Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma

Catherine Igartua\(^1\), Rachel A. Myers\(^1\), Rasika A. Mathias\(^2\), Maria Pino-Yanes\(^3,4\), Celeste Eng\(^3\), Penelope E. Graves\(^5\), Albert M. Levin\(^6\), Blanca E. Del-Rio-Navarro\(^7\), Daniel J. Jackson\(^8\), Oren E. Livne\(^1\), Nicholas Rafael\(^9\), Christopher K. Edlund\(^10\), James J. Yang\(^11\), Scott Huntsman\(^3\), Muhammad T. Salam\(^10\), Isabelle Romieu\(^12\), Raphael Mourad\(^1\), James E. Gern\(^8,13\), Robert F. Lemanske\(^8,14\), Anna Wyss\(^15\), Jane A. Hoppin\(^15,16\), Kathleen C. Barnes\(^2\), Esteban G. Burchard\(^3,17\), W. James Gauderman\(^10\), Fernando D. Martinez\(^5\), Benjamin A. Raby\(^18,19\), Scott T. Weiss\(^18\), L. Keoki Williams\(^20,21\), Stephanie J. London\(^15\), Frank D. Gilliland\(^10\), Dan L. Nicolae\(^1,22\) & Carole Ober\(^1\)

Common variants at many loci have been robustly associated with asthma but explain little of the overall genetic risk. Here we investigate the role of rare (<1%) and low-frequency (1−5%) variants using the Illumina HumanExome BeadChip array in 4,794 asthma cases, 4,707 non-asthmatic controls and 590 case–parent trios representing European Americans, African Americans/African Caribbeans and Latinos. Our study reveals one low-frequency missense mutation in the GRASP gene that is associated with asthma in the Latino sample (\(P = 4.31 \times 10^{-6}\); OR = 1.25; MAF = 1.21%) and two genes harbouring functional variants that are associated with asthma in a gene-based analysis: GSDMB at the 17q12−21 asthma locus in the Latino and combined samples (\(P = 7.81 \times 10^{-8}\) and 4.09 \(\times 10^{-8}\), respectively) and MTHFR in the African ancestry sample (\(P = 1.72 \times 10^{-6}\)). Our results suggest that associations with rare and low-frequency variants are ethnic specific and not likely to explain a significant proportion of the ‘missing heritability’ of asthma.
Asthma is a common, chronic inflammatory disease of the airways with typical onset in childhood. Heritability estimates indicate that approximately half the variation in risk is attributable to genetic factors; yet, the common variants identified by genome-wide association studies (GWAS) account for very little of the genetic risk. This so-called ‘missing heritability’ in asthma and other common diseases has been attributed to many potential causes that generally fall into two categories. First, the genetic variants interrogated by GWAS may not capture all relevant variation. In particular, rare variants, which comprise the bulk of genetic variation in the human genome and are predicted to have larger phenotypic effects than common variants, are not represented or tagged by the single-nucleotide polymorphisms (SNPs) that are included on the genotyping arrays typically used in GWAS. Second, the statistical approaches used in GWAS may be overly simplistic and not adequately model the genetic architecture of asthma, which is probably polygenic and includes many gene–environment interactions.

In this study, we begin to address the first limitation of GWAS by investigating the role of rare (minor allele frequency (MAF) <1%) and low-frequency (MAF 1–5%) variants in asthma. We conducted this meta-analysis study in the largest ethnically diverse sample of asthma cases, non-asthmatic controls and case–parent trios, which include 13 studies from 8 groups of American, African American/African Caribbean and Latino (European American, African American/African Caribbean and Latino) within each of three major ethnic/racial groups (European Americans, African Americans, African Caribbeans, Mexican Americans, Mexicans and Puerto Ricans (Table 1). We tested each low-frequency variant, defined as having a MAF between 1% and 5% within each ethnic group. For the meta-analysis in the combined sample, we excluded variants that were either common (MAF ≥5%) in all three ethnic groups or rare (<1%) in all three ethnic groups (Supplementary Table 1 and Supplementary Table 2). This definition allows us to include SNPs that may be common in one population but rare in the other(s). The quantile–quantile plots for the analyses of low-frequency variants indicated that there was no inflation of test statistics (Supplementary Fig. 1). One missense mutation (c.C1093G, p.R365V) in the GRASP gene was associated with asthma risk in the Latinos after correcting for multiple testing (9,519 tests; Genome-wide Efficient Mixed Model Association (GEMMA)/Family Based Association

### Results

**Single-variant association analyses.** We first evaluated the individual effects of low-frequency, putatively functional exonic variants on asthma risk in 11,225 subjects representing European Americans, African Americans, African Caribbeans, Mexican Americans, Mexicans and Puerto Ricans (Table 1). We tested each low-frequency variant, defined as having a MAF between 1% and 5% within each ethnic group. For the meta-analysis in the combined sample, we excluded variants that were either common (MAF ≥5%) in all three ethnic groups or rare (<1%) in all three ethnic groups (Supplementary Table 1 and Supplementary Table 2). This definition allows us to include SNPs that may be common in one population but rare in the other(s). The quantile–quantile plots for the analyses of low-frequency variants indicated that there was no inflation of test statistics (Supplementary Fig. 1). One missense mutation (c.C1093G, p.R365V) in the GRASP gene was associated with asthma risk in the Latinos after correcting for multiple testing (9,519 tests; Genome-wide Efficient Mixed Model Association (GEMMA)/Family Based Association

