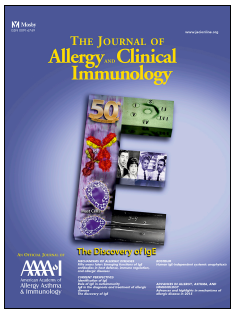
Accepted Manuscript



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

Manuel A.R. Ferreira, PhD, Rick Jansen, PhD, Gonneke Willemsen, PhD, Brenda Penninx, PhD, Lisa M. Bain, BSc, Cristina T. Vicente, BSc, Joana A. Revez, BSc, Melanie C. Matheson, PhD, Jennie Hui, PhD, Joyce Y. Tung, PhD, Svetlana Baltic, PhD, Peter Le Souëf, FRACP, Grant W. Montgomery, PhD, Nicholas G. Martin, PhD, Colin F. Robertson, FRACP, Alan James, FRACP, Philip J. Thompson, FRACP, Dorret I. Boomsma, PhD, John L. Hopper, PhD, David A. Hinds, PhD, Rhiannon B. Werder, BSc, Simon Phipps, PhD

PII: S0091-6749(16)30804-1

DOI: 10.1016/j.jaci.2016.07.017

Reference: YMAI 12294

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 19 February 2016

Revised Date: 8 July 2016

Accepted Date: 12 July 2016

Please cite this article as: Ferreira MAR, Jansen R, Willemsen G, Penninx B, Bain LM, Vicente CT, Revez JA, Matheson MC, Hui J, Tung JY, Baltic S, Le Souëf P, AAGC collaborators, Montgomery GW, Martin NG, Robertson CF, James A, Thompson PJ, Boomsma DI, Hopper JL, Hinds DA, Werder RB, Phipps S, Gene-based analysis of regulatory variants identifies four putative novel asthma risk genes related to nucleotide synthesis and signaling, *Journal of Allergy and Clinical Immunology* (2016), doi: 10.1016/j.jaci.2016.07.017.

Gene-based analysis of regulatory variants identifies four 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,[e] Jennie Hui, PhD,[f-i] Joyce Y. Tung, PhD,[j] Svetlana Baltic, PhD,[k] Peter Le Souëf, FRACP,[l] AAGC collaborators[m], Grant W. Montgomery, PhD,[a] Nicholas G. Martin, PhD,[a] Colin F. Robertson, FRACP,[n] Alan James, FRACP,[i,o,p] Philip J. Thompson, FRACP,[k,o] Dorret I. Boomsma, PhD,[c] John L. Hopper, PhD,[e] David A. Hinds, PhD,[j] Rhiannon B. Werder, BSc,[q*] and Simon Phipps, PhD,[q*]

* Equal contributions.

Affiliations

[a] QIMR Berghofer Medical Research Institute, Brisbane 4029, Australia.

[b] Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands

[c] Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands

[e] Melbourne School of Population and Global Health, The University of Melbourne, Melbourne 3010, Australia.

[f] PathWest Laboratory Medicine of Western Australia (WA), Nedlands 6009, Australia

[g] School of Population Health, The University of WA, Nedlands 6009, Australia

[h] School of Pathology and Laboratory Medicine, The University of WA, Crawley 6009, Australia

[i] Busselton Population Medical Research Foundation, Sir Charles Gairdner Hospital, Nedlands 6009, Australia

[j] 23andMe, Mountain View, California 94041, USA

[k] Institute for Respiratory Health, Harry Perkins Institute of Medical Research, Nedlands 6009, Australia.

[l] School of Paediatrics and Child Health, Princess Margaret Hospital for Children, Subiaco 6008, Australia

[m] Australian Asthma Genetics Consortium (AAGC) collaborators are listed in the Online repository

[n] Respiratory Medicine, Murdoch Children's Research Institute, Melbourne 3052, Australia

[o] School of Medicine and Pharmacology, University of Western Australia, Nedlands 6009, Australia

[p] Department of Pulmonary Physiology and Sleep Medicine, West Australian Sleep Disorders Research Institute, Nedlands 6009, Australia

[q] School of Biomedical Sciences, The University of Queensland, Brisbane 4072, Australia

Corresponding author:

Manuel A R Ferreira, PhD

QIMR Berghofer Medical Research Institute

Locked Bag 2000, Royal Brisbane Hospital,

Herston QLD 4029, Australia

Phone: +61 7 3845 3552

Fax: +61 7 3362 0101

Email: manuel.ferreira@qimrberghofer.edu.au

Sources of funding: This work was 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).

ABSTRACT

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 (GWAS) to identify risk-associated variants are needed.

Objective: 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 two published asthma GWAS (combined N=46,044) and used mouse studies to provide initial functional insights into two genes with novel genetic associations.

Results: We tested the association between asthma and 17,190 genes which were found to have *cis*- and/or *trans*-eQTLs across 16 published eQTL studies. At an empirical false discovery rate 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 six genes not previously implicated in disease aetiology (eg. *LIMS1*, *TINF2* and *SAFB*). The remaining 11 significant genes represent potential novel genetic associations with asthma. The association with four 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 ADP and UDP-sugars, respectively – were up-regulated after allergen challenge, notably in airway epithelial cells, eosinophils and neutrophils. Intranasal exposure with the 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 four genes related to nucleotide synthesis/signaling, and demonstrate the power of gene-based analyses of GWAS.

KEY MESSAGES

- In humans, asthma risk is associated with genetically-determined expression of four genes related to nucleotide synthesis (*B4GALT3, USMG5*) and nucleotide-dependent cell activation (*P2RY13* and *P2RY14*).
- In mice, intranasal exposure with selective agonists for P2ry13 (ADP) or P2ry14 (UDP-glucose) induced the release of IL-33 and eosinophil infiltration into the lungs, in the absence of allergen stimulation.

CAPSULE SUMMARY

Using a new method for gene-based analysis of GWAS results, we identified a genetic association between asthma risk and eQTLs for *B4GALT3* and *USMG5*, which are involved in the production of UDP-galactose and ATP respectively, and *P2RY13* and *P2RY14*, two G protein-coupled receptors activated respectively by ADP and UDP-sugars. Functional studies in the mouse show that activation of P2ry13 or P2ry14 induces the release of IL-33 and eosinophil infiltration into the lungs, in the absence of allergen stimulation. Functional studies that characterize in depth the contribution of these four genes to asthma pathophysiology are warranted.

KEY WORDS

Inflammation, eQTL, transcriptome, predisposition, obesity, EUGENE, VEGAS, PrediXcan, TWAS, ZNF707, AOAH, CLK3, UDP-glucose

ABBREVIATIONS

AECs: airway epithelial cells.

ADP: adenosine 5-diphosphate.

Alt: Alternaria allergen.

ATP: adenosine 5-triphosphate.

BALF: bronchoalveolar lavage fluid.

CRE: cockroach allergen.

DC: dendritic cells.

eQTL: expression quantitative trait locus.

FDR: false discovery rate.

GWAS: Genome Wide Association Study.

HDM: house dust mite.

IL: interleukin.

LD: linkage disequilibrium.

SNP: single nucleotide polymorphism.

UDP: uridine-diphosphoglucose (UDP-glucose)

INTRODUCTION

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 (GWAS), 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 (eg. [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 GWAS to identify variants with a modest but reproducible association with disease risk. These include larger sample sizes, the 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 variants, such as those that regulate gene expression levels [11]. The aim of this study was to develop a method that combined the two latter approaches and apply it to results from a published asthma GWAS to help identify new genes whose expression was 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) , then 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 three 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. And third, this model is used to infer

133 expression levels for a target GWAS data set that includes individuals genotyped for those eQTLs but
134 for whom actual gene expression levels might not be available. The genetically-inferred gene
135 expression levels can then tested for association with the phenotype of interest (eg. asthma).
136
137 As highlighted by Gamazon et al. [12], PrediXcan has several advantages over other gene-based tests,
138 such as VEGAS [9]. However, in our view, it has one major limitation: unlike VEGAS, it is not
139 applicable to GWAS summary statistics, which are typically more readily available, and therefore can
140 be applied to a larger sample size than available GWAS data sets with individual-level genetic data.
141 The TWAS approach developed by Gusev et al. [13] addresses this caveat, but in its current release is
142 applicable only to a relatively small number of genes (4,284 from two blood eQTL studies), *cis*- but not
143 *trans*-acting eQTLs (eg. those located >1 Mb from the target gene), and to a single reference
144 transcriptome dataset at a time.
145
146 In this study, we developed a gene-based association approach, called EUGENE, that combines the
147 biological focus of PrediXcan and TWAS, and the versatility of VEGAS. Our approach also considers
148 eQTL evidence across different tissues and estimates empirical false discovery rates (FDR), while
149 accounting for the LD between variants. We applied this new approach to a published asthma GWAS [14]
150 to try to identify novel genes whose genetic component of gene expression is associated with asthma
151 risk. Finally, we investigated whether results from mouse models of experimental acute allergic asthma
152 are consistent with a contribution of two selected genes to disease pathophysiology.

METHODS

EUGENE approach

The proposed gene-based approach is described in detail in the **Online Repository**. Briefly, for a given gene, our approach includes four 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 sub-set of eQTLs with linkage disequilibrium (LD) r^2<0.1; we refer to these as "independent eQTLs" for a given gene. Second, we extract association results for these independent eQTLs from a disease or trait GWAS of interest and then calculate a gene-based statistic *Q*, as the sum of the 1-df chi-squares 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, false-discovery rate (FDR) thresholds are also estimated empirically to account for multiple testing. Simulations show that the type-I error rate of EUGENE is close to the nominal expectation (**Table E1 in the Online Repository)**. 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 GWAS of asthma

We applied EUGENE to a published asthma GWAS [14] to illustrate the utility of the proposed approach. This GWAS included 6,685 individuals with both asthma and hayfever and 14,091 asthma- and hayfever-free controls, all of European descent, tested for association with 4.9 million SNPs with a

minor allele frequency >1%. In the original analysis of individual SNPs, eleven independent variants were found to be associated with disease risk at a genome-wide significance level of $P<3x10^{-8}$. We used EUGENE to identify genes with an association with disease risk in the Ferreira et al. study [14] at an empirical FDR of 0.05 (corresponding to a *P*-value threshold of $1.9x10^{-4}$). At this FDR level, 5% of genes called significant (ie. with a $P<1.9x10^{-4}$) are expected to be false-positive associations. 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 asthmatics and 15,301 controls.

Predicted direction of effect of gene expression on asthma risk

EUGENE can be used to identify a set of genes whose expression is 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.

Functional studies in the mouse

We selected two putative novel asthma risk genes for preliminary functional studies in the mouse: *P2RY13* and *P2RY14*. The criteria used to select these genes for functional follow-up were as follows: (1) significant gene-based association with asthma in the discovery GWAS at an empirical FDR of 5%;

(2) the eQTLs that contribute to the significant gene-based association were not in LD (r^2<0.1) with established asthma risk variants (those with a $P<5x10^{-8}$ in published GWAS of asthma, hayfever, eczema and/or allergies); and (3) the gene-based association replicated (P<0.05) in an independent GWAS. Four genes satisfied all three criteria: *P2RY13*, *P2RY14*, *USMG5* and *B4GALT3*. We prioritized the former two for follow-up because functional experiments were feasible with available tools/reagents (both are cell-surface receptors). We performed two sets of experiments, which are described in detail in the **Online Repository** and were performed in accordance with the Animal Care and Ethics Committees of the University of Queensland (Brisbane, Australia).

First, we used an established mouse model of acute allergic asthma [16] to identify the cell types in the lung that express P2ry13 and P2ry14 in the context of allergen-induced airway inflammation. Two groups of wild-type C57Bl/6 mice were anesthetized and sensitized intranasally (i.n.) with either saline solution (group 1) or 100 μg of HDM extract on day 0. Subsequently, mice were challenged with either saline (group 1) or 5 μg of HDM (group 2) at day 14, 15, 16 and 17 and sacrificed 3 hours later. Total RNA was isolated from the left lung and quantitative real-time PCR performed to measure overall gene expression. To identify individual cell types in the lung expressing P2ry13 and P2ry14, bronchoalveolar lavage fluid (BALF) was collected and cells stained with anti-P2ry13 or anti-P2ry14antibodies. Cells were then stained with cell-type specific fluorescently labeled antibodies and enumerated using a BD LSR Fortessa cytometer. To assess expression in airway epithelial cells, paraffin-embedded lung sections were prepared as previously described [17] and probed with anti-P2ry13 or anti-P2ry14 antibodies. Photomicrographs were taken at 400x and 1000x magnification at room temperature and acquired using Olympus Image Analysis Software.

We performed a second set of experiments to test the hypothesis that P2ry13 or P2ry14 receptor activation could influence the release of alarmins, such as IL-33, and contribute to airway inflammation. Naïve mice were inoculated i.n. with saline, 10 nM 2-methyl-ADP (P2ry13 agonist), 10

nM UDP-glucose (P2ry14 agonist) or 10 nM ATP (agonist for all P2ry receptors, except P2ry6 and P2ry14), all in 50 uL. For comparison, three additional groups of mice were inoculated with 100 ug of HDM, 100 ug of cockroach extract (*Blattella germanica*) or 25 ug of *Alternaria alternata* extract. Two hours post-challenge, BALF was collected as described above and IL-33 levels measured by ELISA. Seventy-two hours post-challenge, BALF was again collected to obtain immune cell counts and stained for flow cytometry as described above.

