Eleven loci with new reproducible genetic associations with allergic disease risk

Manuel A. R. Ferreira, PhD,a Judith M. Vonk, PhD,b Hansjörg Baurecht, PhD,c Ingo Marenholz, PhD,d,e Chao Tian, PhD,f Joshua D. Hoffman, PhD,g Quinta Helmer, PhD,h Annika Tillander, PhD,i Vilhelmina Ullemar, PhD,i Yi Lu, PhD,j Franz Rüschendorf, PhD,d the 23andMe Research Team, the collaborators of the SHARE study,* David A. Hinds, PhD,3 Norbert Hübner, MD,3 Stephan Weidinger, MD,6 Patrik K. E. Magnusson, PhD,i Eric Jorgenson, PhD,j Young-Ae Lee, MD,de Dorret I. Boomsma, PhD,6 Robert Karlsson, PhD,1 Catarina Almqvist, MD,1k Gerard H. Koppelman, MD,1 and Lavinia Paternoster, PhDm

Brisbane, Australia; Groningen and Amsterdam, The Netherlands; Kiel and Berlin, Germany; Mountain View, San Francisco, and Oakland, Calif; Stockholm, Sweden; and Bristol, United Kingdom

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 between these common allergic diseases remain to be discovered, which could point to new therapeutic opportunities. Objective: We sought 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 subjects. 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 analyzed 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 subjects 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 ($P < .05/30 = .0016$) with disease risk in the replication sample, including 19 that were located in 11 loci not reported to contain allergy risk variants in previous GWASs. Among these were 9 genes with a known function that is directly relevant to allergic disease: FOSL2, VPRBP, IPCEF1, PRR5L, NCF4, APOBR, IL27, ATXN2L, and LAT. For 4 genes (eg, 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 GWASs. Functional studies that investigate the contribution of the 19 associated genes to the pathophysiology of allergic disease and assess their therapeutic potential are warranted. (J Allergy Clin Immunol 2019;143:691-9.)

Key words: Rhinitis, atopic dermatitis, atopy, genome-wide association study, transcriptome

The strong genetic correlations between asthma, hay fever, and eczema estimated from twin studies1–5 in combination with the highly polygenic architecture of these diseases predict that many hundreds, if not thousands, of genetic risk factors are shared among these 3 common allergic diseases. Motivated by this prediction, we recently performed a genome-wide association study

CA112355. L.P. 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 in this article’s Online Repository at www.jacionline.org.

Disclosure of potential conflict of interest: C. Tian and D. A. Hinds both report support from 23andMe during the conduct of the study. C. Almqvist received grant 2017-00641 from Swedish Research Council and Swedish Initiative for Research on Microdata in the Social And Medical Sciences (SIMSAM) framework grant (340-2013-5867) for this work. G. H. Koppelman’s institution received grants from the Lung Foundation of The Netherlands, the Ubbo Emmius Foundation, TEVA (The Netherlands), GlaxoSmithKline, Vertex, and the Tetri Foundation for other works. L. Paternoster received grant MR/J012165/1 from the UK Medical Research Council for this work and personal fees from Merck. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication November 24, 2017; revised February 1, 2018; accepted for publication March 19, 2018.

Available online April 19, 2018.

Corresponding author: Manuel A. R. Ferreira, PhD, QIMR Bergoher Medical Research Institute, Locked Bag 2000, Royal Brisbane Hospital, Herston, QLD 4029, Australia.

