

# Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization

Klaus Bønnelykke<sup>1,54</sup>, Melanie C Matheson<sup>2,54</sup>, Tune H Pers<sup>3-6,54</sup>, Raquel Granell<sup>7</sup>, David P Strachan<sup>8</sup>, Alexsander Couto Alves<sup>9</sup>, Allan Linneberg<sup>10</sup>, John A Curtin<sup>11</sup>, Nicole M Warrington<sup>12</sup>, Marie Standl<sup>13</sup>, Marjan Kerkhof<sup>14</sup>, Ingileif Jonsdottir<sup>15,16</sup>, Blazenska K Bukvic<sup>17</sup>, Marika Kaakinen<sup>18,19</sup>, Patrick Sleimann<sup>20,21</sup>, Gudmar Thorleifsson<sup>15</sup>, Unnur Thorsteinsdottir<sup>15,16</sup>, Katharina Schramm<sup>22</sup>, Svetlana Baltic<sup>23,24</sup>, Eskil Kreiner-Møller<sup>1</sup>, Angela Simpson<sup>11</sup>, Beate St Pourcain<sup>7</sup>, Lachlan Coin<sup>9</sup>, Jennie Hui<sup>25-28</sup>, Eugene H Walters<sup>29</sup>, Carla M T Tiesler<sup>13</sup>, David L Duffy<sup>30</sup>, Graham Jones<sup>31</sup>, Australian Asthma Genetics Consortium (AAGC)<sup>32</sup>, Susan M Ring<sup>7</sup>, Wendy L McArdle<sup>7,8</sup>, Loren Price<sup>23,24</sup>, Colin F Robertson<sup>33</sup>, Juha Pekkanen<sup>34,35</sup>, Clara S Tang<sup>30</sup>, Elisabeth Thiering<sup>13</sup>, Grant W Montgomery<sup>30</sup>, Anna-Liisa Hartikainen<sup>36</sup>, Shyamali C Dharmage<sup>2</sup>, Lise L Husemoen<sup>10</sup>, Christian Herder<sup>37</sup>, John P Kemp<sup>7</sup>, Paul Elliot<sup>9</sup>, Alan James<sup>28,38,39</sup>, Melanie Waldenberger<sup>40</sup>, Michael J Abramson<sup>41</sup>, Benjamin P Fairfax<sup>42</sup>, Julian C Knight<sup>42</sup>, Ramneek Gupta<sup>3</sup>, Philip J Thompson<sup>23,24</sup>, Patrick Holt<sup>43,44</sup>, Peter Sly<sup>45</sup>, Joel N Hirschhorn<sup>4,6,46,47</sup>, Mario Blekic<sup>17</sup>, Stephan Weidinger<sup>48</sup>, Hakon Hakonarsson<sup>20,21</sup>, Kari Stefansson<sup>15,16</sup>, Joachim Heinrich<sup>13</sup>, Dirkje S Postma<sup>49</sup>, Adnan Custovic<sup>11</sup>, Craig E Pennell<sup>12</sup>, Marjo-Riitta Jarvelin<sup>18,19,50-52</sup>, Gerard H Koppelman<sup>53</sup>, Nicholas Timpson<sup>7</sup>, Manuel A Ferreira<sup>30,55</sup>, Hans Bisgaard<sup>1,55</sup> & A John Henderson<sup>7,55</sup> for the EARly Genetics and Lifecourse Epidemiology (EAGLE) Consortium

**Allergen-specific immunoglobulin E (present in allergic sensitization) has a central role in the pathogenesis of allergic disease. We performed the first large-scale genome-wide association study (GWAS) of allergic sensitization in 5,789 affected individuals and 10,056 controls and followed up the top SNP at each of 26 loci in 6,114 affected individuals and 9,920 controls. We increased the number of susceptibility loci with genome-wide significant association with allergic sensitization from three to ten, including SNPs in or near *TLR6*, *C11orf30*, *STAT6*, *SLC25A46*, *HLA-DQB1*, *IL1RL1*, *LPP*, *MYC*, *IL2* and *HLA-B*. All the top SNPs were associated with allergic symptoms in an independent study. Risk-associated variants at these ten loci were estimated to account for at least 25% of allergic sensitization and allergic rhinitis. Understanding the molecular mechanisms underlying these associations may provide new insights into the etiology of allergic disease.**

Allergic sensitization is defined as the presence of allergen-specific immunoglobulin E (IgE) against common environmental antigens<sup>1</sup>. Allergen-specific IgE has a pivotal role in the pathogenesis of allergic reactions through binding allergens and initiating the immunological processes leading to allergic inflammation, and IgE-mediated sensitivity is considered to be central in the initiation of the allergy-associated (atopic) diseases rhinitis, asthma and eczema.

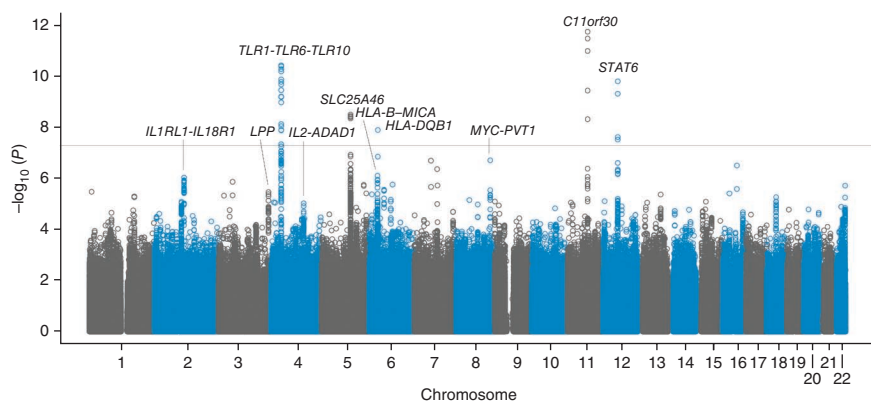
The effort to identify genes for atopic diseases is limited both by the heterogeneity of these diseases<sup>2</sup> and by diagnostic uncertainty. An alternative approach is to study an intermediate phenotype, such as allergic sensitization. This sensitization is likely to be more closely related to a specific physiological mechanism and thereby to the genetic substrate, and it is clearly defined on the basis of objective, standardized assessments. It may, therefore, be a more powerful phenotype in genetic studies. Insight into the genetics of sensitization would also increase understanding of the shared genetic basis for sensitization and atopic diseases, thus also enhancing understanding of their causal relationship.

