Single-cell transcriptomics reveals the impact of sex and age in the healthy human liver

Raza Ur Rahman[1,2], Eliana T. Epstein[3], Shane Murphy[2,4], Liat Amir-Zilberstein[2,4], Cristin McCabe[2,4], Toni Delorey[2,4], Hope Koene[3], Lilly Fernandes[3], Kenneth K. Tanabe[5], Motaz Qadan[5], Christina Ferrone[5], David L. Berger[5], Angela Shih[6], Jacques Deguine[2,4], Alan C. Mullen[1,2,†]

[1]Division of Gastroenterology, University of Massachusetts Chan Medical School, Worcester, MA

[2]Broad Institute of MIT and Harvard, Cambridge, MA

[3]Division of Gastroenterology, Massachusetts General Hospital, Boston, MA

[4]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA

[5]Division of Gastrointestinal and Oncologic Surgery, Harvard Medical School and Massachusetts General Hospital, Boston, MA

[6]Division of Pathology, Massachusetts General Hospital, Boston, MA

[†]Corresponding author: 304 Lazare Research Building, University of Massachusetts Chan Medical School, Worcester, MA 01605, USA. (508) 856-4005. Email: alan.mullen@umassmed.edu

One Sentence Summary: We applied single-nucleus RNA sequencing to define features influenced by sex and age across cell types and subtypes of the human liver.

Abstract

The liver is a vital organ composed of parenchymal, nonparenchymal, and immune cell populations, which are now being mapped at single cell resolution. Here we present a detailed map of the cellular composition of the human liver through single-nucleus RNA sequencing (snRNA-seq), incorporating insights from 32 healthy liver samples. The dataset contains cells from female and male donors spanning more than seven decades of age. Excluding the X and Y chromosome, we identified ~500 genes uniquely enriched in cells of the female liver, and ~600 genes enriched in cells of the male liver. These genes tend to be enriched across multiple cell types in females and are more restricted to individual cell types in males. These genes define unique circuitries within each cell type of the liver between females and males and their impact on cell-cell communication and response to external signals, including enrichment of cholesterol/lipid metabolism and associated angiopoietin-like protein (ANGPTL) signaling in female cells and bone morphogenic protein (BMP) signaling in male cells. With increased age, we observe a greater diversity in gene expression, including enrichment of genes that regulate neuregulin (NGR) signaling. While fewer genes are enriched within cell types of the younger liver, genes regulating insulin growth factor (IGF) signaling are in this group. Together, these results define the activities of healthy cell types within the liver across sex and age and provide a reference for future comparisons across different ancestries, geographies, and disease states.

INTRODUCTION

The liver plays a pivotal role in numerous physiological processes, including glucose and lipid metabolism, coagulation, iron homeostasis, bile secretion, and immune interactions. Dissecting the complex cellular composition of the liver and the influence of factors such as sex and age is essential to translate study findings in both health and disease to more specific patient populations. Advancements in single-cell RNA sequencing (scRNA-seq) technologies have provided unprecedented insights into the cellular landscape of the liver (*1–4*). The introduction of single-nucleus RNA sequencing (snRNA-seq) has also allowed for broader representation of non-immune cell types in the liver compared to scRNA-seq (*5*), as these cell types may be more sensitive to enzymatic digestion. While extensive research has been conducted to examine the liver at the single cell resolution (*1–4*), snRNA-seq data for human liver (*5*) remain limited, and larger datasets are necessary to understand the full extent to which factors such as sex and age affect the function of the diverse cell types and subtypes within the liver.

Sex and age impact liver function and disease progression, but there is limited insight into how these factors influence the liver at the single cell level. Sex-based variations in enzyme activity, gene expression, and steroid hormone sensitivity can influence the ability of the liver to metabolize drugs and hormones (*6–8*). Chronic liver diseases such as fibrosis and cirrhosis are more prevalent and severe in men than in women (*9–11*), and for this reason male mice are used more frequently in models of in vivo liver injury (*12–14*) While single cell analyses have defined the transcriptional activity of individual cell types, cell-cell interactions, and how gene expression and cell interactions change in disease (*2*, *15–18*), our understanding of sex-specific pathways at this resolution remains limited.

Aging is also a well-established risk factor for chronic disease, including liver disease, and the global population is aging rapidly. In 2020, the number of people aged 60 and older surpassed

those under age 5. Furthermore, the percentage of individuals aged 60 and above is projected to nearly double between 2015 and 2050, increasing from 12% to 22% of the global population (https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, October 2024).

Within the liver, aging leads to changes in gene expression, cell function, and morphology (*19–24*) and is linked to reduced regenerative capacity, increased susceptibility to fibrosis, and greater vulnerability to conditions including MASLD, metabolic-dysfunction associated steatohepatitis (MASH), and cirrhosis. Elderly patients with MASLD have a higher prevalence of inflammation/MASH and advanced fibrosis compared to younger counterparts (*25*). One mechanism that may promote the increased prevalence of MASLD with age is cellular senescence, and clearance of senescent hepatocytes can reduce liver steatosis/MASLD in aging and obesity models (*26*). Additional cell types, including endothelial cells and macrophages also show induction of senescence signatures with age (*27*). A more complete understanding of the changes that occur in each cell type and subtype with age will provide insights into the broader implications of liver aging, response to injury, and potential interventions to mitigate its effects.

Despite significant progress in understanding the impact of sex and age on liver function and health, our knowledge at the single cell level remains limited. Here, we present a comprehensive map of the cellular composition of the human liver through snRNA-seq, analyzing data from 32 healthy liver samples (>176,000 nuclei) across seven decades of life from both male and female donors. This study highlights how sex and age influence gene expression, enhances our understanding of cellular heterogeneity within the liver, and provides a foundation for future research on sex- and age-related disease progression.

RESULTS

Charting the human liver landscape with single-nucleus RNA sequencing

We performed snRNA-seq to construct a map of the healthy human liver across 32 samples from female and male donors, spanning more than seven decades of life (**table S1**). Over 176,000 nuclei were analyzed after establishing quality control metrics (**Materials and Methods**) and removing ambient RNA and doublets (**table S1**). The dataset includes 105,983 nuclei from 19 female donors (62.1%) and 64,675 nuclei from 13 male donors (37.9%) (**Fig 1A**). Age distributions across donors are as follows: seven samples from donors aged 18–39, containing 40,771 nuclei (23.9%) labeled as younger; 11 samples from donors aged 40–59, comprising 55,748 nuclei (32.7%) labeled as middle; and 14 samples from donors aged 60 and older, totaling 74,139 nuclei (43.4%) labeled as older (**Fig 1B, fig S1**). Nuclei from all samples were integrated, and clustering was performed at multiple resolutions (**fig S2**). A resolution of 0.7 with 22 clusters (0-21) was selected for downstream analysis due to its ability to capture biologically meaningful sub-clusters, as described below (**Fig 1C, table S2**).

As part of the cluster quality control process, we identified two clusters (11 and 19) that exhibited reduced *MALAT1* reads (**table S3 and fig S3**). *MALAT1* is a long noncoding RNA retained in the nucleus, allowing it to serve as an indicator of high-quality nuclei (*28*, *29*). These two clusters were excluded from downstream analysis because there was enrichment of nuclei with low *MALAT1* expression (fewer than three reads per nucleus).

Clusters were first annotated using the ScType tool (*30*) with liver as the target tissue. This broader annotation assigned clusters into populations of hepatocytes, endothelial cells, hepatic stellate cells (HSCs), cholangiocytes, macrophages, and non-macrophage immune cell populations (**Fig 1D**). Marker genes for each annotated cell type were determined using the

FindAllMarkers function, showcasing the unique gene expression profiles that characterize these populations (**Fig 1E, table S4**). For instance, the top five genes enriched in hepatocytes include cytochrome p450 family members *CYP3A4* and *CYP2B6,* along with *CP* and *TAT*. Endothelial cells are enriched in expression of genes that include the phospho-signaling molecules *FLT1, AKAP12,* and *PTPRB*, while genes most enriched in HSCs include *LAMA2, ADAMTSL1, and CDH19*, which are involved in the extracellular matrix or cell adhesion. For cholangiocytes, the most enriched genes include adhesion molecules *CTNND2* and *PKDH1* along with *ANXA4,* which is linked to exocytosis. Immune cells were separated into macrophages and non-macrophage populations. Macrophages were enriched in expression of the scavenger receptor *CD163*, the phagocytic factor *RAB31*, and *TBXAS1*, responsible for production of thromboxane *A2*. Non-macrophage immune cells as a group show increased expression of genes including *SKAP1* and *CD247 and FYN*, which encode products involved in lymphocyte receptor signaling. Each cell type was represented in each sample, ensuring that downstream analyses included representation from every sample. (**Fig 1F, table S5**).

Evaluating hepatocyte zonation

We next classified hepatocytes into portal, central, and mid-zone based on the expression of key zonal markers: *ASS1*, *ASL*, *GLUL* and *CYP2E1 (31)*. We calculated the Central-to-Portal Ratio (CPR) by subtracting the summed expression of the markers *ASS1* and *ASL* (portal; zone 1) from the summed expression of *GLUL* and *CYP2E1* (central; zone 3) for each hepatocyte sub-cluster (**Fig 1C**). This ratio provides a quantitative measure of the zonal distribution. Clusters were categorized as portal if their CPR values were less than or equal to the lower quantile (clusters 3, 4, 6, 8), central if their CPR values were greater than or equal to the upper quantile (clusters 1, 5, 17, 20), and mid-zone for in between values (cluster 0, 2) (**Fig 2A, table S6**). This classification was validated with UMAP visualization and confirmed with density plots,

demonstrating distinct zonal distributions (**Fig 2B and 2C**). Portal, mid-zone, and central hepatocytes were then annotated within the overall UMAP (**Fig 2D**).

Defining non-hepatocytes cell types and subtypes

We next evaluated each individual non-hepatocyte cluster to define cell types and subtypes (**Fig 1C**). We identified sinusoidal, venous vascular (VV), and arterial vascular (AV) endothelial cells (*32*). HSCs were separated into quiescent HSCs (qHSCs) and portal fibroblasts (*33*), and a population of vascular smooth muscle cells (VSMCs) was annotated. Macrophages were separated into Kupffer-like cells (*CD74*^{high}, *MARCO*^{high}, *CD5L*^{high}) and inflammatory macrophages (*CD74*^{high}, *LYZ*^{high}, *MARCO*^{low}) (*1*) (**Fig 2D-F, table S7**). Each of these cell types was evenly distributed across samples (**fig S4**). These annotations highlight the cellular heterogeneity detected within these populations, indicating functional and phenotypic diversity captured within this dataset across sexes and over many decades of age.

The immune system cell population identified by this analysis (5,239 cells) was then sub-clustered from the raw data using the same analysis pipeline and annotated with the ScType tool using Immune System Cells as the target tissue to define subtypes of non-macrophage immune cells. This approach identified two populations of T cells and two populations of NK T-like cells. (**Fig 2H, table S8** and **Supplementary Materials**). This sub-clustering provides greater resolution of the non-macrophage immune cell landscape within the liver, revealing the diversity of immune system components in this organ defined by snRNA-seq.

Effect of sex on differential gene expression and biological pathways by cell type and subtype

This dataset is composed of cells from 19 females and 13 males, with balanced representation of cell types across sexes **(fig S5**). Cell type proportions were evaluated (*34*) (**Materials and Methods**) and no significant differences were identified between females and males, supporting differential expression analysis to uncover sex-specific gene expression differences (**table S9**). AV endothelial cells and VSMCs were excluded due to low cell counts (601/250 in females, 249/89 in males), and non-macrophage immune cells were analyzed as a group due to small subgroup sizes (**table S9**).

Differential expression analysis across cell types and subtypes revealed 575 genes upregulated in females and 728 genes upregulated in males, with 78 genes demonstrating upregulation in both sexes but in different cell types or subtypes (**table S10**). These findings suggest that while most genes that change in expression are either enriched in female cell types or male cell types, a smaller fraction of genes can be increased in different cell types depending on whether the cells are female or male. When focusing exclusively on autosomal genes (excluding X and Y chromosomes), 479 genes were upregulated in females, 625 in males, and 78 genes were upregulated across both sexes, again in distinct cell types. We then focused on autosomal genes for the subsequent analyses. The top five most enriched genes in female and male cell types and subtypes were plotted (**Fig 3A**), highlighting genes that are enriched in specific cell types and subtypes as well as those shared across multiple categories.

We next examined the biological processes enriched in cell types based on differential expression between female and male cells (**Fig 3B, table S11**). Estrogen Metabolism (WP697) was enriched in all female hepatocytes with the strongest induction in mid-zone hepatocytes, which were also enriched in the Estrogen Receptor Pathway (WP2881). TGF-beta Receptor Signaling (WP560) was enriched in all hepatocytes, with the strongest enrichment also in mid-zone hepatocytes, which were also augmented in the TGF-beta Signaling Pathway (WP366),

along with sinusoidal endothelial cells. Ferroptosis (WP4313) was enriched across VV endothelial, portal fibroblasts, and macrophage populations, while Iron Metabolism Disorders (WP5172) was enriched in Kupffer cells alone. Multiple pathways related to lipid metabolism were increased in female cells. Metabolic Pathway of LDL, HDL and TG (WP4522), Adipogenesis (WP236), Familial Hyperlipidemia (FH) 2 (WP5109), and Lipid Particle Composition (WP3601) were enriched in hepatocyte populations as well as different immune cell populations, while enrichment of FH Type 3 (WP5110) was restricted to portal hepatocytes. (**table S11**).

In male cells, Urea Cycle and Metabolism of Amino Groups (WP497) was enriched in portal hepatocytes, while Chemokine Signaling (WP3929) and Orexin Receptor Pathway (WP5094) were enriched in portal fibroblasts. Glucocorticoid Receptor Pathway (WP2880) was enriched in inflammatory macrophages, whereas Hippo Signaling Regulation Pathway (WP4540) was enriched in both sinusoidal and VV endothelial cells (**table S11**).

There were also examples of common pathways enriched in different cell types or subtypes between females and males, including those involving farnesoid X pathway (WP2879) and VEGFA-VEGFR2 signaling (WP3888). In these examples, the farnesoid X pathway is most enriched in (non-macrophage) immune cells in females (composed primarily of T cells and NK T-like cells), while most enriched in mid-zone hepatocytes in males. The VEGFA-VEGFR2 signaling pathway is enriched in qHSCs and sinusoidal endothelial cells in females compared to VV endothelial cells in males. These findings highlight both female- and male-specific pathways enriched within cell types in the healthy liver and how even the same pathways may modulate the activity of different cell types between female and male individuals.

Impact of sex on cell-cell communication

Given the differences in gene expression and pathways enriched between female and male cells, we predicted signaling pathway strengths by assessing communication probabilities among cell groups using CellChat (v2). This revealed distinct pathways enriched in female and male cells, as well as shared pathways (**Fig 3D**, **table S12**).

Female cells showed enrichment in signaling networks that include ApoA, ANGPTL, FN1, and VISFATIN (**Fig 3D,** orange labels), while male cells were enriched in pathways including NCAM, VECAM, PECAM1/2, IGF, BMP, and PDGF (blue labels). Additional pathways are active but showed no significant sex-specific differences (**Fig 3D,** black labels). These findings suggest sex-based gene expression differences can influence cellular responses and outputs to identical stimuli.

We next visualized cell-cell communication across pathways in female and male cells (**table S12**). In females, ANGPTL signaling was enriched, with qHSCs and portal fibroblasts receiving signals from all hepatocyte zones (portal, mid-zone, and central), while in males qHSCs and portal fibroblasts only received signals from central hepatocytes (**Fig 3E**). Furthermore, cholangiocytes displayed no ANGPTL signaling in males but transmitted signals to sinusoidal and VV endothelial cells and portal fibroblasts in females. Female cells also displayed stronger ANGPTL signaling when pathways are active between the same cell types (**Fig 3E**). ApoA signaling is also highly enriched between female cells, with hepatocytes and nonparenchymal cells signaling to Kupffer cells in females but not males (**Fig 3F**). In contrast, BMP signaling was enriched in males, with the strongest connection from sinusoidal endothelial cells to qHSCs (**Fig 3G)**. Additional signaling from VV endothelial cells was only observed in males. The NCAM pathway was more active in males as well, signaling primarily from non-macrophage immune cells to sinusoidal endothelial cells without active signals identified in females (**Fig 3H**). These

results show that signaling pathways can operate at different strengths and involve different cell types within the female and male liver.

To identify molecules driving these cell-cell interactions, we next visualized ligand and receptor (L-R) pairs enriched in female and male cell types (**Fig 4A-D, table S12**). Pairs were excluded if the receptor-expressing cell type showed higher ligand expression than the ligand-expressing cell type, because we could not exclude that the primary signal was autocrine. We also removed L-R pairs that were enriched in both female and male livers but involved different cell types to simplify visualization. The analysis focused on hepatocytes and endothelial sub-clusters, as these cell types showed the strongest sex-specific gene expression differences. Full results are in **table S12**.

In females, enriched L-R pairs include the combination of ANGPTL1 (ligand) expressed by portal hepatocytes and ITGA1_ITGB1 (receptor complex) expressed by sinusoidal and VV endothelial cells, qHSCs, portal fibroblasts, and hepatocytes (**Fig 4A**). In addition, ANGPTL4 (ligand) from portal hepatocytes pairs with ITGA5_ITGB1 and CDH5 (receptors) in VV endothelial cells and SDC2 and SDC3 (receptors) in Kupffer cells, while APOA1 (ligand) pairs with ABCA1 (receptor) in Kupffer cells and hepatocytes, and VEGFA (ligand) pairs with FLT1, KDR, and FLT1_KDR (receptors) in sinusoidal and with FLT1 in VV endothelial cells.

In contrast, IGF1 (ligand) from portal hepatocytes pairs with IGF1R (receptor) on sinusoidal and VV endothelial cells in males (**Fig 4B**). PLG (ligand) pairs with PARD3 (receptor) in qHSCs, portal fibroblasts, cholangiocytes, and hepatocytes, while NTN1 (ligand) from portal hepatocytes interacts with DSCAM (receptor) on Kupffer cells and UNC5C (receptor) on qHSCs, portal fibroblasts, and Kupffer cells. These findings highlight the diversity of signaling involving hepatocytes between females and males including pathways involved in lipid and cholesterol

metabolism enriched in females (APOA and ANGPTL) and repair/regeneration (PLG and NTN1) in males.

Enrichment of L-R pairs in females involving signaling from sinusoidal endothelial cells include ANGPT2 (ligand) combined with TEK and ITGA5_ITGB1 (receptors) expressed by VV endothelial cells and COL4A1 (ligand) paired with ITGAV_ITGB8 (receptor complex) expressed by cholangiocytes (**Fig 4C**). In males, enriched L-R receptor pairs include BMP6 (ligand) expressed by sinusoidal endothelial cells and BMPR1A/B_BMPR2 (receptor complex) expressed by qHSCs, portal fibroblasts, and cholangiocytes and BMPRA1/AVCR1_BMPR2 (receptor complex) expressed by hepatocytes (**Fig 4D**). In addition, the combination of APP (ligand) expressed by sinusoidal endothelial cells and SORL1 (receptor) expressed by inflammatory macrophages, hepatocytes, and non-macrophage immune cells is also enriched in males. Taken together, the L-R analyses show how individual L-R pairs can differ between liver cell types and subtypes in the female and male liver, establishing different circuitry to respond to both normal physiological signals and those associated with injury and disease.

Influence of sex on protein networks

We next predicted differences in protein-protein interactions between female and male cells in the liver based on gene expression data through STRINGDB network analysis. We focused on networks within hepatocytes and endothelial cells for visualization and included all data in **table S13**. In female portal hepatocytes, we identified a network containing APOA1, APOC1, APOC3, APOE, LDLR (**Fig 4E**); this network was enriched in pathways involved in lipid particle composition (WP3601), metabolic pathways of LDL, HDL, and TG (WP4522), familial hyperlipidemia-related pathways (WP5108, WP5109, WP5110, WP5111, WP5112), and PPAR-alpha signaling (WP2878) (**Fig 4E**). We also identified a protein network in female portal

hepatocytes containing ESR1, RUNX1 FOS, AHR, and KLF6, which was enriched in components of estrogen (WP712) and aryl hydrocarbon receptor pathway (WP2586) (**Fig 4F**).

An additional a protein network containing CYP2B6, CYP3A4, ABCB1, ABCB4, and GSTA1 was also defined in female portal hepatocytes, which was enriched in the farnesoid X receptor (WP2879), pregnane X receptor (WP2876), and estrogen metabolism pathways (WP697) (**fig S6)**. In contrast, male portal hepatocytes contained a network involving heat shock proteins HSP90AA1 and HSPD1 and stress response protein FKBP5 (**Fig 4G**).

Female sinusoidal endothelial cells showed enrichment of a protein network containing ITGB1, FGA, FGB and FGG, proteins linked to VEGFA-VEGFR2 signaling (WP3888), blood clotting cascade (WP272) and fibrin complement receptor 3 signaling pathway (WP4136) (**Fig 4H**). An additional network containing APOB, IGFBP1, HLA-C, HLA-B, PIK3R1 and TGFBR2 also showed enrichment in female sinusoidal endothelial cells (**Fig 4H**), while a network containing IGF1R, FGFR1, and PRKCE, which play a role in Hippo signaling (WP4540), was enriched in male sinusoidal endothelial cells (**Fig 4I**). Together, these results highlight differences in protein-protein interaction networks between female and male cell types that can affect their intrinsic activity and response to external signals.