RESULTS

Application of EUGENE to results from a published asthma GWAS

We applied our proposed gene-based test of association to a published asthma GWAS [14] including 6,685 cases and 14,091 controls to identify genes with eQTLs collectively associated with disease risk. We tested the association with 17,190 genes (**Figure 1**) which were found to have *cis*-eQTLs (N=13,557), *trans*-eQTLs (N=315) or both (N=3,318), across 16 published eQTL studies, representing 12 different cell types or tissues relevant to asthma (**Table E2 in the Online Repository**).

Of the 17,190 genes tested, 48 genes were associated with asthma at an empirical FDR of 0.05 (**Table 1** and **Figure 2**). Of these, 31 (65%) were located within 1 Mb (or on the MHC region) of established risk variants for allergic disease (highlighted with a '+' in **Figure 2** and listed in **Table E3 in the Online Repository**). For example, for *TSLP*[18] (gene-based $P=7x10^{-6}$), we identified six independent *cis*-eQTLs in five tissues, of which four were individually associated with asthma risk at $P<0.05$ (**Table E4 in the Online Repository**). Multiple genes within the same risk locus had significant associations with asthma: 12 in the MHC [15]; 11 on 17q12 [2]; three on 2q12 [15]; and two on 16p13 [14]. Some of these associations resulted from eQTLs being shared between neighboring genes, as observed for *ORMDL3*, *GSDMB* and *ZPBP2* on 17q12 (**Table E5 in the Online Repository)**, and for *CLEC16A* and *SOCS1* on 16p13 (r^2 between rs35441874 and rs7184491 [cf. Table E3] is 0.64). eQTL sharing could arise, for example, if an underlying causal variant disrupts the activity of a regulatory element that controls the expression of multiple genes. But that was not always the case; in the MHC region, the individual *LTA* eQTL that was most strongly associated with asthma ($P=2x10^{-5}$) was in low LD ($r^2 \leq 0.02$) with the eQTLs for the other 11 significant MHC genes (**Table E6 in the Online Repository)**. Similar results were observed for *NEU1*. Therefore, at least in the MHC region, the multiple significant associations observed were not entirely explained by eQTLs shared between genes.

On the other hand, six (12%; *LIMS1*, *AOAH*, *ZNF707*, *CLK3*, *SAFB* and *TINF2*; 'Δ' in **Figure**

2) of the 48 genes significant at an FDR of 0.05 were not located in established risk loci for asthma but the significant gene-based associations were (in most cases, entirely) driven by *trans*-eQTLs located in the MHC region or near *ORMDL3* (**Table E7 in the Online Repository**). These include for example variant rs9268853, which is a *trans*-eQTL for *CLK3* (P=7x10^{-17}) in PBMCs[19], *SAFB* (P=3x10^{-6}) in whole-blood[20] and *AOAH* in three tissues (best P=10^{-61}, [19-21]) (**Table E8 in the Online Repository)**. This variant has also been found to be a *cis*-eQTL (P<5x10^{-8}) for *HLA-DQ* and *HLA-DR* genes across multiple tissues (not shown). These results suggest that MHC and 17q12 variants might contribute to asthma risk not only by directly modulating the expression of nearby genes, but also by indirectly influencing the expression of genes in different chromosomes (eg. through *cis*-mediation [22]).

Of potential greater interest, 11 (23%) of the 48 significant genes were located in potential novel asthma risk loci and the gene-based associations were not driven by established allergy risk variants ('○' in **Figure 2** and **Table 1**). As some of these genes might represent false-positive findings, we studied their association with asthma in an independent GWAS.

Replication of the putative novel gene-based associations in an independent asthma GWAS

To confirm the putative novel associations, we applied EUGENE to an independent GWAS of asthma with publicly available summary statistics [15]. Based on results for 9,967 asthmatics and 15,301 controls, four of the 11 genes selected for replication had a significant gene-based association (P<0.05; **Table E9 in the Online Repository**), when simulations show that on average the expected number of genes significant at this threshold by chance alone given multiple testing was 0.53 (SD=0.77).

We then explored whether the discovery and replication associations for those four genes were consistent by comparing the direction of effect on disease risk for individual eQTLs. Overall, the direction of effect for most eQTLs of a given gene was the same between the two independent GWAS **(Table E10 in the Online Repository**). For example, of the seven eQTLs for *USMG5* that were

individually associated with asthma risk in either study, for six the allele that increased asthma risk was the same (or was on the same haplotype) in both studies; one eQTL was not tested in the replication GWAS, and so the direction of effect could not be compared. Therefore, the association between asthma risk and these four genes is generally consistent at the individual eQTL level between the two independent GWAS. Henceforth, we refer to these four genes with a reproducible gene-based association with asthma as "putative novel asthma risk genes".

Contribution of *cis*- and *trans*-eQTLs to significant gene-based associations

For three (*P2RY13*, *P2RY14*, *USMG5*) of the four putative novel risk genes, the gene-based association with asthma was entirely driven by *cis*-eQTLs. Most of these eQTLs were identified by eQTL studies of whole-blood expression levels (**Table E11 in the Online Repository**). For the fourth gene, *B4GALT3*, three *cis* (in neutrophils, blood and fibroblasts) and one *trans* (in blood) eQTL contributed to the association with asthma (**Table E11 in the Online Repository**). The latter (rs1668873) was located 44 Mb away on chromosome 1 and was previously reported to associate with mean platelet volume and count [23, 24]. This variant is also a *cis*-eQTL for *NUAK2* [20], a nuclear transcriptional modulator that has been shown to induce the expression of *B4GALT5* [25], a galactosyltransferase related to *B4GALT3* [26]. Therefore, these results suggest that both direct (*cis*-eQTLs) and indirect (via transcriptional modulators such as *NUAK2*) genetic effects on *B4GALT3* expression might contribute to asthma risk.

Genetically predicted direction of effect of gene expression on asthma risk

To assess the direction of effect of gene expression on disease risk, we focused on the independent eQTLs for each gene that were individually associated with asthma in the discovery and/or replication GWAS. These variants had the greatest contribution to the significant gene-based tests. When we

compared the direction of effect for each eQTL between asthma risk and expression levels, we found that the allele associated with increased gene expression was also associated with increased asthma risk for all independent eQTLs of *P2RY13* and *P2RY14* (**Table E11 in the Online Repository**). The same pattern of results was observed for six of the seven eQTLs of *USMG5*; for example, the rs1163073:C allele that was associated with asthma risk (OR=1.09, *P*=0.0005), was associated with increased *USMG5* expression in five different cell types or tissues (neutrophils, LCLs, skin, PBMCs and blood). These results suggest that in the tissues or cell types considered in our analysis, a genetically determined increase in gene expression for each of these four genes is associated with increased disease risk. For *B4GALT3*, there was no clear pattern across multiple eQTLs: of the four alleles associated with increased gene expression, two (in neutrophils and whole-blood) were associated with increased and two (in fibroblasts and whole-blood) with decreased disease risk. Such differences between eQTLs could arise, for example, if *B4GALT3* has opposing functional effects on different cell types relevant to asthma (eg. activation in one, inhibition in another). Further studies are required to test this possibility.

Functional studies in the mouse

The four putative novel asthma risk genes identified in our genetic association analysis are involved in nucleotide synthesis (*B4GALT3*, *USMG5*) and nucleotide-dependent cell activation (*P2RY13*, *P2RY14*). Based on this observation, we hypothesise that genetic dysregulation of nucleotide signaling contributes to asthma risk. In depth functional experiments that comprehensively test this hypothesis were beyond the scope of this study. Nonetheless, we carried out two sets of experiments in the mouse to provide preliminary functional support for the involvement in allergic asthma for two of these four nucleotide-related genes: *P2RY13* and *P2RY14*. Both are cell-surface receptors with known agonists, and so were well suited for functional studies.

First, we used an established experimental model of acute allergic asthma[16] to study P2ry13 and

P2ry14 expression in the lungs of C57BL/6 mice sensitized and subsequently challenged with house dust mite (HDM) allergen. In this model, mice develop granulocytic airway inflammation that has a predominant eosinophil contribution[16]. When considering overall lung expression, HDM challenge resulted in a significant increase in *P2ry13* and *P2ry14* expression, relative to control mice challenged with a saline solution (**Figure 3A**). To understand which lung cell types contributed to this increase in gene expression, we used flow cytometry to measure protein expression in airway epithelial cells (AECs) and major immune cell types collected through bronchoalveolar lavage. There was widespread expression of both receptors in AECs, both at baseline and after HDM challenge (**Figure 3B and 3C**). Most eosinophils collected in BALF after HDM challenge stained positive for both receptors (**Figure 3D**); expression in neutrophils was also high (**Figure 3E**). Lymphocytes and dendritic cells had low expression of both receptors (**Figure E2 in the Online Repository**).

Secondly, given the high level of P2ry13 and P2ry14 expression observed in AECs at baseline (ie. in the absence of allergen challenge), and the previously reported pro-inflammatory effect of the respective agonists (eg. [27, 28]), we postulated that receptor activation could promote airway inflammation by inducing the release of alarmins. To test this possibility, we collected BALF from naïve mice 2 and 72 hours after intra-nasal challenge with saline, ADP (selective P2ry13 agonist) or UDP-glucose (selective P2ry14 agonist). At 2 hours post challenge, BALF levels of the alarmin IL-33 were significantly greater in mice exposed to the receptor agonists than in control mice (**Figure 4A**). Of note, nucleotide-induced IL-33 levels were comparable to allergen-induced IL-33 levels, indicating that both ADP and UDP-glucose are sufficient to induce IL-33 release. Furthermore, 72 hours after challenge, the number of BALF eosinophils and lymphocytes were significantly higher in agonist-treated mice (**Figure 4B and 4C**), but this was not the case for neutrophils, dendritic cells and monocytes (**Figure E3 in the Online Repository**). These results demonstrate that selective agonists of P2ry13 and P2ry14 can promote airway inflammation, even in the absence of allergen stimulation.

DISCUSSION

Dysregulation of gene expression is thought to be a common mechanism by which genetic variants can influence cellular function and, ultimately, variation in human traits and disease risk. This proposition is supported by the observation that eQTLs are more likely to be trait-associated than random variants [18] and was the motivation for the gene-based approach developed in this study.

EUGENE has some advantages when compared to other gene-based association approaches, of which we highlight three: VEGAS [9], PrediXcan [12] and TWAS [13]. When compared to VEGAS, EUGENE avoids the requirement to use an arbitrary distance (same for all genes) from the known gene boundaries to define which SNPs to include in the gene-based analysis. If the distance is too small (eg. +/-5 kb), then the contribution of important more distantly located eQTLs might be missed, while a large distance (eg. +/- 1 Mb) could result in testing a large number of variants, many of which are likely to be unrelated to gene expression/function; in both cases, the power to detect a significant gene-based association is reduced. Also, because the analysis in EUGENE is restricted to variants previously shown to influence gene expression, whether in *cis* or in *trans*, a significant gene-based trait association directly implies that genetically-determined differences in gene expression contribute to trait variation. On the other hand, when compared to the recently described PrediXcan approach [12], the main advantage of EUGENE is that it is applicable to GWAS summary statistics, which are typically easier to share than data sets with individual level genetic data. TWAS[13], which is conceptually very similar to PrediXcan, is applicable to summary statistics but its current release only includes weights for a relatively small number of genes obtained from three eQTL studies, although this is likely to increase in the future. Another major difference when compared to both of these approaches is that with EUGENE, eQTLs identified in transcriptome studies of different cell types (and/or upon cell stimulation) can be included in the same association analysis, all contributing with equal weight to the gene-based statistic. This might be important for traits or diseases for which multiple cell types or

tissues are known to play a role in the underlying pathophysiology. In the PrediXcan and TWAS approaches, the weights assigned to different eQTLs in the model used to predict gene expression levels are based on the effect (eg. regression coefficient) of those variants on expression levels measured in a single reference transcriptome data set. The extent to which those weights remain appropriate (ie. yield good prediction) if the reference transcriptome and target data sets are very different (eg. in age composition) is unclear. A disadvantage of EUGENE is that the direction of effect between gene expression and disease risk (or trait variation) is not directly inferred. To do so with the EUGENE approach, the effect of individual eQTLs that contribute to the gene-based test needs to be compared *ad-hoc* between the transcriptome and trait GWAS. These eQTLs also provide a specific small group of variants to test in validation studies. Lastly, EUGENE (but not the other three approaches) estimates FDR thresholds empirically, taking into account the LD between eQTLs of the same or different genes. This is important to account for multiple testing.

When we applied EUGENE to a published GWAS of asthma, we identified 48 genes with a significant gene-based association at an FDR of 0.05, including 11 associations that were not driven by established genetic risk variants for allergic disease. For four of these genes (*B4GALT3*, *USMG5*, *P2RY13* and *P2RY14)*, the association was nominally significant in an independent asthma GWAS and so we refer to these as putative novel asthma risk genes.

B4GALT3 encodes the widely-expressed enzyme β-1,4-galactosyltransferase III that catalyzes the transfer of galactose from UDP-galactose to *N*-acetylglucosamine, to form *N*-acetyllactosamine and UDP [26, 29]. How variation in *B4GALT3* expression might contribute to asthma risk is unclear, but potential mechanisms include activation of β1 integrin [30], which is important in the initiation of T-cell inflammatory responses [31], or by influencing extracellular release of UDP-galactose [32], a P2RY14 agonist.

