LETTERS
https://doi.org/10.1038/s41588-018-0157-1

nature genetics

Corrected: Author Correction

Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis

Allergic rhinitis is the most common clinical presentation of allergy, affecting 400 million people worldwide, with increasing incidence in westernized countries[1,2]. To elucidate the genetic architecture and understand the underlying disease mechanisms, we carried out a meta-analysis of allergic rhinitis in 59,762 cases and 152,358 controls of European ancestry and identified a total of 41 risk loci for allergic rhinitis, including 20 loci not previously associated with allergic rhinitis, which were confirmed in a replication phase of 60,720 cases and 618,527 controls. Functional annotation implicated genes involved in various immune pathways, and fine mapping of the HLA region suggested amino acid variants important for antigen binding. We further performed genome-wide association study (GWAS) analyses of allergic sensitization against inhalant allergens and nonallergic rhinitis, which suggested shared genetic mechanisms across rhinitis-related traits. Future studies of the identified loci and genes might identify novel targets for treatment and prevention of allergic rhinitis.

Allergic rhinitis, an inflammatory disorder of the nasal mucosa, is mediated by allergic hypersensitivity responses to environmental allergens[1]. It has substantial adverse effects on quality of life and health-care expenditures. The underlying causes of allergic rhinitis are still not understood, and prevention of the disease is not possible. The heritability of allergic rhinitis has been estimated to be more than 65% (refs [3,4]). Seven loci have been associated with allergic rhinitis in GWAS on allergic rhinitis, and other loci have been suggested from GWAS on related traits such as self-reported allergy, asthma plus hay fever, or allergic sensitization[5–9]; however only a few of these findings have been replicated.

We carried out a large-scale meta-GWAS on allergic rhinitis including a discovery meta-analysis of 16,531,985 genetic markers from 18 studies comprising 59,762 cases and 152,358 controls of primarily European ancestry (Supplementary Table 1 and cohort-recruitment details in Supplementary Note). We report the genetic heritability on the liability scale of allergic rhinitis to be at least 7.8% (assuming 10% disease prevalence), with a genomic inflation of 1.048 (Supplementary Fig. 1). We identified 42 genetic loci with index markers below genome-wide significance ($P < 5 \times 10^{-8}$), of which 21 have previously been reported in relation to allergic rhinitis or other inhalant allergy[6–9] (Fig. 1, Tables 1 and 2, and Supplementary Figs. 2 and 3).

One study (23andMe) had a proportionally large weight (~80%) in the discovery phase. Overall, there was good agreement between 23andMe and the other studies with respect to effect size and direction and regional association patterns (Supplementary Table 2 and Supplementary Figs. 4 and 5), and the genetic correlation was 0.80 ($P < 2 \times 10^{-17}$). The heterogeneity between 23andMe and the remaining studies was statistically significant ($P < 0.05$) for 7 of 42 loci, in most cases because of a smaller effect size in 23andMe. This finding was probably due to many non-23andMe studies using a more robust phenotype definition of doctor-diagnosed allergic rhinitis (Supplementary Table 3), which tended to result in larger effect sizes (Supplementary Table 4).

The index markers from a total of 25 loci that had not previously been associated with allergic rhinitis or other inhalant allergy were carried forward to the replication phase. These markers included 16 loci that showed genome-wide-significant association in the discovery phase and evidence of association ($P < 0.05$) in both 23andMe and non-23andMe studies (Supplementary Table 2), and an additional nine loci that were selected from the *P*-value stratum between 5×10^{-8} and 1×10^{-6}, on the basis of enrichment of gene sets involved in immunological signaling (Supplementary Table 5). We sought replication in another ten studies with 60,720 cases and 618,527 controls. Of the 25 loci, 20 reached a Bonferroni-corrected significance threshold of 0.05 ($P < 0.0019$) in a meta-analysis of replication studies (Fig. 1 (blue) and Table 2) and also reached genome-wide significance in the combined fixed-effect meta-analysis of discovery and replication studies (Table 2). Evidence of heterogeneity was seen for one of these loci (rs1504215), which did not reach statistical significance in the random-effects model (0.95 (0.92; 0.97), $P = 2.83 \times 10^{-7}$; Supplementary Fig. 3).

A conditional analysis of top loci identified 13 additional independent variants at $P < 1 \times 10^{-5}$, four of which were genome-wide significant (near *WDR36*, *HLA-DQB1*, *IL1RL1*, and *LPP*; Supplementary Table 6 and Supplementary Fig. 5).

To gain insight into the functional consequences of known and novel loci, we used a number of data sources, including: (i) 11 expression quantitative trait locus (eQTL) sets and one methylation quantitative trait locus (meQTL) set from blood and blood subsets; (ii) two eQTL sets and one meQTL set from lung tissue; and (iii) data on enhancer–promoter interactions in 15 different blood subsets. We found support for regulatory effects on coding genes for 33 of the 41 loci. Many loci showed evidence of regulatory effects across a wide range of immune-cell types (including B and T cells), whereas other loci seemed to be cell-type specific (Supplementary Table 7). Calculating the 'credible set' of markers for each locus through a Bayesian approach that selected markers likely to contain the causal disease-associated markers (Supplementary Table 8) and then looking up those markers in the Variant Effect Predictor database generated a list of 17 markers producing amino acid changes, including deleterious changes in NUSAP1, SULT1A1 and PLCL, as predicted by SIFT (Supplementary Table 9).

The major histocompatibility complex (MHC) on chromosome 6p contained some of the strongest association signals in the GWAS, including independent signals located around *HLA-DQB* and *HLA-B*. The top variant at *HLA-DQB* was an eQTL for several HLA genes, including *HLA-DQB1*, *HLA-DQA1*, *HLA-DQA2*, and *HLA-DRB1* in immune and/or lung tissue, and the top variant at *HLA-B* was

A full list of authors and affiliations appears at the end of the paper.

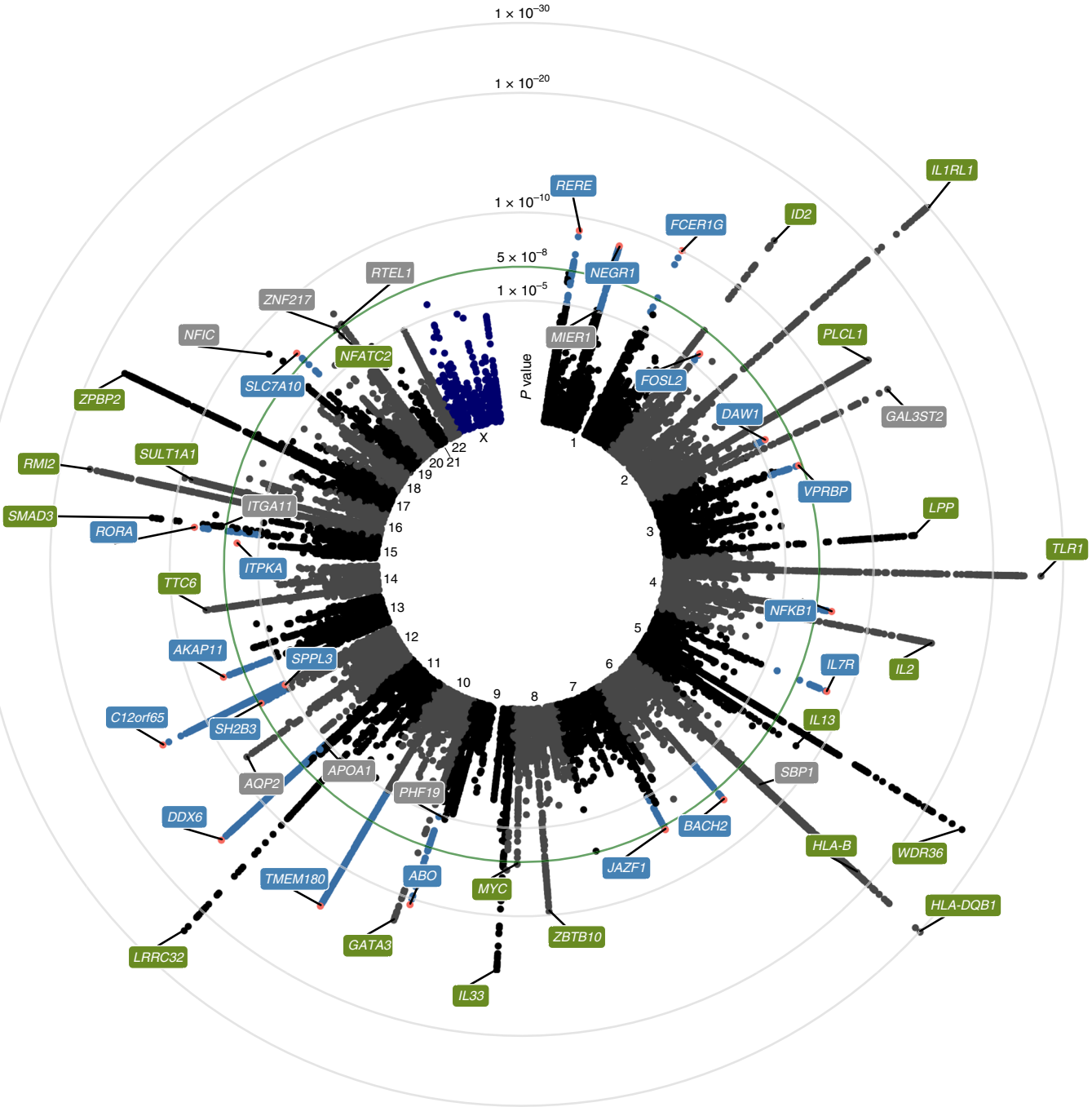

Fig. 1 | Manhattan plot of the meta-GWAS discovery phase. Circular plot of *P* values, calculated from the two-tailed *z* score from an inverse-variance-weighted fixed-effect meta-analysis of association of 16,531,985 genetic markers with allergic rhinitis from the discovery phase, including 212,120 individuals. Only markers with $P<1\times10^{-3}$ are shown. Labels indicate the nearest gene name for the index marker in the locus (the marker with the lowest *P* value). Green labels indicate loci previously associated with allergy; blue labels indicate novel allergic rhinitis–associated loci; gray labels indicate novel loci that were not carried forward to the replication phase. Green line indicates the level of genome-wide significance ($P=5\times10^{-8}$).

