HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2014 February 01.

Published in final edited form as:

Nat Genet. 2013 August ; 45(8): 907–911. doi:10.1038/ng.2686.

A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci

David A. Hinds[1], **George McMahon**[2], **Amy K. Kiefer**[1], **Chuong B. Do**[1], **Nicholas Eriksson**[1], **David M. Evans**[2], **Beate St Pourcain**[3], **Susan M. Ring**[3], **Joanna L. Mountain**[1], **Uta Francke**[1], **George Davey-Smith**[2], **Nicholas J. Timpson**[2,*], and **Joyce Y. Tung**[1,*]

[1]23andMe, Inc., Mountain View, CA 94043, USA

[2]Medical Research Council (MRC) Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

[3]School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

Allergic disease is very common and carries substantial public health burdens. We conducted a meta-analysis of genome-wide association with self-reported cat, dust-mite, and pollen allergy in 53,862 individuals. We used generalized estimating equations to model shared and allergy-specific genetic effects. We identified 16 shared susceptibility loci with $P<5\times10^{-8}$, including 8 loci previously associated with asthma, as well as 4p14 near *TLR1*, *TLR6*, and *TLR10* (rs2101521: $P=5.3\times^{-21}$); 6p21.33 near *HLA-C* and *MICA* (rs9266772: $P=3.2\times10^{-12}$); 5p13.1 near *PTGER4* (rs7720838: $P=8.2\times10^{-11}$); 2q33.1 in *PLCL1* (rs10497813: $P=6.1\times10^{-10}$); 3q28 in *LPP* (rs9860547: $P=1.2\times10^{-9}$); 20q13.2 in *NFATC2* (rs6021270: $P=6.9\times10^{-9}$); 4q27 in *ADAD1* (rs17388568: $P=3.9\times10^{-8}$); and 14q21.1 near *FOXA1* and *TTC6* (rs1998359: $P=4.8\times10^{-8}$). We identified one locus with substantial evidence for differences in effects across allergies, at 6p21.32 in the class II HLA region (rs17533090: $P=1.7\times10^{-12}$), strongly associated with cat allergy. Our study sheds new light on the shared etiology of immune and autoimmune disease.

Correspondence should be addressed to: David A. Hinds, 23andMe, Inc, 1390 Shorebird Way, Mountain View, CA 94043, USA, Phone: 650-963-8900, Fax: 650-960-7300, dhinds@23andMe.com.

*Equal contributions

URLs

BEAGLE, http://faculty.washington.edu/browning/beagle/beagle.html; HapMap, www.hapmap.org; MaCH, http://www.sph.umich.edu/csg/abecasis/MaCH/; Minimac, http://genome.sph.umich.edu/wiki/Minimac; NCBI Gap Plus, http://www.ncbi.nlm.nih.gov/projects/gapplus/sgap_plus.htm; 1000 Genomes, http://www.1000genomes.org; UCSC Genome Browser, http://genome.ucsc.edu.

Author Contributions

D.A.H. and G.McM. analyzed the data. A.K.K. designed the survey for 23andMe. G.McM., D.M.E., and B.StP. were part of the ALSPAC GWAS preparation team. S.M.R. was responsible for ALSPAC sample collection and preparation. C.B.D. and N.E. developed analytical tools. J.L.M., U.F., and G.D.-S. supervised the project. D.A.H., G.McM., N.J.T, and J.Y.T. designed the study and wrote the paper.

Competing Financial Interests

D.A.H., A.K.K, C.B.D., N.E., J.L.M., U.F., and J.Y.T. are current or former employees of, and own stock or stock options in, 23andMe, Inc.

Allergies and allergic asthma are among the most common diseases in the industrialized world. In the United States, a nationwide survey showed that over half the population tested positive for sensitization to at least one common allergen, a considerable increase over results collected approximately ten years earlier[1]. The cause of this apparent increase in prevalence is unknown, but the rapidity with which it has occurred implicates an environmental component[2]. Still, estimates of the heritability of allergy are high[3,4], suggesting that understanding the genetic liability underlying these conditions is key to understanding the disease.

A number of genes implicated in allergy and asthma through association and functional studies belong to pathways involved in immune and inflammatory processes, such as innate immunity, adaptive immunity, and allergic inflammation[5]. These genes belong to a range of gene families that encode toll-like receptors, interleukins, chemokines, and various other signaling molecules and transcription factors. Published genome-wide association studies (GWAS) on allergic conditions have focused on asthma and atopic dermatitis, resulting in a substantial number of loci associated with asthma (*HLA-DQ*, *IL33*, *IL18R1*, *SMAD3*, *GSDMA*, *IL2RB*, *RORA*, *GSDMB*, *IL13*, *SLC22A5*, *DENND1B*, *PDE4D*, *ORMDL3*, *IL6R*, 5q22.1, 11q13.5)[6–11] and a smaller number with atopic dermatitis (*FLG*, *OVOL1*, *ACTL9*, 5q22.1, 11q13.5, 20q13.33)[12–14]. Studies using other measures of atopy have been less definitive, likely due to limited sample sizes[15–17]; the largest, of allergic rhinitis and IgE sensitization to grass pollen, identified three regions with genome-wide significance (class II HLA, 5q22.1, 11q13.5).

We selected three common self-reported allergy phenotypes -- pollen allergy, dust mite allergy, and cat allergy -- for which comparable data were available in the 23andMe participant cohort[18] and in a cohort of mothers from the Avon Longitudinal Study of Parents and Children (ALSPAC)[19] (Table 1). We used generalized estimating equations (GEE) to jointly model genetic effects across all three phenotypes. The GEE approach accounts for the correlations between the phenotypes, and enables us to estimate both shared and allergy-specific effects. We first performed a genome-wide meta-analysis of GEE tests for shared effects. Then, for a set of 3725 markers with nominal evidence of association with at least one allergy, we performed GEE tests for allergy-specific effects (Supplementary Table 1).

