Human Molecular Genetics, 2019, Vol. 28, No. 23 4022–4041

doi: 10.1093/hmg/ddz175
Advance Access Publication Date: 30 July 2019
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema

Åsa Johansson[†], Mathias Rask-Andersen, Torgny Karlsson and Weronica E. Ek[*,‡]

[1]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

[*]To whom correspondence should be addressed at: Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, PO Box 815, 75108, Uppsala, Sweden. Tel: +46703519004; Fax: +46184714931; Email: Weronica.ek@igp.uu.se

Abstract

Even though heritability estimates suggest that the risk of asthma, hay fever and eczema is largely due to genetic factors, previous studies have not explained a large part of the genetics behind these diseases. In this genome-wide association study, we include 346 545 Caucasians from the UK Biobank to identify novel loci for asthma, hay fever and eczema and replicate novel loci in three independent cohorts. We further investigate if associated lead single nucleotide polymorphisms (SNPs) have a significantly larger effect for one disease compared to the other diseases, to highlight possible disease-specific effects. We identified 141 loci, of which 41 are novel, to be associated ($P \leq 3 \times 10^{-8}$) with asthma, hay fever or eczema, analyzed separately or as disease phenotypes that includes the presence of different combinations of these diseases. The largest number of loci was associated with the combined phenotype (asthma/hay fever/eczema). However, as many as 20 loci had a significantly larger effect on hay fever/eczema only compared to their effects on asthma, while 26 loci exhibited larger effects on asthma compared with their effects on hay fever/eczema. At four of the novel loci, *TNFRSF8*, *MYRF*, *TSPAN8*, and *BHMG1*, the lead SNPs were in Linkage Disequilibrium (LD) (>0.8) with potentially casual missense variants. Our study shows that a large amount of the genetic contribution is shared between the diseases. Nonetheless, a number of SNPs have a significantly larger effect on one of the phenotypes, suggesting that part of the genetic contribution is more phenotype specific.

Introduction

Asthma, hay fever and eczema are common complex immunological diseases affecting many people worldwide (1). The prevalence for these diseases varies among populations and has an underlying architecture that includes both environmental and genetic risk factors (1). Comorbidity between asthma, hay fever and eczema is common, and previous genome-wide association (GWA) studies have, apart from identifying a large number of genetic variants associated with risk of disease (2–8), also found evidence of a genetic overlap between the diseases (6, 9).

Family and twin studies have estimated that the contribution of genetic factors, i.e. the heritability for asthma (1, 10, 11) to be

[†]Åsa Johansson, http://orcid.org/0000-0002-2915-4498
[‡]Weronica E. Ek, http://orcid.org/0000-0003-2194-496X
Received: March 7, 2019. **Revised:** May 29, 2019. **Accepted:** June 3, 2019

35%–95%, for hay fever (1, 11) to be 33%–91%, and for eczema (12) to be as high as 90%. A recent large study estimated the single nucleotide polymorphism (SNP)-based heritability, the heritability that can be attributed the genetic variation captured by SNPs in a genome-wide association (GWA) study, to be 15% for asthma, 22% for hay fever, and 9% eczema (6). The same study performed a GWA study that included the first release of UK Biobank ($N = 138\,354$) and analyzed asthma, hay fever and eczema as a combined phenotype and identified 99 significantly associated loci (6). Many of the identified target genes were predicted to influence the function of immune cells, and only six loci were identified to have disease-specific effects (6). Many previous GWA studies for asthma, hay fever and eczema have been conducted in different cohorts that were subsequently meta-analyzed with the purpose of increasing statistical power (2, 4, 6–8, 13).

The aim with this study was to explain a larger part of the genetic background of self-reported asthma, hay fever, and eczema as well as identify possible novel disease-specific effects, compared to what has been done in previous studies. We investigated the genetic background of self-reported asthma, hay fever, and eczema combined to a single phenotype using data from UK Biobank, which is a more homogenous population compared to previous large GWA studies on the same phenotypes (6). We also included the whole UK Biobank ($N = 346\,545$) cohort, in comparison to previous studies (6) that only included the first UK Biobank release ($N = 138\,354$). We also used three independent cohorts to replicate novel loci. Due to the larger sample size, we also had power enough to investigate if associated lead SNPs had a significantly larger effect for one disease phenotype compared to the other phenotypes and thereby highlight possible disease-specific effects. Associated SNPs were also functionally annotated to assess likely causal mechanisms.

Although the phenotypes in the UK Biobank are self-reported, the questions are well defined and identical for all participants.

Results

Association analysis

The UK Biobank includes 502 682 participants, of which 443 068 are Caucasians. The disease prevalence in the Caucasian participants was 11.7% for asthma and 23.2% for hay fever and/or eczema (combined). As many as 45.8% of the asthmatic participants had reported having hay fever and/or eczema, and 23.0% of the hay fever and/or eczema participants had reported being diagnosed with asthma (Fig. 2). We conducted the GWA study using 346 545 unrelated Caucasians (Table 1), who passed the quality control (QC) for the second UK Biobank genetic data release and had no ambiguities with regards to disease status. After QC, 15 688 218 genetic variants were included in the analyses (see Materials and Methods). QC and the final number of included participants in respective analyses are summarized in Materials and Methods and in Table 1. We did not identify any statistically significant associations located on the X-chromosome.

GWA study for self-reported asthma

After QC, 41 926 self-reported asthma cases (independent on hay fever/eczema status) and 239 773 controls were included in the GWA analysis. We identified 75 risk loci located >1 Mb apart and containing at least one significantly associated genetic variant ($P \leq 3 \times 10^{-8}$ after adjusting for Linkage Disequilibrium (LD) score [LDSC] intercept of 1.065), that were associated with self-reported asthma, of which 15 loci were found to be novel asthma loci not previously identified in a GWA study (Table 2; Manhattan plot, Fig. 1; Supplementary Material, Tables S1 and S2; quantile-quantile [QQ] plot, Supplementary Material, Fig. **S1**). Using approximate conditional analysis (14), we identified 116 independent significant associations within these 75 loci (Supplementary Material, Table S1). The strongest associations for asthma were found within the human leukocyte antigen (HLA) locus on chromosome 6 ($P = 2.06 \times 10^{-100}$), including 14 independent genetic variants. Several genes within this region have previously been reported to be associated with asthma (i.e. *HLA-DQB1*, *HLA-G* and *HLA-DRB1*) (1, 3, 13). Among the novel asthma loci, some have previously been associated with other similar phenotypes (Supplementary Material, Table S1). For example, *SDK1*, previously annotated to the nearby *CARD11* gene, has been reported to be associated with atopic dermatitis (15), but this is the first time that the *SDK1* locus has been identified in a GWA study for asthma. Five of the 15 novel lead SNPs were replicated in GABRIEL ($P < 0.05$; Table 2). However, most of the SNPs were not possible to investigate in the replication cohort, due to a low number of overlapping SNPs between the cohorts.

Annotation of asthma-associated SNPs

Associated SNPs were further functionally annotated to assess likely causal mechanisms (see Materials and Methods). Overlap with Genotype-Tissue Expression (GTEx) expression quantitative loci (eQTLs) was found for 15 of the 75 asthma loci. Of these, four eQTLs (*EEFSEC*, *ADAM19*, *HHEX* and *TMEM258*) overlapped with the novel loci, where increased expression of *TMEM258* in cell transformed fibroblasts appears to lower the risk for developing asthma (Table 2; Supplementary Material, Tables S1 and S3). In contrast, increased expression of *EEFSEC* in lung tissue seems to increase the risk for asthma (Supplementary Material, Table S3). However, the *EFFSEC* loci did only reach a probability for pairwise colocalization (PPA) of 0.67 between the Genome Wide Association Study (GWAS) hit and the eQTL. Increased expression of *ADAM19* in whole blood and *HHEX* in cell transformed fibroblasts appears to increase the risk of asthma (Supplementary Material, Table S3). Nineteen probable causal missense variants could be observed within the 75 significant GWA loci, of which four are missense variants for the 15 novel loci (Supplementary Material, Table S4). The latter is located within *TNFRSF8*, *MYRF*, *TSPAN8*, and *BHMG1*. The association at *TNFRSF8* was represented by only one genetic variant, rs2230624 (Supplementary Material, Fig. **S2**). This SNP is a missense variant in two transcripts for *TNFRSF8* and causes a cysteine to a tyrosine substitution which was predicted as "probably damaging" by PolyPhen (16) (PolyPhen score, 0.751–0.921) and had a "deleterious" Sorting intolerant from tolerant (SIFT) score (17) of 0. The lead SNP at the *MYRF* locus, rs174535, is a missense variant in five transcripts for *MYRF*. Rs174535 causes a serine to arginine substitution and was predicted to be "probably damaging" by PolyPhen (16) (PolyPhen score, 0.961–1) and had a "deleterious" SIFT score (17) of 0.04–0.07. However, rs174535 is also in LD with the most significant eQTL for *TMEM258* in cell transformed fibroblasts. The lead SNP in the *BHMG1* locus, rs11671106, is a missense variant for *BHMG1* and was predicted as "probably damaging" by PolyPhen (16) (PolyPhen score, 0.94) and had a "deleterious" SIFT score (17) of 0.01. The lead SNP at the *TSPAN8*

Table 1. Baseline characteristics of Caucasian participants in UK Biobank

	Asthma	Hay fever/ eczema	Asthma/hay fever/eczema combined[b]	Hay fever	Eczema	Controls[c]
N Caucasians[a] (prior to QC)	51 645	102 862	130 865	22 919	9578	294 477
N total included after QC[d]	41 934	84 050	106 772	18 915	7884	239 773
N (%) males after QC	21 730 (51.8%)	42 639 (50.7%)	55 124 (51.6%)	8692 (46.0%)	3365 (42.7%)	138 666 (57.8%)
Age year span (mean)	38–70 (56.1)	39–72 (55.4)	38–72 (55.7)	40–77 (55.0)	40–70 (55.0)	39–73 (57.2)
Townsend deprivation index range (mean)	−6.3 to 10.6 (−1.3)	−6.3 to 10.6 (−1.6)	−6.3 to 10.6 (−1.5)	−6.3 to 10.4 (−1.7)	−6.3 to 9.6 (−1.9)	−6.3 to 10.9 (−1.5)
% Ever smoked (N yes/N no)	60.4% (31 040/203 90)	58.5% (59 983/42 616)	59.3% (77 335/53 121)	56.8% (12 986/9879)	60.9% (5815/3737)	60.4% (177 212/116 275)

The total number of Caucasians is N = 443 068.
Asthma or hay fever and/or eczema combined as one phenotype.
The same controls were used in all analyses.
We removed first- and second-degree relatives, using kinship data (estimated genetic relationship, >0.044), and participants with sex discordance and high heterozygosity, as well as participants with more than 5% missing SNP genotypes, resulting in 346 545 individuals after QC.

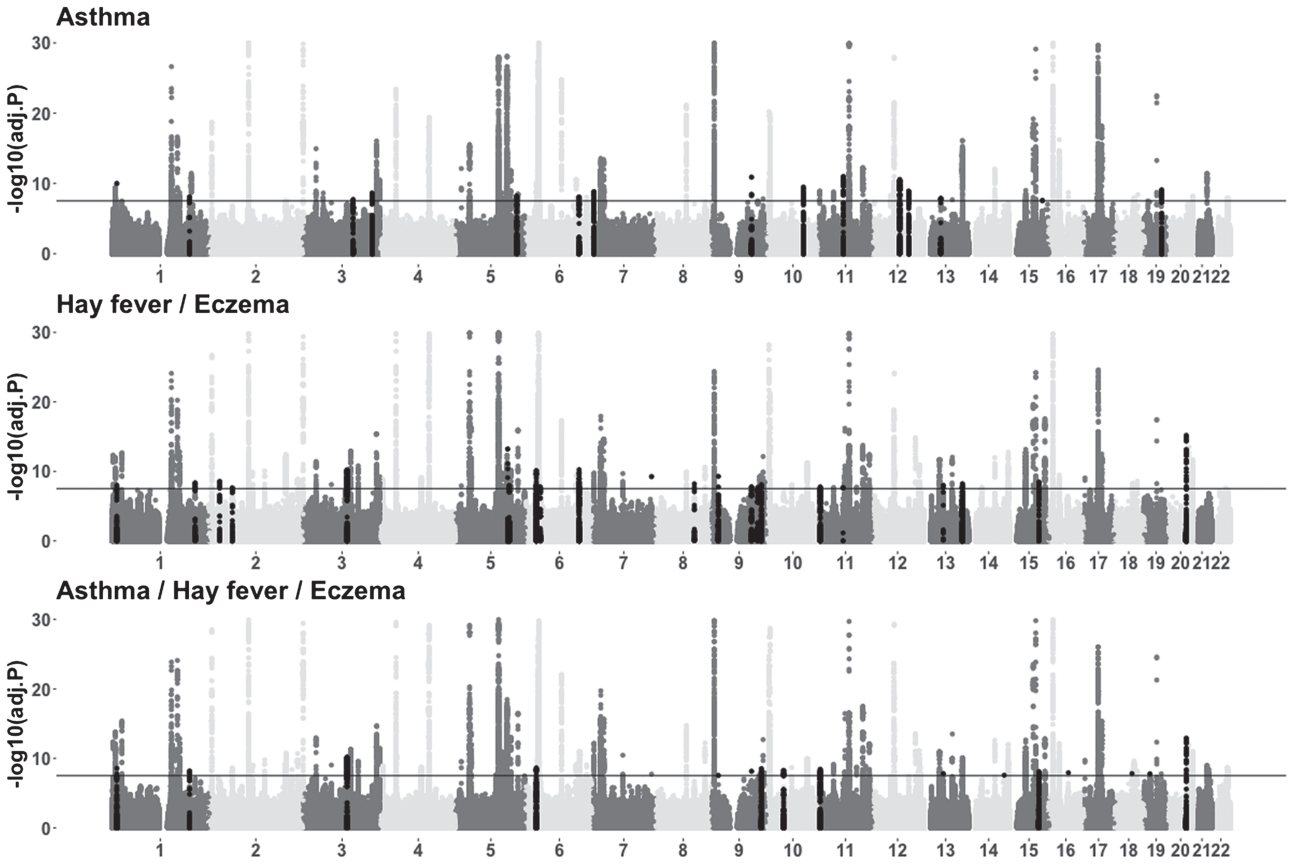

Figure 1. Manhattan plots for asthma, for hay fever and/or eczema, and for asthma and/or hay fever and/or eczema (combined) for autosomal chromosomes. The black horizontal line indicates the genome-wide threshold (3×10^{-8}). The black regions represent novel loci found in this study.

