hofler C, Weiland SK, Fritzsch C, et al. A complete screening of the IL4 gene: novel polymorphisms and their association with asthma and IgE in childhood. J Allergy Clin Immunol 2003;112:893-8.
11. SeattleSNPs. NHLBI HL66682 Program for Genomic Applications, UW-FHCRC, Seattle, WA (URL: http://pga.gs.washington.edu) Accessed June 2009.
12. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308-11.
13. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, et al. Proportionally more deleterious genetic variation in European than in African populations. Nature 2008;451:994-7.
14. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 2002;3:611-21.
15. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673-80.
16. Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 1997;25:2745-51.
17. Chouchane L, Sfar I, Bousaffara R, El Kamel A, Sfar MT, Ismail A. A repeat polymorphism in interleukin-4 gene is highly associated with specific clinical phenotypes of asthma. Int Arch Allergy Immunol 1999;120:50-5.
18. Nagarkatti R, Kumar R, Sharma SK, Ghosh B. Association of IL4 gene polymorphisms with asthma in North Indians. Int Arch Allergy Immunol 2004;134:206-12.
19. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

20. 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.
21. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 2009;5:e1000384.
22. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989;123:585-95.
23. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol 1975;7:256-76.
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. Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. Am J Hum Genet 2008;82:290-303.
26. Li Y, Willer CJ, Sanna S and Abecasis GR. Genotype Imputation. Annu Rev Genomics Hum Genet 2009;10:387-406.
27. Centers for Disease Control and Prevention. National Center for Health Statistics. Available at: http://www.cdc.gov/asthma/asthmadata.htm. Accessed June 2, 2009.
28. Barnes KC, Grant AV, Hansel NN, Gao P, Dunston GM. African Americans with asthma: genetic insights. Proc Am Thorac Soc 2007;4:58-68.
29. 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.
30. Xu X, Fang Z, Wang B, Chen C, Guang W, Jin Y, et al. A genomewide search for quantitative-trait loci underlying asthma. Am J Hum Genet 2001;69:1271-7.
31. Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, et al. Genomic regions exhibiting positive selection identified from dense genotype data. Genome Res 2005;15:1553-65.
32. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 2008;40:695-701.
33. Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or "common gene, common disease"? Methods Mol Biol 2007;376:71-84.
34. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69:124-37.
35. Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, et al. Medical sequencing at the extremes of human body mass. Am J Hum Genet 2007;80:779-91.
36. Azzopardi D, Dallosso AR, Eliason K, Hendrickson BC, Jones N, Rawstorne E, et al. Multiple rare nonsynonymous variants in the adenomatous polyposis coli gene predispose to colorectal adenomas. Cancer Res 2008;68:358-63.
37. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 2005;37:161-5.
38. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264-72.
39. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 2004; 305:869-72.
40. Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci U S A 2006;103:1810-5.
41. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 2008;40:592-9.
42. Kotowski IK, Pertsemlidis A, Luke A, Cooper RS, Vega GL, Cohen JC, et al. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. Am J Hum Genet 2006;78:410-22.
43. Ma X, Liu Y, Gowen BB, Graviss EA, Clark AG, Musser JM. Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoS ONE 2007;2:e1318.
44. Marini NJ, Gin J, Ziegle J, Keho KH, Ginzinger D, Gilbert DA, et al. The prevalence of folate-remedial MTHFR enzyme variants in humans. Proc Natl Acad Sci U S A 2008;105:8055-60.
45. Masson E, Chen JM, Scotet V, Le Marechal C, Ferec C. Association of rare chymotrypsinogen C (CTRC) gene variations in patients with idiopathic chronic pancreatitis. Hum Genet 2008;123:83-91.
46. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009;324:387-9.
47. Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, et al. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet 2007;39:513-6.
48. Zuchner S, Wendland JR, Ashley-Koch AE, Collins AL, Tran-Viet KN, Quinn K, et al. Multiple rare SAPAP3 missense variants in trichotillomania and OCD. Mol Psychiatry 2009;14:6-9.
49. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, et al. Many sequence variants affecting diversity of adult human height. Nat Genet 2008;40:609-15.
50. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007;39:631-7.
51. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 2009;41:18-24.
52. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet 2008;24:133-41.

