



# Further replication studies of the EVE Consortium meta-analysis identifies 2 asthma risk loci in European Americans

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters

Citation	Myers, Rachel A., Blanca E. Himes, Christopher R. Gignoux, James J. Yang, W. James Gauderman, Cristina Rebordosa, Jianming Xie, et al. 2012. Further replication studies of the EVE Consortium meta-analysis identifies 2 asthma risk loci in European Americans. <i>Journal of Allergy and Clinical Immunology</i> 130, no. 6: 1294–1301. doi:10.1016/j.jaci.2012.07.054.
Published Version	doi:10.1016/j.jaci.2012.07.054
Citable link	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:27005776">http://nrs.harvard.edu/urn-3:HUL.InstRepos:27005776</a>
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP</a>



Published in final edited form as:

*J Allergy Clin Immunol.* 2012 December ; 130(6): 1294–1301. doi:10.1016/j.jaci.2012.07.054.

## Further Replication Studies of the EVE Consortium Meta-Analysis Identifies Two Asthma Risk Loci in European Americans

Rachel A. Myers, PhD<sup>a</sup>, Blanca E. Himes, PhD<sup>b</sup>, Christopher R. Gignoux, MS<sup>c</sup>, James J. Yang, PhD<sup>d</sup>, W. James Gauderman, PhD<sup>e</sup>, Cristina Rebordosa, MD<sup>f</sup>, Jianming Xie, PhD<sup>a,g</sup>, Dara G. Torgerson, PhD<sup>c</sup>, Albert M. Levin, PhD<sup>d</sup>, James Baurley, PhD<sup>e</sup>, Penelope E. Graves, PhD<sup>f</sup>, Rasika A. Mathias, ScD<sup>h</sup>, Isabelle Romieu, MD, ScD<sup>i</sup>, Lindsey A. Roth, MA<sup>c</sup>, David Conti, PhD<sup>e</sup>, Lydia Avila, MD<sup>j</sup>, Celeste Eng, BS<sup>c</sup>, Hita Vora, MS<sup>e</sup>, Michael A. LeNoir, MD<sup>k</sup>, Manuel Soto-Quiros, PhD<sup>l</sup>, Jinghua Liu, PhD<sup>e</sup>, Juan C. Celedon, MD, PhD<sup>l</sup>, Harold J. Farber, MD<sup>m</sup>, Rajesh Kumar, MD, MPH<sup>n</sup>, Pedro C. Avila, MD<sup>o</sup>, Kelley Meade, MD<sup>p</sup>, Denise Serebrisky, MD<sup>q</sup>, Shannon Thyne, MD<sup>c</sup>, William Rodriguez-Cintron, MD<sup>r</sup>, Jose R. Rodriguez-Santana, MD<sup>s</sup>, Luisa N. Borrell, DDS, PhD<sup>t</sup>, Robert F. Lemanske Jr., MD<sup>u</sup>, Eugene R. Bleeker, MD<sup>v</sup>, Deborah A. Meyers, PhD<sup>v</sup>, Stephanie J. London, MD, PhD<sup>w</sup>, Kathleen C. Barnes, PhD<sup>h</sup>, Benjamin A. Raby, MD<sup>b</sup>, Fernando D. Martinez, MD<sup>f</sup>, Frank D. Gilliland, PhD<sup>e</sup>, L. Keoki Williams, MD, MPH<sup>x</sup>, Esteban G. Burchard, MD, MPH<sup>c,y</sup>, Scott T. Weiss, MD<sup>b</sup>, Dan L. Nicolae, PhD<sup>a,z,\*</sup>, and Carole Ober, PhD<sup>a,\*</sup>

<sup>a</sup>Department of Human Genetics, University of Chicago, Chicago, USA

<sup>b</sup>Channing Laboratory, Department of Medicine, Harvard Medical School, Boston, USA

<sup>c</sup>Department of Medicine, University of California, San Francisco, San Francisco, USA

<sup>d</sup>Department of Public Health Services, Henry Ford Health System, Detroit, USA

<sup>e</sup>Department of Preventative Medicine, University of Southern California, Los Angeles, USA

<sup>f</sup>Arizona Respiratory Center and BIO5 Institute, University of Arizona, Tucson, USA

<sup>g</sup>School of Biological Sciences and Medical Engineering, School of Biological Sciences and Medical Engineering, Southeast University, Nanjing, China

<sup>h</sup>Department of Medicine, Johns Hopkins University, Baltimore, USA

<sup>i</sup>Instituto Nacional de Salud Publica, Mexico City, Mexico

<sup>j</sup>Hospital Nacional de Niños, San Jose, Costa Rica

<sup>k</sup>Bay Area Pediatrics, Oakland, USA

<sup>l</sup>Division of Pediatric Pulmonology, University of Pittsburgh, Pittsburgh, USA

<sup>m</sup>Section of Pulmonology, Baylor College of Medicine, Houston, USA

<sup>n</sup>Division of Allergy and Immunology, Children's Memorial Hospital, Chicago, USA

© 2012 American Academy of Allergy, Asthma and Immunology. Published by Mosby, Inc. All rights reserved.

Corresponding Author: Carole Ober, 920 E 58<sup>th</sup> St, CLSC room 425, Chicago IL, 60637. Phone: (773) 834-0735, Fax: (773) 834-0505, c-ober@genetics.uchicago.edu.

\*These authors contributed equally to this work.

**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.

<sup>o</sup>Department of Medicine, Northwestern University, Chicago, USA

<sup>p</sup>Children's Hospital of Oakland, Oakland, USA

<sup>q</sup>Jacobi Medical Center and Department of Pediatrics, Albert Einstein College of Medicine, Bronx, USA

<sup>r</sup>Veterans Affairs Medical Center, San Juan, Puerto Rico

<sup>s</sup>Centro de Neumologia Pediatrica, San Juan, Puerto Rico

<sup>t</sup>Department of Health Sciences, Lehman College, City University of New York, New York, USA

<sup>u</sup>Department of Pediatrics, University of Wisconsin, Madison, USA

<sup>v</sup>Center for Human Genomics, Wake Forest School of Medicine, Winston-Salem, USA

<sup>w</sup>National Institute of Environmental Health Sciences, Research Triangle Park, USA

<sup>x</sup>Center for Health Services Research, Department of Internal Medicine, Henry Ford Health System, Detroit, USA

<sup>y</sup>Department of Bioengineering & Therapeutic Sciences, Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco, San Francisco

<sup>z</sup>Departments of Medicine and Statistics, Departments of Medicine and Statistics, University of Chicago, Chicago

## Abstract

**Background**—Genome-wide association studies of asthma have implicated many genetic risk factors, with well-replicated associations at approximately 10 loci that account for only a small proportion of the genetic risk.

**Objectives**—We aimed to identify additional asthma risk loci by performing an extensive replication study of the results from the EVE Consortium meta-analysis.

**Methods**—We selected 3186 SNPs for replication based on the p-values from the EVE Consortium meta-analysis. These SNPs were genotyped in ethnically diverse replication samples from nine different studies, totaling to 7202 cases, 6426 controls, and 507 case-parent trios. Association analyses were conducted within each participating study and the resulting test statistics were combined in a meta-analysis.