In the GEE meta-analysis for shared effects across allergies, we identified 16 genome-wide-significant loci with $P<5\times10^{-8}$ (Table 2, Fig. 1, Supplementary Fig. 1). Of these, 8 had $P<5\times10^{-8}$ in the 23andMe cohort and $P<0.05$ in the ALSPAC cohort (Supplementary Tables 2 and 3). We identified 6 loci with suggestive evidence for association ($5\times10^{-8} < P < 1\times10^{-6}$) (Supplementary Note). Many of these loci have previously been associated with other immunity related phenotypes, and 8 have been associated with asthma in previous GWAS (Supplementary Note). While we describe loci by their proximal genes, in most cases we have no functional evidence for a specific target, and these variants may affect regulation of more distant genes.

To ensure that the results were not confounded by age or differences between genotyping platforms, we tested the index SNPs for platform effects and for interactions with age within the 23andMe cohort, but no tests yielded strong evidence for interaction after adjusting for

multiple comparisons (Supplementary Table 4). We also tested for pairwise interactions among the 22 loci, but none were significant after adjustment for multiple comparisons (Supplementary Table 5).

We assessed whether these associations were supported in a companion study of allergic sensitization[20] (Table 3, Supplementary Table 6). All 22 loci had effects in the same direction, and 10 of our 16 genome-wide-significant loci were supported with $P<0.05$ in the sensitization study. We also annotated our findings based on linkage disequilibrium (LD) with results from published GWAS, coding variation, monocyte expression quantitative trait loci (eQTL)[21], and putative regulatory regions identified by the ENCODE project (Table 3; Supplementary Tables 7, 8, 9, and 10).

We examined evidence for association in our meta-analysis at other loci previously associated with either asthma or atopic dermatitis (Supplementary Tables 11 and 12). We have nominal support ($P < 0.05$, consistent risk allele) for 7 additional asthma loci (*IL6R*, *GAB1*, *RAD50*, *IL13*, *IKZF4*, *RORA*, and *IL2RB*), with a false discovery rate (FDR) of 0.04 across these variants. For atopic dermatitis, we have nominal support at 5 additional loci (*IL13*, *KIF3A*, *CARD11*, *MIR1208*, and *NCF4*), with an FDR of 0.07 for this group. These results indicate substantial overlap among these phenotypes beyond the loci meeting our criteria for significant and suggestive associations.

To test for allergy-specific genetic effects, we included interaction terms for specific allergies in our GEE models (Supplementary Table 1). We found one locus with strong evidence of allergy-specific effects, on chromosome 6 in the MHC region spanning the *HLA-DRA*, *-DRB*, *-DQA1*, and *-DQB2* genes (Fig. 2, Supplementary Fig. 2). Index SNP rs17533090 had a combined $P=1.7\times10^{-12}$ for interaction with allergy type. Effects for the three allergies were consistent across cohorts, and indicated that this locus was specifically associated with cat allergy (Fig. 3). Among SNPs in Table 2, only rs2101521 showed evidence for an allergy-specific effect (unadjusted $P=0.0011$), which was weak compared to the evidence for a shared effect ($P=5.3\times10^{-21}$).

We performed an exploratory analysis to assess associations of allergy loci with symptoms of allergic rhinitis, allergic contact dermatitis, and allergic asthma in the 23andMe cohort. We reclassified cases based on reported symptoms, and used the GEE approach to jointly model genetic effects across symptoms (Supplementary Tables 13 and 14). All effects were in the same direction at all index SNPs. At most loci, we did not see evidence for differential effects ($P>0.05$ for interaction). The exception was at *GSDMB* rs9303280, which was most strongly associated with asthma ($P=0.000035$ for interaction). Effect sizes for contact dermatitis symptoms tended to be smaller than for asthma (20/23, $P=0.0004$) or rhinitis (18/23, $P=0.007$). Effect sizes for asthma tended to be larger than for rhinitis, but not significantly so (14/23, $P=0.30$). However, association tests for rhinitis were more significant than for asthma at most loci (18/23, $P=0.007$), often by several orders of magnitude. Thus, while asthma may be a slightly more specific atopy phenotype, rhinitis appears to be more powerful for discovery of atopy loci in cohort studies, because it is more sensitive and captures more individuals who report allergies.

Genes implicated in our GWAS highlight key pathways in the etiology of common allergy. In the 4p14 region near rs2101521, *TLR1* (Toll-Like Receptor 1) and *TLR6* (Toll-Like Receptor 6) encode pattern-recognition receptors whose role in recognizing external pathogens and activating appropriate immune responses lies at the interface between innate immunity and immunoregulation. Candidate gene studies have identified associations between TLRs and asthma[22–24], and with grass sensitization and rhinitis[17]. However, this region has not been reported as significant in a genome-wide analysis.

We see substantial overlap between loci associated with allergy and loci previously linked to autoimmune disease. In the 5p13.1 region, index SNP rs7720838 is upstream of *PTGER4*, or Prostaglandin E receptor 4, previously implicated as a candidate asthma locus[25]. This SNP is close to a reported association with ankylosing spondylitis (rs10440635, r^2=0.94)[26]. Variants affecting *PTGER4* expression have also been associated with Crohn's disease[27], and mouse studies point to a role in initiating skin immune responses[28,29]. In the 2q33.1 region, our eQTL analysis suggests that index SNP rs10497813 is associated with expression of *PLCL1*, or phospholipase C-like 1, involved in inositol 1,4,5-triphosphate intracellular signaling[30]. Variation in *PLCL1* (rs6738825, r^2=0.97) has also been associated with Crohn's disease[31].

