nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-022-00989-7

Check for updates

Single-cell atlas of human liver development reveals pathways directing hepatic cell fates

Brandon T. Wesley[1,2,15], Alexander D. B. Ross[1,2,3,15], Daniele Muraro[1,2,4], Zhichao Miao[4,5], Sarah Saxton[6,7], Rute A. Tomaz[1,2], Carola M. Morell[1,2], Katherine Ridley[1,3], Ekaterini D. Zacharis[1,2], Sandra Petrus-Reurer[2,8], Judith Kraiczy[3], Krishnaa T. Mahbubani[2], Stephanie Brown[1,2], Jose Garcia-Bernardo[4], Clara Alsinet[4], Daniel Gaffney[4], Dave Horsfall[9], Olivia C. Tysoe[1,2], Rachel A. Botting[10], Emily Stephenson[10], Dorin-Mirel Popescu[10], Sonya MacParland[11], Gary Bader[11], Ian D. McGilvray[12], Daniel Ortmann[1,2], Fotios Sampaziotis[1,2], Kourosh Saeb-Parsy[2,8], Muzlifah Haniffa[4,10,13], Kelly R. Stevens[6,7], Matthias Zilbauer[1,3], Sarah A. Teichmann[4,14,16] and Ludovic Vallier[1,2,16] ✉

The liver has been studied extensively due to the broad number of diseases affecting its vital functions. However, therapeutic advances have been hampered by the lack of knowledge concerning human hepatic development. Here, we addressed this limitation by describing the developmental trajectories of different cell types that make up the human liver at single-cell resolution. These transcriptomic analyses revealed that sequential cell-to-cell interactions direct functional maturation of hepatocytes, with non-parenchymal cells playing essential roles during organogenesis. We utilized this information to derive bipotential hepatoblast organoids and then exploited this model system to validate the importance of signalling pathways in hepatocyte and cholangiocyte specification. Further insights into hepatic maturation also enabled the identification of stage-specific transcription factors to improve the functionality of hepatocyte-like cells generated from human pluripotent stem cells. Thus, our study establishes a platform to investigate the basic mechanisms directing human liver development and to produce cell types for clinical applications.

The liver fulfils a broad spectrum of functions including blood detoxification, metabolite storage, lipid/glucose metabolism and secretion of serum proteins. These critical tasks are performed mainly by the hepatocytes, which are supported by a diversity of cell types. Kupffer cells are tissue-specific resident macrophages responsible for liver homoeostasis and immunity[1]. Hepatic stellate cells sequester vitamin A in healthy organs while promoting fibrosis through collagen secretion during disease[2]. Cholangiocytes form the epithelial lining of the biliary tree, which transports bile into the intestine[3], and play a role in liver repair during chronic injury[4,5]. Finally, sinusoidal endothelial cells provide a permeable interface with circulating blood and promote regeneration after liver damage[6]. Importantly, adult liver cells have been broadly characterized using diverse methods, including detailed single-cell transcriptomic analyses[7–10]. However, the study of these cell types during foetal life remains limited, especially in humans, owing to the scarcity of descriptive studies exploring early liver development at high resolution[11]. This knowledge gap presents a major challenge in the advancement of new therapies, especially for applications of regenerative medicine. In this article, we addressed this limitation by performing single-cell RNA sequencing (scRNA-seq) analyses on human foetal and adult livers (Fig. 1a). This single-cell map uncovered not only the developmental trajectories of the different cell types that make up the liver, but also the cell-to-cell interactions controlling organogenesis. We took advantage of this information to isolate human hepatoblasts, which serve as the early progenitors of the liver parenchyma, and demonstrated that they can be propagated as organoids to model developmental processes. Finally, we utilized this developmental map to assess the differentiation path of human pluripotent stem cells (hPSCs) into hepatocyte-like cells (HLCs) and uncovered transcription factors capable of improving the resemblance of HLCs to adult hepatocytes. Together, our results provide insights into liver development that allow the establishment of an in vitro platform for modelling human liver development while providing the knowledge necessary to improve the production of hepatocytes in vitro[12].

Single-cell transcriptomic map of the developing human liver

To characterize the cellular landscape of the developing human liver (Fig. 1a), single-cell transcriptomes were derived from primary tissue using methods tailored to each stage of liver development as

[1]Wellcome—MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [2]Department of Surgery, University of Cambridge, Cambridge, UK. [3]Department of Paediatrics, University of Cambridge, Cambridge, UK. [4]Wellcome Sanger Institute, Hinxton, UK. [5]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK. [6]Departments of Bioengineering and Pathology, University of Washington, Seattle, WA, USA. [7]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, USA. [8]NIHR Cambridge Biomedical Research Centre, Cambridge, UK. [9]Digital Institute, Newcastle University, Newcastle upon Tyne, UK. [10]Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK. [11]University of Toronto, Toronto, Ontario, Canada. [12]Multi-Organ Transplant Program, Toronto General Hospital Research Institute, Toronto, Ontario, Canada. [13]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [14]Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK. [15]These authors contributed equally: Brandon T. Wesley, Alexander D. B. Ross. [16]These authors contributed equally: Sarah A. Teichmann, Ludovic Vallier. ✉e-mail: lv225@cam.ac.uk

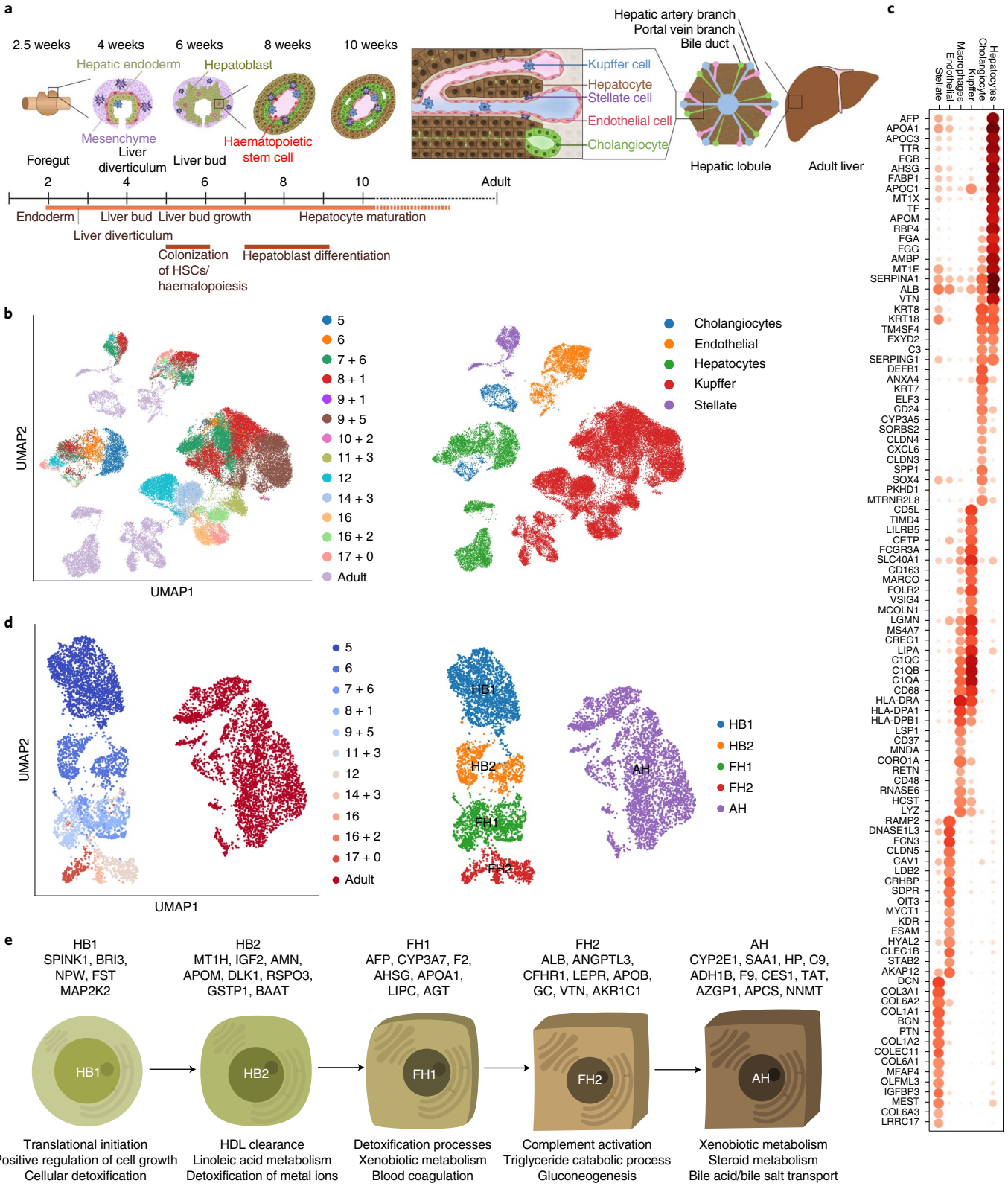


Fig. 1 | Single-cell transcriptomic map of human liver development. a, Schematic representation of human liver development. **b**, UMAP visualization of all integrated single-cell transcriptomic data of foetal and adult human hepatic cells generated using the 10x Genomics workflow; annotation indicates PCW + days (left) and the cell-specific lineages (right). **c**, Gene expression values of selected DEGs for each hepatic cell lineage. Gene-expression frequency (fraction of cells within each cell type expressing the gene) is indicated by dot size and level of expression by colour intensity; colour intensity shows 'gene expression [log-normalized, scaled counts]'. **d**, UMAP visualization of hepatocyte developmental trajectory (left) and annotation of developmental stages based on Louvain analysis (right). **e**, Characteristic genes induced at each stage of hepatocyte differentiation and corresponding Gene Ontology (GO). Plots integrate scRNA-seq data from $n = 17$ independent foetal livers aged 5–17 PCWs and $n = 16$ independent adult livers.

well as from existing datasets (for dataset references, see Methods). Droplet-based scRNA-seq was performed to profile a total of 237,978 hepatic cells[13], of which 87% passed quality control[14–17]. Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction[18] and subclustering[19,20] of these cellular transcriptomes showed that our approach captured the main cell types that make up the liver (Fig. 1b,c). Of note, cholangiocytes were the least represented cell type in our collection, confirming the difficulty of isolating these cells from liver tissue[8]. In addition, cholangiocytes were identified at only 7 post-conceptional weeks (PCWs), reinforcing previous studies indicating that these cells differentiate from hepatoblasts after 7–8 PCWs[21,22]. Concerning endothelial cells, the first cells were captured from 5 PCWs at the time when liver vasculature is known to be established[23,24]. Tissue-resident Kupffer cells could be distinguished from monocyte-derived macrophages by the expression of MARCO, CD163, FCGR3A and CD5L[8] and the absence of LSP1 and CD48 expression (Fig. 1c). Finally, hepatic stellate cells were captured from 5 PCWs, supporting studies in mice suggesting that these cells could be derived from the septum transversum at 3–5 PCWs[2,25,26]. All data generated by this study are available for visualization through our online portal[27]: https://collections.cellatlas.io/liver-development. Importantly, transcriptomic observations were validated by immunostaining on primary human foetal liver (Extended Data Fig. 1a). Collectively, these results show that our single-cell atlas captured the major cell types of the human liver and their dynamic diversity during development.

Developmental trajectory of liver cells

Using this dataset, we examined the developmental trajectory of each cell type, starting with hepatocytes. Principal component analysis (PCA), Louvain clustering and diffusion pseudotime (DPT) analyses[28,29] defined five hepatocyte developmental stages (Fig. 1d and Extended Data Fig. 1): hepatoblast stages 1 and 2 (HB1 at 5 PCWs and HB2 at 6 PCWs), foetal hepatocyte stages 1 and 2 (FH1 at 7–11 PCWs and FH2 at 12–17 PCWs) and adult hepatocytes (AH). Each stage displayed distinct transcriptional changes indicative of unique cell states (Extended Data Fig. 1b–e). Accordingly, this analysis identified markers specific to each stage, including SPINK1 for hepatoblasts, GSTA1 for foetal hepatocytes and haptoglobin (HP) for adult hepatocytes. We also observed that each stage of hepatocyte development was marked by the induction of genes associated with specific liver function (Fig. 1e). Thus, hepatocytes follow a progressive functional maturation during organogenesis corresponding to the acquisition of hepatic activity during foetal life.

We then performed similar analyses on cholangiocytes, stellate cells, endothelial cells and Kupffer cells (Fig. 2). Briefly, only cholangiocytes seemed to gradually differentiate from the HB2 stage, whereas PCA analyses did not reveal major differences among sequential timepoints for most non-parenchymal hepatic cell types. More precisely, Louvain clustering and DPT allowed for the distinction of an embryonic stage at 5–6 PCWs, intermediate foetal stage between 7 and 17 PCWs, and an adult state (Fig. 2a,c,e,g). This suggested that these cell types may not undergo substantial functional maturation during foetal life after their initial embryonic specification. Interestingly, these three stages correspond to major modifications in the liver environment: liver bud formation, colonization by the haematopoietic system at 7 PCWs and the shift from foetal to adult cells[24]. Thus, these data suggest that the developmental trajectory of hepatic cells is influenced by major developmental events while only hepatocytes seem to undergo a progressive functional maturation.

We then further demonstrated the utility of our single-cell map in defining the embryonic origin of specific cell types. We decided to focus on hepatic stellate cells, since previous studies have reached divergent conclusions[2]. Louvain clustering on early stellate cells revealed a population of cells expressing mesenchymal markers, one population with an endothelial bias and a third population expressing markers for both lineages (endothelial: CDH5, LYVE1, KDR and STAB2; stellate cells: PDGFRB, VIM, DES and COL1A1; Fig. 3a–c). DPT analyses confirmed that foetal endothelial and stellate cells could originate from this stellate–endothelial progenitor population, termed SEpro (Fig. 3c–f), while gene expression analyses revealed that these cells expressed genes associated with proliferation and DNA replication characteristic of a stem or progenitor state. Immunohistochemistry validations on primary tissue revealed cells expressing both PDGFRB and CDH5 located within the vasculature of the 6 PCW liver (Fig. 3b), thereby confirming the existence of this progenitor in vivo. Of note, previous studies in model organisms have suggested the existence of such progenitors without functional demonstration[30,31]. To further address this limitation, we decided to validate the existence of such progenitors during differentiation of hPSCs in vitro. We first performed scRNA-seq analyses on hPSCs differentiating into endothelial[32] and hepatic stellate cells[33]. UMAP, PCA and DPT comparison of these differentiations reveal an overlapping stage sharing the expression of markers specific for both lineages (Extended Data Fig. 2a–d). To confirm that this stage could include a common progenitor, hPSCs were differentiated into endothelial cells for 3.5 days and then grown in culture conditions inductive for hepatic stellate cells. The resulting cells were able to transition away from the endothelial pathway characterized by the expression of CDH5, KDR and VWF while acquiring the stellate cells markers PDGFRA, COL1A1, ACTA2 and NCAM (Extended Data Fig. 2e,f). Taken together, these results illustrate how single-cell observations can be combined with in vitro differentiation to further understand the developmental process leading to stellate cell production.

Human HBO derivation and differentiation

We subsequently used our single-cell analyses to direct the isolation and in vitro growth of hepatoblasts, as they represent the natural stem cell of the liver during development. Our single-cell analyses showed that hepatoblast HB1 and HB2 display characteristics of these early liver stem cells. Indeed, these cells expressed WNT target genes associated with adult stem cells such as LGR5, high levels of cell cycle regulators suggesting self-renewal capacity, and markers specific for both biliary and hepatocytic lineages indicative of bipotential capacity of differentiation (Extended Data Fig. 1e). On the basis of these observations and previous reports showing the crucial role of WNT[34], we hypothesized that this signalling could support the growth of hepatoblasts in vitro. To explore this possibility, 6 PCW livers were dissociated into single-cells that were sorted on the basis of EPCAM expression and then grown in 3D culture conditions supplemented with WNT (Fig. 4a). The isolated cells formed branching organoids that could be expanded for more than 20 passages (Extended Data Fig. 3a,b). These hepatoblast organoids (HBOs) homogeneously expressed hepatoblast markers (Fig. 4b and Extended Data Fig. 3c,d), and scRNA-seq analyses demonstrated that they closely resembled their in vivo counterparts, especially the HB2 stage (Fig. 4c and Extended Data Fig. 3e,f). Of note, neither HBOs nor HB1/2 cells express NCAM, thereby excluding the presence of hepatic stem cells in our analyses[35]. To confirm HBO bipotentiality, we transplanted tdTomato-HBO into Fah^{-}/Rag2^{-}/Il2rg^{-} (FRG) mice[36] using an approach developed for primary hepatocytes[37,38]. After 27 days, implants were recovered, and red fluorescent cells could be observed in all the grafts (Extended Data Fig. 3g), indicating that HBOs had engrafted efficiently. Haematoxylin and eosin staining of explanted tissue sections revealed the presence of numerous nodules resembling densely packed hepatocytes, as well as biliary epithelial-like cells assembled into structures resembling bile ducts (Extended Data Fig. 3h). Engrafted organoids stained positive for both KRT18 and AFP at the time of implant, but AFP

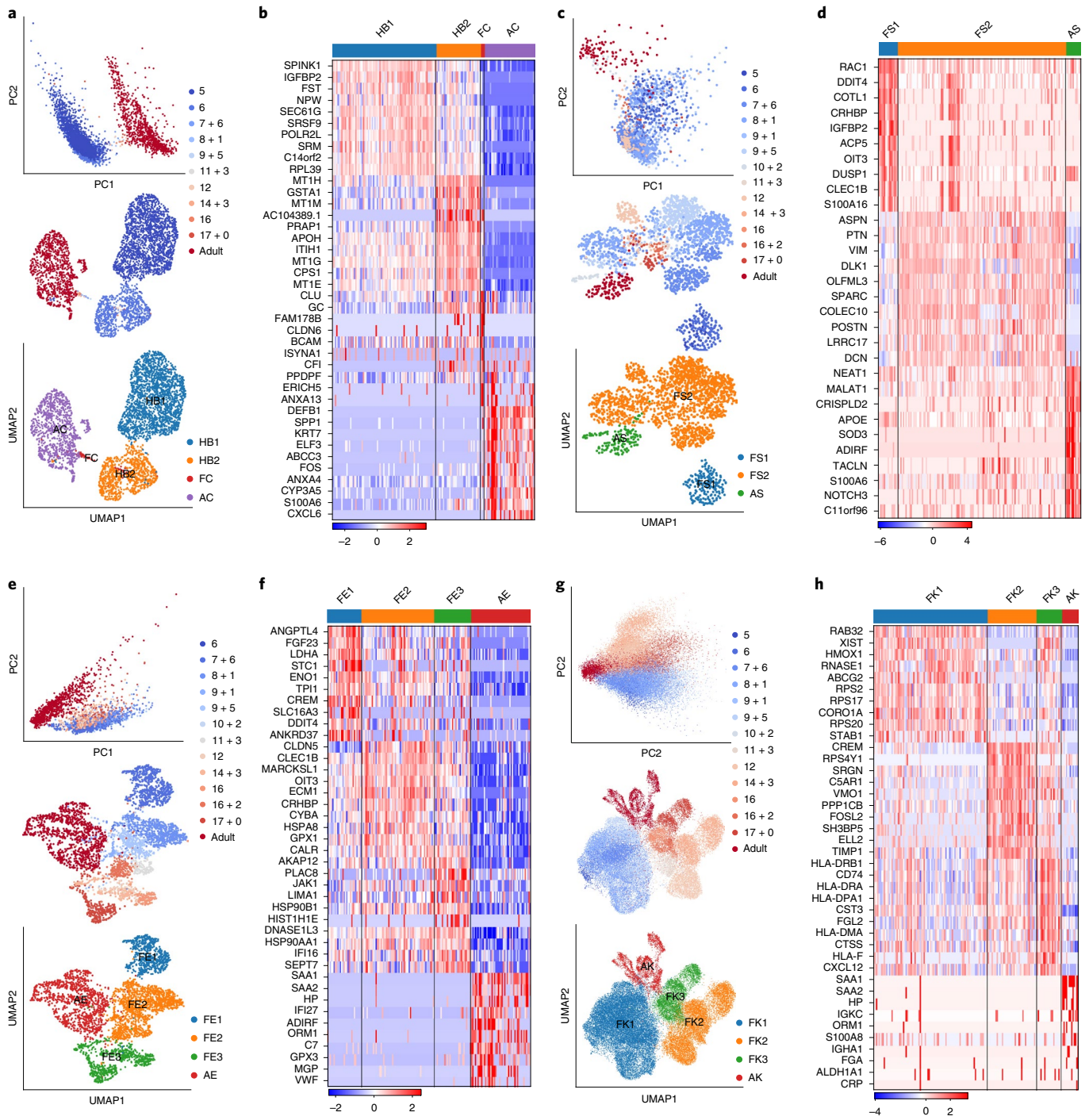


Fig. 2 | Mapping non-parenchymal cell identity during human liver development. a, PCA (top) and UMAP (middle) plots of primary human cholangiocyte sample timepoints and UMAP annotation of discrete cholangiocyte developmental stages (bottom); HB1, hepatoblast 1; HB2, hepatoblast 2; FC, foetal cholangiocyte; AC, adult cholangiocyte. **b**, Heat map showing time-related DEGs of each stage of primary cholangiocyte development. **c**, PCA (top) and UMAP (middle) plots of primary human hepatic stellate cell sample timepoints and UMAP annotation of discrete stellate cell developmental stages (bottom); FS1, foetal stellate cell 1; FS2, foetal stellate cell 2; AS, adult stellate cell. These three developmental stages correlate with the onset of haematopoietic function of the liver and birth. **d**, Heat map showing time-related DEGs of each stage of primary hepatic stellate cell development. **e**, PCA (top) and UMAP (middle) plots of primary human endothelial cell sample timepoints and UMAP annotation of discrete endothelial cell developmental stages (bottom); FE1, foetal endothelial cell 1; FE2, foetal endothelial cell 2; FE3, foetal endothelial cell 3; AE, adult endothelial cell. **f**, Heat map showing time-related DEGs of each stage of primary endothelial cell development. Endothelial cells are closely associated with haematopoietic stem cell differentiation, with changes of function associated with haematopoietic and vascularization events. **g**, PCA (top) and UMAP (middle) plots of primary human Kupffer cell sample timepoints and UMAP annotation of discrete Kupffer cell developmental stages (bottom); FK1, foetal Kupffer cell 1; FK2, foetal Kupffer cell 2; FK3, foetal Kupffer cell 3; AK, adult Kupffer cell. **h**, Heat map showing time-related DEGs specific to each stage of primary Kupffer cell development. Heat map colour scales show 'gene expression [log-normalized, scaled counts]'. Plots integrate scRNA-seq data from $n = 17$ independent foetal livers aged 5–17 PCW and $n = 16$ independent adult livers.