E-mail: manuel.ferreira@qimrberghofer.edu.au.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections 0901-6749/36.00 © 2018 American Academy of Allergy, Asthma & Immunology https://doi.org/10.1016/j.jaci.2018.03.012
(GWAS) designed to identify genetic risk variants that are shared between asthma, hay fever, and eczema. In that GWAS cases (n = 180,129) were defined as subjects who reported having 1 or more allergic diseases, whereas control subjects (n = 180,709) were subjects who reported never having any of these diseases. We identified 136 single nucleotide polymorphisms (SNPs) located in 99 loci (ie, genomic regions located >1 Mb apart) that were independently associated with disease risk at a genome-wide significance threshold P value of less than \(3 \times 10^{-8}\); a threshold that corrects for the number of SNPs tested. 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 GWASs of this multidisease phenotype are underway and 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, MAGMA, and fastBAT. 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 (eg, 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 associated independently 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 (eQTLs). On the other hand, eQTLs have a greater probability of being associated with common diseases and traits. Based on these observations, an additional suite of gene-based methods has been developed recently that only include in the association test functional SNPs, such as eQTLs. These include PrediXcan, EUGENE, and S-PrediXcan. We developed EUGENE because it was not possible with other methods to combine in the same association test information from eQTLs identified in different tissues. This feature is important because multiple tissue types play a role in allergic disease pathophysiology and tissue-specific eQTLs are common.

Based on these observations, an additional suite of gene-based methods has been developed recently that only include in the association test functional SNPs, such as eQTLs. These include PrediXcan, EUGENE, and S-PrediXcan. We developed EUGENE because it was not possible with other methods to combine in the same association test information from eQTLs identified in different tissues. This feature is important because multiple tissue types play a role in allergic disease pathophysiology and tissue-specific eQTLs are common. Furthermore, EUGENE also considers regulatory variants with different mechanisms of action, such as 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.

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 and (2) follow up the top gene-based associations in an independent replication sample ascertained from the UK Biobank study.

**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, which included 360,838 subjects from 13 studies: UK Biobank, 23andMe, GERA, CATSS, NTR, LifeLines, TWINGENE, ALSPAC, SALTY, GENEVA, AAHC, GENUFAD-SHIP-1, and GENUFAD-SHIP-2 (demographics are provided in Table E1 in this article’s Online Repository at www.jacionline.org). In that study single-SNP results were corrected for an inflation factor that reflected technical biases and/or population stratification (specifically, a linkage disequilibrium [LD] Score regression intercept of 1.04). After that correction, there were 136 variants associated independently with allergic disease in the study by Ferreira et al.

Because our aim was to identify new allergy risk loci, we first applied approximate conditional analyses, as implemented in the GCTA tool. The summary statistics of Ferreira et al to adjust the single-SNP results for the effects of the 136 risk variants (a detailed justification for performing conditional analysis before the gene-based analysis is provided in the Methods section in this article’s Online Repository at www.jacionline.org). In the conditional analysis, LD between SNPs was estimated based on genotype data from 5000 subjects from the UK Biobank study.

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

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

We identified eQTLs based on information from 39 published eQTL studies conducted in 19 tissues or cell types relevant to allergic disease (see Table E2 in this article’s Online Repository at www.jacionline.org). For each eQTL study, we (1) downloaded the original publication tables containing results for the eQTLs reported; (2) extracted the SNP identifier, gene name, association P value, and directional effect (if available; \(\beta\); score and effect allele) for all reported eQTLs; and (3) excluded eQTLs located more than 1 Mb of the respective gene (ie, trans-eQTLs) and/or with an association P value of greater than 2.3 \times 10^{-7}, a conservative threshold that corrects for 21,742 genes in the genome, each tested for association with 1,000 SNPs (as suggested by others). We did not include trans-eQTLs in the analysis because often these are thought to involve indirect effects, such as where an 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-eQTLs for a given gene from published studies, we then reduced that list to a set of eQTLs in low LD with each other (specifically, with an \(r^2 < 0.1\)) using the clump command in PLINK, version 1.90. For a given gene, we refer to these as “independent eQTLs,” although we recognize that some pairs of eQTLs will not be in linkage equilibrium. LD was estimated based on genetic data from individuals of European descent from the 1000 Genomes Project (n = 294, release 20130502_v5a). Clumping was not performed separately for each tissue or study but rather applied to the overall list of eQTLs obtained after considering information from all tissues/studies. If an eQTL was identified in multiple tissues/studies, the clumping procedure was performed with the smallest P value reported for that eQTL across all tissues/studies. A file (ASTHMA.20170517.eqtl.proxies.list) containing the independent eQTLs 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 an \(r^2\) value \(>0.8\) from
the GWAS summary statistics, sums the association $\chi^2$ values across those eQTLs, and estimates the significance of the resulting sum test statistic by using the Satterthwaite approximation, which accounts for the LD between eQTLs. This approximation was originally implemented by Baksh et al.\textsuperscript{10} in the GCTA-fastBAT module and is now also available in EUGENE. LD between eQTLs was estimated based on data from 294 European subjects from the 1000 Genomes Project (release 20130502_v5a). The significance threshold required to achieve experiment-wide significance was set at a $P$ value of less than .05/N genes tested.