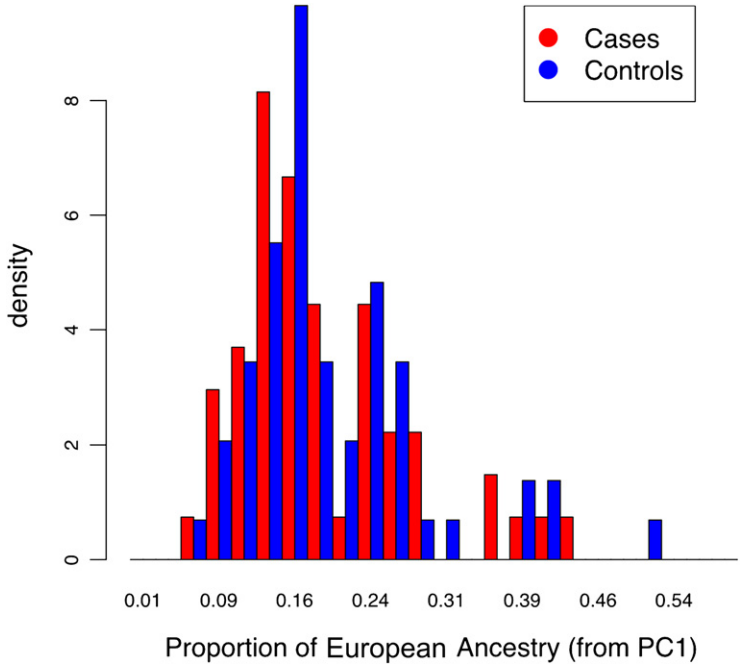

FIG E1. Distribution of genomic European admixture in African American subjects with asthma (n = 54) and controls (n = 58) as estimated from the first principal component (PC) in EIGENSTRAT by using SNPs from the Illumina 1 M genotyping array.

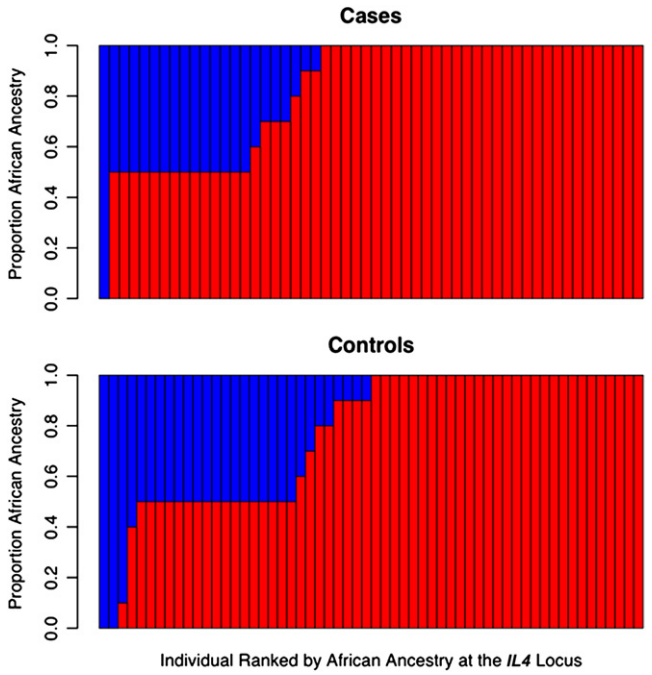

FIG E2. Local ancestry in African American subjects with asthma (n = 54) and controls (n = 58) at the *IL4* locus. The proportion of African ancestry *(red)* and European ancestry *(blue)* was estimated using SNPs from the Illumina 1 M genotyping array with the program LAMP.

