Genome-wide association analysis identifies eleven risk variants associated with asthma-with-hayfever phenotype

ONLINE REPOSITORY

Manuel A.R. Ferreira, PhD,[a]* Melanie C. Matheson, PhD,[b]* Clara S. Tang, PhD,[a,c] Raquel Granell, PhD,[d] Wei Ang, PhD,[e] Jennie Hui, PhD,[f-i] Amy K. Kiefer, PhD,[j] David L. Duffy, PhD,[a] Svetlana Baltic, PhD,[k] Patrick Danoy, PhD,[l] Minh Bui, PhD,[b] Loren Price, PhD,[k] Peter D. Sly, FRACP,[m] Nicholas Eriksson, PhD,[j] Pamela A. Madden, PhD,[n] Michael J. Abramson, PhD,[o] Patrick G. Holt, FRCP,[p] Andrew C. Heath, DPhil,[n] Michael Hunter, PhD,[g,i] Bill Musk, FRACP, [g,i,q,r], AAGC collaborators[s], Colin F. Robertson, FRACP,[t] Peter Le Souëf, FRACP,[u] Grant W. Montgomery, PhD,[a] John A. Henderson, PhD,[d] Joyce Y. Tung, PhD,[j] Shyamali C. Dharmage, PhD,[b] Matthew A. Brown, FRACP,[l] Alan James, FRACP,[i,q,v] Philip J. Thompson, FRACP,[k] Craig Pennell, PhD,[e] Nicholas G. Martin, PhD,[a] David M. Evans, PhD,[d,w] David A. Hinds, PhD,[j] and John L. Hopper, PhD,[b]

Supplementary Methods

Genotyping, quality control (QC) and imputation procedures

AAGC. All 2,137 samples were genotyped with the Illumina 610K array. SNPs were excluded from analysis if the call rate was <95%, minor allele frequency (MAF) < 0.01 and Hardy-Weinberg equilibrium test *P*-value < 10^{-6}. SNPs passing QC were then used to impute with Impute2[E1] 5.7 million variants with a MAF ≥ 0.01 and information > 0.3 using the combined 1000 Genomes Project and HapMap 3 phased data (all ancestral groups) as reference panels. X chromosome SNPs (*N* =140,388 with MAF ≥ 0.01) were imputed with r^2 > 0.3 with MACH/minimac[E2, 3] using the 1000G Phase I Integrated Release Version 3 (*N* = 584 European haplotypes) release of the 1000 Genomes Project.[E4] All subjects were confirmed to be unrelated and of European ancestry through the analysis of genome-wide allele sharing.

23andMe. DNA extraction and genotyping were performed by the National Genetics Institute (NGI), a CLIA-certified clinical laboratory and subsidiary of Laboratory Corporation of America. Samples were genotyped on one of two platforms. About 35% of the participants were genotyped using the Illumina HumanHap550+ BeadChip, which included SNPs from the standard HumanHap550 panel augmented with a custom set of approximately 25,000 SNPs selected by 23andMe. Two slightly different versions of this platform were used, as previously described.[E5] The remaining 65% of participants were genotyped using the Illumina HumanOmniExpress+ Bead Chip. This has a base set of 730,000 SNPs, augmented with approximately 250,000 SNPs to obtain a superset of the HumanHap550+ content, as well as a custom set of about 30,000 SNPs. Every sample that failed to reach a 98.5% call rate for SNPs on the standard platforms was re-analyzed. Persons included in the analysis were selected for having >97% European ancestry, as determined through an analysis of local ancestry via comparison to the three HapMap 2 populations.[E6] A maximal set of unrelated

individuals was chosen for the analysis using a segmental identity-by-descent (IBD) estimation algorithm.[E7] Individuals were defined as related if they shared more than 700 cM IBD, including regions where the two individuals share either one or both genomic segments IBD. This level of relatedness (roughly 20% of the genome) corresponds approximately to the minimal expected sharing between first cousins in an outbred population. Participant genotype data were imputed against the August 2010 release of 1000 Genomes reference haplotypes.[E8] First, we used Beagle[E9] (version 3.3.1) to phase batches of 8,000-9,000 individuals across chromosomal segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs. We excluded SNPs with MAF < 0.001, Hardy-Weinberg equilibrium $P < 10^{-20}$, call rate < 95%, or with large allele frequency discrepancies compared to the 1000 Genomes reference data. We then assembled full phased chromosomes by matching the phase of haplotypes across the overlapping segments. We imputed each batch against the European subset of 1000 Genomes haplotypes using minimac,[E10] using 5 rounds and 200 states for parameter estimation. Analyses were limited to 7.4 million SNPs with imputed $r^2 > 0.5$ averaged across all batches, and $r^2 > 0.3$ in every batch. For the non-pseudoautosomal region of the X chromosome, males and females were phased together in segments, treating the males as already phased; the pseudoautosomal regions were phased separately. We assembled fully phased X chromosomes, representing males as homozygous pseudo-diploids for the non-pseudoautosomal region. We then imputed males and females together using minimac as with the autosomes.

