Spatial Transcriptomics Sequencing of Mouse Liver at 2µm Resolution Using a Novel Spatial DNA Chip

Xun Ding[1], Kendall Hoff[1,2], Radha Swaminathan[3] Scott Pollom[1], Tianlong Huang[4], Xiaochi Li[1], Guoqiang Zhou[4], Zhicong Bai[4], Shizhe Yu[5], Zongping Xia[5], Mikaela Koutrouli[6], Lars Jensen[6], Filip Crnogorac[1], Su Yu[1], Glenn McGall[1], Jeremy Edwards[1,2], Wei Zhou[1]

1. Centrillion Technologies, Palo Alto, CA 94303
2. Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87131
3. Department of Internal Medicine, University of New Mexico, Albuquerque, NM 87131
4. Centrillion Technology, Hangzhou, Zhejiang Province, China 310052
5. Translational Medicine Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China 450052
6. Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Abstract

Spatial transcriptomics has showcased its efficacy in deciphering the intricate relationships between individual cells and tissues. We present spatial transcriptomics data using a novel high-resolution DNA chip with a capture region feature size of 2 x 2µm. Feature-to-feature gap space is zero, maximizing the capture area. Chips are manufactured at wafer scale using photolithography and are transferred to hydrogels, making them compatible with existing sample preparation and analysis workflows for fresh frozen or paraffin-embedded samples. For this report, we examined a fresh frozen sample from adult mouse liver. Using a bin size of 10,

representing a 20µm x 20µm capture area, at 69% sequencing saturation, we obtained over 600 million unique mapped reads, the median number of genes captured was over 8000 per region, which demonstrated potential for obtaining additional unique reads with deeper sequencing. This high-resolution mapping of liver cell types and visualization of gene expression patterns demonstrates significant advances in spatial sequencing technology.

Introduction

The high-throughput exploration of gene expression profiles has transformed our understanding of cellular processes and molecular mechanisms[1]. Traditional transcriptomics methods, such as bulk RNA sequencing, provide invaluable insights into global gene expression patterns within a tissue or cell population. However, they often overlook the intricate spatial organization of cells within complex biological systems[1,2]. The advent of spatial transcriptomics technologies has revolutionized our ability to capture gene expression data within its native spatial context, enabling us to delve deeper into the heterogeneity and interactions of cells within tissues[1,2].

Spatial transcriptomics bridges the gap between high-resolution spatial information and transcriptomic data, allowing researchers to unravel the spatial distribution of gene expression across tissue sections[1–3]. This approach not only provides a detailed map of cellular activity within its anatomical context but also offers a unique opportunity to decipher the molecular signatures that underlie cellular heterogeneity[4], tissue development[5,6], and disease progression. By preserving the intricate relationships between neighboring cells and their surrounding microenvironment, spatial transcriptomics unlocks new dimensions of biological understanding[3], making it a pivotal tool in the fields of cancer research[4,7–11], neuroscience[12], developmental biology[5,6], and drug development[13]. Molecular spatial analysis methods are divided into two approaches: imaging-based approaches and sequencing-based approaches. Imaging-based approaches suffer from low throughput, only capable of capturing the expression of a handful of transcripts per imaging run, while sequencing-based approaches suffer from low resolution[14]. This work demonstrates significant improvement in sequencing-based spatial sequencing.

Microarrays for spatial trancriptomics, also called spatial DNA chips, are manufactured as DNA microarrays with each feature or "spot" containing a poly-T capture sequence for mRNA and a unique barcode sequence to indicate the feature to which the downstream sequencing reads correspond. Due to limitations in the resolution of commercially available microarrays employed in spatial transcriptomics (10-55 µm)[15], additional single-cell sequencing data has been required in order to provide detailed resolution and capture efficiency. Several new technologies have come out in the last two to three years that boast improved resolution but still have large gaps between features[16,17]. Photolithographic microarray fabrication methods are currently used for the production of commercial microarrays with feature spacing in the range of 4-10 µm[18,19] We have previously demonstrated technological advances in wafer-scale manufacturing of 3'-up functional DNA chips in hydrogels with sub-micrometer resolution[20]. Using photoresist technology (proprietary) that uses photoacid generated chemistry in a polymer matrix to spatially deblock DMT protecting groups[19,20], incredibly high-resolution, completely confluent DNA chips can be manufactured at semi-conductor scale.

The work in this paper demonstrates the use of these chips with tissue sections generated from fresh frozen adult mouse liver. The mammalian liver performs a diverse array of functions with approximately 50% of hepatocyte genes expressed in a spatially differentiated manner[21]. The liver is divided into lobules, repeating hexagonal units with approximately 12 concentric layers of hepatocytes[21,22]. Blood flows in from portal nodes and drains from central veins, creating a gradient of oxygen, glucose, and hormones[22]. Periportal hepatocytes function in a microenvironment of high oxygen and hormone levels and primarily function in

gluconeogenesis, ureagenesis, cholesterol biosynthesis, protein secretion, and β-oxidation[21,22]. Conversely, the microenvironment of pericentral hepatocytes has high levels of Wnt (secreted by endothelial cells surrounding central veins[22]), low levels of oxygen, and low levels of circulating hormones[21,22]. The primary functions of these cells is glycolysis, bile acid production, glutamine synthesis, lipogenesis, and xenobiotic metabolism[21,22]. As an organ with such diverse and critical functions, the liver has been extensively studied using scRNA-seq and various spatial methods[21–24]. This makes the liver an excellent candidate for the evaluation of emerging spatial technologies.

Results and Discussion

High-resolution spatial DNA chips were manufactured as previously described[20]. Briefly, semiconductor photolithographic manufacturing techniques were used in conjunction with highly-efficient light-directed DNA synthesis to prepare high-density DNA arrays on silicon wafer substrates. Positional information is encoded in the DNA sequences on the silicon chips photolithographically[20]. High-resolution chips are fabricated by sequential exposures of lithography masks, each one directing a single base addition. Oligonucleotides are subsequently transferred from the donor silicon substrate to a thin polyacrylamide gel covalently cast on a silanized glass slide[20].

Chips are manufactured as 6.5 x 6.5mm or 10 x 10mm with 5,000 x 5,000 (25 million total) features on the 10 x 10mm chips. Each wafer produces 111 10 x 10mm chips. Each feature is approximately square, with a size of 2µm x 2µm. Features are juxtaposed adjacently without interstitial spaces (Fig. 1). Embedded within each feature is a unique predefined sequence,

consisting of the partial Illumina sequencing primer 1 for amplification, a Unique Molecular Identifier (UMI), a zipcode, and a 30 nucleotide poly-T sequence followed by 3'-VN-5', which is optimized to capture the polyA region of an mRNA molecule just before the coding region (Fig. 1). The zipcode contains X and Y coordinates that are decoded during bioinformatics analysis following sequencing using custom PostMaster™ software (Centrillion Tech, Palo Alto, CA). For the work presented in this analysis, 6.5 x 6.5 mm synthesized chips were used.

Zipcodes are designed to reduce errors during DNA zipcode synthesis. A long-range minimum edit distance, which dictates for some given D, if 2 zipcodes are $\geq D$ units apart on the chip, their edit distance must be D. For this particular mask set, $D = 5$. Figure 1 illustrates the method of producing vertical X-zipcodes on the array. For horizontal Y-zipcodes, the masks are rotated 90 degrees counter-clockwise and the process is repeated again, creating a quilt of unique X-Y zipcodes. Each zipcode border is delineated once and only once, alleviating the need for exacting overlay. In effect, neighboring zipcodes only differ by one base (gray-codes), but their spatial extent covers the full chip area eliminating any unused border regions. Such a scheme allows for 100% surface coverage. In terms of resolution, high-resolution chips used in this work have individual features of unique zipcodes that are 2 x 2 μm in size, however we have previously demonstrated this technology at 1μm resolution[20] and have been able to fabricate 0.66 x 0.66 μm features with additional layers of light directed synthesis (data not shown). Notably, any individual mask contains polygons (exposure areas) much larger than the minimum feature size. Their overlay creates the minimum feature size, and, as such, the limiting factor is the overlay accuracy of the mask alignment equipment used (±0.3 μm at 3-sigma).

High-resolution spatial DNA chips are compatible with existing workflows for fresh frozen or paraffin-embedded samples. The data presented here uses a fresh frozen sample from mouse liver. The assay workflow is outlined in Figure 1. Briefly, tissues are fixed, permeabilized, and decrosslinked, and RNA from the tissue is hybridized to the chip. Following hybridization, RNA sequences are copied onto the chip using a reverse transcriptase and a template-switching oligonucleotide[25,26]. The original RNA strand is removed and a second strand is synthesized, stripped from the chip, and used in downstream library preparation, which is then sequenced using an Illumina system.

The spatial transcriptome library was sequenced in four separate batches using the Illumina Novaseq 6000 system. This yielded a total of 3,890,659,083 reads from counts of 354,336,561, 501,709,129, 1,437,827,348, and 1,596,786,045 in each batch, respectively. Of this grand total, 3,478,542,340 reads contained a "zipcode" that was decoded using PostMaster software (Centrillion Tech, Inc., Palo Alto, CA). Briefly, the PostMaster decoder compares putative zipcodes extracted from the reads and designed zipcodes and allows the desired edit distance not larger than 3]. In the spatial transcriptome data analysis, among the 3,563,068,585 read sequences identified with a zipcode structure, a total of 3,467,777,970 were successfully decoded, resulting in a decoding rate of 97.33%. Within this decoded subset, the proportion of sequences with an edit distance of 3 was 1.18% for X zipcodes and 1.29% for Y zipcodes. Upon decoding, 2,399,941,684 of these reads were successfully mapped to a unique location of the mouse genome GRCm39 using the STAR, with 2,100,341,576 of these aligning to a specific gene by featureCounts. Subsequent to this mapping, deduplication was undertaken. Reads that shared the same Ensembl gene ID, X position, Y position, and UMI were classified as duplicates;

only the first occurrence was retained. Post-deduplication, a total of 643,142,336 unique reads remained. This implies a sequencing saturation rate of 69.37%, suggesting the potential for obtaining a higher number of unique reads with additional sequencing. Capture efficiency (Tables 1 and 2) is comparable to previously-reported scRNA-seq methods[27].

