Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells

Keren Bahar Halpern1,4, Rom Shenhav1,4, Hassan Massalha1, Beata Toth1, Adi Egozi1, Efi E Massasa1, Chiara Medgalia2, Eyal David2, Amir Giladi2, Andreas E Moor1, Ziv Porat3, Ido Amit2 & Shalev Itzkovitz1

Spatially resolved single-cell RNA sequencing (scRNAseq) is a powerful approach for inferring connections between a cell’s identity and its position in a tissue. We recently combined scRNAseq with spatially mapped landmark genes to infer the expression zonation of hepatocytes. However, determining zonation of small cells with low mRNA content, or without highly expressed landmark genes, remains challenging. Here we used paired-cell sequencing, in which mRNA from pairs of attached mouse cells were sequenced and gene expression from one cell type was used to infer the pairs’ tissue coordinates. We applied this method to pairs of hepatocytes and liver endothelial cells (LECs). Using the spatial information from hepatocytes, we reconstructed LEC zonation and extracted a landmark gene panel that we used to spatially map LEC scRNAseq data. Our approach revealed the expression of both Wnt ligands and the Dkk3 Wnt antagonist in distinct pericentral LEC sub-populations.

An outstanding challenge in biology is the characterization of the cell types that make up mammalian tissues1–3. Given that the coordinates of a cell in a tissue are a critical determinant of its molecular identity, approaches for spatial transcriptomics are necessary to resolve the connection between location and function4–13. In the mammalian liver, hepatocytes and diverse non-parenchymal cells (NPCs) operate in repeating, hexagonal-shaped anatomical units known as lobules (Fig. 1a). Each lobule is comprised of a central vein, radial sinusoidal networks, and portal nodes that consist of arteries, veins and bile ducts. The lobule blood vessels, which are adjacent to hepatocytes, are lined with LECs and contain diverse resident and circulating immune cells. Blood emanates from the portal nodes and flows toward draining central veins, creating gradients of oxygen, nutrients and hormones14,15. In addition, morphogens, such as Wnt and Rspo3, secreted by central vein LECs create an inverse polarizing field16–18. The graded lobule microenvironment gives rise to spatial division of labor among hepatocytes residing at different radial coordinates19–21. Whether the liver NPCs exhibit similar spatial division of labor is unknown.

LECs make up about 50% of the tissue’s NPCs22 and have critical functions: they form the building blocks of the blood vessels, clear endotoxins, bacteria and other compounds, regulate host immune responses to pathogens, present antigens, and secrete morphogens that shape hepatocyte gene expression22–24. Several studies have identified morphological differences in LECs located at different lobule radial coordinates, including the amounts and sizes of LECs fenestrae and of the cells themselves14. However, we lack a comprehensive picture of LEC spatial diversity in terms of their gene expression signatures.

We recently used spatially resolved single-cell transcriptomics to reconstruct the zonation patterns of all hepatocyte genes9. We used massively parallel single-cell RNA-Seq (MARS-Seq25) to sequence thousands of hepatocytes and constructed a concise panel of zonated hepatocyte landmark genes, quantified with single molecule fluorescence in situ hybridization (smFISH), to retrospectively map the hepatocytes back to their original radial lobule layers. A weighted-average of the hepatocytes’ expression in each layer yielded the expression profiles of all hepatocyte genes. Uncovering similar zonation patterns of LECs is much more challenging. Given that LECs are small cells, and scRNAseq techniques sparsely sample only a small fraction of the total cellular mRNA, transcripts of most genes will not be present in individual LECs. Consequently, a suitable landmark gene panel would need to be large. As we lack prior knowledge regarding the zonation patterns of more than a handful of LEC genes, new approaches will need to be developed to reveal such a suitable landmark gene panel.

To overcome these limitations, we developed paired-cell RNA sequencing (pcRNAseq), which profiles gene expression of hepatocytes and adjacent LECs that are attached to them in the tissue and determines their localization in tissues using the expression of hepatocyte landmark genes. In this way, the spatial zonation patterns of LEC genes can be resolved with high spatial resolution (Fig. 1).

Results

A cell atlas of liver NPCs

To analyze the identities of NPCs, we sorted single perfused mouse liver cells labeled with the pan-immune surface marker CD45 and cells labeled with the endothelial surface marker CD31. We then used MARS-Seq25 to measure the global gene expression of 3,151...
sorted single cells (Fig. 2, Supplementary Fig. 1, and Supplementary Data 1 and 2). The cells clustered into seven groups (Fig. 2a): cells that were enriched for markers of endothelial cells (Fig. 2b), T cells, (Fig. 2c), plasmacytoid dendritic cells (pDCs; Fig. 2d), Kupffer cells (Fig. 2e), liver capsule macrophages (LCM; Fig. 2f), B cells (Fig. 2g) and neutrophils (Fig. 2h).

We used a database of known ligand-receptor pairs (Fig. 2 and Supplementary Data 3). This analysis highlighted LECs as communication hubs that interact with all other liver cell types (Supplementary Fig. 2a), as expected based on their adjacency to hepatocytes on one side and immune cells on the other. Notable ligand receptor pairs included the LEC ligand Csf1 and its Kupffer cell receptor Csf1r, the Kupffer cell ligand C1qa and its LEC receptor Cd93, the LEC-hepatocyte pairs Rspo3-Lgr4 and Wnt2-Fzd8 (Supplementary Fig. 2b and Supplementary Data 3), the LEC-LEC juxtacrine signaling pairs Dll4-Notch1 and Efnb2-Ephb4, and the hepatocyte complement C4a and its Kupffer cell receptor C3ar1 (Supplementary Fig. 2c).

Paired-cell sequencing of hepatocytes and attached LECs
To test whether LEC genes exhibit spatial zonation in their expression along the radial lobule axis, we examined the expression of LEC genes that were previously shown to be zonated: the pericentral Rspo3, Wnt2 and Wnt9b (16–18). The unique molecular identifier (UMI) levels of these genes were too low to enable robust spatial inference of the single sequenced cells (less than 200 of the 1,203 LECs had more than a single UMI of any of these genes). The total numbers of UMIs in LECs were about 23-fold lower than in single hepatocytes sequenced with the same technology (a median of 470 total UMIs in LECs versus a median of 10,710 total UMIs in hepatocytes). Given that these few hundred UMIs were spread over several thousand LEC-expressed genes, we concluded that robust spatial inference with a small panel of landmark genes would be infeasible. We therefore sought to isolate pairs of hepatocytes and directly adjacent LECs, with the aim of identifying their location by the expression of highly abundant hepatocyte landmark genes.

We dissociated liver tissue with collagenase D, an enzyme that we found to be less efficient in tissue dissociation than enzymes such
as Liberase\(^9\) (Supplementary Fig. 3a–c). We used fluorescence-activated cell sorting (FACS) to isolate hepatocytes by gating their size and enriched for cells that also expressed CD31 (Fig. 3a–c and Supplementary Fig. 3). This sorting strategy enriched for hepatocyte-LEC pairs (69%: Fig. 3d,e and Supplementary Fig. 3d). We performed MARStSeq on the isolated pairs and computationally filtered out wells that did not contain both hepatocyte and endothelial cell markers (Online Methods), resulting in a dataset of 4,602 paired-cells (Supplementary Data 4).

To determine whether the hepatocyte-LEC pairs sequenced are directly adjacent in the tissue, rather than coming into contact during the cell isolation procedure, we selected genes that have been shown to be zonated and that were expressed in LECs and not in hepatocytes (Supplementary Data 2): the pericentral genes Rspo3 and Wnt9b\(^16\)–\(^18\) and the periportal genes DLL4 and Efnb2 (ref. 24). Pairs that had transcript reads for either Rspo3 or Wnt9b had significantly higher expression of pericentral hepatocyte genes and lower expression of periportal genes. Conversely, pairs that had transcript reads for either Efnb2 or DLL4 had significantly higher expression of periportal hepatocyte genes and lower expression of pericentral genes (Fig. 3e).