Replication of significant gene-based associations in an independent sample

To confirm novel gene-based associations, we analyzed an independent sample of unrelated subjects of European descent from the UK Biobank study.\textsuperscript{11} The approach used to select subjects for analysis was very similar to that described in detail previously.\textsuperscript{11} 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 subjects, in June 2017; (2) pruned the array data down to a set of 29,446 independent SNPs that had comparable ($P > .005$) allele frequencies between European subjects of the 1000 Genomes Project (CEU, FIN, GBR, and TSI groups) and the UK Biobank subjects of European descent included in the GWAS by Ferreira et al.;\textsuperscript{6} (3) merged the pruned data set with data from 1,092 subjects of known ancestry from the 1000 Genomes Project (release 20130502_v5a); (4) performed multidimensional scaling analysis of identity-by-state allele sharing separately for each of 32 groups of approximately 16,000 subjects (to be computationally feasible), including those from the 1000 Genomes Project; (5) identified and removed subjects who did not cluster closely (within 5 SDs of multidimensional scaling components 1 and 2) to subjects of European ancestry from the 1000 Genomes Project, resulting in 461,885 subjects; and (6) identified and removed any subjects included in or related to (i.e., with a kinship coefficient that indicates third-degree relatedness or closer based on file ukb10007_rel_s488374.dat) the 138,354 subjects of the UK Biobank study included in Ferreira et al.,\textsuperscript{6} as well as subjects (A) with genetically inferred sex different from self-reported sex, (B) who were outliers for SNP missingness or genome-wide heterozygosity levels, and/or (C) with more than 10 third-degree relatives, resulting in 244,395 subjects.

For each subject, allergic disease status was defined as previously described for the UK Biobank study in detail.\textsuperscript{6} To classify asthma status, we combined information from 3 sources: (1) touchscreen questionnaires (data field 6152); (2) noncancer illness code, self-reported during verbal interview (data field 20002); and (3) main (data field 41202) and secondary (data field 41204) International Classification of Diseases, Tenth Revision (ICD-10), diagnoses. Specifically, inclusion criteria for cases were as follows: (1) a report of “asthma” in field 6152 and a code for asthma (1111) in field 20002 or (2) an ICD-10 code for asthma in fields 41202 or 41204. Exclusion criteria for cases were as follows: (1) a report of chronic obstructive pulmonary disease in fields 6152 or 20002 or (2) other self-reported respiratory diseases in field 20002. Inclusion criteria for control subjects were no report of asthma in fields 6152, 20002, 41202, and 41204. Exclusion criteria for control subjects were the same as for cases (no chronic obstructive pulmonary disease or other self-reported respiratory diseases).

To classify hay fever status, we used the same 3 sources of information. Specifically, inclusion criteria for cases were as follows: (1) a report of “hay fever, allergic rhinitis or eczema” in field 6152 and a code for hay fever (1387) in field 20002 or (2) an ICD-10 code for hay fever in fields 41202 or 41204. Inclusion criteria for control subjects 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 as follows: (1) a report of “hay fever, allergic rhinitis or eczema” in field 6152 and a code for eczema (1452) in field 20002 or (2) an ICD-10 code for eczema in fields 41202 or 41204. Inclusion criteria for control subjects were no report of eczema in fields 6152, 20002, 41202, and 41204.

