

DISS. ETH Nr. 29130

*DEVELOPMENT OF SPHERE-SEQUENCING TO STUDY LOCAL
MICROENVIRONMENTS WITHIN COMPLEX TISSUES AT SINGLE-
CELL RESOLUTION*

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

KRISTINA HANDLER

MSc Molekulare Biologie, Universität Wien, Österreich

born on 28th of June, 1993

citizen of Austria

accepted on recommendation of

Prof. Dr. Andreas E. Moor, examiner

Prof. Dr. Barbara Treutlein, co-examiner

Prof. Dr. Roser Vento-Tormo, co-examiner

2023

Table of Contents

| | | |
|------------|---|-----------|
| 1 | Summary | 5 |
| 2 | Zusammenfassung | 7 |
| 3 | Authors Contributions | 11 |
| 4 | Abbreviations | 13 |
| 5 | Introduction | 15 |
| 5.2 | Spatial transcriptomics | 15 |
| 5.2.1 | <i>What is spatial transcriptomics?</i> | 15 |
| 5.2.2 | <i>Historical development of spatial transcriptomics (NGS-based approaches)</i> | 15 |
| 5.2.3 | <i>Spatial transcriptomics approaches</i> | 17 |
| 5.2.3.1 | Imaging-based..... | 18 |
| 5.2.3.1.1 | <i>In situ hybridization (ISH)</i> | 18 |
| 5.2.3.1.2 | <i>In situ sequencing (ISS)</i> | 19 |
| 5.2.3.1.3 | <i>Gene panel design for targeted imaging-based ST</i> | 21 |
| 5.2.3.2 | mRNA capture-based..... | 21 |
| 5.2.3.2.1 | <i>Spatial indexing</i> | 21 |
| 5.2.3.2.2 | <i>Regions of interest</i> | 23 |
| 5.2.3.2.3 | <i>Partial dissociation</i> | 24 |
| 5.2.3.3 | Computational methods to infer the spatial location of a cell..... | 25 |
| 5.2.3.4 | Spatial methods capturing mRNA plus other modalities..... | 25 |
| 5.2.4 | <i>Current limitations</i> | 26 |
| 5.3 | Ligand-receptor interaction analysis | 27 |
| 5.3.1 | <i>Different types of ligand-receptor interactions</i> | 27 |
| 5.3.2 | <i>Inferring ligand-receptor interactions from single-cell transcriptomics data</i> | 28 |
| 5.3.3 | <i>Inferring ligand-receptor interactions from spatial transcriptomics data</i> | 29 |
| 5.3.4 | <i>How to identify cell-cell communications that are biologically most relevant</i> | 30 |
| 5.3.5 | <i>Limitations of current approaches</i> | 30 |
| 5.4 | Mouse liver biology | 30 |
| 5.4.1 | <i>Anatomy of the liver</i> | 30 |
| 5.4.2 | <i>Functions of the liver</i> | 31 |
| 5.4.3 | <i>Immune tolerogenic feature of the liver</i> | 31 |
| 5.4.4 | <i>Cell types within the liver and their functions</i> | 32 |
| 5.4.4.1 | Parenchymal cells..... | 32 |
| 5.4.4.2 | Non-parenchymal cells..... | 32 |
| 5.4.5 | <i>Metabolic zonation</i> | 33 |
| 5.4.6 | <i>Immune zonation</i> | 34 |
| 5.5 | Colorectal liver metastasis | 34 |
| 5.5.1 | <i>Epidemiology and current treatment options</i> | 34 |
| 5.5.2 | <i>Phases of liver metastasis development</i> | 35 |
| 5.5.3 | <i>Influence of hepatic cells in metastasis formation, promotion, and reduction</i> | 35 |
| 5.5.4 | <i>Mouse model of colorectal liver metastasis</i> | 36 |
| 6 | Aims of this thesis | 37 |
| 7 | Material and methods | 39 |
| 7.1 | Experimental protocol of sphere-sequencing | 39 |
| 7.1.1 | Materials | 39 |
| 7.1.1.1 | Equipment..... | 39 |

| | | |
|--------------|---|-----------|
| 7.1.1.2 | Consumables..... | 40 |
| 7.1.1.3 | Reagents..... | 40 |
| 7.1.2 | Procedure..... | 42 |
| 7.1.3 | Anticipated Results..... | 51 |
| 7.1.4 | Troubleshooting..... | 52 |
| 7.2 | Computational analysis approaches for sphere-sequencing..... | 52 |
| 7.2.1 | Packages for analysis..... | 53 |
| 7.2.1.1 | Linux programming..... | 53 |
| 7.2.1.2 | Python programming..... | 53 |
| 7.2.1.3 | R programming..... | 53 |
| 7.2.2 | Procedure..... | 53 |
| 7.2.3 | Anticipated results and troubleshooting..... | 56 |
| 7.3 | Specific methods and materials used in this thesis..... | 57 |
| 7.3.1 | Sphere-sequencing of murine metastatic and healthy livers..... | 57 |
| 7.3.1.1 | Induction of metastasis in mouse liver and tissue collection..... | 57 |
| 7.3.1.2 | Sphere-sequencing experimental procedure..... | 58 |
| 7.3.1.3 | Sphere-sequencing Computational analysis..... | 58 |
| 7.3.2 | Sphere-sequencing of the organoid mixing species experiment..... | 59 |
| 7.3.2.1 | Organoid cultivation and collection..... | 59 |
| 7.3.2.2 | Experimental procedure..... | 60 |
| 7.3.2.3 | Computational analysis..... | 60 |
| 7.3.3 | Sphere-sequencing of mouse healthy spleen..... | 61 |
| 7.3.3.1 | Experimental procedure..... | 61 |
| 7.3.3.2 | Computational analysis..... | 61 |
| 7.3.4 | Sphere-sequencing of Crohn's disease biopsies..... | 62 |
| 7.3.4.1 | Experimental procedure..... | 62 |
| 7.3.4.2 | Computational analysis..... | 62 |
| 7.3.5 | Conventional single-cell RNAseq of murine metastatic liver..... | 62 |
| 7.3.5.1 | Experimental procedure..... | 62 |
| 7.3.5.2 | Computational analysis..... | 63 |
| 7.3.6 | Visium of mouse metastatic liver..... | 63 |
| 7.3.6.1 | Tissue collection..... | 63 |
| 7.3.6.2 | Experimental procedure..... | 63 |
| 7.3.6.3 | Computational analysis..... | 64 |
| 7.3.7 | Public Visium data analysis..... | 65 |
| 7.3.8 | Molecular Cartography of mouse metastatic liver..... | 65 |
| 7.3.8.1 | Tissue collection..... | 65 |
| 7.3.8.2 | Experimental procedure..... | 65 |
| 7.3.8.3 | Image analysis in ImageJ..... | 65 |
| 7.3.8.4 | Computational analysis..... | 65 |
| 8 | Results..... | 67 |
| 8.1 | Overview of sphere-seq and the sorting approach..... | 67 |
| 8.2 | Comparison of sphere-seq to conventional scRNA-seq..... | 70 |
| 8.3 | A mixing species experiment showed a high accuracy of the sphere-seq approach..... | 72 |
| 8.4 | An imaging-based transcriptomics approach was used for the validation of sphere-seq..... | 73 |
| 8.5 | Sphere-seq could be used to identify previously uncharacterized zoned gene expression in LECs and KCs..... | 74 |
| 8.5.1 | Reconstruction of the sphere position along the lobule layer axis..... | 74 |
| 8.5.2 | Determination of spatial areas in Molecular Cartography images..... | 76 |

| | | |
|-------|--|------------|
| 8.5.3 | <i>Sphere-seq identified previously uncharacterized zoned gene expression in LECs</i> | 76 |
| 8.5.4 | <i>Sphere-seq identified zoned gene expression in KCs indicating an involvement in leukocyte recruitment</i> | 79 |
| 8.6 | Spheres could be grouped into areas proximal and distal to metastatic sites to uncover spatially restricted differences | 81 |
| 8.7 | Application of sphere-seq to other tissues and species | 84 |
| 8.7.1 | <i>Application to healthy murine spleen</i> | 84 |
| 8.7.2 | <i>Modified sphere-seq protocol with sequential filtering and manual picking – Example of Crohn’s disease biopsies</i> | 85 |
| 8.7.3 | <i>Sphere-seq could be used to investigate differences in fibrotic and non-fibrotic Crohn’s disease spatial areas</i> | 87 |
| 8.8 | Comparison of sphere-seq to Visium | 88 |
| 8.9 | Potential biases introduced by sphere size | 90 |
| 9 | Discussion | 95 |
| 9.1 | Strength of sphere-seq | 95 |
| 9.2 | Limitations of sphere-seq | 96 |
| 9.3 | Sphere-seq and L-R interaction analysis | 97 |
| 9.4 | Biological knowledge gained with sphere-seq and outlook | 99 |
| 9.5 | Future applications of sphere-seq | 100 |
| 9.6 | Future optimizations/enhancements of sphere-seq | 100 |
| 10 | Conclusion | 103 |
| 11 | Bibliography | 105 |
| 12 | Acknowledgments | 121 |

1 Summary

Within a multicellular organism, cells are organized in a defined manner while communicating with each other. A single cell never just stands alone; its behavior and gene expression are always influenced by its local microenvironment. Disruptions of these organizations cause reciprocal changes in both a cell's transcriptome and its cellular crosstalk with the environment, which can lead to the development of diseases. It is therefore important to not only know the identities and gene expression of single cells, something that can be investigated by single-cell RNAseq (scRNA-seq), but also the spatial arrangements of different cell types and states. Spatial transcriptomics (ST) can be described as the detection of mRNA transcripts within a spatially defined area. In recent years there has been a massive development of ST technologies with constant improvements and optimizations. However, some limitations still affect existing ST methods, namely the need for specialized equipment and expertise, lack of single-cell resolution, or low throughput.

In this thesis, I developed a new ST method called sphere-sequencing (sphere-seq) that addresses these limitations. In short, fresh tissues, which can be from a variety of origins and shapes, are partially dissociated into cellular communities (spheres) of sizes between 100-500 μm . Spheres are then sorted in a size-dependent manner into wells of a 96-well plate using a large fragment biosorter, which works like a conventional flow cytometer but enables the sorting of objects much larger than single cells. Sphere sizes are defined based on prior knowledge of interesting spatial units within complex tissues, for example in the murine liver we were sorting spheres between 200-450 μm to capture spatial differences along central-portal vein axis. After sorting, spheres are dissociated into single cells and labeled with sphere-specific barcodes using a method where barcodes integrate into a cell's membrane via lipid anchors. We established a set of 288 different barcodes, which allows the analysis of 288 spheres per experiment. After cell labeling, all cells are pooled and used for scRNA-seq. After sequencing and demultiplexing a single cell can then be allocated to its sphere of origin and thereby its cellular neighborhood based on the sphere barcode. The anatomical niche of a sphere can then be reconstructed using landmark gene signatures or the presence and absence of a specific cell type. This approach helps to group spheres into defined anatomical regions, either using nominal categories (e.g., metastasis-associated and metastasis-free) or ordinal categories (e.g., specific layers within a zoned organ). Ordinally or nominally labeled spheres can then be analyzed to determine spatially restricted gene expression, cell type abundances, or ligand-receptor (L-R) interactions.

We applied sphere-seq to the mouse liver, to study differences in gene expression between anatomical zones. The liver is a highly spatially organized organ. Hepatocytes, the parenchymal cells of the liver, are organized into repetitive hexagonal structures called lobules. From each corner of the lobule, portal veins and hepatic arteries carry nutrient- and oxygen-rich blood, which then flows through sinusoidal vessels toward the central vein. This generates a gradient of oxygen, nutrients, hormones, and WNT signaling along the portal-to-central axis, which in turn leads to a division of labor between hepatocytes from different zones along this gradient. Non-parenchymal cells (NPCs), like liver endothelial cells (LECs), assist hepatocytic tasks. Different tasks along the central-portal axis require different proteins, therefore, many genes are differentially expressed in different layers of the lobule, which was extensively studied in hepatocytes and LECs. We, therefore, employed a previously identified

gene signature in LECs to reconstruct the spatial position of each sphere along the central-portal axis. This enabled us to study zoned gene expression in other cell types, for example in Kupffer cells (KCs), the tissue-resident macrophages of the liver. This revealed a zoned expression for *Vcam1*, which was enriched in KCs of the periportal regions in metastases-bearing mice. VCAM1 was shown to be involved in immune cell recruitment which in turn can lead to KC activation. L-R interaction analysis further strengthened this assumption because this predicted VCAM1 ligands on KCs interacting with various integrins on immune cells, especially in periportal regions. In general, integrins are important players for immune cells in the processes of binding to and transmigrating through the vascular wall, migrating through tissues, and interacting with other cells. We could therefore speculate that there is a higher need for *Vcam1* to bind incoming immune cells from the portal blood to achieve an increased immune reaction due to the metastatic process.

We further used sphere-seq to assess spatial differences between metastatic niches in a mouse model of colorectal cancer (CRC) liver metastasis. Liver metastasis formation due to CRC is especially common, around 41 % of patients develop metastases in the liver. This is due to the anatomical position and structure of the liver. It is a highly vascular organ and it lies in proximity to the gastrointestinal tract, where blood drains from the gut into the liver, bringing circulating cancer cells from primary tumors into the liver where they can persist and form metastases. However, in the metastatic liver, there are also many areas with healthy tissues and we, therefore, wanted to apply sphere-seq to study the spatial differences between distal and proximal areas to micro-metastatic sites to study what factors are involved in metastasis development. Spheres were grouped based on the presence (= proximal) and absence (= distal) of metastatic cells. This analysis revealed an enrichment of *C1q+* macrophages, which are associated with an anti-inflammatory phenotype, within proximal areas. This subset of macrophages has been previously shown to be involved in T cell exhaustion and is an indicator of poor cancer prognosis. With L-R interaction analysis between macrophages and T cells, we could identify interactions involving SPP1 and FN1 ligands expressed by macrophages, which were shown to be enriched in various cancers and correlate to poor prognosis.

We applied an imaging-based ST technology for multiplexed fluorescence *in situ* hybridization (smFISH) using a 100-plex gene panel to the same samples of metastasis-bearing murine livers. Spatial differences that we identified with sphere-seq analysis could be validated using this imaging approach. This highlighted that sphere-seq truly captures spatial variability of gene expression *in situ*.

Application to other samples like CRC organoids, murine spleen, and Crohn's disease patient biopsies showed that sphere-seq is a versatile method which enables addressing diverse biological questions. Comparison to the most widely used ST method, Visium, revealed that sphere-seq is superior in terms of cellular resolution and the ability to capture different spatial niches with similar quality.

Together, this thesis shows the development of sphere-seq with a comprehensive experimental and computational workflow that can be easily adapted to other tissue types. Its application to liver metastasis revealed interesting new findings that could be characterized further to identify potential therapeutic targets in the future.

2 Zusammenfassung

In multizellulären Organismen organisieren sich Zellen in definierten Bereichen. Eine einzelne Zelle ist niemals alleine, ihr Verhalten und ihre Genexpression ist abhängig von anderen Zellen in ihrer Umgebung. Werden diese zellulären Organisationen gestört oder verändert kann es zur Entwicklung von Krankheiten kommen. Deshalb ist es wichtig nicht nur die Identität und Genexpression einer einzelnen Zelle zu kennen (diese Informationen erhält man durch das Sequenzieren der RNA von einzelnen Zellen), sondern auch die räumlichen Beziehungen von einzelnen Zellen und Zelltypen zueinander. Die Lehre des räumlichen Transkriptoms beschreibt die Detektion von mRNA-Transkripten in räumlich definierten Regionen. In den letzten Jahren gab es eine massive Entwicklung der räumlichen Transkriptom-Technologien mit ständigen Verbesserungen und Optimierungen. Allerdings gehen mit bestehenden Methoden immer noch einige Einschränkungen einher, die hauptsächlich in der Notwendigkeit von spezialisierter Ausrüstung und Fachwissen, mangelhafter Bildgebung von einzelnen Zellen oder niedrigem Durchsatz liegen.

In dieser Doktorarbeit habe ich eine neue Methode entwickelt, sie heisst: «Sphere-sequencing (Spehre-seq)» (Kugel-seq), die einige dieser Limitierungen adressiert. Kurz zusammengefasst, frisches Gewebe von vielfältigem Ursprung und Form, wird partiell zerkleinert in zelluläre Gemeinschaften (Kugeln), die im Durchmesser zwischen 100-500 μm gross sind. Die Kugeln werden dann nach festgelegten Grössen mit einem Biosorter in 96 Multi Well Platten sortiert. Der Biosorter funktioniert wie ein normales Durchflusszytometer, aber erlaubt das Sortieren von grösseren Objekten als einzelnen Zellen. Die Kugel-Grössen werden festgelegt aufgrund von vorherigem Wissen über interessante räumliche Einheiten in komplexen Geweben. Zum Beispiel, in der Mausleber haben wir uns dazu entschieden 200-450 μm grosse Kugeln zu sortieren, damit wir räumliche Unterschiede entlang der verschiedenen Schichten des Leberläppchens analysieren können. Nach dem Sortieren werden die Kugeln in einzelne Zellen dissoziiert und mit Kugel-spezifischen Barcodes gekennzeichnet. Dafür verwenden wir eine Methode, in der die Barcodes mit Hilfe von Lipid-Ankern in die Zellmembran eindringen. Wir haben ein Set mit 288 Barcodes, also können wir 288 Kugeln per Experiment analysieren. Nach dem Sequenzieren und Demultiplexen können wir eine einzelne Zelle zu ihrer originalen Nische zuweisen aufgrund der Kugel-spezifischen Barcodes. Die anatomische Lage der Kugeln kann dann rekonstruiert werden mit Hilfe von Gen-Signaturen die Referenzpunkte darstellen oder durch die Anwesenheit oder Abwesenheit von bestimmten Zelltypen. Dieses Konzept erlaubt uns die Kugeln je nach den räumlichen Orten, von denen sie herkommen, zu gruppieren. Das können entweder nominale Gruppen sein oder die Kugeln können in einem Gradienten in einer Ordinalskala angeordnet werden. Ordinal oder nominal kategorisierte Kugeln können dann verwendet werden um räumliche Unterschiede in der Genexpression, Zelltyp-Anreicherungen oder Interaktionen zwischen Liganden und Rezeptoren zu untersuchen.

Wir haben Sphere-seq angewendet an der Mausleber um Unterschiede in der Gen Expression zwischen unterschiedlichen anatomischen Orten zu analysieren. Die Leber ist strukturell aufgebaut in Nischen. Hepatozyten, der Hauptzelltyp in der Leber, sind organisiert in sich wiederholende sechseckige Strukturen - dem sogenannten «Leberläppchen» - in denen das Blut von der Pfortader und der Leberarterie gemeinsam zur zentralen Vene fliesst. Dieser Blutfluss bewerkstelligt, dass sich ein Gradient zwischen Pfortader und Zentralvene bildet, mit

Variationen in Bezug auf Sauerstoff, Nährstoffe, Hormone und dem WNT-Signalsystem. Dies wiederum geht einher mit einer Arbeitsteilung zwischen Hepatozyten von unterschiedlichen Zonen entlang dieses Gradienten. Nicht-Parenchyme Zellen, wie zum Beispiel die Leber-Endothelial Zellen (LEZ) helfen den Hepatozyten ihre Arbeiten auszuführen. Diese unterschiedlichen Aufgaben entlang der verschiedenen Zonen des Leberläppchens benötigen unterschiedliche Genexpressions-muster. Diese Muster wurden exzessiv untersucht in Hepatozyten und LEZ. Wir haben das ausgenutzt um die räumlichen Positionen innerhalb des Leberläppchens für jede Kugel anhand der Genexpression von LEZ zu rekonstruieren. Dadurch konnten wir räumliche Genexpression in anderen Zelltypen studieren, zum Beispiel in Kupffer-Zellen (KZ), den sessilen Makrophagen der Leber. Wir konnten eine erhöhte Genexpression von *Vcam1* in KZ des Periportalfeldes von Metastasis-tragenden Mäusen identifizieren. VCAM1 ist involviert in der Rekrutierung von Immunzellen, was wiederum zur Aktivierung von KZ führen kann. Die Analyse von Interaktionen der Liganden und Rezeptoren hat diese Annahme bestärkt, weil diese vermehrten Interaktionen zwischen VCAM1 Liganden auf KZ und unterschiedlichen Integrinen auf Immunzellen im Periportalfeld vorhergesagt hat. Integrine sind wichtig für Immunzellen in dem Prozess der Bindung an und Transmigration durch die vaskuläre Wand. Wir spekulieren daher, dass eine höhere Expression von *Vcam1* zur Bindung von Immunzellen führt, welche durch die Pfortader in die Leber hineinkommen, um eine erhöhte Immunantwort aufgrund des metastatischen Prozesses zu bewerkstelligen.

Zusätzlich haben wir Sphere-seq verwendet um unterschiedliche metastatische Nischen zu studieren in einem Mausmodell, das Metastasen in der Leber entwickelt ausgehend von Krebszellen des Dick- und Mastdarms. Die Formation von Metastasen in der Leber durch Krebs im Darm ist sehr weit verbreitet, ungefähr 41 % der Darmkrebs-Patienten entwickeln Metastasen in der Leber. Dies ist so aufgrund der anatomischen Position und Struktur der Leber. Die Leber ist ein sehr vaskuläres Organ, das in der Nähe des Magen-Darm-Trakts liegt. Diese Position führt dazu, dass das Blut direkt vom Darm in die Leber fließt und zirkulierende Krebszellen können daher einfach vom Darm in die Leber fließen, wo sie sich festsetzen können und sich zu Metastasen entwickeln. Allerdings entwickeln sich nur an manchen Stellen in der Leber Metastasen, viele Bereiche bleiben ohne Metastasen, deshalb wollten wir mit Sphere-seq die Unterschiede zwischen Bereichen analysieren, die distal oder proximal zu Metastasen liegen. Wir haben dafür Kugeln gruppiert aufgrund der Anwesenheit (= Proximal) und Abwesenheit (= Distal) von metastatischen Zellen. Unsere Analyse hat gezeigt, dass in proximalen Bereichen *C1q+* Makrophagen angereichert sind, diese haben meistens einen anti-entzündlichen Phänotyp. In vorherigen Studien wurde gezeigt, dass diese spezifischen Makrophagen involviert sind in der Erschöpfung von T-Zellen und sie sind ein Indikator für eine schlechte Prognose in Krebserkrankungen. Mit einer Analyse um die Interaktion von Liganden und Rezeptoren zwischen Makrophagen und T-Zellen vorherzusagen, haben wir einer Erhöhung von Interaktionen mit den Liganden SPP1 und FN1 gefunden, welche von Makrophagen exprimiert sind. Beide sind involviert in mehreren verschiedenen Krebsarten und korrelieren mit einer schlechten Prognose.

Wir haben eine Methode mit Bildgebung angewendet, wo wir 100 verschiedene Gene gleichzeitig visualisieren können (multiplexed smFISH), in unseren Proben von metastatischen Lebern der Maus. Räumliche Unterschiede, die wir in Sphere-seq identifiziert haben, konnten wir validieren mit dieser Bildgebungs-Methode. Dies zeigt, dass wir *in situ* räumliche Unterschiede mit Sphere-seq identifizieren können.

Die Anwendung von Sphere-seq auf andere Proben wie zum Beispiel Darmkrebs-Organoiden, Maus Milz und Biopsien von Morbus Crohn Patienten hat gezeigt, dass diese Methode sehr einfach anwendbar ist für eine Vielzahl unterschiedlicher biologischer Fragestellungen. Der Vergleich mit einer der meist-angewandten Methoden für die Analyse von räumlichen Transkriptomen, Visium, hat gezeigt, dass Sphere-seq besser ist in Bezug auf die Auflösung und der Fähigkeit unterschiedliche räumliche Nischen mit vergleichbarer Qualität zu beschreiben.

Zusammengefasst, diese Arbeit zeigt die Entwicklung von Sphere-seq mit umfassenden Details zum experimentellen und die Analyse betreffenden Ablauf. Dieser kann einfach auf andere Gewebsarten adaptiert werden. Die Anwendung auf Lebermetastasen hat interessante neue Ergebnisse aufgezeigt, die in der Zukunft weiter charakterisiert werden können, um neue Therapiemöglichkeiten zu entwickeln.

3 Authors Contributions

I solely developed and designed the experimental pipeline of sphere-seq from Chapter 7.1. For the Biosorter I got some initial help to set up the sorting process in Chapter 7.1.2, from Erich Brunner (Department of Molecular Life Sciences, UZH, Zürich) and Agnieszka Wengi (Institute of Anatomy, UZH, Zürich). I also had contributed majorly to the development of the computational pipeline in Chapter 7.2, with some help from Dr. Karsten Bach. I conducted the experiments and the data analysis in all chapters, if not stated differently in this section.

Dr. Karsten Bach was writing the code for the DGE analysis between two groups and along the central-to-portal axis of the liver lobule in Chapter 7.2.2 Step 12, and the code for the analysis of the difference in cell type abundance between two groups in Chapter 7.2.2 Step 13. He also did the cell segmentation of Molecular Cartography images and the co-localization analysis between different cell types within different spatial regions in Chapter 7.3.8.4.

Dr. Xenia Ficht proofread the thesis.

Costanza Borrelli generated the GFP-positive AKP organoids for the experiments in Chapter 7.3.2.1 for using them for the organoid mixing species experiment. She was also injecting the mice for metastasis formation in Chapter 7.3.1.1 for generating the metastases-bearing livers for sphere-seq experiments.

Prof. Dr. Andreas E. Moor came up with the idea of the sphere-seq approach and contributed intellectually throughout the progress of this project with supervision and guidance.

4 Abbreviations

| | |
|---|--|
| AE – average expression | MHC – major histocompatibility complex |
| Anti-TNF – anti tumor necrosis factor | MIF – macrophage inhibitory factor |
| APC – antigen-presenting cells | MNN – mutual nearest neighbors |
| ATP – adenosine triphosphate | mRNA – messenger ribonucleic acid |
| BC – barcode | mS – milli seconds |
| BCL – binary base call | mtRNAs – mitochondrial RNAs |
| BCR – B cell receptor | MULTI-seq – multiplexing using lipid-tagged indices for single-cell RNA sequencing |
| Bp – base pair | mW – megawatt |
| BSA – bovine serum albumin | MZ – marginal zone |
| Bulk RNA-seq – bulk ribonucleic acid sequencing | NAC – neoadjuvant chemotherapy |
| CCC – cell-cell communications | NAFLD – non-alcoholic fatty liver disease |
| CD – Cluster of differentiation | NGS – next-generation sequencing |
| CD – Crohn's disease | NK – natural killer |
| cDNA – complementary deoxyribonucleic acid | NKT – natural killer T |
| CEL-seq – Cell Expression by Linear amplification and sequencing | NPCs – non-parenchymal cells |
| CITE-seq – cellular indexing of transcriptomes and epitopes sequencing | PBS – phosphate buffered saline |
| cLM genes – central vein landmark genes | PCR – polymerase chain reaction |
| CRC – colorectal cancer | PCs – principal components |
| Cre – cyclization recombinase | pLM genes – portal vein landmark genes |
| CRISPR – clustered regularly interspaced short palindromic repeats | PLP – padlock probes |
| CUT&Tag – cleavage under targets and tagmentation | PMT – photomultiplier tube |
| CV – central vein | poly(A) – polyadenylated |
| DACs – dopaminergic amacrine cells | poly(A)-tail – polyadenylated tail |
| DAPI – 4',6-diamidino-2-phenylindole | poly(dT) – poly deoxy thymidylate |
| DBiT-seq – Deterministic barcoding in tissue for spatial omics sequencing | psi – pounds per square inch |
| DCs – dendritic cells | PV – portal vein |
| DEGs – differentially expressed genes | qPCR – quantitative polymerase chain reaction |
| DGE analysis – differential gene expression analysis | RCA – rolling circle amplification |
| DMD – digital micromirror device | RNA – ribonucleic acid |
| DNB – DNA nanoball | RNA-seq - ribonucleic acid sequencing |
| EGFR – epidermal growth factor receptor | ROI – region of interest |
| EtOH – Ethanol | RPE- random priming and extension |
| ev/s – events per second | RT – reverse transcription |
| FACS – fluorescence-activated cell sorting | SBH – sequencing by hybridization |
| FBS – fetal bovine serum | SBL – sequencing by ligation |
| FFPE – formalin-fixed paraffin-embedded | SBS – sequencing-by-synthesis |
| FISH – fluorescent <i>in situ</i> hybridization | Sci-RNA-seq – single-cell combinatorial indexing RNA sequencing |
| FISSEQ – fluorescent <i>in situ</i> RNA sequencing | scRNA-seq – single-cell ribonucleic acid sequencing |
| FOCA – Fluidics and Optical Core Assembly | SEDAL – sequencing with error reduction by dynamic annealing and ligation |
| Fz – Frizzled receptor | smFISH – single molecule fluorescent <i>in situ</i> hybridization |
| GEMs – Gel beads-in-emulsions | snRNA-seq – single nuclei ribonucleic acid sequencing |
| GFP – green fluorescent protein | SOLID - sequencing by oligonucleotide ligation and detection |

| | |
|--|--|
| GI – gastrointestinal | SPACECAT – Spatially Photo Activatable Color Encoded Cell Address Tags |
| H&E – hematoxylin and eosin | Sphere-seq – sphere-sequencing |
| HCC – hepatocellular carcinoma | SPLiT-seq – split-pool ligation-based transcriptome sequencing |
| HDST – high-definition spatial transcriptomics | SPOTS – Spatial Protein and Transcriptome Sequencing |
| HSCs – hepatic stellate cells | ST – Spatial transcriptomics |
| HypISS – hybridization-based <i>in situ</i> sequencing | STARmap – spatially-resolved transcript amplicon readout mapping |
| IBD – inflammatory bowel disease | Stereo-seq – Spatial enhanced resolution sonics-sequencing |
| IHC – immunohistochemistry | STRS – spatial total RNA-sequencing |
| inDrop RNA-seq – indexing droplets RNA sequencing | TAMs – tumor associated macrophages |
| ISH – <i>in situ</i> hybridization | TCR – T cell receptor |
| ISS – <i>in situ</i> sequencing | TME – tumor microenvironment |
| KCs – Kupffer cells | TOF – time of flight |
| L-R – ligand-receptor interaction | TPLSM – two-photon laser scanning microscopy |
| LAMs – lipid-associated macrophages | Tregs – regulatory T cells |
| LCM – laser capture microdissection | tSNE – t-distributed stochastic neighbor embedding |
| LCM-seq – laser capture microdissection sequencing | UC – ulcerative colitis |
| LECs – liver endothelial cells | UMAP – Uniform Matrix Approximation and Projection |
| LED – light emitting diode | UMI – unique molecular identifier |
| LMO – lipid-modified oligonucleotide | VCT – volumetric computed tomography |
| loxP – locus of X-over in P1 | VEGF – vascular endothelial growth factor |
| LSECs – liver sinusoidal endothelial cells | VSMCs – vascular smooth muscle cells |
| LVECs – liver vascular endothelial cells | WNT – wntless-related integration site |
| MARS-seq – Massively Parallel Single-Cell RNA-Seq | WTA – whole transcriptome analysis |
| MC – Molecular Cartography | ZC – zonation coordinate |
| MERFISH – multiplexed error-robust FISH | |

5 Introduction

5.2 Spatial transcriptomics

5.2.1 What is spatial transcriptomics?

Transcriptomics is the observation of the transcriptome, the collection of all ribonucleic acid (RNA) transcripts¹. Messenger RNA (mRNA) is of special interest because it encodes proteins and enables insights into the molecular state of a biological system¹. Therefore, “spatial transcriptomics” (ST) describes the analysis of RNAs within different spatial arrangements. Biological tissues are multicellular organizations of different cell types arranged in specific ways to interact with each other to maintain tissue homeostasis. Their compositions can be altered in diseases like cancer, where cells interact in specific ways, different from a homeostatic state, to promote or fight against the diseased state. Investigating the spatial position of cells and their spatial gene expression in perturbed tissues is important for studying cellular mechanisms involved in diseased phenotypes. Using RNA sequencing (RNA-seq) of whole tissues or cell mixtures (currently known as bulk RNA-seq) comparing healthy and perturbed tissues, diverse mechanisms in cancer and other diseases could be decoded². With the development of single-cell RNA-seq (scRNA-seq), which allows studying gene expression in individual cells^{3,4}, scientists got an even closer look at changes in tissue compositions. However, during single-cell dissociation of tissues, a cell's spatial position is lost, therefore these kinds of datasets are difficult to use for structure-function analysis³. Scientists tried to overcome this limitation with the development of ST, which enables the investigation of cellular composition, transcriptional changes, and cellular interactions in a spatial context⁵. The field of ST method development has rapidly evolved to improve spatial resolution and transcriptomic profiling in a high throughput manner.

5.2.2 Historical development of spatial transcriptomics (NGS-based approaches)

As mentioned in the above section, today's use of next-generation sequencing (NGS)-based ST is heavily influenced by prior advancements in bulk RNA-seq and scRNA-seq. RNA-seq was developed more than a decade ago; the usual experimental workflow includes the following steps: RNA extraction, mRNA enrichment, cDNA synthesis, sequencing adapter ligation, library preparation, and sequencing – using primarily the Illumina short-read platform that uses a NGS approach, specifically, sequencing by synthesis with 3' blocked fluorescently labeled nucleotides⁶. Afterwards, sequencing reads are computationally aligned to the transcriptome and reads of different genes are quantified⁶. Currently, this approach is called bulk RNA-seq because whole tissues or samples of cells are sequenced in bulk. Therefore, this approach only allows the analysis of the average gene expression of a mixture of different cells. Unlike bulk RNA-seq, scRNA-seq allows resolving individual cells, their cell types and activation states. The first time scRNA-seq was reported in 2009 when Tang and colleagues sequenced the transcriptome of single mouse blastomeres (the four-cell stage embryo); they called their method “mRNA-Seq”⁷. Their method of capturing mRNA via poly-adenylated (poly(A)) capture⁷, illustrated in figure 1, was the basis of many current scRNA-seq methods. In short, single blastomere cells were manually picked and the poly(A)-tails of mRNAs were bound by poly deoxy thymidylate (poly(dT)) primers⁷. Poly(dT) primers contain an anchor

sequence that can be incorporated at the 5' end during complementary deoxyribonucleotide acid (cDNA) first-strand synthesis⁷. A poly(A)-tail was then added on the 3' end and another poly(dT) primer could bind for second-strand cDNA synthesis incorporating a second anchor sequence⁷. Then double-stranded cDNA was amplified, fragmented and adaptors for sequencing were ligated⁷. After final library amplification, cDNA fragments were sequenced using sequencing by oligonucleotide ligation and detection (SOLiD) sequencing⁷.

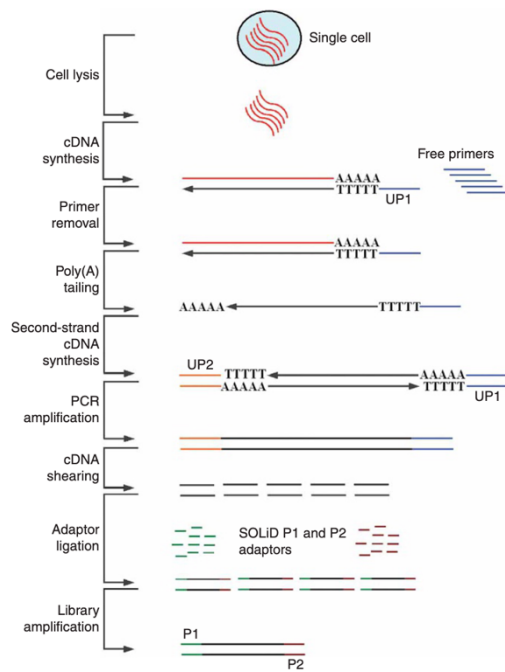


Figure 1: Schematic drawing of the mRNA-Seq workflow.

A single cell, picked manually, is lysed and its mRNA is transcribed into cDNA using poly(dT) primer (=poly(A) capture). An anchor sequence (UP1) is incorporated in the poly(dT) primer resulting in anchor sequence 1 on the 5' end. On the 3' end of the first strand cDNA, a poly(A)-tail is added and during the following second strand cDNA synthesis, a poly(dT) primer with the second anchor sequence (UP2) can bind. UP1 and UP2 primers can then be used for cDNA amplification following fragmentation and adaptor ligation for later SOLiD sequencing. Fragments with adaptors are then amplified in a final library amplification step. Illustration from Tang F. et al., mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6, 377-382, 2009, by permission of Springer Nature (<https://www.nature.com/nmeth/>).

Poly(A) based mRNA capture in single cells was then adapted by many high throughput scRNA-seq technologies to incorporate a cell barcode (BC) and unique molecular identifier (UMI)⁸⁻¹⁴. This results in labeling all mRNA transcripts of a single cell with the same cell BC and each polymerase chain reaction (PCR) transcript with a UMI to account for PCR biases¹⁵. To archive capturing of single cells there are different approaches:

- Microfluidics: A bead with barcoded poly(dT) primers and a single cell are encapsulated in oil droplets and during cell lysis mRNA transcripts are captured by the bead; for example Drop-seq¹⁰, inDrop (indexing droplets) RNA-seq¹¹ and the commercial 10X Genomics Chromium System;
- Plate-based: Single cells are sorted into wells of a 96 or 384 well plate containing poly(dT)-cell-BC-UMI primer; for example MARS-seq (Massively Parallel Single-Cell RNA-Seq)^{13,16} and CEL-seq (Cell Expression by Linear amplification and Sequencing)^{12,17};
- Micro- or nano-well based: Barcoded beads and single cells are trapped into micro- or nano-wells and there is only space for one of each in every single well; for example, Seq-well¹⁴ and the commercial BD Rhapsody Single Cell Analysis System;
- Methods where single cells are labeled in a split-and-pool approach; for example SPLiT-seq (split-pool ligation-based transcriptome sequencing)¹⁸ and sci-RNA-seq (single-cell combinatorial indexing RNA sequencing)¹⁹;

For higher throughput, reduction of costs per sample and batch effect, cell hashing methods were developed where single cells of a sample are labeled with a sample barcode, allowing multiple samples to be pooled for one single-cell capture run^{20,21}. The sample barcode is a stretch of oligonucleotides with an incorporated poly(A) stretch that can be captured with

scRNA-seq poly(A) capture. Two examples are CITE-seq (cellular indexing of transcriptomes and epitopes sequencing), where barcodes are attached to antibodies that can bind to cell surface epitopes²⁰ and MULTI-seq (multiplexing using lipid-tagged indices for single-cell RNA sequencing), where barcodes are attached to lipid anchors that can easily integrate into a cell's phospholipid bilayer membrane²¹.

The advancements in high throughput and increased mRNA capture efficiency per single cell led to many discoveries and atlases from humans²² and mice²³ of cell types, cell subtypes, and cell states within various healthy and diseased tissues (reviewed in Zhang *et al.*, 2021 and Jovic *et al.*, 2022^{24,25}). However, due to single-cell dissociation prior to mRNA capture, the spatial location of a single cell is lost. To preserve spatial information, the field of ST evolved. For example, important progress in the field was made by Raj *et al.* in 2008 by developing a method to visualize single mRNA molecules *in situ*²⁶ using multiple singly labeled probes and Ståhl *et al.* in 2016 who developed a method called “Spatial Transcriptomics” (ST) where mRNA transcripts of cells are captured *in situ*²⁷. Many technologies have been developed since then, with constant improvements in resolution and throughput. They will be discussed in the following section.

5.2.3 Spatial transcriptomics approaches

ST technologies can be broadly categorized into imaging-based and mRNA capture-based (reviewed in Lee *et al.*, 2022²⁸). Imaging-based methods are based on single-molecule *in situ* hybridization (ISH) or *in situ* sequencing (ISS) while mRNA capture-based methods are typically based on capturing mRNAs with spatial barcodes²⁸. A comparison of these two approaches is illustrated in figure 2. Additionally, computational methods reconstruct spatial cellular organizations from scRNA-seq data (reviewed in Kleino *et al.*, 2022)⁵.

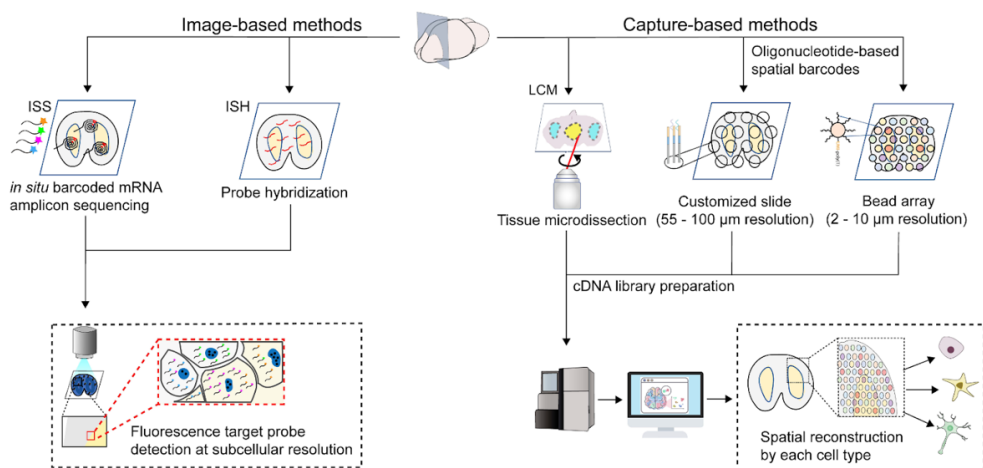


Figure 2: Comparison of image-based (left) and capture-based (right) ST methods. Image-based methods entail *in situ* sequencing (ISS), where barcoded mRNA is identified by mRNA amplicon sequencing, and *in situ* hybridization (ISH), where mRNA transcripts are identified by complementary probe hybridization. Readout of both methods is generated by high-resolution microscopy allowing mRNA molecule detection at subcellular resolution. The capture-based approach can be divided into three subcategories: laser capture microdissection (LCM), where *in situ* regions of interest are obtained directly, capturing of mRNAs of tissue sections via spatial barcodes printed directly on the slide, or via bead arrays. Readouts of these methods are generated via NGS and mapping of transcript reads back to the tissue position. There are methods with various resolutions from single-cell resolution to broad tissue regions. Illustration from Lee J. *et al.*, Recent advances in spatially resolved transcriptomics: challenges and opportunities. *BMB Rep.* 55, 113-124, 2022, under permission of <http://creativecommons.org/licenses/by-nc/4.0>.

5.2.3.1 Imaging-based

Multiplexed imaging-based ST methods can be divided into ISH and ISS approaches; both follow similar workflows (highlighted in figure 3). Detection of transcripts can be targeted by designing hybridization probes complementary to target sequences (adapted from fluorescence *in situ* hybridization [FISH]²⁹ or targeted padlock probes for ISS) or untargeted by using random primers for ISS³⁰. Encoding can be achieved with two different approaches: a linear method, where a single gene is visualized in every round of imaging and an exponential method, where multiple genes are detected in every round of imaging³⁰. Acquired images are then used for downstream analysis that incorporates the following steps: Images of multiple rounds have to be aligned and spots have to be registered; registered spots are then used for spot calling and decoding to assign a barcode for each spot that defines a gene's identity³⁰. The principles of ISH and ISS methods will be explained in the next section with given examples of methodologies.