locus, rs11178649, was in complete LD with rs3763978 ($R^2 = 1$), a missense variant in three transcripts for *TSPAN8*, which causes a glycine to alanine substitution which was predicted as "probably damaging" by PolyPhen (16) (PolyPhen score, 0.989) and had a "deleterious" SIFT score(17) of 0.03.

GWA study for self-reported hay fever/eczema

After QC, 84 034 self-reported hay fever and/or eczema cases that were combined as a single phenotype were included in the analysis. We identified 109 loci to be associated

Table 2. Summary results for the 15 novel loci significantly associated with self-reported asthma in UK Biobank ($P \leq 3 \times 10^{-8}$) with replication in the GABRIEL cohort

Lead SNP	Locus[a] chr:start-end (kbp)	N snps (total[b]/ independent[c])	MAF[d]	Minor/major allele	OR (95% CI) for minor allele	P	Likely target gene(s) (if annotated to an eQTL, tissue type is presented)	GABRIEL P (OR) [95% CI] for effective allele (minor/effective allele)
rs2230624	1:12 175–12 175	1/1	0.02	A/G	0.80 (0.75–0.86)	1.01×10^{-10}	*TNFRSF8*[e]	No proxy
rs2296618	1:198 656–198 670	5/1	0.13	G/A	0.93 (0.91–0.96)	8.03×10^{-9}	*PTPRC*[f]	No proxy
rs10934853	3:127 886–128 075	3/1	0.27	A/C	0.95 (0.94–0.97)	2.20×10^{-8}	*EEFSEC*[g] Cell transformed fibroblasts	0.006 (1.06) [1.02–1.11] A/C
rs6778937	3:176 708–176 868	28/1	0.28	C/T	0.95 (0.93–0.97)	2.54×10^{-9}	*TBL1XR1*[f]	No proxy
rs11466773	5:156 930–156 988	7/1	0.06	T/C	1.09 (1.06–1.13)	6.32×10^{-9}	*ADAM19*[g] Whole blood	No proxy
rs2614266	6:135 691–135 818	6/1	0.44	A/T	1.05 (1.03–1.06)	8.90×10^{-9}	*AHI1*[f]	No proxy
rs10215232	7:3062–3153	12/1	0.12	G/C	0.93 (0.91–0.95)	1.53×10^{-9}	*SDK1*[f]	rs9986945[h], $R^2 = 1.0$[i] 0.03 (1.07) [1.01–1.14] T/G
rs41283642	9:101 915–101 989	3/1	0.03	T/C	0.86 (0.82–0.90)	1.27×10^{-11}	*TGFBR1*[f]	No proxy
rs2497318	10:9434–9444	24/1	0.45	T/C	0.95 (0.94–0.97)	3.21×10^{-10}	*HHEX*[g] Cell transformed fibroblasts	Rs10882091[h], $R^2 = 0.81$[i] 0.84 (1.00) [0.97–1.05] C/T
rs174535	11:61 543–61 623	49/1	0.35	C/T	0.95 (0.93–0.96)	1.02×10^{-11}	*MYRF*[e], *TMEM258*[g] Cell transformed fibroblasts	rs102275[h], $R^2 = 1.0$[i] 0.045 (1.04) [1.00–1.09] C/T
rs11178649	12:71 409–71 585	103/1	0.41	T/G	0.95 (0.93–0.96)	2.68×10^{-11}	*TSPAN8*[e]	rs1051334[h] $R^2 = 1.0$[i] 0.04 (0.95) [0.92–0.99] (G/T)

Continued.

Table 2. Continued

Lead SNP	Locus[a] chr:start-end (kbp)	N snps (total[b]/ independent[c])	MAF[d]	Minor/major allele	OR (95% CI) for minor allele	P	Likely target gene(s) (if annotated to an eQTL, tissue type is presented)	GABRIEL P (OR) [95% CI] for effective allele (minor/effective allele)
rs4761592	12:94 556–94 604	17/1	0.15	T/C	0.93 (0.92–0.95)	1.27×10^{-9}	*PLXNC1*[f]	rs3912394[h], $R^2 = 0.85$[i] 0.021 (1.07) [1.01–1.13] T/C
rs9316059	13:44 475–44 490	5/1	0.20	T/A	1.06 (1.04–1.08)	1.35×10^{-8}	*LINC00284* [f]	rs3764147[h], $R^2 = 0.93$[i] 0.58 (1.01) [0.97–1.06] A/G
rs4842921	15:84 556–84 556	1/1	0.39	A/G	0.96 (0.94–0.97)	2.63×10^{-8}	*ADAMTSL3* [f]	No proxy
rs11671106	19:46 219–46 370	28/1	0.35	T/C	0.95 (0.94–0.97)	8.29×10^{-10}	*BHMG1* [e]	rs7250497[h], $R^2 = 0.97$[i] 0.05 (1.04) [1.00–1.09] G/A

More details can be found in Tables S1–S3 and S4.
[a]Defined as SNPs located <1 Mb apart containing at least one significantly associated genetic variant at $P \leq 3 \times 10^{-8}$.
[b]Total number of SNPs with $P \leq 3 \times 10^{-8}$ within loci.
[c]Total number of independent associations within the locus, based on conditional analysis (14).
[d]Minor allele frequency.
[e]Lead SNP is in LD ($R^2 > 0.8$) with a missense variant.
[f]Gene(s) closest to the lead SNP.
[g]Lead SNP is in LD ($R^2 > 0.8$) with the lead eQTL SNP.
[h]Proxy SNP in LD (>0.8) with lead SNP.
[i]R^2 between lead SNP in UK Biobank and proxy SNP in GABRIEL.

($P \leq 3 \times 10^{-8}$, LDSC intercept = 1.079) with self-reported hay fever/eczema, and 22 of these were novel (Table 3; Manhattan plot, Figure 1; Supplementary Material, Tables S5 and S6; QQ plot, Supplementary Material, Fig. S3). The strongest association was observed for the lead SNP rs5743604 ($P = 7.5 \times 10^{-72}$) located within *TLR1*. This SNP has previously been associated with allergic disease (6, 18). Using conditional analysis, we identified 154 independent significant associations within these 109 loci (Table 3; Supplementary Material, Table S5). Moreover, two of our lead SNPs (rs4845604 and rs9986945, mapped to *RORC* and *SDK1*), observed within previously known loci, were in low LD ($R^2 \leq 0.05$) with the previously reported genetic variants, indicating that they represent novel variants within or close to known loci (Supplementary Material, Table S5). The *UBAC2* locus has previously been reported to be associated with asthma (9), but this is the time that the *UBAC2* locus is reported to be associated with hay fever and/or eczema. We replicated six of the novel lead SNPs in an independent eczema GWA study, the EAGLE consortium ($P \leq 0.05$) (Table 3). However, all SNPs that did not replicate in EAGLE were not significant in UK Biobank either, when analyzing eczema separately (Table 3).

Annotation of hay fever/eczema SNPs. For 11 of the 109 hay fever/eczema associated loci, the lead SNP was in LD with the lead SNP for GTEx eQTLs (Table 3; Supplementary Material, Table S3) and 14 overlapped with possible causal missense variants in genes, including *IL6R*, *IL7R*, *IL13* and *SMAD4* (Supplementary Material, Table S7).

GWA studies for hay fever and eczema analyzed separately. Hay fever and eczema could not be separated for most of the participants, since they had primarily answered yes or no on whether they had either hay fever or eczema. However, to investigate hay fever and eczema individually, we also analyzed hay fever ($N = 18\,915$ hay fever cases) and eczema ($N = 7884$ eczema cases) separately in a smaller subset of UK Biobank participants (Manhattan plot, Supplementary Material, Fig. S4; QQ plots, Supplementary Material, Figs S5 and S6). A total of 27 and 18 loci were identified for hay fever and eczema, respectively. One novel hay fever and one novel eczema locus, which has not been reported in previous GWA studies and that were not significantly associated in the combined hay fever/eczema analysis, were detected when analyzing hay fever and eczema separately (Supplementary Material, Tables S8–S11). The lead SNP, rs12920150 ($P = 1.02 \times 10^{-9}$), at the hay fever locus is located close to *CBLN1* and the lead SNP, rs2485363 ($P = 1.20 \times 10^{-8}$), at the eczema locus is located downstream of *TAGAP*. This novel eczema locus was nominally replicated using the summary statistics from the GWA study on eczema in the EAGLE consortium ($P = 0.018$, odds ratio [OR] = 1.05 [95% confidence interval (CI), 1.02–1.92]). Another locus that was not detected when analyzing hay fever/eczema combined was detected when analyzing eczema separately. The lead SNP for this locus, rs676387 ($P = 2.26 \times 10^{-10}$), is located within *HSD17B1* (Supplementary Material, Tables S10 and S11). This region has previously been reported to be associated with allergic disease (6) and overlap with an eQTL for *TUBG2* in skin, where a decreased expression of *TUBG2* seems to lower the risk for eczema (Supplementary Material, Table S3).

GWA study for asthma/hay fever/eczema (combined as a single phenotype)

For the combined analysis of asthma and/or hay fever and/or eczema ($N = 106\,752$ cases), we identified 110 significant loci (LDSC intercept = 1.081), and 16 of these were novel GWA loci that have not been significantly associated with either asthma, hay fever or eczema in previous GWA studies (Table 4; Manhattan plot, Figure 1; Supplementary Material, Tables S12 and S13; QQ plot, Fig. S7). However, 12 of these 16 novel loci were detected when analyzing asthma and hay fever/eczema separately, while the remaining four novel loci were only found when analyzing asthma, hay fever, and/or eczema together as a single phenotype. Using conditional analysis, we identified 164 independent associations within these 110 loci (Table 4; Supplementary Material, Table S12). The most significant SNP, rs72823641 ($P = 1.14 \times 10^{-78}$), was located within *IL1RL1* and was also significantly associated with asthma and hay fever/eczema when these phenotypes were analyzed separately ($P = 4.09 \times 10^{-61}$ and $P = 9.64 \times 10^{-64}$) (Supplementary Material, Tables S12 and S13). This region has previously been associated with allergic diseases (6) (Supplementary Material, Table S14). We also identified five lead SNPs for the combined phenotype asthma and/or hay fever and/or eczema within previously known loci, which were found to be in low LD ($R^2 \leq 0.05$) with previously reported genetic variants, indicating that they represent novel variants within known loci. These five lead SNPs mapped to *LPP*, *IL31*, *LINC00393*, *CCR7* and *NFATC* (Supplementary Material, Tables S12 and S14). Eight of the 16 novel loci were replicated with the same direction of effect in 23andMe ($P \leq 0.05$) (Table 4).