To create the overall allergic disease phenotype used for analysis, cases were subjects classified as having at least 1 condition (asthma and/or hay fever and/or eczema), as described above. Control subjects were subjects classified as not having any of the 3 conditions. By using this approach, the 244,395 subjects selected for analysis (see above) included 71,807 allergic disease cases, 162,091 allergic disease control subjects, and 10,497 subjects with a missing phenotype.

SNPTEST was then used to test the association between disease status and imputed genotype data for eQTLs of genes selected for replication; age, sex, and SNP chip were included as covariates. We analyzed only SNPs imputed based on the Haplopye Reference Consortium panel, given that variants imputed from the UK10K plus 1000 Genomes panel were not mapped correctly. Association results from 1.2 million HapMap3 SNPs were used to estimate the degree of inflation in test statistics arising because of unaccounted technical biases by using the LD Score approach.\textsuperscript{18} We observed an LD Score intercept of 1.09, which was used to adjust the association results of 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 patients with a single allergic disease

The case-control phenotype analyzed in our GWAS\textsuperscript{5} combined information from asthma, hay fever, and eczema and therefore was expected to improve power to identify risk variants shared between but not specific to any of the 3 diseases.\textsuperscript{21} To understand whether gene-based associations discovered through 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.\textsuperscript{5}

First, we tested the association between eQTLs of selected genes with 3 phenotypes that contrasted 3 nonoverlapping groups of adults with a single allergic disease: asthma-only cases (group 1), asthma-only versus hay fever-only cases (group 2; n = 12,268) versus hay fever-only cases (group 2; n = 33,305); asthma-only cases (group 1) versus eczema-only cases (group 3; n = 6,276); and hay fever-only cases (group 2) versus eczema-only cases (group 3). For a given eQTL, results from these analyses indicate whether the risk allele is more (odds ratio [OR] > 1) or less (OR < 1) common in, for example, group 1 when compared with group 2. For example, if an eQTL contributed similarly to the risks of asthma and hay fever but not eczema, then one would expect an OR of approximately 1 in the asthma-only versus hay fever–only comparison but an OR of greater than 1 in the asthma versus eczema and hay fever versus eczema analyses.

Second, for each gene and phenotype tested, we combined association results across eQTLs by 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 analytic approach used is shown in Fig 1. To identify novel risk loci shared between asthma, hay fever, and eczema, we first adjusted the association results from Ferreira et al.\textsuperscript{6} for the 136 genome-wide significant SNPs (ie, with $P < 3 \times 10^{-5}$) identified in that study by using approximate conditional analysis (Fig 2, A). In the resulting adjusted GWAS, as expected, there were no SNPs associated with disease risk at a $P$ value of less than $3 \times 10^{-5}$ and located greater than 1 Mb from the loci reported in Ferreira et al.\textsuperscript{6} On the other hand, 4 SNPs located in loci reported in that study (in or near CADM3, SLC39A8, LRRC43, and KLF5) were genome-wide significant in the conditional but not original analyses, which is consistent with the presence of additional secondary association signals at
those established risk loci (Fig 2, B). Importantly, there was an enrichment in significant SNP associations in the adjusted GWAS (see Fig E1 in this article’s Online Repository at www.jacionline.org), 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 approach13 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 noncoding RNAs [lncRNA]) that were reported to have 1 or more independent eQTLs in 19 tissues or cell types relevant to allergic disease (see Table E2), 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 I).26 The specific eQTLs included in the gene-based test for each of these 30 genes are listed in Table E3 in this article’s Online Repository at www.jacionline.org. For 21 genes, the association P value obtained with the gene-based test was more significant than the P value obtained with the individual eQTLs most associated with disease risk (range, 2-7 associated eQTLs per gene; Table I). For 8 genes, the difference in significance between the most associated individual eQTLs and the gene-based test exceeded 1 order of magnitude (Fig 3). The most extreme example of this was the SPNS1 gene on chromosome 16p11.2 (gene-based P = $3.5 \times 10^{-9}$ versus the best individual eQTL P = $4.8 \times 10^{-6}$), for which 6 of the 15 eQTLs tested (identified in 5 tissues) were nominally associated with disease risk (see Table E4 in this article’s Online Repository at www.jacionline.org).

