

## CANCER

# XYZeq: Spatially resolved single-cell RNA sequencing reveals expression heterogeneity in the tumor microenvironment

Youjin Lee<sup>1,2,3,4,\*†</sup>, Derek Bogdanoff<sup>5,6†</sup>, Yutong Wang<sup>7,8†</sup>, George C. Hartoularos<sup>9,12,21</sup>, Jonathan M. Woo<sup>1,2,3,4</sup>, Cody T. Mowery<sup>1,2,3,4,10,11</sup>, Hunter M. Nisonoff<sup>8</sup>, David S. Lee<sup>3,12,21</sup>, Yang Sun<sup>3,12,21</sup>, James Lee<sup>13</sup>, Sadaf Mehdizadeh<sup>2</sup>, Joshua Cantlon<sup>14</sup>, Eric Shifrut<sup>1,2,3,4</sup>, David N. Ngyuen<sup>1,2,3,4,15</sup>, Theodore L. Roth<sup>1,2,3,10,11</sup>, Yun S. Song<sup>16,17,18</sup>, Alexander Marson<sup>1,2,3,4,15,18,19,20,21,22\*</sup>, Eric D. Chow<sup>5,6\*</sup>, Chun Jimmie Ye<sup>12,18,20,21,22,23,24\*</sup>

Single-cell RNA sequencing (scRNA-seq) of tissues has revealed remarkable heterogeneity of cell types and states but does not provide information on the spatial organization of cells. To better understand how individual cells function within an anatomical space, we developed XYZeq, a workflow that encodes spatial metadata into scRNA-seq libraries. We used XYZeq to profile mouse tumor models to capture spatially barcoded transcriptomes from tens of thousands of cells. Analyses of these data revealed the spatial distribution of distinct cell types and a cell migration-associated transcriptomic program in tumor-associated mesenchymal stem cells (MSCs). Furthermore, we identify localized expression of tumor suppressor genes by MSCs that vary with proximity to the tumor core. We demonstrate that XYZeq can be used to map the transcriptome and spatial localization of individual cells in situ to reveal how cell composition and cell states can be affected by location within complex pathological tissue.

## INTRODUCTION

Over the past decade, massively parallel single-cell RNA sequencing (scRNA-seq) has emerged as a powerful approach to catalog the remarkable cellular heterogeneity in complex tissues (1, 2). While scRNA-seq can profile the transcriptomes of thousands of cells in a single experiment, it requires the dissociation of tissue into single-cell suspensions before library preparation and sequencing, eliminating any spatial information (3–6). Several strategies have emerged

to obtain molecular and spatial information simultaneously from complex tissue. Imaging-based strategy combines high-resolution microscopy with fluorescence in situ hybridization to achieve sub-cellular resolution and could profile the entire transcriptome (7–10), but this requires lengthy iterative microscopy workflows and large probe panels. Another approach is to hybridize RNA directly from tissue slices onto a microarray containing spatially barcoded oligo(dT) spots or beads to encode location information into RNA-seq libraries. These approaches can sample the entire transcriptome without the need for iterative rounds of hybridization (11), and recent improvements using DNA-barcoded beads (high-definition spatial transcriptomics and Slide-seq<sup>v1/v2</sup>) report spatial resolutions at or below the diameter of a single cell (12–14). However, because of the low numbers of mRNA molecules captured per bead, these spatial transcriptomic approaches often aggregate neighboring beads before downstream analysis, resulting in lower effective resolution and averaging of transcript abundances from multiple cells. As a result, annotation of specific cell types present within each spatial unit of analysis is accomplished by aggregating gene sets computationally defined from orthogonal scRNA-seq datasets (15, 16). While integration methods have demonstrated the ability to localize cell types within the spatial organization of complex tissue, they rely on having available data from two independent assays and have limited ability to infer how spatial context influences the cell state of individual cell types.

## RESULTS

To overcome these limitations, we have developed XYZeq, a method that uses two rounds of split-pool indexing to encode the spatial location of each cell from a tissue sample into combinatorially indexed scRNA-seq libraries (17, 18). Critical for the performance of XYZeq, we fixed tissue slices with dithio-bis(succinimidyl propionate) (DSP), a reversible cross-linking fixative that has been shown

<sup>1</sup>Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>2</sup>Diabetes Center, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>3</sup>Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA 94720, USA. <sup>4</sup>J. David Gladstone Institutes, San Francisco, CA 94158, USA. <sup>5</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA. <sup>6</sup>Center for Advanced Technology, University of California, San Francisco, San Francisco, CA 94158, USA. <sup>7</sup>Graduate Group in Biostatistics, University of California, Berkeley, CA 94720, USA. <sup>8</sup>Center for Computational Biology, University of California, Berkeley, CA 94720, USA. <sup>9</sup>Graduate Program in Biological and Medical Informatics, University of California, San Francisco, San Francisco, CA 94158, USA. <sup>10</sup>Medical Scientist Training Program, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>11</sup>Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>12</sup>Division of Rheumatology, Department of Medicine, University of California, San Francisco, CA 94143, USA. <sup>13</sup>Division of Hematology and Oncology, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>14</sup>Scienion AG, Volmerstrasse 7b, 12489 Berlin, Germany. <sup>15</sup>Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>16</sup>Computer Science Division, University of California, Berkeley, CA 94720, USA. <sup>17</sup>Department of Statistics, University of California, Berkeley, CA 94720, USA. <sup>18</sup>Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. <sup>19</sup>UCSF Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94158, USA. <sup>20</sup>Parker Institute for Cancer Immunotherapy, University of California, San Francisco, San Francisco, CA 94129, USA. <sup>21</sup>Institute for Human Genetics, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>22</sup>Gladstone–UCSF Institute of Genomic Immunology, San Francisco, CA 94158, USA. <sup>23</sup>Institute of Computational Health Sciences, University of California, San Francisco, San Francisco, CA 94143, USA. <sup>24</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94158, USA.

\*Corresponding author. Email: jimmie.ye@ucsf.edu (C.J.Y.); eric.chow@ucsf.edu (E.D.C.); alexander.marson@ucsf.edu (A.M.); youjin.lee@ucsf.edu (Y.L.)

†These authors contributed equally to this work.

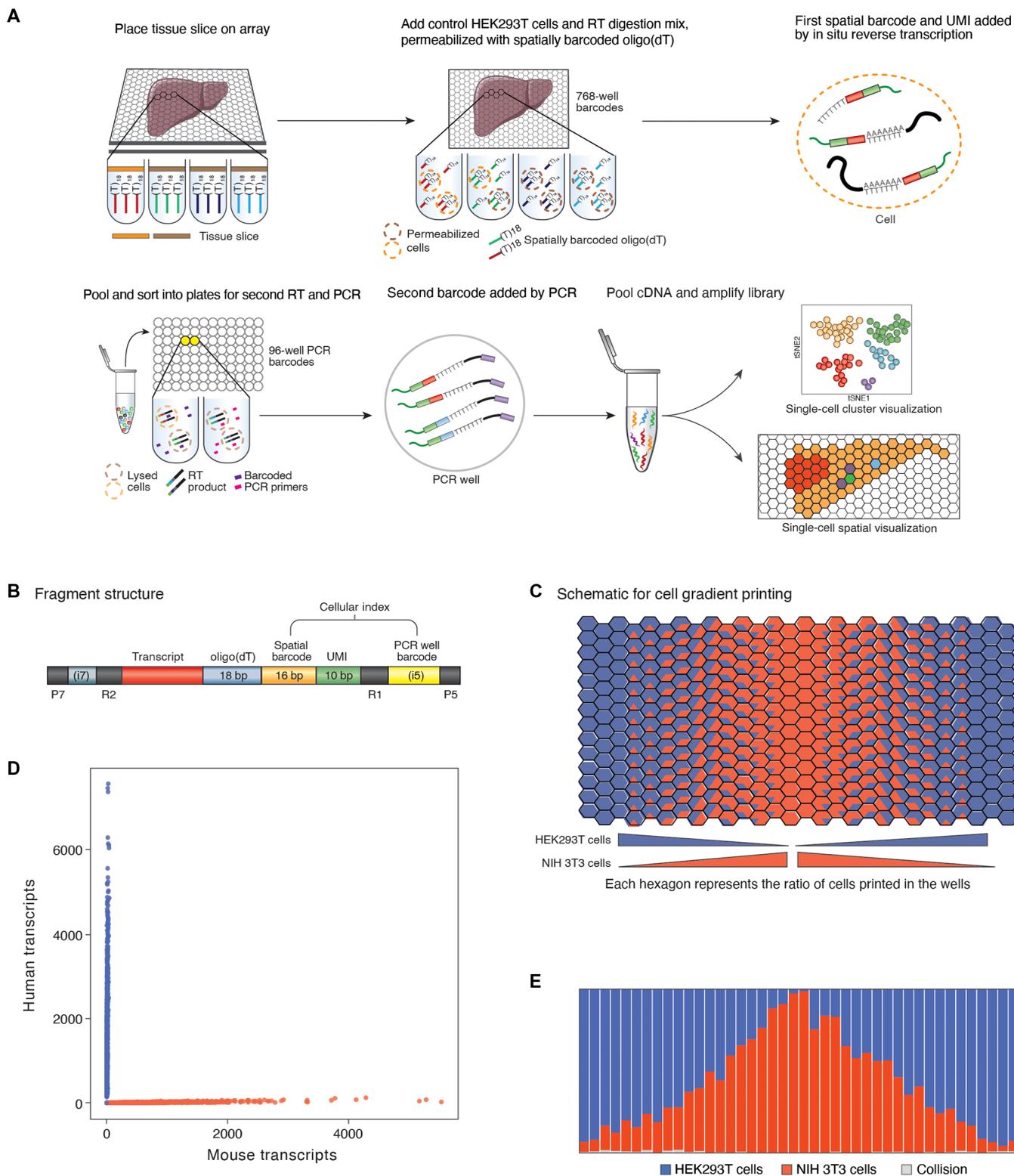
to preserve histological tissue morphology while maintaining RNA integrity for single-cell transcriptomics (19). In the first round of indexing, a fixed and cryo-preserved tissue section is placed on and sealed into an array of microwells spaced 500  $\mu\text{m}$  center to center. The microwells contain distinctly barcoded reverse transcription (RT) primers (spatial barcode). This step physically partitions intact cells from tissue into distinct *in situ* barcoding reactions. After RT, intact cells are removed from the array, pooled, and distributed into wells for a second round of polymerase chain reaction (PCR) indexing, imparting each single cell with a combinatorial barcode (Fig. 1, A and B). After sequencing and demultiplexing, the spatial barcode maps each cell back to its physical location in the array (Fig. 1B). This combinatorial barcoding strategy theoretically could enable spatial transcriptomic analysis of large sets of single cells—with two rounds of split-pool indexing, 768 spatial RT barcodes, and 384 PCR barcodes, up to 294,912 unique single-cell barcodes can be generated.

To determine whether XYZeq can assign transcriptomes to single cells, we performed a mixed-species experiment where a total of 11 distinct ratios of DSP-fixed human [human embryonic kidney (HEK) 293T] and mouse (NIH 3T3) cell mixtures were deposited into each of the 768 barcoded microwells, creating a cell proportion gradient along the columns of the array (Fig. 1C and Materials and Methods). XYZeq was used to generate scRNA-seq data for 6447 cells. A total of 94.8% of cell barcodes were assigned to a single species with an estimated barcode collision rate of 5.1% based on the percentage of cell barcodes with reads mapping to both human and mouse transcriptomes (fig. S1A). We hypothesized that a portion of collisions were due to contamination from ambient RNA released by damaged cells. Using DecontX (20), a hierarchical Bayesian method that assumes the observed transcript counts of a cell is a mixture of counts from two binomial distributions, we removed contaminating transcripts, reducing the collision rate to 0.7% (Fig. 1D and Materials and Methods). After computational decontamination and removal of collision events, we obtained a median of 939 unique molecular identifiers (UMIs) and 439 genes per human cell and 816 UMIs and 336 genes per mouse cell. Mapping each single cell to its originating microwell, we observed a high concordance between the observed and expected cell type proportions along the columns of the wells (Lin's concordance correlation coefficient = 0.91; Fig. 1E and fig. S1B). Together, these results demonstrate that a minimal amount of barcode contamination takes place from single cells in each well and between neighboring wells on the array after pooling, indicating that the XYZeq workflow successfully produces spatially resolved scRNA-seq libraries.

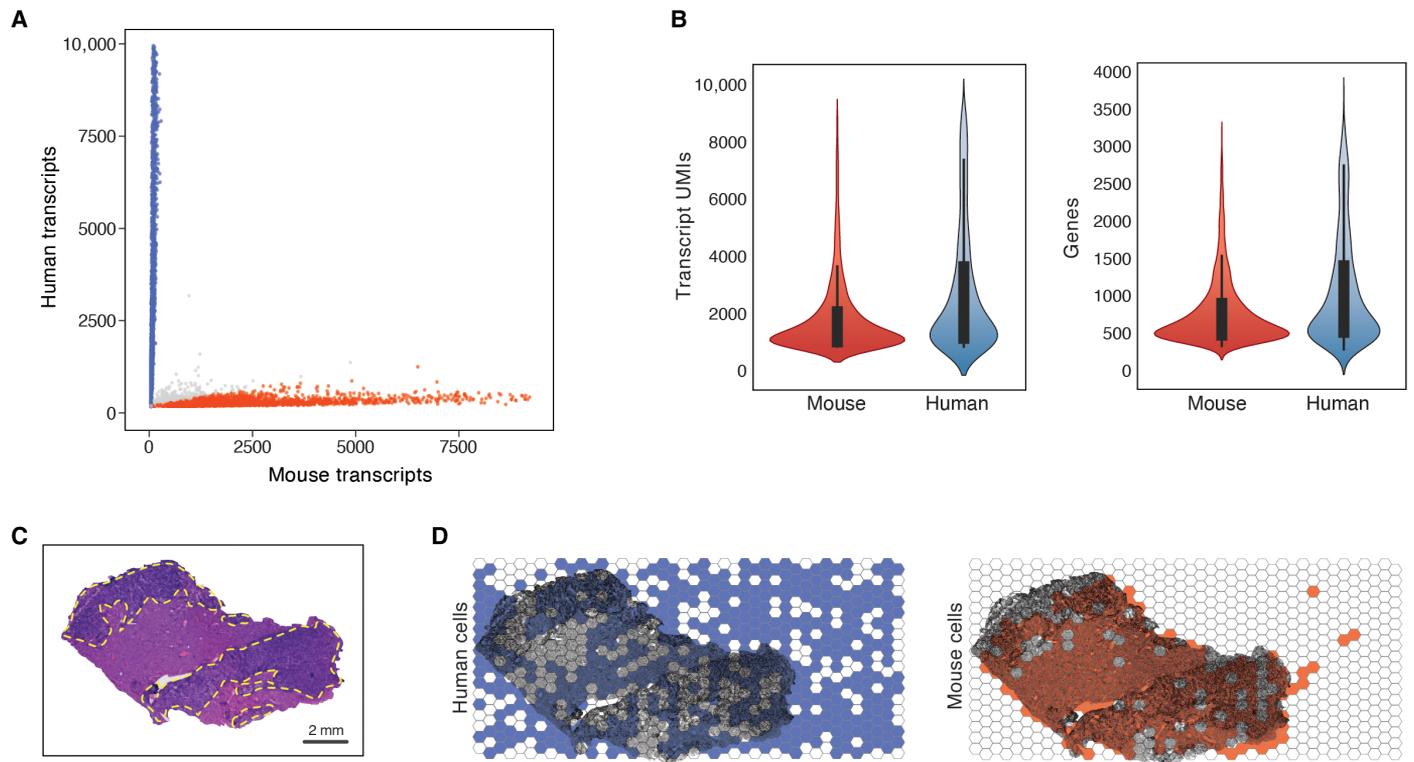
We next applied XYZeq to a fixed and cryopreserved heterotopic murine tumor model established by intrahepatic injections of a syngeneic colon adenocarcinoma cell line, MC38, into immunocompetent mice. This model mimics tissue-infiltrating features of metastatic cancer and is associated with a relatively well-defined tumor boundary (21, 22). MC38 tumor cells also have immunomodulating properties with previous data showing immune cells infiltrating the tumor/tissue interface approximately 10 days after tumor inoculation (23, 24). Thus, we predicted that XYZeq could simultaneously capture the gene expression states and spatial organization of parenchymal liver cells, cancer cells, and tumor-associated immune cell populations. A 25- $\mu\text{m}$  slice of fixed-frozen liver/tumor tissue from a C57BL/6 mouse was placed on top of the prefrozen microwell array while a sequential 10- $\mu\text{m}$  slice was fixed for immunohistochemical staining (fig. S2A and Materials and Methods). We

also deposited fixed human HEK293T cells into the same array at an average of 58 cells per well to serve as a mixed-species internal control to experimentally quantify collision rates. We performed XYZeq and observed an initial collision rate of 7.3% based on comparing the ratio of human versus mouse transcripts (fig. S2B). After computational decontamination and further quality control, which includes filtering cells based on cell counts and mitochondrial expression, the collision rate was reduced to 4.4% (Fig. 2A and Materials and Methods). After removing collisions, we obtained a total of 8746 cells and detected a median of 1596 UMIs and 629 unique genes per HEK293T cell and 1009 UMIs and 456 unique genes per cell from the heterotopic murine tumor model at 46% sequencing saturation (Fig. 2B). A hematoxylin and eosin (H&E)-stained serial section of the tissue revealed a histological boundary between the tumor and adjacent liver/tumor tissue (Fig. 2C). As expected, we observed HEK293T human cells distributed across the entire array, while mouse cells were sequestered within the boundary of the murine tissue (Fig. 2D). Note that empty spatial wells with no cells detected were likely due to a limited number of cells targeted for sequencing ( $\sim 10,000$ ). We obtained a median of 3 human cells per well and 9 mouse cells per well with a total of 13 cells per well expected (fig. S2C).

XYZeq revealed distinct cell types within the murine liver and tumor. Semisupervised Leiden clustering revealed 13 cell populations in the murine tumor model (fig. S3A), from which seven cell types were annotated on the basis of markers that define each population: hepatocytes, cancer cells (MC38), Kupffer cells, liver sinusoidal endothelial cells (LSECs), mesenchymal stem cells (MSCs), lymphocytes, and myeloid cells (Fig. 3A). The annotation of MC38 tumor cells was supported by a high correlation of chromosomal copy numbers estimated from XYZeq scRNA-seq data and publicly available MC38 cytogenetic data (Pearson  $r = 0.78$ ) (25). Notably, a partial amplification of chromosome 15 and a partial deletion of chromosome 14 observed in the XYZeq data were consistent with common chromosomal abnormalities seen in MC38 cells (fig. S3B). As a negative control, we saw low chromosomal copy number correlation when comparing MC38 cells to hepatocytes (26) and immune cells (21) (Pearson  $r = 0.05$  and  $r = 0.17$ , respectively) (fig. S3B). A heatmap showing differentially expressed genes across seven cell types uncovered distinct clusters of cells defined by expression of canonical genes that are relatively exclusive to each cell type (Fig. 3B). Note that we estimated uniformly low rates of contamination of each cell cluster (median under 1%) with the exception of hepatocytes, which had a slightly higher rate at 2.2% (fig. S3C and Materials and Methods). We found comparable median UMIs and genes detected across all cell clusters including immune cell populations that have been difficult to profile using other combinatorial indexing methods (fig. S3, D and E) (27). Cell types expected in non-tumor-bearing liver were identified using markers previously described, which included hepatocytes, Kupffer cells, and LSECs (26). Consistent with the known heterogeneity of hepatocytes, we identified hepatocyte subsets annotated by the expression of pericentral markers (*Glul*, *Oat*, and *Gulo*) (fig. S3F) (26). MC38 adenocarcinoma cells comprised a large uniform cluster and were distinguished by the expression of the known marker *Plec* (22). Myeloid cells were defined by canonical markers *Cd11b* and *Cd74* (28), but other noncanonical markers were also observed, including *Myo1f* (29) and *Tgfb* (30). Lymphocytes showed a similar mix of broad and specific expression patterns of cell type markers, with expression of pan-lymphocyte



**Fig. 1. XYZeq enables single-cell and spatial transcriptome profiling simultaneously.** (A) Schematic of the XYZeq workflow. (B) Schematic of XYZeq sequencing library structure. P5 and P7, Illumina adaptors; bp, base pairs; R1 and R2, annealing sites for Illumina sequencing primers. (C) Schematic representation of the mixed-species cell gradient pattern printed on the chip with 11 unique cell proportion ratios (see Methods for specific cell proportion ratios). (D) Scatterplot of mouse (x axis) and human (y axis) UMI counts detected from a mixture of HEK293T and NIH 3T3 cells after computational decontamination. Blue refers to human cells ( $n = 4182$ ), red refers to mouse cells ( $n = 2220$ ), and gray refers to collisions ( $n = 45$ ). (E) Proportion of HEK293T (blue) cells, NIH 3T3 (red) cells, or collisions (gray) detected by XYZeq for each column of the microwell array.