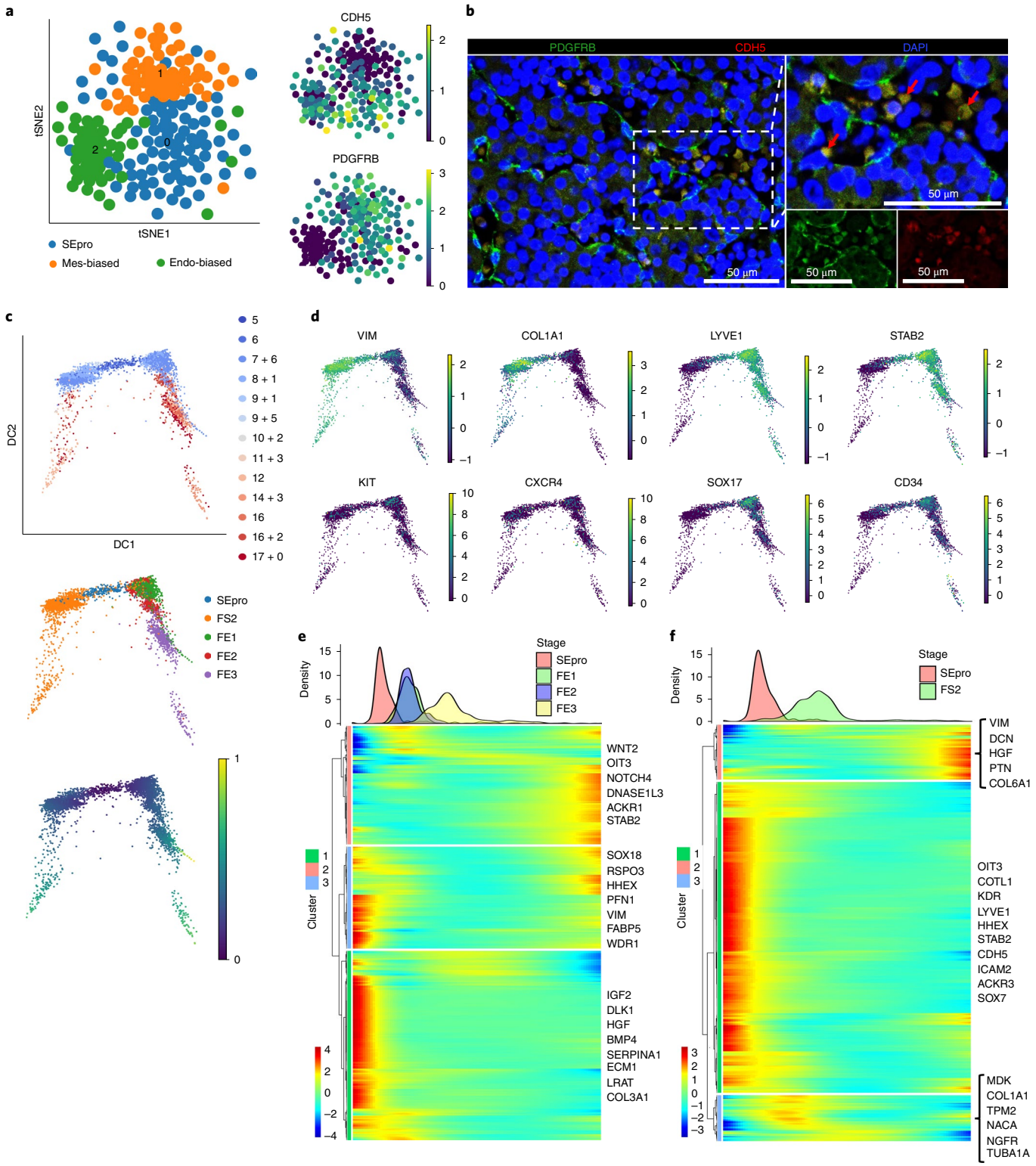
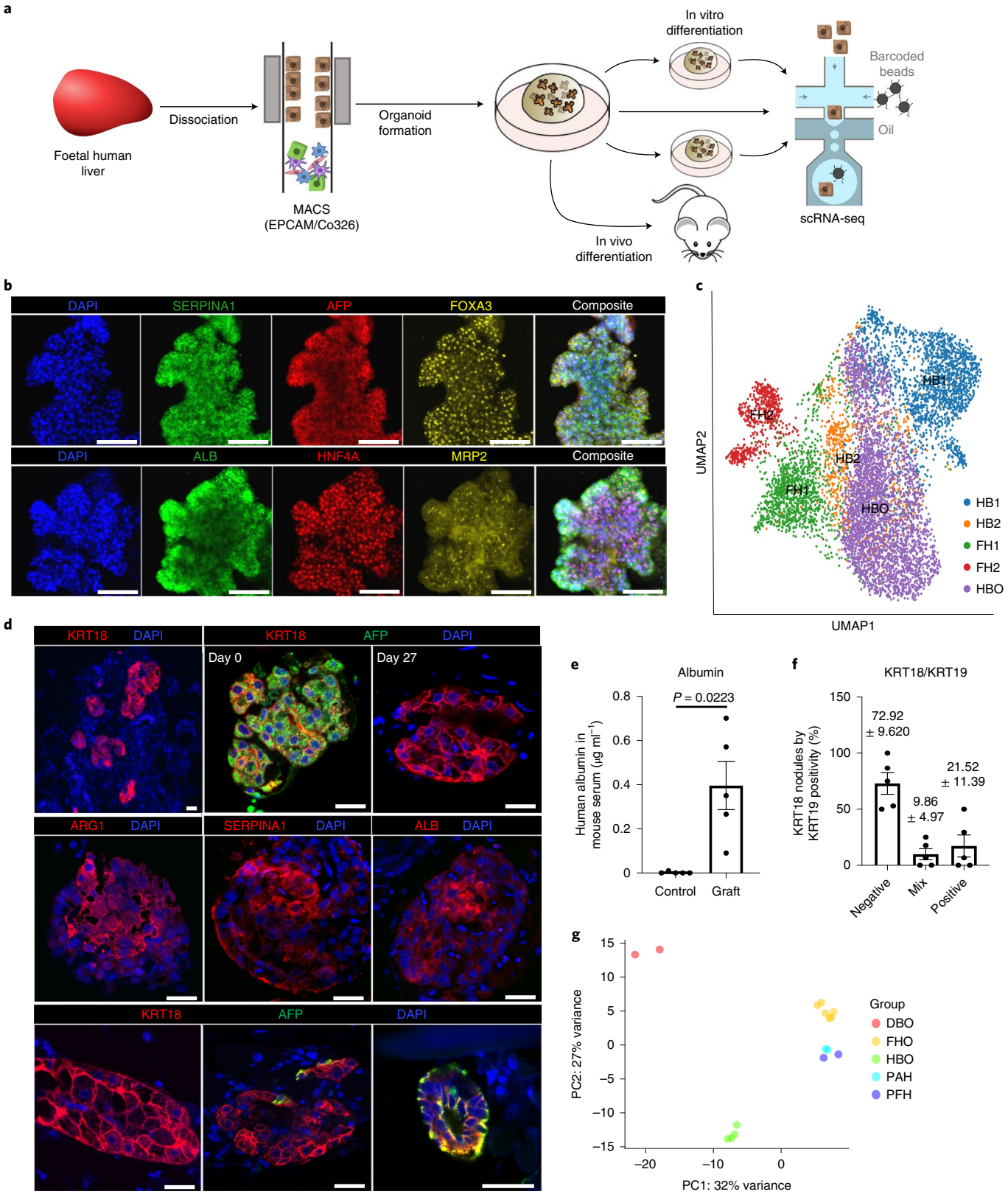


Fig. 3 | Identification of a hepatic stellate and endothelial cell progenitor in the early foetal liver. a, Left: *t*-distributed stochastic neighbour embedding (t-SNE) visualization based on Louvain clustering of 6 PCW human foetal liver cells identifying stellate–endothelial progenitors or 'SEpro'. Right: gene expression t-SNE plots showing the co-expression of specific markers for both hepatic stellate and endothelial lineages by SEpros ($n=3$ independent foetal livers). **b**, Immunofluorescence staining of 6 PCW human liver identifying the SEpro population on the basis of co-expression of stellate (PDGFRB) and endothelial (CDH5) markers; scale bars, 50 µm. **c**, DPT analyses of stellate and endothelial cell developmental trajectories showing that each lineage originated from SEpro (integrated scRNA-seq data from $n=17$ independent foetal livers aged 5–17 PCWs); DC, diffusion component. **d**, DPT analyses of specific markers for each lineage (top row stellate cells, bottom row endothelial cells). **e**,**f**, Heat map of time-related genes during foetal endothelial cell development (**e**) and foetal hepatic stellate cell development (**f**) starting with SEpro and progressing towards 17 PCWs. DPT pseudotime colour scale shows 'geodesic distance [distance between nodes]'; heat map colour scale shows 'gene expression [log-normalized, scaled counts]'.



was markedly decreased by day 27 (Fig. 4d), suggesting differentiation into hepatocytes in vivo. Accordingly, numerous cells in hepatic nodules stained positively for ARG1, A1AT and ALB while significant levels of human albumin were identified in mouse serum compared to control, suggesting functional activity of implanted organoids (Fig. 4e). Finally, some KRT18$^+$ nodules were found to contain cells expressing KRT19 (Fig. 4f), either as a mixed population or as a pure KRT19$^+$ population. Together, our results demonstrate that HB2 hepatoblasts can be grown in vitro while maintaining their capacity to differentiate into hepatocytes and cholangiocytes.

Fig. 4 | Modelling early hepatic development in vitro using HBOs. a, Schematic representation of HBO derivation and subsequent analyses. **b**, Immunostaining of hepatoblast markers in HBOs grown in vitro; scale bars, 100 µm. **c**, UMAP visualization of foetal hepatoblast/hepatocyte differentiation stages along with HBOs, confirming that HBOs share the transcriptional profile of the HB2 stage of hepatocyte development. **d**, Immunostaining showing decrease of the foetal hepatocyte marker (AFP) in HBOs after 27 days of engraftment while hepatocyte markers (KRT18, ALB, ARG1 and SERPINA1) were maintained, indicative of differentiation into mature hepatocytes. Immunostaining for biliary markers identified KRT19-positive cells in a subset of nodules, which organized into bile duct-like structures. Unless otherwise stated, pictures show grafts 27 days post-transplantation; scale bars, 20 µm. **e**, ELISA analyses showing secretion human ALB in the serum of HBO recipient mice 27 days after engraftment ($n=5$ independent animals). **f**, Quantification (percentage) of KRT19-positive cells within KRT18-positive nodules. **g**, PCA showing the divergence in gene expression profile between HBOs ($n=4$ lines derived from four independent foetal livers), DBO ($n=2$), foetal hepatocyte organoids (FHO; $n=6$), primary adult hepatocytes (PAH; $n=2$) and primary foetal liver (PFH; $n=2$). Data are presented as mean ± s.e.m.; unpaired two-tailed *t*-tests.

Of note, two types of human liver organoid system have been described previously by Huch et al. (2015)[39,40] and Hu et al. (2018)[36]. The former is composed of intra-hepatic cholangiocytes that can differentiate towards HLCs[39,41] (differentiated biliary organoids, DBO), while the latter derives organoids from hepatocytes[36]. Therefore, we characterized both systems against HBOs (Fig. 4g). Organoids derived from intra-hepatic cholangiocytes expressed markers such as KRT19, but did not express hepatocyte markers in either the undifferentiated or differentiated state (Extended Data Fig. 3i,j). Transcriptomic comparison also demonstrated the transcriptional divergence between HB2/HBO, DBO and hepatocyte organoids (Fig. 4g). Analysis of the genes driving this separation revealed hepatocyte and biliary markers, with HBOs having intermediate levels of both these sets of markers (Extended Data Fig. 3i–l). Taken together, these data demonstrate that our single-cell analyses have identified a unique self-renewing population of hepatoblasts that can be propagated long-term in vitro.

Dynamic intercellular interactions of the developing liver

To further understand the mechanisms directing liver organogenesis, we captured the interactions between the hepatoblasts/hepatocytes and other liver cell types using the CellPhone database (CellPhoneDB)[42,43] (Fig. 5a and Extended Data Fig. 4). This approach revealed that most interactions began with hepatoblasts, stabilized in foetal hepatocytes and finally disappeared in the adult stage (Extended Data Fig. 4a). Furthermore, a diversity of unknown interactions was captured between stellate cells, Kupffer cells and endothelial cells, indicating potential roles in extracellular matrix organization, haematopoietic development and innate immunity (Extended Data Fig. 4a). Thus, our analysis could reveal the source of signalling pathways controlling liver development. Several of these interactions were validated using RNAScope on primary liver tissues (Extended Data Fig. 4b–d). For example, NOTCH4/DLL4 was expressed by endothelial cells and interacted with DLK1/NOTCH2 on hepatoblasts/hepatocytes (Fig. 5b and Extended Data Fig. 4c). These bidirectional interactions suggest that hepatoblasts may be involved in the vascularization of the liver and thus direct the construction of their own niche. In return, endothelial cells could control hepatoblast differentiation into cholangiocytes, a process known to require NOTCH signalling[44]. Similarly, RSPO3–LGR4/5 interactions were detected between hepatoblasts and stellate cells at 5–6 PCWs (Fig. 5a,b); thus, stellate cells could support hepatoblast self-renewal by boosting WNT signalling.

We then decided to test if these analyses could also be used to identify pathways controlling hepatocyte and cholangiocyte differentiation. To induce hepatocyte differentiation, we selected five growth factors indicated by our CellphoneDB analyses: vascular endothelial growth factor A (VEGFA), complement C3 (C3), neuregulin 1 (NRG1) erythropoietin (EPO) and oncostatin M (OSM) (Fig. 5a and Extended Data Fig. 4). The effect of these factors was then analysed on HBOs grown in the absence of WNT to enable their differentiation. Several factors (EPO, OSM and VEGFA) were associated with a decrease in hepatoblast markers (AFP, LGR5 and MKI67) and an increase in hepatocyte markers (albumin and G6PC) (Figs. 5c and 6a), suggesting a differentiation towards foetal hepatocytes. Other factors either blocked the expression of hepatocyte markers (NRG1) or had limited effect (C3). To further reinforce these observations, we decided to characterize the effect of OSM by performing scRNA-seq on HBOs induced to differentiate into hepatocytes. These analyses confirmed the transcriptional shift of HBO towards hepatocyte after treatment with OSM (Fig. 6b,c). Furthermore, this differentiation was associated with gain of hepatocyte functions including cytochrome P450 activity (Fig. 6d) and lipid accumulation (Fig. 6e). Together, these data confirm that HBOs can differentiate into hepatocytes in the absence of WNT upon stimulation of specific factors such as OSM.

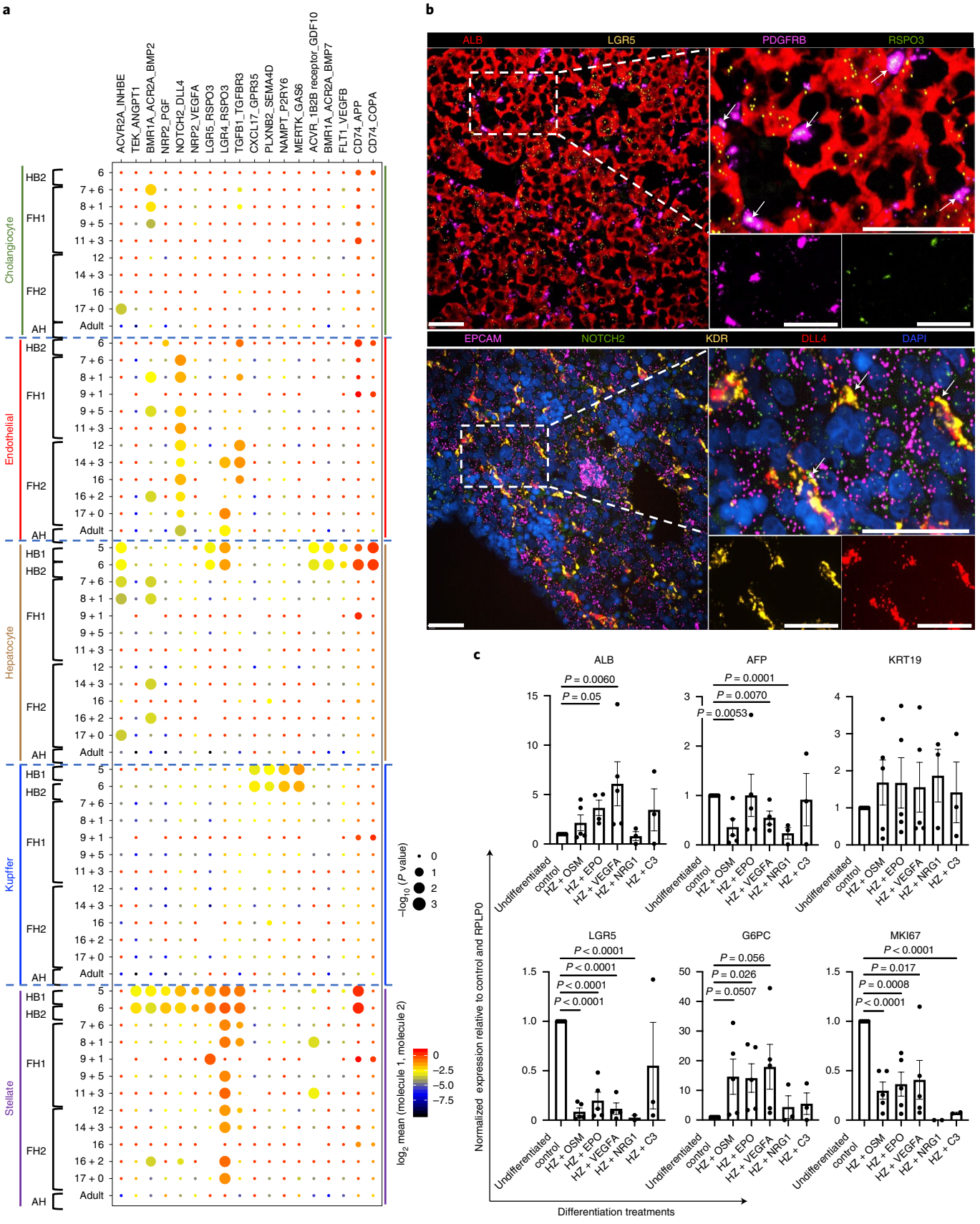
Focusing next on cholangiocyte differentiation, CellphoneDB analyses reinforced previous reports[4,5] suggesting an important function for TGFB in this process[45] (Fig. 6a). To test this hypothesis, we supplemented HBO medium with TGFB for 7 days and observed a marked switch towards cholangiocyte identity illustrated by the induction of biliary markers KRT19 and loss of hepatoblast markers (Fig. 6d,f–h). These observations were confirmed by scRNA-seq analyses showing that the transcriptome of HBOs grown in the presence of TGFB resemble that of cholangiocyte organoids (Fig. 6d,i,j) Taken together, these data confirm the potential of our single-cell analyses to identify cell–cell interactions directing liver development and also the potential of HBOs for validating the function of the signalling pathways involved.

