ARTICLES
https://doi.org/10.1038/s41588-017-0014-7
nature genetics

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[1]. 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[3]. 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[4], 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[5] 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_{fixed}) and random-effects (P_{random}) models (Methods), and a conventional P_{random} (or P_{fixed}) threshold of 5×10^{-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 ($P_{random}\leq5\times10^{-8}$) 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 ($P_{random}\leq5\times10^{-8}$), 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).

Full list of authors and affiliations appear at the end of the paper.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

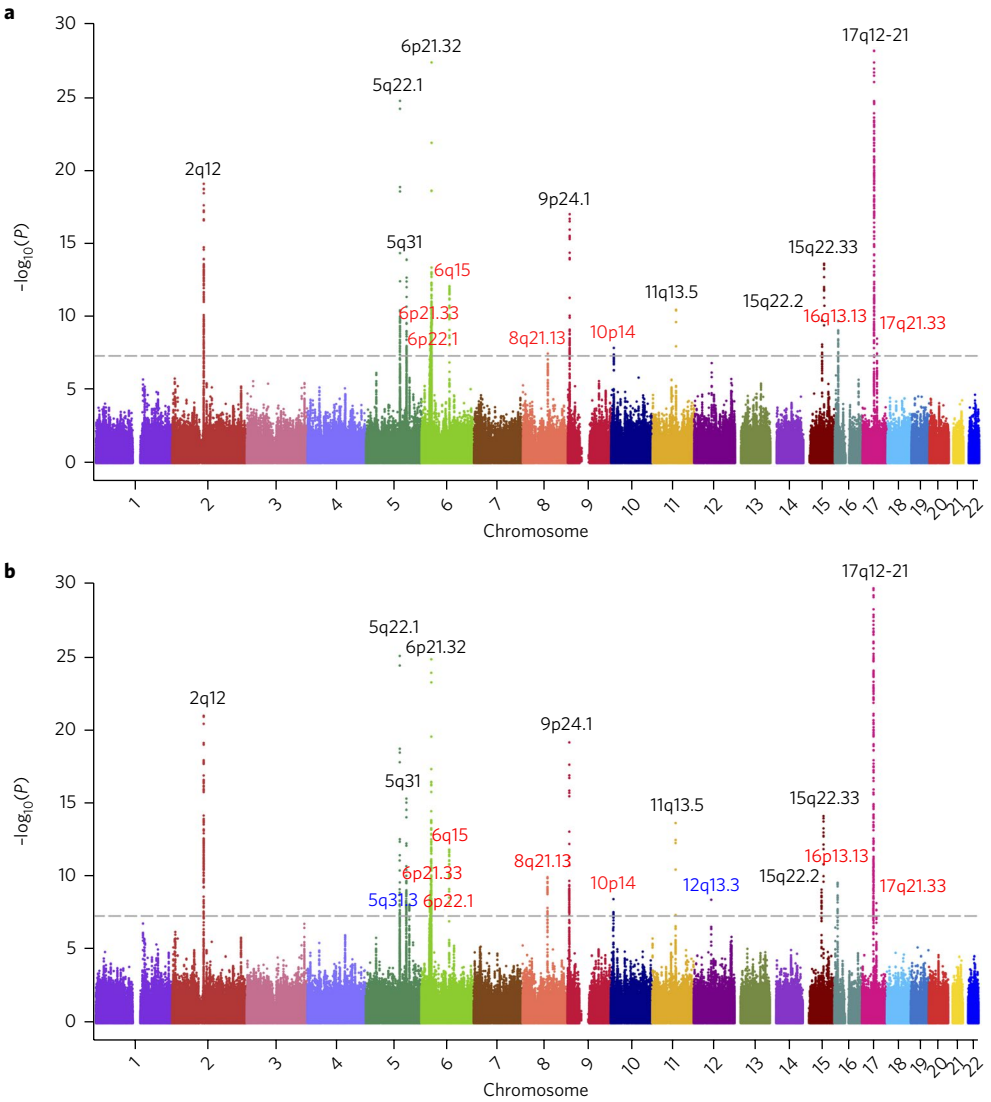

Fig. 1 | Manhattan plots of the results of European-ancestry and multiancestry random-effects meta-analyses of asthma risk. a, The European-ancestry meta-analysis pertains to 19,954 asthma cases and 107,715 controls. **b**, The multiancestry meta-analysis pertains to 23,948 asthma cases and 118,538 controls. Each locus is annotated according to its cytogenetic-band location. The *x* axis represents chromosomal location, and the *y* axis represents $-\log_{10}(P$ value) for tests of association between SNPs and asthma. Black, previously known loci; red, new loci identified in the European-ancestry meta-analysis; blue, additional new loci identified in the multiancestry meta-analysis The dashed horizontal line denotes $P=5\times10^{-8}$.

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[6] and Japanese[7] 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[8]; 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_{het} for Cochran's Q test[9] $=0.008$ across European-ancestry studies and $P_{het}=0.02$ across multiancestry studies; Table 1 and Supplementary Fig. 4). There was also significant evidence of heterogeneity in the ancestry-specific effect sizes ($P_{ethnic}=0.003$) 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 ($P_{random}\leq5\times10^{-8}$) 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* ($P_{random}=2.2\times10^{-30}$ in multiancestry meta-analysis), at least 180 kb from the previously recognized asthma-associated signals at the *GSDMB*/*ORMDL3* haplotype block[3] (Supplementary Fig. 7). This result was attributable to effect-size heterogeneity across studies ($0.001\leq P_{het}\leq0.05$) 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[5]. Indeed, in the pediatric group, the 17q12-21 SNPs did not show heterogeneity ($P_{het}\geq0.09$), and the lead SNP rs8069176 was 3.6 kb proximal to *GSDMB* ($P_{random}=P_{fixed}=4.4\times10^{-26}$), in agreement with results from previous studies[3,5]. The SNP effect sizes in the pediatric and nonpediatric studies showed a significant

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

Table 1 | Genetic loci associated with asthma in European-ancestry and multiancestry meta-analyses

