SEAM: a Spatial single nuclEar metAboloMics method for dissecting tissue microenvironment

Michael Zhang (michael.zhang@utdallas.edu)
University of Texas at Dallas
Zhiyuan Yuan
Tsinghua University
Qiming Zhou
Tsinghua University
Lesi Cai
Tsinghua University
Lin Pan
China-Japan Friendship Hospital
Weiliang Sun
China-Japan Friendship Hospital
Shiwei Qumu
China-Japan Friendship Hospital
Si Yu
Peking Union Medical College Hospital
Yongchang Zheng
Peking Union Medical College Hospital
Shao Li
Tsinghua University
Yang Chen
Tsinghua University
Xinrong Zhang
Tsinghua University

Article

Keywords: SEAM, tissue organization, spatial metabolome

Posted Date: August 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-63938/v1

Version of Record: A version of this preprint was published at Nature Methods on October 4th, 2021. See the published version at https://doi.org/10.1038/s41592-021-01276-3.

SEAM: a Spatial single nuclEar metAboloMics method for dissecting tissue microenvironment

Zhiyuan Yuan[1,9], Qiming Zhou[2,9], Lesi Cai[3,9], Lin Pan[6], Weiliang Sun[6], Shiwei Qumu[7], Si Yu[8], Yongchang Zheng[8], Shao Li[1], Yang Chen[1*], Xinrong Zhang[3*], Michael Q. Zhang[1,4,5*]

1, Ministry of Education Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Department of Automation, BNRist, Tsinghua University, Beijing 100084, China

2, School of Life Sciences, Tsinghua University, Beijing 100084, China

3, Department of Chemistry, Tsinghua University, Beijing, 100084, China

4, Department of Biological Sciences, Center for Systems Biology, The University of Texas, Richardson, TX 75080-3021, USA

5, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China

6, Institute of Clinical Medicine, China-Japan Friendship Hospital; National Clinical Research Center for Respiratory Diseases; Institute of Respiratory Medicine, Chinese Academy of Medical Science, Beijing 100029, China

7, Department of Pulmonary and Critical Care Medicine, China-Japan Friend Hospital; National Clinical Research Center for Respiratory Diseases, Beijing 100029, China

8, Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

9, These authors contributed equally: Zhiyuan Yuan, Qiming Zhou and Lesi Cai

*email:

yc@mail.tsinghua.edu.cn

xrzhang@mail.tsinghua.edu.cn

michael.zhang@utdallas.edu

Abstract

Spatial metabolomics can reveal intercellular heterogeneity and tissue organization. To achieve highest spatial resolution, we reported a novel Spatial single nuclEar metAboloMics (*SEAM*) method, a scalable platform combining high resolution imaging mass spectrometry (IMS) and a series of computational algorithms, that can display multiscale/multicolor tissue tomography together with identification and clustering of single nuclei by their *in situ* metabolic fingerprints. We firstly applied *SEAM* to a range of wild type mouse tissues, then delineate a consistent pattern of metabolic zonation in mouse liver. We further studied spatial metabolome in human fibrotic liver. Intriguingly, we discovered novel subpopulations of hepatocytes with special metabolic features associated with their proximity to fibrotic niche, which was further validated by spatial transcriptomics with Geo-seq. These demonstrations highlight how *SEAM* may be used to explore the spatial metabolome and tissue anatomy at single cell level, hence leading to a deeper understanding of the tissue metabolic organization.

Introduction

The hierarchical organization of multicellular organisms is stably maintained by homeostasis at different levels. At the tissue level, such homeostasis is often further modulated by the combination of intracellular gene expression network and extracellular (microenvironmental) signals[1-4]. Cell and its extracellular environment interact dynamically through various signaling mediators, including metabolites, secretome, and ligand-receptor interactions. Metabolites from extracellular environment can significantly influence cell behavior or even transform its identity. For instance, extensive alcohol intake not only activates the detoxification activity of hepatocytes but also alters the epigenetic landscape of hepatocytes[5]. Conversely, cell releasing metabolites can also have impact on its microenvironment. One classic example is basophils and mast cells releasing histamine to increase the permeability of the capillaries when encountering infection[6]. To facilitate a deeper and more systematic understanding of the multi-scale nature of biological processes (e.g. organ development or tumor microenvironment), various single cell omics-technologies have been rapidly developed and utilized[7]. Currently, advanced imaging mass spectrometry (IMS) based techniques are also being made possible to profile a large number of metabolites spatially and/or temporally, providing new dimensional insights to those hierarchical processes[8,9].

In spatially resolved metabolomics studies, different techniques have been developed including matrix-assisted laser desorption/ionization (MALDI-MS)[10], desorption electrospray ionization (DESI-MS)[11], laser ablation inductively coupled plasma (LA-ICP-MS)[12], and secondary ion mass spectrometry (SIMS)[13]. MALDI-MS utilized t-MALDI ion source for imaging of phospholipids and a few other biomolecule classes in thin, matrix-coated tissue sections and cell cultures at a pixel size of about 1–2 μm[14]. With further improvement, 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 tissue[15]. DESI-MS has been utilized to visualize tissue level metabolomic alterations in 256 esophageal cancer patients[11]. Recently, based on SIMS, 3D OrbiSIMS, a label-free IMS with subcellular lateral resolution, and high mass-resolving power, has been developed[16]. These techniques will increasingly be used in future spatial metabolomics applications.

Although the above techniques achieved unprecedented subcellular resolution, several analytical complications still exist, e.g. 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 might bring exogenous substances leading to sabotaging sample integrity[17]. Another cell segmentation strategy exploited convolution neural network (CNN) trained on pixel-wise annotated cells, demanding for huge human expert labour[18]. As for cell fingerprint extraction, the common practice that took the average of pixel profiles within each cell caused the impairment of distributive information[19,20]. These deficiencies hinder the efforts for the quantification of single cell metabolome while preserving spatial information. Consequently, although there have been instrumental-wise improvements for IMS, the downstream analytical methods still require further development for users to fully exploit spatial metabolomic features.

To overcome those deficiencies, we proposed Spatial single nuclEar metAboloMics (*SEAM*), a novel platform leveraging the spatial metabolome provided by SIMS and a comprehensive series of computational algorithms for delineating *in situ* single cell level metabolome and tissue microenvironment. To our knowledge, this is the first study capable of segmenting and analyzing single nuclear metabolic profiles directly on tissue sections. Importantly, *SEAM* is label-free and only requires minimal experimental preparation, which avoids the introduction of exogenous substance and preserves samples' native state. As a proof of principle, we comprehensively calibrated *SEAM* using popular cell cultures, and then systematically scaled up to various mouse tissues, including wild type mouse lung, kidney, small intestine, and liver. Finally, we discovered different hepatocyte metabolic subpopulations and their spatial network organization within the tissue microenvironment in human fibrotic liver.

Results

Overview

SEAM is an integrated platform for qualitative and quantitative analysis of tissue metabolic cell typing and *in situ* microenvironment. The whole pipeline is composed of two main parts: IMS assay and computational analysis suite (Fig. 1a).

As an IMS technology, time-of-flight secondary ion mass spectrometry (TOF-SIMS) provides both mass spectra (chemical information) and ion images (spatial information), of biomolecules on tissue sections (Fig 1a, top left). Typically, hundreds of peaks in a mass spectrum could be extracted from a 400 × 400 μm^2 scan area on a tissue section. Every experiment outputs multiplex SIMS data with 256×256 pixels in spatial resolution, and each pixel is associated with a vector of over 200 selected *m/z* peaks (Fig. 1a and see Methods). With the reference of H&E staining, to facilitate users with quickly viewing of the metabolic spatial pattern across the full spectrum, rather than manually reading hundreds of *m/z* images one by one, *SEAM* provides *SIMS-View* to compress the multiplex SIMS images from hundreds of channels into three, while preserving local and global structures in the feature space (Fig. 1a, bottom left and middle). Then

the three-channel images are mapped to CIELAB color spaces[21] and can be rapidly surveyed by human vision.

To compensate for the potential information loss of dimensionality reduction by taking the advantage of compositional characteristics and spatial continuity, *SEAM* can further build a spatial single nucleus map and delineate the organization of metabolically distinct *in situ* cell subpopulations (Fig. 1a bottom right). More specifically, *SEAM* provides three additional data analysis modules (see Methods): single nucleus segmentation (*SIMS-Cut*, Fig. 2a), single nucleus representation (*SIMS-ID*, Fig. 2b) and differential metabolite analysis (*SIMS-Diff*).

SEAM can resolve metabolomic profiles at single cell resolution on various tissues with different cell densities

To demonstrate the universality and as a sanity check, we tested *SEAM* using mouse liver (Fig. 1a bottom row), lung, kidney, and small intestine samples(Fig. 1b). Qualitative visualization of *SIMS-View* may illustrate the corresponding tissue structures: e.g. in the liver, the metabolites show gradual changes spreading out from the central vein (CV)[22]; in the lung and kidney, the specific structure of the local metabolic niches, such as bronchioles and glomerulus[23]; and in the small intestine, the characteristic anatomic pattern along the intestinal villus axis[24] (Supplementary Figs. 1-3).

In addition to the spectral projection by UMAP in *SIMS-View*, one can selectively add more histological or functional information back by using those different *SEAM* modules through quantitatively characterizing the spatial and compositional information within the single nuclear metabolome. Compared with the *SIMS-View*, clustering results using the single nuclear representation module *SIMS-ID* can mark strong correspondence to the well-established cell types, for example, hepatocytes and endothelial cells in the liver, Clara cells in the lung, as well as enterocytes and lamina propria in the small intestine (Supplementary Figs. 1-3).

Algorithms design and modular data analysis for _SEAM_

SIMS-View is a fast visualization tool designed for SIMS data, which takes advantage of the efficiency as well as the local and global structure preservation of UMAP[25]. It takes multiplex SIMS data as input and outputs a single human-readable image using three steps. First, SIMS data is regarded as 256×256 independent pixels, each represented by a fixed-length vector, and each pixel is feature-wise normalized to avoid feature bias. Next, the 65536 pixels are fed into UMAP to reduce the dimensionality to 3. 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. *SIMS-View* provides a global view of all the ion distribution features in one single image at the pixel level.

To solve the cell segmentation problem, various in situ works used different methods. Some used matched H&E stain[17], others took one simple measurement as input[18,26]. And most of them used super-

vised segmentation, via either pixel-wise classification or modeling the whole image using CNN. Interestingly, based on the visualization of *SIMS-View* results on different samples, the nuclei of cells showed similar color for most cells yet different from other non-nuclear areas (Fig. 1a, b). Therefore, we decided to isolate the nucleus to demarcating 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 (ROIs) using corresponding metabolic markers, for example, adenine (*m/z* 134) as the nuclear marker[16]. The input data format is multiplex by selecting those ion species highly co-localized with nuclei, which is highly consistent across different samples (Supplementary Fig. 4a). And the core of *SIMS-Cut* is an expectation-maximization (EM) algorithm, aiming to solve an optimization problem of a probabilistic graphical model (PGM)[27] which combines a restricted Boltzmann machine (RBM)[28-31], and a Potts model[32,33] (Fig. 2a, Supplementary Fig. 4d). The RBM (Supplementary Fig. 4c) is suitable for modeling the appearance of a multi-image pixel given its label (foreground/background), and the Potts model (Supplementary Fig. 4b) encourages the resulting segmentation masks to be smooth.

To demonstrate the superior performance of *SIMS-Cut*, we compared with several popular unsupervised segmentation algorithms (Supplementary Fig. 5c), using different cell cultures with adenine (*m/z* 134) as the ground truth (Supplementary Fig. 5a, and see Methods). The results showed that *SIMS-Cut* could consistently outperform contestants in all cases visually and quantitatively (Supplementary Figs. 5a-c, 6). To test the suitability on tissue samples, we also applied *SIMS-Cut* on various wild type mouse tissues, ranging from lung, kidney, small intestine, and liver (Supplementary Figs. 7, 8). For the more challenging case, where cells might display distinct sizes and densities, *SIMS-Cut* was finally applied on human liver fibrosis tissues from multiple patients, and all resulted in consistent and satisfactory performance (Supplementary Figs. 9, 10).

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, extracting the metabolic fingerprint of each cell (both nucleus and cytoplasm) can be done by combining its segmentation mask and corresponding SIMS data. Existing works often represented cells by computing the average of all the pixels containing within each cell[19,20], which required strong assumptions like Gaussian or unimodal, and suffered from loss of pixel variation (Supplementary Fig. 11c). To obtain better results, *SIMS-ID* represents cells using the bipartite graph of pixels and cells constructed by a self-supervised learning algorithm[34-36], which can soften the hard labeling produced by *SIMS-Cut* (Fig. 2b and Supplementary Fig. 11a, b). The resulting representation showed superior discriminative power, noise robustness, and pixel distribution preservation.

To test the distinguishing features mentioned above, we constructed 11 datasets (See Methods) containing both mixed cell populations simulated based on single cell line cultures (Supplementary Figs. 5a, 12), and mixed-cultured cells (Supplementary fig. 13). To compare the discriminative power between *SIMS-ID* and the conventional mean representation, we tested supervised classification using KNN equipped with cross-validation and unsupervised clustering using several standard algorithms (K-Means[37], SC3[38], SIMLR[39], T-SNE[40] followed by K-Means, and UMAP followed by HDBSCAN[41]), then applied them

on both representation methods to compare on datasets 4,5,6,7, each containing 4 cell clusters (Supplementary Fig. 12a), whose ground truth is naturally derived *in silico*; and on datasets 10,11, two mixed-cultured datasets, whose ground truth is provided by BrdU/IdU labeling[42] (Supplementary Fig. 13, and see Methods) without affecting on cell metabolic fingerprint (Supplementary Fig. 21). The results showed superior performance of *SIMS-ID* in both supervised (Supplementary Fig. 17) and unsupervised (Supplementary Fig. 18) cases, even in cases of minor fold changes on two feature dimensions (Supplementary Figs. 12a, 17). To evaluate the sensitivity of capturing pixel distribution of cells, we first tested *SIMS-ID* with dataset 3, where it could identify the change of the pixel distribution from the original data to Gaussian (Supplementary Fig. 16), then on dataset 8 and 9 (Supplementary Fig. 12b,c), where *SIMS-ID* could distinguish cell types with unimodal and multimodal distributions (Supplementary Fig. 19), or different joint distributions even on two feature dimensions (Supplementary Fig. 20). To test the robustness to inaccurate segmentation and pixel-wise multiplicative noise, *SIMS-ID* was applied on dataset 1 and 2, and showed consistently better performance than the mean representation (Supplementary Figs. 14, 15). The *SEAM* analyses of datasets 10,11 are shown in Supplementary Figs. 22, 23.

The resulting representation of *SIMS-ID* lies in high dimensional feature space. *SIMLR*[39] is a popular single cell clustering algorithm, which automatically learns cell to cell affinity with multiple kernel ensemble learning, and shows satisfactory performances when combined with *SIMS-ID* (Supplementary Fig. 18). We simply adopted *SIMLR* as our clustering method.

To characterize the key metabolites 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 earth mover's distance (EMD, see Methods)[43] as 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 (WCV).

SEAM reveals cell spatial metabolic states in wild type mouse liver.

Liver is an important metabolic organ consisting of repeating hexagonal-shaped units called lobules[44]. Spatial heterogeneity of metabolic mechanism has been thoroughly investigated using immunohistochemistry (IHC) analyses[45], transcriptome[22], and epigenome[46], but, to our knowledge, single cell level of direct spatial metabolome has not been reported. This allows us to fill up 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 consists of approximately 200~300 ion species after spectral peak selection and filtering (See Methods), and *SIMS-Cut* detected 724 nuclei in the square. To extract metabolic cell fingerprint, we used *SIMS-ID* to represent each cell with a fixed-length vector, which was fed into *SIMLR* to obtain metabolic distinct cell subpopulations. *SIMLR* reached an optimal k = 8, and the resulting 8 metabolically distinct subpopulations correspond to major liver cell types, including Kupffer cells, 2 subpopulations of endothelial cells, and 4 subpopulations of hepatocytes (Fig. 2c).

