

# Single-cell, single-nucleus, and spatial transcriptomics characterization of the immunological landscape in the healthy and PSC human liver

Tallulah S. Andrews<sup>1,2,3,\*†</sup>, Diana Nakib<sup>1,4,\*†</sup>, Catia T. Perciani<sup>1,4,5</sup>, Xue Zhong Ma<sup>1</sup>, Lewis Liu<sup>1,4</sup>, Erin Winter<sup>1</sup>, Damra Camat<sup>1,4</sup>, Sai W. Chung<sup>1,4</sup>, Patricia Lumanto<sup>1,4</sup>, Justin Manuel<sup>1</sup>, Shantel Mangroo<sup>6</sup>, Bettina Hansen<sup>7,8</sup>, Bal Arpinder<sup>1</sup>, Cornelia Thoeni<sup>5</sup>, Blayne Sayed<sup>1</sup>, Jordan Feld<sup>7</sup>, Adam Gehring<sup>4,7</sup>, Aliya Gulamhusein<sup>7</sup>, Gideon M. Hirschfield<sup>7</sup>, Amanda Ricciuto<sup>5</sup>, Gary D. Bader<sup>9,\*†</sup>, Ian D. McGilvray<sup>1,\*†</sup>, Sonya MacParland<sup>1,4,5,\*†</sup>

Journal of Hepatology 2024. vol. 80 | 730–743



See Editorial, pages 681–683

**Background & Aims:** Primary sclerosing cholangitis (PSC) is an immune-mediated cholestatic liver disease for which there is an unmet need to understand the cellular composition of the affected liver and how it underlies disease pathogenesis. We aimed to generate a comprehensive atlas of the PSC liver using multi-omic modalities and protein-based functional validation.

**Methods:** We employed single-cell and single-nucleus RNA sequencing (47,156 cells and 23,000 nuclei) and spatial transcriptomics (one sample by 10x Visium and five samples with Nanostring GeoMx DSP) to profile the cellular ecosystem in 10 PSC livers. Transcriptomic profiles were compared to 24 neurologically deceased donor livers (107,542 cells) and spatial transcriptomics controls, as well as 18,240 cells and 20,202 nuclei from three PBC livers. Flow cytometry was performed to validate PSC-specific differences in immune cell phenotype and function.

**Results:** PSC explants with parenchymal cirrhosis and prominent periductal fibrosis contained a population of cholangiocyte-like hepatocytes that were surrounded by diverse immune cell populations. PSC-associated biliary, mesenchymal, and endothelial populations expressed chemokine and cytokine transcripts involved in immune cell recruitment. Additionally, expanded CD4<sup>+</sup> T cells and recruited myeloid populations in the PSC liver expressed the corresponding receptors to these chemokines and cytokines, suggesting potential recruitment. Tissue-resident macrophages, by contrast, were reduced in number and exhibited a dysfunctional and downregulated inflammatory response to lipopolysaccharide and interferon- $\gamma$  stimulation.

**Conclusions:** We present a comprehensive atlas of the PSC liver and demonstrate an exhaustion-like phenotype of myeloid cells and markers of chronic cytokine expression in late-stage PSC lesions. This atlas expands our understanding of the cellular complexity of PSC and has potential to guide the development of novel treatments.

© 2024 The Author(s). Published by Elsevier B.V. on behalf of European Association for the Study of the Liver. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Recent efforts to generate a single-cell atlas have revealed a broad diversity of parenchymal, progenitor, tissue-resident and transient circulating immune cells within the liver.<sup>1,2,4</sup> Changes in the frequency and characteristics of these populations have also been observed in atlases of acute liver disease, chronic fibrosis, and hepatic cancers.<sup>3,5,6</sup> Understanding the healthy liver by building a comprehensive reference map with an inventory of cells and their respective spatial localization is crucial in elucidating the transcriptomic and phenotypic changes, rare cell types, and cell-cell interactions that underlie disease development.<sup>7,8</sup>

Primary sclerosing cholangitis (PSC) is an immune-mediated cholestatic liver disease that is characterized by the retention of

bile, destruction of the biliary tree and development of fibrosis.<sup>9</sup> Given the rarity of the disease and the timeline from diagnosis to end-stage liver disease, few studies have attempted to map the PSC liver and characterize its cellular landscape. Previous transcriptomic maps of PSC have only considered sorted populations of immune cells.<sup>10,11</sup> These and other studies assessing PSC and PSC-like murine models that have employed bulk tissue expression have implicated CD4<sup>+</sup> T cells,<sup>10,12</sup> neutrophils,<sup>13</sup> dendritic cells,<sup>11</sup> antibody-producing B cells,<sup>14,15</sup> and macrophages<sup>16–22</sup> in the development of disease. However, the focused scope and low cellular resolution of these studies leave many uncertainties in the pathogenesis and outcome of PSC inflammation, preventing the development of targeted treatments.

Keywords: Liver; Single Cell RNA sequencing; Spatial Transcriptomics; Primary Sclerosing Cholangitis; Myeloid Dysfunction.

Received 1 February 2023; received in revised form 13 December 2023; accepted 20 December 2023; available online 8 January 2024

\* Corresponding authors. Address: MaRS TMDT, Rm 3-302, 101 College St, Toronto, Ontario, M5G 2L7, Canada (416) 581-7442.

E-mail addresses: [s.macparland@utoronto.ca](mailto:s.macparland@utoronto.ca) (S. MacParland), [tandrew6@uwo.ca](mailto:tandrew6@uwo.ca) (T.S. Andrews), [diana.nakib@mail.utoronto.ca](mailto:diana.nakib@mail.utoronto.ca) (D. Nakib), [gary.bader@utoronto.ca](mailto:gary.bader@utoronto.ca) (G.D. Bader), [ian.mcgilvray@uhn.ca](mailto:ian.mcgilvray@uhn.ca) (I.D. McGilvray).

† Equal contribution first authors

‡ Equal contribution senior authors

<https://doi.org/10.1016/j.jhep.2023.12.023>



We present the first single-cell transcriptomic atlas of PSC, and a primary biliary cholangitis (PBC) comparator group, using 5' single-cell, 3' single-nucleus and spatial transcriptomics, as well as a complementary 100,000 cell reference neurologically deceased healthy donor (NDD) liver map. These maps revealed seven disease-associated cell subtypes including T cell- and natural killer (NK) cell-intermediate states, mast cells, dendritic cells, and neutrophils, as well as extensive disease-specific cell-cell interactions between immune and non-immune liver cells. This includes a subpopulation of hepatocytes identified surrounding fibrotic lesions that appeared to have lost their zoned expression and co-expressed cholangiocyte-specific markers, while TREM2<sup>+</sup> and monocyte-like macrophages were concentrated within fibrotic regions. In experimental validations, these macrophages exhibited suppressed inflammatory potential, which could represent a pathway by which these myeloid cells contribute to immune dysregulation in PSC.

## Materials and methods

### Human liver tissue

Healthy human liver tissue from the caudate lobe was obtained from NDD livers which were acceptable for liver transplantation and without evidence of histopathological liver disease. Samples were collected with institutional ethics approval from the University Health Network, Toronto, Canada (REB# 14-7425-AE). PSC and PBC samples were collected from explanted tissue sections (~2 cm<sup>3</sup>) and/or perfusable caudate lobes at the time of transplantation with institutional ethics approval from the University Health Network (REB# 20-5142). All patient clinical characteristics are shown in [Table S1](#).

### NDD, PSC and PBC sample collection and processing for scRNA-seq & snRNA-seq

Samples from NDD (24), PSC (10) and PBC (3) livers were collected and processed for single-cell RNA sequencing (scRNA-seq) fresh or following cryopreservation ([Table S1](#)) with a collagenase dissociation protocol<sup>1,23</sup> ([dx.doi.org/10.17504/protocols.io.m9sc96e](https://doi.org/10.17504/protocols.io.m9sc96e)), in addition to being snap frozen for single-cell RNA sequencing (snRNA-seq), as previously described, by snap-freezing in liquid nitrogen<sup>24,25</sup> ([dx.doi.org/10.17504/protocols.io.261ge34qdl47/v1](https://doi.org/10.17504/protocols.io.261ge34qdl47/v1)) ([Fig. 1A,B](#), details in the [supplementary methods](#)).

### Preprocessing and quality control of 10x chromium samples

Sequencing reads were quantified using cellranger mapping to hg18 (for specific versions see [Table S1](#)). Droplets containing viable cells were identified using EmptyDrops from the DropletUtils (v1.2.0) package<sup>26</sup> ([Fig. 1A](#)). Each sample was scaled individually, and the top 2,000 most highly variable genes were identified in each sample using Seurat (v3.1.3).<sup>27</sup> Cells were initially annotated using a custom algorithm (see the [supplementary methods](#)) using our previous scRNA-seq map.<sup>1</sup>

### Data integration, clustering and annotation

Raw, normalized and scaled expression matrices for all samples were merged and then integrated using Harmony (v1.0),<sup>28</sup> with default parameters. UMAP (uniform manifold approximation and

projection) were recalculated using the integrated lower dimensional space.

The integrated NDD liver map was clustered using Seurat (v3.1.3)<sup>27</sup> and refined with apcluster (v1.4.8)<sup>29,30</sup> to identify the optimal clustering (see the [supplementary methods](#)).

### Liver map subclustering, pathway analysis, and cell-cell communication analysis

The integrated NDD liver map clusters were classified into eight general classes ([Fig. S1](#)). Cells in each of these classes were subset and subclustered by repeating the entire clustering and integration pipeline ([Fig. S2-5](#)). The subclusters were manually annotated using genes significantly expressed in that cluster compared to all other clusters ("marker genes") identified using the Wilcoxon rank-sum test. We used scmap<sup>31</sup> to determine the specific identity of cells comprising doublets ([Fig. S6](#), [Table S2](#)). Potential ligand-receptor interactions were identified using CellPhoneDB (v4.1.0)<sup>32</sup> (see the [supplementary methods](#)). For pathway analysis, genes were ranked by their estimated fold changes and analyzed with the fgsea package (v1.8.0) with 5% false discovery rate (FDR) and 100,000 permutations using the MSigdb Hallmark pathways, MSigdb Immune pathways, and Reactome pathways.

### Cell type comparisons of PSC or PBC vs. NDD

To compare cell type-specific differential expression between PSC, PBC and NDD single-cell and single-nucleus data, we calculated pseudobulk gene expression for each cell type in each sample and were compared between NDD and PSC/PBC using edgeR's exactTest<sup>33</sup> (see the [supplementary methods](#), [Tables S4 and S5](#)).

### 10x Genomics Visium and Nanostring GeoMx DSP experimental protocol

Tissue was prepared for Visium Spatial transcriptomics as previously described.<sup>2</sup> PSC and NDD liver sections were sliced from OCT-embedded sections and submitted to NanoString for staining with selected morphological markers and sequencing with the GeoMx Cancer Transcriptome Atlas (see the [supplementary methods](#)).

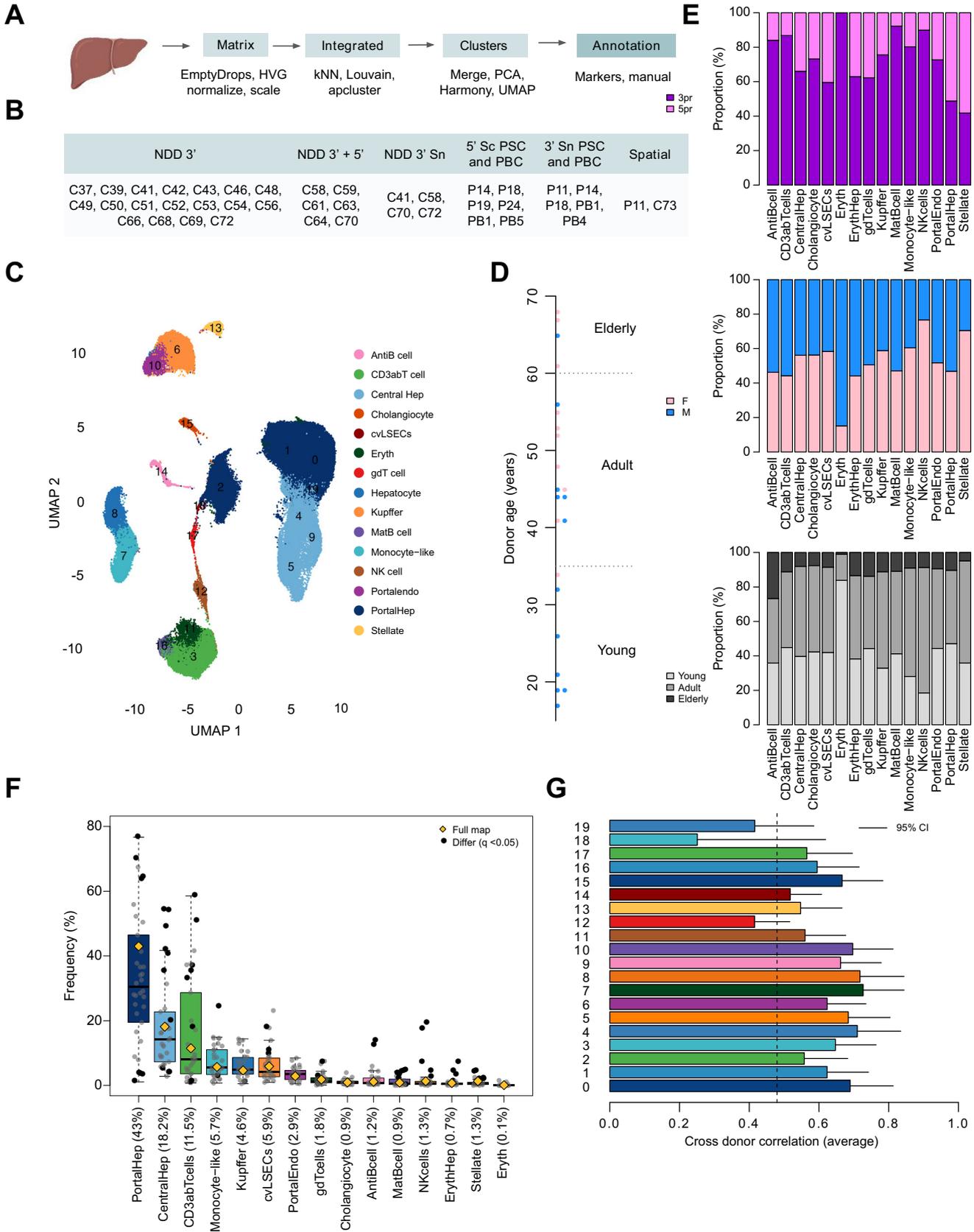
### Visium computational analysis

Visium spatial transcriptomic data was collected for one sample of NDD and one sample of PSC liver. Reads were mapped to the Human genome (GRCh38-2020-A), demultiplexed and quantified according to unique molecular identifiers using spaceranger (v1.1.0). Zonation scores were calculated by first rotating principal components using the base R varimax function to improve their interpretability (see the [supplementary methods](#)). Nanostring GeoMx expression data were deconvolved into cell type composition using Nanostring GeoMx DSP software (v3.0.0.113)<sup>34</sup> and the marker genes from our NDD liver map ([Table S2](#)). Protein validations for spatial transcriptomics were performed via immunofluorescence (IF) ([Fig. S8](#)).

### Flow cytometry and cytokine analysis

Cell suspensions from frozen TLH were stained as previously described<sup>1,35,36</sup> with fluorophore-conjugated monoclonal

# Mapping the healthy and diseased liver



**Fig. 1. 100k single-cell map of NDD human liver reveals conserved cell types across donors.** (A) Single-cell RNA-seq sample collection and analysis workflow (B) Table of samples sequenced in this study. (C) Integrated UMAP, containing 20 clusters annotated to 15 major cell types. (D) Age and sex profiles of the 24 donor livers pink = female, blue = male. (E) Proportion of cells from different technologies, sexes and ages contributing to each cell type of the integrated data. Black points indicate significant enrichment/depletion ( $p < 0.05$ ) based on a hypergeometric test using compositional analysis. There were no significant associations between demographic characteristics and the frequency of different cell types. (F) Proportion of cell types in each sample. The yellow diamond and number in brackets on the x-axis indicates

antibodies detailed in the [supplementary methods](#). Surface markers included CD45, CD206, HLA-DR, and CD68. Intracellular cytokine staining was performed to examine the functional differences in CD68<sup>+</sup> cells that were either CD206<sup>+</sup> or CD206<sup>-</sup> in PSC, PBC and NDD TLH ([supplementary methods](#)). Intracellular secretion of TNF $\alpha$  was examined as previously described.<sup>1,35</sup> Gating strategy is detailed in the [supplementary methods](#) and [Figs S19 and S20](#). Plots were generated and statistical tests were performed in Prism (v10.1.0).

### Histopathology, immunohistochemistry and IF of FFPE samples

To validate *in silico* findings, FFPE tissue sections were employed for histology, immunohistochemistry and IF staining. Disease stage was defined by the Nakanuma score and stage<sup>37</sup> by assessing fibrosis, and bile duct loss. Annotation of fibrotic areas and chronic biliary disease within the liver parenchyma of PSC explant livers were defined based on Masson's trichrome stain and Cytokeratin 7 (CK7) ([Figs S11, S22 and S23](#)).<sup>38</sup> Details and antibodies used are available in the [supplementary methods](#). Analysis of IF experiments and liver immune cell detection was performed on QuPath (v0.4.3) and Prism (v10.1.0), as previously described<sup>39,40</sup> ([dx.doi.org/10.17504/protocols.io.bs6gnhbw](https://dx.doi.org/10.17504/protocols.io.bs6gnhbw); [dx.doi.org/10.17504/protocols.io.e6nvwdmnmwlmk/v1](https://dx.doi.org/10.17504/protocols.io.e6nvwdmnmwlmk/v1)), and plots were generated in Prism (v10.1.0).

## Results

### Global NDD liver map

To generate an expanded reference single-cell liver atlas suitable for identifying PSC-aberrant transcriptomic patterns, we collected single-cell transcriptomes from over 100,000 single cells from 24 different NDD livers with equal representation of males and females spanning a wide age range ([Fig. 1](#)). scRNA-seq was generated using the 10x Chromium with both 3' and 5' chemistries ([Table S1](#); [Fig. 1C](#)). Samples were integrated into a single atlas, which was clustered into 20 coarse-level cell groups. We found that all samples merged together and contributed to almost every cell cluster in our map and that cellular phenotypes were remarkably consistent across demographic factors (Pearson correlation between donors >0.5 in 17/20 clusters) ([Fig. 1G](#); [Table S3](#)). Using compositional analysis (see Methods), we found significant variations in cell type frequencies across samples ([Fig. 1F](#)). However, this is likely due to dissociation and tissue sampling effects rather than donor-specific differences, as significant associations between cell type frequency and donor sex or age were not identified (linear regression Benjamini-Hochberg adjusted *p* value >0.3). These coarse-level cell types were subclustered into a final total of 38 cell states ([Figs S2-5](#)). These included tissue-resident and circulating NK cells, naive and plasma B cells, and an ultra-rare population of mucus-producing cholangiocytes.

### Expanded NDD liver map elucidates human macrophage diversity

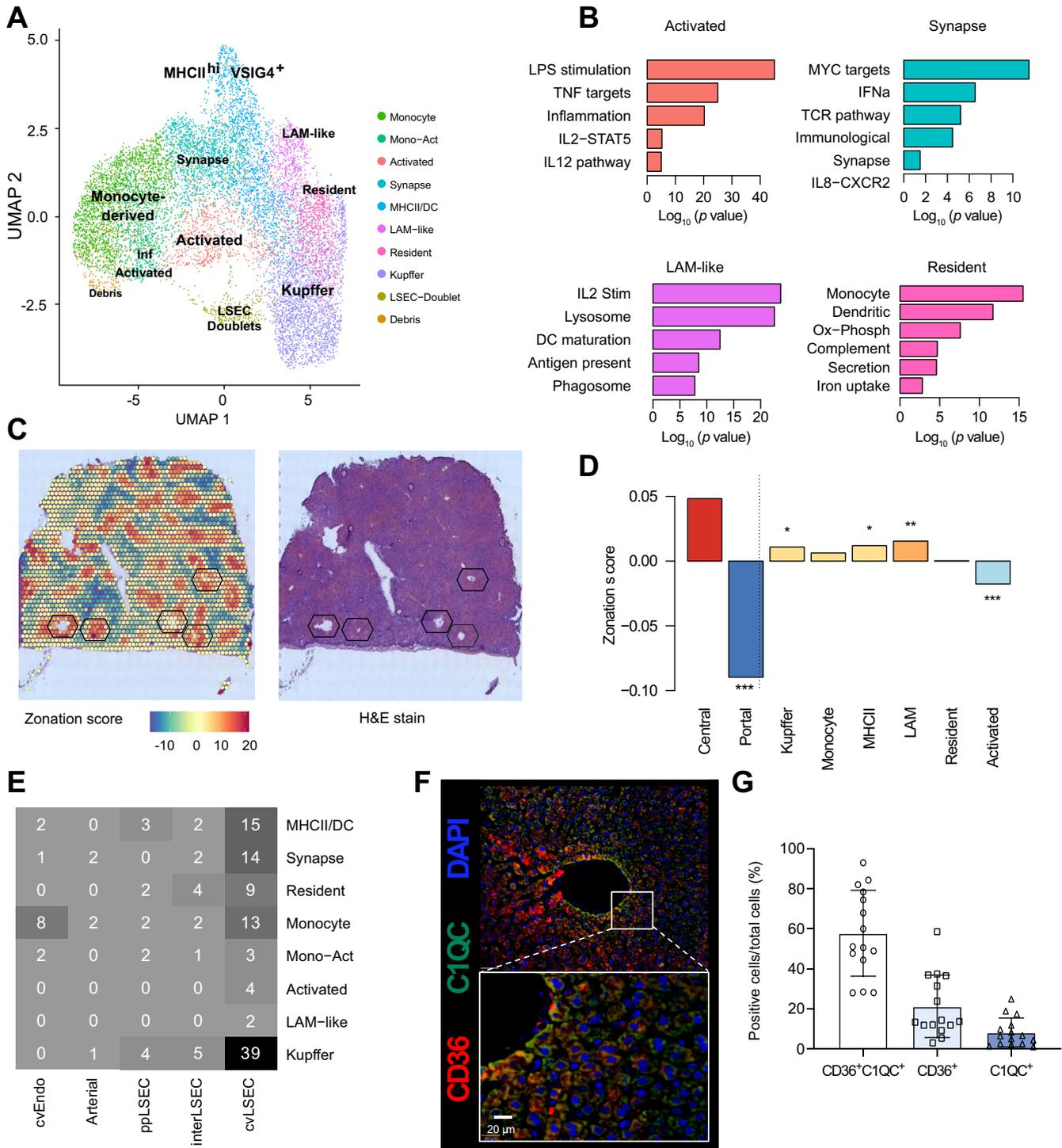
Our original liver atlas, based on five NDD livers, was able to distinguish two different macrophage populations characterized by opposing relationships with inflammation.<sup>1</sup> However, a larger diversity of macrophages has been observed in disease datasets.<sup>3,4</sup> To determine if these macrophage subtypes were present in non-diseased NDD livers, we subclustered our macrophage populations. Subclustering revealed additional macrophage diversity among the 11,127 macrophages of our expanded map ([Fig. 2A](#)). We identified 15 subclusters among the macrophages in our map, several of which exhibited simply different expression in inflammatory (*S100A8*, *S100A9*, *S100A6*, *VCAN*, *LYZ*) or known non-inflammatory (*CD5L*, *MARCO*, *VCAM1*) markers. These clusters were merged together due to their lack of unique marker genes, leaving 10 distinct phenotypes of macrophages that were both conserved across donors and consistent with the macrophage subtypes identified in Williams *et al.*<sup>4</sup> ([Fig. S5](#)). After merging, non-inflammatory Kupffer cells and inflammatory monocyte-derived macrophages accounted for 40% of all macrophages in the NDD liver.

The remaining 60% were spread across eight rarer phenotypes, including LYVE1<sup>+</sup> FOLR2<sup>+</sup> TIMD4<sup>+</sup> non-inflammatory resident macrophages and VSIG4<sup>+</sup> macrophages. The former population was similar to the self-renewing resident cardiac macrophages found in mice<sup>41,42</sup> and was found in 23/24 NDD livers, but was not distinguishable from Kupffer cells in diseased livers. Meanwhile, MHCII<sup>high</sup> VSIG4<sup>+</sup> macrophages were similar to those observed in many mouse tissues,<sup>43</sup> and were found in 23/24 NDD livers and all PSC/PBC livers, as well as in published data on non-alcoholic fatty liver disease (NAFLD)/alcohol-related liver disease (ALD) livers ([Fig. 2](#) and [Fig. S10](#)).

In addition, we identified a novel macrophage subtype characterized by the expression of *PLAC8*, *LST1*, *IFITM3*, *AIF1* and *COTL1* genes, in addition to some inflammatory markers (*FCN1*, *LYZ*, *S100A4*, *S100A8*). This population was found in 23/24 NDD livers but was not distinguishable in PSC/PBC livers. *COTL1* and *LST1* are cytoskeleton-related proteins involved in lamellipodia at immune synapses.<sup>44,45</sup> *CD52*, another top marker of this cluster, is involved in interactions with T cells.<sup>46</sup> Pathway enrichment analysis revealed an upregulation of pathways relating to the interferon (IFN)- $\alpha$  response, TCR signaling, immunological synapses and IL8-CXCR2 signaling ([Fig. 2B](#)). These results suggest that these inflammatory macrophages are specialized in interacting with T cells.

The rarest subtype of macrophages (5%) uniquely expressed *CD9*, *GPNMB*, *TREM2*, *FABP5*, *ACP5*, *PLD3* and *LGMN*, similar to previously described lipid-associated macrophages (LAMs) and cirrhosis scar-associated macrophages,<sup>3,4,47</sup> and are hereafter referred to as LAM-like macrophages. This population was found in 22/24 NDD donor livers and all PSC/PBC livers where scRNA-seq data were collected, and expression of CD9 protein was validated by IF ([Fig. S6D](#)). Pathway analysis revealed

the proportion of the entire map. Black points indicate significant enrichment / depletion based on a hypergeometric test using compositional analysis (see: Methods). There were no significant associations between demographic characteristics and the frequency of different cell-types. (G) Average correlation of the cluster profiles between different donors. The cluster profile is the mean expression of cells within that cluster. Dashed line indicates the average correlation between the cluster profiles of different clusters within the same donor. AntiB cell, antibody secreting B cell; endo, endothelial; F, female; hep, hepatocyte; M, male; NDD, neurologically deceased donor; NK cell, natural killer cell; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis; single-cell RNA-seq, single-cell RNA sequencing.



**Fig. 2. Macrophage diversity revealed by expanded NDD liver map.** (A) 11,127 macrophages were subclustered using Seurat and manually annotated to identify 10 distinct phenotypes. (B) Each subtype was enriched for distinct functional pathways as determined by gene set enrichment analysis. (C) Varimax PCA identifies a component that quantifies the zonation of each spot. Black hexagons indicate pericentral veins of distinct lobules. (D) Using the zonation score across four slices of spatial transcriptomics, we determine non-inflammatory macrophage populations are pericentral while activated macrophages are periportal. Significance obtained from t-test where the t value is computed as the slope divided by the square root of the covariance, and degrees of freedom equal to the number of data points -1, \* $p < 0.01$ , \*\* $p < 0.001$ , \*\*\* $p < 0.0001$ . (E) Number of doublets assigned to each specific pair of subtypes using scmap. (F) Immunofluorescence of LSEC marker CD36 (red) and Kupffer cell marker C1QC (green) confirms presence of colocalized LSECs and Kupffer cells. (G) Percentage of cells over total cells that are co-expressing CD36<sup>+</sup>C1QC<sup>+</sup>, single-positive CD36, and single-positive C1QC. DC, dendritic cell; NDD, neurologically deceased donor; PCA, Principal component analysis; LSEC, liver sinusoidal endothelial cells; LAM-like, lipid-associated macrophage-like; Mono-Act, activated monocyte.

an enrichment of pathways related to the phagosome and lysosome (Fig. 2B).

