Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks

We examined common variation in asthma risk by conducting a meta-analysis of worldwide asthma genome-wide association studies (23,948 asthma cases, 118,538 controls) of individuals from ethnically diverse populations. We identified five new asthma loci, found two new associations at two known asthma loci, established asthma associations at two loci previously implicated in the comorbidity of asthma plus hay fever, and confirmed nine known loci. Investigation of pleiotropy showed large overlaps in genetic variants with autoimmune and inflammatory diseases. The enrichment in enhancer marks at asthma risk loci, especially in immune cells, suggested a major role of these loci in the regulation of immunologically related mechanisms.

Asthma is a complex disease affecting hundreds of millions of people worldwide. The prevalence of asthma varies across populations and ancestral origins; for example, in the US, the prevalence ranges from 3.9% in Mexican Americans to 12.5% in African Americans. The contribution of genetic factors to asthma risk has been demonstrated in family studies, in which heritability estimates range from 25% to 80% (ref. 2). The high variability in prevalence and heritability estimates reflects the roles of environmental exposure in the disease risk and phenotypic heterogeneity that are hallmarks of asthma. These features may explain why genome-wide association studies (GWAS) have identified a smaller number of asthma loci than have been found in similarly sized studies of other multifactorial diseases. Indeed, at the time of analysis, only 21 loci have been associated with asthma per se in 20 studies, and these loci explain only part of the genetic risk. Although an exome-array study has shown no evidence of low-frequency or rare variants with large effects on asthma risk, the role of rare noncoding variants in asthma remains unknown. Future studies based on whole-genome sequencing may clarify the respective influence of common and rare variants on asthma risk. To generate larger sample sizes for GWAS meta-analysis of asthma and thereby enable the discovery of new common risk loci, we established the Trans-National Asthma Genetic Consortium (TAGC), comprising worldwide groups of investigators, which has analyzed genome-wide data available in 142,000 individuals of diverse ancestries. We constructed a comprehensive catalog of asthma risk variants that are robust across populations and environmental-exposure conditions. By combining TAGC meta-analysis results with data from existing databases, we assessed the genetic architecture of asthma risk alleles with respect to functional effects and shared effects with other diseases.

Results

Meta-analysis of asthma GWAS. We combined data from asthma GWAS with high-density genotyped and imputed SNP data (2.83 million SNPs) in the following populations: European ancestry (19,954 asthma cases, 107,715 controls), African ancestry (2,149 asthma cases, 6,055 controls), Japanese ancestry (1,239 asthma cases, 3,976 controls), and Latino ancestry (606 asthma cases, 792 controls) (Supplementary Table 1). After extensive quality control of summary data provided by each participating group (Methods, Supplementary Note and Supplementary Table 2), we conducted ancestry-specific meta-analyses, then performed a multiancestry meta-analysis of all populations (23,948 asthma cases, 118,538 controls) to identify additional loci with panancestry effects. Because childhood-onset asthma may be distinct from later-onset asthma and may represent a more homogeneous subgroup, we also performed analyses on the pediatric subgroup (asthma onset ≤ 16 years; 8,976 asthma cases, 18,399 controls). Meta-analyses of SNP effect sizes obtained from each asthma GWAS were performed with fixed-effects (significance of the combined SNP effect size summarized in $P_{\text{fix}}$ and random-effects ($P_{\text{rand}}$) models (Methods), and a conventional $P_{\text{rand}}$ (or $P_{\text{fix}}$) threshold of $5 \times 10^{-8}$ was used to define genome-wide significance. The results were consistent between methods for detecting loci with at least one SNP significantly associated with asthma. We therefore present the results from the random-effects analysis for the European-ancestry and multiancestry meta-analyses, which included the largest number of studies and allowed for an accurate estimate of the between-study variance, and the results from the fixed-effects analysis for the African-ancestry, Japanese-ancestry, and Latino-ancestry meta-analyses. We observed little evidence of inflation in the test statistics in either the ancestry-specific (European ancestry, $\lambda = 1.031$; African ancestry, $\lambda = 1.014$; Japanese ancestry, $\lambda = 1.021$; Latino ancestry, $\lambda = 1.044$) or multiancestry ($\lambda = 1.046$) meta-analyses (Supplementary Fig. 1).

We identified 673 genome-wide-significant SNPs ($P_{\text{rand}} \leq 5 \times 10^{-8}$) at 16 loci in European-ancestry populations (Fig. 1a, Table 1 and Supplementary Tables 3 and 4; locus definition in Methods). No genome-wide-significant risk loci were detected in African-ancestry, Japanese-ancestry, or Latino-ancestry populations (Supplementary Fig. 2 and Supplementary Tables 5–7), possibly because of a lack of power. In the combined multiancestry meta-analysis, 205 additional SNPs were significant ($P_{\text{rand}} \leq 5 \times 10^{-8}$), including 12 SNPs at two loci not detected in the European-ancestry analysis (Fig. 1b, Table 1 and Supplementary Tables 3 and 8). Altogether, 878 SNPs at 18 loci reached genome-wide significance, of which 69% were significant in both European-ancestry and multiancestry meta-analyses, 23% were significant in only the multiancestry meta-analysis, and 8% were significant in only the European-ancestry meta-analysis (Supplementary Tables 4 and 8; regional plots of the 18 loci in Supplementary Fig. 3). All 18 loci remained genome-wide significant after further genomic control correction of the test statistics, thus confirming the robustness of these results (Supplementary Table 9).
The 18 chromosomal regions included five new loci associated with asthma at 5q31.3, 6p22.1, 6q15, 12q13.3, and 17q21.33; two new associations at 6p21.33 and 10p14 that were independent of previously reported signals at these loci in ancestry-specific populations (Latino and Japanese ancestries, respectively); two associations at 8q21.13 and 16p13.13 that were previously reported for asthma plus hay fever but not for asthma alone in a study of European-ancestry populations; and nine previously identified asthma loci.

None of the lead SNPs at the 18 loci showed evidence of heterogeneity in effect sizes across studies except for the lead variant at 9p24.1 ($P_{\text{het}} = 0.008$ across European-ancestry studies and $P_{\text{het}} = 0.02$ across multiancestry studies; Table 1 and Supplementary Fig. 4). There was also significant evidence of heterogeneity in the ancestry-specific effect sizes ($P_{\text{ethnic}} = 0.003$) for the 6p22.1 lead SNP rs1233578, which consequently did not reach significance in the multiancestry analysis (Table 1 and Supplementary Table 3). The meta-analysis of the pediatric subgroup showed evidence of association ($P_{\text{random}} \leq 5 \times 10^{-8}$) at five of the 18 loci (2q12, 5q31, 6p21.33 9p24.1, and 17q12-21) (Supplementary Figs. 5 and 6 and Supplementary Table 10). No loci specific to the pediatric subgroup were identified.

The results provided genome-wide-significant confirmation of nine previously reported loci in both the European-ancestry and multiancestry meta-analyses (Table 1 and Supplementary Figs. 3b and 4). Our results allowed for detailed analysis of the broad 17q12-21 locus. Notably, the lead SNP (rs2952156) at this locus was within ERBB2 ($P_{\text{random}} = 2.2 \times 10^{-30}$ in multiancestry meta-analysis), at least 180 kb from the previously recognized asthma-associated signals at the GSDMB/ORMDL3 haplotype block (Supplementary Fig. 7). This result was attributable to effect-size heterogeneity across studies ($0.001 \leq P_{\text{het}} \leq 0.05$) that extended over a 200-kb region including ORMDL3 and GSDMB (Supplementary Table 11). This heterogeneity was partly due to the age of asthma onset, as previously reported. Indeed, in the pediatric group, the 17q12-21 SNPs did not show heterogeneity ($P_{\text{het}} \geq 0.09$), and the lead SNP rs8069176 was 3.6 kb proximal to GSDMB ($P_{\text{random}} = P_{\text{fixed}} = 4.4 \times 10^{-28}$), in agreement with results from previous studies. The SNP effect sizes in the pediatric and nonpediatric studies showed a significant
Table 1 | Genetic loci associated with asthma in European-ancestry and multiancestry meta-analyses

| Locus in ENCODE | Significant chromosomal positions | 40′s 5′s | Neighboring gene(s) | Allele | EAF | OR (95% CI) | P (random) | P (het) | OR (95% CI) | P (random) | P (het) | OR (95% CI) |
|-----------------|-----------------------------------|----------|---------------------|--------|-----|-------------|------------|--------|-------------|------------|--------|-------------|
| 5q31.3          | 0/11                              | rs7705042 | 141492419           | NDFIP1, GNDPA1, SPRY4 | C/A   | 0.63 | 1.08 (1.05–1.11) | 1.6 × 10⁻⁶ | 0.07 | 1.09 (1.06–1.12) | 7.9 × 10⁻⁵ | 0.11 | 0.55 |
| 6p22.1          | 8/5                               | rs1233578 | 28712247            | GEPX, TRIM27          | A/G   | 0.13 | 1.11 (1.07–1.15) | 5.3 × 10⁻⁹ | 0.82 | 1.09 (1.05–1.12) | 5.9 × 10⁻⁷ | 0.56 | 0.003 |
| 6q15            | 26/26                             | rs2325291 | 90986686             | BACH2, GJA10, MAP3K7  | G/A   | 0.33 | 0.91 (0.89–0.93) | 8.6 × 10⁻¹⁴ | 0.78 | 0.91 (0.89–0.94) | 2.2 × 10⁻¹⁰ | 0.80 | 0.39 |
| 12q13.3         | 0/1                              | rs167767  | 57503775             | STAT6, NAB2, LRP1     | C/T   | 0.40 | 1.08 (1.05–1.11) | 1.6 × 10⁻⁷ | 0.19 | 1.08 (1.05–1.11) | 3.9 × 10⁻³ | 0.31 | 0.85 |
| 17q21.33        | 4/3                               | rs1763747 | 47461433             | ZNF6S2, PHB           | G/A   | 0.39 | 1.08 (1.05–1.11) | 3.3 × 10⁻⁹ | 0.56 | 1.08 (1.05–1.11) | 6.6 × 10⁻³ | 0.35 | 0.12 |

