



Published in final edited form as:

*J Allergy Clin Immunol.* 2019 February ; 143(2): 691–699. doi:10.1016/j.jaci.2018.03.012.

## Eleven loci with new reproducible genetic associations with allergic disease risk

Manuel AR Ferreira, PhD<sup>1</sup>, Judith M Vonk, PhD<sup>2</sup>, Hansjörg Baurecht, PhD<sup>3</sup>, Ingo Marenholz, PhD<sup>4,5</sup>, Chao Tian, PhD<sup>6</sup>, Joshua D Hoffman, PhD<sup>7</sup>, Quinta Helmer, PhD<sup>8</sup>, Annika Tillander, PhD<sup>9</sup>, Vilhelmina Ullemar, PhD<sup>9</sup>, Yi Lu, PhD<sup>9</sup>, Franz Rüschemendorf, PhD<sup>4</sup>, the 23andMe Research Team<sup>6</sup>, collaborators of the SHARE study<sup>10</sup>, David A Hinds, PhD<sup>6</sup>, Norbert Hübner, MD<sup>4</sup>, Stephan Weidinger, MD<sup>3</sup>, Patrik KE Magnusson, PhD<sup>9</sup>, Eric Jorgenson, PhD<sup>11</sup>, Young-Ae Lee, MD<sup>4,5</sup>, Dorret I Boomsma, PhD<sup>8</sup>, Robert Karlsson, PhD<sup>9</sup>, Catarina Almqvist, MD<sup>9,12</sup>, Gerard H Koppelman, MD<sup>13</sup>, Lavinia Paternoster, PhD<sup>14</sup>

<sup>1</sup>Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Australia <sup>2</sup>Epidemiology, University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands <sup>3</sup>Department of Dermatology, Allergology and Venereology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany <sup>4</sup>Max Delbrück Center (MDC) for Molecular Medicine, Berlin, Germany <sup>5</sup>Clinic for Pediatric Allergy, Experimental and Clinical Research Center of Charité Universitätsmedizin Berlin and Max Delbrück Center, Berlin, Germany <sup>6</sup>23andMe, Inc., Mountain View, California, USA <sup>7</sup>Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA <sup>8</sup>Department Biological Psychology, Netherlands Twin Register, Vrije University, Amsterdam, The Netherlands <sup>9</sup>Department of Medical Epidemiology and Biostatistics and the Swedish Twin Registry, Karolinska Institutet, Stockholm, Sweden <sup>11</sup>Division of Research, Kaiser Permanente Northern California, Oakland, California, USA <sup>12</sup>Pediatric Allergy and Pulmonology Unit at Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden <sup>13</sup>University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Pediatric Pulmonology and Pediatric Allergology, and University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands <sup>14</sup>MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK

### Abstract

**Background**—A recent genome-wide association study (GWAS) identified 99 loci that contain genetic risk variants shared between asthma, hay fever and eczema. Many more risk loci shared

---

Corresponding author: Manuel A R Ferreira, PhD, QIMR Berghofer Medical Research Institute, Locked Bag 2000, Royal Brisbane Hospital, Herston QLD 4029, Australia, Phone: +61 7 3845 3552, Fax: +61 7 3362 0101, manuel.ferreira@qimrberghofer.edu.au.

<sup>10</sup>Collaborators of the SHARE study are listed in this article's Online Repository

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

between these common allergic diseases remain to be discovered, which could point to new therapeutic opportunities.

**Objective**—To identify novel risk loci shared between asthma, hay fever and eczema by applying a gene-based test of association to results from a published GWAS that included data from 360,838 individuals.

**Methods**—We used approximate conditional analysis to adjust the results from the published GWAS for the effects of the top risk variants identified in that study. We then analysed the adjusted GWAS results with the EUGENE gene-based approach, which combines evidence for association with disease risk across regulatory variants identified in different tissues. Novel gene-based associations were followed up in an independent sample of 233,898 individuals from the UK Biobank study.

**Results**—Of the 19,432 genes tested, 30 had a significant gene-based association at a Bonferroni-corrected  $P$ -value of  $2.5 \times 10^{-6}$ . Of these, 20 were also significantly associated ( $P < 0.05/30 = 0.0016$ ) with disease risk in the replication sample, including 19 that were located in 11 loci not reported to contain allergy risk variants in previous GWAS. Amongst these were nine genes with a known function that is directly relevant to allergic disease: *FOSL2*, *VPRBP*, *IPCEF1*, *PRR5L*, *NCF4*, *APOBR*, *IL27*, *ATXN2L* and *LAT*. For four genes (e.g. *ATXN2L*), a genetically-determined decrease in gene expression was associated with decreased allergy risk, and therefore drugs that inhibit gene expression or function are predicted to ameliorate disease symptoms. The opposite directional effect was observed for 14 genes, including *IL27*, a cytokine known to suppress Th2 responses.

**Conclusion**—Using a gene-based approach, we identified 11 risk loci for allergic disease that were not reported in previous GWAS. Functional studies that investigate the contribution of the 19 associated genes to the pathophysiology of allergic disease and assess their therapeutic potential are warranted.

## INTRODUCTION

The strong genetic correlations between asthma, hay fever and eczema estimated from twin studies<sup>1–5</sup>, in combination with the highly polygenic architecture of these diseases, predict that many hundreds if not thousands of genetic risk factors are shared between these three common allergic diseases. Motivated by this prediction, we recently performed a genome-wide association study (GWAS) designed to identify genetic risk variants that are shared between asthma, hay fever and eczema<sup>6</sup>. In that GWAS, cases ( $n=180,129$ ) were defined as individuals who reported having suffered from one or more allergic disease, while controls ( $n=180,709$ ) were individuals who reported never having suffered from any of these diseases. We identified 136 single nucleotide polymorphisms (SNPs) located in 99 loci (*i.e.* genomic regions located  $> 1\text{Mb}$  apart) that were independently associated with disease risk at a genome-wide significance-threshold of  $P < 3 \times 10^{-8}$ , a threshold that corrects for the number of SNPs tested<sup>7</sup>. In our study, we often observed that multiple independent genetic variants within one locus contributed to disease risk – this was the case for 18 of the 99 risk loci identified.

Larger GWAS of this multi-disease phenotype are underway and are expected to identify more risk variants shared between asthma, hay fever and eczema. Here, we report results from another approach that increases power to identify novel risk loci: gene-based instead of SNP-based association analysis. Different gene-based tests have been developed, including VEGAS<sup>8</sup>, MAGMA<sup>9</sup> and fastBAT<sup>10</sup>. For each gene in the genome, these methods combine in a single test the evidence for association with a disease across multiple SNPs, which are typically selected because they are located in or near that gene (*e.g.* within 100 kb). These methods improve power over the alternative approach of testing each SNP at a time (as is done in a GWAS) when multiple SNPs near a gene are independently associated with disease risk. However, not all SNPs near a gene are directly relevant to its function. For example, only some SNPs influence variation in gene expression levels – these are commonly referred to as expression quantitative trait loci (eQTL). On the other hand, eQTL have a greater probability of being associated with common diseases and traits<sup>11</sup>. Motivated by these observations, an additional suite of gene-based methods has been developed recently that only include in the association test functional SNPs, such as eQTL. These include, for example, PrediXcan<sup>12</sup>, EUGENE<sup>13</sup> and S-PrediXcan<sup>14</sup>. We developed EUGENE because it was not possible with other methods to combine in the same association test information from eQTL identified in different tissues. This feature is important because multiple tissue types play a role in allergic disease pathophysiology and tissue-specific eQTL are common<sup>15</sup>. Furthermore, EUGENE also considers regulatory variants with different mechanisms of action, for example, variants that affect splicing but not overall transcription levels. This increased resolution is expected to increase our ability to identify genes that are causally-related to common diseases<sup>16</sup>.

