Single Cell Sequencing Analysis Identifies Genetics-Modulated ORMDL3+ Cholangiocytes Having Higher Metabolic Effects On Primary Biliary Cholangitis

Bingyu Xiang
Wenzhou Medical University
Chunyu Deng
Harbin Institute of Technology
Fei Qiu
Wenzhou Medical University
Jingjing Li
Zhejiang University
Shanshan Li
Wenzhou Medical University
Huifang Zhang
Wenzhou Medical University
Xiuli Lin
Wenzhou Medical University
Yukuan Huang
Wenzhou Medical University
Yijun Zhou
Wenzhou Medical University
Jianzhong Su
Wenzhou Medical University
Mingqin Lu
Wenzhou Medical University
Yunlong Ma (glb-biotech@zju.edu.cn)
Wenzhou Medical University https://orcid.org/0000-0002-1299-4802

Research

Keywords: cholangitis, immune-modulatory, etiology, metabolism, autoimmune

Posted Date: September 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-922624/v1

Version of Record: A version of this preprint was published at Journal of Nanobiotechnology on December 1st, 2021. See the published version at https://doi.org/10.1186/s12951-021-01154-2.

Single cell sequencing analysis identifies genetics-modulated *ORMDL3*$^{+}$ cholangiocytes having higher metabolic effects on primary biliary cholangitis

Bingyu Xiang[1,#], Chunyu Deng[2,#], Fei Qiu[3,#], Jingjing Li[4], Shanshan Li[1], Huifang Zhang[1], Xiuli Lin[1], Yukuan Huang[3], Yijun Zhou[3], Jianzhong Su[3,5], Mingqin Lu[1,*], and Yunlong Ma[3,*]

1. Department of Infectious Diseases, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang, China.
2. School of Life Science and Technology, Harbin Institute of Technology, Harbin 150080, China
3. Institute of Biomedical Big Data, School of Ophthalmology & Optometry and Eye Hospital, School of Biomedical Engineering, Wenzhou Medical University, Wenzhou 325027, Zhejiang, China.
4. 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.
5. Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou 325011, Zhejiang, China.

[#] These authors contributed equally to this work.

[*]**Correspondence should be addressed to**:

1. Yunlong Ma, Ph.D. 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;

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

Abstract

Background: 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.

Results: We constructed a powerful computational framework to integrate GWAS summary statistics with scRNA-seq data to uncover genetics-modulated liver cell subpopulations for PBC. Based on our multi-omics integrative analysis, 29 risk genes including *ORMDL3*, *GSNK2B*, and *DDAH2* were significantly associated with PBC susceptibility. 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%). The *ORMDL3*$^{+}$ cholangiocytes have prominently higher metabolism activity score than *ORMDL3*$^{-}$ cholangiocytes ($P = 1.383\times10^{-15}$). Compared with *ORMDL3*$^{-}$ cholangiocytes, there were 77 significantly differentially expressed genes among *ORMDL3*$^{+}$ cholangiocytes (FDR < 0.05), and these significant genes were associated with autoimmune diseases-related functional terms or pathways. The *ORMDL3*$^{+}$ cholangiocytes exhibited relatively high communications with macrophage and monocyte. Compared with *ORMDL3*$^{-}$ cholangiocytes, the VEGF signaling pathway is specific for *ORMDL3*$^{+}$ cholangiocytes to interact with other cell populations.

Conclusions: 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*$^{+}$ cholangiocytes with higher metabolism activity play important immune-modulatory roles in the etiology of PBC.

Background

Primary biliary cholangitis (PBC), which is formally known as primary biliary cirrhosis until 2016 [1], is a rare chronic cholestatic liver disease characterized by progressive autoimmune-mediated destruction of the small intrahepatic biliary epithelial cells [2, 3]. PBC patients suffering from chronic cholestasis can eventually lead to cirrhosis and hepatic failure without effective treatments [2]. 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 [2]. Previous studies [4] 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 [4]. To date, more than 40 genetic loci with numerous risk genes have been reported [5-10], 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 ($P \leq 5\times10^{-8}$) are reported. There exist many common SNPs with small marginal effects were neglected [11, 12]. Moreover, the vast majority of reported

SNPs were mapped within non-coding genomic regions[12]. 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 (eQTL) [11], DNase I-hypersensitive sites (DHS) [13], and histone marks [14], 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 [15]. An accruing and large number of single cell RNA sequencing (scRNA-seq) studies on autoimmune diseases, including rheumatoid arthritis [16], inflammatory bowel disease [17], and systemic lupus erythematosus [18], 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.

Results

The framework for integrating single-cell transcriptomes and GWAS on PBC

In the current study, we constructed a computational framework to unveil the cell-type specific genetic influence on the etiology of PBC based on multiple omics datasets (Figure 1 and Supplemental Table S1). It included three primary sections: 1) combination of GWAS summary statistics on PBC with scRNA-seq data to recapitulate the genetics-modulated single cell landscape for PBC (Figure 1a), 2) prioritization of genetics-risk genes and pathways contributing to PBC (Figure 1b), and 3) revelation

of the cell-to-cell interactions and metabolic activities of genetics-influenced subset of liver cholangiocytes and its immune microenvironment for 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 performed a gene-level genetic association analysis and found that 563 genes were significantly associated with PBC (FDR ≤ 0.05, Supplemental Figure S1 and Table S2). For example, the top-ranked genes of *HLA-DPA1* ($P = 2.69\times10^{-24}$), *IL12A* ($P = 6.63\times10^{-22}$), and *BTNL2* ($P = 1.65\times10^{-17}$). By using a genome-wide pathway analysis, there were 41 KEGG pathways significantly enriched (FDR ≤ 0.05, Supplemental Table S3 and Figure S2a). Based on the MDS analysis, these 41 pathways were grouped into five clusters, including Th1 and Th2 cell differentiation, allograft rejection, Th17 cell differentiation, cell adhesion molecules, and cytokine-cytokine receptor interactions (Supplemental Figure S2b). The top-ranked pathways, such as Th1 and Th2 cell differentiation, allograft rejection, inflammatory bowel disease, and type I diabetes mellitus, were relevant to autoimmune diseases.

Furthermore, by performing a gene-property analysis based on mouse liver tissue with immune cells, we found that PBC-associated genes were significantly enriched in several immune-related cell types (Supplemental Figure S3 and Table S4), including $Cst3^+$ dendritic cell ($P = 5.8\times10^{-3}$), $Trbc2^+$ T cell ($P = 7.1\times10^{-3}$), $Chil3^+$ macrophage ($P = 0.017$), and $Gzma^+$ T cell ($P = 0.03$), which were in line with the results in previous studies [19, 20].

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 [21] 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). Among these significant genes, there were 52 risk genes having been documented in the GWAS Catalog database (Supplemental Table S5). Moreover, there was a high consistency of results between MAGMA and S-MultiXcan analysis (232/268 = 86.6%, Supplemental Figure S4), such as *HLA-DRB1* ($P = 4.95\times10^{-69}$), *HLA-DRB5* ($P = 1.17\times10^{-42}$), and *BTNL2* ($P = 3.26\times10^{-41}$).

To further validate these risk genes in two PBC-relevant tissues (i.e., liver and blood) using the S-PrediXcan method, we found that 76 and 115 genes were significantly associated with PBC in liver and blood, respectively (FDR ≤ 0.05, Figure 3a-b and Supplemental Tables S6-S7). In total, 29 risk genes were significantly associated with PBC across different methods including MAGMA, S-MultiXcan, and S-PrediXcan analyses (Figure 3c). Using the Pearson correlation analysis, we observed that significant genes from S-MultiXcan analysis showed remarkable correlations with that from MAGMA and S-PrediXcan analyses ($P \leq 0.05$, Figure 3d-f). Moreover, by performing three independent permutation analyses, we found that the number of observed overlapped genes between S-MultiXcan and MAGMA and S-PrediXcan were significantly higher than random events (Empirical $P < 1\times10^{-5}$, Figure 3g-i and Supplemental Figure S5a-c). Overall, we identified 29 genes contribute susceptibility to PBC (Table 1).

Based on three independent expression profiles of liver, blood, and peripheral CD4+T cells, 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 in PBC patients compared with controls (Table 1 and Supplemental Figures S6-S8). The co-expression

patterns among these 29 genes in the peripheral CD4+T cells were prominently altered according to the PBC status (Supplemental Figure S8a). These results further support these identified risk genes have crucial effects on PBC.

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 newly identified (Table 1 and Supplemental Figure S9a). Among these novel identified genes, several genes (e.g., *LY6G5B* and *DDAH2*) were associated with autoimmune-related diseases, including rheumatoid arthritis [22] and type 1 diabetes [23]. By conducting a PPI network enrichment analysis, we observed that these risk genes were significantly interacted with each other in a subnetwork (Figure 3). For example, *TCF19* showed shared protein domains with *IRF5* [24], and *DDAH2* was highly co-expressed with *CSNK2B* [25].

Furthermore, we performed a phenotype-based enrichment analysis, and found that these 29 risk genes were remarkably enriched in several phenotypes relevant to autoimmune diseases (FDR ≤ 0.05, Supplemental Figure S9b and Table S8), such as type I diabetes mellitus, immune system diseases, and juvenile rheumatoid arthritis. We also performed a GO-term enrichment analysis, and found that several GO-terms were notably overrepresented (Supplemental Figures S10-S12), such as interferon-gamma-mediated signaling pathway ($P = 2.8\times10^{-4}$) and neutrophil activation involved in immune response ($P =7.3\times10^{-4}$).

Based on the drug-gene interaction analysis, we identified that 20 of 29 genes (68.96%) were enriched in ten potential “druggable” gene categories (Supplemental Figure S13a and Table S9). Five

genes including *SOCS1, TCF19, CARM1, SH2B3,* and *NAAA* were directly targeted at least one known drug (Supplemental Figure S13b). The gene of *SOCS1* was found to be targeted by insulin and aldesleukin, of which both have been applied to treat autoimmune diseases, including systemic lupus erythematosus [26], type 1 diabetes mellitus [27], and HIV [28]. *SH2B3* was targeted by ruxolitinib and candesartan. Previous studies [29] 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

Using the uniform manifold approximation and projection (UMAP), there were 22 discrete clustered for the liver scRNA-seq dataset (Figure 5a). Using well-known marker genes, these clusters were assigned into 13 distinct cell subpopulations, including portal endothelial cells, cholangiocytes, non-inflammatory macrophages, T cells, γδT 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 (Figure 5b). By calculating the genetic risk score of 29 PBC-associated genes using the *AddModuleScore* function in Seurat, we found that these risk genes were primarily enriched in non-inflammatory macrophage and inflammatory monocyte/macrophage (Figure 5c), suggesting that these risk genes may have crucial functions in innate immunity for PBC. 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 5d and Supplemental Table S10). These results are consistent with previous evidence that an immune-mediated injury of cholangiocytes contributes risk to PBC [30, 31].

Using the specificity algorithm in the MAGMA method [32], we noticed that the risk gene of *ORMDL3* exhibited the highest expression in cholangiocytes than other cell types (Figure 5e 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 [7, 9, 33]. The top-ranked risk SNP associated with *ORMDL3* is rs9303277 ($P = 2.57\times10^{-11}$). This SNP showed significant cell-specific eQTL of *ORMDL3* among naïve B cell ($P = 5.8\times10^{-10}$), $TFH^{+}CD4^{+}$T cell ($P = 3.8\times10^{-9}$), $CD56^{dim}$ $CD16^{+}$ NK cells ($P = 4.4\times10^{-9}$), $TH1^{+}$ $CD4^{+}$ T cells ($P = 6.0\times10^{-6}$), and memory $TREG^{+}$ $CD4^{+}$ T cells ($P = 1.1\times10^{-5}$), and exhibited notable cell-specific promoter-interacting eQTL of *ORMDL3* among naïve B cell ($P = 4.2\times10^{-13}$) and $CD56^{dim}$ $CD16^{+}$ NK cells ($P =5.0\times10^{-12}$) (Supplemental Figure S14). 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 [7, 9, 34]. 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 functions of $ORMDL3^{+}$ cholangiocytes and its cellular communications with immune cells.