In addition to these subtypes of macrophages, we observed a type-independent activation pathway in macrophages. This signature was characterized by high expression of *IL1B*, *CD83*, *CXCL2*, *CXCL3*, *NAMPT*, *THBS1*, and *AREG*, as well as genes involved in pathways associated with *in vitro* lipopolysaccharide (LPS) stimulation of macrophages and TNF $\alpha$  signaling (Fig. 2B). This pathway was observed in both monocyte-like macrophages and macrophages lacking any other phenotype signature. This population was found in 20/24 NDD livers, as well as all PSC/PBC livers (PBC samples having twice the number of activated macrophages as PSC samples) but was not observed in NAFLD/ALD livers.

### Examination of cell-cell communication suggests endothelial cells may be involved in the recruitment of myeloid cells in PSC

Previous work has shown that liver sinusoidal endothelial cells (LSECs) are important in recruiting monocytes to the liver after injury. LSECs have been shown to contribute to inducing a Kupffer cell-like phenotype in recruited monocytes upon artificial depletion of native Kupffer cells in mouse models.<sup>48,49</sup> Interestingly, we found a significant number of doublets between macrophages and LSECs in our NDD human liver that expressed macrophage markers (*CD163*, *CD68*, *TIMP1*, and *C1q*) and LSEC markers (*DNASE1L3*, *ENG*, *SPARC*, *CLEC1B*). These doublet cells accounted for 282 cells (3% of all macrophages), which was 10-fold higher than we expected based on the loading of our samples and frequency of macrophages and LSECs in our map (expect: 0.3%,  $p < 10^{-100}$ ). Using scmap,<sup>31</sup> 142/282 doublets were reliably assigned to both a specific macrophage subcluster and a specific endothelial subcluster (Fig. 2E). Of these, the most common pairwise interaction was between Kupffer cells and central venous LSECs. These doublets were also observed in small numbers in PSC samples (Fig. 3A). IF staining confirmed the colocalization of endothelial (CD36) and Kupffer (C1QC) cells at the protein level in the NDD liver (Fig. 2F).

However, monocyte-like macrophages were specifically enriched in doublets with central venous endothelial cells that line the central vein. To confirm that these doublets represent cell-cell interactions, we used CellPhoneDB.<sup>32</sup> Results were filtered to identify interacting proteins involved in cell-cell adherence or close-contact cell-cell communication that were specific to central venous endothelial cells and monocyte-like macrophages (Fig. 2E and Fig. S6A,B). The 10 top interactions after filtering included SELL-PODXL, which is involved in immune cell rolling-adhesion,<sup>50</sup> *ITGAL/SPN-ICAM1*, which is important for tissue infiltration of immune cells,<sup>51</sup> and *SELP-SELP*, a mediator of leukocyte recruitment<sup>52</sup> (Fig. S6B). These results indicate a potential recruitment pathway for monocytes from the blood binding to endothelial cells in the central vein and being recruited into the liver.

### Spatial transcriptomics identifies spatial location of macrophage populations

To add a spatial dimension to our single-cell atlas, we sequenced four consecutive z-stack slices using 10x Genomics Visium spatial transcriptomics (Fig. 2C,D and Fig. S7). Hepatocyte zonation was captured as either the first or second

principal component in each slice. This score was used as a weight to calculate the average zonation of individual genes, which were summed across cell-type markers to determine the average enrichment of specific cell types. These results confirmed our endothelial cell annotations, and indicated that NK, T and B cells were weakly periportal (Fig. S7E-F). Examining macrophage subsets, we found that MHCII<sup>+</sup>, LAM-like macrophages and Kupffer cells were slightly pericentral, while activated macrophages were periportal (Fig. 2D). IF validation revealed a bimodal distribution of VCAM1<sup>+</sup> cells, *i.e.* putative Kupffer cells, with increased frequency in both the periportal and pericentral region; however, only the increase in periportal regions reached statistical significance (Fig. S8).

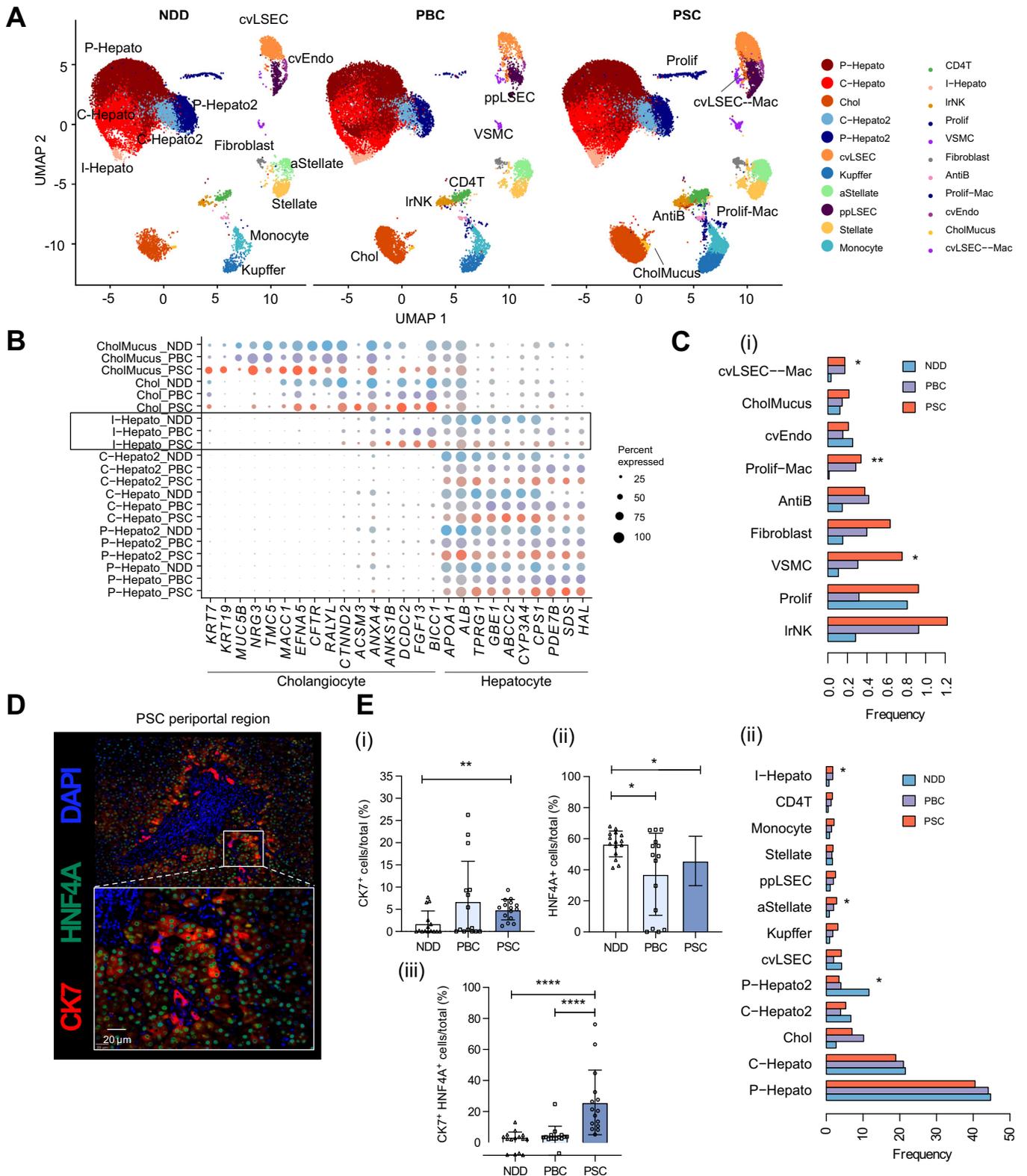
### Cholangiocyte-like hepatocytes are enriched in liver disease

We collected 5' scRNA-seq (Fig. 3) and 3' snRNA-seq (Fig. 4) data from explanted livers from 10 patients with PSC and three patients with PBC. To provide a NDD reference for our 3' snRNA-seq data from PSC and PBC livers, we combined 24,511 single nuclei from four NDD livers from our previously published NDD map<sup>2</sup> with 53,284 nuclei from seven PSC livers, and 20,202 nuclei from two PBC livers. As we observed previously, immune cell populations were poorly captured with snRNA-seq but large numbers of hepatocytes, cholangiocytes, and endothelial cells were observed (Fig. 3A). In comparison to NDD livers, PBC and PSC samples contained fewer periportal hepatocytes (P-Hepato2, 70% loss,  $p = 0.02$ ), a greater number of activated stellate cells (aStellate 3-fold increase,  $p = 0.04$ ), and a greater number of cholangiocytes (2.5-fold increase,  $p = 0.3$ ) (Fig. 3C). Interestingly, we found that interzonal hepatocytes (I-Hepato), hepatocytes that lack pericentral and periportal markers, were significantly increased in PSC and PBC (2.3-fold increase,  $p = 0.04$ ). These hepatocytes upregulate many genes involved in immunological pathways, such as TNF $\alpha$  signaling, IFN $\gamma$  response, and allograft rejection, as well as pathways involved in dedifferentiation, such as KRAS and epithelial-mesenchymal transition signaling (Fig. S9). Many cholangiocyte-related genes, including *KRT7*, *BICC1*, *FGF13*, and *DCDC2*, were upregulated in both PBC and PSC in this hepatocyte population compared to NDD controls (Fig. 3B). In contrast, classical hepatocyte markers, including *ALB*, *CYP3A4*, and *APOA1*, were downregulated. We identified a similar pattern in published scRNA-seq datasets<sup>3</sup> which compared livers affected by ALD and NAFLD to uninjured livers (Fig. S10E). These findings suggest that a shift towards a cholangiocyte-like phenotype is a common feature of disease-associated hepatocytes across many forms of cirrhosis. We directly observed this phenomenon in our PSC livers using IF with HNF4A as a marker of hepatocytes and CK7 as a marker of cholangiocytes. We observed a decrease in HNF4A, an increase in CK7 and an increase in co-expression of the markers in PSC compared to NDD livers (Fig. 3D,E and Fig. S11). This was consistent with the hepatocyte metaplasia observed during disease-scoring of these samples (Figs S22 and S23).

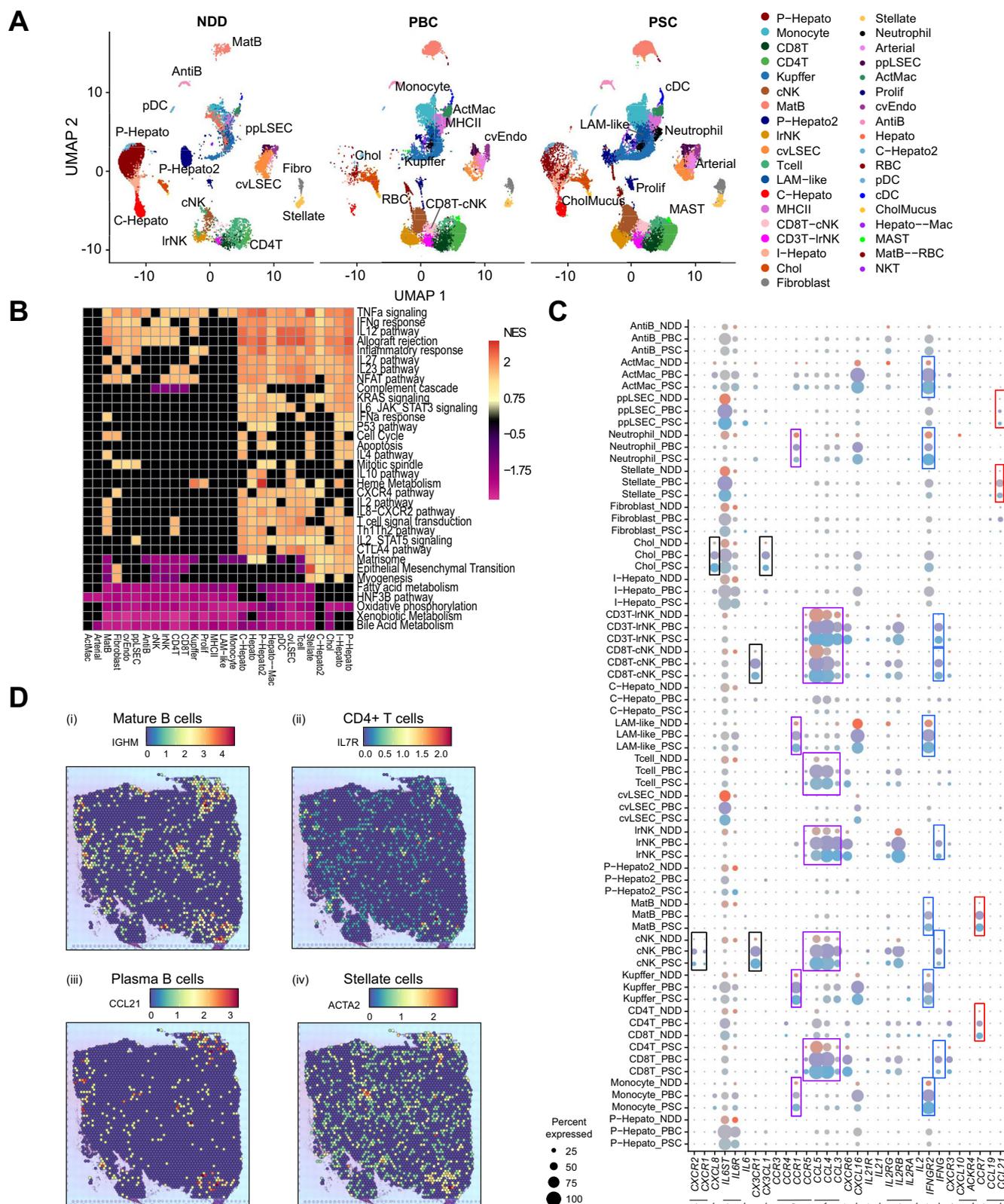
### Immune expansion identified within scar regions

We created a combined 5' scRNA-seq map from 24,007 cells from six livers of our NDD map that were sequenced using 5'

# Mapping the healthy and diseased liver



**Fig. 3. Hepatocyte transdifferentiation in biliary disease.** (A) Integrated map of 24,511 NDD nuclei, 53,284 PSC nuclei, and 20,202 PBC nuclei. (B) Expression of hepatocyte and cholangiocyte markers across hepatocyte and cholangiocyte subtypes. (C) Frequency of each cell type among livers of each aetiology, significance obtained from Wilcoxon rank-sum test across sample: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ; (i) rare cell types making up less than 1% of the total map, (ii) cell types making up >1% of the total map. (D) Representative IF image of a periportal region in the PSC liver. (E) Quantification of (i) CK7<sup>+</sup>, (ii) HNF4A<sup>+</sup>, (iii) double CK7<sup>+</sup>HNF4A<sup>+</sup> cells, significance obtained from Mann-Whitney *U* test across samples: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . NDD, neurologically deceased donor; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis; single-cell RNA-seq.



**Fig. 4. Single-cell RNA-seq reveals immunological complexity of PSC and PBC.** (A) Integrated map of 44,150 cells from PSC, 17,230 cells from PBC, and 24,007 NDD liver cells into a single map. (B) Significantly enriched pathways obtained from gene set enrichment analysis on genes differentially expressed between PSC and NDD livers for each cell type. (C) Expression of ligand-receptor pairs for known cytokines from the KEGG database in PSC, PBC, and NDD cells. Arrows indicate ligand to receptor relationships. (D) Visium spatial transcriptomics of key marker genes of immune populations. (i) Naive B cells, (ii) plasma B cells, (iii-iv) CD4<sup>+</sup>T, (v-vi) ppLSECs, (vii-viii) hepatic stellate cells. NDD, neurologically deceased donor; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis; single-cell RNA-seq.

technology, 44,150 cells from eight PSC livers, and 17,230 cells from two PBC livers (Fig. 4A). We excluded 3' scRNA-seq NDD samples to avoid biasing comparisons with 5' scRNA-seq diseased livers. While hepatocytes, stellate cells and cholangiocytes are well represented in the single-nucleus map, lymphocyte populations were difficult to distinguish. In contrast, our single-cell map revealed the presence of both plasmacytoid and conventional dendritic cells, and a large diversity of lymphocyte populations, including many that were not present in our NDD map (Fig. 4A and Fig. S12B,D). Similar to what has previously been described in fibrotic livers,<sup>53</sup> we observed increases in disease-specific lymphocyte populations including two distinct populations of NKT-like cells (Fig. S12D), one that expressed liver-resident NK cell (lrNK) markers (*EOMES*, *CMC1*, *KLRC1*, *KLRF1*, *XLCL1*)<sup>54</sup> and CD3<sup>+</sup> T cell markers (*CD3D*, *CD3E*, *TRAC*, *TRBC2*), as well as another population that expressed circulating NK cell (cNK) markers (*FCGR3A*, *GZMY*, *GZMK*, *TBX21*, *XLCL2*)<sup>55</sup> and CD8<sup>+</sup> T cell markers (*CD8A*, *CD8B*, *CD3D*, *CD3E*). DoubletFinder<sup>2</sup> determined that these populations are not doublets (estimated 3% doublets for CD3T-lrNK, 10% for CD8T-cNK). These populations were predominantly found in PBC (965 cells, 6%) and PSC (2,879 cells, 6%) livers, but not in NDD livers (<100 cells, 0.2%) nor the published NAFLD/ALD data.<sup>3</sup> Comparing these populations to the most similar NK and T cells, we found that pro-inflammatory CCL3<sup>56</sup> was downregulated in the CD8T-cNK population (log<sub>2</sub> fold-change [FC] -0.4,  $p < 10^{-10}$ ), while anti-inflammatory KLF2<sup>57</sup> was upregulated (log<sub>2</sub>FC 0.5,  $p < 10^{-100}$ ). In contrast, pro-inflammatory cytokines CCL5 (log<sub>2</sub>FC 0.75,  $p < 10^{-100}$ ) and IL23 (log<sub>2</sub>FC 0.39,  $p < 10^{-50}$ ), as well as TRDV1 and TRGV3, were upregulated in the CD3T-lrNK population, suggesting these may be  $\gamma\delta$ T cells.

Differential expression between PSC and NDD livers revealed significant upregulation of many inflammatory pathways including: IL12, IL23, IL27, IFN $\gamma$  and TNF $\alpha$  pathways in Kupffer cells, T and NK cells, and TNF $\alpha$ , IFN $\gamma$ , IL2, KRAS and epithelial-mesenchymal transition pathways in hepatocytes, cholangiocytes, and hepatic stellate cells (Fig. 4B). Most of these pathways were also upregulated in PBC, but the upregulation of IL12 signaling in Kupffer cells and NK cells was only seen in PSC. Moreover, PSC-derived NK cells expressed more genes involved in the IFN $\gamma$  response than PBC (Fig. S13).

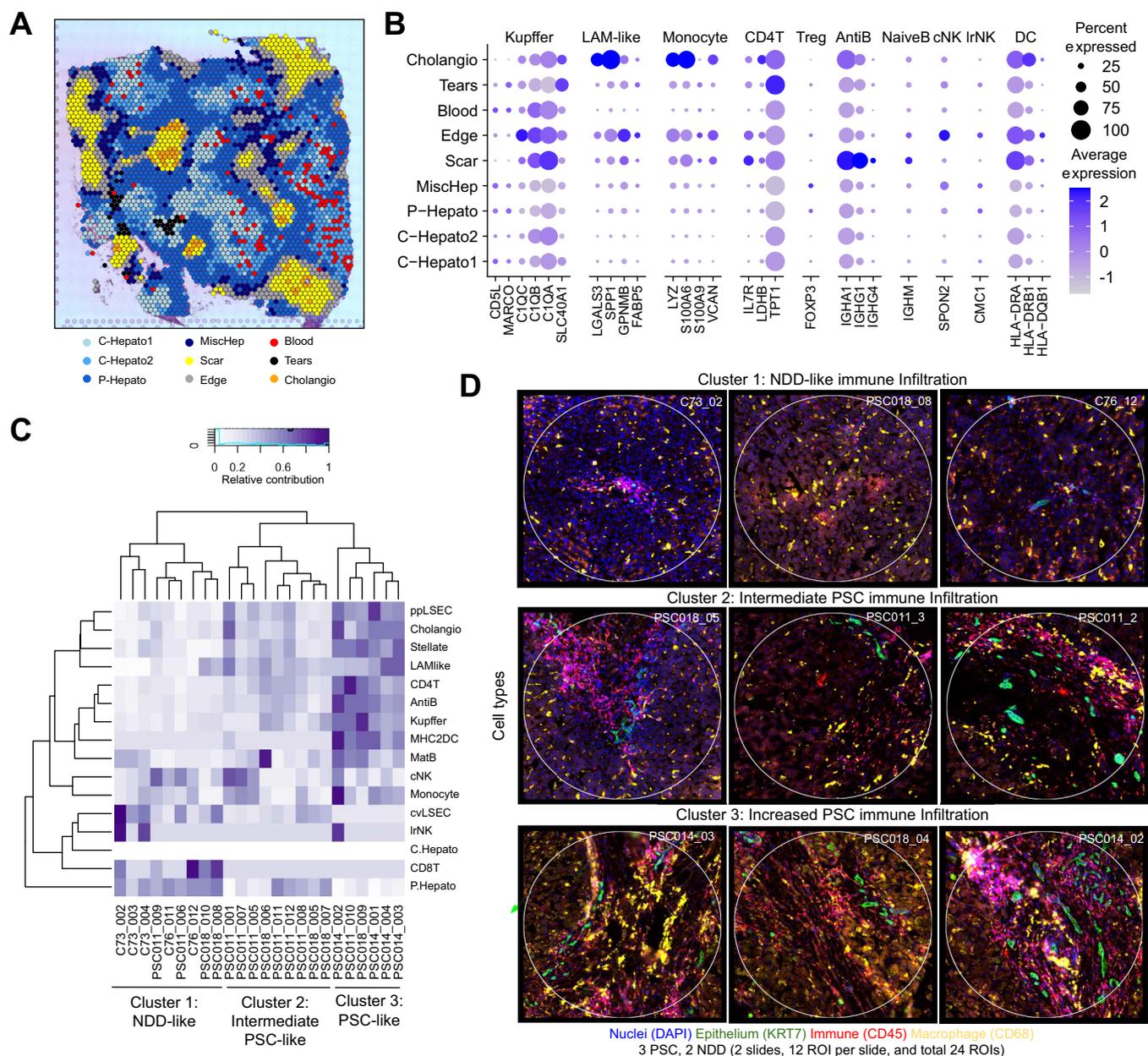
To examine inflammation-related cell-cell signaling, we examined the expression of gene pairs from the KEGG cytokine pathway (hsa04060). We found evidence of interactions between CCR7<sup>+</sup> CD4<sup>+</sup> T cells and CCL19<sup>+</sup> stellate or CCL21<sup>+</sup> periportal LSECs in fibrotic regions in PSC and PBC (Fig. 4C). The colocalization of *CCL19*, *CCL21* and *IL7R* (CD4<sup>+</sup> T cells) was confirmed in our Visium spatial transcriptomics data (Pearson correlation 0.2,  $p < 10^{-30}$ , Fig. 4D), and across multiple replicates of spatial transcriptomics using the Nanostring GeoMx digital spatial profiling platform (Fig. 5C,D and Fig. S14). Using Nanostring, we observed the colocalization of CD4<sup>+</sup> T cells, stellate cells, periportal LSECs and mature B cells in PSC. However, this was only the case for the examined advanced diseased regions. Large-scale heterogeneity was observed between different regions of interest from the same original liver samples.

## Macrophage dysfunction in PSC livers

Unlike the lymphocyte populations, which upregulate inflammatory pathways in PSC, macrophages exhibit increased expression of inhibitory signaling molecules in PSC, including *KLF2*,<sup>57</sup> *OTULINL*, and *IL27RA*. *IL27RA* has been shown to suppress cytokine production in macrophages<sup>58,59</sup> and is upregulated in PSC Kupffer cells (log<sub>2</sub>FC 2.1, FDR 0.03), and in both PSC and PBC LAM-like macrophages (log<sub>2</sub>FC 2.3, FDR 0.001). *OTULINL* is a potent negative regulator of macrophage activation, and deficiency causes autoimmune disease that can be rescued with anti-TNF therapy.<sup>60</sup> We observed significant upregulation of *OTULINL* in all macrophage populations in PSC and PBC livers (log<sub>2</sub>FC 3.4-5.4, FDR <0.001). Monocyte-like macrophages adopted a fibrosis-associated phenotype with the upregulation of *LGALS3*, *SPP1*<sup>3</sup> and *ADA2*,<sup>61</sup> of these, only *ADA2* was upregulated in PBC monocyte-like macrophages. In addition, we observed that TGF- $\beta$ , a potent suppressor of macrophage function,<sup>62</sup> colocalized with these cells within those regions (Fig. S15). We confirmed this scar-associated phenotype of monocyte-like macrophages using our spatial transcriptomics data, which indicated an enrichment of these cells around cholangiocytes in the center of fibrotic regions containing concentric periductal fibrosis (Fig. 5A,B and Fig. S14), whereas Kupffer cells were localized outside of the scar regions (Fig. S15D).

Subclustering of our scRNA-seq map revealed greater macrophage diversity among PSC macrophages (Fig. S16A) with 25 clusters containing at least 50 cells derived from PSC samples vs. 13 clusters from PBC and only seven clusters from NDD macrophages. Furthermore, we consider nine of the clusters as PSC-enriched, as >90% of the cells within the cluster originated from PSC livers (Fig. S16B). In contrast, we observed no PBC-specific cluster; all 13 clusters containing significant numbers of PBC cells also included at least 30% PSC cells. The PSC-specific clusters were predominantly monocyte-like (5/9), whereas macrophages shared between PBC and PSC included 4/13 clusters expressing Kupffer cell markers (*MARCO*, *CD5L*, *VCAM1*) and 2/13 highly expressing MHCII components. Using pathway analysis of the marker genes of each of these subclusters (Fig. S16E), we identified downregulation of 'interferon signaling' in 7/9 PSC-specific clusters and 2/7 clusters shared between PSC and NDD, but only 1/13 clusters containing PBC cells. Similarly, MHCII presentation was only downregulated in the PSC-specific clusters (3/9) but upregulated in clusters shared with PBC (1/13) and NDD clusters (2/7).

