Published in final edited form as:

J Allergy Clin Immunol. 2009 December ; 124(6): 1204–9.e9. doi:10.1016/j.jaci.2009.09.013.

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
Department of Human Genetics, The University of Chicago

Abstract

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 to 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 non-asthmatic controls to identify novel and rare polymorphisms in the *IL4* gene that may be contributing to asthma susceptibility.

Results—We report an excess of private non-coding SNPs in the subjects with asthma compared to non-asthmatic control subjects ($P=0.031$). Tajima's D is significantly more negative in cases (-0.375) compared to controls (-0.073) ($P=0.04$), reflecting an excess of rare variants in the cases.

Conclusions—Our findings indicate that SNPs at the *IL4* locus that are potentially exclusive to African Americans are associated with susceptibility to asthma. Only three of the 26 private SNPs (i.e., SNPs present only in the cases 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.

Keywords

Rare variants; Private alleles; Asthma; *IL4*; IgE; African Americans

Introduction

Interleukin 4 (IL-4) is the major T helper 2 (Th2) 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, (reviewed in⁵) and to

© 2009 American Academy of Allergy, Asthma and Immunology. Published by Mosby, Inc. All rights reserved.

Corresponding author: Emma Thompson, PhD, The University of Chicago, Department of Human Genetics, 920 E. 58th Street, Chicago, IL 60637, Ph. (773) 702-5898, FAX (773) 834-0505, eethomps@uchicago.edu.

Current Address for G.H.: Washington University, Division of Biology and Biomedical Sciences, 660 S. Euclid Ave., Campus Box 8226, St. Louis, MO 63110

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

parasitic infections.^{6, 7} Even though *IL4* is among the most replicated asthma/atopy susceptibility loci,⁵ only two 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, five single nucleotide polymorphisms (SNPs) have been identified in the coding region of the gene (three of which are rare non-synonymous SNPs).¹² On the other hand, an abundance of non-coding SNPs have been identified at the *IL4* locus, several of which have been implicated in disease risk. For example, two common SNPs have been associated with allergic disease (including asthma) in many studies.⁵ These polymorphisms, a functional promoter SNP (–589C/T; rs2243250) and a SNP of unknown function in the 5' untranslated region (UTR) (–33C/T; rs2070874), are in LD (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 (3017T/G, rs2227284 and two 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 to 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 (i.e., SNPs present only in cases or only in controls) and rare noncoding SNPs (minor allele frequency (MAF) <5%) in subjects with asthma compared to 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 (C.A.G.) study were included in this study. The 72 asthma cases met the following criteria: (1) a physician's diagnosis of asthma (with no conflicting diagnosis), (2) the presence of at least two self-reported symptoms (cough, wheeze, shortness of breath), (3) current use of asthma medications, (4) either bronchial hyperresponsiveness, defined as a 20% decrease in FEV₁ after inhalation of 25 mg/mL of methacholine, or reversibility to inhaled bronchodilator, defined as a 15% increase in baseline FEV₁ after inhalation of a bronchodilator (albuterol) or after treatment; (5) birth weight <4.4 pounds, 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 case and control subjects reported at least three grandparents of African or African American descent. Total serum IgE measurements were available for all but 19 asthma cases and 11 non-asthmatic controls; allergen skin prick testing was not performed in the controls 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) (Supplementary Table 1). Nucleotide positions throughout this paper are given with respect to the ATG start site unless otherwise noted. PCR was performed in a total volume of 23 μ L, with 5 mM dNTPs, 37.5 mM MgCl₂, 12.5 μ M of each primer, 5 μ L 5 $\times$ Taq Buffer, 0.2 μ L GoTaq Flexi DNA polymerase (Promega,

Madison WI), and 20 ng of genomic DNA. Unincorporated nucleotides and excess primers were removed from PCR products using Exonuclease (New England Biolabs, Ipswich MA)/Shrimp Alkaline Phosphatase (USBio, Marblehead MA). All amplifications were sequenced in both directions using BigDye Terminator Sequencing Kits (Applied Biosystems, Foster City CA). The chimpanzee consensus sequence (GeneID: 449565) was aligned to the human reference sequence using ClustalW¹⁵ 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 one call for each genotype was often obtained for each position in a sample.

