

**Combination of single cell sequencing data and GWAS summary statistics reveals  
genetically-influenced liver cell types for primary biliary cholangitis**

Bingyu Xiang<sup>1,#</sup>, Chunyu Deng<sup>2,#</sup>, Jingjing Li<sup>3</sup>, Shanshan Li<sup>1</sup>, Huifang Zhang<sup>1</sup>, Xiuli Lin<sup>1</sup>, Mingqin  
Lu<sup>1,\*</sup>, and Yunlong Ma<sup>2,\*</sup>

1. Department of Infectious Diseases, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang, China.
2. Institute of Biomedical Big Data, School of Ophthalmology & Optometry and Eye Hospital, School of Biomedical Engineering, Wenzhou Medical University, Wenzhou 325027, Zhejiang, China.
3. State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University School of Medicine, Hangzhou 310003, Zhejiang, China.

<sup>#</sup> These authors contributed equally to this work.

**\* Correspondence should be addressed to:**

1. Yunlong Ma, Ph.D. [glb-biotech@zju.edu.cn](mailto:glb-biotech@zju.edu.cn);

Institute of Biomedical Big Data, School of Ophthalmology & Optometry and Eye Hospital,  
Wenzhou Medical University, Wenzhou, 325027, China.

2. Mingqin Lu, Professor. [lmq0906@163.com](mailto:lmq0906@163.com);

Department of Infectious Diseases, the First Affiliated Hospital of Wenzhou Medical University,  
Wenzhou 325000, Zhejiang, China.

## Abstract

**Importance:** Primary biliary cholangitis (PBC) is a classical autoimmune disease, which is highly influenced by genetic determinants. Many genome-wide association studies (GWAS) have reported that numerous genetic loci were significantly associated with PBC susceptibility. However, the effects of genetic determinants on liver cells and its immune microenvironment for PBC remain unclear.

**Objective:** To identify genetics-modulated functional liver cell subsets involved in the pathogenesis of PBC.

**Design, Setting, and Participants:** In this present study, 13,239 European participants were collected from IEU open GWAS project on PBC. There were 1,124,241 qualified SNPs used for GWAS analysis. Expression quantitative trait loci (eQTL) data across 49 tissues were downloaded from the GTEx database. Two single cell RNA sequencing (scRNA-seq) profiles and two bulk-based RNA transcriptomes were downloaded from the GEO database. Data collection and analyses were performed from August 2020 to June 2021.

**Main outcomes and measures:** We constructed a powerful computational framework to integrate GWAS summary statistics with scRNA-seq data to uncover genetics-modulated liver cell subpopulations.

**Results:** Based on our multi-omics integrative analysis, we found that 29 risk genes including *ORMDL3*, *GSNK2B*, and *DDAH2* were significantly associated with PBC susceptibility.

Gene-property analysis revealed that four immune cell types, including Cst3<sup>+</sup> dendritic cell, Chil3<sup>+</sup> macrophage, Trbc2<sup>+</sup> T cell, and Gzma<sup>+</sup> T cell, were significantly enriched by PBC- risk genes. By combining GWAS summary statistics with scRNA-seq data, we found that cholangiocytes exhibited a notable enrichment by PBC-related genetic association signals (Permuted P < 0.05). The risk gene of *ORMDL3* showed the highest expression proportion in cholangiocytes than other liver cells (22.38%). Compared with *ORMDL3*<sup>+</sup> cholangiocytes, there were 71 significantly highly-expressed genes among *ORMDL3*<sup>-</sup> cholangiocytes (FDR < 0.05), such as inflammatory cytokine genes *CXCL8*, *CCL3*, *IFI16*, and *IRF1*. These highly-expressed genes were significantly enriched in numerous biological pathways and functional terms associated with autoimmune diseases (FDR < 0.05).

**Conclusions and relevance:** To the best of our knowledge, this is the first study to integrate genetic information with single cell sequencing data for parsing genetics-influenced liver cells for PBC risk. We identified that *ORMDL3*<sup>-</sup> cholangiocytes play important immune-modulatory roles in the etiology of PBC.

**Key points:**

**Question:** Are genetics factors influenced liver cell subpopulations and its immune microenvironment for PBC?

**Findings:** In this comprehensive genomics study based on multi-omics data, genetic determinants were significantly enriched in cholangiocytes and immune cells including subsets of macrophage, dendritic cells, and T cells. *ORMDL3*<sup>-</sup> cholangiocytes have crucial immune-modulatory roles in developing PBC.

**Meaning:** Findings suggest that integration of single cell sequencing data with GWAS summary

statistics contribute to pinpoint PBC-relevant cell types and risk genes.

## Introduction

Primary biliary cholangitis (PBC), which is formally known as primary biliary cirrhosis until 2016<sup>1</sup>, is a rare chronic cholestatic liver disease characterized by progressive autoimmune-mediated destruction of the small intrahepatic biliary epithelial cells<sup>2,3</sup>. PBC patients suffering from chronic cholestasis can eventually lead to cirrhosis and hepatic failure without effective treatments<sup>2</sup>. Although ursodeoxycholic acid has been used as the first-line therapeutic agent for PBC, there exist 10% to 20% of PBC patients resistant to ursodeoxycholic acid and developing to advanced-stage liver disease<sup>2</sup>. Previous studies<sup>4</sup> have reported that a combination of genetic and environmental risk factors have an important influence on the aetiology of PBC. Hence, understanding the genetic mechanisms of PBC is becoming a great interest, which may promote the development of individualized therapeutic strategy for PBC.

Over the past decade, a growing number of genome-wide association studies (GWAS) and ImmunoChip studies based on East Asian and European populations have been performed to uncover the genetic susceptibility loci associated with PBC<sup>4</sup>. To date, more than 40 genetic loci with numerous risk genes have been reported<sup>5-10</sup>, such as *SLC19A3/CCL20*, *IRF8/FOXF1*, *NFKB1/MANBA*, and *PDGFB/RPL3*. Nevertheless, the GWAS approach has generally focused on examining the genetic associations of millions of single nucleotide polymorphisms (SNPs) and only a handful of SNPs with a genome-wide significance ( $P \leq 5 \times 10^{-8}$ ) are reported. There exist many common SNPs with small marginal effects were neglected<sup>11,12</sup>. Moreover, the vast majority of reported SNPs were mapped within non-coding genomic regions<sup>12</sup>. It is plausible to infer that these

non-coding SNPs may modulate the expression levels of corresponding risk genes rather than change the functions of their proteins. Thus, combination of GWAS summary statistics and other different types of data that characterize tissue- and cell-type-specific activity, including expression quantitative trait loci <sup>11</sup>, DNase I-hypersensitive sites (DHS) <sup>13</sup>, and histone marks <sup>14</sup>, contributes to highlight disease-related risk genes and cell types.

With the advance of single cell sequencing techniques, researchers have an effective avenue to discover more refined and novel cell populations for complex diseases <sup>15</sup>. An accruing and large number of single cell RNA sequencing (scRNA-seq) studies on autoimmune diseases, including rheumatoid arthritis <sup>16</sup>, inflammatory bowel disease <sup>17</sup>, and systemic lupus erythematosus <sup>18</sup>, have been reported to parse the heterogeneity of cellular subpopulations at unprecedented resolution. In view of no scRNA-seq study was conducted for uncovering human liver cell types implicated in PBC, we constructed a computational framework to identify risk genes whose genetically expressions associated with PBC and pinpoint cell subpopulations implicated in the etiology of PBC.

## **Methods**

### ***Datasets***

*1. Single-cell transcriptomes of PBC:* We downloaded two independent single cell RNA sequencing profiles (Accession number: GSE93170 and GSE115469) from the GEO database. With regard to the dataset of GSE93170, there were clinically and pathologically diagnosed six healthy controls and six PBC patients enrolled with written informed consent. Peripheral CD4+T cells were used to extract total RNA. The Agilent microarray of SurePrint G3 human GE 8×60K microarray kit was leveraged to produce gene expression profiles according to manufacturer's protocols. The GSE115469 dataset

contained five samples from primary liver patients, which were used for scRNA sequencing based on the 10× Genomics Chromium Single Cell Kits. A total of 8,444 parenchymal and non-parenchymal cells have obtained the transcriptional profiles based on the CellRanger analysis pipeline. The raw digital matrix of gene expression (namely UMI counts per gene per cell) was filtered, normalized and clustered. Cell was omitted if it has a very high (>0.5) mitochondrial genome transcript ratio or a very small library size (<1500). Using the standard Seurat package <sup>19</sup>, there were 20 discrete clusters for the scRNA-seq dataset. Using well-known marker genes, these clusters were assigned into 13 distinct cell subpopulations, including portal endothelial cells, cholangiocytes, non-inflammatory macrophages, T cells,  $\gamma\delta$ T cells, inflammatory monocytes/macrophages, natural killer (NK)-like cells, red blood cells (RBCs), sinusoidal endothelial cells, mature B cells, stellate cells, plasma cells, and hepatocytes.

*2. Bulk-based expression profiles of PBC:* Furthermore, we also downloaded two independent bulk-based expression datasets based on liver tissue (Accession number: GSE159676) and blood (Accession number: GSE119600) from the Gene Expression Omnibus (GEO) database. The dataset of GSE159676 contained six healthy controls and three PBC cases based on fresh frozen liver tissue, which were obtained from explanted livers or diagnostic liver biopsies. The Affymetrix Human Gene 1.0-st array was leveraged to produce bulk-based liver expression profiles with 17,046 probes. With respect to the dataset of GSE119600, there were 47 healthy controls and 90 PBC patients with whole blood samples. The Illumina HumanHT-12 V4.0 expression beadchip was leveraged to produce bulk-based blood transcriptomes with 47,230 probes.