The identified subpopulations showed specific spatial patterns consistent with the known liver organization (Fig. 2c). Kupffer cells are specialized macrophages in the liver, which typically line on the walls of the sinusoids. Endothelial cells correspond to vascular endothelial cells and liver sinusoidal endothelial cells, typically lying between the crevices of hepatocytes and receiving blood from both the hepatic artery and the portal veins into the hepatic parenchyma[47]. 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[44]. *SIMS-Diff* identified differential ion species among the subpopulations (Fig. 3a, b). We found *m/z* 60, 76, and 77 as metabolic markers of endothelial cells, while *m/z* 134, 181, and 91 enriched in Kupffer cells (*m/z* 134 is reported to be adenine, reflecting the higher nucleus-to-cytoplasm ratio). Hepatocytes, which differ from liver non-parenchymal cells, were characterized by *m/z* 255, 279, and 281, corresponding to the fatty acid metabolism of parenchymal tissue. Interestingly, hepatocyte may be sub-classified by C1, C2, C3, and C4 each showing different metabolic fingerprints (Fig. 3a, b).

Hepatocyte metabolic clusters show a consistent but complementary spatial pattern with liver zonation

Having identified the metabolic heterogeneity among hepatocytes in wild type mouse liver lobule, we searched for differential gene expression corroboration in the literature. Hepatocyte C1 was visually localized around CV, and quantitative analysis revealed that the cells in Hepatocyte C1 showed significantly smaller distances from CV compared with the other hepatocytes ($P < 10^{-9}$, one-side Wilcoxon rank sum test) (Fig. 3d). We also found 6 ion species markers and observed the gradual changes along the liver lobule (Fig. 3c), as well as the zonation pattern of each representative metabolite in single cell level, showing consistent pattern with reported spatial transcriptome[22] (Fig. 3e). Additionally, replicate experiments on different CV regions also showed consistent metabolic patterns and cluster-specific metabolites, indicating the robustness and effectiveness of our method (Supplementary Figs. 24a, 25a-f). We reported *SEAM* results of the liver portal node (PN) as our negative control (Supplementary Fig. 25g, h). Consistent with the spatial expression of *GLUL*[22], the spatial pattern of *m/z* 58, 59, 69, 71, 87, and 101 showed higher expression in the nearest 1~2 layers of hepatocytes from CV (Fig. 3c, e). We further conducted the IHC of two liver zonation markers, Glutamine synthetase (GS), the protein encoded by *GLUL*, and Cytochrome P450 2E1 (Cyp2e1), at the adjacent slides and confirmed liver zonation pattern (Supplementary Fig. 24b-d). This example provided *SEAM* with a positive control that it can accurately and comprehensively characterize the spatial heterogeneity within a well-studied tissue microenvironment.

SEAM identified metabolically different hepatocyte subpopulations associated with the fibrotic niche.

Liver cirrhosis has been a major killer, and progressive liver fibrosis often results in liver cirrhosis[48]. Having

been proven effective in the case of wild type mouse liver, *SEAM* was applied to human liver fibrosis to characterize the metabolic microenvironment around a fibrotic niche. We hypothesized that there should be metabolic alterations of hepatocytes around the fibrotic niche, and such alteration might be associated with the distance between hepatocytes and fibrotic boundary (FBD) at a local scale.

To test this hypothesis, we collected 10 non-tumor tissue regions from 3 liver cancer patients (Supplementary Table. 2) and made a sequential 10μm slides for SIMS and other assays. We selected 4 regions from one sample, each containing a fibrotic niche, and conducted SIMS experiments (Fig. 4a, b). The resulting data consists of approximately 200~300 ion species after spectral peak selection and filtering (See 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, 716, 546, and 682 nuclei in four square regions respectively. *SIMS-ID* and *SIMLR* were subsequently performed to get metabolically distinct cell subpopulations. The consistent manifolds and clusters shown by UMAP (Fig. 4c middle column) and the spatial single nucleus map (Fig. 4c right column) confirmed the reliability and robustness of *SEAM*. The identified subpopulations, corresponding to Kupffer cells, immune cells, fibroblasts, endothelial cells, and 3 subpopulations of hepatocytes, exhibited the specific spatial distributions (Fig. 4d) and the matching metabolic fingerprints (Fig. 4e). The correspondence and incongruity between cell subpopulations of human and mouse liver samples were also analyzed (See Methods and Supplementary Fig. 26).

Intriguingly, we observed that Hepatocyte C1 was visually localized near the FBD, and its associated metabolic markers, e.g. *m/z* 69, 55, and 57, showed the consistent spatial pattern across 10 regions (Fig. 4f, g, and Supplementary Figs. 27, 28). To quantify the association between the hepatocyte metabolic alteration and the distance to the FBD, we separately conducted two statistical analyses on 10 regions of 3 patients (Supplementary Table. 2) given defined FBD (see Methods and Supplementary Fig. 30 second column): the distance from FBD to hepatocyte C1/C2 (distance-based analysis), and the normalized count ratio between hepatocyte C1 and C2 (count-based analysis). Using R1 as a demonstration, we first defined 5 zones (zone 0~4) with increasing areas (Fig. 4h left), each representing an accumulative territory between the FBD and the corresponding parallel strip (parallel strips are indicated by gray solid lines, and the accumulative territories of zones are indicated by gray dotted brackets), then the distances from FBD to Hepatocyte C1/C2 within the 5 zones were subsequently summarized by a series of paired boxplots (Fig. 4h right, n=10). Meanwhile, we calculated the normalized count ratio between Hepatocyte C1 and C2 within an area as a function of the distance from the outer edge (indicated by the gray solid line in Fig. 4i left) to the FBD (Fig. 4i right, n=10). The result of the distance-based analysis showed that Hepatocyte C1 was significantly closer to FBD than C2 to FBD within the 5 zones (one-side Wilcoxon rank sum test, Fig. 4h right, n=10), and the relative proximity exhibited high similarity across 10 regions (Supplementary Fig. 30 third column). Complementarily, the count-based analysis showed that the normalized count of C1 is consistently higher than C2, specifically, C1 was about ~30-50% denser than C2 within 100μm (a typical hepatocyte size is ~25μm) to the FBD and reduced quickly to about the same level as C2 after ~350μm (Fig. 4i right, n=10), and this trend was highly similar across 10 regions (Supplementary Fig. 30 fourth column). Detail of FBD determination, zone partition, distance, and normalized count ratio calculation, as well as other necessary terms definition, is exactly described in Methods. The above statistical analyses verified our hypothesis that the metabolic alteration of the hepatocyte subpopulations might be associated with the spatial proximity to the fibrotic niche. To verify the variation of microenvironment was not only reflected at the metabolic level, we subsequently performed Geo-seq, a spatial transcriptome assay at the same ROIs of different hepatocyte subpopulations.

Spatial transcriptome validated metabolism associated gene expression alteration in heterogeneous hepatocytes identified by *SEAM*

To get a deeper understanding of *SEAM* results, we performed Geo-seq with a modified protocol (See Methods) of the transcribed RNA samples isolated from the tissues of the corresponding ROIs from the adjacent slides (Fig. 5a, b, and Supplementary Fig. 31). To increase reproducibility, multiple adjacent slides were used (Supplementary Fig.32-35). 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 fibrotic niche and enriched with ions species *m/z* 69 series were defined as Hepa$^{69\text{-high}}$, whereas Hepatocyte C2, which were distal and not enriched with ions species *m/z* 69 series were defined as Hepa$^{69\text{-low}}$. We also collected the fibrotic regions as the FB samples. In total, 15 cDNA libraries were constructed successfully (Hepa$^{69\text{-high}}$ n=6, Hepa$^{69\text{-low}}$ n=5, and FB n=4). Principle component analysis (PCA) plot indicated that two different groups (Hepa$^{69\text{-high}}$ -proximal and Hepa$^{69\text{-low}}$ -distal) of hepatocytes shared higher similarity relative to FB samples (Fig. 5c). More importantly, Hepa$^{69\text{-high}}$ samples were consistently closer to FB samples than Hepa$^{69\text{-low}}$ to FB samples in PCA space (Fig. 5c and Supplementary Fig. 36). To validate the expression pattern of each group, we first compared gene expression profiles between hepatocytes (i.e. Hepa$^{69\text{-high}}$/ Hepa$^{69\text{-low}}$) and FB, then performed gene ontology (GO) enrichment for both up-regulated and down-regulated differentially expressed genes (DEGs) (See Methods and Supplementary Fig. 37, 38). 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* & *SAA1*), fibrosis (*TGFB1, PDGFB* & *COL4A1*), and immune response (*IGHM*, *IGHG3* & *IGHV4-59*). Both hepatocytes groups showed high levels of hepatocyte marker genes. Whereas genes typically activated in fibrotic regions for fibrosis and immune response were highly expressed in FB samples (Supplementary Fig. 39). There were 718 differentially expressed genes (DEGs) fitting into the criteria of adjust P-value < 0.05 and log fold change (LFC) standard error < 3. The expression heatmap indicated that these genes had different expression patterns between the proximal hepatocytes (Hepa$^{69\text{-high}}$) and the distal (Hepa$^{69\text{-low}}$) (Fig. 5d). We inputted the DEGs for GO enrichment analysis (Fig. 5e). There were 17 genes enriched in the first GO entry, 16 of them were consistently higher in Hepa$^{69\text{-high}}$ than Hepa$^{69\text{-low}}$ (Fig. 5f). Genes of solute carrier transporters families with different functions were enriched in the fibrosis proximal (Hepa$^{69\text{-high}}$) group, indicating the corresponding metabolite transmembrane exchange activities were elevated.

Discussion

In this study, we have developed *SEAM*, a platform combining experiments, and computational algorithms to quantitatively characterize the metabolic intra- or inter-cellular features with multiscale spatial resolution. Unlike other IMS instruments such as DESI (40–60μm)[11], SIMS can provide a high spatial resolution

allowing one to visualize detailed metabolic structures in tissue histology. With fast and minimal sample processing, SIMS maximumly preserves the native state of samples. Given the nature of SIMS, although it breaks most of the molecules into fragments, making it more difficult to annotate (a common challenging issue for MS studies), it produces high multiplexity of metabolic features with the potential of characterizing cell and fine tissue microenvironment. Benefiting from both high spatial resolution and high multiplexity of SIMS, the algorithms of *SEAM* start solely from the features generated by SIMS and run a pipeline enabling metabolic analysis from pixels to single nuclei, then to the selected metabolic molecules with spatial information annotated. Previously, there have been reports on spatial metabolic features at tissue level or *in vitro* single cell level[16]. But, to our knowledge, this is the first study capable of segmenting and analyzing single nuclear metabolic profiles directly on tissue sections. In addition, this algorithmic pipeline is principally scalable to other spatial omics studies based on other IMS platforms, transcriptomics, and proteomics with minimum adjustments, and it's also easy to work together with bioinformatics tools such as CIPHER to predict and prioritize disease-related metabolic molecules[49].

Apart from the scalability of *SEAM*'s algorithms, we have demonstrated that the range of *SEAM* applications could cover from *in vitro* cell culture assays to various tissue samples. Firstly, in the mixed cell-cultured assay, *SEAM* could easily deconvolute the different cell lines co-cultured together. Additionally, in different wild-type murine tissue samples, *SEAM* successfully segmented single nuclei without extra labeling required. The single nuclear metabolic profile analysis was also consistent with conventional tissue histological characterization (Supplementary Figs. 1-3). Specifically, in the liver, a spatially well-orchestrated but complex organ, the CV-PN axis zonation has been well-established at single cell transcriptome level in wild type mouse[22]. We observed consistent zonation patterns at single cell level in CV centered region with the gradational decrease of certain characteristic metabolites. Lastly, we found that hepatocyte subpopulations (among which, to our knowledge, the novel C1 has never been reported before) differentiated by different metabolic features were also transcriptionally distinct shown by Geo-seq (Fig. 5c-f). The elevated expression level of solute carrier genes can potentially explain the enrichment of a list of metabolite species found by *SEAM* (Fig. 4). These genes are involved amino acid transport (*SLC36A4*, *SLC3A2* & *SLC38A9*)[50-52], phosphate transport (*SLC17A2* & *SLC17A4*)[53] and Gamma-Aminobutyric Acid (GABA) transport (*SLC6A12*)[54]. *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 more close to the fibrotic region[51]. It indicates that spatial microenvironment differences could influence cellular metabolic homeostasis, which may in turn further alter the gene regulation and downstream response due to cell adaptation and genetic/epigenetic feedback.

In summary, *SEAM* provides a high spatial resolution single nuclear metabolome profiling pipeline requiring minimal sample preparation and labeling. It is automatically 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 for or complementary to other conventional assays. With future improvement of IMS resolution and molecule annotation capability, *SEAM* would be able to provide more detailed spatial metabolome profiles with higher resolution and broader functionality.

ONLINE METHODS

IMS experiments

TOF-SIMS 5 (ION-TOF GmbH, Münster, Germany) equipped with a Bi liquid metal ion gun (LMIG) is used in this study, collected TOF-SIMS spectra and images of tissue samples using a 30 keV Bi_3^+ LMIG with a high spatial resolution (HSR) mode. The Bi_3^+ current in the HSR mode was 0.1 pA (100 ns pulse width, unbunched beam). The total $Bi3^+$ accumulated ion dose was about 2.0 × 10^{10} ions/cm^2, the typical probe sizes of the Bi_3^+ LMIG was ~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_3^+ LMIG was operated at a cycle time of 150 μs (mass range: 0 ~2000 u). Negative spectra were mass-calibrated using CH_2^-, O^-, OH^-, PO_2^-. A flood gun with low energy electrons was used to compensate for charge buildup on sample surface. A 10-keV Ar_{2500}^+ commercial gas cluster ion gun (GCIB) was used as a sputter gun (rastering over a 550 × 550 μm^2 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, a high mass resolution (HMR) mode was used with 0.8 pA (<1 ns pulse width, bunched beam) Bi_3^+ current, the mass resolutions (measured at C_2H^-) were typically>6000. The total $Bi3^+$ LMIG accumulated ion dose was between 10^{11} and 10^{12} ions/cm^2, rastering over a 300 × 300 μm^2 area with 256 × 256 pixels. The Bi_3^+ LMIG was operated at a cycle time 150 μs (mass range: 0 ~2000 u). Negative spectra were mass-calibrated using CH_2^-, O^-, OH^-, PO_2^-. A flood gun with low energy electrons was used to compensate for charge buildup on sample surface. A 10-keV Ar_{2500}^+ commercial gas cluster ion gun (GCIB) was used as a sputter gun (rastering over a 450 × 450 μm^2 area, incident angle 45°) to carry out the depth profiling. A final 2D image was an overlay of 50-80 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 bellow: mass range 50-500; minimum counts 10000; minimum signal/noise ratio 1000. 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 (ASCII) mode data file by the SIMS built-in data processing software SurfaceLab, which contains three columns corresponding to the X-axis, Y-axis coordinates and signal intensity values.

Biological experiments

Cell culture. Human non-small cell lung cancer cell line A549, human cervix carcinoma cell line Hela, murine hepatoma cell line Hepa 1-6 and murine liver epithelial cell line NCTC 1469 cell lines were grown on microscope cover glass (CITOGLAS, China) with Dulbecco's Modified Eagle Medium (DMEM) (Gibco, USA) containing high glucose, L-glutamine, sodium pyruvate and 10% dialyzed, heat-inactivated FBS (Gibco, USA). Human mammary gland cell line MCF 10A was grown on microscope cover glass (CITOGLAS, China) with DMEM/F12 (1:1) (Gibco, USA) containing insulin 10ug/ml, EGF 20ng/ml, cholera toxin 100ng/ml, hydrocortisone 0.5mg/ml and 5% equine serum. Human breast adenocarcinoma cell line MDA-MB-468 cell line was grown on microscope cover glass (CITOGLAS, China) with L-15 medium containing 10% FBS (Gibco, USA) and free air exchange.