USMG5 encodes a small subunit of ATP synthase [33], an enzyme responsible for ATP synthesis in

399 the mitochondria. *USMG5* knockdown in HeLa cells causes the loss of ATP synthase, resulting in lower
400 ATP synthesis and slower cell growth [34]. In CD4+ T-cells, mitochondria produce the ATP that is rapidly
401 released into the extracellular space upon cell stimulation [35]. In turn, this ATP establishes an autocrine
402 feedback through purinergic receptors that is essential for proper T-cell activation [36]. Given these
403 observations, we speculate that genetically-determined increased *USMG5* expression results in
404 increased mitochondrial production of ATP, increased extracellular ATP release and increased T-cell
405 activation. In turn, this would translate into an increased risk of asthma. ATP synthase has also been
406 detected at the surface of different cell types [37], where it is thought to play different physiological roles,
407 for example, HDL endocytosis in hepatocytes via P2RY13 activation [38] and non-conventional T-cell
408 activation [39]. Whether USMG5 associates with membrane ATP synthase, and so could potentially
409 influence its ectopic roles, remains to be determined.

410 *P2RY13*, also known as *GPR86* or *GPR94*, is a purinergic receptor highly expressed in the
411 immune system, lung and skin, but also in the brain [40-42]; it displays a significant homology with the
412 nearby *P2RY12* and *P2RY14* genes, sharing 48 and 45% amino acid identity [40]. P2RY13 is strongly
413 activated by ADP [40], a degradation product of ATP. Airway epithelial goblet cells are a major source of
414 extracellular ADP, which is released as a co-cargo molecule from mucin-containing granules [43]. In turn,
415 ADP has been shown to enhance antigen-induced degranulation in mast cells, through a P2RY13-
416 dependent mechanism [27]. ADP has also been reported to promote IL-6 release from keratinocytes [42],
417 inhibit TNF-alpha and IL-12 production by mature DCs [44] and promote chemotaxis of immature DCs
418 [45]; however, these studies did not specifically test if the observed ADP effects were mediated by
419 P2RY13. Results from our genetic association analyses indicate that a genetically-determined increase
420 in *P2RY13* expression increases asthma risk, which is consistent with the pro-inflammatory effect
421 suggested for ADP and P2RY13 by these functional studies.

422 *P2RY14*, also known as *GPR105*, encodes a G protein-coupled receptor that is potently and

423 selectively activated by UDP-sugars, especially UDP-glucose [46]. UDP-glucose is thought to be an
424 extracellular pro-inflammatory mediator [47], constitutively released by different cell types including
425 airway epithelial cells [48]. Importantly, infection with respiratory syncytial virus (RSV) or treatment
426 with IL-13 significantly increases UDP-glucose release by airway epithelial cells [49], and this coincides
427 with increased mucus secretion [50]. Known pro-inflammatory effects of UDP-glucose acting through
428 *P2RY14* include inhibition of TLR9-dependent IFN-alpha production [51], increased chemotaxis of
429 neutrophils [52], induction of mast cell degranulation [53], production of IL-8 [28, 54] and STAT3-dependent
430 epidermal inflammation [55]. A small molecule antagonist for P2RY14 was recently developed and
431 shown to effectively block chemotaxis of freshly isolated human neutrophils [56]. Of note, plasma UDP-
432 glucose levels are elevated in mice fed a high-fat diet [57], which raises the possibility that obesity might
433 contribute to chronic P2RY14 activation and, in that way, increase asthma risk and/or severity. Studies
434 that investigate this possibility are underway.

435 When we studied the expression of both P2ry13 and P2ry14 in mice, we found that both genes
436 were highly expressed in AECs, both at baseline and after allergen challenge. High expression was also
437 observed in infiltrating eosinophils after challenge and, to a smaller extent, neutrophils and monocytes.
438 High expression in AECs suggested that receptor activation could contribute to airway inflammation
439 even in the absence of allergen stimulation. This was indeed what we observed when naïve mice were
440 challenged intra-nasally with either ADP or UDP-glucose: 72 hours after challenge, the numbers of
441 BALF eosinophils were significantly increased when compared to control mice. Interestingly,
442 eosinophil influx into the airways was preceded by a significant increase in the levels of the alarmin IL-
443 33. The effect of both receptor agonists on IL-33 release was comparable in magnitude to that observed
444 with allergens known to have a potent effect on IL-33 production, namely the fungus Alternaria
445 alternata [58]. These results demonstrate that activation of P2ry13 and P2ry14, in addition to P2y2 [58], can
446 strongly induce IL-33 release in mice. Interestingly, P2ry13 expression was observed in the nuclei of

AECs after allergen challenge. Given that IL-33 is constitutively stored in the nuclei of AECs [59], it is possible that intracellular activation of P2ry13 expressed on the nuclear membrane plays a role in allergen-induced IL-33 release.

In conclusion, our genetic findings establish an association between asthma risk and genes involved in nucleotide synthesis (*B4GALT3*, *USMG5*) and nucleotide-dependent cell activation (*P2RY13*, *P2RY14*). In mice, *in vivo* activation of P2ry13 and P2ry14 induced IL-33 release and subsequent eosinophilic airway infiltration. These observations suggest that genetic dysregulation of nucleotide signaling contributes to the risk of asthma (allergic and, potentially, also non-allergic) and other related conditions; studies that test this possibility are now warranted. Our results also show that re-analysis of published GWAS with a gene-based test that exclusively focuses on documented eQTLs has the potential to identify novel associations.

ACKNOWLEDGMENTS

We thank all study participants, including customers of 23andMe who answered surveys, as well as the employees of 23andMe, who together made this research possible.

461 **TABLES**

462

463 **Table 1.** Forty eight genes with a significant (FDR < 5%) association with asthma risk in the Ferreira et

464 al. [14] GWAS.

Gene	Position	N eQTLs	N eQTLs tested	N eQTLs with *P*<0.05	Best individual eQTL		EUGENE *P*-value	Potential novel association[a]
					SNP	*P*-value		
HLA-DQB1	6:32627244	78	26	15	rs1063355	1.8×10^{-13}	$<10^{-6}$	No
GSDMB	17:38060848	15	11	5	rs2952140	1.2×10^{-8}	$<10^{-6}$	No
LIMS1	2:109150857	15	14	4	rs1063355	1.8×10^{-13}	$<10^{-6}$	No*
TLR1	4:38792298	9	6	3	rs12233670	1.4×10^{-11}	$<10^{-6}$	No
ORMDL3	17:38077294	19	14	5	rs2952140	1.2×10^{-8}	$<10^{-6}$	No
IKZF3	17:37921198	9	7	3	rs7207600	4.5×10^{-7}	$<10^{-6}$	No
IL18RAP	2:103035149	18	16	6	rs13018263	5.0×10^{-6}	$<10^{-6}$	No
CLEC16A	16:11038345	4	4	2	rs35441874	2.9×10^{-8}	$<10^{-6}$	No
ZPBP2	17:38024417	2	2	1	rs9916765	1.9×10^{-9}	$<10^{-6}$	No
GRB7	17:37894180	2	1	1	rs14050	1.4×10^{-7}	$<10^{-6}$	No
TINF2	14:24708849	4	4	2	rs3135006	1.7×10^{-6}	$<10^{-6}$	No*
TAP2	6:32789610	48	36	11	rs2858312	1.9×10^{-5}	2.0×10^{-6}	No
TAP1	6:32812986	13	12	4	rs6928482	2.0×10^{-8}	2.0×10^{-6}	No
HSPA1B	6:31795512	13	13	6	rs13215091	4.7×10^{-4}	2.0×10^{-6}	No
TSLP	5:110405760	6	6	4	rs17132582	3.2×10^{-4}	2.0×10^{-6}	No
DYNC1H1	14:102430865	6	5	2	rs4906262	1.1×10^{-5}	2.0×10^{-6}	Yes
HLA-DRB1	6:32546546	97	46	18	rs3806156	1.4×10^{-4}	4.0×10^{-6}	No
IL18R1	2:102927989	11	11	5	rs6751967	3.2×10^{-6}	4.0×10^{-6}	No
SOCS1	16:11348262	7	6	5	rs7184491	3.5×10^{-6}	4.0×10^{-6}	No
CISD3	17:36886488	4	4	2	rs2941503	1.6×10^{-7}	5.0×10^{-6}	No
PGAP3	17:37827375	2	2	1	rs903502	1.5×10^{-6}	6.0×10^{-6}	No
IL1RL2	2:102803433	2	2	1	rs9646944	6.7×10^{-7}	9.0×10^{-6}	No
CLK3	15:74890841	2	2	1	rs9268853	1.8×10^{-6}	9.0×10^{-6}	No*
SMAD3	15:67356101	7	7	2	rs17293632	2.0×10^{-7}	1.1×10^{-5}	No
SAFB	19:5623046	2	2	1	rs9268853	1.8×10^{-6}	1.1×10^{-5}	No*
P2RY13	3:151044100	8	8	5	rs9877416	1.2×10^{-4}	1.2×10^{-5}	Yes
AOAH	7:36552456	26	21	3	rs9268853	1.8×10^{-6}	1.3×10^{-5}	No*
SLC44A4	6:31830969	2	2	1	rs9275141	1.1×10^{-6}	1.3×10^{-5}	No
STARD3	17:37793318	7	3	1	rs2941503	1.6×10^{-7}	1.5×10^{-5}	No
LTA	6:31539831	16	13	5	rs2442752	1.5×10^{-5}	1.6×10^{-5}	No
MED24	17:38175350	5	5	2	rs7502514	4.8×10^{-5}	1.7×10^{-5}	No
HIBADH	7:27565061	14	12	6	rs6951856	9.6×10^{-5}	2.9×10^{-5}	Yes
P2RY12	3:151055168	6	6	4	rs17282940	7.0×10^{-5}	3.0×10^{-5}	Yes
NR1D1	17:38249040	5	4	2	rs12150298	2.8×10^{-6}	3.0×10^{-5}	No
ZNF707	8:144766622	9	8	3	rs17609240	1.5×10^{-6}	4.2×10^{-5}	No*
TOP2A	17:38544768	1	1	1	rs2102928	4.1×10^{-5}	4.9×10^{-5}	No
HLA-DRB6	6:32520490	60	13	6	rs522254	6.3×10^{-4}	5.4×10^{-5}	No
REEP3	10:65281123	6	4	2	rs7898489	9.1×10^{-6}	6.1×10^{-5}	Yes
PTCSC3	14:36605314	2	2	2	rs7148603	1.8×10^{-4}	6.5×10^{-5}	Yes
P2RY14	3:150929905	13	12	5	rs10513393	1.1×10^{-4}	7.2×10^{-5}	Yes
HLA-DQA1	6:32595956	79	29	9	rs504594	1.7×10^{-5}	8.1×10^{-5}	No

ACO2	22:41865129	3	3	2	rs960596	1.3×10^{-4}	9.4×10^{-5}	Yes
HCP5	6:31368479	23	22	4	rs2071595	6.7×10^{-6}	9.6×10^{-5}	No
NEU1	6:31825436	8	8	5	rs9267901	9.1×10^{-4}	9.8×10^{-5}	No
MICB	6:31462658	41	30	9	rs9268764	3.3×10^{-5}	1.2×10^{-4}	No
B4GALT3	1:161141100	10	9	3	rs1668873	1.5×10^{-3}	1.2×10^{-4}	Yes
USMG5	10:105148798	16	14	4	rs1163073	4.9×10^{-4}	1.5×10^{-4}	Yes
F12	5:176829141	3	3	2	rs4976765	1.7×10^{-3}	1.5×10^{-4}	Yes

[a] Potential novel genetic associations with asthma (highlighted in grey) are those for which the gene-based association was not driven by eQTLs located in known allergy risk loci. Known allergy loci are defined as those that contain a variant reported to be associated with asthma or other allergic diseases with $P<5x10^{-8}$ in published GWAS. * Genes not located in an asthma risk locus but for which the gene-based association was driven by *trans*-eQTLs in LD with allergy risk variants (see Table E7 for more details).

FIGURE LEGENDS

Figure 1. Outline of analytical procedure.

Figure 2. Summary of association results obtained for 17,190 genes by applying the proposed gene-based test of association to a published asthma GWAS[14]. The red horizontal line shows the *P*-value threshold corresponding to an empirical FDR of 5% (P=1.9x10^{-4}). Forty eight genes exceeded this threshold, including (1) 31 genes located in established risk loci for allergic disease (denoted by '+'; gene name shown in black font); (2) six genes located in new risk loci but with a gene-based association that was driven by *trans*-eQTLs located in the MHC or near *ORMDL3* (denoted by 'Δ'); and (3) 11 genes with a gene-based association that was not driven by eQTLs located in established allergy risk loci (denoted by '○'), including four (green font) for which the association replicated in an independent GWAS[15]. The *y*-axis represents the $-\log_{10}$ of the simulation-derived gene-based *P*-value, which accounts for the residual LD between eQTLs of a given gene. The *P*-value was based on up to 1 million simulations, and so it could not exceed a P=10^{-6} (dashed grey line).