Flow cytometry of primary patient macrophages revealed a significant depletion of CD68<sup>+</sup> cells relative to CD45<sup>+</sup> (PTPRC) cells in PSC, which was confirmed with *in silico* gating of our single-cell data (Fig. 6A-C and Fig. S17-20). Additionally, as previously described in PSC,<sup>20</sup> we observed increased percentages of CD206<sup>+</sup> macrophages in PSC in comparison to NDD (Fig. 6D-E). Intracellular cytokine staining stimulation assays indicated a significantly reduced capacity of PSC macrophages to secrete TNF $\alpha$  in response to LPS and IFN $\gamma$  stimulation, in comparison to NDD and PBC macrophages (Fig. 6F,G). However, when we exposed NDD-derived myeloid populations to recombinant TGF- $\beta$  (Fig. S21), we did not



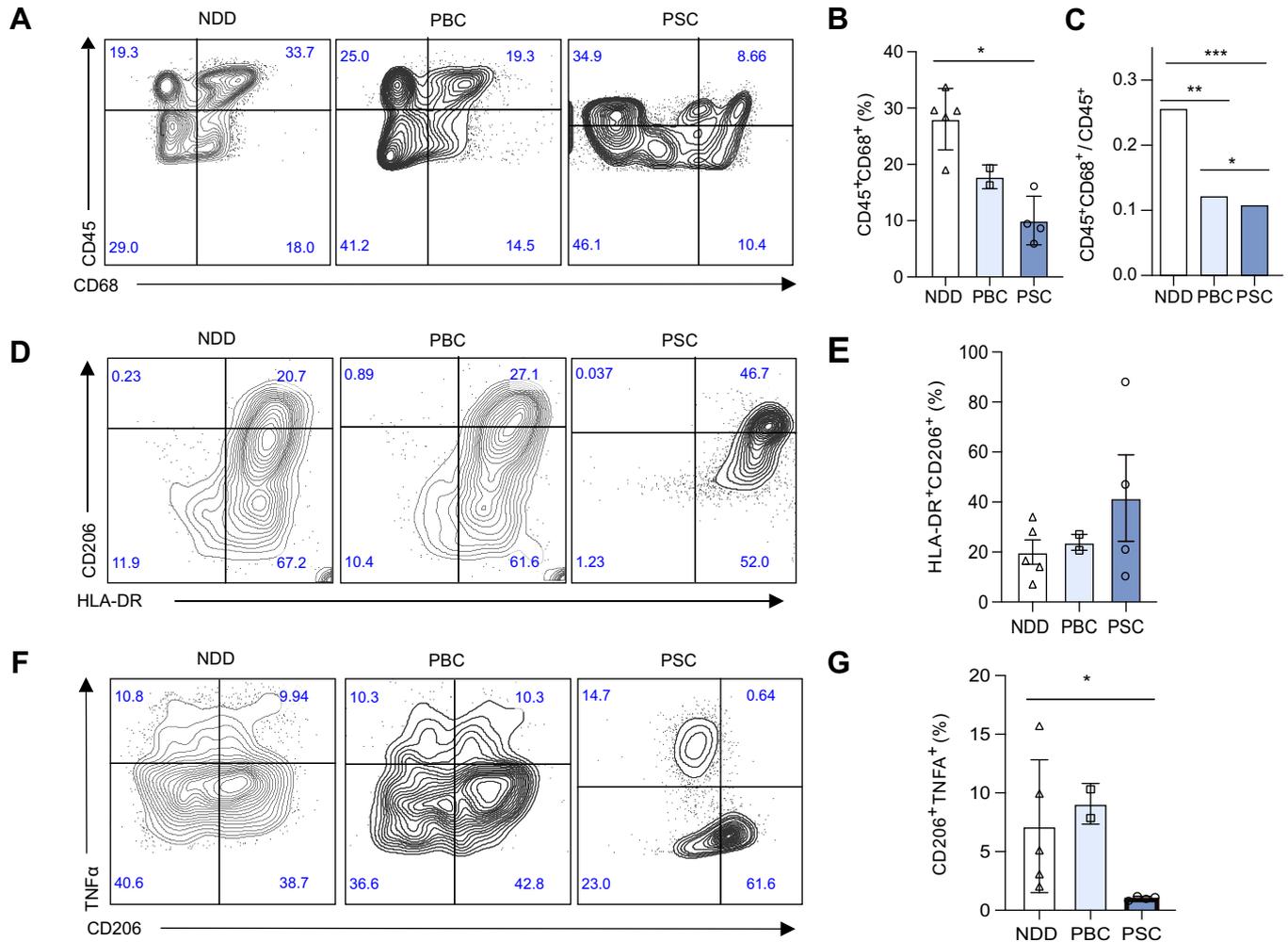
**Fig. 5. Enrichment of immune cells in PSC (fibrotic areas).** (A) Visium spatial transcriptomics was clustered and identified layers of distinct regions around fibrotic areas in PSC. (B) Expression of marker genes for each lymphocyte and macrophage subtype from the PSC and NDD single-cell maps in the PSC Visium data. LAMs and monocytes were enriched in the core of scars near cholangiocytes and periphery of fibrotic areas, while B cells, T cells, and antigen presenting cells were enriched in the center of fibrotic regions, whereas NK cells were present at the periphery. (C) Deconvolution of Nanostring digital expression using cell-type markers from the NDD and PSC single-cell maps. (D) Images of typical example regions for each cluster identified in Nanostring. Nuclei are shown in blue, cholangiocyte epithelium in green, leukocytes in red, and tissue-resident macrophages in yellow. Each circle is 660  $\mu$ m in diameter. Complete NanoString ROIs are in Fig. S14. NDD, neurologically deceased donor; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis.

observe blunted LPS responsiveness in the CD206<sup>+</sup> HLA-DR<sup>+</sup> population.

## Discussion

A major challenge in understanding the cellular ecosystem of PSC is the lack of comprehensive maps of the parenchymal and non-parenchymal cells in the PSC liver at single-cell resolution. Previous transcriptomic studies of the cellular landscape of PSC have employed bulk RNA-seq on liver

tissue<sup>12,13,18,63</sup> and/or scRNA-seq on sorted immune cell populations<sup>10</sup> and have uncovered important aspects of the immunobiology of PSC. Namely, these studies have identified the potential involvement of T helper 17 cells and B cells in PSC pathogenesis,<sup>10,14,15,21</sup> as well as uncovering distinct ductular and fibrosis-associated signatures.<sup>18,63</sup> Taken together, these works imply that there is a complex cellular interplay driving PSC, providing a rationale for our in-depth examination of parenchymal and non-parenchymal cell populations in the PSC liver.



**Fig. 6. PSC macrophages are dysfunctional in cytokine secretion.** (A) Representative flow plots of CD45<sup>+</sup> (PTPRC<sup>+</sup>) and CD68<sup>+</sup> cells in PSC, PBC and NDD total liver homogenate. (B) PSC samples contained significantly fewer double positive cells ( $p < 0.05$ ) in comparison to NDD and PBC. Percentage represents intact cells that are CD68<sup>+</sup>CD45<sup>+</sup> in the PSC, PBC and NDD samples. (C) *In silico* gating of scRNA-seq replicates decrease in CD68<sup>+</sup>CD45<sup>+</sup>/PTPRC<sup>+</sup> cells ( $p = 0.01$ ). (D,E) Flow gating of CD45<sup>+</sup>CD68<sup>+</sup> cells reveals increase in CD206<sup>+</sup> subset in PSC ( $p > 0.05$ ). (F,G) Following LPS stimulation, CD206<sup>+</sup> PSC macrophages (CD45<sup>+</sup>CD68<sup>+</sup>) exhibit reduced production of TNF $\alpha$  ( $p < 0.05$ ). Statistical significance for flow cytometry and intracellular cytokine staining is evaluated using a non-parametric Mann-Whitney  $U$  test, \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ , error bars represent the SEM. Full gating strategy and unstimulated controls are shown in Figs S19 and S20. NDD, neurologically deceased donor; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis.

Our multimodal study centered on linking transcriptomic signatures to spatial positioning and functional validation. Specifically, transitioning cholangiocyte-like hepatocytes were identified in both sc- and snRNA-seq, as well as spatial transcriptomics, wherein the transdifferentiating population was shown to localize around the perimeter of PSC scars. Through IF, we observed that this population is significantly enriched in the PSC liver. Additionally, spatial transcriptomics revealed the distinct and exclusive localization of inflammatory and non-inflammatory myeloid populations in relation to PSC scars, and demonstrated the presence of expanded immune populations within these scars, corroborating the increased immunological diversity shown by scRNA-seq. We observed the enrichment of immunosuppressive transcripts in PSC-specific monocytes but an expansion of activated macrophages in PBC, which is reflected in functional validation experiments demonstrating dysfunction in PSC myeloid populations. Beyond understanding the cellular identities in

the PSC niche, we also examined cellular locations to better link cell states to the degree of histological damage in the liver. Previously, Chung *et al.*<sup>53</sup> found distinct fibrosis-enriched immune populations, a finding that we further refined with the deconvolution of 10x Visium data with our PSC scRNA-seq map. This is supplemented by our examination of the linkage between the heterogeneity of disease through the deconvolution of Nanostring GeoMx DSP data, which revealed the enrichment of immune populations in progressed fibrosis.

In the NDD liver, we uncovered 38 distinct cell types and cell states including a pericentrally located myeloid population similar to the previously described LAMs,<sup>4,47</sup> as well as myeloid subpopulations associated with immunological synapses and a consistent activation signature (Fig. 2). In the PSC liver, we identified a further seven disease-associated cellular states including NKT-like cells, neutrophils, dendritic cells, and a subpopulation of hepatocytes (Figs 3 and 4; Table S6). These

immune populations are likely a result of *in situ* expansion and recruitment from the blood into the diseased liver tissue.<sup>11,13,15</sup> As previously reported for healthy liver tissue, immune sub-populations could only be identified in scRNA-seq (Table S7), whereas the hepatocyte subtype was only detected in snRNA-seq (Table S8).<sup>2</sup> Many of the immune populations were colocalized within fibrotic ‘scar’ regions in PSC, including CD4<sup>+</sup> T cells and antibody-secreting B cells, which is consistent with bulk RNA-seq findings of enrichment of B lymphocytes and FOXP3<sup>+</sup>CD4<sup>+</sup> T cells in the PSC liver.<sup>64</sup>

In contrast, we observed that disease-associated hepatocytes were specifically located in the border regions surrounding the fibrotic ‘scar’ regions in PSC (Fig. 3 and Figs S22, 23). These hepatocytes expressed many markers associated with cholangiocytes suggesting they may be transforming into cholangiocytes. Previous studies have demonstrated hepatocytes undergo this transition in the absence of functional bile ducts in a TGF- $\beta$ -mediated manner.<sup>65,66</sup> Consistent with these results, TGF- $\beta$  was expressed in the PSC fibrotic regions, particularly in scars with transitioning hepatocyte edges (Fig. S13).

We examined the functional capacity of CD206<sup>+</sup> myeloid cells in PSC and observed suppressed TNF $\alpha$  cytokine secretion and reduced inflammatory potential by CD206<sup>+</sup> Kupffer cell-like myeloid cells following stimulation (Fig. 6). A similar cytokine secretion dysfunction has been reported in inflammatory bowel disease, a common comorbidity of PSC.<sup>67,68</sup> This suggests a possible mechanism by which macrophage dysfunction may underlie PSC and inflammatory bowel disease pathogenesis.<sup>13,16,20</sup> We examined the impact of TGF- $\beta$  incubation on

NDD-derived myeloid populations (Fig. S21) and could not recapitulate the hyporesponsiveness of CD206<sup>+</sup> and HLA-DR<sup>+</sup> macrophages in PSC. This further reinforces the complex and multifactorial pathogenesis of PSC, given that spatial transcriptomic signatures demonstrate significant immune recruitment to the PSC scars, such as plasma B cells, inflammatory macrophages and T cells. This demonstrates the importance of future work focusing on early-stage disease to identify early actors in PSC development.

Like most studies on PSC,<sup>10,11,18</sup> much of our data is limited to a single region of each end-stage explanted liver, which was compared to non-diseased livers. However, additional non-PSC disease datasets will serve to further refine these findings. Given the slow progressive nature of PSC development, it remains to be determined whether the findings we report here are PSC-specific and occur during disease pathogenesis, or whether they represent general features of late-stage cholestatic liver disease. In addition, our NanoString spatial transcriptomic data revealed substantial heterogeneity in disease phenotype across relatively small regions of these livers (Fig. 5). Thus, future work will expand this map to cover the temporal and spatial heterogeneity of disease. Our results demonstrate the power of combining multiple high-resolution gene expression platforms including 3’ single-cell, 5’ single-cell, single-nucleus and spatial transcriptomics to fully elucidate the interactions between the tissue microenvironment, cellular identity, and immunological function underlying a complex inflammatory disease. Taken together, this effort provides a framework for investigating the drivers of PSC and uncovering precision medicine targets for immunomodulatory and cell-based therapies.

## Affiliations

<sup>1</sup>Ajmera Transplant Centre, Toronto General Research Institute, University Health Network, Toronto, ON, M5G 2C4, Canada; <sup>2</sup>Department of Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, N6A 5C1, Canada; <sup>3</sup>Department of Computer Science, University of Western Ontario, London, ON, N6A 3K7, Canada; <sup>4</sup>Department of Immunology, University of Toronto, Toronto, ON, M5S 1A8, Canada; <sup>5</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, M5G 1L7, Canada; <sup>6</sup>Division of Gastroenterology, Hepatology and Nutrition, The Hospital for Sick Children, Toronto, ON, M5G 1X8, Canada; <sup>7</sup>Toronto Centre for Liver Disease, University Health Network, Toronto, ON, M5G 2C4, Canada; <sup>8</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, M5T 3M6, Canada; <sup>9</sup>The Donnelly Centre, University of Toronto, Toronto, ON, M5S 3E1, Canada

## Abbreviations

ALD, alcohol-related liver disease; cNK, circulating NK; FDR, false discovery rate; IF, immunofluorescence; IFN, interferon; LAM, lipid-associated macrophage; LPS, lipopolysaccharides; IrNK, liver-resident NK; LSECs, liver sinusoidal endothelial cells; NDD, neurologically deceased donor; NAFLD, non-alcoholic fatty liver disease; NK, natural killer; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis; scRNA-seq, single-cell RNA sequencing; snRNA-seq, single-nucleus RNA sequencing.

## Financial support

Funding provided in part through the UHN foundation; the Chan Zuckerberg Initiative DAF (CZF2019-002429), an advised fund of Silicon Valley Community Foundation; an early researcher award from the Government of Ontario (ERA19-15-210); funds from the University of Toronto’s Medicine by Design initiative, which receives funding from the Canada First Research Excellence Fund (CFREF) to SAM, GDB and IDM; by the NRNB (US National Institutes of Health, grant P41 GM103504) to GDB and by the UHN Foundation. DN has received graduate fellowships from CIHR (CGS-D). CTP has received postdoctoral funds from the Canadian Network on Hepatitis C (CanHepC). CanHepC is funded by a joint initiative of the Canadian Institutes of Health Research (CIHR) (NHC-42832) and the Public Health Agency of Canada (PHAC). CTP received postdoctoral funds from PSC Partners Seeking a Cure Canada which was matched by the UHN Foundation. Aspects of this work were performed in collaboration with the Canadian Autoimmunity Standardization Core (CIHR-HUI-159423). This work was funded in part by CIHR grants PJT162298 (IDM) and PJT162098 and HIT168002 (SAM). SAM holds a CRC Tier 2 in Liver Immunobiology (CRC-2020-00318).

## Conflict of interest

The authors of this study declare that they do not have any conflict of interest. Please refer to the accompanying ICMJE disclosure forms for further details.

## Authors’ contributions

S.A.M., G.D.B. and I.D.M. designed the study. T.S.L., D.N., X.-Z.M., B.K.G., L.L. and P.L. conducted experiments. E.W., B.A., S.M., A.R., and G.H. recruited and consented patients. E.W. S.M., B.S., G.H.M. and I.D.M. harvested the surgical specimen. D.N., C.T.P., L.L., D.C., and S.C., X.-Z.M., and J.M. expedited the processing of the harvested surgical specimens for storage and experiments. C.T. interpreted the histology. T.S.A., D.N., P.L., L.L., S.C., D.C., B.H., A.G., A.G., A.R., G.M.H., S.A.M., G.D.B., and I.D.M. analyzed and interpreted the data. T.S.A., D.N., S.A.M., G.D.B. and I.D.M. prepared the manuscript with critical revision from all authors. All authors reviewed this manuscript.

## Data availability statement

The NDD liver map data and the spatial transcriptomics data are available for exploration with our interaction shiny tools (see links below). Seurat objects and raw count matrices for all of these data and the PSC sc- and sn-RNA-seq data have been uploaded to CELLxGENE Discover and can be found at: <https://cellxgene.cziscience.com/collections/0c8a364b-97b5-4cc8-a593-23c38c6f0ac5>. CELLxGENE Discover allows users to visually explore patterns of gene expression and corresponding clinical metadata using an interactive interface, including differences between healthy and disease samples. The NDD spatial data is currently available at GSE240429, NDD sc- and sn-RNA-seq raw data and raw count matrices are available at GSE243977, previously published NDD liver data included

in this study can be found at GSE115469<sup>1</sup> and GSE185477,<sup>2</sup> and incorporated data from Ramachandran *et al.*<sup>3</sup> can be found at GSE136103. The PSC spatial data is currently available at GSE245620, and the PSC and PBC sc- and snRNA-seq raw count matrices are available at GSE247128. Scripts used to analyze these data are available on our GitHub: <https://github.com/tallulandrews/LiverMap2.0>.

NDD Map:

<https://macparlandlab.shinyapps.io/generalmap/>  
<https://macparlandlab.shinyapps.io/macrophages/>  
<https://macparlandlab.shinyapps.io/lymphocytes/>  
<https://macparlandlab.shinyapps.io/endothelial/>  
<https://macparlandlab.shinyapps.io/cholangiocytes/>  
[https://macparlandlab.shinyapps.io/b\\_cells/](https://macparlandlab.shinyapps.io/b_cells/)  
<https://macparlandlab.shinyapps.io/stellate/>

PSC, PBC and NDD Maps:

[https://macparlandlab.shinyapps.io/shiny\\_sc/](https://macparlandlab.shinyapps.io/shiny_sc/)  
[https://macparlandlab.shinyapps.io/shiny\\_sn/](https://macparlandlab.shinyapps.io/shiny_sn/)

Spatial NDD:

<https://macparlandlab.shinyapps.io/healthylivermapspatialgui/>

Spatial PSC:

[https://macparlandlab.shinyapps.io/shinypsc\\_spatialappgui/](https://macparlandlab.shinyapps.io/shinypsc_spatialappgui/)

### Acknowledgements

This publication is part of the Human Cell Atlas – [www.humancellatlas.org/publications/](http://www.humancellatlas.org/publications/). The authors acknowledge the Princess Margaret Genomics Centre, the Pathology Research Program and the Advanced Optical Microscopy Facility at University Health Network for their support and services. The graphical abstract and figures were created with [Biorender.com](https://biorender.com). We would like to acknowledge PSC Partners Seeking a Cure Canada and Mary Vyas for patient partner participation in this project.

### Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhep.2023.12.023>.

### References

*Author names in bold designate shared co-first authorship*

- [1] **MacParland SA, Liu JC, Ma X-Z**, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun* 2018;9:4383.
- [2] **Andrews TS, Atif J, Liu JC**, et al. Single-cell, single-nucleus, and spatial RNA sequencing of the human liver identifies cholangiocyte and mesenchymal heterogeneity. *Hepatol Commun* 2022;6:821–840.
- [3] Ramachandran P, Dobie R, Wilson-Kanamori JR, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 2019;575:512–518.
- [4] Williams M, Bonnardel J, Haest B, et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* 2022;185:379–396.e38.
- [5] Sharma A, Seow JJW, Dutertre C-A, et al. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. *Cell* 2020;183:377–394.e21.
- [6] Kolodziejczyk AA, Federici S, Zmora N, et al. Acute liver failure is regulated by MYC- and microbiome-dependent programs. *Nat Med* 2020;26:1899–1911.
- [7] Atif J, Thoeni C, Bader GD. Unraveling the complexity of liver disease one cell at a time. *Semin Liver Dis* 2022;42:250–270.
- [8] Cheng ML, Nakib D, Perciani CT. The immune niche of the liver. *Clin Sci* 2021;135:2445–2466.
- [9] Karlsen TH, Folseraas T, Thorburn D. Primary sclerosing cholangitis - a comprehensive review. *J Hepatol* 2017;67:1298–1323.
- [10] Poch T, Krause J, Casar C, et al. Single-cell atlas of hepatic T cells reveals expansion of liver-resident naive-like CD4+ T cells in primary sclerosing cholangitis. *J Hepatol* 2021;75:414–423.
- [11] Müller A-L, Casar C, Preti M, et al. Inflammatory type 2 conventional dendritic cells contribute to murine and human cholangitis. *J Hepatol* 2022;77(6):1532–1544.
- [12] Pratt HE, Wu T, Elhajjaj S. Beyond genome-wide association studies: investigating the role of noncoding regulatory elements in primary sclerosing cholangitis. *Hepatol Commun* 2023;7.
- [13] Zimmer CL, von Seth E, Buggert M, et al. A biliary immune landscape map of primary sclerosing cholangitis reveals a dominant network of neutrophils and tissue-resident T cells. *Sci Transl Med* 2021;13.
- [14] Chung BK, Henriksen EKK, Jørgensen KK. Gut and liver B cells of common clonal origin in primary sclerosing cholangitis-inflammatory bowel disease. *Hepatol Commun* 2018;2:956–967.
- [15] Henriksen EKK, Jørgensen KK, Kaveh F, et al. Gut and liver T-cells of common clonal origin in primary sclerosing cholangitis-inflammatory bowel disease. *J Hepatol* 2017;66:116–122.
- [16] Gucciardi ME, Trussoni CE, Krishnan A, et al. Macrophages contribute to the pathogenesis of sclerosing cholangitis in mice. *J Hepatol* 2018;69:676–686.
- [17] Reich M, Spomer L, Klindt C, et al. Downregulation of TGR5 (GPBAR1) in biliary epithelial cells contributes to the pathogenesis of sclerosing cholangitis. *J Hepatol* 2021;75:634–646.
- [18] Govaere O, Cockell S, Van Haele M, et al. High-throughput sequencing identifies aetiology-dependent differences in ductular reaction in human chronic liver disease. *J Pathol* 2019;248:66–76.
- [19] Wu S, Cao Y, Lu H. Aberrant peribiliary gland niche exacerbates fibrosis in primary sclerosing cholangitis and a potential therapeutic strategy. *Biomed Pharmacother* 2022;153:113512.
- [20] Chen Y-Y, Arndtz K, Webb G. Intrahepatic macrophage populations in the pathophysiology of primary sclerosing cholangitis. *JHEP Rep* 2019;1:369–376.
- [21] Kunzmann LK, Schoknecht T, Poch T, et al. Monocytes as potential mediators of pathogen-induced Th17 differentiation in patients with primary sclerosing cholangitis (PSC). *Hepatology* 2020;72(4):1310–1326.
- [22] Greenman R, Segal-Salto M, Barashi N, et al. CCL24 regulates biliary inflammation and fibrosis in primary sclerosing cholangitis. *JCI Insight* 2023;8.
- [23] MacParland S, Ma X-Z, Manuel J. Human liver caudate lobe dissociation for ScRNA-seq v2. 2018.
- [24] Nakib D, Perciani C, Chung S. Human liver tissue storage methods for multiomic applications v1. 2022.
- [25] Slyper M, Porter CBM, Ashenberg O, et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* 2020;26:792–802.
- [26] Lun ATL, Riesenfeld S, Andrews T, Participants in the 1st Human Cell Atlas Jamboree. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* 2019;20:63.
- [27] Stuart T, Butler A, Hoffman P. Comprehensive integration of single-cell data. *Cell* 2019;177:1888–1902.e21.
- [28] Korsunsky I, Millard N, Fan J. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019;16:1289–1296.
- [29] Frey BJ, Dueck D. Clustering by passing messages between data points. *Science* 2007;315:972–976.
- [30] Meilă M. Comparing clusterings by the variation of information. In: Schölkopf B, Warmuth MK, editors. *Learning theory and kernel machines*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 173–187.
- [31] Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. *Nat Methods* 2018;15:359–362.
- [32] Efremova M, Vento-Tormo M, Teichmann SA. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 2020;15:1484–1506.
- [33] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–140.
- [34] Hernandez S, Lazzano R, Serrano A. Challenges and opportunities for immunoprofiling using a spatial high-plex technology: the NanoString GeoMx<sup>®</sup> digital spatial profiler. *Front Oncol* 2022;12:890410.
- [35] MacParland SA, Tsoi KM, Ouyang B, et al. Phenotype determines nanoparticle uptake by human macrophages from liver and blood. *ACS Nano* 2017;11:2428–2443.
- [36] Tsoi KM, MacParland SA, Ma X-Z, et al. Mechanism of hard-nanomaterial clearance by the liver. *Nat Mater* 2016;15:1212–1221.
- [37] Nakanuma Y, Zen Y, Harada K, et al. Application of a new histological staging and grading system for primary biliary cirrhosis to liver biopsy specimens: interobserver agreement. *Pathol Int* 2010;60:167–174.
- [38] Sarcognato S, Sacchi D, Grillo F. Autoimmune biliary diseases: primary biliary cholangitis and primary sclerosing cholangitis. *Pathologica* 2021;113:170–184.
- [39] Wang X, Perciani C, Ma X-Z. Qupath digital quantification of liver immune cells v1. 2021.
- [40] Lumanto P, Nakib D, Wang S. QuPath immunofluorescence cell detection and Co-localization protocol. 2024.

