Integrated gene-based and pathway analyses using UK Biobank data identify novel genes for chronic respiratory diseases

Gene 767 (2021) 145287

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Integrated gene-based and pathway analyses using UK Biobank data identify novel genes for chronic respiratory diseases

Lijuan Wang [a,1], Rui Fang [a,1], Meng Zhu [a,1], Na Qin [a], Yuzhuo Wang [a], Jingyi Fan [a], Qi Sun [a], Mengmeng Ji [a], Xikang Fan [a], Junxing Xie [a], Hongxia Ma [a,b], Juncheng Dai [a,b,*]

[a] Department of Epidemiology and Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China
[b] Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing 211166, China

ARTICLE INFO

Keywords:
Chronic respiratory diseases
Gene-based analysis
Functional annotation
Pathway enrichment analysis

ABSTRACT

Background: Chronic respiratory diseases have become a non-negligible cause of death globally. Although smoking and environmental exposures are primary risk factors for chronic respiratory diseases, genetic factors also play an important role in determining individual's susceptibility to diseases. Here we performed integrated gene-based and pathway analyses to systematically illuminate the heritable characteristics of chronic respiratory diseases.

Methods: UK (United Kingdom) Biobank is a very large, population-based prospective study with over 500,000 participants, established to allow detailed investigations of the genetic and nongenetic determinants of the diseases. Utilizing the GWAS-summarized data downloaded from UK Biobank, we conducted gene-based analysis to obtain associations of susceptibility genes with asthma, chronic obstructive pulmonary disease (COPD) and pneumonia using FUSION and MAGMA software. Across the identified susceptibility regions, functional annotation integrating multiple functional data sources was performed to explore potential regulatory mechanisms with INQUISIT algorithm. To further detect the biological process involved in the development of chronic respiratory diseases, we undertook pathway enrichment analysis with the R package (clusterProfiler).

Results: A total of 195 susceptibility genes were identified significantly associated with chronic respiratory diseases ($P_{bonferroni} < 0.05$), and 24/195 located out of known susceptibility regions (e.g. *WDPCP* in 2p15). Within the identified susceptibility regions, functional annotation revealed an aggregation of credible variants in promoter-like and enhancer-like histone modification regions and such regulatory mechanisms were specific to lung tissues. Furthermore, 110 genes with INQUISIT score ≥ 1 may influence diseases susceptibility through exerting effects on coding sequences, proximal promoter and distal enhancer regulations. Pathway enrichment results showed that these genes were enriched in immune-related processes and nicotinic acetylcholine receptors pathways.

Conclusions: This study implemented an integrated gene-based and pathway strategy to explore the underlying biological mechanisms and our findings may serve as promising targets for future clinical treatments of chronic respiratory diseases.

1. Background

With an aging global population, chronic respiratory diseases are growing up to be a more prominent cause of death and disability (GBD 2015 Chronic Respiratory Disease Collaborators, 2017). According to the data from the Global Burden of Diseases (GBD) 2017, chronic respiratory diseases have caused 3.91 million deaths, accounting for 15.8% of all-aged deaths globally (GBD 2017 Causes of Death Collaborators, 2018). Among chronic respiratory diseases, chronic obstructive pulmonary disease (COPD) and asthma are common obstructive diseases characterized by persistent airflow limitation and decline of lung function (Drazen et al., 2015). Infectious lung diseases are mainly accompanied by a massively activated inflammatory response (Kellum et al., 2007). Although smoking and environmental exposures are primary risk

* Corresponding author at: Nanjing Medical University, 101 Longmian Road, Nanjing 211166, China.
E-mail address: djc@njmu.edu.cn (J. Dai).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.gene.2020.145287
Received 27 June 2020; Received in revised form 15 October 2020; Accepted 27 October 2020
Available online 10 November 2020

factors for chronic respiratory diseases, genetic factors also play an important role in determining individual's susceptibility to diseases (Duffy et al., 1990; Lichtenstein et al., 2000; Salvi and Barnes, 2009), which could inform drug target identification, risk prediction, and stratified prevention or treatment.

Previous genome-wide association studies (GWASs) have identified dozens of variants associated with chronic respiratory diseases (Moffatt et al., 2010; Rautanen et al., 2015; Wain et al., 2017). However, genome-wide significant loci only account for a small proportion of the genetic variants, which is insufficient to dissect the complex genetic structure. Besides, a single SNP typically has only mild effects while the common diseases are often influenced by the joint effects of multiple loci within a gene or the joint action of multiple genes within a pathway (Lesnick et al., 2007). Thus, by integrating the effects of a group of genetic variants, the gene-based and pathway enrichment analyses can help us holistically unravel the mechanisms of complex diseases (Zhong et al., 2010).

Recently, alternative approaches were developed to perform the gene-based analysis. One way was to integrate functional data with GWAS association results to explore the underlying biological mechanisms. For example, Gusev *et al.* introduced a method, referred as transcriptome-wide association study (TWAS), integrating gene expression data with large scale GWAS data to estimate the association of each gene to disease (Gusev et al., 2016). Another additional approach was to aggregate variants to the level of the whole gene and test the joint association of all variants in the gene with the phenotype. For instance, VEGAS, takes the full set or a defined subset of markers within a gene and accounts for (linkage disequilibrium) LD by using simulations from the multivariate normal distribution (Liu et al., 2010; Mishra and Macgregor, 2015), MAGMA employs multiple linear regression to properly incorporate LD between markers (de Leeuw et al., 2015) and Pascal computes sum and maximum of chi-squared statistics that measure the average and the strongest association signals per gene (Lamparter et al., 2016) to obtain gene-based *P*-values. MAGMA is computationally efficient with better statistical performance compared to other existing gene-based methods. Thus, in this work, we carried out the gene-based analysis using TWAS strategy and MAGMA software to identify novel susceptibility genes for chronic respiratory diseases. Then, the identified genes were utilized to perform pathway enrichment analysis to explore the potential biological process. Our study provides important insight into a deeper understanding of the pathogenesis involved in chronic respiratory diseases.

2. Methods

2.1. Study design and data preparation

2.1.1. UK Biobank

The UK Biobank (UKB) cohort is a major data resource that contains genetic as well as a wide range of phenotypic data of ~500,000 participants of European ancestry aged 39–73 years at recruitment (Bycroft et al., 2018). Genotyping was conducted using the Affymetrix UK BiLEVE Axiom or Affymetrix UK Biobank Axiom array. These arrays were augmented by imputation of ~90 million genetic variants from the 1000 Genomes and UK10K projects, Haplotype Reference Consortium. For detailed information of cohorts, genotyping, imputation, quality control approaches, and association analysis please refer to the published studies (Sudlow et al., 2015; Canela-Xandri et al., 2018).

2.1.2. Phenotype selection

We paid attention to the susceptibility mechanisms modified by genetic mutations in chronic lower respiratory diseases. Phenotypes with more than 5000 cases were selected for subsequent analyses, including asthma with 28,628 cases, COPD with 9266 cases and pneumonia with 9774 cases. All individuals clinically defined from hospital episode statistics were coded as cases, while all other individuals were considered as controls (Canela-Xandri et al., 2018).

2.1.3. GWAS summary datasets

We downloaded the summarized data of genome-wide imputed variants from the GeneATLAS website (http://geneatlas.roslin.ed.ac.uk). We included variants with the quality score of imputation >0.9 and 13,324,371 SNPs remained. We further performed the quality control based on European population from 1000 Genomes Project (Phase 3) with the following criteria: (i) variants having a minor allele frequency (MAF) > 0.01; (ii) variants satisfying the Hardy-Weinberg equilibrium ($P > 1.0 \times 10^{-6}$). After the procedure of quality control, 7,330,104 variants were finally retained for the following analysis. The study design is shown in Fig. 1.

