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

Weronica E. Ek[1*], Mathias Rask-Andersen[1], Torgny Karlsson[1], Åsa Johansson[1]

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

*Corresponding author: Weronica E. Ek

Address: box 815, 75108, Uppsala, Sweden

Telephone: +46703519004.

Fax: +46184714931

Email: Weronica.ek@igp.uu.se

Abstract

The disease risk for asthma, hay fever and eczema include both environmental and genetic risk factors and the comorbidity between the diseases are large. Heritability estimates suggest that the risk of asthma, hay fever and eczema is largely due to genetic factors. In this GWAS, we include 346,545 Caucasian participants from the UK Biobank to increase power to identify novel loci for asthma, hay fever and eczema. We also investigate if associated lead SNPs have a significantly larger effect for one disease phenotype compared to the other phenotypes, to highlight possible disease specific effects.

This study identifies 141 loci, of which 41 are novel to this study, to be associated ($P \leq 3x10^{-8}$) with asthma, hay fever or eczema, analysed separately or combined as a single phenotype. At four of the novel loci, *TNFRSF8, MYRF, TSPAN8,* and *BHMG1,* the lead SNPs were in LD (> 0.8) with potentially casual missense variants. For seven of the novel GWAS loci, the lead SNP was in LD (> 0.8) with genetic variants associated with gene expression (eQTL) where, for example, increased levels of *TMEM258* as well as decreased levels of *HHEX* and *ADAM19* was associated with decreased odds for asthma.

Our study shows that a large amount of the genetic contribution to asthma, hay fever and eczema 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 vary among populations and have an underlying architecture that include both environmental and genetic risk factors (1). Comorbidity between asthma, hay fever and eczema is common, and previous genome-wide association studies (GWAS) have, apart from identifying a large number of genetic variants associated with risk of disease (2–7), also found evidence of a genetic overlap between the diseases (6, 8). In addition, eczema is a major risk factor for developing asthma and hay fever (9). For example, previous studies have shown that children with eczema have an increased odds of developing asthma, called the 'atopic march', as compared to children without eczema (9–12). A previous study has identified genetic factors that underlie the atopic march and has shown that, especially eczema loci, are associated with an increased risk for the atopic march (2).

Family and twin studies have estimated that the contribution of genetic factors, i.e. the heritability for asthma (1, 13, 14) to be 35-95%, for hay fever (1, 14) to be 33-91%, and for eczema (15) to be as high as 90%. A recent large study estimated the SNP-based heritability, the heritability that can be attributed the genetic variation captured by SNPs in a GWAS, to be 15% for asthma, 22% for hay fever, and 9% eczema (6). The same GWAS, which is the largest GWAS so far analysing asthma, hay fever and eczema as a combined phenotype, 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).

The aim with this study was to further investigate the genetic background of self-reported asthma, hay fever, and eczema, as well as hay fever and/or eczema combined

and asthma and/or hay fever and/or eczema combined to a single phenotype using data from the UK Biobank. Although the phenotypes in the UK Biobank are self-reported, the questions are well defined and identical for all participants. Many previous GWAS for asthma, hay fever and eczema have been conducted in different cohorts that were subsequently meta-analysed with the purpose of increasing statistical power (2, 4, 6, 7, 16). Associated SNPs were functionally annotated to assess likely causal mechanisms. We also investigated if associated lead SNPs had a significantly larger effect for one disease phenotype compared to the other phenotypes, to highlight possible disease specific effects.

Results

Association analysis

The UK Biobank includes 502,682 participants, of which 443,068 are Caucasians. The disease prevalence in the Caucasian participants were 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 (Figure 2). We conducted the GWAS 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 where included in the analyses (see methods). QC and the final number of included participants in respective analyses are summarized in the methods and in Table 1.

GWAS 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 GWAS analysis. We identified 75

risk loci located > 1 Mb apart and containing at least one significantly associated genetic variant ($P \leq 3x10^{-8}$ after adjusting for LD-score intercept of 1.065), that were associated with self-reported asthma, of which 15 loci were found to be novel GWAS asthma loci (Table 2; Manhattan plot, Figure 1; S1 Table and S2 Table; quantile-quantile (QQ) plot, S1 Fig). Using approximate conditional analysis (17), we identified 116 independent significant associations within these 75 loci (S1 Table). The strongest associations for asthma were found within the HLA locus on chromosome 6 ($P=2.06x10^{-100}$), 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, 16). Among the novel asthma loci, some have previously been associated with other similar phenotypes (S1 Table). For example, *SDK1*, previously annotated to the nearby *CARD11* gene, have been reported to be associated with atopic dermatitis (18), but this seems to be the first time the *SDK1* locus has been associated with asthma.

Annotation of asthma associated SNPs

Associated SNPs were further functionally annotated to assess likely causal mechanisms (see Methods). Overlap with GTEx eQTLs were 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* appears to lower the risk for developing asthma (Table 2; S1 Table S1 and S3 Table). In contrast, increased expression of *EEFSEC* in lung tissue seem to increase the risk for asthma (S3 Table). Similarly, increased expression of *ADAM19* and *HHEX* appears to increase the risk of asthma (S3 Table). 19 probable causal missense variants could be observed within the 75 significant GWAS loci of which four missense variants for the 15 novel loci (S4 Table). The latter are located within *TNFRSF8*, *MYRF*, *TSPAN8*, and *BHMG1*. The association at *TNFRSF8* was represented by only one genetic

variant, rs2230624 (S2 Fig). 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 (19) (PolyPhen-score 0.751-0.921) and had a 'deleterious' SIFT-score (20) 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(19) (PolyPhen-score 0.961-1) and had a 'deleterious' SIFT-score (20) of 0.04-0.07. However, rs174535 is also in LD with the most significant eQTL for *TMEM258*. The lead SNP in the *BHMG1* locus, rs11671106, is a missense variant for *BHMG1* and was predicted as 'probably damaging' by PolyPhen (19) (PolyPhen-score 0.94) and had a 'deleterious' SIFT-score (20) of 0.01. The lead SNP at the *TSPAN8* 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 (19) (PolyPhen-score 0.989) and had a 'deleterious' SIFT-score(20) of 0.03

<u>GWAS for self-reported hay fever/eczema</u>

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 ($P \leq 3x10^{-8}$, LD-score intercept =1.079) with self-reported hay fever/eczema, and 22 of these were novel (Table 3; Manhattan plot, Figure 1; S5 Table and S6 Table; QQ-plot, S3 Fig). The strongest association was observed for the lead SNP rs5743604 ($P= 7.5x10^{-72}$) located within *TLR1*. This SNP has previously been associated with allergic disease (6, 21). Using conditional analysis, we identified 154 independent significant associations within these 109 loci (Table 3; S5 Table). 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 (S5 Table). The *UBAC2* locus has previously been reported to be associated with asthma (8), but this seems to be the first time the *UBAC2* locus is reported to be associated with hay fever and/or eczema.

Annotation of hay fever/eczema SNPs

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

GWAS for hay fever and eczema analysed 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 analysed hay fever (N hay fever cases = 18,915) and eczema (N eczema cases = 7,884) separately in a smaller subset of UK Biobank participants (Manhattan plot, S4 Fig; QQ-plots, S5 and S6 Fig). 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, was detected when analysing hay fever and eczema separately (S8-S11 Tables). The lead SNP, rs12920150 ($P=1.02x10^{-9}$), at the hay fever locus is located close to *CBLN1* and the lead SNP, rs2485363 ($P=1,20x10^{-8}$), at the eczema locus is located downstream of *TAGAP*. Another locus that was not detected when analysing hay fever/eczema combined was detected when analysing eczema separately. The lead SNP for this locus, rs676387 ($P=2.26x10^{-10}$), is located within *HSD17B1* (S10 and S11 Tables). 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 (S3 Table).

