SEAM is a spatial single nuclear metabolomics method for dissecting tissue microenvironment

Zhiyuan Yuan1,10, Qiming Zhou2,10, Lesi Cai3,10, Lin Pan4, Weiliang Sun4, Shiwei Qumu5, Si Yu6, Jiaxin Feng3, Hansen Zhao3, Yongchang Zheng6, Minglei Shi7, Shao Li1, Yang Chen1,9,11✉, Xinrong Zhang3,11✉ and Michael Q. Zhang6,7,8,11✉

Spatial metabolomics can reveal intercellular heterogeneity and tissue organization. Here we report on the spatial single nuclear metabolomics (SEAM) method, a flexible platform combining high-spatial-resolution imaging mass spectrometry and a set of computational algorithms that can display multiscale and multicolor tissue tomography together with identification and clustering of single nuclei by their in situ metabolic fingerprints. We first applied SEAM to a range of wild-type mouse tissues, then delineated a consistent pattern of metabolic zonation in mouse liver. We further studied the spatial metabolic profile in the human fibrotic liver. We discovered subpopulations of hepatocytes with special metabolic features associated with their proximity to the fibrotic niche, and validated this finding by spatial transcriptomics with Geo-seq. These demonstrations highlighted SEAM’s ability to explore the spatial metabolic profile and tissue histology at the single-cell level, leading to a deeper understanding of tissue metabolic organization.

The hierarchical organization of multicellular organisms is stably maintained by homeostasis at different levels. At the tissue level, such homeostasis is often modulated by the combination of intracellular gene expression network and extracellular (microenvironmental) signals1–4. A cell and its extracellular environment interact dynamically through various signaling mediators, including metabolites, secretome and ligand–receptor interactions. Metabolites from the extracellular environment can substantially influence cell behavior or even transform cell identity. For instance, extensive alcohol intake not only activates the metabolic machinery of hepatocytes but also alters the epigenetic landscape of hepatocytes5. Conversely, the release of metabolites by cells can also have effects on the microenvironment. One classic example is basophils and mast cells releasing histamine to increase the permeability of capillaries when encountering infection6. To facilitate a deeper and more systematic understanding of the multiscale (that is, molecular level, cellular level and tissue level) nature of biological processes (for example, organ development or tumor microenvironment), various single-cell omics technologies have been rapidly developed and used7. Currently, advanced imaging mass spectrometry (IMS)-based techniques are making it possible to profile a large number of metabolites spatially and/or temporally, providing new dimensional insights into the hierarchical processes8–11.

For spatially resolved metabolomics studies, different techniques have been developed including matrix-assisted laser desorption/ionization–mass spectrometry (MALDI–MS)10, desorption electrospray ionization–mass spectrometry (DESI–MS)11, laser ablation–inductively coupled–plasma–mass spectrometry12 and secondary ion mass spectrometry (SIMS)13. Atmospheric pressure MALDI has demonstrated an application on P. caudatum and imaged endogenic biomolecules with 1.4-μm lateral resolution14. MALDI-2 was introduced by adapting a t-MALDI-2 ion source to an Orbitrap mass analyzer and a pixel size of 600 nm was achieved on brain tissue15. DESI–MS has been used to visualize tissue-level metabolic alterations in 256 patients with esophageal cancer16. Using a nanoDESI platform, a 10-μm spatial resolution was achieved on mouse pancreatic islets16. More recently, three-dimensional (3D) OrbitSIMS, a label-free IMS with a subcellular lateral resolution and high mass-resolving power, has been developed17. These techniques will be used increasingly in future spatial metabolomics applications.

Although the above techniques have achieved unprecedented subcellular resolution, several analytical complications still exist, such as single-cell segmentation and cell fingerprint extraction. Previous studies typically segmented cells using hematoxylin–eosin (H&E) staining, which suffered from either inaccurate segmentation due to imperfect registration of adjacent slides, or labeling on the same slides, which could introduce exogenous substances and

1MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic and Systems Biology, BNRist, Institute of TCM-X, Department of Automation, Tsinghua University, Beijing, China. 2MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic and Systems Biology, BNRist, School of Life Sciences, Tsinghua University, Beijing, China. 3Department of Chemistry, Tsinghua University, Beijing, China. 4Institute of Clinical Medicine, China–Japan Friendship Hospital, National Clinical Research Center for Respiratory Diseases, Institute of Respiratory Medicine, Chinese Academy of Medical Science, Beijing, China. 5Department of Pulmonary and Critical Care Medicine, China–Japan Friend Hospital, National Clinical Research Center for Respiratory Diseases, Beijing, China. 6Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. 7MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic and Systems Biology, BNRist, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing, China. 8Department of Biological Sciences, Center for Systems Biology, The University of Texas, Richardson, TX, USA. 9Present address: The State Key Laboratory of Medical Molecular Biology, Department of Molecular Biology and Biochemistry, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, China. 10These authors jointly supervised this work: Yang Chen, Xinrong Zhang, Michael Q. Zhang. ✉email: yc@ibms.pumc.edu.cn; xrzhang@mail.tsinghua.edu.cn; michael.zhang@utdallas.edu

NATURE METHODS | VOL 18 | OCTOBER 2021 | 1223-1232 | www.nature.com/naturemethods

1223
SEAM provides three additional data analysis modules (Methods): 

1. **Compositional Normalization**
   - Takes advantage of compositional characteristics and spatial continuity to build a spatial single-nucleus representation (SIMS-ID, Fig. 2b) and differential metabolite analysis (SIMS-Diff).

2. **Visualization Module**
   - SIMS-View compresses multiplexed SIMS images from three-channel images to CIELAB color spaces and maps all pixels to three. Each resulting dimension is scaled to independent pixels, each represented by a fixed-length vector, and all pixels are feature-wise normalized to avoid feature bias.

3. **Multi-Image Segmentation**
   - SEG provides segmentation masks to be smooth. Detailed performance descriptions can be found in Supplementary Notes and Supplementary Figs. 2–5.

**Algorithm design and modular data analysis.** SIMS-View is a fast visualization tool designed for SIMS data, which takes advantage of the efficiency and the local and global structure preservation of uniform manifold approximation and projection (UMAP). It takes multiplex SIMS data as input and outputs a single human-readable image using three steps. First, SIMS data are regarded as independent pixels, each represented by a fixed-length vector, and each pixel is feature-wise normalized to avoid feature bias. Next, the 65,536 pixels are fed into UMAP to reduce the dimensionality to three. Finally, each of the three resulting dimensions is scaled and color-coded by CIELAB color space, and all pixels are mapped back to their original positions (Supplementary Fig. 1). Combining the advantages of both UMAP and CIELAB color space, and unlike related works, SIMS-View provides a global view of compressed ion distribution features in a single image at the pixel level.

To solve the cell segmentation problem, previous in situ studies used a variety of approaches. Some used matched H&E stain, others took one simple measurement as input, and most of them used supervised segmentation, via either pixel-wise classification or modeling the whole image using convolution neural networks. Based on the visualization of SIMS-View results on different samples, the nuclei of cells showed similar colors for most cells, yet were different from other nonnuclear areas (Supplementary Fig. 2). Therefore, we decided to isolate the nuclei to demarcate every single cell. To avoid extra staining and heavy annotation labor, which would sabotage the original metabolic state of samples, we developed SIMS-Cut, an unsupervised label-free algorithm, to segment regions of interest (ROI) using corresponding metabolic markers, for example, adenine (m/z 134.04) as the nuclear marker. The input data format is multiplexed by selecting ion species that are highly colocalized with nuclei, which is highly consistent across different samples (Supplementary Fig. 6a). At the core of SIMS-Cut is an expectation-maximization algorithm, aiming to solve an optimization problem of a probabilistic graphical model (PGM), which combines a restricted Boltzmann machine (RBM) and a Potts model (Supplementary Fig. 6d). The RBM (Supplementary Fig. 6c) is suitable for modeling the appearance of a multi-image pixel given its label (foreground/background), and the Potts model (Supplementary Fig. 6b) encourages the resulting segmentation masks to be smooth. Detailed performance descriptions can be found in Supplementary Notes and Supplementary Figs. 7–12.

After segmentation, the metabolic fingerprint of each segmented nucleus needs to be extracted and represented. Given the fact that SIMS captures the cumulative intensities along the z axis for each pixel, the metabolic fingerprint of each cell (both nucleus and background) can be rapidly surveyed by human vision (Supplementary Fig. 1).

**Algorithm design and modular data analysis.** SIMS-View is a fast visualization tool designed for SIMS data, which takes advantage of the efficiency and the local and global structure preservation of uniform manifold approximation and projection (UMAP). It takes multiplex SIMS data as input and outputs a single human-readable image using three steps. First, SIMS data are regarded as independent pixels, each represented by a fixed-length vector, and each pixel is feature-wise normalized to avoid feature bias. Next, the 65,536 pixels are fed into UMAP to reduce the dimensionality to three. Finally, each of the three resulting dimensions is scaled and color-coded by CIELAB color space, and all pixels are mapped back to their original positions (Supplementary Fig. 1). Combining the advantages of both UMAP and CIELAB color space, and unlike related works, SIMS-View provides a global view of compressed ion distribution features in a single image at the pixel level.

To solve the cell segmentation problem, previous in situ studies used a variety of approaches. Some used matched H&E stain, others took one simple measurement as input, and most of them used supervised segmentation, via either pixel-wise classification or modeling the whole image using convolution neural networks. Based on the visualization of SIMS-View results on different samples, the nuclei of cells showed similar colors for most cells, yet were different from other nonnuclear areas (Supplementary Fig. 2). Therefore, we decided to isolate the nuclei to demarcate every single cell. To avoid extra staining and heavy annotation labor, which would sabotage the original metabolic state of samples, we developed SIMS-Cut, an unsupervised label-free algorithm, to segment regions of interest (ROI) using corresponding metabolic markers, for example, adenine (m/z 134.04) as the nuclear marker. The input data format is multiplexed by selecting ion species that are highly colocalized with nuclei, which is highly consistent across different samples (Supplementary Fig. 6a). At the core of SIMS-Cut is an expectation-maximization algorithm, aiming to solve an optimization problem of a probabilistic graphical model (PGM), which combines a restricted Boltzmann machine (RBM) and a Potts model (Supplementary Fig. 6d). The RBM (Supplementary Fig. 6c) is suitable for modeling the appearance of a multi-image pixel given its label (foreground/background), and the Potts model (Supplementary Fig. 6b) encourages the resulting segmentation masks to be smooth. Detailed performance descriptions can be found in Supplementary Notes and Supplementary Figs. 7–12.