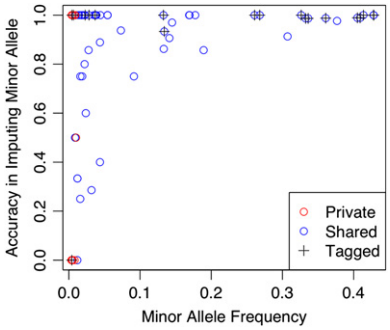

FIG E3. Accuracy in imputing the minor allele for SNPs identified from sequencing studies of 112 African American individuals (54 subjects with asthma, 58 controls) as it relates to the MAF in the sample. *Accuracy,* Number of correctly inferred minor alleles/number of observed minor alleles; *Private,* SNPs that are unique to either subjects with asthma or controls; *Shared,* SNPs present in both subjects with asthma and controls; *Tagged, IL4* SNPs that are tagged by a SNP on the Illumina 1 M genotyping array with an $r^2 > 0.5$.

TABLE E1. *IL4* primer information

Name	Primer sequence	Reference position*	T_m (°C)
IL4_Promoter_F	ggcaaaaccttagcaacaca	10749	58.4
IL4_Promoter_R	gccaatcagcacctctcttc	12302	
IL4_Exon1_F	tgtggcctctcccttctatg	12109	58.4
IL4_Exon1_R	attctcagccgtgtgtttcc	13722	
IL4_Intron1.1_F	cccccaccctctatctgta	13601	58.4
IL4_Intron1.1_R	ggctggattttggaaagatg	15261	
IL4_Intron2.2_Fb	ccctgaacttcatcctcctg	15080	61.8
IL4_Intron2.2_Rb	tggcagatttttgcttctgt	16663	
IL4_Intron2.2_Fc	caaaggtggtatgcagaggaa	15211	61.8
IL4_Intron2.2_Rc	aaacgcattgcacagtggta	16707	
IL4_Intron2.3_F	attctggtgcctcagtctgg	16530	61.8
IL4_Intron2.3_R	catttggaggatgggagaga	18130	
IL4_Exon3_F	cagccttccttcagtggaat	18033	61.8
IL4_Exon3_R	ttgcaagttctgacctctcc	19570	
IL4_Intron3.1_F	aatgaagcaagatggcctgt	19126	55.5
IL4_Intron3.1_R	tttgcctattttgggtgcat	20779	
IL4_Exon4_F	tcaagttccaccctctgagc	20647	61.8
IL4_Exon4_R	atggaaagccgaaagtctcc	22177	
IL4_Exon4_Fa	gctgtgacacacctctccag	20676	61.8
IL4_Exon4_Ra	tttcaccctccttagtcca	22150	
IL4_Intron4.1_F	gggttccctctcgagttagg	21295	55.5
IL4_Intron4.1_R	tacagcagcgcagtcatagc	22786	
IL4_VE_F	acgagtatggcagggaacac	18934	61.8
IL4_VE_R	accgatctgtcagcaaatcc	19321	
IL4_Promoter_Finternal	gggaaggttctgggagaaaa	11147	55.5

*Reference position is with respect to the start of the reference sequence.