Despite the advancements associated with the high-resolution chip, challenges persist. Delineating precise cell boundaries and achieving accurate single-cell genetic mapping remain areas of active research[28]. Potential solutions may involve the deployment of advanced algorithms, possibly integrating machine learning techniques, to enhance spatial transcriptomics data interpretation[29–31]. Additionally, utilizing markers from specific subcellular compartments, such as pre-mRNA, and merging high-resolution microscopy might further refine cellular boundary determinations in tissues with intricate architectures[30].

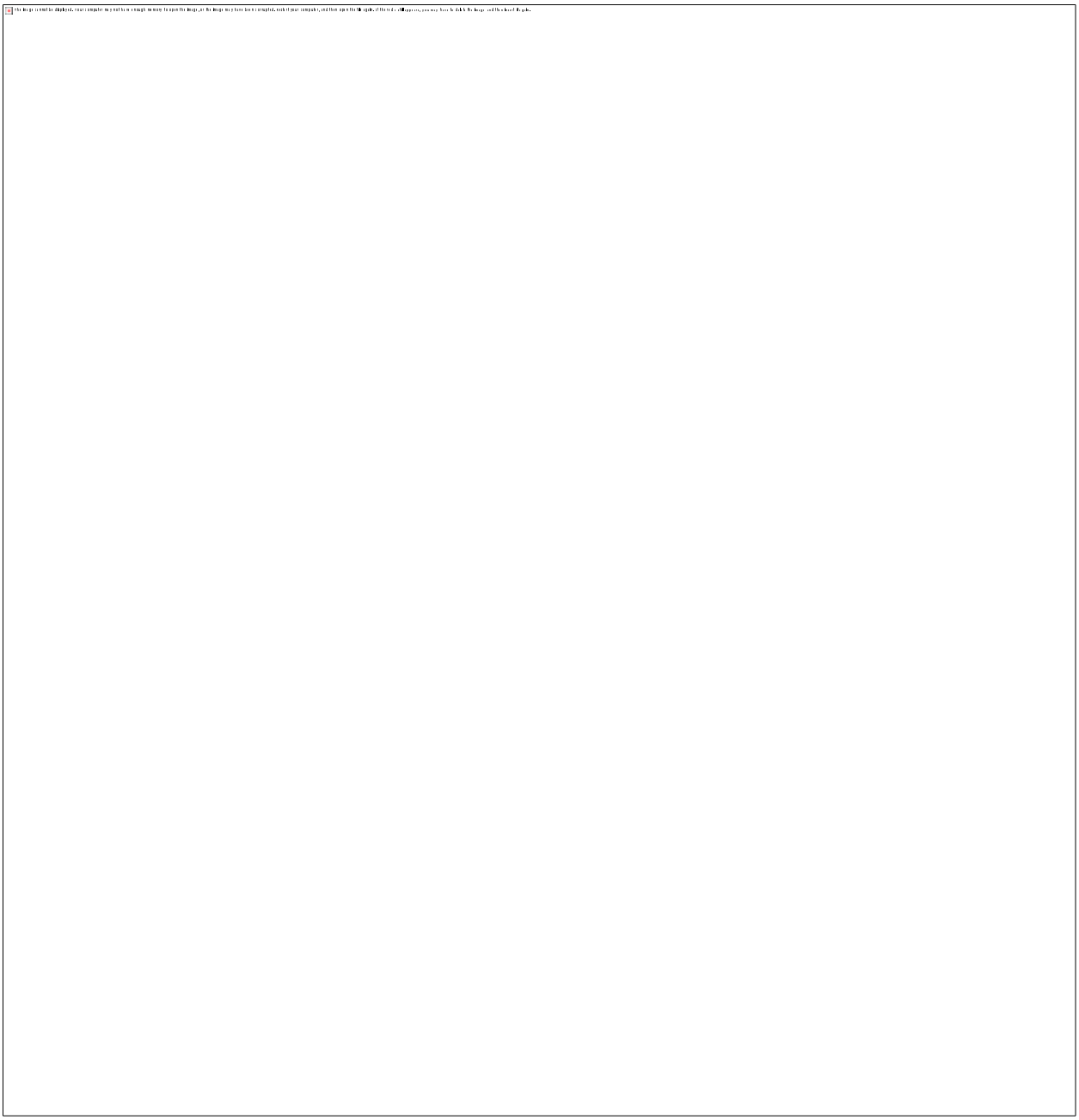

Figure 1: Chip Design and Performance. (A) Manufacturing of individual chips using photolithography. Vertical manufacturing of the zipcode design is shown, giving X positional information. Y positional information is achieved through the same methods with horizontal light-directed synthesis. **(B) Chip Design.** High-resolution spatial DNA chips are first

manufactured using light-directed synthesis on silicon wafers then transferred to hydrogels as previously described[20]. Capture regions are confluent 2 x 2 µm squares. **(C) Assay workflow.** First, the sample is permeabilized and RNA is hybridized to the chips. The next steps of the assay workflow are reverse transcription using a template-switching oligonucleotide (TSO), and second strand synthesis. The second strand is stripped from the chip and used in library preparation and downstream sequencing. **(D) Structure of final sequencing library. (E) Chip performance.** Number of UMI and number of genes captured for binning sizes of 1, 5, 10, and 16. **(F) Sequence saturation.** Duplicated reads, defined as those sharing identical spatial coordinates, UMI, and gene identifiers with prior reads, are excluded to calculate the number of unique reads. The plot shown displays the saturation rate—the ratio of duplicated reads to the total number of reads—emphasizing the incremental accumulation of redundancy with the progression of sequencing depth.

Bin	Size	Mean	Median	Min	Max	IQR
none	2µm x 2µm	84.24173	55	3	1579	88
5	10µm x 10µm	2125.6462	2026	1000	12573	689
10	20µm x 20µm	8552.983	8274	5000	36385	1789
16	32µm x 32µm	21451.65	21040.5	9003	90103	4548.5

Table 1: UMI captured at various bin sizes.

Bin	Size	Mean	Median	Min	Max	IQR
none	2µm x 2µm	63.0764901	46	1	736	64
5	10µm x 10µm	861.295031	840	334	2835	202
10	20µm x 20µm	2210.64091	2181	1495	4982	271
16	32µm x 32µm	3733.74565	3723	2242	7055	391

Table 2: Number of genes captured at various bin sizes.

The liver performs a vast array of metabolic functions that are regulated in spatial[21,22,32], temporal[22,33], and injury-related[32] manners as determined by the microenvironment of each hepatocyte[21,22] in the liver acinus. Zonation in the liver has been studied for over 80 years[34]. Liver lobules are hexagonally shaped with portal arteries and veins supplying blood at each of the 6 points of the hexagon and a central vein, through which blood is drained, at the center of the lobule[21,22,32]. Each of the approximately 12 concentric layers of hepatocytes has a microenvironment slightly different from its neighbors', defined by gradients of oxygen, glucose, pancreatic hormones such as insulin, morphogens, etc[21,32,32].

This structure enables production line patterns of cells, as with the synthesis of bile salts[22]. The first two enzymes in this cascade, CYP7A1 and HSD3B7 are expressed in the most pericentral cell layers where cholesterol, the starting material in the pathway, is most abundant[22]. The next-two enzymes in the cascade, CYP8B1 and CYP27A1, are expressed in the next layer, just periportal to those cells in layer 1[22], indicating different steps of the enzymatic cascade occur in a

spatially-regulated fashion. Researchers have further delineated the intricate zonation of opposing pathways in the liver including hepatic glucose metabolism (glycolysis and gluconeogenesis), glutamine synthesis and ureagenesis, lipogenesis and β-oxidation, protein secretion, and defense against pathogens and xenobiotics[21,22,32,35,36]. These distinct markers and intricate zonal distributions of cells emphasize the specialized roles of hepatocytes across these separate hepatic zones.

A large number of markers have been reported for cells at various hepatocyte layers as well as for other liver cells such as endothelial cells and Kupffer cells[21,22,32,33,37–39]. The dynamic expression patterns of several canonical marker genes can be observed in Figure 2 GLUL is expressed by pericentral hepatocytes[16,21,37,39]. It encodes glutamine synthetase, which converts glutamate into glutamine[38]. OAT encodes organic anion transporters, mainly involved in the uptake of bile acids and exogenous drugs including cancer medication[40,41]. It is also expressed by pericentral hepatocytes[16,38]. HAMP is expressed by midlobule hepatocytes[16,39]. It encodes hepcidin, which blocks the iron export protein, ferroportin, thus controlling iron storage in the liver and systemic iron availability[42–44]. CYP2F2 and CPS1 are periportal hepatocyte markers[16,38,45]. There are 450 human cytochrome P450 proteins, which are involved in xenobiotic metabolism of which CYP2F2 encodes one[38,46]. CPS1 is involved in ammonia metabolysis into urea[45]. JCHAIN, as previously discussed, is expressed by immune cells and liver epithelial cells[47,48]. Expression analysis demonstrates the ability to capture and identify genes that are representative of distinct regions of the mouse liver.