After segmentation, the metabolic fingerprint of each segmented nucleus needs to be extracted and represented. Given the fact that SIMS captures the cumulative intensities along the z axis for each pixel, the metabolic fingerprint of each cell (both nucleus and background) can be rapidly surveyed by human vision (Supplementary Fig. 1).

**Algorithm design and modular data analysis.** SIMS-View is a fast visualization tool designed for SIMS data, which takes advantage of the efficiency and the local and global structure preservation of uniform manifold approximation and projection (UMAP). It takes multiplex SIMS data as input and outputs a single human-readable image using three steps. First, SIMS data are regarded as independent pixels, each represented by a fixed-length vector, and each pixel is feature-wise normalized to avoid feature bias. Next, the 65,536 pixels are fed into UMAP to reduce the dimensionality to three. Finally, each of the three resulting dimensions is scaled and color-coded by CIELAB color space, and all pixels are mapped back to their original positions (Supplementary Fig. 1). Combining the advantages of both UMAP and CIELAB color space, and unlike related works, SIMS-View provides a global view of compressed ion distribution features in a single image at the pixel level.

To solve the cell segmentation problem, previous in situ studies used a variety of approaches. Some used matched H&E stain, others took one simple measurement as input, and most of them used supervised segmentation, via either pixel-wise classification or modeling the whole image using convolution neural networks. Based on the visualization of SIMS-View results on different samples, the nuclei of cells showed similar colors for most cells, yet were different from other nonnuclear areas (Supplementary Fig. 2). Therefore, we decided to isolate the nuclei to demarcate every single cell. To avoid extra staining and heavy annotation labor, which would sabotage the original metabolic state of samples, we developed SIMS-Cut, an unsupervised label-free algorithm, to segment regions of interest (ROI) using corresponding metabolic markers, for example, adenine (m/z 134.04) as the nuclear marker. The input data format is multiplexed by selecting ion species that are highly colocalized with nuclei, which is highly consistent across different samples (Supplementary Fig. 6a). At the core of SIMS-Cut is an expectation-maximization algorithm, aiming to solve an optimization problem of a probabilistic graphical model (PGM), which combines a restricted Boltzmann machine (RBM) and a Potts model (Supplementary Fig. 6d). The RBM (Supplementary Fig. 6c) is suitable for modeling the appearance of a multi-image pixel given its label (foreground/background), and the Potts model (Supplementary Fig. 6b) encourages the resulting segmentation masks to be smooth. Detailed performance descriptions can be found in Supplementary Notes and Supplementary Figs. 7–12.

After segmentation, the metabolic fingerprint of each segmented nucleus needs to be extracted and represented. Given the fact that SIMS captures the cumulative intensities along the z axis for each pixel, the metabolic fingerprint of each cell (both nucleus and background) can be rapidly surveyed by human vision (Supplementary Fig. 1).

**Algorithm design and modular data analysis.** SIMS-View is a fast visualization tool designed for SIMS data, which takes advantage of the efficiency and the local and global structure preservation of uniform manifold approximation and projection (UMAP). It takes multiplex SIMS data as input and outputs a single human-readable image using three steps. First, SIMS data are regarded as independent pixels, each represented by a fixed-length vector, and each pixel is feature-wise normalized to avoid feature bias. Next, the 65,536 pixels are fed into UMAP to reduce the dimensionality to three. Finally, each of the three resulting dimensions is scaled and color-coded by CIELAB color space, and all pixels are mapped back to their original positions (Supplementary Fig. 1). Combining the advantages of both UMAP and CIELAB color space, and unlike related works, SIMS-View provides a global view of compressed ion distribution features in a single image at the pixel level.

To solve the cell segmentation problem, previous in situ studies used a variety of approaches. Some used matched H&E stain, others took one simple measurement as input, and most of them used supervised segmentation, via either pixel-wise classification or modeling the whole image using convolution neural networks. Based on the visualization of SIMS-View results on different samples, the nuclei of cells showed similar colors for most cells, yet were different from other nonnuclear areas (Supplementary Fig. 2). Therefore, we decided to isolate the nuclei to demarcate every single cell. To avoid extra staining and heavy annotation labor, which would sabotage the original metabolic state of samples, we developed SIMS-Cut, an unsupervised label-free algorithm, to segment regions of interest (ROI) using corresponding metabolic markers, for example, adenine (m/z 134.04) as the nuclear marker. The input data format is multiplexed by selecting ion species that are highly colocalized with nuclei, which is highly consistent across different samples (Supplementary Fig. 6a). At the core of SIMS-Cut is an expectation-maximization algorithm, aiming to solve an optimization problem of a probabilistic graphical model (PGM), which combines a restricted Boltzmann machine (RBM) and a Potts model (Supplementary Fig. 6d). The RBM (Supplementary Fig. 6c) is suitable for modeling the appearance of a multi-image pixel given its label (foreground/background), and the Potts model (Supplementary Fig. 6b) encourages the resulting segmentation masks to be smooth. Detailed performance descriptions can be found in Supplementary Notes and Supplementary Figs. 7–12.

After segmentation, the metabolic fingerprint of each segmented nucleus needs to be extracted and represented. Given the fact that SIMS captures the cumulative intensities along the z axis for each pixel, the metabolic fingerprint of each cell (both nucleus and background) can be rapidly surveyed by human vision (Supplementary Fig. 1).
cytoplasm) can be extracted by combining its segmentation mask and corresponding SIMS data. Existing works often represent cells by computing the average of all the pixels contained within each cell12,13, which requires strong assumptions, such as Gaussian or unimodal, and suffers from loss of pixel variation (Supplementary Fig. 15c). To obtain better results, SIMS-ID represents cells using the bipartite graph of pixels and cells constructed by a self-supervised learning algorithm33–35, which can soften the hard labeling produced by SIMS-Cut (Fig. 2b) and Supplementary Fig. 15a,b). The resulting representation showed superior discriminative power, noise robustness and pixel distribution preservation. A detailed performance description can be found in Supplementary Notes and Supplementary Figs. 16–27.

The resulting representation of SIMS-ID lies in high-dimensional feature space. SIMLR41 is a popular single-cell clustering algorithm, which automatically learns cell-to-cell affinity with multiple kernel ensemble learning, and shows satisfactory performance when combined with SIMS-ID (Supplementary Fig. 22). We simply adopted SIMLR as our clustering method.

To characterize key metabolite-differentiating clusters, and account for the variation of pixels within cells, we developed SIMS-Diff as our differential analysis algorithm. SIMS-Diff regards cells as distributions of pixels and uses the earth mover’s distance (Methods)36 for the dissimilarities among cells. Using this, the discriminative power of one feature with respect to a given cluster partition can be measured as the ratio of between-cluster variation (BCV) and within-cluster variation.

Our algorithm modules of SEAM can be freely extended to other platforms, for example multiplexed ion beam imaging by TOF (MIBI–TOF)37, single-cell metabolic regulome profiling40 and sequential fluorescence in situ hybridization (seqFISH)44 (Supplementary Notes, Supplementary Figs. 52–55 and Supplementary Table 5).

SEAM reveals cell metabolic states in mouse liver. The liver is an important metabolic organ consisting of repeating hexagonal-shaped units called lobules45. The spatial heterogeneity of its metabolic mechanisms has been thoroughly investigated using immunohistochemistry (IHC) analyses46, transcriptomics23 and epigenomics47, but, to our knowledge, direct spatial metabolic profiling at the single-cell level has not been reported. This allows us to fill the gap by a proof-of-concept demonstration of SEAM.

To this end, wild-type mice were used to obtain sequential liver sections, and CV-centered regions were selected for SEAM analysis. The SIMS data consist of approximately 200–300 ion species after spectral peak selection and filtering (Methods), and SIMS-Cut detected 724 nuclei in the field of view. To extract metabolic cell fingerprints, we used SIMS-ID to represent each cell with a fixed-length vector, which was fed into SIMLR for clustering. Then SIMLR resulted in metabolically distinct subpopulations corresponding to major liver cell types, including Kupffer cells, endothelial cells and hepatocytes (Fig. 2c).

The identified subpopulations showed specific spatial patterns consistent with the known liver organization (Fig. 2c) and Supplementary Fig. 29a) and were characterized and annotated by experienced pathologists. Kupffer cells are specialized macrophages in the liver, which typically line the walls of the sinusoids. Endothelial cells typically lie between the crevices of hepatocytes and receive blood from both the hepatic artery and the portal veins into the hepatic parenchyma48. Hepatocytes (the parenchymal cells) constitute 80% of the mass and 60% of cell composition in a healthy mammalian liver, performing various metabolic functions strongly...
associated with their positions\(^4\). The identified cell clusters are consistent between replicates (Supplementary Fig. 29b). To further confirm the major cell types identified by pathologists, we performed immunofluorescent staining assays for immunophenotypic markers of CV-enriched hepatocytes (CYP2E1\(^2\))\(^2\), sinusoidal endothelial cells (LYVE1) and Kupffer cells (CD68\(^4\)) on adjacent tissue sections of TOF–SIMS (Supplementary Fig. 28 and Methods). The spatial distribution pattern of endothelial and Kupffer cell clusters identified by SEAM showed high similarity to the spatial distribution of LYVE1 and CD68 (Supplementary Fig. 29c,d). SIMS-Diff identified differential ion species among the subpopulations (Fig. 3a,b and Supplementary Fig. 30a,b). We found that \(m/z\) 59.96, 75.96 and 76.96 colocalized with endothelial cells, while \(m/z\) 134.04, 180.89 and 90.98 were enriched in Kupffer cells (\(m/z\) 134.04 is reported to be adenine, reflecting the higher nucleus-to-cytoplasm ratio). Hepatocytes, which differ from liver nonparenchymal cells, were characterized by \(m/z\) 255.22, 279.22 and 281.23, corresponding to the fatty acid metabolism of parenchymal tissue. Hepatocytes can be subclassified into C1, C2, C3 and C4, each showing a different metabolic fingerprint (Fig. 3a,b).

