Supplement - Shared genetic variants suggest common pathways in allergy and autoimmune diseases
Supplementary Methods

Discovery and meta-analysis

Discovery results from an allergic sensitization GWAS was included as one of the datasets comprising results from 5,809 with allergic sensitization and 9,875 controls with data from the following cohorts: AAGC, ALSPAC\textsuperscript{1}, B58C, COPSAC2000, LISA, MAAS, NFBC 1966, RAINE and PIAMA. Please see discovery paper for ethical statements, cohort profiles and numbers.\textsuperscript{2} Sensitization status was assessed objectively by either elevated levels of allergen-specific IgE (sIgE) in blood or by a positive skin reaction after skin prick test (SPT) against common food and inhalant allergens. The SPT cut off level was 3 mm larger than the negative control for cases, and below 1 mm for controls. For sIgE, cases were defined as levels at or above 3.5 IU/mL and for controls below 0.35 IU/mL. Imputation was independently conducted for each study with reference to HapMap phase 2 or 3 CEU genotypes (Central European ancestry). Study level association analysis was performed using logistic regression models based on an expected additive allelic dosage model for SNPs, adjusting for ancestry-informative principal components as necessary. SNPs with MAF <1\% and/or poor imputation quality (MACH r$^2$<0.3 or IMPUTE proper info <0.4) were excluded. After genomic control at the level of the individual studies, the summary statistics were meta-analyzed using a fixed effects model and inverse variance as weights in METAL (2010-08-01). In total 2,400,129 SNPs were available in three or more cohorts.

GWAS results from the 23andMe study were included in the present study involving 10,509 individuals with self-reported cat-allergy, 9,815 with Dust-mite allergy, 16,133 with Pollen allergy (grasses, trees or weeds) and 26,311 without symptoms in a total of 46,646 individuals. Allergy was defined as those individuals who reported a positive allergy test, difficulty in swallowing or speaking, hives, itchy mouth, itchy eyes, itchy nose or asthma in response to a particular allergen.\textsuperscript{3} The second part of the study sample on self-reported allergy (mothers from the Alspac cohort) was not included in the present study as these individuals are related to individuals in the GWA on sensitization.
Imputation was performed in the 23andMe study using the 1000 Genomes reference (August 2010 release) in batches. SNPs with an imputed $r^2 > 0.5$ averaged across all batches and $r^2 > 0.3$ in every batch were used. SNPs were remapped to B36. A generalized estimating equation (GEE) model was applied to assess the shared genetic effects on all three phenotypes taking into account the correlations between these phenotypes.

The meta-analysis was performed using a fixed effects model using inverse variance as weights in METAL (2010-08-01) after a second genomic control for the meta-analysis of the dataset on self reported allergy and sensitization.

**Enrichment of autoimmune disease-associated loci and allergy**

Candidate loci were chosen from the GWA’s catalog accessed 25th of November 2013 using autoimmune and inflammatory traits of interest (Supplementary Table E1). These reported traits were collapsed into 16 overall autoimmune disease traits. All SNP-trait associations with $P < 5e-8$ were used. We collapsed close SNPs into loci (+/-250kb) and used for each locus the SNP with lowest reported $P$ as index SNP (Supplementary Figure 1). For common loci (listed in table 2), all original publications were checked for effect allele, and any discrepancies with the GWA’S catalog was corrected. After the extraction of SNP-associations, the enrichment Odds Ratio was calculated as the number of observed extracted SNPs with $P < 0.05$ out of total extracted SNPs as compared to total number of independent SNPs with $P < 0.05$ within the GWA discovery results using a Fisher’s exact test.

