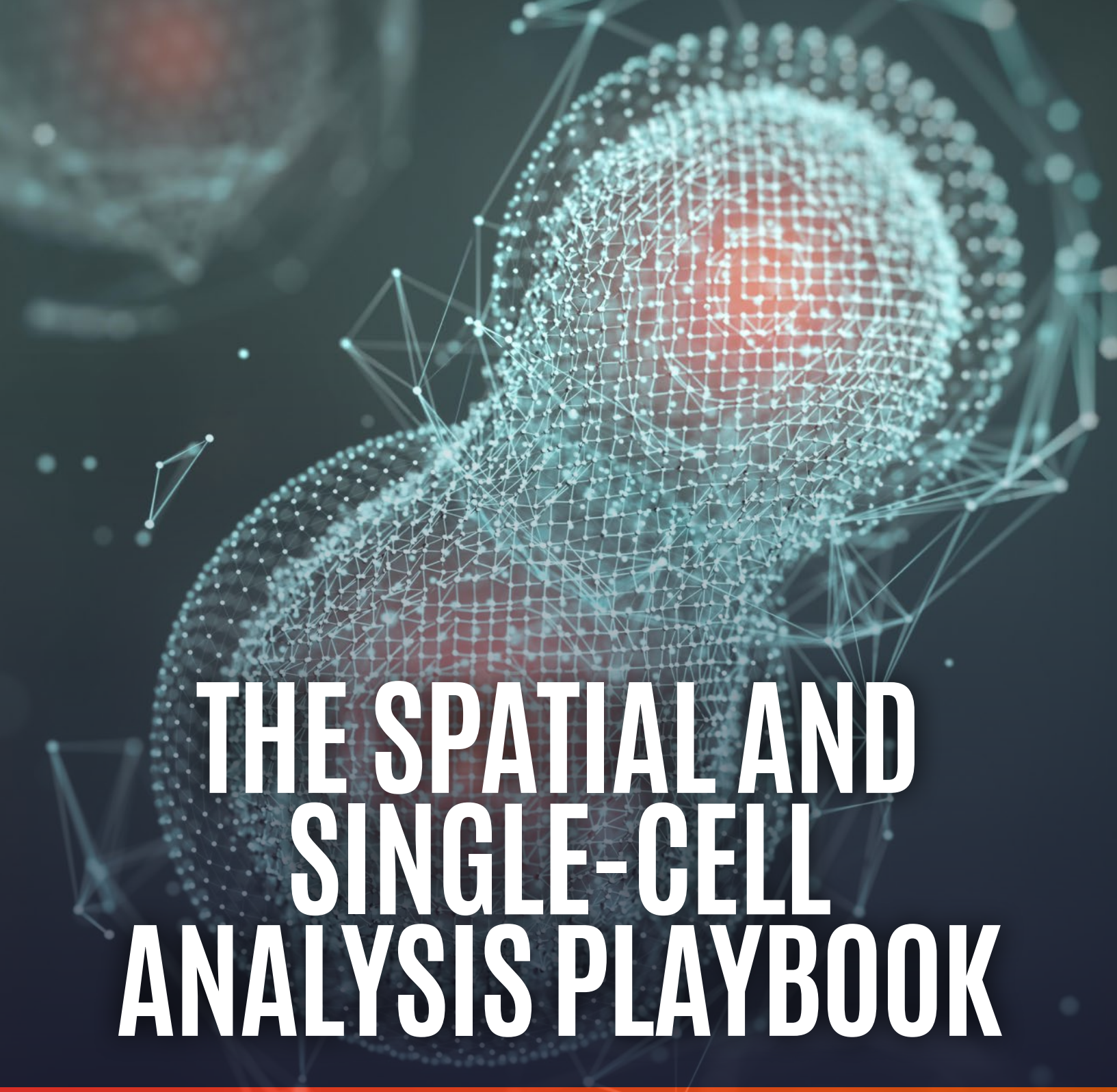


THE SPATIAL AND SINGLE-CELL ANALYSIS PLAYBOOK

PRODUCED BY:

Front Line Genomics

SPONSORED BY

10x GENOMICS

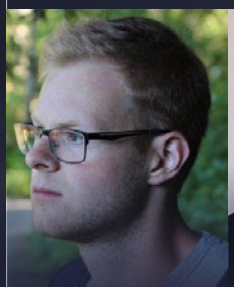
BD

canopy BIOSCIENCES A BRUKER COMPANY

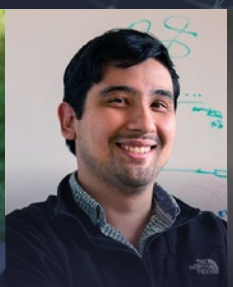
Miltenyi Biotec

SINGLE CELL DISCOVERIES

CONTRIBUTORS

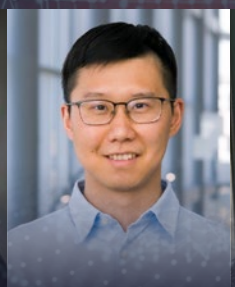


Robin Browaeys
Team Leader – Bio-IT Support
VIB Centre for Inflammation Research, Ghent University



Jason D. Buenrostro
Associate Professor & Broad Institute Member
Harvard University & Broad Institute of MIT and Harvard

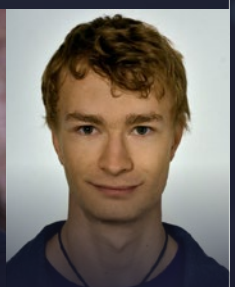
Jared K. Burks
Professor & Co-Director, Flow Cytometry & Cell Imaging Core Facility
The University of Texas MD Anderson Cancer Centre



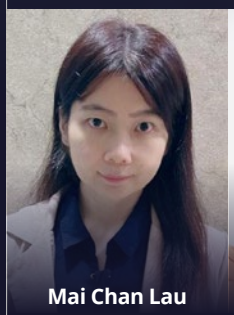
Haiqi Chen
Assistant Professor
UT Southwestern Medical Centre

Yanxiang Deng
Assistant Professor
University of Pennsylvania

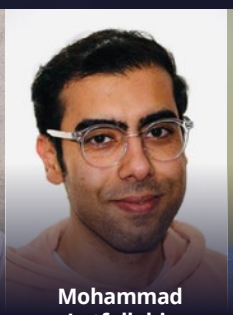
Silvia Domcke
Affiliate Assistant Professor, Dept. of Genome Sciences
University of Washington
Associate Director, Head of Human Genomics
Gordian Biotechnology



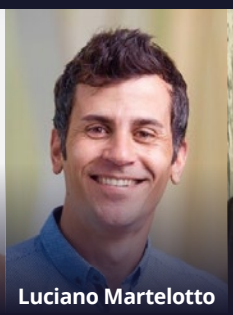
Lukas Heumos
PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich



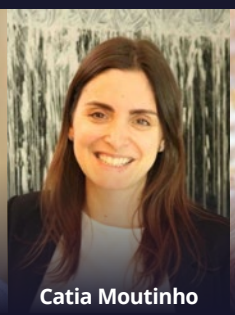
Mai Chan Lau
Assistant Principal Investigator
A*STAR's Bioinformatics Institute (BII) and Singapore Immunology Network (SIgN)



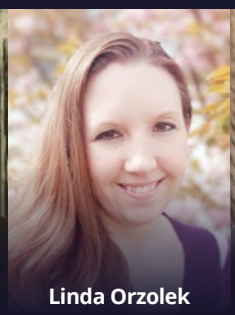
Mohammad Lotfollahi
Scientist
Helmholtz Munich/ Wellcome Sanger institute
Director of Machine Learning
Relation Therapeutics



Luciano Martelotto
Associate Professor & Lab Head, Single Cell and Spatial-Omics Laboratory
University of Adelaide, Australia

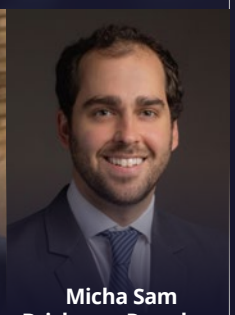


Catia Moutinho
Founder & Scientific Advisor
The Single-Cell World



Linda Orzolek
Director, Single Cell & Transcriptomics Core
Johns Hopkins University

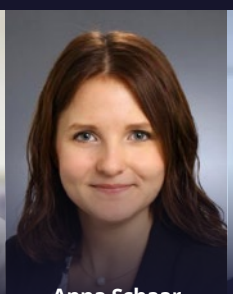
Tancredi Massimo Pentimalli
PhD Student, Nikolaus Rajewsky Lab
Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Centrum (MDC)
Berlin School of Integrative Oncology (BSIO), Charité – Universitätsmedizin Berlin



Micha Sam Brickman Raredon
Research Group Leader, Departments of Anaesthesiology, Pulmonary, Critical Care & Sleep Medicine and Immunobiology
Yale School of Medicine

Andrew Russell
Postdoctoral Fellow, Fei Chen Lab
Broad Institute of MIT and Harvard

Azucena Salas
Inflammatory Bowel Disease Group leader
Fundació de Recerca Clínic Barcelona-IDIBAPS, Hospital Clínic Barcelona



Anna Schaar
PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

Denis Schapiro
Research Group Leader
Heidelberg University Hospital

Emily Stephenson
Senior Research Associate, Haffina Lab
Newcastle University

Jovan Tanevski
Research Area Leader, Saez-Rodriguez Group
Institute for Computational Biomedicine, Heidelberg University and Heidelberg University Hospital

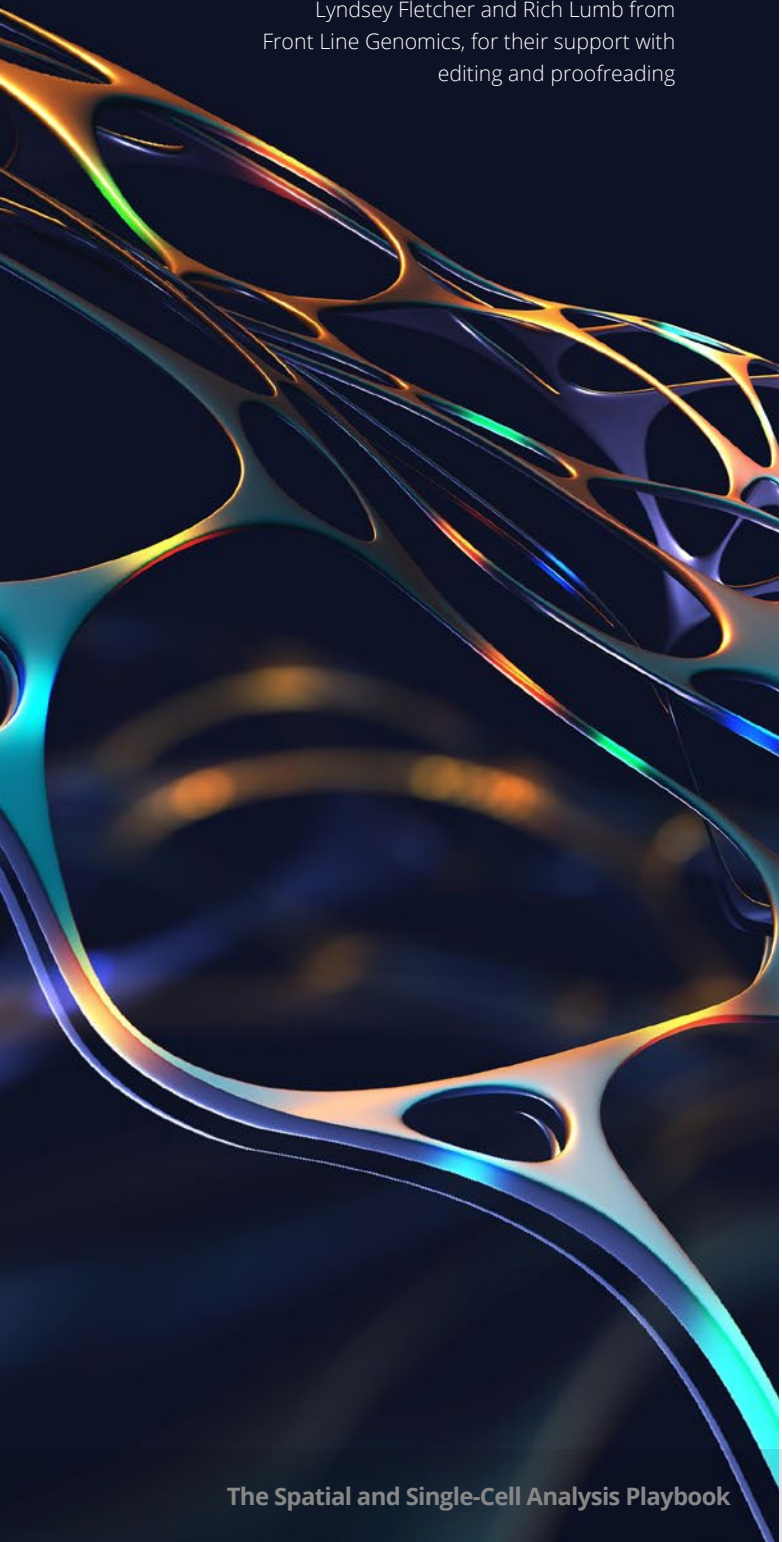
Fan Zhang
Assistant Professor
University of Colorado School of Medicine

FOREWORD

Matt Higgs

Science Writer
Front Line Genomics

Team Acknowledgements:
Lyndsey Fletcher and Rich Lumb from Front Line Genomics, for their support with editing and proofreading



THESE ARE EXCITING TIMES IN THE WORLD OF SPATIAL AND SINGLE-CELL BIOLOGY. DEVELOPMENT APPEARS TO BE HAPPENING ON ALL FRONTS. WHILE SINGLE-CELL TRANSCRIPTOMICS IS BEGINNING TO MATURE INTO A RELIABLE, STANDARDISED TECHNOLOGY, OPTIONS IN SPATIAL BIOLOGY ARE MULTIPLYING.

In this playbook, we cover a breadth of topics in this space. Within, you will find an overview of many of the commercial technologies and computational tools that are available to you, which could improve and assist your work. We also cover some specific applications of these tools, such as: how to get the most out of multiplexed and multimodal data, an overview of cell microenvironment and cell-cell communication methods and a deep dive into single-cell and spatial epigenomics methodologies.

While we cannot claim to have covered every nook and cranny, this playbook provides an up-to-date overview spatial and single-cell analysis for 2023.

Furthermore, by interviewing a series of experts in the field, we have gained unique insights and guidance, which have shaped this playbook. Excerpts from our discussions with these experts are found throughout the chapters. Within them, you will find advice on how to get the most out of specific tools, hard fought wisdom gained from working with these technologies and designing new tools, as well as perspectives and views on current topics in spatial and single-cell.

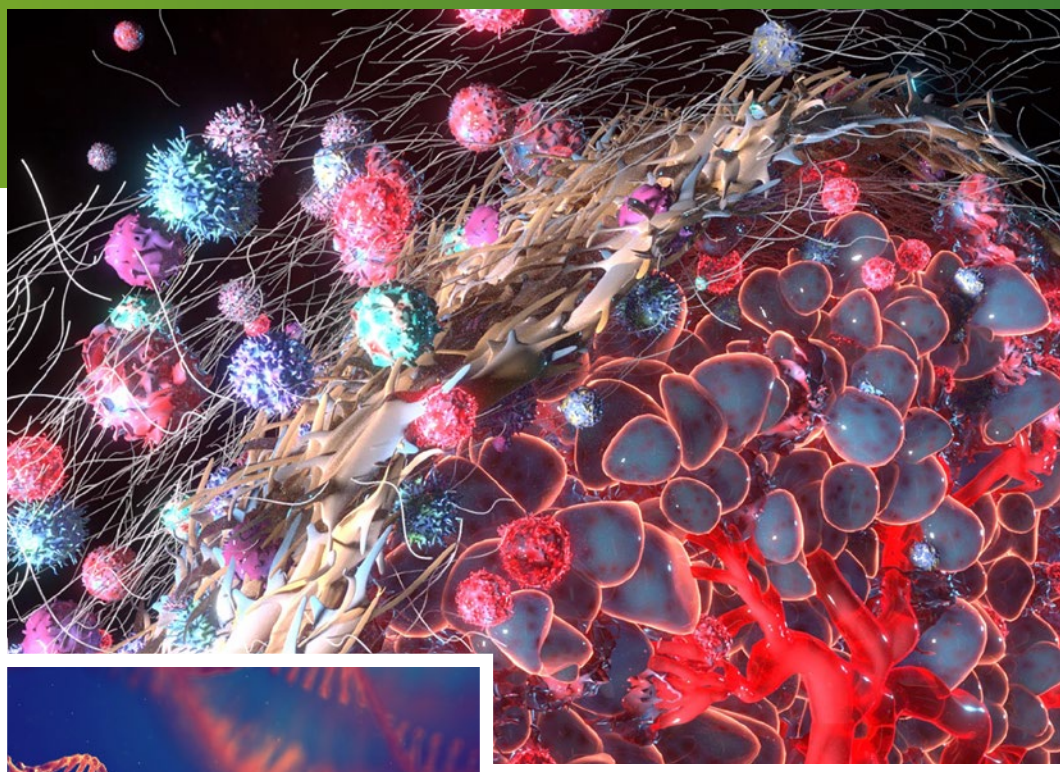
We would like to take this opportunity to thank all of our contributors for their time and insights with writing this playbook.

We would also like to thank the sponsors of this report, 10x Genomics, Canopy Biosciences, Miltenyi Biotec, BD Biosciences & Single Cell Discoveries

We hope you find this playbook a helpful resource.

Thank you for reading.

CONTENTS



5 **CHAPTER 1: LOOKING INTO THE PAST. SINGLE-CELL AND SPATIAL BIOLOGY: A HISTORY**

This first chapter begins by looking back over the history of single-cell and spatial analysis. We will cover several of the key developmental milestones for both single-cell and spatial technologies and look at attempts to merge these technologies to achieve the best of both worlds.

16 **CHAPTER 2: WHAT'S NEW. AN OVERVIEW OF THE LATEST COMMERCIAL TECHNOLOGY DEVELOPMENTS**

Innovation in single-cell and spatial biology often comes from commercial providers. This chapter will give an update on the recent technologies from the key players, highlighting where improvements have been made and the new opportunities these technologies offer.

23 **CHAPTER 3: SIZE MATTERS. SINGLE-CELL AND SPATIAL, MULTIPLEX AND MULTIOMICS**

More cells, more genes, more omics – single-cell and spatial biology has expanded. This chapter will cover some of the biggest initiatives in single-cell biology as well as the unique problems that arise from working with experiments at vast scale, specifically how to handle multiplex and multiomics spatial data.

35 **CHAPTER 4: FRIENDLY NEIGHBOURHOOD CELLS. CELL-CELL COMMUNICATION AND THE CELL MICROENVIRONMENT**

Spatial biology is now seen as key to understanding how complex tissues operate. In this chapter, we will look at the tissue microenvironment and how spatial biology is taking us beyond cell types to help us look at cell neighbourhoods and niches. We will also look at another burgeoning area of analysis, cell-cell communication.

51 **CHAPTER 5: BEYOND THE GENOME. SINGLE-CELL AND SPATIAL EPIGENOMICS**

Single-cell transcriptomics data is not enough to understand cellular identity. This chapter will cover advances in single-cell and spatial epigenetics and how up-to-date methods are allowing researchers to visualise epigenetic profiling alongside RNA expression.

63 **CHAPTER 6: BEDSIDE BENEFITS: SINGLE-CELL AND SPATIAL BIOLOGY IN CLINICAL PRACTICE**

The clinical benefits of single-cell and spatial technologies are clear but integrating them into routine practice is still challenging. This chapter will cover how single-cell methods can be translated to achieve a clinical benefit and the promise of spatial biology for precision medicine.

76 **CHAPTER 7: COMMUNITY DISCUSSIONS. AI, BENCHMARKING AND WHAT IS A CELL TYPE?**

Using expert contributions and insights, the penultimate chapter of the report will cover the hot topics in the single-cell and spatial space. What is the promise of AI? How should we define a cell type? How do we find a gold-standard for this plethora of methods?

91 **CHAPTER 8: GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL LOOK LIKE IN 15 YEARS' TIME?**

The rate of progress in the past 15 years of single-cell biology has been astonishing. Given that the 15-year anniversary of single-cell transcriptome sequencing is just around the corner, we thought we would challenge some experts in the field with a challenging question: What do you think single-cell and spatial methods will look like in 15 years' time?

LOOKING INTO THE PAST. SINGLE-CELL AND SPATIAL TRANSCRIPTOMICS: A HISTORY

THIS CHAPTER BEGINS BY LOOKING BACK OVER THE HISTORY OF SINGLE-CELL AND SPATIAL ANALYSIS. WE WILL COVER SEVERAL OF THE KEY DEVELOPMENTAL MILESTONES FOR BOTH SINGLE-CELL AND SPATIAL TECHNOLOGIES AND LOOK AT ATTEMPTS TO MERGE THESE TECHNOLOGIES TO ACHIEVE THE BEST OF BOTH WORLDS.

Deconstructing the smoothie

Organisms are complex entities, made up of many different tissues. Tissues themselves are also complex entities, made up of many different types of cell. Prior to the first single-cell methodologies, standard genomic practice used bulk tissue to try to understand the genomic and transcriptomic profiles of tissues[1].

A popular analogy has described this as the equivalent of trying to understand fruit by analysing a smoothie (Figure 1). You can learn a lot about fruit in general (i.e., the average abundance of banana) but trying to understand specific fruit becomes a problem when it's all blended together.

Single-cell sequencing is the equivalent of trying to learn about fruit from a fruit salad. With the right tools, you can separate out all the different types of fruit and learn about their differences. Like pulling apart a jigsaw, these tools allow us to identify the cells that make up tissues, which genes they express, what epigenetic alterations they may have and what proteins have been translated.

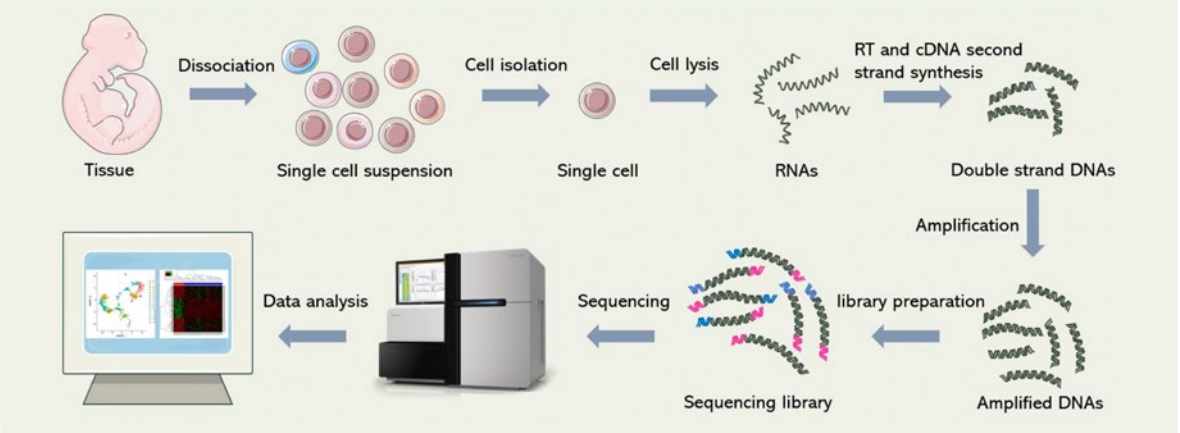
FIGURE 1: FRUIT ANALOGY OF SINGLE-CELL AND SPATIAL BIOLOGY.

Bulk sequencing takes data from all cells in a tissue at once, like sampling a smoothie. Single-cell sequencing allows data from each individual cell in a tissue to be analysed, like separating out fruit in a fruit salad. Spatial can acquire data from individual cells but maintains spatial context, like examining fruit in their proper places on a fruit tart. Images acquired from Canva.

Method of the year 2013, 2019 and 2020

Nature Methods has consistently recognized the power of single-cell and spatial technologies since their inception. In 2013, single-cell transcriptomics was awarded the method of the year with the hope that methods to sequence the DNA and RNA of single cells could transform many areas of biology and medicine[2]. This award was given at the turning point for single-cell methods, when commercial offerings and technological improvements (e.g., microfluidics and microwells) were making the adoption of single-cell methods easier and more efficient.

FIGURE 2: BASIC WORKFLOW OF SINGLE-CELL SEQUENCING.
Image Credit: Pan, et al.[3]



Single-cell technologies overcame the obscuring nature of bulk transcriptomics and allowed scientists to distinguish and characterize rare, but important, cell types from a tissue mass. Figure 2 shows the basic workflow of single-cell technology.

As we will cover below, our ability to sequence single cells has improved in throughput, consistency and cost. However, the analysis of genomics using a single-cell lens diversified in two further directions.

The first direction follows an appreciation that a cell cannot be understood purely from its genome or transcriptome. Instead, the epigenome, proteome and metabolome are needed to build a holistic picture of the molecular makeup of a cell at any one time. The resulting diversification of methods for single-cell multiomics was recognized by Nature Methods with the award of method of the year in 2019[4]. In that time, it has become clear that taking multimodal measurements (such as both RNA and protein) from cells often reveals more about molecular dynamics than the sum of its parts. It can be used to resolve cell populations with subtle transcriptomic differences and truly capture the molecular makeup of a tissue.

The second direction that single-cell transcriptomics has progressed in has been to reintegrate the spatial dimension. Spatially resolved transcriptomics picked up the method of the year award in the following year, 2020[5]. Spatial omics provides essential information about gene expression while retaining spatial context, i.e., which cells are neighbours or close-by and where a rare cell-type is situated. This context is a crucial dimension to all tissues, but has proven particularly valuable for tumour biology, neuroscience and developmental biology. Taking our previous analogy, if single-cell sequencing is the equivalent of analysing a fruit salad, them spatial transcriptomics is the equivalent of analysing a fruit tart. The fruit can be assessed in the same way as in the fruit salad, but this time, in their proper places and in relation to one-another. We can learn all the same things we can with single-cell sequencing (although, currently, with less sophistication) and also learn about the organisation of cells and interactions between them.

Today

We are now in 2023 and these technologies are still developing rapidly. It appears we are progressing towards a holistic methodology - a spatially-resolved, multimodal, single-cell assessment of tissues, or even whole organisms[6]. This would amount to single-cell biology being brought into complete cellular resolution. These data would allow us to visualize the cellular level changes at the onset of disease, infection or any other perturbation to the body.

We do not have this holistic methodology yet. Part of the reason is because these single-cell and spatial technologies still have their challenges. The rest of this report is focused on challenges and solutions from the now, and those we anticipate in the future.

However, it is always worth seeing how far we have come and examining the challenges from the past too. The remainder of this first chapter will outline a condensed overview of the history of these technologies and how the capabilities have expanded to the present day. With the first major single-cell transcriptomics study published in 2009[7], these technologies are approaching their 15th anniversary. So, strap yourself in for this whistlestop tour of single-cell and spatial biology.

2023: Almost 15 years of singularity

Single-cell sequencing has become a powerful tool for understanding complex biological systems[3,8-10]. The story formally began in 2009 with a publication in which scientists were manually isolating mouse blastomeres for sequencing, acquiring transcriptomic information from single cells[7]. From here, there was a gradual but significant advance in the number of cells that could be profiled. In 2011, the advent of cell-specific barcodes allowed for multiplexing and pooling, which meant 100s of cells could be sequenced using a method called STRT[11]. Fluidic circuits improved this further, and 2013 saw the release of the Fluidigm C1, the first single-cell automated prep system[12,13]. MARS-seq[14] in 2014 combined fluorescence activated cell sorting (FACS) and automatic liquid handling to substantially increase the throughput.

The introduction of droplet methods in 2015, inDrop[15] & Drop-seq[16], saw a leap in cell throughput into the 10s of 1000s. The release of 10x Genomics' Chromium in 2016 was pivotal, utilising this technology to create a benchtop option for all scientists. 2017 saw a new methodology emerge, a combinatorial indexing strategy known as sci-RNA-seq[17], currently in its third iteration[18], which, alongside SPLiT-seq[19], is pushing the heights of single-cell throughput into the 100,000s and 1,000,000s.

This advancement in single-cell throughput is proceeding faster the Moore's law[20], and if it continues at this pace it could be feasible that we could eventually see methods capable of sequencing the trillions of cells that make up one human body. Figures 3A and 4 show the trajectory of cell throughput capacity of single-cell technologies and the technologies that have driven this progression.

Alongside the advances in throughput, technologies such as Smart-seq[21,22] and CEL-seq[23,24] that were pioneered in 2012 have been improving the sensitivity of single-cell RNA-sequencing, allowing many more transcripts to be detected while sacrificing the high cell throughput. Both of these techniques have seen updates with the most recent, Smart-Seq[3], released in 2020[25], allowing full length transcript coverage combined with UMI counting strategy to offer the most sensitive single-cell RNA sequencing method.

The most recent single-cell transcriptomic approaches are shifting in two directions. One set of methods are moving away from cell isolation methods towards measuring full transcriptomes from cells in situ with spatial location information retained.

FIGURE 3: DEVELOPMENT OF SINGLE-CELL RNA SEQUENCING TECHNOLOGY.

The direction of progress has seen, (A) an increase in the number of analysed cells, (B) a radical reduction in cost per cell, and (C) an increase in the number of published papers. (D) The technological landmarks that have occurred to increase the number of cells and quality of information gathered from single cells. Image Credit: Jovic, et al.[8]

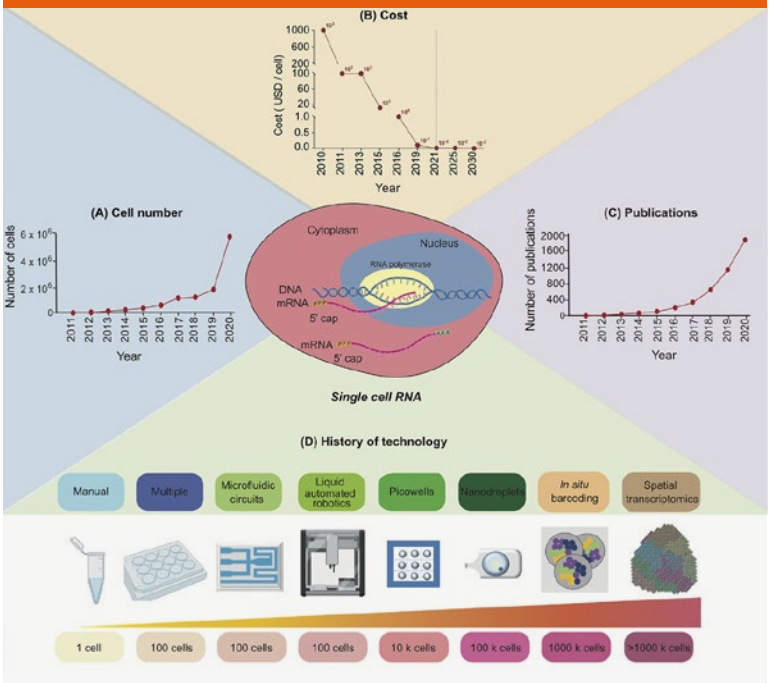
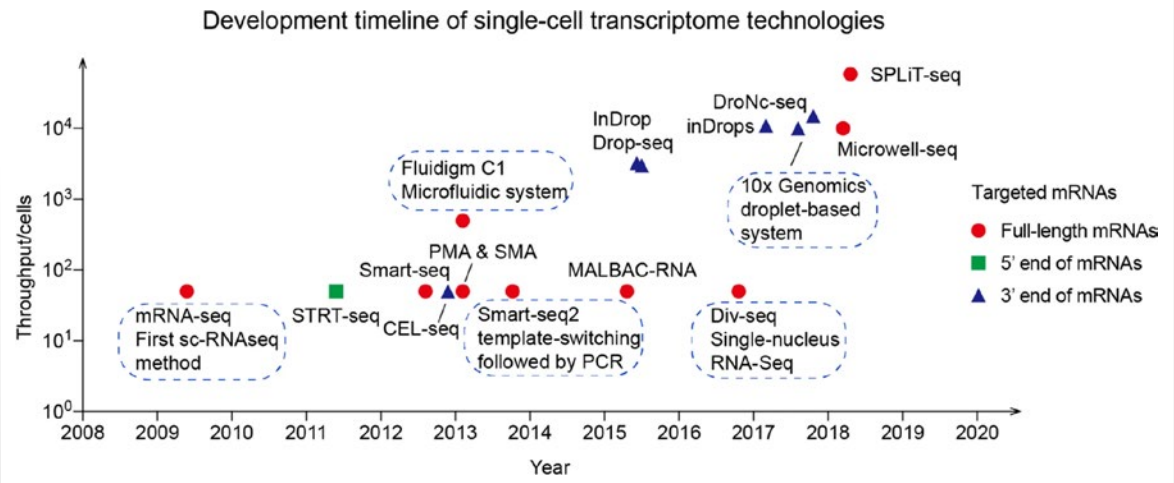


FIGURE 4. TIMELINE OF SINGLE-CELL TECHNOLOGICAL DEVELOPMENTS.

Technologies are displayed by original publication date and the typical throughput. Colours/shapes indicate the type of mRNA that the technology sequences. Significant advances are indicated by dotted borders. Image Credit: Zhang, et al. [28]



This will be covered in further detail in the following section covering spatial technology.

The other direction still involves isolated cells, but is diversifying away from the transcriptome to enable multiomics assessments. This could involve sequencing of the genome, proteome, epigenome, metabolome or multiple of these together[26,27].

Overall, while single-cell sequencing is now almost 15 years old and is maturing into a lab staple, it is still undergoing substantial development each year. We spoke to some experts in the field about what it has been like to witness this technological development from the early days of single-cell to now.

LUCIANO MARTELOTTO

Associate Professor & Lab Head, Single-Cell and Spatial-Omics Laboratory
University of Adelaide, Australia

FLG: We would like to get your perspective as someone who lived through this evolution of the field. What was it like?

Luciano: *It's been remarkable to witness the evolution. The first single-cell experiment was a sequencing of one cell in 2009. That marked the inception. Since then, the field has gradually progressed, moving from sequencing one cell to ten, then a hundred cells. With FACS sorting, it became possible to isolate a small number of cells for copy number profiling, followed by RNA-seq and similar analyses.*

Then there were two consecutive papers from Harvard and MIT. Interestingly, the researchers from these institutions were essentially thinking along the same lines, even though they were just meters apart. They each introduced variations of the same microfluidic technology, both of which became pivotal for subsequent developments.

These two papers laid the groundwork for what we now recognize as microfluidic devices and the Chromium Controller. Once the Chromium Controller was introduced, we saw an increase from thousands to hundreds of thousands of cells. Subsequently, as combinatorial barcoding techniques were adopted (e.g., SPLiT-Seq, now core in PARSE Bio tech), the count reached into the millions. It's remarkable that the technology advanced from a single cell to over a million in just 10 years.

LINDA D. ORZOLEK

Director, Single Cell & Transcriptomics Core
Johns Hopkins University

"I THINK IT'S AN AMAZINGLY EXCITING TIME IN MOLECULAR BIOLOGY, GENETICS, AND MEDICINE. AND IT'S BEEN A FUN RIDE FOR THE LAST 20 YEARS, I WILL DEFINITELY SAY."

FLG: Since you have been part of the transcriptomics core for almost 20 years, what has it been like to go through the evolution of these technologies?

Linda: *It's been really exciting, especially coming in as just an undergraduate without significant research experience and seeing the early days of microarrays. I was originally doing some of the old Clontech arrays, having to use radioactive labelling and things like that. Then seeing the Affymetrix system come out and what could be generated from that. As a naive kid, you're looking at it thinking - 'wow, we are at the pinnacle of understanding here, this is amazing, where do we even go from here?'*

Then, within about five to seven years, I'm being told that microarrays are going to go the way of the dodo, and sequencing is going to take over everything. At the beginning, I was thinking, the data analysis is complex, it's so expensive, there's no way this is going to happen. I had no idea why they were anticipating that within three years we're never going to do another array. Ultimately, of course they were right, but it took a lot longer than 3 years. It was much closer to a decade.

Since then, I've seen a lot of people get excited about new technologies and think that 'this is the pinnacle of everything,' just like I did with arrays. Some people jump into it because that's the new thing. In the past, we had a lot of people come in and ask for sequencing services and we would say, 'why do you need to do sequencing when a microarray will give you the same information cheaper and easier?' The response - 'because that's what people are doing now.' There wasn't a real scientific need for it.

Then it got to the point where sequencing was cheaper, and it did not require any a priori knowledge, meaning we could find novel transcripts. Once costs were down and analysis methods improved, we saw the transition. Then the next thing *comes out, single-cell, and we keep making these giant leaps in resolution of what we can analyze. It's always the same pattern; new technology comes out, certain people jump all over it, because it's the new technology, but they don't understand why. That's where we [the core team] need to be... to push the 'why', so that the right decision is made.*

Given how quickly the field is advancing, it's fun to think about what we are still missing. It is feasible that what we're still missing at this single-cell resolution, we're going to resolve in three to four years. I think it's an amazingly exciting time in molecular biology, genetics, and medicine. And it's been a fun ride for the last 20 years, I will definitely say.

FLG: Spatial methods are clearly developing rapidly, but what about single-cell, is it still maturing? Do you think it's still evolving just as quickly?

Linda: *I think it's still evolving just as quickly. I think there's a lot more that can be done to enable a lot more studies. We always see the limitations of having low cell viability, and clients still come in and say, 'but, can I just pick out the live cells?' Eventually, I believe that there's going to be a method where we incorporate this cell sorting into the processing, so you are only pulling out the live cells. We're starting to see that with some of these micro-fluidics assays that allow for image-based sorting, or these new AI based tools looking at cell morphology to sort out live cells.* ▶

The advantage is that these methods are gentler on the cells, preventing activation or other transcriptional changes that may bias data. We have people who have a 1 in a 1000, or a 1 in 10,000 cell type as their target, and those cells tend to get lost amongst everything. So, being able to find a way to really enrich for those particular cells, I think all of that is still there for development.

Then you couple it with the price reductions that need to continue, so that we can do these much more extensive studies, e.g., time courses, longitudinal studies. I was listening to a presentation earlier this week, talking about retina regeneration, and they were saying that they have done time points during development almost every day, and what a difference it made, and that they would miss the complete picture if they weren't able to do all those time points. But lot of people can't – they have to pick and choose very carefully what they can do, because of funding. As all the prices come down, it enables larger, more extensive projects. So, all of that is still developing, right alongside spatial.

2023: A spatial odyssey

Compared to single-cell sequencing, spatial transcriptomics has not reached the same level of maturity, despite a similar timeframe of development (see Figure 5). This is partly because current spatial technologies can be divided into four different categories, all of which are being developed. These include:

1. **Microdissection methods** – this 'brute-force' method involves dissecting out small regions of interest and sequencing them traditionally. This limits the size of area one can investigate. Methods include: LCM[29], TIVA[30], NICHE-Seq[31] and, most recently, ProximID[32].

2. **In situ Hybridisation (ISH) methods** – these methods visualise RNA directly in their environment through binding probes and fluorescent markers to RNA. A historical challenge has been the limitation on the number of markers in one experiment due to spectral overlap.

3. **In situ Sequencing (ISS) methods** – these methods perform RNA sequencing inside the cell while it remains in tissue context. Due to the spatial limits of the cell, there is a limit to the number of transcripts that can be discriminated simultaneously.

4. **In situ Capture (ISC) methods** – these methods capture transcripts in situ using barcodes, which historically limits RNA capture efficiency.

FIGURE 5: TIMELINE OF THE KEY DEVELOPMENTS IN SPATIAL TRANSCRIPTOMIC TECHNOLOGIES.

The principles of several exemplar technologies are demonstrated including: (A) ProximID, (B) seqFISH+, (C) Stereo-seq, (D) sci-Space, (E) STARmap, (F) 10x Genomics Visium; (G) Slide-seqV2 and (H) Seq-Scope. Image Credit: Zhang, et al. [33]

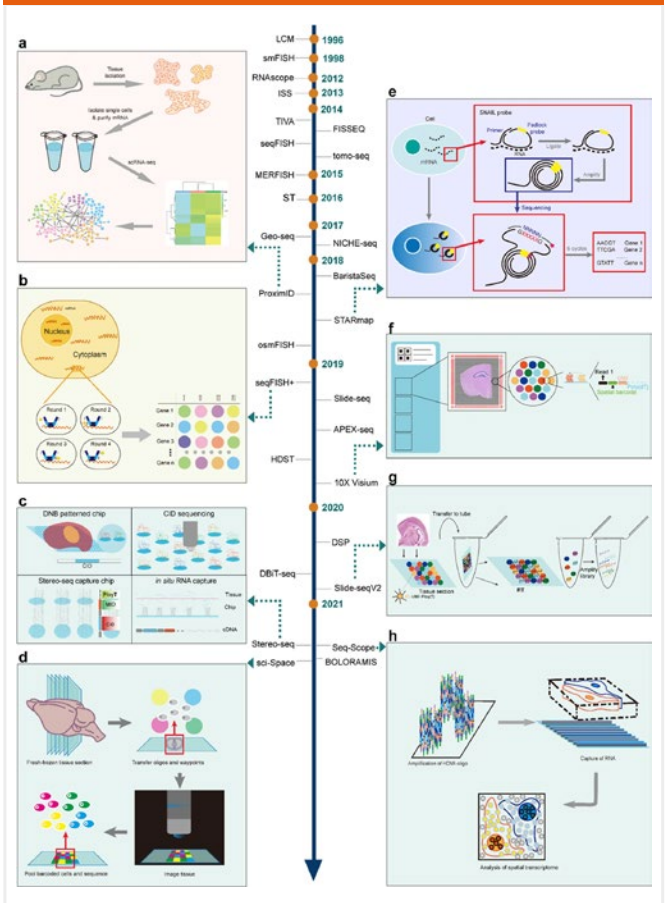


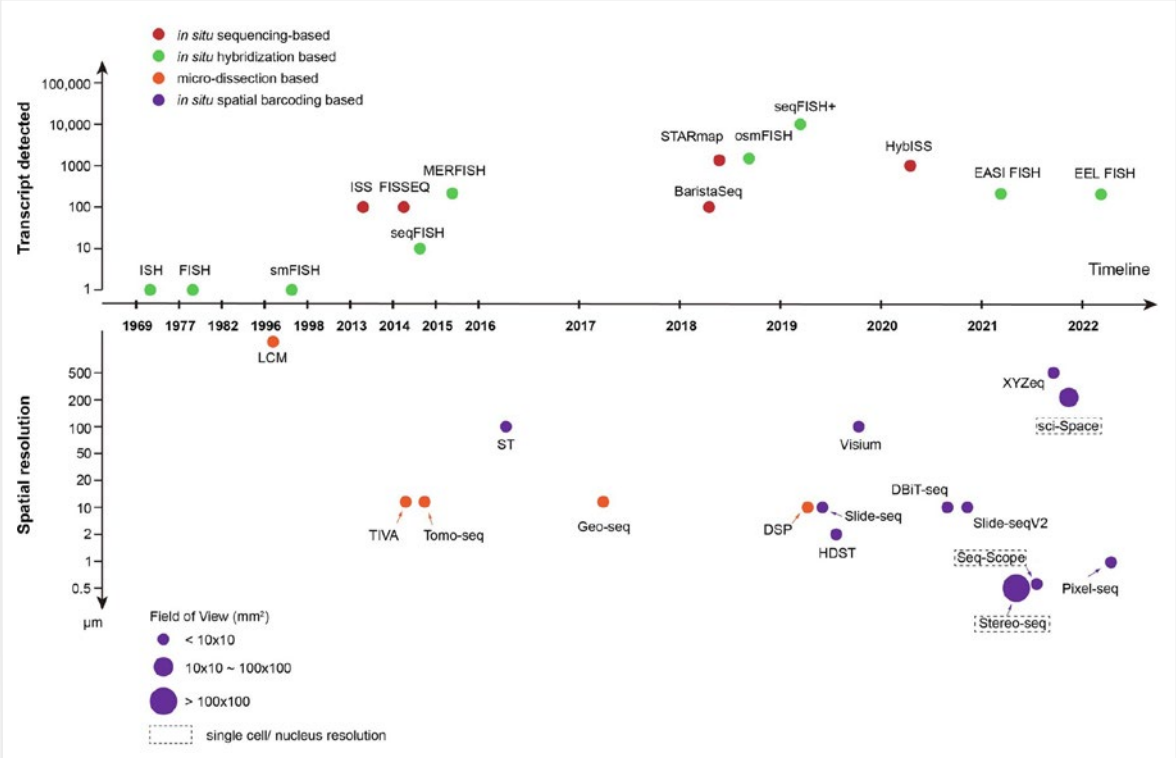
Figure 6 displays a timeline of landmark technologies within these four categories of methods. We will review some of the major advances in ISH, ISS and ISC methodologies.

The first RNA ISH was performed in the early 1990s and what followed was a series of advances that improved the resolution. In 1998, the first single molecule FISH[34] (smFISH) allowed the first subcellular RNA visualisation with singly labelled oligonucleotide probes. This 1998 method suffered technical issues, which meant the first reliable smFISH was published in 2008[35]. This was followed in 2011/12 by the first RNAscope, which used a novel 'Z' probe to bind to RNA transcripts. Later in 2014, seqFISH[36], was published, which used sequential hybridisations to expand the number of targets and in 2015, MERFISH[37], improved on the time and effort of seqFISH using multiple readout hybridisations. SeqFISH+[38] and MERFISH+[39], both published in 2019, mark the current pinnacle of ISH transcript detection allowing for ~10,000 genes to be visualised with confocal microscopy. The historical limitation on the number of markers is overcome using multiple bindings sites and multiple rounds of binding.