						European-ancestry meta-analysis					Multiancestry meta-analysis				
Locus[a]	No. significant SNPs Eur-anc/multi-anc[b]	rsID	Position[c]	Nearby genes[d]	Allele[e]	EAF	OR	95% CI	P_{random}	P_{het}	OR	95% CI	P_{random}	P_{het}	P_{ethnic}
New asthma susceptibility loci															
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^{-9}	0.11	0.55
6p22.1	8/5	rs1233578	28712247	*GPX5*, *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^{-13}	0.78	0.91	0.89–0.94	2.2×10^{-12}	0.80	0.39
12q13.3	0/1	rs167769	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^{-9}	0.31	0.85
17q21.33	4/3	rs17637472	47461433	*ZNF652*, *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^{-9}	0.35	0.12
New signals at loci previously associated with asthma in ancestry-specific populations															
6p21.33	66/53	rs2855812	31472720	*MICB*, *HCP5*, *MCCD1*	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	9046645	*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
Asthma signals previously reported for asthma plus hay fever															
8q21.13	1/28	rs12543811	81278885	*TPD52*, *ZBTB10*	G/A	0.66	0.93	0.91–0.95	3.4×10^{-8}	0.47	0.92	0.90–0.95	1.1×10^{-10}	0.54	0.24
16p13.13	12/13	rs17806299	11199980	*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^{-10}	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^{-21}	0.61	0.64
5q22.1	35/32	rs10455025	110404999	*SLC25A46*, *TSLP*	A/C	0.34	1.15	1.12–1.18	2.0×10^{-25}	0.53	1.15	1.12–1.18	9.4×10^{-26}	0.57	0.27
5q31	33/62	rs20541	131995964	*IL13*, *RAD50*, *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	rs9272346	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*, *IL33*	A/G	0.75	0.85	0.82–0.88	1.1×10^{-17}	0.008	0.86	0.83–0.88	7.2×10^{-20}	0.02	0.57
11q13.5	4/5	rs7927894	76301316	*EMSY*, *LRRC32*	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*, *NARG2*, *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^{-9}	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^{-14}	0.75	1.10	1.08–1.13	7.4×10^{-15}	0.76	0.48
17q12-21	160/198	rs2952156	37876835	*ERBB2*, *PGAP3*, *MIEN1*	A/G	0.70	0.86	0.84–0.88	7.6×10^{-29}	0.55	0.87	0.84–0.89	2.2×10^{-30}	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 ($P_{random}<5\times10^{-8}$). Each locus in this table is represented by the SNP with the strongest evidence of association in the European-ancestry (127,669 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_{het}) and to test for differences among the four ancestry-specific summary effects (P_{ethnic}). EAF, effect allele frequency; OR, odds (log-additive) ratio; 95% CI, 95% confidence interval. [a]Cytogenetic band. [b]Number of genome-wide-significant SNPs ($P_{random}<5\times10^{-8}$) at each locus in European-ancestry meta-analysis/multiancestry meta-analysis. [c]SNP position, build 37. [d]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 *EMSY* and *MIEN1* were previously designated *C11orf30* and *C17orf37*, respectively. [e]Reference/effect allele.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

Table 2 | Main characteristics of the nine loci showing new associations with asthma

Locus[a]	Location of lead SNP[b]	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^{-9})		IBD
6p22.1	Intergenic	B: *ZSCAN12* (3.0×10^{-8}) L: *ZSCAN31* (6.5×10^{-11})	Lung function	
6q15	*BACH2* (intron)	B: *BACH2* (3.0×10^{-10})		MS, T1D, CD, IBD, V, IGG
12q13.3	*STAT6* (intron)	B: *STAT6* (9.8×10^{-198}) L: *STAT6* (3.7×10^{-37})	IgE (total, specific), lung function	Pso, ISP_IFN
17q21.33	Intergenic	B: *GNGT2* (2.1×10^{-52})	Atopic dermatitis	ISP_IL2
New asthma signals at loci previously associated with asthma in ancestry-specific populations				
6p21.33	*MICB* (intron)	B: *TNF* (4.8×10^{-14}), *LST1* (1.0×10^{-13}), *HLA-C* (3.2×10^{-13}), *LTA* (1.0×10^{-10}) L: *MICB* (4.6×10^{-13})	IgE (total, specific), self-reported allergy, atopic dermatitis, lung function	SLE, UC, RA, IBD, BS, GD, SS, AS, Pso, UC, V, WBC, MoC, DS, HIV-1, SJS, HB, HBV, IMN, CD4/CD8 ratio, HIV-1C
10p14	Intergenic	None	Self-reported allergy	RA, ISP_IL1B, ISPV
Asthma signals previously reported for asthma plus hay fever				
8q21.13	Intergenic	None	Atopic dermatitis, asthma plus hay fever, self-reported allergy	RA
16p13.13	*CLEC16A* (intron)	B: *DEXI* (2.2×10^{-43})	Atopic dermatitis, asthma plus hay fever	T1D, 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 1) or SNPs in LD ($r^2 \geq 0.5$) with the lead SNPs were searched in six eQTL databases from whole blood[11,12], lymphoblastoid cell lines[10,13], monocytes[23], and lung tissue[12,14]; only genes with the strongest associations ($P < 5\times10^{-8}$, 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[3]; IBD, inflammatory bowel disease (Crohn's disease), MS, multiple sclerosis, T1D, type 1 diabetes, CD, celiac disease, V, vitiligo, 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, SJS, Stevens–Johnson syndrome, HB, hepatitis B infection, HBV, hepatitis B vaccine response, IMN, idiopathic membranous nephropathy, CD4/CD8, CD4/CD8 lymphocyte ratio, HIV-1C, HIV-1 control, ISP_IL1B, immune response to smallpox (secreted IL-1β), ISPV, immune response to smallpox vaccine (IL-6), PBC, primary biliary cirrhosis, LEP, leprosy.
[a]Cytogenetic band. [b]The protein-coding genes flanking intergenic SNPs are shown in Table 1.

difference for rs8069176 at the *GSDMB* locus ($P_{het} = 7.4\times10^{-4}$) but no difference for rs2952156 at the *ERBB2* locus ($P_{het} = 0.11$). These two SNPs were in only moderate linkage disequilibrium (LD) ($r^2 = 0.30$), and each was in strong LD ($r^2 > 0.9$) with missense variants localized in *ERBB2* for the proxy of rs2952156 and in *ZPBP2* and *GSDMB* for the proxies of rs8069176. Moreover, both rs2952156 and rs8069176 are associated with expression of *GSDMB* and *ORMDL3* in blood[10–13], and with the expression of *GSDMA*, *CDK12*, *GSDMB*, and *ORMDL3* in whole lung tissue[12,14]. However, only rs2952156 is associated with *PGAP3* expression in the lung[12,14] (Supplementary Table 12a). Further exploration of expression quantitative trait loci (eQTL) data from Genotype-Tissue Expression (GTEx)[12] indicated that rs8069176 accounted for a large part of the association of the most significant SNP with *ORMDL3* expression in the blood, whereas rs2952156 accounted for a large part of the association of the most significant SNP with *PGAP3* expression in the lungs (Supplementary Table 12b), thus suggesting that the asthma-associated signals near the *PGAP3*/*ERBB2* and *ORMDL3*/*GSDMB* blocks may affect asthma risk through the expression of different genes in different tissues.