**Fig. 2. Spatially resolved single-cell transcriptomes captured from tissue.** (A) Scatterplot of mouse (*x* axis) and human (*y* axis) UMI counts detected from liver/tumor tissues ( $n=4$ ) at 500 UMI cutoff after decontamination processing. Blue refers to human cells ( $n=2657$ ), red refers to mouse cells ( $n=5707$ ), and gray refers to collisions ( $n=382$ ). (B) Violin plots showing the number of detected UMIs (left) and genes (right) per mouse (red) and human (blue) cell. Median UMI counts for human cells: 1596; mouse cells: 1009. Median gene counts for human cells: 629; mouse cells: 456 across all liver/tumor slices. (C) H&E-stained image of the liver/tumor tissue slice. Tumor region, dark purple with yellow dotted outlines; liver region, pink. Scale bar, 2 mm. (D) Visualization of human (blue) and mouse (red) cell distribution on the XYZeq array overlaid on the H&E-stained slice.

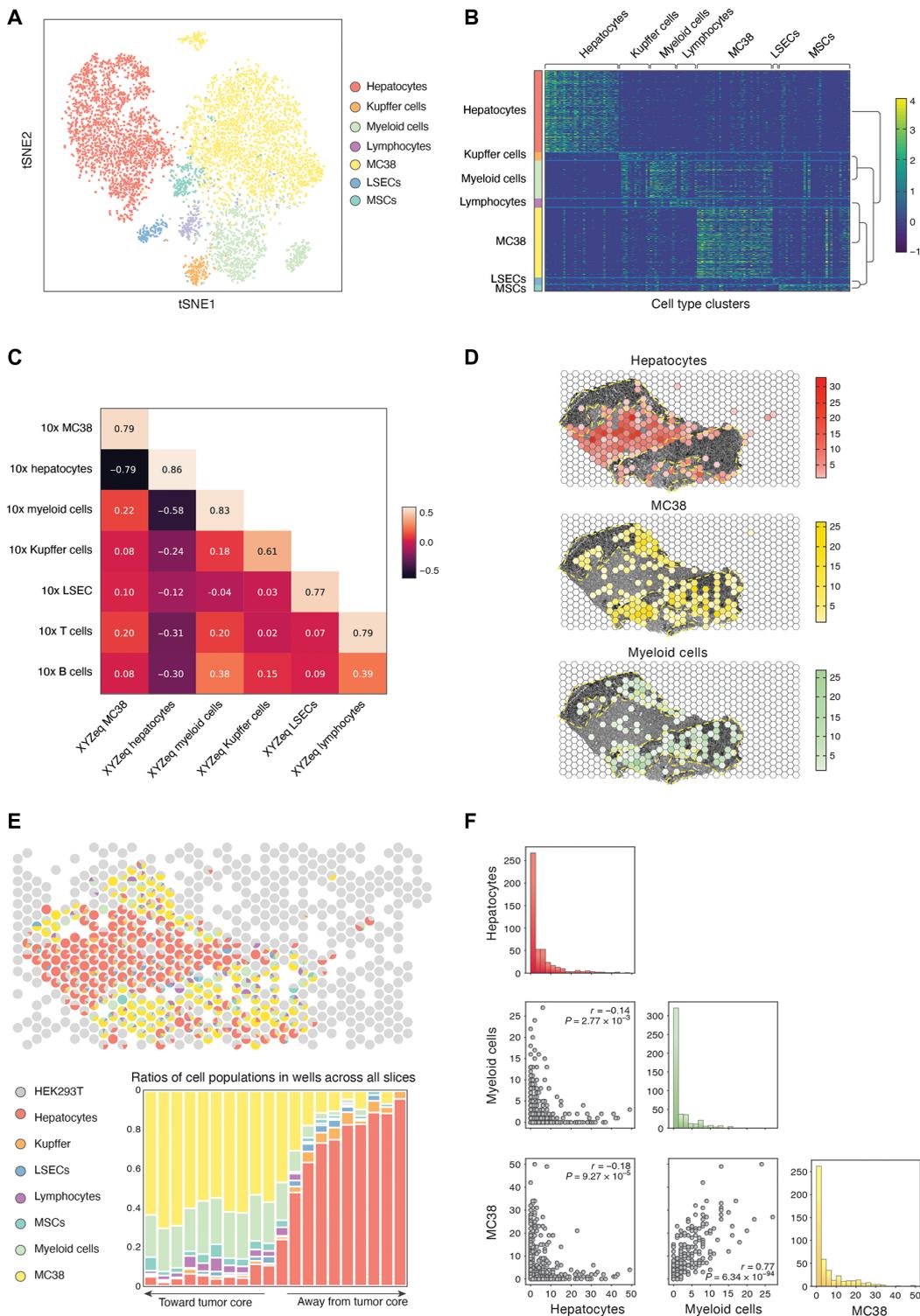
marker *Ili18r1*, T lymphocyte marker *Prkcq*, and cytotoxic T cell marker *Cd8b* (31–33). Last, we detected a cluster of MSCs/stromal cells that expressed both broad mesenchymal cell markers *Rbms3* and *Tshz2* and stem/stromal cell markers *Prkg1* and *Gpc6* (fig. S3F) (34–38).

We next assessed the reproducibility of XYZeq while comparing changes in the transcriptional landscape across the *z*-layer of the organ. Four nonsequential 25- $\mu$ m tissue slices from the same frozen liver/tumor sample block were processed and analyzed. The average expression over all cells for genes detected across all slices was highly correlated between each pair of slices (average pairwise Spearman  $r = 0.93$ ) (fig. S4A). We noted that among the four tissue sections, slices 1 and 2, which were the two most proximal slices in their *z* coordinates (separated by 80  $\mu$ m), had the highest expression correlation (Spearman  $r = 0.96$ ). In contrast, slices 1 and 4, which were the most distal in *z* coordinates (separated by 830  $\mu$ m), had the lowest correlations (Spearman  $r = 0.91$ ). Further, clusters jointly annotated across all four slices consisted of cells from each slice, suggesting that the observed heterogeneity is not due to batch effects (fig. S4B).

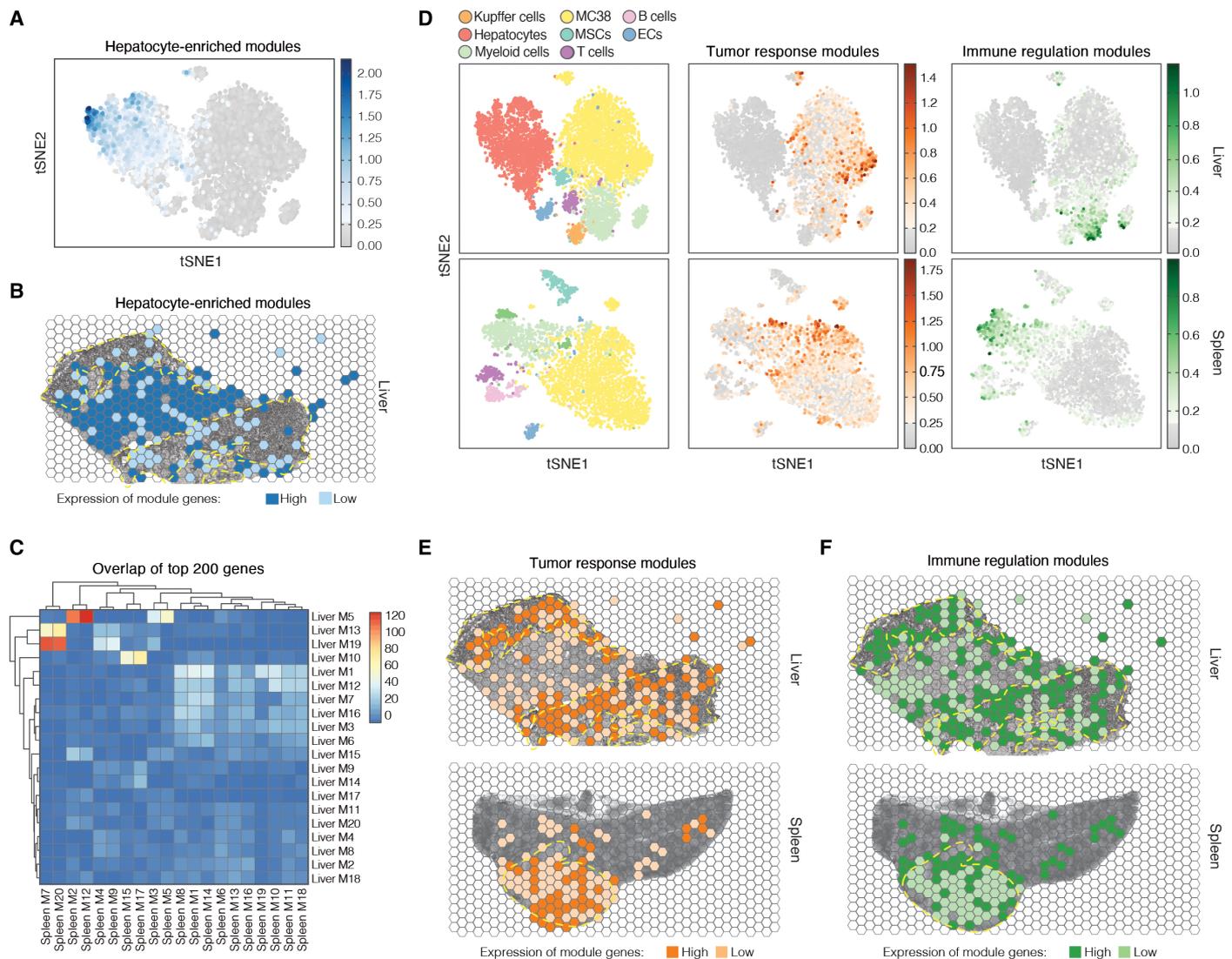
We further compared the quality of the scRNA-seq data generated by XYZeq to another single-cell technology that is commercially available. To accomplish this, we compared the cell type clusters identified from XYZeq to those identified from an independent scRNA-seq dataset of the same liver/tumor model generated using the 10x Genomics droplet-based Chromium system. Most cell

populations detected by 10x were also observed by XYZeq, except neutrophils, erythroid progenitors, and plasma cells (Fig. 3C and fig. S5A), which are immune cell populations known to be sensitive to the cryopreservation (39) required for XYZeq. 10x did not capture MSCs even though cells were isolated from fresh liver/tumor samples. In addition, B cells identified using the 10x platform correlated with the myeloid population detected by XYZeq, likely due to the transcript capture of *Ly86*, *Cd74*, and several class II histocompatibility antigen genes (e.g., *H2ab1* or *H2dmb1*). For the six cell types identified in both the 10x and XYZeq data, we observed high correlations in both the cell type proportions (Lin's concordance correlation coefficient = 0.99; fig. S5B) and the pseudobulk expression profiles of each cell type (Pearson  $r = 0.64$  to 0.86,  $P < 0.01$ ; Fig. 3C).

Next, we turned to the critical question of whether XYZeq can determine the spatial location of each cell. To do this, we compared the spatial localization of each cell cluster to the images of H&E-stained sequential slices. First, to determine that we could accurately define liver from tumor tissue, we confirmed that the density of hepatocytes and cancer cells across the spatial wells overlap with the histological annotation of the adjacent section (Fig. 3D). Projection of other cell types revealed distinct spatial organization patterns for myeloid cells, lymphocytes, Kupffer cells, MSCs, and LSECs (Fig. 3D and fig. S6A). Quantification of cellular composition occupying each spatial well revealed that MSCs, lymphocytes, and myeloid cells



**Fig. 3. Frequency and spatial mapping of single-cell clusters from tissue.** (A) tSNE visualization of the cell types identified from liver/tumor tissue. A total of 6623 total cells were plotted. (B) Heatmap of scaled marker gene expression and hierarchical clustering of genes that define each cell type from liver/tumor tissue. Reference for color bar in (A). (C) Correlations of pseudobulk expression values for matching cell types between XYZeal and 10x Genomics Chromium. (D) Spatial localization of hepatocytes, MC38, and myeloid cells overlaid on a bright-field image of tissue. Yellow dotted outline indicates tumor regions. (E) Pie chart of cell type composition for each XYZeal well from a representative liver/tumor tissue slice (top) and bar chart illustrating combined cell type composition across all four slices of liver/tumor tissue, which tracks with proximity to the tumor (bottom) (see Methods for proximity score). (F) Pairplot showing the frequency of hepatocytes, MC38, and myeloid cells in each well. Scatterplots show the colocalization of two cell types in each well. Histograms show the distribution of number of cells (x axis) per well (y axis) for each cell type. Pearson correlation ( $r$ ) and  $P$  values are annotated.

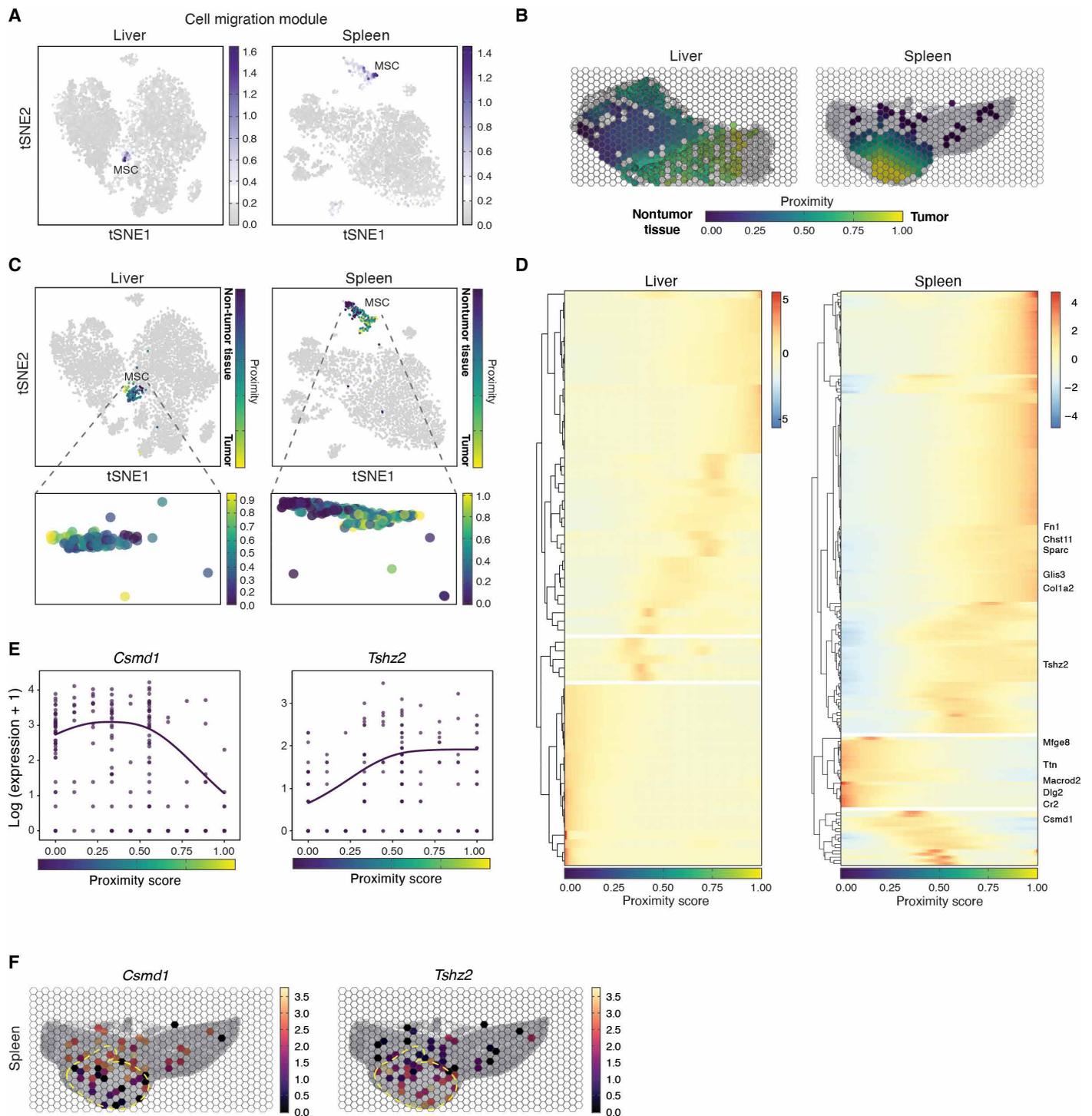


**Fig. 4. Expression of gene modules in space that track with cellular composition.** (A) Projection of average expression of hepatocyte-enriched module (LM14) in tSNE space. Each dot is a cell and colored by the average expression of top contributing module genes (Materials and Methods). (B) Spatial expression of hepatocyte-enriched module (LM14). Each spatial well is colored by the average expression of the top contributing module genes weighted by the number of cells per well. Wells are binarized into high (above weighted average) versus low (all other nonzero expression). Yellow dotted outlines indicate tumor regions. (C) Heatmap representing the number of overlapping genes between each pair of modules in liver/tumor and spleen/tumor. Each row is an LM, and each column is an SM. (D) tSNE projection of XYZeQ scRNA-seq data colored by annotated cell types in liver/tumor (top left) and spleen/tumor (bottom left) and mean gene expression of the top overlapping modules between liver/tumor (top row) and spleen/tumor (bottom row). Tumor response modules correspond to LM5 and SM12, and immune regulation modules correspond to LM19 and SM7. ECs, endothelial cells. Spatial projection visualizes the mean expression of the tumor response modules (E) corresponding to LM5 and SM12 and the immune regulation modules (F) corresponding to LM19 and SM7. Each well in (E) and (F) is colored by the average gene expression of each module weighted by the number of cells per well (high versus low), and yellow dotted outline indicates tumor regions. Wells are binarized into high (above weighted average) versus low (all other nonzero expression).

were colocalized with cancer cells, while Kupffer cells and LSECs colocalized with hepatocytes, suggesting potential regions of cellular interaction in tumor-infiltrated tissue (Fig. 3E and Materials and Methods). These qualitative observations were confirmed by pairwise correlation analysis of cell type proportion across all the wells ( $0.37 \leq \text{Pearson } r \leq 0.77, P < 0.05$ ; Fig. 3F and fig. S6B).