For in situ sequencing, the first approach, called In Situ Sequencing[40] (ISS) was published in 2013 and used padlock probes to sequence targeted genes in tissue sections. This approach achieved subcellular resolution but was limited to ~100 targets. The 2019 release of HybISS[41] saw radical improvement of the target limit using sequencing-by-hybridisation instead of by ligation. ISS is continually seeing improvement with this year's publication of Improved ISS[42] (IISS) using new probing, barcoding and imaging for better gene profiling. Other methods of in situ sequencing including FISSEQ[43], published in 2015, use fluorescence methods to capture genome-wide RNA in an unbiased manner, but, again, do not have the capacity for whole transcriptome level sequencing. More recent methods such as STARmap[44] use padlocks without reverse transcription and DNA nanoballs to sequence an expanded range of targets.

FIGURE 6. TIMELINE OF MAJOR LANDMARKS IN SPATIAL GENOMIC METHODS.

Methodologies are categorised based on type (red, in situ sequencing-based method; green, in situ hybridization based method; orange: micro-dissection based method; purple: in situ spatial barcoding based method). In situ spatial barcoding approaches are quantified by spatial resolution and by Field of View (size of circle). Dashed box indicates single-cell/nuclei resolution. Image Credit: Cheng, et al. [45]



"WHILE BARCODING METHODS SEEM PROMISING AND THE DIRECTION OF FUTURE PROGRESS, ESPECIALLY GIVEN THE IMPRESSIVE SPATIAL RESOLUTION, RNA CAPTURE EFFICIENCY IS STILL POOR COMPARED TO ISOLATING THE CELLS AND SEQUENCING THEM WITH SINGLE-CELL SEQUENCING TECHNOLOGY."

The most recent and most rapidly advancing set of approaches are the in situ capture/barcoding methods. These methods capture transcripts in situ that are then sequenced ex-situ, which avoids the limits of visualising and sequencing targets in situ. However, the challenge for this technology is RNA capture efficiency, particularly with the drive to achieve higher resolution with smaller capture areas. Spatial transcriptomics[46] (ST) was the first of these approaches published in 2016, which was later released commercially as 10x Genomics Visium. It uses barcoded RT primers printed onto glass slides for RNA capture, but initial resolution was 100 micrometres, improving to 55 micrometres with the Visium. Slide-seqV1 and V2[47,48] radically improved the resolution to 10 micrometres and use barcoded beads on a slide.

High-definition spatial transcriptomics[49] (HDST) was released shortly after Slide-seqV1 using smaller beads, enabling an impressive 2 micrometre resolution. Very recent techniques such as Seq-Scope[50], Pixel-seq[51] and Stereo-seq[52] represent the current cutting edge, having refined spatial resolution to under 1 micrometer. Stereo-seq used DNA nanoballs to achieve this and is the first technology to have achieved subcellular resolution and a centimetre-scale field of view. Finally, techniques such as DBiT-seq[53] and sci-Space[54] differ in that they use deterministic barcoding, and the latter method allows for large fields of view at the cost of spatial resolution.

While barcoding methods seem promising and the current direction of future progress, especially given the impressive spatial resolution, RNA capture efficiency is still poor compared to isolating the cells and sequencing them with single-cell sequencing technology. Brand new techniques, such as Slide-tags[55], can address this issue. Slide-tags first barcodes nuclei within tissues with a high spatial resolution. These nuclei can then be dissociated and isolated, meaning the mature single-nuclei technology can be used to sequence them. However, the nuclei still have the spatial barcodes and can be mapped back to their tissue context. We caught up with first author, Dr. Andrew Russell, to hear more about Slide-tags.

INTERVIEW:

ANDREW RUSSELL

POSTDOCTORAL FELLOW, FEI CHEN LAB

BROAD INSTITUTE OF MIT AND HARVARD

FLG: ***Can you describe how Slide-tags works and what gap it is addressing in the field?***

Andrew: For the last decade or more, we've seen the development of a vast repertoire of single-cell sequencing technologies. These are mainly for transcriptomic measurements, but have also expanded to DNA, proteomes and other modalities. These single-cell technologies are really advanced and, combined with some great analysis tools, and commercial offerings, they allow us to do experiments with millions of cells. And that really opens us up to new biology. On the other hand, when we dissociate tissues into single cells with these technologies, we lose the spatial context.

This is what the field has been trying to address in more recent years. We have this emerging branch of spatial sequencing tools. Namely, spatial transcriptomics tools, such as Slide-seq from our lab and other academic and commercial platforms, allow us to measure macromolecules in a spatial context. But they either lose resolution by having to capture these molecules in pixel-based measurements, or they use probe-based methods which need segmentation and aren't transcriptome-wide, so they're not perfect either. Hence, Slide-tags solves this issue between spatially resolved sequencing technologies and single-cell sequencing technologies by essentially allowing you to place single-cell transcriptomes back to their original spatial location within a tissue.

We use the spatially barcoded bead arrays that we've largely developed in the lab for Slide-seq and Slide-DNA-seq. Rather than capturing molecules onto those spatially barcoded bead arrays as we would with those technologies, we actually mobilise these spatial barcodes, we cleave them, and they diffuse into tissues. Then, in these fresh frozen tissues, they associate with nuclei, in a very high spatial resolution manner. We can then isolate these spatially-barcoded nuclei and profile them with established single-cell sequencing technologies. This means we can both capture the transcriptome of those single cells and the spatial barcodes become associated with that single nucleus. That then allows us to both profile the transcriptome, but also use that collection of spatial barcodes to relocate the cell in that tissue.

> "I THINK THAT THE EXCITING THING ABOUT DEVELOPING A TECHNOLOGY IS THAT YOU HAVE THESE IDEAS ABOUT HOW IT CAN BE APPLIED AND WHERE IT CAN BE USEFUL, BUT THEN, AMAZING COLLABORATORS AND SCIENTISTS REALLY DISCOVER THEM FOR YOU."

***FLG:* And what is the spatial resolution like when you map the nuclei back to spatial context?**

Andrew: We've actually done some recent analysis of this. In our paper, we address this in a few different ways. The first thing we did was we checked to see whether we were recapitulating small-scale spatial structures that we saw in a mouse hippocampus, and we found that it was recapitulating the image very faithfully. Secondly, we also tried to measure the accuracy of the placement with the standard error of the spatial barcode distribution of each nucleus. We found this to be 3.5 microns in both x and y. This is quite exciting because this is approximately the size of a nucleus itself. Hence, it makes sense that the accuracy of our spatial placement can't be any better than the diameter of the nucleus.

***FLG:* In summary, Slide-tags allows you to take advantage of how far ahead single-cell sequencing technology is, while maintaining spatial context, is that right?**

Andrew: Yeah, exactly. And back to what I said earlier, we have a lot of different tools to sequence single cells. We think that because this technology is essentially tagging the nuclei with oligonucleotides, any of these sequencing technologies could be applied downstream of slide tags. So, I think there's two main points.

One is that when we're making these spatial technologies, and slide-DNA-seq is an example of that, we have to do a lot of work to effectively measure a new macromolecule in space. But now, with Slide-tags, we could just take that tagged nucleus and we could import it into this whole stream of hundreds of technologies that we now have for single-cell resolved measurements.

The second is that we're still really trying to make the most of these spatial measurements in terms of analysis and we have a vast repertoire of single-cell analysis tools. Hence, it's not just that Slide-tags makes the measurement of new macromolecules easier at a single cell resolution, but it also makes the overall analysis much easier because you don't need to develop a new bespoke analysis for these new tools.

***FLG:* Are there any specific scientific questions or problems that this technology could be applied to?**

Andrew: Yes. What I'm most excited about is the ability to really import this technology into multiomics measurements of single cells. It's something we struggled to do with these existing, spatially resolved sequencing technologies. We have many tools now for measuring the genome and transcriptome in single cells. This offers a platform to do this spatially, which we show in the paper, and we haven't really been able to effectively do this. I think these linked measurements between different modalities allow us to look at options for temporal dynamics. When you can measure open chromatin and transcriptomes simultaneously in the same single cell, you can get a really good dynamic readout. This has been done by others, but we think we can do this in a spatial context now.

Also, if you can measure, say, mutations in a single cell, and the effect on that transcriptome, you get a sense of these genotype/phenotype relationships, which are very important in cancer research, but maybe also important in neurodegeneration and also in aging-associated disease research in general. However, I think that the exciting thing about developing a technology is that you have these ideas about how it can be applied and where it can be useful, but then, amazing collaborators and scientists really discover them for you. I know that people are going to do exciting stuff that we haven't even thought about when we started developing this technology.

Now that we've caught up on the history of single-cell and spatial omics, we can see that progress towards larger capacities and higher resolutions has been fairly constant. Let's move on to chapter 2 to look at the commercial developments in the last two years to get a better picture of the latest improvements in single-cell and spatial technology.

Chapter 1 references

1. Li, X. & Wang, C.-Y. **From bulk, single-cell to spatial RNA sequencing.** International Journal of Oral Science **13**, 36 (2021).
2. **Method of the Year 2013.** Nature Methods **11**, 1-1 (2014).
3. Pan, Y., Cao, W., Mu, Y. & Zhu, Q. **Microfluidics facilitates the development of single-cell RNA sequencing.** Biosensors **12**, 450 (2022).
4. **Method of the Year 2019: Single-cell multimodal omics.** Nature Methods **17**, 1-1 (2020).
5. **Method of the Year 2020: spatially resolved transcriptomics.** Nature Methods **18**, 1-1 (2021).
6. Quake, S.R. **A decade of molecular cell atlases.** Trends in Genetics **38**, 805-810 (2022).
7. Tang, F. et al. **mRNA-Seq whole-transcriptome analysis of a single cell.** Nature Methods **6**, 377-382 (2009).
8. Jovic, D. et al. **Single-cell RNA sequencing technologies and applications: A brief overview.** Clinical and Translational Medicine **12**, e694 (2022).
9. Carangelo, G., Magi, A. & Semeraro, R. **From multitude to singularity: An up-to-date overview of scRNA-seq data generation and analysis.** Frontiers in Genetics **13**(2022).
10. Wang, S. et al. **The evolution of single-cell RNA sequencing technology and application: progress and perspectives.** Int J Mol Sci **24**(2023).
11. Islam, S. et al. **Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq.** Genome research **21**, 1160-1167 (2011).
12. DeLaughter, D.M. **The use of the Fluidigm C1 for RNA expression analyses of single cells.** Current protocols in molecular biology **122**, e55 (2018).
13. Brennecke, P. et al. **Accounting for technical noise in single-cell RNA-seq experiments.** Nature Methods **10**, 1093-1095 (2013).
14. Jaitin, D.A. et al. **Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types.** Science **343**, 776-9 (2014).
15. Klein, A.M. et al. **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.** Cell **161**, 1187-1201 (2015).
16. Macosko, E.Z. et al. **Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.** Cell **161**, 1202-1214 (2015).
17. Cao, J. et al. **Comprehensive single-cell transcriptional profiling of a multicellular organism.** Science **357**, 661-667 (2017).
18. Domcke, S. et al. **A human cell atlas of fetal chromatin accessibility.** Science **370**, eaba7612 (2020).
19. 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).
20. Kharchenko, P.V. **The triumphs and limitations of computational methods for scRNA-seq.** Nature Methods **18**, 723-732 (2021).
21. Picelli, S. et al. **Smart-seq2 for sensitive full-length transcriptome profiling in single cells.** Nature methods **10**, 1096-1098 (2013).
22. Ramsköld, D. et al. **Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.** Nature biotechnology **30**, 777-782 (2012).
23. Hashimshony, T. et al. **CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.** Genome biology **17**, 1-7 (2016).
24. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. **CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.** Cell reports **2**, 666-673 (2012).
25. Hagemann-Jensen, M. et al. **Single-cell RNA counting at allele and isoform resolution using Smart-seq3.** Nature Biotechnology **38**, 708-714 (2020).
26. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. **Methods and applications for single-cell and spatial multi-omics.** Nature Reviews Genetics, 1-22 (2023).
27. Baysoy, A., Bai, Z., Satija, R. & Fan, R. **The technological landscape and applications of single-cell multi-omics.** Nature Reviews Molecular Cell Biology, 1-19 (2023).
28. Zhang, Y., Huang, Y., Hu, L. & Cheng, T. **New insights into human hematopoietic stem and progenitor cells via single-cell omics.** Stem Cell Reviews and Reports **18**, 1322-1336 (2022).
29. Emmert-Buck, M.R. et al. **Laser capture microdissection.** Science **274**, 998-1001 (1996).
30. Lovatt, D. et al. **Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue.** Nat Methods **11**, 190-6 (2014).
31. Medaglia, C. et al. **Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq.** Science **358**, 1622-1626 (2017).
32. Boisset, J.C. et al. **Mapping the physical network of cellular interactions.** Nat Methods **15**, 547-553 (2018).
33. Zhang, L. et al. **Clinical and translational values of spatial transcriptomics.** Signal Transduction and Targeted Therapy **7**, 111 (2022).
34. Femino, A.M., Fay, F.S., Fogarty, K. & Singer, R.H. **Visualization of single RNA transcripts in situ.** Science **280**, 585-590 (1998).
35. 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**, 877-9 (2008).
36. Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M. & Cai, L. **Single-cell in situ RNA profiling by sequential hybridization.** Nat Methods **11**, 360-1 (2014).
37. 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).
38. Eng, C.L. et al. **Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH.** Nature **568**, 235-239 (2019).
39. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. **Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression.** Proc Natl Acad Sci U S A **116**, 19490-19499 (2019).
40. Ke, R. et al. **In situ sequencing for RNA analysis in preserved tissue and cells.** Nat Methods **10**, 857-60 (2013).
41. Gyllborg, D. et al. **Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue.** Nucleic acids research **48**, e112-e112 (2020).
42. Tang, X. et al. **Improved in situ sequencing for high-resolution targeted spatial transcriptomic analysis in tissue sections.** Journal of Genetics and Genomics (2023).
43. Lee, J.H. et al. **Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues.** Nature Protocols **10**, 442-458 (2015).
44. Wang, X. et al. **Three-dimensional intact-tissue sequencing of single-cell transcriptional states.** Science **361**(2018).
45. Cheng, M. et al. **Spatially resolved transcriptomics: A comprehensive review of their technological advances, applications, and challenges.** Journal of Genetics and Genomics (2023).
46. Ståhl, P.L. et al. **Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.** Science **353**, 78-82 (2016).
47. Rodriques, S.G. et al. **Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.** Science **363**, 1463-1467 (2019).
48. Stickels, R.R. et al. **Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2.** Nat Biotechnol **39**, 313-319 (2021).
49. Vickovic, S. et al. **High-definition spatial transcriptomics for in situ tissue profiling.** Nat Methods **16**, 987-990 (2019).
50. Cho, C.-S. et al. **Microscopic examination of spatial transcriptome using Seq-Scope.** Cell **184**, 3559-3572. e22 (2021).
51. Fu, X. et al. **Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain.** Cell **185**, 4621-4633. e17 (2022).
52. Chen, A. et al. **Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell** 185, 1777-1792. e21 (2022).
53. Liu, Y., Enninful, A., Deng, Y. & Fan, R. **Spatial transcriptome sequencing of FFPE tissues at the cellular level.** bioRxiv, 2020.10. 13.338475 (2020).
54. Srivatsan, S.R. et al. **Embryo-scale, single-cell spatial transcriptomics.** Science **373**, 111-117 (2021).
55. Russell, A.J.C. et al. **Slide-tags: scalable, single-nucleus barcoding for multi-modal spatial genomics.** bioRxiv, 2023.04.01.535228 (2023).

WHAT'S NEW. AN OVERVIEW OF THE LATEST COMMERCIAL TECHNOLOGY DEVELOPMENTS

AS SUMMARISED IN CHAPTER 1, SINGLE-CELL AND SPATIAL ANALYSIS HAS DRASTICALLY IMPROVED OVER THE LAST DECADE. EACH PASSING YEAR SEES FURTHER EXPANSION IN THE NUMBER OF COMMERCIAL INSTRUMENTS AVAILABLE AND THE CAPACITIES OF THESE INSTRUMENTS.

For researchers, large-scale single-cell spatial multiomics is the most common short-term goal. The past two years have seen commercial advancements in spatial multiomics methods as well as increases in capacity, multiplexing and resolution to make this a reality.

Despite its maturity, single-cell sequencing hasn't been static and has seen expansion in its multiomics capacity, as well as potentially game-changing instrument-free approaches that could make sequencing more accessible and highly scalable.

In the following sections we will highlight a selection of commercial advances that have happened in the past two years, providing an overview of the directions of progress. This list is not exhaustive and was compiled in August 2023.

MULTIOMICS CAPABILITIES

SPATIAL CO-DETECTION OF RNA AND PROTEIN IN ONE EXPERIMENT

Some of the most established companies in the spatial transcriptomic space have made the same move in the last 24 months; to diversify their core spatial transcriptomic products to co-visualize proteins within a single experiment, making true multiomics spatial platforms. Some key examples include:

10x Genomics Visium Cyt Assist™ - first launched in 2022 to improve sample preparation, has now released a whole transcriptome and 31-plex protein assay as of May 2023. This allows the study of protein and RNA in a single tissue section as well as H&E/IF staining for tissue morphology.

Nanostring CosMx™ SMI - launched in December 2022 as the highest plex in situ imager with 1000-plex RNAs and 64-plex proteins analysed in the same tissue at subcellular resolution. This joins their other products including the first cloud-based spatial data analysis resource, AtoMx™ and their established GeoMx™ that has broader protein and RNA capability with low resolution.

Vizgen MERSCOPE™ - launched a protein co-detection kit in September 2022, allowing users to take full advantage of the subcellular hi-plex nature of the instrument while detecting up to five proteins.

10x GENOMICS

Your sample, your spatial story

Study the samples that are most important to you, whether FFPE-processed or fresh-frozen, TMAs, blocks, or pre-sectioned slides.

Explore CytAssist

Contact us
10xgen.com/spatial-fg | info@10xgenomics.com

SAM000698 - Rev A - Flyer - Your sample, your spatial story

In 2023, Miltenyi Biotec introduced RNAsky™ technology which integrates spatial RNA detection with multiplexed protein in the same experiment on the MACSima™ Spatial Biology Imaging Platform. The automated multiomics workflow can be analyzed with MACS® iQ View image software or other bioinformatics pipelines.

Akoya Bioscience's PhenoCycler®-Fusion was released in January 2022 and offers a high throughput workflow at sub-cellular resolution for 100+ markers, either RNA or protein biomarkers. It is one of the fastest single-cell spatial biology solution, able to map a million cells in 10 minutes.

ACD Biotechne and Standard Biotools have created a workflow (May 2023) to combine the 12-plex RNAscope™ assay with the 40-plus protein Imaging Mass Cytometry™ assay, to create RNA and protein multiomics outputs.

A DIVERSIFICATION OF MULTIOMICS SINGLE-CELL SEQUENCING

Multimodal sequencing capacity has also diversified on the single-cell sequencing front, with new specific omics available to sequence such as:

BioSkryb Genomics' ResolveOME™ kit allows the near-complete summary of the genome and mRNA transcriptome at single cell resolution. With their associated BaseJumper™ data analysis software, this setup creates a unified workflow for DNA and RNA interrogation.

Mission Bio released Tapestri® v3 in May 2023, with DNA as its primary analyte. It allows the analysis of the genotype of a cell plus a variety of other phenotypic data such as CNVs, SNVs or proteins. The new v3 allows up to four times more cells captured per sample, which increases the ability to detect rare cells.

Singleron's PromoScope™ kit, released in November 2022, allows the simultaneous quantification of the whole transcriptome, as well as protein glycosylation at the single-cell level. Relying on their SCOPE-chip® technology, this is the first kit to quantify protein modifications alongside transcriptomics.

SPATIAL METHODS GET SOPHISTICATED

SPATIAL TRANSCRIPTOMICS – MORE TARGETS, MORE RESOLUTION

New spatial transcriptomic tools are increasing the speed, capacity and scope of this technique. This provides researchers with a variety of options to meet their spatial needs. Whether they provide access to more cells, more targets to produce a higher-plex assay or a high resolution to pinpoint tiny molecules, here we highlight a few of the mRNA visualization technologies released in the last 24 months that expand the horizons of spatial biology:

10x Genomics released the Xenium™ platform in December 2022, which is an in situ sequencing platform to compliment the in situ capture platform - Visium™. 'The Xenium™ can sequence the transcripts of 1000s of genes in a high throughput non-destructive manner, which allows proteomics and histopathology to be conducted on the same slide.

Vizgen MERSCOPE™ was released in January 2022 and commands an impressive capacity to visualize 10'000s of RNA in situ with sub-cellular resolution and high sensitivity.

ACD Biotechne's RNAscope™ recently passed a milestone of 40,000 probes (currently over 44,000) in its probe library, making it the broadest in situ technology with high sensitivity.

Curio Bioscience released the Curio Seeker in February 2023, based off of Slide-seq v2 technology, which enables large scale whole transcriptome spatial tissue map at single-cell resolution. This in situ barcoding approach does not require any specific hardware but simply requires the mounting of your tissue on the Curio Seeker Tile. It boasts a 10 micron resolution and no gaps between spots.

The Molecular Cartography™ platform, released by Resolve Biosciences in June 2022 is a multi-analyte, highly multiplexed spatial solution to visualise 100's of genes with subcellular resolution in a single run. The approach uses high-quality optics, which provides the highest resolution and can assess 100's of genes at once. Resolve have stated that future solutions will incorporate additional data layers of analysis (DNA, Protein and Metabolome).

The GenePS system was released by Spatial Genomics in May 2023 and is an automated instrument with capacity to panel over 1,000 genes at exceptional resolution using seqFISH technology. The analysis suite allows easy visualization of molecules from the tissue to the subcellular level as well as multiomics capacity.

SPATIAL PROTEOMICS SOLUTIONS ARE COMING IN FAST

While spatial proteomics still lags behind spatial transcriptomics for number, variety and capacity of solutions, advances were still aplenty in the last 24 months. Highlights include:

MACSima Imaging System was the first platform to publish more than 100 proteomic markers on a single tissue sample, in a single experiment, with off-the-shelf primary antibodies. In 2022, Miltenyi Biotec launched MACS® iQ View image analysis software which allows researchers to easily analyze large complex data stacks. In 2023 Miltenyi Biotec released a 61-marker REAscreen™ IO Plate and custom REAscreen plates for added automation.

Navinci Diagnostics' Naviniflex™ launched the Triflex Cell (November 2022) and Tissue (January 2023) solutions, allowing the visualization of protein-protein interactions and post-translation modifications, revealing the hidden interactions of proteins.

With the launch of Spyre™ antibody panel kits and the HORIZON™ analysis software (in March 2023), Lunaphore's Comet™ became the first end-end spatial proteomic solution able to visualize up to 40 markers on one tissue sample in under a day. The plex-capacity is essentially unlimited, as tissue can be re-run with a new panel of markers.

Canopy Biosciences® launched CellScape™ in March 2022, a benchtop imaging system using ChipCytometry™ technology. It can process 4 samples at once and is technically unlimited-plex due to iterative staining. This presents a rapid, fully automated multiplex spatial proteomics solution.

Standard Biotools released the Hyperion™ XTi Imaging system in April 2023. Relying on the Imaging Mass Cytometry™ technology, it both expands the number of samples and reduces the time taken to perform spatial proteomics. This tool detects 40-plus biomarkers with no autofluorescence interference.

Pixelgen Technologies® launched the first molecular pixelation kit (June 2023) to visualize spatial polarization and colocalization of cell surface proteins at high multiplex in 3D.

These instruments join more established instruments such as Rarecyte's Orion™, and Ionpath's MIBIscope™ in a suite of spatial proteomic solutions.

SINGLE-CELL SEQUENCING STILL SEES IMPROVEMENT

SINGLE-CELL IMPROVEMENT – EVEN HIGHER THROUGHPUT AND FULL-LENGTH TRANSCRIPTOME

As shown in Chapter 1, the throughput of single-cell methods is growing exponentially. Even in the last year, speed, throughput, long-read capacity and consistency have all been improved for established and new instruments, keeping the wheels of progress turning. Some examples include:

BD Biosciences released the BD Rhapsody™ HT Xpress System in February 2023, which makes million-cell studies possible with over 320,000 cells per cartridge and with up to an 80% capture rate. This allows cell capture and barcoding with extremely high throughput. This joins the FACSDiscover S8™ Cell Sorter, which was released in May 2023 and allows cell sorting based on imaging the cells, and the FACSDuet™, launched in July 2023, which automates flow cytometry to automate the entire sample preparation process.

For brand new instruments, Singular Genomics have recently released the G4 sequencer in 2022, one of the most powerful benchtop sequencers available. With the new Max Read™ kit released in February 2023, the instrument now has the power to sequence up to 3.2 billion reads a day.

Pacific Biosciences launched their MAS-Seq kit for single-cell expression analysis in October 2022. It leverages 10x Genomics' single-cell technology with PacBio's HIFI technology for long-read RNA sequencing, to allow researchers to assess novel isoforms and the additional value of long read sequencing in single cells.

Takara presented the SMART-Seq® Pro kit in October 2021, which can perform full-length transcriptome sequencing on single cells isolated with their ICELL8® single-cell system, providing an automated end-to-end solution for sensitive transcriptomic analysis.

Singleron released NEO and Python Junior instruments in June 2023, which are portable instruments for library processing and tissue dissociation respectively. The instruments allow half a million cells in one run, and were designed with standardization and automation in mind.

SCALABLE AND COST-EFFECTIVE INSTRUMENT-FREE APPROACHES

To make single-cell sequencing accessible to any lab is a lofty goal for the research world. In progress towards this goal, the cost of performing a single-cell experiment has rapidly fallen over the last few years, with companies increasing the throughput and capacity of their instruments (see above). However, the inability to initially invest in the instruments leaves many labs unable to participate in the single-cell revolution. This is the space that the following companies have innovated in by supplying instrument-free single-cell solutions.

Parse Biosciences released version 2 of Evercode™ back in August 2022 (following the release of version 1 in February 2021). This new version is more sensitive and robust, and by using the cell or nucleus as the reaction vessel, no hardware is needed, meaning an initial expensive hardware purchase is avoided. Kits can process up to 1 million cells and Evercode™TCR allows T cell receptor profiling too.

Fluent Biosciences also updated their PIPseq™ platform to version 4 in February 2023. With one of the highest cell captures (85%) and better sensitivity, the kit also offers a scalable and cost-effective sequencing solution, compatible with Illumina NGS sequencing instruments. The T100 kit is the first single tube solution with the capacity for 100,000 cells.

Scale Biosciences presents their own technology to escape the need for cell partitioning instruments in single-cell sequencing, through their Scale Bio™ RNA kit (December 2022). Their product is also scalable, affordable and allows deep profiling of cells. The single-cell ATAC kit sets Scale Biosciences apart as offering instrument-free epigenomic pre-indexing.

Honeycomb Biotechnologies released HIVE CLX™ in May 2023, which has 160,000 picowells in their distinctive array for gentle capture of fragile cells. This allows for integration of stable sample storage (up to 9 months) and single-cell profiling without needing specialized instrumentation. Cells are captured quickly and effectively and can be stored as you go - meaning samples can be collected across time without batch effects.

Despite these advances in the capacity of single-cell technologies, we still see some of the same problems for single-cell analysis. We asked some experts in the field, who have a lot of experience in assisting others, about the common problems they have encountered. Furthermore, we asked what advice they had for working with difficult tissues, for sample prep and for deciding how many cells are appropriate for an experiment.

CATIA MOUTINHO

Founder & Scientific Advisor
The Single-Cell World

FLG: As a previous single-cell group leader, and founder of the Single-Cell World platform, what are the most common problems that people have with single-cell technologies?

Catia: *First, the sample preparation. People really struggle to do sample preparation - to get individualized cells in suspension with good viability. This is mainly because of a lack of knowledge. Normally when you look for a bulk protocol, you just do the protocol. And that's it. But the problem with single-cell is that you need happy individual cells in suspension, and how to get them changes a lot between different tissues, and even within the same tissue. For example, if there is fibrosis, necrosis, this will influence the tissue dissociation, and so people will get frustrated. They are not aware that they need to optimize a lot of things. Also, most of the time they don't know what needs to be modified. In standard research, normally, you get a protocol that is published, and you do it. In general, it works without major changes. In single-cell research this is rare.*

Second, the data analysis. One of the major concerns for researchers is who will do the data analysis. It is true that single-cell technology companies are offering software that you can use to analyse the data produced with their technology. But unfortunately, the type of analysis that you get from these "friendly user" programs are not enough at the time of publishing in peer review journals. My advice is always for researchers to establish collaborations with data analysis expert groups, but this is also difficult to get.

FLG: Do you have any advice for people working on difficult tissues, such as human brain?

Catia: *My first advice is optimize, optimize, optimize. Optimize the sample preparation protocol. Even if you find a perfect protocol for that tissue type, try it in the lab with a small tissue piece or a small number of samples. Get used to the protocol. If working with fresh tissue or cryopreserved, check if you can get good quality cells or good quality nuclei.* ▶

If the samples are snap-frozen or frozen, this is where people sometimes lack the knowledge, all the cells are dead. You will never get live cells. This is also true for fixing in ethanol/ methanol etc. In this case, you must go straight for nuclei isolation. You should also perform single nuclei experiments when you have fresh tissues that are difficult to dissociate, like brain, as you mentioned, but also fat tissue or pancreas, for example. You can try to get cells, but with these tissues, in my experience, you will get a lot of debris that then will result in poor quality data. Since people don't have this information, this can be why their experiments fail.

I would like also to mention that single-cell technology is moving in a way where sample preparation logistics are being simplified. There are already kits that allow us to preserve the cells and continue our experiment after some months, or others with which you can rescue archival material that, before, we couldn't use to do single-cell or nuclei experiments.

LINDA D. ORZOLEK

Director, Single Cell & Transcriptomics Core
Johns Hopkins University

FLG: As a core director, one thing I'm sure you've been asked a lot is whether there is a default number of cells and number of samples that you'd recommend?

Linda: *10,000 cells has been the standard number but this is still a relatively arbitrary number. We suggest 500 to 1000 cells per cluster as a starting point to make sure that you have confidence. And we are seeing a push now to run samples in triplicate. It's not because people realized that we need biological replicates for statistical purposes - we always knew that - but the technology was too expensive. And [the push] happens to correspond with the cost coming down; all of a sudden, we need more replicates. And you knew it was always the case, but people couldn't do it. So, that's why things have shifted. We encourage triplicate where you can. Ultimately, we encourage the most cells you can afford. But there are complex power calculations that should be done to determine the real answer and most people don't understand the complexities associated with that in the context of their project, so we have to look for a "starting point."*

And then there are different assays that are coming out, BioSkryb provides whole genome single-cell sequencing now, which is a plate-based method. So, you're really looking at 100 or so cells. Why do you only need 100 cells? Why is that enough when some people are saying that 10,000 isn't? Well, it captures more, hence we get a more complete picture within that single cell, so you don't need as many replicates of the cell to have confidence in what you're calling. There are still different ways of looking at things and approaching them. What the benefits of the different technologies are, that has to come into play with that answer.

"SINGLE-CELL AND SPATIAL TECHNOLOGIES ARE EXPANDING THEIR CAPACITY IN TERMS OF NUMBER OF CELLS, NUMBER OF SAMPLES AND NUMBER OF MULTI-MODAL MARKERS."

EDGING TOWARDS CLINICAL USE

Ultimately, many commercial companies would like to see their instruments and technologies make a difference in a clinical setting. Several companies are incorporating fast turnaround speed and/or a large sample throughput to promote clinical use. Examples include Akoya Bioscience's PhenoImager™, Rarecyte's Orion™ and Singleron's GEXscope™. But as these technologies collectively become cheaper and faster, many of the advanced tools are approaching widespread clinical utility. One example of a technology designed for direct clinic use is Molecular Instrument's HCR™ RNA-CISH (released February 2023), which allows the spatial visualization of a single RNA target in situ in under a 6 hour runtime, with a much reduced cost, presenting a tool that could affordably and practically be deployed in a clinical diagnostic setting.

More generally, due to the large amount of clinical tissue stored as Formalin Fixed Paraffin Embedded (FFPE) samples, which allows long term storage but reduces the quality of omics data available, there has been increasing demand for commercial technology to be powerful enough to derive biological insights from FFPE samples. Most major spatial and sequencing technologies have now included FFPE compatibility in their workflow. This includes Nanostring for whom the CosMx and GeoMx have long had FFPE compatibility and more recently 10x Genomics have released FFPE compatibility for their chromium sequencer in April 2022 and now Vizgen, who released a solution kit for FFPE tissues for MERSCOPE in December 2022.

The valuable progress of using single-cell and spatial technologies for direct clinical applications is explored in depth in Chapter 6 of this report.

In summary, single-cell and spatial technologies are expanding their capacity in terms of number of cells, number of samples and number of multimodal markers. Our next chapter will look at some of the large-scale projects that have been made possible with this enhanced technology.

CHAPTER 3

SIZE MATTERS. SINGLE-CELL AND SPATIAL, MULTIPLEX AND MULTIOMICS

MORE CELLS, MORE SIGNALS, MORE OMICS – SINGLE-CELL AND SPATIAL BIOLOGY IS UNDERGOING EXPANSION. THIS CHAPTER WILL COVER SOME OF THE BIGGEST INITIATIVES IN SINGLE-CELL BIOLOGY AS WELL AS THE UNIQUE PROBLEMS THAT ARISE FROM WORKING AT THIS VAST SCALE. SPECIFICALLY, WE WILL COVER WAYS TO UTILISE SINGLE-CELL ATLASES, WAYS TO HANDLE MULTIPLEX IMAGES AND HOW TO INTEGRATE MULTIOMICS DATA.

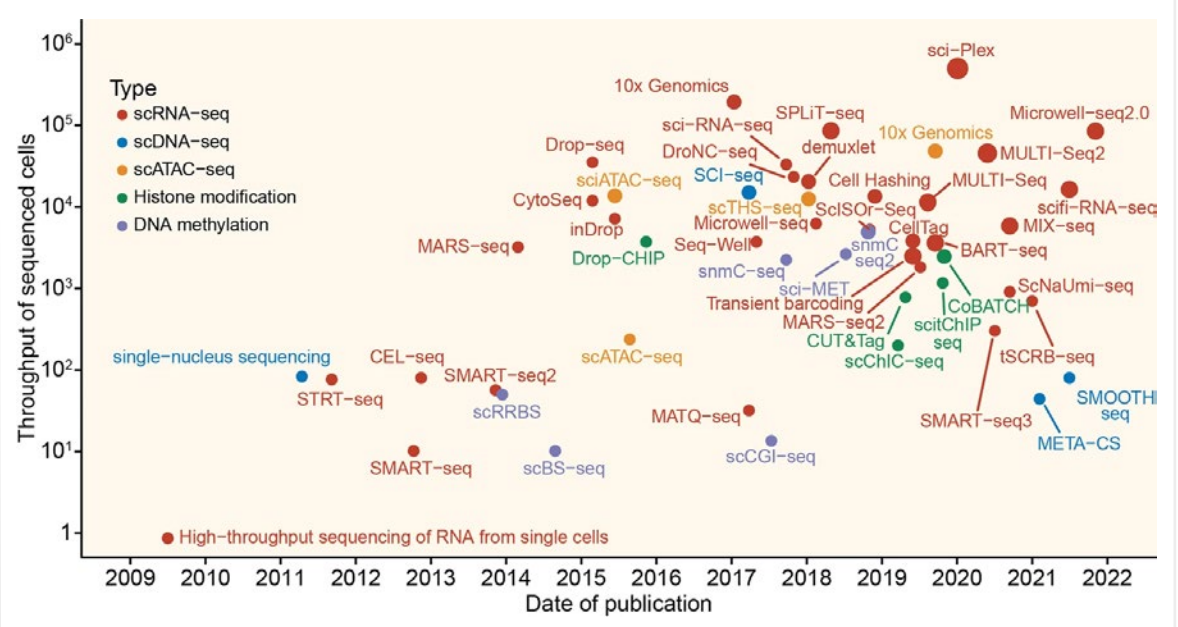
Multi-cell - Atlasing projects using single-cell technology

As we saw in Chapters 1 and 2, a combined academic and commercial effort is producing technology capable of conducting single-cell experiments at huge scale. As a consequence, scientists are using this technology to profile incredible numbers of cells from specific tissues in the human body, creating atlases with increasingly impressive scope.

With combinatorial indexing[1], single-cell experiments are now able to process over a million cells (see Figure 1). This means it is becoming common place to see 100,000s of cells profiled in a single experiment, and the ability to profile whole tissues is open to anyone.

FIGURE 1. TIMELINE OF MAJOR SINGLE-CELL SEQUENCING TECHNOLOGIES WITH THE LEVEL OF THROUGHPUT OF SEQUENCED CELLS.

The most comprehensive single-cell sequencing technologies are capable of a throughput of 100's of 1000's of cells. Colour indicates the different modality of single-cell sequencing. Image Credit: Jia, et al.[2]

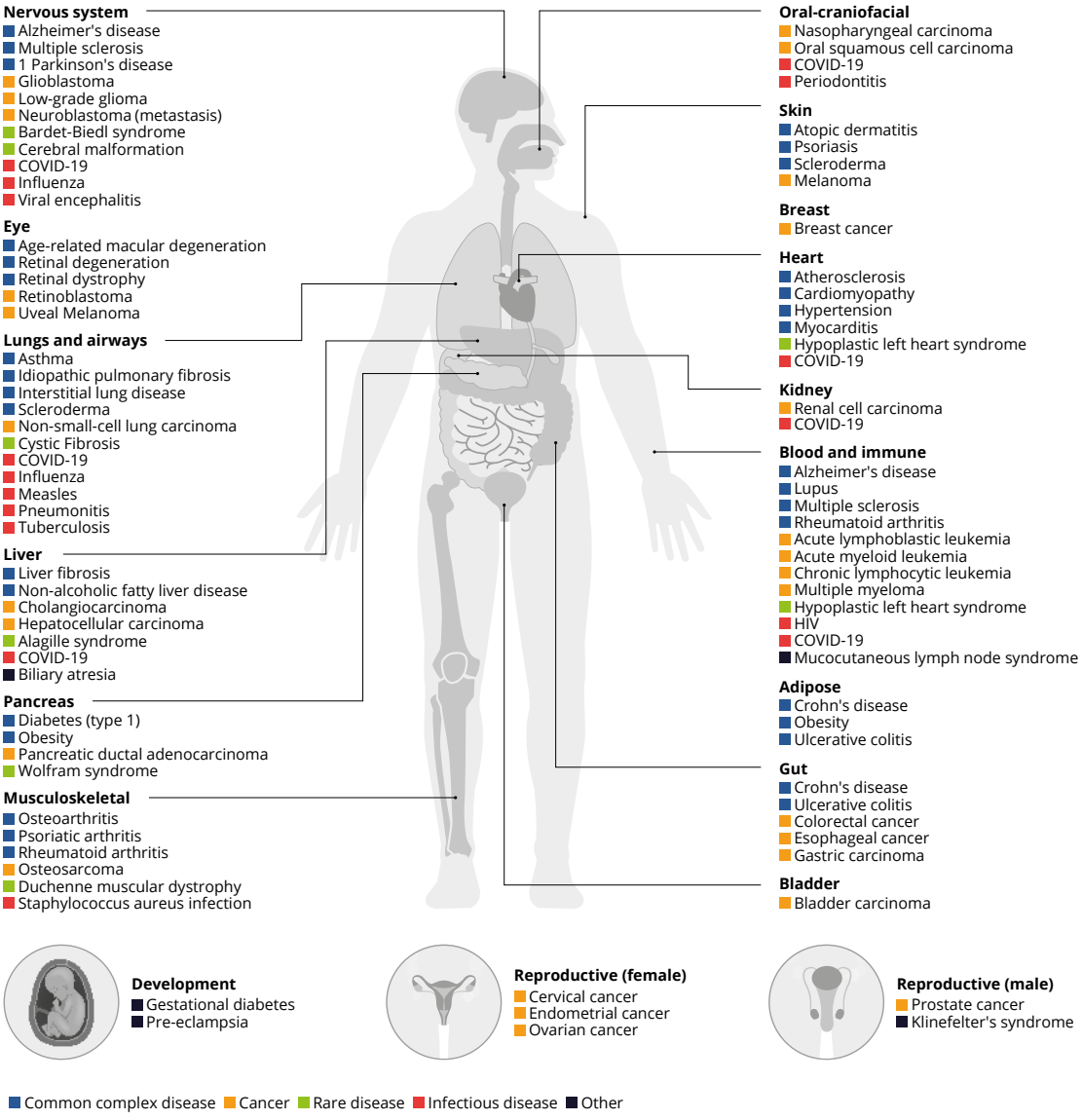


SINGLE-CELL CONSORTIA

The Human Cell Atlas (HCA) initiative is a global consortium using single-cell and spatial technologies to create comprehensive reference maps of the human body. It has over 3,000 members across 95 countries, organised into 18 networks targeting specific organs or systems (see this video for a 'state of'). Figure 2 outlines the selection of tissues that have been profiled as part of the HCA as of the end of 2022. Since then, we have seen developments in the heart cell atlas (700,000 cells and nuclei)[3,4], the kidney (400,000 cells and nuclei)[5] and the human breast (800,000 cells)[6]. Of note is the single-nuclei assessment of human post-mortem brain tissue, which sequenced over 3 million nuclei from all areas of the brain[7]. Also impressive is the new lung atlas[8], which represents one of the largest efforts to combine datasets for this purpose; 49 datasets and 2.4 million cells. This approach relied on transfer learning to effectively integrate so many datasets into one resource. Furthermore, a new roadmap is in place for a Human Gut Cell Atlas, which is proving difficult to map[9].

FIGURE 2. KEY ORGANS AND TISSUES FOR WHICH HEALTHY SAMPLES HAVE BEEN PROFILED AS PART OF THE HUMAN CELL ATLAS INITIATIVE.

Atlases of diseased samples are presented below each tissue and colour coded by condition. Image Credit: Rood, et al.[10]



The HCA isn't the only large-scale single-cell sequencing effort. Atlasing projects have been on scientists' minds for over a decade[11-13]. Examples include:

- The **NIH Human Biomarker Atlas Project** (HuBMAP) is a consortium using single-cell and spatial data to study a selection of tissues just like the HCA[14]. However, the wealth of spatial data sets HuBMAP apart and will be covered in a later section of this chapter.
- The **Human Protein Atlas** is a valuable resource in which single-cell proteomics data is being pooled to build a working resource.
- The **Human Tumour Atlas Network**[15] presents a consortium aiming to profile tumours with a selection of sequencing and spatial technologies to build atlases. The **Pan-Cancer Atlas** from the **Cancer Genome Atlas** consortium mirrors this with over 11,000 tumours from 33 forms of cancer.
- The **Harvard Tissue Atlas** presents a variety of spatial datasets for human tissues along with new tools and standards with which to achieve the most from this technology.
- The **Tabula Sapiens**[16] presents a large scale single experiment atlas of <500,00 nuclei across 24 different tissues. A feat mirrored with 8 tissues and 200,000 nuclei by the HCA[17].
- A **cross-species cross-lifespan cell landscape** of >2.6 million cells was recently put together to begin to explore conserved cell features and effects of aging[18]. This kind of resource is the long-term goal for human studies.
- **Descartes** is a resource collating data from the Shendure lab in the University of Washington, in which RNA and chromatin accessibility data of multiple organisms can be found[19].

This is along with all the individual groups currently producing vast single-cell resources using modern high-throughput technology, which everyone can benefit from.

UTILISING ATLAS DATA

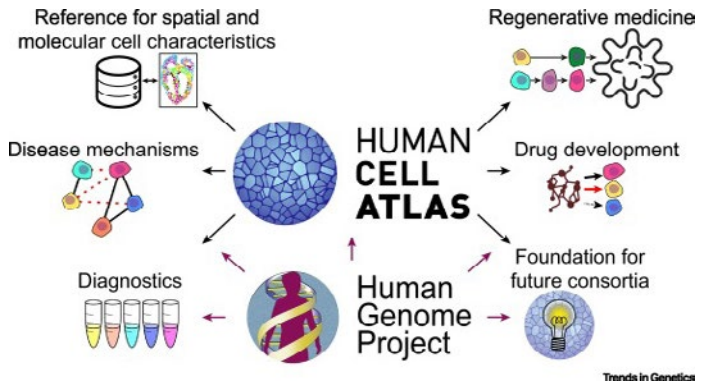
Atlases are a fundamental asset to basic biology. The value they might have for clinical applications is vast and summarised in Figure 3.[10] With novel cell types and cell states being continuously discovered from these efforts, our understanding of tissues in a healthy state, and in a disease state, is always improving. Furthermore, by profiling these organs in a consistent way, the group have been able to draw conclusions about cell types from across the body, able to find common and tissue-specialised versions of cell types such as immune cells and support cells such as fibroblasts.[20]

However, accessing this data is not as straightforward as it might appear. These atlases are so huge that they cannot be casually observed, nor is it a simple task to compare your own generated data to the atlas. This is where computational biology methods have come into their own, and many tools exist to handle this problem.

scArches[21] is a transfer learning tool that directly addresses the problem of matching your data (the query) to a reference atlas, allowing you to do several things such as comparing disease data to healthy atlases and extracting cell types from the atlas to match your query. We spoke to Dr. Mohammad Lotfollahi, the primary developer of this tool, to get his opinion on the value of scArches. He has recently been involved with another tool – expiMap[22] - that expands this process to map cells into gene programs, a more biologically understandable alternative, as well as proposing 'treeArches'[23], to progressively build and update reference atlases with new queries.

FIGURE 3. OVERVIEW OF VARIOUS DIRECTIONS THE HUMAN CELL ATLAS IS CONTRIBUTING TO.