Finally, of the 21 published asthma loci, 12 did not reach genome-wide significance in TAGC (Supplementary Table 13). The most significant SNPs in the GWAS catalog[3] at seven of those loci had *P* values >0.01 in TAGC analyses. Among these seven nonreplicated loci, two (4q31.21 (ref. [7]) and 8q24.11 (ref. [15])) have been reported in Japanese individuals, three (4q12, 9p23, and 10q24.2)[16] had SNPs with low minor allele frequency (MAF ≤ 2%) and have been reported in a childhood-onset asthma study, and two (1q31.3 (ref. [17]) and 5q12.1 (ref. [18])) have been reported in children of European ancestry with asthma defined by current or persistent asthma symptoms with regular use of medication. The most significant SNPs at the remaining five loci had *P* values $\leq 5\times10^{-4}$ in at least one TAGC meta-analysis, thus providing some replication. Among these five loci, (i) the 1q23.1 locus is specific to African-ancestry populations[19]; (ii) the 12q13.2 SNP, reported in a study of Japanese individuals[7], showed heterogeneity in the TAGC Japanese-ancestry meta-analysis as well as the European-ancestry and multiancestry meta-analyses ($P_{het} \leq 0.05$); and (iii) the 7q22.3 SNP, previously reported in European-ancestry populations[20], was associated with a severe form of childhood asthma and also showed heterogeneity across studies in the original publication[20] (in which the P_{random} value did not reach significance) as well as in our study (European-ancestry, multiancestry, and pediatric meta-analyses, $0.006 \leq P_{het} \leq 0.03$). Finally, SNPs at the 1q21.3 and 22q12.3 loci, previously reported in European-ancestry populations[21,22], did not show significant evidence of heterogeneity across TAGC studies in the European-ancestry and multiancestry meta-analyses ($0.11 \leq P_{het} \leq 0.19$). When we repeated these two meta-analyses under a fixed-effects model and separately considered the set of TAGC datasets that were part of the original publication (set P) and the set of remaining TAGC datasets (set R), both 1q21.3 and 22q12.3 SNPs had higher effect sizes in set P than in set R. These differences in effect sizes did not reach significance for the 1q21.3 SNP (P_{het} for Cochran's *Q* test of 0.13 and 0.20 in the European-ancestry and multiancestry analyses, respectively) and were borderline significant for the 22q12.3 SNP ($0.04 \leq P_{het} \leq 0.06$) (Supplementary Table 14). Altogether, these results suggested that the lack of replication was mainly due to heterogeneity attributable to various factors, such as ancestry, specificity of clinical phenotypes, or other factors, as further discussed below.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

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 $P_{random} \leq 10^{-4}$ 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^{-19}
Neuropsychiatric	114	5	1.5×10^{-12}
Cancer	417	7	4.0×10^{-14}
Endocrine system	276	2	4.0×10^{-4}
Digestive system	347	16	1.4×10^{-37}
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[3] was investigated for all TAGC SNPs with $P_{random} \leq 10^{-4}$ in the multiancestry meta-analysis; diseases from the GWAS catalog were grouped according to the disease classification proposed by Wang et al.[37] (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 $P_{random} \leq 10^{-4}$ 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

	Proportion of all cell types (blood cell types) showing enrichment with a given FDR	
Type of regulatory elements	FDR ≤10%	FDR ≤5%
All promoter states	6% (26%)	0
Active promoter states	13% (33%)	0
All enhancer states	57% (100%)	44% (89%)
Active enhancer states	66% (100%)	50% (100%)
DNase I–hypersensitive sites	16% (50%)	12% (40%)

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 6p21.33 and 6p21.32 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 ChromHMM 15-state model applied to functional data of 127 Roadmap and ENCODE reference epigenomes in various cell types (including 27 leukocytes)[24]. DNase I–hypersensitive sites were identified in 51 cell types (including ten leukocytes)[24]. 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[40]; Benjamini–Hochberg FDR was calculated to correct for multiple testing (details in Methods).

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 ($P_{fixed} \leq 5\times10^{-8}$) (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[11,12], lymphoblastoid cell lines[10,13], and monocytes[23]), and whole lung tissue[12,14]. 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[3] while varying the strength of association with asthma (thresholds from 5×10^{-8} to 10^{-3}). 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[24]. Finally, the degree of connectivity among the asthma-associated loci was assessed through text mining[25]. 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 2 and Supplementary Table 16; regional plots are shown in Supplementary Fig. 3a.

New asthma susceptibility loci. Five new loci were identified in this study. The strongest new signal in both the European-ancestry ($P_{random} = 8.6\times10^{-13}$) and multiancestry ($P_{random} = 2.2\times10^{-12}$) meta-analyses was for SNP rs2325291 in an intron of *BACH2* at 6q15, which was strongly correlated with rs10455168 ($r^2 = 0.91$), a cis eQTL altering expression of *BACH2* in the blood[11]. *BACH2* encodes a ZIP transcription factor that regulates nucleic-acid-triggered antiviral responses in human cells[26]. The second-strongest signal in the European-ancestry and multiancestry analyses was for rs17637472 ($P_{random} = 3.3\times10^{-9}$ and 6.6×10^{-9}), which is located between *ZNF652* and *PHB* at 17q21.33 and is a strong cis eQTL for *GNGT2* (173 kb from rs17637472) in the blood[10,11,13,23]. GNGT2 interacts with β-arrestin 1 and consequently promotes G-protein-dependent AKT signaling in NF-κB activation[27].

