Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling

Manuel A. R. Ferreira, PhD,[a] Rick Jansen, PhD,[b] Gonneke Willemsen, PhD,[c] Brenda Penninx, PhD,[b] Lisa M. Bain, BSc,[a] Cristina T. Vicente, BSc,[a] Joana A. Revez, BSc,[a] Melanie C. Matheson, PhD,[d] Jennie Hui, PhD,[e,f,g,h] Joyce Y. Tung, PhD,[i] Svetlana Baltic, PhD,[j] Peter Le Souëf, FRACP,[k] Grant W. Montgomery, PhD,[a] Nicholas G. Martin, PhD,[a] Colin F. Robertson, FRACP,[l] Alan James, FRACP,[h,m,n] Philip J. Thompson, FRACP,[j,m] Dorret I. Boomsma, PhD,[c] John L. Hopper, PhD,[d] David A. Hinds, PhD,[i] Rhiannon B. Werder, B Biomed Sc,[o]* Simon Phipps, PhD,[o]* and the Australian Asthma Genetics Consortium Collaborators‡

Brisbane, Melbourne, Nedlands, Crawley, and Subiaco, Australia; Amsterdam, The Netherlands; and Mountain View, Calif

Background: Hundreds of genetic variants are thought to contribute to variation in asthma risk by modulating gene expression. Methods that increase the power of genome-wide association studies (GWASs) to identify risk-associated variants are needed.

Objective: We sought to develop a method that aggregates the evidence for association with disease risk across expression quantitative trait loci (eQTLs) of a gene and use this approach to identify asthma risk genes.

Methods: We developed a gene-based test and software package called EUGENE that (1) is applicable to GWAS summary statistics; (2) considers both *cis*- and *trans*-eQTLs; (3) incorporates eQTLs identified in different tissues; and (4) uses simulations to account for multiple testing. We applied this approach to 2 published asthma GWASs (combined n = 46,044) and used mouse studies to provide initial functional insights into 2 genes with novel genetic associations.

Results: We tested the association between asthma and 17,190 genes that were found to have *cis*- and/or *trans*-eQTLs across 16 published eQTL studies. At an empirical FDR of 5%, 48 genes were associated with asthma risk. Of these, for 37, the association was driven by eQTLs located in established risk loci for allergic disease, including 6 genes not previously implicated in disease cause (eg, *LIMS1*, *TINF2*, and *SAFB*). The remaining 11 significant genes represent potential novel genetic associations with asthma. The association with 4 of these replicated in an independent GWAS: *B4GALT3*, *USMG5*, *P2RY13*, and *P2RY14*, which are genes involved in nucleotide synthesis or nucleotide-dependent cell activation. In mouse studies, P2ry13 and P2ry14—purinergic receptors activated by adenosine 5-diphosphate and UDP-sugars, respectively—were upregulated after allergen challenge, notably in airway epithelial cells, eosinophils, and neutrophils. Intranasal exposure with receptor agonists induced the release of IL-33 and subsequent eosinophil infiltration into the lungs.

Conclusion: We identified novel associations between asthma and eQTLs for 4 genes related to nucleotide synthesis/signaling and demonstrated the power of gene-based analyses of GWASs.

(J Allergy Clin Immunol 2017;139:1148-57.)

Key words: *Inflammation, expression quantitative trait locus, transcriptome, predisposition, obesity, EUGENE, VEGAS, PrediXcan, TWAS,* ZNF707*,* AOAH*,* CLK3*, UDP-glucose, P2Y14, P2Y13*

Asthma is a highly polygenic disease with potentially hundreds or thousands of risk variants, with small effects contributing to variation in disease risk.[1] A small number of risk-associated variants has been identified through genome-wide association studies (GWASs), but the majority remain to be mapped. Identifying risk-associated variants is important because these could point to genes that were not previously suspected to be involved in disease pathophysiology[2,3] or that could represent drug targets with greater probability of clinical success.[4,5]

