Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations

Michelle Daya et al.#

Asthma is a complex disease with striking disparities across racial and ethnic groups. Despite its relatively high burden, representation of individuals of African ancestry in asthma genome-wide association studies (GWAS) has been inadequate, and true associations in these underrepresented minority groups have been inconclusive. We report the results of a genome-wide meta-analysis from the Consortium on Asthma among African Ancestry Populations (CAAPA; 7009 asthma cases, 7645 controls). We find strong evidence for association at four previously reported asthma loci whose discovery was driven largely by non-African populations, including the chromosome 17q12–q21 locus and the chr12q13 region, a novel (and not previously replicated) asthma locus recently identified by the Trans-National Asthma Genetic Consortium (TAGC). An additional seven loci reported by TAGC show marginal evidence for association in CAAPA. We also identify two novel loci (8p23 and 8q24) that may be specific to asthma risk in African ancestry populations.
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 disproportionately affected by asthma. For example, African Americans and Puerto Ricans have higher asthma-related morbidity and mortality rates compared to European Americans1–3. In addition to environmental, cultural, and socioeconomic 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) has been limited, and in GWAS’s performed to date including individuals of African ancestry, the samples have been modest and underpowered to detect true associations. For example, the largest asthma GWAS focused solely on African ancestry populations included only 819 asthma cases1. Recently, the Trans-National Asthma Genetic Consortium (TAGC) reported 18 asthma-associated loci based on a meta-analysis of 142,486 subjects, but only 2149 cases and 6055 controls included in this study were of African ancestry. Only two genome-wide significant associations5,6 have been reported for asthma from GWAS conducted in African ancestry populations to date4–8. The discovery of genetic risk factors for asthma in African ancestry populations has been further hampered by lack of representation of African ancestry in imputation reference panels, and legacy commercial genotyping arrays that have until very recently not provided adequate coverage of linkage disequilibrium (LD) patterns in African ancestry populations.

To address these disparities in asthma genetics research and the paucity of information on African genetic diversity, we previously established the Consortium on Asthma among African ancestry Populations in the Americas (CAAPA)9. Because of the lack of adequate representation of African ancestry in imputation reference panels, we first performed whole-genome sequencing (WGS) of samples collected from 880 individuals who self-reported African ancestry from 19 North, Central and South American and Caribbean populations (446 individuals from nine African American populations, 43 individuals from Central America, 121 individuals from three South American populations, and 197 individuals from four 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).

Previously we performed coverage analysis of the novel variation identified in the CAAPA sequence data, and found only 69% of common SNP variants and 41% of low-frequency SNP variants identified by CAAPA can be tagged by traditional GWAS arrays at \( r^2 \geq 0.8 \)10. In addition, lack of coverage of low frequency variants (minor allele frequency [MAF] between 0.01–0.05) 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) in partnership with Illumina, Inc., a gene-centric SNP genotyping array designed to complement commercially available genome-wide chips, thereby improving tagging and coverage of African ancestry genetic variation10. The array included ~495,000 SNPs, with a MAF enriched for low frequency variants. Subsequently, the content of the ADPC was incorporated into Illumina’s Multi-Ethnic Global Array (MEGA)10.

Using the ADPC, we genotyped CAAPA participants from nine studies (seven African American studies, one African Caribbean [Barbados], and one South American [Puerto Rico]). We combined the ADPC data with existing commercial genome-wide genotype data, and imputed additional genotypes using the CAAPA reference panel. Subsequently we used the MEGA to genotype additional African ancestry asthma studies with no previously existing genome-wide genotype data (one African American, three South American, and two Caribbean studies), and similarly imputed genotypes on these subjects using the CAAPA reference panel. Sample populations, their ascertainment, and clinical characteristics are described in detail in the Supplementary Note 1, Table 1 and Supplementary Table 1. We then used these data to perform a GWAS of asthma in individuals of African ancestry (7009 asthmatic cases and 7645 controls). We also performed admixture mapping, a technique that leverages local ancestry to identify regions of the genome where ancestry from a particular ancestral population is inherited more frequently among affected compared to unaffected individuals.

Our GWAS results recapitulate 11 of the 18 loci recently reported by TAGC, including the chromosome 9p24 (IL3), 15q22 (ROA), and 17q12–q21 loci, as well as a locus on chromosome 12q13 (STAT6) reported as novel but not replicated by TAGC. We identify two loci on chromosome 8 not previously reported as being associated with asthma, and one local ancestry peak on chromosome 6q22 is genome-wide significant in our admixture mapping study. We speculate that the common asthma-associated variants on chromosome 8p23 implicate the ARHGEF10 or MYOM2 genes, while a low frequency protective variant on chromosome 8q42 is intronic to TATDN1, a gene with increased expression in human airway smooth muscle cells stimulated with interleukin. The TCF21 gene, which has previously been shown to be differentially expressed in bronchial biopsies of asthmatics compared to controls, is a likely candidate in the admixture mapping peak. However, further replication efforts are necessary to provide robust evidence of replication for these chromosome 6 and 8 loci.

Results

Association analysis. Studies included in the asthma association analysis represent a diverse spectrum of African ancestry (Fig. 1a, Supplementary Fig. 1, Supplementary Table 2) with median African ancestry proportions in non-asthmatics as low as 0.17 in subjects from Puerto Rico (GALA II) and as high as 0.90 in subjects from Jamaica and Barbados (JAAS and BAGS). In addition, the studies had different objectives and differed in potentially relevant factors such as age of onset and diagnostic criteria (see Supplementary Note 1 and Supplementary Table 1). For this reason, we performed tests for association separately for each dataset, and then combined the results using MR-MEGA, a novel meta-analysis approach that models allelic effects as a function of axes of genetic variation11. In this way, heterogeneous associations across genetically distinct populations are not penalized, and the degree of heterogeneity due to ancestry, as well as residual effects due to differences in study design, can be assessed. The results of the genome-wide meta-analysis are summarized in Fig. 1b and c, and associations with MR-MEGA association \( p < 10^{-6} \) (the same cut-off employed by the EVE GWAS6) are summarized in Table 2 and Supplementary Table 9. There were seven loci with associations smaller than this threshold. Associations at two loci were genome-wide significant \( (p < 5 \times 10^{-8}) \), including a locus at chromosome 8p23 not previously reported by any asthma GWAS7,12, and the chromosome 17q12–q21 locus, which is regarded as one of the most consistent asthma association findings to date13. Two of the loci with \( p < 10^{-6} \) (but not attaining conventional genome-wide significance) were reported recently by TAGC in the multi-ethnic...
Associations novel to CAAPA. Several SNPs intronic to a gene encoding a long non-coding RNA on chromosome 8p23 have MR-MEGA association $p < 10^{-6}$, and two of these associations were genome-wide significant. While this association was observed in multiple African American samples, the strongest effects were observed in three non-U.S. studies from Barbados (BAGS), Cartagena, Colombia (PGCA), and Puerto Rico (GALA II; Supplementary Fig. 15). While the associated SNPs in this region do not overlap with any expression quantitative trait locis (eQTLs) in the publicly available databases we mined, long-range association between the eQTL and the MR-MEGA association was observed in the European ancestry population. The accuracy of SNP imputation was supported by evidence for replication of this SNP association in BAGS and PGCA, which are likely not the case for the European ancestry populations.