In addition, we sorted single CD31\(^+\) endothelial cells from a mouse constitutively expressing EGFP and hepatocytes from a mouse constitutively expressing DsRed and incubated them together. Following sorting of the mixture, we observed <0.15% CD31\(^+\)CD45\(^-\) pairs (0 cells in paired-cell optimized sorting gate out of 713 analyzed cells versus 1,626 in the same paired-cell gate out of 87,872 analyzed cells, Fisher test, \(P = 3.72 \times 10^{-6}\); Fig. 3c and Supplementary Fig. 3i). These findings suggest that the sequenced hepatocyte-LEC pairs were directly adjacent in the tissue.

To infer the radial lobule coordinates of the sequenced pairs of cells, we assigned each pair a scaled coordinate denoted \(\eta\), which was based on the ratio of the summed expression of 21 pericentral and 30 periportal hepatocyte landmark genes (Supplementary Fig. 4a,b). These zonated landmark genes\(^8\) were selected on the basis of their high expression levels and low inter-mouse variability (Online Methods). Pairs of cells that originated in the pericentral layers had a small \(\eta\) value, whereas pairs from the periporal layers had a large \(\eta\) value (Supplementary Fig. 4c). We used our previous spatial reconstruction of hepatocytes\(^9\) to map each \(\eta\) value to the concentric radial lobule layers (Supplementary Fig. 4d and Online Methods). Finally, to obtain the average expression of each gene in each lobule layer, we summed the expression of all pairs, weighted by their probability to belong to that layer, based on their \(\eta\) values (Supplementary Fig. 4d and Online Methods). This algorithm yielded a zonation table of 33,856 genes (Supplementary Data 5).

Our zonation inference algorithm recapitulated the previous global zonation patterns of hepatocyte genes (Supplementary Fig. 5). Moreover, the reconstructed zonation profiles of hepatocyte genes in the paired-cell data overlapped the zonation profiles that we previously validated using smFISH (Fig. 3f).

**Global zonation of LEC gene expression**

We next turned our attention to LEC genes. We focused on genes that had a relative expression that was at least 23-fold higher in single LECs (Fig. 2) than in single hepatocytes\(^9\) (Supplementary Data 2), so that most of the transcripts in the sequenced pairs would have originated from LECs (Online Methods). To avoid the erroneous inference of a gene’s zonation being based on its zonation in hepatocytes, we also excluded genes that were zonated in single hepatocytes\(^9\) (Online Methods). We further removed genes that were highly expressed in the single immune cells (Fig. 2) and selected genes with a maximal zonation level higher than \(10^{-6}\) of the total cellular transcripts, resulting in 1,303 LEC-specific genes. Our analysis revealed notable spatial heterogeneity of LECs, with around 35% of these genes (475) being significantly zonated (Fig. 4a). We validated the predicted zonation patterns of 12 genes using smFISH (Fig. 4b,c and Online Methods).

Zonated LEC genes included the pericentral ligands Wnt2 and Wnt9b and the Rspo3 ligand (Fig. 4a,c). DLL4 (Fig. 4b,c), which has been shown to be enriched in arterial endothelial cells\(^30,31\), was peripolarly zonated (Fig. 4c), as were Efnb2 (ref. 32) and Cldn5 (ref. 33). Other genes, such as Ecm1, Lyve1 and Ccdn1, exhibited a
non-monotonic zonation pattern, with the highest expression being present in the mid-lobule layers (Fig. 4).

**Molecular signature of pericentral LECs**

The pericentral LECs have recently been shown to be essential for maintaining hepatocyte zonation through the specific secretion of Wnt ligands and Rspo3 (refs. 16–18). Our reconstructed LEC zonation profiles revealed a molecular signature for this pericentral niche, including, in addition to Rspo3, Wnt2 and Wnt9b, thrombomodulin (Thbd), Cdh13, Fabp4 and Kit (Figs. 4c and 5a). The smFISH validations revealed a bimodal expression pattern for these pericentral genes: Rspo3, Thbd and Cdh13 were highly expressed in both the pericentral LECs that line the sinusoidal channels, as well as in those that line the central vein (Figs. 4c and 5b), whereas Wnt2 and Kit were expressed at a higher level in the pericentral sinusoidal LECs and at lower levels in the LECs that line the central vein (Figs. 4c and 5c,d). Other genes, such as Cyp2e1, Dll4, Vhl, and Arg1, were enriched in the Rspo3 = 9.35 × 10^{-13} and depleted in the Rspo3 = 3.3 × 10^{-29} two-sided Wilcoxon rank-sum test). Expression is shown in units of fraction of total cellular UMIs. Box plot elements: center line, median; box limits, first to third quartile (Q1 to Q3); whiskers, extend to the most extreme data point within 1.5x the interquartile range (IQR) from the box; circles, data points. (f) Reconstructed zonation profiles of hepatocyte genes based on the pcRNAseq data overlap profiles validated with smFISH. Patches represent s.e.m. smFISH plots were based on n = 10 lobules from 2 mice, pcRNAseq plots were based on n = 4,602 paired cells.

![Figure 3](image-url)
as Bmp2 and Stab1, also showed specific repression in the pericentral LECs that line the central vein (Fig. 4c). Notably, we identified expression of the Wnt antagonist Dkk3 in a sub-population of pericentral LECs (Fig. 5a,d). This expression was anti-correlated with Wnt2 in the LECs that line the central vein (R = −0.4, P = 3.9 × 10⁻⁴, Fig. 5e).

Spatial reconstruction of single LECs

A limitation of pcRNAseq for spatial reconstruction is the inability to infer LEC zonation for genes that are also abundantly expressed in hepatocytes. To overcome this, we used pcRNAseq to extract a panel of landmark genes that could be used to retrospectively localize the single-sequenced LECs along the lobule radial axis (Fig. 6a,b). Given that individual LECs had very low levels of UMIs as a result of their small size, reliable inference required a large panel of 70 pericentral and 70 periportal LEC genes. As with the paired-cells, we assigned each sequenced LEC a scaled coordinate η that was based on the summed expression of these landmark genes (Online Methods). This coordinate correlated with the cells’ distances from the central vein.

We classified the single LECs into four spatially stratified populations based on the value of η (Online Methods). The lower spatial resolution of four radial layers, compared with the eight layers obtained with the pcRNAseq reconstruction, was required to obtain sufficient transcript representation for each layer. In addition, reliable reconstruction was only possible for 2,145 genes with sufficiently high expression, again as a result of the sparseness of single LEC gene expression (Online Methods).

Our reconstructed profiles (Supplementary Data 6) overlapped the ones obtained with pcRNAseq for the highly zonated LEC genes (Fig. 6c). However, this approach uncovered new zonated LEC profiles that included genes that were also expressed in hepatocytes. Examples include the Notch target Hes1 and Cts1 (Fig. 6d,e and Supplementary Data 6).