3. *GWAS summary statistics on PBC*: We downloaded a GWAS summary dataset on PBC from the IEU open GWAS project (<https://gwas.mrcieu.ac.uk/>)<sup>5</sup>. There were 2,764 PBC patients and 10,475 healthy controls based on European ancestry in this dataset used for performing a meta-analysis of genome-wide association signals. This GWAS dataset was approved by the University Health Network Research Ethics Board, the Mayo Clinic Institutional Review Board, Etico Indipendente IRCCS Istituto Clinico Humanitas, UC Davis Institutional Review Board and the Oxford Research Ethics Committee<sup>5</sup>. A standard quality control (QC) pipeline was applied to remove low-quality SNPs. The software package of MaCH<sup>20</sup> with the reference of HapMap3 CEU+TSI samples was implemented to perform a genome-wide imputation analysis. There were 1,124,241 SNPs with minor allele frequency > 0.005 and imputation quality score  $R^2 > 0.5$  included in follow-up analyses.

### ***Data Processing and Analysis***

1. *Combination of GWAS summary statistics with scRNA-seq data for PBC*: We leveraged a widely-used method of RolyRoly<sup>21</sup>, which was designed to gain the effects of SNPs near protein-coding genes on cell types contributing to complex traits, to explore genetics-influenced liver cell types for PBC. The regression-based polygenic model was used in the RolyRoly to incorporate GWAS summary data with scRNA-seq data (i.e., GSE115469) for identifying PBC-associated liver cell subpopulations. Let  $g(i)$  represents a given gene relevant to SNP  $i$ ,  $S_j = \{i : g(i) = j\}$  represents a given set with multiple SNPs relevant to the gene  $j$ , and  $\beta_{S_j}$  represents a PBC-GWAS-derived effect-size vector of  $S_j$  with a *priori* assumption that  $\beta_{S_j} \sim \text{MVN}(\mathbf{0}, \sigma_j^2 \mathbf{I}_{|S_j|})$ . Under the assumption, RolyPoly offers a polygenic linear model for  $\beta_{S_j}$ :

$$\sigma_j^2 = \gamma_0 + \sum_{i=1}^N \gamma_i \alpha_{ji}$$

where  $\gamma_0$  represents an intercept term,  $\alpha_{j_i} (i=1,2,\dots,N)$  represents a group of annotations, for example, cell-type-specific gene expressions, and  $\gamma_i$  is the annotation's coefficient for  $\alpha_{j_i}$ . To fit the observed and expected sum squared SNP effect sizes relevant to each gene, RolyPoly applies the method-of-moments estimators to estimate  $\gamma_i$  by the following formula:

$$E\left(\sum_{i \in S_j} \hat{\beta}_i^2\right) = \sigma_j^2 \text{Tr}(\mathbf{R}_{S_j}^2) + |S_j| \sigma_e^2 n^{-1}$$

where  $\mathbf{R}_{S_j}$  represents the LD matrix of  $S_j$ . The 1,000 Genome Project European Phase 3 panel<sup>22</sup> was used to calculate the SNP-SNP LD information. The major histocompatibility complex (MHC) region was removed due to the highly extensive LD in this region. RolyPoly utilized the 1,000 iterations of block bootstrap to assess standard errors for calculating t-statistic and corresponding P value. P value  $\leq 0.05$  was interpreted to be of significance.

*2. Systematical computational analyses:* To prioritize candidate risk genes and functional categories for PBC, we also performed systematical bioinformatics analyses (see Supplemental Methods S1-S7 for details). For using the converging effects of multiple SNPs in a given gene, we applied a multi-variant converging regression model to conduct a gene-level association analysis<sup>23,24</sup> of the GWAS summary statistics (Supplemental Methods S1). The disease- and GO-terms enrichment analyses were conducted to reveal the functions of these genetics-risk genes (Supplemental Methods S2). The Multidimensional scaling analysis was used to construct a functional module for identified pathways (Supplemental Methods S3)<sup>11</sup>, and the GeneMANIA tool<sup>25</sup> was used to establish a protein-protein interaction network of these genetics-risk genes (Supplemental Methods S4). To highlight the functional risk genes whose expressions were significantly associated with PBC, we leveraged both S-PrediXcan<sup>26</sup> and S-MultiXcan<sup>27</sup> to combine GWAS summary statistics on PBC



with GTEx eQTL data<sup>28</sup> (Supplemental Methods S5). Furthermore, We referenced the method used in previous studies<sup>11,24,29,30</sup> to perform an *in silico* permutation analysis of 100,000 times of random selections ( $N_{Total}$ ) for validating the consistency of results from S-MultiXcan analysis with other results from three distinct analyses: MAGMA analysis, and S-PrediXcan on liver and blood (Supplemental Methods S6). We also performed a gene-property analysis using the FUMA tool<sup>31</sup> to integrate gene-level association signals from GWAS summary statistics on PBC with scRNA-seq data based on liver tissue from the Mouse Cell Atlas<sup>32</sup>.

### ***Statistical analysis***

Differential gene expression (DGE) analyses between controls and PBC patients of three RNA expression datasets (i.e., GSE93170, GSE159676, and GSE119600) were examined by using the Student's T-test. P value  $\leq 0.05$  was of significance. We also performed a co-expression pattern analysis in the dataset of GSE93170 for genetics-risk genes among PBC and healthy controls to evaluate whether the co-expression patterns were changed due to the disease status. The PLINK (v1.90)<sup>33</sup> was used to calculate the LD values among SNPs. The hypergeometric test was used to evaluate the significant enrichment for the disease- and GO-term enrichment analysis. The Jaccard distance algorithm<sup>34</sup> was used to assess the similarities among pathways.

## **Results**

### ***Workflow for the current genomics analysis***

To explore the genetics-influenced cell types for PBC, we constructed a computational framework by integrating GWAS summary statistics, eQTL data, and single cell RNA sequencing

data. The whole procedure of this framework is shown in Figure 1. There were 8,444 parenchymal and non-parenchymal cells obtained from five samples' transcriptome profiles in the GSE115469 dataset, and we yielded 13 distinct human liver cell subpopulations (see Methods). We also collected a large-scale GWAS summary statistics on PBC (N = 13,239) for extracting useful genetic information and statistical values. Using a regression-based polygenic model<sup>21</sup>, we combined the single cell transcriptomes with genetic association signals from the GWAS summary statistics for genetically mapping the single cell landscape of liver tissue implicated in PBC (Figure 1a, and Supplemental Table S1). Additionally, using the comprehensive bioinformatics analyses, including MAGMA gene-level association analysis, S-MultiXcan-based integrative analysis, S-PrediXcan-based integrative analysis, *in silico* permutation analysis, MDS analysis, network-based enrichment analysis, DEG analysis, drug-gene interaction analysis, pathway enrichment analysis, and functional annotation analysis, we highlighted novel genetics-risk genes associated with PBC susceptibility (Figure 1b). Finally, we used genetic association signals to determine the functional subsets of liver- and immune-based cell types relevant to PBC (Figure 1c).

### ***Identification of gene-level genetic associations for PBC***

To identify the aggregated effects of SNPs in a given gene on PBC, we carried out a gene-level genetic association analysis and found that 563 genes were significantly associated with PBC (FDR  $\leq 0.05$ , Supplemental Figure S1 and Table S2). For example, the top-ranked genes of *HLA-DPA1* (P =  $2.69 \times 10^{-24}$ ), *IL12A* (P =  $6.63 \times 10^{-22}$ ), *HLA-DPB1* (P =  $1.11 \times 10^{-21}$ ), *BTNL2* (P =  $1.65 \times 10^{-17}$ ), *RXRB* (P =  $8.88 \times 10^{-17}$ ), *SLC39A7* (P =  $8.83 \times 10^{-17}$ ), *HLA-DRB1* (P =  $3.51 \times 10^{-13}$ ), *HLA-DQB1* (P =  $6.39 \times 10^{-13}$ ), *IRF5* (P =  $7.18 \times 10^{-11}$ ), *IL12RB2* (P =  $3.03 \times 10^{-10}$ ), *C6ORF10* (P =  $5.58 \times 10^{-10}$ ), *HLA-DRA*

( $P = 5.58 \times 10^{-10}$ ), and *CLEC16A* ( $P = 6.78 \times 10^{-10}$ ). Among these significant genes, there were 64 genes, including *IL12A*, *HLA-DPB1*, *BTNL2*, *HLA-DRB1*, *HLA-DQB1*, *IRF5*, and *IL12RB2*, reported to be associated with PBC documented in the GWAS catalog database<sup>5-9</sup>.

Furthermore, we conducted a genome-wide pathway enrichment analysis, and observed that 41 biological pathways were significantly enriched ( $FDR \leq 0.05$ , Supplemental Table S3). Interestingly, the top-ranked pathways were relevant to autoimmune diseases (Supplemental Figure S2a), including Th1 and Th2 cell differentiation ( $FDR = 3.27 \times 10^{-13}$ ), allograft rejection ( $FDR = 5.18 \times 10^{-7}$ ), inflammatory bowel disease ( $FDR = 1.82 \times 10^{-6}$ ), type I diabetes mellitus ( $FDR = 5.27 \times 10^{-6}$ ), and intestinal immune network for IgA production ( $FDR = 5.27 \times 10^{-6}$ ). Based on the MDS analysis, these 41 significant pathways were grouped into five clusters: Th1 and Th2 cell differentiation, allograft rejection, Th17 cell differentiation, cell adhesion molecules, and cytokine-cytokine receptor interactions (Supplemental Figure S2b). These results suggested that these identified genes have a more likelihood to be risk genes for PBC.

### ***Integrative analysis of GWAS summary statistics with eQTL data for PBC***

To further highlight the functional genes whose expressions are associated with PBC, we leveraged the S-MultiXcan software<sup>27</sup> to meta-analyze tissue-specific associations across 49 GTEx tissues. There were 268 risk genes whose genetically-associated expression showing notable associations with PBC ( $FDR < 0.05$ , Figure 2 and Supplemental Table S4). For example, the top-ranked PBC-associated genes: *HLA-DRB1* ( $P = 4.95 \times 10^{-69}$ ), *HLA-DQA1* ( $P = 1.03 \times 10^{-43}$ ), *HLA-DRB5* ( $P = 1.17 \times 10^{-42}$ ), *BTNL2* ( $P = 3.26 \times 10^{-41}$ ), *EGFL8* ( $P = 1.37 \times 10^{-31}$ ), and *IL12A* ( $P = 5.42 \times 10^{-29}$ ). Among these significant genes, there were 52 risk genes, including *HLA-DRB1*, *BTNL2*,

*HLA-DPBI*, *IL12A*, *HLA-DQB1*, *IRF5* and *IL12RB2*, having been documented in the GWAS Catalog database (Supplemental Table S4). There was a high consistency of results between MAGMA and S-MultiXcan analysis (232/268 = 86.6%, Supplemental Figure S3).

