Genome-wide association study of asthma in individuals of African ancestry reveals novel asthma susceptibility loci

Michelle Daya, Nicholas Rafaels, Sameer Chavan, Henry Richard Johnston, Aniket Shetty, Christopher R. Gignoux, Meher Preethi Boorgula, Monica Campbell, Pissamai Maul, Trevor Maul, Candelaria Vergara, Albert M. Levin, Genevieve Wojcik, Dara G. Torgerson, Victor E. Ortega, Ayo Doumatey, Maria Ilma Araujo, Pedro C. Avila, Eugene Bleecker, Carlos Bustamante, Luis Caraballo, Georgia M. Dunston, Mezbah U. Faruque, Trevor S. Ferguson, Camila Figueiredo, Jean G. Ford, Pierre-Antoine Gourraud, Nadia N. Hansel, Ryan D. Hernandez, Edwin Francisco Herrera-Paz, Eimear E. Kenny, Jennifer Knight-Madden, Rajesh Kumar, Leslie A. Lange, Ethan M. Lange, Antoine Lizee, Alvaro Mayorga, Deborah Meyers, Dan L. Nicolae, Timothy D. O'Connor, Ricardo Riccio Oliveira, Christopher O. Olopade, Olufunmilayo Olopade, Zhaohui S. Qin, Charles Rotimi, Harold Watson, Rainford J. Wilks, L. Keoki Williams, James G. Wilson, Carole Ober, Esteban G. Burchard, Terri H. Beaty, Margaret A. Taub, Ingo Ruczinski, Rasika Ann Mathias, Kathleen C. Barnes
on behalf of CAAPA

Abstract

BACKGROUND: Asthma is a complex disease with striking disparities across racial and ethnic groups, which may be partly attributable to genetic factors. One of the main goals of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) is to discover genes conferring risk to asthma in populations of African descent.

METHODS: We performed a genome-wide meta-analysis of asthma across 11 CAAPA datasets (4,827 asthma cases and 5,397 controls), genotyped on the African Diaspora Power Chip (ADPC) and including existing GWAS array data. The genotype data were imputed up to a whole genome sequence reference panel from n=880 African ancestry individuals for a total of 61,904,576 SNPs. Statistical models appropriate to each study design were used to test for association, and results were combined using the weighted Z-score method. We also used admixture mapping as a complementary approach to identify loci involved in asthma pathogenesis in subjects of African ancestry.

RESULTS: SNPs rs787160 and rs17834780 on chromosome 2q22.3 were significantly associated with asthma (p=6.57×10^{-9} and 2.97×10^{-8}, respectively). These SNPs lie in the intergenic region between the Rho GTPase Activating Protein 15 (*ARHGAP15*) and Glycosyltransferase Like Domain Containing 1 (*GTDC1*) genes. Four low frequency variants on chromosome 1q21.3, which may be involved in the "atopic march" and which are not polymorphic in Europeans, also showed evidence for association with asthma ($1.18 \times 10^{-6} \leq p \leq 3.06 \times 10^{-6}$). SNP rs11264909 on chromosome 1q23.1, close to a region previously identified by the EVE asthma meta-analysis as having a putative African ancestry specific effect, only showed differences in counts in subjects homozygous for alleles of African ancestry. Admixture mapping also identified a significantly associated region on chromosome 6q23.2, which includes the Transcription Factor 21 (*TCF21*) gene, previously shown to be differentially expressed in bronchial tissues of asthmatics and non-asthmatics.

CONCLUSIONS: We have identified a number of novel asthma association signals warranting further investigation.

BACKGROUND

Asthma is a complex disease where the interplay between genetic factors and environmental exposures controls susceptibility and disease progression. In the U.S., ethnic minorities are disproportionally affected by asthma. For example, African Americans and Puerto Ricans have higher asthma-related morbidity and mortality rates compared to European Americans [1-3]. In addition to environmental, cultural and socio-economic risk factors, genetic factors, possibly from a common background ancestry, likely underlie some of these disparities in the health burden of asthma in the U.S. Despite the relatively high burden of disease, representation of African ancestry populations in asthma genome-wide association studies (GWAS) to date has been limited. The largest asthma GWAS focused solely on African ancestry populations included only 908 asthma cases [4], and only 1,612 cases of African ancestry were represented in the largest asthma meta-analysis to date [5]. The Trans-National Asthma Genetic Consortium (TAGC) is currently conducting a meta-analysis of more than 100,000 subjects, of which only 2,149 are cases of African ancestry. The discovery of genetic risk factors for asthma in African ancestry populations has also been further hampered by legacy commercial genotyping arrays that do not provide adequate coverage of linkage disequilibrium (LD) patterns in African ancestry populations.

To address these research disparities and the paucity of information on African genetic diversity, we established the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA). We first performed whole genome sequencing (WGS) of samples collected from individuals (394 asthmatics, 424 non-asthmatics and 60 with unknown asthma status) who self-reported African ancestry from 19 North, Central and South American and Caribbean populations (446 individuals from 9 African American populations, 43 individuals from Central America, 121 individuals from 3 South American populations, and 197 individuals from 4 Caribbean populations), as well as individuals from continental West Africa (45 Yoruba-speaking individuals from Ibadan, Nigeria and 28 individuals from Gabon). These whole-genome sequences were made publicly available through dbGAP (accession code phs001123.v1.p1) and were incorporated into the reference panel on the Michigan imputation server (a free genotype imputation service, https://imputationserver.sph.umich.edu). The average sequencing depth of the CAAPA samples is $\approx$30x across the genome compared to the average depth of $\approx$7x in the 1000 Genomes Project (TGP) [6]. The CAAPA samples also represent 880 unique African genomes compared to 661 African TGP phase 3 genomes, thus making the CAAPA reference panel a valuable resource for imputing genetic variants of African origin.

Ongoing work in the CAAPA consortium has included coverage analysis of the novel variation identified in CAAPA sequencing. We found that only 69% of common SNP variants and 41% of low-frequency SNP variants identified by CAAPA can be tagged by traditional GWAS arrays at $r^2 >= .8$. In addition, lack of coverage of low frequency variants in GWAS arrays negatively impacts the imputation of low frequency variants. To address these issues, we used the CAAPA sequence data to develop the African Diaspora Power Chip (ADPC), a gene-centric SNP genotyping array designed to complement commercially available genome-wide chips, thereby improving tagging and coverage of African ancestry genetic variation [7]. The array included ~495,000 SNPs, with a minor allele frequency (MAF) that is enriched for low frequency variants (MAF between 0.01-0.05).

Using the ADPC, we genotyped CAAPA participants from nine sites (seven providing African American samples, one providing African Caribbean samples from Barbados and one providing samples from Puerto Rico), and then combined ADPC data with existing commercial genome-wide genotype data and imputed using the CAAPA reference panel. One additional African American dataset genotyped on the Affymetrix Axiom AFR [8] array was also imputed using the CAAPA reference panel.

We used these genotyped and imputed data to perform the largest GWAS of asthma in individuals of African ancestry to date (4,827 asthmatic cases and 5,397 controls).

METHODS

Participants and Disease Definitions

CAAPA investigators previously recruited African ancestry participants into asthma genetics studies across nine geographical sites (Table 1), most of which have been described elsewhere [9]. Participants in these studies were unrelated cases and controls, except for subjects from the Barbados Asthma Genetics Study (BAGS), where a family-based study design was used, and the Howard University Family Study (HUFS) where participants were a mixture of unrelated and related subjects. Definition of asthma was based on a doctor's diagnosis and/or standardized questionnaires (see Supplementary Material for details); controls were defined as a reported negative history of asthma. The distributions of age, sex and age of asthma onset are summarized in Table S1.

Study Oversight

NIH guidelines for conducting human genetic research were followed. The Institutional Review Boards (IRB) of Johns Hopkins University (for GRAAD and BAGS), Howard University (for CRAD and HUFS), Wake Forest University (for SARP), the University of California, San Francisco (coordinating center for the SAGE II and GALA II studies), the Western Institutional Review Board for the recruitment in Puerto Rico (GALA II Puerto Ricans), Children's Hospital and Research Center Oakland and Kaiser Permanente-Vallejo Medical Center (for SAGE II), the University of Chicago (for CAG), University of the West Indies, Cave Hill Campus, Barbados (for BAGS), University of Mississippi Medical Center (for JHS), Henry Ford Health System (for SAPPHIRE), all reviewed and approved this study. All participants provided written informed consent.

Genotyping

ADPC

Prior to genotyping on the ADPC, CAAPA samples were separately genotyped on a variety of GWAS platforms (Table 1). Each of the study sites was responsible for and performed their own genotype calling (see details in the Supplementary Material). To improve coverage of genetic variation specific to African ancestry populations, CAAPA subjects with existing GWAS data were genotyped on the ADPC at Illumina, Inc. FastTrack Services [7]. Genotypes from the ADPC were called following recommendations from Illumina Infinium Technical Notes implemented in Illumina's GenomeStudio software (v2011.1) [10].

Quality Control, Data Merging and Imputation

Initial sample and SNP quality filtering of the CAAPA GWAS data sets, including genomic coordinate transformation to hg19, was performed separately by each of the sites (see the Supplementary Material). A custom pipeline was then used to perform additional QC and to merge each of these GWAS data sets with corresponding ADPC genotype data (described in detail in the Supplementary Material). Finally, these combined data sets were used for imputation, a process where genotype posterior probabilities of untyped genetic markers are estimated by matching a set of phased typed markers to reference haplotype panels. SNP-based imputations were done on the Michigan imputation server

(https://imputationserver.sph.umich.edu), separately for each dataset (see the Supplementary Material for more information). Imputing untyped markers in datasets initially typed on a variety of platforms allowed us to combine results for meta-analysis over all datasets. Imputed genetic data sets also provide a higher density of markers and a much larger number of low frequency and rare variants [11, 12].

Genome-wide and local ancestry estimation

The populations included in the asthma association analyses are genetically admixed populations, deriving their origins from populations that were previously demographically isolated (African, European and Native American populations). To account for and leverage the population structure resulting from this admixture, we estimated both genome-wide and local ancestry for all CAAPA subjects.

To identify population outliers, a set of MAF and linkage-disequilibrium (LD) pruned SNPs genotyped across all CAAPA subjects was merged with an overlapping set of markers from unrelated European and African subjects from TGP phase 3 (84+84 CEU and YRI subjects) and unrelated Native American subjects (43 NAT) from Mao et al. [13]. This merged dataset was used to estimate principal components as well as proportions of African, European and Native American ancestry for each CAAPA subject. Principal component analysis was also performed separately by dataset to account for differences in population structure by case-control status. More detailed information can be found in the Supplementary Material (under the heading **Genome-wide ancestry analysis**).

When admixture occurs between different population groups, recombination events result in chromosomes that are a mosaic of blocks of ancestry derived from the different source populations (called local ancestry segments). Local ancestry inference is a statistical technique that use the probability of recombination events to determine the boundaries of each local ancestry segment, and differences in allele and haplotype frequencies between putative source populations to classify the source ancestry of all segments. The local ancestry of chromosomal segments from 7,416 African American and African Caribbean Barbados CAAPA subjects was inferred using the software program RFMix, with African and European TGP phase 3 subjects serving as reference panels [14]. We inferred 15,824 segments of local ancestry across the genomes of these 7,416 subjects. A more detailed description of this process can be found in the Supplementary Material (see **Local ancestry inference**).

Functional annotation

Several steps were used to explore the potential functional relevance of associations observed: 1) The Genotype-Tissue Expression (GTEx) eQTL browser was used to determine if a SNP is an eQTL; 2) RegulomeDB was used to determine if a SNP may have a regulatory effect; and 3) ANNOVAR gene, filter and region based annotations were examined (Supplementary Table S2).

Statistical Analysis

Association analysis

We first tested for association between imputed allele dosages and case-control status. The imputed allele dosage is defined as the expected number of copies of the variant allele each individual carries, and is a continuous number ranging between 0 and 2 (incorporating the uncertainty of the imputed genotypes). Due to different study objectives of participating studies, and differences in potentially relevant factors such as age of onset and diagnostic criteria (see the Supplementary Material), we

performed tests for association separately for each dataset, and then combined the results in a meta-analysis, allowing for the assessment of association signals that may be particular to any given type of study design. Appropriate statistical tests were used for each type of dataset (see Supplementary Table S3 and the Supplementary Material; due to different effect size statistics, association test results were combined using the weighed Z-score method [15]). P-values $< 5 \times 10^{-8}$ were considered statistically significant.

Candidate gene analysis

In addition to the GWAS approach, we also explored associations with known asthma candidate genes in the meta-analysis results. Due to the relatively lower African contribution in the Puerto Rican (GALA II) study, this pipeline was applied twice: Once to the meta-analysis of all CAAPA datasets, and once to a meta-analysis excluding GALA II.

Asthma associations with $p < 10^{-5}$ were selected from the GWAS catalog and Phenotype-Genotype Integrator (PheGenI) databases (http://www.genome.gov/gwastudies/ and http://www.ncbi.nlm.nih.gov/gap/phegenily, respectively). Gene regions were defined as 1) ±1MB from a gene for variants that fell within a single gene, 2) ±1MB from the variant and gene for variants that did not fall within the gene itself but for which the databases reported a gene, 3) ±1MB from a variant for variants with no reported gene, and 4) ±1MB from the first and last gene for SNPs falling within a region with many reported genes. For each identified gene region, imputed data from the SAGE II dataset (the largest African American dataset with unrelated subjects) was used to estimate the number of independent SNPs in each gene region (see the Supplementary Material for more information). The p-values of all SNPs within a candidate gene region were multiplied by the number of independent tests estimated for that region to give Bonferroni corrected p-values. Corrected p-values < 0.05 were considered statistically significant.

Admixture mapping

Admixture mapping identifies putative disease causing variants that occur more frequently on local ancestry segments inherited from a particular source population. We performed admixture mapping by using linear mixed models (implemented in the EMMAX software) to test for association between the number of copies of African ancestry at a local ancestry segment (0/1/2) and asthma case-control status. We fitted 15,824 models (one for each local ancestry segment) to the 6,686 African American and African Caribbean Barbados subjects for whom local ancestry was inferred and where both cases and controls were available. The models used a covariance matrix to account for relatedness between all individuals, as well as population structure, which accounts for global ancestry. The genotype dataset (e.g. GRAAD(1), GRAAD(2), CAG, etc.) was included as a fixed effect in these models. To correct for multiple testing, we estimated the number of effective tests using the method described by Gao et al. [16], and then applied a Bonferroni corrected significance threshold of $0.05/(262$ effective tests$)=1.9 \times 10^{-4}$. See the Supplementary Material for a more detailed description of this analysis (heading **Admixture mapping**).

Stratifying association tests by local ancestry

A logistic regression interaction model was used to determine if an association differed by local ancestry (where case-control status ~ allelic_dosage + dataset + allelic_dosage*nr_copies_African_ancestry, with nr_copies_African_ancestry encoded as factor variable with values 0, 1 or 2, and the *glm()* function in the R package is used with first the binomial family link and then a logit link). SNPs yielding p-values

$<10^{-5}$ in either of the meta-analysis (i.e. including or excluding GALA II), or in the candidate gene analysis, were tested using this model. Kinship matrices were used to exclude closely related subjects from these association tests (i.e. those having kinship coefficient > 0.25 with one or more of the other subjects in the dataset). This analysis was restricted to only those datasets contributing to the meta-analysis (e.g. if a SNP was filtered out for a particular dataset due to low imputation quality, that dataset was excluded from the local ancestry interaction model). Models with interaction p-values < 0.05 (using the anova.glm() function in R, test=Chisq) were further examined by stratifying the dataset according to 0, 1 or 2 copies of African ancestry at that locus, and then using an additive allelic test for association within each stratum, testing for differences between case and control allele count.

RESULTS

Population Structure

Population structure was assessed by principal component analysis (PCA), as well as genome-wide estimates of proportions of ancestry deriving from specified reference populations (ADMIXTURE [17]). These results are summarized in Supplementary Figures S1 and S2. Three African American subjects were identified as PCA outliers, and were excluded from further association analysis. The median percentages of African ancestry in African American, Barbados and Puerto Rican subjects were 82.2, 90.9 and 18.14% respectively.

Association Analysis

The individual dataset and meta-analysis QQ plots showed little evidence of systematic test statistic inflation (Supplementary Figure S3 and Figure 1). Table 2 summarizes association results from two meta-analysis: a meta-analysis of all the CAAPA GWAS datasets, and a meta-analysis excluding Puerto Rican (GALA II) subjects, due to their relatively higher European ancestry. SNPs are shown if they have $p < 10^{-6}$ in either of the meta-analyses and/or were within an asthma candidate gene region and passed a multiple testing significance threshold for that gene region in either of the two meta-analyses. A detailed breakdown of the association results by dataset is shown in Supplementary Table S4, and the results of the candidate gene analysis are summarized in Supplementary Table S5. SNPs with MAF < 0.01 and with fewer than 8 datasets contributing to the meta-analysis are excluded from Table 2, but are included in Supplementary Table S4.

Two SNPs on chromosome 2q22.3, rs787160 and rs17834780, passed the genome-wide significance threshold of $p < 5 \times 10^{-8}$. These SNPs are intergenic and located close to the Glycosyltransferase Like Domain Containing 1 (*GTDC1*) and Rho GTPase Activating Protein 15 (*ARHGAP15*) genes (Supplementary Figure S4). *ARHGAP15* is expressed in lymph node cells, peripheral blood mononuclear cells, CD8 T-cells, and B-lymphocytes, while *GTDC1* is expressed in lung and peripheral blood leukocytes (genecards). Glycosyltransferase plays a role in the adhesion of environmental factors to epithelial cells, which could in turn play some role in risk to asthma [18]. SNP rs787160 falls in a region of weak transcription in fetal lung tissue (15 state model), and rs17834780 falls in a heterochromatin region in lung tissue (18 state model) in the Encyclopedia of DNA Elements (ENCODE) database.

Association between chromosome 17q21 and asthma are regarded as the most consistent finding in asthma genetics research to date [19]. Although SNPs in this region do not pass our genome-wide significant threshold in the meta-analysis results, two chromosome 17q21 SNPs, rs3809717 and rs12150079 (hg 19 base pair positions 37886986 and 38025417 respectively) were identified in the

candidate gene analysis. SNP rs3809717 is upstream from the Migration and Invasion Enhancer 1 (*MIEN1*) gene and SNP rs12150079 is intronic to Zona Pellucida Binding Protein 2 (*ZPBP2*) gene. Interestingly, rs12150079 is annotated as a strong enhancer in the ENCODE B-lymphocyte and lymphoblastoid cell line (ENCODE 15 state model GM12878 cell line), and rs3809717 is annotated as either a strong or weak promoter in multiple ENCODE cell lines (ENCODE 15 state model, active promoter in the GM127878, HepG2, HSMM, HUVEC, NHEK and NHLF cell lines, weak promoter in the H1-hESC, HMEC and K562 cell lines).

A number of genes on chromosome 1q21.3 are putative asthma genes (*CRNN, LCE5A, CRCT1, LCE3E, IL6R, ADAR* and *KCNN3*). This region falls within the epidermal differentiation complex, which is also implicated in the development of atopic dermatitis [20]. Four SNPs falling within this region reached statistical significance in the candidate gene analysis, are in strong LD, and are within introns of the *INTS3* gene. These SNPs have a low MAF in the CAAPA datasets (MAF ~ 0.01, see Supplementary Table 4 for a more detailed breakdown), and are rare in European populations (MAF=0.001 in 1006 TGP phase 3 subjects). According to ENCODE, rs114469574 falls within enhancer regions in B-lymphocytes and lymphoblastoid cell lines, as well as the hepatocellular carcinoma cell line (15 state model, cell lines GM12787 and HepG2).

SNP rs11264909 on chromosome 1q23.1 reached significance in the candidate gene analysis that excluded GALA II. This SNP is located approximately 900 kb downstream from an association near the Pyrin and HIN Domain-Containing Protein 1 (*PYHIN1*) gene, which was reported as an African-ancestry specific association by the EVE asthma meta-analysis (rs1102000 at position 158,932,907 in the EVE meta-analysis, and rs1101999 at position 158,932,555 in the EVE replication; these 2 variants are in almost complete linkage disequilibrium in African Americans) [5]. SNP rs11264909 is intronic to the Kin of IRRE Like (Drosophila) (*KIRREL*) gene. Gene ontology (GO) annotations of *KIRREL* include myosin binding, which could affect bronchospasm [21]. ENCODE annotations for rs11264909 suggest this variant may affect transcription (the variant is annotated as having weak transcription in the H1-hESC cell line, transcription transition in HMEC and HSMM cell lines, insulator in the HUVEC cell line, and transcription elongation in NHEK and NHLF cell lines). The EVE SNPs rs1102000 and rs1101999 had p=2.58×10^{-3} and p=2.58×10^{-3} in the CAAPA meta-analysis when excluding GALA II, which is comparable to the replication p=2.80×10^{-3} reported by EVE for rs1101999 in African ancestry populations. The MAF of EVE SNPs rs1102000 and rs1101999 in the CAAPA datasets (excluding GALA II) ranged between 0.25 and 0.31, which is comparable to the MAF range of 0.26-0.29 reported by EVE. We note that the imputation quality metrics of the EVE SNPs rs1102000 and rs1101999 SNPs in the CAAPA datasets (Rsq ranging between 0.71-1.00, Supplementary Table 4) were on average slightly lower compared to the CAAPA SNP rs11264909 (Rsq ranging between 0.83-0.97, Supplementary Table 4). In addition, the majority of the CAAPA datasets did not have nominally significant p-values (i.e. < 0.05) in the individual datasets for these chromosome 1q23.1 SNPs (EVE SNPs rs1102000 and rs1101999 and CAAPA SNP rs11264909). For the CAAP SNP rs11264909, the direction of effect was consistent for 9 of the 10 datasets (with the smallest CAAPA dataset, GRAAD(2), having an opposite effect), contributing to an overall meta-analysis p=3.98×10^{-6}. In contrast, for the EVE rs1102000 and rs1101999 SNPs, 3 datasets yielded conflicting effect directions (the JHS, GRAAD(2) and SARP datasets, see Supplementary Table S4).

Admixture mapping

The distribution of p-values from the admixture mapping tests for association is summarized in Supplementary Figure S5. The QQ plot shows little evidence for systematic test statistic inflation. The deflated inflation factor of 0.90 appears to be largely driven by 0.1<p<0.5, whereas the number of p <

0.02 is greater than expected, suggesting that the association results are enriched with local ancestry segments showing differences in ancestry between cases and controls.

Only one segment of local ancestry, ranging from base pair positions 134,149,974-134,300,365 on chromosome 6, crossed the multiple testing threshold of $p < 1.9 \times 10^{-4}$. This segment falls within a local ancestry peak on chromosome 6q22.31-23.2, with increased African ancestry associated with increased risk of asthma (Supplementary Table S6). Genes falling in this segment include the Transcription Factor 21 (*TCF21*) and TATA-Box Binding Protein Like 1 (*TBPL1*) genes. The *TCF21* gene has been shown to be differentially expressed in bronchial biopsies of asthmatics compared to controls [22].

The second largest admixture-mapping peak occurred on chromosome 4q27-28.1. On average, asthma cases had higher African ancestry compared to controls in the local ancestry segments within this peak (Supplementary Table S6). The local ancestry segment with the smallest p-value ranged from base pair positions 124,387,497-124,567,967 and did not include any known genes. The closest gene to the segment is the Sprouty RTK Signaling Antagonist 1 (*SPRY1*) gene, which may negatively regulate respiratory organogenesis (genecards).

The admixture-mapping peak observed on chromosome 1q23.1 included both the *KIRREL* and *PYHIN1* genes (Supplementary Table S6), identified by the CAAPA and the EVE asthma meta-analysis, respectively. The effect of African ancestry is protective for all the local ancestry segments within this peak. Cases had on average 79.8% African ancestry and controls had on average 81.9% African ancestry in the segment with the smallest p-value.

Stratifying association tests by local ancestry

SNPs with $p < 10^{-5}$ in either of the meta-analyses (including vs. excluding GALA II) , and SNPs associated with asthma in the candidate gene analysis, were tested for interaction with number of copies of African ancestry. Whereas admixture mapping identifies local ancestry segments that differ in ancestry between cases and controls, this approach stratifies the data by number of copies of African ancestry that an individual carries at each tested SNP. The interaction p-value indicates whether a difference in allele frequency between cases and controls is observed in one class of local ancestry background more than in another. Supplementary Table S7 summarizes these results; only two interactions reached significance, one specific to African ancestry, and one specific to European ancestry. The African specific interaction involved SNP rs11264909 on chromosome 1q23.1. A difference in allele frequency was observed only in those subjects carrying 2 copies of African ancestry in this region, with the minor A allele being observed more frequency in cases (0.12) compared to controls (0.10). For SNP rs10246878 on chromosome 7, allele frequency differences were only observed in carriers who had at least 1 copy of European ancestry.

DISCUSSION

Genome-wide association

Only two associations reached genome-wide significance in our study. The associations are found on chromosome 2q22.3, and to our knowledge, an association at this vicinity have only been reported once, in a GWAS of disocyanate asthma which included 74 cases and 824 controls of European ancestry [23]. Notably, the most significant SNP in our study, rs787160, is located one base pair from an insertion-deletion polymorphism. For this reason, the SNP was not included in the TPG phase 3 reference panel, which may explain why rs787160 has not been identified by any prior GWAS. The variant calling

quality metrics of rs787160 in the CAAPA reference genomes did not indicate genotyping error (mean sequencing depth of 39.07 and mean genotyping quality of 89.54; see Supplementary Figure 6 for the distribution of these statistics). In addition, the imputation quality metrics in the CAAPA GWAS datasets were high (Rsq > 0.92 in all datasets, Supplementary Table S4). To obtain more evidence of the accuracy of this result, we genotyped the 3 most significant SNPs on chromosome 2q22.3 in HUFS samples using a TaqMan Real-Time PCR assay. In 1,792 subjects with at least one SNP to compare, the data was 98.76% concordant (Supplementary Table S8), which further substantiates that the SNPs were imputed with high accuracy. It is potentially interesting that SNP rs787160 was not identified by our stratified local ancestry models as having any African-specific effect and that the minor allele has a higher frequency in the GALA II dataset (i.e., especially among subjects having on average higher European ancestry, Supplementary Table S4). These observations suggest that the association is not specific to individuals with African ancestry. Although the associations with markers on chromosome 2q22.3 still need to be replicated in an independent sample set, our analysis underscores the importance of retaining SNPs adjacent to structural variants in reference panels, should they have variant call quality metrics of high quality.

Candidate gene-analyses

The chromosome 17q21 association with asthma is evident in the meta-analysis including GALA II, but less evident when excluding GALA II (Supplementary Figure S4). This region has primarily been associated with childhood onset asthma [19, 24, 25]. In addition, we note that the two most significant associations with markers on chromosome 17q21 in the meta-analysis that included GALA II had a much higher MAF among Puerto Rican subjects (MAF 0.22 and 0.23 for rs3809717 and rs12150079, respectively) compared to subjects from the other CAAPA datasets (where MAF ranged between 0.05-0.08 for both rs3809717 and rs12150079, with the same minor alleles as in GALA II). Our models stratifying association tests by estimated local ancestry also failed to identify chromosome 17q21 SNPs as showing ancestry specific effects. The inclusion of adult onset asthmatics and reduced statistical power due to a lower MAF for SNPs in this region may explain the failure to achieve genome-wide significance for SNPs on chromosome 17q21 in our study.

Our candidate gene analyses identified a number of low frequency variants on chromosome 1q21.3 among African ancestry subjects that are virtually monomorphic among European ancestry populations. This region falls within the epidermal differentiation complex (EDC). The EDC region contains a very large and diverse family of genes associated with skin barrier dysfunction, and has been implicated in the development of atopic dermatitis (AD) [20]. Risk factors for AD are likely also involved in the pathogenesis of asthma, through the so-called "atopic march", which suggests that allergic sensitization occurs first through a damaged skin barrier, and in turn affects the development of a person's immune response and risk for asthma [26].