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 control subjects genotyped by the UK Biobank study17 who were unrelated to subjects from our initial GWAS.6 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 < .05/30 = .0016; Table I). These included 19 genes located in 11 loci not implicated in allergic disease in previous GWASs: OR10J5 (chromosome 1q23.2), RP11-534L20.5 (1q32.1),

FIG 1. Outline of the analytic approach used.
FOSL2 (2p23.2), RBM15B and VPRBP (3p21.2), IPCEF1 (6q25.2), AC004893.11 (7q22.1), PRRSL (11p13), NSMCE1 (16p12.1), SPNS1 and 7 other nearby genes (16p11.2), and PVALB and NCF4 (22q12.3). Of note, 9 of these genes have a known function that is directly relevant to the pathophysiology of allergic disease (Table II).27-48

Predicted directional effect of gene expression on disease risk

Because the gene-based approach used focuses exclusively on eQTLs, which in turn are associated with expression of specific genes, we were able to identify in a single analysis both novel risk loci and the likely gene or genes 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 whether the allergy-protective allele of eQTLs 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 ($P < .0016$) association in the replication study, we found that for 14 genes, the allergy-protective allele was associated with increased gene expression (see Table E5 in this article’s Online Repository at www.jacionline.org). This includes, for example, the IncRNA RP11-534L20.5, for which information from 6 different tissues (including blood, lung, and skin) indicates that increased gene expression has a protective effect on disease risk. On the other hand, for 4 genes (AC004893.11, ATXN2L, NSMCE1, and RP11-24N18.1), the allergy-protective allele was associated with decreased gene expression, whereas for 2 genes, eQTL directional effects were conflicting between studies (IPCEF1 and SULT1A1).
Association analyses contrasting patients with a single allergic disease

The multiple-disease case-control phenotype analyzed in our GWAS maximizes power to identify risk variants that are shared between asthma, hay fever, and eczema. As such, we expected the 11 novel risk loci identified above to have comparable effects on the 3 individual diseases. To address this possibility, we tested the association between the 20 genes in those 11 loci and 3 phenotypes that compared 3 nonoverlapping groups of adults with 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 < .05/20 genes × 3 phenotypes) = .0008), no single gene had a significant association in the asthma versus hay fever, asthma versus eczema, or hay fever versus eczema analyses (see Table E6 in this article’s Online Repository at www.jacionline.org). Even at a nominal P value of less than .05, which does not correct for multiple testing, only 3 genes (in 2 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 3 individual diseases.

DISCUSSION

By analyzing results from our previous GWAS 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 2 main reasons. First, by testing individual genes rather than SNPs, the multiple testing burden was reduced greatly; this translated into a less stringent threshold required to declare genome-wide significance (P = 2.5 × 10⁻⁶ instead of 3 × 10⁻⁸), which increases power. Second, most of these loci (7/11) contain