ALSPAC. A total of 9,912 children were genotyped using the Illumina HumanHap550 quad array by 23andMe subcontracting the Wellcome Trust Sanger Institute, Cambridge, UK and the NGI, USA. Individuals were excluded from further analysis on the basis of having incorrect gender assignments; minimal or excessive heterozygosity (<0.320 and >0.345 for the Sanger data and <0.310 and >0.330 for the LabCorp data); disproportionate levels of

individual missingness (>3%); evidence of cryptic relatedness (>10% IBD) and being of non-European ancestry (as detected by a multidimensional scaling analysis seeded with HapMap 2 individuals). The resulting data set consisted of 8,365 children. SNPs with a MAF of < 0.01 and call rate of < 95% were removed. Furthermore, only SNPs which passed an exact test of Hardy Weinberg equilibrium ($P > 5 \times 10^{-7}$) were considered for analysis. After cleaning, 500,527 directly genotyped SNPs were available for analysis. HapMap variants were imputed with MACH 1.0.16 Markov Chain Haplotyping software,[E2, 3] using CEPH individuals from phase 2 of the HapMap project as a reference set (release 22).

Raine. All 1,494 samples were genotyped using the Illumina 660K array. SNPs were excluded from analysis if the call rate was <95%, MAF < 0.01 and Hardy-Weinberg equilibrium test *P*-value < 5.7×10^{-7}. SNPs passing QC were then used to impute HapMap SNPs with MACH v1.0.16.[E2, 3] Only SNPs with imputed $r^2 > 0.3$ were retained for analysis. All subjects were confirmed to be unrelated and of Caucasian ancestry through the analysis of genome-wide allele sharing.

Association analyses

AAGC. Autosomal SNPs were tested for association with case-control status using an allelic test of association implemented in PLINK;[E11] for imputed variants, we analysed best-guess genotypes. X chromosome SNPs were analysed with logistic regression under an additive model (dosage for imputed variants) using MACH2DAT[E2, 3] and assuming a dosage compensation model, ie. equating hemizygous males to homozygous females,[E12] such that the allelic dosage extremes for males were 0 (if A/-, as for AA females) and 2 (if B/-, as for BB females).

23andMe. Genome-wide association analyses of asthma-with-hayfever were performed using custom software. We computed likelihood ratio tests for an additive effect

of allelic dosage using logistic regression with covariates for age, gender, and five principal components to account for residual population substructure. The X chromosome was treated the same as the autosomes; allele dosages ranged from 0 to 2 for both men and women, as for the AAGC analysis.

ALSPAC. Genome-wide association analysis of asthma-with-hayfever was carried out using MACH2DAT[E2, 3] by regressing expected allelic dosage on case-control status and assuming an additive model.

Raine. Autosomal SNPs were analysed with logistic regression under an additive model (dosage for imputed variants) using ProbABEL.[E13]

E References

E1. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009; 5:e1000529.

E2. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet 2009; 10:387-406.

E3. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010; 34:816-34.

E4. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491:56-65.

E5. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet 2010; 6:e1000993.

E6. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003; 164:1567-87.

E7. Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, et al. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. PLoS One 2012; 7:e34267.

E8. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010; 467:1061-73.

E9. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 2007; 81:1084-97.

E10. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012; 44:955-9.

E11. 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.

E12. Clayton D. Testing for association on the X chromosome. Biostatistics 2008; 9:593-600.

E13. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 2010; 11:134.

E14. 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.

E15. Ferreira MA, Hottenga JJ, Warrington NM, Medland SE, Willemsen G, Lawrence RW, et al. Sequence variants in three loci influence monocyte counts and erythrocyte volume. Am J Hum Genet 2009; 85:745-9.