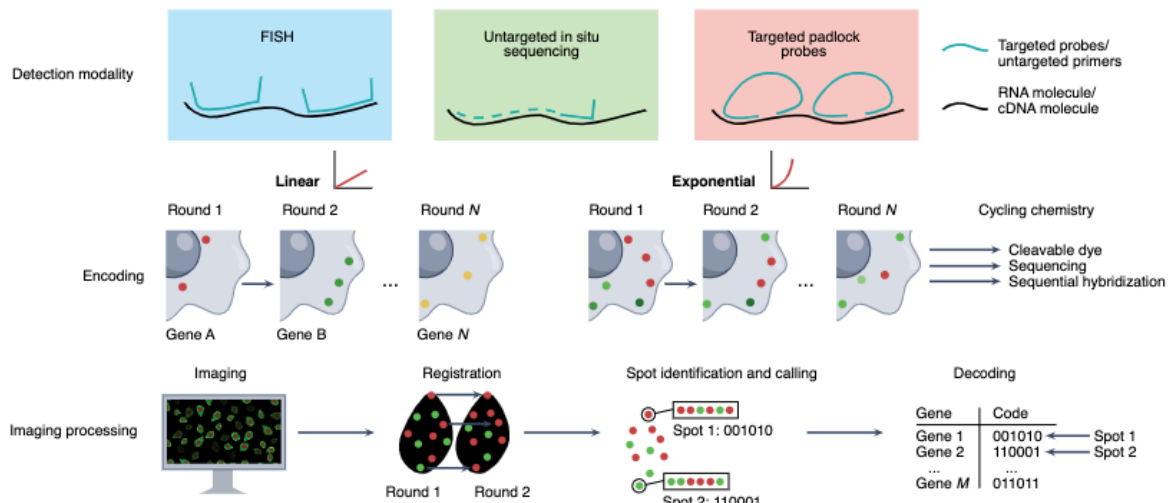


Figure 3: Comparison of different imaging-based ST technologies. mRNAs can be detected with three different approaches: Fluorescent *in situ* hybridization (FISH) approach where a probe complementary to the mRNA molecule hybridizes and can be visualized with either a fluorophore that is bound by the probe or via readout probe that hybridizes to the target probe; *In situ* sequencing can be either untargeted using untargeted primer or targeted using targeted padlock probes. For multiplexing, encoding can be either linear when a different gene is labeled in each round, or exponential when multiple RNAs are labeled in a single round. This can be done using different chemistries, cleavable dyes, *in situ* sequencing, or sequential hybridization. Image processing: Data is collected with fluorescent imaging, different rounds of imaging have to be aligned and spots have to be assigned to each imaging round, then gene identification is defined by spot calling and decoding based on several imaging rounds. Illustration from Tian L. et al., *The expanding vistas of spatial transcriptomics*. *Nature Biotechnology*, 2022, by permission of Springer Nature (<https://www.nature.com/nbt/>).

5.2.3.1.1 *In situ* hybridization (ISH)

ISH-based ST methods are built on fluorescence *in situ* hybridization (FISH)²⁹. In this method, oligonucleotide sequence probes complementary to a target mRNA labeled with multiple fluorophores are designed, which bind the mRNA and can be visualized with fluorescence microscopy²⁹. However, due to the large fluorescent signal of multiple labeled probes, it was difficult to detect single molecules. This was circumvented by Raj *et al.*, developing an approach (single molecule FISH, smFISH) where each target mRNA is bound by 48 or more short singly labeled probes²⁶. However, with this method, only a limited number of different mRNA molecules can be visualized at the same time due to the limitation of how many different fluorophores can be spectrally separated by the available optical setup. To allow a higher

throughput, MERFISH (Multiplexed error-robust FISH) was developed^{31,32}. In short, a set of probes are designed to bind target mRNA molecules; the probes do not only contain complementary oligonucleotide sequences to hybridize to target mRNAs they also contain a readout sequence^{31,32}. The readout sequence is bound by readout probes in successive rounds of hybridization and imaging, resulting in a unique, binary barcode for each target RNA^{31,32}. MERFISH allowed the detection of 140-1001 genes simultaneously due to constraints in optical resolution^{31,32}. SeqFISH+ overcame this limitation by making use of 60 pseudo-colors within three fluorescent channels, which allowed the detection of 10,000 genes³³. There are also other ISH ST approaches, some of which are highlighted and compared in table 1.

| Method | Features | Resolution | Number of genes/UMIs | References |
|-------------------|---|--------------|--|---|
| MERFISH | Sequential rounds of probe hybridization and visualization; error detection and/or correction | Sub-cellular | 140 genes with and 1001 genes without error correction | Chen <i>et al.</i> , 2015 Science ³² ; Moffitt <i>et al.</i> , 2016 PNAS ³¹ |
| seqFISH+ | Sequential rounds of probe hybridization and detection, increased detection quantity by usage of pseudocolors | Sub-cellular | ~10,000 genes | Eng <i>et al.</i> , 2019 Nature ³³ |
| Split-FISH | Multiplexed FISH in combination with split probes where two probes have to bind the target in order to be detected by a bridge probe that is read by a readout probe; reduction of off-target effects | Sub-cellular | ~317 genes | Goh <i>et al.</i> , 2020 Nat. Methods ³⁴ |
| EEL FISH | RNA is transferred to a glass slide prior to probe hybridization using electrical force; reduces time-consuming z-axis imaging | Sub-cellular | ~448 genes per color channel | Borm <i>et al.</i> , 2022 Nat. Biotech. ³⁵ |

Table 1: Comparison of selected ISH ST methodologies.

5.2.3.1.2 *In situ* sequencing (ISS)

In comparison to ISH ST methods, ISS ST methods are not limited by the number of fluorophores and the optical diffraction limit, which is the smallest physical distance required to distinguish between two objects in an optical setup^{26,28,29}. Nucleotide sequences of RNA molecules are detected by sequencing-by-ligation (SBL), for example, in STARmap (spatially-resolved transcript amplicon readout mapping)³⁶, and FISSEQ (fluorescent *in situ* RNA sequencing)³⁷, or by sequencing-by-hybridization (SBH) in HybISS (hybridization-based *in situ* sequencing)³⁸. In short, SBL (shown in figure 4A on the example of FISSEQ) works as follows: Sequencing primer hybridizes to adapter sequences of *in situ* amplified cDNA followed by sequential rounds of fluorescent probe ligation, detection by confocal imaging and cleavage of fluorescent signal; a barcode of signals is then decoded for each spot and gene³⁷. SBH (illustrated in figure 4B on the example of HybISS) uses bridge probes that hybridize to *in situ* amplified cDNA and fluorescent readout-probes hybridize to bridge probes and are read out with fluorescence microscopy³⁸. The hybridization approach of bridge and readout probes in SBH leads to increased fluorescent signals that can be visualized by standard epifluorescence microscopes instead of confocal microscopes for SBL³⁸. mRNA detection follows different

strategies. One strategy is *in situ* reverse transcription (RT) of mRNAs using random hexamers or targeted primers, where adapter sequences for later sequencing are incorporated, following cDNA amplification using rolling circle amplification (RCA) of ligated first strand cDNA (used, for example, in FISSEQ)^{28,37} shown in figure 4C. Another approach is the use of so-called padlock probes (PLP) which are oligonucleotide probes that have two target-complementary segments that are connected by a linker sequence and after the target recognition the ends of the probe are ligated to form a circularizes molecule, this reduces background and results in an increased detection efficiency³⁹. PLPs are then ligated to the target cDNA and RCA is initiated (used in STARmap³⁶ and HybISS³⁸), shown in figure 4D. Different ISS ST technologies are summarized and compared in table 2.

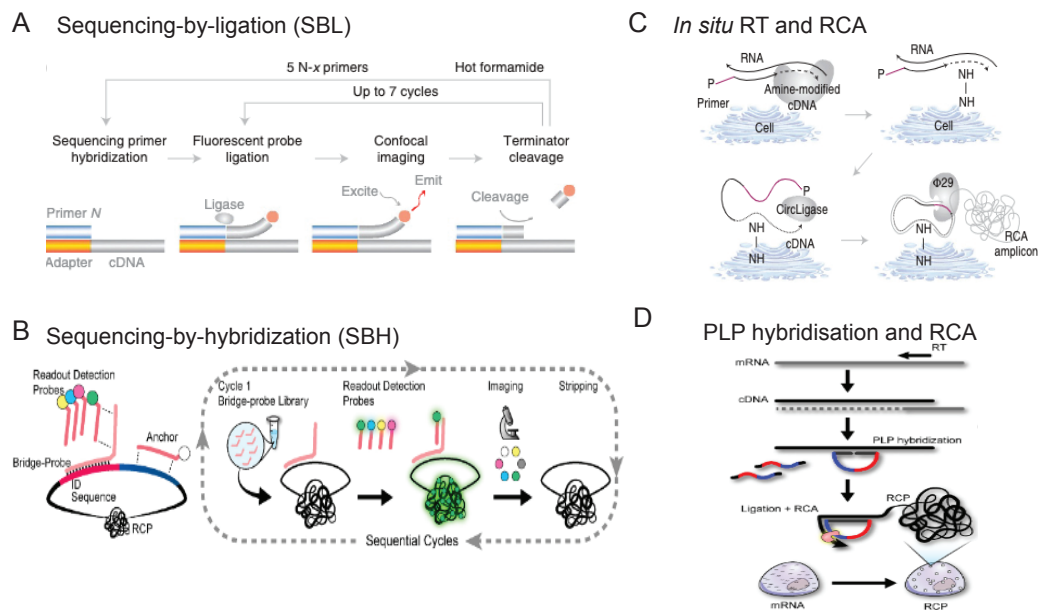


Figure 4: Overview of different approaches for ISS ST methodologies. **A)** Sequencing-by-ligation using sequencing primer hybridization and sequential rounds of fluorescent probe ligation, imaging, and cleavage. **B)** Sequencing-by-hybridization using a combination of bridge- and readout-probes. Signal detection happens by sequential rounds of first bridge-probe hybridization following read-out probe hybridization, imaging of fluorophore from read-out probe, and stripping. **C)** *In situ* RT and rolling circle amplification (RCA): Primer for RT bind in a targeted or untargeted way following RCA. **D)** Padlock probes (PLP) hybridize to cDNA after RT initiating RCA. Illustration A and C from Lee J. H. et al., Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* 10, 442-458, 2015, by permission of Springer Nature (<https://www.nature.com/nmeth/>). Illustration B and D from Gyllborg D. et al., Hybridization-based *in situ* sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissues. *Nucleic Acids Res.* 48, e112, 2020, by permission of Oxford University Press.

| Method | Features | Resolution | Number of transcripts | References |
|----------------|--|--------------|-----------------------|---|
| FISSEQ | <i>In situ</i> RT, RCA and SBL | Sub-cellular | ~200 | Lee et al., 2015 <i>Nat. Protocol</i> ³⁷ |
| STARmap | No RT, PLP hybridization to mRNA, RCA and SBL using SEDAL sequencing (sequencing with error-reduction by dynamic annealing and ligation) | Sub-cellular | 160-1020 | Wang et al., 2018 <i>Science</i> ³⁶ |

| | | | | |
|---------------|---|--------------|-------------|--|
| HybISS | <i>In situ</i> RT, PLP hybridization, RCA and SBH | Sub-cellular | ~119 | Gyllborg <i>et al.</i> , 2020 Nucleic Acids Res. ³⁸ |
| ExSeq | Combines expansion microscopy ⁴⁰ with FISSEQ ³⁷ ; signals very close to each other can be differentiated better | Sub-cellular | Up to 3,000 | Alon <i>et al.</i> , 2021 Science ⁴¹ |

Table 2: Comparison of selected ISS ST methodologies.

5.2.3.1.3 Gene panel design for targeted imaging-based ST

For a targeted imaging-based ST approach, a gene panel has to be designed prior to the experiment. There are many things to consider when choosing the gene panel because the number of genes to detect are limited. First of all, cell type markers have to be included that are unique in order to confidently distinguish cell types. Additionally, the expression level and length of the mRNA has to be taken into account. If the mRNA is too short, only few probes can hybridize which might generate a signal too weak to detect and if the expression level is too high it might be challenging to detect single mRNA molecules^{42,43}. For example, for MERFISH these cutoffs have been determined to be the possibility to design at least 48 hybridization probes and an upper limit of expression level of 100-200 mRNA molecules per cell, which can be converted to 10-20 UMI counts for Drop-seq¹⁰ for example, where the capture rate is 10 %^{42,43}.

5.2.3.2 mRNA capture-based

Methods involving mRNA capture are usually of lower resolution (single cells to multiple cells) than imaging-based methods (sub-cellular to single cell), however, they can be used in a high throughput, unsupervised manner²⁸. In comparison to imaging-based targeted ST approaches they do not rely on gene panel design, mRNAs can be captured in an explorative way. Generally, these methods are based on the same principle as scRNA-seq, the capturing of mRNA with poly(dT) primers introducing a spatial or single cell barcode^{3,28}. However, the mechanism of how mRNA gets in contact with the poly(A) capture primers is different. Different approach can be subdivided broadly into three categories: mRNA molecules can be captured from tissue slices with spatial indices using arrays or microfluidics; mRNA molecules can be captured via regions of interest (ROIs), which can be labeled with for example photoactivatable dyes following single cell extraction, which is then followed by sorting of labeled cells, and scRNA-seq⁴⁴; or mRNA molecules can be captured⁴⁴ from small cellular communities in bulk after partial dissociation.

5.2.3.2.1 Spatial indexing

Spatial indexing of fresh frozen tissue slices is used by various different ST technologies highlighted in figure 5 and table 3. In general, mRNA molecules are captured from thin (10 µm in most methods) fresh frozen tissue slices with spatial primers that include a poly(dT) stretch, UMI, and a spatial barcode sequence, either attached on a glass slide or applied using microfluidics⁴⁴. Different approaches result in different resolutions (compared in table 3). The ST method developed by Ståhl *et al.*, later adopted by the 10X Genomics platform Visium, uses arrays of indexing primers printed in defined spatial positions, so-called spots^{27,45}. However, in Ståhl *et al.* each spot captured an area of a mixture of up to 100 cells, which was improved to 10 cells in Visium²⁸. This has been even further improved by the development of

Slide-seq where barcoded beads like the ones used for Drop-seq¹⁰, are printed on a rubber-coated glass slide⁴⁶. The identity of the spatial barcoded bead is then determined by SOLiD (SBL) sequencing and after the mRNA of frozen tissue slices is captured with beads, the library preparation is done following the Drop-seq procedure⁴⁶. In another approach, the barcoded poly(d)T primer is amplified into clusters either in an arrayed manner, like in HDST (high-definition spatial transcriptomics) where clusters are generated in wells 2 μm of size and spatial positions are identified by sequential hybridization using fluorescent decoder oligonucleotides (SBH)⁴⁷ or in a continuous manner using solid-phase amplification like in Seq-Scope where spatial positions are identified by sequencing-by-synthesis (SBS)⁴⁸. SBS is adopted by the Illumina platforms and decodes oligonucleotide sequences by sequential rounds of adding all four nucleotides labeled with different fluorophores and imaging, the fluorophore is cleaved off and another round of hybridization takes place (reviewed in Goodwin *et al.*, 2016⁴⁹). Stereo-seq (Spatial enhanced resolution sonics-sequencing) uses DNA nanoball (DNB)-patterned arrays, and positions are defined by DNB sequencing, enabling nanometer capture resolution⁵⁰. DNB sequencing was developed by MGI technologies. In short, small DNA fragments are circularized and amplified into DNA nanoballs where sequencing primers can bind to adapter sequences, and then DNA sequences are encoded by sequential rounds of adding fluorescently labeled probes, imaging, and removal.

Another approach to spatially introduce barcoded poly(d)T primer to capture fresh frozen tissue slices is microfluidics, this is employed by DBiT-seq (Deterministic barcoding in tissue for spatial omics sequencing)⁵¹. In this method, a tissue slice is trapped in a gasket where microfluidics channels supply the permeabilized tissue with two sets of barcodes in two perpendicular directions, each spot resulting in having a unique combination of two barcodes⁵¹.

Most of the ST methods require fresh frozen tissue slices, however, there are developments towards the usage of FFPE fixed tissues, for example in an adapted version of Visium⁵².

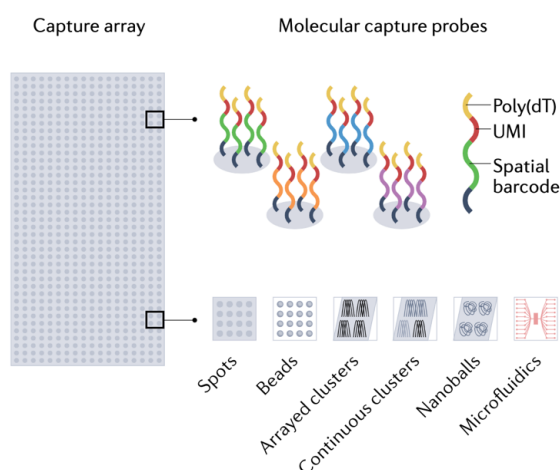


Figure 5: Schematic drawing of spatial indexing approaches using molecular capture probes: a combination of poly(dT), UMI and spatial barcode oligonucleotide sequences. Indexing probes are printed on a glass slide in different organizations: spots, beads, arrayed clusters, continuous clusters, or nanoballs. Probes can also be applied to tissue sections via microfluidics. Illustration from Moffitt J. R. *et al.*, *The emerging landscape of spatial profiling technologies*. *Nat. Rev. Genet.*, 2022, by permission of Springer Nature (<https://www.nature.com/nmeth/>).

| Method | Features | Resolution | Number of transcripts/UMI | References |
|--------|---|---|--|--|
| ST | Spatial barcodes are printed within spots on a glass slide; Spatial barcode position is predefined; | 100 μm , 200 μm center-to-center = 50-100 cells per spot | 1,000-3,000 transcripts per spot | Ståhl <i>et al.</i> , 2016 Science ²⁷ ; Salmen <i>et al.</i> , 2018 <i>Nat. Protocol</i> ⁴⁵ ; |

| | | | | |
|------------------------------|--|---|------------------------------------|--|
| 10X genomics Visium platform | Based on ST by Ståhl <i>et al.</i> , 2016 ²⁷ ; Spatial barcode primers in predefined spots; | 55 μm ; 100 μm center-to-center = 1-10 cells per spot | ~4,500 transcripts per spot | Ståhl <i>et al.</i> , 2016 Science ²⁷ ; Salmen <i>et al.</i> , 2018 Nat. Protocol ⁴⁵ ; |
| Slide-seq | Spatial barcoded beads printed on a glass slide; Spatial position decoded by SBL; | 10 μm ; 10 μm center-to-center = single-cell | ~59 UMIs per bead | Rodrigues <i>et al.</i> , 2019 Science ⁴⁶ |
| HDST | Arrayed clusters of spatial barcoded primers in 2 μm wells; Spatial position decoded by SBH; | 2 μm = sub-cellular | ~190 UMIs per cell | Vickovic <i>et al.</i> , 2019 Nat. Methods ⁴⁷ |
| Slide-seq V2 | Same approach like Slide-seq ⁴⁶ , but higher capture efficiency; Spatial position decoded by SBL; | 10 μm ; 10 μm center-to-center = single-cell | ~500-1,000 UMIs per bead | Stickels <i>et al.</i> , 2021 Nat. Biotech. ⁵³ |
| Seq-Scope | Continuous clusters of spatial barcoded primers; Spatial position decoded by SBS; | 0.5-0.8 μm center-to-center = sub-cellular | 4,700 UMIs per cell | Cho <i>et al.</i> , 2021 Cell ⁴⁸ |
| PIXEL-seq | Continuous clusters of spatial barcoded primers; Spatial position decoded by SBS; | ≤ 1 μm = sub-cellular | <1,000 UMIs /10x10 μm^2 | Fu <i>et al.</i> , 2022 Cell ⁵⁴ |
| Stereo-seq | Spatial barcoded primers arranged in DNB-patterned arrays; Spatial position decoded by DNB sequencing | 220 nm; 500-715 nm center-to-center = sub-cellular | ~1,450 UMIs per 10 μm | Chen <i>et al.</i> , 2022 Cell ⁵⁰ |

Table 3: Comparison of selected spatial indexing mRNA capture-based ST technologies.

5.2.3.2.2 Regions of interest

A second widely applied approach that evolved in ST technology development was the labeling of ROIs using photo-activation followed by sorting of labeled cells and scRNA-seq or isolating of ROIs using other strategies^{28,44}. Selected methods are summarized in table 4. The approach of isolating ROIs was first employed by LCM-Seq (Laser capture microdissection sequencing) which is based on LCM (laser-capture microdissection)⁵⁵ using infrared or ultraviolet lasers to isolate ROIs, following RNA-seq using the SMART-seq2^{56,57} scRNA-seq protocol, which is a plate-based approach for full-length transcriptomic studies⁵⁸. Another approach is labeling of ROIs using photoactivation. These methods rely on the concept of photoactivatable tags that can be illuminated in defined regions^{28,44}. After activation, tissues get dissociated and illuminated cells can be sorted and used for scRNA-seq^{28,44}. The first method with this approach was NICHE-seq using a reporter mouse expressing photoactivatable GFP that could be activated by two-photon laser scanning microscopy (TPLSM)⁵⁹. A similar approach is used by SPACECAT (Spatially Photo Activatable Color Encoded Cell Address Tags) where samples are stained with photoactivatable dyes⁶⁰. In methods like ZipSeq⁶¹ and Light-Seq⁶², illumination releases caged oligonucleotide sequences on tissues that can then be hybridized to fluorophores or primer sequences.

| Method | Features | Resolution | Number of transcripts/UMI | References |
|------------------|---|---------------------------------------|---|--|
| LCM-Seq | ROI extraction using LCM ⁵⁵ ; sequencing of ROIs using SMART-seq2 ^{56,57} ; | Down to one cell | Up to 4,958 transcripts per ROI | Nichterwitz <i>et al.</i> , 2016 Nat. Com. ⁶³ |
| NICHE-seq | ROI extraction by photoactivation of GFP from a transgenic mouse and FACS; sequencing using MARS-seq ¹⁶ ; | Around 100 μ m, then single cells | 200-1,500 per cell | Medaglia <i>et al.</i> , 2017 Science ⁵⁹ |
| SPACECAT | ROI extraction by photoactivation of dyes applied to tissues and FACS; sequencing using Seq-Well ¹⁴ ; | 10-20 cells+, then single cells | ~6,000 transcripts per cell | Genshaft <i>et al.</i> , 2021 Nat. Com. ⁶⁰ |
| ZipSeq | ROI extraction by uncaging a complementary sequence for binding a primer for poly(A) capture; sequencing by Chromium 10X Genomics platform; | 200-400 μ m, then single cells | 500-6,000 transcripts per cell | Hu <i>et al.</i> , 2020 Nat. Methods ⁶¹ |
| Light-Seq | ROI labeling <i>in situ</i> by uncaging DNA barcodes that can then be sequenced using <i>ex situ</i> NGS; | 4-1,000+ cells | 1,000-10,000 UMIs per 10x10 μ m ² area | Kishi <i>et al.</i> , 2022 Nat. Methods ⁶² |

Table 4: Comparison of selected ROI capture-based ST technologies.

5.2.3.2.3 Partial dissociation

The third approach uses fluorescence-activated cell sorting (FACS) to isolate physically interacting pairs or communities of cells (2-10 cells) after partial dissociation of tissues, followed by sequencing of sorted communities in bulk using the MARS-seq scRNA-seq protocol^{13,16,64-66}. After sequencing, transcriptomes of single cells are analyzed using computational deconvolution approaches. The spatial location of cells can then be reconstructed using so-called landmark genes, meaning their expression is specific for certain spatial locations^{65,66}. Furthermore, ligand-receptor interactions between pairs of cells can be investigated⁶⁴. For example, in pcRNAseq (paired-cell sequencing) the authors isolated heterogeneous pairs of hepatocytes and liver endothelial cells (LECs) and then they used zoned landmark gene expression in hepatocytes to reconstruct the position of LECs⁶⁵. A zoned gene expression signature in LECs could then be identified⁶⁵.

| Method | Features | Resolution | Number of transcripts/UMI | References |
|-----------------|--|----------------|---------------------------|--|
| pcRNAseq | FACS sorting of pairs of cells; sequencing using MARS-seq ¹⁶ ; | Pairs of cells | 200-1,500 per cell | Halpern <i>et al.</i> , 2018 Nat. Biotech. ⁶⁵ |
| PIC-seq | FACS sorting of pairs of cells; sequencing using MARS-seq2.0 ¹³ ; | Pairs of cells | ~700 transcripts per cell | Giladi <i>et al.</i> , 2020 Nat. Biotech. ⁶⁴ |
| ClumpSeq | FACS sorting of communities of ~10 cells; sequencing using MARS-seq2.0 ¹³ ; | 2-10 cells | ~700 transcripts per cell | Manco <i>et al.</i> , 2021 Nat. Com. ⁶⁶ |

Table 5: Comparison of selected partial dissociation capture-based ST technologies.

5.2.3.3 Computational methods to infer the spatial location of a cell

Using computational approaches, single cells from scRNA-seq can be integrated with spatial information to infer their spatial location. One approach to do so is using a small set of landmark genes acquired by ISH or methods like LCM-Seq following integration with scRNA-seq data to reconstruct the spatial location of single cells⁶⁷⁻⁶⁹. This approach is highlighted in figure 6 on the example of the small intestinal villus where landmark gene signatures of cells from different zones within the crypt-villus axis are generated using LCM-Seq⁶⁹. These were then used for the reconstruction of the anatomical origin of individual cells from scRNA-seq⁶⁹. Using more extensive mathematical models many other tools were developed to integrate large scRNA-seq datasets with small samples of ST data (reviewed in Kleino *et al*, 2022⁵). Two examples are novoSpaRc, which take into consideration that nearby cells are often more similar⁷⁰, and DEEPsc which uses a deep learning-based algorithm to impute spatial structures of scRNA-seq data using a spatial reference⁷¹.

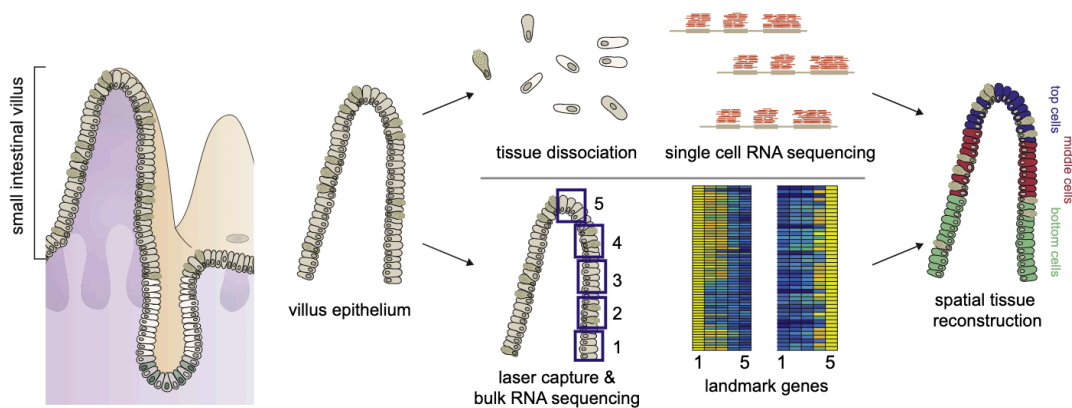


Figure 6: Schematic illustration of spatial reconstruction of single cells from the intestinal villi using landmark gene signatures. Small intestinal villi are used for scRNA-seq and LCM-seq of different zones. Genes detected by LCM-seq of different zones can be used as landmark genes to reconstruct the position of single cells along the crypt-villus axis. Illustration from Moor A.E. *et al.*, *Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis*. *Cell*, 175, 1156-1167, 2018, by permission of Elsevier Science & Technology Journals.

5.2.3.4 Spatial methods capturing mRNA plus other modalities

In most recent ST technology developments, many scientists try to capture mRNA in combination with other modalities. One example is a method capturing all RNA molecules (coding RNAs, noncoding RNAs, and viral RNAs), called STRS (spatial total RNA-sequencing)⁷². This is achieved by enzymatic *in situ* polyadenylations, adding poly(A) tails onto RNA molecules that are not polyadenylated and these can then be captured using the Visium standard protocol⁷². Another example is Spatial CUT&Tag⁷³ where DBiT-seq⁵¹ is combined with cleavage under targets and tagmentation (CUT&Tag) chemistry^{74,75} to detect histone modifications. Thereby, histone modifications are bound by antibodies, followed by the ligation of adapters to the genomic DNA of the position where the antibodies bind⁷³. Then adapters are labeled with a spatial barcode following the DBiT-seq workflow⁷³. DBiT-seq was also adapted to work for proteins in a method called Spatial-CITE-seq, by applying DNA-conjugated antibodies to the tissue slice before spatial barcoding⁷⁶. This workflow was also adapted for Visium in SPOTS (Spatial PrOtein and Transcriptome Sequencing)⁷⁷.

5.2.4 Current limitations

The field of ST is relatively new and although there are constant developments and improvements, there are still many limitations that come along with different ST methods.

Imaging-based methods are limited by the optical diffraction limit and the auto-fluorescent background signal of some tissue types can impair the fluorescent signal of labeled mRNA molecules to be detected properly⁵. Optical crowding is often challenging when analyzing large panels of targeting probes⁵; this problem has been improved by ExSeq, which uses expansion microscopy prior to *in situ* sequencing⁴¹. Probe design is another challenge with targeted imaging-based technologies. SeqFISH+³³ requires at least 1kb large mRNA molecules for probe design, MERFISH^{31,32} needs 100-200 nucleotides⁵. Detection rates can often be increased by multiple probes binding the same mRNA molecule; however, this is limited by the length of mRNA transcripts⁷⁸. Additionally, secondary structures and base composition can impact accessibility for probe binding⁷⁸. ISH and some ISS methods are often targeted; therefore, prior knowledge is needed for panel design.

mRNA capture-based spatial-indexing methods are unsupervised but often lack single-cell resolution. A single spot in Visium^{27,45} captures up to 10 cells. This is especially problematic when studying small cells with low amounts of mRNA molecules which could be crowded out by mRNA from larger cells. Hildebrandt *et al.*, saw that in ST experiments of the liver there were low signals of liver endothelial cells (LECs) because mRNA molecules of the larger hepatocytes were captured preferentially⁷⁹. Halpern *et al.* found that LECs had 23-fold lower total numbers of UMIs compared to hepatocytes in their scRNA-seq experiments⁶⁵, therefore the chance that mRNA molecules of hepatocytes are captured in ST experiments is much higher. Even though array-based methods like Slide-seq⁴⁶ have a single-cell resolution, one capture area could still capture a mix of two cells because its position could lie between a border region of two cells. It is difficult to truly acquire single-cell resolution confidently. Computational methods like SPOTlight⁸⁰, SpatialDecon⁸¹, or Cell2Location⁸² can be used to analyze spot cell type composition or to assign individual cell types to spatial areas⁵, however, these are highly dependent on the quality of single-cell references and differences in gene expression of genes that are expressed within more than one cell type are difficult to identify. Another challenge in regards to that is that during tissue permeabilization mRNA molecules can diffuse⁵ and bind to distant capture areas contaminating other spatial areas.

In comparison, many imaging-based methods even have a sub-cellular resolution. However, to assign signals to single cells, cell-segmentation procedures are required. Standard segmentation tools use specific markers like for example DAPI (4',6-diamidino-2-phenylindole) staining for the nucleus to assign imaging pixels to cytoplasmic, nuclear, and empty regions⁵. Another approach is the usage of membrane markers, for example Lohoff *et al.* used a combination of β -catenin, E-cadherin, pan-cadherin and N-cadherin to label cellular borders in combination with DAPI staining for cell segmentation⁸³. The assignment can be challenging for cells with different shapes, labeling cell borders might be superior over DAPI only staining in this case. Additionally, it might be difficult to segment cells from very densely organized structures. The authors of the development of Seq-Scope, for example, discuss that it is challenging to do segmentation in liver samples to divide NPCs from hepatocytes because hepatocytes are much bigger compared to NPCs⁴⁸. For Stereo-seq, the authors claim that,

especially for densely packed tissue areas of small cells like immune cells, cell segmentation can be challenging⁵⁰.

Many ROI-based methods have single cell resolution by dissociating selected ROIs into single cells previous to single cell capture. However, often, only a limited number of regions can be analyzed at the same time due to constraints in fluorescent signal combinations or different illumination properties. For example, ZipSeq allows analysis of a maximum of four regions simultaneously⁶¹ and Light-Seq allows three rounds of light-directed barcoding, thereby labeling three regions⁶². Additionally, these methods are highly supervised in terms of spatial regions, ROIs are predefined by known microanatomical structures and thereby these methods are blind to unknown spatial tissue heterogeneities.

Partial dissociation-based methods like PIC-seq⁶⁴, pcRNAseq⁶⁵, and Clump-seq⁶⁶ lack single-cell resolution, and thereby they face similar challenges of deconvolution as mentioned for ST array-based methods.

In both, imaging- and mRNA capture-based spatial indexing technologies, transcript detection is often limited by the thickness of 10 μm tissue slices. Sometimes this does not even include an entire cell, for example, the diameter of a hepatocyte within the liver is approximately 20-30 μm ⁸⁴. Three-dimensional information is lost and due to high costs per experiment, studies are often restricted to a couple of hand-picked slices resulting in analysis of a couple of flat areas within an organ. Additionally, tissues have to be freshly frozen or otherwise preserved to maintain mRNA integrity and sectioning is often tricky and difficult.

Many ST technologies are limited by their instruments which are often expensive and unavailable to a broad scientific community. This is especially true for imaging-based methods where high-resolution microscopes with automated fluidics are needed⁵. NGS costs are lower, however, with higher resolution also the sequencing cost increases. Additionally, in both cases, often complex computational methods are needed for image and data analysis, which are often not accessible for many scientists³.

5.3 Ligand-receptor interaction analysis

5.3.1 Different types of ligand-receptor interactions

Cells communicate with each other in defined manners to develop and maintain tissue homeostasis. These interactions are often impaired and changed when diseases like cancer develop. It is therefore important to study how cells communicate to find potential therapeutic targets. The general principle of cell-cell communication (CCC) is a sender cell that releases or presents so-called ligands to its surroundings and a receiver cell that binds the ligand with a so-called receptor, together termed ligand-receptor (L-R) interaction⁸⁵. Ligands can be various different molecules like proteins, small peptides, steroids, amino acids, retinoids, or fatty acid derivatives; they can be either presented on the cell surface, secreted or diffused through the plasma membrane⁸⁵. The binding of ligands by receptors initiates downstream signaling within the receiver cell⁸⁵. L-R interactions can happen across different distances, i) juxtacrine interactions require direct cell-cell contact; ii) paracrine (ligand and receptor on neighboring cells) and autocrine (ligand and receptor on the same cell) signaling require

ligands to travel short distances; iii) and finally endocrine signaling requires molecules to travel long distances within the bloodstream, which can potentially control the behaviors of the whole organism, for example, this includes hormones⁸⁵ (Fig.7). The distance that a cyto- or chemokine can meaningfully propagate during paracrine signaling was estimated at around 250 μm by Francis *et al.*⁸⁶. The term cytokines refers to signaling molecules in the form of peptides, proteins, and glycoproteins, while chemokines are chemo attractive cytokines that mediate communication within the immune system⁸⁷.

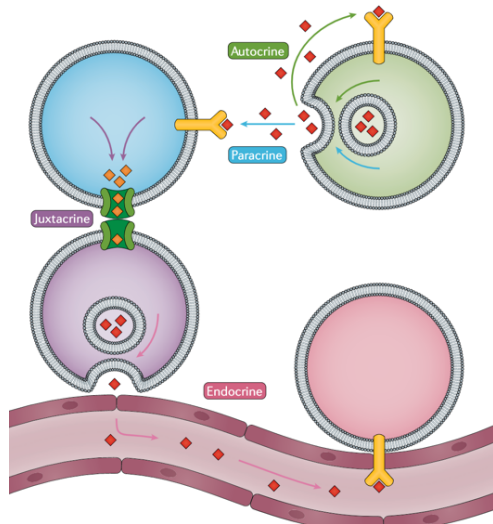


Figure 7: Overview of different types of L-R interactions defined by the distance two cells are apart to be able to communicate. Juxtacrine signaling requires direct cell contact, paracrine (ligand and receptor on neighboring cells) and autocrine (ligand and receptor on the same cell) work for short distances and endocrine signals can travel long distances within the bloodstream. Illustration from Armingol E. *et al.*, *Deciphering cell-cell interactions and communication from gene expression*. *Nat. Rev. Genet.* 22, 71-88, 2021, by permission of Springer Nature (<https://www.nature.com/nmeth/>).

5.3.2 Inferring ligand-receptor interactions from single-cell transcriptomics data

CCC studies are traditionally done with methods of co-immunoprecipitation, proximity labeling proteomics, yeast two-hybrid screening, or fluorescence resonance energy transfer imaging to identify protein-protein interactions (reviewed in Rao *et al.*, 2014 and Zhou *et al.*, 2020^{88,89}). However, the accessibility to transcriptomic studies instead of proteomics is much higher, and therefore computational tools were developed to predict L-R interactions from transcriptomics data (reviewed in Armingol *et al.*, 2021⁹⁰). The resolution of different cell types within scRNA-seq datasets is especially useful for L-R prediction studies because one knows the exact cell type where a ligand or a receptor is expressed⁹⁰. Methods to predict L-R interactions from scRNA-seq data are all based on a similar approach shown in figure 8. In short, after gene expression matrix generation by scRNA-seq and clustering, genes are filtered based on known ligands and receptors from collected resources of known interactions⁹⁰. Based on gene expression of L-R pairs across two clusters, a communication score is calculated⁹⁰.

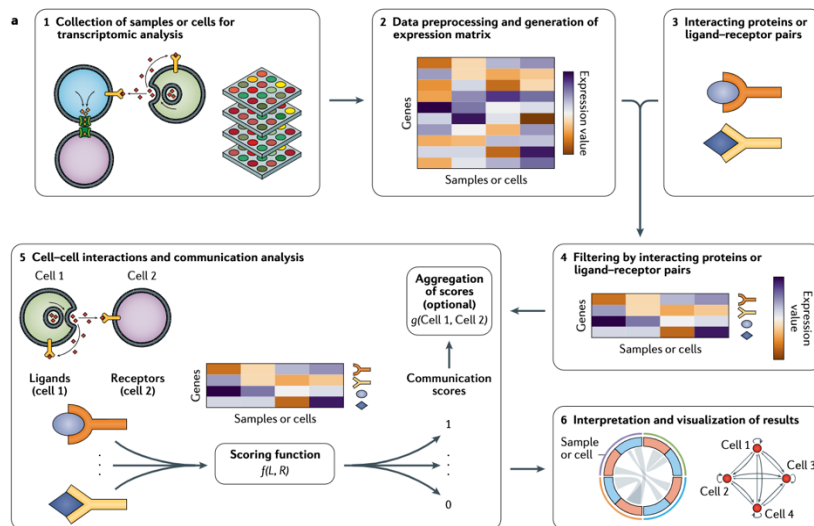


Figure 8: Overview of the approach to predict L-R interactions from single-cell transcriptomics data. 1) Cell or tissue collection for scRNA-seq analysis; 2) Generation of gene expression matrices per cell or sample; 3) Adding information of known L-R pairs; 4) Filtering of gene expression data by known L-R pairs; 5) Calculation of a communication score of L-R pairs between two cell types; 6) Interpretation and visualization of results. Illustration from Armingol E. et al., *Deciphering cell-cell interactions and communication*

from gene expression. *Nat. Rev. Genet.* 22, 71-88, 2021, by permission of Springer Nature (<https://www.nature.com/nmeth/>).

There are many different tools for L-R prediction with slight differences in the type of L-R database that is used and the calculation of communication scores. Many of these methods were compared by Dimitrov et al., 2022⁹¹. Databases from CellPhoneDB⁹² and ICELLNET⁹³ are manually curated, meaning that the interactions were manually checked, while other iTALK⁹⁴ and SingleCellSignalR⁹⁵ are using databases that also include non-curated interactions, which results in larger databases, but higher chances of false-positives. There are also many different approaches how to identify the most important interactions, reviewed in Armingol et al., 2021⁹⁰. Methods like iTALK use a differential combination approach, where differentially expressed genes (DEGs) between clusters are overlapped with known L-R pairs⁹⁴. NicheNet uses a network-based approach by taking downstream targets and transcription factors into consideration for the prediction⁹⁶. CellPhoneDB and ICELLNET are based on permutation tests to evaluate significance; they rely on random shuffling of labels to generate a null hypothesis that can then be used for testing of significance of observed L-R pairs^{92,93}. Ligands and receptors can be heteromeric complexes and only if all subunits are expressed meaningful interactions can take place, therefore, some tools like CellPhoneDB and ICELLNET account for subunit expression, ensuring that only if all subunits are expressed sufficiently an interaction is predicted^{92,93}.

5.3.3 Inferring ligand-receptor interactions from spatial transcriptomics data

Due to the rapid development of different ST approaches, a lot of effort has been dedicated to including spatial information in L-R interaction analysis, which helps to prioritize interactions and to remove false positives because spatially close cells are also more likely to interact with each other. Existing tools like CellPhoneDB (version 3) and NicheNet included the consideration of spatial information in their pipelines^{92,96-98}. Other tools were developed specifically for application to ST data. For example, a feature in the Giotto ST toolbox takes the spatial proximity of cell types into consideration for L-R prediction by using spatial proximity graphs⁹⁹. Another tool, SpaOTsc, constructs space-constrained communication networks by estimating the “diffusivity” of signaling pathways by investigating downstream gene expression in a spatial context¹⁰⁰. And a very recent approach from Fischer *et al.* uses neural networks to compute intercellular communication by taking the influence of the niche composition on gene expression into consideration¹⁰¹.

5.3.4 How to identify cell-cell communications that are biologically most relevant

All L-R prediction tools based on transcriptomics give large lists of potential L-R interactions as outputs. However, it is challenging to decipher the biologically most relevant interactions. Using ST in combination with or instead of scRNA-seq data is already a good effort to extract more relevant interactions because cells are more likely to interact if they are in spatial proximity. An interesting approach was introduced by Williams *et al.* where they hypothesized that biologically most relevant L-R interactions are evolutionarily conserved⁹⁸. Therefore, the authors performed L-R prediction studies on single-cell transcriptomics data from the liver of various different species like macaque, hamster, chicken, zebrafish, humans, and mouse⁹⁸. They could thereby find a highly conserved interaction between Kupffer cells and stellate cells⁹⁸. Other approaches are based on the validation of L-R interactions identified with transcriptomic assays using proteomic assays. For example, experimental approaches like co-occurrence measurements of ligand and receptor proteins using FACS or immunohistochemistry⁹⁰. However, these do not functionally validate L-R interactions; to do so, inhibitors against or activator cytokines for ligands or receptors can be used to block certain interactions *in vivo* or *in vitro* and expressional changes in pathways can be assessed⁹⁰.

5.3.5 Limitations of current approaches

L-R interaction analysis based on transcriptomic data (scRNA-seq and ST) are based on the following assumptions: Gene expression can be directly translated to protein abundance and protein abundance is translatable to interacting potential without taking any other factors like post-translational modifications into consideration⁹⁰. This is a major limitation because predicted L-R interactions might result in false positives that might not be biologically relevant. Some effort has been dedicated to taking on this limitation by combining transcriptomics and proteomics within one assay⁹⁰, for example, scRNA-seq in combination with a CITE-seq antibody panel against potential L-R pairs, or tools like Nativeomics which uses mass spectrometry to identify ligand binding to membrane proteins¹⁰² and INs-seq that integrates scRNA-seq readouts with intracellular protein measurements using cell permeabilization, staining, and FACS¹⁰³. Another consideration is that not all interactions that are within L-R databases are functionally validated properly⁹⁰ and many contributing interactions are likely not annotated and included in databases yet. The inclusion of spatial data reduces the likelihood of false positive L-R predictions⁹⁰. However, with the lack of single-cell resolution it is often difficult to study direct L-R interactions between specific cell types.

5.4 Mouse liver biology

The liver is a highly spatially organized organ with well annotated stereotypic gene expression patterns and functions arranged in defined spatial locations¹⁰⁴. This makes the liver an ideal organ to benchmark our novel ST technology.

5.4.1 Anatomy of the liver

Within the human body, the liver makes up 2-3 % of the body weight, making it the largest internal organ¹⁰⁵. As a combination of two lobes, a larger right lobe and a smaller left lobe, the liver is located in the upper abdominal cavity where the rib cage protects it and the position is maintained by ligamentous attachments that expand into the so-called Glisson capsule or

hepatic capsule that protects the liver's surface¹⁰⁵. The blood entering into the liver is a composition of approximately 75 % of venous blood, a collection from the spleen, stomach, intestine, gallbladder, and pancreas, and 25 % of highly oxygenated blood from the hepatic artery¹⁰⁶. On the microanatomical scale, the liver is built up of homogenous subunits called liver lobules that are hexagonal structures of bi-layered chords of hepatocytes, separated by sinusoidal vessels¹⁰⁴. Within the sinusoids, blood flows from the portal vein (PV) and the hepatic artery radially inwards to the draining central vein (CV)^{104,107}. The PV, hepatic artery, and the bile duct together form the portal triad¹⁰⁷. The blood flow creates a gradient of oxygen, due to influx from the hepatic artery and gradual consumption as blood flows towards the CV¹⁰⁴; the influence of this will be discussed later in the section on metabolic zonation. Endothelial cells are lining the sinusoids and exchange molecules from the blood with hepatocytes in a microenvironment called the space of Disse, which lies in between these two cell types¹⁰⁸. Fluids in the bile duct are going in the opposite direction of the blood flow to transport bile acid to the intestine¹⁰⁴. Figure 9 illustrates the microanatomical structure of the liver.