### Table 1 | Sample composition.

| Sample       | Case | Control | Male | Female | Total |
|--------------|------|---------|------|--------|-------|
| **European American** |      |         |      |        |       |
| ALHS         | 776  | 802     | 859  | 719    | 1,578 |
| CAG          | 186  | 209     | 172  | 223    | 395   |
| CAMP*        | 376  | 706     | 588  | 494    | 1,082 |
| CHS          | 415  | 640     | 546  | 509    | 1,055 |
| COAST        | 134  | 119     | 150  | 103    | 253   |
| Total        | 1,511| 1,770   | 2,315| 2,048  | 4,363 |
| **African Ancestry** |      |         |      |        |       |
| CAG          | 274  | 200     | 170  | 304    | 474   |
| GRADD-AA     | 236  | 288     | 199  | 325    | 524   |
| GRADD-AC†    | 227  | 297     | 250  | 274    | 524   |
| SAPPHIRE     | 553  | 233     | 278  | 508    | 786   |
| Total        | 1,290| 1,018   | 897  | 1,411  | 2,308 |
| **Latino**   |      |         |      |        |       |
| CHS          | 581  | 775     | 713  | 643    | 1,356 |
| GALA II-MA   | 568  | 604     | 568  | 604    | 1,172 |
| GALA II-PR   | 844  | 540     | 731  | 653    | 1,384 |
| MCAAS*       | 214  | 428     | 337  | 305    | 642   |
| Total        | 1,993| 1,919   | 2,349| 2,205  | 4,554 |
| **Combined** |      |         |      |        |       |
| Total        | 4,794| 4,707   | 5,561| 5,664  | 11,225|

ALHS, Agricultural Lung Health Study; CAG, Chicago Asthma Genetics Study; CAMP, Childhood Asthma Management Program; CHS, The Children’s Health Study; COAST, The Childhood Origins of Asthma Study; GALA II, Genes-Environment and Admixture in Latino Asthmatics (-MA, Mexican American; -PR, Puerto Rican); GRAAD, Genomic Research on Asthma in the African Diaspora and Barbados (-AA, African Americans; -AC, African Caribbeans); MCCAS, Mexico City Childhood Asthma Study; SAPPHIRE, The Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-Ethnicity.

Descriptions of the individual study samples have either been published or are described in Supplementary Note 1.

*C*Case-parent trios.

†Samples from 164 families; kinship matrix was used to address family structure as implemented in GEMMA.
Test (FBAT) meta-analysis, \( P = 4.31 \times 10^{-6}; N = 4,554\); odds ratio (OR) = 1.25; MAF = 0.012). Six additional low-frequency variants were modestly associated at suggestive levels of significance \((P < 10^{-3})\) in at least one of the three groups or in the combined sample (Table 2). Of these, five missense mutations (one each in FAM134B, NPC1L1, IKZF1, SLC26A3 and GSDMB) showed evidence of association with asthma in the combined sample, two of which (IKZF1 and SLC26A3) were more significant in the African ancestry sample. The low-frequency variants at FAM134B (c.G1145C, p.S382T; transcript accession NM_001042471), NPC1L1 (c.C919T, p.R307X; transcript accession NM_001042471) were associated with asthma. Second, this gene encodes a protein that interacts PDE4DIP, two asthmatic and two non-asthmatic children also reported a noteworthy for two reasons. First, an exome-sequencing study sample, two of which (GEMMA, \( P \leq 0.05\)) showed evidence of association with asthma in the combined sample. Of these, five missense mutations (one each in NPC1L1, FAM134B, and IKZF1) segregating with asthma. Second, this gene encoding a protein that interacts with PDE4D, a gene that was identified in a GWAS of asthma. Our results further implicate this pathway in asthma pathogenesis.

Table 2 | Results of meta-analyses of single SNP association studies of low-frequency functional variants and asthma risk with association \( P \)-values \(< 10^{-4}\).

| rsID (Chr: position) | Gene | Residue change | Alleles | GEMMA/FBAT meta-analysis \( P \)-values | MAF | OR (95% CI) |
|----------------------|------|----------------|---------|--------------------------------------|-----|-------------|
|                      |      |                | COMB (N = 11,225) | EA (N = 4,363) | AA (N = 2,308) | LAT (N = 4,554) | EA | AA | LAT | COMB | EA | AA | LAT |
| rs14716868 (1:144930790) | PDE4D | c.C919T, p.R307X | T/C | 1.48 \times 10^{-5} | 9.76 \times 10^{-5} | 1.22 \times 10^{-4} | 3.41 \times 10^{-3} | 1.12 | (1.04-1.21) | — | 1.17 | (1.08-1.28) | 0.89 | (0.81-0.97) |
| rs67313818 (5:16475299) | FAM134B | c.A907C, p.I303L | G/C | 1.30 \times 10^{-5} | 6.56 \times 10^{-4} | 3.30 \times 10^{-4} | 0.0175 | 0.0146 | 0.92 | (0.89-0.95) | — | — | 0.89 | (0.81-0.97) |
| rs149782907 (7:44561344) | NPC1L1 | c.C2920T, p.P974S | G/A | 6.55 \times 10^{-5} | 1.68 \times 10^{-5} | 1.28 \times 10^{-3} | 0.0109 | 0.0146 | 0.74 | (0.68-0.87) | — | — | 0.65 | (0.65-0.85) |
| rs957693481 (7:45047774) | SLC26A3 | c.A365G, p.L365V | G/A | 7.89 \times 10^{-5} | 4.97 \times 10^{-5} | 4.87 \times 10^{-4} | 0.0475 | 0.0457 | 1.12 | (1.06-1.21) | 1.12 | (1.07-1.21) | 1.25 | (1.14-1.38) |
| rs7974462 (1:144930790) | GRASP | c.G1145C, p.S382T | G/C | 0.0543 | 0.99 | 4.31 \times 10^{-4} | 0.0337 | 0.0059 | 0.021 | (0.99-1.01) | 0.77 | (0.71-0.89) | 1.07 | (1.03-1.1) |
| rs12450091 (5:16475299) | GSDMB | c.G1145C, p.S382T | G/A | 5.56 \times 10^{-5} | 2.31 \times 10^{-3} | 2.11 \times 10^{-3} | 0.1945 | 0.07 | 0.107 | (0.99-1.01) | 0.98 | (0.91-1.04) | 1.07 | (1.03-1.1) |