E16. Ferreira MA, Mangino M, Brumme CJ, Zhao ZZ, Medland SE, Wright MJ, et al. Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. Am J Hum Genet 2010; 86:88-92.

E17. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet 2012; 44:187-92.

E18. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007; 39:857-64.
E19. Ramasamy A, Curjuric I, Coin LJ, Kumar A, McArdle WL, Imboden M, et al. A genome-wide meta-analysis of genetic variants associated with allergic rhinitis and grass sensitization and their interaction with birth order. J Allergy Clin Immunol 2011; 128:996-1005.
E20. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011; 476:214-9.
E21. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26:2336-7.

E Tables

Table E1. Participants included in the GWAS of asthma-with-hayfever.

Cohort	Cases						Controls		
	N	Age, mean (range)	Sex, N female (%)	Asthma age-of-onset, N (%)			N	Age (mean, range)	Sex, N female (%)
				≤16	>16	Unknown			
AAGC	1,505	41 (4 - 89)	923 (61)	772 (51)	462 (31)	271 (18)	632	37 (12 - 91)	379 (60)
23andMe	4,230	48 (6-102)	2,272 (54)	2,329 (55)	1,898 (45)	3 (0)	10,842	49 (2-97)	4,180 (39)
ALSPAC	668	14 (14-14)	283 (42)	668 (100)	0 (0)	0 (0)	2,132	14 (14-14)	1,118 (52)
Raine	282	17 (16-18)	127 (45)	274 (97)	8 (3)	0 (0)	485	17 (16-18)	247 (51)
Total	**6,685**		**4,055 (61)**	**4,043 (61)**	**2,368 (35)**	**274 (4)**	**14,091**		**5,924 (42)**

Table E2. Screening items used to classify hayfever status.

Cohort	Study	Sub-study	N cases	N controls	Question used to classify hayfever status	Cases	Controls
AAGC	QIMR	ADOL	81	280	Has a doctor ever diagnosed you as suffering from hayfever?	"Yes"	"No"
		ASTHMA	346	0	Have you EVER had hayfever or nasal allergies?	"Yes"	"No"
		ALC1	0	118	How often have you had Hay fever? [before & after age 14]	"Sometimes" or "Often", before or after age 14	"Never", before & after age 14
		CANB	0	25	How often have you had Hay fever?	"Only as a child" or "Quite often"	"Never"
		ECZEMA	71	0	Has a doctor ever diagnosed you as suffering from Hayfever?	"Yes"	"No"
	LIWA		438	37	How often have you had Hay fever?	"Yes"	"No"
	TAHS		296	0	Have you ever had hayfever(that is sneezing running or blocked nose when you do not have a cold or the flu)?	"Yes"	"No"
	BUSS		273	172	Q1: Has your doctor ever told you that you had Hayfever? Q2: Do you sneeze or get an itchy running nose? Q3: Has your chest ever made a wheezing or whistling sound? Q4: Have you ever felt tight in the chest?	"Yes" to Q1	"No" to Q1 & Q2 & Q3 & Q4
Raine			282	485	Has anyone ever told you that your child has hayfever?	"Paediatrician, specialist or doctor diagnosed"	"Not diagnosed by paediatrician, specialist or doctor"
ALSPAC			668	2,132	Q1: Has he/she had hay fever in the past 12 months? [reported at age 11 and 14] Q2: Has he/she ever had hay fever? [reported at age 11 and 14]	"Yes" to Q1 or Q2	"No" to Q1 and Q2
23andMe			4,230	10,842	Have you ever had allergic rhinitis (stuffed or dripping nose caused by allergies)?	"Yes"	"No"
Total			**6,685**	**14,091**			

Table E3. Breakdown of samples used for each set of analyses performed.