Allergic sensitization has an estimated heritability of 0.40–0.85 (refs. 3,4). Heritability seems higher for a general tendency for sensitization (sensitization against any allergen) than for sensitization against one specific allergen<sup>5</sup>. Many candidate genes have been reported<sup>6</sup>, but few have been firmly established. The first GWAS on allergic sensitization<sup>7</sup> identified three putative susceptibility loci (5q22.1 near *TMEM232*, 11q13.5 near *C11orf30* and the human leukocyte antigen (HLA) region) but considered only sensitization to a single allergen (grass), was based on a modest sample size and did not include replication.

We conducted a two-stage GWAS meta-analysis of allergic sensitization including data from 16 studies from the EARly Genetics and Lifecourse Epidemiology (EAGLE) Consortium and the Australian

A full list of author affiliations appears at the end of the paper.

Received 29 October 2012; accepted 10 June 2013; published online 30 June 2013; doi:10.1038/ng.2694



**Figure 1** Manhattan plot for the discovery genome-wide association meta-analysis. The horizontal line indicates the genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ).

Asthma Genetics Consortium (AAGC). Sensitization status was assessed objectively and was defined by either elevated levels of allergen-specific IgE in blood or a positive skin reaction after puncture of the skin through a droplet of allergen extract (skin prick test, SPT). Sensitization status was assessed for common food and inhalant allergens (Supplementary Tables 1–3).

The stage 1 discovery phase included 5,789 individuals who were sensitized (cases) and 10,056 who were not sensitized (controls). *In silico* and *de novo* replication of the top SNP from each of the 26 loci with associations in stage 1 included up to 6,114 independent cases and 9,920 independent controls, all of European descent.

In stage 1, meta-analysis was performed on results after applying study-specific genomic control adjustment. There was minimal overall evidence of population stratification or other unaccounted biases on the full

blood IgE levels) or sex after adjustment for multiple comparisons.

All 10 genome-wide significant SNPs in our study showed evidence of association ( $P < 0.005$ ) with self-reported allergy symptoms in an independent companion study<sup>8</sup> including more than 53,000 individuals (Table 1 and Supplementary Table 5). These findings suggest that these loci increase not only the risk of sensitization but also that of allergic disease.

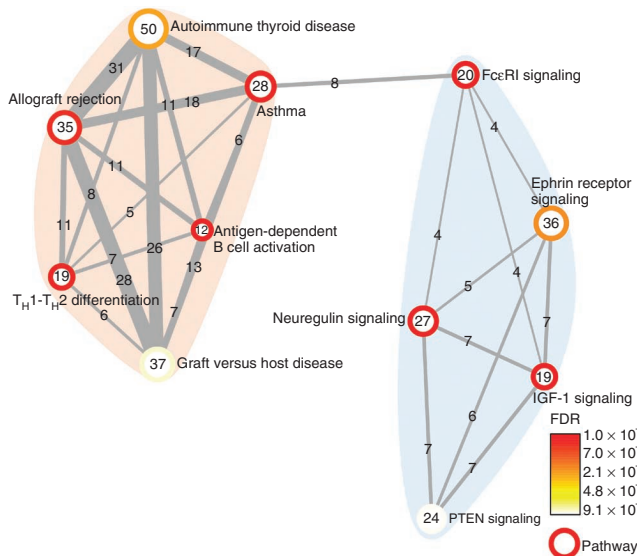
To provide insight into the molecular mechanisms underlying each of the ten associations, we searched for *cis*-acting expression quantitative trait loci (eQTLs) using gene expression data obtained from six cell types or tissues (white blood cells (WBCs), lymphoblastoid cell lines (LCLs), whole blood, adipose tissue, B cells and monocytes<sup>9</sup>). At nine of the ten loci, the sentinel SNP was associated ( $P < 0.001$ ) with the expression of one or more nearby genes (within 1 Mb of