**Results**—Two novel associations were replicated in European Americans: rs1061477 in the *KLK3* gene on chromosome 19 (combined OR = 1.18; 95% CI 1.10 – 1.25) and rs9570077 (combined OR = 1.20 95% CI 1.12–1.29) on chromosome 13q21. We could not replicate any additional associations in the African American or Latino individuals.

**Conclusions**—This extended replication study identified two additional asthma risk loci in populations of European descent. The absence of additional loci for African Americans and Latino individuals highlights the difficulty in replicating associations in admixed populations.

## Keywords

Asthma; genetic risk factors; meta-analysis; *KLK3*

## Introduction

Asthma is a common and complex disease, affecting approximately 300 million people worldwide (1). There is strong evidence for a genetic contribution to asthma risk, with

heritability estimates ranging between 35% and 95% (2, 3) and many genetic associations reported (reviewed in (4)). Two large meta-analyses by the GABRIEL (5) and EVE (6) Consortia and many genome-wide association studies (7–13) for asthma have been reported in the last five years. The meta-analysis by the GABRIEL Consortium focused on populations of European ancestry and included twenty-three different studies from across Europe and Canada. Their findings included associations in or near the genes *IL1RL1-IL1RL18*, *HLA-DQ*, *IL33*, *SMAD3*, and *GSDMA/GSDMB* (17q21). Most of these loci had been previously been implicated in asthma or asthma-related phenotypes (e.g., 17q21 (7), *HLA-DQ* (14–16), *IL33* (17), *IL1RL1* (17, 18)). Most of these associations were also observed in the EVE Consortium meta-analysis, which included both primary and replication individuals from European American, African American and Latino populations. The EVE Consortium study also reported associations with single nucleotide polymorphisms (SNPs) near the *TSLP* gene, which was reported in a sub-analysis in GABRIEL as well as in previous studies (12, 19), and identified a novel association among African American individuals with SNPs in the *PYHINI* gene.

In the original EVE meta-analysis, we attempted replication of 13 SNPs from separate regions. The selected SNPs had association ( $p$ -value  $< 10^{-6}$ ) in either one of the three ethnic groups or in the combined sample. For the current study, we hypothesized that some SNPs with larger association  $p$ -values might also represent true associations. To address this possibility, we performed a deeper replication study of the results of the EVE meta-analysis. To this end, we genotyped a panel of 3,186 candidate SNPs in a replication sample totaling 7,709 individuals comprising European American, African American, and Latino subjects with asthma.

## Methods

### EVE Consortium Meta-Analysis (Primary Analysis)

The detailed methods for the EVE consortium meta-analysis were previously published (6). Briefly, the EVE meta-analysis was comprised of 4,867 European Americans (1,486 cases, 1,539 controls, and 620 trios), 4,644 Latinos (606 cases, 792 controls, and 1,082 trios), and 3,447 African Americans (1,154 cases, 1,054 controls, and 413 trios). Two million HapMap SNPs were either genotyped or imputed (20) and were analyzed for association with asthma. The meta-analysis test statistic was a linear combination of normally distributed test statistics weighted by the square root of the study sample size. Significance of the meta-analysis test statistic was assessed using standard normal approximations. This analysis will henceforth be referred to as the primary analysis.

### SNP Selection

We selected SNPs for additional genotyping based on the results from the original EVE meta-analysis (referred to here as the primary study. This include all SNPs with a  $p$ -value  $< 1 \times 10^{-4}$  in any one group (i.e. European American, African American, or Latino American individuals) or a  $p$ -value  $< 1 \times 10^{-3}$  in the combined meta-analysis of the groups combined. A less conservative  $p$ -value cut-off was used for the combined meta-analysis than the ethnic specific analyses because the combined analysis was more powerful and showed a larger number of small  $p$ -values than the ethnicity-specific analyses. We also genotyped an additional 407 SNPs that were informative for determining genetic ancestry (i.e. ancestry informative markers [AIMs]).

### Subjects and Genotyping

Asthma cases, non-asthma controls, and asthma case-parent trios were recruited from locations across North and Central America by the same investigators contributing samples

to the primary analysis. Descriptions of the replication samples and ascertainment schemes were previously reported (Supplemental Table 4 and Supplemental Note in reference (6)). Genotyping was performed using a custom Illumina iSELECT array at the Southern California Genotyping Consortium (SCGC) at the University of California – Los Angeles.

### Statistical Analysis

Each EVE center performed quality control and association analyses on their own data sets, using protocols identical to those used in the primary analysis (6). The description of each center's quality control and statistical analyses, including covariates, is presented in the supplemental methods. A shared summary file for each replication study population included SNP name, risk allele, alternative allele, risk allele frequency in cases, risk allele frequency in controls, normal or chi-square distributed test statistic, p-value, odds ratio, and the standard error of the log odds ratio. Test statistics were converted to a standard normal distribution and all results were checked for consistent allelic directionality across all study populations. The meta-analysis test statistics were calculated as a linear, weighted combination of the individual study population test statistics. The weights ( $w$ ) were a function of sample size ( $N$ ), proportion of cases ( $v$ ), and allele frequency ( $p$ ) and defined as:  $w = \sqrt{v(1-v)p(1-p)N}$  (21). P-values were obtained using standard normal approximations. The combined odds ratios were calculated as a linear combination of log odds ratios with weights proportional to the standard errors of the log odds ratios. All statistical analyses were completed using the statistical software R.

### Enrichment Analysis

We investigated whether or not there was an enrichment of small p-values at the 0.01 level in each of the four meta-analyses, one each in the European American, African American, Latino, and combined sample. To do this analysis we first constructed a set of p-values that are not correlated or in linkage disequilibrium (LD). For each analysis, we grouped SNPs according to physical location in the genome. Then, for each group of SNPs, we constructed a region-based p-value by taking the smallest p-value from the group of SNPs and multiplied that p-value by the number of SNPs in the group. This resulted in a region-based p-value that has two properties: 1) the region-based p-value is a conservative multiple testing corrected p-value for all of the SNPs within each genomic region and 2) the region-based p-values are independent of each other. A binomial test was used to ascertain enrichments of small ( $< 0.01$ ), independent, region-based p-values. The null hypothesis in this test is the region-based p-values are uniformly distributed and we do not observe any replications, whereas the alternative hypothesis is there exist an excess of small region-based p-values. The assumption in the binomial test is that the all of the region-based p-values are independent of each other, whereas this assumption of independence would be violated if the binomial test was applied to the unadjusted p-values

### Results

Of the 5,560 SNPs selected for replication, 3,186 passed the Illumina array design and were successfully genotyped in 7,202 cases, 6,426 controls, and 507 case-parent trios (Table I). The 3,186 SNPs included 177 SNPs from the five genomic regions that were replicated in our earlier study (6). The remaining 3,009 SNPs (listed in Table EI in the Online Repository) were in 690 genomic regions that showed a signal of association in the primary analysis: 128 SNPs in 68 regions in European Americans, 340 SNPs in 124 regions in African Americans, 258 SNPs in 74 regions in Latinos, and 2,450 SNPs in 595 regions in the combined sample. The SNPs and regions are not entirely mutually exclusive among the ethnic groups, 167 SNPs overlapped between two or more of the samples. The average power to replicate these SNPs was highest among African Americans and lowest in Latinos

(see Table EII in the Online Repository). The Q-Q plots for the 177 SNPs from the five previously replicated regions show substantial inflation of small p-values in the European American, Latino American, and the combined samples, but not in the African American samples (Figure E1A–D in the Online Repository). In contrast, the Q-Q plots (Figure 1A–D) of p-values for the ethnicity-specific replications revealed an inflation of small p-values only in the European Americans, but not in the African American, Latino American, or combined replication samples.