We further validated these risk genes in two PBC-relevant tissues (i.e., liver and blood) using the S-PrediXcan method. 76 and 115 genes were demonstrated to be significantly associated with PBC in liver and blood tissue, respectively (FDR < 0.05, Supplemental Figure S4a-b and Supplemental Tables S5-S6). Notably, 139 genes identified from MAGMA and S-MultiXcan analysis were replicated by using the S-PrediXcan analysis based on liver and blood (Supplemental Figure S4c). Using the Pearson correlation analysis, we observed that significant genes from S-MultiXcan analysis showed remarkable correlations with that from MAGMA and S-PrediXcan analysis on liver and blood ( $r = 0.14\sim 0.64$ ,  $P < 0.05$ , Supplemental Figure S4d-f). Moreover, by performing three independent *in silico* permutation analyses (see Methods), we found that the number of observed overlapped genes between S-MultiXcan and MAGMA and S-PrediXcan were significantly higher than random events (Empirical  $P < 1 \times 10^{-5}$ , Supplemental Figures S4g-i and S5a-c). Overall, based on aforementioned integrative genomics analyses, we identified that there were 29 genes showing supportive evidence to involve in PBC susceptibility (Table 1).

### ***Functional analysis of these 29 PBC-associated risk genes***

Through mining the PubMed literature and GWAS Catalog database, we found that 10 of 29 identified risk genes have been reported to be associated with PBC in previous GWAS studies, and there were 19 novel genes, including *GSNK2B*, *LY6G5B*, *DDAH2*, and *LY6G5C*, newly identified (Supplemental Figure S6a). Among these novel identified genes, several have been demonstrated to

be associated with other autoimmune-related diseases, such as *LY6G5B* and *DDAH2* exhibited significant associations with rheumatoid arthritis<sup>35</sup> and type 1 diabetes<sup>36</sup>. Consistently, we performed a phenotype-based enrichment analysis, and found that these 29 risk genes were remarkably enriched in several phenotypes relevant to autoimmune diseases (Supplemental Figure S6b and Table S7), such as autoimmune disease ( $P = 8.47 \times 10^{-8}$ ), type I diabetes mellitus ( $P = 2.04 \times 10^{-4}$ ), immune system diseases ( $P = 1.41 \times 10^{-3}$ ), and juvenile rheumatoid arthritis ( $P = 2.41 \times 10^{-3}$ ).

By conducting a PPI network enrichment analysis, we observed that these risk genes were significantly interacted with each other in a subnetwork (PPI enrichment  $P = 0.0016$ , Supplemental Figure S6c), suggesting that these identified genes may have collectively biological functions on PBC risk<sup>34,37</sup>. Furthermore, we also performed a GO-term enrichment analysis according to three different categories (i.e., CC, MF, and BP), and found that several GO-terms were notably overrepresented (Supplemental Figures S6d, and S7-S9), such as interferon-gamma-mediated signaling pathway ( $P = 2.8 \times 10^{-4}$ ) and neutrophil activation involved in immune response ( $P = 7.3 \times 10^{-4}$ ).

Based on three independent expression profiles of liver (GSE159676), blood (GSE119600), and peripheral CD4+T cells (GSE93170), we conducted differential gene expression analysis for these 29 risk genes between PBC and matched control group. We found that 22 of 29 genes (75.86%) showed significantly differential expressions among PBC patients compared with controls (Table 1, Supplemental Figures S10-S12). Based on the dataset of peripheral CD4+T cells, the co-expression patterns among these 29 genes were prominently altered according to the PBC status (Supplemental Figure S12a). These results further support these identified genes have functional effects on the

development of PBC.

### ***Drug-gene interaction analysis for 29 risk genes***

Based on the drug-gene interaction analysis, we identified that 20 of 29 genes (68.96%) were enriched in ten potential “druggable” gene categories, including enzyme, druggable genome, kinase, and clinically actionable (Supplemental Figure S6e and Supplemental Table S8). The gene of *SOCS1* was found to be targeted by insulin and aldesleukin (Supplemental Figure S13), of which both have been applied to treat autoimmune diseases, including systemic lupus erythematosus<sup>38</sup>, type 1 diabetes mellitus<sup>39</sup>, and HIV<sup>40</sup>. Additionally, *TCF19* gene was targeted by nevirapine, *NAAA* was targeted by cyclopentyl palmitate, *CARMI* was targeted by BIIB021 and eosin, and *SH2B3* was targeted by ruxolitinib and candesartan. Previous studies<sup>41</sup> have demonstrated that the Janus kinase (JAK)-inhibitor ruxolitinib significantly influenced dendritic cell differentiation and function resulting in impaired T-cell activation, which could be used for the treatment of autoimmune diseases. These results provide a good drug repurposing resource to develop effective therapeutics for PBC.

### ***Identification of genetically-influenced liver cell subpopulations for PBC***

We performed a gene-property analysis based on mouse liver tissue with immune cells, and found that PBC-associated genes were enriched in several immune-related cell types (Figure 3a and Supplemental Table S9), including *Cst3*<sup>+</sup> dendritic cell ( $P = 5.8 \times 10^{-3}$ ), *Trbc2*<sup>+</sup> T cell ( $P = 7.1 \times 10^{-3}$ ), *Chil3*<sup>+</sup> macrophage ( $P = 0.0167$ ), and *Gzma*<sup>+</sup> T cell ( $P = 0.0304$ ), which were in line with the results in previous studies<sup>42,43</sup>. To uncover genetics-regulatory cell subpopulations associated with PBC, we used a regression-based polygenic model to combine GWAS summary statistics on PBC with

scRNA-seq data with 13 distinct human liver cell types. We found that cholangiocytes showed a notable enrichment by PBC-relevant genetic association signals (Permuted  $P < 0.05$ , Figure 3b and Supplemental Table S10). These results are consistent with previous evidence that an immune-mediated injury of cholangiocytes contributes risk to PBC<sup>44,45</sup>.

Using the specificity algorithm in the MAGMA method<sup>23</sup>, we noticed that the risk gene of *ORMDL3* exhibited the highest expression in cholangiocytes than other cell types (Figure 3c and Supplemental Table S11). The majority of *ORMDL3*-expressing cells were cholangiocytes with a relative high percentage of 22.38%, reminiscing that *ORMDL3* was demonstrated to be significantly associated with PBC in earlier studies<sup>7,9,46</sup>. The top-ranked risk SNP associated with *ORMDL3* is rs9303277 ( $P = 2.57 \times 10^{-11}$ , Figure 3d). This SNP showed significant cell-specific eQTL of *ORMDL3* among naïve B cell ( $P = 5.8 \times 10^{-10}$ ), TFH<sup>+</sup>CD4<sup>+</sup>T cell ( $P = 3.8 \times 10^{-9}$ ), CD56<sup>dim</sup> CD16<sup>+</sup> NK cells ( $P = 4.4 \times 10^{-9}$ ), TH1<sup>+</sup> CD4<sup>+</sup>T cells ( $P = 6.0 \times 10^{-6}$ ), and memory TREG<sup>+</sup> CD4<sup>+</sup>T cells ( $P = 1.1 \times 10^{-5}$ ), and exhibited notable cell-specific promoter-interacting eQTL of *ORMDL3* among naïve B cell ( $P = 4.2 \times 10^{-13}$ ) and CD56<sup>dim</sup> CD16<sup>+</sup> NK cells ( $P = 5.0 \times 10^{-12}$ ) (Figure 3e and Supplemental Methods S8). Among these immune cell types, the CC genotype of rs9303277 shows prominent association with higher expression of *ORMDL3* compared with other genotypes (Supplemental Figure S14a-k).

Growing attentions have concentrated on the role of *ORMDL3* in the development of inflammatory diseases, including PBC<sup>7,9,47</sup>. In view of the main goal of current study was to characterize genetics-modulated liver cell subpopulations for PBC, the majority of our subsequent analyses focused on revealing the effects of *ORMDL3*-mediated cholangiocytes.

### ***Characterization of biological functions of ORMDL3-mediated cholangiocytes***

We compared the expression profiles of *ORMDL3*<sup>+</sup> cholangiocytes with *ORMDL3*<sup>-</sup> cholangiocytes, and found that there were 77 significantly differential expressed genes (DEGs) with six up-regulated DEGs and 71 down-regulated DEGs among *ORMDL3*<sup>+</sup> cholangiocytes (FDR < 0.05, Figure 4a and Supplemental Tables S13-S14). With regard to six up-regulated DEGs, two genes of *GALNT1* and *HERPUD2* have been reported to attenuate inflammatory responses<sup>48,49</sup>. There were seven significant KEGG pathways overrepresented, and the most significant one is Herpes simplex virus 1 infection (FDR < 0.05, Figure 4c and Supplemental Table S15). Among these 71 down-regulated DEGs, *HLA-DRA* has been shown to be associated with PBC in previous GWASs<sup>7</sup>, and *CXCL8*, *CCL3*, *CXCL1*, *TIMP1*, *SPP1*, and *IRF1* were inflammatory and cytokine genes, which have been reported to be linked with the chemotaxis of immune cells that efflux to the site of cytokine storms in response to ongoing tissue damage<sup>50,51</sup>. *IFI16* is reported to be an innate immune sensor for intracellular DNA<sup>52</sup>.