Image Credit: Lindeboom, et al.[24]



MOHAMMAD LOTFOLLAHI
Scientist
Helmholtz Munich/Wellcome Sanger institute
Director of Machine Learning
Relation Therapeutics

FLG: First of all, we wanted to talk about scArches. Can you describe why you developed it? What problem was it trying to address and how does it work?

Mo: *The single-cell community and the consortia, such as the Human Cell Atlas and CZI, are trying to build a comprehensive reference of all the cells from our body, tissue by tissue. Once these atlases are built, ideally, you want to share them with the community so that they will be able to use it. However, how to use them is not straightforward. You can think of the atlases as a map, and you need an algorithm that can project the new data on that map. Since the data have been produced in a different lab, there is a batch effect or technical effect between them. The whole purpose of scArches was a to make these atlases usable and to share them with the community. Most people aren't able to build these large resources, and this tool allows us to democratize the usage of atlases.*

How does it work? The concept of the map is the best way to explain it. When you have a map, you want to fix the coordinates, and you don't want to change the map every time you have new data. So, scArches receives a prebuilt atlas and receives the new data, which it maps on the top of this reference atlas by correcting the technical effecting differences between the query and the reference.

By mapping new data to the existing atlas, you can transfer the cell type labels from the reference to the query, and that replaces the couple of months it would take someone to analyse the query data to get those cell type annotations. Also, when you put something in the context of a bigger thing, you might be able to find sub populations. For example, let's say you have beta cells in the query, and there are multiple sub populations of beta cell in the reference. When you map the query to the reference, it either aligns with the ones in the reference, or it separates out from the rest, which might give you a clue that this is a subpopulation that you would have never found out without mapping it to the reference

Another usage is for mapping disease data on the top of the healthy atlas. Diseases affect different cell types differently, and not all the cell types are affected by each disease. And by mapping your disease data on the top of the healthy atlas, you can find the ones that were affected since they will not be aligned. Based on that you can infer the cell states that are risk for disease.

Additional tools have been created this year to leverage reference atlases to automate the processing, analysis and interpretation of single-cell data. CellTypist[25] allows automated cell type annotation using a curated database of cell types and marker genes. Azimuth is a web application and presents a library of 13 datasets, facilitating reference-based mapping using these data. SEAcells[26] allows you to infer RNA and epigenomic states using reference data.

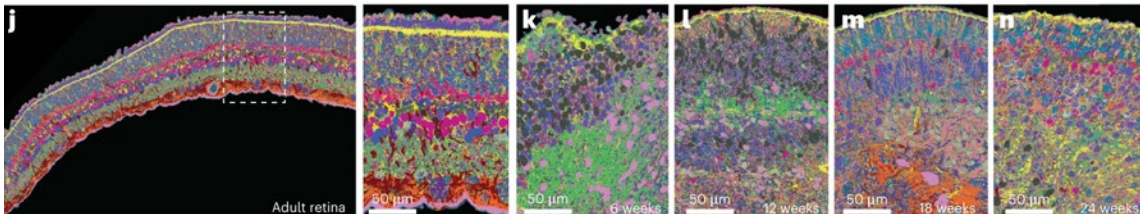
Furthermore, interactive and user-friendly tools to access large-scale and atlas-level datasets are also on the rise. These include tools such as cellxgene[27], a Chan-Zuckerberg Initiative tool, which allows interactive data manipulation of millions of cells. Also, TooManyCellsInteractive[28], which allows users to manipulate single-cell data and visualise it as an interactive radial tree, and Cellar[29], an interactive cell assignment tool.

Multiplex - The dawn of mass panel spatial mapping

Spatial technology has improved on several fronts, including in situ sequencing methods with enhanced resolution to spatially locate whole transcriptomes. Probe-based methodologies have also advanced, allowing large-scale hi-plex spatial experiments.

Commercial technologies, such as Nanostring CosMx™, Resolve Molecular Cartography™, RNAscope Hiplex and Vizgen MERSCOPE™, allow 10's or 100's or even 1000's of RNA probe-based markers. Furthermore, the spatial proteomics tools, such as Canopy CellScape™, Akoya Biosciences PhenoCycler®, Lunaphore Comet™, Standard Biotools the Hyperion™ and Miltenyi Biotec MACSima™, allow 20-100+ protein markers in hi-plex. Figure 4 shows the stunning images being generated with this type of technology; in this case, the human retina.

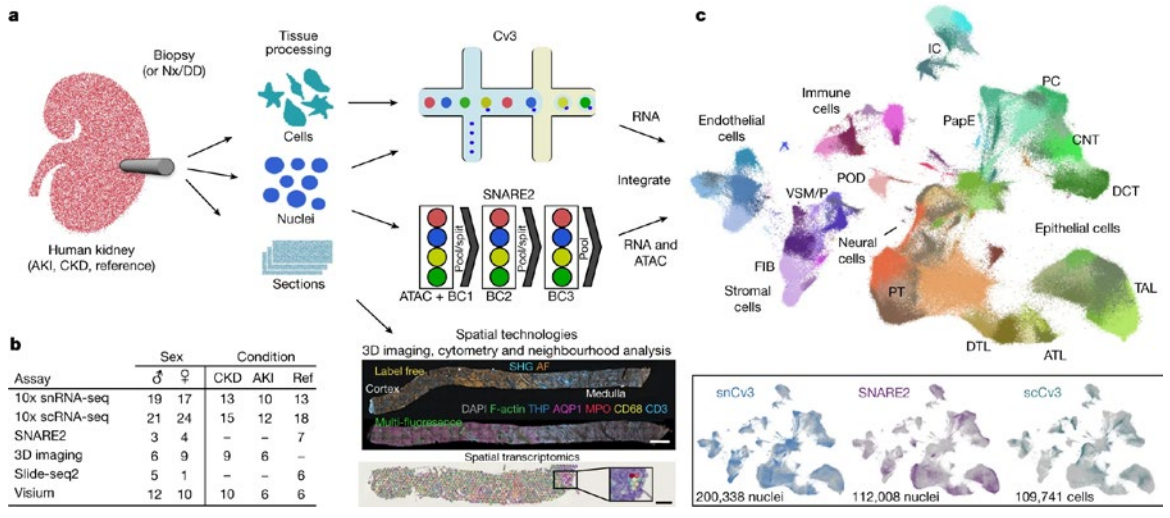
FIGURE 4. HIGHLY MULTIPLEXED IMMUNOHISTOCHEMISTRY OF A PRIMARY ADULT HUMAN RETINA SECTION.
Image Credit: Wahle, et al. [30]



One of the goals of the HCA going forward is to incorporate both multiomics and spatial data into these largely singe-cell RNA based resources. As of 2023, several of these atlases are now adding spatial data. These include the human heart, using 10x Genomics Visium[4], the human lung[31] and placenta[32], using Visium and RNAscope, and the human breast using Visium, Molecular Cartography, RNAscope, MERFISH and Phenocycler[33] - an example of the value of using many spatial technologies in one study.

As mentioned above, the HuBMAP consortium is leading the way in producing spatially resolved cell atlases. With 400 individuals across 60 institutions, the goal is to generate high-resolution spatial atlases of the healthy human adult body[14]. Recent spatial maps include, the intestine, which was mapped using Molecular Cartography and CODEX technology[34], the maternal-fetal interface using Multiplexed Ion Beam Imaging (MIBI) technology[35], and the kidney, using SLIDE-seq2, Visium, RNAscope and large scale 3D tissue cytometry[5]. Figure 5 displays an overview of this experiment on the kidney.

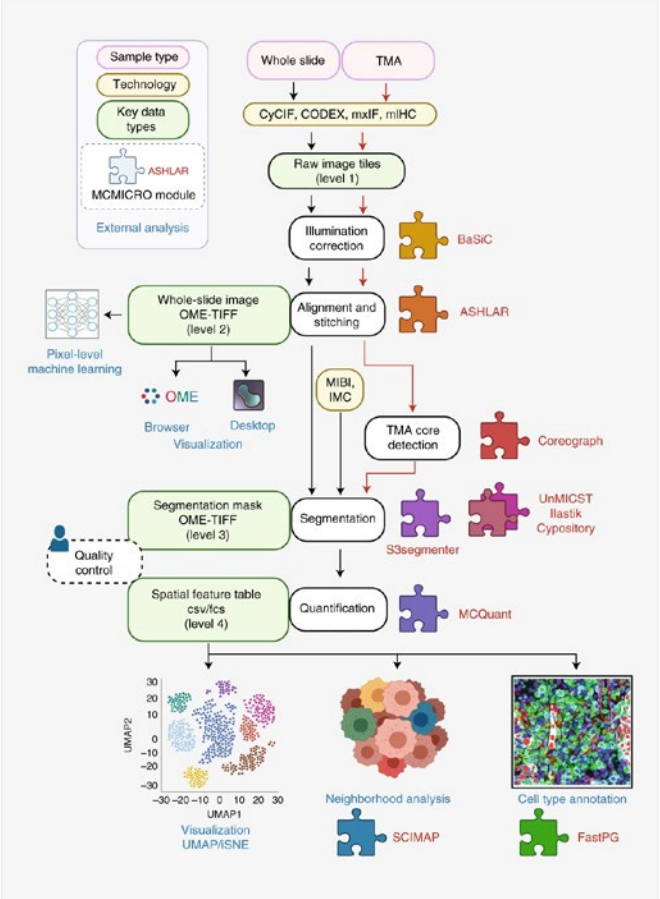
FIGURE 5. HUMAN KIDNEY ATLAS USING SINGLE-CELL AND SPATIAL.
*(**A**) Human kidney samples from deceased donors (DD) or nephrectomies (Nx) were used for single-cell 3D imaging or Slide-seq2 and Visium. (**B**) Summary of samples. (**C**) Integration of Omic RNA data using joint UMAP embedding to align cell types across the different data modalities. Image Credit: Lake, et al. [5]*



Furthermore, there are large scale efforts performed outside of these two consortia. For example, the whole mouse brain (~8 million cells) had a spatial atlas produced with MERFISH[36]. In cancer biology, recent examples of spatial analysis using imaging mass cytometry have examined ~1.1 million brain tumour cells[37] and ~1.6 million lung adenocarcinoma cells[38].

FIGURE 6: THE CANONICAL WORKFLOW OF MCMICRO.

MCMICRO allows end-to-end image processing of multiplexed whole slide images. Image Credit: Schapiro, et al. [50]



HANDLING LARGE-SCALE SPATIAL DATA

The large-scale spatial datasets are amazing resources, like their single-cell counterparts. To help standardise experiments across this space and allow wide usage of these resources, the community has recently created a Minimum Information for highly multiplexed Tissue Imaging[39] (MITI). However, analysing spatial data to make the most of its additional spatial dimension is still a big challenge in the field. We will quickly review some of the newest tools on the block.

One very basic challenge in spatial data is cell segmentation. With spatial methods now achieving sub-cellular resolution, i.e., with 'spots' inside of cells, the boundary of a cell is more important to define than ever. This is difficult when cells are crowded. In spatial proteomics, membrane markers allow distinction of cell boundaries, and tools such as RAMCES[40] help select the optimal markers for your tissue. For transcriptomics, often a nucleus or membrane stain is used and cells are segmented with tools such as Cellpose[41] or using deep-learning models such as Mesmer[42]. Recent tools such as SCS[43] use machine learning to adaptively learn how spots place within cells to more efficiently capture cell boundaries.

Other recent tools address specific challenges in the spatial analysis pipeline. CellSighter[44] uses deep learning to better classify cells in hi-plex images. Bento[45], a toolbox specific for subcellular analysis, can define subcellular domains and gene-gene colocalisation patterns. Visinity[46] is a whole-slide visual analytics system to analyse cell interaction patterns. Finally, STELLAR[47], a deep learning-based cell annotator, 'learns' cell types across different regions, tissues and samples, and automatically labels cells.

Several 'toolboxes' or frameworks exist to perform multiple analysis steps and sophisticated downstream analysis in one platform. There are commercial options, such as HALO, which tend to be more refined but less flexible, or there are several options produced by the community. Recent, up-to-date examples include Squidpy[48], which builds upon an existing package, Scanpy, to allow visualisation of gene expression data at spatial locations. Giotto[49] is a toolbox for spatial data and allows identification of spatially variable genes and gene-expression/cell type colocalization. Finally, MCMICRO[50], which represents an end-to-end pipeline to transform multi-channel whole-slide images into single-cell data. MCMICRO is scalable and allows new packages to be incorporated where desired. We spoke to the lead developer of MCMICRO, Dr. Denis Schapiro, about the issues of cell segmentation and how to handle hi-plex images.

DENIS SCHAPIRO
Research Group Leader
Heidelberg University Hospital

"MCMICRO IS CONVENIENT AS IT PROVIDES ACCESS TO VARIOUS PREPROCESSING, SEGMENTATION AND DOWNSTREAM ANALYSIS ALGORITHMS, ALLOWING THE USER TO TRY OUT DIFFERENT METHODS WITHIN THE SAME ENVIRONMENT."

FLG: One of the surprising things for outsiders is that cell segmentation isn't completely robust yet. How close do you think we are to being able to robustly detect cells of different sizes and types in a tissue section?

Denis: *As you said, cellular heterogeneity in a sample will always remain a challenge and additionally, we are working with a 2D representation of a 3D space. So if by robust segmentation you mean to have a single universal model that is able to perform an accurate segmentation on any given tissue in 2D, then that is still a few steps away. Fortunately, recent developments, allow retraining of segmentation models to make it more suitable for specific samples and cell types and 3D imaging methods are also emerging. At the moment, the question is rather, is the segmentation good enough to answer your specific question. One example would be looking at tumors and detecting 100,000 tumor cells. It doesn't matter whether you find 104,000 or 100,000 tumor cells, but you may care whether you find 10 or 20 cells of a specific rare immune cell type. I would suggest thinking about what you want to discover and where segmentation needs to be very accurate and where it can be more forgiving.*

I think the other part is that 3D is the next frontier. There you could theoretically be able to map all cells very accurately. If you had good membrane markers, it could work very well. It may be challenging for cells that don't have a nucleus, or have multiple nuclei, but if there's a really good membrane staining it should work quite well.

Lastly, we should also be thinking about methods that do not require segmentation. Direct calling of cell types or directly calling specific regions based on machine learning approaches are already emerging.

FLG: For someone who's quite new to the field, what computational tools and advice would you suggest to help them handle a high-plex image?

Denis: *I think the key part is to just get started and run the analysis from the beginning to the end. This means accepting - at first - that your analysis may not be perfect. But once you run it from beginning to the end, you will understand what issues you need to prioritize, which issues are minor and should be addressed last, and what issue cannot be fixed so you may need to stop analysis and rather rerun the experiment. Getting your feet wet and running the data as soon as possible from beginning to the end will help you to really understand the quality and complexity of your data. Therefore, a dynamic, flexible, and scalable workflow is required.*

To address these challenges, we have developed a pipeline called MCMICRO, which enables end-to-end processing of various multiplexed imaging methods. MCMICRO is convenient as it provides access to various preprocessing, segmentation and downstream analysis algorithms, allowing the user to try out different methods within the same environment. My research group as well as the Laboratory of Systems Pharmacology at the Harvard Medical School, which is led by Prof. Peter Sorger, continue to maintain, develop, and expand the pipeline. While MCMICRO is primarily command-line based, a team at OHSU, led by Prof. Jeremy Goecks, has developed a Galaxy implementation of MCMICRO, which provides a graphical-user interface to the pipeline.

However, if you're just starting out, and you only have a few images to analyze, another great tool to start with is QuPath. QuPath has fantastic online tutorials and I think that it can get you very far if you don't have to scale it up crazily and you don't require specific tools for image processing or segmentation. ▶

The other thing is you need to think about downstream analysis and the corresponding questions you want to answer. E.g., how will you identify your cell types of interest, how are you planning to perform neighborhood analysis and so on. Here I would suggest exploring the wide array of open-source tools available to our community. If you get stuck, open an issue either directly on the corresponding Github repository or on image.sc, which is a great resource and community.

Of course, there are also commercial tools, which should be explored if you have the necessary funding or work in a commercial environment.

To summarize, I would start with QuPath and see how far you get. If you need more flexibility and scale up, move to MCMICRO.

Ultimately, the extra dimension of spatial data calls for interesting and novel ways to analyse it. MISTy[51] is a machine learning framework which constructs multiple views of spatial data based on function context (Intrinsic, Local, Tissue – see Figure 7). From here, spatial relationships can be worked out within each of these views from highly multiplexed data. We spoke to lead developer, Dr. Jovan Tanevski, about highly multiplexed data and how MISTy works.

JOVAN TANEVSKI
Research Area Leader, Saez-Rodriguez Group
Institute for Computational Biomedicine, Heidelberg University and Heidelberg University Hospital

FLG: What are some of the unique challenges in having highly multiplexed data, with that many signals all in one spatial area?

Jovan: *Well, spatial biology is not new. In histology and histopathology, it is routinely considered by tissue staining, or by immunohistochemistry or immunofluorescence panels. These are in turn used for diagnosis, estimation of disease progression and assignment of patients to treatment groups. But they capture a limited number of variables or observations at every spatial location, ranging from two to, let's say, ten - on a good day.*

What we have now with the new technologies, capturing 10's, to 100's to potentially 1000's of different molecular markers - this is definitely an unprecedented view of tissue biology. This, again, opens up the opportunity to ask the question - how can we improve our understanding of the relationship between structure and function? With the limited molecular panels, we could identify general cell types, a limited number of functional states of the cells and then reason on top of that. Now we have a more complete view of the molecular making of each cell within the tissue.

Given these new technologies, that are both highly dimensional and deliver high resolution, we need computational tools that are adequate for approaching this type of data. Not necessarily completely data driven, but considering available domain knowledge, building upon current practices and adding an additional layer of information that we can learned from this new data.

This means that the models we are building, based on these extended molecular panels and the extended spatial resolution, have to be interpretable and communicable - in a way that the domain experts that are using these models can apply them in their everyday work. Potentially also for future clinical applications.

FLG: Jumping off from that, could you describe MISTy, the tool you developed, in basic terms? What does it allow you to do with that multiplex spatial data?

Jovan: *With MISTy we can start to explore the relationships that exists in the data at different spatial contexts. This is the basic idea that motivated MISTy. Let's see what makes these cells tick and how do they play a role within the tissue structure. Are there some relationships between the different molecules in space that we can capture?* ▶

So, MISTy is a multi-view explainable machine learning approach. MISTy builds specific views that captures observations in different spatial contexts. These observations can be abundances of molecular markers or be focused on different functional aspects of the data. Given these views, we try to answer three questions.

The first question is, does the spatial context allow us to learn something more about the state of the cell that we are observing? The second question is, from which spatial context does this information come from? Is it coming only from within the cell, within the immediate neighbourhood or from the broader tissue structure? Finally, what are the underlying relationships that lead to the contributions and the performance of the model.

MISTy is a predictive model and uses a regression task to judge whether we learn something from any of the spatial contexts. The underlying assumption being that if we can predict the abundance of a target molecule better by considering the abundance of all other molecules in different spatial contexts independently, then that means that there are potential relationships and additional information available in these different contexts. The significance of these relationships then corresponds to the estimated predictor-target importances in the view specific models.

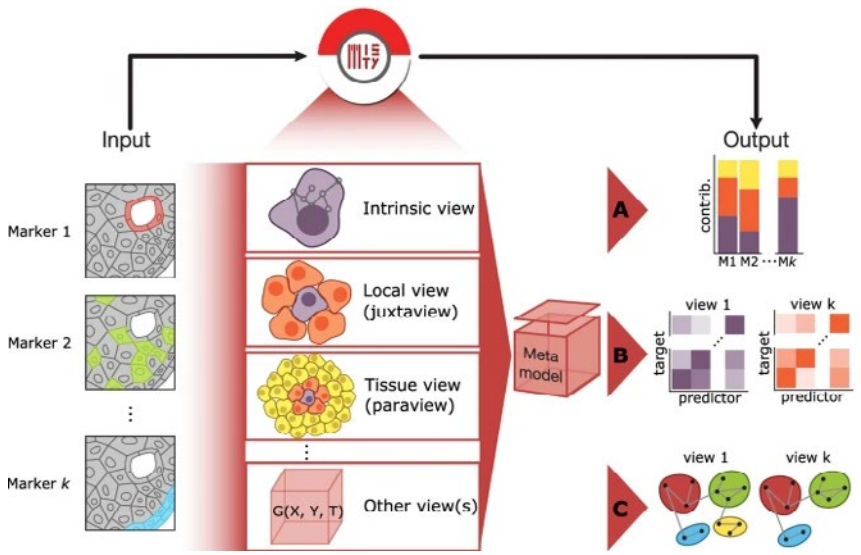
High dimensional spatially resolved data offers measurements of 100's or 1000's of different markers. In order to make our models interpretable, we needed to project this data into functionally relevant spaces. To this end, we use prior domain knowledge and combine it with the data-driven approach, in order to extract these functional relationships from the data. For example, instead of looking at the expression of 1000 genes, we can estimate the pathway activities within the cell given the expression of the genes. We can also represent each cell as the activity of the transcription factors. Making these projected values targets of our models, we can gain a look into the regulatory events that are happening within the cell and in different spatial contexts.

Furthermore, since we now have reference atlases that allow us to determine different cell subtypes, we can make the MISTy models specific to a certain cell and regions of interest. We can then explore what are the specific relationships for a specific cell type of interest and compare them across different spatial regions.

Finally, MISTy is scalable and flexible. It can be applied to the different types of omics data. Unlike some related methods it works with relatively low number of samples, but it also scales well to 100's or 1000's of samples.

FIGURE 7. OVERVIEW OF MISTY, A MULTIVIEW FRAMEWORK FOR MULTIPLEXED SPATIAL DATA.

*MISTy takes multiple views of the tissues to extract (**A**) information about the expression of markers in a spatial unit at each view, (**B**) an estimate of the interactions amongst markers at each view and (**C**) a qualitative view of the communities of markers in each view. Image Credit: Tanevski, et al.* [51]



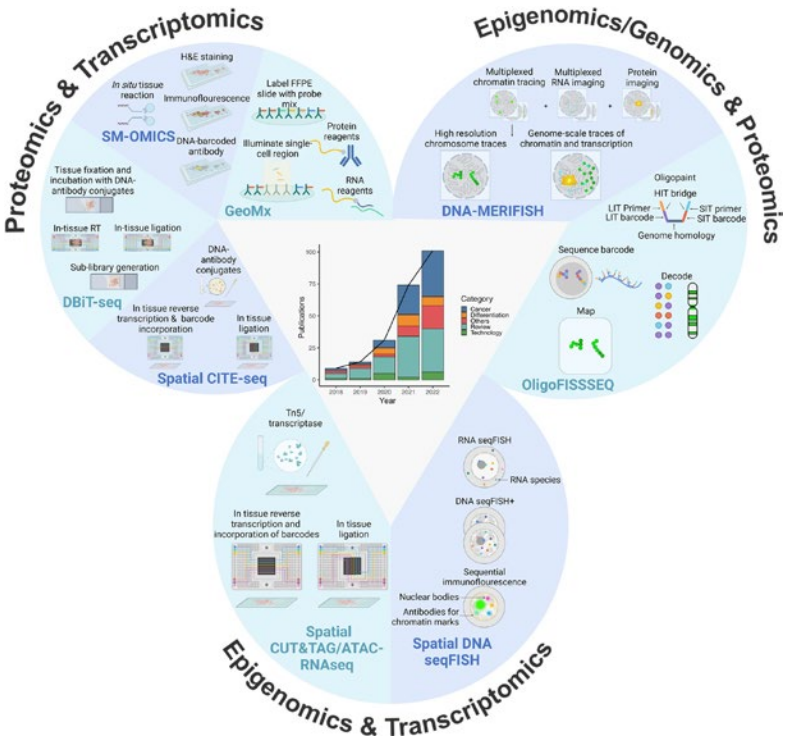
Multimodal - single-cell & spatial multiomics

Finally, we come to the multimodal expansion of single-cell and spatial. While RNA might have been first, DNA, protein, epigenomics and now metabolomics data are being sequenced and viewed spatially. Whether this is alone in mono-omic studies or together in multiomics, we are now trying to view a cell as a complex molecular entity with all the pieces of the puzzle accounted for.

Multiomics is already being integrated into several of the large-scale single-cell projects listed previously. However, the expansion of spatial multiomics methods is the major story of the last year. Figure 8 highlights some of these spatial methodologies for different omic combinations. Highlights include, spatial CITE-seq[52], which allows whole transcriptome and >1000 proteins to be measured at cellular resolution, and spatial-CUT&TAG/ATAC-RNAseq[53]. Chapter 5 will discuss this spatial epigenomic and transcriptomic tool in detail.

FIGURE 8. THE EXPLOSION OF MULTIOMICS SPATIAL TECHNOLOGIES OVER THE LAST 5 YEARS.

Representative spatial technologies for the different multiomics assessments. In the centre is a bar chart of the number of publications reporting spatial multiomics methods in different categories. Image Credit Li [54]



MULTIOMICS INTEGRATION

While each 'omic' has its own challenges for proper preprocessing and analysis, the true challenge in the multiomics space is trying to integrate these omics into one shared space. Data from different modalities varies in dimensions and data type – RNA is measured across 1000's of genes while proteins are measured across 100's, and DNA is a sequence of letters. Multiomics data also varies in distribution; for example, negative binomial (RNA) and Poisson (ATAC). If we want a holistic view of each single cell, this needs to be integrated.

The variety of options for multiomics data integration is beyond the scope of this playbook, for which readers should refer to recent reviews[55-58] and benchmarking[59]. Here we will briefly examine some recently released and popular tools.

Firstly, there are integration tools that integrate multiomics data within the same cell, within the same experiment. Examples of this include: MOFA+[60], TotalVI[61], WNN[62] and scMoMaT[63]. The other set of tools align multimodal profiles from different experiments with tools such as GLUE[64], MultiVI[65], Cobolt[66], Bridge[67] and Multigrate[68]. The advantage of these tools is the ability to impute missing modalities in data with only single modalities. It is worth bearing in mind that many of these tools are tailored to the specific multiomics data (i.e., many are specific for CITE-seq), yet Multigrate is not limited to specific assays, giving it general usability. We caught up with Dr. Lotfollahi again (next page), this time to ask about Multigrate.

Integrating multimodal data spatially is the next frontier for computational biologists. Spatial Glue[69] represents the first major attempt to achieve this, using graph neural network like Glue with dual attention to integrate omics within spatial domains.

MOHAMMAD LOTFOLLAHI

Scientist
Helmholtz Munich/Wellcome Sanger institute
Director of Machine Learning
Relation Therapeutics

FLG: A lot of people are suggesting solutions for the problem of multimodal data integration. Can you briefly describe your tool Multigrate and the general problem with data integration?

Mo: *Advancement in the technology has allowed us to capture data from these different modalities in each cell. This could be from the DNA space, for example, chromatin accessibility, whether that DNA region was transcribed into RNA, and then whether that RNA was translated it to a protein. Each modality gives us different information. An analogy would be that, for a person, you can have speech, you can have text, and you can also have a face image. They give you different information about each person and this is similar thing for the cell too.*

The question is, can we combine them to get the holistic view of the cell and how it behaves at different levels? But these modalities are different in nature, so DNA is a sequence, RNA is basically a matrix of counts and protein is also similar to RNA, but in a really different space. So, the question for multimodal integration is, how to integrate these data modalities together, to have one representation for the cell that combines all of those different modalities?

A challenge to that is that these cell atlases are usually built using RNA alone and not all the datasets have measurements from different modalities. Another challenge here is how to work with this partial information, and you would want an algorithm that can handle missing modalities. One last thing is the question of whether we can we actually understand the effect of each modality; for example, the effect of a disease might not be observable in RNA but might be observable in protein space.

With Multigrate, we wanted to extend these reference building efforts multimodally, to integrate all of these modalities while handling the missing modalities. It's a generative model. It learns a distribution per each modality, and it learns to combine those distributions into one unified distribution, which captures the intersection of all three different modalities. This then allows you to build a multimodal reference which you can project new data to, that shares some modalities with the reference. Then you can use the reference to impute the missing modalities in the query. This can help you to prioritise experiment design, because it's expensive to measure all these modalities together. And these types of algorithms can identifying the important modalities.

Your single source for single-cell multiomics

Introducing the new BD Rhapsody™ HT Xpress System

Learn more at **bdbiosciences.com/HTXpress**

Gene expression analysis: Targeted or WTA

CITE-seq reagents for multiomic research

Immune repertoire profiling

BD

Chapter 3 references

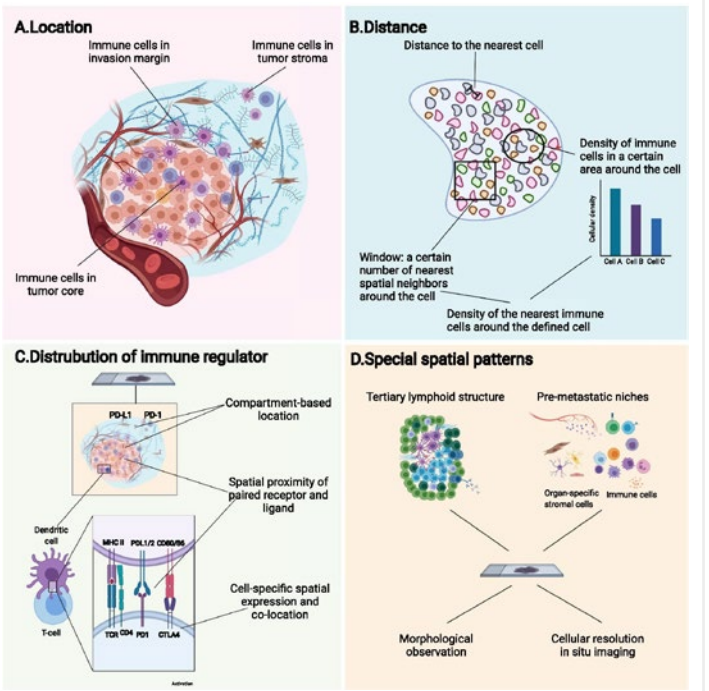
1. Vitak, S.A. et al. **Sequencing thousands of single-cell genomes with combinatorial indexing.** Nature methods **14,** 302-308 (2017).
2. Jia, Q., Chu, H., Jin, Z., Long, H. & Zhu, B. **High-throughput single-cell sequencing in cancer research.** Signal Transduction and Targeted Therapy **7**, 145 (2022).
3. Miranda, A.M.A. et al. **Single-cell transcriptomics for the assessment of cardiac disease.** Nature Reviews Cardiology **20**, 289-308 (2023).
4. Kanemaru, K. et al. **Spatially resolved multiomics of human cardiac niches.** Nature **619**, 801-810 (2023).
5. Lake, B.B. et al. **An atlas of healthy and injured cell states and niches in the human kidney.** Nature **619**, 585-594 (2023).
6. Reed, A.D. et al. **A Human Breast Cell Atlas Mapping the Homeostatic Cellular Shifts in the Adult Breast.** bioRxiv, 2023.04.21.537845 (2023).
7. Siletti, K. et al. **Transcriptomic diversity of cell types across the adult human brain.** bioRxiv, 2022.10.12.511898 (2022).
8. Sikkema, L. et al. **An integrated cell atlas of the lung in health and disease.** Nature Medicine **29**, 1563-1577 (2023).
9. Zilbauer, M. et al. A **Roadmap for the Human Gut Cell Atlas.** Nature Reviews Gastroenterology & Hepatology (2023).
10. Rood, J.E., Maartens, A., Hupalowska, A., Teichmann, S.A. & Regev, A. **Impact of the Human Cell Atlas on medicine.** Nature Medicine **28**, 2486-2496 (2022).
11. Dalerba, P. et al. **Single-cell dissection of transcriptional heterogeneity in human colon tumors.** Nature biotechnology **29**, 1120-1127 (2011).
12. Quake, S.R. **A decade of molecular cell atlases.** Trends in Genetics **38**, 805-810 (2022).
13. Shapiro, E., Biezuner, T. & Linnarsson, S. **Single-cell sequencing-based technologies will revolutionize whole-organism science.** Nature Reviews Genetics **14**, 618-630 (2013).
14. Jain, S. et al. **Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP).** Nature Cell Biology **25**, 1089-1100 (2023).
15. Rozenblatt-Rosen, O. et al. **The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution.** Cell **181**, 236-249 (2020).
16. Consortium*, T.S. et al. **The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans.** Science **376**, eabl4896 (2022).
17. Eraslan, G. et al. **Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function.** Science **376**, eabl4290 (2022).
18. Wang, R. et al. **Construction of a cross-species cell landscape at single-cell level.** Nucleic Acids Research **51**, 501-516 (2022).
19. Domcke, S. et al. **A human cell atlas of fetal chromatin accessibility.** Science **370**, eaba7612 (2020).
20. Elmentaite, R., Domínguez Conde, C., Yang, L. & Teichmann, S.A. **Single-cell atlases: shared and tissue-specific cell types across human organs.** Nature Reviews Genetics **23**, 395-410 (2022).
21. Lotfollahi, M. et al. **Mapping single-cell data to reference atlases by transfer learning.** Nature Biotechnology **40**, 121-130 (2022).
22. Lotfollahi, M. et al. **Biologically informed deep learning to query gene programs in single-cell atlases.** Nature Cell Biology **25**, 337-350 (2023).
23. Michielsen, L. et al. **Single-cell reference mapping to construct and extend cell-type hierarchies.** bioRxiv, 2022.07.07.499109 (2023).
24. Lindeboom, R.G.H., Regev, A. & Teichmann, S.A. **Towards a Human Cell Atlas: Taking Notes from the Past.** Trends Genet **37**, 625-630 (2021).
25. Xu, C. et al. **Automatic cell type harmonization and integration across Human Cell Atlas datasets.** bioRxiv, 2023.05.01.538994 (2023).
26. Persad, S. et al. **SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data.** Nature Biotechnology (2023).
27. Megill, C. et al. **cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices.** bioRxiv, 2021.04.05.438318 (2021).
28. Klamann, C., Lau, C. & Schwartz, G.W. **TooManyCellsInteractive: a visualization tool for dynamic exploration of single-cell data.** bioRxiv, 2023.06.16.544954 (2023).
29. Hasanaj, E., Wang, J., Sarathi, A., Ding, J. & Bar-Joseph, Z. **Interactive single-cell data analysis using Cellar.** Nature Communications **13**, 1998 (2022).
30. Wahle, P. et al. **Multimodal spatiotemporal phenotyping of human retinal organoid development.** Nature Biotechnology (2023).
31. Madissoon, E. et al. **A spatially resolved atlas of the human lung characterizes a gland-associated immune niche.** Nature Genetics **55**, 66-77 (2023).
32. Arutyunyan, A. et al. **Spatial multiomics map of trophoblast development in early pregnancy.** Nature **616**, 143-151 (2023).
33. Kumar, T. et al. **A spatially resolved single-cell genomic atlas of the adult human breast.** Nature **620**, 181-191 (2023).
34. Hickey, J.W. et al. **Organization of the human intestine at single-cell resolution.** Nature **619**, 572-584 (2023).
35. Greenbaum, S. et al. **A spatially resolved timeline of the human maternal–fetal interface.** Nature **619**, 595-605 (2023).
36. Zhang, M. et al. **A molecularly defined and spatially resolved cell atlas of the whole mouse brain.** bioRxiv, 2023.03.06.531348 (2023).
37. Karimi, E. et al. **Single-cell spatial immune landscapes of primary and metastatic brain tumours.** Nature **614**, 555-563 (2023).
38. Sorin, M. et al. **Single-cell spatial landscapes of the lung tumour immune microenvironment.** Nature **614**, 548-554 (2023).
39. Schapiro, D. et al. **MITI minimum information guidelines for highly multiplexed tissue images.** Nature Methods **19**, 262-267 (2022).
40. Dayao, M.T., Brusko, M., Wasserfall, C. & Bar-Joseph, Z. **Membrane marker selection for segmenting single cell spatial proteomics data.** Nature Communications **13**, 1999 (2022).
41. Petukhov, V. et al. **Cell segmentation in imaging-based spatial transcriptomics.** Nature Biotechnology **40**, 345-354 (2022).
42. Greenwald, N.F. et al. **Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning.** Nature Biotechnology **40**, 555-565 (2022).
43. Chen, H., Li, D. & Bar-Joseph, Z. **SCS: cell segmentation for high-resolution spatial transcriptomics.** Nature Methods **20**, 1237-1243 (2023).
44. Amitay, Y. et al. **CellSighter: a neural network to classify cells in highly multiplexed images.** Nature Communications **14**, 4302 (2023).
45. Mah, C.K. et al. **Bento: A toolkit for subcellular analysis of spatial transcriptomics data.** bioRxiv, 2022.06.10.495510 (2023).
46. Warchol, S. et al. **Visinity: Visual Spatial Neighborhood Analysis for Multiplexed Tissue Imaging Data.** IEEE Trans Vis Comput Graph **29**, 106-116 (2023).
47. Brbić, M. et al. **Annotation of spatially resolved single-cell data with STELLAR.** Nat Methods **19**, 1411-1418 (2022).
48. Palla, G. et al. **Squidpy: a scalable framework for spatial omics analysis.** Nature methods **19**, 171-178 (2022).
49. Dries, R. et al. **Giotto: a toolbox for integrative analysis and visualization of spatial expression data.** Genome biology **22**, 1-31 (2021).
50. Schapiro, D. et al. **MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging.** Nature Methods **19**, 311-315 (2022).
51. Tanevski, J., Flores, R.O.R., Gabor, A., Schapiro, D. & Saez-Rodriguez, J. **Explainable multiview framework for dissecting spatial relationships from highly multiplexed data.** Genome biology **23**, 1-31 (2022).
52. Liu, Y. et al. **High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq.** Nature Biotechnology (2023).
53. Zhang, D. et al. **Spatial epigenome–transcriptome co-profiling of mammalian tissues.** Nature **616**, 113-122 (2023).
54. Li, X. **Harnessing the potential of spatial multiomics: a timely opportunity.** Signal Transduction and Targeted Therapy **8**, 234 (2023).
55. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. **Methods and applications for single-cell and spatial multi-omics.** Nature Reviews Genetics, 1-22 (2023).
56. Argelaguet, R., Cuomo, A.S.E., Stegle, O. & Marioni, J.C. **Computational principles and challenges in single-cell data integration.** Nature Biotechnology **39**, 1202-1215 (2021).
57. Miao, Z., Humphreys, B.D., McMahon, A.P. & Kim, J. **Multi-omics integration in the age of million single-cell data.** Nature Reviews Nephrology **17**, 710-724 (2021).
58. Wang, X., Wu, X., Hong, N. & Jin, W. **Progress in single-cell multimodal sequencing and multi-omics data integration.** Biophysical Reviews (2023).
59. Luecken, M.D. et al. **Benchmarking atlas-level data integration in single-cell genomics.** Nature Methods **19**, 41-50 (2022).
60. Argelaguet, R. et al. **MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data.** Genome biology **21**, 1-17 (2020).
61. Gayoso, A. et al. **Joint probabilistic modeling of single-cell multi-omic data with totalVI.** Nature Methods **18**, 272-282 (2021).
62. Hao, Y. et al. **Integrated analysis of multimodal single-cell data.** Cell **184**, 3573-3587.e29 (2021).
63. Zhang, Z. et al. **scMoMaT jointly performs single cell mosaic integration and multi-modal bio-marker detection.** Nature Communications **14**, 384 (2023).
64. Cao, Z.-J. & Gao, G. **Multi-omics single-cell data integration and regulatory inference with graph-linked embedding.** Nature Biotechnology **40**, 1458-1466 (2022).
65. Ashuach, T., Gabitto, M.I., Jordan, M.I. & Yosef, N. **MultiVI: deep generative model for the integration of multi-modal data.** bioRxiv, 2021.08.20.457057 (2021).
66. Gong, B., Zhou, Y. & Purdom, E. **Cobolt: integrative analysis of multimodal single-cell sequencing data.** Genome Biology **22**, 351 (2021).
67. Hao, Y. et al. **Dictionary learning for integrative, multimodal and scalable single-cell analysis.** Nature Biotechnology (2023).
68. Lotfollahi, M., Litinetskaya, A. & Theis, F.J. **Multigrate: single-cell multi-omic data integration.** bioRxiv, 2022.03.16.484643 (2022).
69. Long, Y. et al. **Integrated analysis of spatial multi-omics with SpatialGlue.** bioRxiv, 2023.04.26.538404 (2023).

FRIENDLY NEIGHBOURHOOD CELLS. CELL-CELL COMMUNICATION AND THE CELLULAR MICROENVIRONMENT

CELL TYPES IN EVERY TISSUE OF THE BODY HAVE BEEN ELUCIDATED WITH TRANSCRIPTOMIC AND PROTEOMIC DATA FROM SINGLE CELLS. HOWEVER, CELLS DO NOT ACT IN ISOLATION, AND KNOWING WHAT CELL TYPE YOU ARE LOOKING AT ONLY MAKES UP HALF THE STORY.

FIGURE 1. METRICS THAT CAN BE ISOLATED FROM ANALYSING THE TISSUE MICROENVIRONMENT USING SPATIAL TECHNOLOGY.

*The TME is used as an example here. The spatial architecture of the TME can be described according to the (**A**) location of cells, (**B**) distance between cells, (**C**) distribution of specific markers e.g., immunoregulators, and (**D**) specific spatial patterns. Image Credit: Fu, et al. [1]*



Tissues are living, dynamic entities, and cells are in **constant** communication with each other, taking the form of ligand secretion over short distances or endocrine signalling sent across the body. Measuring cell interactions is the next layer of information to map before we can say we fully understand the cellular composition of a tissue.

In this chapter, we will look at the tissue microenvironment and how spatial biology is taking us beyond cell types to help us look at cell neighbourhoods and niches. We will also look at another burgeoning area of analysis, cell-cell communication. We will showcase some of the methodologies to quantify this from single-cell data, as well as the valuable role that spatial biology has played to identify cells in immediate communicable distance to one another.

Friendly neighbourhoods - the tissue microenvironment

Single-cell sequencing has been revolutionary for understanding cells. Unfortunately, it necessitates the liberation of viable cells from the tissue, meaning context is lost for investigating cell interaction.

Spatial biology addresses this issue. No longer are cells quantified as isolated entities based on their transcriptomic profile, but they are now visualised as members of a community of cells. Their placements and neighbourhoods can be analysed alongside their expression profile.

Spatial transcriptomics, proteomics and multi-omics are all frequently used to study aspects of cellular niches, providing different levels of insight. Cell types can be visualised based on their location in tissue, their co-localization or avoidance of other cell types, and the niches they form with other cells in their vicinity.

Figure 1 illustrates a variety of metrics and information about cells that can be extracted from spatial profiling of tissue microenvironments. With 100s of markers visualised in a single section of tissue, drawing these insights from spatial data requires specialist computational tools. Tools such as MISTy[3] can view expression markers in a local and tissue-wide view, to identify cellular communities and niches based on location and distance. Holistic software frameworks such as Giotto[4] and Squidpy[5] allow for interactive visualisation of spatial data and for identifying cell neighbourhoods, specific spatial patterns and the effect of cell neighbours on gene expression.

Ultimately, downstream analysis is still lagging behind technology advancement, relying mainly on distances between cells to learn about communities. New analysis tools such as SPIAT[6] explicitly address this problem and have a wider range of functions for neighbourhood analysis from basic cell typing and cell colocalizations to neighbourhood composition, niche heterogeneity and tissue region variations. However, it appears we are at the early stages of analyses to identify cell neighbourhood composition and interaction.

Studying cellular neighbourhoods is useful to many areas of biology. We asked Dr. Haiqi Chen about the value that studying cellular niches has for his area of research, reproductive biology.

HAIQI CHEN

Assistant Professor
UT Southwestern Medical Centre

FLG: For your research, what value do spatial technologies provide when studying the reproductive system? For example, what is the value of looking at cell neighbourhoods and niches?

Haiqi: *Reproductive biology is broad enough that I cannot give you a generalised answer. Instead, I can give an example of what we've been studying. We are interested in how stem cells are regulated to produce sperm in the testes. Traditionally, when people study how stem cells are regulated by their microenvironment, they typically do gene deletion using animal models. It's quite powerful, and most of our basic knowledge about gene functions in regulating stem cells behaviours has come from this approach. But you can easily imagine the drawbacks, namely that it's really low throughput, it's historically been hard to generate genetic animal models and its time consuming. Right now, it's not so hard with the wide adoption of CRIPSR technologies, it's just time consuming to generate an animal model with a deletion for every single gene. You just can't do that.*

Now with spatial technologies, we can really recapitulate the microenvironment from the ground up. We can see how the stem cells interact with other cell types surrounding them and then you could easily nominate molecular interactions that may be important. Or at least confirm these molecular interactions are present at a specific spatial location. This is more informative and can tell you that these are the ligand-receptor interactions that you will want to look at, instead of just going blind i.e. 'I'm going to knock out this gene and see what happens.' This is really helpful.

Another thing, there are just not that many ways to study human biology, especially for those tissues that are hard to model. For example, in my field, there's no reliable in vitro model to recapitulate human spermatogenesis - how sperm is produced from the stem cells all the way to the mature sperm. You just cannot do it in vitro. So, having access to human tissue samples and using spatial technology to really understand the molecular interactions within the human tissue sample is really powerful. With access to human tissue, we can look at spatially resolved cell types, spatially resolved transcriptome and other modalities such as protein, all together within intact tissue. That will give you so much more information than has been previously done.

To further put current cell neighbourhood analysis into context, we will now provide an overview of an application of spatial technology to deconvolute a tissue microenvironment; specifically, the tumour microenvironment.