- [41] Dick SA, Wong A, Hamidzada H, et al. Three tissue resident macrophage subsets coexist across organs with conserved origins and life cycles. *Sci Immunol* 2022;7:eabf7777.
- [42] Dick SA, Macklin JA, Nejat S, et al. Self-renewing resident cardiac macrophages limit adverse remodeling following myocardial infarction. *Nat Immunol* 2019;20:29–39.
- [43] Chakarov S, Lim HY, Tan L, et al. Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. *Science* 2019;363.
- [44] Schiller C, Diakopoulos KN, Rohwedder I. LST1 promotes the assembly of a molecular machinery responsible for tunneling nanotube formation. *J Cell Sci* 2013;126:767–777.
- [45] Kim J, Shapiro MJ, Bamidele AO. Coactosin-like 1 antagonizes cofilin to promote lamellipodial protrusion at the immune synapse. *PLoS ONE* 2014;9:e85090.
- [46] Zhao Y, Su H, Shen X. The immunological function of CD52 and its targeting in organ transplantation. *Inflamm Res* 2017;66:571–578.
- [47] Jaitin DA, Adlung L, Thaiss CA, et al. Lipid-associated macrophages control metabolic homeostasis in a trem2-dependent manner. *Cell* 2019;178:686–698.e14.
- [48] Wen Y, Lambrecht J, Ju C. Hepatic macrophages in liver homeostasis and diseases-diversity, plasticity and therapeutic opportunities. *Cell Mol Immunol* 2021;18:45–56.
- [49] Bonnardel J, T'Jonck W, Gaubomme D, et al. Stellate cells, hepatocytes, and endothelial cells imprint the kupffer cell identity on monocytes colonizing the liver macrophage niche. *Immunity* 2019;51:638–654.e9.
- [50] Sasseti C, Tangemann K, Singer MS. Identification of podocalyxin-like protein as a high endothelial venule ligand for L-selectin: parallels to CD34. *J Exp Med* 1998;187:1965–1975.
- [51] Woodman RC, Johnston B, Hickey MJ. The functional paradox of CD43 in leukocyte recruitment: a study using CD43-deficient mice. *J Exp Med* 1998;188:2181–2186.
- [52] Klintman D, Li X, Thorlacius H. Important role of P-selectin for leukocyte recruitment, hepatocellular injury, and apoptosis in endotoxemic mice. *Clin Diagn Lab Immunol* 2004;11:56–62.
- [53] Chung BK, Øgaard J, Reims HM. Spatial transcriptomics identifies enriched gene expression and cell types in human liver fibrosis. *Hepatol Commun* 2022;6:2538–2550.
- [54] Jameson G, Robinson MW. Insights into human intrahepatic NK cell function from single cell RNA sequencing datasets. *Front Immunol* 2021;12:649311.
- [55] Yang C, Siebert JR, Burns R, et al. Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome. *Nat Commun* 2019;10:3931.
- [56] Castellino F, Huang AY, Altan-Bonnet G. Chemokines enhance immunity by guiding naive CD8+ T cells to sites of CD4+ T cell-dendritic cell interaction. *Nature* 2006;440:890–895.
- [57] Das H, Kumar A, Lin Z. Kruppel-like factor 2 (KLF2) regulates proinflammatory activation of monocytes. *Proc Natl Acad Sci USA* 2006;103:6653–6658.
- [58] Peshkova IO, Fatkhullina AR, Mikulski Z. IL-27R signaling controls myeloid cells accumulation and antigen-presentation in atherosclerosis. *Sci Rep* 2017;7:2255.
- [59] Rückerl D, Hessmann M, Yoshimoto T. Alternatively activated macrophages express the IL-27 receptor alpha chain WSX-1. *Immunobiology* 2006;211:427–436.
- [60] Damgaard RB, Walker JA, Marco-Casanova P. The deubiquitinase OTULIN is an essential negative regulator of inflammation and autoimmunity. *Cell* 2016;166:1215–1230.e20.
- [61] Tiwari-Heckler S, Yee EU, Yalcin Y, et al. Adenosine deaminase 2 produced by infiltrative monocytes promotes liver fibrosis in nonalcoholic fatty liver disease. *Cell Rep* 2021;37:109897.
- [62] Letterio JJ, Roberts AB. Regulation of immune responses by TGF-beta. *Annu Rev Immunol* 1998;16:137–161.
- [63] Gindin Y, Chung C, Jiang Z, et al. A fibrosis-independent hepatic transcriptomic signature identifies drivers of disease progression in primary sclerosing cholangitis. *Hepatology* 2021;73:1105–1116.
- [64] Guillot A, Winkler M, Silva Afonso M, et al. Mapping the hepatic immune landscape identifies monocytic macrophages as key drivers of steatohepatitis and cholangiopathy progression. *Hepatology* 2023;78(1):150–166.
- [65] Schaub JR, Huppert KA, Kurial SNT, et al. De novo formation of the biliary system by TGFβ-mediated hepatocyte transdifferentiation. *Nature* 2018;557:247–251.
- [66] Sekiya S, Suzuki A. Hepatocytes, rather than cholangiocytes, can be the major source of primitive ductules in the chronically injured mouse liver. *Am J Pathol* 2014;184:1468–1478.
- [67] Amatullah H, Fraschilla I, Digumarthi S, et al. Epigenetic reader SP140 loss of function drives Crohn's disease due to uncontrolled macrophage topoisomerases. *Cell* 2022;185:3232–3247.e18.
- [68] Fraschilla I, Amatullah H, Rahman R-U. Immune chromatin reader SP140 regulates microbiota and risk for inflammatory bowel disease. *Cell Host Microbe* 2022;30:1370–1381.e5.

**Supplemental information**

**Single-cell, single-nucleus, and spatial transcriptomics characterization of the immunological landscape in the healthy and PSC human liver**

**Tallulah S. Andrews, Diana Nakib, Catia T. Perciani, Xue Zhong Ma, Lewis Liu, Erin Winter, Damra Camat, Sai W. Chung, Patricia Lumanto, Justin Manuel, Shantel Mangroo, Bettina Hansen, Bal Arpinder, Cornelia Thoeni, Blayne Sayed, Jordan Feld, Adam Gehring, Aliya Gulamhusein, Gideon M. Hirschfield, Amanda Ricciuto, Gary D. Bader, Ian D. McGilvray, and Sonya MacParland**

**Single-cell, single-nucleus, and spatial transcriptomics  
characterisation of the immunological landscape in the healthy and  
PSC human liver**

Tallulah S. Andrews, Diana Nakib, Catia T. Perciani, Xue Zhong Ma, Lewis Liu, Erin Winter, Damra Camat, Sai W. Chung, Patricia Lumanto, Justin Manuel, Shantel Mangroo, Bettina Hansen, Bal Arpinder, Cornelia Thoeni, Blayne Sayed, Jordan Feld, Adam Gehring, Aliya Gulamhusein, Gideon M Hirschfield, Amanda Ricciuto, Gary D. Bader, Ian D. McGilvray, Sonya MacParland

Table of contents

Supplementary methods.....	2
Supplementary references.....	8
Supplementary figures.....	9
Supplementary table legends.....	44

## Supplementary methods

### Sample Collection and Processing for sc- and sn- RNA-seq

Within 1 hour of caudate removal from the donor organ, intact caudate and explanted tissue sections were snap frozen, embedded in optimal cutting temperature (OCT) medium to be stored at -80°C. and formalin-fixed and paraffin-embedded (FFPE) for histology, immunohistochemistry (IHC) and immunofluorescence (IF). Remaining total liver homogenate (TLH) was cryopreserved in 90% FBS and 10% DMSO in liquid nitrogen for further experimentation and scRNA-seq. Nuclei for snRNA-seq were extracted from snap frozen tissue, as described previously.

### Processing of each 10X sample

We loaded each Chromium run to capture 10,000 single cells. We allowed generous margins for error due to damaged cells or unusually high capture rate and assumed this loading strategy would plausibly capture between 100 and 20,000 individual cells.

To ensure consistent quality across all samples, scRNA-seq data were further filtered based on genes identified, unique molecular identifiers (UMIs) and mitochondrial RNA. Sequencing reads obtained from each Chromium run were quantified using cellranger and hg18 (for specific versions see Supplementary Table 1). Droplets containing viable cells were identified using EmptyDrops from the Droplet Utils (v1.2.0) package <sup>1</sup>, using custom parameters.

Since we did not perform any washing or sorting prior to loading cells into the 10X Chromium, it was necessary to computationally identify viable cells from cellular debris and ambient RNA. We used EmptyDrops with the following parameters: lower=20,000, niters=100,000, ignore=10, retain=100. Viable cells were defined using a 1% FDR threshold. We assumed cell barcodes with fewer than 10 unique molecular identifiers (UMIs) were likely the result of ambiently floating droplet-barcodes or poor-quality reads resulting in erroneous droplet-barcode readouts, and were thus discarded.

To ensure consistent quality across all samples, cells were further filtered keeping only those cells with > 500 genes detected, > 875 total UMIs, and < 50% mt-RNA. The 50% mitochondrial RNA threshold was applied for the liver due to the presence of cells (i.e: hepatocytes) with naturally high mitochondrial counts<sup>3</sup>. Cells from all samples were library-size normalized to 1500 UMIs/cell and log transformed using a pseudocount of 1. Each sample was scaled individually, and each cell was annotated using scmap-cluster <sup>2</sup> using our 5 liver map as a reference <sup>3</sup>. In addition, the sample was clustered using Seurat's Louvain algorithm using default parameters and clusters were automatically annotated by comparing the enrichment for our previously identified marker genes <sup>3</sup> among the top DE genes for each cluster (see below). Cells were assigned to cell-cycle stages using Seurat's default parameters. The top 2,000 most highly variable genes (HVGs) were identified in each sample using Seurat (v3.1.3) <sup>4</sup>, genes located in the mitochondrial genome were excluded from the set of HVGs and cells in each sample were assigned to cell-cycle stages using Seurat's default parameters.

## **Custom Marker-based Automatic Annotation**

A cell-type specific marker for cluster  $i$  was defined as follows:

1. The mean expression of across all clusters was ordered from highest to lowest
2. The highest change between sequential clusters was calculated.
3. If fewer than half the clusters fell above this threshold this gene was deemed a marker of those clusters
4. Repeat 1-3 for mean detection rate
5. Genes consistently defined as markers using mean expression and detection rate, and where either the change in mean expression exceeded 0.3, or the change in detection rate exceeded 0.1 were kept as marker genes of the cluster(s).

This definition allows genes to be markers of multiple clusters and while prioritizing genes with the largest changes in expression. This was necessary to avoid issues with new liver samples being over clustered - i.e. the same cell-type being spread across multiple clusters - relative to the original liver map.

The above definition was also used to define marker genes in our original 5 Liver map. Reference marker genes that were specific to a single cell-type were labeled as such, while genes shared across related clusters were assigned the high-order annotations: "Hepatocyte", "LSEC", "Macrophage", "T cell", "B cell".

The sets of marker genes for each cluster were tested for enrichment with each set of reference marker genes using a hypergeometric test. The set of reference marker genes with the lowest p-value (highest significance) was used to annotate the cluster.

## **Pseudobulk gene expression**

Pseudobulk gene expression for each cell type in each sample by summing the respective raw UMI counts. Pseudobulk expression was compared between NDD and PSC/PBC using edgeR's exactTest using only genes passing QC thresholds in PSC, PBC and NDD data and cell types with at least 2 samples in each condition containing at least 10 cells of that type. Differentially expressed genes were identified using  $FDR < 0.05$ .

## **Integration**

Raw, normalized and scaled expression matrices for all samples were merged. Consensus highly variable genes were defined as those genes determined to be highly variable in at least two samples and that were detected in all samples (leaving 3,204 highly variable genes). The samples were integrated using harmony<sup>5</sup>, with default parameters, based on the expression of these highly variable genes, and a global integrated UMAP was calculated from the harmony integrated lower dimensional space using default parameters.

## **Clustering**

The integrated liver map was clustered using Seurat's Louvain clustering algorithm varying the resolution parameter from 0.3 to 2 and the  $k$  of the cell-cell graph from 40 to 80. To identify the

most robust clustering solutions from these 30 clusterings, we used affinity propagation<sup>6</sup>, as implemented in apcluster (v1.4.8), with variation of information<sup>7</sup> to measure distances between clusterings. The “p” parameter for apcluster was set at -2.5 as this resulted in the most stable number of clusters which was 3, the exemplar for each of these clusters was selected to define three levels of clustering: Fine, Coarse, and Core (see supplementary). Using the uniformity of automatic cell-type annotation within clusters as a metric, it was determined the Coarse clustering (res = 0.6, k = 50) to be the most consistent with the underlying biology. Clusters were annotated using a combination of the most common automatic annotation obtained at the sample level (see above), and manual curation using known DE genes<sup>3</sup>. DE genes for these clusters were determined using the wilcoxon-rank-sum test method in Seurat comparing each cluster to all others.

The resulting clusters were manually annotated by comparing markers of each cluster from the FindMarkers function in Seurat with known cell type specific genes (**Supplementary Table 2 & 3**). The integrated PSC, PBC and NDD 5' single-cell data were clustered using the top 30 principal components, k=20, res=2 and the integrated PSC, PBC and NDD 3' single-nucleus data were clustered using the top 30 PCs, k=20, res=2. The resolution was increased to separate neutrophils from other myeloid populations. Clusters that represented the same cell types were merged during manual annotation (**Supplementary Table 6-7**).

### **Variability in cell-type frequency**

Since frequencies are a fixed-sum composition, we examine only relative abundance of cell-type relative each other in each sample and compare the relative proportion of each pair of cell-types to the median relative proportion of that pair of cell-types across all samples (See equation 1). A cell-type that has a significantly different relative abundance, as determined with a proportion test, to >75% of other cell-types in a particular sample is deemed significantly different in that sample.

$P_{ij}$  = proportion of cells of type  $i$  in sample  $j$  =  $n_{ij}/N_j$   
Relative proportion  $i$  vs  $k$  in sample  $j$  =  $P_{ij}/(P_{kj} + P_{ij})$

A general linear model was used to test the association of each cell-type frequency with either donor age or donor sex, using a Benjamini-Hochberg FDR correction no association was significant (FDR > 0.3).

### **Subclustering**

The merged data was subsetted into 7 groups based on the global UMAP and core clustering structure (Supplementary Table). The clustering pipeline was rerun on this subset, including: high variable gene detection, PCA, harmony integration, Seurat clustering, and apcluster.

Subclusters were manually annotated using marker genes identified using the Wilcoxon-rank-sum test. Those subclusters that were subclustered further followed the same procedure using

1,000 HVGs, 20 PCs, and Seurat clustering with resolution of 0.5. Subtype markers were identified using the wilcox-rank-sum test, and pathway analysis was performed using fgsea with 5% FDR and 20,000 permutations using the MSigdb Hallmark pathways, MSigdb Immune pathways, and Reactome pathways. The significant pathways were manually curated to identify the most significant and distinct pathways, as well as determine the most appropriate term descriptor when multiple highly overlapping pathways were among the top enriched pathways. Finally manual inspection of a DotPlot of the contributing genes was performed to confirm upregulation as specificity of the associated genes.

## **Pathway Enrichment**

Differentially expressed genes in each subcluster were calculated using the wilcox-rank-sum test after filtering out undetectable genes (expressed in less than 0.1% of all cells of this cell-type), and log<sub>2</sub> fold-changes were calculated. Pathway enrichments were calculated using fgsea on genes ranked by their log<sub>2</sub> fold-changes after excluding ribosomal proteins and transcripts originating from the mitochondrial genome. Only terms with between 15 and 1000 annotated genes were considered and 100,000 permutations were used to calculate significance and ensure reproducibility of enrichments. Pathways significant at a 5% FDR were reported as significant. The top 20 significantly enriched pathways from each geneset source were manually curated to identify the most informative non-redundant pathway(s) from each source. Human pathway annotations were obtained from KEGG, Reactome, and MSigdb (Hallmark and Immune pathways). To ensure reliability of immune related pathways, results were further curated to ensure both up and down regulated gene lists were enriched among up and down regulated DE genes respectively.

Pathway enrichments were calculated as above and cell-cell interactions were identified using ligand-receptor interactions obtained from the KEGG Cytokine pathway (hsa04060). Cell type frequencies were calculated for each sample, and differences between NDD and PSC were determined using the Wilcoxon rank-sum test. Of note, there were too few samples of PBC (N=2) to accurately estimate sample-sample heterogeneity in cell-type frequency.

Note, we only consider Cytokine pathways to avoid these pathways being obfuscated by adhesion and complement cascade pathways that are very common between our hepatocyte and endothelial cells.

## **Macrophage-Endothelial cell-cell communication**

We used scmap to determine the specific identity of macrophages and endothelial cells comprising the macrophage-endothelial doublets. Doublets were independently mapped to the macrophage subpopulations and endothelial subpopulations using the most specific marker genes for each cell population as identified during manual annotation. Only doublets that could be reliably assigned (similarity > 0.6 & consistent between pearson and spearman correlations) to both a macrophage and endothelial subpopulation were counted to determine the number of

doublets containing each possible pair of subtypes. Cellphonedb was used on each subclustered sample individually then results were summed across samples.

### ***In Silico Flow Sorting***

To compare our single-cell RNAseq data to our flow cytometry data, we performed *in silico* gating of our single-cell data using thresholds based on the non-immune fraction: PTPRC > 0.3, CD68 > 0.3. PSC scRNA-seq was enriched in immune expression overall compared to NDD, thus higher thresholds for “on” for each gene were determined based on maximum expression seen in non-immune cell types.

### ***10x Genomics Visium and Nanostring GeoMx DSP Experimental Protocol***

Tissue was prepared for Visium Spatial transcriptomics as previously described. Briefly, liver tissue was embedded in OCT media, frozen at -80°C, and cryosectioned with 16-µm thickness at -10°C (cryostar NX70 HOMP). Sections were placed on a chilled Visium Tissue Optimization Slide (10x Genomics) and processed following the Visium Spatial Gene Expression User Guide. Tissue was permeabilized for 12 minutes, based on an initial optimizations trial. Libraries were prepared according to the Visium Spatial Gene Expression User Guide and samples sequenced on a NovaSeq 6000. PSC and NDD liver sections were sliced from OCT-embedded sections and submitted to NanoString for staining with selected morphological markers: DAPI (nuclei), KRT (epithelial cells), CD45 (immune cells), and CD68 (macrophages). For this experiment, we employed the GeoMx Cancer Transcriptome Atlas that contains over 1800 immune- and cancer-related targets.

### **Spatial Transcriptomics**

Tissue was prepared for Visium Spatial transcriptomics as previously described. Briefly, liver tissue was embedded in OCT media, frozen at -80°C, and cryosectioned with 16-µm thickness at -10°C (cryostar NX70 HOMP). Sections were placed on a chilled Visium Tissue Optimization Slide (10x Genomics) and processed following the Visium Spatial Gene Expression User Guide. Tissue was permeabilized for 12 minutes, based on an initial optimizations trial. Libraries were prepared according to the Visium Spatial Gene Expression User Guide and samples sequenced on a NovaSeq 6000.

Spatial transcriptomics was performed on four sequential slices from a single healthy human liver, using the 10X VISIUM platform according to the manufacturer's protocol. Raw data was processed using spaceranger, and downstream analysis was performed using Seurat (version 4.0.4). Raw counts were normalized using SCTransform, and the proportion of mitochondrial and ribosomal content was calculated using standard methods. Mitochondrial genome transcripts and genes with no more than 2 UMIs in any spot in the entire tissue slice were removed. Gene expression was normalized using sctransform, and data visualized using Seurat (v4.0.2). Transcriptomic data were clustered using the standard Seurat pipeline: highly variable gene detection, PCA, k-nearest-neighbor detection, Louvain clustering, and wilcoxon-rank-sum

test for marker genes, using default parameters. Note, spatially variable genes largely overlapped with highly variable genes but excluded many B cell and red blood cell markers.

Principal component analysis was performed on each slice using default parameters on the normalized data. The top 12 principal components were rotated using the inbuilt varimax method in R to improve the specificity of patterns captured. The gene loading matrix of each set of rotated components were correlated across slices and reciprocal best-matches were determined to identify sets of equivalent components. The set of matching components ranked either 1st or 2nd by total variance explained across slices was determined through visual inspection and examination of gene loadings to represent the spatial zonation of hepatocytes across liver lobules. The highest quality slice, as determined by number of genes per spot and mitochondrial proportion per spot, was used for publication figures.

The zonation score of individual genes was determined by scaling their normalized expression values to sum to 100, which were then used as weights to determine the average zonation of spots expressing the gene. Significance of zonation was determined using linear regression with zonation as the dependent variable and expression as the independent variable.

Cell-type enrichments were calculated using the Hypergeometric method within Giotto<sup>8</sup>, using the marker genes in Supplementary Table 2.

Significance of zonation was calculated using linear regression between zonation score and the normalized gene expression. Colocalization was determined using Pearson correlation.

### ***Integration of Publicly Available Cirrhosis scRNA-seq Data***

Published liver single-cell RNAseq cirrhosis data were obtained from the original authors<sup>9</sup>. Cell types were reannotated using marker genes and cell type names defined in this manuscript (**Fig. S10**). Differential expression and changes in cell type frequencies between cirrhotic and non-cirrhotic (tumor adjacent) tissue was performed, as described above.

### **Flow cytometry and Intracellular Cytokine Staining:**

Cell suspensions ( $2 \times 10^6$  cells) from the non-parenchymal cell (NPC) fraction were stimulated in 12-well plates with 10 ng/ml LPS (L2880-10MG), 25ng/ml IFN $\gamma$  (Thermofisher, PHC4031) for 6 hours in the presence of brefeldin (BD Biosciences, 555029) and monensin (BD Biosciences, 554724).

Live/dead Zombie Violet (Biolegend, 423114) or Zombie NIR (Biolegend, 423106) was employed to assess viable cells and fluorophore-conjugated monoclonal antibodies to the following human cell-surface markers: anti-CD45-BV650 (Biolegend, 304044), anti-CD206-AF647 (Biolegend, 321116), anti-HLADR-AF700 (Biolegend, 307626), anti-CD68-R780 (Biolegend, 333816), and anti-CD68-PeCyp5.5 (Biolegend, 333814). Intracellular TNF $\alpha$  was detected with anti-TNF $\alpha$ -PeCy7 (Biolegend, 502930) and anti-TNF $\alpha$ -PacBlue (Biolegend, 502920). Gating strategy for

cell-surface markers was set based on background auto-fluorescence measured in unstained and fluorescence minus one (FMO) controls, and gating strategy for intracellular markers in stimulated samples was set based on unstimulated controls and FMOs.

### **Immunohistochemistry and Immunofluorescence:**

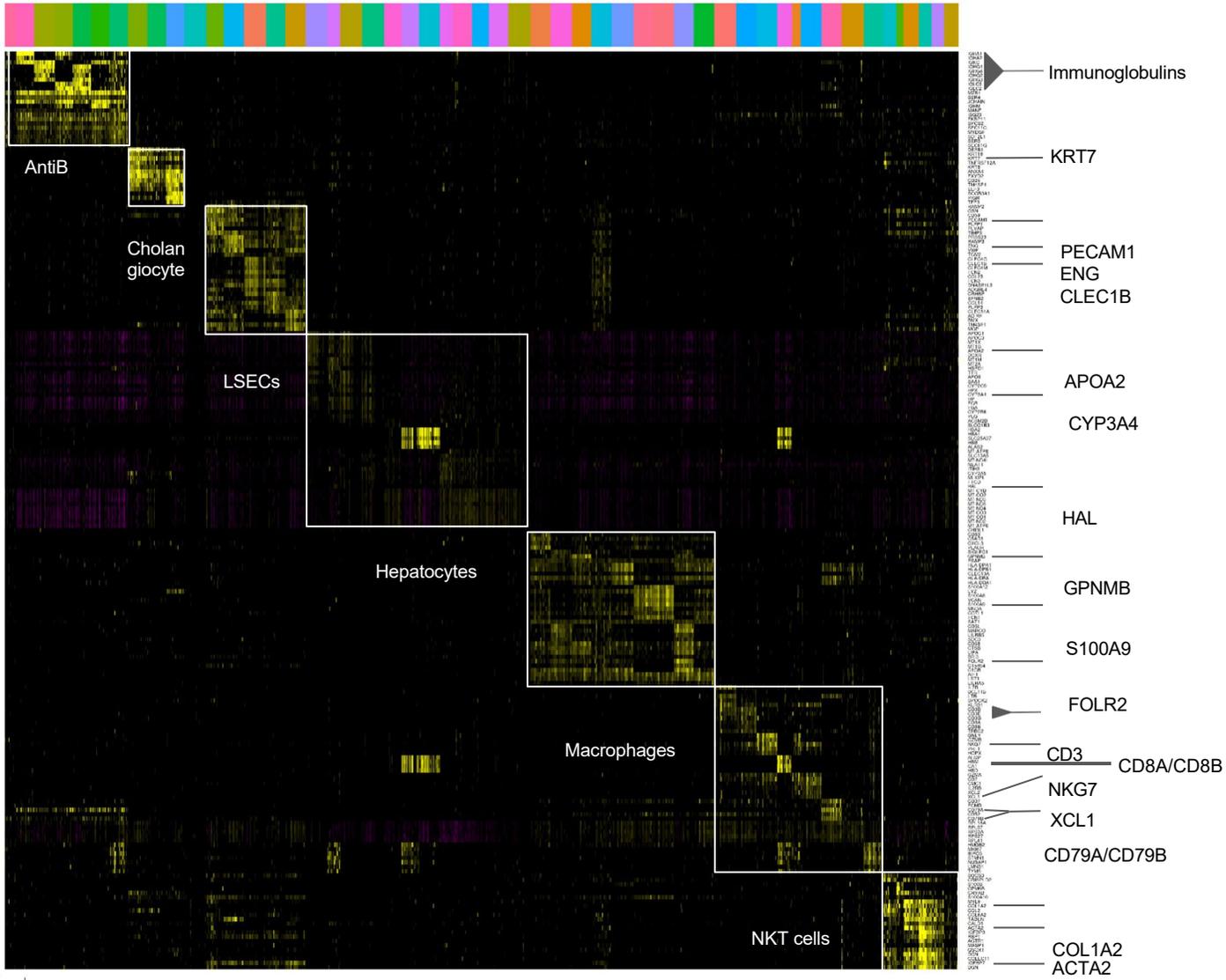
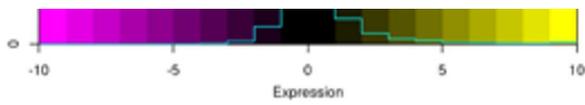
Annotation of fibrotic areas and chronic biliary disease, referred to as ‘scars’ in this study, within the liver parenchyma of PSC explant livers were defined based on Masson’s trichrome stain highlighting fibrosis) and Cytokeratin 7 (CK7) immunohistochemical stain that highlighting biliary metaplasia in hepatocytes as a feature of chronic biliary disease. CK7 staining highlights the hepatocytes with biliary metaplasia near the edges of fibrotic areas (scars) in PSC explant samples (**Fig. S11, 22-23**). Antibodies included anti-VCAM1 (Thermofisher, MA5-11447), anti-CD32 (Thermofisher, MA5-32601), anti-C1QC (Abcam, ab75756), anti-CK7 (Abcam, ab68459), anti-HNF4A (Abcam, ab41898), anti-CD9 (Abcam, ab236630), and anti-CD36 (ab17044).