We identified a putative African-ancestry specific associated SNP, rs11264909, on chromosome 1q23.1. This variant is ~0.9MB downstream from 2 SNPs identified by the EVE asthma meta-analysis. Although the EVE SNPs were not polymorphic in Europeans, the A allele of the CAAPA rs11264909 SNP had a frequency of 0.20 in the CAAPA study populations (excluding GALA II) and a much higher frequency of 0.61 in Europeans (1006 TGP phase 3 subjects). The A allele was observed more frequently in CAAPA cases (frequency of 0.22, excluding GALA II) compared to CAAPA controls (frequency of 0.19, excluding GALA II), suggesting that cases on average have more European ancestry at this locus compared to controls. This was borne out by the admixture mapping results, which pointed to a protective effect of African ancestry in this region. These observations imply that the association observed with asthma is not "due to" African ancestry. However, when stratifying the association tests

by the number of copies of African ancestry carried by each subject, only individuals homozygous for African ancestry at rs11264909 showed differences in allele frequency between cases and controls, with the A allele being observed more frequently among cases. This region may therefore indeed have an African ancestry specific effect, where the combination of African ancestry and the A allele at this locus increase the risk of asthma.

Admixture mapping

The admixture mapping association analysis identified two regions, chromosomes 4q27-28.1 and 6q22.31-23.2, that were not immediately evident from the SNP association results. This corroborated the value of admixture mapping as a complementary approach to association mapping in admixed populations, which may uncover associations not detectable by SNP association tests alone [27]. The chromosome 6 local ancestry segment associated with asthma (after correcting for multiple tests) harbors the *TCF21* gene which is differentially expressed in the bronchial tissues of asthmatics and non-asthmatics [22]. Because no SAPPHIRE controls were typed on the ADPC, asthmatic cases from SAPPHIRE were not included in the admixture mapping tests for association. Consistent with the admixture mapping association results, the SAPPHIRE cases also have excess African ancestry in this local ancestry segment ($>2\pm$ SD from the mean global African ancestry in SAPPHIRE cases), serving as independent verification that increased African ancestry may well increase risk for developing asthma. Furthermore, the SNP with the most significant association in this region, SNP rs111966851 (Supplementary Figure S4), has a MAF of 0.308 and 0.006 in Africans and Europeans respectively (based on 1322 African and 1006 European TGP phase 3 subjects), which also implies that African ancestry in this local ancestry segment increases risk for asthma.

Functional relevance

We used a number of publicly available databases and resources, including GTEx [28], RegulomeDB [29] and ENCODE [30], to explore the potential functional relevance of the associations we identified in our study. We found no functional annotations with immediate or clear relevance to our peak SNPs. While these public resources have had tremendous value in extending our understanding of functional domains within the human genome, we acknowledge the following limitations that may have led to the lack of additional supporting findings for our peak SNPs: the lack of suitable target tissue of relevance in asthma (e.g. lack of peripheral blood mononuclear cells or nasal epithelium in public catalogs), the lack of the inclusion of diseased individuals in the public catalogs (i.e. none of the catalogs include asthmatics where the relevant regulatory signature may indeed be different) and perhaps most importantly the minimal representation of individuals with African ancestry in these public catalogs (i.e. all the added variation represented in our CAAPA genomes and imputed into the African ancestry case control subjects used in our approach is under-represented in the public catalogs and therefore potentially missed). Therefore, while our initial exploration leveraging these public catalogs yield no concrete evidence of functional relevance, we believe that the expansion of these catalogs over time will address some of our concerns.

CONCLUSION

The prevalence, morbidity, and mortality related to asthma in ethnic groups with significant African ancestry, specifically Puerto Ricans and African Americans, is high compared to individuals of European descent and other Hispanic groups such as Mexican Americans. Thus, it is critical to understand the genetic architecture of asthma in these most afflicted subgroups of individuals with asthma. Because of recent admixture events over the past 500 years, these African descent populations

have a complex genetic architecture. We accounted for this complex architecture through the development of our genotyping platforms, based on whole-genome sequencing of individuals of African descent with an overrepresentation of asthma. Our GWAS of >10,000 subjects of African ancestry represent among the largest studies to date, and demonstrates the complexity of asthma susceptibility genes in these recently admixed populations.

Our study has yielded SNPs and regions that may be specific to risk for asthma in individuals of African ancestry. In addition, our study provides evidence for a novel association on chromosome 2q22.3 that may be relevant in other ethnicities, which may have been undetected by previous asthma association studies due to poor representation of the associated variants in traditional imputation reference panels and genotyping arrays. The associations we report here warrant further investigation and replication. Further evaluation of the genes we identified in well-characterized asthma cohorts, already well represented within CAAPA, will provide additional insight into the role of these loci in determining asthma severity and progression. Another logical next step is the use of multi-omic approaches to generate hypotheses regarding their functional relevance, that is only limited now by the lack of multi-omic data sets that are representative of African ancestry.

FIGURES

Figure 1: Summary of the distribution of meta-analysis p-values. QQ and Manhattan plots of the meta-analysis p-values, stratified by minor allele frequency (MAF) for low frequency and common SNPs. Inflation factors were calculated by transforming p-values to 1 degree of freedom (df) Chi-square statistics, and dividing the median of these statistics by the median of the theoretical Chi-square (1 df) distribution. The blue and red horizontal lines in the Manhattan plots represent the suggestive and significant p-value thresholds, respectively. Panel a) shows the QQ and Manhattan plots for the meta-analysis of all CAAPA datasets. Panel b) shows the QQ and Manhattan plots for the meta-analysis excluding GALA II.
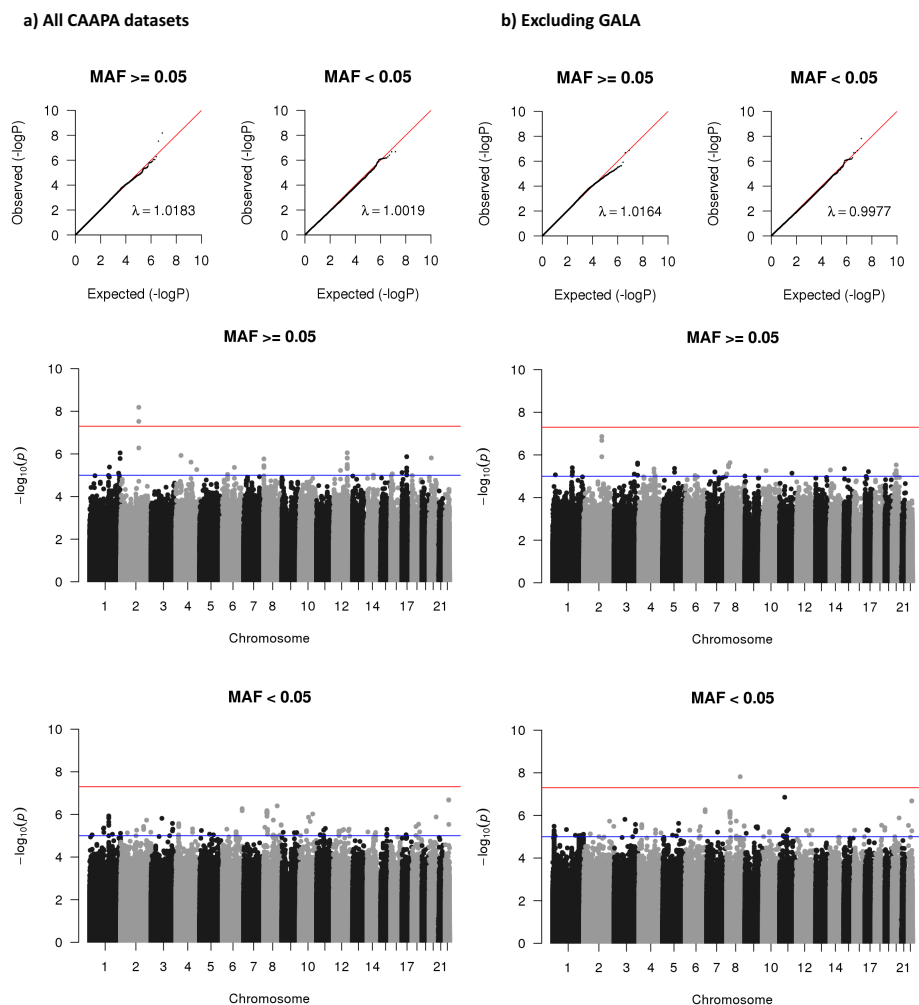
TABLES

Table 1: Studies included in the CAAPA meta-analysis. Studies were genotyped on a variety of GWAS backbones, and coverage of African genetic variation was augmented by additional genotyping on the African Diaspora Power Chip (ADPC). The table summarizes the final number of subjects included in the analysis per study, the number of genotyped SNPs in the combined GWAS backbone and ADPC array data sets, and the number of imputed SNPs that were used in the meta-analysis.

Population	Location	Study	GWAS platform	Total	Asthmatics	Non-asthmatics	GWAS+ ADPC	Tested
African American	Baltimore	GRAAD	Illumina HumanHap 650Y	781	396	385	1,000,719	17,190,019
	Baltimore	GRAAD	Illumina OMNI 2.5	85	64	21	2,253,609	9,056,401
	Chicago	CAG	Illumina HumanHap1M	270	114	156	1,319,903	13,643,921
	Detroit	SAPPHIRE	Affymetrix Axiom AFR	1,891	1,325	566	754,238*	18,657,065
	Jackson	JHS	Affymetrix 6.0	326	44	282	1,047,756	13,525,786
	Jackson	JHS	Affymetrix 6.0	632	101	531	1,115,933	13,806,893
	San Francisco	SAGE II	Affymetrix Axiom LAT	1,692	1,001	691	1,085,595	19,274,579
	Washington	HUFS	Affymetrix 6.0	1,831	303	1,527	1,096,292	19,271,549
	Winston-Salem	SARP	Illumina HumanOmniExpress + HumanHap1M	341	296	45	852,391	13,441,970
	Total			7,849	3,644	4,205		
Barbados	Barbados	BAGS	Illumina HumanHap 650Y	684	282	338	974,505	14,308,079
Puerto Rico	Puerto Rico	GALA II	Affymetrix Axiom LAT	1,754	901	853	1,037,661	16,174,844

* ADPC data available for 730 cases and 0 controls; tests for association were therefore done using genotype data imputed from the Affymetrix data only

Table 2: Summary of meta-analysis results. This table summarizes association results based on two meta-analysis: a meta-analysis of all the CAAPA GWAS datasets, and a meta-analysis that excludes GALA II, due to the relatively higher European ancestry within GALA II. The table summarizes association results that pass any of the following criteria: 1) SNPs with p-values < 10-6 in either of the meta-analyses that include or exclude GALA II, and 2) SNPs that fall within an asthma candidate gene region and that pass a multiple testing significance threshold for that gene region in either of the two meta-analyses. SNPs with MAF < 0.01 and with less than 8 studies contributing to the meta-analysis p-value are excluded from the table.

rsID	chromosome: hg19 position	Major/ minor allele	CAAPA MAF	Nearest gene	Region	Incl. GALA (Y/N)	P-value	Case MAF/ Control MAF
rs77941353[1]	1:153688130	C/T	0.012	GEMIN2P1	intergenic	Y	3.06×10^{-6}	0.010/0.013=0.77
rs114469574[1]	1:153704977	A/G	0.012	INTS3	intron	Y	1.50×10^{-6}	0.010/0.013=0.77
[1]	1:153714954	C/G	0.013	INTS3	intron	Y	1.36×10^{-6}	0.011/0.015=0.73
rs114026160[1]	1:153727808	C/T	0.012	INTS3	intron	Y	1.18×10^{-6}	0.010/0.013=0.77
rs11264909 [1]	1:158056779	T/A	0.204	KIRREL*	intron	N	3.98×10^{-6}	0.216/0.194=1.11
rs10927221[2]	1:244532963	T/C	0.232	C1orf100	intron	Y	8.91×10^{-7}	0.250/0.216=1.16
rs17834780[2]	2:144638809	G/A	0.129	GTDC1	intergenic	Y	2.97×10^{-8}	0.142/0.188=0.76
rs787160[2]	2:144650140	T/A	0.153	GTDC1	intergenic	Y	6.51×10^{-9}	0.169/0.139=1.22
rs787151[2]	2:144655836	G/A	0.240	GTDC1	intergenic	Y	5.26×10^{-7}	0.253/0.227=1.11
rs62365003[1]	5:94736141	C/T	0.096	FAM81B	intron	N	4.28×10^{-6}	0.106/0.087=1.22
rs28601685[1]	8:27338294	C/T	0.212	CHRNA2	upstream-gene	N	2.31×10^{-6}	0.202/0.221=0.91
rs5008317[1]	9:89759848	A/G	0.023	ZCCHC6	upstream-gene	N	3.46×10^{-6}	0.019/0.025=0.76
rs74532409[2]	10:105730220	G/A	0.029	SLK	intron	Y	9.52×10^{-7}	0.034/0.025=1.36
rs9300316[2]	12:111196204	T/C	0.160	PPP1CC	intergenic	Y	8.83×10^{-7}	0.148/0.171=0.87
rs62066359[1]	17:26241077	A/C	0.032	LYRM9	intergenic	N	5.09×10^{-6}	0.039/0.025=1.56
rs3809717[1]	17:37886986	C/A	0.096	MIEN1	upstream-gene	Y	4.54×10^{-6}	0.088/0.104=0.85
rs12150079[1]	17:38025417	G/A	0.102	ZPBP2	intron	Y	1.35×10^{-6}	0.092/0.111=0.83

ACKNOWLEDGEMENTS

AUTHOR INFORMATION

Department of Medicine, University of Colorado Denver, Aurora, CO, 80045, USA
Kathleen C. Barnes, Meher Preethi Boorgula, Monica Campbell, Sameer Chavan, Michelle Daya, Christopher R. Gignoux, Ethan M. Lange, Leslie A. Lange, Nicholas Rafaels, Aniket Shetty

Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD 21205, USA
Terri H. Beaty, Rasika Ann Mathias

Department of Medicine, Einstein Medical Center, Philadelphia, PA 19141, USA
Jean G. Ford

Department of Medicine, Johns Hopkins University, Baltimore, MD 21224, USA
Cassandra Foster, Li Gao, Nadia N. Hansel, Rasika Ann Mathias, Alan F. Scott, Candelaria Vergara

Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD 21205, USA
Ingo Ruczinski, Margaret A. Taub

Departments of Biomedical Engineering and Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA
Steven Salzberg

Data and Statistical Sciences, AbbVie, North Chicago, IL 60064, USA
Jingjing Gao

Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA
Yijuan Hu, Henry Richard Johnston, Zhaohui S. Qin

Department of Public Health Sciences, Henry Ford Health System, Detroit, MI 48202, USA
Albert M. Levin

Department of Microbiology, Howard University College of Medicine, Washington, DC 20059, USA
Georgia M. Dunston

National Human Genome Center, Howard University College of Medicine, Washington, DC 20059, USA
Georgia M. Dunston, Mezbah U. Faruque

Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
Eimear E. Kenny

Illumina, Inc., San Diego, CA 92122, USA
Kimberly Gietzen, Linda Gutierrez, Mark Hansen

Pulmonary and Critical Care Medicine, Morehouse School of Medicine, Atlanta, GA 30310, USA
Marilyn G. Foreman

Department of Medicine, Northwestern University, Chicago, IL 60611, USA
Pedro C. Avila, Leslie Grammer, Robert Schleimer

Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA
Kwang-Youn A. Kim

Department of Pediatrics, Northwestern University, Chicago, IL 60611, USA
Rajesh Kumar

Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
Carlos Bustamante, Francisco M. De La Vega, Christopher R. Gignoux, Shaila Musharoff, Suyash S. Shringarpure, Genevieve Wojcik

Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94143, USA
Esteban G. Burchard, Ryan D. Hernandez, Zachary A. Szpiech

Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, USA
Esteban G. Burchard, Celeste Eng, Maria Pino-Yanes, Dara G. Torgerson

Department of Neurology, University of California, San Francisco, San Francisco, CA 94158, USA
Pierre-Antoine Gourraud, Antoine Lizee

Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA
Raul Torres

Department of Medicine, University of Chicago, Chicago, IL 60637, USA
Dan L. Nicolae, Olufunmilayo Olopade, Oluwafemi Oluwole

Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
Carole Ober

Department of Medicine and Center for Global Health, University of Chicago, Chicago, IL 60637, USA
Christopher O. Olopade

Department of Chemical Pathology, University of Ibadan, Ibadan, 900001, Nigeria
Ganiyu Arinola

Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA
Timothy D. O'Connor, Wei Song

Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA
Adolfo Correa, Solomon Musani

Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA
James G. Wilson

Department of Medicine, Vanderbilt University, Nashville, TN 37232, USA
Tina Hartert, Lorraine B. Ware

Department of Medicine, University of Arizona College of Medicine, Tucson, AZ 85724, USA
Eugene Bleecker, Deborah Meyers

Center for Human Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
Victor E. Ortega

Genetics and Epidemiology of Asthma in Barbados, The University of the West Indies, Chronic Disease Research Centre, Jemmots Lane, St. Michael, BB11115, Barbados
Pissamai Maul, Trevor Maul

Faculty of Medical Sciences, The University of the West Indies, Queen Elizabeth Hospital, Bridgetown, St. Michael, BB11000, Barbados
Harold Watson

Immunology Service, Universidade Federal da Bahia, Salvador, BA 401110170, Brazil
Maria Ilma Araujo

Departamento de Biorregulacao, Universidade Federal da Bahia, Brazil
Camila Figueiredo

Laboratório de Patologia Experimental, Centro de Pesquisas Gonçalo Moniz, Salvador, BA 40296-710, Brazil
Ricardo Riccio Oliveira

Institute for Immunological Research, Universidad de Cartagena, Cartagena, 130000, Colombia
Luis Caraballo, Beatriz Martinez, Javier Marrugo, Catherine Meza

Faculty of Medicine, Universidad Nacional Autonoma de Honduras en el Valle de Sula, San Pedro Sula, 21102, Honduras
Gerardo Ayestas, Allan Saenz, Gloria Varela

Facultad de Medicina, Universidad Catolica de Honduras, San Pedro Sula, 21102, Honduras
Said Omar Leiva Erazo, Edwin Francisco Herrera-Paz, Pamela Landaverde-Torres, Rosella Martinez, Luis F. Mayorga, Hector Ramos

Centro de Neumologia y Alergias, San Pedro Sula, 21102, Honduras
Alvaro Mayorga, Delmy-Aracely Mejia-Mejia, Edwin Francisco Herrera-Paz

Faculty of Medicine, Universidad Nacional Autonoma de Honduras en el Valle de Sula, San Pedro Sula, 21102, Honduras
Gerardo Ayestas, Allan Saenz, Gloria Varela

Centro Medico de la Familia, San Pedro Sula, 21102, Honduras
Olga Marina **Vasquez**

Caribbean Institute for Health Research, The University of the West Indies, Kingston, Jamaica
Trevor S. Ferguson, Jennifer Knight-Madden, Rainford J. Wilks

Department of Child Health, The University of the West Indies, Kingston, Jamaica
Maureen Samms-Vaughan

Centre de Recherches Médicales de Lambaréné, 13901, Gabon
Ayola Akim Adegnika, Ulysse Ateba-Ngoa

Department of Parasitology, Leiden University Medical Center, Leiden, 2333, Netherlands
Maria Yazdanbakhsh

Center for Research on Genomics & Global Health, National Institutes of Health, Bethesda, MD 20892, USA
Ayo Doumatey, Charles Rotimi

Department of Internal Medicine, Henry Ford Health System, Detroit, MI 48202, USA
L. Keoki Williams

National Human Genome Center, Howard University College of Medicine, Washington, DC 20059, USA
Georgia M. Dunston, Mezbah U. Faruque

The Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL 60611, USA
Rajesh Kumar

Institute for Human Genetics, University of California, San Francisco, San Francisco, CA 94143, USA
Ryan D. Hernandez

California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, CA 94143, USA
Ryan D. Hernandez

CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, 28029, Spain
Maria Pino-Yanes

Department of Statistics, University of Chicago, Chicago, IL 60637, USA
Dan L. Nicolae

Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA
Timothy D. O'Connor, Wei Song

Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA
Timothy D. O'Connor, Wei Song

Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, TN 37232, USA
Lorraine B. Ware

Institut für Tropenmedizin, Universität Tübingen, 72074, Germany
Ayola Akim Adegnika, Ulysse Ateba-Ngoa

Department of Parasitology, Leiden University Medical Center, Netherlands
Ayola Akim Adegnika, Ulysse Ateba-Ngoa

CONSORTIA

CAAPA

Ayola Akim Adegnika, Ganiyu Arinola, Ulysse Ateba-Ngoa, Gerardo Ayestas, Adolfo Correa, Francisco M. De La Vega, Celeste Eng, Said Omar Leiva Erazo, Marilyn G. Foreman, Cassandra Foster, Li Gao, Jingjing Gao, Kimberly Gietzen, Leslie Grammer, Linda Gutierrez, Mark Hansen, Tina Hartert, Yijuan Hu, Kwang-Youn A. Kim, Pamela Landaverde-Torres, Javier Marrugo, Beatriz Martinez, Rosella Martinez, Luis F. Mayorga, Delmy-Aracely Mejia-Mejia, Catherine Meza, Solomon Musani, Shaila Musharoff, Oluwafemi Oluwole, Maria Pino-Yanes, Hector Ramos, Allan Saenz, Steven Salzberg, Maureen Samms-Vaughan, Robert Schleimer, Alan F. Scott, Suyash S. Shringarpure, Wei Song, Zachary A. Szpiech, Raul Torres, Gloria Varela, Olga Marina Vasquez, Lorraine B. Ware, Maria Yazdanbakhsh.

REFERENCES

1.   Moorman, J.E., et al., *National surveillance of asthma: United States, 2001-2010.* Vital Health Stat 3, 2012(35): p. 1-58.
2.   Moorman, J.E., et al., *National surveillance for asthma--United States, 1980-2004.* MMWR Surveill Summ, 2007. **56**(8): p. 1-54.
3.   Keet, C.A., et al., *Neighborhood poverty, urban residence, race/ethnicity, and asthma: Rethinking the inner-city asthma epidemic.* J Allergy Clin Immunol, 2015. **135**(3): p. 655-62.
4.   Mathias, R.A., et al., *A genome-wide association study on African-ancestry populations for asthma.* J Allergy Clin Immunol, 2010. **125**(2): p. 336-346 e4.
5.   Torgerson, D.G., et al., *Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations.* Nat Genet, 2011. **43**(9): p. 887-92.
6.   Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

7.   Johnston, H.R., et al., *Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome.* bioRxiv, 2017.

8.   Padhukasahasram, B., et al., *Gene-based association identifies SPATA13-AS1 as a pharmacogenomic predictor of inhaled short-acting beta-agonist response in multiple population groups.* Pharmacogenomics J, 2014. **14**(4): p. 365-71.

9.   Mathias, R.A., et al., *A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome.* Nat Commun, 2016. **7**: p. 12522.

10.   Illumina. *Infinium Genotyping Data Analysis*. Available from: http://www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf.

11.   Huang, J., et al., *Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel.* Nat Commun, 2015. **6**: p. 8111.

12.   Zawistowski, M., et al., *Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes.* Am J Hum Genet, 2010. **87**(5): p. 604-17.

13.   Mao, X., et al., *A genomewide admixture mapping panel for Hispanic/Latino populations.* Am J Hum Genet, 2007. **80**(6): p. 1171-8.

14.   Maples, B.K., et al., *RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference.* Am J Hum Genet, 2013. **93**(2): p. 278-88.

15.   Begum, F., et al., *Comprehensive literature review and statistical considerations for GWAS meta-analysis.* Nucleic Acids Res, 2012. **40**(9): p. 3777-84.

16.   Gao, X., J. Starmer, and E.R. Martin, *A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms.* Genet Epidemiol, 2008. **32**(4): p. 361-9.

17.   Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome Res, 2009. **19**(9): p. 1655-64.

18.   Kauffmann, F., et al., *Associations of blood group-related antigens to FEV1, wheezing, and asthma.* Am J Respir Crit Care Med, 1996. **153**(1): p. 76-82.

19.   Raedler, D. and B. Schaub, *Immune mechanisms and development of childhood asthma.* Lancet Respir Med, 2014. **2**(8): p. 647-56.

20.   Barnes, K.C., *An update on the genetics of atopic dermatitis: scratching the surface in 2009.* J Allergy Clin Immunol, 2010. **125**(1): p. 16-29 e1-11; quiz 30-1.

21.   Fredberg, J.J., et al., *Perturbed equilibrium of myosin binding in airway smooth muscle and its implications in bronchospasm.* Am J Respir Crit Care Med, 1999. **159**(3): p. 959-67.

22.   Vaillancourt, V.T., et al., *From expression pattern to genetic association in asthma and asthma-related phenotypes.* BMC Res Notes, 2012. **5**: p. 630.

23.   Yucesoy, B., et al., *Genome-Wide Association Study Identifies Novel Loci Associated With Diisocyanate-Induced Occupational Asthma.* Toxicol Sci, 2015. **146**(1): p. 192-201.

24.   Ono, J.G., T.S. Worgall, and S. Worgall, *17q21 locus and ORMDL3: an increased risk for childhood asthma.* Pediatr Res, 2014. **75**(1-2): p. 165-70.

25.   Moffatt, M.F., et al., *A large-scale, consortium-based genomewide association study of asthma.* N Engl J Med, 2010. **363**(13): p. 1211-21.

26.   Boguniewicz, M. and D.Y. Leung, *Atopic dermatitis: a disease of altered skin barrier and immune dysregulation.* Immunol Rev, 2011. **242**(1): p. 233-46.

27.   Galanter, J.M., et al., *Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study.* J Allergy Clin Immunol, 2014. **134**(2): p. 295-305.

28.   Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.

29.   Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB.* Genome Res, 2012. **22**(9): p. 1790-7.

30.   Consortium, E.P., *A user's guide to the encyclopedia of DNA elements (ENCODE).* PLoS Biol, 2011. **9**(4): p. e1001046.

SUPPLEMENTARY MATERIAL

CAAPA ADPC Participants and Disease Definitions

The primary selection criteria applied by each CAAPA study was (i) self-reported African ancestry; (ii) a physician's diagnosis of asthma or a negative history of asthma symptoms and asthma medication usage among controls; and (iii) the availability of DNA at a concentration of at least 10 ng/μl from a primary blood sample. 10,827 of the selected participants that had genotype data that passed all quality control criteria were used for the CAAPA association analysis (Table 1).

The **Genomic Research on Asthma in the African Diaspora (GRAAD) consortium** represent 8 studies of asthma in African American children and adults, and one additional study on healthy African Americans. GRAAD study subjects were recruited from the Baltimore-Washington DC metropolitan area, and all subjects used in this GWAS self-reported as African American. Case-control status was defined as previously described [1]. Briefly, standardized questionnaires administered by a clinical coordinator were used to determine whether a subject has a history of both self-reported and physician-diagnosed asthma (asthma case), or had no history of asthma (healthy control). Adult controls were favored to minimize the inclusion of subjects that may still develop asthma.

The **Chicago Asthma Genetics (CAG)** study comprises European American and African American families ascertained through affected sib pairs, case-parent trios (through affected offspring), adults and children with severe persistent asthma, and non-asthmatic control subjects (> 18 years) [2]. Asthma cases and families were recruited in the adult and/or pediatric asthma clinics at University of Chicago Hospital; controls were recruited from the medical center at large.

