Published in final edited form as:
Nature. 2019 June 13; 572(7768): 199–204. doi:10.1038/s41586-019-1373-2.

A Human Liver Cell Atlas reveals Heterogeneity and Epithelial Progenitors

Nadim Aizarani[1,2,3], **Antonio Saviano**[4,#], **Sagar**[1,#], **Laurent Mailly**[4], **Sarah Durand**[4], **Josip S. Herman**[1,2,3], **Patrick Pessaux**[4,5], **Thomas F. Baumert**[4,5,*], **Dominic Grün**[1,6,*]

[1]Max-Planck-Institute of Immunobiology and Epigenetics, D-79108 Freiburg, Germany

[2]Faculty of Biology, University of Freiburg, Freiburg, Germany

[3]International Max Planck Research School for Molecular and Cellular Biology (IMPRS-MCB), Freiburg, Germany

[4]Inserm U1110, Institut de Recherche sur les Maladies Virales et Hepatiques, Université de Strasbourg, F-67000 Strasbourg, France

[5]Pôle Hepato-digestif, Institut Hopitalo-universitaire, Hôpitaux Universitaires, F-67000 Strasbourg, France

[6]CIBBS -Centre for Integrative Biological Signaling Studies, University of Freiburg, Germany

Abstract

The human liver is an essential multifunctional organ, and liver diseases are rising with limited treatment options. However, the cellular composition of the liver remains poorly understood. Here, we performed single-cell RNA-sequencing of ~10,000 cells from normal liver tissue of 9 human donors to construct a human liver cell atlas. Our analysis revealed previously unknown sub-types among endothelial cells, Kupffer cells, and hepatocytes with transcriptome-wide zonation of some of these populations. We reveal heterogeneity of the EPCAM^{+} population, which comprises hepatocyte-biased and cholangiocyte populations as well as a TROP2^{int} progenitor population with strong potential to form bipotent liver organoids. As proof-of-principle, we utilized our atlas to

*Correspondence and requests for materials should be addressed to Dr. Dominic Grün, gruen@ie-freiburg.mpg.de and Prof. Thomas Baumert, MD, Thomas.Baumert@unistra.fr.
#co-second authors

Data availability. Data generated during this study have been deposited in Gene Expression Omnibus (GEO) with the accession code GSE124395. The human liver cell atlas can be interactively explored at http://human-liver-cell-atlas.ie-freiburg.mpg.de/.

Author contributions T.F.B. and D.G. conceived the study. N.A. designed, optimized, and performed cell sorting, scRNA-seq experiments, organoid culture, immunofluorescence, provided validation using the Human Protein Atlas, and performed computational analysis and interpretation of the data. A.S. managed the supply of patient material, isolated single cells from patient tissue, performed animal experiments and immunofluorescence. S. contributed to scRNA-seq analyses and performed single-cell RNA-seq experiments. L.M. performed animal experiments. J.S.H. created the web interface. S.D. isolated single cells from patient tissues. P.P. performed liver resections and provided patient liver tissues. T.F.B. established the liver tissue supply pipeline and supervised the animal experiments. D.G. analyzed and interpreted the data and supervised experiments and analysis of N.A. and S. D.G., N.A., and T.F.B. coordinated and led the study. N.A. and D.G. wrote the manuscript with input from S., A.S., and T.F.B.

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

unravel phenotypic changes in hepatocellular carcinoma cells and in human hepatocytes and liver endothelial cells engrafted into a mouse liver. Our human liver cell atlas provides a powerful resource enabling the discovery of previously unknown cell types in the normal and diseased liver.

Europe PMC Funders Author Manuscripts

The liver is a key organ in the human body. It serves as a central metabolic coordinator with a wide array of essential functions, including regulation of glucose and lipid metabolism, protein synthesis, and bile synthesis. Furthermore, the liver is a visceral organ that is capable of remarkable natural regeneration after tissue loss(1). However, prevalence and mortality of liver disease have risen dramatically within the last decades(2). The liver cellular landscape has barely been explored at single-cell resolution, which limits our molecular understanding of liver function and disease biology. The recent emergence of sensitive single-cell RNA-sequencing (scRNA-seq) methods(3) enables the investigation of cell types in health and disease.

To characterize the human liver at single-cell resolution, we developed a robust pipeline for scRNA-seq of cryopreserved and freshly isolated patient-derived human liver samples and assembled an atlas consisting of 10,372 cells from nine donors. We performed in-depth analysis of all liver cell types with a major focus on epithelial liver cell progenitors.

Single-Cell RNA-seq of the human liver

We applied mCEL-Seq2 (4) for scRNA-seq of non-diseased liver tissue from six patients who underwent liver resections for colorectal cancer metastasis or cholangiocarcinoma without history of chronic liver disease (Fig. 1a, Methods). We sorted and sequenced viable cells in an unbiased fashion as well as specific cell populations on the basis of cell surface markers (Extended Data Fig. 1, Methods). Since fresh liver tissue material is scarce and difficult to preserve, and biobanks with cryopreserved liver samples represent rich resources, we generated scRNA-seq data from cryopreserved cells in addition to single-cell suspensions from freshly prepared liver resection specimens (Methods). We then used RaceID3 for the identification of cell types (Methods) (4),(5).

Cells from different patients, isolated from freshly prepared or cryopreserved single-cell suspensions co-clustered (Extended Data Fig. 1). Furthermore, fresh and cryopreserved cells from the same patient did not reveal pronounced differential gene signatures (Extended Data Fig. 1e-h). However, we observed compositional differences both between fresh and cryopreserved samples derived from the same patient as well as across different fresh (or cryopreserved) samples. We attribute these differences to variability in cell viability and cell type composition across samples.

Since scRNA-seq of randomly sampled populations yielded almost exclusively hepatocytes and immune cells (Extended Data Fig. 1i), we applied additional sorting strategies to enrich for endothelial cells (Extended Data Fig. 1a-c) and $EPCAM^{+}$ cells (see below).

Based on the expression of marker genes, our atlas comprises all the main liver cell types, including hepatocytes, $EPCAM^{+}$ bile duct cells (cholangiocytes), CLEC4G+ liver sinusoidal endothelial cells (LSECs), $CD34^{+}PECAM^{high}$ macro-vascular endothelial cells (MaVECs),

hepatic stellate cells and myofibroblasts, Kupffer cells, and immune cells (Fig. 1b-d and Supplementary Data Table 1). To facilitate interactive exploration of our human liver cell atlas, we created a web interface: http://human-liver-cell-atlas.ie-freiburg.mpg.de/.

Zonation of human liver cell types

Hepatocytes are spatially heterogeneous and zonated along the portal-central axis of the liver lobule(6–8). Based on their metabolic sub-specialization, the liver lobule has been divided into the periportal zone surrounding the portal triad (portal vein, hepatic artery and bile duct), the central zone nearest to the central vein, and the remaining mid zone(6–8). While previous observations suggested sub-specialization of non-parenchymal cells like LSECs and Kupffer cells6, heterogeneity of these cell types has remained elusive, and most studies were carried out in rodents.

We were able to directly compare the signature of MaVECs and LSECs and identified several previously unknown sub-populations (see Extended Data Fig. 2 and Supplementary Note 1).

ScRNA-seq is highly informative on hepatocyte zonation in mouse(9) and a first single-cell analysis of human hepatocyte and endothelial cell zonation at limited resolution was done recently(10). To infer continuous transcriptome-wide zonation, we reasoned that the major axis of variability for a cell type could reflect gene expression changes associated with zonation. Hence, we ordered LSECs and hepatocytes by diffusion pseudo-time (dpt)(11), here interpreted as pseudo-space, along this axis and applied self-organizing maps (SOMs) to infer co-expression modules (Fig. 2, Methods).

We first validated our strategy by recovering previously characterized zonation of mouse hepatocytes(9) (Extended Data Fig. 3a-d). For our human hepatocytes, this approach recovered zonated expression patterns of landmark genes, e.g. *ALB* and *PCK1* (periportal module 1), *CYP1A2* and *CYP2E1* (central/midzonal module 34 and 24, respectively), and *GLUL* (central module 33)(7,9) (Fig. 2a, Extended Data Fig. 3e-g, Supplementary Data Table 2,3). In total, 1,384 out of 3,395 expressed genes (41%) included in the hepatocyte analysis exhibited significant zonation (Benjamini-Hochberg corrected ANOVA $P<0.01$). Pathway enrichment analysis revealed that periportal hepatocyte modules are enriched in genes involved in biological oxidations, consistent with an oxygen gradient peaking in the periportal zone(6–8), and in the glycogen synthesis pathway. (Extended Data Fig. 3h). In accordance with its zonation in mouse hepatocytes, the urea cycle enzyme *CPS1* peaks in periportal hepatocytes (Extended Data Fig. 3g). Midzonal hepatocyte modules are enriched in, e.g., metabolism of xenobiotics by Cytochromes P450. Immunostainings for selected genes validate the predicted zonation on the protein level (Fig. 2a).

LYVE1 and CD14 were previously identified as markers distinguishing midzonal and central LSECs periportal populations(12). Analyzing LSEC zonation, we found that 806 out of 1,198 expressed genes (67%) exhibited significant zonation (Benjamini-Hochberg corrected ANOVA $P<0.01$) (Fig. 2b, Extended Data Fig. 3i, Supplementary Data Table 4, 5). Central and midzonal endothelial cells (modules 1 and 3) exhibited peaked expression of

LYVE1 and *FCN3,* a ficolin protein which can activate the lectin pathway of complement activation. Interestingly, pathway enrichment analysis of the central and midzonal endothelial modules recovered pathways, like binding and uptake of ligands by scavenger receptors, that are shared with midzonal hepatocytes (Extended Data Fig. 3j). Together with a more detailed gene expression analysis (see Supplementary Note 2) this observation suggests the concept of co-zonated genes and functions across hepatocytes and endothelial cells.

Finally, a comparison between mouse(9,13) and human revealed only limited evolutionary conservation of gene expression zonation (see Supplementary Note 3 and (Supplementary Data Tables 6) and (7), reflecting widespread evolutionary changes.

Human liver immune cell populations

A detailed analysis of the *CD163*$^{+}$ *VSIG4*$^{+}$ Kupffer cell compartment revealed sub-populations with distinct gene expression signatures (see Supplementary Note 4 and Extended Data Fig. 4), in agreement with a recent study10. Moreover, we detected shared gene expression and pathways between Kupffer cell subsets and endothelial cells (see Supplementary Note 4 and Extended Data Fig. 4), providing further evidence for functional co-operation of different cell types.

We identified an *MS4A1*$^{+}$*CD37*$^{+}$ B cell subset, which corresponds to circulating B cells upregulating MHC class II components, and a liver resident *MZB1*$^{+}$ B cell subset expressing *DERL3*, *SSR4*, and *IGHG4* (Extended Data Fig. 5).

Finally, we recovered a population of *CD56*$^{+}$ NK cells (cluster 5) and *CD56*$^{-}$ (cluster 3) as well as *CD56*$^{+}$ (cluster 1) *CD8A*$^{+}$ NKT cells, expressing different combinations of chemokine ligands, granzymes, and killer cell lectin-like receptor genes (Extended Data Fig. 6). Clusters 12 and 18 up-regulate a number of heat shock genes. These observations demonstrate an unexpected variety of immune cell subtypes in the human liver.

Putative bipotent epithelial progenitors

Liver regeneration after tissue damage involves the replication of several liver cell types including hepatocytes and cholangiocytes. Furthermore, different types of liver damage lead to specific mechanisms of liver regeneration(14,15). However, the existence of a naïve adult stem cell population in the human liver and its contribution to turnover and regeneration remains controversial. Rare EPCAM^{+} cells have been termed hepatic stem cells(16), which can form dense round colonies when cultured and are bipotent progenitors of hepatoblasts, which differentiate into cholangiocytes or hepatocytes both *in vitro* and *in vivo*(16,17).

In search for genuine liver progenitor cells, we sorted and sequenced single EPCAM^{+} cells from adult human livers. We discovered novel biliary and potential liver progenitor cell surface markers correlating with *EPCAM* or *TROP1* expression; these include *TACSTD2* (*TROP2)*, *FGFR2*, *TM4SF4,* and *CLDN1* and immunohistochemistry confirmed predicted novel markers such as *ANXA4* and the transcriptional co-activator *WWTR1* (Extended Data Fig. 7a).

A focused analysis revealed that the EPCAM+ compartment is transcriptionally heterogeneous and consists of an *ASGR1*$^{+}$ hepatocyte-biased population, *KRT19*^{high}*CFTR*^{high}*ALB*^{low} cholangiocyte populations, and a remaining population of putative naïve progenitor cells (Fig. 3a, Extended Data Fig. 7b,c). The EPCAM^{+} population exhibits only stochastic expression of proliferation markers *MKI67* and *PCNA* and is negative for the hepatoblast marker *AFP* (Extended Data Fig. 7d). Hence, the transcriptional heterogeneity of this population is unlikely to arise as a result of proliferation, and the observed sub-types reside in the normal human liver.

To explore the relatedness of these sub-populations, we reanalyzed the EPCAM^{+} population with RaceID3 and employed StemID2 for lineage reconstruction(4,18) (Fig. 3b, Methods). This analysis revealed that the population in the center of the t-SNE map (clusters 1,2,5,6,7) bifurcates into hepatocyte progenitors and cholangiocytes. To provide further evidence for continuous differentiation trajectories connecting naïve *EPCAM*+ progenitors to cholangiocytes and mature hepatocytes, we performed StemID2 and diffusion map analyses on the combined population of mature hepatocytes and *EPCAM*+ cells (Extended Data Fig. 8a-c).

To better understand the emergence of fate bias towards the two lineages, we applied FateID for the inference of lineage probabilities in each cell(4). Consistently, FateID infers similar probabilities of the central population to differentiate towards hepatocytes and cholangiocytes (Fig. 3c). The fate bias predictions are supported by a differential gene expression analysis revealing up-regulation of common genes comprising several signaling pathway components (*HES1*, *SFRP5*, *FGFR2*, *FGFR3*) in the central population (Fig. 3d), and gradual up-regulation of distinct gene sets towards the hepatocyte-biased and cholangiocyte populations, respectively (Extended Data Fig. 8e). The expression of *TACTSD2* (*TROP2*) anti-correlated with hepatocyte fate bias, exhibiting a gradient that ranges from high expression in mature cholangiocytes to very low expression in the hepatocyte-biased population (Fig. 3e and Extended Data Fig. 7c). Immunostaining of TROP2 in normal human liver tissue showed specific expression in cells of the bile ducts and bile ductules (Fig. 3f). Interestingly, TROP2 expression was reported in amplifying oval cells of injured mouse livers(19).

The central *TROP2*^{int} population is in itself heterogeneous and contains a *MUC6*^{high} population (cluster 7). *MUC6* is highly expressed by pancreatic progenitors and multi-potent bile duct tree stem cells(20), which have been proposed to be the origin of the EPCAM^{+} hepatic stem cells. The *TROP2*^{high} cholangiocyte clusters comprise a *CXCL8*$^{+}$ population (cluster 8) and an *MMP7*$^{+}$ population (clusters 4 and 13) (Extended Data Fig. 7c and Extended Data Fig. 8e,f), while *TROP2*^{low} clusters up-regulate hepatocyte markers such as *ALB*, *HP*, *HNF4A* and *ASGR1* (Extended Data Fig. 7c and 8e,f).

The central *TROP2*^{int} population stratified as bipotent based on the FateID-predicted bias expresses early developmental transcription factors such as *HES1,* which is essential for tubular bile duct formation(21), and *PROX1,* an early specification marker for the developing liver in the mammalian foregut endoderm which is required for hepatocyte proliferation and migration during development(22) (Fig. 3d). Furthermore, we find lower

expression of hepatocytes genes such as *HNF4A*, *HP* and *ALB* and of cholangiocyte markers like *KRT19* and *CFTR* in the population stratified as bipotent compared to the hepatocyte-biased and the mature cholangiocyte populations, respectively (Fig. 3d) and (Extended Data Fig. 7c) and (8f). We speculate that we enriched for the *TROP2*^{int} *KRT19*$^{low/-}$ immature population during cell isolation, since mature bile duct cells require a harsher digestion for their isolation, which can negatively affect other liver cell types. Thus, the actual fraction of *KRT19*^{high} cells in the tissue is presumably higher. We validated the existence of EPCAM^{+}KRT19$^{low/-}$ cells in addition to EPCAM^{+}KRT19$^{high/+}$ cells in situ by immunofluorescence (Fig. 3g and Extended Data Fig. 7e).