Genotyping a Variable Element (VE) in Intron 3

The genotypes for a VE in intron 3 (VE6566),¹⁰ a 70bp copy-number variant, were determined by size separation on 3% agarose gels (1–3 copies). DNA was amplified using PCR primers (Table S1) 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 cases, controls, and the combined sample using Haploview.¹⁹ Single SNPs were tested for association with asthma status using a chi-squared test, as implemented in PLINK.²⁰ The proportion of private SNPs in asthma cases compared to controls was evaluated by permutation test, in which case/control status was permuted 100,000 times in the cases and controls (combined), holding the genotypes constant and preserving the pattern of missing data, in order 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 minor allele frequency in unaffected individuals, and then individuals were given a score consisting of the sum of their weighted alleles (rarer alleles being given greater weight). Individuals were ranked based on 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 $P < 0.05$ (i.e. less than 5,000 of the 100,000 permutations being 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 two estimates (relative to their standard error) is expressed as Tajima's D, which is expected to be zero under neutrality. A positive Tajima's D reflects a larger π than θ_w and indicates an excess of intermediate-frequency variants, while a negative value (larger θ_w than π) indicates an excess of rare variants. Significant deviations from zero in either direction indicate a skew in the allele frequency spectrum, which can be a sign of non-neutral evolution (i.e. selection) or demographic events (i.e. population history). Permutation tests were used to assess whether Tajima's D was significantly more negative in cases compared to the controls. A total of 100,000 permutations were conducted by randomly sampling 72 individuals without replacement to represent asthma cases from the pooled set of cases and controls. The remaining sample of 70 individuals was then taken to represent asthma controls. For each

permutation, Tajima's D was calculated for the sampled cases and controls and a difference between their values was calculated in order to assess the probability that we have observed a larger difference between the values of Tajima's D for cases 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 using genotypes from over 1 million SNPs on the Illumina Human 1M 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 using 63,598 SNPs on chromosome 5 with the program LAMP²⁵ 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 cases and controls, and between individuals harboring private SNPs and those who do not. Genomic European admixture did not differ between cases and controls ($P = 0.64$, Figure S1) or between individuals carrying private polymorphisms when compared to those who do not (WRST $P = 0.83$). At the local scale, there was also no significant difference in European admixture between cases and controls ($P = 0.28$, Figure S2), or between individuals carrying private polymorphisms when compared to those who do not ($P = 0.49$).

Estimates of LD and SNP Imputation

Linkage disequilibrium between SNPs identified from the sequencing of *IL4* and SNPs genotyped on the Illumina 1M genotyping platform in 112 African American individuals (54 cases, 58 controls), were estimated using Haploview 4.1.¹⁹ The *IL4* SNPs from sequencing were considered to be tagged by SNPs on the Illumina 1M if they demonstrated an r^2 value > 0.5 and were within 500 Kb of the *IL4* locus. In order to evaluate the potential for imputing the *IL4* SNPs identified by sequencing, we generated two datasets (a query and reference panel) including the same 112 African American individuals that were typed on the Illumina 1M platform. The query dataset included only the 588 SNPs typed on the Illumina 1M 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 1M and the *IL4* SNPs from the sequencing of the same individuals. The *IL4* SNPs in the query dataset were imputed using Mach 1.0,²⁶ 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 (Table S2). Of these, 86 were SNPs and 6 were insertion/deletion polymorphisms. More than half of the SNPs ($N=57$) had minor allele frequencies (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 asthma cases, and in non-asthmatic controls after applying a Bonferroni correction for multiple testing. Only a single SNP (4400A/G; rs2243270) had a Hardy-Weinberg uncorrected P -value < 0.05 in both the control ($P=0.02$) and combined ($P=0.008$) samples. The MAFs for each of the 92 polymorphisms in asthma cases and controls and in atopic cases and atopic controls are

shown in Table S2. Consistent with our previous report, which included the subjects in this study,⁹ the -589C/T (rs2243250), -33C/T (rs2070874), and 3017T/G (rs2227284) SNPs had similar minor allele frequencies in cases and controls ($P > 0.15$).

Private SNPs and the Allele Frequency Spectrum in Cases and Controls

We identified 26 private SNPs: 18 occurred only in asthma cases and eight occurred only in controls, all with MAF = 0.02. Only two 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 two cases (CA4748 and CA9994), and presumably reside on the same haplotype in these individuals; two additional private SNPs (612, 8982) were each present in two cases, one private SNP (4614) was present in three cases, and one private SNP (-1621) was present in two controls (Figure 1).