**New asthma susceptibility loci**

| 6q13            | 66/53                             | rs2855812 | 31472720             | MICB, HCPS, MCDC1     | G/T   | 0.23 | 1.10 (1.06–1.13) | 1.7 × 10⁻⁸ | 0.23 | 1.10 (1.07–1.13) | 8.9 × 10⁻¹² | 0.39 | 0.58 |
| 10p14           | 3/6                               | rs2589561 | 90466645             | GATA3, CELF2          | A/G   | 0.82 | 0.90 (0.87–0.94) | 1.4 × 10⁻⁸ | 0.78 | 0.91 (0.88–0.94) | 3.5 × 10⁻⁹ | 0.82 | 0.25 |

**New signals at loci previously associated with asthma in ancestry-specific populations**

| 6q13.3          | 12/13                             | rs17806299 | 11999980             | CLEC16A, DEXI, SOCS1  | G/A   | 0.20 | 0.90 (0.88–0.93) | 2.1 × 10⁻¹⁰ | 0.51 | 0.91 (0.88–0.94) | 2.7 × 10⁻¹² | 0.49 | 0.58 |

**Known asthma loci**

| 2q12            | 133/144                           | rs1420101 | 102957716            | IL1RL1, IL1RL2, IL18R1 | C/T   | 0.37 | 1.12 (1.10–1.15) | 9.1 × 10⁻²⁰ | 0.63 | 1.12 (1.09–1.15) | 3.9 × 10⁻²⁰ | 0.61 | 0.64 |
| 5q22.1          | 35/32                             | rs14055023 | 110404999            | STC2A44, TSLP         | A/C   | 0.34 | 1.15 (1.12–1.18) | 2.0 × 10⁻⁵ | 0.53 | 1.15 (1.12–1.18) | 9.4 × 10⁻⁵ | 0.57 | 0.27 |
| 5q31            | 33/62                             | rs20541   | 13199596             | IL13, RADS5, IL4      | A/G   | 0.79 | 0.89 (0.86–0.91) | 1.4 × 10⁻¹⁴ | 0.73 | 0.89 (0.87–0.92) | 5.0 × 10⁻¹⁶ | 0.77 | 0.62 |
| 6p21.32         | 101/124                           | rs972346  | 32604372             | HLA-DRB1, HLA-DQA1    | G/A   | 0.56 | 1.16 (1.13–1.19) | 4.8 × 10⁻²⁸ | 0.46 | 1.16 (1.12–1.19) | 5.7 × 10⁻²⁴ | 0.14 | 0.43 |
| 9p24.1           | 65/71                            | rs992969  | 6209697              | RANBP6, IL3          | A/G   | 0.75 | 0.85 (0.82–0.88) | 1.1 × 10⁻³⁷ | 0.008 | 0.86 (0.83–0.88) | 7.2 × 10⁻³₀ | 0.02 | 0.57 |
| 11q13.5          | 4/5                               | rs7927894 | 76301316             | EMSY, LRRRC2         | C/T   | 0.37 | 1.10 (1.07–1.13) | 3.5 × 10⁻¹¹ | 0.38 | 1.10 (1.08–1.13) | 2.2 × 10⁻¹⁴ | 0.56 | 0.47 |
| 15q22.2          | 9/14                             | rs11071558 | 61069421             | RORA, RARG2, VPS13C  | A/G   | 0.14 | 0.89 (0.85–0.92) | 1.9 × 10⁻¹⁰ | 0.44 | 0.89 (0.86–0.92) | 1.3 × 10⁻³ | 0.19 | 0.06 |
| 15q22.33         | 13/13                            | rs2033784 | 67449660             | SMAD3, SMAD6, AAGAB   | A/G   | 0.30 | 1.11 (1.08–1.14) | 2.5 × 10⁻¹⁳ | 0.75 | 1.10 (1.08–1.13) | 7.4 × 10⁻¹⁵ | 0.76 | 0.48 |
| 17q12-21         | 160/198                           | rs2952156 | 37876835             | ERB82, PGAP3, MGENI   | A/G   | 0.70 | 0.86 (0.84–0.88) | 7.6 × 10⁻²⁴ | 0.55 | 0.87 (0.84–0.89) | 2.2 × 10⁻²⁵ | 0.52 | 0.35 |

*SNP P-values for association with asthma are based on random-effects meta-analysis in Stata. A total of 878 SNPs, belonging to 18 loci, reached genome-wide significance (P < 5 × 10⁻⁸). Each locus in this table is represented by the SNP with the strongest evidence of association in the European-ancestry (127669 subjects) or multiancestry meta-analysis (142,486 subjects from European-ancestry, African-ancestry, Japanese-ancestry and Latino-ancestry populations). Cochran’s Q test was used to test for heterogeneity in SNP effect sizes across studies (P < 0.05) and to test for differences among the four ancestry-specific summary effects (P < 0.05). EAF, effect allele frequency; OR, odds (log-additive) ratio; 95% CI, 95% confidence interval. *Cytogenetic band. **Number of genome-wide significant SNPs (P < 5 × 10⁻⁸) at each locus in European-ancestry meta-analysis/multiancestry meta-analysis. **SNP position, build37. The gene in which the SNP is located is first indicated, followed by the previous gene and next gene (for intergenic SNPs, only the previous and next genes are indicated); note that TNF and MEFN were previously designated CTORF170 and CTORF37, respectively. Reference/ effect allele.*
Table 2 | Main characteristics of the nine loci showing new associations with asthma

| Locusa | Location of lead SNPb | Cis eQTLs in blood (B) and lung tissue (L) | Association with allergy-related and lung-function phenotypes | Association with autoimmune diseases and other immunologically related traits |
|--------|-----------------------|------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
| New asthma susceptibility loci |
| 5q31.3 | NDFIP1 (intron) | B: NDFIP1 (2.7 × 10^{-12}), L: ZSCAN43 (6.5 × 10^{-10}) | Lung function | IBD |
| 6p22.1 | Intergenic | B: ZSCAN2 (3.0 × 10^{-9}), L: ZSCAN43 (6.5 × 10^{-10}) | IgE (total, specific), lung function | MS, T1D, CD, IBD, V, IGG |
| 6q15 | BACH2 (intron) | B: BACH2 (3.0 × 10^{-10}) | Self-reported allergy, atopic dermatitis, lung function | SLE, UC, RA, IBD, BS, GD, SS, AS, Pso, UC, WBC, MoC, DS, HIV-1, SJ5, HB, IVM, CD4/CD8 ratio, HIV-TC |
| 12q13.3 | STA76 (intron) | B: STA76 (9.8 × 10^{-10}), L: STA76 (3.7 × 10^{-10}) | IgE (total, specific), lung function | Pso, ISP_IFN |
| 17q21.33 | Intergenic | B: GNGT2 (2.1 × 10^{-8}) | Atopic dermatitis | ISP_IL2 |

New asthma signals at loci previously associated with asthma in ancestry-specific populations

| Locusa | Location of lead SNPb | Cis eQTLs in blood (B) and lung tissue (L) | Association with allergy-related and lung-function phenotypes | Association with autoimmune diseases and other immunologically related traits |
|--------|-----------------------|------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
| 6p21.33 | MICB (intron) | B: TNF (4.8 × 10^{-10}), L: ST1 (1.0 × 10^{-10}), HLA-C (3.2 × 10^{-10}), LTA (1.0 × 10^{-10}) | IgE (total, specific), self-reported allergy, atopic dermatitis, lung function | SLE, UC, RA, IBD, BS, GD, SS, AS, Pso, UC, WBC, MoC, DS, HIV-1, SJ5, HB, IVM, CD4/CD8 ratio, HIV-TC |
| 10p14 | Intergenic | None | Self-reported allergy | RA, ISP_IL1B, ISPV |

Asthma signals previously reported for asthma plus hay fever

| Locusa | Location of lead SNPb | Cis eQTLs in blood (B) and lung tissue (L) | Association with allergy-related and lung-function phenotypes | Association with autoimmune diseases and other immunologically related traits |
|--------|-----------------------|------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
| 8q21.13 | Intergenic | None | Atopic dermatitis, asthma plus hay fever, self-reported allergy | RA |
| 16p13.13 | CLEC16A (intron) | B: DEK (2.2 × 10^{-10}) | Atopic dermatitis, asthma plus hay fever | TID, PBC, MS, RA, IBD, CD, LEP |