Pathway enrichment analysis revealed that there were 55 significant KEGG pathways enriched by these 71 down-regulated DEGs (FDR < 0.05, Figure 4b and Supplemental Table S14), such as, hematopoietic cell lineage, inflammatory bowel disease, rheumatoid arthritis, Th17 cell differentiation, cytokine-cytokine receptor interaction, and chemokine signaling pathway, of which several have been demonstrated to implicate in inflammatory-related diseases<sup>11,53,54</sup>. To further explore the biological functions of these 71 down-regulated DEGs, we conducted GO-term enrichment analysis according to three categories of biological process (BP), cellular component (CC), and molecular function (MF). There were 19 BP-terms, 7 CC-terms, and 13 MF-terms showing notable enrichments, respectively (FDR < 0.05, Figure 4d-f and Supplemental Tables S16-S18). These functional terms were largely linked with immune-related functions, such as



granulocyte activation, neutrophil mediated immunity, S100 protein binding, and MHC class II protein complex binding. Moreover, these genes were prominently enriched in six disease-related terms, including liver cirrhosis, chemical and drug induced liver injury, acute coronary syndrome, acute kidney injury, neoplasm invasiveness, and chronic obstructive airway disease (FDR < 0.05, Supplemental Figure S15 and Table S19). These results suggest that *ORMDL3*<sup>-</sup> cholangiocytes have immune-modulatory effects on PBC risk.

## Discussion

Using an integrative genomics approach based on multiple layers of evidence, we identified that the genetics-influenced expression of 29 risk genes were remarkably associated with PBC. Among them, 10 genes, including *IRF5*<sup>5,6,8,9</sup>, *SOCS1*<sup>5</sup>, *SYNGR1*<sup>5,7</sup>, *ORMDL3*<sup>7,9,46</sup>, *MANBA*<sup>5</sup>, *IDUA*<sup>5</sup>, *DGKQ*<sup>5</sup>, *FCRL3*<sup>55</sup>, *NAAA*<sup>56</sup>, and *SH2B3*<sup>5</sup>, have been documented to be associated with PBC. There were 19 newly identified PBC-risk genes, such as *GSNK2B*, *LY6G5B*, *DDAH2*, *C6orf48*, and *HLA-DMA*. The novel risk gene of *DDAH2*, which encodes a dimethylarginine dimethylaminohydrolase, has been reported to be associated with the autoimmune diseases of rheumatoid arthritis<sup>57</sup> and type I diabetes<sup>58</sup>. The *HLA-DMA* gene, which belongs to the major histocompatibility complex (MHC) class II alpha chain paralogues, has been widely reported to be associated with autoimmune diseases, including type I diabetes mellitus<sup>59</sup>, rheumatoid arthritis<sup>60</sup>, and systemic lupus erythematosus<sup>61</sup>. The *ORMDL3* gene, which has been associated with PBC susceptibility<sup>7,9,46</sup>, is related to biological functions of innate immune system and metabolism. Moreover, *ORMDL3* has also been extensively reported to be linked with other inflammatory diseases, including childhood asthma<sup>62</sup>, inflammatory bowel diseases<sup>53</sup>, rheumatoid arthritis<sup>54</sup>, and Crohn's disease<sup>63</sup>. Functional enrichment analyses

uncovered that these risk genes were notably enriched in several biological processes or disease-terms that are relevant to autoimmune phenotypes<sup>64,65</sup>. Together, these results suggest these identified genes are more likely to be genuine genes implicated in PBC risk.

To examine the liver and its immune microenvironment for PBC, we performed the gene-property analysis and regression-based polygenic analysis based on human and mouse liver tissues with immune cells. Our gene-property analysis identified that four immune cell subpopulations, *Cst3*<sup>+</sup> dendritic cell, *Chil3*<sup>+</sup> macrophage, *Trbc2*<sup>+</sup> T cell, and *Gzma*<sup>+</sup> T cell, are significantly associated with PBC risk. Dendritic cells have been shown to be relevant to the pathogenesis of PBC<sup>42,66</sup>. Earlier studies have demonstrated that an intense biliary inflammatory CD8<sup>+</sup> and CD4<sup>+</sup> T cell response has been used for characterizing PBC<sup>67</sup>. Moreover, dendritic cells are critical for inducing antigen-specific T-cell tolerance and for activating the self-specific T cells, which have central roles in many aspects of the pathogenesis of autoimmune liver diseases<sup>68</sup>. Macrophages are key components of the innate immune system, and have critical immuno-modulatory and tissue-repairing roles in reducing immune responses and enhancing tissue regeneration<sup>69</sup>. Multiple lines of evidence<sup>70</sup> have demonstrated that the infiltration of macrophages in diseased tissues is considered to be a hallmark of several autoimmune diseases, including PBC<sup>43</sup>. Recently, Dorris et al.<sup>71</sup> reported that the PBC susceptibility gene *C5orf30* modulates macrophage-mediated immune regulations. Overall, these results suggest that three main immune cells, dendritic cells, macrophages, and T cells, are vulnerable to be influenced by PBC-relevant genetic association signals.

Moreover, our regression-based polygenic analysis suggested that cholangiocytes were significantly influenced by PBC-related genetic association signals. The non-parenchymal cholangiocytes have been reported to be injury in numerous human diseases termed as

cholangiopathies<sup>72</sup>, including PBC<sup>72,73</sup>. Recently, Banales and coworkers<sup>72</sup> have demonstrated that cholangiocytes play pivotal roles in innate and adaptive immune responses relevant to immune-mediated cholangiopathies. Erice et al.<sup>74</sup> reported that microRNA-506 induces PBC-like features in human cholangiocytes and promotes the activating processes of immune responses. We further found that *ORMDL3*<sup>+</sup> cholangiocytes predisposed to attenuate the release of inflammatory factors, whereas *ORMDL3*<sup>-</sup> cholangiocytes, containing numerous highly-expressed inflammatory cytokine genes, tend to have genetics-mediated immune regulation for PBC risk. Many biological pathways or functional terms associated with autoimmune-related diseases have been significantly enriched, such as Th17 cell differentiation, cytokine-cytokine receptor interaction, and chemokine signaling pathway, inflammatory bowel diseases, and rheumatoid arthritis<sup>11,53,54</sup>. Overall, these results suggest that down-expression of *ORMDL3* has a vital role in the immune-mediated injury of cholangiocytes conferring high risk to PBC.

There exist some limitations should be cautious. Although we leveraged integrated bioinformatics methods to highlight PBC-associated risk genes based on multiple omics data, there were many potential risk genes with suggestive evidence for PBC as shown in the supplemental tables needed to be further studied. Furthermore, due to the heterogeneity across different datasets used in the present investigation, we leveraged different statistical methods for multiple testing correction for each dataset, such as FDR < 0.05 for MAGMA-based gene association analysis and S-MultiXcan analysis, permuted P < 0.05 for genome-wide pathway enrichment analysis, Bonferroni corrected P < 0.05 for gene-property analysis, and empirical P value < 0.05 for *in silico* permutation analysis. In view of current integrative genomic analysis is only based on samples derived from European ancestries, more studies based on other ancestries should be performed to validate the

effects of these risk genes on PBC.

In conclusion, current study provides multiple lines of evidence to support 29 genes including 19 novel genes are remarkably associated with PBC susceptibility. To the best of our knowledge, this is the first study to parse genetics-influenced human liver cell subpopulations that contribute risk to PBC, and found that *ORMDL3*<sup>+</sup> cholangiocytes potentially play important immune-regulatory roles in the pathogenesis of PBC. Current study gives several highlighted genes and genetics-risk cell type for functional experimentations to reveal the genetic mechanisms of PBC.

## Figure Legends

**Figure 1. The workflow of current integrative genomics analysis.** a) Integrating single cell RNA-sequencing data with GWAS summary statistics on PBC based on a regression-based polygenic model; b) Prioritization of PBC-risk genes by integrating GWAS summary data with GTEx eQTL data; c) Genetics-influenced liver cell subpopulations and its immune microenvironment for PBC.

**Figure 2. Circos plot showing the results of S-MultiXcan integrative genomics analysis.** Note: A circular symbol in the outer ring represents a given gene. Color represents the statistical significance of the gene, where red color marks significant genes with  $FDR \leq 1 \times 10^{-8}$ , orange color marks significant genes with  $FDR$  is between  $1 \times 10^{-8}$  and 0.001, light blue indicates significant genes with  $FDR$  ranging from 0.001 to 0.05, and dark blue indicates non-significant genes with  $FDR > 0.05$ .

**Figure 3. Combination of GWAS summary statistics on PBC with scRNA-seq data on human liver tissue.** a) Results of MAGMA gene-property analysis based on the web-access tool of FUMA. Gray color means non-significant enrichment. Orange color stands for significant enrichment. b) Bar graph showing genetics-influenced liver cell subpopulations for PBC. Orange color represents PBC-relevant genetic association signals showing a significant enrichment in cholangiocytes. c) Dot plot showing the expression percentage of 29 PBC-risk genes for each cell type from human liver tissue. The color stands for the average expression of each gene in each cell type, and the size of circular symbol indicates the percentage of a given gene expressed in each cell type. d) Regional association plot for *ORMDL3* gene of GWAS summary statistics using the *LocusZoom* software. The color bar illustrates SNPs' LD information with rs9303277 as shown in the color legend. e) Immune cell-type specific eQTL and promoter-interacting eQTL (pieQTL) of *ORMDL3* for rs9303277 using the DICE database.

**Figure 4. Differential gene expression analysis between *ORMDL3*<sup>+</sup> and *ORMDL3*<sup>-</sup> cholangiocytes.** a) Volcano plot showing the differentially expressed genes (DEGs) between *ORMDL3*<sup>+</sup> cholangiocytes and *ORMDL3*<sup>-</sup> cholangiocytes. Green color represents 71 significantly down-regulated DEGs, and orange color represents 6 significantly up-regulated DEGs. b)-c) Pathway enrichment analyses of (b) 71 down-regulated DEGs and (c) 6 up-regulated DEGs based on the KEGG pathway resource. d)-f) GO enrichment analyses of 71 down-regulated DEGs according to three categories: (d) biological process, (e) cellular component, and (f) molecular function.

