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,[i] Franz Rüschendorf, PhD,[d] the 23andMe Research Team,[f] the collaborators of the SHARE study,* David A. Hinds, PhD,[f] Norbert Hübner, MD,[d] Stephan Weidinger, MD,[c] Patrik K. E. Magnusson, PhD,[i] Eric Jorgenson, PhD,[j] Young-Ae Lee, MD,[d,e] Dorret I. Boomsma, PhD,[h] Robert Karlsson, PhD,[i] Catarina Almqvist, MD,[i,k] Gerard H. Koppelman, MD,[l] and Lavinia Paternoster, PhD[m] *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 2.5×10^{-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 T_H2 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 studies[1-5] 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

From [a]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane; [b]Epidemiology, University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD; [c]the Department of Dermatology, Allergology and Venereology, University Hospital Schleswig-Holstein, Campus Kiel; [d]Max Delbrück Center (MDC) for Molecular Medicine, Berlin; [e]the Clinic for Pediatric Allergy, Experimental and Clinical Research Center of Charité Universitätsmedizin Berlin and Max Delbrück Center, Berlin; [f]23andMe, Inc, Mountain View; [g]the Department of Epidemiology and Biostatistics, University of California San Francisco; [h]the Department Biological Psychology, Netherlands Twin Register, Vrije University, Amsterdam; [i]the Department of Medical Epidemiology and Biostatistics and the Swedish Twin Registry, Karolinska Institutet, Stockholm; [j]the Division of Research, Kaiser Permanente Northern California, Oakland; [k]the Pediatric Allergy and Pulmonology Unit at Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm; [l]the 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; and [m]MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol.

*Collaborators of the SHARE study are listed in this article's Online Repository at www.jacionline.org.

This research has been conducted using the UK Biobank Resource under Application no. 10074. M.A.R.F. was supported by a Senior Research Fellowship (APP1124501) from the National Health and Medical Research Council (NHMRC) of Australia. J.D.H. was supported by National Institutes of Health (NIH) postdoctoral training grant 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 Berghofer 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

0091-6749/$36.00

© 2018 American Academy of Allergy, Asthma & Immunology

https://doi.org/10.1016/j.jaci.2018.03.012

Abbreviations used

eQTL:	Expression quantitative trait locus
GWAS:	Genome-wide association study
ICD-10:	International Classification of Diseases, Tenth Revision
LD:	Linkage disequilibrium
lncRNA:	Long noncoding RNA
OR:	Odds ratio
SNP:	Single nucleotide polymorphism

(GWAS) designed to identify genetic risk variants that are shared between asthma, hay fever, and eczema.[6] 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 3×10^{-8}, a threshold that corrects for the number of SNPs tested.[7] 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,[8] MAGMA,[9] and fastBAT.[10] 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.[11]

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,[12] EUGENE,[13] and S-PrediXcan.[14] 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.[15] 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.[16]

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[6] and (2) follow up the top gene-based associations in an independent replication sample ascertained from the UK Biobank study.[17]

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,[6] which included 360,838 subjects from 13 studies: UK Biobank, 23andMe, GERA, CATSS, NTR, LifeLines, TWINGENE, ALSPAC, SALTY, GENEVA, AAGC, 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[18] of 1.04). After that correction, there were 136 variants associated independently with allergic disease in the study by Ferreira et al.[6]

Because our aim was to identify new allergy risk loci, we first applied approximate conditional analyses, as implemented in the GCTA tool,[19] to the summary statistics of Ferreira et al[6] 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.[17]

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,[13] a gene-based approach that is applicable to summary statistics (ie, it 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.[15]

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; β/*z* 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 2.3×10^{-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).[20-22] We did not include trans-eQTLs in the analysis because often these are thought to involve indirect effects,[23] 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 procedure in PLINK, version 1.90.[24] 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 χ^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 Bakshi et al[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.[17] The approach used to select subjects for analysis was very similar to that described in detail previously.[6] 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[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 (ie, with a kinship coefficient that indicates third-degree relatedness or closer based on file ukb1007_rel_s488374.dat) the 138,354 subjects of the UK Biobank study included in Ferreira et al,[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.[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 Haplotype 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.[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[6] 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.[25] 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.[6]

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; 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[6] for the 136 genome-wide significant SNPs (ie, with $P < 3 \times 10^{-8}$) 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 3×10^{-8} and located greater than 1 Mb from the loci reported in Ferreira et al.[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