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

For the combined analysis of asthma and/or hay fever and/or eczema (N cases=106,752), we identified 110 significant loci (LD-score intercept=1.081), and 16 of these were novel GWAS loci that have not been significantly associated with either asthma, hay fever or eczema in previous GWA studies (Table 4; Manhattan plot, Figure 1; S12 and S13 Tables; QQ-plot, S7 Fig). However, 12 of these 16 novel GWAS loci were detected when analysing asthma and hay fever/eczema separately, while the remaining four novel loci were only found when analysing 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; S12 Table). The most significant SNP, rs72823641 ($P=1.14x10^{-78}$), was located within *IL1RL1* and was also significantly associated with asthma and hay fever/eczema when these phenotypes were analysed separately ($P = 4.09x10^{-61}$ and $P = 9.64x10^{-64}$) (S12 and S13 Tables). This region has previously been associated with allergic diseases (6) (S14 Table). 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* (S12 and S14 Tables).

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; S3 and S12 Tables). Among

the novel loci, one overlapped with an eQTL for *HIST1H2BD* in whole blood ($P=1.11x10^{-16}$). A decreased level of *HIST1H2BD* seems to increase the risk of this combined phenotype (S3 Table). Probable causal missense variants could be observed at 17 out of 110 significant GWAS 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 (S15 Table).

SNP-based heritability

To quantify the SNP-based heritability for asthma and hay fever/eczema, we used LD score regression analysis (LDSC) (22). These analyses included the same cases and controls as for the association analysis (see Methods). The SNP-based heritability was estimated to be 21% for asthma and 16% for hay fever/eczema (Table 5). Our significant GWAS loci, which were located ≥ 1 Mb apart and contained at least one significantly associated genetic variant at $P\leq3x10^{-8}$, 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 GWASs, we included all individuals reporting either asthma (for the asthma GWAS) or hay fever/eczema (for the hay fever/eczema GWAS) 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 identify 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) (Figure 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 (S16 Table).

A total of 154 lead GWAS SNPs (see method section for a description of the selection of lead SNPs), representing the 138 different loci, identified in the GWAS 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 (Figure 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, 24 loci/SNPs were specific for asthma-only, i.e. had a significantly higher OR for asthma compared to for hay fever/eczema, while 19 loci/SNPs were specific to hay fever/eczema-only. A major part of the loci/SNPs, 106, showed no significant difference in effect between these disease phenotypes. (Figure 3; S16 Table).

When comparing subjects with asthma and hay fever/eczema to subjects with asthma only, 46 loci/SNPs were specific for asthma with hay fever/eczema. No SNP was specific for the asthma only group (Figure 3; S16 Table). For the remaining 103 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 (Figure 3; S16 Table), 63 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 84 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 (Figure 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 (Figure 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 (Figure 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 GWAS. 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) towards having a larger effect in asthma-only compared to hay fever/eczema-only (S16 Table).

Discussion

In this large GWAS, 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 analysed separately or together as combined phenotypes. Out of these loci, 41 are novel to our study and have not been reported to be associated with the same disease phenotype previously. However, some of our novel loci have previously been associated with a related phenotype (S1, S5, S8, S10 and S12 Tables). For example, some of the novel asthma loci has previously been associated with IgE levels, eosinophil counts or dermatitis and some of the novel hay fever/eczema loci with IgE levels or eosinophil counts.

The SNP-based heritability was estimated to be 21% for asthma and 16% 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 GWAS. The SNP-based heritability for asthma has previously been estimated at 15% (6), which is slightly lower than the estimate from our study. In comparison to the high estimates for the heritability (33-95%) from family and twin studies (1, 13, 15), this suggests that a major contribution to the genetic risk for asthma, hay fever and eczema cannot be identified in studies on common genetic variants. However, heritability estimates from family and twin studies have been suggested to be overestimated (23–25) 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.

The largest number of loci were 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 a previous GWAS (6). One study also estimated a large genetic correlation ($r_g = 0.75$, $P = 6.84 \times 10^{-62}$) between asthma and allergic diseases (8). With this combined phenotype (asthma and/or hay fever and/or eczema), four novel loci were identified that were not found for asthma or hay fever/eczema when analysed 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 *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 (26). The variant at the *SMAD7* locus is located within an intron of *SMAD7*. This gene has previously been associated with inflammatory bowel disease (IBD) (27), colorectal cancer (28), and haemoglobin concentration (26). 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 (26), and chronic obstructive pulmonary disease (29).

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 *FGL* 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.R2447X, 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.

For four of the novel loci: 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 (32). To our knowledge, this is the first time that this locus has been associated with asthma and allergy in a GWAS. 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 (33). 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 (34); another has been associated with increased risk of inflammation (35). 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 (36).

For seven of the 41 novel GWAS loci, the lead SNP was in LD (> 0.8) with an eQTL. We could see a positive correlation between expression of *TUBG2, HHEX, EEFSEC,* and *ADAM19* and risk of disease, as well as a negative correlation for *TMEM258 and HIST1H2BD*. Decreased expression of *TMEM258* 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 (37), which agrees with our results. 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.

As highlighted by this study, as well as previous studies (2–7)(38), 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 ($P=7.96x10^{-25}$), found within the *FLG* locus, was more significantly associated with eczema when this phenotype was analysed separately ($P=1.15x10^{-69}$). In contrast, this variant was not associated with hay fever when hay fever was analysed separately. It was, however, associated with

asthma ($P=2.37x10^{-27}$). This is in agreement with the previous GWAS 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 (39). 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 (10, 12, 39). When analysing 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* (40). Autoantibodies against glutamate receptors are involved in the development of autoimmune disease (41). One novel locus was also identified when analysing eczema separately, downstream of *TAGAP*. This locus has previously been associated with celiac disease (42) and multiple sclerosis (43).

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 ($P \leq 0.05$; S14 Table). 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 were presented in our data (S14 Table). 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 (44). 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 (45, 46), populations of different ancestry (18, 47) or small sample sizes (< 10,000) (47, 48). 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 (49) (for more information, see S14 Table). A recent GWAS by Zhu *et al* (8), which was also conducted on the UK Biobank cohort, reported seven novel allergy-associated loci. These loci where not available in the GWAS catalog at the time of writing this article and are therefore not included in S14 Table. When analysing asthma and different allergic disease phenotypes jointly, Zhu *et al* (8) identified seven novel allergy loci (8), five of which were replicated in our study. The two loci that did not replicate in our study where mapped to *ALG9* on chromosome 11(rs659529) and to *EVI5* on chromosome 1 (rs12743520).

In previous GWAS 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 GWAS analyses, but not including the three SNPs identified for the hay fever and eczema phenotypes analysed separately since we did not have power (large enough sample size) enough to include hay fever and eczema separately in these analyse. 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 GWAS 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 19 loci had a significantly larger effect on hay fever/eczema-only compared to their effects on asthma while 24 loci exhibited larger effects on asthma compared with their effects on hay fever/eczema (Figure 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 (50). 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 (51). 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, TNFSF4*. These proteins are mainly expressed in immune cells and regulate immune response and inflammation as well as proliferation, apoptosis and embryogenesis (52).

The largest number of phenotype-specific loci was observed for the group of cases with asthma and hay fever/eczema (Figure 3; S16 Table), 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 considerable 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=6,276 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 analysed 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.

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 principal components and adjusting for the LD-score 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 utilise a large number of smaller cohorts from different countries and age-groups. Analysing 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-, 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 (53).