The aims of the present study were to (1) identify novel risk loci shared between asthma, hay fever and eczema by applying EUGENE to results from our previous GWAS<sup>6</sup>; and (2) follow up the top gene-based associations in an independent replication sample ascertained from the UK Biobank study<sup>17</sup>.

## METHODS

### Adjusting GWAS results for the effects of genome-wide significant SNPs

The starting point for this study was a GWAS of allergic disease reported recently by Ferreira et al.<sup>6</sup>, which included 360,838 individuals from 13 studies: UK Biobank, 23andMe, GERA, CATSS, NTR, LifeLines, TWINGENE, ALSPAC, SALT, GENEVA, AAGC, GENUFAD-SHIP-1 and GENUFAD-SHIP-2 (demographics in Table E1 in this article's Online Repository). In that study, single-SNP results were corrected for an inflation factor that reflected technical biases and/or population stratification (specifically, an LD-score regression intercept<sup>18</sup> of 1.04). After that correction, there were 136 variants independently associated with allergic disease in Ferreira et al.<sup>6</sup>. Because our aim was to identify new allergy risk loci, we first applied approximate conditional analyses as implemented in the GCTA tool<sup>19</sup> to the summary statistics of Ferreira et al.<sup>6</sup> to adjust the single-SNP results for the effects of the 136 risk variants (a detailed justification for performing conditional analysis prior to the gene-based analysis is provided in the Online

Repository). In the conditional analysis, linkage disequilibrium (LD) between SNPs was estimated using genotype data from 5,000 individuals from the UK Biobank study<sup>17</sup>.

### Gene-based analysis of the adjusted GWAS results

To identify novel risk loci shared between asthma, hay fever and eczema, we analysed the adjusted GWAS results with EUGENE<sup>13</sup>, a gene-based approach that is applicable to summary statistics (*i.e.* it does not require individual-level genetic data for subjects included in the GWAS) and combines evidence for association with disease risk across eQTL identified in different tissues. The latter feature is important because multiple tissue types play a role in allergic disease pathophysiology and tissue-specific eQTL are common<sup>15</sup>.

We identified eQTL based on information from 39 published eQTL studies conducted in 19 tissues or cell types relevant to allergic disease (Table E2 in this article's Online Repository). For each eQTL study, we (i) downloaded the original publication tables containing results for the eQTL reported; (ii) extracted the SNP identifier, gene name, association P-value and directional effect (if available; beta/z-score and effect allele) for all reported eQTL; and (iii) excluded eQTL located >1 Mb of the respective gene (*i.e.* *trans* eQTL) and/or with an association  $P > 2.3 \times 10^{-9}$ , a conservative threshold that corrects for 21,742 genes in the genome, each tested for association with 1,000 SNPs (as suggested by others<sup>20-22</sup>). We did not include *trans* eQTLs in the analysis because often these are thought to involve indirect effects<sup>23</sup>, for example, where a SNP influences the expression of a gene in *cis*, which in turn affects the expression of many other genes in *trans*.

Having identified a list of *cis* eQTL for a given gene from published studies, we then reduced that list to a set of eQTL in low LD with each other (specifically, with an  $r^2 < 0.1$ ) using the --clump procedure in PLINK v1.90<sup>24</sup>. For a given gene, we refer to these as "independent eQTL", although we recognise that some pairs of eQTL will not be in linkage equilibrium. LD was estimated based on genetic data from individuals of European descent from the 1000 Genomes Project 23 (n=294, release 20130502\_v5a). Clumping was not performed separately for each tissue or study, but rather applied to the overall list of eQTL obtained after considering information from all tissues/studies. If an eQTL was identified in multiple tissues/studies, the clumping procedure was performed using the smallest P-value reported for that eQTL across all tissues/studies. A file (ASTHMA.20170517.eqtl.proxies.list) containing the independent eQTL identified per gene is available at <https://genepi.qimr.edu.au/staff/manuelF/eugene/main.html>.

For each gene, EUGENE extracts single-SNP association results for each independent eQTL (or, if not available, for a proxy with  $r^2 > 0.8$ ) from the GWAS summary statistics, sums the association chi-square values across those eQTL, and estimates the significance of the resulting sum test statistic using Satterthwaite's approximation, which accounts for the LD between eQTL. This approximation was originally implemented by Bakshi et al.<sup>10</sup> in the GCTA-fastBAT module and is now also available in EUGENE. LD between eQTL was estimated based on data from 294 Europeans from the 1000 Genomes Project (release 20130502\_v5a). The significance threshold required to achieve experiment-wide significance was set at  $P < 0.05/N$  genes tested.

## Replication of significant gene-based associations in an independent sample

To confirm novel gene-based associations, we analysed an independent sample of unrelated individuals of European descent from the UK Biobank study<sup>17</sup>. The approach used to select individuals for analysis was very similar to that described in detail previously<sup>6</sup>. Briefly, we (1) downloaded array (805,426 variants) and imputed (92,693,895 variants) genetic data for the entire UK Biobank study, comprising 488,377 individuals, in June 2017; (2) pruned the array data down to a set of 29,446 independent SNPs that had comparable ( $P > 0.005$ ) allele frequencies between Europeans of the 1000 Genomes Project (CEU, FIN, GBR and TSI groups) and the UK Biobank individuals of European descent included in the Ferreira et al. GWAS<sup>6</sup>; (3) merged the pruned dataset with data from 1,092 individuals of known ancestry from the 1000 Genomes Project (release 20130502\_v5a); (4) performed multi-dimensional scaling (MDS) analysis of identity-by-state allele sharing separately for each of 32 groups of ~16,000 individuals (to be computationally feasible), including those from the 1000 Genomes Project; (5) identified and removed individuals who did not cluster closely (within 5 standard deviations of MDS components 1 and 2) to individuals of European ancestry from the 1000 Genomes Project, resulting in 461,885 individuals; (6) identified and removed any individuals included in, or related to (*i.e.* with a kinship coefficient that indicates 3rd degree relatedness or closer; based on file `ukb1007_rel_s488374.dat`), the 138,354 individuals of the UK Biobank study included in Ferreira et al.<sup>6</sup>, as well as individuals (i) with genetically-inferred sex different from self-reported sex, (ii) who were outliers for SNP missingness or genome-wide heterozygosity levels, and/or (iii) with >10 third-degree relatives, resulting in 244,395 individuals.