AAGC				
	Asthma			Total
Hayfever	+	-	?	
+	1505	317	0	1822
-	717	632	0	1349
?	395	891	0	1286
Total	2617	1840	0	4457

23andMe				
	Asthma			Total
Hayfever	+	-	?	
+	4230	8046	293	12569
-	1138	10842	142	12122
?	51	151	0	202
Total	5419	19039	435	24893

ALSPAC				
	Asthma			Total
Hayfever	+	-	?	
+	668	806	234	1708
-	554	2132	247	2933
?	351	321	0	672
Total	1573	3259	481	5313

Raine				
	Asthma			Total
Hayfever	+	-	?	
+	282	132	0	414
-	367	485	0	852
?	5	4	0	9
Total	654	621	0	1275

Samples included for analyses			
	Asthma (A)		
Hayfever (H)	+	-	?
+	a	b	c
-	d	e	f
?	g	h	i

Analysis	Cases	Controls	Results tables
A+H+ vs A-H-	a	e	1,E5,E6,E13,E14
A+ vs A-	a+d+g	b+e+h	E8
H+ vs H-	a+b+c	d+e+f	E8
A+ vs A-H-	a+d+g	e	E8
H+ vs A-H-	a+b+c	e	E8
A+H- vs A-H-	d	e	E8
A-H+ vs A-H-	b	e	E8
A+H- vs A-H+	d	b	E9

Table E4. Participants included in the replication stage.

Cohort	Cases						Controls		
	N	Age, mean (range)	Sex, N female (%)	Asthma age-of-onset, N (%)			N	Age (mean, range)	Sex, N female (%)
				≤16	>16	Unknown			
23andMe	878	47 (13 - 95)	535 (61)	480 (55)	398 (45)	0 (0)	2,455	46 (13 - 91)	941 (38)

Table E5. Association results for the top 11 variants obtained after applying genomic control (GC) to the observed SNP association test statistics.

.Chr	Position, bp	Nearest gene, kb distance	SNP, risk allele	Risk allele frequency, range	OR (95% CI)	GC-corrected† association *P*-value
6	32734579	*HLA-DQB1*,1	rs9273373,G	0.54-0.58	1.24 (1.17-1.31)	2×10^{-13}
4	38476105	*TLR1*,2*	rs4833095,T	0.74-0.76	1.20 (1.14-1.26)	2×10^{-11}
5	110495398	*WDR36*,1	rs1438673,C	0.49-0.52	1.16 (1.11-1.21)	9×10^{-11}
2	102332981	*IL1RL1*,2*	rs10197862,A	0.85-0.86	1.24 (1.16-1.32)	1×10^{-10}
11	75976842	*LRRC32*,69	rs2155219,T	0.48-0.52	1.16 (1.11-1.21)	2×10^{-10}
17	35376206	*GSDMA*,3*	rs7212938,G	0.46-0.48	1.16 (1.10-1.21)	9×10^{-10}
5	110429771	*TSLP*,6	rs1837253,C	0.71-0.75	1.17 (1.11-1.24)	3×10^{-9}
9	6165855	*IL33*,40	rs72699186,T	0.15-0.16	1.26 (1.17-1.36)	5×10^{-9}
8	81454434	*ZBTB10*,106	rs7009110,T	0.36-0.41	1.14 (1.09-1.20)	8×10^{-9}
15	65255339	*SMAD3*,19*	rs17294280,G	0.23-0.27	1.18 (1.12-1.25)	1×10^{-8}
16	11136213	*CLEC16A*,47*	rs62026376,C	0.72-0.74	1.17 (1.11-1.24)	3×10^{-8}

* SNP is located within reported gene.

†We first adjusted the SE of the beta estimated for each SNP (N = 4,972,397) tested in the 23andMe study for the observed genome-wide λ of 1.055. The AAGC, ALSPAC and Raine GWAS had λ ~ 1 (**Fig. E1**) and so no GC correction was applied to these three studies. Next, we performed the meta-analysis of the 23andMe GC-corrected GWAS results with the original GWAS from the AAGC, ALSPAC and Raine, as described in the main text; this meta-analysis had a λ of 1.016. Lastly, we applied GC correction to the meta-analysis, resulting in a final λ of 1.00. All analyses performed with METAL.

Table E6. Association results for rs7009110 and rs62026376 in the four individual cohorts.

Cohort	N cases	N controls	*ZBTB10* (rs7009110, A)			*CLEC16A* (rs62026376, C) *		
			OR	SE	*P*-value	OR	SE	*P*-value
AAGC	1,505	632	1.199	0.069	0.0086	1.200	0.076	0.0164
23andMe	4,230	10,842	1.127	0.027	7×10^{-6}	1.171	0.030	2×10^{-7}
ALSPAC	668	2,132	1.185	0.065	0.0089	1.025	0.073	0.7326
Raine	282	485	1.169	0.108	0.1471	1.088	0.123	0.5032

*The sentinel SNP for *CLEC16A* (rs62026376, meta-analysis $P = 1 \times 10^{-8}$) was not available in the ALSPAC and Raine studies. Instead, the results shown in this table for those two studies correspond to the proxy SNP ($r^2 = 0.95$) rs12935657 (G is the predisposing allele, in phase with rs62026376:C). For rs12935657, the four-study meta-analysis *P*-value was 5×10^{-8}.