Conclusion

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 analysing asthma, hay fever and eczema together. Two novel loci were also identified when analysing 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.

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 centre once, but some individuals visited the centre 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 is available for most participants. The UK Biobank study was approved by the National Research Ethics Committee (REC reference 11/NW/0382). An application for using data from UK Biobank has been approved

(application nr: 15479). 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).

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 analysed in this study. For hay fever/eczema, 4,881 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=2,143), or reported they had eczema in the interview but not in the touch-screen (N=2,738). 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 (54) and 1000 genomes phase 3 (55) 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 (56) (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 dataset, we only included SNPs in the HRC panel with a MAF ≥ 0.01. We removed SNPs deviating from Hardy-Weinberg (P-value < $1x10^{-20}$) and markers with more than 5% missing genotype data. We only included SNPs with an imputation quality >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, high heterozygosity/missingness (individuals listed in UK Biobank data field 22010 and 22027), and participants with more than 5% missing genotypes. After QC and exclusion, 346,545 unrelated Caucasian participants remained.

Genome-wide association study

GWAS were performed for each phenotype using logistic regression and an additive genetic model implemented in PLINK version 1.90 (57). We performed a GWAS 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), 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, that 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 (TDI) (as a proxy for socioeconomic status), sex, age, smoking, and the first ten 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 LD-score intercept, using the LD score regression software (LDSC) (22), for each phenotype and adjusted the GWAS summary statistics accordingly (22). The genome-wide significance threshold was set to $3x10^{-8}$, as suggested for GWA studies that include variants with a minor allele frequency (MAF) ≥ 0.01(58), which was the threshold used in our study. Individual loci were defined as regions with at least one significantly associated SNP ($P \leq 3x10^{-8}$). 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 ($P \leq 3x10^{-8}$), within 1 Mb), we used an approximate conditional analyses implemented in GCTA (17). LD calculations were based on 5,000 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 ($P \leq 3x10^{-8}$). 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 ($\leq 5x10^{-8}$) reported in the NHGRI-EBI GWAS catalog (downloaded December 2, 2018). We also searched for GWA studies 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 were 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 GWAS 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.

<u>Annotation of target genes and identification of causal genetic effects</u>

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 Genotype-Tissue Expression (GTEx) database (59) to find evidence of overlap with expression quantitative loci (eQTLs). We downloaded significant eQTLs from the Genotype-Tissue Expression (GTEx) database. First, we selected GTEx SNPs that overlapped with the UK Biobank SNPs and used a conservative significance threshold

$P \leq 2.3 \times 10^{-9}$ for cis effects (<1 Mb) form 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 dataset. Third, we estimated the LD between the lead eQTL SNPs and our lead GWAS SNPs. A lead GWAS 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 EBV-transformed lymphocyte, transformed fibroblasts, whole blood, lung and skin (sun exposed and not sun exposed).

We also used the Bioconductor biomaRt (60) 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 9 December 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.

<u>SNP-based heritability</u>

To quantify the SNP-based heritability for asthma and for hay fever/eczema (combined as a single phenotype) we used LD score regression software (LDSC) (22) including the same cases and controls as for the association analysis for each phenotype (22). 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 GWAS 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 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) (Figure 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<=0.8) associated lead SNPs identified in the asthma, hay fever/eczema or asthma/hay fever/eczema GWASs. For some regions, different SNPs, that represent the same signal (R^2>0.8 between the SNPs), were identified in the different GWAs. For these regions, only the SNP with the lowest P-value from the original GWAS was included in these analyses. For regions where, different lead SNPs were identified in the different GWAS, 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·(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 $se_{12}^2 = se_{14}^2 + se_{24}^2$, where the first subscript denotes the outcome of interest while the second subscript denotes the reference outcome.

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 vs cases analyses. In our analyses, an FDR adjustment is to prefer (in favour 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 pair-wise 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).

Ethics

UK Biobank was given ethical approval by the North West Multicentre Research Ethics Committee, 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.

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/). Codes used for this project can be accessed by contacting the corresponding author.

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 LD Score 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

We acknowledge all the participants and the administrative staff at the UK Biobank. 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 (SSMF), 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.

Author contributions

Planned the study (WEE and ÅJ), analysed the data (WEE, ÅJ, TK), literature search (WEE), Figures (WEE, MRA, ÅJ), data interpretation (WEE, ÅJ, TK, MRA), writing of manuscript (WEE, ÅJ, TK, MRA).

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**.

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**, 907–11.

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.*, 10.1038/s41588-017-0014-7.

8. 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.*, 10.1038/s41588-018-0121-0.

9. Gustafsson,D., Sjoberg,O. and Foucard,T. (2000) Development of allergies and asthma in infants and young children with atopic dermatitis--a prospective follow-up to 7 years of age. *Allergy*, 10.1034/j.1398-9995.2000.00391.x.

10. Spergel,J.M. (2010) From atopic dermatitis to asthma: The atopic march. *Ann. Allergy, Asthma Immunol.*, 10.1016/j.anai.2009.10.002.

11. Zheng,T., Yu,J., Oh,M.H. and Zhu,Z. (2011) The atopic march: Progression from atopic dermatitis to allergic rhinitis and asthma. *Allergy, Asthma Immunol. Res.*, **3**, 67–73.

12. 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.*, 10.1016/j.jaci.2004.01.778.

13. 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.

14. 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.

15. 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.

16. 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**, 1211–21.

17. 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.

18. 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.

19. 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.

20. Ng, C.,P. and Henikoff,S. (2001) Predicting deleterious amino acid substitutions. *Genome Res.*, **11**, 863–874.

21. Pickrell,J.K., Berisa,T., Liu,J.Z., Ségurel,L., Tung,J.Y. and Hinds,D.A. (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.*, **48**, 709–717.

22. 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.

23. 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.

24. Visscher,P.M., McEvoy,B. and Yang,J. (2010) From Galton to GWAS: quantitative genetics of human height. *Genet. Res. (Camb).*, **92**, 371–379.

25. 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.

26. 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**, 1415–1429.e19.

27. 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.

28. 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.

29. 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.

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**.

31. Simmonds,M. and Gough,S. (2007) The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Curr. Genomics*, **8**, 453–465.

32. 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.*, 10.1016/j.jaci.2006.07.014.

33. 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.

34. 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.

35. 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.

36. 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**.

37. 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.

38. 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.*, 10.1016/j.jaci.2018.03.012.

39. 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.

40. Hirai,H., Pang,Z., Bao,D., Miyazaki,T., Li,L., Miura,E., Parris,J., Rong,Y., Watanabe,M., Yuzaki,M., *et al.* (2005) Cbln1 is essential for synaptic integrity and plasticity in the cerebellum. *Nat. Neurosci.*, 10.1038/nn1576.

41. Gleichman,A.J. (2012) Glutamate receptor autoimmune diseases: Understanding the mechanisms, determining the epitopes, and developing better tests. *ProQuest Diss. Theses*.

42. 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.*, 10.1038/ng.102.

43. 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*, 10.1038/nature10251.

44. 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.*, 10.1016/j.jaci.2016.02.045.

45. 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.*, 10.1056/NEJMoa0901867.

46. 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.*, 10.1164/rccm.201408-1426OC.

47. 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.*, 10.1016/j.ajhg.2009.04.006.

48. 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.

49. Ioannidis,J.P.A. (2005) Why most published research findings are false. *PLoS Med.*, 10.1371/journal.pmed.0020124.

50. Brocker,C.N., Vasiliou,V. and Nebert,D.W. (2009) Evolutionary divergence and functions of the ADAM and ADAMTS gene families. *Hum. Genomics*, 10.1186/1479-7364-4-1-43.