A summary of the replication results is shown in Table II. SNPs selected for replication based on a group-specific result in the primary meta-analysis were analyzed in their respective replication samples (e.g. primary meta-analysis results seen in European American individuals were analyzed in a European American replication population). In European American individuals, we expected fewer than one of the 128 SNPs to have a p-value < 0.001 by chance. We observed 2 SNPs with a p-value < 0.001 in these samples, which is a significant enrichment after accounting for linkage disequilibrium (LD) among SNPs (p-value = 0.031). Only one SNP had a p-value < 0.001 in the African American individuals and in the combined sample, which does not represent a significant enrichment of low p-values in either group. Finally, none of the 258 SNPs had a p-value < 0.001 in the Latino American samples.

The SNPs with the most significant combined (primary + replication) p-values are shown in Table III. The two most significant associations were in the European American sample, with p-values < 0.001 and combined (primary + replication) p-values approaching genome-wide levels of significance ( $2 \times 10^{-8}$ ). The first replication is with rs1061477 with a combined primary and replication p-value of  $2.3 \times 10^{-8}$  and a combined odds ratio of 1.18 (95% CI 1.10 – 1.25) (Figure 2A). This SNP is in an intron in the *KLK3* gene (encoding kallikrein-related serine peptidase 3) on chromosome 19q13 (Figure 2B). The second replication is with rs9570077, with a combined p-value of  $9.9 \times 10^{-8}$  and combined odds ratio of 1.20 (95% CI: 1.12–1.29) (Figure 3A). This SNP is located on chromosome 13q21, approximately 420 kb downstream of the *DIAPH3* gene, which encodes diaphanous homolog 3 (Figure 3B). Finally, seven SNPs showed small p-values in both the primary sample (p-value <  $1 \times 10^{-4}$ ) and the replication sample (p-value < 0.01) but large combined p-values, indicating that the associations were in opposite directions. One such SNP is rs9891949 with a replication p-value = 0.005 and a primary p-value =  $7.5 \times 10^{-7}$  in European Americans. The risk SNP differed between studies (combined OR = 1.00, 95% CI = 0.94 – 1.07, Figure 4). This SNP is located on chromosome 17p13, 1 kb from the gene *AURKB* which encodes serine/threonine-protein kinase aurora-B.

## Discussion

We report here two additional replicated associations with asthma in European Americans. Despite using a conservative test for enrichment, we observed a significant excess of small p-values in the European American replication samples. Two of the three SNPs match the direction of association seen in the primary analysis.

One of the replicated SNPs, rs1061477, lies in the *KLK3* gene on chromosome 19q13. This region includes other genes in the kallikrein family and lies within a region that was previously implicated in asthma and related phenotypes through linkage studies in European and European American families (14, 22, 23). Although not previously considered as an asthma candidate in this gene rich region on 19q13, the fine-scale resolution of GWAS specifically localized a strong association signal to the *KLK3* gene. *KLK3* encodes PSA (prostate specific antigen) which can invoke an immune response by inducing IFN- $\gamma$  (interferon gamma) secretion (24). IFN- $\gamma$  secretion is a key immunomodulatory molecule in



asthma pathogenesis (25). *KLK3* gene expression is highest in the prostate but at lower levels in other tissues (26), including bronchial epithelium from individuals with asthma (27).

The second replicated association, rs9570077, is located at 13q21 and resides between two pseudogenes, *RPP40P2* and *POLR3KPI*. SNPs in this region show very little LD ( $r^2$  0.075) with SNPs around the closest gene, *DIAPH3*. The 13q21 region has also been implicated in asthma through linkage analysis (28). It is possible that the associated region encodes a long-range enhancer of *DIAPH3* or another gene on chromosome 13q21.

In addition to the two replicated associations, seven SNPs had small p-values in both the primary (p-value  $< 1 \times 10^{-4}$ ) and replication (p-value  $< 0.01$ ) studies, but associations were in opposite directions. The most dramatic example of this was the associations with rs9891949, located 1 kb from the *AURKB* gene on chromosome 17p13. The association signals were strong in both the primary and replication studies; the combined p-value ignoring association directionality =  $5.8 \times 10^{-8}$  in European Americans. However, the combined p-value accounting for directionality = 0.316 in European Americans (Figure 4). This type of association with opposite alleles associated in primary and replication samples has been reported in other studies of asthma, where it has been attributed to a genotype-environment interactions with maternal asthma status (15) or to differences in LD between population groups (13). In our study, it is difficult to attribute this pattern to either explanation because participants in the Children's Health Study (CHS) in Los Angeles were included in both the primary and replication samples, and opposite association patterns were observed within this one study. There are no substantial differences between the CHS children included in the primary vs. replication study. In particular, we ruled out differences in sex ratios, ages of onset of asthma, and indicators of atopy between the European American primary and replication samples that showed association with one allele vs. the other (Table EIII in the Online Repository). Finally, it is possible that these seven associations are purely chance occurrences resulting from random variation in association signals. Thus, it remains to be determined whether the seven associations of this type represent true associations with direction of effect determined by an as yet unknown variable or if they are simply the result of random variation.

Notably, none of the associations were replicated in African Americans or Latinos. This observation is surprising because the power to replicate the primary associations signals was high in all populations, with greatest power in the African Americans and least in the Latinos (Table EII). However, there was a wider range of power estimates in the Latinos and African Americans compared to European Americans. The combination of the lowest average power and increased range of power for the Latino sample may have contributed to the lack of replication in that sample, but is less likely to have contributed to the lack of replication in the African Americans because power was the highest in that group.

It is possible that the association signals observed in the primary analyses were more likely to be false positive associations in the admixed individuals due to reduced accuracy of imputation in these populations. In the primary analysis, we included samples that were genotyped with different platforms (Affymetrix or Illumina) and on arrays with different numbers of SNPs. To enable analyses with the most inclusive set of SNPs, each participating center imputed all SNPs that were not present on their platform but present on at least one of the other platforms. Thus, different SNPs were imputed in different samples, although all imputations were referenced against HapMap release 21 samples. Whole genome sequences were available for a subset of the European American (n = 21), African American (n = 32), and Latino (n = 24) individuals from the primary samples, allowing us to determine concordance rates between imputed genotypes and true genotypes. We observed similar concordance rates (98%) in all three populations, suggesting that our ability to

replicate associations was not influenced by ethnicity-based differences in imputation accuracy.