Consistent with our scRNA-seq data, flow cytometry profiles of EPCAM and TROP2 displayed a gradient of TROP2 expression in EPCAM^{+} cells, and EPCAM expression correlated with TROP2 expression (Fig. 4a). Moreover, forward and side-scatter profiles of EPCAM^{+} cells indicate that the compartment is heterogeneous and consists of populations with different sizes and morphologies (Fig. 4a). Based on the TROP2 expression distribution, we compartmentalized EPCAM^{+} cells into three compartments: TROP2$^{low/-}$, TROP2^{int}, and TROP2^{high} (Fig. 4a). To validate that the TROP2^{int} population harbors the progenitor population, we attempted to culture bipotent organoids(23) from each compartment. In agreement with our prediction, TROP2^{int} cells exhibited the highest organoid forming capacity, while TROP2$^{low/-}$ cells did not form organoids, and TROP2^{high} cells gave rise to much smaller organoids at strongly reduced frequency compared to TROP2^{int} cells (Fig. 4b). Single-cell culture of TROP2^{int} cells demonstrated the organoid-forming capacity of individual cells from this gate, providing evidence for bipotency on the clonal level (Fig. 4c). ScRNA-seq of the input populations for organoid culture from each compartment expectedly showed a marked enrichment of the respective compartment in the original EPCAM^{+} data (Fig. 4d,e and Extended Fig. 8g,h). Strikingly, flow cytometry profiles of EPCAM and TROP2 for organoid cells grown from the TROP2^{int} compartment recover TROP2$^{low/-}$, TROP2^{int} and TROP2^{high} populations in the organoids (Fig. 4f).

To elucidate the cell type composition of the organoids in depth, we performed scRNA-seq. Co-analysis of the organoid cells and the EPCAM^{+} cells sequenced directly from the patients demonstrates marked transcriptome differences (Fig. 4e). While *EPCAM* and *CD24* were expressed in cells from the organoids and patients, organoid cells downregulated various genes such as *AQP1* or the WNT signaling modulator *SFRP5* and upregulated others such as the proliferation marker *MKI67*$^{+}$, reflected by differential enrichment of the corresponding pathways (Fig. 4g and Extended Data Fig. 8i-k). We observed several sub-populations within the organoids, including a non-dividing hepatocyte-biased *SERPINA1*^{high} population and a non-dividing *KRT19*^{high} cholangiocyte-biased population, consistent with the signature of the EPCAM+ cells recovered from the patients (Fig. 4e). This further supports the claim that the TROP2^{int} compartment harbors a bipotent progenitor population, which can give rise to hepatocyte and cholangiocyte populations.

In contrast to patients' cells, organoid cells strongly downregulate *ALB* but express *AGR2* and other mucin family genes like *MUC5AC* and *MUC5B*, normally expressed, e.g., in intestinal cells and gastrointestinal cancers(24,25)(Fig. 4g and Extended Data Fig. 8j). These observations reflect that organoid cells express genes that are expressed in other systems,

acquire a more proliferative state, and appear to upregulate stem cell-related pathways like WNT signaling.

In light of (1) these functional validation experiments, (2) the observed gene signature of the *TROP2*^{int} cells, and (3) the *in situ* location, our data strongly suggest that the putative liver progenitor population can be defined as a subpopulation of bile duct cells.

Perturbed cell states in liver cancer

Hepatocellular carcinoma is the most common type of primary liver cancer(26). In order to demonstrate the value of our atlas as a reference for comparisons with diseased liver cells, we sequenced CD45^{+} and CD45^{-} cells from HCC tissue of three different patients (Extended Data Fig. 9a,b, Methods).

We recovered several cell types from the tumors, including cancer cells, endothelial cells, Kupffer cells, NKT and NK cells (Fig. 5a and Extended Data Fig. 9c) and compared them to the normal liver cell atlas. Differential gene expression analysis and immunohistochemistry revealed that cancer cells lose the expression of cytochrome P450 genes like CYP2E1 and CYP2C8 and the periportally zonated gene CPS1 (Fig. 5b and Extended Data Fig. 9d,e) as well as the metabolic signature of normal hepatocytes (Fig. 5e). They up-regulate *AKR1B10*, a known biomarker of HCC with potential involvement in hepatocellular carcinogenesis(27) (Extended Data Fig. 9d). Moreover, immunohistochemistry confirmed that IL32, a pro-inflammatory TNF-alpha inducing cytokine, is highly upregulated in cancer cells (Fig. 5b). Overall, cancer cells upregulate WNT and Hedgehog signaling pathways, highlighting similarities between *EPCAM*$^{+}$ normal liver progenitors and the observed cancer cell population (Fig. 5c).

Endothelial cells from the tumor upregulate, e.g., extracellular matrix organization genes such as *COL4A2* and *SPARC* (Fig. 5d, Extended Data Fig. 9f). Strikingly, they do not express LSEC markers like *CLEC4G* but upregulate MaVEC markers like *PECAM1*, *AQP1* and *CD34* (Fig. 5e and Extended Data Fig. 9f,g). Moreover, HCC LSECs upregulate PLVAP which makes them less permeable and could potentially restrict the access of lymphocytes and soluble antigens(28) to the tumor (see Supplementary Note 5 and Extended Data Fig. 9f,g).

We conclude that the comparison of scRNA-seq data between the cell populations of HCC and the liver cell atlas allows the inference of perturbed gene expression signatures, biomarkers and modulated functions across cell types.

A human liver chimeric mouse model

Patient derived xenograft mouse liver models are a powerful tool for studying human liver cells and diseases *in vivo*(29). To correctly interpret such experiments, it is crucial to understand the differences between cells directly taken from the human liver and transplanted human cells in the mouse chimeric liver.

To address this issue, we transplanted human liver cells from patient-derived hepatocyte and non-parenchymal cell fractions into an FRG-NOD (*Fah*$^{-/-}$*Rag2*$^{-/-}$*Il2rg*$^{-/-}$) mouse liver model(30) (HMouse) and sorted single human cells in an unbiased fashion and on the basis of hepatocyte and endothelial cell markers post engraftment for scRNA-seq (Fig. 6a and Extended Data Fig. 10a). We then compared these engrafted cells to our reference atlas and observed that we had successfully transplanted both human hepatocytes and endothelial cells (Fig. 6b and Extended Data Fig. 10b,c), which maintained their fundamental gene signatures, such as *ALB* or *PCK1* and *CLEC4G*, *PECAM1* or *CD34*, respectively (Extended Data Fig. 10b-f). Nevertheless, many genes were differentially expressed compared to human liver cells, e.g. *AKR1B10*, which was also expressed by cancer cells from HCC (Fig. 6c and Extended Data Fig. 10g). Gene Set Enrichment Analysis (GSEA) of differentially expressed genes revealed that HMouse hepatocytes and endothelial cells downregulated pathways, such as hemostasis, and upregulated WNT and Hedgehog signaling as well as cell cycle genes (Fig. 6d) akin to what we observed in the HCC cancer cells and cells from the liver organoids.

Discussion

We here established a human liver cell atlas, revealing heterogeneity within major liver cell populations and the existence of an epithelial progenitor in the adult human liver.

Our atlas reveals transcriptome-wide zonation of hepatocytes and endothelial cells, and suggests that different liver cell types may cooperate in order to carry out essential functions. Although we could validate predicted zonation profiles with antibody staining, it will be essential to perform more large scale in situ gene expression analysis. The novel EPCAM $^{+}$TROP2^{int} population provides a strong candidate with potential involvement in homeostatic turnover, liver regeneration, disease pathogenesis, and tumor formation. We point out that, while our *in silico* analysis and the *in vitro* organoid culture experiments provide evidence for bipotency of this population, its lineage potential remains to be demonstrated *in vivo*.

As demonstrated by our HCC analysis, the atlas provides a key reference to investigate liver diseases and will contribute to advance urgently needed human liver models, including organoids as well as humanized liver chimeric mouse models.

Methods

Human liver samples

Human liver tissue samples were obtained from patients who had undergone liver resections between 2015 and 2018 at the Center for Digestive and Liver Disease (Pôle Hépato-digestif) of the Strasbourg University Hospitals University of Strasbourg, France. For the human liver cell atlas, samples were acquired from patients without chronic liver disease (defined as liver damage lasting over a period of at least six months), genetic hemochromatosis with homozygote C282Y mutation, active alcohol consumption (> 20 g/d in women and > 30 g/d in men), active infectious disease, pregnancy or any contraindication for liver resection. All patients provided a written informed consent. The protocols followed the ethical principles of the declaration of Helsinki and were approved by the local Ethics Committee of the

University of Strasbourg Hospitals and by the French Ministry of Education and Research (Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche; approval number DC-2016-2616). Data protection was performed according to EU legislation regarding privacy and confidentiality during personal data collection and processing (Directive 95/46/EC of the European Parliament and of the Council of the 24 October 1995). Samples (BP1) and tissue blocks were obtained from Biopredic International.

Tissue dissociation and preparation of single cell suspensions

Human liver specimens obtained from resections were perfused for 15 minutes with calcium-free 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid buffer containing 0.5 mM ethylene glycol tetraacetic acid (Fluka) followed by perfusion with 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid containing 0.5 mg/mL collagenase (Sigma-Aldrich) and 0.075% $CaCl_2$ at 37°C for 15 min as previously described(32). Then the cells were washed with phosphate-buffered saline (PBS) and nonviable cells were removed by Percoll (Sigma-Aldrich) gradient centrifugation. Part of the isolated cells was further separated into primary human hepatocytes (PHH) and non-parenchymal cells (NPCs) by an additional centrifugation step. The isolated cells were frozen in liquid nitrogen using the CryoStor® CS10 solution (Sigma-Aldrich).

Transplantation of human cells into *Fah*$^{-/-}$/*Rag2*$^{-/-}$/*Il2rg* $^{-/-}$ mice

Fah$^{-/-}$/*Rag2*$^{-/-}$/*Il2rg* $^{-/-}$ *(FRG)* breeding mice were kept at the Inserm Unit 1110 SPF animal facility and maintained with 16 mg/L of 2-(2-nitro-4-trifluoro-methyl-benzoyl)-1,3 cyclohexanedione (NTBC; Swedish Orphan Biovitrum) in drinking water. Six-week old mice were intravenously injected with 1.5 x 10^9 pfu of an adenoviral vector encoding the secreted form of the human urokinase-like plasminogen activator (Ad-uPA)(33). Forty-eight hours later, 10^6 PHH and 2 x 10^5 NPCs from the same liver donor and isolated as previously described, were injected intra-splenically via a 27-gauge needle. For the procedure, the mice were kept under gaseous isoflurane anesthesia and received a subcutaneous injection of buprenorphine at the dose of 0.1 mg/kg. After the transplantation the NTBC was gradually decreased and completely withdrawn in 7 days. The success of the transplantation was evaluated 2 months after the procedure by dosing human albumin in mouse serum as previously described(34). All procedures are consistent with the guidelines set by the Panel on Euthanasia (AVMA) and the NIH Guide for the Care and Use of Laboratory Animals as well as the Declaration of Helsinki in its latest version, and to the Convention of the Council of Europe on Human Rights and Biomedicine. The animal research was performed within the regulations and conventions protecting the animals used for research purposes (Directive 86/609/EEC), as well as with European and national laws regarding work with genetically modified organs. The animal facility at the University of Strasbourg, Inserm U1110 has been approved by the regional government (Préfecture) and granted the authorization number N° D67-482-7, 2012/08/22.

FACS

Liver cells were sorted from mixed, hepatocyte, and non-parenchymal cell fractions on an Aria Fusion I using a 100 μm nozzle. Cells from the HCC samples were not fractionated and

were sorted directly after tissue digestion. Zombie Green (Biolegend) was used as a viability dye. Cells were stained with human specific antibodies against CD45 (Biolegend), PECAM1 (Biolegend), CD34 (Biolegend), CLEC4G (R&D systems), ASGR1 (BD Biosciences), EPCAM (R&D systems), and TROP2 (Biolegend). Organoids were stained with antibodies against EPCAM and TROP2. For the humanized mouse samples, cells were stained either with antibodies against ASGR1 and PECAM1 or human HLA-ABC (BD Biosciences) and mouse H2-Kb (BD Biosciences). Viable cells were sorted in an unbiased fashion or from specific populations based on the expression of markers into the wells of 384 well plates containing lysis buffer.

Single-cell RNA amplification and library preparation

Single cell RNA sequencing was performed according to the mCEL-Seq2 protocol(4,35). Viable liver cells were sorted into 384-well plates containing 240 nL of primer mix and 1.2 μL of PCR encapsulation barrier, Vapor-Lock (QIAGEN) or mineral oil (Sigma-Aldrich). Sorted plates were centrifuged at 2200 g for a few minutes at 4°C, snap-frozen in liquid nitrogen and stored at −80°C until processed. 160nL of reverse transcription reaction mix and 2.2 μL of second strand reaction mix were used to convert RNA into cDNA. cDNA from 96 cells were pooled together before clean up and in vitro transcription, generating 4 libraries from one 384-well plate. 0.8 μL of AMPure/RNAClean XP beads (Beckman Coulter) per 1 μL of sample were used during all the purification steps including the library cleanup. Other steps were performed as described in the protocol. Libraries were sequenced on an Illumina HiSeq 2500 and 3000 sequencing system (pair-end multiplexing run, high output mode) at a depth of ~150,000-200,000 reads per cell.

Quantification of Transcript Abundance

Paired end reads were aligned to the transcriptome using bwa (version 0.6.2-r126) with default parameters(36). The transcriptome contained all gene models based on the human whole genome ENCODE V24 release. All isoforms of the same gene were merged to a single gene locus. The right mate of each read pair was mapped to the ensemble of all gene loci and to the set of 92 ERCC spike-ins in the sense direction. Reads mapping to multiple loci were discarded. The left read contains the barcode information: the first six bases corresponded to the unique molecular identifier (UMI) followed by six bases representing the cell specific barcode. The remainder of the left read contains a polyT stretch. The left read was not used for quantification. For each cell barcode, the number of UMIs per transcript was counted and aggregated across all transcripts derived from the same gene locus. Based on binomial statistics, the number of observed UMIs was converted into transcript counts(37).

Single-Cell RNA Sequencing Data Analysis

10,372 cells passed quality control threshold of >1,000 transcripts (Poisson-corrected UMIs37) for the normal human liver cell atlas. For cells from the organoids, 1052 cells passed the quality control thresholds. For cells from HCC, 1282 cells passed the quality control threshold of >1,000 transcripts. For cells from the humanized mouse, 311 cells passed the quality control threshold of >1,000 transcripts. All the datasets were analyzed using RaceID3(4). For normalization, the total transcript counts in each cell were normalized

to 1 and multiplied by the minimum total transcript count across all cells passing the quality control threshold (> 1,000 transcripts per cell). Prior to normalization, cells expressing >2% of *KCNQ1OT1* transcripts, a previously identified marker of low quality cells(18) were removed from the analysis. Moreover, transcripts correlating to *KCNQ1OT1* with a Pearson's correlation coefficient >0.4 were also removed. RaceID3 was run with the following parameters: mintotal=1000, minexpr=2, minnumber=10, outminc=2, cln=15 and default parameters otherwise. The t-distributed stochastic neighbor embedding (t-SNE) algorithm was used for dimensional reduction and cell cluster visualization.

Diffusion Pseudo-Time Analysis and Self-Organizing Maps

Diffusion pseudotime (dpt) analysis(11) was implemented and diffusion maps generated using the destiny R package. The number of nearest neighbors, k, was set to 100. Self-organizing maps (SOMs) were generated using the FateID package based on the ordering computed by dpt as input. Only genes with >2 counts after size normalization in at least a single cell were included for the SOM analysis. Briefly, smooth zonation profiles were derived by applying local regression on normalized transcript counts after ordering cells by dpt. Next, a one-dimensional SOM with 200 nodes was computed on these profiles after z-transformation. Neighboring nodes were merged if the Pearson's correlation coefficient of the average profiles of these nodes exceed 0.85. The remaining aggregated nodes represent the gene modules shown in the SOM figures.

P-values for the significance of zonation were derived by binning dpt-ordered profiles into three equally-sized bins to perform ANOVA. The resulting p-values were multiple-testing corrected with the Benjamini-Hochberg method. Increasing the number of bins produced similar results.