At each of the nine loci showing new associations with asthma, cis genes whose expression (eQTLs) is associated with the lead asthma-associated SNPs (shown in Table 2) or SNPs in LD (r^2 ≥ 0.5) with the lead SNPs were searched in six eQTL databases from whole blood (11), lymphoblastoid cell lines (10), monocytes (23), and lung tissue (12,14); only genes with the strongest associations (P < 5 × 10^{-8}, as shown in parentheses) are presented here (details in Supplementary Table 16). Overlap of these nine loci with associations with allergy-related and lung-function phenotypes as well as with autoimmune diseases and other immunologically related traits was annotated with the GWAS catalog (18); IBD, inflammatory bowel disease (Crohn’s disease), MS, multiple sclerosis, T1D, type 1 diabetes; CD, celiac disease; V, viremia, IGG, IgG glycosylation, Pso, psoriasis, ISP_IFN, immune response to smallpox (secreted IFN-α), ISP_IL2 immune response to smallpox (secreted IL2), SLE, systemic lupus erythematosus, UC, ulcerative colitis, RA, rheumatoid arthritis; BS, Behçet syndrome; GD, Grave’s disease; SS, systemic sclerosis; AS, ankylosing spondylitis; WBC, white blood cell count; MoC, monocyte count; DS, dengue shock; HIV-1, HIV-1 susceptibility; SJ5, Stevens-Johnson syndrome; HB, hepatitis B infection; HTV, hepatitis B vaccine response; IMN, idiopathic membranous nephropathy; CD4/CD8, CD4/CD8 lymphocyte ratio; HIV-TC, HIV-1 control; ISP_IL1B, immune response to smallpox (secreted IL1B); ISPV, immune response to smallpox vaccine (IL-6); PBC, primary biliary cirrhosis; LEP, leprosy.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.
To investigate whether the 18 asthma loci identified in this study contained multiple distinct signals, we performed approximate conditional regression analysis, based on summary statistics, for all loci (Methods), except for the 9p24.1 region, which showed heterogeneity in SNP effect size across studies over the entire locus. For the 17q12-21 locus, this analysis was restricted to the pediatric subgroup in which there was no heterogeneity. After conditioning on the lead SNP in each investigated region, four secondary signals (2q12, 5q22.1, 5q31, and 6p21.32) remained significant (P_{\text{random}} ≤ 5 × 10^{-8}) (Supplementary Table 15), thus yielding 22 distinct genome-wide-significant signals.

To provide biological insight into our findings, we conducted a comprehensive bioinformatic assessment of the asthma-association signals. To pinpoint the most likely candidate genes at the nine loci with new associations with asthma per se, we interrogated the results of six eQTL studies in tissues relevant to asthma: blood (including peripheral blood\cite{11,12}, lymphoblastoid cell lines\cite{14,13}, and monocytes\cite{15}), and whole lung tissue\cite{16,14}. We also searched for missense variants potentially tagged by the association signals, using the HaploReg v4.1 tool (URLs). To assess the degree of overlap of asthma associations with susceptibility loci for other phenotypes, we interrogated the GWAS catalog\cite{1} while varying the strength of association with asthma (thresholds from 5 × 10^{-4} to 5 × 10^{-9}). To obtain greater insight into how asthma-associated variants might functionally influence disease, we interrogated the Roadmap/Encyclopedia of DNA Elements (ENCODE) functional genomics data generated from a wide range of human cell types\cite{18}. Finally, the degree of connectivity among the asthma-associated loci was assessed through text mining\cite{19}. The results are described below.

Candidate genes at the nine loci showing new associations. A summary of the eQTL analysis for these nine loci is described in Table 3 and Supplementary Table 16; regional plots are shown in Supplementary Fig. 3a.

**Table 3 | Overlap of TAGC asthma-associated SNPs with GWAS-catalog association signals by disease group**

| Disease group                  | Number of GWAS-catalog association signals | Number of SNPs associated with asthma at P_{\text{random}} ≤ 10^{-8} in the multiancestry meta-analysis | P value for overlap |
|-------------------------------|-------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------|
| Cardiovascular                | 743                                       | 20                                                                                              | 7.8 × 10^{-42}    |
| Body size and morphology      | 346                                       | 2                                                                                               | 5.0 × 10^{-4}     |
| Immune/autoimmune             | 480                                       | 49                                                                                              | 3.0 × 10^{-129}   |
| Nervous system                | 242                                       | 4                                                                                               | 1.4 × 10^{-8}     |
| Blood                         | 594                                       | 10                                                                                              | 1.3 × 10^{-39}    |
| Neuropsychiatric              | 114                                       | 5                                                                                               | 1.5 × 10^{-32}    |
| Cancer                        | 417                                       | 7                                                                                               | 4.0 × 10^{-54}    |
| Endocrine system              | 276                                       | 2                                                                                               | 4.0 × 10^{-4}     |
| Digestive system              | 347                                       | 16                                                                                              | 1.4 × 10^{-47}    |
| Eyes                          | 177                                       | 2                                                                                               | 2.0 × 10^{-4}     |
| Respiratory system            | 85                                        | 2                                                                                               | 3.6 × 10^{-5}     |
| Infectious disease/infection  | 104                                       | 2                                                                                               | 5.3 × 10^{-5}     |
| Urinary system                | 144                                       | 1                                                                                               | 1.5 × 10^{-2}     |
| Alcohol, smoking, and illicit substances | 30                                | 0                                                                                               | 1                 |
| Musculoskeletal system        | 132                                       | 0                                                                                               | 1                 |

Overlap of TAGC asthma-associated SNPs with association signals of all diseases/traits in the GWAS catalog\cite{2} was investigated for all TAGC SNPs with P_{\text{random}} ≤ 10^{-8} in the multiancestry meta-analysis; diseases from the GWAS catalog were grouped according to the disease classification proposed by Wang et al.\cite{9} (the ‘digestive system’ group includes Crohn’s disease, a subtype of inflammatory bowel disease). The significance of overlap was estimated with the binomial-tail probability for observing the shown number of TAGC asthma SNPs among the number of SNPs reported in the GWAS catalog for a group of diseases (for example, the probability of observing ≥ 20 asthma SNPs with P_{\text{random}} ≤ 10^{-8} among the 743 cardiovascular SNPs is shown in the last column); a conservative Bonferroni-adjusted significance threshold for enrichment in shared associations is 0.05/15 = 0.003 (for the 15 disease groups investigated).

**Table 4 | Enrichment of asthma risk loci in promoter and enhancer marks and DNAse I-hypersensitive sites**

| Type of regulatory elements | Proportion of all cell types (blood cell types) showing enrichment with a given FDR |
|-----------------------------|--------------------------------------------------------------------------------------|
| FDR ≤ 10%                   | FDR ≤ 5%                                                                             |
| All promoter states         | 6% (26%)                                                                             |
| Active promoter states      | 13% (33%)                                                                            |
| All enhancer states         | 57% (100%)                                                                           |
| Active enhancer states      | 66% (100%)                                                                           |
| DNAse I-hypersensitive sites| 16% (50%)                                                                             |

The colocalization of SNPs at asthma risk loci with regulatory elements (promoters, enhancers, and DNAse I-hypersensitive sites) was assessed at 16 asthma loci identified in this study (Table 1); the 6q21.33 and 6q21.32 loci encompassing the HLA region were excluded because of the high amount of variability and LD in that region. Enhancer and promoter states were defined with the Chrom-HMM 15-state model applied to functional data of 127 Roadmap and ENCODE reference epigenomes in various cell types (including 27 leukocytes). DNAse I-hypersensitive sites were identified in 151 cell types (including 10 leukocytes). Empirical P values for enrichment were obtained through 10,000 Monte Carlo simulations of random sets of SNPs matching the original set of asthma-associated SNPs\cite{25}; Benjamini–Hochberg FDR was calculated to correct for multiple testing (details in Methods).
Articles

Asthma signals reported for asthma plus hay fever. In one study of individuals of European-ancestry, loci on chromosomes 8q21.13 and 16p13.13 have been associated with asthma plus hay fever but not with asthma alone. In our results, the 8q21.13 lead SNP rs12543811 ($P_{\text{random}} = 3.4 \times 10^{-8}$ and $1.1 \times 10^{-10}$ in the European-ancestry and multiancestry analyses) was located between $TPD52$ and $ZBTF1$ and was in strong LD ($r^2 = 0.79$) with the previously reported asthma/hay fever SNP rs7009110. These two SNPs represented the same signal, because the association with rs12543811 became nonsignificant after conditioning on rs7009110. Thus, the 8q21.13 locus is likely to be associated with allergic asthma. A functional analysis of the asthma/hay fever locus pinpointed $PAG1$ as a promising candidate. The chromosome 16p13.13 SNP rs17806299 is within an intron of $CLEC16A$ ($P_{\text{random}} = 2.1 \times 10^{-10}$ and $2.7 \times 10^{-10}$ in European-ancestry and multiancestry meta-analyses). Although it was in moderate LD ($r^2 = 0.66$) with the previously reported asthma/hay fever signal (rs62026376), the association with asthma rs17806299 was removed after conditioning on rs12935657 ($r^2 = 0.96$ with rs62026376), thus indicating that these SNPs represented the same signal and that 16p13.13 was probably also an allergic asthma locus. The SNP rs17806299 is strongly associated with the expression of a nearby gene, $DEXI$ in the blood. Similar observations of associations of $CLEC16A$ SNPs with autoimmune diseases and expression of $DEXI$ together with chromosome-conformation-capture experiments have implicated $DEXI$ as the most likely candidate gene associated with autoimmune diseases. The potential relevance of $DEXI$ in allergic diseases has also been previously discussed.