The proportion of private SNPs (among all SNPs detected) was higher in asthma cases compared to controls ($P = 0.031$; permutation test) (Figure 1). The sex ratios among subjects carrying private SNPs (16 females, 10 males) and among subjects not carrying private SNPs (73 females, 30 males) were not different ($P = 0.358$). Private SNPs were not associated with European admixture in either asthma cases (Fisher's Exact test, $P = 1.0$) or controls ($P = 0.710$) (Table S3), indicating that an excess of private SNPs in the cases is not due to differences in the number of admixed individuals or the proportion of European admixture between the groups. Additionally, the weighted sum method²¹ also revealed an association between variation at the *IL4* locus and asthma status ($P = 0.017$; permutation test). Lastly, 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 ($P = 0.004$), 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 cases and controls. Tajima's D²² considers the difference between two measures of nucleotide diversity, θ_W and π , which is positive when there is an excess of intermediate frequency variants, negative when there is an excess of rare variants (at either high or low frequency derived SNPs), and zero under neutral expectations (where $\theta_W = \pi$). The value for Tajima's D in the asthma 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 asthma cases (-0.375) than in controls (-0.073) ($P = 0.041$; permutation test), reflecting the increased number of low frequency variants in the asthma cases.

Imputing Rare and Private SNPs

None of the private SNPs and only a single rare variant identified in this study are present on the Illumina 1M genotyping array. Furthermore, only three of the private SNPs in *IL4* are tagged by SNPs on that platform ($r^2 > 0.5$, two private to controls and one private to cases) (Table S4). As expected, a greater proportion of private SNPs failed to be correctly imputed at even a single minor allele as compared to SNPs that are shared between cases and controls (57% vs. 1.7%; Fisher's Exact test, $p < 10^{-6}$; Table S5; Figure S3). 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 decades²⁷ 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 (reviewed in¹⁴), 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. Due to 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 to non-asthmatic African American control subjects, suggesting that this may indeed be the case at some loci.

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 asthma cases compared to 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 diseases in populations of African descent may be particularly challenging.

At the present time, 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 cases 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 vs. adult onset. However, resequencing 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 variation in asthma and other common diseases. A final caveat that we note is that sequencing artifacts due to 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

due to PCR artifacts, we would not expect this to occur more commonly in the cases than in the controls. Therefore, we are confident that the excess of rare or private SNPs in cases compared to controls in our study is not due to PCR or sequencing artifacts.

Lastly, we note that only two of the 26 private variants present in our sample (Figure 1) have been reported previously (rs2243272 and rs2243277), and none of the private variants are present on the Affymetrix 6.0 or Illumina 1M genotyping platforms. Furthermore, only three of the 26 private variants are tagged by SNPs on the Illumina 1M genotyping platform (Table S4). We also find that a greater proportion of private SNPs failed to be correctly imputed at even a single minor allele as compared to SNPs that are shared between cases and controls, which reflects the challenges of imputing rare alleles in general (Figure S3). Given that our reference and query panels for imputation are perfectly matched (i.e. 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 one million or more SNP genotypes may still miss the majority of susceptibility alleles for common diseases in populations of African descent.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by R01 HL72414, U01 HL049596 and M01 RR00055. D.G.T. is supported by R01 HL087665 and U10 HL064313 and E.E.T. is supported by T32 HL007605.

Abbreviations Used

C.A.G	Chicago Asthma Genetics
IgE	immunoglobulin E
IL4	interleukin 4
LD	linkage disequilibrium
MAF	minor allele frequency
SNP	single nucleotide polymorphism
Th2	T helper 2
UTR	untranslated region
VE	variable element
WRST	Wilcoxon rank sum test

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; 8:303–333. [PubMed: 1693082]
2. Murphy KM, Reiner SL. The lineage decisions of helper T cells. *Nat Rev Immunol.* 2002; 2:933–944. [PubMed: 12461566]
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; 101:217–221. [PubMed: 9500755]

