



Exploring Spatial Transcriptomics

2020-05-28
Alma Andersson



<https://github.com/almaan>



SPATIAL
research
<https://www.spatialresearch.org>

A brief Introduction

- Alma Andersson
- From : Utterbäck, Sweden
 - Population : 69
- Now : Stockholm, Sweden
 - Population : 1,605,030
- 2017-2018 : Delemotte Lab
 - Molecular Dynamics
 - Membrane proteins (Ion Channels)
- 2018-Current : Lundeberg Lab
 - Spatial Transcriptomics (ST)
 - Computational Method Development
- github : almaan
- Small disclaimer : first online teaching experience

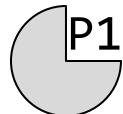


Outline

- ~~Introduction~~
- Notation
- Background
- Data Orientation
- Data Analysis (Overview)
- Break
- Questions
- Cont. Data Analysis (Overview)
- Exercises (Information)
- Questions

Notation

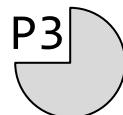
- Exercise session consists of 3 parts
- Symbols below used to indicate when material is included in one of these



- Material in Part 1



- Material in Part 2



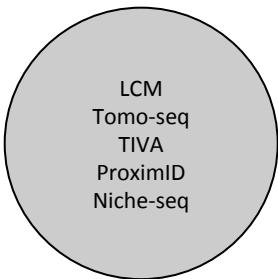
- Material in Part 3

|| ■■ Background ■■||

The spatial space | Overview of techniques

The spatial space | Overview of techniques

Microdissection

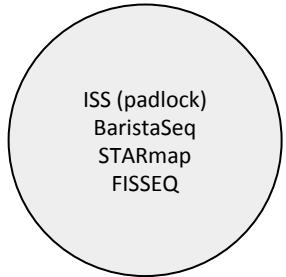


Isolate a region of interest, place isolate in separate well and sequence (either by bulk or single-cell methods).

“Brute Force” approach.

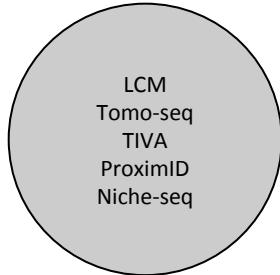
The spatial space

In situ sequencing



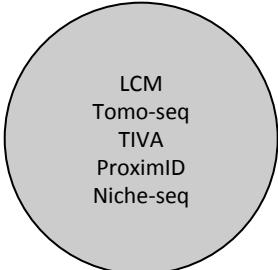
Sequence the transcripts in place. Provides sub-cellular resolution. Tend to rely on gene panels, “*a priori*” defined targets.

Microdissection

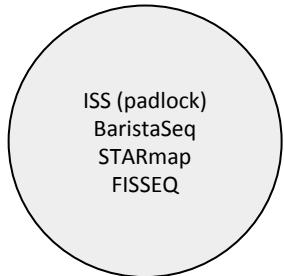


The spatial space

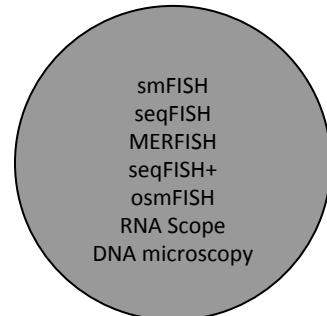
Microdissection



In situ sequencing



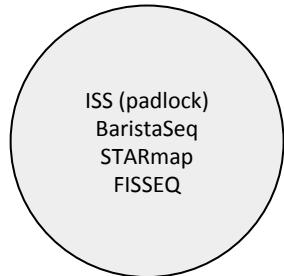
In situ hybridization



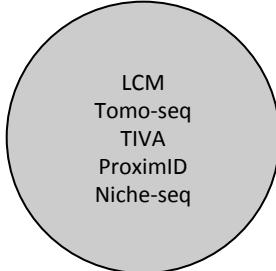
Labeled probes for specific targets, hybridize in place and visualize for spatial information.

The spatial space

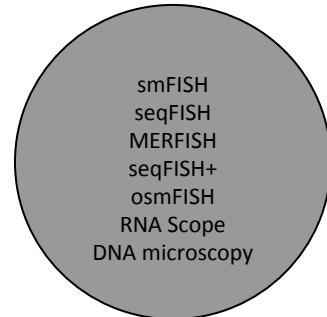
In situ sequencing



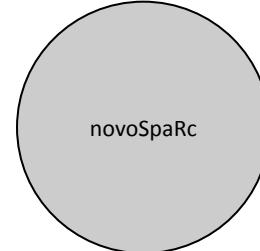
Microdissection



In situ hybridization

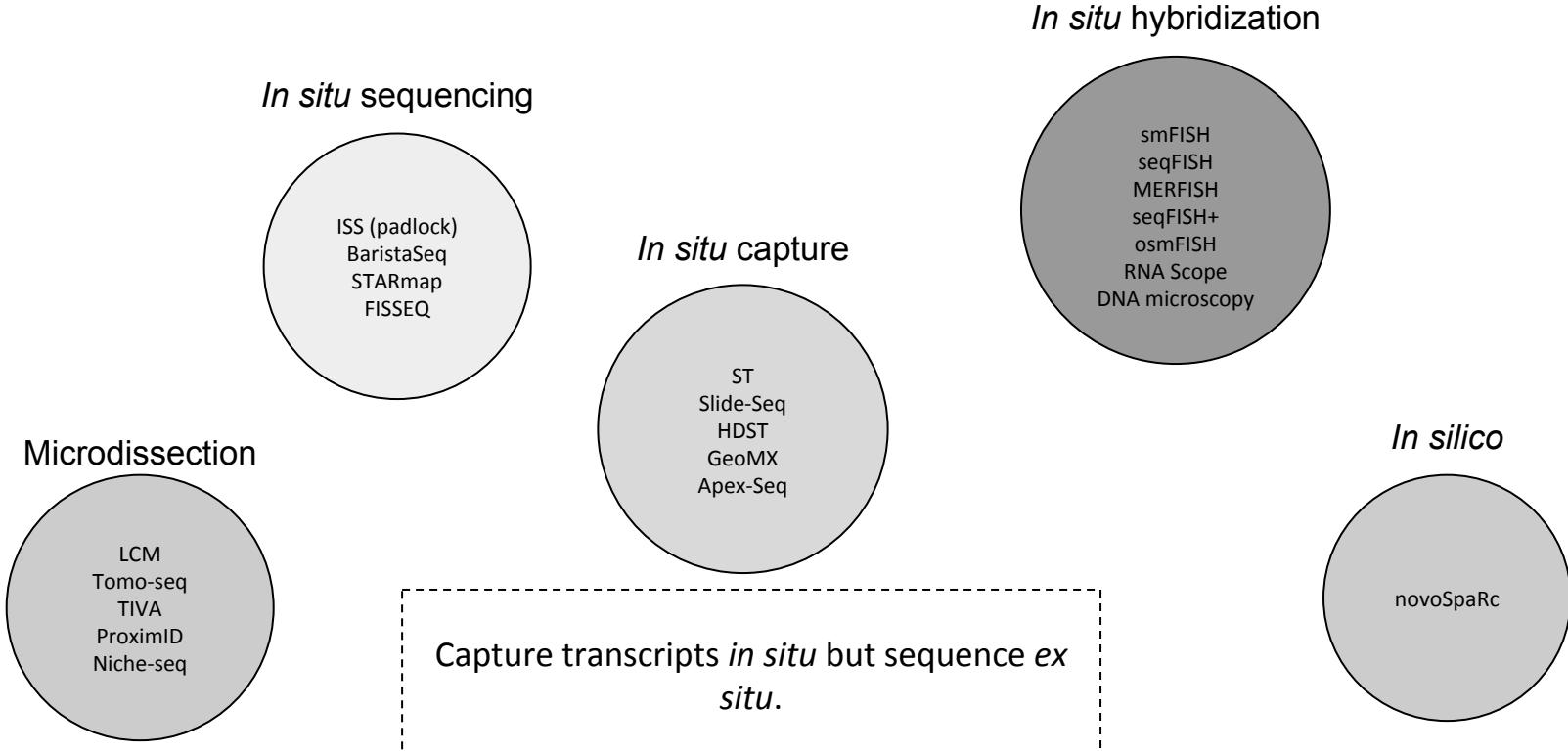


In silico



Infer and reconstruct spatial structure from
single cell data.