## Acknowledgements

We appreciate Cordell and his colleagues who have deposited and shared GWAS summary data on primary biliary cirrhosis in the IEU open GWAS project.

## Conflict of Interest Statement

The authors declare no conflict of interest.

## Funding

This study was funded by the Scientific Research Foundation for Talents of Wenzhou Medical University (KYQD20201001 to Y.M.).

## Data Availability

All the GWAS summary data applied in the present analysis can be obtained from the IEU open GWAS project (<https://gwas.mrcieu.ac.uk/>). The GTEx eQTL data (version 8) can be gained from the Zenodo repository (<https://zenodo.org/record/3518299#.Xv6Z6igzbg1>). All scripts used in the Methods are deposited in the public available GitHub repository: [https://github.com/mayunlong89/PBC\\_project](https://github.com/mayunlong89/PBC_project).

## References

1. Beuers U, Gershwin ME, Gish RG, et al. Changing nomenclature for PBC: From 'cirrhosis' to 'cholangitis'. *J Hepatol*. 2015;63(5):1285-1287.
2. Lindor KD, Bowlus CL, Boyer J, Levy C, Mayo M. Primary Biliary Cholangitis: 2018 Practice Guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2019;69(1):394-419.
3. Rong G, Zhong R, Lleo A, et al. Epithelial cell specificity and epitope recognition by serum autoantibodies in primary biliary cirrhosis. *Hepatology*. 2011;54(1):196-203.

4. Tanaka A, Leung PSC, Gershwin ME. The genetics of primary biliary cholangitis. *Curr Opin Gastroenterol*. 2019;35(2):93-98.
5. Cordell HJ, Han Y, Mells GF, et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun*. 2015;6:8019.
6. Hirschfield GM, Liu X, Xu C, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N Engl J Med*. 2009;360(24):2544-2555.
7. Qiu F, Tang R, Zuo X, et al. A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. *Nat Commun*. 2017;8:14828.
8. Liu X, Invernizzi P, Lu Y, et al. Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. *Nat Genet*. 2010;42(8):658-660.
9. Mells GF, Floyd JA, Morley KI, et al. Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat Genet*. 2011;43(4):329-332.
10. Juran BD, Hirschfield GM, Invernizzi P, et al. Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum Mol Genet*. 2012;21(23):5209-5221.
11. Ma Y, Huang Y, Zhao S, et al. Integrative Genomics Analysis Reveals a 21q22.11 Locus Contributing Risk to COVID-19. *Hum Mol Genet*. 2021.
12. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481-487.
13. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet*. 2014;94(4):559-573.
14. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235.
15. Zhang Y, Ma Y, Huang Y, et al. Benchmarking algorithms for pathway activity transformation of single-cell RNA-seq data. *Comput Struct Biotechnol J*. 2020;18:2953-2961.
16. Zhang F, Wei K, Slowikowski K, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol*. 2019;20(7):928-942.
17. Corridoni D, Chapman T, Antanaviciute A, Satsangi J, Simmons A. Inflammatory Bowel Disease Through the Lens of Single-cell RNA-seq Technologies. *Inflamm Bowel Dis*. 2020;26(11):1658-1668.
18. Nehar-Belaid D, Hong S, Marches R, et al. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat Immunol*. 2020;21(9):1094-1106.
19. Macosko EZ, Basu A, Satija R, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161(5):1202-1214.
20. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816-834.
21. Calderon D, Bhaskar A, Knowles DA, et al. Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. *Am J Hum Genet*. 2017;101(5):686-699.
22. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
23. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219.
24. Xu M, Li J, Xiao Z, Lou J, Pan X, Ma Y. Integrative genomics analysis identifies promising SNPs and genes implicated in tuberculosis risk based on multiple omics datasets. *Aging (Albany NY)*. 2020;12(19):19173-19220.
25. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network

- integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38(Web Server issue):W214-220.
26. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun.* 2018;9(1):1825.
  27. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.* 2019;15(1):e1007889.
  28. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.
  29. Ma X, Wang P, Xu G, Yu F, Ma Y. Integrative genomics analysis of various omics data and networks identify risk genes and variants vulnerable to childhood-onset asthma. *BMC Med Genomics.* 2020;13(1):123.
  30. Dong Z, Ma Y, Zhou H, et al. Integrated genomics analysis highlights important SNPs and genes implicated in moderate-to-severe asthma based on GWAS and eQTL datasets. *BMC Pulm Med.* 2020;20(1):270.
  31. Watanabe K, Umičević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nat Commun.* 2019;10(1):3222.
  32. Han X, Wang R, Zhou Y, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell.* 2018;172(5):1091-1107.e1017.
  33. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
  34. Ma Y, Li J, Xu Y, et al. Identification of 34 genes conferring genetic and pharmacological risk for the comorbidity of schizophrenia and smoking behaviors. *Aging (Albany NY).* 2020;12(3):2169-2225.
  35. Hu HJ, Jin EH, Yim SH, et al. Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis. *Exp Mol Med.* 2011;43(11):613-621.
  36. Tomer Y, Dolan LM, Kahaly G, et al. Genome wide identification of new genes and pathways in patients with both autoimmune thyroiditis and type 1 diabetes. *J Autoimmun.* 2015;60:32-39.
  37. Ma Y, Li MD. Establishment of a Strong Link Between Smoking and Cancer Pathogenesis through DNA Methylation Analysis. *Sci Rep.* 2017;7(1):1811.
  38. von Spee-Mayer C, Siegert E, Abdirama D, et al. Low-dose interleukin-2 selectively corrects regulatory T cell defects in patients with systemic lupus erythematosus. *Ann Rheum Dis.* 2016;75(7):1407-1415.
  39. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet.* 2014;383(9911):69-82.
  40. Sundin DJ, Wolin MJ. Aldesleukin therapy in HIV-infected patients. *Am J Health Syst Pharm.* 1998;55(14):1520-1523.
  41. Heine A, Held SA, Daecke SN, et al. The JAK-inhibitor ruxolitinib impairs dendritic cell function in vitro and in vivo. *Blood.* 2013;122(7):1192-1202.
  42. Hiasa Y, Akbar SM, Abe M, Michitaka K, Horiike N, Onji M. Dendritic cell subtypes in autoimmune liver diseases; decreased expression of HLA DR and CD123 on type 2 dendritic cells. *Hepatol Res.* 2002;22(4):241-249.
  43. Harada K, Shimoda S, Ikeda H, et al. Significance of periductal Langerhans cells and biliary epithelial cell-derived macrophage inflammatory protein-3 $\alpha$  in the pathogenesis of primary biliary cirrhosis. *Liver Int.* 2011;31(2):245-253.
  44. Zhou T, Kyritsi K, Wu N, et al. Knockdown of vimentin reduces mesenchymal phenotype of cholangiocytes in the Mdr2(-/-) mouse model of primary sclerosing cholangitis (PSC). *EBioMedicine.* 2019;48:130-142.
  45. Ronca V, Mancuso C, Milani C, Carbone M, Oo YH, Invernizzi P. Immune system and cholangiocytes: A puzzling affair in primary biliary cholangitis. *J Leukoc Biol.* 2020;108(2):659-671.
  46. Nakamura M, Nishida N, Kawashima M, et al. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. *Am J Hum Genet.* 2012;91(4):721-728.



47. Das S, Miller M, Broide DH. Chromosome 17q21 Genes ORMDL3 and GSDMB in Asthma and Immune Diseases. *Adv Immunol*. 2017;135:1-52.
48. Sun CM, Zhang WY, Wang SY, Qian G, Pei DL, Zhang GM. microRNA let-7i-5p aggravates kidney fibrosis via targeting GALNT1. *Gen Physiol Biophys*. 2021;40(2):147-154.
49. Dong YF, Chen ZZ, Zhao Z, et al. Potential role of microRNA-7 in the anti-neuroinflammation effects of nicorandil in astrocytes induced by oxygen-glucose deprivation. *J Neuroinflammation*. 2016;13(1):60.
50. Rydzynski Moderbacher C, Ramirez SI, Dan JM, et al. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell*. 2020;183(4):996-1012.e1019.
51. Su Y, Chen D, Yuan D, et al. Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. *Cell*. 2020;183(6):1479-1495.e1420.
52. Unterholzner L, Keating SE, Baran M, et al. IFI16 is an innate immune sensor for intracellular DNA. *Nat Immunol*. 2010;11(11):997-1004.
53. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47(9):979-986.
54. Kurreeman FA, Stahl EA, Okada Y, et al. Use of a multiethnic approach to identify rheumatoid-arthritis-susceptibility loci, 1p36 and 17q12. *Am J Hum Genet*. 2012;90(3):524-532.
55. Tanaka A, Ohira H, Kikuchi K, et al. Genetic association of Fc receptor-like 3 polymorphisms with susceptibility to primary biliary cirrhosis: ethnic comparative study in Japanese and Italian patients. *Tissue Antigens*. 2011;77(3):239-243.
56. Zamanou A, Samiotaki M, Panayotou G, Margaritis L, Lymberi P. Fine specificity and subclasses of IgG anti-actin autoantibodies differ in health and disease. *J Autoimmun*. 2003;20(4):333-344.
57. Dimitroulas T, Sandoo A, Hodson J, Smith J, Panoulas VF, Kitis GD. Relationship between dimethylarginine dimethylaminohydrolase gene variants and asymmetric dimethylarginine in patients with rheumatoid arthritis. *Atherosclerosis*. 2014;237(1):38-44.
58. Fogarty RD, Abhary S, Javadiyan S, et al. Relationship between DDAH gene variants and serum ADMA level in individuals with type 1 diabetes. *J Diabetes Complications*. 2012;26(3):195-198.
59. Siegmund T, Donner H, Braun J, Usadel KH, Badenhop K. HLA-DMA and HLA-DMB alleles in German patients with type 1 diabetes mellitus. *Tissue Antigens*. 1999;54(3):291-294.
60. Morel J, Roch-Bras F, Molinari N, Sany J, Eliaou JF, Combe B. HLA-DMA\*0103 and HLA-DMB\*0104 alleles as novel prognostic factors in rheumatoid arthritis. *Ann Rheum Dis*. 2004;63(12):1581-1586.
61. Yen JH, Chen CJ, Tsai WC, Tsai JJ, Ou TT, Liu HW. HLA-DMA and HLA-DMB genotyping in patients with systemic lupus erythematosus. *J Rheumatol*. 1999;26(9):1930-1933.
62. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*. 2007;448(7152):470-473.
63. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008;40(8):955-962.
64. Grayson PC, Kaplan MJ. At the Bench: Neutrophil extracellular traps (NETs) highlight novel aspects of innate immune system involvement in autoimmune diseases. *J Leukoc Biol*. 2016;99(2):253-264.
65. Mo X, Guo Y, Qian Q, et al. Mendelian randomization analysis revealed potential causal factors for systemic lupus erythematosus. *Immunology*. 2020;159(3):279-288.
66. Kita H, Lian ZX, Van de Water J, et al. Identification of HLA-A2-restricted CD8(+) cytotoxic T cell responses in primary biliary cirrhosis: T cell activation is augmented by immune complexes cross-presented by dendritic cells. *J Exp Med*. 2002;195(1):113-123.
67. Coppel RL, Gershwin ME. Primary biliary cirrhosis: the molecule and the mimic. *Immunol Rev*. 1995;144:17-49.