Notably, the lead SNPs at the nine new asthma-associated loci were located in noncoding regions and did not tag missense variants.

Overlap of loci associated with asthma and other phenotypes. We next explored whether the nine loci bearing new signals for asthma per se overlapped with GWAS loci reported for allergy-related phenotypes, lung-function phenotypes, or other immunologically related diseases, by using the GWAS catalog. Six of these nine asthma loci showed overlapping associations with allergy-related phenotypes, and eight showed overlapping associations with autoimmune diseases or infection-related phenotypes (Table 2). Moreover, three asthma loci overlapped with associations with lung-function phenotypes.

We expanded our search of overlap between the asthma-association signals with multiancestry $P_{\text{random}} < 10^{-4}$ in this study and GWAS signals with all phenotypes and diseases in the GWAS catalog. We examined 4,231 unique trait–loci combinations (Methods) and used the disease classification from Wang et al.7 to group traits. We summarized the overlap with GWAS-catalog signals as the proportion of catalog SNPs with asthma $P$ values $< 10^{-4}$ in our analysis. The results showed significant overlap with autoimmune disease (49 out of 480 catalog SNPs (10%) showed evidence for asthma association), in agreement with the hypothesized shared susceptibility; moderate overlap with diseases with an inflammatory component (cardiovascular diseases, cancers, and neuropsychiatric diseases); and little to no overlap with other diseases (Table 3). When investigating specific diseases and traits (Supplementary Table 17), we observed the most significant overlap with allergic phenotypes. There was little to no overlap with other phenotypes that appeared most frequent in the GWAS catalog (for example, no shared associations with type 2 diabetes).

When we broadened our analysis to a larger set of SNPs in the GWAS catalog to identify loci for diseases with potentially shared genetic architecture with asthma (i.e., SNPs associated with asthma at $P_{\text{random}} \leq 10^{-5}$ in our multiancestry meta-analysis), additional pleiotropic signals emerged (Supplementary Table 18). This larger set of associations suggested a broader picture of asthma risk, with a wide range of shared effects with traits ranging from lung cancer...
and multiple sclerosis (with rs3817963 in BTNL2) to coronary heart disease (with rs1333042 near CDKN2B). This analysis also generated an extended set of candidate asthma-associated genes. Indeed, there were 210 SNPs in the GWAS catalog that were associated with asthma in TAGC at a threshold of 10^-3, and the proportion of false positives among these was smaller than 1%.

**Enrichment of asthma risk loci in epigenetic marks.** Because nearly all lead SNPs at the 18 loci identified by this study, except for the IL13 missense variant (rs20541), were located in noncoding sequences, we investigated whether the asthma-associated variants for the missense variant (rs20541), were located in noncoding nearly all lead SNPs at the 18 loci identified by this study, except Because Enrichment of asthma risk loci in epigenetic marks. positives among these was smaller than 1%.

LD partners, and MAF). Empirical for colocalization with enhancer marks, there was only weak enrich- Hochberg false discovery rate (FDR) values were then computed to obtaining enhancers and promoters assayed in all 127 epigenomes and DNa-1–hypersensitive sites available in 51 cell types. To assess enrichment of the asthma risk variants for colocalization with these regulatory elements, we used the Uncovering Enrichment through Simulation (UES) pipeline. This approach generates random SNP sets that match the characteristics of the original asthma-associated SNPs (distance from the nearest transcription start site, number of LD partners, and MAF). Empirical P values for enrichment were calculated by comparing the observed frequency of colocalization of SNPs with a given type of regulatory element in the original asthma-associated SNP set to the co-localization-frequency distribution obtained from the 10,000 random SNP sets generated. Benjamini–Hochberg false discovery rate (FDR) values were then computed to correct for multiple testing (Methods).

Although the asthma-associated variants were strongly enriched for colocalization with enhancer marks, there was only weak enrichment in promoter marks (Table 4 and Supplementary Table 19). This enrichment was highest in leukocytes (27 leukocytes, of which 19 (70%) were lymphocytes and monocytes). For example, an FDR ≤5% for enrichment of asthma loci in active enhancers was observed in 100% of leukocytes compared with 50% of all cell types. The enrichment of asthma risk variants for colocalization with DNa-1–hypersensitive sites was intermediate between the enrichments in promoters and enhancers and was again elevated in blood cells (FDR ≤5% in 40% of leukocytes and 12% of all cell types) (Table 4 and Supplementary Table 20).

The strong enrichment of asthma loci in enhancer marks, especially in immune cells, indicated that the associated genetic variants are likely to be involved in the regulation of immunologically related functions. This finding also suggested that epigenetic mechanisms may be key to promoting asthma, as evidenced by IgE levels, an asthma-associated phenotype.

**Connectivity among asthma-associated loci.** To characterize the degree of connectivity among the 18 asthma-associated loci, we applied the Gene Relationships Across Implicated Loci (GRAIL) text-mining approach. Genes at 11 of these loci showed connections with a GRAIL score P_{GRAIL} ≤5% (and seven of them were highly connected, with P_{GRAIL} <10^-3) (Fig. 2 and Supplementary Table 21). These genes were connected through keywords such as ‘asthma’, ‘allergy’, ‘atopic’, ‘interleukin’, ‘cytokines’, ‘airway’, and ‘inflammation’, thus confirming the central role of immunologically related mechanisms accounting for these connections.

**Discussion**

In this meta-analysis of worldwide asthma GWAS in ethnically diverse subjects, we identified nine new loci influencing asthma risk. Our findings confirm that immunologically related mechanisms are prominent in asthma susceptibility and provide new insights that may open new avenues for future asthma research. The asthma-associated loci identified by TAGC are enriched in enhancer marks and are likely to be involved in gene regulation. Although these findings were observed in immune cells, asthma-associated genes (e.g., IL1RL1, TSLP, IL33, and ORMDL3/GSDMB) are also expressed in the airway epithelium, where they modulate airway inflammation. Investigation of epigenetic marks in airway epithelial cells may provide additional insight. The best candidates at many loci are involved in immune responses to viruses or bacteria, thereby underscoring the importance of infections in asthma risk. This study further provides evidence of an overlap of asthma loci with loci underlying autoimmune diseases and other diseases with an inflammatory component, thereby strengthening the growing understanding of the importance of pleiotropy in multifactorial diseases.

Our meta-analysis doubles the number of asthma cases analyzed in prior genome-wide studies at the time of analysis. We identified 878 SNPs corresponding to 22 distinct association signals at 18 loci meeting criteria for genome-wide significance in European-anxiety and/or multiancestry populations. Pooling data from ethnically diverse populations can increase the power to detect new loci (in this study, two loci reached the genome-wide threshold only in the multiancestry analysis) but may also increase heterogeneity. Beyond differences in the genetic background, varying environmental-exposure conditions can modify genetic risk and result in heterogeneity in SNP effect size, and consequently make the power of multiancestry analysis lower than that of ancestry-specific analysis. If asthma prevalence is assumed to be 10%, the variance in asthma liability explained by the 22 distinct genome-wide-significant variants in this study was estimated to be 3.5% (95% confidence interval 2.0–5.4%) of which 72% was accounted for by the known loci, and 28% was accounted for by the new loci. Notably, the current study was based on HapMap2-imputed data, which were shared within the TAGC consortium and thus allowed for detection of associations with common genetic variants (MAF ≥1%).

The overall relative paucity of asthma risk loci detected by large-scale GWAS, as compared with the number of risk loci identified for other common diseases, may be due to the clinical heterogeneity of asthma and the important etiological role of differing environmental-exposure conditions. Asthma is thought to be not a single disease but a syndrome that varies according to many characteristics, including the age of asthma onset, the severity of disease, the type of cellular inflammatory infiltrates, occupational exposure, and the varying response to treatment. It is thus possible that additional asthma loci may be identified by studies targeting more specific asthma subphenotypes and/or considering environmental exposure.

In conclusion, future discoveries might result from exploring more complex models of asthma phenotypes and from joint analysis of asthma and other immunologically mediated and inflammatory diseases. The central role of gene-regulatory mechanisms highlighted by our study might prompt genome-wide exploration of the epigenome in immune cells and the respiratory epithelium while integrating information on genetic variation and environmental-exposure histories.