Comparison to previous asthma GWAS. We compared results from the recent TAGC meta-ethnic analysis, the largest and most definitive GWAS performed in observed asthma cases and controls to date, with the CAAPA meta-analysis. TAGC reported 18 loci associated with asthma, categorizing nine loci as known asthma susceptibility genes (known), five as new loci (known), and four as new signals at loci previously associated with asthma (new), two as new signals at loci previously associated with cardiovascular disease (new), and two as new signals at loci previously associated with asthma in ancestry-specific populations (ancestry-specific), plus

Table 1 Studies included in the CAAPA association analysis

| Population       | Location         | Study   | GWAS platform          | Non-asthmatics | Asthmatics | Total | Nr SNPs |
|------------------|------------------|---------|------------------------|----------------|------------|-------|---------|
| African American | Baltimore        | BASS    | Illumina MEGA          | 216            | 135        | 351   | 12,403,613|
|                  | Baltimore        | GRAAD(1)| ADPC + Illumina HumanHap 650Y | 385            | 396        | 781   | 15,486,076|
|                  | Baltimore        | GRAAD(2)| ADPC + Illumina OMNI 2.5 | 23             | 65         | 88    | 7,496,303 |
| Chicago          | CAG              | ADPC    | + Illumina HumanHapIIM | 156            | 114        | 270   | 11,604,736|
| Detroit          | SAPPHIRE         | ADPC^4+ | Affymetrix Axiom AFR   | 566            | 1325       | 1891  | 18,768,360|
| Jackson          | JHS(1)           | ADPC    | + Affymetrix 6.0       | 283            | 44         | 327   | 12,124,436|
|                  | JHS(2)           | ADPC    | + Affymetrix 6.0       | 546            | 101        | 647   | 14,777,976|
| San Francisco    | SAGE II          | ADPC    | + Affymetrix Axiom LAT | 691            | 1001       | 1692  | 18,008,099|
| Washington       | HUF5             | ADPC    | + Affymetrix 6.0       | 1527           | 303        | 1830  | 18,102,295|
| Winston-Salem    | SARP             | ADPC    | + Illumina HumanOmniExpress + HumanHapIIM | 45            | 302        | 347   | 12,281,618|

The table summarizes the geographical location and number of subjects for each study included in the association analysis. Bold numbers are the sums of the numbers above. *ADPC data available for 730 asthmatics only.

The table includes the geographical location and number of subjects for each study included in the association analysis. Bold numbers are the sums of the numbers above. *ADPC data available for 730 asthmatics only.
two as asthma signals previously reported for asthma with hay fever (asthma + hay fever) (Table 2 and Supplementary Table 12)7. Because some of the CAAPA studies were included in the TAGC asthma meta-analysis (of the 2149 asthmatics and 6055 non-asthmatics in TAGC, 1601 asthmatics and 2375 non-asthmatics were from studies included in the CAAPA discovery, and 548 asthmatics and 3680 non-asthmatics were from studies included in the CAAPA replication), to contrast African and European asthmatics and 3680 non-asthmatics were from studies included in the CAAPA discovery, and 548 asthmatics in TAGC, 1601 asthmatics and 2375 non-asthmatics were from studies included in the TAGC multi-ethnic meta-analysis is largely due to their strong associations observed in subjects of European ancestry.

TAGC summary statistics were merged with the CAAPA meta-analysis results, and associations of 810 SNPs with significant fixed-effect p-values in TAGC Europeans, Bonferroni corrected for the number of overlapping SNPs in the merged dataset, were assessed for replication in the CAAPA meta-analysis. Three known asthma loci replicated in CAAPA, after a Bonferroni correction for 20 tests (1 locus from EVE, 1 locus from eMERGE, and 18 loci from TAGC). Windows of ±10 KB around the lead SNP at each selected locus are colored blue (EVE and eMERGE loci), red (TAGC loci), and purple (CAAPA loci with lead SNPs having Bonferroni-adjusted alpha level for 20 tests (1 locus from EVE, 1 locus from eMERGE, and 18 loci from TAGC). Windows of ±10 KB around the lead SNP at each selected locus are colored blue (EVE and eMERGE loci), red (TAGC loci), and purple (CAAPA loci with lead SNPs having p < 10^-6). A larger window of ±200 KB is shown for the chromosome 17q12-21 locus.
Correction for 810 association tests: chromosomes 9p24 (RANBP6, IL33), 15q22 (RORA, NARG2, VPS13C), and 17q12–q21 (Supplementary Table 13).

The TAGC lead SNP rs167679 on chromosome 12q13 is intrinsic to STAT6, a transcription factor that affects Th2 lymphocyte responses mediated by IL-4 and IL-13.22,23 This was a new association reported by TAGC, not previously implicated in any asthma GWAS, although we note this SNP has been reported as a putatively causal SNP discovered by GWAS of lung function,23,24 and a number of linkage studies have pinpointed this chromosomal region in the early days of genome-wide investigations of asthma and atopy.25–28 In addition, markers in STAT6 have been identified by a number of candidate gene association studies.29–32 However, prior to the CAAPA analyses, this locus had not been replicated in independent asthma GWAS datasets. The lead SNP rs3122929 in CAAPA at this locus is in strong LD with the TAGC lead SNP rs167769 in Europeans from the TGP (r² = 0.93) and nearly achieves genome-wide significance in CAAPA (MR-MEGA association p = 9.1 × 10⁻⁷). In addition, the association observed in TAGC subjects of African ancestry was marginal (fixed effect p = 0.05). With the increased sample size available through CAAPA, our meta-analysis provides further evidence of the association between the 12q13 region and asthma, confirming its contribution to asthma risk in African ancestry populations.

We also assessed windows ±10 KB from each of the 18 TAGC lead SNPs for replication in CAAPA (±200 KB for the chr17q12–q21 locus, due to the extended LD in this region), as well as 2 SNPs previously reported as achieving genome-wide significance in African ancestry populations5,6 (Fig. 1c). Considering a Bonferroni-corrected significance threshold for 20 tests (one for each of these prior loci), an additional seven TAGC loci showed evidence of association in the CAAPA meta-analysis (Table 2, loci not in bold font). This includes a novel TAGC locus on chromosome 6p21 implicating human leukocyte antigen genes. In addition, for the known chromosome 5q31 TAGC locus, one of the CAAPA associations passing this significance threshold involved a SNP in strong LD with the TAGC lead SNP in TGP Europeans (rs1295686, r² with TAGC lead SNP rs20541 = 0.96, Supplementary Table 18).