| Gene location | eQTL with strongest association in GWAS | Gene-based analysis: discovery (n = 360,838) | Gene-based analysis: replication (n = 233,898) |
|---------------|-------------------------------- -------|------------------------------------------|-----------------------------------------------|
| Locus | Gene | Chromosome | Start | SNP | GWAS P value | No. of eQTLs tested | No. of eQTLs with P < .05 | EUGENE P value | No. of eQTLs tested | No. of eQTLs with P < .05 | EUGENE P value |
| 1 | CASZ1* | 1 | 1 | rs12045923 | 4.9E-07 | 2 | 1 | 1.9E-06 | 2 | 1 | 2.0E-02 |
| 2 | OR10J5 | 1 | 1 | rs2427837 | 1.0E-05 | 3 | 3 | 2.3E-06 | 3 | 3 | 6.3E-07 |
| 3 | RP11-534L20.5 | 1 | 1 | rs11117858 | 8.1E-06 | 12 | 5 | 1.3E-06 | 11 | 3 | 1.1E-03 |
| 4 | FOSL2 | 2 | 28615315 | rs7562 | 8.8E-07 | 1 | 1 | 8.8E-07 | 1 | 1 | 2.6E-04 |
| 5 | RBM15B | 3 | 5142873 | rs73078636 | 1.4E-07 | 1 | 1 | 1.4E-07 | 1 | 1 | 4.0E-04 |
| 6 | VPRBP | 3 | 51433329 | rs73078636 | 1.4E-07 | 1 | 1 | 1.4E-07 | 1 | 1 | 4.0E-04 |
| 7 | TICAM2 | 5 | 114914339 | rs17137937 | 2.9E-06 | 3 | 2 | 1.7E-06 | 3 | 1 | 5.7E-03 |
| 8 | NUP43 | 6 | 15004541 | rs9009158 | 7.5E-06 | 22 | 6 | 2.1E-06 | 22 | 1 | 4.5E-01 |
| 9 | IPCEF1 | 6 | 15475631 | rs9397706 | 2.8E-07 | 14 | 5 | 1.7E-07 | 14 | 6 | 3.8E-10 |
| 10 | ABO* | 9 | 136125788 | rs500057 | 2.8E-07 | 14 | 5 | 1.7E-07 | 14 | 6 | 3.8E-10 |
| 11 | FTPLA | 10 | 17631958 | rs7092926 | 6.8E-07 | 4 | 4 | 8.2E-07 | 4 | 0 | 2.1E-01 |
| 12 | TMEM236 | 11 | 36317838 | rs7925585 | 2.8E-06 | 6 | 2 | 9.7E-07 | 6 | 2 | 1.4E-06 |
| 13 | NSMCE1 | 16 | 27236312 | rs4529392 | 5.2E-06 | 2 | 2 | 6.6E-07 | 2 | 2 | 1.3E-06 |
| 14 | APOBR | 16 | 28509579 | rs151233 | 3.2E-06 | 2 | 2 | 1.6E-08 | 2 | 2 | 3.8E-05 |
| 15 | IL27 | 16 | 28510683 | rs231970 | 4.8E-06 | 2 | 2 | 2.6E-08 | 2 | 2 | 1.6E-04 |
| 16 | SULT1A1 | 16 | 28616903 | rs7539558 | 1.6E-06 | 8 | 3 | 2.0E-07 | 8 | 4 | 1.4E-05 |
| 17 | ATXNL2 | 16 | 28834356 | rs8056890 | 1.7E-06 | 1 | 1 | 1.7E-06 | 1 | 1 | 1.7E-04 |
| 18 | RP11-264B17.4 | 16 | 28986294 | rs8045689 | 6.3E-06 | 3 | 2 | 1.1E-06 | 3 | 2 | 1.5E-03 |
| 19 | LAT | 16 | 28996167 | rs8045689 | 6.3E-06 | 3 | 2 | 1.1E-06 | 3 | 2 | 1.5E-03 |
| 20 | MYO1C | 19 | 1367392 | rs8150750 | 9.2E-05 | 4 | 3 | 1.5E-06 | 4 | 1 | 5.6E-02 |
| 21 | ICAM1 | 19 | 10381511 | rs1799969 | 1.8E-05 | 12 | 7 | 5.1E-07 | 12 | 0 | 4.1E-01 |
| 22 | CTD-2369P2.8 | 19 | 10396476 | rs1799969 | 1.8E-05 | 12 | 7 | 5.1E-07 | 12 | 0 | 4.1E-01 |
| 23 | ICAM4 | 19 | 10397643 | rs1799969 | 1.8E-05 | 12 | 7 | 6.1E-07 | 11 | 0 | 4.9E-01 |
| 24 | FPR1 | 19 | 52248425 | rs7254019 | 1.8E-05 | 14 | 7 | 1.5E-08 | 14 | 3 | 3.2E-03 |
| 25 | PVCL1 | 22 | 37196728 | rs4821544 | 3.6E-07 | 5 | 2 | 1.3E-07 | 5 | 4 | 5.9E-08 |
| 26 | NCF4 | 22 | 37257030 | rs4821544 | 3.6E-07 | 5 | 2 | 4.0E-08 | 5 | 3 | 1.1E-04 |