To address the possibility that signal heterogeneity across the primary studies limited our ability to replicate, we investigated a measure technical heterogeneity: the proportion of variation in odds ratios attributed to signal heterogeneity rather than random chance ( $I^2$ ). In all populations, the heterogeneity ( $I^2$ ) of the odds ratios is very low (<25%) (29), indicating that most of the variation in odds ratio estimates is a result of random variation rather than study based differences in effect sizes. Less than five percent of SNPs had  $I^2$  values considered highly heterogeneous (>75%). Therefore, heterogeneity of the primary association signals does not explain the lack of replication in the admixed samples.

Alternatively, the lack of replication in the non-European, admixed populations may have been due to differences between the primary and replication samples. In particular, differences in ancestry or patterns of linkage disequilibrium (LD) may be limiting replication in these samples. For example, it is well known that asthma prevalences differs within Latinos (30) with highest prevalence among Puerto Ricans and lowest among Mexicans (31). Although both the primary and replication studies included Mexican Americans and Puerto Rican Americans from the U.S., the primary study also included subjects from Mexico City whereas the replication study included subjects from Costa Rica and Honduras. Likewise, our 'African ancestry' group included individuals from Barbados in the primary analysis whereas only African Americans were included in the replication studies. Differences in global and local ancestry between these samples could have impacted our ability to replicate associations, even after adjusting for global ancestry (32). In addition, patterns of LD in admixed populations may also be an important factor, notably if the original association signals at typed or imputed SNPs are due to LD with causal variation. Additional studies in these populations may shed light on these possibilities.

In summary, we report two additional asthma risk loci in European Americans that represent previously unknown genes or regions and possibly novel pathways of pathogenesis. While the effects of genotypes at each locus on risk are small, we provide evidence for two new asthma candidate regions where family-based studies previously showed evidence of linkage. Further investigations are required to identify the specific causal variants and the mechanism of the associations for *KLK3* and the 13q21 locus. The lack of replication in the African and Latino American populations suggests that additional strategies are warranted when studying these groups (33). Careful attention to sample composition and recruitment strategies, environmental risk factors, and admixture corrections may improve the identification and replication of genetic risk factors for asthma in admixed populations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors acknowledge Rebecca Anderson, David Witonsky, Duanny Alva, Gaby Ayala-Rodriguez, Ulysses Burley, Lisa Caine, Elizabeth Castellanos, Grace Y. Chiu, Jaime Colon, Denise DeJesus, Iliana Flexas, Dana B. Hancock, Blanca Lopez, Brenda Lopez, Louis Martos, Vivian Medina, Juana Olivo, Mario Peralta, Esther Pomares, Jihan Quraishi, Blanca del Rio Navarro, Elizabeth Nguyen, Johanna Rodriguez, Shahdad Saeedi, Sandra Salazar, Min Shi, Juan J. Sierra-Monge, Dean Soto, Ana Taveras for data collection, management, and analysis. The authors also acknowledge the support from J. Kiley, S. Banks-Schlegel, and W. Gan at the National Heart, Lung, and Blood Institute, all of the patients and families for their participation in these studies, and the numerous health care providers and community clinics for their support.

## Sources of Support



This work was supported by grants from the Office of the Director, NIH to C.O. and D.L.N. and the National Heart, Lung, and Blood Institute (HL101651 to C.O. and D.L.N.; HL087665 to D.L.N.; HL070831, HL072414 and HL049596 to C.O.; HL064307 and HL064313 to F.D.M.; HL075419, HL65899, HL083069, HL066289, HL087680, HL101543 and HL101651 to S.T.W.; HL079055 to L.K.W.; HL087699, HL49612, HL075417, HL04266 and HL072433 to K.C.B.; HL061768 and HL076647 to F.D.G.; HL087680 to W.J.G.; HL078885 and HL088133 to E.G.B.; HL87665 to D.A.M.); the National Institutes of Allergy and Infectious Disease (AI070503 to C.O.; AI079139 and AI061774 to L.K.W.; AI50024, AI44840 and AI41040 to K.C.B.; and AI077439 to E.G.B.), the National Institute of Diabetes and Digestive and Kidney Diseases to L.K.W. (DK064695); the National Institutes of Environmental Health Sciences (ES09606, ES018176 and ES015903 to K.C.B.; ES007048, ES009581, R826708, RD831861 and ES011627 to F.D.G.; ES015794 to E.G.B.; and the Division of Intramural Research, Z01 ES049019, to S.J.L.); the National Center for Research Resources (RR03048 to K.C.B.), the Environmental Protection Agency (83213901 and R-826724 to K.C.B.), the American Asthma Foundation and the Fund for Henry Ford Hospital (to L.K.W.), Mary Beryl Patch Turnbull Scholar Program (to K.C.B.); and the Flight Attendant Medical Research Institute (FAMRI), Robert Wood Johnson Foundation (RWJF) Amos Medical Faculty Development Award, the American Asthma Foundation, and the Sandler Foundation (to E.G.B.).

## Abbreviations

<b>SNP</b>	Single Nucleotide Polymorphism
<b>OR</b>	Odds Ratio
<b>KLK3</b>	Kallikrein-related serine peptidase 3
<b>PSA</b>	Prostate specific antigen
<b>LD</b>	Linkage Disequilibrium