**URLs.** National Human Genome Research Institute (NHGRI) and European Bioinformatics Institute (EBI) catalog of published genome-wide association, https://www.ebi.ac.uk/gwas/; 1000 Genomes Project Consortium Phase 3, http://www.international-algenome.org/; Genome-wide Complex Trait Analysis (GCTA), http://csgenomics.com/software/gcta/; Blood eQTL browser, https://omictools.com/blood-eqtl-browser-tool; GTEx, http://www.gtexportal.org/; Multiple Tissue Human Expression Resource (MuTHER) database, http://www.muther.ac.uk/; eQTL database in lymphoblastoid cell lines from MRCA and MRCE families,
References
1. Akinbami, L. J. et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001–2010. (U.S. Department of Health and Human Services, Washington, DC, 2012; 1–8. (NCHS Data Brief no. 94).
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. Wellen, K. E. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42, D1001–D1006 (2014).
4. Igaruia, C. et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. Nat. Commun. 6, 9565 (2015).
5. Bouzigon, E. et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. N. Engl. J. Med. 359, 1985–1994 (2008).
6. 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 (2015).
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. Ferreira, M. A. et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J. Allergy Clin. Immunol. 133, 1564–1571 (2014).
9. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1593–1608 (2002).
10. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44, 1084–1089 (2012).
11. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243 (2013).
12. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015).
13. Liang, L. et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res. 23, 716–726 (2013).
14. Hao, K. et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet. 8, e1003029 (2012).
15. Noguchi, E. et al. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. PLoS Genet. 7, e1002170 (2011).
16. Ding, L. et al. Rank-based genome-wide analysis reveals the association of pyranoide receptor-2 gene variants with childhood asthma among human populations. Hum. Genomics 7, 16 (2013).
17. Sleiman, P. M. et al. Variants of DENND1B associated with asthma in children. N. Engl. J. Med. 362, 36–44 (2010).
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. Torterolo, D. et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat. Genet. 43, 887–892 (2011).
20. Benneleykky, K. et al. A genome-wide association study identifies CTHRC3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat. Genet. 46, 51–55 (2014).
21. Ferreira, M. A. et al. Identification of IL6R and chromosome 11q13.3 as risk loci for asthma. Lancet 378, 1066–1074 (2011).
22. Moffatt, M. F. et al. A large-scale, consortium-based genomewide association study of asthma. N. Engl. J. Med. 363, 1211–1221 (2010).
23. Zeller, T. et al. Genetics and beyond: the transcriptome of human monocytes and disease susceptibility. PLoS One 5, e10693 (2010).
24. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).
25. Raychaudhuri, S. et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 5, e100534 (2009).
26. Hong, S. W., Kim, S. & Lee, D. K. The role of Bach2 in nucleic acid-triggered antiviral innate immune responses. Biochem. Biophys. Res. Commun. 365, 426–432 (2008).
27. Yang, M., He, R. L., Benovic, J. L. & Ye, R. D. Beta-Arrestin1 interacts with the G-protein subunits β1y2 and promotes β1y2-dependent Akt signalling for NF-kappaB activation. Biochem. J. 417, 287–296 (2009).
28. Soler Artigas, M. et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat. Genet. 43, 1082–1090 (2011).
29. Goenka, S. & Kaplan, M. H. Transcriptional regulation by STAT6. Immunol. Res. 50, 87–96 (2011).
30. Qian, X., Gao, Y., Ye, X. & Lu, M. Association of STAT6 variants with asthma risk: a systematic review and meta-analysis. Hum. Immunol. 75, 847–853 (2014).
31. Wang, Y., Tong, X. & Ye, N. Ndp1 negatively regulates RIG-I-dependent immune signaling by enhancing E3 ligase Smurf1-mediated MAVS degradation. J. Immunol. 189, 5304–5313 (2012).
32. Venuprasad, K., Zeng, M., Baughan, S. L. & Massoumi, R. Multifaceted role of the ubiquitin ligase Itch in immune regulation. Immunol. Cell Biol. 93, 452–460 (2015).
33. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167, 1369–1384 e149 (2016).
34. Barnes, P. J. Pathophysiology of allergic inflammation. Immunol. Rev. 242, 31–50 (2011).
35. Vicente, C. T. et al. Long-range modulation of PAG1 expression by 8q21 allergy risk variants. Am. J. Hum. Genet. 97, 329–336 (2015).
36. Davison, L. J. et al. Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. Hum. Mol. Genet. 21, 32–33 (2012).
37. Wang, L. et al. CPAG: software for leveraging pleiotropy in GWAS to reveal similarity between human traits links plasma fatty acids and intestinal inflammation. Genome Biol. 16, 190 (2015).
38. Rottem, M. & Shoenfeld, Y. Asthma as a paradigm for autoimmune disease. Int. Arch. Allergy Immunol. 132, 210–214 (2003).
39. Li, X. et al. Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. J. Allergy Clin. Immunol. 130, 861–868.e7 (2012).
40. Hayes, J. E. et al. Tissue-specific enrichment of lymphoma risk loci in regulatory elements. PLoS One 10, e0139360 (2015).
41. Jiang, L. et al. An epigenome-wide association study of total serum immunoglobulin E concentration. Nature 520, 670–674 (2015).
42. Wenzel, S. E. Asthma phenotypes: the evolution from clinical to molecular approaches. Nat. Med. 18, 716–725 (2012).

Acknowledgements
We thank all participants who provided data for each study and also thank our valued colleagues who contributed to data collection and phenotypic characterization of clinical samples, genotyping, and analysis of individual datasets. Detailed acknowledgments and funding for individual studies can be found in the Supplementary Note.

Author contributions
TAGC study management: F.D., K.C.B., W.O.C.C., M.F.M., C.O., and D.L.N.
F.D. and D.L.N. designed the study and wrote the manuscript. F.D., D.L.N., and P.M.-J. designed and conducted the statistical analysis. K.C.B., W.O.C.C., M.F.M., and C.O. designed the study and wrote the manuscript. M.B., A.V., S. Letort, and H.M. carried out the quality control of the data and performed statistical analysis.
AAGC (Australia): study principal investigators (Pis), M.A.E., M.C.M., C.R.F., and P.I.T.; data collection or analysis, M.A.E., M.C.M., C.R.F., G.I., and P.I.T.
ALLERGEN Canadian Asthma Primary Prevention Study (CAPPs) and Study of Asthma, Genes and the Environment (SAGE): study Pis, A.B.B., M.C.Y., D.D., and A.L.K.; data collection or analysis, D.D. and J.E.P.; study phenotyping, A.B.B. and M.C.Y.
Saguenuy-Lac Saint-Jean (SLSI) Study: study Pis, C.I. and T.J.H.; study design and management, C.L.
Analysis in Population-based Cohorts of Asthma Traits (APCAT) Consortium: study Pis, J.N.H., M.-R.J., and V. Salomaa. Framingham Heart Study (FHS): study Pi, G.T.O.; data collection or analysis, S.V. and Z.G. The European Prospective Investigation of Cancer (EPIC)-Norfolk: study Pi, N.J.W.; data collection or analysis, J.H.Z. and R.S.
Northern Finland Birth Cohort of 1966 (NFBC1966): study PIs, M.-R.J.; data collection or analysis, A.C.A. and A.R. FINKRISK: study PI, V. Salomaa; data collection or analysis, M. Kuokkanen and T. Laitinen. Health 2000 (H2000) Survey: study PIs, M.H. and P.J.; data collection or analysis, M. Kuokkanen and T.H. Helsinki Birth Cohort Study (HBCS): study PI, J.G.E.; data collection or analysis, E.W. and A. Palotie. Young Finns Study (YFS): study PI, O.T.R.; data collection or analysis, T. Lehtimäki and M. Kähönen.

African Ancestry Studies from the Candidate Gene Association Resource (CARe) Consortium: study PIs, J.N.H. and S.S.R.; data collection or analysis, C.D.P., D.B.K., I.J.S., R.K., K.M.B., and W.B.W.

Multi-Ethnic Study of Atherosclerosis (MESA): study PIs, R.G.B. and S.S.R.; data collection or analysis, K.M.D. and A.M.

Atherosclerosis Risk in Communities Study (ARIC): study PI, S.J.L.; data collection or analysis, S.J.L. and I.R.L.

Cardiovascular Health Study (CHS): study PIs, S.A.G. and S.R.H.; data collection or analysis, G.L., S.A.G., and S.R.H.

deCode genetics: study PIs, K.S., I.J., D.F.G., U.T., and G.T.; data collection or analysis, I.J., D.F.G., and G.T.; study phenotyping, U.S.B.

Early Genetics and Lifecourse Epidemiology (EAGLE) Consortium: PI, I.B.

Copenhagen Prospective Study on Asthma in Childhood (COPSAC): study PIs, H.B. and K.B.; data analysis, E. Kreiner and J.W.; study phenotyping, K.B. Danish National Birth Cohort (DNBC): study PI, M.M.; data collection or analysis, B.F. and F. Geller. GENERATION R: study PI, J.C.d.L.; data collection or analysis, R.J.P.d.V., I.D., and V.W.J. GENENplus/LISAplus: study PI, J. Heinrich; genotyping, data collection or analysis, M. Standl and C.M.T.T.; study phenotyping, J. Heinrich. Manchester Asthma and Allergy Study (MAAS): study PIs, A.S. and A.C.; data collection or analysis, I.A.C.

Western Australian Pregnancy Cohort Study (RAINE): study PI, P.H.; data collection or analysis, W.A. and C.E.P.

British 1958 Birth Cohort (B58C) Study: PI and statistical analysis, D.P.S.