Figure 3. **Expression levels of *P2ry13* and *P2ry14* in lung of C57Bl/6 mice sensitized and then challenged with a saline solution or a house dust mite extract. (A)** Overall gene expression in lung. Expression levels were normalized to *Hprt* and are expressed as fold-change over saline challenge group. Results show mean +/- SD in each group. * Wilcoxon rank sum test $P < 0.005$ when comparing HDM and saline groups. (**B, D and E**) Expression of P2ry13 and P2ry14 based on flow cytometry analysis in lung epithelial cells, or eosinophils and neutrophils collected in BALF after saline or HDM challenge. **(C)** Expression of P2ry13 and P2ry14 in lung sections of mice challenged with saline or HDM.

Figure 4. *In vivo* exposure to P2ry13 and P2ry14 receptor agonists in naïve C57Bl/6 mice. Mice were challenged with either vehicle, one of three allergens (HDM, CRE, Alt) or one of three nucleotides (ATP, UPD-glucose, ADP), and euthanized 2 and 72 hours after challenge. **(A)** IL-33 expression in bronchoalveolar lavage fluid (BALF) collected 2 hours post challenge. Total number of eosinophils **(B)** and lymphocytes **(C)** recruited to the BALF at 72 hours post challenge, based on flow cytometry analysis. Veh: vehicle. HDM: house dust mite. CRE: cockroach. Alt: alternaria.

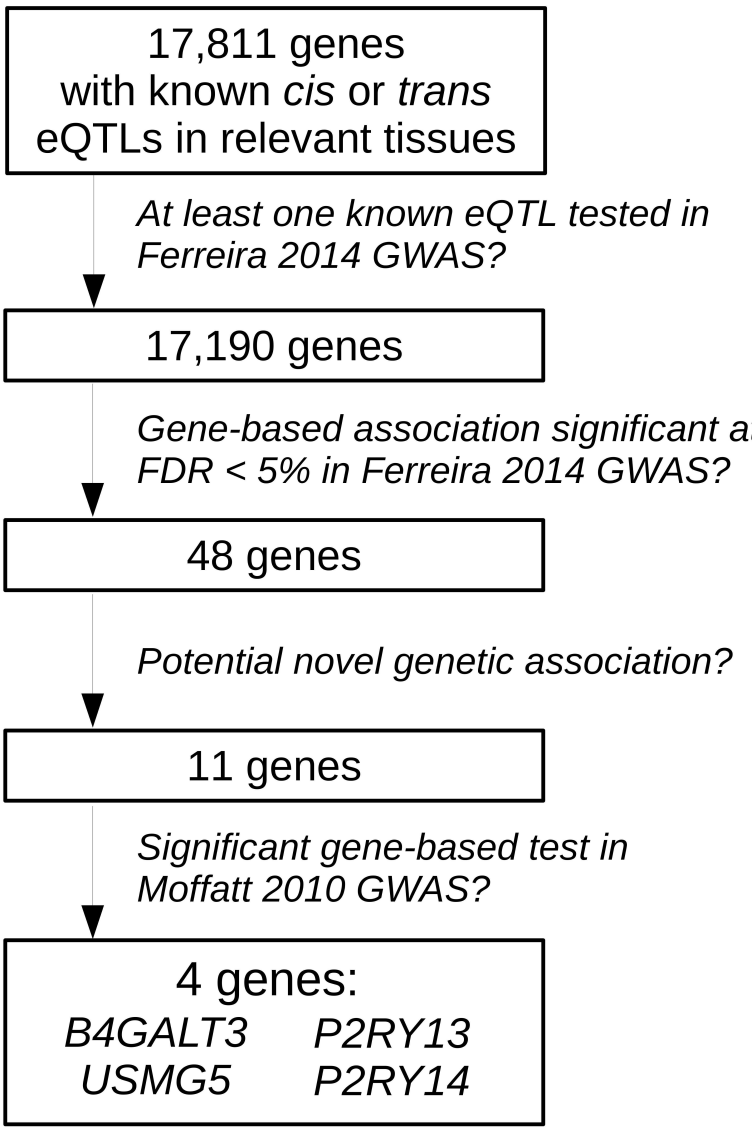
REFERENCES

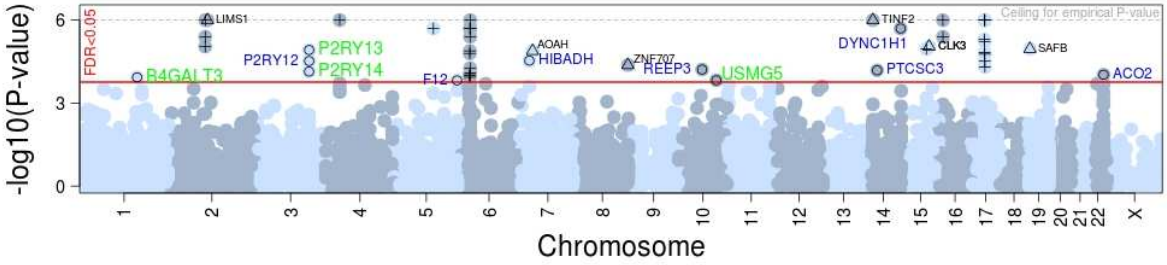
1. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souef P, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet 2011; 378:1006-14.
2. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 2007; 448:470-3.
3. Vicente CT, Edwards SL, Hillman KM, Kaufmann S, Mitchell H, Bain L, et al. Long-Range Modulation of PAG1 Expression by 8q21 Allergy Risk Variants. Am J Hum Genet 2015.
4. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014; 506:376-81.
5. 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.
6. 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.
7. Bonnelykke K, Sleiman P, Nielsen K, Kreiner-Moller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet 2014; 46:51-5.
8. Galesloot TE, van Steen K, Kiemeney LA, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. PLoS One 2014; 9:e95923.
9. 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.
10. Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. Am J Hum Genet 2011; 88:283-93.
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; 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. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016; 48:245-52.
14. Ferreira MA, Matheson MC, Tang CS, Granell R, Ang W, Hui J, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol 2014; 133:1564-71.
15. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010; 363:1211-21.
16. Ullah MA, Revez JA, Loh Z, Simpson J, Zhang V, Bain L, et al. Allergen-induced IL-6 trans-signaling activates gammadelta T cells to promote type 2 and type 17 airway inflammation. J Allergy Clin Immunol 2015; 136:1065-73.
17. Davidson S, Kaiko G, Loh Z, Lalwani A, Zhang V, Spann K, et al. Plasmacytoid dendritic cells promote host defense against acute pneumovirus infection via the TLR7-MyD88-dependent signaling pathway. J Immunol 2011; 186:5938-48.
18. 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.
19. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond--the

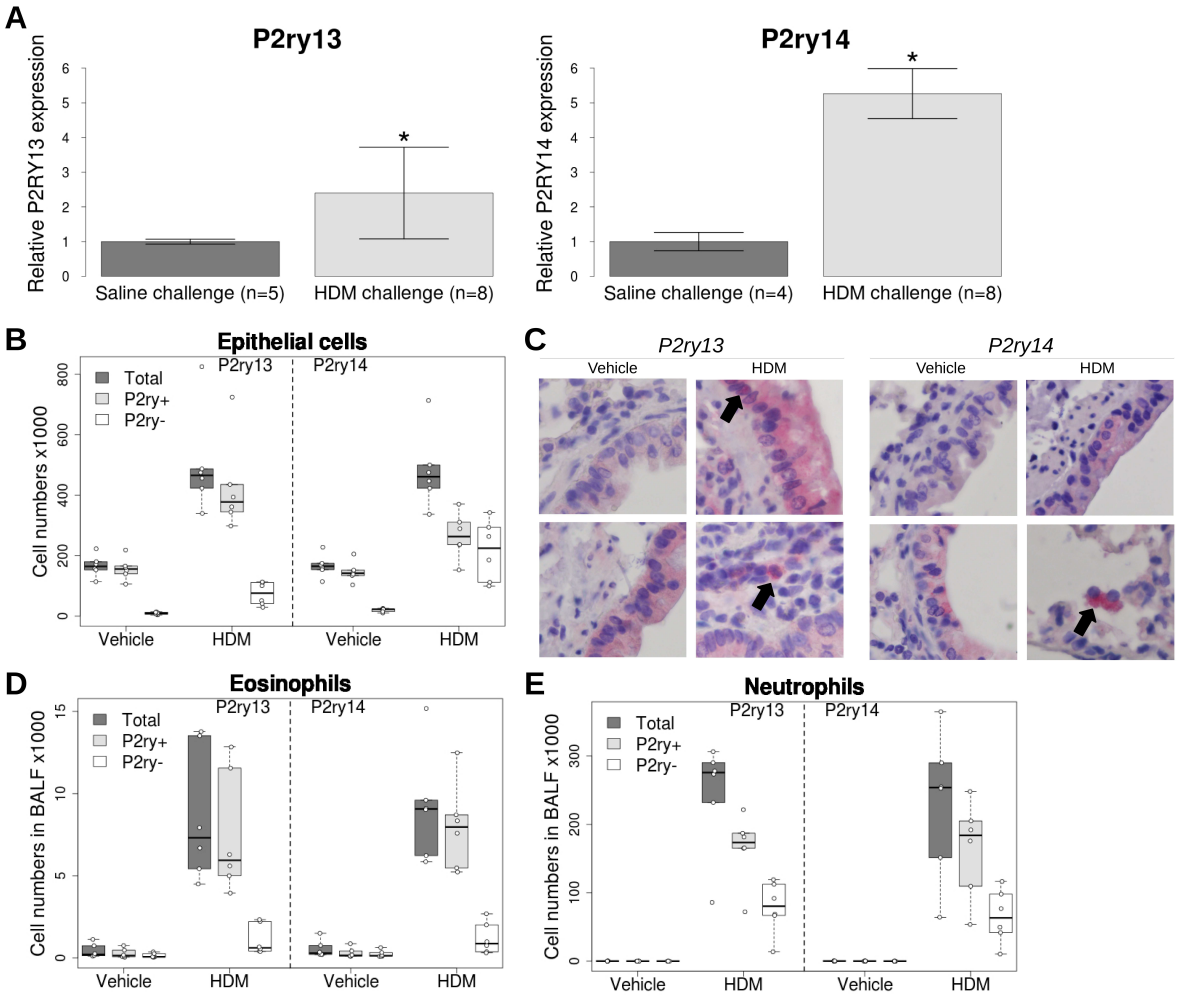
transcriptome of human monocytes and disease susceptibility. PLoS One 2010; 5:e10693.
20. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013; 45:1238-43.
21. Kim S, Becker J, Bechheim M, Kaiser V, Noursadeghi M, Fricker N, et al. Characterizing the genetic basis of innate immune response in TLR4-activated human monocytes. Nat Commun 2014; 5:5236.
22. 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.
23. Berndt SI, Gustafsson S, Magi R, Ganna A, Wheeler E, Feitosa MF, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet 2013; 45:501-12.
24. Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet 2009; 41:1182-90.
25. Kuga W, Tsuchihara K, Ogura T, Kanehara S, Saito M, Suzuki A, et al. Nuclear localization of SNARK; its impact on gene expression. Biochem Biophys Res Commun 2008; 377:1062-6.
26. Lo NW, Shaper JH, Pevsner J, Shaper NL. The expanding beta 4-galactosyltransferase gene family: messages from the databanks. Glycobiology 1998; 8:517-26.
27. Gao ZG, Ding Y, Jacobson KA. P2Y(13) receptor is responsible for ADP-mediated degranulation in RBL-2H3 rat mast cells. Pharmacol Res 2010; 62:500-5.
28. Muller T, Bayer H, Myrtek D, Ferrari D, Sorichter S, Ziegenhagen MW, et al. The P2Y14 receptor of airway epithelial cells: coupling to intracellular Ca2+ and IL-8 secretion. Am J Respir Cell Mol Biol 2005; 33:601-9.
29. Seminario MC, Guo J, Bochner BS, Beck LA, Georas SN. Human eosinophils constitutively express nuclear factor of activated T cells p and c. J Allergy Clin Immunol 2001; 107:143-52.
30. Liao WC, Liu CH, Chen CH, Hsu WM, Liao YY, Chang HM, et al. beta-1,4-Galactosyltransferase III suppresses extravillous trophoblast invasion through modifying beta1-integrin glycosylation. Placenta 2015; 36:357-64.
31. Suzuki K, Okuno T, Yamamoto M, Pasterkamp RJ, Takegahara N, Takamatsu H, et al. Semaphorin 7A initiates T-cell-mediated inflammatory responses through alpha1beta1 integrin. Nature 2007; 446:680-4.
32. Lazarowski ER. Quantification of extracellular UDP-galactose. Anal Biochem 2010; 396:23-9.
33. Meyer B, Wittig I, Trifilieff E, Karas M, Schagger H. Identification of two proteins associated with mammalian ATP synthase. Mol Cell Proteomics 2007; 6:1690-9.
34. Ohsakaya S, Fujikawa M, Hisabori T, Yoshida M. Knockdown of DAPIT (diabetes-associated protein in insulin-sensitive tissue) results in loss of ATP synthase in mitochondria. J Biol Chem 2011; 286:20292-6.
35. Ledderose C, Bao Y, Lidicky M, Zipperle J, Li L, Strasser K, et al. Mitochondria are gate-keepers of T cell function by producing the ATP that drives purinergic signaling. J Biol Chem 2014; 289:25936-45.
36. Yip L, Woehrle T, Corriden R, Hirsh M, Chen Y, Inoue Y, et al. Autocrine regulation of T-cell activation by ATP release and P2X7 receptors. FASEB J 2009; 23:1685-93.
37. Vantourout P, Radojkovic C, Lichtenstein L, Pons V, Champagne E, Martinez LO. Ecto-F(1)-