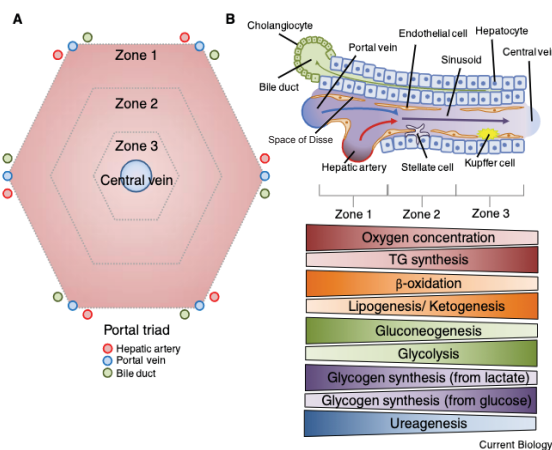


Figure 9: Microanatomical structure of the liver sinusoids and metabolic zonation. A) Hexagonal structure of the liver lobules. They are organized in different zones lining the sinusoids forming a central to portal axis. B) Structure of a liver sinusoid showing the blood flow from the portal vein and hepatic artery draining into the central vein. Different cell types, which support hepatic function, are indicated in the figure. There is a gradient of oxygen, nutrients and hormones within the blood which results in different tasks within different lobule zones shown in different color gradients. This division of labor is a concept called “metabolic zonation”. Illustration from Trefts E. et al., *The liver*. *Curr. Biol.*, 27, R2247-R2251, 2017, by permission from Elsevier.

5.4.2 Functions of the liver

The liver has several critical functions to preserve physiological homeostasis (reviewed in Ben-Moshe and Itzkovitz *et al.*, 2019)¹⁰⁴. The blood drains from the intestine and is therefore full of metabolites and nutrients¹⁰⁴. One of the liver's main tasks is to store nutrients and to manage their release in a controlled manner. For example, the liver helps to keep constant glucose levels throughout the day¹⁰⁴. Other liver tasks are detoxification, defense against pathogens, production of hormones, and protein synthesis¹⁰⁴. Hepatocytes produce bile acids, which are collected in bile ducts and shuttled over the portal triad to the intestine¹⁰⁴. Bile acids are important for solubilizing lipids to promote lipid absorption¹⁰⁹.

5.4.3 Immune tolerogenic feature of the liver

An interesting feature of the liver is its immune tolerance (reviewed in Crispe *et al.*, 2003)¹¹⁰. It has been hypothesized that due to a constant influx of blood from the intestine, which is full of harmless bacterial products and food-derived antigens, the liver's immune system has to have a certain tolerance against them that represses it from getting activated against harmless influx¹¹⁰. However, the hepatic immune system also has to be ready if pathogenic products are entering the liver, therefore different immune cell types have to exert different tasks to regulate immune tolerance and immune activation.

5.4.4 Cell types within the liver and their functions

The liver is composed of 60% parenchymal cells, which are the hepatocytes, and 40% non-parenchymal cells (NPCs), which are mainly liver endothelial cells (LECs), hepatic stellate cells (HSCs), cholangiocytes, Kupffer cells (KCs) and additional immune cells¹¹¹. The most important cell types for this thesis within the murine liver will be described in more detail within this section.

5.4.4.1 Parenchymal cells

Parenchymal is a term that describes cells that are the most abundant and functionally important within a tissue volume¹¹². Within the liver, these are the hepatocytes making up around 80 % of the liver mass¹¹¹. They have diverse functions depending on where the hepatocytes are within the central to portal axis¹⁰⁴. The exact functions will be discussed in the section on metabolic zonation.

5.4.4.2 Non-parenchymal cells

Non-parenchymal describes cells that help parenchymal cells in a supportive way to conduct their functions.

5.4.4.2.1 Liver endothelial cells (LECs)

Endothelial cells in general are lining blood vessel walls where they are involved in the exchange and filtering of for example nutrients¹¹³. Within the liver they can be broadly subdivided into large vessel endothelial cells (LVECs) and liver sinusoidal endothelial cells (LSECs)^{114,115}. LVECs are lining the large blood vessels¹¹⁴. Liver sinusoidal endothelial cells (LSECs) are lining the liver sinusoids, where they function as sieve plates to exchange molecules between the blood and the hepatic cell types¹¹⁶. They have a higher permeability for molecules than other endothelial cells¹¹³. Their open pores, so-called fenestrae, which are around 100-200 nm in diameter lack a diaphragm and basement membrane (reviewed in Braet *et al.*, 2002)¹¹⁷. This makes an bidirectional exchange of particles, fluids and solutes easier because it does not require complex processes of endocytosis and transcytosis, however, only particles smaller than the fenestrae can pass¹¹⁷. Interestingly, the number and diameter of fenestrae can change dynamically due to pathological or physiological changes¹¹⁷. Additionally, LSECs are involved in the tethering of leukocytes so they can enter the hepatic tissue from the blood stream (reviewed in Shetty *et al.*, 2018)¹¹⁸. The processes of initial adhesion, subsequent crawling on the vessel wall, and transmigration are mediated by adhesion molecules, for example, intercellular adhesion molecule 1 (ICAM1) and vascular cell adhesion molecule 1 (VCAM1), that bind leukocytes via integrins¹¹⁸.

5.4.4.2.2 T lymphocytes

T lymphocytes within the liver can be broadly subdivided into CD4+ and CD8+ T cells which increase during inflammation due the recruitment of circulating T cells via for example LSECs or KCs¹¹⁰. T cells can be primed by antigen-presenting cells (APCs) which can be for example LSECs¹¹⁹, KCs^{110,120} or DCs¹¹⁰. Antigen presentation is a process where a cell presents an antigen on its cell surface, which can for example be derived from pathogens¹²¹. For T cells this presentation occurs via a so-called major histocompatibility complex (MHC)¹²¹. The

presented antigen can then be recognized by T cells, which in response can initiate anti-pathogenic effector functions¹²¹.

5.4.4.2.3 Macrophages

KCs are the tissue resident macrophages of the liver and they are involved in the phagocytosis of microorganisms and apoptotic cells^{110,122,123}. During an inflammatory response, KCs can function as APCs and interact with T cells which may lead to their proliferation and cytokine production^{110,120}. Interaction of KCs with lymphocytes can also lead to reciprocal activation of KC¹²⁴. There are also indications that KCs are involved in immune tolerance, e.g. by the production of nitric oxide that suppresses activation of T cells¹²⁵. During pathological conditions, circulating monocytes can be recruited from the blood which can differentiate into macrophages with different immunogenic functions in response to their microenvironment (reviewed in Wen *et al.*, 2021)¹²⁶. Previously, macrophages were subdivided into two categories based on their Ly6C expression (Ly6C^{high} and Ly6C^{low}), however, recent advances in scRNA-seq studies discovered many more subtypes during pathological conditions^{126,127}. Some of them will be discussed in a later section on macrophages in liver metastasis.

5.4.4.2.4 NK and NKT cells

Within the liver, there is an exceptionally high proportion of natural killer (NK) and natural killer T (NKT) cells compared to other lymphoid tissues^{110,128,129}. NK cells are involved in the adaptive immune response by generating cytokines and interacting with other immune cells¹³⁰. NKT cells in general serve as a bridge between the adaptive and the innate immune system¹³¹ and they were found to secrete IL-4 which gives them an anti-inflammatory phenotype^{110,132}.

5.4.5 Metabolic zonation

Spatial heterogeneity between hepatocytes across the central-portal axis was first described in 1933 by Kater *et al.*, where they described gradients of glycogen and fat and differences in mitochondria morphology¹³³. Many other studies followed, grouping different metabolic functions into pericentral (combined lobule layers towards the central veins) and periportal (combined lobule layers towards the portal vein) tasks, a concept named “metabolic zonation”^{104,133–139} (Fig. 9).

- Pericentral tasks: Glycolysis, lipogenesis, alcohol detoxification
- Periportal tasks: Gluconeogenesis, oxidative metabolism, ureagenesis, β -oxidation of fatty acids, cholesterol biosynthesis, amino acid breakdown

In general, periportal tasks are more energy-consuming because the oxygen levels are higher; therefore, hepatocytes can make more adenosine triphosphate (ATP)¹⁰⁴. Another factor that impacts metabolic zonation is the production of WNT (Wingless-related integration site) which is produced by centrally located LECs and its concentration, therefore, gradually decreases towards the portal vein^{65,104,140}. WNT/ β -catenin signaling has been recognized as a master regulator of liver metabolic zonation¹¹⁶. One-third of zoned genes within hepatocytes are regulated by WNT signaling – with WNT-activated and -repressed genes being expressed pericentrally and periportal, respectively¹⁴¹.

Metabolic zonation only imprints within the first week after birth in mice, hepatocytes in perinatal mice show no zonation^{116,142,143}. The first studies on metabolic zonation were done

using methods like histochemistry, immunohistochemistry (IHC), ISH, or a method called digitonin-collagenase perfusion¹⁰⁴. Digitonin is a detergent that lyses eukaryotic cells and perfusion of the liver from either the central or portal vein leads to damaged cells within the central or portal vein respectively^{144–146}. The cells within the undamaged side could then be dissociated using collagenase, an enzyme that digests collagen which keeps the tissue matrix intact, and undamaged cells could be analyzed^{144–146}. With the development of single-cell transcriptomics, the hepatic division of labor could be studied in high throughput assays. Halpern *et al.* used scRNA-seq of hepatocytes and 6 landmark genes, identified previously and verified by smFISH, to reconstruct the spatial position of hepatocytes within 9 lobule layers from central to portal vein¹⁴¹. Reconstructed hepatocytes could then be investigated further and approximately half of the genes were found to be zoned to conduct different tasks¹⁴¹. Information about zoned gene expression in this study was then used further by Halpern *et al.* in 2018 to reconstruct the spatial position of LECs using pcRNAseq⁶⁵. Pairs of LECs and hepatocytes were sequenced together and the position of pairs was then reconstructed using hepatocyte landmark genes, different gene expression patterns in LECs between different lobule layers could then be analyzed further⁶⁵. They found that also in LECs approximately 50 % of the genes are zoned with pericentral genes being mainly associated to WNT signaling⁶⁵. This underlined the important role of LECs in supporting the zonation of hepatocytes.

5.4.6 Immune zonation

Recently, Gola *et al.* employed quantitative multiplex imaging to characterize the immune zonation of the liver, and described the enrichment of NKT cells and KCs in the periportal area¹⁴⁷. This was also seen in the recent effort of Williams *et al.* to characterize all hepatic cells using single-cell and single-nuclei sequencing (snRNA-seq)⁹⁸. Immune zonation in mice manifests only around day 20-25 after birth when the mice are weaned compared to metabolic zonation that is matured around one week after birth^{142,143,147}. Unlike metabolic zonation, WNT signaling does not have any influence on immune zonation, rather the change in gut microbiota during food uptake after weaning induces the periportal enrichment of KCs and NKT cells¹⁴⁷. The periportal enrichment is suggested to be important for pro-inflammatory responses towards incoming bacteria and pathogens from the intestinal blood flow, this might protect important functions of the central veins¹⁴⁷. Even though there is a zonation of KCs as a whole, no difference in gene expression across KCs from different zones within the central-portal axis has been described to date.

5.5 Colorectal liver metastasis

5.5.1 Epidemiology and current treatment options

Liver metastasis formation is especially common in colorectal cancer (CRC), presumably because venous blood drains from the gut and the liver is a highly vascular organ¹⁴⁸. A study from 2006 revealed that of 4,399 CRC patients 41 % developed metastases in the liver¹⁴⁹. Worldwide, CRC constitutes around 10% of all diagnosed cancers yearly and is the second most common type of cancer in women and third in men¹⁵⁰. In patients where it is possible to resect metastatic tissues, patients undergo surgical resection, combined with pre-operative neoadjuvant chemotherapy (NAC)^{151,152}. Patients with non-resectable metastases are treated

with chemotherapy and biological agents targeting EGFR (epidermal growth factor receptor) or VEGF (vascular endothelial growth factor) to reduce tumor growth^{151,152}. However, researchers are constantly trying to find targets for more defined treatment options for different metastatic phases.

5.5.2 Phases of liver metastasis development

Metastasis formation is a multi-phase process (reviewed in Brodt *et al.*, 2016)¹⁵³. The first phase is called the microvascular phase¹⁵³. During that phase tumor cells that derive from primary sites like the colon and circulate within the vasculature get trapped¹⁵³. It is followed by the extravascular pre-angiogenic phase where tumor cells travel into the space of Disse and activate a stromal response¹⁵³. The third phase is the angiogenic phase where de novo vasculature is formed within micrometastases¹⁵³. This is followed by the fourth and last phase, the growth phase which is defined by the expansion of metastases¹⁵³.

5.5.3 Influence of hepatic cells in metastasis formation, promotion, and reduction

Hepatic cells are influencing CRC metastatic development in different ways. Within the tumor microenvironment (TME) different cell types are getting recruited or existing cell types are changing their behavior in order to have pro- or anti-metastatic effects (reviewed in Niu *et al.*, 2022 and Oura *et al.*, 2021)^{154,155}. Cell types most important for this thesis will be explained in more detail.

When a tumor cell enters the liver through the bloodstream, LSECs, KCs, and NK cells are likely among its first encounters¹⁵⁶. These cells have the potential to fight the incoming threat, for example by KC-dependent phagocytosis, NK-dependent cytolysis, and activation of other immune cells through the release of pro-inflammatory factors like TNF- α ^{156,157}.

5.5.3.1 Liver sinusoidal endothelial cells

Liver sinusoidal endothelial cells are involved in recruitment of immune cells from the blood into the TME¹⁵⁵. Later in the angiogenic phase, LSECs are also involved in de-novo vascularization¹⁵³.

5.5.3.2 Macrophages

KCs in general have an anti-metastatic effect through phagocytosis of tumor cells and the release of TNF- α , which has a pro-inflammatory effect^{153,158,159}. Additionally, KCs can recruit other immune cells like NK cells and neutrophils to fight against the metastatic cells¹⁵³. Circulating monocytes that originated in the bone marrow are recruited to the TME, where they differentiate into tumor associated macrophages (TAMs) that can have different functions to promote or fight against the metastatic process^{126,154}. One example is the CCL2|CCR2 recruitment axis between tumor cells and monocytes, where they then acquire a TAM phenotypes that favors tumor growth¹⁵⁴. Recently published scRNA-seq studies from patient samples of CRC and liver metastasis found an enrichment of SPP1+ and MRC1+CCL18+ expressing macrophage which correlated with an immunosuppressive microenvironment^{160,161}. Another example is C1Q+ macrophages which have been shown to be immunosuppressive and were found to be correlated with poor patient prognosis in hepatic metastasis^{127,162}.

5.5.3.3 T lymphocytes

Recruited macrophages can influence the T cell composition within the TME¹⁵³. For example the C1Q+ macrophage subset is correlated with a higher abundance of exhausted T cells¹²⁷. T cell exhaustion is a phenomenon where T cells lose their function to produce effector cytokines, they lose their cytotoxicity and they overexpress inhibitory receptors (reviewed in Jiang *et al.*, 2015)¹⁶³. All these features lead to a pro-tumorigenic TME¹⁶³.

5.5.4 Mouse model of colorectal liver metastasis

Mouse models of liver metastasis can be subdivided into two categories: spontaneous and experimental (reviewed in Oh *et al.*, 2017)¹⁶⁴. The spontaneous approach includes a primary tumor generated by injection of cancer cells into the colon; and after the primary tumor develops it spontaneously metastasizes to the liver¹⁶⁴. This approach represents all phases of the metastatic process, however, metastasis formation is often not predictable and reproducible and it takes a long time until metastases development^{164–166}. Sometimes the primary tumor even has to be resected, before proper metastasis formation, because the tumor burden is too high^{164–166}. During the experimental approach, CRC cells are injected directly into the spleen or the portal vein, the cells then enter the microvasculature and some of them engraft and form liver metastasis¹⁶⁴. The advantage of this model is that the metastasis formation is fast and it is more reproducible than the spontaneous model, however, only the late metastatic phase can be studied and the systemic changes of the immune system due to a primary tumor are bypassed¹⁶⁴. We used an experimental model, inducing liver metastasis via intrasplenic injection of AKPS CRC organoids. These organoids harbor four mutations in the following genes: *Apc*, *Kras*, *Tp53* and *Smad4*, which are the four driver mutations involved in colorectal cancer development¹⁶⁷.

6 Aims of this thesis

The field of ST technology development is quite new and still harbors many limitations. For example, complex technology requirements like high-resolution microscopes, targeting only a few ROIs, the requirement of prior knowledge for gene panel design in commercially available imaging-based methods, and the lack of single-cell resolution in commercially available, easy-to-use, unsupervised technologies¹⁶⁸. Therefore, we wanted to develop a new ST methodology called sphere-sequencing (sphere-seq) that includes the following improvements:

- Single-cell resolution
- Unsupervised transcriptome-capture
- Three-dimensional spatial capture
- High-throughput
- No requirements for tissue slicing
- Simplicity of data analysis

These are the aims to archive that:

1. Establishment of an experimental and computational workflow for sphere-seq.
2. Validation of sphere-seq on metastasis-bearing livers.
3. Evaluation of sphere-seq performance in comparison to one of the most applied mRNA-capture array-based ST method (Visium).
4. Application of sphere-seq on other tissues and species.

7 Material and methods

7.1 Experimental protocol of sphere-sequencing

This section describes the experimental protocol of sphere-seq that was developed for this project. It includes a detailed section for all materials used together with their distributors. The procedure is explained in a continuous manner step by step in Arabic numbering format. Options for some steps are highlighted with capital letters in alphabetical order for different tissues and Latin numbers for different scRNA-seq methodologies. There are also sections highlighting the anticipated results and options for troubleshooting.

7.1.1 Materials

7.1.1.1 Equipment

| Equipment | Distributor/Catalogue number | Tissue |
|--|---------------------------------|------------|
| Big centrifuge 5910R | Eppendorf | All |
| Centrifuge insert for 50 mL tubes | Eppendorf | All |
| Centrifuge insert for 5 mL tubes | Eppendorf | All |
| Centrifuge insert for plates | Eppendorf | All |
| Small centrifuge 5424R | Eppendorf | All |
| Thermal cycler (ProFlex PCR System) | Applied Biosystems | All |
| Thermomixer | Eppendorf | All |
| Thermomixer adapter for 96 well plates | Eppendorf | All |
| Thermomixer adapter for 1.5 mL tubes | Eppendorf | All |
| Large fragment biosorter | Union Biometrica | All |
| 1000 µm nozzle | Union Biometrica | All |
| P20 multichannel pipette | Rainin | All |
| P200 multichannel pipette | Rainin | All |
| P20 pipette | Rainin | All |
| P200 pipette | Rainin | All |
| P1000 pipette | Rainin | All |
| Hemocytometer | Sigma Aldrich | All |
| 10x Magnetic Separator (for small PCR tubes) | 10X Genomics | All |
| Magnetic separation Rack for 1.5 mL tubes | New England Biolabs #S15065 | All |
| Mini centrifuge | Sigma Aldrich | All |
| Qubit 4 fluorometer | Thermo Fisher Scientific | All |
| Tape Station 4200 system | Agilent Technologies | All |
| 10X Vortex Adapter | Chromium 10X #330002 | All if 10X |
| Chromium 10X single-cell capture system | Chromium 10X | All if 10X |
| BD Rhapsody™ Express Single-Cell Analysis system | BD Rhapsody | All if BD |
| BD Rhapsody™ P1200M pipette | BD Rhapsody #633704 | All if BD |
| BD Rhapsody™ P5000M pipette | BD Rhapsody #633705 | All if BD |
| Large magnetic separation stand | V&P Scientific, Inc. #VP772FB-1 | All if BD |
| Heat block for 1.5 mL tubes | Fisher Scientific | All if BD |
| Perfusion pump | Thermo Fisher Scientific | Liver |
| Rocking Shaker | Thermo Fisher Scientific | Organoids |

Table 6: Equipment for sphere-seq.

7.1.1.2 Consumables

| Consumables | Distributor/Catalogue number | Tissue |
|--|--|---------------------|
| 15 and 50 mL falcon tubes | Falcon | All |
| Strainer in different sizes (40, 70, 100, 300, 200, 400, 500 µm) | PluriSelect #43-50040-51, #43-50070-51, #43-50100-51, #43-50200-03, #43-50300-03, #43-50400-03, #43-50500-03 | All |
| Non-binding, V-shaped 96-well plates | Greiner Bio-one #651901 | All |
| Adhesive sealing sheets for plates | Thermo Fisher Scientific #AB0558 | All |
| Low-retention P20 pipette tips | Rainin | All |
| Low-retention P20 pipette tips | Rainin | All |
| P20 pipette tips | Rainin | All |
| P200 pipette tips | Rainin | All |
| Low-retention P1000 pipette tips | Rainin | All |
| P1000 pipette tips | Rainin | All |
| 5 mL FACS tubes | Falcon | All |
| Low-binding 1.5 mL Eppendorf tubes | Eppendorf | All |
| BD Rhapsody™ Cartridge Kit | BD Biosciences #633801 | All |
| Quit tubes | Invitrogen #Q32856 | All |
| High sensitivity D1000 Tape Station tapes | Agilent Technologies #5067-5584 | All |
| High sensitivity D5000 Tape Station tapes | Agilent Technologies #5067-5592 | All |
| Tape Station loading tips | Agilent Technologies #5067-5598 | All |
| P1000 pipette tips | Eppendorf | All if BD |
| 5 mL pipette tips | Eppendorf | All if BD |
| Chromium Next GEM Chip G Single Cell Kit (1 reaction) | 10X Genomics #PN-10000120 | All if Chromium 10X |

Table 7: Consumables for sphere-seq.

7.1.1.3 Reagents

| Reagents | Distributor/Catalogue number | Tissue |
|--|---|--------|
| PBS, pH 7.4 | Gibco | All |
| 60 µm Megabead NIST traceable particles | Polysciences | All |
| 125 µm Megabead NIST traceable particles | Polysciences | All |
| 175 µm Megabead NIST traceable particles | Polysciences | All |
| 10 X PBS | Millipore | All |
| Cleaning Solution for Biosorter | Union Biometrica #300-5072-000 | All |
| 100 % Ethanol | Sigma Aldrich | All |
| BSA | Sigma Aldrich | All |
| MULTI-seq Lipid-Modified Oligos | Sigma Aldrich #LMO001-100R; or lab of Zev Gartner as UCSF | All |
| Trypan Blue Solution, 0.4 % | Gibco | All |
| MULTI-seq primer | IDT, standard desalting, 5'-CTTGGCACCCGAGAATTCC-3' | All |

| | | |
|--|---|-------------------|
| SPRIselect beads | Beckman coulter B23319 | All |
| Elution buffer (EB) | Qiagen | All |
| Nuclease free water | Invitrogen | All |
| HiFi HotStart ReadyMix (2x) | Fisher Scientific #NC0295239 | All |
| Small RNA TruSeq RPI primer different indices | Illumina #RS-200-0012 or IDT TrueGrade purification | All |
| 100% Isopropanol | Sigma Aldrich | All |
| High sensitivity D1000 Tape Station reagents | Agilent Technologies #5067-5585 | All |
| High sensitivity D5000 Tape Station reagents | Agilent Technologies #5067-5593 | All |
| dsDNA high-sensitivity Qubit reagent kit | Life Technologies #Q32854 | All |
| 288 MULTI-seq barcode oligos for 3' capture | IDT, standard desalting, 5'- CCTTGGCACCCGAGAATTCCANNNNNNNNA ₃₀ -3' | All if 3' capture |
| 288 MULTI-seq barcode oligos for 5' capture | IDT, standard desalting, 5'- CCTTGGCACCCGAGAATTCCANNNNNNNNCCC ATATAAGAAA -3' | All if 5' capture |
| BD Rhapsody™ Cartridge Reagent Kit (1 reaction) | BD Biosciences #633731 | All if BD |
| BD Rhapsody™ cDNA Kit (1 reaction) | BD Biosciences #633773 | All if BD |
| BD Rhapsody™ Whole transcriptome Analysis (WTA) Amplification Kit (1 reaction) | BD Biosciences #633801 | All if BD |
| Chromium Next GEM Single Cell 5' library and Gel bead Kit v 1.1 (1 reaction) | 10X Genomics #PN-1000165 | All if 10X |
| Chromium Single Cell 5' Library Construction Kit | 10X Genomics #PN-1000020 | All if 10X |
| DynaBeads MyOne Silan Beads | Thermo Fisher #37002D | All if 10X |
| Low glucose DMEM (1 g/L D-Glucose/L-Glutamine, Pyruvate) | Gibco #31885-023 | Liver |
| 1 M HEPES | Gibco | Liver |
| Liberase | Sigma Aldrich #05401119001 | Liver |
| 10 X TripLE | Gibco | Liver |
| MACS tissue storage solution | Miltenyi Biotec #130-100-008 | Crohn |
| HBSS (-Ca ²⁺ -Mg ²⁺) | Sigma Aldrich | Crohn |
| 0.5 M EDTA | Lonza | Crohn |
| HBSS (+Ca ²⁺ +Mg ²⁺) | Sigma Aldrich | Crohn |
| DNAseI | Roche Diagnostics #10104159001 | Crohn |
| Collagenase from Clostridium histolyticum | Sigma Aldrich #C5138 | Crohn |
| 1 X TripLE | Gibco | Organoids |

Table 8: Reagents needed for sphere-seq.

7.1.2 Procedure

General note: use low-retention pipette tips and low-binding tubes whenever possible to reduce cell loss.

1. Prepare the following solutions

| Solution | Components | Tissue |
|--|--|--------|
| 1 X PBS for sorter (5-10 L) | 10 X PBS diluted with ddH ₂ O | All |
| Standard beads solution (30 mL) | 2-3 drops of each bead size (60, 125, 175 µm) in 30 mL ddH ₂ O | All |
| 70 % EtOH for sorter (0.5 L) | Dilute 100 % EtOH with ddH ₂ O | All |
| EDTA/PBS | 20 mM EDTA in PBS | All |
| Anchor:BC for 100-400 µm spheres (288x) | 50 nM, Anchor:BC 1:1 in 20 µl PBS; | All |
| Anchor:BC for 400-800 µm spheres (288x) | 100 nM, Anchor:BC 1:1 in 20 µl PBS; | All |
| Co-anchor for 100-400 µm spheres (288x) | 50 nM in 20 µl PBS | All |
| Co-anchor 100 nM (288x) | 100 nM in 20 µl PBS | All |
| 10 % BSA/PBS (40 mL) | Dissolve 3 g BSA in PBS; filter through 40 µm strainer before use | All |
| 1 % BSA/PBS (20 mL) | Dilute 2 mL 10 % BSA with 18 mL PBS | All |
| MULTI-seq primer | 2.5 µM in TE | All |
| 80 % EtOH for library preparation (10 mL) | Dilute 100 % EtOH with nuclease free ddH ₂ O | All |
| Murine Liver Dissociation Buffer (9 mL) | low glucose DMEM (1 g/L D-Glucose/L-Glutamine, Pyruvate) with 15 mM HEPES, 32 µg/mL Liberase and 1x TripLE | Liver |
| Crohn Epithelial Dissociation Buffer (30 mL) | HBSS (-Ca ²⁺ -Mg ²⁺) with 10 mM HEPES and 5 mM EDTA | Crohn |
| Crohn Digestion Buffer (30 mL) | HBSS (+Ca ²⁺ +Mg ²⁺) with 0.5 mg/mL DNaseI and 0.5 mg/mL Collagenase | Crohn |
| 10 % BSA/low Glucose DMEM (40 mL) | Dissolve 3 g BSA in low Glucose DMEM; filter through 40 µm strainer before use | All |
| 1 % BSA/low Glucose DMEM (20 mL) | Dilute 2 mL 10 % BSA with 18 mL low Glucose DMEM | All |

Table 9: Solutions that have to be prepared for sphere-seq.

2. Set-up of the large fragment biosorter (follows the BioSorter manual from Union Biometrica)
 - 2.1. Use 1000 µm FOCA (Fluidics and Optics Core Assembly)
 - 2.2. Fill 10 mL ddH₂O into the sample cup (50 mL falcon tube)
 - 2.3. Check the tanks: the waste tank should be empty, the water tank filled with ddH₂O, the sheath flow tank filled with 1x PBS, the ethanol tank filled with 70 % EtOH, and the cleaning solution tank filled with a cleaning solution;
 - 2.4. Turn on the biosorter and the computer
 - 2.5. Open FlowPilot software
 - 2.6. Turn on the lasers using the iBEAM-SMART software: Switch on the emission of channel 1 to 60 mW

- 2.7. Open an existing or new experiment in the drop-down menu “setup”
- 2.8. Set the settings of the biosorter to the following parameters, these can be adjusted depending on the sort requirements, the machine, and the software:
 - Sample cup pressure: 0.42 pounds per square inch (psi)
 - Diverter pressure: 2.5 psi
 - Power: 50 megawatts (mW)
 - Gain: 1.0
 - Photomultiplier tube (PMT) Volts green: 600
 - PMT Volts red: 750
 - Drop width: 7 milliseconds (mS)
 - Sort delay: 23 mS
 - Sheath flow rate: 57 %
 - Select Pinch valve
- 2.9. Use the flushing procedure to fill the pipe that connects the sample cup with the FOCA with fluids:
 - 2.9.1. check all boxes “Sheath on”, “Sample on”, “Diverter pressure” and “Waste tray open”
 - 2.9.2. Depressurize the sample cup by clicking “Refill sample”
 - 2.9.3. When the pipe is filled with fluids depressurize by clicking “Done Refill”
- 2.10. General procedure: when changing the sample cup depressurize beforehand by clicking “Refill sample”, after the sample cup is changed click “Done Refill”
- 2.11. Washing and priming of the system: “Maintenance” drop-down menu, select in the following order:
 - 2.11.1. Water Wash
 - 2.11.2. Prime Flow Cell
 - 2.11.3. Prime Sample Cup
- 2.12. Check if all pipes around the FOCA are filled with fluids, if there are air bubbles left repeat 2.11 until the air bubbles are gone
- 2.13. Check flow rate and stream:
 - 2.13.1. Flow stream
 - 2.13.1.1. Check the boxes “Sheath on” and “Waste tray open”
 - 2.13.1.2. Uncheck “Sample on” and “Diverter pressure”
 - 2.13.1.3. The stream should be a narrow straight stream
 - 2.13.1.4. If not, then adjust the micrometer wheel on the side of the waste tray
 - 2.13.2. Sheath flow rate:
 - 2.13.2.1. can be measured by collecting the flow through for one minute
 - 2.13.2.2. It should be around 45 mL for the 1000 μ m FOCA
 - 2.13.2.3. If not accurate enough, sheath rate can be adjusted
- 2.14. Open an existing or new sample
 - 2.14.1. If new then draw plots from the “Layout” drop-down menu, you need a histogram with TOF on the x-axis and extinction on the y-axis for size selection, and for fluorescence a histogram with TOF on the x-axis and Green/Red/.. on the y-axis.
- 2.15. For sorting in a 96-well plate calibrate plate positions:
 - 2.15.1. Put the lid of a 96-well plate on the X-Y-Z Stage
 - 2.15.2. Open the menu “Calibrate plate positions”
 - 2.15.3. Set the position for plate A well A1 and the last well; do the same for plate B

- 2.15.4. Set positions by repeatedly doing a drop test, checking the position of the drop, adjusting X and Y, and testing again until the drop is central within the well
- 2.15.5. The z-position can be adjusted too in case the sorting should happen on an ice block for example
3. Creation of a standard curve using standard-sized beads (60, 125, and 175 μm) to adapt the sorting gate
 - 3.1. Put the falcon tube with standard beads solution in the sample cup position
 - 3.2. In the "Acquisition" menu select the storage gate "all events" to store and acquire approximately 300-500 events per bead size
 - 3.3. Take acquired data to generate a standard curve with bead size on the x-axis and TOF and/or extinction on the y-axis
 - 3.4. Calculate TOF and extinction based on the linear model for sphere sizes of interest and draw the appropriate gates in the histogram
 - 3.5. The linear model can also be used in reverse to calculate an estimate of sphere sizes after sorting
4. Tissue collection
 - A. Murine liver:
 - a. Euthanize the mouse
 - b. Perfuse the liver with PBS with a perfusion pump (flow rate 2-3 mL/min) over an insertion in the vena cava inferior
 - c. After perfusion remove the gallbladder and collect the liver in PBS on ice
 - B. Murine spleen: Euthanize the mouse and collect the spleen in PBS on ice
 - C. Crohn's biopsy: Surgically resected tissues are collected in MACS tissue storage solution on ice; process further within 4-6 hours for optimal cell quality
 - D. CRC organoids:
 - a. Dissolve Matrigel of domes with ice cold PBS and collecting them in a 50 mL falcon tube, put the tube in an ice bucket and put it on a shaker for 30 min
 - b. Wash two times with cold PBS by spinning at 290 x g at 4 °C
5. Partial dissociation
 - 5.1. Use the following partial dissociation solutions:
 - A. Murine liver: low glucose DMEM (1 g/L D-Glucose/L-Glutamine, Pyruvate)
 - B. Murine spleen: PBS
 - C. Crohn's disease patient biopsy: HBSS (-Ca²⁺-Mg²⁺)
 - 5.2. Use mechanical force of a scissor to partially dissociate collected tissues in a 2 mL Eppendorf tube with partial dissociation solution
 - 5.3. Repeatedly filter partially dissociated spheres with an upper-size strainer (for example 400 μm), wash down chunks that get caught by the filter, re-cut, and filter again. Repeat this until most tissue is cut into small spheres
 - 5.4. To remove single cells and small spheres, filter solution with a lower size limit strainer (for example 200 μm)
 - 5.5. After filtering, flip the strainer upside down and wash off spheres that got caught by the strainer. This is the sphere-seq solution with in this example spheres between approximately 200-400 μm in diameter
 - 5.6. Dilute the solution to approximately 45 mL using partial dissociation solution and dilute further to achieve a dilution of around 5-10 events per second (ev/s) with the sorter
6. Prepare sorting plates
 - 6.1. Take non-binding, V-shape 96-well plates
 - 6.2. Add appropriate dissociation buffer to each well

- A. Murine liver: 30 μ l Murine Liver Dissociation Buffer (see table 9)
 - B. Murine spleen: 30 μ l PBS
 - C. Crohn's biopsy: 100 μ l Crohn Epithelial Dissociation Buffer (see table 9)
 - D. CRC organoids: 30 μ l 1 X TripLE
- 6.3. Keep on ice until use
7. Sphere sorting using a large fragment biosorter
- 7.1. Draw an appropriate sorting gate:
- Option 1: Sorting only based on size:
 - a. Draw a gate in the histogram with TOF on the x-axis and extinction on the y-axis based on values defined by the standard size linear model
 - Option 2: Sorting based on size and fluorescence:
 - a. For green turn on the 488 nm blue laser to a power of 50 (only green has been optimized).
 - b. Draw a size selected gate like in option 1
 - c. Use this gate to make a histogram with TOF on the x-axis and Green/Red/... on the y-axis
 - d. Acquire a negative control with spheres that have similar sizes to the sample that will be sorted
 - e. Draw a gate that is outside of the negative control for all different sizes, the same size distribution is important because larger spheres also have a larger autofluorescence signal. In the histogram therefore with an increase in TOF also the green/red/... signal increases in the negative control
- 7.2. Add the sample cup with the spheres and acquire events
- 7.3. Dilute if flow rate is too high, should be in the range of 5-10 ev/s otherwise sorting might not be pure
- 7.4. Turn on the mixer within the sample cup at around 50 % speed so the spheres do not settle (reduce speed if spheres are fragile)
- 7.5. Adjust the coincidence mode to "NoDoubles" which helps the sorting to be pure (one sphere per well)
- 7.6. Select the calibrated plate in the drop-down menu "Setup/Plate" and add 1 event to the wells that should be sorted
- 7.7. In the sorting window select the sorting gate and tick the box "Store only sorted"
- 7.8. Check the sorting efficiency by sorting a couple of objects on the lid of a 96-well plate, after selecting the wells and adding 1 event each, press "Fill Plate"
- 7.9. After sorting check the presence/absence of objects per well under a microscope, optionally adjust the sort delay in the "Delay Setup Menu". The sort delay defines the time between detection of a particle and giving a sort command.
- 7.10. Add the plate with the dissociation buffer to the X-Y-Z stage
- 7.11. Select all required wells with 1 event and press "Fill Plate"
- 7.12. During sorting make sure the sample cup does not go empty, in case it is almost empty abort the sort, remembering the position of wells that were sorted already, add solution to the sample cup and start sorting again selecting the wells that are left; Do not 'Pause' the sorting because sometimes this leads to re-flush
- 7.13. After sorting, seal the plate and put it on ice until further use for dissociation
- 7.14. Washing of the biosorter in-between samples: Apply flushing procedure (step 2.9) until sheath solution comes into the sample cup, then change sample to fresh 40 mL

- ddH₂O acquire approximately 30 mL until no events are seen anymore, turn sheath flow to 48 in this case so the FOCA gets rinsed with higher speed
- 7.15. End-of-the-sort washing procedure: First apply 7.14, then do washes in the following order:
 - 7.15.1. Cleaning Solution Wash
 - 7.15.2. Ethanol Wash
 - 7.15.3. Water Wash
 - 7.15.4. Rinse the FOCA with another 40 mL ddH₂O from the sample cup
 - 7.15.5. Change a sample cup to 30 mL 70 % EtOH
 - 7.16. Turn off the biosorter and the computer
 - 7.17. Remove the solution from the waste container
 8. Single-cell dissociation of spheres
 - A. Murine liver:
 - a. Incubate on a thermomixer at 37 °C shaking at 300 rpm for 20 min
 - b. For the last 2 minutes shake at 700 rpm and place on ice after
 - B. Murine spleen:
 - a. Manually dissociate spheres by harshly pipetting up and down 50 times on ice using a multichannel P20 pipette
 - C. Crohn's biopsy:
 - a. Incubate two times 15 min at 37 °C and vortex plate in-between
 - b. Spin at 400 x g for 10 min at 4 °C
 - c. Remove supernatant carefully
 - d. Add 100 µl of Crohn's Digestion Buffer (see table 9)
 - e. Incubate at 37 °C and 300 rpm for 30 min
 - f. Inactivate enzyme activity by adding 50 µl of 20 mM EDTA/PBS and incubate at 37 °C, 300 rpm for 5 min
 - g. Spin plates at 4 °C, 400 rpm for 10 min
 - h. Remove supernatant leaving approximately 30-50 µl left in the wells and place on ice
 - D. CRC organoids:
 - a. Incubate at 37 °C, 300 rpm for 12 min
 - b. For the last 2 min shake at 700 rpm and place on ice
 9. Labeling of cells with sphere-specific barcode

Short description of the labeling approach

This step follows the protocol from McGinnis *et al.*, 2019²¹ with small adaptations. The technology was developed for single-cell hashing and its workflow will be summarized. It has three parts, 8 base pair (bp) long BCs with incorporated primer sequences and a poly(A)-stretch, a lipid anchor, and a lipid co-anchor²¹. The lipid anchor is mixed with a BC of interest, which hybridizes to the lipid anchor. This anchor with bound BC (called lipid-modified oligonucleotide, LMO) can easily integrate into a cell's hydrophobic membrane over its hydrophobic lignoceric acid amide on the 5' side²¹. The co-anchor binds the anchor and strengthens the integration of the complex to the membrane via its 3' palmitic acid amide²¹. The poly(A)-stretch of the BC can then be captured with scRNA-seq technologies, acting like any other mRNA²¹. McGinnis *et al.* designed 96 different BCs²¹, we extended this to 288 with a hamming distance of at least 3. The hamming distance measures the number of different nucleotides between two BCs, because of errors in sequencing this should be at least 3 to

reduce misidentifications and misallocations of BCs. During library preparation after cDNA amplification, the library gets split into two parts, the smaller fragments being the MULTI-seq BCs and the larger ones the cDNA transcripts. Later, these two pieces of information can be integrated²¹. The MULTI-seq workflow is summarized in figure 10.

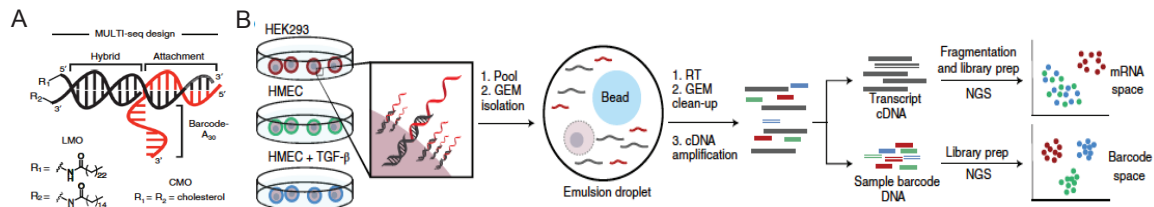


Figure 10: Overview of the MULTI-seq cell hashing approach. A) Structure of the lipid-modified oligonucleotide (LMO) design. B) Workflow: First cells of different samples are labeled with different BCs. Then cells are pooled and captured with scRNA-seq poly(A)-capture. After Gel beads-in-emulsion (GEMs) isolation, RT, clean-up and cDNA amplification the DNA is split into the large cDNA transcript fraction and the small sample barcode DNA fraction, which are used for separate library preparation. After sequencing and demultiplexing a single cell can be allocated to its transcripts and MULTI-seq BC over the Cell BC information and cells can be clustered based on mRNA space and Barcode space. Illustration from McGinnis C.S. et al., MULTI-seq: sample indexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* 16, 619-626, 2019, by permission of Springer Nature (<https://www.nature.com/nmeth/>).

The advantage of the MULTI-seq cell hashing approach for its use for spheres-seq is the possibility to quench the cell labeling approach, which allows the pooling of cells without washing samples individually. This is important for sphere-seq because the cell number of an individual sphere is very small and washing of cells within individual wells would lead to massive cell loss. The reaction can be quenched by adding high concentrations of albumin (bovine serum albumin, BSA; or fetal bovine serum, FBS; BSA is a component of FBS). Albumin binds lipids via its uncharged amino acid side chains over a nonpolar binding to the hydrocarbon chain of lipids¹⁶⁹. Therefore, residual Anchor:BC anchor complexes in solution are bound by albumin and are therefore prevented from integrating into cells from different spheres after pooling.

The MULTI-seq protocol was developed for 10X with large cell numbers. We therefore first adapted the concentration of LMOs to the small cell numbers within spheres. We then adapted the protocol to its usage with the BD Rhapsody Single-Cell Capture method. Therefore, the MULTI-seq protocol was combined with the Sample Tag Library Preparation Protocol from BD. The detailed protocol for BD and 10X will be explained in the following section. Subsection (I.) will be the BD approach, and (II.) will be the 10X approach. The protocol was optimized for 10X 5' v1.1. and BD v1, but might also be compatible with other versions.

9.1. Prepare Anchor:BC and Co-anchor solutions for appropriate sphere sizes (100-400 μ m: 50 nM; 400-800 μ m 100nM)

Tips for solution preparation:

- Anchor:BC: For 288 spheres pre-mix the Anchor with PBS in 3 columns of a 96-well plate and then add to 288 wells containing the BCs
 - For 50nM: in each well of 3 columns mix 13 μ l Anchor (1 μ M stock) and 234 μ l PBS, mix by shaking on the Thermomixer for 1 min at 700 rpm; then put 1 μ l of BC (1 μ M stock) to each of 288 wells and add 19 μ l of previously mixed Anchor/PBS.