Several novel allergy loci are in or near genes involved in T helper cell differentiation. Index SNP rs9860547 in the 3q28 region falls in the *LPP* gene (lipoma-preferred partner). A nearby variant in *LPP* (rs1464510, r^2=0.70) has been associated with celiac disease[32,33] and vitiligo[34]. Our eQTL analysis suggests that our association may be mediated by an effect on expression of *BCL6* (B-cell lymphoma 6), a transcription factor that represses STAT6-mediated response to IL-4 and IL-13, and IgE class switching[35], and inhibits type 2 T helper (Th2) cell differentiation in a mouse model[36]. In the 20q13.2 region, index SNP rs6021270 is in the *NFATC2* gene, encoding a component of the NFAT (nuclear factor of activated T cells) transcription complex, which plays an important role in regulating Th cell differentiation[37]. Variation in *NFATC2* has not been linked to any allergic or autoimmune phenotype, however, mice lacking NFATc2 show increased lung inflammation in experimentally induced allergic asthma[38,39]. In the 4q27 region, index SNP rs17388568 falls in the *ADAD1* gene but evidence for association spans the nearby *IL2* and *IL21* genes. This same SNP has been associated with type I diabetes autoantibodies[40] and ulcerative colitis[41], and a nearby SNP in strong LD (rs2069772, r^2=0.91) has been suggestively associated with allergic rhinitis[17]. IL-2 and IL-21 cytokines are involved in regulation of multiple Th cell types; IL-21 is up-regulated in Th2 and Th17 cells and inhibits IL-2, while IL-2 is required for Th1 differentiation and inhibits differentiation of Th17 cells[42].

In the 14q21.1 region, index SNP rs1998359 is upstream of *FOXA1*, a member of the forkhead box transcription factor family. The closely related FOXA2 and FOXA3 transcription factors have roles in regulation of Th2 mediated inflammation and mucus production in allergic airway disease[43]; while a similar role of FOXA1 has not been established, FOXA1 and FOXA2 are known to have overlapping patterns of expression in respiratory epithelium[44].

In the 6p21.33 region, HLA-B and HLA-C are major histocompatibility complex (MHC) class I molecules expressed on most cell types, responsible for the display of intracellular peptides to T cells. MICA belongs to a family of non-classical MHC molecules that resemble the class I molecules and are thought to be involved in innate antitumor and antiviral surveillance[45]. Alleles of *HLA-B* are associated with severe allergic reactions such as abacavir hypersensitivity and Stevens-Johnsons syndrome[46,47]. SNPs in these three genes have been associated with a number of immune-system-related phenotypes such as psoriasis and HIV-1 control[48,49].

Previous studies have suggested associations between specific allergen sensitivities and HLA class II alleles[50]. However, these studies have been small and have reported inconsistent results[51]. Our finding of a specific association with cat allergy is the first demonstration of allergen specificity in a GWAS context.

We assessed directionality of effects in cases where our index SNPs are in strong LD ($r^2 > 0.5$) with SNPs previously associated with autoimmune disease (Supplementary Table 15). At some loci (*LRRC32*, *PTGER4*, *PLCL1*, *SMAD3*, *ADAD1*, *CLEC16A*), autoimmune disease and allergy are associated with the same risk alleles. At others (*GSDMB*, *LPP*), the risk allele for autoimmune disease appears to be protective for allergy. Many autoimmune diseases are associated with increased activation of type 1 T helper (Th1) responses, while allergy has been associated with Th2 activity[52]. Our results may help to identify elements that influence the balance of Th1 versus Th2 activity, versus elements that contribute to both responses.

Self reported allergy status can be unreliable[53], and the surveys we used were not standardized or validated. In the 23andMe cohort, the high proportion of allergy cases likely reflects responder bias in completing the allergy survey. The ALSPAC cohort was assessed during pregnancy, which can alter allergic disease status[54]. These limitations should not compromise the validity of our genetic associations, but they make functional interpretation more challenging.

Our results demonstrate that self reported allergy can be used to identify disease susceptibility loci, with results consistent with studies of more narrowly defined allergy manifestations and allergic sensitization. Self-directed web-based data collection in the 23andMe cohort yielded results largely consistent with traditional survey methods used in the ALSPAC cohort. Our findings reinforce and extend evidence for a shared genetic etiology of allergic and autoimmune disease, with novel allergy susceptibility loci near *LPP*/*BCL6*, *HLA-C*/*MICA*, *PTGER4*, and *PLCL1*, all previously associated with autoimmune disease. Our findings also highlight the role of the Th2 cell lineage in pathogenesis of allergy, with associations in or near key Th2 genes including *ID2*, *BCL6*, *GATA3*, *IL13*, *IL33*, *TSLP*, and *IL1RL1*. An important next step will be to more carefully characterize the extent to which individual associations lead to a global predisposition to allergy, versus effects on specific targets such as skin, lung, or mucosa.

Methods

23andMe Cohort

Participants in the 23andMe cohort were customers of 23andMe, Inc., a personal genetics company, who had been genotyped as part of the 23andMe Personal Genome Service®. Individuals 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[55]. A maximal set of unrelated individuals was chosen for the analysis using a segmental identity-by-descent (IBD) estimation algorithm[56]. 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 identical-by-descent. This level of relatedness (roughly 20% of the genome) corresponds approximately to the minimal expected sharing between first cousins in an outbred population. The study protocol and consent form were approved by the external AAHRPP-accredited Institutional Review Board, Ethical & Independent Review Services (E&I Review). For a small number of participants (167) under the age of 18, consent was provided by a parent, guardian, or legally authorized adult.

DNA extraction and genotyping were performed on saliva samples by 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 on the Illumina HumanHap550+ BeadChip platform, 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[18]. The remaining 65% of participants were genotyped on the Illumina HumanOmniExpress+ Bead Chip. This platform has a base set of 730,000 SNPs. This was 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. Individuals whose analyses failed repeatedly were re-contacted by 23andMe customer service to provide additional samples, as is done for all 23andMe customers.

Participant genotype data were imputed against the August 2010 release of 1000 Genomes reference haplotypes[57]. First, we used Beagle[58] (version 3.3.1) to phase batches of 8000–9000 individuals across chromosomal segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs. We excluded SNPs with minor allele frequency < 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[59] (2011-10-27), 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.

23andMe participants were able to fill out web-based questionnaires whenever they logged into their 23andMe accounts. Allergy information was derived primarily from an “Allergies and Asthma” survey (Supplementary Note). The survey covers allergic reactions to 38

common allergens, including foods, plants, animals, molds, latex, dust mites, medicines, and vaccines. Cases were defined as those who reported a positive allergy test, difficulty swallowing or speaking, hives, itchy mouth, itchy eyes, itchy nose, or asthma in response to that allergen. Controls were defined as individuals who did not meet these criteria. For pollen allergy, we aggregated reports of allergies to grasses, trees, or weeds. At the time of the analysis, 30% of 171,274 23andMe research participants had taken this allergy survey. We incorporated 7635 additional controls who reported having neither seasonal nor environmental allergies in a medical history survey, or who reported not currently having allergies in an asthma survey. The final analysis included 46,646 participants.