### **Supplementary references**

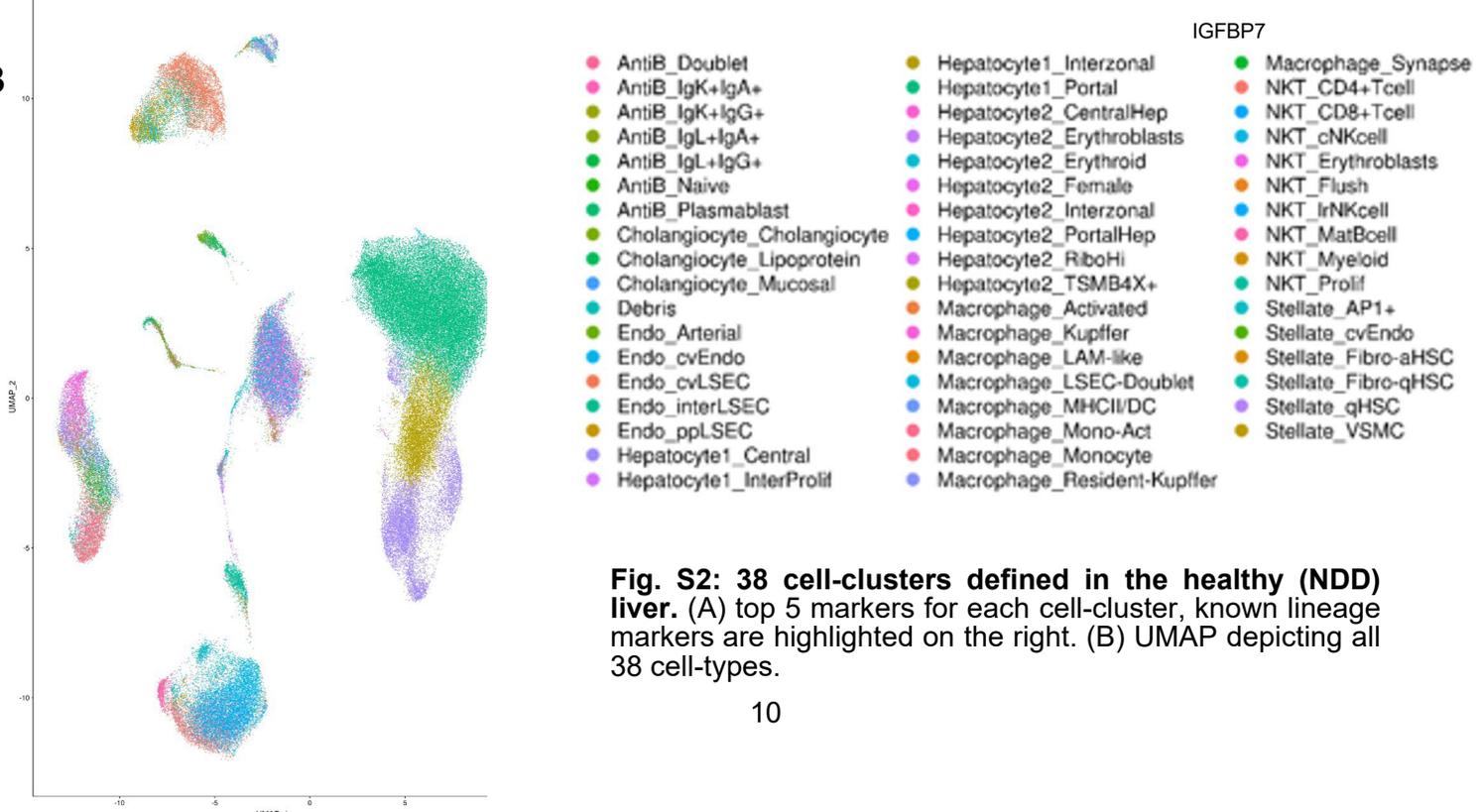
1. Lun ATL, Riesenfeld S, Andrews T, et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* 2019;20(1):63.
2. Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. *Nat Methods.* 2018;15(5):359-362.
3. MacParland SA, Liu JC, Ma X-Z, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun.* 2018;9(1):4383.
4. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol.* 2015;33(5):495-502.
5. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* 2019;16(12):1289-1296.
6. Frey BJ, Dueck D. Clustering by passing messages between data points. *Science.* 2007;315(5814):972-976.
7. Meilă M. Comparing Clusterings by the Variation of Information. In: Schölkopf B, Warmuth MK, eds. *Learning Theory and Kernel Machines.* Vol 2777. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2003:173-187.
8. Dries R, Zhu Q, Dong R, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* 2021;22(1):78.
9. Ramachandran P, Dobie R, Wilson-Kanamori JR, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 2019;575:512–8.



**A**



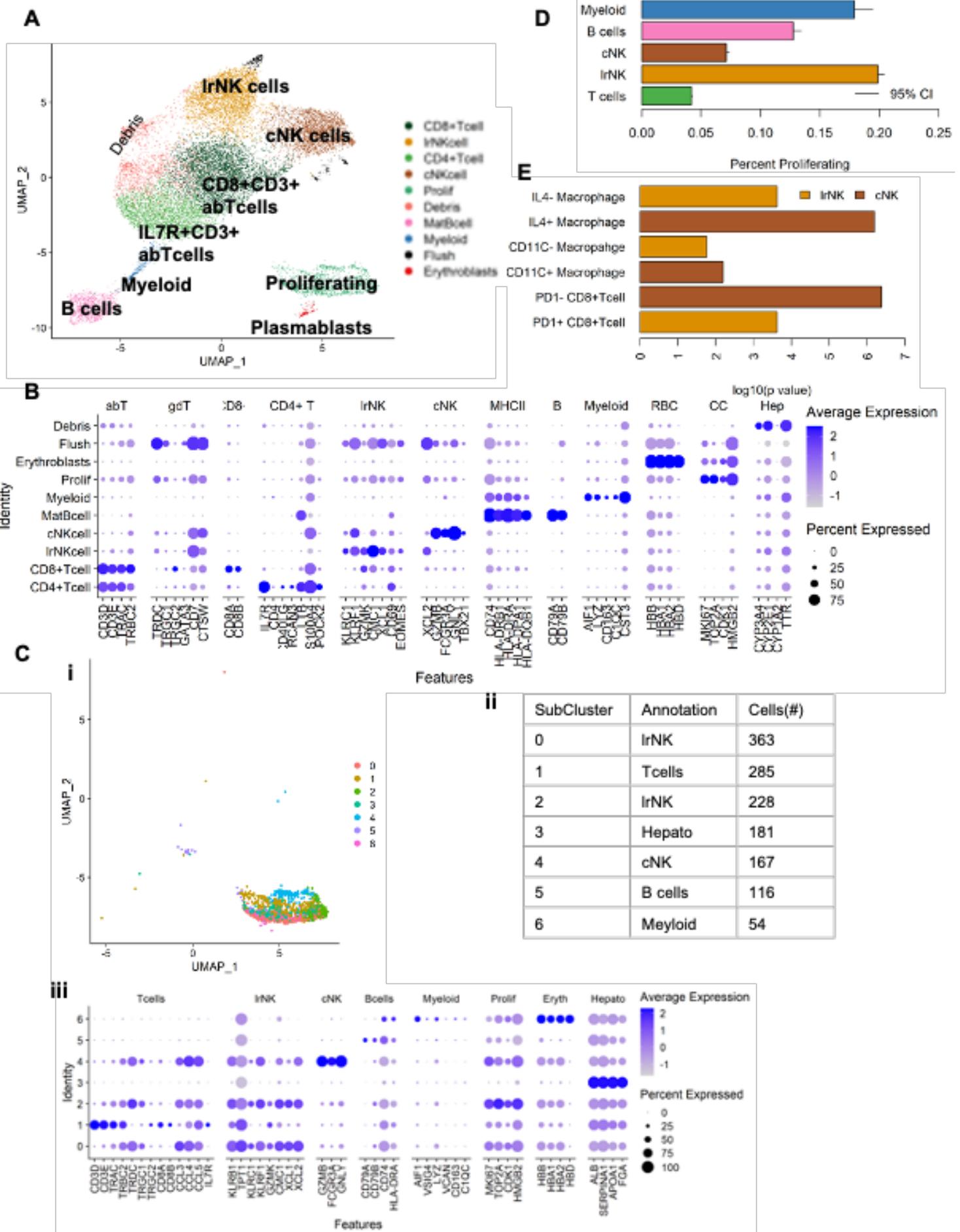
**B**



**Fig. S2: 38 cell-clusters defined in the healthy (NDD) liver.** (A) top 5 markers for each cell-cluster, known lineage markers are highlighted on the right. (B) UMAP depicting all 38 cell-types.

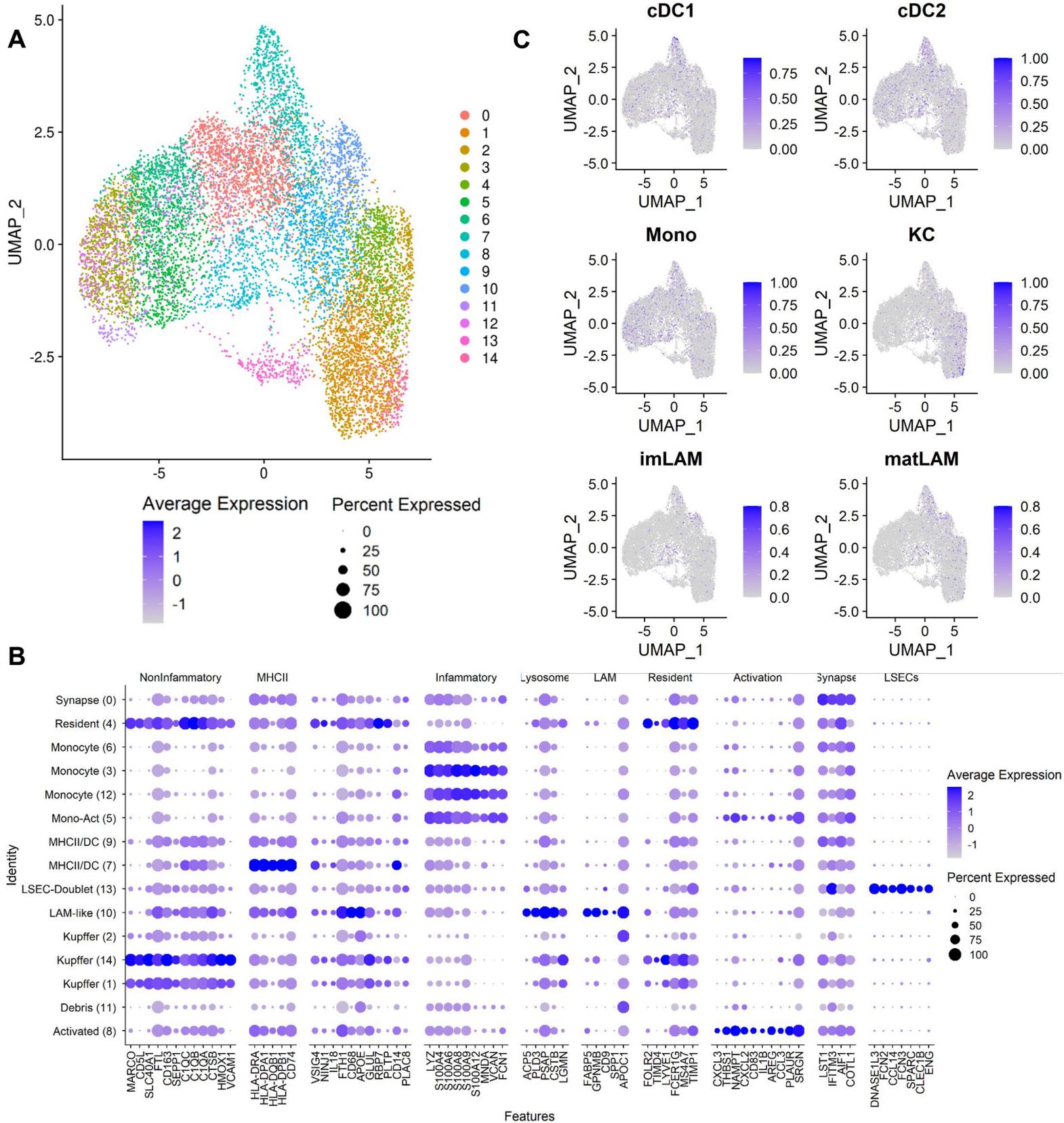


**Extended Data Fig 3 : Subclustering of other populations of the NDD Map.** (A) Endothelial cells. (B) Stellate / Mesenchymal cells (C) Cholangiocytes (D) Antibody secreting B cells. Left panels show UMAP of subclusters, Right panel shows top marker genes for each cluster.

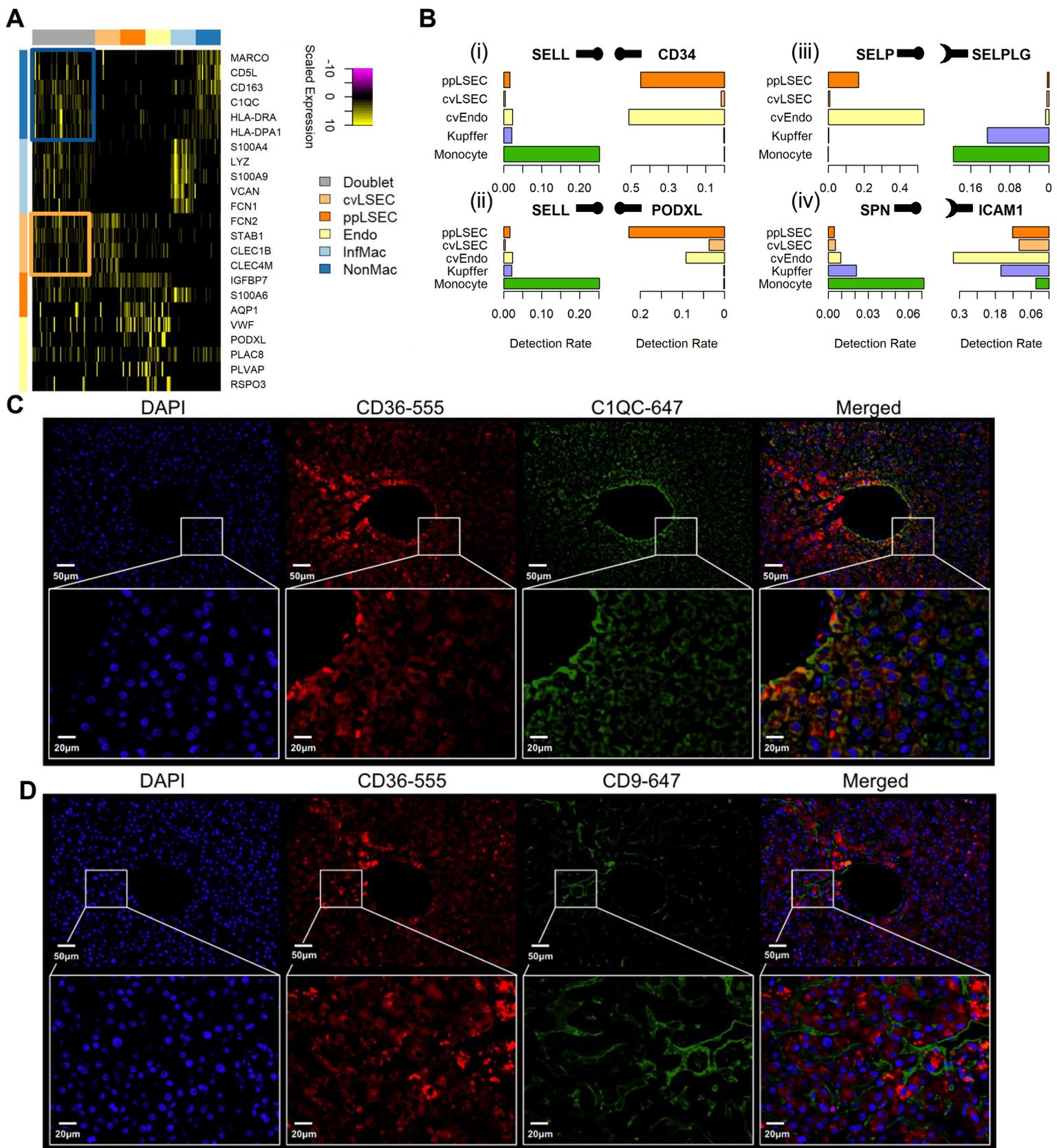


**Fig. S4: Subclustering of lymphocyte-like cells of the NDD Map.** (A) UMAP of subclustered lymphocyte-like cells. (B) Marker genes of lymphocyte subtypes. (C i-iii) Subclustering of the proliferating cluster and annotation using the same markers as B. (D) Proportion of proliferating cells for each lymphocyte group based on C. (E) Differential pathway enrichments between cNK and IrNK cells from GSEA.



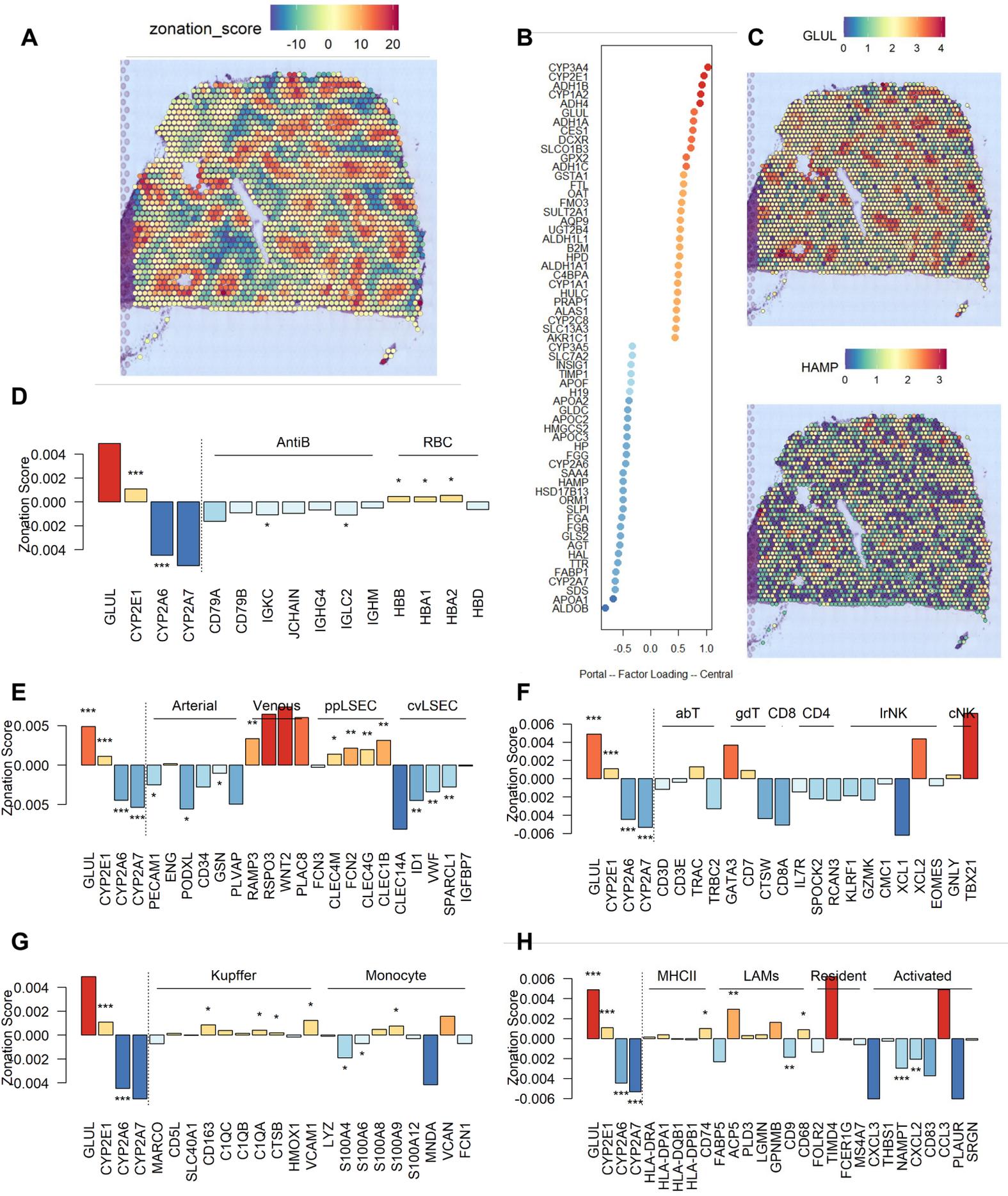


**Fig. S5: Subclustering of Macrophages** (A) UMAP of integrated subclustered macrophages (B) Top marker genes for each cluster used to manually annotate the macrophage map. (C) Average expression of macrophage gene signatures from Guilliams et al 2022.

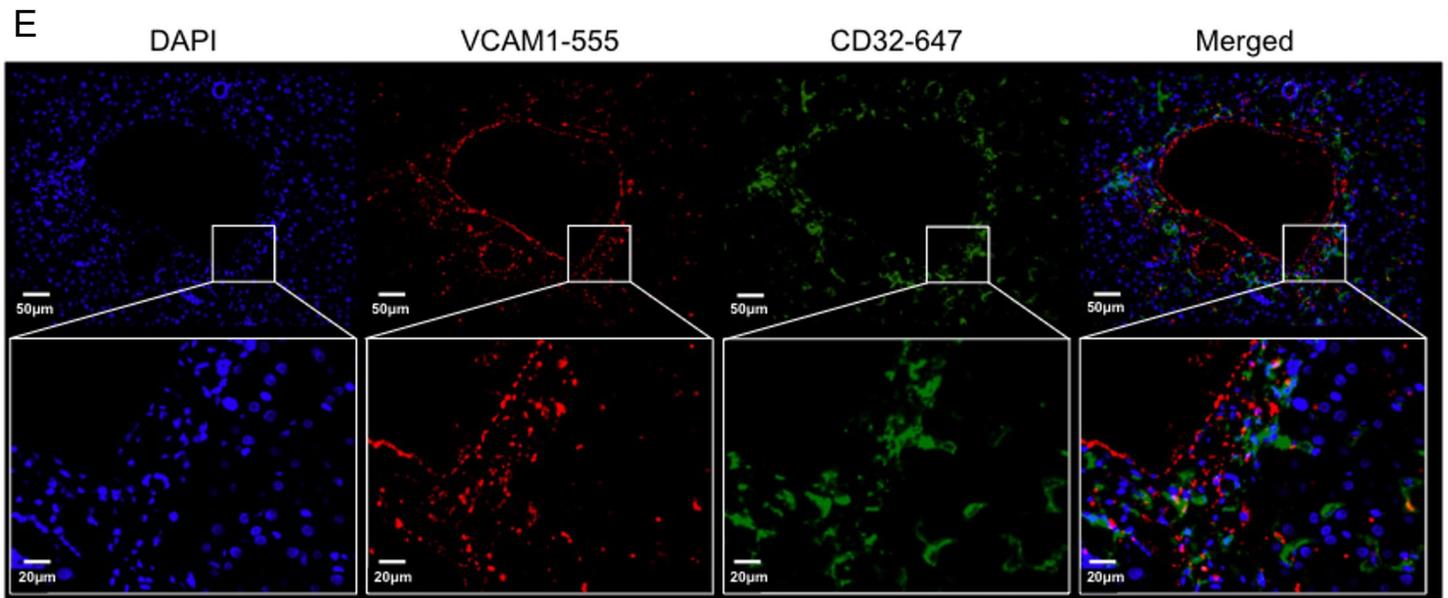
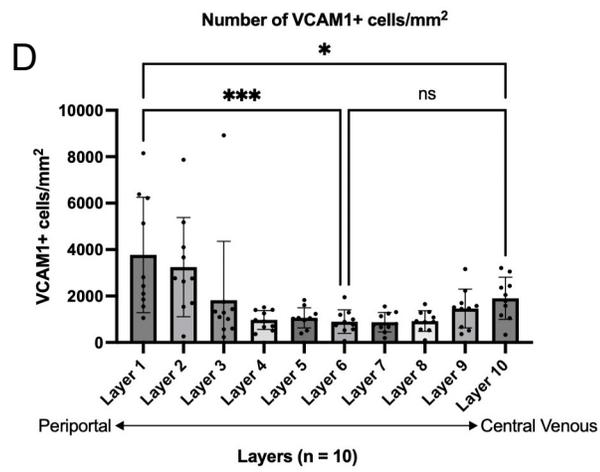
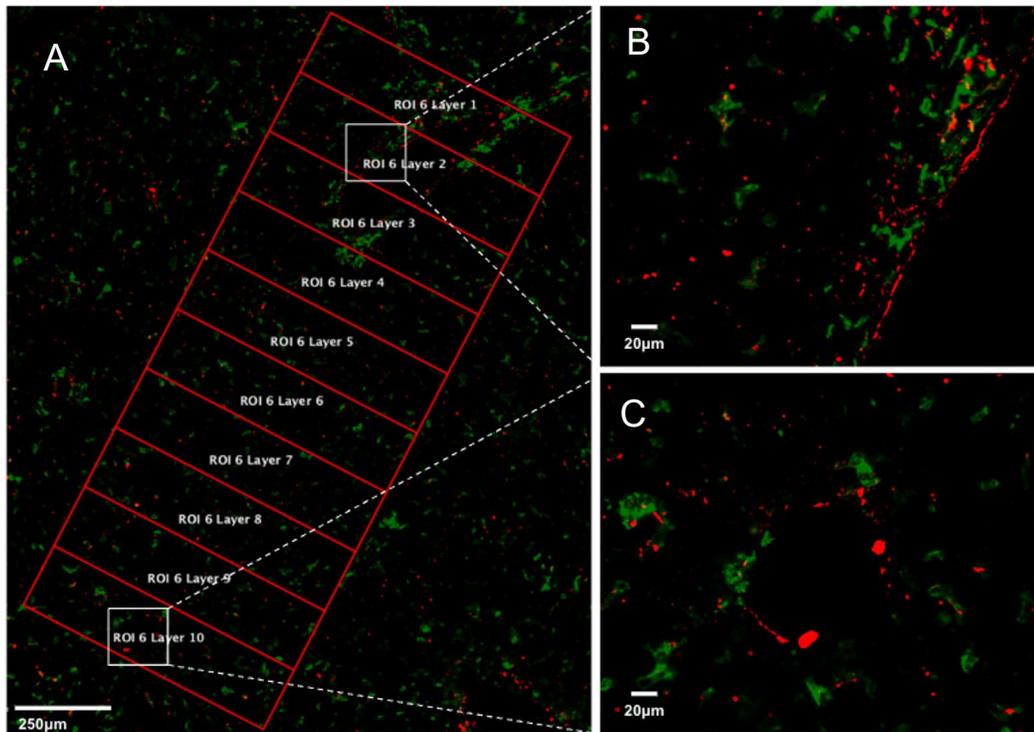


**Fig. S6: Macrophage-LSEC interaction.** (A) Expression of lineage specific genes in a random subset of doublet (grey), and singlet LSEC (orange) and macrophage (blue) cells confirms co-expression in the same droplet. (B) Most specific Monocyte-cvEndo interacting proteins as determined by CellPhoneDB. (C) Demonstrating colocalization of C1QC<sup>+</sup> Kupffer cells and CD36<sup>+</sup> LSECs, Central venous region in the healthy (NDD) liver by individual markers and

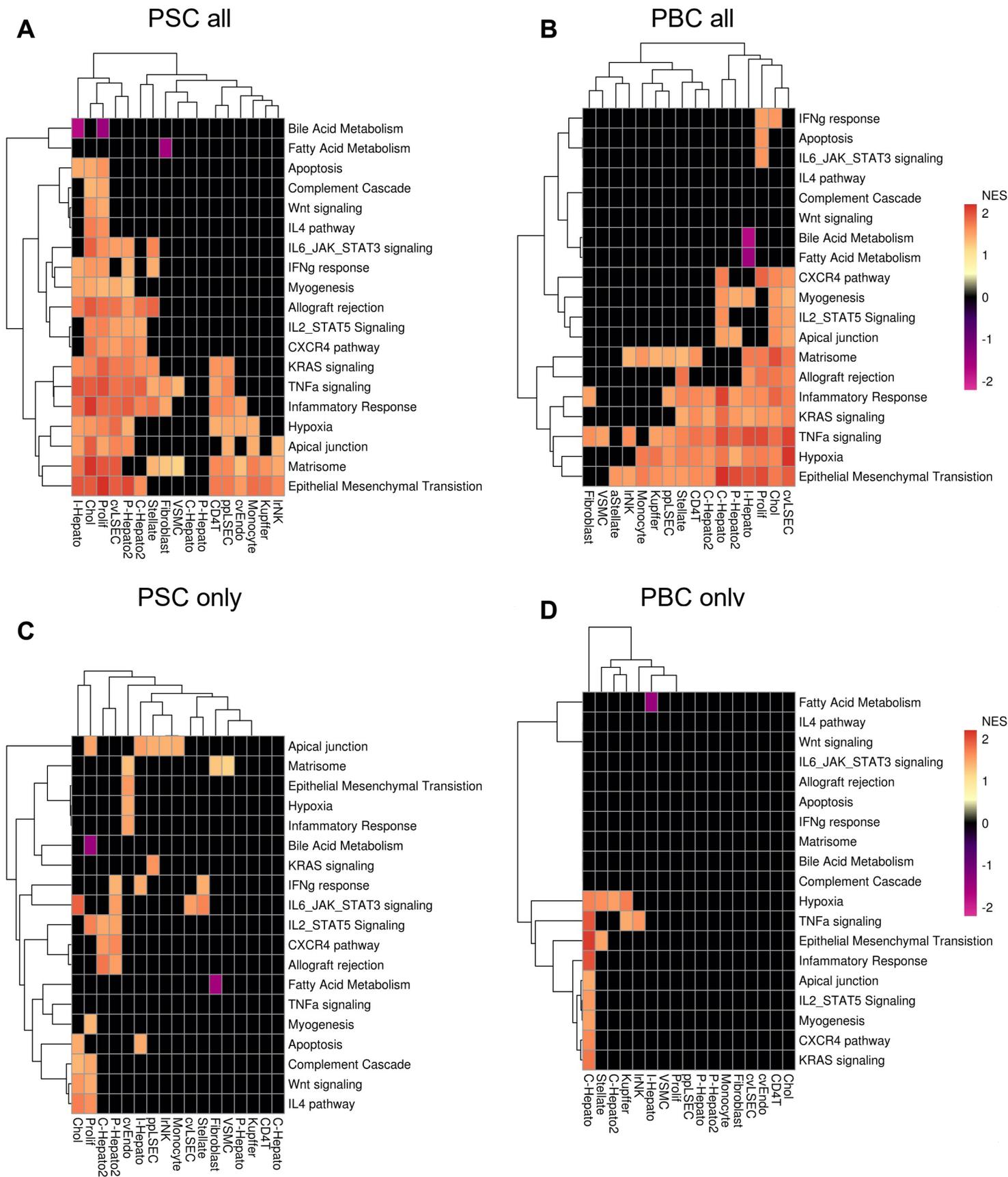
fluorophores (DAPI, CD36-555 and C1QC-647), as well as merged. D) Demonstrating the presence of CD9<sup>+</sup> cells in the healthy (NDD) liver. Central venous region in the healthy (NDD) liver by individual markers and fluorophores (DAPI, CD36-555 and CD9-647), as well as merged.