Liver atlas informs the maturation of hiPSC derivatives

To further exploit our single-cell map and the data generated above, we addressed the challenge of cellular maturation associated with hPSC[46] differentiation. It is well established that most protocols currently available to differentiate hPSCs result in cells

Fig. 5 | Cell-to-cell interaction networks during human liver development. a, CellphoneDB analysis of the receptor–ligand interactions of hepatocytes with other hepatic cells across all developmental timepoints. *y* axis shows ligand–receptor/receptor–ligand interactions, with the hepatocyte protein listed first in each pairing; *x* axis shows developmental timeline of each cell type; dot colour indicates $\log_2$ mean expression of interacting molecules and dot size shows $-\log_{10}P$ values (integrated scRNA-seq data from $n=17$ independent foetal livers ranging in age from 5 to 17 PCWs; $n=16$ independent adult livers). **b**, RNAScope validating ligand–receptor interactions that establish the hepatoblast niche in 6 PCW liver. Top: RSPO3 is expressed in hepatic stellate cells, and its LGR5 receptor is expressed by hepatoblasts. Bottom: DLL4 is expressed by endothelial cells, while NOTCH2 receptor is expressed on hepatoblasts. Scale bars, 50 µm. **c**, qPCR showing the expression of hepatocyte maturation genes following treatment with key signalling molecules discovered using the single-cell liver development atlas ($n=5$ independent experimental replicates); Undiff. control, HBOs grown in upkeep culture conditions to maintain their self-renewal capacity; HZ, HepatoZYME basal medium. Data are presented as mean ± s.e.m.; unpaired two-tailed *t*-tests.

with foetal characteristics[47] rather than fully functional, adult-like cells. Accordingly, single-cell analyses and detailed characterizations have demonstrated the foetal identity of hPSC-derived hepatocytes[48,49]; however, the mechanisms blocking progress towards an adult phenotype remain unclear. To address this question, we performed scRNA-seq on hPSCs differentiating into HLCs



(Fig. 7a). PCA analyses confirmed the progressive process driving the acquisition of a hepatocytic identity (Fig. 7b,c). This differentiation trajectory was then compared with the developmental trajectory of primary hepatoblasts/hepatocytes by DPT alignment (Fig. 7d), UMAP, PAGA analyses and PCA (Extended Data Fig. 5a,b). These comparisons showed that HLCs at day 14 of differentiation aligned to the second hepatoblast stage, after which their differentiation follows an in vitro specific process. Differential gene expression analyses yielded a list of genes related to xenobiotic metabolism, bile acid transport and lipid metabolism pathways, which suggests that the divergence between HLCs and primary cells prevents the acquisition of fully adult function. Importantly, a similar divergence

Fig. 6 | Characterization of the differentiation capacity of HBOs into both hepatocyte and cholangiocyte mature lineages. a, Immunostaining showing that HBOs differentiated into hepatocytes (HBO + HZ) maintain expression of ALB while losing the foetal marker AFP; scale bars, 100 µm. **b**, PCA showing that HBO differentiation in vitro follows the developmental trajectory of foetal primary hepatocyte development. **c**, Violin plots of key functional markers corresponding to the acquisition of an adult hepatocyte phenotype after HBO differentiation. **d**, Cytochrome P450 3A5/7 and cytochrome P450 3A4 activity in HBO, HBO + HZ, and HBO treated with TGFB (HBO + TGF) ($n = 6$ independent experimental replicates using lines derived from two independent foetal livers); RLU, relative light unit. **e**, BODIPY assay showing differences in lipid uptake in HBO compared with HBO + HZ and primary adult hepatocytes (PAH). **f**, Immunocytochemistry of HBO, HBO + TGF, and BO stained for KRT19 and ASGR1; scale bars, 100 µm. **g**, qPCR analyses showing the expression of denoted genes in HBO and HBO + TGF ($n = 5$, each point represents an HBO line derived from a unique primary foetal liver). **h**, ELISA analyses showing the concentration of protein in medium secreted by HBOs ($n = 8$, each line derived from an independent foetal liver), and HBOs treated with TGFB ($n = 3$) after 48 h of freshly applied medium. Values are normalized to cell number (that is, per million cells) with albumin as micrograms per litre, and α-fetoprotein as units per millilitre. **i**, PCA plot of scRNA-seq data comparing the in vitro differentiation of HBOs towards cholangiocytes (HBO + TGF) to adult BOs and in vivo differentiation of hepatoblasts. **j**, scRNA-seq violin plots showing the loss of hepatocyte functional genes and the acquisition of a biliary transcriptome, thus demonstrating the similarity of HBO + TGF cholangiocytes to the positive BO control. Data are presented as mean ± s.e.m.; unpaired two-tailed *t*-tests.

from primary development was observed in other cell types generated from hPSCs including cholangiocytes[50], endothelial cells[32,51], stellate cells[33] and macrophages[52] (Extended Data Fig. 5c–f). Thus, hPSC differentiation could systematically deviate from a natural developmental path after embryonic stages, preventing the production of adult cells.

Comparison of in vivo to in vitro hepatocyte differentiation also revealed transcription factors expressed in foetal hepatocytes (FH1) that were missing in hPSC-derived cells (Extended Data Fig. 6a–c). To validate their functional relevance, these factors were overexpressed in hPSCs differentiated into HLCs for 15 days. The phenotype of the resulting cells was assayed by scRNA-seq 8 days after transduction (Extended Data Fig. 6d). Of particular interest, NFIX or NFIA expression changed the transcriptional profile of HLCs (Fig. 7e,f). This shift was characterized by a decrease in foetal markers and the induction of markers indicative of the adult state (ALB, HP and C3; Fig. 7e,f). Furthermore, pathway enrichment analyses showed an increase in specific functions associated with adult hepatocyte identity, including metabolic and complement-related pathways (Extended Data Fig. 6e). Thus, NFIX or NFIA overexpression during differentiation of human induced pluripotent stem cells (hiPSCs) appears to increase the expression of specific markers probably downstream of these transcription factors. These results were further validated by inducing the expression of NFIX, NFIA and CAR during differentiation of hPSCs towards HLCs using the Opti-OX system[53,54]. Stable induction of NFIX, and to a lesser extent NFIA, upregulated an array of functional markers including ALB, SAA1, SAA2, LRP1, CES2 and AOX1 while downregulating AFP (Extended Data Fig. 7). Taken together, these results show that misexpression of key developmental regulators during HLC differentiation could explain their limited capacity to become adult hepatocytes, and that expression of these factors at an early step of differentiation could augment their similarity to adult cells.

Discussion

Our study provides a detailed map of human liver development and shows that the developmental trajectory of liver cells is influenced by changes in liver environment. Colonization by the haematopoietic system[55] appears to initiate the functional maturation of hepatocytes, which progressively acquire intrinsic hepatic functions. Of note, the activation of the liver after birth[24,56] represents another key change allowing the full maturation of cells. However, this stage could not be included in our analyses since collection of neonatal tissue in human are extremely rare and ethically problematic. Nonetheless, our results established that the step-by-step differentiation of hepatocytes and cholangiocytes originates from a crosstalk among parenchymal and non-parenchymal cells. Of particular interest, stellate cells appear to have an underestimated importance in supporting hepatoblast self-renewal while building the hepatic niche. Such key developmental roles could be shared by many tissue-specific fibroblasts involved in organ fibrosis[57,58] during chronic diseases. In addition, hepatoblasts/hepatocytes are likely to also influence the surrounding cells to establish their own niche and to direct their functional maturation through interactive feedback loops. Utilization of this knowledge has enabled us to develop a culture system to grow hepatoblasts in vitro that provide a promising model system not only to study liver organogenesis, but also to produce cells for clinical applications. Nonetheless, it is important to note that further investigations are necessary to demonstrate the capacity of HBOs to differentiate into fully functional hepatocytes and cholangiocytes after clonal isolation. Finally, our developmental map has revealed that differentiation of hPSCs diverge from a natural path of development at an early stage and then follow an in vitro-specific process. This divergence explains the foetal nature of cells generated by current protocols[59–64] and suggests that improving the intermediate, specification steps of these protocols may be necessary to generate adult cells. Thus, understanding organ development remains the best approach for generating fully functional cell types in vitro. Our study illustrates how single-cell analyses can be combined with in vitro models to uncover the mechanisms driving the generation of functional hepatic cells in vivo and thus paves the way towards identifying factors for improving differentiation in vitro. This methodology and the resulting knowledge are likely to be transferable to other organs and will be useful for generating a diversity of cell types for disease modelling and cell-based therapies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-022-00989-7.

Received: 13 October 2020; Accepted: 29 July 2022;
Published online: 15 September 2022

References

1. Bilzer, M., Roggel, F. & Gerbes, A. L. Role of Kupffer cells in host defense and liver disease. *Liver Int.* **26**, 1175–1186 (2006).
2. Asahina, K. et al. Mesenchymal origin of hepatic stellate cells, submesothelial cells, and perivascular mesenchymal cells during mouse liver development. *Hepatology* **49**, 998–1011 (2009).
3. Maroni, L. et al. Functional and structural features of cholangiocytes in health and disease. *Cell. Mol. Gastroenterol. Hepatol.* **1**, 368–380 (2015).
4. Raven, A. et al. Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration. *Nature* **547**, 350–354 (2017).
5. Lu, W.-Y. et al. Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat. Cell Biol.* **17**, 971–983 (2015).

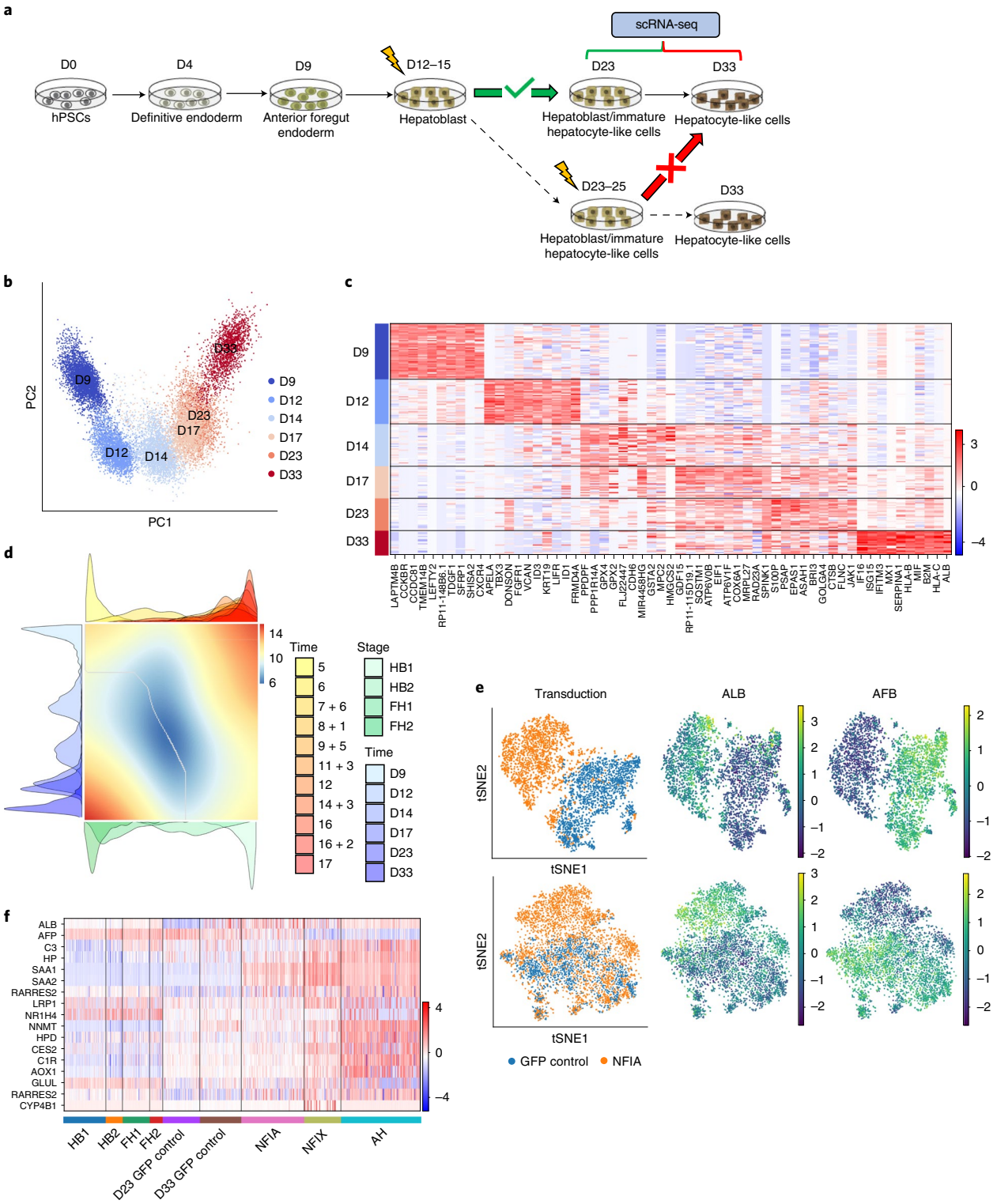


Fig. 7 | Temporal overexpression of key transcription factors in hPSC-derived hepatocytes increases their similarity to adult primary hepatocytes. **a**, Schematic representation of experimental processes to validate the functional transcription factors in hepatocyte differentiation. **b**, PCA showing the step-by-step differentiation of hPSCs into hepatocytes. D, day of differentiation ($n = 6$ sequential differentiation timepoints, with one replicate sequenced per timepoint). **c**, Heat map of top ten DEGs specific between each stage of differentiation; Wilcoxon rank-sum test, z score > 10. **d**, Alignment of primary hepatocyte developmental trajectory to hiPSC differentiation using the CellAlign software; red colour shows regions of misalignment/dissimilarity, and blue colour shows regions of close alignment/similarity (integrated scRNA-seq data from $n = 17$ independent foetal livers ranging in age from 5 to 17 PCWs and $n = 16$ independent adult livers). **e**, UMAP visualization of HLCs transduced with transcription factors NFIX, NFIA and GFP (control) showing that TFs can increase ALB expression while decreasing the expression of the foetal marker AFP. **f**, Heat map showing the acquisition of functional hepatocytes markers in transduced hepatocytes derived from hPSCs ($n = 1$ sample sequenced per transduction). Heat map colour scales show 'gene expression [log-normalized, scaled counts]'.

6. Poisson, J. et al. Liver sinusoidal endothelial cells: Physiology and role in liver diseases. *J. Hepatol.* **66**, 212–227 (2017).
7. Aizarani, N. et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* https://doi.org/10.1038/s41586-019-1373-2 (2019).
8. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
9. Ramachandran, P. et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* https://doi.org/10.1038/s41586-019-1631-3 (2019).
10. Segal, J. M. et al. Single cell analysis of human foetal liver captures the transcriptional profile of hepatobiliary hybrid progenitors. *Nat. Commun.* **10**, 3350 (2019).
11. Wang, X. et al. Comparative analysis of cell lineage differentiation during hepatogenesis in humans and mice at the single-cell transcriptome level. *Cell Res.* https://doi.org/10.1038/s41422-020-0378-6 (2020).
12. Forbes, S. J. & Alison, M. R. Knocking on the door to successful hepatocyte transplantation. *Nat. Rev. Gastroenterol. Hepatol.* **11**, 277–278 (2014).
13. Kegel, V. et al. Protocol for isolation of primary human hepatocytes and corresponding major populations of non-parenchymal liver cells. *J. Vis. Exp.* https://doi.org/10.3791/53069 (2016).
14. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
15. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).
16. Gayoso, A. & Shor, J. DoubletDetection. *Zenodo* https://doi.org/10.5281/ZENODO.2658730 (2018).
17. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
18. Mcinnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw*. https://doi.org/10.21105/joss.00861 (2018).
19. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* https://doi.org/10.1186/s13059-017-1382-0 (2018)..
20. Van Der Maaten, L., Courville, A., Fergus, R. & Manning, C. Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
21. Godlewski, G., Gaubert-Cristol, R., Rouy, S. & Prudhomme, M. Liver development in the rat and in man during the embryonic period (Carnegie stages 11–23). *Microsc. Res. Tech.* **39**, 314–327 (1997).
22. Antoniou, A. et al. Intrahepatic bile ducts develop according to a new mode of tubulogenesis regulated by the transcription factor SOX9. *Gastroenterology* **136**, 2325–2333 (2009).
23. Si-Tayeb, K., Lemaigre, F. P. & Duncan, S. A. Organogenesis and development of the liver. *Dev. Cell* **18**, 175–189 (2010).
24. Collardeau-Frachon, S. & Scoazec, J.-Y. Vascular development and differentiation during human liver organogenesis. *Anat. Rec. Adv. Integr. Anat. Evol. Biol.* **291**, 614–627 (2008).
25. Asahina, K., Zhou, B., Pu, W. T. & Tsukamoto, H. Septum transversum-derived mesothelium gives rise to hepatic stellate cells and perivascular mesenchymal cells in developing mouse liver. *Hepatology* **53**, 983–995 (2011).
26. Loo, C. K. C. & Wu, X. J. Origin of stellate cells from submesothelial cells in a developing human liver. *Liver Int* **28**, 1437–1445 (2008).
27. Horsfall, D. & McGrath, J. Adifa software for single cell insights. *Zenodo* https://doi.org/10.5281/zenodo.5824896 (2022).
28. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 1–9 (2019).
29. Alpert, A., Moore, L. S., Dubovik, T. & Shen-Orr, S. S. Alignment of single-cell trajectories to compare cellular expression dynamics. *Nat. Methods* **15**, 267–270 (2018).
30. Pérez-Pomares, J. M. et al. Contribution of mesothelium-derived cells to liver sinusoids in avian embryos. *Dev. Dyn.* **229**, 465–474 (2004).
31. Lotto, J. et al. Single-cell transcriptomics reveals early emergence of liver parenchymal and non-parenchymal cell lineages. *Cell* **183**, 702–716.e14 (2020).
32. Patsch, C. et al. Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells. https://doi.org/10.1038/ncb3205 (2015).
33. Coll, M. et al. Generation of hepatic stellate cells from human pluripotent stem cells enables in vitro modeling of liver fibrosis.*Cell Stem Cell* **23**, 101–113.e7 (2018).
34. Prior, N. et al. Lgr5+ stem and progenitor cells reside at the apex of a heterogeneous embryonic hepatoblast pool. *Dev.* **146**, 174557 (2019).
35. Schmelzer, E., Wauthier, E. & Reid, L. M. The phenotypes of pluripotent human hepatic progenitors. *Stem Cells* **24**, 1852–1858 (2006).
36. Hu, H. et al. Long-term expansion of functional mouse and human hepatocytes as 3D organoids. *Cell* **175**, 1591–1606.e19 (2018).
37. Li, B. et al. Adult mouse liver contains two distinct populations of cholangiocytes. *Stem Cell Rep.* **9**, 478–489 (2017).
38. Stevens, K. R. et al. In situ expansion of engineered human liver tissue in a mouse model of chronic liver disease. *Sci. Transl. Med.* **9**, aah5505 (2017).
39. Huch, M. et al. Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299–312 (2015).
40. Huch, M. et al. In vitro expansion of single Lgr5^{+} liver stem cells induced by Wnt-driven regeneration. *Nature* **494**, 247–250 (2013).
41. Marsee, A. et al. Building consensus on definition and nomenclature of hepatic, pancreatic, and biliary organoids. *Cell Stem Cell* **28**, 816–832 (2021).
42. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).
43. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
44. Geisler, F. et al. Liver-specific inactivation of Notch2, but not Notch1, compromises intrahepatic bile duct development in mice. *Hepatology* **48**, 607–616 (2008).
45. Clotman, F. et al. Control of liver cell fate decision by a gradient of TGFβ signaling modulated by Onecut transcription factors. *Genes Dev.* **19**, 1849–1854 (2005).
46. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
47. Baxter, M. et al. Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes. *J. Hepatol.* **62**, 581–589 (2015).
48. Takebe, T. et al. Vascularized and complex organ buds from diverse tissues via mesenchymal cell-driven condensation. *Cell Stem Cell* **16**, 556–565 (2015).
49. Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533–538 (2017).
50. Sampaziotis, F. et al. Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nat. Biotechnol.* **33**, 845–852 (2015).
51. Challet Meylan, L., Challet Meylan, L., Patsch, C. & Thoma, E. Endothelial cells differentiation from hPSCs. *Nat. Protoc. Exch*. https://doi.org/10.1038/protex.2015.055 (2015).
52. Wilgenburg, B., van, Browne, C., Vowles, J. & Cowley, S. A. Efficient, long term production of monocyte-derived macrophages from human pluripotent stem cells under partly-defined and fully-defined conditions. *PLoS ONE* **8**, e71098 (2013).
53. Bertero, A. et al. Optimized inducible shRNA and CRISPR/Cas9 platforms for in vitro studies of human development using hPSCs. *Development* **143**, 4405–4418 (2016).
54. Pawlowski, M. et al. Inducible and deterministic forward programming of human pluripotent stem cells into neurons, skeletal myocytes, and oligodendrocytes. *Stem Cell Rep.* **8**, 803–812 (2017).
55. Popescu, D.-M. et al. Decoding human fetal liver haematopoiesis. *Nature* https://doi.org/10.1038/s41586-019-1652-y (2019).
56. Septer, S. et al. Yes-associated protein is involved in proliferation and differentiation during postnatal liver development. *Am. J. Physiol. Liver Physiol.* **302**, G493–G503 (2012).
57. Edeling, M., Ragi, G., Huang, S., Pavenstädt, H. & Susztak, K. Developmental signalling pathways in renal fibrosis: the roles of Notch, Wnt and Hedgehog. *Nat. Rev. Nephrol.* **12**, 426–439 (2016).
58. Martinez, F. J. et al. Idiopathic pulmonary fibrosis. *Nat. Rev. Dis. Prim.* **3**, 17074 (2017).
59. Hannan, N. R. F., Segeritz, C.-P., Touboul, T. & Vallier, L. Production of hepatocyte-like cells from human pluripotent stem cells. *Nat. Protoc.* **8**, 430–437 (2013).
60. Yusa, K. et al. Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391–394 (2011).
61. Touboul, T. et al. Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* **51**, 1754–1765 (2010).
62. Gieseck, R. L. et al. Maturation of induced pluripotent stem cell derived hepatocytes by 3D-culture. *PLoS One* **9**, 0086372 (2014).
63. Gieseck, R. L., Vallier, L. & Hannan, N. R. F. *Generation of Hepatocytes from Pluripotent Stem Cells for Drug Screening and Developmental Modeling* 123–142 (Humana Press, 2015).
64. Berger, D. R., Ware, B. R., Davidson, M. D., Allsup, S. R. & Khetani, S. R. Enhancing the functional maturity of induced pluripotent stem cell-derived human hepatocytes by controlled presentation of cell-cell interactions in vitro. *Hepatology* **61**, 1370–1381 (2015).