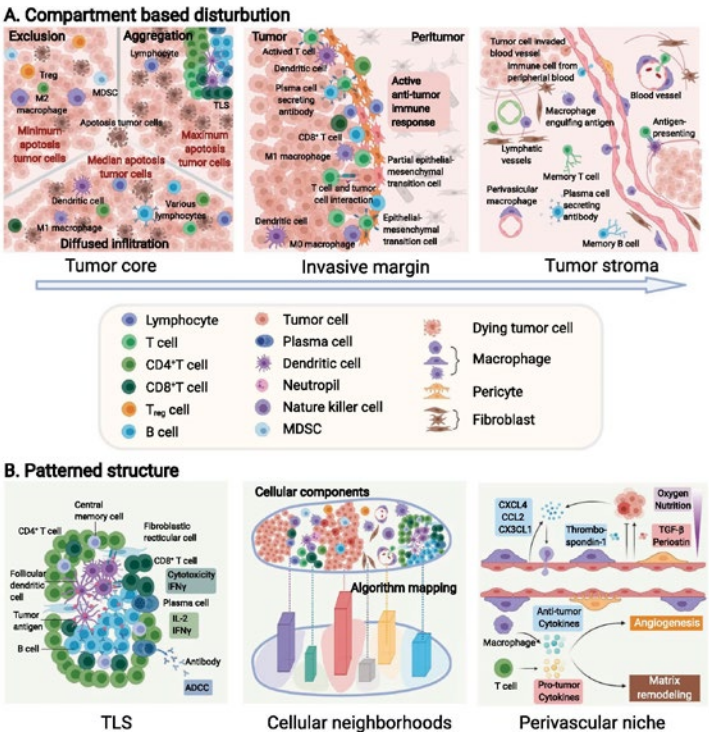
Not-so-friendly neighbourhoods - the tumour microenvironment

How tissue microenvironments differ in disease states is the focal point for many individuals working with spatial technologies. The Tumour Microenvironment (TME) is front and centre amongst targets to deconvolute. Tumour cells depend on the microenvironment for growth and metastasis. The constitution of these microenvironments is tissue-specific, consisting of many cell types clustered together[7]. Information such as the spatial distribution of tumour cells and the location of infiltrating immune cells can tell us a lot about cancer progression and potential treatment responses.

Modern spatial technologies overcome the low-plex and low-throughput limitations of historical approaches, allowing the study of the spatial heterogeneity of cancer. Figure 1 displays four key aspects of the spatial architecture of the TME that can be studied with spatial technologies; cell locations, cell distance/ neighbourhoods, protein distribution and cell microenvironments. Spatial omics not only quantifies and locates cells, but also reveals functions and potential intercellular reactions. Figure 2 demonstrates how this can be used to piece together aspects of the TME.

FIGURE 2. INSIGHTS THAT CAN BE LEARNED ABOUT THE SPATIAL ARCHITECTURE OF TME, SPECIFICALLY IMMUNE CELL LOCALISATIONS.

*(**A**) Tumours can be divided into different spatial compartments: tumour core, stroma and invasion margin. (**B**) Immune localisations within these structures such as tertiary lymphoid structures and perivascular niches have unique patterns which spatial biology has revealed. This means each compartment has unique cellular components and cellular neighbours. Image Credit: Fu, et al.[1]*



Reviews on the topic provide extensive details[2,8,9], but some broad findings from spatial analysis of the TME include:

- Multi-marker analysis has identified potential targets for immune therapy against solid tumours (e.g., Kinkhabwala, et al.[10])
- The tumour interface has a unique microenvironment and has markers that distinguish tumour advancement (e.g., Hunter, et al.[11]).
- There is cellular heterogeneity across the tumour structure with cell composition associated with differences in cancer progression and immunosuppression (e.g., Nirmal, et al.[12]).
- Immune cells that invade the tumour are compartmentalised in a tumour specific way (e.g., Dhainaut, et al.[13]).
- TME cellular composition, spatial location and interactions are associated with disease prognosis (e.g., Danenberg, et al.[14]).

MAI CHAN LAU

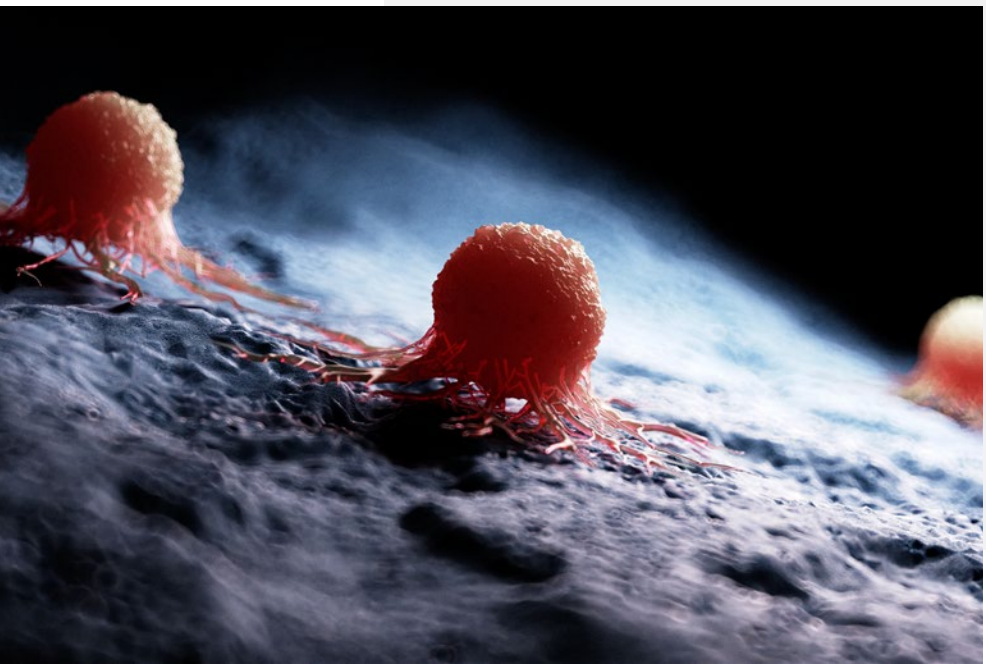
Assistant Principal Investigator
A*STAR's Bioinformatics Institute (BII) and Singapore Immunology Network (SIgN)

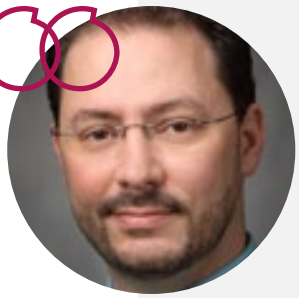
FLG: Could you briefly summarize the value of modern spatial technology for studying the tumour microenvironment? And what is the value of looking at these cellular neighborhoods?

Mai Chan: *Advanced spatial technology enables high-resolution study of the tumor microenvironment, offering insights unattainable through single-cell technology alone. While single-cell methods can describe the behavior and phenotypic characteristics of a particular cell type, they don't reveal how these cells interact with each other. Knowing which immune cells are actually in contact with the tumor is crucial; otherwise, a physical barrier like fibroblasts or fatty tissue could prevent effective tumor cell killing. Further combining spatial data with histological information is therefore invaluable for understanding the role of cell-to-cell interaction and physical location within the tumor microenvironment.*

FLG: There's a variety of spatial omics tools to study the tumor microenvironment, how do people go about selecting one?

Mai Chan: *The existing tools serve different purposes. Microdissection-based spatial transcriptomics technologies are more robust because they are closer to the bulk sequencing. However, these technologies often require a very well-defined hypothesis of what you want to look at, and that always comes with some prior knowledge on the tissue either via histopathological assessment or additional assays on serial tissue sections. It would work if you already knew the regions you are studying and wish to compare them to see how they are different. Hence, it's a hypothesis validation approach. For higher resolution spatial transcriptomics technologies, these are more exploratory. They give you more flexibility and chances to finding novel insights. When selecting technologies, it's essential to consider their compatibility with two types of tissues: fresh frozen and formalin-fixed paraffin-embedded. Additionally, it's important to determine if the technologies are based on NGS (Next-Generation Sequencing) or target probe methods. Whilst spatial proteomics technologies can be divided into two main categories: fluorescence-based and mass-based approaches. These technologies provide detailed molecular insights at a sub-cellular level, capturing information from 5 to 100 plex.*





JARED K. BURKS

Professor & Co-Director, Flow Cytometry & Cell Imaging Core Facility
The University of Texas MD Anderson Cancer Centre

FLG: You must get a lot of people wanting help with investigating the tumour microenvironment. Everyone knows this will depend on the research question, but is there some common advice you would give someone if they brought you a tumour sample for studying the TME?

Jared: *Yes, and no. I have to get the researcher to define the problem, and a lot of them come in with 'I want to see what happens'. We also have the problem that these are like the first images of Mars, we've not mapped a lot of these tissues in a high-plex assay. So, we don't know how the macrophages relate to tumour cells and how they are influenced or communicate with T cells and what happens when we add five B cells to this mix. So, there is some validity to 'see what happens'. But we have to have something that we can measure. So most often, I start with a random list of antibodies/protein targets/genes. We will want to look at structural markers. Structural markers tell us where things occur. We will want to look at phenotypic markers, which tells us which cells are present in those locations. And then ultimately, we will want to look at functional markers, which tells us what the cells are doing. If we can look at a structural, phenotypic and functional set. Then we can look at tissue samples from a long-term survivor versus a short-term survivor. Or we can look at immune therapy patient's tissue versus a traditional therapy patient and ask questions. How are their cellular neighbourhoods physically different and what type of correlative do we find amongst those two different groups? That's something that we can actually put a physical number to, whether that's in the characterization of four or five cells together, or if that's just a single T cell to tumour cell distance metric, which is one of the low hanging fruit measurements.*

Recent work[15] has used spatial in situ methods to generate a 3D model of TME cellular neighborhoods by profiling subsequent sections of tumours. While 2D approaches can help locate a selection of cellular neighborhoods, cells live in 3D space and have neighbors in all directions. Valuable insights can be gained from taking the 3D approach. We spoke to Tancredi Pentimalli, first author of this study, to hear why.

TANCREDI MASSIMO PENTIMALLI

PhD Student, Nikolaus Rajewsky Lab, **Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Centrum (MDC), Berlin School of Integrative Oncology (BSIO), Charité – Universitätsmedizin Berlin**

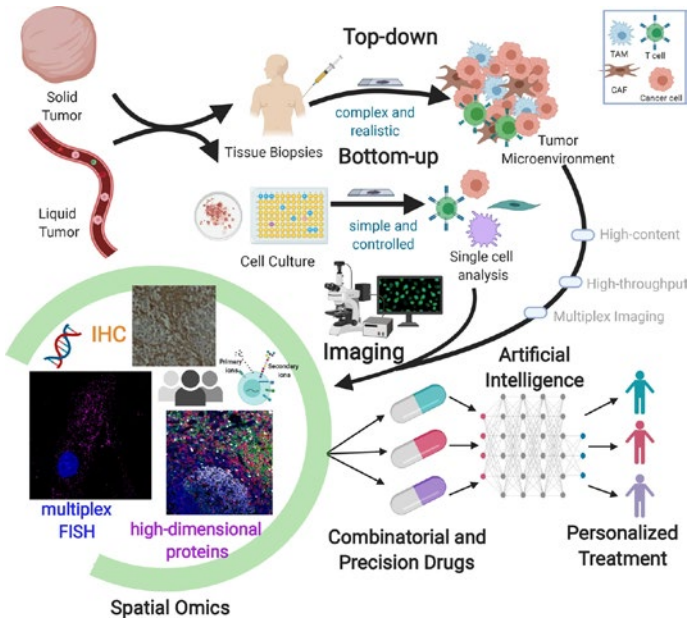
FLG: What are the advantages of viewing cellular neighborhoods in 3D as opposed to 2D? Did you learn anything new about the tumour microenvironment as a result?

Tancredi: *Cells live and communicate in 3D tissues. In our experiment exploring cellular neighbourhoods, we show how profiling these neighbourhoods in 3D better captures the spatial organization of cells into multicellular niches. In the tumour microenvironment, investigating 3D neighbourhoods identified intratumoral immune niches, which were invisible in 2D. These are critical for orchestrating the anti-tumoral immune responses and 3D neighbourhoods identified which interactions could promote tumour immune escape in situ. In the immunotherapy era, 3D approaches will have a central role for studying and therapeutic targeting of patient-specific receptor-ligand interactions.*

"IN THE TUMOUR MICROENVIRONMENT, INVESTIGATING 3D NEIGHBOURHOODS IDENTIFIED INTRATUMORAL IMMUNE NICHES, WHICH WERE INVISIBLE IN 2D. THESE ARE CRITICAL FOR ORCHESTRATING THE ANTI-TUMORAL IMMUNE RESPONSES ..."

FIGURE 3. ANALYSING SOLID AND LIQUID TUMOURS WITH MULTIPLEXED SPATIAL PROTEOMICS.

Tumour samples can be assessed using a top-down approach, studying the tumour microenvironment as a complex heterogeneous entity. Or as a bottom-up approach using cell cultures to analyse cell behaviours individually. Either way, multiplex imaging techniques are applied to both approaches to profile molecular characteristics. Image Credit: Allam, et al.[19]



Spatial proteomics for understanding the TME

To study the TME sufficiently, we need tools that can spatially identify interactions between tumour cells, immune cells and stromal cells within the TME. This requires quantification of high numbers of proteins with spatial organisation. Spatial proteomics has already identified unique TME communities that correlate with cancer severity[16-18] and look promising for refining future treatment strategies.

Spatial proteomics is a valuable method for profiling molecular characteristics of tumours regardless of whether a top-down or bottom-up approach is used (see Figure 3). Spatial proteomics can effectively define cellular neighbourhoods, identify cell-cell communication through ligand-receptor colocalization and direct visualisation of ligand and receptor proteins and ultimately lead the way to personalised cancer treatment. Analysis of cell-cell communication is covered in depth in the next section.

Miltenyi Biotec

CASE STUDY

DEEP PHENOTYPING AND ANALYSIS OF THE IMMUNE MICROENVIRONMENT ACROSS DIFFERENT SOLID TUMORS UTILIZING A COMPLETELY AUTOMATED IMAGING SYSTEM AND STAINING PANEL

TO BETTER UNDERSTAND CANCER BIOLOGY, A THOROUGH CHARACTERIZATION OF THE SPATIAL CONTEXT OF THE HETEROGENEOUS AND DYNAMIC TUMOR MICROENVIRONMENT (TME) IS NEEDED.

The TME includes malignant and non-malignant cell populations that play a role in immunotherapy. A vast number of markers are required to effectively characterize the location and relationships between immune infiltrates, tumor-specific markers, and structural components of the tumor. A spatial biology approach that can elucidate the variety and function of different cell types is critical to unravel this complexity, identify new target candidates, as well as predict and monitor the response to therapeutic intervention.

There are a number of challenges that emerge from performing truly high-multiplexed spatial proteomics. This includes the overall workflow, availability of antibodies to identify the extensive number of markers needed to fully characterize the TME, as well as the ability of a researcher to analyze the data without reliance on the expertise of bioinformatics specialists to begin deriving insights. The MACSima™ Spatial Biology Imaging Platform was developed to overcome these obstacles and allows users to focus on the science as they navigate complex tissue environments, such as the TME. This system is unique in its ability to automatically stain and image a virtually unlimited number of targets using MACSima Imaging Cyclic Staining (MICS) technology with walk-away ease. Once images are acquired, the accompanying MACS® iQ View Image Analysis Software provides a simple yet powerful analysis solution that was designed specifically to analyze the large data stacks acquired with ultrahigh-plexed imaging (Figure 1).

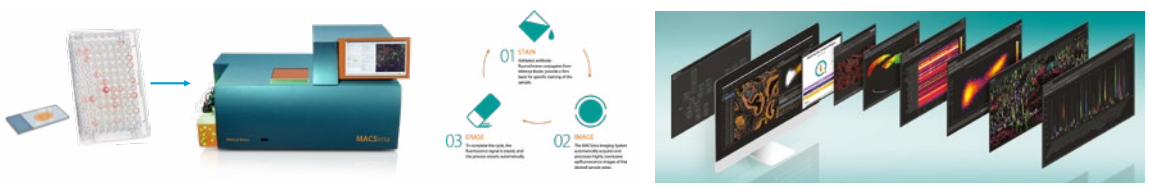
READY-TO-USE MACSIMA ANTIBODY PANELS FOR SPATIAL MULTIPLEX IMAGING

Selecting from a portfolio of hundreds of performance-verified antibodies for immunofluorescence (IF), Miltenyi Biotec developed the REAscreen™ Immuno-oncology Antibody Panel that provides a convenient and standardized staining panel, dried down in a 96-well plate.

The plate is simply placed into the MACSima System and the pre-defined panel is selected from a menu in the experiment set up in the operating software.

FIGURE 1.

Left. *Automated workflow using a pre-optimized REAscreen Antibody Plate for MICS on the MACSima Spatial Biology System.* **Right.** *The MACS iQ View Image Analysis Software Package enables immediate access to data analysis by the user.*



The REAscreen Immuno-oncology Antibody Panel was optimized for human formalin-fixed paraffin-embedded (FFPE) samples and contains 61 essential markers to identify immune cells, tumor stroma including blood and lymphatic vessels, and malignant epithelial cell populations, in either their proliferative or apoptotic state. It enables the analysis of at least 12 potential immune cell subsets within the tumor microenvironment, as well as the activation or checkpoint status of those populations.

Performance was verified on numerous solid tumor tissue samples (Figure 2), including cervical squamous cell carcinoma (CSCC), head and neck squamous cell carcinoma (HNSCC), melanoma, pancreatic ductal adenocarcinoma (PDAC), as well as colorectal adenocarcinoma (CRC) (Figure 3), and tonsil tissue, which was used as an internal control for each MICS experiment.

SPATIAL PROTEOMICS USING MICS REVEALS DISTINCT CELL POPULATIONS WITHIN THE TUMOR MICROENVIRONMENT

The different tissues show distinct distribution of several immune cell populations within the tumor microenvironment. For example, the CRC sample shows a low-grade infiltrating adenocarcinoma with colon mucosa and submucosa and a prominent lymphocyte infiltrate; a tertiary lymphoid structure is also visible towards the left of the region of interest (ROI) (Figure 3).

NEIGHBORHOOD ANALYSIS AND IMMUNE CELL POPULATIONS

Distance mapping can point to signatures in the different populations and neighborhoods where the cells reside and help draw conclusions about relationships, communication and function. The MICS data can be analyzed comprehensively by multiple workflows within MACS iQ View Image Analysis Software or exported to an established bioinformatics pipeline. Here, MACS iQ View was applied to segment the cells based on nuclei staining using an established algorithm followed by identification of the immune cell populations through cell gating of distinct phenotypic cell types and the creation of distance maps.

Intra-tumoral regions were defined by the tumor cells using tumor-specific markers and peri-tumoral and distant regions were established (Figure 4 A-B). Cell populations were gated as follows: Tregs (CD4+ FoxP3+, mast cells - mast cell tryptase+, neutrophils (CD11b-CD66b+), M1 macrophages (CD14+ CD68+CD163-), M2 macrophages (CD14+ CD68+ CD163+), Monocytes (CD11b+), NK cells (CD56+), Plasma cells (CD79a+), B cells (CD20+), helper T cells (CD4+), cytotoxic T cells (CD8+).

FIGURE 2. MAGNIFIED OVERVIEWS OF MULTIPLEX IF STAINING OF HUMAN FFPE CANCER TISSUES.

*Images display expression of relevant markers noted in each figure. (**A**) CSCC, (**B**) HNSCC, (**C**) melanoma, (**D**) PDAC.*

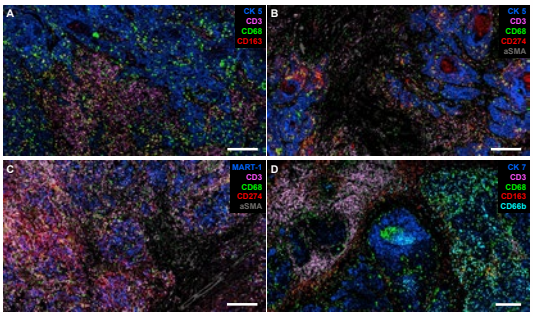
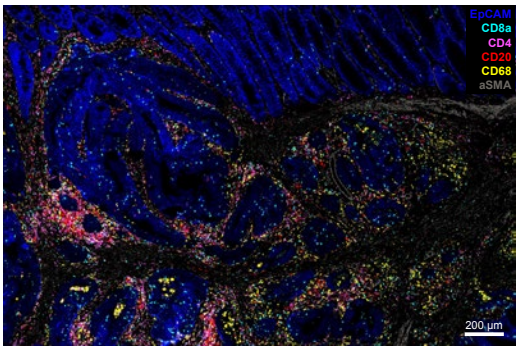
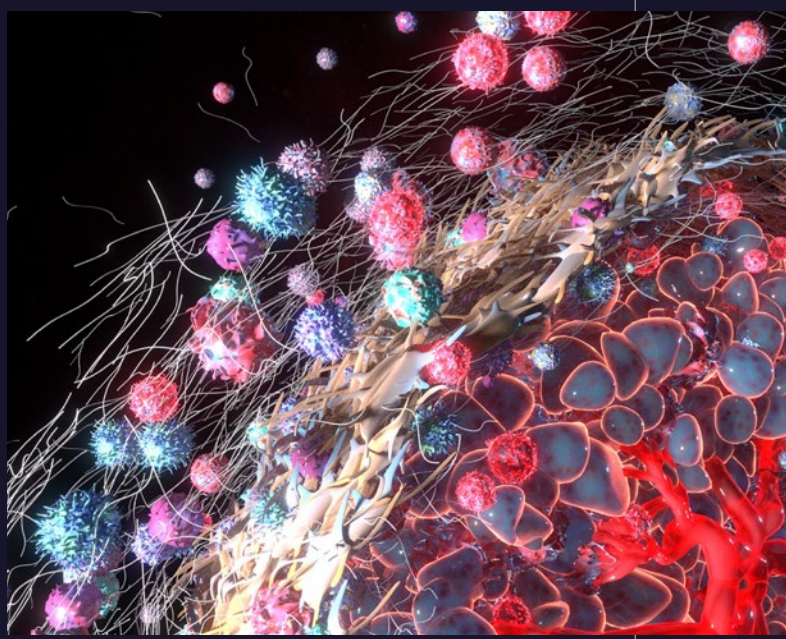


FIGURE 3. WHOLE ROI OVERVIEW OF A MULTIPLEX IF STAINING OF HUMAN CRC TISSUE IMAGES DISPLAY EXPRESSION OF EPCAM, CD8A, CD4, CD20, CD68, AND ASMA.





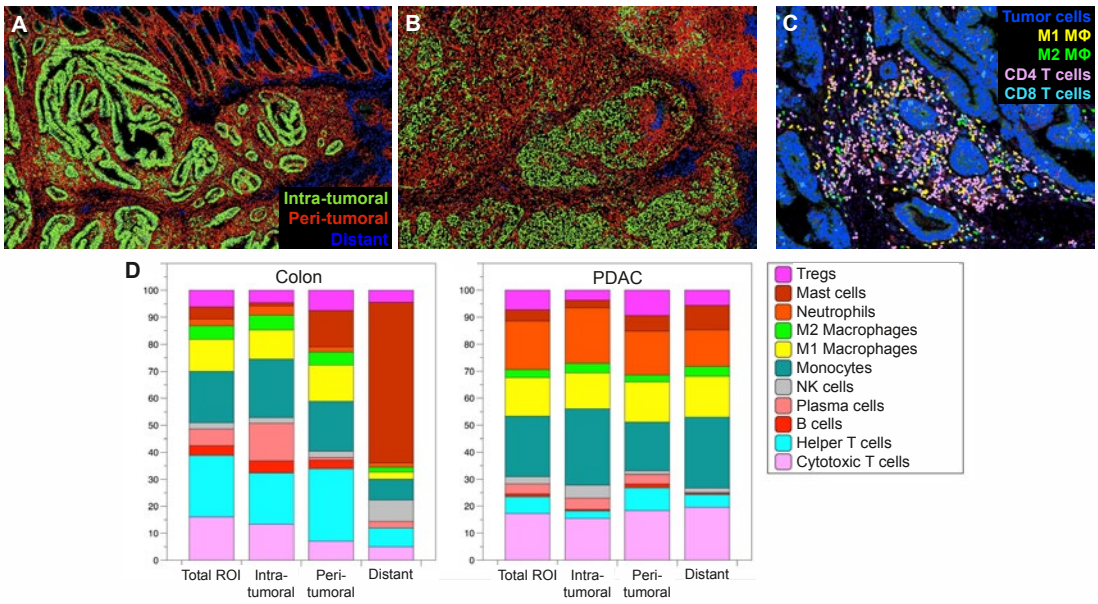
Subsequently, the percent of each immune population within the respective tumor regions could be identified (Figure 4D). The MACS iQ View Software includes additional dimensionality reduction tools such as t-SNE, UMAP, and heat maps, all designed to derive insights from large and complex data stacks.

WHAT'S NEXT? COMBINING SPATIAL PROTEOMIC AND TRANSCRIPTOMICS TOGETHER FOR HIGH SPATIAL-RESOLUTION MULTIOMICS

Miltenyi Biotec introduced a combination of spatial transcriptomics and proteomics on the MACSima Imaging System at the American Association for Cancer Research meeting earlier this year. This multiomics approach, which generates highly specific and quantitative gene expression profiles and protein signatures on the same tissue section and platform, that will allow users to push the boundaries of spatial biology. The resulting data allows for a deep dive into the mechanisms of tumor progression, while providing the spatial context within which these events occur. Such comprehensive spatial tumor profiling has the potential to improve biomarker identification, tumor cell phenotyping, and clinical prognoses for patients.

FIGURE 4. ANALYSIS USING MACS IQ VIEW.

*Distance maps of the human CRC (**A**) and PDAC (**B**) samples. (**C**) Magnification of the CRC sample, showing the identified immune cell populations. (**D**) Percent of the indicated immune populations per total number of immune cells within the respective tumor region.*



Nicole Johnson, Miltenyi Biotec. Data and contributions by: *Julia Femel, Emily Neil, Dongju Park, Fabio El Yassouri, Anijutta Appelshoffer, Erica Lloyd, Michael DiBuono, Henry Sauer, Hanna Lafayette, Hsinyi Smith, Jinling Wang, Dominic Mangiardi, Alex Makrigiorgos, Paurush Praveen, Silvia Rüberg, Werner Müller, Tanya Wantenaar, Robert Pinard, Andreas Bosio*

Cell Gossip - the study of cell-cell communications

The above section highlights the value of viewing cells in context. However, to truly analyse a cell's community, we need to know what they are saying to each other. For this we need to look at the recent advances in cell-cell communication.

Cells have various ways of communicating; an important mechanism is for 'sender' cells to secrete ligands that can bind to corresponding receptor proteins on the plasma membrane of 'receiver' cells.

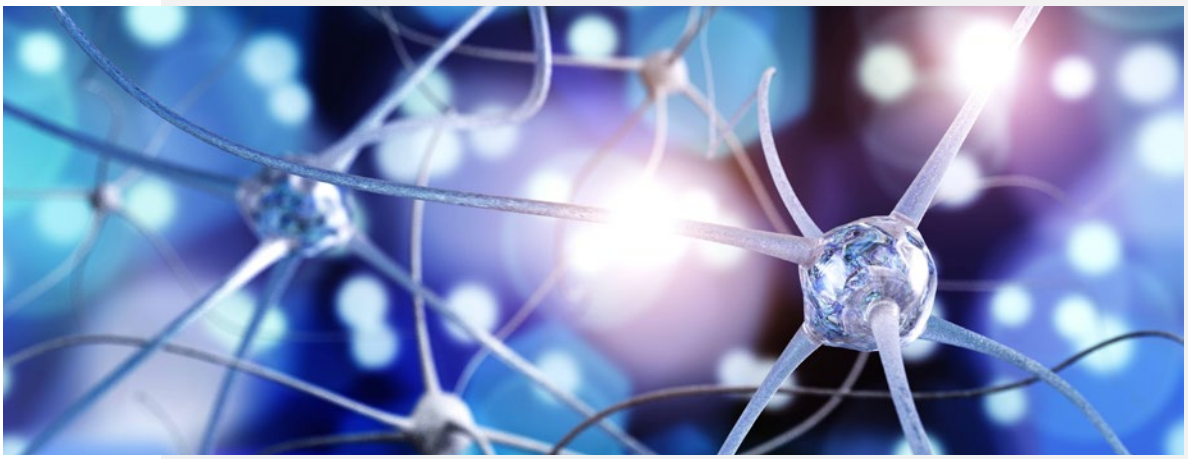
The analysis of cell-cell communication is limited, since we do not yet have the necessary capabilities to track individual ligands as they are produced, released and bind to receptors. Instead, cell-cell communication is an inferred analysis. It is performed downstream from single-cell sequencing and spatial data and uses the expression of genes and proteins for specific cell messengers (ligands) and the complementary receptors for them.

We asked our contributors why studying cell-cell communication matters.

ROBIN BROWAEYS
Team Leader – Bio-IT Support
VIB Centre for Inflammation Research, Ghent University

FLG: For people who aren't familiar with it, why is it valuable to study cell-cell communication in cellular microenvironments?

Robin: *Cell-cell communication is a crucial process in multicellular organisms. Our tissues are comprised of different types of cells and to make sure the tissue is functioning properly, and hence the organism is functioning properly, the cells have to coordinate their functions with each other. Also, cell-cell interactions are very important in development. A classic example are stem cells that will differentiate into functioning differentiated cells, and this is often due to the influence of external signals produced by cells in the environment. Cells are constantly talking to each other within multicellular organisms. Cells within tissues, and sometimes between tissues, are talking to each other by producing several molecules. This is crucial to understanding basic health and different diseases. In diseases, like COVID-19, there is a dysregulation of the immune response, and it is because the immune cells talk differently to each other compared to normal responses to a virus.*



MICHA SAM BRICKMAN RAREDON

Research Group Leader, Departments of Anaesthesiology, Pulmonary, Critical Care & Sleep Medicine and Immunobiology
Yale School of Medicine

FLG: Please can you give a description of cell-cell communication and why it matters?

Sam: *It's a great question, I think it's an emerging area in the field and I don't know how many people even know to ask that question. So, why do we care about cell-cell communication? Why is this important for tissues? When we study biology, either in humans or animals, or we study pathology, like disease versus control tissues, in most cases you are dealing with a tissue, which is made of cells. A lot of diseases that are actively being looked into, at this point, are really tissue level diseases. What's going wrong is not just a single genetic defect and it's not just a single cell type. Rather it's where the entire tissue as a whole isn't behaving properly and isn't able to stay in a healthy state.*

Now, how does a tissue operate? This is the fundamental answer to your question. A tissue is this community of cells that are all talking to each other. And some of them might be directly contacting one another, some of them might be very close by, but they're constantly sending and receiving signals from one another. That signalling between those individual cells is what allows this community of cells to stay in a stable state and operate correctly. When we study cell-to-cell communication, we get a really good quantitative window into that conversation between the community members. We can understand why a tissue might be stable, or why it might be falling into a disease state.

An analogy I use is, imagine if you're studying a city. If you were to study New York City, you could do a census, and you could study all of the individuals in that city, and you could know where they lived, who they were, what they did for a living and everything about them. But all you would have was a static snapshot. And you can actually learn a lot from that kind of a thing.

What connectivity analysis or ligand-receptor-signalling analysis allows you to measure is how all of those individuals are talking to one another. It's like seeing their phone calls or the mail that they're sending or something that gives you a sense of the different relationships between one another. It gives you an architecture to understand how a city is operating rather than just what makes up a city.

Single-cell methods for cell-cell communication

Two classes of methods exist for this analysis in single-cell data. There are (1) methods that measure the levels of ligands and of receptors in cell clusters to ascertain the level of communication these cells are having; and (2) methods that estimate downstream intracellular activities (e.g., gene-expression changes) alongside ligand-receptor expression rates to try to estimate whether a ligand has bound and affected genomic processes in the cell.

All methods operate on a basic assumption; that transcriptomic data is a good proxy for cell-cell communication events. However, cell communication happens at the protein level and is spatially constrained. Hence, a level of caution is necessary when interpreting the output of cell-cell communication methods. Furthermore, crosstalk is often estimated between cell populations rather than between cells of a population, which is realistically where a lot of communication will occur. However, these methods have recently been shown to be robust and concordant with both proteomic and spatial ligand/receptor data[20] so should not be overlooked.

We will now review several of the tools available for evaluating cell-cell communication from single-cell data (see Figure 4 for the principles of this analysis) followed by a discussion of how spatial analysis will accelerate this field drastically.

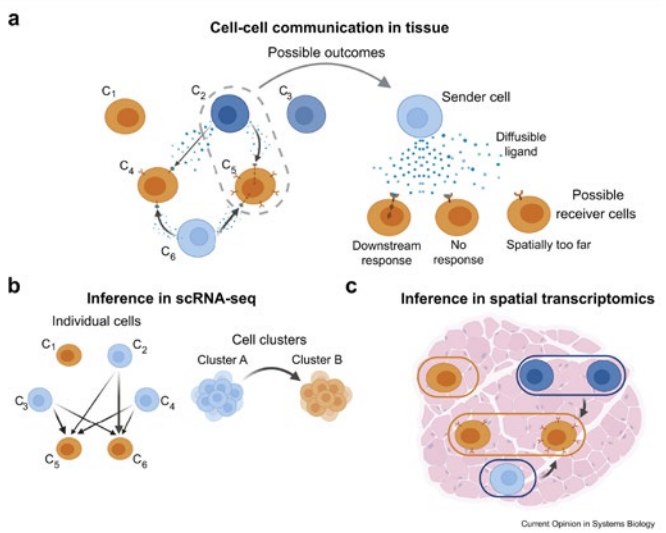
The first class of tools, discussed above, use ligand/receptor expression estimates. They are the most common. An overview of the major tools in the field already exists, which the reader should refer to[20]. Several popular tools in this category include CellPhoneDB[22] and CellChat[23], which have well curated ligand-receptor databases to infer cell communications. The outcome of the review process saw the creation of a new tool – LIANA (Figure 5) – that takes single-cell RNA data and establishes a common interface with all methods and resources that exist for cell-cell communication, providing a consensus ranking for the method's prediction.

For the second class of tool, those which estimate the downstream effects of signalling, NicheNet[24] is the most well-known in this space. NicheNet, like the first class of tools, assumes that a sender cell produces a ligand if the gene for that ligand is expressed. It goes beyond the other methods by also assuming that the receiver cells (with the receptors) will experience a signal propagation affecting master gene regulators and TFs, which can be measured. NicheNet makes use of prior knowledge of downstream gene effects to do this.

Recently, NicheNetv2 and MultiNicheNet have also been released[25]. NicheNetv2 adds experimentally determined target genes for over 100 ligands, grounding the predicted downstream effects in biology. MultiNicheNet tackles a problem in the cell-cell communication field - appropriately analysing multiple samples and conditions. In brief, MultiNicheNet infers the differentially expressed ligand-receptor pairs between conditions and the downstream target genes, while accounting for inter-sample heterogeneity. Dr. Robin Browaeys was first author for both tools. We spoke with Robin to get to grips with this new version of NicheNet.

FIGURE 4. PRINCIPLES OF CELL–CELL COMMUNICATION INFERENCE.

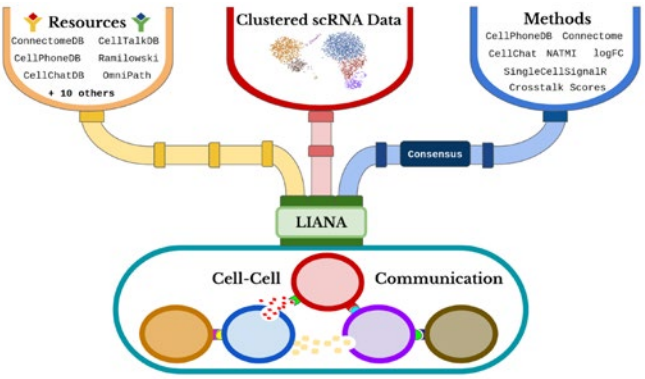
*(**A**) Cells can secrete ligands that diffuse and can bind to receptors. This is likelier to occur for receiver cells that are closest to the sender cell and when there is sufficient receptor expression. The blue and orange cells represent different cell types. For the blue cells, darker shades represent stronger ligand expression. (**B**) Cell–cell communication can be inferred from scRNA-seq at either the individual cell or cell cluster level, but spatial distances between cells are lost. (**C**) Using spatial transcriptomics to infer cell–cell communication preserves spatial distances between cells but potentially at the loss of single-cell or gene resolution. Image and Caption Credit: Almet, et al.[21]*



> "THE ANALYSIS OF CELL-CELL COMMUNICATION IS LIMITED, SINCE WE DO NOT YET HAVE THE NECESSARY CAPABILITIES TO TRACK INDIVIDUAL LIGANDS AS THEY ARE PRODUCED, RELEASED AND BIND TO RECEPTORS."

FIGURE 5. OVERVIEW OF LIANA.

LIANA takes any annotated scRNA dataset and establishes a common interface to all the resources and methods in any combination. LIANA also provides a consensus ranking for the method's predictions. Image Credit: https://saezlab.github.io/liana/

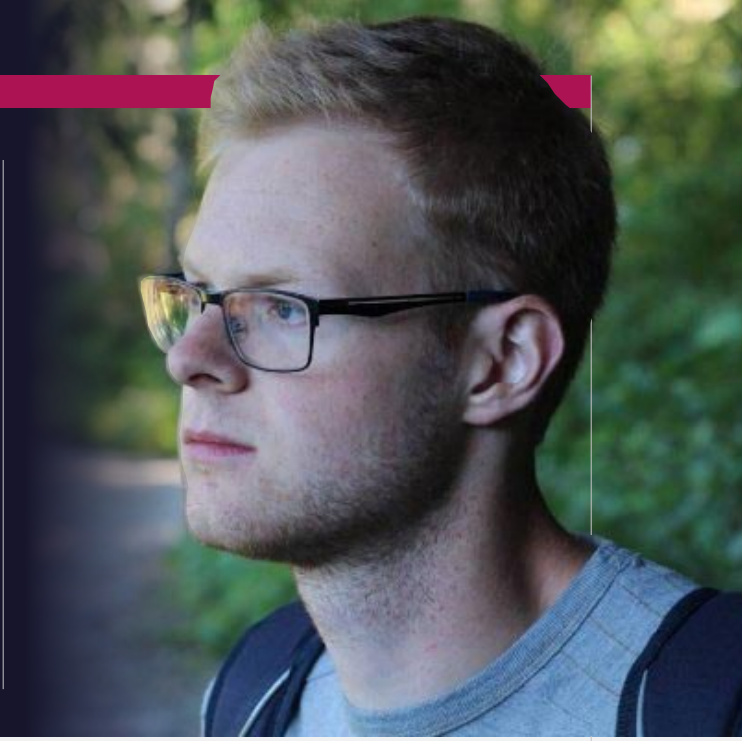


INTERVIEW:

ROBIN BROWAEYS

TEAM LEADER - BIO-IT SUPPORT

VIB CENTRE FOR INFLAMMATION RESEARCH, GHENT UNIVERSITY



***FLG:* Can you briefly describe how you can work out cell-cell communication from transcriptomic data? And how does your tool, NicheNet, work?**

Robin: The boom in methods to study cell-cell communications came after the boom in single-cell experimental technologies, because these data now allow us to identify different cell populations etc. The evident next question is, how do these different populations in my sample communicate with each other?

The most well-known tools for this are CellPhoneDB, CellChat and NicheNet. I think CellPhoneDB was one of the first dedicated tools for studying cell-cell communication on single-cell transcriptomics data. This tool looks at co-expression of ligands and receptors across cell populations and it will use some permutation tests to actually determine which interactions are specific between which cell types. These tools assume that if you have RNA expression of a gene that encodes a ligand protein in cell type A, and you have the RNA expression of a gene that encodes a receptor in cell type B, then there might be communication between cell type A and cell type B through this ligand-receptor interaction. To do this, they integrate expression data of your cells with databases of ligand-receptor pairs constructed using prior knowledge.

"ONE OF THE PREDOMINANT EFFECTS OF CELL-CELL COMMUNICATION IS A CHANGE IN GENE EXPRESSION IN THE RECEIVING CELL TYPE AND WE WANTED TO BUILD A MODEL THAT WOULD ALLOW US TO SAY 'THESE SECRETED LIGANDS CAN POTENTIALLY REGULATE THESE GENES' IN OTHER CELL TYPES."

Parallel to that we developed NicheNet, and we use a different principle. We focus on looking for the enrichment of target genes of the ligand-receptor interactions. One of the predominant effects of cell-cell communication is a change in gene expression in the receiving cell type and we wanted to build a model that would allow us to say 'these secreted ligands can potentially regulate these genes' in other cell types.

How did we do that? Actually, the same principle as these other tools, we combined our expression data with prior knowledge from databases. Now, we are not only looking at ligand-receptor interactions, but also, everything that's happening downstream in terms of transcriptional regulation. Hence, we had to integrate several data sources and databases to do that. Then we came up with a predictive model that predicts target genes regulated by ligands, and we validated this model to check how well it works. This model can then be used on transcriptome data of interacting cells to look at downstream effects of ligand-receptor interactions to prioritize them based on that. Other tools will typically generate quite large lists of interactions because each cell type expresses several ligands, several receptors, meaning there are many combinations and possible redundancy.

Whereas for NicheNet, the goal from the beginning was to prioritize interactions based on what we see in the receiver and predict what might have been the most important signal produced by the sender cell. A limitation is that we have to use this prior knowledge model, which might lead to some false positive and false negative predictions. So, it's a double-edged sword. Why would you use NicheNet? It helps you prioritize, but you might miss some things.

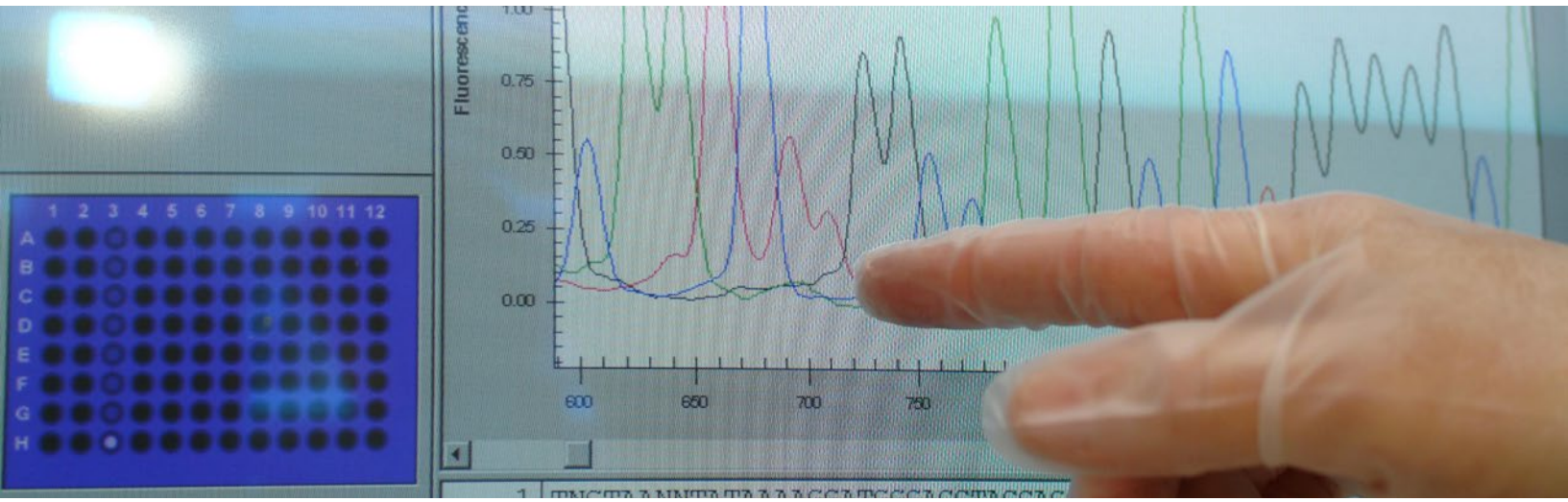
***FLG:* And then the new tool, MultiNicheNet, expands this out to analyse multiple samples?**

Robin: Yes, it was specifically developed to handle datasets with multiple samples. We came to the idea that there was a need for that during the COVID-19 pandemic. I got sent datasets from the university hospital and they wanted to look at cell-cell communication dysregulation between COVID-19 patients and healthy controls. If I applied NicheNet in the normal way, it would not be very appropriate, because we typically pool all the cells of all patients together, and then analyze cell-cell communication. This then ignores variation between patients and that cell-cell communication happens within one patient. To handle that, we produced MultiNicheNet. We based our tool on the principles of differential expression analysis methods that were recently developed to perform proper DE analysis on multi sample datasets. Then we applied these principles to our NicheNet framework. On top of that, we also included some other prioritization criteria, such as cell type specificity and differential expression of the ligand-receptors pairs themselves, as was done in the other tools. So, it's a multi sample extension to NicheNet, but it also incorporates some elements of the other tools.

***FLG:* Are there any plans to incorporate spatial data into NicheNet?**

Robin: I think that's definitely a goal for my previous research group who will now continue working on NicheNet, and its extensions. In practice, this is not straightforward because you have to consider the different spatial technologies, each with their limitations. You have technologies that allow you to profile cells at single cell or sub-cellular resolution, but they're not transcriptome-wide. Transcriptome-wide would be needed for proper NicheNet analysis because the approach is based on enrichment compared to background. On the other hand, you have technologies that generate a transcriptome wide profile of spots, and these spots typically contain transcripts of several cells. People say 2-10 and then in practice it might be between 5-20, or 5-50. It depends on the tissue but it's not at the single cell level. How do you transfer this to cell-cell communication? Are you looking at communication within a spot? Between spots? The conclusion isn't very straightforward. The spatial data with single cell resolution, it's not transcriptome-wide yet, but those data might be interesting to detect which cell types might be co-occurring, and this information can be used in parallel, to better inform you on which cell type combinations you run in your cell-cell communication analysis.