To assess the generalizability of XYZeQ to other tissues, we processed samples from the same heterotopic murine tumor model in the spleen. We recovered a total of 7505 cells at a median of 1312

UMIs and 661 unique genes per HEK293T cell and 1169 UMIs and 577 unique genes per mouse cell at an estimated collision rate of 1.36% (fig. S7, A and B). Similar to the liver/tumor model, XYZeQ was able to reconstruct the boundaries of the splenic mouse tissue with the MC38 tumor region annotated on a sequential H&E-stained slice (fig. S7, C to E). A median of four human cells per well and seven mouse cells per well were detected (fig. S7F). Semisupervised Leiden clustering revealed six distinct cell populations for the spleen/tumor model including B cells, T cells, myeloid cells, MSCs,



**Fig. 5. Differential gene expression within MSCs associated with their spatial proximity to tumor.** (A) Average expression of the cell migration modules (LM10 and SM17) in tSNE space. Each dot is a cell colored by its mean expression of the top module genes between corresponding liver/tumor and spleen/tumor modules. (B) XYSeq array colored by the tumor proximity score. Values near 1 (yellow) indicate regions rich in tumor, values near 0 (purple) indicate regions rich in nontumor cells, and wells capturing the border between the two tissue types take on values around 0.5 (blue/green). (C) MSCs colored by the cell-specific proximity score in tSNE space. (D) Row-clustered heatmap showing the scaled, mean gene expression in MSCs of genes enriched in three spatial regions (intratumor, boundary, and intratissue) along the one-dimensional proximity score. For spleen/tumor, statistically significant genes enriched in the tumor and nontumor regions are highlighted. (E) Log expression (y axis) of *Csmd1* (left) and *Tshz2* (right) along the proximity score (x axis). Each dot corresponds to one MSC cell, and the regression line is fitted using the negative binomial distribution (Materials and Methods). (F) Projection in space of mean expression of *Csmd1* (left) and *Tshz2* (right) in MSCs. Yellow dotted outline indicates tumor region.

endothelial cells, and MC38 tumor cells (fig. S8A). We observed that all four spleen/tumor slices contributed to each cell type cluster, suggesting that the annotated clusters are not due to batch effects (fig. S8B). A heatmap showing differentially expressed genes across the six cell types revealed distinct clusters of cells expressing canonical genes that are relatively exclusive to each type (fig. S8C). Cells from each type could be spatially mapped across the tissue (fig. S8D). Collectively, these results demonstrate that XYZeq can generate spatially resolved scRNA-seq data from different fixed-frozen tissues.

The ability to obtain spatial and single-cell transcriptomic data simultaneously allowed us to assess the effects of cellular composition on gene expression patterns across space. We applied non-negative matrix factorization (NMF) to both the liver/tumor and spleen/tumor scRNA-seq data to define modules of coexpressed genes and associated the expression of each module in each cell type with its expression across spatial wells. Using our approach, we identified 20 modules of coexpressed genes in each tissue (Materials and Methods). As a proof of principle of the approach, we first identified liver module (LM) 14 from the liver/tumor data, which was predominantly expressed by the hepatocyte cluster in the t-distributed stochastic neighbor embedding (tSNE) space (Fig. 4A). As expected, the highest LM14-expressing wells were enriched for hepatocytes, suggesting that the spatial variability of this module is largely driven by the frequency of hepatocytes (Fig. 4B).

Next, we reasoned that because both the liver and spleen were injected with the same tumor cell line, the invading tumors may induce a shared gene expression profile that vary over space, driven in part by the cellular composition of the tumor microenvironment. To test this hypothesis, we first identified pairs of matching gene modules between the two tissues from the NMF analysis (Materials and Methods). We found four distinct LMs that had at least 25% of genes overlapping with spleen/tumor modules (SMs) (Fig. 4C and fig. S9A). Gene ontology analysis of the modules revealed the enrichment of genes implicated in tumor response, immune regulation, and cell migration (figs. S9, B and C, and S10B). Consistent with the enrichment analysis, many of the genes from these modules have been implicated in tumorigenesis (complete gene lists are in table S1). Unlike LM14, further analysis of these matching modules revealed a heterogeneous composition of cell populations that contributed to the expression of specific module genes (fig. S9D and Materials and Methods). For example, the tumor response module LM5 and its matching modules SM2 and SM12 (Fig. 4C and fig. S9A) consisted of genes predominantly expressed in MC38 tumor cells with some expression in myeloid cells and lymphocytes (Fig. 4D, fig. S9D, and Materials and Methods). The immune regulation modules, LM13 and LM19 (matched with SM7 and SM20), consisted of genes expressed primarily in both conventional (e.g., myeloid and lymphocytes) and nonconventional (e.g., Kupffer cells from liver samples) immune cells (Fig. 4, C and D, and fig. S9D). The expression of these overlapping modules was highest in regions densely infiltrated with cancer cells (Fig. 4, E and F). Collectively, these results show that the joint analysis of scRNA-seq and spatial metadata from XYZeq can identify spatially variable gene modules due to differences in cellular composition across tissue samples.

We next focused our analysis on matching modules LM10 and SM15/SM17, which are primarily expressed by MSCs and enriched for genes involved in cell migration (Figs. 4C and 5A and figs. S9D and S10, A and B). Because MSCs are known to have homing

abilities to injured or inflamed sites (40), we hypothesized that LM10 could be differentially expressed in MSCs based on their proximity to the tumor. To test this hypothesis, we first computed a tumor proximity score for each well based on the composition of and distance from nearby wells (Fig. 5B; see Materials and Methods and fig. S11 for score definition). Projecting the proximity score onto MSCs in tSNE space revealed that the transcriptional heterogeneity of the population is associated with spatial proximity to tumor (Fig. 5C). We then analyzed the MSC expression profiles using tradeSeq (41) to identify differentially expressed genes that tracked with the proximity score. We identified and clustered 177 genes from the liver/tumor tissue ( $P < 0.05$ ) and 66 genes from the spleen/tumor tissue ( $P < 0.05$ ) that are associated with the continuous, one-dimensional proximity score (Fig. 5D). The genes were broadly divided into three groups based on the proximity cells to tumor: intratumor, tumor-tissue boundary, and intratissue with statistically significant genes highlighted for the spleen/tumor tissue (Benjamini-Hochberg false discovery rate  $< 0.05$ ) (Fig. 5D). For MSCs found in the intratumor regions of the spleen/tumor, many of the differentially expressed genes are reported to regulate the extracellular matrix (ECM) (Fig. 5D, right) (42–45), suggesting that MC38 cells may induce a local gene expression program in neighboring MSCs that could contribute to malignant remodeling of the ECM.

Last, we leveraged the scRNA-seq data from XYZeq to visualize how individual MSCs expressed *Tshz2* and *Csmd1*, two genes of divergent function that are spatially variable with respect to the tumor in the spleen. Both genes are characterized as tumor suppressor genes and are often silenced in cancer cells to promote malignant growth and metastasis (36, 46, 47). However, we found that spleen/tumor MSCs expressed lower levels of *Csmd1* but higher levels of *Tshz2* in closer proximity to the tumor (Fig. 5E). The mean differential expression of these genes was specific to splenic MSCs and not expressed by MC38 tumor cells. The expression pattern of each of these genes in space revealed a pattern consistent with the aforementioned spatial trajectory analysis, suggesting that their heterogeneous expression in MSCs may be determined by the location of the cells with respect to tumor (Fig. 5F). Together, these results reveal that joint analysis of spatial and single-cell transcriptomic data from XYZeq can detect transcriptionally variable genes within specific cell types (e.g., MSCs) driven by their location within the complex tissue architecture.

## DISCUSSION

We introduce XYZeq, a new scRNA-seq workflow that encodes spatial meta information at 500- $\mu$ m resolution. XYZeq enables unbiased single-cell transcriptomic analysis to capture the full spectrum of cell types and states while simultaneously placing each cell within the spatial context of complex tissue. In murine tumor models, we demonstrate that XYZeq identifies both spatially variable patterns of gene expression determined by cellular composition and heterogeneity within a cell type determined by spatial proximity. Looking forward, XYZeq provides a scalable workflow that can be adapted to multiple z-layers of tissue and can potentially facilitate analysis of entire organs. Large-scale integrated profiling of multiple modalities of single cells mapped to the structural features of their tissue will enable greater understanding of how the tissue microenvironment affects cellular infiltration and interaction in health and disease.

**MATERIALS AND METHODS****Mice, tumor cell line, and tumor inoculation**

Six- to 12-week-old C57BL/6 female mice were purchased from Jackson Laboratories and housed under specific pathogen-free conditions. MC38 colon adenocarcinoma cell line expressing luciferase was a gift from R. D. Beauchamp (Vanderbilt University). MC38 cell line was cultured in complete cell culture medium (RPMI 1640 with GlutaMAX, penicillin, streptomycin, sodium pyruvate, Hepes, non-essential amino acid, and 10% fetal bovine serum). Cell lines were routinely tested for mycoplasma contamination. For experiments, mice were given an anesthetic cocktail of buprenorphine (300  $\mu$ l) and meloxicam (300  $\mu$ l) 30 min before the procedure. At the time of surgery, one drop of bupivacaine was administered, and mice were anesthetized with isoflurane before intrahepatic (or intrasplenic) injection of MC38 colon adenocarcinoma cells (50  $\mu$ l at  $10 \times 10^6$  cells/ml) using a 30-gauge  $1/2$ -inch needle. Incision was stapled closed, and postoperative care was given to the mice. All experiments were conducted in accordance with the animal protocol approved by the University of California, San Francisco Institutional Animal Care and Use Committee.

**Cancer model system**

The intrahepatic and intrasplenic cancer model that we used for the paper is described in great detail in a recently published report by Lee *et al.* (21). Briefly, intrahepatic and intrasplenic tumors were generated by subcapsular injection of the tumor cells directly into the organs. To establish the ideal time point for sacrificing the mice, *in vivo* imaging was done on tumor-inoculated mice. Intraorgan-injected MC38 cells were modified to express the firefly luciferase. Mice were intraperitoneally infected with D-luciferin (150 mg/kg; Gold Biotechnology) 7 min before imaging with the Xenogen IVIS Imaging System. Mice with detectable tumor nodules with at least 5-mm fluorescence were euthanized for tissue harvesting. Organs to be used for XYZeq were fixed with DSP (Thermo Fisher Scientific) and cryopreserved, while organs used for 10x Genomics Chromium single-cell sequencing were digested in RPMI 1640 complete medium that were supplemented with collagenase D (125 U/ml; Roche) and deoxyribonuclease I (20 mg/ml; Roche) and then processed for single-cell suspension using the gentleMACS tissue dissociator per the manufacturer's protocol (Miltenyi).

**10x Genomics Chromium platform**

Cells isolated from tissue were washed and resuspended in phosphate-buffered saline with 0.04% bovine serum albumin at 1000 cells/ $\mu$ l and loaded on the 10x Genomics Chromium platform per the manufacturer's instructions and sequenced on NovaSeq or HiSeq 4000 (Illumina).

**Tissue harvesting and cryopreservation**

At day 10 after tumor inoculation, mice were euthanized and harvested for the tumor-injected liver (or spleen) and incubated for 30 min in ice-cold dimethyl sulfoxide-free freezing media (Bulldog Bio). This was followed by 30 min of incubation in ice-cold DSP (Thermo Fisher Scientific) supplemented with 10% fetal calf serum (FCS) and then neutralized in ice-cold 20 mM tris-HCl (pH 7.5). The organs were placed in a cryomold, sealed airtight, and slowly frozen overnight at  $-80^{\circ}\text{C}$ .

**Cells and reagent dispensing into array**

The sciFLEXARRAYER S3 (Sciencion AG) was used to dispense cells and reagents to the microwell arrays. Drop stability and array

quality were assessed for each experiment. Before dispensing into the microwell arrays slides, Autodrop detection was used to assess drop stability and quantify the velocity, deviations, and drop volume for each reagent. Volume entry was used to determine the number of drops required to reach the total designated well volume. Each well oligo(dT) primer (5'-CTACACGACGCTCTCCGATCTNNNNNNNN[16-base pair unique spatial barcode] TTTT-TTTTTTTTTTTTTT-3', where "N" is any base; IDT) was spotted into a different well in the array. During barcoding, the dew point control software monitored the ambient temperature and humidity, allowing dynamic control of the temperature of the source plate to maintain nominal oligo concentrations through the duration of the run. Barcoded slides were dried in the wells before storage. Reaction mix (Thermo Fisher Scientific) was added to wells and automated with a 10% bleach wash between each probe to eliminate carryover contamination. Dissociation/permeabilization buffer was printed into each well on the day of experiment, and tissue section was loaded onto the microwell array slides. For all tissue experiments, DSP-fixed HEK293T cells were added at 5  $\mu$ l (at  $10 \times 10^6$  cells/ml) to the RT digestion mix before being dispensed across all the wells in the microarray. The average number of HEK293T cells were 58 cells per well; however, the absolute number of cells per well likely varied across the array due to the cells being in suspension inside the dispensing nozzle. Cells harvested from the array after incubation were analyzed on the Aria (BD Biosciences), and datasets were analyzed using FlowJo software (Tree Star Inc.).

**Array fabrication**

Photoresist masters are created by spinning on a layer of photoresist SU-8 2150 (Thermo Fisher Scientific) onto a 3-inch silicon wafer (University Wafer) at 1500 rpm and then soft baking at  $95^{\circ}\text{C}$  for 2 hours. Then, photoresist-layered silicon wafer is exposed to ultraviolet (UV) light for 30 min over a photolithography mask (CAD/Art Sciences, USA) that was printed at 12,000 DPI (dots per inch). After UV exposure, the wafers are hard-baked at  $95^{\circ}\text{C}$  for 20 min and then developed for 2 hours in fresh solution of propylene glycol monomethyl ether acetate (Sigma-Aldrich) to develop, followed by a manual rinse with fresh propylene glycol monomethyl ether acetate then baked at  $95^{\circ}\text{C}$  for 2 min to remove residual solvent. Polydimethylsiloxane (PDMS) mixture (Sylgard 184, Dow Corning, Midland) with prepolymer:curing agent ratio of 10:1 was poured over the SU-8 silicon wafer master. This was placed in a 100-mm petri dish and was cured overnight in a  $70^{\circ}\text{C}$  oven. This PDMS-negative mold was peeled off the SU-8 silicon master the following day. PDMS block was placed on a flat surface, and Norland Optical Adhesive 81 (NOA81) (Thorlabs) was poured into the mold to cover the entire surface. A slide was placed on top of the NOA-poured PDMS mold, and a transparent weight was placed on top. NOA was cured for 2 min under UV light, flipping once halfway through the UV curing time. Last, the PDMS mold was detached from the cured NOA microwell array slide (referred to as microwell array chips). The dimensions of each hexagonal well are approximately 400  $\mu$ m in height and 500  $\mu$ m in diameter with the volume of 0.04  $\text{mm}^3$ , which can hold 40 nl of liquid.

**XYZeq methodology**

Liver/tumor organ was mounted on a cyrostat (Leica) and sliced at 25  $\mu$ m for use as an XYZeq experimental sample or mounted on a histology slide at 10  $\mu$ m for immunohistochemical staining. On the day of experiment, XYZeq microwell array chips were spotted with

an RT cocktail mix that was spiked-in with DSP-fixed HEK293T cells. The microwell array chips were brought down to  $-80^{\circ}\text{C}$ , and a tissue slice was placed on top of the array. A digital image was taken to document the orientation of the tissue before sandwiching a silicone gasket sheet between the XYZeq microwell array chip and a blank histology slide. The chip was placed in a microarray hybridization chamber (Agilent) to ensure an airtight seal while undergoing tissue digestion and RT. To recover high-quality RNA from fixed-frozen tissue, the microarray hybridization chamber housing the chip had to undergo a gradual step-wise temperature increase to  $42^{\circ}\text{C}$  before the 20-min incubation to undergo RT. The chip was removed from the chamber and placed in a 50-ml conical tube with 50 ml of  $1\times$  SSC buffer and 25% FCS. The tube was vortexed and spun down at 1000 rcf for 10 min. Excess volume was removed, and cells were filtered and stained for DAPI (4',6-diamidino-2-phenylindole; Life Technologies) before sorting (BD Aria) into 96-well plates pre-loaded with  $5\ \mu\text{l}$  of the second RT mix. Plates were reverse-transcribed for 1.5 hours at  $42^{\circ}\text{C}$ , followed by PCR using  $2\times$  Kapa HotStart ReadyMix (Kapa Biosystems). PCR amplification was performed with an indexing primer (5'-AATGATACGGCGACCACCGA-GATCTACAC [15]ACACTCTTCCCT ACACGACGCTCTTC-CGATCT-3'; IDT). Contents of the PCR plate were pooled into 2-ml Eppendorf tubes, and complementary DNA (cDNA) was purified with AMPure XP SPRI bead (Beckman). cDNA was tagged and amplified with Illumina Nextera library p7 index (IDT). Final library was analyzed by BioAnalyzer (Agilent) and quantified by Qubit (Invitrogen) and sequenced on a NovaSeq or HiSeq 4000 (Illumina) (read 1: 26 cycles; read 2: 98 cycles; index 1: 8 cycles; index 2: 8 cycles).

### XYZeq decontamination analysis

In our analysis, we recognized that some reads aligning to the mouse genes were present in cells that otherwise had high alignment to the human genome. We suspected that these reads were ambient RNA contamination and sought to remove them. We first removed mouse-aligned transcripts with an extremely high expression in human cell population [ $n = 59$ ,  $\log(\text{counts} + 1) > 6$ ]. The human cell population was considered a control in the contamination detection, because any ambient RNA from lysed cells was expected to contaminate both mouse and human cells. DecontX (20) was then performed to estimate the contamination rate for different cell populations using the human-mouse mixture dataset and therefore derive a decontaminated count matrix from the raw data. Briefly, the algorithm applies variational inference to model the observed counts of each cell as a mixture of true gene expression of its corresponding cell population and the contamination signature (from other cell populations) and then subtracts the contamination signature (fig. S3C). By considering the human-mouse mixed-species experiment, we could remove those counts potentially contributing to collision and effectively account for all potential transcripts in the lysed cells that contribute to ambient RNA. In fig. S3C, the initial estimated contamination rate for each mouse cell type is plotted with the median estimates ranging from 0.06 to 0.31%, with the highest seen in the hepatocyte cell cluster with 2.18% initial contamination fraction. All the downstream analysis was performed on the basis of the decontaminated data after contamination removal.

### How distinctions were made between collision rate and contamination rate

The collision rate is directly calculated from the gene expression of human-mouse mixture dataset based on the ratio between

mouse-aligned and human-aligned transcripts, while the contamination rate for each cell is estimated as a cell-specific parameter in the Bayesian hierarchical model via variational inference from DecontX. To specify the contamination rate, each cell has a beta-distributed parameter modeling its proportion of transcript counts, which come from its native expression distribution. The estimated contamination rate for each cell is the proportion of transcript counts, which come from contamination in the Bayesian model. Each transcript in a cell follows a multinomial distribution parameterized by the native expression distribution of its cell population or contamination from all the other cell populations, given a Bernoulli hidden state, indicating whether the transcript comes from its native expression distribution or from the contamination distribution.