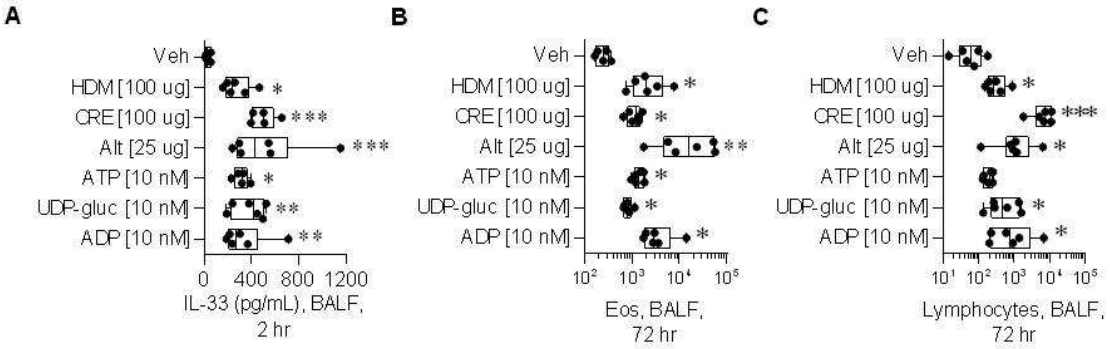
592 ATPase: a moonlighting protein complex and an unexpected apoA-I receptor. World J
593 Gastroenterol 2010; 16:5925-35.
594 38.Jacquet S, Malaval C, Martinez LO, Sak K, Rolland C, Perez C, et al. The nucleotide receptor
595 P2Y13 is a key regulator of hepatic high-density lipoprotein (HDL) endocytosis. Cell Mol Life
596 Sci 2005; 62:2508-15.
597 39.Scotet E, Martinez LO, Grant E, Barbaras R, Jeno P, Guiraud M, et al. Tumor recognition following
598 Vgamma9Vdelta2 T cell receptor interactions with a surface F1-ATPase-related structure and
599 apolipoprotein A-I. Immunity 2005; 22:71-80.
600 40.Communi D, Gonzalez NS, Detheux M, Brezillon S, Lannoy V, Parmentier M, et al. Identification
601 of a novel human ADP receptor coupled to G(i). J Biol Chem 2001; 276:41479-85.
602 41.Zhang FL, Luo L, Gustafson E, Palmer K, Qiao X, Fan X, et al. P2Y(13): identification and
603 characterization of a novel Galphai-coupled ADP receptor from human and mouse. J Pharmacol
604 Exp Ther 2002; 301:705-13.
605 42.Inoue K, Hosoi J, Denda M. Extracellular ATP has stimulatory effects on the expression and release
606 of IL-6 via purinergic receptors in normal human epidermal keratinocytes. J Invest Dermatol
607 2007; 127:362-71.
608 43.Kreda SM, Seminario-Vidal L, van Heusden CA, O'Neal W, Jones L, Boucher RC, et al. Receptor-
609 promoted exocytosis of airway epithelial mucin granules containing a spectrum of adenine
610 nucleotides. J Physiol 2010; 588:2255-67.
611 44.la Sala A, Ferrari D, Corinti S, Cavani A, Di Virgilio F, Girolomoni G. Extracellular ATP induces a
612 distorted maturation of dendritic cells and inhibits their capacity to initiate Th1 responses. J
613 Immunol 2001; 166:1611-7.
614 45.Idzko M, Dichmann S, Ferrari D, Di Virgilio F, la Sala A, Girolomoni G, et al. Nucleotides induce
615 chemotaxis and actin polymerization in immature but not mature human dendritic cells via
616 activation of pertussis toxin-sensitive P2y receptors. Blood 2002; 100:925-32.
617 46.Chambers JK, Macdonald LE, Sarau HM, Ames RS, Freeman K, Foley JJ, et al. A G protein-
618 coupled receptor for UDP-glucose. J Biol Chem 2000; 275:10767-71.
619 47.Harden TK, Sesma JI, Fricks IP, Lazarowski ER. Signalling and pharmacological properties of the
620 P2Y receptor. Acta Physiol (Oxf) 2010; 199:149-60.
621 48.Lazarowski ER, Shea DA, Boucher RC, Harden TK. Release of cellular UDP-glucose as a potential
622 extracellular signaling molecule. Mol Pharmacol 2003; 63:1190-7.
623 49.Okada SF, Zhang L, Kreda SM, Abdullah LH, Davis CW, Pickles RJ, et al. Coupled nucleotide and
624 mucin hypersecretion from goblet-cell metaplastic human airway epithelium. Am J Respir Cell
625 Mol Biol 2011; 45:253-60.
626 50.Kreda SM, Okada SF, van Heusden CA, O'Neal W, Gabriel S, Abdullah L, et al. Coordinated release
627 of nucleotides and mucin from human airway epithelial Calu-3 cells. J Physiol 2007; 584:245-
628 59.
629 51.Shin A, Toy T, Rothenfusser S, Robson N, Vorac J, Dauer M, et al. P2Y receptor signaling regulates
630 phenotype and IFN-alpha secretion of human plasmacytoid dendritic cells. Blood 2008;
631 111:3062-9.
632 52.Sesma JI, Kreda SM, Steinckwich-Besancon N, Dang H, Garcia-Mata R, Harden TK, et al. The
633 UDP-sugar-sensing P2Y(14) receptor promotes Rho-mediated signaling and chemotaxis in
634 human neutrophils. Am J Physiol Cell Physiol 2012; 303:C490-8.
635 53.Gao ZG, Ding Y, Jacobson KA. UDP-glucose acting at P2Y14 receptors is a mediator of mast cell
636 degranulation. Biochem Pharmacol 2010; 79:873-9.

637 54. Arase T, Uchida H, Kajitani T, Ono M, Tamaki K, Oda H, et al. The UDP-glucose receptor P2RY14
638 triggers innate mucosal immunity in the female reproductive tract by inducing IL-8. J Immunol
639 2009; 182:7074-84.
640 55. Jokela TA, Karna R, Makkonen KM, Laitinen JT, Tammi RH, Tammi MI. Extracellular UDP-
641 glucose activates P2Y14 Receptor and Induces Signal Transducer and Activator of
642 Transcription 3 (STAT3) Tyr705 phosphorylation and binding to hyaluronan synthase 2 (HAS2)
643 promoter, stimulating hyaluronan synthesis of keratinocytes. J Biol Chem 2014; 289:18569-81.
644 56. Barrett MO, Sesma JI, Ball CB, Jayasekara PS, Jacobson KA, Lazarowski ER, et al. A selective
645 high-affinity antagonist of the P2Y14 receptor inhibits UDP-glucose-stimulated chemotaxis of
646 human neutrophils. Mol Pharmacol 2013; 84:41-9.
647 57. Xu J, Morinaga H, Oh D, Li P, Chen A, Talukdar S, et al. GPR105 ablation prevents inflammation
648 and improves insulin sensitivity in mice with diet-induced obesity. J Immunol 2012; 189:1992-
649 9.
650 58. Kouzaki H, Iijima K, Kobayashi T, O'Grady SM, Kita H. The danger signal, extracellular ATP, is a
651 sensor for an airborne allergen and triggers IL-33 release and innate Th2-type responses. J
652 Immunol 2011; 186:4375-87.
653 59. Moussion C, Ortega N, Girard JP. The IL-1-like cytokine IL-33 is constitutively expressed in the
654 nucleus of endothelial cells and epithelial cells in vivo: a novel 'alarmin'? PLoS One 2008;
655 3:e3331.
656






A
P2ry13
Relative P2RY13 expression
Saline challenge (n=5)
HDM challenge (n=8)
P2ry14
Relative P2RY14 expression
Saline challenge (n=4)
HDM challenge (n=8)
B
Epithelial cells
Cell numbers x1000
Total
P2ry+
P2ry-
P2ry13
P2ry14
Vehicle
HDM
C
P2ry13
P2ry14
Vehicle
HDM
D
Eosinophils
Cell numbers in BALF x1000
E
Neutrophils



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

Online repository

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,[e] Jennie Hui, PhD,[f-i] Joyce Y. Tung, PhD,[j] Svetlana Baltic, PhD,[k] Peter Le Souëf, FRACP,[l] AAGC collaborators[m], Grant W. Montgomery, PhD,[a] Nicholas G. Martin, PhD,[a] Colin F. Robertson, FRACP,[n] Alan James, FRACP,[i,o,p] Philip J. Thompson, FRACP,[k,o] Dorret I. Boomsma, PhD,[c] John L. Hopper, PhD,[e] David A. Hinds, PhD,[j] Rhiannon B. Werder, BSc,[q*] and Simon Phipps, PhD,[q*]

Supplementary Methods

EUGENE approach

The proposed gene-based approach includes four steps, which for a given gene briefly consists of: (1) identifying variants that influence gene expression; (2) extracting association results for these regulatory variants from a disease or trait GWAS of interest and then calculate a gene-based statistic *Q* – this represents the aggregate evidence for association in that GWAS across all regulatory variants of that gene; and perform simulations using individual-level genetic data to estimate (3) the statistical significance of *Q* and (4) false discovery rate (FDR) thresholds to empirically account for multiple testing. These steps are described in more detail below; the software and input files required to run EUGENE are freely available at https://genepi.qimr.edu.au/staff/manuelF.

Step 1: generate a list of independent single nucleotide polymorphisms (SNPs) that are associated with variation in gene expression levels (ie. expression quantitative trait loci, eQTLs). First, a database of eQTLs associated with gene expression levels in *cis* (located < 1 Mb from gene boundaries) or *trans* (> 1 Mb away or in a different chromosome) was created from 16 published transcriptome GWAS that analysed 12 tissues or cell types relevant for asthma (**Table E2**). Other biologically relevant marks of gene regulation (eg. CpG methylation) will be made available as studies with greater power to identify such effects are published. Second, for each gene, the eQTLs in this database were reduced to a sub-set with linkage disequilibrium (LD) r^2<0.1, using the clump procedure implemented in PLINK [1]. We refer to these as "independent eQTLs" for a given gene; more stringent LD thresholds (eg. low |D'|) or conditional association analyses could be used to generate a more conservative set of independent eQTLs. The average number of independent eQTLs per gene was 5.1 (median=3, maximum=97, IQR=5; **Figure E1**). Third, we used data from the 1000 Genomes project [2] to identify all known proxies (r^2>0.8) for each independent eQTL. This is important because if a specific eQTL is not tested in a given GWAS, a proxy SNP might still be available and could instead be

used for analysis. A file containing independent eQTLs for each gene, and their respective proxy SNPs, is required to run EUGENE and is available at https://genepi.qimr.edu.au/staff/manuelF. Files containing eQTLs in tissues relevant for other diseases or traits are also available but were not analysed in this study.

Step 2: calculate a gene-based association statistic Q *for each gene based on results from individual SNPs in a GWAS of interest.* The 1-df disease association chi-square statistic *q* for each of *k* independent eQTLs of gene *i* is extracted from the trait or disease GWAS of interest. If a specific eQTL was not tested, the most correlated proxy SNP available (with r^2>0.8) is used. Then, for each gene *i*, the overall statistical evidence for association between all *k* independent eQTLs tested for that gene and the trait or disease of interest is simply calculated as $Q_i = \sum_{j=1}^{k} q_{ij}$, that is, the sum of the individual chi-square statistics across all independent eQTLs tested.

Step 3: perform simulations using individual-level genetic data to estimate the statistical significance of Q_i. When eQTLs of the same gene are in linkage equilibrium (eg. $r^2 \sim 0$), a measure of statistical significance of *Qi* could be obtained from a chi-square distribution with *k* degrees of freedom. In this scenario, this asymptotic *P*-value is not inflated under the null hypothesis of no association (not shown). However, as the LD between eQTLs increases, the asymptotic *P*-value becomes inflated under the null: for example, using an r^2 threshold of 0.1 to define independent eQTLs, on average (across 1,000 simulated GWAS) 5.4% of genes had a significant asymptotic *P*-value at *P*<0.05, a 1.08-fold increase over the 5.0% nominal expectation. This is because the assumption of statistical independence between eQTLs of a gene is not strictly achieved with that r^2 threshold. For this reason, we do not calculate an asymptotic *P*-value for Q_i, but instead estimate an empirical *P*-value that accounts for the residual LD between eQTLs. To estimate the empirical *P*-value for Q_i, we analyze the association between gene *i* and a dummy trait with a normal distribution in 379 unrelated individuals of European descent with genotype data available through the 1000 Genomes Project [2].