BrdU cell mix-culture experiment. Following protocol from the previous study, A549 and Hela cell lines were both cultured with and without 20μM BrdU (Sigma, USA) for 48 hours before seeding. A549 with BrdU were then replated with non-BrdU Hela at the same density on microscope cover glass (CITOGLAS, China) for 20 hours and vice versa for non-BrdU A549 and Hela with BrdU. The same mix-culture procedure for IdU (Sigma, USA) was applied at Hepa 1-6 and NCTC 1469 cell lines.

Mice. 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, 22-26℃ with sterile pellet food and water ad libitum.

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)).

Intrahepatic cholangiocarcinoma (ICC) patient non-tumor liver tissues. The ICC non-tumor liver tissues were obtained from leftover pieces from surgery. 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).

Tissue section preparation. Mouse and human tissues were isolated individually and embedded in Optimum Cutting Temperature (O.C.T) compound (SAKURA, USA), then snap-frozen in liquid nitrogen. Cryo-section were performed using CM1900 Cryostat (Leica, Germany) to obtain 3μm ~ 10μm continuously adjacent sections.

Histology staining. Tissue cryo-sections were thawed at room temperature for 5 min then washed in PBS twice, 5min each time. Slides were fixed in 4% paraformaldehyde (PFA) for 20 min at room temperature then washed in PBS once. H&E stainings were then performed using the H&E staining kit (Leagene, China). Images were obtained from Axio Scan. Z1 (ZEISS, Germany) or Cytation5

(Biotek, USA).

Immunohistochemistry. Tissue cryo-sections were thawed at room temperature for 5 min then washed in PBS twice, 5min each time. Samples were permeabilized and blocked in 5% BSA solution (Sigma, USA) with 0.4% Triton-X100 (AMRESCO, USA) for 2h at room temperature. Dilute and apply primary antibody in PBS with 0.1% Triton-X100 with suited concentration according to each antibody and incubate in a humid dark chamber at 4℃ overnight. Wash three times in PBS with 0.1% Triton-X100, 10min each. Dilute and apply secondary antibody in PBS with 0.1% Triton-X100 and incubate in a humid dark chamber at room temperature for 2h. Wash three times in PBS with 0.1% Triton-X100, 10min each. Slides were mounted using ProLong™ Gold Antifade Mountant (ThermoFisher, USA). Images were captured either by LSM780 confocal microscope (ZEISS, Germany) or Cytation5 (Biotek, USA).

Modified Geo-seq. A spatial transcriptome analysis method, Geo-seq, previously described by Chen, Jun, et al[55]. A modified version was adopted. Tissue cryosections were mounted on the PEN membrane slide and stored at -80 degree freezer for short term storage. Slides were stained in 0.5% cresyl violet and dehydrated in serial ethanol. Tissue blocks were obtained in a 0.2 ml PCR tube by LMD7000 (Leica, Germany). Buffer RLT (Qiagen, Germany) with DTT (Sigma, USA) were added and shaken vigorously for tissue lysis and RNA release. RNA Clean beads (Vazyme, China) 1.8x were added to isolate total RNA. Prepare annealing procedure in the same tube with 3ul H2O, 1ul dNTP, 1ul Oligo(dT), and 0.5ul RNase Inhibitor (RI) (Life Technologies, USA). Incubated at 72 degrees 3min and immediately transfer in ice for 2min. Prepare reverse transcription reaction in the same tube with 2ul 5x RT buffer, 0.5ul DTT, 0.5ul RI, 0.5ul Template Switch Oligo (TSO, Sangon Biotech, China), 1ul Maxima reverse transcriptase (Life Technologies, USA). Incubate with 50 degrees with 1 hour and deactivate reverse transcriptase with 85 degrees for 5 min. Amplified the first strand product with 12.5ul 2x KAPA HIFI HotStart ReadyMix (Sigma, USA), 0.5ul TSO-PCR primer (Sangon Biotech, China) and 2ul H2O. The reaction condition was 95 degrees 3min, 98 degree 20s, 67 degree 15s, 72 degrees 6min for 21 cycles, and 72 degrees for 5min. PCR product was purified with 0.8x DNA Clean beads (Vazyme, China). The next generation sequencing (NGS) library was then constructed by TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme, China). Libraries were sequenced by Illumina Xten Pair-end 150bp by Annoroad.

RNA-seq data processing and analysis

RNA-seq data were firstly performed with adaptor removal and quality filtering by Trim Galore[56]. The qualified reads were then mapped to the human gencode reference genome using STAR and generated BAM files[57,58]. Duplication was removed by PICARD (http://broadinstitute.github.io/picard/) for all the BAM files. Read count for each gene was performed by HTSeq-count with reference to gencode human gene annotation, release 32 (GRCh38.p13)[57,59]. Different gene expression analyses were analyzed using DESeq2 in R[60].

***SIMS-Cut* framework**

Given an M×N×N SIMS data, with M filtered metabolic peaks and N×N image as input, *SIMS-Cut* first select m metabolites co-localizing with nucleus (Supplementary Fig. 4a), and then iteratively solves a maximum a posteriori (MAP) problem (Supplementary Fig. 4d) to get an N×N binary matrix Y that indicates a nucleus.

$$Y_{ij} = \begin{cases} 1 & \text{nuclei region} \\ 0 & \text{otherwise} \end{cases} \quad i,j \in [1,N] \tag{1}$$

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

$$Y^* = \underset{Y}{\operatorname{argmax}}\, p(Y|X) \tag{2}$$

where

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \propto P(X|Y)P(Y) \tag{3}$$

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

As with traditional hidden Markov random field (HMRF) based image segmentation[61,62], *SIMS-Cut* uses a similar graphical model, consisting of $P(Y)$, the smoothing model for unknown label field Y before guarantee spatial homogeneity, and $P(X \mid Y)$, the data model for the conditional distribution of pixel metabolic profiles X given corresponding pixel label.

Smoothing model. The label prior, $P(Y)$ is modeled as a special Markov random field (MRF), called Potts model[32]. According to the Hammersley-Clifford theory[63,64], P(X) follows a Gibbs distribution[65]:

$$P(Y) = \frac{1}{Z}\exp\left(-U(Y)\right) \tag{4}$$

Where U is called energy function, which is calculated by summing over the potential of all second-order cliques V, each clique corresponds to a pair of neighboring pixels(e.g. the 4-neighborhood system). Z is called a partition function, making $P(Y)$ a valid probability density function (pdf).

$$U(Y) = \sum_{(i_1,j_1),(i_2,j_2)\in \text{doubletons}} V(y_{i_1,j_1}, y_{i_2,j_2}) \tag{5}$$

V is defined on doubleton, penalizing the heterogeneity of labels.

$$V(y_{i_1,j_1}, y_{i_2,j_2}) = \begin{cases} -1, \text{if } y_{i_1,j_1} = y_{i_2,j_2} \\ +1, \text{if } y_{i_1,j_1} \neq y_{i_2,j_2} \end{cases} \tag{6}$$

Data model. According to the graphical model (Supplementary Fig. 4b), and d-separate[27],

$$P(X|Y) = \prod_{i,j \in [1,N]} P(x_{ij}|y_{ij}) \tag{7}$$

While the multivariate Gaussian distribution is typically suited for the data model of color image segmentation[66,67], its model capacity is limited and its assumptions are too strong for SIMS data. Instead, we use Restricted Boltzmann Machines (RBM)[28-31] to model the conditional distribution of data intensities given label assignment.

RBM as a generative model is typically a two-layer bipartite undirected graph. It's composed of a visible layer which is m dimensional metabolic profile in our case and a hidden layer which is a kind of d dimensional memory providing model capacity. In theory, RBM is a Universal approximation for any pdf with a large enough number of hidden layers[30]. Here we use two separate RBMs to model $P(x_{ij}|y_{ij} = 0)$ and $P(x_{ij}|y_{ij} = 1)$ respectively, and we describe one RBM in the following. For the sake of notation simplicity, in the following, we use $V = [v_p], p \in [1, m]$ to denote x_{ij} (the subscript is removable thanks to the conditional independence given by (7)).

The graphical model of RBM is shown in Supplementary Fig. 4c. $H = [h_q], q \in [1, d]$ is the hidden layer variable, and V is the visible layer variable. $C = [c_q], q \in [1, d]$, $B = [b_p], p \in [1, m]$, and $W = [w_{pq}], p \in [1, m], q \in [1, d]$ are parameters. The joint probability density function is:

$$P(V, H) = \frac{1}{Z} e^{-E(V,H)} \tag{8}$$

where E is the energy function:

$$E(V, H) = -\sum_{p=1}^{m} \sum_{q=1}^{d} w_{pq} h_q v_p - \sum_{p=1}^{m} b_p v_p - \sum_{q=1}^{d} c_q h_q \tag{9}$$

and Z is the partition function:

$$Z = \sum_{V,H} e^{-E(V,H)} \tag{10}$$

The probability that an RBM model assigns a vector V, e.g. x_{ij} is given by (8).

$$p(x_{ij}|y_{ij} = a) = RBM(V; W^a, C^a, B^a) = \frac{1}{Z^a} \sum_{H} e^{-E(V,H)} \tag{11}$$

$$= \frac{1}{Z^a} \prod_{p=1}^{m} e^{b_p^a v_p} \prod_{q=1}^{d} (1 + e^{c_q^a + \sum_{p=1}^{m} w_{pq}^a v_p})$$

Note that the superscripts indicates the parameters of specific RBM.

Partition function of RBMs Estimation. For a specific pixel given its segmentation label a, the log probability that RBM assigns metabolic profiling x_{ij} is computed as:

$$\log P(x_{ij}|y_{ij}=a) = -F^a(x_{ij}) - \log Z^a \tag{12}$$

Here $F^a(x_{ij})$ is the free energy of RBM corresponding to class a, which can be rapidly calculated. To estimate the partition function Z, we build a softmax model to classify x_{ij} at every pixel to its label y_{ij}:

$$\log P(y_{ij}=a|x_{ij}) = \frac{e^{-F^a(x_{ij})-\log Z^a}}{\sum_{y_{ij}} e^{(-F^{y_{ij}}(x_{ij})-\log Z^{y_{ij}})}} \tag{13}$$

MAP. Our objective can be an expression as:

$$\begin{aligned} \operatorname{argmax}_Y \log P(X|Y) + \log P(Y) &= \operatorname{argmax}_Y \sum_{i,j} \log P(x_{ij}|y_{ij}) + \log P(Y) \\ &= \operatorname{argmax}_Y \sum_{i,j} \log \mathrm{RBM}(x_{ij}; W^{y_{ij}}, C^{y_{ij}}, B^{y_{ij}}) + \log P(Y) \end{aligned} \tag{14}$$

It's a nonconvex problem, we develop an EM-style algorithm to alternating between two steps to reach a locally optimal point iteratively.

Each iteration of *SIMS-Cut* consists of three sub-problems, each of which can be solved efficiently. The input of each iteration is the segmentation mask output by the previous iteration, and the first level's input is simply k-means clustering of an input image. The segmentation mask will converge in no more than 20 levels according to our experiments.

In the first sub-problem, the parameters of two RBMs are estimated given the label of each pixel input from the previous level. Estimated as the parameters of RBMs, solving the partition function is time-consuming, thus the second sub-problem bypasses the obstacle and at the same time controls the bias of each iteration with the help of a simple binary classification task[68]. And the third and last sub-problem uses the well-known graph-cut algorithm[32,33,69,70] to obtain the pixel labels, i.e. the segmentation mask for the current iteration. As the process of iterations, the intermediate segmentation masks gradually shrinkage, while local homogeneity and nucleus centralization are simultaneously kept. Finally, the reaping algorithm is used to salvage as many isolated nuclei as possible during the shrinkage process. More details about solving these sub-problems are as follows:

Initialization of Y, C, B, W. The parameters of RBM, e.g. C, B, W is randomly initialized using a Gaussian with zero mean and unit variance. The label assignment Y is initialized using k-means.

Sub-problem 1: Fix Y to update C,B,W. This step is Equivalent to learn two independent RBMs. Since Y is given, the training data for the two RBMs can be extracted from X. An efficient learning algorithm, persistent contrastive divergence (PCD)[71,72] can be applied. Also, PCD algorithm is based on Maximum likelihood estimation, leading to an increase of objective.

Sub-problem 2: Fix Y,C,B,W to update two partition functions. Partition function estimation of

RBM is time consuming even if all its parameters are known. Based on the efficient way to deal with the unknown partition functions[29,68], we build an auxiliary binary classification task and treat the two partition functions as parameters to estimate. Furthermore, a hyper-parameter beta can be tuned to control the process of iteration (see Details and online code).

Sub-problem 3: Fix C, B, W, and two partition functions to update Y. This step is equivalent to an energy minimization problem, and global optimized Y can be efficiently using a graph cut algorithm.

Reaping. Using our parameter setting, the above algorithm converges to all-zeros Y within 20 iterations. Because of the spatially different contrast of SIMS image, some nucleus may be lost during the iteration. We develop an enhancement algorithm to maintain the intermediate identified nucleus.

Due to the bias, as the levels grow high, the region of within-nucleus gets smaller. But the MRF-based segmentation makes the intermediate segmentation mask of each level homogenous and evident. To get the final non-connected nucleus mask, a reaping algorithm is proposed in Algorithm 1.

Algorithm 1

Input: M_k*: segmentation masks for each level;* A_u*: upper bound of nuclei area;* A_l*: lower bound of nuclei area;*

Step 1: Create a queue Q to maintain isolated segments. Create an all-zeros mask M_{rst}

Step 2: Initialize Q by putting all isolated segments of level 2 to the head of Q; initialize M_{rst} *using* M_2*.*

Step 3: pop a segment q from head of Q, set the segment region of M_{rst} *to zeros.*

Step 4: for l from k+1 to K, where q belongs to M_k

if l reaches K

then set the q region of M_{rst} *to ones*

if two or more segments in M_l *belongs to q*

then push these new segments to tail of Q;

set these segment region of M_{rst} *to ones;*

break

Step 5: return to step 3, until Q is empty

Step 6: return M_{rst}

Implementation details and parameters setting. We use correlation distance to select top 20 co-localized ions with Adenine (*m/z* 134), whose conditional probabilities given labels are modeled

by two label-specific RBMs. K-means on a 134 intensity map is used to initialize the segmentation label, we set k=4 and set clusters with the lowest center as background, other 3 clusters as foreground. For the smoothing model, we use the 4-neighborhood system. For the data model, we use two Generative RBMs, each with 20 visible nodes and 50 hidden nodes. For RBM training, persistent contrast divergence (PCD) is used for 10 epochs each level.

For convenience, we use a Matlab toolbox for RBM modeling and training[73]. When optimizing the energy minimization problem, we use the Matlab version of the Boykov-Kolmogorov algorithm[69] provided by https://vision.cs.uwaterloo.ca/code/. The original algorithm takes the smoothing model as a neighbor weights matrix, whose format is described in the code comment, but we modified the matrix by average filtering with a window size of 21 to provide more smooth quality (optional). To weight between the data model and the smoothing model, we divide the weights matrix by a constant (typically 5~10, we use 5 for best practice).