2.2. Gene-based analysis

2.2.1. TWAS analysis using FUSION

We downloaded FUSION (Gusev et al., 2016) software from http://gusevlab.org/projects/fusion/. FUSION firstly estimated the heritability of genes explained by *cis*-SNPs (SNPs within 1 MB region surrounding the TSS) based on individual genotype data derived from the Genotype-Tissue Expression Project (GTEx) database and restricted TWAS analysis to include *cis*-heritable genes ($P_{\text{heritability}} < 0.01$). Then, for these eligible genes, FUSION constructed the effect sizes of *cis*-SNPs on gene expression (i.e. expression weights) using gene expression data in lung tissue in GTEx database with several predictive linear models (Elastic Net, LASSO, GBLUP, and BSLMM). The prepackaged expression weights can be downloaded directly from the FUSION website. Last, FUSION imputed the *cis*-genetic component of expression into large scale GWAS data based on expression weights from the training data while accounting for linkage disequilibrium (LD) among SNPs, and tested the associations between the predicted gene expression and traits of interest. For each gene, FUSION estimated the z-score of the expression and a complex trait (Z_{TWAS}) as a linear combination of the vector of GWAS summary Z scores at a given *cis*-locus with expression weight vector W derived from the reference panels. However, the imputed z-score of expression and trait (WZ) has variance WVW^t, where V is a covariance matrix across SNPs at the locus (i.e., LD), as:

$$Z_{TWAS} = \frac{WZ}{\sqrt{WVW^t}}$$

2.2.2. MAGMA's gene association test

The SNP-based *P* values were used for gene-based analysis using MAGMA (de Leeuw et al., 2015) software (http://ctg.cncr.nl/software/magma), a novel tool for gene and gene-set analysis. Total 19,427 protein-coding genes from the database (NCBI 37.3) were used for SNP annotation. Then, MAGMA used a multiple regression approach to properly incorporate LD between markers and to detect multi-marker effects for a genome-wide gene association analysis.

For both two gene-based methods, we applied a stringent Bonferroni correction to account for multiple testing and associations with $P_{\text{bonferroni}} < 0.05$ were considered as statistically significant.

2.3. Functional annotation

2.3.1. Functional enrichment of defined CRVs

We firstly identify susceptibility genes of these three respiratory traits using gene-based analysis. Then, we focused on SNPs located within 1 MB region surrounding the associated genes and defined SNPs with *P* values within two orders of magnitude of the most significant SNP in each locus as credible risk variants (CRVs). To investigate the enrichment of CRVs in chromatin regulatory regions, Fisher's exact test was used to estimate the distribution of the above CRVs in active promoter and enhancer regions defined in human lung fibroblast cell line NHLF and human lung cancer cell line A549 by calculating the fold-

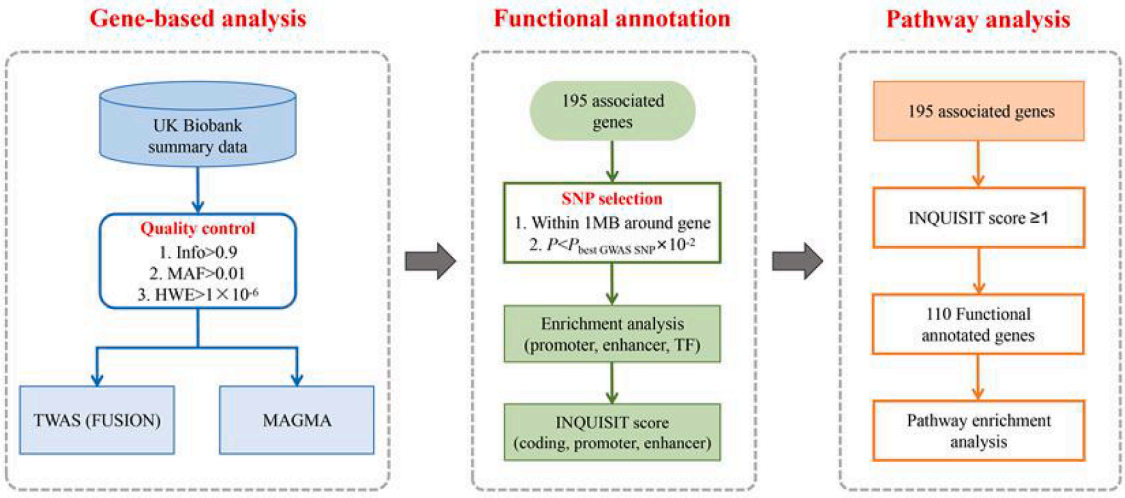

Fig. 1. Flowchart for the study design. (1) Gene-based analysis. GWAS-summarized data downloaded from UKB website was used to perform gene-based analysis using FUSION and MAGMA softwares. Standard quality control was conducted for eligibility. (2) Functional annotation. A total of 195 genes were identified significantly associated with chronic respiratory disease risk ($P_{bonferroni} < 0.05$). Functional annotation of these genes was performed using multiple functional data sources (promoter, enhancer, TF) with INQUISIT algorithm. (3) Pathway analysis. A total of 110 genes with INQUISIT score ≥1 were included in pathway enrichment analysis using R package (clusterProfiler).

enrichment against the background of 1000 genomes (other SNPs in the defined locus). Chromatin state data in four human cell types of other tissues including human skeletal muscle myoblast (HSMM), human embryonic stem cell (HESC), normal human epidermal keratinocyte (NHEK), and lymphoblastoid (GM12878) were also included in our analysis for comparison. All the histone modifications of promoter (H3K4me3 and H3K9ac) and enhancer marks (H3K4me1 and K3K27ac) were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/).

2.3.2. *Functional annotation using INQUISIT algorithm*

We then performed functional annotation of identified genes to further detect the underlying regulation mechanisms using the INQUISIT algorithm (Michailidou et al., 2017). We calculated a score for each gene by assessing the potential impact of each CRV on distal regulation, proximal regulation and the coding sequences. The INQUISIT score was contributed with multiple lines of evidence including Hi-C chromatin interaction information, enhancer-target gene predictions, topologically associated domains (TAD), histone modification marks, transcription factor binding sites, expression quantitative trait loci (eQTLs) in lung-related tissues and coding impact predictions. For the distally regulated gene, if the CRV was located in an enhancer element predicted by FANTOM5 (Andersson et al., 2014) or PreSTIGE (Corradin et al., 2014), or resided in an enhancer element containing transcription factor binding sites (TFBS), or was an eQTL for that gene based on the GTEx database, one score was added. In particular, two scores were added if the CRV was also located in an enhancer element that predicted by Hi-C experiment. However, if the gene was separated from CRV by a TAD boundary, the score was down-weighted by multiplying the sum of the above by 0.05. Likewise, for the proximally regulated gene, we added one score if the CRV was located in a promoter region represented by H3K4me3 or H3K9ac histone modification marks, or the promoter histone modification peak that the CRV resided in was also interacted with the TFBS, or was an eQTL of that gene in GTEx database. For coding impact, if the CRV located in the exonic region had an impact on protein function including non-synonymous, deleterious missense and truncating, we grouped the associated gene to the coding category and added one score to that gene. Finally, expression level was also considered for all identified genes and we multiplied scores by 0.1 when genes showed a lack of expression in the GTEx normal lung tissues.

2.3.3. *Functional exploration for the best GWAS SNP*

To investigate the genetic effects of variants on gene regulation, we conducted functional annotation for the most significant GWAS SNP within cis-locus (1 MB region surrounding the susceptibility genes) in each novel susceptibility region. We annotated the best GWAS SNPs as well as variants in high LD relationships ($r^2 > 0.8$) with them to promising genes using ANNOVAR software (Wang et al., 2010). The functional effects of missense variations were predicted using the SIFT (Ng and Henikoff, 2003) and PolyPhen (Adzhubei et al., 2010) databases. To investigate the potential function of the association at non-coding regions, we utilized data from the GTEx database (http://www.gtexportal.org/) to perform the eQTL analyses in 383 lung tissue samples. To further map the variants to potential regulatory elements, we annotated SNPs according to the histone Chip-seq (H3K4me3, H3K9ac, H3K4me1 and K3K27ac) peaks and DNase I hypersensitivity sites (DHS) based on features measured in NHLF cell line.