**Table 1** Discovery and replication results for ten loci associated with allergic sensitization

Locus	Effect allele	Nearest gene	Effect allele frequency	Discovery		Replication		Combined			Replication in study of allergy symptoms <i>P</i> value category <sup>a</sup>
				OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	OR (95% CI)	<i>P</i> value	<i>P</i> <sub>het</sub> / <i>R</i> <sup>2</sup> (%)	
11q13.5	rs2155219[T]	<i>C11orf30</i>	0.47	1.20 (1.14–1.24)	<b><math>1.8 \times 10^{-12}</math></b>	1.15 (1.09–1.21)	$1.1 \times 10^{-7}$	1.18 (1.13–1.22)	<b><math>1.4 \times 10^{-18}</math></b>	0.57/0	+++
12q13.3	rs1059513[T]	<i>STAT6</i>	0.90	1.34 (1.22–1.47)	<b><math>1.6 \times 10^{-10}</math></b>	1.24 (1.13–1.37)	$1.0 \times 10^{-5}$	1.30 (1.21–1.39)	<b><math>1.0 \times 10^{-14}</math></b>	0.25/17	+
5q22.1	rs10056340[T]	<i>SLC25A46</i>	0.83	0.82 (0.77–0.88)	<b><math>3.2 \times 10^{-9}</math></b>	0.83 (0.77–0.90)	$4.4 \times 10^{-6}$	0.83 (0.78–0.87)	<b><math>5.2 \times 10^{-14}</math></b>	0.93/0	+++
6p21.32	rs6906021[T]	<i>HLA-DQB1</i>	0.55	0.86 (0.82–0.91)	<b><math>1.3 \times 10^{-8}</math></b>	0.87 (0.81–0.93)	$4.8 \times 10^{-5}$	0.87 (0.83–0.90)	<b><math>2.2 \times 10^{-12}</math></b>	0.51/0	+++
2q12.1	rs3771175[A]	<i>IL1RL1-IL18R1</i>	0.14	0.79 (0.72–0.88)	$9.1 \times 10^{-6}$	0.83 (0.77–0.90)	$1.1 \times 10^{-6}$	0.83 (0.78–0.88)	<b><math>4.9 \times 10^{-11}</math></b>	0.39/5	+++
4p14	rs17616434[T]	<i>TLR1-TLR6-TLR10</i>	0.78	1.24 (1.16–1.32)	<b><math>3.8 \times 10^{-11}</math></b>	1.22 (1.15–1.31)	$3.8 \times 10^{-10}$	1.23 (1.18–1.29)	<b><math>5.2 \times 10^{-11}</math></b> <sup>b</sup>	0.04/40	+++
3q28	rs9865818[A]	<i>LPP</i>	0.59	0.88 (0.84–0.93)	$3.4 \times 10^{-6}$	0.90 (0.85–0.94)	$2.2 \times 10^{-5}$	0.89 (0.86–0.92)	<b><math>2.7 \times 10^{-10}</math></b>	0.49/0	++
8q24.21	rs4410871[T]	<i>MYC-PVT1</i>	0.28	1.16 (1.09–1.23)	$2.0 \times 10^{-7}$	1.12 (1.05–1.20)	$6.7 \times 10^{-4}$	1.14 (1.09–1.19)	<b><math>5.4 \times 10^{-10}</math></b>	0.45/1	+
4q27	rs17454584[A]	<i>IL2-ADAD1</i>	0.74	0.88 (0.83–0.93)	$9.5 \times 10^{-6}$	0.86 (0.80–0.92)	$1.4 \times 10^{-5}$	0.87 (0.83–0.91)	<b><math>5.5 \times 10^{-10}</math></b>	0.74/0	++
6p21.33	rs6932730[T]	<i>HLA-B-MICA</i>	0.82	1.18 (1.10–1.26)	$7.9 \times 10^{-7}$	1.10 (1.02–1.18)	0.0075	1.14 (1.09–1.20)	<b><math>4.2 \times 10^{-8}</math></b>	0.26/16	++

*P* values for discovery and combined analysis are shown in bold if genome-wide significant ( $P < 5 \times 10^{-8}$ ). *P*<sub>het</sub>, heterogeneity *P* value for Cochran's *Q* statistic. OR, odds ratio; CI, confidence interval.

<sup>a</sup>*P*-value levels are categorized for the replication in an independent study on allergic symptoms<sup>8</sup>: +, significant after Bonferroni correction for the ten genome-wide significant loci,  $P < 0.005$ ; ++,  $P < 1 \times 10^{-4}$ ; +++, genome-wide significant,  $P < 5 \times 10^{-8}$ . <sup>b</sup>This *P* value was calculated by random-effects model owing to evidence of heterogeneity between studies (*P*<sub>het</sub> for Cochran's *Q* statistic  $< 0.05$ , *I*<sup>2</sup>  $> 25\%$ ).



**Figure 2** Gene set enrichment map for the 11 significant gene sets from MAGENTA analysis. Circles depict pathways. Numbers inside the circles indicate the number of genes in the pathway, and the color indicates false discovery rate, FDR. Numbers on the connecting lines indicate overlapping genes, suggesting that significant pathways are highly overlapping.

the SNP; **Supplementary Tables 6–8**). Furthermore, at all ten loci, we identified variants located within Encyclopedia of DNA Elements (ENCODE)-predicted regulatory regions that were in high linkage disequilibrium (LD,  $r^2 > 0.8$ ) with the sentinel SNP (**Supplementary Table 9**). Taken together, these results suggest that, at these ten associated loci, the underlying causal variant(s) is likely to influence DNA transcription and/or mRNA degradation. Using data from the 1000 Genomes Project, we also identified missense variants in high LD with the sentinel SNPs in *TLR1* at 4p14 and *MICA* at 6p21.32 (**Supplementary Table 10**).

The 8q24 region does not include any gene previously implicated in allergic disease or related traits. The lead SNP, rs4410871, is located in the *PVT1* transcript downstream of the *MYC* gene. *MYC* is a transcription factor involved in multiple cellular processes, including cell growth, differentiation and apoptosis in lymphocytes, and has an established role in several cancers, including B cell malignancies<sup>10</sup>. *PVT1* is a noncoding RNA that hosts several microRNAs<sup>11</sup>, which have been suggested to be as important as *MYC* in T cell activation<sup>12</sup>. The same variant (rs4410871) was reported to associate with multiple sclerosis, with the risk allele in our study (T) being the protective allele for multiple sclerosis<sup>13</sup>.

The 11q13.5 and 5q22.1 loci do not harbor any genes with an established role in the immune system. However, both loci were identified recently in a GWAS of allergic sensitization<sup>7</sup>, and the 11q13.5 locus seems to have pleiotropic effects of importance in the development of both atopic disease<sup>7,14,15</sup> and autoimmune disease<sup>16,17</sup>. The correlation structure and previous findings in the 5q22.1 region suggest that this region harbors several independent loci for atopic traits and that our top SNP represents a new risk variant (**Supplementary Table 11** and **Supplementary Note**).

The 4p14 locus harbors the genes encoding Toll-like receptor 1 (*TLR1*), *TLR6* and *TLR10*. This locus was associated with asthma and IgE levels in candidate gene studies<sup>18–20</sup>. Toll-like receptors are pattern-recognition molecules involved in innate immunity and immune responses to microbial exposure<sup>21</sup>. The top SNP at this locus

showed evidence of heterogeneity between studies. This may indicate modification of its effect by environmental factors and points to a focus for future gene-environment interaction studies.

The six remaining loci with genome-wide significant association all implicate genes with established roles in the immune system, including *STAT6* (ref. 22), *IL1RL1* (refs. 23,24), *BCL6* (ref. 25), *IL2* (ref. 26), *HLA-DQB1* (refs. 6,27) and *HLA-B-MICA*<sup>28–30</sup> (**Supplementary Note**).

