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

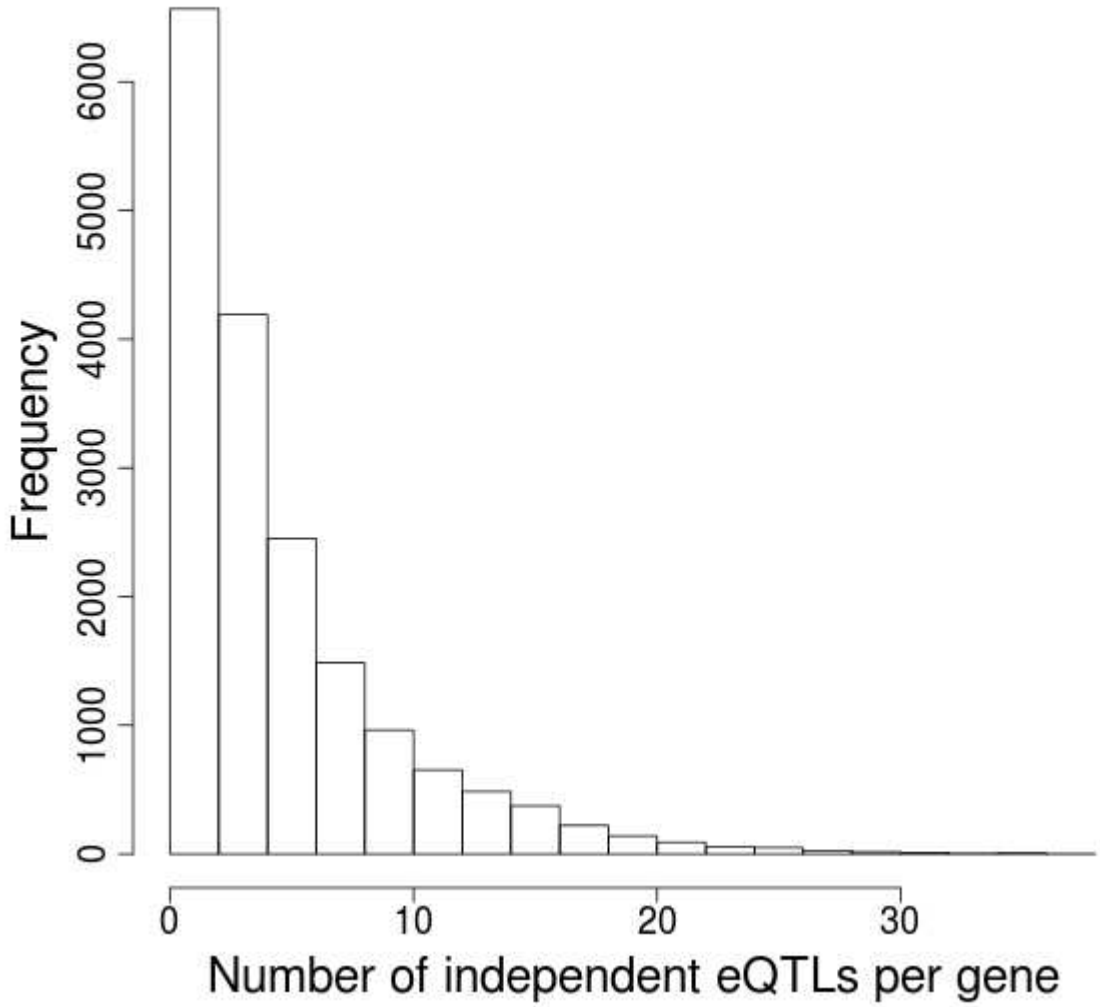
[n] Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Perth, Australia