## References

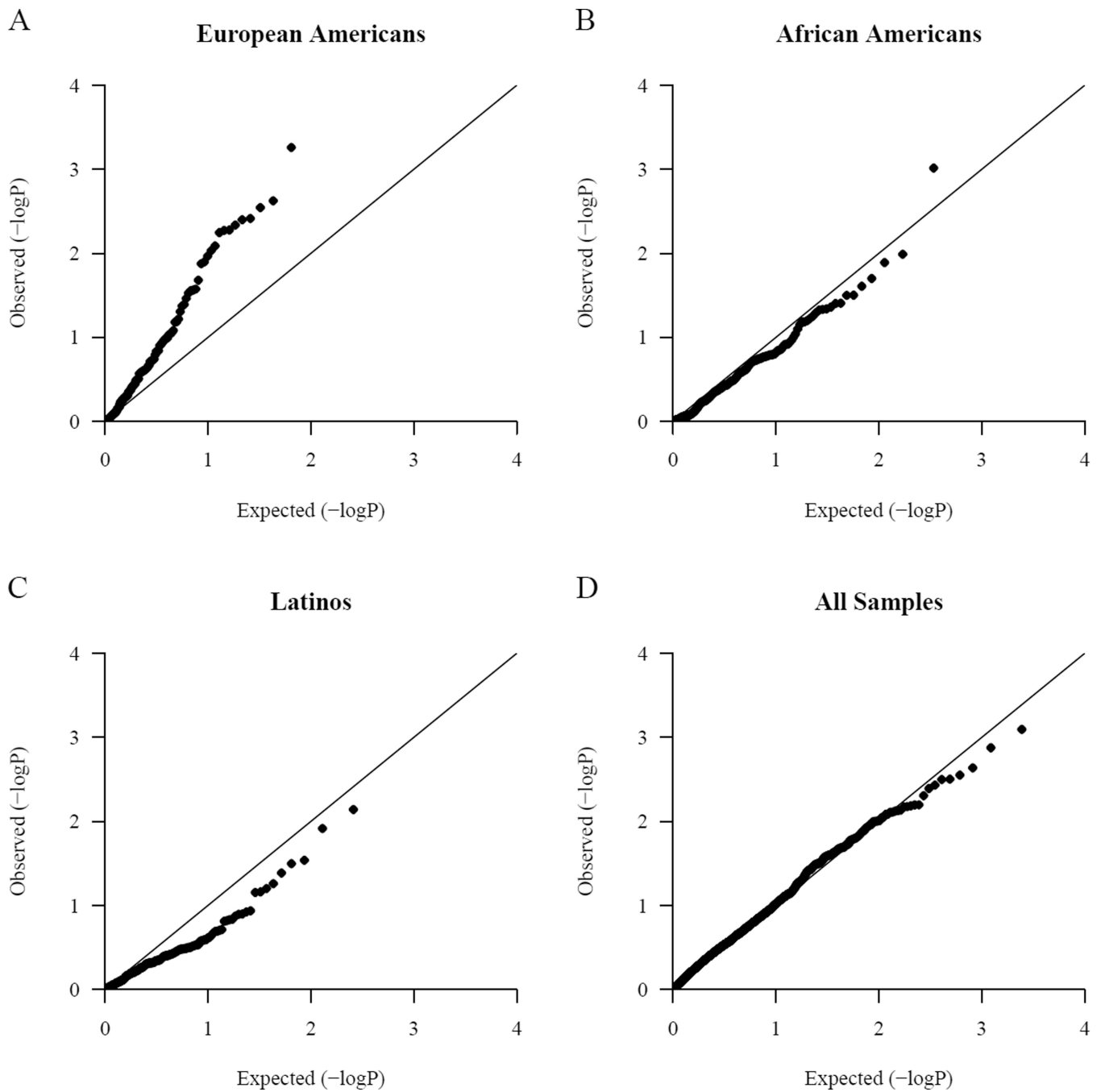
- Masoli M, Fabian D, Holt S, Beasley R. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy*. May; 2004 59(5):469–478. [PubMed: 15080825]
- Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis*. Dec; 1990 142(6 Pt 1):1351–1358. [PubMed: 2252253]
- Nieminen MM, Kaprio J, Koskenvuo M. A population-based study of bronchial asthma in adult twin pairs. *Chest*. Jul; 1991 100(1):70–75. [PubMed: 2060393]
- Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective. *Immunol Rev*. Jul; 2011 242(1):10–30. [PubMed: 21682736]
- Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *The New England journal of medicine*. Sep 23; 2010 363(13):1211–1221. [PubMed: 20860503]
- Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet*. Sep; 2011 43(9):887–892. [PubMed: 21804549]
- 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*. Jul 26; 2007 448(7152):470–473. [PubMed: 17611496]
- Choudhry, S.; Taub, M.; Mei, R.; Rodriguez-Santana, J.; Rodriguez-Cintron, W.; Shriver, MD., et al. Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region; Validation Study 2008 Jul 01 Contract No. p. 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. *American journal of human genetics*. Jun 01; 2009 84(5):581–593. [PubMed: 19426955]
- Hancock DB, Romieu I, Shi M, Sienna-Monge J-J, Wu H, Chiu GY, et al. Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in Mexican children. *PLoS Genetics*. Aug 01. 2009 5(8):e1000623. [PubMed: 19714205]
- Mathias, RA.; Grant, AV.; Rafaels, N.; Hand, T.; Gao, L.; Vergara, C., et al. A genome-wide association study on African-ancestry populations for asthma; Multicenter Study 2010 Mar 01 Contract No. p. 2

12. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet.* Sep; 2011 43(9):893–896. [PubMed: 21804548]
13. 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.* Jan 7; 2010 362(1):36–44. [PubMed: 20032318]
14. Ober C, Tsalenko A, Parry R, Cox NJ. A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. *Am J Hum Genet.* Nov; 2000 67(5):1154–1162. [PubMed: 11022011]
15. Nicolae D, Cox NJ, Lester LA, Schneider D, Tan Z, Billstrand C, et al. Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21. *American journal of human genetics.* Mar 01; 2005 76(2):349–357. [PubMed: 15611928]
16. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, et al. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol.* Feb; 2010 125(2):328–335. e11. [PubMed: 20159242]
17. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadóttir A, Sulem P, Jonsdóttir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet.* Mar; 2009 41(3):342–347. [PubMed: 19198610]
18. Ferreira, MAR.; Merae, AF.; Medland, SE.; Nyholt, DR.; Gordon, SD.; Wright, MJ., et al. Association between ORMDL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia. *European journal of human genetics : EJHG.* Dec 08. 2010
19. Hunninghake GM, Soto-Quirós ME, Avila L, Kim HP, Lasky-Su J, Rafaels N, et al. TSLP polymorphisms are associated with asthma in a sex-specific fashion. *Allergy.* Dec 01; 2010 65(12):1566–1575. [PubMed: 20560908]
20. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using Sequence and Genotype Data to Estimate Haplotypes and Unobserved Genotypes. *Genetic Epidemiology.* Dec; 2010 34(8):816–834. [PubMed: 21058334]
21. Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genetic Epidemiology.* May 01; 2010 34(3):254–257. [PubMed: 19918758]
22. Venanzi S, Malerba G, Galavotti R, Lauciello MC, Trabetti E, Zanoni G, et al. Linkage to atopy on chromosome 19 in north-eastern Italian families with allergic asthma. *Clin Exp Allergy.* Aug; 2001 31(8):1220–1224. [PubMed: 11529891]
23. Xu J, Meyers DA, Ober C, Blumenthal MN, Mellen B, Barnes KC, et al. Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S populations: collaborative study on the genetics of asthma. *Am J Hum Genet.* Jun; 2001 68(6):1437–1446. [PubMed: 11349227]
24. Kodak JA, Mann DL, Klyushnenkova EN, Alexander RB. Activation of innate immunity by prostate specific antigen (PSA). *Prostate.* Nov 1; 2006 66(15):1592–1599. [PubMed: 16683268]
25. Finn PW, Bigby TD. Innate immunity and asthma. *Proc Am Thorac Soc.* May 1; 2009 6(3):260–265. [PubMed: 19387027]
26. Lawrence MG, Lai J, Clements JA. Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus. *Endocr Rev.* Aug; 2010 31(4):407–446. [PubMed: 20103546]
27. Bochkov YA, Hanson KM, Keles S, Brockman-Schneider RA, Jarjour NN, Gern JE. Rhinovirus-induced modulation of gene expression in bronchial epithelial cells from subjects with asthma. *Mucosal Immunol.* Jan; 2010 3(1):69–80. [PubMed: 19710636]
28. The Collaborative Study on the Genetics of Asthma. A genome-wide search for asthma susceptibility loci in ethnically diverse populations The Collaborative Study on the Genetics of Asthma (CSGA). *Nat Genet.* Apr; 1997 15(4):389–392. [PubMed: 9090385]
29. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003; 327:557–560. [PubMed: 12958120]