Among the other new signals, the lead SNP rs1233578 at 6p22.1 ($P_{random} = 5.3\times10^{-9}$ in European-ancestry populations) was located between *TRIM27* and *GPX5*. This SNP was not associated with gene expression in the blood or lungs but was in LD ($r^2 = 0.6$ in European-ancestry populations) with rs7766356 (312 kb from rs1233578), which is a cis eQTL for *ZSCAN12* in the blood[13] and *ZSCAN31* in the lungs[14]. These genes encoding zinc-finger proteins are associated with lung function[28]. The two SNPs rs1233578 and rs7766356 represented the same association signal in European-ancestry populations (the association with rs7766356 became nonsignificant after conditioning on the lead SNP rs1233578). The 12q13.3 lead SNP (rs167769), which was significant only in the multiancestry analysis ($P_{random} = 3.9\times10^{-9}$), was located within an intron of *STAT6* and is strongly associated with *STAT6* expression in the blood[10,11,13] and lungs[14]. STAT6 is a transcription factor that is essential for the T_H2-lymphocyte functional responses mediated by IL-4 and IL-13 (ref. [29]). This result established the association of *STAT6* with asthma risk that has been disputed in candidate-gene studies[30]. The 5q31.3 lead SNP rs7705042 ($P_{random} = 7.9\times10^{-9}$ in multiancestry analysis) was

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

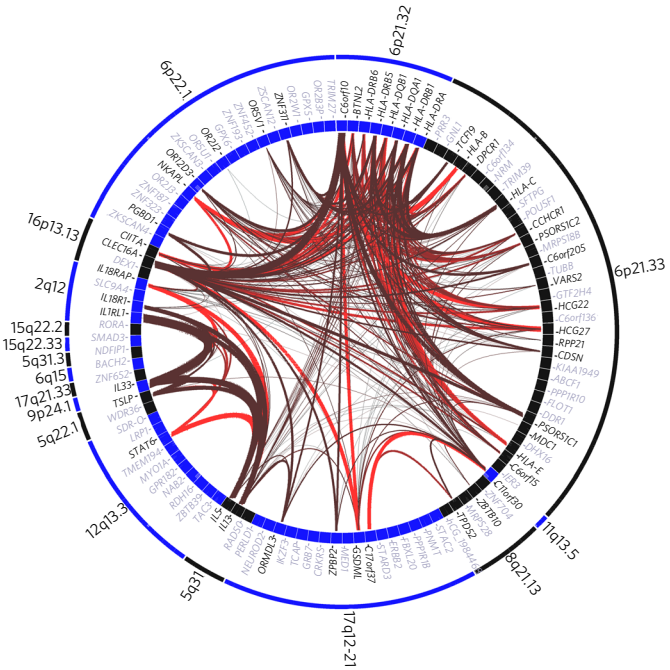

Fig. 2 | GRAIL circle plot of connectivity among genes at asthma risk loci. The 17 asthma risk loci are along the outer ring (the 10p14 locus was ignored because it corresponds to a gene desert); the internal ring represents the genes at these loci. The widths of the lines drawn between genes correspond to the strength of the literature-based connectivity, with thicker lines representing stronger connections.

located within an intron of *NDFIP1* and is associated with *NDFIP1* expression in the blood[11–13]. NDFIP1 is a potent inhibitor of the antiviral response[31] and inflammation processes[32].

New asthma signals at loci reported in specific populations. Two associations in our study were with new SNPs at loci previously reported to be associated with asthma in people of Latino[6] and Japanese[7] ancestry. The first one, at 6p21.33, has previously been reported in an admixture mapping study in Latino individuals[6]. The lead TAGC SNP rs2855812 ($P_{random} = 8.9 \times 10^{-12}$ in the multiancestry analysis; $P_{random} = 1.7 \times 10^{-8}$ in the European-ancestry meta-analysis) was located within an intron of *MICB*. This SNP was not correlated ($r^2 = 0$) with any of the SNPs reported in the study of Latino individuals[6]. The 6p21.33 region contains many genes whose transcripts are associated with TAGC asthma signals, including *TNF*, *LST1*, *HLA-C*, and *LTA* in the blood[10,11,13], and *MICB* in the lungs[12,14]. These genes are involved in immunologically related mechanisms. This 6p21.33 locus is approximately 600 kb from the previously reported 6p21.32 locus that spans HLA class II genes. Intensive sequencing efforts will be needed to further clarify the HLA-region associations. The second association was at the 10p14 locus, where a GWAS in Japanese individuals[7] has reported an association (lead SNP rs10508372) with adult asthma. We detected a new signal, rs2589561, in European-ancestry ($P_{random} = 1.4 \times 10^{-8}$) and multiancestry meta-analyses ($P_{random} = 3.5 \times 10^{-9}$) that was not correlated with rs10508372 in either European-ancestry or Japanese-ancestry populations. The SNP rs2589561 is in a gene desert, 929 kb from *GATA3*. However, recently published promoter-capture Hi-C data in hematopoietic cells[33] has shown that two proxies of rs2589561 ($r^2 > 0.9$) are located in a region that interacts with the *GATA3* promoter, especially in CD4$^+$ T cells. These findings suggest that the SNP may be in a distal regulator of *GATA3*, which encodes a transcription factor that is a master regulator of differentiation of T_H2 cells and type 2 innate lymphoid cells (ref. [34]).

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[8]. In our results, the 8q21.13 lead SNP rs12543811 ($P_{random} = 3.4 \times 10^{-8}$ and 1.1×10^{-10} in the European-ancestry and multiancestry analyses) was located between *TPD52* and *ZBTB10* 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[35]. The chromosome 16p13.13 SNP rs17806299 is within an intron of *CLEC16A* ($P_{random} = 2.1 \times 10^{-10}$ and 2.7×10^{-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)[8], the association of asthma with 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[11,23]. 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[36]. The potential relevance of *DEXI* in allergic diseases has also been previously discussed[8].