SEAM reveals metabolic patterns consistent with liver zonation. Having identified the metabolic heterogeneity among hepatocytes in wild-type mouse liver lobule, we focused on hepatocyte C1, one

---

**Fig. 2 | Algorithm design and performance.**

a. Schematic diagram of SIMS-Cut, leveraging the Potts model as the prior for pixel labels and RBMs as a conditional distribution of pixel intensities. Left, top 20 nucleus-localized ions. Middle, iterative optimization between subproblems (Methods). Right, nucleus segmentation mask. Scale bar, 50 \(\mu\)m.

b. Schematic diagram of SIMS-ID, learning vector-formed representation for each segmented nucleus using self-representation learning. Left, multiplex SIMS data combined with nucleus segmentation mask. Middle, a neural network for an auxiliary classification task. Right, single nucleus representation output.

c. Demonstration of algorithms on the CV of wild-type mouse liver. Left, UMAP visualization of single nuclei represented using SIMS-ID, colored by clusters. Middle, spatial single-nucleus map. White arrow indicates CV. Scale bar, 100 \(\mu\)m. Right top, respective layout of cell populations. Scale bar, 100 \(\mu\)m. Right bottom, zoomed-in images of each population merged with gray-scale image of \(m/z\) 134.04. The red dotted area indicates liver sinusoid. Scale bar, 10 \(\mu\)m.
of the hepatocyte subpopulations whose spatial distribution was of special interest. First, we found that hepatocyte C1 (Fig. 3c left and Supplementary Fig. 30c left) as well as one of its ion species markers, m/z 87.00 (Fig. 3c middle and Supplementary Fig. 30c middle), localized around CV. The gradient orientation analysis showed that the gradient of m/z 87.00 was spreading from CV toward the outer region (Fig. 3c right, Supplementary Fig. 30c right and Methods), and the 3D surface plot also supported the enrichment of m/z 87.00.
Further quantitative analysis revealed that the cells in hepatocyte C1 showed significantly smaller distances from the CV compared with the other hepatocyte groups (P < 10^-4, one-sided Wilcoxon rank sum test) (Fig. 3e). We also found a series of ion species markers of hepatocyte C1 that displayed gradual changes along the liver lobule, and this zonation pattern showed consistency with the reported spatial transcriptome52 (Fig. 3f). Additionally, replicate experiments on different CV regions also showed consistent metabolic patterns and cluster-specific metabolic profiles but no such pattern was observed in the portal node regions (Supplementary Figs. 34a and 35). Consistent with the spatial expression of GLUL53, the spatial pattern of m/z 58.00, 59.01, 69.00, 71.02, 87.00 and 101.03 showed higher expression in the nearest 1–2 layers of hepatocytes from CV (Fig. 3f). We further conducted IHC of two liver zonation markers, glutamine synthetase (the protein encoded by GLUL) and cytochrome P450 2E1 (CYP2E1), on adjacent slides, and confirmed the liver zonation pattern (Supplementary Fig. 34b–d). To identify potential metabolites contributing to this liver zonation pattern, metabolite standards were analyzed using TOF–SIMS (Supplementary Table 9 and Supplementary Fig. 36). The series of CV-enriched ion species was consistent with the abundant ion species from the glucose mass spectrum in TOF–SIMS analysis (Fig. 3g). This suggests that the metabolite enriched at the CV region was most likely to be glucose. This finding was consistent with previous knowledge that glucose uptake takes action at the CV51. This example provided SEAM with a positive control that can accurately and comprehensively characterize the spatial heterogeneity within a well-studied tissue microenvironment.

SEAM identified hepatocyte subpopulations associated with fibrotic niche. Liver cirrhosis is a major killer, and progressive liver fibrosis often results in liver cirrhosis51. Having been proved effective in the case of wild-type mouse liver, SEAM was applied in the case of wild-type mouse liver, SEAM was applied from CV (Fig. 3f). We further conducted IHC of two liver zonation markers, glutamine synthetase (the protein encoded by GLUL) and cytochrome P450 2E1 (CYP2E1), on adjacent slides, and confirmed the liver zonation pattern (Supplementary Fig. 34b–d). To identify potential metabolites contributing to this liver zonation pattern, metabolite standards were analyzed using TOF–SIMS (Supplementary Table 9 and Supplementary Fig. 36). The series of CV-enriched ion species was consistent with the abundant ion species from the glucose mass spectrum in TOF–SIMS analysis (Fig. 3g). This suggests that the metabolite enriched at the CV region was most likely to be glucose. This finding was consistent with previous knowledge that glucose uptake takes action at the CV51. This example provided SEAM with a positive control that can accurately and comprehensively characterize the spatial heterogeneity within a well-studied tissue microenvironment.

To test this hypothesis, we collected ten nontumor tissue regions from three patients with liver cancer (Supplementary Table 2) and made sequential 10-μm slides for SIMS and other assays. We selected four regions from one sample, each containing a fibrotic niche, and conducted SIMS experiments (Fig. 4a,b). The resulting data consist of approximately 200–300 ion species after spectral peak selection and filtering (Methods). The color-coded pixel visualizations produced by SIMS-View depicted a qualitative spatial pattern within each region (Fig. 4c left column). To quantitatively characterize the cell composition and spatial organization, SIMS-Cut detected 902, 69.00 series, was defined as Hepa<sub>69-high</sub>, whereas hepatocyte C2, which was distal and not enriched with the ion species m/z 69.00 series, was defined as Hepa<sub>69-low</sub>. We also collected the fibrotic regions, as was done for the FBD samples. In total, 15 complementary DNA libraries were constructed successfully (Hepa<sub>69-high</sub> N = 6, Hepa<sub>69-low</sub> N = 5 and FBD N = 4). Principal component analysis (PCA) indicated that two different groups (Hepa<sub>69-high</sub>-proximal and Hepa<sub>69-low</sub>-distal) of hepatocytes shared higher similarity relative to FBD samples (Fig. 5c). More importantly, Hepa<sub>69-high</sub> samples were consistently closer to FBD samples than Hepa<sub>69-low</sub> to FBD samples in PCA space (Fig. 5c and Supplementary Fig. 47). To validate the expression pattern of each group, we first compared gene expression profiles between hepatocytes (that is, Hepa<sub>69-high</sub> or Hepa<sub>69-low</sub>) and FBD, then performed gene ontology (GO) enrichment for both up- and down-regulated differentially expressed genes (DEGs) (Methods and Supplementary

Transcriptome validation on hepatocyte subpopulations. To get a deeper understanding of the SEAM results, we performed Geo-seq of the transcribed RNA samples isolated from the tissues of the corresponding ROI from adjacent slides (Fig. 5a,b and Supplementary Fig. 42) with a modified protocol (Methods). To increase reproducibility, multiple adjacent slides were used (Supplementary Figs. 43–46). The Geo-seq slides showed high continuity with the corresponding SIMS slides in terms of spatial histology (Fig. 5b). Hepatocyte C1 from SEAM’s result, which was proximal to the fibrotic niche and enriched with the ion species m/z 69.00 series, was defined as Hepa<sub>69-high</sub>, whereas hepatocyte C2, which was distal and not enriched with the ion species m/z 69.00 series, was defined as Hepa<sub>69-low</sub>. We also collected the fibrotic regions, as was done for the FBD samples. In total, 15 complementary DNA libraries were constructed successfully (Hepa<sub>69-high</sub> N = 6, Hepa<sub>69-low</sub> N = 5 and FBD N = 4). Principal component analysis (PCA) indicated that two different groups (Hepa<sub>69-high</sub>-proximal and Hepa<sub>69-low</sub>-distal) of hepatocytes shared higher similarity relative to FBD samples (Fig. 5c). More importantly, Hepa<sub>69-high</sub> samples were consistently closer to FBD samples than Hepa<sub>69-low</sub> to FBD samples in PCA space (Fig. 5c and Supplementary Fig. 47). To validate the expression pattern of each group, we first compared gene expression profiles between hepatocytes (that is, Hepa<sub>69-high</sub> or Hepa<sub>69-low</sub>) and FBD, then performed gene ontology (GO) enrichment for both up- and down-regulated differentially expressed genes (DEGs) (Methods and Supplementary
Fig. 4 | SEAM identifies hepatocyte subtypes with differential metabolic states associated with spatial localization. **a**, H&E staining of human liver sample post TOF–SIMS analysis. Scale bar, 500 µm. **b**, Zoomed-in H&E images of four TOF–SIMS analyzed regions. White arrows indicate fibrotic and inflammation niches. Scale bar, 100 µm. **c**, SEAM results of four regions. Left, color-coded pixel visualizations. Scale bar, 100 µm. Middle, UMAP colored by cell clusters. Right, spatial single-nucleus map. **d**, Spatial single-nucleus maps of respective clusters merged with gray-scale ion image of m/z 134.04 in R1. Scale bar, 100 µm. **e**, Differential metabolic profile analysis of cell clusters, full m/z list in Supplementary Table 10. **f**, Hepatocyte C1 enriched ion species. Scale bar, 100 µm. **g**, Left, merged ion image of m/z 69.00 (red) and m/z 134.04 (blue). Middle and right, spatial localization of hepatocytes C1 and C2, respectively, merged with gray-scale ion image of m/z 69.00. Scale bar, 100 µm. **h**, Hepatocyte C1 is consistently closer to FBD than C2 within all five zones. Left, schematic diagram of zone definition and distance calculation. Right, paired boxplots (elements defined in Methods) of distances between C1/C2 and FBD in N=10 regions. For one-sided Wilcoxon rank sum test, P > 0.05 is not shown on the plot, *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001 and ****P ≤ 0.0001 are shown. The exact number of cells N and P values of boxplots are listed in the legend of Supplementary Fig. 41. **i**, Normalized count of hepatocyte C1 is consistently higher than C2. Left, schematic diagram of normalized count ratio calculation. Right, normalized count ratio between C1 and C2 is a function of the distance from the outer edge (indicated by the gray line in the left part of i) to the FBD in N=10 regions.
Figs. 48 and 49). Up-regulated DEGs were mainly involved in liver biosynthesis pathways for both Hepa$^{69 \text{high}}$ and Hepa$^{69 \text{low}}$ groups and down-regulated DEGs were highly enriched in lymphocyte activation and humoral immune response pathways. We further looked at the well-known marker genes specific for hepatocytes (ASL, HP and SAA1), fibrosis (TGFBI, PDGFB and COL4A1) and immune response (IGHM, IGHG3 and IGHV4-59). Both hepatocyte groups showed high levels of hepatocyte marker genes, whereas genes typically activated in fibrotic regions for fibrosis and immune response were highly expressed in FBD samples (Supplementary Fig. 50).