Spatial sorting of LECs

In light of the global spatial heterogeneity of LECs, we sought to identify zonated surface markers for prospective isolation of bulk...
Figure 5  Expression signature of pericentral LECs. (a) Zonation profiles of representative pericentral genes, selected out of the 60 zonated LEC genes with the highest ratio of expression between pericentral layer 1 and pericentral layer 8. Patches represent s.e.m. Plots were based on n = 4,602 paired cells. (b) Rspo3 (green dots) was highly expressed in both LECs that line the central vein (CV, arrows) and in sinusoidal pericentral LECs (arrowheads). Scale bar represents 10 \( \mu \)m. (c) Kit (green dots) was repressed in LECs that line the central vein (arrows) and upregulated in sinusoidal pericentral LECs (arrowheads). Red dots are mRNAs of the LEC marker Aqp1, yellow blobs are hepatocyte lipofuscins that fluoresce in both red and green channels. Scale bar represents 10 \( \mu \)m. (d) Dkk3 (green dots) was expressed in a subset of LECs that line the central vein (green arrow), distinct from cells that expressed Wnt2 (red dots, cell marked by red arrow). Scale bar represents 5 \( \mu \)m. In b–d, micrographs are representative of ten lobules and two mice exhibiting similar results. (e) Expression of Dkk3 and Wnt2 was anticorrelated in LECs that line the central vein (Spearman \( R = -0.4, P = 3.9 \times 10^{-4}, n = 92 \) cells).
spatially resolved LEC populations. Indeed, our analysis revealed several zoned transcription factors (Supplementary Fig. 6a) and surface markers (Supplementary Fig. 6b). Kit, which encodes the CD117 surface marker, exhibited a zonated profile with a gradual decline in expression from the central vein towards the portal node (Figs. 4c and 5a). We used an anti-CD117 antibody to sort CD31⁺CD45⁻ LECs according to their CD117 levels. Quantitative PCR measurements of these bulk spatially sorted populations validated the identities of these spatially resolved LECs (Supplementary Fig. 6c,d). This sorting approach can therefore yield massive amounts of LECs from distinct lobule layers, which can be used for future interrogation of their genome, epigenome, proteome and

![Figure 6](image_url)

**Figure 6** Spatial reconstruction of single LECs using landmark genes obtained from pcRNAseq. (a,b) pcRNAseq-based zonation profiles of the panel of LEC landmark genes used to localize cells in the scRNAseq data (70 pericentral LEC genes, a; 70 periporal LEC genes, b). Profiles were normalized by their maximum. (c) Overlap of zonation profiles based on scRNAseq (blue, n = 1,163 single cells) and pcRNAseq (red, n = 4,602 paired cells). Patches represent s.e.m. Zonation profiles were normalized to their means across all layers (pcRNAseq profiles contain eight layers, whereas scRNAseq profiles contain four). (d) Ctsl (green dots) was highly expressed in hepatocytes (white dashed outline) and was also expressed in the adjacent LECs (arrows). Micrographs are representative of ten lobules and two mice exhibiting similar results. (e) scRNAseq spatial reconstruction revealed zonated expression of Ctsl in LECs, with a reduced expression level in pericentral LECs (1) compared with mid-lobule LECs (2). Box plot elements: center line, median; box limits, first to third quartile (Q1 to Q3) of the smFISH expression, horizontal lines are medians. Quantification was based on n = 30 cells from each layer from 2 mice. Gray patches mark the pericentral layers (CV, left) and periporal layers (PN, right). Blue represents the scRNAseq-based zonation profile. Patches are s.e.m. Panels on the left show representative images. Blue is DAPI nuclear stain. Scale bars represent 5 µm in d and e.
other cellular features that cannot currently be robustly measured at the single-cell level.

DISCUSSION

An outstanding challenge in spatial transcriptomics is the expression mapping of small non-parenchymal tissue cells. We developed paired-cell sequencing, a method that extracts spatial information from attached adjacent parenchymal cells. We used this approach to uncover a high degree of spatial heterogeneity of LECs, with around 35% of LEC genes zonated. We also inferred the molecular signature of pericentral LECs, an important liver niche that secretes key morphogens such as Wnts and Rspo3, and identified the expression of the Wnt antagonist Dkk3 in a sub-population of the LECs that line the central vein that was distinct from the Wnt-producing cells (Fig. 5d,e). Expression of both Wnt ligands and Dkk antagonists has been demonstrated in niche cells of the intestine and hair follicle. Our findings highlight a potential balancing effect of these positive and negative regulators of liver Wnt signaling by pericentral LECs.

To extend the zonation profiles to genes that are also expressed in hepatocytes, we extracted a large panel of LEC landmark genes from the pcRNAseq data and used it to localize the single LECs along the lobule radial axis. This revealed zonation profiles for genes that were masked by the hepatocyte expression in the paired-cell data. Although more genes could be examined with this approach, the spatial resolution of this reconstruction was lower (four layers versus eight layers for the pcRNAseq data). Technologies that provide more sequenced cells along with higher capture rate could facilitate a higher spatial resolution and statistical power for identifying additional zonated layer 1 LECs. Thus, this study demonstrated the utility of pcRNAseq not only for inferring zonation profiles of NPC-specific genes, but also as a method for unbiased detection of a large panel of NPC landmark genes, to be used in single-cell-based spatially resolved transcriptomics.

Although our reconstruction had high spatial resolution, it did not capture the entire complexity of LECs, as there are limitations related to the parenchymal cells used for spatial inference. Endothelial cells are comprised of arterial, venous and sinusoidal cells. A hepatocyte at a given radial lobule layer is often spatially adjacent to more than one type of LEC. For example, sorted pericentral hepatocytes can either carry an attached pericentral LEC that lines the radial sinusoids or one that lines the central vein (Fig. 5b,c). Thus, our layer 1 LEC expression is an average of these distinct pools of cells. Indeed, using smFISH measurements, we found that some pericentral LEC genes, such as Rspo3 and Cdh13, were highly expressed in the LECs that line the central vein, whereas others, such as Wnt2 and Kit, were more highly expressed in the pericentral sinusoidal LECs. Moreover, LECs at the same tissue location can exhibit distinct sub-populations and additional variability, as exemplified in the pericentral LECs that express either Dkk3 or Wnt2. The fact that the spatial structure of LECs did not naturally emerge from the single-cell data (Fig. 2b) may also indicate that LEC identity is defined by a complex overlay of different factors, with the lobule coordinate being only one. Our spatial blueprint should therefore serve as a basis for exploring these additional layers of variability in individual layers.

Modeling hepatocyte functions ex vivo has been a challenging field because of the difficulties of emulating the complex liver microenvironment. For example, the expression of the pericentrally zonated xenobiotic metabolism enzymes is often lost in several hours after extraction of hepatocytes, but partially retrieved by co-culturing hepatocytes with LECs. Our identification of zonated LEC surface markers such as CD117 (Supplementary Fig. 6) could be used to obtain co-cultures of hepatocytes and spatially stratified LECs, which could potentially be even more efficient in reconstructing zonated hepatocyte functions of interest ex vivo. Such spatial sorting could also be used to explore other cellular features of spatially stratified LEC populations, such as histone modifications, DNA methyllations and protein content.

pcRNAseq can be generically applied to other tissues and cell types. In the liver, natural candidates for similar reconstructions are hepatic stellate cells, which are physically adjacent to hepatocytes in the space of Disse (Supplementary Fig. 7e and Supplementary Data 7). Enteroctyes, the most abundant epithelial cells in the intestine, have recently been shown to exhibit global expression gradients along the crypt-villus axis. Several important intestinal cell types, such as enteroidendocrine cells, goblet cells, tuft cells and intra-epithelial lymphocytes, are interleaved and physically attached to enterocytes in the epithelial sheet (Supplementary Fig. 7a–d). PcrNAseq could readily be used to uncover the spatial expression gradients of these cells. Other examples of relevant cell pairs include neurons and either astrocytes, glia and microglia in the brain, a tissue in which neurons exhibit broad zonation, as well as peripheral neurons and diverse cell types such as hepatocytes, enterocytes and adipose cells (Supplementary Data 7). PcrNAseq can also be used to explore the interactions between diverse tumor cell populations and their adjacent stromal neighbors, such as cancer-associated fibroblasts or immune cells (Supplementary Fig. 7f). In such samples, the ‘landmark genes’ expressed in the cancer cells could either intrinsically encode spatial landmarks, such as core versus periphery of the tumor or adjacency to blood vessels, or alternatively clonally or transcriptionally distinct populations that do not have a clear spatial correlate.