For each individual, allergic disease status was defined as previously described for the UK Biobank study in detail<sup>6</sup>. To classify asthma status, we combined information from three sources: (1) touchscreen questionnaire (data-field 6152); (2) Non-cancer illness code, self-reported during verbal interview (data-field 20002); (3) main (data-field 41202) and secondary (data-field 41204) ICD10 diagnoses. Specifically, inclusion criteria for cases were: (i) a report of “Asthma” in field 6152 and a code for asthma (1111) in field 20002; or (ii) an ICD10 code for asthma in fields 41202 or 41204. Exclusion criteria for cases were: (i) a report of COPD in fields 6152 or 20002; or (ii) other self-reported respiratory diseases in field 20002. Inclusion criteria for controls were no report of asthma in fields 6152, 20002, 41202 and 41204. Exclusion criteria for controls were the same as for cases (no COPD or other self-reported respiratory diseases). To classify hay fever status, we used the same three sources of information. Specifically, inclusion criteria for cases were: (i) a report of “Hay fever, allergic rhinitis or eczema” in field 6152 and a code for hay fever (1387) in field 20002; or (ii) an ICD10 code for hay fever in fields 41202 or 41204. Inclusion criteria for controls were no report of hay fever in fields 6152, 20002, 41202 and 41204. The eczema phenotype was created very similarly to the hay fever phenotype. Inclusion criteria for cases were: (i) a report of “Hay fever, allergic rhinitis or eczema” in field 6152 and a code for eczema (1452) in field 20002; or (ii) an ICD10 code for eczema in fields 41202 or 41204. Inclusion criteria for controls were no report of eczema in fields 6152, 20002, 41202 and 41204. To create the overall allergic disease phenotype used for analysis, cases were individuals classified as suffering from at least one condition (asthma and/or hay fever and/or eczema), as described above. Controls were individuals classified as not having

suffered from all three conditions. Using this approach, the 244,395 individuals selected for analysis (see above) included 71,807 allergic disease cases, 162,091 allergic disease controls and 10,497 individuals with a missing phenotype.

SNPTEST was then used to test the association between disease status and imputed genotype data for eQTL of genes selected for replication; age, sex and SNP chip were included as covariates. We only analysed SNPs imputed based on the Haplotype Reference Consortium panel, given that variants imputed from the UK10K + 1000 Genomes panel were not mapped correctly. We also tested the association with 1.2 million HapMap3 SNPs in order to estimate the degree of inflation in test statistics arising because of unaccounted technical biases, using the LD Score approach<sup>25</sup>. We observed an LD Score intercept of 1.09, which was used to adjust the association results for the eQTL tested. Lastly, EUGENE was used as described above to perform the gene-based analysis for all genes selected for replication.

### Association analyses contrasting individuals suffering from a single allergic disease

The case-control phenotype analysed in our GWAS<sup>6</sup> combined information from asthma, hay fever and eczema, and so was expected to improve power to identify risk variants shared between, but not specific to any of, the three diseases<sup>26</sup>. To understand if gene-based associations discovered through the analysis of that multiple-disease phenotype were indeed likely to represent risk factors shared across allergic diseases, we performed case-only association analyses as described in detail previously<sup>6</sup>. First, we tested the association between eQTL of selected genes with three phenotypes that contrasted three non-overlapping groups of adults who suffered from a single allergic disease: asthma only cases (g1;  $n=12,268$ ) vs. hay fever only cases (g2;  $n=33,305$ ); asthma only cases (g1) vs. eczema only cases (g3;  $n=6,276$ ); and hay fever only cases (g2) vs. eczema only cases (g3). For a given eQTL, results from these analyses indicate if the risk allele is more (odds ratio [OR]  $>1$ ) or less (OR $<1$ ) common in *e.g.* group 1 (g1) when compared to group 2 (g2). For example, if an eQTL contributed similarly to the risks of asthma and hay fever but not eczema, then one would expect an OR $\sim 1$  in the asthma only vs. hay fever only comparison, but an OR $>1$  in the asthma vs. eczema and hay fever vs. eczema analyses. Second, for each gene and phenotype tested, we combined the association results across eQTL using EUGENE as we did in the analysis of the adjusted GWAS results described above.

This study was approved by the Human Ethics Committee of the QIMR Berghofer Medical Research Institute.

## RESULTS

### Identification of novel risk loci for allergic disease through gene-based association analysis

An overview of the analytical approach used is shown in Figure 1. To identify novel risk loci shared between asthma, hay fever and eczema, we first adjusted the association results from Ferreira et al.<sup>6</sup> for the 136 genome-wide significant SNPs (*i.e.* with  $P<3\times 10^{-8}$ ) identified in that study using approximate conditional analysis (Figure 2A). In the resulting adjusted

GWAS, as expected there were no SNPs associated with disease risk at  $P < 3 \times 10^{-8}$  and located  $> 1 \text{ Mb}$  from the loci reported in Ferreira et al. <sup>6</sup>. On the other hand, four SNPs located in loci reported in that study (in/near *CADM3*, *SLC39A8*, *LRRC43* and *KLF5*) were genome-wide significant in the conditional but not in the original analyses, consistent with the presence of additional secondary association signals at those established risk loci (Figure 2B). Importantly, there was an enrichment in significant SNP associations in the adjusted GWAS (Figure E1 in the Online Repository), suggesting that many of these associations are likely to represent true-positive findings.

To identify loci that were likely to contain true-positive associations, we applied the EUGENE gene-based approach <sup>13</sup> to the adjusted GWAS results. Specifically, we tested the association between disease risk and 19,432 genes (or other types of transcripts, such as long non-coding RNAs [lncRNA]) that were reported to have one or more independent eQTL in 19 tissues or cell types relevant to allergic disease (Table E2 in this article's Online Repository), including whole blood, lung, skin and individual immune cell types.

We identified 30 significant gene-based associations at a Bonferroni-corrected  $P$ -value of  $2.5 \times 10^{-6}$ , which were located in 18 loci (Table 1). The specific eQTL included in the gene-based test for each of these 30 genes are listed in Table E3 in this article's Online Repository. For 21 genes, the association  $P$ -value obtained with the gene-based test was more significant than the  $P$ -value obtained with the individual eQTL most associated with disease risk (Figure 3), indicating that multiple eQTL for the same gene were associated with disease risk (range 2 to 7 associated eQTL per gene; Table 1). For eight genes, the difference in significance between the most associated individual eQTL and the gene-based test exceeded one order of magnitude (Figure 3). The most extreme example of this was the *SPNS1* gene on chromosome 16p11.2 (gene-based  $P = 3.5 \times 10^{-9}$  versus best individual eQTL  $P = 4.8 \times 10^{-6}$ ), for which 6 of the 15 eQTL tested (identified in five tissues) were nominally associated with disease risk (Table E4 in this article's Online Repository).

### Replication of significant gene-based associations in an independent sample

Next, we performed a replication study to determine which of the 30 significant gene-based associations were likely to represent true-positive findings. To this end, we first identified 71,807 cases and 162,091 controls genotyped by the UK Biobank study<sup>17</sup> who were unrelated to individuals from our initial GWAS <sup>6</sup>. We then used EUGENE to test the association in this independent sample between case-control status and each of the 30 genes identified above.

There were 20 significant gene-based associations at a conservative significance threshold that accounts for the 30 genes tested ( $P < 0.05/30 = 0.0016$ ; Table 1). These included 19 genes located in 11 loci not implicated in allergic disease in previous GWAS: *OR10J5* (chromosome 1q23.2); *RP11-534L20.5* (1q32.1); *FOSL2* (2p23.2); *RBM15B* and *VPRBP* (3p21.2); *IPCEF1* (6q25.2); *AC004893.11* (7q22.1); *PRR5L* (11p13); *NSMCE1* (16p12.1); *SPNS1* and seven other nearby genes (16p11.2); *PVALB* and *NCF4* (22q12.3). Of note, nine of these genes have a known function that is directly relevant to the pathophysiology of allergic disease (Table 2).