### Cell species mixing experiment

Mixture of HEK293T and NIH 3T3 cells were deposited into wells in a gradient pattern across the columns of the array with a total of 11 distinctive cell proportion ratios. Specifically, columns on the array were spotted with human cell-to-mouse cell ratio of  $100/0$ ;  $90/10$ ;  $80/20$ ;  $70/30$ ;  $60/40$ ;  $50/50$ ;  $40/60$ ;  $30/70$ ;  $20/80$ ;  $10/90$ ;  $0/100$ ;  $10/90$ ;  $20/80$ ;  $30/70$ ;  $40/60$ ;  $50/50$ ;  $60/40$ ;  $70/30$ ;  $80/20$ ;  $90/10$ ; and  $100/0$ , with only human cells flanking the end columns and only mouse cells in the center columns. The ratio of UMI deduplicated reads aligning to either human or mouse reference genomes was calculated for each cell, and those with less than 66% aligning to a single species were deemed barcode collision cells.

### XYZeq single-cell analysis

Single-cell RNA sequence data processing was performed where sequencing reads were processed as previously described (17). Briefly, raw base calls were converted to FASTQ files and demultiplexed on the second combinatorial index using bcl2fastq v2.20. Reads were trimmed using trim galore v0.6.5, aligned to a mixed human (GRCh38) mouse (mm10) reference genome and UMI deduplicated. Reads were then assigned to single cells by demultiplexing on the first combinatorial index, before the construction of a gene by cell count matrix. The count matrix was processed using the Scanpy toolkit. Cells with less than 500 UMIs and greater than 10,000 UMIs, as well as cells expressing less than 100 unique genes or more than 15,000, were discarded. Cells with more than 1% mitochondrial read percentage were also discarded. Gene counts were normalized to 10,000 per cell, log-transformed, and further filtered for high mean expression and high dispersion using the filter genes dispersion function, with a minimum mean of 0.35, maximum mean of 7, and minimum dispersion of 1. Gene counts were then corrected using the regress out function with total counts per cell and the percentage mitochondrial UMIs per cell as covariates. Subsequent dimensionality reduction was done by scaling the gene counts to a mean of 0 and unit variance, followed by principal components analysis, computing of a neighborhood graph, and tSNE. Leiden clustering was performed with a resolution of 0.8, and cells were grouped to reveal distinct murine cell types and human HEK293T cells.

### 10x data processing

Count matrices were generated using the “count” tool from Cell Ranger version 3.1.0, using the combined human and mouse reference dataset (version 3.1.0) and the “chemistry” flag set to “five-prime.” The count matrix was processed using the Scanpy toolkit. Cells with less than 500 UMIs and greater than 75,000 UMIs, as well

as cell expressing less than 100 unique genes and greater than 10,000, were discarded. Cells with more than 7.5% mitochondrial read percentage were also discarded. Gene counts were normalized to 10,000 per cell, log-transformed, and further filtered for high mean expression and high dispersion using the filter genes dispersion function, with a minimum mean of 0.2, maximum mean of 7, and minimum dispersion of 1. Gene counts were then corrected using the regress out function with total counts per cell and the percentage mitochondrial UMIs per cell as covariates. Subsequent dimensionality reduction was done by scaling the gene counts to a mean of 0 and unit variance, followed by principal components analysis, computing of a neighborhood graph, and tSNE. Leiden clustering was performed with a resolution of 1, and cells were grouped to reveal major murine cell types and human HEK293T cells.

### Heatmap for XYZeq

Mouse cells were subsetted from the XYZeq processed data matrix. The processed gene expression values were plotted in a heatmap with a minimum fold change of 1.5 and hierarchically clustered using the heatmap function from Scanpy, with the default settings of Pearson correlation method and complete linkage.

### XYZeq gene pairplot

Four slices of liver/tumor tissue were processed using the XYZeq assay (with HEK293T cells spiked in) and aligned to a joint human and mouse reference. All genes with at least one count in each slice were kept, and the counts across the common set of genes between pairwise slices were plotted in the lower triangle, with the Spearman correlation for the data shown in the upper triangle. Along the diagonal, histograms were plotted showing the distribution of counts per gene for all the nonzero genes for each slice.

### XYZeq cells per well pairplot

Pairplot shows the number of microwells containing pairwise combinations of cell types. For scatterplots, each point in the plot represents a well, and its coordinate positions indicate the number of cells of each cell type present in that well. Every dot on the scatterplot is a gene representing mean per gene for common genes across all cells in the slices. Along the diagonal of the figure are histograms, showing the univariate distribution of cell number per well for the given cell type.

### Heatmap comparing 10x to XYZeq

Mouse cells were subsetted from each of the processed data matrices. For pairwise mouse Leiden clusters found between XYZeq and 10x, the scaled and log-transformed gene expression values of common genes were plotted. For each comparison, a Pearson correlation was calculated and plotted in the heatmap. Row/column labels were ordered according to their corresponding cell types.

### Correlation plot

Mouse cells were subsetted from each of the processed data matrices. Proportions for each cell type (as determined by the Leiden clustering and visualized using tSNE) were plotted, and the coefficient of determination was calculated by fitting to the model that assumes proportions are equal between the two assays.

### Gene module analysis of top contributing genes

To identify gene modules using NMF, genes expressed in fewer than five cells and cells expressing fewer than 100 genes were filtered out.

Variance stabilizing transformation was performed on count data, and confounding covariates including number of counts per cell, batch, and mitochondrial read percentage were regressed out by a regularized negative binomial regression model using the SCTransform (48) function in the Seurat R package. Pearson residual values from the regression model were centered, and all negative values were converted to zero. Nonsmooth NMF (nsNMF) was performed on the resulting expression data with a rank value of 20 using the nmf (49) function in NMF R package. In each module, genes were sorted by their magnitude in the corresponding coefficient matrix in a descending order. Gene ontology enrichment analysis was performed for the sorted genes in each module using GOrilla (50). For each module, the top consecutive genes with higher coefficients in this module compared to all the other modules were further selected as genes contributing the most to the module (51) in the tissue-specific analysis. Binary spatial plots were generated by first calculating the median expression across all the cells for each well within each batch based on the log-normalized gene expression data. We then extracted the mean expression across all the genes within one module for each well and calculated the average of mean expression across selected module genes for each well weighted by the number of cells in each well. The wells with a mean expression across genes above the weighted average were labeled as highly expressing for that gene module, and all the other wells with nonzero expression of those selected module genes were labeled as lowly expressing that gene module. tSNE plots representing the gene modules were colored by their mean expression of genes within the annotated module.

### Overlapping analysis between the gene modules identified in liver/tumor and spleen/tumor

Gene modules were first identified using nsNMF with a rank value of 20 for the two tissues, liver/tumor and spleen/tumor, respectively. The top 200 genes in each sorted gene list for a module were selected as having high association with the module. For each module in the liver/tumor tissue, the spleen/tumor module with the largest gene overlap was initially matched as functionally similar. We then removed those matched pairs with fewer than 25% overlapping genes out of top 200 genes in the liver/tumor module. To calculate cell type fractions that make up each module, the average gene expression for each gene across all the cells was calculated. Median expression across all the overlapping genes for each cell type was further computed, which was later transformed into fractions by dividing by the sum of median expression across all the cell types.

### Defining the proximity score by wells

We sought to define a score for each well of the hexagonal well array that would capture how centrally located a well was within either the tumor or nontumor tissue domains. Central to the method was the determination of successive concentric “layers” of wells that were adjacent to a well in question: those corresponding to its immediate neighbors (layer 1), those wells exactly two wells away (layer 2), and so on, for  $n$  layers. In the spleen/tumor, we selected several wells on the far side of the tumor region and set the score of these wells to 1. We then took 10 successive layers of wells and decreased the score linearly with each layer, with the wells in layers 10 and beyond set to 0. In the liver, MC38 cells were found in different locations, and therefore, unlike the spleen, there was no single unidirectional spatial dimension to place all MC38 cells at one end and all nontumor tissue cells at the other. Therefore, we used an alternative approach

to calculate these scores in the liver/tumor tissue. For each well  $w_{x,y}$ , annotated by their  $x, y$  position on the hexagonal well array, we calculated the proportion of hepatocytes,  $p_{x,y}$ , since the hepatocytes were the most abundant parenchymal cell type in and strictly associated with the nontumor liver tissue

$$t_{x,y} = \# \text{ of total hepatocytes and MC38 cells in } w_{x,y}$$

$$h_{x,y} = \# \text{ of hepatocytes in } w_{x,y}$$

$$p_{x,y} = \frac{h_{x,y}}{t_{x,y}}$$

Then, for each well in question  $w_{x,y}$ , we tabulated the surrounding wells in each of the successive concentric 10 layers. We denote these wells  $w_{x',y'}$  to differentiate from the well in question. For each of those layers  $l$ , we took its constituent wells'  $p_{x',y'}$  and calculated a cell number-weighted average  $p_{x,y,l}$

$$w_{x,y,l} = \{w_{x',y'} \in \text{layer } l \text{ of } w_{x,y}\}$$

$$t_{x,y,l} = \# \text{ of total hepatocytes and MC38 cells in } w_{x,y,l}$$

$$p_{x,y,l} = \sum_{x',y'} \frac{w_{x,y,l} t_{x',y'}}{t_{x,y,l}} p_{x',y'}$$

Then, for the well in question  $w_{x,y}$ , we calculated a distance-weighted average of all the  $p_{x,y,l}$  and this became the proximity score  $s_{x,y}$  for the well in question. The distance weights for each layer,  $u_l$ , were based on an exponential decay, terminated to 10 terms and then normalized to 1 by dividing by the sum of all weights  $u_s$ . We give equal weight to  $p_{x,y}$  and the value for the layer 1 neighbors  $p_{x,y,1}$ . A decay factor  $d$  of 1.05 was chosen empirically, as it seemed to create the most uniform-like distribution of the scores across all wells

$$d = 1.05, u_s = \sum_{l=1}^{10} \frac{1}{d^l}$$

$$u_l = \frac{\left(\frac{1}{d^l}\right)}{u_s}$$

$$s_{x,y} = u_1 p_{x,y} + \sum_{l=1}^{10} u_l p_{x,y,l}$$

These calculations were repeated for all wells containing at least one murine cell.

### Trajectory inference analysis

Genes expressed in fewer than five cells and cells expressing fewer than 100 genes were excluded. Variance stabilizing transformation was performed using the SCTransform (48) function in the R Seurat package. The resulting corrected count data in MSC in one tissue were used as the count matrix input in trajectory inference analysis, using the tradeSeq (41) package in R. Genes whose expression is associated with the proximity score were identified by the associationTest function in tradeSeq, based on a Wald test under the negative binomial generalized additive model. The  $P$  values were corrected using

Benjamini-Hochberg multiple testing procedure, and genes with corrected  $P$  values smaller than 0.05 were considered to be significantly associated with the proximity score.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/17/eabg4755/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

1. A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suva, A. Regev, B. E. Bernstein, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
2. S. V. Puram, I. Tirosh, A. S. Parikh, A. P. Patel, K. Yizhak, S. Gillespie, C. Rodman, C. L. Luo, E. A. Mroz, K. S. Emerick, D. G. Deschler, M. A. Varvares, R. Mylvaganam, O. Rozenblatt-Rosen, J. W. Rocco, W. C. Faquin, D. T. Lin, A. Regev, B. E. Bernstein, Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
3. C. Ziegenhain, B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, H. Leonhardt, H. Heyn, I. Hellmann, W. Enard, Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643.e4 (2017).
4. I. C. Macaulay, C. P. Ponting, T. Voet, Single-cell multiomics: Multiple measurements from single cells. *Trends Genet.* **33**, 155–168 (2017).
5. M. L. Suva, I. Tirosh, Single-cell RNA sequencing in cancer: Lessons learned and emerging challenges. *Mol. Cell* **75**, 7–12 (2019).
6. V. Svensson, R. Vento-Tormo, S. A. Teichmann, Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
7. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
8. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
9. C.-H. L. Eng, M. Lawson, Q. Zhu, R. Dries, N. Kouloula, Y. Takeji, J. Yun, C. Cronin, C. Karp, G.-C. Yuan, L. Cai, Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
10. S. Shah, E. Lubeck, W. Zhou, L. Cai, seqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. *Neuron* **94**, 752–758.e1 (2017).
11. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, J. Frisén, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
12. S. G. Rodrigues, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, E. Z. Macosko, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
13. S. Vickovic, G. Eraslan, F. Salmén, J. Klughammer, L. Stenbeck, D. Schapiro, T. Äijö, R. Bonneau, L. Bergenstråhle, J. F. Navarro, J. Gould, G. K. Griffin, Å. Borg, M. Ronaghi, J. Frisén, J. Lundeberg, A. Regev, P. L. Ståhl, High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
14. R. R. Stickels, E. Murray, P. Kumar, J. Li, J. L. Marshall, D. J. Di Bella, P. Arlotta, E. Z. Macosko, F. Chen, Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2020).
15. K. Achim, J. B. Pettit, L. R. Saraiva, D. Gavriouchkina, T. Larsson, D. Arendt, J. C. Marioni, High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
16. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
17. J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure, Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
18. A. B. Rosenberg, C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T. Grayback, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, G. Seelig, Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).

19. M. Attar, E. Sharma, S. Li, C. Bryer, L. Cubitt, J. Broxholme, H. Lockstone, J. Kinchen, A. Simmons, P. Piazza, D. Buck, K. J. Livak, R. Bowden, A practical solution for preserving single cells for RNA sequencing. *Sci. Rep.* **8**, 2151 (2018).
20. S. Yang, S. E. Corbett, Y. Koga, Z. Wang, W. E. Johnson, M. Yajima, J. D. Campbell, Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* **21**, 57 (2020).
21. J. C. Lee, S. Mehdizadeh, J. Smith, A. Young, I. A. Mufazalov, C. T. Mowery, A. Daud, J. A. Bluestone, Regulatory T cell control of systemic immunity and immunotherapy response in liver metastasis. *Sci. Immunol.* **5**, eaba0759 (2020).
22. M. Yadav, S. Jhunjunhwal, Q. T. Phung, P. Lupardus, J. Tanguay, S. Bumbaca, C. Franci, T. K. Cheung, J. Fritsche, T. Weinschenk, Z. Modrusan, I. Mellman, J. R. Lill, L. Delamarre, Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).
23. K. N. Kodumudi, J. Siegel, A. M. Weber, E. Scott, A. A. Sarnaik, S. Pilon-Thomas, Immune checkpoint blockade to improve tumor infiltrating lymphocytes for adoptive cell therapy. *PLOS ONE* **11**, e0153053 (2016).
24. H. Tang, Y. Liang, R. A. Anders, J. M. Taube, X. Qiu, A. Mulgaonkar, X. Liu, S. M. Harrington, J. Guo, Y. Xin, Y. Xiong, K. Nham, W. Silvers, G. Hao, X. Sun, M. Chen, R. Hannan, J. Qiao, H. Dong, H. Peng, Y. X. Fu, PD-L1 on host cells is essential for PD-L1 blockade-mediated tumor regression. *J. Clin. Invest.* **128**, 580–588 (2018).
25. M. Efreanova, D. Rieder, V. Klepsch, P. Charoentong, F. Finotello, H. Hackl, N. Hermann-Kleiter, M. Löwer, G. Baier, A. Krogsdam, Z. Trajanoski, Targeting immune checkpoints potentiates immunoeediting and changes the dynamics of tumor evolution. *Nat. Commun.* **9**, 32 (2018).
26. Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
27. J. Ding, X. Adiconis, S. K. Simmons, M. S. Kowalczyk, C. C. Hession, N. D. Marjanovic, T. K. Hughes, M. H. Wadsworth, T. Burks, L. T. Nguyen, J. Y. H. Kwon, B. Barak, W. Ge, A. J. Kedaigle, S. Carroll, S. Li, N. Hacohen, O. Rozenblatt-Rosen, A. K. Shalek, A.-C. Villani, A. Regev, J. Z. Levin, Systematic comparative analysis of single cell RNA-sequencing methods. *bioRxiv*, 632216 (2019).
28. M. J. C. Jordão, R. Sankowski, S. M. Brendecke, Sagar, G. Locatelli, Y.-H. Tai, T. L. Tay, E. Schramm, S. Armbruster, N. Hagemeyer, O. Groß, D. Mai, Ö. Çiçek, T. Falk, M. Kerschensteiner, D. Grün, M. Prinz, Single-cell profiling identifies myeloid cell subsets with distinct fates during neuroinflammation. *Science* **363**, eaat7554 (2019).
29. S. V. Kim, W. Z. Mehal, X. Dong, V. Heinrich, M. Pypaert, I. Mellman, M. Dembo, M. S. Mooseker, D. Wu, R. A. Flavell, Modulation of cell adhesion and motility in the immune system by Myo1f. *Science* **314**, 136–139 (2006).
30. X. Yu, A. Buttgerit, I. Lelios, S. G. Utz, D. Cansever, B. Becher, M. Greter, The cytokine XGF-beta promotes the development and homeostasis of alveolar macrophages. *Immunity* **47**, 903–912.e4 (2017).
31. H. Helgeland, I. Gabrielsen, H. Akselsen, A. Y. M. Sundaram, S. T. Flåm, B. A. Lie, Transcriptome profiling of human thymic CD4+ and CD8+ T cells compared to primary peripheral T cells. *BMC Genomics* **21**, 350 (2020).
32. O. J. Harrison, N. Srinivasan, J. Pott, C. Schiering, T. Krausgruber, N. E. Iltot, K. J. Maloy, Epithelial-derived IL-18 regulates Th17 cell differentiation and Foxp3(+) Treg cell function in the intestine. *Mucosal Immunol.* **8**, 1226–1236 (2015).
33. N. Isakov, A. Altman, PKC-theta-mediated signal delivery from the TCR/CD28 surface receptors. *Front. Immunol.* **3**, 273 (2012).
34. L. E. Oikari, R. K. Okolicsanyi, A. Qin, C. Yu, L. R. Griffiths, L. M. Haupt, Cell surface heparan sulfate proteoglycans as novel markers of human neural stem cell fate determination. *Stem Cell Res.* **16**, 92–104 (2016).
35. D. Fritz, B. Stefanovic, RNA-binding protein RBMS3 is expressed in activated hepatic stellate cells and liver fibrosis and increases expression of transcription factor Prx1. *J. Mol. Biol.* **371**, 585–595 (2007).
36. M. Riku, S. Inaguma, H. Ito, T. Tsunoda, H. Ikeda, K. Kasai, Down-regulation of the zinc-finger homeobox protein TSHZ2 releases GLI1 from the nuclear repressor complex to restore its transcriptional activity during mammary tumorigenesis. *Oncotarget* **7**, 5690–5701 (2016).
37. H. Kalyanaraman, N. Schall, R. B. Pilz, Nitric oxide and cyclic GMP functions in bone. *Nitric Oxide* **76**, 62–70 (2018).
38. N. Schall, J. J. Garcia, H. Kalyanaraman, S. P. China, J. J. Lee, R. L. Sah, A. Pfeifer, R. B. Pilz, Protein kinase G1 regulates bone regeneration and rescues diabetic fracture healing. *JCI Insight* **5**, e135355 (2020).
39. J. Baboo, P. Kilbrede, M. Delahaye, S. Milne, F. Fonseca, M. Blanco, J. Meneghel, A. Nancekivill, N. Gaddum, G. J. Morris, The impact of varying cooling and thawing rates on the quality of cryopreserved human peripheral blood T cells. *Sci. Rep.* **9**, 3417 (2019).
40. Q. Wang, T. Li, W. Wu, G. Ding, Interplay between mesenchymal stem cell and tumor and potential application. *Hum. Cell* **33**, 444–458 (2020).
41. K. Van den Berge, H. R. de Bézieux, K. Street, W. Saelens, R. Cannoodt, Y. Saeys, S. Dudoit, L. Clement, Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
42. J. Soikkeli, P. Podlasz, M. Yin, P. Nummela, T. Jahkola, S. Virolainen, L. Krogerus, P. Heikkilä, K. von Smitten, O. Saksela, E. Hölttä, Metastatic outgrowth encompasses COL-1, FN1, and POSTN up-regulation and assembly to fibrillar networks regulating cell adhesion, migration, and growth. *Am. J. Pathol.* **177**, 387–403 (2010).
43. Y. Wang, H. Xu, B. Zhu, Z. Qiu, Z. Lin, Systematic identification of the key candidate genes in breast cancer stroma. *Cell. Mol. Biol. Lett.* **23**, 44 (2018).
44. J. Li, Z. Zeng, X. Jiang, N. Zhang, Y. Gao, Y. Luo, W. Sun, S. Li, J. Ren, Y. Gong, C. Xie, Stromal microenvironment promoted infiltration in esophageal adenocarcinoma and squamous cell carcinoma: A multi-cohort gene-based analysis. *Sci. Rep.* **10**, 18589 (2020).
45. Y. Gao, S. P. Yin, X. S. Xie, D. D. Xu, W. D. Du, The relationship between stromal cell derived SPARC in human gastric cancer tissue and its clinicopathologic significance. *Oncotarget* **8**, 86240–86252 (2017).
46. A. Escudero-Esparza, M. Bartoschek, C. Gialeli, M. Okroj, S. Owen, K. Jirstrom, A. Orimo, W. G. Jiang, K. Pietras, A. M. Blom, Complement inhibitor CSMD1 acts as tumor suppressor in human breast cancer. *Oncotarget* **7**, 76920–76933 (2016).
47. S. Roperio, F. Setien, J. Espada, M. F. Fraga, M. Herranz, J. Asp, M. S. Benassi, A. Franchi, A. Patiño, L. S. Ward, J. Bovee, J. C. Cigudosa, W. Wim, M. Esteller, Epigenetic loss of the familial tumor-suppressor gene exostosin-1 (EXT1) disrupts heparan sulfate synthesis in cancer cells. *Hum. Mol. Genet.* **13**, 2753–2765 (2004).
48. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
49. R. Gaujoux, C. Seoighe, A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
50. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
51. P. Carmona-Saez, R. D. Pascual-Marqui, F. Tirado, J. M. Carazo, A. Pascual-Montano, Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics* **7**, 78 (2006).