ALSPAC Cohort

The Avon Longitudinal Study of Parents and Children[19] is a large birth cohort that has recruited 14,541 pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992. Mothers enrolled in the study, filled out a questionnaire at the end of the third month of pregnancy, which included questions on allergies. Mothers were asked if they were allergic to cat, pollen or dust with the option each time of indicating yes or no. The questions did not specify current or past allergy.

Centre National de Génotypage (CNG) carried out DNA genotyping on the Illumina human660W-quad array and genotypes were called with Illumina GenomeStudio. PLINK[60] (v1.07) was used to carry out quality control measures on an initial set of 10,015 subjects and 557,124 directly genotyped SNPs. SNPs were removed if they displayed more than 5% missingness or a Hardy-Weinberg equilibrium $P<10^{-6}$. Additionally SNPs with a minor allele frequency of less than 1% were removed. Samples were excluded if they displayed more than 5% missingness, had indeterminate X chromosome heterozygosity or extreme autosomal heterozygosity. We restricted the analysis to individuals with European ancestry; samples showing evidence of population stratification were identified by multidimensional scaling of genome-wide identity-by-state pairwise distances using the four HapMap populations as a reference, and then excluded. Cryptic relatedness was assessed using a Pi hat of more than 0.125 which is expected to correspond to roughly 12.5% alleles shared IBD or a relatedness at the first cousin level.

A total of 8,340 subjects and 526,688 SNPs passed these quality control filters. We imputed autosomal SNPs against the HapMap[61] CEU population (release 22) using MaCH[62] (v1.0.16, Li 2010). A combination of MaCH and Minimac[59] (v4.4.3, 2010-12-13) was used to impute X chromosome genotypes against the HapMap CEU population (release 21). Analyses were limited to 2.5 million SNPs with imputed $r^2 > 0.3$. Out of 8,340 subjects with genotype data, 7,216 had allergy phenotype data and were used in the GWAS.

Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and the Local Research Ethics Committees.

Single Phenotype GWAS

We performed traditional genome-wide tests for association with each of the three allergy phenotypes, using logistic regression, assuming an additive model for genetic effects. The

23andMe analyses were adjusted for age, gender, and the top 5 principal components of the genotype data matrix. The ALSPAC analyses were not adjusted for covariates. ALSPAC GWAS results were remapped to NCBI Build 37 using the liftOver tool[63]. For each allergy phenotype, we used METAL[64] to perform an inverse-variance-weighted fixed-effects meta-analysis across 2.4 million SNPs in the intersection of the 23andMe and ALSPAC results. We applied genomic control corrections to the individual GWAS result sets (23andMe: λ=1.06 to 1.08; ALSPAC: λ=1.00 to 1.02). The meta-analysis results showed no inflation (λ=1.001 to 1.004).

Multiple Phenotype GWAS

We jointly modeled association across the three allergens using generalized estimating equations (GEE)[65]. We used an unstructured correlation matrix for the three outcomes. In each cohort, we first fit GEE models with the same covariates used in the single phenotype GWAS, with additional terms for interactions between each covariate and allergen, and a single shared genotype effect, using a fast approximate method[66]. Results were adjusted for genomic control (23andMe: λ=1.07; ALSPAC: λ=1.02). We performed an inverse-variance-weighted fixed-effects meta-analysis of the shared effects across the 23andMe and ALSPAC cohorts. Then, for a subset of 3725 SNPs with either a single-phenotype $P<10^{-4}$ with any allergy, or $P<10^{-4}$ in the approximate GEE meta-analysis, we refit GEE models using the R package *geepack*[67]. In addition to refitting the shared effects, we incorporated interactions between genotype and allergy type, and used analysis of deviance to assess significance of the interactions. We used Fisher's method[68] to compute combined *P* values from the 23andMe and ALSPAC interaction tests. This test combines evidence for an interaction in each cohort but does not assess directional consistency of the interactions.

Heterogeneity Assessment

In the 23andMe cohort, we assessed genotyping platform effects by logistic regression of platform against 5 principal components and the imputed allele dosage, and performed a likelihood ratio test to assess significance of the allele dosage term. We assessed the index SNPs for age effects in the 23andMe cohort by fitting a GEE model with an age by dosage interaction, and testing significance with a Wald test on the interaction term. We tested index SNPs for heterogeneity across cohorts using Cochran's *Q* statistic, and used I^2 to measure the extent of heterogeneity[69]. We determined confidence intervals for I^2 using the non-central χ^2 method (Supplementary Table 4). While several SNPs have large I^2, confidence intervals are very wide and remain consistent with the null hypothesis of no heterogeneity.

Assessment of SNP Interactions

In the 23andMe cohort, we fit GEE models assuming shared effects across allergy types with allele dosages and interactions for all pairwise combinations of the 22 shared-effect index SNPs and rs17533090 (Supplementary Table 5). We used Wald tests to assess significance of the interaction terms. Given a conditioning SNP1 and tested SNP2, we also computed a joint test of both the main effect of SNP2 and the interaction SNP1×SNP2 being equal to zero.

Functional Annotation

We used publicly available bioinformatic resources to annotate putative associations. Generally, we required that an annotated variant be within 500 kb of our index SNP, with $r^2>0.5$ based on the European subset of 1000 Genomes haplotypes. We used the NCBI Gap Plus resource to identify nearby GWAS findings (Supplementary Table 7). We used tables from the UCSC Genome Browser to identify nearby coding SNPs (Supplementary Table 8). We identified nearby expression quantitative trait loci (eQTL) from a study of monocytes[21] (Supplementary Table 9). We also used HaploReg[70] to identify nearby annotations from the ENCODE project[71] (Supplementary Table 10). Finally, we took all reported associations with asthma or atopic dermatitis from the NHGRI GWAS Catalog[72], and looked up our corresponding meta-analysis results (Supplementary Tables 11 and 12).