2.4. *Pathway enrichment analysis*

We performed pathway enrichment analysis on genes with integrated score ≥1 defined by INQUISIT algorithm to further explore the biological process. Considering the genetic correlations between these three traits based on the UK Biobank data (asthma-COPD: rG = 0.288; asthma-pneumonia: rG = 0.138; COPD-pneumonia: rG = 0.245), we conducted the pathway analysis using combined genes as input. We used the Reactome Pathway Database (Croft et al., 2011) as a reference, which was implemented in R package "clusterProfiler" (Yu et al., 2012). Bonferroni method was used for multiple correction and pathways with adjusted *P*-value < 0.05 were considered statistically significant.

3. Results

3.1. *Gene-based analysis identified significant susceptibility genes*

In the gene-based analysis, we totally identified 221 significant associations for the three respiratory diseases at Bonferroni *P*-value < 0.05, including 195 unique genes (Supplementary Table 1). Of these genes, 24 were located in novel susceptibility regions independent of know regions identified by GWAS or candidate gene strategies (Table 1). We performed Venn diagrams to visually present the overlapped susceptibility genes of these three respiratory traits identified by two gene-

Table 1
Significant genes in novel susceptibility regions identified by gene-based analysis.

Method	Trait	Region	Gene	Chr	Locus start	Locus end	Gene-based P value	Bonferroni P value
TWAS	Asthma	8p23.1	*FAM85B*	8	8,025,341	8,084,136	2.53E-06	1.96E-02
	COPD	9q33.3	*PBX3*	9	128,509,624	128,729,656	8.91E-08	6.91E-04
MAGMA	Asthma	1p36.12	*LACTBL1*	1	23,279,536	23,291,831	2.48E-06	4.45E-02
		2p15	*WDPCP*	2	63,348,518	64,054,977	1.19E-06	2.14E-02
		3q26.32	*TBL1XR1*	3	176,737,143	176,915,261	1.04E-06	1.87E-02
		6q25.1	*ZC3H12D*	6	149,768,794	149,806,197	2.18E-06	3.92E-02
		7p22.3	*ADAP1*	7	940,573	995,043	4.69E-08	8.44E-04
		8p23.1	*MFHAS1*	8	8,640,864	8,751,155	9.95E-07	1.79E-02
		9q33.3	*DENND1A*	9	126,145,934	126,692,431	1.09E-06	1.96E-02
		10p12.31	*MLLT10*	10	21,823,094	22,032,559	4.47E-07	8.04E-03
		11q12.2	*TMEM258*	11	61,535,973	61,558,075	4.01E-08	7.20E-04
			FEN1	11	61,562,813	61,564,716	4.23E-07	7.61E-03
		13q12.11	*PSPC1*	13	20,248,896	20,357,142	9.61E-07	1.73E-02
		13q32.3	*UBAC2*	13	99,853,028	100,038,688	6.18E-10	1.11E-05
			GPR183	13	99,946,784	99,959,659	4.46E-07	8.01E-03
		14q32.12	*RIN3*	14	92,980,118	93,155,339	8.82E-12	1.59E-07
		17q21.32	*ZNF652*	17	47,366,568	47,439,478	2.34E-07	4.20E-03
	COPD	8p23.1	*XKR6*	8	10,753,555	11,058,875	1.13E-06	2.03E-02
			SOX7	8	10,587,706	10,588,022	1.34E-06	2.40E-02
			PINX1	8	10,622,473	10,691,291	1.39E-06	2.50E-02
		9q33.3	*PBX3*	9	128,509,624	128,729,656	2.03E-09	3.65E-05
		9q34.13	*MED27*	9	134,735,494	134,955,295	8.46E-08	1.52E-03
	Pneumonia	2p15	*AHSA2*	2	61,404,553	61,413,216	1.06E-06	1.91E-02
		15q25.1	*IREB2*	15	78,729,773	78,793,798	1.17E-06	2.11E-02

based approaches (Fig. 2). Manhattan plots of genes as well as SNPs associations are presented in Fig. 3.

For asthma, approximately half of the identified genes resided in the MHC region, indicating the crucial role of immune response in asthma pathogenesis. Nicotinic acetylcholine receptors including *CHRNA5*, *CHRNA3,* and *CHRNB4* at 15q25 locus, were found significantly linked to COPD susceptibility, suggesting that genetic variants integrating environmental exposures such as smoking contributed to the development of COPD. Besides, *IREB2* in this locus was identified significantly associated with the susceptibility to both COPD ($P = 2.94 \times 10^{-13}$) and pneumonia ($P = 1.17 \times 10^{-6}$). Since cigarette smoking is one of the main risk factors for respiratory diseases especially COPD, ignoring smoking behavior information in the estimation of the GWAS summary statistics may leads to bias in the identification of susceptibility genes. We thus excluded 10,239 smoking associated variants extracted from the GWAS catalog website and those identified in the UK Biobank database together with SNPs in high LD ($r^2 > 0.6$) relationships with them and re-conducted the gene-based analysis. No obvious changes were observed for the identified susceptibility genes (Supplementary Table 1).

3.2. Functional annotation

We firstly defined a set of CRVs in each identified locus and annotated these variants with publicly available genomic data. Then, we systematically evaluated these CRVs for evidence of enrichment of genomic features such as histone modification marks. Interestingly, we observed significant over-representation of these CRVs in promoter-like (H3K4me3 and H3K9ac) and enhancer-like (H3K4me1 and H3K27ac) histone modification regions, and such enrichment was especially identified in lung (NHLF) or lung cancer (A549) related cell lines (Fig. 4, Supplementary Table 2).

We applied the INQUISIT functional annotation strategy to further detect the regulatory mechanisms underlying identified genes by evaluating the impact of CRVs on coding sequences, proximal promoter and distal enhancer regulations in each gene locus. Among these genes, coding impact evaluation aligned CRVs to two genes, proximal regulatory gene mapping matched CRVs to 80 genes and distal regulatory gene mapping annotated CRVs to 103 genes. This resulted in 110 unique mapped genes, and only *TSPAN8* was identified by all three mapping strategies (Supplementary Table 3).

To unravel the effects of genetic variants on gene regulatory changes,

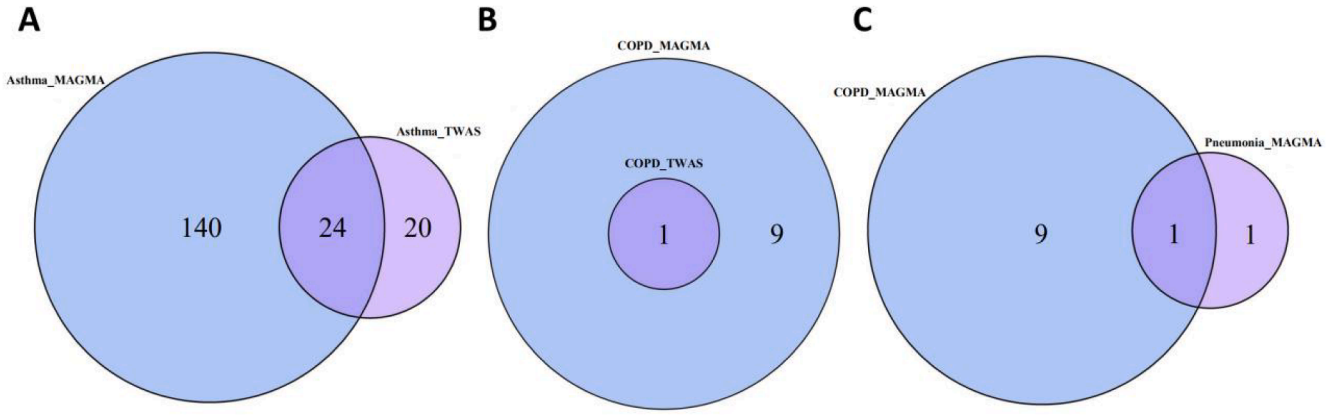

Fig. 2. Venn diagrams showing (A) the number of overlapped susceptibility genes identified related to asthma using FUSION and MAGMA; (B) the number of overlapped susceptibility genes identified related to COPD using FUSION and MAGMA; (C) the number of overlapped susceptibility genes identified related to COPD and pneumonia using MAGMA.