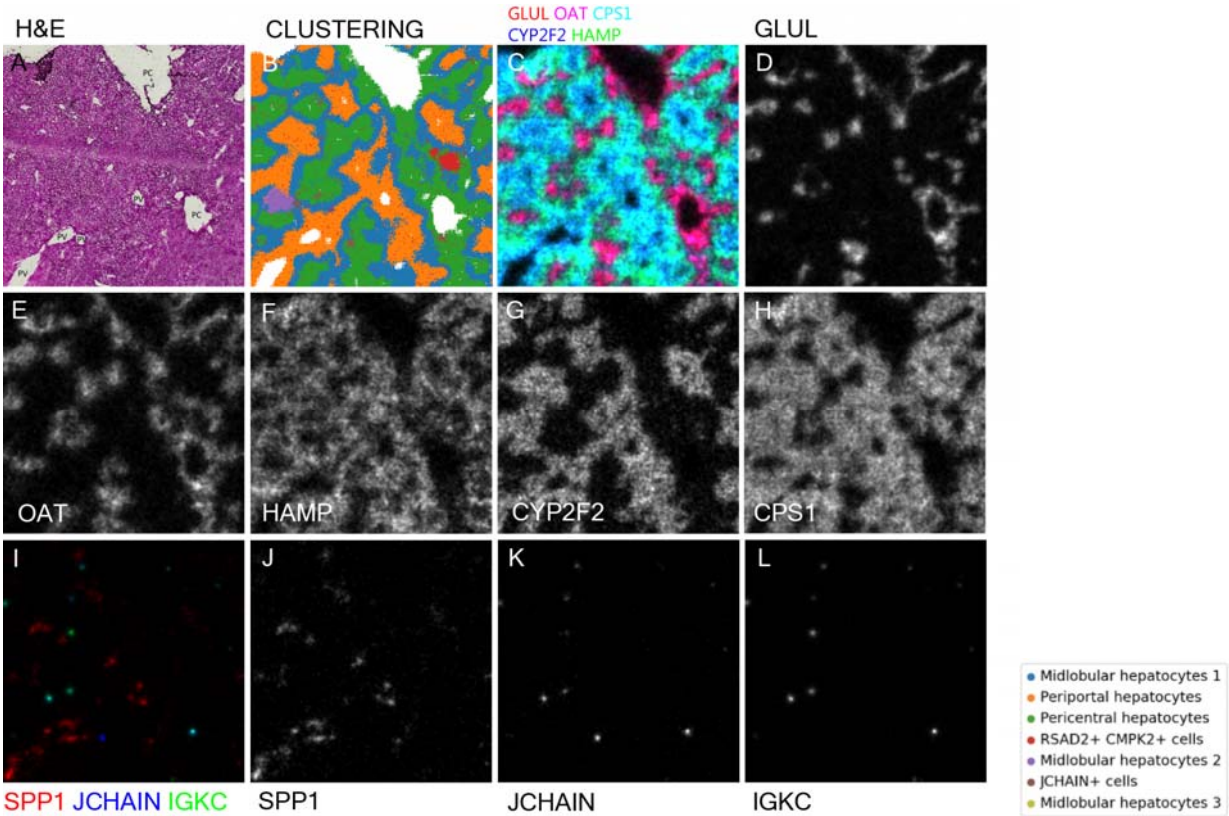

Figure 2: Spatial expression of cell-specific markers. For comparison, an H&E staining of a neighboring section is shown in panel A and clustering results from Figure 3 are shown in panel B. The key on the right displays cell types determined through clustering. Transcriptomics results are shown for markers for periportal hepatocytes (CPS1, CYP2F2), pericentral hepatocytes (GLUL, OAT), and midlobular hepatocytes (HAMP) are shown in panels C through H. Panel C offers a composite view, merging the information from Panels D through H. Additionally, Panels I through L display markers found in other liver cells. SPP1 is a marker found in biliary epithelial cells (BECs), Kupffer cells, and activated hepatic stellate cells (HSCs). JCHAIN is expressed in immune cells and liver endothelial cells. Immunoglobulin IGKC is expressed by immune cells. Panel I is a merged representation of Panel I through L.

From Figure 3, it can be observed that the gene expression pattern closely follows the shape of the tissue section. This would indicate the capture of mRNA was specifically restricted to the area where the tissue section was present. Unsupervised clustering[49] identified seven distinct cell types (Fig. 3): periportal hepatocytes, pericentral hepatocytes, three distinct midlobular hepatocyte cell types, and two clusters of cells expressing unique markers, the first expressing JCHAIN and the second expressing RSAD2 and CMPK2. While midlobular clusters 2 and 3 have markers solely expressed in these clusters, genes expressed in midlobular cluster 1 are not exclusive to midlobular cluster 1 (Fig. 3).

JCHAIN encodes a small polypeptide (joining chain) that governs multimerization of secretory immunoglobulins M and A[47]. While it is an immune cell marker, it is also expressed by liver epithelial cells[47,48]. The JCHAIN+ cluster also shows strong expression of CD74, the cognate receptor for macrophage migration inhibitory factor (MIF)[50]. CD74 is expressed by hepatocytes and hepatic stellate cells[50,51]. CD74 plays a role in antigen presentation in the liver in models of fibrosis following toxic liver injury and both MIF and CD74 are overexpressed in liver during hepatocarinogenesis[50–52]. In the liver, CD74 is expressed on hepatocytes as well as hepatic stellate cells (HSC)[51]. Specifically, the MIF/CD74 axis was identified to exert hepatoprotective effects in a model of NASH, as well as liver fibrosis induced by chronic toxic liver injury [52,53].

RSAD2 (also known as Viperin) and CMPK2 are interferon (IFN) inducible genes[54]. IFNs are produced in response to viral infection[54]. However, viperin has been demonstrated to be highly expressed in liver in the absence of IFN-stimulation, suggesting a need for further research to better understand the role of these IFN-inducible genes within the context of liver[55]. Further

investigation is required to determine whether the tissue can be used to generate additional meaningful clusters. For reference, a bin size of 5, corresponding to 10 x 10 µm regions, was used to generate the images in Figures 2 and 3. Hepatocytes are approximately 80000 μm^3 or 20 x 20 x 20 µm in diameter in adult mice[56,57] .

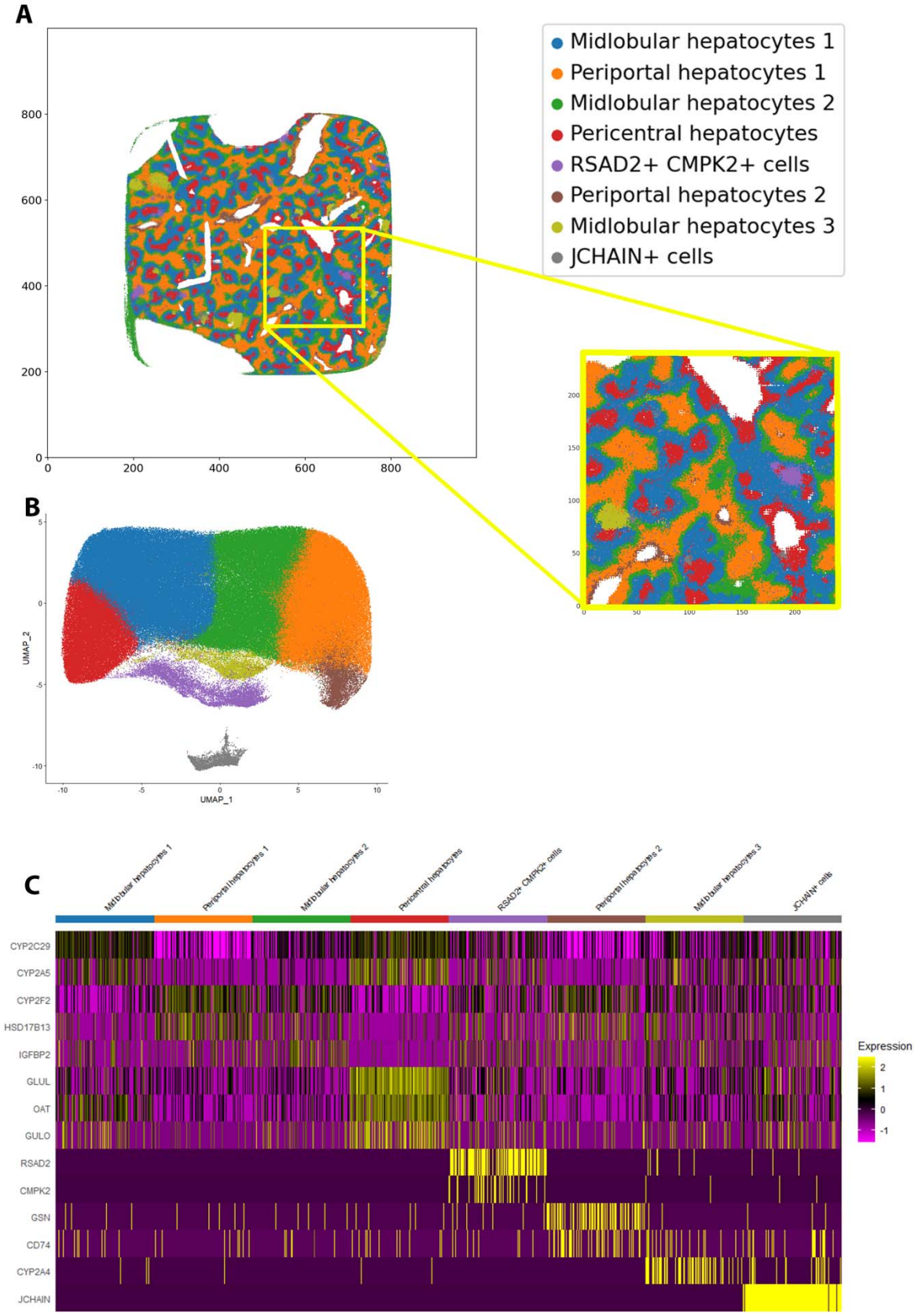
A
Midlobular hepatocytes 1
Periportal hepatocytes 1
Midlobular hepatocytes 2
Pericentral hepatocytes
RSAD2+ CMPK2+ cells
Periportal hepatocytes 2
Midlobular hepatocytes 3
JCHAIN+ cells
B
UMAP_1
UMAP_2
C
Midlobular hepatocytes 1
Periportal hepatocytes 1
Midlobular hepatocytes 2
Pericentral hepatocytes
RSAD2+ CMPK2+ cells
Periportal hepatocytes 2
Midlobular hepatocytes 3
JCHAIN+ cells
CYP2C29
CYP2A5
CYP2F2
HSD17B13
IGFBP2
GLUL
OAT
GULO
RSAD2
CMPK2
GSN
CD74
CYP2A4
JCHAIN
Expression

Fig 3. Successful capture of the mouse liver transcriptome. (A) Reconstructed map of spatial expression based on unsupervised clustering data. Hash marks represent barcode positions with each barcode being 0.2µm in height and width. **(B)** UMAP plot for clustering data. **(C)** Differential metabolic profile analysis of cell clusters. Full expression levels are provided in Supplementary materials.