Assessment of SNP Effects on Allergy Symptoms

In the 23andMe cohort, we reclassified cases based on self report of allergic symptoms representing allergic rhinitis, allergic asthma, and/or allergic contact dermatitis (Supplementary Methods). We performed a GEE analysis across these multiple outcomes, including the same controls used in the GWAS (Supplementary Tables 13 and 14). The model included the same covariates used in the GWAS (age, gender, 5 principal components), as well as interactions between these covariates and symptom type.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the customers of 23andMe who answered surveys, as well as the employees of 23andMe, who together made this research possible. We also thank Kari Nadeau for assistance in survey development, and Stan Nelson and Russ Altman for reviewing surveys. This work was supported in part by the National Heart Lung and Blood Institute of the National Institutes of Health under grant number 1R43HL115873-01.

We are extremely grateful to all the families who took part in ALSPAC, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council (Grant ref: 74882) the Wellcome Trust (Grant ref: 092731) and the University of Bristol provide core support for ALSPAC.

This publication is the work of the authors and D.A.H, N.J.T. and J.Y.T. will serve as guarantors for the contents of this paper.

References

1. Arbes S Jr, Gergen P, Elliott L, Zeldin D. Prevalences of positive skin test responses to 10 common allergens in the US population: results from the third National Health and Nutrition Examination Survey. J Allergy Clin Immunol. 2005; 116:377–83. [PubMed: 16083793]
2. Strachan DP. Family size, infection, and atopy: the first decade of the hygiene hypthesis. Thorax. 2000; 55(Suppl 1):S2–S10. [PubMed: 10943631]
3. Fagnani C, et al. Heritability and shared genetic effects of asthma and hay fever: an Italian study of young twins. Twin Res Hum Genet. 2008; 11:121–131.10.1375/twin.11.2.121 [PubMed: 18361712]

4. Willemsen G, van Beijsterveldt TCEM, van Baal CGCM, Postma D, Boomsma DI. Heritability of self-reported asthma and allergy: a study in adult Dutch twins, siblings and parents. Twin Res Hum Genet. 2008; 11:132–142.10.1375/twin.11.2.132 [PubMed: 18361713]
5. Vercelli D. Discovering susceptibility genes for asthma and allergy. Nat Rev Immunol. 2008; 8:169–82. nri2257. [PubMed: 18301422]
6. Himes BE, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. Am J Hum Genet. 2009; 84:581–593. [PubMed: 19426955]
7. Sleiman PMA, et al. Variants of DENND1B associated with asthma in children. N Engl J Med. 2010; 362:36–44. [PubMed: 20032318]
8. Moffatt MF, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010; 363:1211–1221. [PubMed: 20860503]
9. Hirota T, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. Nat Genet. 2011; 43:893–896. [PubMed: 21804548]
10. Torgerson DG, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet. 2011; 43:887–892. [PubMed: 21804549]
11. Ferreira MAR, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet. 2011; 378:1006–1014. [PubMed: 21907864]
12. Esparza-Gordillo J, et al. A common variant on chromosome 11q13 is associated with atopic dermatitis. Nat Genet. 2009; 41:596–601. [PubMed: 19349984]
13. Sun LD, et al. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. Nat Genet. 2011; 43:690–694. [PubMed: 21666691]
14. Paternoster L, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2012; 44:187–192. [PubMed: 22197932]
15. Andiappan AK, et al. Genome-wide association study for atopy and allergic rhinitis in a Singapore Chinese population. PLoS One. 2011; 6:e19719. [PubMed: 21625490]
16. Wan YI, et al. A genome-wide association study to identify genetic determinants of atopy in subjects from the United Kingdom. J Allergy Clin Immunol. 2011; 127:223–231.e3. [PubMed: 21094521]
17. Ramasamy A, 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. [PubMed: 22036096]
18. Eriksson N, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet. 2010; 6:e1000993. [PubMed: 20585627]
19. Fraser A, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 201210.1093/ije/dys066
20. Bønnelykke K, et al. Meta-analysis of genome-wide association studies identifies 10 loci influencing allergic sensitization. Nat Genet. in press.
21. Zeller T, et al. Genetics and beyond - the transcriptome of human monocytes and disease susceptibility. PLoS One. 2010; 5:e10693. [PubMed: 20502693]
22. Eder W, et al. Toll-like receptor 2 as a major gene for asthma in children of European farmers. J Allergy Clin Immunol. 2004; 113:482–488. [PubMed: 15007351]
23. Tantisira K, et al. Toll-like receptor 6 gene (TLR6): single-nucleotide polymorphism frequencies and preliminary association with the diagnosis of asthma. Genes Immun. 2004; 5:343–346. [PubMed: 15266299]
24. Kormann MS, et al. Toll-like receptor heterodimer variants protect from childhood asthma. J Allergy Clin Immunol. 2008; 122:86–92. [PubMed: 18547625]
25. Kurz T, et al. Fine mapping and positional candidate studies on chromosome 5p13 identify multiple asthma susceptibility loci. J Allergy Clin Immunol. 2006; 118:396–402. [PubMed: 16890764]
26. Evans DM, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011; 43:761–767. [PubMed: 21743469]