The genome-wide meta-analysis of asthma in multi-ethnic populations previously performed by the EVE consortium reported a genome-wide significant African ancestry-specific association for SNP rs1101999, which is intrinsic to the PYHIN1 gene.6 This was the first African ancestry asthma association reported by a GWAS. This SNP has a high minor allele frequency (MAF < 0.05) in TGP phase 3 AFR population (n = 1322), but a very rare or low frequency in other populations (MAF < 0.005 in TGP phase 3 EAS, EUR, and SAS, n = 1008, 1006, and 978, respectively, MAF = 0.04 in AMR, n = 694), and the role this gene may play in asthma remains unclear. SNP rs1101999 was marginally associated with asthma in the CAAPA meta-analysis (MR-MEGA association p = 6.4 × 10⁻³), but the strength of association was much reduced from its original report (Supplementary Fig. 13, Supplementary Tables 12 and 14). Recently, the eMERGE (electronic medical records and genomics) network conducted an asthma GWAS in biobank subjects, and reported one genome-wide significant association in African Americans, SNP rs11785591, intrinsic to the PTGES gene. This association did not replicate in the CAAPA meta-analysis (MR-MEGA association p = 0.23, Supplementary Table 12, Fig. 1c, Supplementary Fig. 13).

### Table 2 Lead SNP summary

| GWAS | Locus | Descr. | rsID | hg19 position | Genes | RA/EA | CAAPA | EAF | OR [95% CI] | P | TAGC | EAF | OR [95% CI] | P |
|------|-------|--------|------|---------------|-------|-------|-------|-----|------------|---|------|-----|------------|---|
| CAAPA | 8p23 | new | rs13278710 | 2,550,802 | LOC10972815 (intronic) | C/T | 0.12 | 1.26 | [1.16-1.37] | 3.2 × 10⁻⁶ | 0.19 |
| | 8q24 | new | rs14467718 | 125,531,357 | TATDNI (intronic) | C/T | 0.01 | 0.48 | [0.37-0.63] | 2.7 × 10⁻⁷ | 0.00 |
| | 12q23 | known | rs3122929 | 57,509,102 | STAT6, LR1P1 (intronic) | G/A | 0.27 | 1.17 | [1.01-1.24] | 9.1 × 10⁻⁴ | 0.34 |
| | 15a22 | known | rs10519067 | 610,668,347 | RORA (intronic) | G/A | 0.29 | 0.85 | [0.80-0.90] | 1.8 × 10⁻⁷ | 0.14 |
| | 17q12–21 | new | rs907092 | 37,922,259 | ORMDL3, GSDMB, ZBP2, ERBB2 | G/A | 0.20 | 0.80 | [0.75-0.85] | 4.3 × 10⁻¹² | 0.47 |
| | 9p24 | known | rs999296 | 6,209,697 | RANBP6, IL33 | A/G | 0.70 | 0.88 | [0.84-0.94] | 1.1 × 10⁻⁴ | 0.75 |
| | 11q13 | known | rs7927730 | 110,404,999 | IL2C, IL12B, IL12R | A/G | 0.34 | 1.08 | [1.02-1.14] | 9.0 × 10⁻³ | 0.37 |
| | 12q13 | known | rs1107558 | 53,507,775 | STAR6, NAB2, LR1P1 | C/T | 0.11 | 1.10 | [1.04-1.16] | 6.9 × 10⁻⁴ | 0.40 |
| | 17q12–21 | known | rs20541 | 37,922,259 | HLA-DPB1, HLA-DOA1 | A/G | 0.56 | 1.07 | [1.02-1.13] | 0.03 | 0.56 |
| | 15q22 | known | rs2033784 | 67,449,660 | IL2C, IL12B, IL12R | A/G | 0.97 | 0.89 | [0.86-0.91] | 2.7 × 10⁻⁷ | 0.37 |
| | 15q22 | known | rs167769 | 57,503,775 | SMAD3, SMAD6, VPS13C, RANBP6, IL33 | A/G | 0.35 | 1.02 | [0.97-1.08] | 0.32 | 0.13 |
| | 17q12–21 | known | rs20541 | 37,922,259 | HLA-DPB1, HLA-DOA1 | A/G | 0.56 | 1.07 | [1.02-1.13] | 0.03 | 0.56 |
| | 17q12–21 | new | rs2033784 | 67,449,660 | SMAD3, SMAD6, AAGAB, GSDMB, ORMDL3, GSDMB | A/G | 0.42 | 0.94 | [0.89-1.10] | 0.33 | 0.11 |
| | 17q12–21 | new | rs2952156 | 37,876,835 | ERBB2, PGAP3, C1orf57 | A/G | 0.58 | 0.92 | [0.87-0.97] | 0.01 | 0.70 |

CAAPA lead SNPs (MR-MEGA association p < 10⁻⁶) and genome-wide significant associations reported by TAGC with evidence for replication in CAAPA are listed. Because some of the CAAPA studies were included in the TAGC asthma meta-analysis, and in order to contrast African and European asthma susceptibility loci, the table summarizes the associations in TAGC Europeans only (alleles from the TAGC random effects analysis, reported for the CAAPA and TAGC results). TAGC loci with strong evidence for association in CAAPA are highlighted in bold font. 1000 Genomes Project phase III European allele frequencies are reported as the estimated EAF for TAGC Europeans.

New asthma GWAS result not reported prior to the corresponding GWAS. Known asthma GWAS result reported prior to the corresponding GWAS. RA reference allele, EA effect allele. EAF allele frequency, P P asthma association
American (20%), and a very small European component (3%). Interestingly, the lead SNP in CAAPA was the same SNP reported by a meta-analysis of asthma in Puerto Rican children\textsuperscript{33}, distinct from the lead SNPs reported by the multi-ethnic EVE and TAGC GWAS\textsuperscript{6,7}. The most significant and largest effect size magnitudes were observed for the studies that had higher European and Native American components; however, we speculate that this may reflect risk for asthma inherited from a Native American genetic background given the minimal European component in the HONDAS population. We note this trend is not as strong in PGCA, the CAAPA study with the highest proportion of Native American ancestry (29%), which may be due in part to the heterogeneous patterns of LD in this chromosomal region.

In addition to the smaller effect size magnitude in studies with high African ancestry, we also observed much weaker associations (Supplementary Fig. 11 vs. Supplementary Fig. 12), and no chr17q12–q21 associations with \( p < 10^{-6} \) were observed in an inverse-variance meta-analysis of African American samples (Supplementary Table 4), despite a relatively large sample size of 3651 cases and 4222 controls. These observations are consistent with a recent report by Stein et al. summarizing the chromosome 17q12–q21 locus, which included an analysis and discussion of the relatively weak associations observed at this locus among African Americans\textsuperscript{13}. They posited that the reduced strength of association may be due to an overall lower MAF spectrum in African Americans in this region (which reduces statistical power to detect association), breakdown of LD on African haplotypes, and different asthma endotypes (viral exposures) in children. Based on the reduced effect size magnitude observed in studies with higher average African ancestry, we additionally posit a smaller effect size magnitude is observed on African ancestry haplotypes (which could in turn be due to breakdown of LD and reduced correlation between tagging and causal variants). To investigate this, we extracted 17 putatively causal SNPs reported in the Stein et al. review, plus an additional five SNPs from the CAAPA meta-analysis with MR-MEGA association \( p < 10^{-6} \) and \( r^2 < 0.8 \) between all 17 SNPs in TGP European and African populations. We then stratified subjects for whom local ancestry estimates were available based on the number of copies of African ancestry at each of these SNPs, and tested for association between each SNP and asthma separately for each local ancestry group (0, 1, or 2 copies of African ancestry, Supplementary Data 2). Figure 2c shows a trend of decreasing effect size magnitude as the number of copies of African ancestry increased, for most of these 22 candidate SNPs, suggesting (but not conclusively proving) smaller effect sizes magnitudes on African ancestry haplotypes.