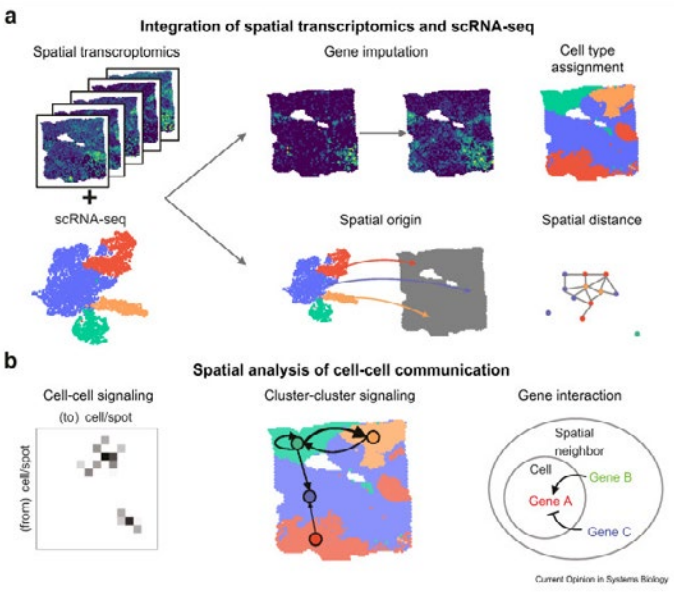
NICHES[26], tackles another problem that cell-cell communication methods have; not taking advantage of the single-cell data. Most methods average ligand expression across a cell cluster as a measure of the cell type's ligand expression. NICHES, by comparison, operates at the single-cell level, measuring the interactions of every cell with another cell, comparing them in an iterative pairwise manner (see interview below for more details).



Spatial methods for cell-cell communication

FIGURE 6. SPATIAL ANALYSIS OF CELL-CELL COMMUNICATION

*(**A**) Integrating spatial and scRNA-seq data provides combined information and can allow cell type assignment and spatial distance estimate data (**B**) Current spatial cell–cell communication inference methods output: a cell–cell or cluster–cluster network due to ligand–receptor binding and more general intercellular gene regulatory networks in space. Image Credit: Almet, et al.[21]*



A biological reality of ligands is that they can only travel a certain distance once secreted[21]. With the influx of spatial data, the opportunity is arising for a more accurate assessment of cell-cell communication, one which calculates communication within a defined niche of cells to which ligands can actually disperse. Figure 6 displays the tasks and outputs of spatial analysis of cell-cell communication.

Not only does NICHES bring new insights to single-cell data, but it can also be used for spatial data. By limiting the analysis of ligands and receptors to spatial neighbours, microenvironment signalling can be assessed. We spoke to Dr. Micha Sam Brickman Raredon, lead author of NICHES, for further details.

INTERVIEW:

MICHA SAM BRICKMAN RAREDON

RESEARCH GROUP LEADER, DEPARTMENTS OF ANAESTHESIOLOGY, PULMONARY, CRITICAL CARE & SLEEP MEDICINE AND IMMUNOBIOLOGY

YALE SCHOOL OF MEDICINE

***FLG:* Can you describe how your new tool, NICHES, works?**

Sam: NICHES can be applied to either single cell or spatial data, but it is always doing the same fundamental operation, which is that it's pairing cells together, taking one cell and treating it as a sending cell and another cell and treating it as a receiving cell. Then it looks at the ligand expression of the sending cell and the receptor expression on the receiving cell. These are like connected edges and you can infer potential communication there. It's like shortwave radio operators, many people can be tuned to the same radio wavelength, but if two individuals tuned to a very specific wavelength, they can have quite a degree of privacy. You know, other people are listening, but it shows you that there's a connection there. NICHES looks for that kind of pattern. If they're the only two individuals within this population that are doing that, there's a very high likelihood that they are leveraging that communication for some kind of a biological function.

Previous techniques that came before NICHES would average cellular populations. So, you'd perform single-cell for a tissue, and you'd get 50 different cell types that are very different phenotypically. The previous tools would take one cluster and look at the average ligand expression in that whole cluster, and then another cluster and they would take the average receptor expression of that cluster. They would then just create an edge between those two mean values. This is extremely lossy. It's great, it's very powerful and it's been used very extensively. However, you're obscuring all of that fine patterning and the subtlety around how these two populations are communicating with one another. You just can't see it because you've regressed both to the mean. You also destroy the beautiful statistical power of single-cell technologies; we're not just measuring something once; we're measuring hundreds of thousands of times.

We built NICHES to specifically solve this problem. As I said, it takes one-to-one cell pairs. But if you zoom out from that a little bit, it takes one population and another population, and it randomizes both, and then it creates those one-on-one pairs, and it does it thousands of times. So, you get really high statistical confidence that a certain set of wavelengths is exclusively or preferentially being used to communicate information between these two populations.

The spatial data has this added benefit that you have x,y coordinates, which means you can limit the edges that you're analysing just to local neighbourhoods. This is really the way that tissues work, and it is what makes them so beautiful. A lot of these ligand-receptor mechanisms, they have a maximum diffusion distance. They can only really send a message over 300 microns or 500 microns, and after that the signal has diffused down so that it's not really there anymore. If you can say these cells are co-localized, and they're highly connected, that's much stronger evidence than if you just started doing it in the abstract.

> "YOU ALSO DESTROY THE BEAUTIFUL STATISTICAL POWER OF SINGLE-CELL TECHNOLOGIES; WE'RE NOT JUST MEASURING SOMETHING ONCE; WE'RE MEASURING HUNDREDS OF THOUSANDS OF TIMES."

Other methods have been developed to take advantage of spatial transcriptomic data. Fundamental packages such as Squidpy[5] and Giotto[4] have ligand and receptor analysis methods in a spatial context. There are also graph-based methods such as SpaTalk[27], which evaluates intercellular and intracellular communication. Newer mathematical methods, such as COMMOT[28], look promising. Rather than assessing ligand and receptor levels, these methods work out optimal transport systems for ligands within spatial constraints, and use them to predict cell communication patterns across a tissue. Combining single-cell and spatial data from the same tissue is another promising avenue, using the cell type annotations of the former and the spatial context of the latter (See Figure 6A). A tool, Renoir[29], has recently been released for this purpose, and can infer communication niches, as well as identify the major ligands in each niche, uncovering key players in cell communications.

Ultimately, to get the most out of spatial data for cell-cell communication, it needs to be hi-plex, allowing the visualisation of many ligand and receptor genes and/or proteins in the same section. As single-cell and spatial proteomics develops to include more targets, we can hope that direct ligand and receptor protein assessment will be commonplace for assessing cell-cell communication.

Chapter 4 references

1. Fu, T. et al. **Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response.** Journal of hematology & oncology **14**, 98 (2021).
2. Elhanani, O., Ben-Uri, R. & Keren, L. **Spatial profiling technologies illuminate the tumor microenvironment.** Cancer cell (2023).
3. Tanevski, J., Flores, R.O.R., Gabor, A., Schapiro, D. & Saez-Rodriguez, J. **Explainable multiview framework for dissecting spatial relationships from highly multiplexed data.** Genome biology **23**, 1-31 (2022).
4. Dries, R. et al. **Giotto: a toolbox for integrative analysis and visualization of spatial expression data.** Genome biology **22**, 1-31 (2021).
5. Palla, G. et al. **Squidpy: a scalable framework for spatial omics analysis.** Nature methods **19**, 171-178 (2022).
6. Feng, Y. et al. **Spatial analysis with SPIAT and spaSim to characterize and simulate tissue microenvironments.** Nature Communications **14**, 2697 (2023).
7. Baghban, R. et al. **Tumor microenvironment complexity and therapeutic implications at a glance.** Cell Communication and Signaling **18**, 1-19 (2020).
8. van Dam, S., Baars, M.J. & Vercoulen, Y. **Multiplex tissue imaging: spatial revelations in the tumor microenvironment.** Cancers **14**, 3170 (2022).
9. Yu, Q., Jiang, M. & Wu, L. **Spatial transcriptomics technology in cancer research.** Frontiers in Oncology **12**, 1019111 (2022).
10. Kinkhabwala, A. et al. **MACSima imaging cyclic staining (MICS) technology reveals combinatorial target pairs for CAR T cell treatment of solid tumors.** Scientific Reports **12**, 1911 (2022).
11. Hunter, M.V., Moncada, R., Weiss, J.M., Yanai, I. & White, R.M. **Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface.** Nature communications **12**, 6278 (2021).
12. Nirmal, A.J. et al. **The spatial landscape of progression and immunoediting in primary melanoma at single-cell resolution.** Cancer discovery **12**, 1518-1541 (2022).
13. Dhainaut, M. et al. **Spatial CRISPR genomics identifies regulators of the tumor microenvironment.** Cell **185**, 1223-1239. e20 (2022).
14. Danenberg, E. et al. **Breast tumor microenvironment structures are associated with genomic features and clinical outcome.** Nature genetics **54**, 660-669 (2022).
15. Pentimalli, T.M. et al. **High-resolution molecular atlas of a lung tumor in 3D.** bioRxiv, 2023.05.10.539644 (2023).
16. Piyadasa, H., Angelo, M. & Bendall, S.C. **Spatial proteomics of tumor microenvironments reveal why location matters.** Nature Immunology **24**, 565-566 (2023).
17. Karimi, E. et al. **Single-cell spatial immune landscapes of primary and metastatic brain tumours.** Nature **614**, 555-563 (2023).
18. Sorin, M. et al. **Single-cell spatial landscapes of the lung tumour immune microenvironment.** Nature **614**, 548-554 (2023).
19. Allam, M., Cai, S. & Coskun, A.F. **Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics.** npj Precision Oncology **4**, 11 (2020).
20. Dimitrov, D. et al. **Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data.** Nature communications **13**, 3224 (2022).
21. Almet, A.A., Cang, Z., Jin, S. & Nie, Q. **The landscape of cell–cell communication through single-cell transcriptomics.** Current opinion in systems biology **26**, 12-23 (2021).
22. 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.** Nature protocols **15**, 1484-1506 (2020).
23. Jin, S. et al. **Inference and analysis of cell-cell communication using CellChat.** Nature communications **12**, 1088 (2021).
24. Browaeys, R., Saelens, W. & Saeys, Y. **NicheNet: modeling intercellular communication by linking ligands to target genes.** Nature methods **17**, 159-162 (2020).
25. Browaeys, R. et al. **MultiNicheNet: a flexible framework for differential cell-cell communication analysis from multi-sample multi-condition single-cell transcriptomics data.** bioRxiv, 2023.06. 13.544751 (2023).
26. Raredon, M.S.B. et al. **Comprehensive visualization of cell–cell interactions in single-cell and spatial transcriptomics with NICHES.** Bioinformatics **39**, btac775 (2023).
27. Shao, X. et al. **Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk.** Nature Communications **13**, 4429 (2022).
28. Cang, Z. et al. **Screening cell–cell communication in spatial transcriptomics via collective optimal transport.** Nature Methods **20**, 218-228 (2023).
29. Rao, N. et al. **Charting spatial ligand-target activity using Renoir.** bioRxiv, 2023.04. 14.536833 (2023).

CHAPTER 5

BEYOND THE GENOME. SINGLE-CELL AND SPATIAL EPIGENOMICS

DNA, RNA AND PROTEIN MAKE UP THE CENTRAL DOGMA OF GENOMICS. THEY ARE THE TYPICAL TARGETS OF MOST SEQUENCING EXPERIMENTS THAT TRY TO UNDERSTAND A CELL'S IDENTITY AND FUNCTION. HOWEVER, JUST AS IMPORTANT TO UNDERSTANDING A CELL'S STATE IS THE EPIGENOMIC PROFILE OF THE CELL.

The epigenome is represented by a complex interplay of mechanisms, including DNA methylation, histone modifications and chromatin conformation, that control a gene's expression (See Figure 1). Historically, these markers have been measured from bulk-sequencing, but we now understand that these metrics vary between individual cells and, consequently, the last decade has seen significant advancement in using these technologies for single cells. In this chapter, we will examine some of the methods to measure the epigenome at a single-cell level, and conclude with the big advance in this field, spatial epigenomics.

Why does the epigenome matter?

Sequencing epigenetic information is valuable for several reasons. If we are interested in cellular identify, then the set of transcription factors, chromatin conformations and RNA regulators are more persistent than RNA for determining cell fate. Ultimately, if you are interested in gene regulation and expression, epigenetic information (in either mono-omic or multiomics form) provides insights not possible to gain from RNA, DNA or protein alone. We asked experts in the field why the epigenome matters.

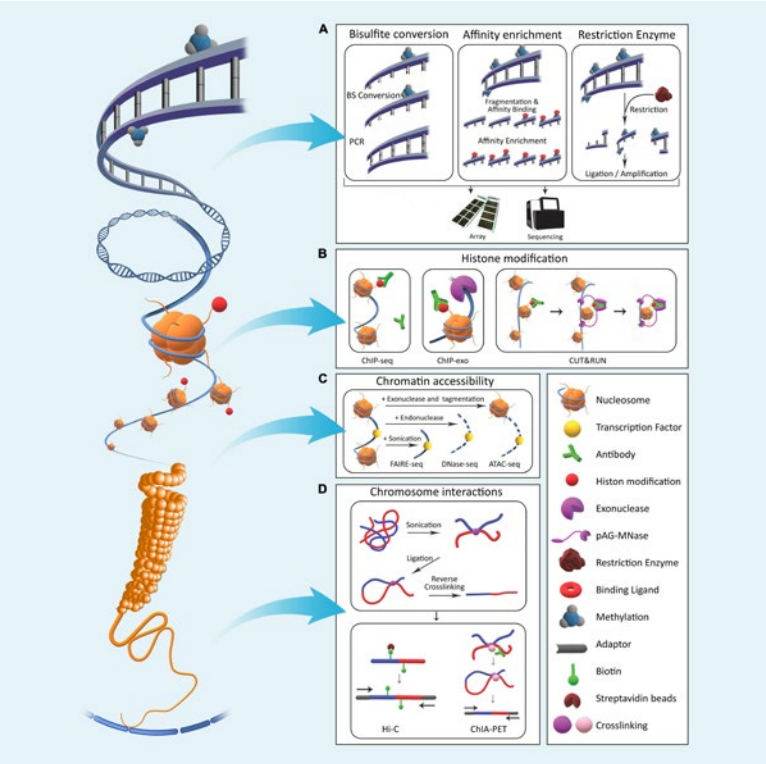


FIGURE 1. THE FOUR MAJOR EPIGENETIC LAYERS AND HOW TO PROFILE THEM.

*(**A**) DNA methylation means the addition of a methyl group to certain bases in DNA. Methods for assessment of genome-wide DNA methylation are broadly categorized into bisulfite conversion-based, affinity enrichment-based, and restriction enzyme-based techniques. (**B**) Histones undergo a variety of chemical modifications on their tail domains to affect DNA accessibility and subsequent expression. Methods for detecting these modifications rely on antibodies specifically designed to bind modified histone tails for immunoprecipitation with varying levels of resolution. (**C**) Genomic regions differ with respect to nucleosome occupancy and accessibility of DNA molecule to proteins. Various methods have been developed that quantify these characteristics across the genome. (**D**) Long-range interactions exist between regulatory elements across the genome. To identify and characterize them in a genome-wide fashion, various methods based on crosslinking and ligation have been developed with varying levels of coverage and specificity. Image and Caption Credit: Mehrmohamadi, et al.*[1]

JASON D. BUENROSTRO

Associate Professor & Broad Institute Member
Harvard University & Broad Institute of MIT and Harvard

FLG: Why is chromatin accessibility a valuable omic to study just by itself, and to include in multiomics studies alongside RNA and DNA?

Jason: *Just like most areas of biology or in most areas of any other industry, people first start with broad exploration. Single-cell RNA-seq provides a broad survey of cell types, which gives a sense of the function of the cells. But, as soon as you start to generate that kind of data, you start asking some new questions, you want to validate your results. That's one thing to do, and having multiple omics technologies allows you to really validate that that cell type is a cell type. And I think that's principally what most people use epigenomics for. The most common use case is 'I found this weird cluster or this weird edge of my cell type, is that a robust thing?' Well, it should be robust in the epigenome space as well and it should have a unique epigenomic identity. If you see a cluster there as well, you can be feeling pretty confident. If you don't, it might be an artifact of your RNA measurement, which is still an issue for the field.*

So, that's kind of the vanilla answer, but for the nuanced answer, it allows you to do new things. One of them is to understand gene regulation, to build these interaction networks that say, this transcription factor, which we know encodes the identity of the cell, was expressed. That's an RNA measurement. But we infer it to bind to these elements, that's a DNA measurement, because they're open and there's a sequence there that defines that interaction. Now, these elements, which are DNA elements, loop to a gene to activate gene expression, to create RNA. This kind of triangulation of functional units really allows us to understand not just what cells are, but also how we might push them to become healthy, or push them to create new identities, which is again the premise of therapeutics. In my mind, having the knowledge of why a cell is what it is, that's going to be a cornerstone of manipulating those identities to help the disease, and so on and so forth.

HAIQI CHEN

Assistant Professor
UT Southwestern Medical Centre

FLG: What is the value of studying epigenomics alongside RNA?

Haiqi: *This question can be applied to any multiomics really. The transcriptome is just one feature of a cell's identity. It can reflect the cellular state to some extent, but how the genome of the cell is organised, how the epigenome is organised, and how they determine the cell fate is also super important. If people want to get a multifaceted view of cells, tissues or organisms, having those modalities would also be great.*

SILVIA DOMCKE
Affiliate Assistant Professor, Dept. of Genome Sciences
University of Washington
Associate Director, Head of Human Genomics
Gordian Biotechnology

FLG: Do you think epigenome and transcriptome sequencing should become the standard way to profile the gene expression of cells now that technologies exist to profile both? If so, can you explain the added value of profiling chromatin?

Silvia: *I don't think there is a 'standard' way, since the most informative measurement will always depend on the question. Measuring RNA has many advantages, such as a high dynamic range and a defined set of genes to map to, which makes it easier to determine differential expression, annotate cell types etc. Chromatin profiling is especially useful when thinking about gene regulation, i.e., trying to understand which transcription factors (TF) are important for regulating or maintaining certain cell states. In addition, most GWAS SNPs fall into non-coding regulatory regions of the genome. Accordingly, measuring cell type-specific chromatin accessibility can help identify the cell types that certain variants are most likely to act in, as well as point towards potential underlying mechanisms of non-coding disease-associated variants, e.g., if a key TF binding site is disrupted.*

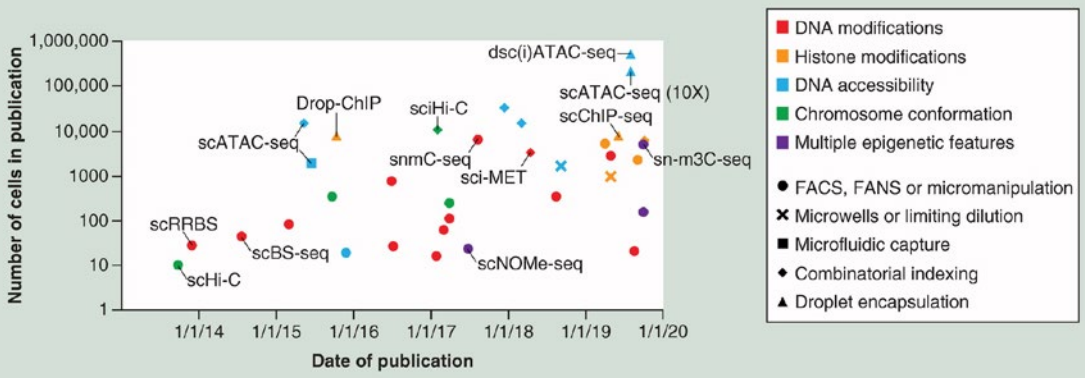
Single-cell epigenomic methods

A variety of methods exist for profiling single-cell epigenomic data. The depth and throughput capacities of these methods have only been improving[2] (see Figure 2). However, the main challenges with single-cell epigenomics remain the same: high variability, low coverage per cell, limited throughput (total number of cells analysed from a sample) and high costs[1]. We will consider the major methodologies for each epigenetic layer in turn.

SINGLE-CELL DNA METHYLATION ASSAYS

DNA Methylation is an important regulator of gene expression. It occurs when methyl groups are added to regions of the genome called CpG island. It is measured by bisulfate sequencing[3]. Early approaches, such as single-cell whole genome bisulfite sequencing (scWGBS-seq - Farlik, et al.[4]) and the reduced range version (scRRBS-seq - Guo, et al.[5]), used library prep methods to overcome the DNA loss caused by bisulfate sequencing to enable methylation reads from single cells. These early methods typically had high coverage of major CpG islands, but at low throughput and low coverage of sparse CpGs.

FIGURE 2. PUBLICATION DATES AND CELL NUMBERS OF SINGLE-CELL EPIGENOMIC METHODS.
Colour indicates epigenetic layer being profiled by each technology and symbol for single-cell isolation technology. Image credit: Bond, et al. [2]



Modern methods for DNA methylation, such as sci-MET[6] and scCGI-seq[7], use combinatorial sequencing to increase the throughput but at a lower coverage per cell.

SINGLE-CELL HISTONE MODIFICATION ASSAYS

Modifications to histone proteins (the proteins that DNA is wound around) are known to regulate genomic characteristics and transcriptional states. Mapping these modifications is informative for understanding epigenetic programs and differentiation trajectories of cells. Classically, these modifications have been studied using chromatin immunoprecipitation sequencing (ChIP-seq), which uses antibodies that are specific to known histone marks. scChIP-seq[8] is the single-cell alternative, passing cells through micrococcal nuclease to reduce background noise and make the technique viable for individual cells. It is a high-throughput technique but at the cost of lower coverage per cell.

An alternative method, called CUT&RUN, involves chromatin immune-cleavage sequencing called scChIC-seq in single cells. It uses specific antibodies to target DNA-bound proteins and then cuts off the DNA to which the antibody is bound for sequencing. This allows a high number of reads per cell but at lower throughput. CUT&Tag addresses shortcomings of CUT&RUN by preventing DNA loss. scCUT&Tag[9], the single-cell implementation, is fast becoming a popular method for this assessment in single cells. It has recently been expanded to scMulti-CUT&Tag[10] and scCUT&Tag2for1[11] to allow for sequencing of multiple chromatin factors and for the active and repressive genomic elements respectively. A similar assay, scGET-seq[12], also profiles whether chromatin is in an open and closed state to compute a new metric, chromatin velocity, which measures epigenetic plasticity of cells.

SINGLE-CELL CHROMATIN ACCESSIBILITY ASSAYS

Chromatin accessibility refers to the availability of DNA regions for regulatory proteins to bind and has many biological insights (see Figure 3). Methodologies such as DNAse-seq (and the single-cell equivalent scDNAse-seq) are used to measure this by using enzymes to fragment DNA regions without chromatin protection and inferring chromatin coverage based on sequencing the fragmented DNA. However, chromatin accessibility is dominated by one methodology with enhanced sensitivity - ATAC-seq. This method uses Tn5 transposases, which only insert into open regions of the genome (without chromatin coverage) and fragments them. By sequencing just these fragments, it is possible to identify 'open' genomic regions. A single-cell implementation was released in 2015, sc-ATAC-seq[13], and is performed on isolated cells, which allows high read coverage but low throughput. However, a combinatorial form was later introduced, sci-ATAC-seq, which allows high throughput at the cost of coverage[14]. sci-Atac-seq3 is the most recent iteration of the technology, using a three-level combinatorial indexing assay to profile cells at extraordinary throughput and low-cost to allow chromatin profiling of whole embryos[15].

SINGLE-CELL NUCLEAR ORGANISATION

Finally, an appreciation of the 3D nature of genomic regulation is often lacking in epigenomic analyses. Chromosomes are folded into domains that can interact and chromatin regions from different chromosomes can interact too, allowing crosstalk between genetic elements across the genome. Learning about chromatin conformation can help understand how these changes are associated with gene regulation and cellular function. The predominant method to measure this higher-order chromatin structure in single-cells, is a single-cell adaption of the Hi-C protocol. Like ATAC-seq, this was adapted to work on single cells in the conventional manner, scHi-C[16], which is performed on isolated cells and has a low throughput, and in combinatorial form – sci-Hi-C[17], which improves the throughput at the cost of depth. These methods are genome-wide and rely on cross-linking and ligating physically interacting chromosome regions to identify interacting DNA domains.[18]

MULTIMODAL EPIGENOMIC METHODS

As of 2023, we are entering a situation in which combinations of all the previously detailed methods are available for multiomics sequencing. This is an area of rapid growth and detailed lists of all major multiomics methods exist[19]. Below we will outline some key tools combining epigenomics methods together or epigenomics with other omics.

Epigenomics has most often been coupled with transcriptomics. For example, DNA methylation can be co-profiled with RNA using scMT-seq[20] and scM&T-seq[21]. Chromatin availability can be profiled with RNA such as with SNARE-seq[22], Paired-seq[23], SHARE-seq[24] and ISSAAC-seq[25]. Chromatin accessibility can also be profiled with proteins using ASAP-seq[26], which combines popular ATAC- and CITE-seq methods. Profiling histone modifications with scCUT&Tag has been paired with several other omics measurements such as transcriptome (CoTECH[27] & Paired-tag[28]) and surface proteins (scCUT&TAGPro)[29].

Epigenomic methods have also been paired together in multiomics form to create a more holistic epigenomic profile of individual cells. For example, scCool-Seq[30] is a method that profiles DNA methylation and chromatin accessibility with a medium throughput and high coverage. The updated version, iscCOOL-seq[31] improves that throughput further. Recent methods such as scNOMeRe-seq[32] profiles the methylome, chromatin accessibility and the transcriptome.

Nuclear organisation (HiC) has been paired with methylome profiling using methods such as scMethyl-Hic[33] and snm3C-seq[34].

Finally, methods exist to profile epigenomics alongside two other omics. This includes Neat-seq[35], DOGMA-seq[26] and TEA-seq[36], which profiles chromatin, RNA and protein (the later focusing on surface proteins). ScTrio-Seq[37] profiles the methylome alongside RNA and DNA. Finally, ScNMT-seq[38], and the very recent scChaRM-seq[39], profile chromatin accessibility, methylation and RNA.

Valuable lists of methodologies and advice for epigenomic and multimodal sequencing can be found in reviews from this year - Vandereyken, et al.[19], Preissl, et al.[40], Baysoy, et al.[41].

Single-Cell Epigenomic Analysis

Single-cell epigenomic data is uniquely challenging to analyse, with its high dimensionality, sparsity, cell-cell variability and batch effects[40]. Unique computational tools exist for each of the major tasks (e.g. data processing, clustering and downstream characterization) and for each of the major epigenomic modalities. These go beyond the scope of this chapter, but an up-to-date overview can be found in Preissl, et al.[40] and Heumos, et al.[42].

Of general utility are the recent toolkits that encompass tools for several epigenomic computational tasks in one package tasks such as Signac[43], SnapATAC[44], ArchR[45] and EpiScanpy[46], which would be a good place to start to analyse single-cell epigenomics data. As with most forms of single-cell and spatial analysis, machine learning and neural networks are being deployed to enhance this analysis (see scBasset[47] as a very recent example, which has been used to improve cell clustering, data integration and scATAC profile denoising).

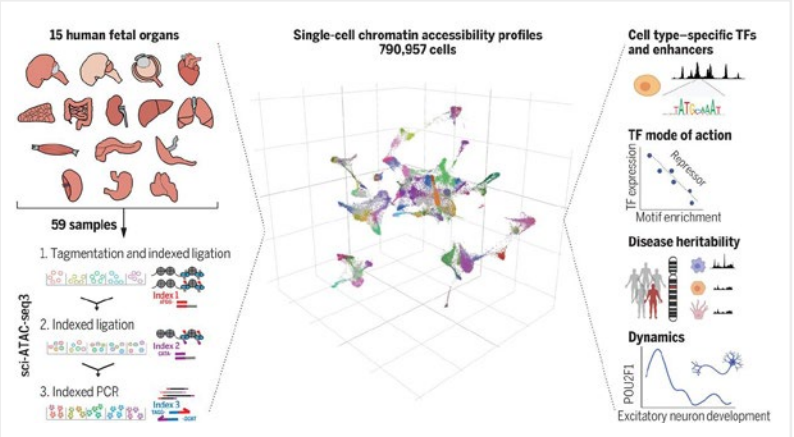
Applications of Single-Cell Epigenomics

Currently, single-cell epigenomic technologies are widely used to profile tissues during adulthood and development. The International Human Epigenome Consortium is leading the way in building single-cell epigenomic atlases to broaden our understanding of gene regulation in development, health and disease.

Some impressive examples of this profiling include the human cell atlas of fetal chromatin accessibility[15] using sci-ATAC-seq3 to profile chromatin accessibility on ~800,000 cells (See Figure 3) and the adult chromatin atlas[48] with ~615,000 cells. Efforts to build DNA methylation atlases of humans[49] have also been carried out this year.

FIGURE 3. OVERVIEW OF THE HUMAN CELL ATLAS OF FETAL CHROMATIN ACCESSIBILITY.

15 fetal organs were processed using sci-ATAC-seq3 allowing 800,000 cells to be profiled providing insights int disease heritability and dynamics. Image Credit: Domcke, et al.[15]



SILVIA DOMCKE

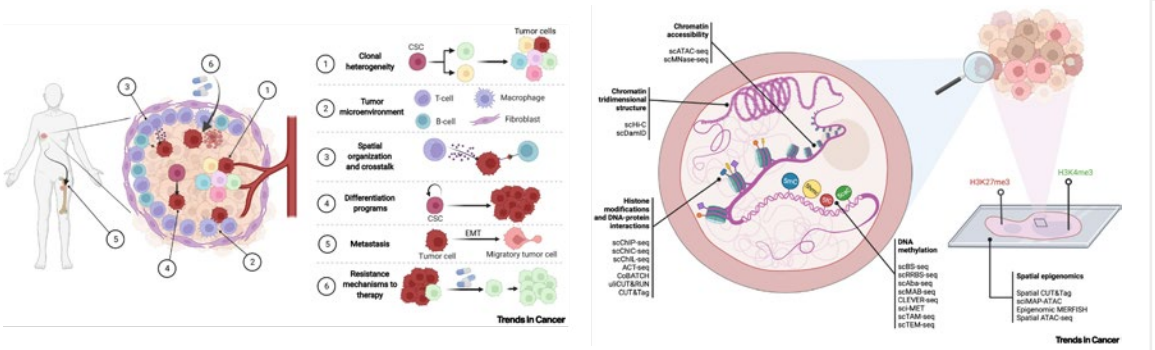
Affiliate Assistant Professor, Dept. of Genome Sciences
University of Washington
Associate Director, Head of Human Genomics
Gordian Biotechnology

FLG: If someone was to take up sci-ATAC-seq3 for the first time, are there any unique challenges/things to be aware of?

Silvia: *The key challenge with sci-ATAC-seq3 is that there is some loss of material during the protocol due to the centrifugation steps, and accordingly it is not the method of choice for very limited samples that require a near perfect nuclei recovery. But if you start with a sufficient amount of input material, don't be too discouraged if you can't see your nuclei pellet in the process – they will still be there in the sequencing!*

"THE KEY CHALLENGE WITH SCI-ATAC-SEQ3 IS THAT THERE IS SOME LOSS OF MATERIAL DURING THE PROTOCOL"

FIGURE 4. SINGLE-CELL AND SPATIAL EPIGENOMICS AND CANCER.
Left. *There are six different aspects of cancer for which epigenomic methods can provide much needed information.* **Right.** *Examples of single-cell and spatial technologies to study the different epigenetic mechanisms of cancer. Image Credit: Casado-Pelaez, et al.*[50]



Single-cell epigenomics also has a lot of promise for research into disease states. Tumour biology, in particular, is a complex heterogenous biological system. Epigenomic methods provide necessary delineation to help understand cancer clonal heterogeneity, tumour microenvironment interactions and the epigenetic programs of malignant tumour cells undergoing growth or metastasis (See Figure 4) [50,51]. Clinically, this would enable the discovery of new biomarkers, early detection of metastasis, and help to tailor personalized therapies. We asked our contributors what the clinical value of epigenomic profiling was.

"EPIGENOMIC METHODS PROVIDE NECESSARY DELINEATION TO HELP UNDERSTAND CANCER CLONAL HETEROGENEITY, TUMOUR MICROENVIRONMENT INTERACTIONS AND THE EPIGENETIC PROGRAMS OF MALIGNANT TUMOUR CELLS UNDERGOING GROWTH OR METASTASIS"

JASON D. BUENROSTRO

Associate Professor & Broad Institute Member
Harvard University & Broad Institute of MIT and Harvard

FLG: What is the clinical value of profiling chromatin accessibility?

Jason: *Like most things, when you think about the clinic, technologies have to be pretty mature before they go into the clinic. And I think right now, a lot of the examples that we have are early stage and proof of principle. But in the near future, I really imagine that epigenomics might be used for diagnostic purposes. For example, this tumour is a 'really bad' tumour because it has an invasive signature, or even my immune system is healthy, but somebody else's might have an autoimmune disease, and there's a signature associated with it. Of course, some of that can be done with other genomic technologies, but I do think that the epigenome, as we've seen in other examples, is a really robust measurement of what the cell is doing and also tends to tell us a lot about what the cell might do in a future challenge. Hence you tend to see in many examples that there are signatures that are really robust on the epigenome but are hard to nail down using transcriptomes. One great example of this is T cell exhaustion. There's a really well defined ATAC-seq signature associated with exhaustion, but it can be challenging to define the transcriptome signature.*

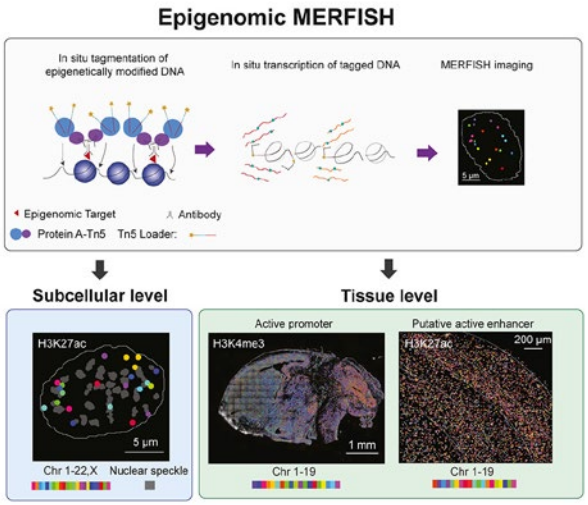
SILVIA DOMCKE

Affiliate Assistant Professor, Dept. of Genome Sciences
University of Washington
Associate Director, Head of Human Genomics
Gordian Biotechnology

FLG: Is there clinical value to chromatin profiling?

Silvia: *Chromatin profiling can help prioritize non-coding disease-associated variants for further investigation of disease mechanisms or to identify potential therapeutic targets. It can help nominate transcription factors that might be causally involved in various diseases as well as derive gene regulatory networks underlying disease states. Currently its use case is likely more focused on learning the gene regulatory relationships underlying disease states per se by profiling cohorts of diseased and normal samples, rather than as a per-patient standard readout in the clinic.*

FIGURE 5. OVERVIEW OF EPIGENOMIC MERFISH.
This technique allows the visualisation of epigenetically modified DNA at subcellular and tissue level resolutions. Image Credit: Lu, et al.[54]



Spatial epigenomic methods

A significant advance in the epigenomic space in the last 12-18 months has been the blossoming of spatial epigenomic methods to profile chromatin accessibility and histone modifications.

Two significant early methods in this space were Spatial-ATAC-seq[52] and Spatial-CUT&Tag[53]. These methods rely on the original CUT&Tag and ATAC-seq chemistry but capture that information with a spatial resolution of 20µm.

The Spatial-ATAC assay allows unbiased genome-wide mapping of chromatin accessibility and has been used to profile chromatin in the whole mouse embryo to identify spatial organization of cell types as well as their states and fates[53].

SpatialCUT&Tag provides a targeted approach, visualizing spatial histone modifications using specific antibodies. This provides insights into protein-DNA interactions, transcription factor binding, and the epigenetic modifications linked to the targeted proteins. This can provide a more granular understanding of epigenetic biomarkers of disease and histone modifications controlling chromatin structure.

Alternative methods exist for histone modifications, such as epigenomic MERFISH[54], an imaging based method that combines the CUT&Tag methodology with MERFISH. This method achieves remarkable subcellular resolution (see Figure 5) and has been used to identify new promoter-enhancer hubs in the mouse brain. However, this method lacks the unbiased nature of the above methods, requiring prior knowledge to select epigenomic loci.

Subcellular resolution has been achieved for another technique, Photoselective Sequencing (PSS)[55], which can profile subcellular regions. PSS is a cost-effective and simple technique using light microscopy to profile chromatin accessibility for specific selected regions of tissue. This approach was spear-headed by Sarah Mangiameli, but we caught up with one of the authors of PSS manuscript, Dr. Haiqi Chen, to hear more about PSS.

HAIQI CHEN
Assistant Professor
UT Southwestern Medical Centre

FLG: How does Photo-Selective Sequencing (PSS) work? What can it be used for and why might someone choose to use PSS over other similar approaches for profiling the genome or epigenome?

Haiqi: *The tool was initially developed to address a question - how do you obtain spatial information of DNA sequences, especially DNA sequences associated with epigenomic modifications? Next generation sequencing is powerful because it can get sequence information in a very high throughput fashion, but you don't have the spatial information. And in the meantime, high resolution imaging, like microscopy, is really powerful in terms of achieving high spatial resolution, especially at the sub cellular level, but without getting the sequence information of the genome. So, we (Sarah Mangiameli and others from the Labs of Fei Chen and Jason Buenrostro at Harvard) thought, what if we can combine those two technologies together and get both the spatial context as well as DNA sequence information? That's how PSS was born. Basically, you use microscopy to image cells within intact tissue sections at the sub-cellular resolution, and you use laser with a diffraction limit at a nanometre scale, to spatially select DNA that you're interested in (say cells in a specific location or a specific region within a cell) and only sequence that selected DNA. That's the whole idea of PSS.*

PSS really comes into address the accessibility challenges associated with a lot of current spatial technologies that require specialized instruments because you only need a modern-day microscope, which most lab or most core facility have, and also a FRAP module, which is something that is used for fluorescence recovery after photobleaching, which is a very standard assay that people use in the imaging core. No other specialized equipment is required. And sample preparation is simple enough that all you need are enzymes called Tn5 transposases and some DNA oligo adapters that allow you to photo-cleave cells. So, relatively easy to use and not very expensive if you have a core facility that already have a microscope and the FRAP module, and it has some stellar resolution, which I think most technologies cannot provide at this point."

"SPATIALCUT&TAG-RNA-SEQ AND SPATIAL-ATAC&RNA-SEQ WERE SUCCESSFULLY APPLIED TO DEVELOPING MOUSE AND ADULT HUMAN BRAINS, REVEALING GENOME-WIDE GENE REGULATION MECHANISMS IN A SPATIAL TISSUE CONTEXT.

Finally, 3D genome organisation has also been captured spatially. Methods relying on fluorescence such as DNA-MERFISH[56], DNA-seqFISH+[57] have the ability to profile 1000s of genomic loci allowing for the characterization of chromatin domains and chromosomal interactions. OligoFISSEQ[58] is another alternative. It uses in situ sequencing methods, which allow for higher throughput but with fewer possible targets.

All methods can only probe a limited amount of the genome, and it is hoped these methods can be expanded to chart 3D genome variability across thousands of cells in healthy and diseased tissues and organs in different species to find variable features of 3D genome organisation.[18]

JASON D. BUENROSTRO
Associate Professor & Broad Institute Member
Harvard University & Broad Institute of MIT and Harvard

FLG: What is the value of viewing the epigenomic spatially? And what are your hopes for this approach?

Jason: *There are two categories of methods and I'll just disentangle them a little. Category one is where you're looking within a cell's spatial structure. So, you might be interested in asking where those DNA Elements are spatially located within the cell. And that's super important. That's what PSS (Photoselective Sequencing) allows you to do. Because you're able to now know more about the function of our genome. By knowing these two elements of DNA are touching or interacting, we can now paint a clearer picture as to what those elements are doing. So, that's one category. And that's what we've been mostly pioneering. And we've been very excited to do. The other category is to understand epigenomes in a tissue context. You might be interested in asking how this cell's epigenome, or the cell in general, interacts with that other cell. And by understanding those interactions, it gives you a much better sense of functional units within our tissue. This nerve cell is interacting with this fibroblast. That's probably important for the functioning of that fibroblast. Now, let's understand that circuit better. So, spatial epigenomics will tell us a lot about how our individual cells and our cell types interact to form functional units, that then drive unique outcomes. So, spatial is useful for figuring out the cell type, and then really making sure that that's real. But also, it's telling you what the cells might do in response to being pushed, and being able to build gene regulatory networks, and these sorts of things.*

CO-PROFILING THE EPIGENOME AND TRANSCRIPTOME

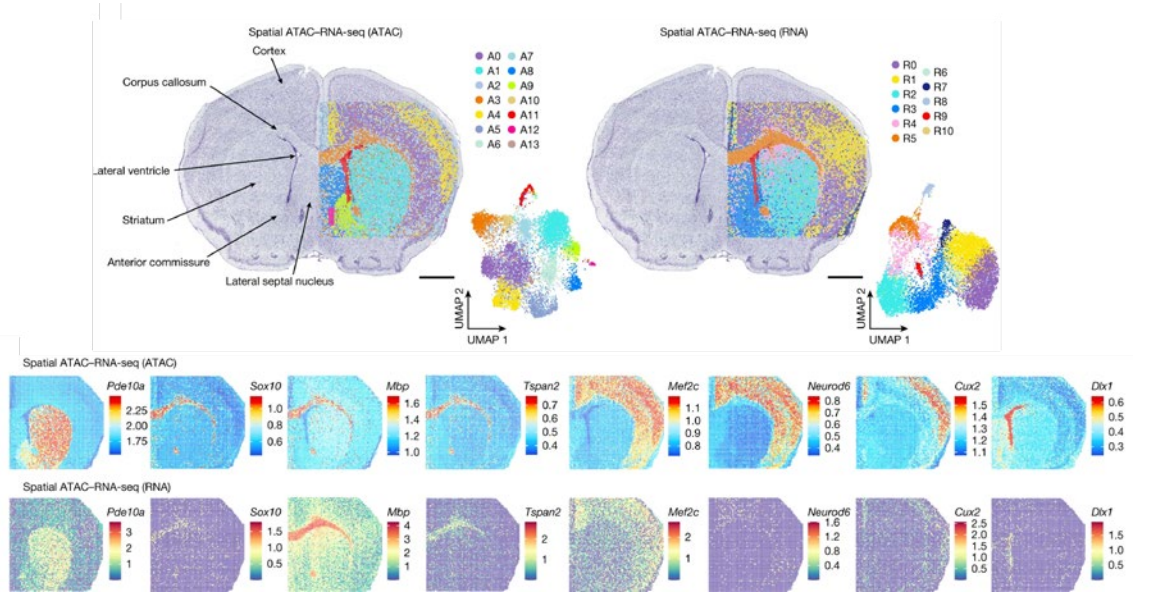
In March 2023, SpatialCUT&Tag-RNA-seq and Spatial-ATAC&RNA-seq were published[59]. These methods provide an exciting spatial multiomics methodology to simultaneously profile chromatin accessibility or histone modifications alongside RNA. These methods are hi-plex with a respectable 20µm resolution.

SpatialCUT&Tag-RNA-seq and Spatial-ATAC&RNA-seq were successfully applied to developing mouse and adult human brains, revealing genome-wide gene regulation mechanisms in a spatial tissue context[59]. They identified a new cluster of neurons not identified by RNA or ATAC-seq alone. Using a pseudo-time series to map chromatin to gene expression, it was possible to identify loss of chromatin accessibility and expression of genes across time.

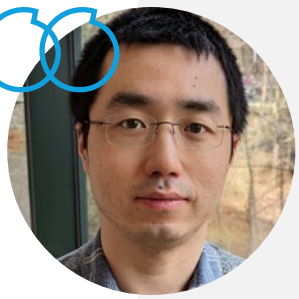
Furthermore, the authors observed an interesting discordance between genes whose chromatin was 'active', but for which limited RNA was found (See Neurod6 in Figure 6). This subset of genes was seen as 'primed' for expression presenting another unique marker to explore in health and disease.

FIGURE 6. CO-PROFILING OF EPIGENOME AND TRANSCRIPTOME IN P22 MOUSE BRAIN.

Top. *Spatial distribution and UMAP of all clusters for ATAC and RNA in spatial ATAC–RNA-seq of mouse brain. Pixel size, 20 μm; scale bars, 1 mm.* **Bottom.** *Spatial mapping of gene activity scores and gene expression for selected marker genes in different clusters for ATAC and RNA in spatial ATAC–RNA-seq. Image Credit: Zhang, et al.*[59]



This shows that to truly understand gene expression in individual cells, these different layers of molecular information are needed. The authors of the paper state, 'In summary, spatially resolved, genome-wide co-sequencing of epigenome and transcriptome at the cellular level represents one of the most informative tools in spatial biology and can be applied to a wide range of biological and biomedical research.' We asked Dr. Yanxiang Deng about this method.