Conservation of zonation between human and mouse

Expression data from Halpern et al.(9) (GEO accession code GSE84498) were used for analyzing evolutionary conservation of hepatocyte zonation between human and mouse. The transcript count data were analyzed with RaceID3 to determine cell types, with parameter mintotal=1,000 and cln=6. A subgroup of clusters was identified as hepatocytes based on marker gene expression and used for pseudo-time (dpt) and sub-sequent SOM analysis as was done for the human data. To obtain a similar number of genes, only genes with at least 1.5 counts after size normalization in at least a single cell were included. To identify orthologs between human and mouse for the references used in this study and by Halpern et al. as provided by the authors, we first identified pairs of orthologs based on identical gene identifiers upon capitalization of all letters. We further computed mutual blastn (run with default) best hits. The final list comprises 16,670 pairs of orthologs.

Conservation of zonation was assessed based on the Pearson's correlation of zonated expression profiles after binning the human data into nine equally-sized bins akin to the nine zones derived in Halpern et al. Conservation of zonation of endothelial cells was evaluated based on published mouse data from Halpern et al.(13) utilizing the classification into four spatially stratified population. To calculate Pearson's correlation coefficient between human

Europe PMC Funders Author Manuscripts

and mouse endothelial cells, a dpt analysis was performed for all human cells mapping to endothelial cell clusters and these profiles were discretized into four equally-sized bins.

Lineage Analysis of the *EPCAM*$^{+}$ compartment

For a separate analysis of the *EPCAM*$^{+}$ population, all cells from clusters 4, 7, 24 and 39 were extracted and reanalyzed with RaceID3(4) using the parameters mintotal=1000 and minexpr=2, minnumber=10 outminc=2, and default parameters otherwise. StemID(24) was run on these clusters with cthr=10 and nmode=TRUE and knn=3. FateID(4) was run on the filtered and feature-selected expression matrix from RaceID3, with target clusters inferred by FateID using *ASGR1* plus *ALB* and *CXCL8* plus *MMP7* as markers for hepatocyte and cholangiocyte lineage target clusters. Using the *KRT19* and *CFTR* as mature cholangiocyte markers yields highly similar results.

Differential Gene Expression Analysis

Differential gene expression analysis between cells and clusters was performed using the diffexpnb function from the RaceID package. First, negative binomial distributions reflecting the gene expression variability within each subgroup were inferred based on the background model for the expected transcript count variability computed by RaceID3. Using these distributions, a p-value for the observed difference in transcript counts between the two subgroups was calculated and multiple testing corrected by the Benjamini-Hochberg method using the same strategy as described in Anders et al(38).

Pathway Enrichment Analysis and Gene Set Enrichment Analysis

Symbol gene IDs were first converted to Entrez gene IDs using the clusterProfiler(39) package. Pathway enrichment analysis and GSEA(40,41) were implemented using the ReactomePA(42) package. Pathway enrichment analysis was done on genes taken from the different modules in the SOMs. GSEA was done using the differentially expressed genes inferred by the diffexpnb function from the RaceID package.

Validation of protein expression using the Human Protein Atlas

Immunostaining images were collected from the Human Protein Atlas(31) (https://www.proteinatlas.org).

Immunofluorescence

Human liver tissue was fixed overnight in 3.7% formaldehyde (Figure 3j) or cryosectioned and fixed in 2.5% paraformaldehyde for 20 minutes (Extended Data Fig. 7e). The tissue was embedded in OCT and stored at -80°C. The tissue was cryosectioned into 7 micron sections. The tissue was washed twice for 5 min in 0.025% Triton 1x PBS. The tissue was then blocked in 10% FBS with 1% BSA in 1x PBS for 2 hours at room temperature. The dilution used for the anti-human KRT19 (HPA002485, Sigma) and EPCAM (SAB4200704, Sigma) was 1:100 in 100 μL of 1xPBS with 1% BSA. The antibody was incubated overnight at 4°C in the dark. The tissue was washed twice with 0.025% Triton 1x PBS and then incubated with secondary antibodies (donkey anti-rabbit IgG-AF488 (A21206, Thermo Fisher Scientific) and goat anti-mouse IgG-AF568 (A11019) (Fig. 3g) or sheep anti-mouse IgG-

AF488 (515-545-062, Jackson ImmunoResearch) and donkey anti-rabbit IgG-RRX (711-295-152, Jackson ImmunoResearch) (Extended Data Fig. 7e) at a 1:200 dilution in 1x PBS with 1% BSA for 1 hour at room temperature. The tissue was then washed twice with 0.025% Triton 1x PBS. DAPI Fluoromount-G (Southern Biotech) was added to the tissue and a coverslip placed on top. Imaging was done using the Zeiss confocal microscope LSM780 (Fig. 3g) or ZEISS Axio Vert.A1. Images were taken at 63x magnification.

Organoid culturing

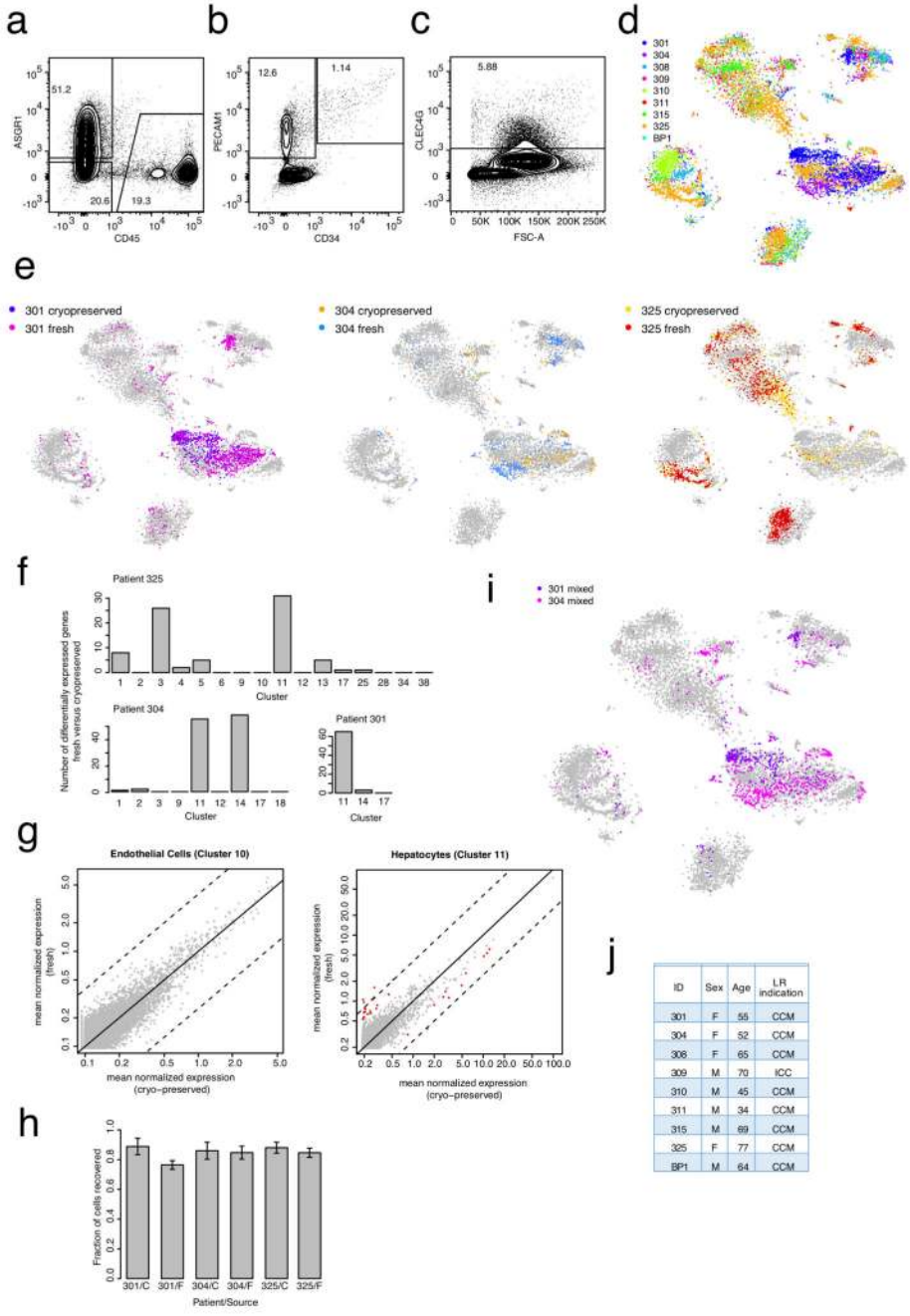
Organoid culturing was done as previously described(43). The cell populations from the $EPCAM^{+}$ compartment were sorted on an Aria Fusion 1 using a 100 μm nozzle into tubes containing culture medium supplemented with 10 μM ROCK inhibitor (Y-27632) (Sigma-Aldrich). After sorting, cells were centrifuged in order to remove the medium and then resuspended in 25 μL of Matrigel. Droplets of the Matrigel solution containing the cells were added to the wells of a 24 well suspension plate and incubated for 5-10 min at 37°C until the Matrigel solidified. Droplets were overlaid with 250 μL of liver isolation medium and then incubated at 37°C, 5% CO_2. After 3-4 days, the liver isolation medium was replaced with liver expansion medium. For the single cell culture, from each patient, single cells from the $TROP2^{int}$ gate were sorted into the wells of a non-tissue culture treated 96 well plate containing medium with 5% Matrigel. Organoids were passaged 14 days after isolation and then passaged multiple times 5-7 days after splitting. For FACS, single cell suspensions were prepared from the organoids by mechanical dissociation followed by TrypLE (Life Technologies) digestion as previously described(43). Organoid cells were sequenced 5 days after splitting and 17 days after initially sorting the cells for the culture.

Step-by-Step Protocol

A detailed protocol for scRNA-seq of cryopreserved human liver cells is available at Protocol Exchange(44).

Extended Data

Europe PMC Funders Author Manuscripts



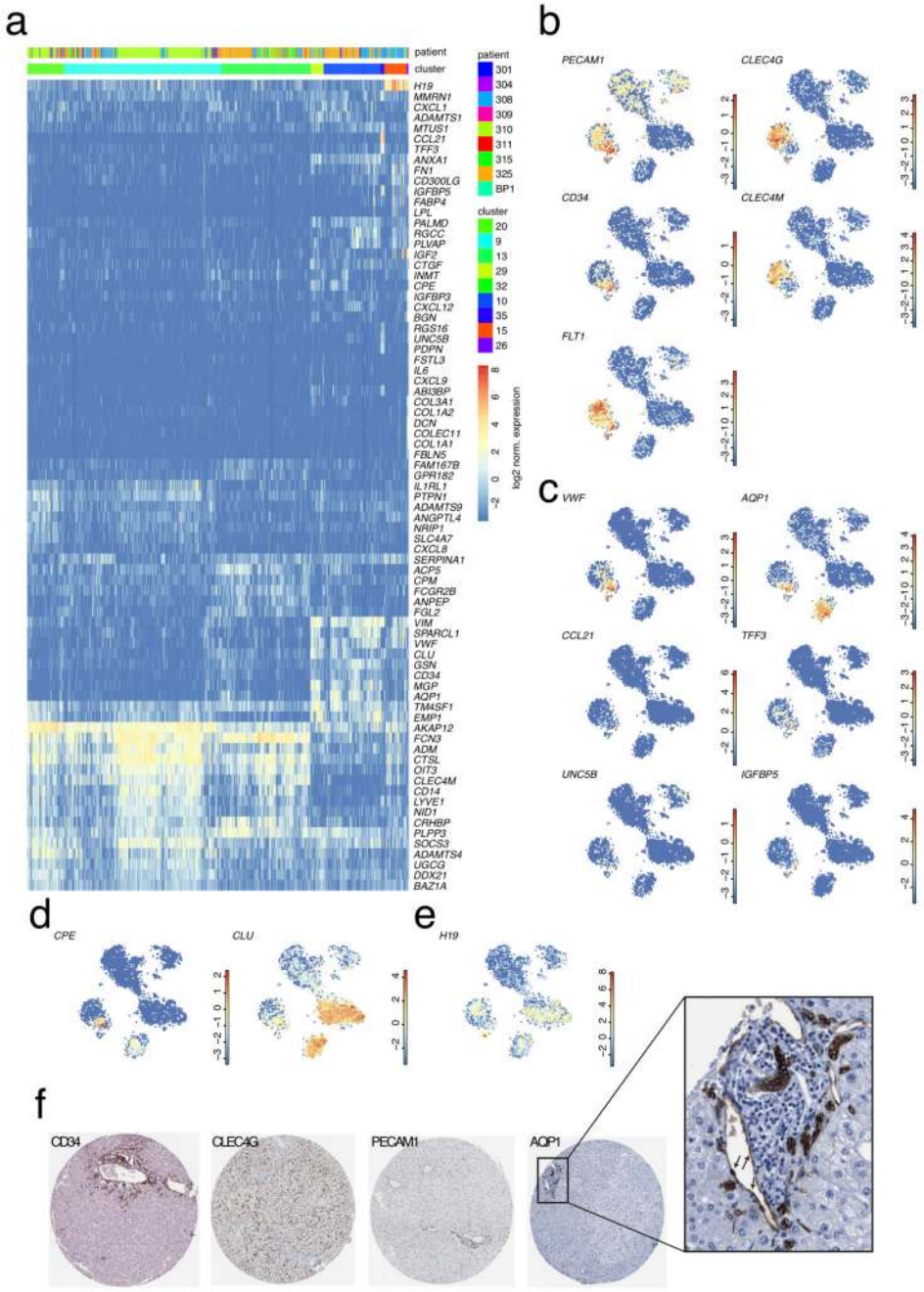
Extended Data Figure 1. ScRNA-seq analysis of normal liver resection specimens from nine adult patients.

a, FACS plot for CD45 and ASGR1 staining from a mixed fraction (hepatocyte and non-parenchymal cells). **b**, FACS plot for PECAM1 and CD34 staining from a mixed fraction. **c**, FACS plot for CLEC4G staining from a mixed fraction. (a-c) n=6 independent experiments. **d**, t-SNE map showing the IDs of the 9 patients from which the cells were sequenced. Cells were sequenced from freshly prepared single-cell suspensions for patients 301, 304, 325, and BP1, and from cryopreserved single-cell suspensions for patients 301, 304, 308, 309,

Europe PMC Funders Author Manuscripts

310, 311, 315, and 325. Cells were sorted and sequenced mainly in an unbiased fashion from non-parenchymal cell, hepatocyte and mixed fractions of patients 301 and 304. Non-parenchymal and mixed fractions were used to sort specific populations on the basis of markers. $CD45^-$ and $CD45^+$ positive cells were sorted from all patients. CLEC4G+ LSECs were sorted by FACS from patients 308, 310, 315, and 325. $EPCAM^+$ cells were sorted by FACS from patients 308, 309, 310, 311, 315, and 325. **e**, t-SNE map highlighting data for fresh and cryopreserved cells from patients 301, 304 and 325. Although minor shifts of frequencies within cell populations are visible, transcriptomes of fresh and cryopreserved cells co-cluster. Differential gene expression analysis of fresh versus cryopreserved cells, e.g. for endothelial cells of patient 325 in cluster 10 (f), did not reveal any differentially expressed genes. (d,e) n=10,372 cells. **f**, Barplot showing the number of differentially expressed genes (Benjamini-Hochberg corrected $P<0.01$; Methods) between fresh and cryopreserved cells within each cluster for patient 325 (upper panel; n=2,248 cells) and patients 304 (n=959 cells) and 301 (n=1,329 cells) (lower panel). Only clusters with >5 cells from fresh and cryopreserved samples were included for the computation. **g**, Scatter plot of mean normalized expression across fresh and cryopreserved cells from patient 325 in endothelial cells of cluster 10 (no differentially expressed genes, left) (n= 101 cells) and cluster 11 (maximal number of differentially expressed genes across all clusters, right) (n=272 cells). Red dots indicate differentially expressed genes (Benjamini-Hochberg corrected $P<0.01$; Methods). Diagonal (solid black line) and log2 fold changes of four (broken black lines) are indicated. Almost all differentially expressed genes for cluster 11 exhibit log2-foldchanges < 4. **h**, Barplot showing the fraction of sorted cells which passed quality filtering (see Methods) after scRNA-seq. Error bars are derived from the sampling error assuming binomial counting statistics. F, fresh samples; C, cryopreserved samples. **i**, t-SNE map highlighting cells sequenced from mixed plates representing unbiased samples for patient 301 and 304. Without any enrichment strategy, hepatocytes and immune cells strongly dominate and endothelial cells as well as $EPCAM^+$ cells are rarely sequenced. **j**, Table of patient information. CCM: colon cancer metastasis; ICC: intrahepatic cholangiocarcinoma; LR: liver resection.