- For 100 nM: premix 3 columns with 26 μ l Anchor (1 μ M stock) and 208 μ l PBS, then put 2 μ l of BC (1 μ M stock) to each of 288 wells and add 16 μ l of previously mixed Anchor/PBS.
 - Co-anchor: Dilute Co-Anchor in PBS for 288 spheres in 3 columns of a 96-well plate.
 - For 50 nM mix 13 μ l Co-anchor (1 μ M stock) with 247 μ l PBS;
 - For 100 nM mix 26 μ l with 234 μ l PBS.
- 9.2. Add 20 μ l Anchor:BC solution to the plates with dissociated spheres
 - 9.3. Mix plates using a thermomixer at 20 °C at 700 rpm for 1 min
 - 9.4. Put on ice and incubate for 5 min
 - 9.5. Add 20 μ l of Co-anchor solution
 - 9.6. Mix the plates using a thermomixer at 20 °C at 700 rpm for 1 min
 - 9.7. Put on ice and incubate for 5 min
 - 9.8. Add 100 μ l 10 % BSA/PBS (10 % BSA in low Glucose DMEM for liver)
 - 9.9. Mix the plates using a thermomixer at 20 °C at 700 rpm for 1 min
 - 9.10. Put on ice and incubate for 5 min
10. Pool cells and wash using FACS tubes
 - 10.1. Spin down the first time using appropriate speed and time
 - A. Murine liver: 300 x g, 10 min
 - B. Murine spleen: 400 x g, 10 min
 - C. Crohn's biopsy: 400 x g, 10 min
 - D. CRC organoids: 300 x g, 5 min
 - 10.2. Remove supernatant by decanting and pool leftover of all FACS tubes into one
 - 10.3. Wash 2 times with 1 % BSA/PBS (or 1 % BSA/low Glucose DMEM for liver), after 1st wash filter through a 40 or 70 μ m strainer depending on the cell types
 - A. Murine liver: 70 μ m
 - B. Murine spleen: 40 μ m
 - C. Crohn's biopsy: 40 μ m
 - D. CRC organoids: 40 μ m
 - 10.4. After the 2nd wash transfer cell suspension into a 1.5 mL low-bind Eppendorf tube and spin down a last time
 - 10.5. Resuspend in 1 % of BSA/PBS
 - I. BD: resuspend in 55 μ l
 - II. 10X: resuspend in 37.5 μ l
 11. Quality control of cell suspension:
 - 11.1. Mix 2.5 μ l of cell suspension with 2.5 μ l Trypan Blue Solution
 - 11.2. Load suspension into a hemocytometer chamber and check under a microscope
 - 11.3. Continue if the viability is at least 70 % and at least 10,000 cells are left in the solution
 12. Single-cell capture and cDNA synthesis
 - 12.1. Load a maximum amount of cells
 - I. BD:
 - a. Load up to 80,000 cells
 - b. Follow the protocol: Single Cell Capture and cDNA Synthesis
 - c. In short: single cells are loaded into a microwell chip; they will settle down within 15 min; then magnetic capture beads are applied; in each well, there is only space for one cell and one bead; after both are loaded lysis buffer is applied for 2 min, during which poly(A) mRNAs and MULTI-seq BCs bind to capture beads; using magnetic force, beads with captured cellular transcripts are

retrieved; retrieved beads are then used for RT followed by exonuclease I treatment

II. 10X:

- a. Load up to 30,000 cells
- b. Follow the protocol: Chromium Next GEM Single Cell V(D)J Reagent Kits v1.1
- c. In short: cells are mixed with master mix containing reagents for lysis and reverse transcription; Cell:Master mix solution and Gel beads are loaded separately into the 10X Chip; Beads and cells are then fused to generate single-cell GEMs; within GEMs, cells are lysed and RNA is reverse transcribed

13. cDNA amplification and splitting of libraries into whole-transcriptome analysis (WTA) and MULTI-seq fractions

I. BD:

- a. Recover the MULTI-seq product following the first steps of the 'Rhapsody WTA SampleTag' protocol
- b. The Sample Tag product in this case is the MULTI-seq product
- c. The MULTI-seq product is denatured off the beads during incubation at 95 °C
- d. Follow instructions for cDNA amplification for the WTA fraction following the protocol's sections of Random priming and extension (RPE), RPE purification, RPE PCR, and purification of RPE PCR amplification product using SPRI instead of AMPure beads for the cleanups

II. 10X:

- a. Follow instructions from Chromium Next GEM Single Cell V(D)J Reagent Kits v1.1 until cDNA Amplification
- b. During cDNA Amplification add the MULTI-seq primer
 - 50 µl Amplification Master Mix
 - 14 µl SC5' Feature cDNA Primers
 - 1 µl MULTI-seq primer (2.5 µM)
- c. Perform cDNA amplification
- d. Split cDNA fractions during 0.6X SPRI clean-up
- e. After incubation with SPRI beads, transfer the supernatant containing the small cDNA fragments to a new tube, this is the MULTI-seq fraction
- f. Continue cleanup of the large cDNA fragments, this is the WTA fraction

14. WTA library preparation

I. BD:

- a. Follow instructions from BD 'Rapsody WTA SampleTag' protocol for WTA index PCR using SPRI instead of AMPure beads for the cleanups

II. 10X:

- a. Follow instructions from Chromium Next GEM Single Cell V(D)J Reagent Kits v1.1 for 5' Gene Expression (GEX) Library Construction (Fragmentation, End Repair & A-tailing, followed by Adapter Ligation and Sample Index PCR)

15. MULTI-seq library preparation

I. BD:

- a. Follow BD instructions from BD 'Rapsody WTA SampleTag' protocol for Purifying Sample Tag PCR 1 products using SPRI instead of AMPure XP beads; use 1.8x SPRI for the cleanup
- b. Quantify concentration on Qubit using a dsDNA high-sensitivity kit
- c. Follow MULTI-seq protocol
- d. Take 3.5 ng for indexing PCR

- e. Mix the following
 - 26.25 μ l Kapa HiFi HotStart ReadyMix (2X)
 - 2.5 μ l Library Forward Primer
 - 2.5 μ l RPI primer (choose a unique one for each sample)
 - 3.5 ng DNA
 - Nuclease-free water to 50 μ l final volume
 - f. Perform indexing PCR:
 1. 95 $^{\circ}$ C – 5 min
 2. 98 $^{\circ}$ C – 15 sec
 3. 60 $^{\circ}$ C – 30 sec
 4. 72 $^{\circ}$ C – 30 sec
 5. Repeat steps 2-4 (8-12 times)
 6. 72 $^{\circ}$ C – 1 min
 7. 4 $^{\circ}$ C – hold
 - g. Perform clean-up with 1.6X SPRI beads:
 1. Add 80 μ l of SPRI beads
 2. Mix and incubate for 5 min at room temperature
 3. Put on a magnet and wait until the solution is clear
 4. Remove and discard the supernatant
 5. Wash 2 times with 80 % EtOH without disturbing the beads
 6. Let the beads dry for 1-2 min
 7. Resuspend beads in 25 μ l elution buffer (EB)
- II. 10X:
- a. Follow MULTI-seq protocol
 - b. Take the supernatant from the cDNA cleanup
 - c. Use 3.2X SPRI beads and 1.8X Isopropanol
 - d. Mix and incubate for 5 min at room temperature
 - e. Put on a magnet and discard the supernatant
 - f. Wash 2 times with 80 % EtOH without disturbing the beads
 - g. Dry beads for 2 min and resuspend in 50 μ l EB
 - h. Quantify concentration on a Qubit using a dsDNA high-sensitivity kit
 - i. Take 3.5 ng for indexing PCR
 - j. Mix the following
 - 26.25 μ l Kapa HiFi HotStart ReadyMix (2X)
 - 2.5 μ l Library Forward Primer
 - 2.5 μ l RPI primer (choose a unique one for each sample)
 - 3.5 ng DNA
 - Nuclease-free water to 50 μ l final volume
 - k. Perform indexing PCR:
 1. 95 $^{\circ}$ C – 5 min
 2. 98 $^{\circ}$ C – 15 sec
 3. 60 $^{\circ}$ C – 30 sec
 4. 72 $^{\circ}$ C – 30 sec
 5. Repeat steps 2-4 (8-12 times)
 6. 72 $^{\circ}$ C – 1 min
 7. 4 $^{\circ}$ C – hold
 - l. Perform clean-up with 1.6X SPRI beads:
 1. Add 80 μ l of SPRI beads

2. Incubate for 5 min at room temperature
 3. Put on a magnet and wait until the solution is clear
 4. Remove and discard the supernatant
 5. Wash 2 times with 80 % EtOH without disturbing the beads
 6. Let the beads dry for 1-2 min
 7. Resuspend beads in 25 µl elution buffer (EB)
16. Check the quality and concentration of the libraries:
- 16.1. Check the quality with Tape Station high sensitivity D1000 reagents and tapes
 - 16.2. Check the concentration with a Qubit 4 fluorometer using a dsDNA high-sensitivity kit
 - 16.3. Calculate the nM concentration of the libraries using the concentration from Qubit and the average peak size from Tape Station and the following formula:
 - $(\text{Concentration in ng/}\mu\text{l} / (660 \text{ g/mol} * \text{average peak size in bp})) * 1,000,000$
 - The 660 g/mol is the average weight of one bp
17. Sequencing
- Paired-end sequencing
 - WTA library: 30,000-50,000 reads/cell
 - I. BD:
 - Read 1: 60 bp
 - Read 2: at least 62 bp
 - Index read: 8 bp
 - II. 10X:
 - Read 1: 26 bp
 - Read 2: at least 88 bp
 - Index read: 8 bp
 - MULTI-seq library: 3,000-5,000 reads/cell
 - I. BD:
 - Read 1: 60 bp
 - Read 2: at least 8 bp
 - Index read: 6 bp
 - II. 10X:
 - Read 1: 26 bp
 - Read 2: at least 8 bp
 - Index read: 6 bp

7.1.3 Anticipated Results

The WTA library should have a peak around 250-1,000 bp; the concentration should be > 1 ng/µl. The MULTI-seq library should have a peak around 175 bp for 10X and around 230-250 bp for BD; the concentration should be in a range between 0.8-3 ng/µl. figure 11 shows examples of DNA profiles from Tape Station measurements.

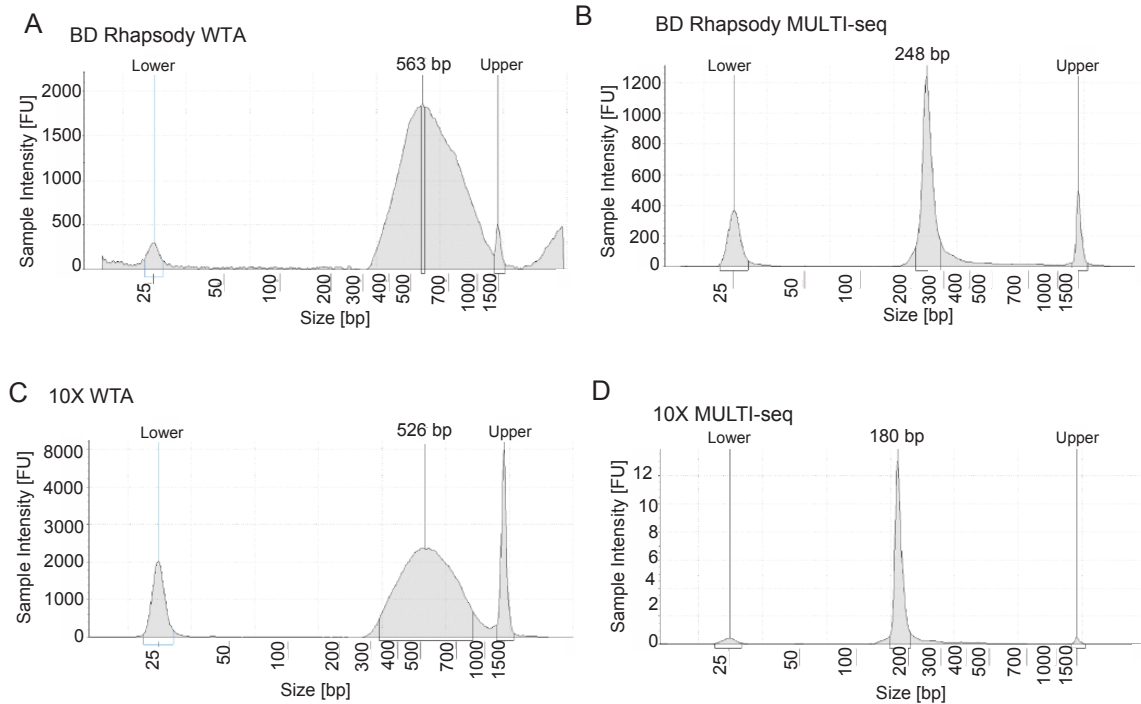


Figure 11: Examples of Tape Station profiles for A) BD Rhapsody WTA library, B) BD Rhapsody MULTI-seq library, C) 10X WTA library, D) 10X MULTI-seq library.

7.1.4 Troubleshooting

There are not enough cells at the end of the sphere-seq procedure to run a scRNA-seq experiment: The reasons for that could be cell loss during the washing steps or the spheres being too small. This could be addressed by being more careful during washing and decanting. Additionally, the centrifugation steps could be adapted, smaller cells like immune cells need longer (10 min at 400 x g) compared to epithelial cells (5 min at 300 x g). For very small spheres, one could consider carrier cells, which are of different cell types so they can be subdivided later on. Another reason for a low cell count at the end of the procedure could be that the sorting is not efficient, resulting in empty wells, to improve that one could optimize the sort delay.

Cell viability is too low: One could optimize the single-cell dissociation. With practice, the sphere-seq procedure is faster and viability might increase.

There is only a very small or no MULTI-seq peak: Make sure to remove all residual albumin (FBS and BSA) before MULTI-seq labeling of spheres. Make sure the dissociation is complete because if cells are still within clumps the MULTI-seq lipids cannot integrate into the cell's membrane because its surface is hidden by other cells. Reduce the freeze-thaw cycles of the MULTI-seq BC dilutions, the single-stranded DNA might degrade.

7.2 Computational analysis approaches for sphere-sequencing

This section describes the computational approach developed for the sphere-seq data analysis pipeline. It includes a detailed section for analysis packages used for different programming languages (Linux, R and Python). The procedure is explained in a continuous

manner step by step in an Arabic numbering format. Options for some steps are highlighted with Latin numbers or letters in alphabetical order. There are also sections highlighting the anticipated results and options for troubleshooting. The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

7.2.1 Packages for analysis

7.2.1.1 Linux programming

| Package | Version | Reference |
|-------------|------------|--|
| Bcl2fastq | 2.20.0.422 | Illumina |
| zUMIs | 2.9.4 | Parekh <i>et al.</i> , 2018 Gigascience ¹⁷⁰ |
| STAR | 2.5.2b | Dobin <i>et al.</i> , 2013 Bioinformatics ¹⁷¹ |
| Cell Ranger | 5.0.0 | 10X Genomics |

Table 10: Linux packages used in the sphere-seq analysis pipeline.

7.2.1.2 Python programming

Python version 3.8.10 was used

| Package | Version | Reference |
|-------------|---------|--|
| CellPhoneDB | 4.0.0 | Efremova <i>et al.</i> , 2020 Nat. Protoc. ⁹² |

Table 11: Python packages used in the sphere-seq analysis pipeline.

7.2.1.3 R programming

R version 4.1.0 was used

| Package | Version | Reference |
|----------------------|---------|--|
| biomaRt | 2.50.3 | Durinck <i>et al.</i> , 2005 Bioinformatics ¹⁷² ; Durinck <i>et al.</i> , 2009 Nat. Protoc. ¹⁷³ |
| deMULTIplex | 1.0.2 | McGinnis <i>et al.</i> , 2019 Nat. Biotech. ²¹ |
| Seurat | 4.0.3 | Hao <i>et al.</i> , 2021 Cell ¹⁷⁴ |
| scraper | 1.22.1 | Lun <i>et al.</i> , 2016 F1000Res ¹⁷⁵ |
| SingleCellExperiment | 1.16.0 | Amezquita <i>et al.</i> , 2020 Nat. Methods ¹⁷⁶ |
| dplyr | 1.0.7 | Wickham <i>et al.</i> , 2020 ¹⁷⁷ |
| tidyverse | 1.3.1 | Wickham <i>et al.</i> , 2019 J. Open Source Softw. ¹⁷⁸ |
| ggplot2 | 3.3.5 | Wickham <i>et al.</i> , 2016 Springer Link ¹⁷⁹ |
| batchelor | 1.10.0 | Haghverdi <i>et al.</i> , 2018 Nat. Biotechnol. ¹⁸⁰ |
| edgeR | 3.36.0 | Robinson <i>et al.</i> , 2010 Bioinformatics ¹⁸¹ ; McCarthy <i>et al.</i> , 2012 Nucleic Acids Res. ¹⁸² ; Chen <i>et al.</i> , 2016 F1000Res. ¹⁸³ |

Table 12: R packages used in the sphere-seq analysis pipeline.

7.2.2 Procedure

1. Demultiplexing of raw data after Illumina sequencing: Demultiplex the binary base call (BCL) files with Bcl2fastq to convert them to FASTQ files
2. FASTQ file processing:
 - 2.1. WTA scRNA-seq FASTQ files: generate count matrices with genes as rows and cell IDs as columns
 - I. 10X: Use the Cell Ranger pipeline with a GRCm38 v2020-A gene code for mouse data
 - II. BD:

- a. Use the zUMIs pipeline with the following gene alignments generated with STAR, for human samples GRCh38 v2020-A and for mouse GRCm38 vM25 with the option to fuse it to a GFP 3' UTR sequence in case there is GFP within the data
 - b. Convert ensemble IDs to gene names using the biomaRt package with musmusculus gene ensemble or hsapiens gene ensemble version 95
- 2.2. MULTI-seq FASTQ files: process with the deMULTIplex package: generates a MULTI-seq BC UMI matrix per cell; then an identity (MULTI-seq BC, Doublet or Negative) is assigned to each cell following the classification workflow; the instructions can be found on the GitHub page: <https://github.com/chrismcginnis-ucsf/MULTI-seq>
3. Use the Seurat package for scRNA-seq analysis: Generate a Seurat object
4. Integration of MULTI-seq BC and WTA information: match the cell IDs of the MULTI-seq BC and the gene count matrix of the WTA; this assigns each cell to a MULTI-seq BC and therefore sphere of origin
5. Normalize and scale WTA data using the Seurat function 'SCTransform'
6. OPTIONAL: Apply batch effect correction using mutual nearest neighbours (MNN) correction within the batchelor package
7. Quality control:
 - 7.1. Remove low-quality cells with lower than 200 features
 - 7.2. Remove low-quality cells with high mitochondrial content defined by the ratio of mitochondrial genes to cytoplasmic genes. This cut-off can be chosen differently for different samples, for example, epithelial cells from patient biopsies tend to have higher mitochondrial content compared to immune cells. Optionally chose different cut-offs for different cell type clusters. As a general rule of thumb use 10% as a cut-off.
 - 7.3. Remove doublets by applying an upper feature per cell cut-off, for example, 7,500. Chose an appropriate value based on a FeatureScatter plot plotting nCount_RNA against nFeature_RNA per cell
8. Clustering and annotation:
 - 8.1. For clustering choose the top 2,000 highly variable features, when doing batch effect correction uses a combined number of variable features, for example, 6,000 for 10 samples.
 - 8.2. Choose significant principal components (PCs) using an Elbow plot
 - 8.3. Cluster cells first broadly using a low resolution and annotate cell clusters into broad cells like T, B, Myeloid, Stromal,...
 - 8.4. Then use the Seurat function 'FindSubCluster' to re-cluster broadly annotated clusters and annotate subtypes based on known marker genes or DEGs identified with the Seurat function 'FindAllMarkers', it uses a non-parametric Wilcoxon Rank Sum test (default parameter: min.pct 0.25 and logfc.threshold 0.25).
9. Integration of sphere size and fluorescence signal from the Biosorter:
 - 9.1. Generate a standard curve with TOF measurement from standard-size beads
 - 9.2. Fit TOF measurements of sorted spheres into the linear model to calculate sphere size in diameter
 - 9.3. Normalize the fluorescence signal with the sphere size
 - 9.4. Overlay the plate ID and well position with MULTI-seq BC identity and integrate sphere size and normalized fluorescence information with the MULTI-seq BC column of the Seurat object

10. Cells per sphere cut-off: Apply a cut-off of a minimum cell number per sphere and remove all other spheres
11. Reconstruction of the spatial position of spheres:
 - I. Based on a continuous gene expression axis: Explained on the example of the lobule layer axis from central to portal veins within the liver
 - a. Follow the workflow adapted from the previously described pipeline in Ben-Moshe *et al.*, 2022¹¹⁵
 - b. Extract only specific cell types where the signature is expressed, in this example LECs
 - c. Produce pseudo-bulks of each sphere from the cell type of interest (f.e. LECs) by calculating the average expression of normalized signature genes; signature A and B; in this example of CV landmark (cLM) and PV landmark (pLM) genes from Halpern *et al.*, 2018⁶⁵
 - d. Use two normalization steps: First log-normalize genes and then divide the normalized expression by the highest expression across spheres; this ensures that all genes contribute equally to the reconstruction algorithm
 - e. Calculate a zonation coordinate (ZC) for each sphere by dividing the sum of double normalized signature A genes by the sum of signature A and B genes; in this example $ZC = pLM/(pLM+cLM)$
 - f. Rescale ZCs so that 0 is the most signature A and 1 is the most signature B. In this example, 0 will be the most central and 1 will be the most portal. This can be done because the ZC is a relative score and depends on the cells and landmark genes. Additionally, signature A and B genes are not exclusively expressed in cells of one or the other group, there is a gradient of gene expression.
 - g. ZCs can then be grouped into different zones or areas. In this example lobule layers L1-L10 (L1: $ZC < 0$, L2: $ZC < 0.2$, L3: $ZC < 0.3$,...L10: $0.9 > ZC \leq 1.0$)
 - II. Based on cell type presence:
 - a. Choose a cell type of interest
 - b. Spheres containing the cell type of interest are spatial group A, spheres lacking the cell type of interest are spatial group B
 - c. For example, spheres with metastatic cells are group A (proximal to micro-metastasis), and spheres lacking metastatic cells are group B (distal to micro-metastasis)
12. Analysis of DGE between different spatial groups:
 - I. Within a continuous zonation axis:
 - a. Use the edgeR package
 - b. Make pseudobulks per sphere of the cell type of interest (Average single cell counts per sphere)
 - c. Only consider spheres with at least 2 of the cells of interest or 5 if there are many cells
 - d. Remove lowly expressed genes with the function 'filterByExp'
 - e. Fit a negative binomial generalized log-linear model with the continuous zonation axis as an ordered factor covariate (in the example of the lobule layer $L1 < L2 < L3 < L4 < L5 < L6 < L7 < L8 < L9 < L10$)
 - f. Use the sample identity as a covariate blocking factor to account for batch effects
 - g. Use 'glmQLFTest' function to identify genes that are different from 0 and have a Benjamin-Hochberg adjusted p-value smaller than 0.05. For studying linear

zonation patterns use `coef = '.L'` and `coef = '.Q'` for studying parabolic patterns (genes high in lobule layer L5 and low in L1 and L10 for example)

- II. Between two groups:
 - a. Use the edgeR package
 - b. Follow the description of 12.I.a-f, however, instead of using an ordered factor, use two groups as factor covariates
 - c. Also, a non-parametric Wilcoxon signed-rank test using the 'ggsignif' function from the ggplot2 R package can be applied
13. Analysis of differences in cell type abundance between two groups:
 - a. Follow the workflow previously described in Lun *et al.*, 2017¹⁸⁴
 - b. Compute normalized log counts from cluster abundances within two groups using the 'cpm' function of edgeR, accounting for the total number of cells from each sample
 - c. Specify a design matrix, use group labels as covariates
 - d. Use sample identity as a blocking factor
 - e. Estimate the dispersion parameter using the 'estimateDisp' function with `trend = 'none'`
 - f. For each cell type fit a negative binomial generalized log-linear model with 'glmQLFit' function (`robust = TRUE`, `abundance.trend = FALSE`)
 - g. Use the function 'glmQLFTest' to identify cell types significantly different from 0 with a Benjamin-Hochberg adjusted p-value of 0.05
14. Ligand-receptor interaction analysis between two groups:
 - a. Use the python package CellPhoneDB
 - b. CellPhoneDB uses a human L-R database, if data is mouse convert mouse gene symbols to human gene symbols using the biomaRt package
 - c. Follow the workflow on the Github page:
<https://github.com/ventolab/CellphoneDB>
 - d. Group the sphere into two spatial groups and perform CellPhoneDB analysis separately on grouped objects
 - e. CellPhoneDB takes cell cluster labels and gene expression as input to match with known L-R interaction pairs from the CellPhoneDB public repository
 - f. Use the default parameters
 - g. Use the following outputs from each of the two analyses: A data.frame with average L-R expression between two cell types represented by a mean value, and a data.frame of p-values for each L-R pair determined using a null distribution of means from randomly permuted annotated cluster labels
 - h. Only consider significant interactions ($p\text{-value} \leq 0.05$)
 - i. Calculate the difference in L-R interaction scores (mean values) between two groups

7.2.3 Anticipated results and troubleshooting

In the MULTI-seq BC classification process from step 2.2, to allocate cells to spheres of origin, there should be a clear clustering of MULTI-seq BCs. Positive cells in red should cluster in a clear differentiated cluster (Fig. 12, left). If there is no clear differentiation of BC signals within clusters (Fig. 12, right) or if there is no signal at all (Fig. 12, middle) BC should be removed from classification schema.

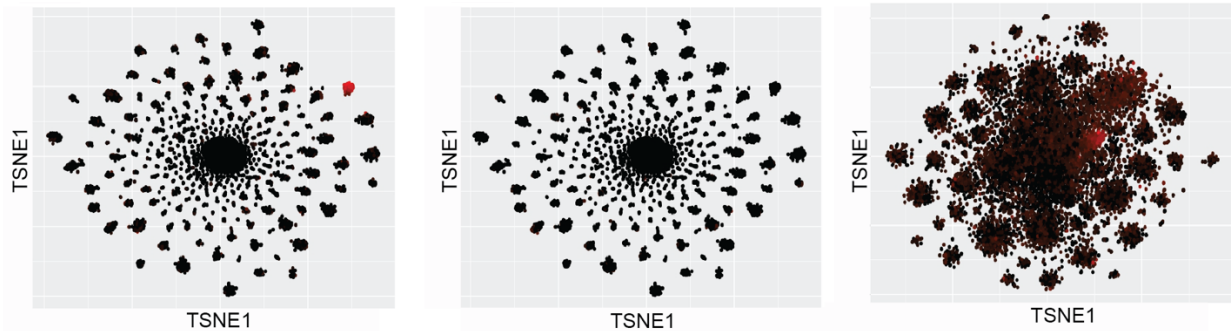


Figure 12: *t*-distributed stochastic neighbor embedding (tSNE) plot of cells clustered by MULTI-seq BCs. The left plot shows cells positive for one BC, the middle plot shows a BC that is not present and should therefore be removed from the classification schema and the right plot shows BC signal in more than one cluster and should therefore also be removed from the classification process.

7.3 Specific methods and materials used in this thesis

7.3.1 Sphere-sequencing of murine metastatic and healthy livers

7.3.1.1 Induction of metastasis in mouse liver and tissue collection

VilCreERT2; APCfl/fl; Tp53fl/fl; KrasG12D/wt (AKP) organoids were obtained from Owen Sansom (Beatson Institute for Cancer Research in Glasgow) and modified to harbor an additional knockout in *Smad4* which results in VilCreERT2; APCfl/fl; Tp53fl/fl; KrasG12D/wt; Smad4KO (AKPS) organoids. Organoids were cultured in Matrigel (Corning) as stated previously in Sato *et al.*¹⁸⁵. Supplement culture medium (Advanced DMEM/F12, Life Technologies™) with the following compounds: 10 mM HEPES (Life Technologies™), 1 x N2 supplement (Life Technologies™), 100 mg/mL Penicillin/streptomycin, 2 mM L-Glutamine (Life Technologies™), 1mM N-acetylcysteine (Sigma-Aldrich) and 1x B27 supplement (Life Technologies™). AKPS organoids (4 domes per recipient mouse) were washed in ice-cold PBS repeatedly to eliminate all Matrigel. Then the organoids were dissociated into small cellular communities and resuspended in 50 µl PBS. They were then loaded into a syringe (BD, MicroFine, 0.3 ml, 30 G). C57BL/6 mice (8 weeks old, male, from Janvier-Labs) were injected subcutaneously with Carprofen (5 mg/kg) 30 min before surgery following anesthetization with isoflurane gas and they were put on a 37 °C thermal pad to keep warm. The mice were then shaved and disinfected with betadine (mix of 0.5% lidocaine (5 mg/ml), and 0.25 % bupivacaine (2.5 mg/ml) over a subcutaneous injection at the planned incision line. After disinfection, the organoids were injected beneath the capsule of the spleen and incubated for 10 min followed by splenic resection by ligation. Sterile PBS was then used to wash the wound and using an absorbable polyglactin suture (Vicryl 4- 0 or 5-0 coated) the peritoneal wall was closed. The skin was closed with wound clips. During the wake-up phase, Buprenorphine (0.1 mg/kg) was injected subcutaneously. During the wake-up phase, the animals were monitored and also in the following days after surgery. After 14 days the experiment was terminated. All procedures were performed according to the Swiss Guidelines upon the approval of the Cantonal Veterinary Office Basel-City. Tissue was collected as stated in the sphere-seq experimental protocol in section 7.1 step 4.A.

7.3.1.2 Sphere-sequencing experimental procedure

The experimental procedure was performed according to the protocol in section 7.1. Sphere-seq was done with liver of 9 injected mice, from which 3 had macroscopically visible metastasis, and one healthy liver. Spheres were sorted with sizes between 200–450 μm and single cells were captured using the BD Rhapsody platform. 50 nM MULTI-seq lipid Anchor:BC and CoAnchor were used. Sphere-seq libraries were sequenced on a NovaSeq 6000 system (Illumina) using NovaSeq SP Reagent Kits (100 cycles) v1.5 and S4 Reagent kits (200 cycles) v1.5 with XP workflow.

Testing of sorting purity of single spheres per well. Liver tissue was partially dissociated and used for sorting of spheres between 200 and 450 μm in diameter. Nine plates were sorted and used for imaging on a Leica Thunder Imaging System. Plate scans were then visually inspected to count wells with none, one or multiple fragments.

7.3.1.3 Sphere-sequencing Computational analysis

For the computational analysis the pipeline described in section 7.2 was followed with specific parameters described in this section. The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

Estimation of true cell number per liver sphere. spheres were assumed to be spherical, and their volume was calculated for different sizes (50, 100, 150, 200, 250, 300, 350, 400, 450 μm) using the formula for spherical volume ($\frac{4}{3} * \pi * r^3 * 1.07 \text{ g/mL}$, r is the radius, 1.07 g/mL is the density). The density of the liver was estimated to be around 1.07 g/mL¹⁸⁶ and 1 g of liver was estimated to have 135 million hepatocytes¹⁸⁷. 20 % NPC¹¹¹ were added resulting in 162 million cells per gram liver. It was then calculated that there are approximately 151,401,869 million cells per cm^3 and this value was then used to calculate the estimate amounts of cells in $x \mu\text{m}^3$ spheres.

Single cell count matrix generation. zUMIs was used to generate the single cell count matrix using the GRCm38 vM25 gene code. For three samples (S1-3) deeper sequencing was required. FASTQ files were merged and then used as input for zUMI.

Batch effect correction and normalization. 10 samples were integrated while correcting for batch effects. After integration, the combined object underwent SCT normalization.

Quality control. Low quality cells (< 200 features, > 20 % of reads mapped to mitochondrial genes) and doublets (> 7,500 features) were removed.

Clustering and annotation. 10 MNN corrected PCs were used for clustering in a Uniform Matrix Approximation and Projection (UMAP) two-dimensional space. A resolution of 1 was used to cluster and annotate broad cell types. Sub clustering was then used to annotate cell subtypes. For annotation, markers from the Liver Cell Atlas from Guilliams *et al.* were used⁹⁸.

Cells per sphere cutoff. Spheres with less than 5 cells were removed.

Reconstruction of spheres into lobule layer positions. ZC per sphere was calculation from LECs with central and portal vein landmark genes⁶⁵. Spheres were then grouped into 8 lobule layers because there were not many spheres from the most central and most portal (L1-L3, L4, L5, L6, L7, L8-L10). To control the accuracy of the algorithm, landmark gene expression in hepatocytes from grouped spheres of pericentral (L1-L5) and periportal (L6-L10) areas was analyzed.

Split of dataset for different analysis. Samples were split into three datasets, healthy liver (n=1), livers from mice that were injected with CRC organoids (n=9) for liver zonation analysis, and liver samples with visible metastasis and high abundance of metastatic cells (≥ 20 cells) (n=3, S3, S6, S7) were used for the analysis of metastatic niches.

Grouping based on cell type presence – classification of metastatic distance. Only taking samples S3, S6 and S7 into consideration; Spheres with metastatic cells were defined as 'proximal' and spheres lacking metastatic cells as 'distal'.

Analysis of zonation specific genes. DEGs were analyzed within a continuous zonation axis (L1-L3<L4<L5<L6<L7<L8-L10) for LECs and KCs. Genes that were identified in perturbed liver samples were plotted in pericentrally and periportally grouped spheres from healthy liver with a non-parametric Wilcoxon signed-rank test applied from the 'ggsignif' function in ggplot2¹⁷⁹ R package.

Analysis of differences in cell type abundance. Analyses was done between pericentral (L1-L5) and periportal (L6-L10), and between proximal and distal areas.

L-R interaction analysis. Analysis was performed between pericentral (L1-L5) and periportal (L6-L10), and between proximal and distal areas.

Assessment of bias between different sphere sizes and cell number cutoffs. The differences between sphere size and cell counts of spheres between two groups (pericentral and periportal, distal and proximal) were assessed with a non-parametric Wilcoxon signed-rank test using the 'ggsignif' function from ggplot2 R package¹⁷⁹. Spheres were then grouped into two different size ranges (211-325 μm and 326-457 μm) to test the influence of different sphere sizes. For testing the influence of different cell counts an object with spheres having at least 5 cells/sphere was compared to an object with spheres that have at least 20 cells/sphere. Differences in cell type proportions were assessed by UMAP plotting, split by different scenarios. DEGs of KCs and LECs were assessed between two groups as described previously with batch and lobule layer as blocking factors covariates. Zonated genes were assessed in LECs from different scenarios as described previously.

7.3.2 Sphere-sequencing of the organoid mixing species experiment

7.3.2.1 Organoid cultivation and collection

Cultivating human CRC organoids. Human CRC organoids were obtained from the laboratory lead by Salvatore Piscuoglio (Visceral Surgery Research Laboratory, University of Basel) and cultivated as previously described¹⁸⁵. The following supplements were added to the culture medium (Advanced DMEM/F12, Life Technologies™): 10 mM HEPES (Life Technologies™), 2 mM Glutamax (Gibco), 50 ng/ml Human epidermal growth factor (hEGF) (Stemcell

#78006.1), 10 mM Nicotinamide (Sigma Aldrich), 500 nM A83-01 (Stemcell #100-0245), 1x B27 supplement (Life Technologies™), 100 ng/ml human R-spondin (LuBioScience #120-38-20), 1.25 mM N-acetyl-L-cysteine (Sigma Aldrich), 100 ng/ml Recombinant human Noggin (Stemcell #78060), 10 μM Y-27632 dihydrochloride (Rock inhibitor) (Stemcell #72304), 10 nM Prostaglandin E2 (PGE2) (Stemcell #72634), 10 nM Gastrin (Sigma Aldrich, #G9145) and 10 μM SB202190 (Stemcell #72634).

Cultivating mouse CRC organoids. ViiCreERT2;APCfl/fl;Tp53fl/fl;KrasG12D/wt (AKP) organoids were obtained from Owen Sansom (Beatson Institute for Cancer Research in Glasgow). The plasmid pMSCV-loxp-dsRed-loxp-eGFP-Ruo-WPRE¹⁸⁸ was used to label them with GFP. They were cultured as described in section '7.3.1.1 Induction of metastasis in mouse liver and tissue collection' while adding 100 μg/ml murine recombinant Noggin to the medium (LuBioScience, #250-38-250).

Organoid collection. Collect in PBS as described in section 7.1 step 4.D and mix mouse and human organoids in a ratio 1:1.

7.3.2.2 Experimental procedure

For the experimental procedure the protocol in section 7.1 was followed. Mixed organoids were used for sorting based on GFP signal. 144 GFP+ mouse and 144 GFP- human organoids were sorted. Single cells were captured using the BD Rhapsody platform. 50 nM MULTI-seq lipid Anchor:BC and CoAnchor were used. Sphere-seq libraries were sequenced on a NovaSeq 6000 system (Illumina) using NovaSeq SP Reagent Kits (100 cycles) v1.5 kit.

7.3.2.3 Computational analysis

For the computational analysis the pipeline described in section 7.2 was followed with specific parameters described in this section. The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

Single cell count matrix generation. zUMIs was used to generate the single cell count matrix using a fused gene code of GRCm38 v2020-A, GRCh38 v202-A and GFP 3' UTR sequence.

Quality control, clustering and annotation. First low-quality cells (< 200 features, > 30 % of mitochondrial genes) were removed. Followed by UMI count normalization and scaling using 'SCTransform' from the Seurat package. Single cells were then clustered in UMAP space using 10 PCs. Clusters were then annotated to human and mouse depending on expression of species-specific genes. Cells were also annotated based on their sphere BC and GFP signal of sorted spheres.

Removal of cell free RNA with DecontX. When we explored the data, we realized that there were contaminations of human or mouse reads in cells of the opposite species due to cell free RNA, even after low quality cells were removed. Therefore *decontX* function within the R package *celda*¹⁸⁹ (v1.12.0) with default parameters was used to remove the contamination due to cell free RNA.

Analysis of the amount of correctly and incorrectly assigned cells. Cells were annotated based on two species specific identifies. First the annotation based on the gene expression and secondly based on the GFP signal due to the sphere BC. These two annotations were matched and proportions of correctly and incorrectly assigned cells per sphere could be calculated. A cell was incorrect for example if the GFP signal and sphere BC annotated it as mouse, however based on gene expression it was human. There were three spheres that were 100 % incorrectly assigned cells and they were, therefore, allocated to the opposite GFP condition because most probably there was an error in fluorescent sorting.

7.3.3 Sphere-sequencing of mouse healthy spleen

7.3.3.1 Experimental procedure

Sphere-seq of murine spleens. For the experimental procedure the protocol in section 7.1 was followed. Two healthy spleens were processed for sphere-seq with 96 BC each. Sphere sizes were between 200-400 µm. Single cells were captured using the 10X Genomics Platform. 25 nM for sample 1 and 50 nM for sample 2 of MULTI-seq lipid Anchor:BC and CoAnchor were used. Sphere-seq libraries were sequenced on a NovaSeq 6000 system (Illumina) using NovaSeq SP Reagent Kits (100 cycles) v1.5 kit.

Conventional scRNA-seq of spleen with 5 MUTLI-seq BC. Spleen was squished through a 70 µm strainer and single cells were collected by centrifugation at 400 x g for 10 min at 4 °C. Cell pellet was resuspended in red blood lysis buffer (ACK Lysing Buffer, Thermo Fisher Scientific). Cells were split into 5 samples with the same cell numbers each and they were labeled with 2 µM MULTI-seq reagents (barcodes used: Bar8, Bar46, Bar54, Bar62 and Bar84) following the MULTI-seq protocol²¹. Single cells were then given to the Functional Genomics Facility Zürich (FGCZ) at UZH in Zürich who performed the single cell capture using the 10X Genomics Platform and 3' capture beads and sequencing on a NovaSeq 6000 system (Illumina) using NovaSeq SP Reagent Kits (100 cycles) v1.5 kit.

7.3.3.2 Computational analysis

For the computational analysis the pipeline described in section 7.2 was followed with specific parameters described in this section. The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

Single cell count matrix generation. Cell Ranger was used to generate the single cell count matrix using the GRCm38 v2020-A gene code.

Quality control, clustering and annotation. Two sphere-seq experiments of healthy spleen were merged without batch effect correction. Low quality cells and doublets (< 200 features, > 10 % mitochondrial reads, > 6,000 features) were removed. UMI counts were scaled and normalized using 'SCTransform' from the Seurat¹⁷⁴ package and then clustered in UMAP space using 10 PCs. Marker gene from Medaglia *et al.*⁵⁹ were used for annotation and sphere with less than 5 cells were removed.

Conventional scRNA-seq of spleen with 5 MUTLI-seq BC. Cells were allocated to their sample BC following the workflow from the deMULTIplex²¹ R package (v1.0.2).

7.3.4 Sphere-sequencing of Crohn's disease biopsies

7.3.4.1 Experimental procedure

For the experimental procedure the protocol in section 7.1 was followed. Two patient biopsies from the ileocecal valve, which encompasses parts of the ileum (small intestine) and the caecum (colon) were processed, from each two samples (inflamed and one non-inflamed). Samples were processed 4-6 hours after surgery. Spheres of around 500 μm diameter in size were manually picked instead of sorted with the biosorter. For both samples, spheres from inflamed and non-inflamed biopsies were picked: Patient 1: non-inflamed Bar145-Bar192 and inflamed Bar1-144, 193-288; Patient 2: non-inflamed Bar97-192 and inflamed Bar1-96, 193-288. 100 nM MULTI-seq lipid Anchor:BC and CoAnchor were used for labelling and single cells were processed using the BD Rhapsody platform. Sphere-seq libraries were sequenced on a NovaSeq 6000 system (Illumina) using NovaSeq SP Reagent Kits (100 cycles) v1.5 kit.

7.3.4.2 Computational analysis

For the computational analysis the pipeline described in section 7.2 was followed with specific parameters described in this section. The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

Single cell count matrix generation. zUMIs was used to generate the single cell count matrix using the GRCh38 v202-A gene code was used.

Quality control, clustering and annotation. Sphere-seq data of both patients were merged without batch effect correction. Low quality cells (< 200 features) were removed before clustering in UMAP space using 10 PCs. No mitochondrial cutoff was applied because the epithelial cells in biopsies have much higher mitochondrial genes because samples cannot be processed fresh, and we did not want to lose all of them. Cell types were first broadly annotated at resolution 0.5 and then sub-clustered to annotate subtypes while removing clusters with an overrepresentation of mitochondrial gene in their top 5 DEGs (function 'FindAllMarkers' from Seurat¹⁷⁴ with default parameters for DGE analysis). Cell type markers were used from Martin *et al.*¹⁹⁰ and Smillie *et al.*¹⁹¹. Spheres with less than 5 cells were removed from further analysis.

Fibrotic vs. Non-fibrotic analysis in patient 1. Patient 1 was used for grouping spheres into fibrotic and non-fibrotic spheres based on the presence of fibroblasts, defined by S100A4 expression¹⁹². These two groups were then used to analyze DEGs, cell type abundance and L-R interactions.

7.3.5 Conventional single-cell RNAseq of murine metastatic liver

7.3.5.1 Experimental procedure

Two samples of metastatic livers (each one lobe) were used for *ex vivo* digestion after perfusion with PBS (Gibco). Samples were cut into small pieces and incubated in 1.5 ml Eppendorf tubes with 1 ml of dissociation cocktail (the same as is used for sphere-seq) at 37 °C shaking with 300 rpm. After incubation samples were squished through a 100 μm strainer

(Falcon) using the plunger from a syringe and cells were collected in low glucose DMEM medium (1 g/L D-Glucose/L-Glutamine, Pyruvate; Gibco) containing 1 % BSA (Sigma Aldrich). Cells were counted using a hemocytometer and 15,000 cells were loaded on a BD Rhapsody Single Cell capture machine. Library preparation was done following instructions from BD 'Rapsody Whole Transcriptome analysis' protocol using SPRI instead of AMPure beads for the cleanups. Final libraries were then sequenced on a NovaSeq 6000 system (Illumina) using S4 Reagent kits (200 cycles) v1.5 with XP workflow (Read 1: 60 bp, read 2: 100 bp, index read: 8 bp).

7.3.5.2 Computational analysis

Demultiplexing and single-cell matrix generation. Demultiplexing of BCL files was done using bcl2fastq (v2.20.0.422, Illumina). FASTQ files were then processed to a single-cell matrix using zUMIs¹⁷⁰ (v2.9.4) with a gene alignment generated with STAR¹⁷¹ (v2.5.2b) using mouse GRCm38 vM25 fused to GFP 3' UTR. Ensemble IDs were then converted to gene names using the biomaRt^{172,173} (v2.50.3) package with musmusculus gene ensemble version 95. Demultiplexing of raw data after Illumina sequencing: Demultiplex the binary base call (BCL) files with Bcl2fastq to convert them to FASTQ files.

Quality control, clustering and annotation. The two samples were merged without batch effect correction. Low quality cells and doublets (< 200 features, > 20 % mitochondrial reads, > 7,500 features) were removed. UMI counts were scaled and normalized using 'SCTransform' from the Seurat¹⁷⁴ package and then clustered in UMAP space using 10 PCs. Marker gene from the liver cell atlas⁹⁸ were used for annotation.

Comparison with sphere-seq. Before applying any quality cut-offs, the median percentage of mitochondrial genes was analyzed and compared between both protocols (conventional scRNA-seq and sphere-seq). Additionally, the cell type presence between both protocols were analyzed in UMAP split plots.