The replication P value is shown in boldface if significant after correcting for the 30 genes tested (ie, P < .05/30 = .0016) and in italics if .0016 < P < .05 (ie, nominally significant).

*Located in a locus identified in a previous GWAS of allergic disease.25

The replication P value is shown in boldface if significant after correcting for the 30 genes tested (ie, P < .05/30 = .0016) and in italics if .0016 < P < .05 (ie, nominally significant).

*Located in a locus identified in a previous GWAS of allergic disease.25

**The replication P value is shown in boldface if significant after correcting for the 30 genes tested (ie, P < .05/30 = .0016) and in italics if .0016 < P < .05 (ie, nominally significant).**
genes for which multiple independent eQTLs were individually associated with disease risk. Under this scenario, the gene-based test used improves power over the alternative approach of testing individual eQTLs 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.\(^{10,49}\)

Our results point to 19 genes as being the likely candidates underlying the association between the 11 new loci and allergic disease risk. However, we stress that functional studies are now required to confirm that the expression of these genes (1) is determined by (not simply associated with) the eQTLs included in the respective gene-based test and (2) influences disease pathophysiology. In other words, for a 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 eQTLs 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\(^ {46}\) and therefore could plausibly contribute to the pathophysiology of allergic disease. However, there were 4 additional genes within 1 Mb of  OR10J5  that had a statistically (although not genome-wide) significant gene-based association in our discovery analysis:  FCER1A  \((P = 1.2 \times 10^{-5})\),  DARC  \((P = 0.005)\),  IFI16  \((P = 0.009)\), and  PYHIN1  \((P = 0.03)\), with the latter reported previously to contain asthma risk variants in populations of African ancestry.\(^ {51}\) The association with the first 2 replicated in the independent UK Biobank sample \((P = 4.8 \times 10^{-7}\) and  \(P = 0.001\), respectively\)\). Therefore it is possible that  FCER1A, DARC, or both represent causal genes underly- ing 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.\(^ {52}\)

Why did our analysis identify  OR10J5  and not, for example,  FCER1A? The gene-based test for  OR10J5  included 3 eQTLs (see Table E2), all of which were 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  FCER1A  included these same 3 eQTLs (ie, these eQTLs were shared between  OR10J5  and  FCER1A) plus an additional 6 independent eQTLs, none of which were individually associated with disease risk. Because the additional eQTLs tested for  FCER1A  were not associated with disease risk, the resulting gene-based association was weaker when compared with  OR10J5. An interesting question is why some but not all eQTLs of  FCER1A  (and other genes) are associated with disease risk; for example, it could be that disease associations are restricted to eQTLs that influence gene expression in a specific subset of immune cells or to eQTLs 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 or genes, we note that 9 of the 19 genes located in novel risk loci encode proteins with a known function that is directly relevant to allergic disease (Table II). For example,  FOSL2  is involved in B-cell and epidermal differentiation.\(^ {27,26}\) Furthermore, it has a critical yet complex role in Th17 differentiation and function\(^ {29}\): on the one hand, it represses Th17 signature genes (eg,  IL17A), but on the other hand, it promotes expression of genes that drive Th17 maintenance and survival (eg,  IL6R).\(^ {29}\) When we compared the directional effect of  FOSL2  eQTLs between disease risk and gene expression, we found that the allele associated with reduced disease risk was associated with

\[\text{Table II. Genes with a known function that is relevant to the pathophysiology of allergic disease}\

