Shared Genetic Architecture between Asthma and Allergic Diseases: A Genome-Wide Cross Trait Analysis of 112,000 Individuals from UK Biobank

Zhaozhong Zhu[1,2,3,4], Phil H. Lee[3,4,5,6], Mark D. Chaffin[5], Wonil Chung[1], Po-Ru Loh[1,7], Quan Lu[2], David C. Christiani[2], Liming Liang[1,7]

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.
[2] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.
[3] Center for Human Genetics Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.
[4] Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.
[5] Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
[6] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
[7] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

Corresponding author: Liming Liang, Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. (Email: lliang@hsph.harvard.edu ; Tel: 617-432-5896; Fax: 617-432-1722)

Authors' contributions:
ZZ, LL, PHL, QL, DCC designed the study.
ZZ, LL, MC, WC, PRL performed the statistical analysis.
ZZ, LL, WC, MC, QL wrote the manuscript.
All authors helped interpret the data, reviewed and edited the final paper, and approved the submission.
ZZ and LL had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests: The authors declare no competing interests.

Abstract

Clinical and epidemiological data suggest that asthma and allergic diseases are associated. And may share a common genetic etiology. We analyzed genome-wide single-nucleotide polymorphism (SNP) data for asthma and allergic diseases in 35,783 cases and 76,768 controls of European ancestry from the UK Biobank. Two publicly available independent genome wide association studies (GWAS) were used for replication. We have found a strong genome-wide genetic correlation between asthma and allergic diseases (r_g = 0.75, P = 6.84×10^{-62}). Cross trait analysis identified 38 genome-wide significant loci, including novel loci such as D2HGDH and GAL2ST2. Computational analysis showed that shared genetic loci are enriched in immune/inflammatory systems and tissues with epithelium cells. Our work identifies common genetic architectures shared between asthma and allergy and will help to advance our understanding of the molecular mechanisms underlying co-morbid asthma and allergic diseases.

Introduction

Asthma is a chronic respiratory syndrome that is characterized by abnormal and inflamed mucosa of the airways, wheezing, and shortness of breath. Allergic diseases are immune responses for allergies, such as allergic rhinitis and atopic dermatitis (eczema). Asthma, allergic rhinitis and eczema all belong to type I hypersensitivity, which is an immune response to foreign antigen and often associates with immunoglobulin E (IgE)-mediated inflammation[1,2]. Genetic studies offer a structured means of understanding the causes of asthma and allergic diseases, as well as identifying targets that can be used to treat the syndrome[3-7].

Clinical and epidemiological studies suggest that asthma and allergy are associated[8,9]. Several studies have identified allergic diseases, such as allergic rhinitis and eczema, as a risk factor for asthma, with the prevalence of allergic rhinitis in asthmatic patients being 80% to 90%[8,10]. These studies demonstrate that the coexistence of asthma and allergy is frequent, and that allergy usually precedes asthma. Also our previous epigenetic study has identified methylation loci linked to asthma and allergy via IgE pathway[11].

One hypothesis to account for the similar symptoms and conditions is that these diseases share a common genetic etiology. Cotsapas et al discovered nearly half of loci in genome-wide association studies (GWAS) of an individual disease influence risk to at least two diseases, indicating the shared genetic architecture of immune-mediated inflammatory and autoimmune diseases[12]. As each of the shared or similar risk factors has strong genetic influences on disease risk, the observed clustering of multiple risk factors could be due to an overlap in the causal genes and pathways[13-20]. In addition, grouping variants by the traits they influence should provide insight into the specific biological processes underlying comorbidity and disease risk.

Clinical and epidemiological studies have found asthma and allergic diseases can occur either in the same individual or in closely related family members[21-24], suggesting potential pleiotropic effect. The heritability has been estimated at varying between 35% and 95% for asthma[21,25] and 33% and 91% for allergic rhinitis, 71% and 84% for atopic dermatitis (eczema), 34% and 84% for serum IgE levels, which suggests that the genetic contribution to them can be significant[21]. While Hinds and colleagues investigated the shared genetic etiology for 38 allergic diseases including asthma[26], understanding has been limited to only 16 genome-wide shared susceptibility loci which were based on self-reported phenotype.

In order to increase our knowledge of shared genetic determinants influencing doctor diagnosed allergic diseases and asthma, and to potentially discover novel loci, we investigated the genetic commonality among asthma and allergic diseases (hay fever/allergic rhinitis or eczema). We conducted a large scale GWAS analysis based on these traits to explore genetic correlations and shared genetic components among these diseases using data from the UK Biobank, which is the largest and most complete European Biobank available at present. Replications were done in two independent public available GWAS studies, the GABRIEL asthma GWAS study[4] and the EArly Genetics & Lifecourse Epidemiology (EAGLE) eczema consortium study[27].

Methods

Study population and design.

Study participants were from the UK Biobank study, described in detail elsewhere[28-30]. In brief, the UK Biobank is a prospective study of > 500,000 people living in the UK. All people in the National Health Service registry who were aged 40–69 years and living <25 miles from a study

center were invited to participate from 2006–2010. In total, 503,325 participants were recruited from over 9.2 million mailed invitations. Self-reported baseline data were collected by questionnaire, and anthropometric assessments were performed. For the current analysis, individuals of non-white ethnicity were excluded to avoid confounding effects. All participants provided informed consent to the UK Biobank.

Overall study design was shown in Figure 1. To identify genetic variants that contribute to doctor-diagnosed asthma and allergic diseases and link them with other conditions, we performed GWAS using phenotype measures in UK Biobank participants (N = 152,249). We removed 39,698 samples that are non-European, related, withdrew from UK Biobank or are missing phenotype information. Thus, a total of 112,551 European descents with high-quality genotyping and complete phenotype/covariate data were used for these analyses, including 25,685 allergic diseases (hay fever/allergic rhinitis or eczema), 14,085 asthma and 76,768 controls for the analysis. There were 53,094 males and 59,457 females in the population.

Furthermore, we have included summary statistics for a replication analysis from an independent asthma GWAS (GABRIEL study, 10,365 cases 16,110 controls)[4] and an independent eczema GWAS (EAGLE eczema consortium, 13,287 cases and 41,345 controls)[27].

Data summary, quality control (QC) and imputation

Detailed genotyping and QC procedures of UK Biobank were described previously (http://biobank.ctsu.ox.ac.uk/) and in supplementary material. In summary, after QC procedures were applied, the current released UK Biobank data contains 806,466 SNPs that passed SNP QC in at least one batch. Post-imputation QC was performed as previously outlined (http://biobank.ctsu.ox.ac.uk/); and genotype imputation procedure provided a total of

73,355,677 SNPs. In addition, we performed stringent QC standards by PLINK 1.90. [31] We selected variants that did not deviate from Hardy-Weinberg Equilibrium (HWE) ($P > 1\times10^{-12}$), per variant missing call rates<5%, per-sample missing rate<40%, minor allele frequency (MAF) > 1% and an imputation quality score (INFO) > 0.8. Quantile-Quantile (QQ) plots were produced and checked for each phenotype. R package GenABEL was used to compute genomic inflation values. A total of 7,489,529 SNPs passed QC on whole genome and with complete information in all three phenotypes, which were eligible for statistical association analyses.

GWAS analysis

We performed additive logistic regression genetic association analysis adjusting for age, sex, genotyping array and ten ancestry principal components using PLINK 1.90[31] to assess association between phenotype and genotype on each individual disease. After association analysis, we applied PLINK clumping function (parameters: --clump-p1 5e-8 --clump-p2 1e-5 --clump-r2 0.2 --clump-kb 500) to determine top loci that are independent to each other, i.e. variants with p-value less than 1×10^{-5}, has r^2 more than 0.2 and less than 500 kb away from the peak will be assigned to that peak's clump. The top loci that were annotated to the same gene were then further combined as one independent loci for downstream analysis. We have used GWAS catalog (https://www.ebi.ac.uk/gwas, search date: 4/1/2017) to identify previously reported genes by other GWAS studies.