Methods

General statement on experimental design. Data collection and analysis were not performed blind to the conditions of the experiments. In addition, there was no randomization in the organization of the experimental conditions or stimulus presentation.

Adult human liver collection and dissociation. Liver samples were obtained under sterile conditions from deceased transplant organ donors as rapidly as possible after cessation of circulation. Tissue samples were transferred to the laboratory at 4 °C in University of Wisconsin organ preservation solution. Biopsy tissue was taken from the patient, placed in room temperature (RT) HepatoZYME-SFM medium and processed immediately. Protocols were developed from information in a previously published dissociation method[13]. The liver tissue was washed twice with warm DPBS with $Ca^{+2}Mg^{+2}$ + 0.5 mM EDTA, transferred into a Petri dish and diced into small pieces (roughly 1 cm^2 for resections and 0.25 cm^2 for biopsies) using a scalpel. The tissue was distributed evenly into multiple GentleMacs Tissue Dissociation C Tubes, or a single 1.5 ml Eppendorf tube for biopsies. 37 °C 0.2 Wünsch ml^{-1} Liberase enzymatic digestion solution reconstituted in HepatoZYME medium (without growth factors) containing DNAse I (2,000 U ml^{-1}) was added to each tube: 5 ml per tissue dissociator C tube and 1 ml per Eppendorf tube. The enzymatic digestion occurred in an incubating shaker at 37 °C and 200 RPM for 30 min. The partially degraded extracellular tissue matrix was mechanically dissociated by running two 'B' cycles using the 'C tube' in the Miltenyi Biotec GentleMACS tissue dissociator. A 1:1 ratio of 20% FBS to 80% DPBS was added to terminate the enzymatic reaction and the cell suspensions were filtered through 70 µm filters, with large pieces gently mashed through the filter. Cells dissociated from donor resections were centrifuged at 50*g*, 4 °C for 5 min to pellet the hepatocyte fraction, followed by centrifugation of the supernatant at 300*g*, 4 °C for 5 min and subsequently at 650*g*, 4 °C for 7 min to collect non-parenchymal cell fractions. The hepatocytes were pooled in 5 ml of DPBS, pipetted onto cold 25% percoll solution and centrifuged at 1,250*g*, 4 °C for 20 min without brake. The purified hepatocyte pellet was incubated for 10 min in 5 ml of red blood cell (RBC) lysis solution (Miltenyi Biotec, 130-094-183) and pelleted by centrifuging at 50*g*, 4 °C for 5 min, yielding a cleaned hepatocyte fraction that was resuspended in cold HepatoZYME-SFM for scRNA-seq. The remaining two non-parenchymal cell (NPC) fractions were pooled together in 3.1 ml of DPBS, to which 900 µl of debris removal solution is added. The Miltenyi Biotec Debris Removal (Miltenyi Biotec, 130-109-398) protocol was followed according to the manufacturer's guidelines to yield a clean cell pellet. The NPC pellets were incubated for 7–10 min in 2 ml of RBC lysis buffer at RT and subsequently pelleted by centrifuging at 650*g*, 4 °C for 5 min, yielding a clean non-parenchymal cell fraction that was resuspended in cold HepatoZYME-SFM. Non-parenchymal cell types were isolated from liver biopsies by first centrifuging the total cell suspension at 400*g*, 4 °C for 5 min and incubating for 10 min in RBC lysis solution at RT. Debris removal solution was used to clean remaining debris and dead cells according to the manufacturer's instructions. The clean pelleted cells were resuspended in cold HepatoZYME-SFM medium for analyses.

Foetal human liver collection and dissociation. Primary human foetal tissue was obtained from patients undergoing elective terminations (ethical approval obtained from East of England—Cambridge Central Research Ethics Committee REC-96/085). The liver was dissected from the abdominal cavity and placed into a solution containing Hanks' buffered saline solution (HBSS) supplemented with 1.07 Wünsch units ml^{-1} Liberase DH (Roche Applied Science) and 70 U ml^{-1} hyaluronidase (Sigma-Aldrich), and placed on a microplate shaker at 37 °C, 750 RPM, for 15 min. The sample was subsequently washed three times in HBSS using centrifugation at 400*g* for 5 min each. The single-cell suspension was then sorted for EPCAM/CD326-positive cells using CD326 microbeads (Miltenyi Biotec, 130-061-101), according to the manufacturer's guidelines.

Establishment of HBOs. The single-cell suspension was resuspended in hepatoblast organoid medium (HBO-M); Advanced DMEM/F12 supplemented with HEPES, penicillin–streptomycin and GlutaMAX, 2% B27, 20 mM nicotinamide, 2 mM *N*-acetylcysteine, 50% WNT3A conditioned medium, 10% R-Spondin, 50 ng ml^{-1} EGF and 50 µM A83-01. For long-term culture (more than four passages) 10 µM Y27632 was required. To the resuspended cells was added a volume of Growth Factor Reduced Phenol Free Matrigel (Corning) to make the final solution up to 55% Matrigel by volume, and the mixture pipetted into 48-well plates (20 µl per well). The plates were placed at 37 °C for 15 min to allow the mixture to set, and subsequently 200 µl of fresh HBO medium applied to each well.

Establishment of BOs and differentiation. Biliary organoids (BOs) were derived from samples taken from the livers of adult human deceased donors (National Research Ethics Committee East of England—Cambridge South 15/EE/0152). These were maintained and differentiated as per the author's guidelines[39,40].

HBO differentiation into hepatocytes. Hepatocyte differentiation medium was made using complete HepatoZYME (Life Technologies, 17705-021). This medium comprised basal HepatoZYME supplemented with non-essential amino acids (Thermo Fisher, 11140050), chemically defined lipid concentrate (Thermo Fisher, 11905031), L-glutamine (2 mM), insulin (14 ng ml^{-1}) and transferrin (10 ng ml^{-1}). To this medium was added either OSM (Sigma Aldrich) at a concentration of 20 ng ml^{-1}, EPO (R&D systems) at a concentration of 50 U ml^{-1} or VEGFA (R&D Systems) at a concentration of 50 ng ml^{-1}. This supplemented HepatoZYME was applied to HBOs for 7 days, applying fresh medium every 48 h.

HBO differentiation into cholangiocytes. HBO maintenance medium was supplemented with TGF-β (2 ng ml^{-1}) and applied to established HBO lines for a total of 7 days, applying fresh medium every 48 h.

Cytochrome P450 enzyme activity. Cytochrome P450 enzymatic activity was assessed using Promega P450-Glo assay systems using Luciferin–IPA for and Luciferin–PFBE as surrogate markers for cytochrome P450 3A4 and cytochrome p450 3A5/7, respectively. Each substrate was diluted 1:1,000 in freshly applied medium as per the manufacturer's guidelines and placed on organoids in three-dimensional culture. After incubation at 37 °C for 4 h, the medium was collected and 50 µl placed into a detection plate with 50 µl detection reagent. The solution was then left for 20 min before being read via a luminometer. Readings were adjusted for the average of three control readings, with the control consisting of the medium and reagents that was kept at 37 °C for 4 h without contact with cells.

Medium protein analysis. Albumin, α-fetoprotein, apolipoprotein-B and α-1-antitrypsin were detected in medium by ELISA (performed by core biomedical assay laboratory, Cambridge University Hospitals). Concentrations were normalized to cell number.

Murine blood protein analysis. Blood was drawn retro-orbitally for human albumin ELISA (Bethyl Laboratories) immediately before mice were killed at the termination of the experiment (27 days). Serum was separated by centrifugation, and levels of human albumin were determined by ELISA using goat polyclonal capture and horseradish peroxidase (HRP)-conjugated goat anti-human albumin detection antibodies (Bethyl Laboratories). Non-implanted FNRG ($Fah^{-/-}$, $Rag2^{-/-}$, $Il2rg^{-/-}$, on a NOD background) mouse blood serum was included as a negative control.

Implantation of HBOs and induction of liver injury in mice. All surgical procedures were conducted according to protocol 4388-01 approved by the University of Washington Institutional Animal Care and Use Committees. Seven female FRGN mice aged 14–18 weeks were used for implant procedures. Fourteen- to 18-week-old female FNRG mice (Yecuris) were administered sustained-release buprenorphrine and anaesthetized with isofluorane. Three organoid tissues were sutured onto the inguinal fat pads of each mouse. Three mice received organoid tissues with Donor 4 hepatoblasts, and two mice received organoid tissues with Donor 5 hepatoblasts. Incisions were closed aseptically. Nitisinone was withdrawn from animals' drinking water immediately after implantation of organoid tissues and for 14 days after implantation to induce liver injury. Nitisinone was then re-introduced to the drinking water to allow for recovery, and then removed again after 4 days for the remainder of the experiment. Animals were killed 27 days after implantation of organoid tissues. No statistical methods were used to pre-determine sample sizes as this was not relevant for our study; our sample sizes are similar to those reported in previous publications[38].

Immunostaining of collected constructs from mice. Implants were collected and fixed in 4% paraformaldehyde for 48 h at 4 °C. Excess fat was trimmed off of the implants, which were then dehydrated in graded ethanol (50–100%), embedded in paraffin and sectioned using a microtome (6 mm). Some sections were histochemically stained with haematoxylin and eosin. For immunostaining, sections were blocked with normal donkey serum and incubated with primary antibodies against human (for antibodies, see Supplementary Tables). To semi-quantify KRT19 distribution in nodules, graft nodules in which all cells were $KRT18^+/KRT19^+$ were tallied as '+'. Nodules with both $KRT18^+/KRT19^+$ and $KRT18^+/KRT19^-$ cells were tallied as '+/−'. Nodules with only $KRT18^+/KRT19^-$ cells were tallied as '−'. Nodules in each category were summed across each tissue section and divided by total $KRT18^+$ grafts in the section to acquire percentages in each animal, with each data point representing one animal.

Fluorescent in situ mRNA hybridization. Primary foetal human liver tissue was fixed in 10% formalin and embedded in paraffin, sectioned on to slides and stored for smFISH. Formalin-fixed paraffin-embedded slides were baked at 65 °C for 1 h. Slides were deparaffinized with two washes for 10 min each in xylene solution (Bond Dewax Solution, Leica AR9222) and two washes for 5 min each in PBS and air dried.

Multiplex smFISH was performed on a Leica BondRX fully automated stainer, using RNAScope Multiplex Fluorescent V2 technology (Advanced Cell Diagnostics 322000). Slides underwent heat-induced epitope retrieval with Epitope Retrieval Solution 2 (pH 9.0, Leica AR9640) at 95 °C for 5 min. Slides were then incubated in RNAScope Protease III reagent (ACD 322340) at 42 °C for 15 min, before being treated with RNAScope Hydrogen Peroxide (ACD 322330) for 10 min at RT to inactivate endogenous peroxidases.

All double-Z mRNA probes were designed by ACD for RNAScope on Leica Automated Systems. Slides were incubated in RNAScope 2.5 LS probes (designed against human genes *RSPO3*, *LGR5*, *ALB*, *PDGFRB*, *LRP5*, *DES*, *DKK1*, *EPCAM*, *NOTCH2*, *KDR*, *DLL4*, *WNT2B* and *WNT4)* for 2 h at RT. DNA amplification trees were built through consecutive incubations in AMP1 (pre-amplifier, ACD 323101), AMP2 (background reduction, ACD 323102) and AMP3 (amplifier, ACD 323103) reagents for 15–30 min each at 42 °C. Slides were washed in LS Rinse buffer (ACD 320058) between incubations.

After amplification, probe channels were detected sequentially via HRP–TSA labelling. To develop the C1–C3 probe signals, samples were incubated in channel-specific HRP reagents for 30 min, TSA fluorophores for 30 min and HRP-blocking reagent for 15 min at 42 °C. The probes in C1, C2 and C3 channels were labelled using Opal 520 (Akoya FP1487001KT), Opal 570 (Akoya FP1488001KT) and Opal 650 (Akoya FP1496001KT) fluorophores (diluted 1:500), respectively. The C4 probe complexes were first incubated with TSA–biotin (Akoya NEL700A001KT, 1:250) for 30 min at RT, followed by streptavidin-conjugated Atto425 (Sigma 56759, 1:400) for 30 min at RT. Samples were then incubated in DAPI (Sigma, 0.25 μg ml^{-1}) for 20 min at RT, to mark cell nuclei. Slides were briefly air dried and manually mounted using ~90 μl of Prolong Diamond Antifade (Fisher Scientific) and standard coverslips (24 × 50 mm^2; Fisher Scientific). Slides were dried at RT for 24 h before storage at 4 °C for >24 h before imaging.

Single-molecule fluorescence in situ hybridization (smFISH)-stained foetal liver slides were imaged on an Operetta CLS high-content screening microscope (Perkin Elmer). Image acquisition was controlled using Perkin Elmer's Harmony software. High-resolution smFISH images were acquired in confocal mode using an sCMOS camera and ×40 NA 1.1 automated water-dispensing objective. Each field and channel were imaged with a *z*-stack of 20 planes with a 1 μm step size between planes. All appropriate fields of the tissue section were manually selected and imaged with an 8% overlap.

Differentiation of hPSCs into HLCs. The differentiation of hiPSCs to HLCs followed the protocol previously published by the Vallier lab[59,60]. Before differentiation, 12-well plates were coated with 0.1% gelatin for an hour at 37 °C followed by MEF medium overnight at 37 °C and washed with DPBS before use. The hiPSCs were passaged at 70–90% confluency by incubating at 37 °C for 4 min with accutase cell dissociation reagent (Thermo Fisher, A1110501). The cell suspension was diluted at a 1:1 ratio with complete E8 medium (Gibco A1517001), pipetted gently to mechanically break any clumps into single cells, and pelleted by centrifuging at ~350*g* for 3 min. The hiPSCs were plated at a density of 50,000–60,000 cells cm^{-2} in 12-well plates in complete E8 medium supplemented with Y27632 ROCK inhibitor (10 μM) on 0.1% gelatin/MEF-coated plates. The medium was changed the following day to complete E8 without ROCK inhibitor. On day 1 of differentiation (2 days after plating), CDM–PVA medium supplemented with activin (100 ng ml^{-1}), FGF2 (80 ng ml^{-1}), BMP4 (10 ng ml^{-1}), LY294002 (10 μM) and CHIR99021 (3 μM) was added to the cells to induce endoderm formation (for all complete medium compositions, see Supplementary Tables). CHIR99021 was removed from this medium on day 2 of differentiation. On day 3, RPMI-1640 with non-essential amino acids (Thermo Fisher, 11140050) and B27 supplement (Thermo Fisher, 17504044) was supplemented with activin (100 ng ml^{-1}) and FGF2 (80 ng ml^{-1}). Foregut differentiation was initiated on day 4 and carried out until day 8 by changing medium to complete RPMI medium supplemented with activin (50 ng ml^{-1}). The hepatoblast and subsequent hepatocyte phenotype was induced by changing medium to HepatoZYME-SFM (Life Technologies, 17705-021) supplemented with non-essential amino acids (Thermo Fisher, 11140050), chemically defined lipid concentrate (Thermo Fisher, 11905031), L-glutamine (2 mM), insulin (14 ng ml^{-1}), transferrin (10 ng ml^{-1}), OSM (20 ng ml^{-1}) and hepatocyte growth factor (50 ng ml^{-1}) from day 9 to day 33.

Differentiation of hPSCs into CLCs. Differentiation of hPSCs to cholangiocyte-like cells (CLCs) followed the protocol previously published in our lab[50]. Before differentiation, 12-well plates were coated with 0.1% gelatin for 1 h at 37 °C followed by MEF medium overnight at 37 °C and washed with DPBS before use. The hiPSCs were passaged at 70–90% confluency by incubating at 37 °C for 4 min with accutase cell dissociation reagent (Thermo Fisher, A1110501). The cell suspension was diluted at a 1:1 ratio with complete E8 medium (Gibco A1517001), pipetted gently to mechanically break any clumps into single cells, and pelleted by centrifuging at ~350*g* for 3 min. The hPSCs were plated in complete E8 medium supplemented with ROCK inhibitor Y27632 (10 μM) at a density of 50,000–60,000 cells cm^{-2} on 0.1% gelatin/MEF-coated plates. The medium was changed the following day to E8 without ROCK inhibitor, and 2 days after plating, the differentiation was started by changing the medium to CDM–PVA supplemented with activin (100 ng ml^{-1}), FGF2 (80 ng ml^{-1}), BMP4 (10 ng ml^{-1}), LY294002 (10 μM) and CHIR99021 (3 μM) to induce endoderm formation. The same medium was used the following day (day 2) without CHIR99021. On day 3, RPMI-1640 with non-essential amino acids and B27 supplement (Thermo Fisher, 17504044) (RPMI + medium) was supplemented with activin (100 ng ml^{-1}) and FGF2 (80 ng ml^{-1}) only. Hepatoblasts were induced from day 9 to day 12 using RPMI + medium supplemented with SB (10 μM) and BMP4 (50 ng ml^{-1}). This bipotent progenitor was directed towards the cholangiocyte lineage from day 13 to day 16 by feeding with RPMI + medium supplemented with retinoic acid (3 μM), FGF10 (50 ng ml^{-1}) and activin (50 ng ml^{-1}). The mature cholangiocyte phenotype was induced by re-plating the cells in 3D culture using a 1:2 ratio of cell suspension to Matrigel. The cells matured from day 17 to day 26 in 3D culture with common bile duct medium supplemented with EGF (50 ng ml^{-1}) and forskolin (10 μM), yielding CLCs at day 26 of differentiation.