For this we used Hapmap, CEU panel, to define independent loci. This was performed using PLINK (--indep-pairwise 200 5 0.5) with a sliding window on 200 SNPs at steps on 5 SNPs pruning the datasets to contain only one of 2 correlated SNPs with a $r^2 > 0.5$. To plot enrichment, we equally plotted observed $P$-values against expected under the null hypothesis (QQ plots). Enrichment and QQ plots were plotted overall for all 290 loci and separately for each of the 16 autoimmune diseases, however only for those diseases with more than 10 loci reported in the GWA catalogue. For extracted SNPs a False Discovery Rate corrected $P$-value $< 0.05$ was considered significant. Analysis were performed in Plink and R project (3.0.1).
Functional evaluation

Enrichment of SNPs falling in DHS sites:
DHS sites were downloaded from the ENCODE project\(^9\) and from the Epigenomics Roadmap\(^{10}\) selecting only cell types (or cell lines, herefrom “cell types”) with duplicates, removing transformed cell types, and removing redundant cell lines, based on manual curation. Huh7 was an outlier in all analyses, and was removed. DHS sites were set to a fixed width of 150bp from center of region for all cell types. Allergy and Crohn’s Disease SNPs were split in bins of increasing p-value cutoff, starting at 1 (including all SNPs and setting baseline for enrichment) and decreasing one decimal digitor each bin (1, 0.1, 0.001 etc). Each bin was overlapped with DHS regions using bedtools v2.19.0, and enrichment for each bin was calculated for each cell type as compared to p =1. For GWAS Catalogue SNPs, SNPs were selected for traits with > 30 reported associated SNPs. For identical SNPs for the same trait, the SNP with the lowest p-value was chosen. SNPs were then overlapped with DHS regions using bedtools v2.19.0, and ratio of overlapping SNPs was calculated. To filter out non-informative cell-types, only cell-types with the highest quartile of overlap ratios was included. For immune cell hierarchical clustering the manhattan distance of square root transformed ratios were used. For PCA, log10 transformed ratios were used.

Enrichment of SNPs falling in FANTOM enhancers:
FANTOM cell specific enhancers were downloaded from ‘http://enhancer.binf.ku.dk/’ and were set to a fixed width of 150bp from center of region for all cell types. The ratio between overlaps of all SNPs (p <= 1) and SNPs at p<= 1e-5 was calculated, and a p-value for this ratio calculated using a binominal test with the genomic overlap frequency as null frequency, calculated as the number of total enhancers per cell type times enhancer length (150bp), divided by the total number of base pairs shown to be bound by transcription factors in the human genome across cell types in the ENCODE project (231mb)\(^9\). FDR values were calculated adjusting p-values with the Benjamini-Hochberg method.
Data-driven Enrichment-Prioritized Integration for Complex Traits (DEP ICT):

For details of this method please refer to Pers et al.\textsuperscript{11}. DEP ICT facilitates gene set enrichment analysis by testing whether genes in associated regions enrich for reconstituted versions of known pathways, gene sets, as well as protein complexes. The gene-set enrichment analyses in DEP ICT contains three steps: first a scoring step; second a bias correcting step taking into account gene density that possibly could inflate results due to gene length and finally estimating experiment-wide FDR’s.

Gene set reconstitution is accomplished by identifying genes that are co-expressed with genes in a given gene set based on a panel of 77,840 gene expression microarrays; genes that co-express with genes from a given gene set are likely to be part of that gene set.\textsuperscript{12} In DEP ICT, several types of gene sets were reconstituted: 5,984 protein complexes that were derived from 169,810 high-confidence experimentally-derived protein-protein interactions\textsuperscript{13}; 2,473 phenotypic gene sets derived from 211,882 gene-phenotype pairs from the Mouse Genetics Initiative\textsuperscript{14}; 737 Reactome database pathways\textsuperscript{15}; 184 KEGG database pathways\textsuperscript{16}; and 5,083 Gene Ontology database terms\textsuperscript{17}. In addition, the DEP ICT also facilitates tissue and cell type enrichment analysis, by testing whether genes in associated regions are highly expressed in any of 209 Medical Subject Heading annotations of 37,427 microarrays from the Affymetrix U133 Plus 2.0 Array platform. We used DEP ICT to test enrichment in a total of 14,461 reconstituted gene sets and enrichment of 209 tissue and cell type annotations.