an eQTL for *MICA* (Supplementary Table 7). In addition, we found associations with several classical HLA alleles, including *HLA-DQB1*02:02*, *HLA-DQB1*03:01*, *HLA-DRB1*04:01*, and *HLA-C*04:01*, which were in weak linkage disequilibrium (LD) ($r^2<0.1$) with the GWAS top SNPs (Supplementary Tables 10 and 11), and strong associations with well-imputed amino acid variants, including HLA-DQB1 His30 ($P=2.06\times10^{-28}$, odds ratio (OR) = 0.91) and HLA-B Asp116/His116-/Leu116 ($P=6.00\times10^{-13}$, OR = 1.06) (Supplementary Tables 12 and 13). Within HLA-DQB1, the amino acid variant was in moderate LD ($r^2=0.71$) with the GWAS top SNP and accounted for most of the SNP association (rs34004019, $P=2.18\times10^{-28}$, OR = 0.88, conditional $P=1.35\times10^{-3}$). Within *HLA-B*, the strongest associated amino acid variant was only in weak LD ($r^2=0.23$) with the top SNP and accounted for a small part of the SNP association (rs2428494, $P=3.99\times10^{-15}$, OR = 1.07, conditional $P=3.23\times10^{-10}$). Importantly, the strongest associated amino acid variants in HLA-DQB1 and HLA-B, respectively, were both located in the peptide-binding pockets with a high likelihood of affecting MHC–peptide interaction (Fig. 2). MHC class II molecules, including HLA-DQ, are known for their roles in allergen binding and type

Table 1 | Association results of index markers (variant with lowest *P* value for each locus) previously reported in relation to allergic rhinitis or other inhalant allergy

			Discovery						
Variant	Locus	Nearest genes	EA/OA	EAF	*n* (studies)	OR	95% CI	*P*	Het. *P*
Known									
rs34004019	6p21.32	*HLA-DQB1, HLA-DQA1*	G/A	0.27	196,951 (11)	0.89	0.87–0.90	1.00×10^{-30}	0.41
rs950881	2q12.1	*IL1RL1, IL1RL1*	T/G	0.15	212,120 (18)	0.88	0.87–0.90	1.74×10^{-30}	0.91
rs5743618	4p14	*TLR1, TLR10*	A/C	0.27	210,652 (17)	0.90	0.89–0.92	4.38×10^{-27}	0.70
rs1438673	5q22.1	*CAMK4, WDR36*	C/T	0.50	212,120 (18)	1.08	1.07–1.10	3.15×10^{-26}	0.26
rs7936323	11q13.5	*LRRC32, C11orf30*	A/G	0.48	212,120 (18)	1.08	1.06–1.09	6.53×10^{-24}	0.0001
rs2428494	6p21.33	*HLA-B, HLA-C*	A/T	0.42	195,753 (12)	1.08	1.06–1.09	7.01×10^{-19}	0.25
rs11644510	16p13.13	*RMI2, CLEC16A*	T/C	0.37	212,120 (18)	0.93	0.92–0.95	1.58×10^{-17}	0.65
rs12939457	17q12	*GSDMB, ZPBP2*	C/T	0.44	212,120 (18)	0.94	0.92–0.95	2.35×10^{-17}	0.02
rs148505069	4q27	*IL21, IL2*	G/A	0.33	212,120 (18)	1.07	1.05–1.08	2.54×10^{-15}	0.02
rs13395467	2p25.1	*ID2, RNF144A*	G/A	0.28	212,120 (18)	0.94	0.92–0.95	9.93×10^{-15}	0.61
rs9775039	9p24.1	*IL33, RANBP6*	A/G	0.16	212,120 (18)	1.08	1.06–1.10	2.22×10^{-14}	0.40
rs2164068	2q33.1	*PLCL1*	A/T	0.49	212,120 (18)	0.94	0.93–0.96	4.21×10^{-14}	0.82
rs2030519	3q28	*TPRG1, LPP*	G/A	0.49	212,120 (18)	1.06	1.04–1.07	1.83×10^{-13}	0.12
rs11256017	10p14	*CELF2, GATA3*	T/C	0.18	212,120 (18)	1.07	1.05–1.09	2.72×10^{-12}	0.60
rs17294280	15q22.33	*AAGAB, SMAD3*	G/A	0.25	212,120 (18)	1.07	1.05–1.09	5.97×10^{-12}	0.07
rs7824993	8q21.13	*ZBTB10, TPD52*	A/G	0.37	212,120 (18)	1.05	1.04–1.07	1.86×10^{-10}	0.56
rs9282864	16p11.2	*SULT1A1, SULT1A2*	C/A	0.33	208,761 (16)	0.94	0.93–0.96	4.69×10^{-10}	0.03
rs9687749	5q31.1	*IL13, RAD50*	T/G	0.44	207,604 (16)	1.06	1.04–1.09	1.84×10^{-9}	0.19
rs61977073	14q21.1	*TTC6*	G/A	0.22	212,120 (18)	1.06	1.04–1.08	5.78×10^{-9}	0.05
rs6470578	8q24.21	*TMEM75, MYC*	T/A	0.28	212,120 (18)	1.05	1.03–1.07	4.36×10^{-8}	0.02
rs3787184	20q13.2	*NFATC2, KCNG1*	G/A	0.19	207,604 (16)	0.94	0.93–0.96	4.76×10^{-8}	0.69

'Nearest genes' denotes the nearest up- and downstream gene (for intergenic variants with two genes listed) or surrounding gene (for intronic variants with one gene listed), with the exception of rs5743618, an exonic missense variant within *TLR1*. EA/OA, effect allele/other allele; EAF, effect-allele frequency. *P* values were calculated from two-tailed *z* scores from an inverse-variance-weighted fixed-effect meta-analysis. Het. *P*, *P* value for heterogeneity obtained from Cochrane's *Q* test.

2 helper T cell (T_H2)-driven immune responses[10]; our results therefore suggest that the GWAS signal at this locus involves structural changes related to allergen binding properties. This effect might be in addition to gene-regulatory effects similar to those found for autoimmune disease[11,12]. Most of the 20 loci not previously associated with allergic rhinitis implicate genes with a known role in the immune system, including *IL7R*[13,14], *SH2B3* (ref. [15]), *CEBPA-CEBPG*[16,17], *CXCR5* (ref. [18]), *FCER1G*, *NFKB1* (ref. [19]), *BACH2* (refs [20,21]), *TYRO3* (ref. [22]), *LTK*[23], *VPRBP*[24], *SPPL3* (ref. [25]), *OASL*[26], *RORA*[27], and *TNFSF11* (ref. [28]). Other loci implicated genes with no clear function in allergic rhinitis pathogenesis. These included one of the strongest associated loci in this meta-analysis at 12q24.31, with the top signal located between *CDK2AP1* and *C12orf65*, and containing cis-eQTLs in blood and lung tissue for several genes and evidence of enhancer–promoter interaction with *DDX55* in various immune cells (Supplementary Table 14 and further locus description in Supplementary Note). Concomitantly with the current study, a GWAS combining asthma, eczema, and allergic rhinitis was conducted[29]. Most (15/20) of the identified allergic rhinitis–associated loci in our study were also suggested in the previous, less specific GWAS[29] (as indicated in Table 2), whereas many suggested loci from the previous GWAS were not identified in our study. Asthma, eczema, and allergic rhinitis are related but distinct disease entities, often with separate disease mechanisms; for example, allergic sensitization is present in only 50% of children with asthma[30] and 35% of children with eczema[31]. Our results therefore complement those from the less specific 'atopic phenotype' GWAS[29] by pinpointing loci specifically associated and replicated in relation to allergic rhinitis.

Allergic rhinitis–associated loci were significantly enriched ($P<1\times10^{-5}$) in variants reported to be associated with autoimmune disorders. Reported autoimmune variants were located within 1 Mb of 31 (76%) of the 41 allergic rhinitis–associated loci. For 24 of these, an autoimmune top SNP was also associated with allergic rhinitis, and for 12 of these, the autoimmune top SNP was in LD ($r^2>0.5$) with the allergic rhinitis top SNP (Supplementary Table 15). For approximately half of these, the direction of the effect was the same for the autoimmune and allergic rhinitis top SNP, in line with findings from a previous study[32], thus underlining the complex genetic relationship between allergic rhinitis and autoimmunity, which might involve shared as well as diverging molecular mechanisms.

Assessment of the enrichment of allergic rhinitis–associated variant burden in open chromatin, as defined by DNase-hypersensitive sites, showed a clear enrichment in several blood and immune-cell subsets, with the largest enrichment in T cells (CD3 expressing), B cells (CD19 expressing), and T and NK cells (CD56 expressing) (Fig. 3, Supplementary Table 16 and Supplementary Fig. 6). We also probed tissue enrichment by using gene-expression data from a wide number of sources and observed enrichment of allergic rhinitis–associated genes in blood and immune-cell subsets, as well as in tissues of the respiratory system, including the oropharynx and the respiratory and nasal mucosa (Supplementary Table 17).