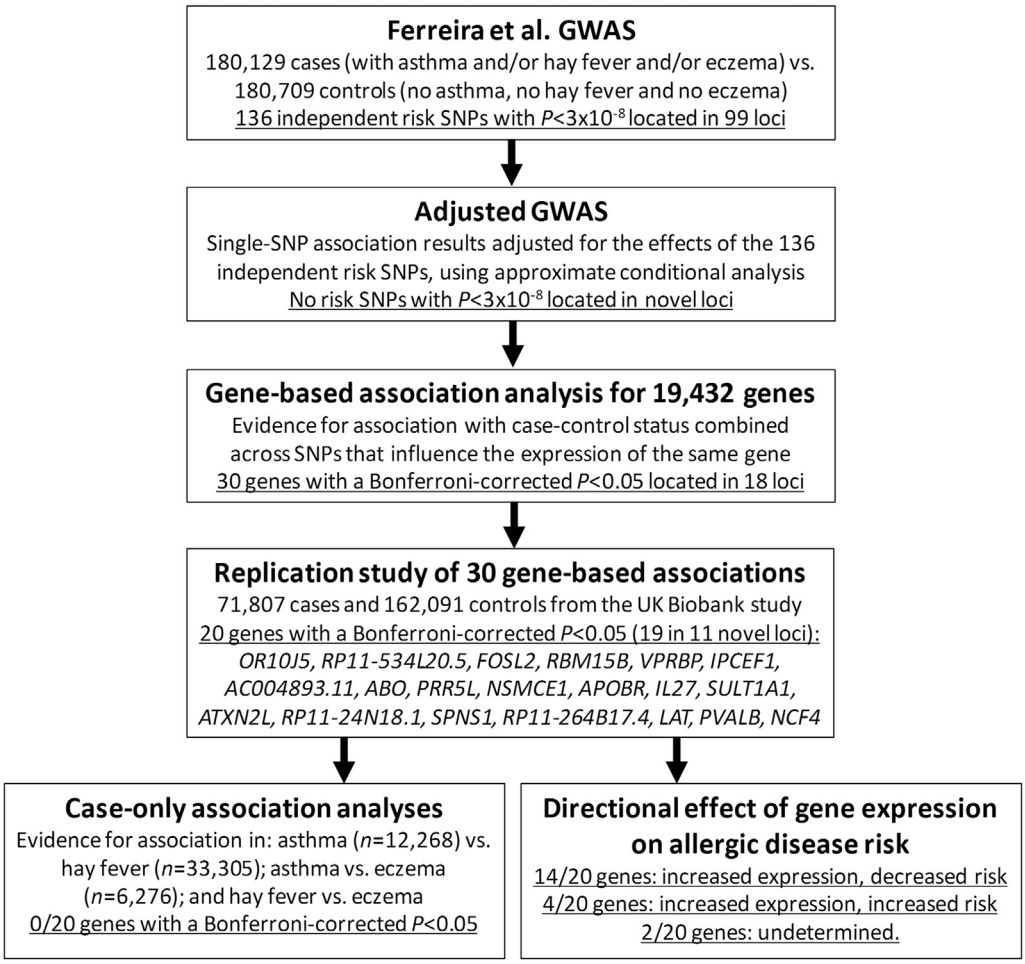


FIG 1. Outline of the analytic approach used.

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 approach[13] 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 2.5×10^{-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 (Fig 3), indicating that multiple eQTLs for the same gene were 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 study[17] 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),