There were 718 DEGs fitting the criteria of adjusted $P < 0.05$ and log fold-change standard error $< 3$. The expression heatmap indicated that these genes had different expression patterns between the proximal (Hepa$^{69 \text{high}}$) and the distal (Hepa$^{69 \text{low}}$) hepatocytes (Fig. 5d). We used the DEGs for GO enrichment analysis (Fig. 5e). There were 17 genes enriched in the first GO entry, of which 16 were consistently higher in Hepa$^{69 \text{high}}$ than Hepa$^{69 \text{low}}$ (Fig. 5f). Genes for solute carrier transporter families with different functions were enriched in the fibrosis-proximal (Hepa$^{69 \text{high}}$) group, indicating that the corresponding metabolite transmembrane exchange activities were elevated.

**Discussion**

In this study, we developed SEAM, a platform combining experiments and computational algorithms to quantitatively characterize metabolic intra- or intercellular features with multiscale spatial resolution. Unlike other IMS instruments such as DESI (40–60 $\mu$m)$^{11}$, SIMS can provide high spatial resolution (HSR), allowing one to visualize detailed metabolic structures in tissue histology. With fast and minimal sample processing, SIMS maximally preserves the
native state of samples. Given the nature of SIMS, although it breaks most of the molecules into fragments, thereby making it more difficult to annotate (a common challenging issue for mass spectrometry studies), it produces a high multiplicity of metabolic features with the potential of characterizing cell and fine tissue microenvironment. Benefiting from both HSR and high multiplicity of SIMS, the algorithms of SEAM start solely from the features generated by SIMS and run a pipeline enabling metabolic fingerprint analysis from pixels to single nuclei, then to the selected metabolic features with spatial information annotated. Previously, there have been reports on spatial metabolic features at tissue level or in vitro single-cell level. This study shows the segmentation and analysis of single nuclear metabolic feature profiles directly on tissue sections. Also, this algorithmic pipeline is, in principal, scalable to other spatial-omics studies based on other IMS platforms, transcriptomics and proteomics with minimum adjustments (Supplementary Notes and Supplementary Figs. 52–55), and it easily works together with additional bioinformatics tools such as CIPHER to predict and prioritize disease-related metabolic molecules.

Apart from the scalability of SEAM’s algorithms, we have demonstrated that the range of SEAM applications could cover in vitro cell culture assays to various tissue samples. First, in the mixed cell-cultured assay, SEAM could easily deconvolute the different cell lines cocultured together. Additionally, in different wild-type mouse tissue samples, SEAM successfully segmented single nuclei without requiring extra labeling. The single-nuclear metabolic profile analysis was also consistent with conventional tissue histological characterization (Supplementary Figs. 3–5). Specifically, in the liver, a spatially well-orchestrated but complex organ, the CV–portal node axis zonation has been well-established at single-cell transcriptome level in wild-type mouse. We observed consistent zonation patterns at the single-cell level in the CV-centered region with a gradiational decrease in certain characteristic metabolites. Last, we found that hepatocyte subpopulations (including C1) differentiated by different metabolic features were also transcriptionally distinct, as shown by Geo-seq (Fig. 5c–f). The elevated expression of solute carrier genes can potentially explain the enrichment of a list of metabolite species found by SEAM (Fig. 4). These genes are involved in amino acid transport (SLC36A4, SLC3A2 and SLC38A9) and phosphate transport (SLC17A2 and SLC17A4) and -aminobutyric acid transport (SLC6A12). SLC3A2 has already been reported to play a central role in fibronectin matrix assembly, which also concurs with our result, as the proximal samples were closer to the fibrotic region. More importantly, SLC3A2 (4F2hc) and SLC7A6 (y+LAT2) have been reported to dimerize together68 and mediate arginine, leucine and glutamine uptake, and their expression was elevated in the Hepa69+ group (Fig. 5f and Supplementary Fig. 51a). These findings indicate that the transmembrane transport of amino acids was increased in these hepatocytes, and might be caused by the fibrogenesis adjacent to the Hepa69+ group. Further, we have made putative annotation of ion species relating to small biomolecules and fatty acids by searching databases and published studies17,59–64 (Supplementary Tables 7 and 8). Ion peak m/z 145.04 has been assigned to glutamine. More importantly, this peak was also abundant in our glutamine standard mass spectrum (Supplementary Fig. 36). Therefore, we were confident to assign m/z 145.04 as glutamine. We compared the relative level of glutamine (m/z 145.04) between the Hepa69+ and Hepa69− group. The Hepa69+ group has a significantly higher level of glutamine compared to the Hepa69− group (Supplementary Fig. 51b,c). Therefore, combining the evidence from SEAM and Geo-seq suggests that glutamine level was elevated in the Hepa69+ group, and that this was potentially mediated by up-regulation of SLC3A2 and SLC7A6 transmembrane amino acid transport activity. In summary, the results indicate that changes in the spatial microenvironment (for example, fibrogenesis) could influence cellular metabolic homeostasis (for example, glutamine elevation), as well as alteration of gene regulation (for example, SLC3A2 up-regulation).

Although the investigation of lipids, amino acids, carbohydrates and nucleotides can be achieved with TOF–SIMS, there might still exist challenges in metabolite annotation and coverage. However, this does not affect the usability of SEAM. On one hand, SEAM’s advantage is the identification of tissue histology patterns differentiated by high-resolution metabolic features, and such detected metabolic features provide necessary clues for further identifying possible metabolites by annotation tools60–64. On the other hand, SEAM can be combined with other IMS technologies and spatial-omics (Supplementary Notes). We believe SEAM is not restricted by certain IMS platforms and will gain more influence in various spatial-omics profiling applications.

In summary, SEAM provides a HSR single-nuclear metabolic profiling and analysis pipeline that requires minimal sample preparation and is label free. It is scalable to different biological samples ranging from cell culture assays to complex tissue samples. It can have a great impact on differentiating subtle tissue metabolic changes undetectable by or complementary to other conventional assays. With future improvements in IMS resolution and molecule annotation capability, SEAM would be able to provide more detailed spatial metabolome profiles with higher resolution and broader functionality.

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/s41592-021-01276-3.

Received: 22 August 2020; Accepted: 18 August 2021; Published online: 4 October 2021

References
1. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. Nat. Med. 19, 1423–1437 (2013).
2. Riquelme, P. A., Drapeau, E. & Doetsch, F. Brain micro-ecologies: neural stem cell niches in the adult mammalian brain. Philos. Trans. R. Soc. B 363, 123–137 (2008).
3. Swain, P. S., Elovitz, M. B. & Sigia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc. Natl Acad. Sci. USA 99, 12795–12800 (2002).
4. Zhang, J. W. & Li, L. H. Stem cell niche: microenvironment and beyond. J. Biol. Chem. 283, 9499–9503 (2008).
5. Shukla, S. D. & Lim, R. W. Epigenetic effects of ethanol on the liver and gastrointestinal system. Alcohol Res. 35, 47 (2013).
6. Beny, P. Role of histamine in acute inflammation. J. Pharm. Sci. Res. 7, 373–376 (2015).
7. Stuart, T. & Satija, R. Integrative single-cell analysis. Nat. Rev. Genet. 20, 257–272 (2019).
8. Pareek, V., Tian, H., Winogrud, N. & Benkovic, S. J. Metabolomics and mass spectrometry imaging reveal channeled de novo purine synthesis in cells. Science 368, 0036–8075 (2020).
9. Kennedy, D. E. et al. Novel specialized cell state and spatial compartments within the germinal center. Nat. Immunol. 21, 660–670 (2020).
10. Stoeckli, M., Chaurand, P., Hallahan, D. E. & Caprioli, R. M. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. Nat. Med. 7, 493–496 (2001).
11. Sun, C. et al. Spatially resolved metabolomics to discover tumor-associated metabolic alterations. Proc. Natl Acad. Sci. USA 116, 52–57 (2019).
12. Hare, D. J. et al. Three-dimensional atlas of iron, copper, and zinc in the brain tissue. Anal. Chem. 86, 4721–4728 (2014).
13. Kompaier, M., Hetles, S. & Spengler, B. Atmospheric pressure MALDI mass spectrometry imaging of tissues and cells at 1.4-μm lateral resolution. Nat. Methods 14, 90–96 (2017).
15. Niehaus, M., Soltwisch, J., Belov, M. E. & Dreisewerd, K. Transmission-mode MALDI-2 mass spectrometry imaging of cells and tissues at subcellular resolution. Nat. Methods 16, 925–931 (2019).
16. Yin, R. et al. High spatial resolution imaging of mouse pancreatic islets using nanospray desorption electrospray ionization mass spectrometry. Anal. Chem. 90, 6548–6555 (2018).
17. Passarelli, M. K. et al. The 3D OrbISIMS-label-free metabolic imaging with subcellular lateral resolution and high mass-resolving power. Nat. Methods 14, 1175 (2017).
18. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. Nat. Methods 16, 987–990 (2019).
19. Keren, L. et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. Cell 174, 1373 (2018).
20. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc. Natl Acad. Sci. USA 116, 19490–19499 (2019).
21. Keren, L. et al. MIBI–TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure. Sci. Adv. https://doi.org/10.1126/sciadv.aax5853 (2019).
22. Robertson, A. R. The CIE 1976 color-difference formulae. Color Res. Appl. 2, 0361–2317 (1977).
23. Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature 542, 352–356 (2017).
24. Park, J. et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science 360, 758–763 (2018).
25. Moor, A. E. et al. Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. Cell 175, 1156–1167 (2018).
26. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. 37, 38–44 (2019).
27. Fonvillé, J. M. et al. Hyperspectral visualization of mass spectrometry imaging data. Anal. Chem. 85, 1415–1423 (2013).
28. Abdelmoula, W. M. et al. Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of mass spectrometry imaging data. Proc. Natl Acad. Sci. USA 113, 12244–12249 (2016).
29. Smeets, T. et al. Evaluation of distance metrics and spatial autocorrelation in uniform manifold approximation and projection applied to mass spectrometry imaging data. Anal. Chem. 91, 5706–5714 (2019).
30. Wang, Y. et al. Multiplexed in situ imaging mass cytometry analysis of the human endocrine pancreas and immune system in type 1 diabetes. Cell Metab. 29, 769–783 e764 (2019).
31. Koller, D. & Friedman, N. Probabilistic Graphical Models: Principles and Techniques (MIT Press, 2009).
32. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. Science 313, 504–507 (2006).
33. Hinton, G. in Neural Networks: Tricks of the Trade (eds Montavon, G., Orr, G. B. & Müller, K. R.) (Springer, 2010).
**Methods**