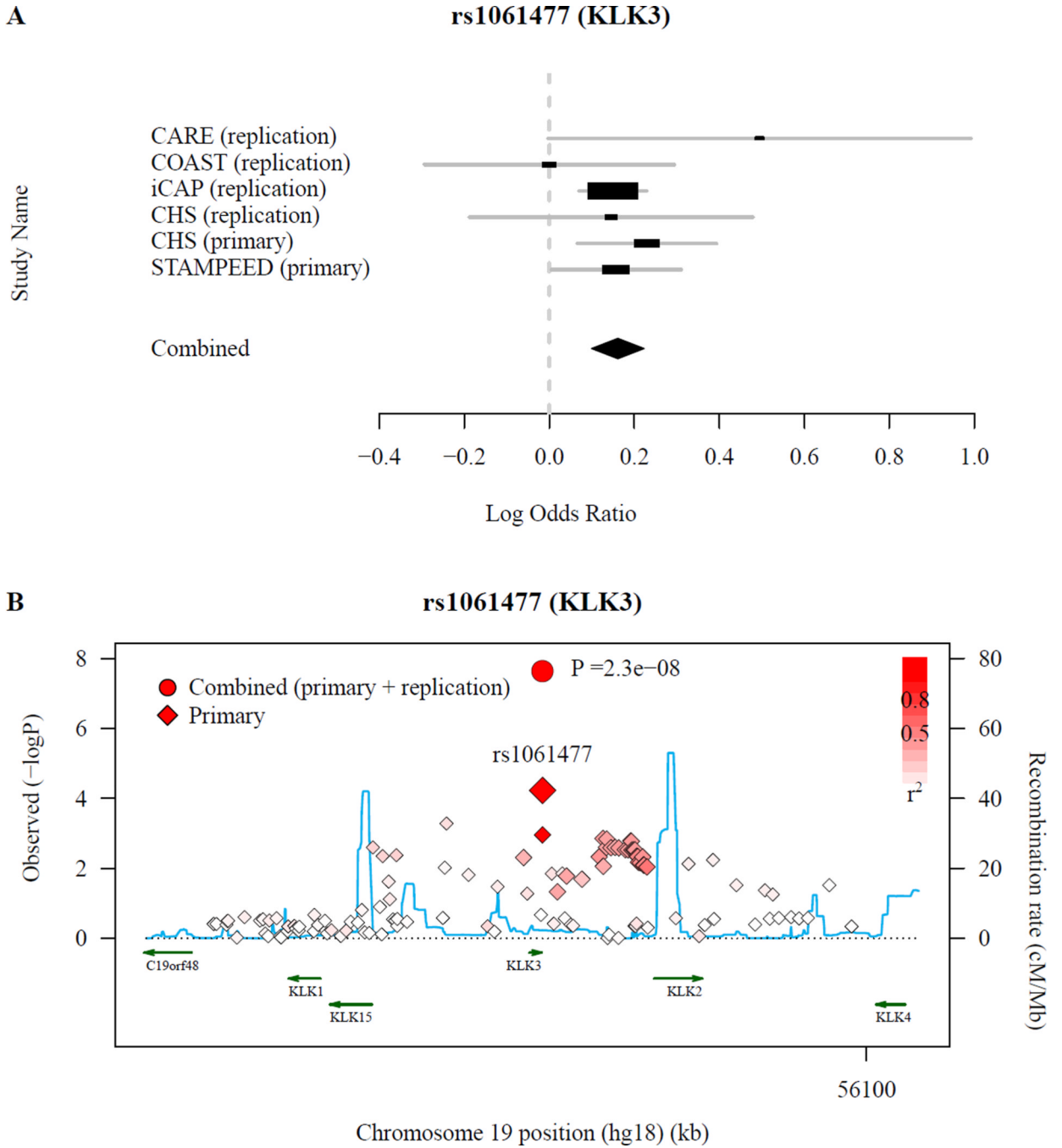
30. Ledogar RJ, Penchaszadeh A, Garden CC, Iglesias G. Asthma and Latino cultures: different prevalence reported among groups sharing the same environment . *Am J Public Health*. Jun; 2000 90(6):929–935. [PubMed: 10846511]
31. Moorman JE, Rudd RA, Johnson CA, King M, Minor P, Bailey C, et al. National surveillance for asthma--United States, 1980–2004. *MMWR Surveill Summ*. Oct 19; 2007 56(8):1–54. [PubMed: 17947969]
32. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet*. May; 2004 36(5):512–517. [PubMed: 15052271]
33. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature*. Jul 14; 2011 475(7355):163–165. [PubMed: 21753830]
34. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, et al. SCAN: SNP and copy number annotation. *Bioinformatics*. Jan 15; 2010 26(2):259–262. [PubMed: 19933162]
35. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. Dec 15; 2008 24(24):2938–2939. [PubMed: 18974171]
36. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. Jun 1; 2007 316(5829):1331–1336. [PubMed: 17463246]

### Key Messages

- Variation in *KLK3* and 13q21 show association with asthma in European Americans.
- Replication in admixed populations such as African Americans and Latinos is sensitive to genetic and technical heterogeneity.