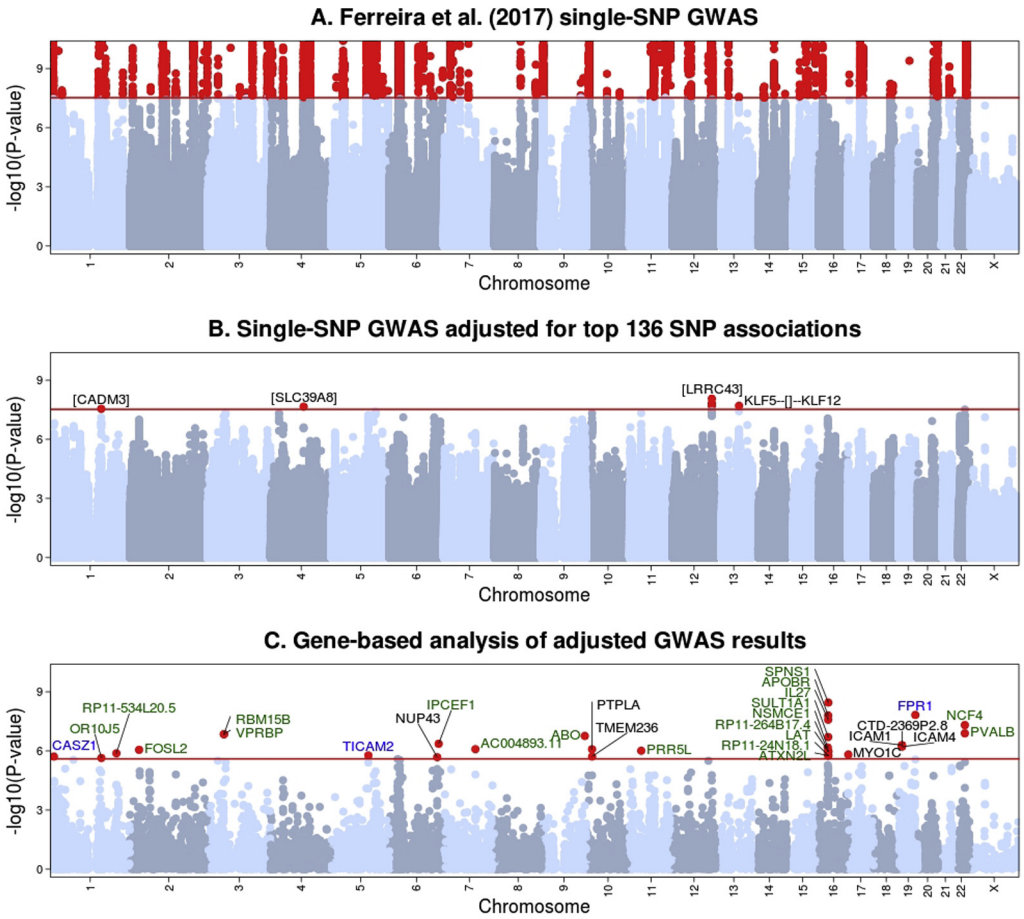


FIG 2. Summary of association results between allergic disease status and single SNPs or individual genes. **A,** Single-SNP association results published by Ferreira et al.[6] **B,** Single-SNP association results obtained after adjusting the results of Ferreira et al[6] for the effect of the 136 genome-wide significant associations reported in that study. Variants in 4 loci (in *CADM3*, *SLC39A8*, and *LRRC43* and between *KLF5* and *KLF12*) were genome-wide significant (ie, $P < 3 \times 10^{-8}$) in the adjusted but not original GWAS results and therefore represent secondary single-SNP association signals in loci identified in Ferreira et al.[6] **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* value of less than 2.5×10^{-6} are listed, with font color reflecting the evidence for association with those genes in the UK Biobank replication study: green, $P < .0016$ (ie, significant after correcting for 30 genes tested); blue, $.0016 < P < .05$ (ie, nominally significant); and black, $P > .05$ (ie, not significant).

FOSL2 (2p23.2), *RBM15B* and *VPRBP* (3p21.2), *IPCEF1* (6q25.2), *AC004893.11* (7q22.1), *PRR5L* (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 lncRNA *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*).

TABLE I. Association results for 30 genes with a gene-based association *P* value of less than 2.5×10^{-6} in the discovery analysis

		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	10696661	rs12045923	4.9E-07	2	1	1.9E-06	2	1	*2.0E-02*
2	*OR10J5*	1	159504793	rs2427837	1.0E-05	3	3	2.3E-06	3	3	**6.3E-07**
3	*RP11-534L20.5*	1	206677281	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	51428731	rs73078636	1.4E-07	1	1	1.4E-07	1	1	**4.0E-04**
	VPRBP	3	51433298	rs73078636	1.4E-07	1	1	1.4E-07	1	1	**4.0E-04**
6	*TICAM2*	5	114914339	rs17137937	2.9E-06	3	2	1.7E-06	3	1	*5.7E-03*
7	*NUP43*	6	150045451	rs6909158	7.5E-06	22	6	2.1E-06	22	1	4.5E-01
8	*IPCEF1*	6	154475631	rs9397706	4.7E-07	4	3	4.3E-07	4	3	**4.1E-07**
9	*AC004893.11*	7	98610788	rs4236540	8.1E-07	1	1	8.1E-07	1	1	**4.3E-04**
10	*ABO**	9	136125788	rs550057	2.8E-07	14	5	1.7E-07	14	6	**3.8E-10**
11	*PTPLA*	10	17631958	rs7092926	6.8E-07	4	4	8.2E-07	4	0	2.1E-01
	TMEM236	10	17794251	rs7092926	6.8E-07	3	3	1.9E-06	3	0	3.6E-01
12	*PRR5L*	11	36317838	rs7925585	2.8E-06	6	2	9.7E-07	6	2	**1.4E-06**
13	*NSMCE1*	16	27236312	rs4523932	5.2E-06	2	2	6.6E-07	2	2	**1.3E-06**
14	*APOBR*	16	28505970	rs151233	3.2E-06	2	2	1.6E-08	2	2	**3.8E-05**
	IL27	16	28510683	rs231970	4.8E-06	2	2	2.6E-08	2	2	**1.6E-04**
	SULT1A1	16	28616903	rs75539558	1.6E-06	8	3	2.0E-07	8	4	**1.4E-05**
	ATXN2L	16	28834356	rs8056890	1.7E-06	1	1	1.7E-06	1	1	**1.7E-04**
	RP11-24N18.1	16	28841933	rs8056890	1.7E-06	1	1	1.7E-06	1	1	**1.7E-04**
	SPNS1	16	28985542	rs2726040	4.8E-06	15	6	3.5E-09	14	5	**7.4E-04**
	RP11-264B17.4	16	28986294	rs8045689	6.3E-06	3	2	1.1E-06	3	2	**1.5E-03**
	LAT	16	28996147	rs8045689	6.3E-06	3	2	1.1E-06	3	3	**2.4E-04**
15	*MYO1C*	17	1367392	rs56157500	9.2E-05	4	3	1.5E-06	4	1	5.6E-02
16	*ICAM1*	19	10381511	rs1799969	1.8E-05	12	7	5.1E-07	12	0	4.1E-01
	CTD-2369P2.8	19	10396477	rs1799969	1.8E-05	12	7	5.1E-07	12	0	4.1E-01
	ICAM4	19	10397643	rs1799969	1.8E-05	12	7	6.1E-07	11	0	4.9E-01
17	*FPR1*	19	52248425	rs7254019	1.8E-05	14	7	1.5E-08	14	3	*3.2E-03*
18	*PVALB*	22	37196728	rs4821544	3.6E-07	5	2	1.3E-07	5	4	**5.9E-05**
	NCF4	22	37257030	rs4821544	3.6E-07	5	2	4.8E-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.[26]