Annotation of Asthma/hay fever/eczema (combined as a single phenotype) SNPs. For 16 of the 110 asthma and/or hay fever and/or eczema associated loci, the lead SNPs overlapped with a lead SNP for an eQTL (Table 4; Supplementary Material, Tables S3 and S12). Among the novel loci, one overlapped with an eQTL for *HIST1H2BD* in whole blood ($P = 1.11 \times 10^{-16}$). A decreased level of *HIST1H2BD* seems to increase the risk of this combined phenotype (Supplementary Material, Table S3). However, this locus did not replicate in 23andMe ($P = 0.87$). Probable causal missense variants could be observed at 17 out of 110 significant loci, and one of these was observed at one of the novel loci located within *TNFRSF8* and was also identified in the asthma analysis above and replicated in 23andMe ($P = 0.046$) (Table 4 and Supplementary Material, Table S15).

SNP-based heritability

To quantify the SNP-based heritability for asthma and hay fever/eczema, we used LD score (LDSC) regression analysis (19). These analyses included the same cases and controls as for the association analysis (see Materials and Methods). The SNP-based heritability was estimated to be $21 \pm 3\%$ for asthma and $16 \pm 3\%$ for hay fever/eczema (Table 5). Our significant loci, which were located ≥ 1 Mb apart and contained at least one significantly associated genetic variant at $P \leq 3 \times 10^{-8}$, explained 4.2% of the heritability for asthma and 3.6% of the heritability for hay fever/eczema (Table 5).

Identification of phenotype-specific loci (SNP)

In our GWA studies, we included all individuals reporting either asthma (for the asthma GWA study) or hay fever/eczema (for the hay fever/eczema GWA study) as cases, independent on if they reported having the other disease phenotype (i.e. asthma cases could have reported having asthma and hay fever/eczema or only asthma). To investigate possible phenotype-specific SNPs,

Table 3. Summary results for the 22 novel loci significantly associated with self-reported hay fever and/or eczema in UK Biobank ($P \leq 3 \times 10^{-8}$) with replication in the EAGLE cohort

Lead SNP	Locus[a] chr:start-stop (kbp)	N snps (total[b]/ independent[c])	MAF[d]	Minor/major allele	P (OR)[e] [95% CI] hay fever/ eczema	P (OR)[e] [95% CI] hay fever	P (OR)[e] [95% CI] eczema	Likely target gene (if annotated to an eQTL, tissue type is presented)	EAGLE P (OR) [95% CI] estimated for the minor allele (minor/major allele)
rs1201113	1:12 100–12 147	3/1	0.12	A/G	1.05×10^{-8} (0.95) [0.93–0.97]	4.67×10^{-3} (0.95) [0.92–0.99]	1.69×10^{-2} (0.94) [0.90–0.99]	*TNFRSF*[f]	0.18 (1.04) [0.98–1.10] A/G
rs906363	1:212 858–212 877	6/1	0.15	C/T	4.44×10^{-9} (1.05) [1.03–1.07]	8.56×10^{-4} (0.97) [1.02–1.08]	2.54×10^{-4} (1.09) [1.04–1.13]	*BATF3*[f]	1.52×10^{-6} (1.19) [1.07–1.17] C/T
rs13405815	2:28 623–28 644	9/1	0.46	T/C	3.01×10^{-9} (0.97) [0.95–0.98]	4.04×10^{-5} (0.96) [0.94–0.98]	3.70×10^{-3} (0.95) [0.92–0.98]	*RP11-373D23.3*[g] Skin, not sun exposed	Proxy rs6547850[h], $R^2 = 1.0$[i] 0.0014 (0.95) [0.92–0.98] T/C T/G
rs10185028	2:61 112–61 161	6/1	0.23	G/A	2.74×10^{-8} (1.04) [1.03–1.05]	9.47×10^{-4} (1.04) [1.02–1.07]	1.17×10^{-2} (1.05) [1.01–1.09]	*REL*[f]	0.0007 (1.08) [1.03–1.12] G/A
rs11717778	3:112 526–112 693	144/1	0.34	A/G	6.04×20^{-11} (0.96) [0.95–0.97]	2.67×10^{-4} (0.96) [0.94–0.98]	5.38×10^{-6} (0.92) [0.89–0.96]	*CD200R1L*[f]	0.0007 (0.94) (0.91–0.98) A/G
rs62379371	5:133 439–133 639	4/1	0.05	A/G	6.08×10^{-14} (0.90) [0.87–0.92]	7.83×10^{-6} (0.89) [0.84–0.94]	4.18×10^{-5} (0.85) [0.78–0.92]	*TCF7*[f]	0.62 (0.95) [0.77–1.17] A/G
rs13185930	5:137 461–137 605	10/1	0.25	A/G	1.12×10^{-8} (1.04) [1.03–1.05]	3.19×10^{-2} (1.03) [1.00–1.05]	4.64×10^{-1} (1.01) [0.98–1.05]	*GFRA3*[f]	0.42 (1.02) [0.98–1.06] A/G
rs2229768	6:25 823–26 239	24/1	0.24	C/T	8.20×10^{-11} (1.05) [1.03–1.06]	8.99×10^{-6} (1.06) [1.03–1.08]	2.74×10^{-1} (1.02) [0.98–1.06]	*U91328.19*[g] Cell transformed fibroblasts	0.80 (1.00) [0.96–1.04] C/T
rs1998266	6:36 349–36 380	5/1	0.14	T/C	1.70×10^{-8} (0.95) [0.94–0.97]	1.23×10^{-3} (0.95) [0.92–0.98]	1.68×10^{-3} (0.93) [0.88–0.97]	*ETV7*[f]	0.08 (0.95) [0.91–1.00] T/C
rs2746438	6:135 624–135 950	36/1	0.44	T/A	5.70×10^{-11} (1.04) [1.03–1.05]	9.24×10^{-7} (1.05) [1.03–1.08]	2.73×10^{-5} (1.07) [1.04–1.11]	*AHI1*[f]	0.21 (1.02) [0.99–1.06] T/A
rs3918226	7:150 690–150 690	1/1	0.08	T/C	5.65×10^{-10} (0.93) [0.91–0.95]	1.54×10^{-7} (0.90) [0.86–0.93]	6.09×10^{-1} (0.98) [0.93–1.05]	*NOS3*[f]	0.33 (1.04) [0.97–1.11] T/C
rs6986151	8:101 514–101 519	2/1	0.20	C/T	6.19×10^{-9} (1.04) [1.03–1.06]	5.39×10^{-6} (1.06) [1.04–1.09]	2.44×10^{-3} (1.06) [1.02–1.11]	*ANKRD46* [f]	0.31 (1.00) [0.95–1.05] C/T

Continued.

Table 3. Continued

Lead SNP	Locus[a] chr:start-stop (kbp)	N snps (total[b]/ independent[c])	MAF[d]	Minor/major allele	P (OR)[e] [95% CI] hay fever/ eczema	P (OR)[e] [95% CI] hay fever	P (OR)[e] [95% CI] eczema	Likely target gene (if annotated to an eQTL, tissue type is presented)	EAGLE P (OR) [95% CI] estimated for the minor allele (minor/major allele)
rs1330303	9:16 715–16 756	2/1	0.35	T/C	5.21×10^{-10} (0.96) [0.95–0.97]	2.95×10^{-8} (0.94) [0.92–0.96]	3.90×10^{-1} (0.99) [0.95–1.02]	*BNC2*[f]	0.11 (1.01) [0.99–1.06] T/C
rs4743311	9:101 790–101 820	3/1	0.25	G/A	1.69×10^{-8} (1.04) [1.03–1.05]	7.93×10^{-6} (1.06) [1.03–1.08]	1.27×10^{-1} (1.03) [0.99–1.07]	*COL15A1*[f]	0.88 (1.00) [0.96–1.04] G/A
rs12343737	9:117 804–117 834	2/1	0.10	T/C	2.09×10^{-8} (0.95) [0.93–0.95]	1.28×10^{-6} (0.92) [0.89–0.95]	9.59×10^{-2} (0.96) [0.91–1.01]	*TNC*[f]	0.25 (0.97) [0.92–1.03] T/C
rs10986320	9:127 022–127 095	13/1	0.37	C/G	8.81×10^{-9} (1.04) [1.02–1.05]	4.13×10^{-5} (1.05) [1.02–1.07]	7.32×10^{-2} (1.03) [1.00–1.07]	*NEK6*[f]	0.20 (0.98) [0.95–1.01] C/G
rs4076542	11:2237–2296	4/1	0.37	A/G	1.74×10^{-8} (1.04) [1.02–1.05]	4.40×10^{-3} (1.04) [1.02–1.06]	7.78×10^{-1} (1.02) [0.98–1.05]	*ASCL2*[f]	0.70 (0.99) [0.96–1.03] A/G
rs4939490	11:60 793–60 793	4/1	0.39	G/C	2.15×10^{-8} (0.97) [0.96–0.98]	2.15×10^{-3} (0.97) [0.95–0.99]	9.76×10^{-2} (0.97) [0.94–1.01]	*CD6*[f]	0.25 (1.02) [0.99–1.06] G/C
rs3116590	13:50 808–50 811	2/1	0.21	G/A	1.00×10^{-8} (1.04) [1.03–1.06]	1.11×10^{-2} (1.03) [1.01–1.06]	7.24×10^{-3} (1.05) [1.01–1.10]	*DLEU1*[f]	0.38 (1.02) [0.98–1.06] G/A
rs4771332	13:99 839–100 070	7/1	0.31	T/C	6.21×10^{-9} (0.96) [0.95–0.98]	6.09×10^{-3} (0.97) [0.95–0.99]	1.29×10^{-1} (0.97) [0.94–1.01]	*UBAC2*	0.40 (0.98) [0.95–1.02] C/T
rs4381563	15:75 399–75 448	7/1	0.34	A/T	3.37×10^{-9} (0.96) [0.95–0.98]	2.98×10^{-3} (0.97) [0.95–0.99]	3.37×10^{-2} (0.96) [0.93–1.00]	*PPCDC*[f]	0.89 (1.00) [0.97–1.04] A/T
rs6066184	20:45 232–45 716	33/1	0.26	G/C	6.55×10^{-16} (0.95) [0.93–0.96]	5.25×10^{-11} (0.92) [0.90–0.94]	2.03×10^{-2} (0.96) [0.92–0.99]	*EYA2*[f]	0.02 (1.05) [1.01–1.09] G/C

More details can be found in Tables S3 and S5–S7.
[a]Defined as SNPs located <1 Mb apart containing at least one significantly associated genetic variant at $P \leq 3 \times 10^{-8}$.
[b]Total number of SNPs with $P \leq 3 \times 10^{-8}$ within loci.
[c]Total number of independent associations within the locus, based on conditional analysis (14).
[d]Minor allele frequency.
[e]OR for minor allele.
[f]Gene(s) closest to the lead SNP.
[g]Lead SNP is in LD ($R^2 > 0.8$) with the lead eQTL SNP.
[h]Proxy SNP in LD (>0.8) with lead SNP.
[i]R^2 between lead SNP in UK Biobank and proxy SNP in EAGLE.

Table 4. Summary results for the 16 novel loci significantly associated with self-reported asthma and/or hay fever and/or eczema (combined) in UK Biobank ($P \leq 3 \times 10^{-8}$) with replication in 23andMe

Lead SNP	Locus[a] chr:start-stop (kbp)	*N* snps (total[b]/ independent[c])	MAF[d]	Minor/major allele	*P* (OR[e]) [95% CI] combined	*P* (OR[e]) [95% CI] asthma	*P* (OR[e]) [95% CI] hay fever/ eczema	Likely target gene (if annotated to an eQTL, tissue type is presented)	23andMe *P* (OR) [95% CI] estimated for the minor allele (minor/major allele)
rs2230624	1:12 080–12 175	2/2	0.02	A/G	2.64×10^{-9} (0.87) [0.84–0.91]	1.01×10^{-10} (0.80) [0.75–0.86]	1.99×10^{-7} (0.88) [0.84–0.92]	*TNFRSF8*[f]	0.046 (0.99) [0.97–1.01] (A/G)
rs7410883	1:198 640–198 670	5/1	0.11	C/T	7.00×10^{-9} (0.95) [0.93–0.97]	2.26×10^{-8} (0.93) [0.91–0.95]	1.78×10^{-6} (0.96) [0.94–0.97]	*PTPRC*[g]	0.018 (0.97) [0.94–0.99] C/T
rs9816107	3:112 526–112 693	160/1	0.34	A/C	6.80×10^{-11} (0.96)	1.51×10^{-6} (0.96) [0.95–0.97]	8.39×10^{-11} (0.96) [0.95–0.97]	*CD200R1L*[g]	0.41 (0.99) [0.97–1.01] A/C
rs62379371	5:133 439–133 639	4/1	0.05	A/G	1.13×10^{-13} (0.91) [0.89–0.93]	1.23×10^{-7} (0.91) [0.88–0.94]	6.08×10^{-14} (0.90) [0.87–0.92]	*VDAC1*[g]	No pro × y
rs9379828	6:26 038–26 184	13/1	0.37	G/C	2.71×10^{-9} (1.03) [1.02–1.05]	1.07×10^{-10} (1.05) [1.04–1.07]	1.42×10^{-6} (1.03) [1.02–1.04]	*HIST1H2BD*[h] Whole blood	0.87 (1.00) [0.98–1.03] G/C
rs1330303	9:16 715–16 715	1/1	0.35	T/C	2.84×10^{-8} (0.97) [0.96–0.98]	0.016 (0.98) [0.97–1.00]	5.21×10^{-10} (0.96) [0.95–0.97]	*BNC2*[g]	0.59 (0.99) [0.98–1.01] T/C
rs41283642	9:101 915–101 915	1/1	0.03	T/C	7.03×10^{-9} (0.92) [0.89–0.94]	1.27×10^{-11} (0.86) [0.82–0.90]	6.42×10^{-6} (0.93) [0.90–0.96]	*TGFBR1*[g]	0.036 (0.93) [0.87–0.99] T/C
rs3758212	9:127 002–127 178	14/1	0.37	T/C	2.99×10^{-9} (1.03) [1.02–1.05]	1.80×10^{-5} (1.04) [1.02–1.05]	2.18×10^{-8} (1.04) [1.02–1.05]	*NEK6*[g]	4.7×10^{-4} (1.03) [1.01–1.05] T/C
rs2505504	**10:43 728–43 763**	**17/1**	**0.29**	**A/G**	**5.45×10^{-9} (1.04) [1.02–1.05]**	**1.96×10^{-6} (1.04) [1.02–1.06]**	**4.61×10^{-7} (1.03) [1.02–1.05]**	***RASGEF1A***[g]	0.049 (1.02) [1.00–1.04] A/G

Continued.