51. 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.*, 10.1073/pnas.95.2.588.

52. Clark,J., Vagenas,P., Panesar,M. and Cope,A.P. (2005) What does tumour necrosis factor excess do to the immune system long term? In *Annals of the Rheumatic Diseases*.

53. Pawankar,R., Mori,S., Ozu,C. and Kimura,S. (2011) Overview on the

pathomechanisms of allergic rhinitis. *Asia Pac. Allergy*, 10.5415/apallergy.2011.1.3.157.

54. UK10K Consortium (2015) The UK10K project identifies rare variants in health and disease. *Nature*, **526**, 82–90.

55. 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.

56. 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.

57. 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.

58. 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.

59. Carithers,L.J. and Moore,H.M. (2015) The Genotype-Tissue Expression (GTEx) Project. *Biopreserv. Biobank.*, **13**, 307–308.

60. 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.

Figure Legends

Figure 1. Manhattan plots for asthma, for hay fever and/or eczema, and for asthma and/or hay fever and/or eczema (combined). The black horizontal line indicates the genome wide threshold ($3x10^{-8}$). The black regions represent novel GWAS loci found in this study.

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

Figure 3. Venn diagram showing the phenotype-specificity of the GWAS 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 (R2<0.8) lead GWAS SNP has been analysed 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, 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).

S1 Figure QQ-plot for Asthma in UK Biobank. The red line denotes the expected null-line of no association.

S2 Figure Regional plot for the missense variant rs2230624 in asthma.

S3 Figure QQ-plot for Hay fever and/or Eczema (combined) in UK Biobank. The red line denotes the expected null-line of no association.

S4 Figure Manhattan plots for Hay Fever and/or Eczema (combined), Hay Fever (only) and Eczema (only) analysed in UK Biobank. The black horizontal line indicates the genome wide threshold ($3x10^{-8}$). The black regions in the hay fever and/or eczema (combined) plot represent novel GWAS loci found in this study and the black regions in the hay fever (only) and eczema (only) plot represent two novel loci, not found in previous GWAS studies or in our combined hay fever / eczema analysis.

S5 Figure QQ-plot for Hay fever (only) in UK Biobank. The red line denotes the expected null-line of no association.

S6 Figure QQ-plot for Eczema (only) in UK Biobank. The red line denotes the expected null-line of no association.

S7 Figure QQ-plot for Asthma and/or Hay fever and/or Eczema (combined) in UK Biobank. The red line denotes the expected null-line of no association.

Tables

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	9,578	294,477
N total included after QC[d]	41,934	84,050	106,772	18,915	7,884	239,773
N (%) males after QC	21,730 (51.8%)	42,639 (50.7%)	55,124 (51.6%)	8,692 (46.0%)	3,365 (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)	-6.3-10.6 (-1.3)	-6.3-10.6 (-1.6)	-6.3-10.6 (-1.5)	-6.3-10.4 (-1.7)	-6.3-9.6 (-1.9)	-6.3-10.9 (-1.5)
% Ever smoked (N yes / N no)	60.4% (31,040 / 20,390)	58.5% (59,983 / 42,616)	59.3% (77,335/53,121)	56.8% (12,986/9,879)	60.9% (5,815/3,737)	60.4% (177,212/116,275)

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

Table 2. Summary results for the 15 novel loci significantly associated with self-reported asthma in UK

lead SNP	locus[a] chr:start-end (kbp)	N snps (total[b]/ independent[c])	MAF[d]	Minor/ major allele	OR for minor allele	P	Likely target gene (s)
rs2230624	1:12,175-12,175	1/1	0.02	A/G	0.80	1.01×10^{-10}	*TNFRSF8*[e]
rs2296618	1:198,656-198,670	5/1	0.13	G/A	0.93	8.03×10^{-9}	*PTPRC*[f]
rs10934853	3:127,886-128,075	3/1	0.27	A/C	0.95	2.20×10^{-8}	*EEFSEC*[g]
rs6778937	3:176,708-176,868	28/1	0.28	C/T	0.95	2.54×10^{-9}	*TBL1XR1*[f]
rs11466773	5:156,930-156,988	7/1	0.06	T/C	1.09	6.32×10^{-9}	*ADAM19*[g]
rs2614266	6:135,691-135,818	6/1	0.44	A/T	1.05	8.90×10^{-9}	*AHI1*[f]
rs10215232	7:3,062-3,153	12/1	0.12	G/C	0.93	1.53×10^{-9}	*SDK1*[f]
rs41283642	9:101,915-101,989	3/1	0.03	T/C	0.86	1.27×10^{-11}	*TGFBR1*[g]
rs2497318	10:94,34-94,44	24/1	0.45	T/C	0.95	3.21×10^{-10}	*HHEX*[g]
rs174535	11:61,543-61,623	49/1	0.35	C/T	0.95	1.02×10^{-11}	*MYRF*[e], *TMEM258*[g]
rs11178649	12:71,409-71,585	103/1	0.41	T/G	0.95	2.68×10^{-11}	*TSPAN8*[e]
rs4761592	12:94,556-94,604	17/1	0.15	T/C	0.93	1.27×10^{-9}	*PLXNC1*[f]
rs9316059	13:44,475-44,490	5/1	0.20	T/A	1.06	1.35×10^{-8}	*LINC00284*[f]
rs4842921	15:84,556-84,556	1/1	0.39	A/G	0.96	2.63×10^{-8}	*ADAMTSL3*[f]
rs11671106	19:46,219-46,370	28/1	0.35	T/C	0.95	8.29×10^{-10}	*BHMG1*[e]

Biobank ($P \leq 3 \times 10^{-8}$).

More details can be found in S1-S3 and S4 Tables.

[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 (17).

[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.

Table 3. Summary results for the 22 novel loci significantly associated with self-reported hay fever and/or eczema in UK Biobank ($P \leq 3x10^{-8}$).