### Predicted directional effect of gene expression on disease risk

Because the gene-based approach used focuses exclusively on eQTL, which in turn are associated with the expression of specific genes, we were able to identify in a single analysis both novel risk loci as well as the likely gene(s) underlying each association. Furthermore, often (but not always) the directional effect of an eQTL on gene expression can be obtained from published eQTL studies. Based on this information, for each gene, we determined if the allergy-protective allele of eQTL included in the gene-based test was associated with increased or decreased gene expression. This is important because drugs that mimic the directional effect of the allergy-protective allele on gene expression might be expected to attenuate (rather than exacerbate) allergic disease symptoms.

When we performed this analysis for the 20 genes with a significant ( $P < 0.0016$ ) association in the replication study, we found that for 14 genes the allergy protective allele was associated with increased gene expression (Table E5 in this article's Online Repository). This includes, for example, the long non-coding RNA *RP11-534L20.5*, for which information from six different tissues (including blood, lung and skin) indicates that increased gene expression has a protective effect on disease risk. On the other hand, for four genes (*AC004893.11*, *ATXN2L*, *NSMCE1*, *RP11-24N18.1*) the allergy protective allele was associated with decreased gene expression, while for two genes eQTL directional effects were conflicting between studies (*IPCEF1*, *SULT1A1*).

### Association analyses contrasting individuals suffering from a single allergic disease

The multiple-disease case-control phenotype analysed in our GWAS<sup>6</sup> maximizes power to identify risk variants that are shared between asthma, hay fever and eczema<sup>26</sup>. As such, we expected the 11 novel risk loci identified above to have comparable effects on the three individual diseases. To address this possibility, we tested the association between the 20 genes in those 11 loci and three phenotypes that compared three non-overlapping groups of adults who suffered from a single allergic disease: (1) asthma only cases ( $n=12,268$ ) versus hay fever only cases ( $n=33,305$ ); (2) asthma only cases versus eczema only cases ( $n=6,276$ ); and (3) hay fever only cases versus eczema only cases. After correcting for the number of tests performed ( $P < 0.05 / (20 \text{ genes} \times 3 \text{ phenotypes}) = 0.0008$ ), no single gene had a significant association in the asthma versus hay fever, asthma versus eczema or hay fever versus eczema analyses (Table E6). Even at a nominal  $P < 0.05$ , which does not correct for multiple testing, only three genes (in two loci) had a significant association in these analyses: *OR10J5*, *RP11-264B17.4* and *LAT*. These results indicate that most (if not all) of the 11 novel risk loci identified in this study do not have differential effects on the three individual diseases.

## DISCUSSION

By analyzing results from our previous GWAS<sup>6</sup> with a gene-based test of association, followed by replication of top findings in an independent sample, we identified 11 loci that contain previously unrecognized genetic risk variants for allergic disease. Results from case-only association analyses indicate that these loci have similar effects on asthma, hay fever and eczema risk.



The 11 novel risk loci were not reported in the original GWAS because they did not contain any single SNP associated with disease risk at a significance threshold that accounted for the number of SNPs tested. We were able to identify these loci in the present study for two main reasons. First, by testing individual genes rather than SNPs, the multiple testing burden was greatly reduced; this translated into a less stringent threshold required to declare genome-wide significance ( $P$ -value of  $2.5 \times 10^{-6}$  instead of  $3 \times 10^{-8}$ ), which increases power. Second, most of these loci (7 out of 11) contain genes for which multiple independent eQTL were individually associated with disease risk. Under this scenario, the gene-based test used improves power over the alternative approach of testing individual eQTL separately. Overall, our results support the use of gene-based eQTL-centric approaches to identify novel risk loci for human diseases and traits, as reported by others<sup>10, 27</sup>.

Our results point to 19 genes as being the likely candidates underlying the association between the 11 new loci and allergic disease risk. We stress, however, that functional studies are now required to confirm that the expression of these genes (1) is determined by (not simply associated with) the eQTL included in the respective gene-based test; and (2) influences disease pathophysiology. In other words, for given locus, we cannot exclude the possibility that a gene with a significant association is not causally-related to disease pathophysiology. Instead, for example, it might simply share eQTL with a nearby causal gene that was not identified in our analysis. We highlight one possible example of this. We observed a genome-wide significant association with a single gene in the 1q23.2 locus: the olfactory receptor gene *OR10J5* ( $P=2.3 \times 10^{-6}$ ). This gene is thought to regulate lipid accumulation<sup>28</sup>, and so could plausibly contribute to the pathophysiology of allergic disease. However, there were four additional genes within 1 Mb of *OR10J5* that had a statistically (albeit not genome-wide) significant gene-based association in our discovery analysis: *FCERIA* ( $P=1.2 \times 10^{-5}$ ), *DARC* ( $P=0.0005$ ), *IFI16* ( $P=0.009$ ) and *PYHIN1* ( $P=0.03$ ), the latter reported previously to contain asthma risk variants in populations of African ancestry<sup>29</sup>. The association with the first two replicated in the independent UKB sample:  $P=4.8 \times 10^{-7}$  and  $P=0.0001$ , respectively. Therefore, it is possible that *FCERIA* and/or *DARC* represent causal gene(s) underlying the observed association at this locus. Both are biologically plausible candidate allergy risk genes, encoding respectively the alpha subunit of the high-affinity IgE receptor and an atypical chemokine receptor<sup>30</sup>. Why did our analysis identify *OR10J5* and not, for example, *FCERIA*? The gene-based test for *OR10J5* included three eQTL (see Table E2 in this article's Online Repository), all individually associated ( $P < 0.05$ ) with disease risk; the strongest disease association was observed for eQTL rs2427838 (single-SNP  $P=1.0 \times 10^{-5}$ ). On the other hand, the gene-based test for *FCERIA* included these same three eQTL (*i.e.* these eQTL were shared between *OR10J5* and *FCERIA*) plus an additional six independent eQTL, none of which were individually associated with disease risk. Because the additional eQTL tested for *FCERIA* were not associated with disease risk, the resulting gene-based association was weaker when compared to *OR10J5*. An interesting question is why some but not all eQTL of *FCERIA* (and other genes) are associated with disease risk; it could be, for example, that disease associations are restricted to eQTL that influence gene expression in a specific subset of immune cells, or to eQTL that influence multiple relevant genes. Future studies that address this question are warranted.

With the caveat in mind that significant gene-based associations pinpoint causal risk loci but not necessarily the right causal gene(s), we note that nine of the 19 genes located in novel risk loci encode proteins with a known function that is directly relevant to allergic disease (Table 2). For example, *FOSL2* is involved in B cell and epidermal differentiation<sup>31, 32</sup>. Furthermore, it has a critical yet complex role in Th17 differentiation and function<sup>33</sup>: on the one hand, it represses Th17 signature genes (e.g. *IL17A*), but on the other hand, it promotes the expression of genes that drive Th17 maintenance and survival (e.g. *IL6R*)<sup>33</sup>. When we compared the directional effect of *FOSL2* eQTL between disease risk and gene expression, we found that the allele associated with reduced disease risk was associated with increased gene expression. These genetic findings suggest that increased *FOSL2* expression results in attenuated allergic inflammatory responses. Consistent with this possibility, deletion of a *FOSL2* repressor in mice decreased the capacity of CD4+ T cells to develop into the pro-inflammatory T follicular helper cell lineage following immunization with ovalbumin, viral infection, or in the context of low-grade chronic inflammation<sup>34</sup>. Based on these observations, we suggest that therapeutic strategies that increase *FOSL2* expression should be considered for the treatment of allergic diseases.