Applying machine learning to evaluate male and female cell populations

To determine if liver cells could be classified as female or male by machine learning, we applied a Random Forest classifier across all cell types using 80% of the data for training and 20% for testing. Two models were trained: one using all genes as a baseline (full model) and another excluding X and Y chromosome genes (autosomal model). The full model achieved an accuracy of 81.8% and recall of 99.5%, while the autosomal model still performed well, with 76.7%

accuracy and 99.0% recall (**Fig 4J**). We then evaluated classification performance for individual cell types and subtypes. Both models performed robustly, with the full model generally achieving slightly higher accuracy and recall. For example, portal hepatocytes were classified with 86.7% accuracy and 100% recall in the full model and 81.1% accuracy and 99.9% recall in the autosomal model (**Fig 4K**). Sinusoidal endothelial cells showed similar trends, with 77.2% accuracy and 99.5% recall in the full model and 73.4% accuracy and 99.1% recall in the autosomal model (**Fig 4L**). Other cell types also demonstrated robust classification performance (**fig S7A-I**). The autosomal model achieved similar accuracy to the full gene model across all cell types, demonstrating that cell type and subtype classification can be effectively performed using only autosomal gene expression (**table S14 and fig S7A-I**).

These results highlight the key role of autosomal genes in differentiating female and male cells. When the full model was run using all cell types, X- and Y-linked genes such as *XIST*, *UTY*, *TTTY14*, and *USP9Y* were among the top features, as expected. *XIST*, known for its critical role in X-chromosome inactivation in females (*35*), also exhibited the highest feature importance score. However, even after excluding X and Y chromosome genes, the classifier highlighted key genes contributing sex differences within the liver, including *CUX2*, which is expressed at higher levels in female livers in murine models (*36*), as well as *A2M*, found at higher plasma levels in females (*37*), alongside *ECHD2, LINC01554 (C5ORF27) and ALB*. These results highlight the abundance of differentially expressed X- and Y-linked genes in the female and male liver, while also reinforcing the role of autosomal genes in maintaining unique expression programs (**Fig 4M** and **table S14**).

Effect of age on differential gene expression and biological pathways by cell type and subtype

We next took advantage of the age distribution of our samples to investigate how gene expression changes with age across cell types and subtypes (**Fig 1B, fig S1**). Although all age groups (younger, middle, and older) showed similar representation of cell types **(fig S8)**, there were more female than male nuclei in each age group, and the ratio of female to male nuclei was not constant across groups (**fig S9**)(**table S15**). To limit the impact of sex in this analysis, we downsampled the female nuclei so that there were equal numbers of female and male nuclei within each cell type and age group (**fig S9** and **Supplementary Materials**). We then performed differential expression analysis across all cell types and subtypes to identify age-specific changes in gene expression (**table S16**). We identified 581 genes that increase in expression from younger to middle-age to older, and 18 genes that decrease with age when analyzed by cell types and subtypes. The top five most enriched genes in older (red) versus younger cells (blue) for each cell type are shown (**Fig 5A**). Genes enriched with age include *PDK4* and *CPT1A* in hepatocytes, which, encode mitochondrial proteins that regulate glucose and lipid metabolism and fatty acid beta-oxidation, respectively (*38*, *39*) and *NEK10*, a regulator of mitochondrial function (*40*). Sinusoidal endothelial cells, portal fibroblasts, cholangiocytes, and Kupffer cells all show differential expression of unique genes with changes in age (**Fig 5A** and **table S16**).

Pathways enriched with increasing age were also identified (**Fig 5B)**, but the smaller set of genes that decrease with age were not sufficient to identify pathways in this group. Portal hepatocytes showed enrichment for Endothelin Pathways (WP2197) and Fatty Acid Biosynthesis (WP357), while all hepatocytes were enriched for Mitochondrial Fatty Acid Oxidation Disorders (WP5123), Mitochondrial Long Chain Fatty Acid Beta Oxidation (WP368) and Fatty Acid Beta Oxidation (WP143). Sinusoidal endothelial cells were enriched in pathways including Hippo Merlin Signaling (WP4540, WP4541) and Focal Adhesion PI3K-Akt-mTOR Signaling (WP3932). VEGFA-VEGFR2 Signaling (WP3888) showed significant enrichment in

portal fibroblasts and Kupffer cells with age, and the NRF2 Pathway (WP2884) was enriched in multiple cell types. Collectively, this analysis underscores the cell-type-specific adaptations across key metabolic and signaling pathways with increasing age. (**table S17)**.

Senescence signatures with increasing age

Expression of senescence genes increases with age in many organs, but we did not identify a senescence signature in our pathway analysis (**Fig 5B**). We next evaluated expression of genes associated with senescence (*41*) to better understand how these genes change with increasing age across cell types and subtypes (**Fig 5C, fig S10**). We found that CXCL2 was enriched with age in inflammatory macrophages and portal fibroblasts (**Fig 5C**), consistent with induction of *CXLC2* in macrophages in the aging mouse liver (*42*)**.** Genes induced with age also included *STAT3* in both Kupffer cells and portal fibroblasts, as well as *IGFBP2* in all hepatocytes and inflammatory macrophages (**Fig 5B, fig S10**). While only nine senescence-associated genes were significantly enriched with age within individual cell types or subtypes, different combinations of senescence genes followed the trend of increasing expression from younger to middle to older age groups for each cell type and subtype (**table S18** and **Supplementary Material**). Using the set of genes that increase with age within each cell type and subtype, we then defined cell-type-specific senescence signatures (**Fig 5D, table S18**). While hepatocyte subsets contained the highest number of nuclei (portal 41,104, mid-zone 26,074 and central 25,584; **table S2 and S7**), it is notable that hepatocytes contained the smallest number of senescence genes that increase with age compared to all other cell types (portal 13, mid zone 15 and central 28). For example, 63 senescence genes exhibited an upward trend in non-macrophage immune cells (4112 cells), 49 in Kupffer cells (3530 cells), and 47 in inflammatory macrophages (5544 cells). Notably, these cell types make up less than one-fifteenth of the total number of hepatocytes in this dataset.

Impact of age on cell-cell interactions

We next evaluated signaling pathway strength from younger to older liver cells using CellChat. In younger cells, pathways involving ANNEXIN, VCAM, and IGF were enriched (**Fig 6A**, green labels), while older cells showed enrichment in pathways related to SEMA3, CDH1, and NRG (purple labels). Additional pathways (black labels) were active but did not significantly differ between younger and older cells (**table S19**).

We then visualized cell-cell communication across cell types and subtypes with increasing age. For example, the IGF signaling pathway, enriched in younger cells (**Fig 6B**), shows robust signaling from hepatocytes to sinusoidal and VV endothelial cells, with signaling to qHSCs and portal fibroblasts present only in younger cells. In contrast, older cells show enrichment in pathways including NRG, with stronger signals between hepatocytes and new signals developing from hepatocytes to cholangiocytes, as well as from sinusoidal and VV endothelial cells to cholangiocytes (**Fig 6C**). These findings highlight how signaling pathways vary in strength and involve different cell types as the liver ages.

To better understand the individual molecules responsible for these cell-cell interactions, we next visualized L-R pairs. We focused on portal hepatocytes and sinusoidal endothelial cells (**Fig 6D-G),** and included data for each cell type in **table S19**. L-R pairs, including the IGF1 (ligand) expressed by portal hepatocytes and IGF1R (receptor) expressed by sinusoidal endothelial cells, VV endothelial cells, qHSCs, and portal fibroblasts are enriched in younger donors (**Fig 6D**). In addition, the combination of VEGFA (ligand), expressed by portal hepatocytes, and FLT1 (receptor), expressed by sinusoidal and VV endothelial cells, or KDR (receptor), expressed by sinusoidal endothelial cells, are also enriched in younger donors.

With older age, enriched L-R pairs include the combination of NRG1 (ligand), expressed in portal hepatocytes, and ERBB3 (receptor), expressed in other hepatocyte populations, or ERBB4 (receptor), expressed by cholangiocytes (**Fig 6E**). Additionally, the combination of PLG (ligand), expressed by portal hepatocytes, and PARD3 (receptor) expressed by qHSCs, portal fibroblasts, cholangiocytes, and other hepatocytes is also enriched in older donors. These results demonstrate change in L-R pairs that occur with age, including enrichment of IGF1 signaling in younger livers and NRG1 in older livers.

We observed fewer L-R pairs signaling to portal hepatocytes enriched in younger or older donors (**fig S11A-B)**, as compared to L-R pairs signaling from hepatocytes. In younger donors, the interactions primarily involved the combination of HGF (ligand) expressed in qHSCs and MET (receptor) expressed by portal hepatocytes and the combination of collagens expressed by portal fibroblasts and multiple receptors expressed by portal hepatocytes (**fig S11A**). Older donors showed enrichment of L-R pairs involving BMP-6 (ligand) expressed by VV endothelial cells and ACVR1_BMPR2 and BMPR1A_BMPR2 receptor complexes expressed by portal hepatocytes (**fig S11B**).

Analysis of L-R pairs involving sinusoidal endothelial cells also displayed age-dependent shifts, with the greatest diversity observed in signals received, predominantly from hepatocytes, portal fibroblasts, qHSCs, and cholangiocytes (**Fig 6F, 6G, fig S11C-D**). In younger donors, the combination of VEGFA (ligand) expressed by hepatocytes and FLT1 and KDR (receptors) expressed by sinusoidal endothelial cells was enriched along with the combination of TENM2 (ligand) expressed by hepatocytes and ADGRL2 (receptor) expressed by sinusoidal endothelial cells. L-R pairs involving THBS1 (ligand) expressed by qHSCs and portal fibroblasts and CD36 (receptor) expressed by sinusoidal endothelial cells were also enriched (**Fig 6F**).

In older donors, the signaling landscape shifted (**Fig 6G**). The combination of ANGPTL4 (ligand), expressed by hepatocytes and CDH5 (receptor) expressed by sinusoidal endothelial cells was enriched. In addition, the combinations of PROS1 (ligand) and MERTK (receptor) along with TENM1 (ligand) and ADGRL2 (receptor) were enriched to signal from hepatocytes to sinusoidal endothelial cells. The combination of SEMA3C (ligand) expressed by portal fibroblasts and the receptors NRP1_PLXNA2, NRP2_PLXNA2, NRP1_NRP2, and PLXND1 expressed by sinusoidal endothelial cell was also enriched.

While there were fewer changes in L-R pairs with age that involved signaling from sinusoidal endothelial cells, we did observe an enrichment of EFNB2 (ligand) expressed by sinusoidal endothelial cells and EPHA4 (receptor) expressed by VV endothelial cells, along with ANGPT2 (ligand) expressed by sinusoidal endothelial cells and TEK and ITGA5_ITGB1 (receptors) expressed by VV endothelial cells (**fig S11C**). In older donors there was an enrichment of the combination of NRG3 (ligand) expressed by sinusoidal endothelial cells and ERBB4 (receptor) expressed by cholangiocytes and the combination of ADM (ligand) expressed by sinusoidal endothelial cells and CALCRL (receptor) expressed by VV endothelial cells and qHSCs (**fig S11D**). Together, these findings highlight L-R pairs that change with age involving portal hepatocytes and sinusoidal endothelial cells, while also cataloging the combination of L-R pairs enriched between all cell types throughout the liver.

DISCUSSION

The liver plays a key role in numerous physiological processes. While single-cell studies have highlighted the diversity of liver cell types and disease-related changes (*1–4*), how sex and age influence healthy liver cells remains poorly understood. Using snRNA-seq, we analyzed over 176,000 nuclei from 32 adult donors spanning seven decades to explore liver cell composition and heterogeneity across sex and age.

We analyzed gene expression, pathways, L-R interactions, and protein networks in female and male cells to uncover sex-specific differences. Estrogen signaling was prominent in female cells (*43*, *44*), alongside enrichment in cholesterol, lipid metabolism, and ferroptosis pathways across hepatocytes and non-parenchymal cells (**table S11**). Male cells showed enrichment of Hippo signaling in sinusoidal and VV endothelial cells and chemokine pathways in portal fibroblasts (*45–48*). These findings confirm known pathways and reveal additional mechanisms modulating cell activities in female and male liver cells.

Analysis of signaling pathways and L-R pairs in female and male liver cells highlights sex-specific differences in lipid metabolism and cholesterol regulation. ApoA signaling from hepatocytes to Kupffer cells expressing ABCA1, along with enriched cholesterol metabolism networks in female hepatocytes, highlights sex-specific differences in lipid metabolism (*49*) and MASLD risk factors (*50*). Enrichment of pathways involving ANGPTL further suggest differences in lipid metabolism (*51*, *52*) between female and male livers. These findings align with the role of estrogen in modulating lipid metabolism in female hepatocytes, enhancing cholesterol clearance via LDL receptor upregulation and promoting the uptake of LDL cholesterol from the bloodstream (13, 52). Additionally, estrogen stimulates the production of APOA-1/II, key components of HDL particles, facilitating cholesterol transport (*53*).

While not identified in the initial analysis of sex-specific differential gene regulation, BMP emerged as a male-enriched signaling pathway, driven by BMP6 expression in sinusoidal endothelial cells signaling to qHSCs, with weaker signals to hepatocytes, cholangiocytes, and portal fibroblasts (*54–56*). BMP6 is induced with MASLD and has activity against inflammation and fibrosis (*57*), in addition to its role in regulating hepcidin (HAMP) expression in hepatocytes (*58*). BMP family members also modulate DGAT2 and mTORC1 in hepatocytes to regulate

lipid/triglyceride turnover (*59*, *60*) and interact with Gremlin1 to control hepatocyte senescence during progression of NAFLD/NASH (*61*). Together, these findings highlight pathways by which enrichment of BMP6 signaling in males may influence susceptibility to NAFLD and related liver diseases.

Comparative analysis of protein-protein interaction networks in female and male liver cells reveals sex-specific molecular differences that affect cellular functions and responses. Female-enriched networks include estrogen signaling, pregnane X receptor, and metabolic pathways of LDL and HDL, while male-enriched networks focus on stress response and selenium pathways. These distinctions highlight the role of sex hormones and metabolic regulation in liver disease outcomes, offering insights for precision medicine and sex-specific interventions.

Machine learning analysis of gene expression in females and males, with and without X and Y chromosome genes, highlighted significant predictive power in both donor and cell type classifications. Even without sex chromosome genes, machine learning revealed distinct autosomal gene expression differences (**table S14**), emphasizing the importance of genes like *CUX2, A2M, ECHDC2, ALB, and LINC01554* in distinguishing sex. It is notable that our analysis identified increased expression of *CUX2* in males, while murine models detect increased *CUX2* in female livers (*36*). These murine studies showed CUX2 expression was responsive to hormone signals, and it is possible that differences in hormone signaling could explain different patterns between humans and mice, or the findings could be influenced by analysis of nuclear-enriched RNA compared to total cellular RNA. Overall, these findings enrich our understanding of sex classification and open avenues for further research into the intricate biological pathways that govern sex differences at the cellular level.

Age-related shifts in liver cell signaling revealed that younger livers show enriched IGF and VEGFA pathways linked to growth, vascular interactions, and cellular maintenance. IGF and its receptor IGF1R prevent apoptosis, promote cell proliferation, and induce vascular endothelial growth factor (VEGF) expression (*62*). Vascular endothelial growth factor A (VEGFA) regulates angiogenesis, inflammatory responses, and energy metabolism, including hepatic lipid and glycogen metabolism (*63*). In contrast, older livers exhibit increased NRG1 and PLG signaling, suggesting adaptations in repair and metabolism. Neuregulins (NRGs) regulate glucose and lipid homeostasis (*64*) and protect against NAFLD by modulating ERBB3 phosphorylation and PI3K-AKT signaling (*65*). These findings highlight age-related shifts from growth and vascular signals in younger livers to repair-focused signaling in older livers.

We analyzed sex and age separately due to limited sample size but identified intersecting trends when viewing the results together. For example, ANGPT2 and VEGFA signaling are enriched in younger age and females, while HGF and IGF1 are enriched in younger age and males. In contrast, ANGPTL4 and NRG1 are enriched in older age and females, whereas BMP6 and PLG are enriched in older age and males. These results identify pathways that may be further enriched within subgroups, where, for example, a pathway promoted with aging may be further augmented in one sex compared to another.

Limitations

While this study provides key insights into sex- and age-related differences in liver signaling pathways, limitations exist. The cohort of 32 adult donors, mostly Caucasian (30/32), may limit generalizability to broader populations. Samples from a single center reflect local geography, and although histologically assessed as healthy, tissues were derived from excess surgical samples, potentially not representing completely healthy livers. Despite these constraints, the study highlights the broad influence of sex and age on liver cell types and subtypes.

In conclusion, our snRNA-seq analysis reveals sex- and age-related changes in gene expression, signaling, and cell-cell communication in the healthy human liver at the cell-type and subtype level. These findings offer a valuable reference for studying the impact of sex and age on liver disease progression.

MATERIALS AND METHODS

Tissue collection

All liver samples were collected from excess surgical tissue at Massachusetts General Hospital in accordance with protocols approved by the Mass General Brigham (MGB) or Dana Farber Cancer Institute (DFCI) Institutional Review Boards (IRB).

Sequencing

TST (Tween with salts and Tris) buffer was prepared as previously described (*66*) with final concentrations of 146mM NaCl, 10mM Tris-HCl ph7.5, 1mM CaCl2, 21mM MgCl2, 0.03% Tween-20 and 0.01% BSA in ultrapure water. From dry ice, a piece of frozen tissue was placed into a gentleMACS C Tube on dry ice (Miltenyi Biotec, cat. no. 130-093-237) with 2ml of TST buffer including 1U/ml Protector RNase inhibitor (Sigma-Aldrich, cat. no. 3335402001). Tissue was immediately dissociated using a gentleMACS Dissociator (Miltenyi Biotec, cat. no. 130-096-427) with the m_Spleen_01_01 program twice and incubated on ice for 5min to complete a 10min incubation with the TST buffer. C tubes were spun at 4 °C for 2min at 500g. The pellet was resuspended in TST buffer, filtered through a 40µm Falcon cell strainer (VWR International, LLC, cat. no. 43-50040-51) into 50ml conical tube. The strainer was washed with 1ml 1TST buffer including 0.5U/ml Protector RNase inhibitor before use. Additional 1ml of 1TST buffer including 0.5U/ml Protector RNase inhibitor was used to wash the gentleMACS C Tubes and filter. 1ml of 1XST buffer including 0.5U/ml Protector RNase inhibitor was added to the filter for a final wash. The sample was transferred to a 15ml conical tube and centrifuged at 4 °C for 10min

at 500g. The pellet was resuspended in 1XPBS (-Mg/-Ca) 1% BSA, and 1U/ml Protector RNase inhibitor buffer (volume depends on pellet size, between 100-200 µl). The nucleus solution was filtered through a 35 µm Falcon cell strainer (Corning, cat. no. 352235). Nuclei were counted using a INCYTO C-chip disposable hemocytometer (VWR International, cat. no. 22-600-100). 8,000 - 12,000 nuclei of the single-nucleus suspension were loaded onto the Chromium Chips for the Chromium Single Cell 3′ Library (Chromium Next GEM Single Cell 3' Kit v3.1, cat. no. PN-1000268, PN-1000120, PN-1000215) according to the manufacturer's recommendations (10x Genomics). Gene expression libraries were constructed and indexed according to manufacturer's instructions and pooled for sequencing on either a NovaSeq 6000 or NovaSeqX sequencer (Illumina). All libraries were sequenced to a targeted depth of 400 million reads in the following configuration: R1: 28bp, R2: 90bp, I1, I2: 10bp.

Single-nucleus RNA-seq quality control and analysis

Samples were demultiplexed using Cell Ranger mkfastq or bclconvert and aligned to the GENCODE v42 reference genome using Cell Ranger v7.0.1. Ambient RNA contamination was removed with SoupX version 1.6.2 (**Supplementary Material**), and doublets or multiplets were filtered out using DoubletFinder v2.0.4 (**Supplementary Material**). High-quality cells were integrated with Harmony v1.2.0 and clustered at different resolutions using Seurat (v4.3.0), followed by UMAP visualization (**Supplementary Material**). Clusters identified were assigned cell type labels using the `sctype_score` function from the ScType R package (**Supplementary Material**), utilizing the 'ScTypeDB_full.xlsx' database downloaded on 26/06/2023 (**Supplementary Material**). Marker genes were selected through differential expression analysis for each cell type. Cell-type proportions were analyzed with the speckle R package v0.99.7 using metadata for sex and age group (**Supplementary Material**).

Differential expression, gene ontology enrichment and cell-cell communication analysis

Differential expression analysis was performed for sex (male vs. female) and age groups (younger: 18-40, middle: 41-60, older: 61-90 years) using the MAST test implemented in the 'FindMarkers' function of Seurat version 4.3.0, incorporating donor as a latent variable to account for batch effects. Age-related comparisons focused on pairwise analyses between younger, middle-aged, and older groups within each cell type. Genes consistently increasing or decreasing across age groups were identified as age-associated markers (**Supplementary Material) (table S10, table S11**). Gene ontology enrichment for biological processes was conducted using EnrichR version 3.2 with the WikiPathway_2023_Human library (**Supplementary Material)** (**table S16** and **table S17**). CellChat v2.1.2 was used to analyze communication patterns across sex (males vs. females) and age groups (younger: 18-40 years, middle: 41-60 years, older: 61-90 years). CellChat was applied independently to each group and then merged for comparison using the mergeCellChat function (male vs. female for sex, and younger, middle, and older groups for age) (**Supplementary Material)**.

STRING-DB network analysis

Protein-protein interactions (PPIs) were analyzed using the R package STRINGdb v2.10.1. The STRING database (species = 9606, version = 11.5) was accessed with the network type set to physical (**Supplementary Material)**. Gene ontology enrichment analysis was conducted on the STRING online server (https://version-11-5.string-db.org/ accessed in November 2024) to identify WikiPathways associated with these genes (**Supplementary Material)**.