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[3]. 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 $P_{random} < 10^{-4}$ in this study and GWAS signals with all phenotypes and diseases in the GWAS catalog[3]. We examined 4,231 unique trait–loci combinations (Methods) and used the disease classification from Wang et al.[37] 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[38,39]; 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 $P_{random} \leq 10^{-3}$ 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

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

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 and their proxies ($r^2 \geq 0.80$) might be concentrated in cis-regulatory DNA elements. We explored only 16 of 18 identified asthma loci, excluding the two loci spanning the HLA region because of the region's high variability and extensive LD. We interrogated the 111 Roadmap and 16 ENCODE reference epigenomes in a wide range of human cell types[24], focusing on histone marks characterizing enhancers and promoters assayed in all 127 epigenomes and DNase I–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[40]. 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 DNase I–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[41].

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[25]. 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[21,22] 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-ancestry 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[42], 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.internationalgenome.org/; Genome-wide Complex Trait Analysis (GCTA), http://cnsgenomics.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,

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

https://www.hsph.harvard.edu/liming-liang/software/eqtl/; GHS-Express, http://genecanvas.ecgene.net/; HaploReg v4.1, http://archive.broadinstitute.org/mammals/haploreg/haploreg.php/; Roadmap and ENDCODE epigenomics data, http://egg2.wustl.edu/roadmap/web_portal/; UES pipeline, https://github.com/JamesHayes/uesEnrichment/; GRAIL, https://software.broadinstitute.org/mpg/grail/; Visualizing GRAIL connections (VIZ-GRAIL), http://software.broadinstitute.org/mpg/grail/vizgrail.html; LocusZoom, http://locuszoom.org/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0014-7.

Received: 20 October 2016; Accepted: 8 November 2017;
Published online: 22 December 2017

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. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
4. Igartua, C. et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. *Nat. Commun.* **6**, 5965 (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 (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. 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**, 1539–1558 (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 ryanodine 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. Torgerson, D. G. et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat. Genet.* **43**, 887–892 (2011).
20. Bønnelykke, 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).
21. Ferreira, M. A. et al. Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006–1014 (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**, e1000534 (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 β1γ2 and promotes β1γ2-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, X. Ndfip1 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 e19 (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**, 322–333 (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. Liang, 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.F., M.C.M., C.F.R., and P.J.T.; data collection or analysis, M.A.F., M.C.M., C.F.R., G.J., and P.J.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.

Saguenay-Lac-Saint-Jean (SLSJ) Study: study PIs, C.L. 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.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

Northern Finland Birth Cohort of 1966 (NFBC1966): study PI, M.-R.J.; data collection or analysis, A.C.A. and A.R. **FINRISK**: 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., L.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 L.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, H.B. **Cophenhagen 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.J.; data collection or analysis, R.J.P.v.d.V., L.D., and V.W.V.J. **GINIplus/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, J.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.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. **Kursk State Medical University (KSMU) Study**: study PI, A. Polonikov; data collection or analysis, M. Solodilova and V.I.; **Medical Research Council-funded Collection of Nuclear Families with Asthma (MRCA-UKC)**: study PIs, W.O.C.C. and M.M.; data collection or analysis, L. Liang. **Multicentre Asthma Genetics in Childhood Study (MAGICS)**: study PI, M. Kabesch; data collection or analysis, A.V.B. and S.M. **German Multicentre Allergy Study (MAS)**: study PI, Y.-A.L.; data collection or analysis, S. Lau and I.M. **Prevention and Incidence of Asthma and Mite Allergy (PIAMA) cohort** : study PIs, G.H.K. and D.S.P.; data collection or analysis, G.H.K., D.S.P., and U.G. **Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults (SAPALDIA)**: study PI, N.P.-H.; data collection or analysis, M.I. and A.K. **Tomsk Study**: study PIs, L.M.O. and V.P.P.; data collection or analysis, M.B.F. and P.A.S. **UFA Study**: study PI, E. Khusnutdinova; data collection or analysis, A.S.K. and Y.F. **Industrial Cohorts Research Group (INDUSTRIAL)**: study PIs, D.H. and T.S.; data collection or analysis, I.M.W. and V. Schlünssen. **Severe Asthma Cohorts (SEVERE)**: study PIs, A.B., K.F.C., and C.E.B.

Netherlands Twin Register (NTR) Study: study PI, D.I.B.; genotyping, data collection or analysis, J.J.H., H.M., and G.W.

Rotterdam Study: study PIs, A.H., B.H.S., and G.G.B.; genotyping, data collection or analysis, G.G.B., B.H.S., D.W.L., L. Lahousse, and A.G.U.

Dutch Asthma Genetics Consortium (DAGC): study PIs, G.H.K. and D.S.P.; genotyping, data collection or analysis, G.H.K., D.S.P., J.A., M.A.E.N., and J.M.V.

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.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