IMS experiments. TOF–SIMS 5 (ION-TOF GmbH) equipped with a Bi liquid metal ion gun (LMIG) was used in this study, and we collected TOF–SIMS spectra and images of tissue samples using a 30 keV Bi+, LMIG with a HSR mode. The Bi+ current in the HSR mode was 0.1 pA (100 ns pulse width, HSR mode, unblunched beam). The total Bi+3 accumulated ion dose was about 2.0 × 10^14 ions per cm^2, and the typical probe sizes of the Bi+5 was roughly 200 nm in HSR mode. The secondary ion images were acquired using Bi+5 LMIG rastering over a 400 × 400 μm^2 area with 256 × 256 pixels. The Bi+5 LMIG was operated at a time cycle of 150 μs (mass range ∼2–2,000 u). Negative spectra were mass calibrated using CH2O, O, OH– and PO4–. A flood gun with low energy electrons was used to compensate for charge buildup on the sample surface. A 10–keV Ar+ (256 pixels) × 256 pixels. The Bi3+550 μm2 area, incident angle 45°) to carry out the depth profiling. A final 2D image was an overlay of 80–120 layers of depth profiling scan images.

In initial cell analysis, high mass resolution (HMR) mode was used with 0.8 pA (<1 ns pulse width, bunched beam) Bi+, current, and the mass resolutions (measured at C2H4+) were typically >6,000. The total Bi3+550 μm2 area, incident angle 45°) to carry out the depth profiling. A final 2D image was an overlay of 80–120 layers of depth profiling scan images.

**Peak selection.** To avoid noise interference and improve follow-up analysis efficiency and accuracy, picking out peaks from a full spectrum was necessary. A peak search process in SurfaceLab was carried out with the parameters as follows: mass range 50–500, minimum counts 10,000 and minimum signal to noise ratio 1,000. Typically, 200–500 peaks were picked out from a full spectrum.

**SIMS data preprocessing.** Each peak corresponds to a highly spatially resolved and spectrally filtered ion image: the former originated from a specific one or a class of chemical substances in the tissue sample, while the latter shows its characteristic spatial distribution features in this tissue square (Fig. 1a, top right).

For further data analysis, each ion image can be exported as an American Standard Code for Information Interchange mode data file by the SIMS built-in data processing software SurfaceLab, which contains three columns corresponding to the xy axes coordinates and signal intensity values. Overlaying ion images were processed by ImageJ FIJI (v1.53a).

**Metabolite TOF–SIMS analysis.** Most metabolites from Supplementary Table 9 were dissolved into double distilled water. Cholic acid was dissolved into methanol (high-performance liquid chromatography level, Sigma) then diluted with double distilled water. Metabolite solution was dropped onto gold plate (CITOGLAS), dried in a sterile airflow cabinet and then processed into TOF–SIMS for analysis using the same parameters as for the HMR mode. Peaks were filtered without low intensity (<mean of intensity) and background peaks from the glass slides.

**Biological experiments. Cell culture.** Human non-small-cell lung cancer cell line A549, human cervical cancer cell line HeLa, and BP cells (human epithelial cell line Hepa 1–6 and mouse liver epithelial cell line NCTC 1469) were grown on gelatin LMIG was roughly 200 nm in HSR mode. The secondary ion images were acquired using Bi+3 LMIG rastering over a 400 × 400 μm^2 area with 256 × 256 pixels. The Bi+5 LMIG was operated at a time cycle 150 μs (mass range 0–2,000 u). Negative spectra were mass calibrated using CH2O, O, OH– and PO4–. A flood gun with low energy electrons was used to compensate for charge buildup on the sample surface. A 10–keV Ar+ (256 pixels) × 256 pixels. The Bi3+550 μm2 area, incident angle 45°) to carry out the depth profiling. A final 2D image was an overlay of 80–120 layers of depth profiling scan images.

**Bromodeoxyuridine (BrdU) cell mix-culture experiment.** Following protocol from the previous study, A549 and HeLa cell lines were cultured with and without 20 μM BrdU (Sigma) for 48 h before seeding. A549 with BrdU were then replated with non-BrdU HeLa cells at a density of 2.10^5 cells per cm^2. The bromodeoxyuridine (BrdU) cell mix-culture experiment was performed using Cytofix (Leica) to obtain 3–10 μm continuously adjacent sections.

**Tissue section preparation.** Mouse and human tissues were isolated individually and embedded in Optimum Cutting Temperature compound (SAKURA), then snap-frozen in liquid nitrogen. Cryosections were then washed with CM1900 Cryostat (Leica) for 3–5 min then washed in PBS twice for 5 min each time. Slides were fixed in 4% PFA for 20 min at room temperature then washed in PBS once. Samples were permeabilized and blocked in 5% BSA solution (Sigma) with 0.4% Triton-X-100 (AMRESCO) for 2 h at room temperature. We carried out primary antibody incubation using glutamine synthetase antibody (ab176562, Abcam, 1:200 dilution) and cyclochrome P450 2E1 antibody (ab28146, Abcam, 1:300 dilution) in PBS with 0.1% Triton-X-100, and incubated this in a humid dark chamber at 4°C overnight. Next, it was washed three times in PBS with 0.1% Triton-X100 for 10 min each time. We carried out secondary antibody incubation using HRP conjugated goat-anti-rabbit antibody (7074, Cell Signaling, 1:500 dilution) in PBS with 0.1% Triton-X100 and incubated this in a humid dark chamber at room temperature for 2 h. We washed this three times in PBS with 0.1% Triton-X100 for 10 min each time. We applied 1x DAB (3,3’-diaminobenzidine) solution using a DAB staining kit (PA110, Tiangen) for 3–5 min. Slides were mounted using ProLong Gold Antifade Mountant (ThermoFisher). Images were obtained from Axio Scan. Z1 (Zeiss) and processed via ZEN (Blue edition) v.3.1 or obtained from Cytation5 (Biotek) and processed via its software Gen 5.

**Immunofluorescence.** Tissue cryosections were thawed at room temperature for 5 min then washed in PBS twice for 5 min each time. Slides were fixed in 4% PFA for 20 min at room temperature then washed in PBS once. Samples were permeabilized and blocked in 5% BSA solution (Sigma) with 0.4% Triton-X100 (AMRESCO) for 2 h at room temperature. We carried out primary antibody incubation using cytochrome P450 2E1 antibody (ab28146, Abcam, 1:300 dilution), LVVE1 antibody (ab218535, Abcam, 1:800 dilution) and CD68 antibody (ab53444, Abcam, 1:300 dilution) in PBS with 0.1% Triton-X100 and incubated this in a humid dark chamber at 4°C overnight, then washed it three times in PBS with 0.1% Triton-X100 for 10 min each time. We carried out secondary antibody incubation using Alexa Fluor 488 conjugated goat-anti-rabbit antibody (A11007, Thermofisher Scientific, 1:500 dilution) and Alexa Fluor 594 conjugated goat-anti-rat antibody (A11007, Thermofisher Scientific, 1:500 dilution) in PBS with 0.1% Triton-X100 and incubated this in a humid dark chamber at 4°C overnight, then washed it three times in PBS with 0.1% Triton-X100 for 10 min each time. Slides were mounted using ProLong Gold Antifade Mountant (ThermoFisher). Images were captured either by LSM780 confocal microscope (Zeiss) or Axio Scan. Z1 (Zeiss) and processed via ZEN (Blue edition) v.3.1.

**Modified Geo-seq.** We modified a spatial transcriptome analysis method, Geo-seq, previously described by Chen et al. To perform this, tissue cryosections were mounted on the polyethylene naphthalate membrane slide and stored at −80°C for short-term storage. Slides were stained in 0.5% cresyl violet and dehydrated in serial ethanol. Tissue samples were then obtained in a 0.2 ml PCR tube by LMD7000 (Leica). Buffer for RLT (Qiagen) with dithiothreitol (DTT) (Sigma) was added and shaken vigorously for tissue lysis and RNA release. RNA Clean beads (Vazyme) were added to isolate total RNA. We prepared the annealing procedure in the same tube with 3 μl of H2O, 1 μl of dNTP, 1 μl of Oligo(dT) and 0.5 μl of RNase Inhibitor (Life Technologies). This was incubated at 72°C for 5 min and immediately transferred to −80°C. We prepared reverse transcription at this temperature with 2 μl of 5X RT buffer, 0.5 μl of DTT, 0.5 μl of RI, 0.5 μl of Template Switch Oligo (Sangon Biotech) and 1 μl of maxima reverse transcriptase (Life Technologies).
This was incubated at 50 °C for 1 h and reverse transcriptase was deactivated at 85 °C for 5 min. We amplified the first strand product with 12.5 µl of 2x KAPA HiFi HotStart ReadyMix (Sigma), 0.5 µl of TS-PCR primer (Sanogen Biotech) and 2 µl of H2O. The reaction conditions were 95°C for 3 min, 98 °C for 20 s, 67 °C for 15 s, 72 °C for 6 min for 21 cycles and 72°C for 5 min. The PCR product was purified with 0.8X DNA Clean beads (Vazyme). The next-generation sequencing library was then constructed by TruePrep DNA Library Prep Kit v2 for Illumina (Vazyme). Libraries were sequenced by Illumina Xten Pair-end 150 bp by Annoroad.

**Cell-type annotation**. The cell-type annotation was performed by experienced pathologists with reference to the H&E staining. To double confirm the cell-type annotations with morphology and protein markers, we performed SIMS and H&E staining on the same tissue sections, and further conducted immunofluorescence staining of CYP2E1, CD68 and LYVE1 on the adjacent section.