Another example of a gene identified in our analysis and previously implicated in the pathophysiology of allergic disease was *IL27*. This was one of eight genes identified on chromosome 16p11.2, a region that overlaps a large (~0.45 Mb) and common (49% frequency in Europeans) genomic inversion previously reported to be associated with the joint occurrence of asthma and obesity<sup>35</sup>. Of the two eQTL included in the gene-based test for *IL27*, one (rs7191548) is in high LD ( $r^2=0.72$ ) with a SNP that tags the inversion (rs4788101), suggesting that the association observed with *IL27* in our study is partly explained by that large structural variant. IL-27 has been shown to suppress Th2 responses<sup>36, 37</sup>, and so has been suggested as a potential new therapy for asthma. However, to our knowledge, no clinical trials have been performed to test this possibility. Our observation that, for both *IL27* eQTL tested, the allele associated with reduced disease risk was associated with increased gene expression in blood provides further support for an anti-inflammatory effect of IL-27 for the treatment of allergic conditions.

Lastly, we identified four significant gene-based associations with non-coding RNAs of unknown function. Of these, the lncRNA *RP11-534L20.5* is of particular interest, as it is located in close proximity (8kb) to *IKBKE*, a regulator of the NF-kappaB pathway that plays a role in immune-related mechanisms<sup>38, 39</sup>. Using data from release v7 of the GTEx project<sup>40</sup>, we found a highly significant positive correlation in gene expression between *RP11-534L20.5* and *IKBKE* in the skin ( $P=10^{-11}$ ), with a weaker but consistent effect in blood and lung (not shown). Such an association could arise, for example, if *RP11-534L20.5* regulates the expression of *IKBKE* or if both transcripts share a regulatory element. Consistent with the latter hypothesis, the 5' end of *RP11-534L20.5* overlaps a peak of H3K27 acetylation (a mark for active enhancers) in multiple cell lines and physically interacts with the *IKBKE* promoter in a B-cell line<sup>41</sup>. Further studies are warranted to investigate the function of *RP11-534L20.5*, as well as the other non-coding RNAs identified in our analysis.

Two additional caveats should be considered when interpreting results from our study. First, our original GWAS<sup>6</sup> included only individuals of European ancestry, and so it is unclear if the risk loci identified in our current study extend to individuals of other ancestries. Second, the association analyses performed to compare individuals suffering from a single allergic disease were based on a relatively small sample size, and so it is possible that the lack of significant associations reflects the lower power of these analyses.

In conclusion, we identified significant and reproducible gene-based associations with 19 genes located in 11 loci not previously reported in GWAS of allergic disease. Our genetic findings suggest that drugs that target these genes might have an increased probability of success if prioritised for clinical development<sup>42</sup>. Our results further demonstrate the utility of applying gene-based tests of association to results from existing GWAS.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 10074. MARF was supported by a Senior Research Fellowship (APP1124501) from the National Health and Medical Research Council (NHMRC) of Australia. JDH was supported by NIH postdoctoral training grant CA112355. LP was funded by a UK MRC fellowship award (MR/J012165/1) and works in a unit funded by the UK MRC (MC\_UU\_12013). Detailed acknowledgments are provided in the Online Repository.

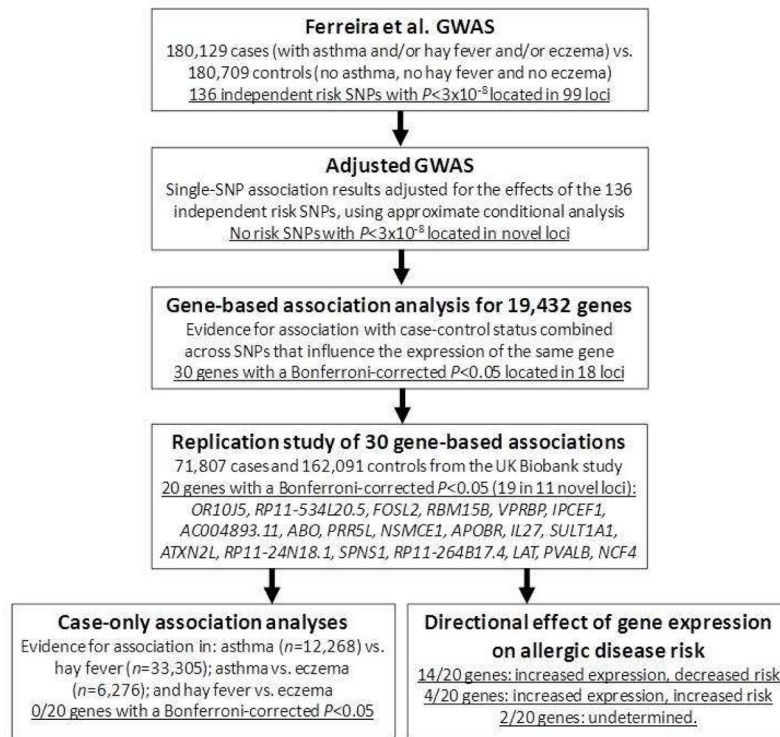
## References

1. Fagnani C, Annesi-Maesano I, Brescianini S, D'Ippolito C, Medda E, Nistico L, et al. Heritability and shared genetic effects of asthma and hay fever: an Italian study of young twins. *Twin Res Hum Genet.* 2008; 11:121–31. [PubMed: 18361712]
2. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis.* 1990; 142:1351–8. [PubMed: 2252253]
3. Thomsen SF, Ulrik CS, Kyvik KO, Skadhauge LR, Steffensen I, Backer V. Findings on the atopic triad from a Danish twin registry. *Int J Tuberc Lung Dis.* 2006; 10:1268–72. [PubMed: 17131787]
4. van Beijsterveldt CE, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J.* 2007; 29:516–21. [PubMed: 17215318]
5. Ullemar V, Magnusson PK, Lundholm C, Zettergren A, Melen E, Lichtenstein P, et al. Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy.* 2016; 71:230–8. [PubMed: 26786172]
6. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet.* 2017; 49:1752–7. [PubMed: 29083406]
7. Fadista J, Manning AK, Florez JC, Groop L. The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet.* 2016; 24:1202–5. [PubMed: 26733288]
8. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet.* 2010; 87:139–45. [PubMed: 20598278]
9. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015; 11:e1004219. [PubMed: 25885710]
10. Bakshi A, Zhu Z, Vinkhuyzen AA, Hill WD, McRae AF, Visscher PM, et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci Rep.* 2016; 6:32894. [PubMed: 27604177]