**Fig. S7: Zonation of cell types in NDD liver.** (A) First varimax principal component captures hepatocyte zonation in healthy liver spatial transcriptomics. (B) Topmost contributing genes to the zonation associated component. (C) Example gene expression of a pericentral (GLUL) and periportal (HAMP) gene. (D - H) zonation score of individual marker genes for each of the major cell-types identified in the healthy map. Red = pericentral, blue = periportal. \*p < 0.01, \*\*p < 10<sup>-10</sup>, \*\*\*p < 10<sup>-100</sup>.



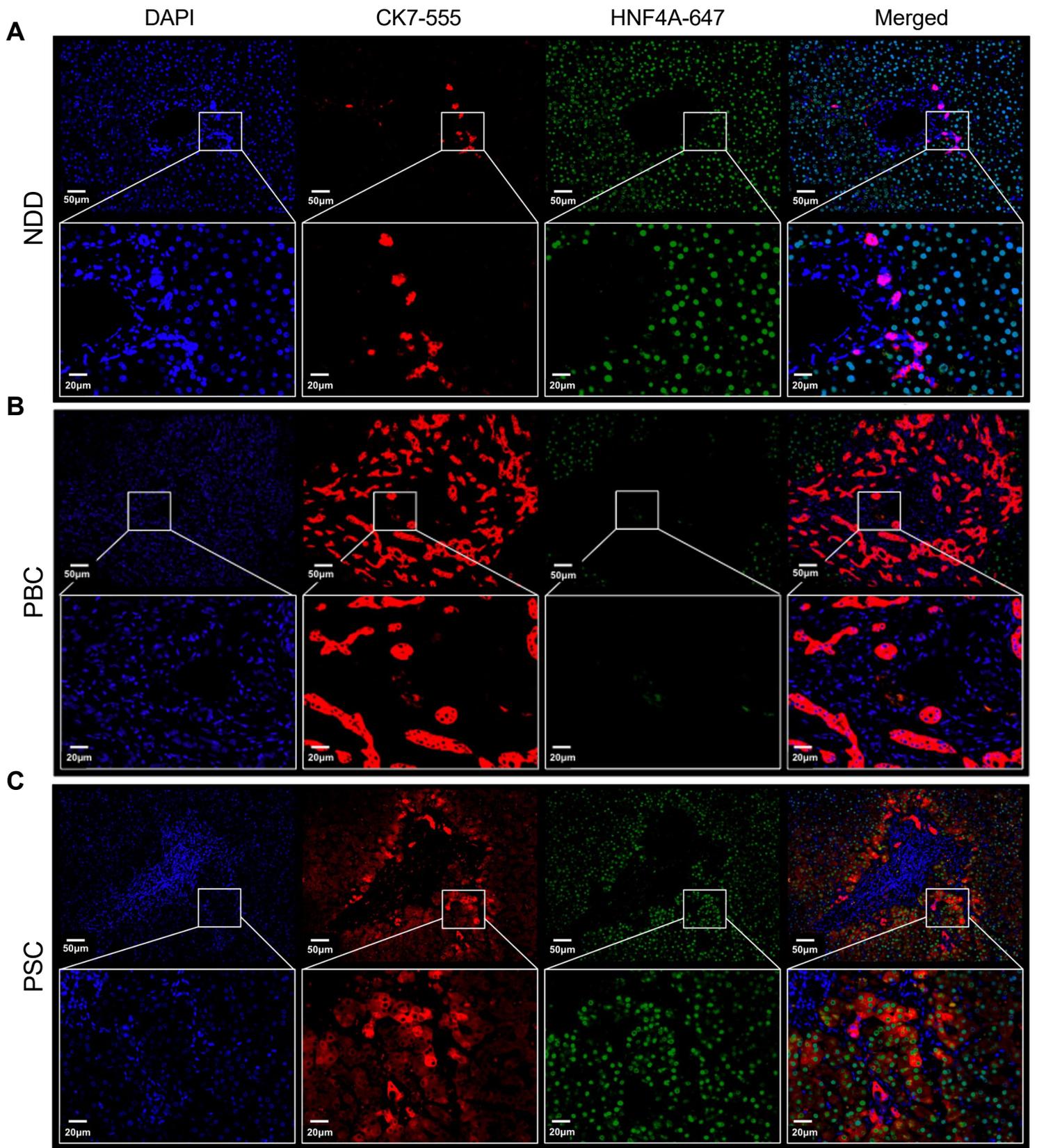
**Extended Data Fig 8. Immunofluorescence identifies bimodal distribution of VCAM1+ cells across the liver sinusoid.** Red= VCAM1, Green = CD32, Blue = DAPI. (A) Individual sinusoids were divided into 10 layers from periportal (Layer 1) to central venous (Layer 10) regions. (B) Example images of a periportal, and (C) pericentral layer. (D) Quantification across 10 replicates. Statistical significance was assessed a one-way analysis of variance (ANOVA) with a Bonferroni post-test, \*\*\* P < 0.001, \*\*P < 0.01, \*P < 0.05. (E) Periportal region in healthy (NDD) liver by individual markers and fluorophores (DAPI, VCAM1-555 and CD32-647), as well as merged.



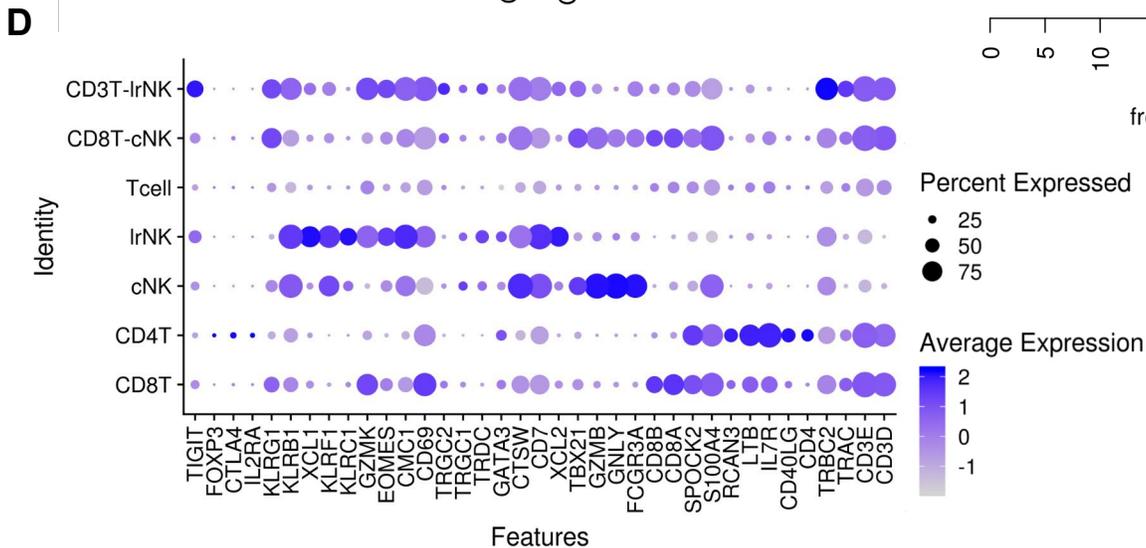
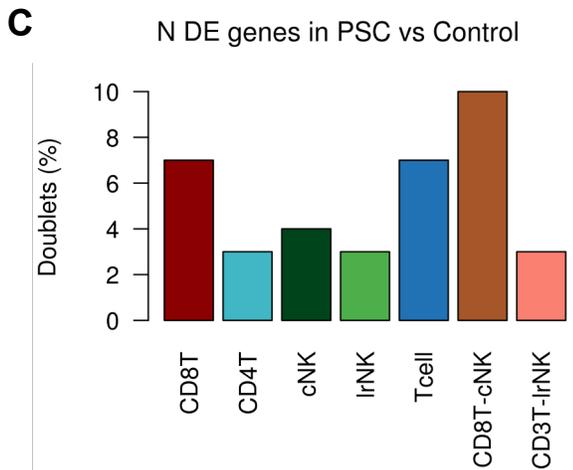
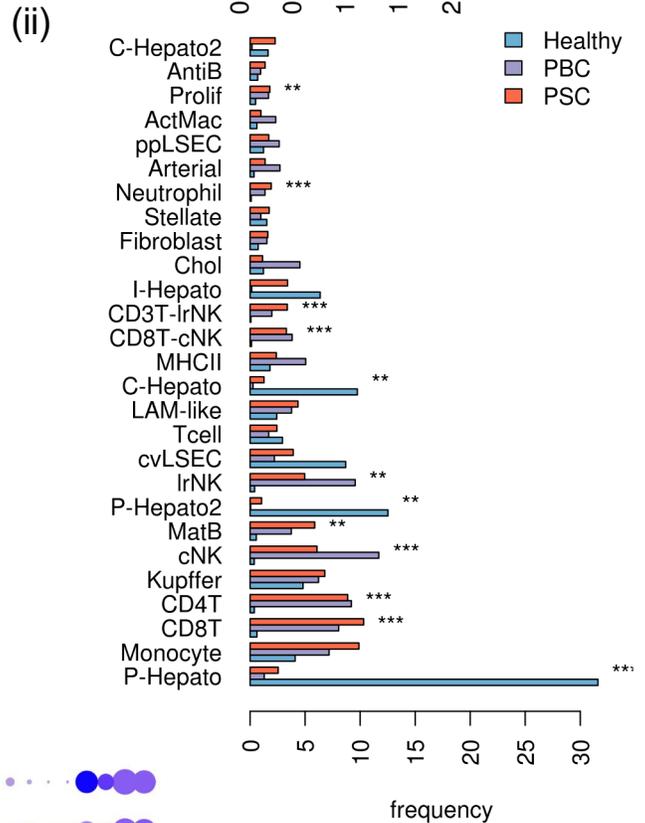
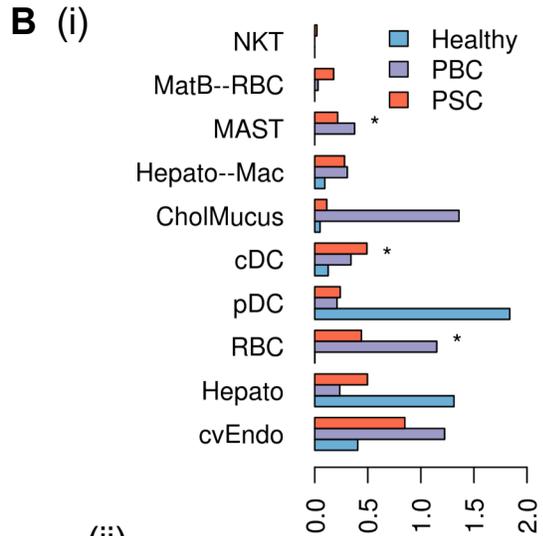
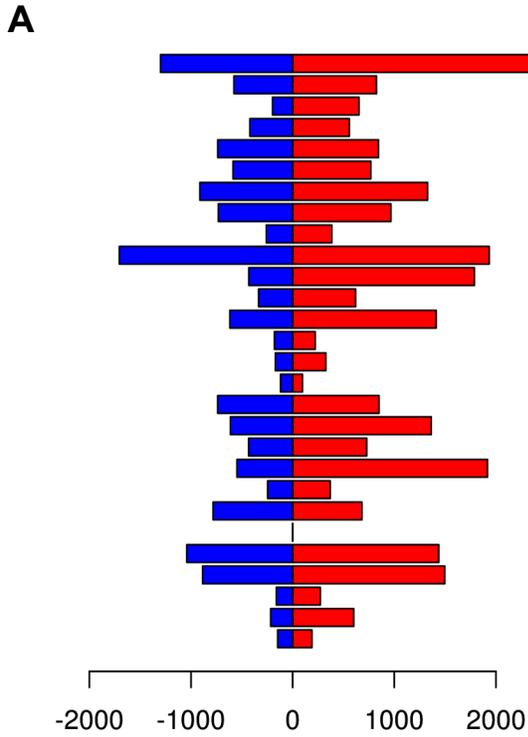
**Fig. S9: Disease-associated pathways in single-nucleus RNAseq.** (A) Significantly enriched pathways among genes differentially expressed between PSC and NDD livers (B) Significantly enriched pathways among genes differentially expressed between PBC and NDD livers (C) Pathways from A filtered to remove those also found in B. (D) Pathways from B filtered to remove those also found in A.



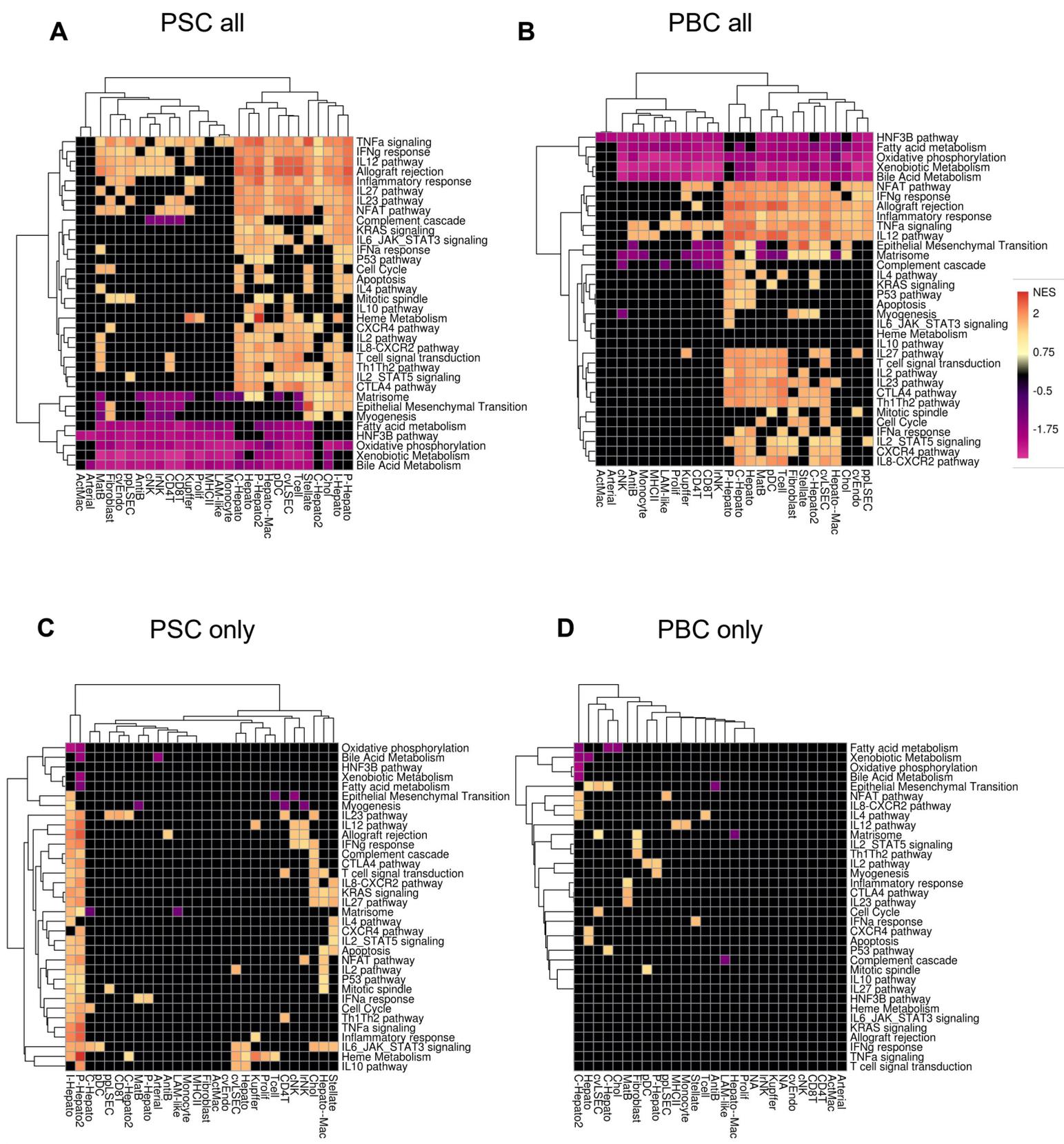
**Fig. S10: Integrated UMAP of cirrhosis liver data from Ramachandran et al. (2019).** (A) Combined UMAP split by aetiology. (B) number of genes up (red) or down (blue) regulated in cirrhotic livers compared to uninjured livers in each cell-type. (C) frequency of each cell-type in each aetiology, due to small sample sizes, significance was tested comparing all cirrhotic to uninjured liver. (D) Cytokine expression in non-inflammatory cirrhotic liver. (E) Expression of hepatocyte and cholangiocyte marker genes from this study in the Ramachandran et al (2019) data. Note there was only 1 Hepatocyte captured in the PBC data thus it has been excluded.



**Extended Data Fig 11. Immunofluorescence reveals increased CK7+HNF4A+ co-expressing cells in the PSC periportal area in comparison to the PBC and NDD liver.** Periportal region by individual and merged protein markers: DAPI (blue), CK7 (red) and HNF4A (green) in (A) NDD, (B) PBC, and (C) PSC.



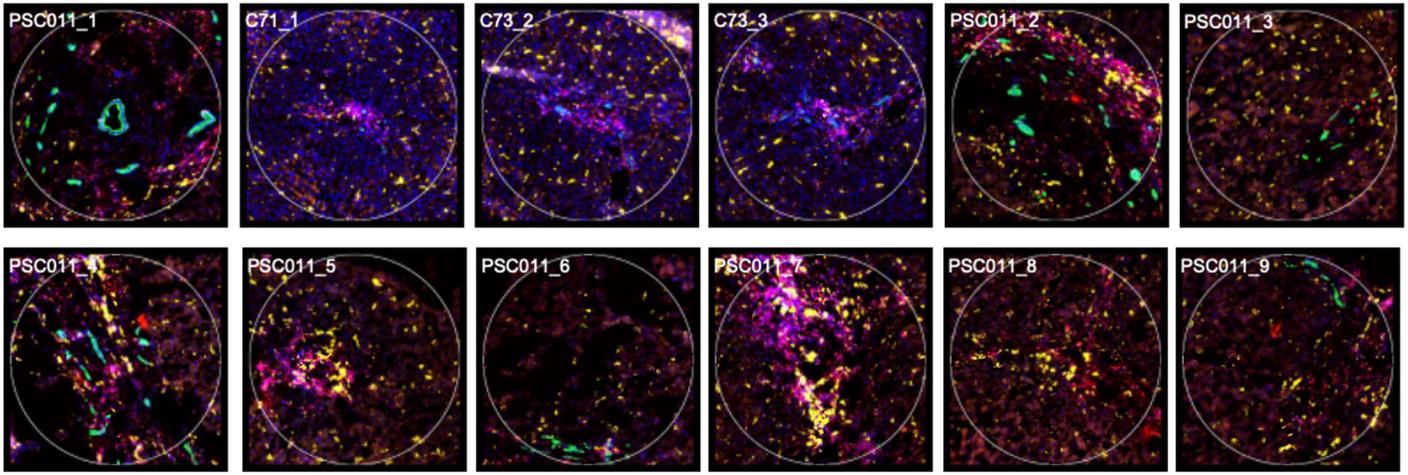
**Fig. S12: Cholestatic disease immune diversity captured in scRNA-seq.** (A) number of genes up (red) or down (blue) regulated in PSC livers compared to NDD in each cell-type. (B) frequency of each cell-type in each aetiology, (i) rare cell-types that accounted for <1% of all the cells, (ii) cell-types accounted for  $\geq 1\%$ . Frequency of each cell-type among livers of each aetiology, significance obtained from wilcoxon rank sum test across sample: \* p value < 0.05, \*\* p value < 0.01, \*\*\*p value < 0.001. (C) Proportion of doublets as estimated by Doublet Finder. (D) Expression of lymphocyte marker genes across T & NK cell populations.



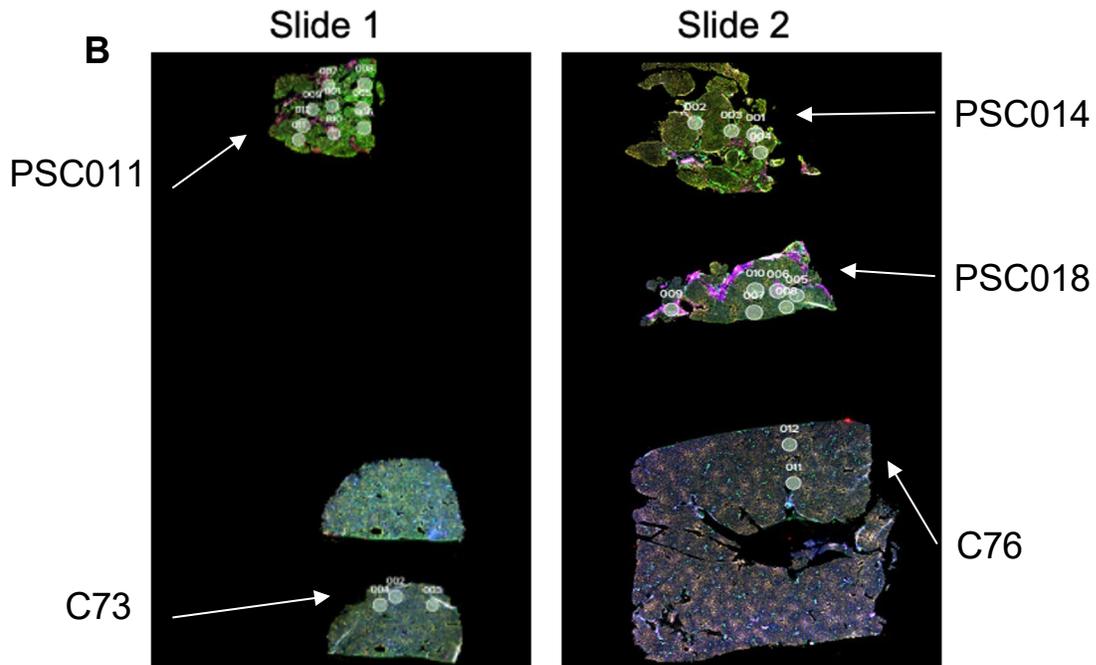
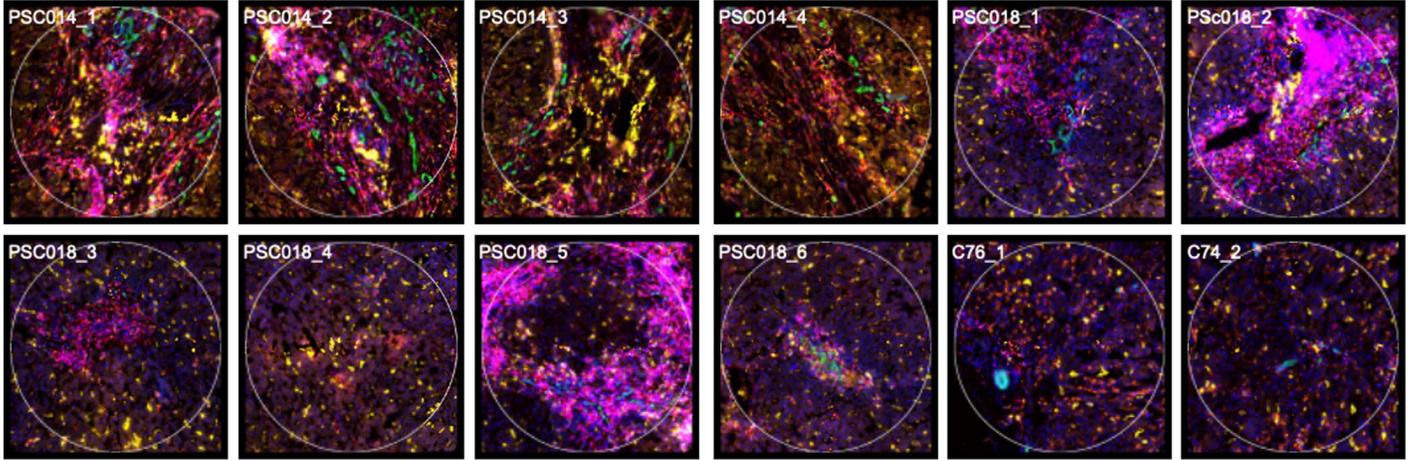
**Fig. S13: Disease-associated pathways in sc-RNAseq.** (A) Significantly enriched pathways among genes differentially expressed between PSC and NDD livers (B) Significantly enriched pathways among genes differentially expressed between PBC and NDD livers (C) Pathways from A filtered to remove those also found in B. (D) Pathways from B filtered to remove those also found in A.