Lead SNP	Locus[a] chr:start-	N snps (total[b]/ independent[c])	MAF[d]	Minor/ major	P (OR[e]) hay fever/	P (OR[e]) hay	P (OR[e]) eczema	Likely target gene
rs1201113	1:12,100-12,147	3/1	0.12	A/G	$1.05x10^{-8}$ (0.95)	$4.67x10^{-3}$ (0.95)	$1.69x10^{-2}$ (0.94)	*TNFRSF*[f]
rs906363	1:212,858-212,877	6/1	0.15	C/T	$4.44x10^{-9}$ (1.05)	$8.56x10^{-4}$ (0.97)	$2.54x10^{-4}$ (1.09)	*BATF3*[f]
rs13405815	2:28,623-28,644	9/1	0.46	T/C	$3.01x10^{-9}$ (0.97)	$4.04x10^{-5}$ (0.96)	$3.70x10^{-3}$ (0.95)	*RP11-373D23.3*[g]
rs10185028	2:61,112-61,161	6/1	0.23	G/A	$2.74x10^{-8}$ (1.04)	$9.47x10^{-4}$ (1.04)	$1.17x10^{-2}$ (1.05)	*REL*[f]
rs11717778	3:112,526-112,693	144/1	0.34	A/G	$6.04x20^{-11}$ (0.96)	$2.67x10^{-4}$ (0.96)	$5.38x10^{-6}$ (0.92)	*CD200R1L*[f]
rs62379371	5:133,439-133,639	4/1	0.05	A/G	$6.08x10^{-14}$ (0.90)	$7.83x10^{-6}$ (0.89)	$4.18x10^{-5}$ (0.85)	*TCF7*[f]
rs13185930	5:137,461-137,605	10/1	0.25	A/G	$1.12x10^{-8}$ (1.04)	$3.19x10^{-2}$ (1.03)	$4.64x10^{-1}$ (1.01)	*GFRA3*[f]
rs2229768	6:25,823-26,239	24/1	0.24	C/T	$8.20x10^{-11}$ (1.05)	$8.99x10^{-6}$ (1.06)	$2.74x10^{-1}$ (1.02)	*U91328.19*[g]
rs1998266	6:36,349-36,380	5/1	0.14	T/C	$1.70x10^{-8}$ (0.95)	$1.23x10^{-3}$ (0.95)	$1.68x10^{-3}$ (0.93)	*ETV7*[f]
rs2746438	6:135,624-135,950	36/1	0.44	T/A	$5.70x10^{-11}$ (1.04)	$9.24x10^{-7}$ (1.05)	$2.73x10^{-5}$ (1.07)	*AHI1*[f]
rs3918226	7:150,690-150,690	1/1	0.08	T/C	$5.65x10^{-10}$ (0.93)	$1.54x10^{-7}$ (0.90)	$6.09x10^{-1}$ (0.98)	*NOS3*[f]
rs6986151	8:101,514-101,519	2/1	0.20	C/T	$6.19x10^{-9}$ (1.04)	$5.39x10^{-6}$ (1.06)	$2.44x10^{-3}$ (1.06)	*ANKRD46*[f]
rs1330303	9:16,715-16,756	2/1	0.35	T/C	$5.21x10^{-10}$ (0.96)	$2.95x10^{-8}$ (0.94)	$3.90x10^{-1}$ (0.99)	*BNC2*[f]
rs4743311	9:101,790-101,820	3/1	0.25	G/A	$1.69x10^{-8}$ (1.04)	$7.93x10^{-6}$ (1.06)	$1.27x10^{-1}$ (1.03)	*COL15A1*[f]
rs12343737	9:117,804-117,834	2/1	0.10	T/C	$2.09x10^{-8}$ (0.95)	$1.28x10^{-6}$ (0.92)	$9.59x10^{-2}$ (0.96)	*TNC*[f]
rs10986320	9:127,022-127,095	13/1	0.37	C/G	$8.81x10^{-9}$ (1.04)	$4.13x10^{-5}$ (1.05)	$7.32x10^{-2}$ (1.03)	*NEK6*[f]
rs4076542	11:2,237-2,296	4/1	0.37	A/G	$1.74x10^{-8}$ (1.04)	$4.40x10^{-3}$ (1.04)	$7.78x10^{-1}$ (1.02)	*ASCL2*[f]
rs4939490	11:60,793-60,793	4/1	0.39	G/C	$2.15x10^{-8}$ (0.97)	$2.15x10^{-3}$ (0.97)	$9.76x10^{-2}$ (0.97)	*CD6*[f]
rs3116590	13:50,808-50,811	2/1	0.21	G/A	$1.00x10^{-8}$ (1.04)	$1.11x10^{-2}$ (1.03)	$7.24x10^{-3}$ (1.05)	*DLEU1*[f]
rs4771332	13:99,839-100,070	7/1	0.31	T/C	$6.21x10^{-9}$ (0.96)	$6.09x10^{-3}$ (0.97)	$1.29x10^{-1}$ (0.97)	*UBAC2*
rs4381563	15:75,399-75,448	7/1	0.34	A/T	$3.37x10^{-9}$ (0.96)	$2.98x10^{-3}$ (0.97)	$3.37x10^{-2}$ (0.96)	*PPCDC*[f]
rs6066184	20:45,232-45,716	33/1	0.26	G/C	$6.55x10^{-16}$ (0.95)	$5.25E^{-11}$ (0.92)	$2.03x10^{-2}$ (0.96)	*EYA2*[f]

More details can be found in S3 and S5-S7 Table.

[a] Defined as SNPs located < 1 Mb apart containing at least one significantly associated genetic variant at $P \leq 3x10^{-8}$.

[b] Total number of SNPs with $P \leq 3x10^{-8}$ within loci.

[c] Total number of independent associations within the locus, based on conditional analysis (17).

[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.

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 ($P \leq 3x10^{-8}$).

Lead SNP	Locus[a] chr:start-stop (kbp)	N snps (total[b]/ independent[c])	MAF[d]	Minor/ major allele	P (OR[e]) combined	P (OR[e]) asthma	P (OR[e]) hay fever/ eczema	Likely target gene
rs2230624	1:12,080-12,175	2/2	0.02	A/G	$2.64x10^{-9}$ (0.87)	$1.01x10^{-10}$ (0.80)	$1.99x10^{-7}$ (0.88)	*TNFRSF8*[f]
rs7410883	1:198,640-198,670	5/1	0.11	C/T	$7.00x10^{-9}$ (0.95)	$2.26x10^{-8}$ (0.93)	$1.78x10^{-6}$ (0.96)	*PTPRC*[g]
rs9816107	3:112,526-112,693	160/1	0.34	A/C	$6.80x10^{-11}$ (0.96)	$1.51x10^{-6}$ (0.96)	$8.39x10^{-11}$ (0.96)	*CD200R1L*[g]
rs62379371	5:133,439-133,639	4/1	0.05	A/G	$1.13x10^{-13}$ (0.91)	$1.23x10^{-7}$ (0.91)	$6.08x10^{-14}$ (0.90)	*VDAC1*[g]
rs9379828	6:26,038-26,184	13/1	0.37	G/C	$2.71x10^{-9}$ (1.03)	$1.07x10^{-10}$ (1.05)	$1.42x10^{-6}$ (1.03)	*HIST1H2BD*[h]
rs1330303	9:16,715-16,715	1/1	0.35	T/C	$2.84x10^{-8}$ (0.97)	0.016 (0.98)	$5.21x10^{-10}$ (0.96)	*BNC2*[g]
rs41283642	9:101,915-101,915	1/1	0.03	T/C	$7.03x10^{-9}$ (0.92)	$1.27x10^{-11}$ (0.86)	$6.42x10^{-6}$ (0.93)	*TGFBR1*[g]
rs3758212	9:127,002-127,178	14/1	0.37	T/C	$2.99x10^{-9}$ (1.03)	$1.80x10^{-5}$ (1.04)	$2.18x10^{-8}$ (1.04)	*NEK6*[g]
rs2505504	**10:43,728-43,763**	**17/1**	**0.29**	**A/G**	**$5.45x10^{-9}$ (1.04)**	**$1.96x10^{-6}$ (1.04)**	**$4.61x10^{-7}$ (1.03)**	***RASGEF1A***[g]
rs7114923	11:2,237-2,305	21/1	0.37	T/C	$3.74x10^{-9}$ (1.03)	$1.20x10^{-5}$ (1.04)	$2.65x10^{-8}$ (1.04)	*ASCL2*[g]
rs3116590	13:50,808-50,808	1/1	0.21	G/A	$1.52x10^{-8}$ (1.04)	0.0064 (1.03)	$1.00x10^{-8}$ (1.04)	*DLEU1*[g]
rs61975764	**14:93,014-93,014**	**1/1**	**0.47**	**A/G**	**$2.65x10^{-8}$ (1.03)**	**$1.46x10^{-7}$ (1.04)**	**$7.84x10^{-7}$ (1.03)**	***RIN3***[g]
rs4381563	15:75,275-75,448	19/1	0.33	A/T	$7.81x10^{-9}$ (0.97)	0.00027 (0.97)	$3.37x10^{-9}$ (0.96)	*PPCDC*[g]
rs12956924	**18:46,451-46,451**	**1/1**	**0.31**	**A/G**	**$1.52x10^{-8}$ (1.03)**	**$2.64x10^{-5}$ (1.04)**	**$2.09x10^{-7}$ (1.03)**	***SMAD7***[g]
rs10419921	**19:16,412-16,412**	**1/1**	**0.30**	**T/C**	**$1.67x10^{-8}$ (1.03)**	**$5.62x10^{-6}$ (1.04)**	**$1.09x10^{-6}$ (1.03)**	***KLF2***[g]
rs6066184	20:45,228-45,716	26/2	0.26	G/C	$1.19x10^{-13}$ (0.95)	$4.27x10^{-6}$ (0.96)	$6.55x10^{-16}$ (0.95)	*EYA2*[d]

Loci found only when analysing all three diseases as one phenotype are marked as bold. More details can be found in S3, S12-S13 and S15 Tables.