LD score regression analysis

Post-GWAS genome-wide genetic correlation analysis by LD score regression (LDSC) [32] was conducted using all UK Biobank SNPs after imputation of 1000 Genomes Project variants and QC [33]. LDSC estimates genetic correlation between two traits (ranging from −1 to 1) from

summary statistics using the facts that the GWAS effect size estimate for each SNP incorporates the effects of all SNPs in linkage disequilibrium with that SNP. SNPs in a high linkage disequilibrium region would have higher $\chi 2$ statistics than SNPs in a low linkage disequilibrium region, and a similar relationship is observed when single-study test statistics are replaced with the product of the *z* scores from two studies of traits with some correlation [32]. LDSC applied a self-estimated intercept during the analysis to account for shared subjects between studies[32].

Cross trait meta-analysis

After assessing genetic correlations among all traits, we applied cross trait GWAS meta-analysis by using the R package Association analysis based on SubSETs (ASSET) to combine the association evidence for asthma and allergic diseases at individual variants[34]. This method combines effect estimate and standard error of the GWAS summary statistics to test hypothesis of association between the SNP with any subset of studies. It can also account the correlation among studies/subjects that might arise due to shared subjects across distinct studies or due to correlation among related traits in the same study by using case-control overlap matrices.

GTEx Tissue Specific Expression Analysis

In order to determine if shared asthma and allergy genes are overly expressed in disease-relevant tissues, we conducted tissue specific expression analysis (TSEA)[35,36]. Gene lists were generated from lists obtained from analyses for asthma, allergic diseases and cross-disease, and by including those which have a matching HUGO Gene Nomenclature Committee (HGNC) name. Genes identified in this way were analyzed for tissue enrichment using publicly available expression data from pilot phase of the Genotype-Tissue Expression project (GTEx)[37] (www.gtexportal.org, 1/31/2013). In the GTEx project, postmortem samples from a wide variety

of tissues and donors have been used for bulk RNA sequencing according to a unified protocol. All samples were sequenced using Illumina 76 base-pair paired-end reads. The pilot phase of the GTEx project consists of 1,839 samples derived from 189 post-mortem subjects, including samples from 45 different tissues, with some tissues offering multiple 'sub-tissue' types. TSEA averaged collapsed reads per kilobase per million mapped reads (RPKM) values for the sub-tissue types (25 'whole-tissue' types). We filtered for unique HGNC IDs (n = 213 shared genes mapped by identified loci from meta-analysis result existing in tissue expression dataset). For each tissue, transcripts from the processed GTEx transcripts that are specifically expressed or enriched have been identified by using the TSEA pSI R package function to calculate the specificity index probability (pSI). Significance level of shared genes between asthma and allergic diseases enriched in each tissue were identified by by Fisher's exact test with Benjamini-Hochberg correction.

Replication analysis

In order to confirm our findings, we have further included two public available independent GWAS data: asthma GWAS, the GABRIEL consortium study [4] and eczema GWAS, the EArly Genetics & Lifecourse Epidemiology (EAGLE) eczema consortium study [27], for LD score regression analysis, cross trait analysis and GTEx tissue expression enrichment analysis. We have used LiftOver tool (http://genome.sph.umich.edu/wiki/LiftOver) to convert asthma GWAS reference genome from hg18 to hg19. ImpG-Summary was used to impute asthma GWAS summary statistics to the 1000 Genomes Project variants (phase 3 release v5)[38]. SNPs with imputation quality r2pred < 0.8 were removed. Cross trait analysis between two replication GWAS datasets was conducted using R package Cross-Phenotype Meta-Analysis (CPMA),

which tests for the hypothesis that each independent SNP has multiple phenotype association and combines the empirical evidence based on their P-values[12].

Overrepresentation Enrichment Analysis

In order to understand the biological insights of the shared genes, we have used the WebGestalt tool [39] to assess overrepresented enrichment of the identified shared gene set between asthma and allergic diseases in KEGG[40] pathways and Gene Ontology (GO) biological process[41,42].

Sensitivity analysis

To estimate the effect of using non-allergic asthma individuals, we performed additive logistic regression analysis adjusting for the same covariates as the primary analysis to assess the association between phenotype and genotype on each individual disease using only non-overlapping samples between asthma and allergic diseases from the UK Biobank (excluding overlapped cases from asthma and allergic diseases, N = 6,177). We carried out cross-disease heritability analysis and meta-analysis to identify the loci that are associated with both diseases. Due to the large reduction in sample size of asthma cases and allergic diseases cases, the power of detecting association between phenotypes and SNP is expected to decrease.

Cross tissue heritability of gene expression

The GTEx project[37] (phs000424.v6.p1) was also used to quantify the cis SNP cross-tissue heritability ($h^2_{g1g2,cis}$) of gene expression across a range of tissues. All estimates were obtained using restricted maximum likelihood (REML) as implemented in Genome-wide Complex Trait Analysis (GCTA) [43]. Only common variants that fall within 1 MB of the transcription start site of a gene were considered in deriving these estimates. Enrichment of $h^2_{g1g2,cis}$ was determined by

comparing average $h^2_{g1g2,cis}$ for a given gene list to 1,000,000 random gene lists of equal size in the same tissue pair.

Results

The genetic correlation was calculated using common SNPs. There was strong positive genetic correlation between asthma and allergic diseases (r_g = 0.75, P = 6.84×10^{-62}). We have also extended our LDSC analysis to another type of immune complex mediated (Type III) disease: rheumatoid arthritis[44], and two types of delayed cell mediated (Type IV) immune diseases: Crohn's disease and ulcerative colitis[45]; however we did not find significant genetic correlation between asthma and other immune traits (rheumatoid arthritis, Crohn's disease and ulcerative colitis), we confirmed the high genetic correlation between Crohn's disease and ulcerative colitis (r_g = 0.57, P = 7.98×10^{-23}) (Table 1). Thus this empirical evidence of shared genetic etiology for asthma and allergic diseases encourages the investigation of common pathophysiology. Estimates of the SNP-based heritability by using GWAS summary statistics are presented in Supplementary Table 1. The proportion of variance in the traits explained by common genetic variants at the observed scale is 7.2% (s.e. 0.7%) for asthma and 7.5% (s.e. 0.7%) for allergic diseases.

The phenotype-genotype association test was carried out on all 112,551 samples and 7,489,529 SNPs from UK Biobank after QC. We identified 32 genome-wide significant (P < 5×10^{-8}) independent loci for asthma and 33 for allergic diseases (See Method, Supplementary Table 2 and 3, Supplementary Fig. 4 and 5). The genomic control parameter λ was 1.13 for asthma and 1.14 for allergic diseases (Supplementary Fig. 1 and 2)[46-48].

For asthma, we confirmed most of the previously reported association loci, including IL1R1, IL1RL1, IL18R1, IL13, SLC25A46, HLA regions, SMAD3, GSDMB and many others (Supplementary Table 2). Out of the 32 independent loci reported here, 7 are novel associations with asthma and 4 of these new loci were within protein-coding genes body (Supplementary Table 2). For allergic diseases, genes such as C11orf30, FLG, SLC25A46, TMEM232, CAMK4, TSLP, WDR36, CLEC16A, DEXI, HLA regions and many others were confirmed in our analysis (Supplementary Table 2). Out of the 33 independent loci identified for allergic diseases, 11 are novel, of which 8 were mapped to protein-coding genes (Supplementary Table 3). Six SNPs were found to exactly overlap between asthma and allergic diseases (rs34290285 on D2HGDH and GAL3ST2, rs7705653 on SLC25A46 and TMEM232, rs12413578 near HV745896, rs7936070 on C11orf30, rs56062135 on SMAD3 and rs10414065 near SLC7A10), and 22 loci contains the same protein-coding genes between asthma and allergic diseases even though not at the same peak SNP.

Among 6 exactly overlapping SNPs, 4 were mapped to protein-coding genes. TMEM232 and SLC25A46 (rs7705653) encode transmembrane protein 232 that belongs to tetraspanin family and a solute carrier protein involved in transport of metabolites respectively. C11orf30 (rs7936070) is associated with total serum IgE levels in asthma[49]; also the locus C11orf30 increases susceptibility to poly-sensitization[50]. SMAD3 (rs56062135) is involved in the development of asthmatic inflammation due to responses of immune cells, such as cytokines and T-helper 2 (Th2) cells, which are known to be important in generating an inflammation that characterizes asthma and allergic disease[51]. The fourth SNP, rs34290285 falls on two novel genes for asthma and allergic diseases, D2HGDH and GAL3ST2, which are responsible for

encoding metabolism related D-2hydroxyglutarate dehydrogenase enzyme[52] and tumor metastasis related Galactose-3-O-sulfotransferase 2 enzyme[53].