Table 4. Continued

Lead SNP	Locus[a] chr:start-stop (kbp)	N snps (total[b]/ independent[c])	MAF[d]	Minor/major allele	P (OR[e]) [95% CI] combined	P (OR[e]) [95% CI] asthma	P (OR[e]) [95% CI] hay fever/ eczema	Likely target gene (if annotated to an eQTL, tissue type is presented)	23andMe P (OR) [95% CI] estimated for the minor allele (minor/major allele)
rs7114923	11:2237–2305	21/1	0.37	T/C	3.74×10^{-9} (1.03) [1.02–1.05]	1.20×10^{-5} (1.04) [1.02–1.05]	2.65×10^{-8} (1.04) [1.02–1.05]	*ASCL2*[g]	0.11 (1.02) [1.00–1.03] T/C
rs3116590	13:50 808–50 808	1/1	0.21	G/A	1.52×10^{-8} (1.04) [1.03–1.05]	0.0064 (1.03) [1.01–1.05]	1.00×10^{-8} (1.04) [1.03–1.06]	*DLEU1*[g]	0.0038 (1.03) [1.01–1.06] G/A
rs61975764	**14:93 014–93 014**	**1/1**	**0.47**	**A/G**	**2.65×10^{-8} (1.03) [1.02–1.04]**	**1.46×10^{-7} (1.04) [1.03–1.06]**	**7.84×10^{-7} (1.03) [1.02–1.04]**	***RIN3***[g]	**0.13 (0.92) [0.83–1.02] A/G**
rs4381563	15:75 275–75 448	19/1	0.33	A/T	7.81×10^{-9} (0.97) [0.96–0.98]	0.00027 (0.97) [0.96–0.99]	3.37×10^{-9} (0.96) [0.95–0.97]	*PPCDC*[g]	0.38 (0.99) [0.97–1.01] A/T
rs12956924	**18:46 451–46 451**	**1/1**	**0.31**	**A/G**	**1.52×10^{-8} (1.03) [1.02–1.05]**	**2.64×10^{-5} (1.04) [1.02–1.05]**	**2.09×10^{-7} (1.03) [1.02–1.05]**	***SMAD7***[g]	**0.023 (1.02) [1.00–1.05] A/G**
rs10419921	**19:16 412–16 412**	**1/1**	**0.30**	**T/C**	**1.67×10^{-8} (1.03) [1.02–1.05]**	**5.62×10^{-6} (1.04) [1.02–1.06]**	**1.09×10^{-6} (1.03) [1.02–1.04]**	***KLF2***[g]	**0.099 (1.02) [1.00–1.04] T/C**
rs6066184	20:45 228–45 716	26/2	0.26	G/C	1.19×10^{-13} (0.95) [0.94–0.97]	4.27×10^{-6} (0.96) [0.94–0.98]	6.55×10^{-16} (0.95) [0.93–0.96]	*EYA2*[d]	0.036 (0.98) [0.96–1.00] G/C

Loci found only when analyzing all three diseases as one phenotype are marked as bold. More details can be found in Tables S3, S12, S13 and S15.
[a]Defined as SNPs located <1 Mb apart containing at least one significantly associated genetic variant at $P \leq 3 \times 10^{8}$.
[b]Total number of SNPs with $P \leq 3 \times 10^{-8}$ within loci.
[c]Total number of independent associations within the locus, based on conditional analysis (14).
[d]Minor allele frequency.
[e]OR for minor allele.
[f]Lead SNP is in LD ($R^2 > 0.8$) with a missense variant.
[g]Gene(s) closest to the lead SNP.
[h]Lead SNP is in LD ($R^2 > 0.8$) with the lead eQTL SNP.

Table 5. SNP-based heritability in UK Biobank for asthma and hay fever/eczema (combined as one phenotype) estimated with LDSC (19)

Phenotype	N cases	N controls	Prevalence used in LDSC[a]		(A) All SNPs		(B) Without significant loci		h2 explained by significant loci
			Population	Sample	h2	SE	h2	SE	Absolute terms: (A)–(B)
Asthma	41 926	239 751	0.117	0.117	0.210	0.017	0.168	0.010	0.042
Hay fever/ eczema	84 034	239 751	0.232	0.232	0.160	0.010	0.124	0.007	0.036

[a]LDSC requires values for population and sample prevalence when estimating SNP heritability. In these analyses, we used the prevalence in the UK Biobank cohort (Table 1) as the sample prevalence. However, due to the cross-sectional design of the UK Biobank, we used the same values as population prevalence.

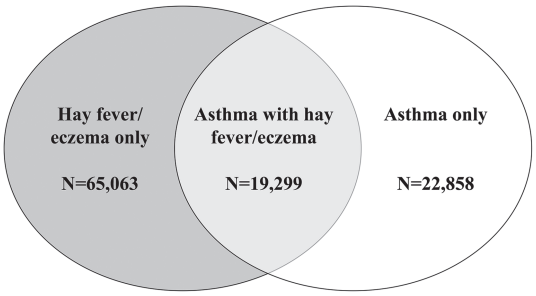

Figure 2. Comorbidity between asthma and hay fever/eczema.

we performed polytomous (multinomial) logistic regression to identify whether the effect of a locus (lead SNP) was significantly (FDR ≤ 0.05) larger for one disease phenotype as compared to another. These effects can therefore be considered as being disease/phenotype specific. To conduct these analyses, we used four non-overlapping groups: 1) asthma cases without hay fever/eczema ($N = 22\,858$), 2) hay fever/eczema cases without asthma ($N = 65\,063$), 3) asthma cases with hay fever/eczema (only including $N = 19\,299$ participants that had reported asthma in combination with hay fever or eczema), and 4) controls without asthma, hay fever and eczema ($N = 240\,817$) (Fig. 2). Hay fever and eczema were not separated in this analysis due to the small sample size (Table 1). Groups were compared in a pairwise fashion (Supplementary Material, Table S16).

A total of 154 lead SNPs (see Materials and Methods for a description of the selection of lead SNPs), representing the 138 different loci, identified in the GWA study for asthma, hay fever/eczema, or for asthma/hay fever/eczema, were included in polytomous logistic regression analyses. To illustrate the specificity in the Venn diagrams (Fig. 3), each SNP was assigned to an area that represents either a phenotype-specific effect (significantly larger in one group of cases) or a shared effect (no significant difference between the two groups of cases).

In the comparison of asthma only, i.e. without hay fever and eczema, to hay fever/eczema only, i.e. without asthma, 26 loci/SNPs were specific for asthma only, i.e. had a significantly higher OR for asthma compared to for hay fever/eczema, while 20 loci/SNPs were specific to hay fever/eczema only. A major part of the loci/SNPs, 103, showed no significant difference in effect between these disease phenotypes. (Fig. 3; Supplementary Material, Table S16).

When comparing subjects with asthma and hay fever/eczema to subjects with asthma, 53 loci/SNPs were specific for asthma with hay fever/eczema. No SNP was specific for the asthma-only group (Fig. 3; Supplementary Material, Table S16). For the remaining 96 loci/SNPs, there was no significant difference in effect between subjects with asthma only and subjects with asthma as well as hay fever/eczema.

Finally, when comparing cases of hay fever/eczema only with cases of hay fever/eczema combined with asthma (Fig. 3; Supplementary Material, Table S16), 64 loci/SNPs had significantly larger effect in the group with hay fever/eczema combined with asthma. No locus had a larger effect in the hay fever/eczema without asthma group. As many as 83 loci/SNPs had no detectable difference in effect between these two disease phenotypes.

For some loci, multiple, possibly independent ($R^2 \leq 0.8$) SNPs were included in the analyses. For most of the analyses, such independent SNPs within the same locus showed the same phenotype specificity or lack of specificity. That is, all independent SNPs within one locus belong to the same area in the Venn diagram (Fig. 3). However, for a number of loci, the effect for the different independent SNPs showed different phenotype specificity. This resulted in 149 independent loci/SNPs when comparing the asthma-only group to the hay fever/eczema-only group and when comparing subjects with asthma and hay fever/eczema to subjects with asthma only (Fig. 3). For the last group, when comparing cases of hay fever/eczema only with cases of hay fever/eczema combined with asthma, 147 independent loci/SNPs were identified and included in the Venn diagram (Fig. 3). For example, two uncorrelated SNPs ($R^2 < 0.05$) were found to be located within the same intron of *IL2RA*: rs61839660, which was associated with hay fever/eczema, and rs12722547, which was associated with asthma in the GWA study. The rs61839660 SNP has a significantly larger effect in both hay fever/eczema only and hay fever/eczema with asthma compared to asthma only, but no difference in effect between hay fever/eczema with or without asthma. The effect of rs12722547 was instead significantly larger in the hay fever/eczema with asthma group compared to the hay fever/eczema without asthma group. Rs12722547 also exhibited a trend (nominal $P = 0.05$) toward having a larger effect in asthma only compared to hay fever/eczema only (Supplementary Material, Table S16).

Discussion

In this large GWA study, including 346 545 unrelated Caucasian participants from UK Biobank, we identified 141 unique loci that are associated with self-reported asthma, hay fever, and/or eczema when these traits are analyzed separately or together as combined phenotypes. In comparison with previous studies based on UK Biobank and similar disease phenotypes, our study has several strengths and presents additional results. Out of all identified loci, as many as 41 are novel to our study and have not been reported to be associated with the same disease phenotype previously. Compared to Ferreira *et al.* (6) and Zhu *et al.* (9), who