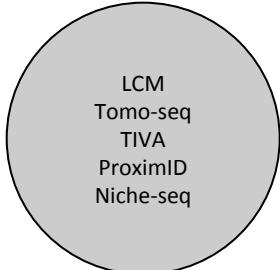
The spatial space



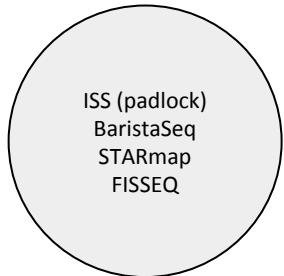
Credit to J.Bergenstråhle and M.Asp for categorization of techniques

The spatial space

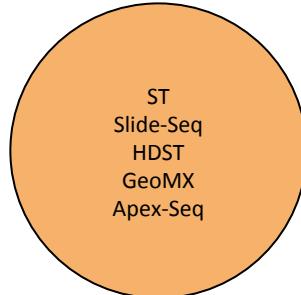
Microdissection



In situ sequencing

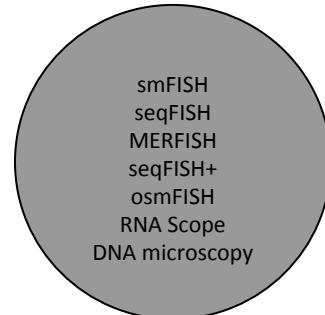


In situ capture

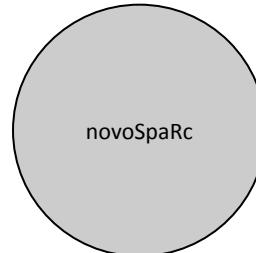


In situ hybridization

In situ hybridization



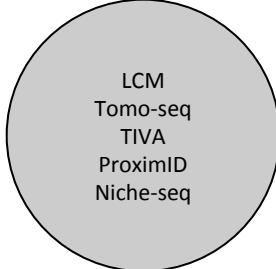
In silico



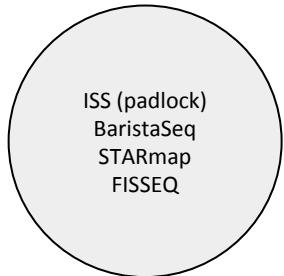
Credit to J.Bergenstråhle and M.Asp for categorization of techniques

The spatial space

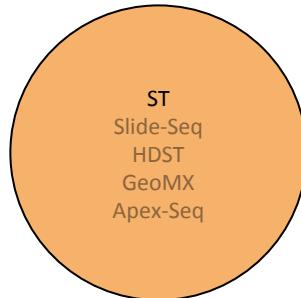
Microdissection



In situ sequencing

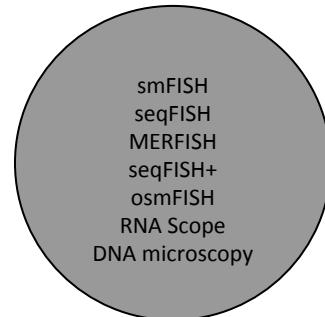


In situ capture

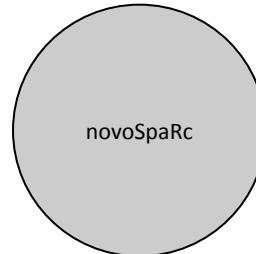


In situ hybridization

In situ hybridization



In silico



Credit to J.Bergenstråhle and M.Asp for categorization of techniques

Spatial Transcriptomics (ST)

Mid 2016

TRANSCRIPTION

Visualization and analysis of gene expression in tissue sections by spatial transcriptomics

Patric L. Ståhl,^{1,2,*} Fredrik Salminen,^{3,7} Sanja Vicković,⁷ Anna Lundmark,^{2,3} José Fernández-Nebro,^{1,3} Jens Magnusson,¹ Stefania Giannella,⁷ Michaela Aspli,² Joakim Lundeberg,^{1,3} Mikael Lindström,¹ Annika Söderström,¹ Simon Cudkowicz,^{4,5} Åke Borgz,⁴ Fredrik Poulton,⁴ Paul Igiv Coates,⁴ Petter Salminen,² Jan Muster,² Olaf Bergmann,¹ Joakim Lundeberg,^{1,2} Jonas Frisell,²

An analysis of the distribution of proteins or messenger RNAs (mRNAs) in *Neurogranin* (Ngn) sections is a key task in the study of gene expression regulation. This usually involves the quantification of a few proteins or expressed genes at a time. In this work, we developed a strategy, which we call "spatial transcriptomics," that allows visualization and quantitative analysis of the transcriptome with spatial resolution. We used a combination of two-dimensional positional barcoding and multiplexed reverse sequencing primers with unique political barcodes, we demonstrate high-quality RNA sequencing data with maintained two-dimensional positional information. The mouse brain tissue sections were processed with a modified version of the standard quantitative gene expression and visualization of the distribution of mRNAs within tissue sections and enable novel types of bioinformatics analyses, valuable in research and diagnostics.

In tissue sequencing (RNA-seq) (*i*) of homogenized samples, the spatial context of gene expression is lost due to averaging transcriptome and loss of spatial information. The positional context of gene expression is of key importance to understand-

ing tissue functionality and a-

Several strategies have rec-

ognized the need to maintain the number of transcripts data

relative to the total amount of RNA

in the standard research setting of regular histological sections (*ii*). However, it is very difficult to introduce positional molecular information into the standard workflow within the context of an RNA-seq. We first analyzed the distribution of mRNA (mRNA) in tissue sections o-

mobilized reverse transcriptase and sequencing (RT-Seq) (*iii*). We found that adult mouse olfactory bulb tissue sections provided a good expression reference data. Tissue sections were fixed, permeabilized, stained, and imaged (Fig. 1A).

The tissue sections were then treated with fluorescent reagents on top of fluorescently labeled nucleic acids

and then extracted cDNA (Fig. 1A, and fig. S1). The tissue

was then enzymatically removed, which left cDNA coupled to the service oligonucleotides on the slide (Fig. 1B). After fragmentation, the cDNA was then corresponding to the tissue structure revealed by the general histology (Fig. 1B and C), and the cDNA was sheared and sequenced (Fig. 1C, and Fig. 2A, B, C, D to G). By comparing the hematoxylin-and-eosin and fluorescence signals, we could measure the average distance between the center of a cell nucleus of a cell to 1.7 ± 2 μm (mean ± SD) (Fig. S1, E to H).

We then used a similar approach to measure mRNA in tissue sections with minimal diffusion and maintained positional representation methodology to obtain a high-resolution spatial barcode (Fig. 2A), and we denoted this strategy "spatial transcriptomics." We deposited ~20 million

sequencing reads per cell nucleus (with a diameter of 100 μm) and a center-to-center distance of 200 μm over an area of 0.1 mm by 0.1 mm (Fig. S2).

We generated sequencing libraries based on mRNA extracted from tissue sections (Fig. 2A and B) (G). Comparison with data from RNA extracted and fragmented in solution revealed a high correlation between the two methods. The method was also found with the other (Fig. S3).

The correlation between the surface and *in-situ* RNA sequencing data was found to be high (Fig. S4).

*These authors contributed equally to this work. These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden; ²SciLifeLab, KTH Royal Institute of Technology, Division of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden; ³Department of Dental Medicine, Division of Periodontology, Karolinska Institutet, Stockholm, Sweden; ⁴Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden; ⁵Department of Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden; ⁶Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden; ⁷Swedish Institute of Otolaryngology and Pathology, Department of Otolaryngology, Lund University, Lund, Sweden; ⁸Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; ⁹Science for Life Laboratory, Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

¹⁰These authors contributed equally to this work. ¹¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

Science Publication

Ståhl et.al

DOI: 10.1126/science.aaf2403

synthesized cDNA (Fig. 1A, and fig. S1). The tissue

was then enzymatically removed, which left cDNA coupled to the service oligonucleotides on the slide (Fig. 1B). After fragmentation, the cDNA was then corresponding to the tissue structure revealed by the general histology (Fig. 1B and C), and the cDNA was sheared and sequenced (Fig. 1C, and Fig. 2A, B, C, D to G). By comparing the hematoxylin-and-eosin and fluorescence signals, we could measure the average distance between the center of a cell nucleus of a cell to 1.7 ± 2 μm (mean ± SD) (Fig. S1, E to H).

We then used a similar approach to measure mRNA in tissue sections with minimal diffusion and maintained positional representation methodology to obtain a high-resolution spatial barcode (Fig. 2A), and we denoted this strategy "spatial transcriptomics." We deposited ~20 million sequencing reads per cell nucleus (with a diameter of 100 μm) and a center-to-center distance of 200 μm over an area of 0.1 mm by 0.1 mm (Fig. S2).

We generated sequencing libraries based on mRNA extracted from tissue sections (Fig. 2A and B) (G). Comparison with data from RNA extracted and fragmented in solution revealed a high correlation between the two methods. The method was also found with the other (Fig. S3).

The correlation between the surface and *in-situ* RNA sequencing data was found to be high (Fig. S4).

*These authors contributed equally to this work. These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

²⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

³⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁴⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁵⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁶⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁷⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁸⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

⁹⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁰⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹¹⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹²⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹³⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁴⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵⁰These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵¹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵²These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵³These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵⁴These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵⁵These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵⁶These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵⁷These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵⁸These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se

¹⁵⁹These authors contributed equally to this work. Corresponding author Email: jonasfrisell@kth.se</

Spatial Transcriptomics (ST) → Visium

Mid 2016

TRANSCRIPTION

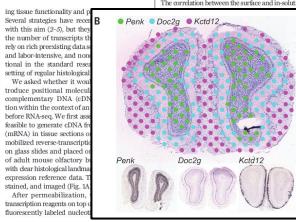
Visualization and analysis of gene expression in tissue sections by spatial transcriptomics

Patrik L. Ståhl,^{1,2,*} Fredrik Salhün,^{2,*} Sanja Vicković,^{2,*} Anna Lundmark,^{2,3,4} José Fernández-Nebro,^{1,5} Jens Magnusson,¹ Stefania Giannouli,² Michaela Aspli,² Joakim Lundeberg,^{1,6} Mikael Lindström,¹ Åsa Berg,² Anders Sandelin,² Simon Codeluppi,^{2,7} Åke Borg,² Fredrik Pontén,² Paul Igivu Costa,² Petter Salhün,² Jan Muster,² Olaf Bergmann,² Joakim Lundeberg,² Jonas Frisell,²

Analytic of the distribution of proteins and messenger RNAs (mRNAs) in *Nestin*-positive brain sections is a key tool in the study of tissue regeneration and disease. This study introduces the visualization of few proteins or expressed genes at a time, and have defined a strategy, which we call "spatial transcriptomics," that allows visualization and quantitative analysis of the transcriptome with single-cell resolution. By using a combination of spatial transcriptomics and 10x Genomics' reverse transcription primers with unique political barcodes, we demonstrate high-quality RNA sequencing data with maintained two-dimensional positional information from the mouse brain. We also show that the same approach can be used to analyze gene expression data and visualization of the distribution of mRNAs within tissue sections and enables novel types of bioinformatics analyses, valuable in research and diagnostics.