Furthermore, we have strengthened this finding in a cross-trait analysis and identified 38 loci contain SNPs with $P < 5\times10^{-8}$ (Fig. 2 and Table 2). The genomic control parameter λ is 1.31 for cross trait meta-analysis (Supplementary Fig. 3). Figure 2 shows the Manhattan plot of the cross-trait analysis between asthma and allergic diseases. The strongest association signal was observed on chromosome 6 at HLA-DQ region (rs9273374, $P_{meta} = 7.87\times10^{-35}$)(Fig. 2 and Table 2), the first asthma-susceptibility locus identified[54], and extended haplotypes encompassing HLA-DQ and HLA-DR have been studied for their effects on specific allergen sensitization[26,55]. The second strongest signal was mapped to C11orf30 on chromosome 11q13 (rs7936070, $P_{meta} = 2.81\times10^{-28}$) (Fig. 2 and Table 2), a gene associated with total serum IgE levels and increased susceptibility to poly-sensitization. This SNP was also an overlapping SNP between asthma and allergic diseases single GWAS model. The third strongest signal was observed on IL1R1 genes (rs72823641, $P_{meta} = 1.58\times10^{-27}$)[56] (Fig. 2 and Table 2). Cytokine receptors have an impact on many cytokine induced immune and inflammatory responses, such as asthma and allergy[4,27,57]. The fourth strongest signal observed on SMAD3 (rs56062135, $P_{meta} = 1.56\times10^{-22}$) is another loci found in both asthma and allergic diseases, which is known to have a role in immune cell response[51]. We have also found the FLG (Filaggrin) gene (rs61816766, $P_{meta} = 4.63\times10^{-12}$) to be associated with both asthma and allergic diseases, which is crucial for maintaining normal skin layer function[58]. Our cross-trait results showed most loci are significant for both diseases, and suggests that it is these shared genetic loci between two diseases driving the overall significant positive genetic correlation.

To understand whether the 38 loci as a group are enriched for expression in certain tissue types, we used the GTEx pilot data[37]. The GTEx TSEA enrichment analysis identified five independent tissues expression that were significantly enriched (after Benjamin-Hochberg correction) for expression of cross-trait associated genes (Fig. 3). The most strongly enriched tissue was part of the integumentary system (skin). The other significantly enriched tissues included esophagus, vagina, lung and whole blood. We also found that the identified shared genes between asthma and allergic diseases have significant cross tissue heritability between each pair of tissues, particularly skin and lung have relatively high cross tissue heritability (Supplementary Fig. 9).

In our replication analysis, we have confirmed a significant level of genetic correlation between asthma and eczema by using two independent GWAS data: the GABRIEL study[4] and EAGLE study[27]. LD score regression results showed there was a strong positive genetic correlation between asthma (GABRIEL) and eczema (EAGLE) (r_g = 0.46, P = 0.0044) (Table 1). We have then used CPMA to investigate the shared genes between asthma and one allergic disease, eczema. Thirteen loci demonstrated genome-wide significance. Of these, genes IL6R, IL1RL1, IL13, IL4, ACSL6, TSLP, IL33, C11orf30 and GSDMB were replicated comparing with the UK Biobank results (Supplementary Table 4). SMAD3 was found as a suggestive gene (rs744910, $P_{\text{meta}} = 1.46\times10^{-7}$) even though it does not reach to the genome-wide significance (Supplementary Table 5). While the phenotypes in our replication phase are not exactly same as UK Biobank's, these results underscore the emerging overlap in genetic bases of asthma and allergy.

GO analysis highlighted several biological processes significantly enriched of the shared genes between asthma and allergic diseases, including epithelium development, keratinization, skin development and T helper 1 type immune response (Fig.4 and Supplementary Table 6). Also

KEGG pathway analysis found the shared genes were significantly enriched in immune related pathway (Supplementary Table 7).

Discussion

In this study, we have identified strong genetic correlation between asthma and allergic diseases (hay fever/allergic rhinitis or eczema) and shared genetic loci between them. Further we have replicated our finding in two independent GWAS studies[4,27]. We have also found shared genetic loci between asthma and allergic diseases were significantly enriched in several tissues, including skin, esophageal, vagina, lung and whole blood tissue expression.

A total of 7 new loci were identified for asthma and 11 for allergic diseases. Of these, *D2HGDH* and *GAL3ST2* are associated with both asthma and allergic diseases, and have significant single tissue expression quantitative trait loci (eQTL) including lung, esophagus mucosa, skin (Supplementary Table 8 and 9). D2HGDH encodes D-2hydroxyglutarate dehydrogenase enzyme regulate alpha-ketoglutarate[52] and can regulate alpha-ketoglutarate levels[59], which also regulate Th1 cell and regulatory T cell[60]. Also Lian et al showed down-regulation of isocitrate dehydrogenase 2 (IDH2) is likely one of the mechanisms underlying altered 5-position of cytosine expression within epidermal basal cell layer[61]. The defection of such layer can also progress to asthma by releasing cytokine into systemic circulation and sensitizing to the lung[62]. GAL3ST2 encodes tumor metastasis related Galactose-3-O-sulfotransferase 2 enzyme53 and is expressed in the human small airway epithelium[63]. We also found rs12413578, near HV745896, a new regulatory genes for human telomerase reverse transcriptase (hTERT) gene expression, to be associated with both traits. Telomerase can be activated by antigen, leading to allergen-

specific immune responses[64]. In terms of asthma, we found rs6461503, near integrin subunit beta 8 (ITGB8) as a novel loci. This gene was previously reported involving in T-cell mediated inflammation and autoimmunity[65,66].

To diminish heterogeneity of significant loci potentially attributable to overlap samples in UK Biobank, we performed sensitivity analysis of excluding allergic asthma subjects (overlap sample between asthma and allergic diseases). When analyzing the 7,908 asthma and 19,508 allergic diseases subjects with 76,768 control subjects, we still observed significant genetic correlation between two traits (r_g = 0.42, P = 6.64×10^{-7}). In meta-analysis, most of the loci still showed robust genome wide shared associations between both traits, such as HLA-DQ, IL1R1, C11orf30, SLC25A46, D2HGDH and many others (Supplementary Table 10 and Supplementary Fig. 7). Although the significance level was reduced, this was mainly due to the loss of sample size for 6,177 individuals. In our sensitivity analysis for effect estimate, which should not be biased by sample size, we observed highly similar results between including and excluding allergic asthma subjects (Supplementary Table 10 and Supplementary Fig. 8).

We found genes that are associated with IgE function can consist of another crucial pathway for atopic diseases, such as asthma and eczema. We have found C11orf30 on chromosome 11 (Supplementary Table 2-4) as another shared genes between asthma and allergic diseases. C11orf30 was reported to be associated with total serum IgE levels and increases susceptibility to poly-sensitization[49]. When individuals are exposed to allergen, IgE are released from the immune system and travel to local organs or tissues to release type 2 cytokines such as IL-4,IL-13 and other inflammatory mediators[67]. These cytokines can further cause allergic diseases and asthma, such as reddish/dry skin and coughing, wheezing or shortness of breath[68,69].

Clinical and epidemiologic observations have shown that immune-mediated inflammatory and autoimmune diseases can be associated due to overlap of genes[12]. Indeed, our LDSC analysis showed high genome-wide genetic correlation between asthma and allergic diseases and identified shared genes between them. However, we did not find significant genetic correlation between asthma and three other immune related diseases, rheumatoid arthritis, Crohn's disease and ulcerative colitis, suggesting that the patterns of genetic architecture among immune related traits can still be distinct. Asthma and allergic diseases both belong to IgE mediated hypersensitivity (Type I). While the other three traits are immune related, rheumatoid arthritis belongs to immune complex mediated (Type III) hypersensitivity, and Crohn's disease and ulcerative colitis belong to delayed cell mediated (Type IV) hypersensitivity. Thus we found strong genetic correlation for the traits within each hypersensitivity category. Different types of hypersensitivity are characterized by different types of gene sets.