**A**

Slide 1

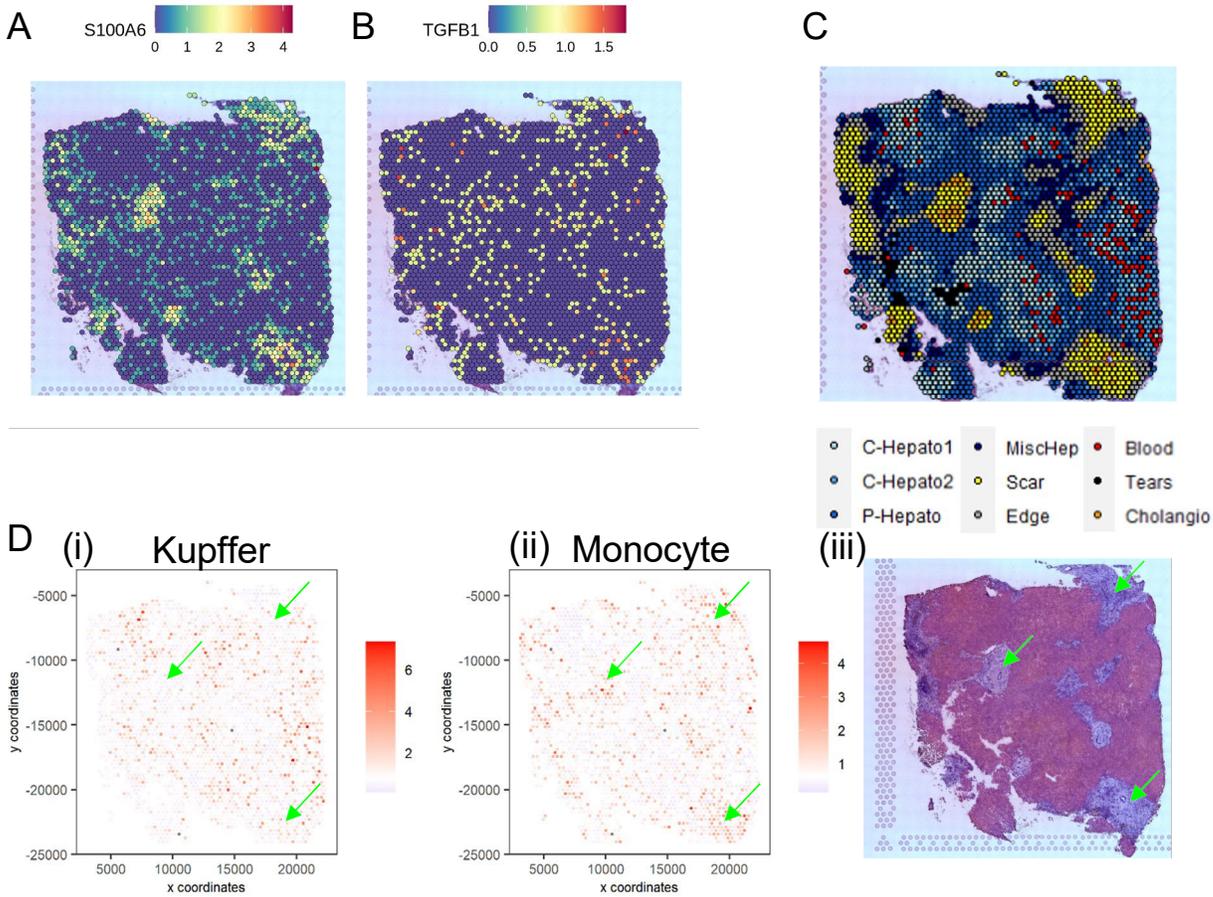


Slide 2

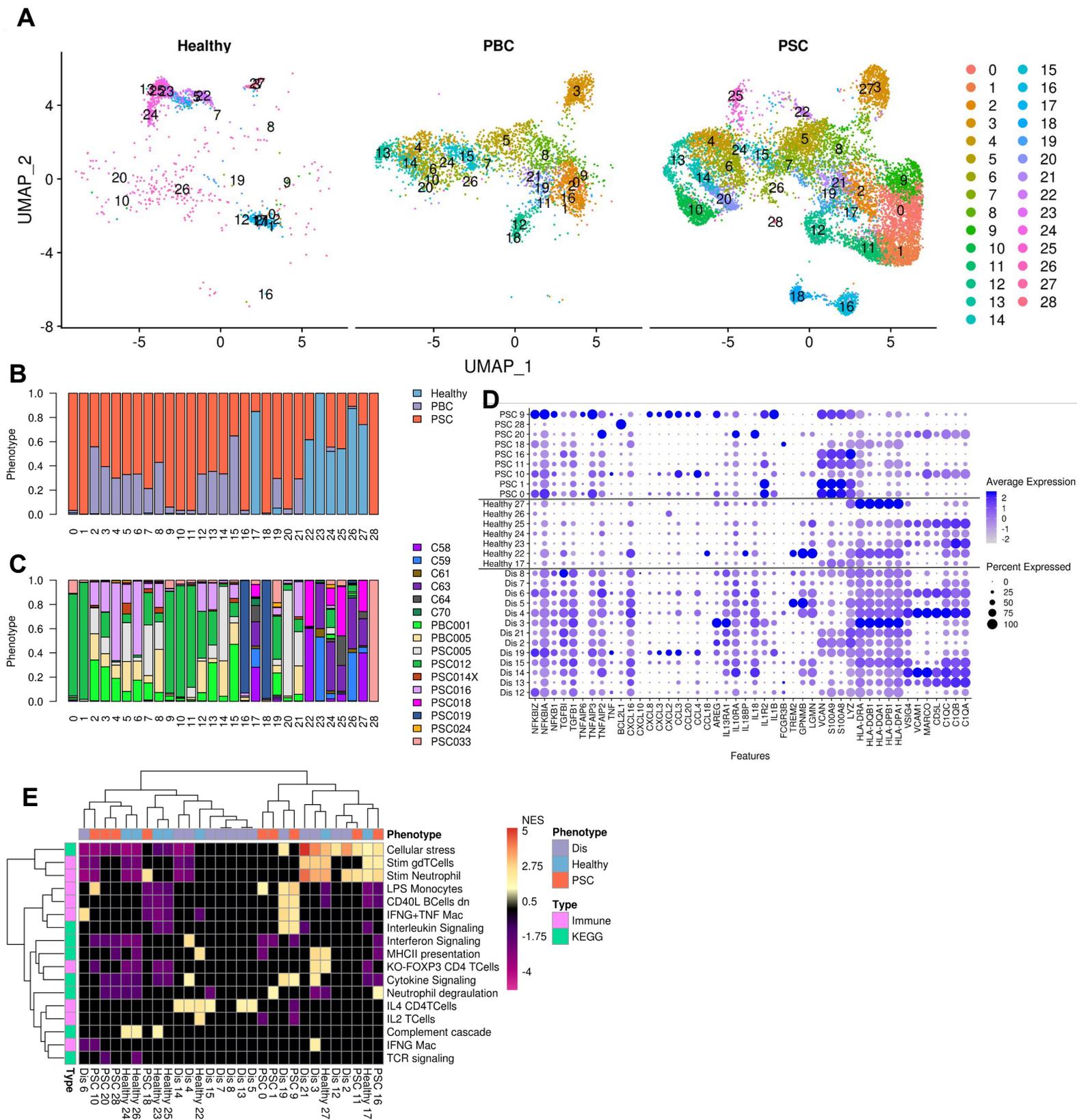


**Nuclei (DAPI) Epithelium (KRT7) Immune (CD45) Macrophage (CD68)**  
 3 PSC, 2 NDD (2 Slides, 12 ROI per slide, and total 24 ROIs)

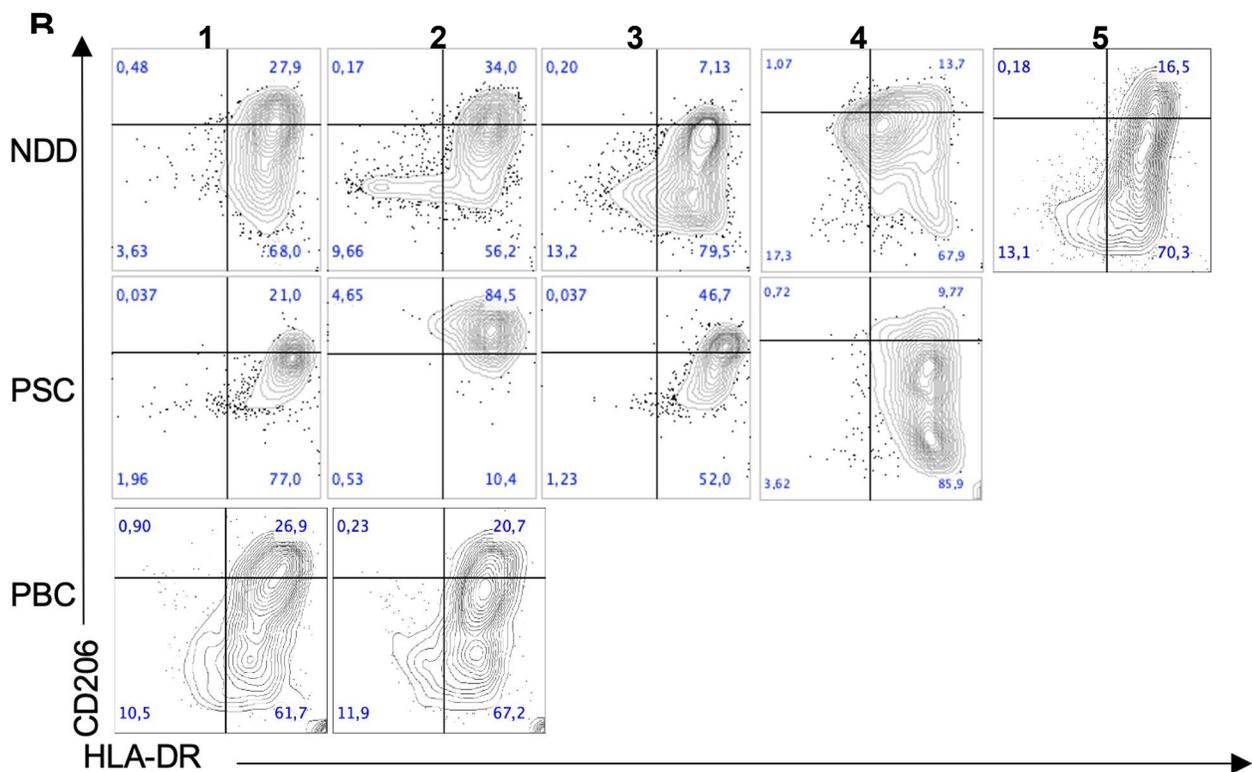
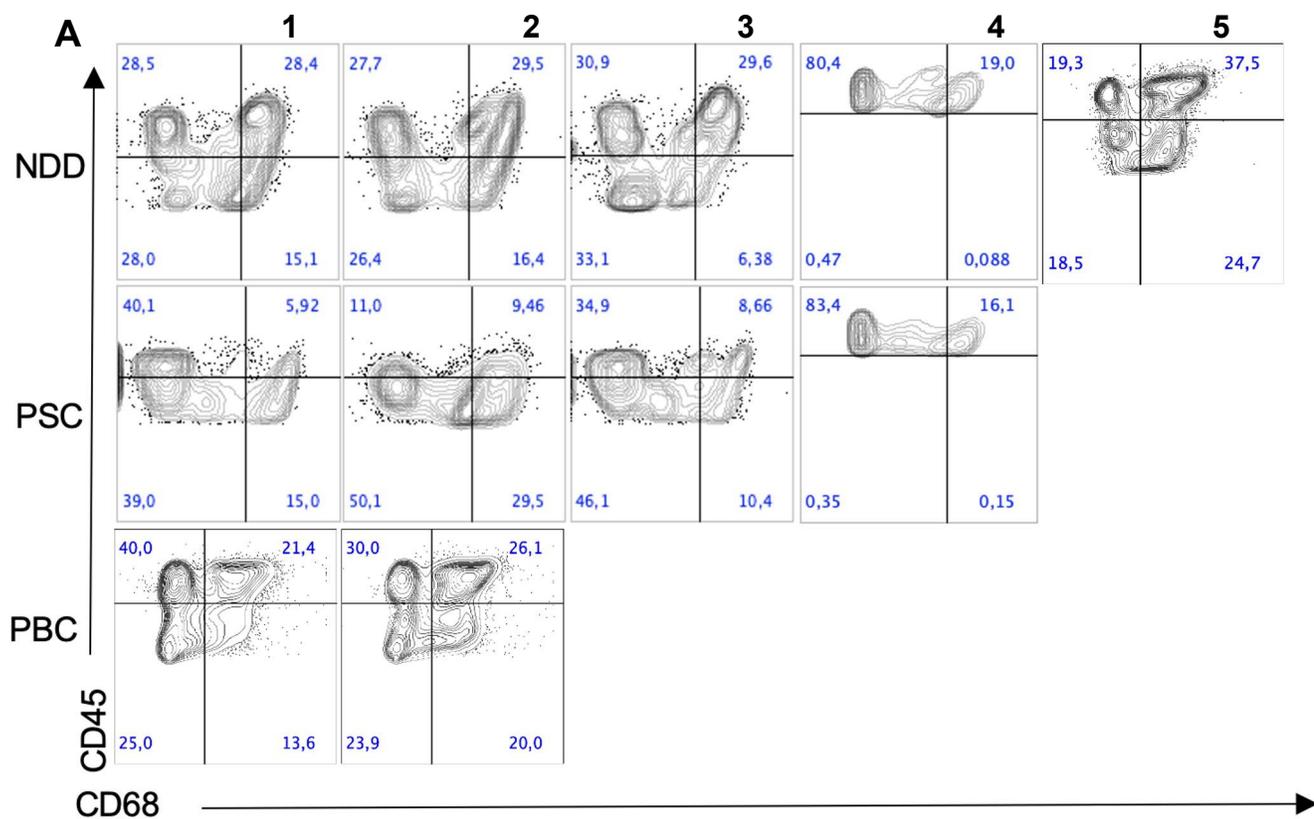
**Fig. S14: 24 Nanostring Regions of Interest (ROIs).** Images from the 24 regions of interest (ROIs) selected from the Nanostring GeoMx Digital Spatial Profiling platform. (A) PSC and NDD regions of interest in Slide 1 (PSC011\_1-9, C71\_1-4) and Slide 2 (PSC014\_1-4, PSC018\_1-6, C76\_1-2). (B) Entire scanned Slide 1 (PSC011, C73) and Slide 2 (PSC014, PSC018, C76). White circles are 660  $\mu\text{m}$  in diameter.



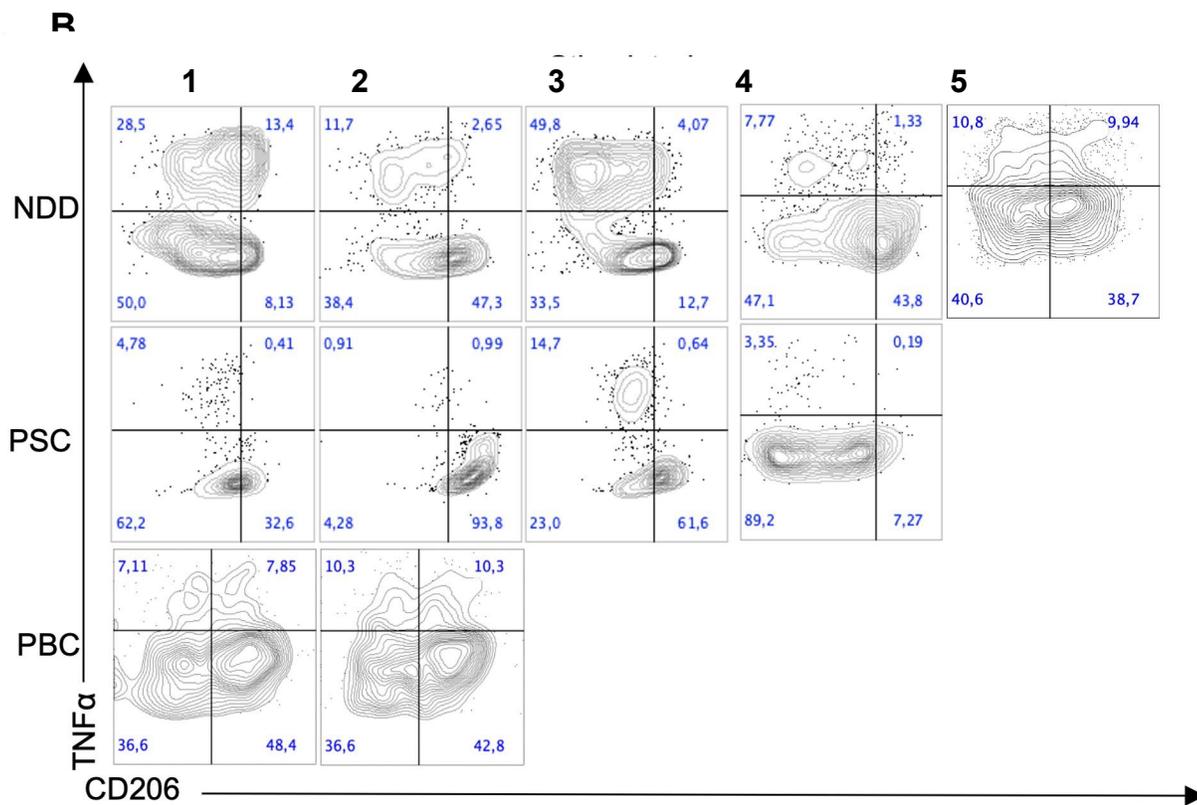
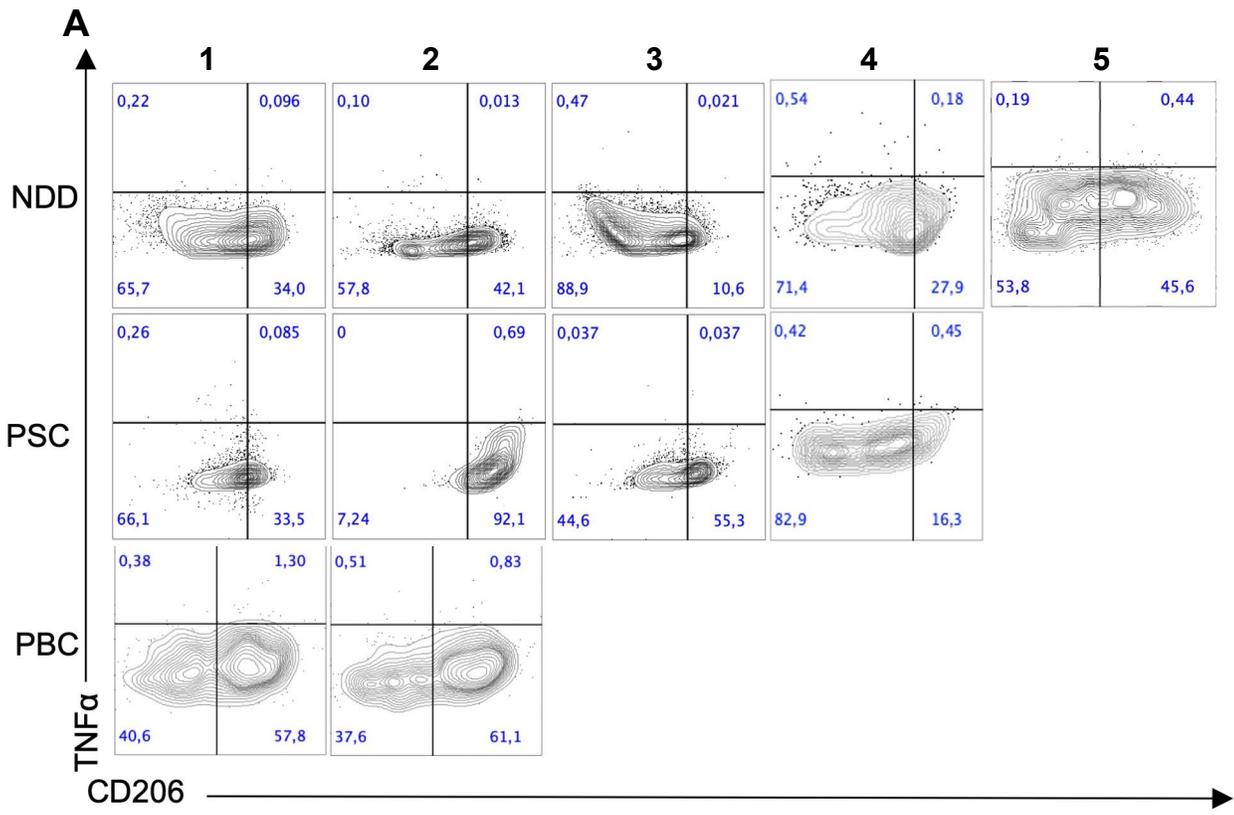
**Fig. S15. TGFβ is expressed in large scars where it co-localizes with Monocyte-like macrophages.** (A) Expression of the Monocyte-like Macrophage marker S100A6 in VISIUM spatial transcriptomics of PSC liver. (B) Expression of TGFβ. (C) Annotated clusters in PSC. (D) Macrophage signature enrichments as calculated using Giotto, using gene signatures for Kupffer (i) and Monocytes (ii) from Supplementary Table 2 within spatial transcriptomics of PSC liver. (iii) H&E-stained tissue. Green arrows indicate scar areas.



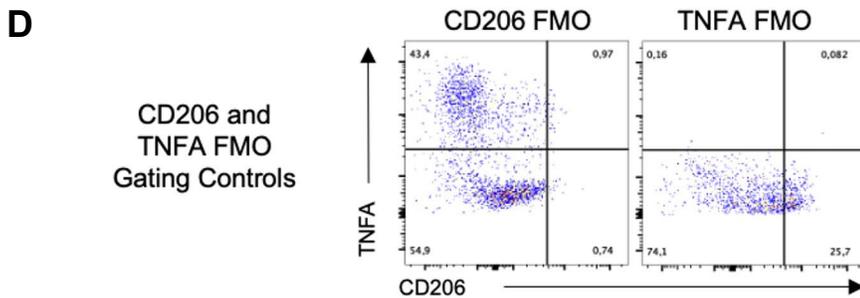
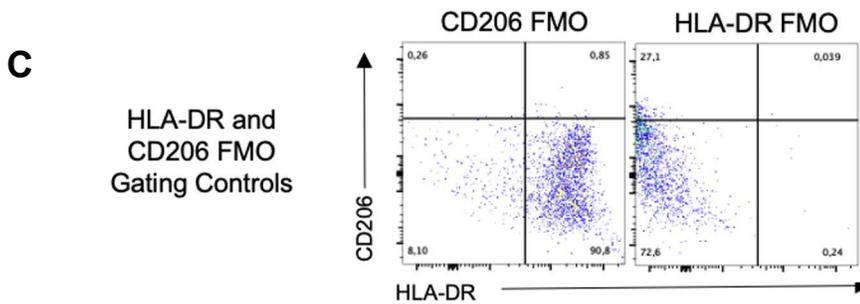
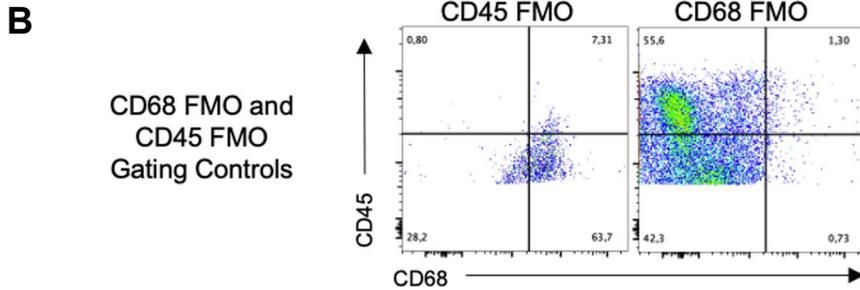
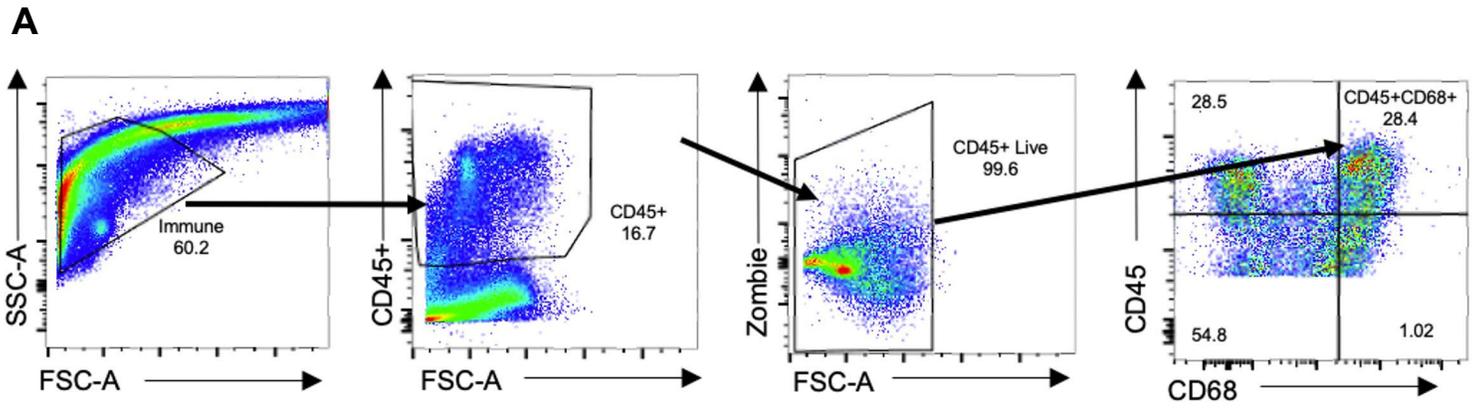
**Fig. S16: Subclustering reveals macrophage diversity in PSC.** (A) Significantly enriched pathways among genes differentially expressed between PSC and NDD livers. (B) Significantly enriched pathways among genes differentially expressed between PBC and NDD livers. (C) Pathways from A filtered to remove those also found in B. (D) Pathways from B filtered to remove those also found in A.



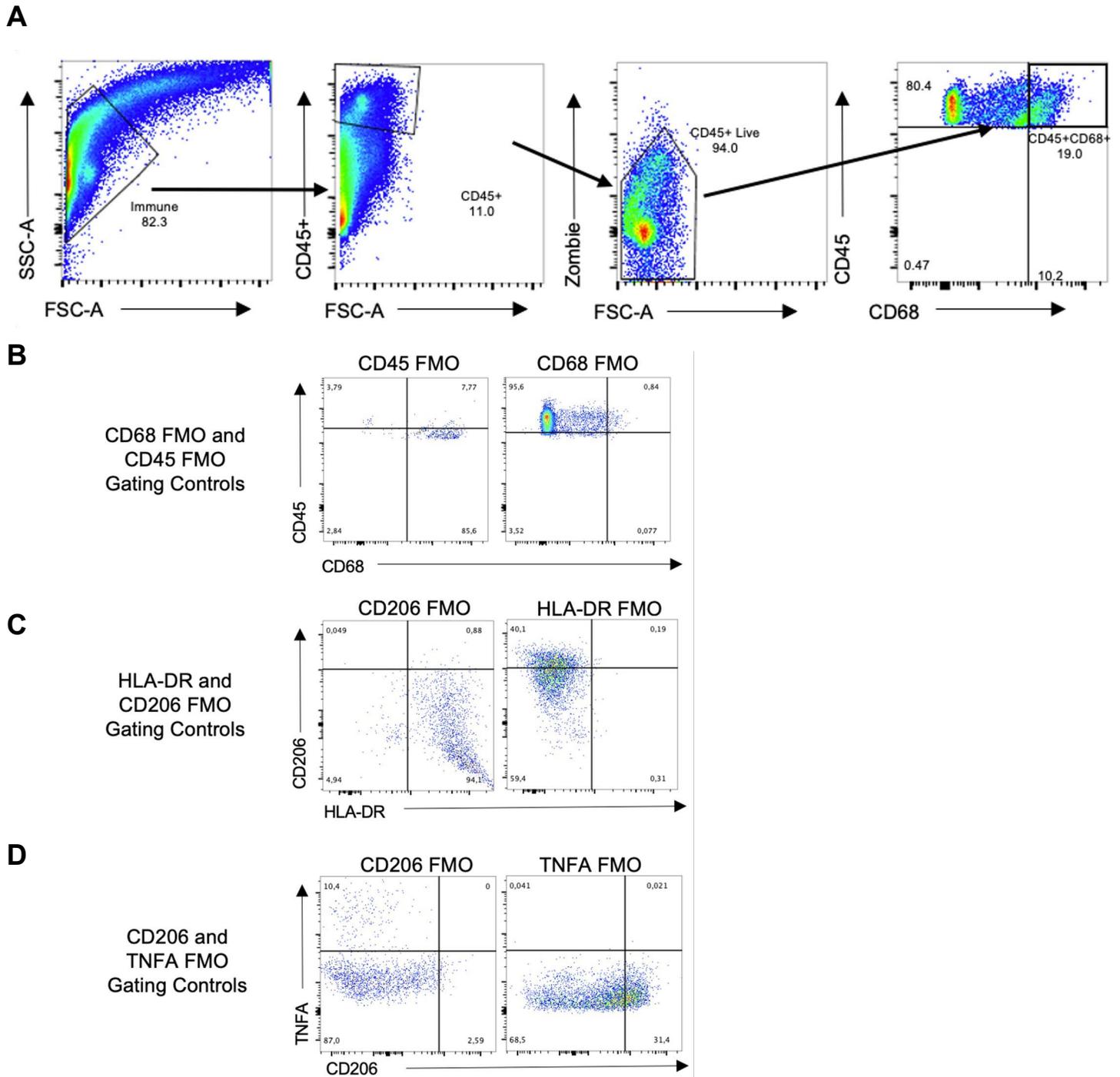
**Fig. S17: Expanded CD45/CD68 and HLA-DR/CD206 flow cytometry plots.** A) Flow cytometry plots of CD68 vs. CD45 for NDD (n=5), PSC (n=4) and PBC (n=2) depicted in Figure 6A-B. B) Flow cytometry plots of HLA-DR vs. CD206 for NDD (n=5), PSC (n=4) and PBC (n=2) depicted in Figure 6D-E.