To explore biological connections and identify new pathways associated with allergic rhinitis, we combined all genes suggested from the eQTL and meQTL analyses, enhancer–promoter interactions, and localization within the top loci. The resultant prioritized gene set consisted of 255 genes, of which 89 (~36%) were

Table 2 | Association results of index markers (variant with lowest *P* value for each locus) not previously associated with allergic rhinitis reaching a Bonferroni-corrected significance threshold of 0.05 in the replication phase

				Discovery						Replication					Combined				
Variant	Locus	Nearest genes	EA/OA	EAF	*n* (studies)	OR	95% CI	*P*	Het. *P*	*n* (studies)	OR	95% CI	*P*	FWER	*n* (studies)	OR	95% CI	*P*	Het. *P*
rs7717955[a]	5p13.2	*CAPSL, IL7R*	T/C	0.27	212,120 (18)	0.95	0.93–0.96	1.50×10^{-9}	0.24	679,247 (10)	0.93	0.91–0.94	4.09×10^{-25}	1.06×10^{-23}	891,367 (28)	0.94	0.93–0.95	3.78×10^{-32}	0.09
rs63406760[a]	12q24.31	*CDK2AP1, C12orf65*	G/–	0.26	210,652 (17)	0.93	0.91–0.95	5.12×10^{-14}	0.91	675,338 (7)	0.95	0.93–0.96	3.27×10^{-12}	8.51×10^{-11}	885,990 (24)	0.94	0.93–0.95	2.54×10^{-24}	0.89
rs1504215[a]	6q15	*BACH2, GJA10*	A/G	0.34	207,604 (16)	0.95	0.94–0.97	1.49×10^{-8}	0.02	679,247 (10)	0.95	0.94–0.97	1.99×10^{-11}	5.17×10^{-10}	886,851 (26)	0.95	0.94–0.96	1.54×10^{-18}	0.05
rs28361986[a]	11q23.3	*CXCR5, DDX6*	A/T	0.20	212,120 (18)	0.93	0.91–0.95	1.81×10^{-14}	0.87	675,919 (8)	0.94	0.93–0.96	7.92×10^{-11}	2.06×10^{-9}	888,039 (26)	0.94	0.92–0.95	2.32×10^{-23}	0.91
rs2070902[a]	1q23.3	*AL590714.1, FCER1G*	T/C	0.25	212,120 (18)	1.06	1.04–1.08	1.03×10^{-10}	0.18	679,247 (10)	1.05	1.03–1.06	7.27×10^{-10}	1.89×10^{-8}	891,367 (28)	1.05	1.04–1.06	6.19×10^{-19}	0.23
rs111371454[a]	15q15.1	*ITPKA, RTF1*	G/A	0.21	212,120 (18)	1.06	1.03–1.08	1.65×10^{-7}	0.17	675,338 (7)	1.04	1.03–1.06	8.47×10^{-9}	2.20×10^{-7}	887,458 (25)	1.05	1.03–1.06	1.28×10^{-14}	0.22
rs12509403[a]	4q24	*MANBA, NFKB1*	T/C	0.32	212,120 (18)	0.95	0.94–0.97	9.97×10^{-9}	0.27	679,247 (10)	0.96	0.95–0.97	1.86×10^{-8}	4.84×10^{-7}	891,367 (28)	0.96	0.95–0.97	1.17×10^{-15}	0.39
rs9648346[a]	7p15.1	*JAZF1, TAX1BP1*	G/C	0.22	207,604 (16)	1.05	1.03–1.07	3.62×10^{-8}	0.74	679,247 (10)	1.04	1.03–1.06	1.39×10^{-7}	3.63×10^{-6}	886,851 (26)	1.05	1.03–1.06	3.30×10^{-14}	0.48
rs35350651[a]	12q24.12	*ATXN2, SH2B3*	C/–	0.49	206,136 (15)	1.04	1.03–1.06	6.63×10^{-8}	0.60	672,701 (6)	1.04	1.02–1.05	1.41×10^{-7}	3.66×10^{-6}	878,837 (21)	1.04	1.03–1.05	5.82×10^{-14}	0.43
rs2519093[a]	9q34.2	*ABO, OBP2B*	T/C	0.20	212,120 (18)	1.06	1.04–1.09	4.96×10^{-11}	0.38	675,919 (8)	1.04	1.03–1.06	2.96×10^{-7}	7.68×10^{-6}	888,039 (26)	1.05	1.04–1.07	2.79×10^{-16}	0.61
rs62257549	3p21.2	*VPRBP*	A/G	0.20	212,120 (18)	0.95	0.93–0.97	7.13×10^{-8}	0.45	677,615 (9)	0.96	0.94–0.97	3.37×10^{-7}	8.76×10^{-6}	889,735 (27)	0.95	0.94–0.97	1.84×10^{-13}	0.53
rs11677002	2p23.2	*FOSL2, RBKS*	C/T	0.45	212,120 (18)	0.96	0.95–0.98	3.80×10^{-7}	0.21	679,247 (10)	0.97	0.96–0.98	3.54×10^{-7}	9.20×10^{-6}	891,367 (28)	0.97	0.96–0.97	7.08×10^{-13}	0.36
rs35597970[a]	10q24.32	*ACTR1A, TMEM180*	–/A	0.45	210,652 (17)	1.06	1.04–1.07	1.34×10^{-13}	0.96	676,970 (8)	1.03	1.02–1.05	4.37×10^{-7}	1.14×10^{-5}	887,622 (25)	1.04	1.03–1.05	5.42×10^{-18}	0.53
rs2815765	1p31.1	*LRRIQ3, NEGR1*	T/C	0.37	212,120 (18)	0.95	0.94–0.97	1.18×10^{-9}	0.59	679,247 (10)	0.97	0.95–0.98	6.16×10^{-7}	1.60×10^{-5}	891,367 (28)	0.96	0.95–0.97	9.45×10^{-15}	0.52
rs11671925[a]	19q13.11	*CEBPA, SLC7A10*	A/G	0.17	206,136 (15)	0.94	0.92–0.96	1.80×10^{-8}	0.97	677,551 (9)	0.96	0.94–0.98	2.80×10^{-6}	7.29×10^{-5}	883,687 (24)	0.95	0.94–0.96	5.91×10^{-13}	0.60
rs2461475[a]	12q24.31	*SPPL3, ACADS*	C/T	0.47	212,120 (18)	1.04	1.02–1.05	9.19×10^{-7}	0.97	677,551 (9)	1.03	1.02–1.04	6.52×10^{-6}	0.0002	889,671 (27)	1.03	1.02–1.04	3.81×10^{-11}	0.83
rs6738964[a]	2q36.3	*SPHKAP, DAW1*	G/T	0.24	212,120 (18)	0.96	0.94–0.97	4.51×10^{-7}	0.72	679,247 (10)	0.97	0.96–0.98	4.96×10^{-5}	0.0013	891,367 (28)	0.96	0.95–0.97	1.86×10^{-10}	0.87
rs10519067[a]	15q22.2	*RORA*	A/–	0.13	212,120 (18)	0.93	0.91–0.96	1.78×10^{-9}	0.37	442,354 (7)	0.93	0.90–0.96	7.53×10^{-5}	0.0020	654,474 (25)	0.93	0.92–0.95	5.53×10^{-13}	0.36
rs138050288[a]	1p36.23	*RERE, SLC45A1*	–/CA	0.29	210,652 (17)	1.05	1.04–1.07	5.96×10^{-10}	0.71	675,338 (7)	1.03	1.01–1.04	0.0002	0.0046	885,990 (24)	1.04	1.03–1.05	6.62×10^{-12}	0.63
rs7328203	13q14.11	*TNFSF11, AKAP11*	G/T	0.46	212,120 (18)	1.05	1.03–1.06	5.94×10^{-9}	0.90	677,551 (9)	1.02	1.01–1.04	0.0005	0.0134	889,671 (27)	1.03	1.02–1.04	1.28×10^{-10}	0.78

'Nearest genes' denotes the nearest up- and downstream gene (for intergenic variants with two genes listed) or surrounding gene (for intronic variants with one gene listed), with the exception of rs1504215, an exonic synonymous variant within *BACH2*. EA/OA, effect allele/other allele; EAF, effect-allele frequency. *P* values were calculated from two-tailed *z* scores from an inverse-variance-weighted fixed-effect meta-analysis. Het. *P*, *P* value for heterogeneity obtained from Cochrane's Q test. FWER, family-wise error rate. [a]Variants also reported associated with a combined asthma/eczema/hay fever phenotype by Ferreira et al.[29] (within ±1 Mb).

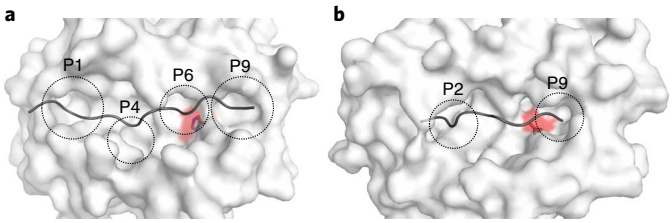

Fig. 2 | Structural visualization of amino acid variants associated with allergic rhinitis. The surface of the MHC molecule is shown in white, and the backbone of the bound peptide is shown in dark gray. The amino acid variant in focus is highlighted in red, and the peptide-binding pockets of the MHC molecule are indicated with dashed circles and are annotated P1–P9. **a**, The amino acid variant with the strongest association with allergic rhinitis was HLA-DQB1 His30 (MHC class II), located close to P6 with a distance of 6 Å to the peptide (excluding the peptide side chain). The protective amino acid variant at this location in relation to allergic rhinitis was histidine, whereas the risk variant was serine. Histidine is positively charged and has a large aromatic ring, whereas serine is not charged and not aromatic. Therefore, this alteration results in a substantial change in the binding-pocket environment. **b**, The strongest allergic rhinitis–associated amino acid variants in HLA-B (MHC class I) were HLA-B Asp116/His116-/Leu116, located close to P9 with a distance of 7 Å to the peptide (excluding the peptide side chain). The proximity to the bound peptide for both variants indicates that they are likely to affect the MHC–peptide interaction and thereby affect which peptides are presented.

present in more than one set (Supplementary Fig. 7). Overall, the full set was enriched in pathways involved in type 1 helper T cell (T_H1) and T_H2 activation (Fig. 4), antigen presentation, cytokine signaling, and inflammatory responses (Supplementary Table 18).

Using the 255 prioritized genes in combination with the STRING database to identify proteins interacting with the proteins encoded by the high-priority genes, we demonstrated a high degree of interaction at the protein level, and several of these proteins, including TNFSF11, NDUFAF1, PD-L1, IL-5, and IL-13, are targets of approved drugs or drugs in development (Fig. 4).