27. Libioulle C, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet. 2007; 3:e58. [PubMed: 17447842]
28. Kabashima K, et al. Prostaglandin E2-EP4 signaling initiates skin immune responses by promoting migration and maturation of Langerhans cells. Nat Med. 2003; 9:744–749. [PubMed: 12740571]
29. Yao C, et al. Prostaglandin E2-EP4 signaling promotes immune inflammation through Th1 cell differentiation and Th17 cell expansion. Nat Med. 2009; 15:633–640. [PubMed: 19465928]
30. Lin X, et al. Control of calcium signal propagation to the mitochondria by inositol 1,4,5-trisphosphate-binding proteins. J Biol Chem. 2005; 280:12820–12832. [PubMed: 15644334]
31. Franke A, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010; 42:1118–1125. [PubMed: 21102463]
32. Hunt KA, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet. 2008; 40:395–402. [PubMed: 18311140]
33. Dubois PC, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010; 42:295–302. [PubMed: 20190752]
34. Jin Y, et al. Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo. N Engl J Med. 2010; 362:1686–1697. [PubMed: 20410501]
35. Harris MB, et al. Transcriptional repression of Stat6-dependent interleukin-4-induced genes by BCL-6: specific regulation of iepsilon transcription and immunoglobulin E switching. Mol Cell Biol. 1999; 19:7264–7275. [PubMed: 10490661]
36. Kusam S, Toney LM, Sato H, Dent AL. Inhibition of Th2 differentiation and GATA-3 expression by BCL-6. J Immunol. 2003; 170:2435–2441. [PubMed: 12594267]
37. Hermann-Kleiter N, Baier G. NFAT pulls the strings during CD4+ T helper cell effector functions. Blood. 2010; 115:2989–2997. [PubMed: 20103781]
38. Karwot R, et al. Protective role of nuclear factor of activated T cells 2 in CD8+ long-lived memory T cells in an allergy model. J Allergy Clin Immunol. 2008; 121:992–999. [PubMed: 18329088]
39. Karwot R, Ubel C, Bopp T, Schmitt E, Finotto S. Increased immunosuppressive function of CD4(+)CD25(+)Foxp3(+)GITR+ T regulatory cells from NFATc2(−/−) mice controls allergen-induced experimental asthma. Immunobiol. 2012; 217:905–911.
40. Plagnol V, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. PLoS Genet. 2011; 8:e1002216. [PubMed: 21829393]
41. Anderson CA, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet. 2011; 43:246–252. [PubMed: 21297633]
42. Liao W, Lin JX, Wang L, Li P, Leonard WJ. Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. Nat Immunol. 2011; 12:551–559. [PubMed: 21516110]
43. Park SW, et al. Distinct roles of FOXA2 and FOXA3 in allergic airway disease and asthma. Am J Respir Crit Care Med. 2009; 180:603–610. [PubMed: 19628779]
44. Wan H, et al. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. J Biol Chem. 2005; 280:13809–13816. [PubMed: 15668254]
45. Bauer S, et al. Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science. 1999; 285:727–729. [PubMed: 10426993]
46. Martin AM, et al. Predisposition to abacavir hypersensitivity conferred by HLA-B*5701 and a haplotypic Hsp70-Hom variant. Proc Natl Acad Sci USA. 2004; 101:4180–4185. [PubMed: 15024131]
47. Ferrell PB Jr, McLeod HL. Carbamazepine, HLA-B*1502 and risk of Stevens-Johnson syndrome and doxic epidermal necrolysis: US FDA recommendations. Pharmacogenomics. 2008; 9:1543–1546. [PubMed: 18855540]
48. Strange A, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet. 2010; 42:985–990. [PubMed: 20953190]
49. Pereyra F, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010; 330:1551–1557. [PubMed: 21051598] Marsh, DG., et al. HLA and specific immune responsiveness to allergens. In: Tsuji, K.; Alzawa, M.; Sazazuki, T., editors. HLA

1991: Proceedings of the Eleventh International Histocompatibility Workshop and Conference. Vol. 1. Oxford University Press; Oxford: 1992. p. 765-767.

50. Young RP, et al. HLA-DR and HLA-DP genotypes and immunoglobulin E responses to common major allergens. Clin Exp Allergy. 1994; 24:431–439. [PubMed: 8087654]
51. Howell WM, Holgate ST. HLA genetics and allergic disease. Thorax. 1995; 50:815–818. [PubMed: 7570428]
52. Romagnani S. The Th1/Th2 paradigm. Immunol Today. 1997; 18:263–266. [PubMed: 9190109]
53. Smith HE, et al. Is structured allergy history sufficient when assessing patients with asthma and rhinitis in general practice? J Allergy Clin Immunol. 2009; 123:646–650. [PubMed: 19135237]
54. Schatz M, Zeiger RS. Asthma and allergy in pregnancy. Clin Perinatal. 1997; 24:407–432.
55. Falush D, Stephens M, Pritchard J. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003; 164:1567–1587. [PubMed: 12930761]
56. Henn BM, et al. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. PLoS One. 2012; 7:e34267. [PubMed: 22509285]
57. Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–1073. [PubMed: 20981092]
58. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. Am J Hum Genet. 2007; 81:1084–1097. [PubMed: 17924348]
59. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. NatGenet. 2012; 44:955–959.
60. Purcell S, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
61. Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449:851–861. [PubMed: 17943122]
62. 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; 34:816–834. [PubMed: 21058334]
63. Hinrichs AS, et al. The UCSC Genome Browser Database: update 2006. Nuc Acids Res. 2006; 34:D590–D598.
64. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
65. Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.
66. Voorman A, Rice K, Lumley T. Fast computation for genome-wide association studies using boosted one-step statistics. Bioinformatics. 2012; 28:1818–1822. [PubMed: 22592383]
67. Halekoh U, Hojsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Soft. 2006; 15:1–11.
68. Fisher, RA. Statistical methods for research workers. Oliver and Boyd; Edinburgh: 1932.
69. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002; 21:1539–1558. [PubMed: 12111919]
70. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nuc Acids Res. 2012; 40:D930–D934.
71. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
72. Hindorff LA, et al. Potential etiologic and functional implications of genome wide association loci for human diseases and traits. Proc Natl Acad Sci USA. 2009; 106:9362–9367. [PubMed: 19474294]

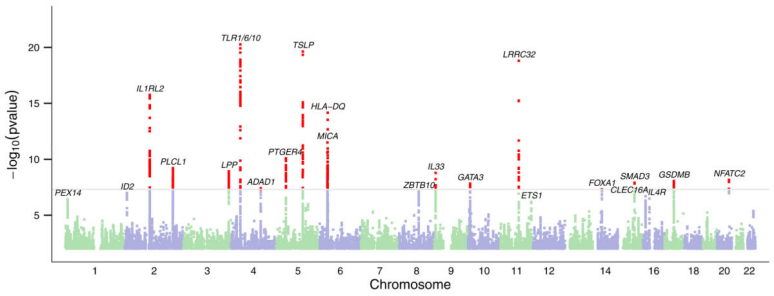

Figure 1.