List of Supplementary Materials

Materials and Methods

Fig S1 to to S11 supplementary figures

Data files S1 to S19 (Excel files)

References

1. S. A. MacParland, J. C. Liu, X.-Z. Ma, B. T. Innes, A. M. Bartczak, B. K. Gage, J. Manuel, N. Khuu, J. Echeverri, I. Linares, R. Gupta, M. L. Cheng, L. Y. Liu, D. Camat, S. W. Chung, R. K. Seliga, Z. Shao, E. Lee, S. Ogawa, M. Ogawa, M. D. Wilson, J. E. Fish, M. Selzner, A. Ghanekar, D. Grant, P. Greig, G. Sapisochin, N. Selzner, N. Winegarden, O. Adeyi, G. Keller, G. D. Bader, I. D. McGilvray, Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 1–21 (2018).

2. P. Ramachandran, R. Dobie, J. R. Wilson-Kanamori, E. F. Dora, B. E. P. Henderson, N. T. Luu, J. R. Portman, K. P. Matchett, M. Brice, J. A. Marwick, R. S. Taylor, M. Efremova, R. Vento-Tormo, N. O. Carragher, T. J. Kendall, J. A. Fallowfield, E. M. Harrison, D. J. Mole, S. J. Wigmore, P. N. Newsome, C. J. Weston, J. P. Iredale, F. Tacke, J. W. Pollard, C. P. Ponting, J. C. Marioni, S. A. Teichmann, N. C. Henderson, Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512–518 (2019).

3. N. Aizarani, A. Saviano, Sagar, L. Mailly, S. Durand, J. S. Herman, P. Pessaux, T. F. Baumert, D. Grün, A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199–204 (2019).

4. A. Sharma, J. J. W. Seow, C. A. Dutertre, R. Pai, C. Blériot, A. Mishra, R. M. M. Wong, G. S. N. Singh, S. Sudhagar, S. Khalilnezhad, S. Erdal, H. M. Teo, A. Khalilnezhad, S. Chakarov, T. K. H. Lim, A. C. Y. Fui, A. K. W. Chieh, C. P. Chung, G. K. Bonney, B. K. Goh, J. K. Y. Chan, P. K. H. Chow, F. Ginhoux, R. DasGupta, Onco-fetal Reprogramming of Endothelial Cells Drives Immunosuppressive Macrophages in Hepatocellular Carcinoma. *Cell* **183** (2020).

5. T. S. Andrews, J. Atif, J. C. Liu, C. T. Perciani, X. Ma, C. Thoeni, M. Slyper, G. Eraslan, A. Segerstolpe, J. Manuel, S. Chung, E. Winter, I. Cirlan, N. Khuu, S. Fischer, O. Rozenblatt□Rosen, A. Regev, I. D. McGilvray, G. D. Bader, S. A. MacParland, Single□Cell, Single□Nucleus, and Spatial RNA Sequencing of the Human Liver Identifies Cholangiocyte and Mesenchymal Heterogeneity. *Hepatology Communications* **6**, 821 (2022).

6. R. Marcos, C. Lopes, F. Malhão, C. Correia-Gomes, S. Fonseca, M. Lima, R. Gebhardt, E. Rocha, Stereological assessment of sexual dimorphism in the rat liver reveals differences in hepatocytes and Kupffer cells but not hepatic stellate cells. *J. Anat.* **228**, 996–1005 (2016).

7. G. Rando, W. Wahli, Sex differences in nuclear receptor-regulated liver metabolic pathways. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1812**, 964–973 (2011).

8. S. D. Tadic, M. S. Elm, H.-S. Li, G. J. Van Londen, V. M. Subbotin, D. C. Whitcomb, P. K. Eagon, Sex differences in hepatic gene expression in a rat model of ethanol-induced liver injury. *Journal of Applied Physiology* **93**, 1057–1068 (2002).

9. M. Pinzani, R. G. Romanelli, S. Magli, Progression of fibrosis in chronic liver diseases: time to tally the score. *J. Hepatol.* **34**, 764–767 (2001).

10. J. D. Browning, L. S. Szczepaniak, R. Dobbins, P. Nuremberg, J. D. Horton, J. C. Cohen, S. M. Grundy, H. H. Hobbs, Prevalence of hepatic steatosis in an urban population in the

United States: impact of ethnicity. *Hepatology* **40**, 1387–1395 (2004).

11. W. Dai, L. Ye, A. Liu, S. W. Wen, J. Deng, X. Wu, Z. Lai, Prevalence of nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus: A meta-analysis. *Medicine (Baltimore)* **96**, e8179 (2017).

12. C. Burelle, V. Clapatiuc, S. Deschênes, A. Cuillerier, M. De Loof, M.-È. Higgins, H. Boël, C. Daneault, B. Chouinard, M.-É. Clavet, N. Tessier, I. Croteau, G. Chabot, C. Martel, M. G. Sirois, S. Lesage, Y. Burelle, M. Ruiz, A genetic mouse model of lean-NAFLD unveils sexual dimorphism in the liver-heart axis. *Commun. Biol.* **7**, 356 (2024).

13. M. Schiffrin, C. Winkler, L. Quignodon, A. Naldi, M. Trötzmüller, H. Köfeler, H. Henry, P. Parini, B. Desvergne, F. Gilardi, Sex dimorphism of nonalcoholic fatty liver disease (NAFLD) in pparg-null mice. *Int. J. Mol. Sci.* **22**, 9969 (2021).

14. A. Smiriglia, N. Lorito, M. Serra, A. Perra, A. Morandi, M. A. Kowalik, Sex difference in liver diseases: How preclinical models help to dissect the sex-related mechanisms sustaining NAFLD and hepatocellular carcinoma. *iScience* **26**, 108363 (2023).

15. O. Krenkel, J. Hundertmark, T. P. Ritz, R. Weiskirchen, F. Tacke, Single cell RNA sequencing identifies subsets of hepatic stellate cells and myofibroblasts in liver fibrosis. *Cells* **8**, 503 (2019).

16. B. A. J. Tamburini, J. M. Finlon, A. E. Gillen, M. S. Kriss, K. A. Riemondy, R. Fu, R. P. Schuyler, J. R. Hesselberth, H. R. Rosen, M. A. Burchill, Chronic liver disease in humans causes expansion and differentiation of liver lymphatic endothelial cells. *Front. Immunol.* **10**, 1036 (2019).

17. T. Zhou, M. Kiran, K. O. Lui, Q. Ding, Decoding liver fibrogenesis with single-cell technologies. *Life Medicine* **1**, 333–344 (2022).

18. X. Xiong, H. Kuang, S. Ansari, T. Liu, J. Gong, S. Wang, X.-Y. Zhao, Y. Ji, C. Li, L. Guo, L. Zhou, Z. Chen, P. Leon-Mimila, M. T. Chung, K. Kurabayashi, J. Opp, F. Campos-Pérez, H. Villamil-Ramírez, S. Canizales-Quinteros, R. Lyons, C. N. Lumeng, B. Zhou, L. Qi, A. Huertas-Vazquez, A. J. Lusis, X. Z. S. Xu, S. Li, Y. Yu, J. Z. Li, J. D. Lin, Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. *Mol. Cell* **75**, 644–660.e5 (2019).

19. R. P. Thomas, M. Guigneaux, T. Wood, B. M. Evers, Age-associated changes in gene expression patterns in the liver. *J. Gastrointest. Surg.* **6**, 445–53; discussion 454 (2002).

20. M. Porukala, P. K. Vinod, Network-level analysis of ageing and its relationship with diseases and tissue regeneration in the mouse liver. *Sci. Rep.* **13**, 4632 (2023).

21. S. N. Hilmer, V. C. Cogger, R. Fraser, A. J. McLean, D. Sullivan, D. G. Le Couteur, Age-related changes in the hepatic sinusoidal endothelium impede lipoprotein transfer in the rat. *Hepatology* **42**, 1349–1354 (2005).

22. A. J. McLean, V. C. Cogger, G. C. Chong, A. Warren, A. M. A. Markus, J. E. Dahlstrom, D. G. Le Couteur, Age-related pseudocapillarization of the human liver. *J. Pathol.* **200**, 112–117 (2003).

23. R. Maeso-Díaz, M. Ortega-Ribera, A. Fernández-Iglesias, D. Hide, L. Muñoz, A. J. Hessheimer, S. Vila, R. Francés, C. Fondevila, A. Albillos, C. Peralta, J. Bosch, F. Tacke, V. C. Cogger, J. Gracia-Sancho, Effects of aging on liver microcirculatory function and sinusoidal phenotype. *Aging Cell* **17**, e12829 (2018).

24. A. Warren, V. C. Cogger, R. Fraser, L. D. Deleve, R. S. McCuskey, D. G. Le Couteur, The effects of old age on hepatic stellate cells. *Curr. Gerontol. Geriatr. Res.* **2011**, 439835 (2011).

25. M. Noureddin, K. P. Yates, I. A. Vaughn, B. A. Neuschwander-Tetri, A. J. Sanyal, A. McCullough, R. Merriman, B. Hameed, E. Doo, D. E. Kleiner, C. Behling, R. Loomba, NASH CRN, Clinical and histological determinants of nonalcoholic steatohepatitis and advanced fibrosis in elderly patients: HEPATOLOGY, Vol. 00, No. X, 2013 NOUREDDIN ET AL. *Hepatology* **58**, 1644–1654 (2013).

26. M. Ogrodnik, S. Miwa, T. Tchkonia, D. Tiniakos, C. L. Wilson, A. Lahat, C. P. Day, A. Burt, A. Palmer, Q. M. Anstee, S. N. Grellscheid, J. H. J. Hoeijmakers, S. Barnhoorn, D. A. Mann, T. G. Bird, W. P. Vermeij, J. L. Kirkland, J. F. Passos, T. von Zglinicki, D. Jurk, Cellular senescence drives age-dependent hepatic steatosis. *Nat. Commun.* **8**, 15691 (2017).

27. L. Grosse, N. Wagner, A. Emelyanov, C. Molina, S. Lacas-Gervais, K.-D. Wagner, D. V. Bulavin, Defined p16High senescent cell types are indispensable for mouse healthspan. *Cell Metab.* **32**, 87–99.e6 (2020).

28. Z. A. Clarke, G. D. Bader, *MALAT1*expression indicates cell quality in single-cell RNA sequencing data, *bioRxiv* (2024). https://doi.org/10.1101/2024.07.14.603469.

29. T. Montserrat-Ayuso, A. Esteve-Codina, Revealing the prevalence of suboptimal cells and organs in reference cell atlases: An imperative for enhanced quality control, *bioRxiv* (2024). https://doi.org/10.1101/2024.04.18.590104.

30. A. Ianevski, A. K. Giri, T. Aittokallio, Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat. Commun.* **13**, 1246 (2022).

31. K. B. Halpern, R. Shenhav, O. Matcovitch-Natan, B. Toth, D. Lemze, M. Golan, E. E. Massasa, S. Baydatch, S. Landen, A. E. Moor, A. Brandis, A. Giladi, A. S. Avihail, E. David, I. Amit, S. Itzkovitz, Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).

32. J. M. Gómez-Salinero, F. Izzo, Y. Lin, S. Houghton, T. Itkin, F. Geng, Y. Bram, R. P. Adelson, T. M. Lu, G. Inghirami, J. Z. Xiang, R. Lis, D. Redmond, R. Schreiner, S. Y. Rabbany, D. A. Landau, R. E. Schwartz, S. Rafii, Specification of fetal liver endothelial progenitors to functional zonated adult sinusoids requires c-Maf induction. *Cell Stem Cell* **29**, 593–609.e7 (2022).

33. W. Yang, H. He, T. Wang, N. Su, F. Zhang, K. Jiang, J. Zhu, C. Zhang, K. Niu, L. Wang, X. Yuan, N. Liu, L. Li, W. Wei, J. Hu, Single-cell transcriptomic analysis reveals a hepatic stellate cell-activation roadmap and myofibroblast origin during liver fibrosis in mice. *Hepatology* **74**, 2774–2790 (2021).

34. B. Phipson, C. B. Sim, E. R. Porrello, A. W. Hewitt, J. Powell, A. Oshlack, Propeller: Testing for differences in cell type proportions in single cell data. *Bioinformatics* **38**, 4720–4726 (2022).

35. G. Pacini, I. Dunkel, N. Mages, V. Mutzel, B. Timmermann, A. Marsico, E. G. Schulz, Integrated analysis of Xist upregulation and X-chromosome inactivation with single-cell and single-allele resolution. *Nat. Commun.* **12**, 3638 (2021).

36. E. V. Laz, M. G. Holloway, C.-S. Chen, D. J. Waxman, Characterization of three growth hormone-responsive transcription factors preferentially expressed in adult female liver. *Endocrinology* **148**, 3327–3337 (2007).

37. V. R. Varma, S. Varma, Y. An, T. J. Hohman, S. Seddighi, R. Casanova, A. Beri, E. B. Dammer, N. T. Seyfried, O. Pletnikova, A. Moghekar, M. R. Wilson, J. J. Lah, R. J. O'Brien, A. I. Levey, J. C. Troncoso, M. S. Albert, M. Thambisetty, Alpha-2 macroglobulin in Alzheimer's disease: a marker of neuronal injury through the RCAN1 pathway. *Mol. Psychiatry* **22**, 13–23 (2017).

38. Y. Zhao, M. Tran, L. Wang, D.-J. Shin, J. Wu, PDK4-deficiency reprograms intrahepatic glucose and lipid metabolism to facilitate liver regeneration in mice. *Hepatol. Commun.* **4**, 504–517 (2020).

39. S. M. Houten, S. Violante, F. V. Ventura, R. J. A. Wanders, The biochemistry and physiology of mitochondrial fatty acid β-oxidation and its genetic disorders. *Annu. Rev. Physiol.* **78**, 23–44 (2016).

40. A. P. de Oliveira, C. D. C. Navarro, P. R. F. Dias, T. Arguello, B. R. Walker, S. R. Bacman, L. M. Sousa, R. F. Castilho, S. R. Consonni, C. T. Moraes, J. Kobarg, NEK10 kinase ablation affects mitochondrial morphology, function and protein phosphorylation status. *Proteome Sci.* **22**, 8 (2024).

41. D. Saul, R. L. Kosinsky, E. J. Atkinson, M. L. Doolittle, X. Zhang, N. K. LeBrasseur, R. J. Pignolo, P. D. Robbins, L. J. Niedernhofer, Y. Ikeno, D. Jurk, J. F. Passos, L. J. Hickson, A. Xue, D. G. Monroe, T. Tchkonia, J. L. Kirkland, J. N. Farr, S. Khosla, A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nat. Commun.* **13**, 4827 (2022).

42. Y. Liu, J. Xiao, J. Cai, R. Li, X. Sui, J. Zhang, T. Lu, H. Chen, G. Chen, H. Li, C. Jiang, X. Zhao, C. Xiao, Y. Lei, J. Yao, G. Lv, J. Liang, Y. Zhang, J.-R. Yang, J. Zheng, Y. Yang, Single-cell immune profiling of mouse liver aging reveals Cxcl2+ macrophages recruit neutrophils to aggravate liver injury. *Hepatology* **79**, 589–605 (2024).

43. B. T. Palmisano, L. Zhu, J. M. Stafford, Role of estrogens in the regulation of liver lipid metabolism. *Adv. Exp. Med. Biol.* **1043**, 227–256 (2017).

44. T. Li, G. Wang, H. Zhao, F. Liu, D. Chen, X. Zhou, Z. Deng, Y. Cao, W. Fu, H. Zhang, J. Yang, Estrogen Receptor 1 Signaling in Hepatic Stellate Cells Designates Resistance to Liver Fibrosis, *Pathology* (2024). https://www.biorxiv.org/content/10.1101/2024.10.10.617541v1.full.

45. C. Gan, U. Yaqoob, J. Lu, M. Xie, A. Anwar, N. Jalan-Sakrikar, S. Jerez, T. S. Sehrawat, A. Navarro-Corcuera, E. Kostallari, N. W. Habash, S. Cao, V. H. Shah, Liver sinusoidal

endothelial cells contribute to portal hypertension through collagen type IV-driven sinusoidal remodeling. *JCI Insight* **9** (2024).

46. J. O. Russell, F. D. Camargo, Hippo signalling in the liver: role in development, regeneration and disease. *Nat. Rev. Gastroenterol. Hepatol.* **19**, 297–312 (2022).

47. G. Kolios, V. Valatas, E. Kouroumalis, Role of Kupffer cells in the pathogenesis of liver disease. *World J. Gastroenterol.* **12**, 7413–7420 (2006).

48. F. Marra, F. Tacke, Roles for chemokines in liver disease. *Gastroenterology* **147**, 577–594.e1 (2014).

49. D. S. Freedman, J. D. Otvos, E. J. Jeyarajah, I. Shalaurova, L. A. Cupples, H. Parise, R. B. D'Agostino, P. W. F. Wilson, E. J. Schaefer, Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham Study. *Clin. Chem.* **50**, 1189–1200 (2004).

50. B. T. Palmisano, L. Zhu, R. H. Eckel, J. M. Stafford, Sex differences in lipid and lipoprotein metabolism. *Mol. Metab.* **15**, 45–55 (2018).

51. A. K. Singh, B. Chaube, X. Zhang, J. Sun, K. M. Citrin, A. Canfrán-Duque, B. Aryal, N. Rotllan, L. Varela, R. G. Lee, T. L. Horvath, N. Price, Y. Suárez, C. Fernandez-Hernando, Hepatocyte-specific suppression of ANGPTL4 improves obesity-associated diabetes and mitigates atherosclerosis in mice. *J. Clin. Invest.* **131** (2021).

52. M. Deng, S. Kersten, Characterization of sexual dimorphism in ANGPTL4 levels and function. *J. Lipid Res.* **65**, 100526 (2024).

53. H. Szafran, W. Smielak-Korombel, The role of estrogens in hormonal regulation of lipid metabolism in women. *Przegl. Lek.* **55**, 266–270 (1998).

54. E. Charlebois, C. Fillebeen, J. Presley, G. Cagnone, V. Lisi, V.-P. Lavallée, J.-S. Joyal, K. Pantopoulos, Liver sinusoidal endothelial cells induce BMP6 expression in response to non-transferrin-bound iron. *Blood* **141**, 271–284 (2023).

55. S. Colucci, S. Altamura, O. Marques, K. Müdder, A. R. Agarvas, M. W. Hentze, M. U. Muckenthaler, Iron-dependent BMP6 regulation in liver sinusoidal endothelial cells is instructed by hepatocyte-derived secretory signals. *HemaSphere* **6**, e773 (2022).

56. S. Wang, C. Chen, L. Yu, J. Mueller, V. Rausch, S. Mueller, Bone morphogenetic protein 6-mediated crosstalk between endothelial cells and hepatocytes recapitulates the iron-sensing pathway in vitro. *J. Biol. Chem.* **297**, 101378 (2021).

57. S. Arndt, E. Wacker, C. Dorn, A. Koch, M. Saugspier, W. E. Thasler, A. Hartmann, A. K. Bosserhoff, C. Hellerbrand, Enhanced expression of BMP6 inhibits hepatic fibrosis in non-alcoholic fatty liver disease. *Gut* **64**, 973–981 (2015).

58. B. Andriopoulos Jr, E. Corradini, Y. Xia, S. A. Faasse, S. Chen, L. Grgurevic, M. D. Knutson, A. Pietrangelo, S. Vukicevic, H. Y. Lin, J. L. Babitt, BMP6 is a key endogenous regulator of hepcidin expression and iron metabolism. *Nat. Genet.* **41**, 482–487 (2009).

59. T. E. Thayer, C. L. Lino Cardenas, T. Martyn, C. J. Nicholson, L. Traeger, F. Wunderer, C. Slocum, H. Sigurslid, H. R. Shakartzi, C. O'Rourke, G. Shelton, M. D. Buswell, H. Barnes,

L. R. Neitzel, C. D. Ledsky, J. P. Li, M. F. Burke, E. Farber-Eger, D. S. Perrien, R. Kumar, K. E. Corey, Q. S. Wells, K. D. Bloch, C. C. Hong, D. B. Bloch, R. Malhotra, The role of bone morphogenetic protein signaling in non-alcoholic fatty liver disease. *Sci. Rep.* **10**, 9831 (2020).

60. Q. Peng, B. Chen, H. Wang, Y. Zhu, J. Wu, Y. Luo, G. Zuo, J. Luo, L. Zhou, Q. Shi, Y. Weng, A. Huang, T.-C. He, J. Fan, Bone morphogenetic protein 4 (BMP4) alleviates hepatic steatosis by increasing hepatic lipid turnover and inhibiting the mTORC1 signaling axis in hepatocytes. *Aging (Albany NY)* **11**, 11520–11540 (2019).

61. R. K. Baboota, A. Rawshani, L. Bonnet, X. Li, H. Yang, A. Mardinoglu, T. Tchkonia, J. L. Kirkland, A. Hoffmann, A. Dietrich, J. Boucher, M. Blüher, U. Smith, BMP4 and Gremlin 1 regulate hepatic cell senescence during clinical progression of NAFLD/NASH. *Nat. Metab.* **4**, 1007–1021 (2022).

62. N. Reinmuth, F. Fan, W. Liu, A. A. Parikh, O. Stoeltzing, Y. D. Jung, C. D. Bucana, R. Radinsky, G. E. Gallick, L. M. Ellis, Impact of insulin-like growth factor receptor-I function on angiogenesis, growth, and metastasis of colon cancer. *Lab. Invest.* **82**, 1377–1389 (2002).