Europe PMC Funders Author Manuscripts

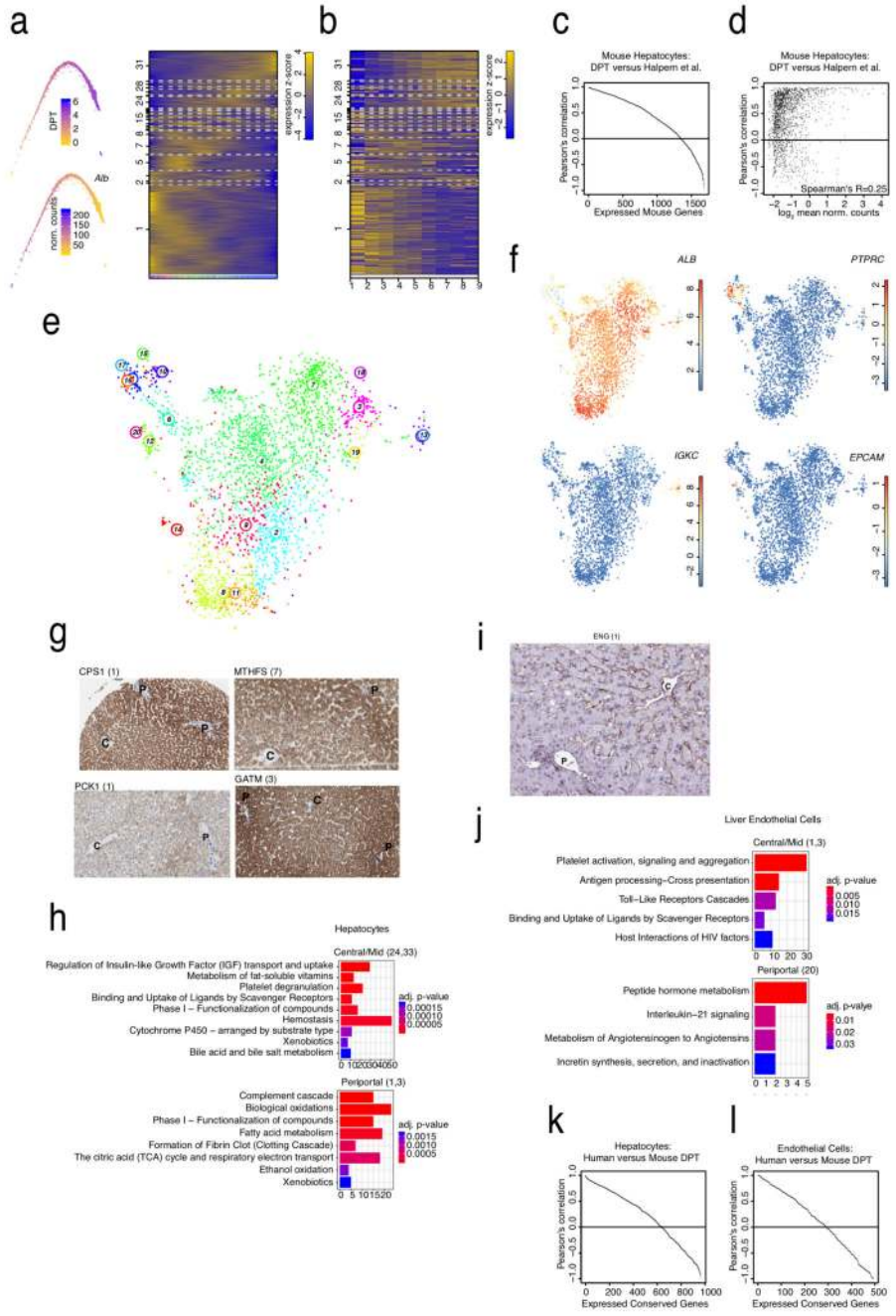


Extended Data Figure 2. The endothelial cell compartment is a heterogeneous mixture of sub-populations.

a, Expression heatmap of genes up-regulated in endothelial cell clusters (Benjamini-Hochberg corrected $P<0.01$; n=1,830 cells; Methods). For each cluster the top ten up-regulated genes were extracted and expression of the joint set is shown in the heatmap across all endothelial cell clusters. Genes were ordered by hierarchical clustering. **b**, Expression t-SNE maps for LSEC and MaVEC markers *PECAM1*, *CLEC4G*, *CD34*, *CLEC4M* and *FLT1*. **c**, Expression t-SNE maps for *VWF*, *AQP1*, *CCL21*, *TFF3* and *UNC5B* and *IGFBP5*.

d, Expression t-SNE maps for *CPE* and *CLU*. **e**, Expression t-SNE map for *H19*. The color bar in (b-e) indicates log2 normalized expression. (b-e) n=10,372 cells. **f**, Immunostaining of CD34, AQP1, CLEC4G and PECAM1 in normal liver tissue from the Human Protein Atlas. The portal area for the AQP1 was enlarged to show positive staining of both bile duct cells and portal MaVECs (black arrows).

Europe PMC Funders Author Manuscripts

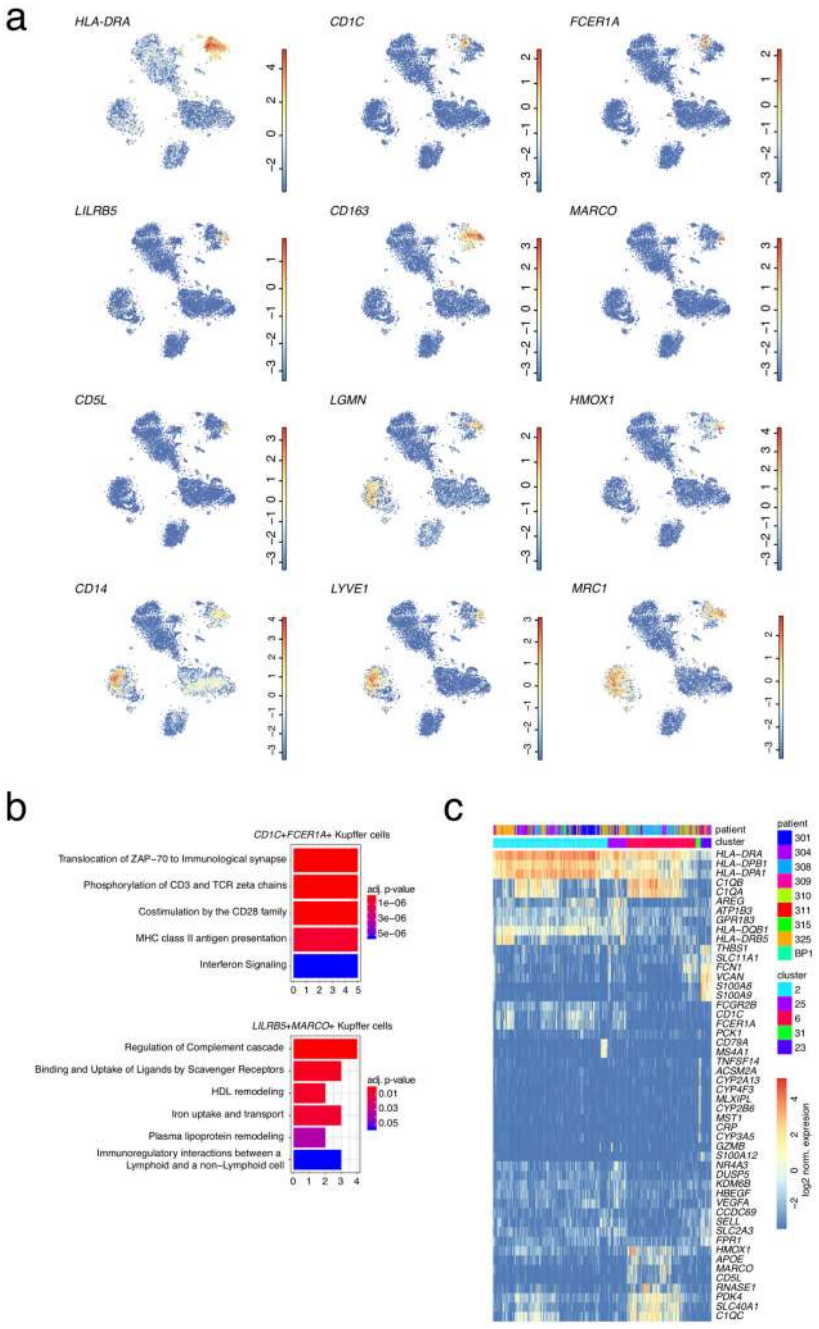


Extended Data Figure 3. Evolutionary conservation of zonation profiles.

a, Diffusion maps highlighting inferred differentiation-pseudotime (dpt) and *Alb* expression (left), and a self-organizing map for mouse hepatocyte single-cell RNA-seq data(9) (Methods). Compare to Figure 2 for details. **b**, Heatmap showing the spatial hepatocyte zonation profiles (nine zones) inferred by Halpern et al. using the same ordering of genes as in (a). **c**, Pearson's correlation coefficient of zonation profiles inferred by Halpern et al. and our dpt approach after discretizing dpt-inferred zonation profiles into 9 equally-sized bins. 1,347 out of 1,684 genes (80%) above the expression cutoff exhibit a positive correlation

Europe PMC Funders Author Manuscripts

between the two methods. **d**, Pearson's correlation coefficient as a function of average normalized expression. Negative correlations are enriched at low expression, and Pearson's correlation of zonation profiles positively correlates with expression (Spearman's R=0.25; n=1,684 genes). **e**, t-SNE map of single-cell transcriptomes highlighting the clusters generated by RaceID3, run separately on hepatocytes (cluster 11, 14, and 17 in Fig. 1c). The map reveals a major group of hepatocyte clusters and a number of small cluster co-expressing T cell related genes, B cell related genes, or progenitor genes. **f**, t-SNE maps highlighting the expression of *ALB*, the immune cell marker *PTPRC*, the B cell marker *IGKC*, and the progenitor marker *EPCAM*. The color bar indicates log2 normalized expression. Co-expression of hepatocyte and immune cell markers could either indicate the presence of doublets or could be due to spill-over of highly expressed genes such as *ALB* during library preparation between cells. For the zonation analysis (Figure 2) only cells in clusters 3, 7, 19, 4, 2, 9, 8, and 11 from the map in (e) were included. (e,f) n=3,040 cells. **g**, Immunostaining of periportal genes CPS1, PCK1, MTHFS, and GATM from the Human Protein Atlas(31) are shown. The zonation module containing each gene in the self-organizing map (Fig. 2a) is indicated in parentheses. The portal tracts and central veins in the immunostainings are denoted as "P" and "C", respectively. **h**, Pathways enriched for the genes in hepatocyte central/mid modules 24 and 33 (top; n=659 genes) and periportal modules 1 and 3 (bottom; n=422 genes) are shown (cf. Fig. 2a). **i**, Immunostaining of central gene ENG from the Human Protein Atlas31 are shown. The zonation module in the self-organizing map (Fig. 2b) is indicated in parentheses. The portal tracts and central veins in the immunostainings are denoted as "P" and "C", respectively. **j**, Pathways enriched for the genes in endothelial central/mid modules 1 and 3 (top; n=422 genes) and periportal module 20 (bottom; n=73 genes) are shown (cf. Fig. 2b). (h,j) P-values in the pathway enrichment analysis were calculated based on a hypergeometric model and adjusted using the Benjamini-Hochberg method (Methods). **k**, Pearson's correlation coefficient of hepatocyte zonation profiles of orthologues pairs of human and mouse genes. Mouse data are from Halpern et al.(9) (n=967 genes) **l**, Pearson's correlation coefficient of endothelial cell (including MVECs and LSECs) zonation profiles of orthologues pairs of human and mouse genes (n=977 genes). Mouse data are from Halpern et al. (13) See Methods for details.



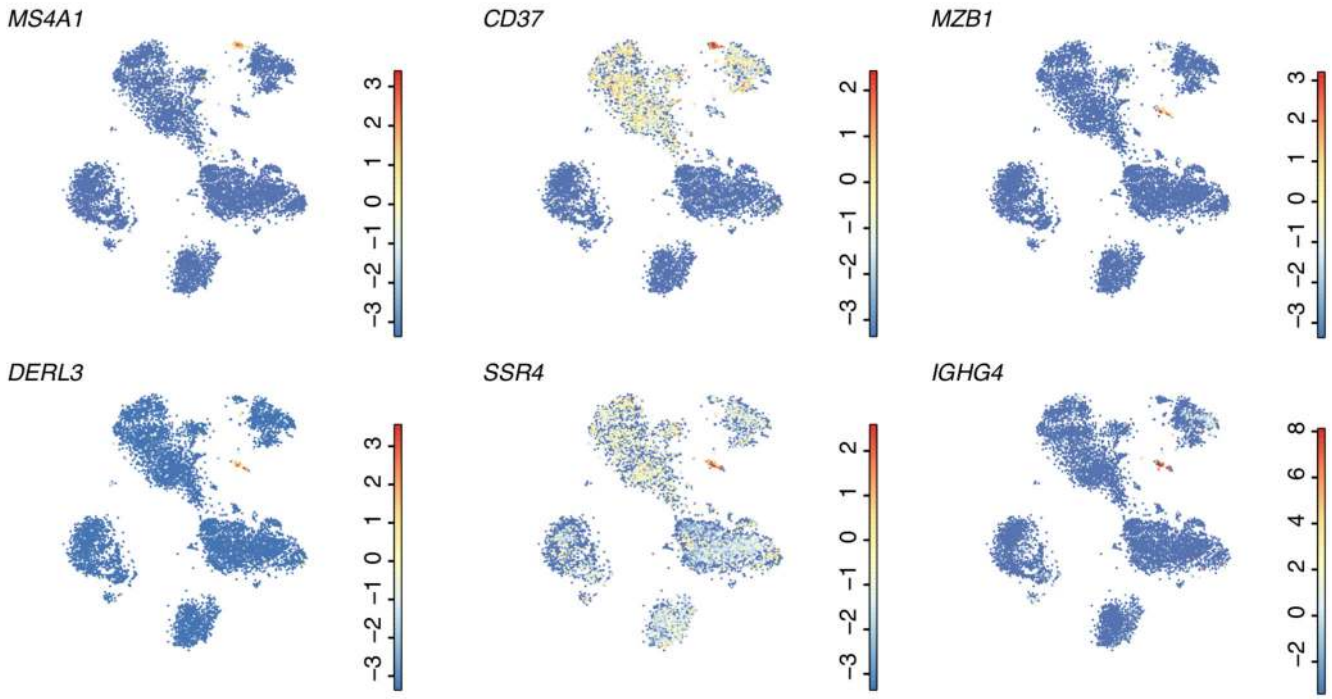
Extended Data Figure 4. The human liver contains different Kupffer cell populations.

a, Expression t-SNE maps of the markers for the Kupffer cell subtypes. The color bar indicates log2 normalized expression (n=10,372 cells). **b**, Major pathways up-regulated in the *CD1C*$^{+}$ antigen presenting (n=12 genes) and *LILRB5*$^{+}$ metabolic/immunoregulatory (n= 35 genes) Kupffer cell subsets as revealed by Reactome pathway analysis. The number of genes in each pathway is shown on the x-axis. P-values are calculated based on a hypergeometric model and adjusted using the Benjamini-Hochberg method. **c**, Expression heatmap of genes up-regulated in Kupffer cell clusters (Benjamini-Hochberg corrected

Europe PMC Funders Author Manuscripts

P<0.01, Methods). For each cluster the top ten up-regulated genes were extracted and expression of the joint set is shown in the heatmap across all Kupffer cell clusters. Genes were ordered by hierarchical clustering.