Table E7. Association with the *ZBTB10* locus in eczema-free people.

Cohort	N cases	N controls	***ZBTB10* (rs7009110, T)**		
			OR	SE	*P*-value
AAGC	674	607	1.172	0.082	0.0520
23andMe	2,648	9,955	1.159	0.032	4×10^{-6}
ALSPAC	96	821	1.078	0.161	0.64
Raine	150	373	1.259	0.136	0.0914
Overall	3,568	11,756	1.162	0.029	1×10^{-7}

Table E8. Association analyses of the top 11 variants stratified by asthma and hayfever status.

Locus	SNP[†], risk allele	**A+** (N=10,263) **vs** **A-** (N=24,759)		**H+** (N=16,513) **vs** **H-** (N=17,256)		**A+** (N=10,263) **vs** **A-H-** (N=14,091)		**H+** (N=16,513) **vs** **A-H-** (N=14,091)		**A+H-** (N=2,776) **vs** **A-H-** (N=14,091)		**A-H+** (N=9,301) **vs** **A-H-** (N=14,091)	
		OR, 95% CI	*P*-value	OR, 95% CI	*P*-value	OR, 95% CI	*P*-value	OR, 95% CI	*P*-value	OR, 95% CI	*P*-value	OR, 95% CI	*P*-value
HLA-DQB1	rs9273373,G	1.17 (1.12-1.22)	$5x10^{-12}$	1.12 (1.08-1.16)	$2x10^{-08}$	1.22 (1.16-1.28)	$2x10^{-14}$	1.13 (1.09-1.18)	$2x10^{-09}$	1.14 (1.05-1.25)	0.0019	1.07 (1.02-1.12)	0.0052
TLR1	rs4833095,T	1.11 (1.06-1.15)	$7x10^{-07}$	1.14 (1.10-1.18)	$4x10^{-12}$	1.15 (1.10-1.21)	$5x10^{-10}$	1.15 (1.10-1.19)	$2x10^{-12}$	1.08 (1.00-1.16)	0.0467	1.11 (1.07-1.16)	0.0000
WDR36	rs1438673,C	1.11 (1.07-1.15)	$2x10^{-09}$	1.09 (1.06-1.12)	$7x10^{-08}$	1.14 (1.09-1.18)	$7x10^{-11}$	1.11 (1.07-1.14)	$9x10^{-10}$	1.10 (1.04-1.17)	0.0022	1.07 (1.03-1.11)	0.0002
IL1RL1	rs10197862,A	1.16 (1.11-1.22)	$3x10^{-09}$	1.14 (1.10-1.20)	$2x10^{-09}$	1.21 (1.15-1.28)	$2x10^{-11}$	1.16 (1.10-1.21)	$8x10^{-10}$	1.13 (1.03-1.23)	0.0086	1.11 (1.05-1.17)	0.0001
LRRC32	rs2155219,T	1.08 (1.05-1.12)	$4x10^{-06}$	1.10 (1.07-1.14)	$7x10^{-10}$	1.11 (1.07-1.15)	$2x10^{-07}$	1.11 (1.07-1.14)	$5x10^{-10}$	1.00 (0.94-1.06)	0.9839	1.08 (1.04-1.12)	0.0001
GSDMA	rs7212938,G	1.15 (1.11-1.19)	$8x10^{-15}$	1.05 (1.02-1.09)	0.0020	1.16 (1.11-1.20)	$2x10^{-12}$	1.07 (1.03-1.11)	0.0002	1.14 (1.07-1.22)	0.0001	1.02 (0.98-1.06)	0.3297
TSLP	rs1837253,C	1.15 (1.10-1.19)	$4x10^{-11}$	1.06 (1.02-1.10)	0.0022	1.16 (1.11-1.21)	$3x10^{-10}$	1.08 (1.04-1.12)	$9x10^{-05}$	1.15 (1.07-1.24)	0.0002	1.03 (0.98-1.07)	0.2448
IL33	rs72699186,T	1.19 (1.12-1.26)	$1x10^{-08}$	1.13 (1.07-1.19)	$2x10^{-05}$	1.25 (1.16-1.34)	$5x10^{-10}$	1.15 (1.08-1.22)	$2x10^{-06}$	1.19 (1.06-1.34)	0.0028	1.08 (1.02-1.16)	0.0163
ZBTB10	rs7009110,T	1.09 (1.05-1.13)	$5x10^{-06}$	1.08 (1.04-1.11)	$5x10^{-06}$	1.11 (1.06-1.15)	$1x10^{-06}$	1.09 (1.05-1.13)	$1x10^{-06}$	1.05 (0.99-1.12)	0.1054	1.06 (1.02-1.10)	0.0049
SMAD3	rs17294280,G	1.12 (1.07-1.17)	$3x10^{-07}$	1.08 (1.04-1.12)	0.0002	1.13 (1.08-1.19)	$1x10^{-06}$	1.08 (1.04-1.13)	0.0002	1.06 (0.98-1.14)	0.1468	1.03 (0.98-1.08)	0.2882
CLEC16A	rs62026376,C	1.12(1.07-1.17)	$7x10^{-07}$	1.09 (1.04-1.13)	$3x10^{-05}$	1.15 (1.09-1.21)	$4x10^{-08}$	1.10 (1.05-1.14)	$6x10^{-06}$	1.08 (0.99-1.17)	0.0957	1.05 (1.01-1.10)	0.0227