AA, African Ancestry; Chr, chromosome; CI, confidence interval; COMB, combined; EA, European American; LAT, Latino; MAF, minor allele frequency; OR, odds ratio. Positions are from NCBI build 37. For residue changes, the transcript accession number is presented in parenthesis when multiple transcripts exist. Alleles are presented as minor allele/major allele.

Ethnic specific meta-analyses were performed for all low-frequency variants (MAF <1%; 8,249 in European American, 17,861 in African American and 9,519 in Latinos); the combined meta-analysis included variants that were neither common (≥5%) nor rare (<3%) in all three ethnic groups (32,689 variants). ORs are calculated for case-control studies only (case-parent trio studies excluded) and presented for the minor allele only. All the association with the GRASP variant in Latinos was significant after Bonferroni correction (9,519 variants, GEMMA/FBAT meta-analysis \( P < 4.31 \times 10^{-4}\)).

Power analyses for low-frequency variants. After testing for nearly 33,000 functional low-frequency exonic variants in 11,225 subjects, only 1 variant remained statistically significant in Latinos after correcting for testing of 9,519 of variants and only 6 were modestly associated at \( P \)-values \(< 10^{-4}\) in the African ancestry or the combined sample. To assess whether the paucity of associations was due to lack of power, we performed simulations to determine the power of our studies to detect associations with asthma for low-frequency variants. We considered ORs of 1.5, 2, 2.5, 3 and 4, within the expected ranges of effect sizes for

with the common variants at the 17q12–21 locus that have been associated with asthma in previous GWAS\(^5,19,20\), we performed conditional analyses of rs12450091 in the case–control samples only (case–parent trio studies were excluded and meta \( P \)-value recalculated: Latino GEMMA meta-analysis \( P = 6.89 \times 10^{-5}; N = 3,912\); combined GEMMA meta-analysis \( P = 5.52 \times 10^{-5}; N = 9,501\)). For this analysis, we first identified three common variants (one synonymous, one missense and one intronic) at this locus that were included on the exome array and associated with asthma in previous reports (G at rs907092 (refs 4,5,19, G at rs2305480 (refs 5,20) and T at rs7216389 (refs 5,21,22)). In Latinos, the risk allele at rs12450091 (C) tags 28% of the high-risk \( GGT \) haplotypes (\( GGT \) frequency 18.09%, \( GTT \) frequency 45.84%) and is not in LD with any of the known expression quantitative trait loci reported at this locus\(^2,12,21-24\). We then repeated the association studies with rs12450091 (C) in GSDMB in three separate analyses, including each of the common variants as a covariate in each analysis to 'remove' the effects of the common variant on the association signal. In each of these three analyses, the evidence for association between rs12450091 and asthma was reduced from \( P = 6.89 \times 10^{-5} \) to \( P \leq 0.022 \) in Latinos and from \( P = 5.52 \times 10^{-5} \) to \( P \leq 0.019 \) in the combined sample (Table 3). These results suggest that although most of the evidence for association with this variant was shared with the associated common variants, the residual signal in conditional analyses indicates that this missense substitution or a variant in strong LD with the missense substitution contributes additional, independent risk to asthma in Latinos.

Conditional analyses at the 17q12–21 locus. To determine whether the association with this variant (C; A365G, p.E122G; rs12450091) in Latinos and the combined sample represents independent risk for asthma or is in linkage disequilibrium (LD)
low-frequency variants with moderate-to-high effects. In fact, we had at least 80% power to detect associations with variants with frequencies ≥1% and ORs ≥2.5 in the combined sample. In the European American and Latino samples, we had 80% power to detect variants with frequencies ≥1% and ORs ≥4. We had lowest power in the African ancestry sample, in which there was 80% power to detect variants with frequencies >2% and ORs ≥4 (Supplementary Fig. 2). Thus, the lack of significant associations with low-frequency variants in our study could be due to low power in the ethnic/racial groups, small effect sizes of variants in this frequency range, or both.