In order to better visualize the spatial composition of the tissue, Cell2location[29] and U-CIE[58] were used to generate cell type spatial and cluster maps. Cell2location is a spatial deconvolution method used here to interpret gene expression data within liver tissue sections, as referenced from the Liver Cell Atlas. By applying Cell2location, we quantify the cell type composition at each tissue location, integrating this information to form a detailed cellular map. The analysis translates high-dimensional transcriptomic data into a resolved spatial context, revealing the distribution and abundance of cell types across the liver section. The result is a color-coded representation that provides insights into the tissue's cellular architecture (Figure 4A), with the intensity of each color corresponding to the presence of specific cell types. This approach highlights the spatial heterogeneity of the liver and can be used to identify regions with distinct cellular functions or compositions, offering a window into the intricate structure and functionality of liver tissue at the cellular level.

U-CIE[58] is utilized for visualizing gene expression with UMAP[59] condensing high-dimensional data into three dimensions[60]. This maintains the spatial integrity of gene expression patterns. Post-reduction, U-CIE translates the spatial data into the CIELAB color space, highlighting gene

expression variations. The visualization turns gene expression at each liver location into a color (Figure 4B), demonstrating the heterogeneity of the liver's cellular landscape. Further, single-cell data from the Liver Cell[61] is projected in the three UMAP dimensions of the spatial data, with U-CIE coloring applied to depict cell types (Figure 4C-D). This combination reveals the distribution of cell types across the liver, providing a color-coded map of cellular organization.

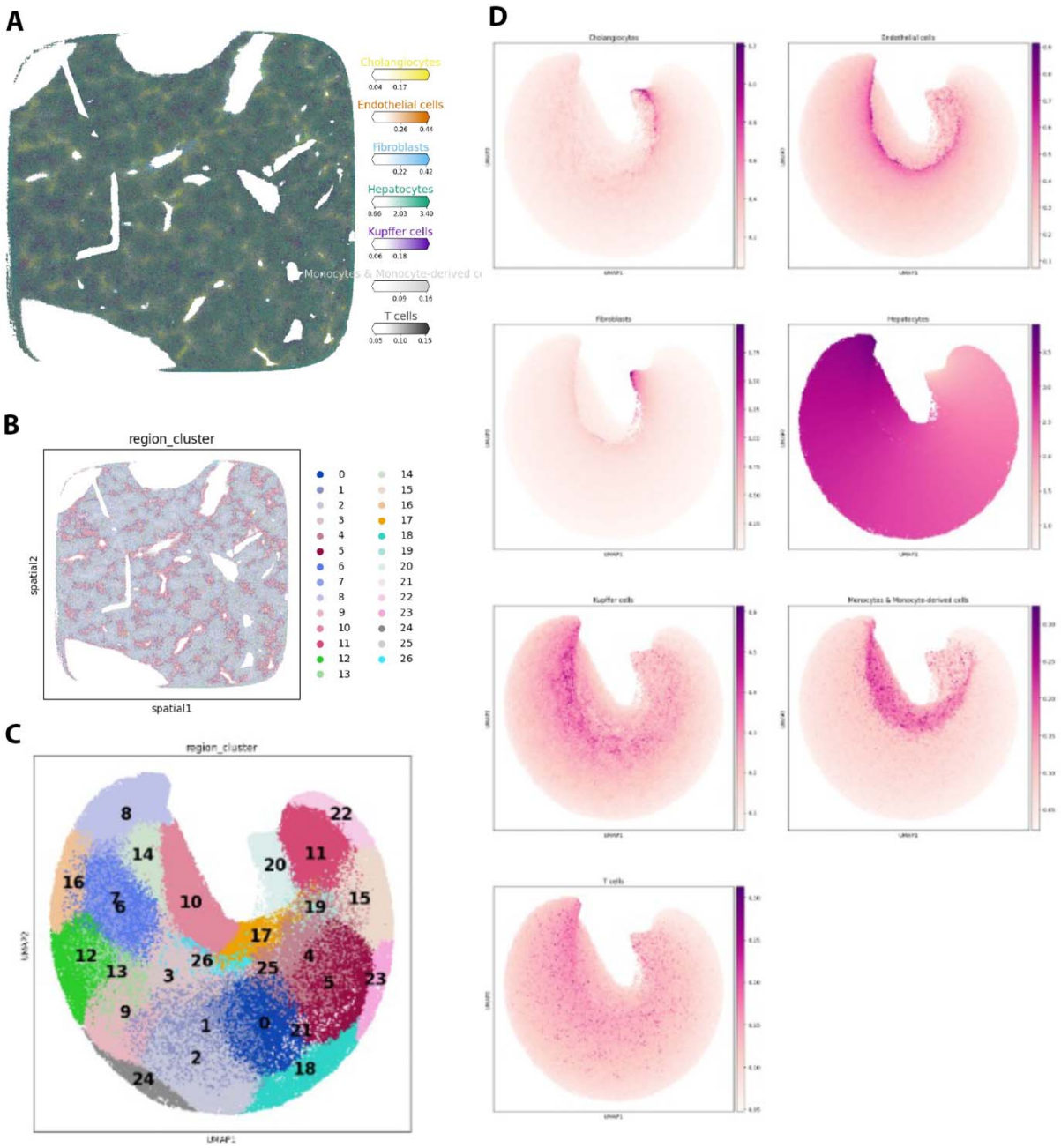

Figure 4: Cell type composition and U-CIE. (A) Cell2Location analysis of cell types. Cholangeocytes, endothelial cells, fibroblasts, hepatocytes, Kuppfer cells, macrophages and monocytes, and T cells are color-coded and plotted using the spatial zipcode. **(B-D)** Clustering data using UMAP and U-CIE.

Materials and Methods

Chip Fabrication and Decoding. The high-resolution chips, fabricated by Centrillion Technologies, are synthesized on 150mm silicon wafers using aligner model CENAL1000 (Centrillion Taiwan), generating 111 10mm x 10mm chips per wafer. Chips are singulated and subsequently inverted and transferred to a thin polyacrylamide gel (10-20 µm) cast on a silanized glass slide. This process has been previously described[20]. Each of the 10mm x 10mm chips contains 5,000 x 5,000 features that are 2µm x 2µm each and juxtaposed adjacently without interstitial spaces. Embedded within each feature is a predefined sequence, consisting of a fragment of the Illumina read primer 1 for amplification, a Unique Molecular Identifier (UMI), a zipcode, and a 3' sequence, 3'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-5', which is optimized to capture the polyadenylated region of an mRNA molecule closest to the coding region. The zipcode is constructed from a 15-16 nucleoside sequence indicative of the X coordinate, a 'GGG' separator, and another 15-16 nucleoside sequence corresponding to the Y coordinate. This zipcode can be effectively decoded using the PostMaster software. The sequences common to all barcode probes (poly-T, UMI & primer), which required no spatial definition, were synthesized using standard 5'-DMT deoxynucleoside synthesis reagents & coupling protocols[20,62]. The spatially-defined zipcode sequences were synthesized photolithographically using high-efficiency 5'-(2-nitronaphth-1-yl)benzyloxycarbonyl ("NNBOC") phosphoramidite reagents with standard coupling, masking, alignment and exposure protocols.[(use two refs above)] Vacuum contact lithography was performed using chrome on quartz

masks. Two-micron features achieved by employing a modified masking protocol as described in Figure 1.

The barcodes were designed such that the “embeddings” of adjacent barcodes have an edit distance of 1. The term "embedding" refers to the binary sequence that specifies when the barcode's position on the chip is exposed to light during the chip synthesis via photolithography[63]. The embeddings of all the chip barcodes form a two-dimensional gray code[63]. Another important property of pairs of DNA barcodes on the chip is the minimum edit distance between barcodes that are sufficiently far apart. DNA barcodes that are 1, 2, 3, or 4 units apart have an edit distance equal to the number of units apart, whereas DNA barcodes that are at least 5 units apart have an edit distance of at least 5. This property holds true for both sections of the DNA barcodes, meaning that if the X and Y coordinates of a pair of barcodes each differ by at least 5, the edit distance of the first sections will be at least 5, and the edit distance of the second sections will be at least 5. The combination of using gray codes and maintaining a minimum edit distance for distant DNA barcodes enables the chip to be fully covered by DNA barcodes, ensuring efficient transcript capture, and allowing for error correction during both chip synthesis and subsequent decoding.