Manhattan plot of meta-analysis results for shared effects. The plotted values represent the most-significant scores from the meta-analyses of cat, pollen, and dust mite allergy, with all results with $P<10^{-4}$ recomputed using generalized estimating equations to assess effects shared across allergens. Results with $P<5\times10^{-8}$ are shown in red. Gene labels are provided for cross referencing with other results and are not intended to suggest that we have established a causal basis for the observed associations.

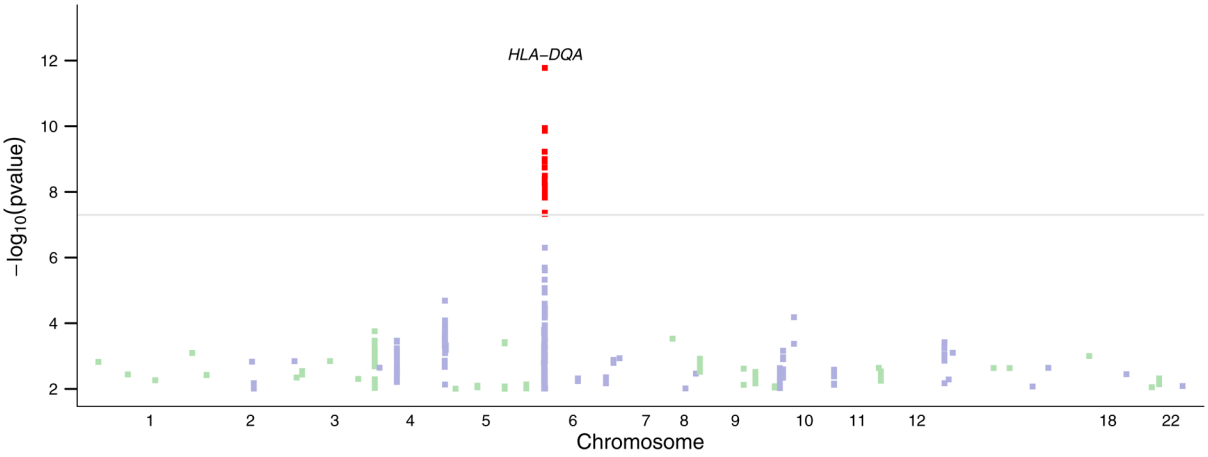

Figure 2.

Manhattan plot of meta-analysis results for interactions with allergen. Results with $P<5\times10^{-8}$ are shown in red. Interaction tests were performed for markers with $P<$1e-4 for association with at least one of cat, pollen, or dust mite allergy.

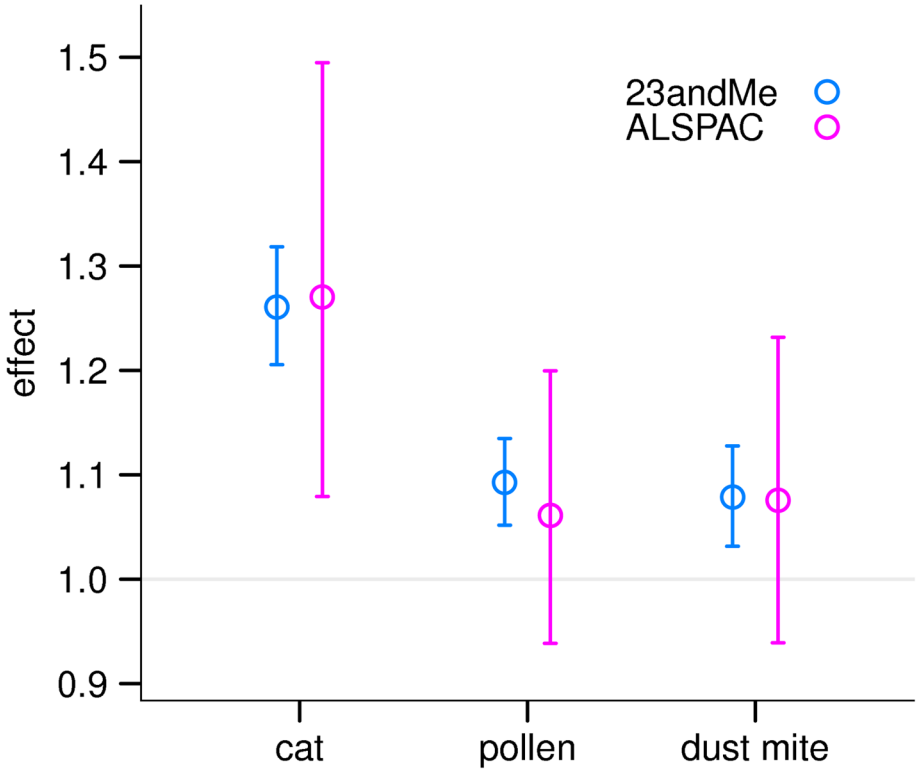

Figure 3.

Marginal effect sizes and 95% confidence intervals for rs17533090 for cat, pollen, and dust mite allergy, in the 23andMe and ALSPAC cohorts. Effects are odds ratios for the high risk G allele of rs17533090.

Table 1

Demographic characteristics of cohorts.

	23andMe		ALSPAC	
	N	%	N	%
Total	46646	100.0	7216	100.0
Gender				
Male	26344	56.5	0	0.0
Female	20302	43.5	7216	100.0
Age				
age <= 30	4300	14.6	4829	67.0
30 < age <= 45	8088	31.1	2382	33.0
45 < age <= 60	6282	25.6	0	0.0
60 < age	6428	28.7	0	0.0
Allergy status				
cat allergy	10509	22.5	704	9.8
dust mite allergy	9815	21.0	964	13.4
pollen allergy	16133	34.6	1201	16.6
Number of allergies				
three allergies	4947	10.6	328	4.6
any two allergies	6228	13.3	536	7.4
any one allergy	9160	19.6	813	11.3
no allergies	26311	56.4	5539	76.8

Table 2

Index SNPs from meta-analysis for shared effects.