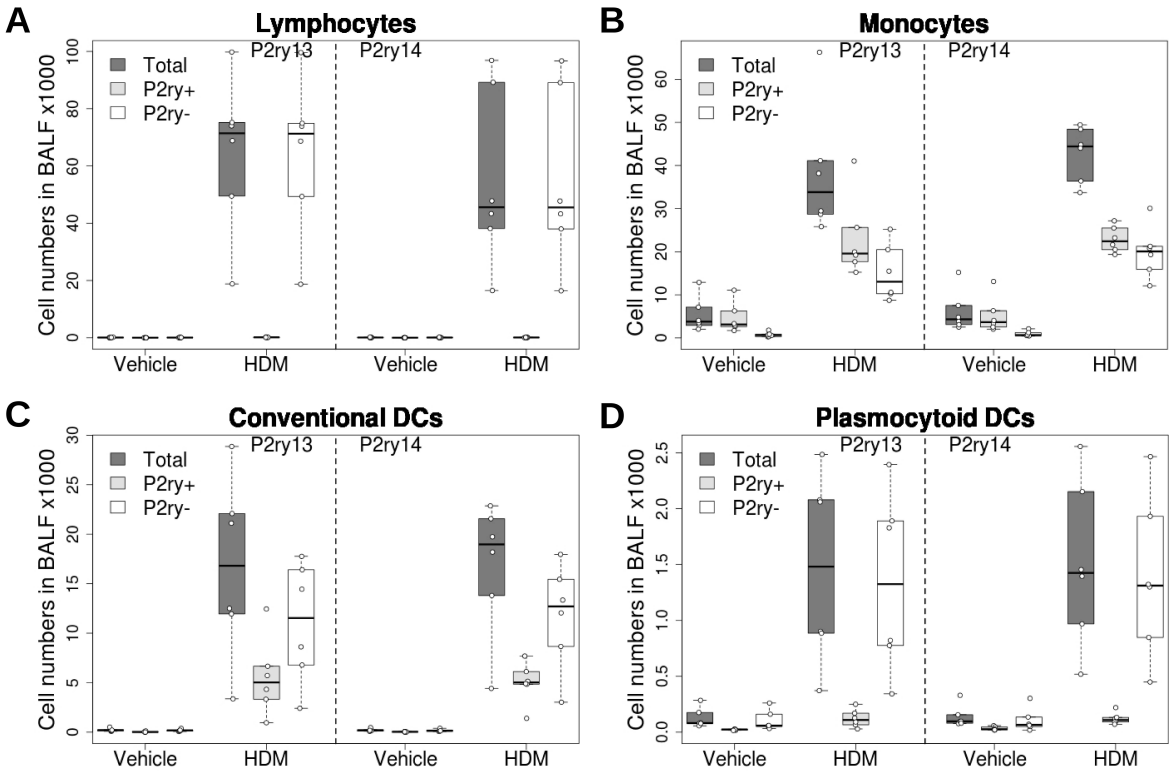
[o] The University of Queensland Diamantina Institute, Translational Research Institute, Australia

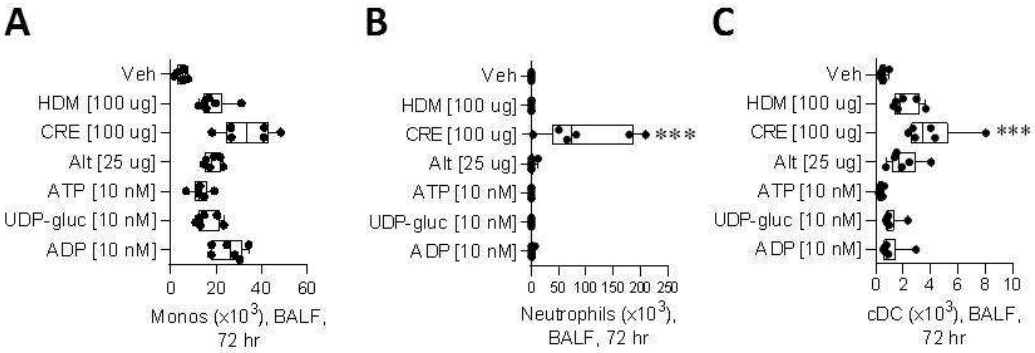
SUPPLEMENTAL REFERENCES

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


A
Lymphocytes
B
Monocytes
C
Conventional DCs
D
Plasmocytoid DCs
Total
P2ry+
P2ry-
P2ry13
P2ry14
Cell numbers in BALF x1000
Vehicle
HDM
Vehicle
HDM


A
Veh
HDM [100 ug]
CRE [100 ug]
Alt [25 ug]
ATP [10 nM]
UDP-gluc [10 nM]
ADP [10 nM]
0 20 40 60
Monos (x10^3), BALF, 72 hr
B
Veh
HDM [100 ug]
CRE [100 ug]
Alt [25 ug]
ATP [10 nM]
UDP-gluc [10 nM]
ADP [10 nM]
***
0 50 100 150 200 250
Neutrophils (x10^3), BALF, 72 hr
C
Veh
HDM [100 ug]
CRE [100 ug]
Alt [25 ug]
ATP [10 nM]
UDP-gluc [10 nM]
ADP [10 nM]
***
0 2 4 6 8 10
cDC (x10^3), BALF, 72 hr