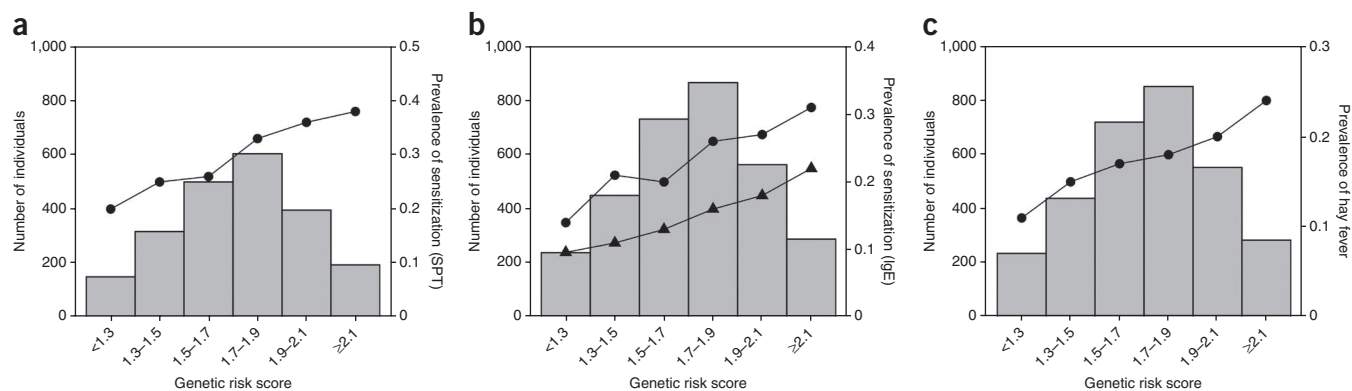
To explore biological connections and potentially identify new pathways associated with allergic sensitization, we applied gene set enrichment analysis (MAGENTA)<sup>31</sup> to the discovery-stage GWAS results and identified significant enrichment of 11 partially overlapping gene sets (**Fig. 2** and **Supplementary Table 12**). These results suggest the enrichment of two cohesive clusters of biologically related gene sets. The largest cluster included gene sets related to immune function, all sharing multiple genes, suggesting a broad overlap between autoimmunity and allergy (**Supplementary Fig. 4**). The second cluster included potentially new sensitization pathways, such as *PTEN* signaling, ephrin receptor signaling and insulin growth factor (IGF)-1 signaling, all characterized by involvement in cell growth and cancer. Three of 19 SNPs not in LD ( $r^2 < 0.1$ ) with the loci identified in our discovery GWAS (in *PTPN11*, *IL5* and *CD86*; **Supplementary Table 13**) showed evidence of association ( $P < 0.05$ ) with self-reported allergy in the companion paper<sup>8</sup>, suggesting that these are true, albeit weaker, allergy susceptibility loci.

To further highlight biological connections between genes located within the ten loci with genome-wide significant associations, we applied the GRAIL pathway analysis tool<sup>32</sup> and identified biologically connected genes related to immune function. Notably, we found that *STAT6*, *TLR6*, *IL18RAP* and *HLA-DQB1* were interrelated, and expression of these genes was associated with the top SNPs, suggesting their involvement in a common mechanism that is likely to be causally associated with sensitization (**Supplementary Fig. 5**).

We tested the potential use of the ten sensitization-associated loci in determining asthma or eczema risk by testing for association with these diseases in two large published GWAS<sup>33,34</sup>. There was evidence of association with asthma with directional consistency for nine of the ten SNPs (**Supplementary Table 14**). This overlap was considerably greater than the previously reported minimal overlap between total IgE levels and asthma-associated gene variants<sup>33</sup>, consistent with a potentially causal role for sensitization in the pathogenesis of asthma. In eczema, the associations were weaker but had consistent directions of effect (**Supplementary Table 15**).

Our analysis also provided the opportunity to investigate whether previously reported susceptibility loci for atopic diseases and related intermediate phenotypes are associated with allergic sensitization (**Supplementary Tables 16** and **17**). Many reported atopy-related genes were associated with sensitization. However, several loci strongly associated with total IgE levels, such as *FCER1A* and *HLA-A*, as well as loci strongly associated with asthma, such as the 17q12–21 locus and *IL33*, were only weakly or not significantly associated with sensitization, suggesting that these loci increase asthma risk by mechanisms other than allergic sensitization.

Lastly, we estimated the proportion of allergic sensitization in the general population caused by the ten identified loci (**Supplementary Table 18**). We obtained a conservative population-attributable risk fraction estimate above 25% by considering the 10% of the population with the lowest risk scores to represent an ‘unexposed’ group. Results were similar for allergic rhinitis, which is the strongest clinical correlate of sensitization. The capacity to predict individual risk was limited (**Supplementary Note**). Sensitization prevalence plotted by genetic risk score (**Fig. 3**) showed approximately two times higher



**Figure 3** Combined impact of risk alleles from the ten genome-wide significant loci on prevalence of allergic sensitization and allergic rhinitis (hay fever) in the population-based Health2006 replication study. (**a–c**) For each individual, we calculated a genetic risk score by applying per-allele risk estimates from the replication sets to the number of higher-risk alleles at the ten genome-wide significant loci (one SNP per locus). The risk score thus represents an index of the number of weighted risk alleles. Along the x axis, individuals in each risk score interval are shown (with the lower bound included and the upper bound excluded in each interval), and the prevalence of sensitization or hay fever in each interval is plotted (right y axis). The histogram (left y axis) shows the number of individuals in each risk score interval. (**a**) Sensitization measured by SPT. (**b**) Sensitization measured by allergen-specific IgE in blood (circles and triangles depict sensitization prevalence with IgE level cutoffs of 0.35 IU/ml and 3.5 IU/ml, respectively). (**c**) Hay fever.

prevalence of sensitization in the 9.2% of the population with the highest genetic risk scores compared to the 7.6% of the population with the lowest risk scores.

Several of the identified loci for sensitization are also reported susceptibility loci for autoimmune disease (**Supplementary Table 19**). We observed examples both of opposing directions of effect (for multiple sclerosis at the *MYC-PVT1* locus) and the same direction of effect (for diabetes and Crohn's disease at the *IL2* and *C11orf30* loci) relative to the effects observed in sensitization. Allergic and autoimmune disorders are classically described as representatives of helper T cell type 1 ( $T_H1$ ) and type 2 ( $T_H2$ ) immunity associated with the production of cross-regulatory cytokines. Our results suggest that the  $T_H1$  and  $T_H2$  paradigm<sup>35</sup> does not adequately explain the relationship between these disorders.