For the allergy meta-analysis, and Crohns disease, DEP ICT was performed on all loci P<10e-5. For PCA GWAS catalogue data, DEP ICT was performed on all traits with more than 30 reported associated SNPs. For identical SNPs for the same trait, the SNP with the lowest p-value was chosen.

Pathway analysis and visualization for allergy and Crohn’s disease:

To account for difference in GWAS study sizes, a linear model was fitted between logged p-values of DEP ICT results for allergy and Chron’s disease, and the estimator was used to adjust the p-value thresholds for the largest study, Crohn’s disease. The inflation estimate for Crohn’s disease was 1.21. Shared pathways were set at $p_{\text{allergy}}$ and $p_{\text{crohns\_adjusted}} < 0.001$, allergy specific pathways were set at $p_{\text{allergy}} < 0.001$ and $p_{\text{crohns\_adjusted}} > 0.05$ and Crohn’s disease specific pathways were set at $p_{\text{allergy}} > 0.05$ and $p_{\text{crohns\_adjusted}} < 0.001$. 

\textsuperscript{11} Pers et al.
\textsuperscript{12} Experimentally-derived protein-protein interactions.
\textsuperscript{13} Phenotypic gene sets.
\textsuperscript{14} Mouse Genetics Initiative.
\textsuperscript{15} Reactome database pathways.
\textsuperscript{16} KEGG database pathways.
\textsuperscript{17} Gene Ontology database terms.
DHS genomic location:

Genomic regions for 186 cell types, of which 14 cell types (CD14_Primary_Cells, CD19_Primary_Cells_Peripheral_UW, CD20, CD3_Primary_Cells_Cord_BI, CD3_Primary_Cells_Peripheral_UW, CD34Mobilized, CD56_Primary_Cells, Th0, Th1, Th17, Th2, Treg, GM12864, and Fetal_Thymus) were annotated as immune cells, were downloaded from the ROADMAP and ENCODE tracks (June 2014) in the UCSC Genome Browser and processed by bedtools, ensuring no redundancy between exons, introns, promoters (defined as 5000 bases upstream and 200 bases downstream of transcription start sites), and intergenic sites. DHS sites were overlapped with genomic regions, requiring 1bp of overlap. Enrichment of markers in DHS regions was calculated for GWAS catalog traits with 30 or more reported variants, and was normalized by trait SNP count and cell type specific DNAse sequence lengths.

The uneven distribution of cell types within the Roadmap ENCODE dataset could possibly contribute to the separation of immune-mediated diseases from other diseases. However, repeated iterative removal of ¼ of cell types continuously produced a statistical significant separation of autoimmune diseases, allergy and asthma vs. other traits (results not shown), hence supporting the finding of common SNPs and cell types to congregate in allergy and autoimmune diseases. In addition, hierarchical clustering was performed on the full ENCODE Roadmap set to investigate if this clustering was facilitated by similar DHS-profiles in different immune cells, basically representing a single “immune-system-footprint”, but this was not the case, as different immune cell types also separated internally, comparable to other non-immune cell types (Supplementary Figure 18).

A further cluster analysis of the cell-type specific genomic DHS location in all cells (intronic, exonic, intergenic, promoter) was performed revealing that the DHS sites in immune cells tend to fall within promoter and exonic regions (Supplementary Figure 19).
Transcription factor binding sites:
Transcription factor binding sites for 161 transcription sites were downloaded in BED format from ENCODE for the hg19 build, and were intersected with 28 independent shared loci, expanded to included markers with r2>= 0.5 in the 1000g CEU panel, using BedTools. Enrichment and one-sided p-values were calculated in relation to an empirical null distribution of loci overlap for each TF, generated by 10,000 random permutations of random genomic loci with the same length characteristics as the 28 LD-expanded shared loci. Random locus LD-structure was assumed to have no effect on TF-binding probability as a function of locus length.
Acknowledgements and funding

**23andMe:** We thank the customers of 23andMe who answered surveys, as well as the employees of 23andMe, who together made the 23andMe research possible. This work was supported in part by the National Heart, Lung, and Blood Institute of the US National Institutes of Health under grant 1R43HL115873-01.