[a] Defined as SNPs located<1 Mb apart containing at least one significantly associated genetic variant at $P \leq 3x10^{8}$.

[b] Total number of SNPs with $P \leq 3x10^{-8}$ within loci.

[c] Total number of independent associations within the locus, based on conditional analysis (17).

[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 (22).

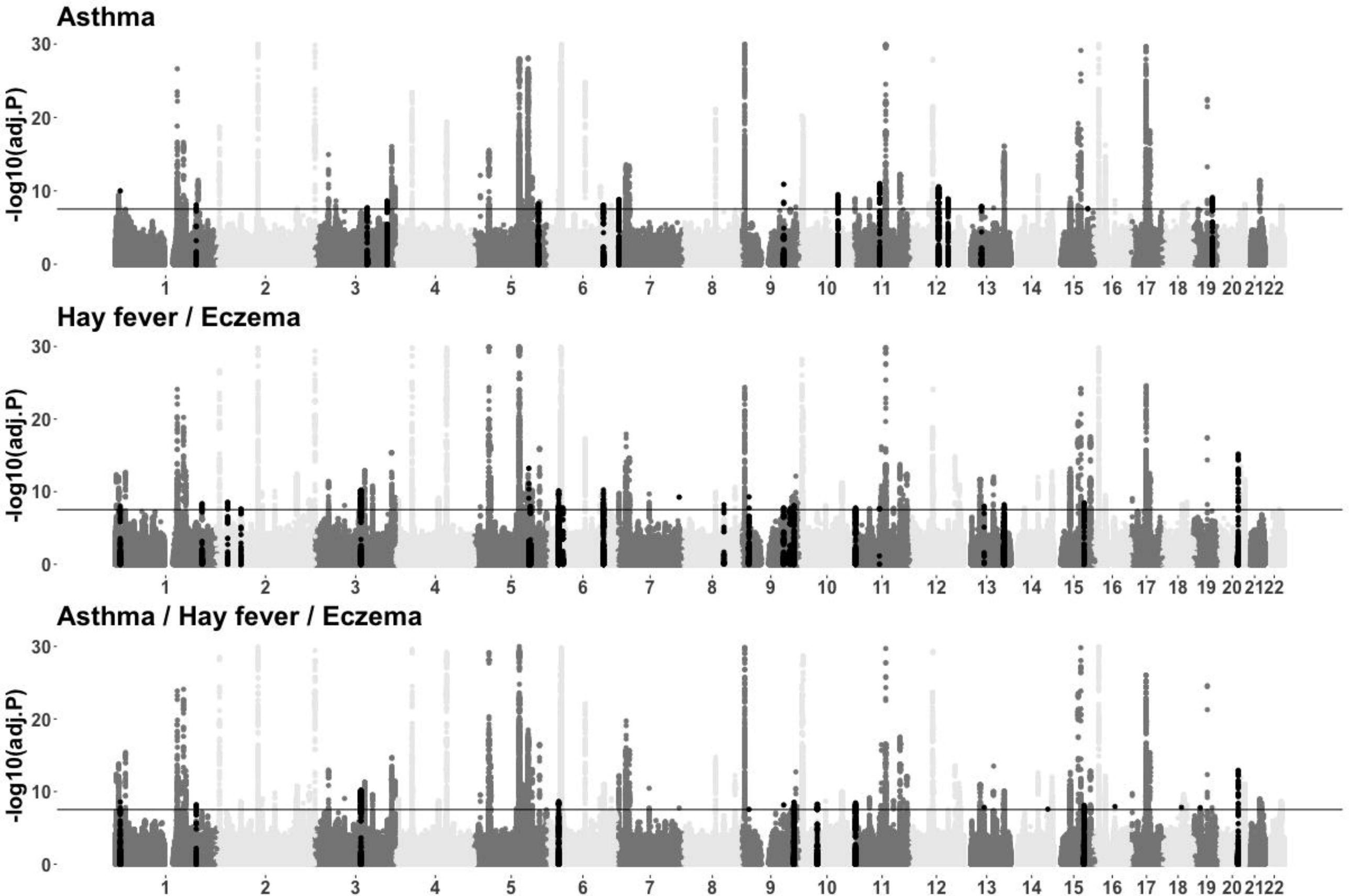
			Prevalence used in LD-score[a]		(A) All SNPs		(B) Without significant loci		h2 explained by significant loci
Phenotype	**N cases**	**N controls**	**Population**	**Sample**	**h2**	**SE**	**h2**	**SE**	**Absolute terms: (A) - (B)**
Asthma	41,926	239,751	0.117	0.117	0.210	0.017	0.168	0.010	0.042
Hay fever/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

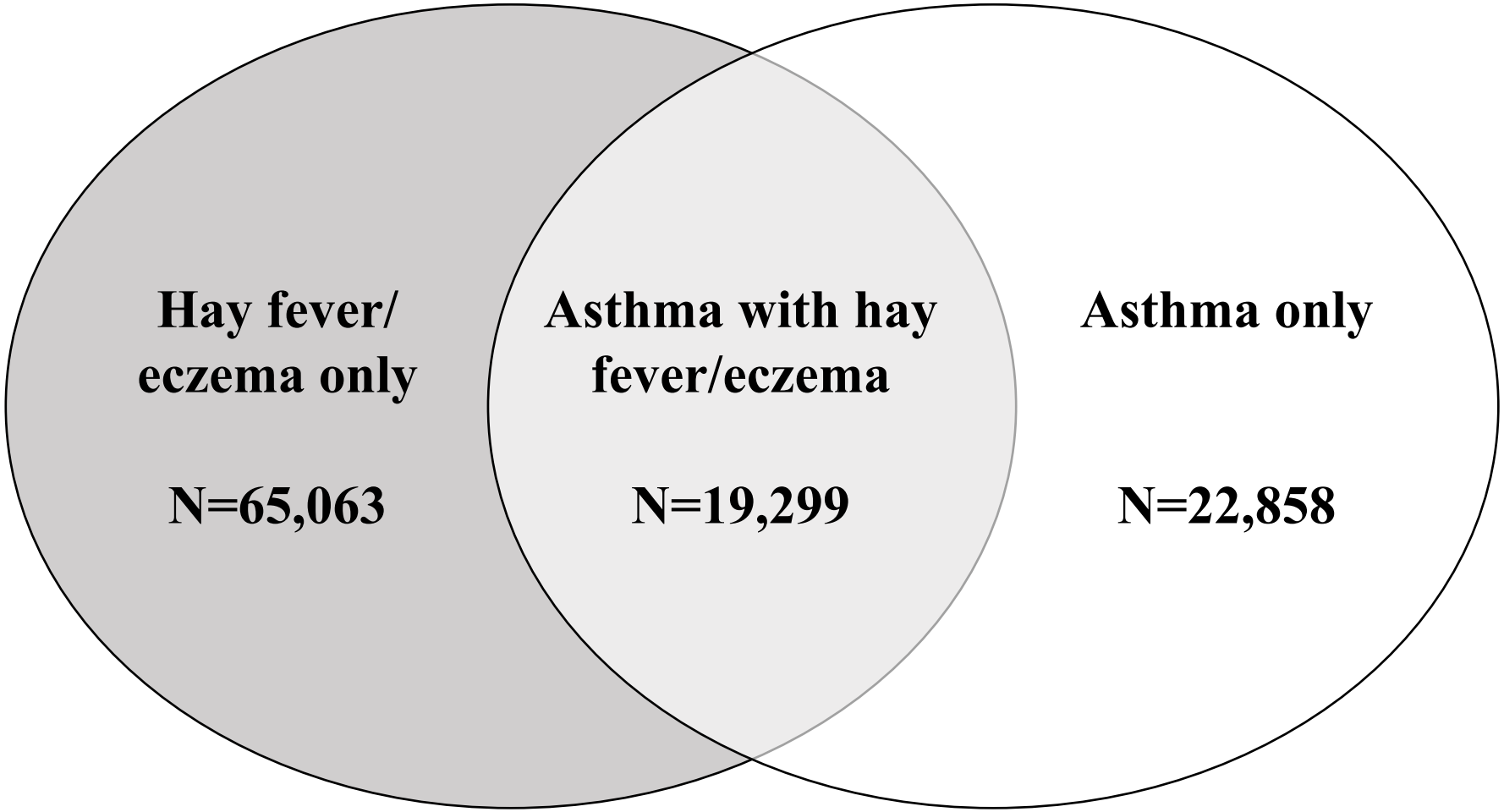
[a]LDSC requires values for population and sample prevalence when estimating SNP heritability. In these analyses we used the prevalence in the UK Biobank cohort (Table 1) as the sample prevalence. However, due to the cross-sectional design of the UK Biobank we used the same values as population prevalence.