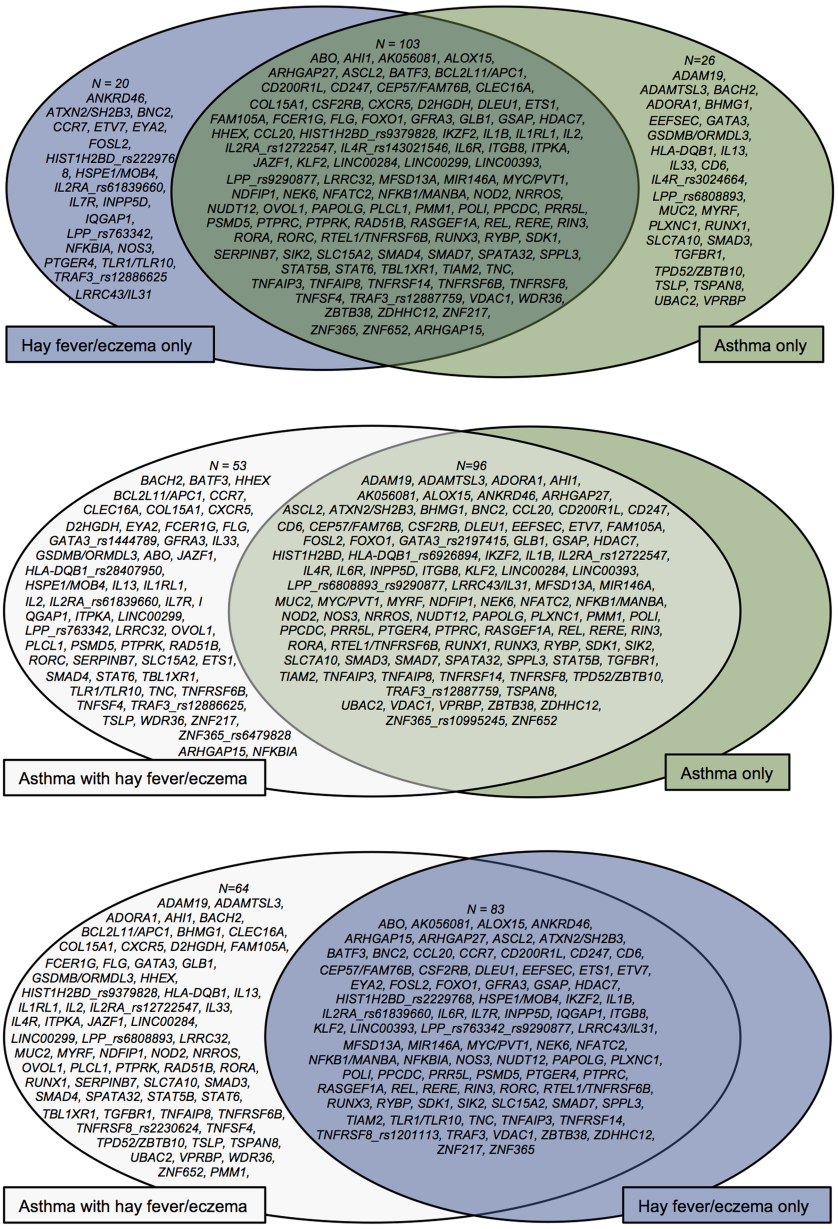

Figure 3. Venn diagram showing the phenotype specificity of the GWA loci, based on the results from the polytomous logistic regression analyses. The Venn diagram show loci (SNPs) that are specific (significantly larger effect) to or shared between (no significant difference in effects) between two non-overlapping groups of cases. The name of each locus is denoted by the most likely gene(s). At some of the loci (e.g. *IL2RA*, *LPP*, and *IL4R*), more than one independent ($R^2 < 0.8$) lead SNP has been analyzed in the polytomous logistic regression. If those showed different specificity pattern, they have been included twice in the figure with the name of respective lead SNP(s) also included in the locus name. *P*-values and estimates for the genes can be found in Table S16 where the area number 1 (green in the figure) indicates specificity for the asthma only; area number 2 (blue in the figure), specificity for hay fever/eczema only; and area number 3 (white in the figure), specificity for asthma with hay fever/eczema (significantly larger estimate in the asthma with hay fever/eczema group of cases).

only included the first release of UK Biobank, we included the full UK Biobank cohort. We also present five different GWA studies for five different phenotypes and further had the strength to identify a number of possible phenotype-specific effects that had not been discussed previously.

The largest number of loci was associated with combined phenotype (asthma and/or hay fever and/or eczema), most likely due to the larger sample size of this group. However, this is in agreement with a shared genetic contribution between diseases, as has been shown in Ferreira *et al.* (6) and Zhu *et al.* (9). With this combined phenotype (asthma and/or hay fever and/or eczema), we identified four novel loci that were not found for asthma or hay fever/eczema when analyzed separately. Three of these loci appear to be highly relevant to the pathogenies of all three diseases: *SMAD7*, *KLF2* and *RIN3*. The variant at the *SMAD7* locus is located within an intron of *SMAD7* and was replicated in 23andMe ($P = 0.023$). This gene has previously been associated with inflammatory bowel disease (18), colorectal cancer (20), and hemoglobin concentration (21). The variant at the *KLF2* locus is located in the 5' UTR of *KLF2*. This gene plays a role in processes during development including epithelial integrity, inflammation, and T-cell viability. Previous studies have found

associations between this locus and lymphocyte percentage of white cells, neutrophil percentage of white cells, white blood cell counts, monocyte percentage of white cells, and eosinophil percentage of granulocytes (21). The variant at the *RIN3* locus is located within an intron of *RIN3* and is also associated with *RIN3* expression. This gene has previously been associated with myeloid white cell count, eosinophil basophil counts (21), and chronic obstructive pulmonary disease (22). However, the *KLF2* and *RIN3* loci did not replicate in 23andMe (Table 4). It is worth noting that some of our novel loci have previously been associated with a related phenotype (Supplementary Material, Tables S1, S5, S8, S10 and S12). For example, some of the novel asthma loci has previously been associated with Immunoglobulin E (IgE) levels, eosinophil counts or dermatitis and some of the novel hay fever/eczema loci with IgE levels or eosinophil counts.

For four of the novel loci that were also replicated in independent cohorts: near *TNFRSF8*, *MYRF*, *TSPAN8*, and *BHMG1*, the lead SNP was in LD with potentially deleterious missense variants. The lead genetic variant at the *TNFRSF8* locus, rs2230624, which is associated with asthma as well as the combined asthma/hay fever/eczema phenotype, is a potentially causal missense variant that causes a cysteine to a tyrosine substitution in the TNFRSF8 protein. This protein, which is also referred to as *CD30*, is a receptor that is expressed on activated T and B cells and has been shown in clinical studies to have a role in the development of allergic asthma (23). To the best of our knowledge, this is the first time that this locus has been associated with asthma and allergy in a GWA study. The lead SNP at the *MYRF* asthma locus, rs174535, is a missense variant within the myelin regulatory factor protein (*MYRF*) that causes a serine to arginine substitution near the end of the protein. This gene lies within the fatty acid desaturase (*FADS*) cluster on a fatty acid synthesis-associated haplotype (24). Variants on this haplotype are also strongly associated with expression of *FADS1* and *FADS2*, two genes that are involved in the desaturation of polyunsaturated fatty acids in the biosynthesis of long chain polyunsaturated fatty acids (LC-PUFAs). One of these variants has previously been shown to modulate the effect of breast-feeding on asthma (25); another has been associated with increased risk of inflammation (26). Reduced capacity to desaturase omega-6 LC-PUFAs due to *FADS* polymorphisms has been shown to be nominally associated with reduced risk for development of eczema, potentially due to a pathogenic role of omega-6 LC-PUFAs in development of allergy (27).

For eight of the 41 novel GWA loci, the lead SNP was in LD (>0.8) with an eQTL. We could see a positive correlation between expression of *TUBG2* (in skin), *HHEX* (in cell transformed fibroblasts), *EEFSEC* (in lung and cell transformed fibroblasts), and *ADAM19* (in whole blood) and risk of disease, as well as a negative correlation for *TMEM258* and *HIST1H2BD*. Decreased expression of *TMEM258* in cell transformed fibroblasts was associated with increased risk of asthma. In transgenic experiments in mice, it has been shown that a lower expression of *TMEM258* leads to severe intestinal inflammation (28), which agrees with our results. We further evaluated pairwise colocalization between eQTLs from GTEx and our GWA signals using the program coloc (29). As many as 83% of the overlapping eQTLs also showed a pairwise colocalization (≥0.75 probability) (Supplementary Material, Table S3). A possible limitation of this analysis is that it relied solely on the GTEx database. Additional sources of information on eQTLs may increase the total number of eQTLs that are associated with asthma, hay fever and eczema.

For 16 loci that were associated with asthma, 20 loci associated with hay fever/eczema and for 21 loci associated with asthma/hay fever/eczema, we identified multiple independently associated variants. This indicates that several of the asthma-, hay fever-, and eczema-associated loci represent multiple independent disease-associated variants. As an example, the *FLG* locus contains three independent asthma and asthma/hay fever/eczema-associated variants. This gene has previously been shown to contain loss-of-function mutations that are causal for skin barrier deficiency and strongly predispose to both eczema and asthma (30). The four most prevalent European *FLG* mutations are c.2282del4, p.R501X, p.$R^2$447X, and p.S3247X (30). An additional example is the *HLA* region whose association with immune diseases is particularly complex and which has previously been suggested to include several independent regulatory factors (31). In our analyses, we identify as many as 21 independent associations within this locus.

As highlighted by this study, as well as previous studies (2–7, 32), many disease-associated loci overlap between asthma, hay fever and eczema. However, several loci were only significantly associated with only one of the investigated phenotypes. By testing for association with hay fever and eczema separately in a smaller set of participants, we were able to resolve some of these signals. Interestingly, one of the strongest associations for hay fever/eczema ($P = 7.96 \times 10^{-25}$), found within the *FLG* locus, was more significantly associated with eczema when this phenotype was analyzed separately ($P = 9.75 \times 10^{-65}$). In contrast, this variant was not associated with hay fever when hay fever was analyzed separately. It was, however, associated with asthma ($P = 2.37 \times 10^{-27}$). This is in agreement with the previous GWA study by Ferreira *et al*., where a SNP at the *FLG* locus was shown to be specifically associated with eczema (6). However, a different study has shown that mutations within the *FLG* locus are associated with eczema starting in the first year of life, and that these mutations are associated with a later development of both asthma and hay fever (33). This is an example of the typical progression of allergic diseases that often begin early in life, which is commonly referred to as the atopic march (33–35). When analyzing hay fever separately, we identified one novel locus near *CBLN1*. Studies on transgenic mice have shown that knock-out of *CBLN1* mimics loss-of-function mutations that occur in the orphan glutamate receptor, *GRID2* (36). Autoantibodies against glutamate receptors are involved in the development of autoimmune disease (37). One novel locus was also identified and replicated when analyzing eczema separately, downstream of *TAGAP*. This locus has previously been associated with celiac disease (38) and multiple sclerosis (39).

We further investigated our novel asthma, hay fever/eczema and eczema loci for replication in three independent cohorts. We were only able to replicate six SNPs out of the 15 identified to be novel for asthma with the summary statistics from the GABRIEL asthma consortium. The GABRIEL GWA study only included 582 802 SNPs genotyped with the Illumina Human610 quad array, and therefore, a large number of SNPs did not overlap between our studies. However, for most of the loci where we identified the same SNP or a proxy in LD ($R^2 \geq 0.8$), we did find a nominal replication ($P \leq 0.05$) (Table 2). The GABRIEL study was based on childhood asthma, while the UK Biobank asthma phenotype is based on adult asthma, and these two disease phenotypes may therefore have some different underlying genetic effects. It is also important to remember that a lack in replication may also be due to a lower power to detect associated SNPs in GABRIEL due to a smaller sample size. Five lead SNPs identified for the combined analysis hay fever/eczema were replicated in

the EAGLE study (Table 3). The lack of replication is most probably due to differences between phenotypes. While the EAGLE study only included eczema cases, our study also included hay fever. All SNPs that did not replicate in EAGLE were neither statistically significant when analyzing eczema independently in UK Biobank (Table 3). With 23andMe, we had a large power to replicate the combined analysis asthma/hay fever/eczema but only replicated eight SNPs of the 16 novel loci. However, one locus did not have a proxy in LD with our lead variant and could therefore not be included in the replication. Also, The *CD200R1L* locus did not replicate in 23andMe but replicated in the EAGLE cohort, which suggest that the effect is larger in eczema compared to asthma and hay fever. We cannot rule out that the loci that did not replicate might be false positives in our analysis. However, it is also possible that the lack of replication between cohorts is due to the discrepancy in the disease definitions.

Out of all asthma and/or allergic disease-associated loci that have been reported to the GWAS catalog as of December 2, 2018, the majority ($N = 108$) were nominally replicated in our study ($P \leq 0.05$; Supplementary Material, Table S14). Twelve associations were not possible to test due to lack of data, i.e. neither the reported SNP nor any SNP in LD with the reported SNP was presented in our data (Supplementary Material, Table S14). Asthma, hay fever, and eczema are known to be heterogeneous diseases in which environmental factors play an important role (1). Genetic variants associated with asthma, hay fever and eczema are likely to be population specific (40). It is therefore possible that population-specific variants are not detected in our study. Many of the previous associations that were not replicated in our study have been identified in studies that have used a somewhat different phenotype (41, 42), populations of different ancestry (15, 43) or small sample sizes (<10 000) (43, 44). Research findings from studies on smaller cohorts are more likely to be false positives, especially when no replication of primary findings has been performed, and are thereby less likely to represent true causative mechanisms (45) (for more information, see Supplementary Material, Table S14). A recent GWA study by Zhu *et al*. (9), which was also conducted on the UK Biobank cohort, however using a different combination of allergies as a phenotype, reported seven novel allergy-associated loci, five of which were replicated in our study. These loci where not available in the GWAS catalog at the time of writing this article and are therefore not included in Supplementary Material, Table S14. The two loci that did not replicate in our study were mapped to *ALG9* on chromosome 11(rs659529) and to *EVI5* on chromosome 1 (rs12743520).