**AAGC/QIMR:** We thank the twins and their families for their participation; Dixie Statham, Ann Eldridge, Marlene Grace, Kerrie McAloney (sample collection); Lisa Bowdler, Steven Crooks (DNA processing); David Smyth, Harry Beeby, Daniel Park (IT support). **Busselton:** The Busselton Health Study acknowledges the numerous Busselton community volunteers who assisted with data collection and the study participants from the Shire of Busselton.

The NHMRC (including grants 613627 and 1036550), Asthma Foundations in Tasmania, Queensland and Victoria, The Clifford Craig Trust in Northern Tasmania, Lew Carty Foundation, Royal Hobart Research Foundation and the University of Melbourne, Cooperative Research Centre for Asthma, New South Wales Department of Health, Children’s Hospital Westmead, University of Sydney.

Contributions of goods and services were made to the CAPS study by Allergopharma Joachim Ganzer KG Germany, John Sands Australia, Hasbro, Toll refrigerated, AstraZeneca Australia, and Nu-Mega Ingredients Pty Ltd. Goods were provided at reduced cost to the CAPS study by Auspharm, Allersearch and Goodman Fielder Foods. MCM, SCD and MAB are supported by the NHMRC Fellowship Scheme.

The Busselton Health Study acknowledges the generous support for the 1994/5 follow-up study from Healthway, Western Australia.

**ALSPAC:** We are extremely grateful to all the families who took part in this study, 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 and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors Nicholas Timpson and John Henderson will serve as guarantors for the contents of this paper. This research was specifically funded by The MRC Centre for Causal Analyses in Translational Medicine (grant G0600705) provided funding for N Timpson.

GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

**B58C:** We acknowledge use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. (http://www.b58cgene.sgul.ac.uk/). Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research.

**COPSAC2000:** We thank all the families participating in the COPSAC cohort for their effort and commitment; and the whole COPSAC study team including computer and laboratory technicians,
research scientists, managers, receptionists, and nurses. We also wish to thank Bjarne Kristensen and Inger Pedersen (Phadia Denmark) for their dedicated work with the IgE-analyses.

COPSAC is funded by: the Lundbeck Foundation, the Danish Council for Strategic Research, , the Pharmacy Foundation, the Danish Agency for Science, Technology and Innovation, the EU Seventh Framework Programme, Ronald McDonald House Charities, the Global Excellence in Health award Programme, the Danish Medical Research Council, the Director K. GAD and family Foundation, the A. P. Møller og Hustru Chastine Mc-Kinney Møller General Purpose Foundation, the Aage Bang Foundation, the Health Insurance Foundation, the East Danish Medical Research Council, the Copenhagen City Council Research Foundation, the Kai and Gunhild Lange Foundation, the Dagmar Marshall Foundation, the Ville Heise legacy, the Region of Copenhagen, the Ib Henriksen foundation, the Birgit and Svend Pock-Steen foundation, the Danish Ministry of the Interior and Health’s Research Centre for Environmental Health, the Gerda and Aage Hensch foundation, the Rosalie Petersen Foundation, the Hans and Nora Buchard Foundation, the Gangsted Foundation, the Danish Medical Association, Asthma-Allergy Denmark, the Danish Otolaryngology Association, the Oda Pedersen legacy, the Højmosegaard Legacy, the A. P. Møller og Hustru Chastine Mc-Kinney Møller Foundation for the advancement of Medical Knowledge, the Jacob and Olga Madsen Foundation, the Aase and Einar Danielsen Foundation, and Queen Louise’s Children’s’ Hospital Research Foundation.

The IgE analyses were sponsored by Phadia ApS.