In conclusion, we increased the number of established susceptibility loci for allergic sensitization from three to ten, which contribute substantially to the population-attributable risk fraction. Understanding the molecular mechanisms underlying associations with these sequence variants may provide new insights into the etiology of allergy and other immune diseases.

**URLs.** MaCH, <http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G-PhaseI-Interim.html>; HaploReg, <http://www.broadinstitute.org/mammals/haploreg/haploreg.php>; GHS\_Express, <http://genecanvas.ecgene.net/uploads/ForReview/>.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

*Note: Supplementary information is available in the online version of the paper.*

## ACKNOWLEDGMENTS

A full list of acknowledgments for each study is given in the **Supplementary Note**.

## AUTHOR CONTRIBUTIONS

**Study-level data analysis:** K.B., T.H.P. (meta-analysis and systems biology analyses), R. Granell (meta-analysis), D.P.S., A.C.A. (systems biology analyses), A.L., J.A.C., N.M.W., M.S., M. Kerkhof, B.K.B., M. Kaakinen, P. Sleimann, G.T.,

K. Schramm, E.K.-M., A.S., L.C., C.S.T., B.P.F., R. Gupta, M.B., S.W., H.H., D.S.P., A.C., G.H.K. and N.T. **Study design:** K.B., M.C.M., T.H.P., D.P.S., A.L., B.K.B., E.K.-M., A.S., C.F.R., G.W.M., S.C.D., P.E., M.J.A., J.C.K., R. Gupta, P.J.T., J.N.H., M.B., S.W., H.H., J. Heinrich, D.S.P., A.C., C.E.P., M.-R.J., G.H.K., N.T., M.A.F., H.B. and A.J.H. **Writing manuscript:** K.B., T.H.P., R. Granell, D.P.S., A.C.A., A.L., J.A.C., G.H.K., N.T., M.A.F., H.B. and A.J.H. **Data collection:** K.B., M.C.M., D.P.S., A.L., I.J., B.K.B., U.T., A.S., J. Hui, E.H.W., D.L.D., G.J., L.P., C.F.R., J.P., E.T., A.-L.H., S.C.D., L.L.H., C.H., A.J., B.P.F., J.C.K., M.J.A., P.J.T., P.H., P. Sly, M.B., H.H., K. Stefansson, J. Heinrich, D.S.P., A.C., M.-R.J., G.H.K., H.B. and A.J.H. **Genotyping:** K.B., A.L., J.A.C., I.J., P. Sleimann, U.T., S.B., E.K.-M., A.S., B.S.P., C.M.T.T., S.M.R., W.L.M., G.W.M., L.L.H., J.P.K., M.W., H.H., K. Stefansson, A.C., C.E.P., M.-R.J., G.H.K. and H.B. **Revising and reviewing manuscript:** K.B., M.C.M., T.H.P., R. Granell, D.P.S., A.C.A., A.L., J.A.C., N.M.W., M.S., M. Kerkhof, I.J., B.K.B., M. Kaakinen, P. Sleimann, G.T., U.T., K. Schramm, S.B., E.K.-M., A.S., B.S.P., L.C., J. Hui, E.H.W., C.M.T.T., D.L.D., G.J., S.M.R., W.L.M., L.P., C.F.R., J.P., C.S.T., E.T., G.W.M., A.-L.H., S.C.D., L.L.H., C.H., J.P.K., P.E., A.J., M.W., M.J.A., B.P.F., J.C.K., R. Gupta, P.J.T., P.H., P. Sly, J.N.H., M.B., S.W., H.H., K. Stefansson, J. Heinrich, D.S.P., A.C., C.E.P., M.-R.J., G.H.K., N.T., M.A.F., H.B. and A.J.H.

## COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Johansson, S.G. *et al.* Revised nomenclature for allergy for global use: report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J. Allergy Clin. Immunol.* **113**, 832–836 (2004).
- Anderson, G.P. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* **372**, 1107–1119 (2008).
- Los, H., Postmus, P.E. & Boomsma, D.I. Asthma genetics and intermediate phenotypes: a review from twin studies. *Twin Res.* **4**, 81–93 (2001).
- Thomsen, S.F., Ulrik, C.S., Kyvik, K.O., Ferreira, M.A. & Backer, V. Multivariate genetic analysis of atopy phenotypes in a selected sample of twins. *Clin. Exp. Allergy* **36**, 1382–1390 (2006).
- Thomsen, S.F., van der Sluis, S., Kyvik, K.O. & Backer, V. A study of asthma severity in adult twins. *Clin. Respir. J.* **6**, 228–237 (2012).
- Vercellii, D. Discovering susceptibility genes for asthma and allergy. *Nat. Rev. Immunol.* **8**, 169–182 (2008).
- 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.* **128**, 996–1005 (2011).
- Hinds, D.A. *et al.* A meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat. Genet.* published online; doi:10.1038/ng.2686 (30 June 2013).
- Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).