To bypasses the time-consuming partition function estimating problem of the two RBMs, a simple classifier is performed during each iteration. Note that the exact value of the two partition functions needn't be known[68,74], the difference matters instead. We first calculate the free energy of all $N \times N$ pixels separately using the parameters of the two RBMs and sort the difference. Then sort the difference and take every N-1 interval as classification cutoff. At the same time, one confusion matrix for each cut off is maintained, so N-1 F measures controlled by beta corresponding to every interval can be calculated. Finally, the partition function difference with the best F measure is selected. The beta parameter (typically 0.5~1) is tuned to control the convergence process.

During the *SIMS-Cut* procedure provided in the methods section, due to the beta parameter, as the levels grow high, the region of within-nucleus gets smaller. But the MRF-based segmentation makes the intermediate segmentation mask of each level homogenous and evident. To get the final non-connected nucleus mask, a reaping algorithm is proposed. The detail is as follows: Suppose after L level's segmentation, *SIMS-Cut* converges to an all-background segmentation mask. Since each level is an intermediate segmentation mask given beta and upper level's estimated parameters. The hierarchical structure can be modeled as a tree, whose nodes are nucleus of all levels, root is a dummy node, the second highest level is the nucleus of first segmentation. Node i is the child of node j if i belong to the next level of j, and the segmentation region of i is a subset of segmentation region of j. The leaf nodes are nucleus in the lowest level, the last level of *SIMS-Cut* procedure. From top to bottom, nodes are split alongside the tree structure, and the reaping algorithm can capture nodes that are optimally split (i.e. according to *m/z* 134 intensity).

SIMS-ID Framework

After *SIMS-Cut*, hundreds of separated nuclei has 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[75], 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 analysis, like clustering, visualization, and so on.

Data preprocessing. Due to the variability of tissue thickness, and variation in ionization and detector efficiency, SIMS data need to be preprocessed. We use Variance-stabilizing normalization[76], specifically, the median spectrum is used to estimate the normalization factor, and logarithm was used as variance-stabilizing transformation.

Motivation. *SIMS-ID* is based on the observation that the outputs of a trained neural network contain much richer information than just a one-hot classifier. Hinton, G. et al observe that mutual similarity between classes can be distilled from a trained softmax based neural network classifier, e.g. an image of a BMW, may only have a very small chance of being mistaken for a garbage truck, but that mistake is still many times more probable than mistaking it for a carrot[75]. Lu, Y. applies factor analysis to reveal the visual similarity of image classes[77]. Wu, Z. utilizes a similar concept to train an instance-level classifier as an auxiliary task for unsupervised representation learning[78].

Auxiliary classifier construction. *SIMS-ID* first constructed a multiple-layer dense neural network armed with a softmax activation at the last layer for classification, then preprocessed pixel data are input to classify each pixel to the right nuclei, after training, the temperature of softmax output is raised to a user-set value to soften the probabilistic distribution, and finally the distilled softmax output of each input pixel can be considered a similarity between the nuclei to which that pixel belongs and other nuclei, from that pixel's point of view. Further experiments showed that the overfitting of the auxiliary classifier doesn't hurt the performance of afterward analysis.

Interpretation. The auxiliary classifier can naturally capture apparent similarity among classes, i.e. nucleus without being directed to do so. The distilled information, i.e. the high-temperature softmax output of each pixel can be expressed as a $\mathrm{P} \times \mathrm{C}$ matrix PCM, where P is the number pixels within all nucleus, and C is the number of the identified nuclei. The matrix can be interpreted using three distinct ways.

Nucleus Similarity measure from each pixel's view. Each row of PCM can be considered as a similarity measure between the corresponding nuclei and other nuclei. If the i-th pixel belongs to the j-th nuclei, for the i-th row of PCM, after dividing each element by the j-th element of the row, we can get a normalized similarity vector, whose j-th element is 1. Moreover, in the auxiliary classification phase, the more easily confused with the correct class, i.e. nuclei, the higher the corresponding element of normalized PCM is.

Nucleus representation of multiple instance learning. In the multiple instance learning (MIL) literature[79,80], a bag of instances can typically be represented by similarities between this bag and all instances. A column of normalized PCM can be considered as the probability of each pixel

belonging to that nucleus.

The adjacency matrix of nucleus-pixel bipartite graph. The original one-hot pixel-nucleus relationship doesn't provide any information between nuclei. After knowledge distillation, the one-hot relationship is shattered to a more smooth knowledge, from which nucleus relationship can be discovered. The normalized PCM can be interpreted as an adjacency matrix identifying to the bipartite graph, and the (i,j)-th entry of PCM is the weight between the i-th pixel and j-th nucleus.

Parameters setting and network structure. The pixel classification network structure is shown in Supplementary Fig. 11b. We use multiple layer perceptrons except for the last layer, ReLU[81] activation function for each layer, softmax as probability output, and Adam[82] as an optimizer. The number of neurons of the first layer is M, the number of observed metabolites and the number of neurons of the last layer is the same as the number of the nucleus. Since overfitting doesn't hurt the representation performance according to our experiment, we set all the pixels as training data, and the number of training epochs is set to 100~300.

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 400~1000, and the length of the representation vector for each cell is equal to the number of pixels within segmented cells, typically ranging from 5000~15000. With the consideration of both data characteristics and experimental performance (Supplementary Fig. 18), we apply SIMLR[39], a single cell clustering algorithm, which automatically learns the low-rank similarity matrix by means of multiple kernel ensemble. Besides, SIMLR also provides 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 the feature's discriminative power to tell clusters apart. Due to the nature of our data, the traditional two-sample test can’t 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 WCV to evaluate the compactness of the similarity matrix, where BCV is between cluster variation, and WCV is within cluster variation. For each feature, we use EMD (earth mover's distance)[43] as a metric for two nuclei represented by histograms, and the variation can be simply evaluated by summing all pairwise distances.

Earth mover’s distance as a valid metric for histograms. EMD originally arose in the field of

optimal transporting problems, recent studies show that it can be fruitfully applied to compare histograms. Thus, if one thinks of a histogram as a pile of dirt, then the EMD between two histograms is the minimum cost required to move the dirt in one pile to the other. Here, the cost is defined as the amount of dirt moved multiplied by the distance it is moved. Univariate EMD has several nice properties: (1) it's a true distance; (2) it doesn't need to assume the distribution form of histograms; (3) it's computationally efficient.

Discriminative feature identification using EMD. For each feature, a C×C EMD matrix can be calculated, whose (i,j)-th entry is the distance between i-th nuclei histogram and j-th nuclei histogram. Then we use the given clustering result to sort the rows and columns, and discriminative features may pose a block diagonal EMD matrix. The ratio between BCV and WCV can be used to evaluate the feature's discriminative power between two clusters. BCV can be simply calculated by summing over all pairwise distance between the two clusters, and similarly, WCV can be simply calculated by summing over all pairwise distance within two clusters independently.

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 (MIA)[83]. Specifically, we ranked metabolites by the score computed using SCANPY[84], which is z-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. And the significance of the intersection of gene sets between any pair of clusters was inferred using the hypergeometric distribution. The MIA 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 following terms: $\mathrm{FBD_{Ri}}$ is the fibrotic boundary of region Ri; $\mathrm{PSP(j,FBD_{Ri})}$ is a parallel strip whose distance to $\mathrm{FBD_{Ri}}$ is equal to j μm; $\mathrm{AREA(j,i)}$ is the territory between $\mathrm{FBD_{Ri}}$ and $\mathrm{PSP(j,FBD_{Ri})}$; $\mathrm{Zone(j,i)}$ is short for $\mathrm{AREA((j+1)\times 100,i)}$; $\mathrm{CFBD(cell_i,Zone(j,k))}$ is the distance (μm) between $\mathrm{cell_i}$ and $\mathrm{FBD_{Rk}}$ within $\mathrm{Zone(j,k)}$; $\mathrm{NCC(population_i,area_1,area_2)}$ is the ratio between the number of cells in $\mathrm{population_i}$ within $\mathrm{area_1}$ and the number of cells in $\mathrm{population_i}$ within $\mathrm{area_2}$.

The FBD is approximated according to *SIMS-View* and spatial single nucleus map (Supplementary Fig. 30). Coming to cases where FBD couldn't be well fitted by a single line segment, polylines are used, and the distance to FBD is simply adjusted to be the smallest among distances to all line segments.

The statistical analysis of Fig. 4h is conducted as following: In zone j, $j \in \{0,1,2,3,4\}$, the red boxplot is the summarization of $\{\mathrm{CFBD}(\mathrm{cell_i}, \mathrm{Zone(j,k)}) \mid k \in \{1,2,3,4,5,6,7,8,9,10\}, \mathrm{cell_i} \in \mathrm{Hepatocyte\ C1}\}$, and the green boxplot is the summarization of $\{\mathrm{CFBD}(\mathrm{cell_i}, \mathrm{Zone(j,k)}) \mid k \in \{1,2,3,4,5,6,7,8,9,10\}, \mathrm{cell_i} \in \mathrm{Hepatocyte\ C2}\}$. The P-value is based on Wilcoxon rank sum test.

The statistical analysis of Fig. 4i is conducted as following: the x-axis is the distance between $\mathrm{PSP(j,FBD_{Ri})}$ and corresponding fibrotic boundary $(\mathrm{FBD_{Ri}})$, $i \in \{1,2,3,4,5,6,7,8,9,10\}, j \in [0,450]$; the y-axis is the normalized count ratio between C1 and C2, which is $\frac{\mathrm{NCC(C1,AREA(j,i),AREA(j_{max},i))}}{\mathrm{NCC(C2,AREA(j,i),AREA(j_{max},i))}}$, $i \in \{1,2,3,4,5,6,7,8,9,10\}, j \in [0,450]$.

All parameters of boxplots are set as default using Seaborn (https://seaborn.pydata.org), a Python statistical data visualization toolbox.

Datasets

Simulated datasets: Four different human cell lines are cultured as a source of simulation (Supplementary Fig. 5a), and all the following datasets are manual alteration and a combination of the four cell lines.

Dataset 1: Use 4 cell lines as 4 clusters, for each cell, randomly add $noise\_ratio \times \#pixels$ number of all-zero pixels.

Dataset 2: Use 4 cell lines as 4 clusters, for each pixel, multiply it with a random number drawn from $U(0, noise\_ratio)$.

Dataset 3: Use 4 cell lines as cluster1, and the altered version of 4 cell lines as cluster2. Alteration method: for each cell, first randomly select $noise\_ratio \times \#pixel$ pixels, then replace these pixels with samples drawn from feature-independent Gaussian fitted with original data.

Dataset 4: Use A549 cell line as cluster1, and use the 3 differently altered version as the other 3 clusters. Alteration method: First, randomly select 2 dimensions, i and j. Then, for cluster2, multiply fold_change to the i-th dimension of all pixels of 10A cells, and the j-th dimension remains unchanged. For Cluster3, multiply fold_change to both i-th and j-th dimension of all pixels of 10A cells. For cluster4: multiply fold_change to the j-th dimension of all pixels of 10A cells, and the i-th dimension remains unchanged. The simulating method of dataset 4 is illustrated as (supplementary Fig. 12a).

Dataset 5: Similar to dataset 4 but using Hela cell line.

Dataset 6: Similar to dataset 4 but using SK-BR-3 cell line.

Dataset 7: Similar to dataset 4 but using MCF 10A cell line.

Dataset 8: Use 10A cell line as cluster1, and use the 3 differently altered version as the other clusters. Alteration method: First, randomly select 2 dimensions, i and j, and calculate the $mean\_i$ and $variance\_i$ for each cell. Second, for each cell, randomly divide pixels into two partitions of an equal number of pixels. Next, for cluster2, for each cell, replace the i-th dimension of the first partition with data drawn from $Gaussian(fold\_chang \times mean\_i,\ variance\_i)$, and replace the i-th dimension of the second partition with data drawn from $Gaussian((2-fold\_change) \times mean\_i,\ variance\_i)$. The j-th dimension remains unchanged. For cluster4, for each cell, replace the j-th dimension of the first partition with data drawn from $Gaussian(fold\_change \times mean\_j,\ variance\_j)$, and replace the j-th dimension of the second partition with data drawn from $Gaussian((2-fold\_change) \times mean\_j,\ variance\_j)$. The i-th dimension remains unchanged. For cluster3, the alteration for the i-th dimension is the same as cluster2, and the j-th dimension is the same with cluster4. The simulating method of dataset 8 is illustrated as (supplementary Fig. 12b).

Dataset 9: Use 2 differently altered versions of 10A cell line as two clusters. Alteration method: First, randomly select 2 dimensions, i and j, and calculate the $mean\_i$ and $variance\_i$ for each cell. Second, for each cell, randomly divide pixels into two partitions of an equal number of pixels. Next, for cluster1, for each cell, replace the i-th dimension of the first partition with data drawn from $Gaussian(fold\_chang \times mean\_i,\ variance\_i)$, and replace the i-th dimension of the second partition with data drawn from $Gaussian((2-fold\_change) \times mean\_i,\ variance\_i)$. And replace the j-th dimension of the first partition with data drawn from $Gaussian(fold\_change \times mean\_j,\ variance\_j)$, and replace the j-th dimension of the second partition with data drawn from $Gaussian((2-fold\_change) \times mean\_j,\ variance\_j)$. For cluster2, for each cell, replace the i-th dimension of the first partition with data drawn from $Gaussian(fold\_chang \times mean\_i,\ variance\_i)$, and replace the i-th dimension of the second partition with data drawn from $Gaussian((2-fold\_change) \times mean\_i,\ variance\_i)$. And replace the j-th dimension of the second partition with data drawn from $Gaussian(fold\_change \times mean\_j,\ variance\_j)$, and replace the j-th dimension of the first partition with data drawn from $Gaussian((2-fold\_change) \times mean\_j,\ variance\_j)$. The simulating method of dataset 9 is illustrated as (supplementary Fig. 12c).

Mixture cell datasets: Mixture cell culture uses BrdU/IdU as ground truth label (Supplementary Fig. 13), and the BrdU/IdU stain does not affect the cell metabolic profiling (Supplementary Fig. 21).

Dataset 10: A549 cell line stained with BrdU is mixed with Hela cell line (Supplementary Fig. 13a, b).

Dataset 11: NCTC1469 cell line stained with IdU is mixed with Hepa1-6 cell line (Supplementary Fig. 13c, d).

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 and lung (Fig. 1,2,3), and human liver R1 (Fig. 4) are available at Github (https://github.com/yuanzhiyuan/SEAM/tree/master/SEAM/data/raw_tar). The rest of raw SIMS data and processed SIMS data are available at figshare (10.6084/m9.figshare.12622883, 10.6084/m9.figshare.12622841, 10.6084/m9.figshare.12622838 and 10.6084/m9.figshare.12622922). Geo-seq (Fig. 5) raw sequencing data and processed data have been deposited to NCBI GEO with accession number GSE153463.

Code availability

An open-source Python and MATLAB implementation of SEAM is available at GitHub (https://github.com/yuanzhiyuan/SEAM, and https://github.com/yuanzhiyuan/SIMS-Cut).

References

1 Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* **19**, 1423-1437, doi:10.1038/nm.3394 (2013).

2 Riquelme, P. A., Drapeau, E. & Doetsch, F. Brain micro-ecologies: neural stem cell niches in the adult mammalian brain. *Philos T R Soc B* **363**, 123-137, doi:10.1098/rstb.2006.2016 (2008).

3 Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *P Natl Acad Sci USA* **99**, 12795-12800, doi:10.1073/pnas.162041399 (2002).

4 Zhang, J. W. & Li, L. H. Stem cell niche: Microenvironment and beyond. *J Biol Chem* **283**, 9499-9503, doi:10.1074/jbc.R700043200 (2008).

5 Shukla, S. D. & Lim, R. W. Epigenetic effects of ethanol on the liver and gastrointestinal system. *Alcohol research: current reviews* **35**, 47 (2013).

6 Benly, P. Role of histamine in acute inflammation. *Journal of Pharmaceutical Sciences and Research* **7**, 373 %@ 0975-1459 (2015).