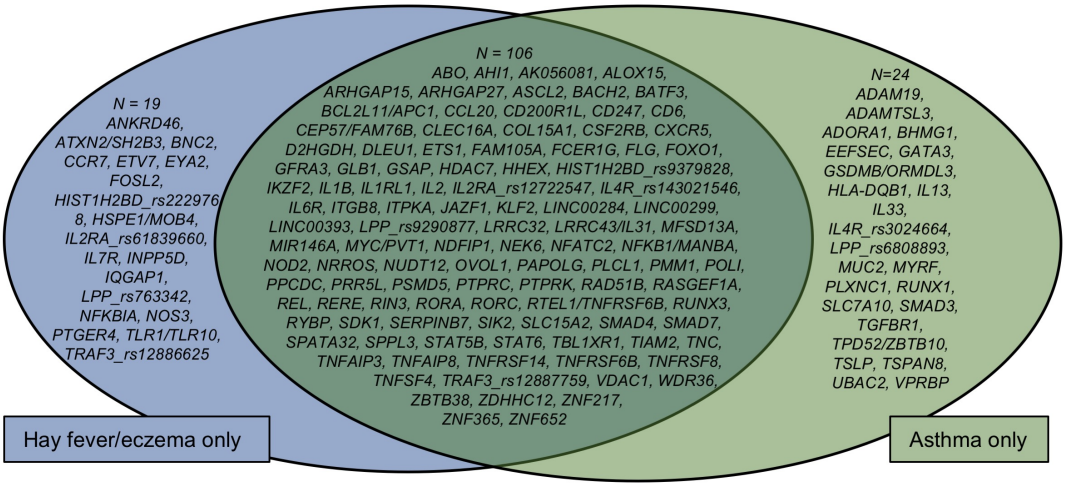
Abbreviations

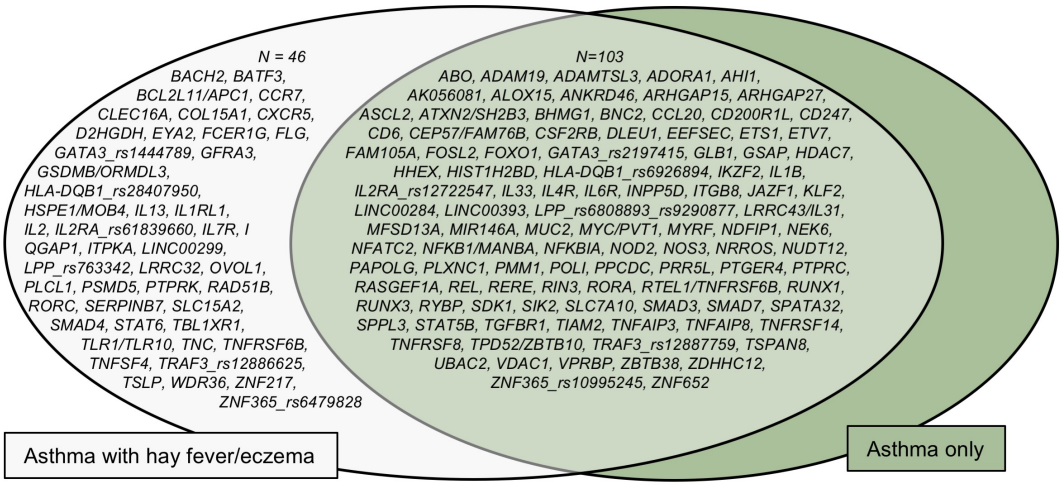
SNP Single nucleotide polymorphism

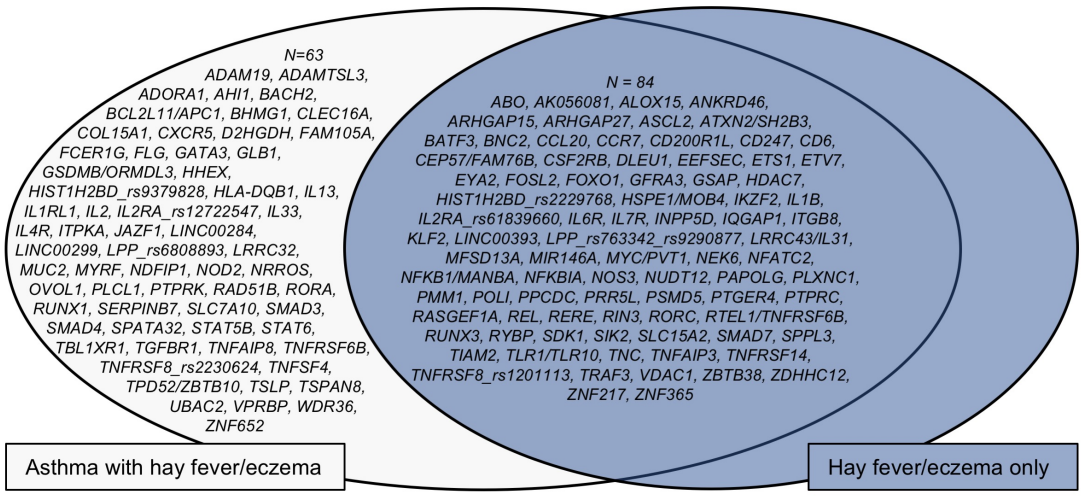
GWAS Genome Wide Association Study


Asthma
Hay fever / Eczema
Asthma / Hay fever / Eczema
-log10(adj.P)
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22


Hay fever/
eczema only
N=65,063
Asthma with hay
fever/eczema
N=19,299
Asthma only
N=22,858