TABLE E2. *IL4* polymorphisms

SNP ID	Position with respect to ATG	Ancestral allele	Derived allele	Derived allele frequency in the total sample	Derived allele frequency in subjects with asthma	Derived allele frequency in controls without asthma
—	-1665	G	A	0.054	0.070	0.037
—	-1621	G	A	0.007	0.000	0.014
rs10080170	-1608	T	G	0.879	0.910	0.848
—	**-1603**	**C**	**G**	**0.004**	**0.007**	**0.000**
—	-1587	G	A	0.035	0.035	0.036
—	-1543	G	A	0.014	0.007	0.022
rs10065221	-1535	A	G	0.085	0.056	0.116
rs10058157	-1532	T	C	0.167	0.132	0.203
rs2243242	-1508	G	—	0.377	0.355	0.400
rs2243247	-1137	G	A	0.176	0.176	0.177
rs2243248	-1099	T	G	0.158	0.155	0.162
rs2243249	-1046	T	C	0.026	0.021	0.030
—	**-996**	**T**	**C**	**0.007**	**0.014**	**0.000**
—	-869	A	C	0.004	0.000	0.008
—	**-593**	**T**	**C**	**0.007**	**0.014**	**0.000**
rs2243250	-589	C	T	0.691	0.704	0.677
rs2070874	-33	T	C	0.580	0.542	0.619
rs2243251	44	A	G	0.183	0.190	0.176
—	240	G	A	0.004	0.000	0.007
—	**612**	**C**	**T**	**0.007**	**0.014**	**0.000**
—	**708**	**C**	**T**	**0.004**	**0.007**	**0.000**
rs2243252	845	T	C	0.022	0.021	0.023
—	**856**	**G**	**A**	**0.004**	**0.007**	**0.000**
—	**969**	**C**	**T**	**0.004**	**0.007**	**0.000**
rs734244	983	C	T	0.419	0.458	0.377
rs2243253	1018	C	T	0.139	0.134	0.144
—	1113	C	T	0.011	0.007	0.015
—	1304	C	T	0.014	0.014	0.014
—	**1420**	**G**	**C**	**0.004**	**0.007**	**0.000**
rs11479198	1424	C	—	0.426	0.441	0.410
—	2187	T	C	0.015	0.023	0.008
—	2307	A	G	0.019	0.022	0.015
rs2243258	2401	C	T	0.030	0.036	0.024
rs2243259	2402	C	T	0.019	0.021	0.016
—	2992	C	A	0.014	0.022	0.007
rs2227284	3017	T	G	0.131	0.116	0.147
rs2243260	3041	A	T	0.015	0.014	0.015
rs2243261	3097	G	T	0.164	0.167	0.162
—	3150	—	TC	0.026	0.036	0.015
—	3212	G	A	0.004	0.000	0.007
—	3261	G	C	0.004	0.000	0.008
—	**3384**	**G**	**A**	**0.004**	**0.007**	**0.000**
—	**3407**	**T**	**C**	**0.004**	**0.007**	**0.000**
—	**3444**	**G**	**A**	**0.004**	**0.007**	**0.000**
rs2227282	3470	C	G	0.118	0.114	0.121
rs2243263	3590	G	C	0.174	0.167	0.182
—	3675	C	T	0.008	0.007	0.008
—	3726	C	A	0.037	0.050	0.023
rs2243264	3938	A	G	0.018	0.021	0.015
rs2243265	4000	C	A	0.019	0.014	0.023
rs2243266	4080	G	A	0.343	0.371	0.313
—	**4142**	**T**	**A**	**0.004**	**0.007**	**0.000**
rs2243267	4177	G	C	0.353	0.371	0.333
rs2243268	4254	A	C	0.275	0.297	0.254
rs9282745	4291	T	A	0.047	0.051	0.043
rs9282746	4294	G	A	0.051	0.043	0.058
rs2243269	4303	AA	—	0.274	0.284	0.265
rs2243270	4400	A	G	0.673	0.657	0.689
—	4470	T	C	0.011	0.015	0.007
rs2243271	4590	G	A	0.037	0.043	0.030

(Continued)

TABLE E2. (*Continued*)