**Acknowledgments:** We thank C. Her for isolating hepatocytes; J. Fischer for discussion on differential gene expression analysis; J. Bolen and K. Wong for assistance in immunohistochemical staining; A. Shaw and C. Apostolopolou for tissue-related discussion and imaging; and J. Haliburton, L. Lu, R. Cole, and L. Clark for discussion on soft lithography and microwell array fabrication. We also want to thank Scienion for assistance in customizing software for the sciFLEXARRAYER and setting up nozzles for dispensing. **Funding:** Y.L. was supported by the T32 Ruth L. Kirschstein Fellowship T32AI00733429. C.T.M. is supported by NIH grants T32GM007618, T32DK007418, and F30AI157167 and is a UCSF ImmunoX Computational Immunology Fellow. G.C.H. is supported by NSF Graduation Research Fellow under grant number 1650113. T.L.R. is supported by NIH IF30DK120213-01. D.N.N. is supported by NIH grants L40AI140341 and K08AI153767 and CIRM Alpha Stem Cell Clinic Fellowship. E.D.C. is supported by the NIH grant RM1 HG009490. A.M. and C.J.Y. are supported by NIH grant DP3DK111914 and funding from the Northern California JDRF Center of Excellence. A.M. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund, is an investigator at the Chan Zuckerberg Biohub, and has received funding from the Innovative Genomics Institute (IGI), the American Endowment Foundation, the Cancer Research Institute (CRI), the Lloyd J. Old STAR award, a gift from the Jordan family, a gift from B. Bakar, and is a member of the Parker Institute for Cancer Immunotherapy (PICI). C.J.Y. is further supported by NIH grants R01AR071522, R01AI136972, and R01HG011239 and the Chan Zuckerberg Initiative; is an investigator at the Chan Zuckerberg Biohub; and is a member of the Parker Institute for Cancer Immunotherapy (PICI). Y.W. and Y.S.S. are supported in part by NIH grant R35-GM134922. Y.S.S. is a Chan Zuckerberg Biohub Investigator. **Author contributions:** Y.L., D.B., C.J.Y., and E.D.C. conceived the idea. Y.L. and D.B. fabricated the microwell array slides, developed library preparation pipeline, and completed the development of the method. E.D.C., A.M., C.J.Y., and Y.S.S. supervised the work. H.M.N. and Y.W. carried out differential gene expression analysis under the supervision of Y.S.S. J.M.W. and Y.L. developed the tissue fixing and freezing method. J.C. and D.B. developed the spotting protocol on the sciFLEXARRAYER. Y.W. performed decontamination analysis under the supervision of C.J.Y. and Y.S.S. Y.W. and D.B. performed the gene module analysis and trajectory inference analysis under the supervision of C.J.Y. and Y.S.S. D.B., Y.W., and G.C.H. developed the analytical pipeline and performed all sequencing analysis under the supervision of C.J.Y. and Y.S.S. J.L. and S.M. developed the mouse tumor model and performed the tumor injections. E.S. designed primers. D.N.N. provided input on the design of hexagonal wells. T.L.R. and Y.L. determined evaporation threshold for the method. C.T.M., D.S.L., Y.L., and Y.S. prepared the 10x libraries. Y.L., Y.W., A.M., E.D.C., Y.S.S., and C.J.Y. wrote the manuscript. **Competing interests:** Y.L., D.B., A.M., E.D.C., C.J.Y., and J.M.W. are listed as inventors on a

patent application related to this work filed by the UCSF (no. 62/979,235, filed 20 February 2020). A.M. is a cofounder, member of the Board of Directors, and member of the Scientific Advisory Boards of Spotlight Therapeutics and Arsenal Biosciences. A.M. has served as an advisor to Juno Therapeutics, was a member of the scientific advisory board at PACT Pharma, and was an advisor to Trizell. A.M. has received an honorarium from Merck and Vertex, a consulting fee from AlphaSights, and is an investor in and informal advisor to Offline Ventures. A.M. owns stock in Arsenal Biosciences, Spotlight Therapeutics, and PACT Pharma. The Marson laboratory has received research support from Juno Therapeutics, Epinomics, Sanofi, GlaxoSmithKline, Gilead, and Anthem. C.J.Y. is a Scientific Advisory Board member for and holds equity in Related Sciences and ImmunAI, a consultant for and holds equity in Maze Therapeutics, and a consultant for TReX Bio. C.J.Y. has received research support from Chan Zuckerberg Initiative, Chan Zuckerberg Biohub, and Genentech. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper

and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. Raw and processed sequencing data are available at NCBI Gene Expression Omnibus under accession code GSE164430.

Submitted 8 January 2021

Accepted 4 March 2021

Published 21 April 2021

10.1126/sciadv.abg4755

**Citation:** Y. Lee, D. Bogdanoff, Y. Wang, G. C. Hartoularos, J. M. Woo, C. T. Mowery, H. M. Nisonoff, D. S. Lee, Y. Sun, J. Lee, S. Mehdizadeh, J. Cantlon, E. Shifrut, D. N. Ngyuen, T. L. Roth, Y. S. Song, A. Marson, E. D. Chow, C. J. Ye, XYZeq: Spatially resolved single-cell RNA sequencing reveals expression heterogeneity in the tumor microenvironment. *Sci. Adv.* **7**, eabg4755 (2021).

## XYZeq: Spatially resolved single-cell RNA sequencing reveals expression heterogeneity in the tumor microenvironment

Youjin Lee, Derek Bogdanoff, Yutong Wang, George C. Hartoularos, Jonathan M. Woo, Cody T. Mowery, Hunter M. Nisonoff, David S. Lee, Yang Sun, James Lee, Sadaf Mehdizadeh, Joshua Cantlon, Eric Shifrut, David N. Ngyuen, Theodore L. Roth, Yun S. Song, Alexander Marson, Eric D. Chow and Chun Jimmie Ye

*Sci Adv* 7 (17), eabg4755.  
DOI: 10.1126/sciadv.abg4755

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/7/17/eabg4755>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2021/04/19/7.17.eabg4755.DC1>

### REFERENCES

This article cites 50 articles, 9 of which you can access for free  
<http://advances.sciencemag.org/content/7/17/eabg4755#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

## Supplementary Materials for

### **XYZeq: Spatially resolved single-cell RNA sequencing reveals expression heterogeneity in the tumor microenvironment**

Youjin Lee\*, Derek Bogdanoff, Yutong Wang, George C. Hartoularos, Jonathan M. Woo, Cody T. Mowery, Hunter M. Nisonoff, David S. Lee, Yang Sun, James Lee, Sadaf Mehdizadeh, Joshua Cantlon, Eric Shifrut, David N. Ngyuen, Theodore L. Roth, Yun S. Song, Alexander Marson\*, Eric D. Chow\*, Chun Jimmie Ye\*

\*Corresponding author. Email: [jimmie.ye@ucsf.edu](mailto:jimmie.ye@ucsf.edu) (C.J.Y.); [eric.chow@ucsf.edu](mailto:eric.chow@ucsf.edu) (E.D.C.); [alexander.marson@ucsf.edu](mailto:alexander.marson@ucsf.edu) (A.M.); [youjin.lee@ucsf.edu](mailto:youjin.lee@ucsf.edu) (Y.L.)

Published 21 April 2021, *Sci. Adv.* **7**, abg4755 (2021)  
DOI: 10.1126/sciadv.abg4755

#### **This PDF file includes:**

Figs. S1 to S11  
Table S1

**Figure legends for supplementary figs:**

**Fig. S1: Single cell mixed species experiment reveals strong correlation to estimated cell gradient proportions.**

(A) Scatter plot of mouse (x-axis) and human (y-axis) UMI counts detected from a mixture of HEK293T and NIH3T3 cells. Blue refers to human cells (n=4,389) and red refers to mouse cells (n=1,728) and gray refers to collisions (n=330). (B) Scatter plots revealing high concordance between observed and expected cell type proportions in each column of the XYZeq array. Lin's Concordance Correlation = 0.91.

**Fig. S2: Quantification of cells captured per well from liver/tumor tissue.**

(A) Image of the liver/tumor tissue slice on top of the XYZeq microarray with reagents spotted in wells (frozen: white). Photo credit: Youjin Lee (UCSF). (B) Scatter plot of transcripts (n=4) from human (blue: n=2,667), mouse cells (red: n=6,854), and collisions (gray: n=747). (C) Median cell number in wells across XYZeq array for HEK293T human (top) and liver/tumor mouse (bottom) cells.

**Fig. S3: Distinct cell types clusters identified from XYZeq of liver/tumor tissue.**

(A) t-distributed stochastic neighbor embedding (tSNE) visualization of Leiden cluster to annotated cell types. (B) Correlation of mean chromosomal expression of MC38 cells observed in XYZeq compared to MC38 cells from Efremova, et al. (25), hepatocytes from *Tabula Muris* (26), and immune cells enriched from liver/tumor tissue from an independent internal experiment (21). Both the x-axis and y-axis denotes average

expression of all genes on a given chromosome. (C) Violin plot representing estimated contamination fraction for each cell type from our liver/tumor XYZeq data. (D,E) Violin plot showing the number of detected UMIs and genes per cell cluster. Median UMI counts (log) and gene counts for each cell cluster: hepatocytes (3.04 and 552), Kupffer cells (2.92 and 420), lymphocytes (2.97 and 454), mesenchymal stem cells (MSCs) (3.08 and 594), myeloid cells (3.03 and 511), MC38 (3.22 and 851), and liver sinusoidal endothelial cells (LSECs) (2.94 and 431). (F) Annotated cell-identity clusters. Feature plot of cells that are positive for each individual marker genes to identify hepatocytes (*Cps1*, *Glul*), MC38 (*Plec*), myeloid cells (*Cd11b*, *Cd74*), LSECs (*Stab2*, *Ptprb*), lymphocytes (*Cd8b*, *Ii18r1*), Kupffer cells (*Cd5l*, *Timd4*), MSCs (*Rbms3*, *Tshz2*), pericentral hepatocytes (*Glul*, *Gluo*, *Oat*) from low expression (purple) to high expression (yellow).

**Fig. S4: Reproducibility of XYZeq across tissue slices.**

Four non-sequential z-layer slices of liver/tumor tissue processed with XYZeq (with HEK293T cells spiked-in as control). (A) Pairplot showing the expression of common genes between different slices of liver/tumor. Scatter plots show the UMI counts for common expressed (UMIs > 0) genes. Histograms show the distribution of number of UMIs (x-axis) per gene (y-axis) for each slice. (B) tSNE visualization of Leiden clusters across four liver/tumor slices.

**Fig. S5: Cell type clusters captured from XYZeq found comparable to 10x Genomics platform.**

(A) tSNE representation of liver/tumor tissue data generated with the 10X Chromium V3 kit. 2,703 total cells were plotted. (B) Scatter plot comparing the proportion of each cell type found in XYZeq and 10X Chromium V3. Lin's concordance coefficient = 0.988.

**Fig. S6: Distinct spatial localization pattern across tissue for each cell type cluster.**

(A) Spatial density plot showing localization of lymphocytes (purple), MSCs (green), Kupffer cells (orange) and LSECs (blue) in the spatial array. Yellow dotted outline indicates tumor region. (B) Pairplot showing the frequency of cell types found in each well across the XYZeq array. Scatter plots show the co-localization of the cell types in each well. Histograms show the distribution for number of cells (x-axis) per well (y-axis) for each cell type.  $r$  and  $p$  values annotated.

**Fig. S7: XYZeq of spleen/tumor tissue reveals comparable data quality to liver/tumor tissue.**

(A) Scatter plot of mouse (x-axis) and human (y-axis) UMI counts detected from spleen/tumor tissues ( $n=4$ ). Blue refers to human cells ( $n=4,007$ ) and red refers to mouse cells ( $n=3,394$ ) and gray refers to collisions ( $n=104$ ). (B) Violin plot showing the number of detected UMIs and genes per cell. Median UMI counts for human cells (1,312) and mouse cells (1,169). Median gene counts for human cells (661) and mouse cells (577). (C) Hematoxylin and Eosin (H&E) stained image of the spleen/tumor tissue slice. Tumor region (light purple with yellow dotted outline); spleen region (dark pink with dark purple features). Scale showing 2mm. (D) Image of spleen/tumor tissue on frozen XYZeq

microarray with reagents in wells (white). Photo credit: Youjin Lee (UCSF) (E) Visualization of human (blue) and mouse (red) cell distribution on the XYZeq array with 500 UMI cutoff overlaid on the image of H&E stained tissue slice. (F) Median cell number in wells across XYZeq array for HEK293T human (top) and spleen/tumor mouse (bottom) cells.

**Fig. S8: Identification and spatial mapping of cell type clusters from spleen/tumor tissue.**

(A) tSNE projection of the spleen/tumor XYZeq data. 3,394 total cells were plotted. (B) tSNE visualization of Leiden cluster to annotate cell types for spleen/tumor. (C) Heat map of a scaled expression of marker genes and hierarchical clustering that define each cell type from XYZeq spleen/tumor tissue. (D) Image of spleen/tumor tissue overlaid with spatial plot of the XYZeq array showing localization of cell type clusters from (A) with 500 UMI cutoff. Yellow dotted outline indicates tumor region.

**Fig. S9: Cell type contribution and functional annotation of gene modules.**

(A) Barplot showing percent fraction of overlapping genes in liver/tumor modules compared to corresponding spleen/tumor modules. Dotted line represents the threshold used to determine significant overlap between the modules. (B,C) GO annotations for tumor response module (LM5, SM2/12) (B) and immune regulation module (LM19, SM7/20) (C). P-values computed using GOrilla (50) and adjusted by Benjamini-Hochberg correction. LM and SM denote liver/tumor modules and spleen/tumor modules

respectively. (D) Pie chart representation of cell type fractions that make up each module (See **Methods**)

**Fig. S10: Expression of the cell migration gene module enriched in MSCs.**

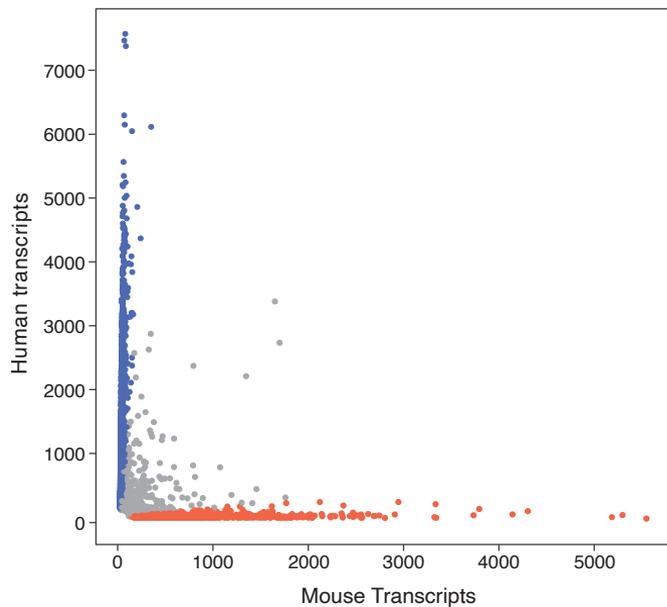
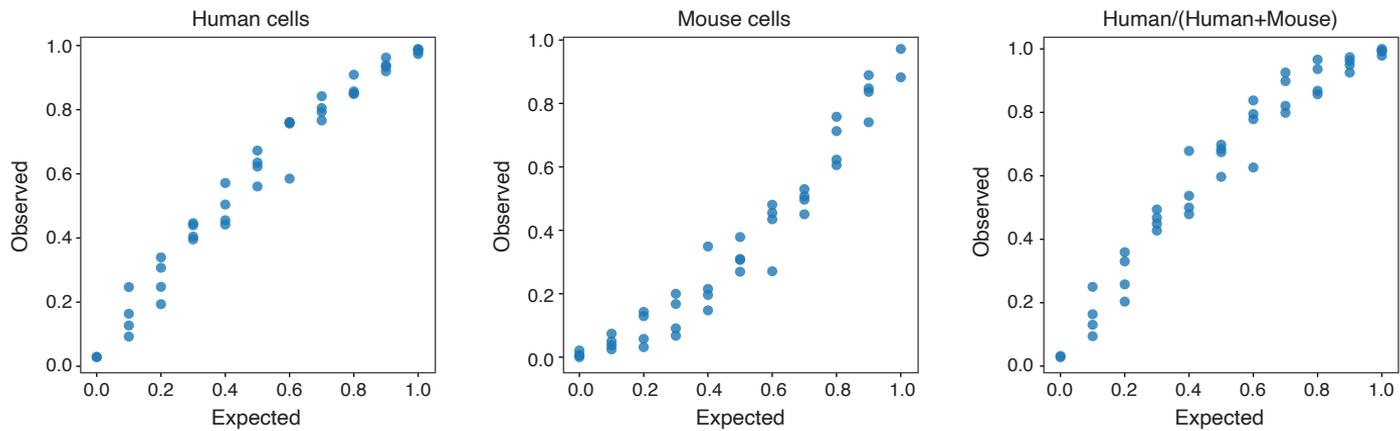
(A) Matrix plot of top overlapping genes in the cell migration module (LM10) across all cell types in liver/tumor. (B) GO annotation for cell migration module from LM10 and SM15/17. P-values computed using GOrilla (50) and adjusted by Benjamini-Hochberg correction.