4. Rosenwasser LJ, Klemm DJ, Dresback JK, Inamura H, Mascali JJ, Klennert M, et al. Promoter Polymorphisms in the Chromosome-5 Gene-Cluster in Asthma and Atopy. *Clinical and Experimental Allergy*. 1995; 25:74–78. [PubMed: 8590350]
5. Ober C, Hoffjan S. Asthma genetics 2006: the long and winding road to gene discovery. *Genes Immun*. 2006; 7:95–100. [PubMed: 16395390]
6. Blackwell JM, Mohamed HS, Ibrahim ME. Genetics and visceral leishmaniasis in the Sudan: seeking a link. *Trends Parasitol*. 2004; 20:268–274. [PubMed: 15147677]
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. [PubMed: 19190772]
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; 114:80–87. [PubMed: 15241348]
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. [PubMed: 16336695]
10. Kabesch M, Tzotcheva I, Carr D, Hofler C, Weiland SK, Fritzsche 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; 112:893–898. [PubMed: 14610476]
11. SeattleSNPs. Seattle, WA: UW-FHCRC; NHLBI HL66682 Program for Genomic Applications. (URL: <http://pga.gs.washington.edu>).
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; 29:308–311. [PubMed: 11125122]
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; 451:994–997. [PubMed: 18288194]
14. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet*. 2002; 3:611–621. [PubMed: 12154384]
15. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. 2002; Chapter 2(Unit 2):3. [PubMed: 18792934]
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; 25:2745–2751. [PubMed: 9207020]
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; 120:50–55. [PubMed: 10529588]
18. Nagarkatti R, Kumar R, Sharma SK, Ghosh B. Association of IL4 gene polymorphisms with asthma in North Indians. *Int Arch Allergy Immunol*. 2004; 134:206–212. [PubMed: 15178889]
19. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21:263–265. [PubMed: 15297300]
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; 81:559–575. [PubMed: 17701901]
21. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet*. 2009; 5:e1000384. [PubMed: 19214210]
22. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989; 123:585–595. [PubMed: 2513255]
23. Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 1975; 7:256–276. [PubMed: 1145509]
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; 38:904–909. [PubMed: 16862161]

25. Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. *Am J Hum Genet.* 2008; 82:290–303. [PubMed: 18252211]
26. Li Y, Ding J, Abecasis GR. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet.* 2006; 79:S2290.
27. Centers for Disease Control and Prevention. [Accessed June 2, 2009] National Center for Health Statistics. Health Data Interactive. www.cdc.gov/nchs/hdi.htm.
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. [PubMed: 17202293]
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; 84:581–593. [PubMed: 19426955]
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; 69:1271–1277. [PubMed: 11673820]
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; 15:1553–1565. [PubMed: 16251465]
32. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet.* 2008; 40:695–701. [PubMed: 18509313]
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. [PubMed: 17984539]
34. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet.* 2001; 69:124–137. [PubMed: 11404818]
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; 80:779–791. [PubMed: 17357083]
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; 68:358–363. [PubMed: 18199528]
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; 37:161–165. [PubMed: 15654334]
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; 354:1264–1272. [PubMed: 16554528]
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; 305:869–872. [PubMed: 15297675]
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; 103:1810–1815. [PubMed: 16449388]
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; 40:592–599. [PubMed: 18391953]
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; 78:410–422. [PubMed: 16465619]
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. [PubMed: 18091991]
44. Marini NJ, Gin J, Ziegler J, Kehoe 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; 105:8055–8060. [PubMed: 18523009]

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. [PubMed: 18172691]
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; 324:387–389. [PubMed: 19264985]
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; 39:513–516. [PubMed: 17322881]
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. [PubMed: 19096451]
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; 40:609–615. [PubMed: 18391951]
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; 39:631–637. [PubMed: 17401366]
51. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadóttir 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. [PubMed: 19079260]
52. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 2008; 24:133–141. [PubMed: 18262675]

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.

Figure 1.

Private SNP Haplotypes in Asthma Cases 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) (i.e., 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.

Descriptive Statistics for SNPs at the *IL4* Locus.

Table 1

	Proportion of Private SNPs	Permutation P-value	θ_W^a	π^b	Tajima's D	Permutation P-value
Combined Sample (n=142)	-	-	12.53×10^{-4}	11.09×10^{-4}	-0.356	-
Non-Asthma Controls (n=70)	0.076		11.26×10^{-4}	11.00×10^{-4}	-0.073	
Asthma Cases (n=72)	0.211	0.031	12.71×10^{-4}	11.20×10^{-4}	-0.375	0.0129

^a Watterson's 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.²³

^b Average number of pairwise differences between samples.²²