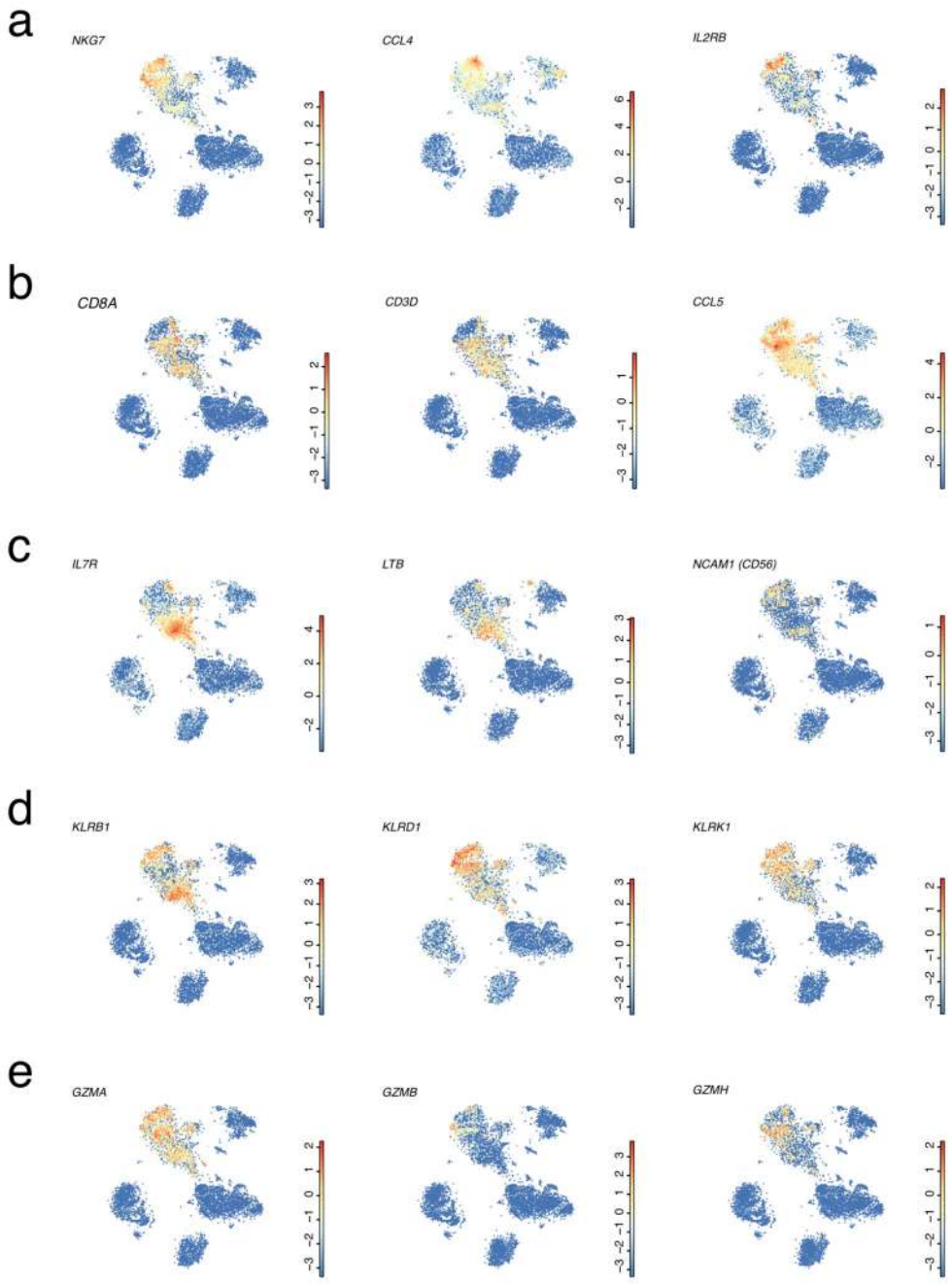
Europe PMC Funders Author Manuscripts



Extended Data Figure 5. The human liver contains different B cell populations.
Expression t-SNE maps of the markers for the B cell subtypes. The color bar indicates log2 normalized expression (n=10,372 cells).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

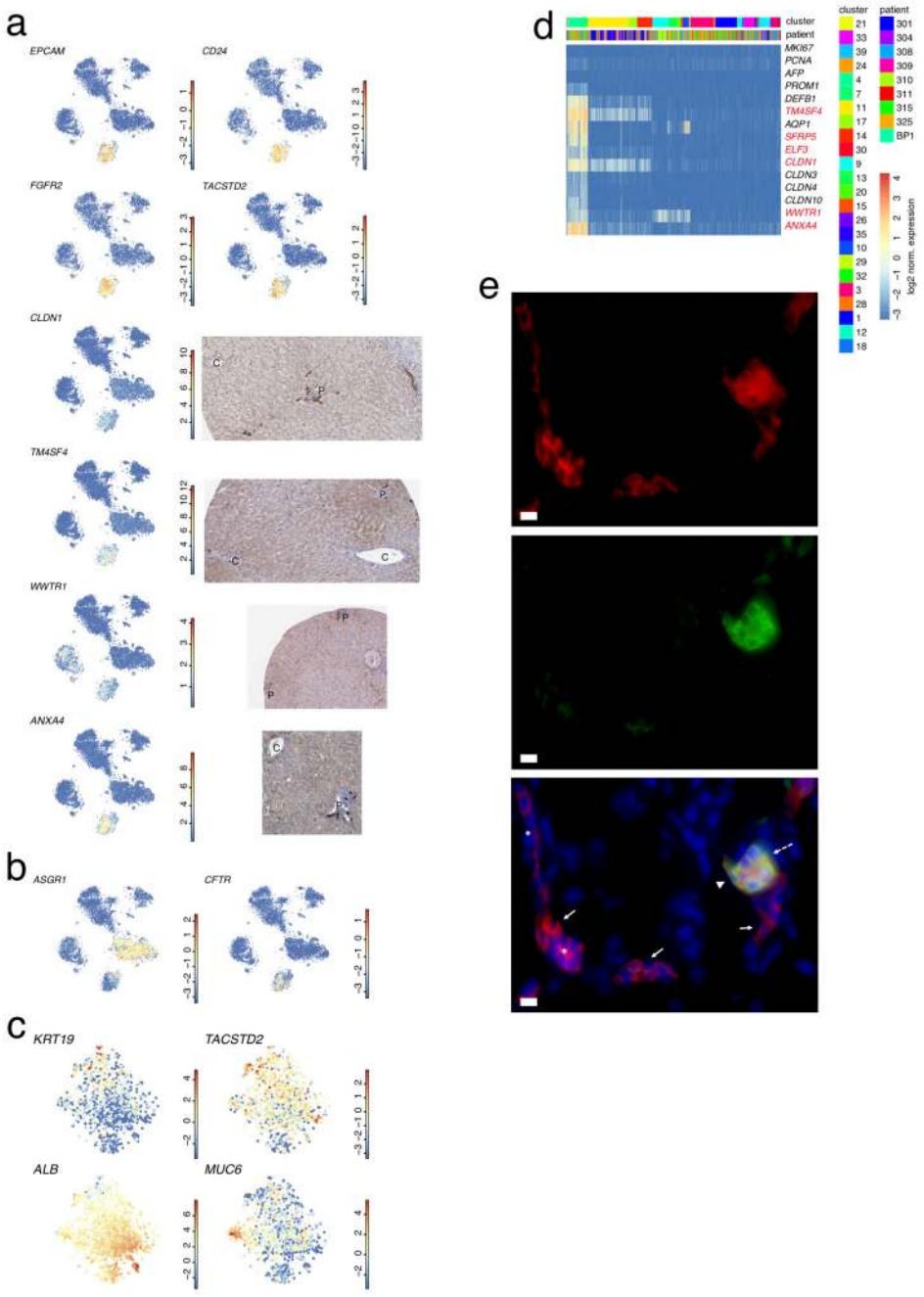


Extended Data Figure 6. Heterogeneity of NK and NKT cells in the human liver.

(**a – c**), Expression t-SNE maps of inferred markers of (a) cluster 5 (b) cluster 1, and (c) cluster 3. Cluster 5 comprises mainly *CD56*$^{+}$ *CD8A*$^{-}$ NK cells, some of which up-regulate *CCL4*. Cluster 1 comprises *CD56*$^{-}$ *CD8A*$^{+}$ NKT cells, which upregulate *CCL5*. Cluster 3 consists of both *CD56*$^{+}$ and *CD56*$^{-}$ *CD8A*$^{+}$ NKT cells. Clusters 1 and 3 express T cell receptor components exemplified by *CD3D* co-receptor expression. *CD56* is encoded by *NCAM1*. **d**, Differential expression of killer cell lectin-like receptor genes across the different populations shown in (a-c). **e**, Differential expression of granzyme genes across the

different populations shown in (a-c). The color bar in (a-e) indicates log2 normalized expression. (a,e) n=10,372 cells.

Europe PMC Funders Author Manuscripts



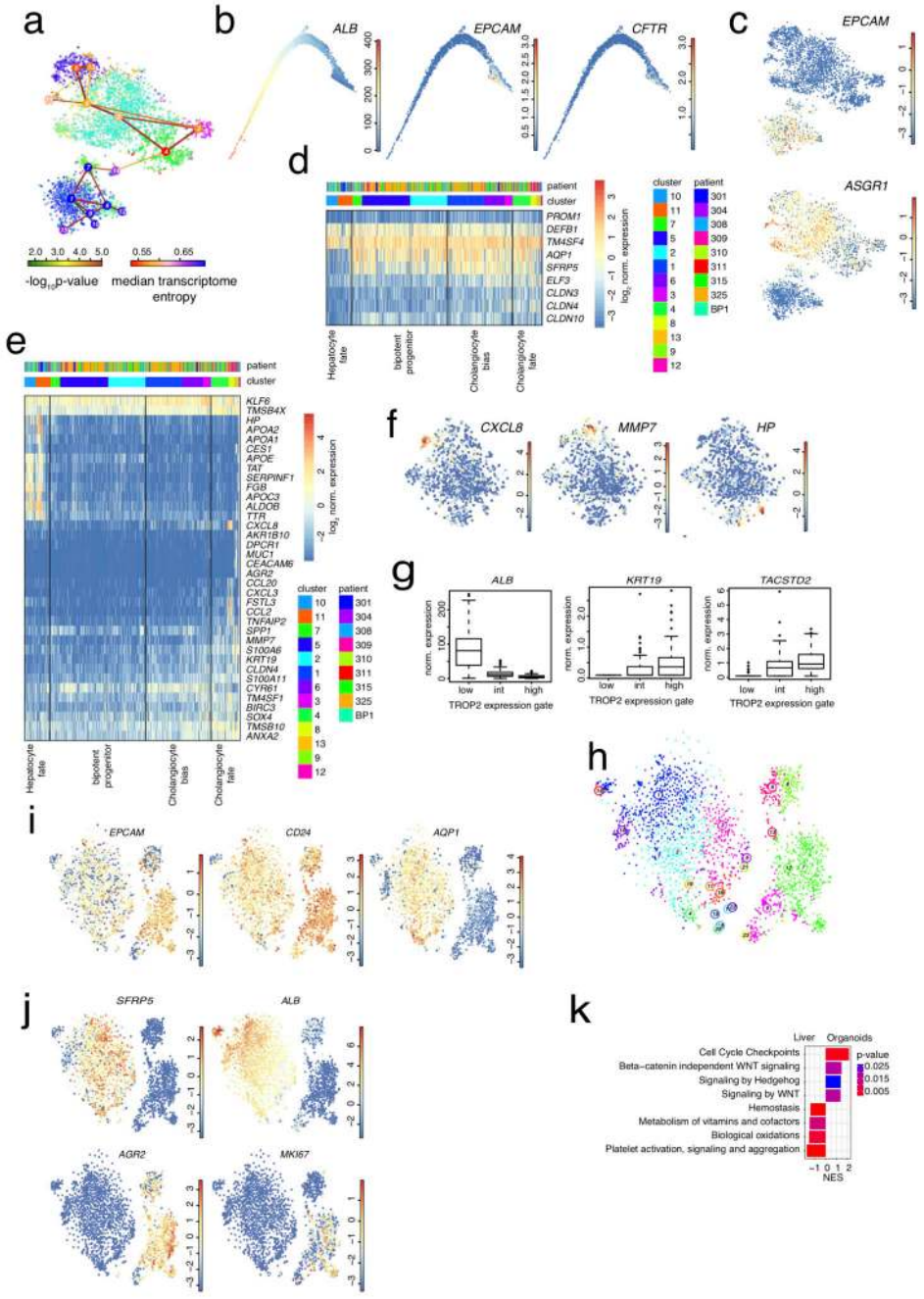
Extended Data Figure 7. ScRNA-Seq reveals novel marker genes expressed by *EPCAM*$^{+}$ cells.

a, Expression t-SNE maps (left) for *EPCAM*, *CD24*, *FGFR2*, *TACSTD2*, *CLDN1*, *TM4SF4*, *WWTR1*, and *ANXA4* (n=10,372 cells) and immunohistochemistry from the Human Protein Atlas (right) for CLDN1, TM4SF4, WWTR1, and ANXA4. The color bar in the expression t-SNE maps indicates log2 normalized expression. **b**, Expression t-SNE maps for *ASGR1* and *CFTR* (n=10,372 cells). The color bar in the expression t-SNE maps indicates log2 normalized expression. **c**, t-SNE maps showing expression of *KRT19, ALB*, *TACSTD2* and *MUC6* expression in the *EPCAM*$^{+}$ compartment (n=1,087 cells). The color bar indicates

Europe PMC Funders Author Manuscripts

log2 normalized expression. **d**, Expression heatmap of proliferation markers (*MKI67*, *PCNA*), *AFP*, and identified markers of the *EPCAM*$^{+}$ compartment. Genes highlighted in red correspond to newly identified markers of the EPCAM+ compartment. The heatmap comprises all clusters to show the specificity of the markers for the progenitor compartment. The expression analysis confirms the absence of proliferating and *AFP*$^{+}$ cells. **e**, Immunofluorescence labeling of EPCAM and KRT19 on human liver tissue. EPCAM $^{+}$KRT19$^{low/-}$ (solid arrow) in the canals of Hering (*) and EPCAM^{+}KRT19^{+} (broken arrow) cells in the biliary duct (arrowhead) are indicated. Nuclei are stained with DAPI. scale bar, 10μm (n=3 independent experiments).

Europe PMC Funders Author Manuscripts



Extended Data Figure 8. The EPCAM$^+$ compartment segregates into different major sub-populations.