Differentiation of hPSC into HSLCs. This differentiation protocol was gathered from a previously published paper[33]. Before differentiation, 12-well plates were coated with a 1:50 ratio of reduced growth factor Matrigel and low-glucose DMEM medium overnight at 37 °C and washed with DPBS before use. The hiPSC were passaged at 70–90% confluency by incubating at 37 °C for 4 min with accutase cell dissociation reagent (Thermo Fisher, A1110501). The cell suspension was diluted at a 1:1 ratio with complete E8 medium (Gibco A1517001), pipetted gently to mechanically break any clumps into single cells, and pelleted by centrifuging at ~350*g* for 3 min. The hiPSCs were plated in complete E8 medium supplemented with Y27632 (10 μM) on 12-well plates coated with a 1:50 ratio of reduced growth factor Matrigel and low-glucose DMEM overnight at a concentration of 90,000 cells cm^{-2}. The medium was changed to E8 without ROCK inhibitor the following day. The cells were differentiated to mesoderm by adding DMEM–MCDB 201 medium supplemented with BMP4 (20 ng ml^{-1}) on day 1 and day 3 of differentiation. On day 5, a mesenchymal phenotype was induced by adding DMEM–MCDB 201 medium with BMP4 (20 ng ml^{-1}), FGF1 (20 ng ml^{-1}) and FGF3 (20 ng ml^{-1}). These cells transitioned to liver mesothelium by adding DMEM–MCDB 201 supplemented with retinoic acid (5 μM), palmitic acid (100 μM), FGF1 (20 ng ml^{-1}) and FGF3 (20 ng ml^{-1}) on day 7. From days 9 to 13, the cells were fed every 2 days with retinoic acid (5 μM) and palmitic acid (100 μM) to attain a foetal hepatic stellate-like cell (HSLC) phenotype.

Differentiation of hPSCs into ELCs. This protocol was adapted from previously published papers[32,51]. Before differentiation, 12-well plates were coated with 0.1% gelatin for 1 h at 37 °C followed by MEF medium overnight at 37 °C and washed with DPBS before use. The hiPSCs were passaged at 70–90% confluency by incubating at 37 °C for 4 min with accutase cell dissociation reagent (Thermo Fisher, A1110501). The cell suspension was diluted at a 1:1 ratio with complete E8 medium (Gibco, A1517001), pipetted gently to mechanically break any clumps into single cells, and pelleted by centrifuging at ~350*g* for 3 min. The hiPSCs were plated in E8 medium supplemented with Y27632 (10 μM) at a density of 45,000 cells cm^{-2} on 0.1% gelatin/MEF-coated plates. Differentiation was begun the follow day (day 1) by inducing mesoderm using CDM–PVA medium supplemented with FGF2 (20 ng ml^{-1}), BMP4 (10 ng ml^{-1}) and LY294002 (10 μM). On day 2.5, the cells were fed with Stempro-34 medium with VEGFA (200 ng ml^{-1}), forskolin (2 μM) and L-ascorbic acid (1 mM). The medium was changed every day until day 5.5 to yield foetal endothelial-like cells (ELCs).

Differentiation of hiPSCs into macrophage-like cells. This protocol was adapted from a previously published protocol[52]. The hiPSCs were passaged at 70–90% confluency and plated onto an ultra-low adherence 96-well plate with embryoid body (EB) medium consisting of E8 supplemented with Y27632 (10 μM), BMP4 (50 ng ml^{-1}), SCF (20 ng ml^{-1}) and VEGF (50 ng ml^{-1}). The cells were incubated for 4 days with half of the medium in each well being replaced with fresh medium after 2 days. On day 4, the EBs were transferred to a six-well plate coated with 0.1% gelatin in DPBS, and X-VIVO 15 media supplemented with GlutaMAX (2 mM), 2-mercaptoethanol (55 μM), M-CSF (100 ng ml^{-1}) and IL-3 (25 ng ml^{-1}) was added to the wells. Every 5 days, for roughly 10 days, two-thirds of the medium in each well was changed. On day 14, the EBs began production of macrophage progenitors. These floating progenitors were collected with the supernatant and plated on uncoated dishes in RPMI + 10% FBS supplemented with M-CSF (100 ng ml^{-1}) On day 7 after macrophage progenitor plating, the differentiated macrophages were collected for downstream analyses.

hIPSC-derived stellate–endothelial progenitor cells. Cells were plated and differentiated according to the ELC differentiation protocol described (Methods). At day 3.5 of this protocol, the SEpro cells were present in culture and collected for analyses. To test the bipotentiality of these cells, they were dissociated from the plate into single-cell suspensions by incubating with TryplE for 20 min at 37 °C, resuspending in culture medium and replating at a density of 100,000 cells cm^{-2}. These cells were subjected to the HSLC differentiation conditions beginning at day 7 of this protocol and continuing for at least 4 days to produce HSLCs.

Lentiviral transduction. Lentiviral aliquots were thawed on ice and added to the desired wells of D15 or D23 differentiating HLCs under sterile conditions. Polybrene was added to a final concentration of 10 μg ml^{-1} to each of the wells, and the plates were rocked gently to ensure even distribution of lentivirus on the adherent cells. The cells were incubated at 37 °C for 24 h to allow the lentivirus to infect the cells and stably integrate the transgene into the host genome. After 24 h, the cells were washed with DPBS and complete HepatoZYME–SFM with OSM (20 ng ml^{-1}) and hepatocyte growth factor (50 ng ml^{-1}) was added to the cells. This washing procedure was repeated again at 48 h and 72 h post-transduction to ensure

the removal of any remaining lentivirus. hiPSC-derived HLCs were transduced on day 15 of differentiation and assayed on day 23 or 25.

scRNA-seq. Single-cell suspensions from primary tissue and in vitro culture were loaded onto the Chromium controller by 10x Genomics, which is a droplet-based single-cell capture platform. The individual cells flowed through the microfluidic chip, were lysed and tagged by a bead containing unique molecular identifiers and were encapsulated in an oil droplet. This resulting emulsion was amplified through reverse transcription, followed by library preparation as dictated by the 10x Genomics manual. The resulting libraries were sequenced on the Illumina HiSeq 4000 platform. These files were aligned to the GRCh38 human genome and pre-processed using the CellRanger 10x Genomics software for downstream analyses.

Quality control. Cells expressing fewer than 1,000 counts, fewer than 500 genes or more than 40% mitochondrial content were excluded. Application of this filter selected a total of 237,978 cells, which is 87% of the raw number of cells. Genes expressed in fewer than 3 cells were filtered out, leaving 29,907 genes (89% of the total number of genes). Doublets were identified by applying two doublet prediction methods: Doubletdetection[16] and Scrublet[17].

Pseudotemporal ordering and alignment. Time-related genes were selected as markers in collection time (Wilcoxo rank-sum test, z score >10). DPT of time-related genes was derived using the DPT[19] routine implemented in SCANPY. Comparison of pseudotemporal trajectories was performed within the cellAlign[29] framework. cellAlign applies dynamic time warping to compare the dynamics of two single-cell trajectories using a common gene set and to identify local areas of highly conserved expression. The algorithm calculates pairwise distances between ordered points along the two trajectories in gene expression space. cellAlign then finds an optimal path through the matrix of pairwise distances that preserves pseudotemporal ordering and minimizes the overall distance between the matched cells. We applied cellAlign onto all corresponding pairs of hiPSCs and primary cell types by selecting genes used for calculating DPT both in hiPSCs and primary cell types. Cells whose distances were lower than a 0.25 quantile threshold were annotated as aligned in Fig. 7a.

CellPhoneDB analyses. The magnitude and significance of cellular interactions between cell types were calculated using a publicly available repository of curated receptors, ligands and their interactions (CellPhoneDB[43], v2.0). Normalized data for primary tissues at each collection timepoint were used as input. P values of interactions were calculated on the basis of random permutations of cluster labels to generate a null distribution; interactions were considered statistically significant on the basis of the default P value threshold (P value <0.05).

Statistics and reproducibility. Primary cell-specific stage identification was determined on the basis of the standard Louvain clustering routine of scRNA-seq data in SCANPY. Calculation of differentially expressed genes (DEGs) was calculated using a P value threshold ($P<0.01$) and absolute log-fold change threshold (|log-fold change|>1.0). Calculation of time-related genes was selected on the basis of z score (z score >10) thresholds for statistical significance using the Wilcoxon rank-sum test, as stated above. DPT was calculated using the SCANPY computational routine. Significant values for CellPhoneDB ligand–receptor interactions were selected on the basis of a P value threshold for significance (P value <0.05). Traditional over-representation analysis (ORA) pathway enrichment of the interactions revealed by CellPhoneDB were selected for statistical significance on the basis of a P value threshold ($P<1\times10^{-16}$). Transcriptomic shifts between control hPSC-derived cells and control, untreated cells were determined on the basis of Louvain clustering with standard parameters in SCANPY, as well as through differential gene expression on the basis of a P value threshold ($P<0.01$) and absolute log-fold change threshold (|log-fold change|>1.0).

Statistical tests comparing groups in quantitative PCR (qPCR) analyses were calculated using GraphPad Prism. Unpaired samples were compared for each condition using unpaired, two-tailed t-tests, as annotated in the corresponding figure legends (data points and error bars correspond to mean ± standard error of the mean (s.e.m.)). Data distribution was assumed to be normal unless stated otherwise, but this was not formally tested. Outlying data points were excluded on the basis of known experimental error or statistical significance of an outlier test ($P<0.05$). All immunofluorescence and histology stains are representative and correlate to sequencing results, with each micrograph repeated at least twice. All biologically independent replicates are stated explicitly in their respective figure legends.

Cell lines. The hiPSC CA1ATD was was published by Yusa et al. (2011)[60]. The hiPSC lines FSPS13B, YEMZ and KOLF were produced and extensively characterized by the Wellcome Trust Sanger Institute HipSci initiative. The HEK239T line is commercially available from ATCC (CRL-11268), and the HUVEC line is available from Lonza (C2519A). The human foetal HBO lines were generated and characterized thoroughly in this study. None of the cross-contaminated or misidentified cell lines on the list maintained by the International Cell Line Authentication Committee was used in this study.

Ethical approval. Primary human adult liver samples were obtained from deceased transplant organ donors (National Research Ethics Committee East of England—Cambridge South 15/EE/0152) following consent by family following organ donation procedure. Primary human adult liver biopsies were collected from living patients under ethical consent to be used in research (North West—Preston Research Ethics Committee 14/NW/1146) following informed consent by the patient. Primary human foetal liver tissue was obtained from patients undergoing elective terminations (East of England—Cambridge Central Research Ethics Committee, REC-96/085) after informed consent. Detailed information was provided on the objective of the research. Ethical approval covers the use of liver tissues for research and follow the UK rules for the use of human foetal tissues, which exclude compensation for tissue donation. All human tissue was used after obtaining informed consent for use in research following the organ transplantation procedure.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequencing data that support the findings of this study have been deposited in ArrayExpress under accession code E-MTAB-8210. Foetal liver sequencing data have been deposited to ArrayExpress under accession E-MTAB-7189. Previously published foetal liver scRNA-seq data from Popescu et al. have been deposited in ArrayExpress under accession E-MTAB-7407. Previously published adult liver scRNA-seq data from MacParland et al. and Ramachandran et al. have been deposited in NCBI GEO under accession GSE115469 and GSE136103, respectively. All data sources are described in the Supplementary Tables. All additional raw numerical source data presented in plots and graphs in this study are found in the source data files. Any additional data are available upon reasonable request. Source data are provided with this paper.

Code availability

All Python and R scripts supporting the findings of this paper are available upon reasonable request.

Acknowledgements

We thank K. Nayak (University Department of Paediatrics, Cambridge) for help with maintenance of cell lines and technical support and C. Usher for the illustrations in Figs. 1a,e and 4a. We thank the Cellular Genetics department at Cambridge University Hospital (I. Simonic) for performing comparative genomic hybridization, the Core Biochemical Assay Laboratory at Cambridge University Hospitals (K. Burling) for ELISA analysis of culture media, R. Barker and X. He (John Van Geest Centre for Brain Repair, University of Cambridge) for their help accessing tissue, F. Johansson (University of Washington) for surgical and animal support and the Tietze Foundation for funding support. C.A. and D.G. received funding from the Open Targets consortium (OTAR026 project) and the Wellcome Sanger core funding (WT206194). We acknowledge the Cambridge Biorepository for Translational Medicine for the provision of human tissue used in the study. We acknowledge the NHS Addenbrooke's Hospital Tissue Bank for sectioning samples for histology and the Histopathology and Cytology service for immunohistochemistry staining of primary liver tissue. We thank the Cambridge Stem Cell Institute and the Imaging facility. B.T.W. was supported by the Gates Cambridge funding programme. A.D.B.R. was supported by Wellcome Translational Medicine and Therapeutics Clinician PhD programme. The L.V. lab is funded by the ERC Proof of Concept grant Relieve-Chol, by the ERC advanced grant New-Chol, the Cambridge University Hospitals National Institute for Health Biomedical Research Centre and the core support grant from the Wellcome Trust and Medical Research Council of the Wellcome–Medical Research Council Cambridge Stem Cell Institute. C.M.M. is funded by the NC3Rs Training Fellowship. F.S., L.V. and K.S.-P. gratefully acknowledge support from the Rosetrees Trust (REAG/240 and NMZG/233).

Author contributions

B.T.W. performed experimental design, data generation of primary human liver and hPSC differentiation scRNA-seq, tissue processing and analyses, data analysis and biological interpretation, and manuscript preparation; A.D.B.R. performed experimental design, generation of organoid systems and their differentiation, data generation, tissue dissection, processing, and analyses, data analysis and biological interpretation, and manuscript preparation; D.M. performed computational analyses and interpretation; S.S. and K.R.S. conceived of and performed the in vivo experiments, contributed to manuscript preparation and generated figures; J.K. performed experiments and provided analytical support and supervision; Z.M. performed computational analyses and interpretation; R.A.T. performed experimental design, tissue processing, critically appraised the manuscript and provided intellectual contributions; C.M.M. performed tissue dissociation protocol development, immunofluorescence analysis and provided intellectual contributions; K.R. performed smFISH immunostainings and imaging on primary tissue; K.T.M. provided primary adult human liver tissue directly from deceased donors in clinic; J.G.-B. provided technical support while experiments were planned and performed; R.A.B. provided published data to include in the study and critically appraised the manuscript with assistance from E.S. and D.-M.P.; C.A.

provided data and intellectual contributions with advice from D.G.; S.M. and G.B. provided data and intellectual contributions with support from I.D.M.; D.O. provided intellectual contributions, performed flow cytometry analysis and critically appraised the manuscript; E.D.Z. contributed to in vitro experiments; S.B. contributed to experimental planning; F.S. performed tissue processing, provided data, provided intellectual contributions and critically appraised the manuscript; D.H. curated the online data portal viewer; S.P.-R., O.C.T. and K.S.-P. performed direct hepatic and renal mouse injections, K.S.-P. additionally provided primary human liver tissue and intellectual contributions; M.H. provided published datasets and intellectual contributions and critically appraised the manuscript; M.Z. supervised organoid generation and provided intellectual contributions, S.A.T. performed computational analyses planning, data interpretation, provided intellectual contributions and critically appraised the manuscript; and L.V. conceived the study, performed experimental design, interpreted data and analyses and prepared the manuscript.

Competing interests

All authors declare no competing interests.

Additional information

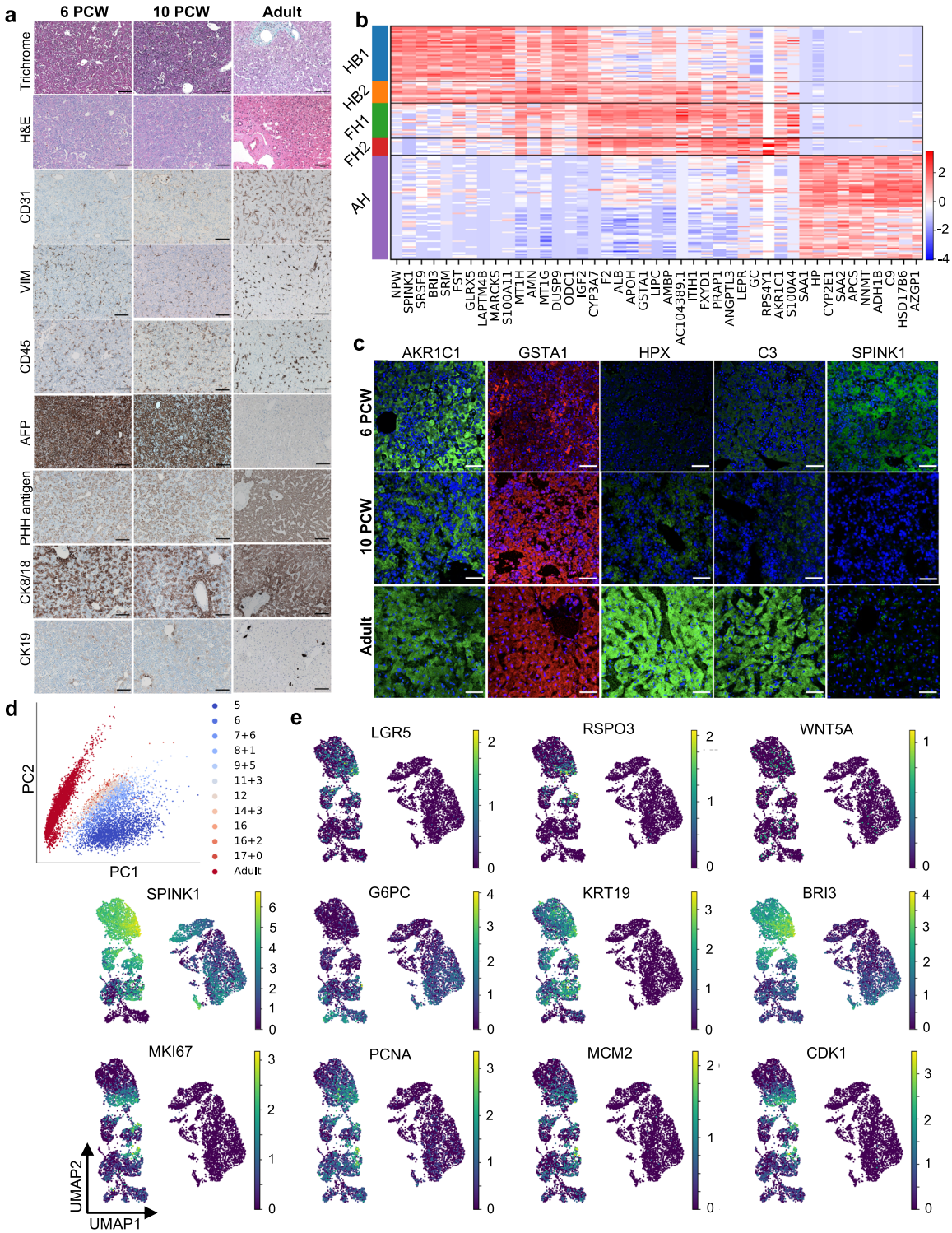
Extended data is available for this paper at https://doi.org/10.1038/s41556-022-00989-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-022-00989-7.

Correspondence and requests for materials should be addressed to Ludovic Vallier.