Characterization of biological functions of $ORMDL3^{+}$ cholangiocytes

The subset of *ORMDL3*$^{+}$ cholangiocytes account for 22.38% (635/2837) of cholangiocytes (Figure 6a). By calculating the metabolism activity score among all annotated cells, we found that the cholangiocytes showed remarkably higher metabolism activity score than other cells (Figure 6b). Interestingly, we noticed that the metabolism activity score of *ORMDL3*$^{+}$ cholangiocytes was significantly higher than that in *ORMDL3*$^{-}$ cholangiocytes ($P = 1.383\times10^{-15}$, Figure 6c). Furthermore, we compared the expression profiles of *ORMDL3*$^{+}$ cholangiocytes with *ORMDL3*$^{-}$ cholangiocytes. There were 77 significantly DEGs with six up-regulated DEGs and 71 down-regulated DEGs among *ORMDL3*$^{+}$ cholangiocytes (FDR < 0.05, Figure 6d and Supplemental Tables S12-S13). With regard to six up-regulated DEGs, there were seven significant pathways overrepresented (FDR < 0.05, Supplemental Figure S15a and Table S14).

Among the 71 down-regulated DEGs, *HLA-DRA* has been shown to be associated with PBC [7], 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 [35, 36]. *IFI16* is reported to be an innate immune sensor for intracellular DNA [37]. Pathway enrichment analysis revealed that there were 55 significant KEGG pathways enriched by these 71 down-regulated DEGs (FDR < 0.05, Figure 6e and Supplemental Table S15), of which several such as inflammatory bowel disease, rheumatoid arthritis, Th17 cell differentiation, cytokine-cytokine receptor interaction, and chemokine signaling pathway have been demonstrated to implicate in inflammatory-related diseases [11, 38, 39].

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, Supplemental Figure S15b-d and Tables S16-S18). These functional terms were largely linked with immune and metabolism functions, such as granulocyte activation, neutrophil mediated immunity, and S100 protein binding. These results suggest that *ORMDL3*$^+$ cholangiocytes have immune-modulatory effects on PBC risk.

Cellular communications of ORMDL3*$^+$ *cholangiocytes with other cells

To gain refined insights into *ORMDL3*$^+$ cholangiocytes, we performed a cell-to-cell interaction analysis among cell populations in liver tissue using the CellChat algorithm. By calculating the aggregated cell-cell communication network by counting the number of links, we found that *ORMDL3*$^+$ cholangiocytes showed the highest outgoing (sources) interaction strength than other cell types (Figure 6e and Supplemental Figure S16). By further summarizing the communication probability among cellular interactions, we observed a high connective network of *ORMDL3*$^+$ cholangiocytes with other cells (Figure 6f and Supplemental Figures S17-S18). The *ORMDL3*$^+$ cholangiocytes showed relatively high communications with non-inflammatory macrophage and inflammatory monocyte/macrophages (Figure 6f), recalling that these two types of immune cells have been remarkably enriched by PBC-risk genes in our above analysis. There were 35 significant ligand-receptor interactions (including source and target) of *ORMDL3*$^+$cholangiocytes were predicted, including MIF-(CD74+CXCR4), MIF-(CD74+CD44), VEGFA-VEGFR2, and NAMPT-INSR (Figure 7a-b). There existed several unique ligand-receptor pairs in *ORMDL3*$^+$ cholangiocytes compared with *ORMDL3*$^-$ cholangiocytes, including VEGFA-VEGFR2, VEGFA-VEGFR1R2 and VEGFA-VEGFR1 (Supplemental Figure S19).

By performing a pattern recognition analysis based on the non-negative matrix factorization to

uncover the global communication patterns among all cell populations, we observed that five communication patterns that connect cell populations with signaling pathways in the context of both outgoing signaling (i.e., treating cells as source, Figure 8a) and incoming signaling (i.e., treating cells as target, Supplemental Figure S20a), respectively. We found that a large portion of outgoing signaling of *ORMDL3*$^{+}$ cholangiocytes is characterized by pattern #1, which represents multiple signaling pathways, including MIF, PARs, VISFATIN, COMPLEMENT, ANGPTL, PROS, CHEMERIN, AGT, VEGF, and IGF (Figure 8b). Compared with *ORMDL3*$^{-}$ cholangiocytes, the VEGF signaling pathway is specific for *ORMDL3*$^{+}$ cholangiocytes as source cells to interact with other cells (Figure 8b). On the other hand, the communication patterns of both *ORMDL3*$^{+}$and *ORMDL3*$^{-}$ cholangiocytes as target cells showed similar patterns (Supplemental Figure S20b). Network centrality analysis confirmed that VEGF signaling pathway is not only a significant sender (i.e., source cells) but also a prominent influencer controlling the communications for *ORMDL3*$^{+}$ cholangiocytes (Figure 8c). Notably, among all known ligand-receptor pairs, VEGF signaling pathway is dominated by VEGFA ligand and its VEGFR1 receptor (Supplemental Figure S21). Using the expression levels of all genes in the VEGF signaling pathway to calculate a VEGF activity score, we found that *ORMDL3*$^{+}$ cholangiocytes exhibited a significantly higher activity score than that among *ORMDL3*$^{-}$ cholangiocytes ($P = 1.28\times10^{-15}$, Figure 8d). These results suggest that VEGF signaling pathway play an important role in cellular communications of *ORMDL3*$^{+}$ cholangiocytes.

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* [5, 6, 8, 9], *SOCS1* [5], *SYNGR1* [5, 7], *ORMDL3* [7, 9, 33], *MANBA* [5], *IDUA* [5], *DGKQ* [5], *FCRL3* [40], *NAAA* [41], and *SH2B3* [5], 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*. Some of these novel risk genes such as *DDAH2* and *HLA-DMA* have been reported to be associated with the autoimmune diseases, including type I diabetes [42], rheumatoid arthritis [43], and systemic lupus erythematosus [44]. The previously-reported gene of *ORMDL3* 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 [45], inflammatory bowel diseases [38], rheumatoid arthritis [39], and Crohn's disease [46]. Functional enrichment analyses uncovered that these genetics-risk genes were notably enriched in several biological processes or disease-terms, which are relevant to autoimmune phenotypes [47, 48]. Together, these identified genes are more likely to be genuine genes implicated in PBC risk.

Our gene-property analysis unveiled that PBC-relevant genetic association signals were significantly enriched in several immune-related cell populations, including dendritic cells, T cells, and macrophage. Consistently, based on the genetic risk score, we revealed that these 29 PBC-associated genes were prominently enriched in non-inflammatory macrophage and inflammatory monocyte/macrophage. Dendritic cells have been shown to be relevant to the pathogenesis of PBC [19, 49]. Earlier studies have demonstrated that an intense biliary inflammatory CD8+ and CD4+ T cell response has been used for characterizing PBC [50]. Moreover, dendritic cells are crucial 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 [51]. Macrophages and monocytes are

key components of the innate immune system, and have tissue-repairing roles in reducing immune responses and enhancing tissue regeneration [52]. Multiple lines of evidence [53] have demonstrated that the infiltration of macrophages and monocytes in diseased tissues is considered to be a hallmark of several autoimmune diseases, including PBC [20, 54]. Recently, Dorris et al. [55] reported that the PBC susceptibility gene *C5orf30* modulates macrophage-mediated immune regulations. Overall, these results suggest that genetics-risk genes have critical immuno-modulatory roles in the development of PBC.

To further examine the effects of liver cell populations and its immune microenvironments on PBC, we performed a regression-based polygenic analysis based on human liver tissues with immune cells. Cholangiocytes were significantly enriched by PBC-related genetic association signals. Previously, the non-parenchymal cholangiocytes have been reported to be injury in numerous human diseases termed as cholangiopathies [56], including PBC [56, 57]. Recently, Banales and coworkers [56] have demonstrated that cholangiocytes play pivotal roles in innate and adaptive immune responses relevant to immune-mediated cholangiopathies. Erice et al. [58] reported that microRNA-506 induces PBC-like features in human cholangiocytes and promotes the activating processes of immune responses. Moreover, we found that there existed higher metabolism activity score of cholangiocytes than other cells among liver tissues, indicating that abnormal metabolic pathways may involve in the etiology of PBC, which is in line with previous results [59]. Interestingly, the cell subset of *ORMDL3*$^{+}$ cholangiocytes have strikingly higher metabolic activity than *ORMDL3* negative cells.

Compared with *ORMDL3*$^{-}$ cholangiocytes, we found 77 significant DEGs among *ORMDL3*$^{+}$ cholangiocytes, which contain numerous cytokine and chemokine genes, such as *CXCL8*, *CCL3*, and *CXCL1*, that may involve in mediating the immune-regulation for PBC risk. Furthermore, based on

cellular communications analysis, we identified *ORMDL3*$^{+}$ cholangiocytes exhibited high interactions with two innate immune cell types of non-inflammatory macrophage and inflammatory monocyte/macrophages. Several vital signaling pathways, including MIF, PARs, VEGF, and IGF, were predicted to implicate in the cellular interactions of *ORMDL3*$^{+}$ cholangiocytes with other cells. The VEGF signaling pathway showed a higher specificity for *ORMDL3*$^{+}$ cholangiocytes than *ORMDL3*$^{-}$ cholangiocytes. VEGF is a potent stimulating factor for angiogenesis and vascular permeability, which have been reported to involve in PBC [60]. Multiple lines of evidence have demonstrated that VEGF has an important role in pathological conditions that are associated to autoimmune diseases, such as systemic lupus erythematosus, rheumatoid arthritis, inflammatory bowel disease, and multiple sclerosis [61].

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. 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.

Conclusions

In summary, 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 and its immune microenvironments that contribute risk to PBC, and found that *ORMDL3*$^{+}$ cholangiocytes potentially play important immune-regulatory roles in the pathogenesis of PBC. Current study gives several highlighted genetics-risk genes and disease-relevant cell types for unveiling the genetic mechanisms 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).

2. Bulk-based expression profiles of PBC: 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/) [5]. 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 GWAS signals. A standard quality control (QC) pipeline was applied to remove low-quality SNPs. The software package of MaCH [62] 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 the current analyses.

Combination of GWAS summary statistics with scRNA-seq data for PBC

We leveraged a widely-used method of RolyRoly [63], 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 $\boldsymbol{\beta}_{S_j}$ represents a PBC-GWAS-derived effect-size vector of S_j with a *priori* assumption that $\boldsymbol{\beta}_{S_j} \sim \text{MVN}(\mathbf{0}, \sigma_j^2 \boldsymbol{I}_{|S_j|})$. Under the assumption, RolyPoly offers a polygenic linear model for $\boldsymbol{\beta}_{S_j}$:

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

where γ_0 represents an intercept term, $\alpha_{ji} (i = 1,2,...,N)$ represents a group of annotations, for example, cell-type-specific gene expressions, and γ_i is the annotation's coefficient for α_{ji}. To fit the observed and expected sum squared SNP effect sizes relevant to each gene, RolyPoly applies the method-of-moments estimators to estimate γ_i by the following formula:

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

where $\boldsymbol{R}_{S_j}$ represents the linkage disequilibrium (LD) matrix of S_j. The 1,000 Genome Project European Phase 3 panel [64] was used to calculate the LD among SNPs. 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 ≤ 0.05 was interpreted to be of significance.

Integrative genomic analysis of combining GWAS summary statistics with eQTL data

To highlight the functional risk genes whose expressions were significantly associated with PBC, we leveraged the S-PrediXcan tool [65] to combine GWAS summary data on PBC with eQTL data for

liver and blood tissues from the GTEx Project (version 8) [66]. A MASHR-based prediction method with two linear regression models was used to estimate gene expression weights:

$$\boldsymbol{Y} = \boldsymbol{\alpha}_1 + \boldsymbol{X}_l \beta_l + \boldsymbol{\varepsilon}_1$$

$$\boldsymbol{Y} = \boldsymbol{\alpha}_2 + \boldsymbol{G}_g \gamma_g + \boldsymbol{\varepsilon}_2$$

where $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are independent error terms, $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are intercepts, $\boldsymbol{Y}$ is the n dimensional vector for n samples, $\boldsymbol{X}_l$ is the allelic dosage for SNP l in n samples, β_l is the effect size of SNP l, $\boldsymbol{G}_g = \sum_{i \in gene(g)} \omega_{ig} \boldsymbol{X}_i$ is the predicted expression calculated by ω_{lg} and $\boldsymbol{X}_l$, in which ω_{lg} is derived from the GTEx Project , and γ_g is the effect size of $\boldsymbol{G}_g$. The Wald-statistic Z score for each association is transformed as:

$$Z_g = \frac{\hat{\gamma}_g}{\text{se}(\hat{\gamma}_g)} \approx \sum_{i \in gene(g)} \omega_{ig} \frac{\hat{\sigma}_i}{\hat{\sigma}_g} \frac{\hat{\beta}_i}{\text{se}(\hat{\beta}_i)}$$

where $\hat{\sigma}_g$ is the standard deviation of $\boldsymbol{G}_g$ and can be calculated from the 1,000 Genomes Project European Phase 3 Panel [64], $\hat{\beta}_l$ is the effect size from GWAS on PBC and $\hat{\sigma}_l$ is the standard deviation of $\hat{\beta}_l$.