Other GWAS datasets with available individual level genetic data can be used in this step, including those of non-European ancestry. Using a GWAS dataset with a larger sample size (>4,000 individuals) did not influence the performance of this step (not shown), and so we used data from the 1000 Genomes Project given its availability to other researchers and decreased computation time. Briefly, in this analysis we (1) simulate a normally-distributed phenotype for the genotyped individuals under the null hypothesis of no association with any eQTLs of gene *i*; (2) test this phenotype for association with each eQTL – any significant associations in this analysis are due to chance only; and (3) calculate a gene-based association statistic *Q* as described above (Q_{i_null}). This procedure is repeated a large number of times (eg. 1 million simulations). The empirical *P*-value for Q_i is then calculated as the proportion of simulations for which $Q_{i\_null} \geq Q_i$.

Step 4: perform simulations using individual-level genetic data to estimate FDR thresholds to account for multiple testing. Typically, EUGENE will be used to test the association between a trait and many genes, and so it is important to address the impact of multiple testing on false positive findings. To achieve this, we adopt the false-discovery rate (FDR) quantity advocated by Storey and Tibshirani [3]. For a given threshold *t* (eg. 0.05), FDR is approximated by the expected number of genes with a *P*-value $\leq t$ when the null is true ($E[F_t]$), divided by the expected total number of genes with a *P*-value $\leq t$ ($E[S_t]$). A simple estimate of $E[S_t]$ is the observed S_t, that is, the number of genes with a *P*-value $\leq t$. To estimate $E[F_t]$, we use simulations generated under the null using individual-level genetic data as described above. Specifically, we (1) simulate a normally-distributed phenotype for the genotyped individuals under the null hypothesis of no association with any eQTLs of all genes tested; (2) test this phenotype for association with all eQTLs analysed across all genes; and (3) calculate a gene-based association statistic *Q* and its empirical *P*-value for each gene tested as described above. This procedure is repeated 100 times. For each of these 100 simulations, we count the number of genes significant at different *P*-value thresholds *t* (10^{-6} down to 0.1); for a given *t*, the average count across 100 simulations

is taken as an estimate of $E[F_t]$. For each *P*-value threshold *t* considered, we estimate FDR_t as $E[F_t] / E[S_t]$; based on these FDR_t, we determine the minimum *P*-value threshold *t* that would result in an FDR of 0.05. At this *P*-value threshold *t*, 5% of genes called significant are estimated to be false-positives. As a concrete example, in the discovery GWAS described in the main text, the *P*-value threshold that resulted in an FDR of 0.05 was 1.9×10^{-4}. At this threshold, 48 genes were associated with asthma risk, of which about 3 (48 x 0.05) are expected to be false-positives due to multiple testing.

Assessment of the type-1 error rate of EUGENE

To assess whether the proposed gene-based test had an appropriate type-I error rate, we (1) simulated a dummy normally-distributed phenotype for 379 genotyped individuals, as described above, and tested its association with all available SNPs – that is, we simulated results from a GWAS under the null hypothesis of no association; (2) applied EUGENE to the resulting summary statistics and retained the gene-based *P*-value for each of 17,190 genes; and (3) repeated steps (1) and (2) to simulate and analyse results for 1,000 GWAS generated under the null. The type-I error rate for a given nominal α was taken as the mean (across the 1,000 null GWAS) proportion of genes with a significant association at that α level.

Functional studies in the mouse

Experiment set 1: Expression of P2ry13 and P2ry14 in a mouse model of acute experimental asthma. We used an established mouse model of acute allergic asthma [4] to identify the cell types that express P2ry13 and P2ry14 in the context of allergen-induced airway inflammation. Experiments were performed in accordance with the Animal Care and Ethics Committees of the University of Queensland (Brisbane, Australia). Briefly, two groups of eight- to twelve-week old wild-type C57Bl/6 mice were lightly anesthetized with isoflurane and sensitized intranasally with either saline solution (group 1) or

100 μg of HDM extract (Dermatophagoides pteronyssinus; Greer Laboratories, Lenoir, NC, USA; group 2) on day 0. Subsequently, mice were challenged with either saline (group 1) or 5 μg of HDM (group 2) at day 14, 15, 16 and 17 and sacrificed 3 hours later. In this model, mice challenged with HDM develop all the hallmark features of asthma, including airway hyperresponsiveness, mucous cell hyperplasia and granulocytic airway inflammation [4].

To measure overall gene expression in lung, total RNA was isolated from the left lung with TriReagent solution (Ambion) and phenol-chloroform extraction. DNAse digestion was performed with Turbo DNAse (Ambion), according to the manufacturer's instructions. Reverse transcription was performed using M-MLV reverse transcriptase and random primers (Invitrogen). Quantitative real-time PCR was performed with SYBR Green (Life Technologies) and the primers described in **Table E13**. Expression values were normalized to *Hprt* and expressed as fold change over saline mice.

To identify individual cell types in the lung expressing P2ry13 and P2ry14, a bronchoalveolar lavage was performed by flushing the lungs with 600 ul of ice-cold PBS. The recovered fluid was centrifuged (1600 rpm for 5 minutes) and the bronchoalveolar lavage fluid (BALF) stored at -80oC until analysis. Lung lobes were dissected and single cell suspensions prepared by mechanical digestion through a cell strainer as described [4]. Following red blood cell lysis with Gey's lysis buffer, cells were counted then incubated with Fc block for 30 minutes at 4°C. Cells were then stained with anti-P2ry13 (Acris Antibodies) or anti-P2ry14 (Acris Antibodies), followed by appropriate fluorescently-labelled secondary. Then cells were stained with the following fluorescently labeled antibodies: FITC conjugated Ly6G (clone 1A8), PerCP-Cy5.5 conjugated CD11b (clone M1/70), AF647 conjugated Siglec F (clone E50-2440) (all BD Biosciences), BV570 conjugated Ly6C (clone HK1.4), BV785 conjugated CD11c (clone N418), AF488 conjugated Epcam (CD326) (clone G8.8), BV421 conjugated CD45 (clone 30-F11) (all Biolegend), PE conjugated B220 (clone RA3-6B2) and CD3ε (clone145-2C11), APC-eFluor 780 or PE conjugated MHCII (clone M5/114.15.2) (all eBioscience). Cells were

enumerated using a BD LSR Fortessa cytometer (BD Biosciences, San Jose, CA, USA) and the data analyzed with FACSDiva v8 (BD Biosciences) and FlowJo v8.8 (Treestar).

To assess expression in airway epithelial cells, paraffin-embedded lung sections were prepared as previously described [5]. Lung sections were pretreated with 10% normal goat serum for 30 min. Sections were probed with anti-P2ry13 (Acris Antibodies) or anti-P2ry14 (Acris Antibodies) overnight at 4°C. Following incubation with appropriate secondary antibodies, immunoreactivity was developed with Fast Red (Sigma-Aldrich) and counterstained with Mayer's hematoxylin. Photomicrographs were taken at 400x and 1000x magnification using an Olympus BX-51 microscope with an Olympus DP-72 camera at room temperature and acquired using Olympus Image Analysis Software.

Experiment set 2: Effect of in vivo exposure to P2ry13 and P2ry14 agonists on airway inflammatory profile in mice. Given the high expression of both receptors on airway epithelial cells in naïve mice, we hypothesized that receptor activation could influence the release of alarmins, such as IL-33, and contribute to airway inflammation. To test this possibility, 6 naïve mice per group were inoculated via intra nasal (i.n.) route with saline, 10 nM 2-methyl-ADP (P2ry13 agonist; R&D), 10 nM UDP-glucose (P2ry14 agonist; Abcam) or 10 nM ATP (agonist for all P2ry receptors, except P2ry6 and P2ry14; Sigma), all in 50 uL (ie. total dose of 0.5 pmol for all nucleotides). For comparison, three additional groups of mice were inoculated with 100 ug of HDM (source as above), 100 ug of cockroach extract (*Blattella germanica,* Greer Laboratories) or 25 ug of alternaria alternata extract (*Alternaria tenuis,* Greer Laboratories). Two hours post-challenge, BALF was collected as described above and IL-33 levels measured by ELISA (R&D Systems). Seventy-two hours post-challenge, BALF was again collected to obtain immune cell counts and stained for flow cytometry as described above.

E Tables

Table E1. Type-I error rate of EUGENE.

Nominal type-I error rate (α)	Observed type-I error rate
0.1000	0.09818
0.0500	0.04873
0.0100	0.00950
0.0050	0.00468
0.0010	0.00090
0.0005	0.00043

The observed type-I error rate corresponds to the average proportion of genes (out of 17,190 tested) with an association P-value ≤ α, when analyzing 1,000 GWAS simulated under the null hypothesis of no association.

Table E2. Number of genes with significant cis (+/- 1 Mb) eQTLs in published GWAS of gene expression that analyzed tissues relevant to asthma.

N Genes	Reference	Tissue	Experiment/eQTL type
11047	[6]	Whole-blood	
10142	[7]	Whole-blood[a]	
9271	[8]	Lung	
8752	[9]	Fibroblasts	
7461	[10]	Monocytes	
7294	[11]	LCLs	
7225	[9]	Lung	
6823	[10]	B-cells	
6783	[9]	Whole-blood	
5184	[12]	PBMCs	
3175	[13]	Neutrophils	
2954	[9]	LCLs	
2754	[9]	Spleen	
2098	[14]	Fibroblasts	
2097	[14]	LCLs	
1992	[14]	T-cells	
1732	[15]	LCLs	
1133	[16]	LCLs	
1074	[17]	Monocytes	Baseline
992	[15]	Skin	
916	[17]	Monocytes	LPS
889	[6]	Whole-blood	Splice eQTLs
831	[18]	Neutrophils	
528	[6]	Whole-blood	ASE eQTLs
508	[19]	Skin	Normal
484	[20]	Small airways	
404	[19]	Skin	Uninvolved
381	[19]	Skin	Lesional
313	[21]	LCLs	
81	[17]	Monocytes	Differential

[a] Including all *cis* SNP-gene associations significant at FDR of 0.5 (listed in file 2012-12-21-CisAssociationsProbeLevelFDR0.5.txt released with the original publication).
LCLs: lymphoblastoid cell lines. PBMCs: peripheral blood mononuclear cells.

Table E3. Thirty one genes associated with asthma at an empirical FDR of 0.05 and located within 1 Mb (or on the MHC region) of established risk variants for allergic disease.

Gene	N eQTLs	*cis*-eQTLs				*trans*-eQTLs				EUGENE *P*-value
		N Tested	N with *P*<0.05	Best individual eQTL		N Tested	N with *P*<0.05	Best individual eQTL		
				SNP	P-value			SNP	P-value	
Chromosome 2q12										
IL1RL2	2	2	1	rs9646944	6.7E-07	0	0	NA	NA	4.0E-6
IL18R1	11	11	5	rs6751967	3.2E-06	0	0	NA	NA	7.0E-6
IL18RAP	18	16	6	rs13018263	5.0E-06	0	0	NA	NA	<1E-6
Chromosome 4p14										
TLR1	9	6	3	rs12233670	1.4E-11	0	0	NA	NA	<1E-6
Chromosome 5q22										
TSLP	6	6	4	rs17132582	3.2E-04	0	0	NA	NA	7.0E-6
Chromosome 6p21										
HCP5	23	20	4	rs2071595	6.7E-06	2	0	rs891140	0.5289	9.1E-5
MICB	41	28	8	rs9268764	3.3E-05	2	1	rs647316	0.0249	1.1E-4
LTA	16	13	5	rs2442752	1.5E-05	0	0	NA	NA	2.9E-5
HSPA1B	13	13	6	rs13215091	4.7E-04	0	0	NA	NA	3.0E-6
NEU1	8	7	5	rs9267901	9.1E-04	1	0	rs975666	0.4341	1.1E-4
SLC44A4	2	2	1	rs9275141	1.1E-06	0	0	NA	NA	2.0E-5
HLA-DRB6	60	13	6	rs522254	6.3E-04	0	0	NA	NA	5.8E-5
HLA-DRB1	97	16	7	rs9272230	5.2E-04	30	11	rs3806156	0.0001	5.0E-6
HLA-DQA1	79	23	5	rs504594	1.7E-05	6	4	rs1235162	0.0044	8.1E-5
HLA-DQB1	78	22	12	rs3129719	2.6E-06	4	3	rs1063355	1.8E-13	<1E-6
TAP2	48	36	11	rs2858312	1.9E-05	0	0	NA	NA	3.0E-6
TAP1	13	11	3	rs6928482	2.0E-08	1	1	rs653178	0.0363	<1E-6
Chromosome 15q22										
SMAD3	7	7	2	rs17293632	2.0E-07	0	0	NA	NA	1.7E-5
Chromosome 16p13										
CLEC16A	4	4	2	rs35441874	2.9E-08	0	0	NA	NA	2.0E-6
SOCS1	7	5	5	rs7184491	3.5E-06	1	0	rs1219648	0.141	<1E-6
Chromosome 17q12										
CISD3	4	4	2	rs2941503	1.6E-07	0	0	NA	NA	<1E-6
STARD3	7	3	1	rs2941503	1.6E-07	0	0	NA	NA	2.7E-5
PGAP3	2	2	1	rs903502	1.5E-06	0	0	NA	NA	1.4E-5
GRB7	2	1	1	rs14050	1.4E-07	0	0	NA	NA	<1E-6
IKZF3	9	7	3	rs7207600	4.5E-07	0	0	NA	NA	<1E-6
ZPBP2	2	2	1	rs9916765	1.9E-09	0	0	NA	NA	<1E-6
GSDMB	15	11	5	rs2952140	1.2E-08	0	0	NA	NA	<1E-6
ORMDL3	19	12	5	rs2952140	1.2E-08	2	0	rs4836703	0.4763	<1E-6
MED24	5	5	2	rs7502514	4.8E-05	0	0	NA	NA	1.1E-5
NR1D1	5	4	2	rs12150298	2.8E-06	0	0	NA	NA	4.4E-5
TOP2A	1	1	1	rs2102928	4.1E-05	0	0	NA	NA	4.6E-5

Table E4. Individual independent eQTLs contributing to a significant gene-based association test for *TSLP*.

eQTL	Study	Tissue	eQTL effect on gene expression			eQTL effect on asthma risk				Proxy-eQTL phase	Predicted effect of increased gene expression on asthma risk
			A1	Beta	*P*	Proxy tested	A1	OR	*P*		
rs12110124	[11]	LCLs	C	-3.3	6E-06	rs12110124	T	1.08	0.0022	Same	Increase
	[9]	Fibroblasts	C	-0.2	9E-09					Same	Increase
	[9]	Spleen	C	-0.8	2E-12					Same	Increase
rs17132582	[7]	Whole-blood	A	4.0	5E-05	rs17132582	A	1.16	0.0003	Same	Increase
rs2289278	[9]	Fibroblasts	G	0.3	5E-06	rs17132762	A	0.87	0.0028	AG/GC	Increase
rs252858	[7]	Whole-blood	T	-3.5	4E-04	rs252858	T	0.92	0.0167	Same	Increase

Table E5. Linkage disequilibrium (r^2) between the eQTLs most associated with asthma for each of the 11 genes in the 17q12 region that had a significant association with asthma.