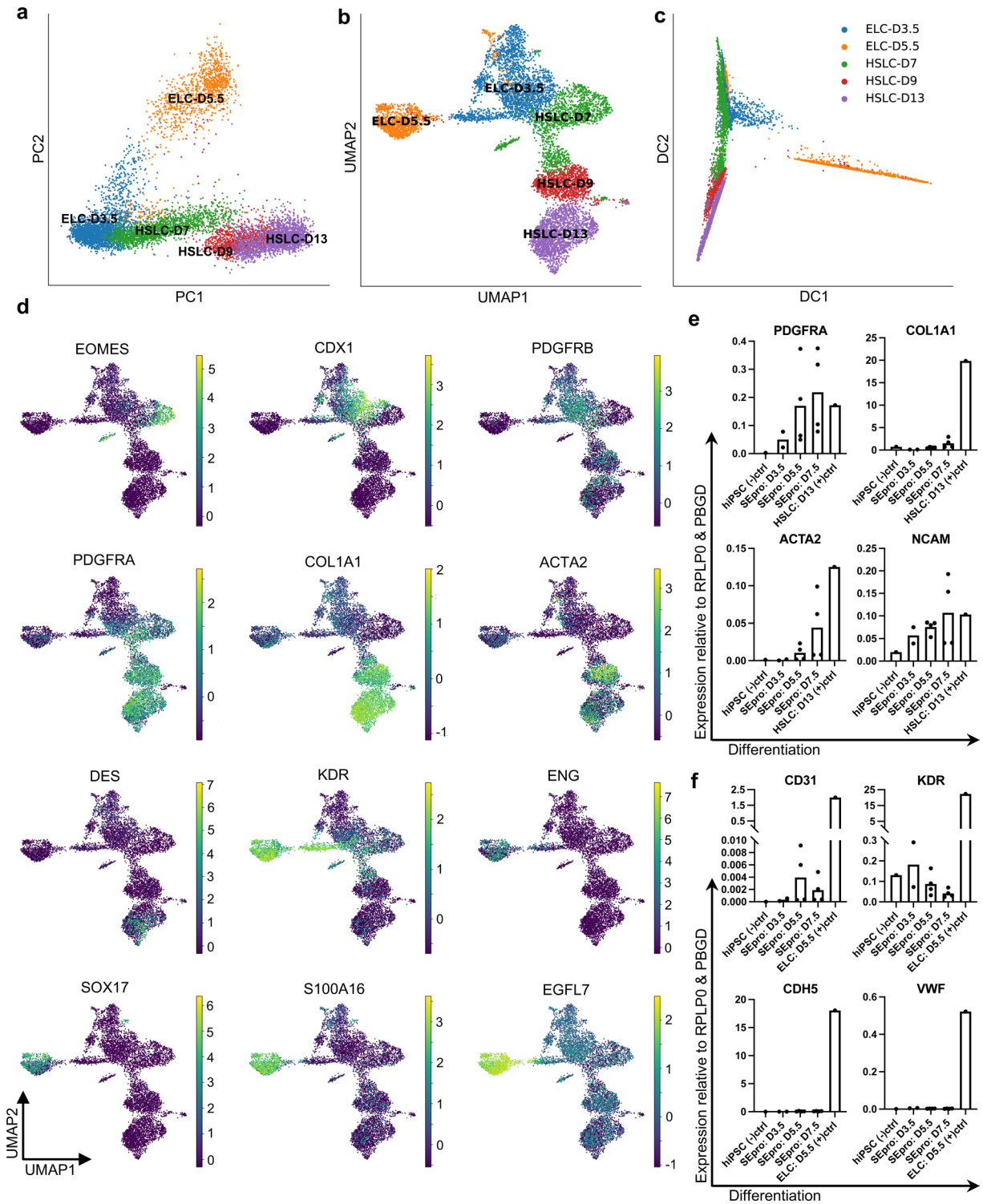
Peer review information *Nature Cell Biology* thanks Valerie Gouon-Evans and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

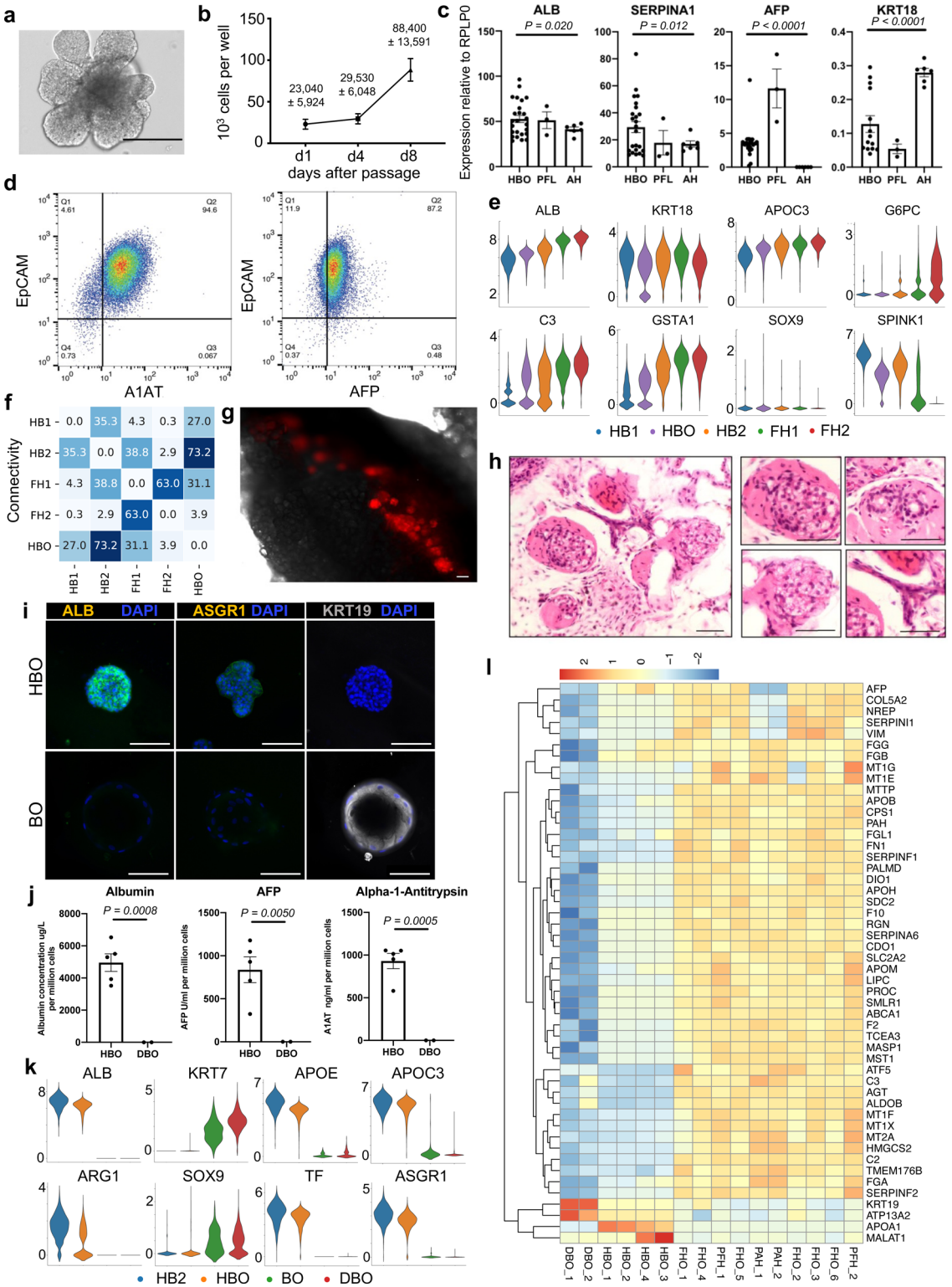


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Characterisation of the proliferation and differentiation of cells within the developing human liver. a, Histology sections of key transcriptional and morphological events of primary foetal and adult liver spanning the key developmental ages analysed in this study; early foetal (6 PCW ± 4-5 days), intermediate foetal (10 PCW) and adult developmental stages; scale bars = 100 um. These immunostainings show transcriptional and morphological events including loss of AFP and increase of hepatocyte-specific antigen during development. Induction of CK19 occurs as cholangiocytes are specified from hepatoblasts, with observable increases in expression in areas of duct formation in the 10 PCW sample. Note the increase of CK8/18 expressing cells as the liver develops, and the necessary increase in vasculature indicated by CD31 expression. Importantly, VIM and CD45, staining for mesenchymal (stellate) cells and resident immune (Kupffer) cells respectively, are present throughout all stages of liver development. Scale bars = 100 um. **b**, Heatmap of the top 10 time-related differentially expressed genes (DEG) at each primary human hepatocyte developmental stage; Wilcoxon-Rank-Sum test, z-score>10; HB1 = hepatoblast stage 1, HB2 = hepatoblast stage 2, FH1 = foetal hepatocyte stage 1, FH2 = foetal hepatocyte stage 2, AH = adult hepatocyte. **c**, Immunostaining analyses showing the expression of stage-specific markers on liver tissue sections; scale bars = 50 um. HB, sample aged between 5-7 PCW; FH, sample dated at 11 PCW; AH, adult liver. **d**, PCA of foetal hepatoblasts/hepatocytes from 5-17 PCW and adult hepatocytes showing the progression of stages correlating to developmental time. **e**, UMAP representation of hepatoblast/foetal hepatocyte markers (SPINK1, G6PC, and BRI3) and WNT pathway markers (WNT5A, LGR5 and RSPO3). UMAP visualization also shows cell proliferation markers and cell cycle regulators across the hepatocyte developmental stages, showing progressive loss of proliferative capacity until the FH2 stage, which may mark commitment to the hepatocyte lineage (integrated scRNA-seq data $n = 17$ independent foetal livers ranging in age from 5 to 17 post-conceptional weeks and $n = 16$ independent adult livers). Heatmap and feature plot colour scales show "gene expression [log-normalized, scaled counts]".

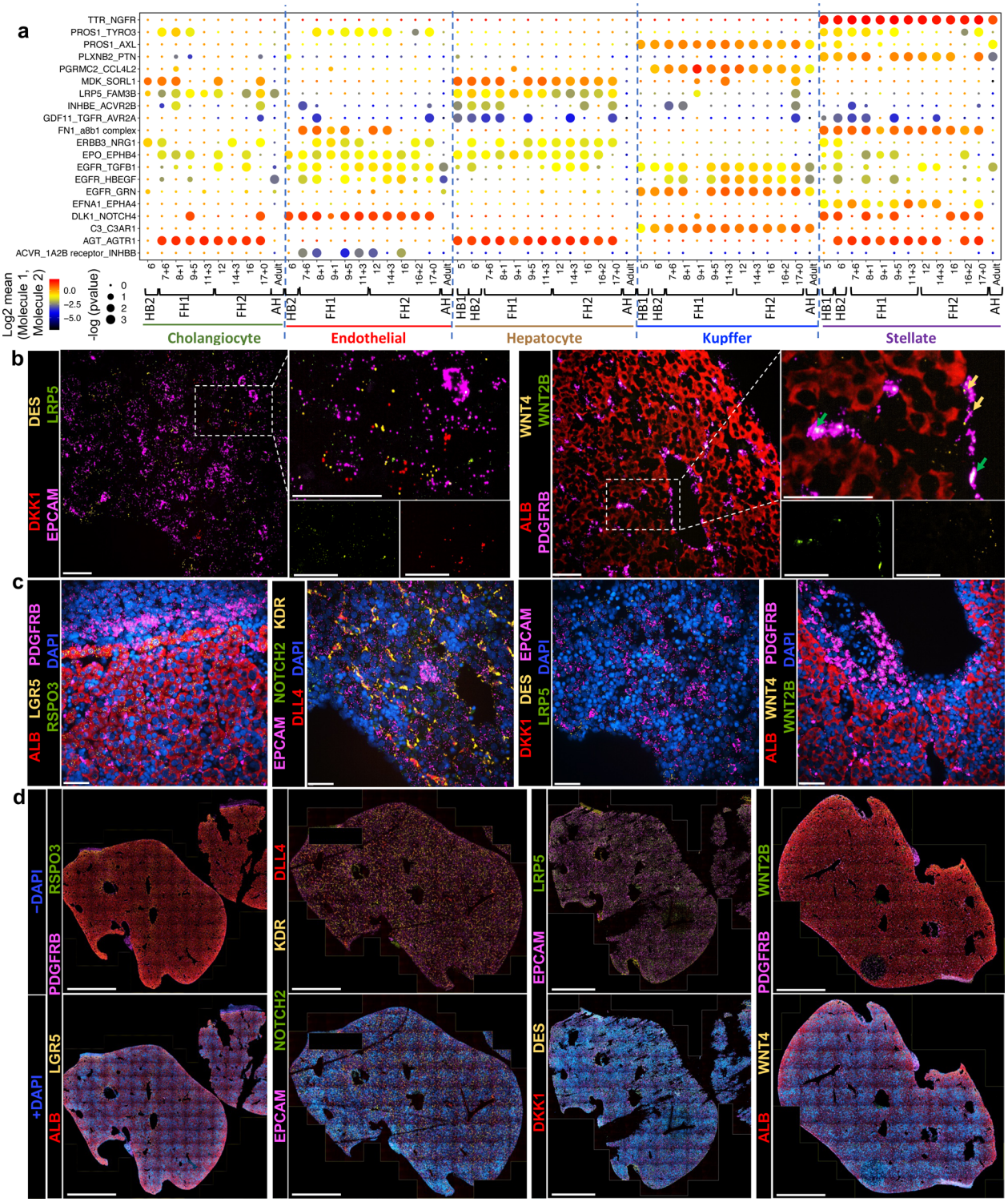


Extended Data Fig. 2 | SEpro validation using an *in vitro* developmental model. a, PCA analysis of hiPSC-derived hepatic stellate cells (HSLC) and hiPSC-derived endothelial cells (ELC) confirming their common bipotent stellate-endothelial progenitor (SEpro) stage following mesoderm induction and before cell lineage specification (scRNA-seq integrates $n = 5$ differentiation timepoints). **b**, UMAP visualization (30 neighbours, 500 PCs) demonstrating the highly correlative relationship of SEpros to the transcriptomically similar ELC-D3.5 and HSLC-D7 stages of *in vitro* differentiation. **c**, Diffusion pseudotime confirming the specification of foetal-like HSLCs and ELCs from their common bipotent progenitor. **d**, UMAP visualizations showing the co-expression of key mesenchymal and endothelial markers such as CDX1, PDGFRB and KDR during the *in vitro* SEpro stage, followed by upregulation of lineage-specific markers and loss of co-expression. **e**, QPCR analyses of hiPSC differentiated first toward endothelial cells, then transitioned into culture conditions to specify hepatic stellate cells upon reaching the bipotential SEpro stage. These qPCR analyses show the acquisition of hepatic stellate cells markers and **f**, the loss of endothelial markers ($n = 2$ independent cell lines). Data are presented as mean values +/- SEM. Feature plot colour scales show "gene expression [log-normalized, scaled counts]".

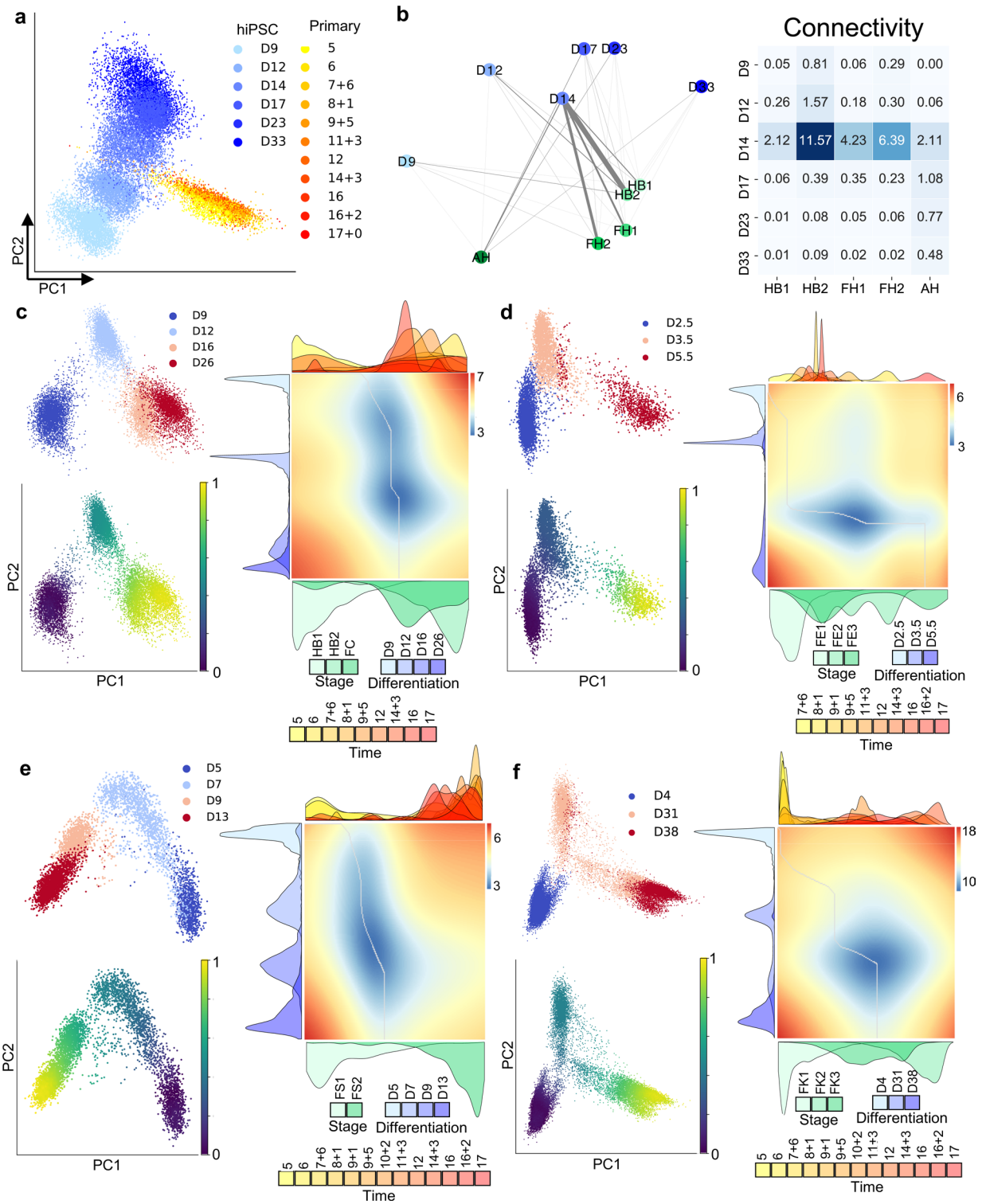


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Characterisation of progenitor hepatoblast organoids and their unique developmental identity. a, Representative brightfield image of HBO; scale bar = 200 um. **b**, Total number of cells per well averaged over three HBO lines at days 1, 4 and 8 post-passage ($n=5$). **c**, QPCR analyses confirming the expression of key hepatoblast markers in HBO. Gene expression is shown as values relative to housekeeping gene RPLP0. Each dot represents a biological replicate; HBO: ALB, SERPINA1, AFP ($n=25$: derived from $n=9$ primary foetal livers at progressive passages from 3 to 15); HBO: KRT18 ($n=14$); primary foetal liver (PFL) ($n=3$); primary adult hepatocytes (PAH) ($n=6$). **d**, Fluorescence activated cell sorting (FACS) analysis on HBO (passage 11) using antibodies to EPCAM (647) and A1AT (488) (left), and EPCAM (647) and AFP (488) (right). **e**, ScRNA-seq violin plots comparing key functional markers of different stages of *in vivo* hepatocyte development (integrated scRNA-seq data from $n=17$ independent foetal livers ranging in age from 5 to 17 post-conceptional weeks and $n=16$ independent adult livers) to *in vitro* hepatoblast organoids ($n=2$ cell lines), confirming their similarity to the HB2 stage. **f**, PAGA connectivity analysis confirming the resemblance of HBO and HB2. **g**, TdTomato-positive cell grafts (red) were identified in mouse fat (black, phase) upon explant of organoid tissues after 27 days of engraftment; scale bar = 20 um. **h**, Hematoxylin & Eosin staining of explanted grafts showing nodules (right, 4 images) with cells that morphologically resembled either densely packed hepatocytes (left) or biliary ductal-like structures (transverse section, right top; longitudinal section, right bottom); scale bars = 100 um. **i**, Immunostaining of HBO and BO for albumin (ALB), asialoglycoprotein receptor 1 (ASGR1), and cytokeratin-19 (KRT19); scale bars = 100 um. **j**, Concentration per litre of secreted proteins by HBO ($n=5$, each line derived from an independent foetal liver) and DBO ($n=2$, each line derived from an independent foetal liver) after 48 hours of freshly applied medium. Values are normalised to cell number (that is per million cells). **k**, ScRNA-seq violin plots highlighting the similarity in key hepatoblast functional gene expression between HB2 and HBO, and their dissimilarity to the cholangiocytic BO and DBO cultures. **l**, Heatmap of the top 50 absolute loadings in principal component one (32% variance) comparing DBO, FHO, HBO, PAH and PFH. Heatmap colour scales show "gene expression [log-normalized, scaled counts]". Data are presented as mean values +/- SEM; unpaired two-tailed t-tests.