Florence Demenais[1,2*], Patricia Margaritte-Jeannin[1,2], Kathleen C. Barnes[3], William O. C. Cookson[4], Janine Altmüller[5], Wei Ang[6], R. Graham Barr[7], Terri H. Beaty[8], Allan B. Becker[9], John Beilby[10], Hans Bisgaard[11], Unnur Steina Bjornsdottir[12], Eugene Bleecker[13], Klaus Bønnelykke[11], Dorret I. Boomsma[14], Emmanuelle Bouzigon[1,2], Christopher E. Brightling[15], Myriam Brossard[1,2], Guy G. Brusselle[16,17,18], Esteban Burchard[19], Kristin M. Burkart[20], Andrew Bush[21,22], Moira Chan-Yeung[23], Kian Fan Chung[21,24], Alexessander Couto Alves[25], John A. Curtin[26], Adnan Custovic[27], Denise Daley[23,28], Johan C. de Jongste[29], Blanca E. Del-Rio-Navarro[30], Kathleen M. Donohue[7], Liesbeth Duijts[31], Celeste Eng[32], Johan G. Eriksson[33], Martin Farrall[34,35], Yuliya Fedorova[36], Bjarke Feenstra[37], Manuel A. Ferreira[38], Australian Asthma Genetics Consortium (AAGC) collaborators[39], Maxim B. Freidin[40], Zofia Gajdos[41,42], Jim Gauderman[43], Ulrike Gehring[44], Frank Geller[37], Jon Genuneit[45], Sina A. Gharib[46], Frank Gilliland[43], Raquel Granell[47,48], Penelope E. Graves[49], Daniel F. Gudbjartsson[50,51], Tari Haahtela[52], Susan R. Heckbert[53], Dick Heederik[44], Joachim Heinrich[54,55], Markku Heliövaara[56], John Henderson[47,48], Blanca E. Himes[57], Hiroshi Hirose[58], Joel N. Hirschhorn[42,59,60], Albert Hofman[17,61], Patrick Holt[62], Jouke Hottenga[14], Thomas J. Hudson[63,64], Jennie Hui[10,65,66], Medea Imboden[67,68], Vladimir Ivanov[69], Vincent W. V. Jaddoe[70], Alan James[71,72], Christer Janson[73], Marjo-Riitta Jarvelin[74,75,76,77], Deborah Jarvis[21,78], Graham Jones[79], Ingileif Jonsdottir[50,80], Pekka Jousilahti[56], Michael Kabesch[81], Mika Kähönen[82], David B. Kantor[83,84], Alexandra S. Karunas[36,85], Elza Khusnutdinova[36,85], Gerard H. Koppelman[86,87], Anita L. Kozyrskyj[88], Eskil Kreiner[11], Michiaki Kubo[89], Rajesh Kumar[90,91], Ashish Kumar[67,68,92], Mikko Kuokkanen[56,93], Lies Lahousse[17,94], Tarja Laitinen[95], Catherine Laprise[96,97], Mark Lathrop[98], Susanne Lau[99], Young-Ae Lee[100,101], Terho Lehtimäki[102], Sébastien Letort[1,2], Albert M. Levin[103], Guo Li[46], Liming Liang[61,104], Laura R. Loehr[105], Stephanie J. London[106], Daan W. Loth[17], Ani Manichaikul[107], Ingo Marenholz[100,101], Fernando J. Martinez[49], Melanie C. Matheson[108], Rasika A. Mathias[109], Kenji Matsumoto[110], Hamdi Mbarek[14], Wendy L. McArdle[111], Mads Melbye[37,112,113], Erik Melén[92,114,115], Deborah Meyers[13], Sven Michel[81], Hamida Mohamdi[1,2], Arthur W. Musk[116,117], Rachel A. Myers[118], Maartje A. E. Nieuwenhuis[87,119], Emiko Noguchi[120], George T. O'Connor[121,122], Ludmila M. Ogorodova[123], Cameron D. Palmer[42,59], Aarno Palotie[93,124,125], Julie E. Park[23], Craig E. Pennell[6], Göran Pershagen[92,114], Alexey Polonikov[69], Dirkje S. Postma[87,119], Nicole Probst-Hensch[67,68], Valery P. Puzyrev[40], Benjamin A. Raby[126], Olli T. Raitakari[127], Adaikalavan Ramasamy[25,128], Stephen S. Rich[107], Colin F. Robertson[129], Isabelle Romieu[130,131], Muhammad T. Salam[43,132], Veikko Salomaa[56], Vivi Schlünssen[133], Robert Scott[134], Polina A. Selivanova[135], Torben Sigsgaard[133], Angela Simpson[26,136], Valérie Siroux[137,138], Lewis J. Smith[139], Maria Solodilova[69], Marie Standl[55], Kari Stefansson[50,80], David P. Strachan[140], Bruno H. Stricker[17,141,142], Atsushi Takahashi[89], Philip J. Thompson[143], Gudmar Thorleifsson[50], Unnur Thorsteinsdottir[50,80], Carla M. T. Tiesler[55,144], Dara G. Torgerson[32], Tatsuhiko Tsunoda[89,145], André G. Uitterlinden[142], Ralf J. P. van der Valk[146], Amaury Vaysse[1,2], Sailaja Vedantam[41,42], Andrea von Berg[147], Erika von Mutius[148,149], Judith M. Vonk[87,150], Johannes Waage[11], Nick J. Wareham[134], Scott T. Weiss[126], Wendy B. White[151], Magnus Wickman[92,152], Elisabeth Widén[93], Gonneke Willemsen[14], L. Keoki Williams[153,154], Inge M. Wouters[44], James J. Yang[155], Jing Hua Zhao[134], Miriam F. Moffatt[4], Carole Ober[156] and Dan L. Nicolae[157*]

[1]Genetic Variation and Human Diseases Unit (UMR-946), Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France. [2]Institut Universitaire d'Hématologie, Université Paris Diderot, Université Sorbonne Paris Cité, Paris, France. [3]Division of Biomedical Informatics and Personalized Medicine, Colorado Center for Personalized Medicine, University of Colorado, Denver, CO, USA. [4]Section of Genomic Medicine, National Heart and Lung Institute, London, UK. [5]Cologne Center for Genomics and Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany. [6]School of Women's and Infants' Health, University of Western Australia, Perth, Western Australia, Australia. [7]Departments of Medicine and Epidemiology, Columbia University, New York, NY, USA. [8]Division of Genetic Epidemiology, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [9]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