Allergic rhinitis is strongly correlated with allergic sensitization (presence of allergen-specific IgE), but sensitization is often present without allergic rhinitis, thus suggesting that specific mechanisms determine the progression from sensitization to disease. We therefore conducted a GWAS on sensitization to inhalant allergens, comprising 8,040 cases and 16,441 controls from 13 studies (Supplementary Table 1), in what is, to our knowledge, the largest GWAS on allergic sensitization to date[7]. A total of ten loci reached genome-wide significance, including one novel locus near *FASLG* (Supplementary Table 19). The genetic heritability on the liability scale was 17.75% (10% prevalence), a value considerably higher than the heritability of allergic rhinitis and consistent with a more homogeneous phenotype. Look-up of top allergic rhinitis–associated loci in the allergic-sensitization GWAS demonstrated high agreement: 40 of the 41 allergic rhinitis markers showed the same direction of effect, and 28 also showed nominal significance for allergic sensitization (Supplementary Table 20). This result suggests that allergic

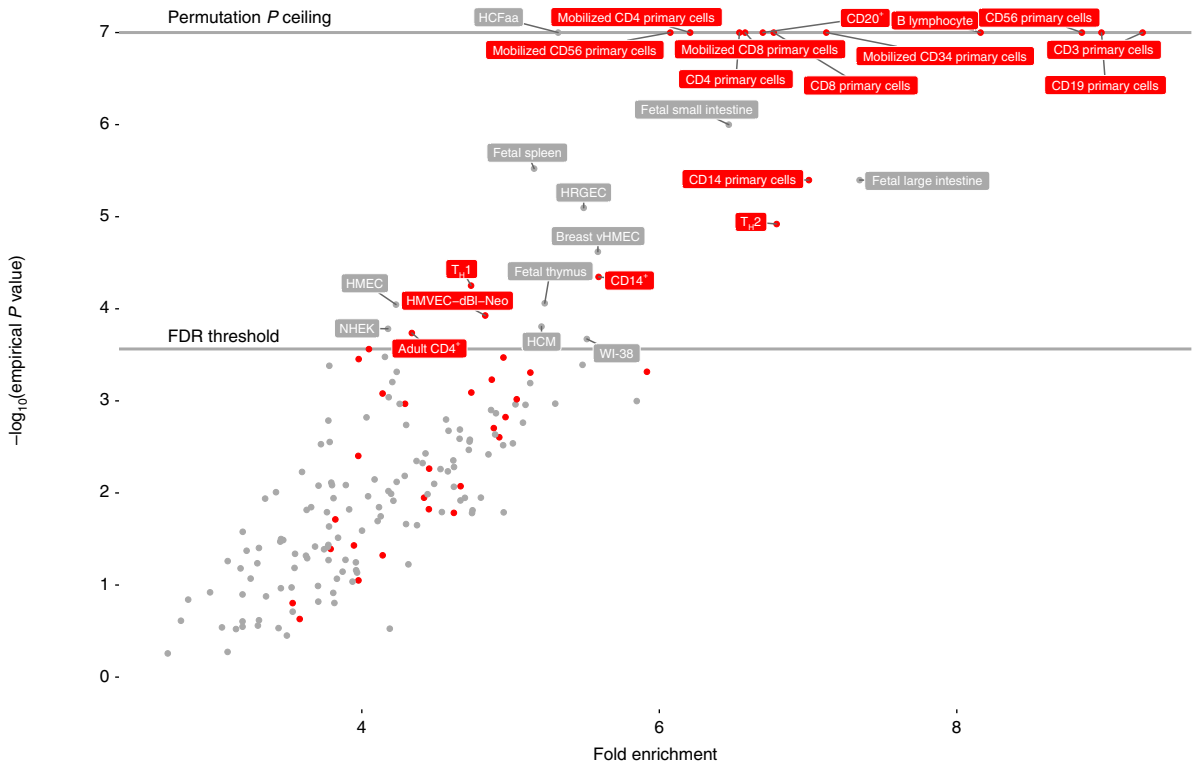

Fig. 3 | Enrichment of allergic rhinitis-associated variants in tissue-specific open chromatin. Enrichment of 16,531,985 genetic variants associated with allergic rhinitis in 212,120 individuals (at $P < 1 \times 10^{-8}$ as a threshold for marker association) in 189 cell types from ENCODE and Roadmap epigenomics data. Enrichment and *P* values were calculated empirically against a permuted genomic background with the GARFIELD tool. Red labels indicate blood and blood-related cell types, and gray labels indicate other cell types. Owing to the number of permutations (1×10^7), empirical *P* values reached a minimum ceiling of $1/(1 \times 10^7)$. The false discovery rate (FDR) threshold was 0.00026. For Epstein–Barr virus–transformed B-lymphocyte cell types (cell type GM****), only the most enriched instance is shown ('B lymphocyte'). NHEK, normal human epidermal keratinocytes; HMEC/vHMEC, mammary epithelial cells; HCM, human cardiac myocytes; WI-38, lung fibroblast derived; HRGEC, human renal glomerular endothelial cells; HCFaa, human cardiac fibroblast–adult atrial cells; HMVEC–dBl–Neo, human microvascular endothelial cells.

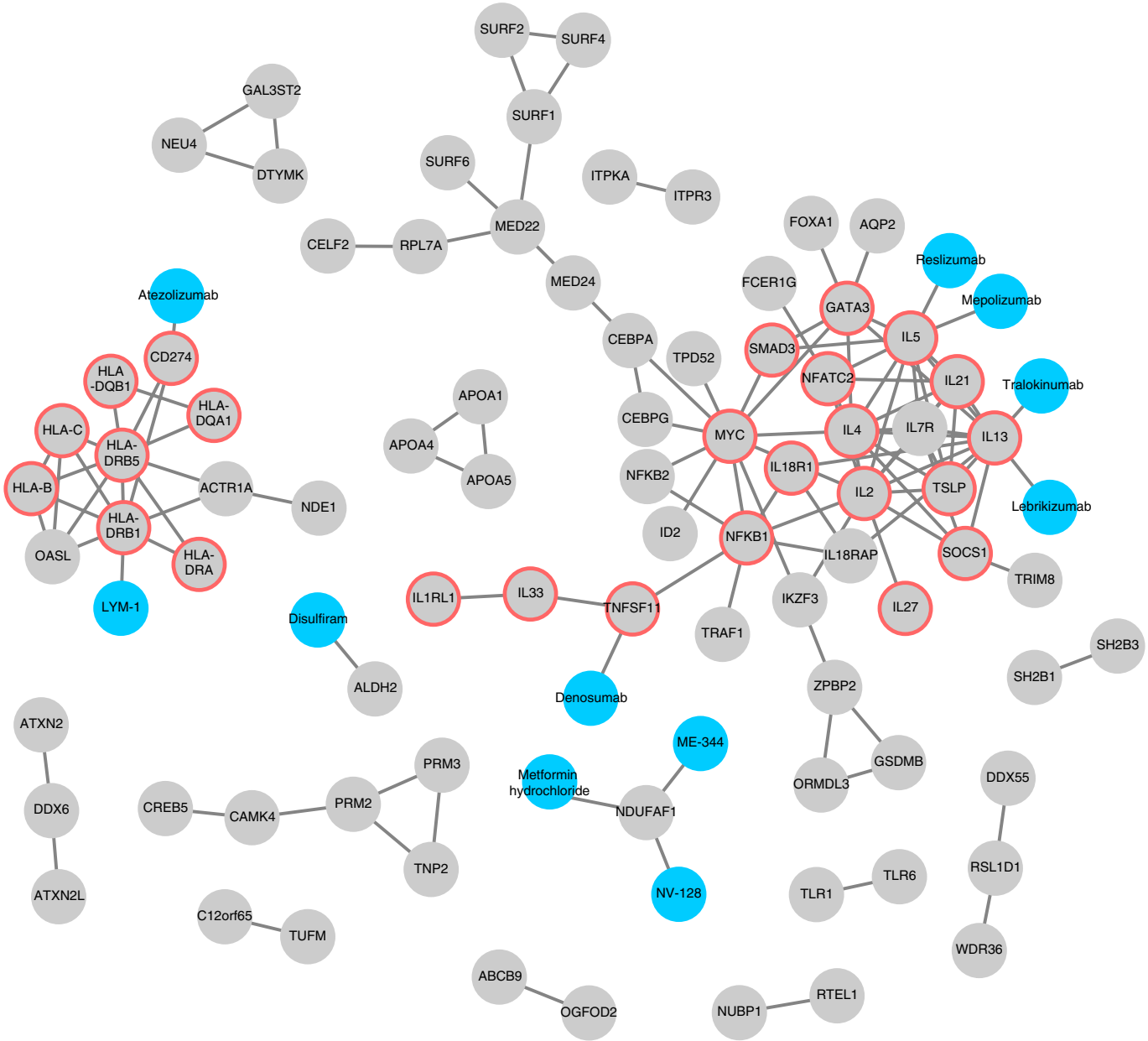

Fig. 4 | Interaction network among drugs and proteins from genes associated with allergic rhinitis. Gray nodes represent locus genes as well as genes prioritized from eQTL/meQTL and PCHiC sources, on the basis of genetic association of 16,531,985 markers with allergic rhinitis in 212,120 individuals. Blue nodes represent drugs from the ChEMBL drug database. Edges represent very high-confidence interactions from the STRING database (for locus–locus interactions) and drug-target evidence (for drug–locus interactions). Red borders indicate genes with protein products that were significantly enriched in the 'T_H1 and T_H2 activation' pathway ($-\log(P) > 19.1$) from ingenuity pathway analysis.

rhinitis and allergic sensitization share biological mechanisms, and that allergic-sensitization loci generally affect systemic allergic sensitization. We compared genetic pathways of allergic rhinitis and allergic sensitization by using the DEPICT tool, which showed overlap in enriched pathways but also differences among the top gene sets: allergic rhinitis gene sets were characterized by B cell, T_H2, and parasite responses, whereas allergic-sensitization gene sets were characterized by a broader activation of cells (Supplementary Fig. 8 and Supplementary Tables 21 and 22).