**RNA-sequencing (RAN-seq) data processing and analysis**. RNA-seq data were first obtained through adapter removal and quality filtering by ‘Trim Galore’. The qualified reads were then mapped to the human genome reference genome using STAR and used to generate BAM files. Samtools was used to index the BAM files. Duplication was removed by PICARD (http://broadinstitute.github.io/picard/) for all the BAM files. The read count for each gene was performed by HTSeq-count with reference to gencode human gene annotation, release 32 (GRCh38.p12). Different gene expression analyses were analyzed using DESeq2 in R. We carried out GO analysis using clusterProfiler and annotated using org.Hs.eg.db (Carlson v.2019, org.Hs.eg.db: Genome wide annotation for human.

**Statistics and reproducibility**. For the HMR negative ion mode in TOF–SIMS analysis, the replication numbers, N, of independent cell line experiments (Supplementary Figs. 7, 17b, 25b and 26a,b,h) were as follows: A549 (N = 3), HeLa (N = 3), SK-BR-3 (N = 3), MCF-10A (N = 3), A549/HeLa mixture (N = 3) and MDA-MB-468 mixture (N = 1). The replication numbers, N, of mouse tissue experiments (Supplementary Fig. 35a–b) were as follows: liver (N = 11 on two independent mice). For HSR negative ion mode in TOF–SIMS analysis, the N of independent cell line experiments (Supplementary Figs. 7, 17d and 27a,b,a) as follows were: MCF-7 (N = 11) and NCTC 1469/Hepa1-6 mixture (N = 1). The replication numbers, N, of mouse tissue experiments (Figs. 1b, 2c and 3c and Supplementary Figs. 3a,c,e, 4a,c,e, 5a,c,e, 6, 8, 9, 28, 29c and 30c) were as follows: liver (N = 7 on three independent mice), lung (N = 5 on three independent mice), small intestine (N = 4 on two independent mice), kidney (N = 4 on two independent mice), stomach (N = 2 on two independent mice) and pancreas (N = 4 on two independent mice). The N of human tissue experiments (Figs. 4c,d,f and 5b and Supplementary Figs. 38, 39, 40b, 41 and 51b) were as follows: ICC nonumor liver tissues (N = 10 on three independent patient samples). For Geo-seq assays, the N of human tissue experiments (Supplementary Figs. 42, 43, 44, 45 and 46a) as follows were: ICC nonumor liver tissues (N = 5 (Hepa1-6), 6 (Hepa69-low), 8 (Hepa69-high) and 4 (FBD) from two slides on one patient sample). For H&E staining assays, the N of mouse tissue experiments (Fig. 1 and Supplementary Figs. 28, 34a and 35a–b) were as follows: liver (N = 3 independent mice), kidney (N = 2 independent mice), stomach (N = 2 independent mice) and pancreas (N = 2 independent mice). The N of human tissue experiments (Figs. 4a,b and 5b) were as follows: ICC nonumor liver tissues (N = 3 independent patient samples). For IHC and immunofluorescence staining assays, the N of mouse tissue experiments (Supplementary Figs. 28a,b and 29c and 34b–d) were as follows: glutamine synthetase on liver (N = 2 independent mice), Cyp2e1 on liver (N = 4 independent mice), CD68 on liver (N = 2 independent mice) and LYVE1 on liver (N = 2 independent mice). For F-actin/neural lipid staining assays, the N of human tissue experiments (Supplementary Fig. 40a) were as follows: ICC nonumor liver tissues (N = 2 on one patient sample). For SIMS-Cut comparison on public MIBI data, the N of MIBI data (Supplementary Figs. 52 and 53) were as follows: triple-negative samples from patients with breast cancer (N = 36 fields of view on 36 patients).

**SIMS-Cut framework**. Given an M × N × N SIMS data, with M filtered metabolic peaks and N × N images as input, SIMS-Cut first select M metabolites colocalizing with nucleus (Supplementary Fig. 6a), and then iteratively solves a maximum a posteriori problem (Supplementary Fig. 6d) to get an N × N binary matrix Y that indicates a nucleus.

\[
Y_j^\text{Nuc} = \begin{cases} 
1 & i,j \in [1, N] \\
0 & \text{otherwise}
\end{cases}
\]  

Since the SIMS data is superimposed on a certain thickness of biological slice in its nature, we regard the segmented nucleus region as a cell containing molecular fragments in both cytoplasm and nucleus. The main part of SIMS-Cut can be formulated by finding an optimal Y:

\[
Y^* = \arg \max_Y \frac{p(Y|X)}{p(Y)} \propto p(Y|X)
\]

where

\[
p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}
\]

\[X = [x_i], i,j \in [1, N], \text{ and } x_N \in \mathbb{R}^n, \text{ which is the M-dimensional metabolic density at the coordinate of } (i,j). \text{ This Bayesian formulation aims to find the optimal label assignment } Y^* \text{ that produces the maximum posterior probability given } X.\]

As with traditional hidden Markov random field-based image segmentation, SIMS-Cut uses a similar graphical model, consisting of p(Y), the smoothing model for unknown label field Y before guaranteed spatial homogeneity and p(Y|X), the data model for the conditional distribution of pixel metabolic profiles X given corresponding pixel labels.

**Smoothing model**. The label prior, p(Y) is modeled as a special Markov random field called the Potts model. For HSR negative ion mode in TOF–SIMS analysis, the label prior, p(Y) is modeled as a special Markov random field called the Potts model.

\[
p(Y) = \frac{1}{Z} \exp(-U(Y))
\]

where U is the energy function, which is calculated by summing over the potential of all second-order cliques V, and each clique corresponds to a pair of neighboring pixels (for example, the four-neighborhood system). Z is a partition function, making p(Y) a valid probability density function:

\[
U(Y) = \sum_{(i,j), (j,k) \in \text{doubletons}} V(y_{i,j}, y_{j,k})
\]

V is defined on doubletons, penalizing the heterogeneity of labels:

\[
V(y_{i,j}, y_{j,k}) = \left\{ \begin{array}{ll} 
-1, & \text{if } y_{i,j} \neq y_{j,k} \\
+1, & \text{if } y_{i,j} = y_{j,k}
\end{array} \right.
\]

**Data model**. According to the graphical model (Supplementary Fig. 6b) and d separate,

\[
p(X|Y) = \prod_{i,j \in [1, N]} p(x_i|y_j)
\]

While the multivariate Gaussian distribution is typically suited to the data model of color image segmentation, its model capacity is limited and its assumptions are too strong for SIMS data. Instead, we use RBMs to model the conditional distribution of data intensities given label assignment.

**RBMs as a generative model**. RBMs are typically a two-layer bipartite undirected graph. It is composed of a visible layer that is an M-dimensional memory providing model capacity. In theory, RBM is a universal approximation for any probability density function with a large enough number of hidden layers. Here we use two separate RBMs to model p(x_i|y_j = 0) and p(x_i|y_j = 1), respectively, and we describe each RBM in the following. For the sake of notational simplicity, in the following, we use V = [v_k], p \in [1, m] to denote x_k (the subscript is removable thanks to the conditional independence given by equation (7)).

The graphical model of RBM is shown in Supplementary Fig. 6c.

\[
H = [h], \quad q \in [1, d] \text{ is the hidden layer variable, and } V \text{ is the visible layer variable.}
\]

\[
C = [c], q \in [1, d], \quad B = [b], p \in [1, m] \text{ and } W = [w], p \in [1, m], q \in [1, d] \text{ are parameters.}
\]

The joint probability density function is

\[
p(V, H, F) = \frac{1}{Z} e^{−E(V,H)}
\]

where E is the energy function

\[
E(V, H, F) = −m \sum_{p=1}^{m} \sum_{q=1}^{d} w_{pq} h_q y_p − \sum_{p=1}^{m} b_p y_p − \sum_{q=1}^{d} c_q h_q
\]

and Z is the partition function

\[
Z = \sum_{V,H,F} e^{−E(V,H)}
\]
The probability that an RBM model assigns a vector \( V \), for example \( x_0 \), is given by

\[
p(x_0|y=a) = \text{RBM}(V; W^a, C^a, B^a)
\]

\[
= \frac{1}{Z} \sum_{a'} e^{-(V;W^a)} = \frac{1}{Z} \sum_{a'} \left( \prod_i e^{w_{ia} x_i} \prod_j e^{c_{ja} y_j} \right)
\]

Note that the superscripts indicate the parameters of a specific RBM.

**Partition function of RBMs estimation.** For a specific pixel given its segmentation label \( a \), the log probability that the RBM assigns metabolic profiling \( x_i \) is computed as

\[
\log p(x_i|y=a) = -E(x_i) - \log Z
\]

Here \( E(x_i) \) is the free energy of the RBM corresponding to class \( a \), which can be rapidly calculated. To estimate the partition function \( Z \), we build a softmax model to classify \( x_i \) at every pixel to its label \( y \):}

\[
\log p(y = a|x_i) = \frac{e^{-E(x_i)} - \log Z}{\sum_j e^{-E(x_i)} - \log Z}
\]

**Maximum a posteriori.** Our objective can be expressed as

\[
\text{argmax}_x \log p(X|Y) + \log p(Y) = \text{argmax}_x \sum_i \log p(x_i|y_j) + \log p(Y) = \text{argmax}_x \sum_i \log \text{RBM}(x_i; W, C, B) + \log p(Y)
\]

As it is a nonconvex problem, we develop an EM-style algorithm to alternate between subproblems to reach a locally optimal point iteratively (Supplementary Notes).

**SIMS-ID framework.** After SIMS-Cut, hundreds of separated nuclei have been detected from an \( N \times N \) image, each pixel containing \( M \)-dimensional metabolic profiles. Thus, each nucleus contains a diverse number of connecting pixels, represented by fixed dimensional vectors. SIMS-ID conducted an auxiliary classification task to assign a single fixed dimensional vector to each nucleus, which is robust to over/under segmentation in SIMS-Cut. The representation learned by SIMS-ID compresses all the pixel metabolic information using a distilled softmax space, regarding a nucleus as a whole while including distribution information of pixels. A fixed dimensional representation of the nucleus helps further analysis of single-nuclei data, such as clustering, visualization and so on. More detailed information can be found in the Supplementary Notes.

**Clustering.** Represented by fixed-length vectors, the nuclei can be straightforwardly clustered and visualized in low-dimensional space. The number of cells that one SIMS experiment captures typically ranges from 400 to 1,000, and the length of the representation vector for each cell is equal to the number of pixels within segmented cells, typically ranging from 5,000 to 15,000. With the consideration of both data characteristics and experimental performance (Supplementary Fig. 32), we apply SIMLR, a single-cell clustering algorithm, which automatically learns the low-rank similarity matrix by means of multiple kernel ensemble. SIMLR also provides a means of estimating the number of clusters, which we can take as a guideline to explore populations of metabolic cell states in different scales.