68. Swiecki M, Colonna M. The multifaceted biology of plasmacytoid dendritic cells. *Nat Rev Immunol*. 2015;15(8):471-485.
69. Ma WT, Gao F, Gu K, Chen DK. The Role of Monocytes and Macrophages in Autoimmune Diseases: A Comprehensive Review. *Front Immunol*. 2019;10:1140.
70. Carlsen HS, Baekkevold ES, Morton HC, Haraldsen G, Brandtzaeg P. Monocyte-like and mature macrophages produce CXCL13 (B cell-attracting chemokine 1) in inflammatory lesions with lymphoid neogenesis. *Blood*. 2004;104(10):3021-3027.
71. Dorris ER, Tazzyman SJ, Moylett J, et al. The Autoimmune Susceptibility Gene C5orf30 Regulates Macrophage-Mediated Resolution of Inflammation. *J Immunol*. 2019;202(4):1069-1078.
72. Banales JM, Huebert RC, Karlsen T, Strazzabosco M, LaRusso NF, Gores GJ. Cholangiocyte pathobiology. *Nat Rev Gastroenterol Hepatol*. 2019;16(5):269-281.
73. Lleo A, Maroni L, Glaser S, Alpini G, Marzioni M. Role of cholangiocytes in primary biliary cirrhosis. *Semin Liver Dis*. 2014;34(3):273-284.
74. Erice O, Munoz-Garrido P, Vaquero J, et al. MicroRNA-506 promotes primary biliary cholangitis-like features in cholangiocytes and immune activation. *Hepatology*. 2018;67(4):1420-1440.

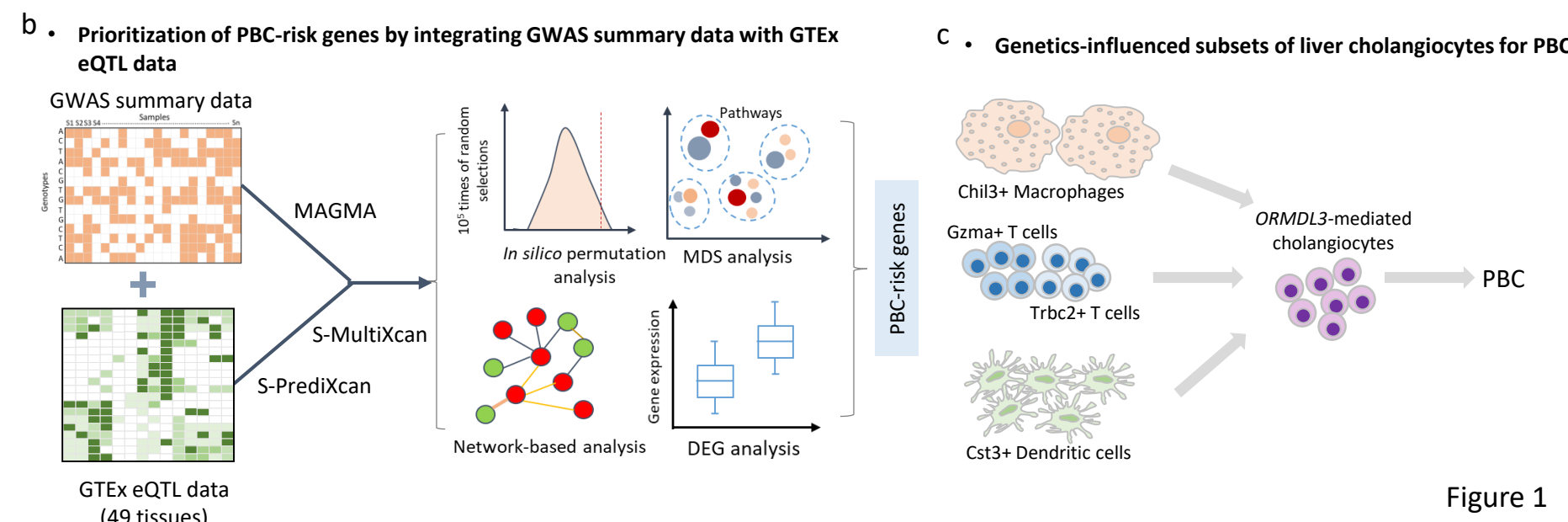
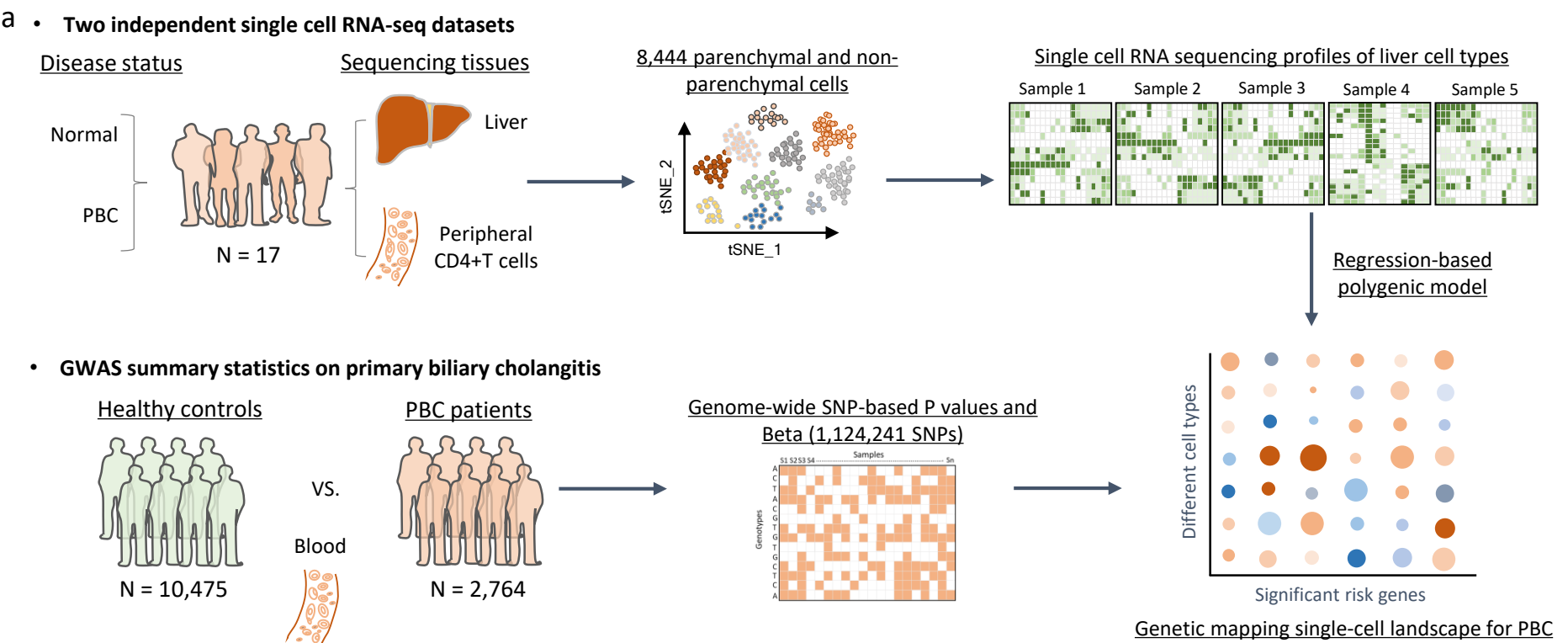