Nonallergic rhinitis, defined as having rhinitis symptoms without evidence of allergic sensitization, is a common but poorly understood disease entity[33]. We performed what is, to our knowledge, the first GWAS on this phenotype, with the goal of identifying specific rhinitis mechanisms. The analysis included 2,028 cases and 9,606 controls from nine studies but did not identify any risk loci at the genome-wide-significant level. Comparison with the allergic rhinitis results suggested some overlap in susceptibility loci (Supplementary Note and Supplementary Table 23).

We estimated the proportion of allergic rhinitis in the general population that was attributable to the 41 identified allergic rhinitis–associated loci and obtained a conservative population-attributable risk-fraction estimate of 39% (95% confidence interval (CI) 26–50%), by considering the 10% of the population with the lowest genetic-risk scores to represent an 'unexposed' group. The allergic rhinitis prevalence plotted by genetic-risk score (Supplementary Fig. 9) showed a prevalence approximately two times higher in the 7% of the population with the highest risk score than the 7% with the lowest risk score.

Finally, we investigated the genetic correlation of allergic rhinitis with allergic sensitization, asthma[34], and eczema[35] by using LD-score regression. There was a strong correlation between allergic rhinitis and allergic sensitization ($r^2 = 0.73$, $P < 2 \times 10^{-34}$), a moderate correlation with asthma ($r^2 = 0.60$, $P < 3 \times 10^{-14}$), and a weaker correlation with eczema ($r^2 = 0.40$, $P < 2 \times 10^{-7}$).

We tested the identified allergic rhinitis–associated loci for association with allergic rhinitis in non-European cohorts, which showed only a nominally significant association for one locus, but this analysis had limited statistical power, owing to population sizes (Supplementary Table 24).

In conclusion, we expanded the number of established susceptibility loci for allergic rhinitis and highlighted the involvement of allergic rhinitis–susceptibility loci in diverse immune-cell types and both innate and adaptive IgE-related mechanisms. Future studies of novel allergic rhinitis–associated loci might identify targets for the treatment and prevention of this disease.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0157-1.

Received: 30 June 2017; Accepted: 10 May 2018;
PublishePublished online: 16 July 2018

References

1. Greiner, A. N., Hellings, P. W., Rotiroti, G. & Scadding, G. K. Allergic rhinitis. *Lancet* **378**, 2112–2122 (2011).
2. Björkstén, B., Clayton, T., Ellwood, P., Stewart, A. & Strachan, D. Worldwide time trends for symptoms of rhinitis and conjunctivitis: phase III of the International Study of Asthma and Allergies in Childhood. *Pediatr. Allergy Immunol.* **19**, 110–124 (2008).
3. Willemsen, G., van Beijsterveldt, T. C. E. M., van Baal, C. G. C. M., Postma, D. & Boomsma, D. I. Heritability of self-reported asthma and allergy: a study in adult Dutch twins, siblings and parents. *Twin Res. Hum. Genet.* **11**, 132–142 (2008).
4. Fagnani, C. et al. Heritability and shared genetic effects of asthma and hay fever: an Italian study of young twins. *Twin Res. Hum. Genet.* **11**, 121–131 (2008).
5. Ramasamy, A. et al. A genome-wide meta-analysis of genetic variants associated with allergic rhinitis and grass sensitization and their interaction with birth order. *J. Allergy Clin. Immunol.* **128**, 996–1005 (2011).
6. Hinds, D. A. et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat. Genet.* **45**, 907–911 (2013).
7. Bønnelykke, K. et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat. Genet.* **45**, 902–906 (2013).
8. Ferreira, M. A. R. 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. Bunyavanich, S. et al. Integrated genome-wide association, coexpression network, and expression single nucleotide polymorphism analysis identifies novel pathway in allergic rhinitis. *BMC Med. Genom.* **7**, 48 (2014).
10. Jahn-Schmid, B., Pickl, W. F. & Bohle, B. Interaction of allergens, major histocompatibility complex molecules, and T cell receptors: a 'ménage à trois' that opens new avenues for therapeutic intervention in type I allergy. *Int. Arch. Allergy Immunol.* **156**, 27–42 (2011).
11. Cavalli, G. et al. MHC class II super-enhancer increases surface expression of HLA-DR and HLA-DQ and affects cytokine production in autoimmune vitiligo. *Proc. Natl Acad. Sci. USA* **113**, 1363–1368 (2016).
12. Hayashi, M. et al. Autoimmune vitiligo is associated with gain-of-function by a transcriptional regulator that elevates expression of HLA-A*02:01 in vivo. *Proc. Natl Acad. Sci. USA* **113**, 1357–1362 (2016).
13. Puel, A., Ziegler, S. F., Buckley, R. H. & Leonard, W. J. Defective IL7R expression in $T^-B^+NK^+$ severe combined immunodeficiency. *Nat. Genet.* **20**, 394–397 (1998).
14. Lundmark, F. et al. Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. *Nat. Genet.* **39**, 1108–1113 (2007).
15. Mori, T. et al. Lnk/Sh2b3 controls the production and function of dendritic cells and regulates the induction of IFN-γ-producing T cells. *J. Immunol.* **193**, 1728–1736 (2014).
16. Scott, L. M., Civin, C. I., Rorth, P. & Friedman, A. D. A novel temporal expression pattern of three C/EBP family members in differentiating myelomonocytic cells. *Blood* **80**, 1725–1735 (1992).
17. Gao, H., Parkin, S., Johnson, P. F. & Schwartz, R. C. C/EBP gamma has a stimulatory role on the IL-6 and IL-8 promoters. *J. Biol. Chem.* **277**, 38827–38837 (2002).
18. León, B. et al. Regulation of T_H2 development by $CXCR5^+$ dendritic cells and lymphotoxin-expressing B cells. *Nat. Immunol.* **13**, 681–690 (2012).
19. Lawrence, T. The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harb. Perspect. Biol.* **1**, a001651 (2009).
20. Shinnakasu, R. et al. Regulated selection of germinal-center cells into the memory B cell compartment. *Nat. Immunol.* **17**, 861–869 (2016).
21. Roychoudhuri, R. et al. BACH2 regulates $CD8^+$ T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat. Immunol.* **17**, 851–860 (2016).
22. Rothlin, C. V., Ghosh, S., Zuniga, E. I., Oldstone, M. B. A. & Lemke, G. TAM receptors are pleiotropic inhibitors of the innate immune response. *Cell* **131**, 1124–1136 (2007).
23. Chan, P. Y. et al. The TAM family receptor tyrosine kinase TYRO3 is a negative regulator of type 2 immunity. *Science* **352**, 99–103 (2016).
24. Kassmeier, M. D. et al. VprBP binds full-length RAG1 and is required for B-cell development and V(D)J recombination fidelity. *EMBO J.* **31**, 945–958 (2012).
25. Hamblet, C. E., Makowski, S. L., Tritapoe, J. M. & Pomerantz, J. L. NK cell maturation and cytotoxicity are controlled by the intramembrane aspartyl protease SPPL3. *J. Immunol.* **196**, 2614–2626 (2016).
26. Andersen, J. B., Strandbygård, D. J., Hartmann, R. & Justesen, J. Interaction between the 2′-5′ oligoadenylate synthetase-like protein p59 OASL and the transcriptional repressor methyl CpG-binding protein 1. *Eur. J. Biochem.* **271**, 628–636 (2004).
27. Halim, T. Y. F. et al. Retinoic-acid-receptor-related orphan nuclear receptor alpha is required for natural helper cell development and allergic inflammation. *Immunity* **37**, 463–474 (2012).
28. Anderson, D. M. et al. A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
29. Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757 (2017).
30. Pearce, N., Pekkanen, J. & Beasley, R. How much asthma is really attributable to atopy? *Thorax* **54**, 268–272 (1999).
31. Böhme, M., Wickman, M., Lennart Nordvall, S., Svartengren, M. & Wahlgren, C. F. Family history and risk of atopic dermatitis in children up to 4 years. *Clin. Exp. Allergy* **33**, 1226–1231 (2003).
32. Kreiner, E. et al. Shared genetic variants suggest common pathways in allergy and autoimmune diseases. *J. Allergy Clin. Immunol.* **140**, 771–781 (2017).
33. Bousquet, J. et al. Important research questions in allergy and related diseases: nonallergic rhinitis: a GA2LEN paper. *Allergy* **63**, 842–853 (2008).
34. Moffatt, M. F. et al. Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007).
35. Paternoster, L. et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* **47**, 1449–1456 (2015).

Acknowledgements

Detailed acknowledgements and funding details for each contributing study are provided in the Supplementary Note.

Author contributions

Study design and management: K.B., J.W., M.S., and D.P.S. Meta-analyses: M.S. and J.W. Manuscript writing: K.B., J.W., M.S., J.A.C., J.T., L.E.J., and M.A.F. Systems biology analyses: J.W., J.A.C., J.T., L.E.J., J.M.M., S.B.-G., and D.T. Data collection, analysis, and design in the individual contributing studies: K.B., J.W., M.S., J.A.C., C.F., A. Abdellaoui, T.S.A., A.C.A., A.F.S.A., J.M.A., A. Arnold, A.B.-L., H. Baurecht, C.E.M.v.B., E.R.B., D.I.B., S. Bunyavanich, E.B., Z.C., I.C., A.C., H.T.d.D., S.C.D., J.D., L.D., M.J.E., W.J.G., C.G., F.G., R.G., H.G., T.H., J. Heinrich, J. Henderson, N.H.-P., D.A.H., P.H., M.I., V.W.V.J., M.-R.J., D.L.J., I.J., M.K., J.K., A.K., Y.-A.L., A.M.L., X.L., F.L.-D., E.M., D.A.M., R.M., D.L.N., E.A.N., T.P., L.P., C.E.P., G.P., M.P.-Y., N.M.P.-H., F.R., A.S., K.S., J.S., G.S., E.T., P.J.T., C.T., M.T., J.Y.T., C.A.W., S. Weidinger, S. Weiss, G.W., L.K.W., C.O., M.A.F., H. Bisgaard, D.P.S., The 23andMe Research Team and AAGC collaborators. Immunological interpretation: N.S. and S. Brix. Gene-expression analysis: M.G. and J.D. Protein modeling: K.K.J.