**Fig. S11: Tumor proximity score defined for both liver and spleen tissue.**

(A) The proximity scores for each tissue relied on the annotation of successive concentric layers of neighbors for a well in question. (B) The set of wells neighboring each well in the array were tabulated for up to 10 layers. (C) The cell-containing wells of representative spleen/tumor slice, where bright green indicates a higher proportion of tumor cells, and bright red indicates a higher proportion of non-tumor cells. (D) The wells containing cells of a representative liver/tumor slice. Bright green indicates higher proportion of tumor cells, bright red indicates higher proportion of hepatocytes. (E) The proximity score values annotated on each well (left top/bottom panels) from the liver tissue, where bright green is closer to the minimum value and bright red is closer to the maximum value. The scores are visualized for different values of  $l$  and  $d$ . The values of  $l = 10$  and  $d = 1.05$  were chosen as they rendered the distribution of scores (right top/bottom panels) more uniform across all wells.

**Supplementary Table 1: Lists of overlapping genes between select liver/tumor and spleen/tumor modules.**

Four liver/tumor modules (LM) that passed the cutoff of having at least 25% of genes overlapping with spleen/tumor modules.

**A****B**

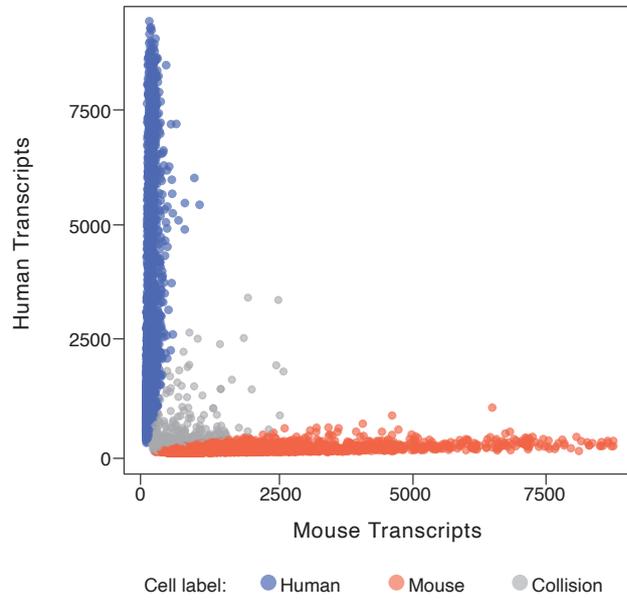
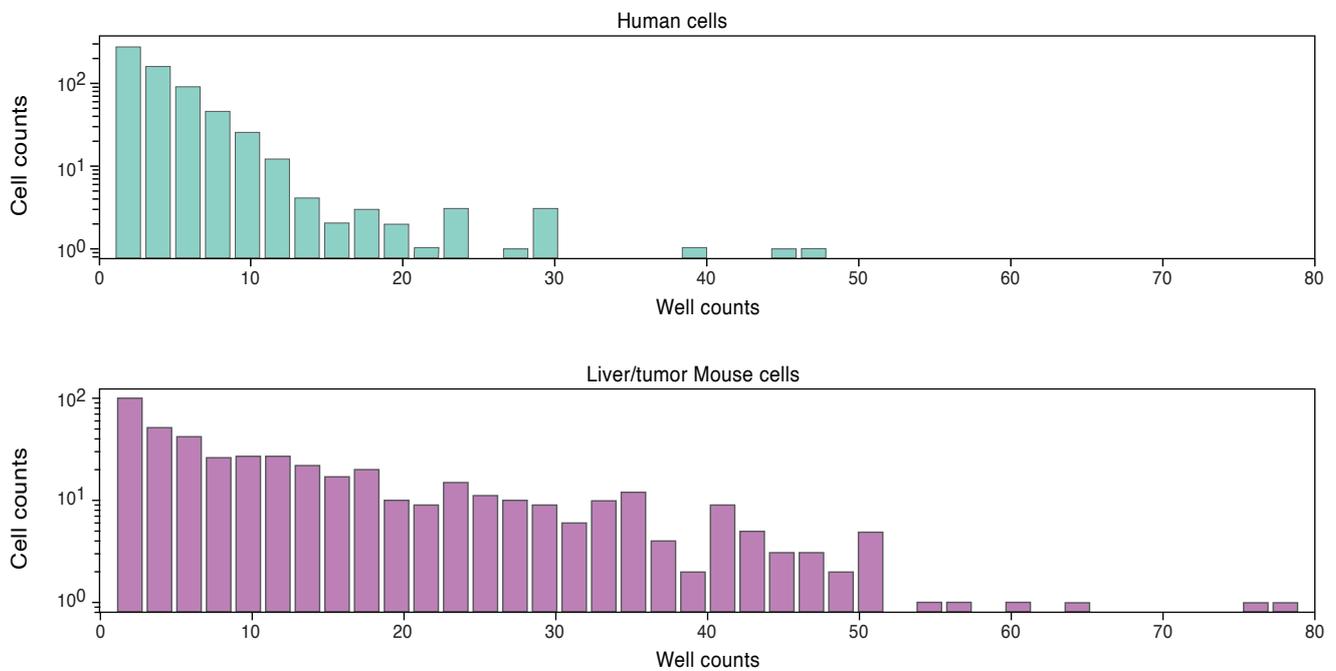
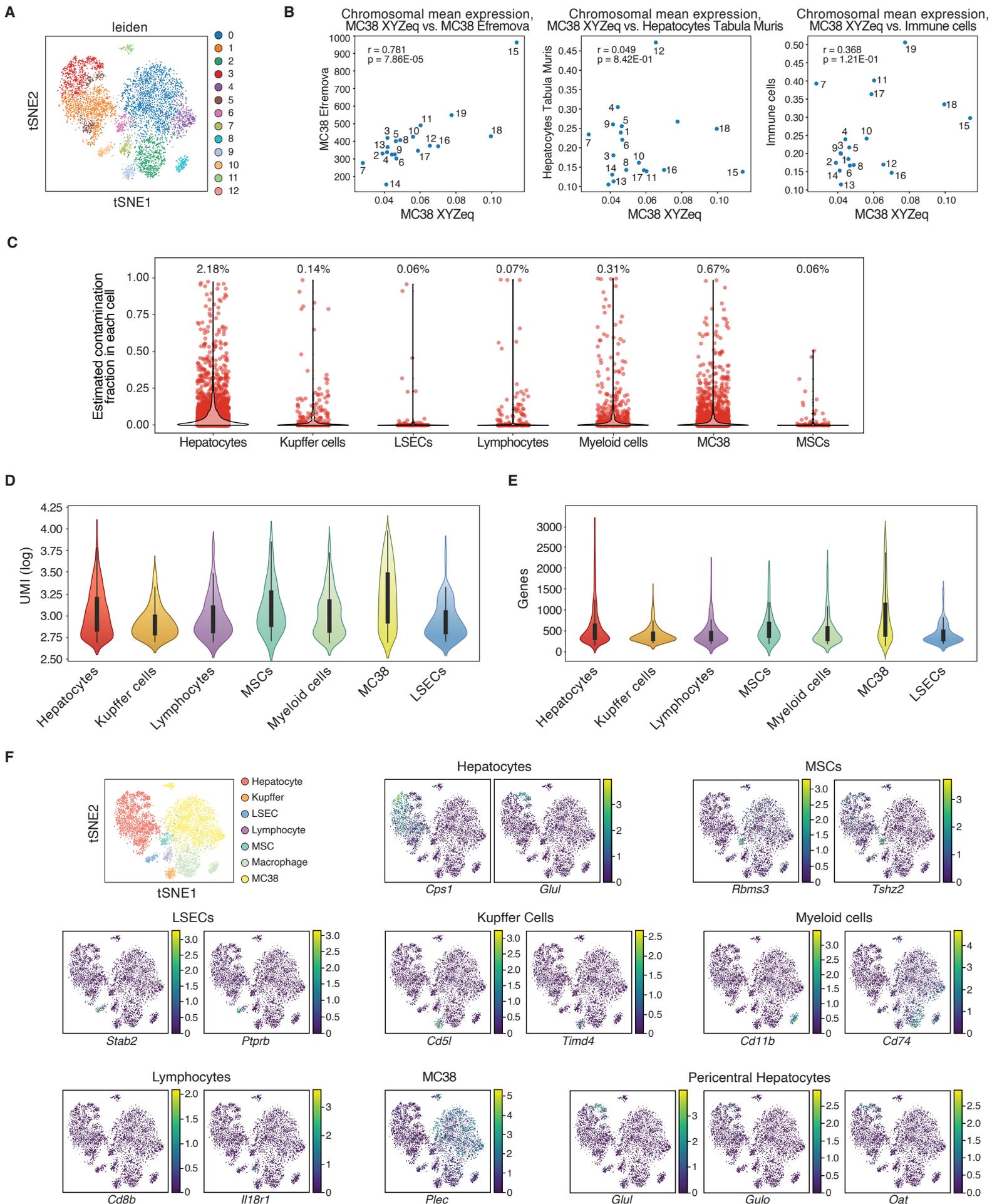
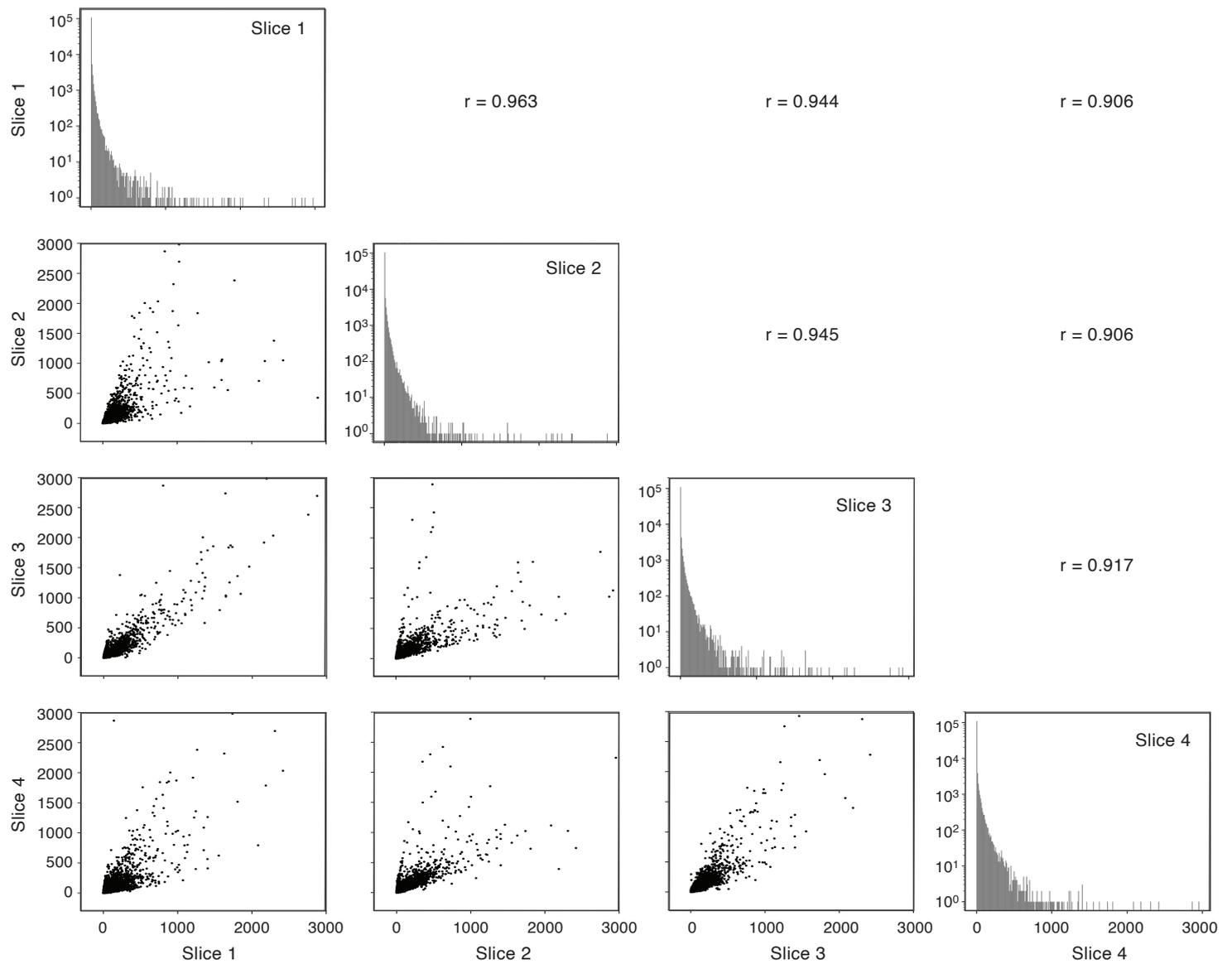
**A****B****C**

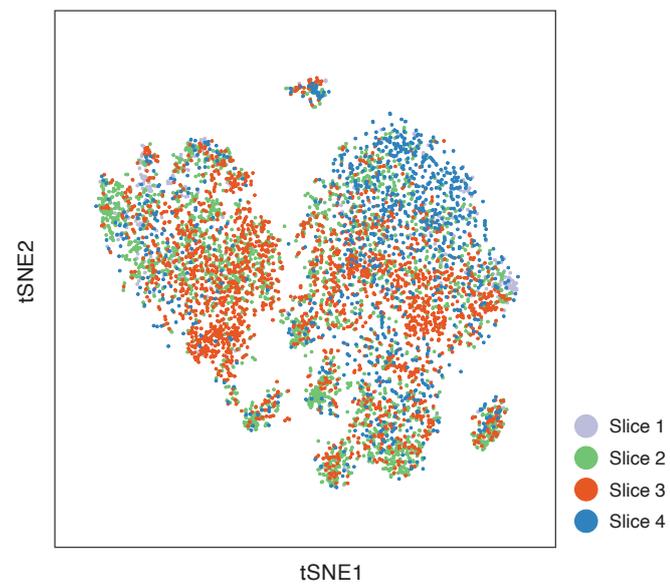
Figure S3

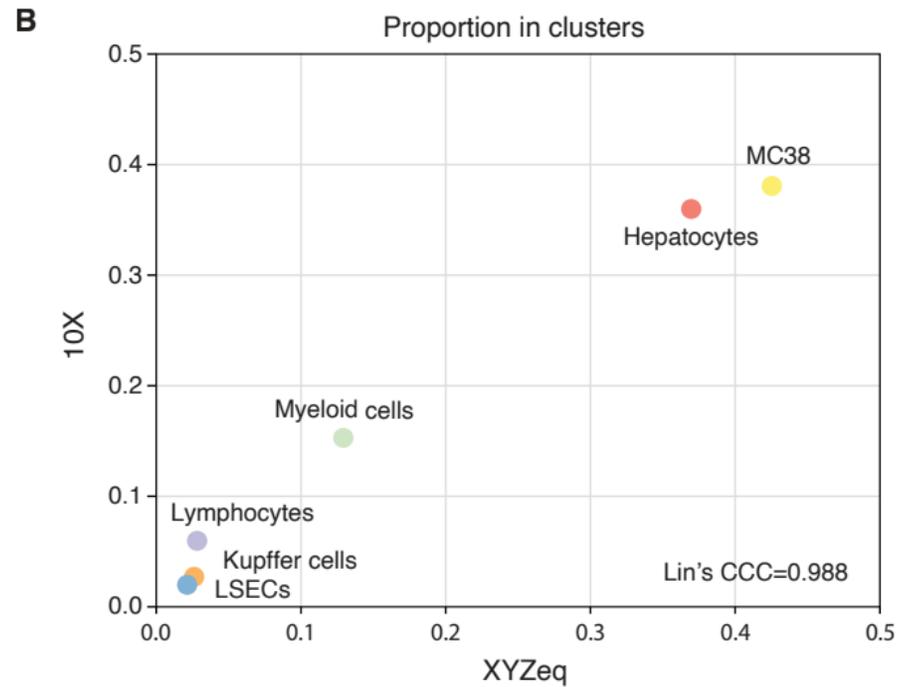
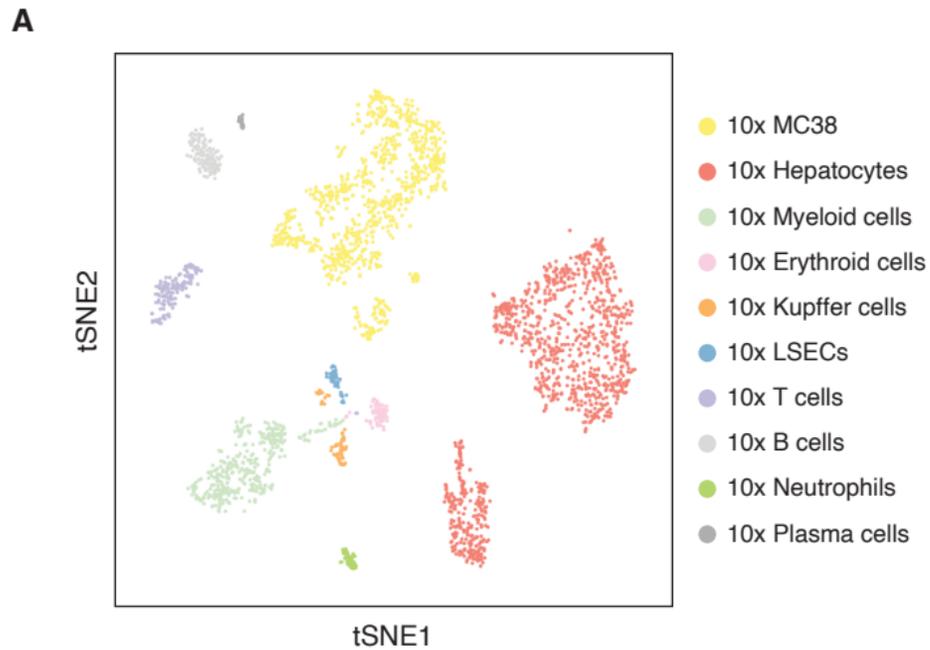