10. Adhikary, S. & Eilers, M. Transcriptional regulation and transformation by Myc proteins. *Nat. Rev. Mol. Cell Biol.* **6**, 635–645 (2005).
11. Huppi, K. *et al.* The identification of microRNAs in a genomically unstable region of human chromosome 8q24. *Mol. Cancer Res.* **6**, 212–221 (2008).
12. Beck-Engeser, G.B. *et al.* Pvt1-encoded microRNAs in oncogenesis. *Retrovirology* **5**, 4 (2008).
13. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
14. Esparza-Gordillo, J. *et al.* A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nat. Genet.* **41**, 596–601 (2009).
15. Ferreira, M.A. *et al.* Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006–1014 (2011).
16. Anderson, C.A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246–252 (2011).
17. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
18. Kormann, M.S. *et al.* Toll-like receptor heterodimer variants protect from childhood asthma. *J. Allergy Clin. Immunol.* **122**, 86–92 (2008).
19. Lazarus, R. *et al.* TOLL-like receptor 10 genetic variation is associated with asthma in two independent samples. *Am. J. Respir. Crit. Care Med.* **170**, 594–600 (2004).
20. Reijmerink, N.E. *et al.* Toll-like receptors and microbial exposure: gene-gene and gene-environment interaction in the development of atopy. *Eur. Respir. J.* **38**, 833–840 (2011).
21. Tesse, R., Pandey, R.C. & Kabisch, M. Genetic variations in Toll-like receptor pathway genes influence asthma and atopy. *Allergy* **66**, 307–316 (2011).
22. Potaczek, D.P. & Kabisch, M. Current concepts of IgE regulation and impact of genetic determinants. *Clin. Exp. Allergy* **42**, 852–871 (2012).
23. Gudbjartsson, D.F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **41**, 342–347 (2009).
24. Savenije, O.E. *et al.* Interleukin-1 receptor-like 1 polymorphisms are associated with serum IL1RL1-a, eosinophils, and asthma in childhood. *J. Allergy Clin. Immunol.* **127**, 750–756 (2011).
25. Ichii, H. *et al.* Role for Bcl-6 in the generation and maintenance of memory CD8<sup>+</sup> T cells. *Nat. Immunol.* **3**, 558–563 (2002).
26. Christensen, U. *et al.* Family based association analysis of the *IL2* and *IL15* genes in allergic disorders. *Eur. J. Hum. Genet.* **14**, 227–235 (2006).
27. Howell, W.M. & Holgate, S.T. HLA genetics and allergic disease. *Thorax* **50**, 815–818 (1995).
28. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–1125 (2010).
29. Martin, A.M. *et al.* Predisposition to abacavir hypersensitivity conferred by *HLA-B\*5701* and a haplotypic *Hsp70-Hom* variant. *Proc. Natl. Acad. Sci. USA* **101**, 4180–4185 (2004).
30. Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nat. Genet.* **42**, 985–990 (2010).
31. Segrè, A.V., Groop, L., Mootha, V.K., Daly, M.J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).
32. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
33. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363**, 1211–1221 (2010).
34. Paternoster, L. *et al.* Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat. Genet.* **44**, 187–192 (2012).
35. Rabin, R.L. & Levinson, A.I. The nexus between atopic disease and autoimmunity: a review of the epidemiological and mechanistic literature. *Clin. Exp. Immunol.* **153**, 19–30 (2008).

<sup>1</sup>COPSAC (Copenhagen Prospective Studies on Asthma in Childhood), Health Sciences, University of Copenhagen & Copenhagen University Hospital, Gentofte, Denmark. <sup>2</sup>Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Victoria, Australia. <sup>3</sup>Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark. <sup>4</sup>Division of Endocrinology, Children's Hospital, Boston, Massachusetts, USA. <sup>5</sup>Center for Basic and Translational Obesity Research, Children's Hospital, Boston, Massachusetts, USA. <sup>6</sup>Broad Institute, Cambridge, Massachusetts, USA. <sup>7</sup>School of Social and Community Medicine, University of Bristol, Bristol, UK. <sup>8</sup>Division of Population Health Sciences & Education, St. George's, University of London, London, UK. <sup>9</sup>Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. <sup>10</sup>Research Centre for Prevention and Health, Glostrup Hospital, Glostrup, Denmark. <sup>11</sup>Manchester Academic Health Science Centre, University of Manchester, University Hospital of South Manchester National Health Service (NHS) Foundation Trust, Manchester, UK. <sup>12</sup>School of Women's and Infants' Health, The University of Western Australia, Crawley, Western Australia, Australia. <sup>13</sup>Institute of Epidemiology I, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. <sup>14</sup>Department of Epidemiology, Groningen Research Institute for Asthma and COPD (GRIAC), University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. <sup>15</sup>deCODE Genetics, Reykjavik, Iceland. <sup>16</sup>University of Iceland, Reykjavik, Iceland. <sup>17</sup>Department of Pediatrics, General Hospital Dr. Josip Benčević Slavonki Brod, University of Osijek, Osijek, Croatia. <sup>18</sup>Institute of Health Sciences, University of Oulu, Oulu, Finland. <sup>19</sup>Biocenter Oulu, University of Oulu, Oulu, Finland. <sup>20</sup>Center for Applied Genomics, The Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. <sup>21</sup>Division of Human Genetics, The Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. <sup>22</sup>Institute of Human Genetics, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. <sup>23</sup>Lung Institute of Western Australia, The University of Western Australia, Perth, Western Australia, Australia. <sup>24</sup>Centre for Asthma, Allergy and Respiratory Research, The University of Western Australia, Perth, Western Australia, Australia. <sup>25</sup>PathWest Laboratory Medicine of Western Australia, Nedlands, Western Australia, Australia. <sup>26</sup>School of Population Health, The University of Western Australia, Nedlands, Western Australia, Australia. <sup>27</sup>School of Pathology and Laboratory Medicine, The University of Western Australia, Nedlands, Western Australia, Australia. <sup>28</sup>Busselton Population Medical Research Foundation, Sir Charles Gairdner Hospital, Perth, Western Australia, Australia. <sup>29</sup>School of Medicine, University of Tasmania, Hobart, Tasmania, Australia. <sup>30</sup>Queensland Institute of Medical Research, Brisbane, Queensland, Australia. <sup>31</sup>School of Science and Health, University of Western Sydney, Penrith, New South Wales, Australia. <sup>32</sup>Details appear in the **Supplementary Note**. <sup>33</sup>Respiratory Medicine, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. <sup>34</sup>Department of Environmental Health, National Institute for Health and Welfare (THL), Kuopio, Finland. <sup>35</sup>Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. <sup>36</sup>Department of Clinical Sciences, Obstetrics and Gynecology, Institute of Clinical Medicine, University of Oulu, Oulu, Finland. <sup>37</sup>Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>38</sup>School of Medicine and Pharmacology, The University of Western Australia, Nedlands, Western Australia, Australia. <sup>39</sup>Department of Pulmonary Physiology, West Australian Sleep Disorders Research Institute, Nedlands, Western Australia, Australia. <sup>40</sup>Research Unit of Molecular Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. <sup>41</sup>Department of Epidemiology & Preventive Medicine, Monash University, Melbourne, Victoria, Australia. <sup>42</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>43</sup>Telethon Institute for Child Health Research, The University of Western Australia, Perth, Western Australia, Australia. <sup>44</sup>Centre for Child Health Research, The University of Western Australia, Perth, Western Australia, Australia. <sup>45</sup>Queensland Children's Medical Research Institute, University of Queensland, World Health Organization (WHO) Collaborating Centre for Research on Children's Environmental Health, Brisbane, Queensland, Australia. <sup>46</sup>Division of Genetics, Children's Hospital, Boston, Massachusetts, USA. <sup>47</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. <sup>48</sup>Department of Dermatology, Allergy and Venerology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. <sup>49</sup>Department of Pulmonology, GRIAC, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. <sup>50</sup>Department of Epidemiology and Biostatistics, School of Public Health, Medical Research Council—Health Protection Agency (MRC-HPA) Centre for Environment and Health, Imperial College London, London, UK. <sup>51</sup>National Institute of Health and Welfare, Oulu, Finland. <sup>52</sup>Unit of Primary Care, Oulu University Hospital, Oulu, Finland. <sup>53</sup>Department of Pediatric Pulmonology and Pediatric Allergy, Beatrix Children's Hospital, GRIAC, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. <sup>54</sup>These authors contributed equally to this work. <sup>55</sup>These authors jointly directed this work. Correspondence should be addressed to K.B. (kb@copsac.com).