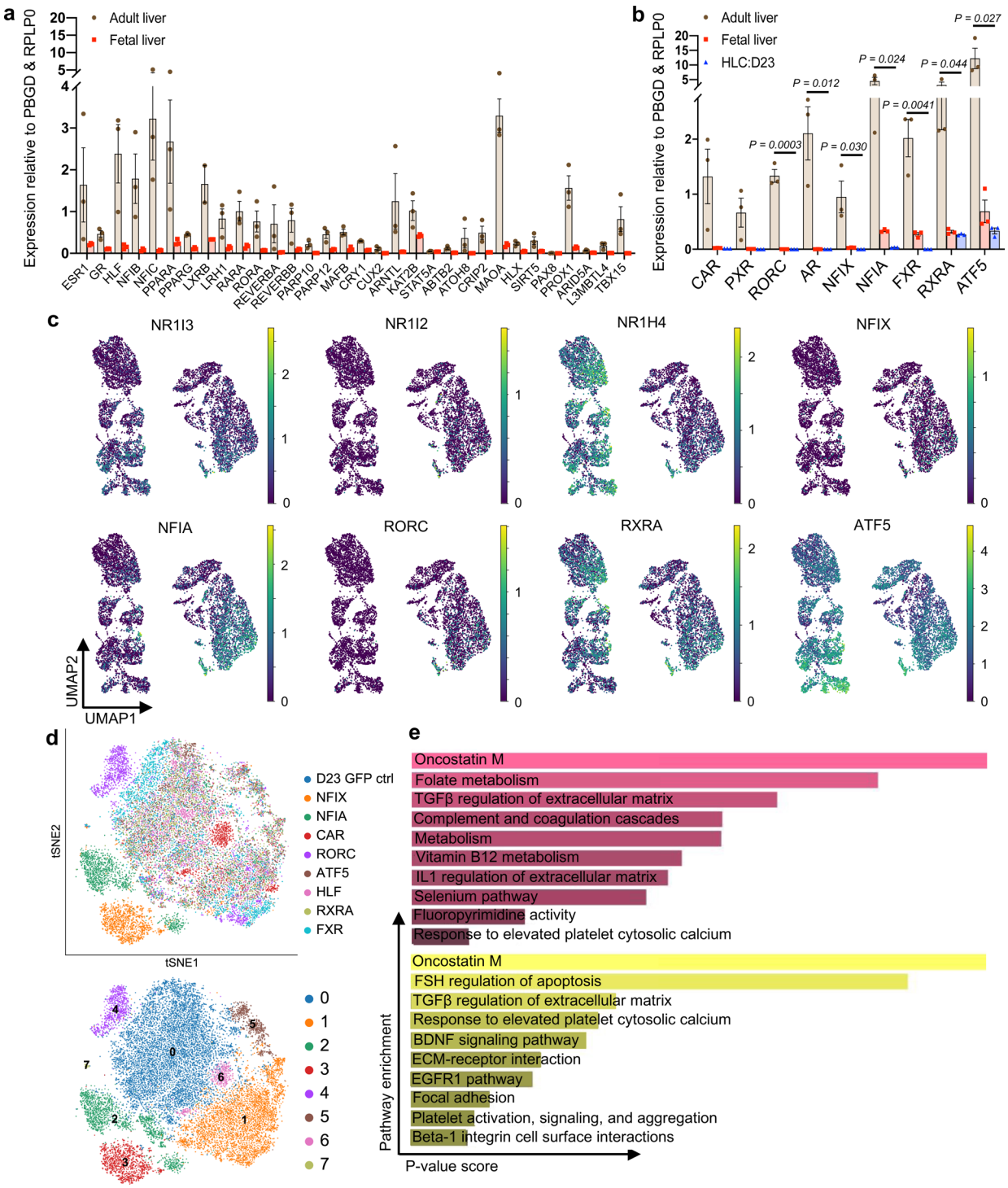


Extended Data Fig. 4 | Identifying dynamic intercellular interactions during development that enable hepatocyte differentiation in the human liver. **a**, CellphoneDB analysis of hepatocytes receptors-ligand interactions with cholangiocytes, endothelial cells, hepatocytes, Kupffer cells and stellate cells across all developmental timepoints. These dynamic, temporal interactions with the nonparenchymal cells establish the hepatoblast/hepatocyte developmental extracellular environment. The Y-axis shows ligand-receptor/receptor-ligand interactions, with the hepatocyte protein listed first in the pair; x-axis shows developmental time of each interacting cell type; intensity shows log2 mean of interacting molecules; size of dot shows -log10(P) values ($n=16$ independent foetal livers ranging in age from 5 to 17 post-conceptional weeks and $n=16$ independent adult livers). **b**, RNAScope images showing DES + hepatic stellate cells expressing DKK1 which interacts with the LRP5 receptor on EPCAM + hepatoblasts (left panel). RNAScope images also reveal hepatoblasts expressing WNT4 and stellate cells expressing WNT2B, which is necessary for the proliferation and expansion of the early foetal liver hepatoblasts (right panel); scale bars = 50 um. **c**, RNAScope images including DAPI nuclear stain to visualise the co-expression of markers demonstrated by the RNAScope images in Fig. 5; scale bars = 50 um. **d**, RNAScope of entire human foetal liver tissue sections, both without (top row) and with (bottom row) DAPI nuclear staining; thus demonstrating the tissue architecture and validity of selected RNAScope imagines in Fig. 5; scale bars = 1 mm.

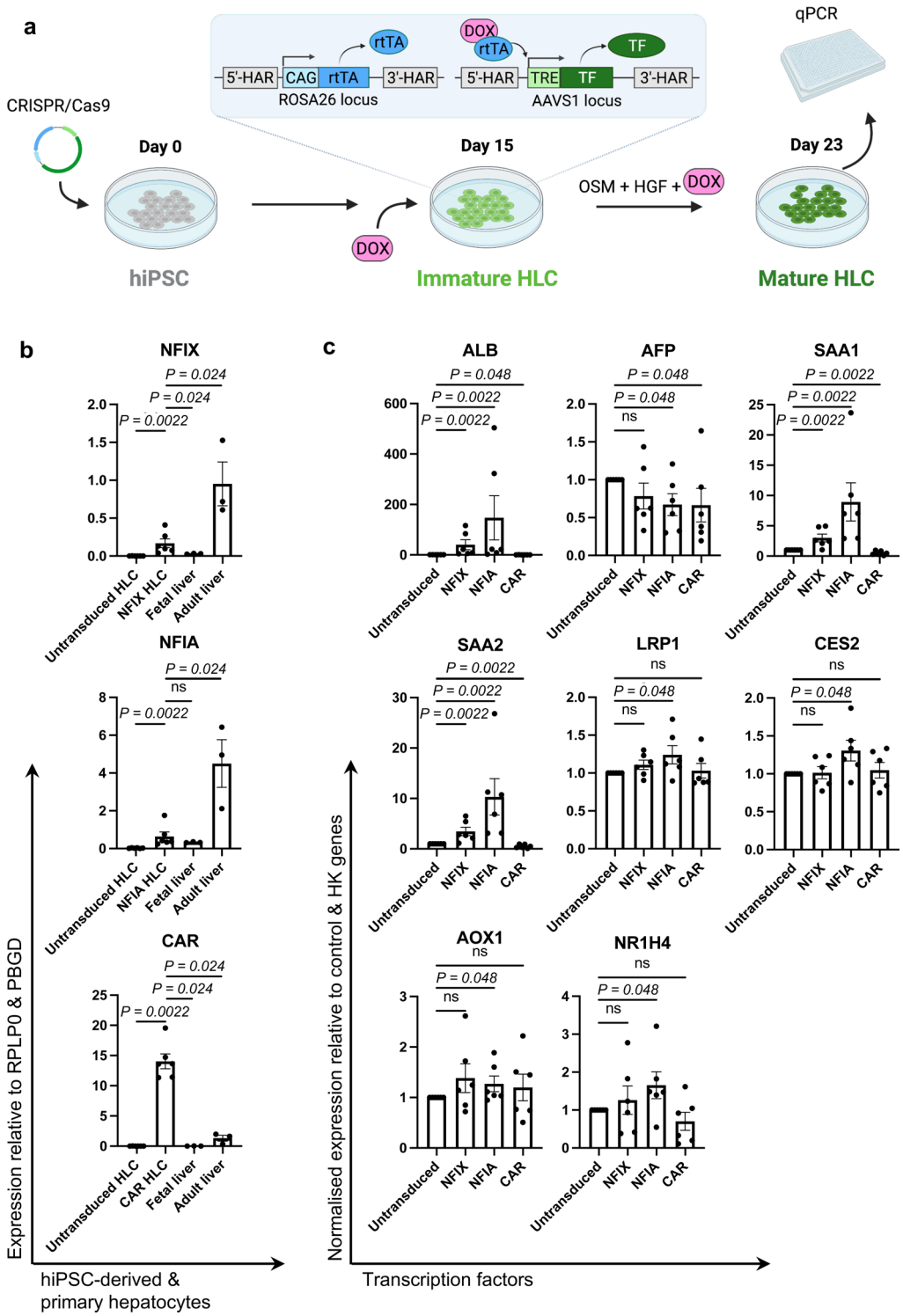


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Comparison of hPSC-derived liver cell types to their primary counterparts during human liver development. a, PCA plot of HLC diffusion map alignment to primary hepatoblast/hepatocyte development, validating the similarity of D14 HLC to HB2 shown in Fig. 7. **b**, PAGA connectivity plots and connectivity values comparing HLC timepoints to primary developmental timepoints/stages, confirming the similarity of the differentiation state of HLCs at day 14 (D14) to primary HB2. **c**, PCA plot of cholangiocyte-like cell (CLC) differentiation timepoints (top) and pseudotime (bottom) with 0.0 being the earliest pseudo-timepoint and 1.0 being the latest for all pseudotime analyses. Alignment of CLCs differentiation time course ($n = 4$ sequential timepoints) to primary cholangiocyte development (right), revealing the divergence of in vitro and in vivo differentiation at an early timepoint explaining the inability of CLCs to fully resemble primary adult cells. **d**, PCA plot of endothelial-like cells (ELC) differentiation timepoints (top) and pseudotime (bottom). Alignment of ELC differentiation time course ($n = 3$ sequential timepoints) to primary endothelial cell development (right) revealing the divergence of in vitro and in vivo differentiation at an early timepoint. **e**, PCA plot of hepatic stellate-like cell (HSLCs) differentiation timepoints (top) and pseudotime (bottom). Alignment of HSLCs differentiation time course ($n = 4$ sequential timepoints) to primary hepatic stellate cell development (right) revealing the divergence of in vitro and in vivo differentiation at an early timepoint. **f**, PCA plot of macrophage-like cell (MLC) differentiation timepoints (top) and pseudotime (bottom). Alignment of MLC differentiation time course ($n = 3$ sequential timepoints) to primary Kupffer cell development (right) revealing the divergence of in vitro and in vivo differentiation at an early timepoint. Dpt pseudotime colour scale shows "geodesic distance [distance between nodes]"; cell alignment plot red colour shows regions of misalignment/dissimilarity, blue colour shows regions of close alignment/similarity. Sequential hPSC differentiation timepoints integrate scRNA-seq data from one replicate per timepoint; plots of primary liver development integrate scRNA-seq data from $n = 16$ independent foetal livers ranging in age from 5 to 17 post-conceptional weeks.



Extended Data Fig. 6 | Improving *in vitro* functionality of hPSC-derived hepatocytes using the human liver developmental map. a, QPCR analyses showing the expression of transcription factors known to affect hepatocyte maturation in adult ($n=3$) and foetal ($n=3$, 6-10 PCW) human liver as well as **b**, factors identified from the scRNA-seq human liver development map (full lists included in Source Data). **c**, UMAP visualization of transcription factors identified in Fig. 7 during all stages of hepatocyte development annotated in Fig. 1 (integrated scRNA-seq data from $n=17$ independent foetal livers ranging in age from 5 to 17 post-conceptional weeks and $n=16$ independent adult livers). Note that NFIX and NFIA are expressed at low levels in foetal stages and reach a maximum in adult hepatocytes. Colour scale shows "gene expression [log-normalized, scaled counts]". **d**, t-SNE visualization of hPSC-derived hepatocytes transduced with selected transcription factors (TFs) or day 23 (D23) GFP control (top). Unique, separate Louvain clusters signify differences in expression profile among transduced cells compared to the GFP control (bottom) ($n=1$ sample sequenced per transduction). **e**, Pathway enrichment of top 150 DEGs comparing transduced HLCs to D23 GFP control for NFIA (top) and NFIX (bottom) using the NCATS BioPlanet pathway database. Feature plot colour scales show "gene expression [log-normalized, scaled counts]". Data are presented as mean values +/- SEM; unpaired two-tailed t-tests.



Extended Data Fig. 7 | Validating the maturation of *in vitro* derived hepatocytes using an inducible expression culture system. a, Schematic of the experimental design for doxycycline (DOX) induction of key transcription factors during hepatocyte-like cell differentiation from hPSCs. HiPSC cells were edited to include the transcription factor in an inducible cassette, which was upregulated at day 15 of differentiation using DOX and continued to be expressed until day 23 when the HLCs were assays for maturation via qPCR; created with BioRender.com. **b**, QPCR analysis showing expression of NFIX, NFIA or CAR by day 23 hepatocyte-like cells (HLCs) that were induced to express their respective transcription factors since day 15 of differentiation using a doxycycline-inducible expression system ($n \geq 3$). The induced expression levels are compared to primary foetal and adult human liver, showing that expression is within physiological bounds. **c**, In addition to the increase in mature albumin expression and the decrease of foetal AFP expression, qPCR shows the increase in a wide array of functional and metabolic hepatocyte markers, thus, indicating maturation due to temporally-relevant expression of key transcription factors ($n = 6$). Data are presented as mean values +/- SEM; Mann-Whitney unpaired two-tailed t-tests.

nature research

Corresponding author(s): Ludovic Vallier

Last updated by author(s): May 18, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

The l0X single cell sequencing data were mapped with the Cell ranger program (version 2.0.2)15 to the reference 'refdata-cellranger-GRCh38-l.2.0', which is the human genome 'GRCh38' downloaded from (https://support.l0xgenomics.com/single-cell-gene-expression/software/downloads/latest). Reads for each sample were first imported into an 'AnnData' object (https://anndata.readthedocs.io/en/latest/ index.html) in the SCANPY30 framework, count matrices were then concatenated into a total matrix of 273,033 cells and 33,694 genes.

Data analysis

All packages used throughout the computational analyses:
Scanpy (vl.4), Anndata (v0.6.18), Numpy (vl.15.4), Scipy (vl.1.0), Pandas (v0.25.1), Scikit-learn (v0.21.3), Statsmodels (v0.10.1), Python-igraph (v0.7.1), Louvain (v0.6.1), CellphoneDB (V2.0.0)
R (v3.5.0), dplyr (v0.8.0.1), RColorBrewer (vl.1-2), pheatmap (vl.0.12), reshape2 (vl.4.3), ggplot2 (v3.1.1), cellAlign (v0.1.0), packrat (v0.5.0)
Read counts were log-normalised in the SCAN PY framework using total count normalisation (scaling factor 10,000). Highly variable genes were selected for downstream analysis based on their normalised dispersion, obtained by scaling with the mean and standard deviation of the dispersions for genes falling into a given bin for mean expression of genes ('highly_variable_genes' routine in SCANPY). Technical variation was regressed out using a linear regression model ('regress_out' function from the python package of 'NaiveDE'), which allows the user to specify wanted variance when cell types or stages are known. Alternatively, the 'COMBAT' function from the svaseq package could have been equivalently used. Regressed values were scaled by maintaining the default maximum value of 10 for each gene expression in each cell ('scale' routine in SCANPY).
Clusters were detected by applying the graph-based, community detection method 'Louvain'. Annotation of clusters was based on the expression of known markers. Since Louvain clustering does not necessarily correlate with biologically meaningful clusters and can over-cluster cells associated with the same cell type, clusters with the same annotation were merged. Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (tSNE) and Uniform Manifold Approximation and Projection (UMAP) plots were calculated with the SCAN PY routines.

Transcriptional similarity among stages was quantified and visualised using the method: Partition-Based Graph Abstraction (PAGA). PAGA

generates a graph-like map of cells based on estimated connectivities which preserves both the continuous and disconnected structure of the data at multiple resolutions. In a PAGA graph, nodes correspond to cell groups and edge weights quantify the connectivity between groups. Such weights represent the confidence in the presence of an actual connection by allowing discarding spurious, noise-related connections. The highest the connectivity values, the highest the confidence of a connection. In our PAGA graphs, we used connectivity estimates to analyse the relative similarity among stages.

Time-related genes were selected as markers in collection time (Wilcoxon-Rank-Sum test, z-score>l0). Diffusion pseudo-time of time-related genes was derived using the DPT46 routine implemented in SCAN PY. Comparison of pseudo-temporal trajectories was performed within the cellAlign framework. cellAlign applies dynamic time warping to compare the dynamics of two single-cell trajectories using a common gene set and to identify local areas of highly conserved expression. The algorithm calculates pairwise distances between ordered points along the two trajectories in gene expression space. cellAlign then finds an optimal path through the matrix of pairwise distances which preserves pseudotemporal ordering and minimises the overall distance between the matched cells. We applied cellAlign onto all corresponding pairs of hiPSCs and primary cell types by selecting genes used for calculating diffusion pseudo-time both in hiPSCs and primary cell types. Cells whose distances were lower than a 0.25 quantile threshold were annotated as aligned in Fig. 7.
The significance of cellular interactions between cell types was calculated with a publicly available repository of curated receptors, ligands and their interactions (CellPhoneDB48, v2.0). Normalised data for primary tissues at each collection time point were used as input. Significance of interactions was calculated based on random permutations of cluster labels to generate a null distribution. Interactions were considered significant based on the default p-value threshold (p-value<0.05).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated by this study is available for visualization through our online portal: https://collections.cellatlas.io/liver-development

Source Data files include complete lists of outputs from all computational analyses and parameters used for each specific analysis used throughout the main figures and extended data figures. The Supplementary Tables list all markers used to annotate cell type lineages (Fig. 1). Fig. 1 Source Data annotates all differentially expressed genes (DEGs) and gene ontology (GO) terms for the 5 hepatoblast/hepatocyte developmental stages discovered in this study using single-cell RNA sequencing (scRNA-seq) (Fig. 1d,e; Extended Data Fig. 1). Fig. 2 Source Data annotates all DEGs and GO terms for other primary human liver cell types discussed in this study: cholangiocytes, hepatic stellate cells, endothelial cells and Kupffer cells (Fig. 2). Fig. 3 Source Data shows the DEG and the cell clustering corresponding to the pseudotemporal specification of the stellate-endothelial progenitor cells (SEpro) toward endothelial and hepatic stellate lineages (Fig. 3). Validation of this finding was performed by generating this stellate-endothelial progenitor in vitro and differentiating this cell population into hepatic stellate-like cells (Extended Data Fig. 2). Fig. 4 Source Data shows the DEG comparison of hepatoblast organoids generated in vitro (HBOs) to the in vivo developmental stages to hepatoblasts to hepatocytes, thus confirming their similarity to the bipotential progenitor stem cells in the developing liver (Fig. 4). HBOs were differentiated toward hepatocytes in vitro, and the scRNA-seq DEG analyses comparing the transcriptomes of these differentiated organoids to the primary hepatocyte developmental stages are annotated in Fig. 5 Source Data (Fig. 5). Similarly, confirming the bipotentiality of these cells, they were differentiated in vitro toward cholangiocytes, with scRNA-seq DEG analyses comparing these differentiated cells to their primary counterparts annotated in Fig. 5 Source Data (Fig. 5). These complete differentially expressed gene lists show how the HBOs progress toward either hepatocytes or cholangiocytes by decreasing their similarities to the primary hepatoblasts and increasing their transcriptomic similarities to the mature lineages. Source Data Fig. 6 provides all interactions calculated from CellPhoneDB, with an emphasis on how each cell type interactions with the developing hepatocyte throughout liver organogenesis (Fig. 6, Extended Data Fig. 4). Fig. 6 Source Data also annotates ORA pathway enrichment using the Reactome database (www.reactome.org) for all interactions calculated using CellPhoneDB. The in-depth pairwise DEG characterisation of human pluripotent stem cell-derived (hPSC-derived) hepatocytes at each stage of differentiation, and the comparison of this differentiation timecourse to the developmental path of primary hepatocytes is annotated in Fig. 7 Source Data (Fig. 7 and Extended Data Fig. 5). The DEGs associated with each stage of hPSC differentiation for the other liver cell lineages are annotated in full in Fig. 7 Source Data. DEG analyses of hPSC-derived hepatocytes that have been transduced with key transcription factors and profiled using scRNA-sequencing are annotated in Fig. 7 Source Data (Fig. 7 and Extended Data Fig. 6).