63. Y. Wang, J. Ni, A. Wang, R. Zhang, L. Wang, Vascular endothelial growth factor A (VEGFA) regulates hepatic lipid and glycogen metabolism in Schizothorax prenanti. *Int. J. Mol. Sci.* **24**, 15171 (2023).

64. P. Zhang, H. Kuang, Y. He, S. O. Idiga, S. Li, Z. Chen, Z. Yang, X. Cai, K. Zhang, M. J. Potthoff, Y. Xu, J. D. Lin, NRG1-Fc improves metabolic health via dual hepatic and central action. *JCI Insight* **3** (2018).

65. D. Meng, H. Pan, Y. Chen, J. Ding, Y. Dai, Roles and mechanisms of NRG1 in modulating the pathogenesis of NAFLD through ErbB3 signaling in hepatocytes (NRG1 modulates NAFLD through ErbB3 signaling). *Obes. Res. Clin. Pract.* **15**, 145–151 (2021).

66. M. Slyper, C. B. M. Porter, O. Ashenberg, J. Waldman, E. Drokhlyansky, I. Wakiro, C. Smillie, G. Smith-Rosario, J. Wu, D. Dionne, S. Vigneau, J. Jané-Valbuena, T. L. Tickle, S. Napolitano, M.-J. Su, A. G. Patel, A. Karlstrom, S. Gritsch, M. Nomura, A. Waghray, S. H. Gohil, A. M. Tsankov, L. Jerby-Arnon, O. Cohen, J. Klughammer, Y. Rosen, J. Gould, L. Nguyen, M. Hofree, P. J. Tramontozzi, B. Li, C. J. Wu, B. Izar, R. Haq, F. S. Hodi, C. H. Yoon, A. N. Hata, S. J. Baker, M. L. Suvà, R. Bueno, E. H. Stover, M. R. Clay, M. A. Dyer, N. B. Collins, U. A. Matulonis, N. Wagle, B. E. Johnson, A. Rotem, O. Rozenblatt-Rosen, A. Regev, A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792–802 (2020).

Acknowledgements

The authors thank Ramnik Xavier and Åsa Segerstolpe for helpful discussions, Jasneet Aneja, Helena Lau, Georg Lauer, Carlos Fernandez-del Castillo, and Genevieve Boland for assistance in collection of liver tissue and Jonathan Eddy, Oiza Dimowo, James Cheney, and Mantero

Valentino Moreno-Cheek for assistance with tissue preservation. We thank Adam Slamin, Salome Maldonado, Dan Dubinsky, and the Broad Genomics Platform for help with generation of single-nucleus RNA sequencing data. A.C.M was supported by grants from the Chan Zuckerberg Initiative.

Funding

Chan Zuckerberg Initiative.

Author contribution

Conceptualization: R.R, J.D, A.C.M

Funding acquisition: A.C.M.

Sample collection and preparation: A.C.M, E.T.E, H.K, L.F, A.S, K.T, M.Q, C.F, D.B,

snRNA-sequencing: S.M, L.A.-Z, C,M., T.D., J.D.

Computational analyses: R.R, A.C.M.

Writing – original draft: R.R, A.C.M.

Writing – review & editing: R.R, J.D, A.C.M.

Competing Interest/conflict of interest

A.C.M. has received research funding from Boehringer Ingelheim and GlaxoSmithKline for unrelated projects. R.R. is a co-founder of deepnostiX, based in Germany and Pakistan, and founder of VitalEdge in the USA. Additionally, R. R. serves as a consultant for Ibis Therapeutics. No other authors have conflicts to declare.

Data and materials availability

Sequencing data used for this study are included in the Gene Expression Omnibus (GEO) under accession code: GSE210077

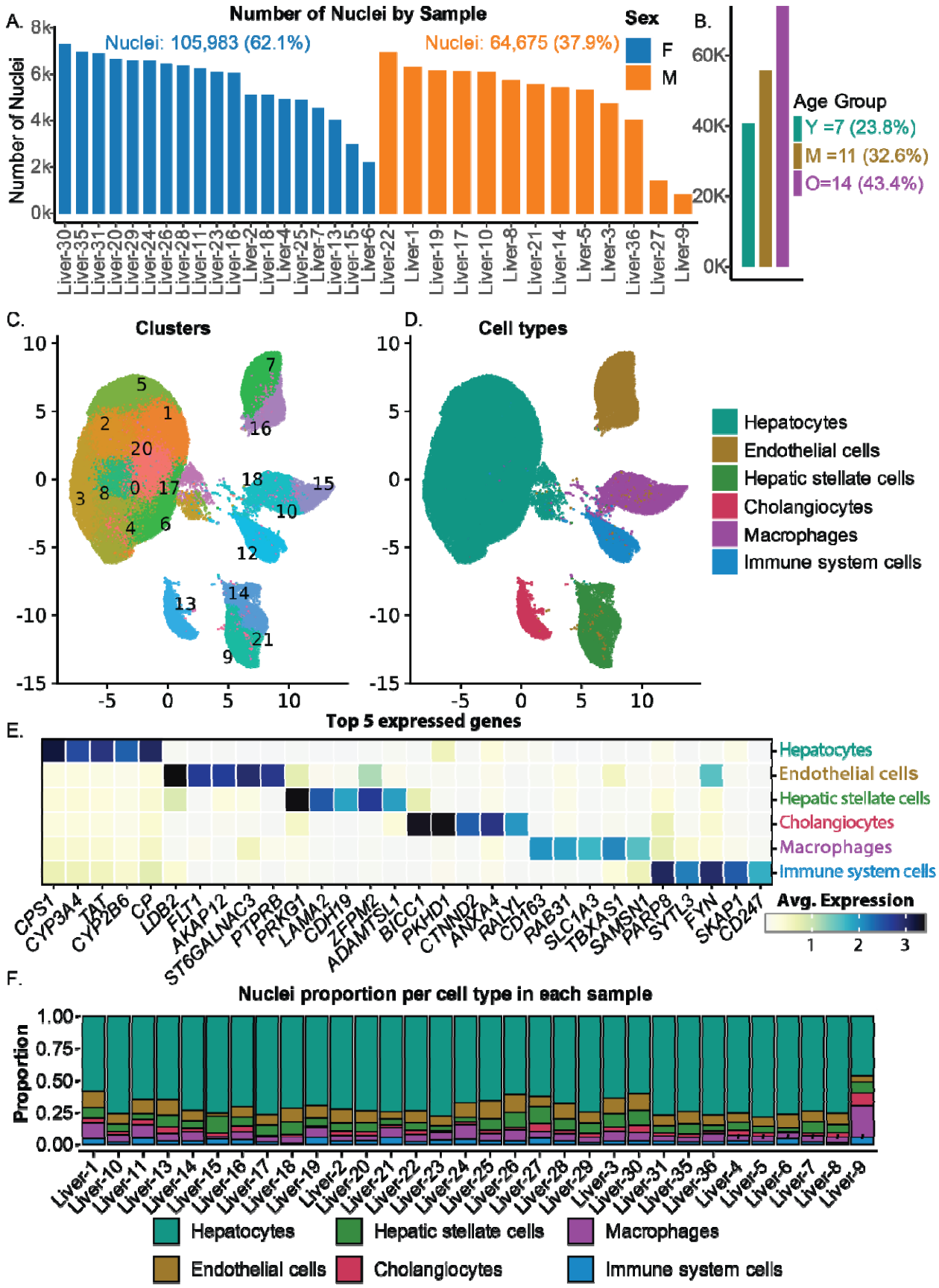

A.
Number of Nuclei by Sample
Nuclei: 105,983 (62.1%)
Nuclei: 64,675 (37.9%)
Sex
F
M
Number of Nuclei
8k
6k
4k
2k
0k
Liver-30
Liver-35
Liver-31
Liver-20
Liver-29
Liver-24
Liver-26
Liver-28
Liver-11
Liver-23
Liver-16
Liver-2
Liver-18
Liver-4
Liver-25
Liver-7
Liver-13
Liver-15
Liver-6
Liver-22
Liver-1
Liver-19
Liver-17
Liver-10
Liver-8
Liver-21
Liver-14
Liver-5
Liver-3
Liver-36
Liver-27
Liver-9
B.
60K
40K
20K
0K
Age Group
Y =7 (23.8%)
M =11 (32.6%)
O=14 (43.4%)
C.
Clusters
D.
Cell types
Hepatocytes
Endothelial cells
Hepatic stellate cells
Cholangiocytes
Macrophages
Immune system cells
E.
Top 5 expressed genes
CPS1
CYP3A4
TAT
CYP2B6
CP
LDB2
FLT1
AKAP12
ST6GALNAC3
PTPRB
PRKG1
LAMA2
CDH19
ZFPM2
ADAMTSL1
BICC1
PKHD1
CTNND2
ANXA4
RALYL
CD163
RAB31
SLC1A3
TBXAS1
SAMSN1
PARP8
SYTL3
FYN
SKAP1
CD247
Avg. Expression
1
2
3
F.
Nuclei proportion per cell type in each sample
Proportion
1.00
0.75
0.50
0.25
0.00
Liver-1
Liver-10
Liver-11
Liver-13
Liver-14
Liver-15
Liver-16
Liver-17
Liver-18
Liver-19
Liver-2
Liver-20
Liver-21
Liver-22
Liver-23
Liver-24
Liver-25
Liver-26
Liver-27
Liver-28
Liver-29
Liver-3
Liver-30
Liver-31
Liver-35
Liver-36
Liver-4
Liver-5
Liver-6
Liver-7
Liver-8
Liver-9
Hepatocytes
Hepatic stellate cells
Macrophages
Endothelial cells
Cholangiocytes
Immune system cells

Fig. 1. Sample composition, clustering, and cell-type proportions in the human liver. (A) Bars indicate nuclei counts within individual samples. Female (F) are in blue and male (M) are in orange. **(B)** Bars indicate nuclei counts by age group. Age categories are grouped as Younger (18–30 years, teal), Middle (40–59 years, yellow), and Older (≥ 60 years, purple). **(C)** Clustering of nuclei at resolution 0.7 visualized using UMAP. Each point represents an individual nucleus, colored according to its assigned cluster. Cluster identities are labeled numerically to correspond with specific populations identified in the analysis. **(D)** Cell-type assignments for each cluster (from C), identified using ScType, are displayed on the UMAP plot. **(E)**Top 5 marker genes defining each cell type identified in panel D are shown. **(F)** This panel illustrates the proportions of each cell type across all samples.

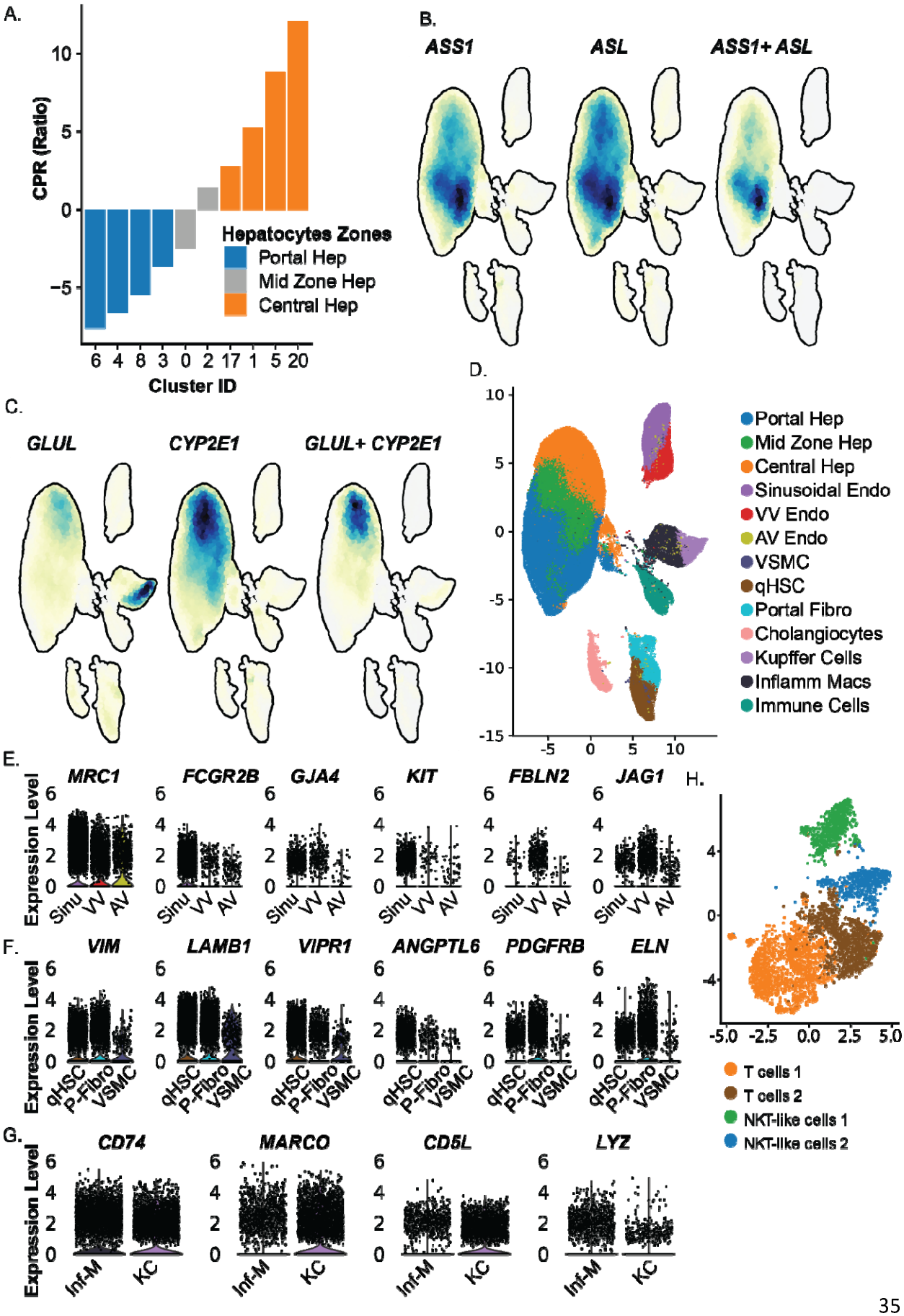

A.
CPR (Ratio)
Cluster ID
6 4 8 3 0 2 17 1 5 20
Hepatocytes Zones
Portal Hep
Mid Zone Hep
Central Hep
B.
ASS1
ASL
ASS1+ ASL
C.
GLUL
CYP2E1
GLUL+ CYP2E1
D.
Portal Hep
Mid Zone Hep
Central Hep
Sinusoidal Endo
VV Endo
AV Endo
VSMC
qHSC
Portal Fibro
Cholangiocytes
Kupffer Cells
Inflamm Macs
Immune Cells
E.
MRC1
FCGR2B
GJA4
KIT
FBLN2
JAG1
Expression Level
Sinu
VV
AV
F.
VIM
LAMB1
VIPR1
ANGPTL6
PDGFRB
ELN
qHSC
P-Fibro
VSMC
G.
CD74
MARCO
CD5L
LYZ
Inf-M
KC
H.
T cells 1
T cells 2
NKT-like cells 1
NKT-like cells 2

Fig. 2. Defining hepatocyte zones and subtypes in non-hepatocytes. (A) Central-to-Portal Ratio (CPR) for hepatocyte clusters, calculated using zonal markers ASS1, ASL (portal zone), and GLUL, CYP2E1 (central zone). Clusters were classified as portal (CPR values ≤ lower quantile: clusters 3, 4, 6, 8), central (CPR values ≥ upper quantile: clusters 1, 5, 17, 20), or mid-zone (intermediate CPR values: clusters 0, 2). **(B)** Density plots showing expression of GLUL and CYP2E1 overlaid on the UMAP, along with a combined (joint) density plot reflecting their co-expression distribution. **(C)** Density plots showing expression of ASS1 and ASL overlaid on the UMAP, alongside a combined (joint) density plot for their co-expression. **(D)** UMAP visualization of cell types and subtypes, detailing hepatocytes, non-parenchymal cells, and immune populations. **(E-G)** Marker genes identifying subtypes of indicated populations as shown in panel D. **(H)** UMAP visualization of non-macrophage immune subtypes.

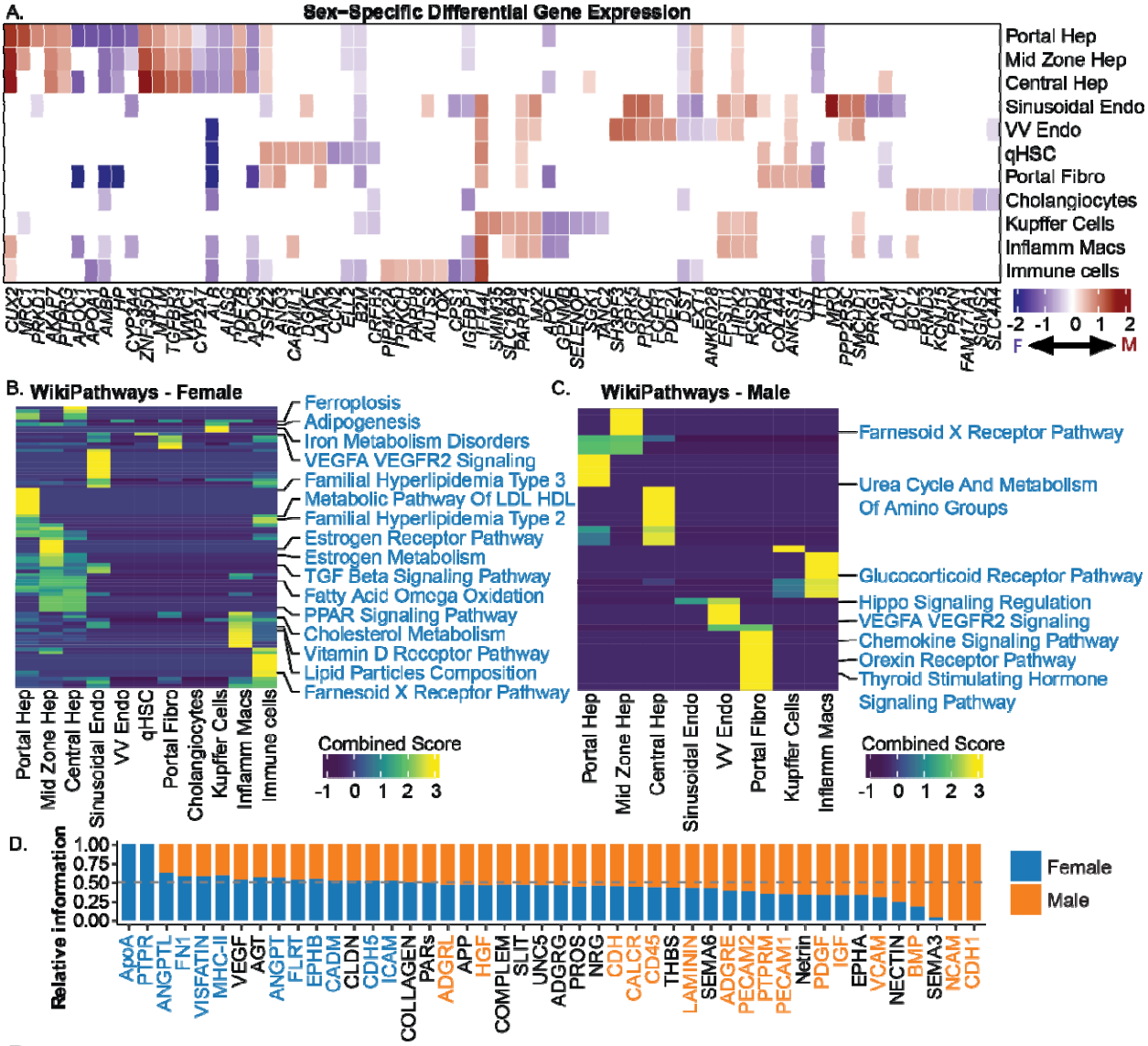

A.
Sex-Specific Differential Gene Expression
Portal Hep
Mid Zone Hep
Central Hep
Sinusoidal Endo
VV Endo
qHSC
Portal Fibro
Cholangiocytes
Kupffer Cells
Inflamm Macs
Immune cells
-2 -1 0 1 2
F
M
B. WikiPathways - Female
Ferroptosis
Adipogenesis
Iron Metabolism Disorders
VEGFA VEGFR2 Signaling
Familial Hyperlipidemia Type 3
Metabolic Pathway Of LDL HDL
Familial Hyperlipidemia Type 2
Estrogen Receptor Pathway
Estrogen Metabolism
TGF Beta Signaling Pathway
Fatty Acid Omega Oxidation
PPAR Signaling Pathway
Cholesterol Metabolism
Vitamin D Receptor Pathway
Lipid Particles Composition
Farnesoid X Receptor Pathway
Combined Score
-1 0 1 2 3
C. WikiPathways - Male
Farnesoid X Receptor Pathway
Urea Cycle And Metabolism Of Amino Groups
Glucocorticoid Receptor Pathway
Hippo Signaling Regulation
VEGFA VEGFR2 Signaling
Chemokine Signaling Pathway
Orexin Receptor Pathway
Thyroid Stimulating Hormone Signaling Pathway
Combined Score
-1 0 1 2 3
D.
Relative information
1.00
0.75
0.50
0.25
0.00
Female
Male
ApoA PTPR ANGPTL FN1 VISFATIN MHC-II VEGF AGT ANGPT FLRT EPHB CADM CLDN CDH5 ICAM COLLAGEN PARs ADGRL APP HGF COMPLEM SLIT UNC5 ADGRG PROS NRG CDH CALCR CD45 THBS LAMININ SEMA6 ADGRE PECAM2 PTPRM PECAM1 Netrin PDGF IGF EPHA VCAM NECTIN BMP SEMA3 NCAM CDH1
E.
Female
ANGPTL
Male
qHSC
Mid Zone Hep
Immune Cells
Central Hep
Kupffer Cells
VV Endo
Portal Hep
Inflamm Macs
Cholangiocytes
Portal Fibro
Sinusoidal Endo
F.
ApoA
Female (Only)
G.
Female
BMP
Male
H.
NCAM
Male (Only)

Fig. 3. Sex-specific gene expression, pathway enrichment, and signaling. (A) Heatmap displaying the log2 fold change of gene expression between females and males across cell types and subtypes within the liver. Red indicates higher expression levels in males, while blue signifies higher expression levels in females. The top 5 differentially expressed genes, upregulated in each group (female or male) and expressed in at least 25% of cells are displayed for each cell type and subtype. **(B)** Biological processes (Wiki pathways) enriched in differentially expressed genes across cell types in females. The combined score was calculated by logarithmically transforming the p-value from Fisher's exact test and multiplying it by the z-score. **(C)** Biological processes (Wiki pathways) enriched in differentially expressed genes across cell types in males. **(D)** Significant signaling pathways were ranked based on differences in overall information flow within the inferred networks between females and males. The top signaling pathways labeled in blue are more enriched in females. Pathways labeled in orange are more enriched in males. Pathways labeled in black are not statistically enriched in either category. **(E)** Chord circle plots display significantly interacting pathways and communication probabilities of ANGPTL signaling (higher in females compared to males). **(F)** Chord circle plot displays significantly interacting pathways and communication probabilities of ApoA signaling, (present only in females). **(G)** Chord circle plot displays significantly interacting pathways and communication probabilities of BMP signaling(higher in males as compared to females). **(H)** Chord circle plot displays significantly interacting pathways and communication probabilities of NCAM signaling (present only in males).