7.3.6 Visium of mouse metastatic liver

7.3.6.1 Tissue collection

Mouse liver tissues were collected as described in 7.1.2/4a and then a small part of one lobe was embedded in O.C.T.TM compound (Tissue-Tek) snap-frozen in liquid nitrogen, by floating the embedded tissue within a metal beaker filled with isopentane (Sigma Aldrich). Tissues were then stored at -80 °C.

7.3.6.2 Experimental procedure

Library preparation. Embedded and snap-frozen tissues were used for slicing of 10 µm sections with a cryostat (Leica CM3050S). The Slices were put within the capture area of a 10X Visium Spatial Gene expression slide that was calibrated to -20 °C within the cryostat. The slide was processed the following day and stored at -80 °C in a slide container until then. For the generation of cDNA libraries, the manufacturer's instructions were followed. In short: Tissue on slides underwent methanol fixation and hematoxylin and eosin (H&E) staining to assess tissue morphology and quality. The following steps were performed after: tissue lysis,

RT, second strand synthesis and cDNA denaturation. These steps were done on the slide. Permeabilization time, previously assessed with the Tissue Optimization Kit and Protocol, was 10 min. After that cDNA was transferred into PCR tubes following qPCR to measure its concentration. The next step was cDNA amplification with qPCR defined cycle numbers, followed by the final steps of: end repair, A-tailing, adapter ligation and index PCR.

Quality assessment and sequencing. Quantity and quality of libraries were measured using a dsDNA high-sensitivity (HS) kit (Life Technologies #Q32854) on a Qubit 4 fluorometer (Thermo Fisher) and a high sensitivity D1000 reagents and tapes (Agilent #5067-5585, #5067-5584) or high sensitivity D5000 reagents and tapes (Agilent #5067-5593, #5067-5592) on a TapeStation 4200 system (Agilent Technologies). Paired-cell sequencing was performed: NovaSeq 6000 system (Illumina) with NovaSeq SP Reagent Kits (100 cycles) v1.5 using the following read configuration: read 1: 28 bp, read 2: 82 bp, index read 1: 10 bp, index read 2: 10 bp and 50,000 reads/spot.

7.3.6.3 Computational analysis

The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

Demultiplexing. BCL files were demultiplexed with Bcl2fastq v2.20.0.422 from Illumina which converts them to FASTQ files.

Pre-processed using Space Ranger. FASTQ files were pre-processed with the Space Ranger pipeline (v1.2.0) (10X Genomics) using GRCm38 v2020-A gene code.

Clustering and spatial area annotation. Both samples underwent separate processing. They were normalized and scaled using 'SCTransform' from Seurat¹⁷⁴ and clustered in UMAP space with 10 PCs. Annotation of clusters was based on landmark genes of hepatocytes and done to allocate spots into portal vein, central vein and metastatic areas. Portal and central vein areas were considered 'distal' and metastatic areas as 'proximal' for analysis of metastatic areas.

Assessment of the number of gene features per spatial area. Mean values of gene feature counts were calculated from proximal and distal areas to metastatic sites. These were compared to mean values from sphere-seq.

Batch effect correction. Visium samples were then merged using the MNN batch correction method within the batchelor¹⁸⁰ package (v1.10.0).

Deconvolution. Visium spots were deconvoluted using sphere-seq data as a reference. Top 20 genes from each cell type were used which also showed sufficient expression in Visium datasets. Genes were then used to deconvolute spots with the SCDC¹⁹³ package (v0.0.0.9000). At least 75 % of genes had to be assigned to a specific cell type to annotate spots, others were annotated as mixed.

7.3.7 Public Visium data analysis

The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

Visium datasets from Williams et al.⁹⁸ were used (healthy liver and NAFLD (non-alcoholic fatty liver disease) and spots were annotated for central, mid, periportal and portal zones based on zoned marker genes from hepatocytes to test with sphere-seq identified zoned gene expression.

7.3.8 Molecular Cartography of mouse metastatic liver

7.3.8.1 Tissue collection

Tissue collection was done as described for Visium.

7.3.8.2 Experimental procedure

Sample preparation and pre-processing. The procedure has been previously described⁹⁸. In short, 4 liver samples (two with visible metastasis, two that were injected with CRC organoids, but no visible metastasis were formed) were used for sectioning 10 µm thick slices and placed on capture areas of slides from Resolve BioSciences, as explained for Visium. These slides were stored at -80 °C and sent on dry ice to Resolve BioSciences. There, the samples were processed further using the following procedure: Sample fixation and 100-plex combinatorial smFISH. During every cycle of smFISH colored probes were developed, imaged and decolorized. This resulted in a unique combinatorial BC of each target gene. We chose 100 genes from sphere-seq analysis (20 genes for defining cell types and 80 genes to validate genes found in spatial analysis of sphere-seq). Resolve BioSciences used a Zeiss Celldiscoverer 7 microscope for imaging with a magnification of 25x imaging 9 rounds and 16 z-stacks per region. Computationally spots were then segmented, fluorescence background was removed and alignment of images from different rounds was done. The result was a profile from each pixel including the information of 16 values (two color channels, 8 imaging rounds = 16 images).

7.3.8.3 Image analysis in ImageJ

ImageJ with genexyz PolyLux tool plugin from Resolve BioSciences was used for image analysis.

7.3.8.4 Computational analysis

The detailed code is stored in the GitHub repository: <https://github.com/Moors-Code/Sphere-sequencing>.

Cell segmentation. Cellpose¹⁹⁴ (v.2.0.4) was used for nuclei segmentation based on DAPI images with a model of pretrained nuclei and the following parameters: flow_threshold 0.5, cellprob_threshold -0.2. Nuclear segments were expanded to 10 pixels (1.38 µm, 1 pixel = 138 nm) using the function 'expand_labels' from scikit-image. Then transcripts were assigned

to each expanded segment which represents a cell. All segments that were larger than the sum of 4 median absolute deviation (MAD) and the median segment area were excluded from downstream analysis. Low quality cells were removed during clustering and annotation. Low quality cells or clusters were identified if they could not be annotated properly to a certain cell type or the clusters were very small only showing very view DEGs.

Clustering and annotation. Generated count matrix from cell segmentation was used for normalization and scaling within the Seurat¹⁷⁴ workframe ('SCTransform'). Clustering in UMAP space was done with 10 PCs, followed by annotation using cell type marker genes. Clusters that could not be annotated properly were removed. Annotation was then projected onto DAPI images in ImageJ based on x-y coordinates of segmented cells.

Integration of spatial feature area. Landmark genes for central vein (*Cyp2e1*), portal vein (*Cyp2f2*) and metastatic sites (*Gpx2*) were used for area visualization. Followed by manually drawing and exporting of x and y coordinates of spatial areas. X and y coordinates were then matched with these of segmented cells to allocate single cells to spatial areas.

DGE analysis between two groups. This analysis was following the workflow of sphere-seq, but instead of generating pseudo bulks from spheres, they were generated from spatial feature areas.

Investigation of differences in cell type abundance comparing two groups. The workflow of sphere-seq was followed for this analysis.

Colocalization analysis. The 2D space of the centroids from segmented areas was used to generate a spatial neighborhood graph based on the Euclidean distance. Vertices show the cells connected by edges if their distance is $< 10 \mu\text{m}$. A kd-tree based nearest neighbor search was used to construct the graph within a pre-defined radius of $10 \mu\text{m}$. Therefore, the R function 'nn2' (RANN v.2.6.1, searchtype='radius') was used with a sufficiently large k ($k=41$). This analysis outputs an adjacency matrix that could then be used for graph construction with the igraph¹⁹⁵ package (v.1.3.4). For each ROI, the number of edges from two cell types was calculated and normalized by the sum of cells in each region. This was done for both spatial groups (proximal and distal) and then the difference between the normalized amounts of edges was computed. For testing the significance of an enrichment, an empirical null distribution was generated by random permutation of cell type labels ($m = 1000$) per slide and compared to the tested value. The P value was then computed with the formula $P=(b+1)/(m+1)$, taking tissue composition and spatial identifies into account. B is the number of permutations where a higher number of edges between cell types was observed and m is the total amount of permutations¹⁹⁶. This was computed for each slide and cell-cell pair to calculate a score representing co-localization or avoidance, depending on the sign. Visualization was done using an adaptation from⁸³.

8 Results

8.1 Overview of sphere-seq and the sorting approach

The sphere-seq approach is a combination of well-established methods: flow cytometry, single-cell hashing, and scRNA-seq. The method is based on previously developed methodologies, Paired-cell sequencing⁶⁵, PIC-seq⁶⁴, and Clump sequencing⁶⁶, that sequence cellular communities of 2-10 cells, requiring computational deconvolution procedures to allocate transcripts to different cells. We wanted to extend these approaches to single cell resolution and larger cellular communities. The established experimental workflow starts with partial dissociation of tissues into cellular communities, so-called spheres. Spheres are then sorted into wells, one sphere per well, of a 96-well plate using a large fragment biosorter. After sorting, individual spheres are dissociated into single cells within their respective wells, followed by labeling of cells with sphere-specific lipid-tagged BCs²¹. After labeling, cells are pooled and used for scRNA-seq using the BD Rhapsody or 10X Genomics platform. The sphere label allows us to allocate a single cell to its sphere and thereby its cellular neighborhood of origin. The outline of the sphere-seq approach is illustrated in fig. 13. We applied it to a murine model of liver metastasis, murine healthy liver, murine healthy spleen, CRC organoids (human and mouse), and Crohn's disease patient biopsies.

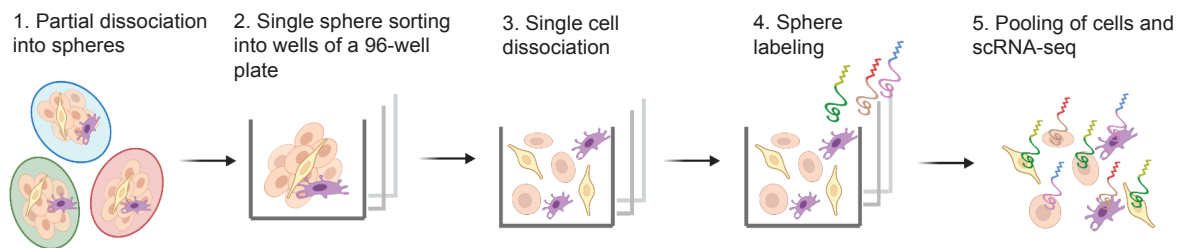


Figure 13: Schematic drawing of the sphere-seq experimental workflow. Created with BioRender.com.

Sorting with the large fragment biosorter was optimized to sort single spheres into 96-well plates. Using a linear model of standard sized beads TOF and extinction measurements could be calculated for drawing a sorting gate for sphere sizes of interest, in case of the murine liver between 200 and 450 μm (Fig. 14A, B). The accuracy of the sorting approach was assessed by sorting, plate imaging and counting of fragments per single wells of 9 plates. 92 % of wells had one fragment, 7.5 % no fragment and 0.5 % had two fragments (Fig. 14 C, D). We can therefore conclude that the sorting process is sufficiently accurate. The wells with zero fragments will lead to cell loss, but they would not contaminate our dataset. The two wells we identified with two fragments each, could be either due to a sorting error, or the fragment split into two due to the sorting pressure while sorting. Because we applied an index sorting approach, TOF measurements of each sorted event was recorded and could, therefore, be used to estimate the size of each individual sphere using the linear model of standard size beads. We sorted spheres with different size distributions for different tissue types (Fig. 14E). The size-gating of the biosorter was pre-defined in a way so we could capture biologically relevant subunits which was 200-450 μm in case of the murine liver, 100-400 μm for the murine spleen and CRC organoids had sizes between 100-300 μm .

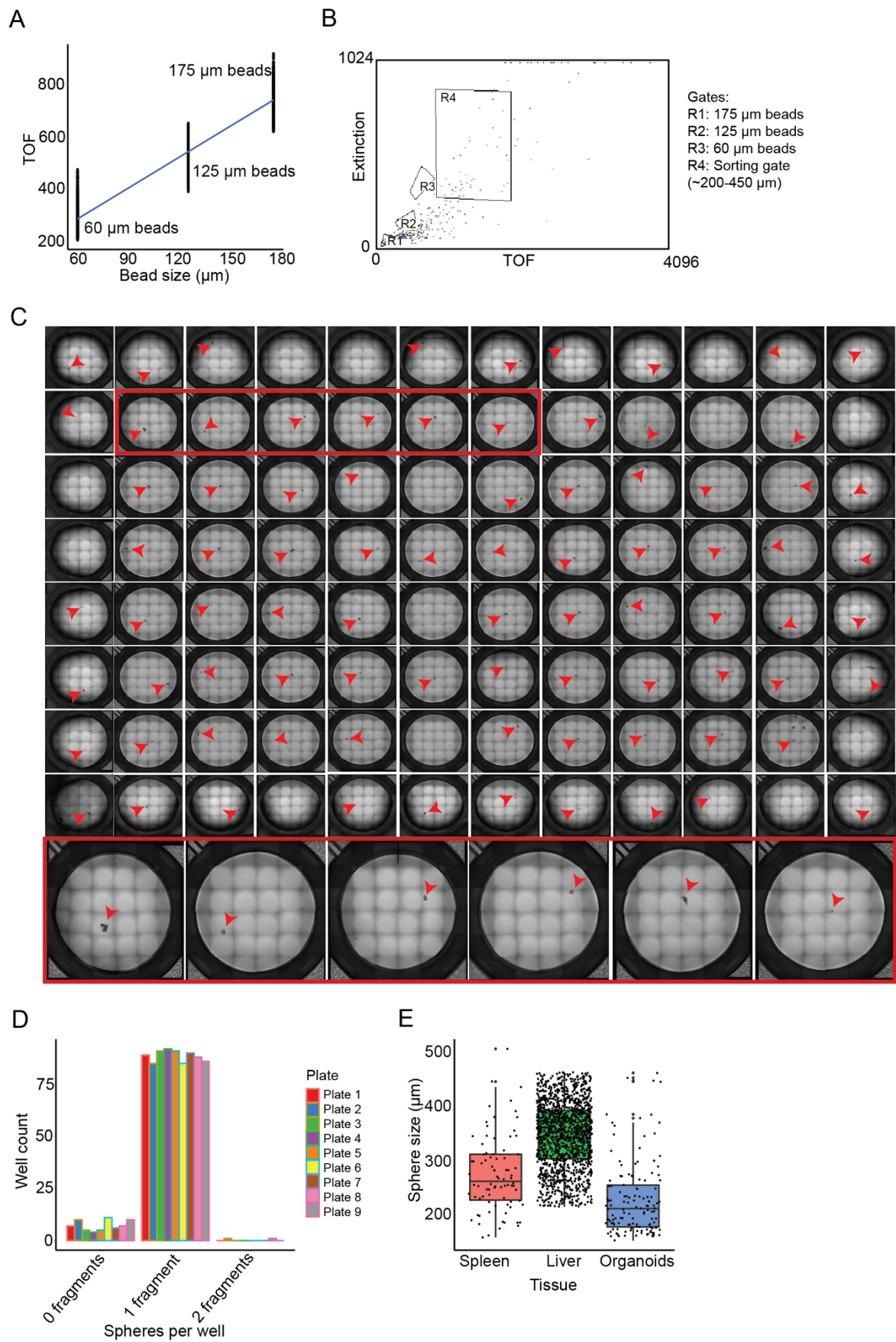


Figure 14: Overview of the sphere sorting approach. A) Linear standard curve of standard sized beads [60 μm ($n = 982$), 125 μm ($n = 736$), 175 μm ($n = 923$)] to calculate sphere sizes. Black dots are measurements of individual

beads; TOF = Time of flight. **B)** Image showing the gating from the large fragment sorter. R1-R3 represent standard sized beads and R4 shows the sorting gate of spheres between 200 and 450 μm . **C)** Brightfield microscopy image of 96 wells with single sorted sphere. Red arrows indicate spheres and red squares indicate zoomed-in wells. **D)** Barplot showing the number of spheres within wells ($n=864$ wells across 9 plates). **E)** Sphere size distribution of sorted spheres from different tissues [$n = 82$ (murine spleen), 1468 (murine liver), 138 (CRC organoids) spheres].

The theoretical number of cells per sphere could be quantified using the estimated amounts of cells per gram mouse liver which was estimated to be around 135 million hepatocytes¹⁸⁷. Due to the fact that 20% of the liver mass are hepatocytes and 20% NPCs¹¹¹, we added 27 million NPCs, which results in 162 million cells per gram liver. The density of the liver was estimated to be around 1.07 g/mL based on volumetric computed tomography (VCT)¹⁸⁶. Therefore, the volume of 1g liver could be calculated to be 1.07 cm^3 and the number of cells per cm^3 was estimated to be 151,401,869 million cells. The volume of different sphere sizes was calculated based on their diameter, calculated from the linear model of standard-sized beads, using the formula for spherical volume ($\frac{4}{3} * \pi * r^3 * 1.07 \text{ g/mL}$, r is the radius, 1.07 g/mL is the density) with the assumption that the sphere is entirely round. Even though spheres were not entirely round and had different shapes (Fig. 15A), we used this assumption to estimate the theoretical number of liver cells per sphere size (Fig. 15B). The true cell number of cells per liver sphere of different sizes was determined by integrating estimated size from biosorter and sequenced single-cell transcriptomes per sphere (Fig. 15C). Cell numbers were quite variable between spheres, possibly due to irregular shapes and sample processing. However, the mean values of sequenced cells were increasing with sphere size which showed some kind of robustness of the sphere-seq approach because the larger the sphere the more cells should be there for sequencing. Approximately 0.4-1.2 % of cells from theoretical amounts of cells got sequenced (Fig. 15D). The low amounts of recovered cells per sphere make sphere-seq sparse. The reasons for the low percentage could be cellular loss during the experimental procedure, most likely due to washing steps. Also, the MULTI-seq classification seems to be less efficient in sphere-seq. There was some variation between samples, but approximately 50 % of cells (accounting for around 3000-5000 cells in most cases) could be assigned to a single MULTI-seq/sphere BC (= positive cells), the rest were negative or doublets (Fig. 15E). We run an experiment with murine splenocytes that were split into 5 samples, labeled with different MUTLI-seq BC. This revealed an efficiency of around 60 % positives, which is 10 % more efficient than sphere-seq (Fig. 15F). A balance in terms of concentration and washing steps had to be found to archive sufficient labeling while preventing background labeling that impacts the MULTI-seq classification algorithm. A too small amount leads to negative cells, but too much leads to doublets. In the beginning phase of sphere-seq development we used 100-200 nM per sphere and only one washing step which led to massive amounts of doublets impacting the MULTI-seq algorithm to find positively labeled cells. We then reduced the concentration to 25-50 nM and 2-3 washes which led to more efficient classifications (data not shown). Another factor that influences the labeling is connective tissue that holds together the cells within a sphere. It usually gets filtered out before MULTI-seq labeling, however in sphere-seq this could impact the workflow negatively, unnecessary pipetting steps should be reduced to prevent cell loss. Therefore, connective tissue could hinder proper lipid anchor integration all around a cell's surface. This could also be a possible explanation of the decreasing cell coverage with larger sphere sizes (Fig. 15D), because larger spheres harbor larger amounts of connective tissue. For downstream analysis of sphere-seq, we only consider spheres with at least 5 cells which in our experience is a sufficient cutoff to analyze sphere-seq data. This is an arbitrary cut-off, as are thresholds for sphere size - the influence and potential biases introduced by both of these parameters will be discussed in

detail within a later section. All in all, these quality metrics demonstrate that our processing and sorting strategy is a fast and precise way to isolate spatial niches from complex tissues.

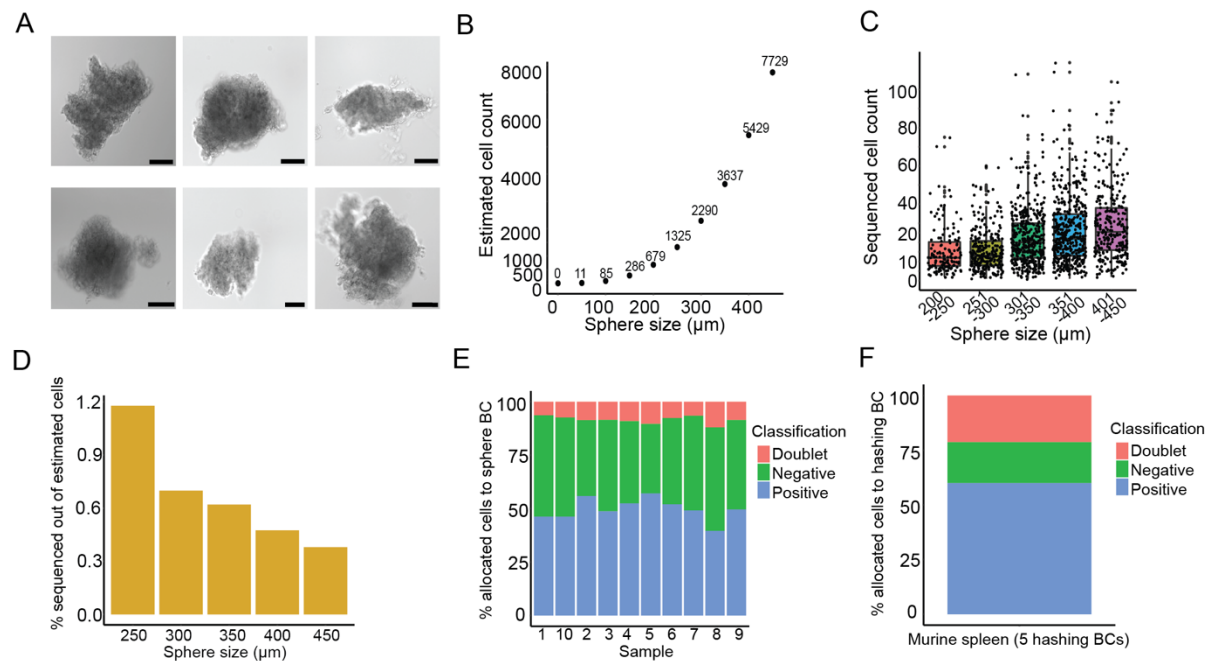


Figure 15: Comparison of sequenced and theoretically estimated cell counts per sphere of murine liver samples. **A)** Representative brightfield microscopy images of liver spheres. Scale bar of 100 μm is displayed in the right bottom corners. **B)** Scatterplot showing the theoretically calculated cell counts for different sphere sizes. **C)** Distribution of sequenced cell counts from different spheres grouped within different size. Dots represent individual spheres ($n = 200\text{-}250: 144, 251\text{-}300: 266, 301\text{-}350: 410, 351\text{-}400: 450, 401\text{-}450: 272$ spheres). The middle line represents the median; upper and lower line are the first and third quartile (25th and 75th percentiles); the upper and lower whiskers extend to $1.5 \times$ the distance between first and third quartile; dots outside of the whiskers are considered outliers. **D)** Barplot comparing the percentage of sequenced cells from the theoretical estimate of cell counts ($n = 200\text{-}250: 144, 251\text{-}300: 266, 301\text{-}350: 410, 351\text{-}400: 450, 401\text{-}450: 272$ spheres). **E)** Grouped barplot showing the proportions of positive (blue), negative (green), and doublets (red) allocated to sphere BCs after MULTI-seq classification ($n = S1: 3257$ positive, 446 doublet, 3326 negative; $S2: 4809$ positive, 738 doublet, 3048 negative; $S3: 3572$ positive, 621 doublet, 3119 negative; $S4: 4561$ positive, 790 doublet, 3338 negative; $S5: 3594$ positive, 652 doublet, 2037 negative; $S6: 4162$ positive, 610 doublet, 3220 negative; $S7: 2765$ positive, 367 doublet, 2485 negative; $S8: 2721$ positive, 827 doublet, 3315 negative; $S9: 5167$ positive, 887 doublet, 4353 negative; $S10: 3166$ positive, 504 doublet, 3169 negative cells). **F)** Grouped barplot showing the proportions of positive (blue), negative (green) and doublets (red) allocated to hashing BCs after MULTI-seq classification from a murine spleen scRNA-seq experiment combining 5 individually labeled samples (Positives: BC8: 795, BC46: 795, BC54: 780, BC62: 346, BC84: 2409; doublets: 1818; negatives: 1592 cells).

8.2 Comparison of sphere-seq to conventional scRNA-seq

We applied sphere-seq to a total of 10 murine livers. One from a healthy mouse and nine mice were injected with CRC organoids over an intrasplenic injection following splenic resection. After 2 weeks livers were harvested and we found that three of them had visible metastases. All samples were merged and single cells were annotated using marker genes from the recently published liver cell atlas⁹⁸ (Fig. 16A, B). Single cells could be allocated to their sphere of origin and after removing spheres with less than 5 cells, we could analyze different cell type proportions per sphere (Fig. 16C).

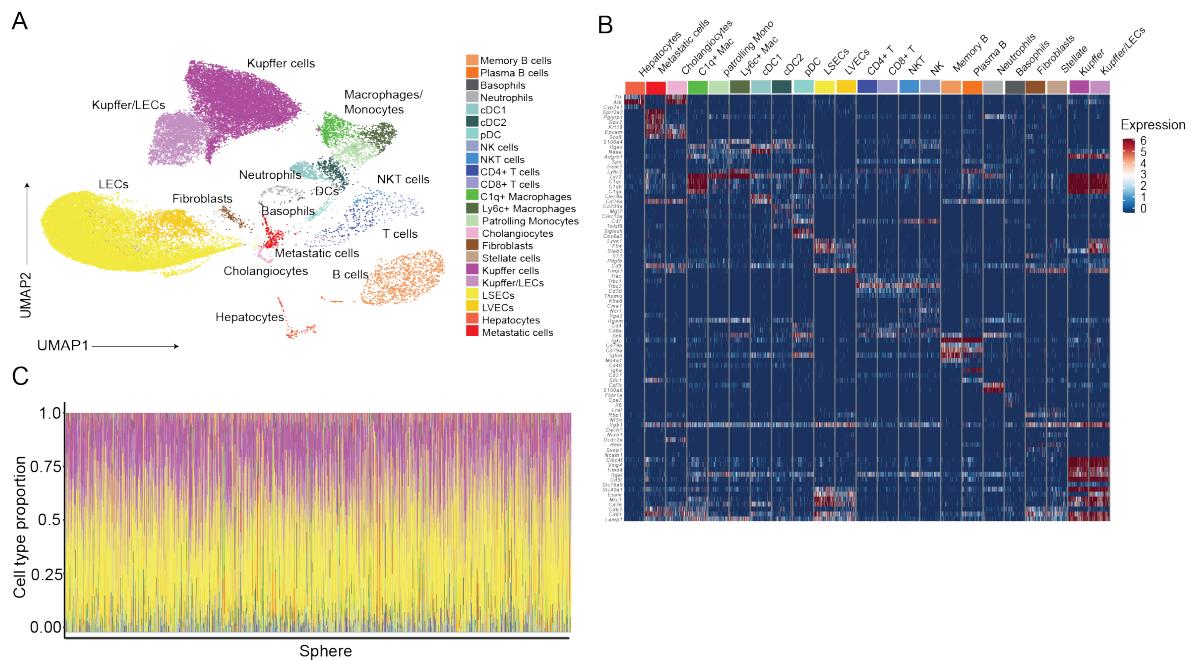


Figure 16: Cell type annotation of sphere-seq experiments from murine livers. **A)** Uniform manifold approximation and projection (UMAP) from all liver sphere-seq samples combined ($n = 10$ samples: 9 injected with CRC organoids, 1 healthy). Cells are colored based on their cell type. **B)** Heatmap of cell type marker gene expression of annotated cells ($n = 10$ samples: 9 injected with CRC organoids, 1 healthy). **C)** Barplot showing different proportions of cell types per sphere; only spheres with at least 5 cells are included. ($n = 1568$ spheres across 10 samples: 9 injected with CRC organoids, 1 healthy).

To assess the quality of sphere-seq compared to conventional scRNA-seq we conducted two conventional scRNA-seq experiments of CRC injected murine livers using the same *ex vivo* dissociation protocol like for sphere-seq. Cells were annotated using marker genes from the liver cell atlas⁹⁸ (Fig. 17A). The ratio of genes mapped to mitochondrial genes compared to all other genes is an indicator of cell quality. Low-quality cells have high ratios because their cell membrane becomes leaky leading to cytoplasmic mRNAs diffusing out of the cells, while mitochondrial RNAs (mtRNAs) will be contained within the mitochondria that are too large to diffuse out of the cells broken membrane¹⁹⁷. Therefore, in low quality cells, a higher amount of mtRNAs is being captured compared to cytoplasmic RNAs. The comparison of the median values of the percentages of mitochondrial to cytoplasmic genes between sphere-seq and conventional scRNA-seq revealed that sphere-seq might even be better in terms of cellular quality. Additionally, the number of UMI counts and genes are slightly better in sphere-seq (Fig. 17B). However, both protocols show similar values for these quality measurements and it has to be considered that there is much variability in cellular quality between scRNA-seq experiments in general. We were also wondering, if during the random sphere-seq sorting process, we introduce any kind of bias towards one cell type or another. This could arise because attachments of some cell types might be stronger or weaker and cell types with lower connectivity might be lost during partial dissociation. Therefore, we compared cell types of sphere-seq and conventional scRNA-seq, which were comparable across protocols (Fig. 17C).

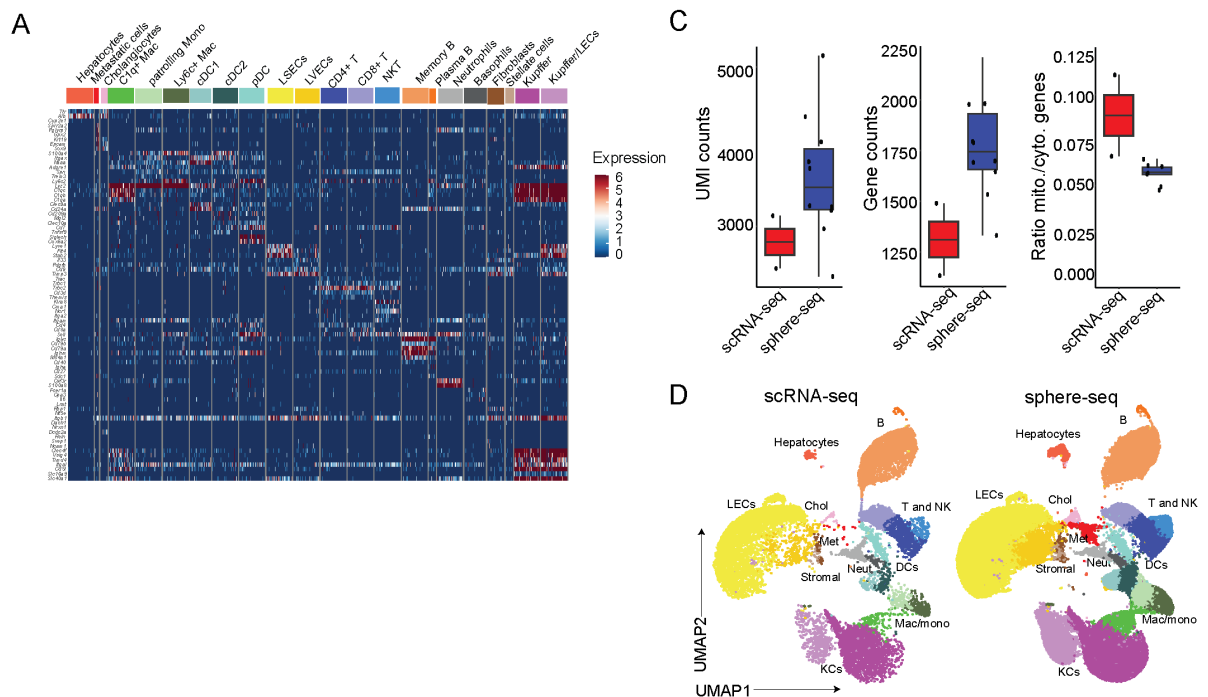


Figure 17: Comparison of sphere-seq to conventional scRNA-seq of murine metastatic livers. A) Heatmap of cell type marker gene expression of annotated cells from conventional scRNA-seq ($n = 2$ sampled injected with CRC organoids). **B)** Boxplots comparing UMI counts, gene counts and ratios of mitochondrial to cytoplasmic genes from scRNA-seq and sphere-seq. Dots represent samples (scRNA-seq: $n = 2$; Sphere-seq: $n = 9$). The middle line represents the median; upper and lower line are the first and third quartile (25th and 75th percentiles); the upper and lower whiskers extend to $1.5 \times$ the distance between first and third quartile; dots outside of the whiskers are considered outliers. **C)** Uniform manifold approximation and projection (UMAP) comparing cell types in scRNA-seq and sphere-seq (scRNA-seq: $n = 2$; Sphere-seq: $n = 9$).

In sum, we have found no evidence indicating that the sphere-seq procedure impacted the quality of single cell transcriptomes. Additionally, we could show that we did not introduce a specific bias towards certain cell types (at least not a bias that is not shared with conventional scRNA-seq). Of course, this could be slightly different for other tissues, than the liver, that are more prone to shear stress during sorting or that require faster single-cell processing procedures for good quality reads.

8.3 A mixing species experiment showed a high accuracy of the sphere-seq approach

To assess the accuracy of our sphere-seq approach in terms of cell allocation to the sphere of origin, we conducted a species-mixing experiment. To this end, we mixed GFP+ mouse and GFP- human CRC organoids. Using fluorescent index sorting we then sorted 144 wells with GFP+ and 144 wells with GFP- organoids. In this case, a sphere was an organoid. We then applied the sphere-seq procedure and analyzed the purity of human and mouse cells matching the index fluorescent sorting. A schematic drawing of the experimental approach is highlighted in fig. 18A. The normalized GFP signal of spheres classified as human and mouse based on GFP-expression is indicated in fig. 18B. To assess the accuracy of sphere-seq, we analyzed the proportion of incorrectly assigned cells per sphere (Fig. 18C). This revealed that 95 % of cells possessed species specific UMIs that matched the species of the index sorting – these cells were considered to be correctly assigned (Fig. 18D). This experiment showed that sphere-seq is a highly accurate approach because the vast majority of cells could be correctly

allocated. Miss-allocation might have been either due to unbound MULTI-seq BCs that introduce background staining of cells after pooling of individual wells, or loose single cells attached to organoids of the opposite species during sorting.

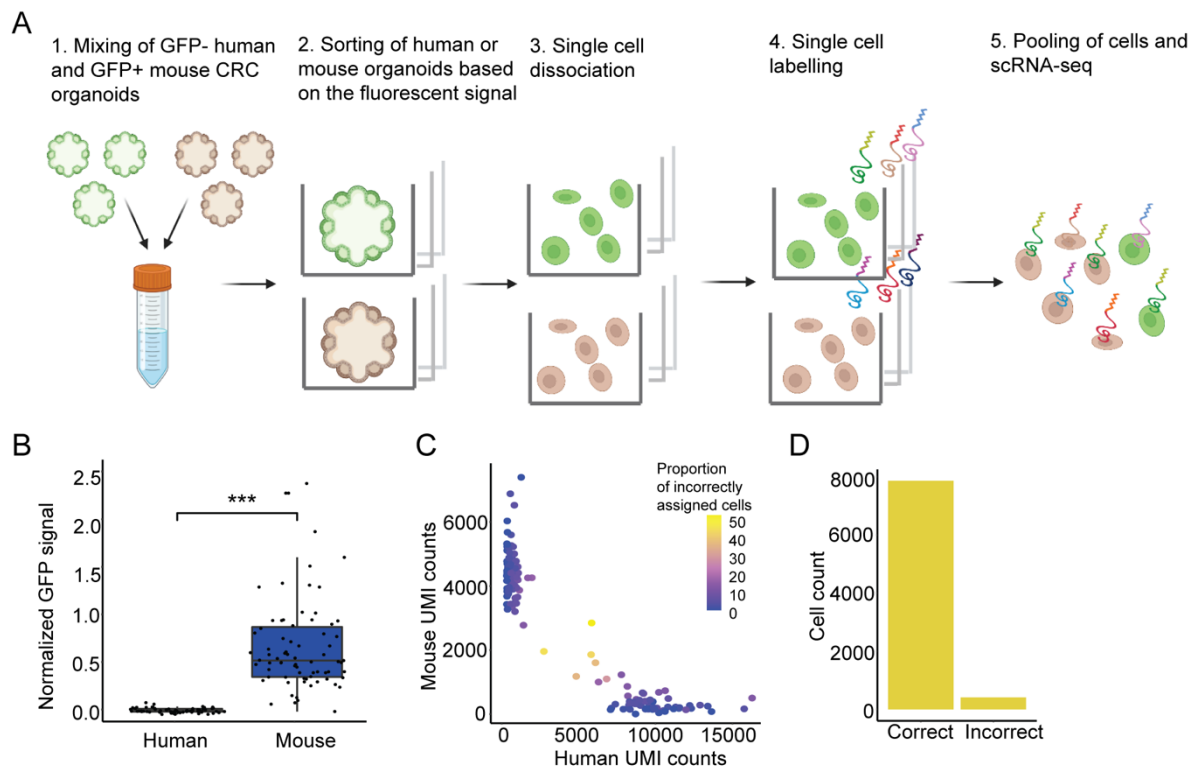


Figure 18: Species mixing experiment with CRC organoids to assess sphere-seq barcoding accuracy. A) Schematic illustration of the experimental workflow. Created with BioRender.com. **B)** Boxplots showing the normalized GFP signal of sorted GFP- human and GFP+ mouse organoids ($n = 288$ organoids, each species 144). Black dots show individual organoids. The middle line represents the median; upper and lower line are the first and third quartile (25th and 75th percentiles); the upper and lower whiskers extend to $1.5 \times$ the distance between first and third quartile; dots outside of the whiskers are considered outliers. **C)** Scatter plot highlighting spheres with incorrectly assigned cells. Dots represent individual spheres and their color shows the proportion of incorrectly assigned cells (mouse cells in human spheres or human cells in mouse spheres) ($n = 139$ spheres). **D)** The fraction of cells that are correctly and incorrectly assigned is shown in a barplot ($n = 139$ spheres, 7862 cells correct, 410 cells incorrect).

8.4 An imaging-based transcriptomics approach was used for the validation of sphere-seq

To validate findings from sphere-seq of the metastasis-bearing mouse liver, we wanted to use an imaging-based ST approach. We designed a 100-gene panel that included marker genes for different cell types and genes that showed spatially differentiated gene expression in sphere-seq. This panel was used for highly-multiplexed FISH (Molecular Cartography, MC) for 4 samples of CRC injected livers, from which two showed visible metastasis. DAPI staining of images could be used for cell segmentation. Segmented single cells could then be used for dimensionality reduction, clustering, and cell type annotation (Fig. 19A, B). Comparison of broad cell type annotation between MC and sphere-seq revealed that immune cells and LECs are well reflected in sphere-seq (Fig. 19C). However, we found a bias of sphere-seq towards losing hepatocytes, stellate cells, fibroblasts and metastatic cells. This was probably due to the *ex vivo* digestion protocol of livers in sphere-seq. A phenomenon, we also saw in conventional scRNA-seq datasets from livers that were digested in the same way and which

was previously discussed by Gulliams *et al.*⁹⁸. Some cell types (neutrophils, DCs, basophils, cholangiocytes) could not be resolved with MC because the gene panel was restricted to 100 genes and selected genes were not specific enough to annotate these cell types.

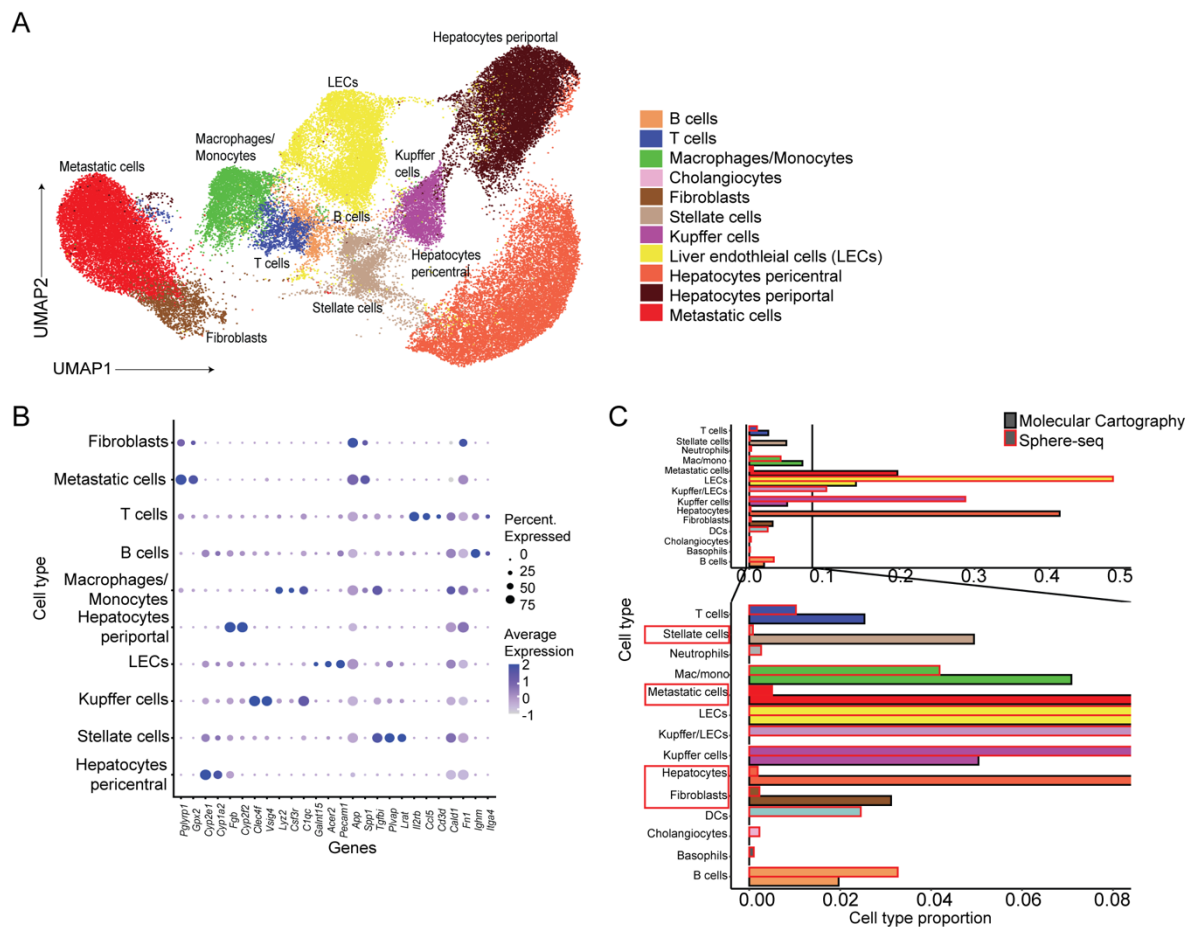


Figure 19: Annotation of Molecular Cartography imaging-based ST data and its comparison to sphere-seq. **A)** UMAP visualization of cells after cell segmentation ($n = 4$ samples). Cells are colored based on their cell type. **B)** Dotplot showing the cell type marker expression within annotated clusters. **C)** Zoomed-in barplot comparing cell type proportions between MC and sphere-seq (MC: $n = 4$ samples, Sphere-seq: $n = 9$ samples). Cell types highlighted in red boxes represent the largest differences between the datasets.

8.5 Sphere-seq could be used to identify previously uncharacterized zoned gene expression in LECs and KCs

8.5.1 Reconstruction of the sphere position along the lobule layer axis

To study zoned gene expression within the liver we wanted to first reconstruct the spatial origin of each sphere. Therefore, we applied a previously generated algorithm that uses a zoned gene expression signature in LECs^{65,115} to assign spheres to a specific position within the central-portal axis. For each sphere we calculated a zonation coordinate (ZC) based on pericentral and periportal landmark gene expression signatures (cLM and pLM) in LECs. Average expression (AE) across LECs from a sphere was calculated separately for cLM and pLM genes and used for the calculation of ZC by dividing the AE pLM genes by the sum of AE cLM and AE pLM genes. This generated a ZC between 0 and 1 where 0 is the most central and 1 is most portal. Using this approach, spheres could be placed within the central-portal axis (Fig. 20A). Spheres could be grouped into different lobule layers, the most central (L1-3)

and most portal (L8-10) were combined because there were not many spheres from the most central or portal layer. Furthermore, they could be grouped into pericentral (L1-5) and periportal (L6-10) zones (Fig. 20B). We could speculate that the sphere size might be too big to capture the most distant layers, the middle layers are more of a continuum in terms of zoned gene expression. To assess the accuracy of this reconstruction approach in our sphere-seq dataset, we analyzed zoned landmark gene expression in hepatocytes from spheres grouped into pericentral and periportal areas. The pericentral-specific landmark genes *Cyp2e1* and *Cyp1a2*, which are P450 enzymes involved in the oxidation of drugs¹⁰⁴, were highly expressed in hepatocytes of spheres from the pericentral group. The periportal genes *Cyp2f2*, and *Alb*, – of note, albumin production is a highly energetically demanding task and therefore needs a lot of ATP resources only present in portal areas¹⁰⁴ – were almost exclusively expressed in hepatocytes of periportal spheres (Fig. 20C). This showed that our assignment of spheres to specific anatomical compartments worked reliably and this information could be used for further analysis.