**SIMS-Diff framework.** The goal of this algorithm is quantification of the feature’s discriminative power to tell clusters apart. Owing to the nature of our data, the traditional two-sample test cannot be directly applied. We assume that discriminative features can produce a similarity matrix with a block diagonal structure. Therefore, we use the ratio between BCV and within-cluster variation to evaluate the compactness of the similarity matrix. For each feature, we use the earth mover’s distance as a metric for two nuclei represented by histograms, and the variation can be simply evaluated by summing all pairwise distances. More detailed information can be found in the Supplementary Notes.

**Pixel-neighborhood analysis.** Pixel-neighborhood analysis is a tool for computationally identifying spatial niches or microenvironments within a SIMS field of view. The main idea came from the cellular neighborhoods analysis. We modified some portions and added some visualizations to fit our case. Pixel-neighborhood analysis contains four steps, and more detailed information can be found in the Supplementary Notes.

**Gradient orientation analysis.** Gradient orientation analysis is a tool to visualize the gradient quantification of metabolic features within a SIMS field of view. More detailed information can be found in the Supplementary Notes.

**Multimodal intersection analysis between mouse and human liver samples.** To access the correspondence between clusters identified in mouse and human samples, we adopted modified multimodal intersection analysis\(^5\). Specifically, we ranked metabolites by the score computed using SCANPP\(^6\), which is a score underlying the computation of a \( P \) value (Student’s \( t \)-test) for each gene for each cluster. Next, gene sets of each cluster were defined as genes with the top 20 associated scores. The significance of the intersection of gene sets between any pair of clusters was inferred using the hypergeometric distribution. The multimodal intersection analysis map was finally displayed as a heatmap, with each element defined as the negative logarithm \( P \) value (hypergeometric test) of the corresponding cluster pair.

**Statistical analysis of human samples.** To exactly describe the statistical analysis in Fig. 4, we defined the following terms: \( FBD_i \) is the FBD of region \( R_i \); \( \text{PSP}(R_i) \) is a parallel strip whose distance to \( FBD_i \) is equal to \( j \mu \); \( \text{AREA}(j) \) is the territory between \( FBD_i \) and \( \text{PSP}(j) \); \( \text{CFBD}(cell\text{-}j) \) is the distance (\( \mu \)) between cell and \( FBD_i \) within \( zone(k) \); \( \text{NCCI}(population, area, area) \) is the ratio between the number of cells in population, within area, and the number of cells in population, within area.

The FBD is approximated according to SIMS-View and the spatial single-nucleus map (Supplementary Fig. 41). Coming to cases where FBD could not be well fitted by a single line segment, polylines are used and the distance to FBD is simply adjusted to be the smallest among all line segments.

The statistical analysis in Fig. 4b was conducted as follows: in zone \( j \), \( i \in \{0, 1, 2, 3, 4\} \), the red boxplot is the summary of \( \{\text{CFBD}(cell\text{-}j)\} \) \( i \in \{0, 2, 4, 5, 6, 7, 8, 9, 10\} \), cell, \( \in \) hepatocyte \( C_1 \), and the green boxplot is the summary of \( \{\text{CFBD}(cell\text{-}j)\} \) \( i \in \{0, 2, 4, 5, 6, 7, 8, 9, 10\} \), cell, \( \in \) hepatocyte \( C_2 \). The \( P \) value is based on a Wilcoxon rank sum test.

**Boxplots.** All boxplots in the main and supplementary figures share the same axis, the lower and upper hinges show the first and third quartiles (the 25th and 75th percentiles); the center lines correspond to the median; the upper whisker extends from the upper hinge to the largest value, which should be less than 1.5 \( \times \) the interquartile range (or distance between the first and third quartiles) and the lower whisker extends from the lower hinge to the smallest value, which is at most 1.5 \( \times \) the interquartile range. Data beyond the end of the whiskers are ‘outlying’ points and are plotted individually.

**Datasets.** Detailed information of simulated datasets (Datasets 1 to 9) and mixture cell datasets (Datasets 10 and 11) can be found in Supplementary Notes.

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

**Data availability**

Raw SIMS data for mouse liver, stomach, pancreas, kidney (Figs. 1–3), and human liver R1 (Fig. 4) are available at GitHub (https://github.com/yuanzhiyuan/SEAM/tree/master/SEAM/data/raw_tar). The rest of the raw SIMS data and processed SIMS data are available at figshare (https://doi.org/10.6084/m9.figshare.12622885.v1, https://doi.org/10.6084/m9.figshare.12622841.v1, https://doi.org/10.6084/m9.figshare.12622383.v1 and https://doi.org/10.6084/m9.figshare.12622922.v1). Geo-seq (Fig. 5) raw sequencing data and processed data have been deposited to NCBI GEO with accession number GSE153463. The MIBI–TOF data can be downloaded from https://mibi-share.ionpath.com. Single-cell metabolic regulome profiling data can be downloaded from https://doi.org/10.5281/zenodo.3951613. The seqFISH data can be downloaded from https://doi.org/10.5281/zenodo.5025068.

**Code availability**

An open-source Python and MATLAB implementation of SEAM is available at Zenodo (https://doi.org/10.5281/zenodo.5025068).
References
69. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).
70. Chen, J. et al. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. Nat. Protoc. 12, 566–580 (2017).
71. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17, 10–12 (2011).
72. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 47, D766–D773 (2019).
73. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
74. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
75. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
76. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, https://doi.org/10.1186/s13059-014-0550-8 (2014).
77. Yu, G. C., Wang, L. G., Han, Y. Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16, 284–287 (2012).
78. Zhang, Y. Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57 (2001).
79. Panjwani, D. K. & Healey, G. Markov random-field models for unsupervised segmentation of textured color images. IEEE Trans. Pattern Anal. 17, 939–954 (1995).
80. Hammersley, J. M. & Clifford, P. Markov Fields on Finite Graphs and Lattices (Unpublished manuscript 46, 1971).
81. Clifford, P. in Disorder in Physical Systems: A Volume in Honour of John M. Hammersley (eds Grimmett, G. R. & Welsh, D. J. A.) (Oxford Univ. Press, 1990).
82. Besag, J. Spatial interaction and statistical-analysis of lattice systems. J. R. Stat. Soc. B. Met. 36, 192–236 (1974).
83. Panjwani, D. K. & Healey, G. Markov random-field model for unsupervised segmentation of textured color images. IEEE T. Pattern Anal. 17, 1128–1128 (1995).
84. Kato, Z. & Pong, T. C. A Markov random field image segmentation model for color textured images. Image Vision Comput. 24, 1103–1114 (2006).
85. Hinton, G., Vinyals, O. & Dean, J. Distilling the knowledge in a neural network. Preprint at arXiv https://arxiv.org/abs/1503.02531 (2015).
86. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. Cell https://doi.org/10.1016/j.cell.2020.07.005 (2020).
87. Moncada, R. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat. Biotechnol. https://doi.org/10.1038/s41587-019-0392-8 (2020).
88. Wolf, F. A., Angerer, P. & Theis, F. J. SCANGPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018).
89. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272 (2020).

Acknowledgements
We acknowledge the Imaging Core Facility, Technology Center for Protein Sciences, Tsinghua University for assistance with using the LMD7000. We also thank Y. Chen from the Imaging Core Facility for her detailed instructions on the LMD7000. We thank the Center of Laboratory Animal Resources, Tsinghua University for maintenance of mice and providing the CM1900 Cryostat. We thank H. Li for computing resource support. We thank H. Zhang for help with ethics material preparation. We thank Y. Li, Z. Ye, R. Qi and all other members of our laboratory for valuable comments and discussions. We thank M. Qian for helpful advice on algorithm development. This work was supported by the National Basic Research Program of China (2018YFA0801402, 2017YFA0505030 (Y.C.)); National Nature Science Foundation of China (81888994, 31871343 (Y.C.), 21974078, 21727813 (X.Z.), 62050152 (M.S.), 81630103, 6206116069 (S.L.), CAMS Innovation Fund for Medical Sciences (2020-RC318-009 (Y.C.)) and foundation of BNRst (BNR2019TD01020 (S.L.)). M.Q.Z. is supported by the Cecil H. and Ida Green Distinguished Chair.

Author contributions
Y.C., M.Q.Z. and X.Z. conceived and designed the project. L.C. designed and implemented the algorithms under the guidance of M.Q.Z. and Y.C., and was assisted by Q.Z. Z.Y. analyzed the SIMS data and Q.Z. analyzed the spatial transcriptome data. Y.Z. and S.L. provided the clinical samples. L.F and S.Q. guided the histological annotation. S.L. and M.S. gave suggestions on the application of the method. J.F. and H.Z. helped with the metabolite annotation. Z.Y., Q.Z. and L.C. completed the figures and writing of the paper with the guidance of Y.C., X.Z. and M.Q.Z.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-021-01276-3.
Correspondence and requests for materials should be addressed to Yang Chen, Xinxong Zhang or Michael Q. Zhang.
Peer review information Nature Methods thanks Benjamin Balluff and the other anonymous, reviewer(s) for their contribution to the peer review of this work. Arunima Singh was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.
Reprints and permissions information is available at www.nature.com/reprints.
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 our Editorial Policies 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

TOF-SIMS 5 (ION-TOF GmbH, Münster, Germany) equipped with a Bi liquid metal ion gun (LMIG) is used for SIMS experiments, and the built-in software SurfaceLab 6.7 is used for SIMS data collection. Cytation 5 built-in software Gen 5. ZEN [Blue edition] V3.1.

Data analysis

Custom software and associated dependencies are provided on Github: https://github.com/yuanzhixuan/SEAM, and https://github.com/yuanzhixuan/SIMS-Cut. We have also uploaded all the code in Zenodo [10.5281/zenodo.5025068]. Python v3.7.5, Matlab R2015b, and R v3.5.3 were used for data analysis. Spatial single cell data structure and figure generation: SCANPY v1.4.6 and Seaborn v0.10.1. SIMLR used for clustering analysis is publicly available on https://github.com/bowang87/SIMLR_PY. Statistical hypothesis testing: Scipy v1.4.1. Neural network design: Keras v2.3.1. RNAseq data analysis: Trim galore v2.6, STAR v2.7.2b, samtools v1.9, picard v2.18.7, htseq v0.11.2. R v3.5.1. DESeq2_1.22.2, clusterProfiler_3.10.1, org.Hs.eg.db_3.11.4. ImageJ 1.53a.