For each tissue and for each pair of cell types, pcRNAseq would require optimization of the dissociation protocols to enable isolation of pairs rather than single cells or clumps of more than two cells (Supplementary Fig. 3). Combining the approach with DNA stains such as Hoechst could facilitate pair-enrichment in cases in which FACs cell size properties are insufficient. Moreover, cell-type-specific surface markers could be used for the enrichment of the relevant cell types. In summary, pcRNAseq opens avenues for detailed spatial characterization of cells in diverse tissues.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank M. Kolesnikov and S. Jung (Weizmann Institute) for the C57BL/6-Acrb-DsRed.T3 mice and D. Jaitin (Weizmann Institute) for the C57BL/6-Tg(CAG-EGFP) mice. S.I. is supported by the Henry Chanoch Kreter Institute for Biomedical Imaging and Genomics, The Leir Charitable Foundations, Richard Jakubskind Laboratory of Systems Biology, Cymerman-Jakubskind Prize, The Lord Sief of Brompton Memorial Fund, the I-CORE program of the Planning and Budgeting Committee and the Israel Science Foundation (grants 1902/12 and 1796/12), the Israel Science Foundation grant No. 1486/16, the EMBO Young Investigator Program and the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement number 335122, the Bert L. and N. Kuggie Vallee Foundation and the Howard Hughes Medical Institute (HHMI) international research scholar award. S.I. is the incumbent of the Philip Harris and Gerald Ronson Career Development Chair.

AUTHOR CONTRIBUTIONS

K.B.H., C.M., R.T., A.E., A.E.M., E.E.M. and Z.P. performed the experiments. R.S., S.I., H.M., A.G. and E.D. performed the data analysis. I.A. contributed to project...
design. S.I. supervised the study. S.L., K.B.H. and R.S. wrote the paper. All of the authors discussed the results and commented on the manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Regev, A. et al. The Human Cell Atlas. eLife 6, e27041 (2017).
2. Tabula Muris Consortium, Quake, S.R., Wys-Coray, T. & Darmanis, S. Transcriptomic characterization of 20 organs and tissues from mouse at single cell resolution creates a Tabula Muris. Preprint at https://www.biorxiv.org/content/eary/2017/12/20/237446 (2017).
3. Han, X. et al. Mapping the Mouse Cell Atlas by Microwell-Seq. Cell 172, 1091–1107.e17 (2018).
4. Crosetto, N., Bresolin, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. Nat. Rev. Genet. 16, 57–66 (2015).
5. Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C. & Teichmann, S.A. The technology and biology of single-cell RNA sequencing. Mol. Cell 58, 610–620 (2015).
6. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. & Regev, A. Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502 (2015).
7. Achim, K. et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat. Biotechnol. 33, 503–509 (2015).
8. Karaiskos, N. et al. The Drosophila embryo at single-cell transcriptome resolution. Science 358, 194–199 (2017).
9. Halpern, K.B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature 542, 352–356 (2017).
10. Ståhl, P.L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353, 78–82 (2016).
11. Lein, E., Born, L.E. & Linnarsson, S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science 358, 64–69 (2017).
12. Lee, J.H. De Novo Gene Expression Reconstruction in Space. Trends Mol. Med. 23, 583–593 (2017).
13. Moor, A.E. & Itzkovitz, S. Spatial transcriptomics: paving the way for tissue-level systems biology. Curr. Opin. Biotechnol. 46, 126–133 (2017).
14. Jungermann, K. & Kietzmann, T. Zonation of parenchymal and non-parenchymal metabolism in liver. Annu. Rev. Nutr. 16, 179–203 (1996).
15. Gebhardt, R. Metabolic zonation of the liver: regulation and implications for liver function. Pharmacol. Ther. 53, 275–354 (1992).
16. Wang, B., Zhao, L., Fish, M., Logan, C.Y. & Nusse, R. Self-renewing diploid Axin(2+) cells fuel homeostatic renewal of the liver. Nature 524, 180–185 (2015).
17. Planas-Paz, L. et al. The RSPO-LGR4/5-ZNRF3/RNF43 module controls liver zonation and size. Nat. Cell Biol. 18, 467–479 (2016).
18. Rocha, A.S. et al. The angionecrosis factor rondonin is a key determinant of liver zonation. Cell Reports 13, 1757–1764 (2015).
19. Braeuning, A. et al. Differential gene expression in periportal and perivenous mouse hepatocytes. FEBS J. 273, 5051–5061 (2006).
20. Gebhardt, R. & Matz-Soja, M. Zonation: novel aspects of its regulation and its impact on homeostasis. World J. Gastroenterol. 20, 8491–8504 (2014).
21. Ciolot, S. & Perret, C. Liver Zonation. in Molecular Pathology of Liver Diseases (ed. Monga, S.P.S.) 7–16 (Springer US, 2011).
22. Aird, W.C. Phenotypic heterogeneity of the endothelium. II. Representative vascular beds. Circ. Res. 100, 174–190 (2007).
23. Strauss, O., Phillips, A., Ruggiero, K., Barlett, A. & Dunbar, P.R. Immunofluorescence identifies distinct subsets of endothelial cells in the human liver. Sci. Rep. 7, 44356 (2017).
24. Rafii, S., Butler, J.M. & Ding, B.S. Angiocrine functions of organ-specific endothelial cells. Nature 529, 316–325 (2016).
25. Jatin, D.A. et al. Massively parallel single-cell RNA-seq for marker-free deconvolution of tissues into cell types. Science 343, 776–779 (2014).
26. Chistiakov, D.A., Orekhov, A.N., Sobenin, I.A. & Bobryshev, Y.V. Plasmacytoid dendritic cells: development, functions, and role in atherosclerotic inflammation. Front. Physiol. 5, 279 (2014).
27. Sierra, F. et al. A liver capsular network of monocyte-derived macrophages restricts hepatic dissemination of intraperitoneal bacteria by neutrophil recruitment. Immunity 47, 374–388.e6 (2017).
28. Graeber, T.G. & Eisenberg, D. Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. Nat. Genet. 29, 295–300 (2004).
29. Zhou, J.X., Taramelli, R., Pedrini, E., Knijnenburg, T. & Huang, S. Extracting intercellular signaling network of cancer tissues using ligand-receptor expression patterns from whole-tumor and single-cell transcriptomes. Sci. Rep. 7, 8815 (2017).
30. Shutter, J.R. et al. Dl14, a novel B7-related ligand expressed in arterial endothelium. Genes Dev. 14, 1313–1318 (2000).
31. Hellström, M. et al. Dl14 signalling through Notch1 regulates formation of tip cells during angiogenesis. Nature 445, 776–780 (2007).
32. Adam, R.H. & Klein, R. Eph receptors and ephrin ligands. essential mediators of vascular development. Trends Cardiovasc. Med. 10, 183–188 (2000).
33. Rahner, C., Mitic, L.L. & Anderson, J.M. Heterogeneity in expression and subcellular localization of claudins 2, 3, 4, and 5 in the rat liver, pancreas, and gut. Gastroenterology 120, 411–422 (2001).
34. FANTOM Consortium and the RIKEN PMI and CLST (DGT) & Forrest, A.R. et al. A promoter-level mammalian expression atlas. Nature 507, 462–470 (2014).
35. Moore, K.A. & Lemischka, I.R. Stem cells and their niches. Science 311, 1880–1885 (2006).
36. Gregorieff, A. et al. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology 129, 626–638 (2005).
37. Sick, S., Reinker, S., Timmer, J. & Schlake, T. Wnt and DKK determine hair follicle spacing through a reaction-diffusion mechanism. Science 314, 1447–1450 (2006).
38. Macosko, E.Z. et al. Highly parallel genome-wide expression profiling of individual cells using droplet droplets. Cell 161, 1202–1214 (2015).
39. Klein, A.M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201 (2015).
40. Bagnoli, J.W. et al. mcSCRB-seq: sensitive and powerful single-cell RNA sequencing. Preprint at https://www.biorxiv.org/content/early/2017/10/18/188367 (2017).
41. Vanlandewijck, M. et al. A molecular atlas of cell types and zonation in the brain vasculature. Nature 554, 475–480 (2018).
42. Godoy, P. et al. Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. Arch. Toxicol. 87, 1319–1530 (2013).
43. Hernandez-Gea, V. & Friedman, S.L. Pathogenesis of liver fibrosis. Annu. Rev. Pathol. 6, 425–456 (2011).
44. Moor, A.E. et al. Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. Preprint at https://www.biorxiv.org/content/early/2018/02/07/261529 (2018).
45. Olivares-Villagomez, D. & Van Kaer, L. Intestinal intraepithelial lymphocytes: sentinels of the mucosal barrier. Trends Immunol. 39, 264–275 (2018).
46. Boisset, J.-C. et al. Mapping the physical network of cellular interactions. Nat. Methods 15, 547–553 (2018).
47. Shah, S., Lubeck, E., Zhou, W. & Cai, L. seqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. Neuron 94, 752–758.e1 (2017).
48. Codemardini, S. et al. Spatial organization of the somatosensory cortex revealed by cyclic smFISH. Preprint at https://www.biorxiv.org/content/early/2018/03/04/276097 (2018).
49. Zeng, W. et al. Sympathetic neuro-adipose connections mediate leptin-driven lipolysis. Cell 163, 84–94 (2015).
50. Bhowmick, N.A. & Moses, H.L. Tumor-stroma interactions. Curr. Opin. Genet. Dev. 15, 97–101 (2005).
ONLINE METHODS