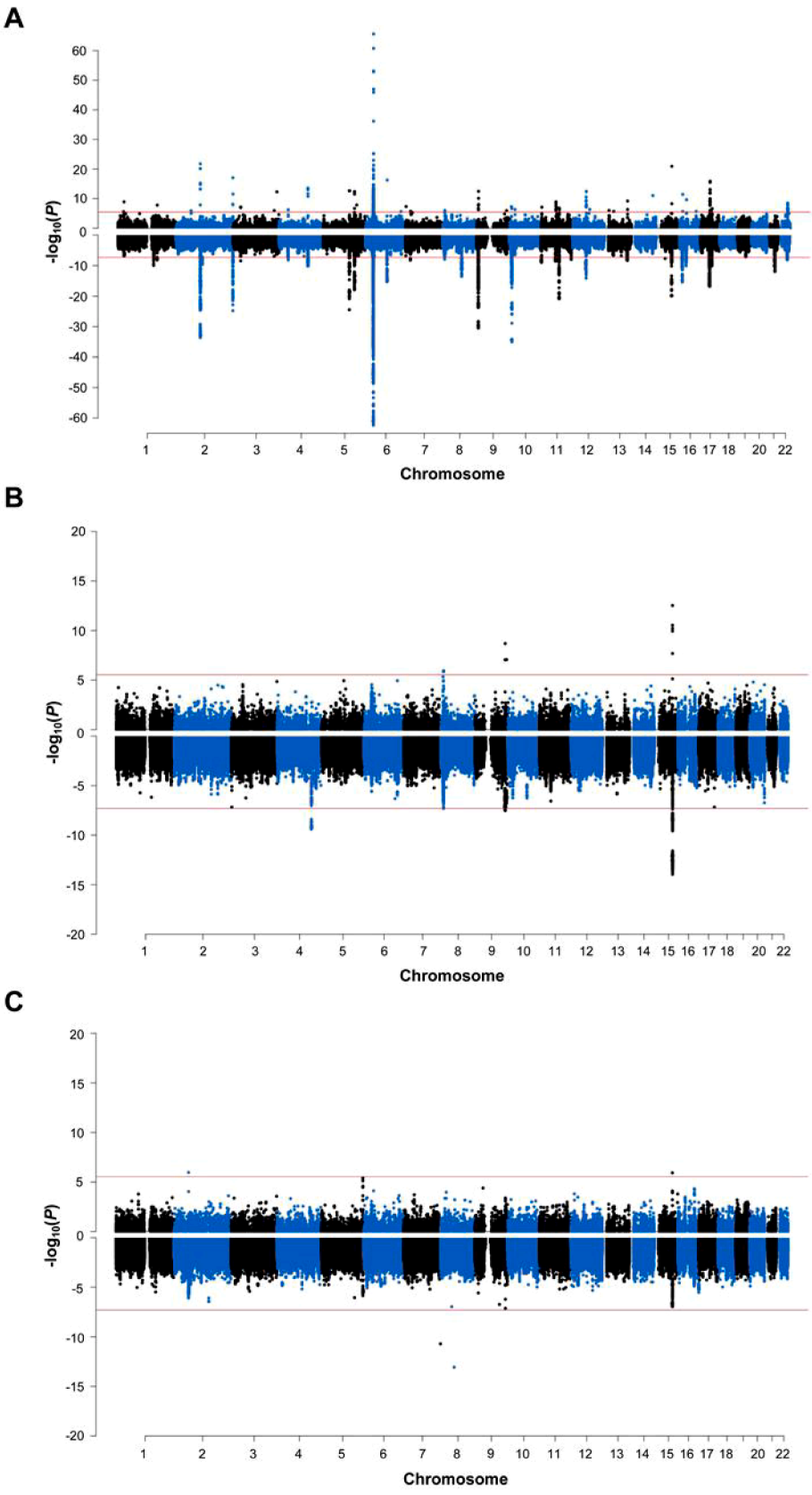

Fig. 3. Manhattan plots for (A) Asthma, (B) COPD and (C) Pneumonia GWAS and gene-based associations. The top figure is Manhattan plot for gene-based associations. Each point corresponds to an association test between gene with asthma/COPD/Pneumonia risk. The red line represents the boundary for significance (2.78×10^{-6}). The bottom figure is the GWAS Manhattan plot where each point is the result of a SNP association test with asthma/COPD/Pneumonia risk. The red line corresponds to the traditional genome-wide significant boundary (5.00×10^{-8}). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

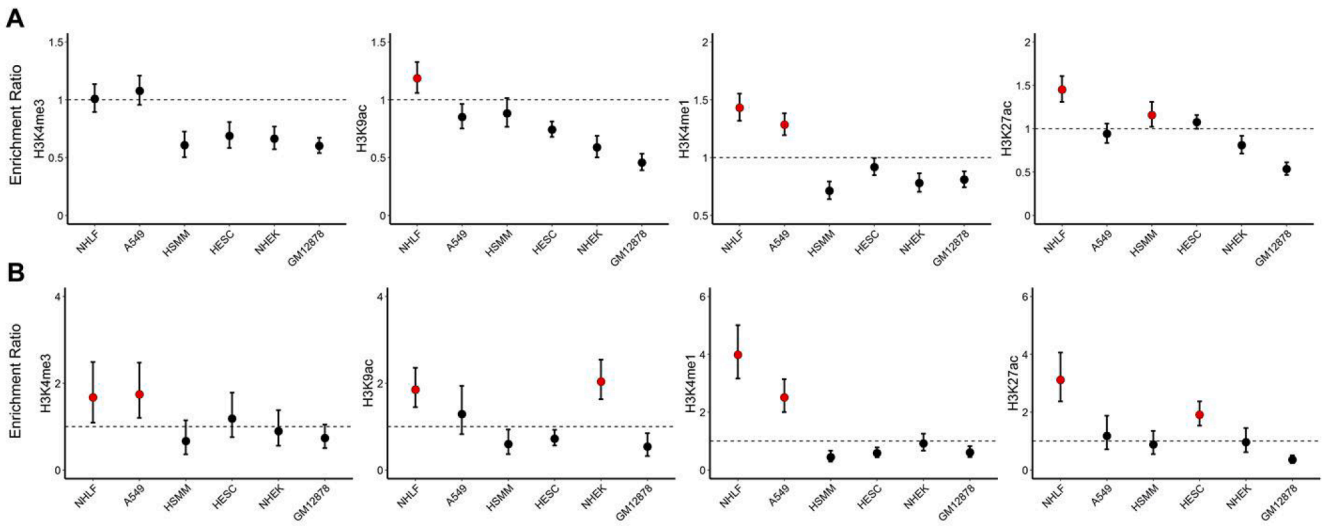

Fig. 4. Functional enrichment of CRVs in histone modification regions for (A) asthma and (B) COPD. 2 cell types in human lung tissue, NHLF and A549, as well as 4 human cell types of other tissues (HSMM, HESC, NHEK, and GM12878) were included to investigate the enrichment in H3K4me3, H3K9ac, H3K4me1 and H3K27ac signals, respectively.