In previous GWA studies for asthma, the disease phenotype commonly contained other disease phenotypes as well, e.g. participants with asthma commonly also report hay fever/eczema. In contrast, our polytomous logistic regression approach allowed for identification of genetic variants with differing effects between the different sub-phenotypes. These effects can therefore be considered as being disease/phenotype specific. This was achieved by subdividing the participants in four non-overlapping groups depending on asthma and hay fever/eczema status. The SNPs that were included in these analyses were selected from our main GWA analyses, but not including the two SNPs identified for the hay fever and eczema phenotypes analyzed separately since we did not have power (large enough sample size) enough to include hay fever and eczema separately in these analyses. This means that a locus that was defined as specific for asthma only has already been associated with any of the combined phenotypes and/or with asthma, independent of hay fever/eczema status. The association for such variants may have been due to comorbidity between asthma and the other diseases, e.g. a larger fraction of asthma cases in the hay fever/eczema group compared to the controls, or that the effect of the asthma-only specific variants was only partly diluted by being combined with other disease phenotypes. A large number of loci exhibited differential effects between hay fever/eczema only and asthma only. As many as 20 loci had a significantly larger effect on hay fever/eczema only compared to their effects on asthma while 26 loci exhibited larger effects on asthma compared with their effects on hay fever/eczema (Fig. 3). Among the loci that were specific for asthma only, we find *ADAM19* and *ADAMTSL3* which are proteins with multiple biological roles within the cell and believed to be important in a number of diseases, including asthma (46). Among the loci that were specific for hay fever/eczema, we find the toll-like receptor loci, *TLR1/TLR10*, which also showed a larger effect on hay fever compared to asthma only in the Ferreira *et al*. study (6). Most associated variants at this locus are located within the promoter region of *TLR1*, which encodes the toll-like receptor 1. This protein constitutes a component of the innate immune response to microbial pathogens (47). Several loci that overlap between asthma only and hay fever/eczema only were annotated to genes related to tumor necrosis factor (TNF) function, such as *TNFAIP3*, *TNFAIP8*, *TNFRSF11A*, *TNFRSF14*, *TNFRSF6B*, *TNFRSF8*, and *TNFSF4*. These proteins are mainly expressed in immune cells and regulate immune response and inflammation as well as proliferation, apoptosis and embryogenesis (48).

The largest number of phenotype-specific loci was observed for the group of cases with asthma and hay fever/eczema (Fig. 3; Supplementary Material, Table S16), a group of cases that has not been included in similar analyses in previous studies (6). This is a group of participants with an allergic disease in combination with asthma, which could to some degree represent participants with allergic asthma. The number of phenotype-specific loci is considerably larger in our study compared to previous studies that have performed similar analyses, such as the study by Ferreira *et al*.(6), which only identified six disease-specific loci. This is not surprising since our analyses included larger sample sizes: $N = 65\,063$, $N = 22\,858$, and $N = 19\,299$ compared to $N = 33\,305$, $N = 12\,268$, and $N = 6276$ in the study by Ferreira *et al*. (6) for the three sub-groups included in the analyses of disease-specific effects. In addition, since only genome-wide significant SNPs were taken forward to the polytomous logistic regression analyses, we used the False Discovery Rate by Benjamini-Hochberg to adjust for multiple testing. This increases the power to pinpoint as many positive findings as possible, still with a small false-discovery rate (5% in our case), compared to the more conservative Bonferroni method used in the previous study by Ferreira *et al*. (6). The previous study also separated hay fever and eczema and compared the three groups hay fever only, eczema only and asthma only. Since different subgroups of cases were analyzed in our study, our results do not disagree with that of Ferreira *et al*. (6) that found six disease-specific SNPs: near *FLG*, *RPTN-HRNR* (close to *FLG*), *IL2RA*, *IL1RL2- IL8R1*, *WDR36-CAMK4* and *GSDMB*, where five of them were significantly different between hay fever and eczema.

The SNP-based heritability was estimated to be $21 \pm 3\%$ for asthma and $16 \pm 3\%$ for hay fever/eczema. These percentages represent the portion of heritability that can be captured by the common genetic variants that were included in the GWA study. The SNP-based heritability for asthma has previously been estimated at 15%, which, at face value, is slightly lower than the estimate from our study. The observed difference may be a result

of a difference in disease definition and/or simply a consequence of statistical uncertainty, due to a much smaller sample size in the previous study (6). In comparison to the high estimates for the heritability (33%–95%) from family and twin studies (1, 10, 12), our results suggest that a major contribution to the genetic risk for asthma, hay fever and eczema might not be identified in studies using common genetic variants or need cohorts with even larger sample sizes. However, heritability estimates from family and twin studies have been suggested to be overestimated (49–51) due to the fact that these estimates often are based on simplistic models that ignore shared environmental factors. Our estimate might also be lower due to the presence of disease-associated rare variants that are not captured by the SNP-based heritability estimate.

A possible limitation of the present study is the self-reported phenotypes, which might lead to a recall bias and misclassification. Another limitation is that the UK Biobank cohort traits are not independent since there are shared cases between asthma, hay fever and eczema and completely shared controls. However, findings presented in this article apply to a single large population of individuals of similar age. Population stratification was also controlled for by filtering for Caucasian participants, including ancestry-derived principal components and adjusting for the LDSC intercept in our analyses. Participants of the UK Biobank are also more likely to be exposed to more similar environmental factors, compared to the participants of previous meta-analyses that utilize a large number of smaller cohorts from different countries and age-groups.

Analyzing hay fever and eczema as a combined phenotype is another limitation in our study, which prohibits identification of hay fever- and eczema-specific SNPs. We therefore refer to SNPs as phenotype-specific rather than disease-specific in the polytomous logistic regression analyses. However, both hay fever and eczema are IgE-mediated hypersensitivities and therefore probably share similar physiology (52).

In summary, we describe 15 novel loci for asthma, 22 novel loci for hay fever and/or eczema, and an additional four novel loci were found when analyzing asthma, hay fever and eczema together. Two novel loci were also identified when analyzing hay fever and eczema separately. Pinpointing candidate genes for common diseases are important for tailor-made studies that want to prioritize candidate genes for developing novel therapeutic strategies. This study further highlights a large amount of shared genetic contribution to these diseases, indicating that the comorbidity between asthma, hay fever and eczema is partly due to shared genetic factors. However, we also show that a number of SNPs have a significantly larger effect on one of the phenotypes, suggesting that part of the genetic contribution is phenotype specific.

Materials and Methods

Study population

The UK Biobank includes 502 682 participants recruited from all across the UK. Participants were between 37 and 73 years old at time of recruitment between 2006 and 2010. Most participants visited the center once, but some individuals visited the center at up to three times. Participants answered questions about self-reported medical conditions, diet, and lifestyle factors. A total of 820 967 genotyped SNPs and up to 90 million imputed variants are available for most participants. We included 346 545 unrelated Caucasians (see selection of participants and sample QC below) with genotypes from the second UK Biobank genotype release (Table 1).

UK Biobank was given ethical approval by the North West Multicentre Research Ethics Committee (REC reference 11/NW/0382), the National Information Governance Board for Health and Social Care and the Community Health Index Advisory Group. UK Biobank holds a generic Research Tissue Bank approval granted by the National Research Ethics Service (http://www.hra.nhs.uk/) that lets applicants conduct research on UK Biobank data without obtaining ethical approvals for each separate project. Access to UK Biobank genetic and phenotypic data was given through the UK Biobank Resource under application number 15479. All participants provided signed consent to participate in UK Biobank.

Disease phenotypes: asthma and hay fever/eczema

Self-reported asthma as well as self-reported hay fever and/or eczema (combined) were assessed using the UK Biobank touch screen question number (data field 6152), which asked the participants the following question: “Has a doctor ever told you that you have had any of the following conditions? (You can select more than one answer): 1) asthma and 2) hay fever, allergic rhinitis or eczema, 3) none of the above or 4), prefer not to answer.” Because hay fever and eczema diagnosis could not be separated, we called this variable hay fever/eczema (i.e. participants reported hay fever and/or eczema). All participants were also invited to participate in an interview. At first, nurses (trained UK Biobank staff member) confirmed with each participant that the information they provided on the screen or questionnaire was correct if they had answered that a doctor had told them they had one or more of the following diseases: heart attack, angina, stroke, high blood pressure, blood clot in leg, blood clot in lung, emphysema/chronic bronchitis, asthma, or diabetes. Due to the confirmation of asthma cases, the overlap in asthma variables between the touch-screen questionnaire and verbal interview was very high. For asthma, only 622 individuals were removed due to conflicting answers between the touch screen and verbal interview. Using a drop-down menu, the nurses could also add other diagnoses. These diagnoses (UK Biobank data field 20002) were used to define hay fever and eczema cases separately. However, the disease prevalence in this variable appears to be largely underreported, as many individuals reported hay fever or eczema in the touch-screen questionnaire but did not report hay fever or eczema during the interview. For this reason, the touch-screen data variables hay fever/eczema, with a much larger sample size (Table 1) compared to hay fever and eczema separately, was used as one of the primary phenotypes analyzed in this study. For hay fever/eczema, 4881 individuals were removed due to conflicting answers between the touch-screen questionnaire and the interview, for individuals reported they had hay fever during the interview but not on the touch-screen questionnaire ($N = 2143$), or reported they had eczema in the interview but not in the touch screen ($N = 2738$). We further removed 22 individuals who had asked to be removed from the UK Biobank.

Controls

Controls ($N = 239\,773$) were selected as individuals answering “none of the above” in question 6152 and who did not report asthma, hay fever or eczema in variable number 20002. The same controls were used for all phenotypes.

Genotyping

The UK Biobank Axiom array had been used to genotype 438 417 of the 502 682 UK Biobank participants. The other 49 994 samples (all from the interim release) had been genotyped on the closely related UK BiLEVE array. The UK BiLEVE cohort and the rest of the UK Biobank differ only in small details of the DNA processing stage. The two arrays have 95% common marker content. We included a variable for array type (UK BiLEVE or UK Biobank Axiom) as covariate. SNPs in UK Biobank were imputed using UK10K (53) and 1000 genomes phase 3 (54) as reference panels. Imputation in the second release resulted in 92 693 895 SNPS (released in June 2017). However, because the UK Biobank reported problems with imputation quality for a subset of the SNPs (caused by mismatch in coordinates between the UK10 and the 1000 genomes reference panels), we followed the recommendation to only include genetic variants included on the HRC panel (55) ($N = 39\,727\,058$).

Quality control

Quality control of genotype data and imputation of genotypes had already been carried out centrally by UK Biobank. From the imputed data set, we only included SNPs in the HRC panel with a minor allele frequency (MAF) of ≥ 0.01. We removed SNPs deviating from Hardy-Weinberg (P-value $< 1 \times 10^{-20}$) and markers with more than 5% missing genotype data. We only included SNPs with an imputation quality of >0.3. After QC, a total of 15 688 218 autosomal SNPs and SNPs on the X-chromosome were included in our analyses. We only included Caucasian participants who were clustering according to the genetic principal components (56 180 non-Caucasians were removed: individuals listed in UK Biobank data file 22006). We further removed first- and second-degree relatives ($N = 32\,751$), using kinship data (estimated genetic relationship, >0.044), and participants with sex discordance and high heterozygosity/missingness (individuals listed in UK Biobank data field 22010 and 22027), as well as participants with more than 5% missing genotypes. After QC and exclusion, 346 545 unrelated Caucasian participants remained.

Genome-wide association study

A GWA study was performed for each phenotype using logistic regression and an additive genetic model implemented in PLINK version 1.90 (56). We performed GWA studies for five sets of phenotypes: 1) asthma (independent on hay fever and eczema status), 2) hay fever/eczema (hay fever and/or eczema independent on asthma status), and 3) hay fever and/or eczema and/or asthma, as well as 4) hay fever (independent on asthma and eczema status) and 5) eczema (independent on asthma and hay fever status). The same controls, which have reported that they did not have any of the disease phenotypes, were used for all analyses ($N = 239\,773$). The following covariates were included in our analysis: townsend deprivation index (as a proxy for socioeconomic status), sex, age, smoking, and the first 10 ancestry-derived principal components. In addition, to adjust for the different genotyping chips, we included a binary indicator variable for UK Biobank Axiom versus UK BiLEVE genotyping array. We calculated the LDSC intercept, using the LDSC regression software (19), for each phenotype and adjusted the summary statistics accordingly (19). The genome-wide significance threshold was set to 3×10^{-8}, as suggested for GWA studies that include variants with a MAF of ≥ 0.01(57), which was the threshold used in our study. Individual loci were defined as regions with at least one significantly associated SNP ($P \leq 3 \times 10^{-8}$). Start and stop positions for each locus were where no additional significantly associated SNPs could be found (upstream for start position or downstream for stop position) within 1 Mb.

Identification of additional independent variants within associated loci

To identify independently associated variants within each defined locus (significant SNPs ($P \leq 3 \times 10^{-8}$), within 1 Mb), we used an approximate conditional analyses implemented in GCTA (14). LD calculations were based on 5000 randomly selected Caucasian participants from UK Biobank (after sample QC). For each locus, the most significant top SNP was identified and the summary statistics of all SNPs within the same locus was adjusted by the effect of the lead SNP. After adjusting for the lead SNP, we identified the most significantly associated SNP within the locus that remained significant ($P \leq 3 \times 10^{-8}$). In the next step, we once again adjusted the summary statistics of all SNPs within the same locus, by including the effect of both the original lead SNP and the conditional lead SNP form the first iteration. This process was thereafter repeated until no other SNPs within the locus were found significant after adjusting for all previously detected independent lead SNPs.