Mice and tissues. All animal studies were approved by the Institutional Animal Care and Use Committee of WIS. C57/b6 male mice age 6–16 weeks were housed under reverse phase cycle, and fasted for 2 h starting at 8 a.m. (note that the change from 5 h fasting in ref. 9 to 2 h fasting did not changed the spatial reconstruction (Supplementary Fig. 5a)). C57BL/6-Tg(CAG-EGFP)51 and C57BL/6-Actb-DsRed.T3 (ref. 52) mice were used for the dual color FACS (Supplementary Fig. 3h). All mice were anesthetized with an intraperitoneal injection of a ketamine (100 mg/kg) and xylazine (10 mg/kg) mixture. For smFISH, liver tissues were harvested and fixed in 4% paraformaldehyde for 3 h, incubated overnight with 30% sucrose in 4% paraformaldehyde and then embedded in OCT. 7 µm cyrosections were used for hybridization. Mouse liver cells for RNAseq were extracted from five mice (3 for pcRNAseq and 2 for scRNAseq). All smFISH quantifications were performed on at least 2 mice. smFISH images in Figure 5 were done on ad libitum mice.

Antibodies used in this study. The following antibodies were used for cell isolation: CD31 (APC 102510, PE-Cy7 102418), CD45 (APC-Cy7 103116, PE-Cy7 103114), CD3 (PE 100205), CD19 (PE 152407). CD117 (Kit, APC 105815) All antibodies were purchased from BioLegend and were diluted 1:100 in the final antibodies staining solution.

Hybridization and imaging. Probe library constructions, hybridization procedures and imaging conditions were previously described33-34. SmFISH probe libraries (Supplementary Data 8) were coupled to Cy5 or Alexa594. Endothelial cells were detected by Aqp1 staining. To detect cell borders alexa fluor 488 conjugated phalloidin (Rhenium A12379) was added to the GLOX buffer wash34. Portal nodes were identified morphologically using the DAPI channel based on bile ductile, central vein was identified using smFISH for Glul in TMR, included in all hybridizations. For zonation validation profiles, images were taken as scans extending from the portal node to the central vein. Endothelial cells were classified into seven layers as follows: cells that were in contact with Glul+ cells and that lined the central vein were classified as layer 1, cells that were in contact with Glul+ cells but resided in the sinusoidal channels were classified as layer 2, cells that surrounded the portal vessels at a distance of up to 1 hepatocyte were classified as layer 6, cells that lined the inside of the vessels in the portal node were classified as layer 7, sinusoidal cells in the mid-lobeule area were assigned layers 3–5 using equal distances from layer 2 to layer 6. To validate the predicted zonation we selected 15 genes that were significantly zonated and had an average expression in the scRNAseq data of LECs of more than 10^-4 of the cellular UMI. Rspo3, Cd113, Thbd, Wnt2, Ki67, Bmpa, Bmp4, Ccnd1, Lymph, Stab1, Dll4, Zeb2, Efnb2, Efnb4 and Kif4. SmFISH did not yield a signal for Bmp4, Zeb2 and Kif4, whereas the remaining 12 genes are presented in Figure 4c. The results in Figure 4c were based on at least 30 cells from each layer and from two mice. Quantification of smFISH data was done using ImageM34. Dots were counted in the first 5 µm of the Z-stack, and divided by the segmented cell volume to obtain the mRNA concentration per cell.

Hepatocyte isolation. Mouse liver cells were isolated by a modification of a previously described two-step collagenase perfusion method32 from 2-h fasted, 6–7-week-old male C57/b6 mice for the pcRNAseq (three mice) or 3–4 months for the scRNAseq (two mice). Digestion step was performed with Collagenase D (Sigma 11088885001) for the pcRNAseq or Liberase Blendzyme 3 recombinant collagenase (Roche Diagnostics) for the scRNAseq according to the manufacturer's instruction. Isolated cells were stained and taken directly to sorting. We found that Collagenase D was superior to Liberase for isolating pairs, since it is less efficient in tissue dissociation (Supplementary Fig. 3).

Multispectral imaging flow cytometry analysis (ImageStream). Cells were imaged using a multispectral Imaging Cytometer (ImageStreamX Mark II Imaging flow-cytometer; Amnis, Part of EMD Millipore). At least 5 X 10^5 cells were collected from each sample and data were analyzed using the manufacturer's image analysis software (IDEAS 6.2; Amnis). Images were compensated for fluorescent dye overlap by using single-stain controls. Cells were gated for focused cells, using the Gradient RMS feature, as previously described26. Cropped cells were further eliminated by plotting the cell area of the bright field image against the Centroid X feature (the number of pixels in the horizontal axis from the left corner of the image to the center of the cell mask). Only cells negative for PI staining were included in the analysis. Cells were then gated for single cells or doublets, using the area and aspect ratio features of the bright field channel. Additional singles were further gated for area of the bright field and the aspect ratio of the Hoechst staining (normalized for intensity), and added to the 'single' population. CD31-positive cells were gated according to the intensity (total fluorescence within the image) and Max pixel (the highest intensity pixel within the image) values of the CD31 staining. We found that 65% of the cells were hepatocyte-LEC pairs, 23% were single hepatocytes, 8% had more than one hepatocyte attached to a single LEC (Supplementary Fig. 3d). We computationally selected only pairs that had both hepatocytes and endothelial cells markers (below). While pairs that had more than a single hepatocyte (that is, triplets of two hepatocytes and one LEC) could not be omitted computationally, they most probably originated from hepatocytes that were attached in the tissue. Since the hepatocyte gene expression is used to localize the pairs, such triplets should not introduce a significant bias.