Tissue transcriptomes are typically studied by RNA sequencing (RNA-seq) (*i*) of homogenized samples. This approach loses averaged transcriptome and loss of spatial information. The positional context of gene expression is of key importance to understand-

ing tissue functionality and a

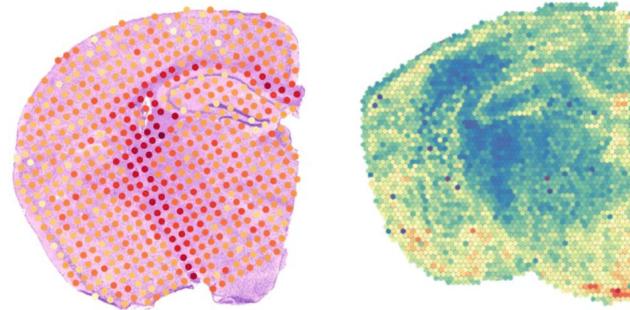


spatialized cDNA (Fig. 1A, and fig. S1). The tissue was then enzymatically removed, which left cDNA coupled to the service oligonucleotides on the slide (Fig. 1B). After this, the tissue was stained with antibodies corresponding to the tissue structure revealed by the general histology (Fig. 1B and C), and the cDNA was sheared and fragmented (Fig. 1C, and fig. S1, D to G). By comparing the hematoxylin-and-eosin and fluorescence signals, we could measure the average distance between the center of each cell to 1.7 ± 2 μm (mean ± SD) (fig. S1, E to H).

We then used the same approach to measure mRNA in tissue sections with minimal diffusion and maintained positional representation notwithstanding the presence of hematoxylin and barcodes (Fig. 2A), and we denoted this strategy "spatial transcriptomics." We deposited ~20 million mRNA molecules in a tissue section with a diameter of 100 μm and a center-to-center distance of 0.200 μm over an area of 0.13 cm² (fig. S2). To validate our approach, we extracted total cellular DNA, we generated sequencing libraries based on mRNA and DNA in the same tissue section (Fig. 2A and B) (G). Comparison with data from RNA extracted and fragmented in solution revealed a high correlation between the two methods. The correlation was also found with the other (fig. S2C). The correlation between the surface and in-situ

Late 2018
10x Genomics™
acquisition

Late 2019



Launch of **Visium** Spatial Gene Expression Platform

Science Publication
Ståhl et.al

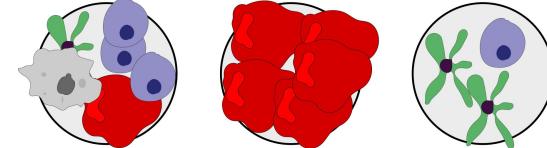
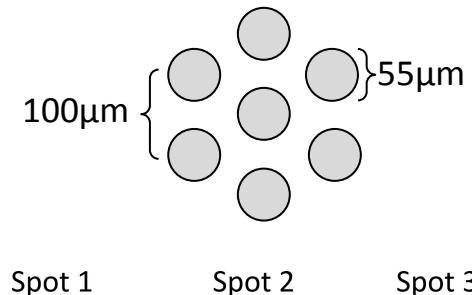
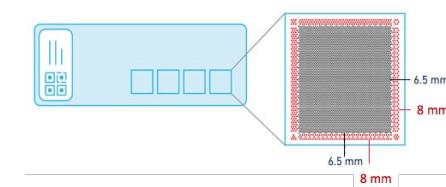
DOI: 10.1126/science.aaf2403

SciLifeLab

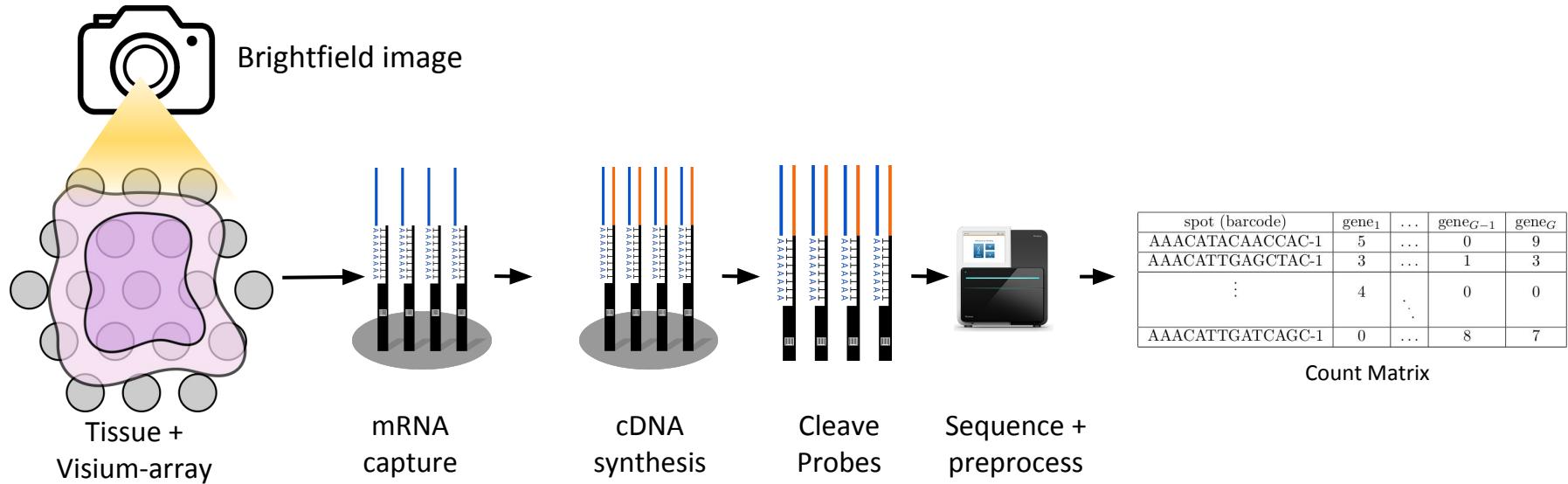


Visium Platform

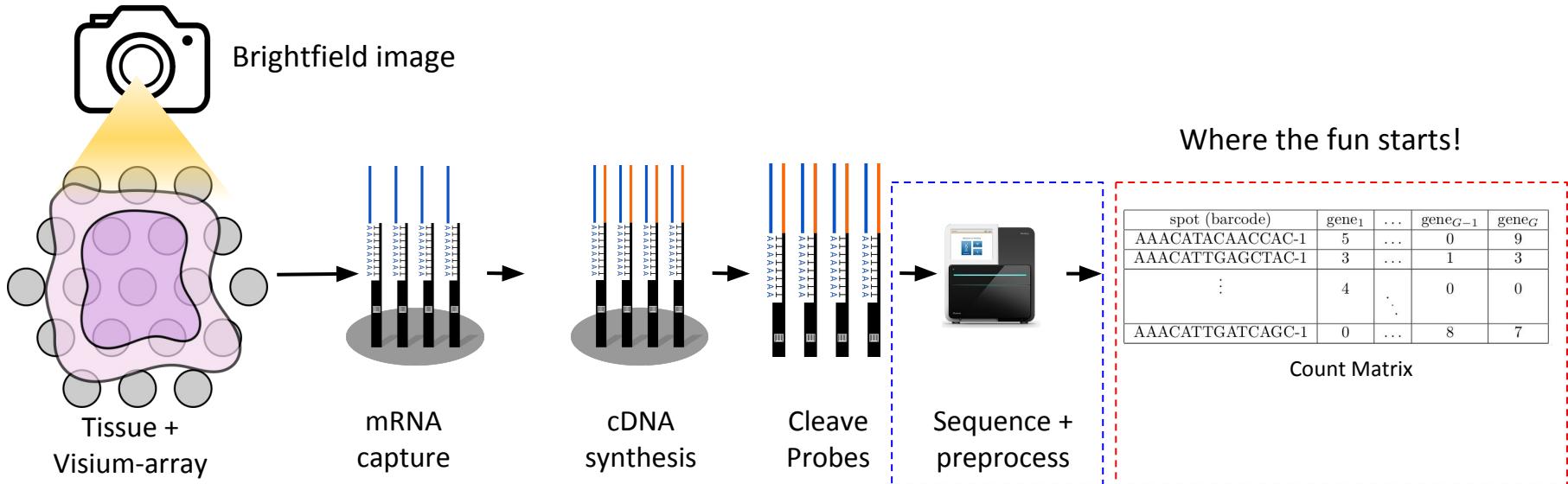
- Array based technique
- 6.5mm x 6.5mm area to put sample on
- 4992 spots arranged in regular grid
- Spot specs:
 - Spot diameter : 55 μm
 - Center to center distance : 100 μm
- Each spot has millions of capture probes
 - spatial barcode
 - polyT sequence
 - captures polyadenylated mRNA
 - Full transcriptome(-ish)
- ~ 1-10 cells contribute to each spot
 - **NOTE :** Not single cell resolution!