The funding agencies did not have any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**LISA/GINI:** The authors thank all the families for their participation in the GINIplus and LISAplus studies. Furthermore, we thank all members of the GINIplus and LISAplus Study Groups for their excellent work. The LISAplus Study Group consists of the following: The study team wishes to acknowledge the following: Helmholtz Zentrum Muenchen - German Research Center for Environment and Health, Institute of Epidemiology I, Neuherberg (Heinrich J, Wichmann HE,
Sausenthaler S, Chen C-M; University of Leipzig, Department of Pediatrics (Borte M), Department of Environmental Medicine and Hygiene (Herbarth O); Department of Pediatrics, Marien-Hospital, Wesel (von Berg A); Bad Honnef (Schaaf B); UFZ-Centre for Environmental Research Leipzig-Halle, Department of Environmental Immunology (Lehmann I); IUF – Leibniz Research Institute for Environmental Medicine, Düsseldorf (Krämer U); Department of Pediatrics, Technical University, Munich (Bauer CP, Hoffman U); Centre for Allergy and Environment, Technical University, Munich (Behrendt H).

The GINIplus Study Group consists of the following: The study team wishes to acknowledge the following: Helmholtz Zentrum Muenchen - German Research Center for Environmental Health, Institute of Epidemiology I, Munich (Heinrich J, Wichmann HE, Sausenthaler S, Chen C-M, Thiering E, Tiesler C, Standl M, Schnappinger M, Rzehak P); Department of Pediatrics, Marien-Hospital, Wesel (Berdel D, von Berg A, Beckmann C, Groß I); Department of Pediatrics, Ludwig Maximilians University, Munich (Koletzko S, Reinhardt D, Krauss-Etschmann S); Department of Pediatrics, Technical University, Munich (Bauer CP, Brockow I, Grübl A, Hoffmann U); IUF – Leibniz Research Institute for Environmental Medicine, Düsseldorf (Krämer U, Link E, Cramer C); Centre for Allergy and Environment, Technical University, Munich (Behrendt H).

Personal and financial support by the Munich Center of Health Sciences (MCHEALTH) as part of the Ludwig-Maximilians University Munich LMU innovative is gratefully acknowledged.

**Manchester Asthma and Allergy Study (MAAS):** We would like to thank the children and their parents for their continued support and enthusiasm. We greatly appreciate the commitment they have given to the project. We would also like to acknowledge the hard work and dedication of the study team (post-doctoral scientists, research fellows, nurses, physiologists, technicians and clerical staff).
MAAS was supported by the Asthma UK Grants No 301 (1995-1998), No 362 (1998-2001), No 01/012 (2001-2004), No 04/014 (2004-2007) and The Moulton Charitable Foundation (2004-current); age 11 years clinical follow-up is funded by the Medical Research Council (MRC) Grant G0601361.

NFBC 1966: We thank Professor Paula Rantakallio (launch of NFBC 1966 and 1986), Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking).

Financial support was received from the Academy of Finland (project grants 104781, 120315 and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland, the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council (G0500539, PrevMetSyn/SALVE). The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. A. Couto Alves acknowledges the European Commission, Framework 7, grant number 223367. Jess L Buxton acknowledges the Wellcome Trust fellowship grant, number WT088431MA.

RAINE: The authors are grateful to the Raine Study participants, their families, and to the Raine Study research staff for cohort coordination and data collection. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). The following Institutions provide funding for Core Management of the Raine Study: The University of Western Australia (UWA), Raine Medical Research Foundation, UWA Faculty of Medicine, Dentistry and Health Sciences, The Telethon Institute for Child Health Research, Curtin University, Edith Cowan University and Women and Infants Research Foundation. This study was supported by project grants from the National Health and Medical Research Council of
Australia (Grant ID 403981 and ID 003209; http://www.nhmrc.gov.au/) and the Canadian Institutes of Health Research (Grant ID MOP-82893; http://www.cihr-irsc.gc.ca/e/193.html).