The **Study of Asthma Phenotypes and Pharmacogenomics Interactions by Race-ethnicity (SAPPHIRE)** is a longitudinal, population-based study aiming to identify the genetic predictors of asthma medication in a multi-ethnic patient [3]. Eligible patients received their care from a large healthcare system serving southeast Michigan and the greater Detroit metropolitan area. Patients with asthma were ≥12 years of age, had a documented physician diagnosis of asthma in the medical record, and had no prior diagnosis of congestive heart failure or chronic obstructive pulmonary disease [4, 5]. Control patients had the same enrollment criteria as case patients with the exception of having no prior diagnosis of asthma.

The **Jackson heart study (JHS)** was designed to study risk factors for of cardiovascular disease as well as other chronic disorders, such as asthma, in African American adults [6]. Detailed information on the study design and data collection has been published previously [7-9]. Participants were classified as asthmatic based on self-reported physician diagnosis of asthma or use of asthma medication [6].

The **Study of African Americans, Asthma, Genes & Environments (SAGE II)** is an ongoing population-based, case-control study recruiting African American participants from clinics in the San Francisco Bay Area [10]. Subjects were eligible if they were 8-21 years of age, self-identified all four grandparents as African Americans, and had <10 pack-years of smoking history. Asthma was defined by physician diagnosis and report of symptoms in the 2 years preceding enrollment. Controls had no reported history of asthma, eczema, hives, hay fever, allergic rhinitis, no reported use of medication for allergies and no reported symptoms of wheezing or shortness of breath during their lifetime

The **Howard University Family Study (HUFS)** is a population based family study of African Americans in the Washington metropolitan area. African American families, as well as unrelated individuals, were randomly

ascertained to study the genetic and environmental basis of common complex diseases [11]. Asthma was defined based on self-reported physician diagnosis and medication history [12]. Controls were defined based on absence of physician-diagnosed asthma and no history of asthma medication.

The NHLBI-sponsored **Severe Asthma Research Program** (SARP1-2) is a multi-ethnic cohort of comprehensively characterized subjects with mild to severe asthma enriched for severe disease. SARP1-2 is an integrated study of the clinical and biological features of asthma with the goal of understanding the cellular and molecular mechanisms underlying asthma and its endotypes. Study participants that self-reported as African American were included in the CAAPA GWAS. A diagnosis of asthma in SARP was based on evidence of methacholine bronchial hyperresponsiveness or bronchodilator reversibility in combination with asthma symptoms [13].

Probands from the **Barbados Asthma Genetics Study (BAGS)** were recruited through referrals as described previously [1, 14, 15]. Nuclear and extended family members were also enrolled in the study. Asthma was defined based on a history of both self-reported and physician-diagnosed asthma, as well as a history of wheezing without an upper respiratory tract inflection [1]. Controls were selected based on no history of asthma.

The **Genes-environments & Admixture in Latino Americans (GALA II) is** an ongoing population-based, case-control study recruiting Latino children from five centers (Chicago, Illinois; Bronx, New York; Houston, Texas; San Francisco Bay Area, California; and Puerto Rico) [16]. Subjects were eligible if they were 8-21 years of age and all four grandparents self-identified as Latino. The definition of asthma cases and controls followed the same protocols as SAGE II. Only subjects recruited in Puerto Rico were included in the CAAPA GWAS.

CAAPA WGS Reference Panel Participants and Disease Definitions

The whole genome sequence (WGS) reference panel that was used for genotype imputation is comprised of 880 unrelated subjects. Participant selection and disease definition is summarized in detail in Supplementary Appendix of Mathias et al. [17], but the note only describes the 642 subjects in the data freeze used for that study's analysis. The final data freeze of 880 subjects used for genotype imputation in our association study is summarized in Supplementary Table S9. Samples collected from JHS, Gabon and Palenque were not available in the previous data freeze (642 subjects), but are available in the final data freeze (880 subjects). The JHS is described above (under **CAAPA ADPC Participants and Disease Definitions**). Asthma was not defined for the subjects from Gabon and Palenque.

GWAS backbone genotyping, SNP calling and quality control

CAAPA studies were genotyped on a variety of GWAS backbones (Table 1). Each of the study sites was responsible for and performed their own genotype calling and genomic coordinate transformation where applicable, and for most datasets, this has been described previously (GRAAD(1)+BAGS: [1], CAG: [18], SAPPHIRE: [3], JHS: [19], SAGE II: [20], HUFS: [21], SARP [], GALA II [16]). Exceptions are described below.

GRAAD(2)

Samples from GRAAD participants that were recruited after the initial GWAS study[1] were genotyped on the Illumina Omni2.5 BeadChip at the Lowe Family Genomics core, Johns Hopkins. Genotypes were called following recommendations from Illumina Infinium Technical Notes implemented in Illumina's GenomeStudio software (v2011.1).

SNPs that were out of HWE (p-value < 10^{-6}) and SNPs with a missing call rate > 0.1 were dropped from the dataset. Samples with a missing call rate > 0.01 and samples with gender discrepancies (using F-statistics estimated by PLINK, samples with F(inbreeding males)<0.2 or F(inbreeding females)>0.5 defined as miss-classified) were dropped from the dataset. PLINK was used to estimate IBD between all pairs of subjects, using a set of LD-pruned SNPs. Duplicates (PLINK PI_HAT > 0.9, remove both individuals) and related individuals (PLINK PI_HAT > 0.35, remove at least one individual) were then dropped from the dataset.

JHS

After receiving the JHS datasets from the study coordinators, genomic coordinates were translated using CrossMap (http://crossmap.sourceforge.net) and the hg18tohg19 chain file published with this software. Genotyping on the Affymetrix 6.0 array was performed in two batches, and these two genotype data batches were therefore imputed and analyzed separately.

SARP

Genotypes were called using the GenomeStudio clustering algorithm. DNA samples with call rates <0.90 were excluded. SNPs with call frequencies <0.90 were excluded and clustering was repeated followed by exclusion of samples with a call rate <0.98. After updating SNP statistics, a second round of SNP exclusion was performed for cluster separation scores <0.20 followed by SNPs with score <0.30 and a GenTrain score <0.75. After clustering and genotype calling, QC was performed using PLINK. SNPs with a MAF >0.01 or with HWE p<0.001 in unrelated subjects were removed. Sex was determined for all subjects using chromosome X data and subjects with a mismatch with self-reported sex data were excluded. IBD tests were also performed to exclude identical or related samples.

ADPC genotyping, SNP calling and quality control

10,411 CAAPA samples, 346 HapMap controls and 217 internal samples were genotyped at Illumina on the ADPC. 306 samples failed QC at Illumina. For the remaining 9,888 CAAPA samples, sample quality control was performed in the following order:
1. Remove IBD duplicates (n=193; performed on 113,324 LD-pruned SNPs filtered for SNPs that have Hardy-Weinberg equilibrium (HWE) p-values > 1e-6 and MAF > 0.05; samples with PLINK PI_HAT < 0.9 were classified as IBD duplicates)
2. Remove samples with a high proportion of missing genotypes (n=0, more than 1% missing genotypes)
3. Remove samples with gender discrepancies (n=28, using F-statistics estimated by PLINK, samples with F(inbreeding males)<0.2 or F(inbreeding females)>0.5 were removed)
4. Remove samples with excess heterozygosity (n=2, using F-statistics estimated by PLINK, samples with F(inbreeding)>0.25).
An additional 364 samples without existing GWAS data, and 71 samples from two studies too small for accurate imputation (the imputation server requires at least 50 samples) were also removed, leaving 9230 samples.

Preliminary SNP QC was performed using GenomeStudio. 130,637 SNPs that failed one or more of the following criteria were removed, as recommended by Illumina:
1. Call frequency < 0.97
2. Replicate errors > 2
3. Parent-parent-child (PPC) errors > 2
4. Cluster separation <= 0.3

5. Mean normalized intensity of first homozygote cluster (AA_R_Mean) <= 0.2 and/or mean normalized intensity of heterozygote cluster (AB_R_Mean) <= 0.2 and/or mean normalized intensity of second homozygote cluster (BB_R_Mean) <=0.2
6. Mean of normalized theta value for first homozygote cluster (AA_T_Mean) > 0.3
7. Standard deviation of normalized theta value for first homozygote cluster (AA_D_Mean) > 0.06
8. Mean of normalized theta value for second homozygote cluster (BB_T_Mean) < 0.7
9. Standard deviation of normalized theta value for second homozygote cluster (BB_D_Mean) > 0.06.

Merging and imputation of GWAS and ADPC datasets

After receiving GWAS SNP array data from the various study centers, a pipeline of custom scripts was used to apply consistent quality control (QC) metrics to the GWAS datasets and to merge autosomal content from the GWAS and ADPC datasets. The pipeline was applied separately to each of the GWAS datasets listed in Table 1, and each of the datasets were uploaded separately to the imputation server hosted by the University of Michigan (https://imputationserver.sph.umich.edu; for the imputation configuration, the CAAPA - African American reference panel was used, the configured population was AA (CAAPA), and ShapeIT was used for phasing).

The steps applied in the pipeline were as follows:
1. Non-autosomal SNPs and SNPs with duplicate positions were removed from the GWAS dataset. Samples that were not typed successfully on the ADPC array were also removed. SNPs that failed a test for Hardy-Weinberg equilibrium (HWE; p-value < 0.0001), monomorphic SNPs and SNPs that had more than 5% missing genotypes were then removed from the GWAS dataset.
2. Non-autosomal SNPs and four SNPs with duplicate positions were removed from the ADPC dataset. SNPs that failed a test for Hardy-Weinberg equilibrium (p-value < 0.0001), monomorphic SNPs and SNPs that had more than 5% missing genotypes were removed.
3. AT/CG SNPs were handled as follows, in both the GWAS and ADPC datasets: SNPs not in the reference panel were deleted; SNPs with MAF > 0.4 in either the dataset or reference panel were deleted; SNPs that differ in MAF more than 1% compared to the reference panel were deleted; for the remaining SNPs, if the minor allele was different from the reference panel minor allele, the strand of the SNP was flipped. These thresholds are stringent and a large number of SNPs were deleted (roughly 30 000 SNPs on the ADPC); however this strategy was preferred as incorrect AT/CG SNPs could potentially have a severe impact on imputation quality.
4. Next, the GWAS and ADPC datasets were uploaded to the imputation server for pre-imputation QC (Quality Control Only option). Based on the QC report downloaded from the server, SNPs typed on alternate strands compared to the reference panel were identified, and the alleles of these SNPs were flipped.
5. SNPs with common positions in the GWAS and ADPC datasets were extracted in order to identify discordant samples, and to verify the genotype concordance between the datasets. The proportion of alleles shared identical by-descent (IBD) between all pairs of samples were estimated using PLINK (PI_HAT output column), and GWAS and ADPC samples with identical sample identifiers but sharing less than 0.9 alleles IBD were removed from the GWAS and ADPC datasets (21 samples). The concordance between the remaining GWAS and ADPC samples was then checked (concordance was > 99.7% in all datasets).
6. Next, genotypes for corresponding GWAS and ADPC samples were merged into a single file. Discordant SNPs were handled as follows (same SNP in the ADPC and GWAS datasets, for the same sample, with different genotype calls): if a SNP was discordant in >= 5% of samples, the SNP was deleted from the dataset, else the SNP value was set to missing for the sample(s) it was discordant for.
7. Merged genotype files were uploaded to the imputation server. Imputed datasets and reports were downloaded after the job runs completed.
8. Based on the findings of a study that quantified imputation accuracy in African Americans, post-imputation SNP filtering was done based on minor allele frequency (MAF) and the per-SNP estimation of the squared

correlation between imputed allele dosages and true unknown genotypes (Rsq) [22]. SNPs with MAF ≤ 0.005 were excluded if Rsq ≤ 0.5 and SNPs with MAF > 0.005 were excluded if Rsq ≤ 0.3.

The above process is depicted in Supplementary Figure S7 for the HUFS dataset. The pipeline automatically generated flow diagrams and the diagrams were manually checked for each dataset to verify the correctness of each step in the process. Results for the data merging steps and results scraped from reports downloaded from the imputation server are summarized in Supplementary Table S10.

Imputation pipeline

Creating the reference panel

1. **Subject removal:** Identity-by-descent (IBD) tests were performed on 917 CAAPA samples to test for unexpected relatedness. IBD tests were generated from the multi-sample VCF by KNOME, Inc. Subjects were removed in order to ensure that no duplicate, parent-offspring or sibling relationships are present in the data.
2. **Variant filtering:** Variants with genotyping quality scores (GQ) less than 20, depth (DP) less than 7, and in regions of segmental duplication were removed. Only bi-allelic variants with call rates of at least 95% were retained.
3. **Phasing:** Phasing was performed using ShapeIT (v2.r790) using 0.5 Mb windows.

As part of collaboration with the University of Michigan, this phased reference panel was uploaded to the Michigan imputation server.

Imputation

Full details of the pipeline is documented at https://imputationserver.sph.umich.edu/index.html#!pages/pipeline. Briefly, the steps in the pipeline is as follows:

1. **Create chunks** of size 20 Mb**.**
2. **Do quality checks on chunks.** For chunks, the amount of valid variants, the amount of variants found in the reference panel, and the sample call rate are checked. For variants, valid alleles (A/C/G/T), SNP call rate, allele frequency differences between study and reference panel, and allele switches and strand flips are checked.
3. **Phase genotypes.** 3 phasing algorithms are available: ShapeIt, HapiUR and Eagle. For this study, we used ShapeIt to phase genotypes.
4. **Impute genotypes per chunk** using Minimac3 and a window size of 500 Kb**.**
5. **Merge chunks into a single file, per chromosome.**

Quantifying imputation accuracy

Post-imputation SNP filtering was done based on MAF and the per-SNP Rsq [11]. On average, 5.5% SNPs with MAF > 0.005 were dropped from each study data set. Excluding the GRAAD(2) dataset, 72.6% of SNPs with MAF ≤ 0.005 were deleted (Supplementary Table S11). In contrast, 96% of GRAAD(2) SNPs with MAF ≤ 0.005 were dropped, most likely due to the small size of this sample (88 individuals), which may have affected the accuracy of the pre-phasing step performed during imputation. (Note: a MAF of 0.005 translates to the occurrence of a rare allele in 1 out of 100 individuals; imputation of allelic dosages rather than integer allele counts means it is still possible to observe a rare allele in only 88 individuals). A summary of the median Rsq

values by chromosome and dataset is shown in Supplementary Figure S8. Datasets with the largest density of SNPs performed best overall (i.e. the GRAAD(2) and CAG datasets had the highest Rsq values, despite their small sample sizes). Similar to the results of Liu et al., imputation quality was lowest for chromosome 19, which may be due to the high gene density of this chromosome [11].

To further gauge imputation accuracy, 15% of the overlapping SNPs genotyped in the GRAAD(2), HUFS and GALA II datasets on commercial arrays (i.e. not on the ADPC) were selected at random and omitted from the combined datasets (10,751 SNPs in total that are also present in the CAAPA reference panel), and the imputations were run again. In terms of sample size, the first two datasets are the smallest and largest African American sample sizes respectively, and the GRAAD(2) dataset had the highest density of genotyped SNPs. GALA II participants are from Puerto Rico, and this dataset was examined in detail due to differences in the population structure of Puerto Rico compared to the other studies (Supplementary Figures S1 and S2). Imputation error can be quantified by comparing typed SNPs and most probable imputed genotype at these masked SNPs. This comparison was done per SNP, yielding an accuracy proportion per SNP (number of imputed genotypes that matched the typed genotypes exactly, divided by the number of total genotypes that were compared). The median and interquartile range accuracies (i.e. the proportion of matching imputed and typed genotypes for a particular SNP) for GRAAD(2), GALA II and HUFS were 0.9886 [0.9545 – 1.0000], 0.9618 [0.9236 – 0.9829] and 0.9610 [0.9194 – 0.9838], respectively, indicating high imputation accuracy.

Genome-wide ancestry analysis

Genome-wide ancestry of subjects was assessed in order to identify population outliers to exclude from association analysis, as well as gauging the degree of population structure in the various studies (differences in ancestry between cases and controls).

<u>Methods</u>

First, SNPs with MAF < 1% were removed, followed by a linkage-disequilibrium (LD) pruning step. These steps were applied separately for each of the three ethnicities represented by the CAAPA studies (African American, African Caribbean = Barbados and Puerto Rican), and the datasets were then restricted to pruned SNPs common to all (122,206 SNPs). The algorithm implemented in the PLINK software package was used for the LD pruning step; SNPs within a window were recursively removed using a variance inflation factor threshold (window size 50 SNPs, step size 5 SNPs, variance inflation factor threshold 2). Note that for the SAPPHIRE dataset, the combined genome-wide ancestry analysis described here was restricted to the 730 asthmatic cases for which ADPC data were available.

Unrelated phase 3 thousand genomes project (TGP) subjects of European ancestry (CEU, n=84) and African ancestry (YRI, n=84), as well as unrelated Native American (NAT) subjects from Mao et al. [18] (n=43), were used as reference panels for the genome-wide ancestry analysis. Using the combined dataset of SNPs common to the reference panels plus the pruned CAAPA SNPs (16,410 SNPs), principal components were formed from the genotypes of the CAAPA and reference subjects by the R Bioconductor package. This software package calculates principal components by partitioning subjects into related and unrelated subsets (where the partitioning is based on a kinship matrix, see **Genetic covariance** structure). PC1 distinguished African ancestry from European and Native American ancestry and PC2 distinguished Native American ancestry from European and African ancestry. Based on this analysis, three outlier individuals that clustered close to the TGP CEU (European) subjects were identified, and these subjects were removed from the datasets prior to association testing.

The combined dataset (16,410 SNPs) was also used to estimate the proportion of genome-wide ancestry deriving from the three source populations represented by the reference panels, for each CAAPA subject. This was done using the software program ADMIXTURE. Because the BAGS and HUFS studies include families, and ADMIXTURE assumes that subjects are unrelated, only the founders in these studies (n=226 and n=997 respectively) were included for the ADMIXTURE analysis. The results are summarized in Supplementary Figure S2 and Supplementary Table S12.

Finally, the GENESIS PCA was repeated separately for each of the CAAPA datasets using the CAAPA pruned SNPs (122,206 SNPs), in order to test for differences in ancestry between cases and controls within each dataset. Tests for association between the top 10 principal components and case-control status were done separately for each principal component using logistic regression (*glm()* function in the R *base* package, binomial family with a logit link). For the JHS and GALA II study participants, association between ADMIXTURE estimates of genome-wide ancestry and case-control status was also tested using logistic regression (*glm()* function in the R *base* package, binomial family with a logit link), separately for each study and ancestry.

Results

The first principal component explained most of the variance in the genetic data (Supplementary Figure S9) in all of the datasets, which was similar to previous results in a GWAS of asthma in African ancestry populations [1]. Furthermore, differences in genome-wide ancestry between cases and controls in majority African ancestry populations were minimal. These differences were only evident on the first principal component in 2 of the 9 African American datasets, and were not detectable in Barbados. Only CAG, SAPPHIRE and HUFS had lower order principal components associated with asthma status (Supplementary Table S12). Differences in genome-wide ancestry were however observed between African American cases and controls from JHS (as evident on principal component 1, see Supplementary Table S12), and the Puerto Rican cases and controls from GALA II (as evident on principal components 1-3, see Supplementary Table S12). For these 3 datasets, cases had on average slightly, yet statistically significant, higher proportions of African ancestry compared to controls (Supplementary Figure S10).

Association analysis

Linear mixed effects models were used for association testing of the CAG, GRAAD, GALA II and SAGE datasets (as implemented in the EMMAX software package). This model uses a covariance matrix to account for relatedness between all individuals, as well as the observed population structure. This linear model has been shown to yield a good approximation for significance testing in case-control GWAS, when the number of cases and controls in a dataset is balanced [23]. However, the approximation breaks down for an unbalanced number of cases and controls, which was the case for some of the CAAPA datasets. For this reason, logistic regression was used to test for association in the JHS(1), JHS(2), GRAAD(2), SAPPHIRE and SARP datasets (as implemented in the PLINK software package), after ensuring the coefficient of relatedness between each pair of subjects was below 0.25. The first principal component, as well as any of the top 10 principal components associated with asthma status (p < 0.05, Supplementary S12), were included as covariates in these logistic regression models. Unrelated and family-based designs were used to recruit the HUFS subjects, and although this dataset has an unbalanced number of cases and controls, due to the high degree of relatedness between family members, fixed effects only regression models are not suited for association analysis in this dataset. Instead, logistic (and not linear) mixed effects models were used to test for association (as implemented in the GENESIS R Bioconductor package). This test uses a kinship matrix to encapsulate relatedness and excludes other sources of variance such as population structure, and therefore principal components were also included as fixed effect covariates, using the same inclusion strategy described above for the logistic regression models implemented in PLINK. The BAGS cohort is also comprised of families ascertained through

asthmatic cases, and includes 64 subjects with unknown phenotypes. To maximize statistical power, a quasi-likelihood score test was used to analyze BAGS families (as implemented in the MQLS software package) [24], using a specified disease prevalence of 0.2 [25]. This test allows both linkage and association to contribute to the test statistic, and also leverages genetic information from subjects with unknown phenotypes. Additional filtering of the association results was required prior to performing the meta-analysis. A more stringent Rsq filter was applied to the BAGS data to remove spurious p-values (Supplementary Figure S11). In addition, based on recommendations from the GIANT consortium [26], SNPs with a minor allele count (MAC) $\leq$ 6 were removed, which diminished test statistic inflation initially observed in some of these CAAPA datasets (Supplementary Figure S12). After performing these steps, the individual dataset and meta-analysis QQ plots showed little evidence of test statistic inflation (Supplementary Figure S3 and Figure 1).

Candidate gene analysis

Asthma associations with $p < 10^{-5}$ were selected from the GWAS catalog and Phenotype-Genotype Integrator (PheGenI) databases (http://www.genome.gov/gwastudies/ and http://www.ncbi.nlm.nih.gov/gap/phegenily, respectively). Gene regions were defined as 1) ±1MB from a gene for variants that fell within a single gene, 2) ±1MB from the variant and gene for variants reported for a gene but did not fall within the gene itself, 3) ±1MB from a variant for variants with no reported gene, and 4) ±1MB from the first and last gene for SNPs falling within a multi-genic region. Gene boundaries were based on USCS coordinates, and for those genes not in USCS, based on ENSEMBL BIOMART coordinates, and for those genes not in USCS or ENSEMBL BIOMART, based on the NCBI gene database coordinates.

Imputed data from the SAGE II dataset (the largest African American dataset with unrelated subjects) was used to estimate the number of independent SNPs in each identified candidate gene region. This was done using PLINK's sliding window LD-prune algorithm (window size of 50 SNPs, step size of 5 SNPs, and a variance inflation factor threshold of 5, corresponding to a multiple correlation coefficient of ≤0.8 between SNPs). The number of SNPs selected by the PLINK algorithm for a region was used as a proxy for the number of independent tests in that region. The p-values of all SNPs within a candidate gene region were multiplied by the number of independent tests estimated for that region to give Bonferroni corrected p-values. Corrected p-values < 0.05 were considered statistically significant.

Genetic covariance structure

Kinship matrices were estimated employing the same set of 122,206 CAAPA pruned SNPs used for genome-wide ancestry analysis.

Balding-Nichols kinship matrices were generated using the EMMAX software package and were included in the linear mixed effect models used to test for association between SNPs and asthma (CAG, GRAAD, BAGS, GALA II and SAGE; EMMAX software), as well as for the admixture mapping association tests. The matrices were used to account for relatedness between individuals as well as population structure (differences in ancestry between cases and controls).

Kinship matrices were also estimated using the KING-robust algorithm implemented in the KING software package [27], and was specified in the GENESIS logistic mixed effects models used for association tests of HUFS. The matrices are robust to the presence of population structure, and the KING kinship matrices were therefore also estimated and used to identify pairs of related subjects to exclude from logistic regression association tests. For each pair of subjects that shared a kinship coefficient > 0.25 (KING Kinship column × 2), the subject with the least represented phenotype were deleted (controls for GRAAD(2), SAPPHIRE, JHS(1), JHS(2), cases for SARP); if both subjects had the same phenotype the sample with the highest missing

genotype rate in the ADPC dataset were deleted. 2 controls and 1 case, 1 control, 15 controls, and 6 cases were deleted from the GRAAD(2), JHS(1), JHS(2) and SARP datasets respectively.

Local ancestry inference

First, GWAS array and ADPC genotyped autosomal SNPs that overlapped in 7,416 African American and Barbados subjects were merged in order to perform haplotype phasing using the ShapeIT software package (361,411 SNPs). After performing phasing, the 7,416 genomes were merged with African and European TGP phase 3 genomes (99 CEU and 108 YRI subjects, files downloaded from *ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ to data/raw/tgp_release_20130502*). The genetic map required as input for local ancestry inference was downloaded from *https://github.com/joepickrell/1000-genomes-genetic-maps/tree/master/interpolated_from_hapmap*. Custom scripts were used to create input files for local ancestry inference according to the RFMix input file specification. The final phased input files comprised 273,738 markers.

RFMix inferred 15,824 local ancestry segments across the CAAPA genomes. The mean proportion of African ancestry inferred for each of these segments is summarized graphically in Supplementary Figure S13. Deviations between mean local African ancestry and mean genome-wide African ancestry are normally distributed, as expected (Supplementary Figure S14). Estimates of genome-wide ancestry from ADMIXTURE and RFMix were also highly concordant (Supplementary Figure S15), with a mean estimate difference of only 0.0009 (less than 0.1%).

Admixture mapping

Custom scripts were used to convert the RFMix local ancestry calls to an EMMAX dosage TPED file, where the encoded "dosage" is defined as having 0, 1 or 2 copies of African ancestry at a particular local ancestry segment. Tests for association between the local ancestry dosage values and asthma case-control status were then done using the linear mixed effect models implemented in the EMMAX software package. Phenotypes for subjects from SAPPHIRE, which included only asthmatics and no controls on the ADPC, as well as the 3 population outliers identified by the genome-wide ancestry analysis, were set to missing. The models included dataset as a fixed effect covariate, and a Balding-Nichols kinship matrix as random effect.

The method described by Gao et al. was used to estimate the number of effective tests that should be used in a Bonferroni correction for multiple testing [28]. Briefly, a local ancestry correlation matrix, and corresponding eigenvalues, was calculated using local ancestry dosage values (0/1/2 copies of African ancestry) per chromosome. The number of effective tests for a chromosome is then set to the number of eigenvectors that explain 99.5% of the variance in the local ancestry data (for n local ancestry segments, find the largest k such that $\sum_{i=1}^{k} eigenvalues / \sum_{i=1}^{n} eigenvalues \leq 0.995$). The total number of effective tests is then set to the sum of the number of effective tests calculated for each of the 22 autosomes. Using this method, a Bonferroni corrected p-value threshold of 0.05/(262 total number of effective tests)= 1.9×10^{-4} was used to claim statistical significance.

TaqMan genotyping and imputation data concordance

Genotyping was performed on the ABI 7900HT Sequence Detection System on 3 single nucleotide polymorphisms (SNPs). Using samples from HUFS, the top three SNP associations on chromosome 2q22.3 were genotyped in order to check for concordance with imputed genotypes (rs17834780, rs787151 and rs787151).

SNP Genotyping

1. Genomic DNA was extracted from peripheral blood samples.

2. Genotyping was performed using the TaqMan™ probe based, 5' nuclease assay with minor groove binder (MGB) chemistry (ABI, Foster City,CA). Genotyping was performed at a total volume of 5 µL per well, in 384-well MicroAmp Optical plates, using real-time detection followed by an endpoint read (Allelic discrimination) using the ABI 7900 HT Sequence Detection System software SDS2.4 (Applied Biosystems, Foster City, CA, USA).

3. Fluorescence data files from each plate were analyzed on scatter plots in the allelic discrimination window by automated allele calling software SDS 2.4 (ABI, Foster City, CA) and visually inspected.