**Gene-based analyses.** Studies of rare variants have virtually no power to detect individual associations in samples of even many thousands of individuals.\(^{25-27}\) Rather, gene-based tests that combine the evidence for association among all variants within each gene provide more power to detect associations. To test the hypothesis that the overall burden of functional variants at each gene confers risk for asthma, we performed a gene-based association test as implemented in MetaSKAT-O\(^{28}\) for each of the 10,208 genes represented on the exome array that had three or more functional variants present in at least two case-control samples. We included in these analyses all putatively functional variants on the array regardless of allele frequency. We performed two sets of analyses. In the first we weighted all variants equally, and in the second we gave larger weights to rarer variants (MetaSKAT-O default). For those analyses to be performed, we considered all potentially functional variants (missense, nonsense and splice site) and then considered only variants predicted to be 'probably damaging' by Polymorphism Phenotyping v2 (PolyPhen-2).\(^{29}\) In the analysis with equal weights and defining functional variants more broadly, associations with GSDMB and ZBPB2, two genes at the 17q12–21 asthma locus, were significant in the combined (MetaSKAT-O: \(P = 6.10 \times 10^{-10}\), 16 variants and \(P = 1.34 \times 10^{-6}\)), 7 variants, respectively; \(N = 9,501\) and Latino (MetaSKAT-O: \(P = 1.02 \times 10^{-6}\), 12 variants and \(P = 1.73 \times 10^{-6}\), 7 variants, respectively; \(N = 3,912\)) samples after Bonferroni correction for 9,534 and 10,439 genes, respectively. In addition, the MTHFR gene was significant (MetaSKAT-O: \(P = 1.72 \times 10^{-6}\), 11 variants; \(N = 2,308\)) in the African ancestry sample after multiple testing correction for 10,342 genes. When we considered only variants with PolyPhen-2 prediction of 'probably damaging', the association with GSDMB remained significant (combined MetaSKAT-O: \(P = 4.09 \times 10^{-6}\), 7 variants, \(N = 9,501\); Latino MetaSKAT-O: \(P = 7.81 \times 10^{-8}\), 5 variants, \(N = 3,912\), but the associations with the ZBPB2 and MTHFR genes did not. No genes were significant when larger weights were applied to rare variants in either of the analyses (Table 4, Supplementary Table 3, Supplementary Figs 3–6), suggesting that these gene-based associations are driven largely by the more common functional variants that are present on the array. In fact, when we remove one common missense SNP in GSDMB (c.G844A, p.G282R; rs2305479; MAF 0.335 in Latinos), the evidence for association of this gene in the test that considered only PolyPhen-2 predicted damaging variants becomes nonsignificant (equal weights; MetaSKAT-O: \(P = 1.0\) in Latinos, \(N = 3,912\) and \(P = 0.276\) in the combined sample, \(N = 9,501\)). Similar to GSDMB, the associations with ZBPB2 were no longer significant after removing the one common missense variant (c.G518T p.S173I; rs11557467; MAF 0.357 in Latinos) and testing the association of the remaining six variants in the Latino (MetaSKAT-O: \(P = 0.54\), \(N = 3,912\)) and the combined sample (MetaSKAT-O: \(P = 0.12\), \(N = 9,501\)). This suggests that the association with ZBPB2 is driven by this common variant. In contrast, removing the one common missense variant in MTHFR (c.C665T p.A222V; rs1801133; MAF = 0.11 in African ancestry) does not eliminate the gene-based association signal (nine rare, one low-frequency variant; MetaSKAT-O: \(P = 0.002\), \(N = 2,308\)), indicating that the association with variants in the MTHFR gene is attributable to common, low-frequency and rare variants.

**Assessing coverage of exonic variation on the exome array.** Lastly, we assessed how well the variants on the exome array captured the low-frequency and rare functional variation present in the genomes of asthmatic individuals. We had available whole-genome sequences (WGS) for 278 asthmatics (101 European Americans, 93 African Americans and 84 Latinos) as part of ongoing studies of the EVE Consortium,\(^{1,10}\) including 172 of the individuals in this study (43 European Americans, 44 African Americans and 85 Latinos). We selected all functional exonic variants with MAF<5% in any of the three groups from the WGS and compared the overlap between the missense, nonsense and splice site mutations present in the WGS with the variants on the exome array. Overall, there were 105,943 rare or low-frequency variants in 16,359 genes in the WGS of the 278 asthmatic individuals, of which 65,170 (61.51%) variants in 14,222 genes (86.93%) were on the array (Table 5). Surprisingly, the proportion of functional variants on the array was similar in the three ethnic

| rsID (Chr:Position) | Gene | Residue Change | Alleles | GEMMA meta-analysis P (COMB) | Conditional GEMMA meta-analysis P for rs12450091 (c.A365G, p.E122G) | LAT (no covariate P = 6.89 × 10^{-5}) |
|---------------------|------|----------------|---------|----------------------------|-------------------------------------------------|-----------------------------------|
| rs907092 (17:37922259) | IKZF3 | c.C1314T, p.S438S | A/G | 2.08 × 10^{-10} | 0.102/0.97 | 0.0190 | 0.0215 |
| rs2305480 (17:38062196) | GSDMB | c.C892T, p.P298S | T/C | 1.09 × 10^{-8} | 0.104/0.99 | 0.0139 | 0.0160 |
| rs7216389 (17:38069949) | GSDMB | Intronic C/T | 1.23 × 10^{-7} | 0.13/1 | 0.0131 | 0.0149 |
The table below shows genes significantly associated with asthma.