Association analyses contrasting patients with a single allergic disease

The multiple-disease case-control phenotype analyzed in our GWAS[6] maximizes power to identify risk variants that are shared between asthma, hay fever, and eczema.[25] 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 ($P < .05/[20 \text{ genes} \times 3 \text{ phenotypes}] = .0008$), 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[6] 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 ($P = 2.5 \times 10^{-6}$ instead of 3×10^{-8}), which increases power. Second, most of these loci (7/11) contain

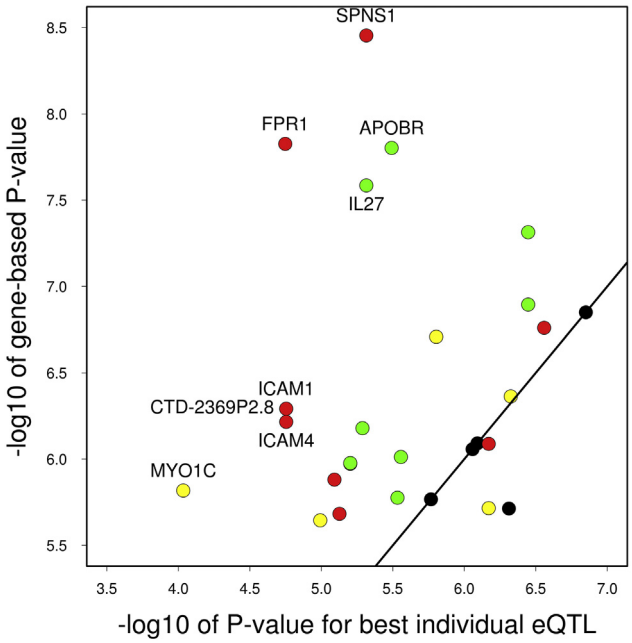


FIG 3. Comparison of results obtained with single-SNP and gene-based analyses for the 30 genes with a significant gene-based *P* value (ie, $P < 2.5 \times 10^{-6}$) in the discovery study. For each gene, the *x-axis* shows the statistical evidence for association ($-\log_{10}$ *P* value) obtained with 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 8 genes, for which the latter was at least 1 order of magnitude more significant than the former. The color of each circle indicates the number of independent eQTLs for that gene that were individually associated with disease risk at a *P* value of less than .05: black (1 associated eQTL), green (2 associated eQTLs), yellow (3 associated eQTLs), and red (≥3 associated eQTLs).

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[50] 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* ($P = .0005$), *IFI16* ($P = .009$), and *PYHIN1* ($P = .03$), 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 $P = .0001$, respectively). Therefore it is possible that *FCER1A*, *DARC*, or both represent causal genes underlying the observed association at this locus. Both are biologically plausible candidate allergy risk genes, encoding, respectively the α subunit of the high-affinity IgE receptor and an atypical chemokine receptor.[52]

TABLE II. Genes with a known function that is relevant to the pathophysiology of allergic disease

Gene	Relevant function	References
FOSL2	B cell and epidermal differentiation, T_H17 and adipocyte function	27-30
VPRBP	T-cell function, B-cell development, viral replication	31-33
IPCEF1	Binds cytohesin 2, which is involved in IL-1β signaling and cell adhesion	34-36
PRR5L	mTOR interactor that regulates cell death	37
NCF4	Phagocyte oxidative burst, antigen presentation	38, 39
APOBR	Macrophage receptor for apolipoprotein B48	40
IL27	Regulation of T_H1 and T_H2 responses, Treg cell function, epithelial cell proliferation	41-44
ATXN2L	Cytokine signaling	45
LAT	T-cell development and activation	46-48

mTOR, Mammalian target of rapamycin.

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 ($P < .05$) 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.[27,28] Furthermore, it has a critical yet complex role in T_H17 differentiation and function[29]: on the one hand, it represses T_H17 signature genes (eg, *IL17A*), but on the other hand, it promotes expression of genes that drive T_H17 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

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.[53] 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.[54] 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 T_H2 responses[41,55] 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.[56,57] Using data from release v7 of the GTEx project,[58] 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.[59] 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[6] 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 an increased probability of success if prioritized for clinical development.[60] 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, APOBR, IL27, ATXN2L,* 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

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.
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.
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.
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.
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.
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.
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.
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.
9. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11:e1004219.
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.
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.
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.
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.
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; https://doi.org/10.1101/045260.
15. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648-60.
16. Odhams CA, Cunninghame 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.
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; https://doi.org/10.1101/166298.
18. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression

distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291-5.
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.
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.
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.
22. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506-11.
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.
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.
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.
26. 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.
27. Ubieta K, Garcia M, Grotsch 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. Venkateswarlu K. Interaction protein for cytohesin exchange factors 1 (IPCEF1) binds cytohesin 2 and modifies its activity. J Biol Chem 2003;278:43460-9.
35. 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.
36. Oh SJ, Santy LC. Differential effects of cytohesins 2 and 3 on beta1 integrin recycling. J Biol Chem 2010;285:14610-6.
37. Thedieck K, Polak P, Kim ML, Molle KD, Cohen A, Jeno P, et al. PRAS40 and PRR5-like protein are new mTOR interactors that regulate apoptosis. PLoS One 2007;2:e1217.
38. Ellson CD, Davidson K, Ferguson GJ, O'Connor R, Stephens LR, Hawkins PT. Neutrophils from p40phox-/- mice exhibit severe defects in NADPH oxidase regulation and oxidant-dependent bacterial killing. J Exp Med 2006;203:1927-37.
39. 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.
40. 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.
41. 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.
42. Do J, Kim D, Kim S, Valentin-Torres A, Dvorina N, Jang E, et al. Treg-specific IL-27Ralpha 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.
43. 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.
44. Yang B, Suwanpradid 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.
45. 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.
46. 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.
47. 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.
48. Zhang W, Sloan-Lancaster J, Kitchen J, Trible RP, Samelson LE. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. Cell 1998;92:83-92.
49. Xu Z, Wu C, Wei P, Pan W. A powerful framework for integrating eQTL and GWAS summary data. Genetics 2017;207:893-902.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.
55. Jirmo 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.
56. 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.
57. Ng SL, Friedman BA, Schmid S, Gertz J, Myers RM, Tenoever 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.
58. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. Nature 2017;550:204-13.
59. 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.
60. 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.

Eleven loci with new reproducible genetic associations with allergic disease risk

ONLINE REPOSITORY

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

E Tables

Tables E1 to E5 are provided in a separate Excel workbook.

Rationale for performing the conditional analysis prior to the gene-based analysis

The goal of our gene-based analysis was to identify genes with a significant association with allergic disease risk that was not driven by the 136 sentinel allergy risk SNPs reported by Ferreira et al. [1]. In practice, applying the gene-based test to the original results of Ferreira et al. [1] (i.e. before conditional analysis) would not be the best way to achieve that goal. Had we done so, there would be hundreds (exactly 362) of genes with a significant association at the genome-wide significance threshold of $P<2.5\text{x}10^{-6}$ (Figure E2). Most of the 362 genes are located in close proximity to (<1 Mb) the 136 sentinel allergy risk SNPs that we reported by Ferreira et al. [1], and so the observed gene-based associations are very likely to be driven by those known allergy risk SNPs. As indicated above, the goal of our analysis was not to identify such genes, but instead those with a gene-based association driven by potentially new allergy risk SNPs.

To do so, we could have simply filtered out genes that were located "close" to the 136 sentinel risk SNPs reported by Ferreira et al. [1]. But that is not an ideal approach, because LD patterns are not uniform throughout the genome, with substantial LD between variants found at distances of >1 Mb in some regions [2]. So if we had defined "close" as <1 Mb from a sentinel variant, then we could potentially fail to filter out genes located at a greater distance but still with an association driven by a known allergy risk SNP. Conversely, had we used a distance threshold that would ensure linkage equilibrium across the genome between the Ferreira et al. [1] sentinel risk SNPs and eQTL (e.g. >10 Mb), we would potentially encounter the opposite problem, and filter out genes with an association not driven by a known allergy risk SNP.

Another approach could have been to remove results for the 136 sentinel risk SNPs (and for any other variants in LD with those) from the Ferreira et al. [1] GWAS. Then apply the gene-based test after those

exclusions. In our view, this is more appropriate than the "distance-based" approach described above, but still not ideal, because it requires defining a threshold to identify variants in LD with the sentinel risk SNPs. Should this be an r^2 of 0.10, 0.05, or 0.01? The threshold selected could in practice turn out to be too liberal or too conservative.

Instead of the distance- or LD-based filtering approaches described above, we argue that a more efficient (and still simple and robust) approach to identify genes with an association not driven by a sentinel risk SNP identified in Ferreira et al. [1], is to first adjust the single-SNP results of that GWAS for the effects of the sentinel risk SNPs identified in that GWAS. Then apply the gene-based test to the adjusted GWAS results. Approximate conditional association analysis can be (and was in our study) performed with the GCTA tool [2], requiring only summary statistics from the GWAS and genotype data from individuals with an ancestry that is representative of the studies that were included in the GWAS (we used 5,000 Europeans from the UK Biobank).

In practice, the *P*-value for any variant located >10 Mb from a sentinel risk SNP is the same before and after the conditional analysis performed with GCTA, because both variants are considered to be in linkage equilibrium. For variants located <10 Mb from a sentinel variant, the *P*-value after the conditional analysis in GCTA will be attenuated if in (and in proportion to the) LD with the nearby sentinel risk SNP(s). At one extreme (when r^2=0), the *P*-value will be (almost if not) identical before and after conditional analysis. On the other hand, if the r^2 between the test variant and the sentinel variant is 1, then the *P*-value after conditioning on the sentinel variant will be 1 (or very close to).

When we applied the gene-based test to the adjusted GWAS results, we identified 30 genes with a $P<2.5\text{x}10^{-6}$. For most of these genes, the gene-based *P*-value obtained with the original GWAS results

(i.e. prior to the conditional analysis) was virtually unchanged (see Figure E2), confirming that the conditional analysis had not introduced biases that affected our results.