3. Quality control: 5 – 10% of the randomly selected samples were repeated as routine quality control procedure.

Genotype concordance

TaqMan genotype call quality was assessed by means of manual inspection of the allelic discrimination plots, checking the proportion of successful genotype calls, and tests for HWE. The imputed genotype calls for corresponding samples were then converted to "hard" genotype calls, for those genotypes with imputation probabilities of at least 80%. The percentage of matching TaqMan and imputed genotype calls was then calculated. These results are summarized in Supplementary Table S8.

SUPPLEMENTAL FIGURES

Supplementary Figure S1: CAAPA principal components. This figure summarizes the first three principal components generated from a combined dataset of 16,410 overlapping SNPs in 84 African (YRI) and 84 European (CEU) TGP phase 3 subjects, 43 Native American (NAT) subjects and 9,170 CAAPA subjects. Three CAAPA outliers that clustered with CEU subjects (bottom leftmost corner of the plot) were identified based on this principal component analysis, and were excluded from the association analysis.
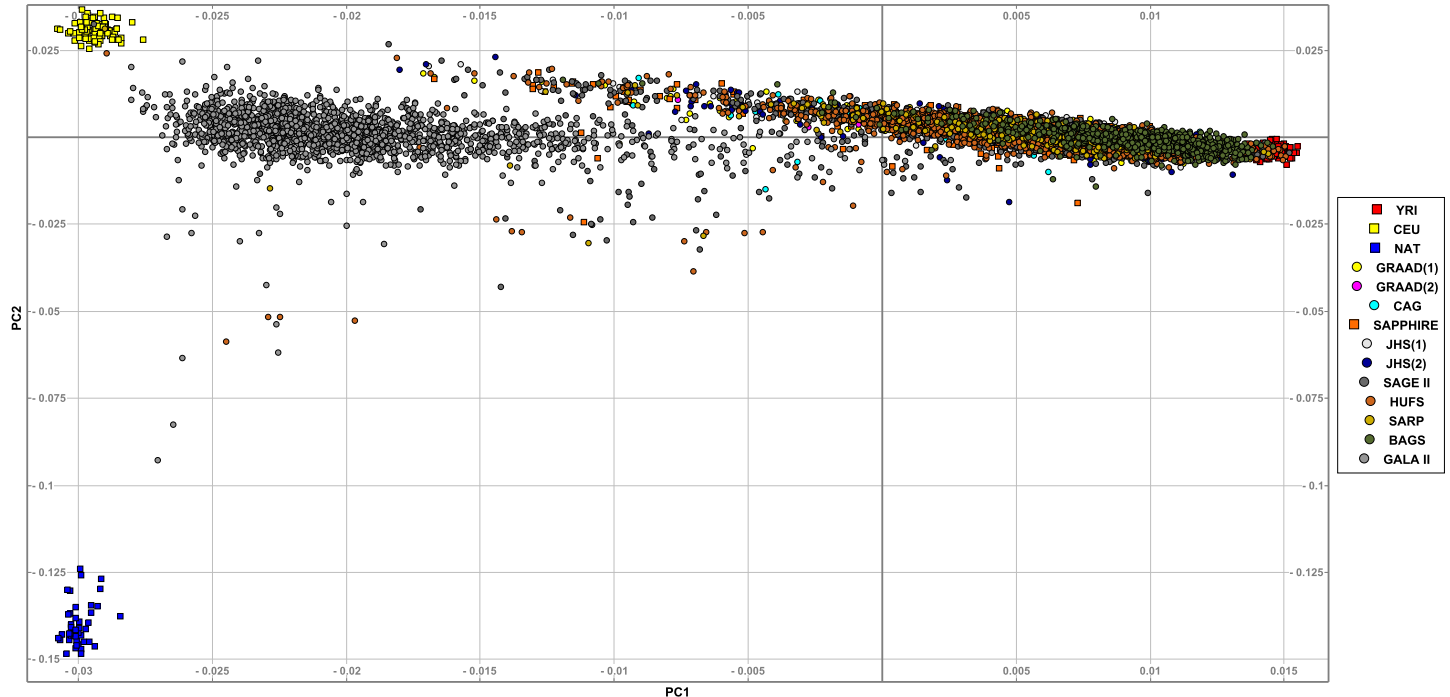
Supplementary Figure S2: CAAPA ADMIXTURE estimates. This figure summarizes the genome-wide proportions of ancestry for K=3 populations, as estimated by the software program ADMIXTURE. A combined dataset of 16,410 overlapping SNPs in 84 African (YRI, red) and 84 CEU (European, yellow) TGP phase 3 subjects, 43 Native American (NAT, blue) subjects and 7,861 putatively unrelated CAAPA subjects were used to generate these estimates.
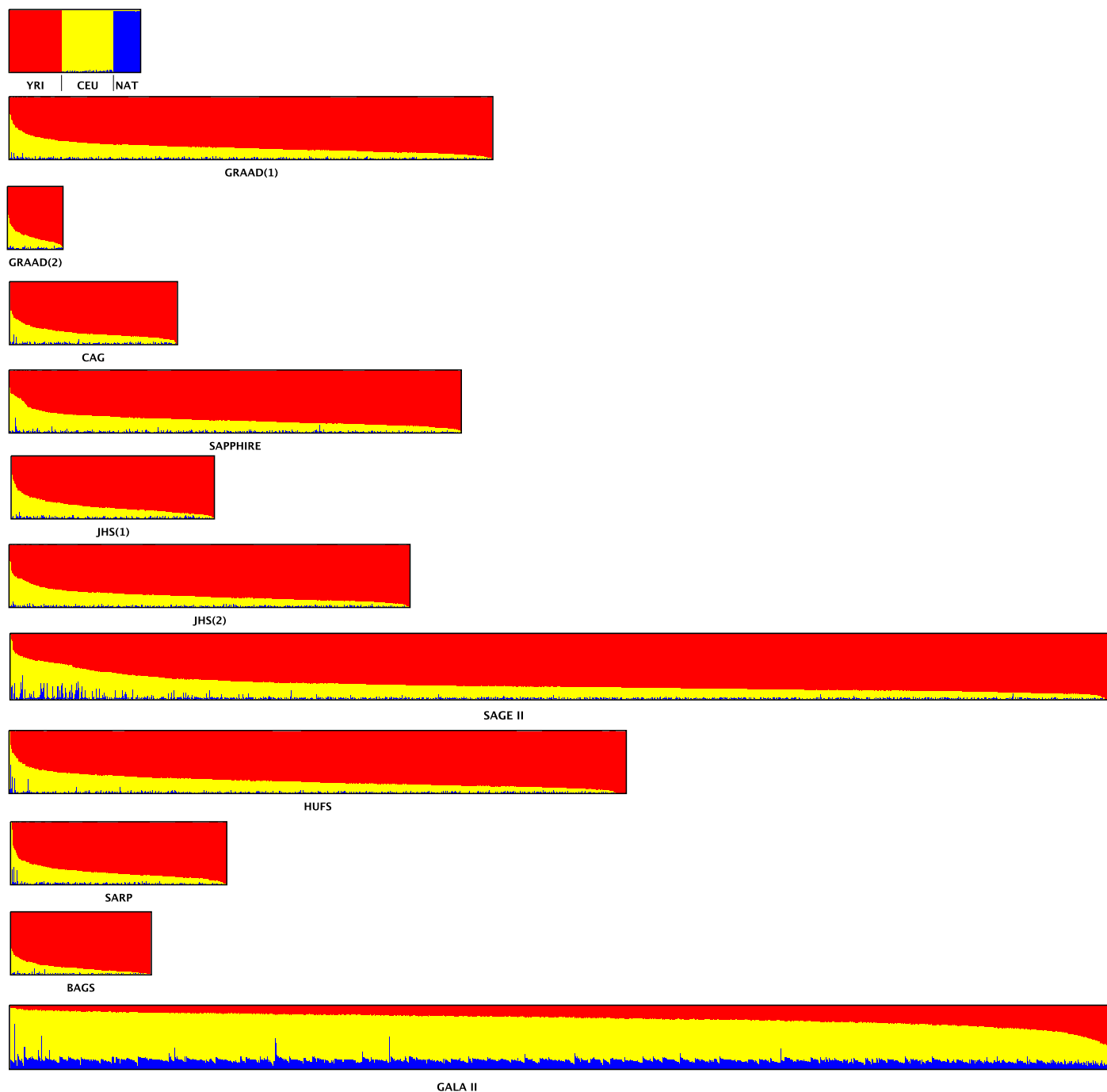
Supplementary Figure S3: QQ plots summarizing the distribution of p-values per dataset**, stratified by minor allele frequency (MAF).** SNPs with minor allele count ≤ 6 were filtered out from the result sets. Inflation factors were calculated by transforming p-values to 1 degree of freedom (df) Chi-square statistics, and dividing the median of these statistics by the median of the theoretical Chi-square (1 df) distribution.
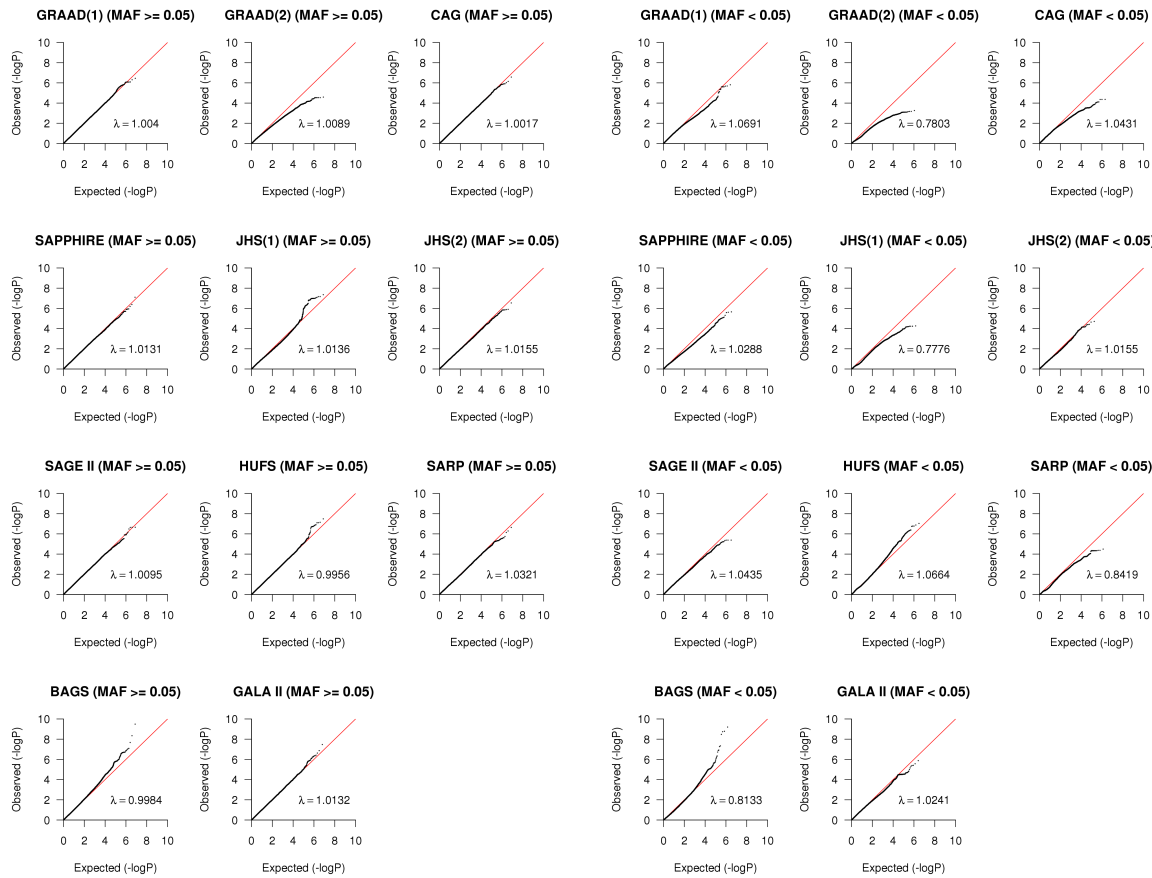
Supplementary Figure S4: Locus zoom plots of the association p-values for the chromosome 1q21.3, 1q23.1, 2q22.3, 6q23.2 and 17q21 regions. In brackets, the meta-analysis result set used to generate the figures: "all" is the result set including all datasets, "excluding GALA II" is the result set excluding GALA II.
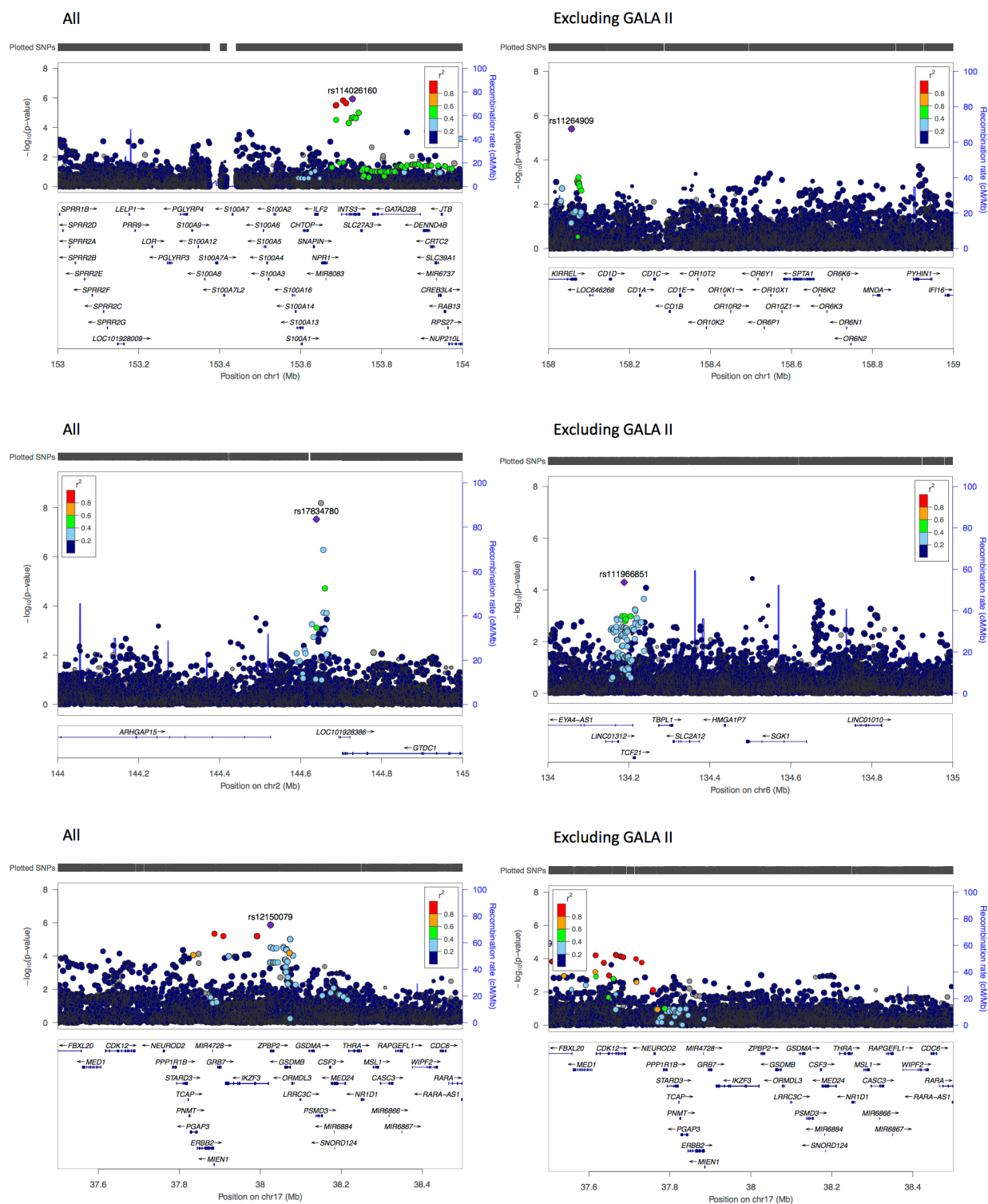
Supplementary Figure S5: Summary of the distribution of admixture mapping p-values. A QQ plot is shown in the top panel, followed by a histogram of p-values, and the Manhattan plot is shown in the bottom panel. The inflation factor was calculated by transforming p-values to 1 degree of freedom (df) Chi-square statistics, and dividing the median of these statistics by the median of the theoretical Chi-square (1 df) distribution. The vertical red line in the histogram indicates the median p-value, and the horizontal red line represent the theoretical uniform distribution that the p-values are expected to follow. The horizontal red line in the Manhattan plot represents the p-value significance threshold of 1.9×10^{-4}.
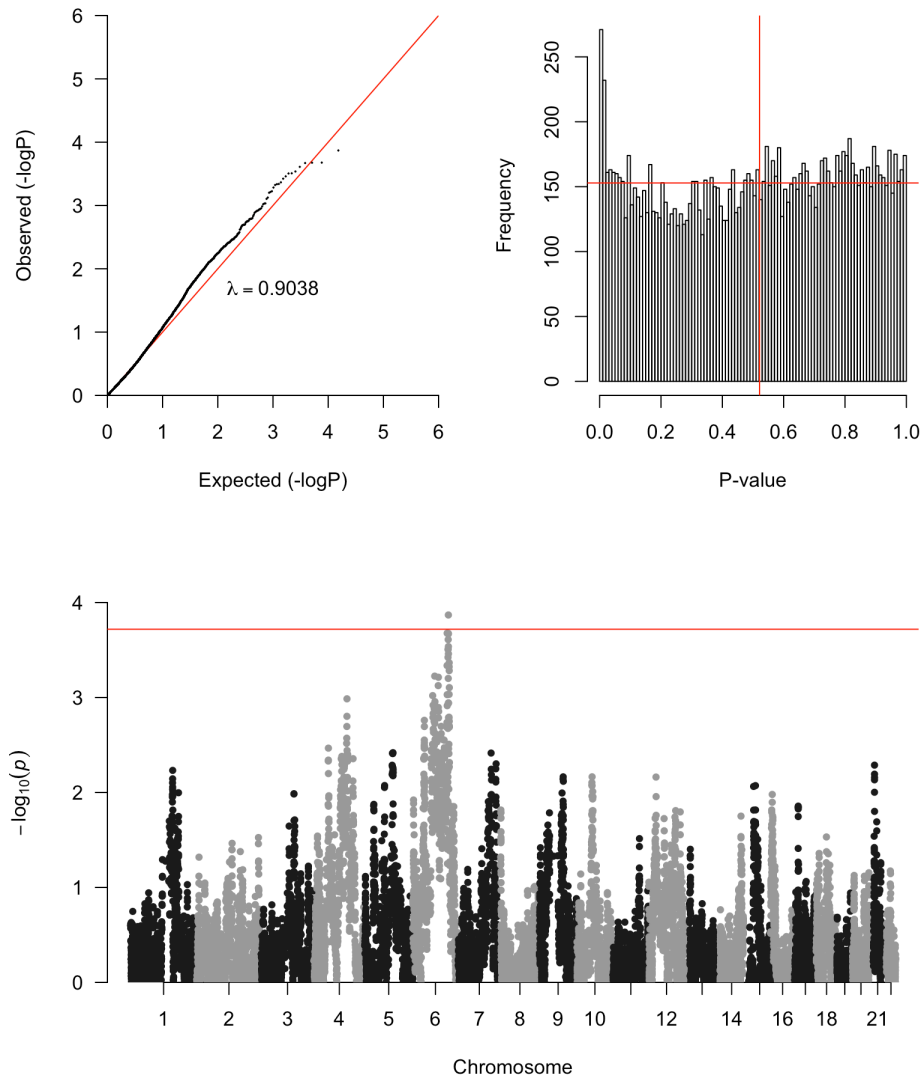
Supplementary Figure S6: Summary of the distribution of the DP and GQ variant call quality metrics in the CAAPA WGS reference panel, for rs787160.