The experimental workflow (in a nutshell)



The experimental workflow (in a nutshell)





|| ■ ■ Data Orientation ■ ■ ||

After sequencing (brief)

- **spaceranger mkfastq** | BCL files to FASTQ
- **spaceranger count** | tissue detection/alignment, UMI counting



```
-bash-4.2$ tree -L 2
.
├── analysis
│   ├── clustering
│   ├── diffexp
│   ├── pca
│   ├── tsne
│   └── umap
└── cloupe.clope
    ├── filtered_feature_bc_matrix
    │   ├── barcodes.tsv.gz
    │   ├── features.tsv.gz
    │   └── matrix.mtx.gz
    ├── filtered_feature_bc_matrix.h5
    ├── metrics_summary.csv
    ├── molecule_info.h5
    ├── possorted_genome_bam.bam
    ├── possorted_genome_bam.bam.bai
    ├── raw_feature_bc_matrix
    │   ├── barcodes.tsv.gz
    │   ├── features.tsv.gz
    │   └── matrix.mtx.gz
    └── raw_feature_bc_matrix.h5
    └── spatial
        ├── aligned_fiducials.jpg
        ├── detected_tissue_image.jpg
        ├── scaleFactors.json.json
        ├── tissue_hires_image.png
        ├── tissue_lowres_image.png
        └── tissue_positions_list.csv
    └── web_summary.html
```

Automatically generated analysis

filtered = spots under tissue

raw = all spots

For mapping between image coordinates

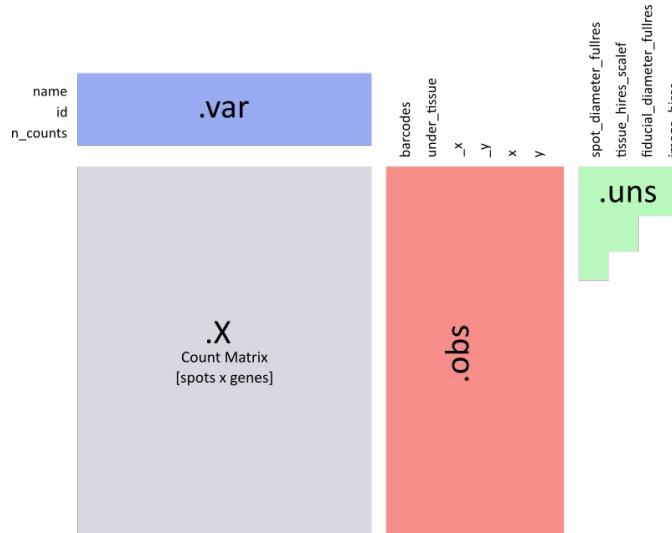
Resized images

Maps barcode to coordinate

Example of **spaceranger count** output

Processed data

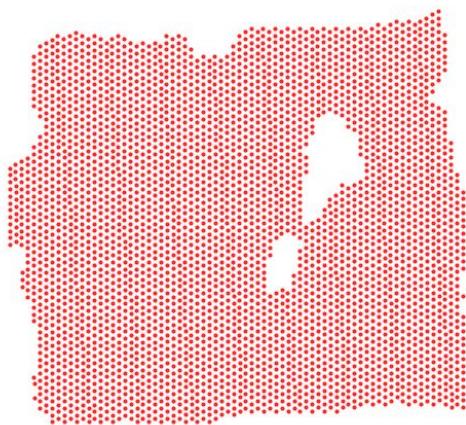
- Either use .mtx files or .h5 files to assemble a data object to work with
 - No standardized format
- Personal preference : convert to .h5ad file (will be using in exercises)
 - scanpy/anndata teams working on - soon to release - their own (similar) format



An initial assessment

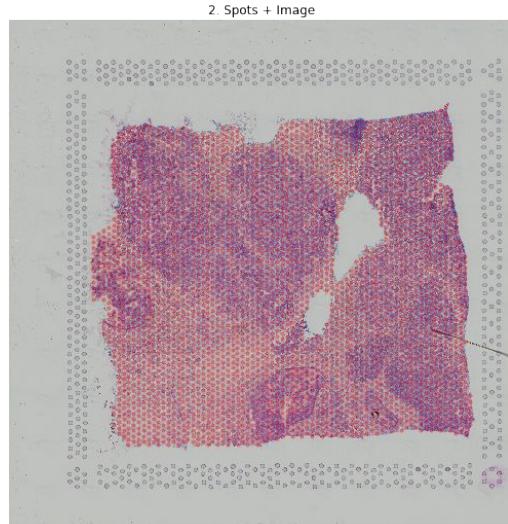
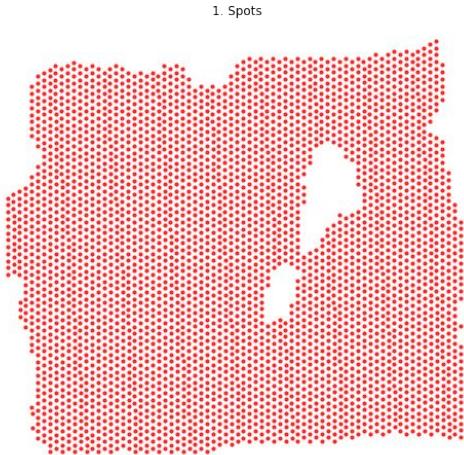
- Example with Human Breast cancer data
 - Public data : Available at 10x website

1. Spots



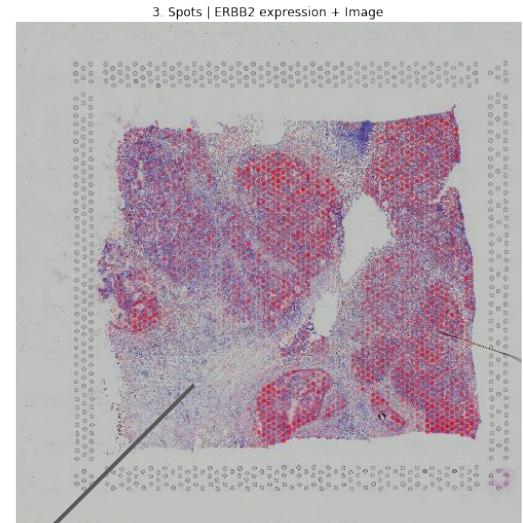
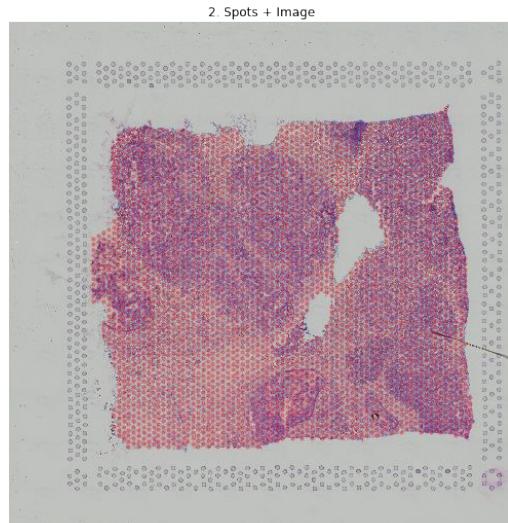
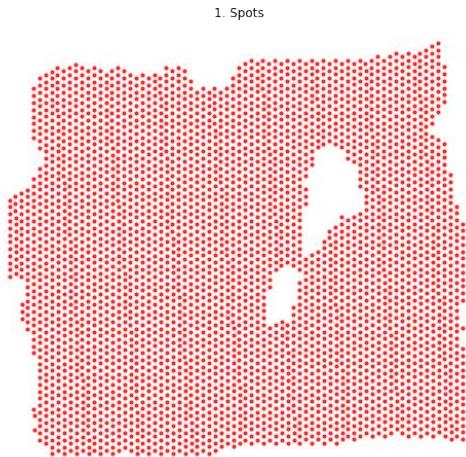
An initial assessment

- Example with Human Breast cancer data
 - Public data : Available at 10x website



An initial assessment

- Example with Human Breast cancer data
 - Public data : Available at 10x website



Facecolor intensity proportional
to gene expression value

■ ■ | Visualizing high dimensional spatial data

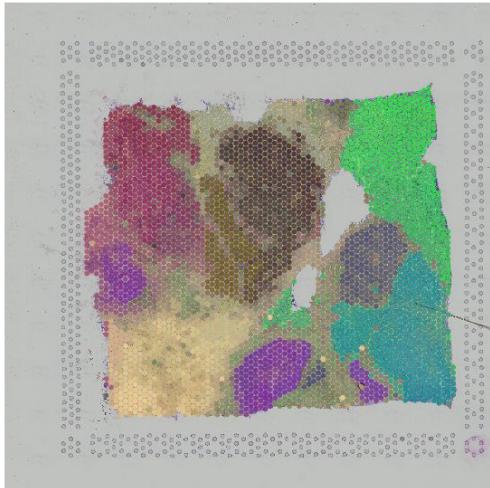
- We visualized one feature/gene (*ERBB2*)
- How do we handle multiple non-mutually exclusive features?

■ ■ | Visualizing high dimensional spatial data

- We visualized one feature/gene (*ERBB2*)
- How do we handle multiple non-mutually exclusive features?
- One idea :
 - Embed gene expression data in 3 dimensional space (e.g. using UMAP)
 - Do affine transformation to unit cube
 - Consider values as RGB values (or other colorspace) and color spots accordingly

■ ■ ■ Visualizing high dimensional spatial data

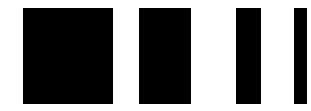
- We visualized one feature/gene (*ERBB2*)
- How do we handle multiple non-mutually exclusive features?
- One idea :
 - Embed gene expression data in 3 dimensional space (e.g. using UMAP)
 - Do affine transformation to unit cube
 - Consider values as RGB values (or other colorspace) and color spots accordingly



Regions with similar colors have similar gene expression.



Data Analysis



■ ■ | Before the analysis | Filtering, Normalization, Batch correction, etc.

- No magic recipe to give
 - How to process your data is very much dependent on the samples and objective
 - Much can be learnt from analysis of single cell data
 - Will give some general advice

■ ■ | Before the analysis | Filtering, Normalization, Batch correction, etc.

- No magic recipe to give
 - How to process your data is very much dependent on the samples and objective
 - Much can be learnt from analysis of single cell data
 - Will give some general advice
- Consider filtering :
 - Genes based on expression levels (total expression > thrs)
 - Genes based on spot presence (#spots gene is observed at > thrs)
 - Spots based on expression levels (total gene expression at spot > thrs) †
 - Ribosomal (RP) and mitochondrial genes (MT) tend to exhibit spurious expression patterns. Exclusion of these is common.