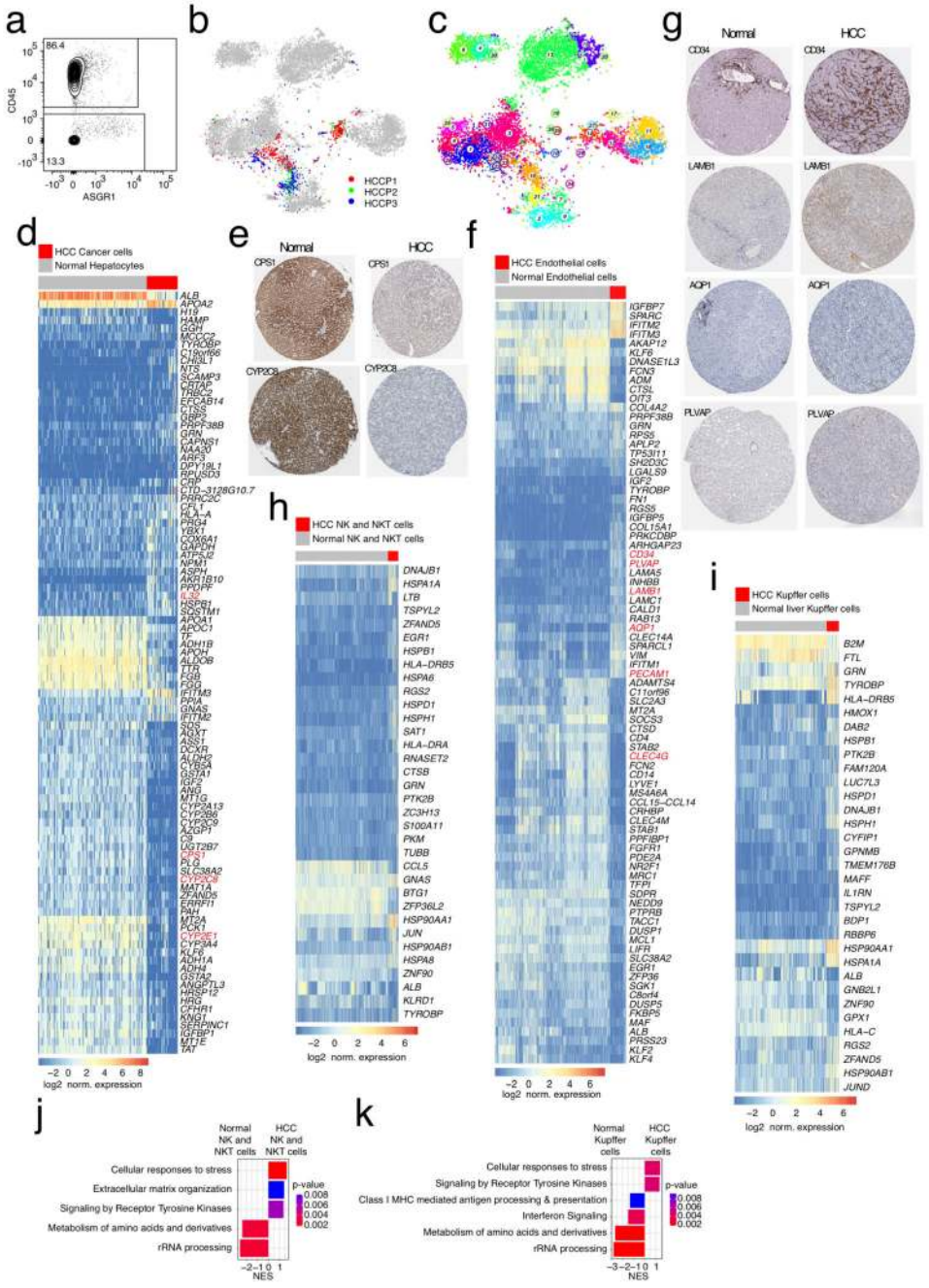
a, Separate RaceID3 and StemID2 analysis of the *EPCAM*$^+$ and the hepatocyte population reveals a lineage tree connecting *EPCAM*$^+$ cells to mature hepatocytes via an EPCAM+ hepatocyte progenitor cluster (part of the EPCAM+ population in Fig. 3b. Shown are links with StemID2 P<0.05. The node color highlights transcriptome entropy. **b**, Expression t-SNE map of *EPCAM* (left) and the hepatocyte marker *ASGR1* (right) for the population shown in (a). The color bar indicates log2 normalized expression. **c**, Two-dimensional

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

diffusion map representation of the population shown in (a) highlighting the expression of the hepatocyte marker *ALB* (left), *EPCAM* (center), and the mature cholangiocyte marker *CFTR* (right). The maps suggest continuous transitions from the *EPCAM*$^{+}$ compartment towards hepatocytes and mature cholangiocytes. (b,c) n=3,877 cells. **d**, Expression heatmap of *de novo* identified markers of the *EPCAM*$^{+}$ compartment, highlighting the expression distribution within clusters of this population only (see Fig. 3). **e**, Expression heatmap of all genes differentially expressed in the more mature clusters, belonging to the groups denoted as “Hepatocyte fate” and “Cholangiocyte fate”. For each of these clusters, the top ten up-regulated genes (Benjamini-Hochberg corrected $P<0.01$) were selected, and the joint set of these genes is shown in the figure. **f,** Expression t-SNE maps of *CXCL8, MMP7,* and *HP*. The color bar indicates log2 normalized expression. (d-f) n=1,087 cells. **g,** Normalized expression counts of *ALB, KRT19,* and *TACSTD2* in cells sequenced from the gates in (Fig. 4a) (n=293 cells). Boxes, interquartile range; whiskers, 5%- and 95%- quantile; data points, outliers. **h**, t-SNE map of RaceID3 clusters for the organoid cells and EPCAM+ cells from the patients (from Fig. 3) including cells sorted from the gates in (a). **i,** Expression t-SNE maps of *EPCAM*, *CD24,* and *AQP1* in the organoids cells and the patients’ EPCAM^{+} cells. **j,** Expression t-SNE maps of *SFRP5*, *ALB*, *AGR2*, and *MKI67*. The color bar indicates log2 normalized expression. **k,** GSEA of differentially expressed genes between organoid and EPCAM^{+} liver cells from the patients (Benjamini-Hochberg corrected $P<0.01$; Methods). NES, normalized enrichment score (h-k) n=11,610 genes.

Europe PMC Funders Author Manuscripts

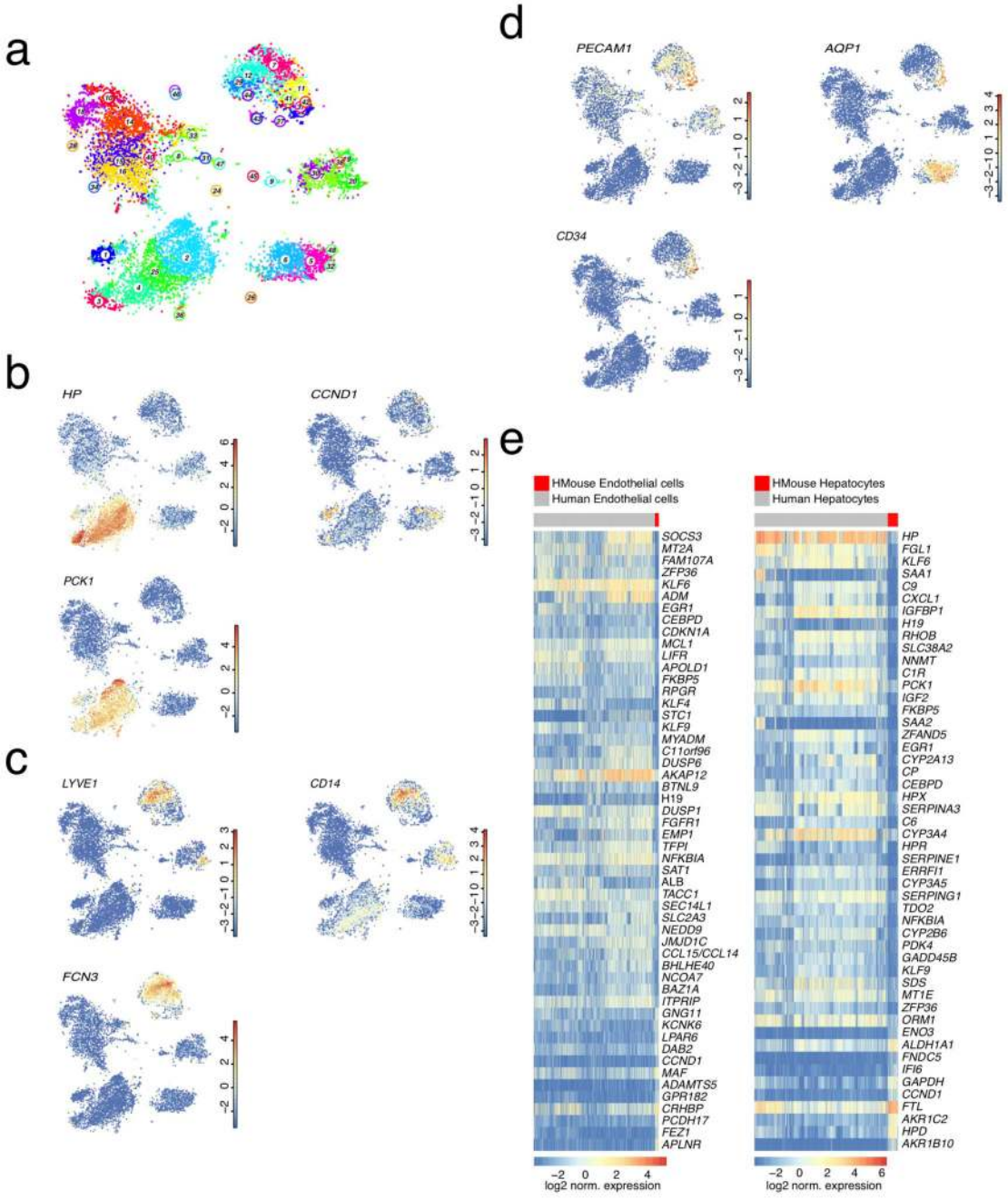


Extended Data Figure 9. Cell types from patient-derived HCC exhibit perturbed gene expression signatures.

a, FACS plot of CD45 and ASGR1 staining on cells from HCC samples (n=3 independent experiments). **b**, Symbol t-SNE map showing the IDs of the HCC patients (n=11,654 cells). **c**, t-SNE map showing RaceID3 clusters for normal liver cells co-analyzed with cells from HCC tissues (n=3 patients). **d**, Expression heatmap of differentially expressed genes between cancer cells from HCC and normal hepatocytes (Benjamini-Hochberg corrected $P<0.05$ and log2-foldchange >1.6; n=256 cells; Methods). Genes highlighted in red

correspond to differentially expressed genes validated by immunohistochemistry. **e**, Immunostaining of CPS1 and CYP2C8 in normal liver and HCC tissues from the Human Protein Atlas. **f**, Expression heatmap of differentially expressed genes between endothelial cells from HCC and normal endothelial cells from MaVEC and LSEC clusters. (Benjamini-Hochberg corrected $P<0.05$; log2-foldchange >1.5; n=1,936 cells; Methods). Genes highlighted in red correspond to differentially expressed genes validated by immunohistochemistry. **g**, Immunostaining of CD34, LAMB1, AQP1 and PLVAP in normal liver and HCC tissues from the Human Protein Atlas. **h,** Heatmap of differentially expressed genes between normal and HCC-resident NK and NKT cells (Benjamini-Hochberg corrected $P<0.05$; n=2754 cells; Methods). **i**, Heatmap of differentially expressed genes between normal and HCC-resident Kupffer cells (Benjamini-Hochberg corrected $P<0.05$; n=991 cells; Methods)**. j,** GSEA of differentially expressed genes between normal and HCC-resident NK and NKT cells (n=15,442 genes). **k,** GSEA of differentially expressed genes between normal and HCC-resident Kupffer cells (n=15,442 genes). (j,k) The bar chart shows the normalized enrichment score (NES) and highlights the p-value (Methods).

Europe PMC Funders Author Manuscripts



Extended Data Figure 10. Transplanted human liver cells in a humanized mouse model exhibit a distinct gene signature compared to cells within the human liver.

a, t-SNE map of RaceID3 clusters of liver cells from the patients co-analyzed with cells from the humanized mouse liver model. **b**, Expression t-SNE maps of the hepatocyte markers *ALB*. **c**, Expression t-SNE maps of the endothelial markers *CLEC4G*. **d**, Expression t-SNE maps of *HP*, *PCK1* and *CCND1*. **e**, Expression t-SNE maps of the liver endothelial cell zonated genes *LYVE1*, *FCN3* and *CD14*. **f**, Expression t-SNE maps of *PECAM1*, *CD34* and *AQP1*. The color bar in (a-f) indicates log2 normalized expression. (a-f) n= 10,683 cells.

g, Heatmaps of differentially expressed genes between hepatocytes (n=3,175 cells) and endothelial (n=1,710 cells) cells from the patients (Human Hepatocytes and Human Endothelial cells) and from the humanized mouse model (HMouse Hepatocytes and HMouse Endothelial cells). Benjamini-Hochberg corrected $P<0.05$; negative binomial (Methods).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the Max Planck Society, the German Research Foundation (DFG) (SPP1937 GR4980/1-1, GR4980/3-1, and GRK2344 MeInBio), by the DFG under Germany's Excellence Strategy (CIBSS – EXC-2189 – Project ID 390939984), and by the Behrens-Weise-Foundation (all to D.G.). This work was supported by ARC, Paris and Institut Hospitalo-Universitaire, Strasbourg (TheraHCC and TheraHCC2.0 IHUARC IHU201301187 and IHUARC2019 to T.F.B.), the European Union (ERC-AdG-2014-671231-HEPCIR to T.F.B., EU H2020-667273-HEPCAR to T.F.B, ERC-PoC-2016-PRELICAN to T.F.B), ANRS and the Foundation of the University of Strasbourg. This work was done under the framework of the LABEX ANR-10-LABX-0028_HEPSYS and Inserm Plan Cancer and benefits from funding from the state managed by the French National Research Agency as part of the Investments for the future. We thank Sebastian Hobitz and Konrad Schuldes from the FACS facility and Dr. Ulrike Bönisch from the Deep Sequencing facility of the Max Planck Institute of Immunobiology and Epigenetics. We acknowledge the CRB (Centre de Ressources Biologiques-Biological Resource Centre of the Strasbourg University Hospitals) for the management of regulatory requirements of patient-derived liver tissue. We thank Dr. Catherine Fauvelle and Laura Heydmann (Inserm U1110, University of Strasbourg) for their contributions to the initial single cell isolations used in the study. We thank Drs. Frank Juehling, François Duong and Catherine Schuster (Inserm U1110, University of Strasbourg) for helpful discussions. We thank the patients for providing informed consent to participate in the study and the nurses, technicians and medical doctors of the hepatobiliary surgery and pathology services of the Strasbourg University Hospitals for their support.

References

1. Michalopoulos GK, DeFrances MC. Liver regeneration. Science. 1997; 276:60–66. [PubMed: 9082986]
2. Ryerson AB, et al. Annual Report to the Nation on the Status of Cancer, 1975-2012, featuring the increasing incidence of liver cancer. Cancer. 2016; 122:1312–1337. DOI: 10.1002/cncr.29936 [PubMed: 26959385]
3. Grun D, van Oudenaarden A. Design and Analysis of Single-Cell Sequencing Experiments. Cell. 2015; 163:799–810. DOI: 10.1016/j.cell.2015.10.039 [PubMed: 26544934]
4. Herman JS, Sagar, Grun D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat Methods. 2018; doi: 10.1038/nmeth.4662
5. Grun D, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature. 2015; 525:251–255. DOI: 10.1038/nature14966 [PubMed: 26287467]
6. Jungermann K, Kietzmann T. Zonation of parenchymal and nonparenchymal metabolism in liver. Annu Rev Nutr. 1996; 16:179–203. DOI: 10.1146/annurev.nu.16.070196.001143 [PubMed: 8839925]
7. Gebhardt R. Metabolic zonation of the liver: regulation and implications for liver function. Pharmacol Ther. 1992; 53:275–354. [PubMed: 1409850]
8. Kietzmann T. Metabolic zonation of the liver: The oxygen gradient revisited. Redox Biol. 2017; 11:622–630. DOI: 10.1016/j.redox.2017.01.012 [PubMed: 28126520]
9. Halpern KB, et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature. 2017; 542:352–356. DOI: 10.1038/nature21065 [PubMed: 28166538]
10. MacParland SA, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nature Communications. 2018; 9doi: 10.1038/s41467-018-06318-7