Regarding the observed inflation of test statistics after the conditional analysis (lambda of 1.17), it is important to note that lambda was greater (1.19) in our original GWAS[1] (i.e. prior to the conditional analysis). Therefore, the conditional analysis did not cause a systematic inflation of test statistics. We showed in Ferreira et al. [1] that most of this inflation is likely to reflect the polygenic nature of allergic disease risk and the large sample size used, which results in many thousands of SNPs being truly associated with disease risk (which increases lambda). This is likely to be the case because the inflation due to technical biases was only 1.04 (intercept from LD score regression analysis [2]). The Ferreira et al. GWAS results were adjusted by this inflation prior to the gene-based analysis reported in our current study.

Collaborators of the SHARE study

Jorge Esparza-Gordillo[1,2,3], Oliver Hummel[1], Sarah Grosche[1,2], John S Witte[4], Jouke-Jan Hottenga[5], Gonneke Willemsen[5], Elke Rodríguez[9], Melanie Hotze[9], Andre Franke[10], Melanie C Matheson[11], Shyamali C Dharmage[11], Andreas Arnold[12], Georg Homuth[13], Carsten O Schmidt[14], Philip J Thompson[15], Nicholas G Martin[16], David L Duffy[16], Natalija Novak[17], Holger Schulz[18,19], Stefan Karrasch[18,20], Christian Gieger[21], Konstantin Strauch[22,23], Ronald B Melles[24] and David A Hinds[25]

1 Max Delbrück Center (MDC) for Molecular Medicine, Berlin, Germany

2 Clinic for Pediatric Allergy, Experimental and Clinical Research Center of Charité Universitätsmedizin Berlin and Max Delbrück Center, Berlin, Germany

3 Current address: GlaxoSmithKline, Stevenage, UK

4 Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

5 Department Biological Psychology, Netherlands Twin Register , Vrije University, Amsterdam, The Netherlands

9 Department of Dermatology, Allergology and Venereology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

10 Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany

11 Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

12 Clinic and Polyclinic of Dermatology, University Medicine Greifswald, Greifswald, Germany

13 Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany

14 Institute for Community Medicine, Study of Health in Pomerania/KEF, University Medicine

Greifswald, Greifswald, Germany

15 Institute for Respiratory Health, Harry Perkins Institute of Medical Research, University of Western Australia, Nedlands, Australia

16 Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Australia

17 Department of Dermatology and Allergology, University-Hospital Bonn, Bonn, Germany

18 Institute of Epidemiology I, Helmholtz Zentrum Munchen - German Research Center for Environmental Health , Neuherberg, Germany

19 Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research, Munich, Germany

20 Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany

21 Research Unit of Molecular Epidemiology and Institute of Epidemiology II,, Helmholtz Zentrum Munchen - German Research Center for Environmental Health , Neuherberg, Germany

22 Institute of Genetic Epidemiology, Helmholtz Zentrum Munchen - German Research Center for Environmental Health , Neuherberg, Germany

23 Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany

24 Division of Research, Kaiser Permanente Northern California, Oakland, California, USA

25 23andMe, Inc., Mountain View, California, USA

Acknowledgments

23andMe: We would like to thank the research participants and employees of 23andMe for making this work possible. We particularly thank the following members of the 23andMe Research Team: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson.

LifeLines: The LifeLines Biobank initiative has been made possible by funds from FES (Fonds Economische Structuurversterking), SNN (Samenwerkingsverband Noord Nederland) and REP (Ruimtelijk Economisch Programma). The authors wish to acknowledge the services of the LifeLines Cohort Study, the contributing research centres delivering data to LifeLines, and all the study participants.

SALTY/TWINGENE/CATSS: We acknowledge funding from the Swedish Research Council. Financial support was provided by the Swedish Research Council through the Swedish Initiative for research on Microdata in the Social and Medical Sciences (SIMSAM) framework grant number 340-2013-5867 and grants from the Swedish Heart-Lung Foundation. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX.

ALSPAC: The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. ALSPAC GWAS data was generated by

Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

AAGC: The AAGC was funded by a grant from the NHMRC (project ID 613627).

GENEVA: The project received infrastructure support through the DFG Clusters of Excellence "Inflammation at Interfaces" (grants EXC306 and EXC306/2), and was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (sysINFLAME, grant # 01ZX1306A), and the PopGen 2.0 network (01EY1103). The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

NTR: This study was supported by multiple grants from the Netherlands Organization for Scientific Research (NWO: 016-115-035, 463-06-001, 451- 04-034); ZonMW (31160008, 911-09-032); and NWO 480-15-001/674: Netherlands Twin Registry Repository: researching the interplay between genome and environment; The Amsterdam Public Health Institute (APH) and Neuroscience Campus Amsterdam (NCA); Biomolecular Resources Research Infrastructure (BBMRI–NL, 184.021.007), European Research Council (ERC-230374); Genotyping was made possible by grants from NWO/SPI 56-464-14192, Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health, Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls (USA) and the National Institutes of Health (NIH R01 HD042157-01A1,

MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995).