† Only necessary if you expect “defunct” spots under the tissue

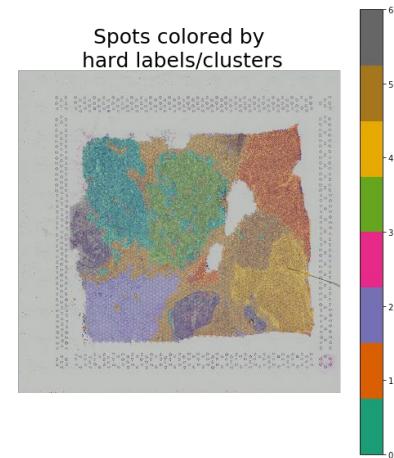
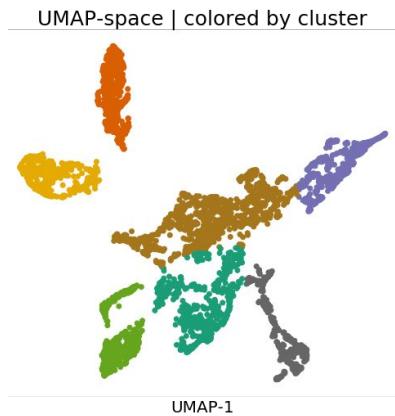
■ ■ | Before the analysis | Filtering, Normalization, Batch correction, etc.

- No magic recipe to give
 - How to process your data is very much dependent on the samples and objective
 - Much can be learnt from analysis of single cell data
 - Will give some general advice
- Consider filtering :
 - Genes based on expression levels (total expression > thrs)
 - Genes based on spot presence (#spots gene is observed at > thrs)
 - Spots based on expression levels (total gene expression at spot > thrs) †
 - Ribosomal (RP) and mitochondrial genes (MT) tend to exhibit spurious expression patterns. Exclusion of these is common
- Normalization / batch correction :
 - Recommend to account for spot “library size” - varying cell density
 - Include slide/array as covariate (sometimes big variation is observed)
 - Popular tools for batch correction : sctransform and Harmony

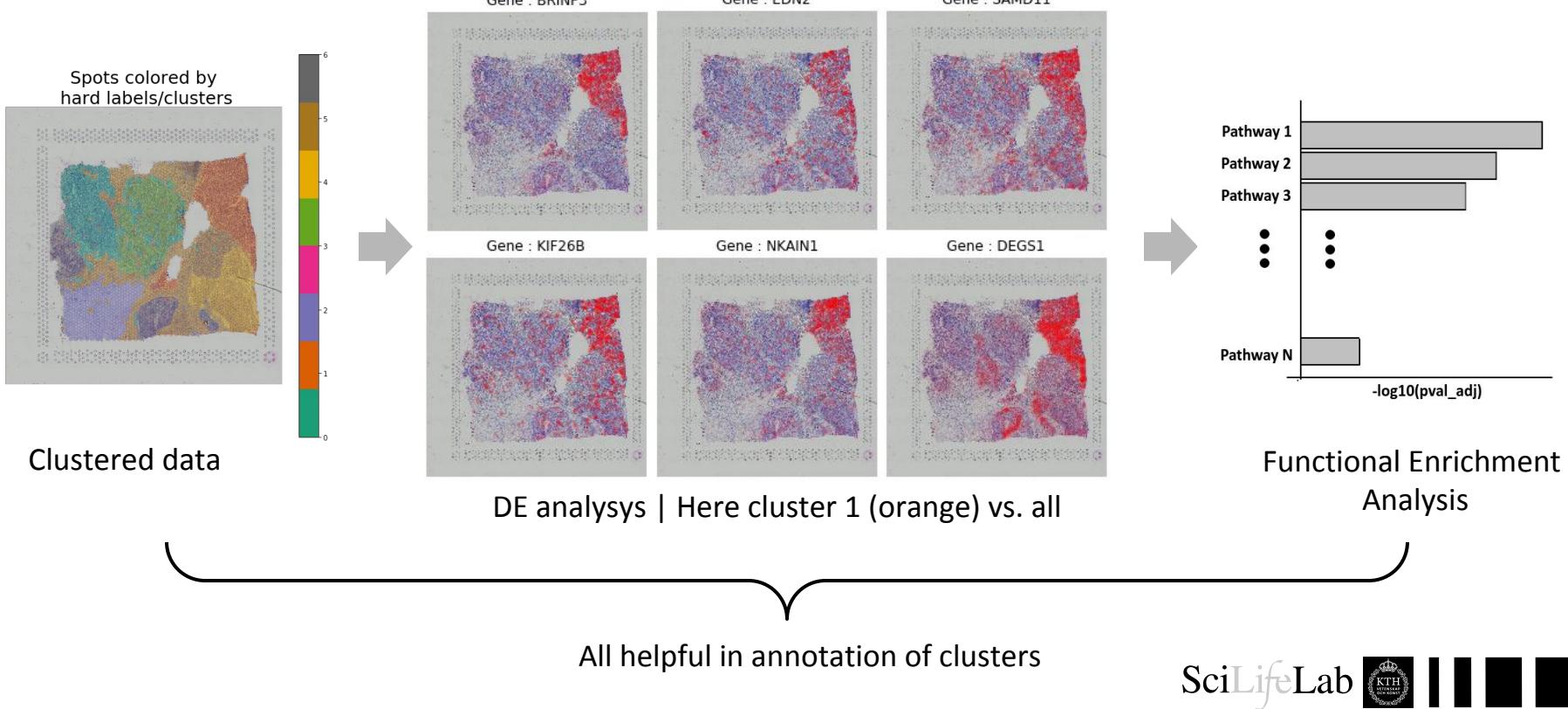
† Only necessary if you expect “defunct” spots under the tissue

Example : Basic Analysis

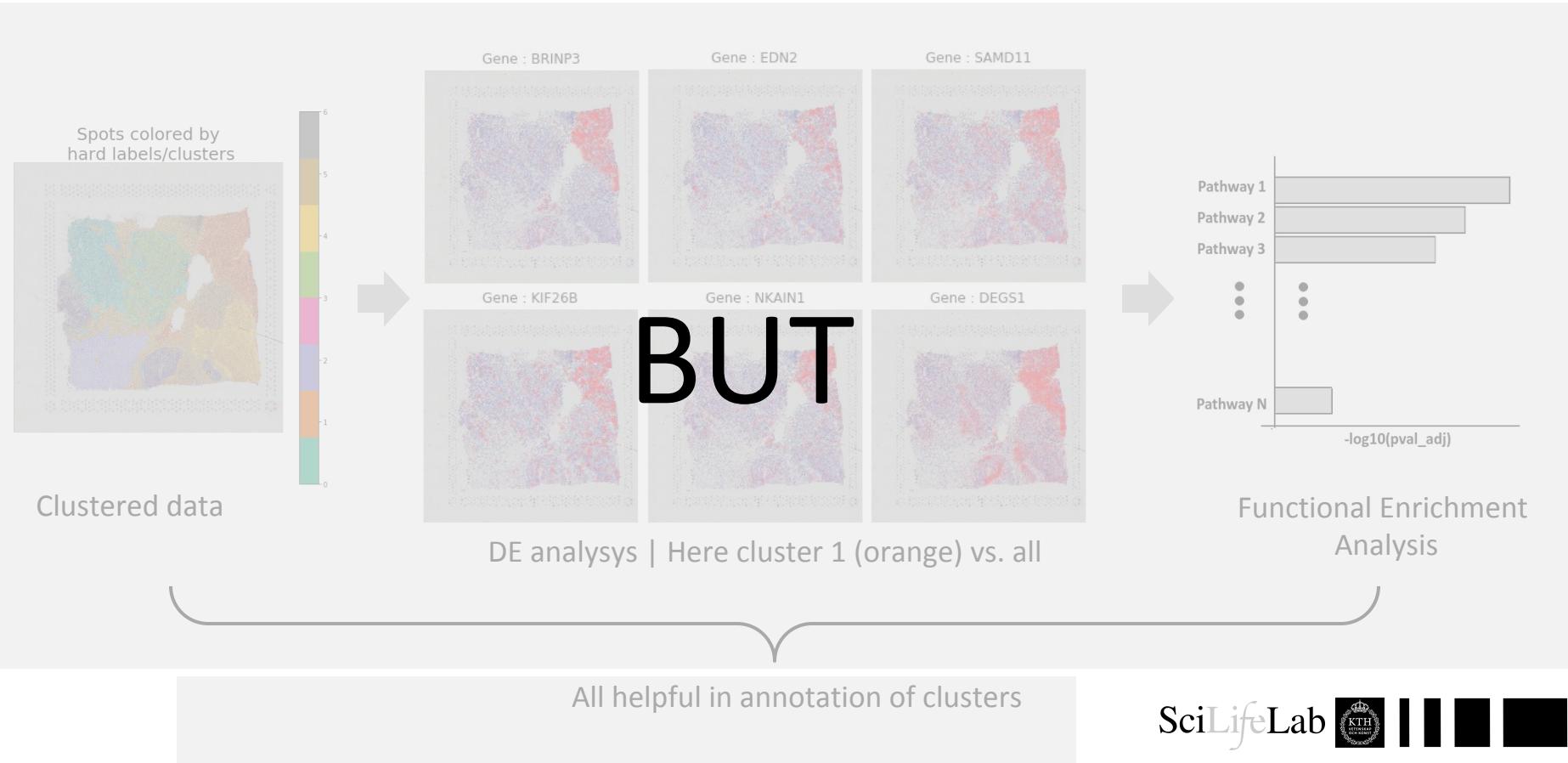
- Cluster the spots based on gene expression
 - Normalize → PCA_{n=20} → UMAP → GMM
- Backmap clusters onto tissue
- Use HE-image as reference
 - Sanity check - does it make sense?
 - Valuable resource
- Next, annotate clusters
 - Find genes associated with cluster
 - Functional enrichment analysis