[†] SNPs rs9273373, rs72699186 and rs62026376 were not tested in ALSPAC and Raine, and so results for these three SNPs are based on the AAGC and 23andMe studies. The sample sizes for these three SNPs are N=8,036 for A+, N=20,879 for A-, N=14,391 for H+, N=13,471 for H-, N=11,474 for A-H-, N=1,855 for A+H- and N=8,363 for A-H+.

Table E9. Comparison of risk allele frequency for the top 11 variants between individuals with asthma but not hayfever (A+H-, coded as cases) and individuals with hayfever but not asthma (A-H+, coded as controls).

Locus	SNP[†], risk allele	**A+H-** (N=2,409) **vs** **A-H+** (N=9,169)	
		OR, 95% CI	*P*-value
HLA-DQB1	rs9273373,G	1.09 (1.00-1.20)	0.0560
TLR1	rs4833095,T	0.95 (0.87-1.03)	0.2362
WDR36	rs1438673,C	1.02 (0.95-1.10)	0.5341
IL1RL1	rs10197862,A	1.01 (0.91-1.12)	0.8552
LRRC32	rs2155219,T	0.92 (0.85-0.98)	0.0139
GSDMA	rs7212938,G	1.15 (1.07-1.24)	0.0003
TSLP	rs1837253,C	1.13 (1.04-1.23)	0.0037
IL33	rs72699186,T	1.07 (0.95-1.21)	0.2672
ZBTB10	rs7009110,T	0.97 (0.90-1.04)	0.3740
SMAD3	rs17294280,G	1.00 (0.92-1.10)	0.9329
CLEC16A	rs62026376,C	1.02 (0.93-1.12)	0.6811

[†] The Raine study did not contribute to these analyses. SNPs rs9273373, rs72699186 and rs62026376 were not tested in ALSPAC and Raine, and so results for these three SNPs are based on the AAGC and 23andMe studies. The sample sizes for these three SNPs are N=1,855 for A+H- and N=8,363 for A-H+.

Table E10. Association between the *ZBTB10* and *CLEC16A* loci with asthma risk in the GABRIEL GWAS[E14].

Locus	Sentinel SNP, risk allele	Proxy SNP, risk allele	r^2	Risk alleles in phase?	GABRIEL results	
					OR	*P*-value
ZBTB10	rs7009110, A	rs6473226, T	0.96	yes	1.067	0.0029
ZBTB10	rs7009110, A	rs1543857, G	0.97	yes	1.059	0.0085
CLEC16A	rs62026376, C	rs17673553, A	0.96	yes	1.073	0.0042

Table E11. Variants associated with variation in the numbers of peripheral blood leukocyte populations[E15] or lymphocyte subtypes[E16] and in linkage disequilibrium ($r^2 > 0.3$) with the *CLEC16A* sentinel SNP.