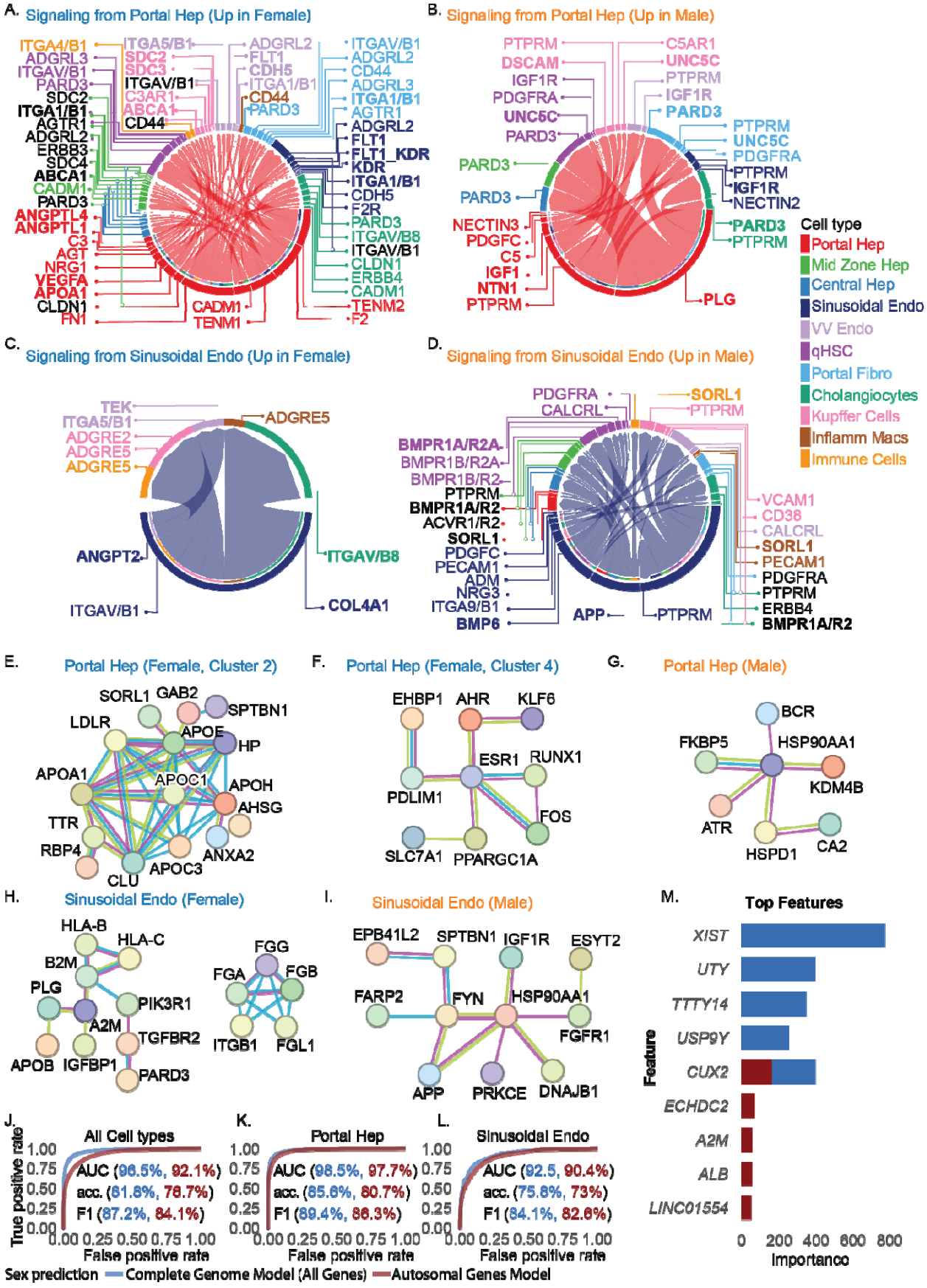

A. Signaling from Portal Hep (Up in Female)
ITGA4/B1
ADGRL3
ITGAV/B1
PARD3
SDC2
ITGA1/B1
AGTR1
ADGRL2
ERBB3
SDC4
ABCA1
CADM1
PARD3
ANGPTL4
ANGPTL1
C3
AGT
NRG1
VEGFA
APOA1
CLDN1
FN1
ITGA5/B1
SDC2
SDC3
ITGAV/B1
C3AR1
ABCA1
CD44
ADGRL2
FLT1
CDH5
ITGA1/B1
CD44
PARD3
CADM1
TENM1
ITGAV/B1
ADGRL2
CD44
ADGRL3
ITGA1/B1
AGTR1
ADGRL2
FLT1
FLT1_KDR
KDR
ITGA1/B1
CDH5
F2R
PARD3
ITGAV/B8
ITGAV/B1
CLDN1
ERBB4
CADM1
TENM2
F2
B. Signaling from Portal Hep (Up in Male)
PTPRM
DSCAM
IGF1R
PDGFRA
UNC5C
PARD3
PARD3
PARD3
NECTIN3
PDGFC
C5
IGF1
NTN1
PTPRM
C5AR1
UNC5C
PTPRM
IGF1R
PARD3
PTPRM
UNC5C
PDGFRA
PTPRM
IGF1R
NECTIN2
PARD3
PTPRM
PLG
Cell type
Portal Hep
Mid Zone Hep
Central Hep
Sinusoidal Endo
VV Endo
qHSC
Portal Fibro
Cholangiocytes
Kupffer Cells
Inflamm Macs
Immune Cells
C. Signaling from Sinusoidal Endo (Up in Female)
TEK
ITGA5/B1
ADGRE2
ADGRE5
ADGRE5
ADGRE5
ANGPT2
ITGAV/B1
ITGAV/B8
COL4A1
D. Signaling from Sinusoidal Endo (Up in Male)
PDGFRA
CALCRL
BMPR1A/R2A
BMPR1B/R2A
BMPR1B/R2
PTPRM
BMPR1A/R2
ACVR1/R2
SORL1
PDGFC
PECAM1
ADM
NRG3
ITGA9/B1
BMP6
SORL1
PTPRM
VCAM1
CD38
CALCRL
SORL1
PECAM1
PDGFRA
PTPRM
ERBB4
BMPR1A/R2
APP
PTPRM
E. Portal Hep (Female, Cluster 2)
SORL1
GAB2
SPTBN1
LDLR
APOE
HP
APOA1
APOC1
APOH
AHSG
TTR
RBP4
ANXA2
CLU
APOC3
F. Portal Hep (Female, Cluster 4)
EHBP1
AHR
KLF6
ESR1
RUNX1
PDLIM1
FOS
SLC7A1
PPARGC1A
G. Portal Hep (Male)
BCR
FKBP5
HSP90AA1
KDM4B
ATR
HSPD1
CA2
H. Sinusoidal Endo (Female)
HLA-B
HLA-C
B2M
PLG
PIK3R1
A2M
TGFBR2
APOB
IGFBP1
PARD3
FGG
FGA
FGB
ITGB1
FGL1
I. Sinusoidal Endo (Male)
EPB41L2
SPTBN1
IGF1R
ESYT2
FARP2
FYN
HSP90AA1
FGFR1
APP
PRKCE
DNAJB1
M. Top Features
XIST
UTY
TTTY14
USP9Y
CUX2
ECHDC2
A2M
ALB
LINC01554
Feature
0
200
400
600
800
Importance
J. All Cell types
AUC (96.5%, 92.1%)
acc. (81.8%, 78.7%)
F1 (87.2%, 84.1%)
K. Portal Hep
AUC (98.5%, 97.7%)
acc. (85.6%, 80.7%)
F1 (89.4%, 86.3%)
L. Sinusoidal Endo
AUC (92.5, 90.4%)
acc. (75.8%, 73%)
F1 (84.1%, 82.6%)
True positive rate
False positive rate
Sex prediction
Complete Genome Model (All Genes)
Autosomal Genes Model

Fig. 4. Sex-specific intercellular interactions, protein networks, and application of machine learning. (A-D) Chord diagrams representing sex-specific ligand-receptor (L-R) interactions. Outer rings denote different cell types, while links illustrate interactions, with ligands expressed by the source cell type and receptors expressed by the target cell type. Labels connected to more than one cell type by colored lines in the diagram are indicated in black. Examples discussed in the main text are in bold. L-R signaling originating from portal hepatocytes enriched in females (**A**), and males (**B**). L-R signaling originating from sinusoidal endothelial cells enriched in females (**C**) and males (**D**). **(E-H)** STRINGDB network analysis of protein-protein interactions (PPIs) Nodes represent proteins, while edges (lines) depict interactions, with different colors indicating evidence from distinct sources (Red: fusion evidence, Green: neighborhood evidence, Blue: co-occurrence evidence, Purple: experimental evidence, Yellow: text mining evidence, Light blue: database evidence, Black: co-expression evidence). **(E)** Protein network in female portal hepatocytes enriched in pathways related to lipid metabolism. **(F)** Protein network in female portal hepatocytes enriched in estrogen signaling and oncostatin M signaling pathways. **(G)** Protein network in male portal hepatocytes. **(H)** Protein network in female sinusoidal endothelial cells. **(I)** Protein network in male sinusoidal endothelial cells. **(J-L)** Machine learning-based prediction of sex using **(J)** all cell types combined, **(K)** portal hepatocytes, and **(L)** sinusoidal endothelial cells. The blue line indicates the AUC for models trained on all genes (including sex chromosomes), while the red line represents models trained exclusively on autosomal genes. **(M)** The plot shows the top 5 features for predicting sex in the Complete Genome Model (blue bars) and the Autosomal Gene Model (red bars). The y-axis represents feature importance scores, with higher values indicating greater contributions.

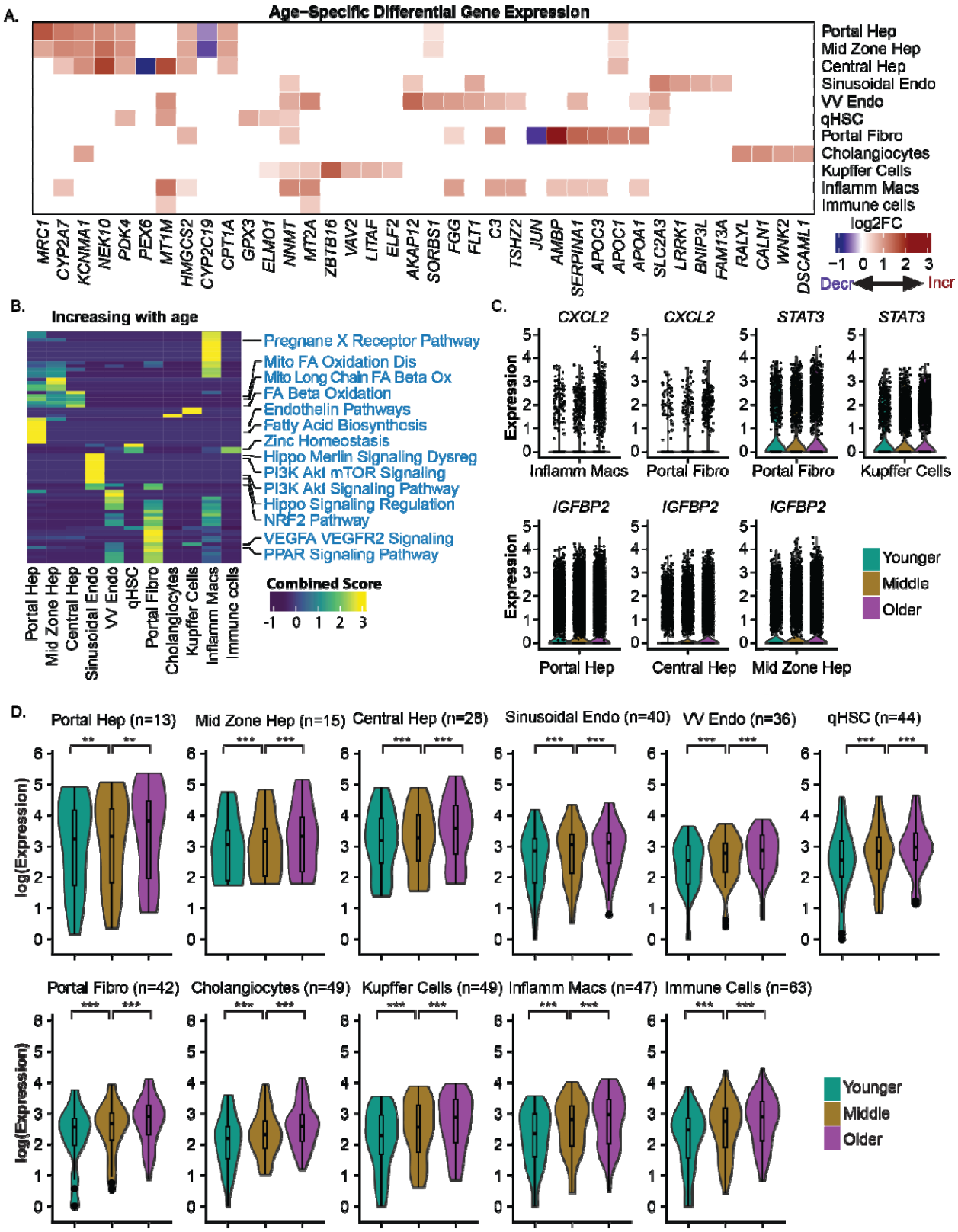

A.
Age-Specific Differential Gene Expression
Portal Hep
Mid Zone Hep
Central Hep
Sinusoidal Endo
VV Endo
qHSC
Portal Fibro
Cholangiocytes
Kupffer Cells
Inflamm Macs
Immune cells
MRC1
CYP2A7
KCNMA1
NEK10
PDK4
PEX6
MT1M
HMGCS2
CYP2C19
CPT1A
GPX3
ELMO1
NNMT
MT2A
ZBTB16
VAV2
LITAF
ELF2
AKAP12
SORBS1
FGG
FLT1
C3
TSHZ2
JUN
AMBP
SERPINA1
APOC3
APOC1
APOA1
SLC2A3
LRRK1
BNIP3L
FAM13A
RALYL
CALN1
WNK2
DSCAML1
log2FC
-1 0 1 2 3
Decr
Incr
B.
Increasing with age
Pregnane X Receptor Pathway
Mito FA Oxidation Dis
Mito Long Chain FA Beta Ox
FA Beta Oxidation
Endothelin Pathways
Fatty Acid Biosynthesis
Zinc Homeostasis
Hippo Merlin Signaling Dysreg
PI3K Akt mTOR Signaling
PI3K Akt Signaling Pathway
Hippo Signaling Regulation
NRF2 Pathway
VEGFA VEGFR2 Signaling
PPAR Signaling Pathway
Portal Hep
Mid Zone Hep
Central Hep
Sinusoidal Endo
VV Endo
qHSC
Portal Fibro
Cholangiocytes
Kupffer Cells
Inflamm Macs
Immune cells
Combined Score
-1 0 1 2 3
C.
CXCL2
CXCL2
STAT3
STAT3
Expression
Inflamm Macs
Portal Fibro
Portal Fibro
Kupffer Cells
IGFBP2
IGFBP2
IGFBP2
Portal Hep
Central Hep
Mid Zone Hep
Younger
Middle
Older
D.
Portal Hep (n=13)
Mid Zone Hep (n=15)
Central Hep (n=28)
Sinusoidal Endo (n=40)
VV Endo (n=36)
qHSC (n=44)
log(Expression)
Portal Fibro (n=42)
Cholangiocytes (n=49)
Kupffer Cells (n=49)
Inflamm Macs (n=47)
Immune Cells (n=63)
Younger
Middle
Older

Fig. 5. Age-related gene expression changes and senescence signatures. (A) Heatmap displaying the log2 fold change of gene expression between age groups across cell types. Red indicates increase in expression with age and blue decrease in expression with age. A maximum of 5 differentially expressed genes, upregulated in each group (increasing or decreasing with age) and expressed in at least 25% of cells are displayed for each cell type and subtype. **(B)** Heatmap presents a gene ontology enrichment analysis of up-regulated genes across different age groups, focusing on pathways that increase with age. Enrichr combined score is calculated by the logarithmic transformation of the p-value obtained from Fisher's exact test, multiplied by the z-score. **(C)** Violin plots in each panel show gene expression across three age groups: Younger, Middle, and Older. These plots highlight statistically significant changes in senescence-associated genes (Saul et al., 2022) identified in the dataset. **(D)** Cell type-specific senescence signatures. Violin plots of senescence gene expression for differentially expressed genes across indicated groups, with average expression values displayed on a log scale. ** indicates $p < 0.005$ and *** $p < 0.001$ as obtained from pairwise Wilcoxon tests.