Example : Basic Analysis



Example : Basic Analysis



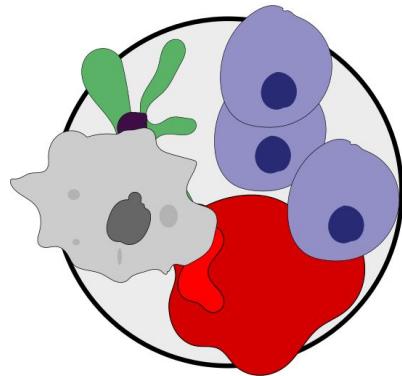


Example : Basic Analysis

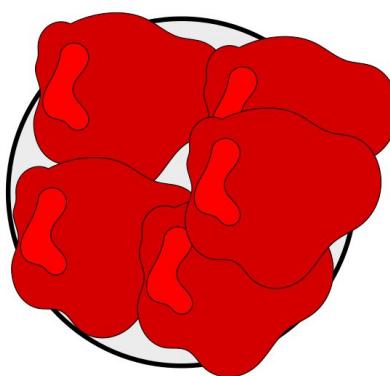
Remember :

“each spot is a mixture of multiple cells,
i.e., one spot may contain multiple cell types”

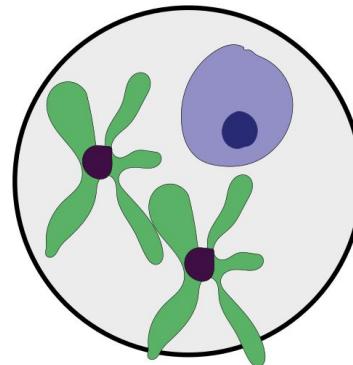
Spot 1



Spot 2

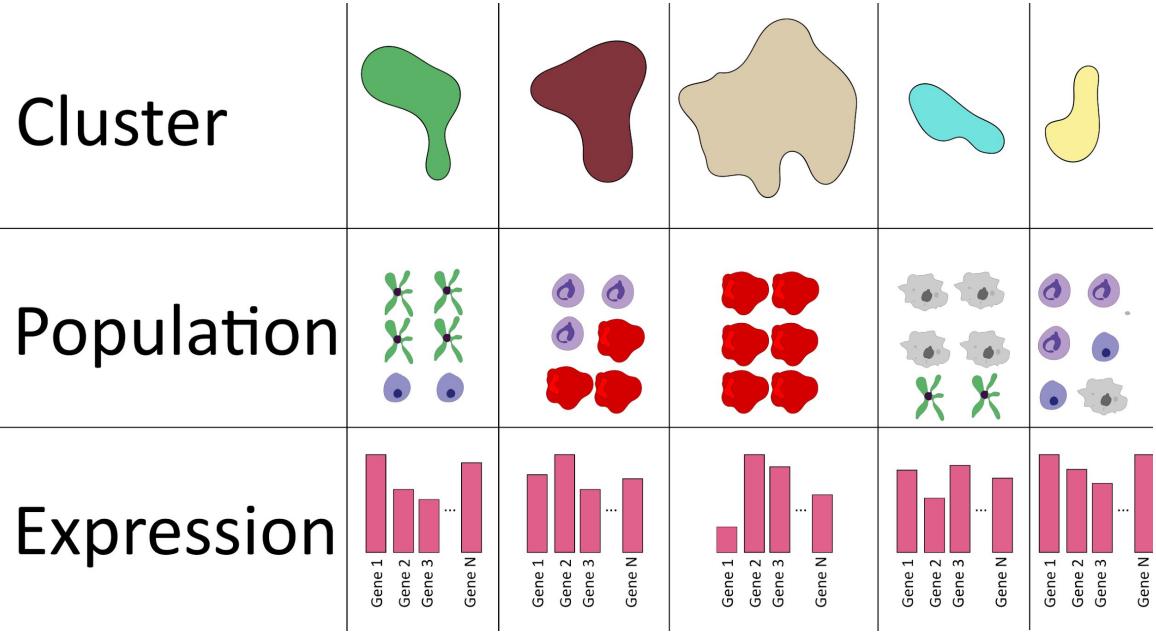
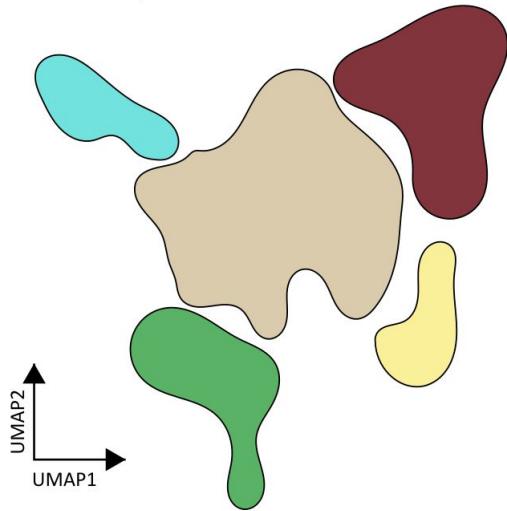


Spot 3



████████Cluster ≠ cell type

Clustered Spatial Gene Expression Data



■ ■ | So where are my cell types?

■ ■ || Break

- 10 min Break
- During :
 - Read through questions (for me)
- After :
 - Discuss (some) questions when everyone is back
 - Data Analysis Cont.
 - Information about the exercises
 - One more question session

(()) C []

■ ■ || Questions

■ ■ | So where are my cell types?

■ ■ | So where are my cell types?

- Marker genes?

So where are my cell types?

- Marker genes? Easy and straightforward, but
 - Requires known marker genes
 - Risk for overlap among marker genes
 - How do we interpret expression values?
 - Lowly expressed markers genes may not always be observed

So where are my cell types?

- Marker genes? Easy and straightforward, but
 - Requires known marker genes
 - Risk for overlap among marker genes
 - How do we interpret expression values?
 - Lowly expressed markers genes may not always be observed
- Alternative solution - Integrate single cell (SC) and spatial data!

So where are my cell types?

- Marker genes? Easy and straightforward, but
 - Requires known marker genes
 - Risk for overlap among marker genes
 - How do we interpret expression values?
 - Lowly expressed markers genes may not always be observed
- Alternative solution - Integrate single cell (SC) and spatial data!
 - Extract information of cell types from SC data and apply to spatial data
 - **Big challenge : deconvolution required (on Visium data)**
 - Multiple approaches have emerged, (e.g., MIA by Itai and Seurat's module)



Integration of Single Cell and Spatial Data

Single Cell Data
Clusters of cell types

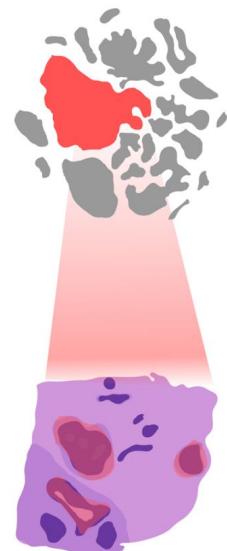


Integration

Spatial Data



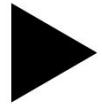
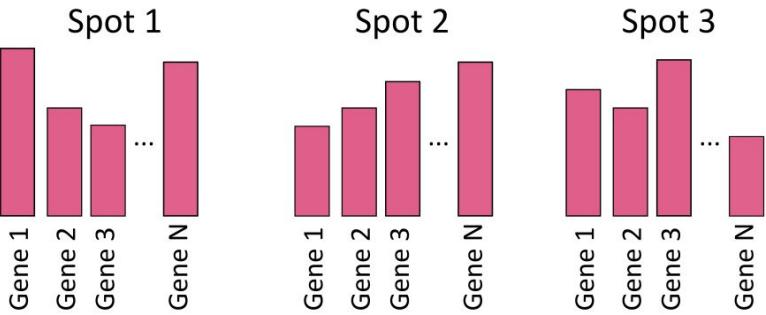
Spatially map each cell type



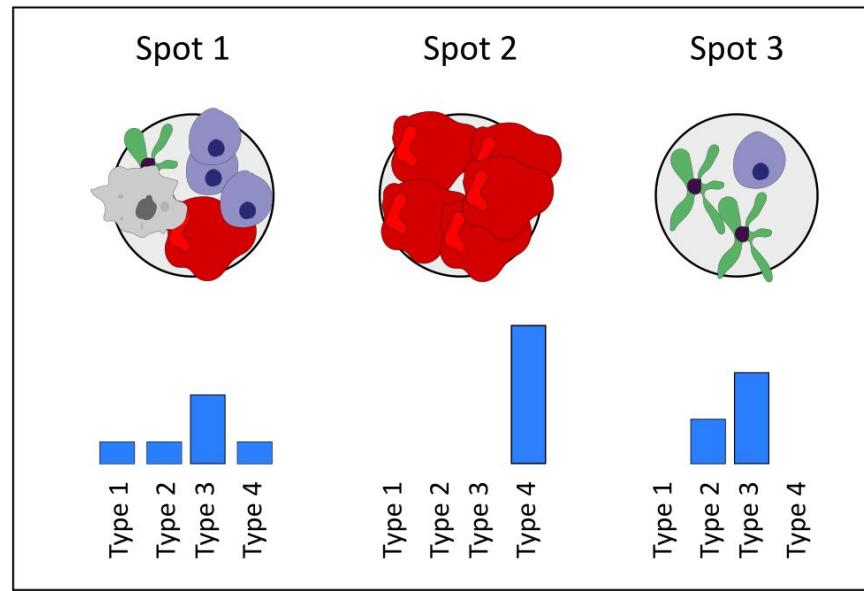


Our objective : deconvolve expression data

From this



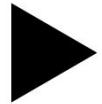
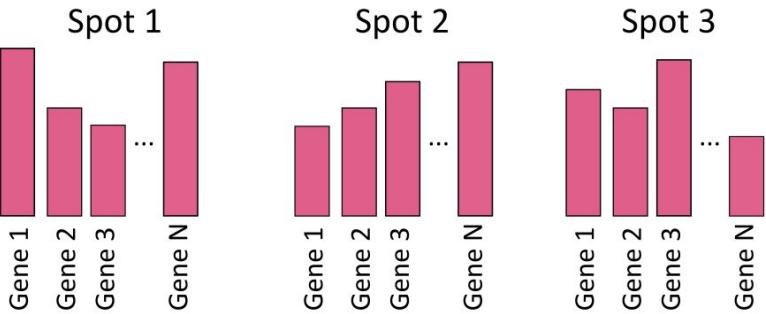
We want this