Consistent with Stein et al., the degree to which LD breaks down in CAAPA samples increased with average African ancestry (Fig. 2a and Supplementary Fig. 18). One example of how this breakdown could affect association is the rs12936231-G haplotype and its association with ORMDL3 expression. Specifically, Stein et al. notes that the rs12936231-C and rs4065275-G alleles were associated with high expression of ORMDL3 in peripheral blood cells from non-African populations, but the rs4065275-G allele was not associated with expression of ORMDL3 in Yoruban (African) lymphoblastoid cell lines. This is possibly due to the rs12936231-C and rs4065275-G haplotype almost always being present in this combination on non-African haplotypes, while the rs12936231-G and rs4065275-G haplotype in non-Africans was rare (3%). In contrast, the rs12936231-G and rs4065275-G haplotype is common in Africans (19%)\textsuperscript{13}. We examined the rs12936231-rs4065275 haplotype for association with asthma in the three local ancestry groups (0, 1, or 2 copies of African ancestry), but no significant associations were observed in any of the groups (Supplementary Table 16).

Finally, we note genome-wide significant SNPs at the chromosome 17q12–q21 region in CAAPA ranged between positions 37,908,667–38,089,717 (Fig. 2a), which included the ORMDL3/GSDMB haplotype block (hg19, see Supplementary Table 9), at least 32 kb from the most significant SNP (at position 37,876,835) in TAGC. However, in the TAGC pediatric sub-group analysis the strongest association was observed 3.6 kb proximal to GSDMB, so we speculate the relatively strong association in CAAPA at the ORMDL3/GSDMB region reflects the large proportion of childhood onset asthmatics included in CAAPA (Supplementary Table 1). The ORMDL3/GSDMB region is strongly associated with childhood onset asthma\textsuperscript{13,34}, and to rule out that the weaker associations in high African ancestry populations may be due to the inclusion of adult onset asthmatics in some of the CAAPA studies, we re-tested the association between the 22 candidate SNPs and asthma, using only those studies where age of onset was available, or that were pediatric studies. We compared the association results when including versus excluding adult onset asthmatics, and the strength of the associations remained marginal when adult onset asthmatics were excluded (Supplementary Table 15).

### Table 3 Summary of tests for replication for two novel loci on chromosome 8 with \( p \)-values < \( 10^{-6} \) in the CAAPA meta-analysis

| SNP          | Ethnicity       | Study       | Number non-asthmatics | Number asthmatics | \( p \)-value | Effect direction |
|--------------|----------------|-------------|-----------------------|-------------------|--------------|------------------|
| rs13277810   | African ancestry| CAAPA       | 7009                  | 7645              | 3.20E-08     | +                |
|             | African American| CARDIA      | 109                   | 860               | 3.16E-01     | +                |
|             | African American| MESA        | 200                   | 1437              | 3.44E-01     | +                |
|             | African American| ARIC        | 89                    | 1636              | 5.24E-01     | +                |
|             | African American| BioMe       | 391                   | 1550              | 1.07E-01     | +                |
|             | Hispanic        | BioMe       | 519                   | 1775              | 7.15E-04     | +                |
|             | European        | TAGC        | 19,954                | 107,715           | 9.55E-02     | +                |
| rs13269769a  | African ancestry| CAAPA       | 7009                  | 7645              | 2.70E-07     | +                |
| rs114647118  | African ancestry| CAAPA       | 7009                  | 7645              | 2.70E-07     | +                |

rs13277810 is a common variant that reached genome-wide significance, and rs114647118 is a low frequency variant with MR-MEGA association \( p \)-value < \( 10^{-6} \). rs13277810 was not available in the TAGC meta-analysis summary statistics, but rs13269769, a SNP that has high LD with rs13277810 in the 1000 Genomes Project Europeans, was available. The rs114647118 association had low imputation quality in BioMe, and was not included in the TAGC meta-analysis summary statistic dataset. Effect direction is defined in terms of the minor allele. The CAAPA discovery associations are highlighted in bold font. *with rs13277810 in TGP EUR = 0.994

**Associations with total serum IgE.** We also examined whether genetic associations with asthma overlap with atopy by testing lead SNPs from Table 2 for association with total serum IgE (tIgE) using 4132 subjects for whom this phenotype was available (CAAPA lead SNPs and lead SNPs from TAGC that replicated in...
CAAPA; associations were tested separately in cases and controls and then combined using meta-analysis). Four of the SNPs correlated with levels of IgE in asthmatics (with the asthma risk allele associated with increased levels of IgE, Supplementary Table 19), suggesting that these asthma risk alleles may lead to an increased Th2 immune response. This includes SNP rs1420101 in the IL1R1 locus (a gene that is known to correlate with eosinophilia35), SNP rs10519067 intronic to the RORA gene (encoding a transcription factor that regulates the growth of group 2 innate lymphoids, a key cell type in the memory Th2 cell response36), and SNPs rs907092 and rs2952156 in the chr17q12–21 locus. These results are consistent with previous studies that have shown both shared and unique associations between these phenotypes37-41.

Admixture mapping. In addition to performing an asthma GWAS, we also leveraged local ancestry to identify regions of the genome where ancestry from a particular ancestral population was inherited more frequently in affected versus unaffected individuals. This technique, called admixture mapping, can be used as a complementary approach to association mapping in admixed populations to uncover associations not detectable by SNP tests alone42. Because the genetic structure of the Barbados population is similar to that of African Americans (Fig. 1a and Supplementary Fig. 1), and because of overlap in genotype array coverage, we combined the African American studies genotyped on the ADPC and the Barbados study (BAGS) and used this combined dataset for admixture mapping discovery. CAAPA studies genotyped on the MEGA, additional BAGS subjects genotyped on Illumina’s Omni array, and African American subjects from BioMe were leveraged to replicate our admixture mapping discoveries. The distribution of p-values from the discovery admixture mapping tests for association is summarized in Supplementary Fig. 20 (local ancestry dose association p-values reported by EMMAX). The QQ plot shows little evidence of 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 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 (see Methods). 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 20). 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. The most significant SNP showing association in this region in the admixture mapping discovery group (rs111966851, Supplementary Fig. 6) has a MAF of 0.308 and 0.006 in Africans and Europeans, respectively ($n = 1322$ AFR and $n = 1006$ EUR TGP phase 3), which corroborates the idea that African ancestry in this segment may increase risk for asthma. For replication, local ancestry segments overlapping with the peak segments were tested for association with asthma in CAAPA studies with similar genetic structure to the discovery studies (i.e., African Americans and Jamaicans, see population structure in Fig. 1a and Supplementary Fig. 1), additional BAGS (Barbados) subjects, and African American subjects from BioMe. None of these segments were associated with asthma (Supplementary Table 21), and SNP rs111966851 (the most significant SNP in this region in the admixture mapping discovery) was also not associated with asthma in the CAAPA studies excluded from the admixture mapping discovery (Supplementary Fig. 16).