YANXIANG DENG
Assistant Professor
University of Pennsylvania

Yanxiang: *These new spatial technologies can simultaneously spatially map two crucial components of our genetic makeup, the epigenome and the transcriptome. The epigenome controls the switching mechanisms that turn genes on and off; the transcriptome is the result of those gene expressions and is what defines the cell. The co-profiling strategy enabled us to see the mechanisms of how the genes are regulated or switched on and off in the tissue context.*

Chapter 5 references

1. Mehrmohamadi, M., Sepehri, M.H., Nazer, N. & Norouzi, M.R. **A comparative overview of epigenomic profiling methods.** Frontiers in cell and developmental biology **9**, 714687 (2021).
2. Bond, D.R., Uddipto, K., Enjeti, A.K. & Lee, H.J. **Single-cell epigenomics in cancer: charting a course to clinical impact.** Epigenomics **12**, 1139-1151 (2020).
3. Moore, L.D., Le, T. & Fan, G. **DNA methylation and its basic function.** Neuropsychopharmacology **38**, 23-38 (2013).
4. Farlik, M. et al. **Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics.** Cell Rep **10**, 1386-97 (2015).
5. Guo, H. et al. **Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing.** Nature Protocols **10**, 645-659 (2015).
6. Mulqueen, R.M. et al. **Highly scalable generation of DNA methylation profiles in single cells.** Nat Biotechnol **36**, 428-431 (2018).
7. Han, L. et al. **Bisulfite-independent analysis of CpG island methylation enables genome-scale stratification of single cells.** Nucleic Acids Res **45**, e77 (2017).
8. Rotem, A. et al. **Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state.** Nature biotechnology **33**, 1165-1172 (2015).
9. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. **Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues.** Nature biotechnology **39**, 825-835 (2021).
10. Gopalan, S., Wang, Y., Harper, N.W., Garber, M. & Fazzio, T.G. **Simultaneous profiling of multiple chromatin proteins in the same cells.** Molecular cell **81**, 4736-4746. e5 (2021).
11. Janssens, D.H. et al. **CUT&Tag2for1: a modified method for simultaneous profiling of the accessible and silenced regulome in single cells.** Genome biology **23**, 1-19 (2022).
12. Tedesco, M. et al. **Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin.** Nature biotechnology **40**, 235-244 (2022).
13. Buenrostro, J.D. et al. **Single-cell chromatin accessibility reveals principles of regulatory variation.** Nature **523**, 486-490 (2015).
14. Cusanovich, D.A. et al. **Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing.** Science **348**, 910-914 (2015).
15. Domcke, S. et al. **A human cell atlas of fetal chromatin accessibility.** Science **370**, eaba7612 (2020).
16. Nagano, T. et al. **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.** Nature **502**, 59-64 (2013).
17. Ramani, V. et al. **Sci-Hi-C: A single-cell Hi-C method for mapping 3D genome organization in large number of single cells.** Methods **170**, 61-68 (2020).
18. Bouwman, B.A.M., Crosetto, N. & Bienko, M. **The era of 3D and spatial genomics.** Trends Genet **38**, 1062-1075 (2022).
19. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. **Methods and applications for single-cell and spatial multi-omics.** Nature Reviews Genetics, 1-22 (2023).
20. Hu, Y. et al. **Simultaneous profiling of mRNA transcriptome and DNA methylome from a single cell.** Single Cell Methods: Sequencing and Proteomics, 363-377 (2019).
21. Angermueller, C. et al. **Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity.** Nature methods **13**, 229-232 (2016).
22. Chen, S., Lake, B.B. & Zhang, K. **High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell.** Nature biotechnology **37**, 1452-1457 (2019).
23. Zhu, C. et al. **An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome.** Nature structural & molecular biology **26**, 1063-1070 (2019).
24. Ma, S. et al. **Chromatin potential identified by shared single-cell profiling of RNA and chromatin.** Cell **183**, 1103-1116. e20 (2020).
25. Xu, W. et al. **ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells.** Nature Methods **19**, 1243-1249 (2022).
26. Mimitou, E.P. et al. **Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells.** Nature biotechnology **39**, 1246-1258 (2021).
27. Xiong, H., Luo, Y., Wang, Q., Yu, X. & He, A. **Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions.** Nature Methods **18**, 652-660 (2021).
28. Zhu, C. et al. **Joint profiling of histone modifications and transcriptome in single cells from mouse brain.** Nature methods **18**, 283-292 (2021).
29. Zhang, B. et al. **Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro.** Nature biotechnology **40**, 1220-1230 (2022).
30. Guo, F. et al. **Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells.** Cell research **27**, 967-988 (2017).
31. Gu, C., Liu, S., Wu, Q., Zhang, L. & Guo, F. **Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes.** Cell research **29**, 110-123 (2019).
32. Wang, Y. et al. **Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos.** Nature communications **12**, 1247 (2021).
33. Li, G. et al. **Joint profiling of DNA methylation and chromatin architecture in single cells.** Nature methods **16**, 991-993 (2019).
34. Lee, D.-S. et al. **Simultaneous profiling of 3D genome structure and DNA methylation in single human cells.** Nature methods **16**, 999-1006 (2019).
35. Chen, A.F. et al. **NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells.** Nature Methods **19**, 547-553 (2022).
36. Swanson, E. et al. **Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq.** Elife **10**, e63632 (2021).
37. Hou, Y. et al. **Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas.** Cell research **26**, 304-319 (2016).
38. Clark, S.J. et al. **scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells.** Nature communications **9**, 781 (2018).
39. Yan, R. et al. **Decoding dynamic epigenetic landscapes in human oocytes using single-cell multi-omics sequencing.** Cell Stem Cell **28**, 1641-1656. e7 (2021).
40. Preissl, S., Gaulton, K.J. & Ren, B. **Characterizing cis-regulatory elements using single-cell epigenomics.** Nature Reviews Genetics **24**, 21-43 (2023).
41. Baysoy, A., Bai, Z., Satija, R. & Fan, R. **The technological landscape and applications of single-cell multi-omics.** Nature Reviews Molecular Cell Biology, 1-19 (2023).
42. Heumos, L. et al. **Best practices for single-cell analysis across modalities.** Nature Reviews Genetics **24**, 550-572 (2023).
43. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R. **Single-cell chromatin state analysis with Signac.** Nature methods **18**, 1333-1341 (2021).
44. Fang, R. et al. **Comprehensive analysis of single cell ATAC-seq data with SnapATAC.** Nature communications **12**, 1337 (2021).
45. Granja, J.M. et al. **ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis.** Nature genetics **53**, 403-411 (2021).
46. Danese, A. et al. **EpiScanpy: integrated single-cell epigenomic analysis.** Nature Communications **12**, 5228 (2021).
47. Yuan, H. & Kelley, D.R. **scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional neural networks.** Nature Methods **19**, 1088-1096 (2022).
48. Zhang, K. et al. **A single-cell atlas of chromatin accessibility in the human genome.** Cell **184**, 5985-6001.e19 (2021).
49. Loyfer, N. et al. **A DNA methylation atlas of normal human cell types.** Nature **613**, 355-364 (2023).
50. Casado-Pelaez, M., Bueno-Costa, A. & Esteller, M. **Single cell cancer epigenetics.** Trends Cancer **8**, 820-838 (2022).
51. Nam, A.S., Chaligne, R. & Landau, D.A. I**ntegrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics.** Nat Rev Genet **22**, 3-18 (2021).
52. Deng, Y. et al. **Spatial profiling of chromatin accessibility in mouse and human tissues.** Nature **609**, 375-383 (2022).
53. Deng, Y. et al. **Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level.** Science **375**, 681-686 (2022).
54. Lu, T., Ang, C.E. & Zhuang, X. **Spatially resolved epigenomic profiling of single cells in complex tissues.** Cell **185**, 4448-4464.e17 (2022).
55. Mangiameli, S.M. et al. **Photoselective sequencing: microscopically guided genomic measurements with subcellular resolution.** Nature Methods **20**, 686-694 (2023).
56. Su, J.H., Zheng, P., Kinrot, S.S., Bintu, B. & Zhuang, X. **Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin.** Cell **182**, 1641-1659.e26 (2020).
57. Takei, Y. et al. **Integrated spatial genomics reveals global architecture of single nuclei.** Nature **590**, 344-350 (2021).
58. Nguyen, H.Q. et al. **3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing.** Nature Methods **17**, 822-832 (2020).
59. Zhang, D. et al. **Spatial epigenome–transcriptome co-profiling of mammalian tissues.** Nature **616**, 113-122 (2023).

CHAPTER 6

BEDSIDE BENEFITS. SINGLE-CELL AND SPATIAL IN CLINICAL PRACTICE

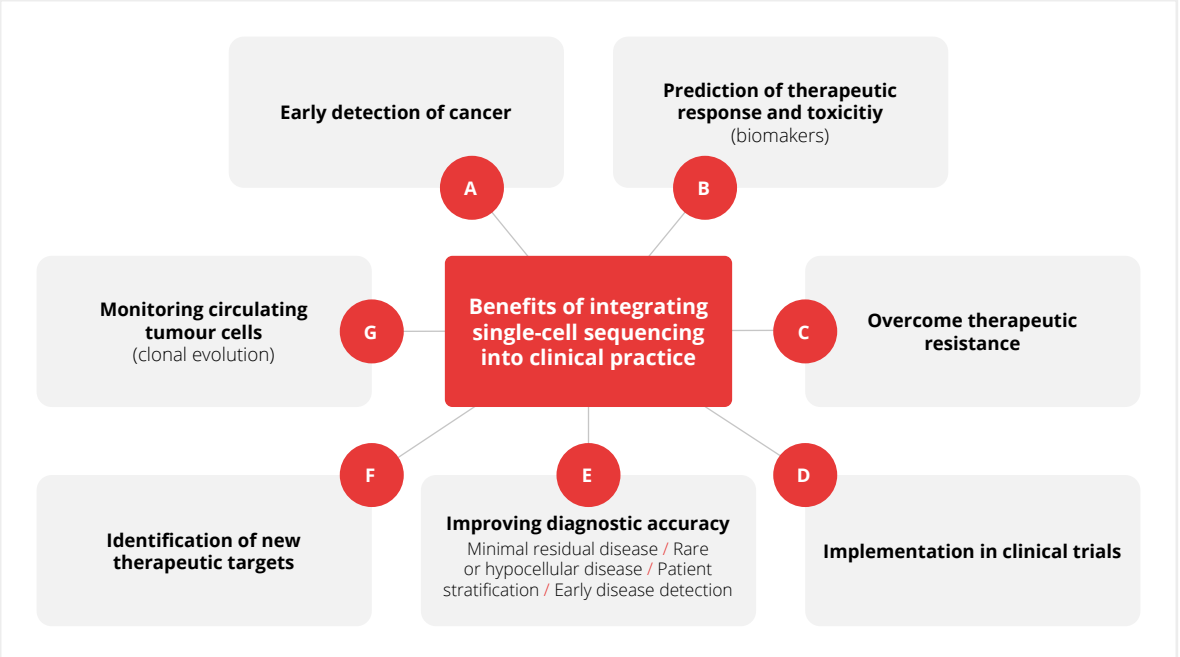
HOW DO WE LEVERAGE THESE POWERFUL TECHNOLOGIES FOR CLINICAL BENEFIT? THIS IS A PERTINENT QUESTION TO CONSIDER AS THESE TECHNOLOGIES MATURE. IN THIS CHAPTER WE WILL REVIEW SOME OF THE LATEST APPLICATIONS OF SINGLE-CELL AND SPATIAL TECHNOLOGIES TO UNDERSTAND DISEASE, WHETHER THAT'S USING THESE TECHNOLOGIES TO FIND NEW BIOMARKERS, TO UNDERSTAND WHY TREATMENTS DON'T WORK OR EVEN FOR EARLY DETECTION AND DIAGNOSIS. WE WILL DISCUSS THE PROMISES AND THE CHALLENGES OF BRINGING THIS TECHNOLOGY TO THE CLINIC.

Single-cell technologies in the clinic

As we have seen in Chapter 1, single-cell sequencing technologies have been undergoing rapid development for more than a decade. As a consequence, these technologies are much closer to clinical translation than spatial. Figure 1 displays some anticipated benefits of these technologies, some of which are already occurring!

FIGURE 1. ANTICIPATED BENEFITS OF INTEGRATING SINGLE-CELL SEQUENCING INTO CLINICAL PRACTICES.

*These include. (**A**) The early detection of cancer by applying to circulating tumour cells. (**B**) Predicting immunotherapy response and toxicity through cellular genomic signatures of specific immune cell subsets. (**C**) Identifying immunotherapy resistance through selective transcriptional loss of targeted HLA. (**D**) A range of applications in clinical trials. (**E**) Improved diagnostics through approaches such as patient stratification. (**F**) Identifying novel therapeutic targets through profiling subsets of cell types. (**G**) The monitoring of the evolution of cancer by deep profiling of circulating tumour cells. Image Adapted From: Lim, et al.*[1]



Chromium Gene Expression Flex

Flex your single cell muscle

Power up your possibilities for single cell studies

What can you do with Flex?

- Profile more samples: run scRNA-seq on fresh, frozen, and fixed samples—even FFPE
- Work on your own schedule: fix, batch, and run samples later without losing data quality
- Go further without increasing your spend: process up to 1M cells per run and save up to 75% on per-sample costs
- Dive deeper: perform high-throughput capture of protein alongside transcriptomic analysis

Learn how:
10xgen.com/flex-fg

10x GENOMICS

Current barriers to bringing single-cell to the clinic

To achieve clinically-actionable single-cell methods, three major developments are needed[1].

Firstly, we need to see the **standardisation of tissue collection in the clinic in a way that protects the omics data of individual cells. Most single-cell technologies require live cells.** This means that clinics need to invest in the necessary equipment to enable quick cell capture following fresh tissue biopsy. This will require a shift away from the clinical standard - formalin fixation for long-term storage. A fresh biopsy process needs to be simple and should occur on samples within hours of biopsy and followed by cryopreservation for storage[2].

There is already a wealth of formalin fixed and FFPE tissue samples collected from routine clinical practice. Being able to utilise these samples with modern single-cell technology is a challenge that is much closer to being addressed. There has been a recent influx of commercial and academic solutions to make use of the FFPE-based tissue samples collected routinely in clinical scenarios. 10x Genomics produced a Flex kit, which allows FFPE and fixed samples to be run on the Chromium X Series.

In academia, methods to isolate nuclei from FFPE tissues were reported last year, such as snPATHO-Seq[3] and snFFPE-seq[4]. These allow single nucleus sequencing, but lack sensitivity to detect low-quality RNA. Earlier this year, snRandom-seq[5] was developed, allowing full-length snRNA sequencing in FFPE tissues, opening the door for high-level single nuclei profiling of FFPE clinical samples. New solutions also released this year, such as FixNCut[6], allow for reversible fixation of tissues, meaning cells can be preserved at collection and analysed later without artefacts or RNA degradation.

The second development that we need for clinically actionable single-cell methods is the introduction of **cost effective sequencing platforms with high specificity, sensitivity and throughput.** Some platforms already meet these criteria, with high profile platforms from 10x Genomics and Mission Bio emerging in clinical labs, but there is still extremely limited accessibility. Affordable instrument-free approaches may provide a solution to the cost-effectiveness problem, but overall technological development is heading in the right direction.

Thirdly, we need **standardised, user-friendly, bioinformatic approaches that yield clinically interpretable findings**[7]. The current bioinformatic pipelines are robust but most areas of analysis are still undergoing enhancement, and there is disagreement over the gold standard (see Chapter 7). However, the road to identifying a gold standard is shortening, and the development of reference atlases and the algorithms to use them (see Chapter 3) will help make cell identification faster and more robust[8].

Other issues also need addressing. For example, the outcomes of these single-cell assessments need to be reported in a clear and concise manner for understandable and interpretable insights[9].

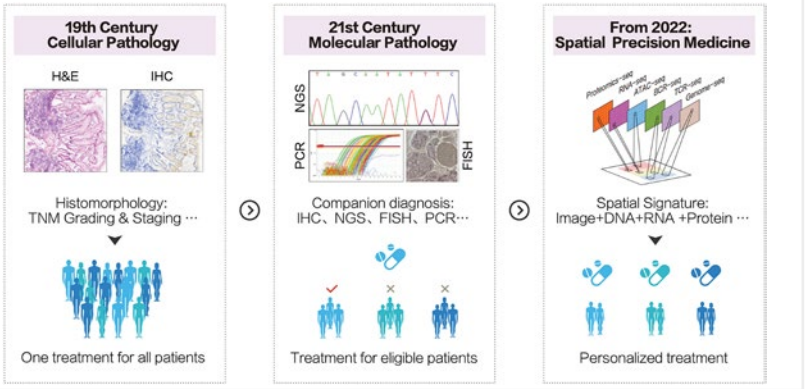
The promise of spatial technology

As a younger technology, the current spatial methods are even less well utilised in clinical labs, but they have a lot of promise to redefine medicine as we know it (see Figure 2). Pathology is a principal diagnostic tool for clinical care and is gold standard to guide the treatment and prognosis of cancer[10]. However, cancer, among other diseases, is intensely heterogenous[11]. To combat this heterogeneity, new diagnostic methods that assay multi-dimensional clinical biomarkers (i.e., multimodal genomic measurements in space and time) are necessary to truly assess these conditions.

While single-cell technology can provide deep genomic profiling, the persistence of histological pathology shows how important measurements of space are to diagnosis. Hence, modern spatial technology presents the first set of tools with the potential to truly meet that multi-dimensional biomarker standard, able to quantify the levels of different omes in space and even across time. This could create a medical paradigm in which personalised spatial signatures are a powerful diagnostic and treatment-selection tool. Figure 3 highlights a selection of enhanced metrics that can be acquired from spatial transcriptomic data as opposed to classical histology.

FIGURE 2. HOW SPATIAL BIOLOGY COULD CHANGE THE FUTURE OF PATHOLOGY.

Historical pathology (19th century) has relied on histomorphology assessment of classical stains such as H&E. Modern (21st century) pathology is now molecular, using companion diagnostics such as immunohistochemistry, FISH and NGS to sort patients into eligible treatment groups. Modern spatial technology promises a future of precision medicine, in which multimodal spatial signatures can be used to allocate personalised treatments to individuals. Image Credit: Zhang, et al.[10]



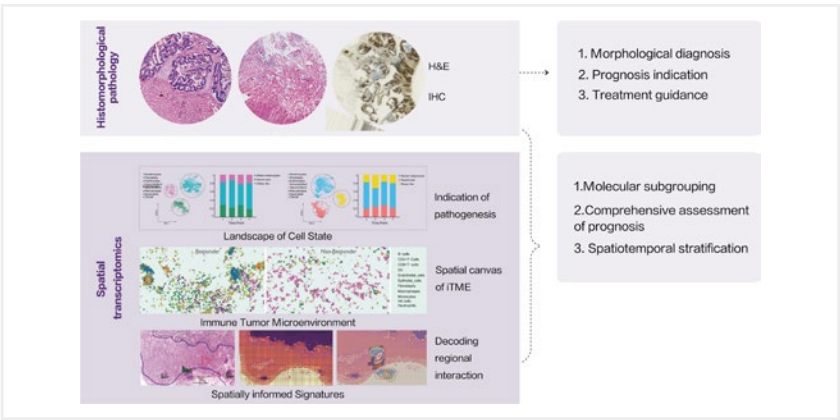


FIGURE 3. DIFFERENT INSIGHTS GAINED FROM SPATIAL TRANSCRIPTOMICS COMPARED TO HISTOMORPHOLOGICAL PATHOLOGY.

Spatial signatures would allow the identification of different pathological cell states, the landscape of the immune TME (iTME) and for spatiotemporal intracellular crosstalk analysis. This is compared to the restricted morphological and prognosis information that can be gained from standard pathology workflows. Image Credit: Zhang, et al.[10]

Furthermore, if we can associate spatial signatures with clinical metadata (i.e., survival, prognosis, treatment response), we will likely be able to predict clinical outcomes from a very early stage of disease[12]. As one recent review summarised - "An end-to-end solution of spatiotemporal omics with satisfactory reproducibility, sensitivity, throughput, and price will assist physicians in making accurate diagnoses, determining an effective treatment, avoiding wasting time, and minimizing toxic impact. Breakthroughs in precision medicine can be expected to improve care delivery, health outcomes, and quality of life."[10]

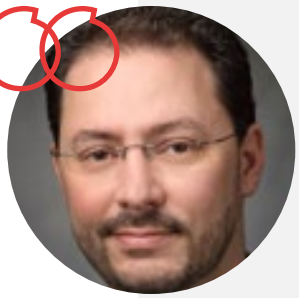
Challenges still remain

Spatial methodologies still have a number of challenges that need addressing before clinical adoption. There are still a plethora of spatial technology options that have limits on the size of tissue section that can be analysed, the number of molecular markers and the spatial resolution of these markers. This variability means that choosing the technology to adopt in a clinic to handle many different sample types is challenging.

As with single-cell technologies, tissue acquisition is also a barrier to wide adoption of spatial technologies. Leading spatial technologies are trying to make use of the current histological workflow, allowing spatial solutions on FFPE sections at typical thicknesses. Again, as with single-cell, cost, time and throughput are all problematic for the adoption of spatial technologies.

Biologically, the stability and reproducibility of spatial transcriptomic profiles also needs to be confirmed before confident clinical diagnoses can be made from spatial microenvironment analyses[13]. How tissue microenvironments and cell-cell communication strategies can be impacted by differences in disease complexity, stage, severity and pathology has not yet been confirmed.

We asked some of our contributors what they thought about spatial technology being adopted into the clinic, and some of the barriers that might slow progression.



JARED K. BURKS

Professor & Co-Director, Flow Cytometry & Cell Imaging Core Facility
The University of Texas MD Anderson Cancer Centre

FLG: Do you see spatial technology making its way to the clinic eventually? Perhaps performing real time assessments of tumour samples?

Jared: *I do. If we just look at immunotherapies, at the moment, a lot of them fail. And we're not always clear, on the clinical side, why they fail. And I believe they fail because the right immune cell is not in the right location to be stimulated to perform its role better. Or, there's an M2 macrophage that's immunosuppressive but the tumour has proactively surrounded itself with mechanisms to evade immune cells, so it's going to nullify it when it gets there. If we better characterised these patients, could we better prescribe either a combination therapy or a staggered therapy, where we try to first take out some of these hostile cells or hostile environments, and then later come in with an immunotherapy once that's optimised.*

I think we are at a point in pathology, which we were at with surgery a few years back, when all these robotics and lasers and other high-tech gadgets entered the surgical suite. We're entering a very computer-based and AI tool pathology, and that's not the way we've been training our pathologists. So, they're very uncomfortable with what we're showing them. They're so used to seeing what's in the tissue through the microscope lens and judging it all by eye. And now they're having to depend upon that camera, and a particular monitor and a whole bunch of things that they don't even know how to standardise. For example, if they go to a different monitor, do they see it the same way?

So, there's a lot of nuance, and we've got to figure that out. But I think that this is going to come to the clinic through companion diagnostics with pharmaceuticals. And then ultimately, the patients are going to start demanding it because they're reading the journal articles. They're coming in as informed as possible because their life's on the line; trying to figure out what's best for them so they can self-advocate.

DENIS SCHAPIRO

Research Group Leader
Heidelberg University Hospital

FLG: I wanted to ask you about the translational aspects of your work. What are the clinical applications you are working on and what are the challenges that you're finding in terms of actually getting spatial technology into the clinic?

Denis: *For the last two years, my group in Heidelberg has been focusing on spatial technologies in the context of oncology, but also now extending to immunology, as well as cardiology and other disease types.*

The key challenges we are currently working on are experimental design strategies as well as workflow standardization. E.g., do we need a tissue microarray, or do we require whole slide images? How many patients, samples, images and cells do we need to have enough statistical power? And what technology should we use for each question that we have? I think that's really critical. The other part is, we want to stay as close to the routine applications as possible without disrupting them. Ideally, we would work either on a consecutive tissue section or even on the same section, so the pathologists can continue doing their routine diagnosis. This would enable us, in parallel, to provide an additional layer of information. Hopefully this will help us to move spatial omics technologies closer to patient care. ▶

In terms of data analysis challenges, I think the key here is thinking about what Quality Control steps do we need? How do we know our analysis works and how can we do this with minimal human intervention? How do we control for variations in our experiments? What are the internal controls? Also, how do we automate the individual steps to the extent that, one day, we can just run it and only check the Quality Control report if required?

"IDEALLY, WE WOULD WORK EITHER ON A CONSECUTIVE TISSUE SECTION OR EVEN ON THE SAME SECTION, SO THE PATHOLOGISTS CAN CONTINUE DOING THEIR ROUTINE DIAGNOSIS."

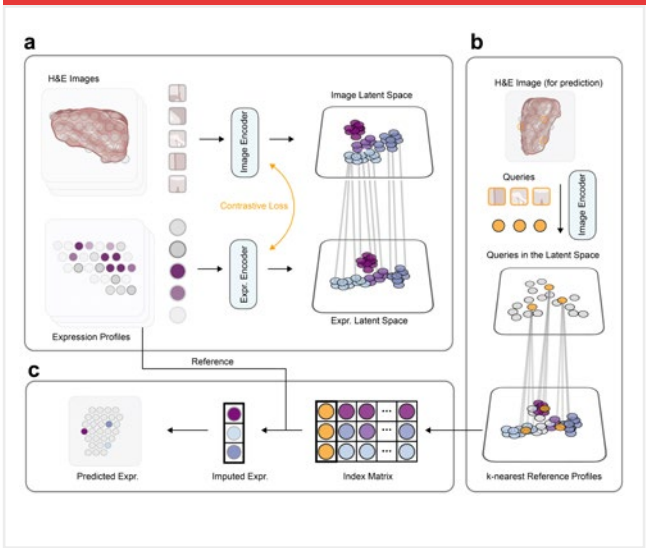
Can AI lend a helping hand?

One barrier to the adoption of both single-cell and spatial technologies is the difficulty and time-consuming nature of performing the downstream analysis to produce useful clinical insights. Parallel to the radical developments in single-cell and spatial technology is the rapid expansion of AI algorithm development to analyse the data produced by these tools[14]. In Chapter 7, we will take a deeper look at the advancement of AI tools in single-cell and spatial biology.

Since it will be difficult to directly use spatial technology in the clinic in the near future, AI presents alternative options. One promising avenue has been to develop machine learning algorithms that can predict single-cell and spatial transcriptomic features using an image that is routinely taken in the clinic, a H&E-stained image. These algorithms are trained on H&E and single-cell and spatial transcriptomics data of the same tissue and have had success in relating spatial transcriptomic features to H&E images with tools such as SCHAF[15], BLEEP[16] and MOMA[17]. Figure 4 displays the workflow for BLEEP as an exemplar. This approach could be used as a high throughput technique to quickly identify samples with morphological features that are suggestive of pathology. We spoke to Dr. Mai Chan Lau, who's work specialises in clinical applications of AI, machine learning and spatial omics.

FIGURE 4. WORKFLOW FOR BLEEP, A METHOD TO PREDICT SPATIAL EXPRESSION PROFILES FROM H&E IMAGES.

*(**A**) BLEEP learns a bimodal embedding from expression profiles and H&E image patches, (**B**) image queries are projected into the joint embedding space to index the k nearest reference expression profiles, and (**C**) the indexed reference expression profiles are linearly combined to produce the imputed gene expressions for queries. Image and Caption credit: Xie, et al. [16]*



INTERVIEW:

MAI CHAN LAU

ASSISTANT PRINCIPAL INVESTIGATOR
A*STAR'S BIOINFORMATICS INSTITUTE (BII) AND SINGAPORE IMMUNOLOGY NETWORK (SIGN)

***FLG:* What value do you think AI brings to the clinic?**

Mai Chan: AI has become increasingly vital for various applications, including the analysis of high-dimensional, large-scale, multi-omics, and spatial omics data. It's also instrumental in building predictive models. Although we haven't yet reached the point where AI can offer spatial predictions that lead to clinically actionable insights, progress is being made in that direction. At present, the clinical applications of AI are predominantly found in fields like radiology and cardiology, especially in the analysis of CT scans and X-ray images. This prevalence is mainly because these fields have access to a vast number of training images. In contrast, spatial omics technologies lack such extensive data, largely due to their high costs, specialized technical needs, and the invasive requirement for resected tissues. So, while we are some distance away from the desired outcomes, the future looks promising, and we are headed in the right direction.

***FLG:* What are the challenges of bringing AI to clinic?**

Mai Chan: Acquiring sufficient and high-quality data is often the most significant challenge in training robust AI models. The data must be both diverse and representative to enable the model to generalize to new, unseen situations. This task is complicated not only by resource limitations but also by ethical considerations surrounding data collection and use. Especially in Singapore, the clinical trial cohort size is usually quite small, at most close to 100. For AI training, we require huge amounts of data. However, emerging AI approaches could help address such issues, such as semi-supervised training models that require less ground truth for model training as well as generative approaches to generate simulated data to help increase the data size.

"SPATIAL OMICS TECHNOLOGIES LACK SUCH EXTENSIVE DATA, LARGELY DUE TO THEIR HIGH COSTS, SPECIALIZED TECHNICAL NEEDS, AND THE INVASIVE REQUIREMENT FOR RESECTED TISSUES."

***FLG:* What is the promise of AI for spatial omics and the clinic?**

Mai Chan: AI and spatial omics are rapidly emerging fields. The marrying of these fields will accelerate the future of precision medicine.

***FLG:* And this would allow the power of spatial omics to reach people in the clinic in real time?**

Mai Chan: Yes, and I think the key is to leverage the Hematoxylin and Eosin (H&E) stained and digitized images which are abundant. That's why our team at A*STAR is focused on developing H&E-based models that can predict (multiple) spatial omics signals. By obtaining a good prediction from H&E, we can do retrospective analysis to find robust biomarkers. This approach will allow us to examine a larger sample size compared to costly technologies which typically analyse only 10 to 20 samples. One ongoing project is the creation of a web-based visualization tool that enables users to view molecular markers within H&E-stained tissue. The tool is still in development and can be accessed at https://mspc.bii.a-star.edu.sg/minhn/he2.html.

Single-cell and spatial technologies vs. disease heterogeneity

Individuals differ in how their disease presents, how it evolves and how treatment affects it. This heterogeneity is a huge problem for clinicians. As already discussed in this chapter, single-cell and spatial technologies are currently broadening the selection of biomarkers that are available for diseases. Hence, these technologies present an opportunity to classify disease heterogeneity by simultaneously profiling this selection of biomarkers. Here we will review some recent examples of how single-cell and spatial methods have provided insights into heterogenous disease.

CANCER AND THE TUMOUR MICROENVIRONMENT (TME)

How spatial and single-cell methods are providing insights into cancer has been covered throughout this report. Tumours are notoriously heterogeneous, and this leads to misdiagnosis and misidentification of tumour margins. As a tour de force of the single-cell efforts to characterise cancer cells, a 2023 study collated data on over 1,000 tumours and identified hallmarks of intratumour heterogeneity[18]. They revealed gene expression programs that are co-ordinately upregulated in subpopulations of cells within many types of tumour. It is a pan-cancer single-cell resource, which each new tumour sample can be compared to.

However, the TME is a spatial entity and the work going into deconvoluting it is truly an incredible demonstration of the power these technologies can have for clinical diagnoses. A summary of how spatial biology is helping to understand the TME has already been discussed in Chapter 4.

Spatial biomarkers of the TME will be an unparalleled resource to diagnosticians concerning cancer stage and evolution[19]. For example, a combination of pan-cancer single-cell analysis and spatial transcriptomics identified recurrent gene expression programs for cancer cell states, such as immune suppression and angiogenesis, which could inform disease state[20]. Another example saw hi-plex 3D CyCIF technology used to identify unique biomarkers for different stages of melanoma development[21]. Figure 5 shows that the localisation of cytotoxic T-cells to melanocytes and the expression of markers of T-cell exhaustion and Pdl1+ macrophages distinguished the different stages. This combination of expression and protein markers located in space is exactly the kind of precise biomarker that clinicians need to tackle heterogeneity.

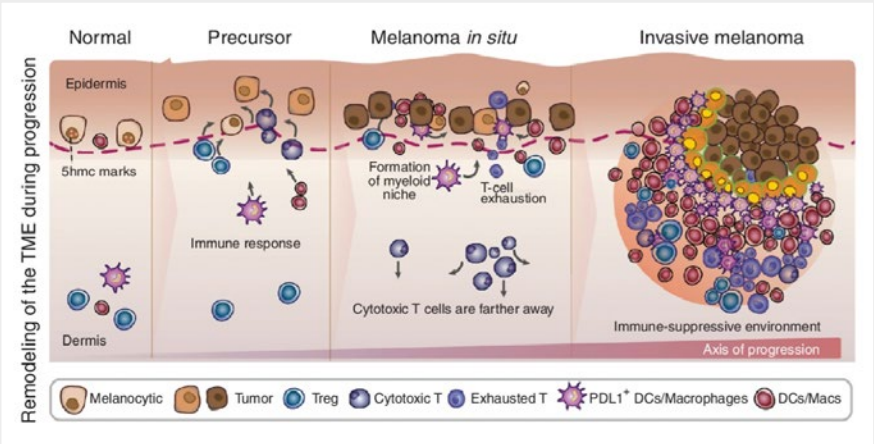


FIGURE 5. SCHEMATIC OF THE REMODELLING OF THE MELANOMA TME AS THE DISEASE PROGRESSES WITH SPATIAL/GENOMIC FEATURES THAT DISTINGUISH THE STAGES.

Image Credit: Nirmal, et al.[21]

INFLAMMATORY BOWEL DISEASE (IBD)

Disease such as ulcerative colitis and Crohn's disease are perplexingly heterogeneous. Single-cell and spatial technologies present an immensely valuable tool for addressing this. Hence, there is an influx of data and approaches to help us understand diseases of the gut. For example, this year has seen the release of several single-cell atlases of the gut[22-24]. These studies have found unique gene expression in inflamed vs. non-inflamed tissue in patients and unique changes in the small and large intestine.

Profiling the Tumor Microenvironment of Neuroblastoma using ChipCytometry™

Who

Margarida Neves M.Sc., Ph.D. Student, Department of Pathology, UCL Cancer Institute, University College London and Department of Translational Medicine, Autolus Limited

Product Focus

ChipCytometry Platform
ChipCytomety Custom Panel

Background and Objective

More precise and effective treatments are needed to treat patients with cancer. CAR T-cell therapy, where a patient's T cells are removed from their blood and genetically engineered in the laboratory to find and kill cancer cells, has the potential to deliver life-changing benefits. Patients with neuroblastomas have particularly poor prognoses and may greatly benefit from novel therapies.

A primary aim of my PhD research is to analyze the complex interactions occurring in the tumor-immune microenvironment in tissue samples from patients with various types of cancer. I have a joint appointment at UCL Cancer Institute and Autolus Limited under the supervision of Professor Teresa Marafioti and Dr. Mathieu Ferrari. Our goal is to investigate how to improve the design of future CAR T-cell therapies to ultimately improve patient outcomes.

Study Summary and Results

We used frozen tissue sections embedded in optimal cutting temperature (OCT) compound from patients with neuroblastomas. Tissue sections were loaded on microfluidic chips and protein abundance for 34 targets were quantified using a custom panel for immune cells, with a focus on T cell markers. We characterized key cell types and analyzed their spatial relationship within the tumor.

Initially, we deployed various open-source tools for computational pathology to analyze ChipCytometry datasets including QuPath, Scanpy, and Squidpy. The ability to export ChipCytometry data in open-source OME-TIFF file format has facilitated analysis by allowing us to test various pipelines. At the same time, our group is working with a bioinformatician to develop new computational modules to expand analysis into new areas.

Next Steps

Our custom ChipCytometry panel was an incredibly useful tool for spatial analysis of cell types based on protein expression. We are currently writing a manuscript to describe our findings. Our data highlights the necessity of understanding the spatial context of cells in the tissues to allow for more accurate biological comprehension of the data.

Our group will also perform transcriptional profiling with GeoMx Digital Spatial Profiler from NanoString. Each of the two platforms – ChipCytometry and GeoMx Digitial Spatial Profiler – will offer a unique set of complementary data that provides a more complete picture of the state and function of cells in the tumor microenvironment. We anticipate that our multi-omic approach will yield interesting insights into the molecular mechanisms at play in neuroblastoma and ultimately allow us to develop more effective CAR T-cell therapies.

To learn more, visit CanopyBiosciences.com/chipcytometry

For research use only. Not for use in diagnostic procedures.

Explore Cellular Landscapes Like Never Before

Meet CellScape™

The Next Generation of ChipCytometry™ Instrumentation

CellScape is an end-to-end solution for highly multiplexed spatial omics. Combining an advanced, purpose-built imaging system with easy-to-use fluidics for walk-away automation, the CellScape system accelerates exploration in the rapidly evolving field of spatial biology.

- **Highly Multiplexed:** Measure virtually unlimited protein biomarkers on a single sample
- **Throughput and Automation:** Analyze four samples at a time with walk-away automation
- **Precision Imaging:** Combine High Resolution and innovative High Dynamic Range for true single-cell quantification
- **Versatile:** Open-source reagents, no proprietary conjugations

CanopyBiosciences.com/CellScape

canopy
BIOSCIENCES® A BRUKER COMPANY

Spatial transcriptomics adds another layer of understanding to the gut[25]. One extremely recent example combined the power of single-cell and spatial technology to understand the inflamed gut in IBD[26]. Single-cell technologies identified macrophages and neutrophils as the major cell type differing between patients and controls. Nanostring CosMx was used to spatially locate the macrophage and neutrophil subsets and further classify the disordered macrophages on a spatial dimension. In conclusion, the authors state that "intestinal macrophages, which sense changes in the microenvironment, could act as reliable indicators of patient-specific molecular patterns and thus, promising targets"[26].

We reached out to the senior author, Dr. Azucena Salas, who recently presented her work at a Front Line Genomics webinar, to discuss the additional benefits of merging single-cell and spatial technologies in her work on IBD.

AZUCENA SALAS

Inflammatory Bowel Disease Group leader

Fundació de Recerca Clínic Barcelona-IDIBAPS, Hospital Clínic Barcelona

Azucena: *Here we combined spatial and single-cell RNA-seq from IBD patients, which I think is particularly helpful when you want to distinguish rare cell types that are not very transcriptionally distinct from other cell types. Those cell types or states may not be revealed by the spatial data alone, which may measure reduced gene sets compared to scRNA-seq. So, you can, I think, provide more resolution to these spatial data sets. By combining these two datasets, not only do we validate that diverse macrophage populations are found in the tissue, but also, by placing them in tissue structures we can understand a little bit more about their role, and their potential crosstalk to other cells.*

> "THOSE CELL TYPES OR STATES MAY NOT BE REVEALED BY THE SPATIAL DATA ALONE, WHICH MAY MEASURE REDUCED GENE SETS COMPARED TO SCRNA-SEQ."

AUTOIMMUNE DISEASE'S

Autoimmune disorders are caused by abnormal immune regulation or deficiencies which damage host tissue. The complicated nature of the disease and difficulties with diagnosis and disease monitoring mean there are a number of insights to gain using single-cell methodologies[27]. Conditions such as Multiple Sclerosis, Vitiligo, endometriosis and the IBD's discussed above have all had new insights gained from single-cell and spatial methods[28].

We will focus on one disease. Rheumatoid arthritis (RA) is a common autoimmune disease affecting ~1% of the population. It is characterised by inflammation in the synovium of joints, leads to loss of cartilage and bone and ultimately pain and disability[28].

Spatial technologies have been used to study the synovium microenvironment in patients with RA, finding spatial patterns of leukocyte infilration[29]. Combined single-cell and spatial transcriptomics found unique B cell populations in the joints of early RA patients[30]. Finally, a multimodal approach using CITE-seq and histology of 79 donors has recently enabled researchers to construct a single-cell RNA synovial cell atlas[31]. This resource found six unique cell type abundance phenotypes in patients, showing the RA's heterogeneity can be captured at the cell population level.

FAN ZHANG
Assistant Professor
University of Colorado School of Medicine

FLG: What insights into Autoimmune disease pathogenesis have you acquired using single-cell computational omics, and what is the potential clinical value of your work?

Fan: *Patients with autoimmune diseases are often refractory to standard therapies and do not achieve remission, so there is an unmet need for novel and personalized therapies that can be gained from digging into the molecular and cellular heterogeneity of human disease samples. Our single-cell computational strategies, combined with systems immunology approaches, deconstructed the inflammatory cellular components of rheumatoid arthritis, a prototypical autoimmune disease, and determined whether certain states are enriched only in certain subsets of patients[31] (Zhang et al., Nature, In press, 2023). These analyses provided a molecular and a tissue-based stratification, which could inform novel targeted-treatment approaches in other autoimmune diseases.*

NEUROLOGICAL DISORDERS - ALZHEIMER'S DISEASE

The brain is an immensely complex tissue, made up of neurons and supporting cells. Subtle variations in the genomic profiles of these cells can create radically different functions. Given that the organ is intensely heterogeneous, understanding neurological disorders is equally complex. It is not possible to cover all the value that single-cell and spatial technology has for deciphering the brain in this report, but it is covered in depth in this 2023 review[32].

Figure 6 shows some key studies using single-cell technologies to understand neurological disorders. Alzheimer's disease has been a point of focus amongst these disorders. Single-cell studies are converging on a set of biomarkers and phenotypes for the different cells in the brain in Alzheimer's disease[33,34] (see Figure 7). Furthermore, spatial technologies are adding biomarkers to this assessment and significant effort is now going into characterising the Amyloid Plaque Niche[35]. Multi-omics assessments will also be immensely valuable in pulling apart the heterogeneity of brain disorders. Recent reviews highlight the value that multi-omics single-cell and spatial methods will have in Alzheimer's research and are worth a read[36,37].

Single-cell and spatial studies in the brain have also helped with understanding the headaches and cognitive symptoms associated with COVID-19 infection. Analysis of post-mortem brain structures has shown aberrant features of specific cells such as microglia and astrocytes, as well as inappropriate localisation of T-cells in the choroid plexus[39,40]. This leads us to the final disease exemplar, COVID-19.

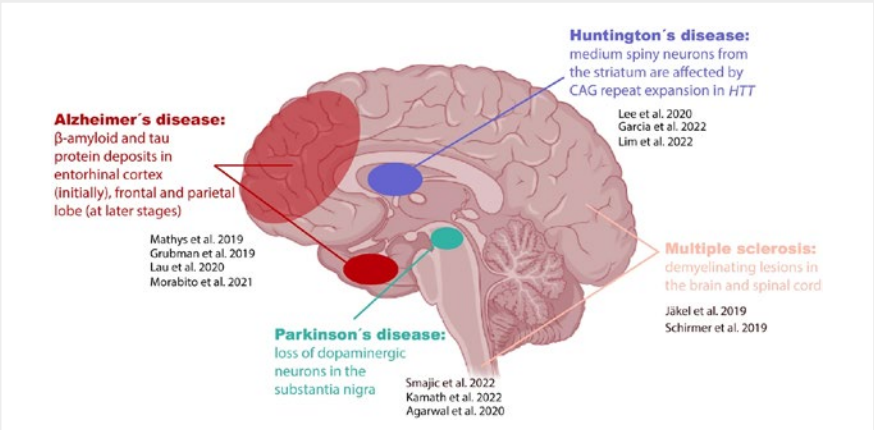


FIGURE 6. MOST COMMON NEURODEGENERATIVE DISORDERS WITH THEIR AFFECTED BRAIN REGIONS ALONG WITH SIGNIFICANT PUBLICATIONS USING SINGLE-CELL TECHNOLOGIES.

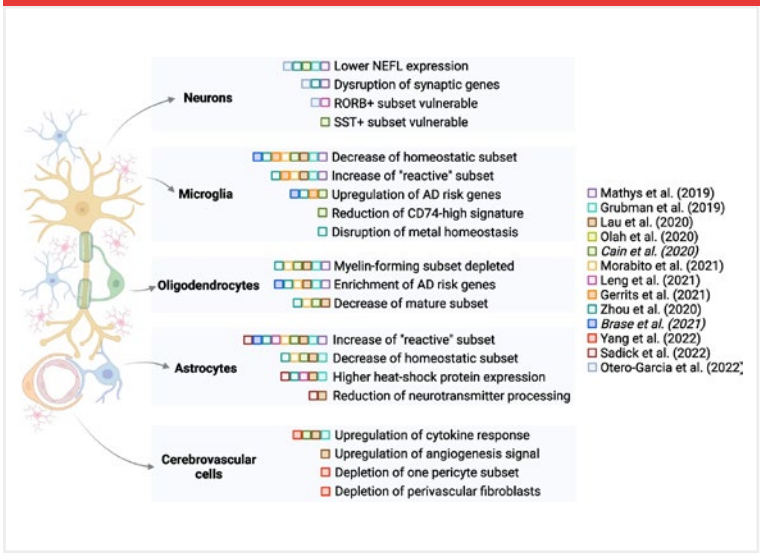
Please refer to Image source for full reference details of any publications listed. Image Credit: Pozojevic and Spielmann[38]

COVID-19

Beyond the impact on the brain, understanding the biological impact of COVID-19 infection on the human lung and circulatory system is another general area in which single-cell and spatial technologies have shown promising advances. Spatially-resolved imaging mass cytometry of the infected lung has shown unique features such as increased neutrophil infiltration into the lung, and that the predominant infection target of the virus is alveolar epithelial cells, which creates a hyper-inflammatory cell state[41]. Studies from this year are furthering this approach identifying novel immune crosstalk mechanisms[42].

FIGURE 7. OVERVIEW OF CELL TYPE-SPECIFIC BIOMARKERS IDENTIFIED USING SINGLE-CELL TECHNOLOGIES ON ALZHEIMER'S DISEASE SAMPLES.

Please refer to Image source for full reference details of any publications listed. Image Credit: Luquez, et al.[33]



A single-cell multi-omics approach on ~780,000 blood cells from 130 patients recently found several COVID-19 specific phenotypes, such as increased monocytes with receptors for interacting with platelets, predicted to replenish alveolar macrophages, and an expansion of CD8+ effector T cells[43]. First author of this study, Dr. Emily Stephenson, recently spoke at a Front Line Genomics webinar and had the following to say.