N = 19
ANKRD46, ATXN2/SH2B3, BNC2, CCR7, ETV7, EYA2, FOSL2, HIST1H2BD_rs2229768, HSPE1/MOB4, IL2RA_rs61839660, IL7R, INPP5D, IQGAP1, LPP_rs763342, NFKBIA, NOS3, PTGER4, TLR1/TLR10, TRAF3_rs12886625
N = 106
ABO, AHI1, AK056081, ALOX15, ARHGAP15, ARHGAP27, ASCL2, BACH2, BATF3, BCL2L11/APC1, CCL20, CD200R1L, CD247, CD6, CEP57/FAM76B, CLEC16A, COL15A1, CSF2RB, CXCR5, D2HGDH, DLEU1, ETS1, FAM105A, FCER1G, FLG, FOXO1, GFRA3, GLB1, GSAP, HDAC7, HHEX, HIST1H2BD_rs9379828, IKZF2, IL1B, IL1RL1, IL2, IL2RA_rs12722547, IL4R_rs143021546, IL6R, ITGB8, ITPKA, JAZF1, KLF2, LINC00284, LINC00299, LINC00393, LPP_rs9290877, LRRC32, LRRC43/IL31, MFSD13A, MIR146A, MYC/PVT1, NDFIP1, NEK6, NFATC2, NFKB1/MANBA, NOD2, NRROS, NUDT12, OVOL1, PAPOLG, PLCL1, PMM1, POLI, PPCDC, PRR5L, PSMD5, PTPRC, PTPRK, RAD51B, RASGEF1A, REL, RERE, RIN3, RORA, RORC, RTEL1/TNFRSF6B, RUNX3, RYBP, SDK1, SERPINB7, SIK2, SLC15A2, SMAD4, SMAD7, SPATA32, SPPL3, STAT5B, STAT6, TBL1XR1, TIAM2, TNC, TNFAIP3, TNFAIP8, TNFRSF14, TNFRSF6B, TNFRSF8, TNFSF4, TRAF3_rs12887759, VDAC1, WDR36, ZBTB38, ZDHHC12, ZNF217, ZNF365, ZNF652
N=24
ADAM19, ADAMTSL3, ADORA1, BHMG1, EEFSEC, GATA3, GSDMB/ORMDL3, HLA-DQB1, IL13, IL33, IL4R_rs3024664, LPP_rs6808893, MUC2, MYRF, PLXNC1, RUNX1, SLC7A10, SMAD3, TGFBR1, TPD52/ZBTB10, TSLP, TSPAN8, UBAC2, VPRBP
Hay fever/eczema only
Asthma only


N = 46
BACH2, BATF3, BCL2L11/APC1, CCR7, CLEC16A, COL15A1, CXCR5, D2HGDH, EYA2, FCER1G, FLG, GATA3_rs1444789, GFRA3, GSDMB/ORMDL3, HLA-DQB1_rs28407950, HSPE1/MOB4, IL13, IL1RL1, IL2, IL2RA_rs61839660, IL7R, I QGAP1, ITPKA, LINC00299, LPP_rs763342, LRRC32, OVOL1, PLCL1, PSMD5, PTPRK, RAD51B, RORC, SERPINB7, SLC15A2, SMAD4, STAT6, TBL1XR1, TLR1/TLR10, TNC, TNFRSF6B, TNFSF4, TRAF3_rs12886625, TSLP, WDR36, ZNF217, ZNF365_rs6479828
N=103
ABO, ADAM19, ADAMTSL3, ADORA1, AHI1, AK056081, ALOX15, ANKRD46, ARHGAP15, ARHGAP27, ASCL2, ATXN2/SH2B3, BHMG1, BNC2, CCL20, CD200R1L, CD247, CD6, CEP57/FAM76B, CSF2RB, DLEU1, EEFSEC, ETS1, ETV7, FAM105A, FOSL2, FOXO1, GATA3_rs2197415, GLB1, GSAP, HDAC7, HHEX, HIST1H2BD, HLA-DQB1_rs6926894, IKZF2, IL1B, IL2RA_rs12722547, IL33, IL4R, IL6R, INPP5D, ITGB8, JAZF1, KLF2, LINC00284, LINC00393, LPP_rs6808893_rs9290877, LRRC43/IL31, MFSD13A, MIR146A, MUC2, MYC/PVT1, MYRF, NDFIP1, NEK6, NFATC2, NFKB1/MANBA, NFKBIA, NOD2, NOS3, NRROS, NUDT12, PAPOLG, PLXNC1, PMM1, POLI, PPCDC, PRR5L, PTGER4, PTPRC, RASGEF1A, REL, RERE, RIN3, RORA, RTEL1/TNFRSF6B, RUNX1, RUNX3, RYBP, SDK1, SIK2, SLC7A10, SMAD3, SMAD7, SPATA32, SPPL3, STAT5B, TGFBR1, TIAM2, TNFAIP3, TNFAIP8, TNFRSF14, TNFRSF8, TPD52/ZBTB10, TRAF3_rs12887759, TSPAN8, UBAC2, VDAC1, VPRBP, ZBTB38, ZDHHC12, ZNF365_rs10995245, ZNF652
Asthma with hay fever/eczema
Asthma only


N=63
ADAM19, ADAMTSL3, ADORA1, AHI1, BACH2, BCL2L11/APC1, BHMG1, CLEC16A, COL15A1, CXCR5, D2HGDH, FAM105A, FCER1G, FLG, GATA3, GLB1, GSDMB/ORMDL3, HHEX, HIST1H2BD_rs9379828, HLA-DQB1, IL13, IL1RL1, IL2, IL2RA_rs12722547, IL33, IL4R, ITPKA, JAZF1, LINC00284, LINC00299, LPP_rs6808893, LRRC32, MUC2, MYRF, NDFIP1, NOD2, NRROS, OVOL1, PLCL1, PTPRK, RAD51B, RORA, RUNX1, SERPINB7, SLC7A10, SMAD3, SMAD4, SPATA32, STAT5B, STAT6, TBL1XR1, TGFBR1, TNFAIP8, TNFRSF6B, TNFRSF8_rs2230624, TNFSF4, TPD52/ZBTB10, TSLP, TSPAN8, UBAC2, VPRBP, WDR36, ZNF652
N = 84
ABO, AK056081, ALOX15, ANKRD46, ARHGAP15, ARHGAP27, ASCL2, ATXN2/SH2B3, BATF3, BNC2, CCL20, CCR7, CD200R1L, CD247, CD6, CEP57/FAM76B, CSF2RB, DLEU1, EEFSEC, ETS1, ETV7, EYA2, FOSL2, FOXO1, GFRA3, GSAP, HDAC7, HIST1H2BD_rs2229768, HSPE1/MOB4, IKZF2, IL1B, IL2RA_rs61839660, IL6R, IL7R, INPP5D, IQGAP1, ITGB8, KLF2, LINC00393, LPP_rs763342_rs9290877, LRRC43/IL31, MFSD13A, MIR146A, MYC/PVT1, NEK6, NFATC2, NFKB1/MANBA, NFKBIA, NOS3, NUDT12, PAPOLG, PLXNC1, PMM1, POLI, PPCDC, PRR5L, PSMD5, PTGER4, PTPRC, RASGEF1A, REL, RERE, RIN3, RORC, RTEL1/TNFRSF6B, RUNX3, RYBP, SDK1, SIK2, SLC15A2, SMAD7, SPPL3, TIAM2, TLR1/TLR10, TNC, TNFAIP3, TNFRSF14, TNFRSF8_rs1201113, TRAF3, VDAC1, ZBTB38, ZDHHC12, ZNF217, ZNF365
Asthma with hay fever/eczema
Hay fever/eczema only