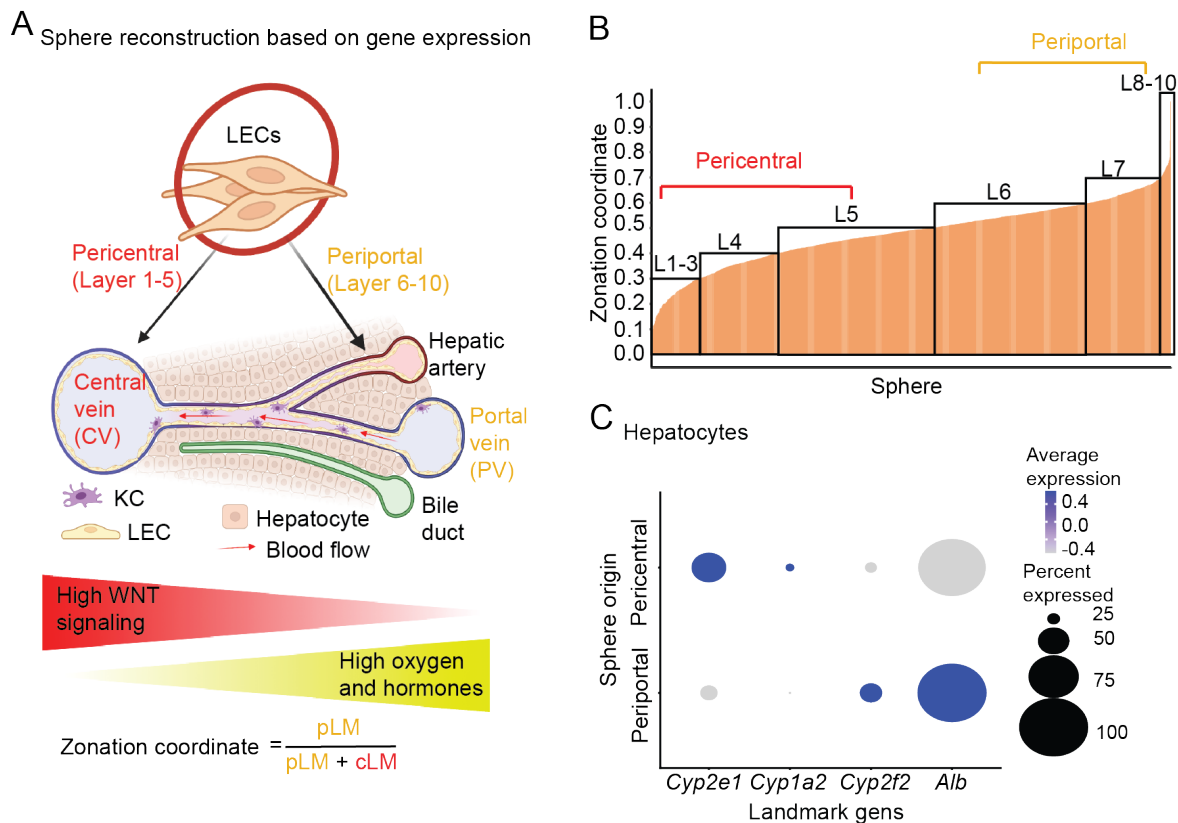


Figure 20: Allocation of spheres to different lobule layers based on a previously identifies zoned gene expression signature in LECs. A) Schematic illustration of the strategy to reconstruct sphere position based on gene expression in this case a zonation coordinate is calculated based on CV and PV landmark genes (cLM and pLM) in LECs. Created with BioRender.com. **B)** Barplot showing the distribution of zonation coordinates of spheres across different lobule layers ($n = L1-3: 143, L4: 243, L5: 452, L6: 478, L7: 207, L8-10: 29$ spheres across 10 samples). **C)** Dotplot of landmark gene expression in hepatocytes of grouped spheres into pericentral and periportal regions: pericentral landmark genes: *Cyp2e1* and *Cyp1a2*; periportal landmark genes: *Cyp2f2* and *Alb* ($n =$ pericentral: 46, periportal: 31 spheres across 10 samples).

8.5.2 Determination of spatial areas in Molecular Cartography images

To use MC data for validation of spatially distinct findings of sphere-seq, we first needed to group cells of MC images into periportal, pericentral or metastatic areas. Therefore, we used marker genes for periportal (*Cyp2f2*), pericentral (*Cyp2e1*), and metastatic (*Gpx2*) regions. We visualized these transcripts on MC images and manually drew areas surrounding cells expressing specific landmark genes (Fig. 21A). Information of previously annotated pericentral and periportal hepatocytes from single-cell segmented images could then be grouped based on manually assigned spatial areas (Fig. 21B). This showed the accuracy of this approach to determine spatial areas.

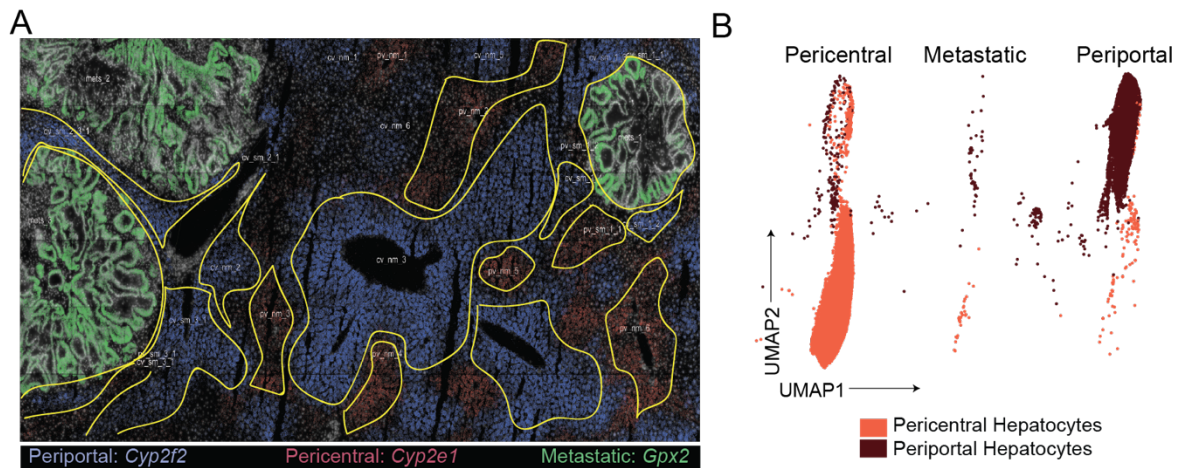


Figure 21: Determination of spatial areas in Molecular Cartography images. A) Representative image from MC of periportal (*Cyp2f2* in blue), pericentral (*Cyp2e1* in red) and metastatic (*Gpx2* in green) areas. Yellow lines show manually drawn areas to group cells into different spatial areas. B) UMAP visualization of pericentral and periportal hepatocytes from different spatial areas ($n = 4$ samples).

8.5.3 Sphere-seq identified previously uncharacterized zoned gene expression in LECs

After the reconstruction of spheres based on the zoned gene expression in LECs, we wanted to analyze zoned gene expression in LECs of genes that were not present in the signature that was used for reconstruction. We did a DGE analysis of LECs from spheres grouped into lobule layers and we could find many genes that were significantly differentially expressed along the central-portal axis (Fig. 22A, left). We included some of the zoned genes into our gene panel for MC. We then extracted LECs after cell segmentation and annotation and analyzed DEGs between periportal and pericentral areas. Thereby, we could validate our findings of zoned genes (Fig. 22A, right). For example, we found *Plpp1*, a phospholipid phosphatase, enriched pericentrally, matching the known pericentral zonation of lipogenesis¹⁰⁴. *Galnt15*, a gene involved in O-linked oligosaccharide biosynthesis, was periportal zoned, which makes sense because gluconeogenesis is periportal zoned¹⁰⁴ (Fig. 22A-D). Other examples were the pericentrally zoned *Lhx6* (Fig. 22A-D), which was previously identified to be centrally zoned in LECs¹⁹⁸, it is involved in the regulation of canonical WNT signaling which is centrally zoned^{104,199}. And *Cd36* (fatty acid translocase/cluster of differentiation 36) (Fig. 22A-D) which is involved in fatty acid transport and has an established periportal zonation in LECs²⁰⁰⁻²⁰², was also found to be periportal zoned in LECs in our dataset (Fig. 22B-D). Furthermore, we assessed the zonation of these

four genes in LECs from healthy liver, and found them to be significantly zoned as well (Fig. 22E), illustrating, that these genes are most likely involved in homeostatic processes. In conclusion, with the help of sphere-seq, we could identify previously uncharacterized zone-specific gene expression in LECs. Additionally, we could confirm the previously published pericentral and periportal expression of *Lhx6* and *Cd36*, respectively^{198,201,202}, which gave us confidence that our approach of assigning spheres to lobule layers was working reliably.

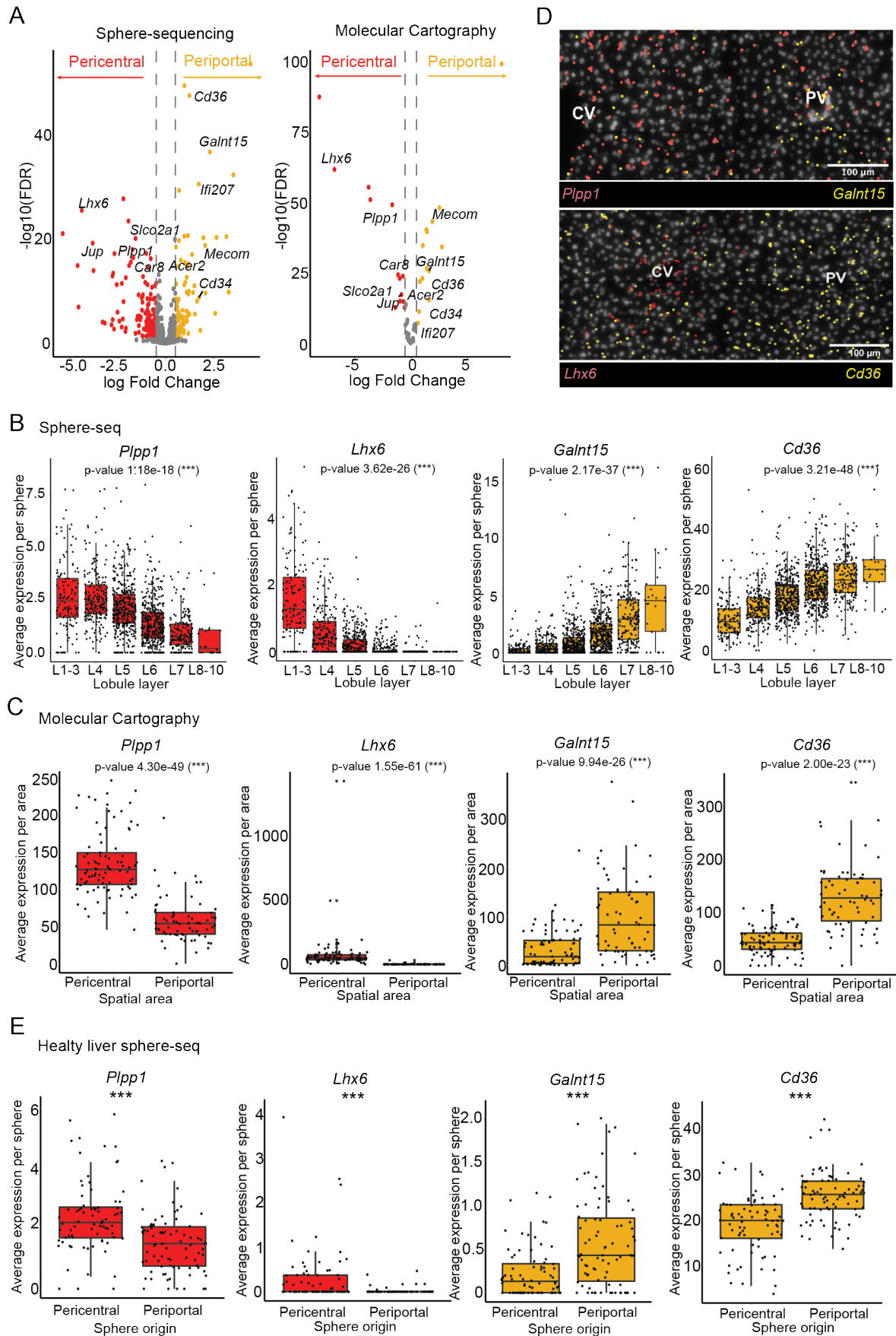


Figure 22: Analysis of zoned gene expression in LECs. A) Volcano plots of DEGs between spatially ordered spheres (left, sphere-seq, $n = 1384$ spheres across 9 samples) or spatially grouped areas (right, Molecular

Cartography, $n = 155$ areas across 4 samples) in pericentral (red) and periportal (yellow) zones. Colored dots show significantly upregulated DEGs in one or the other condition and labels indicate genes that were present also in the MC gene panel and could be validated. **B)** Boxplots of selected genes of spatially ordered spheres ($n = L1-L3: 137, L4: 214, L5: 402, L6: 409, L7: 196, L8-L10: 26$ spheres across 9 samples). Black dots represent individual spheres. **C)** Boxplots of selected genes of spatially grouped areas from MC data ($n =$ pericentral 89, periportal 66 areas across 4 samples). Black dots represent individual manually drawn areas. **D)** Representative MC images of selected zoned genes. DAPI stain (white) is overlaid with gene expression in colors. The scale bar is shown in the lower left corner. **E)** Boxplot of selected genes of spatially grouped spheres from healthy murine livers ($n =$ pericentral: 76, periportal: 83 spheres across 1 sample). A non-parametric Wilcoxon signed-rank test was applied ($*** < 0,001$, not significant > 0.05). Black dots represent individual spheres. For **A**, **B**, and **C** a negative binomial generalized log-linear model was applied for statistical analysis; p -values (Benjamin-Hochberg adjusted) < 0.05 were considered as significant ($*** < 0.001$, $** < 0.005$, $* < 0.05$). In **B**, **C** and **E** the middle line represents the median; upper and lower line are the first and third quartile (25th and 75th percentiles); the upper and lower whiskers extend to 1.5 * the distance between first and third quartile; dots outside of the whiskers are considered outliers.

8.5.4 Sphere-seq identified zoned gene expression in KCs indicating an involvement in leukocyte recruitment

Next, we wanted to use spheres, which were spatially-sorted based on LECs, to analyze zoned gene expression in other cell types, for example, KCs. We identified that *Vcam1* (Vascular Cell Adhesion Molecule-1) is periportally zoned in KCs which was confirmed independently with MC (Fig. 23A, B). VCAM1 function has been mostly characterized in endothelial cells where it facilitates binding of immune cells and can promote their extravasation²⁰³. Interestingly, *Vcam1* was not significantly zoned in KCs of healthy liver (Fig. 23C), suggesting its upregulation/zonation was a response to the perturbation of introducing metastases. In LECs it has been previously described that *Vcam1* expression is upregulated during metastasis²⁰⁴. Not much information about the function of *Vcam1* in KCs has been revealed yet, however, a study by Okada *et al.* showed that VCAM1 mediates crosstalk between KCs and lymphocytes, which in return leads to activation of KCs¹²⁴.

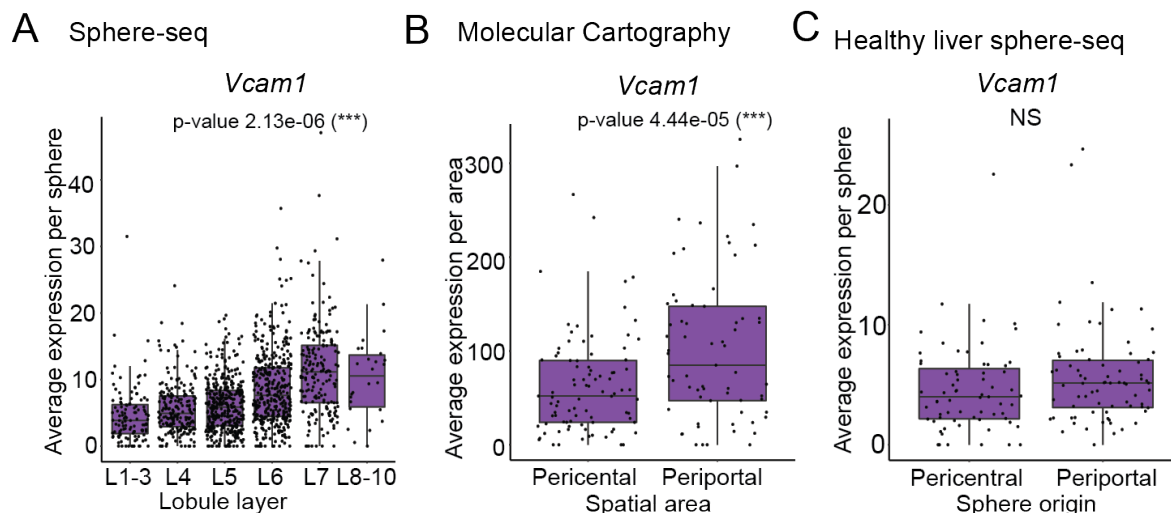


Figure 23: Analysis of zoned gene expression in KCs. Boxplots of *Vcam1* of **A)** spatially ordered spheres from sphere-seq data ($n = L1-L3: 115, L4: 206, L5: 379, L6: 387, L7: 182, L8-L10: 26$ spheres across 9 samples), **B)** spatial areas from MC data ($n =$ pericentral 89, periportal 66 areas across 4 samples) and **C)** spatially grouped spheres from sphere-seq data of untreated murine liver ($n =$ pericentral: 67, periportal: 74 spheres across 1 sample). The middle line represents the median; upper and lower line are the first and third quartile (25th and 75th percentiles); the upper and lower whiskers extend to 1.5 * the distance between first and third quartile; dots outside of the whiskers are considered outliers. For **A** and **B**, a negative binomial generalized log-linear model was applied for statistical analysis; p -values (Benjamin-Hochberg adjusted) < 0.05 were considered significant ($*** < 0.001$, $**$

< 0.005, * < 0.05). For **C** a non-parametric Wilcoxon signed-rank test was applied (*** < 0,001, not significant > 0.05). Black dots represent individual spheres (**A** and **C**) and individual manually drawn areas (**B**).

Therefore, we wanted to further investigate L-R interactions between KCs and lymphocytes (B and T cells) for both periportal and pericentral areas. We grouped spheres from pericentral (ZC = 0-05) and periportal (ZC = 0.51-1) areas and performed separate L-R interaction analyses using the CellPhonDB⁹² algorithm. This algorithm uses gene expression and cell type cluster information to predict L-R interactions between two cell types by comparing their expression to a database of know interactions. We could find many interactions that were spatially restricted to or enriched in pericentral or periportal areas (Fig. 24A, B). We could verify the zoned expression of some ligands and receptors by MC (Fig. 24C). The analysis unveiled a periportally enriched L-R interaction between KC-expressing ligands *Vcam1* and various integrin complex receptors on B and T cells ($\alpha 9\beta 1$, $\alpha 4\beta 7$, and $\alpha 4\beta 1$) (Fig. 24A-C). Integrins on T cells are adhesion molecules that help T cells to adhere to the vessel wall^{205,206}. The upregulation of these interactions periportally indicated that KCs might promote lymphocyte recruitment within this zone. This matched with the periportally enriched *Ccl3/Ccr5* interaction of KCs and T cells because CCR5 is reportedly involved in T cell recruitment during acute liver inflammation^{205,206}.

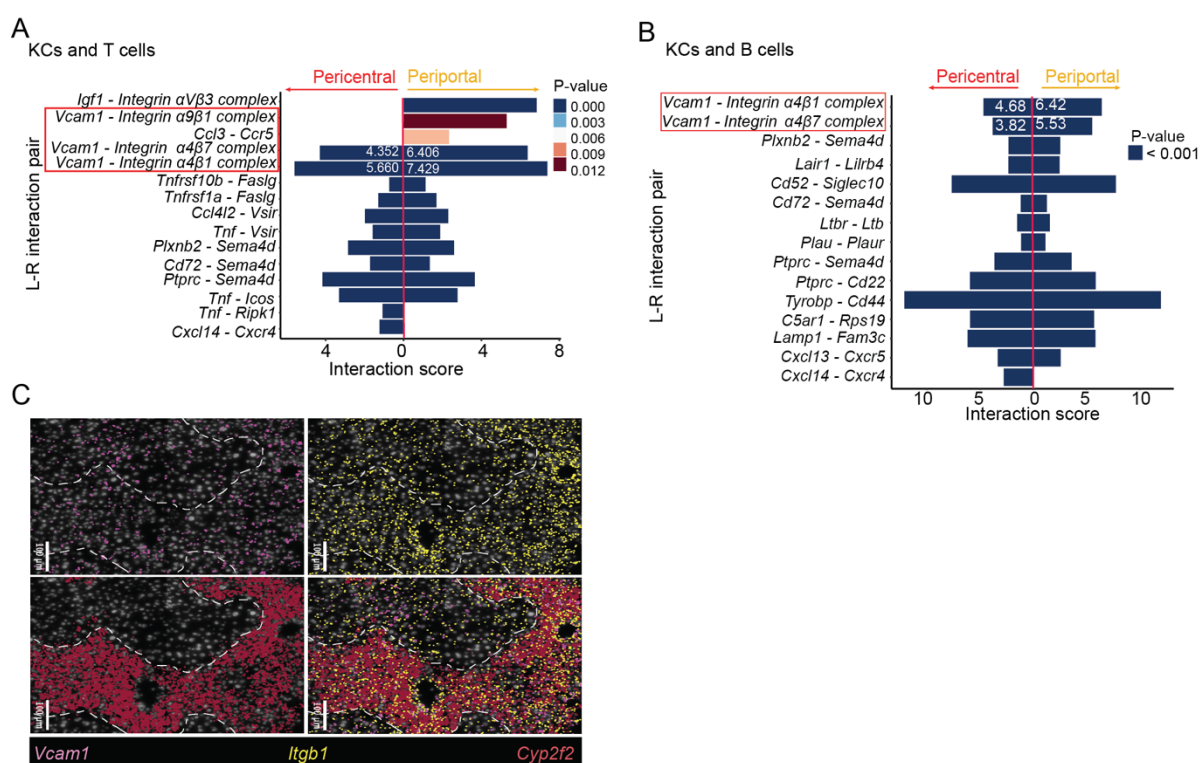


Figure 24: L-R interactions analysis comparing pericentral and periportal areas. A) Predicted L-R interactions of KCs and T cells comparing interaction scores from pericentral and periportal grouped spheres (n = 9 samples). Interaction scores were generated with CellPhoneDB with a permutation test to calculate p-values. Interactions explained in the text are highlighted with a red square. **B)** Same as for A, but KCs and B cells. **C)** Representative MC image of indicated L-R genes. DAPI stain (white) is overlaid by gene expression in colors. Scale bars are shown in the bottom left corner.

All in all, these results suggested the importance of *Vcam1* in periportal KCs to archive increased lymphocyte recruitment. Through the periportal blood flow, KCs and lymphocytes

encounter each other periportally for the first time and therefore it is important to have strong recruitment there.

8.6 Spheres could be grouped into areas proximal and distal to metastatic sites to uncover spatially restricted differences

We next wanted to analyze the spatial impact of metastatic cells and therefore we grouped spheres into spheres containing metastatic cells, which we called 'proximal', and spheres lacking metastatic cells, which we called 'distal' to metastases (Fig. 25A, B). For this analysis we only considered three CRC injected mouse liver samples. They showed visible metastasis within the liver tissue and we could analyze at least 20 metastatic cells within each of these three samples (Fig. 25C, D). There were more spheres from distal areas (Fig. 25E) which could be due to a higher abundance of non-metastatic tissue within the liver and/or the lower efficiency to isolate and capture metastatic cells as we have also seen in conventional scRNA-seq experiment. We also inspected a high variability in the number of metastatic cells per sphere (Fig. 25F), which could be due to the location of spheres either from within a metastasis or from the border region or due to the technical constraints to isolate metastatic cells. To account for that, we grouped all spheres with metastatic cells to analyze spatial differences in cell type abundance and L-R interactions to distal grouped spheres.

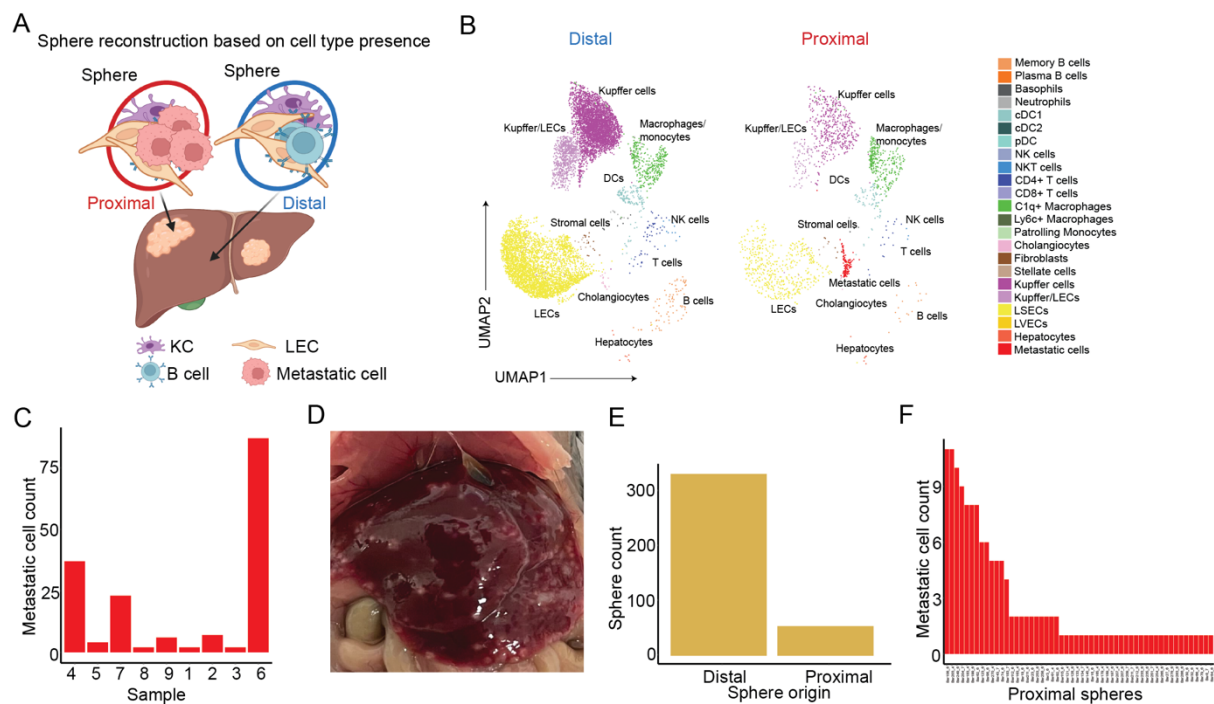


Figure 25: Grouping of spheres into proximal and distal areas to micro-metastatic sites based on the presence/absence of metastatic cells. **A)** Schematic drawing of the reconstruction approach. Created with BioRender.com. **B)** UMAP of cells between distal (left) and proximal (right) spheres ($n = 3$ samples). Cells are colored by their annotated cell types. **C)** Barplot comparing the number of metastatic cells in different samples ($n = 9$ samples). **D)** Representative photographic image of a mouse liver with visible metastasis. **E)** Barplot comparing the sphere counts between distal and proximal areas ($n =$ distal: 332, proximal 54 spheres across 3 samples). **F)** Barplot comparing metastatic cell counts in proximal spheres ($n = 54$ spheres across 3 samples).

After grouping spheres into proximal and distal areas, we first wanted to determine differences in cell type abundances. Differential abundance analysis of broadly annotated cell clusters revealed a significant enrichment of macrophages/monocytes in proximal areas, while in distal

areas KCs and LECs were significantly enriched (Fig. 26A, B). We again used MC to validate these findings. To this end, we pooled CV and PV areas as defined in Fig. 21 as distal, and metastatic areas as proximal spatial regions. We saw similar trends in differential cell-type abundances in sphere-seq and MC, highlighting that sphere-seq conscientiously recapitulated real cell-type proportions (Fig. 26A, C, D). Monocytes/macrophages could be annotated into three subtypes, *Ly6c*⁺ macrophages expressing *Ly6c2*, *C1q*⁺ macrophages expressing *C1qc*, and patrolling monocytes expressing *Spn* (Fig. 26E). We saw a trend of enrichment *C1q*⁺ macrophages in proximal metastatic areas (Fig. 26F), a subtype has been previously identified to have anti-inflammatory characteristics¹²⁹. This could be confirmed by MC where we did DGE analysis between macrophages/monocytes from proximal and distal areas and we could uncover a significant enrichment of complement genes *C1qc* and *C1qb* in proximal areas (Fig. 26G). *Tgfb1* (TGF beta-induced), which was also enriched in macrophages/monocytes from proximal areas, is known to be involved in immunosuppression^{207–210}, matching our hypothesis that there is an enrichment of anti-inflammatory macrophages in proximal sites.

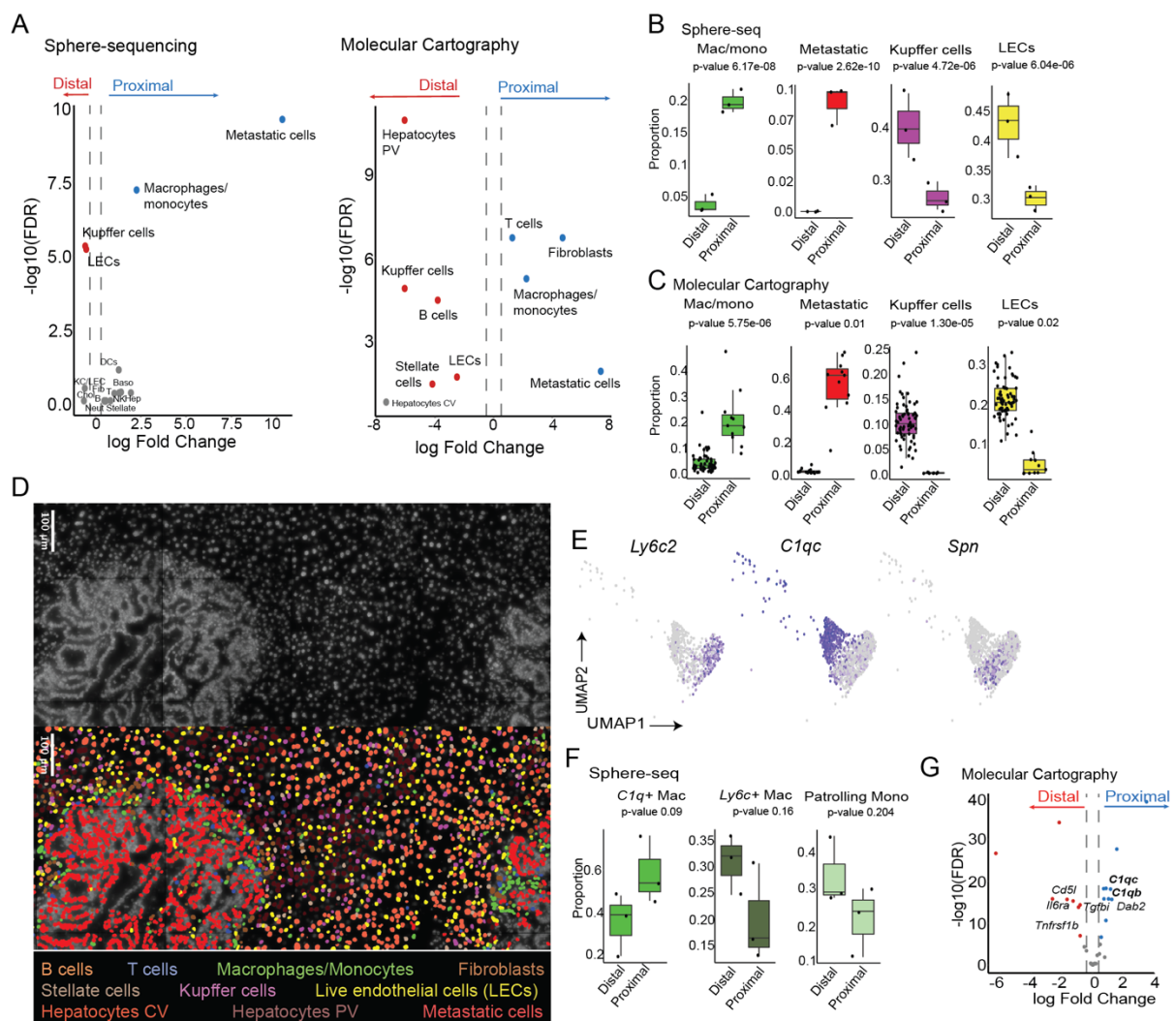
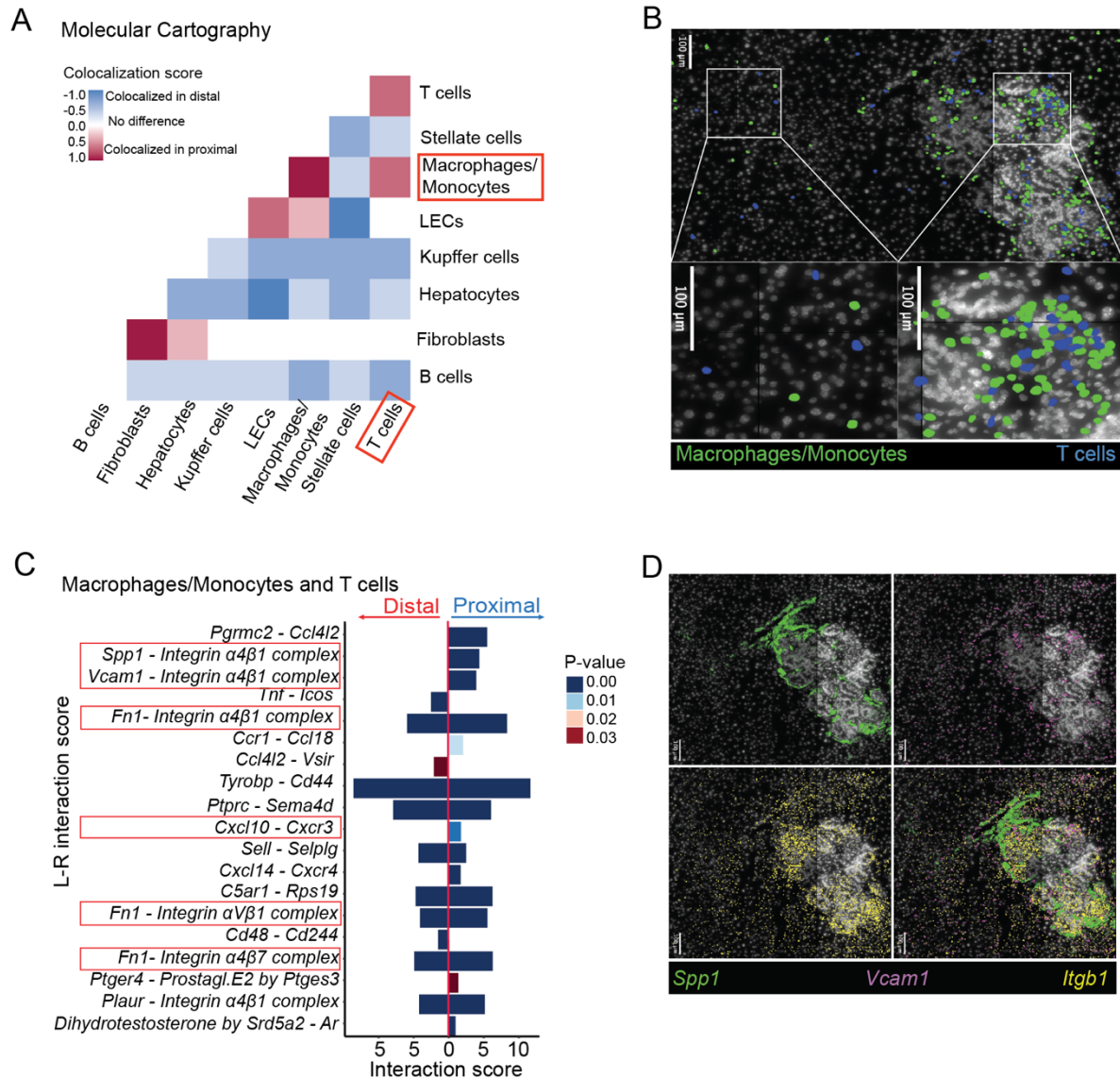


Figure 26: Cell type abundance analysis of metastatic-proximal compared to -distal grouped spheres. A) Volcano plot showing the results of a differential abundance analysis (DAA) of cell types between metastatic areas from sphere-seq (left, $n = 3$ samples) and Molecular Cartography (right, $n =$ distal: 71, proximal: 11 areas across 2 samples). Colored dots represent significant cell type enrichment. **B)** Boxplots of selected cell types comparing distal and proximal groups from sphere-seq data ($n = 3$ samples). Each dot represents a sample. **C)** Boxplots of selected cell types comparing distal and proximal areas from MC data ($n =$ distal: 71, proximal: 11 areas across 2

samples). Dots represent spatial areas. **D)** Representative image of MC data of proximal and distal areas. DAPI stain (white) is overlaid by cell type annotation in colors. Scale bar is shown in the top left corner. **E)** Feature UMAP plots showing marker gene expression of different macrophage/monocyte subtypes ($n = 3$ samples). **F)** Boxplots of macrophage/monocyte subtypes comparing distal and proximal groups from sphere-seq data ($n = 3$ samples). Each dot represents a sample. **G)** Volcano plot showing DEGs from distal and proximal areas of MC data. Colored dots represent significant enrichment. Bold gene names highlight complement genes indicative of $C1q+$ macrophages. For **A**, **B**, **C**, **F**, and **G**, a negative binomial generalized log-linear model was applied for statistical testing; p -values (Benjamin-Hochberg adjusted) < 0.05 were considered significant. For **B**, **C** and **F**, the middle line represents the median; upper and lower line are the first and third quartile (25th and 75th percentiles); the upper and lower whiskers extend to $1.5 * \text{the distance between first and third quartile}$; dots outside of the whiskers are considered outliers.

$C1q+$ macrophages have further been described to be involved in T cell exhaustion¹²⁷. Therefore, we wanted to take a closer look into the cellular crosstalk of macrophages/monocytes and T cells by analyzing cell-cell co-localizations in MC imaging data and L-R interactions in sphere-seq data between these two cell types. We decided to only consider broadly annotated cell types, i.e., macrophages/monocytes and T cells because i) the limited number of genes within the MC panel did not allow us to properly divide clusters into sub-cell types and ii) L-R interaction analysis on the sphere-seq data likely has a higher statistical power when using larger clusters of cells. The co-localization analysis revealed significantly more co-localizations between macrophages/monocytes and T cells in proximal compared to distal areas (Fig. 27A, B). To identify potential interactions driving this phenomenon we performed L-R interaction analysis between these two cell types comparing proximal and distal areas. On the proximal side, we found an enrichment of Secreted Phosphoprotein 1 (*Spp1*), Fibronectin-1 (*), and *Vcam1* ligands derived from macrophages/monocytes interacting with various integrin complexes on T cells (Fig. 27C). Enrichment within proximal areas of some of these ligands and receptors could be validated by MC (Fig. 27D). The interaction involving *Vcam1* is most likely involved in adhesive reactions of macrophages/monocytes and T cells. SPP1 and FN1 ligands derived from macrophages/monocytes, here found to be enriched in proximal areas, were reported to be involved in various cancers and correlate with a poor prognosis^{211–213}. FN1 reportedly correlates with anti-inflammatory macrophage infiltration²¹² and SPP1 was reported to be involved in T cell suppression within the TME²¹⁴. Interactions of both of these ligands with integrins were found to be increased in CRC and might have an impact on tumor progression²¹⁵. Additionally, we found the interaction *Cxcl10|Cxcr3* to be proximally enriched, which is well-known for mediating the recruitment of effector T cells²¹⁶. In primary CRC this interaction between $C1q+$ macrophages and other lymphocytes is increased²¹⁵, matching our results. This shows that with sphere-seq we can dissect small spatial niches within complex tissues and identify locally enriched L-R interactions that might be important in metastatic progression or suppression.*



8.7 Application of sphere-seq to other tissues and species

After establishing the sphere-seq protocol we wanted to test its applicability in other tissues. Therefore, we generated proof-of-concept datasets from the healthy murine spleen and Crohn's disease biopsy samples.

8.7.1 Application to healthy murine spleen

We first applied the sphere-seq protocol to healthy murine spleen where we wanted to dissect different immunological zones which can be identified by an enrichment of distinct immune

cell types. The T cell zone is enriched for T cell subsets, the B follicle for B cells and the marginal zone (MZ) contains other immune cells like macrophages^{217,218}. Sphere-seq was performed gating for sphere sizes between 200 and 400 μm (Fig. 28A) and cells were annotated based on marker genes from Medaglia *et al.*⁵⁹ (Fig. 28C-D). Cells were then allocated to their sphere of origin to analyze different cell type and immunological zone proportions (Fig. 28E). However, there were almost no spheres that had their origin in a specific zone only, spheres had mixtures of cells all across the three zones. We speculate that reducing the sphere size to 100-200 μm could increase the spatial resolution. Another approach could be the incubation of spheres with surface antibodies against T or B cell markers respectively before sorting. This could help to enrich for spheres from T or B cell areas exclusively.

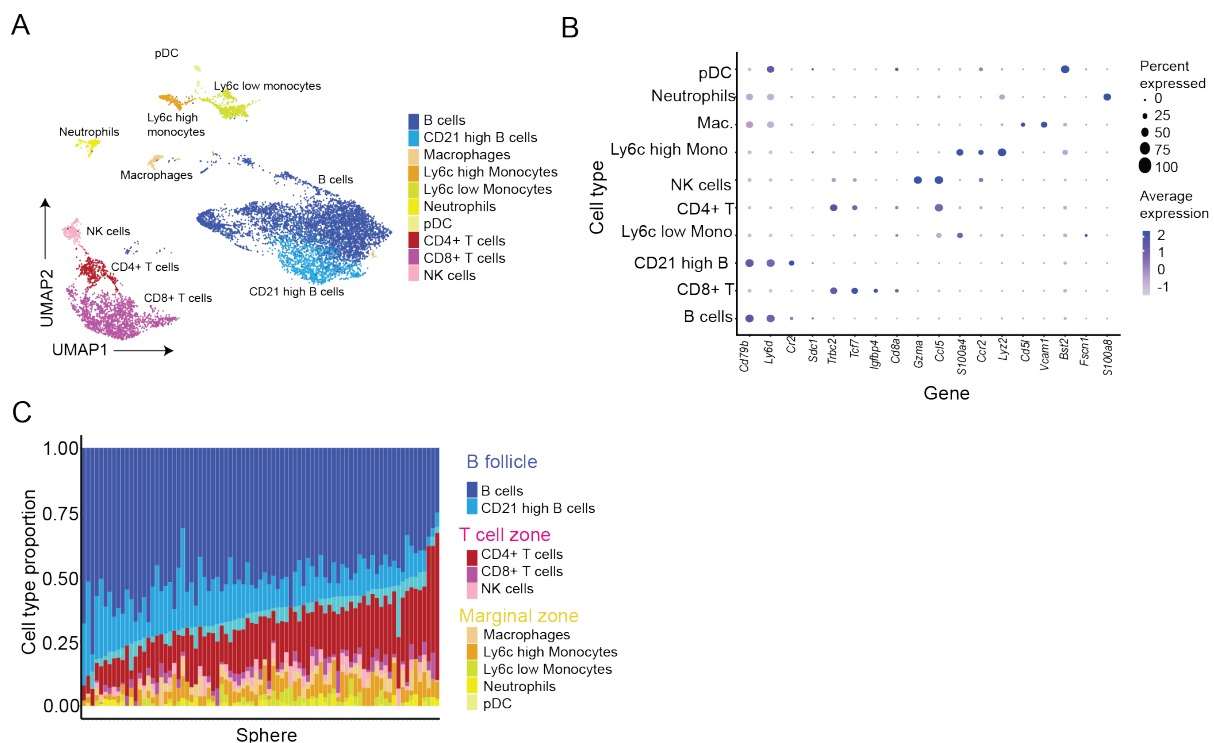


Figure 28: Sphere-seq of healthy murine spleen. A) Uniform manifold approximation and projection (UMAP) from all spleen sphere-seq samples combined ($n = 2$). Cells are colored based on their cell type. **B)** Dotplot of cell type marker expression annotated clusters ($n = 2$ samples). **C)** Barplot showing different proportions of cell types per sphere; only spheres with at least 5 cells ($n = 83$ spheres across 2 samples).