Though the association between respiratory diseases like asthma and skin tissue diseases may, at first, seem counterintuitive, exploring the link between the two apparently distinct phenotypes from a gene function perspective reveals a noticeable overlap in underlying genetics and molecular pathways in their etiologies. We have confirmed FLG gene on chromosome 1 (Supplementary Table 2-4) was associated with both allergic diseases and asthma. FLG is important for the formation of the stratum corneum and also for the hydration of this barrier layer[58]. People who have mutations in the FLG gene can be sensitive to external allergens and produce dry and flaky skin[58]. Meanwhile it can activate a strong allergic immune response through peripheral blood to many organs, such as lung, causing the inflammation in the lungs that results in shortness of breath[58].

Our predominant finding in whole blood tissue enrichment indicated that these diseases were caused by some malfunction of the shared immune system. Genes such asCD247 and RUNX3 play important roles in T-helper cells, $CD8^+$ T cells, NK and DCs [70] were previously reported as a risk variant for many immune-related inflammatory diseases, such as asthma, Crohn's disease, ulcerative colitis and eczema[70-73].

Our findings suggest that the shared pathway between asthma and allergic diseases might have significant functions in tissues like skin, esophagus, vagina, lung and whole blood. Skin, esophagus, vagina and lung tissues are all characterized by various type of epithelium or mucosa[74,75], which share similarities in terms of functions and mechanisms. For example, the epithelial cells lining the airway, lungs and skin serve as the primary defense to inflammatory stimuli and antigens infiltrating these tissues through inducing cell signaling pathways in response to contact with the irritants. Indeed, a hallmark of asthma and allergic rhinitis is allergic sensitization by means of epithelial activation in the respiratory tract[76,77]. Also epithelial cells can transdifferentiate into motile mesenchymal cells, which can behave like stem cells and differentiate into multiple cell lineages and migrate to damage tissues to repair[78]. Thus, the enrichment of asthma and allergy genes expressed in epithelial components is not specific to a certain organ or tissue; rather, it seems to be generalizable across epithelium types. Also whole blood is responsible for production of anti-inflammatory cytokines and circulating them in the whole body[79].

We also acknowledge the limitations of our study. First, more allergic diseases will be needed to investigate if genes are specific to certain types of allergy shared with asthma, such as food allergy. Also, our allergic symptom phenotype contained both allergic rhinitis/hay fever and eczema. The allergy genes we found may not be associated with a specific type of allergy.

However, since allergic rhinitis and eczema are both type I hypersensitivity and characterized by epithelial cells, they should share highly similar physiology[80,81]. Finally, our UK Biobank cohort traits are not independent. There are shared cases and complete shared controls between asthma and allergic diseases. However, for all of our analysis, we have applied methods that can adjust potential effect due to overlapping samples; and we have also applied the sensitivity analysis and two independent GWAS data to approve the robustness of shared loci. A surprising finding in replication analysis was we did not observe a significant association in the HLA region. This may be due to different GWAS platforms in the GABRIEL and EAGLE studies, or that 1000 genome imputation was conducted for GABRIEL GWAS summary statistics. This process provided fewer total SNPs that are common for both traits than in UK Biobank, which could be another reason that only part of the shared genes were found in this analysis.

These shared genetic components can provide opportunity to investigate asthma and allergy co-morbidity on molecular level. Drugs targeting shared genetic loci might benefit treatment of both diseases. Our findings can also inspire patients and doctors to be more cautious about co-morbidity of asthma and allergic diseases in the clinical practice.

Conclusion

Understanding the genetic overlap between asthma and allergic diseases can be beneficial to treat disease outcome more efficiently. Our GWAS study has highlighted shared genetic pathway in immune and inflammatory systems and epithelium tissues by asthma and allergic diseases. This work reinforces the idea that asthma and allergy implicate shared common biological processes and open new avenues of treatment for them.

URLs. 1000 Genomes Project, http://www.1000genomes.org/; BEAGLE, http://faculty.washington.edu/browning/beagle/beagle.html; UK Biobank, http://biobank.ctsu.ox.ac.uk/; LiftOver, http://genome.sph.umich.edu/wiki/LiftOver; ImpG, https://github.com/huwenboshi/ImpG; LDSC, https://github.com/bulik/ldsc; GTEx, http://www.gtexportal.org; TSEA, http://genetics.wustl.edu/jdlab/tsea/; WebGestalt, http://www.webgestalt.org/option.php; PLINK, https://www.cog-genomics.org/plink2; GenABEL, http://www.genabel.org/packages/GenABEL; R, https://www.r-project.org/; GWAS catalog, https://www.ebi.ac.uk/gwas; ASSET, https://bioconductor.org/packages/release/bioc/html/ASSET.html; CPMA, http://coruscant.itmat.upenn.edu/software.html; GCTA, http://cnsgenomics.com/software/gcta; GABRIEL consortium, https://www.cng.fr/gabriel/index.html; EAGLE consortium, http://data.bris.ac.uk/data;

Acknowledgments

This research has been conducted using the UK Biobank Resource under application number 16549. We would like to thank the participants and researchers from the UK Biobank who significantly contributed or collected data. We thank to GABRIEL consortium and EAGLE consortium for providing GWAS summary statistic data. We also thank to Dr Vernerri Anttila and Dr Steven Gazal for their statistical advice. We thank Elizabeth A. Loehrer for editorial assistance.

Figure Legend

Figure 1. Overall study population and design. UK Biobank cohort consists of 112,551 subjects, including doctor diagnosed 14,085 asthma subjects and 25,685 allergic diseases, 76,768 controls. Two publicly available independent GWASs, an asthma GWAS (N=26,475) and an eczema GWAS (N=54,632) were used for replication.

Figure 2. Manhattan plot of cross trait meta-analysis. The horizontal axis illustrates chromosome position, the vertical axis illustrates significance level -$\log_{10}$(P-value) of asthma and allergic diseases' shared markers. Total of 38 independent loci was identified to be associated with both diseases. Dashed black line indicate genome-wide significance ($P < 5\times10^{-8}$)

Figure 3. GTEx Tissue Specific Expression Analysis (TSEA). The horizontal axis illustrates tissue expression enrichment P-values after after Benjamin-Hochberg correction. The vertical axis illustrates 25 independent tissue types from GTEx.

Figure 4. Gene Ontology (GO) biological process enrichment analysis. Red box indicates significantly regulated biological process (after false discovery rate).

Reference

1. Lenz, H.J. Management and preparedness for infusion and hypersensitivity reactions. *Oncologist* **12**, 601-9 (2007).
2. Wallace, D.V. *et al.* The diagnosis and management of rhinitis: an updated practice parameter. *J Allergy Clin Immunol* **122**, S1-84 (2008).
3. Torgerson, D.G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887-92 (2011).
4. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-21 (2010).
5. Bunyavanich, S. *et al.* Integrated genome-wide association, coexpression network, and expression single nucleotide polymorphism analysis identifies novel pathway in allergic rhinitis. *BMC Med Genomics* **7**, 48 (2014).
6. 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).
7. Pillai, S.G. *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* **5**, e1000421 (2009).
8. Leynaert, B., Neukirch, F., Demoly, P. & Bousquet, J. Epidemiologic evidence for asthma and rhinitis comorbidity. *J Allergy Clin Immunol* **106**, S201-5 (2000).
9. Brauer, M. *et al.* Air pollution from traffic and the development of respiratory infections and asthmatic and allergic symptoms in children. *Am J Respir Crit Care Med* **166**, 1092-8 (2002).
10. Pariente, P.D., LePen, C., Los, F. & Bousquet, J. Quality-of-life outcomes and the use of antihistamines in a French national population-based sample of patients with perennial rhinitis. *Pharmacoeconomics* **12**, 585-95 (1997).
11. Liang, L. *et al.* An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* **520**, 670-4 (2015).
12. Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet* **7**, e1002254 (2011).
13. Criswell, L.A. *et al.* Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* **76**, 561-71 (2005).
14. Cross-Disorder Group of the Psychiatric Genomics, C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-9 (2013).
15. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).
16. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).

17. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet* **99**, 139-53 (2016).
18. Lane, J.M. *et al.* Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat Genet* (2016).
19. Hobbs, B.D. *et al.* Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* **49**, 426-432 (2017).
20. Emdin, C.A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626-634 (2017).
21. Ober, C. & Yao, T.C. The genetics of asthma and allergic disease: a 21st century perspective. *Immunol Rev* **242**, 10-30 (2011).
22. Cookson, W. The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nat Rev Immunol* **4**, 978-88 (2004).
23. Belsky, D.W. *et al.* Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. *Lancet Respir Med* **1**, 453-61 (2013).
24. Holgate, S.T. The epidemic of allergy and asthma. *Nature* **402**, B2-4 (1999).
25. Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. & Mathews, J.D. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* **142**, 1351-8 (1990).
26. Hinds, D.A. *et al.* A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet* **45**, 907-11 (2013).
27. Paternoster, L. *et al.* Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat Genet* **47**, 1449-56 (2015).
28. Collins, R. What makes UK Biobank special? *Lancet* **379**, 1173-4 (2012).
29. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
30. Allen, N.E., Sudlow, C., Peakman, T., Collins, R. & Biobank, U.K. UK biobank data: come and get it. *Sci Transl Med* **6**, 224ed4 (2014).
31. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
32. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
33. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
34. Bhattacharjee, S. *et al.* A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am J Hum Genet* **90**, 821-35 (2012).
35. Dougherty, J.D., Schmidt, E.F., Nakajima, M. & Heintz, N. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res* **38**, 4218-30 (2010).
36. Xu, X., Wells, A.B., O'Brien, D.R., Nehorai, A. & Dougherty, J.D. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci* **34**, 1420-31 (2014).
37. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
38. Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* **30**, 2906-14 (2014).
39. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77-83 (2013).
40. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

41. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
42. Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049-56 (2015).
43. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2 (2012).
44. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-81 (2014).
45. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-24 (2012).
46. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
47. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* **17**, R122-8 (2008).
48. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).
49. Li, X. *et al.* The C11orf30-LRRC32 region is associated with total serum IgE levels in asthmatic patients. *J Allergy Clin Immunol* **129**, 575-8, 578 e1-9 (2012).
50. Amaral, A.F. *et al.* The locus C11orf30 increases susceptibility to poly-sensitization. *Allergy* **70**, 328-33 (2015).
51. Anthoni, M. *et al.* Smad3 -signalling and Th2 cytokines in normal mouse airways and in a mouse model of asthma. *Int J Biol Sci* **3**, 477-85 (2007).
52. Struys, E.A. *et al.* Mutations in the D-2-hydroxyglutarate dehydrogenase gene cause D-2-hydroxyglutaric aciduria. *Am J Hum Genet* **76**, 358-60 (2005).
53. Seko, A., Nagata, K., Yonezawa, S. & Yamashita, K. Down-regulation of Gal 3-O-sulfotransferase-2 (Gal3ST-2) expression in human colonic non-mucinous adenocarcinoma. *Jpn J Cancer Res* **93**, 507-15 (2002).
54. Marsh, D.G., Meyers, D.A. & Bias, W.B. The epidemiology and genetics of atopic allergy. *N Engl J Med* **305**, 1551-9 (1981).
55. Moffatt, M.F. *et al.* Association between quantitative traits underlying asthma and the HLA-DRB1 locus in a family-based population sample. *Eur J Hum Genet* **9**, 341-6 (2001).
56. Dale, M. & Nicklin, M.J. Interleukin-1 receptor cluster: gene organization of IL1R2, IL1R1, IL1RL2 (IL-1Rrp2), IL1RL1 (T1/ST2), and IL18R1 (IL-1Rrp) on human chromosome 2q. *Genomics* **57**, 177-9 (1999).
57. Gudbjartsson, D.F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* **41**, 342-7 (2009).
58. McLean, W.H. The allergy gene: how a mutation in a skin protein revealed a link between eczema and asthma. *F1000 Med Rep* **3**, 2 (2011).
59. Lin, A.P. *et al.* D2HGDH regulates alpha-ketoglutarate levels and dioxygenase function by modulating IDH2. *Nat Commun* **6**, 7768 (2015).
60. Klysz, D. *et al.* Glutamine-dependent alpha-ketoglutarate production regulates the balance between T helper 1 cell and regulatory T cell generation. *Sci Signal* **8**, ra97 (2015).
61. Lian, C.G. *et al.* Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* **150**, 1135-46 (2012).
62. Demehri, S., Morimoto, M., Holtzman, M.J. & Kopan, R. Skin-derived TSLP triggers progression from epidermal-barrier defects to asthma. *PLoS Biol* **7**, e1000067 (2009).
63. Hackett, N.R. *et al.* RNA-Seq quantification of the human small airway epithelium transcriptome. *BMC Genomics* **13**, 82 (2012).

64. Haruta, Y. *et al.* Activation of telomerase is induced by a natural antigen in allergen-specific memory T lymphocytes in bronchial asthma. *Biochem Biophys Res Commun* **259**, 617-23 (1999).
65. Travis, M.A. *et al.* Loss of integrin alpha(v)beta8 on dendritic cells causes autoimmunity and colitis in mice. *Nature* **449**, 361-5 (2007).
66. Worthington, J.J. *et al.* Integrin alphavbeta8-Mediated TGF-beta Activation by Effector Regulatory T Cells Is Essential for Suppression of T-Cell-Mediated Inflammation. *Immunity* **42**, 903-15 (2015).
67. Novak, N., Kraft, S. & Bieber, T. IgE receptors. *Curr Opin Immunol* **13**, 721-6 (2001).
68. Wynn, T.A. Type 2 cytokines: mechanisms and therapeutic strategies. *Nat Rev Immunol* **15**, 271-82 (2015).
69. Barnes, P.J. Th2 cytokines and asthma: an introduction. *Respir Res* **2**, 64-5 (2001).
70. Lotem, J. *et al.* Runx3 at the interface of immunity, inflammation and cancer. *Biochim Biophys Acta* **1855**, 131-43 (2015).
71. Laprise, C. The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young founder population. *Genes Immun* **15**, 247-55 (2014).
72. Guo, C. *et al.* Chromatin immunoprecipitation and association study revealed a possible role of Runt-related transcription factor 3 in the ulcerative colitis of Chinese population. *Clin Immunol* **135**, 483-9 (2010).
73. Esparza-Gordillo, J. *et al.* A functional IL-6 receptor (IL6R) variant is a risk factor for persistent atopic dermatitis. *J Allergy Clin Immunol* **132**, 371-7 (2013).
74. Presland, R.B. & Dale, B.A. Epithelial structural proteins of the skin and oral cavity: function in health and disease. *Crit Rev Oral Biol Med* **11**, 383-408 (2000).
75. Ganesan, S., Comstock, A.T. & Sajjan, U.S. Barrier function of airway tract epithelium. *Tissue Barriers* **1**, e24997 (2013).
76. Wang, Y., Bai, C., Li, K., Adler, K.B. & Wang, X. Role of airway epithelial cells in development of asthma and allergic rhinitis. *Respir Med* **102**, 949-55 (2008).
77. Rackley, C.R. & Stripp, B.R. Building and maintaining the epithelium of the lung. *J Clin Invest* **122**, 2724-30 (2012).
78. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-96 (2014).
79. Meijer, H., Reinecke, J., Becker, C., Tholen, G. & Wehling, P. The production of anti-inflammatory cytokines in whole blood by physico-chemical induction. *Inflamm Res* **52**, 404-7 (2003).
80. Alonso, L. & Fuchs, E. Stem cells of the skin epithelium. *Proc Natl Acad Sci U S A* **100 Suppl 1**, 11830-5 (2003).
81. Pawankar, R., Mori, S., Ozu, C. & Kimura, S. Overview on the pathomechanisms of allergic rhinitis. *Asia Pac Allergy* **1**, 157-67 (2011).

Table 1. Genetic correlation between immune diseases

Phenotype	Asthma	Allergic diseases	Rheumatoid arthritis	Crohn's disease	Ulcerative colitis	Asthma (Moffat et al)	Eczema (Paternoster et al)
Asthma	1	0.7455	0.0942	0.0428	-0.0514	0.6307	0.4138
Allergic diseases	6.84×10^{-62}	1	0.0042	0.1415	0.0611	0.4738	0.5308
Rheumatoid arthritis	0.0793	0.9259	1	0.067	0.1003	0.0369	-0.0463
Crohn's disease	0.5925	0.0371	0.318	1	0.5711	-0.007	0.119
ulcerative colitis	0.4942	0.3721	0.1032	7.98×10^{-23}	1	-0.0814	-0.0121
Asthma (Moffat et al)	1.09×10^{-11}	1.99×10^{-8}	0.6478	0.9477	0.4753	1	0.4611
Eczema (Paternoster et al)	1.29×10^{-9}	4.04×10^{-15}	0.5846	0.1982	0.9134	0.0044	1

Upper off-diagonal shows genetic correlation score (Rg, ranges from -1 to 1). Lower off-diagnal shows genetic correlation P value.