[10]Department of the Diagnostic Genomics Laboratory, PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia. [11]Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. [12]Department of Medicine, Landspitali, National University Hospital of Iceland, Reykjavik, Iceland. [13]Center for Genomics, Wake Forest University School of Medicine, Winston-Salem, NC, USA. [14]Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrjie Universiteit, Amsterdam, The Netherlands. [15]University of Leicester, Institute for Lung Health, Glenfield Hospital, Leicester, UK. [16]Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium. [17]Department of Epidemiology, Erasmus MC–University Medical Center Rotterdam, Rotterdam, The Netherlands. [18]Department of Respiratory Medicine, Erasmus MC–University Medical Center Rotterdam, Rotterdam, The Netherlands. [19]Department of Bioengineering & Therapeutic Sciences and Medicine, University of California, San Francisco, San Francisco, CA, USA. [20]Division of Pulmonary, Allergy and Critical Care, College of Physicians and Surgeons, Columbia University, New York, NY, USA. [21]National Heart and Lung Institute, Imperial College London, London, UK. [22]Royal Brompton Harefield National Health Service (NHS) Foundation Trust, London, UK. [23]Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada. [24]Biomedical Research Unit, Royal Brompton & Harefield National Health Service (NHS) Trust, London, UK. [25]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [26]Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK. [27]Department of Paediatrics, Imperial College London, London, UK. [28]Centre for Heart and Lung Innovation, University of British Columbia, Vancouver, BC, Canada. [29]Department of Pediatrics, Division of Respiratory Medicine, Erasmus MC–University Medical Center Rotterdam, Rotterdam, The Netherlands. [30]Hospital Infantil de Mexico Federico Gomez, Mexico City, Mexico. [31]Department of Pediatrics, Division of Respiratory Medicine, and Department of Pediatrics, Division of Neonatology, Erasmus MC–University Medical Center Rotterdam, Rotterdam, The Netherlands. [32]Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. [33]Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [34]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [35]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [36]Institute of Biochemistry and Genetics, Ufa Scientific Center of the Russian Academy of Sciences, Ufa, Russian Federation. [37]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [38]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [39]A full list of members and affiliations appears in the Supplementary Note. [40]Population Genetics Laboratory, Research Institute of Medical Genetics, Tomsk NRMC, Tomsk, Russian Federation. [41]Divisions of Genetics and Endocrinology, Children's Hospital, Boston, MA, USA. [42]Broad Institute, Cambridge, MA, USA. [43]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [44]Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands. [45]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany. [46]Department of Medicine, University of Washington, Seattle, WA, USA. [47]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [48]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [49]Asthma and Airway Disease Research Center and BIO5 Institute, University of Arizona, Tucson, AZ, USA. [50]deCODE genetics, Amgen Inc., Reykjavik, Iceland. [51]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [52]Skin and Allergy Hospital, University of Helsinki, Helsinki, Finland. [53]Department of Epidemiology, University of Washington, Seattle, WA, USA. [54]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, University Hospital in Munich, Munich, Germany. [55]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany. [56]National Institute for Health and Welfare (THL), Helsinki, Finland. [57]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA. [58]Health Center, Department of Internal Medicine, Keio University, Tokyo, Japan. [59]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. [60]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, USA. [61]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [62]Cell Biology Telethon Kids Institute, University of Western Australia, Subiaco, Western Australia, Australia. [63]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [64]AbbVie Inc., Redwood City, CA, USA. [65]Busselton Population Medical Research Institute, Perth, Western Australia, Australia. [66]School of Population and Global Health, University of Western Australia, Nedlands, Western Australia, Australia. [67]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland. [68]University of Basel, Basel, Switzerland. [69]Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russian Federation. [70]The Generation R Study Group, Department of Pediatrics and Department of Epidemiology, Erasmus MC–University Medical Center Rotterdam, Rotterdam, The Netherlands. [71]Department of Pulmonary Physiology and Sleep Medicine, Busselton Population Medical Research Institute, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. [72]School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia. [73]Department of Medical Sciences: Respiratory, Allergy & Sleep Research, Uppsala University, Uppsala, Sweden. [74]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK. [75]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. [76]Biocenter Oulu, University of Oulu, Oulu, Finland. [77]Unit of Primary Care, Oulu University Hospital, Oulu, Finland. [78]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [79]School of Science and Health, Western Sydney University, Sydney, New South Wales, Australia. [80]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [81]Department of Pediatric Pneumology and Allergy, University Children's Hospital Regensburg (KUNO), Regensburg, Germany. [82]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland. [83]Department of Anesthesiology, Perioperative and Pain Medicine, Division of Critical Care Medicine, Boston Children's Hospital, Boston, MA, USA. [84]Department of Anaesthesia, Harvard Medical School, Boston, MA, USA. [85]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russian Federation. [86]University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, The Netherlands. [87]Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, The Netherlands. [88]Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada. [89]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [90]Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA. [91]Divison of Allergy and Clinical Immunology, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [92]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [93]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [94]Pharmaceutical Care Unit, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium. [95]Department of Pulmonary Medicine, University of Turku and Turku University Hospital, Turku, Finland. [96]Département des Sciences Fondamentales, Université du Québec à Chicoutimi, Chicoutimi, Quebec, Canada. [97]Centre de Santé et de Services Sociaux du Saguenay-Lac-Saint-Jean, Saguenay, QC, Canada. [98]McGill University and Genome Quebec Innovation Centre, Montréal, QC, Canada. [99]Pediatric Pneumology and Immunology, Charité Universitätsmedizin, Berlin, Germany. [100]Max-Delbrück-Centrum (MDC) for Molecular Medicine, Berlin, Germany. [101]Pediatric Allergology, Experimental and Clinical Research Center, Charité Universitätsmedizin, Berlin, Germany. [102]Department of Clinical Chemistry, Fimlab Laboratories, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [103]Department of Public Health Sciences, Henry Ford Health System, Detroit, MI, USA. [104]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [105]Division of General Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [106]National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park,

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

NC, USA. [107]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [108]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia. [109]Division of Allergy & Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA. [110]Department of Allergy and Clinical Immunology, National Research Institute for Child Health and Development, Tokyo, Japan. [111]Bristol Bioresource Laboratories, School of Social and Community Medicine, University of Bristol, Bristol, UK. [112]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [113]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [114]Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden. [115]Sachs Children's Hospital, Stockholm, Sweden. [116]Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia. [117]Schools of Population Health and of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia. [118]Center for Applied Genomics and Precision Medicine, Duke University School of Medicine, Durham, NC, USA. [119]University Medical Center Groningen, Department of Pulmonology, University of Groningen, Groningen, The Netherlands. [120]Department of Medical Genetics, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. [121]Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [122]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA. [123]Department of Faculty Pediatrics, Siberian State Medical University, Tomsk, Russian Federation. [124]Analytic and Translational Genetics Unit, Departments of Medicine, of Neurology and of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [125]The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [126]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [127]Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland. [128]Department of Medical and Molecular Genetics, King's College London, London, UK. [129]Respiratory Medicine, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [130]Hubert Department of Global Health, Mory University, Atlanta, GA, USA. [131]Center for Population Health Research, National Institute of Public Health, Cuernavaca, Mexico. [132]Department of Psychiatry, Kern Medical, Bakersfield, CA, USA. [133]Department of Public Health, Section for Environment, Occupation & Health, Aarhus University, Aarhus, Denmark. [134]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK. [135]Department of Faculty Therapy, Siberian State Medical University, Tomsk, Russian Federation. [136]University Hospital of South Manchester, National Health Service (NHS) Foundation Trust, Manchester, UK. [137]Institut National de la Santé et de la Recherche Médicale (INSERM) U1209, Institute for Advanced Biosciences, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health, Grenoble, France. [138]Université de Grenoble Alpes/CNRS UMR5309, Institute for Advanced Biosciences, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health, Grenoble, France. [139]Division of Pulmonary and Critical Care Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [140]Population Health Research Institute, St George's University of London, London, UK. [141]Netherlands Healthcare Inspectorate, The Hague, The Netherlands. [142]Department of Internal Medicine, Erasmus MC–University Medical Center Rotterdam, Rotterdam, The Netherlands. [143]Institute for Respiratory Health and Harry Perkins Institute of Medical Research, University of Western Australia and The Lung Health Clinic, Nedlands, Western Australia, Australia. [144]Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany. [145]Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan. [146]The Generation R Study Group, Department of Pediatrics, Division of Respiratory Medicine and Department of Epidemiology, Erasmus MC–University Medical Center Rotterdam, Rotterdam, The Netherlands. [147]Department of Pediatrics, Marien-Hospital Wesel, Wesel, Germany. [148]Dr. Von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany. [149]German Center for Lung Research, Munich, Germany. [150]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands. [151]Undergraduate Training and Education Center (UTEC), Jackson Heart Study, Tougaloo College, Jackson, MI, USA. [152]Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden. [153]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI, USA. [154]Department of Internal Medicine, Henry Ford Health System, Detroit, MI, USA. [155]School of Nursing, University of Michigan, Ann Arbor, MI, USA. [156]Department of Human Genetics, University of Chicago, Chicago, IL, USA. [157]Departments of Statistics, Human Genetics and Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL, USA. Florence Demenais, Patricia Margaritte-Jeannin, Kathleen C. Barnes, William O. C. Cookson, Miriam F. Moffatt, Carole Ober and Dan Nicolae contributed equally to this work. *e-mail: florence.demenais@inserm.fr; nicolae@galton.uchicago.edu