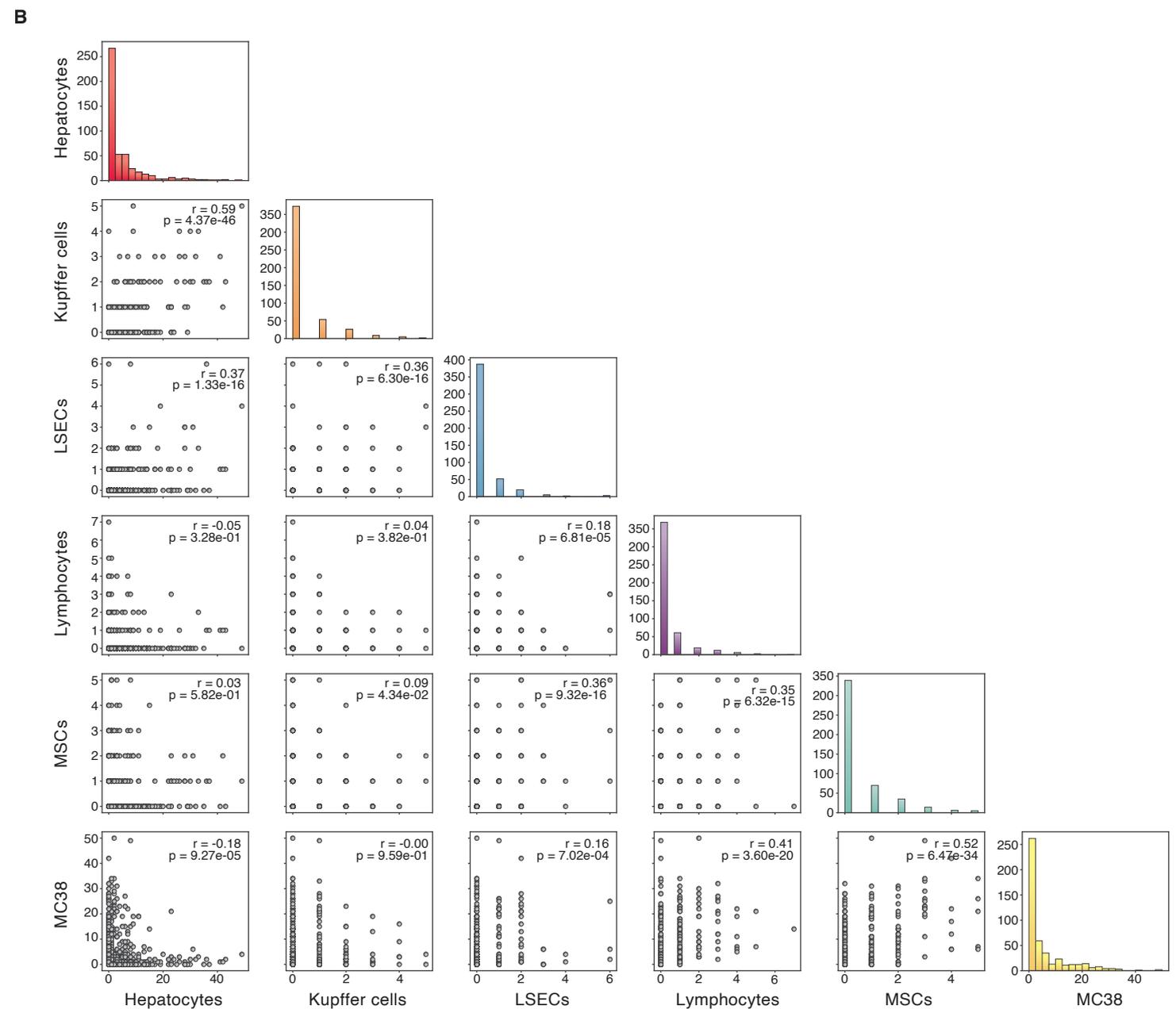
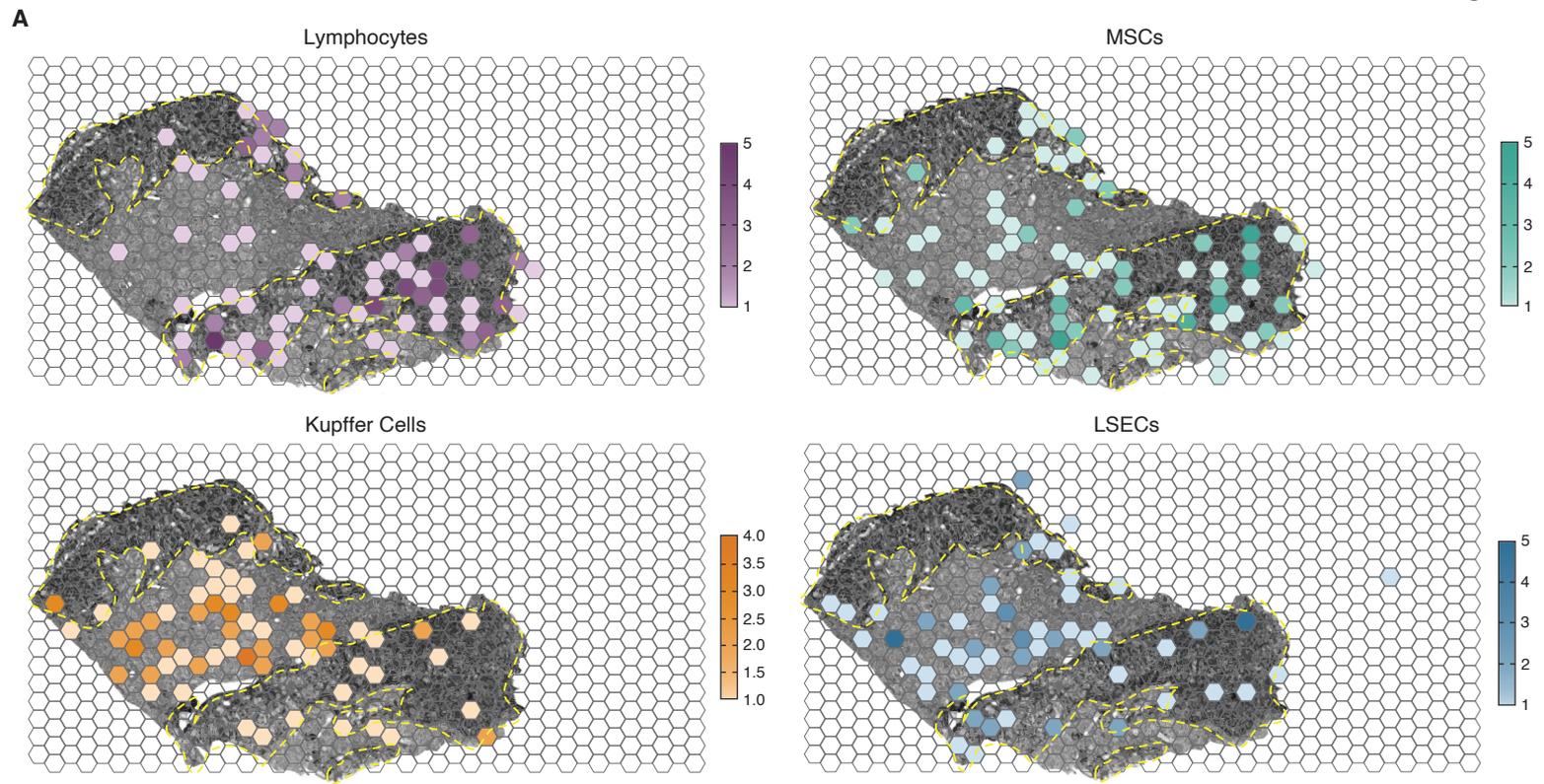
A

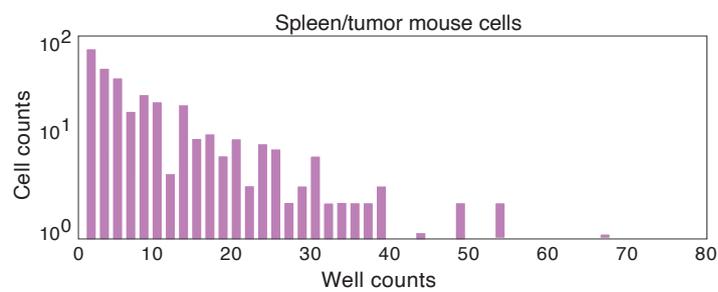
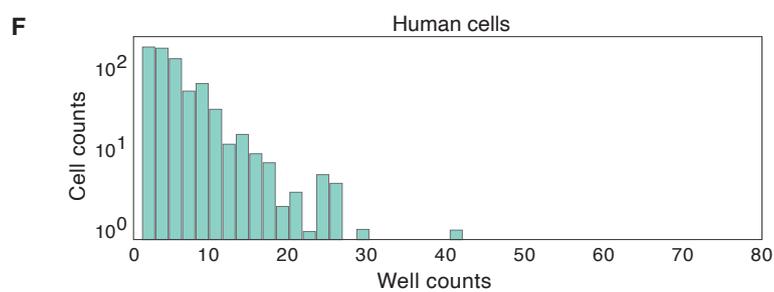
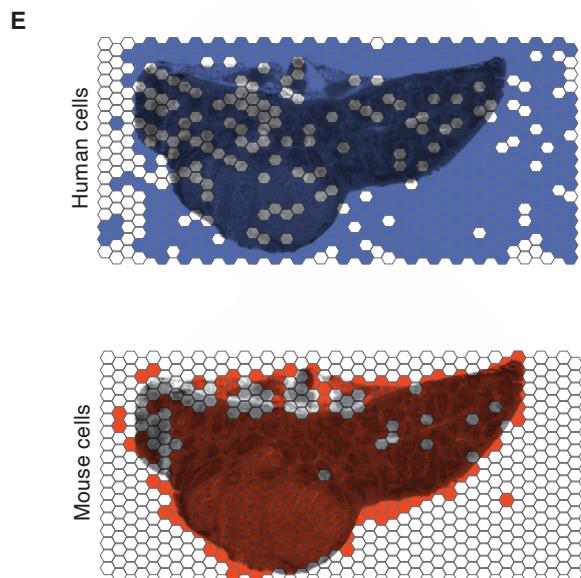
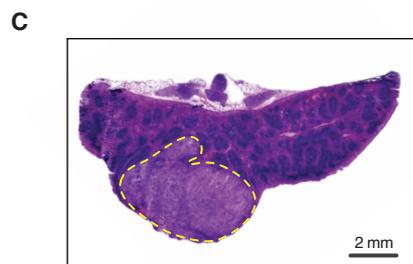
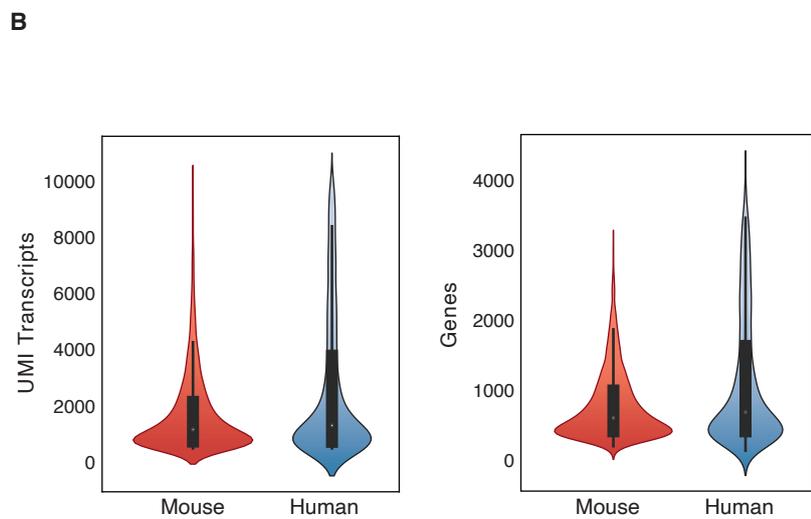
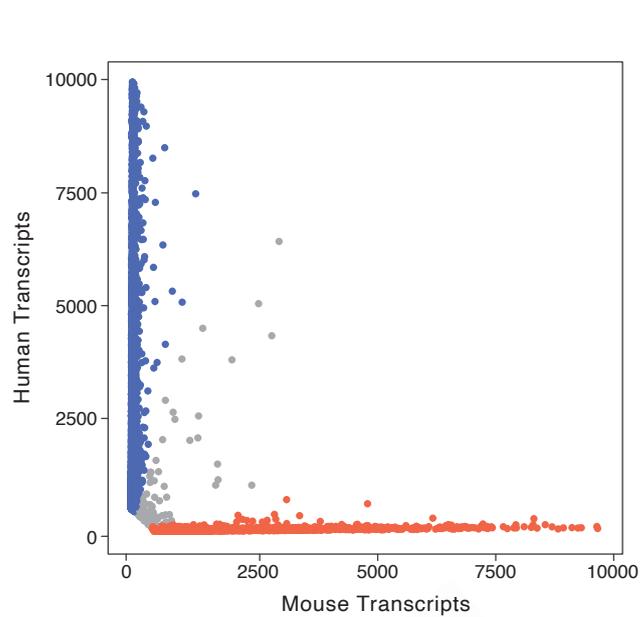