EMILY STEPHENSON
Senior Research Associate,
Haffina Lab
Newcastle University

FLG: What have you learned about COVID-19 using single-cell multiomics technologies such as Cite-Seq, and how could this have clinical impact?

Emily: *Using single-cell multi-omics has enabled us to study the immune response to COVID-19 at unprecedented resolution. We have learned that this response, in milder cases, is driven by the co-ordinated actions of adaptive lymphocytes such as T and B cells. However, this response can become uncoupled in severe cases, and can contribute to worsening disease. These actions can serve as therapeutic targets and further investigations into the treatment of the disease.*

Our findings in individuals with asymptomatic COVID-19 were that they had an expansion of a specific sub-type of B cell, one that is associated with mucosal linings such as the upper airway of the nose and throat. This could have significance when designing future vaccines or prophylaxis."

This list is by no means a comprehensive selection of every disease that stands to become better characterised with the development of single-cell and spatial methods. Instead, they present a valuable set of examples for the type of specific biomarker that will be acquirable to enable the precision therapy approaches that are on the horizon.

CASE STUDY

AAV BIODISTRIBUTION AT SINGLE-CELL RESOLUTION

VIRAL TROPISM ASSESSMENT AT THE SINGLE-CELL LEVEL IS CRUCIAL FOR DEVELOPING SAFE AND EFFECTIVE OPTIMIZED AAV VECTORS FOR GENE THERAPY, AS HIGHER AAV SPECIFICITY LEADS TO LOWER DOSAGES, FEWER SIDE EFFECTS AND MORE AFFORDABLE STRATEGIES

As of October 2023, over fifteen *in vivo* gene therapy drugs have entered the US/EU pharmaceutical market. The majority utilize adeno-associated virus (AAV) vectors as a delivery vehicle, preferred for their lower integration risk, stable gene expression, and relatively broad tropism (i.e., capacity to transduce many cell types). Nevertheless, many AAV-based drugs fail due to low tissue specificity, which makes high dosing necessary and consequently increases safety risks[1]. This case study shows the single-cell solution to that issue.

A DEEPER VALIDATION OF TROPISM

Commonly, researchers use immunohistochemistry and *in situ* hybridization techniques to analyze vector tropism. However, these approaches have downsides. They explore a limited number of AAVs simultaneously, often only one or two, and rely on known marker genes to identify cell types. This prevents an unbiased study of tropism and precludes the accurate detection of hard-to-detect cell types.

Moreover, these approaches complicate tracing back findings to downstream changes in single transduced cells. Single-cell sequencing can overcome these limitations, providing a higher-resolution view of vector tropism and downstream effects of the expressed transgene.

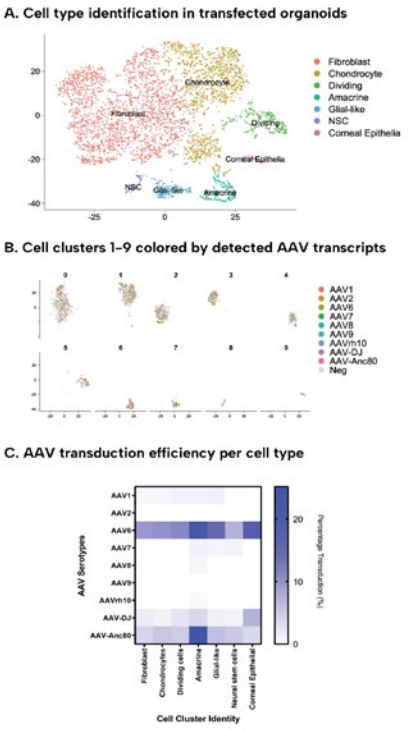
DETECT AAV TRANSCRIPTS IN SINGLE CELLS

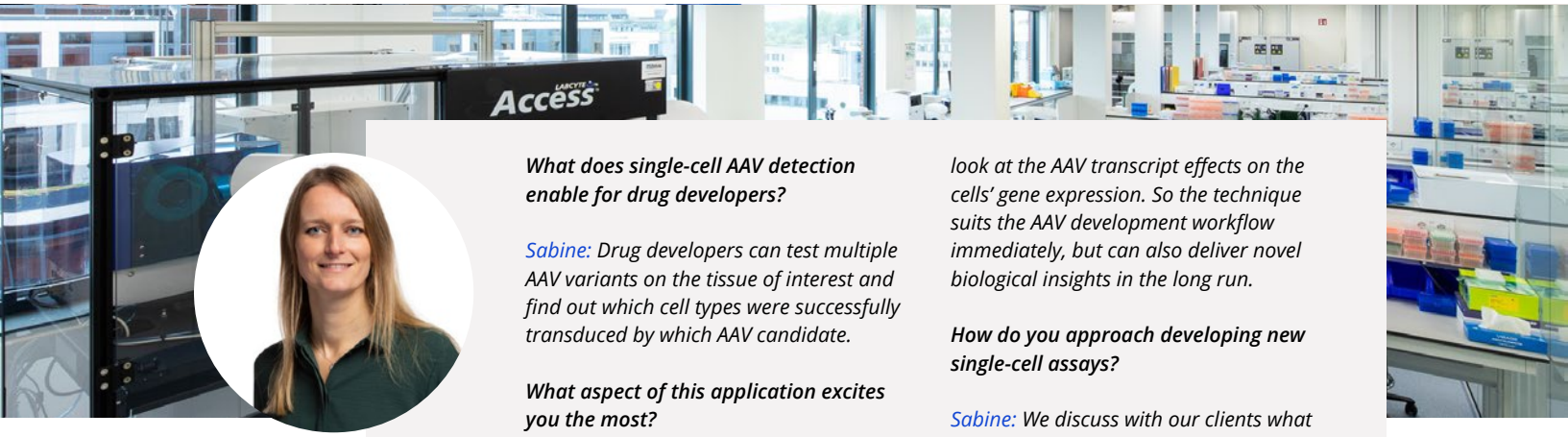
Two methods have shown succesfull AAV transcript detection in single cells: (1) AAV transcript RNA can be detected directly from single-cell mRNA sequencing libraries[2,3]; (2) custom enrichment strategies can 'fish out' AAV transcripts from single cells with targeted amplification[5].

Both methods enable multiplex measurement of transduction efficiency and specificity by measuring how libraries of delivery vectors transduce diverse cell types (see figure 1). Multiplexing is common practice in AAV biodistribution studies, for instance when analyzing different AAV serotypes or testing engineered novel capsids in high throughput.

FIGURE 1. MULTIPLEX AAV TROPISM IN OCULAR ORGANOIDS.

*(**A**) Full transcriptome sequencing clusters cells on gene expression. (**B**) Simultaneous AAV transcript detection identifies succesfully transfected cells. (**C**) Matched results measure viral tropism and quantify transfection efficiency[2].*





SABINE TANIS
Team Lead Method Development
Single Cell Discoveries

What does single-cell AAV detection enable for drug developers?

Sabine: Drug developers can test multiple AAV variants on the tissue of interest and find out which cell types were successfully transduced by which AAV candidate.

What aspect of this application excites you the most?

Sabine: The multimodal readout. The gene expression profile enables accurate cell typing. By projecting AAV detection on this, we can characterize AAV tropism at cell type-resolution. Then, the single-cell gene expression profiles also enables a deeper look at the AAV transcript effects on the cells' gene expression. So the technique suits the AAV development workflow immediately, but can also deliver novel biological insights in the long run.

How do you approach developing new single-cell assays?

Sabine: We discuss with our clients what they need and work on a bespoke solution. For this, we leverage our experience and diverse in-house expertise for protocol development and optimization. That way, clients can trust us to return custom-fit solutions for their challenging projects.

DELIVERING A SINGLE-CELL SOLUTION

Single Cell Discoveries' R&D unit has further developed these methods to reveal AAV transduction efficiency, specificity and downstream effect in single cells. Our current protocols allow our clients to measure cells' transcriptomes parallel to AAV and transgene transcripts at single-cell resolution.

This assay reveals:

1. Transduction efficiency as measured by transgene expression;
2. Cell-type-level specificity of AAV tropism enabled by cell-type annotation, based on full transcriptome data;
3. Gene expression changes in the cells that express the AAV transcript, showing the pharmacodynamic effects of the gene therapy.

By combining these readouts in a single assay, drug developers can make a better-informed selection of AAV candidates for further optimization, fuel capsid engineering efforts, and rule out vectors with low specificity earlier on in the pipeline. We are eager to continue applying single-cell RNA sequencing in collaboration with our clients to develop safer, more effective, and more accessible gene therapies.

About Single Cell Discoveries

We are focused on providing cutting-edge single-cell sequencing services and assay development to bio-pharmaceutical companies, health systems, and academic research centers globally.

In our brand-new, purpose-built laboratory in Utrecht, the Netherlands, our team of PhD-level scientists is dedicated to developing customized solutions to your unique scientific questions, all while ensuring high quality and rapid turnaround times. We receive samples from US-based clients via our sample collection point in the US.

Case study references

1. Mével, M., Bouzelha, M., Leray, A., Pacouret, S., Guilbaud, M., Penaud-Budloo, M., Alvarez-Dorta, D., Dubreil, L., Gouin, S. G., Combal, J.P., Hommel, M., Gonzalez-Aseguinolaza, G.,Blouin, V., Moullier, P., Adjali, O., Deniaud, D.,& Ayuso, E. (2020). Chemical modification ofthe adeno-associated virus capsid to improve gene delivery. Chemical Science, 11(4), 1122–1131.
2. Keng, C.T., Guo, K., Liu, Y.C., Shen, K.Y., Lim, D.S., Lovatt, M., Ang, H.P., Mehta, J.S., Chew, W.L. (2022) Multiplex viral tropism assay in complex cell populations with single-cell resolution. Gene Ther 29, 555–565.
3. Xi, Z., Ozturk, B.E., Johnson, M.E., Turunc, S., Stauffer, W.R., Byrne, L.C. (2022) Quantitative single-cell transcriptome-based ranking of engineered AAVs in human retinal explants. Molecular Therapy Methods & Clinical Development, 30(35), 476-489.
4. Brown, D. A., Altermatt, M., Dobreva, T., Chen, S., Wang, A., Thomson, M., & Gradinaru, V.. (2021) Deep Parallel Characterization of AAV Tropism and AAV-Mediated Transcriptional Changes via Single-Cell RNA Sequencing. Front Immunol, 12, Sec. Vaccines and Molecular Therapeutics.

Chapter 6 references

1. Lim, J. et al. **Transitioning single-cell genomics into the clinic.** Nature Reviews Genetics **24**, 573-584 (2023).
2. Guillaumet-Adkins, A. et al. **Single-cell transcriptome conservation in cryopreserved cells and tissues.** Genome biology **18**, 1-15 (2017).
3. Vallejo, A.F. et al. **snPATHO-seq: unlocking the FFPE archives for single nucleus RNA profiling.** bioRxiv, 2022.08.23.505054 (2022).
4. Chung, H. et al. **SnFFPE-Seq: towards scalable single nucleus RNA-Seq of formalin-fixed paraffin-embedded (FFPE) tissue.** bioRxiv, 2022.08.25.505257 (2022).
5. Xu, Z. et al. **High-throughput single nucleus total RNA sequencing of formalin-fixed paraffin-embedded tissues by snRandom-seq.** Nature Communications **14**, 2734 (2023).
6. Jiménez-Gracia, L. et al. **FixNCut: Single-cell genomics through reversible tissue fixation and dissociation.** bioRxiv, 2023.06.16.545221 (2023).
7. Fan, J., Slowikowski, K. & Zhang, F. **Single-cell transcriptomics in cancer: computational challenges and opportunities.** Experimental & Molecular Medicine **52**, 1452-1465 (2020).
8. Rood, J.E., Maartens, A., Hupalowska, A., Teichmann, S.A. & Regev, A. **Impact of the Human Cell Atlas on medicine.** Nature Medicine **28**, 2486-2496 (2022).
9. Deans, Z.C. et al. **Recommendations for reporting results of diagnostic genomic testing.** European Journal of Human Genetics **30**, 1011-1016 (2022).
10. Zhang, J. et al. **Spatiotemporal Omics-Refining the landscape of precision medicine.** Life Medicine **1**, 84-102 (2022).
11. Dagogo-Jack, I. & Shaw, A.T. **Tumour heterogeneity and resistance to cancer therapies.** Nature Reviews Clinical Oncology **15**, 81-94 (2018).
12. Moffitt, J.R., Lundberg, E. & Heyn, H. **The emerging landscape of spatial profiling technologies.** Nature Reviews Genetics **23**, 741-759 (2022).
13. Zhang, L. et al. **Clinical and translational values of spatial transcriptomics.** Signal Transduction and Targeted Therapy **7**, 111 (2022).
14. Zeng, Z., Li, Y., Li, Y. & Luo, Y. **Statistical and machine learning methods for spatially resolved transcriptomics data analysis.** Genome Biology **23**, 83 (2022).
15. Comiter, C. et al. **Inference of single cell profiles from histology stains with the Single-Cell omics from Histology Analysis Framework (SCHAF).** bioRxiv, 2023.03.21.533680 (2023).
16. Xie, R., Pang, K., Bader, G.D. & Wang, B. **Spatially Resolved Gene Expression Prediction from H&E Histology Images via Bi-modal Contrastive Learning.** ArXiv **abs/2306.01859**(2023).
17. Tsai, P.-C. et al. **Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients.** Nature Communications **14**, 2102 (2023).
18. Gavish, A. et al. **Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours.** Nature **618**, 598-606 (2023).
19. Seferbekova, Z., Lomakin, A., Yates, L.R. & Gerstung, M. **Spatial biology of cancer evolution.** Nature Reviews Genetics **24**, 295-313 (2023).
20. Barkley, D. et al. **Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment.** Nature Genetics **54**, 1192-1201 (2022).
21. Nirmal, A.J. et al. **The spatial landscape of progression and immunoediting in primary melanoma at single-cell resolution.** Cancer discovery **12**, 1518-1541 (2022).
22. Hickey, J.W. et al. **Organization of the human intestine at single-cell resolution.** Nature **619**, 572-584 (2023).
23. Shu, W. et al. **Single-cell Expression Atlas Reveals Cell Heterogeneity in the Creeping Fat of Crohn's Disease.** Inflammatory Bowel Diseases **29**, 850-865 (2023).
24. Kong, L. et al. **The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon.** Immunity **56**, 444-458. e5 (2023).
25. Danan, C.H., Katada, K., Parham, L.R. & Hamilton, K.E. **Spatial transcriptomics add a new dimension to our understanding of the gut.** American Journal of Physiology-Gastrointestinal and Liver Physiology **324**, G91-G98 (2023).
26. Garrido-Trigo, A. et al. **Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease.** Nature Communications **14**, 4506 (2023).
27. Yang, X., Hou, X., Zhang, J., Liu, Z. & Wang, G. **Research progress on the application of single-cell sequencing in autoimmune diseases.** Genes & Immunity (2023).
28. Zeng, L. et al. **Research progress of single-cell transcriptome sequencing in autoimmune diseases and autoinflammatory disease: A review.** Journal of Autoimmunity **133**, 102919 (2022).
29. Vickovic, S. et al. **Three-dimensional spatial transcriptomics uncovers cell type localizations in the human rheumatoid arthritis synovium.** Communications Biology **5**, 129 (2022).
30. Hardt, U. et al. **Integrated single cell and spatial transcriptomics reveal autoreactive differentiated B cells in joints of early rheumatoid arthritis.** Scientific Reports **12**, 11876 (2022).
31. Zhang, F. et al. **Cellular deconstruction of inflamed synovium defines diverse inflammatory phenotypes in rheumatoid arthritis.** bioRxiv, 2022.02.25.481990 (2022).
32. Piwecka, M., Rajewsky, N. & Rybak-Wolf, A. **Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease.** Nature Reviews Neurology **19**, 346-362 (2023).
33. Luquez, T. et al. **Cell type-specific changes identified by single-cell transcriptomics in Alzheimer's disease.** Genome Medicine **14**, 136 (2022).
34. Murdock, M.H. & Tsai, L.-H. **Insights into Alzheimer's disease from single-cell genomic approaches.** Nature Neuroscience **26**, 181-195 (2023).
35. Mallach, A. et al. **Applying high-resolution spatial transcriptomics to characterize the amyloid plaque cell niche in Alzheimer's Disease.** bioRxiv, 2023.06.29.546675 (2023).
36. Clark, C., Rabl, M., Dayon, L. & Popp, J. **The promise of multi-omics approaches to discover biological alterations with clinical relevance in Alzheimer's disease.** Frontiers in Aging Neuroscience **14**(2022).
37. Marshall, C.R., Farrow, M.A., Djambazova, K.V. & Spraggins, J.M. **Untangling Alzheimer's disease with spatial multi-omics: a brief review.** Frontiers in Aging Neuroscience **15**(2023).
38. Pozojevic, J. & Spielmann, M. **Single-Cell Sequencing in Neurodegenerative Disorders.** Molecular Diagnosis & Therapy **27**, 553-561 (2023).
39. Yang, A.C. et al. **Dysregulation of brain and choroid plexus cell types in severe COVID-19.** Nature **595**, 565-571 (2021).
40. Fullard, J.F. et al. **Single-nucleus transcriptome analysis of human brain immune response in patients with severe COVID-19**. Genome medicine **13**, 1-13 (2021).
41. Rendeiro, A.F. et al. **The spatial landscape of lung pathology during COVID-19 progression.** Nature **593**, 564-569 (2021).
42. Cross, A.R. et al. **Spatial transcriptomic characterization of COVID-19 pneumonitis identifies immune circuits related to tissue injury.** JCI Insight **8**(2023).
43. Stephenson, E. et al. **Single-cell multi-omics analysis of the immune response in COVID-19.** Nature Medicine **27**, 904-916 (2021).

COMMUNITY DISCUSSIONS . AI, BENCHMARKING AND WHAT IS A CELL TYPE?

OUR PENULTIMATE CHAPTER COVERS SOME OF THE EXCITING FUTURE PROSPECTS WITHIN SINGLE-CELL AND SPATIAL RESEARCH THAT COULD ADDRESS CHALLENGES FACED BY THE COMMUNITY.

First, we summarize how artificial intelligence is being used to improve single cell and, especially, spatial analysis and its ability to predict unseen aspects of single-cell biology in silico. Then, we review answers to a pressing question in single-cell biology - how do we define a cell type? Finally, we have a look at some of the community efforts to improve the uptake, understanding and to standardize the plethora of single-cell and spatial tools.

How Artificial Intelligence is shaping single-cell and spatial analysis

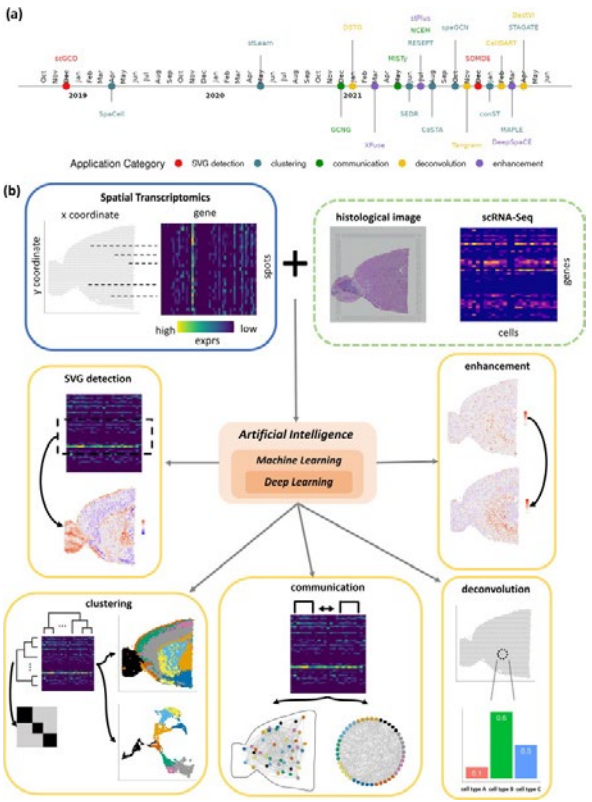
Artificial intelligence and machine learning models are a hot topic of conversation. Whilst most readers will be aware of ChatGPT and its ability to synthesize information in conversational form, perhaps less well known is the various ways artificial intelligence is being applied to biological sciences, and specifically to address the challenges faced within single-cell and spatial methods.

Machine learning methods, the most powerful forms of deep learning, have been applied to combat several of the data analysis problems in single-cell and spatial analysis (see Figure 1). Their utility comes from handling this high dimensional data in a way that utilizes most of the data compared to traditional methods. The most widely employed algorithms are autoencoders, which can capture features and improve signal/noise ratios and can learn to improve most aspects of single-cell analysis - such as removing batch effects, dimensionality reduction, clustering of cells and cell annotation. The various algorithms for the above have all been reviewed extensively[1,2].

Machine learning models for more specific challenges in single-cell analysis have already been discussed in this report in chapters 3 and 6, but deep learning has been applied very recently to more targeted problems. These include predicting stemness from single-cell data[3], for specific challenges in drug discovery such as drug-drug interactions[4], and we are seeing the use of ML models expanding into spatial analysis, with algorithms to handle the challenging problem of cell segmentation[5].

FIGURE 1. OVERVIEW OF AI METHODOLOGIES AND APPLICATION AREAS FOR SPATIAL TRANSCRIPTOMICS.

*(**A**) Timeline of AI methods for ST analysis. (**B**) applications of spatial data for which machine learning and deep learning have assisted. Image Credit: Li, et al.[7]*



June 2023 saw the release of scGPT[6] , which presents a unified solution to single-cell analysis. It is the first utilization of large-scale generative pre-trained models, such as is used in ChatGPT, to perform single-cell analysis from start to finish.

The model was trained on over 33 million cells and negated the need for dimensionality reduction. When trained on gene and cell embeddings, scGPT subsequently improved cell type annotations from the data it was trained on and was finetuned to identify new gene regulatory networks and to perform multibatch and multimodal data integration better than current popular tools. With plans to train scGPT on more single-cell data including perturbations, it is likely a matter of time before pre-trained models are integrated into standard single-cell research.

Machine learning has great promise, but there are drawbacks. For instance, how poorly understood these models are compared to the statistical models they replace, and hence why most people lack the knowledge necessary to finetune these models to get the best results. When these models are treated as a black box, there will be consequences for the biological conclusions drawn from them.

Regardless, many elements of spatial analysis still present fundamental computational problems. We asked some of our contributors what they thought the promise of machine learning was for spatial analysis.

MOHAMMAD LOTFOLLAHI

Scientist, **Helmholtz Munich/ Wellcome Sanger institute,**
Director of Machine Learning, **Relation Therapeutics**

FLG: What is the hope for deep learning models with spatial analysis? Is dealing with spatial data a different kind of problem?

Mo: *We know the location of the cell is important. For example, if two cells are close to each other, they are likely communicating, and their transcriptome is actually affected by this communication. Spatial genomics gives you the power to understand tissue organization. AI approaches allow us to understand this tissue organization. For example, we can identify clusters of cells that are always sitting together, and we can understand why those cells are actually sitting close to each other and what type of communications they're having.*

Another challenge with spatial transcriptomic is that often the location in space doesn't align for different tissue sections. If you take a slice from a region in the brain and then another region, and another, they will have completely different coordinates. One big challenge is how to put all of these things into one big map by having a common coordinate system. AI could help with that because it can give you a holistic view of the whole tissue, and maybe even for the whole body. So, you can have a 3D co-ordinate system that you can manoeuvre and then understand how the composition of the cells changes, and how the transcriptome changes. This would effectively be a 3D simulator, in which you can change something and then see the effect on the surrounding cells and also the whole tissue.

MACHINE LEARNING HAS GREAT PROMISE, BUT THERE ARE DRAWBACKS. FOR INSTANCE, HOW POORLY UNDERSTOOD THESE MODELS ARE COMPARED TO THE STATISTICAL MODELS THEY REPLACE, AND HENCE WHY MOST PEOPLE LACK THE KNOWLEDGE NECESSARY TO FINETUNE THESE MODELS TO GET THE BEST RESULTS."

JOVAN TANEVSKI

Research Area Leader, Saez-Rodriguez Group
Institute for Computational Biomedicine, Heidelberg University and Heidelberg University Hospital

FLG: What is the promise of AI and machine learning for spatial biology?

Jovan: *Spatial biology combined with machine learning will contribute to the discovery of novel biological insights. Where I expect to see most progress, is in the re-establishment of the relationship between structure and function at a higher resolution, and the improvement of our understanding of the underlying biology. Interpretable and explainable models can facilitate the establishment of new theories of tissue biology which could translate to clinical practice, complementing current histological methods and positively impacting clinical outcomes.*

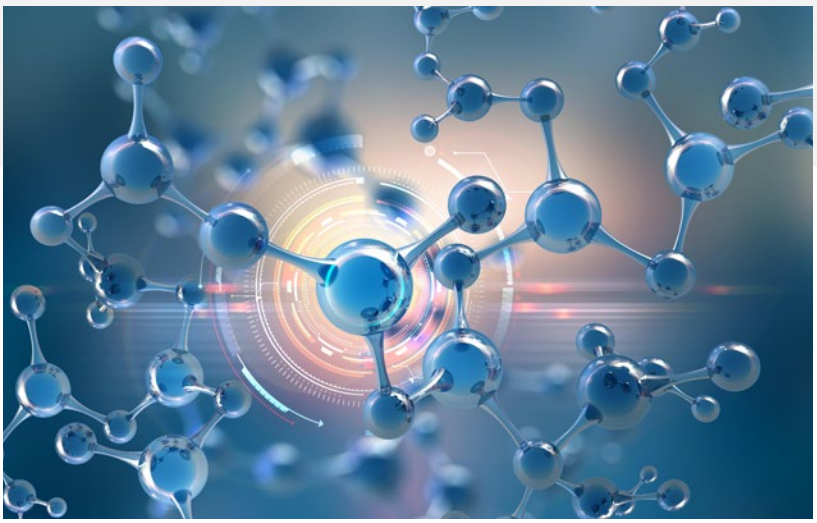
> INTERPRETABLE AND EXPLAINABLE MODELS CAN FACILITATE THE ESTABLISHMENT OF NEW THEORIES OF TISSUE BIOLOGY WHICH COULD TRANSLATE TO CLINICAL PRACTICE, COMPLEMENTING CURRENT HISTOLOGICAL METHODS AND POSITIVELY IMPACTING CLINICAL OUTCOMES."

ANNA SCHAAR

PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

FLG: What is the promise of AI for spatial biology?

Anna: *I believe that eventually, it will change the field. However, there are two aspects we need to keep in mind. The first one is data. Especially in spatial omics, we are still looking at either very limited feature spaces or limited resolution. Ideally, this issue will be minimized in the feature or resolved in efficient ways through AI. The second, very important aspect, is understanding what models learned from spatial data. AI-based methods work well according to various different benchmarks. But it is important to investigate whether the outperformance is simply due to a feature of the data or because the model in fact learned some interesting biology. This also highlights the need for well-thought-through metrics and benchmarks that assess what AI models have learned.*



There are also several challenges in making these tools relevant to the clinic[8], such as the limited available patient data to train models on. Typically, these tools are trained on atlases which tend to translate poorly to individual clinical data. This means the accumulation of high-quality single-cell data will have a great impact on the utility of the AI models, and should be a priority for the community.

One very exciting future prospect for machine learning goes further than learning to perform a specific aspect of single-cell analysis better than humans. Instead, models have the potential to learn about single-cell biology in a way that they can predict unseen aspects of it.

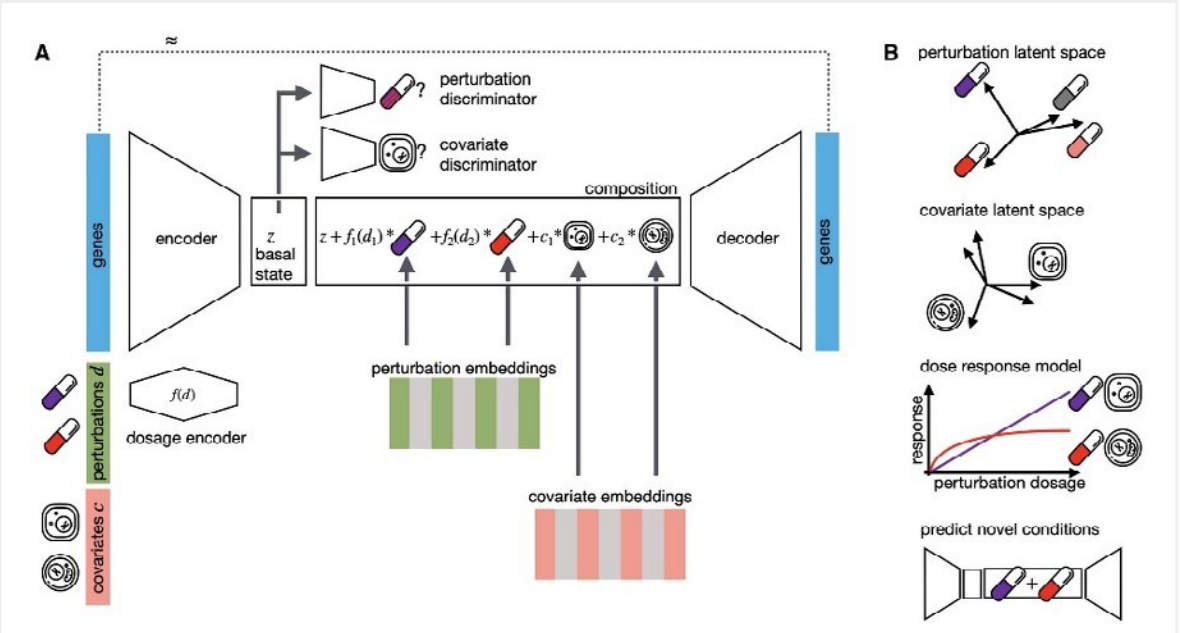
We spoke to Dr. Mohammed Lotfollahi about his most recent tool to predict cellular responses to complex perturbations, such as predicting the impact of unseen drug combinations upon cells. This highlights one of the great strengths of deep learning models; using current data to learn and predict the best avenue to explore for future experiments, so that one can escape the luck element when having to choose which of 1000's of potential experiments/drug targets/therapies to pursue.

"THIS HIGHLIGHTS ONE OF THE GREAT STRENGTHS OF DEEP LEARNING MODELS; USING CURRENT DATA TO LEARN AND PREDICT THE BEST AVENUE TO EXPLORE FOR FUTURE EXPERIMENTS,"

Their deep generative model, CPA, can predict gene expression profiles for cells that would result from unseen combinations of perturbations that it has previously learned (Figure 2)[9,10]. As mentioned in our discussion with Dr. Mohammad Lotfollahi below, this leads us closer towards the goal of training learning models so extensively on single-cell biology that they operate as in silico drug and disease testing apparatus. Whatever the future of deep learning in single-cell biology, it is already drastically improving data analysis and helping to finetune experimental procedure.

FIGURE 2. REPRESENTATION OF CPA.

*(**A**) Workflow overview. The encoder takes a matrix of gene expression and isolates the basal state before adding perturbations to output a new gene-expression matrix based upon the perturbations it assigns. (**B**) The features of CPA are plotted in latent space ad allows dose response interpolation and predictions to novel unseen drug combinations. Image Credit: Lotfollahi, et al.[11]*



INTERVIEW:

MOHAMMAD LOTFOLLAHI

SCIENTIST, **HELMHOLTZ MUNICH/WELLCOME SANGER INSTITUTE,** DIRECTOR OF MACHINE LEARNING, **RELATION THERAPEUTICS**

FLG: ***The last tool I want to discuss, the one you've most recently put out, is CPA and ChemCPA. How do they work? And what is the value of these tools to the scientific community and for drug discovery?***

Mohammad: For most diseases, we don't have a drug target. We know the phenotype of the disease i.e., how the person or sample has changed compared to healthy control, but we often don't know the underlying mechanism. For drug discovery and pharma, they're using high throughput screens, meaning that a disease sample is being tested with 1000's of drugs to see which one of those drugs can push the phenotype toward the desired healthy state. Since single-cell technology is expensive, you cannot keep doing this with 1000's of drugs and then when you go to drug combinations. That's even worse because these drug combinations represent a combinatorial space.

So, you need in silico algorithms that will be able to generate single cells and predict how transcripts will change in response to a specific drug or genetic manipulation. That's what we were doing with CPA and ChemCPA. CPA is the general framework that allows you to predict single-cell behaviour changes (i.e., transcription changes) to both drugs and genetic manipulations.

The way it works is a bit like a Lego. If you can imagine a cell as being in a basal state. This is a single Lego block, and then you can add a second Lego block to it and that makes it a neuron, and then you can add a third piece of Lego and that represents the effect of a drug. This algorithm would separate these three things out. It reverses the effect of the cell type, it reverses the effect of the drug, and it tries to then put them back together. Once it has learned to separate things, then you could replace one of those components with a new component.

So, you can replace the drug with another drug that you haven't tested, and CPA actually allows that. Then you can predict the transcriptome effect of that drug or a specific genetic manipulation.

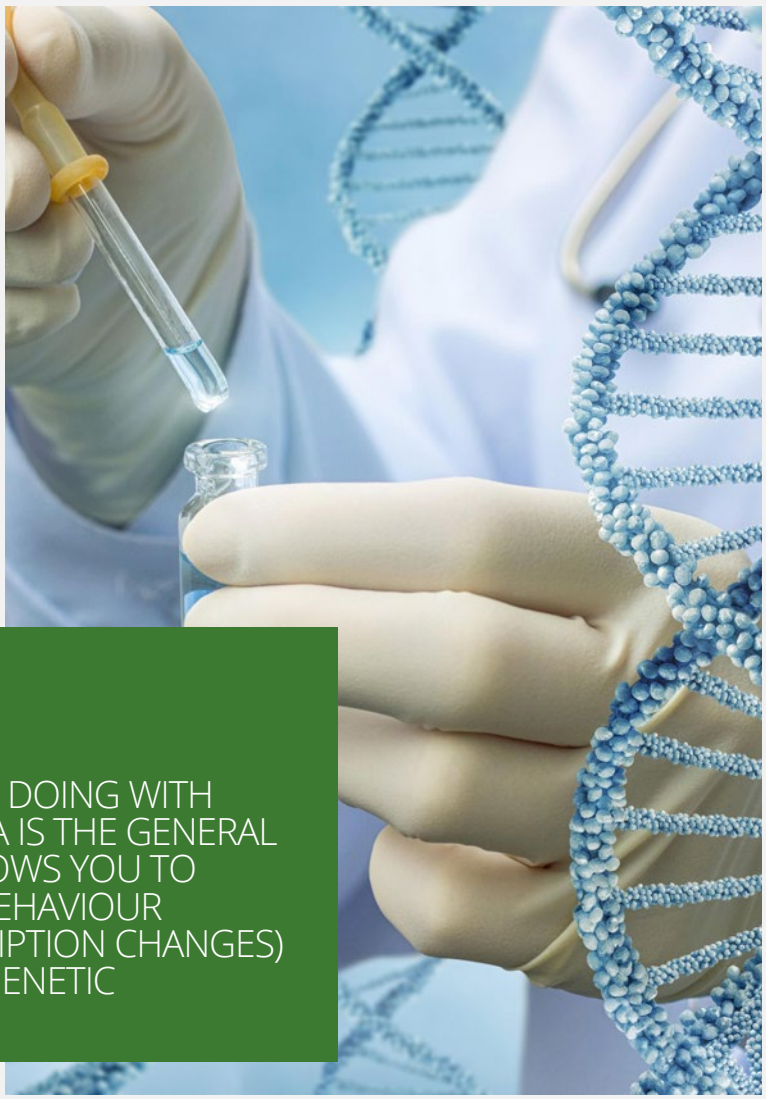
Then with ChemCPA, we incorporated the structural identity of the drugs. So now the model is not just conditioned on the transcriptome effect, but also on the structure of the drug. It's hypothesized that drugs with similar structure might induce a similar effect, and we leverage that assumption here. If the model sees enough drugs with a diverse set of structures, it might be able to predict the effect of drugs it has never seen.

***FLG:* And did you see validation for the algorithm?**

Mohammad: Yes. In the paper we applied the algorithm to three cancer cell lines. We had the responses of these cancer cell lines, breast cancer, leukaemia and lung cancer, to 188 different drugs. Then we trained CPA, which inferred a perturbation map of which drugs would induce similar responses. If two drugs induce similar behaviours in the cells, they will be embedded close to each other in this perturbation space. We use that perturbation space to select drug candidates that were actually effective on the cells to design a second experiment, but this time with combinatorial drugs. We subsetted the 188 drugs, down to 32 mono, and then, combo drugs. Again, we trained the model with the new data set, but this time we left out bunch of these combinatorial drugs. We showed that the model predicted the right transcriptome effect for them, it matched what the experimental data told us.

***FLG:* That's really promising! Would you say the ultimate goal for this kind of work is to build an in silico drug testing suite?**

Mohammad: Exactly. That's the ultimate goal. One where you have a gigantic in silico space of chemicals or genetic perturbations, and then you can narrow down the space to say, three drugs that are really effective against the disease. Then you'd go and test them and then this data could be fed back into the algorithm. If the algorithm predicted an effect, you go and generate data, you feed it back to the algorithm and it gets better and better for multiple iterations. At some point, you reach the phenotype that you desire. That would be the ultimate goal. It could give you a simulated space for perturbing the tissue and the cell, and that will help you understand and design your experiments.



"THAT'S WHAT WE WERE DOING WITH CPA AND CHEMCPA. CPA IS THE GENERAL FRAMEWORK THAT ALLOWS YOU TO PREDICT SINGLE-CELL BEHAVIOUR CHANGES (I.E., TRANSCRIPTION CHANGES) TO BOTH DRUGS AND GENETIC MANIPULATIONS."

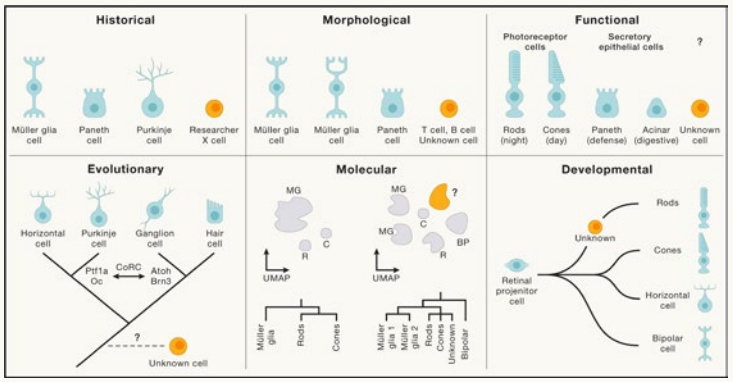
What is a cell type?

Classifying cells into cell types is both practical and informative in the era of single-cell transcriptomics. However, given the centuries-long interest biologists have had in understanding organisms at the cellular level, and the immense value of a standardized cellular taxonomy, it is quite remarkable how poorly defined a 'cell type' still is.

Do we classify cells by their shape, locale, development/ontogeny or how they interact with other cells? Historically, all of the above have been used (see Figure 3 for overview of classification schemes).

FIGURE 3. VARIOUS WAYS TO CLASSIFY CELLS. CELL TYPE NOMENCLATURE USES A SELECTION OF ALL OF THESE, BUT NONE ARE SUFFICIENT ALONE AND A CENTRALIZED SYSTEM IS NEEDED GOING FORWARD.

Image Credit: Domcke and Shendure[12]



With the advent of single-cell transcriptomics, and the ability to explore genomics in individual cells, our ability to understand cell types has been transformed from observational and small scale to building whole atlases of cell types based on transcriptomic differences.

The majority of single-cell sequencing efforts have used the most accessible metric, RNA levels, as the basis of cellular identity. This results in portioning out clusters of cells based on transcriptomic similarities. This has revealed the unexpected cellular heterogeneity of multiple tissues when it comes to RNA expression.

But is RNA a sufficient marker of cellular identity?

The answer is no. RNA is short-lived and not a reliable marker of identity, particularly in snapshot form (such as in a single-cell sequencing experiment) since RNA can vary significantly between cell states. Ultimately, transcription shows a picture of a cell's potential to make proteins, not its actual protein constituent nor its function.

Transcriptomics has been essential for moving cell classification away from identifying cells based on shape and size, but converging expert opinion suggests that[13], for our definition of a cell type to be meaningful, it must be associated with what cells do, which means transcriptomics must be linked to anatomical and functional information too[14].

Several attempts have been made to create a new standard to classify cells by. Morris[15] recently suggested three central pillars which when taken account of, would construct a high-resolution, dynamic cell identify landscape[15]:

Phenotype – the physical, molecular and functional features of a cell.
Lineage – the developmental origins of cell.
State – although cell identify is stable, it must be defined in a flexible enough way to remain consistent across the variety of states a cell may be in.

To try to capture this information, it has been suggested that a periodic table of cells could be used to classify cells. In this table, each row would capture the cell differentiation trajectory of a cell type, organised by developmental origin and each column would capture cells fates from conception to maturity. This periodic table could operate as a central resource for identifying cell types, and could be used to predict missing cell types following development and differentiation logic[16].

How do we build this table?

A practical step by step guide[11] suggested we first use transcriptomics information as an anchor to which other multimodal information (namely epigenomic and spatial) can be incorporated to distinguish these types further. These multimodal transcriptomic clusters can then undergo anatomical and functional assessments, followed by developmental trajectory analysis and cross-species comparison to reflect the reality that cell types are the product of evolution.

However, the most recent foray into this topic argues against atlases and periodic tables, which still rely heavily on pre-defined marker genes to segment the original cell types. It instead suggests a new organizing principle for cell types - a reference cell tree - which is data driven and based on consensus ontogeny[12].

The authors of this publication tackle a few cloudy elements when it comes to cellular identity. Firstly, they highlight how muddled the terminology is around cell types and provide clear definitions of the terms cell type, state, identity etc. (See box). Secondly, they provide clear aims for a cellular classification system.

Cell type: A recurring pattern of developmental origin and potential within and across cell lineage trees of individuals of a given species, generally reflected in shared molecular properties.

Cell state: Variations in molecular phenotypes within a cell type that do not impact its developmental potential (e.g., cell cycle, stochastic fluctuations).

Cell identity: An individual cell as characterized solely by its molecular phenotypes at a given moment in time.

Cell lineage: The relationships among cells of an individual organism as defined solely by the series of cell divisions that begins with a single zygote.

Cell trajectory: Ordering of cells' developmental relationships inferred solely from similarity in molecular phenotypes, which might or might not recapitulate developmental cell lineage relationships.

They claim we should be looking for a cell type nomenclature that (1) identifies all cells during a lifecycle, (2) accounts for inter-individual variation and disease state changes, (3) is biological/ functionally meaningful but - most importantly-, (4) a nomenclature for cells which will last 100's of years into the future so that we can successfully compare the same cells across studies.

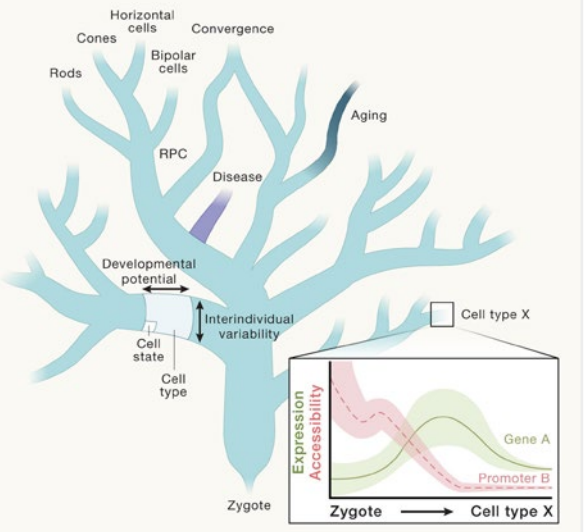
When evaluating the different organizing principles detailed in Figure 3, none of them are sufficient to classify cells to meet this standard. However, if the data-driven power of molecular profiling was combined with a form of lineage tracing, would this meet the criteria?

With effective time-resolved lineage tracing, molecular profiling of cells at multiple developmental time points, and methods to link cell states between different developmental time points, we could effectively build developmental/molecular reference guides for cell types which could act as a unifying reference for all cell typing. This would allow us to move away from organizing cells by identity and difference (Tables and Atlases) and instead towards relating cells to one another through functional and temporal relationships (Trees). See Figure 4 for an exemplar tree.

Whether in the form of a periodic table or reference tree, the cataloguing of cells will not persist as unidimensional transcriptional clusters in the age of spatial biology, multiomics and live sequencing. To get some further insights into reference cell trees we reached out to the paper's first author, Dr. Silvia Domcke, for more clarity on how these trees can be constructed and the value they would provide.

FIGURE 4. EXEMPLAR CELL REFERENCE TREE INCORPORATING THE ENTIRETY OF ORGANISM DEVELOPMENT ALONGSIDE MOLECULAR MEASUREMENTS.

Image Credit: Domcke and Shendure[12]



INTERVIEW:

SILVIA DOMCKE

AFFILIATE ASSISTANT PROFESSOR, DEPT. OF GENOME SCIENCES, **UNIVERSITY OF WASHINGTON,** ASSOCIATE DIRECTOR, HEAD OF HUMAN GENOMICS, **GORDIAN BIOTECHNOLOGY**

FLG: ***In brief, how do we transition from single-cell atlases to single-cell reference trees? What steps need to happen?***

Silvia: We are increasingly adept at profiling molecular measurements (such as gene expression or chromatin accessibility) at single cell resolution for an ever expanding breadth of tissues and even whole organisms (i.e., 'atlases'). However, it is challenging to ascertain how these cells relate to one another, both within one study or modality, and even more so across studies or modalities (e.g., gene expression and chromatin accessibility). If we measured the 'lineage', the relationships among cells, as defined by the series of cell divisions that begins with a single zygote alongside the molecular profiles in each cell, we could use this to derive a 'reference cell tree' for each organism. This would represent a ground truth framework onto which all other datasets could be projected and this would be analogous to projection of local chromatin states onto the reference genomes constructed by the Human Genome Project.