11. Haghverdi L, Buttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods. 2016; 13:845–848. DOI: 10.1038/nmeth.3971 [PubMed: 27571553]
12. Strauss O, Phillips A, Ruggiero K, Bartlett A, Dunbar PR. Immunofluorescence identifies distinct subsets of endothelial cells in the human liver. Sci Rep. 2017; 7doi: 10.1038/srep44356
13. Halpern KB, et al. Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. Nat Biotechnol. 2018; 36:962–970. DOI: 10.1038/nbt.4231 [PubMed: 30222169]
14. Raven A, et al. Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration. Nature. 2017; 547:350–354. DOI: 10.1038/nature23015 [PubMed: 28700576]
15. Michalopoulos GK, Barua L, Bowen WC. Transdifferentiation of rat hepatocytes into biliary cells after bile duct ligation and toxic biliary injury. Hepatology (Baltimore, Md). 2005; 41:535–544.
16. Schmelzer E, et al. Human hepatic stem cells from fetal and postnatal donors. J Exp Med. 2007; 204:1973–1987. DOI: 10.1084/jem.20061603 [PubMed: 17664288]
17. Turner R, et al. Human hepatic stem cell and maturational liver lineage biology. Hepatology (Baltimore, Md). 2011; 53:1035–1045.
18. Grun D, et al. De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. Cell Stem Cell. 2016; 19:266–277. [PubMed: 27345837]
19. Okabe M, et al. Potential hepatic stem cells reside in EPCAM+ cells of normal and injured mouse liver. Development. 2009; 136:1951–1960. DOI: 10.1242/dev.031369 [PubMed: 19429791]
20. Cardinale V, et al. Multipotent stem/progenitor cells in human biliary tree give rise to hepatocytes, cholangiocytes, and pancreatic islets. Hepatology (Baltimore, Md). 2011; 54:2159–2172.
21. Kodama Y, et al. Hes1 is required for the development of intrahepatic bile ducts. Gastroenterology. 2003; 124:A123–A123. DOI: 10.1016/S0016-5085(03)80604-2
22. Sosa-Pineda B, Wigle JT, Oliver G. Hepatocyte migration during liver development requires Prox1. Nat Genet. 2000; 25:254–255. DOI: 10.1038/76996 [PubMed: 10888866]
23. Huch M, et al. Long-term culture of genome-stable bipotent stem cells from adult human liver. Cell. 2015; 160:299–312. DOI: 10.1016/j.cell.2014.11.050 [PubMed: 25533785]
24. Betge J, et al. MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. Virchows Arch. 2016; 469:255–265. [PubMed: 27298226]
25. Park S-W, et al. The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. Proc Natl Acad Sci U S A. 2009; 106:6950–6955. [PubMed: 19359471]
26. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. 2018; 391:1301–1314. DOI: 10.1016/S0140-6736(18)30010-2 [PubMed: 29307467]
27. Matkowskyj KA, et al. Aldoketoreductase family 1B10 (AKR1B10) as a biomarker to distinguish hepatocellular carcinoma from benign liver lesions. Hum Pathol. 2014; 45:834–843. [PubMed: 24656094]
28. Rantakari P, et al. The endothelial protein PLVAP in lymphatics controls the entry of lymphocytes and antigens into lymph nodes. Nat Immunol. 2015; 16:386–396. DOI: 10.1038/ni.3101 [PubMed: 25665101]
29. Grompe M, Strom S. Mice with human livers. Gastroenterology. 2013; 145:1209–1214. DOI: 10.1053/j.gastro.2013.09.009 [PubMed: 24042096]
30. Azuma H, et al. Robust expansion of human hepatocytes in Fah-/-/Rag2-/-/Il2rg-/- mice. Nat Biotechnol. 2007; 25:903–910. DOI: 10.1038/nbt1326 [PubMed: 17664939]
31. Uhlen M, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015; 347doi: 10.1126/science.1260419
32. Krieger SE, et al. Inhibition of hepatitis C virus infection by anti-claudin-1 antibodies is mediated by neutralization of E2-CD81-claudin-1 associations. Hepatology. 2010; 51:1144–1157. DOI: 10.1002/hep.23445 [PubMed: 20069648]
33. Lieber A, Peeters MJ, Gown A, Perkins J, Kay MA. A modified urokinase plasminogen activator induces liver regeneration without bleeding. Hum Gene Ther. 1995; 6:1029–1037. DOI: 10.1089/hum.1995.6.8-1029 [PubMed: 7578415]

34. Mailly L, et al. Clearance of persistent hepatitis C virus infection in humanized mice using a claudin-1-targeting monoclonal antibody. Nat Biotechnol. 2015; 33:549–554. DOI: 10.1038/nbt.3179 [PubMed: 25798937]
35. Hashimshony T, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016; 17:77.doi: 10.1186/s13059-016-0938-8 [PubMed: 27121950]
36. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. DOI: 10.1093/bioinformatics/btp698 [PubMed: 20080505]
37. Grun D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. Nat Methods. 2014; 11:637–640. DOI: 10.1038/nmeth.2930 [PubMed: 24747814]
38. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106.doi: 10.1186/gb-2010-11-10-r106 [PubMed: 20979621]
39. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012; 16:284–287. [PubMed: 22455463]
40. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–15550. DOI: 10.1073/pnas.0506580102 [PubMed: 16199517]
41. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34:267–273. DOI: 10.1038/ng1180 [PubMed: 12808457]
42. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016; 12:477–479. DOI: 10.1039/c5mb00663e [PubMed: 26661513]
43. Broutier L, et al. Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. Nat Protoc. 2016; 11:1724–1743. DOI: 10.1038/nprot.2016.097 [PubMed: 27560176]
44. Aizarani N, et al. Protocol for Single-Cell RNA-Sequencing of Cryopreserved Human Liver Cells. Protoc Exch. 2019

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

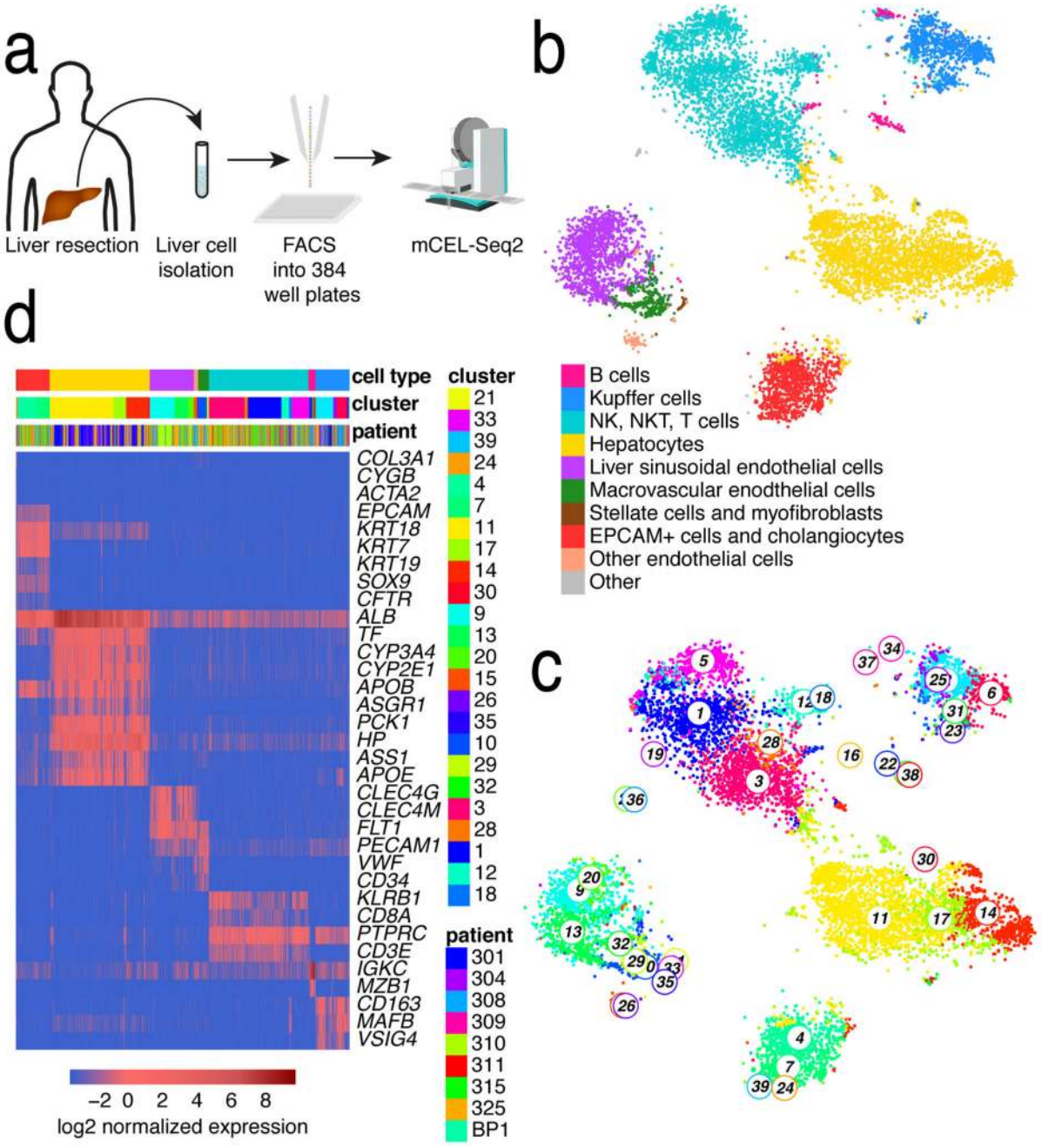


Figure 1. ScRNA-seq reveals cell types in the adult human liver.

a, Outline of the protocol used for scRNA-seq of human liver cells. Specimens from liver resections were digested to prepare single cell suspensions. Cells were sorted into 384-well plates and processed according to the mCEL-Seq2 protocol. **b**, t-SNE map of single-cell transcriptomes from normal liver tissue of nine different donors highlighting the main liver cell compartments. **c**, t-SNE map of single-cell transcriptomes highlighting RaceID3 clusters, revealing sub-type heterogeneity in all major cell populations of the human liver. **d**, Heatmap showing the expression of established marker genes for each cell compartment.

Color bars indicate patient, major cell type, and RaceID3 cluster. Scale bar, log2-transformed normalized expression. “Other” in the legend of (b) denotes various small populations comprising 22 red blood cells and 46 cells that cannot be unambiguously annotated. “Other endothelial cells” cannot be unambiguously classified as LSECs or MaVECs. (b,c) n= 10,372 cells.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

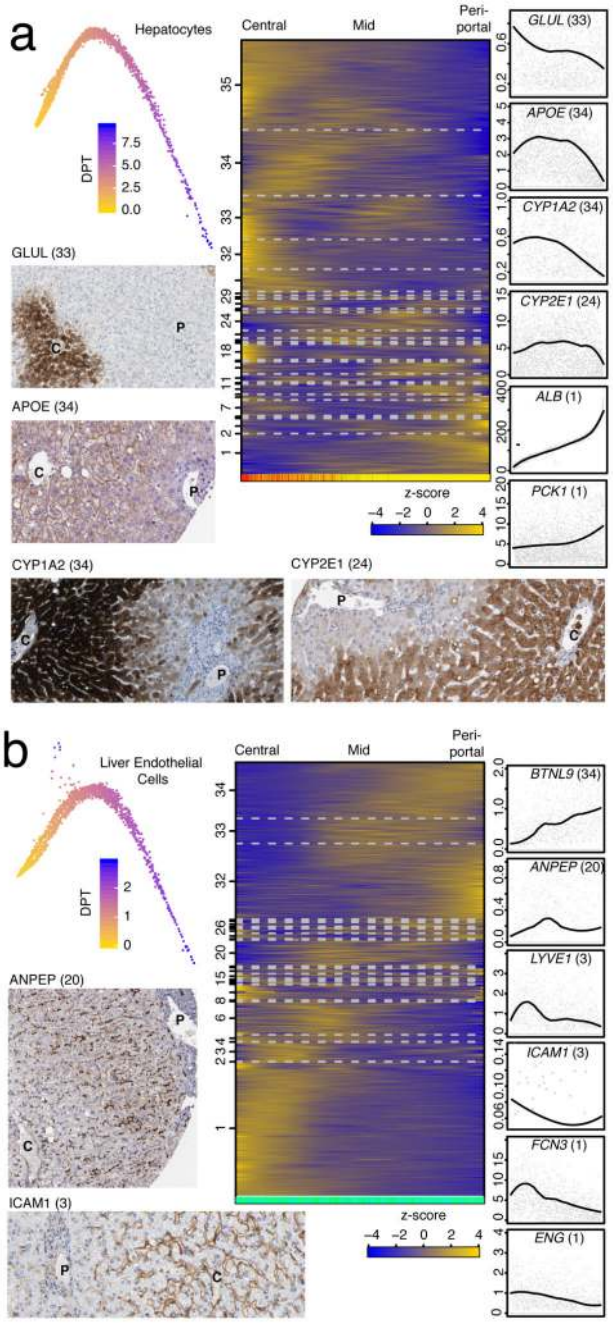


Figure 2. Heterogeneity and zonation of hepatocytes and endothelial cells.

a, Diffusion maps (left) and self-organizing maps (SOM, middle) of single-cell transcriptome-derived zonation profiles for hepatocytes (n=2,534 cells). DPT indicates diffusion-pseudotime and is here interpreted as a spatial zonation coordinate. Zonation profiles of *GLUL* (central), *APOE* (midzonal), *CYP1A2* and *CYP2E1* (central/midzonal), *ALB* and *PCK1* (periportal), and immunostaining (bottom left) of GLUL, APOE, CYP1A2, and CYP2E1 from the Human Protein Atlas31. See Extended Data Fig. 3g for additional stainings. **b,** Diffusion maps (left) and self-organizing maps (SOM, middle) of the single-cell

transcriptome-derived zonation profiles for endothelial cells (n=1,361 cells). Zonation profiles of *BTNL9* and *ANPEP* (periportal), *LYVE1 and FCN3* (midzonal), and *ICAM1*, *FCN3*, and *ENG* (central), and immunostaining of, ICAM1 and ANPEP from the Human Protein Atlas (bottom left). (a,b) P, portal tracts; C, central. color bar, RaceID3 cluster. The y-axis of the zonation profiles indicates normalized expression.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

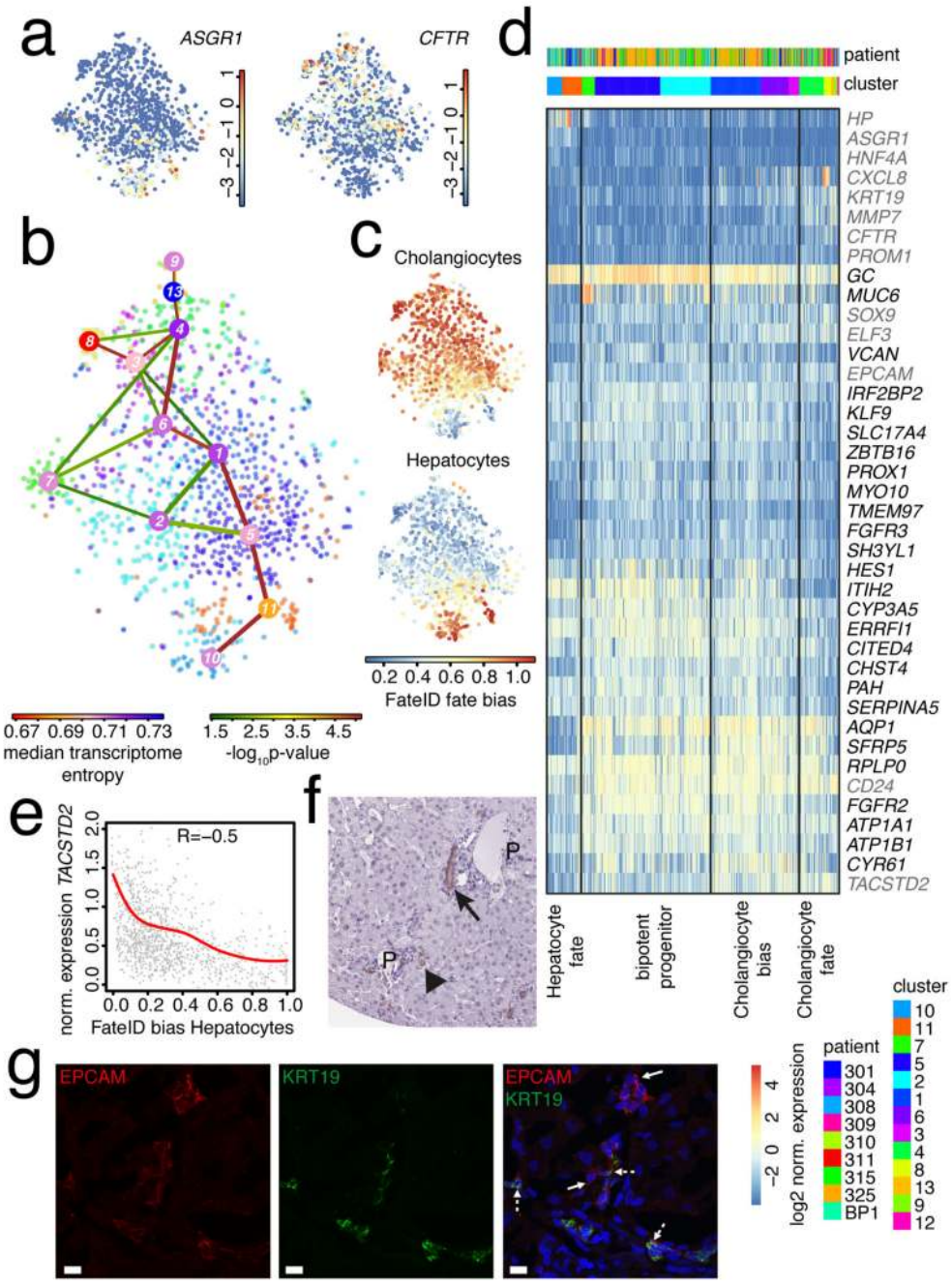


Figure 3. Identification of a putative progenitor population in the adult human liver.

a, Expression t-SNE maps of *ASGR1* and *CFTR* for the EPCAM$^+$ compartment only. The color bar indicates log2 normalized expression. **b**, StemID218 analysis of the *EPCAM*$^+$ compartment. Shown are links with StemID2 P<0.05. node color, transcriptome entropy. **c**, FateID analysis of the *EPCAM*$^+$ compartment highlights populations that are preferentially biased towards hepatocyte progenitors and cholangiocytes, respectively, and reveals similar bias towards both lineages in the central population (clusters 1,2,5,6,7). The color bar indicates lineage probability. **d**, Expression heatmap of selected hepatocyte markers (*HP*,

ASGR1), mature cholangiocyte genes (*KRT19*, *CFTR, CXCL8, MMP7*), additional progenitor markers (highlighted in grey), and all genes up-regulated in the central population (clusters 1,2,5,6,7) within the *EPCAM*$^{+}$ compartment (Benjamini-Hochberg corrected $P<0.01$; foldchange>1.33; Methods). Four compartments are indicated resolving the predicted fate bias (see Extended Data Fig. 8). **e**, Correlation of nearest-neighbor-imputed (k=5) expression (using RaceID3) of *TACSTD2* and hepatocyte bias predicted by FateID. red line, loess regression. R, Spearman's rank correlation. (a,e) n=1,087 cells**. f**, Immunostaining of TROP2 from the Human Protein Atlas (n=3 biologically independent samples). The arrow points to a bile duct and the arrowhead to a bile ductule. **g.** Immunofluorescence labeling of EPCAM and KRT19. EPCAM^{+}KRT19$^{low/-}$ (solid arrow) and EPCAM^{+}KRT19+ (broken arrow) cells are indicated. Nuclei are stained with DAPI. Images are maximum z-stack projections of 6μm. scale bar, 10μm. (n=3 independent experiments).