Histogram of DP

Histogram of GQ
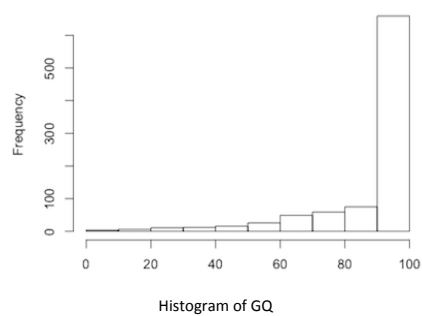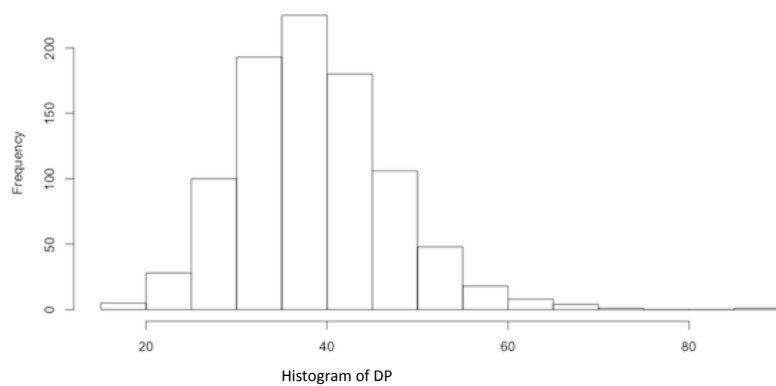
Supplementary Figure S7: Flow diagram of GWAS+ADPC dataset merging and imputation. This figure depicts the steps used to merge and impute the GWAS backbone and ADPC datasets, as applied to HUFS.
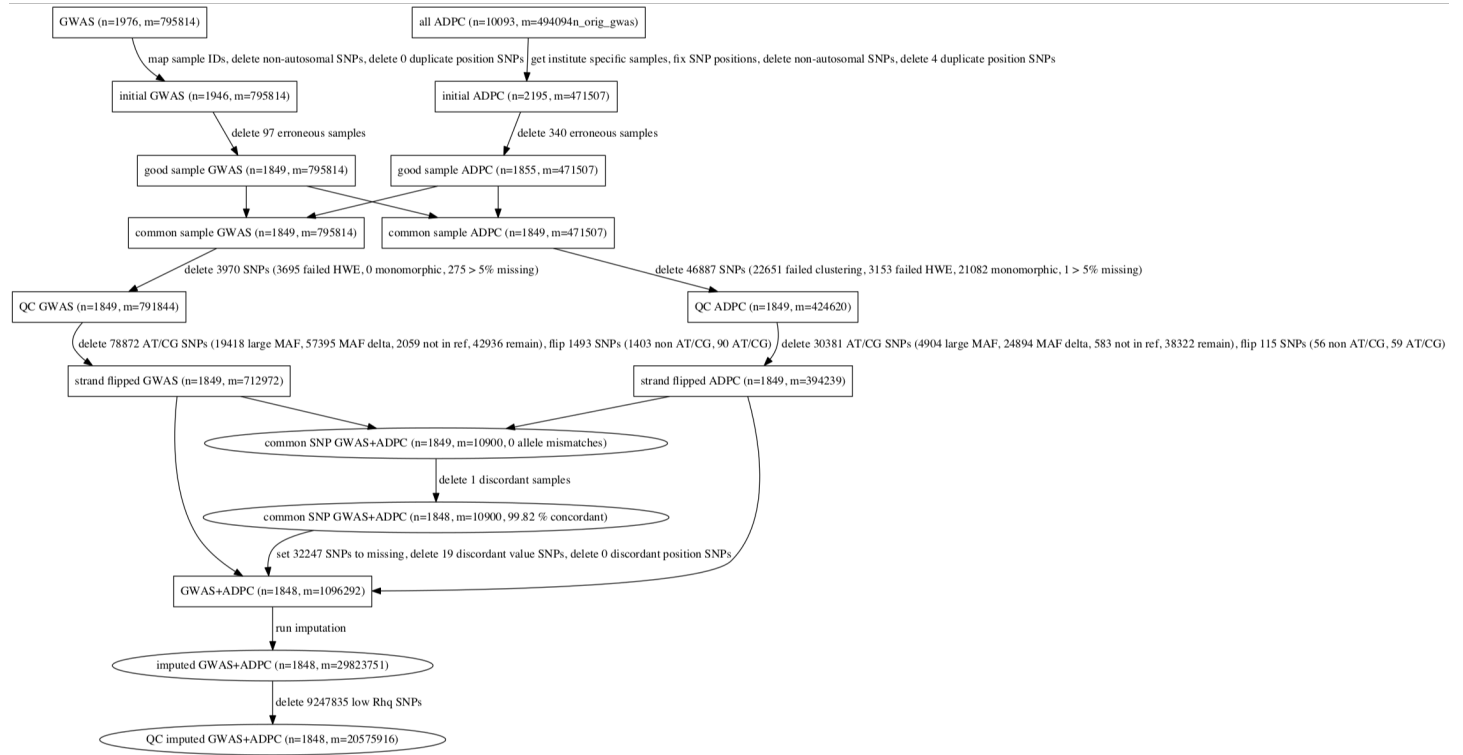
Supplementary Figure S8: Median Rsq values by chromosome and dataset. The first panel represents the median Rsq values across all SNPs, the second panel represents the median Rsq values when sub-setting to SNPs with MAF ≤ 0.005, and the third panel represents the median Rsq values when sub-setting to SNPs with MAF > 0.005.
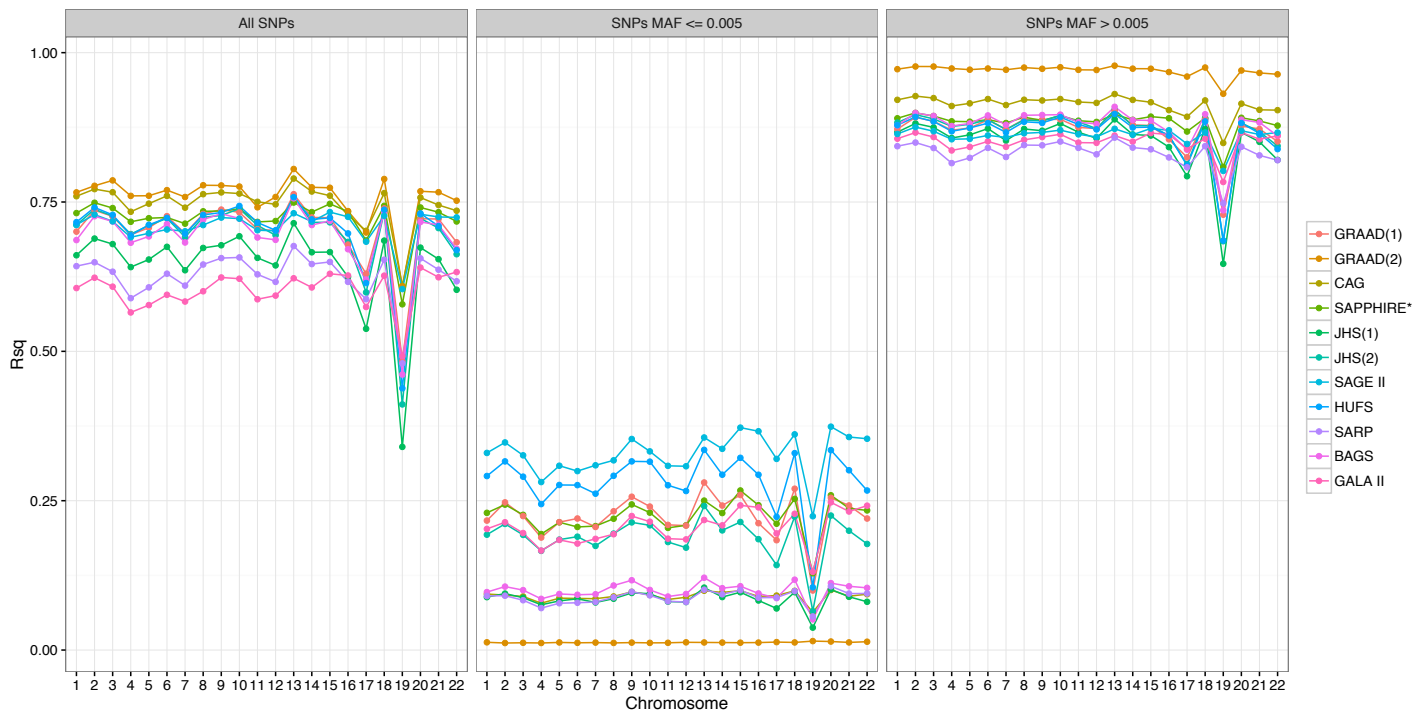
Supplementary Figure S9: Scree plots of the within-dataset principal component analysis. Principal components were calculated using GENESIS separately for each dataset, excluding reference population panels. The plots show the eigenvalues of each principal component. Principal component 1 explains most of the variance in the data for all the CAAPA datasets.
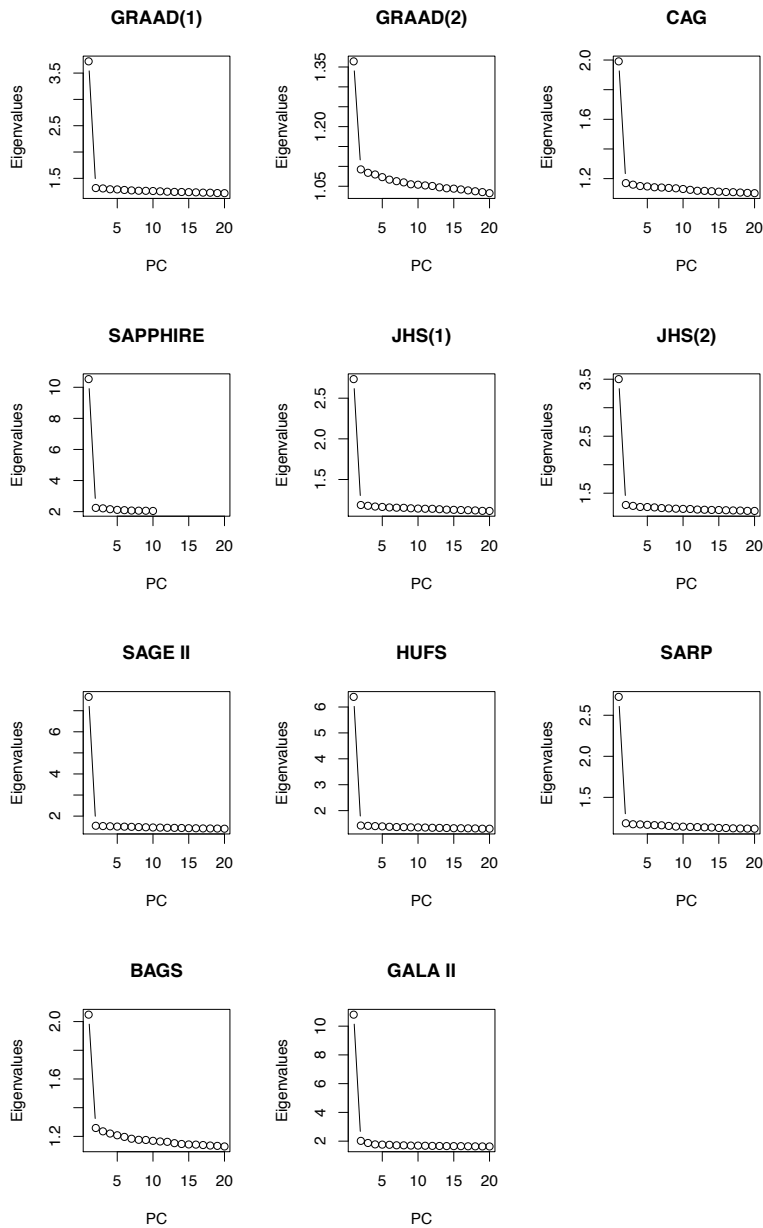
Supplementary Figure S10: Distribution of ADMIXTURE estimates in the JHS and GALA II datasets. The association between ADMIXTURE estimates of genome-wide ancestry and case-control status was tested using logistic regression, separately for each dataset and ancestry. The ADMIXTURE estimates shown in Supplementary Figure S4 were used to generate these boxplots and association p-values.
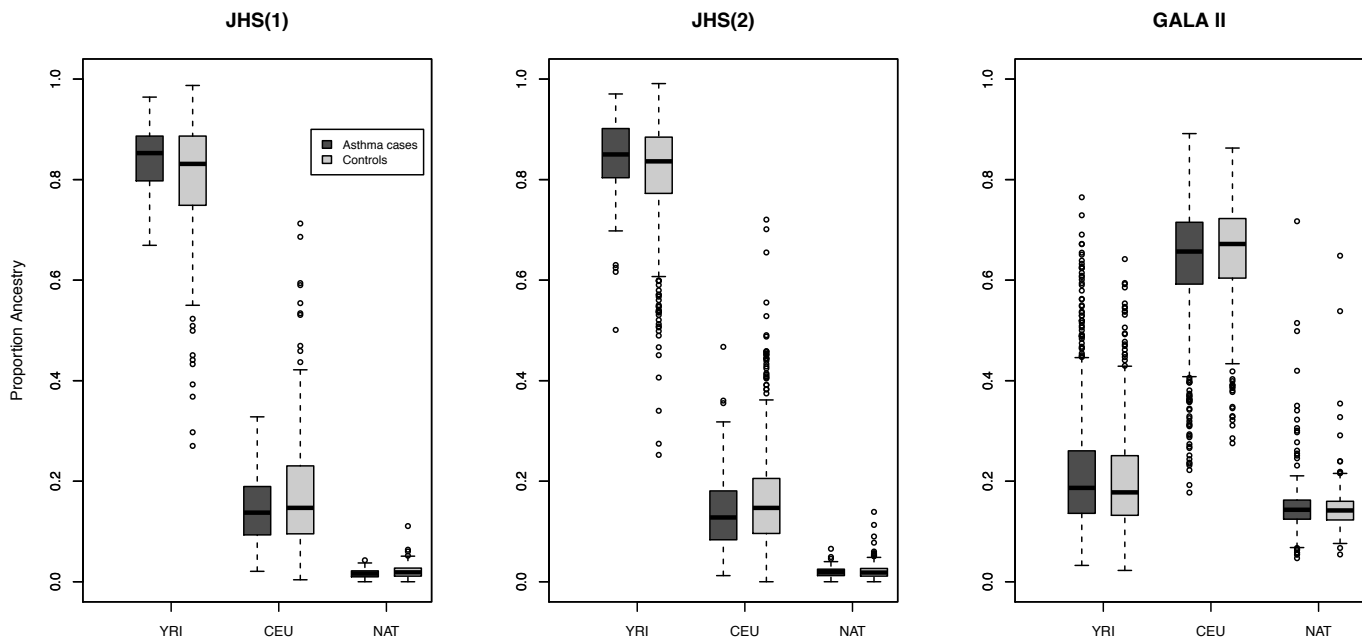
Supplementary Figure S11: Summary of the distribution of BAGS p-values, estimated using MQLS, for different Rsq filters. Inflation factors were calculated by transforming p-values to 1 degree of freedom (df) Chi-square statistics, and dividing the median of these statistics by the median of the theoretical Chi-square (1 df) distribution. A.) QQ plots filtered using Minimac Rsq and MAF estimates. SNPs with MAF ≤ 0.005 were deleted if Rsq ≤ 0.5 and SNPs with MAF > 0.005 were deleted if Rsq ≤ 0.3. B.) QQ plots filtered using MQLS Rsq and control MAF estimates, that takes into account the correlation between subjects due to relatedness. SNPs with MAF ≤ 0.005 were deleted if Rsq ≤ 0.5 and SNPs with MAF > 0.005 were deleted if Rsq ≤ 0.3. C.) QQ plots filtered using MQLS Rsq and control MAF estimates. SNPs with MAF ≤ 0.005 were deleted if Rsq ≤ 0.7 and SNPs with MAF > 0.005 were deleted if Rsq ≤ 0.5. This filter was applied prior to including BAGS in the meta-analysis.
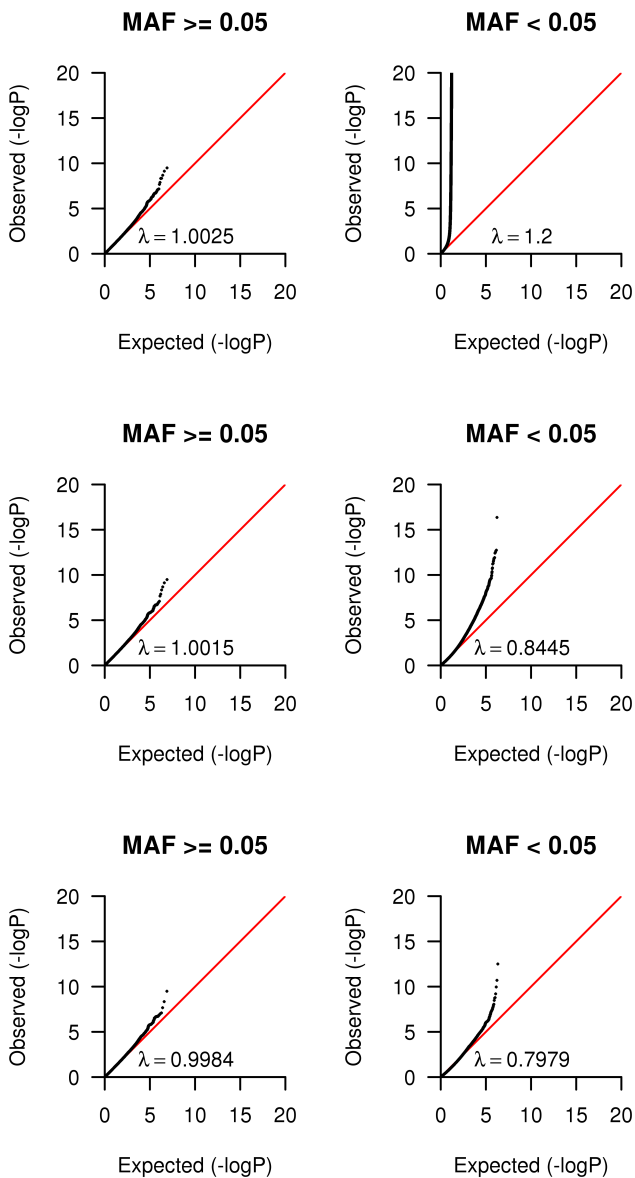
Supplementary Figure S12: Summary of the distribution of p-values for SNPs with MAF < 0.05, prior to applying a minor allele count (MAC) filter. QQ plots and histograms of p-values are shown in the left and right panels respectively. Inflation factors were calculated by transforming p-values to 1 degree of freedom (df) Chi-square statistics, and dividing the median of these statistics by the median of the theoretical Chi-square (1 df) distribution. The vertical red lines in the histograms indicate the median p-value, and the horizontal red lines represent the theoretical uniform distribution that the p-values are expected to follow.
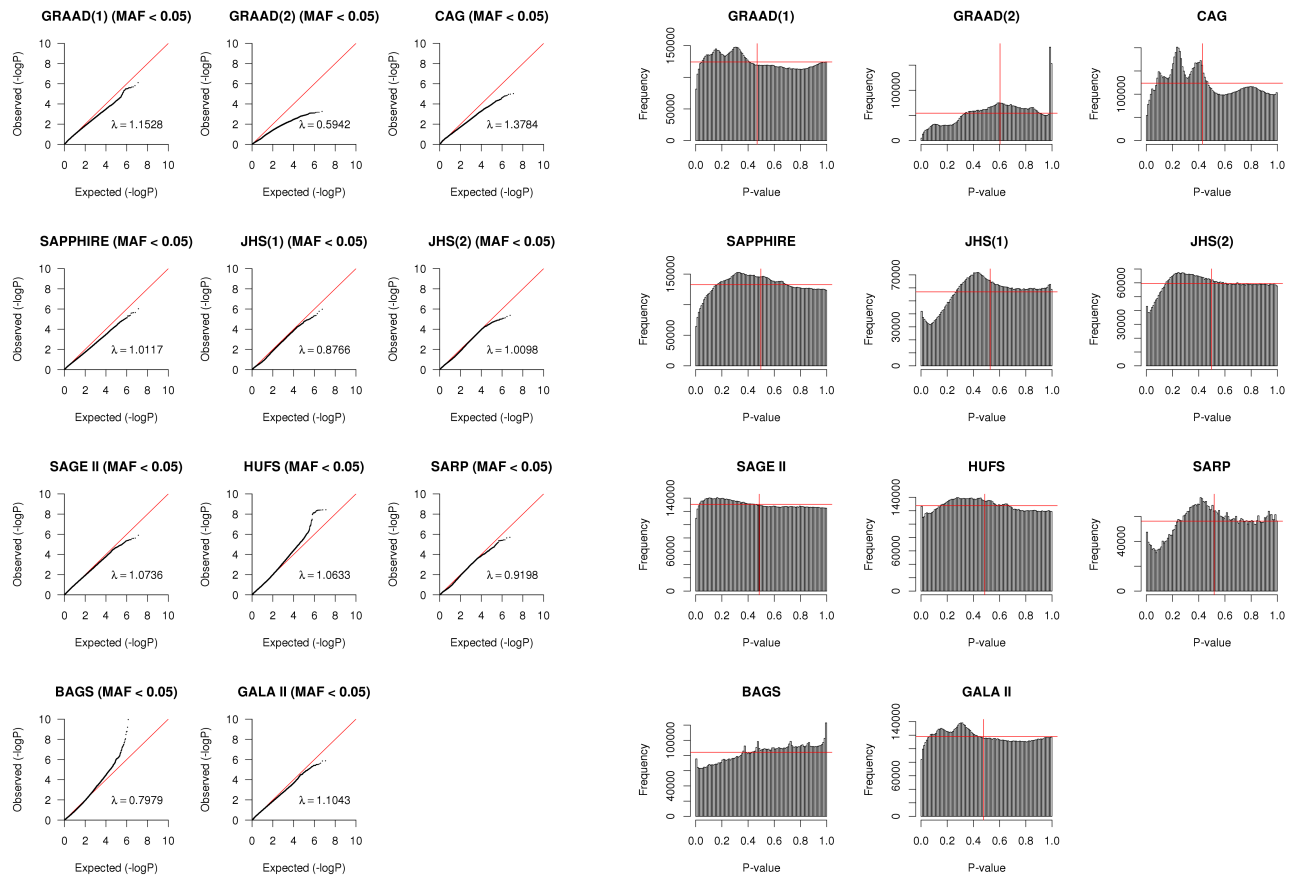
Supplementary Figure S13: The mean proportion of Africa ancestry across the genome of 7,146 CAAPA subjects, for the 15,824 local ancestry segments inferred by RFMix. The mean local African ancestry for a segment is the proportion of haplotypes called as African for that segment.
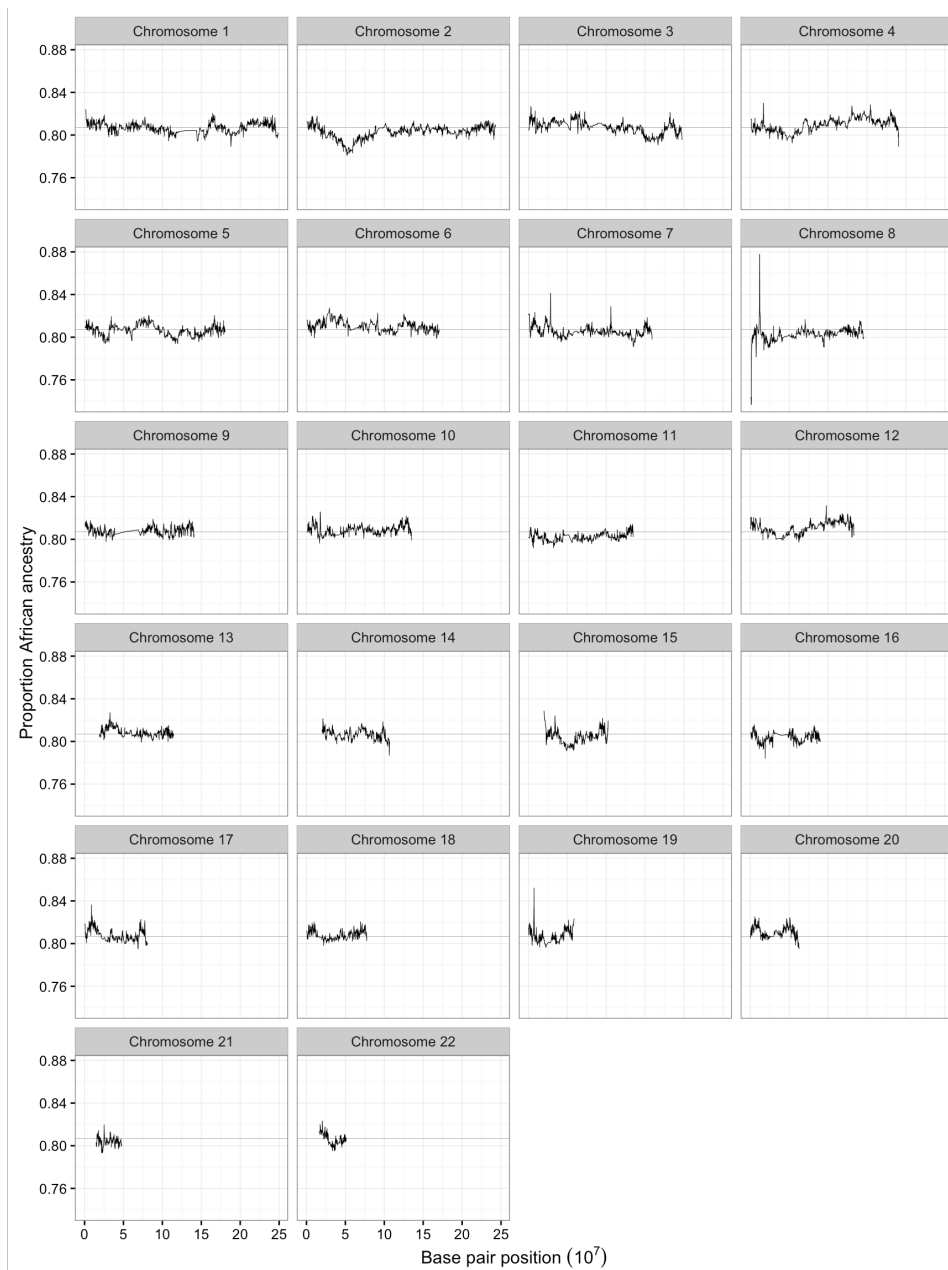
Supplementary Figure S14: Histogram of the standardized deviations between mean local African ancestry and mean genome-wide African ancestry, for the 15,824 local ancestry segments inferred by RFMix. The mean local African ancestry for a segment is the proportion of haplotypes called as African for that segment. The mean genome-wide African ancestry was calculated by dividing the number of SNPs called by RFMix as African, by the total number of SNPs for which local ancestry was called, across all genomes.
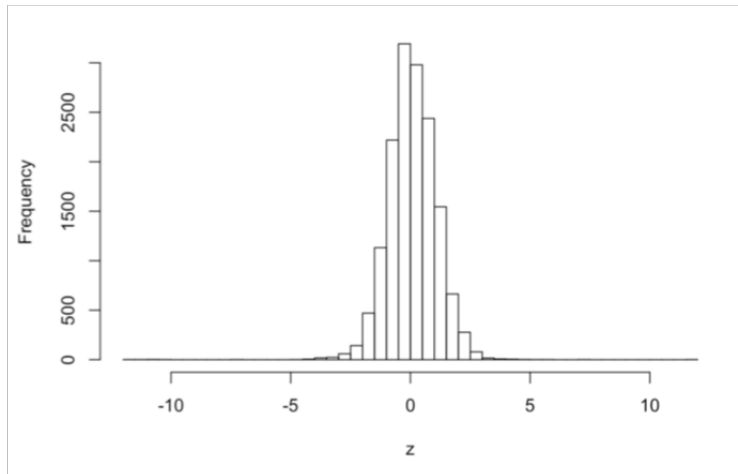
Supplementary Figure S15: Histogram of the differences between genome-wide ancestry estimated using ADMIXTURE, and genome-wide ancestry estimated using RFMix. The difference (delta) was calculated for each of the 7,416 individuals for which local ancestry was called. The per-individual RFMix mean genome-wide African ancestry was calculated by dividing the number of SNPs called by RFMix as African, by the total number of SNPs for which local ancestry was called across the genome, for that individual.
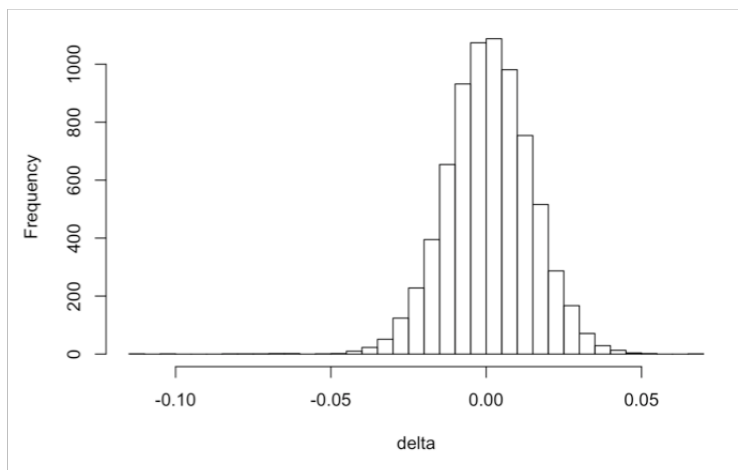
SUPPLEMENTAL TABLES

Supplementary Table S1: Clinical characteristics summary. The distribution of age, sex and age of asthma onset is summarized in this table by dataset used in the meta-analysis. SEM=standard error of the mean. Childhood onset asthma is defined as having been diagnosed with asthma before 16 years of age. The information summarized in this table was not available for all subjects included in the association analysis; therefore the number of subjects (n) used to calculate these statistics is given in brackets.

	Mean age (± SEM)		% Males		% Childhood onset
	Cases	Controls	Cases	Controls	
GRAAD(1)	23.5 (±0.9) (n=395)	35.0 (±0.9) (n=385)	46.1% (n=395)	39.2% (n=385)	68.8% (n=109)
GRAAD(2)	44.0 (±1.5) (n=64)	38.7 (±2.6) (n=23)	27.7% (n=65)	30.4% (n=23)	55.7% (n=61)
CAG	23.6 (±1.7) (n=113)	38.0 (±0.8) (n=156)	46.0% (n=113)	26.3% (n=156)	76.1% (n=113)
SAPPHIRE	32.6 (±0.4) (n=1,243)	39.1 (±0.6) (n=485)	36.0% (n=1,243)	34.8% (n=485)	69.0% (n=1,243)
JHS(1)	66.0 (±0.9) (n=44)	65.9 (±0.3) (n=283)	15.9% (n=44)	29.7% (n=283)	26.8% (n=41)
JHS(2)	49.6 (±1.2) (n=101)	49.5 (±0.5) (n=546)	24.8% (n=101)	37.7% (n=546)	48.3% (n=89)
SAGE II	14.7 (±0.2) (n=998)	18.4 (±0.3) (n=690)	50.6% (n=1001)	42.5% (n=691)	97.1% (n=837)
HUFS	38.1 (±0.9) (n=303)	39.6 (±0.4) (n=1527)	31.7% (n=303)	41% (n=1527)	61.1% (n=293)
SARP	31.2 (±0.8) (n=302)	36.0 (±1.6) (n=45)	42.4% (n=302)	26.7% (n=45)	70.3% (n=296)
BAGS	22.7 (±0.8) (n=282)	37.3 (±0.9) (n=338)	49.3% (n=282)	46.7% (n=338)	80.6% (n=247)
GALA II	12.4 (±0.1) (n=900)	13.5 (±0.1) (n=853)	55.4% (n=901)	46.8% (n=853)	99.7% (n=717)
Total	20.64 (n=4,745)	22.83 (n=5,331)	45.2% (n=4,750)	40.3% (n=5,332)	79.6% (n=4,046)

Supplementary Table S2: ANNOVAR file versions.

FILTER BASED:

build	Table Name	Explanation	Date
hg19	avsnp147	dbSNP147 with allelic splitting and left-normalization	20160606
hg19	snp138	dbSNP with ANNOVAR index files	20140910
hg19	snp138NonFlagged	dbSNP with ANNOVAR index files, after removing those flagged SNPs (SNPs < 1% minor allele frequency (MAF) (or unknown), mapping only once to reference assembly, flagged in dbSnp as "clinically associated")	20140222
hg19	gerp++elem	conserved genomic regions by GERP++	20140223
hg19	cadd13	CADD version 1.3	20160607
hg19	ljb26_all	whole-exome SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, MetaSVM, MetaLR, VEST, CADD, GERP++, PhyloP and SiPhy scores from dbNSFP version 2.6	20140925
hg19	dbnsfp30a	whole-exome SIFT, PolyPhen2 HDIV, PolyPhen2 HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, MetaSVM, MetaLR, VEST, CADD, GERP++, DANN, fitCons, PhyloP and SiPhy scores from dbNSFP version 3.0a	20151015
hg19	dbnsfp31a_interpro	protein domain for variants	20151219
hg19	dbscsnv11	dbscSNV version 1.1 for splice site prediction by AdaBoost and Random Forest	20151218
hg19	cg69	allele frequency in 69 human subjects sequenced by Complete Genomics	20120222
hg19	cosmic70	COSMIC database version 70 on WGS data	20140911
hg19	exac03	ExAC 65000 exome allele frequency data for ALL, AFR (African), AMR (Admixed American), EAS (East Asian), FIN (Finnish), NFE (Non-finnish European), OTH (other), SAS (South Asian)). version 0.3. Left normalization done.	20151129
hg19	exac03nontcga	ExAC on non-TCGA samples (updated header)	20160423
hg19	exac03nonpsych	ExAC on non-Psychiatric disease samples (updated header)	20160423
hg19	kaviar_20150923	170 million Known VARiants from 13K genomes and 64K exomes in 34 projects	20151203
hg19	hrcr1	40 million variants from 32K samples in haplotype reference consortium	20151203
hg19	1000g2015aug (6 data sets)	alternative allele frequency data in 1000 Genomes Project for autosomes (ALL, AFR (African), AMR (Admixed American), EAS (East Asian), EUR (European), SAS (South Asian)). Based on 201409 collection v5 (based on	20150824

Build	Table Name	Explanation	Date
		201305 alignment) but including chrX and chrY data finally! The 1000G team fixed a bug in chrX frequency calculation. Based on 201508 collection v5b (based on 201305 alignment)	
hg19	clinvar_20160302	CLINVAR database with Variant Clinical Significance (unknown, untested, non-pathogenic, probable-non-pathogenic, probable-pathogenic, pathogenic, drug-response, histocompatibility, other) and Variant disease name	20160303
hg19	popfreq_all_20150413	A database containing all allele frequency from 1000G, ESP6500, ExAC and CG46	20150413

REGION BASED:

Cell	Description	Lineage	Tissue	Karyotype	Date for the HMM model
GM12878	B-lymphocyte, lymphoblastoid, International HapMap Project - CEPH/Utah - European Caucasion, Epstein-Barr Virus	mesoderm	blood	normal	20110626
H1-hESC	embryonic stem cells	inner cell mass	embryonic stem cell	normal	20110626
K562	leukemia, "The continuous cell line K-562 was established by Lozzio and Lozzio from the pleural effusion of a 53-year-old female with chronic myelogenous leukemia in terminal blast crises." - ATCC	mesoderm	blood	cancer	20110626
HepG2	hepatocellular carcinoma	endoderm	liver	cancer	20110626
HUVEC	umbilical vein endothelial cells	mesoderm	blood vessel	normal	20110626
HMEC	mammary epithelial cells	ectoderm	breast	normal	20110626
HSMM	skeletal muscle myoblasts	mesoderm	muscle	normal	20110626
NHEK	epidermal keratinocytes	ectoderm skin	normal		20110626
NHLF	lung fibroblasts	endoderm	lung	normal	20110626

Other region based databases	Date
Encode CTCF binding site for Gm12878 tissues	20130512
Variants located in DNase I hypersensitivity sites from ENCODE	20141102
Annotate ENCODE transcription factor ChIP-Seq data	20110523
Variant located in a "hot spot" of DNAse I hypersensitivity sites in the GM12878 cell line	20101226

GENE BASED:

Build	Table Name	Explanation	Date
hg19	refGene	FASTA sequences for all annotated transcripts in RefSeq Gene	20151211
hg19	knownGene	FASTA sequences for all annotated transcripts in UCSC Known Gene	20151211
hg19	ensGene	FASTA sequences for all annotated transcripts in ENSEMBL Gene	20151211

Supplementary Table S3: Statistical models used for association testing. For the logistic regression models that included only fixed effects (no kinship matrix), for each pair of subjects that shared a kinship coefficient > 0.25, the subject with the least represented phenotype were deleted. For all logistic regression models, the first principal component as well as any of the top 10 principal components that were associated with asthma status (p-value < 0.05), were included as covariates.