EVE Consortium: study PIs, C.O., D.L.N., K.C.B., E. Bleecker, E. Burchard, J. Gauderman, F. Gilliland, S.J.L., F.J.M., D.M., I.R., S.T.W., L.K.W., and B.A.R.; data collection or analysis, D.L.N.; J. Gauderman, S.J.L., D.M., D.G.T., B.A.R., B.E.H., P.E.G., M.T.S., C.E., B.E.D.-R.-N., J.J.Y., A.M.L., R.A. Myers, R.A. Mathias, and T.H.B.

Japanese Adult Asthma Research Consortium (JAARC): study PI, T.T.; data collection or analysis, T.T., A.T., and M. Kubo.

Japan Pediatric Asthma Consortium (JPAC): study PI, E.N.; data collection or analysis, H.I.H. and K.M.

GABRIEL Consortium: study PIs, W.O.C.C. and E.V.M.; genotyping, M.L.; data analysis, E.B., F.D., M.F., and D.P.S. Epidemiological study on the Genetics and Environment of Asthma (EGEA): study PIs, V. Siroux and F.D.; genotyping, data collection or analysis, M.L. and E. Bouzigon. Avon Longitudinal Study of Parents and Children (ALSPAC): study PI, J. Henderson; genotyping, data collection or analysis, W.L.M. and R.G.; study phenotyping, J. Henderson. European Community Respiratory Health Survey (ECRHS): study PI, D.J.; data collection or analysis, C.J. and J. Heinrich. Children, Allergy, Milieu, Stockholm, Epidemiology (BAMSE) study: study PIs, E.M., M.W., and G.P. Busselton Health Study: study PIs, A.W.M., A.J., and J.B.; genotyping, data collection or analysis, A.W.M., A.J., J. Hui, and J.B. GABRIEL Advanced Surveys: study PI, E.V.M.; data collection or analysis, M. Kabesch and J. Genuneit. Kusk State Medical University (KSMU) Study: study PI, A. Polonikov; data collection or analysis, M. Solodilova and V.I.

All authors provided critical review of the manuscript.

Competing interests
The authors affiliated with deCODE (D.F.G., I.J., K.S., U.T., and G.T.) are employees of deCODE genetics/Amgen. All other coauthors have no conflicts of interest to declare.

 Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-017-0014-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to F.D. or D.L.N.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Florence Demenais1,2*, Patricia Margaritte-Jeannin1,2, Kathleen C. Barnes3, William O. C. Cookson4, Janine Altmüller5, Wei Ang6, R. Graham Barr7, Terri H. Beaty8, Allan B. Becker9, John Beilby10, Hans Bisgaard1, Unnur Steina Bjornsdottir11, Eugene Bleecker12, Klaus Bønnelykke11, Dorret I. Boomsma13, Emmanuelle Bouzigon1,2, Christopher E. Brightling15, Myriam Brossard1,2, Guy G. Brusselle16,17,18, Esteban Burchard19, Kristin M. Burkart20, Andrew Bush21,22, Moira Chan-Yeung23, Kian Fan Chung21,24, Alessander Couto Alves25, John A. Curtin26, Adnan Custovic27, Denise Daley23,28, Johan C. de Jongste29, Blanca E. Del-Rio-Navarro30, Kathleen M. Donohue31, Liesbeth Duijts32, Celeste Eng32, Johan G. Eriksson33, Martin Farrall34,35, Yuliya Fedorova36, Bjarke Feenstra37, Manuel A. Ferreira38, Australian Asthma Genetics Consortium (AAGC) collaborators39, Maxim B. Freidin40, Zofia Gajdos41,42, Jim Gauderman43, Ulrike Gehring44, Frank Geller37, Jon Genuneit45, Sina A. Gharib46, Frank Gilliland43, Raquel Granell47,48, Penelope E. Graves49, Daniel F. Gudbjartsson50,51, Tari Haathtela52, Susan R. Heckbert53, Dick Heederik44, Joachim Heinrich54,55, Markku Heliövaara56, John Henderson47,48, Blanca E. Himes57, Hiroshi Hirose58, Joel N. Hirschhorn42,59,60, Albert Hofman17,61, Patrick Holt62, Jouke Hottenga14, Thomas J. Hudson63,64, Jennie Hu10,65,66, Medea Imboden50, Vladimir Ivanov69, Vincent W. J. Jaddoe70, Alan James71,72, Christier Janson73, Marjo-Riitta Jarvelin74,75,76,77, Deborah Jarvis21,78, Graham Jones79, Ingileif Jonsdottir80, Pekka Jousilahti26, Michael Kabesch81, Mika Kähönen82, David B. Kantor83,84, Alexandra S. Karunas36,85, Elza Khusnutdinova36,85, Gerard H. Koppelman86,87, Anita L. Kozyrskyj88, Eskil Kreiner11, Michaaki Kubo89, Rajesh Kumar90,91, Ashish Kumar67,68,92, Mikko Kuokkanen56,93, Lies Lahousse17,94, Tarja Laitinen95, Catherine Laprise96,97, Mark Lathrop98, Susanne Lau99, Young-ae Lee100,101, Terho Lehtimäki102, Sébastien Letort1,2, Albert M. Levin103, Guo Li46, Lining Liang61,103, Laura R. Loehr105, Stephanie J. London106, Daan W. Loth17, Ani Manichaikul107, Ingo Marenholz100,101, Fernando J. Martinez49, Melanie C. Matheson108, Rasika A. Mathias109, Kenji Matsumoto110, Hamdi Mbarek10,14, Wendy L. Mc Ardle111, Mads Melbye37,112,113, Erik Melén92,114,115, Deborah Meyers13, Sven Michel81, Hamida Mohamd1,2, Arthur W. Musk116,117, Rachel A. Myers118, Maartje A. E. Nieuwenhuis87,119, Emiko Noguchi120, George T. O’Connor121,122, Ludmila M. Ogorodova123, Cameron D. Palmer42,59, Aarno Palotie93,124,125, Julie E. Park10,12, Craig E. Pennell6, Göran Pershagen92,114, Alexey Polonikov69, Dirkje S. Postma87,119, Nicole Probst-Hensch67,68, Valery P. Puzyrev40, Benjamin A. Raby126, Olli T. Raitakari127, Adaikalan Ramanasamy25,128, Stephen S. Rich107, Colin F. Robertson129, Isabelle Romieu130,131, Muhammad T. Salam43,132, Veikko Salomaa26, Vivi Schlüssen133, Robert Scott134, Polina A. Selivanova135, Torben Sigsgaard133, Angela Simpson36,136, Valérie Siroux37,138, Lewis J. Smith139, Maria Solodilova69, Marie Standl55, Kari Stefansson50,80, David P. Strachan140, Bruno H. Stricker17,141,142, Atsushi Takahashi89, Philip J. Thompson143, Guðmar Thorleifsson50, Unnur Thorsteinsdottir50,80, Carla M. T. Tiesler55,144, Dara G. Torgerson12, Tatsukiho Tsunoda89,145, André G. Uitterlinden142, Ralf J. P. van der Valk146, Amaury Vaysse1,2,12, Sailaja Vedantam41,42, Andrea von Berg147, Erika von Mutius148,149, Judith M. vonk87,150, Johannes Waage11, Nick J. Wareham134, Scott T. Weiss126, Wendy B. White151, Magnus Wickman92,152, Elisabeth Widén93, Gonneke Willemse14, L. Keoki Williams153,154, Inge M. Wouters14, James J. Yang155, Jing Hua Zhao134, Miriam F. Moffatt4, Carole Ober156 and Dan L. Nicolae157*.

*Genetic Variation and Human Diseases Unit (UMR-946), Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France. 1Institut Universitaire d’Hématologie, Université Paris Pardoret, Université Sorbonne Paris Cité, Paris, France. 2Division of Biomedical Informatics and Personalized Medicine, Colorado Center for Personalized Medicine, University of Colorado, Denver, CO, USA. 3Section of Genomic Medicine, National Heart and Lung Institute, London, UK. 4Cologne Center for Genomics and Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany. 5School of Women’s and Infants’ Health, University of Western Australia, Perth, Western Australia, Australia. 6Departments of Medicine and Epidemiology, Columbia University, New York, NY, USA. 7Division of Genetic Epidemiology, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. 8Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada.
Methods

GWAS and shared data. All 66 GWAS from the TAGC consortium are described in the Supplementary Note and are summarized in Supplementary Table 1. These GWAS included 36 studies of individuals of European ancestry (19,954 asthma cases, 91,895 controls), seven studies of individuals of African ancestry (2,149 asthma cases, 6,055 controls), two studies of individuals of Japanese ancestry (1,239 asthma cases, 3,976 controls), and one study of individuals of Latino ancestry (606 asthma cases, 792 controls), with a total of 23,948 asthma cases and 118,538 controls. There were 27 studies including only childhood-onset asthma (defined as asthma diagnosed at or before 16 years of age), thus allowing us to separately analyze a pediatric subgroup (8,976 asthma cases, 18,399 controls). All subjects provided informed consent to participate in genetic studies, and the local ethics committee for each individual study approved the study protocol. The definition of asthma was based on physicians’ diagnoses and/or standardized questionnaires (details in Supplementary Note). The samples were genotyped on a variety of commercial arrays. Analyses described in the Supplementary Note and Supplementary Table 2. GWAS were performed on imputed SNP data that were generated with HapMap2 as the reference panel in one of the commonly used imputation programs (Supplementary Note and Supplementary Table 2). In each dataset, the effect of each individual SNP on asthma, assuming an additive genetic model, was estimated through a logistic-regression-based approach and is expressed in terms of a regression coefficient with its standard error; the detailed methodology and software used for analysis in each study can be found in the Supplementary Note and Supplementary Table 2.