## ONLINE METHODS

**Phenotype definition.** We studied the general tendency of allergen-specific IgE production without restriction by assessment method or allergen. Cases were defined as individuals with objectively measured sensitization against at least one of the allergens tested for in the respective studies, and controls were defined as individuals who were not sensitized against any of the allergens tested (**Supplementary Tables 1 and 2**). We included sensitization assessed by skin reaction after puncture of the skin with a droplet of allergen extract (SPT) and/or by detection of the levels of circulating allergen-specific IgE in the blood. The SPT wheal diameter cutoffs were 3 mm larger than the negative control for cases and smaller than 1 mm for controls. To optimize case specificity and the correlation between methods, we chose a high cutoff of specific IgE levels for cases (3.5 IU/ml) and a low cutoff for controls (0.35 IU/ml). We combined data from children and adults but chose a lower age limit of 6 years, as sensitization status at younger ages shows poorer correlation with sensitization later in life, both owing to transient sensitization and frequent development of sensitization during late childhood.

**Discovery analysis.** The discovery analysis was based on 5,798 individuals who were sensitized (cases) and 10,056 controls of European descent from 10 studies. Individual studies were approved by the local research ethics committees, and informed consent was obtained from all participants when necessary. Details on sample recruitment as well as phenotypes and summary details for each collection are given in the **Supplementary Note** and in **Supplementary Tables 1 and 2**. Genome-wide genotyping was performed independently in each cohort with the use of various standard genotyping technologies (**Supplementary Table 3** and **Supplementary Note**). Imputation was independently conducted for each study with reference to HapMap phase 2 or 3 CEU genotypes (Utah residents of Northern and Western European ancestry), and association analysis was performed using logistic regression models based on an expected allelic dosage model for SNPs, adjusting for ancestry-informative principal components as necessary. SNPs with minor allele frequency (MAF) of <1% and/or poor imputation quality ( $r^2 < 0.3$ , if using the imputation algorithm MaCH, or proper info <0.4, if using the IMPUTE imputation algorithm) were excluded. After genomic control at the level of the individual studies, we performed meta-analysis of association data for ~2.5 million imputed and genotyped autosomal SNPs using additive model fixed-effects meta-analysis as implemented in METAL (version 2010-08-01). For a given SNP, study-specific effect size estimates (transformed to  $z$  scores by METAL) were weighted inversely with their variance.

**Replication analysis.** We selected for replication the most strongly associated SNP in each of the 26 most strongly associated loci in the discovery meta-analysis (all with  $P < 1 \times 10^{-5}$  in stage 1; **Table 1** and **Supplementary Table 4**). These SNPs were analyzed using *in silico* data from six studies with GWAS data not included in the discovery meta-analysis and *de novo* genotyping data from three additional studies (**Supplementary Tables 1–3**). The largest possible replication sample size amounted to 6,114 cases and 9,920 controls from 9 studies, all of European descent. Association analyses were again conducted for each study using a logistic regression model with similar covariate adjustments, based on expected allelic dosage for the *in silico* studies and allele counts in the *de novo* genotyping studies. We applied a threshold of  $P < 5 \times 10^{-8}$  for genome-wide significance. Informed consent was obtained from all parents (and children, when appropriate) participating in the individual studies, which were approved by the respective ethics committee. Details can be found in the **Supplementary Note**.

**Association and Manhattan plots.** The LocusZoom tool was used to generate association plots with the 400-kb region centered on the index SNP<sup>36</sup>. Manhattan plots were generated with R Bioconductor software.

**Functional annotation.** To help identify the potential causal gene(s) underlying each of the ten associations, we searched for *cis* eQTLs using gene expression results obtained from in-house data sets of four cell types or tissues: WBCs, LCLs, whole blood and adipose tissue.

*Avon Longitudinal Study of Parents and Children (ALSPAC).* mRNA quantified expression data were collected for 947 unrelated ALSPAC participants

from LCLs established from blood samples taken when the children were 9 years old. All cell lines were from a first passage. Cells were harvested when confluent, washed in PBS and frozen in RNAlater (Ambion, Life Technologies). RNA was extracted using the Qiagen RNeasy extraction kit and amplified using the Illumina TotalPrep-96 RNA Amplification kit. Expression was surveyed using Illumina HT-12 v3 BeadChip arrays. Expression data from ALSPAC are available upon request in accordance with the ALSPAC data sharing policy (<http://www.bris.ac.uk/alspac/researchers/data-access/policy/>). Each individual sample was run with two replicates. Expression data were normalized by quantile normalization between replicates and then by median normalization across individuals.

*KORA.* Expression analysis in this study was based on whole-blood samples of KORA (Cooperative Health Research in the Region of Augsburg) F4 participants aged 62 to 81 years<sup>37</sup>. Gene expression profiling was performed on RNA isolated from whole blood using the Illumina HumanHT-12 v3 BeadChip as described elsewhere<sup>38</sup> (deposited at ArrayExpress under accession E-MTAB-1174).