Single-cell and paired-cell sorting. Paired cells were sorted with SORP-FACSARiaII machine using a 130-µm nozzle, single NPCs were sorted using 100-µm nozzle. Dead cells were excluded on the basis of 500 ng/ml Dapi incorporation. To enrich for hepatocytes a 1.5 ND filter was used, whereas a 1 ND filter was used to enrich for NPCs. To enrich for hepatocyte-LEC pairs, cells were gated according to size that match the hepatocytes distribution by FSC-A and SSC-A. The cells were next gated by FSC-W to remove clusters of hepatocytes, and also gated for CD45 negative CD31 positive staining to enrich for pairs of hepatocyte-endothelial cells. We selected the top 1.85% of cells according to CD31 staining, to ensure the inclusion of LECs. Cells were sorted into 384-well cell capture plates containing 2 µl of lysis solution and barcode poly(T) reverse-transcription (RT) primers for single-cell RNA-seq25. Barcoded single cell capture plates were prepared with a Bravo automated liquid handling platform (Agilent) as described previously24. Four empty wells were kept in each 384-well plate as a no-cell control during data analysis. To enrich for pericentral pairs, which are a minority due to the hexagonal lobule geometry, in 4 out of the 17 384-well plates of pairs sorted we collected pairs that were also positive for CD73, a hepatocyte-specific surface marker that we previously found to be pericentrically zonated38. Immediately after sorting, each plate was spun down to ensure cell immersion into the lysis solution, snap frozen on dry ice and stored at ~80 °C until processed.

Spatial sorting. Livers were perfused with Liberase Blendzyme 3 recombinant collagenase (Roche Diagnostics). Cells were enriched for NPCs by 3 min of centrifugation at 30g, supernatant was centrifuged at 300g for 5 min. Cell pellet was treated with red blood cell lysis buffer (Sigma R7757) according to the manufacturer's instruction. Isolated cells were stained and taken to sorting. Cells were gated for CD31 positive, CD45 negative population to enrich for LECs. CD117-positive LECs (relative to the fluorophore isotype control) were divided into four equal populations according to fluorescent levels where highest expression represented the most pericentral population (Supplementary Fig. 6). Cells were sorted into lysis buffer.

Quantitative PCR quantification. RNA was isolated from cell lysates by Dynabeads mRNA DIRECT Micro Kit (Invitrogen 61021) according to the manufacturer's instruction. RNA was reverse transcribe with Maxima H minus RT enzyme (Thermo scientific E0753) and subjected to quantitative PCR with fast SYBR green on selected genes (Supplementary Data 9).

MARS-seq library preparation. Single cell libraries were prepared, as described previously25. Briefly, mRNA from cells sorted into MARS-Seq capture plate were barcoded and converted into cDNA and pooled using an automated pipeline. The pooled sample was then linearly amplified by T7 in vitro transcription and the resulting RNA was fragmented and converted into sequencing ready library by tagging the samples with pool barcodes and Illumina sequences during ligation, reverse transcription and PCR. Each pool of cells was tested for library quality and concentration was assessed as described in25. Mapping of single-cell reads to mouse reference genome (mm9) was done using HISAT version 0.1.6-beta and reads with multiple mapping
For each single cell and for each gene we first subtracted the estimated background expression. Background was calculated for each 384-well plate separately, as the mean gene expression in the four empty wells and in 'drop-out' wells that had low signal. These drop-out wells were defined as wells with both a number of expressed genes that was lower than the minimal number of expressed genes in the empty wells and a total UMI count that was smaller than the maximal total UMI count in the empty wells. After subtraction, negative values were set to zero. Next, cells with total UMI counts lower than 200 or higher than 6,000 were removed.

We used Seurat v2.0.1 package in R⁷ to visualize and cluster the single cell RNAseq data (Fig. 2). Gene expression measurements (UMIs per gene) were normalized for each cell by the summed UMI, multiplied by a scale factor 10,000, and then log-transformed. To avoid undesired sources of variation in gene expression, we used Seurat to regress out cell-cell variation driven by batch, total number of UMIS, and mouse ID. For detection of variable genes we set a bottom cutoff of 0.2 and a top cutoff of 3.5 on the log-transformed average gene expression, as well as a bottom cutoff of 0.6 on the dispersion. Cell clustering was based on PCA dimensionality reduction using the first 15 PCs, and a resolution value of 0.5.

We used cell type-specific markers to interpret the resulting seven clusters, based on literature search and the Immunological Genome project database⁸, Ptprb, Igfbp7, Clec4g, Agpl (ref. 22, and Ehd3 were highly expressed in the endothelial cell cluster, C1qa, C1qb, C1qc, Clec4f and Csf1r in the Kupffer cell cluster, Cc15, Trbc2, Cd3e, Cd3d, Nkg7 and Thy1 in the T cell cluster, Igk, Cd22, Cd79b, Cd19, Cd79a, Ebf1 and Pax5 in the B cell cluster, S100a9, S100a8, Csf3r, Sipi, Selp and Retnlg in the neutrophil cluster, Bcl1a, Runx2, Ccr9, Siglec, Spih, and Irf8 in the Plasmacytoid Dendritic cell (pDC) cluster⁹ and Cd11c (Itgax) and Cx3cr1 in the liver capsular macrophage cluster⁷.

Ligand–receptor analysis. To examine the interactions between liver cell types we used Seurat v2.0.1 R package to obtain the clustering and tSNE coordinates of a combined dataset, that included the single cells sequenced in this study as well as the single cell sequencing data presented in ref. 9. Data in Seurat was regressed on UMI counts and mice (however, hepatocytes were regressed in a pooled manner). Data was normalized and clustered as described in the previous section for the single NPC data, with the exception that the resolution value here was 0.35. This analysis resulted in eight clusters representing endothelial cells, hepatocytes, kupffer cells. T cells, B cells, pDCs, LCMs and neutrophils.

Ligand-receptor murine pairs were extracted from ref. 28 (708 unique ligands and 691 unique receptors). For each gene g and each cluster c we calculated the average expression \( x_{cg} \). We next computed a Z-score, \( Z_{cg} \), representing the enrichment of each ligand and receptor in each cluster

\[
Z_{cg} = \frac{x_{cg} – \text{mean}(x_{cg})}{\text{std}(x_{cg})}
\]

where the means \( \text{mean}(x_{cg}) \) and \( \text{std}(x_{cg}) \) were taken over the eight cluster values. We next defined an interaction score as

\[
Z_{\text{interaction}} = \sqrt{Z_i^2 + Z_R^2}
\]

where, \( Z_i \) is the ligand Z-score for cluster c, and \( Z_R \) is the receptor Z-score for cluster c. This resulted in 6,689 interactions annotated by ligand, receptor, source cluster and target cluster (Supplementary Data 3, all pairs sheet). We next considered only interactions for which both the average ligand and expression in the source cluster, and the average receptor expression in the target cluster were above 0.1 and for which the fraction of positive cells for the ligand/receptor were higher than 2% in the source/target cluster respectively (note that expression values here are in units of Seurat's log-transformed normalized data). In addition we removed pairs in which either the ligand/receptor were expressed in less than ten cells in the source/target cluster. This resulted in 2,895 interactions (Supplementary Data 3, expression-filtered sheet). To highlight cluster-specific interactions we reported all interactions with ligand/receptor Z-score above 1.5 and \( Z_{\text{interaction}} > 2 \) (Supplementary Data 3, cluster-specific sheet). Visualization of representative pairs was done with Matlab (Supplementary Fig. 2). Supplementary Figure 2a shows the network of cluster interactions, visualized with Cytoscape.

Processing of paired-cell data. For the CD31+ hepatocyte pairs, we used the same background subtraction method used for the single NPCs. We selected cells with total number of UMIs between 2,500 and 50,000, and at least 500 expressed genes. For each well, we normalized the expression by the total UMIs in that well, so that the expression was in units of fraction of total UMIs to obtain the expression matrix \( D \in \mathbb{R}^{g \times h \times c} \). To ensure that each pair includes an hepatocyte we removed pairs with an expression less than 0.01 for Alb⁹. To ensure that each pair contains an endothelial cell, we excluded wells for which the summed expression of the endothelial markers: Ptprb, Igfbp7, Clec4g, Agpl and Ehd3 was lower than \( 5 \times 10^{-5} \). In addition, we excluded wells for which the summed expression of the Kupffer cell markers C1qa, C1qb, C1qc, Clec4f and Csf1r was higher than 0.0005. These threshold values were chosen as the 80% values in the respective clusters in the single NPC dataset, after correcting for the 23-fold dilution of NPC transcripts that occurs when they are mixed with the large hepatocytes. This procedure resulted in 4,602 paired cells that were retained for further analysis.