**PIAMA:** The PIAMA study would like to thank all participants, and co-investigators of the study.

The PIAMA study is supported by the Dutch Asthma Foundation (grant 3.4.01.26, 3.2.06.022, 3.4.09.081 and 3.2.10.085CO), the ZonMw (a Dutch organization for health research and development; grant 912-03-031), and the ministry of the environment.

Genome-wide genotyping was funded by the European Commission as part of GABRIEL (A multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract number 018996 under the Integrated Program LSH-2004-1.2.5-1 Post genomic approaches to understand the molecular basis of asthma aiming at a preventive or therapeutic control.
Supplementary Figure 1
Flowchart of the selected known autoimmune disease associated SNPs/loci for lookup in the allergy GWAS identified within the GWA’s catalog (accessed 25th of November 2013). Please also see methods section here in the supplement. A detailed description for each step in the flow chart:
1) The GWA’s catalog\(^2\) were accessed 25th of November 2013.
2) All autoimmune diseases and associations to SNPs were selected with \(p<5 \times 10^{-8}\). The chosen traits were collapsed into 16 overall autoimmune disease categories (see supplemental table 1)
3) We collapsed close SNPs into loci (+/-250kb)\(^6\) and used for each locus only the SNP with lowest reported \(P\) as index SNP and as representative for the specific locus.
4) For several of the SNPs we had to use a proxy SNP as the index SNP were not present within the allergy GWAS. Proxy SNPs were chosen on highest \(r^2\) to index SNP and if two or more proxies had the same \(r^2\) the SNP closest in physical distance to the index SNP were chosen. In total 290 SNPs were available for look up/extraction within the allergy GWAS.

---

**GWAs Catalog**
(25\(^{th}\) of November 2013)

- 15100 phenotype-variant associations:
  - 946 traits, 1757 Papers

**Selecting phenotypes of interest**
collapsed in 16 overall phenotypes with \(p<5 \times 10^{-8}\)

- 870 phenotype-SNP associations:
  - 648 unique SNPs, 108 Papers

**Collapsed into loci**
using as index most significant reported SNP

- 296 loci

**Extracting SNPs**
from sensitization gwas

- 290 SNPs (33 proxies, \(r^2>0.5\))
Supplementary Figure 2
QQ plot of the meta-analysed 2,284,215 SNPs and association to 1) Sensitization 2) Self-reported allergy 3) These two data-sets meta-analysed and 4) Without reported known loci
Supplementary Figure 3
Manhattan plot of the of the meta-analysed 2,284,215 SNPs and association to allergy. Red dots indicate novel loci not described in the discovery papers (grey)\textsuperscript{2,3}, with p<5\times10\textsuperscript{-6}. Dashed line: 10\textsuperscript{-6}. Solid line: 5\textsuperscript{-8}.
**Supplementary Figure 4**
LocusZoom plots of the suggestive novel loci from the allergy meta-analyses
rs9790601

Plotted SNPs:

- SLC39A8
- NFKB1
- MANBA
- UBE2D3
- NHE4C1
- CISD2

Position on chr4 (Mb):

Position of SNPs:

103.4 103.6 103.8 104

Recombination rate (K/Mb):

0 20 40 60 80 100

- $r^2$
- 0.0
- 0.2
- 0.4
- 0.6
- 0.8

Plot of $-\log(p)$ vs. Recombination rate for rs9790601.
Plot of SNP rs848. The graph shows the recombination rate (cM/Mb) on the y-axis and the position on chr5 (Mb) on the x-axis. The SNP rs848 is highlighted, and other SNPs are plotted along with their corresponding genes.
rs7072398

Plotted SNPs

Position on chr10 (Mb)

Recombination rate (cM/Mb)

\[ r^2 \]

\[ 0.2 \]

\[ 0.4 \]

\[ 0.6 \]

\[ 0.8 \]

\[ 1.0 \]
**Supplementary Figure 5**

QQ plots of the autoimmune disease associated loci within the combined allergy meta-analysis as well as allergic sensitization and self-reported allergy separately. The numbers in the figures show enrichment Odds Ratio and P-value for enrichment.
**Supplementary Figure 6**
Separate QQ plots of the autoimmune disease associated loci within the allergy meta-analysis with printed calculated enrichment Odds Ratio and P-value for enrichment. Only plotted for autoimmune diseases with at least 10 loci associated. Solid line reflects the P-value distribution under the null while the dashed is the distribution of all SNPs from the allergy meta-analysis.