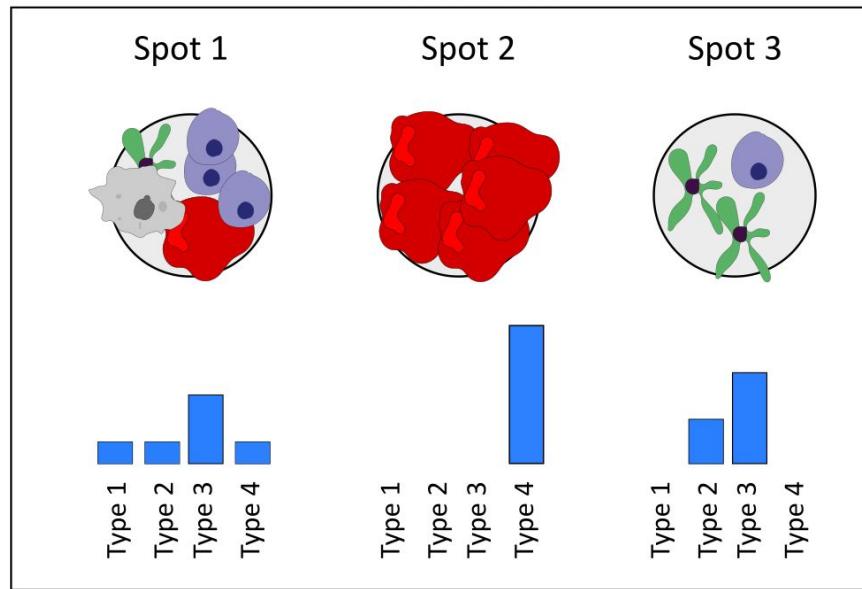


Our objective : deconvolve expression data

From this



We want this

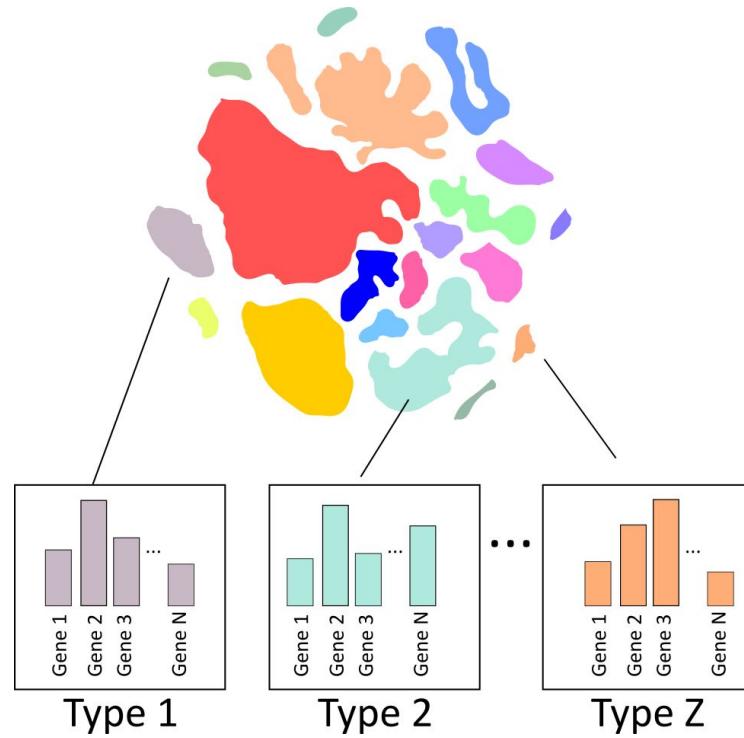


Suggested approach : Model-based Probabilistic Inference



It's as easy as 1-2-3

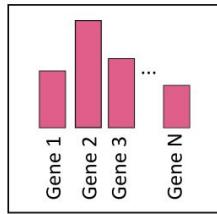
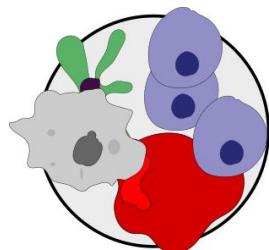
1. Infer cell type expression parameters from SC data





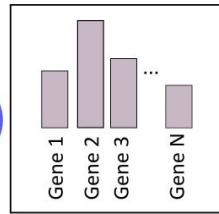
It's as easy as 1-2-3

2. Use inferred parameters to find optimal combination **combination** of cell types in spot



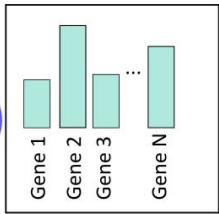
=

W_1



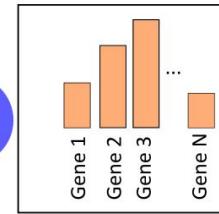
+

W_2



+ ... +

W_z



Spot 1

Type 1

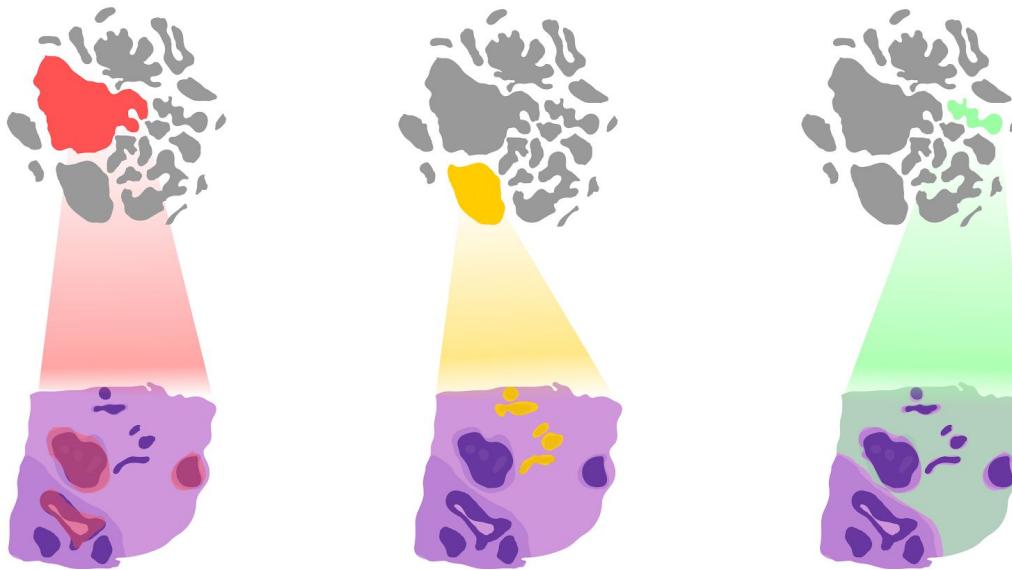
Type 2

Type Z



It's as easy as 1-2-3

3. Map cell type proportions back onto the tissue



■ ■ | Slightly more complex than adding bar graphs..

- Probabilistic Model



Slightly more complex than adding bar graphs..

- Probabilistic Model

Observed counts for gene g at spot s

$$x_{gs} = \sum_{c \in C_s} x_{gsc},$$

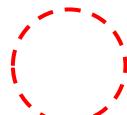
$$x_{gsc} \sim NB(\beta_g \alpha_s r_{gz_c}, p_g)$$

Gene Bias

Scaling factor

Cell type
specific Rate

Success
probability



= Learn from Single Cell Data

$$y_{gc} \sim \mathcal{NB}(s_c r_{gz_c}, p_g)$$

Slightly more complex than adding bar graphs..

- Probabilistic Model

$$x_{gs} \sim NB\left(\alpha_s \sum_{c \in C_s} \beta_g r_{gz_c}, p_g\right)$$



Change Index of summation

$$x_{gs} \sim NB\left(\alpha_s \sum_{z \in Z} \beta_g n_{sz} r_{gz}, p_g\right)$$



Number of cells from type z at spot s

Slightly more complex than adding bar graphs..

- Probabilistic Model

$$x_{gs} \sim NB\left(\alpha_s \sum_{z \in Z} \beta_g n_{sz} r_{gz}, p_g\right)$$



join scaling factor
and cell counts

$$v_{sz} = \alpha_s n_{sz}$$

$$x_{gs} \sim NB\left(\sum_{z \in Z} \beta_g v_{sz} r_{gz}, p_g\right)$$

■ ■ | Slightly more complex than adding bar graphs..

- Probabilistic Model

$$w_{sz} = \frac{v_{sz}}{\sum_{z \in Z} v_{sz}} = \frac{\alpha_s n_{sz}}{\alpha_s \sum_{z \in Z} n_{sz}} = \frac{n_{sz}}{\sum_{z \in Z} n_{sz}}$$

Slightly more complex than adding bar graphs..

- Probabilistic Model

$$w_{sz} = \frac{v_{sz}}{\sum_{z \in Z} v_{sz}} = \frac{\alpha_s n_{sz}}{\alpha_s \sum_{z \in Z} n_{sz}} = \frac{n_{sz}}{\sum_{z \in Z} n_{sz}}$$

Number of cells from cell type z at spot s

n_{sz}

Total number of cells at spot s

Slightly more complex than adding bar graphs..