To enhance the statistical power to identify significant genes whose expression has similar functions across different tissues, we leveraged the S-MultiXcan tool [21] to incorporate convergent evidence across 49 different GTEx tissues. The S-MultiXcan fits a multivariate regression model from multiple tissue models jointly:

$$\boldsymbol{Y} = \sum_{j=1}^{p} \boldsymbol{T}_j g_j + \boldsymbol{e} = \boldsymbol{T}\boldsymbol{g} + \boldsymbol{e}$$

where $\tilde{\boldsymbol{T}}_j = \sum_{i \in gene(j)} \omega_i \boldsymbol{X}_i$ is the predicted expression of tissue j, and $\boldsymbol{T}_j$ is the standardization of $\tilde{\boldsymbol{T}}_j$ to $mean = 0$ and $standard\ deviation = 1$. g_j is the effect size for the

predicted gene expression in tissue j, $\boldsymbol{e}$ is an error term with variance σ_e^2, and p is the number of chosen tissues. There were 22,279 genes used in the multivariate regression model. $P < 6.89\times10^{-4}$ (FDR < 0.05) was considered to be significant.

Gene-based genetic association analysis

We conducted a gene-level association analysis of the GWAS summary statistics on PBC by leveraging a multi-variant converging regression model in the Multi-marker Analysis of GenoMic Annotation (MAGMA) [32, 67]. The analyzed SNP set among each gene was depended on whether the SNP mapped into the body of the gene or its extended regions (+/- 20 kb downstream or upstream). The 1,000 Genomes Project Phases 3 European Panel [64] was used to compute the LD between SNPs. The P value threshold was set to 0.0016, and the method of Benjamini-Hochberg false discovery rate (FDR) was applied for multiple testing correction [68]. Furthermore, to investigate the biological processes involved in PBC, we performed genome-wide pathway enrichment analysis using the *build-in* function of the MAGMA gene-set method [32]. The competitive P value of each pathway was computed by leveraging the results that the incorporated effect of genes in a given pathway are prominently larger than the incorporated effect of all rest genes. The *build-in* function of 10,000 permutations in MAGMA was used for adjusting competitive P values. We utilized the widely-used pathway resource of KEGG with latest version [69] for the MAGMA-based pathway enrichment analysis.

Multidimensional scaling analysis

To obtain the similarity of enriched pathways identified from MAGMA-based pathway

enrichment analysis, we carried out a multidimensional scaling (MDS) analysis to group these biological pathways. First, we arranged a *pathway.txt* file included all the significant enriched pathways, and then utilized the Jaccard distance algorithm [70] to calculate the pathway-pathway distance scores according to overlapped genes. Using the distance scores among pathways, we obtained the first two components of results from the MDS analysis (i.e., MDS1 and MDS2). Subsequently, we plotted the clusters of these identified pathways via MDS1 and MDS2 using the *symbols* function in *R* platform [71]. The most significant pathway (i.e., pathway has the lowest P value in each cluster) was used to mark each cluster.

In silico permutation analysis

We referenced the method used in previous studies [11, 67, 68, 72] 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 (Geneset #1, N = 308, FDR ≤ 0.05) with other results from three distinct analyses: MAGMA analysis (Geneset #2: N = 563, FDR ≤ 0.05), S-PrediXcan on liver (Geneset #3: N = 76, FDR ≤ 0.05), and S-PrediXcan on blood (Geneset #4: N = 115, FDR ≤ 0.05). First, we separately calculated the number of overlapped genes between Geneset #1 and Genesets #2, #3 and #4 ($N_{Observation}$). Second, we used the total number of genes in MAGMA, S-PrediXcan on liver, and S-PrediXcan on blood as the background genes ($N_{Background}$ = 18,068, 12,033, and 12,070, respectively). By randomly selecting the same number of genes as Genesets #2, #3, and #4 from the background genes to overlap with Geneset #1, we computed the number of overlapped genes in each time (N_{Random}). The empirical P value was calculated by using the formula: $P = \frac{N_{Random} \geq N_{Observation}}{N_{Total}}$. The empirically permuted P value ≤ 0.05 is interpreted as significance. By using the same method as

above, we also performed an *in silico* permutation analysis for results from S-MultiXcan, MAGMA, and S-PrediXcan (on liver and blood) at a significant level of P value ≤ 0.05 to further explore the consistency.

Functional enrichment analysis of disease- and GO-terms

We conducted a disease-based enrichment analysis for these identified risk genes associated with PBC by using the WEB-based Gene SeT AnaLysis Toolkit (WebGestalt: http://www.webgestalt.org) [73] based on the GLAD4U database [74]. Moreover, we used the WebGestalt tool to perform a functional enrichment analysis for these common PBC-genetic risk genes based on the Gene Ontology (GO) database [75]. There were three categories of GO-terms: molecular function (MF), cellular component (CC), and biological process (BP). The number of genes in each GO-term is between 5 and 2,000. The over-representation algorithm was adopted for evaluating the significant level for these enrichment analyses, and the Benjamini-Hochberg FDR method was used for multiple correction. In addition, the WebGestalt tool was used to conduct functional enrichment analyses for these significantly differentially expression genes (DEGs) between *ORMDL3*$^{+}$ and *ORMDL3*$^{-}$ cholangiocytes.

Protein-Protein interaction network analysis

To explore the functional interactions of these identified PBC-associated genes, we conducted a protein-protein interaction (PPI) network-based analysis by leveraging the GeneMANIA tool (http://www.genemania.org/), which is a widely-used *plug-in* of Cytoscape platform [76-78]. By using PBC-associated genes as an input list, the GeneMANIA would predict genes with similar biological

functions and construct interacted links by incorporating current existing genomics and proteomics information, containing co-expression associations and shared protein domains.

MAGMA gene-property analysis

To identify PBC-associated single cell subpopulations enriched by MAGMA-identified genes, we conducted MAGMA gene property analysis using the web-access tool of FUMA (https://fuma.ctglab.nl/) [79], which is a highly efficient and easy-to-use software, to integrate gene-level association signals from GWAS summary data on PBC [5] with scRNA-seq data based on liver tissue from the Mouse Cell Atlas [80]. There were 20 distinct cell types of liver tissue in the Mouse Cell Atlas. The Bonferroni method [81] was used for multiple testing correction.

Immune cell-specific expression quantitative trait loci

To find the eQTL for PBC-risk SNPs, we first extracted genetic variants statistics around upstream and downstream 400 kb of the targeted genes from GWAS summary data. We used the *LocusZoom* (http://locuszoom.org/) to figure a regional association plot for each gene. Then, we explored the cis-eQTL and promotor-interacting eQTL (pieQTL) among the Database of Immune Cell Expression quantitative trait loci and Epigenomics (DICE, https://dice-database.org/landing) [82] for annotating functional genetic variants.

Defining cell activity scores

We applied the cell activity score (CAS) to evaluate the immunological degree and metabolic activity of each cell type by using a pre-defined expression gene set [83]. The CASs were obtained by

using a widely-used equation: $CAS_j(n)$ = average(RLE(PGS_j, n)) - average(RLE(CGS_m, n)), where PGS_j is a pre-curated gene set j in a given cell n, and CGS_m is a control gene set that was randomly selected from aggregate expression levels bins, which gain a comparable distribution of expression levels and over size to that of the pre-defined gene set. RLE represents the relative expression of PGS_j or CGS_m. The *AddModuleScore* function in the Seurat tool [84] was leveraged to calculate the CAS with default parameters. We used pre-collected 85 metabolism-related pathways (N = 1,667 genes) in KEGG database [69], such as primary bile acid biosynthesis (ID: 00120), pyruvate metabolism (ID: 00620), and oxidative phosphorylation (ID: 00640), to define inflammatory score and metabolic activity score, respectively.

Cellular interactions among different cell types

To unveil potential cell-to-cell communications of *ORMDL3*$^+$ and *ORMDL3*$^-$ cholangiocytes with other liver and immune cells, we leveraged the CellChat R package [85] to infer the cellular interactions based on the normalized scRNA-seq dataset (i.e., GSE115469). The algorithm of CellChat could examine the ligand-receptor interactions significance among different types of cells based on the expression of soluble agonist, soluble antagonist, and stimulatory and inhibitory membrane-bound co-receptors. By summing the probabilities of the ligand-receptor interactions among a given signaling pathway, we could calculate the communication probability for the pathway.

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 ≤ 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) [86] 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 [70] was used to assess the similarities among pathways.

List of abbreviations

PBC = Primary biliary cholangitis, Genome-wide association study = GWAS, SNP = Single nucleotide polymorphism, eQTL = expression quantitative trait loci, DHS = DNase I-hypersensitive site, scRNA-seq = single cell RNA sequencing, GEO = the Gene Expression Omnibus, QC = Quality control, LD = linkage disequilibrium, MHC = Major histocompatibility complex, MAGMA = the Multi-marker Analysis of GenoMic Annotation, FDR = False discovery rate, MDS = multidimensional scaling, WebGestalt = the WEB-based Gene SeT AnaLysis Toolkit, GO = Gene ontology, MF = Molecular function, BP = Biological process, CC = Cellular component, DEG = Differential expression genes, PPI = Protein-protein interaction, pieQTL = promoter-interacting expression quantitative loci, DICE = the Database of Immune Cell Expression quantitative trait loci and Epigenomics, CAS = Cell activity score, PGS = Pre-curated gene set, CGS = Control gene set, RLE = Relative expression.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors are agreed to publish this paper

Competing interests

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.).

Author Contributions

Y.M. and M.L. conceived and designed the study. Y.M., B.X., C.D., F.Q. S.L., Y.Z., and J.L. contributed to management of data collection. Y.M., B.X., C.D., F.Q., J.L., H.Z., X.L., and Y.H., conducted bioinformatics analysis and data interpretation. Y.M., J.S., and M.L. wrote the manuscripts. All authors reviewed and approved the final manuscript.

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. We also thank to all authors who have deposited single cell RNA sequencing data and bulk-based RNA expression data on primary biliary cirrhosis in the NCBI GEO database.

Availability of data and materials

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#.Xv6Z6igzbgl). The single cell RNA sequencing data and bulk-based RNA expression data were downloaded from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/gds/). All scripts used in the Methods are deposited in the public available GitHub repository: https://github.com/mayunlong89/PBC_project.

References

1. Beuers U, Gershwin ME, Gish RG, Invernizzi P, Jones DE, Lindor K, Ma X, Mackay IR, Parés A, Tanaka A *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, Leung PS, Bowlus CL, Yang GX, Yang CY, Coppel RL, Ansari AA, Cuebas DA *et al*: **Epithelial cell specificity and apotope 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, Li Y, Hirschfield GM, Greene CS, Xie G, Juran BD, Zhu D, Qian DC *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, Lu Y, Xie G, Lu Y, Gu X, Walker EJ, Jing K, Juran BD *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, Shi X, Wei Y, Zheng X, Dai Y, Gong Y, Wang L, Xu P *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, Kosoy R, Lu Y, Bianchi I, Podda M, Xu C, Xie G, Macciardi F *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, Cordell HJ, Franklin CS, Shin SY, Heneghan MA, Neuberger JM, Donaldson PT, Day DB *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, Atkinson EJ, Li Y, Xie G, Kosoy R, Ransom M, Sun Y, Bianchi I *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, Yao Y, Zhang Y, Qu J, Wu N, Su J: **Integrative Genomics Analysis Reveals a 21q22.11 Locus**

Contributing Risk to COVID-19. *Hum Mol Genet* 2021.

12. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM *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, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K *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, Zhang Y, Jiang Q, Zhou M, Su J: **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, Fonseka CY, Rao DA, Kelly S, Goodman SM, Tabechian D, Hughes LB, Salomon-Escoto 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, Chen G, Bolisetty M, Baisch J, Walters L, Punaro M, Rossi RJ, Chung CH *et al*: **Mapping systemic lupus erythematosus heterogeneity at the single-cell level**. *Nat Immunol* 2020, **21**(9):1094-1106.
19. 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.
20. Harada K, Shimoda S, Ikeda H, Chiba M, Hsu M, Sato Y, Kobayashi M, Ren XS, Ohta H, Kasashima S *et al*: **Significance of periductal Langerhans cells and biliary epithelial cell-derived macrophage inflammatory protein-3α in the pathogenesis of primary biliary cirrhosis**. *Liver Int* 2011, **31**(2):245-253.
21. 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.
22. Hu HJ, Jin EH, Yim SH, Yang SY, Jung SH, Shin SH, Kim WU, Shim SC, Kim TG, Chung YJ: **Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis**. *Exp Mol Med* 2011, **43**(11):613-621.
23. Tomer Y, Dolan LM, Kahaly G, Divers J, D'Agostino RB, Jr., Imperatore G, Dabelea D, Marcovina S, Black MH, Pihoker C *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.
24. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI *et al*: **InterPro in 2019: improving coverage, classification and access to protein sequence annotations**. *Nucleic Acids Res* 2019, **47**(D1):D351-d360.
25. Wang Q, Diskin S, Rappaport E, Attiyeh E, Mosse Y, Shue D, Seiser E, Jagannathan J, Shusterman S, Bansal M *et al*: **Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number**. *Cancer Res* 2006, **66**(12):6050-6062.
26. von Spee-Mayer C, Siegert E, Abdirama D, Rose A, Klaus A, Alexander T, Enghard P, Sawitzki B, Hiepe F, Radbruch A *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.
27. Atkinson MA, Eisenbarth GS, Michels AW: **Type 1 diabetes**. *Lancet* 2014, **383**(9911):69-82.
28. Sundin DJ, Wolin MJ: **Aldesleukin therapy in HIV-infected patients**. *Am J Health Syst Pharm* 1998, **55**(14):1520-1523.
29. Heine A, Held SA, Daecke SN, Wallner S, Yajnanarayana SP, Kurts C, Wolf D, Brossart P: **The JAK-inhibitor**

ruxolitinib impairs dendritic cell function in vitro and in vivo. *Blood* 2013, **122**(7):1192-1202.

30. Zhou T, Kyritsi K, Wu N, Francis H, Yang Z, Chen L, O'Brien A, Kennedy L, Ceci L, Meadows V *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.
31. 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.
32. de Leeuw CA, Mooij JM, Heskes T, Posthuma D: **MAGMA: generalized gene-set analysis of GWAS data**. *PLoS Comput Biol* 2015, **11**(4):e1004219.
33. Nakamura M, Nishida N, Kawashima M, Aiba Y, Tanaka A, Yasunami M, Nakamura H, Komori A, Nakamuta M, Zeniya 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.
34. Das S, Miller M, Broide DH: **Chromosome 17q21 Genes ORMDL3 and GSDMB in Asthma and Immune Diseases**. *Adv Immunol* 2017, **135**:1-52.
35. Rydyznski Moderbacher C, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, Belanger S, Abbott RK, Kim C, Choi J *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.
36. Su Y, Chen D, Yuan D, Lausted C, Choi J, Dai CL, Voillet V, Duvvuri VR, Scherler K, Troisch P *et al*: **Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19**. *Cell* 2020, **183**(6):1479-1495.e1420.
37. Unterholzner L, Keating SE, Baran M, Horan KA, Jensen SB, Sharma S, Sirois CM, Jin T, Latz E, Xiao TS *et al*: **IFI16 is an innate immune sensor for intracellular DNA**. *Nat Immunol* 2010, **11**(11):997-1004.
38. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T *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.
39. Kurreeman FA, Stahl EA, Okada Y, Liao K, Diogo D, Raychaudhuri S, Freudenberg J, Kochi Y, Patsopoulos NA, Gupta N *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.
40. Tanaka A, Ohira H, Kikuchi K, Nezu S, Shibuya A, Bianchi I, Podda M, Invernizzi P, Takikawa H: **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.
41. 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.
42. Fogarty RD, Abhary S, Javadiyan S, Kasmeridis N, Petrovsky N, Whiting MJ, Craig JE, Burdon KP: **Relationship between DDAH gene variants and serum ADMA level in individuals with type 1 diabetes**. *J Diabetes Complications* 2012, **26**(3):195-198.
43. 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.
44. 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.
45. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E *et al*: **Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma**. *Nature* 2007, **448**(7152):470-473.
46. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM *et al*: **Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease**. *Nat Genet*

2008, **40**(8):955-962.

47. 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.
48. Mo X, Guo Y, Qian Q, Fu M, Lei S, Zhang Y, Zhang H: **Mendelian randomization analysis revealed potential causal factors for systemic lupus erythematosus**. *Immunology* 2020, **159**(3):279-288.
49. Kita H, Lian ZX, Van de Water J, He XS, Matsumura S, Kaplan M, Luketic V, Coppel RL, Ansari AA, Gershwin ME: **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.
50. Coppel RL, Gershwin ME: **Primary biliary cirrhosis: the molecule and the mimic**. *Immunol Rev* 1995, **144**:17-49.
51. Swiecki M, Colonna M: **The multifaceted biology of plasmacytoid dendritic cells**. *Nat Rev Immunol* 2015, **15**(8):471-485.
52. 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.
53. 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.
54. Peng A, Ke P, Zhao R, Lu X, Zhang C, Huang X, Tian G, Huang J, Wang J, Invernizzi P *et al*: **Elevated circulating CD14(low)CD16(+) monocyte subset in primary biliary cirrhosis correlates with liver injury and promotes Th1 polarization**. *Clin Exp Med* 2016, **16**(4):511-521.
55. Dorris ER, Tazzyman SJ, Moylett J, Ramamoorthi N, Hackney J, Townsend M, Muthana M, Lewis MJ, Pitzalis C, Wilson AG: **The Autoimmune Susceptibility Gene C5orf30 Regulates Macrophage-Mediated Resolution of Inflammation**. *J Immunol* 2019, **202**(4):1069-1078.
56. Banales JM, Huebert RC, Karlsen T, Strazzabosco M, LaRusso NF, Gores GJ: **Cholangiocyte pathobiology**. *Nat Rev Gastroenterol Hepatol* 2019, **16**(5):269-281.
57. 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.
58. Erice O, Munoz-Garrido P, Vaquero J, Perugorria MJ, Fernandez-Barrena MG, Saez E, Santos-Laso A, Arbelaiz A, Jimenez-Agüero R, Fernandez-Irigoyen J *et al*: **MicroRNA-506 promotes primary biliary cholangitis-like features in cholangiocytes and immune activation**. *Hepatology* 2018, **67**(4):1420-1440.
59. Bell LN, Wulff J, Comerford M, Vuppalanchi R, Chalasani N: **Serum metabolic signatures of primary biliary cirrhosis and primary sclerosing cholangitis**. *Liver Int* 2015, **35**(1):263-274.
60. Medina J, Sanz-Cameno P, García-Buey L, Martín-Vílchez S, López-Cabrera M, Moreno-Otero R: **Evidence of angiogenesis in primary biliary cirrhosis: an immunohistochemical descriptive study**. *J Hepatol* 2005, **42**(1):124-131.
61. Carvalho JF, Blank M, Shoenfeld Y: **Vascular endothelial growth factor (VEGF) in autoimmune diseases**. *J Clin Immunol* 2007, **27**(3):246-256.
62. 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.
63. Calderon D, Bhaskar A, Knowles DA, Golan D, Raj T, Fu AQ, Pritchard JK: **Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression**. *Am J Hum Genet* 2017, **101**(5):686-699.
64. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR: **A global reference for human genetic variation**. *Nature* 2015, **526**(7571):68-74.
65. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards

TL *et al*: **Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics**. *Nat Commun* 2018, **9**(1):1825.

66. **The Genotype-Tissue Expression (GTEx) project**. *Nat Genet* 2013, **45**(6):580-585.
67. 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.
68. 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.
69. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic acids research* 2000, **28**(1):27-30.
70. Ma Y, Li J, Xu Y, Wang Y, Yao Y, Liu Q, Wang M, Zhao X, Fan R, Chen J *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.
71. Zhong Y, Chen L, Li J, Yao Y, Liu Q, Niu K, Ma Y, Xu Y: **Integration of summary data from GWAS and eQTL studies identified novel risk genes for coronary artery disease**. *Medicine (Baltimore)* 2021, **100**(11):e24769.
72. Dong Z, Ma Y, Zhou H, Shi L, Ye G, Yang L, Liu P, Zhou L: **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.
73. Wang J, Duncan D, Shi Z, Zhang B: **WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013**. *Nucleic Acids Res* 2013, **41**(Web Server issue):W77-83.
74. Jourquin J, Duncan D, Shi Z, Zhang B: **GLAD4U: deriving and prioritizing gene lists from PubMed literature**. *BMC Genomics* 2012, **13 Suppl 8**(Suppl 8):S20.
75. **Gene Ontology Consortium: going forward**. *Nucleic Acids Res* 2015, **43**(Database issue):D1049-1056.
76. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT *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.
77. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res* 2003, **13**(11):2498-2504.
78. Ma Y, Li MD: **Establishment of a Strong Link Between Smoking and Cancer Pathogenesis through DNA Methylation Analysis**. *Sci Rep* 2017, **7**(1):1811.
79. Watanabe K, Umićević 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.
80. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F *et al*: **Mapping the Mouse Cell Atlas by Microwell-Seq**. *Cell* 2018, **172**(5):1091-1107.e1017.
81. **Etymologia: Bonferroni correction**. *Emerg Infect Dis* 2015, **21**(2):289.
82. Chandra V, Bhattacharyya S, Schmiedel BJ, Madrigal A, Gonzalez-Colin C, Fotsing S, Crinklaw A, Seumois G, Mohammadi P, Kronenberg M *et al*: **Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants**. *Nat Genet* 2021, **53**(1):110-119.
83. Ren X, Wen W, Fan X, Hou W, Su B, Cai P, Li J, Liu Y, Tang F, Zhang F *et al*: **COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas**. *Cell* 2021, **184**(7):1895-1913.e1819.
84. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell transcriptomic data across different conditions, technologies, and species**. *Nat Biotechnol* 2018, **36**(5):411-420.
85. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q: **Inference and analysis of cell-cell communication using CellChat**. *Nat Commun* 2021, **12**(1):1088.

86. 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.

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. Circus 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 $\leq 1\times10^{-8}$, orange color marks significant genes with FDR is between1×10^{-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. Results from integrative genomics analyses. a)-b) Circus plot demonstrates the results of S-PrediXcan integrative genomics analysis on (a) liver tissue and (b) blood tissue. c) Venn plot showing the overlapped genes between S-MultiXcan, MAGMA, and S-PrediXcan on both liver and blood. d)-f) Pearson correlation of top-ranked genes identified from S-MultiXcan analysis with that from (d) MAGMA analysis, (e) S-PrediXcan analysis on liver, and (f) S-PrediXcan analysis on blood. g-i) *In silico* permutation analysis of 100,000 random selections for the overlapped genes between S-MultiXcan and (g) MAGMA, (h) S-PrediXcan on liver, and (i) S-PrediXcan on blood.

Figure 4. PPI network analysis of 29 PBC-risk gene. Note: This network analysis was performed using the GeneMANIA tool. Orange node represents risk genes identified to be associated with PBC in our current analysis. Gray node represents predicted genes that connected with PBC-risk genes. Purple edge represents the co-expression interactions (account for 69.52%). Light orange edge represents the shared protein domain interactions (account for 30.48%).

Figure 5. Genetics-influenced cell populations for PBC. a) UMAP dimensionality reduction embedding of liver and immune cells among all five samples from primary liver patients (n = 8,444 cells) colored by each cluster. b) UMAP embedding of liver and immune cells colored by orthogonally generated clusters labeled by manual cell type annotation (13 cell types). c) UMAP embedding of all cells among 13 cell populations colored by the genetic risk score of 29 PBC-risk genes. d) Bar graph showing genetics-influenced liver cell subpopulations for PBC. Orange color represents PBC-relevant genetic association signals showing a significant enrichment in cholangiocytes. e) 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.

Figure 6. The biological functions of *ORMDL3*$^{+}$ cholangiocytes. a) UMAP projections of liver and immune cells colored by *ORMDL3*$^{+}$ and *ORMDL3*$^{-}$ cholangiocytes. Blue color represents *ORMDL3*$^{-}$ cholangiocytes, red color represents *ORMDL3*$^{+}$ cholangiocytes, and gray color stands for the rest of all other cell types. b) UMAP embedding of all cells among 13 cell populations colored by the metabolism activity score of all metabolism-related pathways collected from the KEGG pathway

resource. c) Violin plot showing the difference in metabolism activity score between *ORMDL3*$^{+}$ and *ORMDL3*$^{-}$ cholangiocytes. Two-side Wilcoxon test was used for assessing significance. d) Volcano plot showing the differentially expressed genes (DEGs) between *ORMDL3*$^{+}$ cholangiocytes and *ORMDL3*$^{-}$ cholangiocytes. Green color represents 71 significantly down-regulated DEGs, and orange color represents 6 significantly up-regulated DEGs. e) Pathway enrichment analysis of 71 down-regulated DEGs based on the KEGG pathway resource. f) Scatter plot showing the dominant senders (sources) and receivers (targets) in a 2D space. y axis represents incoming interaction strength, and x axis represents outgoing interaction strength. The size of each node represents the count of cellular interactions. g) The cellular communications of *ORMDL3*$^{+}$ Cholangiocyte with other cell populations in liver tissues. The width represents the number of cellular interactions.