Table 4 | Genes significantly associated with asthma.

| Sample | Gene | Chr | No. variants | MetaSKAT-O P |
|--------|------|-----|--------------|-------------|
| a. Functional variants: missense, nonsense and splice site | GSDMB | 17 | 16 | 6.10 × 10^-10 |
| Combined | ZPPB2 | 17 | 7 | 1.34 × 10^-6 |
| Latino | GSDMB | 17 | 12 | 1.02 × 10^-6 |
| Latino | ZPPB2 | 17 | 7 | 1.73 × 10^-6 |
| African Ancestry | MTHFR | 1 | 11 | 1.72 × 10^-6 |
| b. Functional variants: PolyPhen-2 probably damaging variants | GSDMB | 17 | 7 | 4.09 × 10^-8 |
| Latino | GSDMB | 17 | 5 | 7.81 × 10^-8 |

**Discussion**

Our study is the largest to date, to assess the effects of rare variants on asthma risk in ethnically diverse individuals. Although we were limited by the variation present on the array and by relatively small samples within each ethnic/racial group, our results provide several insights into the genetic architecture of asthma. First, associations with rare or low-frequency variants are likely to be ethnic specific. The few significant associations that we detected were all limited to a single ethnic group, despite having the greatest power to detect associations in the combined sample. Two novel associations, one with a missense mutation in the GRASP gene in Latinos and one with the many rare missense mutations in the MTHFR gene in African Americans/African Caribbeans, were revealed. The MTHFR gene harbours 11 missense mutations in African Americans, only one of which is common (rs1801133), and showed a gene-level association with asthma. Although most mutations in this gene were not predicted to be ‘damaging’ by PolyPhen-2 scores, the sheer number of potentially functional variants and the strong association with asthma in the African ancestry sample makes this gene a good candidate for further studies. Common variants in MTHFR, encoding methylenetetrahydrofolate reductase, have been associated with many common conditions, including neural tube defects, vascular disease and pregnancy complications, among others. There have been conflicting reports as to whether defects, vascular disease and pregnancy complications, among others. There have been conflicting reports as to whether defects, vascular disease and pregnancy complications, among others. There have been conflicting reports as to whether defects, vascular disease and pregnancy complications, among others. There have been conflicting reports as to whether...
in GRASP that is associated with asthma in Latinos occurs at a frequency of 0.012 in this population, at higher frequency (0.034) in European Americans and very low-frequency (0.006) in the African ancestry samples. The fact that the association was confined to Latinos even though we had greater power to detect associations in the European Americans suggests that the effects of this variant on asthma risk might involve interactions with other variants or environmental exposures that are more common in Latinos, but this single-variant association requires replication in an independent Latino sample or functional validation of this variant’s role in asthma risk. GRASP encodes GRP1 (general receptor for phosphoinositides-1)-associated scaffold protein and has been shown to be highly expressed in the brain with lower levels of expression in the lung, heart, embryo, kidney and ovary. Although little is known about the function of this gene or its potential role in asthma, scaffold proteins have been implicated in asthma signalling pathways. Second, among the six low-frequency variants showing suggestive evidence for association with asthma (Table 2), two variants were associated with decreased risk and three were associated with increased risk. The discovery of potentially protective alleles in three genes, FAM134B (family with sequence similarity 134, member B1), NPC1L1 (Niemann-Pick disease, type C1, gene-like 1) and IKZF1 (IKAROS family zinc finger 1), is intriguing because elucidating the function of these variants could identify novel therapeutic targets for asthma.

Third, the GSDMB (gadermin B) locus on chromosome 17q12–21 is within an LD block that has been robustly associated with asthma. The GSDMB gene is encoded by its alternative splice forms, other Gasdermin genes are involved in apoptosis in epithelial cells, processes and cells that are relevant with asthma; yet, the causal gene and its function are still unknown because elucidating the function of these variants could identify novel therapeutic targets for asthma. It is additionally notable that in contrast to other potential causal variants, a common SNP within a splicing defined using our study samples and was completed independently for each project.

Because of the different array design used for the CHS samples, the union of 239,909 (96.79%) variants was genotyped in the 13 sample sets. An additional 7,961 variants were genotyped in all the other samples and were subjected to quality control and meta-analysed independent of CHS.

**SNP quality control.** Among the variants present on the Illumina HumanExome BeadChip array interrogates 247,870 variants discovered from the exome and genome sequences of ~12,000 individuals of European American, African American, Latino and Asian American samples. Variants included on the array are predicted to have a functional effect on the protein and are enriched for missense mutations, nonsense mutations and variants within predicted splice sites, as previously described.

**Exome array.** The Illumina Infinium HumanExome BeadChip genotyping array interrogates 247,870 variants discovered from the exome and genome sequences of ~12,000 individuals of European American, African American, Latino and Asian American samples. Variants included on the array are predicted to have a functional effect on the protein and are enriched for missense mutations, nonsense mutations and variants within predicted splice sites, as previously described.