we performed functional annotation of 18 best GWAS SNPs in novel susceptibility loci (Supplementary Table 4). The risk allele of the most significant SNP rs2084200 within *WDPCP* locus was significantly associated with up-regulated expression of *WDPCP* in 383 GTEx lung tissues (Beta = 0.15, $P = 2.20 \times 10^{-5}$, Supplementary Table 5, Supplementary Fig. 1). Additionally, variants in high LD ($r^2 > 0.8$) relationships with rs2084200 resided in the promoter region of *CDH1*, where exhibited strong interaction with *WDPCP* identified by ChIA-PET. The protective

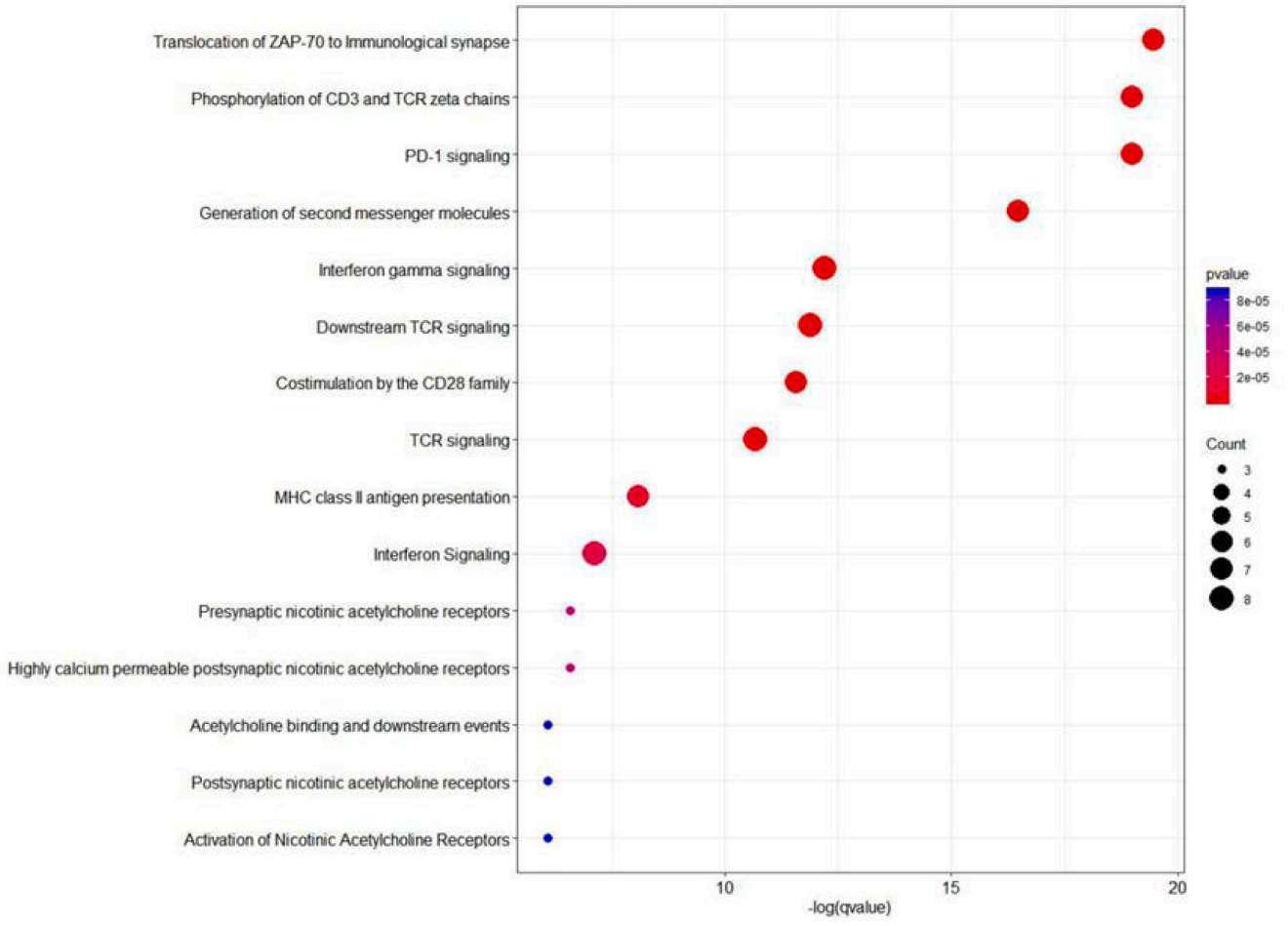

Fig. 5. Pathway enrichment of genes with integrated score ≥1 defined by INQUISIT algorithm. We used the combined results from three traits as input considering the highly genetic correlations between traits. The intensity of color represents the magnitude of *P* value.

allele of the tag SNP rs2416984 at 9q33.3 showed significant association with increased *PBX3* expression (Beta = 0.21, $P = 1.80 \times 10^{-7}$, Supplementary Fig. 2), and its related SNPs ($r^2 > 0.8$) were mainly located in the histone modification marks targeting both promoters and enhancers in NHLF and A549 cell types. The risk allele G of SNP rs17484235 was significantly correlated with the decreased expression of *IREB2* (Beta = -0.051, $P = 3.80 \times 10^{-2}$, Supplementary Fig. 3), along with its related variants presenting strong regulatory signals by targeting the *IREB2* promoter.

3.3. Pathway enrichment analysis

To further explore the biological pathways involved in the process of chronic respiratory diseases, we performed pathway enrichment analysis with 110 functional annotated genes defined by INQUISIT (integrated score ≥ 1). The result revealed the enrichment of 15 pathways ($P_{adj} < 0.05$) that were involved in the immune system and nicotinic acetylcholine receptors signaling, such as *PD-1* signaling ($P_{adj} = 1.74 \times 10^{-8}$), interferon gamma signaling ($P_{adj} = 2.56 \times 10^{-5}$) and presynaptic nicotinic acetylcholine receptors ($P_{adj} = 1.72 \times 10^{-2}$, Fig. 5, Supplementary Table 6).

4. Discussion

UK Biobank is a very large and detailed prospective study with over 500,000 participants aged 39–73 years when recruited in 2006–2010. Utilizing the GWAS-summarized data downloaded from UK Biobank website, we performed gene-based analysis and identified 24 susceptibility genes within novel regions associated with chronic respiratory diseases susceptibility. Functional annotation revealed that abnormal regulation of gene expression plays an important role in the development of respiratory diseases. Pathway enrichment analysis demonstrated that the implicated genes were mainly aggregated in immune-related processes and nicotinic acetylcholine receptors pathways. By utilizing an integrated gene-based and pathway strategy, our study systematically evaluated susceptibility genes and potential biological processes for chronic respiratory diseases, providing important insights for the etiology and clinical treatment of chronic respiratory diseases.

Gene *WDPCP* at 2p15 was newly identified associated with the susceptibility of asthma. The expression of *WDPCP* is significantly regulated by the best GWAS SNP rs2084200 and its related variants within the cis-locus. *WDPCP* plays a critical role in regulating planar cell polarity and ciliogenesis by mediating septin localization (Cui et al., 2013). As the ciliated epithelium that covers the surface of the airways forms an immunologically active natural barrier to invasion and injury, ciliary dysfunction could increase susceptibility to infection and inflammation and has been found as a feature of moderate to severe asthma (Thomas et al., 2010). While there is a lack of detailed information regarding the mechanisms of this gene underlying asthma development and further studies are warranted.