Figure 7. Dot plots showing that predicted cellular interactions of *ORMDL3*$^{+}$ Cholangiocyte with other cell populations in liver tissues. (a) *ORMDL3*$^{+}$ Cholangiocyte as a source (outgoing) communicating with other cells, (b) *ORMDL*$^{+}$ Cholangiocyte as a target (incoming) communicating with other cells. The circular size of represents the statistical significance of each ligand-receptor pair, and color represents the communication probability.

Figure 8. The inferred outgoing (source) communication patterns of cell populations in liver tissue. a) Five inferred outgoing communications patterns of 13 cell populations, which shows the correspondence between the inferred latent patterns and cell groups, as well as signaling pathways, b) The contribution of signaling pathways among the five outgoing communications patterns to liver and immune cell types. The red dashed box marked the VEGF signaling pathway. c) Heatmap showing the

relative importance of each cell group based on the computed four network centrality measures of VEGF signaling pathway network. d) Violin plot showing the difference in VEGF signaling score between *ORMDL3*$^{+}$ and *ORMDL3*$^{-}$ cholangiocytes. The expression levels of all genes in VEGF signaling pathway were used to calculate the VEGF signaling score. Two-side Wilcoxon test was used for assessing significance.

Table 1. Identification of 29 significant risk genes by using integrative genomics analysis

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
CSNK2B	1.40	7.17×10^{-19}	8.41×10^{-16}	3.22×10^{-11}	2.24×10^{-6}	1.67×10^{-11}	4.96×10^{-4} (Blood), 2×10^{-2} (CD14+T cell)	Novel gene
LY6G5B	-7.92	2.43×10^{-16}	1.93×10^{-13}	5.52×10^{-11}	1.54×10^{-10}	1.19×10^{-11}	Non-significant	Novel gene
DDAH2	-5.57	1.39×10^{-15}	9.99×10^{-13}	2.65×10^{-9}	1.56×10^{-11}	1.69×10^{-9}	2.70×10^{-3} (CD14+T cell)	Novel gene
LY6G5C	-7.82	2.40×10^{-15}	1.62×10^{-12}	8.64×10^{-13}	9.05×10^{-12}	1.41×10^{-11}	2.80×10^{-2} (Blood)	Novel gene
IRF5	7.11	2.62×10^{-15}	1.72×10^{-12}	4.29×10^{-13}	1.12×10^{-12}	7.18×10^{-11}	7.10×10^{-3} (Liver)	Reported gene
SOCS1	-3.26	9.33×10^{-13}	4.16×10^{-10}	5.25×10^{-10}	7.34×10^{-10}	1.06×10^{-8}	3.50×10^{-2} (CD14+T cell)	Reported gene
SYNGR1	-1.30	2.83×10^{-9}	8.88×10^{-7}	7.97×10^{-4}	1.30×10^{-7}	4.13×10^{-2}	1.40×10^{-2} (CD14+T cell)	Reported gene
C6orf48	-4.46	9.57×10^{-9}	2.66×10^{-6}	7.83×10^{-4}	6.19×10^{-4}	4.83×10^{-11}	4.16×10^{-6} (Blood)	Novel gene
HLA-DMA	-3.99	8.87×10^{-8}	1.90×10^{-5}	1.27×10^{-3}	9.16×10^{-4}	1.62×10^{-8}	2.10×10^{-3} (Liver)	Novel gene
SMC4	-1.34	1.00×10^{-7}	2.10×10^{-5}	9.38×10^{-5}	2.25×10^{-6}	2.07×10^{-4}	4.80×10^{-4} (Blood)	Novel gene
TCF19	4.55	1.49×10^{-7}	2.99×10^{-5}	9.46×10^{-5}	2.24×10^{-6}	1.98×10^{-4}	Non-significant	Novel gene

ORMDL3	-2.92	7.07×10^{-7}	1.22×10^{-4}	2.35×10^{-3}	7.38×10^{-5}	2.33×10^{-6}	Non-significant	Reported gene
KPNA4	-1.22	1.07×10^{-6}	1.82×10^{-4}	2.10×10^{-2}	3.58×10^{-3}	4.67×10^{-4}	3.40×10^{-3} (CD14+T cell)	Novel gene
MANBA	-1.45	1.48×10^{-6}	2.35×10^{-4}	2.46×10^{-6}	2.39×10^{-2}	8.30×10^{-8}	9.67×10^{-5} (Blood)	Reported gene
MFSD6	-0.32	1.61×10^{-6}	2.51×10^{-4}	2.55×10^{-2}	4.57×10^{-2}	2.06×10^{-3}	3.60×10^{-2} (Liver)	Novel gene
MED1	-2.13	1.76×10^{-6}	2.70×10^{-4}	1.65×10^{-2}	1.53×10^{-2}	4.42×10^{-5}	Non-significant	Novel gene
IDUA	3.95	2.37×10^{-6}	3.47×10^{-4}	3.94×10^{-2}	3.87×10^{-2}	8.95×10^{-4}	8.10×10^{-3} (Blood)	Reported gene
DGKQ	3.24	3.86×10^{-6}	5.37×10^{-4}	4.90×10^{-4}	3.36×10^{-4}	2.35×10^{-4}	8.34×10^{-7} (Blood)	Reported gene
FCRL3	-4.40	8.14×10^{-6}	1.07×10^{-3}	2.84×10^{-3}	2.84×10^{-3}	5.81×10^{-4}	Non-significant	Reported gene
ZCRB1	3.22	9.34×10^{-6}	1.20×10^{-3}	4.83×10^{-3}	5.02×10^{-3}	1.52×10^{-3}	5.90×10^{-3} (Liver); 2.8×10^{-3} (CD14+T cell);	Novel gene
AGAP5	-4.28	1.29×10^{-5}	1.59×10^{-3}	4.50×10^{-3}	3.40×10^{-3}	3.65×10^{-2}	Non-significant	Novel gene
CARM1	-2.90	2.11×10^{-5}	2.49×10^{-3}	1.14×10^{-3}	8.92×10^{-4}	1.86×10^{-4}	2.60×10^{-2} (Blood)	Novel gene
UBE2D3	-4.18	2.83×10^{-5}	3.23×10^{-3}	2.76×10^{-3}	5.08×10^{-3}	5.81×10^{-4}	3.40×10^{-2} (Blood); 1.6×10^{-4} (CD14+T cell)	Novel gene
NAAA	4.08	6.73×10^{-5}	6.88×10^{-3}	1.25×10^{-2}	6.85×10^{-3}	7.41×10^{-4}	3.20×10^{-2} (Liver)	Reported gene
SH2B3	-1.69	1.20×10^{-4}	1.17×10^{-2}	2.35×10^{-2}	1.87×10^{-2}	2.44×10^{-3}	1.40×10^{-2} (Blood)	Reported gene
CTSH	1.75	2.05×10^{-4}	1.84×10^{-2}	8.53×10^{-3}	3.46×10^{-2}	2.19×10^{-5}	Non-significant	Novel gene

TTC34	3.56	2.82×10^{-4}	2.39×10^{-2}	5.86×10^{-3}	5.02×10^{-3}	4.98×10^{-3}	9.80×10^{-3} (Blood)	Novel gene
METTL1	2.19	3.11×10^{-4}	2.59×10^{-2}	4.95×10^{-2}	3.81×10^{-2}	1.68×10^{-2}	5.86×10^{-7} (Blood)	Novel gene
TSFM	3.45	3.14×10^{-4}	2.61×10^{-2}	4.87×10^{-2}	3.87×10^{-2}	1.54×10^{-2}	2.80×10^{-4} (Blood); 1.5×10^{-3} (CD14+T cell)	Novel gene

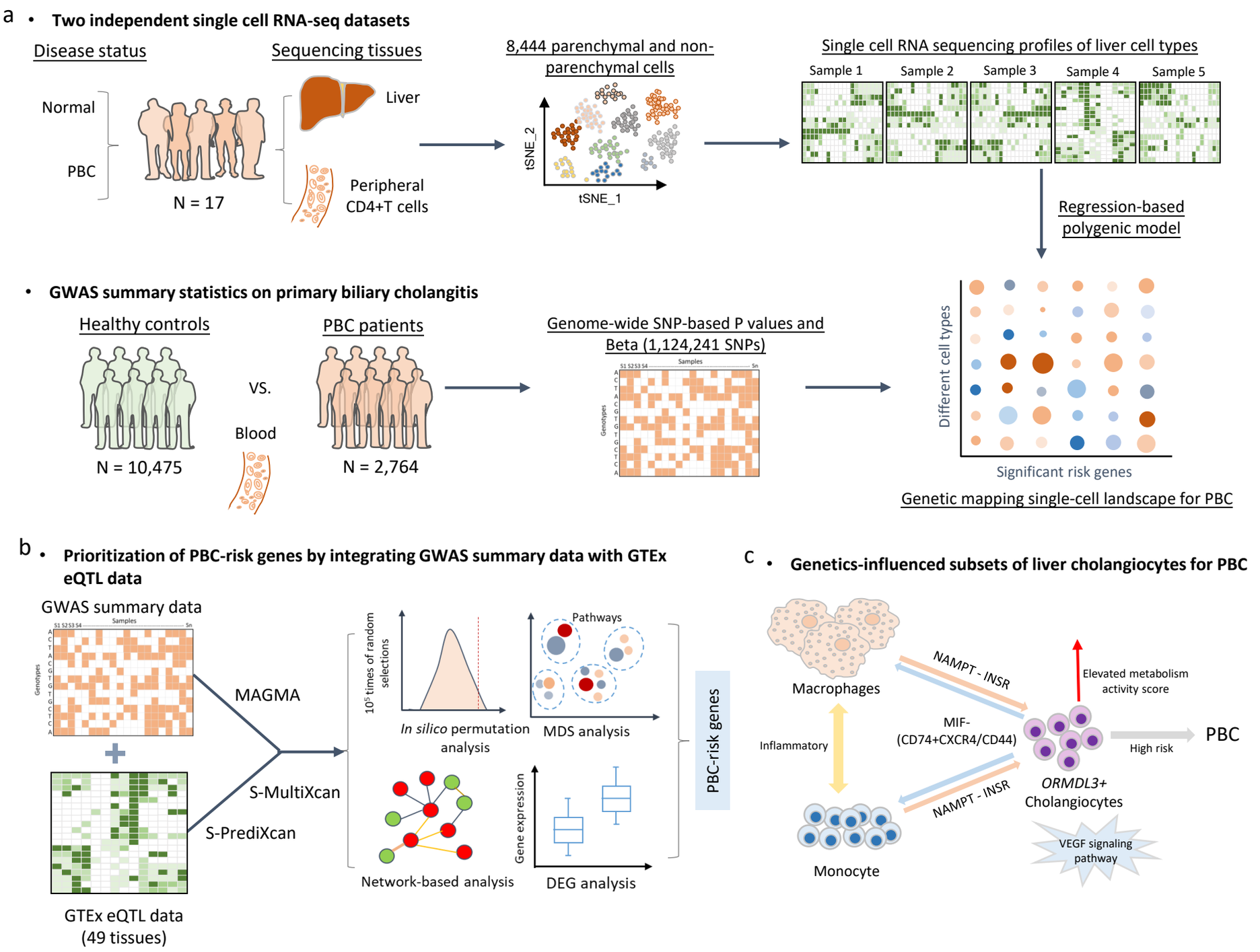
a
Two independent single cell RNA-seq datasets
Disease status
Sequencing tissues
Normal
PBC
N = 17
Liver
Peripheral CD4+T cells
8,444 parenchymal and non-parenchymal cells
tSNE_2
tSNE_1
Single cell RNA sequencing profiles of liver cell types
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Regression-based polygenic model
GWAS summary statistics on primary biliary cholangitis
Healthy controls
VS.
Blood
PBC patients
N = 10,475
N = 2,764
Genome-wide SNP-based P values and Beta (1,124,241 SNPs)
Different cell types
Significant risk genes
Genetic mapping single-cell landscape for PBC
b
Prioritization of PBC-risk genes by integrating GWAS summary data with GTEx eQTL data
GWAS summary data
MAGMA
S-MultiXcan
S-PrediXcan
GTEx eQTL data (49 tissues)
10^5 times of random selections
In silico permutation analysis
Pathways
MDS analysis
Network-based analysis
Gene expression
DEG analysis
PBC-risk genes
c
Genetics-influenced subsets of liver cholangiocytes for PBC
Macrophages
NAMPT - INSR
Elevated metabolism activity score
Inflammatory
MIF-
(CD74+CXCR4/CD44)
High risk
PBC
ORMDL3+
Cholangiocytes
NAMPT - INSR
Monocyte
VEGF signaling pathway