Competing interests

G.S., I.J., and K.S. are affiliated with deCODE genetics/Amgen and declare competing financial interests as employees. C.T., D.A.H., J.Y.T., and the 23andMe Research Team are employees of and hold stock and/or stock options in 23andMe, Inc. L.P. has received a fee for participating in a scientific-input engagement meeting from Merck Sharp & Dohme Limited, outside of this work.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0157-1.

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

Correspondence and requests for materials should be addressed to K.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Johannes Waage[1,74], Marie Standl[2,74], John A. Curtin[3], Leon E. Jessen[1], Jonathan Thorsen[1], Chao Tian[4], Nathan Schoettler[5], The 23andMe Research Team[6], AAGC collaborators[6], Carlos Flores[7,8], Abdel Abdellaoui[9,10], Tarunveer S. Ahluwalia[1], Alexessander C. Alves[11], Andre F. S. Amaral[12], Josep M. Antó[13,14,15,16], Andreas Arnold[17], Amalia Barreto-Luis[7], Hansjörg Baurecht[18], Catharina E. M. van Beijsterveldt[9], Eugene R. Bleecker[19], Sílvia Bonàs-Guarch[20], Dorret I. Boomsma[9,21], Susanne Brix[22], Supinda Bunyavanich[23], Esteban G. Burchard[24,25], Zhanghua Chen[26], Ivan Curjuric[27,28], Adnan Custovic[29], Herman T. den Dekker[30,31,32], Shyamali C. Dharmage[33], Julia Dmitrieva[34], Liesbeth Duijts[30,32,35], Markus J. Ege[36], W. James Gauderman[26], Michel Georges[34], Christian Gieger[2,37], Frank Gilliland[26], Raquel Granell[38], Hongsheng Gui[39], Torben Hansen[40], Joachim Heinrich[2,41], John Henderson[38], Natalia Hernandez-Pacheco[7,42], Patrick Holt[43], Medea Imboden[27,28], Vincent W. V. Jaddoe[30,31,44], Marjo-Riitta Jarvelin[11,45,46,47], Deborah L. Jarvis[12], Kamilla K. Jensen[48], Ingileif Jónsdóttir[49,50], Michael Kabesch[51], Jaakko Kaprio[52,53,54], Ashish Kumar[27,28,55], Young-Ae Lee[56,57], Albert M. Levin[58], Xingnan Li[59], Fabian Lorenzo-Diaz[42], Erik Melén[55,60], Josep M. Mercader[20,61,62], Deborah A. Meyers[19], Rachel Myers[5], Dan L. Nicolae[5], Ellen A. Nohr[63], Teemu Palviainen[53], Lavinia Paternoster[38], Craig E. Pennell[64], Göran Pershagen[55,65], Maria Pino-Yanes[7,8,42], Nicole M. Probst-Hensch[27,28], Franz Rüschendorf[56], Angela Simpson[3], Kari Stefansson[49,50], Jordi Sunyer[13], Gardar Sveinbjornsson[49], Elisabeth Thiering[2,66], Philip J. Thompson[67], Maties Torrent[68], David Torrents[20,69], Joyce Y. Tung[4], Carol A. Wang[64], Stephan Weidinger[18], Scott Weiss[70], Gonneke Willemsen[9], L. Keoki Williams[39,71], Carole Ober[5], David A. Hinds[4], Manuel A. Ferreira[72], Hans Bisgaard[1], David P. Strachan[73] and Klaus Bønnelykke[1*]

[1]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. [2]Institute of Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [3]Division of Infection, Immunity and Respiratory Medicine, University of Manchester, Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust, Manchester, UK. [4]23andMe, Inc., Mountain View, CA, USA. [5]Department of Human Genetics, University of Chicago, Chicago, IL, USA. [6]A list of members and affiliations appears in the Supplementary Note. [7]Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Tenerife, Spain. [8]CIBER de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain. [9]Department of Biological Psychology, Netherlands Twin Register, VU University, Amsterdam, The Netherlands. [10]Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [11]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK. [12]Population Health and Occupational Disease, National Heart and Lung Institute, Imperial College London, London, UK. [13]ISGlobal, Barcelona, Spain. [14]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [15]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [16]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. [17]Clinic and Polyclinic of Dermatology, University Medicine Greifswald, Greifswald, Germany. [18]Department of Dermatology, Venereology and Allergology, University-Hospital Schleswig-Hostein, Campus Kiel, Kiel, Germany. [19]Divisions of Pharmacogenomics and Genetics, Genomics and Precision Medicine, Department of Medicine, University of Arizona College of Medicine, Tucson, AZ, USA. [20]Barcelona Supercomputing Center (BSC), Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona, Spain. [21]APH Amsterdam Public Health, Amsterdam, The Netherlands. [22]Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark. [23]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [24]Department of Medicine, University of California San Francisco, San Francisco, CA, USA. [25]Department of Bioengineering & Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA. [26]Department of Preventive Medicine, University of Southern California, Keck School of Medicine, Los Angeles, CA, USA. [27]University of Basel, Basel, Switzerland. [28]Swiss Tropical and Public Health Institute, Basel, Switzerland. [29]Department of Paediatrics, Imperial College London, London, UK. [30]The Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands. [31]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [32]Department of Pediatrics, Division of Respiratory Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. [33]Allergy and Lung Health Unit, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia. [34]Laboratory of Animal Genomics, Unit of Medical Genomics, GIGA Institute, University of Liège, Liège, Belgium. [35]Department of Pediatrics, Division of Neonatology, Erasmus Medical Center, Rotterdam, The Netherlands. [36]LMU Munich, Dr von Hauner Children's Hospital, Munich, and German Center for Lung Research (DZL), Munich, Germany. [37]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [38]MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [39]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI, USA. [40]Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Department of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [41]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, University of Munich Medical Center, Ludwig-Maximilians-Universität München, Munich, Germany. [42]Genomics and Health Group, Department of Biochemistry, Microbiology, Cell Biology and Genetics, Universidad de La Laguna, La Laguna, Tenerife, Spain. [43]Telethon Kids Institute (TKI), Perth, Western Australia, Australia. [44]Department of Pediatrics, Erasmus Medical Center, Rotterdam, The Netherlands. [45]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. [46]Biocenter Oulu, University of Oulu, Oulu, Finland. [47]Unit of Primary Care, Oulu University Hospital, Oulu, Finland. [48]Department of Bio and Health

Informatics, Technical University of Denmark, Kongens Lyngby, Denmark. [49]deCODE genetics/Amgen Inc, Reykjavik, Iceland. [50]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [51]Department of Pediatric Pneumology and Allergy, University Children's Hospital Regensburg (KUNO), Regensburg, Germany. [52]Department of Public Health, University of Helsinki, Helsinki, Finland. [53]Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland. [54]National Institute for Health and Welfare, Helsinki, Finland. [55]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [56]Max-Delbrück-Center (MDC) for Molecular Medicine, Berlin, Germany. [57]Clinic for Pediatric Allergy, Experimental and Clinical Research Center, Charité Universitätsmedizin Berlin, Berlin, Germany. [58]Department of Public Health Sciences, Henry Ford Health System, Detroit, MI, USA. [59]Divisions of Genetics, Genomics and Precision Medicine, Department of Medicine, University of Arizona College of Medicine, Tucson, AZ, USA. [60]Sachs' Children's Hospital, Stockholm, Sweden. [61]Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [62]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [63]Institute of Clinical Research, University of Southern Denmark, Department of Obstetrics & Gynecology, Odense University Hospital, Odense, Denmark. [64]School of Medicine and Public Health, Faculty of Medicine and Health, University of Newcastle, Callaghan, New South Wales, Australia. [65]Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden. [66]Ludwig-Maximilians-University of Munich, Dr. von Hauner Children's Hospital, Division of Metabolic Diseases and Nutritional Medicine, Munich, Germany. [67]Institute for Respiratory Health, Harry Perkins Institute of Medical Research, University of Western Australia, Nedlands, Western Australia, Australia. [68]Ib-Salut, Area de Salut de Menorca, Institut d'Investigacio Sanitaria Illes Balears (IdISBa), Palma de Mallorca, Spain. [69]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [70]Channing Division of Network Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA, USA. [71]Department of Internal Medicine, Henry Ford Health System, Detroit, MI, USA. [72]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [73]Population Health Research Institute, St George's, University of London, London, UK. [74]These authors contributed equally: Johannes Waage, Marie Standl. *e-mail: kb@copsac.com

Methods

Phenotype definition. *Allergic rhinitis.* Cases were defined as individuals who had ever received a diagnosis or experienced symptoms of allergic rhinitis, on the basis of available phenotype definitions in the included studies (Supplementary Table 3 and cohort-recruitment details in Supplementary Note). All relevant ethical regulations were followed, as specified in relation to the individual studies in the Supplementary Note. To maximize numbers and optimize statistical power, we did not require doctor-diagnosed allergic rhinitis or verification by allergic sensitization. This approach was confirmed by a sensitivity analysis in 23andMe on the basis of association with known risk loci for allergic rhinitis (data not shown). Controls were defined as individuals who had never received a diagnosis or experienced symptoms of allergic rhinitis.

Allergic sensitization. We considered specific IgE production against inhalant allergens without restricting by assessment method or type of inhalant allergen. Cases were defined as individuals with objectively measured sensitization against at least one of the inhalant allergens tested for in the respective studies, and controls were defined as individuals who were not sensitized against any of the allergens tested for. We included sensitization assessed by skin reaction after puncture of the skin with a droplet of allergen extract (SPT) and/or by detection of the levels of circulating allergen-specific IgE in the blood. The SPT wheal-diameter cutoffs were ≥ 3 mm for cases and < 1 mm for controls. To optimize case specificity and the correlation between methods, we chose a high cutoff of specific IgE levels for cases (0.7 IU/ml) and a low cutoff for controls (0.35 IU/ml).