<table>
<thead>
<tr>
<th>Gene</th>
<th>Relevant function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOSL2</td>
<td>B cell and epidermal differentiation, Th17 and adipocyte function</td>
<td>27-30</td>
</tr>
<tr>
<td>VPBRB</td>
<td>T-cell function, B-cell development, viral replication</td>
<td>31-33</td>
</tr>
<tr>
<td>PCFEI</td>
<td>Binds cytokines 2, which is involved in IL-1β signaling and cell adhesion</td>
<td>34-36</td>
</tr>
<tr>
<td>PRPSL</td>
<td>mTOR interactor that regulates cell death</td>
<td>37</td>
</tr>
<tr>
<td>NCF4</td>
<td>Phagocyte oxidative burst, antigen presentation</td>
<td>38, 39</td>
</tr>
<tr>
<td>APOBR</td>
<td>Macrophage receptor for apolipoprotein B48</td>
<td>40</td>
</tr>
<tr>
<td>IL27</td>
<td>Regulation of Th1 and Th2 responses, Treg cell function, epithelial cell proliferation</td>
<td>41-44</td>
</tr>
<tr>
<td>ATXN2L</td>
<td>Cytokine signaling</td>
<td>45</td>
</tr>
<tr>
<td>LAT</td>
<td>T-cell development and activation</td>
<td>46-48</td>
</tr>
</tbody>
</table>

*mTOR, Mammalian target of rapamycin.*
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 proinflammatory T follicular helper cell lineage after immunization with ovalbumin or viral infection or in the context of low-grade chronic inflammation. 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 8 genes identified on chromosome 16p11.2, a region that overlaps a large (approximately 0.45 Mb) and common (49% frequency in Europeans) genomic inversion previously reported to be associated with the joint occurrence of asthma and obesity. Of the 2 eQTLs included in the gene-based test for IL27, one (rs7191548) is in high LD ($r^2 = 0.72$) with an SNP that tags the inversion (rs4788101), suggesting that the association observed with IL27 in our study is explained partly by that large structural variant. IL-27 has been shown to suppress Th2 responses and therefore 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 the allele associated with reduced disease risk was associated with increased gene expression in blood for both IL27 eQTLs tested provides further support for an anti-inflammatory effect of IL-27 in the treatment of allergic conditions.

Lastly, we identified 4 significant gene-based associations with noncoding RNAs of unknown function. Of these, the lncRNA RP11-534L20.5 is of particular interest because it is located in close proximity (8kb) to IKBKE, a regulator of the nuclear factor κB pathway that plays a role in immune-related mechanisms. Using data from release v7 of the GTEx project, 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 tissue (data not shown). Such an association could arise, for example, if RP11-534L20.5 regulates IKBKE expression 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. Further studies are warranted to investigate the function of RP11-534L20.5, as well as the other noncoding RNAs identified in our analysis.

Two additional caveats should be considered when interpreting results from our study. First, our original GWAS included only subjects of European ancestry, and therefore it is unclear whether the risk loci identified in our current study extend to subjects of other ancestries. Second, the association analyses performed to compare subjects with a single allergic disease were based on a relatively small sample size, and therefore 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 GWASs of allergic disease. Our genetic findings suggest that drugs that target these genes might have increased probability of success if prioritized for clinical development. Our results further demonstrate the utility of applying gene-based tests of association to results from existing GWASs.

Key messages
- The risk of developing asthma, hay fever, and/or eczema is associated with genetic variants that influence the expression of 19 genes located in 11 novel risk loci for allergic disease, including FOSL2, VPRBP, IPCEF1, PRR5L, NCF4, AP0BR, IL27, ATXN2L1, and LAT.
- Gene-based association methods that focus on regulatory variants can help identify genetic risk factors for human traits and diseases that were missed by single-variant analyses reported in published GWASs.

REFERENCES


25. 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.