7 Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat Rev Genet* **20**, 257-272, doi:10.1038/s41576-019-0093-7 (2019).

8 Pareek, V., Tian, H., Winograd, N. & Benkovic, S. J. Metabolomics and mass spectrometry imaging reveal channeled de novo purine synthesis in cells. *Science* **368**, 283-290 %@ 0036-8075 (2020).

9 Kennedy, D. E. *et al.* Novel specialized cell state and spatial compartments within the germinal center. *Nature Immunology*, 1-11 %@ 1529-2916 (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 %@ 1546-1170X (2001).

11 Sun, C. *et al.* Spatially resolved metabolomics to discover tumor-associated metabolic alterations. *Proc Natl Acad Sci U S A* **116**, 52-57, doi:10.1073/pnas.1808950116 (2019).

12 Hare, D. J. *et al.* Three-dimensional atlas of iron, copper, and zinc in the mouse cerebrum and brainstem. *Anal Chem* **84**, 3990-3997 %@ 0003-2700 (2012).

13 Sjövall, P., Lausmaa, J. & Johansson, B. Mass spectrometric imaging of lipids in brain tissue. *Anal Chem* **76**, 4271-4278 %@ 0003-2700 (2004).

14 Zavalin, A., Yang, J. & Caprioli, R. Laser beam filtration for high spatial resolution MALDI imaging mass spectrometry. *J Am Soc Mass Spectr* **24**, 1153-1156 %@ 1044-0305 (2013).

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, doi:10.1038/s41592-019-0536-2 (2019).

16 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-+, doi:10.1038/Nmeth.4504 (2017).

17 Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods* **16**, 987-990, doi:10.1038/s41592-019-0548-y (2019).

18 Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-+,

doi:10.1016/j.cell.2018.08.039 (2018).
19 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 U S A* **116**, 19490-19499, doi:10.1073/pnas.1912459116 (2019).
20 Keren, L. *et al.* MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci Adv* **5**, doi:10.1126/sciadv.aax5851 (2019).
21 Robertson, A. R. The CIE 1976 color-difference formulae. *Color Research & Application* **2**, 7-11 %@ 0361-2317 (1977).
22 Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352-356, doi:10.1038/nature21065 (2017).
23 Park, J. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758-763, doi:10.1126/science.aar2131 (2018).
24 Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-+, doi:10.1016/j.cell.2018.08.063 (2018).
25 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38-+, doi:10.1038/nbt.4314 (2019).
26 Wang, Y. J. *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, doi:10.1016/j.cmet.2019.01.003 (2019).
27 Koller, D. & Friedman, N. *Probabilistic graphical models: principles and techniques*. (MIT press, 2009).
28 Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504-507, doi:10.1126/science.1127647 (2006).
29 Hinton, G. A Practical Guide to Training Restricted Boltzmann Machines. (2010).
30 Le Roux, N. & Bengio, Y. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput* **20**, 1631-1649, doi:10.1162/neco.2008.04-07-510 (2008).
31 Bengio, Y., Courville, A. & Vincent, P. Representation Learning: A Review and New Perspectives. *Ieee T Pattern Anal* **35**, 1798-1828, doi:10.1109/Tpami.2013.50 (2013).
32 Boykov, Y., Veksler, O. & Zabih, R. Markov random fields with efficient approximations. *Proc Cvpr Ieee*, 648-655, doi:Doi 10.1109/Cvpr.1998.698673 (1998).
33 Boykov, Y., Veksler, O. & Zabih, R. Fast approximate energy minimization via graph cuts. *Ieee T Pattern Anal* **23**, 1222-1239, doi:10.1109/34.969114 (2001).
34 Doersch, C., Gupta, A. & Efros, A. A. Unsupervised Visual Representation Learning by Context Prediction. *2015 Ieee International Conference on Computer Vision (Iccv)*, 1422-1430, doi:10.1109/Iccv.2015.167 (2015).
35 Noroozi, M. & Favaro, P. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. *Computer Vision - Eccv 2016, Pt Vi* **9910**, 69-84, doi:10.1007/978-3-319-46466-4_5 (2016).
36 Noroozi, M., Pirsiavash, H. & Favaro, P. Representation Learning by Learning to Count. *2017 Ieee International Conference on Computer Vision (Iccv)*, 5899-5907, doi:10.1109/Iccv.2017.628 (2017).
37 Hartigan, J. A. & Wong, M. A. Algorithm AS 136: A k-means clustering algorithm. *Journal*

of the Royal Statistical Society. Series C (Applied Statistics) **28**, 100-108 %@ 0035-9254 (1979).

38 Kiselev, V. Y. *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* **14**, 483-486, doi:10.1038/nmeth.4236 (2017).

39 Wang, B., Zhu, J. J., Pierson, E., Ramazzotti, D. & Batzoglou, S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat. Methods* **14**, 414-+, doi:10.1038/nmeth.4207 (2017).

40 Maaten, L. v. d. & Hinton, G. Visualizing data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).

41 McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* **2**, 205 %@ 2475-9066 (2017).

42 Brison, J. *et al.* ToF-SIMS imaging and depth profiling of HeLa cells treated with bromodeoxyuridine. *Surface and Interface Analysis* **43**, 354-357, doi:10.1002/sia.3415 (2011).

43 Ramdas, A., Trillos, N. G. & Cuturi, M. On Wasserstein Two-Sample Testing and Related Families of Nonparametric Tests. *Entropy-Switz* **19**, doi:ARTN 47

10.3390/e19020047 (2017).

44 Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. *Nat Rev Gastroenterol Hepatol*, doi:10.1038/s41575-019-0134-x (2019).

45 Sano, K. *et al.* Distributional Variation of P-450 Immunoreactive Hepatocytes in Human-Liver Disorders. *Hum Pathol* **20**, 1015-1020, doi:Doi 10.1016/0046-8177(89)90274-8 (1989).

46 Brosch, M. *et al.* Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control. *Nat Commun* **9**, 4150, doi:10.1038/s41467-018-06611-5 (2018).

47 Shetty, S., Lalor, P. F. & Adams, D. H. Liver sinusoidal endothelial cells - gatekeepers of hepatic immunity. *Nat Rev Gastro Hepat* **15**, 555-567, doi:10.1038/s41575-018-0020-y (2018).

48 Ramachandran, P. *et al.* Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature*, doi:10.1038/s41586-019-1631-3 (2019).

49 Wu, X., Jiang, R., Zhang, M. Q. & Li, S. Network-based global inference of human disease genes. *Mol Syst Biol* **4**, 189, doi:10.1038/msb.2008.27 (2008).

50 Pillai, S. M. & Meredith, D. SLC36A4 (hPAT4) is a high affinity amino acid transporter when expressed in Xenopus laevis oocytes. *J Biol Chem* **286**, 2455-2460 %@ 0021-9258 (2011).

51 Féral, C. C. *et al.* CD98hc (SLC3A2) participates in fibronectin matrix assembly by mediating integrin signaling. *The Journal of cell biology* **178**, 701-711 %@ 1540-8140 (2007).

52 Wang, S. *et al.* Lysosomal amino acid transporter SLC38A9 signals arginine sufficiency to mTORC1. *Science* **347**, 188-194 %@ 0036-8075 (2015).

53 Reimer, R. J. SLC17: a functionally diverse family of organic anion transporters. *Mol Aspects Med* **34**, 350-359 %@ 0098-2997 (2013).

54 Kempson, S. A., Zhou, Y. & Danbolt, N. C. The betaine/GABA transporter and betaine: roles in brain, kidney, and liver. *Frontiers in physiology* **5**, 159 %@ 1664-1042X (2014).

55 Chen, J. *et al.* Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-

seq. *Nature Protocols* **12**, 566-580, doi:10.1038/nprot.2017.003 (2017).
56 Krueger, F. Trim galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. **516**, 517 (2015).
57 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766-D773, doi:10.1093/nar/gky955 (2019).
58 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
59 Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
60 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, doi:ARTN 550
10.1186/s13059-014-0550-8 (2014).
61 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 T Med Imaging* **20**, 45-57, doi:Doi 10.1109/42.906424 (2001).
62 Panjwani, D. K. & Healey, G. Markov Random-Field Models for Unsupervised Segmentation of Textured Color Images. *Ieee T Pattern Anal* **17**, 939-954, doi:Doi 10.1109/34.464559 (1995).
63 Hammersley, J. M. & Clifford, P. (1971).
64 Clifford, P. Markov random fields in statistics. *Disorder in physical systems: A volume in honour of John M. Hammersley* **19** (1990).
65 Besag, J. Spatial Interaction and Statistical-Analysis of Lattice Systems. *J Roy Stat Soc B Met* **36**, 192-236 (1974).
66 Panjwani, D. K. & Healey, G. Markov Random-Field Model for Unsupervised Segmentation of Textured Color Images (Vol 17, Pg 939, 1995). *Ieee T Pattern Anal* **17**, 1128-1128 (1995).
67 Kato, Z. & Pong, T. C. A Markov random field image segmentation model for color textured images. *Image Vision Comput* **24**, 1103-1114, doi:10.1016/j.imavis.2006.03.005 (2006).
68 Chen, F. Q., Wu, Y., Bu, Y. D. & Zhao, G. D. Spectral Classification Using Restricted Boltzmann Machine. *Publications of the Astronomical Society of Australia* **31**, doi:10.1017/pasa.2013.38 (2014).
69 Boykov, Y. & Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Ieee T Pattern Anal* **26**, 1124-1137, doi:Doi 10.1109/Tpami.2004.60 (2004).
70 Boykov, Y. & Funka-Lea, G. Graph Cuts and Efficient N-D Image Segmentation. *Int J Comput Vision* **70**, 109-131, doi:10.1007/s11263-006-7934-5 (2006).
71 Hinton, G. E. Training products of experts by minimizing contrastive divergence. *Neural Computation* **14**, 1771-1800, doi:Doi 10.1162/089976602760128018 (2002).
72 Tieleman, T. 1064-1071.
73 Keyvanrad, M. A. & Homayounpour, M. M. A brief survey on deep belief networks and introducing a new object oriented toolbox (DeeBNet). *arXiv preprint arXiv:1408.3264* (2014).
74 Larochelle, H. & Bengio, Y. 536-543.

75 Hinton, G., Vinyals, O. & Dean, J. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* (2015).
76 Veselkov, K. A. *et al.* Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery. *Anal Chem* **83**, 5864-5872, doi:10.1021/ac201065j (2011).
77 Lu, Y. Unsupervised learning on neural network outputs: with application in zero-shot learning. *arXiv preprint arXiv:1506.00990* (2015).
78 Wu, Z., Xiong, Y., Yu, S. X. & Lin, D. 3733-3742.
79 Duin, R. P. W. & Pękalska, E. The dissimilarity space: Bridging structural and statistical pattern recognition. *Pattern Recogn Lett* **33**, 826-832 %@ 0167-8655 (2012).
80 Tax, D. M. J., Loog, M., Duin, R. P. W., Cheplygina, V. & Lee, W.-J. 222-234 (Springer).
81 Glorot, X., Bordes, A. & Bengio, Y. 315-323.
82 Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
83 Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol*, doi:10.1038/s41587-019-0392-8 (2020).
84 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).

Acknowledgements

The authors would like to acknowledge Imaging Core Facility, Technology Center for Protein Sciences, Tsinghua University for assistance of using LMD7000. We also thank Yalan Chen from Imaging Core Facility for her detailed instruction on LMD7000. We thank Center of Laboratory Animal Resources, Tsinghua University for mice maintenance and providing CM1900 Cryostat. We thank Hongjun Li for the computing resource supporting. We thank Hui Zhang for the help of ethics material preparation. We thank Minglei Shi, Yisi Li, Zhaofeng Ye, Rui Qi and all other members of our lab for valuable comments and discussions. We thank Minping Qian for helpful advice on algorithm development. This work was partly supported by National Basic Research Program of China (2018YFA0801402, 2018YFB0704304, 2017YFA0505503), National Nature Science Foundation of China (31871343, 21974078, 21727813, 21621003) and foundation of BNRist (BNR2019TD01020).

Author contributions

Y.C, M.Q.Z and X.Z conceived and designed the project. L.C designed the IMS experiment and generated the IMS data. Q.Z processed the mouse and human sample assisted by W.S, and generated IHC and HE imaging data. Q.Z and L.C designed and conducted the cell culture and BrdU staining experiment. Q.Z

designed and conducted the modified Geo-seq experiment. Z.Y developed and implemented the algorithms under the guidance of M.Q.Z and Y.C, and assisted by Q.Z. Z.Y analyzed the SIMS data, and Q.Z analyzed the spatial transcriptome data. Y.Z and S.Y provided the clinical samples. L.P and S.Q guided the histological annotation. S.L gave suggestions on the application of the method. Z.Y, Q.Z, and L.C completed the figures and manuscript with the guidance of Y.C. X.Z and M.Q.Z.

Competing interests

The authors declare no competing interests.

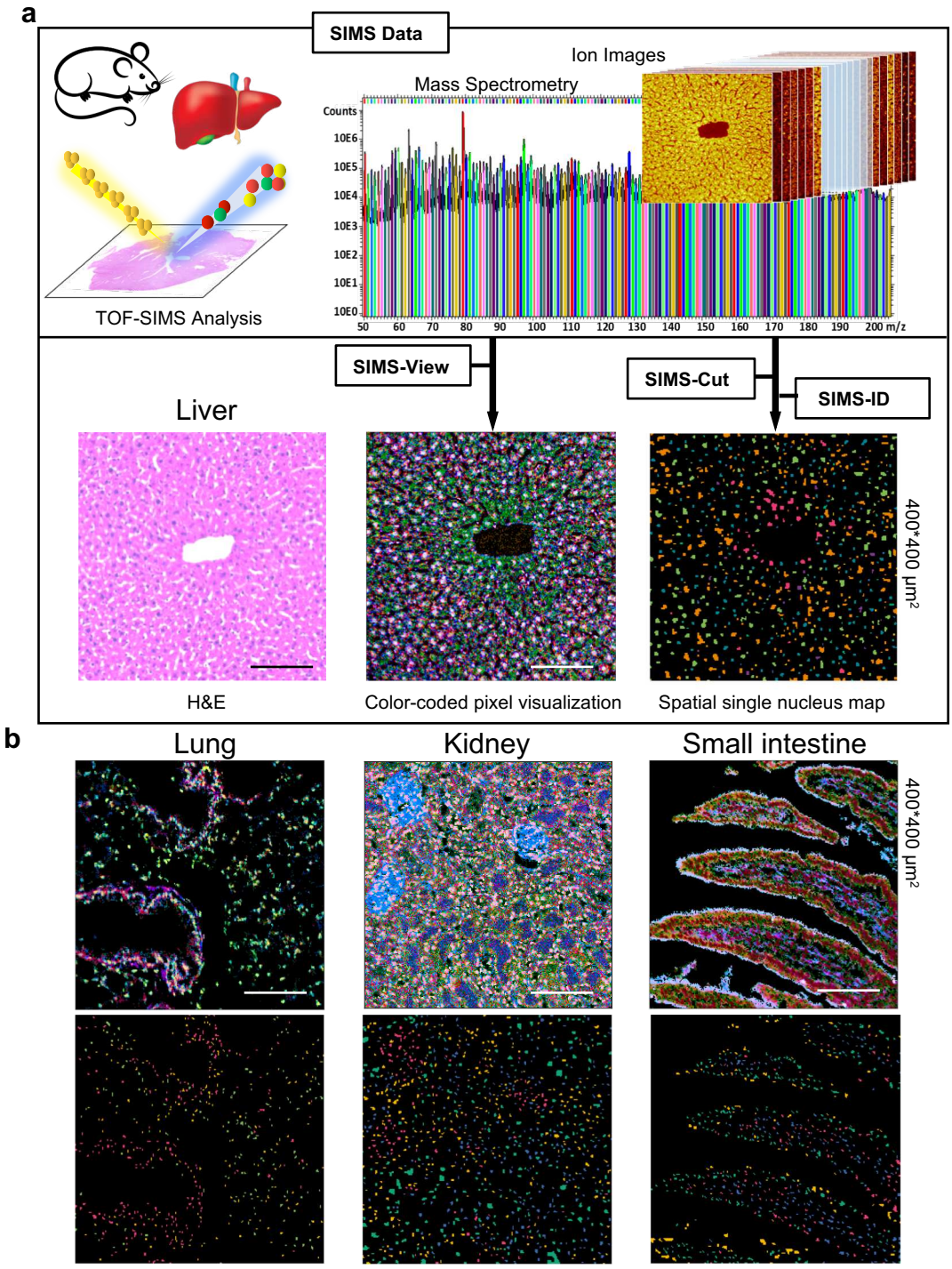

Fig. 1 | SEAM captures spatial metabolic heterogeneity at single nucleus resolution. **a,** Overview of SEAM. (Left) Tissue samples on glass slides are analyzed by TOF-SIMS to generate multiplex SIMS data containing mass spectrometry and ion images (Right). (Bottom left) H&E staining of mouse liver central vein region. (Bottom middle) Color-coded pixel visualization is obtained by SIMS-View. (Bottom right) Spatial single nucleus map is obtained by a sequential of algorithms: SIMS-Cut (segmentation), SIMS-ID (representation), and SIMS-Cluster (clustering). **b,** SEAM scales to different mouse tissues with different cell density and distribution pattern. First row is color-coded pixel visualization by SIMS-view to differentiate metabolic patterns at pixel level. Second row is spatial single nucleus map for cell type visualization at original tissue space. Scale bar 100μm. In Fig. 1a, Mouse illustration: Image by OpenClipart-Vectors from Pixabay. Liver illustration: Image by zachvanstone8 from Pixabay.