Nonallergic rhinitis. Cases were defined as individuals with current allergic rhinitis symptoms (within the previous 12 months) and no allergic sensitization (negative specific IgE (<0.35 IU/ml) and/or negative skin-prick test (<1 mm) for all allergens and time points tested).

Controls were defined as individuals who had never had symptoms of allergic rhinitis and showed no allergic sensitization (negative specific IgE (<0.35 IU/ml) and/or negative skin-prick test (<1 mm) for all allergens and time points tested)

For all three phenotypes, we combined data from children and adults but chose a lower age limit of 6 years, because allergic rhinitis and sensitization status at younger ages show poorer correlation with status later in life, owing to both transient symptoms/sensitization status and frequent development of symptoms/sensitization during late childhood.

GWAS quality control and cohort summary-data harmonization. For allergic rhinitis, allergic sensitization, and nonallergic rhinitis, data were imputed separately for each cohort with the 1000 Genomes Project (1KGP) phase 1, version 3 release, and the genome-wide association analysis was adjusted for sex and, if necessary, for age and principal components (Supplementary Table 3). All studies included individuals of European descent, except Generation R and RAINE, which comprised a mixed, multiancestral population. We used EasyQC v. 9.2 (ref. [36]) for quality control and marker harmonization for cohort-level meta-GWAS summary files. Cohort data were harmonized to genome build GRCh37 and were checked against the 1KGP phase 3 reference-allele frequencies for processing problems. GWAS summary 'karyograms' were visually inspected to catch cohorts with incomplete data. Distributions of estimate coefficients and errors, as well as plots of standard error versus sample size and of P value versus Z score, were inspected for each cohort for systematic errors in statistical models. Ambiguous markers that were nonunique in terms of both genomic position and allele coding were removed. A minimum imputation score of 0.3 (r^2) or 0.4 (proper_info) was required for markers. A minimum minor-allele count of 7 was required for each marker in each cohort, as suggested by the GIANT consortium and EasyQC.

Meta-analysis. For allergic rhinitis, allergic sensitization, and nonallergic rhinitis, meta-analysis for the discovery phase was conducted with GWAMA[37] with an inverse-variance-weighted fixed-effect model with genomic-control correction of the individual studies. Each locus is represented by the variant showing the strongest evidence within a 1-Mb buffer. Loci were inspected visually by plotting genomic neighborhood and coloring for 1KGP r^2 values. From the pool of genome-wide-significant markers in the discovery, one locus with index marker rs193243426 without a credible LD structure was removed from further analysis (Supplementary Fig. 10). Heterogeneity was assessed with Cochran's Q test.

Meta-analysis of replication candidates from the allergic rhinitis discovery phase was carried out in R version 3.4.0 and the *meta* package version 4.8-2 with an inverse-variance-weighted fixed-effect model. For a subset of markers, cohorts reported suitable proxies ($r^2 > 0.85$), and followed-up markers were not present or had insufficient imputation or genotyping quality (Supplementary Table 25).

Gene-set-overrepresentation analysis, discovery phase. To facilitate the selection of biologically relevant discovery candidates in the sub-genome-wide significant stratum ($5 \times 10^{-8} < P < 1 \times 10^{-6}$), we used a custom algorithm for gene-set-overrepresentation analysis implemented in R, with a scoring and permutation regime modeled after MAGENTA[38]. Genes with lengths less than 200 bp, with copies on multiple chromosomes, and with multiple copies on the same chromosome more than 1 Mb apart were removed from analysis. Gene models (GENCODE v 19) were downloaded from the UCSC Table Browser[39] and expanded 110 kb upstream and 40 kb downstream, similarly to the MAGENTA procedure. The HLA region was excluded from analysis (chromosome 6: 29691116–33054976). Similarly to the MAGENTA procedure, gene scores were adjusted for the number of markers per gene, gene width, recombination hotspots, genetic distance, and number of independent markers per gene, all with updated data from the UCSC Table Browser. For the gene-set-overrepresentation permutation calculation, gene sets from the MSigDB collections c2, c3, c5, c7, and hallmark were included[40]. A MAGENTA-style enrichment cutoff at 95% was used. Gene sets with FDR <0.05 were considered.

Conditional analyses. To identify additional independent markers at each discovery genomic region, we used genome-wide complex trait analysis (GCTA) v. 1.26.0 (ref. [41]). Within a window of ±1 Mb of each discovery-phase index marker, all markers were conditioned on the index by using the --cojo-cond feature of GCTA with default parameters. Plink v. 1.90b3.42 (ref. [42]) was used to calculate r^2 for GCTA with the UK10K full genotype panel[43] as a reference. A total of 42 of 52 markers from the full discovery phase were present in UK10K. Because a MAF-dependent inflation of conditional P values was observed (data not shown), only conditional markers with MAF ≥10% were selected.

Locus definition and credible sets for VEP annotation. Discovery loci were defined as index markers extended with markers in LD ($r^2 \geq 0.5$), on the basis of 1KGP phase 3. Protein-coding gene transcript models (GENCODE v. 24) were downloaded from the UCSC Table Browser, and the nearest upstream and downstream genes, as well as all genes within the extended loci, were annotated. Credible sets for each locus were calculated with the method of Morris[44].

LD was calculated for each discovery index variant within ±500 kb, and markers with $r^2 < 0.1$ were excluded. For the remaining markers, the Bayesian factor (ABF) values and the posterior probabilities (PostProb) were calculated, and cumulative posterior probability values were generated on the basis of ranking markers on ABF. Finally, variants were included in the 99%-credible set until the cumulative posterior probability was ≥0.99.

Credible sets for each locus were annotated with information on mutation effects in coding regions by using the Variant Effect Predictor (VeP) REST API[45], exporting only the nonsynonymous substitutions.

GWAS-catalog lookup. For annotation of markers with identification in previous GWA studies, the GWAS catalog was downloaded from NHGRI-EBI (v.1.0.1, 28 November 2016). For this analysis, allergic rhinitis–associated loci were lifted from genomic build GRCh37 to GRCh38 and extended with ±1 Mb in each direction before being overlapped with GWAS-catalog annotations. Relevant GWAS-catalog overlap traits were binned into the trait groups 'allergic rhinitis', 'asthma', 'autoimmune', 'eczema', 'infectious diseases', 'lung-related traits', and 'other allergy'. One million random genomic intervals of the same length (2 Mb) were obtained to generate a background overlap distribution, and P values were calculated from this background.

Classical HLA-allele analysis. Analyses of imputed classical HLA alleles were performed in the 23andMe study (AR discovery population) comprising 49,180 individuals with allergic rhinitis and 124,102 controls.

HLA imputation was performed with HIBAG v. 1.2.3 (ref. [46]). We imputed allelic dosage for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, and *HLA-DRB1* loci at four-digit resolution with the default settings of HIBAG, for a total of 292 classical HLA alleles.

Using an approach suggested by de Bakker[47], we downloaded the files mapping HLA alleles to amino acid sequences from https://www.broadinstitute.org/mpg/snp2hla/ and mapped our imputed HLA alleles at four-digit resolution to the corresponding amino acid sequences; in this way, we translated the imputed HLA allelic dosages directly to amino acid dosages. We encoded all amino acid variants in the 23andMe European samples as 2,395 biallelic amino acid polymorphisms, as previously described[48].

Similarly to the SNP imputation, we measured imputation quality by using r^2, which is the ratio of the empirically observed variance of the allele dosage to the expected variance, assuming Hardy–Weinberg equilibrium.

To test associations among imputed HLA alleles, amino acid variants, and phenotypes, we performed logistic regression using the same set of covariates used in the SNP-based GWAS. We applied a forward stepwise strategy within each type of variant to establish statistically independent signals in the HLA region. Within each variant type, we first identified the most strongly associated signals (lowest P value) and performed forward iterative conditional regression to identify other independent signals. All analyses were controlled for sex and five principal components of genetic ancestry. The P values were calculated with a likelihood-ratio test.

Structural visualization of amino acid variants. Structural visualization of amino acid variants was performed for the strongest associated variants in HLA-DQB1 (position 30) and HLA-B (position 116), respectively (Supplementary Table 10) and was based on X-ray structures from the Protein Data Bank (PDB)[49]. To find

the best structure, we used the specialized search function in the Immune Epitope Database[50], selecting only X-ray-crystalized structures for the specific MHC classes HLA-DQB1 (class II) and HLA-B (class I). Using this criterion, we found 17 crystallized structures for HLA-DQB1 and 164 structures for HLA-B. From these lists, we selected the structure with the lowest resolution and the amino acids encoded by the reported top SNPs. The accession codes for the selected structures were PDB 4MAY[51] for HLA-DQB1 and PDB 2A83 (ref. [52]) for HLA-B. Both structures were visualized in PyMOL v. 1.8.2.1 (http://www.pymol.org/). Furthermore, we used PyMOL to measure intramolecular distances from the side chains of the amino acids associated with allergic rhinitis to the Cα atoms in the peptide. This distance measure was chosen to accommodate the possibility of different amino acids in the peptide. For two amino acids to interact, the distance should be approximately 4 Å or less. We measured distances of 6 Å (HLA-DQB1) and 7 Å (HLA-B). However, these distances do not include the peptide side chains, which range from 1.5 to 8.8 Å. Therefore, we estimate that physical interaction between the amino acids associated with allergic rhinitis and the peptide is likely.