Imputation, quality control (including adjustments for population stratification), and analysis were performed by each group independently, and on a predefined set of 3,952,683 autosomal SNPs were shared. These SNPs were those of the HapMap Phase 2, release 21 panel in subjects of European, Asian, and African ancestry that were filtered through SNP annotation from build 37.3 of the reference sequence and dbSNP build 135 (31,387 SNPs (0.8% of all SNPs) from previous annotations that showed discrepancies with the chosen annotation were deleted). The variables that were shared contained the study name, general information on SNPs (rsID, chromosome, position, alleles (baseline and effect alleles as used in the analysis by each study), SNP status (imputed or genotyped SNP and whether the SNP genotype or imputed value was used in computation), quality control (QC) indicators (call rate and P value for the Hardy–Weinberg (HW) equilibrium test for genotyped SNPs, software used for imputation, and imputation quality score for imputed SNPs), allele frequencies in individuals with asthma and control individuals, and information on association statistics (regression coefficient for SNP effect, standard error of regression coefficient, Z scores, and P values associated with Z-score statistics).

Quality control of shared data. For each SNP, the alleles on the HapMap2 template (reference and alternate alleles on the positive strand) were compared with the alleles (baseline and effect alleles) used in the analysis by each group. When necessary, the association variables (allele frequencies, regression coefficient for SNP effect, and Z score) were switched to match the reference/alternate alleles of the template. Data for each SNP showing any ambiguity or error in assignment to the template were set to missing. In addition, several QC checks were performed regarding the name, format, range of possible values for all shared variables mentioned in the previous paragraph, and consistency across variables. Any problem or inconsistency was corrected; otherwise, the data for that SNP were set to missing. After this first stage of QC, association statistics of all SNPs had P > 0.01 in both controls and affected individuals. When an imputed SNP value was used in the analysis, the criteria were imputation quality score ≥ 0.5 and MAE ≥ 0.01 in both controls and asthma cases. The distribution of the summary statistics (regression coefficient for SNP effect, standard error, and Z score) of all SNPs passing QC was examined for each study; SNPs that still showed extreme Z scores (≥ 7) after QC were excluded.

Meta-analysis of asthma GWAS. We conducted fixed-effects meta-analysis with inverse variance weighting and random-effects meta-analysis, using the DerSimonian and Laird estimator of the between-study variance, when the meta-analyses included a large number of studies (European-ancestry, multicaseancestry and pediatric-subgroup meta-analyses), thus allowing for an accurate estimate of the between-study variance. We used a fixed-effects model for the meta-analyses of the African-ancestry, Japanese-ancestry, and Latino-ancestry populations. For all these meta-analyses, we used the SNP regression coefficient and standard error from each study for which the SNP passed QC. All meta-analyses were done with Stata version 14.1. To minimize the false-positive findings and to obtain robust results, we examined the combined results for SNPs for which at least two-thirds of the studies had a Z score > 4.1 (and the meta-analysis corresponding effect sizes were performed by using a standard normal distribution. We applied a threshold of P < 0.001 (or PMeta ≤ 5 × 10–8) to declare a combined SNP effect as genome-wide significant. To verify the robustness of the results, we applied a genomic control correction to the association test statistics. The lead SNP at a locus was the variant with the strongest evidence of association in the European-ancestry or multicaseancestry meta-analysis. We defined a support interval around the lead SNP designated as ‘locus’; the bounds of this interval were the positions of the two most extreme SNPs among all SNPs that were located within 500 kb on each side of the lead SNP and had PMeta (or PMeta) ≤ 10–6. The heterogeneity of per-SNP effect sizes across all studies in a meta-analysis was assessed with Cochran’s Q test. Differences among the four ethnic-specific summary effects were also tested with Cochran’s Q statistic.

Conditional analysis of asthma-associated loci. GCTA software (URLs) was used to perform approximate conditional analysis for all loci with at least one SNP reaching the genome-wide-significance level. This approximate conditional analysis is based on the summary meta-analysis statistics obtained under a fixed-effects model and takes into account the correlations among SNPs that are estimated from a large reference population. In the meta-analysis of the conditional analysis, we defined a set of all SNPs that were genome-wide significant. To verify the robustness of the results, we applied a threshold of P < 0.001 (or PMeta ≤ 5 × 10–8) to the lead SNP and had PMeta (or PMeta) ≤ 10–6. The heterogeneity of per-SNP effect sizes across all studies in a meta-analysis was assessed with Cochran’s Q test. Differences among the four ethnic-specific summary effects were also tested with Cochran’s Q statistic.

Identification of cis eQTLs at new asthma risk loci. To obtain greater insight into the genes potentially driving the association signals at the new asthma loci, we used the following strategy: (i) a meta-analysis of the transcriptional profiles from peripheral blood cells of 5,311 individuals of European ancestry (the blood eQTL browser); (ii) gene expression data from 777 lymphoblastoid cell lines from the MuTHER consortium; (iii) transcriptional profiles of 405 and 550 lymphoblastoid cell lines and lung tissue: (i) a meta-analysis of the transcriptional profiles from peripheral blood cells of 5,311 individuals of European ancestry (the blood eQTL browser); (iv) GTeX data from 1,490 individuals included in the GHS-express database; (v) GTeX eQTL Browser data from multiple tissues including the brain and lungs; and (vi) transcriptional profiles from the lung tissues of 1,111 subjects (URLs).

Search for missense variants at new asthma risk loci. To complement the eQTL analysis, we searched whether the lead asthma-associated SNPs and secondary signals within 500 kb had any overlap with non-synonymous variants. For each investigated SNP, we interrogated six publically available eQTL databases, giving priority to cell types more likely to be involved in asthma biology (blood cell types and lung tissue): (i) a meta-analysis of the transcriptional profiles from peripheral blood cells of 5,311 individuals of European ancestry (the blood eQTL browser); (ii) gene expression data from 777 lymphoblastoid cell lines from the MuTHER consortium (606 and 792 controls); (iii) transcriptional profiles of 405 and 550 lymphoblastoid cell lines from UK asthma (ARCA) and eczema (MRCE) family members, respectively; (iv) GTeX data from 1,490 individuals included in the GHS-express database; (v) GTeX eQTL Browser data from multiple tissues including the brain and lungs; and (vi) transcriptional profiles from the lung tissues of 1,111 subjects (URLs).

Overlap of loci associated with asthma and other phenotypes. Overlap of new asthma risk loci with associations with allergy-related phenotypes/diseases and immunologically related diseases as well as lung-function phenotypes was first annotated by using the 24 March 2015 version of the NHGRI–EBI GWAS catalog (URLs). We then used this catalog to identify potentially overlapping asthma signals with PMeta ≤ 10–8 in the multicaseancestry meta-analysis with association signals of all diseases and traits in the catalog. That version of the catalog comprised 19,080 SNP entries, 16,047 of which had a TAGC asthma-association P value. To investigate pleiotropy, we filtered out SNPs associated with asthma in the database, SNPs with a reported GWAS P value > 10–8 (with the intent of removing some of the potential false positives in the catalog) and SNPs that were not investigated. However, for the 17q12-21 locus, where there was no heterogeneity in the pediatric subgroup, GCTA was restricted to the European-ancestry pediatric subgroup. We used the large ECRHS dataset as the reference sample to estimate LD. This dataset was genotyped with the Illumina Human610Quad array and included 2,101 unrelated individuals to establish QC. After genotyping was performed with MAGIC software and the HapMap2, release 21 panel; only well-imputed SNPs (imputation quality score r2 > 0.8) with MAF ≥ 1% were retained in this reference sample. For each asthma-associated locus, the region explored by conditional analysis extended 500 kb on each side of the two extreme SNPs defining the support interval around the lead SNP (described in preceding paragraph). However, we decreased that extension to 250 kb for the 6p21.33 and 6p21.32 loci to avoid overlap. The length of the regions explored by conditional analysis varied from 0.61 Mb to 1.63 Mb. Within each region investigated by conditional analysis, summary effects for SNPs belonging to that region were adjusted for the lead SNP by using the --cojo-cond option; tests for the adjusted SNP effects were based on the two-sided Wald test. If there was an additional SNP meeting the Bonferroni-corrected threshold for the total number of SNPs over all regions investigated by GCTA (P = 4.1 × 10–8), after adjustment for the lead SNP, we performed an additional round including both SNPs. If the remaining SNPs had P > 4.1 × 10–8, no further analysis was performed. The results of this analysis are reported in Supplementary Table 15.
For some diseases or quantitative traits, there were multiple SNPs in the same region reported in the catalog, thus potentially yielding redundant information. Some of the SNPs might have been in strong LD, whereas others might have reflected independent signals. To avoid possible duplication of signals, we retained only unique trait–loci combinations, as reflected by the variables ‘disease trait’ and ‘region’ in the catalog. There were 4,231 unique entries remaining after this filtering step. Diseases/traits in the GWAS catalog were grouped according to the classification from Wang et al. The overlap was estimated as the binomial-tail probability for observing the number of TAGC SNPs with \( P_{\text{random}} \leq 10^{-4} \) among the number of SNPs reported in the GWAS catalog for a group of diseases. The significance threshold for enrichment in shared associations between a disease group and asthma was set to 0.05 divided by the number of disease groups investigated, through a Bonferroni correction. Finally, we examined a larger set of SNPs in the GWAS catalog that showed an association with asthma at \( P_{\text{random}} \leq 10^{-4} \) in TAGC multiancestry meta-analysis and estimated the proportion of false positives among those SNPs.