GENUFAD-SHIP-1.We thank all individuals and families for their participation in this study. We thank all physicians and nurses involved in patient recruitment for their valuable contribution to the study. We are grateful to the laboratory technicians C. Flachmeier and T. Thuss for their work. The study was funded by the German Ministry of Education and Research (BMBF) through the Clinical Research Group for Allergy at Charité Berlin, the National Genome Research Network (NGFN). The SHIP authors are grateful to Mario Stanke for the opportunity to use his server cluster for SNP imputation. We thank all staff members and participants of the SHIP studies, as well as all of the genotyping staff for generating the SHIP SNP data set. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network ‘Greifswald Approach to Individualized Medicine (GANI_MED)’ funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data were supported by the Federal Ministry of Education and Research (grant 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany, and the Federal State of Mecklenburg–West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH.

GENUFAD-SHIP-2. We thank all individuals and families for their participation in this study. We thank all physicians and nurses involved in patient recruitment for their valuable contribution to the study. We are grateful to the laboratory technicians C. Flachmeier and T. Thuss for their work. The study was funded by the German Ministry of Education and Research (BMBF) through the Clinical Research

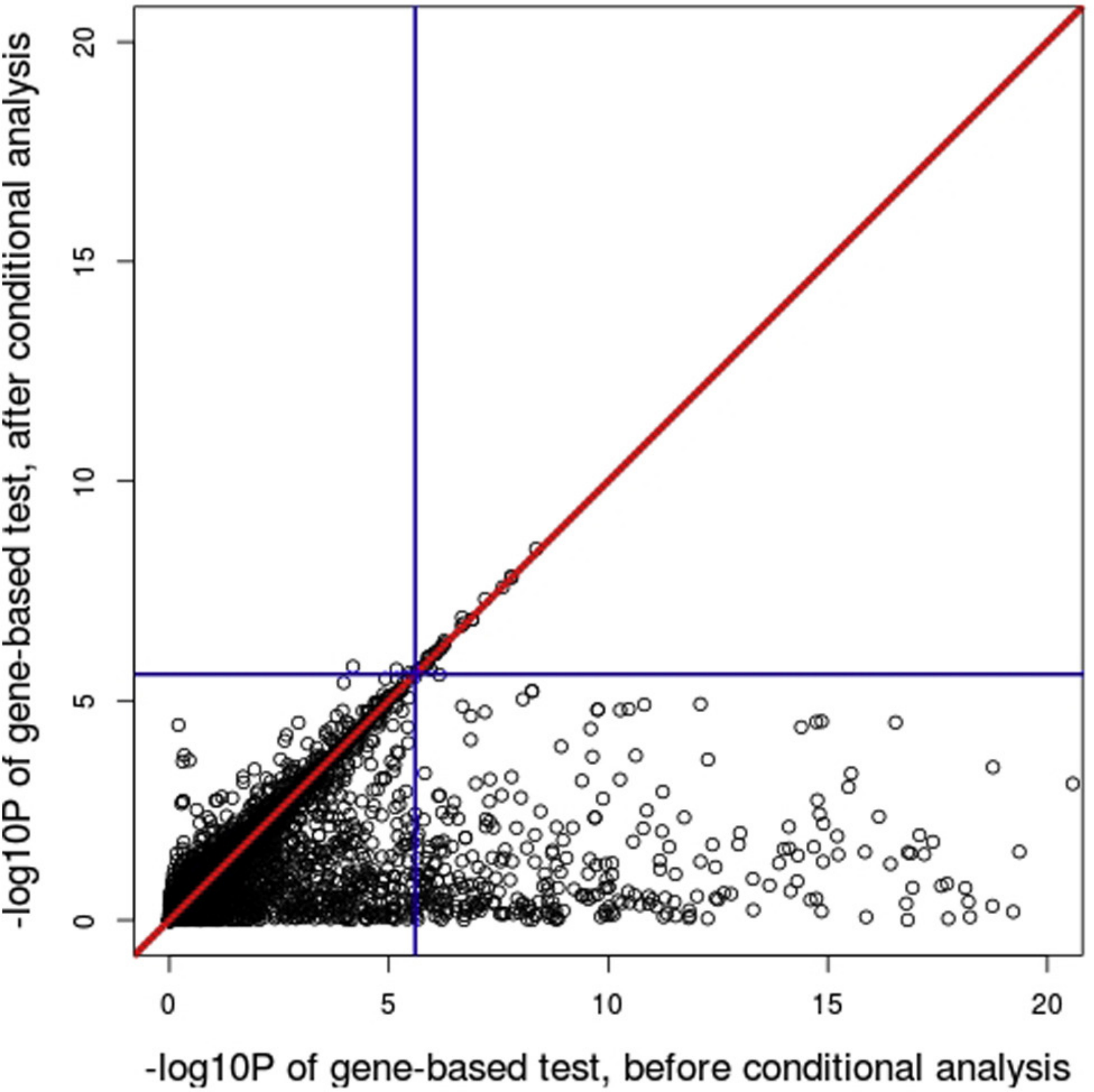
Group for Allergy at Charité Berlin, the National Genome Research Network (NGFN). The SHIP authors are grateful to Mario Stanke for the opportunity to use his server cluster for SNP imputation. We thank all staff members and participants of the SHIP studies, as well as all of the genotyping staff for generating the SHIP SNP data set. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network ‘Greifswald Approach to Individualized Medicine (GANI_MED)’ funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data were supported by the Federal Ministry of Education and Research (grant 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany, and the Federal State of Mecklenburg–West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH.