CRISPR-based lineage-tracing systems can perform continuous recording of cell lineage for weeks with multiple events per cell division, and capture scRNA-seq profiles from the same cells[17] To build a reference tree, we would need to:

- Subject a series of progressively older embryos, as well as adult organ systems, to a flavour of molecular recording that yields a comprehensive lineage tree, with rich molecular states for terminal nodes (e.g., scRNA-seq, scATACseq etc.).

> "ONE IMMEDIATE ADVANTAGE WOULD BE THAT CELLS FROM DIFFERENT STUDIES CAN BE RELATED TO ONE ANOTHER AND BE UNEQUIVOCALLY NAMED."

- Bridge gaps using methods that additionally record information about the molecular identities of their ancestors18, as well as inference.
- Such trees would then be merged across individuals to yield a molecularly annotated consensus ontogeny of a given species using phylogenetic algorithms19, statistically bounded by the invariant aspects of this organism's development.
- A data-driven consensus nomenclature could be applied to unequivocally define and name cell types within branch segments of the tree, based on objective criteria such as maximum information gain, and related to currently used cell type names as 'synonyms'.
- Once the reference tree is constructed, we would not need to measure lineage alongside each molecular measurement to enable projection onto the tree.

***FLG:* What are the biggest problems that we will run into when trying to make trees over atlases?**

Silvia: Although methods exist to measure lineage and different molecular profiles in the same cell, they need to be further optimized to achieve the necessary resolution. In addition, this information will need to be collected for many individuals of the same organism.

A further key challenge is that genome-editing based lineage tracing is - of course - not an option in humans. There are at least four avenues to address this challenge:

(1) Generation of consensus ontogenies of closely related organisms.

(2) In vitro human "stembryo" models.

(3) Lineage tracing based on somatic mutations in chromosomal or mitochondrial DNA.

(4) Lineage inference based on molecular profiles from human tissues, the accuracy of which can be assessed on organisms for which lineage data exists.

Similarly, when it is not possible to apply lineage-tracing approaches to pathological states, cell states with only molecular information could still be mapped onto the lineage tree by using inferred trajectories.

There are further practical challenges, such as the logistics of data storage, analysis and maintenance.

***FLG:* What would be the direct advantage to a researcher in the field or a clinical researcher to having cell reference trees (consensus ontogeny) as opposed to atlases?**

Silvia: One immediate advantage would be that cells from different studies can be related to one another and be unequivocally named. In addition, it enables the nomination and validation of genes, such as transcription factors, that shape specific cell type transitions. For diseases in which developmental processes are involved, it provides a whole new understanding of how disease phenotypes arise and helps characterize both inter- and intra-individual phenotypic variation. It also supports the systematic 'placement' of in vitro systems (e.g., organoids), which are increasingly used in preclinical research, in relation to wild-type development. Moreover, it would facilitate aligning cell types or organ systems across species and could thus highlight potential key differences when translating research findings from animal studies to the clinic.

Community Efforts to teach and standardize single-cell and spatial

With fast-paced progress comes an accumulation of problems and challenges. Current users of single-cell and spatial methods have various levels of understanding when it comes to how best to utilise these methodologies. One of the recent movements in the field has been a push towards standardization and benchmarking methods to enable comparable single-cell and spatial experiments to be performed across the world, and to identify a gold standard approach.

Many members of the community have put their energy into this, as well as into producing resources and fostering communities to assist technology users with how to use computational tools, how best to prepare samples, or simply a place to try and troubleshoot problems. Below we list some of these initiatives as well as comments from organizers within them to showcase the essential progress the community is making to standardize and broaden single-cell and spatial methodology.

Single Cell Ninjas. Perhaps the most well-known community in the single-cell space is the Single-Cell Ninjas, a group of researchers with extensive expertise on performing single-cell experiments and analysing the data. Established by Luciano Martelotto and Catia Moutinho, the community consists of 100's of single-cell and spatially minded individuals and currently operates on twitter at the following handle @Sc_Ninjas. Their aim is to help people tackle the challenges they have in running their single-cell experiments, to troubleshoot problems and provide advice on best technology and protocols. Luciano is also part of a new platform called GESTALT - @GESTALT_sp with a focus on spatial technologies, we spoke with him to get an insight into these two endeavours.

LUCIANO MARTELOTTO
Associate Professor & Lab Head, Single Cell and Spatial-Omics Laboratory
University of Adelaide, Australia

FLG: Can you describe yours and Catia's project, the Single Cell Ninjas?

Luciano: *The Single Cell Ninjas is a platform we maintain on Twitter. We established it after realizing how much knowledge we were gaining from the 'wild'—from hands-on experience with these technologies and through interactions with various individuals. We wondered, why not share this knowledge with those who might not have access? And I'm not just talking about those in Boston or major USA, Australian or European institutions. I'm referring to individuals in underrepresented countries with limited access to infrastructure and experts. This was the genesis of Single Cell Ninjas. Over the past few years, we've provided valuable advice and have worked diligently to democratize knowledge.*

A group of us, unrelated to the Single Cell Ninjas, recently launched a platform named 'GESTALT' or Global Alliance for Spatial Technologies, which is primarily focused on spatial omics. You can also find us on Twitter @GESTALT_sp. We felt it was crucial to establish a group that functions as a global alliance for spatial technologies. The primary objective is to eliminate any ambiguity or confusion in terms of what's possible now and what likely happen in the future, and at the same time democratize spatial-omics knowledge and access. Once we achieve that clarity, our aim is for this platform to serve as a hub for collaboration and knowledge sharing. This will be a place where we can work together, form partnerships, standardize processes, and foster understanding about interoperability. That's the essence of our initiative.

The Single-Cell World. Catia Moutinho has another platform she runs to disseminate information called The Single-Cell World. The goal of this platform is to disentangle 'single-cell technology in a way that anyone will understand'. She also produces a podcast focused on understanding single-cell sequencing processes and things you might want to consider when deciding the best approach for your experiment.

CATIA MOUTINHO

Founder & Scientific Advisor
The Single-Cell World

FLG: As part of the Single-Cell World project, you produce an excellent podcast covering topics and trends in single-cell, what was the motivation behind the Single-Cell World Podcast?

Catia: *I love podcasts. I love to read but I don't love to read science papers. It was always an obligation, to be honest. I love to read to disconnect. But if I'm listening to a science podcast, it's easier. I can go for a walk and enjoy life while I am learning. And since the Single-Cell World is my project, I can decide to do what I like most, so I thought 'okay, let's do a podcast.' This was the motivation. And also, I like to talk with people. And it's easier for me to explain through talking rather than with writing. I also write the Single-Cell World blog.*

FLG: How do we help researcher's struggling to keep up with the fast-paced development of these tools?

Catia: *Right now, as a freelancer scientific advisor, one of my tasks is to be updated on single-cell technology. Before when I was a Group Leader, it was different, and I remember I didn't had time-to keep myself on top of things. I knew about most of the updates because companies were contacting our lab with new products and technologies. Otherwise, I couldn't get the time to read, to attend the different webinars, to check all social media information. Back then I thought, it will be amazing if there is one place where I can find all the new information. This was one of my goals with the Single-Cell World project, besides sharing my knowledge and help scientists. So, this is my strategy - to try to help.*

Sanbomics. While some people like to listen, other people like to watch. An excellent video resource to understand and visualize single-cell analysis is the Sanbomics YouTube channel run by Mark Sanborn. Along with statistical and computational walkthroughs, there are also in-depth guides on performing different aspects of single-cell analysis for beginners, including a video that is over an hour long of Mark performing single-cell RNAseq analysis from start to end.

scRNA-tools. Switching focus to community-led resources for standardization, there are some very useful resources such as scRNA-tools, which is a catalogue of all single-cell RNA tools (currently with over 1500 tools) with information such as the language they're written in, the number of citations using the tool, and the broad areas of single-cell analysis the tool works in. Here you can find all available tools for clustering, dimensionality reduction or visualization with ease.

Sajita Lab. The Sajita lab website, led by Rahul Sajita, Associate Professor, NYU, is a useful resource. The Sajita lab hosts a yearly Single Cell Genomics Day, a virtual practical workshop where you can hear about new single-cell tools, as well as the yearly presentation of the top 10 single-cell developments of the year. There are resources such as guidance for their popular tool - Seurat, a slide deck to help choose the right single-cell tools and a cost per cell calculator to help plan your experiments

Open Problems. Benchmarking computational tools is a pressing community effort and Open Problems in Single-Cell Analysis is another community effort targeting the vast quantity of software available for single-cell analysis by providing an open source, community driven platform for continuously updated benchmarking of defined tasks in single-cell analysis. Here you can find updated benchmarking for spatial decomposition, denoising, batch integration and more.

Scverse. The scverse[20] is a multi-institution open-source software project in which a group of highly talented early career computational biologist, who have produced widely used software packages, gather to address storage and analysis needs of single-cell data and provide software-specific troubleshooting. The goal is to build an essential infrastructure for single-cell analysis, and to attempt to counter the issue of an ever expanding set of scattered and overlapping tools by centralizing their tool set and performing long-term maintenance. A community is also building in the scverse with events such as Hackathons and community meetings regularly scheduled, and an active discourse forum to discuss general and software-specific points of interest. . We spoke to Anna and Lukas, two core members of the scverse team.

ANNA SCHAAR

PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

LUKAS HEUMOS

PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

FLG: Can you describe what the scverse is? Why was it set up? And what do you hope it will achieve?

Lukas: *A postdoc in our lab, Luke Zappia, published a paper that visualized a graph of all major packages in Python for single-cell and highlighted dependencies. This figure highlighted that the packages are very centralized because the whole community had built on top of our 'anndata' and our 'scanpy' packages, which are hugely popular for analysis with Python. We knew that these core packages are only really being maintained by our lab, alongside a few external contributors. It had grown a bit out of our scope, it had become too big. We also noticed other major packages, such as scvi-tools, muon, scirpy, in the community that other people have built that we thought must also be available and maintained in the future.*

Hence, we talked to the other developers and thought, 'Okay, can we find a consortium that maintains all of these packages jointly', to ensure that it's not only in our lab's hands anymore, but in the hands of everyone together? This would enable us to share our resources and jointly define the future of these packages. That's why we founded scverse. You can see it as a Bioconductor but for Python, and initially with a single-cell focus, but this might change in the future. Currently, we've hosted several hackathons and are strongly focused on community building. We recently became a NUMFOCUS-sponsored project, which means that we are moving even further away from our labs and becoming our own entity.

FLG: Great, and can you also tell me a little bit about the community aspect and the teaching aspect of scverse?

Anna: *Scverse is more than just the core developers that have built these foundational tools for analysing single-cell data. For us, the community that wants to interact and engage with developers and others in the field is the main focus of our efforts. We actively interact on multiple different communication channels with our community, namely GitHub, discourse, and also on Zulip. People can join these channels, raise issues and ask questions.*

Beyond these channels, we host the scverse open community meetings every second Tuesday. In these meetings, we invite people that developed packages that are part of the ecosystem of scverse. People can come and listen to the core developers, ask questions, and learn how scverse packages are built and can be used. They are a great way of getting to know the developers behind the projects. ▶

Additionally, our YouTube channel hosts all the community meetings for those who couldn't attend or joined our community later. For in-person engagement, we also organize scverse hackathons. Currently, these are invite-only and mainly driven by core members of the scverse, but we are also planning to open them up to everyone in the community. And last, but not least, we will soon also host scverse workshops, the scverse training program for learning how to perform single-cell analysis in Python, and our first scverse conference in 2023.

Single-Cell Best Practices. Anna and Lukas have spearheaded another highly valuable community resource. They used the expertise of scverse and many other experts to build a guide of Single-cell Best Practices[21]. Naturally, every experimental design requires adjustments from a standard guide, but the community has been calling out for a guidebook of how to perform generic single-cell analysis to use as a baseline.. The advice in the book ranges from analysis methods for single-cell RNA sequencing to tackling the challenges in spatial omics and multimodal integration. This book is informed and reviewed by experts and can be used by anyone. The goal is for this to be a living resource which will be updated as the technology improves.

ANNA SCHAAR

PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

LUKAS HEUMOS

PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

FLG: Let's talk about your resource, the 'best practices in single-cell'. Do you think community-led movements such as scverse and 'best practices in single-cell' are the best way to keep on top of the mass production of analysis tools? Also, for the 'best practices in single-cell', you did a really good job of picking out a typical/starting point toolset to use for single-cell analysis. How do you plan to kind of keep on top of all these new methods that are emerging?

Anna: *One reason why Lukas and I started the best practices in single-cell was in fact the immense collection of tools. As Lukas mentioned, our colleague Luke Zappia maintains a website (scrna-tools.org) tracking the number of analysis tools available for single-cell, and it recently exceeded 1,500 tools. And it's super challenging to stay up to date on what works best and why, not only for new joiners in the field. On top of that, our colleague Dr. Malte Luecken published a paper in 2019 on current best practices in single-cell RNA-seq analysis - we internally call it the best practices 1.0 paper. This paper and its recommended workflow were a great success, but only covered scRNA-seq analysis. Since 2019, the field grew massively and there was an increasing need for an updated version also accounting for additional modalities. This gave us the motivation to write a whole new book on this.*

Of course, it is challenging to maintain the best practices, but that's also where we want to engage the community. We encourage everyone to propose updates, new notebooks, or chapters to the best practices online book. Internally, we update the online book whenever there is a new benchmark published, so we ensure our recommendations are up-to-date. We also implemented a review process for the online book to ensure no bias in recommended methods or tools. Our goal for the best practices book is that it's a living resource from the community to the community and not just driven by a small group of people in Munich.

FLG: What was the motivation to create Single-Cell Best Practice book, 2.0?

Lukas: *As Anna mentioned, first of all, it's the number of tools. We can't keep track of them anymore, and we also need to know which ones work best.* ▶

The second reason is that we want to formulate an opinion that is mostly neutral and less subjective. So, an approach that is benchmarking-based is what Dr. Malte Luecken (Group leader, Helmholtz Munich) pioneered. However, that inspiring effort is outdated because he only performed it on RNA-seq, and the field has gone multimodal. We wanted to improve on that and create a living standard, so we wrote the book together with the associated paper. We just want to ensure that we have a resource that can be updated and that guides users.

The third reason is teaching. Why do we constantly have to redevelop our resources? Because the tools change, the approaches to analysis change, and we just want to develop these practices with the community so that it's not just us updating the teaching materials. Instead, we are updating them with the community, which is way more efficient.

FLG: And having put together a massive resource like this book, are there any areas of analysis that you feel are coming to a peak and are there obvious areas that are still rapidly developing?

Anna: *Interestingly, all areas are still moving in the field. For the best practices online book, we differentiate tools into benchmark validated, well-described, or well-established analysis steps and novel analysis steps that still need to be fully explored by the community. Interestingly, we see movements and new developments in all three categories. One example is a classical preprocessing step for single-cell RNA-seq data: normalization. A benchmark published this year claimed they found the best working normalization techniques. But since those well-established analysis steps are already commonly used by the community, it is, of course, easier to describe in the best practices book. The more challenging aspect is novel modalities and analysis steps. We simply still need to learn and investigate how to best analyse new single-cell measurements and what methods and tools work best. So, there's more freedom and movement in general. For the best practices online book, however, we see movement and development in all aspects.*

Benchmarking in proteomics and spatial. The less mature fields such as spatial transcriptomics and proteomics have yet to build centralized benchmarking criteria and gold standards, but this process is well underway. One example is the benchmarking and community-based guidance that is being accumulated for single-cell proteomics[22], spearheaded by Nikolai Slavov, Associate Professor, Northeastern University. Furthermore, there are resources that have begun benchmark specific aspects of spatial analysis, such as cellular deconvolution[23].

Chapter 7 references

1. Brendel, M. et al. **Application of deep learning on single-cell RNA sequencing data analysis: a review.** Genomics, Proteomics & Bioinformatics **20**, 814-835 (2022).
2. Bao, S. et al. **Deep learning-based advances and applications for single-cell RNA-sequencing data analysis.** Brief Bioinform **23**(2022).
3. Dezem, F.S. et al. **A machine learning one-class logistic regression model to predict stemness in single cell transcriptomics and spatial omics datasets.** bioRxiv, 2023.05.04.539461 (2023).
4. Qi, R. & Zou, Q. **Trends and potential of machine learning and deep learning in drug study at single-cell level.** Research **6**, 0050 (2023).
5. Chen, H., Li, D. & Bar-Joseph, Z. **Cell segmentation for high-resolution spatial transcriptomics.** bioRxiv, 2023.01.11.523658 (2023).
6. Cui, H. et al. **scGPT: Towards building a foundation model for single-cell multi-omics using generative AI.** bioRxiv, 2023.04.30.538439 (2023).
7. Li, Y., Stanojevic, S. & Garmire, L.X. **Emerging artificial intelligence applications in Spatial Transcriptomics analysis.** Computational and Structural Biotechnology Journal **20**, 2895-2908 (2022).
8. Ma, A., Wang, J., Xu, D. & Ma, Q. **Deep learning analysis of single-cell data in empowering clinical implementation.** Clinical and Translational Medicine **12**, e950 (2022).
9. Lotfollahi, M. et al. **Biologically informed deep learning to query gene programs in single-cell atlases.** Nature Cell Biology **25**, 337-350 (2023).
10. Welch, J. **Multiplying insights from perturbation experiments: predicting new perturbation combinations.** Mol Syst Biol **19**, e11667 (2023).
11. Lotfollahi, M. et al. **Predicting cellular responses to complex perturbations in high-throughput screens.** Mol Syst Biol **19**, e11517 (2023).
12. Domcke, S. & Shendure, J. **A reference cell tree will serve science better than a reference cell atlas.** Cell **186**, 1103-1114 (2023).
13. **What Is Your Conceptual Definition of a Cell Type - in the Context of a Mature Organism?** Cell Systems **4**, 255-259 (2017).
14. Zeng, H. **What is a cell type and how to define it?** Cell **185**, 2739-2755 (2022).
15. Morris, S.A. **The evolving concept of cell identity in the single cell era.** Development **146**(2019).
16. Xia, B. & Yanai, I. **A periodic table of cell types.** Development **146**(2019).
17. Choi, J. et al. **A time-resolved, multi-symbol molecular recorder via sequential genome editing.** Nature **608**, 98-107 (2022).
18. Chen, W. et al. **Multiplex genomic recording of enhancer and signal transduction activity in mammalian cells.** bioRxiv, 2021.11.05.467434 (2021).
19. Stadler, T., Pybus, O.G. & Stumpf, M.P.H. **Phylodynamics for cell biologists.** Science **371**(2021).
20. Virshup, I. et al. **The scverse project provides a computational ecosystem for single-cell omics data analysis.** Nature Biotechnology **41**, 604-606 (2023).
21. Heumos, L. et al. **Best practices for single-cell analysis across modalities.** Nature Reviews Genetics **24**, 550-572 (2023).
22. Gatto, L. et al. **Initial recommendations for performing, benchmarking and reporting single-cell proteomics experiments.** Nature Methods **20**, 375-386 (2023).
23. Li, H. et al. **A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics.** Nature Communications **14**, 1548 (2023).

GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL BIOLOGY LOOK LIKE IN 15 YEARS' TIME?

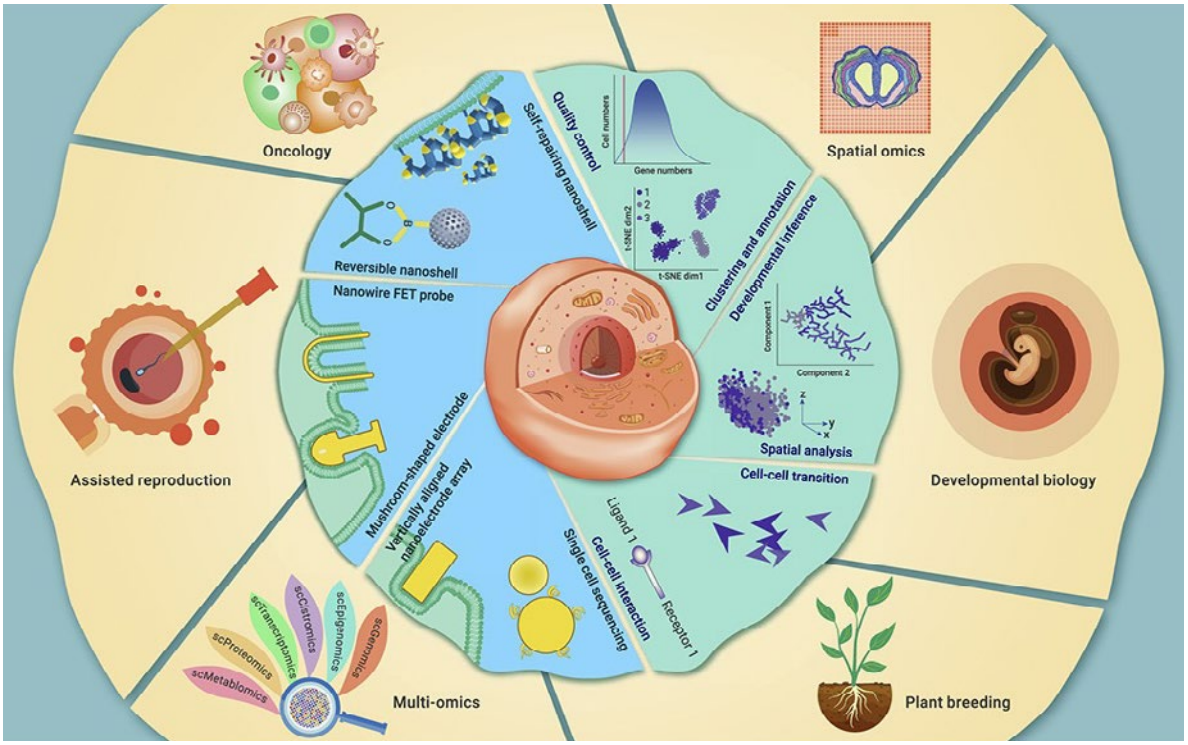


Image Credit: Wen, et al.[1]

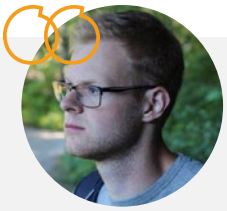
THE RATE OF PROGRESS IN THE PAST 15 YEARS OF SINGLE-CELL BIOLOGY HAS BEEN ASTONISHING. THE UPTAKE OF SPATIAL BIOLOGY TO VIEW SINGLE CELLS IN CONTEXT HAS BEEN EVEN FASTER. ULTIMATELY, WE HAVE SEEN THE POWER THIS TECHNOLOGY HAS TO RESOLVE CELL HETEROGENEITY IN HEALTH AND DISEASE AND BRING US CLOSER TO TREATMENTS.

A recent editorial[2] asked what does the future hold for single-cell biology? An impossible question to answer with any accuracy in anything but the immediate short term. The editorial envisioned a single-cell world in which large scale data is standardized and brought together to help understand disease states and perturbations caused by drugs. A world in which multimodal data are integrated to understand the spatiotemporal steps that individual cells take when developing into tissues. A world with the ability to expand Live-seq to perform real-time measurements of multimodal molecules across many cells. A world where, perhaps ultimately, we create a machine learning model of the cell that we can interact with and test the impacts of diseases in silico.

A conservative view of the future might expect better models and bigger data. It would expect more modalities to be integrated into the ever expanding atlases, including data from perturbations and altered phenotypes.

The 15-year anniversary of single-cell transcriptome sequencing is just around the corner. To the eyes of someone in 2009, the cell-throughput and multiomics capacities of these tools today look like centuries of progress rather than the work of one and a half decades. We thought we would challenge our contributors with the question: **What do you think single-cell and spatial methods will look like in 15 years' time?**

FLG: What do you think single-cell and spatial methods might look like in 15 years' time?



ROBIN BROWAEYS

Team Leader – Bio-IT Support
VIB Centre for Inflammation Research, Ghent University

Robin: I'm hoping to see clinical impacts at that point. In single-cell, we have way more datasets generated for many more biological systems, mouse models, diseases. The technology has become much more accessible to people and that also sparks tremendous interest in computation tool development. I think the same will happen for the spatial field. Many more datasets from many more interesting biological cases that will be studied by those technologies. That will then spark an interest in computational tools to analyse it.

It's already started, of course. In terms of technology development, I hope we will have single cell resolved transcriptome-wide spatial data. It would be a tremendous wealth for cell-cell communication modelling because you can then really start modelling interactions between cells in the tissue context.

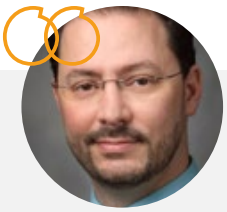
JASON D. BUENROSTRO

Associate Professor & Broad Institute Member
Harvard University & Broad Institute of MIT and Harvard

Jason: What I hope for, and it's hard to know if we'll actually get there, is that that single cell biology becomes boring, because if it becomes boring, that means it's mature enough that people will just use it without necessarily having to consider the question, 'what analysis tool do I use?'. I'm hoping it'll be as robust as some of the alignment tools we use for aligning bits of the genome. Single-cell tools will be like that, because when that happens, then we can start to have a real impact in the clinic and can use this for diagnostic purposes.

That's where I like to think that single-cell and spatial genomics will be in 15 years. It'll be commonplace and routine within research with a really strong, mature sense of what's a robust measurement, and what is an artifact, and how we analyse the data. That area will be so robust that we'll be tempted to start using it for diagnostics to improve patient outcomes. That's my vision for the future.

GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL BIOLOGY LOOK LIKE IN 15 YEARS' TIME?



JARED K. BURKS

Professor & Co-Director, Flow Cytometry & Cell Imaging Core Facility
The University of Texas MD Anderson Cancer Centre

Jared: So, I started giving talks about spatial biology with a Lego sphere. I described all of these different colour blocks as different cell types, and in this meeting, I threw this up into the air and smashed it on the stage. I picked up a few of the parts, and I said, tell me who was next to who ,tell me who influenced who, and how they communicated with one another. And who can put it back together in the same sequence that it was in?

That captivated everybody, partly because it was just after lunch, and everybody was a little drowsy, and when it hit that wooden stage, it made the biggest noise, and everybody's head shot up, 'whoa, what just happened?'. But it really drove home the point that these disaggregation technologies, where we're really looking at straight composition, lack all of the spatial information, they lack how our neighbours influence us.

I don't know where you live, but clearly you have neighbours. And you can imagine if you have a loud, noisy neighbour; how that might impact your sleep, which would impact your job performance, which would have this long spiralling effect. Our neighbours influence us. Where we live influences us. Where's your closest place to get food? Where do your kids go to school? All of these things are really relevant.

So where do I hope spatial will be in 15 years? I hope we can figure out how to put this Lego sphere back together and we can functionally understand how this works. We've got to understand all of these networks, the communication outside of the cell, the communications direct cell-to-cell, the neighbour effect, the construction effect. I'm hoping that in 15 years we'll have that data then we can really predict and understand functional neighbourhoods and non-functional neighbourhoods. And that we can develop personalised therapies based on your particular neighbourhood, and not just in general. I'm hoping that, by then, the clinic will have accepted these technologies.

HAIQI CHEN

Assistant Professor
UT Southwestern Medical Centre

Haiqi: I think there are a couple of things that will become very routine by that point.

Firstly, as long as there's not an even more superior technology that will replace spatial technologies all together, I'm assuming that multimodal spatial capture will become routine. You'll be able to capture protein, RNA, DNA and all the other important cellular modalities all at once plus the spatial information too.

Secondly, spatial analysis will go 3D. Currently, analysis of tissue is mostly done on thin tissue slices (2D). One approach is to do spatial analysis of a series of 2D slices and then reconstruct them back into the 3D tissue. Moreover, we can even achieve 4D with a time component in it.

The last thing that I think will become quite common will be spatially resolved functional analysis by combining spatial capture technologies with CRISPR screens, or other functional analyses. This would allow not only the assessment of cell intrinsic effects of a gene perturbation, but also examination of its extracellular effects. That's something that I'm personally really interested in, and my lab is trying to work on something like that.

GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL BIOLOGY LOOK LIKE IN 15 YEARS' TIME?

YANXIANG DENG

Assistant Professor
University of Pennsylvania

Yanxiang: I think we will be able to profile more omics at the same time in the tissue section. I think we will be able to reconstruct spatial omics in 3D. I.e., the reconstruction of three-dimensional positional tissue anatomy. Finally, we need to add the dimension of time to spatial omics, so that we can measure the cellular dynamics and behaviour.

SILVIA DOMCKE

Affiliate Assistant Professor, Dept. of Genome Sciences, **University of Washington**
Associate Director, Head of Human Genomics, **Gordian Biotechnology**

Silvia: It is becoming increasingly feasible to systematically dissect gene regulation in both healthy and diseased states by combining in silico predictions with large scale experimental perturbations and a variety of phenotypic readouts, involving both spatial and temporal components. If the ultimate goal is the ability to intentionally manipulate cell states, e.g., for therapeutic purposes, it will be fascinating to see which single-cell measures besides gene expression will be most informative, and required, along this path.

LUKAS HEUMOS

PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

Lukas: I think that people will still be arguing about the same basic fundamental problems in 15 years. Differential gene expression predates single-cell data, but people still argue about the best tools. I wouldn't be surprised if that's also the case in 15 years. Also, I believe that foundational models will become a significant share of analysis tools. Furthermore, the analysis of single-cell data will be more automated, and we will likely be able to better integrate modalities. Currently, they are still being analysed quite separately, and it's complex to analyse them jointly. Finally, the spatial field will have matured in 15 years, and it will just be another analysis that you routinely conduct, so you will always look at single-cell measurements resolved in space because it makes more sense.

ANNA SCHAAR

PhD Candidate, Fabian Theis Lab
Institute of Computational Biology, Helmholtz Munich

Anna: I agree with Lukas on all three points, especially the last one. Additionally, there will be new assays coming up, we already see this trend now with spatial-ATAC or spatial maps of receptors in single-cell. Some of the analysis issues we are currently facing in spatial biology will be hopefully resolved, like integrating across space or normalizing with respect to a cell's physical location. There are still a lot of open questions in the single-cell field, but I hope at the speed we're currently developing methods and insights, that we get to the core of how cells truly function with respect to their physical environment and depending on various different measures.

MAI CHAN LAU

Assistant Principal Investigator
A*STAR's Bioinformatics Institute (BII) and Singapore Immunology Network (SIgN)

Mai Chan: I hope to see the sensitivity and the resolution of spatial technologies improve by leaps and bounds. I would expect the technology to be more affordable, and the handling of tissues and workflow to be more robust. The handling process can be subjective, and failure of certain techniques could result in wastage of tissue, along with associated costs and efforts. Ultimately, all the resolutions of these issues will converge. For spatial omics, or at least spatial transcriptomics, the resolution is getting higher and higher. Now, it's down to single cell level. Thus, when all these technical and cost issues are addressed, single cell sequencing may not be necessary anymore. I also believe that as researchers begin exploring other areas of spatial omics – like spatial metabolomics, spatial TCR, and ribo-sequencing – the future will be dominated by precision medicine informed by AI-enabled spatial multi-omics.

MOHAMMAD LOTFOLLAHI

Scientist, **Helmholtz Munich/Wellcome Sanger institute**
Director of Machine Learning , **Relation Therapeutics**

Mo: I love this situation with AI and text, where we are for example, with ChatGPT and deep generative models. The power here is in training a model on millions of samples and examples so that it learns. So, the ultimate goal is having these gigantic AI algorithms that we can train on all of these datasets. Can you understand the specialised or global behaviour of all cells across all tissues using images that you can get from spatial, while also considering the effects of drugs? Then you have this gigantic algorithm that you can simply ask 'what is the effect of changing this gene if the cell was in this location?' This algorithm would then generate that for you. Or down the line, given a transcriptome from a patient, you could just give it to this algorithm and the algorithm identifies the cell types that were affected by disease and the potential drugs that you can prescribe for this patient. Combined with patient information, it might facilitate augmented decision making for clinical applications. That's my hope here. And these consortium efforts, such as the Human Cell Atlas are actually a really cool resource for building these type of things. I think this would be the future. I'm super positive.

GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL BIOLOGY LOOK LIKE IN 15 YEARS' TIME?

LUCIANO MARTELOTTO

Associate Professor & Lab Head, Single Cell and Spatial-Omics Laboratory
University of Adelaide, Australia

Luciano: I'm hoping that things like genetic perturbations are going to be much easier. I also anticipate that computational biology will be offered as a standalone subject in universities. This is crucial because, currently, there is a noticeable shortage of computational biologists. While tools in molecular biology are advancing rapidly, integrating computational expertise remains vital.

I expect the amount of transcript per cell to increase along with enhanced multiplexing capabilities. It would be more beneficial to focus on sequencing a broader range of samples rather than delving deeper into individual cells, although both approaches hold their significance.

As for spatial technologies, I believe prices will become more affordable. However, sequencing technologies, given their versatility and the potential for innovation, will likely remain dominant for a considerable period. Imaging technologies have their limitations, especially with challenges like molecular crowding.

Addressing these issues is essential before we can progress further. This isn't to say that sequencing is without its problems, but it does offer a broader range of possibilities. I envision spatial technologies evolving to accommodate more 'omes' from identical sections and allowing multi-dimensional segmentations. Being able to create a 3D reconstruction would be a game-changer.

CATIA MOUTINHO

Founder & Scientific Advisor
The Single-Cell World

Catia: I don't plan more than one year in advance. If you ask me, what will I be doing in five years? No idea. Well, I can tell you what I wish. In 15 years, I would love to put a biological tissue or cells in a machine, with which we could extract all the information - multi omics information and also spatial information in 3D - from the tissue or cells in one go. This is what I would love to see.

GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL BIOLOGY LOOK LIKE IN 15 YEARS' TIME?

LINDA ORZOLEK

Director, Single Cell & Transcriptomics Core
Johns Hopkins University

Linda: I don't know what it's going to look like in two years, let alone 15. I think in 15 years, you'll have a single cell that you'll be able to identify the location of all of the transcripts in the cells. You'll then be able to pull out the transcripts and do full length sequencing on them, not just rely on these probe-based methods. Maybe it's still the naïve undergrad in me, but I think that these technologies are evolving so rapidly that, in 15 years, I struggle to see a limitation. With AI advancing as rapidly as it is, what we can't comprehend, AI may be able to.

I also think we will be consistently working at subcellular resolution. New in situ sequencing methods are already allowing us to target hundreds of genes at once, with claims that thousands are coming in the next year. So, a year or two after that, is 18,000 that much of a problem? I don't think that it should be. I think they will figure out the lasers and the fluorescence etc. to advance it. I think that it'll combine with AI to make inferences that are beyond what we can see and understand. I think that everybody will be able to afford it, and I think it will be much more readily available in a clinical, personalized medicine setting. And I think, and I hope and pray really, that it will be used as a standard for testing and diagnostic treatments for a variety of situations within 15 years.

TANCREDI MASSIMO PENTIMALLI

PhD Student, Nikolaus Rajewsky Lab
Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Centrum (MDC)
Berlin School of Integrative Oncology (BSIO), Charité – Universitätsmedizin Berlin

Tancredi: Single-cell and spatial technologies are developing so rapidly that the field could be radically different already in few years from now. Things that seemed impossible a few years ago, such as transcriptome-wide, single-cell resolved, spatial transcriptomics, are now coming within reach - even for FFPE tissues. Importantly, FFPE compatibility will drastically speed up the translational impact of single-cell and spatial omics technologies, enabling the analysis of routine clinical samples at unprecedented resolution.

The simultaneous profiling of cell types, states and crosstalk along disease trajectories will speed up the identification of mechanism-based, novel therapeutic strategies. With the growing clinical relevance of spatial technologies, approaches to spatially analyse samples in their full thickness will be urgently needed. Furthermore, methods to spatially perturb tissues (e.g., optogenetics) will be required to test specific hypotheses.

GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL BIOLOGY LOOK LIKE IN 15 YEARS' TIME?

MICHA SAM BRICKMAN RAREDON

Research Group Leader, Departments of Anaesthesiology, Pulmonary, Critical Care & Sleep Medicine and Immunobiology, **Yale School of Medicine**

Sam: I think the resolution on spatial is going to get better and better. I do think that's possible. I think that we're going to start to have three-dimensional spatial data generated in the same way that microscopy has shifted in the last decade from largely two-dimensional slices of tissues to whole organ clearing. I suspect we may start to see techniques that are three-dimensional omics technologies where you can see the whole organ, and that's probably very important and necessary, because these organs are 3D. If you're going to study spatial patterns, you need to see what's in and out of plane. That's probably going to be the biggest advance.

ANDREW RUSSELL

Postdoctoral Fellow, Fei Chen Lab
Broad Institute of MIT and Harvard

Andrew: I think about this a lot, actually. If you are into sci-fi, or if you like being optimistic about the future, which I think we have to be at the moment, then you know we can come up with some really amazing ideas. You can be excited about the future that humanity might have with these technologies, both in the healthcare realm and in other areas. One thing that I do think we'll be able to do is to be able to profile 3D tissues. I don't know what format this technology will take, but if we could profile large 3D thick tissue areas, we could take out a region of a tumour or another bit of tissue and we would put it into this technology and we would be able to profile that tissue completely in 3D and reconstruct that. We're not very good at doing it now. We can take thin sections of a tissue, and we can profile them fairly well, but I think that three-dimensional component is largely untouched.

In addition to that, we could get better at understanding small molecules. We are pretty good with macromolecules, proteins and DNA and RNA, but the exciting spaces will be in between that. We will be looking at metabolites, at lipids and at signalling molecules. Closely integrated with that is the question of how we build efficient models of those systems from the data we collect using new technologies to perturb them in silico. And whether we could do that at scale. That would be very fast and cost efficient - it would be amazing. We could take out a tissue biopsy, we could run millions of simulations on how you treat that tissue, and then we would probably, within an hour or two, would then be able to find a course of treatment that would be most effective. Maybe 15 years is too soon, but things can grow exponentially.

GLIMPSING THE FUTURE. WHAT MIGHT SINGLE-CELL AND SPATIAL BIOLOGY LOOK LIKE IN 15 YEARS' TIME?

AZUCENA SALAS

Inflammatory Bowel Disease Group leader

Fundació de Recerca Clínic Barcelona-IDIBAPS, Hospital Clínic Barcelona

Azucena: I think that we are just starting to understand the potential of spatial biology, which I am convinced will bring upon a revolution in our understanding of diseases. Unravelling disease complexity (heterogeneity across patients), which is one of the most important current challenges in medicine, will require information not only on cell states, but on their context and interacting partners in a way that spatial biology can provide. Based on the speed at which the field (both technologies and computational tools) is moving, I am convinced that the future of precision medicine will be spatial. My personal hope is that we see the routine application of these technologies to patient care.

DENIS SCHAPIRO

Research Group Leader

Heidelberg University Hospital

Denis: I believe that all omics will be spatial so I think we will be able to analyze the full proteome, transcriptome, metabolome, translatome, lipidome and more with single cell resolution. Additionally, we will be able to do so in 3D. For translational applications in the next 15 years, I hope that we will have very robust methods that will enable actionable insights for patients. I believe there's a huge potential for something like pathology 2.0 or 3.0, with the combination of digital pathology and deep molecular cellular profiling. I hope my predictions will age well, so that we will be able to discover biomarkers which will translate to better diagnosis and treatments as soon as possible.

EMILY STEPHENSON

Senior Research Associate, Haffina Lab

Newcastle University

Emily: It is a very exciting time for single-cell and spatial technologies, and I think that over the next 15 years we will continue to see improvements to the technologies and the development of new methods. In particular, we will see higher-resolution multimodal maps in which single-cell and spatial technologies will be seamlessly integrated.

Three-dimensional spatial omics technologies will become more refined, enabling the visualisation of intricate cellular interactions within intact tissues. This will provide a deeper understanding of cell-to-cell communication, immune responses, and microenvironmental influences in their natural context.

Finally, I think single-cell and spatial technologies have the potential to be instrumental in tailoring medical treatments to individuals. Cellular heterogeneity data could guide the selection of therapies, optimising their effectiveness while minimizing side effects.

JOVAN TANEVSKI

Research Area Leader, Saez-Rodriguez Group, **Institute for Computational Biomedicine, Heidelberg University and Heidelberg University Hospital**

Jovan: Firstly, I expect these technologies to consolidate and standardize, which would greatly facilitate the downstream analyses. I'm expecting a lot of improvement both in in single-cell and in spatial technologies in terms of reliability and quality of the data. There are a lot of synergies between these technologies but there is still a lot to explore at their intersection. For example, in the direction of reducing costs by efficient and complementary experimental designs.

Secondly, I expect further developments in single-cell and spatial multiomics, opening up new venues for spatially aware integrative analyses and exploration of context specific relationships between the modalities.

Finally, I expect that the quantity and the quality of perturbational and spatiotemporally resolved data that will be available, will allow us to start building comprehensive mechanistic models of tissue biology.

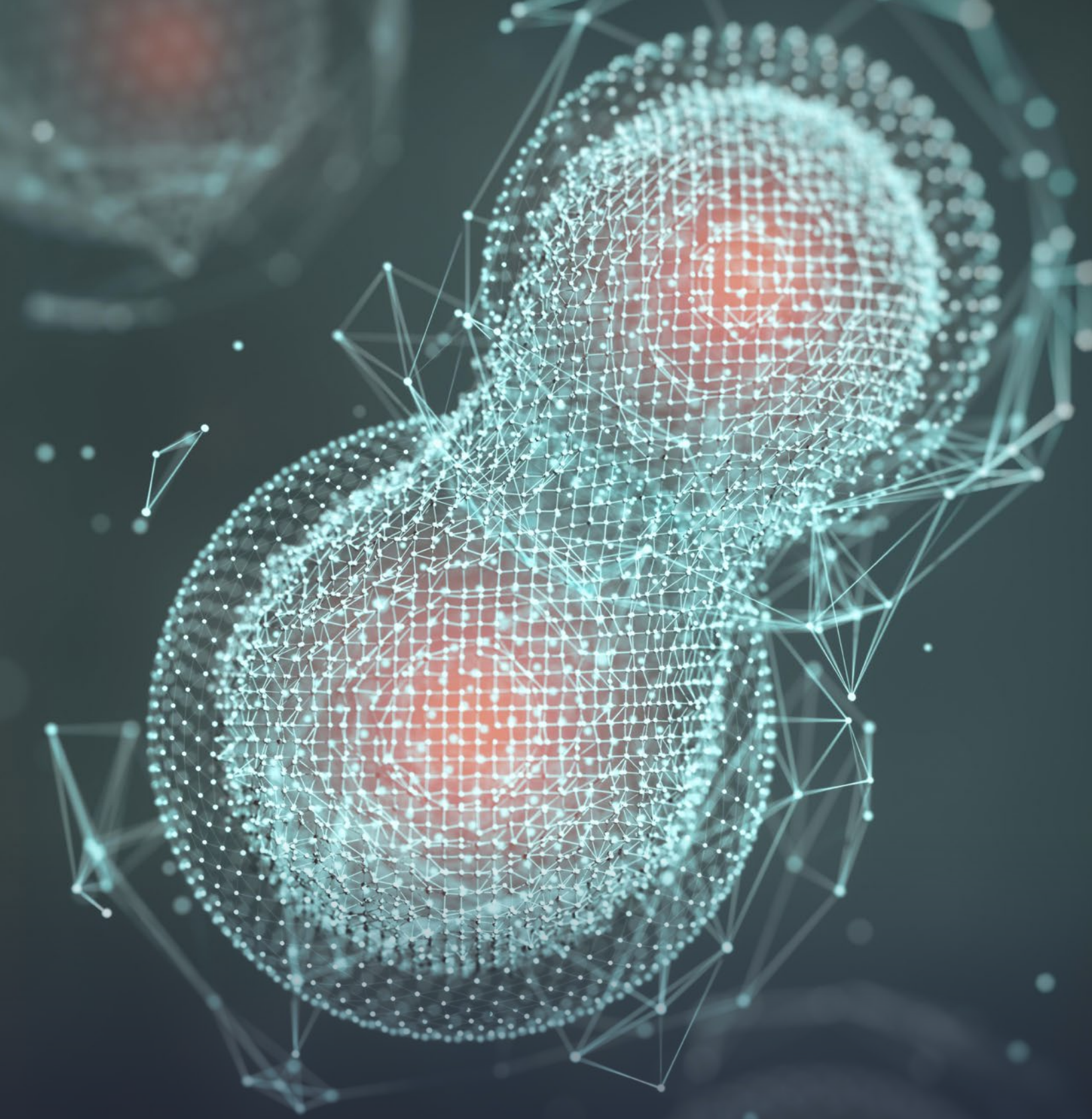
FAN ZHANG

Assistant Professor
University of Colorado School of Medicine

Fan: Single-cell multi-omics approaches are revolutionizing our understanding of disease heterogeneity and pathogenesis. I expect that approaches such as single-cell spatial transcriptomics will help answer the key immunopathological questions in the next 15 years: how different cell phenotypes communicate with each other to create diverse biological niches in diseases. How to bring the spatial and temporal axes together into single-cell biology. How to use single-cell biological findings more efficiently to impact clinical interpretation to further provide insights into disease progression. I expect that single-cell biology approaches could push the field towards these directions with fundamental principles in a multi-disciplinary manner.

Chapter 8 references

1. Wen, L. et al. **Single-cell technologies: From research to application.** The Innovation **3**, 100342 (2022).
2. Polychronidou, M. et al. **Single-cell biology: what does the future hold?** Molecular Systems Biology **19**, e11799 (2023).



Front Line Genomics

frontlinegenomics.com

info@frontlinegenomics.com
@FLGenomics
@frontlinegenomics

Festival of Genomics: festivalofgenomics.com