**Enrichment of asthma risk loci in epigenetic marks.** To obtain greater insight into the functional role of the genetic variants at the new and known asthma loci identified in this study, we investigated whether the lead SNPs and their proxies \( (r^2 \geq 0.8) \) were concentrated in cis-regulatory DNA elements. We used the USES pipeline44 (URLs) that was adapted to the current study. This approach tests whether GWAS-identified SNPs are enriched in particular functional annotations through use of Monte Carlo simulations. The original set of asthma-associated SNPs included the lead SNPs at each asthma risk locus (i.e., one SNP per asthma-associated locus, as recommended by Hayes et al.40). We excluded the two associated loci spanning the HLA region (6p21.33 and 6p21.32), because of the high amount of variability and LD in that region. Each of the original lead SNPs was categorized according to its distance from the nearest transcription start site (TSS) and the number of LD partners \( (r^2 \geq 0.8) \). Quartiles for both the TSS distance and LD-partner count were calculated, and the initial SNPs were binned accordingly. Then, SNPs from the entire set of imputed SNPs used for analysis were binned according to the original SNP criteria (distance from the closest TSS, number of LD partners, and MAF). Random SNP sets were chosen, matching the original bin frequencies. LD partners \( (r^2 \geq 0.8) \) for both the original lead SNPs and random SNPs were retrieved. The SNP data, including the original and random sets of SNPs and their corresponding LD partners \( (r^2 \geq 0.8) \), were intersected with the cell-specific epigenome tracks of regulatory elements with BedTools intersectBed48, to determine which SNPs colocalized with a given type of regulatory elements (for example, enhancers or promoters). The resultant SNPs were then collapsed into loci that colocalized with marks according to LD structure. We computed an empirical \( P \) value for a specific track by using 10,000 random SNP sets (this \( P \) value was equal to \( n_{\text{rand}}/n \), where \( n_{\text{rand}} \) is the number of instances in which the frequency of colocalization of the random SNP sets with the regulatory feature was greater than or equal to the frequency of colocalization with the feature for the original SNP set, and \( n \) is the number of random SNP sets generated (here, 10,000)). We used Benjamini–Hochberg FDRs to correct for multiple testing. We interrogated the functional data from 111 Roadmap reference epigenomes and 16 additional epigenomes from ENCODE that are available in a wide range of human cell and tissue types45 (URLs). We focused on enhancers and promoters that were defined with the ChromHMM 15-state model and assayed in all 127 epigenomes. We also examined enrichment in DNase I–hypersensitive sites that were available in 51 cell types.

**Connectivity among asthma-associated loci.** We used GRAIL25 to assess the relatedness among asthma-associated loci. As previously described in detail40, to define the genes near each SNP GRAIL finds the furthest neighboring SNPs in the 3′ and 5′ direction that are in LD \( (r^2 \geq 0.5) \) and proceeds outward in each direction to the nearest recombination hotspot. All genes that overlap that interval are considered to be implicated by the SNP. If there are no genes in that region, the interval is extended by 250 kb in either direction. We used the genome-wide-significant signals identified by this study as seeds and queried loci to investigate the biological connectivity among those loci. The connectivity between genes belonging to these loci was assessed through text mining of PubMed abstracts. Each gene at each locus was scored for enrichment in GRAIL connectivity to genes located at the other loci by using statistical text-mining methods, as previously described45. The interconnectivity among genes at asthma risk loci was visualized using VIZGRAIL46 (URLs).

**Variance explained by the asthma-associated genetic variants.** We estimated the variance in asthma liability explained by the 22 distinct genome-wide-significant SNPs (18 lead SNPs plus four secondary signals identified by approximate conditional analysis) at the 18 asthma-associated loci, by using a method based on the liability threshold model48 and assuming a prevalence of asthma of 10%. The variance in asthma liability explained by individual SNPs was summed over all 22 significant variants. For the loci that included two SNPs (lead SNP and secondary signal), we used the SNP effect sizes estimated by approximate joint analysis by using GTCTA44. We also estimated the variance in asthma liability explained by the nine lead SNPs at the nine new asthma loci and by the 13 distinct genome-wide-significant signals at the nine known loci.

**Life Sciences Reporting Summary.** Further information on experimental design is available in the Life Sciences Reporting Summary.

**Data availability.** The summary statistics of the meta-analysis that support the findings of this study are available through a link from the GWAS Catalog entry for the TAGC study on the EMBL–EBI (European Bioinformatics Institute) website (https://www.ebi.ac.uk/gwas/downloads/summary-statistics).

**References**

43. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. Control. Clin. Trials 7, 177–188 (1986).
44. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, 369–375 (2012).
45. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. Annu. Rev. Genomics Hum. Genet. 10, 387–406 (2009).
46. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
47. Raychaudhuri, S. VIZ-GRAIL: visualizing functional connections across disease loci. Bioinformatics 27, 1589–1590 (2011).
48. So, H. C., Gui, A. H., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet. Epidemiol. 35, 310–317 (2011).
Experimental design

1. Sample size

Describe how sample size was determined.

This study is a meta-analysis of asthma genome-wide association studies (GWAS) that was conducted in the framework of the Trans-National Asthma Genetic Consortium (TAGC). This consortium brought together worldwide groups of investigators with genome-wide data available in a total of 142,486 individuals (23,948 cases, 118,538 controls) of diverse ancestries, thus providing enough power to discover new asthma loci, based on results from GWAS of similar size for other complex diseases. The sample sizes were reported by the groups forming the consortium.

2. Data exclusions

Describe any data exclusions.

The meta-analysis included a total of 66 GWAS based on HapMap2 imputed SNPs. Imputation, quality control (QC) and analysis was done by each group independently. Data (summary statistics for association between each SNP and asthma and QC criteria for each SNP) on a predefined set of 3,952,683 autosomal SNPs was shared. From this SNP panel, we excluded 620,238 ambiguous SNPs (for which the DNA strand cannot be determined) and 501,370 SNPs that did not pass the QC criteria for all 66 studies, thus making a total of 2,831,075 SNPs for the meta-analysis. To minimize the false-positive findings and to obtain robust results, we examined the combined results for 2 million SNPs for which at least two-thirds of the studies contributed to the meta-analysis (ie SNPs passed QC in at least two-thirds of the studies).

3. Replication

Describe whether the experimental findings were reliably reproduced.

Because the meta-analysis included almost all asthma GWAS that had been conducted worldwide when TAGC was formed, no reasonable replication could be performed. This meta-analysis combined summary statistics from various populations and, thus, took into account different sample sizes, SNP effect sizes, variances and allele frequencies from each of the populations under investigation. We assessed whether the effect sizes of newly discovered variants in this analysis were not statistically different across studies by testing for heterogeneity between them.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Randomization does not apply to our meta-analysis of summary statistics of asthma GWAS shared by the TAGC consortium. This is an observational study.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The meta-analysis was done by combining summary statistics and applying QC criteria (based on mathematical grounds) in a systematic manner for all studies. This is an observational study where no blinding was applied.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters
For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|-----|-----------|
| ☑   | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☑   | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☑   | A statement indicating how many times each experiment was replicated |
| ☑   | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☑   | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☑   | The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted |
| ☑   | A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☑   | Clearly defined error bars |

See the web collection on statistics for biologists for further resources and guidance.

7. Software
Describe the software used to analyze the data in this study.

The meta-analysis was done using Stata version 14.1 (STATA Corp., College Station, Texas, USA). All software are indicated in the URL section. Approximate conditional analysis was done using the Genome-wide Complex Trait Analysis (GCTA) software (see URLs). The analysis of co-localization of asthma risk variants with epigenetic marks was done using the Uncovering Enrichment through Simulation pipeline (see URLs). Connectivity between asthma-associated loci was investigated using the GRAIL software (see URLs). The statistical analysis of pleiotropy was done using the statistical software R.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

8. Materials availability
Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

The summary statistics of the meta-analysis that support the findings of this study are available through a link from the GWAS Catalog entry for the TAGC study on the EMBL-EBI (European Bioinformatics Institute) web site (https://www.ebi.ac.uk/gwas/downloads/summary-statistics).

9. Antibodies
Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

NA

10. Eukaryotic cell lines
a. State the source of each eukaryotic cell line used.

NA

b. Describe the method of cell line authentication used.

NA
c. Report whether the cell lines were tested for mycoplasma contamination.

NA
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

NA
Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

NA

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study involved combining summary statistics from individual analyses of association of SNPs with asthma performed by each participating group. An overview of all studies included in the meta-analysis is presented in Supplementary Table 1. Methods used for the individual analyses are shown in Supplementary Table 2. A brief description of the participants in each study is presented in the Supplementary Note.

As stated in the Online Methods, all subjects provided informed consent to participate in genetic studies and local ethics committees for each of the individual studies approved the study protocol.