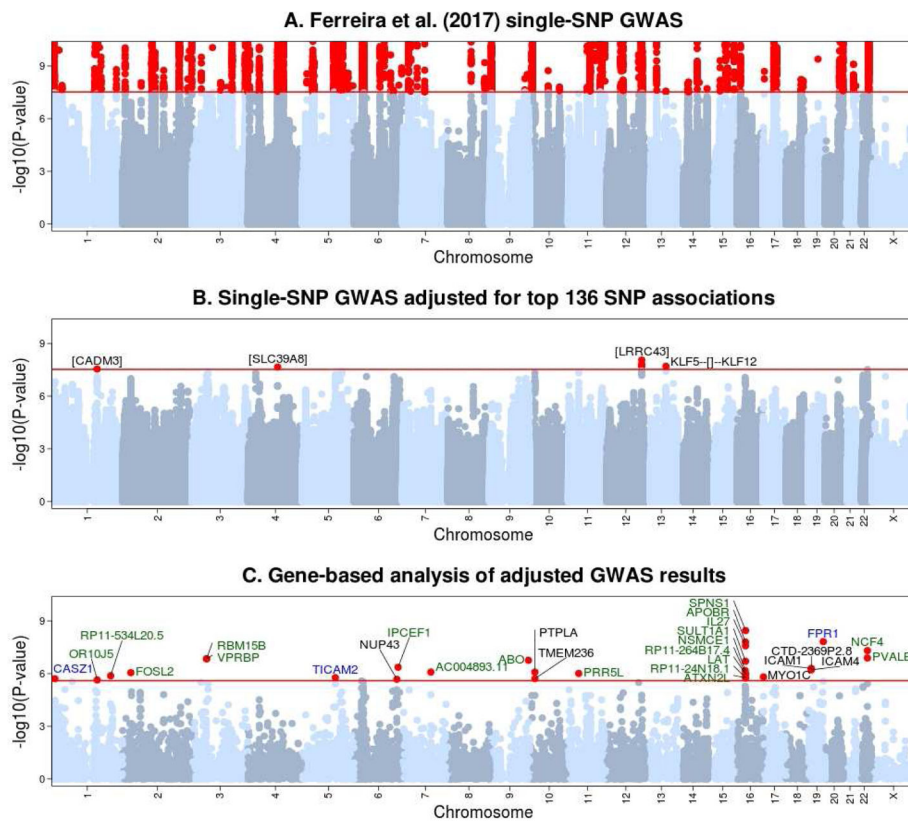
11. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 2010; 6:e1000888. [PubMed: 20369019]
12. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015; 47:1091–8. [PubMed: 26258848]
13. Ferreira MA, Jansen R, Willemsen G, Penninx B, Bain LM, Vicente CT, et al. Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling. *J Allergy Clin Immunol.* 2017; 139:1148–57. [PubMed: 27554816]
14. Barbeira A, Dickinson SP, Torres JM, Bonazzola R, Zheng J, Torstenson ES, et al. Integrating tissue specific mechanisms into GWAS summary results. *bioRxiv.* 2017
15. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015; 348:648–60. [PubMed: 25954001]
16. Odhams CA, Cunnigham Graham DS, Vyse TJ. Profiling RNA-Seq at multiple resolutions markedly increases the number of causal eQTLs in autoimmune disease. *PLoS Genet.* 2017; 13:e1007071. [PubMed: 29059182]
17. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv.* 2017
18. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. Schizophrenia Working Group of the Psychiatric Genomics C. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015; 47:291–5. [PubMed: 25642630]
19. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012; 44:369–75. S1–3. [PubMed: 22426310]
20. Davis JR, Fresard L, Knowles DA, Pala M, Bustamante CD, Battle A, et al. An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am J Hum Genet.* 2016; 98:216–24. [PubMed: 26749306]
21. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature.* 2010; 464:773–7. [PubMed: 20220756]
22. Lappalainen T, Sammeth M, Friedlander MR, Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature.* 2013; 501:506–11. [PubMed: 24037378]
23. Pierce BL, Tong L, Chen LS, Rahaman R, Argos M, Jasmine F, et al. Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. *PLoS Genet.* 2014; 10:e1004818. [PubMed: 25474530]
24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015; 4:7. [PubMed: 25722852]
25. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. Schizophrenia Working Group of the Psychiatric Genomics C. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015
26. Ferreira MA. Improving the power to detect risk variants for allergic disease by defining case-control status based on both asthma and hay fever. *Twin Res Hum Genet.* 2014; 17:505–11. [PubMed: 25296694]
27. Xu Z, Wu C, Wei P, Pan W. A Powerful Framework for Integrating eQTL and GWAS Summary Data. *Genetics.* 2017
28. Tong T, Ryu SE, Min Y, de March CA, Bushdid C, Golebiowski J, et al. olfactory receptor 10J5 responding to alpha-cedrene regulates hepatic steatosis via the cAMP-PKA pathway. *Sci Rep.* 2017; 7:9471. [PubMed: 28842679]
29. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet.* 2011; 43:887–92. [PubMed: 21804549]
30. Graham GJ, Locati M, Mantovani A, Rot A, Thelen M. The biochemistry and biology of the atypical chemokine receptors. *Immunol Lett.* 2012; 145:30–8. [PubMed: 22698181]

31. Ubieta K, Garcia M, Grottsch B, Uebe S, Weber GF, Stein M, et al. Fra-2 regulates B cell development by enhancing IRF4 and Foxo1 transcription. *J Exp Med*. 2017; 214:2059–71. [PubMed: 28566276]
32. Wurm S, Zhang J, Guinea-Viniegra J, Garcia F, Munoz J, Bakiri L, et al. Terminal epidermal differentiation is regulated by the interaction of Fra-2/AP-1 with Ezh2 and ERK1/2. *Genes Dev*. 2015; 29:144–56. [PubMed: 25547114]
33. Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, et al. A validated regulatory network for Th17 cell specification. *Cell*. 2012; 151:289–303. [PubMed: 23021777]
34. Hu R, Kagele DA, Huffaker TB, Runtsch MC, Alexander M, Liu J, et al. miR-155 promotes T follicular helper cell accumulation during chronic, low-grade inflammation. *Immunity*. 2014; 41:605–19. [PubMed: 25367574]
35. Gonzalez JR, Caceres A, Esko T, Cusco I, Puig M, Esnaola M, et al. A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. *Am J Hum Genet*. 2014; 94:361–72. [PubMed: 24560518]
36. Jirno AC, Daluege K, Happle C, Albrecht M, Dittrich AM, Busse M, et al. IL-27 Is Essential for Suppression of Experimental Allergic Asthma by the TLR7/8 Agonist R848 (Resiquimod). *J Immunol*. 2016; 197:4219–27. [PubMed: 27799314]
37. Yoshimoto T, Yoshimoto T, Yasuda K, Mizuguchi J, Nakanishi K. IL-27 suppresses Th2 cell development and Th2 cytokines production from polarized Th2 cells: a novel therapeutic way for Th2-mediated allergic inflammation. *J Immunol*. 2007; 179:4415–23. [PubMed: 17878337]
38. Alves BN, Tsui R, Almaden J, Shokhirev MN, Davis-Turak J, Fujimoto J, et al. IkappaBepsilon is a key regulator of B cell expansion by providing negative feedback on cRel and RelA in a stimulus-specific manner. *J Immunol*. 2014; 192:3121–32. [PubMed: 24591377]
39. Ng SL, Friedman BA, Schmid S, Gertz J, Myers RM, Tenover BR, et al. IkappaB kinase epsilon (IKK(epsilon)) regulates the balance between type I and type II interferon responses. *Proc Natl Acad Sci U S A*. 2011; 108:21170–5. [PubMed: 22171011]
40. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017; 550:204–13. [PubMed: 29022597]
41. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet*. 2015; 47:598–606. [PubMed: 25938943]
42. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015; 47:856–60. [PubMed: 26121088]
43. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016; 48:709–17. [PubMed: 27182965]
44. Wrann CD, Eguchi J, Bozec A, Xu Z, Mikkelsen T, Gimble J, et al. FOSL2 promotes leptin gene expression in human and mouse adipocytes. *J Clin Invest*. 2012; 122:1010–21. [PubMed: 22326952]
45. Guo Z, Kong Q, Liu C, Zhang S, Zou L, Yan F, et al. DCAF1 controls T-cell function via p53-dependent and -independent mechanisms. *Nat Commun*. 2016; 7:10307. [PubMed: 26728942]
46. Kassmeier MD, Mondal K, Palmer VL, Raval P, Kumar S, Perry GA, et al. VprBP binds full-length RAG1 and is required for B-cell development and V(D)J recombination fidelity. *EMBO J*. 2012; 31:945–58. [PubMed: 22157821]
47. Sadewasser A, Paki K, Eichelbaum K, Bogdanow B, Saenger S, Budt M, et al. Quantitative Proteomic Approach Identifies Vpr Binding Protein as Novel Host Factor Supporting Influenza A Virus Infections in Human Cells. *Mol Cell Proteomics*. 2017; 16:728–42. [PubMed: 28289176]
48. Venkateswarlu K. Interaction protein for cytohesin exchange factors 1 (IPCEF1) binds cytohesin 2 and modifies its activity. *J Biol Chem*. 2003; 278:43460–9. [PubMed: 12920129]
49. Zhu W, London NR, Gibson CC, Davis CT, Tong Z, Sorensen LK, et al. Interleukin receptor activates a MYD88-ARNO-ARF6 cascade to disrupt vascular stability. *Nature*. 2012; 492:252–5. [PubMed: 23143332]
50. Oh SJ, Santy LC. Differential effects of cytohesins 2 and 3 on beta1 integrin recycling. *J Biol Chem*. 2010; 285:14610–6. [PubMed: 20223830]