Determining the novelty status of significant loci

To determine whether significant loci were novel to any of the diseases, we compiled a list of all asthma, hay fever, eczema and allergy risk SNPs with genome-wide significant association ($\leq 5 \times 10^{-8}$) reported in the NHGRI-EBI GWAS catalog (downloaded December 2, 2018). We also searched for GWA study results using PubMed and bioRxiv. We classified a locus to be "novel" if the locus was >1 Mb from any of the previously reported loci/variants for the disease. We also estimated LD between each lead SNP and all genome-wide significant associations found in the NHGRI-EBI GWAS catalog, to determine whether the lead SNP was a novel variant in a known locus (if the locus was <1 Mb from any of the previously reported loci/variants for the disease). We considered our associated SNP to be a novel variant if R^2 was smaller than 0.05 between our top associated variant and previously reported variants within the same locus. A locus was also reported as novel for a specific disease (i.e. asthma) if previous GWA studies only reported association to a different allergic disease (for example hay fever). If the locus was previously reported for a combined phenotype, i.e. in studies combining different allergic diseases, including the one tested, it was not reported as a novel locus.

Annotation of target genes and identification of causal genetic effects

To identify likely target genes for associated variants, we first reported the closest gene(s) to the lead SNP for each locus and reported if the SNP was intronic or exonic using the Human Genome Browser (GRCh37). We also performed additional analyses to potentially better define plausible target genes. To examine the relationship between the lead SNP for each locus and gene expression, we used the GTEx database (58) to find evidence of overlap with eQTLs. We downloaded significant eQTLs from

the GTEx database. First, we selected GTEx SNPs that overlapped with the UK Biobank SNPs and used a conservative significance threshold of $P \leq 2.3 \times 10^{-9}$ for *cis* effects (<1 Mb) from the GTEx data, in agreement with previous studies (6). Second, we identified the most significant eQTL SNP for each tissue and gene in the GTEx data set. Third, we estimated the LD between the lead eQTL SNPs and our lead GWA study SNPs. A lead GWA SNP in LD ($R^2 > 0.8$) with a lead GTEx eQTL SNP was considered to overlap with the eQTL. Only cells or tissues that were relevant for our disease phenotypes were considered when searching for eQLTs, including Epstein-Barr virus (EBV)-transformed lymphocyte, transformed fibroblasts, whole blood, lung and skin (sun exposed and not sun exposed). To further evaluate how good the evidence is for shared causal variants between the GWA signal and eQTLs, we also evaluated pairwise colocalization using the program coloc (29). Here we used GWA and eQTL summary statistics for SNPs within a 1 Mb window around each gene for which we detected an overlap between its lead eQTL and a lead GWA SNP. We considered a probability of colocalization (PPA) of ≥0.75 to be evidence for one underlying causal SNP being responsible for both (eQTL and GWA) associations.

We also used the Bioconductor biomaRt (59) package in R for functional annotation of associated SNPs. In BiomaRt, lead SNPs and all SNPs in LD ($R^2 > 0.8$) with a lead SNP were cross-referenced against: Ensembl Genes, Ensembl Variation, and Ensembl Regulation version 91 (accessed December 9, 2017, using the human assembly GRCh37). Here we checked whether the lead SNPs were in LD ($R^2 > 0.8$) with a potentially functional genetic variant by investigated regulatory features for the SNPs (i.e. promoters, enhancers etc.), binding motifs (i.e. if any of the SNPs were found within a motif for a transcription factor), and if the SNPs were possibly damaging variants (i.e. missense, stop gained, stop lost, or splice acceptor/donor variants) and if the variants were predicted to be deleterious by SIFT or PolyPhen.

Replication

We replicated our novel asthma, hay fever/eczema, asthma/hay fever/eczema and eczema loci in three independent cohorts from the EAGLE eczema consortium, the GABRIEL asthma consortium and in 23andMe ($P \leq 0.05$). Our novel loci identified for asthma was replicated using the summary statistics from the GABRIEL consortium which consisted of 10 365 physician-diagnosed asthmatic cases and 16 100 healthy controls (13). All individuals in GABRIEL were genotyped for 582 892 SNPs using the Illumina Human610 quad array. More information on this cohort has been published elsewhere (13). The EAGLE consortium GWA summary statistics consists of 21 000 atopic dermatitis (eczema) cases and 96 000 controls (60) and were used to replicate novel loci for hay fever/eczema and eczema analyzed separately. Further information about this cohort has been published previously (60). Novel loci identified in the combined analysis asthma/hay fever/eczema were replicated in 23andMe which included 83 335 cases that had reported asthma, hay fever and/or eczema and 34 934 controls. Participants in 23andMe were genotyped with four genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550+ BeadChip with a total of about 560 000 SNPs. The V3 platform was based on the Illumina OmniExpress+ Bead Chip including about 950 000 SNPs. The V4 platform is a custom array, including a lower redundancy subset of V2 and V3 SNPs with additional coverage of lower-frequency coding variation including about 570 000 SNPs. Participant genotype data were further imputed against the March 2012 "v3" release of 1000 genomes reference haplotypes. Further information about the 23andMe data used for replication in this study has been published previously (6). If the lead SNP from our study was not found in GABRIEL, EAGLE or 23andMe, we search for a proxy in LD ($R^2 \geq 0.8$) with the lead SNP.

SNP-based heritability

To quantify the SNP-based heritability for asthma and for hay fever/eczema (combined as a single phenotype) we used LDSC regression software (19) including the same cases and controls as for the association analysis for each phenotype (19). To calculate the heritability on the liability scale, we needed to adjust for disease prevalence. Since this was a population-based study, we set the Caucasian population and sample prevalence to the one calculated for each disease in UK Biobank. We included 1 108 908 HapMap SNPs to calculate the heritability for asthma and hay fever/eczema. We also removed all significant loci from each individual GWA study result to estimate how much of the heritability was explained by the significant loci reported in this study.

Identification of phenotype-specific loci (SNP)

To identify possible phenotype-specific SNPs, we performed polytomous (multinomial) logistic regression to identify whether the effect of a locus (lead SNP) was significantly (FDR ≤ 0.05) larger for one disease phenotype as compared to another. These effects can therefore be considered as being disease/phenotype specific. To conduct these analyses, we used four non-overlapping groups: 1) asthma cases without hay fever/eczema ($N = 22\,858$), 2) hay fever/eczema cases without asthma ($N = 65\,063$), 3) asthma cases with hay fever/eczema (only including $N = 19\,299$ participants that had reported asthma in combination with hay fever or eczema), and 4) controls without asthma, hay fever and eczema ($N = 240\,817$) (Fig. 2). Hay fever and eczema were not separated in this analysis due to the small sample size (Table 1).

We performed polytomous logistic regression for all possibly independent ($R^2 \leq 0.8$) associated lead SNPs identified in the asthma, hay fever/eczema or asthma/hay fever/eczema GWA studies. For some regions, different SNPs, which represent the same signal ($R^2 > 0.8$ between the SNPs), were identified in the different GWA studies. For these regions, only the SNP with the lowest *P*-value from the original GWA study was included in these analyses. For regions where different lead SNPs were identified in the different GWA studies and where these lead SNPs were not in strong LD ($R^2 < =0.8$), all lead SNPs were included in the analyses.

The polytomous (multinomial) logistic regression was performed with the response variable, *Y*, being categorically distributed with $K = 4$ non-overlapping groups/outcomes (the four non-overlapping groups are explained above). Out of $K \cdot (K-1)/2 = 6$ comparisons in total, there are $K - 1 = 3$ independent comparisons. The logit function is defined as the logarithm of the quotient between the probability of a given outcome (e.g. $P(Y = 1)$) and the probability of a reference or pivot outcome (i.e. $P(Y = 4)$ in our case). This function is assumed to be linear in all explanatory variables, including covariates and the specific SNP under consideration. Note that the beta estimates (i.e. the log-odds ratios) are unique for each comparison. The polytomous (multinomial) regression was performed using multinom in the

R library nnet for the three independent odds: $P(Y=1)/P(Y=4)$, $P(Y=2)/P(Y=4)$, and $P(Y=3)/P(Y=4)$. Beta estimates, standard errors, and P-values (two-sided, normal approximation) for the remaining comparisons between phenotypic outcomes (i.e. $P(Y=1)/P(Y=2)$, $P(Y=1)/P(Y=3)$, and $P(Y=2)/P(Y=3)$) were calculated from the model output such that, e.g. $beta_{12} = beta_{14} - beta_{24}$ and $se_{12}{}^2 = se_{14}{}^2 + se_{24}{}^2 - 2*Cov\ (beta_{14},\ beta_{24})$, where the first subscript denotes the outcome of interest, while the second subscript denotes the reference outcome. The covariance terms, which only constituted about 10% of the total variance, were taken from the inverse of the Hessian matrix.

To determine whether the lead SNPs were specific to one disease phenotype or shared among phenotypes, we identified for which disease phenotype the OR was the highest (we used the value of the OR rather than the most significant P-value in order not to be influenced by the different power in the phenotype groups due to different sample-sizes), and whether the OR was significantly (FDR ≤ 0.05) higher compared to the other disease phenotypes. As a threshold for significance, we used an FDR (Benjamini-Hochberg) value of 0.05, corresponding to a nominal P-value of < 0.017 in the three sets of cases versus cases analyses. In our analyses, an FDR adjustment is to prefer (in favor of Bonferroni) due to its power to pinpoint as many positive findings as possible, while retaining a low false-discovery rate (5% in our case).

Results were plotted as Venn diagrams to show the pairwise overlap between disease phenotypes. If two SNPs from the same locus that were not in LD with each other ($R^2 <= 0.8$) were assigned to the same area, the locus only occurs once in the Venn diagram. However, for a few loci, multiple unlinked ($R^2 <= 0.8$) SNPs from the same locus were assigned to different areas. Such loci were included at multiple locations in the Venn diagram together with the name of the SNP (i.e. gene_SNP).

Data availability

The genotypes and phenotypes included in the current study are available from the UK Biobank data, which can be accessed by researchers upon application (https://www.ukBiobank.ac.uk/). Summary statistics and codes used for this project can be accessed by contacting the corresponding author. Summary statistics from the EAGLE cohort can be downloaded from https://data.bris.ac.uk/data/dataset/, and summary statistics from the GABRIEL cohort can be downloaded from https://www.cnrgh.fr/gabriel/results.html. GWAS summary statistics for the 23andMe replication set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit https://research.23andme.com/dataset-access/for more information and to apply to access the 23andMe data.

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflicts of Interest statement

All authors declare no conflicts of interest.

Funding

Swedish Society for Medical Research; Kjell and Märta Beijers Foundation; Göran Gustafssons Foundation; Swedish Medical Research Council [Project Number 2015-03327]; Marcus Borgström Foundation; Åke Wiberg Foundation; Borgström Hedström Foundation; Swedish Heart-Lung Foundation.

URLs

UK Biobank, http://www.ukBiobank.ac.uk; PLINK, https://www.cog-genomics.org/plink2; NHGRI-EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/; Software tool for LDSC estimation and estimation of variance components from summary statistics, https://github.com/bulik/ldsc/; GCTA, http://cnsgenomics.com/software/gcta/; BiomaRt, http://www.bioconductor.org; GTEx, https://www.gtexportal.org/home/; DGIdb, http://www.dgidb.org/.

Acknowledgement

The authors acknowledge all the participants and the administrative staff at the UK Biobank and 23andMe. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under projects b2016021, b2017059, sens2017538, and sens2017541. The work was supported by grants from the Swedish Society for Medical Research, the Kjell and Märta Beijers Foundation, Göran Gustafssons Foundation, the Swedish Medical Research Council (Project Number 2015-03327), the Marcus Borgström Foundation, the Åke Wiberg Foundation, the Borgström Hedström Foundation and the Swedish Heart-Lung Foundation.