Sentinel SNP	Proxy SNP	r^2	Associated trait	*P*-value
rs62026376	rs9652582	0.43	Eosinophils	0.0029
rs62026376	rs3901386	0.42	CD56+ NK cells	0.0031

Table E12. Variants in the *ZBTB10* or *CLEC16A* regions associated with other inflammatory or immune-related diseases and in linkage disequilibrium ($r^2 > 0.3$) with the asthma-with-hayfever sentinel SNP.

Locus	Sentinel SNP	Proxy SNP	r^2	Trait	*P*-value	Reference
ZBTB10	rs7009110	rs7000782	0.51	Atopic dermatitis	$1x10^{-6}$	E17
CLEC16A	rs62026376	rs12708716	0.55	Type 1 diabetes	$3x10^{-18}$	E18
CLEC16A	rs62026376	rs887864	0.53	Allergic rhinitis	$1x10^{-6}$	E19
CLEC16A	rs62026376	rs7200786	0.31	Multiple sclerosis	$9x10^{-17}$	E20

Table E13. Nineteen variants associated with the risk of having asthma-with-hayfever at the suggestive significance level ($3 \times 10^{-8} < P \leq 5 \times 10^{-6}$).

Chr	Position, bp	Nearest gene, kb distance	SNP, risk allele	Risk allele frequency, range	OR (95% CI)	Association *P*-value	Heterogeneity test *P*-value (I^2, 95% CI)
10	6164627	*RBM17*,6	rs41295115,C	0.05-0.06	1.32 (1.19-1.47)	1×10^{-7}	0.66 (0, 0-89)
5	110182208	*SLC25A46*,56	rs3853750,C	0.15-0.18	1.17 (1.10-1.23)	1×10^{-7}	0.98 (0, 0-85)
12	60536687	*FAM19A2*,148*	rs17605016,G	0.07-0.10	1.22 (1.14-1.32)	2×10^{-7}	0.78 (0, 0-85)
9	6058077	*RANBP*6,52	rs343496,A	0.81-0.82	1.16 (1.10-1.23)	2×10^{-7}	0.52 (0, 0-85)
17	35226234	*IKZF*3,48*	rs12450323,T	0.17-0.19	1.17 (1.10-1.25)	5×10^{-7}	0.78 (0, 0-89)
12	28086392	*PTHLH*,70	rs11049300,A	0.96-0.98	1.43 (1.24-1.65)	8×10^{-7}	0.62 (0, 0-85)
21	36929813	*SIM2*,64	rs6517368,T	0.71-0.74	1.13 (1.08-1.19)	9×10^{-7}	0.59 (0, 0-85)
8	10850884	*XKR6*,60*	rs6982751,C	0.86-0.86	1.19 (1.11-1.28)	1×10^{-6}	0.15 (50, 0-91)
2	218385831	*TNS1*,1*	rs76043829,G	0.87-0.89	1.22 (1.13-1.33)	2×10^{-6}	0.92 (0, 0-89)
14	102308211	*TRAF3*,5	rs8010932,A	0.81-0.84	1.15 (1.09-1.22)	3×10^{-6}	0.20 (34, 0-77)
12	98119078	*FAM71C*,447	rs2712665,T	0.67-0.69	1.12 (1.07-1.17)	3×10^{-6}	0.58 (0, 0-85)
6	148622701	*SASH1*,83	rs4896981,G	0.78-0.79	1.16 (1.09-1.23)	3×10^{-6}	0.55 (0, 0-89)
1	108124058	*VAV3*,185*	rs7521681,A	0.14-0.16	1.15 (1.08-1.22)	3×10^{-6}	0.95 (0, 0-85)
1	197031259	*PTPRC*,38	rs2759643,A	0.67-0.72	1.12 (1.07-1.17)	3×10^{-6}	0.37 (3, 0-85)
11	76047835	*LRRC32*,2*	rs1320644,A	0.33-0.36	1.11 (1.06-1.17)	5×10^{-6}	0.89 (0, 0-85)
11	120711564	*SC5DL*,22	rs2060009,A	0.68-0.72	1.12 (1.07-1.18)	5×10^{-6}	0.35 (7, 0-86)
1	67823218	*GADD45A*,100	rs787538,T	0.59-0.61	1.11 (1.06-1.16)	5×10^{-6}	0.83 (0, 0-85)
5	83942985	*EDIL3*,227	rs72766477,A	0.04-0.04	1.34 (1.18-1.52)	5×10^{-6}	0.73 (0, 0-89)
17	4958496	*USP6*,2	rs9912347,G	0.51-0.53	1.11 (1.06-1.16)	5×10^{-6}	0.52 (0, 0-85)

* SNP is located within reported gene.