	Design	% Cases	Statistical model	Covariates	Software
GRAAD(1)	Unrelated	51%	Linear mixed model	Balding-Nichols kinship matrix	EMMAX
GRAAD(2)	Unrelated	75%	Logistic regression	Principal component 1	PLINK
CAG	Unrelated	42%	Linear mixed model	Balding-Nichols kinship matrix	EMMAX
SAPPHIRE	Unrelated	70%	Logistic regression	Principal components 1 & 8	PLINK
JHS(1)	Unrelated	14%	Logistic regression	Principal component 1	PLINK
JHS(2)	Unrelated	16%	Logistic regression	Principal component 1	PLINK
SAGE II	Unrelated	59%	Linear mixed model	Balding-Nichols kinship matrix	EMMAX
HUFS	Unrelated and families	17%	Logistic mixed model	KING kinship matrix, Principal components 1 & 8	GENESIS, KING
SARP	Unrelated	87%	Logistic regression	Principal component 1	PLINK
BAGS	Asthmatics and their nuclear and extended families	41%	MQLS		MQLS
GALA II	Unrelated	51%	Linear mixed model	Balding-Nichols kinship matrix	EMMAX
Combined		47%	Weighted Z-score		METAL

Supplementary Table S5: Candidate gene association results. This table summarizes the candidate genes that are associated with asthma after applying a Bonferroni correction using the number of independent markers in the candidate gene region. Results are shown for the meta-analysis of all the CAAPA studies, and the meta-analysis that excludes the GALA II study, due to the relatively higher European ancestry within GALA II.

Meta-analysis of all CAAPA studies					
Chromosome: hg19 position	Candidate gene region	Start position	End position	Meta-analysis p-value	Corrected p-value
1:153688130	CRNN_LCE5A_CRCT1_LCE3E_IL6R_ADAR_KCNN3	151381718	155842754	3.06E-06	0.0423
1:153704977	CRNN_LCE5A_CRCT1_LCE3E_IL6R_ADAR_KCNN3	151381718	155842754	1.50E-06	0.0208
1:153714954	CRNN_LCE5A_CRCT1_LCE3E_IL6R_ADAR_KCNN3	151381718	155842754	1.36E-06	0.0189
1:153727808	CRNN_LCE5A_CRCT1_LCE3E_IL6R_ADAR_KCNN3	151381718	155842754	1.18E-06	0.0163
2:144650140	LRP1B_LOC101928386_LOC105373654	139988995	145598432	6.51E-09	0.0002
2:207007157	PARD3B_CPO_KLF7	204410515	209031970	1.86E-06	0.0363
10:105730220	SH3PXD2A	104348284	106615164	9.53E-07	0.0079
17:37886986	IKZF3_GSDMB_ORMDL3_LRRC3C_GSDMA	36913968	39134019	4.54E-06	0.0402
17:38025417	IKZF3_GSDMB_ORMDL3_LRRC3C_GSDMA	36913968	39134019	1.35E-06	0.0120
18:31058259	CCDC178_ASXL3_LOC107985157	29517365	32390868	3.70E-06	0.0380
Meta-analysis excluding GALA					
Chromosome: hg19 position	Candidate gene region	Start position	End position	Meta-analysis p-value	Corrected p-value
1:158056779	PYHIN1_OR10J3_OR10J7P	157901336	160321816	3.98E-06	0.0345
2:144650140	LRP1B_LOC101928386_LOC105373654	139988995	145598432	1.37E-07	0.0031
2:207007157	PARD3B_CPO_KLF7	204410515	209031970	1.86E-06	0.0350
2:240496929	LOC101928111_LOC401040	239441158	241441158	3.26E-06	0.0339
5:94736141	KIAA0825	92486555	94954309	4.28E-06	0.0344
8:27338294	CHRNA2_EPHX2	26317277	28402439	2.31E-06	0.0229
9:89759848	ZCCHC6_LOC105376124_GAS1	87902647	90562104	3.46E-06	0.0385
17:26241077	NOS2	25083791	27220409	5.09E-06	0.0247
1:158056779	PYHIN1_OR10J3_OR10J7P	157901336	160321816	3.98E-06	0.0345
2:144650140	LRP1B_LOC101928386_LOC105373654	139988995	145598432	1.37E-07	0.0031
2:207007157	PARD3B_CPO_KLF7	204410515	209031970	1.86E-06	0.0350

Supplementary Table S6: Summary of the admixture mapping peaks on chromosomes 1, 4 and 6. A linear mixed model (EMMAX software) was used to test for association between the number of copies of African ancestry ("dosage" value of 0, 1, 2) and case-control status (2,608 cases and 3,994 controls). Since ADPC data was only available for SAPPHIRE cases, the study was not included in the model. The most significant segments are highlighted in **bold font**. Z is the standardized deviation between the mean African ancestry of a local ancestry segment in cases/controls/SAPPHIRE, and the grand African ancestry mean in cases/controls/SAPPHIRE (the means and Z were calculated separately for each of these 3 groups).

Chr	Start hg19 position	End hg19 position	Linear model Beta	P-value	Proportion African ancestry					
					Case mean	Control mean	SAPPHIRE mean	Case Z	Control Z	SAPPHIRE Z
1	158036570	158081768	-0.0239	0.0126	0.7916	0.8096	0.7733	-2.6315	0.0859	-1.4213
1	158094726	158375751	-0.0228	0.0183	0.7943	0.8112	0.7760	-2.2167	0.3739	-1.2083
1	158385985	158743270	-0.0258	0.0079	0.7928	0.8133	0.7753	-2.4537	0.7506	-1.2615
1	158743369	158903174	-0.0252	0.0093	0.7929	0.8143	0.7733	-2.4241	0.9279	-1.4213
1	158903759	158911302	-0.0260	0.0072	0.7952	0.8158	0.7733	-2.0686	1.1937	-1.4213
1	158941270	159111113	-0.0258	0.0088	0.8002	0.8200	0.7767	-1.2983	1.9249	-1.1550
1	**159116681**	**159172854**	**-0.0270**	**0.0059**	**0.7983**	**0.8189**	**0.7760**	**-1.5946**	**1.7255**	**-1.2083**
1	159174683	159623952	-0.0251	0.0109	0.7993	0.8195	0.7760	-1.4465	1.8363	-1.2083

Supplementary Table S4: Summary of association results by meta-analysis result set and individual CAAPA dataset. The ALL rows refers to the meta-analysis result including all CAAPA datasets. The ALL EXCL GALA II rows refer to the meta-analysis result excluding the GALA II dataset. SNPs are identifed by their chr:hg19_position. SNPs with the following criteria are summarized: SNPs with p-values < 10-6 in either of the meta-analysis result sets that include or exclude GALA II, and SNPs that fall within an asthma candidate gene region and that pass a multiple testing significance threshold for that gene region in either of the two meta-analysis result sets. Two SNPs close to the PYHIN1 gene, reported by the EVE meta-analysis, are also summarized (1:158932555 and 1:158932907 respectively).

1:153688130	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			3.06E-06	Z=-4.667	C	T	T	0.012	0.010	0.013	-
ALL EXCL GALA II			5.09E-05	Z=-4.051	C	T	T	0.013	0.013	0.015	-
JHS(1)	0.6543	N	2.78E-01	OR=0.000 [95% CI 0.000-0.000]	C	T	T	0.014	0.000	0.016	-
JHS(2)	0.5167	N			C	T	T	0.004	0.005	0.004	?
GRAAD(2)	1.0054	Y	9.04E-01	OR=1.149 [95% CI 0.119-11.074]	C	T	T	0.028	0.031	0.022	?
SARP	0.7432	N	3.34E-01	OR=0.493 [95% CI 0.117-2.069]	C	T	T	0.026	0.023	0.044	-
CAG	0.8914	N	9.13E-01	Beta=0.0164	C	T	T	0.020	0.022	0.019	+
GRAAD(1)	0.6550	N	2.82E-01	Beta=-0.1473	C	T	T	0.012	0.009	0.016	-
SAGE	0.8177	N	2.92E-02	Beta=-0.1699	C	T	T	0.015	0.012	0.020	-
HUFS	0.6855	N	2.86E-01	Score=-2.199 Var=4.251	C	T	T	0.011	0.007	0.012	-
SAPPHIRE	0.6982	N	8.41E-03	OR=0.424 [95% CI 0.224-0.802]	C	T	T	0.016			-
BAGS	0.7129	N	1.23E-01	Freq1=0.011 Freq0=0.021	C	T	T	0.014	0.009	0.021	-
GALA II	0.7723	N	2.03E-02	Beta=-0.3570	C	T	T	0.004	0.001	0.006	-

1:153704977	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			1.50E-06	Z=4.811	G	A	G	0.012	0.010	0.013	+
ALL EXCL GALA II			3.61E-05	Z=4.131	G	A	G	0.014	0.013	0.015	+
JHS(1)	0.7017	N	2.50E-01	OR=0.000 [95% CI 0.000-0.000]	A	G	G	0.015		0.018	+
JHS(2)	0.5669	N			A	G	G	0.004	0.005	0.004	?
GRAAD(2)	1.0029	N	9.03E-01	OR=1.151 [95% CI 0.119-11.151]	A	G	G	0.028	0.031	0.022	?
SARP	0.8396	N	3.24E-01	OR=0.519 [95% CI 0.141-1.909]	A	G	G	0.026	0.023	0.044	+
CAG	0.9790	N	9.24E-01	Beta=0.0134	A	G	G	0.020	0.022	0.019	-
GRAAD(1)	0.8001	N	2.73E-01	Beta=-0.1331	A	G	G	0.012	0.009	0.016	+
SAGE	0.8466	N	3.13E-02	Beta=-0.1629	A	G	G	0.015	0.012	0.020	+
HUFS	0.7437	N	2.48E-01	Score=-2.488 Var=4.634	A	G	G	0.012	0.007	0.013	+
SAPPHIRE	0.7911	N	8.00E-03	OR=0.459 [95% CI 0.258-0.815]	A	G	G	0.017			+
BAGS	0.8384	N	1.05E-01	Freq1=0.011 Freq0=0.022	A	G	G	0.014	0.009	0.021	+
GALA II	0.7992	N	1.32E-02	Beta=-0.3731	A	G	G	0.004	0.001	0.006	+

1:153714954	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			1.36E-06	Z=4.83	G	C	G	0.013	0.011	0.015	+

	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL EXCL GALA II			4.05E-05	Z=4.105	G	C	G	0.015	0.013	0.016	+
JHS(1)	0.6695	N	2.93E-01	OR=0.001 [95% CI 0.000-354.012]	C	G	G	0.017	0.000	0.019	+
JHS(2)	0.6074	N	2.63E-01	OR=2.163 [95% CI 0.560-8.359]	C	G	G	0.006	0.015	0.005	-
GRAAD(2)	0.9676	N	9.23E-01	OR=0.901 [95% CI 0.107-7.603]	C	G	G	0.028	0.031	0.022	?
SARP	0.8586	N	3.28E-01	OR=0.533 [95% CI 0.151-1.882]	C	G	G	0.026	0.023	0.044	+
CAG	0.9861	N	9.40E-01	Beta=0.0104	C	G	G	0.020	0.022	0.019	-
GRAAD(1)	0.7774	N	2.63E-01	Beta=-0.1309	C	G	G	0.012	0.009	0.016	+
SAGE	0.8470	N	5.00E-02	Beta=-0.1366	C	G	G	0.017	0.014	0.022	+
HUFS	0.7001	N	1.72E-01	Score=-3.123 Var=5.238	C	G	G	0.014	0.007	0.015	+
SAPPHIRE	0.8025	N	1.25E-03	OR=0.407 [95% CI 0.236-0.702]	C	G	G	0.019			+
BAGS	0.7646	N	6.28E-02	Freq1=0.012 Freq0=0.027	C	G	G	0.014	0.009	0.021	+
GALA II	0.7612	N	9.95E-03	Beta=-0.3733	C	G	G	0.004	0.001	0.006	+

					ALLELES			MAF			METAL
1:153727808	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			1.18E-06	Z=-4.859	C	T	T	0.012	0.010	0.013	-
ALL EXCL GALA II			3.40E-05	Z=-4.145	C	T	T	0.014	0.013	0.015	-
JHS(1)	0.7320	N	2.88E-01	OR=0.000 [95% CI 0.000-0.000]	C	T	T	0.015	0.000	0.018	-
JHS(2)	0.6105	N			C	T	T	0.004	0.005	0.004	?
GRAAD(2)	0.9711	N	9.55E-01	OR=0.940 [95% CI 0.109-8.133]	C	T	T	0.028	0.031	0.022	?
SARP	0.8993	N	3.14E-01	OR=0.528 [95% CI 0.153-1.828]	C	T	T	0.026	0.023	0.044	-
CAG	0.9922	N	9.31E-01	Beta=0.0120	C	T	T	0.020	0.022	0.019	+
GRAAD(1)	0.8597	N	2.74E-01	Beta=-0.1290	C	T	T	0.012	0.009	0.016	-
SAGE	0.8627	N	3.23E-02	Beta=-0.1569	C	T	T	0.015	0.012	0.020	-
HUFS	0.7781	N	2.48E-01	Score=-2.547 Var=4.869	C	T	T	0.012	0.007	0.013	-
SAPPHIRE	0.8474	N	7.79E-03	OR=0.475 [95% CI 0.274-0.821]	C	T	T	0.018			-
BAGS	0.8783	N	8.81E-02	Freq1=0.010 Freq0=0.022	C	T	T	0.014	0.009	0.021	-
GALA II	0.8045	N	1.06E-02	Beta=-0.3746	C	T	T	0.004	0.001	0.006	-

					ALLELES			MAF			METAL
1:158056779	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			1.23E-05	Z=4.373	T	A	A	0.267	0.284	0.253	+
ALL EXCL GALA II			3.98E-06	Z=4.612	T	A	A	0.204	0.216	0.194	+
JHS(1)	0.8966	N	2.38E-01	OR=1.421 [95% CI 0.793-2.548]	T	A	A	0.213	0.239	0.208	+
JHS(2)	0.8934	N	8.01E-01	OR=1.058 [95% CI 0.683-1.639]	T	A	A	0.177	0.163	0.179	+
GRAAD(2)	0.9662	N	9.23E-01	OR=0.954 [95% CI 0.367-2.476]	T	A	A	0.182	0.169	0.217	-
SARP	0.8875	N	8.50E-02	OR=1.872 [95% CI 0.917-3.823]	T	A	A	0.187	0.195	0.122	+
CAG	0.9340	N	1.62E-01	Beta=0.0771	T	A	A	0.189	0.215	0.170	+
GRAAD(1)	0.8897	N	1.47E-01	Beta=0.0499	T	A	A	0.197	0.212	0.182	+
SAGE	0.9196	N	8.28E-02	Beta=0.0385	T	A	A	0.211	0.220	0.196	+
HUFS	0.9047	N	2.11E-02	Score=18.820 Var=66.560	T	A	A	0.212	0.248	0.206	+
SAPPHIRE	0.8321	N	1.33E-02	OR=1.308 [95% CI 1.058-1.617]	T	A	A	0.222			+

| BAGS | 0.9092 | N | 2.33E-01 | Freq1=0.191 Freq0=0.180 | T | A | A | 0.181 | 0.191 | 0.183 | + |
| GALA II | 0.9317 | N | 5.52E-01 | Beta=0.0103 | T | A | T | 0.467 | 0.464 | 0.470 | + |

1:158932555	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			1.42E-03	Z=3.19	C	T	C	0.213	0.202	0.222	+
ALL EXCL GALA II			2.58E-03	Z=3.014	C	T	C	0.248	0.239	0.255	+
JHS(1)	0.8984	N	2.06E-02	OR=2.150 [95% CI 1.125-4.111]	C	T	C	0.246	0.159	0.260	+
JHS(2)	0.8961	N	5.24E-01	OR=0.894 [95% CI 0.634-1.262]	C	T	C	0.268	0.287	0.265	-
GRAAD(2)	1.0057	Y	8.89E-01	OR=0.948 [95% CI 0.443-2.028]	C	T	C	0.312	0.331	0.261	-
SARP	0.7800	N	6.07E-01	OR=0.856 [95% CI 0.472-1.550]	C	T	C	0.261	0.267	0.233	-
CAG	0.8308	N	7.04E-01	Beta=0.0204	C	T	C	0.250	0.237	0.260	+
GRAAD(1)	0.7520	N	6.72E-03	Beta=0.0914	C	T	C	0.250	0.220	0.281	+
SAGE	0.7784	N	8.52E-02	Beta=0.0384	C	T	C	0.247	0.237	0.263	+
HUFS	0.8934	N	1.85E-01	Score=11.867 Var=80.169	C	T	C	0.232	0.210	0.236	+
SAPPHIRE	0.7091	N	6.27E-01	OR=1.049 [95% CI 0.864-1.273]	C	T	C	0.268			+
BAGS	0.7455	N	3.60E-01	Freq1=0.752 Freq0=0.717	C	T	C	0.250	0.222	0.271	+
GALA II	0.5833	N	2.64E-01	Beta=0.0432	C	T	C	0.066	0.064	0.069	+

1:158932907	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			1.41E-03	Z=3.192	G	A	G	0.213	0.201	0.222	+
ALL EXCL GALA II			2.58E-03	Z=3.014	G	A	G	0.248	0.239	0.255	+
JHS(1)	0.8984	N	2.02E-02	OR=2.156 [95% CI 1.127-4.125]	G	A	G	0.246	0.159	0.260	+
JHS(2)	0.8961	N	5.28E-01	OR=0.895 [95% CI 0.634-1.263]	G	A	G	0.268	0.287	0.265	-
GRAAD(2)	1.0055	N	8.89E-01	OR=0.948 [95% CI 0.443-2.028]	G	A	G	0.312	0.331	0.261	-
SARP	0.7826	N	6.07E-01	OR=0.856 [95% CI 0.473-1.549]	G	A	G	0.261	0.267	0.233	-
CAG	0.8327	N	7.07E-01	Beta=0.0202	G	A	G	0.250	0.237	0.260	+
GRAAD(1)	0.7545	N	6.72E-03	Beta=0.0912	G	A	G	0.250	0.220	0.281	+
SAGE	0.7804	N	8.46E-02	Beta=0.0384	G	A	G	0.248	0.236	0.265	+
HUFS	0.8932	N	1.85E-01	Score=11.856 Var=80.125	G	A	G	0.232	0.210	0.236	+
SAPPHIRE	0.7110	N	6.33E-01	OR=1.048 [95% CI 0.864-1.272]	G	A	G	0.268			+
BAGS	0.7480	N	3.59E-01	Freq1=0.752 Freq0=0.717	G	A	G	0.250	0.222	0.271	+
GALA II	0.5854	N	2.62E-01	Beta=0.0434	G	A	G	0.066	0.064	0.069	+

1:244532963	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			8.91E-07	Z=-4.914	C	T	C	0.232	0.250	0.216	-
ALL EXCL GALA II			1.03E-04	Z=-3.884	C	T	C	0.195	0.206	0.185	-
JHS(1)	0.77404	N	9.08E-01	OR=0.961 [95% CI 0.490-1.886]	T	C	C	0.185	0.159	0.189	+
JHS(2)	0.7821	N	9.68E-02	OR=1.403 [95% CI 0.941-2.092]	T	C	C	0.186	0.223	0.179	-
GRAAD(2)	0.99973	N	3.52E-01	OR=1.584 [95% CI 0.601-4.177]	T	C	C	0.210	0.231	0.152	-

	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
SARP	0.964	N	6.20E-01	OR=0.870 [95% CI 0.501-1.509]	T	C	C	0.198	0.194	0.211	+
CAG	0.97039	N	1.48E-01	Beta=-0.0779	T	C	C	0.194	0.167	0.215	+
GRAAD(1)	0.96398	N	1.60E-01	Beta=0.0462	T	C	C	0.196	0.207	0.184	-
SAGE	0.97256	N	3.85E-02	Beta=0.0437	T	C	C	0.202	0.213	0.187	-
HUFS	0.79563	N	1.44E-01	Score=11.300 Var=59.866	T	C	C	0.193	0.211	0.189	-
SAPPHIRE	0.96571	N	3.16E-03	OR=1.346 [95% CI 1.105-1.639]	T	C	C	0.207			-
BAGS	0.94938	N	3.78E-01	Freq1=0.170 Freq0=0.156	T	C	C	0.162	0.170	0.158	-
GALA II	0.97808	N	1.42E-03	Beta=0.0554	T	C	C	0.388	0.413	0.362	-

					ALLELES			MAF			METAL
2:127422031	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			5.50E-06	Z=-4.545	G	A	G	0.008	0.012	0.005	-
ALL EXCL GALA II			1.12E-02	Z=-2.538	G	A	G	0.003	0.004	0.002	-
JHS(1)	0.5575	N			A	G	G	0.005	0.000	0.005	?
JHS(2)	0.4194	N			A	G	G	0.002	0.000	0.003	?
GRAAD(2)	0.5603	N			A	G	G	0.011	0.008	0.022	?
SARP	0.4328	N			A	G	G	0.001	0.002	0.000	?
CAG	0.2425	N			A	G	G	0.000	0.000	0.000	?
GRAAD(1)	0.3748	N			A	G	G	0.003	0.005	0.000	?
SAGE	0.5437	N	2.97E-01	Beta=0.1587	A	G	G	0.003	0.004	0.001	-
HUFS	0.6058	N	5.44E-03	Score=3.529 Var=1.612	A	G	G	0.004	0.010	0.002	-
SAPPHIRE	0.3772	N	7.18E-01	OR=1.403 [95% CI 0.224-8.766]	A	G	G	0.005			?
BAGS	0.2209	N	1.71E-02	Freq1=0.001 Freq0=0.002	A	G	G	0.001	0.000	0.001	?
GALA II	0.8027	N	4.46E-05	Beta=0.2179	A	G	G	0.032	0.043	0.020	-

					ALLELES			MAF			METAL
2:144638809	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			2.97E-08	Z=5.543	G	A	A	0.129	0.142	0.118	+
ALL EXCL GALA II			2.09E-07	Z=5.191	G	A	A	0.122	0.134	0.112	+
JHS(1)	0.8916	N	5.12E-01	OR=1.255 [95% CI 0.637-2.474]	G	A	A	0.113	0.125	0.111	+
JHS(2)	0.9328	N	1.14E-01	OR=1.448 [95% CI 0.914-2.291]	G	A	A	0.122	0.149	0.117	+
GRAAD(2)	1.0057	Y	6.24E-01	OR=1.349 [95% CI 0.408-4.462]	G	A	A	0.114	0.123	0.087	+
SARP	1.0013	Y	3.53E-01	OR=0.732 [95% CI 0.378-1.415]	G	A	A	0.116	0.113	0.144	-
CAG	1.0018	Y	9.68E-01	Beta=-0.0027	G	A	A	0.109	0.110	0.109	-
GRAAD(1)	0.9991	Y	2.37E-01	Beta=0.0466	G	A	A	0.122	0.130	0.113	+
SAGE	0.9991	Y	8.17E-03	Beta=0.0657	G	A	A	0.128	0.141	0.109	+
HUFS	0.9371	N	3.19E-04	Score=25.091 Var=48.607	G	A	A	0.123	0.170	0.113	+
SAPPHIRE	0.9979	Y	2.67E-02	OR=1.311 [95% CI 1.032-1.665]	G	A	A	0.121			+
BAGS	0.9989	Y	4.24E-02	Freq1=0.135 Freq0=0.101	G	A	A	0.115	0.133	0.099	+
GALA II	1.0001	Y	4.27E-02	Beta=0.0468	G	A	A	0.158	0.169	0.146	+

					ALLELES			MAF			METAL

2:144650140	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			6.51E-09	Z=-5.803	T	A	T	0.153	0.169	0.139	-
ALL EXCL GALA II			1.37E-07	Z=-5.269	T	A	T	0.141	0.156	0.129	-
JHS(1)	0.9039	N	3.69E-01	OR=0.752 [95% CI 0.404-1.400]	T	A	T	0.133	0.159	0.129	-
JHS(2)	0.9275	N	1.12E-01	OR=0.700 [95% CI 0.452-1.086]	T	A	T	0.138	0.163	0.134	-
GRAAD(2)	1.0002	N	9.74E-01	OR=0.982 [95% CI 0.328-2.941]	T	A	T	0.131	0.131	0.130	-
SARP	0.9665	N	4.68E-01	OR=1.271 [95% CI 0.666-2.425]	T	A	T	0.131	0.127	0.156	+
CAG	0.9744	N	7.91E-01	Beta=0.0169	T	A	T	0.126	0.123	0.128	+
GRAAD(1)	0.9918	N	1.28E-01	Beta=-0.0546	T	A	T	0.145	0.158	0.131	-
SAGE	0.9717	N	6.61E-04	Beta=-0.0815	T	A	T	0.149	0.166	0.122	-
HUFS	0.9399	N	2.37E-04	Score=-26.889 Var=53.509	T	A	T	0.142	0.195	0.132	-
SAPPHIRE	0.9627	N	1.50E-01	OR=0.850 [95% CI 0.681-1.060]	T	A	T	0.143			-
BAGS	0.9831	N	6.10E-02	Freq1=0.850 Freq0=0.884	T	A	T	0.130	0.147	0.115	-
GALA II	0.9786	N	1.46E-02	Beta=-0.0522	T	A	T	0.202	0.217	0.186	-

2:144655836	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			5.26E-07	Z=5.016	G	A	A	0.240	0.253	0.227	+
ALL EXCL GALA II			1.22E-06	Z=4.853	G	A	A	0.238	0.252	0.225	+
JHS(1)	0.8932	N	5.70E-01	OR=1.176 [95% CI 0.673-2.055]	A	G	A	0.235	0.216	0.239	-
JHS(2)	0.9364	N	1.65E-01	OR=0.771 [95% CI 0.534-1.113]	A	G	A	0.239	0.272	0.233	+
GRAAD(2)	1.0057	Y	2.27E-01	OR=1.674 [95% CI 0.725-3.865]	A	G	A	0.244	0.231	0.283	-
SARP	1.0014	Y	3.95E-01	OR=0.779 [95% CI 0.438-1.385]	A	G	A	0.223	0.228	0.189	+
CAG	1.0019	Y	6.65E-01	Beta=0.0217	A	G	A	0.244	0.237	0.250	-
GRAAD(1)	1.0006	Y	9.73E-01	Beta=-0.0010	A	G	A	0.234	0.235	0.232	+
SAGE	1.0001	Y	3.30E-05	Beta=-0.0797	A	G	A	0.245	0.270	0.207	+
HUFS	0.9397	N	8.25E-04	Score=-30.082 Var=80.916	A	G	A	0.232	0.290	0.221	+
SAPPHIRE	0.9993	Y	2.34E-02	OR=0.809 [95% CI 0.673-0.972]	A	G	A	0.225			+
BAGS	1.0007	Y	5.10E-01	Freq1=0.743 Freq0=0.762	A	G	A	0.250	0.254	0.238	+
GALA II	1.0003	Y	1.25E-01	Beta=-0.0299	A	G	A	0.247	0.257	0.236	+

2:207007157	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			1.86E-06	Z=-4.768	G	T	G	0.002	0.003	0.002	-
ALL EXCL GALA II			1.86E-06	Z=-4.768	G	T	G	0.002	0.003	0.001	-
JHS(1)	0.3752	N			T	G	G	0.002	0.000	0.002	?
JHS(2)	0.3820	N			T	G	G	0.002	0.000	0.002	?
GRAAD(2)	0.8332	N			T	G	G	0.006	0.000	0.022	?
SARP	0.0102	N			T	G	G	0.000	0.000	0.000	?
CAG	0.8330	N	2.57E-01	Beta=0.3386	T	G	G	0.006	0.009	0.003	?
GRAAD(1)	0.4706	N			T	G	G	0.001	0.003	0.000	?
SAGE	0.1441	N			T	G	G	0.000	0.000	0.001	?