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 and 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

Raw SIMS data for mouse liver and lung (Fig. 1, 2,3), and human liver R1 (Fig. 4) are available at Github [https://github.com/yuanzhixuan/SEAM/tree/master/SEAM/data/raw_tar]. The rest of raw SIMS data and processed SIMS data are available at figshare [10.6084/m9.figshare.12622883.v1, 10.6084/m9.figshare.12622841.v1,
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-list.pdf

Life sciences study design

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

Sample size
No statistical methods were used to predetermine sample size. One tissue sample can be used for multiple regions of SEAM analysis. Therefore, at a single cell level analysis, independent biological samples can be reduced to minimal to 3 to reduce the amount of patient samples required.

Data exclusions
No data were excluded

Replication
Biological independent replications and technical replications supported the reproducibility and performance of SEAM. Detailed measurement of replication for each experiment was described in the manuscript. "Statistics and Reproducibility" section in Methods section. In Supplementary Figure 41, for R3, the distance-based analysis, the pattern was partially consistent with other 9 replicates. The reason could be the fibrotic area was not large enough.
For cell culture experiment in Supplementary Figure 7(MCF 7 only), 25, 27, no replication was made due to limited TOF-SIMS analysis availability caused by instrument malfunction. However, experiments from similar setup using different cell lines with replication have shown reproducible results.

Randomization
For mature cell culture experiments, two cell lines were well-mixed and co-culture at 50-50 cell number ratio with cell line labeled. TOF-SIMS scan was on randomly chosen regions.
The mice were purchased randomly from different time points but at same age. For mouse liver TOF-SIMS analysis, central vein and portal node regions were allocated into experimental groups based on Glutamine synthetase and Cyp2e1 staining.
The human biopsy were randomly collected but only if ICC was confirmed by histological examination.
For Geo-seq, samples were allocated into experimental groups based on their spatial positions corresponding to SEAM results.

Blinding
For mature cell culture experiments, investigators were blinded during data collection and analysis, followed by validation of positive control labeling for cell line type.
For mouse liver central vein (CV) and portal node PN regions TOF-SIMS analysis, investigators were not blinded to CV and PN regions allocation, because difference was to be assessed between two regions. TOF-SIMS results supported the existence of reproducible difference.
For mouse stomach, kidney, lung, pancreas and small intestine TOF-SIMS cell type clustering analysis, investigators were blinded during data collection and analysis. Results were only further validated by H&E staining afterwards.
For Geo-seq, investigators were not blind to group allocation during data collection and analysis, because transcriptomic difference was to be assessed between subgroups used for SEAM validation. Sample group assignments were further validated using Principal Component Analysis on all sequencing data.

Behavioural & social sciences study design

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

Study description
Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample
State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.

Sampling strategy
Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.

Data collection
Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.

Timing
Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions
If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Non-participation
State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.

Randomization
If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design
All studies must disclose on these points even when the disclosure is negative.

Study description
Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample
Describe the research sample (e.g. a group of tagged Passer domesticus, all Serocerus Thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy
Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection
Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale
Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken.

Data exclusions
If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility
Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization
Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding
Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? Yes No

Field work, collection and transport

Field conditions
Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location
State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access & export
Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance
Describe any disturbance caused by the study and how it was minimized.

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

- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Human research participants
- Clinical data
- Dual use research of concern

Methods

- Involved in the study
- ChiP-seq
- Flow cytometry
- MRI-based neuroimaging

Antibodies

- **Antibodies used**
  - Recombinant Anti-Glutamine Synthetase antibody [EPR13022(B)] (Abcam, ab176562)
  - Anti-Cytochrome P450 2E1 antibody (Abcam, ab28146)
  - Anti-rabbit IgG, HRP-linked Antibody (Cell Signaling Technology, #7074)
  - Anti-CD68 antibody [FA 11] (Abcam, ab53444)
  - Recombinant Anti-LYVE1 antibody [EPR21771] (Abcam, ab218535)

- **Validation**
  - All primary antibodies are commercially available and validated by the manufacturer. All primary antibodies were validated by the manufacturer for IHC or IF staining assay, especially Glutamine Synthetase and LYVE1 antibodies were stained on mouse liver tissue sections. Knock-out validation was also conducted by manufacturer for Glutamine Synthetase. Our staining staining images (Supplementary Figure 28, 29, 34) of these antibodies show similar patterns in mouse liver sections, comparing to previous reports (references below) of their expression pattern.

  Preciosi, M., Okabe, H., Poddar, M., Singh, S. & Monga, S. P. Endothelial Wnt5a regulates β-catenin signaling in murine liver zonation and regeneration: A sequel to the Wnt–Wnt situation. Hepatology communications 2, 845-860 (2018).
  Sekine, S., Ogawa, R., Mcmanus, M. T., Kanai, Y. & Hebbok, M. Dicer is required for proper liver zonation. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland 219, 365-372 (2009).

Eukaryotic cell lines

- **Policy information about cell lines**
  - Cell line source(s): A549; MCF 7; MCF 10A; Hela; SK-BR-3; Hepa 1-6; NCTC 1469; MDA-MB-468
  - Authentication: None of the cell lines used were authenticated.
  - Mycoplasma contamination: Cell lines were not tested for mycoplasma contamination.
  - Commonly misidentified lines (See ICLAC register): No commonly misidentified cell lines were used.

Palaeontology and Archaeology

- **Specimen provenance**
  - Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

- **Specimen deposition**
  - Indicate where the specimens have been deposited to permit free access by other researchers.

- **Dating methods**
  - If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

- Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

- **Ethics oversight**
  - Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

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

Animals and other organisms

- **Policy information about studies involving animals, ARRIVE guidelines** recommended for reporting animal research

- **Laboratory animals**
  - Both 8-week old female and male C57BL/6N mice were purchased from Charles River. All mice were housed in isolated ventilated cages (maxima six mice per cage) barrier facility at Tsinghua University. The mice were maintained on a 12/12-hour light/dark cycle, 42% humidity, 22-25°C with sterile pellet food and water ad libitum.

- **Wild animals**
  - The study did not involve wild animals.
Field-collected samples
The study did not involve samples collected from the field.

Ethics oversight
The laboratory animal facility has been accredited by AAALAC (Association for Assessment and Accreditation of Laboratory Animal Care International) and the IACUC (Institutional Animal Care and Use Committee) of Tsinghua University approved all animal protocols used in this study (Animal Welfare Assurance Number F16-00228 [A5061-01]).

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
This study involved 3 participants, who were diagnosed with intrahepatic cholangiocarcinoma and received no treatment before surgery. Patient characteristics in this study are provided in detail in Supplementary table 2.

Recruitment
Participants were included if they had histologically confirmed intrahepatic cholangiocarcinoma and received no treatment before surgery. Blind and random selection were made based on previous criteria, therefore there is no potential self-selection bias.

Ethics oversight
The protocol of this study was compliant with the principles of the Declaration of Helsinki and was also approved by the Institutional Review Board (IRB) and Ethics Committee (EC) of Peking Union Medical College Hospital (PUMCH) (JS-2492). Informed consent was obtained from all sample donors.

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

Clinical data
Policy information about clinical studies
All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration
Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol
Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection
Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes
Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

Dual use research of concern
Policy information about dual use research of concern

Hazards
Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

| No | Yes |
|----|-----|
| ☐  | ☑️  | Public health
| ☐  | ☑️  | National security
| ☐  | ☑️  | Crops and/or livestock
| ☐  | ☑️  | Ecosystems
| ☐  | ☑️  | Any other significant area
Experiments of concern

Does the work involve any of these experiments of concern:

| No | Yes |
|----|-----|
| □  | □   | Demonstrate how to render a vaccine ineffective |
| □  | □   | Confer resistance to therapeutically useful antibiotics or antiviral agents |
| □  | □   | Enhance the virulence of a pathogen or render a nonpathogen virulent |
| □  | □   | Increase transmissibility of a pathogen |
| □  | □   | Alter the host range of a pathogen |
| □  | □   | Enable evasion of diagnostic/detection modalities |
| □  | □   | Enable the weaponization of a biological agent or toxin |
| □  | □   | Any other potentially harmful combination of experiments and agents |

ChiP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| Data access links | For “Initial submission” or “Revised version” documents, provide reviewer access links. For your “Final submission” document, provide a link to the deposited data. |
|-------------------|--------------------------------------------------------------------------------------------------|
| Files in database submission | Provide a list of all files available in the database submission. |
| Genome browser session (e.g. UCSC) | Provide a link to an anonymized genome browser session for “Initial submission” and “Revised version” documents only, to enable peer review. Write “no longer applicable” for “Final submission” documents. |

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChiP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChiP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChiP-seq data. For custom code that has been deposited into a community repository, provide accession details.

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

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.
### Magnetic resonance imaging

#### Experimental design

| Design type | Indicate task or resting state; event-related or block design. |
|-------------|---------------------------------------------------------------|
| Design specifications | Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials. |
| Behavioral performance measures | State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects). |

#### Acquisition

| Imaging type(s) | Specify: functional, structural, diffusion, perfusion. |
|-----------------|-------------------------------------------------------|
| Field strength | Specify in Tesla |
| Sequence & imaging parameters | Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/TI angle. |
| Area of acquisition | State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined. |

#### Preprocessing

| Preprocessing software | Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.). |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Normalization | If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization. |
| Normalization template | Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI152, ICBM152) OR indicate that the data were not normalized. |
| Noise and artifact removal | Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration). |
| Volume censoring | Define your software and/or method and criteria for volume censoring, and state the extent of such censoring. |

#### Statistical modeling & inference

| Model type and settings | Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation). |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| Effect(s) tested | Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used. |

Specify type of analysis:  
- Whole brain  
- ROI-based  
- Both

Statistica type for inference  
(See Eklund et al. 2016)  
Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correlation  
Describe the type of correlation and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).
**Models & analysis**

| n/a | Involved in the study |
|-----|-----------------------|
|     | Functional and/or effective connectivity |
|     | Graph analysis |
|     | Multivariate modeling or predictive analysis |

**Functional and/or effective connectivity**

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

**Graph analysis**

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

**Multivariate modeling and predictive analysis**

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.