Gene	eQTL	Asthma *P*-value	rs12150298	rs14050	rs2102928	rs2941503	rs2952140	rs7207600	rs7502514	rs903502	rs9916765
NR1D1	rs12150298	3.E-06	1.00								
GRB7	rs14050	1.E-07	0.84	1.00							
TOP2A	rs2102928	4.E-05	0.01	0.01	1.00						
CISD3, STARD3	rs2941503	2.E-07	0.85	0.99	0.01	1.00					
GSDMB, ORMDL3	rs2952140	1.E-08	0.38	0.36	0.02	0.37	1.00				
IKZF3	rs7207600	4.E-07	0.08	0.08	0.18	0.09	0.21	1.00			
MED24	rs7502514	5.E-05	0.01	0.02	0.10	0.02	0.05	0.22	1.00		
PGAP3	rs903502	1.E-06	0.94	0.89	0.01	0.90	0.38	0.08	0.01	1.00	
ZPBP2	rs9916765	2.E-09	0.35	0.35	0.03	0.35	0.92	0.23	0.06	0.36	1.00

Table E6. Linkage disequilibrium (r^2) between the eQTLs most associated with asthma for each of the 13 genes in the MHC region that had a significant association with asthma.

Gene	eQTL	Asthma *P*-value	rs1063355	rs13215091	rs2071595	rs2442752	rs2858312	rs3806156	rs504594	rs522254	rs6928482	rs9267901	rs9268764	rs9275141	rs9276595
HLA-DQB1	rs1063355	2.E-13	1.00												
HSPA1B	rs13215091	5.E-04	0.00	1.00											
HCP5	rs2071595	7.E-06	0.02	0.40	1.00										
LTA	rs2442752	2.E-05	0.00	0.01	0.02	1.00									
TAP2	rs2858312	2.E-05	0.18	0.00	0.18	0.01	1.00								
HLA-DRB1	rs3806156	1.E-04	0.15	0.02	0.03	0.01	0.07	1.00							
HLA-DQA1	rs504594	2.E-05	0.13	0.01	0.01	0.01	0.00	0.23	1.00						
HLA-DRB6	rs522254	6.E-04	0.05	0.01	0.01	0.00	0.00	0.26	0.58	1.00					
TAP1	rs6928482	2.E-08	0.67	0.00	0.01	0.01	0.33	0.30	0.20	0.10	1.00				
NEU1	rs9267901	9.E-04	0.02	0.00	0.00	0.00	0.01	0.02	0.01	0.02	0.03	1.00			
MICB	rs9268764	3.E-05	0.23	0.01	0.03	0.00	0.19	0.22	0.24	0.19	0.33	0.00	1.00		
SLC44A4	rs9275141	1.E-06	0.54	0.00	0.01	0.02	0.41	0.31	0.17	0.08	0.86	0.03	0.27	1.00	
HLA-DQB1-AS1	rs9276595	4.E-02	0.03	0.00	0.02	0.00	0.18	0.02	0.01	0.02	0.10	0.00	0.01	0.13	1.00

Table E7. Six genes associated with asthma at an empirical FDR of 0.05 that were not located in established risk loci for asthma but the significant gene-based associations were driven by *trans*-eQTLs located in the MHC region or near *ORMDL3*.

Gene	Chr	N eQTLs	***cis*-eQTLs**				***trans*-eQTLs**				**EUGENE *P*-value**
			N Tested	N with *P*<0.05	Best individual eQTL		N Tested	N with *P*<0.05	Best individual eQTL		
					SNP	P-value			SNP	P-value	
LIMS1	2	15	4	0	rs1522021	0.4581	10	4	rs1063355	1.8E-13	<1E-6
AOAH	7	26	17	0	rs2718180	0.0509	4	3	rs9268853	1.8E-06	1.5E-05
ZNF707	8	9	7	2	rs11778657	0.0067	1	1	rs17609240	1.5E-06	4.5E-05
TINF2	14	4	3	1	rs2273301	0.0010	1	1	rs3135006	1.7E-06	<1E-6
CLK3	15	2	1	0	rs4646421	0.5147	1	1	rs9268853	1.8E-06	1.3E-05
SAFB	19	2	1	0	rs2184854	0.6355	1	1	rs9268853	1.8E-06	1.0E-05

Table E8. Individual independent eQTLs contributing to a significant gene-based association test for *LIMS1, AOAH, ZNF707, TINF2, CLK3* and *SAFB.*

eQTL[a]	Study	Tissue	eQTL effect on gene expression			eQTL effect on asthma risk				Proxy-eQTL phase	Predicted effect of increased gene expression on asthma risk
			A1	Beta	*P*	Proxy tested	A1	OR	*P*		
					LIMS1						
rs13192471	[7]	Whole-blood	C	-8.8	1E-18	rs13192471	T	1.08	0.0147	Same	Increase
rs443198	[7]	Whole-blood	G	5.5	5E-08	rs443198	A	0.94	0.0076	Same	Increase
rs7765379	[7]	Whole-blood	G	5.6	2E-08	rs7765379	T	0.92	0.0229	Same	Increase
rs9272346	[7]	Whole-blood	G	-14	4E-47	rs1063355	T	0.83	2E-13	TG/GA	Increase
					AOAH						
rs674313	[7]	Whole-blood	T	6.1	1E-09	rs617058	T	1.11	0.0035	TT/GC	Increase
rs9268853	[17]	Monocytes	C	-6.3	4E-09	rs9268853	T	0.89	2E-06	Same	Decrease
	[17]	Monocytes-LPS	C	-6.2	6E-09	rs9268853	T	0.89	2E-06	Same	Decrease
	[7]	Whole-blood	C	-14	1E-44	rs9268853	T	0.89	2E-06	Same	Decrease
	[12]	PBMCs	T	Pos.	1E-61	rs9268853	T	0.89	2E-06	Same	Decrease
rs9357155	[7]	Whole-blood	A	-4.6	4E-06	rs9357155	A	0.91	0.0047	Same	Increase
					TINF2						
rs2273301	[7]	Whole-blood	A	3.1	2E-03	rs2273301	A	0.85	0.0010	Same	Decrease
rs3135006	[17]	Monocytes	T	-8.1	5E-13	rs3135006	T	0.88	2E-06	Same	Increase
	[17]	Monocytes-LPS	T	-9.4	3E-16	rs3135006	T	0.88	2E-06	Same	Increase
					ZNF707						
rs10097337	[7]	Whole-blood	A	5.4	5E-08	rs10097337	A	0.95	0.0353	Same	Decrease
rs11778657	[7]	Whole-blood	G	5.3	1E-07	rs11778657	A	0.93	0.0067	Same	Increase
rs17609240	[7]	Whole-blood	T	-4.6	4E-06	rs17609240	T	0.90	1E-06	Same	Increase
					CLK3						
rs9268853	[12]	PBMCs	T	Pos.	7E-17	rs9268853	T	0.89	2E-06	Same	Decrease
					SAFB						
rs9268853	[7]	Whole-blood	C	4.7	3E-06	rs9268853	T	0.89	2E-06	Same	Increase

[a]In bold: *trans*-eQTLs located in the MHC region. Underlined: *cis*-eQTLs. Italic: *trans*-eQTL located near ORMDL3. Pos.: positive beta.

Table E9. Association results in the Moffatt et al. [22] GWAS for the eleven genes that represent a potential novel genetic association with asthma.

Gene	Position	N eQTLs tested	N eQTLs with *P*<0.05	Best individual eQTL		EUGENE *P*-value
				SNP	*P*-value	
B4GALT3	1:161141100	7	2	rs11587213	0.0102	0.0175
P2RY14	3:150929905	9	1	rs2870518	0.0172	0.0412
P2RY13	3:151044100	8	1	rs2870518	0.0172	0.0484
P2RY12	3:151055168	6	0	rs3732765	0.0598	0.1165
F12	5:176829141	3	0	rs2731672	0.4332	0.8810
HIBADH	7:27565061	11	1	rs16874305	0.0442	0.4670
REEP3	10:65281123	1	0	rs7915849	0.6660	0.6640
USMG5	10:105148798	11	3	rs7897947	0.0004	0.0014
PTCSC3	14:36605314	1	0	rs1766142	0.9076	0.9052
DYNC1H1	14:102430865	3	0	rs12590618	0.6588	0.9530
ACO2	22:41865129	2	0	rs132902	0.1161	0.1583

Table E10. Direction of effect on asthma risk for asthma-associated eQTLs in *P2RY13*, *P2RY14*, *USMG5* and *B4GALT3*, in two independent GWAS.

eQTL	eQTL effect in Ferreira et al 2014				eQTL effect in Moffatt et al 2010				LD between Ferreira and Moffatt proxies		Consistency in direction of effect
	Proxy tested	A1	OR	*P*	Proxy tested	A1	OR	*P*	r^2	Phase	
					P2RY13						
rs2870518	rs2870518	T	0.95	0.0364	rs2870518	T	0.95	0.0172	1.00	Same SNP	Same
rs6440732[a]	rs6440732	A	0.93	0.0102	rs1466684	G	1.02	0.4071	1.00	CG/AA	Same
rs6440742	rs6440742	T	0.94	0.0199	rs6781302	G	0.99	0.6044	0.95	GG/TA	Opposite
rs9814936	rs9814936	A	1.12	0.0001	rs9848789	T	0.96	0.1189	1.00	GT/AC	Same
rs9877416	rs9877416	A	0.90	0.0001	rs9877416	G	1.05	0.0770	1.00	Same SNP	Same
					P2RY14						
rs10513393	rs10513393	A	0.89	0.0001	rs9848789	T	0.96	0.1189	1.00	AT/GC	Same
rs17204536	rs2276765	A	0.94	0.0125	rs3732765	G	0.96	0.0598	0.98	GA/AG	Same
rs2870518	rs2870518	T	0.95	0.0364	rs2870518	T	0.95	0.0172	1.00	Same SNP	Same
rs7616382	rs7616382	A	0.93	0.0271	rs1907637	G	0.98	0.3918	1.00	TA/AG	Same
rs9843590	rs9843590	A	1.08	0.0040	NA	NA	NA	NA	NA	NA	NA
					USMG5						
rs11191724	rs11191724	A	0.95	0.0495	NA	NA	NA	NA	NA	NA	NA
rs1163073	rs1163073	T	0.92	0.0005	rs1163073	T	0.97	0.2240	1.00	Same SNP	Same
rs1572530	rs1572530	A	1.05	0.0644	rs7897947	T	0.91	0.0004	0.82	AG/GT	Same
rs17784294	rs17784294	A	0.95	0.0693	rs7904252	T	0.93	0.0020	0.97	AT/CG	Same
rs2250580	rs2250580	C	1.05	0.0734	rs2486757	T	1.06	0.0135	0.99	CT/GC	Same
rs2271750	rs2271750	A	0.91	0.0048	rs2271750	G	1.02	0.4629	1.00	Same SNP	Same
rs999867	rs999867	T	1.10	0.0211	rs999867	T	1.04	0.2720	1.00	Same SNP	Same
					B4GALT3						
rs11579627	rs11579627	A	1.07	0.0124	rs11581556	G	1.04	0.1137	0.89	AA/GG	Opposite
rs11587213	rs11587213	A	1.04	0.1997	rs11587213	G	0.93	0.0102	1.00	Same SNP	Same
rs4233366	rs4233366	T	1.08	0.0033	rs4233366	T	1.05	0.0236	1.00	Same SNP	Same
rs1668873[b]	rs1668873	A	0.93	0.0015	rs7531256	T	1.02	0.3060	0.91	AG/GT	Same

[a]rs6440732 is in LD (r^2=0.94) with a missense SNP (rs1466684; T158M) in *P2RY13*.
[b]rs1668873 is a *trans*-eQTL; all others are *cis*-eQTLs.