	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
HUFS	0.5461	N	3.73E-05	Score=3.732 Var=0.819	T	G	G	0.003	0.012	0.001	-
SAPPHIRE	0.2485	N	2.46E-01	OR=6.070 [95% CI 0.288-128.031]	T	G	G	0.005			?
BAGS	0.7098	N	1.37E-02	Freq1=0.009 Freq0=0.002	T	G	G	0.004	0.007	0.001	-
GALA II	0.4798	N			T	G	G	0.004	0.006	0.002	?

					ALLELES			MAF			METAL
2:240496929	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			6.84E-06	Z=4.499	G	A	A	0.007	0.009	0.005	+
ALL EXCL GALA II			3.26E-06	Z=4.654	G	A	A	0.005	0.007	0.003	+
JHS(1)	0.5889	N			G	A	A	0.003	0.000	0.004	?
JHS(2)	0.4363	N			G	A	A	0.002	0.000	0.003	?
GRAAD(2)	0.8139	N	6.11E-01	OR=0.584 [95% CI 0.073-4.643]	G	A	A	0.034	0.031	0.043	?
SARP	0.6020	N			G	A	A	0.004	0.003	0.011	?
CAG	0.1744	N			G	A	A	0.000	0.000	0.000	?
GRAAD(1)	0.7293	N	1.52E-02	Beta=0.4655	G	A	A	0.005	0.009	0.001	+
SAGE	0.6840	N	4.01E-04	Beta=0.3949	G	A	A	0.007	0.011	0.001	+
HUFS	0.7310	N	5.31E-02	Score=3.361 Var=3.020	G	A	A	0.006	0.013	0.005	+
SAPPHIRE	0.3464	N	6.54E-01	OR=1.540 [95% CI 0.234-10.146]	G	A	A	0.006			?
BAGS	0.7623	N	9.00E-01	Freq1=0.003 Freq0=0.002	G	A	A	0.001	0.002	0.001	?
GALA II	0.7302	N	2.20E-01	Beta=0.0920	G	A	A	0.014	0.017	0.011	+

					ALLELES			MAF			METAL
3:170768649	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			2.64E-06	Z=-4.697	C	T	T	0.004	0.002	0.005	-
ALL EXCL GALA II			2.64E-06	Z=-4.697	C	T	T	0.004	0.002	0.006	-
JHS(1)	0.4736	N			C	T	T	0.002	0.011	0.000	?
JHS(2)	0.5140	N			C	T	T	0.007	0.010	0.006	?
GRAAD(2)	0.4836	N			C	T	T	0.006	0.008	0.000	?
SARP	0.4937	N			C	T	T	0.006	0.005	0.011	?
CAG	0.6536	N	2.23E-02	Beta=-0.5541	C	T	T	0.009	0.000	0.016	-
GRAAD(1)	0.5520	N	1.26E-02	Beta=-0.4220	C	T	T	0.008	0.003	0.013	-
SAGE	0.2483	N			C	T	T	0.000	0.000	0.000	?
HUFS	0.5206	N	1.50E-01	Score=-1.842 Var=1.638	C	T	T	0.006	0.003	0.007	-
SAPPHIRE	0.6208	N	1.02E-03	OR=0.183 [95% CI 0.067-0.503]	C	T	T	0.008			-
BAGS	0.5805	N	2.37E-94	Freq1=0.003 Freq0=0.008	C	T	T	0.007	0.002	0.009	?
GALA II	0.2060	N			C	T	T	0.000	0.000	0.000	?

					ALLELES			MAF			METAL
5:94736141	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			3.51E-04	Z=3.575	C	T	T	0.087	0.093	0.081	+
ALL EXCL GALA II			4.28E-06	Z=4.597	C	T	T	0.096	0.106	0.087	+
JHS(1)	0.8081	N	1.29E-01	OR=1.715 [95% CI 0.855-3.441]	C	T	T	0.109	0.148	0.102	+

	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
JHS(2)	0.8078	N	8.38E-01	OR=0.947 [95% CI 0.565-1.589]	C	T	T	0.101	0.104	0.101	-
GRAAD(2)	0.8111	N	2.25E-01	OR=2.602 [95% CI 0.556-12.171]	C	T	T	0.119	0.131	0.087	+
SARP	0.6966	N	9.37E-01	OR=0.965 [95% CI 0.394-2.362]	C	T	T	0.076	0.076	0.078	-
CAG	0.8043	N	1.06E-01	Beta=0.1171	C	T	T	0.111	0.132	0.096	+
GRAAD(1)	0.7858	N	4.79E-03	Beta=0.1279	C	T	T	0.101	0.117	0.084	+
SAGE	0.8555	N	3.63E-03	Beta=0.0909	C	T	T	0.090	0.101	0.073	+
HUFS	0.7665	N	1.21E-01	Score=8.954 Var=33.385	C	T	T	0.087	0.106	0.083	+
SAPPHIRE	0.8142	N	2.60E-01	OR=1.167 [95% CI 0.892-1.527]	C	T	T	0.107			+
BAGS	0.8197	N	1.81E-01	Freq1=0.129 Freq0=0.100	C	T	T	0.112	0.124	0.095	+
GALA II	0.8234	N	2.63E-01	Beta=-0.0473	C	T	T	0.050	0.046	0.053	-

					ALLELES			MAF			METAL
7:87275641	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			2.03E-04	Z=-3.715	G	A	A	0.038	0.033	0.041	-
ALL EXCL GALA II			6.09E-06	Z=-4.523	G	A	A	0.036	0.030	0.041	-
JHS(1)	0.59415	N	8.52E-01	OR=0.869 [95% CI 0.200-3.777]	A	G	A	0.035	0.034	0.035	+
JHS(2)	0.59776	N	6.18E-01	OR=1.357 [95% CI 0.410-4.490]	A	G	A	0.027	0.020	0.028	-
GRAAD(2)	1.00526	Y	8.22E-01	OR=1.242 [95% CI 0.189-8.172]	A	G	A	0.034	0.031	0.043	?
SARP	0.86034	N	3.37E-01	OR=1.662 [95% CI 0.589-4.696]	A	G	A	0.046	0.043	0.056	-
CAG	1.00155	Y	4.12E-01	Beta=0.1005	A	G	A	0.033	0.026	0.038	-
GRAAD(1)	1.00017	Y	5.96E-01	Beta=0.0338	A	G	A	0.043	0.040	0.045	-
SAGE	0.63161	N	1.08E-03	Beta=0.1598	A	G	A	0.042	0.032	0.056	-
HUFS	0.61343	N	6.83E-02	Score=6.382 Var=12.256	A	G	A	0.035	0.023	0.037	-
SAPPHIRE	0.22068	N	7.38E-01	OR=0.880 [95% CI 0.418-1.853]	A	G	A	0.052			?
BAGS	1.00026	Y	1.06E-03	Freq1=0.984 Freq0=0.959	A	G	A	0.034	0.021	0.041	-
GALA II	0.31982	N	7.35E-01	Beta=-0.0191	A	G	A	0.045	0.046	0.044	+

					ALLELES			MAF			METAL
8:27338294	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			3.84E-05	Z=-4.117	C	T	T	0.199	0.190	0.207	-
ALL EXCL GALA II			2.31E-06	Z=-4.724	C	T	T	0.212	0.202	0.221	-
JHS(1)	0.9623	N	1.33E-01	OR=0.614 [95% CI 0.324-1.161]	C	T	T	0.209	0.159	0.217	-
JHS(2)	0.9527	N	1.48E-01	OR=1.313 [95% CI 0.908-1.897]	C	T	T	0.208	0.248	0.201	+
GRAAD(2)	1.0001	Y	1.85E-01	OR=2.121 [95% CI 0.698-6.443]	C	T	T	0.170	0.192	0.109	+
SARP	0.9543	N	2.27E-03	OR=0.467 [95% CI 0.286-0.761]	C	T	T	0.244	0.225	0.378	-
CAG	0.9692	N	2.25E-01	Beta=-0.0618	C	T	T	0.241	0.215	0.260	-
GRAAD(1)	0.9613	N	7.54E-02	Beta=-0.0559	C	T	T	0.218	0.199	0.238	-
SAGE	0.9666	N	1.45E-03	Beta=-0.0656	C	T	T	0.213	0.195	0.240	-
HUFS	0.9613	N	3.81E-02	Score=-18.109 Var=76.224	C	T	T	0.204	0.172	0.210	-
SAPPHIRE	0.9530	N	1.35E-01	OR=0.873 [95% CI 0.731-1.043]	C	T	T	0.233			-
BAGS	0.9615	N	9.93E-02	Freq1=0.172 Freq0=0.216	C	T	T	0.194	0.172	0.213	-
GALA II	0.9020	N	8.56E-01	Beta=0.0046	C	T	T	0.145	0.147	0.142	+

8:108979259	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			3.98E-07	Z=-5.07	C	A	A	0.009	0.007	0.011	-
ALL EXCL GALA II			1.52E-08	Z=-5.66	C	A	A	0.010	0.008	0.012	-
JHS(1)	0.7481	N			C	A	A	0.006	0.011	0.005	?
JHS(2)	0.6978	N			C	A	A	0.005	0.010	0.005	?
GRAAD(2)	0.6349	N	6.80E-01	OR=2.095 [95% CI 0.062-70.416]	C	A	A	0.011	0.015	0.000	?
SARP	0.7579	N			C	A	A	0.010	0.012	0.000	?
CAG	0.8803	N	4.80E-01	Beta=-0.1238	C	A	A	0.017	0.013	0.019	-
GRAAD(1)	0.7900	N	1.53E-01	Beta=-0.2068	C	A	A	0.008	0.006	0.010	-
SAGE	0.8572	N	4.07E-04	Beta=-0.2859	C	A	A	0.012	0.007	0.020	-
HUFS	0.8176	N	3.86E-01	Score=-1.840 Var=4.499	C	A	A	0.011	0.008	0.012	-
SAPPHIRE	0.6521	N	2.20E-03	OR=0.300 [95% CI 0.139-0.648]	C	A	A	0.012			-
BAGS	0.7403	N	2.61E-04	Freq1=0.002 Freq0=0.019	C	A	A	0.011	0.002	0.018	-
GALA II	0.8484	N	8.33E-01	Beta=-0.0314	C	A	A	0.003	0.003	0.004	-

9:89759848	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			1.77E-03	Z=3.126	G	A	G	0.021	0.019	0.023	+
ALL EXCL GALA II			3.46E-06	Z=4.641	G	A	G	0.023	0.019	0.025	+
JHS(1)	0.4040	N	2.57E-01	OR=0.175 [95% CI 0.009-3.570]	A	G	G	0.015	0.000	0.018	+
JHS(2)	0.4564	N	1.60E-01	OR=0.283 [95% CI 0.048-1.650]	A	G	G	0.018	0.010	0.019	+
GRAAD(2)	0.5271	N	2.45E-01	OR=0.230 [95% CI 0.019-2.742]	A	G	G	0.017	0.015	0.022	?
SARP	0.5245	N	7.47E-01	OR=0.778 [95% CI 0.169-3.579]	A	G	G	0.026	0.025	0.033	+
CAG	0.6469	N	8.94E-01	Beta=-0.0194	A	G	G	0.026	0.022	0.029	+
GRAAD(1)	0.5251	N	5.42E-01	Beta=-0.0573	A	G	G	0.026	0.025	0.027	+
SAGE	0.5014	N	4.47E-03	Beta=-0.1683	A	G	G	0.024	0.018	0.033	+
HUFS	0.5380	N	1.42E-01	Score=-4.383 Var=8.927	A	G	G	0.021	0.012	0.023	+
SAPPHIRE	0.5254	N	8.23E-04	OR=0.405 [95% CI 0.238-0.687]	A	G	G	0.033			+
BAGS	0.5957	N	4.82E-01	Freq1=0.043 Freq0=0.049	A	G	G	0.032	0.027	0.033	+
GALA II	0.4233	N	3.02E-02	Beta=0.1854	A	G	G	0.015	0.020	0.010	-

9:119020314	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
ALL			7.11E-06	Z=4.49	G	A	G	0.002	0.001	0.003	+
ALL EXCL GALA II			6.17E-03	Z=2.739	G	A	G	0.001	0.001	0.001	+
JHS(1)	0.0905	N			A	G	G	0.000	0.000	0.000	?
JHS(2)	0.5616	N			A	G	G	0.002	0.010	0.000	?
GRAAD(2)	0.0153	N			A	G	G	0.000	0.000	0.000	?
SARP	0.0132	N			A	G	G	0.000	0.000	0.000	?
CAG	0.6229	N	7.36E-01	Beta=-0.1459	A	G	G	0.004	0.004	0.003	?

	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
GRAAD(1)	0.0783	N			A	G	G	0.000	0.000	0.000	?
SAGE	0.5121	N	6.17E-03	Beta=-0.6677	A	G	G	0.002	0.000	0.004	+
HUFS	0.0684	N	6.09E-01	Score=0.063 Var=0.015	A	G	G	0.000	0.000	0.000	?
SAPPHIRE	0.0767	N	3.15E-01	OR=345.048 [95% CI 0.004-30891241.2	A	G	G	0.001			?
BAGS	0.6617	N	8.44E-01	Freq1=0.003 Freq0=0.002	A	G	G	0.001	0.002	0.001	?
GALA II	0.8361	N	3.21E-04	Beta=-0.3973	A	G	G	0.007	0.002	0.011	+

					ALLELES			MAF			METAL
10:105730220	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			9.52E-07	Z=4.901	G	A	A	0.029	0.034	0.025	+
ALL EXCL GALA II			4.00E-05	Z=4.107	G	A	A	0.033	0.038	0.028	+
JHS(1)	0.7736	N	7.94E-01	OR=1.217 [95% CI 0.279-5.302]	G	A	A	0.028	0.034	0.027	+
JHS(2)	0.7874	N	6.58E-02	OR=2.144 [95% CI 0.951-4.831]	G	A	A	0.031	0.054	0.027	+
GRAAD(2)	0.6992	N	5.76E-01	OR=4.227 [95% CI 0.027-659.953]	G	A	A	0.011	0.015	0.000	?
SARP	0.7281	N	2.77E-01	OR=2.666 [95% CI 0.455-15.637]	G	A	A	0.037	0.040	0.022	+
CAG	0.8136	N	9.55E-01	Beta=0.0073	G	A	A	0.030	0.031	0.029	+
GRAAD(1)	0.7400	N	3.78E-01	Beta=0.0677	G	A	A	0.035	0.038	0.031	+
SAGE	0.7718	N	1.29E-02	Beta=0.1241	G	A	A	0.036	0.041	0.029	+
HUFS	0.8045	N	2.03E-02	Score=8.165 Var=12.368	G	A	A	0.032	0.048	0.029	+
SAPPHIRE	0.7228	N	5.52E-02	OR=1.673 [95% CI 0.989-2.828]	G	A	A	0.033			+
BAGS	0.7198	N	8.36E-01	Freq1=0.030 Freq0=0.030	G	A	A	0.028	0.023	0.028	+
GALA II	0.7562	N	6.39E-03	Beta=0.2285	G	A	A	0.013	0.017	0.008	+

					ALLELES			MAF			METAL
12:111196204	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			8.83E-07	Z=4.916	C	T	C	0.160	0.148	0.171	+
ALL EXCL GALA II			7.74E-05	Z=3.952	C	T	C	0.175	0.164	0.184	+
JHS(1)	0.9939	N	7.20E-01	OR=1.117 [95% CI 0.611-2.043]	T	C	C	0.167	0.182	0.164	-
JHS(2)	0.9902	N	6.15E-02	OR=0.661 [95% CI 0.429-1.020]	T	C	C	0.185	0.139	0.194	+
GRAAD(2)	1.0033	N	6.42E-01	OR=0.791 [95% CI 0.295-2.121]	T	C	C	0.159	0.154	0.174	+
SARP	0.9897	N	1.56E-01	OR=0.654 [95% CI 0.364-1.176]	T	C	C	0.149	0.142	0.200	+
CAG	1.0006	N	2.94E-01	Beta=-0.0546	T	C	C	0.215	0.193	0.231	+
GRAAD(1)	0.9639	N	1.31E-01	Beta=-0.0508	T	C	C	0.168	0.152	0.184	+
SAGE	0.9146	N	2.28E-01	Beta=-0.0279	T	C	C	0.175	0.170	0.183	+
HUFS	0.9941	N	7.60E-04	Score=-27.019 Var=64.399	T	C	C	0.167	0.119	0.176	+
SAPPHIRE	0.9441	N	3.34E-01	OR=0.906 [95% CI 0.741-1.107]	T	C	C	0.170			+
BAGS	0.9775	N	5.16E-01	Freq1=0.185 Freq0=0.194	T	C	C	0.188	0.186	0.195	+
GALA II	0.8671	N	2.19E-03	Beta=-0.0906	T	C	C	0.097	0.085	0.110	+

					ALLELES			MAF			METAL
17:26241077	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			9.60E-05	Z=-3.901	C	A	C	0.044	0.050	0.038	-

	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL EXCL GALA II			5.09E-06	Z=-4.561	C	A	C	0.032	0.039	0.025	-
JHS(1)	0.6188	N	7.74E-01	OR=0.776 [95% CI 0.137-4.392]	A	C	C	0.026	0.023	0.027	+
JHS(2)	0.6732	N	1.38E-02	OR=2.668 [95% CI 1.222-5.824]	A	C	C	0.029	0.050	0.026	-
GRAAD(2)	0.4319	N	9.49E-01	OR=1.136 [95% CI 0.022-58.062]	A	C	C	0.011	0.008	0.022	?
SARP	0.6966	N	2.95E-01	OR=2.948 [95% CI 0.390-22.289]	A	C	C	0.036	0.035	0.022	-
CAG	0.7019	N	5.91E-01	Beta=0.0750	A	C	C	0.035	0.044	0.029	-
GRAAD(1)	0.7641	N	6.10E-01	Beta=0.0384	A	C	C	0.038	0.043	0.032	-
SAGE	0.7769	N	8.01E-04	Beta=0.1676	A	C	C	0.034	0.043	0.021	-
HUFS	0.7000	N	6.69E-01	Score=1.414 Var=10.916	A	C	C	0.028	0.028	0.028	-
SAPPHIRE	0.6592	N	2.19E-03	OR=2.498 [95% CI 1.392-4.483]	A	C	C	0.042			-
BAGS	0.7632	N	4.27E-01	Freq1=0.021 Freq0=0.010	A	C	C	0.016	0.020	0.006	-
GALA II	0.7954	N	7.44E-01	Beta=-0.0101	A	C	C	0.096	0.092	0.101	+

					ALLELES			MAF			METAL
17:37886986	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			4.54E-06	Z=-4.585	C	A	A	0.096	0.088	0.104	-
ALL EXCL GALA II			2.17E-03	Z=-3.065	C	A	A	0.068	0.060	0.075	-
JHS(1)	0.6386	N	8.91E-01	OR=1.078 [95% CI 0.370-3.139]	C	A	A	0.080	0.068	0.081	+
JHS(2)	0.6213	N	2.46E-01	OR=0.628 [95% CI 0.286-1.379]	C	A	A	0.073	0.045	0.079	-
GRAAD(2)	0.8680	N	5.37E-01	OR=1.637 [95% CI 0.343-7.810]	C	A	A	0.068	0.077	0.043	+
SARP	0.5221	N	4.06E-01	OR=1.858 [95% CI 0.431-8.017]	C	A	A	0.059	0.060	0.044	+
CAG	0.8343	N	5.96E-02	Beta=-0.1808	C	A	A	0.054	0.035	0.067	-
GRAAD(1)	0.6308	N	8.45E-01	Beta=-0.0122	C	A	A	0.068	0.068	0.068	-
SAGE	0.8958	N	5.72E-02	Beta=-0.0644	C	A	A	0.074	0.065	0.086	-
HUFS	0.6484	N	4.60E-02	Score=-9.427 Var=22.324	C	A	A	0.071	0.051	0.074	-
SAPPHIRE	0.8178	N	5.20E-01	OR=0.904 [95% CI 0.666-1.228]	C	A	A	0.071			-
BAGS	0.6319	N	8.51E-02	Freq1=0.051 Freq0=0.068	C	A	A	0.053	0.035	0.058	-
GALA II	0.8924	N	4.87E-05	Beta=-0.0877	C	A	A	0.215	0.190	0.241	-

					ALLELES			MAF			METAL
17:38025417	RSQ	GENOTYPED	P	EFFECT	REF	EFFECT	MINOR	ALL	CASES	CONTROLS	DIRECTION
ALL			1.35E-06	Z=-4.832	G	A	A	0.102	0.092	0.111	-
ALL EXCL GALA II			9.77E-04	Z=-3.297	G	A	A	0.072	0.063	0.079	-
JHS(1)	1.0009	Y	9.10E-01	OR=0.949 [95% CI 0.380-2.370]	G	A	A	0.078	0.068	0.080	-
JHS(2)	1.0005	Y	1.80E-01	OR=0.636 [95% CI 0.328-1.233]	G	A	A	0.079	0.050	0.084	-
GRAAD(2)	1.0055	Y	3.27E-01	OR=2.169 [95% CI 0.462-10.187]	G	A	A	0.080	0.092	0.043	+
SARP	1.0009	Y	4.09E-01	OR=1.587 [95% CI 0.530-4.754]	G	A	A	0.063	0.065	0.044	+
CAG	1.0017	Y	4.74E-02	Beta=-0.1755	G	A	A	0.059	0.035	0.077	-
GRAAD(1)	1.0002	Y	8.93E-01	Beta=-0.0071	G	A	A	0.065	0.066	0.065	-
SAGE	0.9997	Y	9.48E-02	Beta=-0.0527	G	A	A	0.081	0.073	0.092	-
HUFS	0.9973	Y	2.31E-02	Score=-13.270 Var=34.123	G	A	A	0.078	0.053	0.082	-
SAPPHIRE	0.9267	N	3.79E-01	OR=0.881 [95% CI 0.666-1.167]	G	A	A	0.075			-

	RSQ	GENOTYPED	P	EFFECT	ALLELES REF	EFFECT	MINOR	MAF ALL	CASES	CONTROLS	METAL DIRECTION
BAGS	1.0004	Y	4.74E-02	Freq1=0.038 Freq0=0.058	G	A	A	0.053	0.037	0.056	-
GALA II	1.0001	Y	3.30E-05	Beta=-0.0848	G	A	A	0.229	0.199	0.260	-
18:31058259	**RSQ**	**GENOTYPED**	**P**	**EFFECT**	**REF**	**EFFECT**	**MINOR**	**ALL**	**CASES**	**CONTROLS**	**DIRECTION**
ALL			3.70E-06	Z=-4.628	C	T	T	0.005	0.003	0.006	-
ALL EXCL GALA II			5.12E-04	Z=-3.475	C	T	T	0.002	0.002	0.003	-
JHS(1)	0.2914	N			C	T	T	0.002	0.000	0.002	?
JHS(2)	0.3936	N			C	T	T	0.002	0.000	0.003	?
GRAAD(2)	0.1790	N			C	T	T	0.000	0.000	0.000	?
SARP	0.4895	N			C	T	T	0.004	0.003	0.011	?
CAG	0.4789	N			C	T	T	0.002	0.004	0.000	?
GRAAD(1)	0.2890	N			C	T	T	0.001	0.001	0.001	?
SAGE	0.5570	N	2.50E-02	Beta=-0.3494	C	T	T	0.004	0.002	0.007	-
HUFS	0.2920	N	8.23E-01	Score=-0.175 Var=0.613	C	T	T	0.002	0.003	0.001	?
SAPPHIRE	0.5069	N	7.47E-03	OR=0.162 [95% CI 0.043-0.614]	C	T	T	0.005			-
BAGS	0.5083	N	2.06E-04	Freq1=0.001 Freq0=0.005	C	T	T	0.002	0.000	0.003	?
GALA II	0.6423	N	2.03E-03	Beta=-0.2895	C	T	T	0.014	0.008	0.020	-