Discussion
We report a large GWAS of asthma in African ancestry populations; prior studies included only 763–3037 asthmatic subjects. Eleven of the 18 loci recently identified in the TAGC meta-analysis show evidence of association in CAAPA, including strong evidence for four different regions. This includes the region around STAT6 on chromosome 12q13, a novel region identified by TAGC and not replicated to date, as well as the well-known chromosome 17q12–21 region, which reached genome-wide significance in our analysis. It has been posited that disparities in asthma susceptibility can partly be explained by genetic risk factors. In recapitulating associations mainly discovered in European ancestry populations (a result that has previously not been well-quantified in the literature), our results suggest that at the very least, common genetic variation may not strongly contribute to asthma disparities. However, our data show the chromosome 17q12–21 associations have smaller effect size magnitudes on African ancestry haplotypes. In addition, we built a genetic risk score for the 18 asthma loci reported by TAGC, and found that although asthmatics had a statistically significant higher risk score compared to controls (Supplementary Fig. 24), the effect was too small to build a predictive risk score for asthma in CAAPA.

In addition to recapitulating asthma genes discovered largely in non-African populations, we identified two loci on chromosome 8 not previously reported by asthma GWAS, and through admixture mapping identified a region on chromosome 6q22. In our inclusion of the largest sample of African ancestry individuals in our discovery GWAS, we were unfortunately limited in sample sizes of African ancestry individuals available for replication with consequent limitations in power. The most significant SNP on chromosome 8p23 reached genome-wide significance and was replicated in Hispanics from BioMe. However, our attempts to replicate this same locus in African Americans were unsuccessful; the number of cases compared to controls was considerably smaller (4–18 times smaller) as the replication studies were not primarily ascertainment for asthma. Similarly, we also failed to replicate the low frequency variant on chromosome 8q24 (only 398 African American asthmatics with genetic data were available) and the admixture mapping signal (only 845 cases were available, of which 498 were African American).

The CAAPA meta-analysis includes data from 15 independent studies and is the largest asthma GWAS focused on African ancestry populations to date. Unfortunately, as is the case for other complex diseases for which morbidity disproportionately affects underrepresented and underserved populations, a legacy of underrepresentation or exclusion of minorities from federally-funded studies has rendered comparatively robust non-European datasets rare to nonexistent. A recent analysis of ancestry represented in the GWAS Catalog concluded that non-European, non-Asian groups combined account for less than 4% of individuals included in the catalog. The analysis also found that African ancestry individuals contributed 7% of all catalog associations, despite only comprising 2.4% of the catalog, highlighting the value of GWAS conducted in African ancestry populations for enabling scientific discoveries. Furthermore, the authors stressed the importance of assessing the generalizability of genetic disease associations across populations, and the value that low-LD African ancestry individuals contribute to multi-ethnic fine-mapping of genetic associations. Despite the considerable federal support for the CAAPA initiative, we do recognize that the CAAPA sample size falls considerably short of the recent mega studies comprised of asthma datasets in the hundreds of thousands. Furthermore, replication had to be sought in studies not primarily ascertainment for asthma and with limited sample sizes. Important insights drawn from this study include the demonstration that many of the genetic loci associated with asthma in European ancestry populations may also be at play in African ancestry populations, and a clearer understanding of the LD patterns among African ancestry populations in 17q21. Potentially novel loci discovered by this meta-analysis are as yet not replicated within this study, but warrant follow-up.

Importantly, the advent of institutional biobanks with access to multi-ethnic patient populations, as well as efforts by institutions such as the National Institute of Health to reduce health and research disparities promise to greatly expand representation of well-characterized patients of African ancestry in the near future, allowing for robust follow-up of these CAAPA findings. The improved availability of African ancestry whole-genome sequence imputation reference panels available through initiatives such as the NHLBI-supported TOPMed program should also provide high quality imputation of low and rare frequency variation in African ancestry populations, which will empower future studies. Lastly, we note better availability of other -omics datasets representing diverse ethnicities, such as transcriptomic data in tissue types relevant to asthma, will be needed to enable discoveries by utilizing the next generation of analysis tools.
**Genome-wide ancestry estimation and analysis.** Unrelated phase 3 1000 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.38 (n = 43), were used as reference panels for the genome-wide ancestry analysis. A combined dataset of SNP's common to the reference panels plus all genotyped SNPs in the CAAPA datasets (with <1% missing genotypes in each sample) were created, after which a LD pruning step was performed (leaving 20,482 SNPs). (Note for the SAPPHIRE dataset, the combined genome-wide ancestry analysis described here was restricted to the 730 asthmatic cases for which ADPC [African Diaspora Power Chip] data were available.) Principal components were then formed from the genotypes of the CAAPA and reference subjects by the R Bioconductor package GENESIS. GENESIS uses PC-AiR to calculate principal components accounting for cryptic improvement if the SNP is added to the null model. GENESIS uses PC-relate to dosage and asthma, using the PLINK software package59,60. The GENESIS R statistics, logistic regression was used to test for association between imputed allelic this study, and shared summary statistics with CAAPA. To generate the summary NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-08469-7 | www.nature.com/naturecommunications

**Processing and assessment of individual association results.** First, association results of SNPs with low imputation accuracy were removed from the individual datasets. The chosen filter was informed by a study that quantified imputation accuracy in African Americans64 and filtering was done based on minor allele frequency (MAF) and the per-SNP estimation of the squared correlation between imputed allelic dosages and true unknown genotypes (Rsq). Associations between asthma and SNPs with MAF ≤0.005 were excluded if Rsq ≤0.5, as were associations with SNPs with MAF >0.005 if Rsq ≤0.3. SNPs with a minor allele count ≤10 were also excluded. Next, inflation factors and QQ plots were used to assess the individual study association results (Supplementary Figs. 3 and 4). Large inflation of these test statistics was not observed. All of the CAAPA datasets were therefore included in the meta-analysis, described below.

**Meta-analysis.** MR-MEGA uses multi-dimensional scaling to infer genetic axes of variation across studies, and models allelic effect across studies using a linear regression model of the allelic effect of each study along each genetic axis of variation, weighting the contribution of a study by the inverse variance of the allelic effect from the study. The model is described in detail in the Materials and Methods section of the MR-MEGA publication11. As input, the software requires the association odds ratio (OR) and its 95% confidence interval (CI), which were calculated per SNP and study as follows, where U is the Score statistic of the model fitted by GENESIS for the SNP, and i is its variance:

\[ OR = e^{\hat{U}/(2i)} \]

\[ 95\% CI = e^{\hat{U} + 1.96SE_{\hat{U}}}/(2i) \]

These formulas are based on recommendations from Zhou et al.62. For SAPPHIRE, the odds ratio reported by the software was used as is, and the 95% CI was calculated as follows, using the standard error (SE) reported by the software:

\[ 95\% CI = [\hat{U} - 1.96SE_{\hat{U}}]/(2i) \]

The MR-MEGA software requires specification of the number of genetic axes of variation to be used in the meta-analysis. Most of the CAAPA populations have ancestry from African and European populations, with some Native American ancestry (Fig. 1A). Because the mean Native American ancestry represented is small, however, and the first axis generally separates populations of high and low African ancestry (Supplementary Fig. 5), only 1 axis of genetic variation was used in the meta-analysis.