**Methods**

**Study subjects.** The subjects in the study were from 13 individual studies from eight investigators that comprise the EVE Consortium on the Genetics of Asthma. This sample has large overlap with our previous asthma meta-analyses but includes additional studies: GALA II, COAST and ALHS. Descriptions of each of the studies can be found in prior published work or in Supplementary Note 1. The results reported here are based on analyses of 4,794 asthma cases, 4,707 asthma controls and 590 case–parent trios (Table 1). All participants, or the parents of minors, provided written informed consent. The relevant local institutional research ethics committee approved each contributing study. Albert Einstein College of Medicine of Yeshiva University, Baylor College of Medicine, Brigham and Women’s Hospital, Children’s Memorial Hospital, Columbia University, Henry Ford Health System, Johns Hopkins University, Kaiser Permanente, NIHES (USA), National Institute of Public Health, Mexico (NIPH), Northwestern University, San Juan VA Medical Center, University of California San Francisco, University of Chicago, University of Southern California, University of Wisconsin Madison, WIRB for Torre Medica Auxilio Mutuo.

**Genotyping.** A total of 7,879 samples from 7 studies (CAG, CAMP, COAST, GALA II, GRAAD, MCCAS and SAPPHIRE), including 107 blind duplicates, were genotyped at the WestGen Genomics Center at the University of Washington. An additional 2,440 CHS samples were genotyped at the Molecular Genomics Core at the University of Southern California, using a different version of the array, and 1,585 ALHS samples were genotyped at the Center for Inherited Disease Research at Johns Hopkins University. Genotype calling was performed using Genotyping Module version 1.9, GenTrain version 1.0 in GenomeStudio version 2011.1 for all except for CHS that used GenTrain version 2.0. Genotype clustering was defined using our study samples and was completed independently for each project.

Because of the different array design used for the CHS samples, the union of 239,909 (96.79%) variants was genotyped in the 13 sample sets. An additional 7,961 variants were genotyped in all the other samples and were subjected to quality control and meta-analysed independent of CHS.

**SNP quality control.** Among the variants present on the Illumina HumanExome BeadChip array, 244,794 (98.76%) passed quality control (Supplementary Table 5). We excluded from analyses variants with genotype call rates <95%, Hardy–Weinberg equilibrium P-values <10⁻⁴ in any of the studied populations (African American, African Caribbean, European American and Mexican/Mexican American or Puerto Rican) and 333 caution sites reported as problematic by the exome array design group (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Among the variants passing quality control, 197,339 (79.61%) were variable in the individuals under study. In the combined sample, 23,206 (11.76%) variants were common (MAF ≥5%) to all ethnic groups, 36,413 (18.45%) were low-frequency (MAF 1–5%) and 137,720 (69.79%) were rare (MAF <1%) in all three groups. Private variants for each group and counts of variants in Supplementary Table 1 and Supplementary Table 2. Additional probe quality and genotype-calling evaluations were performed and are described in Supplementary Note 2.

**Sample quality control.** Six hundred and seventy-nine subjects were excluded from analysis (Supplementary Table 6), resulting in a final sample size of 11,225 individuals. Sample exclusions included gender discordances, sample duplications, Mendelian errors in trios, as implemented in PLINK, missing case–control status and incomplete trios. In addition, we completed principal component analysis using the R function pcapp to identify ancestry outliers (Supplementary Figs 7–9). Principal component analysis was carried out by using the ancestry informative markers present on the exome array and including HapMap6 CEU, YRI, CHB, JPT and MEX subjects. Fifty-six samples were excluded because one or both of the first two principle components were more than six s.d. from the expected ethnic group. These inferred principle components were applied as covariates in the gene-based tests.
Single variant analyses. We subjected functional low-frequency variants to single variant association analysis for each of the three ethnic groups (European American, African American and Latino) and for the combined sample. Low-frequency variants (1–5%) were included in each of the ethnic-specific meta-analyses (8,249 in the European Americans, 17,861 in the African ancestry and 9,519 in the Latinos) and the variants that were neither common (≥5%) in all 3 ethnic groups nor rare (<1%) in all three ethnic groups (32,681 variants) were meta-analysed in the combined sample (Supplementary Table 2). To control for population stratification and relatedness, single SNP association tests for case–control studies were completed using the linear mixed model GEMMA13. For the samples comprising case–parent trios, we used the FBAT16. Z-score statistics obtained from GEMMA and FBAT were combined in a meta-analysis as a weighted sum to reflect the power of each independent study. The weight (w) used for each SNP is a function of allele frequency (p), proportion of cases (v) and sample size (N) \( w = \sqrt{v(1-v)(1-p)2N} \), which was calculated using standard normal approximations and Bonferroni significance thresholds were applied (z = 0.05) to correct for the number of variants tested within each group and in the combined sample. ORs were obtained for samples analysed with GEMMA as a linear combination of log ORs with weights proportional to the s.e. of the log ORs using the R package meta.

Power analyses for single-variant analyses. For each ethnic group and for the combined sample, we performed the following simulations. We randomly selected 1,000 variants with MAF between 0.5% and 5%, and simulated genotypes 1,000 times for each of 5 ORs (OR = 1.5, 2, 2.5, 3 and 4). We used the same study structure (sample sizes and case–control or parent offspring designs) and observed MAF distributions that were present in our data. Z-scores were then meta-analysed with the weighted approach described in the single-variant meta-analysis methods and power was calculated with a predefined \( z = 0.05 \). We plotted power estimates versus MAF and implemented a local linear regression fit with the R function loess (Supplementary Fig. 2).