1	159644595	159741698	-0.0237	0.0153	0.7979	0.8177	0.7788	-1.6538	1.5261	-0.9953
1	159756562	159796090	-0.0249	0.0098	0.7997	0.8187	0.7808	-1.3872	1.7033	-0.8356
1	159797418	159819313	-0.0205	0.0344	0.8035	0.8192	0.7842	-0.7947	1.7919	-0.5693
1	159821495	159890459	-0.0195	0.0458	0.8066	0.8220	0.7897	-0.3207	2.2794	-0.1434
1	159912090	159953521	-0.0143	0.1394	0.8039	0.8176	0.7890	-0.7355	1.5039	-0.1966
4	122167001	122397344	0.0202	0.0379	0.8131	0.8073	0.7966	0.6379	-0.9664	-0.3045
4	122431943	122517516	0.0240	0.0143	0.8156	0.8071	0.7979	0.9100	-1.0135	-0.1706
4	122544756	122657735	0.0263	0.0074	0.8154	0.8066	0.7952	0.8891	-1.1077	-0.4383
4	122661260	122901933	0.0282	0.0041	0.8152	0.8042	0.7959	0.8682	-1.5550	-0.3714
4	122921858	123564119	0.0290	0.0030	0.8158	0.8043	0.7938	0.9310	-1.5315	-0.5721
4	123564458	123772304	0.0293	0.0027	0.8156	0.8043	0.7911	0.9100	-1.5315	-0.8398
4	123841789	124285756	0.0302	0.0020	0.8160	0.8035	0.7911	0.9519	-1.6963	-0.8398
4	124288412	124387175	0.0309	0.0016	0.8173	0.8048	0.7918	1.0984	-1.4373	-0.7728
4	**124387497**	**124567967**	**0.0318**	**0.0010**	**0.8152**	**0.8016**	**0.7911**	**0.8682**	**-2.0495**	**-0.8398**
4	124592179	125248603	0.0284	0.0036	0.8161	0.8060	0.7973	0.9728	-1.2254	-0.2375
4	125261530	125485503	0.0281	0.0040	0.8171	0.8063	0.7966	1.0775	-1.1548	-0.3045
4	125488642	125657653	0.0263	0.0073	0.8186	0.8078	0.7986	1.2450	-0.8722	-0.1037
4	125667725	125863974	0.0240	0.0144	0.8179	0.8085	0.7986	1.1612	-0.7545	-0.1037
4	125864789	126165729	0.0240	0.0152	0.8194	0.8110	0.8014	1.3287	-0.2836	0.1639
4	126182392	126409265	0.0266	0.0069	0.8200	0.8085	0.7973	1.3915	-0.7545	-0.2375
4	126412575	126515531	0.0263	0.0075	0.8196	0.8081	0.7979	1.3497	-0.8252	-0.1706
4	126522864	126722949	0.0256	0.0096	0.8202	0.8100	0.7986	1.4125	-0.4720	-0.1037
4	126750168	127030304	0.0281	0.0047	0.8232	0.8117	0.8041	1.7474	-0.1424	0.4316
4	127089467	127238874	0.0283	0.0044	0.8238	0.8122	0.8027	1.8102	-0.0482	0.2977
4	127241211	127274850	0.0286	0.0040	0.8255	0.8126	0.8096	1.9986	0.0224	0.9669
4	127281188	127404747	0.0247	0.0119	0.8194	0.8101	0.8096	1.3287	-0.4484	0.9669
6	132167290	132217687	0.0261	0.0078	0.8240	0.8078	0.8130	1.2247	-0.0302	1.0881
6	132253941	132475231	0.0246	0.0124	0.8230	0.8075	0.8103	1.0782	-0.0909	0.8581
6	132509041	132647736	0.0267	0.0069	0.8259	0.8095	0.8116	1.5178	0.2328	0.9731
6	132664724	132753715	0.0303	0.0021	0.8265	0.8080	0.8171	1.6058	-0.0100	1.4331
6	132765427	132889619	0.0345	0.0004	0.8244	0.8037	0.8123	1.2833	-0.6978	1.0306
6	132891570	132908750	0.0349	0.0003	0.8230	0.8017	0.8089	1.0782	-1.0215	0.7432
6	132928872	133036860	0.0341	0.0005	0.8236	0.8030	0.8151	1.1661	-0.8192	1.2606
6	133052616	133334308	0.0354	0.0003	0.8242	0.8046	0.8219	1.2540	-0.5562	1.8355
6	133370646	133727983	0.0362	0.0002	0.8250	0.8052	0.8253	1.3713	-0.4551	2.1230
6	133731403	133892116	0.0336	0.0006	0.8221	0.8028	0.8253	0.9316	-0.8394	2.1230
6	133897107	134149081	0.0359	0.0002	0.8211	0.8008	0.8212	0.7850	-1.1631	1.7780
6	**134149974**	**134300365**	**0.0370**	**0.0001**	**0.8230**	**0.8013**	**0.8274**	**1.0782**	**-1.0822**	**2.2955**
6	134319531	134377316	0.0334	0.0003	0.8131	0.7917	0.8144	-0.4461	-2.6399	1.2031
6	134380130	134413537	0.0339	0.0002	0.8119	0.7904	0.8075	-0.6220	-2.8423	0.6282
6	134416282	134470574	0.0329	0.0003	0.8133	0.7924	0.8075	-0.4168	-2.5186	0.6282
6	134485169	134517263	0.0301	0.0011	0.8137	0.7936	0.8055	-0.3581	-2.3365	0.4557
6	134538881	134629970	0.0290	0.0018	0.8144	0.7952	0.8103	-0.2409	-2.0735	0.8581
6	134685508	134737470	0.0310	0.0011	0.8219	0.8017	0.8192	0.9023	-1.0215	1.6056
6	134740669	134845333	0.0301	0.0016	0.8209	0.8016	0.8171	0.7557	-1.0418	1.4331
6	134873987	135015339	0.0282	0.0035	0.8213	0.8040	0.8158	0.8143	-0.6574	1.3181

| 6 | 135051669 | 135237744 | 0.0265 | 0.0056 | 0.8175 | 0.8023 | 0.8144 | 0.2281 | -0.9204 | 1.2031 |
| 6 | 135243026 | 135559215 | 0.0237 | 0.0135 | 0.8163 | 0.8046 | 0.8185 | 0.0522 | -0.5562 | 1.5481 |

Supplementary Table S7: Summary of association tests stratified by local ancestry. A logistic regression interaction model was used to determine if an association differs by local ancestry. SNPs that had p-values $< 10^{-5}$ in either of the meta-analyses (including vs. excluding GALA), or that were identified by our candidate gene analysis, were tested using this model. SNPs with MAF < 0.01 and with less than 8 studies contributing to the meta-analysis p-value were excluded from the test. Models with interaction p-values < 0.05 were further examined by stratifying the data set according to 0, 1 or 2 copies of African ancestry at that locus, and then using an additive allelic test for association within each strata, to test for differences in case and control allele frequency.

Chromosome: hg19 position (rsID)	Interaction P-value	0 copies of African ancestry				1 copy of African ancestry				2 copies of African ancestry		
		Case Frequency	Control Frequency	P-value		Case Frequency	Control Frequency	P-value		Case Frequency	Control Frequency	P-value
1:158056779 (rs11264909)	0.0019	0.572	0.574	0.8163		0.359	0.378	0.3555		0.120	0.097	2.80E-05
7:87275641 (rs10246878)	0.0076	0.867	0.798	0.0007		0.915	0.896	0.0001		0.989	0.988	0.1556

Supplementary Table S8: Summary of concordance between chromosome 2q22.3 Taqman and imputed genotypes. Samples from HUFS were typed using a TaqMan genotyping assay in order to do this comparison. The data is 98.76% concordant across 1792 patients with at least one SNP to compare.

SNP	hg19 base pair position	Minor allele	Major allele	MAF	Taqman call rate	Taqman HWE p-value	Taqman-ADPC concordance
rs17834780	144638809	A	G	0.128	0.922	0.546	99.21%
rs787160	144650140	T	A	0.161	0.888	0.027	97.29%
rs787151	144655836	A	G	0.231	0.914	0.566	99.75%

Supplementary Table S9: CAAPA whole genome sequence (WGS) reference panel. This table summarizes the ethnicity and recruitment sites of subjects used to build the CAAPA WGS reference panel.

Ethnicity & Site	Study	Total (% Male)	Asthmatics (% Male)	Non-asthmatics (% Male)
African American (Atlanta)	SARP/COPDGene	50 (66%)	25 (64%)	25 (68%)
African American (Chicago)	CAG	50 (34%)	25 (44%)	25 (24%)
African American (Baltimore-DC)	GRAAD	48 (50%)	24 (50%)	24 (50%)
African American (Nashville)	BREATHE/VALID	49 (45%)	25 (40%)	24 (50%)
African American (NYC)	REACH	50 (54%)	20 (55%)	30 (53%)
African American (Detroit)	SAPPHIRE	49 (35%)	24 (46%)	25 (24%)
African American (San Francisco)	SAGE II	50 (60%)	25 (72%)	25 (48%)
African American (Winston-Salem)	SARP	50 (10%)	25 (8%)	25 (12%)
African American (Jackson)	JHS	50 (52%)	25 (52%)	25 (52%)
Barbados	BAGS	47 (49%)	23 (48%)	22 (50%)
Jamaica	JAAS	50 (50%)	25 (52%)	25 (48%)
Dominican Republic (New York, Texas)	GALA II	47 (40%)	22 (45%)	25 (36%)
Honduras	HONDAS	43 (51%)	22 (55%)	21 (48%)
Colombia	PGCA	50 (48%)	26 (50%)	24 (46%)
Puerto Rico	GALA II	53 (45%)	28 (61%)	25 (28%)
Brazil	BIAS	39 (44%)	7 (29%)	32 (47%)
Nigeria	AEGS	45 (49%)	23 (30%)	22 (68%)
Gabon	Leiden University	28 (50%)		
Palenque (San Basilio de Palenque)	University of Cartagena	32 (63%)		
Total		**880**	**394**	**424**

Abbreviations: *BAGS*, Barbados Asthma Genetics Study; *BIAS*, Brazilian Immunogenetics of Asthma & Schistosomiasis study (Brazil, rural); *CAG*, Chicago Asthma Genetics; *COPDGene*, Genetic Epidemiology of COPD study (Atlanta); *GALA II*; Genetics of Asthma in Latino Americans (Dominican, Puerto Rican); *GRAAD*, Genomic Research on Asthma in the African Diaspora (Baltimore); *JAAS*, Jamaican Adolescent Asthma Study; *JHS*, Jackson Heart Study (Jackson, Mississippi); *PGCA*, Proyecto Genes Candidatos en Asma (Colombia); *REACH*, Reducing Emergency Asthma Care in Harlem (NYC); *SAGE II*, Study of African Americans, Asthma, Genes, & Environments (San Francisco); *SAPPHIRE*, Study of Asthma Phenotypes and Pharmacogenomics Interactions by Race-ethnicity (Detroit); *SARP*, Severe Asthma Research Program (Winston-Salem, Atlanta)

Supplementary Tables S10. Summaries of GWAS+ADPC data set merging and imputation process results.

The GWAS+ADPC data set merging table summarizes the number of subjects (n) and number of SNPs (m) available at different stages of the data merging process. The Imputation QC tables summarize the imputation QC report downloaded from the Michigan imputation server, for pre-imputation QC (Quality Control Only option) of the GWAS data sets and ADPC data sets, and impuation QC of the merged GWAS+ADPC data set.

GWAS+ADPC data set merging

	n initial GWAS	n common samples GWAS	n concordant samples GWAS	m initial GWAS	m QC GWAS	m no ambiguous AT/CG GWAS	m initial ADPC	m QC ADPC	m no ambiguous AT/CG ADPC	m merged
GRAAD(1)	862	785	781	617,061	615,501	615,501	471,507	425,048	392,965	1,000,719
GRAAD(2)	92	88	88	2,307,742	1,991,775	1,949,840	471,507	411,594	369,483	2,253,609
CAG	282	272	270	1,014,303	947,441	945,288	471,507	423,438	387,642	1,319,903
SAPPHIRE	751	733	730	812,088	802,699	798,809	471,507	425,277	394,267	1,165,496
JHS(1)	338	331	327	834,211	751,360	670,004	471,507	423,972	388,795	1,047,756
JHS(2)	654	647	647	834,211	825,930	736,818	471,507	424,987	391,415	1,115,933
SAGE II	1,713	1,693	1,693	772,135	732,697	728,329	471,507	425,157	397,902	1,085,595
HUFS	1,946	1,849	1,848	795,814	791,844	712,972	471,507	424,620	394,239	1,096,292
SARP	352	348	348	464,305	463,635	462,894	471,507	424,329	389,822	852,391
BAGS	840	689	684	617,212	602,691	602,691	471,507	421,659	379,319	974,505
GALA II	1,799	1,762	1,754	723,535	721,317	715,032	471,507	424,621	362,406	1,037,661

Imputation QC

GWAS

	n	m excluded	m remaining	chunks excluded	chunks remaining	m match	reference overlap
GRAAD(1)	785	107,948	507,553	1	289	361,012	97.94%
GRAAD(2)	88	933,074	1,016,766	3	289	769,895	96.14%
CAG	272	163,776	781,512	3	289	552,079	96.81%
SAPPHIRE	733	928	797,881	2	289	643,235	97.08%
JHS(1)	331	123	669,881	3	288	471,169	97.50%
JHS(2)	647	148	736,670	3	288	519,214	97.46%
SAGE II	1693	1,151	727,178	3	289	575,820	96.79%
HUFS	1849	1,553	711,419	3	288	501,190	97.55%
SARP	348	83,528	379,366	3	289	262,387	97.40%
BAGS	689	105,786	496,905	1	289	351,355	97.98%

| GALA II | 1762 | 1,170 | 713,862 | 3 | 289 | 572,136 | 96.79% |

	m allele frequencies > 0.5	m allele switch	m strand flip	m strand flip and allele switch	m AT/CG	m allele mismatch
GRAAD(1)	0	133,834	66,183	41,659	0	106
GRAAD(2)	0	169,899	760,155	170,143	1,780	2,776
CAG	0	199,095	101,438	61,658	148	350
SAPPHIRE	0	130,953	18	3	354	907
JHS(1)	0	174,670	0	0	7,321	123
JHS(2)	0	190,622	1	0	8,084	147
SAGE II	0	127,368	0	0	605	1,151
HUFS	0	183,403	895	508	9,383	150
SARP	0	104,868	50,793	32,717	58	18
BAGS	0	133,355	64,591	41,090	0	105
GALA II	0	118,687	0	0	90	1,170

ADPC

	n	m excluded	m remaining	chunks excluded	chunks remaining	m match	reference overlap
GRAAD(1)	785	272	392,693	7	286	338,041	95.63%
GRAAD(2)	88	255	369,228	7	286	316,285	95.57%
CAG	272	267	387,375	7	286	333,183	95.63%
SAPPHIRE	733	271	393,996	7	286	339,347	95.65%
JHS(1)	331	267	388,528	7	286	334,142	95.62%
JHS(2)	647	267	391,148	7	286	336,675	95.62%
SAGE II	1693	269	397,633	7	286	342,791	95.67%
HUFS	1849	270	393,969	7	286	339,510	95.67%
SARP	348	266	389,556	7	286	335,105	95.63%
BAGS	689	262	379,057	7	286	325,050	95.58%
GALA II	1762	272	362,134	7	286	311,310	95.57%

	m allele frequencies > 0.5	m allele switch	m strand flip	m strand flip and allele switch	m AT/CG	m allele mismatch
GRAAD(1)	0	35,627	34	22	1,857	216
GRAAD(2)	0	35,359	38	17	1,226	200
CAG	0	35,611	35	20	1,633	212

SAPPHIRE		0	35,476	35		21	2,004	215
JHS(1)		0	35,732	34		22	1,639	211
JHS(2)		0	35,642	36		20	1,695	211
SAGE II		0	35,338	35		21	2,278	213
HUFS		0	35,462	34		22	1,932	214
SARP		0	35,693	34		22	1,710	210
BAGS		0	36,121	35		21	1,106	206
GALA II		0	34,416	34		22	367	216

GWAS+ADPC

	n	m excluded	m remaining	chunks excluded	chunks remaining	m match	reference overlap
GRAAD(1)	781	330	1,000,389	5	288	759,714	97.02%
GRAAD(2)	88	2,961	2,250,648	5	288	1,792,321	96.01%
CAG	270	954	1,318,949	5	288	977,477	96.46%
SAPPHIRE	730	1,129	1,164,367	5	288	959,241	96.58%
JHS(1)	327	372	1,047,384	6	287	796,892	96.80%
JHS(2)	647	357	1,115,576	6	287	846,539	96.81%
SAGE II	1693	1,350	1,084,245	5	288	883,845	96.33%
HUFS	1848	364	1,095,928	5	288	833,314	96.87%
SARP	348	228	852,163	7	287	648,130	96.59%
BAGS	684	315	974,190	6	287	735,686	97.04%
GALA II	1754	1,372	1,036,289	5	288	849,240	96.32%

	m allele frequencies > 0.5	m allele switch	m strand flip	m strand flip and allele switch	m AT/CG	m allele mismatch
GRAAD(1)	0	209,045	0	0	1,857	321
GRAAD(2)	0	365,553	0	0	2,822	2,958
CAG	0	292,991	0	0	1,771	562
SAPPHIRE	0	162,890	0	0	2,349	1,117
JHS(1)	0	208,187	0	0	8,821	334
JHS(2)	0	223,829	0	0	9,603	357
SAGE II	0	157,745	0	0	2,866	1,350
HUFS	0	217,151	0	0	11,118	363
SARP	0	173,176	0	0	1,767	228

BAGS	0	208,511	0	0	1,106	310
GALA II	0	148,387	0	0	449	1,370

Supplementary Table S11: The proportion of SNPs deleted and retained in the imputed datasets by dataset and MAF category.
SNPs with MAF ≤ 0.005 were deleted if Rsq ≤ 0.5 and SNPs with MAF > 0.005 were deleted if Rsq ≤ 0.3.

	Initial nr imputed SNPs	Nr MAF ≤ 0.005 SNPs deleted (%)	Nr MAF ≤ 0.005 SNPs retained (%)	Nr MAF > 0.005 SNPs deleted (%)	Nr MAF > 0.005 SNPs retained (%)	Nr SNPs remaining
GRAAD(1)	0.69	29,823,751	3,851,948 (31%)	931,948 (5%)	16,424,959 (95%)	20,276,907
GRAAD(2)	0.96	29,823,751	446,637 (4%)	543,562 (3%)	17,166,766 (97%)	17,613,403
CAG	0.76	29,823,751	2,988,080 (24%)	669,391 (4%)	16,726,200 (96%)	19,714,280
SAPPHIRE*	0.69	29,823,751	3,863,294 (31%)	843,889 (5%)	16,523,446 (95%)	20,386,740
JHS(1)	0.79	29,807,007	2,666,990 (21%)	1,116,492 (6%)	16,124,611 (94%)	18,791,601
JHS(2)	0.71	29,807,007	3,657,343 (29%)	952,901 (5%)	16,402,994 (95%)	20,060,337
SAGE II	0.65	29,823,751	4,396,838 (35%)	887,008 (5%)	16,466,243 (95%)	20,863,081
HUFS	0.67	29,823,751	4,146,355 (33%)	956,849 (6%)	16,429,561 (94%)	20,575,916
SARP	0.8	29,807,007	2,578,587 (20%)	1,194,584 (7%)	16,003,849 (93%)	18,582,436
BAGS	0.76	29,807,007	3,069,717 (24%)	947,981 (6%)	16,221,403 (94%)	19,291,120
GALA II	0.7	29,823,751	4,406,580 (30%)	1,218,368 (8%)	13,745,640 (92%)	18,152,220

* Using ADPC data of 730 cases

Supplementary Table S12: Summary of ancestry distribution in the CAAPA datasets. The median and interquartile range (IQR) of the proportion of African ancestry, estimated using the software program ADMIXTURE, are listed by dataset and asthma case-control status. The p-values of univariate association tests between asthma case-control status and each of the first 10 principal components (PCs) are also listed by study.

	IQR asthma cases	IQR controls	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
GRAAD(1)	0.83 [0.77 - 0.87]	0.83 [0.77 - 0.88]	0.231	0.185	0.728	0.841	0.138	0.625	0.377	0.816	0.159	0.532
GRAAD(2)	0.84 [0.77 - 0.88]	0.79 [0.72 - 0.89]	0.154	0.175	0.653	0.499	0.536	0.363	0.912	0.081	0.141	0.745
CAG	0.84 [0.77 - 0.88]	0.84 [0.77 - 0.87]	0.985	0.227	0.028	0.216	0.440	0.025	0.235	0.875	0.197	0.596
SAPPHIRE	0.81 [0.75 - 0.86]*		0.375	0.601	0.178	0.747	0.763	0.379	0.698	0.014	0.784	0.085
JHS(1)	0.85 [0.80 - 0.88]	0.83 [0.75 - 0.89]	0.040	0.429	0.521	0.420	0.563	0.473	0.698	0.597	0.702	0.239
JHS(2)	0.85 [0.80 - 0.90]	0.84 [0.77 - 0.88]	0.021	0.142	0.163	0.207	0.830	0.735	0.338	0.956	0.765	0.427
SAGE II	0.81 [0.75 - 0.86]	0.81 [0.74 - 0.85]	0.487	0.624	0.807	0.553	0.114	0.779	0.762	0.654	0.810	0.691
HUFS	0.82 [0.75 - 0.88]**	0.83 [0.76 - 0.88]**	0.814	0.358	0.473	0.187	0.439	0.348	0.161	0.049	0.085	0.550
SARP	0.83 [0.76 - 0.88]	0.84 [0.79 - 0.88]	0.953	0.547	0.459	0.796	0.088	0.925	0.847	0.329	0.385	0.106
BAGS	0.91 [0.86 - 0.95]**	0.90 [0.85 - 0.95]**	0.914	0.813	0.856	0.107	0.620	0.403	0.407	0.188	0.062	0.437
GALA II	0.19 [0.14 - 0.26]	0.18 [0.13 - 0.25]	0.004	0.029	0.054	0.367	0.705	0.490	0.744	0.352	0.926	0.477

* Using 730 subjects for which ADPC data was available

** Estimated using founder individuals only

Supplementary Table S13: Software used for data analysis. With the exception of software packages used for calling genotype data, this table summarizes the software packages used for processing and analyzing data.

Software package	URL	Application	Parameters
ADMIXTURE 1.3.0	https://www.genetics.ucla.edu/software/admixture/	Estimating genome-wide proportions of ancestry	K=3
ANNOVAR	http://annovar.openbioinformatics.org	Gene, filter and region based annotation	See Supplementary Table S2
CrossMap v0.2.1	http://crossmap.sourceforge.net	Lifting over genomic coordinates of the JHS(1) and JHS(2) datasets	Chain file used: *hg18ToHg19.over.chain*
EMMAX 07Mar2010	http://genetics.cs.ucla.edu/emmax/	Tests for association in the CAG, GRAAD(1), GALA II and SAGE II studies.	Balding-Nichols kinship matrix estimation
Genesis	http://www.bioinf.wits.ac.za/software/genesis/	PC and ADMIXTURE plots	
KING 1.4	http://people.virginia.edu/~wc9c/KING/index.html	Estimating the kinship matrix for the GRAAD(2), JHS(1), JHS(2), HUFS and SARP studies	Default parameters were used
LocusZoom v0.4.8	http://locuszoom.sph.umich.edu	Locus zoom plots	
Minimac3	https://imputationserver.sph.umich.edu	Imputation of GWAS+ADPC data	*Reference Panel:* CAAPA – African American Panel *Phasing:* ShapeIT *Population:* AA (CAAPA) Imputations were run March-April 2016, and imputations with masked genotypes (to assess imputation accuracy) were run July 2016.
METAL 2011-03-25	http://csg.sph.umich.edu/abecasis/Metal/download/	Meta-analysis	P-value based analysis was done, with the default weight of each cohort set to $2/(1/N_{cases} + 1/N_{controls})$
MQL_dose	http://csg.sph.umich.edu/liang/MQLS/	Tests for association for BAGS	Disease prevalence = 20% were used
PLINK v1.9	https://www.cog-genomics.org/plink2/	Sample and SNP QC, LD filtering, MAF filtering, strand flipping, data set merging, logistic regression association tests.	
R 3.2.5 and 3.3.0	https://cran.r-project.org	Data visualization and data set merging; statistical tests	
GENESIS R Bioconductor package	https://www.bioconductor.org/packages/devel/bioc/html/GENESIS.html	Tests for association in the HUFS cohort; Principal component analysis	
RFMix v1.5.4	https://sites.google.com/site/rfmixlocalancestryinference/	Local ancestry inference	Default parameters were used
ShapeIt v2.r790 & v2.r837	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html	Used for genotype phasing of CAAPA WGS imputation reference panel (r790) and for local ancestry inference (r837)	b37 genetic map files were used for local ancestry inference phasing
VCFtools 0.1.13	https://vcftools.github.io/index.html	SNP filtering based on Rsq values.	

REFERENCES

1. Mathias, R.A., et al., *A genome-wide association study on African-ancestry populations for asthma.* J Allergy Clin Immunol, 2010. **125**(2): p. 336-346 e4.

2. Torgerson, D.G., et al., *Genome-wide ancestry association testing identifies a common European variant on 6q14.1 as a risk factor for asthma in African American subjects.* J Allergy Clin Immunol, 2012. **130**(3): p. 622-629 e9.

3. Padhukasahasram, B., et al., *Gene-based association identifies SPATA13-AS1 as a pharmacogenomic predictor of inhaled short-acting beta-agonist response in multiple population groups.* Pharmacogenomics J, 2014. **14**(4): p. 365-71.

4. Gould, W., et al., *Factors predicting inhaled corticosteroid responsiveness in African American patients with asthma.* J Allergy Clin Immunol, 2010. **126**(6): p. 1131-8.

5. Rumpel, J.A., et al., *Genetic ancestry and its association with asthma exacerbations among African American subjects with asthma.* J Allergy Clin Immunol, 2012. **130**(6): p. 1302-6.

6. Hickson, D.A., et al., *Asthma and asthma severity among African American adults in the Jackson Heart Study.* J Asthma, 2009. **46**(4): p. 421-8.

7. Carpenter, M.A., et al., *Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study.* Am J Med Sci, 2004. **328**(3): p. 131-44.

8. Payne, T.J., et al., *Sociocultural methods in the Jackson Heart Study: conceptual and descriptive overview.* Ethn Dis, 2005. **15**(4 Suppl 6): p. S6-38-48.

9. Taylor, H.A., Jr., et al., *Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study.* Ethn Dis, 2005. **15**(4 Suppl 6): p. S6-4-17.

10. Flores, C., et al., *African ancestry is associated with asthma risk in African Americans.* PLoS One, 2012. **7**(1): p. e26807.

11. Adeyemo, A., et al., *A genome-wide association study of hypertension and blood pressure in African Americans.* PLoS Genet, 2009. **5**(7): p. e1000564.

12. Faruque, M.U., et al., *Transferability of genome-wide associated loci for asthma in African Americans.* J Asthma, 2017. **54**(1): p. 1-8.

13. Moore, W.C., et al., *Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program.* Am J Respir Crit Care Med, 2010. **181**(4): p. 315-23.

14. Barnes, K.C., et al., *Linkage of asthma and total serum IgE concentration to markers on chromosome 12q: evidence from Afro-Caribbean and Caucasian populations.* Genomics, 1996. **37**(1): p. 41-50.

15. Zambelli-Weiner, A., et al., *Evaluation of the CD14/-260 polymorphism and house dust endotoxin exposure in the Barbados Asthma Genetics Study.* J Allergy Clin Immunol, 2005. **115**(6): p. 1203-9.

16. Kumar, R., et al., *Factors associated with degree of atopy in Latino children in a nationwide pediatric sample: the Genes-environments and Admixture in Latino Asthmatics (GALA II) study.* J Allergy Clin Immunol, 2013. **132**(4): p. 896-905 e1.

17. Mathias, R.A., et al., *A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome.* Nat Commun, 2016. **7**: p. 12522.

18. Li, X., et al., *Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases.* J Allergy Clin Immunol, 2012. **130**(4): p. 861-8 e7.

19. Musunuru, K., et al., *Candidate gene association resource (CARe): design, methods, and proof of concept.* Circ Cardiovasc Genet, 2010. **3**(3): p. 267-75.

20. White, M.J., et al., *Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study.* Immunogenetics, 2016. **68**(6-7): p. 391-400.

21. Chen, G., et al., *Development of admixture mapping panels for African Americans from commercial high-density SNP arrays.* BMC Genomics, 2010. **11**: p. 417.

22. Liu, E.Y., et al., *Genotype imputation of Metabochip SNPs using a study-specific reference panel of ~4,000 haplotypes in African Americans from the Women's Health Initiative.* Genet Epidemiol, 2012. **36**(2): p. 107-17.

23. Pirinen, M., P. Donnelly, and C.C.A. Spencer, *Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies.* Annals of Applied Statistics, 2013. **7**(1): p. 369-390.

24. Thornton, T. and M.S. McPeek, *Case-control association testing with related individuals: a more powerful quasi-likelihood score test.* Am J Hum Genet, 2007. **81**(2): p. 321-37.

25. Howitt, M.E., T.C. Roach, and R. Naidu, *Prevalence of Childhood Asthma and Allergy in Barbados.* American Journal of Respiratory and Critical Care Medicine. **157**(A642).

26. Winkler, T.W., et al., *Quality control and conduct of genome-wide association meta-analyses.* Nat Protoc, 2014. **9**(5): p. 1192-212.

27. Manichaikul, A., et al., *Robust relationship inference in genome-wide association studies.* Bioinformatics, 2010. **26**(22): p. 2867-73.

28. Gao, X., J. Starmer, and E.R. Martin, *A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms.* Genet Epidemiol, 2008. **32**(4): p. 361-9.