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

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 56 studies of individuals of European ancestry (19,954 asthma cases, 107,715 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, as detailed 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 data 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 b135 (31,587 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 for at least one SNP in at least one study were available for 2.83 million autosomal SNPs. Strict QC criteria were used for inclusion of a SNP in the analysis. When a SNP genotype was used in the study analysis, these criteria were call rate ≥99%, *P* value for HW test $\geq 10^{-6}$, and MAF ≥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 MAF ≥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 or ≤−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 Der Simonian and Laird[43] estimator of the between-study variance, when the meta-analyses included a large number of studies (European-ancestry, multiancestry 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 contributed to a meta-analysis. Tests of significance of the combined effect sizes were performed by using a standard normal distribution. We applied a threshold of P_{random} (or P_{fixed}) of 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 multiancestry 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 P_{random} (or P_{fixed}) $\leq10^{-6}$. The heterogeneity of per-SNP effect sizes across all studies in a meta-analysis was assessed with Cochran's *Q* test[9]. Differences among the four ethnic-specific summary effects were also tested with Cochran's *Q* statistic.

Conditional analysis of asthma-associated loci. GCTA software[44] (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 included in the meta-analysis. Approximate conditional analysis was performed in only the European-ancestry group, which could be assumed to share a similar LD pattern and was both the largest ancestry-specific dataset and the only one showing genome-wide-significant results. Because this analysis assumes no heterogeneity in SNP effect size across studies, the 9p24.1 and 17q12-21 loci, which showed significant heterogeneity ($P_{het}\leq0.05$, Cochran's *Q* test) for a large portion of each locus, 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 after QC[22]. Imputation was performed with MACH software[45] and the HapMap2, release 21 panel; only well-imputed SNPs (imputation quality score *rsq* >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 1.01 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 ($P=4.1\times10^{-6}$), after adjustment for the lead SNP, we performed an additional round including both SNPs. If the remaining SNPs had $P>4.1\times10^{-6}$, no further analysis was performed. The results of this analysis are reported in Supplementary Table 15.

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 defined a list of SNPs to be interrogated that included the lead SNPs, the secondary signals identified by conditional analysis, and all SNPs in LD with these SNPs (r^2 between 0.5 and 1). To search for cis eQTLs within up to 1 Mb of 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[11]); (ii) gene expression data from 777 lymphoblastoid cell lines from the MuTHER database[10]; (iii) transcriptional profiles of 405 and 550 lymphoblastoid cell lines from UK asthma (MRCA) and eczema (MRCE) family members, respectively[13]; (iv) eQTL data from monocytes from 1,490 individuals included in the GHS-express database[23]; (v) GTEx eQTL Browser data from multiple tissues including the blood and lungs[12]; and (vi) transcriptional profiles from the lung tissues of 1,111 subjects[14] (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 were in LD (r^2 >0.5) with missense variants by using the HaploReg v4.1 tool (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[3] (URLs). We then used this catalog to systematically investigate the overlap of asthma signals with $P_{random}\leq10^{-4}$ in the multiancestry 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 $>10^{-7}$ (with the intent of removing some of the potential false positives in the catalog) and SNPs that were duplicated (i.e., to remove disease-SNP duplications). This procedure decreased the number of entries to 5,927. Notably, this process did not remove either SNPs in perfect LD associated with the same disease or SNPs that were present multiple times in the database because of their association with different phenotypes.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

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.[37]. We summarized the overlap of GWAS-catalog signals with asthma signals according to the proportion of catalog SNPs with asthma P values $<10^{-4}$ in our analysis. The significance of overlap was estimated as the binomial-tail probability for observing the number of TAGC SNPs with $P_{random} \leq 10^{-4}$ 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 $P_{random} \leq 10^{-3}$ 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 ($r^2 \geq 0.80$) were concentrated in cis-regulatory DNA elements. We used the UES pipeline[40] (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 intersectBed[46], 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 r_{loci}/n, where r_{loci} 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 types[24] (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 GRAIL[25] to assess the relatedness among asthma-associated loci. As previously described in detail[25], to define the genes near each SNP, GRAIL finds the furthest neighboring SNPs in the 3′ and 5′ direction that are in LD ($r^2 > 0.5$) 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 described[25]. The interconnectivity among genes at asthma risk loci was visualized using VIZGRAIL[47] (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 model[48] 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 GCTA[44]. 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).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Florence Demenais

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

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 assesed 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.

Software

Policy information about availability of computer code

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 sofwtare 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.

Materials and reagents

Policy information about availability of materials

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.