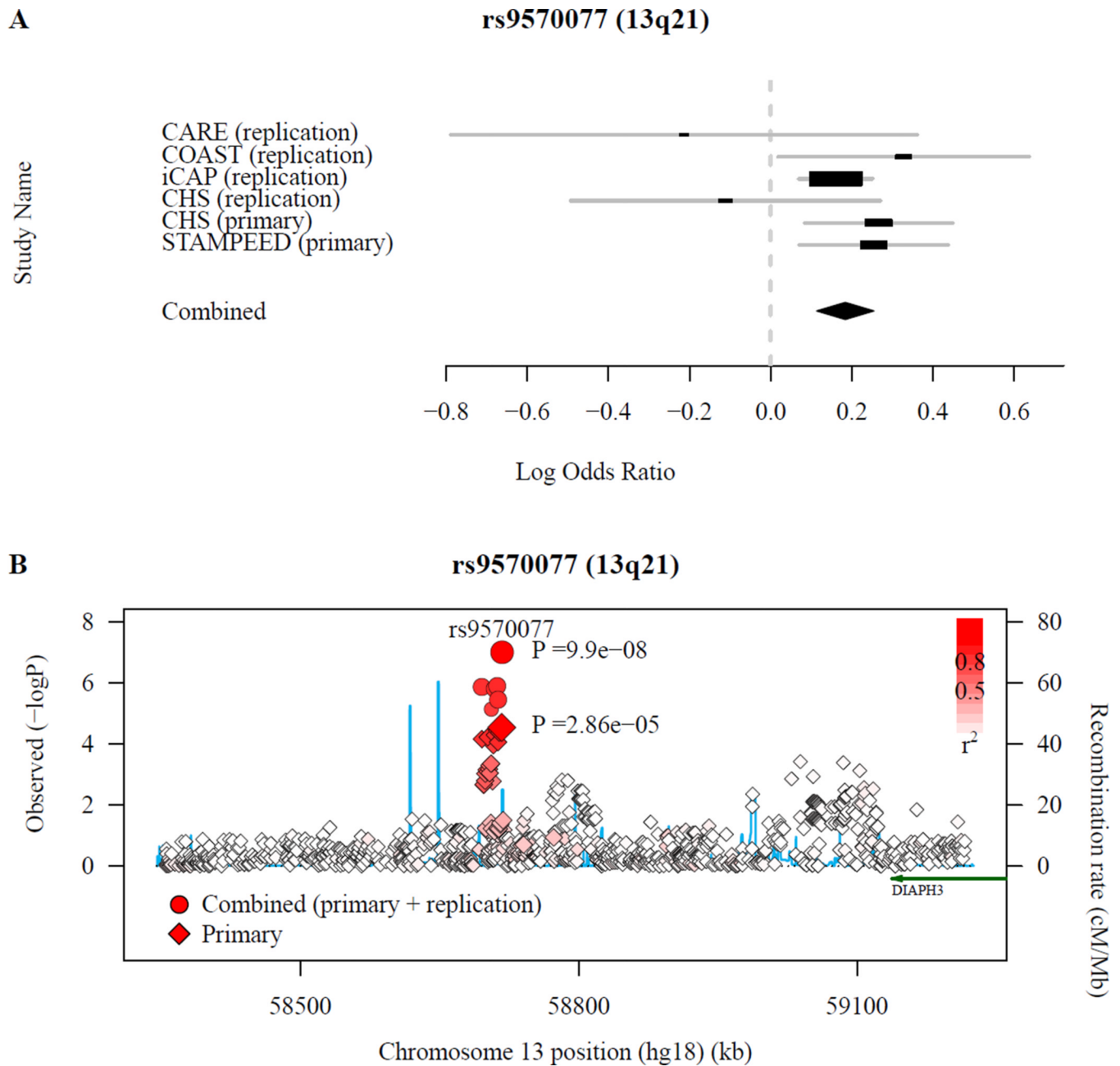
**Figure 1.** Q-Q plots of the replication results. In each plot (A–D), the observed  $-\log(p\text{-value})$  is plotted against the expected  $-\log(p\text{-value})$ . A) The 128 SNPs selected for replication in the European American samples, B) 340 SNPs selected for replication in the African Americans, C), 258 SNPs selected for replication in the Latinos, and D) 2450 SNPs selected for replication in all samples.



**Figure 2.** Results for rs1061477 in the *KLK3* gene. A) Forest plot showing the odds ratios (OR) and 95% confidence intervals for all European American studies in the primary and replication meta-analyses. Two European American studies included in the primary analysis (CARE and CAMP) were family-based studies and did not report odds ratios; those studies are not included in the figure. The R package *meta* was used for determining the combined OR and *rmeta* was used to generate the plot. B) Regional plot of the European American association p-values surrounding rs1061477. Diamond represents p-values for SNPs from the primary study and circles represent the combined (primary + replication) p-values. SNPs are shaded according to LD with the SNP rs1061477, red indicating high LD. The light blue line



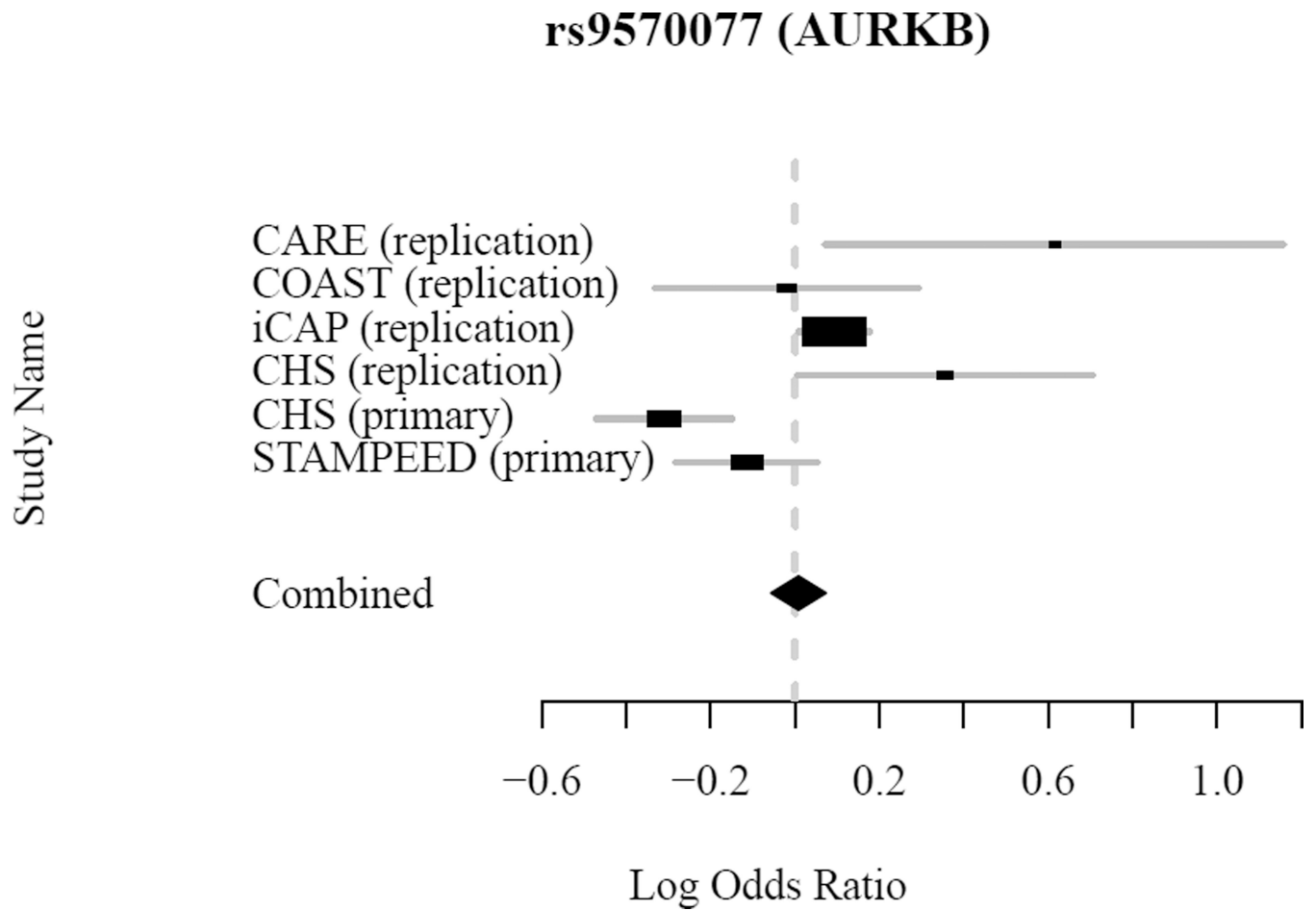
represents CEU 1000 genomes pilot recombination rates. This figure was generated using SNAP (35, 36).