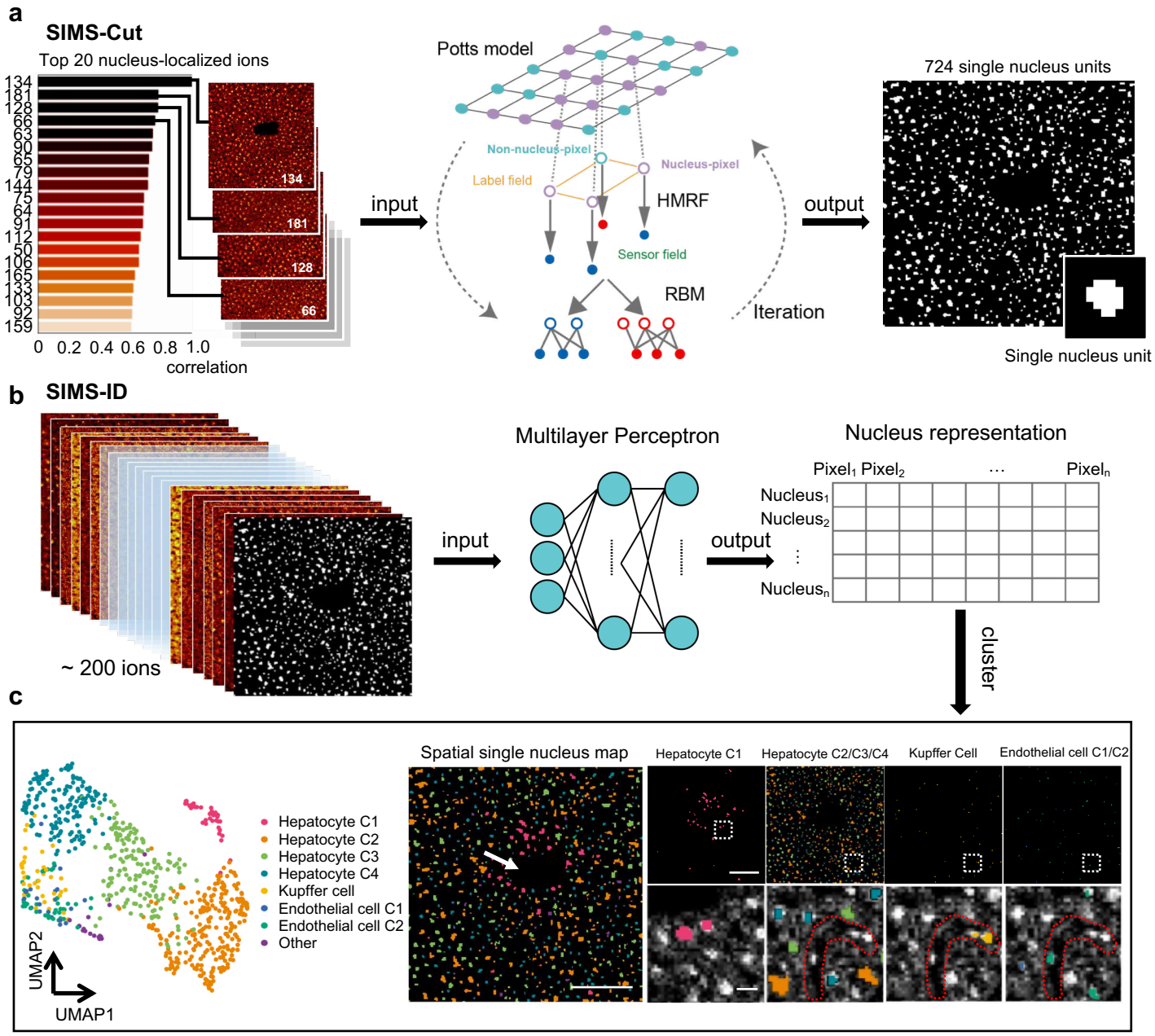

Fig. 2 | Algorithms design and performance. a, Sketch of SIMS-Cut, leveraging Potts model as prior for pixel labels and Restricted Boltzmann Machines as conditional distribution of pixel intensities. (Left) Top 20 nucleus-localized ions. (Middle) Iterative optimization between subproblems (See Methods). (Right) Cell segmentation mask. **b,** Sketch of SIMS-ID, learning vector-formed representation for each segmented cells using self-representation learning. (Left) multiplex SIMS data combined with cell segmentation mask. (Middle) A neural network for a auxiliary classification task. (Right) Single nucleus representation output. **c,** Demonstration of algorithms on central vein (CV) of wild type mouse liver. (Left) UMAP visualization of single nucleus using SIMS-ID representation, colored by SIMS-Cluster identified cell types. (Middle) Spatial single nucleus map. White arrow indicates CV. Scale bar 100μm. (Right top) Respective layout of cell populations. Scale bar 100μm. (Right bottom) Zoom in images of each population merged with grey scaled image of m/z 134.Red dotted area indicate liver sinusoid. Scale bar 10μm.

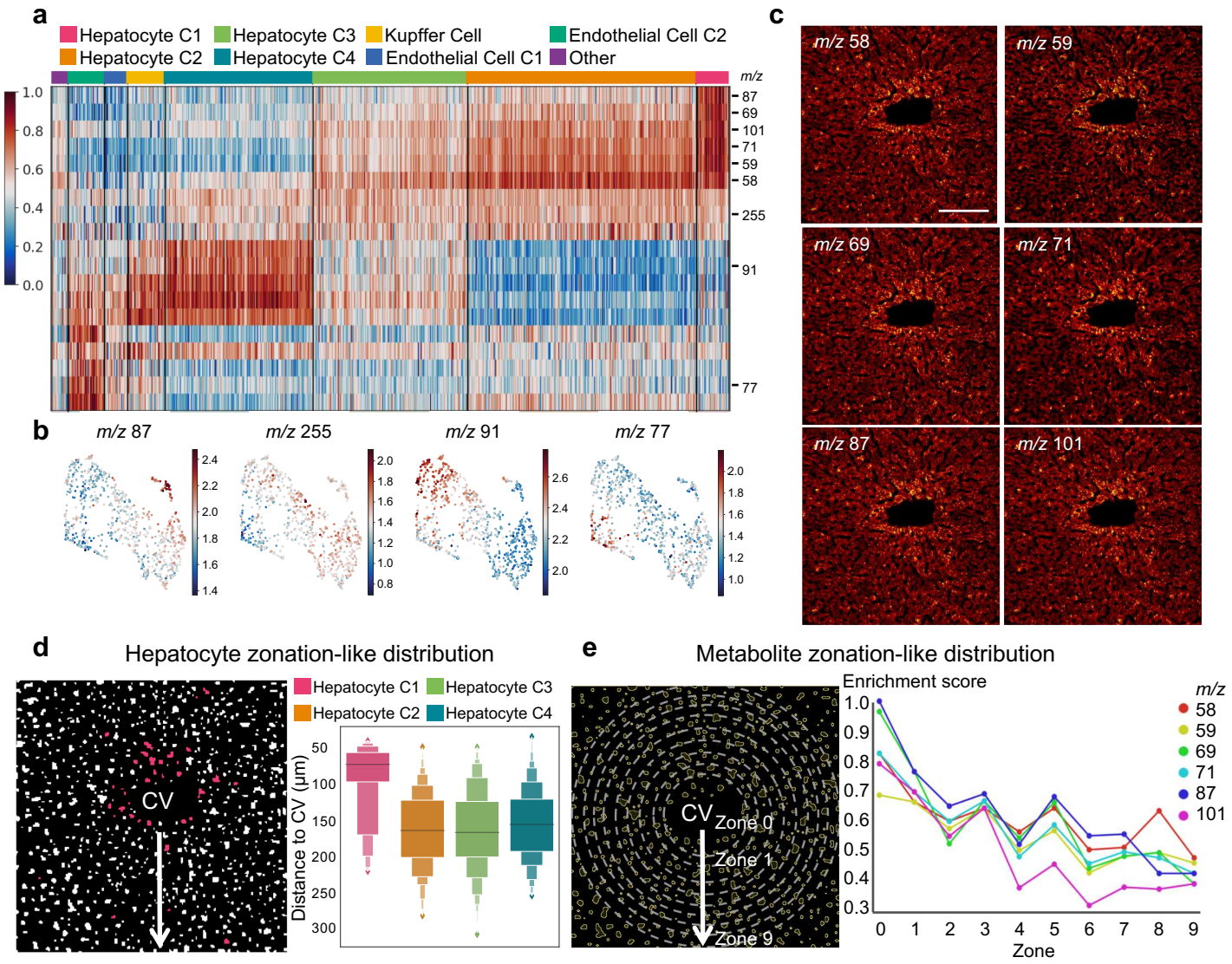

Fig. 3 | SEAM detects zonation-like metabolic pattern in wild type mouse liver. a, Differential metabolite analysis of mouse liver tissue in Fig. 2c. **b,** UMAP colored by abundance of representative differential metabolites. **c,** Ion images of a ion series with zonation-like distribution identified by differential analysis in **Fig. 3a**. Scale bar 100μm. **d,** Hepatocyte C1 subpopulation shows zonation-like distribution. (Left) Schematic diagram of strategy of measuring cell-to-CV distance. (Right) Hepatocyte C1 shows significantly smaller distance to CV than other clusters. **e,** Metabolite series show zonation-like distribution. (Left) Schematic diagram of strategy of measuring metabolite-to-CV distance: Concentric circles with distance of arithmetic sequence from CV partition the liver lobule into 9 zones. (Right) 6 metabolic markers of Hepatocyte C1 show gradient decrease away from CV. X-axis: zone number, Y-axis: enrichment score of each metabolites, which is the proportion of hepatocytes that highly express each metabolites in each zones.

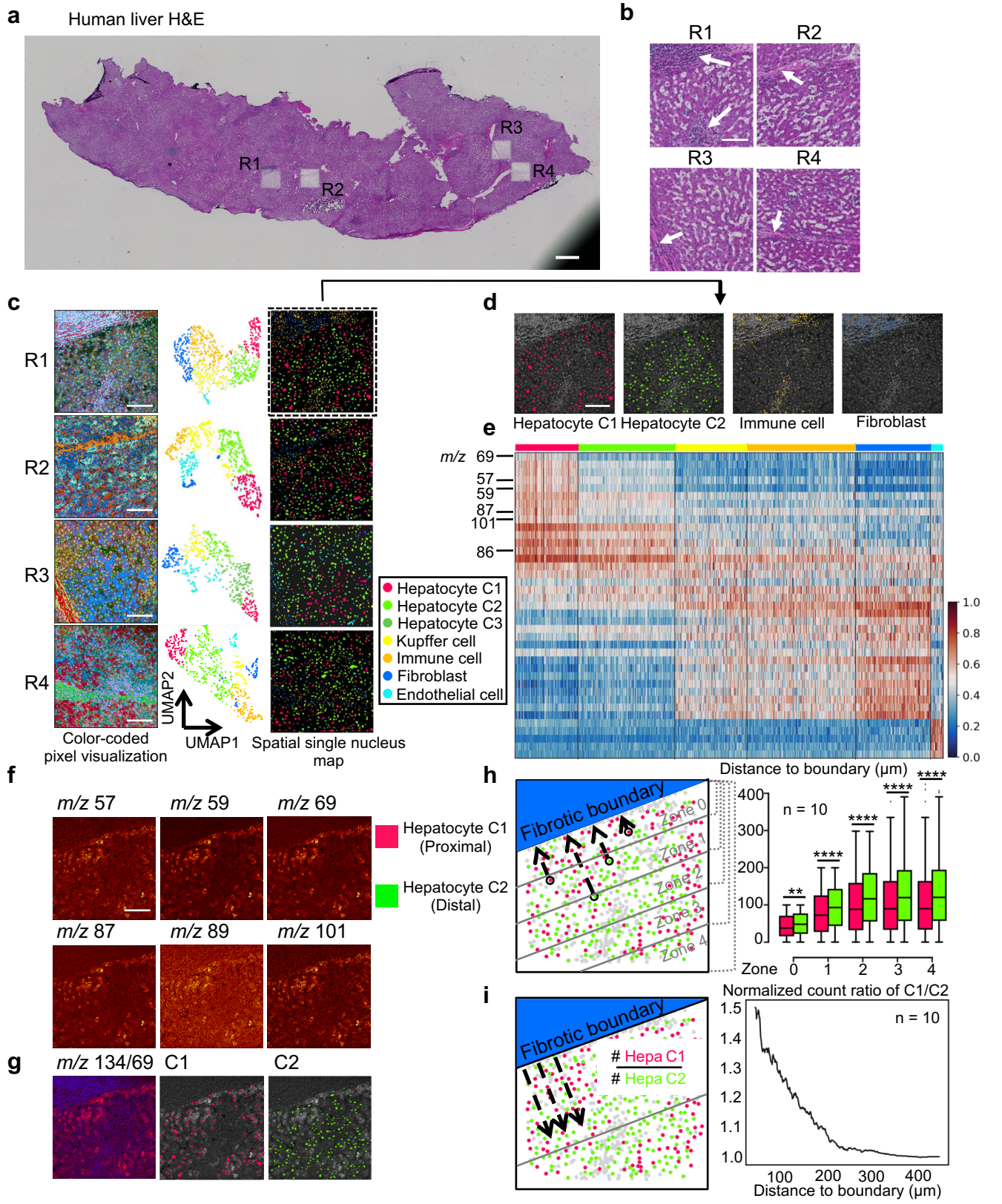
a
Human liver H&E
R1
R2
R3
R4
b
R1
R2
R3
R4
c
R1
R2
R3
R4
UMAP2
UMAP1
Color-coded pixel visualization
Spatial single nucleus map
Hepatocyte C1
Hepatocyte C2
Hepatocyte C3
Kupffer cell
Immune cell
Fibroblast
Endothelial cell
d
Hepatocyte C1
Hepatocyte C2
Immune cell
Fibroblast
e
m/z
69
57
59
87
101
86
1.0
0.8
0.6
0.4
0.2
0.0
Distance to boundary (μm)
f
m/z 57
m/z 59
m/z 69
m/z 87
m/z 89
m/z 101
Hepatocyte C1 (Proximal)
Hepatocyte C2 (Distal)
g
m/z 134/69
C1
C2
h
Fibrotic boundary
Zone 0
Zone 1
Zone 2
Zone 3
Zone 4
n = 10
400
300
200
100
0
**
****
****
****
****
Zone
0
1
2
3
4
i
Fibrotic boundary
Hepa C1
Hepa C2
Normalized count ratio of C1/C2
n = 10
1.5
1.4
1.3
1.2
1.1
1.0
100
200
300
400
Distance to boundary (μm)

Fig. 4 | SEAM identifies hepatocyte subtypes with differential metabolic state associated with spatial localization. a, H&E staining of human liver sample post TOF-SIMS analysis. Scale bar 500μm. **b,** Zoom in H&E images of 4 different regions. Bottom. White arrows indicate fibrotic and inflammation niche. Scale bar 100μm. **c,** SEAM results of 4 regions. First column is color-coded pixel visualizations. Scale bar 100μm. Second column is UMAP colored by cell clusters. Third column is spatial single nucleus map. **d,** Spatial single nucleus maps of respective clusters merged with grey scaled ion image of *m/z* 134. Scale bar 100μm. **e,** Differential metabolite analysis of cell clusters. **f, (**Top and middle row) Hepatocyte C1 enriched metabolites. Scale bar 100μm. (Bottom row left) Merged ion image of *m/z* 69(Red) and *m/z* 134 (Blue). (Bottom row middle and right) Spatial localization of hepatocyte C1 and C2 respectively merged with grey-scaled ion image of *m/z* 69. **h,** Hepatocyte C1 is consistently closer to fibrotic boundary (FBD) than C2 within all 5 zones. (Left) Schematic diagram of zone definition and distance calculation. (Right) Paired boxplots of distances between C1/C2 and FBD. For Wilcoxon Rank Sum test, P-value > 0.05 is not shown on the plot. P-value ≤ 0.05 (*), P-value ≤ 0.01 (**), P-value ≤ 0.001 (***) and P-value ≤ 0.0001 (****) are shown. **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 of the outer edge (indicated by the gray line in the left part of Fig. 4i) to the FBD.