SNP	region	position	alleles	RAF	OR	95% CI	*P*	gene context
rs2101521	4p14	38811551	A/G	0.766	1.15	[1.11,1.18]	5.3×10^{-21}	*TLR1*-[]--*TLR6*
rs1438673	5q22.1	110467499	T/C	0.498	1.12	[1.09,1.14]	2.3×10^{-20}	*WDR36*-[]--*CAMK4*
rs2155219	11q13.5	76299194	G/T	0.511	1.11	[1.09,1.14]	1.6×10^{-19}	*C11orf30*--[]--*LRRC32*
rs10189629	2q12.1	102879464	A/C	0.857	1.16	[1.12,1.20]	1.8×10^{-16}	*IL1RL2*--[]--*IL1RL1*
rs6906021	6p21.32	32626311	T/C	0.475	1.10	[1.07,1.13]	7.1×10^{-15}	*HLA-DQA1*--[]-*HLA-DQB1*
rs9266772	6p21.33	31352113	T/C	0.193	1.11	[1.08,1.14]	3.2×10^{-12}	*HLA-C*[]--*MICA*
rs7720838	5p13.1	40486896	G/T	0.580	1.08	[1.06,1.11]	8.2×10^{-11}	[]*PTGER4*
rs10497813	2q33.1	198914072	T/G	0.483	1.08	[1.05,1.10]	6.1×10^{-10}	[*PLCL1*]
rs9860547	3q28	188128979	G/A	0.462	1.08	[1.05,1.10]	1.2×10^{-9}	[*LPP*]
rs7032572	9p24.1	6172380	A/G	0.167	1.12	[1.08,1.16]	1.7×10^{-9}	*RANBP6*[]--*IL33*
rs6021270	20q13.2	50141264	C/T	0.939	1.16	[1.11,1.23]	6.9×10^{-9}	[*NFATC2*]
rs9303280	17q12	38074031	T/C	0.517	1.07	[1.05,1.10]	8.9×10^{-9}	[*GSDMB*]
rs17228058	15q22.33	67450305	A/G	0.240	1.08	[1.05,1.11]	1.2×10^{-8}	[*SMAD3*]
rs962993	10p14	9053132	T/C	0.576	1.07	[1.05,1.10]	1.5×10^{-8}	*GATA3*[]
rs17388568	4q27	123329362	G/A	0.275	1.08	[1.05,1.10]	3.9×10^{-8}	[*ADAD1*]
rs1998359	14q21.1	38077148	C/G	0.246	1.08	[1.05,1.12]	4.8×10^{-8}	*FOXA1*--[]*TTC6*
rs6473223	8q21.13	81268155	C/T	0.358	1.07	[1.04,1.10]	7.7×10^{-8}	*TPD52*[]*ZBTB10*
rs10174949	2p25.1	8442248	A/G	0.724	1.07	[1.05,1.10]	1.0×10^{-7}	[]*ID2*
rs7203459	16p13.13	11230703	C/T	0.734	1.07	[1.04,1.10]	2.0×10^{-7}	[*CLEC16A*]
rs2107357	16p12.1	27410829	G/A	0.138	1.09	[1.06,1.13]	3.3×10^{-7}	*IL4R*--[]--*IL21R*
rs2056417	1p36.22	10581658	A/G	0.694	1.07	[1.04,1.10]	3.7×10^{-7}	[*PEX14*]
rs10893845	11q24.3	128186882	T/G	0.493	1.06	[1.04,1.09]	6.4×10^{-7}	[]*ETS1*

Region, cytogenetic band; position, build 37 map position of SNP; alleles, low/high risk alleles on genomic reference strand; RAF, risk allele frequency across all study participants; OR, meta-analysis odds ratio for the risk allele; CI, confidence interval; gene context, gene(s) spanning or flanking (<1mb) the index SNP; brackets indicate the position of the SNP and dashes indicate distance to a flanking gene (‘-‘, >1kb; ‘--‘, >10kb; ‘---‘, >100kb).

Table 3

Summary of supporting evidence for allergy loci.

SNP	region	repl[a]	atopy[b]	auto[c]	nsSNP[d]	eQTL[e]	gene context[f]
rs2101521	4p14	***			*	*TLR6*	*TLR1*-[]--*TLR6*
rs1438673	5q22.1	***	**				*WDR36*-[]--*CAMK4*
rs2155219	11q13.5	***	***	**			*C11orf30*--[]--*LRRC32*
rs10189629	2q12.1	**	*				*IL1RL2*--[]--*IL1RL1*
rs6906021	6p21.32	***					*HLA-DQA1*--[]-*HLA-DQB1*
rs9266772	6p21.33	**					*HLA-C*[]--*MICA*
rs7720838	5p13.1	*		*			[]*PTGER4*
rs2117339	2q33.1			*		*PLCL1*	[*PLCL1*]
rs9860547	3q28	***		***		*BCL6*	[*LPP*]
rs7032572	9p24.1		**				*RANBP6*[]--*IL33*
rs6021270	20q13.2						[*NFATC2*]
rs9303280	17q12		***	***	*	*IKZF3*	[*GSDMB*]
rs17293632	15q22.33			*			[*SMAD3*]
rs962993	10p14	*					*GATA3*[]
rs17388568	4q27	***	*	***			[*ADAD1*]
rs9671863	14q21.1						*FOXA1*--[]*TTC6*
rs2202749	8q21.13	***					*TPD52*[]*ZBTB10*
rs13416555	2p25.1	***				*ID2*	[]*ID2*
rs7203459	16p13.13		*	*			[*CLEC16A*]
rs2107357	16p12.1						*IL4R*--[]--*IL21R*
rs2056417	1p36.22					*PEX14*	[*PEX14*]
rs970924	11q24.3						[]*ETS1*

[a] Strength of replication for allergic sensitization[20]:

* *P*<0.05,

** *P*<0.005,

*** *P*<0.0005 (Supplementary Table 4).

[b] Nearby (<500kb, r^2>0.5) GWAS findings for atopy phenotypes (Supplementary Table 7).

[c] Nearby (<500kb, r^2>0.5) GWAS findings for autoimmune disease phenotypes (Supplementary Table 7).

[d] nsSNP: * r^2>0.5 with nonsynonymous SNP (Supplementary Table 8).

[e] Association with expression of gene listed (Supplementary Table 9).

[f] As defined in Table 2.