8.7.2 Modified sphere-seq protocol with sequential filtering and manual picking – Example of Crohn’s disease biopsies

The large fragment biosorter is an expensive machine with unique applications and therefore may not be accessible to many researchers. Additionally, sometimes it is not possible to use it for patient-derived biopsy samples due to biosafety constraints. We, therefore, wanted to develop an adaptation of the sphere-seq workflow that can be used in the absence of a biosorter. This protocol includes size-dependent filtering of a defined range, for example between 300 and 500 μm followed by manual picking using a stereo microscope (Fig. 29A). We used this approach to generate sphere-seq data from two Crohn’s disease (CD) patient biopsies. CD is an inflammatory bowel disease (IBD) characterized by chronic inflammation within so-called lesions within the gastrointestinal (GI) tract^{219,220}.

We processed samples from two patients from who we received tissues from the ileocecal valve, which is the part where the small intestine and the colon are connected. We received two samples each: from the inflamed part and an adjacent non-inflamed part. Single cells could be clustered and annotated using known marker genes from previously published studies in CD and Ulcerative colitis (UC)^{190–192} (Fig. 29B, C). Differences in cell type proportions between inflamed and non-inflamed biopsies could be determined, with larger proportions of fibroblasts and macrophages in inflamed tissues (Fig. 29D), which is in line with a recent scRNA-seq study by Martin *et al.* that revealed, for example, an enrichment of inflamed macrophages and activated fibroblasts in inflamed compared to non-inflamed tissues from CD patients¹⁹⁰. We then decided to only consider spheres from the inflamed tissues to analyze different niches within inflamed sites. Like for murine samples, spheres with less than 5 cells were excluded. We could then determine cell type proportions within different spheres (Fig. 29E). A significant inter-patient heterogeneity was apparent between the two patients which could be due to different stages of CD (one might have been more inflamed than the other) or different areas where the tissue was taken from, since a biopsy represents is a very small part of a lesion. It is therefore important to have a large pool of patient samples to make clear conclusions.

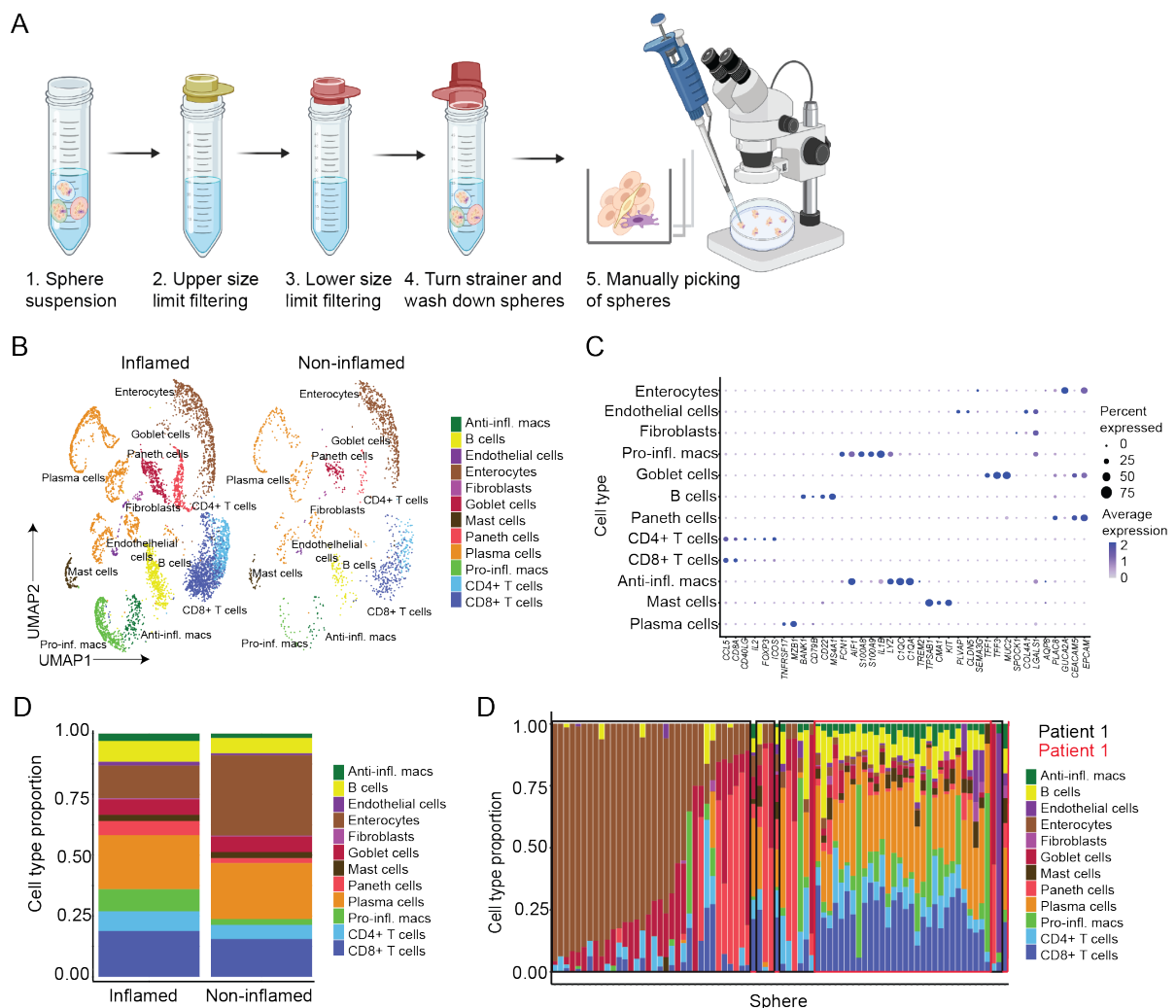


Figure 29: Sphere-seq of Crohn's disease patient biopsies. A) Schematic illustration of workflow that was used for manual picking of spheres instead of sorting with the biosorter. Created with BioRender.com. **B)** UMAP visualization of cell clusters, showing inflamed and non-inflamed biopsies side-by-side ($n =$ inflamed: 78, non-inflamed: 28 spheres across 2 samples). Cells are colored based on their cell type annotation. **C)** Dotplot of cell

type maker genes in different cell type clusters ($n =$ inflamed: 78, non-inflamed: 28 spheres across 2 samples). **D)** Barplot comparing overall cell type proportions from inflamed and non-inflamed tissue samples ($n =$ inflamed: 78, non-inflamed: 28 spheres across 2 samples). **E)** Barplot showing different cell type proportions per sphere from only the inflamed biopsies; spheres with ≥ 5 cells ($n =$ Patient 1: 45, Patient 2: 33 spheres).

8.7.3 Sphere-seq could be used to investigate differences in fibrotic and non-fibrotic Crohn's disease spatial areas

Even though the CD dataset is small, and any results from its analysis should therefore be considered preliminary, we wanted to test if we would still be able to analyze spatial differences within an inflamed CD biopsy. We decided to focus on the sample from patient 1 because it seemed to be that its cell type proportions indicated a higher degree of inflammation (Fig. 30A). We assessed different cell types and their DEGs and uncovered a population of fibroblasts that highly expressed *S100A4*, which has been previously identified to be enriched in activated fibroblasts in CD fibrosis¹⁹². To study spatial heterogeneity between fibrotic and non-fibrotic tissue areas within one biopsy we defined spheres with fibroblasts as 'fibrotic' and spheres without fibroblasts as 'non-fibrotic' (Fig. 30B). Fibrotic areas showed higher abundances of endothelial cells and plasma cells (Fig. 30C, D). We, therefore, wanted to study differences in endothelial cells from fibrotic and non-fibrotic areas further and we found, for example, *ACKR1* to be enriched in fibrotic areas (Fig. 30E). Martin *et al.* found an enrichment of *ACKR1* in endothelial cells from inflamed CD biopsies, which correlated with the presence of activated fibroblasts¹⁹⁰. They additionally found *CCL14* to be expressed in *ACKR1+* fibroblasts¹⁹⁰, a gene we also found to be enriched in fibrosis-associated endothelial cells (Fig. 30E), strengthening our hypothesis that endothelial cells from fibrotic areas were of the *ACKR1+* phenotype as described by Martin *et al.*¹⁹⁰. *ACKR1* (Atypical Chemokine Receptor) also known as Duffy antigen receptor for chemokines (DARC) is involved in chemokine transport and thereby immune cell recruitment²²¹. Therefore, we wanted to analyze the crosstalk of immune cells and endothelial cells, focusing on T cells, by doing an L-R interaction analysis⁹² and comparing results from fibrotic and non-fibrotic spheres (Fig. 30F). The interaction with the highest interaction score in fibrotic areas was between the ligand *CCL5* on T cells and the receptor *ACKR1* on endothelial cells. *ACKR1* is involved in the transport of chemokines like *CCL5* from one compartment to another which leads to immune cell recruitment²²¹. We could therefore speculate that the increased interaction of *CCL5* expressing T cells with *ACKR1* expressing endothelial cells leads to increased recruitment of immune cells by *ACKR1* in fibrotic areas. This hypothesis was strengthened by the increased prediction of *CCL20|CCR6* in fibrotic areas, which is an interaction involved in inflammatory conditions to chemoattract various immune cells like effector and memory T cells²²².

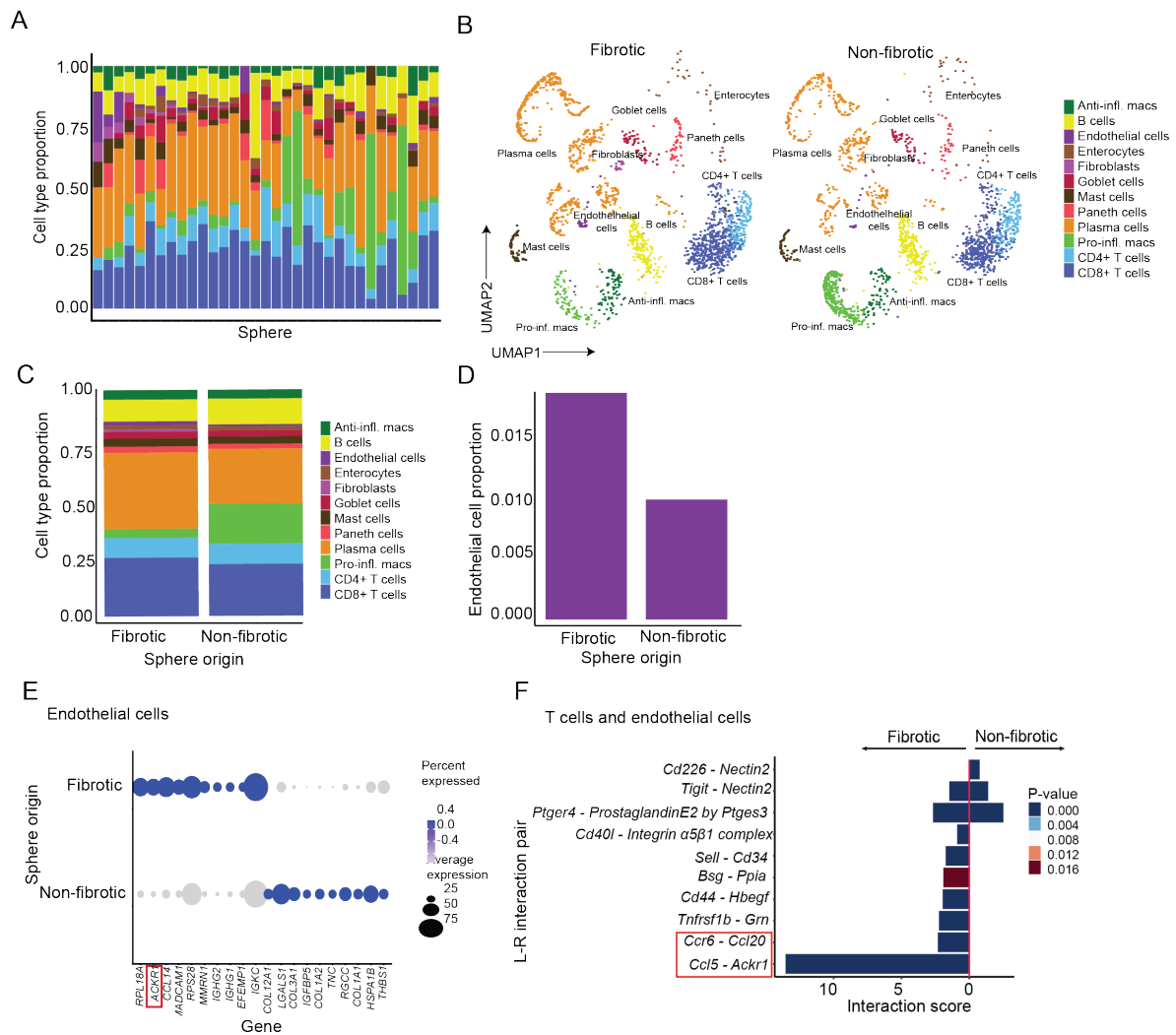


Figure 30: Sphere-seq of Crohn's disease patient 1 biopsy to investigate differences in fibrotic and non-fibrotic tissue areas. A) Barplot showing different cell type proportions per sphere; spheres with ≥ 5 cells ($n = 33$ spheres). **B)** UMAP visualization of cell clusters between spheres grouped into fibrotic and non-fibrotic spatial areas ($n =$ fibrotic: 13, non-fibrotic: 20 spheres across 1 sample). Cells are colored based on their cell type annotation. **C)** Barplot comparing cell type proportions from grouped spheres of fibrotic and non-fibrotic areas ($n =$ fibrotic: 13, non-fibrotic: 20 spheres across 1 sample). **D)** Barplot comparing endothelial cell type proportions from grouped spheres of fibrotic and non-fibrotic areas ($n =$ fibrotic: 13, non-fibrotic: 20 spheres across 1 sample). **E)** Dotplot showing DEGs between endothelial cells of fibrotic and non-fibrotic grouped spheres ($n =$ fibrotic: 11, non-fibrotic: 10 spheres across 1 sample). **F)** Predicted L-R interactions of T and endothelial cells comparing interaction scores in fibrotic and non-fibrotic grouped spheres ($n = 1$ sample). Interaction scores were generated with CellPhoneDB with a permutation test to calculate p-values. Interactions explained in the text are highlighted with a red square.

We could show that with sphere-seq we could generate good quality data from patient biopsies which allowed us to determine spatially restricted differences in cell type abundance, gene expression, and L-R interactions. Of course, it has to be noted, that these results are highly preliminary and have to be confirmed with a larger patient cohort.

8.8 Comparison of sphere-seq to Visium

To benchmark sphere-seq against one of the most widely used ST methods, we wanted to compare it to Visium^{27,45}. We generated Visium libraries from two metastatic liver samples from which we also performed sphere-seq. *In situ* permeabilization for Visium, which leads to mRNA release and capture, has to be optimized for heterogenous tissue compositions. For

example, in metastatic liver, areas of metastasis have different tissue densities than the healthy parts. However, only one permeabilization time can be used for one tissue slice; one has to find a compromise to match the needs of all tissue areas. This can lead to uneven data quality which was far more consistent in sphere-seq experiments as we showed in a comparison of gene feature counts between different spatial regions in Visium and sphere-seq data (Fig. 30A-C). The biggest advantage of sphere-seq compared to array-based ST is single-cell resolution. Visium data needs deconvolution procedures to split spots into single cells. In our data this exposed a clear bias towards the bigger cell type, hepatocytes (Fig. 30D), reducing the ability to analyze spatial features of smaller cells like LECs. We also wanted to test the zoned gene expression we identified in sphere-seq with Visium. Unfortunately, we could not detect reliable gene expression of *Plpp1*, *Galnt15*, and *Vcam1* in our Visium data. Therefore, we used publicly available Visium datasets from Guilliams *et al.* from healthy and non-alcoholic fatty liver disease (NAFLD) livers⁹⁸. This revealed zonation of all three genes (Fig. 30E, F), with *Vcam1* expression being increased and more zoned in NAFLD, aligning with its function in inflammation²⁰⁴. However, because we were lacking single-cell resolution in Visium data, we could not determine whether gene expression came from KCs, LECs or other cell types. In sum, this comparison with Visium showed that sphere-seq is comparatively more suited to studying zoned gene expression in cell types smaller than hepatocytes.

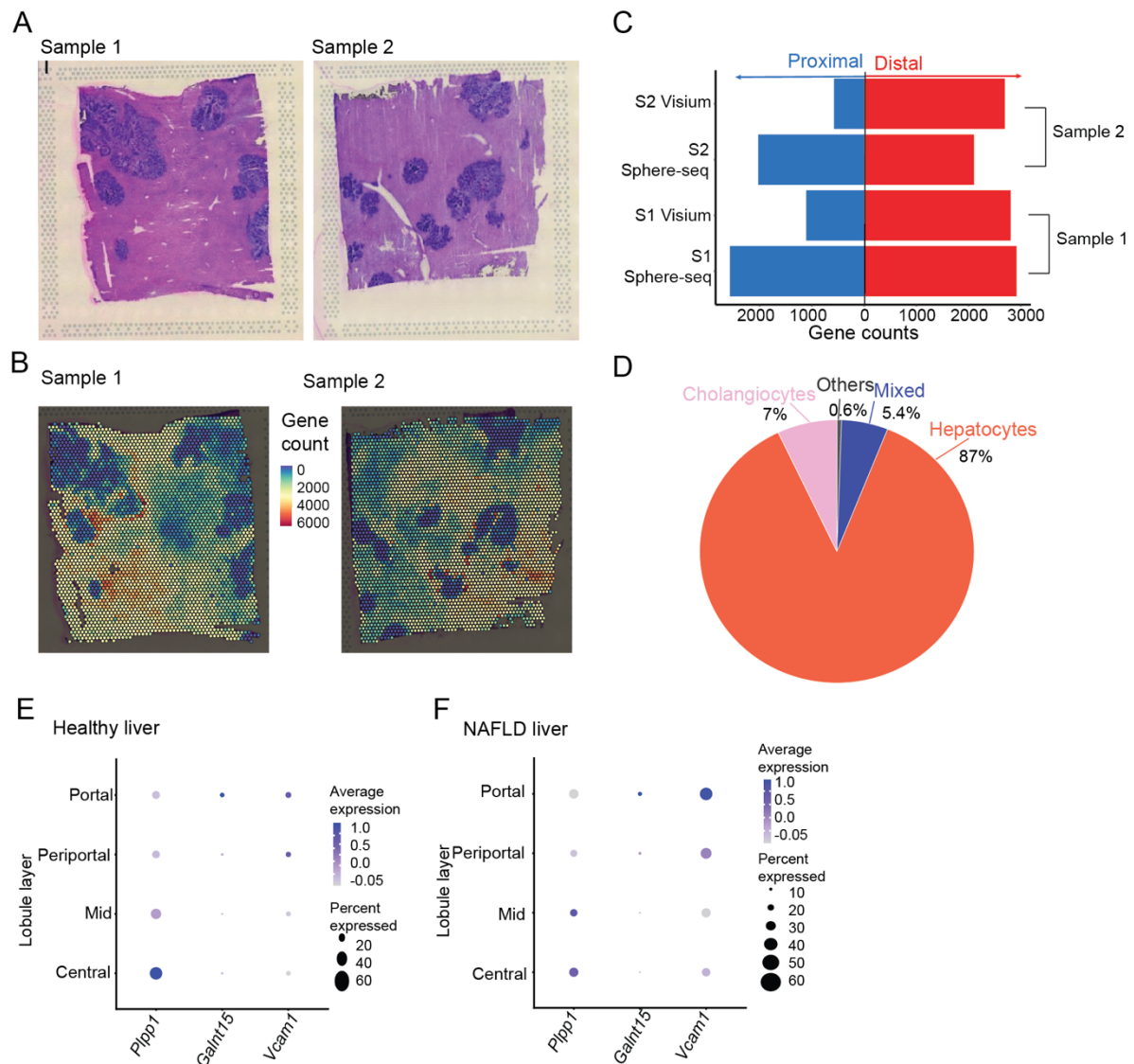


Figure 30: Comparison of sphere-seq with Visium. **A)** H&E staining of mouse metastatic liver samples ($n = 2$). Metastatic areas are dark purple and non-metastatic areas are light purple. **B)** Amounts of gene features in individual spots projected on H&E staining from **A** ($n = 2$ samples). **C)** Barplot showing a comparison of gene counts from different spatial areas of Visium and sphere-seq data ($n = 2$ each). **D)** Piechart showing the average proportions of different cell types after deconvolution of Visium spots. Spots were assigned to individual cell types if at least 75 % of genes were assigned to a specific cell type; the rest of the spots were considered to be mixed ($n = 2$ samples). **E)** Dotplot of average expression of selected genes in publicly available data of Visium experiments from healthy mouse liver⁹⁸. **F)** Same as **E** but from NAFLD mouse liver Visium data⁹⁸.

8.9 Potential biases introduced by sphere size

We wanted to determine if any biases were introduced by different sphere size and cell counts per sphere. We first analyzed if there were overall differences between spatial areas (Fig. 31A, B). There was no significant difference in sphere sizes between spatial areas. Additionally, for metastatic proximal and distal spheres there was no significant difference in cell counts per sphere. Periportal spheres had slightly less cells compared to pericentral spheres. However, we can overall conclude that there is no high variability between grouped sphere conditions. We then wanted to investigate if our zoned gene expression analysis in KCs or LECs was influenced by different size ranges or cell counts. Therefore, we analyzed DEGs between LECs (Fig. 31C) and KCs (Fig. 31D) taking the lobule layers as covariates into account. We

could not find any significant differences in gene expression. Moreover, we did DEGs of LECs from spatially ordered spheres of objects from different technical cutoffs and compared genes that were described before in Fig. 22 (Fig. 31E). We could detect the same genes with slightly different log Fold Changes, but all genes were significantly differentially expressed in all technical subgroups. We, furthermore, could not find any differences in cell type abundances between technical cutoffs (Fig. 31F). Additionally, predicted L-R interactions of separate analyses did not show much variability (Fig. 31G).

All in all, we can conclude that a sphere size of 200-450 μm and a cell count cut-off of 5 cells/sphere was appropriate to capture spatial heterogeneity within the murine metastatic liver. However, this should be tested for different tissues and sample types.

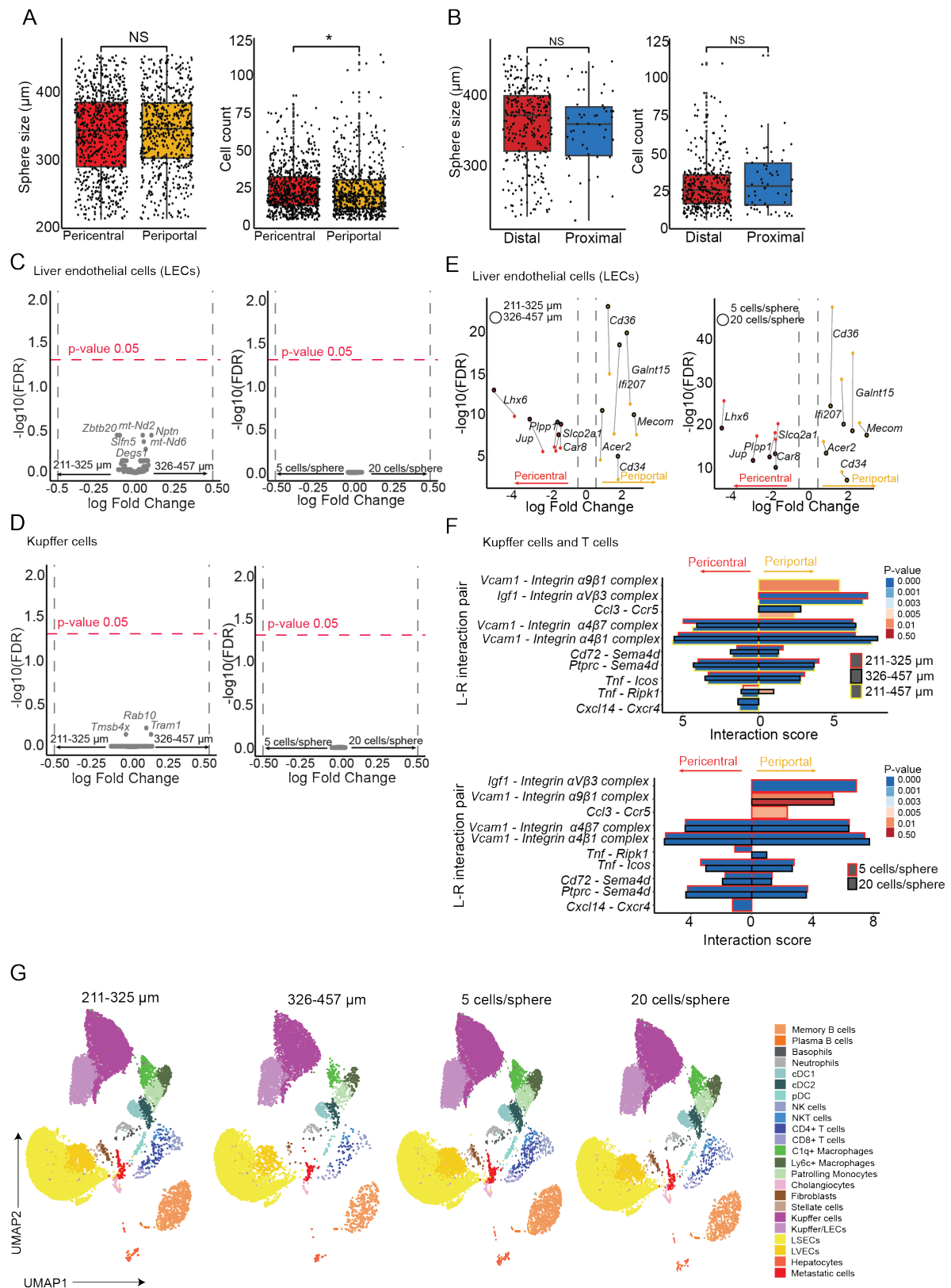


Figure 31: Influence of sphere size and cell counts per sphere. **A)** Boxplot comparing sphere size and cell counts between spheres from pericentral and periportal spheres ($n = n =$ pericentral: 829, periportal: 714 spheres across 9 samples). **B)** Boxplot comparing sphere size and cell counts between distal and proximal spheres ($n =$ distal: 332, proximal: 54 spheres across 3 samples). **C)** DEGs from LECs comparing different technical cutoffs [211-325 ($n=495$ spheres) and 326-457 μm ($n=814$ spheres); 5 cells/sphere ($n=1543$ spheres) and 20 cells/sphere ($n=800$ spheres)]. **D)** DEGs from KCs comparing different technical cutoffs [211-325 ($n=442$ spheres) and 326-457 μm ($n=778$ spheres); 5 cells/sphere ($n=794$ spheres) and 20 cells/sphere ($n=1436$)]. **E)** DEGs in LECs of spatially

ordered spheres between different cutoffs, left different sizes and right different cell count [$n = 211-325$ ($n=495$ spheres), $326-457 \mu\text{m}$ ($n=814$ spheres), 5 cells/sphere ($n=1543$ spheres), 20 cells/sphere ($n=800$ spheres)]. Colored dots represent significantly enriched genes; red, enriched in pericentral zones; yellow, enriched in periportal zones. Gene labels indicate the same genes between analysis. Dots from $326-457 \mu\text{m}$ and 20 cells/sphere analysis are highlighted with black borders around dots. Dots from the same gene between analysis of technical cutoffs are connected with lines. **F**) Predicted ligand-receptor (L-R) interactions between KCs and T cells in pericentral or periportal zones ($n=9$ samples). Interaction scores were calculated from sphere-seq data of different technical cutoffs by CellPhoneDB, which uses a permutation test to generate p -values indicating significantly enriched L-R interactions. Interaction scores from spheres of different technical cutoffs are highlighted by different colors surrounding the bars (red: $211-325 \mu\text{m}$ and 5 cells/sphere, black: $326-457 \mu\text{m}$ and 20 cells/sphere, yellow: $211-457 \mu\text{m}$). **G**) UMAP visualization of annotated cell types comparing different technical cutoffs. For **A** and **B**, dots represent individual spheres; the middle line represents the median; upper and lower line are the first and third quartile (25^{th} and 75^{th} percentiles); the upper and lower whiskers extend to $1.5 * \text{the distance between first and third quartile}$; dots outside of the whiskers are considered outliers; p -values are analyzed using a non-parametric Wilcoxon signed-rank test was applied (not significant > 0.05). For **C**, **D** and **E** a negative binomial generalized log-linear model was applied for statistical analysis with taking lobule layers as covariates into account in **C** and **D**, p -values (Benjamin-Hochberg adjusted) < 0.05 were considered as significant.

9 Discussion

ST is a fast-developing field and even though there is constant improvement in technology development there are still many limitations scientists have to face when considering which ST technology best fits their biological question. We developed sphere-seq, a method that analyses small cellular niches at single-cell resolution. The spatial location of these niches is reconstructed using either landmark gene expression signatures or cell type abundances. We applied sphere-seq to study spatial features of the CRC metastasis-bearing mouse liver and we could identify previously uncharacterized zoned gene expression in LECs and KCs as well as differences in cell type abundances between areas, which were distal or proximal to micro-metastatic sites. We further found spatially restricted L-R interaction pairs. Furthermore, we applied sphere-seq to other tissues and could demonstrate its broad applicability. Comparison with the most widely used ST platform, Visium, highlighted some of the strengths of sphere-seq. However, as with every other ST method, there are also limitations with sphere-seq. To get a clear picture, these strengths and weaknesses will be discussed by comparing sphere-seq to other ST technologies of different categories. The integration of spatial information we gained from sphere-seq with L-R interaction analysis also bears some advantages that will be addressed. Application of sphere-seq to mouse metastatic liver helped us to get new biological insights into liver zonation and metastasis formation. These new insights will be described together with an outlook of experiments to test these further. Options for enhancement and improvements of the sphere-seq approach will be discussed at the end of this section.

9.1 Strength of sphere-seq

Sphere-seq falls under the umbrella of mRNA-capture-based partial dissociation methods^{64–66}, however we extended the approach to single cell resolution and larger cellular communities. Therefore, sphere-seq does not rely computational deconvolution and it enables the analysis of larger, still spatially relevant, cellular subunits in three dimensions. When comparing to most imaging-based approaches like SeqFISH+³³ and MERFISH^{31,32}, which are based on mRNA probe hybridization and visualization, mRNA capture is unsupervised and no prior knowledge is needed for gene panel design, therefore, also previously unknown gene expression can be identified. Of course, there are also unsupervised imaging-based ST methods that are based on *in situ* sequencing like STARmap³⁶ or FISSEQ³⁷, but these technologies depend on sophisticated technologies like high-resolution microscopes that require certain expertise which is not always available for a broad scientific community. This is also true for array-based methods like DBiT-seq⁵¹ that uses complicated microfluidic channels to introduce spatial barcodes, or illumination-based ROI ST methods like ZipSeq⁶¹, Light-seq⁶², or SPACECAT⁶⁰ that require microscopic devices. Sphere-seq also uses a specific machine, the large fragment biosorter, however, we also adapted the protocol to work without the sorter. Spheres can be size selected by sequential filtering through size-dependent strainers following manual picking. This approach is tedious and takes more time, but it makes sphere-seq available to a broader scientific community. For very small and precious samples this strategy can also be used to make sure not to lose any tissue parts. All the other reagents used for sphere-seq like the lipid hashing anchors, and sphere BCs are commercially available and scRNA-seq approaches are often available on an institute level within facilities. However, if such facilities are not available, single cells could be captured with

PIP-seq, an approach recently published as a preprint, which uses a microfluidics-free capture approach using standard laboratory tools generating bead-single cell droplet encapsulations by solely vortexing²²³.

Unlike imaging-based and array-based mRNA-capture technologies, sphere-seq does not require tissue slices. These slices are most of the time only 10 μm in thickness and they, therefore, sometimes do not even capture entire cells. Sphere-seq allows the analysis of niches in three dimensions and spheres can be sampled in a random or targeted process from a whole tissue sample instead of handpicking one flat area within a large tissue block. Tissue fixation and embedding for slice-based methods are sometimes difficult if tissue samples are small or have shapes that do not properly fit the capture area. One example is punch biopsies, for example in dermatology, for which a metal cylinder is introduced into a patient's tissue of interest removing a small piece for pathological diagnostics²²⁴. The size of these biopsies is usually approximately 2-4 mm in diameter²²⁴. It, therefore, can be challenging to properly embed these small and fragile tissues to archive a certain directory within a tissue block for proper slicing. Additionally, for example in Visium the capture area is 6.5x6.5 mm and therefore most spots would be uncovered from the punch biopsy tissue. Sphere-seq could process punch biopsies easier because it does not depend on a specific sample size and shape. Additionally, in comparison the Visium sphere-seq does not rely on tissue permeabilization allowing to capture mRNAs from different spatial areas at similar rates, as we have shown.

The biggest advantage of sphere-seq is single-cell resolution compared to other techniques^{27,45,64-66}. This could be highlighted on the example of zoned *Vcam1* gene expression in KCs. We tested *Vcam1* in Visium datasets of healthy and NAFLD⁹⁸ and found zoned gene expression. However, due to the lack of single-cell resolution, it was difficult to know if *Vcam1* was coming from LECs or KCs. Another phenomenon, due to the nature of Visium to capture more than one cell is a bias towards larger cells like hepatocytes, highlighted in the deconvolution of our Visium data where 87 % of spots had at least 75 % of genes allocated to hepatocytes. Sphere-seq shows higher amounts of little cells like LECs and immune cells, which allows us to analyze the spatial heterogeneity of these small cells in a highly confident manner. Many imaging-based methods also have single-cell resolution, even sub-cellular resolution in most cases, like HybISS³⁸, ExSeq⁴¹, or MERFISH³¹. However, it is often challenging to do cell segmentation, especially in very dense tissue areas.

Another advantage of sphere-seq is that it enables the analysis of inter-spatial neighborhood differences. Other ROI based methods only allow the analysis of a couple of ROIs (for example up to three in SPACECAT⁶⁰, up to eight in ZipSeq⁶¹), and thereby cells can only be subdivided into broad neighborhoods, while sphere-seq potentially allows subdividing broad neighborhoods because of the spheres-BC.

Sphere-seq is based on scRNA-seq and therefore, it has the advantage, that computational tools that were developed for scRNA-seq analysis can be used for its analysis too.

Taking all of this together, sphere-seq has many advantages over other ST technologies, with single-cell resolution, transcriptome-wide interrogation, high throughput and data analysis simplicity as some of the strongest.

9.2 Limitations of sphere-seq

Like any other ST technology, sphere-seq has limitations and weaknesses. First of all, the method is sparse, during the sphere-seq experimental procedure many cells get lost and the

starting amount is often small. We acquire single-cell transcriptomes of a fraction of cells per sphere; around 0.4-1.2 %. We try to circumvent this problem by grouping spheres from different spatial regions and doing in-depth analysis on grouped spheres. As demonstrated, the results are consistent when comparing a cutoff of 5 and 20 cells per sphere.

There are also certain cell type biases introduced with sphere-seq. For example, in liver samples, we mainly lose hepatocytes, cholangiocytes, metastatic cells and fibroblasts, which is, however, in line with results from conventional single-cell RNA-seq experiments due to an *ex vivo* digestion approach. Therefore, we can conclude that this is not a limitation introduced specifically by sphere-seq, but a general pitfall when preparing single cells from liver tissues. There are also certain biases introduced by the single-cell capture method, for example, between the 10X Genomics and BD Rhapsody single-cell capture approach. Fragile cells like granulocytes, for example, eosinophils, cannot be properly captured in 10X Genomics, while in BD Rhapsody, which is a microwell-based approach and therefore more gentle because there are no microfluidic pressures involved, can capture these fragile cells in higher quantities²²⁵. Because sphere-seq utilizes aspects of conventional scRNA-seq protocols, the single-cell capture platform can be changed depending on the cell types of interest.

Another disadvantage of sphere-seq is that it only works for fresh tissues, which is especially detrimental for patient biopsy samples because it takes some time from the surgery until the sample arrives in the lab, because there are usually some diagnostic steps happening in-between. This also makes sphere-seq disadvantageous over other methods that can use samples from biobanks. Biobanks store tissues sometimes fresh frozen, but more often samples are formalin-fixed paraffin-embedded (FFPE). Fresh frozen tissues can be used for many other ST technologies that require tissue sections like Visium^{27,45}, FISSEQ+³⁷ or STARmap³⁶, to only name a few. In addition, new methods are being developed that permit the application of ST to FFPE fixed tissues⁵². Sphere-seq cannot be used for FFPE samples. Fresh frozen tissues cannot be used for scRNA-seq, however they can be used for snRNA-seq by extracting nuclei by using lysing reagents and mechanical force²²⁶. It might be possible to adapt the sphere-seq protocol for snRNA-seq (will be discussed further in the section of sphere-seq optimizations/enhancements).

Sphere-seq is mostly a random sampling process and does not consider microanatomical structure when sorting spheres. This could lead to having individual spheres that cover multiple spatial niches as shown in the example of sphere-seq on the spleen. Other mRNA-capture-based ROI ST methods are better in this by selecting ROIs by visually looking and choosing micro-anatomical structures under a microscope.

Sphere-seq has a disadvantage in detecting rare cell types that are sensitive to single-cell dissociation. Methods like Light-seq for example, which includes *in situ* RT and spatial barcoding without the need for single-cell dissociation can capture rare cells - as has been shown by the analysis of difficult-to-isolate dopaminergic amacrine cells (DACs) from the retina in the mouse⁶².

All in all, sphere-seq has some limitations mostly due to the need for fresh samples and due to the loss of cells during the experimental procedure.

9.3 Sphere-seq and L-R interaction analysis

Approaches that integrate spatial information with L-R interaction prediction often take gene expression from Visium spots into account, and therefore enrichment of L-R interactions from one spatial area to another can be analyzed, but it cannot properly be determined which cell types are involved in the communication, they can only be assumed by deconvolution analysis.

Sphere-seq provides single-cell identities and information about spatial areas and is therefore superior.

L-R prediction tools usually result in extensive lists of many L-R pairs and it is difficult to narrow down the most important ones to do further functional testing. Sphere-seq can help with that by identifying spatially most relevant interactions by analyzing differences between spatial locations.

L-R interaction analysis, however, relies on transcriptomics data and this might differ from ligand and receptor expression on the protein level which is the biggest limitation of inferring L-R interactions from transcriptomics data. In the case of sphere-seq, one could combine it with CITE-seq²⁰ or adapt sphere-seq for a plate FACS readout (will be discussed further in the sphere-seq enhancements/optimizations section). For validating predicted L-R interactions, one could perform colocalization analysis of microscopic images from protein-stains of ligands and receptors.

In general, however, protein expression of ligands and receptors and also their colocalization do not mean that these proteins indeed interact, functional validation experiments are needed. This can be achieved by inhibiting or knocking out one interacting partner (ligand or receptor) and analyzing differences in the phenotype of downstream functions⁹⁰. *In vitro* L-R interactions could be functionally tested in co-culture experiments by for example transiently overexpressing a ligand in one cell type and analyzing the influence of receptor expression on the other cell type²²⁷. *In vivo*, conditional cell-type-specific Cre-loxP systems could be used to delete the expression of ligands or receptors followed by an analysis of phenotypic and downstream signaling effects. In this system, the expression of a bacteriophage *cre* (cyclization recombinase) gene is controlled by a cell-specific promoter and once it is activated, Cre catalyzes DNA recombination of two *loxP* (locus of X-over in P1) sites to for example mediate a genetic deletion²²⁸ - in our case this could be the deletion of exons within *Vcam1* leading to gene inactivation under the control of a KC-specific promoter. There are many transgenic mice available with cell type-specific Cre drivers. Drivers within NPC in the liver are especially interesting to this Ph.D. work. For example, in LECs, the vascular endothelial cadherin Cre driver *Cdh5-PAC-CreER^{T2}* was successfully used to knock down C-X-C chemokine receptor (CXCR)4 and CXCR7²²⁸ and in KCs *LysM-Cre* was successfully used to test knockdown of *Myd88* in a recent study to determine its importance in KCs for immune zonation¹⁴⁷.

The fact that we could identify spatially specific L-R interactions between metastatic-proximal and -distal sites between macrophages and T cells motivated us to identify L-R interactions at the interface between metastatic cells and LECs in a follow-up study in the laboratory. Thereby only cells from proximal areas have been taken into account removing unwanted variation due to LECs which are spatially more distant. Uncovered L-R interactions are currently being functionally validated using an *in vivo* CRISPRa (Clustered Regulatory Interspaced Short Palindromic Repeats activation) screen targeting and thereby activating ligand and receptor genes on LECs in combination with a metastatic seeding system where cancer cells secrete lipid-soluble tagged mCherry to label the surrounding cells²²⁹; an approach that has been optimized to study the interface between metastatic cells and hepatocytes in the laboratory (Borrelli *et al.*, in preparation). First the liver is perturbed with the library of ligand and receptor target genes generating a so-called mosaic liver. This is followed by an intrasplenic injection of CRC organoids that graft and form metastatic sites which secrete mCherry, labeling the surrounding cells. The surrounding LECs can then be sorted and enriched guides can be identified, which can be interpreted as ligands/receptors that are suspected to play an important functional role in the metastatic process.

9.4 Biological knowledge gained with sphere-seq and outlook

Comparisons of different spatial areas within the mouse metastatic liver resulted in many findings, for example, we uncovered cell type abundances, gene expression, and L-R interactions that were spatially restricted. Some of these findings, and strategies and how to investigate and test them further will be discussed in more depth within this section.

Application of sphere-seq to study liver zonation in the metastasis-bearing mouse liver helped us to describe zoned gene expression in KCs for the first time to our knowledge. To date, zoned gene expression in the liver was described for hepatocytes and LECs, which acquire different metabolic functions depending on where they are within the central-portal axis^{65,104,111,116,141}, and HSCs, where a pericentrally zoned gene expression signature suggests an involvement in collagen production during centrilobular injury-induced fibrosis²³⁰. Furthermore, there is immune zonation, particularly of KCs and NKT cells, which is induced by environmental signals provided by incoming bacteria and pathogens from the gut when mice are weaned¹⁴⁷. Thanks to reconstructing the original lobule layer of spheres based on gene signatures in LECs, we could now accomplish the analysis of zoned gene expression in all kinds of other cell types and while we did not find any significantly zoned gene expression in non-tissue resident immune cells, we found a significant zonation of *Vcam1* in KCs. These results were obtained in a mouse model of liver metastasis and it would be interesting to determine *Vcam1* zonation in KCs also in other mouse models with perturbed livers, for example, NAFLD, hepatitis or HCC to determine if the zonation is a general effect or specific for liver metastasis. We analyzed *Vcam1* expression in NAFLD Visium data⁹⁸ and found it to be zoned, however, the lack of single-cell resolution did not allow us to confidently allocate this zonation to KC only.

We found predicted L-R interactions between the ligand VCAM1 on KCs and integrins on T and B cells to be enriched periportally. It would be interesting to functionally validate this interaction by deleting VCAM1 on KCs and determining the effect on T and B cells. They might be reduced in number, which could favor liver metastasis formation. One strategy to knock down VCAM1 ligand expression on KCs could be the usage of the previously mentioned constitutive Cre-loxP system²²⁸. The influence on lymphocyte recruitment and metastasis formation or progression could then be analyzed. The same approach could be used to functionally validate other L-R interactions between KCs and lymphocytes if transgenic mice are available.