Figure 1

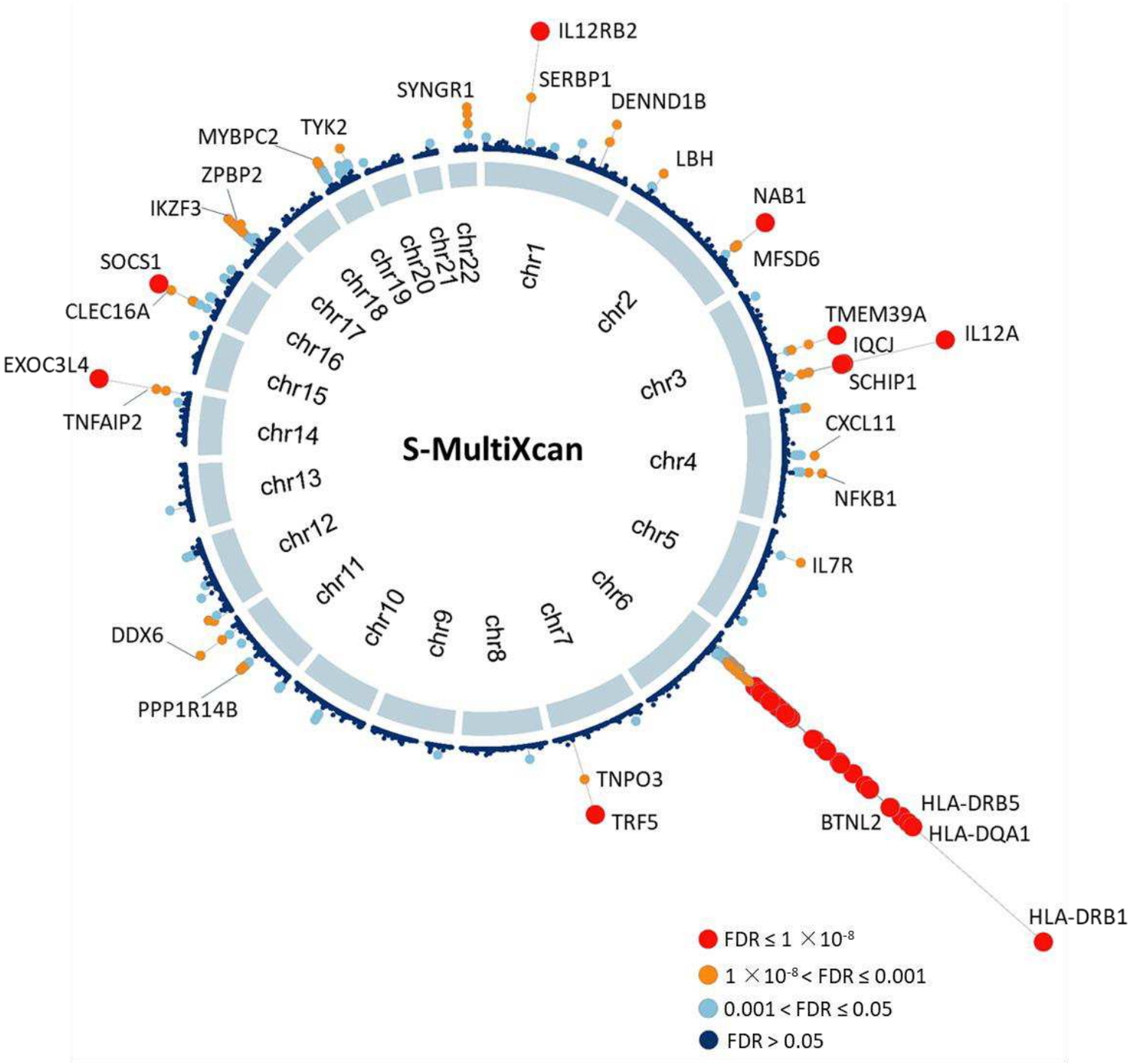
S-MultiXcan
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
IL12RB2
SERBP1
SYNGR1
DENND1B
LBH
NAB1
MFSD6
TMEM39A
IQCJ
IL12A
SCHIP1
CXCL11
NFKB1
IL7R
HLA-DRB5
BTNL2
HLA-DQA1
HLA-DRB1
TNPO3
TRF5
DDX6
PPP1R14B
EXOC3L4
TNFAIP2
SOCS1
CLEC16A
IKZF3
ZPBP2
MYBPC2
TYK2
FDR ≤ 1 × 10-8
1 × 10-8 < FDR ≤ 0.001
0.001 < FDR ≤ 0.05
FDR > 0.05

Figure 2

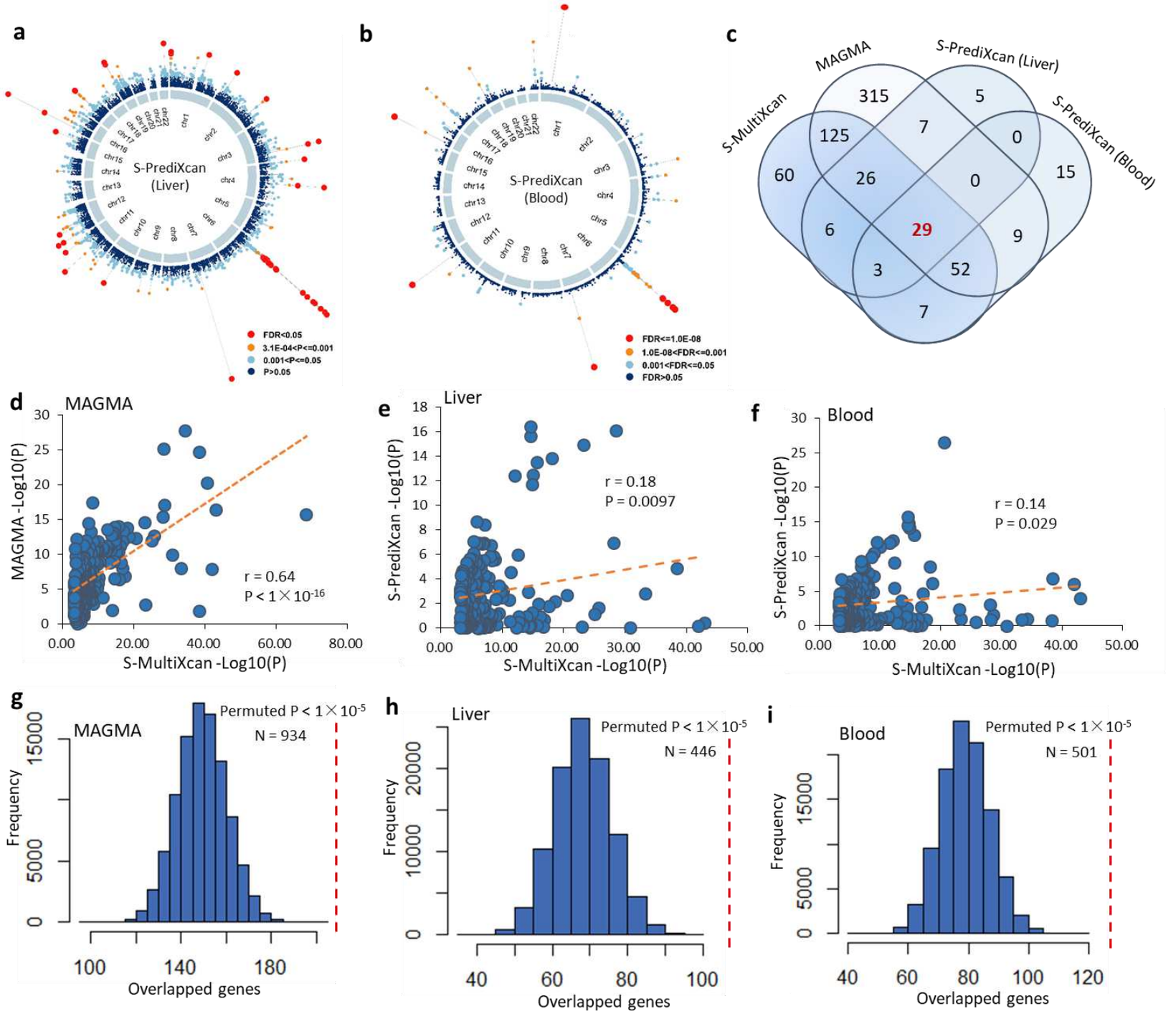
a
S-PrediXcan (Liver)
FDR<0.05
3.1E-04<P<=0.001
0.001<P<=0.05
P>0.05
b
S-PrediXcan (Blood)
FDR<=1.0E-08
1.0E-08<FDR<=0.001
0.001<FDR<=0.05
FDR>0.05
c
MAGMA
S-PrediXcan (Liver)
S-MultiXcan
S-PrediXcan (Blood)
315
5
125
7
0
60
26
0
15
6
29
9
3
52
7
d
MAGMA
MAGMA -Log10(P)
S-MultiXcan -Log10(P)
r = 0.64
P < 1×10-16
e
Liver
S-PrediXcan -Log10(P)
S-MultiXcan -Log10(P)
r = 0.18
P = 0.0097
f
Blood
S-PrediXcan -Log10(P)
S-MultiXcan -Log10(P)
r = 0.14
P = 0.029
g
MAGMA
Permuted P < 1×10-5
N = 934
Frequency
Overlapped genes
h
Liver
Permuted P < 1×10-5
N = 446
Frequency
Overlapped genes
i
Blood
Permuted P < 1×10-5
N = 501
Frequency
Overlapped genes

Figure 3

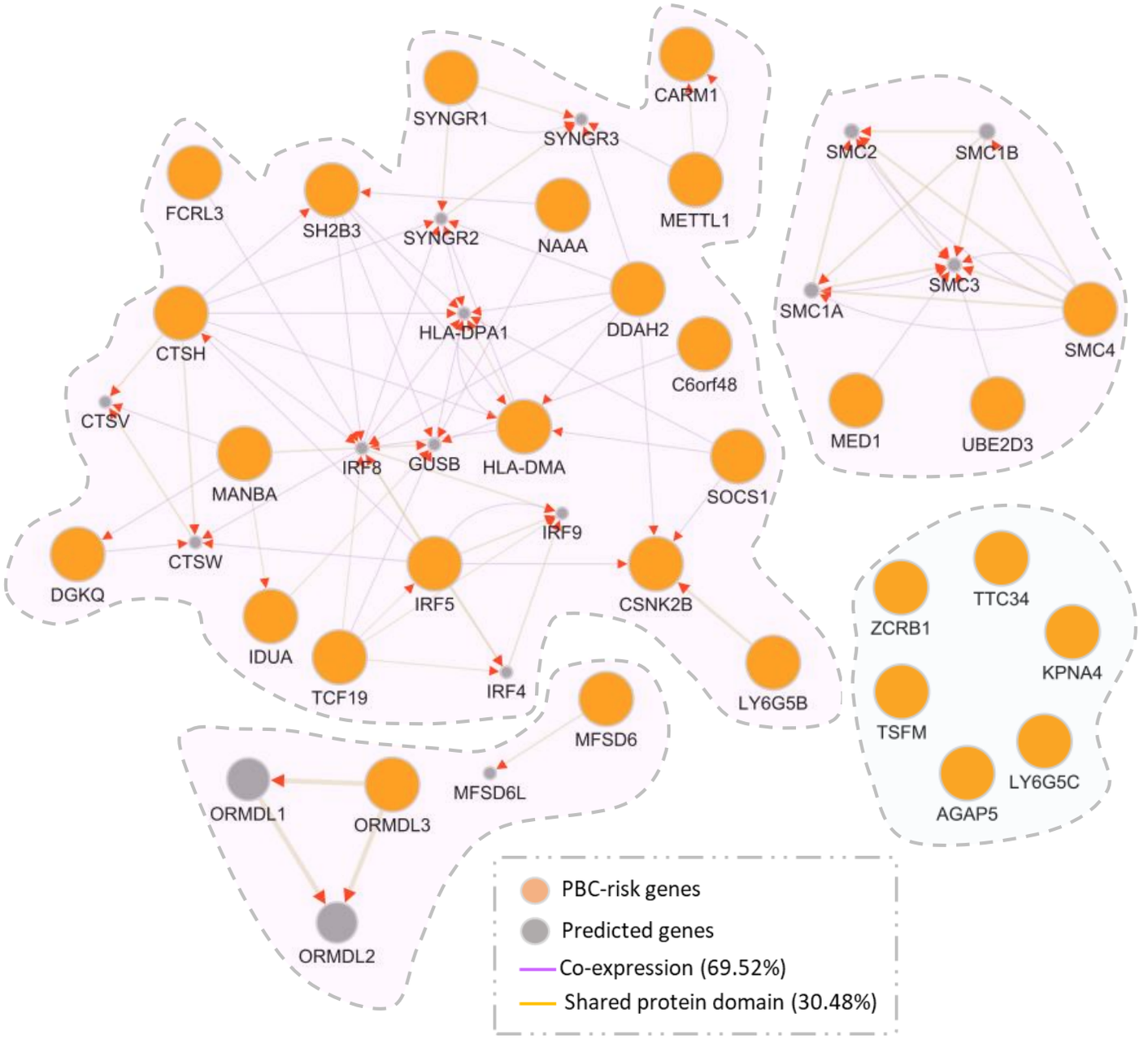
SYNGR1
CARM1
SYNGR3
SMC2
SMC1B
FCRL3
SH2B3
SYNGR2
NAAA
METTL1
SMC3
SMC1A
SMC4
HLA-DPA1
DDAH2
CTSH
C6orf48
CTSV
MED1
UBE2D3
IRF8
GUSB
HLA-DMA
MANBA
SOCS1
IRF9
CTSW
DGKQ
IRF5
CSNK2B
ZCRB1
TTC34
IDUA
KPNA4
TCF19
IRF4
LY6G5B
MFSD6
TSFM
MFSD6L
LY6G5C
ORMDL1
ORMDL3
AGAP5
ORMDL2
PBC-risk genes
Predicted genes
Co-expression (69.52%)
Shared protein domain (30.48%)