Several approaches have been proposed to increase the power of GWASs to identify variants with a modest but reproducible association with disease risk. These include larger sample sizes, analysis of more refined phenotypes,[6,7] multivariate association analysis of related phenotypes,[8] gene-based association analyses,[9,10] and association analyses restricted to functional

From [a]QIMR Berghofer Medical Research Institute, Brisbane; [b]the Department of Psychiatry, VU University Medical Center, Amsterdam; [c]the Department of Biological Psychology, Vrije University Amsterdam, Amsterdam; [d]Melbourne School of Population and Global Health, University of Melbourne, Melbourne; [e]PathWest Laboratory Medicine of Western Australia, Nedlands; [f]the School of Population Health, [g]the School of Pathology and Laboratory Medicine, and [m]the School of Medicine and Pharmacology, University of Western Australia, Nedlands; [h]the Busselton Population Medical Research Foundation, Sir Charles Gairdner Hospital, Nedlands; [i]23andMe Inc, Mountain View; [j]the Institute for Respiratory Health, Harry Perkins Institute of Medical Research, Nedlands; [k]the School of Paediatrics and Child Health, Princess Margaret Hospital for Children, Subiaco; [l]Respiratory Medicine, Murdoch Children's Research Institute, Melbourne; [n]the Department of Pulmonary Physiology and Sleep Medicine, West Australian Sleep Disorders Research Institute, Nedlands; and [o]the School of Biomedical Sciences, University of Queensland, Brisbane.

*These authors contributed equally to this work.

‡Australian Asthma Genetics Consortium (AAGC) collaborators are listed in the Online Repository at www.jacionline.org.

Supported in part by the National Human Genome Research Institute of the National Institutes of Health (R44HG006981) and the National Health and Medical Research Council of Australia (613627 and APP1036550).

Disclosure of potential conflict of interest: M. A. R. Ferreira receives grant support from the National Health and Medical Research Council of Australia (grants 613627 and APP1036550). S. Phipps receives grant support from Pfizer. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication February 19, 2016; revised July 8, 2016; accepted for publication July 12, 2016.

Available online August 20, 2016.

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

http://dx.doi.org/10.1016/j.jaci.2016.07.017

Abbreviations used
ADP: Adenosine 5-diphosphate
AEC: Airway epithelial cell
ATP: Adenosine 5-triphosphate
BALF: Bronchoalveolar lavage fluid
eQTL: Expression quantitative trait locus
FDR: False discovery rate
GWAS: Genome-wide association study
HDM: House dust mite
LD: Linkage disequilibrium
SNP: Single nucleotide polymorphism
UDP: Uridine-diphosphoglucose (UDP-glucose)

variants, such as those that regulate gene expression levels.[11] The aim of this study was to develop a method that combined the 2 latter approaches and apply it to results from a published asthma GWAS to help identify new genes with expression associated with genotype and related to disease risk.

Specifically, we hypothesized that if the expression of a gene is causally related to asthma and gene expression is regulated by multiple independent expression quantitative trait loci (eQTLs), a gene-based approach that captures the aggregate signals from these eQTLs would be expected to improve power over the alternative approach of testing each variant individually. Recently, Gamazon et al[12] described a gene-based association method based on the same concept, called PrediXcan. Briefly, this approach includes 3 steps. First, for a given gene, eQTLs are identified from transcriptome data sets. Second, a model that can be used to predict gene expression levels based on the aggregate effect of those eQTLs is trained on a reference transcriptome data set. Third, this model is used to infer expression levels for a target GWAS data set that includes subjects genotyped for those eQTLs but for whom actual gene expression levels might not be available. The genetically inferred gene expression levels can then by tested for association with the phenotype of interest (eg, asthma).

As highlighted by Gamazon et al,[12] PrediXcan has several advantages over other gene-based tests, such as VEGAS.[9] However, in our view it has 1 major limitation: unlike VEGAS, it is not applicable to GWAS summary statistics, which are typically more readily available, and therefore can be applied to a larger sample size than available GWAS data sets with individual-level genetic data. The TWAS approach developed by Gusev et al[13] addresses this caveat but in its current release is applicable only to a relatively small number of genes (4,284 from 2 blood eQTL studies), *cis*- but not *trans*-acting eQTLs (eg, those located >1 Mb from the target gene), and to a single reference transcriptome data set at a time.