Table 2. Summary of the 38 genomic loci associated with the asthma and allergic disease in cross trait meta-analysis

SNP	*P* value	N	Position	KB	Peak SNP Position	Ref Allele	Alt Allele	INFO	Odds Ratio	Genes
rs61816766	4.63E-12	10	chr1:151976836..152729815	752.98	152319572	C	T	0.92	1.14	C1orf68,CRCT1,CRNN,FLG,FLG-AS1,FLG2,HRNR,LCE2A,LCE2B,LCE2C,LCE2D,LCE3A,LCE3B,LCE3C,LCE3D,LCE3E,LCE4A,LCE5A,NBPF18P,RPTN,S100A11,TCHH,TCHHL1
rs1214598	5.14E-11	33	chr1:167405418..167437953	32.536	167426424	A	G	0.99	0.95	CD247
rs61815704	5.16E-09	24	chr1:152770613..153094297	323.685	152893891	G	C	0.96	1.14	IVL,LCE1A,LCE1B,LCE1C,LCE1D,LCE6A,SMCP,SPRR1A,SPRR1B,SPRR2A,SPRR2B,SPRR2D,SPRR2E,SPRR2F,SPRR3,SPRR4
rs301805	6.43E-09	31	chr1:8431607..8698277	266.671	8481016	T	G	1	0.96	RERE
rs742230	1.02E-08	19	chr1:25249252..25262022	12.771	25251424	G	A	0.99	0.96	RUNX3
rs12743520	3.83E-08	77	chr1:93037112..93272398	235.287	93037112	A	C	0.99	0.93	EVI5
rs72823641	1.58E-27	151	chr2:102734551..103277862	543.312	102936159	A	T	0.99	0.89	IL1R1,IL1RL1,IL1RL2,IL18R1,IL18RAP,MIR4772,SLC9A2,SLC9A4
rs34290285	5.17E-17	73	chr2:242674803..242740090	65.288	242698640	A	G	0.99	0.93	D2HGDH,GAL3ST2
rs10174949	1.70E-16	59	chr2:8438693..8495097	56.405	8442248	A	G	0.99	0.94	LINC00299
rs4916533	1.66E-08	9	chr3:196359973..196377051	17.079	196373582	T	C	0.98	0.93	NRROS
rs28393318	2.14E-19	200	chr4:38708403..38906951	198.549	38784267	G	A	1	0.92	FAM114A1,MIR574,TLR1,TLR6,TLR10
rs56267605	2.56E-12	196	chr4:123024769..123560939	536.171	123363109	C	A	1	1.05	ADAD1,IL2,IL21,IL21-AS1,KIAA1109
rs1837253	4.38E-21	1	chr5:110401872..110401872	0.001	110401872	T	C	1	0.93	TSLP*
rs7705653	1.12E-19	349	chr5:109800443..110368612	568.17	110142816	G	A	0.98	1.14	SLC25A46,TMEM232
rs6881270	1.53E-14	71	chr5:35803577..36029754	226.178	35879095	T	C	1	0.91	CAPSL,IL7R,LOC100506406,SPEF2,UGT3A1
rs2706362	3.75E-11	96	chr5:131881605..131998413	116.809	131925187	C	T	0.99	1.06	IL13,RAD50
rs10074523	2.37E-10	132	chr5:132009154..132128644	119.491	132060583	C	A	0.99	1.10	CCNI2,IL4,KIF3A,SEPT8
rs2548992	4.54E-10	88	chr5:131561758..131833599	271.842	131808668	A	G	0.99	1.10	C5orf56,IRF1,LOC553103,MIR3936,P4HA2,PDLIM4,SLC22A4,S

										LC22A5
rs9273374	7.87E-35	65	chr6:32522576..32804570	281.995	32626614	A	G	0.99	0.84	HLA-DOB,HLA-DQA1,HLA-DQA2,HLA-DQB1,HLA-DQB2,HLA-DRB1,HLA-DRB6,TAP2
rs6461503	3.19E-10	89	chr7:20501204..20628847	127.644	20560996	T	C	0.99	1.05	ITGB8*
rs2136016	2.11E-09	95	chr8:81246659..81315490	68.832	81300681	G	A	0.99	1.09	ZBTB10*
rs9775039	4.42E-22	109	chr9:6086913..6535500	448.588	6177453	A	G	0.99	1.18	GLDC,IL33,TPD52L3,UHRF2
rs2169282	1.80E-10	42	chr9:6348295..6542674	194.38	6350235	A	G	0.99	1.09	GLDC,UHRF2
rs12413578	1.09E-14	24	chr10:9043457..9153428	109.972	9049253	T	C	1	0.91	HV745896*
rs61839660	2.30E-11	47	chr10:6061499..6185344	123.846	6094697	T	C	1	1.12	IL2RA,RBM17
rs10795656	4.07E-10	36	chr10:8565990..8605553	39.564	8595839	A	G	0.99	1.05	GATA3*
rs7936070	2.81E-28	77	chr11:76020398..76302073	281.676	76293527	T	G	1	1.08	C11orf30,LOC100506127,PRKRIR
rs659529	6.03E-11	192	chr11:111366844..111756286	389.443	111436896	T	A	1	0.95	ALG9,BTG4,C11orf1,C11orf88,FDXACB1,LAYN,MIR34B,MIR34C,PPP2R1B,SIK2
rs10876864	1.41E-13	61	chr12:56364321..56518408	154.088	56401085	G	A	1	1.05	CDK2,ERBB3,IKZF4,PA2G4,RAB5B,RPL41,RPS26,SUOX,ZC3H10
rs1059513	7.65E-13	15	chr12:57324513..57510785	186.273	57489709	C	T	1	0.92	GPR182,MYO1A,NAB2,RDH16,SDR9C7,STAT6,TAC3,TMEM194A,ZBTB39
rs8008961	1.24E-08	66	chr14:68739301..68863020	123.72	68752643	T	C	0.99	1.05	RAD51B
rs56062135	1.56E-22	25	chr15:67437863..67466599	28.737	67455630	T	C	1	1.16	SMAD3
rs36045143	1.83E-21	259	chr16:11031741..11297273	265.533	11224966	G	A	1	0.93	CLEC16A,DEXI
rs869402	4.15E-17	225	chr17:37810218..38134889	324.672	38068043	C	T	1	1.12	ERBB2,GRB7,GSDMA,GSDMB,IKZF3,LRRC3C,MIEN1,MIR4728,ORMDL3,PGAP3,PNMT,STARD3,TCAP,ZPBP2
rs9911533	9.70E-16	104	chr17:38729588..38875684	146.097	38775476	C	T	1	0.92	KRT24,KRT222,SMARCE1
rs10414065	2.63E-10	5	chr19:33721455..33726578	5.124	33721455	T	C	0.98	0.91	SLC7A10*
rs2766664	8.07E-11	14	chr20:52170783..52186784	16.002	52171241	A	G	0.99	1.08	LOC101927770,ZNF217
rs3208007	3.22E-10	133	chr20:62260636..62400021	139.386	62322288	T	C	0.99	0.95	ARFRP1,LIME1,RTEL1,RTEL1-TNFRSF6B,SLC2A4RG,STMN3,TNFRSF6B,ZBTB46,ZGPAT

*The nearest gene to this locus

N, number of SNPs clumped with peak SNP; INFO, imputation score;