Selection of hepatocyte landmark genes for pcRNA-seq reconstruction. We selected a large panel of hepatocyte landmark genes for inferring the original lobule coordinates of the pairs based on our previous study of hepatocyte zonation⁷. We chose zonated genes with a maximum expression level of at least \( 2 \times 10^{-4} \) across layers, with a ratio between the maximum and minimum zonation values that exceeded 1.5. We excluded genes that were previously found to vary in a circadian manner (JTK q-value lower than 0.01 (ref. 59), to reduce mouse-to-mouse variability in the expression of the landmark genes. We also removed genes that had an average expression that was at least twofold higher in the LEC scRNAseq data compared to pcRNAseq of hepatocytes⁸. Given the 23-fold dilution of LEC transcripts in the pairs, this ensured that any LEC contamination among the landmark genes would not be higher than 10%. The pericentral signature genes were selected from the remaining gene list, as genes with maximum expression in the most pericentral layer and minimum in most periporal layer, as well as center of mass lower than layer 4.5 (21 genes, Supplementary Fig. 4a). The periporal genes were selected as genes with maximum expression in the most periporal layer and minimum in most central layer, as well as center of mass higher than 5 (30 genes; Supplementary Fig. 4b). While this panel did not include some classic periporal genes such as Pck1, Cps1 and Arg1, reconstruction is essentially unchanged when including these landmark genes (data not shown).

Zonation reconstruction algorithm. To reconstruct the zonation profiles from the pcRNAseq data we used the combined expression of the landmark gene (LM) panel. Each cell i was assigned a coordinate \( \eta_i \), which reflected its location along the radial lobule axis, where \( \eta_i = 0 \) was the most pericentral coordinate and \( \eta_i = 1 \) the most periporal coordinate. We normalized the expression of every landmark gene by the maximum across all cells, to avoid giving excessive weight to highly expressed hepatocyte genes. This resulted in a normalized expression matrix \( E \in \mathbb{R}^{g \times \#cells} \).

For each cell i, we divided the summed expression of the portal LM genes (pLM), by the summed expression of the central (cLM) and portal LM genes to yield a number \( x_i \), which we normalized between 0 and 1 to obtain \( \eta_i \)

\[
x_i = \frac{\sum_{g \in \text{pLM}} E_{g,i}}{\sum_{g \in \text{cLM}} E_{g,i} + \sum_{g \in \text{pLM}} E_{g,i}}
\]

\[
\eta_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad \text{where } X = \{x_1, x_2, ..., x_{\#cells}\}
\]
To map η values to lobule layers, we computed the η of the single hepatocyte from ref. 9. We combined lobule layers 8–9 to yield eight layers since there were only few cells in layer 9. We fitted a Gamma distribution to the histogram of η values for each layer L∈[1,8] (Supplementary Fig. 4d). We used this Gamma distribution to compute the probability that each pair ηi belongs to each lobule layer L, P(η=ηi,layeri,L).

According to Bayes' law

$$ P(\text{cell}|P(\text{layer}=L|\eta=\eta_L)) = \frac{P(\eta=\eta_i|layer=L) \cdot P(\eta=\eta_L)}{\sum_{j \in [1,8]} P(\eta=\eta_j|layer=j) \cdot P(\eta=\eta_L)} $$

(5)

We assumed a uniform prior P(\text{layer}=L) rather than a prior that incorporates the hexagonal geometry of the lobule\(^9\) since we used CD73 in some of the sorted plates to enrich for pericentral pairs.

$$ \forall i, L \in [1,8], P(\text{layer}=L) = 1/8 $$

(6)

We thus calculated a matrix of probabilities \( M_{i,L} = P(\text{layer}=L|\eta=\eta_L) \) such that

$$ \forall i, L \in [1,8], \sum_{j \in [1,8]} M_{i,j} = 1 $$

(7)

To transform \( M \) into a weight matrix \( W \in R^{9 \times cells \times layers} \), we divided each value by the sum of its column \( W_{j,L} = \frac{M_{j,L}}{\sum_{j \in [1,8]} M_{j,L}} \). The final zonation matrix \( Z \in R^{genes \times layers} \) was obtained by multiplying the weight matrix by the expression matrix \( Z = D \times W \). As in the expression matrix \( D \) the units of the zonation matrix \( Z \) are fraction of total cellular UMIs, however here they represent the average over all cells in the layer.

We used 500 bootstrap iterations to obtain standard errors for the mean zonation profiles. To assign zonation significance we extracted a summary statistic for each gene as the difference between the maximum and minimum values of the mean-normalized profile and compared it to the summary statistics in 1,000 datasets in which the cells' η was randomly reshuffled. We calculated Z-scores for the summary statistic of each gene and used the normal distribution to obtain p-values. We used Benjamini–Hochberg multiple hypothesis correction to assign a q-value for each profile.

Determining the set of genes expressed uniquely in endothelial cells. Our paired-cell sequencing yields UMI counts that may originate from either a hepatocyte or its paired endothelial cell. Since around 50% of the genes expressed in hepatocytes exhibit zonation\(^9\) we first excluded genes that were found to be zonated in hepatocytes\(^9\), as well as genes with a maximal zonation level lower than 10\(^{-4}\) in the endothelial zonation matrix Z. For each gene we next calculated the mean over all cells of the expression (in fraction of total UMIs) in single hepatocytes\(^9\) and in every NPC cluster in the scRNAseq data. We removed genes for which the mean expression in one of the non-endothelial cell types was higher than fivefold the mean in the endothelial scRNAseq cluster. When computing hepatocyte mean expression in the data of\(^6\) we removed cells with a summed expression higher than 5 × 10\(^4\) of endothelial markers (Ptprb, Iglbp7, Clec4g, Aqp1, Ehd3) or Kupffer cell markers (C1qa, C1qb, C1qc, Clec4f, Csf1r). We computed the ratio between the median total number of UMIs in single hepatocytes (\( T_{H} = 10,710 \) and in single endothelial cells \( T_{E} = 470, \text{UMI}_{E} = \frac{T_{H}}{25} = 23 \)). This higher mRNA content in hepatocytes is compatible with their much larger size (Fig. 3d).

We selected genes with endothelial expression that was 23-fold higher than the hepatocyte expression, and for which the fraction of cells with nonzero expression was higher in endothelial cells than in hepatocytes. This resulted in 1,303 endothelial genes. Genes with zonation q-value lower than 0.2 were considered significantly zonated, resulting in 475 zonated endothelial genes.

Zonation reconstruction of single LECs. To spatially reconstruct the scRNAseq LEC data we established a panel of 140 LEC landmark genes, extracted from the pcRNAseq data. We chose among the 475 zonated LEC genes the ones with maximal zonation values above 2 × 10\(^{-4}\). We sorted the zonation profiles by their center of mass, and chose the portal and central landmark genes as the top and bottom 70 genes. We used equations (3) and (4) to compute η for every cell over this LEC landmark panel and partitioned the cells into four layers (0 ≤ η ≤ 0.25, 0.25 < η ≤ 0.5, 0.5 < η ≤ 0.75, 0.75 < η ≤ 1). By averaging the expression of cells in each group we established the spatial expression patterns of all genes. We removed 40 of the 1,203 sequenced LECs that had no expression of any of the 140 landmark genes. Due to the sparseness of the scRNAseq data we focused on 2,158 genes with a mean expression higher than 10\(^{-4}\). In addition, we removed 11 genes with mean expression higher than 0.01 in the single hepatocytes, assuming that expression of these genes in LEC could be contamination from hepatocyte mRNA. Significance of zonation was computed using Kruskal–Wallis test followed by Benjamini–Hochberg multiple hypothesis correction.