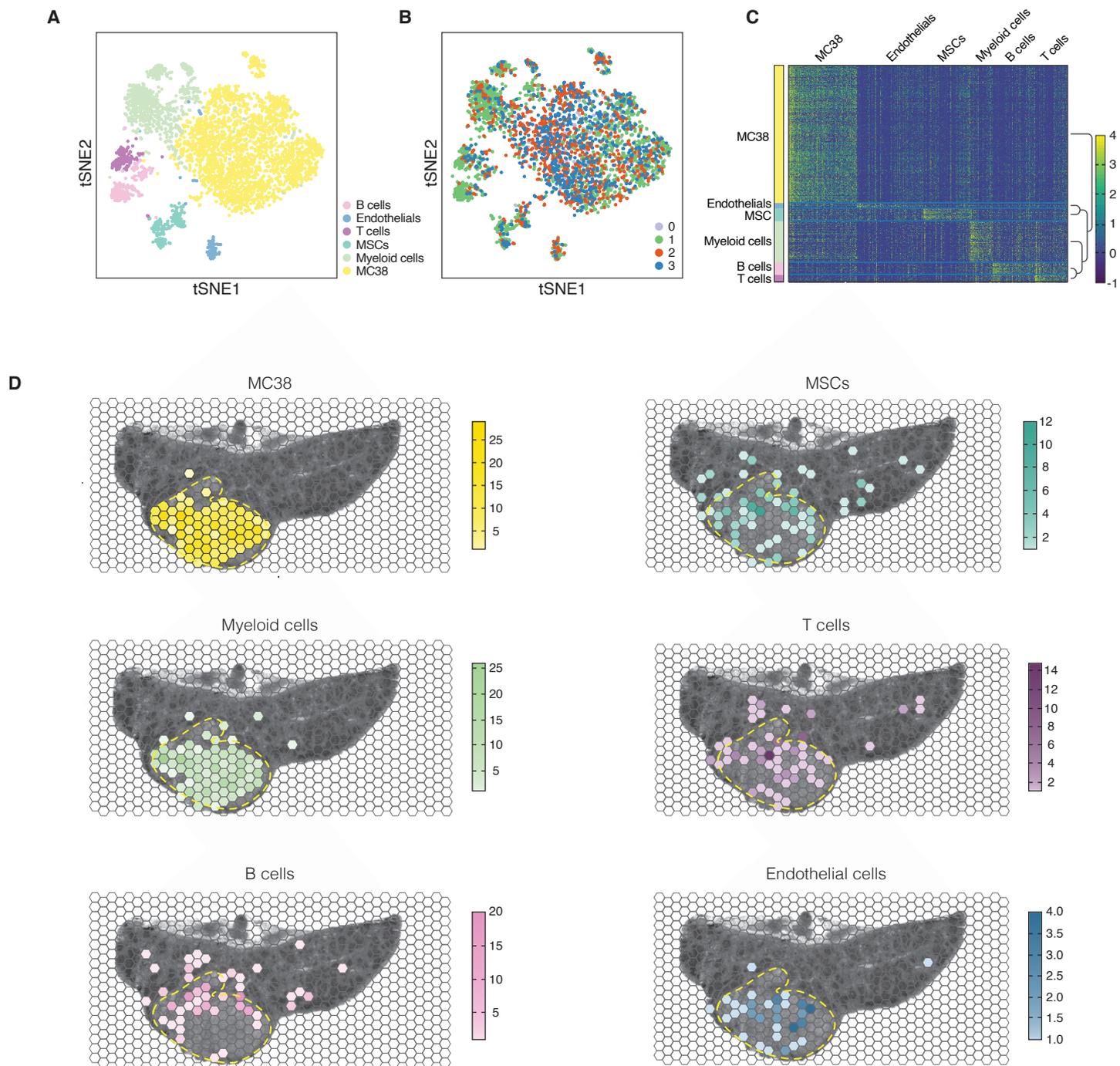
B

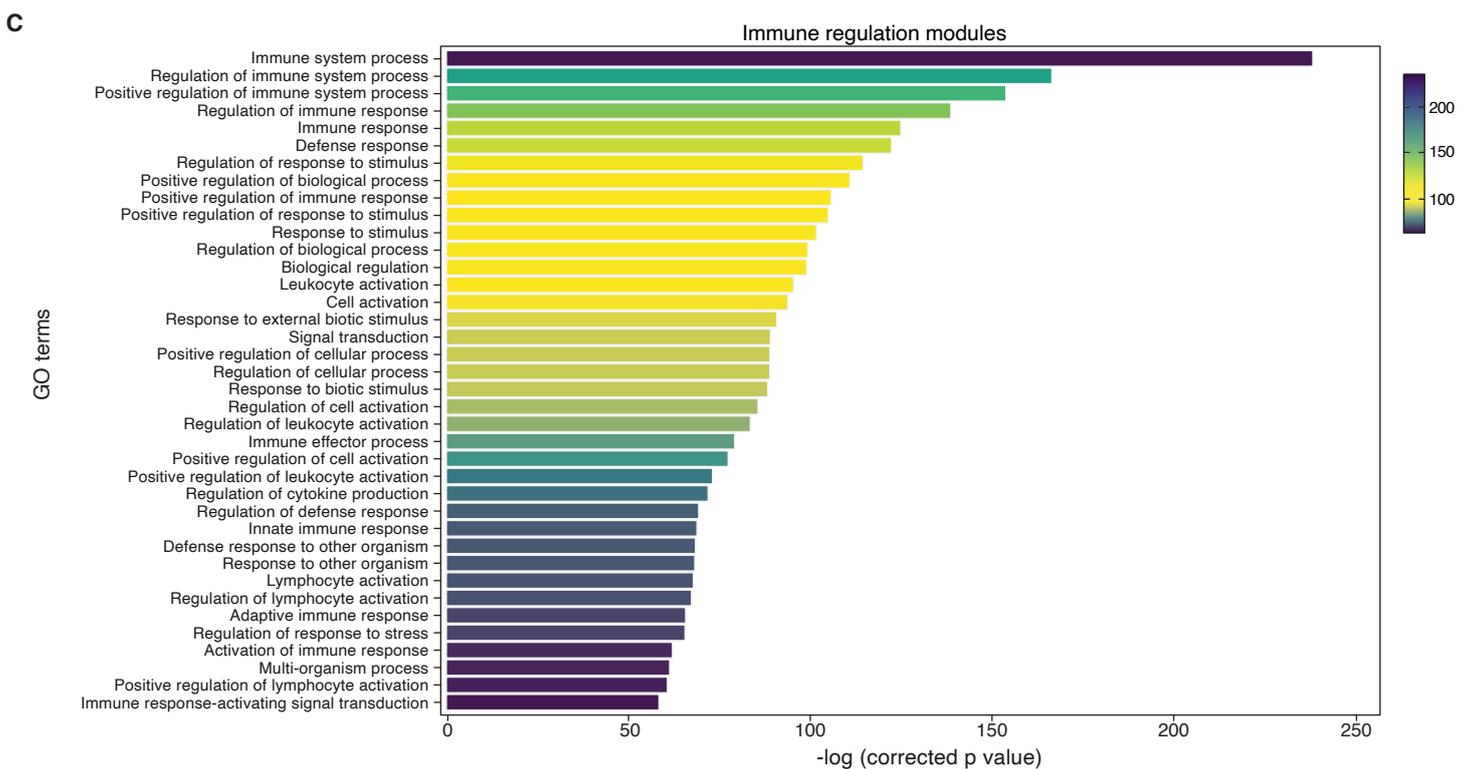
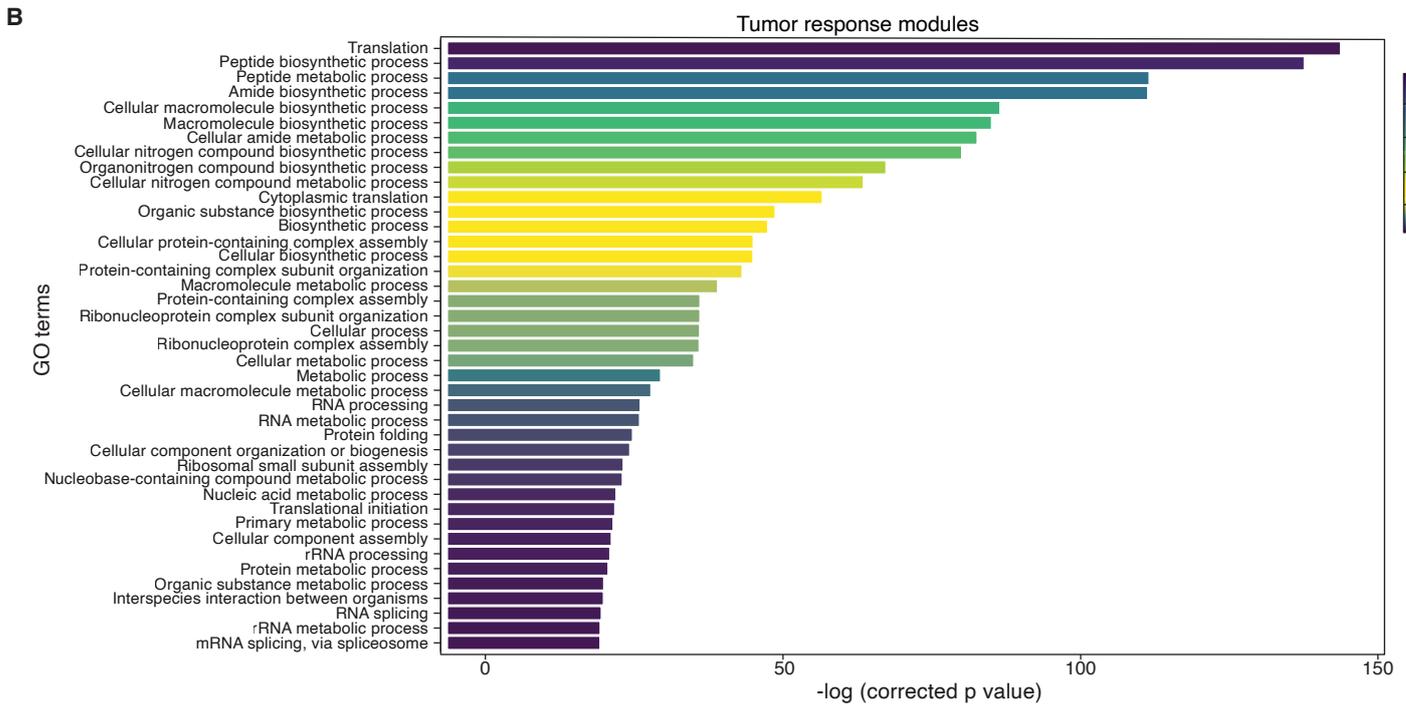
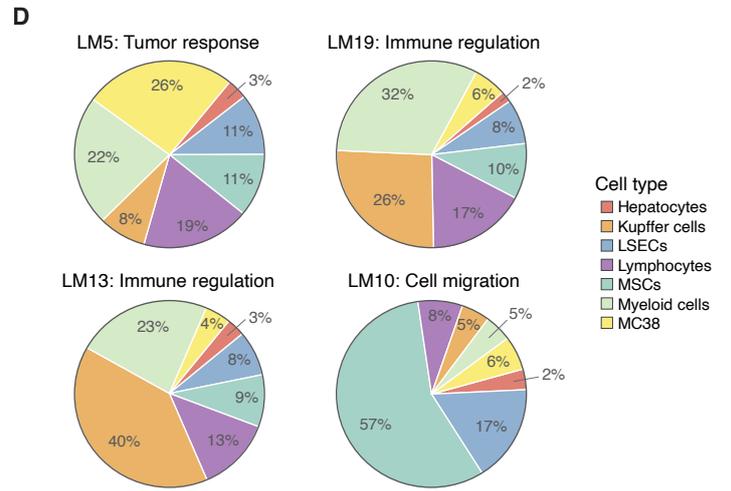
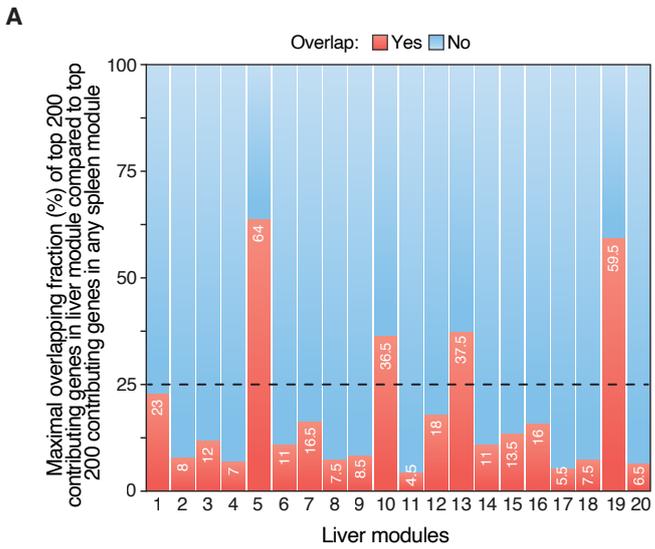




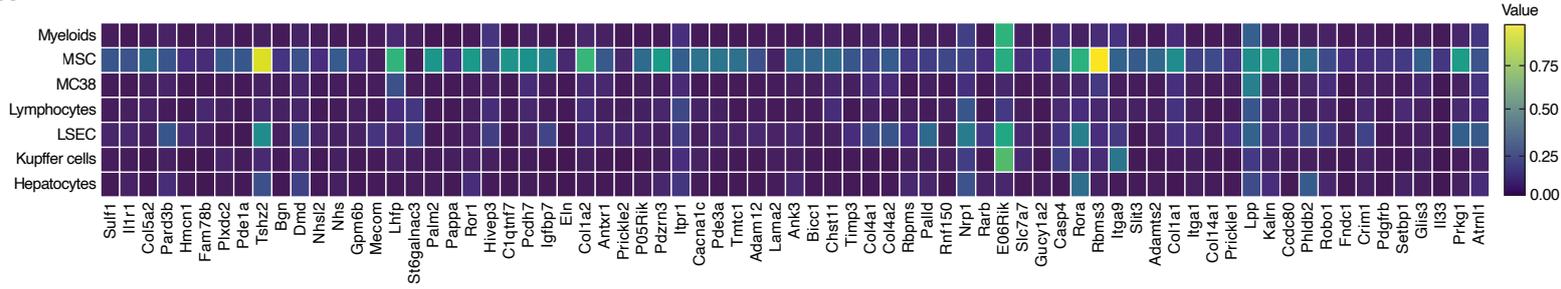




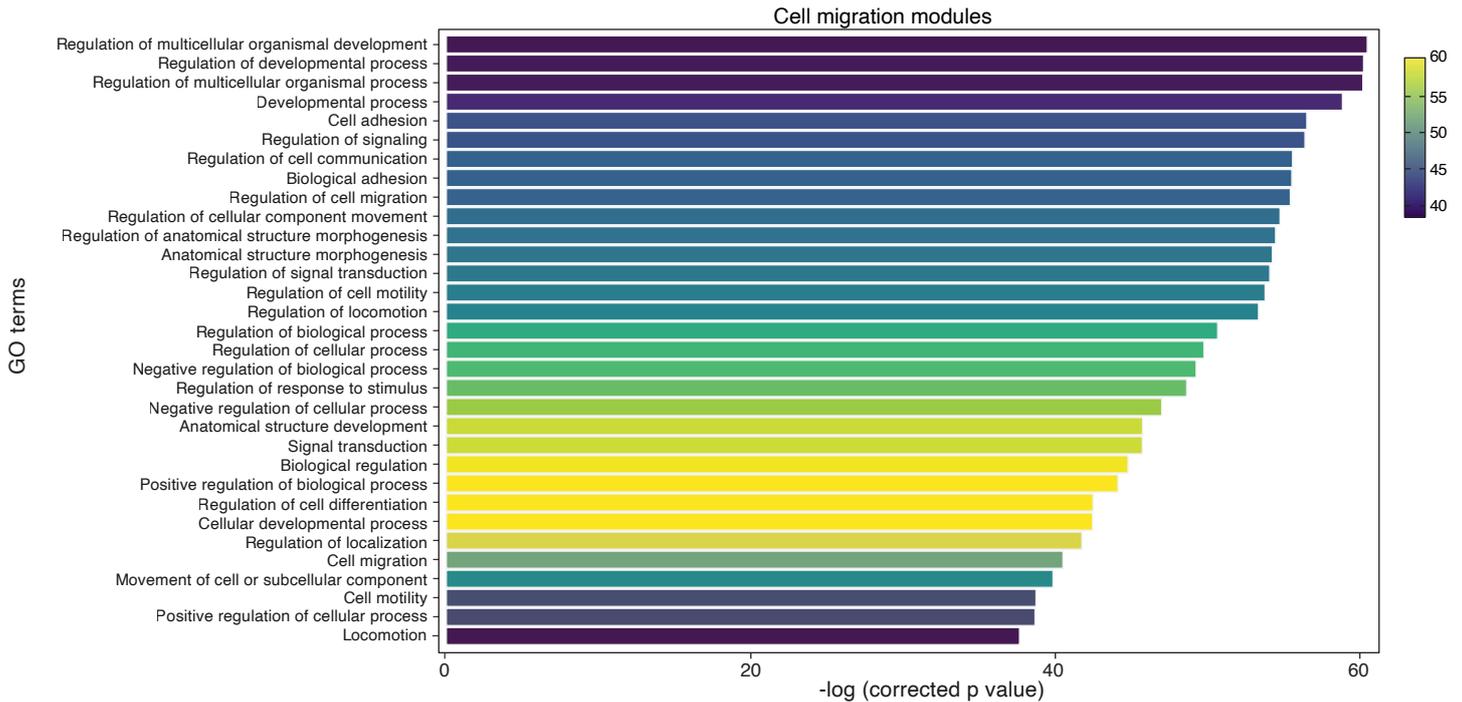




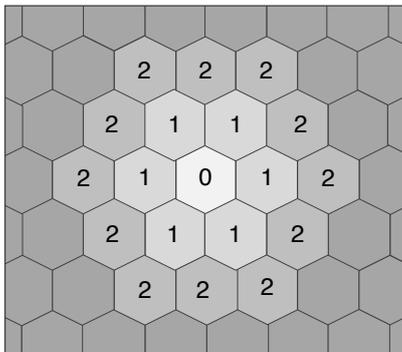
A



B



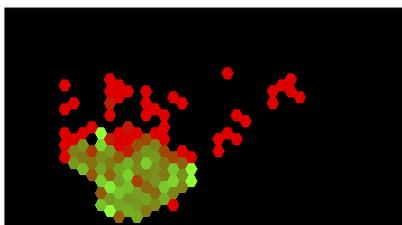
A



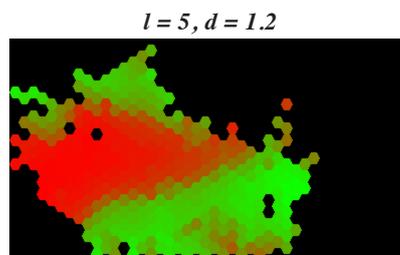
B

$x, y$	$n_1$	$n_2$	$\dots$	$n_{10}$
1, 1	$\{(1, 2), (2, 1)\}$	$\{(3, 2), (1, 3), (3, 1), (2, 2)\}$	$\dots$	$\{(3, 2), (1, 3), (3, 1), (2, 2)\}$
1, 2	$\{(1, 3), (1, 1), (2, 1), (2, 2)\}$	$\{(3, 2), (3, 3), \dots, (1, 4), (2, 3)\}$	$\dots$	$\{(3, 2), (1, 3), (3, 1), (2, 2)\}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
42, 18	$\{(41, 18), (42, 17)\}$	$\{(40, 17), (42, 16), (41, 17), (40, 18)\}$	$\dots$	$\{(32, 13), (32, 14), \dots, (41, 9), (42, 8)\}$

C



E



D

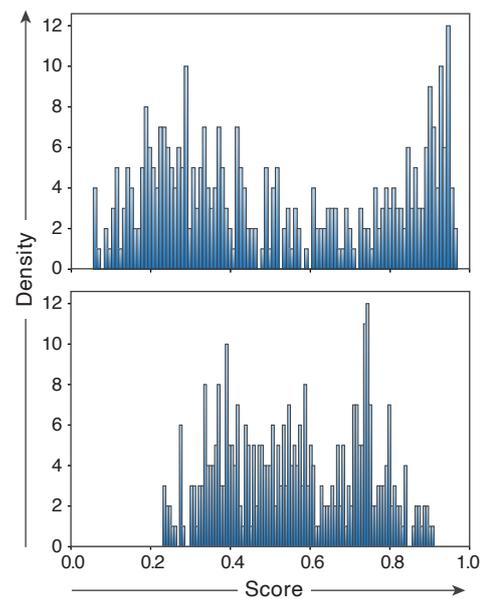
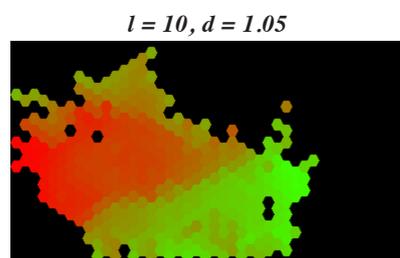
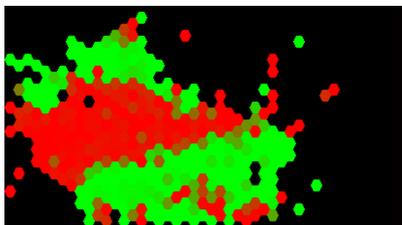


Table 1

Lists of overlapping genes among top 200 contributing genes between liver and spleen

Modules	Overlapping genes
Tumor response modules (128 genes, LM5)	Lgals1 Rpl8 Rpsa Ly6e Rps12 S100a4 Rplp1 Tapbp S100a6 Rpl26 Ptma Rpl14 Rpl38 Rps24 Rps8 Vim Hsp90ab1 B2m Gnb2l1 Rpl21 Ptrf Rplp0 Col3a1 Rpl30 Rps20 Rps26 Rps19 Npm1 Rpl13 Rpl37 Rps5 Rps29 Ly6a Rpl32 Pfn1 Rps15 Rps15a Pfdn5 Ncl Tuba1c Rpl23a Ahnak Rpl13a Rps21 Arpc1b Eef1g Rpl41 Cct7 Rpl37a Rpl23 Calm1 Myh9 Psmc5 Hdgf Msn Pkm Rpl12 Fau Rps27 Rps14 Serpinh1 Cd63 Rpl9 Pdia3 Eef2 Rpl17 Arhgdia Cfl1 Rpl4 Rpn2 Edf1 Ccnd1 Ft1l Sparc Rps25 Hspa8 Rplp2 Plec Rps10 Scd2 Cyb5r3 Serbp1 Rps17 Cct5 Rcn1 Rpl5 Rpl19 H1f0 Grn Rps27a Rpl11 Tuba1a P4hb Rpl36 Hnrnpab Rpl36a Pabpc1 Eif5a Eef1b2 Hspa9 Bsg Rpl27 Rps3 Rpl22 Ybx1 Psmb10 Tmsb10 Col6a1 Ecm1 Fth1 Set Cald1 Canx Tubb5 Cox4i1 Vcp Eef1a1 S100a10 Rpl34 Ldha Gabarap Rpl18a Hint1 Ppp1ca Kpnb1 Tcp1 Rps28 Pdap1
Immune regulation modules (75 genes, LM13)	Ptprj Fyb Mertk Aoah Adap2 Pou2f2 Adgre1 Myo9a Unc93b1 Dock10 Frmd4b Zd-hhc14 Myo1f Prkcb Zeb2 Tmcc3 Zfp710 Slc9a9 Itga9 Lcp2 Hck Fcer1g Mrc1 Rreb1 Spi1 Ccr5 1700112E06Rik Elmo1 March1 Abcg3 Trpm2 Lyn Gm5150 Slc8a1 Nrros Ctsc Gab2 Sirpa Ly86 P2rx4 Pla2g4a Entpd1 Cd84 Fgr Acer3 Cd300lf Cadm1 Tmsb4x Msr1 Maf Mitf Lgmn Csf1r Ncf2 Ptprc Gmip Pik3ap1 Ctsb Lacc1 Arhgap15 Dock8 Plcl2 Shtn1 Prex1 Diaph2 Exoc6 Dock4 Mgat5 Slc43a2 Arrb2 Ppm1h Slc15a3 Pld4 Tbxas1 Vrk2
Immune regulation modules (119 genes, LM19)	Tgfb1 Cd74 Arhgap15 Ccr5 Ctss Ly86 Myo1f 1700112E06Rik Dock10 Elmo1 Tgfb1 Inpp5d Epsti1 Zeb2 Ccr2 H2-Ab1 Ptprc Slc8a1 Dock2 Dock8 Laptm5 Arhgap30 Hck Ms4a6c Cd84 Lyn Ctsc Cyth4 Fyb Itga4 Pld4 Gab2 Lcp1 Cybb Unc93b1 Tbxas1 Gpr141 Hmha1 C1qa Tmsb4x Plcg2 March1 Cx3cr1 Gm4955 Ctsb Pde7b Cd300lf Sirpa Dock4 Zfp710 Nlrp1b Pik3cd Fam49b Klra2 Ms4a6b Slfn5 H2-Eb1 Aoah Lcp2 Csf1r Arhgap24 Abr Cers6 Entpd1 H2-Aa Ptk2b Sp100 H2-DMb1 Slfn2 Mpeg1 Aim2 Apobec1 Rreb1 Etv6 Adap2 Irf7 Spi1 Ms4a4c Gm2a Lyz2 Lgmn Stat1 Ms4a7 Tnfrsf11a Apba1 Pik3ap1 Fam105a Nckap1l Psap Arid5b Mndal Slfn8 Trim30a Lair1 Picalm Il10ra Ncf1 Ccdc88b Coro1a C1qc Gatm Tgm2 Ctsh C1qb Pip4k2a Chd7 Mx1 Samhd1 Fcgr2b Diaph2 Mitf Mef2a Pou2f2 Pid1 Ifi203 Wdfy4 Fnbp1 Dennd1b Pik3r5
Cell migration modules (73 genes, LM10)	Tshz2 Col1a2 Rbms3 Igfbp7 Kalrn Prkg1 Ccdc80 Rora Palm2 Timp3 Pdzn3 Pde3a Adamts2 Ror1 Sulf1 Palld Pcdh7 C1qtnf7 Tmtc1 Pde1a Slit3 Adam12 Col14a1 Col5a2 Casp4 Col1a1 Ank3 Bicc1 Col4a2 Plxdc2 Il1r1 1700112E06Rik Lhfp Rnf150 Pard3b Cacna1c Crim1 Itga1 Itga9 9530026P05Rik Gucy1a2 Col4a1 Pdgfrb Nhs Hivep3 Fndc1 Prickle1 Chst11 Rarb Robo1 Il33 Gpm6b Itpr1 Bgn Mecom Slc7a7 Atrnl1 Rbpms Hmcn1 Pappa Lpp Phldb2 Dmd Antxr1 Lama2 St6galnac3 Fam78b Nrp1 Setbp1 Eln Nhsl2 Prickle2 Glis3