51. Thedieck K, Polak P, Kim ML, Molle KD, Cohen A, Jenö P, et al. PRAS40 and PRR5-like protein are new mTOR interactors that regulate apoptosis. *PLoS One*. 2007; 2:e1217. [PubMed: 18030348]
52. Ellson CD, Davidson K, Ferguson GJ, O'Connor R, Stephens LR, Hawkins PT. Neutrophils from p40phox<sup>-/-</sup> mice exhibit severe defects in NADPH oxidase regulation and oxidant-dependent bacterial killing. *J Exp Med*. 2006; 203:1927–37. [PubMed: 16880254]
53. Crotzer VL, Matute JD, Arias AA, Zhao H, Quilliam LA, Dinauer MC, et al. Cutting edge: NADPH oxidase modulates MHC class II antigen presentation by B cells. *J Immunol*. 2012; 189:3800–4. [PubMed: 22984083]
54. Brown ML, Ramprasad MP, Umeda PK, Tanaka A, Kobayashi Y, Watanabe T, et al. A macrophage receptor for apolipoprotein B48: cloning, expression, and atherosclerosis. *Proc Natl Acad Sci U S A*. 2000; 97:7488–93. [PubMed: 10852956]
55. Do J, Kim D, Kim S, Valentin-Torres A, Dvorina N, Jang E, et al. Treg-specific IL-27 $\alpha$  deletion uncovers a key role for IL-27 in Treg function to control autoimmunity. *Proc Natl Acad Sci U S A*. 2017; 114:10190–5. [PubMed: 28874534]
56. Muallem G, Wagage S, Sun Y, DeLong JH, Valenzuela A, Christian DA, et al. IL-27 Limits Type 2 Immunopathology Following Parainfluenza Virus Infection. *PLoS Pathog*. 2017; 13:e1006173. [PubMed: 28129374]
57. Yang B, Suwanpradit J, Sanchez-Lagunes R, Choi HW, Hoang P, Wang D, et al. IL-27 Facilitates Skin Wound Healing through Induction of Epidermal Proliferation and Host Defense. *J Invest Dermatol*. 2017; 137:1166–75. [PubMed: 28132857]
58. Meunier C, Bordereaux D, Porteu F, Gisselbrecht S, Chretien S, Courtois G. Cloning and characterization of a family of proteins associated with Mpl. *J Biol Chem*. 2002; 277:9139–47. [PubMed: 11784712]
59. Bacchelli C, Moretti FA, Carmo M, Adams S, Stanescu HC, Pearce K, et al. Mutations in linker for activation of T cells (LAT) lead to a novel form of severe combined immunodeficiency. *J Allergy Clin Immunol*. 2017; 139:634–42. e5. [PubMed: 27522155]
60. Aguado E, Richelme S, Nunez-Cruz S, Miazek A, Mura AM, Richelme M, et al. Induction of T helper type 2 immunity by a point mutation in the LAT adaptor. *Science*. 2002; 296:2036–40. [PubMed: 12065839]
61. Zhang W, Sloan-Lancaster J, Kitchen J, Tribble RP, Samelson LE. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell*. 1998; 92:83–92. [PubMed: 9489702]



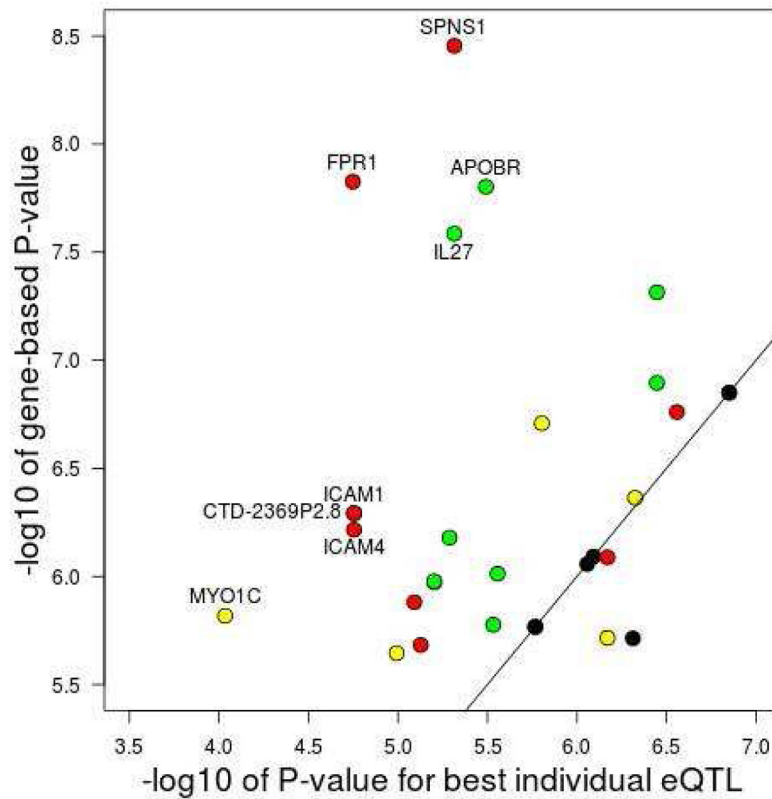
**Figure 1.**  
Outline of analytical approach used.



**Figure 2. Summary of association results between allergic disease status and single SNPs or individual genes**

(A) Single-SNP association results published by Ferreira et al.<sup>6</sup> (B) Single-SNP association results obtained after adjusting the results of Ferreira et al. for the effect of the 136 genome-wide significant associations reported in that study. Variants in four loci (in *CADM3*, *SLC39A8*, *LRRC43* and between *KLF5* and *KLF12*) were genome-wide significant (*i.e.*  $P < 3 \times 10^{-8}$ ) in the adjusted but not in the original GWAS results, and so represent secondary single-SNP association signals in loci identified in Ferreira et al.<sup>6</sup> (C) Gene-based association results obtained for 19,432 genes after applying the EUGENE approach to results from the adjusted GWAS. Genes with a gene-based association  $P < 2.5 \times 10^{-6}$  are listed, with font color reflecting the evidence for association with those genes in the UK Biobank replication study: green -  $P < 0.0016$  (*i.e.* significant after correcting for 30 genes tested); blue -  $0.0016 < P < 0.05$  (*i.e.* nominally significant); and black -  $P > 0.05$  (*i.e.* not significant).