Bioinformatics. The spatial sequencing data processing pipeline was constructed utilizing proprietary software, PostMaster, developed by Centrillion Technologies, Palo Alto. PostMaster parses Unique Molecular Identifiers (UMIs), and X, Y coordinates from the raw fastq files. It then decodes these coordinates, comparing them to the theoretical sequence of the decoded zipcodes. This decoding information, including edit distances and UMI, is stored within the

sequence identifiers of a new R2.fastq.gz file. Only coordinates with an edit distance less than or equal to three (compared to the theoretical sequence) are retained for further analysis. Subsequently, the trimmed reads are mapped to the reference genome using the STAR aligner. These mapped reads are then processed by FeatureCounts to assign them to genomic features, thus generating a count matrix.

Our custom script 'bamreader.sh' parses the decoding information stored in the sequence identifiers, creating a CSV file with gene ID, UMI, and spatial coordinates.
Finally, an in-house Python script converts this data into an h5ad file, a format suitable for spatial transcriptomic analysis. This script also removes duplicates, identified by UMI, X-Pos, Y-Pos, and Gene-id, ensuring the output consists of unique spatial gene expression data. The script constructs a sparse matrix, effectively integrating the gene expression and spatial data.

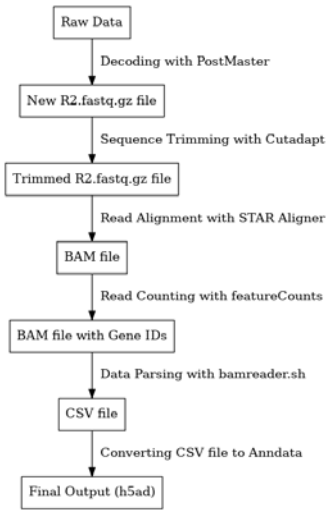

Figure 4: The bioinformatics pipeline. This study implemented a custom bioinformatics pipeline for the processing of spatial transcriptomics data obtained from paired-end FASTQ files.

Our method primarily consisted of five key steps: 1. Sequence Parsing and Decoding with PostMaster: Our initial preprocessing step involved the utilization of PostMaster software. This tool was specifically designed to parse sequencing data from read 1 sequence, extract unique molecular identifiers (UMIs), and decode x and y zipcodes. PostMaster then provides the x and y coordinates as well as the edit distance of each zipcode relative to the theoretical sequences. The output of this step is a new R2.fastq.gz file with decoding information stored in the sequence ID (SeqID). 2. Sequence Trimming with seqtk: Following decoding, we applied the seqtk tool to trim the sequences in the newly formed R2.fastq.gz file. This process is vital in the removal of sequencing adapters and other potential sequence contaminants. 3. Read Alignment with STAR Aligner: Subsequent to trimming, we aligned the cleaned sequences to a reference genome using the Spliced Transcripts Alignment to a Reference (STAR) software. The alignment parameters were set to ensure only unique matches to the reference genome were retained, thereby eliminating multi-mapping reads. This step resulted in a Binary Alignment Map (BAM) file. 4. Read Counting with featureCounts: After alignment, we employed the featureCounts tool to perform a gene-level quantification. Specifically, this step involved counting the number of reads that were successfully mapped to each gene in the reference genome. 5. Data Parsing with bamreader.sh: After featureCounts, we utilized a custom script, bamreader.sh, to parse the decoding information embedded within the sequence IDs in the BAM file. The output of this step was a CSV file containing detailed information such as gene_id, x and y coordinates, edit distance, and UMI, which can then be used for subsequent downstream analyses. 6. Converting CSV file to Anndata: The final step of the data processing pipeline converts the CSV output into a h5ad file, which is a binary format used to efficiently store large, multi-dimensional arrays. This is accomplished through a custom Python script using the Scanpy library, named h5ad.py.

In this step, any duplicate gene expression data, based on Unique Molecular Identifiers (UMI), X-Pos, Y-Pos, and Gene-id, are removed to ensure only unique spatial gene expression information is retained. The script then transforms this data into a sparse matrix representation, which helps to efficiently store and process the large amount of zero entries typical of gene expression data. Each unique gene and its corresponding spatial coordinates are recorded in this matrix, which is organized by gene symbols. Finally, an AnnData object is created with the sparse matrix, enabling the integration of gene expression and spatial data. This h5ad file provides a convenient and efficient format for subsequent data analysis tasks.

Spatial Library Construction. The spatial transcriptomics procedure is initiated with tissue fixation, permeabilization using pepsin, and decrosslinking. mRNA was captured through hybridization then copied using reverse transcription with a template-switching oligo. Templating RNA molecules were removed and the cDNA was copied using second-strand synthesis. The second strand was denatured from the chip and used in downstream library preparation. The library was subsequently purified and size-selected.

Pre-processing of Tissue

Tissue slides, retrieved from a -80□ freezer, were briefly warmed at 37□ for 1 minute. Slides were fixed using 100μl of 4% formaldehyde for 30 minutes at room temperature. Subsequently, decrosslinking was performed at 70□ over 20 minutes in TE Buffer (10mM Tris-HC1;1mM EDTA). The slides were then swiftly rinsed in a 1×PBS-filled staining jar. For permeabilization,

after removing excess water, the slides were treated with 100μL of 1/1000 dilution of 0.1% Pepsin and incubated at 37□ for 5 minutes.

Hybridization

For hybridization, a 20μL solution was formulated on ice with 9.4μL nuclease-free water, 5 μL 20× SSC, 2.4 μL 25mM MgCl2, 0.2μL 0.1M DTT, 2μL 10mM dNTPs, and 1μL RNase Inhibitor A (NEB, M0314L). A new Sequoia 2 chip was retrieved from the original container. After drying the surrounding of the chip's gel area, any residue liquid was quickly aspirated. Both the chip gel and the designated tissue area on the slide received 10μL of Hybridization Buffer. The slide was subsequently inverted onto the chip, ensuring a snug, bubble-free contact. This configuration was incubated at 42□ for 4 hours in a humidity chamber.

Reverse Transcription

For reverse transcription on the chip, a 40 μL RT MIX was prepared on ice, containing 4μL PEG8000, 0.4μL Triton X, 1.5μL TSO Oligo with sequence 5'-AAGCAGTGGTATCAACGCAGAGTACATrGrGrG, 2μL dNTPs, 1.6μL Tris HCl (pH 8.3), 1.2μL NaCl, 4μL MgCl2, 0.53μL GTP, 3.2μL DTT, 1μL RNase inhibitor (NEB, M0314L), 0.8μL Maxima H RT Enzyme (Thermo Fisher, EP0751), and 19.77μL nuclease free Water. After hybridization, the chip was rinsed in a staining dish with 0.1×SSC until the tissue slide released. A frame-seal (Biorad, SLF0201) was then aligned with the dashed region of the chip gel. The holes of the frame were filled with the RT Mix. The prepared chip underwent incubation in a pre-heated preservation box at 42□ for 90 minutes and subsequently at 53□ for 30 minutes.

Second Strand Synthesis

For second strand synthesis, a 50μL reaction mix was prepared on ice, using 24μL nuclease-free water, 5μL Thermopol Buffer, 4μL Bst Pol Full Length (NEB, M0328S), 7μL 10mM dNTPs, and 10μL of a primer with the sequence AAGCAGTGGTATCAACGCAGAG. A 20μL Wash Buffer SSB: 13.2μL nuclease-free water, 1μL Thermol Buffer, 2.8μL 10mM dNTPs, and 4μL primer.

The chip was sequentially cleaned: thrice using 75μL nuclease-free water, incubated for 2 minutes; twice with 75μL 0.08M KOH for 5 minutes; and twice with 75μL Tris (pH=7.5) for 2 minutes.

For synthesis, chips were washed in 20μL Wash Buffer SSB and 50μL Second Strand Synthesis Mix was applied. The chip underwent a 30-minute incubation in a 65□humidity chamber. The chip was then rinsed twice with 100μL Buffer EB and treated with 20μL 0.08M KOH for 10 minutes. After mixing, 20μL was moved to a 0.2ml PCR tube. This KOH process was repeated, combined in the tube, adjusted with 5μL of 1M Tris (pH 7.0), and stored on ice. The chip was finally preserved in 4xSSC at 4□.

cDNA Amplification and Purification

For cDNA pre-amplification, a mixture consisting of 50μL KAPA HiFi Hot Start Ready Mix (Roche, KK2601), 35μL cDNA, and 15μL of primers (Primer-f: CTACACGACGCTCTTCCGATCT; Primer-r: AAGCAGTGGTATCAACGCAGAG) was prepared in a total volume of 100μL. The PCR regimen included: initial denaturation at 98□ for 3 minutes, followed by 14 cycles at 98□ for 15 seconds, 63□ for 20 seconds, and 72□ for 1 minute, concluding with a 5-minute final extension at 72□. Purification of cDNA was performed

using 0.7X Clean NGS-R Beads (CleanNA, CNGS-0050The final elution yielded 20μL of size-selected cDNA, which was transferred to a fresh PCR tube for further analysis.

Fragmentation, End Repair, A-tailing and Ligation

For fragmentation, end repair, and A-tailing, an FS DNA Mix was prepared, consisting of 20μL Input DNA, 10μL Smearase® Mix (Yeasen, 12619ES24), and 30μL ddH2O. After gentle mixing and brief centrifugation, the mixture was subjected to a PCR program with a lid temperature set at 85□. The regimen began with fragmentation at 30□ for 2 minutes, followed by end repair and A-tailing at 72□ for 30 minutes, and then held at 4□.

For ligation, a Ligation Mix consisting of 60μL dA-tailed DNA, 20μL 5×Novel Ligation Buffer, 5μL Novel T4 DNA Ligase (Yeasen, 12626ES24), 3μL DNA Adapter (6μM, annealed from adapter_A (/5'Phosph/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC) and adapter_B (5’-GCTCTTCCGATCT)), and 12μL ddH2O was incubated at 25°C for 15 minutes.