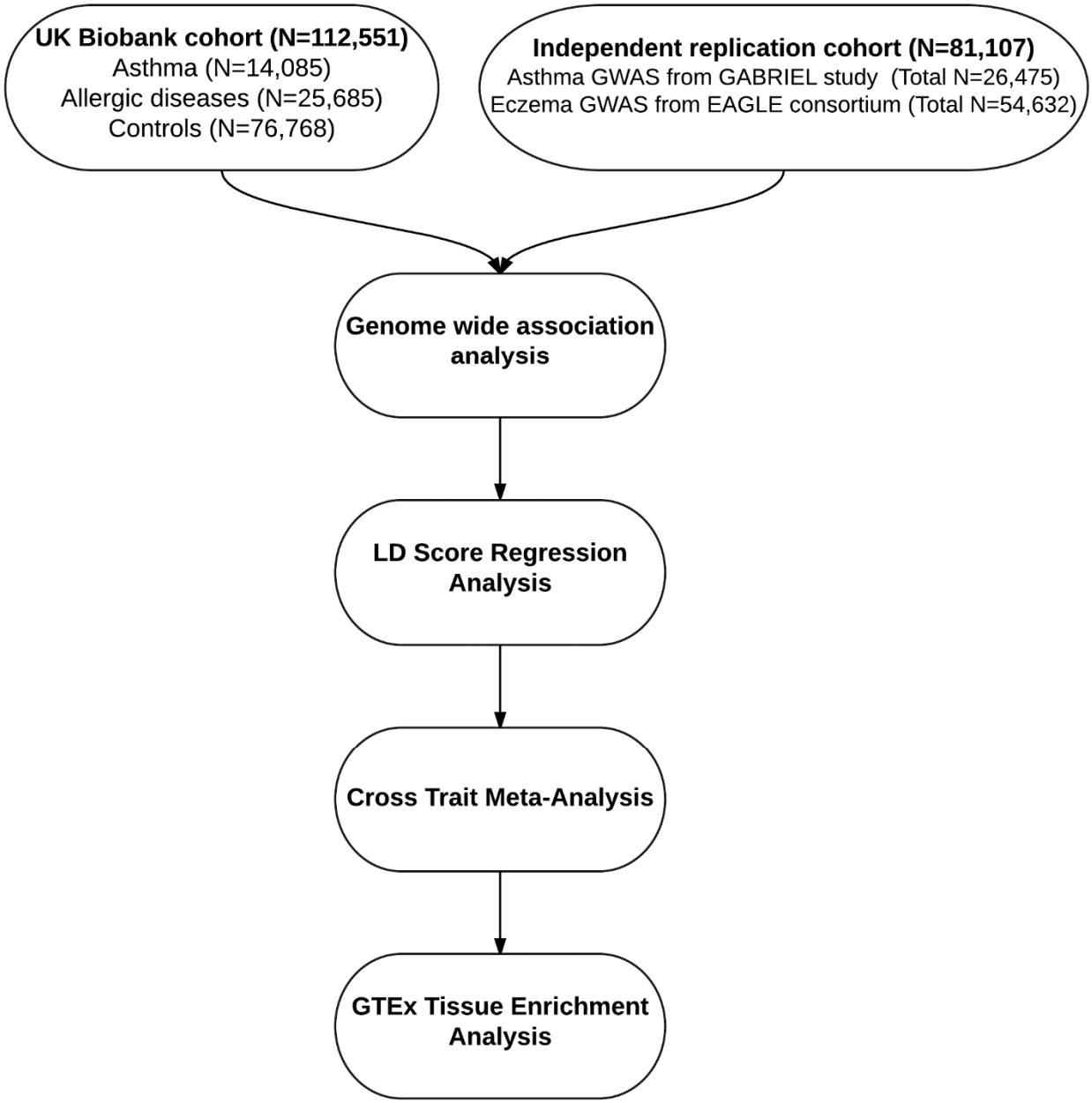
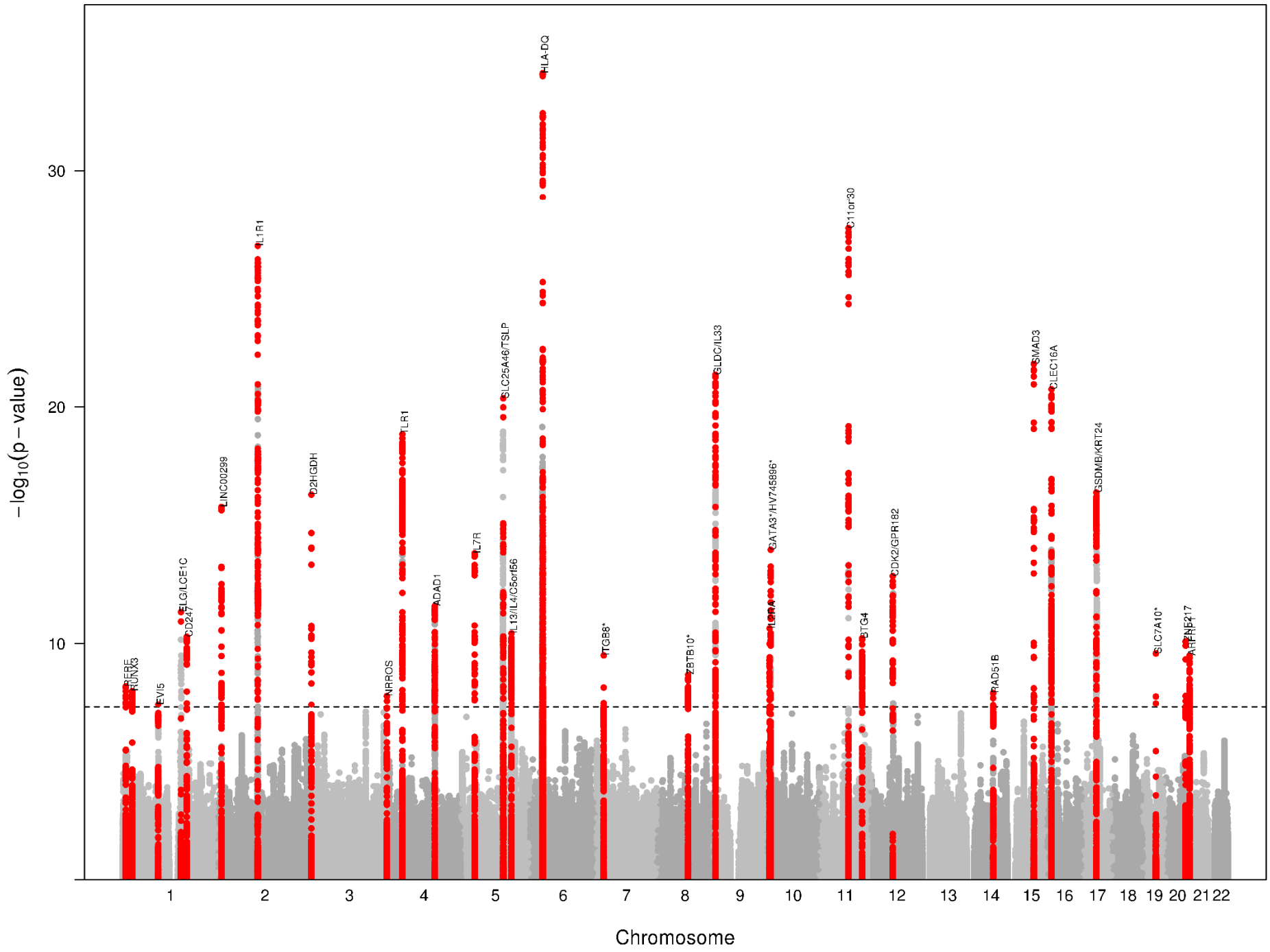


Figure 1. Overall study population and design. UK Biobank cohort consists of 112,551 subjects, including doctor diagnosed 14,085 asthma subjects and 25,685 allergic diseases, 76,768 controls. Two publicly available independent GWASs, an asthma GWAS (N=26,475) and an eczema GWAS (N=54,632) were used for replication.