Figure 1

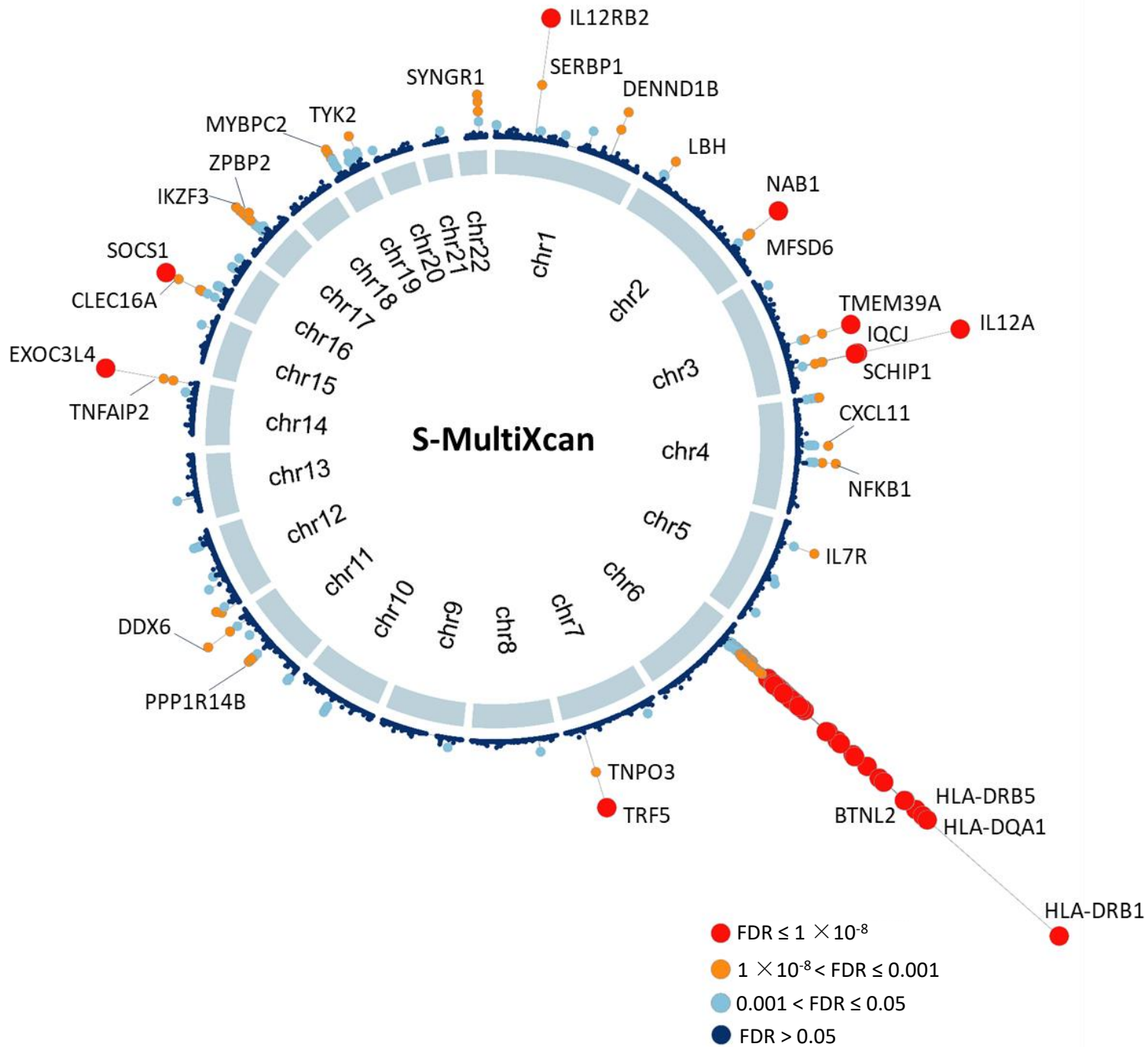
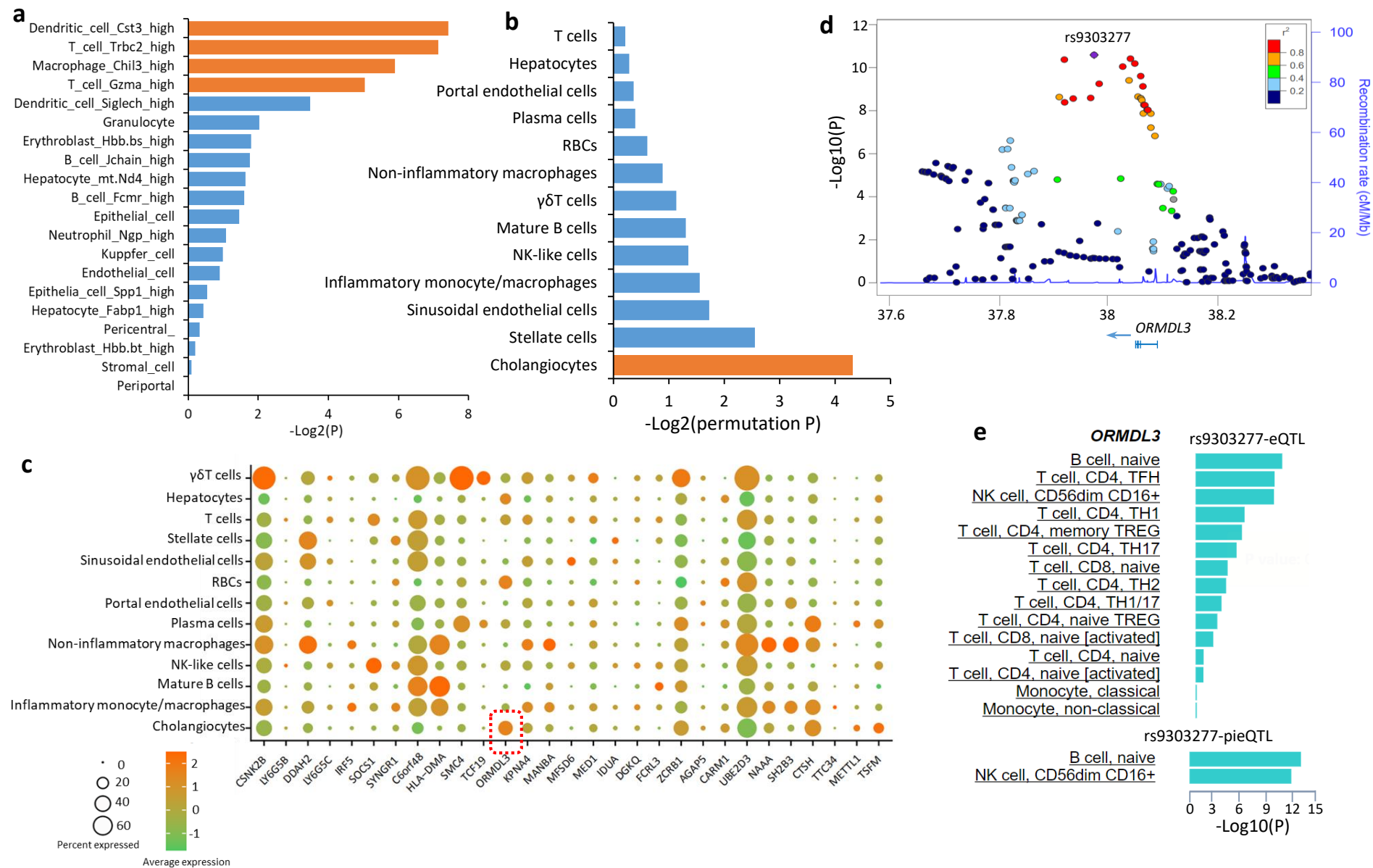


Figure 2



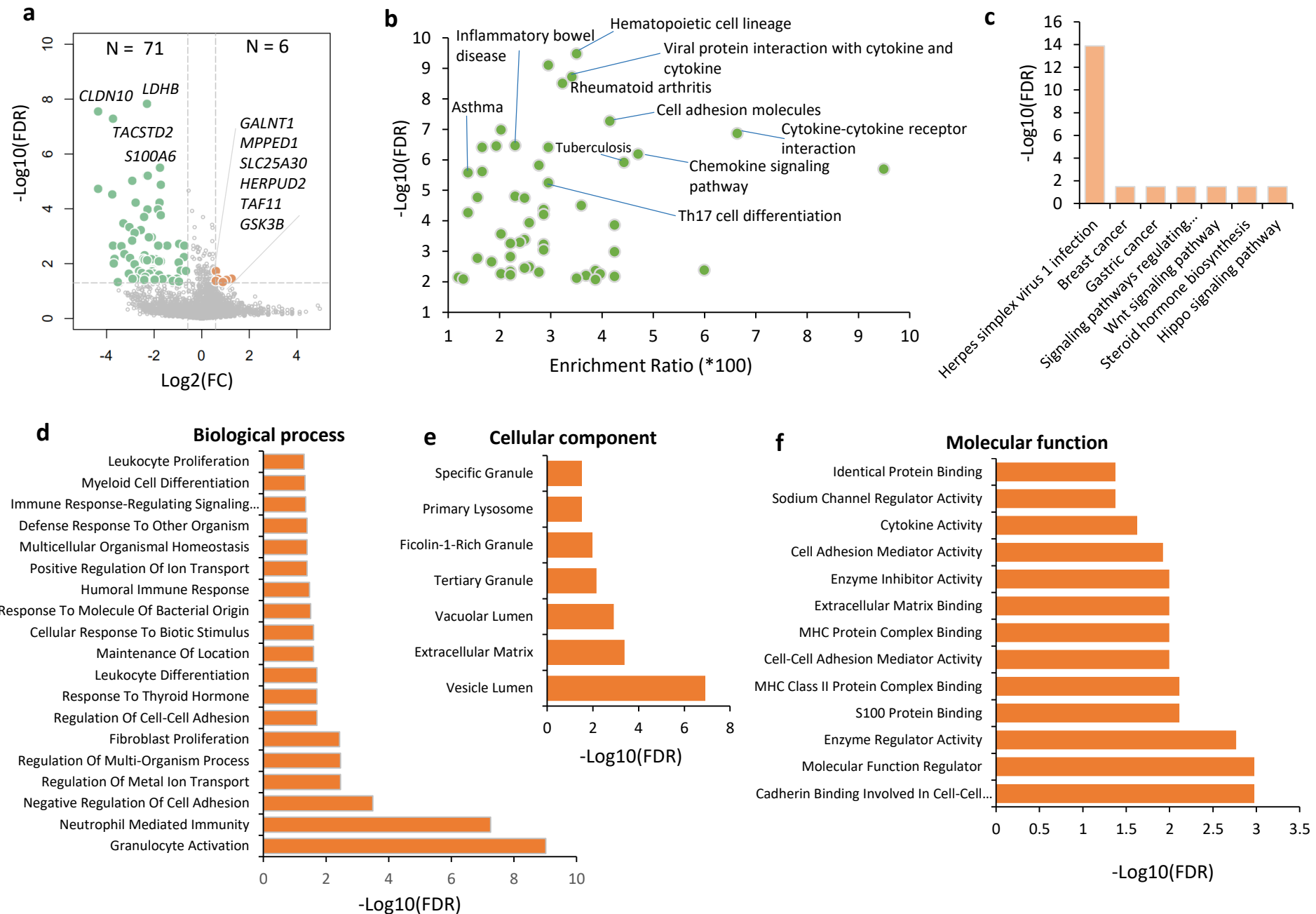


Figure 4

**Table 1. Identification of 29 significant risk genes by using integrative genomics analysis based on GWAS summary statistics and eQTL data**