**Inverse-variance meta-analysis.** MR-MEGA estimates \( p \)-values by comparing the deviance of the regression model with coefficients equal to zero, to the deviance of the model with the coefficients unconstrained, and combined odds ratios/effect sizes across studies are not estimated. For this reason, all cross-study odds ratios reported in this paper, e.g., Table 2 and the forest plots in the supplementary material, were estimated using an inverse-variance meta-analysis. Inverse-variance meta-analysis was also used to combine association results of all the African American studies in order to generate Supplementary Table 4, combined SNP association results for the admixture mapping peak in the admixture mapping discovery dataset (Supplementary Fig. 6), and assessing associations on chromosome 17q12–21 in studies with high and low African ancestry (Supplementary Figs. 11 and 12, Supplementary Data 2).

**Replication in cardiovascular disease studies.** Imputed GWAS array data (described previously63) from three cardiovascular disease studies with asthma information were used to assess replication (see Supplementary Note 2 and Table 3). These summary statistics were generated using the logistic regression implemented in the R software packages) was used to test for association between allelic dose and asthma. Principal components associated with asthma (\( p \)-value < 0.1) were included as covariates in the models.

**Replication in BioMe.** Summary statistics of tests for association between GWAS array data and asthma in African Americans and Hispanics from BioMe were also used to assess replication (see Supplementary Note 2). The same summary statistics were generated using the logistic regression implementation in PLINK69. ImpG-Summary67 and the CAAPA WGS reference panel (Supplementary Table 3, Supplementary Note 3) were used to impute associations surrounding rs13277810, separately for African Americans and Hispanics (predicted \( R^2 \) for rs13277810 was high, 0.911 in both the African American and Hispanic datasets). Associations for the region surrounding the low frequency SNP rs11467118 were also imputed, but the imputation quality of this SNP, as well as other low frequency SNPs in high LD

---

**Statistical models fitted to the MEGA datasets.** Asthma association tests in the MEGA datasets used the same statistical models described for the GWAS + ADPC datasets. However, because some of the populations genotyped on the MEGA have a broader ancestry spectrum (Fig. 1A, Supplementary Fig. 1), the wide ancestry analysis pipeline included as fixed-effect covariates the first principal component, any other principal components identified by the elbow method (Supplementary Fig. 2) as explaining a large percentage of variance, as well as any of the top 20 principal components associated with asthma status [\( p \)-value < 0.05].

**Statistical models fitted to the GWAS + ADPC datasets.** Logistic mixed effects models were used to test for association between imputed allelic dosage and asthma, using the GENESIS R Bioconductor package. GENESIS uses a penalized quasi-likelihood approximation to the generalized linear mixed model. SNP association tests were performed using a score test, which tests for model fit improvement if the SNP is added to the null model. GENESIS uses PC-relate to estimate a kinship matrix excluding other sources of variance such as population structure38, and PC-AiR to calculate principal components accounting for cryptic and known relatedness between subjects38. The kinship matrix and principal components were calculated using a dataset of LD pruned genotyped SNPs. This kinship matrix was included as random effect in the null model, and principal components were included as fixed-effect covariates (the first principal component, as well as any of the top 10 principal components associated with asthma status [\( p \)-value < 0.05], were included).
with it, was insufficient to assess replication (predicted $r^2 = 0.586$ and 0.581 for all SNPs assessed, in the African American and Hispanic datasets, respectively).

TAGC SNP-by-SNP comparison. The CAAPA meta-analysis was compared to the TAGC multi-ethnic meta-analysis, the largest and most definitive asthma GWAS to date\textsuperscript{10}. The association results of the lead SNPs reported by TAGC were compared by extracting the corresponding CAAPA meta-analysis $p$-values, and using inverse-variance meta-analysis to calculate odds ratios and 95% CI for all SNPs (Table 2 and Supplementary Table 12). Note that two of the TAGC lead SNPs were not present in CAAPA: rs28581212 was filtered out of the CAAPA WGS reference panel because this SNP is located in a segmental duplication region, and rs11071558 was filtered out from the CAAPA reference panel because it is tri-allelic. Because the largest population represented in TAGC is European (19,954 of the 23,948 cases and 107,715 of the 118,538 controls), and because some of the CAAPA studies were included in the TAGC multi-ethnic meta-analysis, a direct comparison of association significance and effect size was only done for the TAGC European meta-analysis, and not the multi-ethnic meta-analysis. Together with the ancestry heterogeneity pattern described above, the TAGC summary statistics were downloaded and merged with the CAAPA studies included in the TAGC multi-ethnic meta-analysis, a direct comparison of association significance and effect size was only done for the TAGC European meta-analysis, and not the multi-ethnic meta-analysis. With the ancestry heterogeneity pattern described above, the TAGC summary statistics were downloaded and merged with the CAAPA meta-analysis results. Associations of all SNPs with significant fixed-effect $p$-values in TAGC Europeans were assessed in the CAAPA meta-analysis, after applying a Bonferroni correction for the number of overlapping SNPs in the merged dataset. A Bonferroni correction for the number of SNPs assessed in CAAPA was applied for stringent correction of the promising replication of loci reported by TAGC in CAAPA (Supplementary Table 13).

Additional replication in CAAPA. In addition to the SNP-by-SNP comparison described above, the TAGC summary statistics were downloaded and merged with the CAAPA meta-analysis results. Associations of all SNPs with significant fixed-effect $p$-values in TAGC Europeans were assessed in the CAAPA meta-analysis, after applying a Bonferroni correction for the number of overlapping SNPs in the merged dataset. A Bonferroni-corrected significance threshold for 20 tests (one for each of these prior loci) was used as a measure of additional evidence of replication in the CAAPA meta-analysis (Fig. 1c and Table 2, loci not in bold font).

LD block replication in CAAPA. SNPs within the same LD block and in high LD with the TAGC lead SNPs ($r^2 > 0.8$) in Europeans from the TGP were also selected for comparison, and their association results were inspected (Supplementary Table 18). In this way, no or marginal replication of TAGC lead SNPs in CAAPA due to differences in LD patterns between Europeans and Africans can be assessed. The LD block surrounding each lead TAGC SNP was identified using TGP phase 3 data, population EUR, representing European ancestry. Thus, the effect of SNPs that may be causal in Europeans but not the lead SNP reported in the TAGC European analysis was tested for replication in CAAPA. For SNPs $\pm 10$ KB from each TAGC lead SNP with $\text{MAF} > 0.05$ in the CAAPA WGS reference panel and that intersected with the CAAPA meta-analysis association results as well as TGP phase 3 variants were selected, and Gabriel’s algorithm implemented in Haploview\textsuperscript{80} was then used to identify LD blocks present in the window of SNPs. CAAPA meta-analysis $p$-values for SNPs falling in the LD block containing the lead SNP, and that had $r^2 > 0.8$ with the lead SNP, were extracted, and these SNPs and their asthma associations are summarized in Supplementary Table 18. The chr17q12–21 locus achieved genome-wide significance in CAAPA and was therefore excluded from this analysis.