−log10(p − value)
Chromosome
30
20
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
RERE
RUNX3
EVI5
ELG/LCE1C
CD247
LINC00299
IL1R1
D2HGDH
NRROS
TLR1
ADAD1
IL7R
SLC25A46/TSLP
IL13/IL4/C5orf56
HLA-DQ
TGB8*
ZBTB10*
GLDC/IL33
GATA3*/HV745896*
IL2RA
C11or30
BTG4
CDK2/GPR182
RAD51B
SMAD3
CLEC16A
GSDMB/KRT24
SLC7A10*
ARFRP1

Figure 2. Manhattan plot of cross trait meta-analysis. The horizontal axis illustrates chromosome position, the vertical axis illustrates significance level -$\log_{10}$(P-value) of asthma and allergic diseases' shared markers. Total of 38 independent loci was identified to be associated with both diseases. Dashed black line indicate genome-wide significance ($P < 5\times10^{-8}$)

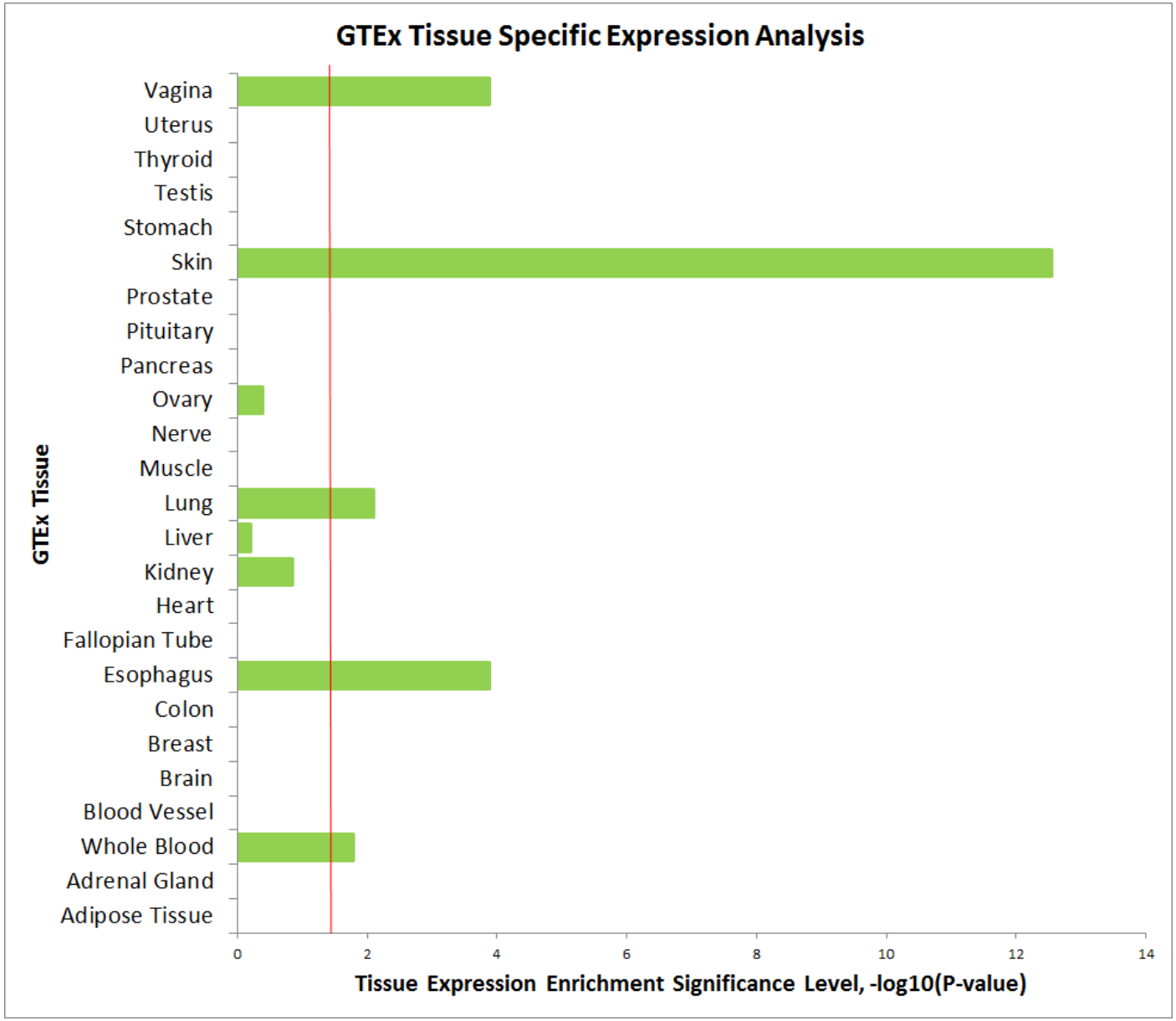


Figure 3. GTEx Tissue Specific Expression Analysis (TSEA). The horizontal axis illustrates tissue expression enrichment P-values after after Benjamin-Hochberg correction. The vertical axis illustrates 25 independent tissue types from GTEx.

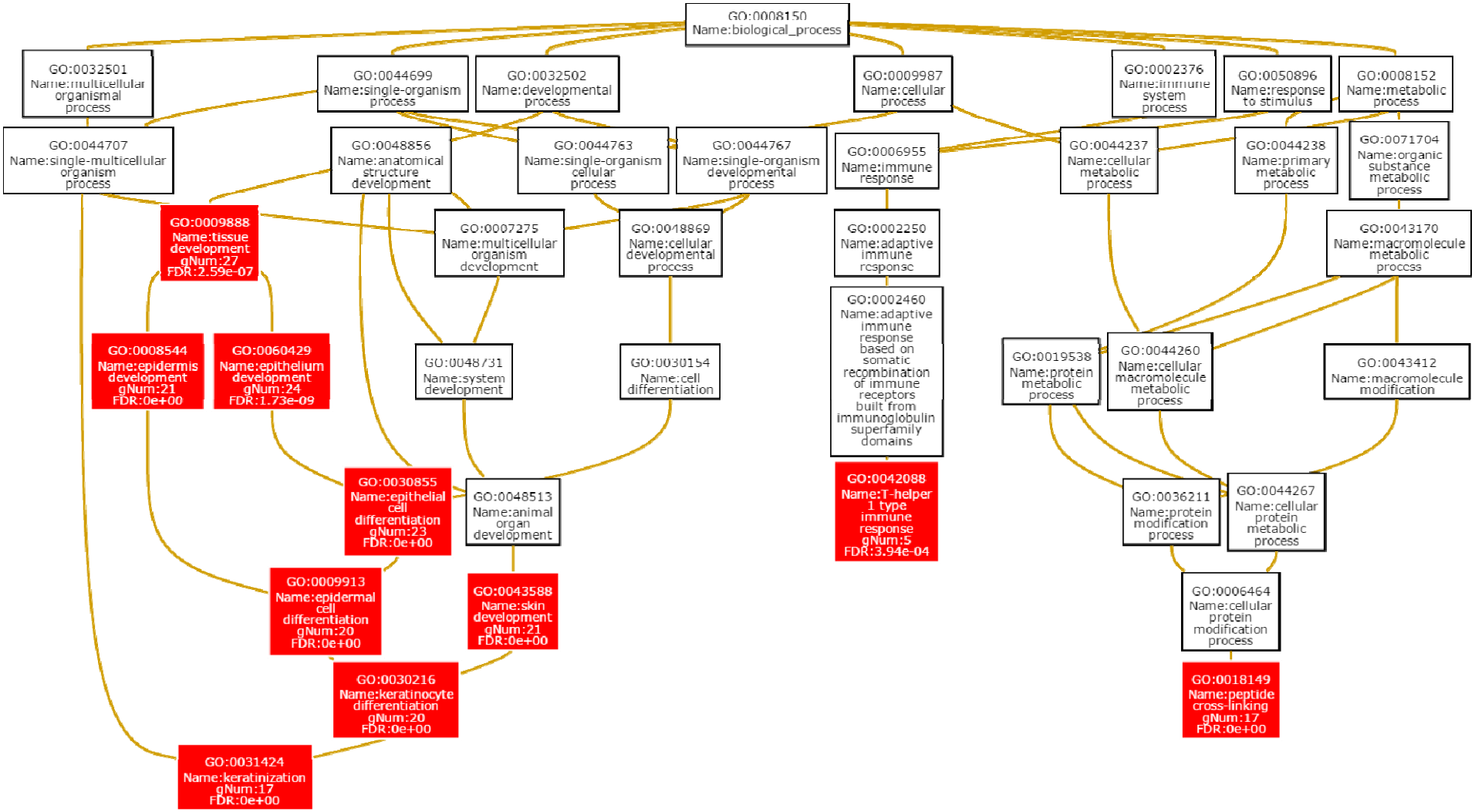


Figure 4. Gene Ontology (GO) biological process enrichment analysis. Red box indicates significantly regulated biological process (after false discovery rate).