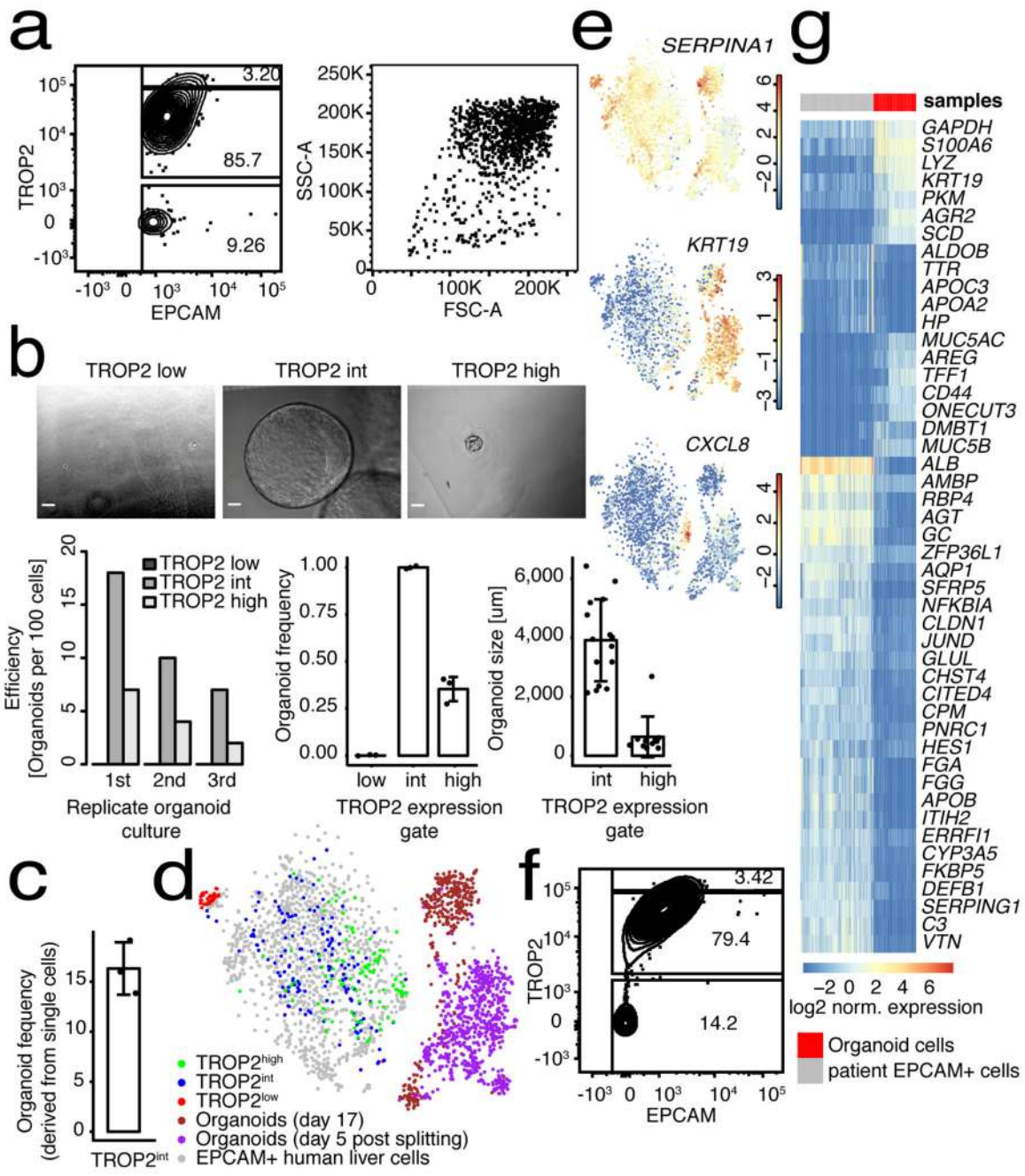


Figure 4. TROP2^{int} cells are a source of liver organoid formation.

a, FACS plots for EPCAM^{+} cells showing EPCAM and TROP2 expression (left) and forward and side scatter (right) (n=6 independent experiments). The gates for the three compartments are shown. **b,** Top panel: Organoid culturing of cells from the TROP2$^{low/-}$, TROP2^{int}, and TROP2^{high} compartments (n=3 independent experiments). Bottom panel: Number of organoids (left), the organoid frequency relative to the TROP2^{int} compartments (center), and size of organoids (right); n=3 patients, 100 seeded cells each. scale bar, 400 μm. **c**, Organoid frequency in single-cell cultures of TROP2^{int} cells (n=3 independent

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

experiments, 96-cells each). Due to the small number of cells we were unable to purify single cells from the other gates for culture. (b,c) Measure of center, mean. Error bars, standard deviation. **d,** Symbol t-SNE map showing the organoid cells, the original $EPCAM^{+}$ data (from Fig. 3) and the cells sorted from the gates in (a). **e,** Expression t-SNE maps of *SERPINA1*, *KRT19*, and *CXCL8*. The color bar indicates log2 normalized expression. **f,** FACS plot of EPCAM and TROP2 expression for organoid cells grown from the $TROP2^{int}$ compartment 17 days after initial culture (n=3 independent experiments). **g,** Expression heatmap of differentially expressed genes between patient and organoid cells (Benjamini-Hochberg corrected $P<0.05$ (Method), mean expression >0.7, log2- foldchange >2). (d-e, g) n=2,870 cells.

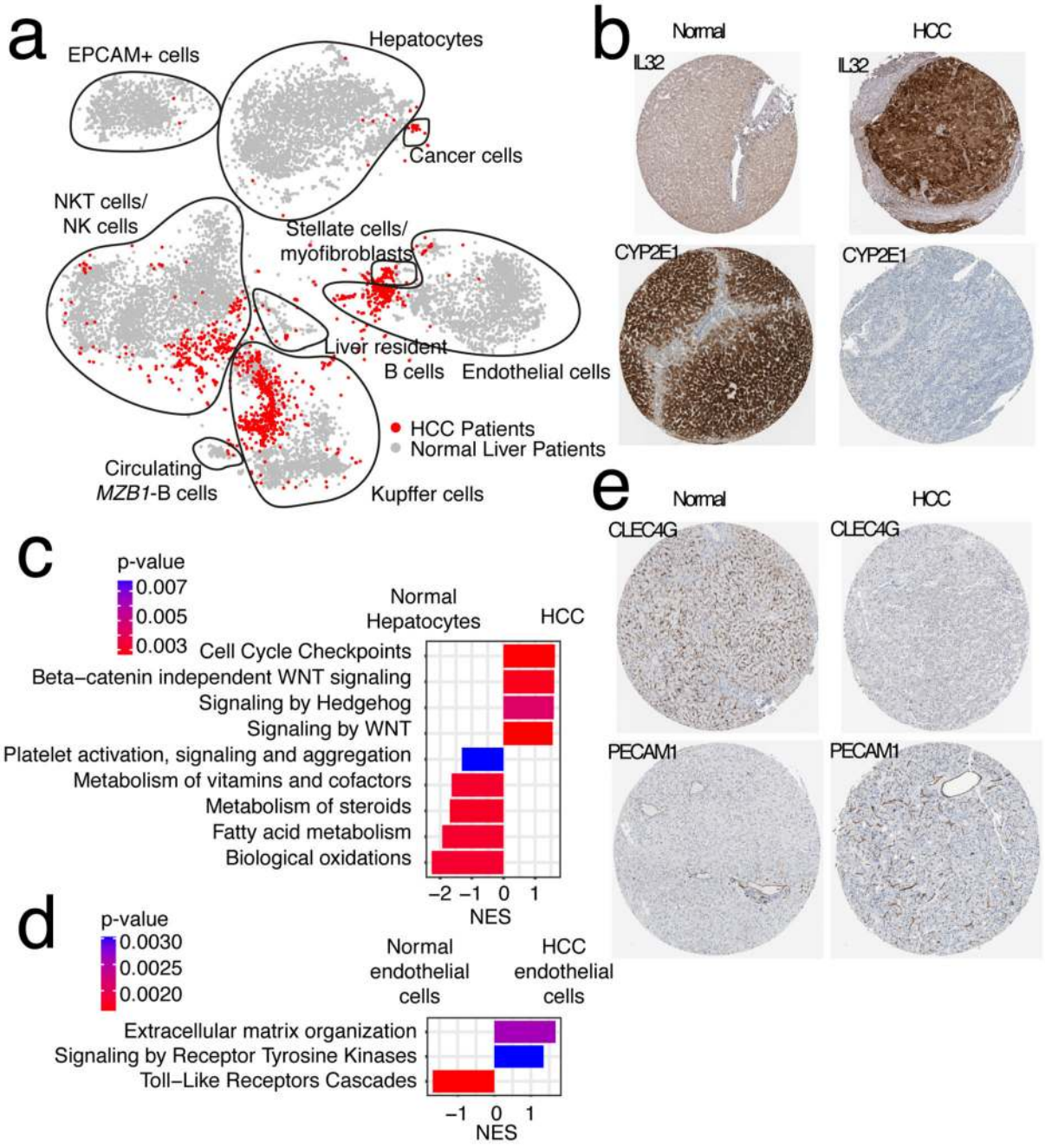


Figure 5. ScRNA-seq of patient-derived HCC reveals cancer-specific gene signatures and perturbed cellular phenotypes.

a, Symbol t-SNE map highlighting normal liver cells and cells from HCC. n=11,654 cells from n=3 different patients. **b,** Immunostaining of IL32 and CYP2E1 in normal liver and HCC tissue. **c,** GSEA for differentially expressed genes between cancer cells from HCC and normal hepatocytes (n=15,442 genes). **d**, GSEA for differentially expressed genes between normal endothelial cells and endothelial cells from HCC (n=15,442 genes). (c, d) Benjamini-Hochberg corrected $P<0.01$; NES, normalized enrichment score; Methods. **e,**

Immunostaining of CLEC4G and PECAM1 in normal liver tissue and HCC tissue. All stainings are taken from the Human Protein Atlas(31).

Europe PMC Funders Author Manuscripts

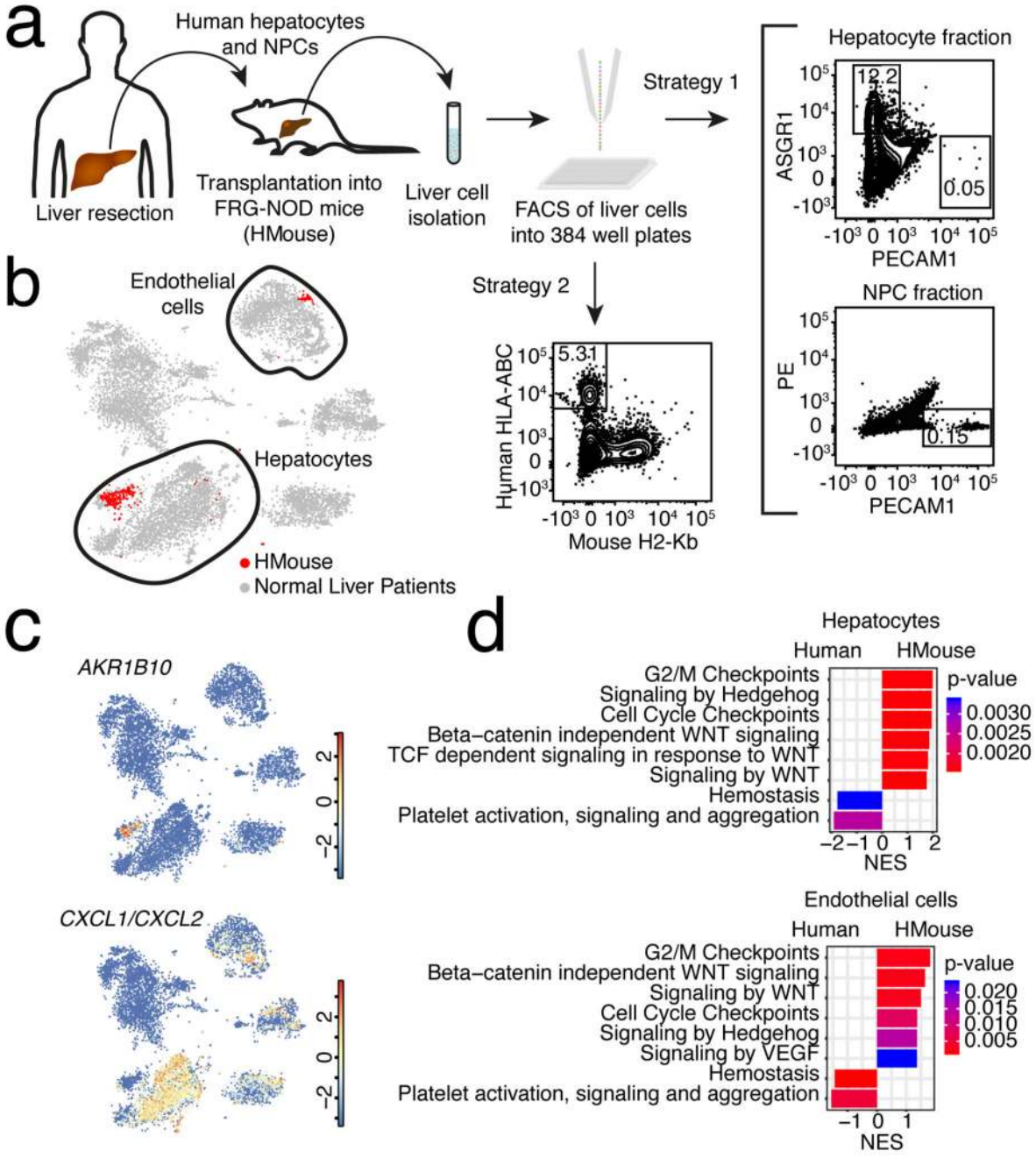


Figure 6. Exploring the gene expression signature of human liver cells in a humanized mouse model.

a, Outline of the transplantation of human liver cells (hepatocytes and non-parenchymal cells) into the FRG-NOD mouse and the two sorting strategies of human cells from the mouse liver. **b**, Symbol t-SNE map highlighting normal liver cells and cells from the humanized mouse model. The main engrafted cell types (hepatocytes and endothelial cells) are circled. **c**, Expression t-SNE maps of *AKR1B10* and *CXCL1*/*CXCL2*. The color bar indicates log2 normalized expression. n=10,683 cells. **d**, GSEA of differentially expressed

genes between hepatocytes and endothelial cells from the humanized mouse (HMouse) and from the patients (Human). n=13,614 genes; Benjamini-Hochberg corrected $P<0.01$; NES, normalized enrichment score; Methods.