Gene name	Z score	P value (S-MultiXcan)	FDR (S-MultiXcan)	FDR (S-PrediXcan on liver)	FDR (S-PrediXcan on blood)	FDR (MAGMA analysis of GWAS)	P value (Differential gene expression analysis)	GWAS_Catalog database or PubMed database
<i>CSNK2B</i>	1.40	$7.17 \times 10^{-19}$	$8.41 \times 10^{-16}$	$3.22 \times 10^{-11}$	$2.24 \times 10^{-6}$	$1.67 \times 10^{-11}$	$4.96 \times 10^{-4}$ (Blood), $2 \times 10^{-2}$ (CD14+T cell)	Novel gene
<i>LY6G5B</i>	-7.92	$2.43 \times 10^{-16}$	$1.93 \times 10^{-13}$	$5.52 \times 10^{-11}$	$1.54 \times 10^{-10}$	$1.19 \times 10^{-11}$	Non-significant	Novel gene
<i>DDAH2</i>	-5.57	$1.39 \times 10^{-15}$	$9.99 \times 10^{-13}$	$2.65 \times 10^{-9}$	$1.56 \times 10^{-11}$	$1.69 \times 10^{-9}$	$2.70 \times 10^{-3}$ (CD14+T cell)	Novel gene
<i>LY6G5C</i>	-7.82	$2.40 \times 10^{-15}$	$1.62 \times 10^{-12}$	$8.64 \times 10^{-13}$	$9.05 \times 10^{-12}$	$1.41 \times 10^{-11}$	$2.80 \times 10^{-2}$ (Blood)	Novel gene
<i>IRF5</i>	7.11	$2.62 \times 10^{-15}$	$1.72 \times 10^{-12}$	$4.29 \times 10^{-13}$	$1.12 \times 10^{-12}$	$7.18 \times 10^{-11}$	$7.10 \times 10^{-3}$ (Liver)	Reported gene
<i>SOCS1</i>	-3.26	$9.33 \times 10^{-13}$	$4.16 \times 10^{-10}$	$5.25 \times 10^{-10}$	$7.34 \times 10^{-10}$	$1.06 \times 10^{-8}$	$3.50 \times 10^{-2}$ (CD14+T cell)	Reported gene
<i>SYNGR1</i>	-1.30	$2.83 \times 10^{-9}$	$8.88 \times 10^{-7}$	$7.97 \times 10^{-4}$	$1.30 \times 10^{-7}$	$4.13 \times 10^{-2}$	$1.40 \times 10^{-2}$ (CD14+T cell)	Reported gene
<i>C6orf48</i>	-4.46	$9.57 \times 10^{-9}$	$2.66 \times 10^{-6}$	$7.83 \times 10^{-4}$	$6.19 \times 10^{-4}$	$4.83 \times 10^{-11}$	$4.16 \times 10^{-6}$ (Blood)	Novel gene

<i>HLA-DMA</i>	-3.99	8.87×10 <sup>-8</sup>	1.90×10 <sup>-5</sup>	1.27×10 <sup>-3</sup>	9.16×10 <sup>-4</sup>	1.62×10 <sup>-8</sup>	2.10×10 <sup>-3</sup> (Liver)	Novel gene
<i>SMC4</i>	-1.34	1.00×10 <sup>-7</sup>	2.10×10 <sup>-5</sup>	9.38×10 <sup>-5</sup>	2.25×10 <sup>-6</sup>	2.07×10 <sup>-4</sup>	4.80×10 <sup>-4</sup> (Blood)	Novel gene
<i>TCF19</i>	4.55	1.49×10 <sup>-7</sup>	2.99×10 <sup>-5</sup>	9.46×10 <sup>-5</sup>	2.24×10 <sup>-6</sup>	1.98×10 <sup>-4</sup>	Non-significant	Novel gene
<i>ORMDL3</i>	-2.92	7.07×10 <sup>-7</sup>	1.22×10 <sup>-4</sup>	2.35×10 <sup>-3</sup>	7.38×10 <sup>-5</sup>	2.33×10 <sup>-6</sup>	Non-significant	Reported gene
<i>KPNA4</i>	-1.22	1.07×10 <sup>-6</sup>	1.82×10 <sup>-4</sup>	2.10×10 <sup>-2</sup>	3.58×10 <sup>-3</sup>	4.67×10 <sup>-4</sup>	3.40×10 <sup>-3</sup> (CD14+T cell)	Novel gene
<i>MANBA</i>	-1.45	1.48×10 <sup>-6</sup>	2.35×10 <sup>-4</sup>	2.46×10 <sup>-6</sup>	2.39×10 <sup>-2</sup>	8.30×10 <sup>-8</sup>	9.67×10 <sup>-5</sup> (Blood)	Reported gene
<i>MFSD6</i>	-0.32	1.61×10 <sup>-6</sup>	2.51×10 <sup>-4</sup>	2.55×10 <sup>-2</sup>	4.57×10 <sup>-2</sup>	2.06×10 <sup>-3</sup>	3.60×10 <sup>-2</sup> (Liver)	Novel gene
<i>MED1</i>	-2.13	1.76×10 <sup>-6</sup>	2.70×10 <sup>-4</sup>	1.65×10 <sup>-2</sup>	1.53×10 <sup>-2</sup>	4.42×10 <sup>-5</sup>	Non-significant	Novel gene
<i>IDUA</i>	3.95	2.37×10 <sup>-6</sup>	3.47×10 <sup>-4</sup>	3.94×10 <sup>-2</sup>	3.87×10 <sup>-2</sup>	8.95×10 <sup>-4</sup>	8.10×10 <sup>-3</sup> (Blood)	Reported gene
<i>DGKQ</i>	3.24	3.86×10 <sup>-6</sup>	5.37×10 <sup>-4</sup>	4.90×10 <sup>-4</sup>	3.36×10 <sup>-4</sup>	2.35×10 <sup>-4</sup>	8.34×10 <sup>-7</sup> (Blood)	Reported gene
<i>FCRL3</i>	-4.40	8.14×10 <sup>-6</sup>	1.07×10 <sup>-3</sup>	2.84×10 <sup>-3</sup>	2.84×10 <sup>-3</sup>	5.81×10 <sup>-4</sup>	Non-significant	Reported gene
<i>ZCRB1</i>	3.22	9.34×10 <sup>-6</sup>	1.20×10 <sup>-3</sup>	4.83×10 <sup>-3</sup>	5.02×10 <sup>-3</sup>	1.52×10 <sup>-3</sup>	5.90×10 <sup>-3</sup> (Liver); 2.8×10 <sup>-3</sup> (CD14+T cell);	Novel gene



<i>AGAP5</i>	-4.28	1.29×10 <sup>-5</sup>	1.59×10 <sup>-3</sup>	4.50×10 <sup>-3</sup>	3.40×10 <sup>-3</sup>	3.65×10 <sup>-2</sup>	Non-significant	Novel gene
<i>CARMI</i>	-2.90	2.11×10 <sup>-5</sup>	2.49×10 <sup>-3</sup>	1.14×10 <sup>-3</sup>	8.92×10 <sup>-4</sup>	1.86×10 <sup>-4</sup>	2.60×10 <sup>-2</sup> (Blood)	Novel gene
<i>UBE2D3</i>	-4.18	2.83×10 <sup>-5</sup>	3.23×10 <sup>-3</sup>	2.76×10 <sup>-3</sup>	5.08×10 <sup>-3</sup>	5.81×10 <sup>-4</sup>	3.40×10 <sup>-2</sup> (Blood); 1.6×10 <sup>-4</sup> (CD14+T cell)	Novel gene
<i>NAAA</i>	4.08	6.73×10 <sup>-5</sup>	6.88×10 <sup>-3</sup>	1.25×10 <sup>-2</sup>	6.85×10 <sup>-3</sup>	7.41×10 <sup>-4</sup>	3.20×10 <sup>-2</sup> (Liver)	Reported gene
<i>SH2B3</i>	-1.69	1.20×10 <sup>-4</sup>	1.17×10 <sup>-2</sup>	2.35×10 <sup>-2</sup>	1.87×10 <sup>-2</sup>	2.44×10 <sup>-3</sup>	1.40×10 <sup>-2</sup> (Blood)	Reported gene
<i>CTSH</i>	1.75	2.05×10 <sup>-4</sup>	1.84×10 <sup>-2</sup>	8.53×10 <sup>-3</sup>	3.46×10 <sup>-2</sup>	2.19×10 <sup>-5</sup>	Non-significant	Novel gene
<i>TTC34</i>	3.56	2.82×10 <sup>-4</sup>	2.39×10 <sup>-2</sup>	5.86×10 <sup>-3</sup>	5.02×10 <sup>-3</sup>	4.98×10 <sup>-3</sup>	9.80×10 <sup>-3</sup> (Blood)	Novel gene
<i>METTL1</i>	2.19	3.11×10 <sup>-4</sup>	2.59×10 <sup>-2</sup>	4.95×10 <sup>-2</sup>	3.81×10 <sup>-2</sup>	1.68×10 <sup>-2</sup>	5.86×10 <sup>-7</sup> (Blood)	Novel gene
<i>TSFM</i>	3.45	3.14×10 <sup>-4</sup>	2.61×10 <sup>-2</sup>	4.87×10 <sup>-2</sup>	3.87×10 <sup>-2</sup>	1.54×10 <sup>-2</sup>	2.80×10 <sup>-4</sup> (Blood); 1.5×10 <sup>-3</sup> (CD14+T cell)	Novel gene