**Figure 3.**

Results for rs9570077 on 13q21. A) Forest plot showing the odds ratios (OR) and 95% confidence intervals for all European American studies in the primary and replication meta-analyses. Two European American studies included in the primary analysis (CARE and CAMP) were family-based studies and did not report odds ratios; those studies are not included in the figure. The R package *meta* was used for determining the combined OR and *rmeta* was used to generate the plot. B) Regional plot of the European American association p-values surrounding rs9570077. Diamond represents p-values for SNPs from the primary study and circles represent the combined (primary + replication) p-values. SNPs are shaded according to LD with the SNP rs9570077, red indicating high LD. The light blue line

represents CEU 1000 genomes pilot recombination rates. This figure was generated using SNAP (35, 36).



**Figure 4.**

Forest plot of rs9891949 near *AURKB*. The odds ratios (OR) and 95% confidence intervals for all European American studies in the primary and replication meta-analyses. Two European American studies included in the primary analysis (CARE and CAMP) were family-based studies and did not report odds ratios; those studies are not included in the figure.

**Table I**

Sizes of the samples included in the EVE replication studies.

Sample*	EVE Site	N Cases	N Controls	N Trios
<i>European American individuals</i>				
CARE	Arizona	NA	NA	69
CHS	USC	155	157	NA
COAST	Chicago	174	188	NA
iCAP	Harvard	2,690	2,166	NA
<i>African American individuals</i>				
iCAP	Harvard	613	1,460	NA
SAGE	UCSF	755	494	NA
SAPPHIRE	Henry Ford	1,163	250	NA
CHS	USC	108	88	NA
<i>Latino individuals</i>				
CARE	Arizona	NA	NA	51
CHS	USC	289	287	NA
Costa Rica	Harvard	NA	NA	387
GALA2	UCSF	1,184	1,241	NA
HOLA	UCSF	71	95	NA

\* CARE - Childhood Asthma Research and Education, CHS - Children's Health Study, COAST-Childhood Origins of Asthma, iCAP - The i2b2 Crimson Asthma Project, SAGE – The Study of African Americans, Asthma, Genes and Environment, SAPPHIRE- Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity, Costa Rica- Genetics of Asthma in Costa Rica Study, GALA2-Genes-Environment and Admixture in Latino Asthmatics, HOLA-Honduran Latino Asthma Study

**Table II**

Enrichments of region-adjusted small P-values. Enrichment p-values, or the probability of observing  $X$  or more regions with p-value  $< 0.01$ , was determined using a binomial test.

Study Group	Number of Region-Based p-values $< 0.01$	Total Number of Regions Studied	Enrichment p-value	SNP Contributing to Each Region-Based p-value $< 0.01$
European American	3	68	0.031	rs1061477, rs9570077, rs2415036*
African American	1	124	0.71	rs8052139
Latino	1	74	0.52	rs17097180*
All samples	3	595	0.94	rs12734326, rs984176, rs11968112

\* SNPs for which the risk allele differs between the primary and replication studies; these are not included in Table III.



Table III

SNPs with combined (primary + replication) P-value  $< 1 \times 10^{-5}$  and replication P-value  $< 0.01$ . SNPs are ordered according to ethnicity and by chromosomal location\*.

Ethnicity**	Chr	SNP	Gene	Feature	Left Gene	Right Gene	Primary P-value	Replication P-value	Combined P-value
AA	16	rs8052139	ITGAX	intron	ITGAM	ITGAD	$4.5 \times 10^{-5}$	0.00097	$3.2 \times 10^{-7}$
EA	13	rs9570077	NA	NA	LOC100129308	DIAPH3	$2.7 \times 10^{-5}$	0.00055	$9.9 \times 10^{-8}$
EA	13	rs9538345	NA	NA	LOC100129308	DIAPH3	$3.6 \times 10^{-5}$	0.0040	$1.2 \times 10^{-6}$
EA	13	rs13378924	NA	NA	LOC100129308	DIAPH3	$6.9 \times 10^{-5}$	0.0029	$1.3 \times 10^{-6}$
EA	13	rs9538344	NA	NA	LOC100129308	DIAPH3	$5.1 \times 10^{-5}$	0.0038	$1.5 \times 10^{-6}$
EA	13	rs9538346	NA	NA	LOC100129308	DIAPH3	$8.6 \times 10^{-5}$	0.0053	$3.3 \times 10^{-6}$
EA	19	rs1061477	KLK3	intron	KLK15	KLK2	$5.9 \times 10^{-5}$	$8.4 \times 10^{-5}$	$2.3 \times 10^{-8}$
ALL	1	rs499913	NA	NA	CRNN	LCE5A	$6.2 \times 10^{-5}$	0.0032	$7.2 \times 10^{-7}$
ALL	1	rs1885531	NA	NA	CRNN	LCE5A	0.00016	0.0074	$3.7 \times 10^{-6}$
ALL	1	rs12734326	NA	NA	FMN2	GREM2	0.0010	0.0013	$7.1 \times 10^{-6}$
ALL	6	rs11964349	NA	NA	MYB	AH11	0.00011	0.0037	$1.5 \times 10^{-6}$
ALL	6	rs11968112	NA	NA	MYB	AH11	0.00077	0.0023	$9.0 \times 10^{-6}$
ALL	7	rs11489283	NA	NA	UBA52PI	NFE2L3	0.00078	0.00081	$3.6 \times 10^{-6}$
ALL	7	rs2214976	NA	NA	LOC646588	UBA52PI	0.00060	0.0041	$8.7 \times 10^{-6}$
ALL	9	rs12682865	ZNF618	intron	LOC100132609	AMBIP	$7.5 \times 10^{-6}$	0.0090	$2.2 \times 10^{-6}$
ALL	9	rs4979311	ZNF618	intron	LOC100132609	AMBIP	$2.4 \times 10^{-5}$	0.0078	$6.1 \times 10^{-7}$
ALL	11	rs984176	NA	NA	BARX2	LOC729717	0.00011	0.0077	$2.8 \times 10^{-6}$
ALL	16	rs11642009	CLEC16A	intron	LOC100130276	LOC100132071	$8.9 \times 10^{-5}$	0.0063	$1.7 \times 10^{-6}$
ALL	16	rs17673553	CLEC16A	intron	LOC100130276	LOC100132071	$1.5 \times 10^{-4}$	0.0031	$1.7 \times 10^{-6}$
ALL	16	rs12935657	CLEC16A	intron	LOC100130276	LOC100132071	0.00023	0.0065	$5.0 \times 10^{-6}$
ALL	20	rs734685	WFDC5	intron	KCNS1	WFDC12	0.00052	0.0028	$6.9 \times 10^{-6}$

\* SNPs were annotated using SCAN (34). SNPs previously replicated (e.g. SNPs in 17q21) are not shown.

\*\* AA denotes African American; EA, European American; ALL, all individuals