Following ligation, the products underwent purification with Clean NGS-R Beads. Briefly, beads were mixed with the product, washed, and the DNA eluted. A size selection targeting 300-700 bp fragments was achieved using a stepwise application of 0.6× and then 0.8× bead volumes. The final purified DNA was resuspended, with 25μL of the resultant solution transferred to a new PCR tube for subsequent analyses.

Amplification and Adapter Addition.

For library construction, a 50μL reaction mix was prepared using 25μL KAPA HiFi Hot Start Ready Mix, 2.5μL NEBNext Index primer (NEB, E7335S), 2.5μL NEB Universal Primer, and

20μL of purified ligated product. The PCR amplification was executed with an initial denaturation at 98°C for 45 seconds, followed by 14 cycles of 98°C for 20 seconds, 67°C for 30 seconds, and 72°C for 20 seconds. This was finalized with a 72°C extension for 1 minute and then held at 4°C indefinitely. Post-amplification, the library underwent size selection using Clean NGS-R Beads. Briefly, beads (0.6× and 0.8× volumes) were applied in successive steps, with washes and magnetic separation. The purified library was eluted in 30.5μL nuclease-free water, with 30μL transferred to a new PCR tube.

Animal Husbandry and tissue preparation. Male C57BL/6 mice, eight weeks postnatal (PN), were used in this study. All procedures involving animals were conducted in strict accordance with the ethical guidelines established by the Institutional Animal Care and Use Committee of Zhengzhou University and were in compliance with the current laws regarding animal use in scientific research in China. Following perfusion, liver was harvested, embedded in OCT, stored at -80ºC, and sectioned at a thickness of 10μm.

Hematoxylin and Eosin (H&E) Staining. A section 20μm away from the section used for spatial transcriptomics was selected for H&E staining. Using the Beyotime C0105S Hematoxylin and Eosin Staining Kit, tissue samples were fixed with 4% paraformaldehyde at room temperature for 10 minutes. They were then stained with hematoxylin for 5 minutes, rinsed in water for 2 minutes, differentiated in 0.5% hydrochloric acid ethanol for 1 minute, followed by a 1-minute soak in 1xPBS. Finally, eosin staining was applied for 1 minute, and excess dye was removed with a 2-minute water rinse.

Computational analysis: Data binning. Cho et al. showed that 10µm sided grids produced a much less noisy UMAP relative to 5µm square grids and was simultaneously able to discover clusters with single cell resolution[16]. Therefore, we combined counts of 5 consecutive features, performed convolution using Gaussian kernel with a width of 1 feature resulting in 10µm-sided binned features.

Identifying regions within mouse liver tissue using unsupervised clustering. A threshold for total transcript count was applied to include only those binned features that were overlaid by the tissue. This binned count matrix was processed using Seurat v4 R package to identify regions present within the tissue. Briefly, the binned count matrix was normalized using Seurat's SCTransform function. Clusters using the normalized binned count matrix were identified using Seurat's shared nearest neighbor graph-based algorithm implemented in the FindNeighbors and FindClusters functions. A resolution of 0.2 was used to perform clustering.

Visualization and annotation of clusters. A custom python script was used to visualize clusters in the context of the tissue section. Marker genes for each cluster were identified using Seurat's FindAllMarkers function. Genes that have a positive logFC >0.25, p-adj < 0.05 and were present in at least 20% of cells were classified as marker genes. The marker genes identified in this study were compared to other studies to annotate clusters generated by unsupervised clustering.

Spatial Transcriptomics Data Analysis Using Cell2location. The analysis was performed using the Cell2location library, specifically employed for spatial transcriptomics data deconvolution. The spatial transcriptomics data were loaded using Scanpy, targeting the H5AD file that contains the preprocessed spatial gene expression matrix. Similarly, annotated single-

cell RNA-seq data from the Liver Cell Atlas was loaded as a reference dataset. Preliminary examination and preprocessing were performed on the spatial transcriptomics data. This included inspecting variable features and filtering out mitochondrial genes to clean the data for subsequent analysis. The single-cell RNA-seq data was prepared similarly, ensuring it was in an appropriate format for model input.

The Cell2location model was applied to deconvolute the spatial transcriptomics data, estimating cell type abundances at each spatial location. This involved integrating the spatial data with the reference single-cell RNA-seq data to create a comprehensive spatial map of cell distribution. Model parameters, fitting procedures, and diagnostic checks were carried out following best practices as recommended in Cell2location documentation and tutorials[64]. The function cell2location.models.RegressionModel() prepares the AnnData object for compatibility with the Cell2location model, ensuring that the input data conforms to the expected format and structure. The function RegressionModel() initializes the RegressionModel object with the preprocessed and formatted AnnData. This step is crucial for specifying the model to be fitted to the spatial transcriptomics data. The rest of the parameters were the default model configurations.

Upon successful model convergence and estimation, the cell type abundance results were visualized using Matplotlib, with particular attention to the spatial distribution of different cell types within the liver sections. This step is critical for interpreting the cellular composition and architecture of the liver tissue, as well as understanding the heterogeneity and spatial organization of cell types.

Conclusions

This work demonstrates significant advances in the photolithographic manufacturing of spatial DNA chips with a highly optimized capture surface. High-resolution spatial DNA chips are manufactured at semi-conductor scale using photolithography. This production technique is suitable for large-scale, highly efficient manufacturing, leveraging the principles of Moore's Law. The development of high-resolution chips represents a pivotal advancement in spatial transcriptomic technologies, merging positional information with extensive tissue coverage and efficient mRNA capture. To achieve these results, we utilized a combination of chemistry, chip design, and bioinformatics development.

The mammalian liver is a well-characterized, complex tissue containing many different cell types that perform coordinated metabolic functions. This makes it an excellent tissue to demonstrate the capabilities of new spatial sequencing technologies. We were able to validate the use of the chips with mouse liver, achieving sequencing results comparable to previously reported scRNA-seq data. This underscores the chip’s potential in comprehensively understanding tissue architectures and in early-stage pathological detection.

At 69% sequencing saturation, our sample yielded over 643 million effective reads from a single section with a median of 8,274 UMIs and 2,181 genes per 20 x 20 µm area (the average size of an adult mouse liver cell), demonstrating exceptional molecular capture capacity. Additional sequencing is ongoing and we expect to obtain additional unique reads. By integrating the analysis techniques of U-CIE and Cell2Location, we have been able to decode complex gene expression patterns within liver tissues, enhancing our understanding of cellular diversity and

structure. We believe the advancements demonstrated here establish a new benchmark for future research in the field. As the field evolves, integrating computational strategies and advanced microscopic methodologies with platforms like this will be paramount to deepening our insights into tissue dynamics and cellular intricacies.

Competing interest statement:

The authors declare the following competing financial interest(s): The Centrillion affiliated authors are employees and the company may commercialize the work described herein.

References

(1) Williams, C. G.; Lee, H. J.; Asatsuma, T.; Vento-Tormo, R.; Haque, A. An Introduction to Spatial Transcriptomics for Biomedical Research. *Genome Med.* **2022**, *14* (1), 1–18.

(2) Marx, V. Method of the Year: Spatially Resolved Transcriptomics. *Nat. Methods* **2021**, *18* (1), 9–14.

(3) Rao, A.; Barkley, D.; França, G. S.; Yanai, I. Exploring Tissue Architecture Using Spatial Transcriptomics. *Nature* **2021**, *596* (7871), 211–220.

(4) Lawson, D. A.; Kessenbrock, K.; Davis, R. T.; Pervolarakis, N.; Werb, Z. Tumour Heterogeneity and Metastasis at Single-Cell Resolution. *Nat. Cell Biol.* **2018**, *20* (12), 1349–1360. https://doi.org/10.1038/s41556-018-0236-7.

(5) Fawkner-Corbett, D.; Antanaviciute, A.; Parikh, K.; Jagielowicz, M.; Gerós, A. S.; Gupta, T.; Ashley, N.; Khamis, D.; Fowler, D.; Morrissey, E.; Cunningham, C.; Johnson, P. R. V.; Koohy, H.; Simmons, A. Spatiotemporal Analysis of Human Intestinal Development at Single-Cell Resolution. *Cell* **2021**, *184* (3), 810-826.e23. https://doi.org/10.1016/j.cell.2020.12.016.

(6) Asp, M.; Giacomello, S.; Larsson, L.; Wu, C.; Fürth, D.; Qian, X.; Wärdell, E.; Custodio, J.; Reimegård, J.; Salmén, F. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **2019**, *179* (7), 1647-1660. e19.

(7) Grauel, A. L.; Nguyen, B.; Ruddy, D.; Laszewski, T.; Schwartz, S.; Chang, J.; Chen, J.; Piquet, M.; Pelletier, M.; Yan, Z. TGFβ-Blockade Uncovers Stromal Plasticity in Tumors by Revealing the Existence of a Subset of Interferon-Licensed Fibroblasts. *Nat. Commun.* **2020**, *11* (1), 6315.

(8) Ji, A. L.; Rubin, A. J.; Thrane, K.; Jiang, S.; Reynolds, D. L.; Meyers, R. M.; Guo, M. G.; George, B. M.; Mollbrink, A.; Bergenstråhle, J. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* **2020**, *182* (2), 497-514. e22.

(9) Ahmed, R.; Zaman, T.; Chowdhury, F.; Mraiche, F.; Tariq, M.; Ahmad, I. S.; Hasan, A. Single-Cell RNA Sequencing with Spatial Transcriptomics of Cancer Tissues. *Int. J. Mol. Sci*. **2022**, *23* (6), 3042. https://doi.org/10.3390/ijms23063042.