The main finding from the comparison of metastasis-distal and -proximal sites was the enrichment of *C1q+* macrophages within proximal areas, *Ly6c+* macrophages and patrolling monocytes showed similar ratios between both areas. *C1q+* macrophages have been proposed to have an anti-inflammatory phenotype²³¹ while *Ly6c+* macrophages are considered to be pro-inflammatory²³². Patrolling monocytes are involved in the removal of damaged cells and debris²³³. A lot is already known about the involvement of *C1q+* macrophages in T cell exhaustion and cancer progression¹²⁷. Unfortunately, we could not differentiate T cells into different subtypes for L-R interaction analysis, because cell numbers were too sparse. However, we could try to characterize T cells within metastatic proximal areas better by either increasing sample size and thereby T cell numbers or by using imaging-based approaches to label exhausted T cells which are characterized by expression of for example *Pd-1* or *Tigit*²³⁴.

Our L-R interaction analysis between distal and proximal sites identified high proximal enrichment of interactions that include *Spp1* and *Fn1* ligand expression on macrophages/monocytes. Both were associated with potential CRC tumor progression²¹⁵.

Recently, there is a high research interest in SPP1 expressing macrophages in CRC^{215,235}, hepatocellular carcinoma (HCC), and liver metastasis¹⁶⁰. SPP1+ macrophages were found to be involved in angiogenesis²³⁶ and hypoxia²³⁵. Qi *et al.* found that SPP1+ macrophages are interacting with FAP-expressing fibroblasts in CRC which limits T cell infiltration and therefore a patient's responsiveness to immunotherapy²³⁵. In addition, it has been proposed that SPP1 directly exerts inhibitory function on T cells within the TME²¹⁴. The direct influence of SPP1 ligand expression of macrophages on T cells could be tested using a co-culture system of macrophages and T cells *in vitro*²³⁷. SPP1 overexpression could be induced *in vitro* in bone marrow-derived macrophages, using for example a transient lipofectamine transfections^{227,238} and the influence on T cell exhaustion could be assessed. Alternatively recombinant SPP1 could be exposed to T cells. These assays could likewise be done with FN1 or other ligands which were found to be enriched in interactions with T cells within proximal metastatic areas to study the functional influence of these ligands on T cells.

9.5 Future applications of sphere-seq

Our preliminary results on CD biopsy samples show a great potential of sphere-seq to discover spatial heterogeneity within inflamed tissues. Fibrosis formation and progression in CD are especially poorly understood and there is no anti-fibrotic therapy available at the moment²⁴¹. Therefore, understanding the genes and cell types involved in fibrosis could be important to develop therapies in the future. Using sphere-seq, we have already shown that we could divide a CD lesion into fibrotic and non-fibrotic areas. Within a resected lesion there might be different stages of fibrosis we could dissect with sphere-seq and differences in each respective microenvironment could be identified. It would be interesting to further analyze different subtypes of immune cells and their cellular crosstalk in fibrotic areas. However, it is important to increase the sample size, at the moment the analysis of fibrotic against non-fibrotic areas has been done within only one biopsy sample. Therefore, it is difficult to make definitive conclusions. However, we could already see a great potential for sphere-seq to address some of the outstanding questions in CD fibrosis research.

In liver metastasis, sphere-seq could be further used to study metastasis at different stages of disease progression. Spheres could be grouped into areas from within the center of a metastasis across the border to regions of healthy liver tissue, provided that landmark genes within different distances could be identified in advance for example by array-based ST methods like Visium^{27,45}. However, this would require a larger number of spheres from metastatic samples than we are currently able to obtain.

9.6 Future optimizations/enhancements of sphere-seq

Sphere-seq has the potential to be developed further to be applied to other sequencing modalities or to increase its throughput and efficiency.

The sphere-seq experimental workflow has some parts that could be further optimized. The labeling efficiency, which is only 50 % at the moment, could be further increased by introducing a filtering step of cells before sphere labeling using 96 well strainer plates which would help to remove connective tissue that impacts cell labeling. However, with this approach it has to be considered that this might lead to cell loss. To address the problem of cell loss, washing steps could be reduced. Thereby, although, it has to be considered that this could increase MULTI-seq background labeling, which leads to false doublet classifications of cells. One could try to

filter labeled cells through a 1 µm strainer (pluriSelect), where only unbound lipid anchors:BCs can pass through, and labeled cells that get stuck in the strainer could be recovered. The sphere-size could be further reduced for a higher spatial resolution, we successfully single sorted mouse colon crypts that were only 70 µm in size. Lower cell numbers per sphere, however, impact the overall cell recovery.

At the moment, 288 spheres can be analyzed per single experiment, this number could be increased by designing a larger cohort of sphere BCs and spheres could be sorted into 384-well plates instead of 96-well plates. Manual pipetting after sphere sorting would be challenging, but sphere-seq could be adapted to an automated workflow using a pipetting robot.

More than one modality could be analyzed with sphere-seq, scRNA-seq could be combined with surface antibody staining using the CITE-seq²⁰ method or V(D)J sequencing²⁴² to analyze B cell or T cell clonal expansion, in combination with the scRNA-seq readout.

As already mentioned earlier, sphere-seq could be adapted for snRNA-seq. However, this requires extensive changes within many steps of the protocol. The most challenging part would be the nuclei isolation while keeping spheres in their wells. For snRNA-seq nuclei are extracted using mechanical homogenization with a mortar and pestle while applying detergents in low concentrations²⁴¹. The detergents disrupt the cellular membrane while keeping the nuclear membrane intact, but the right concentration has to be found to reduce damage to the nuclear membrane^{241,242}. While there are many options to apply mechanical force to a whole sample, it is more difficult to do so to small sphere samples in a 96-well plate. One option, however, would be low-intensity sonication using a Multi-Sample Sonicator for 96 well plates^{244,245}. For labeling single nuclei with sphere-specific BC, lipid anchors cannot be used because the nuclear membrane has a different composition. While the cellular membrane is made up of phospholipids that allow integration of lipids, the nuclear membrane is richer in cholesterol^{246,247}, therefore, McGinnis *et al.* developed a cholesterol anchor for their MULTI-seq method²¹, which can be used to label extracted nuclei.

More extensive biosorter strategies could be employed, for example, to enrich spheres with more than one cell type or gene expression condition. Spheres could be stained with fluorescently labeled surface antibodies to enrich for spheres with specific cell types. For example, in the spleen for B cell or T cell areas. Alternatively, a reporter mouse that expresses a fluorophore in distinct spatial areas like metastatic areas could be used. However, this is quite challenging, we were trying to use this approach: metastatic sites were GFP positive because the CRC organoids expressed GFP, however, we observed very low GFP signals (data not shown), precluding us from using this measure for sorting. We hypothesize that GFP signal could have been hidden by GFP negative cells surrounding the metastatic cells. Additionally, the sorter also has a function to sort for specific shapes, this could be used to sort spheres with shapes that are specific for certain tissue units, like for example intact crypts within a colon patient biopsy.

Sphere-seq could be combined with CITE-seq²⁰ or the protocol could be adapted for a plate FACS readout: in brief, after single-cell dissociation, each sphere could be stained with a cocktail of surface antibodies and then they could be read out by FACS. While CITE-seq enables the analysis of the transcriptome and proteome, it only works for ligands and receptors that are expressed on the surface. If only adapting sphere-seq with a FACS readout also intracellular ligand-receptor protein expression could be analyzed.

Another idea is to adapt sphere-seq for sphere bulk RNA-seq, spheres could be sorted and then processed in-plate following the SMART-seq2 workflow^{56,57}, which allows the analysis of full-length transcripts. Each well and thereby sphere would be labeled with a different BC. This

would be similar to Clump-seq⁶⁶, however, our biosorter approach allows for the sorting of larger communities than 10 cells. One possible application for this would be a pooled CRISPR screen to disrupt functions in, for example, structural units *in vitro* like organoids or even *in vivo* within a certain tissue.

10 Conclusion

In my Ph.D. project, the goal was to develop a broadly applicable new ST technology which captures spatial heterogeneity, but has the capacity of a single-cell readout. We wanted to be able to identify spatial differences in gene expression and cell type abundances. Additionally, sphere-seq readouts should help to narrow down relevant L-R interactions. The integration of spatial information from sphere-seq together with a scRNA-seq readout allows us to identify spatially relevant interactions that could be used for further functional investigations. Many currently available ST technologies are lacking either single cell resolution^{27,45,64–66}, need extensive expertise in high-resolution microscopy^{31,36–38,41}, or are restricted to a low number of ROIs^{60–62}. The sphere-seq workflow combines sorting of cellular communities, single-cell dissociation, sphere labeling, and scRNA-seq and thereby addresses all these limitations. Single-cell resolution is archived by conventional scRNA-seq, the only expertise that is needed is flow cytometry and scRNA-seq, which are commonly used techniques. We even developed an approach that uses manual sphere picking instead of using the biosorter and therefore sphere-seq can be deployed in every laboratory that has standard equipment and a facility providing scRNA-seq. Furthermore, the fact that sphere-seq utilizes the conventional computational workflows of scRNA-seq make it easy to use for a scientist that is not proficient in programming.

We could uncover zoned gene expression in KCs for the first time to our awareness, which might have opened a new branch of liver zonation research, that employs zoned gene expression in this cell type in different disease contexts, and also in human livers. Comparing areas which are distal or proximal to metastatic sites lead to the identification of spatially restricted L-R interactions that might promote metastasis or limit its spread. We uncovered a set of potentially relevant interactions that could be investigated further to eventually identify novel therapeutic targets. Applying sphere-seq to various tissues and species, like the murine liver and spleen, CRC organoids, and Crohn's disease patient biopsies, showed the great potential of sphere-seq to answer various biological questions in a broad spectrum of research of homeostatic and pathological conditions.

There are also limitations to sphere-seq, like the fact that it can only be applied to fresh tissues at the moment, or the need for landmark signatures for reconstructing the spatial location. However, every ST methodology has strengths and limitations, which have to be considered when being chosen to address a biological problem. For example, sphere-seq might not be helpful to study spatial differences in hepatocytes, but it seems to be superior when studying much smaller cells like LECs or KCs in a spatial context.

To conclude, even though there are many different ST methodologies available already, sphere-seq shows many features that are superior to other methods, and therefore it is a valuable addition to the field of ST.

11 Bibliography

1. LaRossa, R. A. Transcriptome. in *Brenner's Encyclopedia of Genetics (Second Edition)* (eds. Maloy, S. & Hughes, K.) 101–103 (Academic Press, 2013).
2. Ergin, S., Kherad, N. & Alagoz, M. RNA sequencing and its applications in cancer and rare diseases. *Mol. Biol. Rep.* **49**, 2325–2333 (2022).
3. Nagle, M. P., Tam, G. S., Maltz, E., Hemminger, Z. & Wollman, R. Bridging scales: From cell biology to physiology using in situ single-cell technologies. *Cell Syst* **12**, 388–400 (2021).
4. Wen, L. *et al.* Single-cell technologies: From research to application. *Innov. J.* **3**, 100342 (2022).
5. Kleino, I., Frolovaité, P., Suomi, T. & Elo, L. L. Computational solutions for spatial transcriptomics. *Comput. Struct. Biotechnol. J.* **20**, 4870–4884 (2022).
6. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).
7. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
8. Nayak, R. & Hasija, Y. A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. *Genomics* **113**, 606–619 (2021).
9. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine* vol. 50 1–14 Preprint at <https://doi.org/10.1038/s12276-018-0071-8> (2018).
10. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
11. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* vol. 161 1187–1201 Preprint at <https://doi.org/10.1016/j.cell.2015.04.044> (2015).
12. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biology* vol. 17 Preprint at <https://doi.org/10.1186/s13059-016-0938-8> (2016).
13. Keren-Shaul, H. *et al.* MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nature Protocols* vol. 14 1841–1862 Preprint at <https://doi.org/10.1038/s41596-019-0164-4> (2019).
14. Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
15. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
16. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).

17. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
18. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
19. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
20. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
21. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
22. He, S. *et al.* Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol.* **21**, 294 (2020).
23. Tabula Muris Consortium *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
24. Zhang, Y. *et al.* Single-cell RNA sequencing in cancer research. *J. Exp. Clin. Cancer Res.* **40**, 81 (2021).
25. Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clin. Transl. Med.* **12**, e694 (2022).
26. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
27. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
28. Lee, J., Yoo, M. & Choi, J. Recent advances in spatially resolved transcriptomics: challenges and opportunities. *BMB Rep.* **55**, 113–124 (2022).
29. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of Single RNA Transcripts in Situ. *Science* vol. 280 585–590 Preprint at <https://doi.org/10.1126/science.280.5363.585> (1998).
30. Tian, L., Chen, F. & Macosko, E. Z. The expanding vistas of spatial transcriptomics. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-01448-2.
31. Moffitt, J. R. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11046–11051 (2016).
32. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
33. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).

34. Goh, J. J. L. *et al.* Highly specific multiplexed RNA imaging in tissues with split-FISH. *Nat. Methods* **17**, 689–693 (2020).
35. Borm, L. E. *et al.* Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-01455-3.
36. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, (2018).
37. Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
38. Gyllborg, D. *et al.* Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res.* **48**, e112 (2020).
39. Nilsson, M. *et al.* Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088 (1994).
40. Chen, F., Tillberg, P. W. & Boyden, E. S. Optical imaging. Expansion microscopy. *Science* **347**, 543–548 (2015).
41. Alon, S. *et al.* Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, (2021).
42. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
43. Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).
44. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet.* (2022) doi:10.1038/s41576-022-00515-3.
45. Salmén, F. *et al.* Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* **13**, 2501–2534 (2018).
46. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
47. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
48. Cho, C.-S. *et al.* Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559-3572.e22 (2021).
49. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
50. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792.e21 (2022).

51. Liu, Y. *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681.e18 (2020).
52. Gracia Villacampa, E. *et al.* Genome-wide spatial expression profiling in formalin-fixed tissues. *Cell Genomics* **1**, 100065 (2021).
53. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
54. Fu, X. *et al.* Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell* **185**, 4621-4633.e17 (2022).
55. Espina, V. *et al.* Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603 (2006).
56. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* vol. 10 1096–1098 Preprint at <https://doi.org/10.1038/nmeth.2639> (2013).
57. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols* vol. 9 171–181 Preprint at <https://doi.org/10.1038/nprot.2014.006> (2014).
58. Nichterwitz, S., Benitez, J. A., Hoogstraaten, R., Deng, Q. & Hedlund, E. LCM-Seq: A Method for Spatial Transcriptomic Profiling Using Laser Capture Microdissection Coupled with PolyA-Based RNA Sequencing. *Methods Mol. Biol.* **1649**, 95–110 (2018).
59. Medaglia, C. *et al.* Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
60. Genshaft, A. S. *et al.* Live cell tagging tracking and isolation for spatial transcriptomics using photoactivatable cell dyes. *Nat. Commun.* **12**, 4995 (2021).
61. Hu, K. H. *et al.* ZipSeq: barcoding for real-time mapping of single cell transcriptomes. *Nat. Methods* **17**, 833–843 (2020).
62. Kishi, J. Y. *et al.* Light-Seq: light-directed in situ barcoding of biomolecules in fixed cells and tissues for spatially indexed sequencing. *Nat. Methods* (2022) doi:10.1038/s41592-022-01604-1.
63. Nichterwitz, S. *et al.* Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nature Communications* vol. 7 Preprint at <https://doi.org/10.1038/ncomms12139> (2016).
64. Giladi, A. *et al.* Dissecting cellular crosstalk by sequencing physically interacting cells. *Nat. Biotechnol.* **38**, 629–637 (2020).
65. Halpern, K. B. *et al.* Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nat. Biotechnol.* **36**, 962–970 (2018).
66. Manco, R. *et al.* Clump sequencing exposes the spatial expression programs of intestinal secretory cells. *Nat. Commun.* **12**, 3074 (2021).
67. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* vol. 33 495–502 Preprint at <https://doi.org/10.1038/nbt.3192> (2015).

68. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
69. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).
70. Moriel, N. *et al.* NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat. Protoc.* **16**, 4177–4200 (2021).
71. Maseda, F., Cang, Z. & Nie, Q. DEEPsc: A Deep Learning-Based Map Connecting Single-Cell Transcriptomics and Spatial Imaging Data. *Front. Genet.* **12**, 636743 (2021).
72. McKellar, D. W. *et al.* Spatial mapping of the total transcriptome by in situ polyadenylation. *Nat. Biotechnol.* 1–8 (2022).
73. Deng, Y. *et al.* Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level. *Science* **375**, 681–686 (2022).
74. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
75. Henikoff, S., Henikoff, J. G., Kaya-Okur, H. S. & Ahmad, K. Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *eLife* vol. 9 Preprint at <https://doi.org/10.7554/elife.63274> (2020).
76. Liu, Y. *et al.* Spatial-CITE-seq: spatially resolved high-plex protein and whole transcriptome co-mapping. *bioRxiv* 2022.04.01.486788 (2022) doi:10.1101/2022.04.01.486788.
77. Ben-Chetrit, N. *et al.* Integrated protein and transcriptome high-throughput spatial profiling. *bioRxiv* 2022.03.15.484516 (2022) doi:10.1101/2022.03.15.484516.
78. Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges. *Mol. Cell* **71**, 468–480 (2018).
79. Hildebrandt, F. *et al.* Spatial Transcriptomics to define transcriptional patterns of zonation and structural components in the mouse liver. *Nat. Commun.* **12**, 7046 (2021).
80. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
81. Danaher, P. *et al.* Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. *Nat. Commun.* **13**, 385 (2022).
82. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
83. Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* **40**, 74–85 (2022).
84. Qin, L. & Crawford, J. M. 1 - Anatomy and Cellular Functions of the Liver. in *Zakim and Boyer's Hepatology (Seventh Edition)* (eds. Sanyal, A. J., Boyer, T. D., Lindor, K. D. & Terrault, N. A.) 2-19.e4 (Elsevier, 2018).

85. Alberts, B. *et al.* *General Principles of Cell Communication*. (Garland Science, 2002).
86. Francis, K. & Palsson, B. O. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12258–12262 (1997).
87. Morán, G. A. G., Parra-Medina, R., Cardona, A. G., Quintero-Ronderos, P. & Rodríguez, É. G. *Cytokines, chemokines and growth factors*. (El Rosario University Press, 2013).
88. Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. S. Protein-protein interaction detection: methods and analysis. *Int. J. Proteomics* **2014**, 147648 (2014).
89. Zhou, Y. *et al.* Evaluation of Single-Cell Cytokine Secretion and Cell-Cell Interactions with a Hierarchical Loading Microwell Chip. *Cell Rep.* **31**, 107574 (2020).
90. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
91. Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).
92. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
93. Noël, F. *et al.* Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* **12**, 1089 (2021).
94. Wang, Y. *et al.* iTALK: an R Package to Characterize and Illustrate Intercellular Communication. *bioRxiv* 507871 (2019) doi:10.1101/507871.
95. Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
96. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
97. Garcia-Alonso, L. *et al.* Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711 (2021).
98. Guilliams, M. *et al.* Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379-396.e38 (2022).
99. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).
100. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).

101. Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-01467-z.
102. Gault, J. *et al.* Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nature Methods* vol. 17 505–508 Preprint at <https://doi.org/10.1038/s41592-020-0821-0> (2020).
103. Katzenelenbogen, Y. *et al.* Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer. *Cell* **182**, 872-885.e19 (2020).
104. Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 395–410 (2019).
105. Abdel-Misih, S. R. Z. & Bloomston, M. Liver anatomy. *Surg. Clin. North Am.* **90**, 643–653 (2010).
106. Eipel, C., Abshagen, K. & Vollmar, B. Regulation of hepatic blood flow: the hepatic arterial buffer response revisited. *World J. Gastroenterol.* **16**, 6046–6057 (2010).
107. Trefts, E., Gannon, M. & Wasserman, D. H. The liver. *Curr. Biol.* **27**, R1147–R1151 (2017).
108. Krishna, M. Microscopic anatomy of the liver. *Clin. Liver Dis.* **2**, S4–S7 (2013).
109. Hofmann, A. F. The continuing importance of bile acids in liver and intestinal disease. *Arch. Intern. Med.* **159**, 2647–2658 (1999).
110. Crispe, I. N. Hepatic T cells and liver tolerance. *Nature Reviews Immunology* vol. 3 51–62 Preprint at <https://doi.org/10.1038/nri981> (2003).
111. Gebhardt, R. Metabolic zonation of the liver: Regulation and implications for liver function. *Pharmacology & Therapeutics* vol. 53 275–354 Preprint at [https://doi.org/10.1016/0163-7258\(92\)90055-5](https://doi.org/10.1016/0163-7258(92)90055-5) (1992).
112. McClelland, R., Dennis, R., Reid, L. M., Palsson, B. & Macdonald, J. M. 7 - TISSUE ENGINEERING. in *Introduction to Biomedical Engineering (Second Edition)* (eds. Enderle, J. D., Blanchard, S. M. & Bronzino, J.) 313–402 (Academic Press, 2005).
113. Potente, M. & Mäkinen, T. Vascular heterogeneity and specialization in development and disease. *Nat. Rev. Mol. Cell Biol.* **18**, 477–494 (2017).
114. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764-779.e20 (2020).
115. Ben-Moshe, S. *et al.* The spatiotemporal program of zonal liver regeneration following acute injury. *Cell Stem Cell* **29**, 973-989.e10 (2022).
116. Torre, C., Perret, C. & Colnot, S. Molecular determinants of liver zonation. *Prog. Mol. Biol. Transl. Sci.* **97**, 127–150 (2010).
117. Braet, F. & Wisse, E. Structural and functional aspects of liver sinusoidal endothelial cell fenestrae: a review. *Comp. Hepatol.* **1**, 1 (2002).
118. Shetty, S., Lalor, P. F. & Adams, D. H. Liver sinusoidal endothelial cells - gatekeepers of hepatic immunity. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 555–567 (2018).

119. Lohse, A. W. *et al.* Antigen-presenting function and B7 expression of murine sinusoidal endothelial cells and Kupffer cells. *Gastroenterology* **110**, 1175–1181 (1996).
120. Rogoff, T. M. & Lipsky, P. E. Antigen presentation by isolated guinea pig Kupffer cells. *J. Immunol.* **124**, 1740–1744 (1980).
121. Janeway, C. A., Jr, Travers, P., Walport, M. & Shlomchik, M. J. *Antigen Presentation to T Lymphocytes*. (Garland Science, 2001).
122. Shi, J., Fujieda, H., Kokubo, Y. & Wake, K. Apoptosis of neutrophils and their elimination by Kupffer cells in rat liver. *Hepatology* **24**, 1256–1263 (1996).
123. Filice, G. A. Antimicrobial properties of Kupffer cells. *Infect. Immun.* **56**, 1430–1435 (1988).
124. Okada, T. *et al.* Liver Resident Macrophages (Kupffer Cells) Share Several Functional Antigens in Common with Endothelial Cells. *Scand. J. Immunol.* **83**, 139–150 (2016).
125. Roland, C. R., Walp, L., Stack, R. M. & Flye, M. W. Outcome of Kupffer cell antigen presentation to a cloned murine Th1 lymphocyte depends on the inducibility of nitric oxide synthase by IFN-gamma. *J. Immunol.* **153**, 5453–5464 (1994).
126. Wen, Y., Lambrecht, J., Ju, C. & Tacke, F. Hepatic macrophages in liver homeostasis and diseases-diversity, plasticity and therapeutic opportunities. *Cell. Mol. Immunol.* **18**, 45–56 (2021).
127. Revel, M., Sautès-Fridman, C., Fridman, W.-H. & Roumenina, L. T. C1q macrophages: passengers or drivers of cancer progression. *Trends in Cancer* vol. 8 517–526 Preprint at <https://doi.org/10.1016/j.trecan.2022.02.006> (2022).
128. Crispe, I. N. & Mehal, W. Z. Strange brew: T cells in the liver. *Immunol. Today* **17**, 522–525 (1996).
129. Norris, S. *et al.* Resident human hepatitis lymphocytes are phenotypically different from circulating lymphocytes. *Journal of Hepatology* vol. 28 84–90 Preprint at [https://doi.org/10.1016/s0168-8278\(98\)80206-7](https://doi.org/10.1016/s0168-8278(98)80206-7) (1998).
130. Peng, H., Wisse, E. & Tian, Z. Liver natural killer cells: subsets and roles in liver immunity. *Cell. Mol. Immunol.* **13**, 328–336 (2016).
131. Bandyopadhyay, K., Marrero, I. & Kumar, V. NKT cell subsets as key participants in liver physiology and pathology. *Cell. Mol. Immunol.* **13**, 337–346 (2016).
132. Vicari, A. P., Mocci, S., Openshaw, P., O'Garra, A. & Zlotnik, A. Mouse gamma delta TCR+NK1.1+ thymocytes specifically produce interleukin-4, are major histocompatibility complex class I independent, and are developmentally related to alpha beta TCR+NK1.1+ thymocytes. *Eur. J. Immunol.* **26**, 1424–1429 (1996).
133. Kater, J. M. Comparative and experimental studies on the cytology of the liver. *Zeitschrift für Zellforschung und Mikroskopische Anatomie* **17**, 217–246 (1933).
134. Chiquoine, A. D. The distribution of glucose-6-phosphatase in the liver and kidney of the mouse. *J. Histochem. Cytochem.* **1**, 429–435 (1953).

135. Jungermann, K., Heilbronn, R., Katz, N. & Sasse, D. The glucose/glucose-6-phosphate cycle in the periportal and perivenous zone of rat liver. *Eur. J. Biochem.* **123**, 429–436 (1982).
136. Jungermann, K. & Katz, N. Functional specialization of different hepatocyte populations. *Physiol. Rev.* **69**, 708–764 (1989).
137. Andersen, B., Zierz, S. & Jungermann, K. Alteration in zonation of succinate dehydrogenase, phosphoenolpyruvate carboxykinase and glucose-6-phosphatase in regenerating rat liver. *Histochemistry* **80**, 97–101 (1984).
138. Sasse, D., Katz, N. & Jungermann, K. Functional heterogeneity of rat liver parenchyma and of isolated hepatocytes. *FEBS Lett.* **57**, 83–88 (1975).
139. Nauck, M., Wölfle, D., Katz, N. & Jungermann, K. Modulation of the glucagon-dependent induction of phosphoenolpyruvate carboxykinase and tyrosine aminotransferase by arterial and venous oxygen concentrations in hepatocyte cultures. *Eur. J. Biochem.* **119**, 657–661 (1981).
140. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2(+) cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
141. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).
142. Notenboom, R. G., Moorman, A. F. & Lamers, W. H. Developmental appearance of ammonia-metabolizing enzymes in prenatal murine liver. *Microsc. Res. Tech.* **39**, 413–423 (1997).
143. Shiojiri, N. *et al.* Heterogeneous hepatocellular expression of glutamine synthetase in developing mouse liver and in testicular transplants of fetal liver. *Lab. Invest.* **72**, 740–747 (1995).
144. Quistorff, B., Grunnet, N. & Cornell, N. W. Digitonin perfusion of rat liver. A new approach in the study of intra-acinar and intracellular compartmentation in the liver. *Biochem. J* **226**, 289–297 (1985).
145. Quistorff, B. & Grunnet, N. Dual-digitonin-pulse perfusion. Concurrent sampling of periportal and perivenous cytosol of rat liver for determination of metabolites and enzyme activities. *Biochem. J* **243**, 87–95 (1987).
146. Lindros, K. O. & Penttilä, K. E. Digitonin-collagenase perfusion for efficient separation of periportal or perivenous hepatocytes. *Biochem. J* **228**, 757–760 (1985).
147. Gola, A. *et al.* Commensal-driven immune zonation of the liver promotes host defence. *Nature* **589**, 131–136 (2021).
148. Williamson, T., Sultanpuram, N. & Sendi, H. The role of liver microenvironment in hepatic metastasis. *Clin. Transl. Med.* **8**, 21 (2019).
149. Hess, K. R. *et al.* Metastatic patterns in adenocarcinoma. *Cancer* **106**, 1624–1633 (2006).
150. Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M. & Wallace, M. B. Colorectal cancer. *Lancet* **394**, 1467–1480 (2019).
151. Kow, A. W. C. Hepatic metastasis from colorectal cancer. *J. Gastrointest. Oncol.* **10**, 1274–1298 (2019).

152. Ciner, A. T., Jones, K., Muschel, R. J. & Brodt, P. The unique immune microenvironment of liver metastases: Challenges and opportunities. *Semin. Cancer Biol.* **71**, 143–156 (2021).
153. Brodt, P. Role of the Microenvironment in Liver Metastasis: From Pre- to Prometastatic Niches. *Clin. Cancer Res.* **22**, 5971–5982 (2016).
154. Niu, Y., Yang, W., Qian, H. & Sun, Y. Intracellular and extracellular factors of colorectal cancer liver metastasis: a pivotal perplex to be fully elucidated. *Cancer Cell Int.* **22**, 341 (2022).
155. Oura, K., Morishita, A., Tani, J. & Masaki, T. Tumor Immune Microenvironment and Immunosuppressive Therapy in Hepatocellular Carcinoma: A Review. *Int. J. Mol. Sci.* **22**, (2021).
156. Braet, F. *et al.* The hepatic sinusoidal endothelial lining and colorectal liver metastases. *World J. Gastroenterol.* **13**, 821–825 (2007).
157. Van den Eynden, G. G. *et al.* The multifaceted role of the microenvironment in liver metastasis: biology and clinical implications. *Cancer Res.* **73**, 2031–2043 (2013).
158. Wen, S. W., Ager, E. I. & Christophi, C. Bimodal role of Kupffer cells during colorectal cancer liver metastasis. *Cancer Biol. Ther.* **14**, 606–613 (2013).
159. Knittel, T. *et al.* Expression patterns of matrix metalloproteinases and their inhibitors in parenchymal and non-parenchymal cells of rat liver: regulation by TNF- α and TGF- β 1. *Journal of Hepatology* vol. 30 48–60 Preprint at [https://doi.org/10.1016/s0168-8278\(99\)80007-5](https://doi.org/10.1016/s0168-8278(99)80007-5) (1999).
160. Wu, Y. *et al.* Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. *Cancer Discov.* **12**, 134–153 (2022).
161. Liu, Y. *et al.* Immune phenotypic linkage between colorectal cancer and liver metastasis. *Cancer Cell* **40**, 424–437.e5 (2022).
162. Yang, J. *et al.* Integrated genomic and transcriptomic analysis reveals unique characteristics of hepatic metastases and pro-metastatic role of complement C1q in pancreatic ductal adenocarcinoma. *Genome Biol.* **22**, 4 (2021).
163. Jiang, Y., Li, Y. & Zhu, B. T-cell exhaustion in the tumor microenvironment. *Cell Death Dis.* **6**, e1792–e1792 (2015).
164. Oh, B. Y., Hong, H. K., Lee, W. Y. & Cho, Y. B. Animal models of colorectal cancer with liver metastasis. *Cancer Lett.* **387**, 114–120 (2017).
165. Lee, W. Y., Hong, H. K., Ham, S. K., Kim, C. I. & Cho, Y. B. Comparison of colorectal cancer in differentially established liver metastasis models. *Anticancer Res.* **34**, 3321–3328 (2014).
166. Heijstek, M. W., Kranenburg, O. & Borel Rinkes, I. H. M. Mouse models of colorectal cancer and liver metastases. *Dig. Surg.* **22**, 16–25 (2005).
167. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
168. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).

169. Spector, A. A. Fatty acid binding to plasma albumin. *J. Lipid Res.* **16**, 165–179 (1975).
170. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* **7**, (2018).
171. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
172. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
173. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
174. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
175. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* **5**, 2122 (2016).
176. Amezquita, R. A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**, 137–145 (2020).
177. Wickham, H., François, R., Henry, L. & Müller, K. A Grammar of Data Manipulation [R package dplyr version 1.0.2]. <https://www.semanticscholar.org> › ...<https://www.semanticscholar.org> › ... (2020).
178. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
179. Wickham, H. ggplot2: elegant graphics for data analysis Springer-Verlag New York; 2009. Preprint at (2016).
180. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
181. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
182. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
183. Chen, Y., Lun, A. T. L. & Smyth, G. K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res.* **5**, 1438 (2016).
184. Lun, A. T. L., Richard, A. C. & Marioni, J. C. Testing for differential abundance in mass cytometry data. *Nat. Methods* **14**, 707–709 (2017).
185. Sato, T. & Clevers, H. Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. *Science* **340**, 1190–1194 (2013).
186. Niehues, S. M. *et al.* Liver volume measurement: reason of the difference between in vivo CT-volumetry and intraoperative ex vivo determination and how to cope it. *Eur. J. Med. Res.* **15**, 345–350 (2010).
187. Sohlenius-Sternbeck, A.-K. Determination of the hepatocellularity number for human, dog, rabbit, rat and mouse livers from protein concentration measurements. *Toxicol. In Vitro* **20**, 1582–1586 (2006).
188. Koo, B.-K. *et al.* Controlled gene expression in primary Lgr5 organoid cultures. *Nat. Methods* **9**, 81–83 (2011).

189. Wang, Z., Yang, S., Koga, Y., Corbett, S. E. & Johnson, W. E. Celda: A Bayesian model to perform co-clustering of genes into modules and cells into subpopulations using single-cell RNA-seq data. *bioRxiv* (2021).
190. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e20 (2019).
191. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).
192. Cunningham, M. F., Docherty, N. G., Burke, J. P. & O'Connell, P. R. S100A4 expression is increased in stricture fibroblasts from patients with fibrostenosing Crohn's disease and promotes intestinal fibroblast migration. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **299**, G457–G466 (2010).
193. Dong, M. *et al.* SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief. Bioinform.* **22**, 416–427 (2021).
194. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
195. Csardi & Nepusz. The igraph software package for complex network research. *InterJournal, complex systems*.
196. Phipson, B. & Smyth, G. K. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, Article39 (2010).
197. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016).
198. Inverso, D. *et al.* A spatial vascular transcriptomic, proteomic, and phosphoproteomic atlas unveils an angiocrine Tie–Wnt signaling axis in the liver. *Dev. Cell* **56**, 1677-1693.e10 (2021).
199. He, J. *et al.* Lhx6 regulates canonical Wnt signaling to control the fate of mesenchymal progenitor cells during mouse molar root patterning. *PLoS Genet.* **17**, e1009320 (2021).
200. Ibrahim, A. & Abumrad, N. A. Role of CD36 in membrane transport of long-chain fatty acids. *Curr. Opin. Clin. Nutr. Metab. Care* **5**, 139–145 (2002).
201. Schleicher, J. *et al.* Zonation of hepatic fatty acid metabolism - The diversity of its regulation and the benefit of modeling. *Biochim. Biophys. Acta* **1851**, 641–656 (2015).
202. Schleicher, J., Dahmen, U., Guthke, R. & Schuster, S. Zonation of hepatic fat accumulation: insights from mathematical modelling of nutrient gradients and fatty acid uptake. *J. R. Soc. Interface* **14**, (2017).
203. Kong, D.-H., Kim, Y. K., Kim, M. R., Jang, J. H. & Lee, S. Emerging Roles of Vascular Cell Adhesion Molecule-1 (VCAM-1) in Immunological Disorders and Cancer. *Int. J. Mol. Sci.* **19**, (2018).
204. Derosa, G. & Maffioli, P. Vascular Cell Adhesion Molecule-1 (VCAM-1) Expression in Liver Disease. in *Biomarkers in Liver Disease* (ed. Preedy, V. R.) 1–11 (Springer Netherlands, 2016).

205. Fu, H., Ward, E. J. & Marelli-Berg, F. M. Mechanisms of T cell organotropism. *Cell. Mol. Life Sci.* **73**, 3009–3033 (2016).
206. Chaudhry, S., Emond, J. & Griesemer, A. Immune Cell Trafficking to the Liver. *Transplantation* **103**, 1323–1337 (2019).
207. Peng, P. *et al.* TGFBI secreted by tumor-associated macrophages promotes glioblastoma stem cell-driven tumor growth via integrin $\alpha\beta 5$ -Src-Stat3 signaling. *Theranostics* **12**, 4221–4236 (2022).
208. Fico, F. & Santamaria-Martínez, A. TGFBI modulates tumour hypoxia and promotes breast cancer metastasis. *Mol. Oncol.* **14**, 3198–3210 (2020).
209. Steitz, A. M. *et al.* Tumor-associated macrophages promote ovarian cancer cell migration by secreting transforming growth factor beta induced (TGFBI) and tenascin C. *Cell Death Dis.* **11**, 1–15 (2020).
210. Lecker, L. S. M. *et al.* TGFBI Production by Macrophages Contributes to an Immunosuppressive Microenvironment in Ovarian Cancer. *Cancer Res.* **81**, 5706–5719 (2021).
211. Wei, T. *et al.* The Significance of Secreted Phosphoprotein 1 in Multiple Human Cancers. *Front Mol Biosci* **7**, 565383 (2020).
212. Wang, H. *et al.* FN1 is a prognostic biomarker and correlated with immune infiltrates in gastric cancers. *Frontiers in Oncology* vol. 12 Preprint at <https://doi.org/10.3389/fonc.2022.918719> (2022).
213. Zhang, X.-X., Luo, J.-H. & Wu, L.-Q. FN1 overexpression is correlated with unfavorable prognosis and immune infiltrates in breast cancer. *Front. Genet.* **13**, 913659 (2022).
214. Klement, J. D. *et al.* An osteopontin/CD44 immune checkpoint controls CD8⁺ T cell activation and tumor immune evasion. *J. Clin. Invest.* **128**, 5549–5560 (2018).
215. Zhang, L. *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **181**, 442-459.e29 (2020).
216. Groom, J. R. & Luster, A. D. CXCR3 in T cell function. *Exp. Cell Res.* **317**, 620–631 (2011).
217. Cesta, M. F. Normal structure, function, and histology of the spleen. *Toxicol. Pathol.* **34**, 455–465 (2006).
218. Bronte, V. & Pittet, M. J. The spleen in local and systemic regulation of immunity. *Immunity* **39**, 806–818 (2013).
219. Roda, G. *et al.* Crohn's disease. *Nature Reviews Disease Primers* **6**, 1–19 (2020).
220. Nayar, S. & Cho, J. H. From single-target to cellular niche targeting in Crohn's disease: intercepting bad communications. *EBioMedicine* **74**, 103690 (2021).
221. Bonecchi, R. & Graham, G. J. Atypical Chemokine Receptors and Their Roles in the Resolution of the Inflammatory Response. *Front. Immunol.* **7**, 224 (2016).
222. Schutysse, E., Struyf, S. & Van Damme, J. The CC chemokine CCL20 and its receptor CCR6. *Cytokine Growth Factor Rev.* **14**, 409–426 (2003).

223. Clark, I. C. *et al.* Microfluidics-free single-cell genomics with templated emulsification. *bioRxiv* 2022.06.10.495582 (2022) doi:10.1101/2022.06.10.495582.
224. Nischal, U., Nischal Kc & Khopkar, U. Techniques of skin biopsy and practical considerations. *J. Cutan. Aesthet. Surg.* **1**, 107–111 (2008).
225. Gurtner, A. *et al.* Single-cell RNA sequencing unveils intestinal eosinophil development and specialization. *bioRxiv* 2021.10.27.466053 (2021) doi:10.1101/2021.10.27.466053.
226. Grindberg, R. V. *et al.* RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences* **110**, 19802–19807 (2013).
227. Fus-Kujawa, A. *et al.* An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Front Bioeng Biotechnol* **9**, 701031 (2021).
228. Greenhalgh, S. N., Conroy, K. P. & Henderson, N. C. Cre-activity in the liver: transgenic approaches to targeting hepatic nonparenchymal cells. *Hepatology* **61**, 2091–2099 (2015).
229. Ombrato, L. *et al.* Metastatic-niche labelling reveals parenchymal cells with stem features. *Nature* **572**, 603–608 (2019).
230. Dobie, R. *et al.* Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme during Liver Fibrosis. *Cell Rep.* **29**, 1832-1847.e8 (2019).
231. Spivia, W., Magno, P. S., Le, P. & Fraser, D. A. Complement protein C1q promotes macrophage anti-inflammatory M2-like polarization during the clearance of atherogenic lipoproteins. *Inflamm. Res.* **63**, 885–893 (2014).
232. Kratofil, R. M., Kubes, P. & Deniset, J. F. Monocyte Conversion During Inflammation and Injury. *Arterioscler. Thromb. Vasc. Biol.* **37**, 35–42 (2017).
233. Thomas, G., Tacke, R., Hedrick, C. C. & Hanna, R. N. Nonclassical patrolling monocyte function in the vasculature. *Arterioscler. Thromb. Vasc. Biol.* **35**, 1306–1316 (2015).
234. Pawelec, G. Is There a Positive Side to T Cell Exhaustion? *Front. Immunol.* **10**, 111 (2019).
235. Qi, J. *et al.* Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. *Nat. Commun.* **13**, 1742 (2022).
236. Huang, H. *et al.* Cell-cell contact-induced gene editing/activation in mammalian cells using a synNotch-CRISPR/Cas9 system. *Protein Cell* **11**, 299–303 (2020).
237. Zhu, J., Cao, J., Liesz, A. & Roth, S. A macrophage-T cell coculture model for severe tissue injury-induced T cell death. *STAR Protoc* **2**, 100983 (2021).
238. Dokka, S., Toledo, D., Shi, X., Ye, J. & Rojanasakul, Y. High-efficiency gene transfection of macrophages by lipoplexes. *Int. J. Pharm.* **206**, 97–104 (2000).
239. Alfredsson, J. & Wick, M. J. Mechanism of fibrosis and stricture formation in Crohn's disease. *Scand. J. Immunol.* **92**, e12990 (2020).

240. De Simone, M., Rossetti, G. & Pagani, M. Single Cell T Cell Receptor Sequencing: Techniques and Future Challenges. *Front. Immunol.* **9**, 1638 (2018).
241. Nadelmann, E. R. *et al.* Isolation of Nuclei from Mammalian Cells and Tissues for Single-Nucleus Molecular Profiling. *Curr Protoc* **1**, e132 (2021).
242. Nabbi, A. & Riabowol, K. Rapid Isolation of Nuclei from Cells In Vitro. *Cold Spring Harb. Protoc.* **2015**, 769–772 (2015).
243. Annunziata, R. *et al.* An optimised method for intact nuclei isolation from diatoms. *Sci. Rep.* **11**, 1681 (2021).
244. Harayama, T. & Riezman, H. Understanding the diversity of membrane lipid composition. *Nat. Rev. Mol. Cell Biol.* **19**, 281–296 (2018).
245. Stoneburg, C. A. LIPIDS OF THE CELL NUCLEI. *J. Biol. Chem.* **129**, 189–196 (1939).

12 Acknowledgments

I would like to take the opportunity to thank everyone that helped me along the way to finish my Ph.D. project.

First of all, I would like to thank my Ph.D. supervisor Prof. Dr. Andreas Moor. Thank you, Andreas, for giving me the opportunity to work on this exciting project. For taking me as your first Ph.D. student and for always motivating me along the way if I had doubts that this project will even work. It has been very interesting and sometimes challenging 4 years seeing the lab expand and grow.

I would also like to thank my Ph.D. committee members, Prof. Dr. Barbara Treutlein and Prof. Dr. Roser Vento-Tormo for their valuable inputs during my committee meeting.

I would also like to thank the whole Moor lab for providing such a supportive and fun work environment. A special thanks to Karsten Bach, who helped with my computational analysis and reviewing my code. Thanks to Erkin Acar, for taking care of our computing cluster and R Studio Server and for always responding within minutes to help if something is not working properly. Thanks to Costanza Borrelli for providing metastatic mouse livers, showing me how to handle organoid cultures, and for the fun of doing spontaneous experiments. Thanks to Davide Eletto for managing the lab, for the always smooth process in ordering reagents, and for organizing the best lab retreat on Ischia. To Farah Mahmedi for organizing the lab and building such a warm atmosphere by organizing birthday presents and social events like the Silvesterlauf. To Dr. Atefeh Lafzi, for discussions and helping with computational problems. To Jonas Kretz for showing me how to use Visium. A special thanks to Xenia Ficht, for proofreading my thesis. And thanks to all the others for interesting discussions, fun lab retreats, and other fun activities in and outside of the lab.

A special thanks to the first generation of Moor lab members (Minkyong Lee, Karelia Velez, Jonas Kretzt, and Costanza Borrelli). Your company made the start of my Ph.D. especially fun.

I would also like to thank the IMCR Institute at UZH in Zürich for providing such a welcoming atmosphere when I started my Ph.D.

I would also like to thank Erich Brunner and Agnieszka Wengi for helping me to establish sorting with the Biosorter, and for always being available when I run into problems during sorting.

Finally, I would like to thank my family who has always supported me, for leaving me the freedom to follow my interests no matter where around the globe, and for providing a loving home that is always nice to return to, to take some relaxing time off. And I would also like to thank my friends who made this Ph.D. time a much more enjoyable experience.