In this study we developed a gene-based association approach, called EUGENE, that combines the biological focus of PrediXcan and TWAS and the versatility of VEGAS. Our approach also considers eQTL evidence across different tissues and estimates empirical false discovery rates (FDRs) while accounting for the linkage disequilibrium (LD) between variants. We applied this new approach to a published asthma GWAS[14] to try to identify novel genes with a genetic component of gene expression associated with asthma risk. Finally, we investigated whether results from mouse models of experimental acute allergic asthma are consistent with a contribution of 2 selected genes to disease pathophysiology.

METHODS

EUGENE approach

The proposed gene-based approach is described in detail in the Methods section in this article's Online Repository at www.jacionline.org. Briefly, for a given gene, our approach includes 4 steps. First, we identify a set of variants that influence gene expression in any cell type or tissue relevant to the disease or trait of interest based on results from published eQTL studies. Including eQTLs identified in tissues not thought to be relevant for the disease of interest might improve power, but this is something we did not consider in our study. We include in this list eQTLs located in *cis* (≤1 Mb from the target gene) or *trans* (>1 Mb away or in a different chromosome). This list is then reduced to a subset of eQTLs with an LD r^2 value of less than 0.1; we refer to these as "independent eQTLs" for a given gene (see Fig E1 in this article's Online Repository at www.jacionline.org).

Second, we extract association results for these independent eQTLs from a disease or trait GWAS of interest and then calculate the gene-based statistic Q as the sum of the 1-df χ^2 values for the individual eQTLs. This represents the aggregate evidence for association in that GWAS across the independent eQTLs of that gene.

Third, we perform simulations using individual-level genetic data to estimate the statistical significance of Q while accounting for the residual LD between eQTLs.

Fourth, FDR thresholds are also estimated empirically to account for multiple testing. Simulations show that the type 1 error rate of EUGENE is close to the nominal expectation (see Table E1 in this article's Online Repository at www.jacionline.org). The software and input files required to run EUGENE are freely available at https://genepi.qimr.edu.au/staff/manuelF.

Application of EUGENE to published GWASs of asthma

We applied EUGENE to a published asthma GWAS[14] to illustrate the utility of the proposed approach. This GWAS included 6,685 patients with both asthma and hay fever and 14,091 asthma- and hay fever–free control subjects, all of European descent, tested for association with 4.9 million single nucleotide polymorphisms (SNPs) with a minor allele frequency of greater than 1%. In the original analysis of individual SNPs, 11 independent variants were found to be associated with disease risk at a genome-wide significance level of a P value of less than 3×10^{-8}. We used EUGENE to identify genes with an association with disease risk in the study by Ferreira et al[14] at an empirical FDR of 0.05 (corresponding to a P value threshold of 1.9×10^{-4}). At this FDR level, 5% of genes considered significant (ie, $P < 1.9 \times 10^{-4}$) are expected to be false-positive associations due to multiple testing.

To confirm putative novel associations, we then applied EUGENE to an independent asthma GWAS, the GABRIEL study,[15] for which summary statistics are publicly available. After excluding overlapping samples (the Busselton study), results from the GABRIEL study were based on 9,967 asthmatic patients and 15,301 control subjects.

Predicted direction of effect of gene expression on asthma risk

EUGENE can be used to identify a set of genes with expression levels determined by eQTLs and for which the eQTLs are collectively associated with disease risk. However, unlike PrediXcan[12] or TWAS,[13] EUGENE does not directly provide the predicted direction of effect of gene expression on disease risk. To understand whether a genetically determined increase in gene expression levels was predicted to increase or decrease disease risk, we compared the direction of effect of each eQTL on gene expression reported on the transcriptome GWAS with the effect on asthma risk reported in the Ferreira et al[14] asthma GWAS. Based on this information, for each eQTL, we report whether the allele associated with increased gene expression is associated with an increased or decreased asthma risk.