Genes-based analyses. Functional variants sets were defined two ways: (1) missense, nonsense or variants within splice sites or (2) coding variants with Phen-2 (ref.29) prediction scores (s) used for each SNP as implemented in MetaSKAT-O28. Analyses were completed with two sets of functional variants: (1) low-frequency variants in the European Americans, (2,316 genes) in the combined sample, we performed the following simulations. We randomly selected 1,000 variants with MAF between 0.5% and 5%, and simulated genotypes 1,000 times for each of 5 ORs (OR = 1.5, 2, 2.5, 3 and 4) and in the combined sample. ORs were obtained for samples analysed with GEMMA as a linear combination of log ORs with weights proportional to the s.e. of the log ORs using the R package meta.

Assessment of functional variants on the Illumina exome array. WGS for the 278 asthmatics was carried out at Complete Genomics (CGD) and resulted in an average read depth of 95.4% at 20 × or greater across the exome target and 91.4% across the genome. We excluded singleton deleterious from this analysis that were discovered in the 278 WGS but not present in dbSNP137 or in the Exome Sequencing Project variant server. We selected functional variants (missense, nonsense and splice site mutations) that were common in low-frequency or rare in the ethnic groups in the WGS and identified the overlap with the variants present on the exome array.

References

1. Thomsen, S. F., van der Sluis, S., Kyvik, K. O., Skytte, A. & Backer, V. Estimates of asthma heritability in a large twin sample. Clin. Exp. Allergy 40, 1054–1061 (2010).
2. Duffy, D. L., Martin, N. G., Battistutta, D., Hopper, J. L. & Mathews, J. D. Genetics of asthma and hay fever in Australian twins. Am. Rev. Respir. Dis. 142, 1351–1358 (1990).
3. Palmer, L. J., Rubin, P. R., James, A. L., Musk, A. W. & Cookson, W. O. Familial aggregation and heritability of asthma-associated quantitative traits in a population-based sample of nuclear families. Eur. J. Hum. Genet. 8, 853–860 (2000).
4. Torgerson, D. G. et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat. Genet. 43, 873–891 (2011).
5. Moffatt, M. F. et al. A large-scale, consortium-based genomewide association study of asthma. N. Engl. J. Med. 363, 1211–1221 (2010).
6. Bennelykke, K. et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat. Genet. 46, 51–55 (2014).
7. Hirota, T. et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. Nat. Genet. 43, 893–896 (2011).
8. Manolio, T. A. et al. Finding the missing heritability of complex diseases. Nature 461, 747–753 (2009).
9. Ober, C. & Vercelli, D. Gene–environment interactions in human disease: nuance or opportunity? Trends Genet. 27, 107–115 (2011).
10. Much, R. A. et al. Further replication studies of the EVE Consortium meta-analysis identifies 2 asthma risk loci in European Americans. J. Allergy Clin. Immunol. 130, 1294–1301 (2012).
11. Alavanja, M. C. et al. The Agricultural Health Study. Environ. Health Perspect. 104, 362–369 (1996).
12. Halapi, E. et al. A sequence variant on 17q21 is associated with age at onset and severity of asthma. Eur. J. Hum. Genet. 18, 902–908 (2010).
13. Lluís, A. et al. Asthma-associated polymorphisms in 17q21 influence cord blood ORMD3 and GSDMA gene expression and IL-17 secretion. J. Allergy Clin. Immunol. 127, 1587–94.e6 (2011).
14. Moffatt, M. F. et al. Genetic variants regulating ORMEL3 expression contribute to the risk of childhood asthma. Nature 460, 470–473 (2007).
15. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–824 (2012).
16. Horvath, S., Xu, X. & Laird, N. M. The family based association test method: strategies for studying general genotype–phenotype associations. Eur. J. Hum. Genet. 9, 301–306 (2001).
17. DeWan, A. T. et al. Whole-exome sequencing of a pedigree segregating asthma. BMC Med. Genet. 13, 95 (2012).
18. Himes, B. E. et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. Am. J. Hum. Genet. 84, 581–593 (2009).
19. 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. 134, 295–305 (2014).
20. Granell, R. et al. Examination of the relationship between variation at 17q21 and childhood wheeze phenotypes. J. Allergy Clin. Immunol. 131, 685–694 (2013).
21. Całyński, M. et al. Rhinovirus wheezing illness and genetic risk of childhood- onset asthma. N. Engl. J. Med. 368, 1398–1407 (2013).
22. Hao, K. et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet. 8, e1003029 (2012).
23. Murphy, A. et al. Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. Hum. Mol. Genet. 19, 4745–4757 (2010).
24. Sharma, S. et al. A genome-wide survey of CD4+ lymphocyte regulatory genetic variants identifies novel asthma genes. J. Allergy Clin. Immunol. 134, 1153–1162 (2014).
25. Zuko, O. et al. Searching for missing heritability: designing rare variant association studies. Proc. Natl Acad. Sci. USA 111, E455–E464 (2014).
26. Torgerson, D. G. et al. Resequencing candidate genes implicates rare variants in asthma susceptibility. Am. J. Hum. Genet. 90, 273–281 (2012).
27. Cirulli, E. T. & Goldstein, D. B. Uncovering the role of rare variants in common disease through whole-genome sequencing. Nat. Rev. Genet. 11, 415–425 (2010).
28. Lee, S., Teslovich, T. M., Boehnke, M. & Lin, X. General framework for meta-analysis of rare variants in sequencing association studies. Am. J. Hum. Genet. 91, 42–53 (2013).
29. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. Methods 7, 248–249 (2010).
30. Huyghe, J. R. et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat. Genet. 45, 197–201 (2013).
31. Irisle, P. N. et al. Impact of the MTHFR C677T polymorphism on risk of neural tube defects: case-control study. BMJ 328, 1533–1536 (2004).
32. Newton-Cheh, C. et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat. Genet. 41, 666–676 (2009).
33. Wu, X. et al. Association between the MTHFR C677T polymorphism and recurrent pregnancy loss: a meta-analysis. Genet. Test Mol. Biomarkers 16, 806–811 (2012).
34. Husemoen, L. L. et al. The association between atopy and factors influencing folate metabolism: is low folate status causally related to the development of atopy? Int. J. Epidemiol. 35, 954–961 (2006).
35. Thuesen, B. H., Husemøen, L. L. N., Fenger, M. & Linneberg, A. Lack of association between the MTHFR (C677T) polymorphism and atopic disease. Clin. Respir. J. 3, 102–108 (2009).
36. Granell, R. et al. The association between mother and child MTHFR C677T polymorphisms, fetal folate intake and childhood atopy in a population-based, longitudinal birth cohort. Clin. Exp. Allergy 38, 320–328 (2008).
37. Håberg, S. E. et al. Maternal folate levels in pregnancy and asthma in children at age 3 years. J. Allergy Clin. Immunol. 127, 262–4–264.e1 (2011).
38. Håberg, S. E., London, S. J., Stigum, H., Naftad, P. & Nystad, W. Folic acid supplements in pregnancy and early childhood respiratory health. Arch. Dis. Child. 94, 180–184 (2009).
39. Nevrivy, D. J. Interaction of GRASP, a protein encoded by a novel retinoic acid-induced gene, with members of the cytohesin family of guanine nucleotide exchange factors. J. Biol. Chem. 275, 16827–16836 (2000).
40. Zhang, Y. et al. Roles of scaffolding protein Gab1 in regulation of myeloid dendritic cell-mediated allergic asthma (HYP7P.281). J. Immunol. 192, 119.4–119.4 (2014).
41. Walker, J. K. L. et al. Beta-arrestin-2 regulates the development of allergic asthma. J. Clin. Invest. 112, 566–574 (2003).
42. Verlaan, D. J. et al. Targeted screening of cis-regulatory variation in human haplotypes. Genome Res. 19, 118–127 (2009).
43. Morrison, F. S. et al. The splice site variant rs11078928 may be associated with a genotype-dependent alteration in expression of GSDMB transcripts. BMC Genomics 14, 627 (2013).
44. Saeki, N. et al. Distinctive expression and function of four GSDM family genes (GSDMA-D) in normal and malignant upper gastrointestinal epithelium. Genes Chromosomes Cancer 48, 261–271 (2009).
45. 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).
46. Gibbs, R. A. et al. The International HapMap Project. Nature 426, 789–796 (2003).