Figure 4

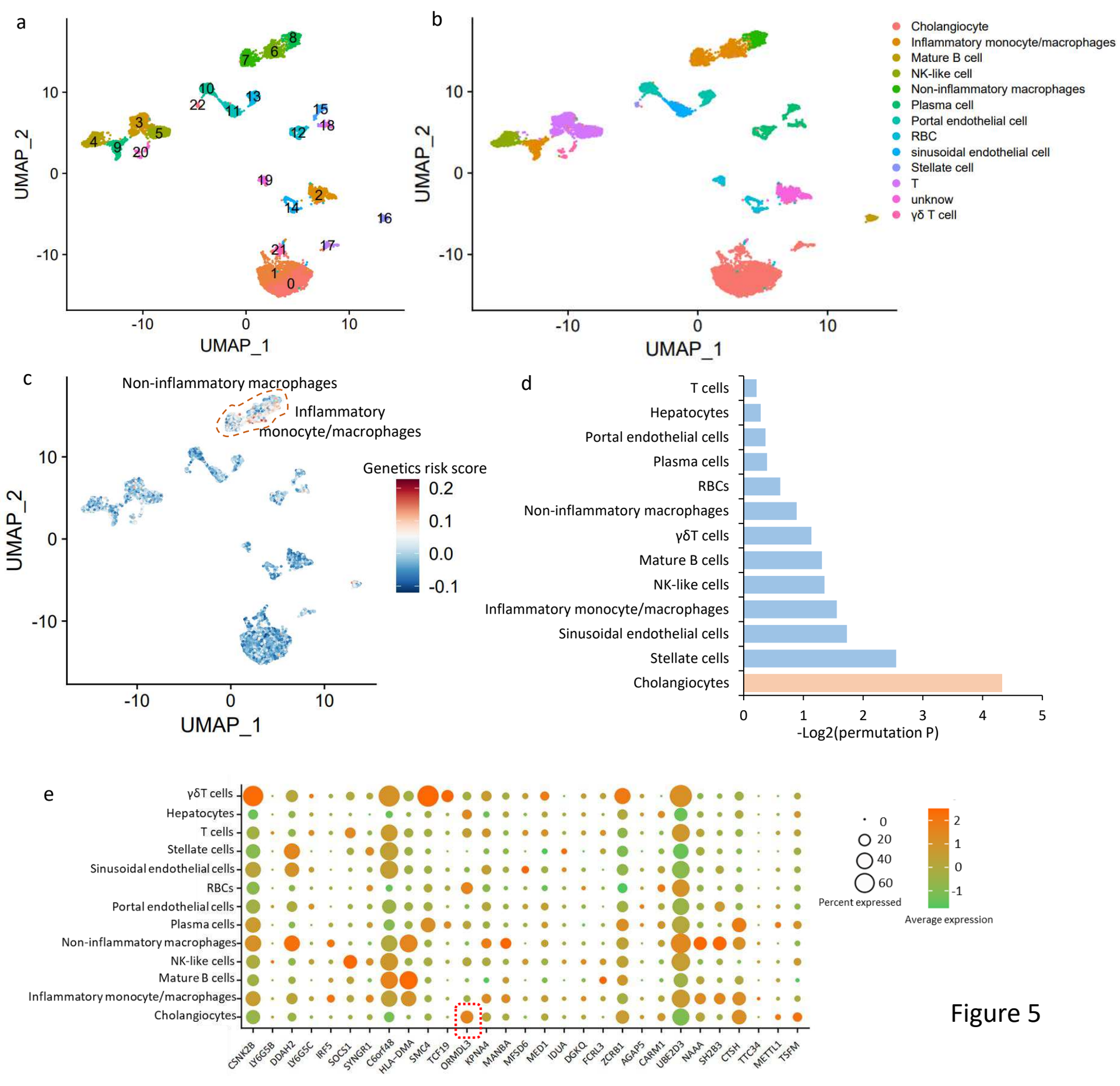
a
b
c
d
e
UMAP_1
UMAP_2
Cholangiocyte
Inflammatory monocyte/macrophages
Mature B cell
NK-like cell
Non-inflammatory macrophages
Plasma cell
Portal endothelial cell
RBC
sinusoidal endothelial cell
Stellate cell
T
unknow
γδ T cell
Non-inflammatory macrophages
Inflammatory monocyte/macrophages
Genetics risk score
0.2
0.1
0.0
-0.1
T cells
Hepatocytes
Portal endothelial cells
Plasma cells
RBCs
Non-inflammatory macrophages
γδT cells
Mature B cells
NK-like cells
Inflammatory monocyte/macrophages
Sinusoidal endothelial cells
Stellate cells
Cholangiocytes
-Log2(permutation P)
γδT cells
Hepatocytes
T cells
Stellate cells
Sinusoidal endothelial cells
RBCs
Portal endothelial cells
Plasma cells
Non-inflammatory macrophages
NK-like cells
Mature B cells
Inflammatory monocyte/macrophages
Cholangiocytes
CSNK2B
LY6G5B
DDAH2
LY6G5C
IRF5
SOCS1
SYNGR1
C6orf48
HLA-DMA
SMC4
TCF19
ORMDL3
KPNA4
MANBA
MFSD6
MED1
IDUA
DGKQ
FCRL3
ZCRB1
AGAP5
CARM1
UBE2D3
NAAA
SH2B3
CTSH
TTC34
METTL1
TSFM
Percent expressed
Average expression

Figure 5

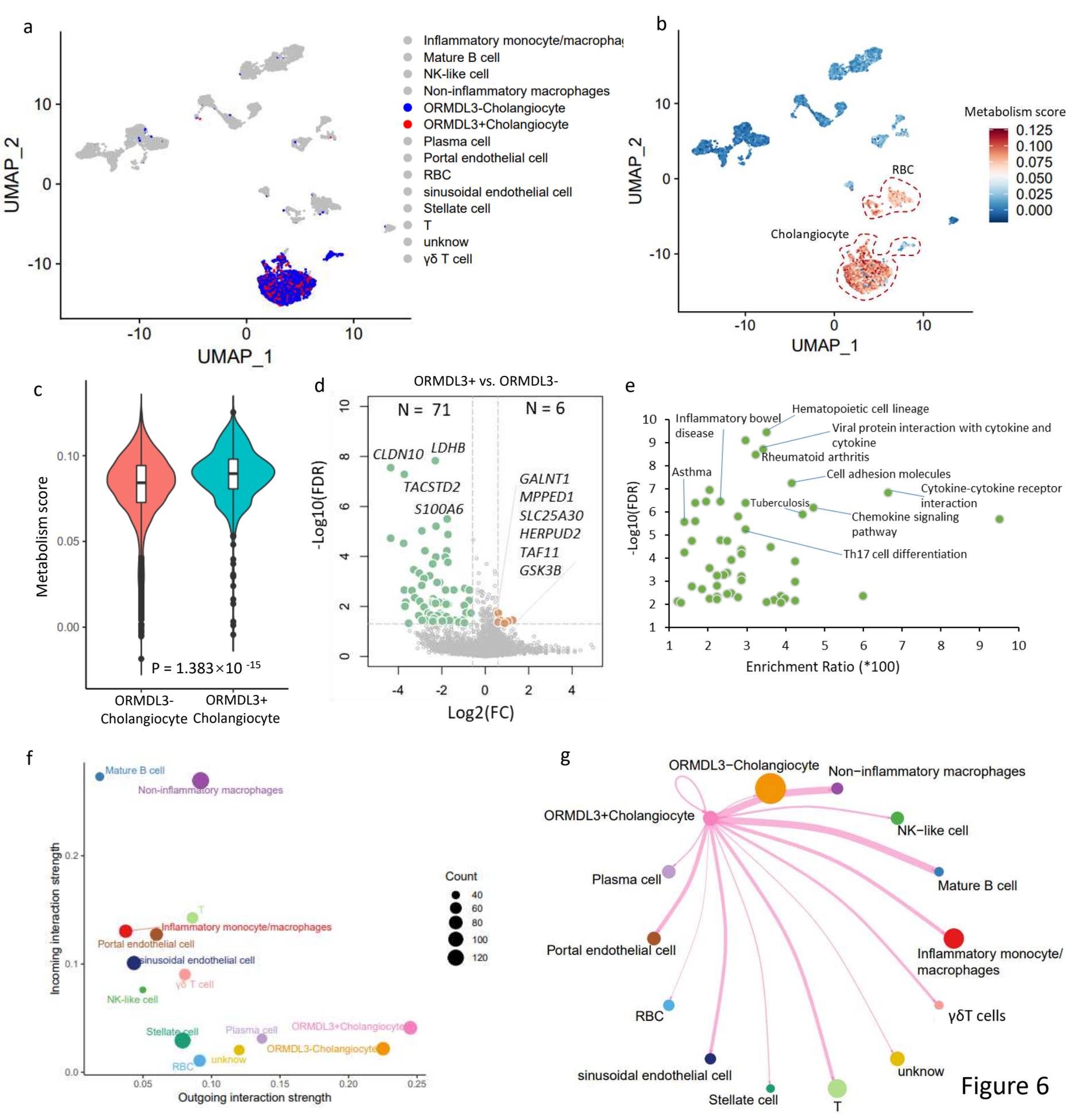
a
Inflammatory monocyte/macropha
Mature B cell
NK-like cell
Non-inflammatory macrophages
ORMDL3-Cholangiocyte
ORMDL3+Cholangiocyte
Plasma cell
Portal endothelial cell
RBC
sinusoidal endothelial cell
Stellate cell
T
unknow
γδ T cell
UMAP_1
UMAP_2
b
Metabolism score
0.125
0.100
0.075
0.050
0.025
0.000
RBC
Cholangiocyte
c
Metabolism score
P = 1.383×10 -15
ORMDL3- Cholangiocyte
ORMDL3+ Cholangiocyte
d
ORMDL3+ vs. ORMDL3-
N = 71
N = 6
CLDN10
LDHB
TACSTD2
S100A6
GALNT1
MPPED1
SLC25A30
HERPUD2
TAF11
GSK3B
-Log10(FDR)
Log2(FC)
e
Inflammatory bowel disease
Hematopoietic cell lineage
Viral protein interaction with cytokine and cytokine
Rheumatoid arthritis
Asthma
Cell adhesion molecules
Cytokine-cytokine receptor interaction
Tuberculosis
Chemokine signaling pathway
Th17 cell differentiation
-Log10(FDR)
Enrichment Ratio (*100)
f
Mature B cell
Non-inflammatory macrophages
T
Inflammatory monocyte/macrophages
Portal endothelial cell
sinusoidal endothelial cell
γδ T cell
NK-like cell
Stellate cell
Plasma cell
ORMDL3+Cholangiocyte
ORMDL3-Cholangiocyte
RBC
unknow
Count
40
60
80
100
120
Incoming interaction strength
Outgoing interaction strength
g
ORMDL3−Cholangiocyte
Non−inflammatory macrophages
ORMDL3+Cholangiocyte
NK−like cell
Plasma cell
Mature B cell
Portal endothelial cell
Inflammatory monocyte/ macrophages
RBC
γδT cells
sinusoidal endothelial cell
unknow
Stellate cell
T