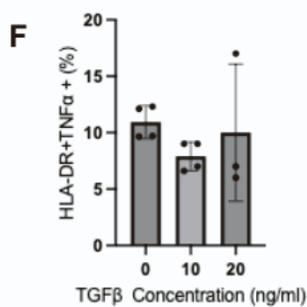
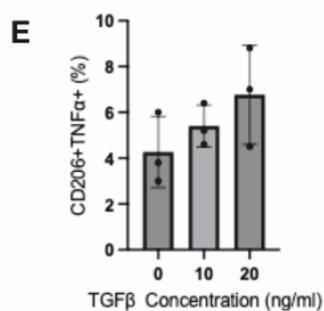
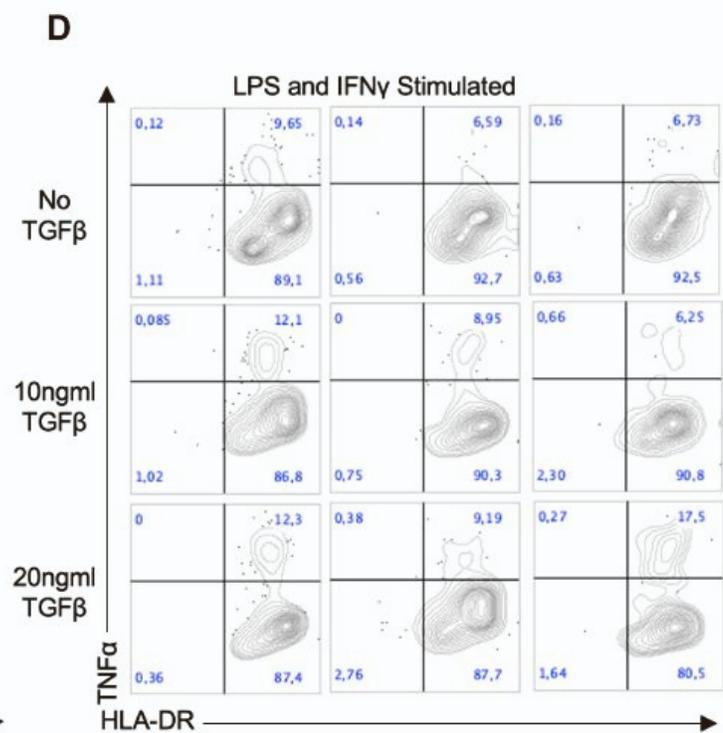
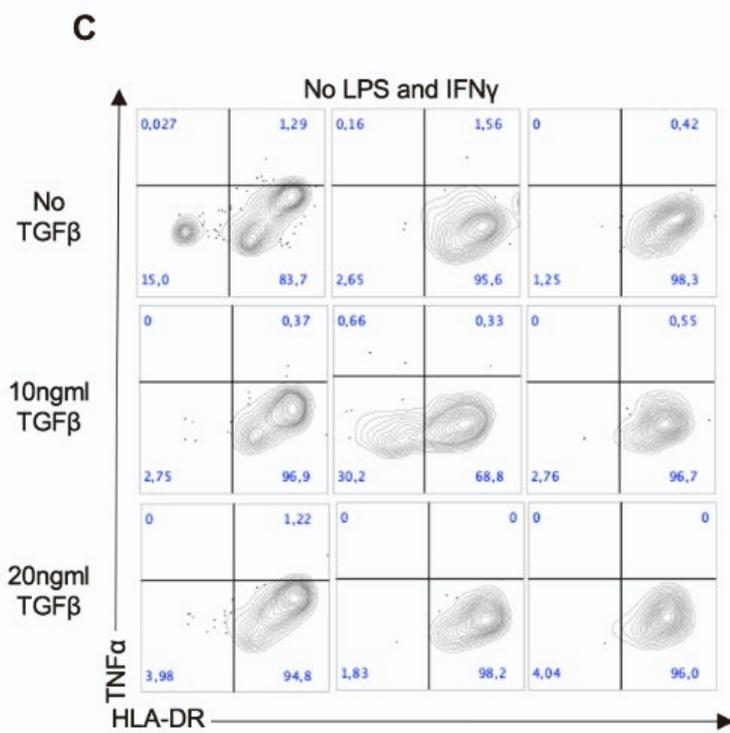
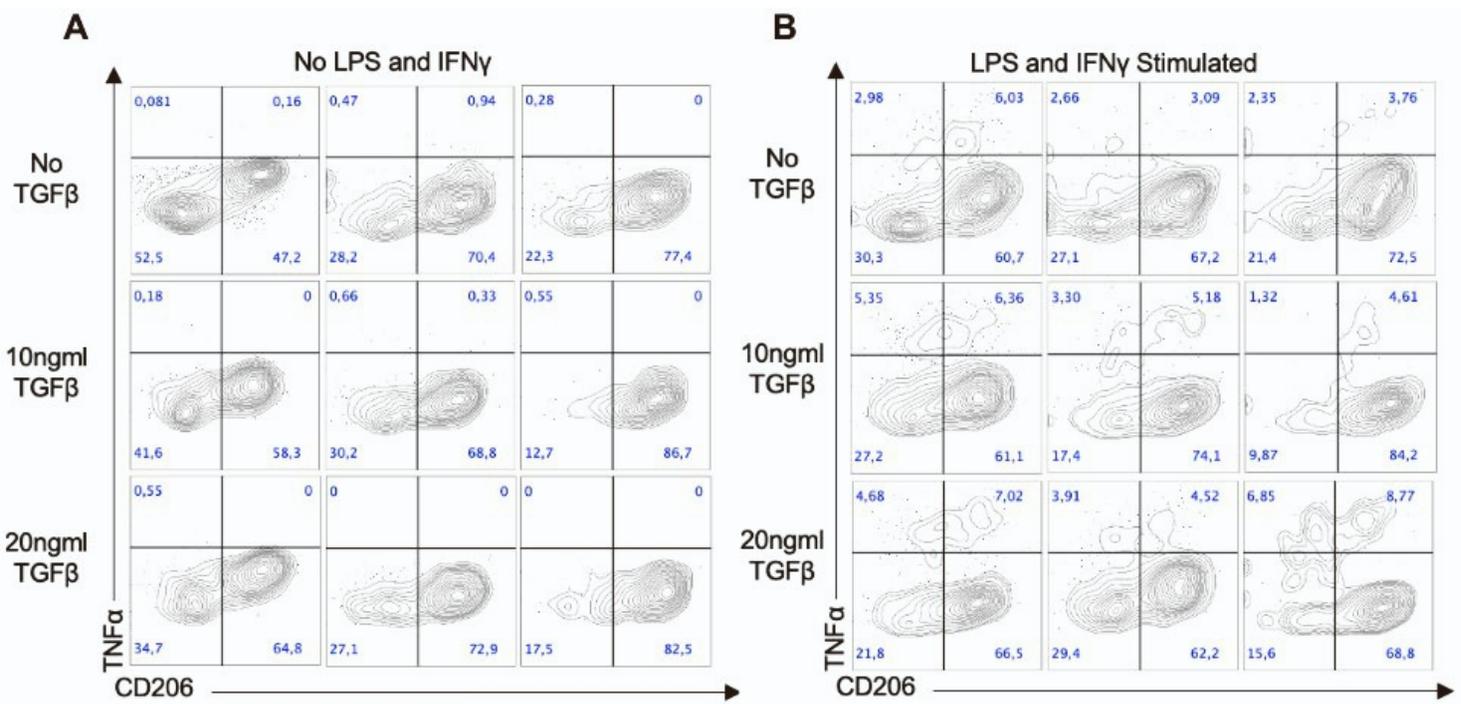
**Fig. S18: Expanded CD206/TNF $\alpha$  flow cytometry plots.** A) Flow cytometry plots of CD206/TNF $\alpha$  for unstimulated NDD (n=5), PSC (n=4) and PBC (n=2) TLH. B) Flow cytometry plots of CD206/TNF $\alpha$  for LPS and IFN $\gamma$  stimulated NDD (n=5), PSC (n=4) and PBC (n=2) TLH depicted in Figure 6F-G.



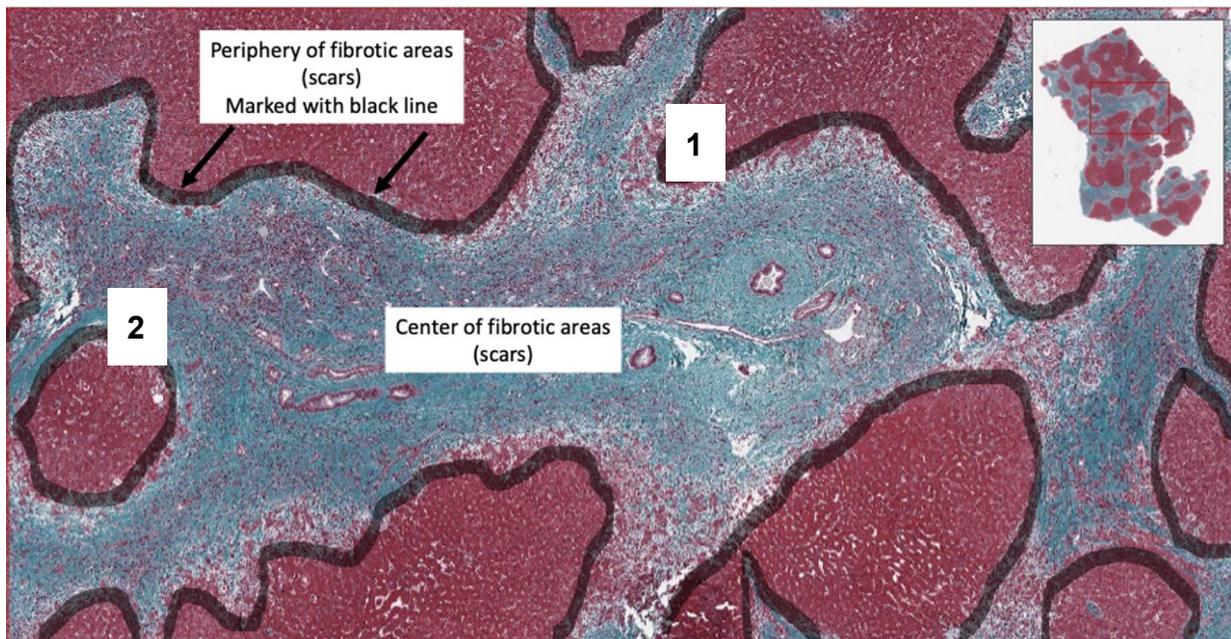
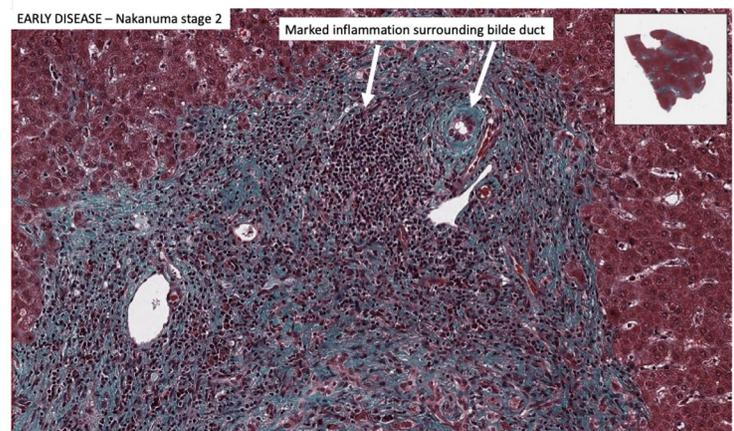
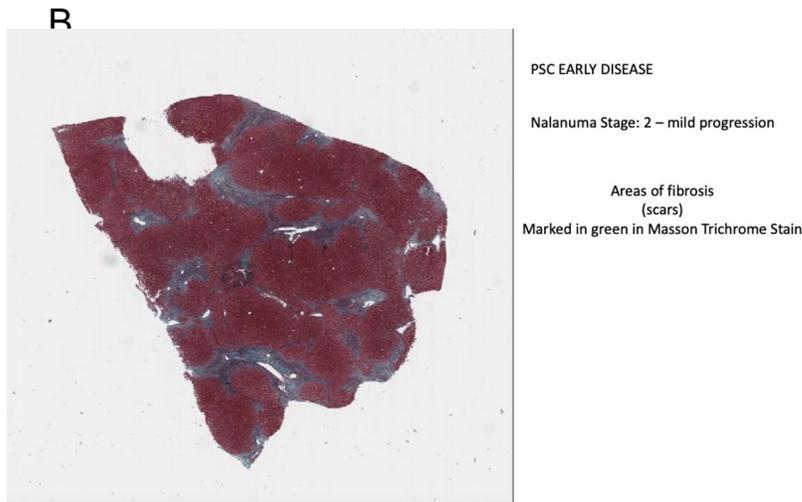
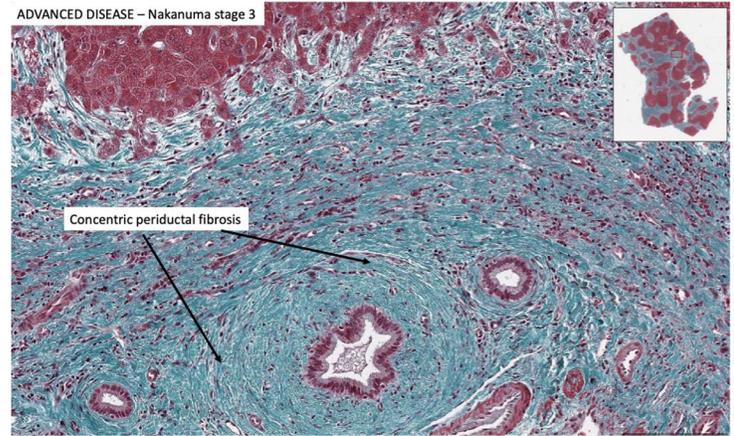
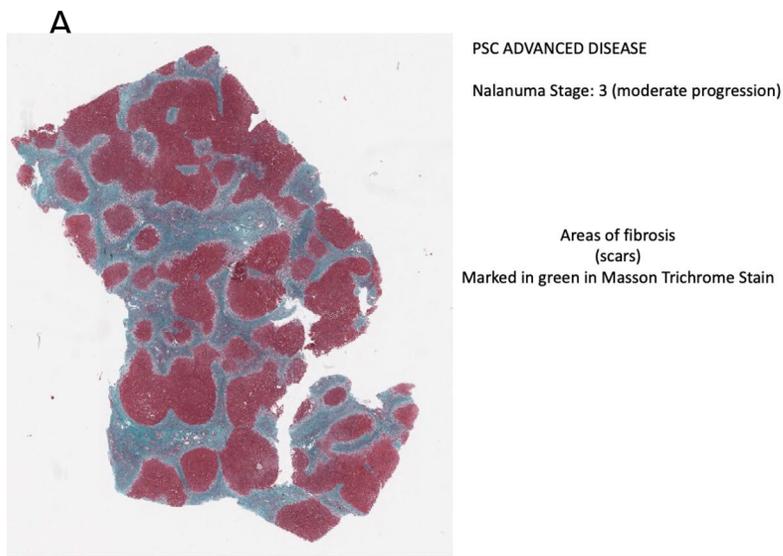
**Fig. S19: Gating Strategies and FMOs** (fluorescence minus one) for flow cytometry immunophenotyping and intracellular cytokine staining of NDD (1-3) and PSC (1-3) TLH in Extended Data Dig. 17 and 18. A) Representative gating strategy of NDD sample. Gating on the immune fraction in a FSC-A vs SSC-A plot, followed by a gate include CD45<sup>+</sup> cells in an FSC-A vs CD45 plot, followed by live cells based on a live/dead Zombie stain in an FSC-A vs Zombie stain, and then gating on CD45<sup>+</sup>CD68<sup>+</sup> cells for macrophage analysis. B) CD45 and CD68 FMO plots used for the gating of CD68<sup>+</sup>CD45<sup>+</sup> cells. C) CD206 and HLA-DR FMOs used to gate on CD206<sup>+</sup> and HLA-DR<sup>+</sup> cells. C) CD206 and TNF $\alpha$  FMOs used to gate on CD206<sup>+</sup>TNF $\alpha$ <sup>+</sup> cells.



**Fig. S20: Gating Strategies and FMOs** (fluorescence minus one) for flow cytometry immunophenotyping and intracellular cytokine staining of NDD (4-5), PSC (4), and PBC (1-2) TLH in **Extend Data Fig. 17 and 18**. A) Representative gating strategy of NDD sample. Gating on the immune fraction in a FSC-A vs SSC-A plot, followed by a gate include CD45<sup>+</sup> cells in an FSC-A vs CD45 plot, followed by live cells based on a live/dead Zombie stain in an FSC-A vs Zombie stain, and then gating on CD45<sup>+</sup>CD68<sup>+</sup> cells for macrophage analysis. B) CD45 and CD68 FMO plots used for the gating of CD68<sup>+</sup>CD45<sup>+</sup> cells. C) CD206 and HLA-DR FMOs used to gate on CD206<sup>+</sup> and HLA-DR<sup>+</sup> cells. C) CD206 and TNF $\alpha$  FMOs used to gate on CD206<sup>+</sup> TNF $\alpha$ <sup>+</sup> cells.

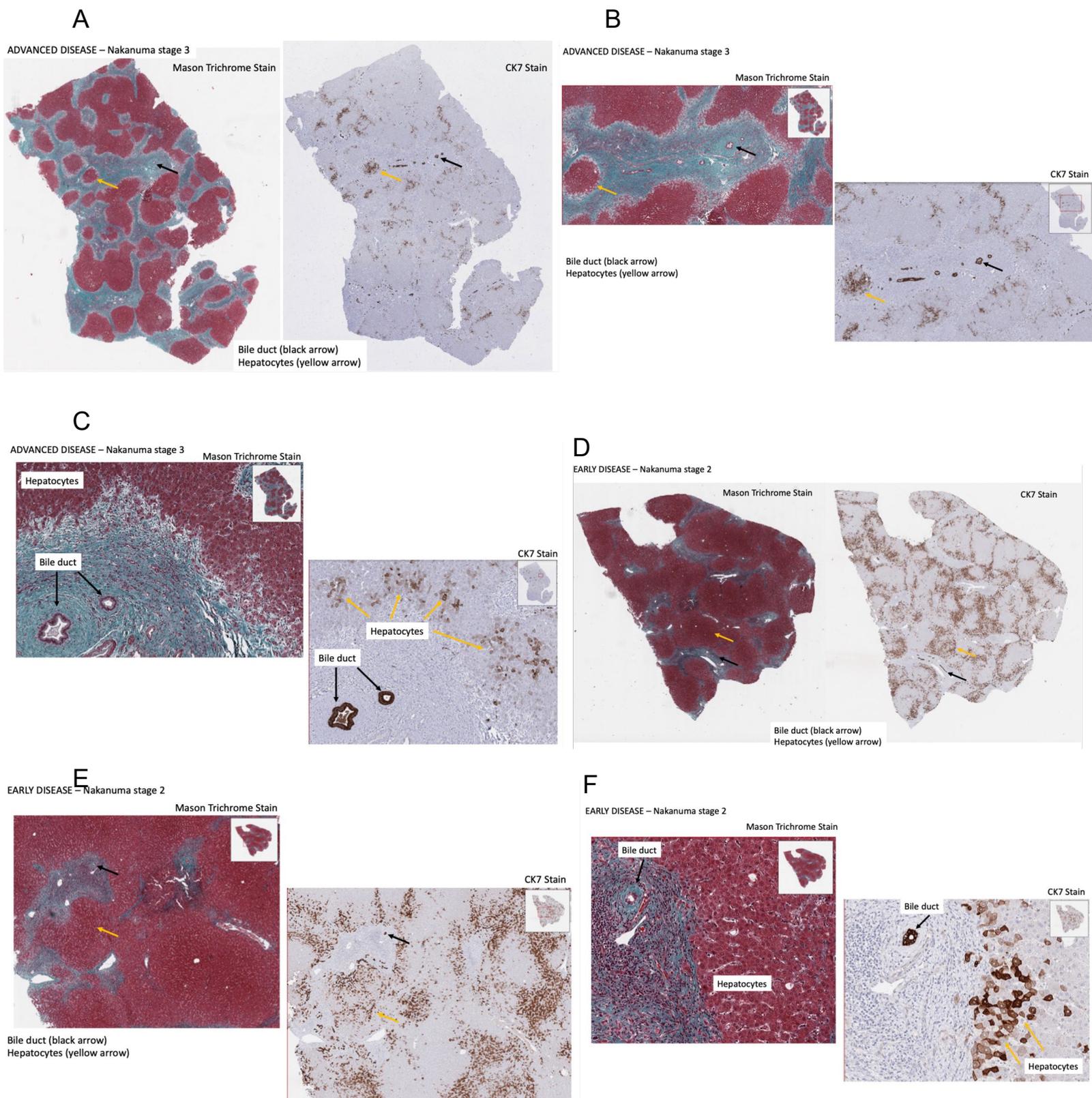


**Fig. S21: TGF $\beta$  incubation does not alter stimulation potential of NDD-derived myeloid populations.** Flow cytometry plots of CD206/TNF $\alpha$  for NDD TLH incubated for 24 hours with 0 ng/ml TGF $\beta$  (n=3), 10 ng/ml TGF $\beta$  (n=3) and 20 ng/ml TGF $\beta$  (n=3) A) without LPS and IFN $\gamma$  stimulation, and B) with LPS and IFN $\gamma$  stimulation. Flow cytometry plots of HLA-DR/TNF $\alpha$  for NDD TLH incubated for 24 hours with 0 ng/ml TGF $\beta$  (n=3), 10 ng/ml TGF $\beta$  (n=3) and 20 ng/ml TGF $\beta$  (n=3), C) without LPS and IFN $\gamma$  stimulation, and D) with LPS and IFN $\gamma$  stimulation. E) Percentage of double positive population CD206 $^+$ TNF $\alpha$  $^+$  across treatment conditions in B. F) Percentage of double positive population HLA-DR $^+$ TNF $\alpha$  $^+$  across treatment conditions in D. Gating strategy employed is described in Extended Data Fig 19 and 20.



**Fig. S22. Disease stage according to fibrosis was defined by the Nakanuma score and stage (see reference) assessing the following components: Fibrosis and Bile Duct Loss.** A) PSC advanced disease: Masson Trichrome stain assessed the extent of fibrosis and identified areas of scarring (stained in green) throughout the liver parenchyma and around individual bile ducts (concentric periductal fibrosis). B) PSC early disease: Early disease PSC livers showed less areas of fibrosis and prominent mononuclear cell inflammation surrounding individual bile ducts. C) Annotation of fibrotic areas (scars) within the liver parenchyma of explanted

PSC tissue based on Masson Trichrome stain: The fibrotic areas (highlighted in green in Masson Trichrome Stain) are divided into a central and peripheral zone. The peripheral zone is defined as either the interphase between periportal and zone 1 hepatocytes (1) or the periseptal zone of regenerative nodules (2) in cirrhotic liver parenchyma.



**Fig. S23. Cytokeratin 7 (CK7) immunohistochemical stain highlights hepatocytes with biliary metaplasia as a feature of chronic biliary disease within fibrotic areas (scars) in PSC explant livers.** Bile duct epithelium is marked by a black arrow and metaplastic hepatocytes are marked by a yellow arrow, mainly at the periphery of fibrotic areas (scars). (A)PSC advanced disease: Low power view of Masson Trichrome Stain and immunohistochemical stain for CK7. (B) PSC advanced disease: Intermediate power view of Masson Trichrome Stain and immunohistochemical stain for CK7. (C) PSC advanced disease: High power view of Masson Trichrome Stain and immunohistochemical stain for CK7. (D) PSC early disease: Low power view of Masson Trichrome Stain and immunohistochemical stain for CK7. (E) PSC early disease: Intermediate power view of Masson Trichrome Stain and immunohistochemical stain for CK7. (F) PSC early disease: High power view of Masson Trichrome Stain and immunohistochemical stain for CK7.

## Supplementary table legends

### [Table S1](#). Summary of Samples & Sequencing

[Table S2](#). **Cell type specific markers identified in the NDD Liver map.** These were determined through two-layer hierarchical clustering, first by clustering data into major cell-types as depicted in Figure 1, then by subclustering each of these clusters as shown in Fig. S2-5. Marker genes for subclusters were determined in comparison to the other cells in the same coarse level cluster, not in comparison to the entire map. Log2 fold-changes, P-values, and cluster information in this table presents the differential expression results for the most significant cluster when comparing the expression across all 38 cell-types in the entire map. Marker genes were selected based on specificity to each cell-type and expert knowledge in the field.

[Table S3](#). **Summary of contributions of each sample to each cluster in the NDD Liver Map as well as PSC/PBC maps.**

[Table S4](#). **Cell-type specific differential expression between PSC or PBC and NDD using 5' scRNAseq.** DE was performed using pseudobulks and edgeR. Negative logFCs indicate genes upregulated in PSC/PBC and Positive logFCs indicate genes downregulated in PSC/PBC. Only genes passing 5% FDR multiple testing correction are included.

[Table S5](#). **Cell-type specific differential expression between PSC and NDD using 3' snRNAseq.** DE was performed using pseudobulks and edgeR. Negative logFCs indicate genes upregulated in PSC and Positive logFCs indicate genes downregulated in PSC. Only genes passing 5% FDR multiple testing correction are included.

[Table S6](#). **All cell-type markers from 5' PSC/PBC/NDD scRNAseq.** DE was performed using wilcox-rank-sum test.

[Table S7](#). **All cell-type markers from 3' PSC/PBC/NDD snRNAseq.** DE was performed using wilcox-rank-sum test.