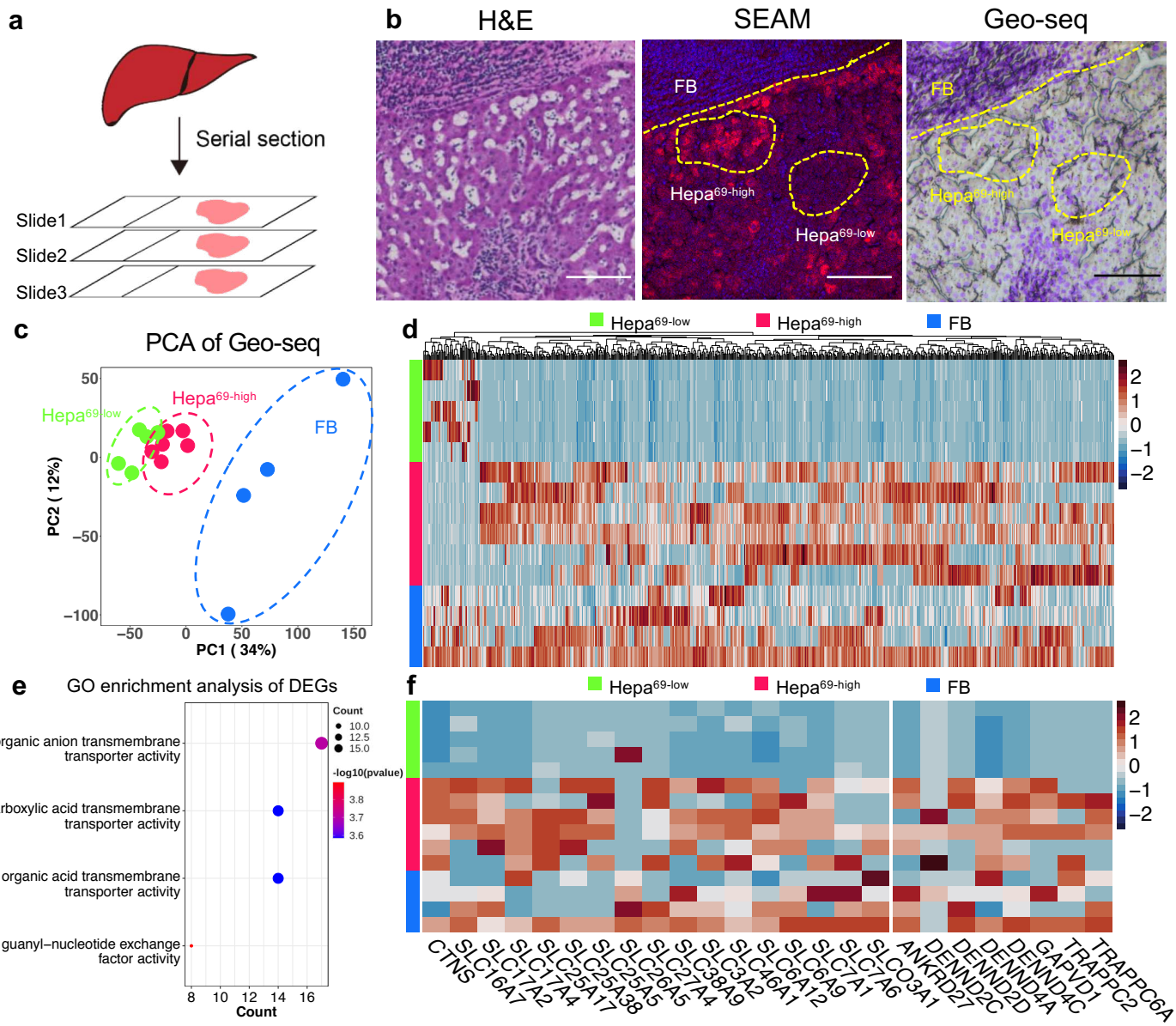

Fig. 5 | Spatial transcriptome validated metabolism associated gene expression alteration in heterogeneous hepatocyte subtypes identified by SEAM. a, Serial sections were made for cross validation among different assays. **b,** Geo-seq was performed at same location (Right) in the adjacent slide of SEAM assay (Middle, *m/z* 134 in blue and *m/z* 69 in red) to obtain continuous tissue spatial structure. Yellow dashed area representatively indicate the captured regions for Geo-seq. Scale bar 100μm. **c,** PCA plot of transcriptomic profiles from a total 15 samples of different regions. **d,** Heatmap of filtered differentially expressed genes (DEGs) between Hepa^{69high} and Hepa^{69low} cells. **e,** GO enrichment of DEGs. **f,** Heatmap of DEGs enriched in GO terms in **e**. Upper part is consensus 14 genes in top 3 GO terms, and lower part is 8 genes enriched in last GO term.

Figures

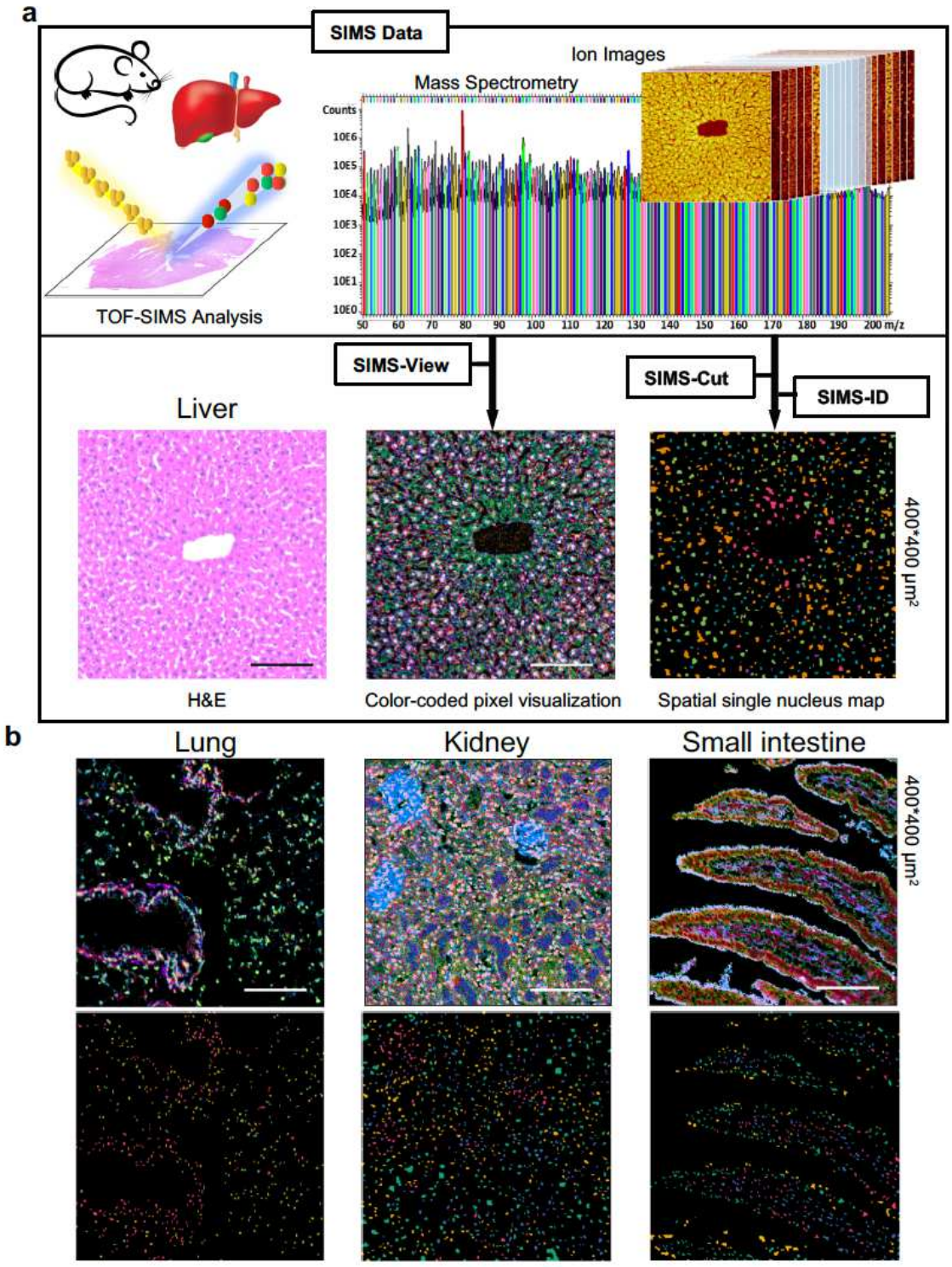

Figure 1

SEAM captures spatial metabolic heterogeneity at single nucleus resolution. a, Overview of SEAM. (Left) Tissue samples on glass slides are analyzed by TOF-SIMS to generate multiplex SIMS data containing mass spectrometry and ion images (Right). (Bottom left) H&E staining of mouse liver central vein region.

(Bottom middle) Color-coded pixel visualization is obtained by SIMS-View. (Bottom right) Spatial single nucleus map is obtained by a sequential of algorithms: SIMS-Cut (segmentation), SIMS-ID (representation), and SIMS-Cluster (clustering). b, SEAM scales to different mouse tissues with different cell density and distribution pattern. First row is color-coded pixel visualization by SIMS-view to differentiate metabolic patterns at pixel level. Second row is spatial single nucleus map for cell type visualization at original tissue space. Scale bar 100μm. In Fig. 1a, Mouse illustration: Image by OpenClipart-Vectors from Pixabay. Liver illustration: Image by zachvanstone8 from Pixabay.

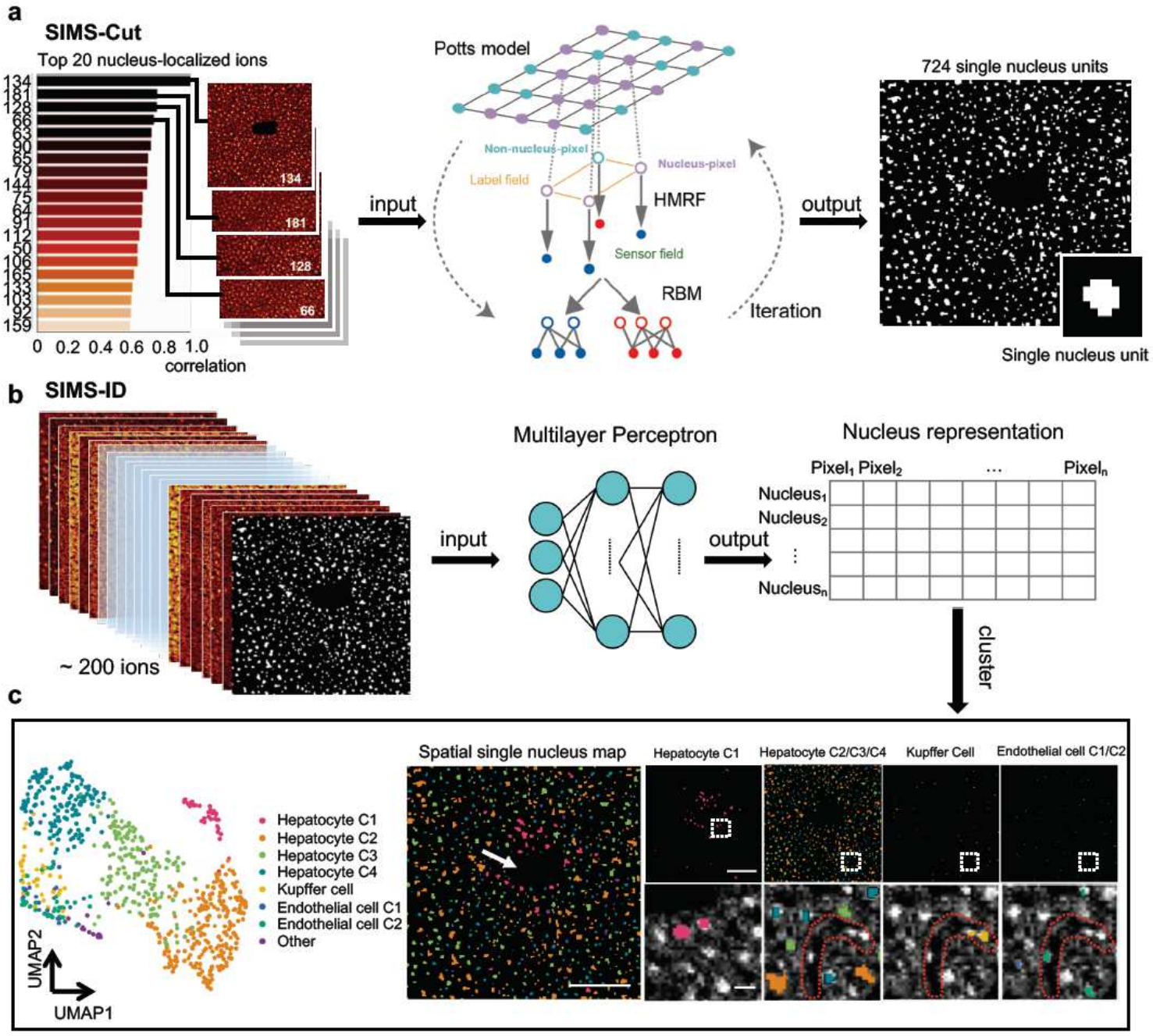

Figure 2

Algorithms design and performance. a, Sketch of SIMS-Cut, leveraging Potts model as prior for pixel labels and Restricted Boltzmann Machines as conditional distribution of pixel intensities. (Left) Top 20 nucleus-localized ions. (Middle) Iterative optimization between subproblems (See Methods). (Right) Cell segmentation mask. b, Sketch of SIMS-ID, learning vector-formed representation for each segmented cells using self-representation learning. (Left) multiplex SIMS data combined with cell segmentation mask. (Middle) A neural network for a auxiliary classification task. (Right) Single nucleus representation output. c, Demonstration of algorithms on central vein (CV) of wild type mouse liver. (Left) UMAP visualization of single nucleus using SIMS-ID representation, colored by SIMS-Cluster identified cell types. (Middle) Spatial single nucleus map. White arrow indicates CV. Scale bar 100μm. (Right top) Respective layout of cell populations. Scale bar 100μm. (Right bottom) Zoom in images of each population merged with grey scaled image of m/z 134.Red dotted area indicate liver sinusoid. Scale bar 10μm.