Table E11. Direction of effect on gene expression and disease risk for asthma-associated eQTLs in *P2RY13*, *P2RY14*, *USMG5* and *B4GALT3*.

eQTL[a]	Study	Tissue	eQTL effect on gene expression			eQTL effect on asthma risk				Proxy - eQTL phase	Predicted effect of increased gene expression on asthma risk
			A1	Beta	*P*	Proxy tested	A1	OR	*P*		
					P2RY13						
rs2870518	[7]	Whole-blood	C	4.7	2E-06	rs2870518	T	0.95	0.0364	Same	Increase
rs6440732	[7]	Whole-blood	C	24	2E-125	rs6440732	A	0.93	0.0102	Same	Increase
rs6440742	[7]	Whole-blood	G	10	2E-25	rs6440742	T	0.94	0.0199	Same	Increase
rs9814936	[7]	Whole-blood	G	-8.5	3E-17	rs9814936	A	1.12	0.0001	Same	Increase
rs9877416	[7]	Whole-blood	G	5.6	2E-08	rs9877416	A	0.90	0.0001	Same	Increase
					P2RY14						
rs10513393	[6]	Whole-blood	NA	NA	4E-56	rs10513393	A	0.89	0.0001	Same	Increase
	[7]	Whole-blood	A	-14	2E-42					Same	Increase
rs17204536	[7]	Whole-blood	T	5.7	1E-08	rs2276765	A	0.94	0.0125	GT/AC	Increase
rs2870518	[7]	Whole-blood	C	5	5E-07	rs2870518	T	0.95	0.0364	Same	Increase
rs7616382	[7]	Whole-blood	T	3.4	8E-04	rs7616382	A	0.93	0.0271	Same	Increase
rs9843590	[7]	Whole-blood	A	3.7	3E-04	rs9843590	A	1.08	0.0040	Same	Increase
					USMG5						
rs11191724	[7]	Whole-blood	A	8.8	1E-18	rs11191724	A	0.95	0.0495	Same	Decrease
rs1163073	[18]	Neutrophils	C	Pos.	5E-25	rs1163073	T	0.92	0.0005	Same	Increase
	[11]	LCLs	C	73	4E-119					Same	Increase
	[15]	LCLs	T	-1.2	4E-116					Same	Increase
	[15]	Skin	T	-0.6	3E-90					Same	Increase
	[7]	Whole-blood	C	64	1E-197					Same	Increase
	[12]	PBMCs	C	Pos.	0E+00					Same	Increase
rs1572530	[12]	PBMCs	G	Neg.	2E-15	rs1572530	A	1.05	0.0644	Same	Increase
rs17784294	[11]	LCLs	A	-19	2E-06	rs17784294	A	0.95	0.0693	Same	Increase
	[12]	PBMCs	C	Pos.	3E-14					Same	Increase
rs2250580	[12]	PBMCs	C	Pos.	2E-12	rs2250580	C	1.05	0.0734	Same	Increase
rs2271750	[11]	LCLs	A	-30	2E-06	rs2271750	A	0.91	0.0048	Same	Increase
	[7]	Whole-blood	A	-15	1E-48					Same	Increase
	[12]	PBMCs	G	Pos.	2E-13					Same	Increase
rs999867	[15]	LCLs	T	0.6	3E-13	rs999867	T	1.10	0.0211	Same	Increase
	[15]	Skin	T	0.3	1E-10					Same	Increase
	[12]	PBMCs	T	Pos.	5E-24					Same	Increase
					B4GALT3						
rs11579627	[13]	Neutrophils	A	-0.1	3E-05	rs11579627	A	1.07	0.0124	Same	Decrease
rs11587213	[9]	Fibroblasts	G	Neg.	1E-07	rs11587213	A	1.04	0.1997	Same	Increase
	[7]	Whole-blood	G	-4.8	2E-06					Same	Increase
rs4233366	[9]	Fibroblasts	T	Pos.	2E-26	rs4233366	T	1.08	0.0033	Same	Increase
rs1668873	[7]	Whole-blood	A	5	6E-07	rs1668873	A	0.93	0.0015	Same	Decrease

[a]In bold: *trans*-eQTL. Pos.: positive beta. Neg.: negative beta.

Table E12. Primers used in gene expression analyses of RNA extracted from mouse lung.

Gene	Oligonucleotide Primer
P2ry13	Forward: 5'- GTGGGTTGAGCTAGTAACTGCC-3' Reverse: 5'- CATCCCAGTGGTGTTGATTG-3'
P2ry14	Forward: 5'- TCCTCCAGACACACTGATGC-3' Reverse: 5'- AAAGGCAAGCTTCGTCAACA-3'
Hprt	Forward: 5'- AGGCCAGACTTTGTTGGATTTGAA-3' Reverse: 5'-CAACTTGCGCTCATCTTAGGCTTT-3'

E FIGURE LEGENDS

Figure E1. Number of independent eQTLs per gene identified from published GWAS of gene expression.

Figure E2. Expression of P2ry13 and P2ry14 in lymphocytes **(A)**, monocytes (**B**), conventional dendritic cells (**C**) and plasmocytoid dendritic cells **(D)** collected in BALF after saline or HDM challenge.

Figure E3. **In vivo exposure to P2ry13 and P2ry14 receptor agonists in naïve C57Bl/6 mice.** Total number of monocytes **(A)**, neutrophils **(B)** and conventional dendritic cells **(C)** recruited to the BALF based on flow cytometry analysis. Veh: vehicle. HDM: house dust mite allergen. CRE: cockroach allergen. Alt: alternaria allergen.

Collaborators of the Australian Asthma Genetics Consortium (AAGC)

David L. Duffy[a] , Dale R. Nyholt[a], John Beilby[b-d], Loren Price[e], Faang Cheah[e], Desiree Mészáros[f], Scott D. Gordon[a], Melissa C. Southey[g], Margaret J. Wright[a], James Markos[h], Li P. Chung[e], Anjali K. Henders[a], Graham Giles[i], Suzanna Temple[e], John Whitfield[a], Brad Shelton[e], Chalermchai Mitrpant[e], Minh Bui, PhD,[j] Mark Jenkins[j], Haydn Walters[f], Michael J. Abramson[k], Michael Hunter[l,d], Bill Musk [l,d,m,n], Matthew A Brown[o], Shyamali C. Dharmage[j]

[a] QIMR Berghofer Medical Research, Brisbane, Australia.

[b] PathWest Laboratory Medicine of Western Australia (WA), Nedlands, Australia.

[c] School of Pathology and Laboratory Medicine, The University of WA, Nedlands, Australia.

[d] Busselton Population Medical Research Foundation, Sir Charles Gairdner Hospital, Perth, Australia.

[e] Lung Institute of WA and Centre for Asthma, Allergy and Respiratory Research, University of WA, Perth, Australia.

[f] Menzies Research Institute, Hobart, Australia.

[g] Department of Pathology, The University of Melbourne, Melbourne, Australia.

[h] Launceston General Hospital, Launceston, Australia.

[i] Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Australia.

[j] Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Australia.

[k] Department of Epidemiology & Preventive Medicine, Monash University, Melbourne, Australia

[l] School of Population Health, The University of WA, Nedlands, Australia

[m] School of Medicine and Pharmacology, University of Western Australia, Nedlands, Australia

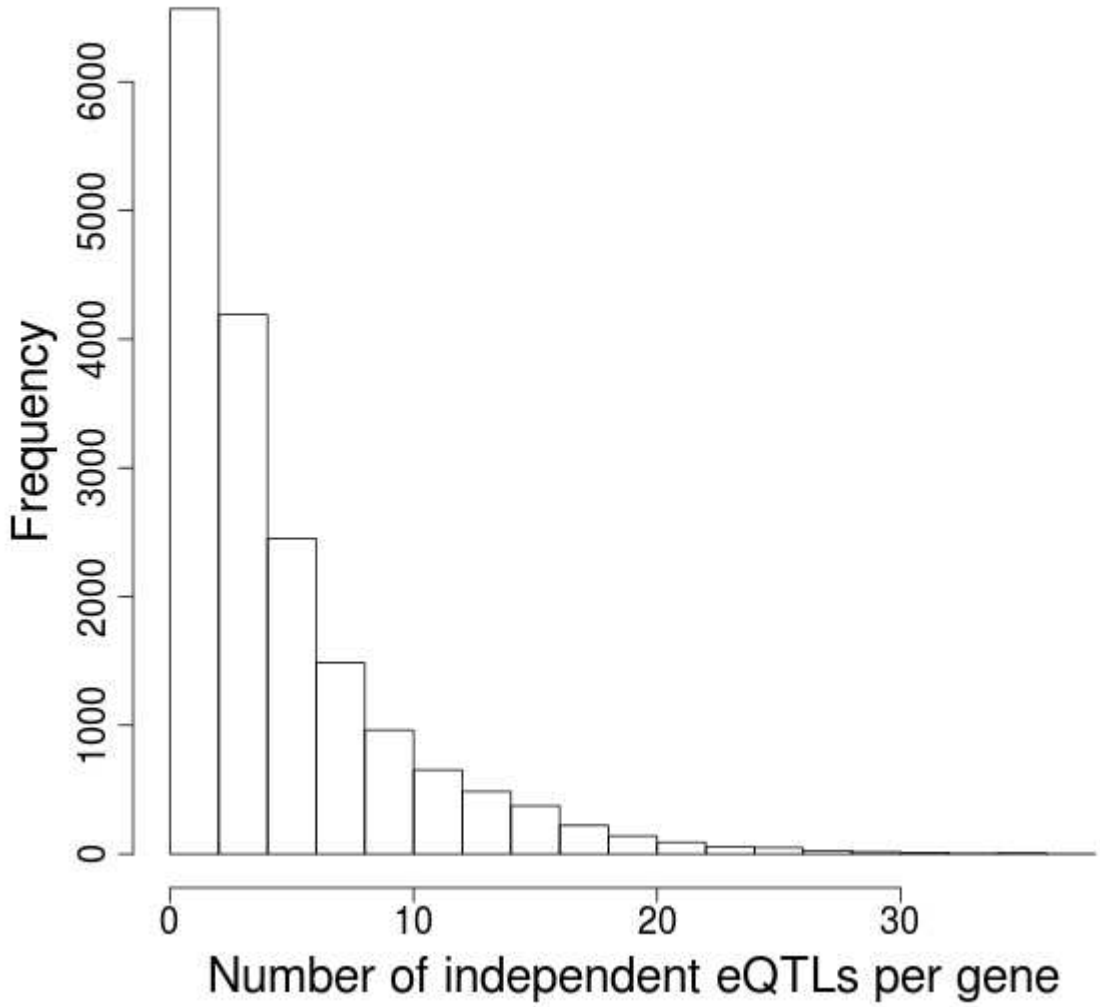
[n] Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Perth, Australia

[o] The University of Queensland Diamantina Institute, Translational Research Institute, Australia

SUPPLEMENTAL REFERENCES

1. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559-75.
2. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491:56-65.
3. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003; 100:9440-5.
4. Ullah MA, Revez JA, Loh Z, Simpson J, Zhang V, Bain L, et al. Allergen-induced IL-6 trans-signaling activates gammadelta T cells to promote type 2 and type 17 airway inflammation. J Allergy Clin Immunol 2015; 136:1065-73.
5. Davidson S, Kaiko G, Loh Z, Lalwani A, Zhang V, Spann K, et al. Plasmacytoid dendritic cells promote host defense against acute pneumovirus infection via the TLR7-MyD88-dependent signaling pathway. J Immunol 2011; 186:5938-48.
6. arc OC, arc OC, Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, et al. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet 2012; 380:815-23.
7. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013; 45:1238-43.
8. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet 2012; 8:e1003029.
9. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015; 348:648-60.
10. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 2012; 44:502-10.
11. 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.
12. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One 2010; 5:e10693.
13. Naranbhai V, Fairfax BP, Makino S, Humburg P, Wong D, Ng E, et al. Genomic modulators of gene expression in human neutrophils. Nat Commun 2015; 6:7545.
14. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007; 447:799-816.
15. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature; 467:832-8.
16. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, et al. A genome-wide association study of global gene expression. Nat Genet 2007; 39:1202-7.
17. Kim S, Becker J, Bechheim M, Kaiser V, Noursadeghi M, Fricker N, et al. Characterizing the genetic basis of innate immune response in TLR4-activated human monocytes. Nat Commun

2014; 5:5236.
18. Andiappan AK, Melchiotti R, Poh TY, Nah M, Puan KJ, Vigano E, et al. Genome-wide analysis of the genetic regulation of gene expression in human neutrophils. Nat Commun 2015; 6:7971.
19. Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, Chen W, et al. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. Am J Hum Genet 2010; 87:779-89.
20. Luo W, Obeidat M, Di Narzo AF, Chen R, Sin DD, Pare PD, et al. Airway Epithelial Expression Quantitative Trait Loci Reveal Genes Underlying Asthma and Other Airway Diseases. Am J Respir Cell Mol Biol 2016; 54:177-87.
21. Brumpton BM, Ferreira MA. Multivariate eQTL mapping uncovers functional variation on the X-chromosome associated with complex disease traits. Hum Genet 2016.
22. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010; 363:1211-21.


Frequency
0
1000
2000
3000
4000
5000
6000
0
10
20
30
Number of independent eQTLs per gene