*deCODE.* Expression analyses were based on WBCs ( $n = 745$ ) and adipose tissue ( $n = 600$ ) from randomly selected Icelandic individuals. Gene expression profiling was performed as previously described<sup>39</sup> (deposited in the Gene Expression Omnibus (GEO) under accessions GSE7965 and GPL3991).

*B cell and monocyte eQTL study.* Expression quantitative trait analysis was performed on gene expression data from CD14<sup>+</sup> monocyte ( $n = 282$ ) and CD19<sup>+</sup> B cell samples that were positively selected from freshly purified peripheral blood mononuclear cell fractions as previously described<sup>40</sup> (deposited at ArrayExpress under accession EGAD00010000144). Genotyping data for these European individuals were generated using HumanOmniExpress-12v1.0 BeadChips (Illumina), with whole-genome imputation performed using IMPUTE2 against the 1000 Genomes Project phase I integrated variant set (March 2012). Expression data were obtained using HumanHT-12 v4 BeadChips (Illumina), with variance-stabilizing transformation and robust spline normalization between chips performed using the R package lumi<sup>41</sup>. To identify putative *cis* eQTLs, linear analysis was performed using Matrix eQTL<sup>42</sup> with a 1-Mb window between each SNP and probe and with the first 15 principal components of the expression data included as covariates.

Furthermore, we used publicly available bioinformatics resources to annotate putative associations with gene expression, coding SNPs and regulatory elements. We annotated SNPs in LD ( $r^2 > 0.8$ , within 2 Mb) with the top SNP from each of the genome-wide significant loci on the basis of 1000 Genomes Project Interim Phase I data (b37) from 381 Europeans. Chromosomal positions were compared against UCSC dbSNP132 (hg19) to find alternative SNP IDs (for example, chromosome 4: 38,774,337 to rs111437368). Regulatory elements were annotated by HaploReg in terms of predicted ENCODE chromatin state (promoter and enhancer histone modification signals) and DNase I hypersensitivity. We used the GHS\_Express database<sup>9</sup> to identify variants associated ( $P < 1 \times 10^{-5}$ ) with gene expression levels (eQTL) and in LD ( $r^2 > 0.3$ ) with the ten top SNPs; the RefGene tract of the UCSC Genome Browser was used for gene annotation.

**Population-attributable risk fractions (PARFs).** Population-attributable risk fractions were estimated from three large general-population samples: B58C (participant ages of 44–45 years), Health2006 (participant ages of 18–69 years) and ALSPAC (participant age of 7.5 years). Health2006 and the B58C subsample used for this purpose were part of the replication stage only, whereas ALSPAC samples were part of both the discovery and replication stages. Within each cohort, a genetic risk score was calculated by applying the pooled per-allele coefficients (ln(OR) values) from the replication sets to the number of higher-risk alleles of each of the ten genome-wide significant loci (one SNP per locus). Population-attributable risk fractions were estimated in two ways. First, the prevalence of sensitization at a genetic risk score of zero (no higher-risk alleles for any of the ten SNPs) was predicted, with 95% confidence intervals, from the intercept term in a logistic regression model with risk score as a continuous explanatory variable. Each population-attributable risk fraction was then derived (with 95% confidence interval) by expressing the difference between the observed prevalence and the predicted (unexposed) prevalence as a percentage of the observed prevalence. Second, because there were no individuals observed with zero higher-risk alleles, the prevalence of sensitization

for individuals in the lowest decile of the genetic risk score distribution was estimated in a similar manner and was used to derive population-attributable risk fraction on the assumption that this 10% of the population was unexposed. The second method has the advantage that it does not predict beyond the bounds of the data, but its results are conservative compared to those from the first method. Population-attributable risk fractions were estimated for sensitization measured by SPT and specific IgE levels (cutoffs of 0.35 and 3.5 IU/ml) and for allergic rhinitis.

**Pathway enrichment analyses.** *MAGENTA pathway analysis.* To explore potentially new pathways associated with allergic sensitization in the full GWAS data set, we applied an adaption of gene set enrichment analysis (meta-analysis gene set enrichment of variant associations, MAGENTA version 2)<sup>31</sup>. Briefly, each gene in the genome was mapped to a single SNP with the lowest *P* value within a 110-kb window upstream or a 40-kb window downstream of the SNP. Gene *P* values were corrected for confounding factors, such as gene size, LD patterns, SNP density and other genetic factors. Adjusted *P* values were ranked, and the observed number of genes in a given pathway above a given significance threshold (95th percentile) was calculated. This statistic was compared with that generated with 10,000 randomly permuted pathways of identical size to generate an empirical gene set enrichment analysis *P* value for each pathway. In our study, individual pathways that reached FDR < 0.01 were deemed significant, and, unless otherwise stated, results for the 95th-percentile cutoff analysis are reported.

To identify the degree of overlap between gene sets, we plotted an enrichment map of the significant pathways. An enrichment map is a network where

vertices represent pathways, vertex size is proportional to the number of genes and edge width represents the number of genes shared by two pathways. Vertices are colored according to enrichment FDR. To further identify which significant pathways were central, we plotted a bipartite network of significant genes (adjusted *P* < 0.0273) connected to their corresponding pathways.

*GRAIL pathway analysis.* GRAIL<sup>32</sup> was run with gene size correction enabled and was based on PubMed abstracts for papers published in or before December 2006. We specified genes at each of the ten validated loci explicitly by considering all genes within the 1-Mb flanking region on each side of the given lead SNP. The VIZ-GRAIL tool<sup>43</sup> was used to visualize the GRAIL results.

36. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
37. Rathmann, W. *et al.* Incidence of type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study. *Diabet. Med.* **26**, 1212–1219 (2009).
38. Mehta, D. *et al.* Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *Eur. J. Hum. Genet.* **21**, 48–54 (2013).
39. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
40. Fairfax, B.P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
41. Du, P., Kibbe, W.A. & Lin, S.M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
42. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
43. Raychaudhuri, S. VIZ-GRAIL: visualizing functional connections across disease loci. *Bioinformatics* **27**, 1589–1590 (2011).