EVE replication sensitivity analysis. A sensitivity analysis of the African ancestry-specific genome-wide association reported by the EVE consortium for SNP rs11019999\textsuperscript{8} was also done (Supplementary Table 12). CAAPA studies included in the EVE meta-analysis were meta-analyzed together, and all CAAPA studies not included in the EVE meta-analysis were meta-analyzed together. MR-MEGA was used for this meta-analysis, as CAAPA studies with relatively high proportions of European ancestry should be included, without penalizing the strength of the association should these high European ancestry populations show a different pattern of association compared to populations with high African and low European ancestry.

Tests for association with total serum IgE. Lead SNPs from Table 2 (reported by CAAPA, as well as TAGC SNPs with evidence for replication in CAAPA) were tested for association with total serum IgE (tlgE)\textsuperscript{71} by adopting the approach in the EVE consortium’s meta-analysis of genetic association with tlgE\textsuperscript{70}. The analysis was stratified based on asthma case-control status and originating study (eight groups of asthmatics and six groups of controls, see Supplementary Table 1 for a summary of the distribution of tlgE in these groups). Linear mixed-effect models implemented in the GENESIS software package were used for association testing between the allelic dose at each SNP and Studentized residuals of log10 transformed tlgE (adjusted for age and sex). Similar to the asthma association models, the models included a kinship matrix as random effect and principal components as fixed-effect covariates. Finally, association test statistics from the analysis strata were combined using inverse-variance meta-analysis, to yield combined statistics for non-asthmatics, asthmatics, and non-asthmatics + asthmatics (Supplementary Table 19).

Admixture mapping discovery. Local ancestry inference was performed using RFMix\textsuperscript{72}, and the pipeline is described in Supplementary Note 12. 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 SAPPPIRE, which included only asthmatics and no controls as well as the three population outliers identified by the genome-wide ancestry analysis, were set to missing. These 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\textsuperscript{71}. 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 was 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 $\sum_{i=1}^{k} \text{eigenvalues} / \sum_{i=1}^{n} \text{eigenvalues} < 0.995$). The total number of effective tests was 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) $\approx 1.9 \times 10^{-4}$ was used to claim statistical significance.

Admixture mapping replication. The segments including the start and end position of the admixture mapping discovery peak (chr1:134,149,974–134,300,365) were selected for replication. Logistic regression was used to test for association between the number of copies of African ancestry (dosage value of 0, 1, 2) and case-control status, separately for each replication dataset (BASS, BAGS, JAAS, and BioMe). The base R package was used to fit the model to the BioMe data, adjusting for the first and fifth principal components, as these principal components were also associated with asthma status ($p$-values $< 0.1$). The R Bioconductor package GENESIS was used to fit models to the BASS, BAGS, and JAAS dataset, including a kinship matrix as random effect and principal components as fixed effects, as described for the SNP association analysis. The results for the segment including/ closest to the midpoint of the admixture mapping discovery peak (134,225,170) was combined using inverse-variance meta-analysis.

Admixture mapping power calculations. We used the Genetic Association Study (GAS) Power Calculator\textsuperscript{72} to perform post-hoc power calculations for admixture mapping. As BAGS and HUFS include related subjects, their effective number of cases and controls were estimated using TGP phase 3 data, population EUR, representing European ancestry. The effective number of cases and controls were estimated using TGP phase 3 data, population EUR, representing European ancestry. For admixture mapping power calculations, the number of effective tests for a chromosome was then set to the sum of the number of effective tests calculated for each of the 22 autosomes. For the GAS Power Calculator, the effective number of cases and controls was estimated as having 0, 1, or 2 copies of African ancestry (dosage value of 0, 1, 2) and case-control status, separately for each replication dataset (BASS, BAGS, JAAS, and BioMe). The base R package was used to fit the model to the BioMe data, adjusting for the first and fifth principal components, as these principal components were also associated with asthma status ($p$-values $< 0.1$). The R Bioconductor package GENESIS was used to fit models to the BASS, BAGS, and JAAS dataset, including a kinship matrix as random effect and principal components as fixed effects, as described for the SNP association analysis. The results for the segment including/ closest to the midpoint of the admixture mapping discovery peak (134,225,170) was combined using inverse-variance meta-analysis.

Data availability. The legacy GWAS array ADPC and MEGA data that support the findings of this study have been deposited in dbGAP with the accession code phs001123.v2.p1. These data can be accessed through dbGAP. Specific data use limitations: GRU–IRB (Gene逞ory U,. Use, IRB approval required). Only ADPC data are available for the SAGE II GALA II, and SAPPHIRE datasets; in addition, no phenotype data are available for these three datasets. Summary statistics from the meta-analysis are also available through the GWAS catalog\textsuperscript{12} [https://www.ncbi.nlm.nih.gov/gwas/downloads/summary-statistics].

Received: 21 June 2018 Accepted: 8 January 2019
Published online: 20 February 2019
References