The COPD susceptibility gene, *PBX3*, is one member of TALE class homeodomain family that are implicated in developmental gene expression through their abilities to form hetero-oligomeric DNA-binding complexes and function as transcriptional regulators in numerous cell types (Monica et al., 1991). The expression of *PBX3* was greatly influenced by the lead SNP rs2416984 as well as its related variants at the *PBX3* locus. By developing a *PBX3*-deficient mice model, Rhee JW et al identified that *PBX3* was essential for the proper development of medullary respiratory control mechanisms and mutations of *PBX3* may promote the pathogenesis of central hypoventilation (Rhee et al., 2004). Besides, Heguy A's research observed that *PBX3* was up-regulated in the alveolar macrophages of smokers compared to nonsmokers, indicating the potential association between *PBX3* and cigarette smoking or COPD (Heguy et al., 2006). However, the biological mechanism involved in COPD pathogenesis remains unclear, which needs further investigations.

Consistent with previous study (DeMeo et al., 2009), *IREB2* was identified as a susceptibility gene for COPD in our study. Interestingly, we also observed an association between *IREB2* and pneumonia susceptibility. *IREB2* encodes an RNA binding protein that acts to maintain human cellular iron metabolism by modulating the expression of those proteins relevant to iron uptake, export, and sequestration (Rouault, 2006). *IREB2* variants are in tight LD with SNPs associated with nicotine addiction. As we all know, conventional cigarette smoking is an important factor for developing both COPD and pneumonia, and prior studies have demonstrated that smoking was associated with lung iron imbalance in pulmonary inflammation (O'Brien-Ladner et al., 2000; Ghio et al., 2008), supporting a role for *IREB2* in the pathogenesis of COPD and pneumonia.

Previous studies have identified that a subset of current and former smokers developed an asthma-COPD overlap condition that was associated with gene expression markers of Th2 inflammation in the airway (Barnes, 2008; Christenson et al., 2015). Torres A *et al.* found that adults with chronic conditions and other risk factors such as COPD, asthma, and smoking were at increased risk of pneumonia (Torres et al., 2015), indicating shared pathogenic mechanisms among chronic respiratory diseases. Consistent with previous findings, our study revealed an aggregation of susceptibility genes in immune-related processes and nicotinic acetylcholine receptors pathways. Furthermore, we conducted functional annotation to explore potential regulatory mechanisms and found that the credible variants were primarily mapped to non-coding regions, and showed a strong over-representation in eQTLs as well as functional regions such as histone modification marks. These results suggest that the variants may contribute to the development of chronic respiratory diseases through regulating the expression of target genes. More importantly, we proved that such genetic mechanisms were specific to lung tissues.

5. Conclusions

In this study, we applied an integrated gene-based and pathway enrichment strategy to explore the potential susceptibility genes and biological processes involved in the development of chronic respiratory diseases. Our study identified 24 genes within novel susceptibility regions and 15 pathways that are statistically significantly associated with respiratory diseases, providing important insights into the genetic causes of diseases and giving suggestions to future clinical treatment. However, other respiratory traits with limited cases such as interstitial lung disease were not included, for the sample size may have an influence on the reliability of the results. Another limitation was that we only focused on protein-coding genes in the identification of susceptibility genes, which could lead to miss relevant ones such as lncRNA genes. Therefore, additional studies on the exploration of noncoding susceptibility genes for respiratory diseases are needed. Also, further functional experiments are warranted to unravel the biological mechanisms behind the associations.

6. Availability of data and materials

All summary results from the analyses performed are available at the GeneATLAS website, http://geneatlas.roslin.ed.ac.uk/.

CRediT authorship contribution statement

Lijuan Wang: Conceptualization, Formal analysis, Data curation, Writing - original draft. **Rui Fang:** Conceptualization, Formal analysis, Data curation. **Meng Zhu:** Conceptualization, Methodology. **Na Qin:** Methodology, Investigation. **Yuzhuo Wang:** Methodology, Investigation. **Jingyi Fan:** Investigation. **Qi Sun:** Investigation. **Mengmeng Ji:** Investigation. **Xikang Fan:** Investigation. **Junxing Xie:** Investigation. **Hongxia Ma:** Supervision, Writing - review & editing. **Juncheng Dai:** Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the study participants and research staffs for their contributions and commitment to this study.

Funding

This work was funded by the National Key Research and Development Program of China (2017YFC0907900, 2017YFC0907905, 2019YFC1315700 and 2019YFC1315704), National Natural Science of China (81820108028, 81521004, 81803306 and 81941020), Science Foundation for Distinguished Young Scholars of Jiangsu (BK20160046), Natural Science Foundation of Jiangsu Province (BK20180675), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (18KJB330002 and 19KJA360002), the Priority Academic Program for the Development of Jiangsu Higher Education Institutions [Public Health and Preventive Medicine] and Top-notch Academic Programs Project of Jiangsu Higher Education Institutions (PPZY2015A067).

Ethics approval and consent to participate

All participants gave informed consent for data provision and linkage. The UK Biobank project was approved by the National Research Ethics Service Committee North West-Haydock (REC reference: 11/NW/0382).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2020.145287.

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7 (4), 248–249.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., Bagger, F.O., Jørgensen, M., Andersen, P.R., Bertin, N., Rackham, O., Burroughs, A.M., Baillie, J. K., Ishizu, Y., Shimizu, Y., Furuhata, E., Maeda, S., Negishi, Y., Mungall, C.J., Meehan, T.F., Lassmann, T., Itoh, M., Kawaji, H., Kondo, N., Kawai, J., Lennartsson, A., Daub, C.O., Heutink, P., Hume, D.A., Jensen, T.H., Suzuki, H., Hayashizaki, Y., Müller, F., Forrest, A.R.R., Carninci, P., Rehli, M., Sandelin, A., 2014. An atlas of active enhancers across human cell types and tissues. Nature 507 (7493), 455–461.

Barnes, P.J., 2008. Immunology of asthma and chronic obstructive pulmonary disease. Nat. Rev. Immunol. 8 (3), 183–192.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J., 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562 (7726), 203–209.

Canela-Xandri, O., Rawlik, K., Tenesa, A., 2018. An atlas of genetic associations in UK Biobank. Nat. Genet. 50 (11), 1593–1599.

Christenson, S.A., Steiling, K., van den Berge, M., Hijazi, K., Hiemstra, P.S., Postma, D.S., Lenburg, M.E., Spira, A., Woodruff, P.G., 2015. Asthma–COPD overlap. Clinical relevance of genomic signatures of type 2 inflammation in chronic obstructive pulmonary disease. Am. J. Respir. Crit. Care Med. 191 (7), 758–766.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal{middle dot}lari, R., Lupien, M., Markowitz, S., Scacheri, P.C., 2014. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. 24 (1), 1–13.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., Stein, L., 2011. Reactome: a database of reactions, pathways and biological processes. Nucl. Acids Res. 39, D691–D697.

Cui, C., Chatterjee, B., Lozito, T.P., Zhang, Z., Francis, R.J., Yagi, H., Swanhart, L.M., Sanker, S., Francis, D., Yu, Q., San Agustin, J.T., Puligilla, C., Chatterjee, T., Tansey, T., Liu, X., Kelley, M.W., Spiliotis, E.T., Kwiatkowski, A.V., Tuan, R., Pazour, G.J., Hukriede, N.A. and Lo, C.W., 2013. Wdpcp, a PCP protein required for ciliogenesis, regulates directional cell migration and cell polarity by direct modulation of the actin cytoskeleton. PLoS Biol 11, e1001720.

de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D., 2015. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 11, e1004219.

DeMeo, D.L., Mariani, T., Bhattacharya, S., Srisuma, S., Lange, C., Litonjua, A., Bueno, R., Pillai, S.G., Lomas, D.A., Sparrow, D., Shapiro, S.D., Criner, G.J., Kim, H.P., Chen, Z., Choi, A.M.K., Reilly, J., Silverman, E.K., 2009. Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. Am. J. Human Genetics 85 (4), 493–502.

Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L., Mathews, J.D., 1990. Genetics of asthma and hay fever in Australian twins. Am. Rev. Respir. Dis. 142, 1351–1358.

GBD 2015 Chronic Respiratory Disease Collaborators, 2017. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir. Med. 5, 691-706.

GBD 2017 Causes of Death Collaborators, 2018. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392, 1736-1788.

Ghio, A.J., Hilborn, E.D., Stonehuerner, J.G., Dailey, L.A., Carter, J.D., Richards, J.H., Crissman, K.M., Foronjy, R.F., Uyeminami, D.L., Pinkerton, K.E., 2008. Particulate matter in cigarette smoke alters iron homeostasis to produce a biological effect. Am. J. Respir. Crit. Care Med. 178 (11), 1130–1138.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., Sullivan, P.F., Nikkola, E., Alvarez, M., Civelek, M., Lusis, A.J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Raitakari, O.T., Kuusisto, J., Laakso, M., Price, A.L., Pajukanta, P., Pasaniuc, B., 2016. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. 48 (3), 245–252.

Heguy, A., O'Connor, T.P., Luettich, K., Worgall, S., Cieciuch, A., Harvey, B.-G., Hackett, N.R., Crystal, R.G., 2006. Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. J. Mol. Med. 84 (4), 318–328.

Kellum, J.A., Kong, L., Fink, M.P., Weissfeld, L.A., Yealy, D.M., Pinsky, M.R., Fine, J., Krichevsky, A., Delude, R.L., Angus, D.C., Gen, I.M.S.I., 2007. Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the Genetic and Inflammatory Markers of Sepsis (GenIMS) Study. Arch. Intern. Med. 167, 1655–1663.

Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. and Bergmann, S., 2016. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. PLoS Comput Biol 12, e1004714.

Lesnick, T.G., Papapetropoulos, S., Mash, D.C., Ffrench-Mullen, J., Shehadeh, L., de Andrade, M., Henley, J.R., Rocca, W.A., Ahlskog, J.E. and Maraganore, D.M., 2007. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. PLoS Genet 3, e98.

Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., Hemminki, K., 2000. Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. N. Engl. J. Med. 343 (2), 78–85.

Liu, J.Z., Mcrae, A.F., Nyholt, D.R., Medland, S.E., Wray, N.R., Brown, K.M., Hayward, N. K., Montgomery, G.W., Visscher, P.M., Martin, N.G., Macgregor, S., 2010. A versatile gene-based test for genome-wide association studies. Am. J. Human Genetics 87 (1), 139–145.

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M.K., Wang, Q., Tyrer, J., Dicks, E.d., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., French, J.D., Qing Chen, X., Fachal, L., McCue, K., McCart Reed, A.E., Ghoussaini, M., Carroll, J.S., Jiang, X., Finucane, H., Adams, M., Adank, M.A., Ahsan, H., Aittomäki, K., Anton-Culver, H., Antonenkova, N.N., Arndt, V., Aronson, K.J., Arun, B., Auer, P.L., Bacot, F., Barrdahl, M., Baynes, C., Beckmann, M.W., Behrens, S., Benitez, J., Bermisheva, M., Bernstein, L., Blomqvist, C., Bogdanova, N.V., Bojesen, S.E., Bonanni, B., Børresen-Dale, A.-L., Brand, J.S., Brauch, H., Brennan, P., Brenner, H., Brinton, L., Broberg, P., Brock, I.W., Broeks, A., Brooks-Wilson, A., Brucker, S.Y., Brüning, T., Burwinkel, B., Butterbach, K., Cai, Q., Cai, H., Caldés, T., Canzian, F., Carracedo, A., Carter, B.D., Castelao, J.E., Chan, T.L., David Cheng, T.-Y., Seng Chia, K., Choi, J.-Y., Christiansen, H., Clarke, C.L., Collée, M., Conroy, D.M., Cordina-Duverger, E., Cornelissen, S., Cox, D.G., Cox, A., Cross, S.S., Cunningham, J.M., Czene, K., Daly, M. B., Devilee, P., Doheny, K.F., Dörk, T., dos-Santos-Silva, I., Dumont, M., Durcan, L., Dwek, M., Eccles, D.M., Ekici, A.B., Eliassen, A.H., Ellberg, C., Elvira, M., Engel, C., Eriksson, M., Fasching, P.A., Figueroa, J., Flesch-Janys, D., Fletcher, O., Flyger, H., Fritschi, L., Gaborieau, V., Gabrielson, M., Gago-Dominguez, M., Gao, Y.-T., Gapstur, S.M., García-Sáenz, J.A., Gaudet, M.M., Georgoulias, V., Giles, G.G., Glendon, G., Goldberg, M.S., Goldgar, D.E., González-Neira, A., Grenaker Alnæs, G. I., Grip, M., Gronwald, J., Grundy, A., Guénel, P., Haeberle, L., Hahnen, E., Haiman, C.A., Håkansson, N., Hamann, U., Hamel, N., Hankinson, S., Harrington, P., Hart, S.N., Hartikainen, J.M., Hartman, M., Hein, A., Heyworth, J., Hicks, B., Hillemanns, P., Ho, D.N., Hollestelle, A., Hooning, M.J., Hoover, R.N., Hopper, J.L., Hou, M.-F., Hsiung, C.-N., Huang, G., Humphreys, K., Ishiguro, J., Ito, H.,

Iwasaki, M., Iwata, H., Jakubowska, A., Janni, W., John, E.M., Johnson, N., Jones, K., Jones, M., Jukkola-Vuorinen, A., Kaaks, R., Kabisch, M., Kaczmarek, K., Kang, D., Kasuga, Y., Kerin, M.J., Khan, S., Khusnutdinova, E., Kiiski, J.I., Kim, S.-W., Knight, J.A., Kosma, V.-M., Kristensen, V.N., Krüger, U., Kwong, A., Lambrechts, D., Le Marchand, L., Lee, E., Lee, M.H., Lee, J.W., Neng Lee, C., Lejbkowicz, F., Li, J., Lilyquist, J., Lindblom, A., Lissowska, J., Lo, W.-Y., Loibl, S., Long, J., Lophatananon, A., Lubinski, J., Luccarini, C., Lux, M.P., Ma, E.S.K., MacInnis, R.J., Maishman, T., Makalic, E., Malone, K.E., Kostovska, I.M., Mannermaa, A., Manoukian, S., Manson, J.E., Margolin, S., Mariapun, S., Martinez, M.E., Matsuo, K., Mavroudis, D., McKay, J., McLean, C., Meijers-Heijboer, H., Meindl, A., Menéndez, P., Menon, U., Meyer, J., Miao, H., Miller, N., Taib, N.A.M., Muir, K., Mulligan, A.M., Mulot, C., Neuhausen, S.L., Nevanlinna, H., Neven, P., Nielsen, S.F., Noh, D.-Y., Nordestgaard, B.G., Norman, A., Olopade, O.I., Olson, J.E., Olsson, H., Olswold, C., Orr, N., Pankratz, V.S., Park, S.K., Park-Simon, T.-W., Lloyd, R., Perez, J. I.A., Peterlongo, P., Peto, J., Phillips, K.-A., Pinchev, M., Plaseska-Karanfilska, D., Prentice, R., Presneau, N., Prokofyeva, D., Pugh, E., Pylkäs, K., Rack, B., Radice, P., Rahman, N., Rennert, G., Rennert, H.S., Rhenius, V., Romero, A., Romm, J., Ruddy, K.J., Rüdiger, T., Rudolph, A., Ruebner, M., Rutgers, E.J.T., Saloustros, E., Sandler, D.P., Sangrajrang, S., Sawyer, E.J., Schmidt, D.F., Schmutzler, R.K., Schneeweiss, A., Schoemaker, M.J., Schumacher, F., Schürmann, P., Scott, R.J., Scott, C., Seal, S., Seynaeve, C., Shah, M., Sharma, P., Shen, C.-Y., Sheng, G., Sherman, M.E., Shrubsole, M.J., Shu, X.-O., Smeets, A., Sohn, C., Southey, M.C., Spinelli, J.J., Stegmaier, C., Stewart-Brown, S., Stone, J., Stram, D.O., Surowy, H., Swerdlow, A., Tamimi, R., Taylor, J.A., Tengström, M., Teo, S.H., Beth Terry, M., Tessier, D.C., Thanasitthichai, S., Thöne, K., Tollenaar, R.A.E.M., Tomlinson, I., Tong, L., Torres, D., Truong, T., Tseng, C.-C., Tsugane, S., Ulmer, H.-U., Ursin, G., Untch, M., Vachon, C., van Asperen, C.J., Van Den Berg, D., van den Ouweland, A.M. W., van der Kolk, L., van der Luijt, R.B., Vincent, D., Vollenweider, J., Waisfisz, Q., Wang-Gohrke, S., Weinberg, C.R., Wendt, C., Whittemore, A.S., Wildiers, H., Willett, W., Winqvist, R., Wolk, A., Wu, A.H., Xia, L., Yamaji, T., Yang, X.R., Har Yip, C., Yoo, K.-Y., Yu, J.-C., Zheng, W., Zheng, Y., Zhu, B., Ziogas, A., Ziv, E., Lakhani, S.R., Antoniou, A.C., Droit, A., Andrulis, I.L., Amos, C.I., Couch, F.J., Pharoah, P.D.P., Chang-Claude, J., Hall, P., Hunter, D.J., Milne, R.L., García-Closas, M., Schmidt, M.K., Chanock, S.J., Dunning, A.M., Edwards, S.L., Bader, G.D., Chenevix-Trench, G., Simard, J., Kraft, P., Easton, D.F., 2017. Association analysis identifies 65 new breast cancer risk loci. Nature 551 (7678), 92–94.