Extended Data Fig. 2 Source Data and Extended Data Fig. 3 Source Data annotate the data used to generate the bar plots in these respective figures. Extended Data Fig. 5 Source Data catalogues the DEGs associated with each stage of hPSC differentiation toward the cholangiocyte, hepatic stellate cell, endothelial cell and macrophage lineages. This comprises a characterization of these hiPSC differentiation timecourses, which was then used to align the pseudotemporal lineages with the developmental timecourses of their respective primary human liver cell types. Extended Data Fig. 6 Source Data shows the qPCR data for the comparison of transcription factor expression levels between primary adult human liver to primary human fetal liver. Extended Data Fig. 7 Source Data shows the qPCR data for the maturation of HLCs by overexpressing temporally-specific transcription factors using an inducible system

The Supplementary Tables describes the antibodies used for histology and immunochemistry, the probes used for RNAScopy and all media compositions used to differentiate and upkeep cells in culture for this study. The Supplementary Tables also annotate all primers used for qPCR experiments including the comparison of adult human liver to fetal human liver and hepatocyte-like cells on day 23 of differentiation, the characterisation of the hepatoblast organoids and their differentiation into hepatocyte and cholangiocyte lineages, and the validation of novel transcription factors essential to the maturation of hPSC-derived hepatocytes in vitro.

All raw single-cell RNA sequencing data has been deposited to ArrayExpress under accession E-MTAB-8210. Fetal liver sequencing data has been deposited to ArrayExpress under accession E-MTAB-7189. Fetal liver sequencing data from Popescu et al. has been deposited to ArrayExpress under accession E-MTAB-7407. Adult liver sequencing data from MacParland et al. and Ramachandran et al. have been deposited to NCBI GEO under accession GSE115469 and GSE136103, respectively. All additional raw numerical source data presented in plots and graphs in this study are found in the Source Data files.

CellPhoneDB is available at www.CellPhoneDB.org, along with lists of ligands and receptors and heteromeric complexes.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined based on previous literature and information from the Human Cell Atlas white paper (www.humancellatlas.org/news/13); no independent statistical models were used to determine sample size.
Data exclusions	Exclusion of data in computational and downstream analyses is described in Methods, with all raw data uploaded to ArrayExpress. As per standard single-cell RNA sequencing quality control parameters, we excluded cells based on high mitochondrial percentage, algorithms and high and low number of genes and counts, and doublet detection algorithms: cells expressing fewer than 1000 counts, fewer than 500 genes or more than 40% mitochondrial content were excluded. Application of such filter selected a total number of 237978 cells, which is 87% of the raw number of cells. Genes expressed in fewer than 3 cells were filtered out, leaving 29907 genes (89% of the total number of genes). Doublets were identified by applying two doublet prediction methods: Doubletdetection and Scrublet. Specific datasets were subset during analyses to include only relevant cells (desired cell types for primary data, on-target differentiation for hPSC-derived cells as shown in the Extended Data Figs. and successfully transduced cells for overexpression experiments in hPSC-derived cells).
Replication	This study includes scRNA-sequencing data from a total of n=19 human adult livers and n=15 human fetal livers of both male and female origin to capture all of the relevant cell types and minimize patient background variation. Here, no further attempts at replication were made, and all data were included.
Randomization	Adult liver that were rejected for transplant were compared against patient data and assessed to be healthy. Fetal liver samples were healthy. No randomization was performed. The background of the individuals that consented for research is anonymous to the researchers, and as such, all samples were only identified through numerical identification. All eligible participants for research were made available to researchers, with no other factors available to facilitate a decision process that could introduce bias.
Blinding	Blinding to the origin of the tissue was not possible, as the tissue was provided by the Cambridge Biorepository for Translational Medicine (CBTM) from deceased adult donors. Blinding was not possible for cell lines nor for data received from collaborating groups. Tissue was handled systematically in a similar manner based on source/age.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	All antibodies are listed in detail in Supplementary Table 22. Anti-SPINK1 antibody [EPR17585-116] SPINK1 Hepatoblast Rb 1:200 IHC-P Abcam ab207302 Anti-Hemopexin antibody [EPR5610] HPX Hepatocyte Rb 1:100-1:250 ICC/IF, IHC-P Abcam ab124935 Anti-C3 antibody C3 Hepatocyte Rb 1:100-1:1000 ICC/IF, IHC-P Abcam ab97462 Anti-AKR1C1 antibody AKR1C1 Hepatocyte Rb 7.5 ug/ml IHC-P Abcam ab192785 Anti-GSTA1 antibody GSTA1 Hepatocyte Go 1:1000 ICC/IF, IHC-P Abcam ab53940 Anti-Cytokeratin 7 antibody KRT7 Cholangiocyte Rb 1:200 IHC-P, ICC Abcam ab68459 Anti-Cytokeratin 19 antibody KRT19 Cholangiocyte Mo 1:200 IHC-P/Fr, ICC Abcam ab7754 Anti-α1-AT SERPINA1 Hepatocyte Rb NA IHC-P Dako A0012 Ansti-CD31 PECAM1 Endothelial Mo NA IHC-P Leica PA0414

Anti-CD45 PTPRC Immune cells Mo NA IHC-P Dako M070101
Anti-CD68 K CD68 Macrophage/Kupffer Mo NA IHC-P Dako M081401
Anti-CK8/18 KRT8/18 Hepatocyte Mo NA IHC-P Leica PA0067
Anti-CK19 KRT19 Cholangiocyte Mo NA IHC-P Leica PA0799
Anti-Human hepatocyte NA Hepatocyte Mo NA IHC-P Dako M715801
Anti-Vimentin VIM Mesenchyme Mo NA IHC-P Dako M702001
Albumin ALB Hepatoblast/Hepatocyte Go 1:100 IHC/IF Bethyl A50-103A
Alpha-1-Antitrypsin SERPINA1 Hepatoblast/Hepatocyte Rb 1:100 IHC/IF Dako AA0012
Alpha-Fetoprotein AFP Hepatoblast Rb 1:100 IHC/IF Dako GA500
Arginase 1 ARG1 Hepatocyte Rb 1:400 IHC/IF Sigma-Aldrich SAB5500008
Cytokeratin-18 KRT18 Hepatocyte/cholangiocyte Mo 1:25 IHC/IF Dako GA618
hepatocyte nuclear factor-4-alpha HNF4A Hepatoblast/hepatocyte Mo 1:100 IHC/IF R & D systems PP-H1415-00
SRY-box transcription factor 9 SOX9 cholangiocyte Rb 1:100 IHC/IF Abcam ab185230

SECONDARY ANTIBODIES:
Secondary Antibodies

Goat anti-Mouse IgG1 Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568. ThermoFisher. Ref: Ham DJ et al., Distinct and additive effects of calorie restriction and rapamycin in aging skeletal muscle. Nat Commun. 2022 Apr 19;13(1):2025. doi: 10.1038/s41467-022-29714-6.

Goat anti-Mouse IgG2a Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488. ThermoFisher. Ref: Liao CC, Chiu CJ, Yang YH, Chiang BL. Neonatal lung-derived SSEA-1+ cells exhibited distinct stem/progenitor characteristics and organoid developmental potential. iScience. 2022 Apr 16;25(5):104262. doi: 10.1016/j.isci.2022.104262. PMID: 35521516; PMCID: PMC9062680.

Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488. ThermoFisher. Ref: Lee HJ, Jung DH, Kim NK, Shin HK, Choi BT. Effects of electroacupuncture on the functionality of NG2-expressing cells in perilesional brain tissue of mice following ischemic stroke. Neural Regen Res. 2022 Jul;17(7):1556-1565. doi: 10.4103/1673-5374.330611. PMID: 34916441; PMCID: PMC8771106.

Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488. ThermoFisher. Shue YT, Drainas AP, Li NY, Pearsall SM, Morgan D, Sinnott-Armstrong N, Hipkins SQ, Coles GL, Lim JS, Oro AE, Simpson KL, Dive C, Sage J. A conserved YAP/Notch/REST network controls the neuroendocrine cell fate in the lungs. Nat Commun. 2022 May 16;13(1):2690. doi: 10.1038/s41467-022-30416-2. PMID: 35577801; PMCID: PMC9110333.

Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568. ThermoFisher. Ref: van Gurp L, Fodoulian L, Oropeza D, Furuyama K, Bru-Tari E, Vu AN, Kaddis JS, Rodríguez I, Thorel F, Herrera PL. Generation of human islet cell type-specific identity genesets. Nat Commun. 2022 Apr 19;13(1):2020. doi: 10.1038/s41467-022-29588-8. PMID: 35440614; PMCID: PMC9019032.

Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568. ThermoFisher. Ref: Yan M, Yu W, Lv Q, Lv Q, Bo T, Chen X, Liu Y, Zhan Y, Yan S, Shen X, Yang B, Hu Q, Yu J, Qiu Z, Feng Y, Zhang XY, Wang H, Xu F, Wang Z. Mapping brain-wide excitatory projectome of primate prefrontal cortex at submicron resolution and comparison with diffusion tractography. Elife. 2022 May 20;11:e72534. doi: 10.7554/eLife.72534. PMID: 35593765; PMCID: PMC9122499.

Donkey anti-Sheep IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568. ThermoFisher. Ref: Nabi D, Drechsler H, Pschirer J, Korn F, Schuler N, Diez S, Jessberger R, Chacón M. CENP-V is required for proper chromosome segregation through interaction with spindle microtubules in mouse oocytes. Nat Commun. 2021 Nov 11;12(1):6547. doi: 10.1038/s41467-021-26826-3. PMID: 34764261; PMCID: PMC8586017.

Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647. ThermoFisher. Ref: Wells KM et al., Brucella activates the host RIDD pathway to subvert BLOS1-directed immune defense. Elife. 2022 May 19;11:e73625. doi: 10.7554/eLife.73625. PMID: 35587649; PMCID: PMC9119680.

Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647. ThermoFisher. Samanta MK, Gayen S, Harris C, Maclary E, Murata-Nakamura Y, Malcore RM, Porter RS, Garay PM, Vallianatos CN, Samollow PB, Iwase S, Kalantry S. Activation of Xist by an evolutionarily conserved function of KDM5C demethylase. Nat Commun. 2022 May 11;13(1):2602. doi: 10.1038/s41467-022-30352-1. PMID: 35545632; PMCID: PMC9095838.

Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568. ThermoFisher. Mike JK, Wu KY, White Y, Pathipati P, Ndjamen B, Hutchings RS, Losser C, Vento C, Arellano K, Vanhatalo O, Ostrin S, Windsor C, Ha J, Alhassen Z, Goudy BD, Vali P, Lakshminrusimha S, Gobburu JV, Long-Boyle J, Chen P, Wu YW, Fineman JR, Ferriero DM, Maltepe E. Defining longer term outcomes in an ovine model of moderate perinatal hypoxia-ischemia. Dev Neurosci. 2022 May 19. doi: 10.1159/000525150. Epub ahead of print. PMID: 35588703.

Validation

All antibodies used are commercially available and were validated by the manufacturer for use in human primary tissue histology. As required, antibodies were validated against negative controls.

Validation:
Anti-SPINK1/P12 antibody [EPR17585-116] - Produced recombinantly (animal-free) for high batch-to-batch consistency and long term security of supply. Manufacturer positive control validation: Human pancreas and hepatocellular carcinoma tissues.

anti-Hemopexin antibody [EPR5610] Ref: Canesin G et al. Scavenging of Labile Heme by Hemopexin Is a Key Checkpoint in Cancer Growth and Metastases. Cell Rep 32:108181 (2020).. Synthetic peptide within Human Hemopexin aa 400-500. The exact sequence is proprietary. Manufacturer positive control validation: Human liver tissue.

Anti C3 Ref: Young AMH et al. A map of transcriptional heterogeneity and regulatory variation in human microglia. Nat Genet 53:861-868 (2021).

Anti AKR Ref: Nie J et al. A Novel Ferroptosis Related Gene Signature for Prognosis Prediction in Patients With Colon Cancer. Front Oncol 11:654076 (2021).

GSTA
Ref: Liu C et al. MiR-144-3p promotes the tumor growth and metastasis of papillary thyroid carcinoma by targeting paired box gene 8. Cancer Cell Int 18:54 (2018). WB, IHC-P ; Human .

KRT7 Ref: Chen H et al. Potential role of FoxO3a in the regulation of trophoblast development and pregnancy complications. J Cell Mol Med 25:4363-4372 (2021).

KRT19: Ref: Chumduri C et al. Opposing Wnt signals regulate cervical squamocolumnar homeostasis and emergence of metaplasia. Nat Cell Biol 23:184-197 (2021).

A1AT: ref: Clausen PP. Immunohistochemical demonstration of alpha-1-antitrypsin in liver tissue. A methodological investigation. Acta path microbiol scand Sect A 1980,88:299-306.

PECAM1: Validated by manufacturer (Leica) using human lymphoma tissue (membrane staining of endothelial cells) ß– images available at manufacturer's website

Anti-CD45 Validated by manufacturer (Agilent, Dako) using tonsillar tissue – images available at manufacturer's website.

Anti-CD68 - Validated by manufacturer (Agilent, Dako) using tonsillar tissue – images available at manufacturer's website.

Anti-CK8/18 - Validated by manufacturer (Leica) using colon mucosa cells – images available at manufacturer's website.

Anti-CK19 - Validated by manufacturer (Leica) using rectal adenocarcinoma – images available at manufacturer's website.

Anti-human hepatocyte Ref: 1. Minervini MI, Demetris AJ, Lee RG, Carr BI, Madariaga J, Nalesnik MA. Utilization of hepatocyte-specific antibody in the immunocytochemical evaluation of liver tumors. Mod Pathol 1997,10:686-92.

Anti-VIM: Validated by manufacturer (Agilent, Dako) using primary liver tissue – images available at manufacturer's website.

A1AT Ref: 1. Clausen PP. Immunohistochemical demonstration of alpha-1-antitrypsin in liver tissue. A methodological investigation. Acta path microbiol scand Sect A 1980,88:299-306.

AFP: Validated by manufacturer (Agilent, Dako) using embryonal carcinoma – images available at manufacturer's website.

AFP: Validated by manufacturer (Sigma Aldrich) using hepatocellular carcinoma – images available at manufacturer's website.

KRT18 Ref: Lee, J. H., Lee, S., et al.. Human Liver Stem Cell Transplantation Alleviates Liver Fibrosis in a Rat Model of CCl4-Induced Liver Fibrosis. International Journal of Stem Cells. 30 November 2021

HNF4a. Ref: SH Lee, V Veeriah, F Levine. Liver fat storage is controlled by HNF4-alpha through induction of lipophagy and is reversed by a potent HNF4-alpha agonist.
Cell Death & Disease, 2021;12(6):603.

SOX9 – Ref: Russell JP, Lim X, Santambrogio A, Yianni V, Kemkem Y, Wang B, Fish M, Haston S, Grabek A, Hallang S, Lodge EJ, Patist AL, Schedl A, Mollard P, Nusse R, Andoniadou CL. Pituitary stem cells produce paracrine WNT signals to control the expansion of their descendant progenitor cells. Elife. 2021 Jan 5;10:e59142.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	CA1ATD - published in Yusa & Rashid et al., 2011 FSPS13B - Wellcome Sanger Institute HipSci initiative, UK YEMZ - Wellcome Sanger Institute HipSci initiative, UK KOLF - Wellcome Sanger Institute HipSci initiative, UK HEK 293T - ATCC (CRL-11268) HUVEC - Lonza (C2519A) Human fetal hepatoblast - University of Cambridge, UK (see below) None of the cross-contaminated or misidentified cell lines on the list maintained by the International Cell Line Authentication Committee (ICLAC) were used in this study.
Authentication	Genotyping, morphology, and pluripotent ability to differentiate into 3 germ lineages.

Mycoplasma contamination	All cell lines were tested for mycoplasma and found negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Female FRGN (Fah-/-, Rag2-/-, Il2rg-/-, NOD) mice aged 12-20 weeks old (Yecuris). Mice were kept in 16h light/24h , temperature 20-25.5ºC, and humidity 30-70%.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field collected samples were used in the study.
Ethics oversight	All surgical procedures were conducted according to protocol 4388-01 approved by the University of Washington Institutional Animal Care and Use Committees. 7 female FRGN (Fah−/−, Rag2−/−, Il2rg−/−, on a NOD background) mice aged 14 to 18-weeks were used for implant procedures.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Adult liver samples were collected from deceased, declined transplant patients that were screened for infection and free of liver disease (healthy). All fetal liver samples were collected from elective terminations and ranged from 5-17 post-conceptional weeks (PCA) in age. All samples used, and their accession numbers, are provided in Extended Data Fig. 2c. All clinical characteristics for these patients are included in the respective repositories.
Recruitment	Human adult liver samples were obtained from deceased organ donors declined for transplant from the Cambridge Biorepository for Translational Medicine (CBTM) at Addenbrooke's Hospital, Cambridge. Human fetal liver samples were obtained from patients undergoing elective terminations at Addenbrooke's Hospital, Cambridge (REC-96/085)
Ethics oversight	Primary human adult liver samples were obtained from deceased transplant organ donors (National Research Ethics Committee East of England – Cambridge South 15/EE/0152) following consent by family following organ donation procedure. Primary human adult liver biopsies were collected from living patients under ethical consent to be used in research (North West - Preston Research Ethics Committee 14/NW/1146) following informed consent by the patient. Primary human fetal liver tissue was obtained from patients undergoing elective terminations (East of England - Cambridge Central Research Ethics Committee, REC-96/085) after informed consent. Detailed information was provided on the objective of the research. Ethical approval covers the use of liver tissues for research and follow the UK rules for the use of human foetal tissues which exclude compensation for tissue donation. All human tissue was used after obtaining informed consent for use in research following the organ transplantation procedure.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Adult liver tissue or in vitro organoids were dissociated as described in Methods. Adult liver tissue was collected from deceased transplant organ donors and dissociated using enzymatic and mechanical means to obtain hepatocyte and nonparenchymal single-cell populations, whilst in vitro organoids were removed from matrigel using cell recovery solution and dissociated using TrypLE enzymatic digestion. Each fraction was pelleted by centrifuging at 350 x g for 3 mins, resuspended in PBS + 1% w/v BSA, and centrifuged again at 350 x g for 3 mins to pellet. The cells were resuspended in FIX & PERM solution (ThermoFisher GAS003), incubated for 20 mins at room temperature in the dark and subsequently centrifuged at 350 x g for 3 mins. The cells were washed twice using PBS + 1% w/v BSA and a third time using PERM & WASH solution. After the third wash, the cells were

	pelleted by centrifuging at 350 x g for 3 mins and resuspended in 1:100 of primary antibody diluted in PERM & WASH. The primary antibody was incubated on the cells 1 hr to overnight and washed 3 times with PERM & WASH. The secondary antibodies were diluted 1:1,000 in PERM & WASH and used to resuspend the cell fractions. After 1 hour of incubation at room temperature in the dark, the cells were washed 3 times in PERM & WASH and resuspended in PBS for flow cytometry analysis.
Instrument	Flow cytometry was performed on the Beckman Coulter CyAn ADP instrument.
Software	FlowJo (v10.4.2)
Cell population abundance	The abundance of EPCAM+/A1AT+ and EPCAM+/AFP+ cells for single-cell RNA sequencing and in vitro organoid formation was determined from the FACS sorting, with the total population of cells counted after the sort. The purity of each suspensions was determined through propagation of the organoid cultures and subsequent QPCR and IF validations. Single-cell RNA sequencing also confirmed the effectiveness of isolating these populations.
Gating strategy	For all flow experiments, cells were identified from cellular debris based on the FSC/SSC plot and singlets were discriminated from doublets by gating the lower population on the pulse width axis (FSC-W). The expression of EPCAM, A1AT and AFP was determined by staining with the indicated antibodies. The voltages and gates for this analysis were set according to cells stained with the respective secondary antibodies alone. All EPCAM+ stained cells were used for single-cell RNA sequencing, to ensure that the resulting data was representative of the entire EPCAM+ fraction of the liver (primarily consisting of hepatoblasts). Quadrant gating for EPCAM+/A1AT+ and EPCAM+/AFP+, as to include all cells both positively and negatively stained, are shown in Supplementary Fig. 3d.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.