(10) Berglund, E.; Maaskola, J.; Schultz, N.; Friedrich, S.; Marklund, M.; Bergenstråhle, J.; Tarish, F.; Tanoglidi, A.; Vickovic, S.; Larsson, L. Spatial Maps of Prostate Cancer Transcriptomes Reveal an Unexplored Landscape of Heterogeneity. *Nat. Commun.* **2018**, *9* (1), 2419.

(11) Yu, Q.; Jiang, M.; Wu, L. Spatial Transcriptomics Technology in Cancer Research. *Front. Oncol.* **2022**, *12*, 1019111.

(12) Ortiz, C.; Navarro, J. F.; Jurek, A.; Märtin, A.; Lundeberg, J.; Meletis, K. Molecular Atlas of the Adult Mouse Brain. *Sci. Adv.* **2020**, *6* (26), eabb3446.

(13) Lyubetskaya, A.; Rabe, B.; Fisher, A.; Lewin, A.; Neuhaus, I.; Brett, C.; Brett, T.; Pereira, E.; Golhar, R.; Kebede, S. Assessment of Spatial Transcriptomics for Oncology Discovery. *Cell Rep. Methods* **2022**, *2* (11).

(14) Madissoon, E.; Oliver, A. J.; Kleshchevnikov, V.; Wilbrey-Clark, A.; Polanski, K.; Orsi, A. R.; Mamanova, L.; Bolt, L.; Richoz, N.; Elmentaite, R. A Spatial Multi-Omics Atlas of the Human Lung Reveals a Novel Immune Cell Survival Niche. *Biorxiv* **2021**, 2021.11. 26.470108.

(15) Werner, E.; Roe, F.; Bugnicourt, A.; Franklin, M. J.; Heydorn, A.; Molin, S.; Pitts, B.; Stewart, P. S. Stratified Growth in Pseudomonas Aeruginosa Biofilms. *Appl. Environ. Microbiol.* **2004**, *70* (10), 6188–6196. https://doi.org/10.1128/AEM.70.10.6188-6196.2004.

(16) Cho, C.-S.; Xi, J.; Si, Y.; Park, S.-R.; Hsu, J.-E.; Kim, M.; Jun, G.; Kang, H. M.; Lee, J. H. Microscopic Examination of Spatial Transcriptome Using Seq-Scope. *Cell* **2021**, *184* (13), 3559-3572.e22. https://doi.org/10.1016/j.cell.2021.05.010.

(17) Xia, K.; Sun, H.-X.; Li, J.; Li, J.; Zhao, Y.; Chen, L.; Qin, C.; Chen, R.; Chen, Z.; Liu, G. The Single-Cell Stereo-Seq Reveals Region-Specific Cell Subtypes and Transcriptome Profiling in Arabidopsis Leaves. *Dev. Cell* **2022**, *57* (10), 1299-1310. e4.

(18) Hoff, K.; Ding, X.; Carter, L.; Duque, J.; Lin, J.-Y.; Dung, S.; Singh, P.; Sun, J.; Crnogorac, F.; Swaminathan, R.; Alden, E. N.; Zhu, X.; Shimada, R.; Posavi, M.; Hull, N.; Dinwiddie, D.; Halasz, A. M.; McGall, G.; Zhou, W.; Edwards, J. S. Highly Accurate Chip-Based Resequencing of SARS-CoV-2 Clinical Samples. *Langmuir* **2021**, *37* (16), 4763–4771. https://doi.org/10.1021/acs.langmuir.0c02927.

(19) Epstein, J. R.; Leung, A. P. K.; Lee, K.-H.; Walt, D. R. High-Density, Microsphere-Based Fiber Optic DNA Microarrays. *Biosens. Bioelectron.* **2003**, *18* (5), 541–546. https://doi.org/10.1016/S0956-5663(03)00021-6.

(20) Costa, J. A.; Dentinger, P. M.; McGall, G. H.; Crnogorac, F.; Zhou, W. Fabrication of Inverted High-Density DNA Microarrays in a Hydrogel. *ACS Appl. Mater. Interfaces* **2019**. https://doi.org/10.1021/acsami.9b07755.

(21) Ben-Moshe, S.; Itzkovitz, S. Spatial Heterogeneity in the Mammalian Liver. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16* (7), 395–410.

(22) Cunningham, R. P.; Porat-Shliom, N. Liver Zonation–Revisiting Old Questions with New Technologies. *Front. Physiol.* **2021**, *12*, 732929.

(23) He, J.; Deng, C.; Krall, L.; Shan, Z. ScRNA-Seq and ST-Seq in Liver Research. *Cell Regen.* **2023**, *12* (1), 1–11.

(24) Saviano, A.; Henderson, N. C.; Baumert, T. F. Single-Cell Genomics and Spatial Transcriptomics: Discovery of Novel Cell States and Cellular Interactions in Liver Physiology and Disease Biology. *J. Hepatol.* **2020**, *73* (5), 1219–1230.

(25) Zhu, Y. Y.; Machleder, E. M.; Chenchik, A.; Li, R.; Siebert, P. D. Reverse Transcriptase Template Switching: A SMART^{TM} Approach for Full-Length cDNA Library Construction. *Biotechniques* **2001**, *30* (4), 892–897.

(26) Hudson, W. H.; Sudmeier, L. J. Localization of T Cell Clonotypes Using the Visium Spatial Transcriptomics Platform. *STAR Protoc.* **2022**, *3* (2), 101391.

(27) Ziegenhain, C.; Vieth, B.; Parekh, S.; Reinius, B.; Guillaumet-Adkins, A.; Smets, M.; Leonhardt, H.; Heyn, H.; Hellmann, I.; Enard, W. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **2017**, *65* (4), 631-643.e4. https://doi.org/10.1016/j.molcel.2017.01.023.

(28) Fang, S.; Chen, B.; Zhang, Y.; Sun, H.; Liu, L.; Liu, S.; Li, Y.; Xu, X. Computational Approaches and Challenges in Spatial Transcriptomics. *Genomics Proteomics Bioinformatics* **2023**, *21* (1), 24–47. https://doi.org/10.1016/j.gpb.2022.10.001.

(29) Kleshchevnikov, V.; Shmatko, A.; Dann, E.; Aivazidis, A.; King, H. W.; Li, T.; Elmentaite, R.; Lomakin, A.; Kedlian, V.; Gayoso, A.; Jain, M. S.; Park, J. S.; Ramona, L.; Tuck, E.; Arutyunyan, A.; Vento-Tormo, R.; Gerstung, M.; James, L.; Stegle, O.; Bayraktar,

O. A. Cell2location Maps Fine-Grained Cell Types in Spatial Transcriptomics. *Nat. Biotechnol.* **2022**, *40* (5), 661–671. https://doi.org/10.1038/s41587-021-01139-4.

(30) Littman, R.; Hemminger, Z.; Foreman, R.; Arneson, D.; Zhang, G.; Gómez-Pinilla, F.; Yang, X.; Wollman, R. Joint Cell Segmentation and Cell Type Annotation for Spatial Transcriptomics. *Mol. Syst. Biol.* **2021**, *17* (6), e10108. https://doi.org/10.15252/msb.202010108.

(31) Chen, A. Spatiotemporal Transcriptomic Atlas of Mouse Organogenesis Using DNA Nanoball-Patterned Arrays. *OPEN ACCESS*.

(32) Paris, J.; Henderson, N. C. Liver Zonation, Revisited. *Hepatology* **2022**, *76* (4), 1219–1230.

(33) Droin, C.; Kholtei, J. E.; Bahar Halpern, K.; Hurni, C.; Rozenberg, M.; Muvkadi, S.; Itzkovitz, S.; Naef, F. Space-Time Logic of Liver Gene Expression at Sub-Lobular Scale. *Nat. Metab.* **2021**, *3* (1), 43–58. https://doi.org/10.1038/s42255-020-00323-1.

(34) Deane, H. W. A Cytological Study of the Diurnal Cycle of the Liver of the Mouse in Relation to Storage and Secretion. *Anat. Rec.* **1944**, *88* (1), 39–65.

(35) Zhu, S.; Rao, X.; Qian, Y.; Chen, J.; Song, R.; Yan, H.; Yang, X.; Hu, J.; Wang, X.; Han, Z.; Zhu, Y.; Liu, R.; Jong-Leong Wong, J.; McCaughan, G. W.; Zheng, X. Liver Endothelial Heg Regulates Vascular/Biliary Network Patterning and Metabolic Zonation Via Wnt Signaling. *Cell. Mol. Gastroenterol. Hepatol.* **2022**, *13* (6), 1757–1783. https://doi.org/10.1016/j.jcmgh.2022.02.010.

(36) Berndt, N.; Kolbe, E.; Gajowski, R.; Eckstein, J.; Ott, F.; Meierhofer, D.; Holzhütter, H.-G.; Matz□Soja, M. Functional Consequences of Metabolic Zonation in Murine Livers: Insights for an Old Story. *Hepatology* **2021**, *73* (2), 795–810.

(37) *SEAM is a spatial single nuclear metabolomics method for dissecting tissue microenvironment | Nature Methods*. https://www.nature.com/articles/s41592-021-01276-3 (accessed 2023-09-18).

(38) Braeuning, A.; Ittrich, C.; Köhle, C.; Hailfinger, S.; Bonin, M.; Buchmann, A.; Schwarz, M. Differential Gene Expression in Periportal and Perivenous Mouse Hepatocytes. *FEBS J.* **2006**, *273* (22), 5051–5061.

(39) Ben-Moshe, S.; Shapira, Y.; Moor, A. E.; Manco, R.; Veg, T.; Bahar Halpern, K.; Itzkovitz, S. Spatial Sorting Enables Comprehensive Characterization of Liver Zonation. *Nat. Metab.* **2019**, *1* (9), 899–911. https://doi.org/10.1038/s42255-019-0109-9.