**Ankylosing Spondylitis:**

![QQ plot showing OR=3.96, P=1.1e-01](image)

**OR=3.96, P=1.1e-01**
Psoriasis:

OR = 4.75, \( P = 1.6 \times 10^{-2} \)
Arthritis:

OR = 5.18, P = 2e-04
Celiac Disease:

OR=7.85, P=1.7e-06
Primary Biliary Cirrhosis:

\[ \text{OR}=7.8, \quad P=1.4\times10^{-4} \]
Ulcerative colitis:

OR=6.48, P=6.4e-08
Crohn’s Disease:

OR=6.53, P=5e-12
Graves Disease:

OR=4.46, P=4.2e-02
Inflammatory Bowel Disease:

OR=5.51, P=2.6e-16
Systemic Lupus Erythematosus:

$\text{OR}=3.1, \text{ } P=5.3\times10^{-2}$
Multiple Sclerosis:

\[
\text{OR}=5.09, \ P=2.1\times10^{-5}
\]
Type-1 Diabetes:

\[ \text{OR}=4.07, \ P=1.7 \times 10^{-3} \]
Supplementary Figure 7

QQ plot of 57 Migraine loci extracted from the allergy meta-analysis results.

Migraine:
Supplementary Figure 8

QQ plot of 77 loci associated with the combined phenotype of schizophrenia and bipolar disorder extracted from the allergy meta-analysis results.

Bipolar disorder and schizophrenia:
Supplementary Figure 9
Principal component plot of GWAS Catalogue SNPs’ perturbation of gene networks, based on the DEPICT tool, PC1 vs PC3

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 10
Principal component plot of GWAS Catalogue SNPs’ perturbation of gene networks, based on the DEPICT tool, PC1 vs PC2, all trait names.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 11

Allergy related loci and their resemblance to autoimmune disease and other types of disease loci were assessed by principal component analysis by analyzing the tendency of each trait-locus to fall in DHS sites in specific cell lines. This plot shows PC1 vs. PC2 and has the outlier "lipid metabolism phenotypes" omitted, and only names for autoimmune diseases, asthma and allergy are printed. The blue area represents the shared minimal ellipsoid area of immune-mediated diseases.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 12
Allergy related loci and their resemblance to autoimmune disease and other types of disease loci were assessed by principal component analysis by analyzing the tendency of each trait-locus to fall in DHS sites in specific cell lines. This plot shows PC1 vs. PC2 for the full data set.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 13

Allergy related loci and their resemblance to autoimmune disease and other types of disease loci were assessed by principal component analysis by analyzing the tendency of each trait-locus to fall in DHS sites in specific cell lines. This plot shows PC1 vs. PC2 overlayed with cell- and tissue type loadings.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 14
Hierarchical clustering of all NHRGI GWAS catalog diseases’ associated SNPs’ tendency to fall within DHS sites for immune cell types within the Encode data set.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 15
Enrichment of DHS sites in SNPs associated to allergy and Crohn’s disease.
X-axis denominates all SNPs associated to the given trait at $-\log10(p) \leq x$, and y gives the enrichment of DHS sites for a given cell/tissue-type for those SNPs, as compared to all SNPs (x=0). Immune cells are indicated in blue.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 16
Enrichment of SNPs falling in FANTOM enhancers

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 17
PCA plot of DEPICT pathway perturbation analysis, showing names for all gene sets.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 18
Enrichment of shared loci with ENCODE ChIP-seq based transcription factor binding sites. Green line indicates FDR < 0.05. Transcription factors in blue have FDR < 0.05 and enrichment >= 3.