Figure 6

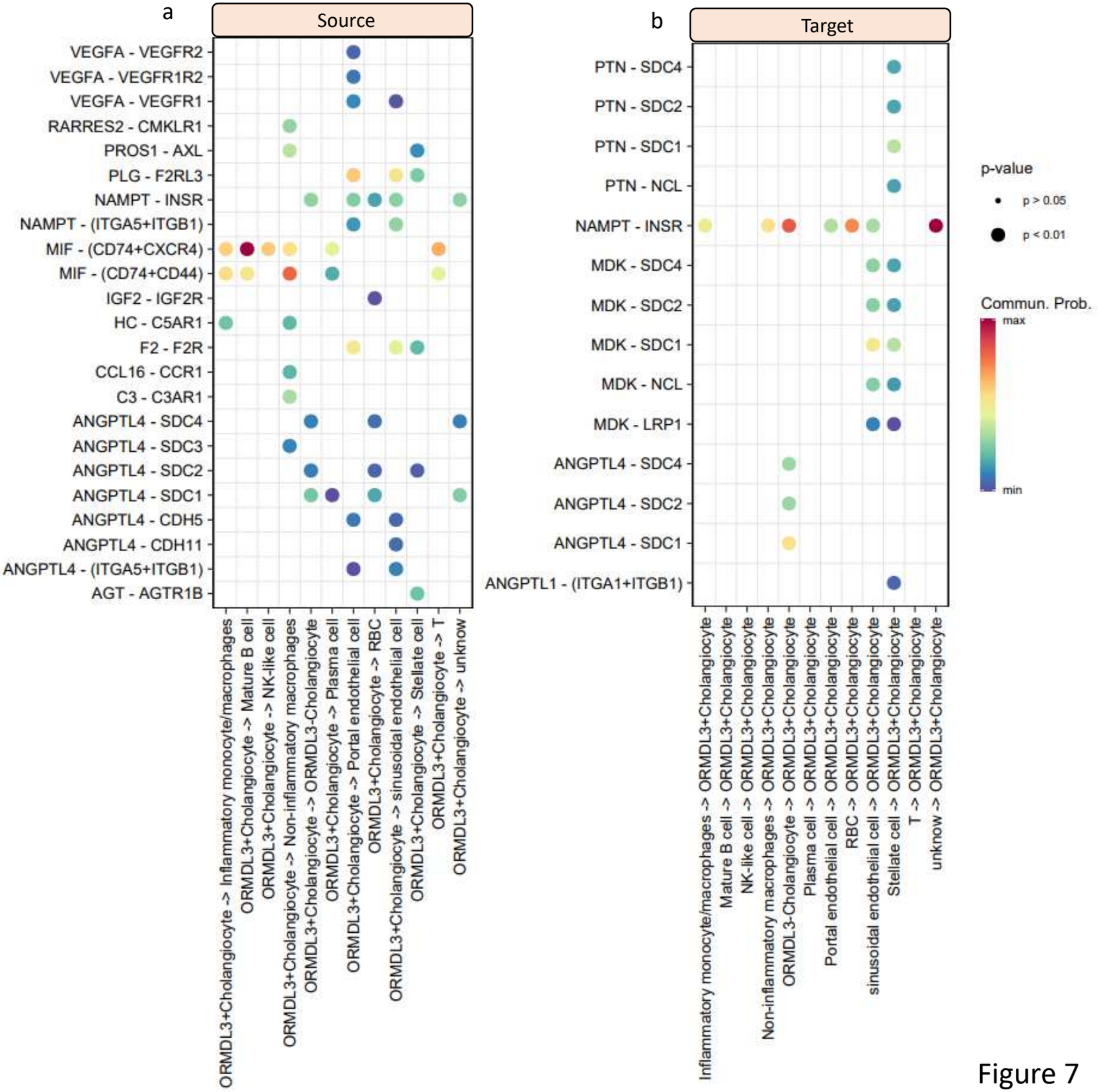
a
Source
VEGFA - VEGFR2
VEGFA - VEGFR1R2
VEGFA - VEGFR1
RARRES2 - CMKLR1
PROS1 - AXL
PLG - F2RL3
NAMPT - INSR
NAMPT - (ITGA5+ITGB1)
MIF - (CD74+CXCR4)
MIF - (CD74+CD44)
IGF2 - IGF2R
HC - C5AR1
F2 - F2R
CCL16 - CCR1
C3 - C3AR1
ANGPTL4 - SDC4
ANGPTL4 - SDC3
ANGPTL4 - SDC2
ANGPTL4 - SDC1
ANGPTL4 - CDH5
ANGPTL4 - CDH11
ANGPTL4 - (ITGA5+ITGB1)
AGT - AGTR1B
ORMDL3+Cholangiocyte -> Inflammatory monocyte/macrophages
ORMDL3+Cholangiocyte -> Mature B cell
ORMDL3+Cholangiocyte -> NK-like cell
ORMDL3+Cholangiocyte -> Non-inflammatory macrophages
ORMDL3+Cholangiocyte -> ORMDL3-Cholangiocyte
ORMDL3+Cholangiocyte -> Plasma cell
ORMDL3+Cholangiocyte -> Portal endothelial cell
ORMDL3+Cholangiocyte -> RBC
ORMDL3+Cholangiocyte -> sinusoidal endothelial cell
ORMDL3+Cholangiocyte -> Stellate cell
ORMDL3+Cholangiocyte -> T
ORMDL3+Cholangiocyte -> unknow
b
Target
PTN - SDC4
PTN - SDC2
PTN - SDC1
PTN - NCL
NAMPT - INSR
MDK - SDC4
MDK - SDC2
MDK - SDC1
MDK - NCL
MDK - LRP1
ANGPTL4 - SDC4
ANGPTL4 - SDC2
ANGPTL4 - SDC1
ANGPTL1 - (ITGA1+ITGB1)
Inflammatory monocyte/macrophages -> ORMDL3+Cholangiocyte
Mature B cell -> ORMDL3+Cholangiocyte
NK-like cell -> ORMDL3+Cholangiocyte
Non-inflammatory macrophages -> ORMDL3+Cholangiocyte
ORMDL3-Cholangiocyte -> ORMDL3+Cholangiocyte
Plasma cell -> ORMDL3+Cholangiocyte
Portal endothelial cell -> ORMDL3+Cholangiocyte
RBC -> ORMDL3+Cholangiocyte
sinusoidal endothelial cell -> ORMDL3+Cholangiocyte
Stellate cell -> ORMDL3+Cholangiocyte
T -> ORMDL3+Cholangiocyte
unknow -> ORMDL3+Cholangiocyte
p-value
p > 0.05
p < 0.01
Commun. Prob.
max
min

Figure 7

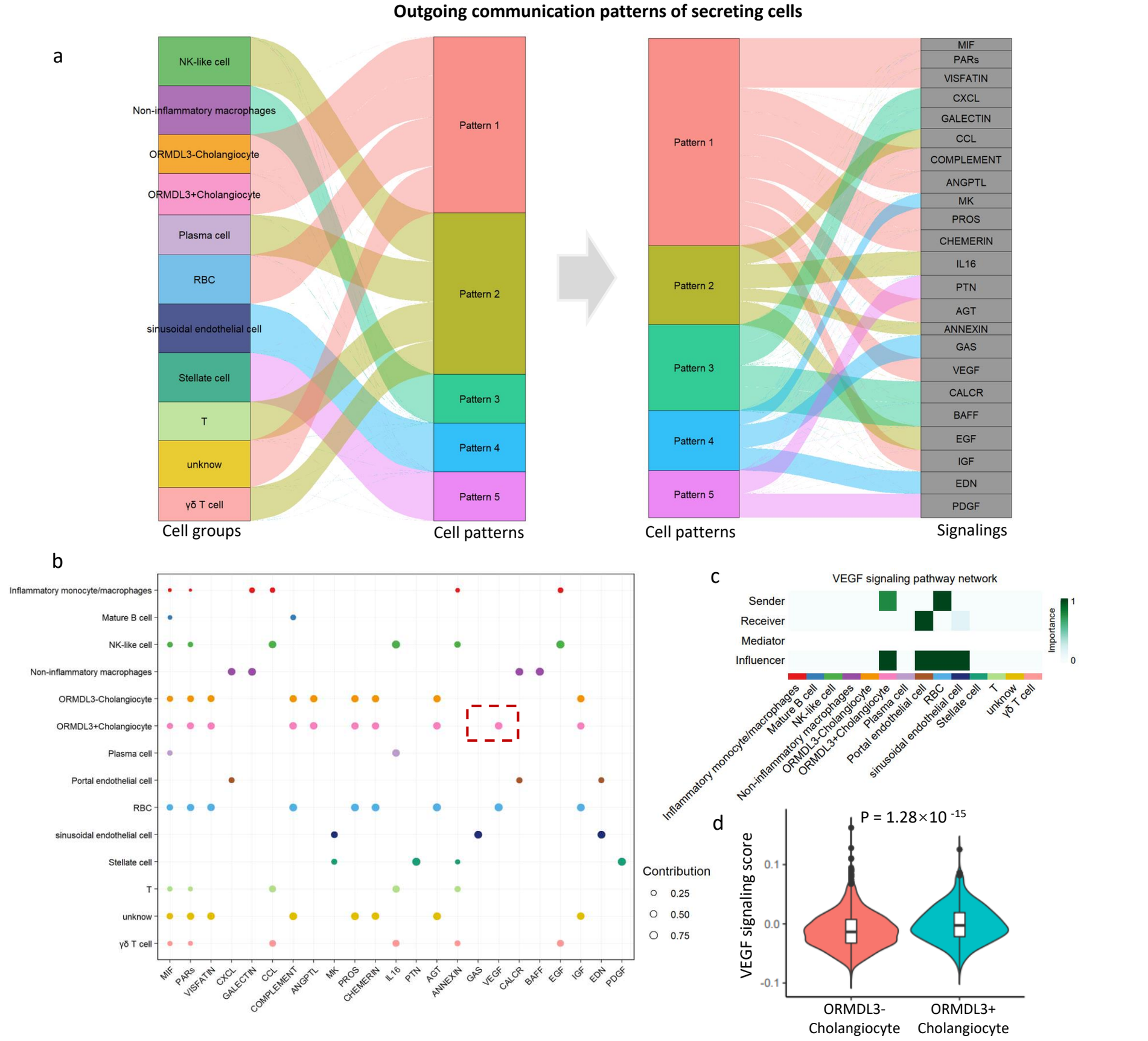
Outgoing communication patterns of secreting cells
a
NK-like cell
Non-inflammatory macrophages
ORMDL3-Cholangiocyte
ORMDL3+Cholangiocyte
Plasma cell
RBC
sinusoidal endothelial cell
Stellate cell
T
unknow
γδ T cell
Cell groups
Pattern 1
Pattern 2
Pattern 3
Pattern 4
Pattern 5
Cell patterns
Pattern 1
Pattern 2
Pattern 3
Pattern 4
Pattern 5
Cell patterns
MIF
PARs
VISFATIN
CXCL
GALECTIN
CCL
COMPLEMENT
ANGPTL
MK
PROS
CHEMERIN
IL16
PTN
AGT
ANNEXIN
GAS
VEGF
CALCR
BAFF
EGF
IGF
EDN
PDGF
Signalings
b
Inflammatory monocyte/macrophages
Mature B cell
NK-like cell
Non-inflammatory macrophages
ORMDL3-Cholangiocyte
ORMDL3+Cholangiocyte
Plasma cell
Portal endothelial cell
RBC
sinusoidal endothelial cell
Stellate cell
T
unknow
γδ T cell
MIF
PARs
VISFATIN
CXCL
GALECTIN
CCL
COMPLEMENT
ANGPTL
MK
PROS
CHEMERIN
IL16
PTN
AGT
ANNEXIN
GAS
VEGF
CALCR
BAFF
EGF
IGF
EDN
PDGF
Contribution
0.25
0.50
0.75
c
VEGF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
1
0
Inflammatory monocyte/macrophages
Mature B cell
NK-like cell
Non-inflammatory macrophages
ORMDL3-Cholangiocyte
ORMDL3+Cholangiocyte
Plasma cell
Portal endothelial cell
RBC
sinusoidal endothelial cell
Stellate cell
T
unknow
γδ T cell
d
P = 1.28×10 -15
VEGF signaling score
0.1
0.0
-0.1
ORMDL3-
Cholangiocyte
ORMDL3+
Cholangiocyte

Figure 8

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Graphicabstract.pdf
- SupplementalFigures.docx
- SupplementalTables.docx