References

1. Ober, C. and Yao, T.C. (2011) The genetics of asthma and allergic disease: a 21st century perspective. *Immunol. Rev.*, **242**, 10–30.
2. Marenholz, I., Esparza-Gordillo, J., Ruschendorf, F., Bauerfeind, A., Strachan, D.P., Spycher, B.D., Baurecht, H., Margaritte-Jeannin, P., Saaf, A., Kerkhof, M. *et al.* (2015) Meta-analysis identifies seven susceptibility loci involved in the atopic march. *Nat. Commun.*, **6**, 8804.
3. Ferreira, M.A.R., Matheson, M.C., Tang, C.S., Granell, R., Ang, W., Hui, J., Kiefer, A.K., Duffy, D.L., Baltic, S., Danoy, P. *et al.* (2014) Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J. Allergy Clin. Immunol.*, **133**, 1564–1571.
4. Hinds, D.A., McMahon, G., Kiefer, A.K., Do, C.B., Eriksson, N., Evans, D.M., St Pourcain, B., Ring, S.M., Mountain, J.L., Francke, U. *et al.* (2013) A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat. Genet.*, **45**, 907–911.
5. Portelli, M.A., Hodge, E. and Sayers, I. (2015) Genetic risk factors for the development of allergic disease identified by genome-wide association. *Clin. Exp. Allergy*, **45**, 21–31.
6. Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemar, V., Van Dongen, J. *et al.* (2017) Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.*, **49**, 1752–1757.
7. Demenais, F., Margaritte-Jeannin, P., Barnes, K.C., Cookson, W.O.C., Altmüller, J., Ang, W., Barr, R.G., Beaty, T.H., Becker, A.B., Beilby, J. *et al.* (2018) Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.*, **50**, 42–53.
8. Waage, J., Standl, M., Curtin, J.A., Jessen, L.E., Thorsen, J., Tian, C., Schoettler, N., Flores, C., Abdellaoui, A., Ahluwalia, T.S.

et al. (2018) Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nat. Genet.*, **50**, 1072–1080.
9. Zhu, Z., Lee, P.H., Chaffin, M.D., Chung, W., Loh, P.R., Lu, Q., Christiani, D.C. and Liang, L. (2018) A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nat. Genet.*, **50**, 857–864.
10. Ullemar, V., Magnusson, P.K.E., Lundholm, C., Zettergren, A., Melén, E., Lichtenstein, P. and Almqvist, C. (2016) Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy Eur. J. Allergy Clin. Immunol.*, **71**, 230–238.
11. Fagnani, C., Annesi-Maesano, I., Brescianini, S., D'Ippolito, C., Medda, E., Nisticò, L., Patriarca, V., Rotondi, D., Toccaceli, V. and Stazi, M.A. (2008) Heritability and shared genetic effects of asthma and hay fever: an Italian study of young twins. *Twin Res. Hum. Genet.*, **11**, 121–131.
12. Bataille, V., Lens, M. and Spector, T.D. (2012) The use of the twin model to investigate the genetics and epigenetics of skin diseases with genomic, transcriptomic and methylation data. *J. Eur. Acad. Dermatol. Venereol.*, **26**, 1067–1073.
13. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M. and Cookson, W.O.C.M. (2010) A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.*, **363**, 1211–1221.
14. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J. *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, **44**, 369–375.
15. Hirota, T., Takahashi, A., Kubo, M., Tsunoda, T., Tomita, K., Sakashita, M., Yamada, T., Fujieda, S., Tanaka, S., Doi, S. *et al.* (2012) Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.*, **44**, 1222–1226.
16. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
17. Ng, C.P. and Henikoff, S. (2001) Predicting deleterious amino acid substitutions. *Genome Res.*, **11**, 863–874.
18. De Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.G. *et al.* (2017) Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.*, **49**, 256–261.
19. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L. and Neale, B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.
20. Tenesa, A., Farrington, S.M., Prendergast, J.G.D., Porteous, M.E., Walker, M., Haq, N., Barnetson, R.A., Theodoratou, E., Cetnarskyj, R., Cartwright, N. *et al.* (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.*, **40**, 631–637.
21. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A. *et al.* (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*, *167*, **e19**, 1415–1429.
22. Cho, M.H., Castaldi, P.J., Hersh, C.P., Hobbs, B.D., Barr, R.G., Tal-Singer, R., Bakke, P., Gulsvik, A., San José Estépar, R., Van Beek, E.J.R. *et al.* (2015) A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *Am. J. Respir. Crit. Care Med.*, **192**, 559–569.
23. Polte, T., Behrendt, A.-K. and Hansen, G. (2006) Direct evidence for a critical role of CD30 in the development of allergic asthma. *J. Allergy Clin. Immunol.*, **118**, 942–948.
24. Ameur, A., Enroth, S., Johansson, A., Zaboli, G., Igl, W., Johansson, A.C., Rivas, M.A., Daly, M.J., Schmitz, G., Hicks, A.A. *et al.* (2012) Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am J Hum Genet*, **90**, 809–820.
25. Standl, M., Sausenthaler, S., Lattka, E., Koletzko, S., Bauer, C.P., Wichmann, H.E., Von Berg, A., Berdel, D., Krämer, U., Schaaf, B. *et al.* (2012) FADS gene cluster modulates the effect of breastfeeding on asthma. Results from the GINIplus and LISAplus studies. *Allergy Eur. J. Allergy Clin. Immunol.*, **67**, 83–90.
26. Martinelli, N., Girelli, D., Malerba, G., Guarini, P., Illig, T., Trabetti, E., Sandri, M., Friso, S., Pizzolo, F., Schaeffer, L. *et al.* (2008) FADS genotypes and desaturase activity estimated by the ratio of arachidonic acid to linoleic acid are associated with inflammation and coronary artery disease. *Am. J. Clin. Nutr.*, **88**, 941–949.
27. Barman, M., Johansson, S., Hesselmar, B., Wold, A.E., Sandberg, A.S. and Sandin, A. (2013) High levels of both n-3 and n-6 long-chain polyunsaturated fatty acids in cord serum phospholipids predict allergy development. *PLoS One*, **8**, e67920.
28. Graham, D.B., Lefkovith, A., Deelen, P., de Klein, N., Varma, M., Boroughs, A., Desch, A.N., Ng, A.C.Y., Guzman, G., Schenone, M. *et al.* (2016) TMEM258 is a component of the oligosaccharyltransferase complex controlling ER stress and intestinal inflammation. *Cell Rep.*, **17**, 2955–2965.
29. Franceschini, N., Giambartolomei, C., de Vries, P.S., Finan, C., Bis, J.C., Huntley, R.P., Lovering, R.C., Tajuddin, S.M., Winkler, T.W., Graff, M. *et al.* (2018) GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat. Commun.*, **9**, 5141.
30. Esparza-Gordillo, J., Matanovic, A., Marenholz, I., Bauerfeind, A., Rohde, K., Nemat, K., Lee-Kirsch, M.A., Nordenskjöld, M., Winge, M.C.G., Keil, T. *et al.* (2015) Maternal Filaggrin mutations increase the risk of atopic dermatitis in children: an effect independent of mutation inheritance. *PLoS Genet.*, **11**, e1005076.
31. Simmonds, M. and Gough, S. (2007) The HLA region and autoimmune disease: associations and mechanisms of action. *Curr. Genomics*, **8**, 453–465.
32. Ferreira, M.A.R., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemar, V., Lu, Y. *et al.* (2018) Eleven loci with new reproducible genetic associations with allergic disease risk. *J. Allergy Clin. Immunol.*, **143**, 691–699.
33. Schuttelaar, M.L.A., Kerkhof, M., Jonkman, M.F., Koppelman, G.H., Brunekreef, B., De Jongste, J.C., Wijga, A., McLean, W.H.I. and Postma, D.S. (2009) Filaggrin mutations in the onset of eczema, sensitization, asthma, hay fever and the interaction with cat exposure. *Allergy Eur. J. Allergy Clin. Immunol.*, **64**, 1758–1765.
34. Spergel, J.M. (2010) From atopic dermatitis to asthma: the atopic march. *Ann. Allergy, Asthma Immunol.*, **105**, 99–106.

35. Illi, S., Von Mutius, E., Lau, S., Nickel, R., Grüber, C., Niggemann, B., Wahn, U., Wahn, V., Groeger, M., Zepp, F. *et al.* (2004) The natural course of atopic dermatitis from birth to age 7 years and the association with asthma. *J. Allergy Clin. Immunol.*, **113**, 925–931.
36. Hirai, H., Pang, Z., Bao, D., Miyazaki, T., Li, L., Miura, E., Parris, J., Rong, Y., Watanabe, M., Yuzaki, M. *et al.* (2005) Cbln 1 is essential for synaptic integrity and plasticity in the cerebellum. *Nat. Neurosci.*, **8**, 1534–1541.
37. Gleichman, A.J. (2012) Glutamate receptor autoimmune diseases: understanding the mechanisms, determining the epitopes, and developing better tests. ProQuest Dissertation Theses, https://repository.upenn.edu/dissertations/AAI3550953.
38. Hunt, K.A., Zhernakova, A., Turner, G., Heap, G.A.R., Franke, L., Bruinenberg, M., Romanos, J., Dinesen, L.C., Ryan, A.W., Panesar, D. *et al.* (2008) Newly identified genetic risk variants for celiac disease related to the immune response. *Nat. Genet.*, **40**, 395–402.
39. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C.A., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E. *et al.* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, **476**, 214–219.
40. Gupta, J., Johansson, E., Bernstein, J.A., Chakraborty, R., Khurana Hershey, G.K., Rothenberg, M.E. and Mersha, T.B. (2016) Resolving the etiology of atopic disorders by using genetic analysis of racial ancestry. *J. Allergy Clin. Immunol.*, **138**, 676–699.
41. Sleiman, P.M.A., Flory, J., Imielinski, M., Bradfield, J.P., Annaiah, K., Willis-Owen, S.A.G., Wang, K., Rafaels, N.M., Michel, S., Bonnelykke, K. *et al.* (2010) Variants of DENND1B associated with asthma in children. *N. Engl. J. Med.*, **362**, 36–44.
42. Israel, E., Lasky-Su, J., Markezich, A., Damask, A., Szefler, S.J., Schuemann, B., Klanderman, B., Sylvia, J., Kazani, S., Wu, R. *et al.* (2015) Genome-wide association study of short-acting β2-agonists a novel genome-wide significant locus on chromosome 2 near ASB3. *Am. J. Respir. Crit. Care Med.*, **191**, 530–537.
43. Himes, B.E., Hunninghake, G.M., Baurley, J.W., Rafaels, N.M., Sleiman, P., Strachan, D.P., Wilk, J.B., Willis-Owen, S.A.G., Klanderman, B., Lasky-Su, J. *et al.* (2009) Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am. J. Hum. Genet.*, **84**, 581–593.
44. Yucesoy, B., Kaufman, K.M., Lummus, Z.L., Weirauch, M.T., Zhang, G., Cartier, A., Boulet, L.-P., Sastre, J., Quirce, S., Tarlo, S.M. *et al.* (2015) Genome-wide association study identifies novel loci associated with diisocyanate-induced occupational asthma. *Toxicol. Sci.*, **146**, 192–201.
45. Ioannidis, J.P.A. (2005) Why most published research findings are false. *PLoS Med.*, **2**, e124.
46. Brocker, C.N., Vasiliou, V. and Nebert, D.W. (2009) Evolutionary divergence and functions of the ADAM and ADAMTS gene families. *Hum. Genomics*, **4**, 43–55.
47. Rock, F.L., Hardiman, G., Timans, J.C., Kastelein, R.A. and Bazan, J.F. (1998) A family of human receptors structurally related to Drosophila Toll. *Proc. Natl. Acad. Sci.*, **95**, 588–589.
48. Clark, J., Vagenas, P., Panesar, M. and Cope, A.P. (2005) What does tumour necrosis factor excess do to the immune system long term. *Ann. Rheum Dis.*, **64**, iv70–iv76
49. Zuk, O., Hechter, E., Sunyaev, S.R. and Lander, E.S. (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci.*, **109**, 1193–1198.
50. Visscher, P.M., McEvoy, B. and Yang, J. (2010) From Galton to GWAS: quantitative genetics of human height. *Genet. Res. (Camb).*, **92**, 371–379.
51. Golan, D., Lander, E.S. and Rosset, S. (2014) Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl. Acad. Sci.*, **111**, E5272–E5281.
52. Pawankar, R., Mori, S., Ozu, C. and Kimura, S. (2011) Overview on the pathomechanisms of allergic rhinitis. *Asia Pac. Allergy*, **1**, 157–167.
53. UK10K Consortium, Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R., Xu, C., Futema, M. *et al.* (2015) The UK10K project identifies rare variants in health and disease. *Nature*, **526**, 82–90.
54. Huang, J., Ellinghaus, D., Franke, A., Howie, B. and Li, Y. (2012) 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur. J. Hum. Genet.*, **20**, 801–805.
55. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K. *et al.* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
56. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
57. Fadista, J., Manning, A.K., Florez, J.C. and Groop, L. (2016) The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur. J. Hum. Genet.*, **24**, 1202–1205.
58. Carithers, L.J. and Moore, H.M. (2015) The Genotype-Tissue Expression (GTEx) project. *Biopreserv. Biobank.*, **13**, 307–308.
59. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A. and Huber, W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
60. Paternoster, L., Standl, M., Waage, J., Baurecht, H., Hotze, M., Strachan, D.P., Curtin, J.A., Bønnelykke, K., Tian, C., Takahashi, A. *et al.* (2015) 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.