Mishra, A., Macgregor, S., 2015. VEGAS2: software for more flexible gene-based testing. Twin Res. Hum. Genet. 18 (1), 86–91.

Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M., Cookson, W.O.C.M., 2010. A large-scale, consortium-based genomewide association study of asthma. N. Engl. J. Med. 363 (13), 1211–1221.

Monica, K., Galili, N., Nourse, J., Saltman, D., Cleary, M.L., 1991. PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. Mol. Cell. Biol. 11 (12), 6149–6157.

Ng, P.C., 2003. SIFT: predicting amino acid changes that affect protein function. Nucl. Acids Res. 31 (13), 3812–3814.

O'Brien-Ladner, A.R., Nelson, S.R., Murphy, W.J., Blumer, B.M., Wesselius, L.J., 2000. Iron is a regulatory component of human IL-1beta production. Support for regional variability in the lung. Am. J. Respir. Cell Mol. Biol. 23, 112–119.

Drazen, J.M., Postma, D.S., Rabe, K.F., 2015. The asthma–COPD overlap syndrome. N. Engl. J. Med. 373 (13), 1241–1249.

Rautanen, A., Mills, T.C., Gordon, A.C., Hutton, P., Steffens, M., Nuamah, R., Chiche, J.-D., Parks, T., Chapman, S.J., Davenport, E.E., Elliott, K.S., Bion, J., Lichtner, P., Meitinger, T., Wienker, T.F., Caulfield, M.J., Mein, C., Bloos, F., Bobek, I., Cotogni, P., Sramek, V., Sarapuu, S., Kobilay, M., Ranieri, V.M., Rello, J., Sirgo, G., Weiss, Y.G., Russwurm, S., Schneider, E.M., Reinhart, K., Holloway, P.A.H., Knight, J.C., Garrard, C.S., Russell, J.A., Walley, K.R., Stüber, F., Hill, A.V.S., Hinds, C.J., 2015. Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study. Lancet Respiratory Med. 3 (1), 53–60.

Rhee, J.W., Arata, A., Selleri, L., Jacobs, Y., Arata, S., Onimaru, H., Cleary, M.L., 2004. Pbx3 deficiency results in central hypoventilation. Am. J. Pathol. 165 (4), 1343–1350.

Rouault, T.A., 2006. The role of iron regulatory proteins in mammalian iron homeostasis and disease. Nat. Chem. Biol. 2 (8), 406–414.

Salvi, S.S., Barnes, P.J., 2009. Chronic obstructive pulmonary disease in non-smokers. The Lancet 374 (9691), 733–743.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T. and Collins, R., 2015. 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.

Thomas, B., Rutman, A., Hirst, R.A., Haldar, P., Wardlaw, A.J., Bankart, J., Brightling, C. E., O'Callaghan, C., 2010. Ciliary dysfunction and ultrastructural abnormalities are features of severe asthma. J. Allergy Clin. Immunol. 126 (4), 722-729.e2.

Torres, A., Blasi, F., Dartois, N., Akova, M., 2015. Which individuals are at increased risk of pneumococcal disease and why? Impact of COPD, asthma, smoking, diabetes, and/or chronic heart disease on community-acquired pneumonia and invasive pneumococcal disease. Thorax 70, 984–989.

Wain, L.V., Shrine, N., Artigas, M.S., Erzurumluoglu, A.M., Noyvert, B., Bossini-Castillo, L., Obeidat, Ma'en, Henry, A.P., Portelli, M.A., Hall, R.J., Billington, C.K., Rimington, T.L., Fenech, A.G., John, C., Blake, T., Jackson, V.E., Allen, R.J., Prins, B. P., Campbell, A., Porteous, D.J., Jarvelin, M.-R., Wielscher, M., James, A.L., Hui, J., Wareham, N.J., Zhao, J.H., Wilson, J.F., Joshi, P.K., Stubbe, B., Rawal, R., Schulz, H., Imboden, M., Probst-Hensch, N.M., Karrasch, S., Gieger, C., Deary, I.J., Harris, S.E., Marten, J., Rudan, I., Enroth, S., Gyllensten, U., Kerr, S.M., Polasek, O., Kähönen, M., Surakka, I., Vitart, V., Hayward, C., Lehtimäki, T., Raitakari, O.T., Evans, D.M., Henderson, A.J., Pennell, C.E., Wang, C.A., Sly, P.D., Wan, E.S., Busch, R., Hobbs, B. D., Litonjua, A.A., Sparrow, D.W., Gulsvik, A., Bakke, P.S., Crapo, J.D., Beaty, T.H., Hansel, N.N., Mathias, R.A., Ruczinski, I., Barnes, K.C., Bossé, Y., Joubert, P., van den Berge, M., Brandsma, C.-A., Paré, P.D., Sin, D.D., Nickle, D.C., Hao, K.e., Gottesman, O., Dewey, F.E., Bruse, S.E., Carey, D.J., Kirchner, H.L., Jonsson, S., Thorleifsson, G., Jonsdottir, I., Gislason, T., Stefansson, K., Schurmann, C., Nadkarni, G., Bottinger, E.P., Loos, R.J.F., Walters, R.G., Chen, Z., Millwood, I.Y., Vaucher, J., Kurmi, O.P., Li, L., Hansell, A.L., Brightling, C., Zeggini, E., Cho, M.H., Silverman, E.K., Sayers, I., Trynka, G., Morris, A.P., Strachan, D.P., Hall, I.P., Tobin, M.D., 2017. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat. Genet. 49 (3), 416–425.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucl. Acids Res 38, e164.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16 (5), 284–287.

Zhong, H., Yang, X., Kaplan, L.M., Molony, C., Schadt, E.E., 2010. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. Am. J. Human Genetics 86 (4), 581–591.