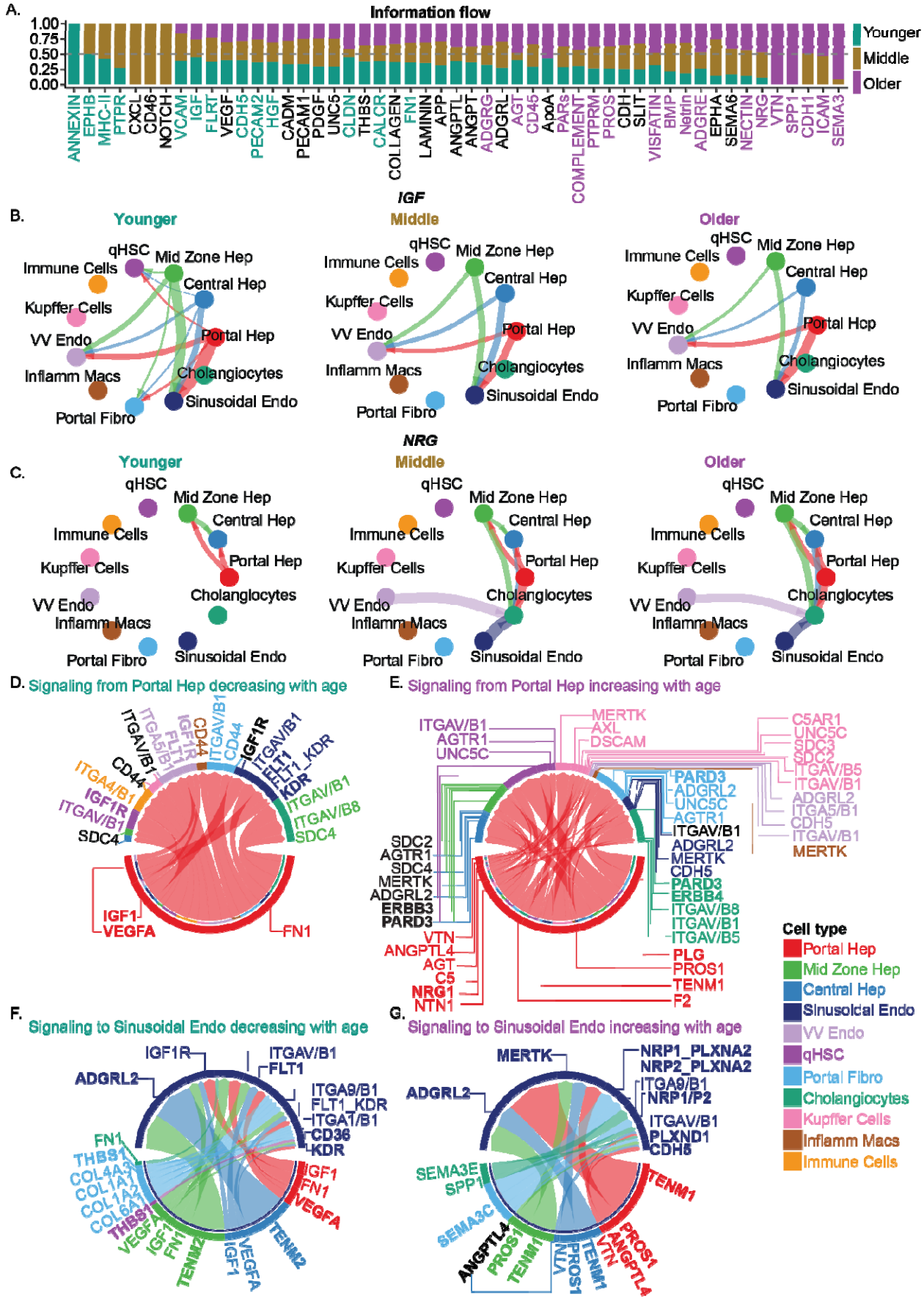

A.
Information flow
1.00
0.75
0.50
0.25
0.00
Younger
Middle
Older
ANNEXIN
EPHB
MHC-II
PTPR
CXCL
CD46
NOTCH
VCAM
IGF
FLRT
VEGF
CDH5
PECAM2
HGF
CADM
PECAM1
PDGF
UNC5
CLDN
THBS
CALCR
COLLAGEN
FN1
LAMININ
APP
ANGPTL
ANGPT
ADGRG
ADGRL
AGT
CD45
ApoA
PARs
COMPLEMENT
PTPRM
PROS
CDH
SLIT
VISFATIN
BMP
Netrin
ADGRE
EPHA
SEMA6
NECTIN
NRG
VTN
SPP1
CDH1
ICAM
SEMA3
IGF
B.
Younger
Middle
Older
qHSC
Mid Zone Hep
Immune Cells
Central Hep
Kupffer Cells
VV Endo
Portal Hep
Inflamm Macs
Cholangiocytes
Sinusoidal Endo
Portal Fibro
NRG
C.
D. Signaling from Portal Hep decreasing with age
E. Signaling from Portal Hep increasing with age
F. Signaling to Sinusoidal Endo decreasing with age
G. Signaling to Sinusoidal Endo increasing with age
Cell type
Portal Hep
Mid Zone Hep
Central Hep
Sinusoidal Endo
VV Endo
qHSC
Portal Fibro
Cholangiocytes
Kupffer Cells
Inflamm Macs
Immune Cells

Fig. 6. Age-related changes in signaling pathways and ligand-receptor interactions. (A) Signaling pathways were ranked according to differences in overall information flow within inferred networks across younger, middle, and older age groups. Pathways labeled in purple are more enriched in older age groups. Pathways labeled in teal are more enriched in younger individuals.Pathways labeled in black show no significant enrichment between younger and older cells. **(B)** The chord circle plot illustrates IGF signaling pathways across cell types in younger, middle, and older populations. Strength of communication is indicated by the thickness of the line, and arrows indicate direction. **(C)** NRG signaling across age groups. **(D-G)** Chord diagrams represent aging-specific ligand-receptor (L-R) interactions. Outer rings denote different cell types, while the links illustrate interactions, with ligands expressed by the source cell type and receptors expressed by the target cell type. Labels connected to more than one cell type by colored lines in the diagram are indicated in black. Examples discussed in the main text are in bold. L-R signaling originating from portal hepatocytes enriched with decreased **(D)** and increased age **(E)** L-R signaling to sinusoidal endothelial cells enriched with decreased (**F**) and increased age (**G**) age.

Single-cell transcriptomics reveals the impact of sex and age in the healthy human liver

Raza Ur Rahman, Eliana T. Epstein, Shane Murphy, Liat Amir-Zilberstein, Cristin McCabe, Toni Delorey, Hope Koene, Lilly Fernandes, Kenneth K. Tanabe, Motaz Qadan, Christina Ferrone, David L. Berger, Angela Shih, Jacques Deguine, Alan C. Mullen[†]

[†]Corresponding author. email: alan.mullen@umassmed.edu

Table of Contents

Supplementary Materials and Methods……..2

Supplementary Figures. ……..12

References…….25

Supplementary Materials and Methods

Single-nucleus RNA-seq quality control and analysis

Samples were demultiplexed using either Cell Ranger (*1*) mkfastq (for NovaSeq 6000) or bclconvert (for NovaSeqX). The reference for alignment was created using the GENCODE v42 annotation GTF and GRCh38.p13 genome FASTA files. Sequencing reads were aligned to the reference genome (including introns) using 10x Genomics Cell Ranger v7.0.1, resulting in count matrices for each sample. Seurat (v4.3.0) (*2*) objects were created for each sample, requiring each gene to be detected in at least 10 cells. Cells with fewer than 500 genes detected, fewer than 500 UMIs, or more than 10% of reads mapping to mitochondrial or ribosomal genes were excluded from the analysis.

To remove ambient RNA contamination, we used SoupX version 1.6.2 (*3*), which corrects for ambient RNA in single-cell RNA sequencing data. Since our dataset is predominantly composed of hepatocytes (**Fig 1D, table S2 and table S7**), we focused on genes highly expressed in hepatocytes. We excluded ambient RNA corresponding to hepatocyte marker genes from MacParland et al (*4*), except for ALB, as well as genes from the cytochrome P450 (CYP) family highly expressed in our dataset. The MacParland et al (*4*) gene set includes CYP2E1, ASGR1, CYP3A4, SCD, SLP1, CYP2B6, CPS1, CRP, and A2M. Additionally, the CYP family genes highly expressed in our dataset include CYP27A1, CYP4A11, CYP4V2, CYP2B7P, CYP2A6, CYP2A7, CYP3A5, CYP2C9, CYP2C8, CYP4F3, and CYP3A43. Each sample was processed individually using SoupX (*3*) to filter ambient RNA contamination. SoupX did not detect any ambient RNA contamination in samples Liver-10, Liver-14, Liver-18, Liver-23, Liver-25, Liver-28, and Liver-30, as no cells were identified as acceptable for estimation using the specified gene sets. For sample Liver-16, the analysis was performed in default mode without providing gene sets, as ambient RNA contamination was overestimated using the selected gene sets.

Furthermore, each sample was processed individually, and doublets or multiplets were filtered out using DoubletFinder v2.0.4 (*5*).

After these filtering steps, the dataset contained high-quality cells suitable for downstream analyses. Individual Seurat objects were merged and normalized using the SCTransform method in Seurat v4.3.0, with mitochondrial reads regressed out. Principal component analysis (PCA) was conducted on the normalized data using the 'RunPCA' function. Integration and batch correction across samples were performed using Harmony v1.2.0 (*6*) with the RunHarmony function. A shared-nearest-neighbor graph was created using the FindNeighbors function, and cells were clustered with the Louvain algorithm using the FindClusters function. A resolution of 0.7 was selected, resulting in 22 clusters (0-21), chosen from a range of tested values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0). Cluster visualization in two dimensions was performed using UMAP with the `RunUMAP` function, employing the euclidean distance measure. Cell-type proportions were analyzed with the speckle R package v0.99.7 (*7*) using the 'propeller' function, with cell-type annotations, sample IDs, and metadata columns (Sex or AgeGroup) as input.

Cell annotations

Clusters identified were assigned cell type labels using the `sctype_score` function from the ScType R package (*8*), utilizing the 'ScTypeDB_full.xlsx' database downloaded on 26/06/2023. This package assigns labels based on marker genes from the ScType database specific to the tissue of interest.

Clusters were categorized using "Liver" as the reference tissue with ScType. This resulted in two hepatic stellate cell clusters (clusters 9 and 14), one immune system cluster (cluster 12), two macrophage clusters (clusters 10 and 15), four endothelial cell clusters (clusters 7, 16, 18, and 21), one cholangiocyte cluster (cluster 13), and an unknown cluster (cluster 19 containing

555 nuclei) that could not be assigned a specific cell type. The remaining clusters were annotated as hepatocytes using ScType. Marker genes were selected through differential expression analysis for each cell type.

Post-cluster quality control

In the post-cluster quality control process, we utilized MALAT1 expression as a measure of high-quality nuclei. MALAT1, a long noncoding RNA retained in the nucleus, is considered a marker for high-quality cells (*9*, *10*). To identify potential low-quality cells, we examined the expression of MALAT1 across clusters. Cells with fewer than three reads for MALAT1 were considered as having low expression, which could indicate lower-quality nuclei. This was observed in Cluster 11 (hepatocytes with 5574 nuclei) and Cluster 19 (unknown cell type with 555 nuclei), where the majority of nuclei showed MALAT1 expression below this threshold. Therefore, we excluded these two clusters (containing 6129 nuclei) from further downstream analysis.

Hepatocytes zonation analysis

Hepatocytes were classified into periportal, pericentral, and intermediate zones based on the expression of key zonal markers: ASS1, ASL, GLUL, and CYP2E1 (*11*). To differentiate hepatocyte clusters into these zones, a Central-to-Portal Ratio (CPR) was calculated for each cluster by subtracting the summed expression of portal markers ASS1 and ASL from the summed expression of central markers GLUL and CYP2E1. This CPR value serves as a quantitative measure of zonal distribution, reflecting the relative expression levels of these markers within each cluster.

Clusters were categorized into three zones based on their CPR values. Clusters with CPR values less than or equal to the lower quantile were classified as "portal" hepatocytes, representing the periportal zone. Clusters with CPR values greater than or equal to the upper

quantile were classified as "central" hepatocytes, representing the pericentral zone. Clusters with CPR values falling between the lower and upper quantiles were classified as "mid-zone" hepatocytes, representing an intermediate zone. The classification process was applied to all of hepatocyte clusters, with portal zones corresponding to Clusters 3, 4, 6, and 8; central zones corresponding to Clusters 1, 5, 17, and 20; and mid-zone hepatocytes corresponding to Clusters 0 and 2.

To assess zonal classification, we employed UMAP (Uniform Manifold Approximation and Projection) visualization, which allowed for the spatial separation of the hepatocyte zones. To assess the enrichment of portal markers (ASS1 and ASL) and central markers (GLUL and CYP2E1) within the hepatocyte clusters, we generated density plots for individual genes and their joint density using the 'do_NebulosaPlot' function in SCpubr R package (version 2.0.0.9000). These plots allowed us to visually assess the spatial distribution of marker expression within the UMAP embedding. The joint density plots provided an integrated view of marker co-expression, offering additional validation for the zonal classification derived from the quantitative Central-to-Portal Ratio (CPR) scores. This approach highlighted the enrichment of portal and central markers in the respective clusters and aligned with the CPR-based classification, providing additional support for the zonal annotation strategy.

Immune system cell sub-clustering and annotation

The immune system cell population (cluster 12 with 5,239 cells) identified in the analysis was further sub-clustered from the raw data to achieve greater resolution of immune cell subtypes. Sub-clustering was performed using the same analysis pipeline described above. Cell clusters were identified by testing resolution values from 0.1 to 1.0, in increments of 0.1, and a resolution of 0.2 was selected based on the ability to define five distinct clusters (0-4). Annotation of the clusters was performed using the ScType tool with "Immune System Cells" set as the target

tissue. Cluster 0, consisting of 1,684 cells, was annotated with equal scores as Memory/Effector CD4/CD8 T cells and general T cells and was renamed as T cells 1. Cluster 1, containing 1,226 cells, was annotated as Granulocytes. Cluster 2, which contained 944 cells, was unannotated and marked as unknown by ScType. It was also enriched in hepatocyte genes. This cluster was excluded since it was not a clear immune population. Clusters 3 and 4, containing 776 and 609 cells respectively, were annotated with equal scores as CD4+ and CD8+ NKT-like cells and were renamed NKT-like cells-1 and NKT-like cells-2, respectively. We then examined the expression of immune system cell marker genes from ScType across each cluster. T cell markers including CD2, CD44, and CD69 were expressed in both Cluster 0 and Cluster 1. CD44 and CD69 were the only granulocyte markers from ScType expressed in Cluster 1, but these genes are also expressed in T cells. Similarly, LCK, a T cell marker from ScType, was expressed in Cluster 1, while PTPRC (expressed in Clusters 0, 1, 3, 4) and GZMA (expressed in Clusters 1, 3, 4) were found across multiple immune clusters. Based on these findings, we renamed Cluster 1 as a second T cell population, designated it T cells 2.

Differential expression and gene ontology enrichment analysis

We conducted differential expression analysis separately for both sex and age groups to explore how gene expression patterns vary. To enhance statistical robustness and reduce the number of comparisons, we categorized our samples into distinct age brackets: younger (18-40 years), middle (41-60 years), and older (61-90 years). Using Seurat version 4.3.0, we compared gene expression profiles between sexes (males vs. females) and across different age groups. Differential gene expression analysis was performed using the MAST test implemented in the 'FindMarkers' function of Seurat version 4.3.0, incorporating donor as a latent variable to account for batch effects. Differentially expressed genes between sexes (males vs. females) and across different age groups (younger, middle and older) were determined for each cell type.

For age-related analysis, pairwise comparisons were performed between younger and middle-aged groups, and between middle-aged and older groups, for each cell type. Relatively few genes showed statistical significance across comparisons (younger < middle < older) or (younger > middle > older) within cell type. To define genes that increase with age (younger < middle < older) or decrease with age (younger > middle > older), we first identified statistically significant differentially expressed genes between the younger and older-aged groups. From this list, we retained genes that were following the increasing (younger < middle < older) or decreasing (younger > middle > older) expression in the middle aged group within the same cell type. Gene ontology enrichment analysis for biological processes was performed using EnrichR version 3.2 (*12*) library WikiPathway_2023_Human. Enrichr combined score is calculated by the logarithmic transformation of the p-value obtained from Fisher's exact test, multiplied by the z-score representing the deviation from the expected rank (**table S10, table S11, table S16** and **table S17**).

STRING-DB network analysis

We analyzed protein-protein interactions (PPIs) based on sex differences using the R package STRINGdb v2.10.1 (*13*). The STRING database (species = 9606, version = 11.5) was accessed with the network type set to physical. After performing differential expression analysis, STRING was used to generate PPI networks for the upregulated genes in males and females (excluding mitochondrial and ribosomal genes), enabling the assessment of potential biological interactions between differentially expressed genes. Gene ontology enrichment analysis was conducted on the STRING online server (https://version-11-5.string-db.org/ accessed in November 2024) to identify WikiPathways associated with these genes.

Cell-cell communication analysis

We performed cell-cell communication analysis to investigate how intercellular signaling varies by sex and age. Using CellChat v2.1.2 (*14*), we analyzed communication patterns across sex (males vs. females) and age groups (younger: 18-40 years, middle: 41-60 years, older: 61-90 years). CellChat was applied independently to each group and then merged for comparison using the mergeCellChat function (male vs. female for sex, and younger, middle, and older groups for age). We employed key functions such as computeCommunProb, filterCommunication, computeCommunProbPathway, aggregateNet, and netAnalysis_computeCentrality with default parameters, using the CellChatDB.human ligand-receptor interaction database. Interactions where ligands were more highly expressed in the receiver cell type than in the sender were filtered.

To rank the resulting signaling networks in sex, we used the rankNet function to perform a paired Wilcoxon test (with do.stat = TRUE) in CellChat to compare differences in signaling information flow in female vs. male. This test assessed whether the overall communication probability, defined by the sum of the weights in the network, differed significantly between male vs. female for each signaling pathway. Significant pathways were ranked based on the overall information flow, with pathways enriched in males or females highlighted accordingly.

To rank the signaling networks across the age groups, we used the rankNet function in CellChat to analyze signaling information flow among younger (18–40 years), middle-aged (41–60 years), and older (61–90 years) cells. Since CellChat does not provide statistical significance for comparisons involving more than two groups, we focused on pairwise comparisons between younger and older groups to identify pathways with significant differences. These significant pathways were then highlighted in the three-group comparison to provide a comprehensive view across all age groups. These results were used to identify sex- and age-specific differences in

cell-cell communication. Circle plots generated by the netVisual_aggregate function visualized communication patterns across groups.

To identify differentially regulated ligand-receptor (LR) pairs between sexes, we used females as the positive dataset (positive fold change compared to males) in the identifyOverExpressedGenes function. LR pairs upregulated in both sexes in different cell types were excluded.

For age-specific signaling and ligand-receptor (LR) pairs analysis, we aimed to identify ligand-receptor (LR) pairs that were differentially enriched in younger or older age groups compared to the other groups. Specifically, the younger group (18-40 years) was compared against both the middle-aged (41-60 years) and older (61-90 years) groups, treating the younger group as the positive dataset in the `identifyOverExpressedGenes` function. Similarly, the older group was compared against the younger and middle-aged groups, with the older group as the positive dataset. This allowed us to identify LR pairs uniquely enriched in either younger or older age groups. To ensure the specificity of our results, we excluded LR pairs that were enriched in both younger and older groups, as these could result from different cell type contributions rather than true age-specific signaling differences.

Nuclei sex prediction using machine learning

The analysis of sex classification in liver cells was performed using a Random Forest classifier implemented through the ranger package version 0.16.0 to assess whether sex could be predicted based on gene expression profiles. The dataset was split into training (80%) and testing (20%) sets using the createDataPartition() function from the caret package version 6.0-93.

Two Random Forest models were trained: one using all expressed genes in the dataset, and the other excluding X and Y chromosome genes to evaluate the performance of autosomal genes alone in predicting sex.

To optimize the Random Forest model, hyperparameter tuning was performed using a grid search with the parameters `num.trees` (number of trees in the forest), `mtry` (the number of features sampled for each split), and `min.node.size` (minimum number of observations in a node). The grid search considered three values for each of these parameters: num.trees (100, 200, 300), mtry (5, 10, 20), and min.node.size (1, 5, 10). The best-performing model was selected based on the highest accuracy, and the optimal hyperparameters were used to train the final model. After selecting the best hyperparameters, the final model was trained using the entire training set and then evaluated on the test set. The ROC curve and AUC were computed and visualized using the pROC package version 1.18.0. The importance of individual features in the model was assessed using the built-in feature importance functionality in the ranger package. The features were ranked by their importance in predicting sex.

Sex balanced downsampling in age groups

To minimize the impact of sex bias in the analysis, we performed a downsampling approach to ensure equal representation of female and male nuclei within each cell type and age group. For each combination of cell type and age group, the total number of nuclei was determined for the less abundant sex (either female or male). Nuclei from the more abundant sex were then randomly downsampled to match this number, ensuring a balanced dataset for downstream analyses. This approach preserved the proportional distribution of nuclei across cell types and age groups while eliminating potential confounding effects of unequal sex representation.

Senescence genes increasing with age

To examine age-related changes in senescence-associated gene expression, we analyzed the dataset using the gene set defined by (*15*). Gene expression was evaluated across individual cell types and subtypes to identify genes whose expression consistently increased across the three age groups (younger < middle < older). Pairwise Wilcoxon tests were performed to assess the statistical significance of differences in expression between age groups (younger vs. middle and middle vs. older). For each comparison, p-values were calculated and visualized on violin plots. A threshold of $p < 0.05$ was used to define significant differences, with p-values further categorized as follows: $p < 0.001$ (***), $p < 0.01$ (**), $p < 0.05$ (*), and $p \geq 0.05$ (ns, not significant).

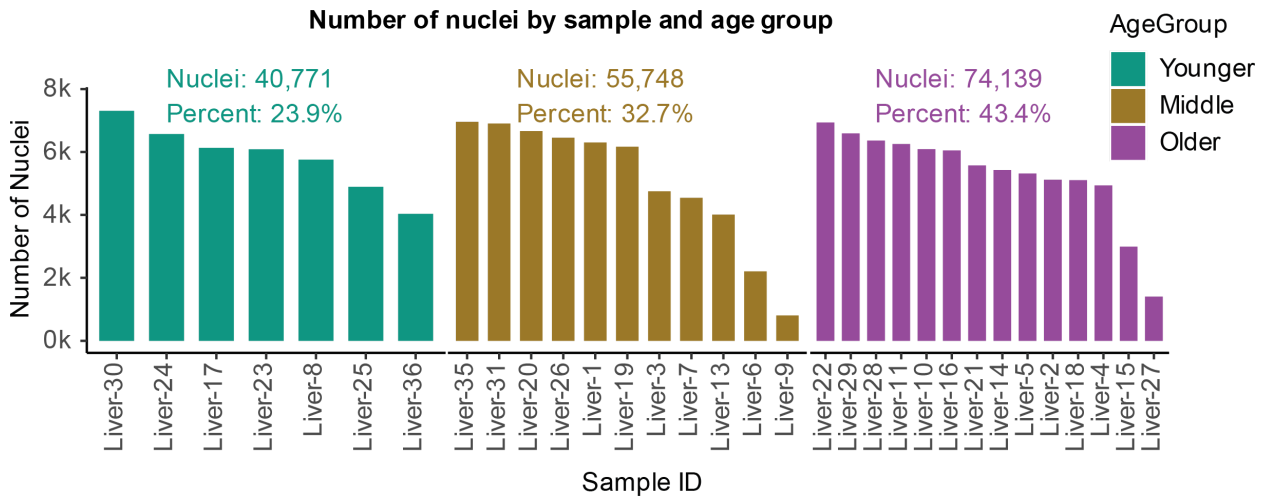


Fig S1. Number of nuclei in each sample, stratified by age group (young, middle, old). Bars represent the nuclei counts within individual samples, with young (colored teal), middle (colored mustard yellow), and old (colored purple) samples. This visualization provides an overview of sample composition by age group across the dataset.