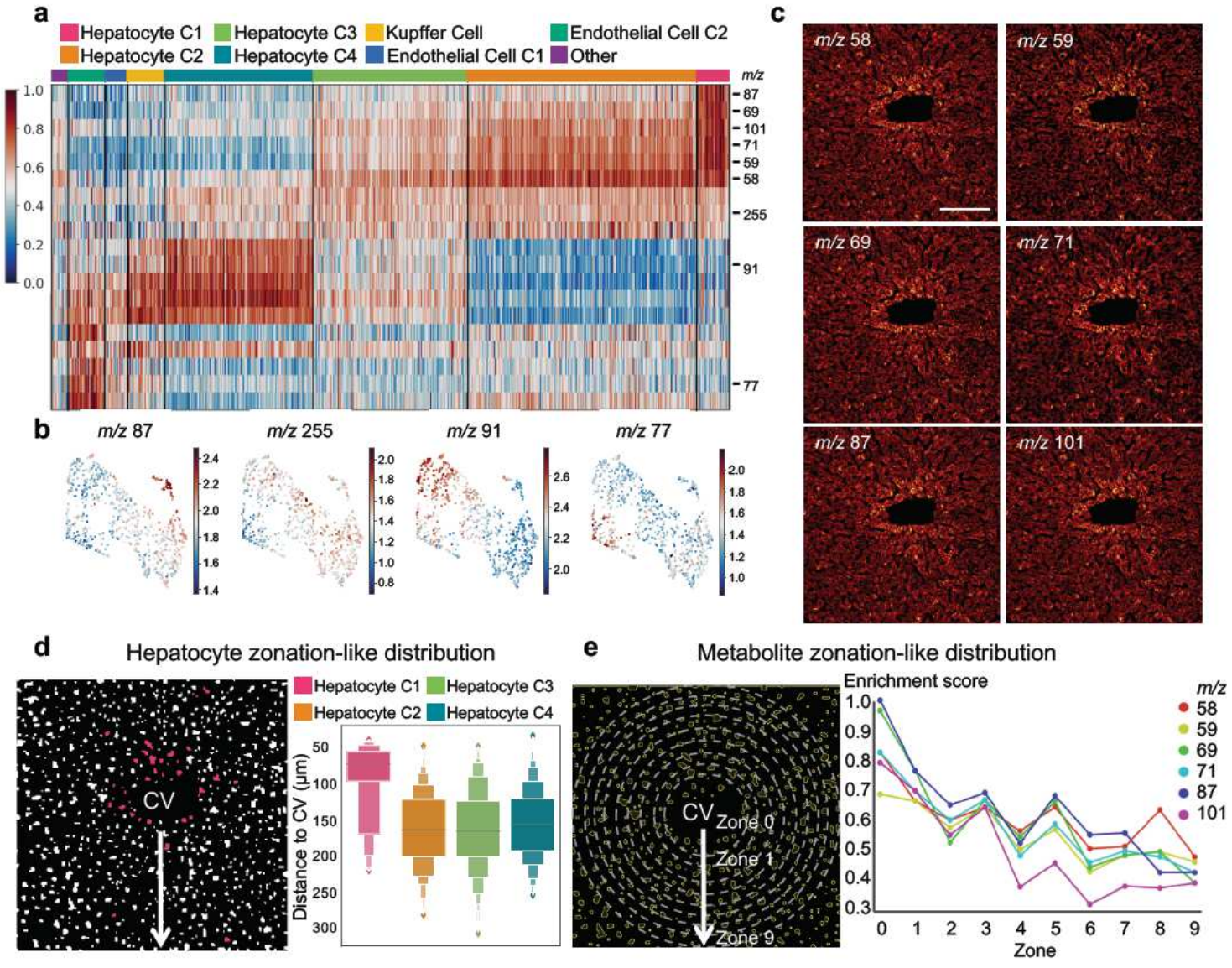

Figure 3

SEAM detects zonation-like metabolic pattern in wild type mouse liver. a, Differential metabolite analysis of mouse liver tissue in Fig. 2c. b, UMAP colored by abundance of representative differential metabolites. c, Ion images of a ion series with zonation-like distribution identified by differential analysis in Fig. 3a. Scale bar 100μm. d, Hepatocyte C1 subpopulation shows zonation-like distribution. (Left) Schematic diagram of strategy of measuring cell-to-CV distance. (Right) Hepatocyte C1 shows significantly smaller distance to CV than other clusters. e, Metabolite series show zonation-like distribution. (Left) Schematic diagram of strategy of measuring metabolite-to-CV distance: Concentric circles with distance of arithmetic sequence from CV partition the liver lobule into 9 zones. (Right) 6 metabolic markers of Hepatocyte C1 show gradient decrease away from CV. X-axis: zone number, Y-axis: enrichment score of each metabolites, which is the proportion of hepatocytes that highly express each metabolites in each zones.

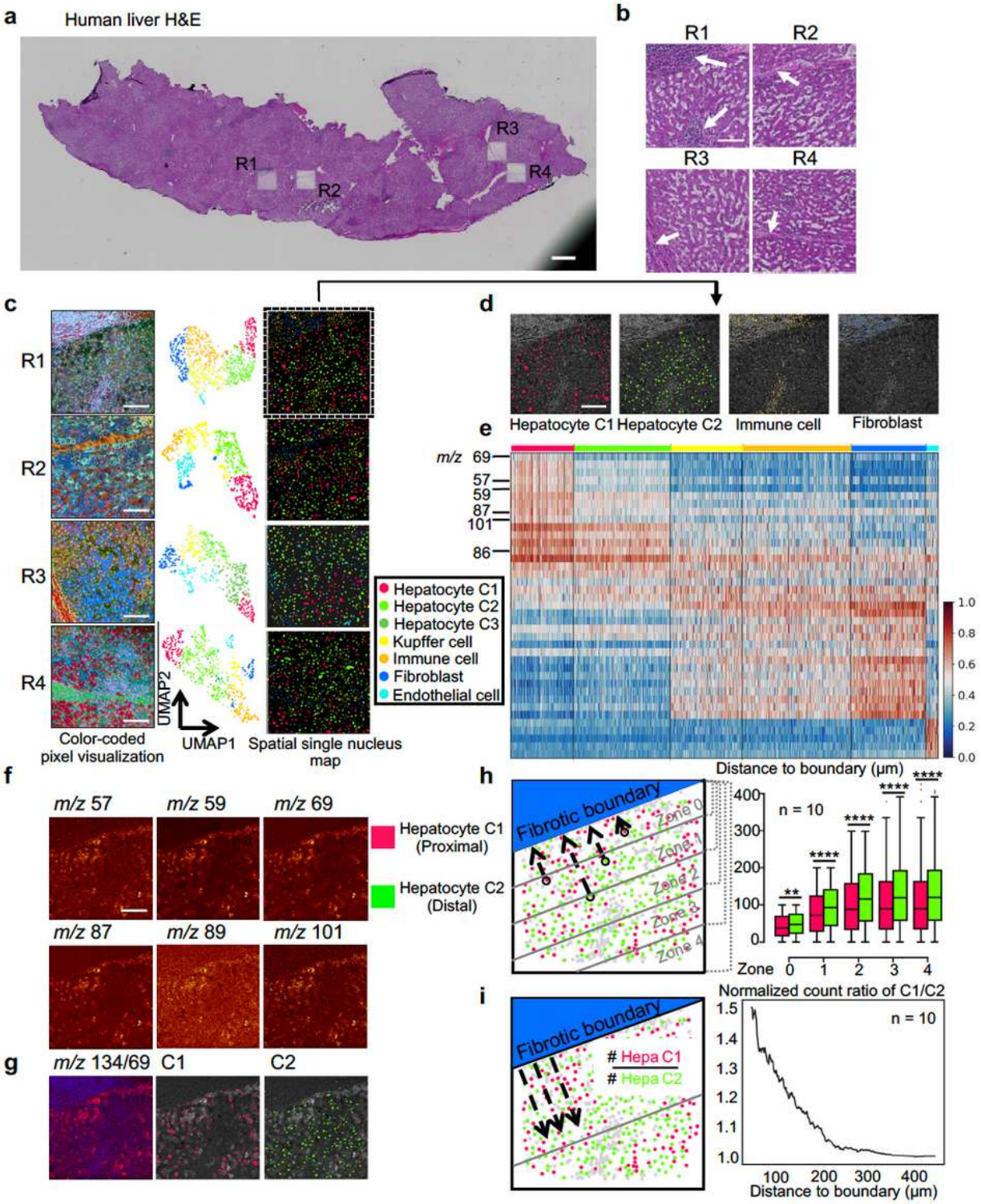

Figure 4

SEAM identifies hepatocyte subtypes with differential metabolic state associated with spatial localization. a, H&E staining of human liver sample post TOF-SIMS analysis. Scale bar 500μm. b, Zoom in H&E images of 4 different regions. Bottom. White arrows indicate fibrotic and inflammation niche. Scale bar 100μm. c, SEAM results of 4 regions. First column is colorcoded pixel visualizations. Scale bar 100μm. Second column is UMAP colored by cell clusters. Third column is spatial single nucleus map. d,

Spatial single nucleus maps of respective clusters merged with grey scaled ion image of m/z 134. Scale bar 100μm. e, Differential metabolite analysis of cell clusters. f, (Top and middle row) Hepatocyte C1 enriched metabolites. Scale bar 100μm. (Bottom row left) Merged ion image of m/z 69(Red) and m/z 134 (Blue). (Bottom row middle and right) Spatial localization of hepatocyte C1 and C2 respectively merged with greyscaled ion image of m/z 69. h, Hepatocyte C1 is consistently closer to fibrotic boundary (FBD) than C2 within all 5 zones. (Left) Schematic diagram of zone definition and distance calculation. (Right) Paired boxplots of distances between C1/C2 and FBD. For Wilcoxon Rank Sum test, P-value > 0.05 is not shown on the plot. P-value ≤ 0.05 (*), P-value ≤ 0.01 (**), P-value ≤ 0.001 (***) and Pvalue ≤ 0.0001 (****) are shown. 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 of the outer edge (indicated by the gray line in the left part of Fig. 4i) to the FBD.

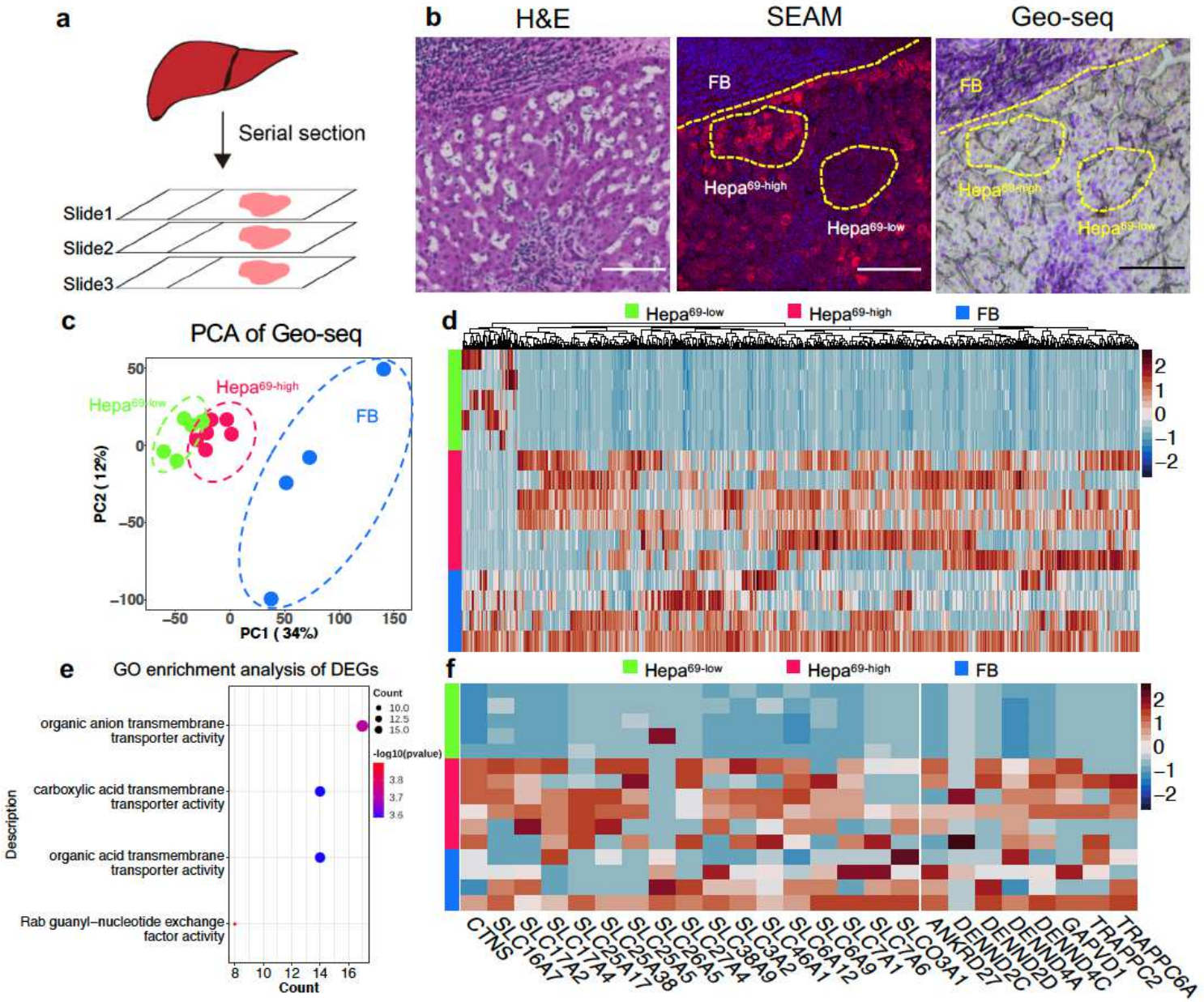

Figure 5

Spatial transcriptome validated metabolism associated gene expression alteration in heterogeneous hepatocyte subtypes identified by SEAM. a, Serial sections were made for cross validation among different assays. b, Geo-seq was performed at same location (Right) in the adjacent slide of SEAM assay (Middle, m/z 134 in blue and m/z 69 in red) to obtain continuous tissue spatial structure. Yellow dashed area representatively indicate the captured regions for Geoseq. Scale bar 100μm. c, PCA plot of transcriptomic profiles from a total 15 samples of different regions. d, Heatmap of filtered differentially expressed genes (DEGs) between Hepa69high and Hepa69low cells. e, GO enrichment of DEGs. f, Heatmap of DEGs enriched in GO terms in e. Upper part is consensus 14 genes in top 3 GO terms, and lower part is 8 genes enriched in last GO term.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarySEAM.pdf
- SupplementaryTable2SEAM.xlsx
- nrreportingsummarySEAM.pdf
- nreditorialpolicychecklistSEAM.pdf