Genetic heritability and genetic correlation. For calculating genetic heritability and genetic correlation between allergic rhinitis and allergic sensitization, as well as between clinical cohorts and 23andMe within allergic rhinitis, we used the LD-score regression-based method, as implemented in LDSC v. 1.0 (refs [45,53]). The population prevalence was set to 10% for allergic rhinitis and allergic sensitization. Genetic correlation analysis among allergic rhinitis, allergic sensitization, and published GWAS studies was carried out in LDHUB platform v. 1.3.1 (ref. [54]) against all traits but excluding metabolites[55].

eQTL sources and analysis. From GTEx V6p[56], all significant variant-gene cis-eQTL pairs for whole blood, lung, and EBV-transformed lymphocytes were downloaded from https://gtexportal.org/ and used in analysis. From Westra et al.[57], we downloaded both cis and trans eQTLs in whole blood, then used variant-gene pairs with FDR <0.1 in subsequent analyses. From Fairfax et al.[58], we downloaded cis-eQTLs from monocytes and B cells, then used variant-gene pairs with FDR < 0.1 in subsequent analyses. From Bonder et al.[58], we downloaded meQTLs from whole blood and used variant-probe pairs with FDR <0.05 in subsequent analyses. From Nicodemus-Johnson et al.[59], we downloaded cis-eQTLs and meQTLs from the lung and used variant-gene pairs with FDR <0.1 in subsequent analyses. From Momozawa et al.[60], we downloaded cis-eQTLs from blood cell types CD14, CD15, CD19, CD4, and CD8 used and variant-gene pairs with a weighted correlation ≥0.6 in analyses. For Supplementary Table 14 priority genes, protein-coding information was downloaded from the UCSC Table Browser, with the 'transcriptClass' field from the 'wgEncodeGencodeAttrsV24lift37' table.

Promoter-capture Hi-C gene prioritization. To assess spatial promoter interactions in the discovery set, we performed a capture Hi-C gene prioritization (CHIGP) as described in Javierre et al.[61] and https://github.com/ollyburren/CHIGP/ by using the following recommended settings and data sources: 0.1 cM recombination blocks, 1KGP EUR reference population, coding markers from the GRCh37 Ensembl assembly, and the CHICAGO-generated[62] promoter-capture Hi-C peak matrix data from 17 human primary blood cell types supplied in the original paper. The resulting protein-coding prioritized genes (gene score >0.5) were used in the downstream network analysis, from cell types 'fetal thymus', 'total CD4 T cells', 'activated total CD4 T cells', 'nonactivated total CD4 T cells', 'naive CD4 T cells', 'total CD8 T cells', 'naive CD8 T cells', 'total B cells', 'naive B cells', 'endothelial precursors', 'macrophages M0', 'macrophages M1', 'macrophages M2', 'monocytes', and 'neutrophils'.

Gene-set-overrepresentation analysis of known and replicating novel loci. All high-confidence gene symbols from eQTL and meQTL sources, PCHiC, as well as genes (models extended 110 kb upstream, and 40 kb downstream) within each r^2-based locus definition from known and replicating novel loci were input into the pathway-based set over-representation analysis module in ConsensusPathDB (CPDB) database and tools[63] with 229 of 277 gene identifiers translated. In addition, these same symbols were used for ingenuity pathway analysis (IPA; http://www.ingenuity.com/; a curated database of the relationships among genes obtained from published articles, and genetic and expression data repositories), to identify biological pathways common to genes. IPA determines whether the associated genes are significantly enriched in a specific biological function or network by assessing direct interactions. We assigned significance if the right-tailed Fisher's exact test $P < 0.05$. eQTL/meQTL, PCHiC and locus–gene intersections were visualized with the UpSetR package (v1.3.2)[64].

Tissue overrepresentation. To assay the enrichment of variants associated with allergic rhinitis in tissue-specific gene-expression sets, we used the DEPICT enrichment method[65], using a P-value threshold of 1×10^{-5} and standard settings.

Enrichment of regulatory regions. To assay the enrichment of variants associated with allergic rhinitis in regions of open chromatin and specific histone marks, we used the GWAS analysis of regulatory or functional information enrichment with LD correction (GARFIELD v. 1) method[66]. In essence, GARFIELD performs greedy pruning of GWAS markers (LD $r^2 > 0.1$) and then annotates them according to functional information overlap. Next, it quantifies fold enrichment at various GWAS significance cutoffs and assesses them by permutation testing while adjusting for the minor allele frequency, distance to nearest transcription start site and number of LD proxies ($r^2 > 0.8$). GARFIELD was run with 10,000,000 permutations and otherwise default settings.

Population-attributable-risk fraction. Population-attributable-risk fractions (PARFs) were estimated from B58C, a general-population sample with participant ages 44–45 years also contributing to the discovery stage. The genetic-risk score was calculated by applying the pooled per-allele coefficients (ln(OR) values) from the allergic rhinitis discovery set to the number of higher-risk alleles of each of the 41 established loci (known genome-wide-significant and novel replicated loci), with one SNP per locus. Because there were no individuals observed with zero higher-risk alleles, the prevalence of sensitization for individuals in the lowest decile of the genetic-risk score distribution was used to derive PARF estimates on the assumption that this 10% of the population was unexposed. This method has the advantage that it does not predict beyond the bounds of the data, but its results are conservative. The PARF was then derived (with 95% confidence intervals) by expressing the difference between the observed prevalence and the predicted (unexposed) prevalence as a percentage of the observed prevalence. PARFs were estimated by using the 41 allergic rhinitis–associated loci in relation to allergic rhinitis, allergic sensitization, and nonallergic rhinitis.

Protein-network and drug interactions. To analyze protein–protein–drug interaction networks, STRING (V10)[67] was used. Protein-network data (9606.protein.links.v10.txt.gz) and protein alias data (9606.protein.aliases.v10.txt) files were downloaded from the string db website (http://string-db.org/). GWAS hits stratified on 'all', 'blood', and 'lung' were converted to Ensembl protein IDs by using the protein alias data. The interactors were subsequently identified by using the link data at a 'high-confidence cutoff of >0.7', as described in the STRING FAQ. The interactor Ensembl protein IDs were then converted to UniProt gene names, and both hits and interactors were then analyzed for interactions with FDA approved drugs by using the ChEMBL Database v. 22 (ref. [68]) API via Python (v. 2.7.12). Finally, stratified networks consisting of GWAS hits connected to interactors and drugs connected to both GWAS hits and interactors were visualized with GGraph (v. 1.0.0), iGraph (v. 1.0.1), and TidyVerse (v. 1.1.1) in R (v. 3.3.2).

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Genome-wide results, excluding 23andMe, are available from the corresponding author on request and will be available on LDhub after July 2018. The full GWAS summary statistics for the 23andMe discovery dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please contact D.A.H. (dhinds@23andme.com) for more information and to apply for access to the 23andMe data.

References

36. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
37. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinforma.* **11**, 288 (2010).
38. Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
39. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
40. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
41. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
42. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
43. UK10K Consortium. et al. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
44. Wellcome Trust Case Control Consortium. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
45. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
46. Zheng, X. et al. HIBAG—HLA genotype imputation with attribute bagging. *Pharm. J.* **14**, 192–200 (2014).
47. Jia, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* **8**, e64683 (2013).

48. Tian, C. et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* **8**, 599 (2017).
49. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
50. Vita, R. et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43**, D405–D412 (2015).
51. Sethi, D. K., Gordo, S., Schubert, D. A. & Wucherpfennig, K. W. Crossreactivity of a human autoimmune TCR is dominated by a single TCR loop. *Nat. Commun.* **4**, 2623 (2013).
52. Rückert, C. et al. Conformational dimorphism of self-peptides and molecular mimicry in a disease-associated HLA-B27 subtype. *J. Biol. Chem.* **281**, 2306–2316 (2006).
53. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
54. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
55. Kettunen, J. et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat. Genet.* **44**, 269–276 (2012).
56. Lucas, A. O. Surveillance of communicable diseases in tropical Africa. *Int. J. Epidemiol.* **5**, 39–43 (1976).
57. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
58. Fairfax, B. P. et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
59. Nicodemus-Johnson, J. et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight* **1**, e90151 (2016).
60. Momozawa, Y. et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 2427 (2018).
61. 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).
62. Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* 17, (127 (2016).
63. Kamburov, A. et al. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* **39**, D712–D717 (2011).
64. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
65. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
66. Iotchkova, V. et al. GARFIELD: GWAS analysis of regulatory or functional information enrichment with LD correction. https://www.biorxiv.org/content/early/2016/11/07/085738 (2016).
67. Szklarczyk, D. et al. STRINGv10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
68. Bento, A. P. et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, D1083–D1090 (2014).

nature research

Corresponding Author: Klaus Bønnelykke

Date: Apr 28, 2018

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, 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.

   We aimed to include all available data-sets on allergic rhinitis worldwide. In this study we achieved a sample size that was several times larger than previous GWAS- studies, which was sufficent for providing genome-wide significant findings.

2. Data exclusions

   Describe any data exclusions.

   Other than the preliminary analyses/experimental optimization, no data were excluded.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   No experimental studies were performed.

4. Randomization

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

   No experimental studies were performed. Cases and controls for genome-wide association were analyzed based upon predefined criteria.

5. Blinding

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

   No blinding was performed and not considered relevant in this meta-Genome-wide association study.

   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 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 summary of the descriptive 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.	This is described in details in the methods section. In short, for AR, AS, and NAR, meta-analysis for the discovery phase was conducted using GWAMA, while meta-analysis of replication candidates from the AR discovery phase was carried out using R version 3.4.0, and the meta package version 4.8-2 with an inverse variance weighted fixed-effect model. Additional software include EasyQC (v9.2), STRING (v10), ChEMBL Database (v22), GGraph (v1.0.0), iGraph (v1.0.1), TidyVerse (v1.1.1), LDHUB platform (v1.3.1), GENCODE (v19 and v24, Genome-wide Complex Trait Analysis (GCTA) (v. 1.26.0), Plink (v1.90b3.42), PyMOL (v1.8.2.1), LDSC (v.1.0), LDHUB platform (v1.3.1), GTEx (V6p), UpSetR package (v1.3.2), GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD) method (v1), Python (v2.7.12), HIBAG (v1.2.3).

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

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.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No eukaryotic cell lines were used.

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.	No animals were used.

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.	Relevant information on research participants in the individual studies is provided in the supplementary material.