Statistics and reproducibility. No statistical method was used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to animal allocation during experiments and outcome assessment. All replicates were technical replicates performed in independent experiments, data is represented as s.e.m. In Figure 3e, two-sided Wilcoxon rank-sum test was used. In Figure 4c and Supplementary Figure 6d, P values were calculated using Kruskal–Wallis method. In Figure 5e and Supplementary Figure 5b Spearman correlation test was performed. In Supplementary Figure 3i two-sided Fisher test was performed. A Life Sciences Reporting Summary is available

Code availability. Supplementary Code contains ZONATION_pcRNAseq-MATLAB code that loads expression data for pairs of hepatocytes and endothelial cells, localizes each pair along the lobule axis according to hepatocyte landmark genes and outputs a table of zonated expression and a list of zonated endothelial genes.

Data availability. Data generated in this study have been deposited in Gene Expression Omnibus with the accession code: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108561. Data referenced in ref. 9 is available in GEO with the accession code GSE84498. Data referenced in Supplementary Figure 2 is available in http://dx.doi.org/10.1038/nfc755. Data referenced in Supplementary Figure 6a is available in Supplementary Table 8 of http://dx.doi.org/10.1038/nature13182.

51. Okabe, M., Ikawa, M., Kominami, K., Nakashiki, T. & Nishimune, Y. ‘Green mice’ as a source of ubiquitous green cells. FEBS Lett. 407, 313–319 (1997).
52. Vintersten, K. et al. Mouse in red: red fluorescent protein expression in mouse ES cells, embryos, and adult animals. Genesis 40, 241–246 (2004).
53. Bahar Halpern, K. et al. Bursty gene expression in the intact mammalian liver. Mol. Cell 58, 147–156 (2015).
54. Lyubimova, A. et al. Single-molecule mRNA detection and counting in mammalian tissue. Nat. Protoc. 8, 1743–1758 (2013).
55. Seglen, P.O. Preparation of rat liver cells. 3. Enzymatic requirements for tissue dispersion. Exp. Cell Res. 82, 391–398 (1973).
56. George, T.C. et al. Cardinal principles of circadian biology. J. Biol. Rhythms 31, 117–129 (2006).
57. Butler, A. & Satija, R. Integrated analysis of single cell transcriptomic data across conditions, technologies, and species. Nat. Biotechnol. 36, 411–420 (2017).
58. Shay, T. & Kang, J. Immunological Genome Project and systems immunology. Trends Immunol. 34, 602–609 (2013).
59. Hughes, M.E., Hogenesch, J.B. & Kornacker, K. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. J. Biol. Rhythms 25, 372–380 (2010).
Experimental design

1. Sample size
   Describe how sample size was determined.
   
   The sample sizes (number of mice and total number of sequenced cells) where chosen as to be sufficient to obtain sufficient power to discern gene expression differences between distinct lobule layers.

2. Data exclusions
   Describe any data exclusions.
   
   No data was excluded from the analysis.

3. Replication
   Describe the measures taken to verify the reproducibility of the experimental findings.
   
   3 mice were used for the pcRNAseq all attempts were successful, 2 mice were used for the scRNAseq all attempts were successful, 4 mice were used for spatial sorting all attempts were successful, 1 mouse was used for dual color pairs formation 4 attempts at replication failed.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.

   All mice used were wild type C57bl6. For each experiment cage mates were taken on the same date. mice allocation was random.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   No blinded groups were allocated in this study. no conditional treatment were done in this study.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   n/a
   Confirmed

   - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   - A statement indicating how many times each experiment was replicated
   - The statistical test(s) used and whether they are one- or two-sided
   - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   - Test values indicating whether an effect is present
   - Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.
   - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   - Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.
7. Software

Describe the software used to analyze the data in this study.

The softwares used to analyze the data are Matlab 2016b, R 3.4.3 and RStudio 1.1.383, ImageM 0.31, IDEAS 6.2, HISAT version 0.1.6-beta, Seurat v2.0.1 package.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

---

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

There are no restrictions on availability of unique materials.

The following antibodies were used for cell isolation: CD31 (APC 102510 clone MEC13.3, PE-Cy7 102418 clone 390, LOT B233286), CD45 (APC-Cy7 103116 clone 30-F11, LOT B242535), CD3 (PE 100205 clone 17A2, LOT B210713), CD19 (PE 152407 clone 1D3/CD19, LOT B228633), CD117 (APC 105812, LOT B249344). All antibodies binding were validated by Flow Cytometry with isotype control and by multispectral Imaging Cytometer (ImageStreamX mark II imaging flow-cytometer). All antibodies were purchased from Biolegend. All antibodies were used in dilution of 1:100.

---

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

The following antibodies were used for cell isolation: CD31 (APC 102510 clone MEC13.3, PE-Cy7 102418 clone 390, LOT B233286), CD45 (APC-Cy7 103116 clone 30-F11, LOT B242535), CD3 (PE 100205 clone 17A2, LOT B210713), CD19 (PE 152407 clone 1D3/CD19, LOT B228633), CD117 (APC 105812, LOT B249344). All antibodies binding were validated by Flow Cytometry with isotype control and by multispectral Imaging Cytometer (ImageStreamX mark II imaging flow-cytometer). All antibodies were purchased from Biolegend. All antibodies were used in dilution of 1:100.

---

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. No cell lines were used in this study.

b. Describe the method of cell line authentication used. No cell lines were used in this study.

c. Report whether the cell lines were tested for mycoplasma contamination. No cell lines were used in this study.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. No cell lines were used in this study.

---

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

C57Bl/6 male mice age 6-12 weeks were housed under reverse phase cycle, and fasted for 2 hours starting at 8AM. Liver tissues were harvested and fixed in 4% paraformaldehyde for 3 hours, incubated overnight with 30% sucrose in 4% paraformaldehyde and then embedded in OCT. Mouse liver cells for RNAseq were extracted from five mice (3 for pcRNASeq and 2 for scRNASeq) by a modification of the two-step collagenase perfusion method of Seglen. C57BL/6-Tg(CAG-EGFP) and C57BL/6-Actb-DsRed.T3 male mice age 4 months were used for the dual color FACS.

---

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No human research participants were used in this study.
Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
2. 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).
3. All plots are contour plots with outliers or pseudocolor plots.
4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.
Mouse liver cells were isolated by a modification of the two-step collagenase perfusion method of Seglen from 2 hours fasted, 6-7 weeks old male C57bl6 mice for the pcRNAseq (3 mice) or 3-4 months for the scRNAseq (2 mice). Digestion step was performed with Collagenase D for the pcRNAseq or Liberase Blendzyme 3 recombinant collagenase for the scRNAseq according to the manufacturer’s instruction. Isolated cells were stained and taken directly to sorting.

6. Identify the instrument used for data collection.
Paired Cells were sorted with SORP-FACSAriaII machine using a 130 μm nozzle, single NPCs were sorted using 100 μm nozzle.

7. Describe the software used to collect and analyze the flow cytometry data.
The software used to collect the data is FACSDiva. the flow cytometry data was analyzed by Flow_Jo_V10

8. Describe the abundance of the relevant cell populations within post-sort fractions.
The abundance of the CD31+ LEC-Hepatocyte paired cells in the population was 1.85%

9. Describe the gating strategy used.
Dead cells were excluded on the basis of 500 ng/ml Dapi incorporation. To enrich for hepatocytes a #1.5 ND filter and was used, whereas a #1 ND filter was used to enrich for NPCs. To enrich for hepatocyte-LEC pairs, cells were gated according to size that match the hepatocytes distribution by FSC-A and SSC-A. The cells were next gated by FSC-W to remove clusters of hepatocytes, and also gated for CD45 negative CD31 positive staining to enrich for pairs of hepatocyte-endothelial cells.

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