Table E14. Replication results for 19 variants associated with the risk of having asthma-with-hayfever at the suggestive significance level ($5x10^{-8} < P \leq 5x10^{-6}$) in the discovery analysis.

Locus	SNP	Allele	Replication analysis			Discovery + replication			
			OR	SE	*P*-value	OR	SE	*P*-value	Direction
SLC25A46	rs3853750	C	1.0562	0.0766	0.4764	1.1533	0.0276	2×10^{-7}	--
PTHLH	rs11049300	G	0.7624	0.1835	0.1322	0.7070	0.0674	3×10^{-7}	++
IKZF3	rs12450323	T	1.0883	0.0757	0.2652	1.1600	0.0292	4×10^{-7}	++
XKR6	rs6982751	G	0.8698	0.0861	0.1022	0.8419	0.0338	4×10^{-7}	++
RBM17	rs41295115	C	1.0563	0.1249	0.6623	1.2769	0.0486	5×10^{-7}	--
EDIL3	rs72766477	A	1.3478	0.1437	0.0406	1.3402	0.0583	5×10^{-7}	++
TNS1	rs76043829	A	0.8684	0.0990	0.1503	0.8242	0.0388	6×10^{-7}	--
VAV3	rs7521681	A	1.1452	0.0770	0.0799	1.1491	0.0280	7×10^{-7}	++
SIM2	rs6517368	C	0.9611	0.0644	0.5376	0.8928	0.0236	2×10^{-6}	++
FAM19A2	rs17605016	G	0.9793	0.0994	0.8328	1.1878	0.0359	2×10^{-6}	-+
RANBP6	rs343496	T	1.0235	0.0721	0.7473	0.8810	0.0268	2×10^{-6}	+-
TRAF3	rs8010932	G	0.9937	0.0790	0.9362	0.8829	0.0283	1×10^{-5}	++
SASH1	rs4896981	A	1.0028	0.0711	0.9691	0.8858	0.0285	2×10^{-5}	-+
FAM71C	rs2712665	C	1.0433	0.0642	0.5100	0.9102	0.0227	3×10^{-5}	+-
GADD45A	rs787538	C	1.0442	0.0585	0.4599	0.9189	0.0212	6×10^{-5}	+-
PTPRC	rs2759643	T	1.0554	0.0605	0.3727	0.9142	0.0225	7×10^{-5}	+-
LRRC32	rs1320644	A	0.9511	0.0612	0.4115	1.0919	0.0222	7×10^{-5}	+-
SC5DL	rs2060009	G	1.0734	0.0628	0.2603	0.9155	0.0230	1×10^{-4}	+-
USP6	rs9912347	A	1.0924	0.0576	0.1251	0.9262	0.0207	2×10^{-4}	-+

E Figure Legends

Figure E1. Quantile-quantile (QQ) plots for the GWAS of asthma-with-hayfever in the four individual cohorts, as well as in the meta-analysis of all studies.

Figure E2. Main association results from the meta-analysis of four GWAS of asthma-with-hayfever.

Figure E3. Regional association results ($-\log_{10}$(P-value), y-axis) for chromosome 8q21. The most-associated SNP is shown in purple, and the color of the remaining markers reflects the linkage disequilibrium (r^2) with the top SNP. The recombination rate (second y-axis) is plotted in light blue and is based on the CEU 1000G population. Plots were generated using LocusZoom.[E21]

Figure E4. Regional association results ($-\log_{10}$(P-value), y-axis) for chromosome 16p13. The most-associated SNP is shown in purple, and the color of the remaining markers reflects the linkage disequilibrium (r^2) with the top SNP. The recombination rate (second y-axis) is plotted in light blue and is based on the CEU 1000G population. Plots were generated using LocusZoom.[E21]

Collaborators of the Australian Asthma Genetics Consortium (AAGC)

Dale R. Nyholt[a], John Beilby[b-d], 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], Mark Jenkins[j], Haydn Walters[f]

[a] Queensland Institute of 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, Lauceston, 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.