**Figure 3. Comparison of results obtained with single-SNP and gene-based analyses for the 30 genes with a significant gene-based P-value (i.e.  $P < 2.5 \times 10^{-6}$ ) in the discovery study**  
 For each gene, the x-axis shows the statistical evidence for association ( $-\log_{10}P$ -value) obtained with the eQTL most strongly associated with disease risk. The y-axis shows the evidence for association ( $-\log_{10}P$ -value) obtained for each gene with the EUGENE gene-based approach. Gene names are shown for eight genes for which the latter was at least one order of magnitude more significant than the former. The color of each circle indicates the number of independent eQTL for that gene that were individually associated with disease risk at a  $P < 0.05$ : black (one associated eQTL), green (two), yellow (three) and red (three or more).

**Table 1**

Association results for 30 genes with a gene-based association  $P < 2.5 \times 10^{-6}$  in the discovery analysis.

Locus	Gene	Gene location		eQTL with strongest association in GWAS		Gene-based analysis: discovery (n=360,838)			Gene-based analysis: replication (n=233,898)		
		Chr	Start	SNP	GWAS P-value	N eQTLs tested	N eQTLs with P<0.05	EUGENE P-value	N eQTLs tested	N eQTLs with P<0.05	EUGENE P-value
1	<i>CASZ1*</i>	1	10696661	rs12045923	4.9E-07	2	1	1.9E-06	2	1	2.0E-02
2	<i>OR10J5</i>	1	159504793	rs2427837	1.0E-05	3	3	2.3E-06	3	3	6.3E-07
3	<i>RPI1-534L20.5</i>	1	206677281	rs111117858	8.1E-06	12	5	1.3E-06	11	3	1.1E-03
4	<i>FOSL2</i>	2	28615315	rs7562	8.8E-07	1	1	8.8E-07	1	1	2.6E-04
5	<i>RBM15B</i>	3	51428731	rs73078636	1.4E-07	1	1	1.4E-07	1	1	4.0E-04
	<i>VPRBP</i>	3	51433298	rs73078636	1.4E-07	1	1	1.4E-07	1	1	4.0E-04
6	<i>TTCAM2</i>	5	114914339	rs17137937	2.9E-06	3	2	1.7E-06	3	1	5.7E-03
7	<i>NUP43</i>	6	150045451	rs6909158	7.5E-06	22	6	2.1E-06	22	1	4.5E-01
8	<i>IPCEFI</i>	6	154475631	rs9397706	4.7E-07	4	3	4.3E-07	4	3	4.1E-07
9	<i>AC004893.11</i>	7	98610788	rs4236540	8.1E-07	1	1	8.1E-07	1	1	4.3E-04
10	<i>ABO*</i>	9	136125788	rs550057	2.8E-07	14	5	1.7E-07	14	6	3.8E-10
11	<i>PTPLA</i>	10	17631958	rs7092926	6.8E-07	4	4	8.2E-07	4	0	2.1E-01
	<i>TMEM236</i>	10	17794251	rs7092926	6.8E-07	3	3	1.9E-06	3	0	3.6E-01
12	<i>PRR3L</i>	11	36317838	rs7925585	2.8E-06	6	2	9.7E-07	6	2	1.4E-06
13	<i>NSMCE1</i>	16	27236312	rs4523932	5.2E-06	2	2	6.6E-07	2	2	1.3E-06
14	<i>APOBR</i>	16	28505970	rs151233	3.2E-06	2	2	1.6E-08	2	2	3.8E-05
	<i>IL27</i>	16	28510683	rs231970	4.8E-06	2	2	2.6E-08	2	2	1.6E-04
	<i>SULT1A1</i>	16	28616903	rs75539558	1.6E-06	8	3	2.0E-07	8	4	1.4E-05
	<i>ATXN2L</i>	16	28834356	rs8056890	1.7E-06	1	1	1.7E-06	1	1	1.7E-04
	<i>RPI1-24N/18.1</i>	16	28841933	rs8056890	1.7E-06	1	1	1.7E-06	1	1	1.7E-04
	<i>SPNS1</i>	16	28985542	rs2726040	4.8E-06	15	6	3.5E-09	14	5	7.4E-04
	<i>RPI1-264B/17.4</i>	16	28986294	rs8045689	6.3E-06	3	2	1.1E-06	3	2	1.5E-03
	<i>LAT</i>	16	28996147	rs8045689	6.3E-06	3	2	1.1E-06	3	3	2.4E-04
15	<i>MYO1C</i>	17	1367392	rs56157500	9.2E-05	4	3	1.5E-06	4	1	5.6E-02

Locus	Gene	Gene location		eQTL with strongest association in GWAS		Gene-based analysis: discovery (n=360,838)			Gene-based analysis: replication (n=233,898)		
		Chr	Start	SNP	GWAS P-value	N eQTLs tested	N eQTLs with P<0.05	EUGENE P-value	N eQTLs tested	N eQTLs with P<0.05	EUGENE P-value
16	<i>IC-AMI</i>	19	10381511	rs1799969	1.8E-05	12	7	5.1E-07	12	0	4.1E-01
	<i>CTD-2369P2.8</i>	19	10396477	rs1799969	1.8E-05	12	7	5.1E-07	12	0	4.1E-01
		19	10397643	rs1799969	1.8E-05	12	7	6.1E-07	11	0	4.9E-01
17	<i>FPR1</i>	19	52248425	rs7254019	1.8E-05	14	7	1.5E-08	14	3	3.2E-03
18	<i>PVALB</i>	22	37196728	rs4821544	3.6E-07	5	2	1.3E-07	5	4	5.9E-05
	<i>NCF4</i>	22	37257030	rs4821544	3.6E-07	5	2	4.8E-08	5	3	1.1E-04

\* Located in a locus identified in a previous GWAS of allergic disease 43

The replication P-value is in bold if significant after correcting for the 30 genes tested (*i.e.*  $P < 0.05/30 = 0.0016$ ) and in italic if  $0.0016 < P < 0.05$  (*i.e.* nominally significant).

**Table 2**  
Genes with a known function that is relevant to the pathophysiology of allergic disease.

Gene	Relevant function	References
<i>FOSL2</i>	B cell and epidermal differentiation, Th17 and adipocyte function	31–33, 44
<i>VPRBP</i>	T cell function, B cell development, viral replication	45–47
<i>IPCEFI</i>	Binds cytohesin 2, which is involved in IL-1beta signaling and cell adhesion	48–50
<i>PRR5L</i>	mTOR interactor that regulates cell death	51
<i>NCF4</i>	Phagocyte oxidative burst, antigen presentation	52, 53
<i>APOBR</i>	Macrophage receptor for apolipoprotein B48	54
<i>IL27</i>	Regulation of Th1 and Th2 responses, Treg function, epithelial cell proliferation	37, 55–57
<i>ATXN2L</i>	Cytokine signaling	58
<i>LAT</i>	T cell development and activation	59–61