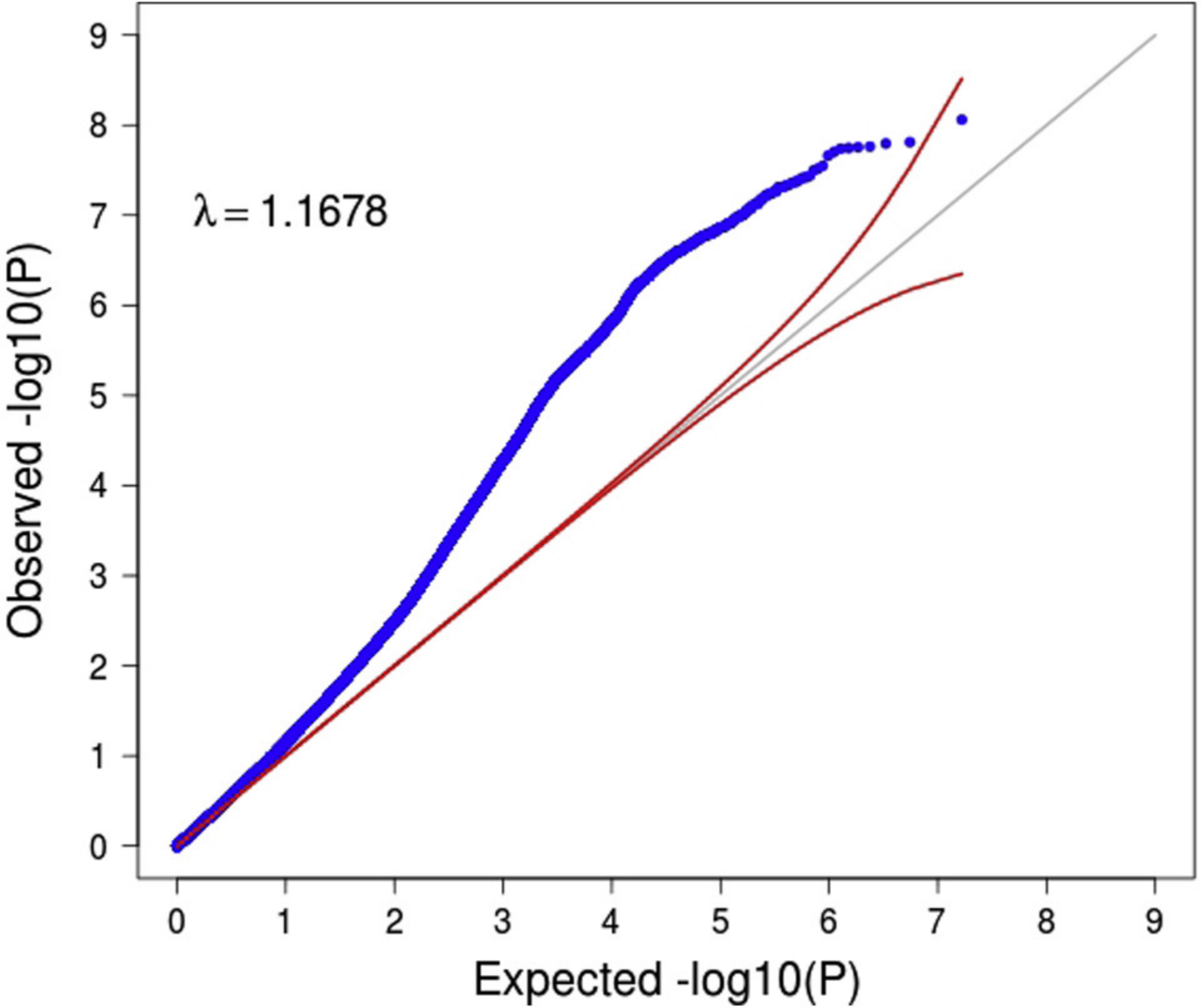
References

1. 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.
2. 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.

Figure E1. Distribution of the observed and expected single-SNP association P-values obtained after adjusting the results from Ferreira et al. [1] (which included a correction for an inflation factor of 1.04; see Methods for details) for the effects of the 136 independent associations identified in that study. The genomic inflation factor (λ, estimated as the median chi-square divided by 0.4549) of the single-SNP results after the conditional analysis is also shown (λ = 1.17); before the conditional analysis, λ was 1.19. The intercept of LD-score regression was 1.00, both before and after the conditional analysis.

Figure E2. Comparison of results from the EUGENE gene-based test obtained before (x-axis) and after (y-axis) adjusting the Ferreira et al.[1] GWAS results for the effects of the sentinel risk SNPs identified in that GWAS. The vertical and horizontal blue lines indicate the P-value threshold used to identify significant associations after accounting for multiple testing (P=2.5x10^{-6}).


-log10P of gene-based test, after conditional analysis
-log10P of gene-based test, before conditional analysis
0
5
10
15
20
0
5
10
15
20


λ = 1.1678
Observed -log10(P)
Expected -log10(P)
0
1
2
3
4
5
6
7
8
9