(40) Rocha, A. S.; Vidal, V.; Mertz, M.; Kendall, T. J.; Charlet, A.; Okamoto, H.; Schedl, A. The Angiocrine Factor Rspondin3 Is a Key Determinant of Liver Zonation. *Cell Rep.* **2015**, *13* (9), 1757–1764. https://doi.org/10.1016/j.celrep.2015.10.049.

(41) Li, T.-T.; An, J.-X.; Xu, J.-Y.; Tuo, B.-G. Overview of Organic Anion Transporters and Organic Anion Transporter Polypeptides and Their Roles in the Liver. *World J. Clin. Cases* **2019**, *7* (23), 3915–3933. https://doi.org/10.12998/wjcc.v7.i23.3915.

(42) Muckenthaler, M.; Roy, C. N.; Custodio, Á. O.; Miñana, B.; deGraaf, J.; Montross, L. K.; Andrews, N. C.; Hentze, M. W. Regulatory Defects in Liver and Intestine Implicate Abnormal Hepcidin and Cybrd1 Expression in Mouse Hemochromatosis. *Nat. Genet.* **2003**, *34* (1), 102–107. https://doi.org/10.1038/ng1152.

(43) Kämmerer, L.; Mohammad, G.; Wolna, M.; Robbins, P. A.; Lakhal-Littleton, S. Fetal Liver Hepcidin Secures Iron Stores in Utero. *Blood* **2020**, *136* (13), 1549–1557. https://doi.org/10.1182/blood.2019003907.

(44) Nemeth, E.; Tuttle, M. S.; Powelson, J.; Vaughn, M. B.; Donovan, A.; Ward, D. M.; Ganz, T.; Kaplan, J. Hepcidin Regulates Cellular Iron Efflux by Binding to Ferroportin and Inducing Its Internalization. *science* **2004**, *306* (5704), 2090–2093.

(45) Martini, T.; Naef, F.; Tchorz, J. S. Spatiotemporal Metabolic Liver Zonation and Consequences on Pathophysiology. *Annu. Rev. Pathol. Mech. Dis.* **2023**, *18*, 439–466.

(46) Nagata, K.; Martin, B. M.; Gillette, J. R.; Sasame, H. A. Isozymes of Cytochrome P-450 That Metabolize Naphthalene in Liver and Lung of Untreated Mice. *Drug Metab. Dispos.* **1990**, *18* (5), 557–564.

(47) Erlandsson, L.; Akerblad, P.; Vingsbo-Lundberg, C.; Kallberg, E.; Lycke, N.; Leanderson, T. Joining Chain–Expressing and–Nonexpressing B Cell Populations in the Mouse. *J. Exp. Med.* **2001**, *194* (5), 557–570.

(48) Shao, W.; Zhang, C.; Liu, E.; Zhang, L.; Ma, J.; Zhu, Z.; Gong, X.; Qin, Z.; Qiu, X. Identification of Liver Epithelial Cell-Derived Ig Expression in μ Chain-Deficient Mice. *Sci. Rep.* **2016**, *6* (1), 23669. https://doi.org/10.1038/srep23669.

(49) Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W. M.; Zheng, S.; Butler, A.; Lee, M. J.; Wilk, A. J.; Darby, C.; Zager, M.; Hoffman, P.; Stoeckius, M.; Papalexi, E.; Mimitou, E. P.; Jain, J.; Srivastava, A.; Stuart, T.; Fleming, L. M.; Yeung, B.; Rogers, A. J.; McElrath, J. M.; Blish, C. A.; Gottardo, R.; Smibert, P.; Satija, R. Integrated Analysis of Multimodal Single-Cell Data. *Cell* **2021**, *184* (13), 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

(50) Wirtz, T. H.; Saal, A.; Bergmann, I.; Fischer, P.; Heinrichs, D.; Brandt, E. F.; Koenen, M. T.; Djudjaj, S.; Schneider, K. M.; Boor, P.; Bucala, R.; Weiskirchen, R.; Bernhagen, J.; Trautwein, C.; Berres, M.-L. Macrophage Migration Inhibitory Factor Exerts Pro-

Proliferative and Anti-Apoptotic Effects via CD74 in Murine Hepatocellular Carcinoma. *Br. J. Pharmacol.* **2021**, *178* (22), 4452–4467. https://doi.org/10.1111/bph.15622.

(51) Maubach, G.; Lim, M. C. C.; Kumar, S.; Zhuo, L. Expression and Upregulation of Cathepsin S and Other Early Molecules Required for Antigen Presentation in Activated Hepatic Stellate Cells upon IFN-γ Treatment. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2007**, *1773* (2), 219–231. https://doi.org/10.1016/j.bbamcr.2006.11.005.

(52) Heinrichs, D.; Knauel, M.; Offermanns, C.; Berres, M.-L.; Nellen, A.; Leng, L.; Schmitz, P.; Bucala, R.; Trautwein, C.; Weber, C. Macrophage Migration Inhibitory Factor (MIF) Exerts Antifibrotic Effects in Experimental Liver Fibrosis via CD74. *Proc. Natl. Acad. Sci.* **2011**, *108* (42), 17444–17449.

(53) Heinrichs, D.; Berres, M.-L.; Coeuru, M.; Knauel, M.; Nellen, A.; Fischer, P.; Philippeit, C.; Bucala, R.; Trautwein, C.; Wasmuth, H. E. Protective Role of Macrophage Migration Inhibitory Factor in Nonalcoholic Steatohepatitis. *FASEB J.* **2014**, *28* (12), 5136.

(54) El-Diwany, R.; Soliman, M.; Sugawara, S.; Breitwieser, F.; Skaist, A.; Coggiano, C.; Sangal, N.; Chattergoon, M.; Bailey, J. R.; Siliciano, R. F.; Blankson, J. N.; Ray, S. C.; Wheelan, S. J.; Thomas, D. L.; Balagopal, A. CMPK2 and BCL-G Are Associated with Type 1 Interferon–Induced HIV Restriction in Humans. *Sci. Adv.* **2018**, *4* (8), eaat0843. https://doi.org/10.1126/sciadv.aat0843.

(55) Eom, J.; Kim, J. J.; Yoon, S. G.; Jeong, H.; Son, S.; Lee, J. B.; Yoo, J.; Seo, H. J.; Cho, Y.; Kim, K. S.; Choi, K. M.; Kim, I. Y.; Lee, H.-Y.; Nam, K. T.; Cresswell, P.; Seong, J. K.; Seo, J.-Y. Intrinsic Expression of Viperin Regulates Thermogenesis in Adipose Tissues. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (35), 17419–17428. https://doi.org/10.1073/pnas.1904480116.

(56) Jiang, X.; Xu, J.; Gore, J. C. Mapping Hepatocyte Size in Vivo Using Temporal Diffusion Spectroscopy MRI. *Magn. Reson. Med.* **2020**, *84* (5), 2671–2683. https://doi.org/10.1002/mrm.28299.

(57) Morales-Navarrete, H.; Segovia-Miranda, F.; Klukowski, P.; Meyer, K.; Nonaka, H.; Marsico, G.; Chernykh, M.; Kalaidzidis, A.; Zerial, M.; Kalaidzidis, Y. A Versatile Pipeline for the Multi-Scale Digital Reconstruction and Quantitative Analysis of 3D Tissue Architecture. *eLife 4*, e11214. https://doi.org/10.7554/eLife.11214.

(58) Koutrouli, M.; Morris, J. H.; Jensen, L. J. U-CIE [/Ju□ ‘si□/]: Color Encoding of High-Dimensional Data. *Protein Sci.* **2022**, *31* (9), e4388. https://doi.org/10.1002/pro.4388.

(59) McInnes, L.; Healy, J.; Melville, J. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv Prepr. ArXiv180203426* **2018**.

(60) Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.-A.; Kwok, I. W.; Ng, L. G.; Ginhoux, F.; Newell, E. W. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat. Biotechnol.* **2019**, *37* (1), 38–44.

(61) Guilliams, M.; Bonnardel, J.; Haest, B.; Vanderborght, B.; Wagner, C.; Remmerie, A.; Bujko, A.; Martens, L.; Thoné, T.; Browaeys, R.; Ponti, F. F. D.; Vanneste, B.; Zwicker, C.; Svedberg, F. R.; Vanhalewyn, T.; Gonçalves, A.; Lippens, S.; Devriendt, B.; Cox, E.; Ferrero, G.; Wittamer, V.; Willaert, A.; Kaptein, S. J. F.; Neyts, J.; Dallmeier, K.; Geldhof, P.; Casaert, S.; Deplancke, B.; Dijke, P. ten; Hoorens, A.; Vanlander, A.; Berrevoet, F.; Nieuwenhove, Y. V.; Saeys, Y.; Saelens, W.; Vlierberghe, H. V.; Devisscher, L.; Scott, C. L. Spatial Proteogenomics Reveals Distinct and Evolutionarily Conserved Hepatic Macrophage Niches. *Cell* **2022**, *185* (2), 379-396.e38. https://doi.org/10.1016/j.cell.2021.12.018.

(62) McGall, G. H.; Fidanza, J. A. Photolithographic Synthesis of High-Density Oligonucleotide Arrays. *DNA Arrays Methods Protoc.* **2001**, 71–101.

(63) Raman Kumar, M.; Naveen Kumar, V. A Numerical Representation Method for a DNA Sequence Using Gray Code Method. In *Soft Computing for Problem Solving*; Das, K. N., Bansal, J. C., Deep, K., Nagar, A. K., Pathipooranam, P., Naidu, R. C., Eds.; Advances in Intelligent Systems and Computing; Springer: Singapore, 2020; pp 645–654. https://doi.org/10.1007/978-981-15-0184-5_55.

(64) *Welcome to cell2location's documentation! — cell2location documentation*. https://cell2location.readthedocs.io/en/latest/ (accessed 2024-01-04).