1. Moorman, J. E. et al. National surveillance of asthma: United States, 2001-2010. Vital Health Stat. 3, 1–58 (2012).
2. Moorman, J. E. et al. National surveillance for asthma—United States, 1980–2004. Mmwr. Surveill. Summ. 56, 1–54 (2007).
3. Keet, C. A. et al. Neighborhood poverty, urban residence, race/ethnicity, and asthma: rethinking the inner-city asthma epidemic. J. Allergy Clin. Immunol. 129, 555–662 (2012).
4. Mathias, R. A. et al. A genome-wide association study on African-ancestry populations for asthma. J. Allergy Clin. Immunol. 125, 336–346 e4 (2010).
5. Almoguera, B. et al. Identification of four novel loci in asthma in European American and African American populations. Am. J. Respir. Crit. Care. Med. 195, 456–463 (2017).
6. Torgeson, D. G. et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat. Genet. 43, 887–892 (2011).
7. Demenais, F. et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer markers. Nat. Genet. 50, 42–53 (2018).
8. White, M. J. et al. Novel genetic risk factors for asthma in African American children: precision medicine and the SAGE II study. Immunogenetics 68, 391–400 (2016).
9. Mathias, R. A. et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat. Commun. 7, 12522 (2016).
10. Jochums, H. R. et al. Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome. Sci. Rep. 7, 46398 (2017).
11. Magi, R. et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. Hum. Mol. Genet. 26, 3639–3650 (2017).
12. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D901 (2017).
13. Stein, M. M. et al. A decade of research on the 17q12–21 asthma locus: piecing together the puzzle. J. Allergy Clin. Immunol. 142, 749–764 (2018).
14. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-08469-7 | www.nature.com/naturecommunications
15. Godava, M., Vrtel, R. & Vodicka, R. STAT6—polymorphisms, haplotypes and epistasis in relation to atopy and asthma. Biomed. Pap. Med. Fac. Univ. Palacky. Olomouc Czech. Repub. 157, 172–180 (2013).
16. Yang, Q. et al. A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. Eur. Respir. J. 49, 160195 (2017).
17. Kreiner-Moller, E. et al. 17q21 gene variation is not associated with asthma in adulthood. Allergy 70, 107–114 (2015).
18. Einaudi, S. et al. Identification of genes expressed by human airway eosinophils after an in vivo allergen challenge. PloS One 8, e67560 (2013).
19. Lima, L. C. et al. Genetic variants in RORA are associated with asthma and allergy markers in an admixed population. Cytokine 113, 177–184 (2019).
20. Sunyer, J., Antó, J. M., Castellsague, J., Soriano, J. B. & Roca, J. Total serum IgE is associated with asthma independently of specific IgE levels. The Spanish Group of the European Study of Asthma. Eur. Respir. J. 9, 1880–1884 (1996).
21. Sears, M. R. et al. Relation between airway responsiveness and serum IgE in children with asthma and in apparently normal children. N. Engl. J. Med. 325, 1067–1071 (1991).
22. Palmer, L. J. et al. Independent inheritance of serum immunoglobulin E concentrations and airway responsiveness. Am. J. Respir. Crit. Care. Med. 161, 1836–1843 (2000).
23. Levin, A. M. et al. A meta-analysis of genome-wide association studies for serum total IgE in diverse study populations. J. Allergy Clin. Immunol. 131, 1176–1184 (2013).
24. Burrows, B., Martinez, F. D., Halonen, M., Barber, R. A. & Cline, M. G. Association of asthma with serum IgE levels and skin-test reactivity to allergens. N. Engl. J. Med. 320, 271–277 (1989).
25. Galindo, M. et al. Genomewide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latin Americans study. J. Allergy Clin. Immunol. 134, 295–305 (2014).
26. Vaillancourt, V. T., Bordeleau, M., Laviolette, M. & Laprise, C. From expression pattern to genetic association in asthma and asthma-related phenotypes. BMC Res. Notes 5, 630 (2012).
27. Gould, W. et al. Factors predicting inhaled corticosteroid responsiveness in African American patients with asthma. J. Allergy Clin. Immunol. 126, 1131–1138 (2010).
28. Sulovari, A., Chen, Y. H., Hudziak, J. J. & Li, D. Atlas of human diseases influenced by genetic variants with extreme allele frequency differences. Hum. Genet. 136, 39–54 (2017).
29. Morales, J. et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. Genome Biol. 19, 21 (2018).
30. West, K. M., Blacksher, E. & Burke, W. Genomics, health disparities, and missed opportunities for the nation’s research agenda. JAMA 317, 1831–1832 (2017).
31. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. Nature 538, 161–164 (2016).
32. Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat. Genet. 49, 1752–1757 (2017).
33. Wolford, B. N., Willer, C. J. & Surakka, I. Electronic health records: the next wave of complex disease genetics. Hum. Mol. Genet. 27, R14–R21 (2018).
34. National Institutes of Health. Health Disparities (NIH, UK, 2010).
35. Brody, J. A. et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. Nat. Genet. 49, 1560–1563 (2017).
36. SchlauCh, D., Glass, K., Hersh, C. P., Silverman, E. K. & Quackenbush, J. Estimating drivers of cell state transitions using gene regulatory network models. BMC Syst. Biol. 11, 139 (2017).
37. Ferreira, M. A. et al. Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling. J. Allergy Clin. Immunol. 139, 1148–1157 (2017).
38. GamaZon, E. R. et al. Genome-wide association method for mapping traits using reference transcriptome data. Nat. Genet. 47, 1091–1098 (2015).
39. Mao, X. et al. A genomewide admixture mapping panel for Hispanic/Latino populations. Am. J. Hum. Genet. 80, 1171–1178 (2007).
40. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet. Epidemiol. 39, 278–293 (2015).
41. Conomos, M. P., Reiner, A. P., Weiss, B. S. & Thornton, T. A. Model-free estimation of recent genetic relatedness. Am. J. Hum. Genet. 98, 127–148 (2016).
42. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
43. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
Acknowledgements

We thank Gonzalo Abecasis for coordinating inclusion of the CAAPA reference panel on the Michigan Imputation Server, Todd Deppe, Estelle Giraud, Cindy Lawley from Illumina for genotyping services, and Pat Oldewurtel for administrative and technical support.

Author contributions

M.D., N.R., and T.M.B. analysed the data, interpreted the data, and wrote the paper. S.C., A.M.L., A.S., M.P.B., G.W., H.R.J. and G.B. analysed and interpreted the data. M.C. performed genotyping. K.C.B., R.A.M., T.H.B. and M.A.T. conceived the experiments, interpreted the data, and wrote the paper. C.V., C.R.G., V.E.O., A.D., D.G.T., N.A., M.I.A., P.C.A., E.B., C.B., L.C., A.C., G.M.D., C.E., M.U.F., T.S.F., C.F., J.G.F., W.G., P.A.G., N.N.H., R.D.H., E.F.H., S.J., E.E.K., J.K., R.K., L.A.L., E.M.L., A.I., P.M., T.M., A.M., D.M., D.L.N., T.D.O., R.R.O., C.O.O., O.O., Z.S.Q., C.R., N.V., H.W., R.J.W., I.G.W., S.S., C.O., E.G.B., L.K.W. and I.R. contributed to interpretation of results and critically reviewed the manuscript. CAAPA provided infrastructure, biospecimens and associated phenotype data, and intellectual input for the overall design and execution of the study.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08469-7.

Competing interests: All authors declare no competing interests.

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/
CAAPA

Ayola Akim Adegnika, Ganiyu Arinola, Ulysses Ateba-Ngoa, Gerardo Ayestas, Hrafnhildur Bjarnadóttir, Adolfo Correa, Said Omar Leiva Erazo, Marilyn G. Foreman, Cassandra Foster, Li Gao, Jingjing Gao, Leslie Grammer, Mark Hansen, Tina Hartert, Yijuan Hu, Iain Königsberg, 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, Maureen Samms-Vaughan, Robert Schleimer, Alan F. Scott, Suyash S. Shringarpure, Wei Song, Zachary A. Szpiech, Raul Torres, Gloria Varela, Olga Marina Vasquez, Francisco M. De La Vega, Lorraine B. Ware & Maria Yazdanbakhsh

Centre de Recherches Médicales de Lambaréné, BP:242, Lambaréné 13901, Gabon. 42Department of Chemical Pathology, University of Ibadan, Ibadan 900001, Nigeria. 43Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. 44Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA. 45Pulmonary and Critical Care Medicine, Morehouse School of Medicine, Atlanta, GA 30310, USA. 46Data and Statistical Sciences, AbbVie, North Chicago, IL 60064, USA. 47Department of Medicine, Vanderbilt University, Nashville, TN 37232, USA. 48Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA. 49Instituto de Investigaciones Immunológicas, Universidad de Cartagena, Cartagena 130000, Colombia. 50Facultad de Ciencias de la Salud, Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula 21102, Honduras. 51Department of Child Health, The University of the West Indies, Kingston 00007, Jamaica. 52Department of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA. 53Biomedical Sciences Graduate Program, University of California San Francisco, San Francisco, CA 94158, USA. 54Centro Medico de la Familia, San Pedro Sula 21102, Honduras. 55Department of Parasitology, Leiden University Medical Center, Leiden 02333, Netherlands.