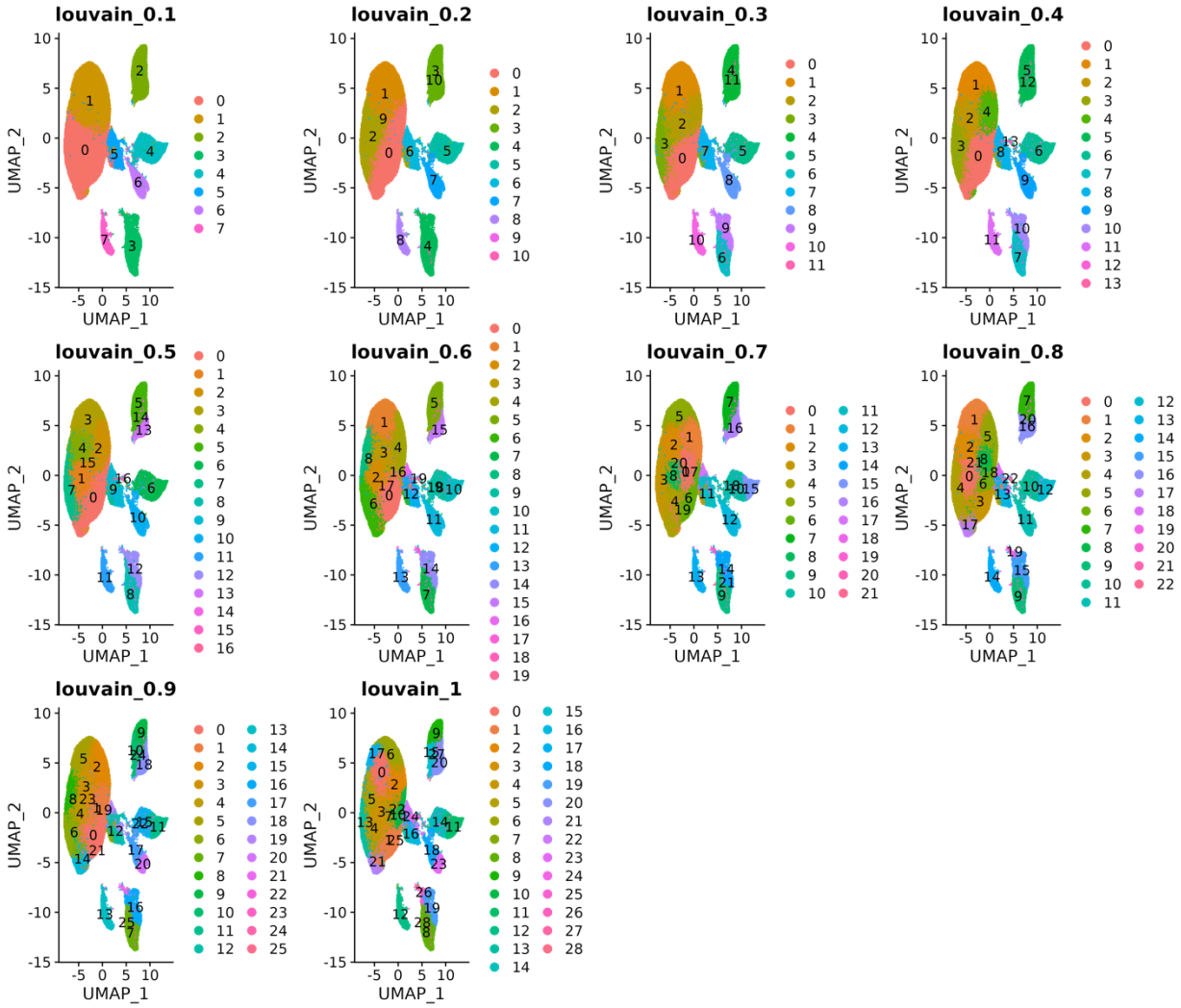
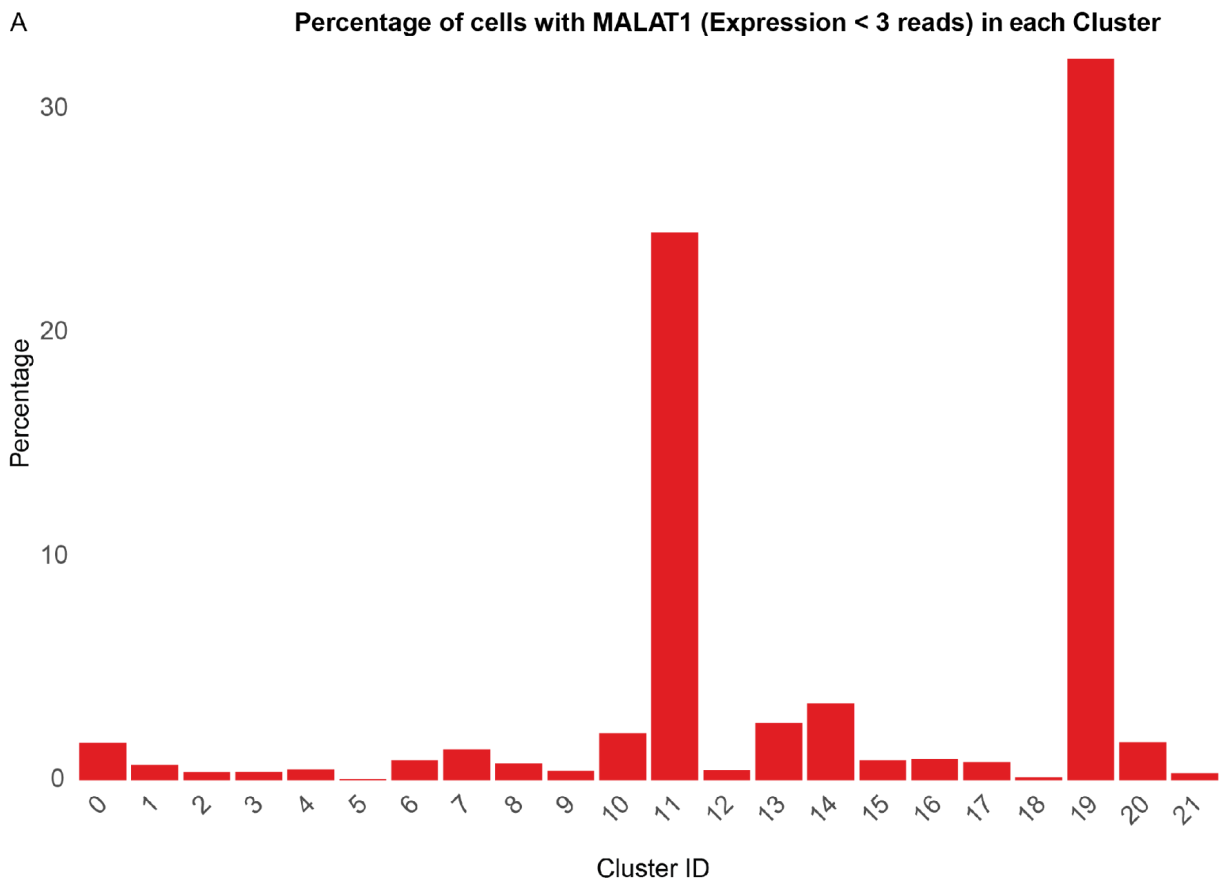


Fig S2. UMAP visualization of nuclei clusters from integrated samples across resolutions ranging from 0.1 to 1.0. Each resolution highlights the clustering patterns and the corresponding number of identified clusters. A resolution of 0.7, which produced 22 clusters (0–21), was chosen for downstream analysis due to its capacity to capture biologically meaningful sub-clusters, as shown in Fig 1C and Table S2.

A


Percentage of cells with MALAT1 (Expression < 3 reads) in each Cluster
Percentage
30
20
10
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Cluster ID

B

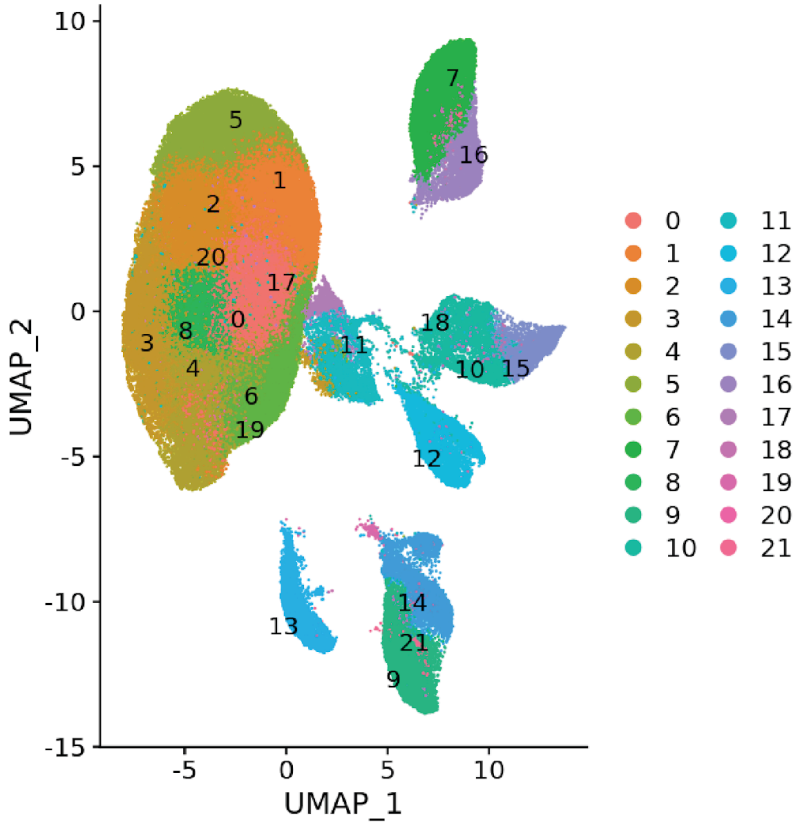

UMAP_2
UMAP_1
10
5
0
-5
-10
-15
-5 0 5 10
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Fig S3. (A) Histogram illustrating the distribution of *MALAT1* reads per nucleus across all clusters. Clusters 11 and 19 showed a notable reduction in *MALAT1* expression, with many nuclei exhibiting fewer than three detected reads. (B) UMAP visualization displaying all clusters, including clusters 11 and 19, which were excluded from downstream analyses due to low *MALAT1* expression (as shown in Fig 1C). These two clusters, containing 6,129 nuclei, were removed as *MALAT1* serves as a marker of nuclear quality (Clarke and Bader 2024; Montserrat-Ayuso and Esteve-Codina 2024).

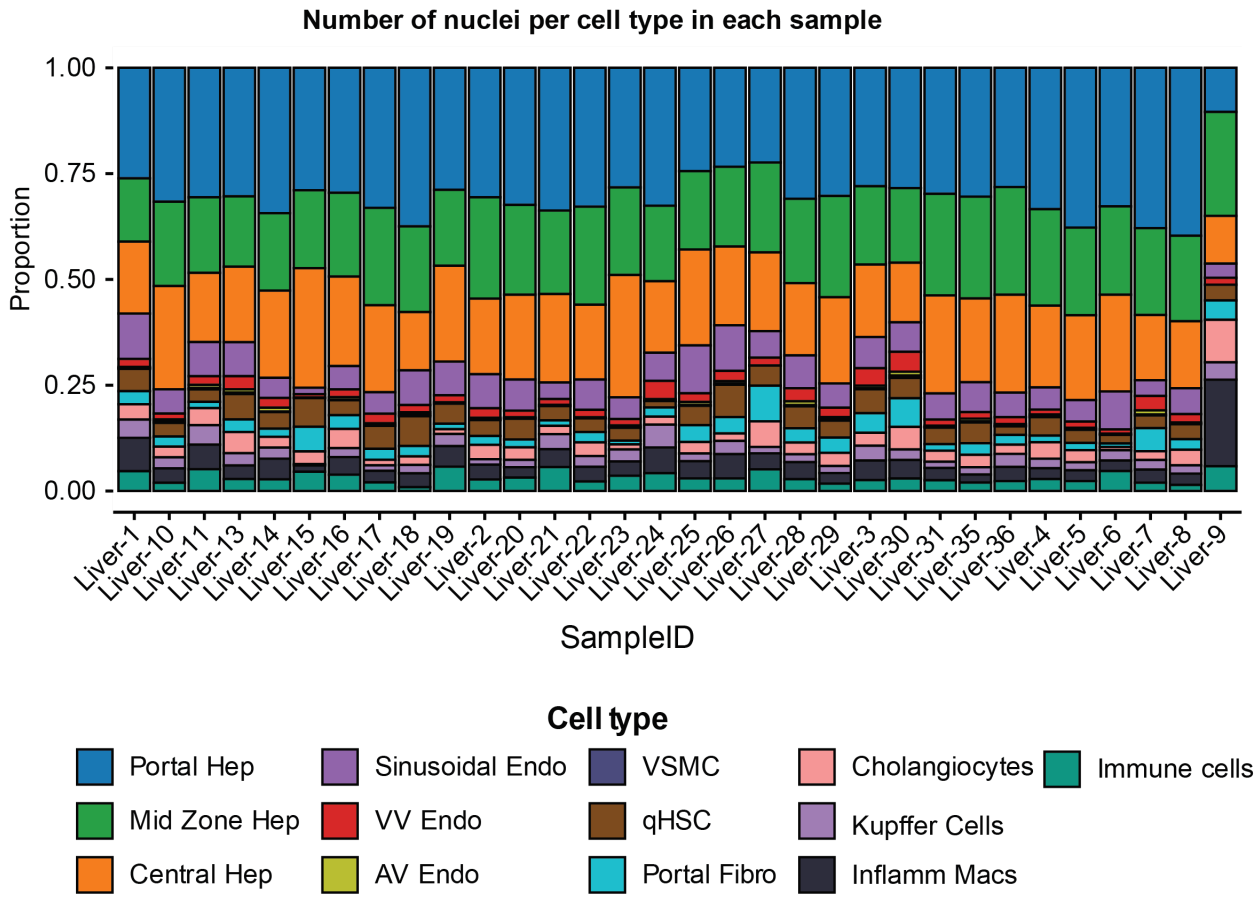


Fig S4. Cell type proportions across samples. This panel illustrates the proportions of each cell type across all samples, showing a uniform representation of cell types with no sample-specific biases. The even distribution confirms that cell types are consistently represented across samples, indicating well-balanced sampling and minimal influence of technical sample artifacts.

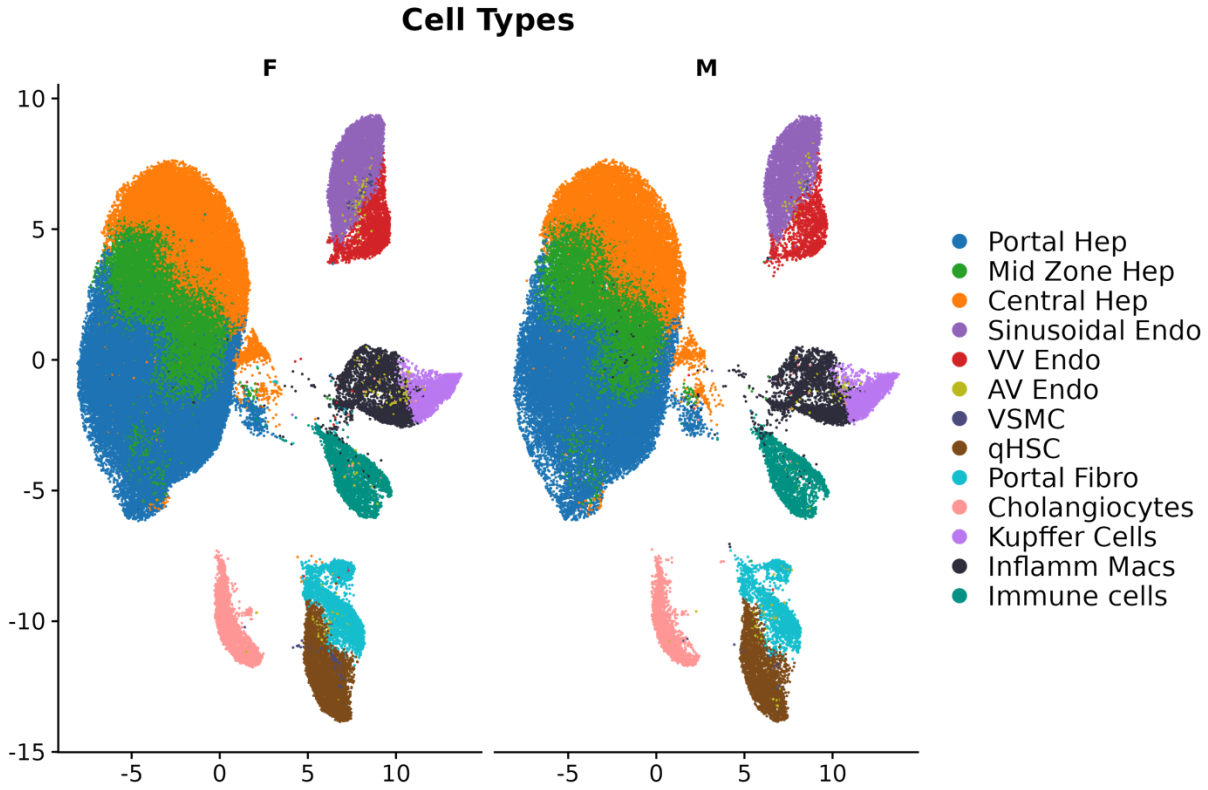


Fig S5. UMAP representation of cell types, split by sex, with female (F) cells on the left and male (M) cells on the right. The plot demonstrates a similar distribution of cell types across clusters, indicating comparable patterns of cell type composition between sexes.

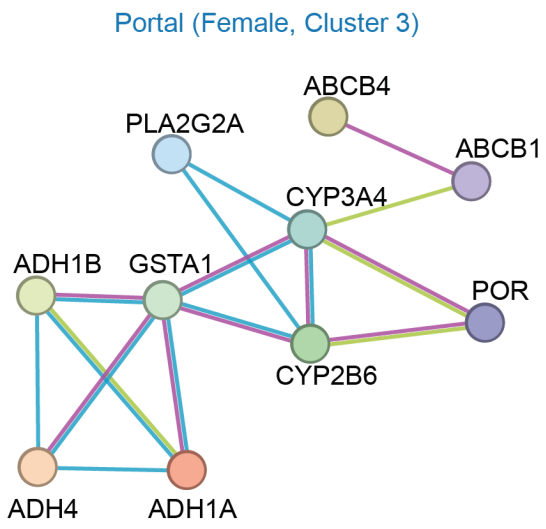


Fig S6. STRINGDB network analysis of protein-protein interactions (PPIs) in female liver cells. The networks visualize physical interactions between proteins based on evidence from STRINGDB. Nodes represent proteins, while edges (lines) depict interactions, with different colors indicating evidence from distinct sources. Protein network in female portal hepatocytes enriched in pathways related to farnesoid X receptor, pregnane X receptor, and estrogen metabolism.