(PLEASE SEE SEPARATE FILE)
Supplementary Figure 19
LocusZoom plots of the autoimmune disease associated loci within the allergy meta-analysis. Each dot represents the association between allergy and the particular SNP. The purple SNP is the index SNP for which the remaining SNPs are colored with respect to the r2 value to the index SNP. The position on the Y-axis represents the P-value (left handside Y-axis). The blue line represents recombination rates (right handside Y-axis)
Plotted SNPs

Position on chr3 (Mb)

Recombination rate (cM/Mb)

-log(p-value)

 rs1464510

FLJ42993

LPP

M2

0.2

0.4

0.6

0.8

1.0

r^2
Plotted SNPs

Position on chr2 (Mb)

rs6738825

- log(p-value)

Recombination rate (cM/Mb)

rs6738825

Position on chr2 (Mb)

RFTN2
BCOL
PLCL1
MARS2

198.4 198.6 198.8 199

100

80

60

40

20

0

0

0.2

0.4

0.6

0.8

1.0

RFTN2
BCOL
PLCL1
MARS2

P os i t i on on c h r 2 ( M b)
Plotted SNPs

rs12708716

Position on chr16 (Mb)

-log(p-value)

Recombination rate (c/Mb)

TEKT5
CIFTA
CLEC16A
SOCS1
C16orf75

NUBP1
DEXI
FAM15A

TNP2
PRM3
PRM2
PRM1

10.8 11 11.2 11.4

Mb
Plotted SNPs

Position on chr14 (Mb)

Recombination rate (cM/Mb)

RAD51L1 → rs911263

Plot r^2 vs. log10(p-value)

0.2 0.4 0.6 0.8 1.0

0 2 4 6 8 10
Plot of SNP positions on chromosome 6 (Mb) showing recombination rate (cM/Mb) with a focus on rs2281388.
Supplementary Figure 20
Paired LocusZoom plots within a Crohn’s data\textsuperscript{17} meta-analysis (top panel) and the allergy meta-analysis (bottom panel) for the 5 most significant shared loci.
**Supplementary Figure 21**

ENCODER Roadmap DHS region overlap with genomic features. DHS regions for each cell type (vertical lines) were overlapped with genomic features (exons, introns, promoters, and intergenic (remaining)) (horizontal lines). Overlaps were z-scaled within each feature, and a heatmap was generated after hierarchical clustering. Immune cells are marked in red at bottom.
Supplementary Figure 22
Association plot for rs11122898 with added enhancer regions for four cell types, as well as enhancer-to-gene regulatory associations, from the FANTOMS data repository\textsuperscript{19}.
Supplement references

Asthma
Celiac disease
Crohn's
Inflammatory bowel disease
Multiple sclerosis
Primary biliary cirrhosis
Psoriasis
Rheumatoid arthritis
Systemic lupus erythematosus
Type 1 diabetes
Ulcerative colitis

Principal component 3: 3.667%
Principal component 1: 10.856%

Principal component 1: 10.856%
Principal component 3: 5.667%
Asthma  IBD  Multiple Sclerosis  Psoriasis  Rheumatoid Arthritis  Allergy  Systemic Lupus Erythematosus  Type1 Diabetes  Ulcerative Colitis  Celiac Disease  Crohn's Disease

IgG Glycosylation  Platelet Count  Hiv1 Control  IR to IL6

Principal Component 1

Principal Component 2

A

B

Allergy

Crohn's Disease

Enrichment

-log10(P)

CD20+ B-cell  T-reg cell  B-cell  CD3+ T-cell

CD3+ T-cell  Th2 T-cell  Thymus  CD14+ monoc.  NK-cell

B
Enrichment vs. -log10(FDR) chart with FDR < 0.01 threshold.