Acknowledgements
This work was supported by NIH grants RC2 HL10165, P30 ES007048, P01 ES09581, P01 ES01162, R01 HL118267 and R01 AI079139, the Intramural Research Program of the NIH (NIEHS Z01 CP010119, Z01 ES049030 and ES102385), the American Asthma Foundation, the Hastings Foundation, the Fund for Henry Ford Hospital and through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory (S10 RR029030-01). Genotyping services were provided by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under US Federal Government contract number HHSN26820110037C from the National Heart, Lung and Blood Institute, and the Molecular Genomics Core at the University of Southern California (P30ES007048 and P01ES009581) and the Center for Inherited Disease Research at Johns Hopkins University. M.P.-Y. was funded by a Postdoctoral Fellowship from Fundación Ramón Areces. C.I. was supported by National Institutes of Health Grant T32 GM007197 and by the Ruth L. Kirschstein National Research Service Award (PA-11-112). KCB was supported in part by the Mary Beryl Patch Turnbull Scholar Program.

Author contributions
B.E.D.-R.-N., D.J.J., J.E.G., R.F.L., A.W., J.A.H., K.C.B., E.G.B., W.J.G., F.D.M., B.A.R., S.T.W., L.K.W., S.J.L., F.D.G., D.L.N. and C.O. provided samples and clinical data. C.I., R.A. Myers, M.P.-Y., O.E.L. and R.M. completed data processing. R.A. Mathias, C.E., P.E.G., A.M.L., B.E.D.-R.-N., D.J.J., N.R., C.K.E., J.J.Y., S.H., M.T.S. and I.R. generated input data. C.I., D.L.N. and C.O. conceived/designed the study and wrote the paper. A.M.L., B.E.D.-R.-N., D.J.J., J.E.G., R.F.L., A.W., J.A.H., K.C.B., E.G.B., W.J.G., F.D.M., B.A.R., S.T.W., L.K.W., S.J.L., F.D.G., D.L.N. and C.O. generated input data. R.A. Mathias, C.E., P.E.G., A.M.L., B.E.D.-R.-N., D.J.J., N.R., C.K.E., J.J.Y., S.H., M.T.S. and I.R. performed all analyses. All authors discussed results and commented on manuscript.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications
Competing financial interests: The authors declare no competing financial interests.
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/
How to cite this article: Igartua, C. et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. Nat. Commun. 6:5965 doi: 10.1038/ncomms6965 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/