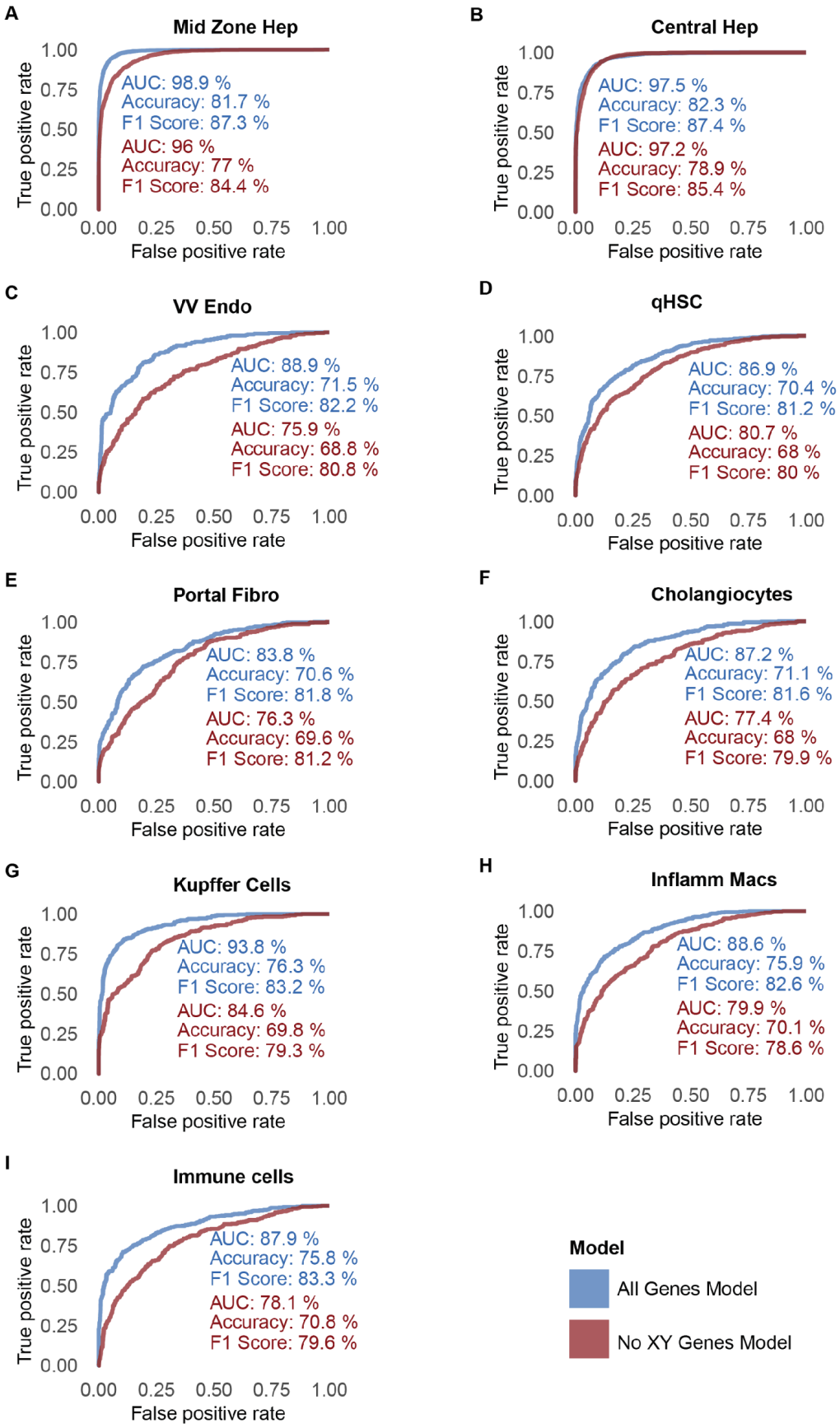

A
Mid Zone Hep
AUC: 98.9 %
Accuracy: 81.7 %
F1 Score: 87.3 %
AUC: 96 %
Accuracy: 77 %
F1 Score: 84.4 %
B
Central Hep
AUC: 97.5 %
Accuracy: 82.3 %
F1 Score: 87.4 %
AUC: 97.2 %
Accuracy: 78.9 %
F1 Score: 85.4 %
C
VV Endo
AUC: 88.9 %
Accuracy: 71.5 %
F1 Score: 82.2 %
AUC: 75.9 %
Accuracy: 68.8 %
F1 Score: 80.8 %
D
qHSC
AUC: 86.9 %
Accuracy: 70.4 %
F1 Score: 81.2 %
AUC: 80.7 %
Accuracy: 68 %
F1 Score: 80 %
E
Portal Fibro
AUC: 83.8 %
Accuracy: 70.6 %
F1 Score: 81.8 %
AUC: 76.3 %
Accuracy: 69.6 %
F1 Score: 81.2 %
F
Cholangiocytes
AUC: 87.2 %
Accuracy: 71.1 %
F1 Score: 81.6 %
AUC: 77.4 %
Accuracy: 68 %
F1 Score: 79.9 %
G
Kupffer Cells
AUC: 93.8 %
Accuracy: 76.3 %
F1 Score: 83.2 %
AUC: 84.6 %
Accuracy: 69.8 %
F1 Score: 79.3 %
H
Inflamm Macs
AUC: 88.6 %
Accuracy: 75.9 %
F1 Score: 82.6 %
AUC: 79.9 %
Accuracy: 70.1 %
F1 Score: 78.6 %
I
Immune cells
AUC: 87.9 %
Accuracy: 75.8 %
F1 Score: 83.3 %
AUC: 78.1 %
Accuracy: 70.8 %
F1 Score: 79.6 %
True positive rate
False positive rate
0.00
0.25
0.50
0.75
1.00
Model
All Genes Model
No XY Genes Model

Fig S7A-I. Machine learning-based prediction of nuclear sex using **(A)** mid zone hepatocytes, **(B)** central hepatocytes, (**C)** venous vascular endothelial cells, **(D)** quiescent HSCs, **(E)** portal fibroblasts, **(F)** cholangiocytes**, (G)** Kupffer like, **(H)** inflammatory macrophages, and **(I)** Non-macrophage immune cells. The blue line indicates the AUC for models trained on all genes (including sex chromosomes), while the red line represents models trained exclusively on autosomal genes. This analysis highlights that although sex chromosome genes improve performance, autosomal genes independently capture significant sex-specific expression patterns.

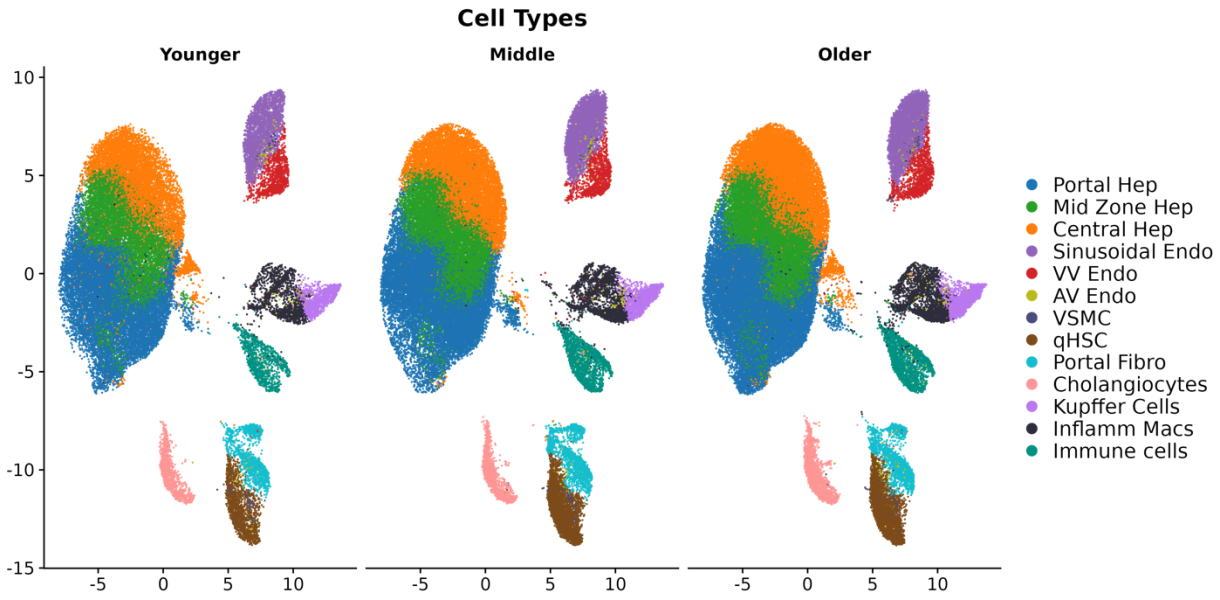


Fig S8. UMAP representation of cell types, split by age group, with younger cells on the left, middle-aged cells in the center, and older cells on the right. The plot demonstrates a similar distribution of cell types across clusters, indicating comparable patterns of cell type composition between age groups.

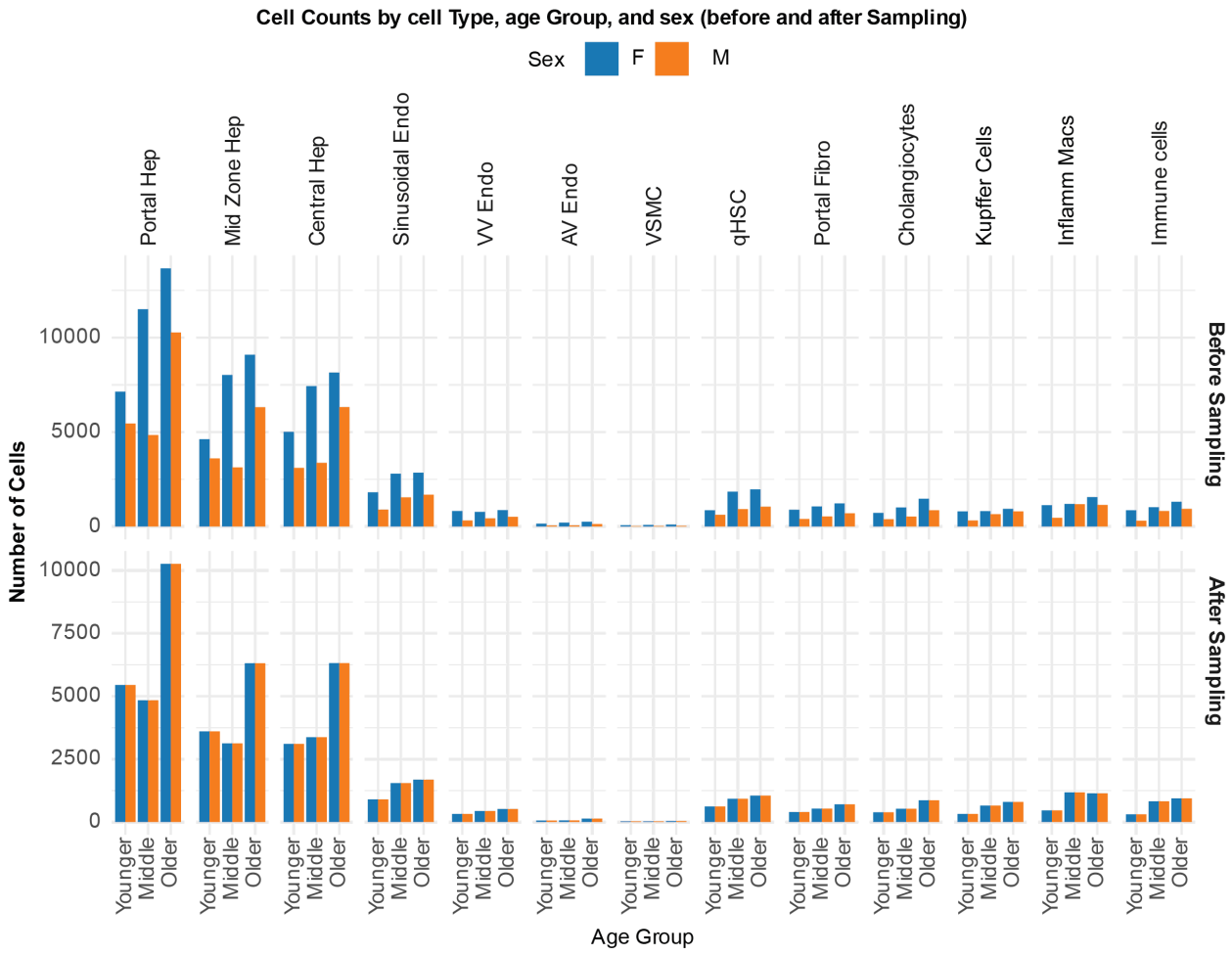


Fig S9. Comparison of cell counts by cell type, age group, and sex before and after down-sampling. The bar plots display the number of cells for each age group (younger, middle, and older) on the x-axis, grouped by sex (female in blue and male in orange). Panels are organized by cell type, with rows representing "Before Sampling" and "After Sampling," ordered to emphasize the impact of down-sampling. The colors distinguish between female and male nuclei, illustrating the balanced representation achieved by the down-sampling process within each cell type and age group. Notably, the ratio of nuclei from each sex (female and male) varies across different age groups within the cell type, highlighting differences in the initial distribution.

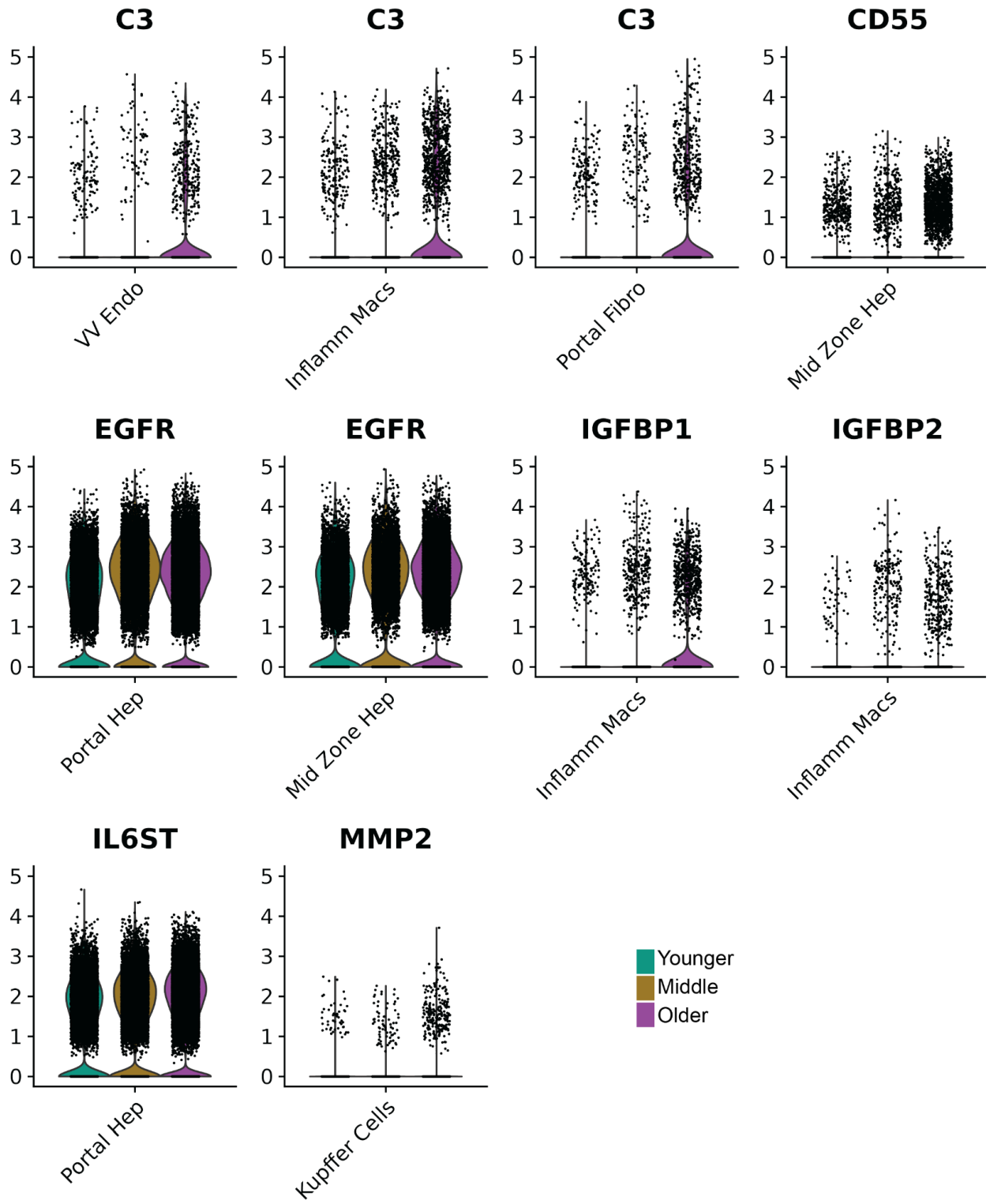


Fig S10. Violin plots in each panel show gene expression across three age groups: Younger, Middle, and Older. These plots highlight statistically significant changes in senescence-associated genes (Saul et al., 2022) identified in the dataset. Genes include C3 in

VV endothelial cells, inflammatory macrophages, and portal fibroblasts; CD55 in mid zone hepatocytes; EGFR in portal and mid-zone hepatocytes; IGFBP1 and IGFBP2 in inflammatory macrophages; IL6ST in portal hepatocytes; and MMP2 in Kupffer cells.

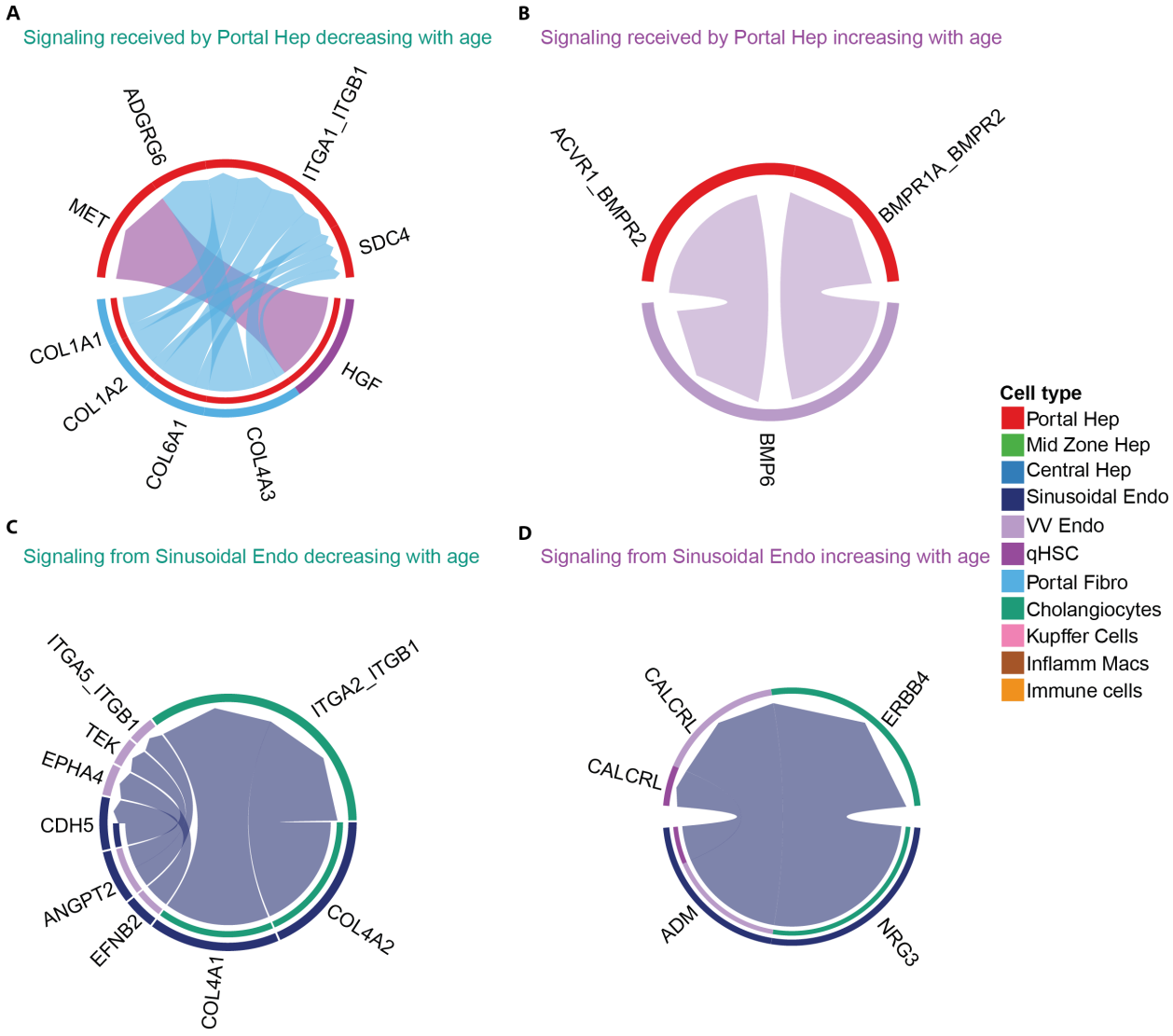


Fig S11. A-D. Chord diagrams representing aging-specific ligand-receptor (L-R) interactions in liver cell types and subtypes. In these diagrams, the outer rings denote different cell types, while the links illustrate interactions, with ligands expressed by the source cell type and receptors expressed by the target cell type. These visualizations highlight changes in intercellular communication networks with age. **A.** L-R signaling to portal hepatocytes decreasing with age. **B**. L-R signaling to portal hepatocytes increasing with age. **C.** L-R signaling originating from sinusoidal endothelial cells decreasing with age. **D**. L-R signaling originating from sinusoidal endothelial cells increasing with age.

References

1. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

2. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

3. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020).

4. S. A. MacParland, J. C. Liu, X.-Z. Ma, B. T. Innes, A. M. Bartczak, B. K. Gage, J. Manuel, N. Khuu, J. Echeverri, I. Linares, R. Gupta, M. L. Cheng, L. Y. Liu, D. Camat, S. W. Chung, R. K. Seliga, Z. Shao, E. Lee, S. Ogawa, M. Ogawa, M. D. Wilson, J. E. Fish, M. Selzner, A. Ghanekar, D. Grant, P. Greig, G. Sapisochin, N. Selzner, N. Winegarden, O. Adeyi, G. Keller, G. D. Bader, I. D. McGilvray, Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 1–21 (2018).

5. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).

6. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

7. B. Phipson, C. B. Sim, E. R. Porrello, A. W. Hewitt, J. Powell, A. Oshlack, Propeller: Testing for differences in cell type proportions in single cell data. *Bioinformatics* **38**, 4720–4726 (2022).

8. A. Ianevski, A. K. Giri, T. Aittokallio, Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat. Commun.* **13**, 1246 (2022).

9. Z. A. Clarke, G. D. Bader, *MALAT1* expression indicates cell quality in single-cell RNA sequencing data, *bioRxiv* (2024). https://doi.org/10.1101/2024.07.14.603469.

10. T. Montserrat-Ayuso, A. Esteve-Codina, Revealing the prevalence of suboptimal cells and organs in reference cell atlases: An imperative for enhanced quality control, *bioRxiv* (2024). https://doi.org/10.1101/2024.04.18.590104.

11. K. B. Halpern, R. Shenhav, O. Matcovitch-Natan, B. Toth, D. Lemze, M. Golan, E. E. Massasa, S. Baydatch, S. Landen, A. E. Moor, A. Brandis, A. Giladi, A. S. Avihail, E. David, I. Amit, S. Itzkovitz, Single-cell spatial reconstruction reveals global division of labour in the

mammalian liver. *Nature* **542**, 352–356 (2017).

12. Z. Xie, A. Bailey, M. V. Kuleshov, D. J. B. Clarke, J. E. Evangelista, S. L. Jenkins, A. Lachmann, M. L. Wojciechowicz, E. Kropiwnicki, K. M. Jagodnik, M. Jeon, A. Ma'ayan, Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).

13. D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

14. S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C.-H. Kuan, P. Myung, M. V. Plikus, Q. Nie, Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

15. D. Saul, R. L. Kosinsky, E. J. Atkinson, M. L. Doolittle, X. Zhang, N. K. LeBrasseur, R. J. Pignolo, P. D. Robbins, L. J. Niedernhofer, Y. Ikeno, D. Jurk, J. F. Passos, L. J. Hickson, A. Xue, D. G. Monroe, T. Tchkonia, J. L. Kirkland, J. N. Farr, S. Khosla, A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nat. Commun.* **13**, 4827 (2022).