SNP ID	Position with respect to ATG	Ancestral allele	Derived allele	Derived allele frequency in the total sample	Derived allele frequency in subjects with asthma	Derived allele frequency in controls without asthma
rs2243272	**4614**	**G**	**T**	**0.011**	**0.023**	**0.000**
rs2243273	5051	C	T	0.051	0.043	0.058
rs2243274	5123	A	G	0.404	0.368	0.441
rs2243275	5212	T	C	0.022	0.029	0.014
rs2243276	5262	T	C	0.007	0.007	0.007
rs2243277	**5342**	**C**	**T**	**0.004**	**0.007**	**0.000**
—	**5976**	**C**	**G**	**0.004**	**0.007**	**0.000**
rs2243278	6058	ACTAAAG ACACGCA GGCCGAGTC	—	0.018	0.021	0.014
—	6566	1 copy	2 or 3 copies	0.650	0.632	0.669
rs2243281	6686	T	C	0.067	0.045	0.088
rs2243285	7284	G	T	0.164	0.144	0.184
—	**7767**	**C**	**T**	**0.004**	**0.007**	**0.000**
—	7999	G	C	0.011	0.014	0.007
rs2243286	8043	C	T	0.029	0.029	0.029
rs2243287	8082	G	C	0.014	0.014	0.014
—	8214	C	G	0.004	0.000	0.007
rs2243288	8235	A	G	0.638	0.679	0.596
rs2243289	8423	A	G	0.270	0.301	0.239
rs2243290	8460	C	A	0.339	0.375	0.304
—	8968	C	T	0.036	0.050	0.022
—	**8982**	**T**	**C**	**0.007**	**0.014**	**0.000**
—	9040	C	T	0.037	0.050	0.023
—	9226	T	C	0.011	0.014	0.008
—	9274	C	G	0.404	0.357	0.455
—	9332	T	C	0.004	0.000	0.008
—	9512	T	C	0.041	0.043	0.039
—	9874	C	T	0.011	0.014	0.007
—	10082	A	G	0.622	0.643	0.601
—	10098	G	A	0.331	0.300	0.362
—	**10099**	**T**	**C**	**0.004**	**0.007**	**0.000**
—	10173	A	C	0.029	0.043	0.015
—	10238	C	T	0.004	0.000	0.007

Private SNPs found only in subjects with asthma appear in *boldface*. RS numbers are given for polymorphisms previously reported in dbSNP.
*The variable element is triallelic in the general population. Only 2 and 3 copies of the variable element were found in our population. The 2-copy allele is taken as the derived allele in this table.

TABLE E3. Number of individuals with private SNPs in admixed and nonadmixed African Americans at the *IL4* locus

	Admixed	Nonadmixed
Controls ($P = .71$)		
With private SNP	3	5
Without private SNP	26	24
Subjects with asthma ($P = 1.0$)		
With private SNP	6	9
Without private SNP	16	21

Admixed individuals are those inferred as having <2 African chromosomes at the *IL4* locus.

TABLE E4. Number of *IL4* SNPs identified from sequencing studies of 112 African Americans that are tagged by a SNP on the Illumina 1 M genotyping array ($r^2 > 0.5$; within 500 kb upstream and downstream of the transcription start and stop site)

	Private		
	Subjects with asthma (rare)	Controls (rare)	Shared (rare)
Tagged, $r^2 > 0.5$	1 (1)	2 (2)	16 (3)
Untagged	13 (13)	5 (5)	43 (31)

The number in *parentheses* represents the subset of rare SNPs in the cell counts (MAF $< 5\%$).

TABLE E5. Number of *IL4* SNPs identified from sequencing studies of 112 African Americans that failed to be correctly imputed at even a single minor allele by using genotypes from the Illumina 1 M genotyping platform (accuracy = 0)

	Private: subjects with asthma (rare)	Private: controls (rare)	Shared (rare)
Accuracy = 0	7 (7)	5 (5)	1 (1)
Accuracy > 0	7 (7)	2 (2)	58 (32)

Accuracy was defined as the number of times the minor allele was correctly imputed/ the number of minor alleles observed in the data. The number in *parentheses* represents the subset of rare SNPs in the sample (MAF < 5%).