- Probabilistic Model

$$w_{sz} = \frac{v_{sz}}{\sum_{z \in Z} v_{sz}} = \frac{\alpha_s n_{sz}}{\alpha_s \sum_{z \in Z} n_{sz}} = \frac{n_{sz}}{\sum_{z \in Z} n_{sz}}$$

Number of cells from cell type z at spot s

Proportion of cell type “ z ” at spot “ s ”

Total number of cells at spot s

The diagram illustrates the components of the equation. A red dashed box highlights the term v_{sz} , which is labeled "Proportion of cell type ‘z’ at spot ‘s’". A green dashed box highlights the denominator $\sum_{z \in Z} n_{sz}$, which is labeled "Total number of cells at spot s". An arrow points from the text "Number of cells from cell type z at spot s" to the term n_{sz} in the final fraction.

Slightly more complex than adding bar graphs..

- Probabilistic Model
 - Use MLE estimate to find unadjusted proportions ($V = [v_{sz}]$)
 - Minimize :
$$l(\mathbf{V}, \beta) = -\log[L(\mathbf{V}, \beta | \mathbf{r}, \mathbf{p}, \mathbf{X})]$$
 Likelihood function
 - Stochastic optimization (PyTorch)

Slightly more complex than adding bar graphs..

- Probabilistic Model
 - Assumes single cell and spatial data are both NB distributed
- Tool : *stereoscope*
- Output : [spot] x [cell_type] matrix
 - Elements are proportion of cell belonging to the given cell type at each spot

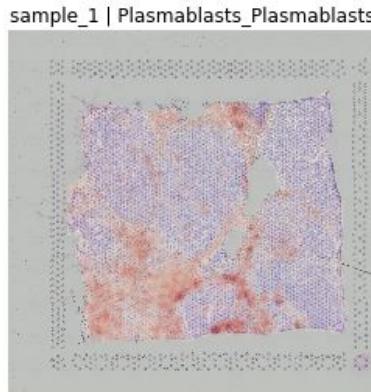
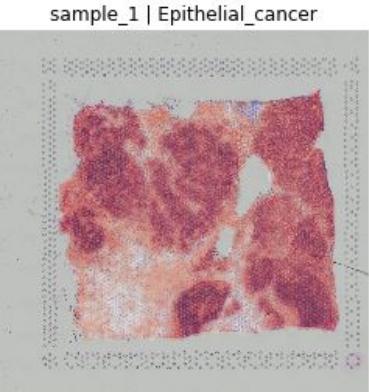
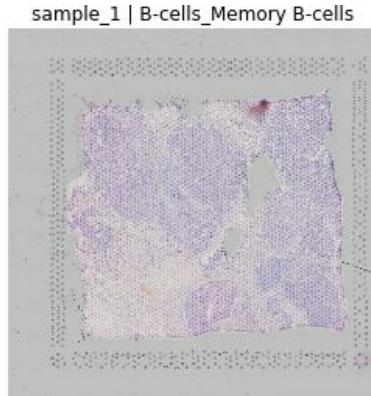
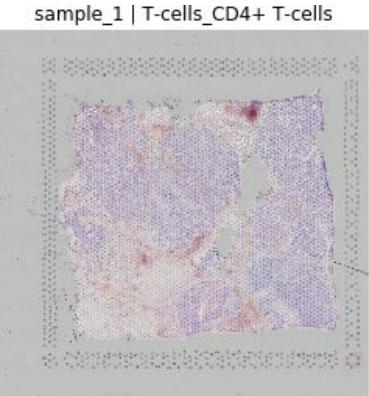


<https://github.com/almaan/stereoscope>

#shameless self-advertising

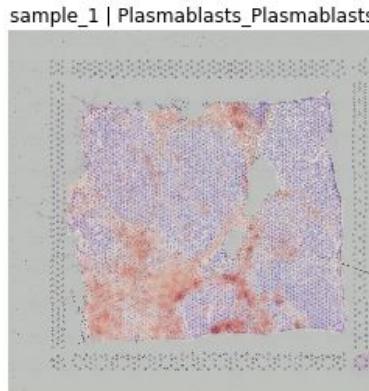
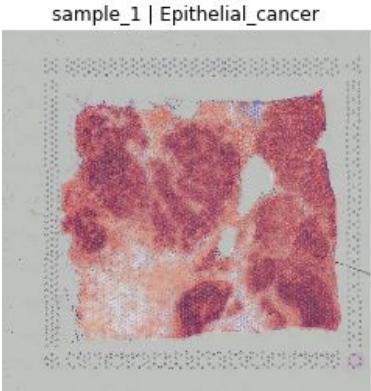
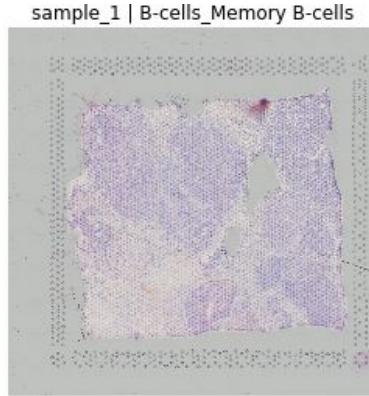
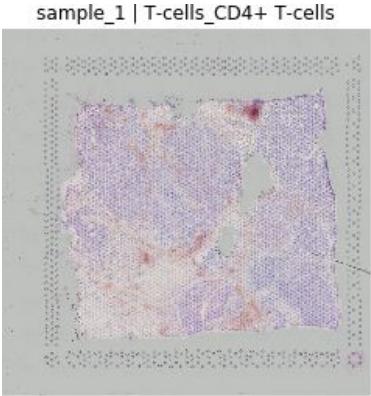
Applying it to our breast cancer data

Proportion estimates

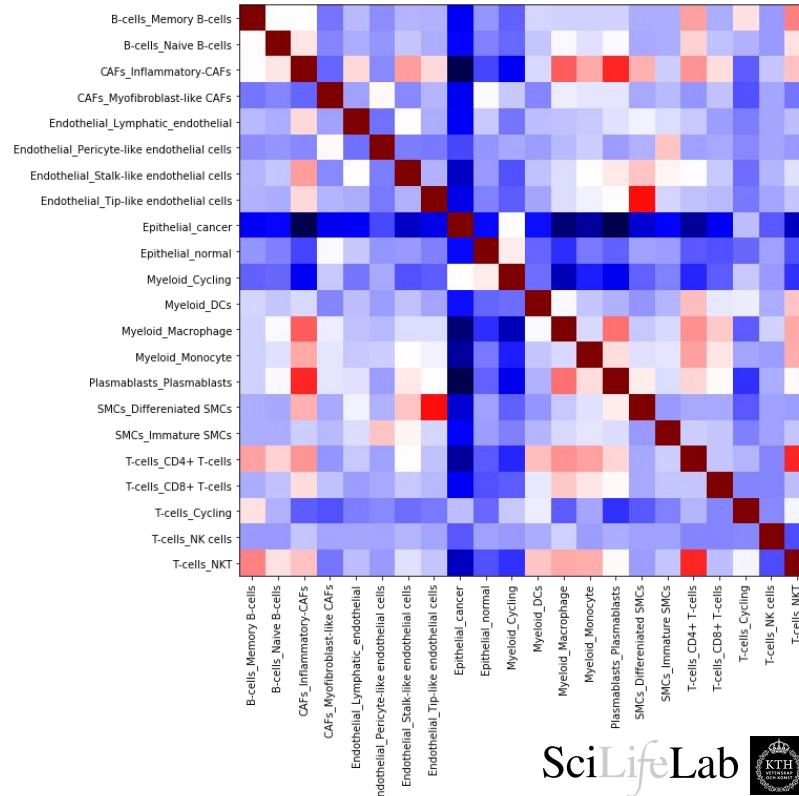


Applying it to our breast cancer data

Proportion estimates



Cell type co-localization



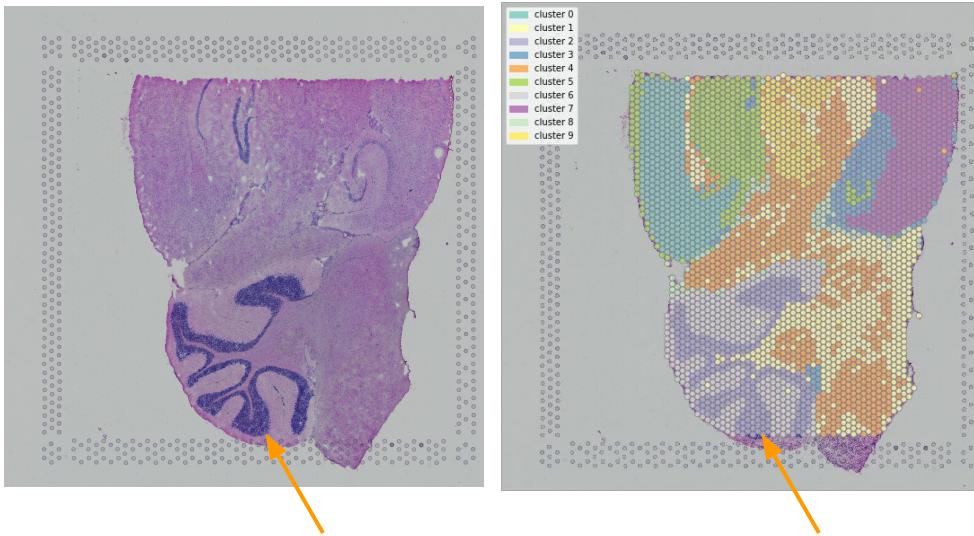


Summary : Integration of single cell and spatial data

- Leverages strengths from respective technique
 - Spatial resolution of well defined cell types
- Can be used as basis in subsequent analyses
 - Patterns of cell type co-localization
- Solution until experimental techniques reach single cell res.
- Atlases are exciting!

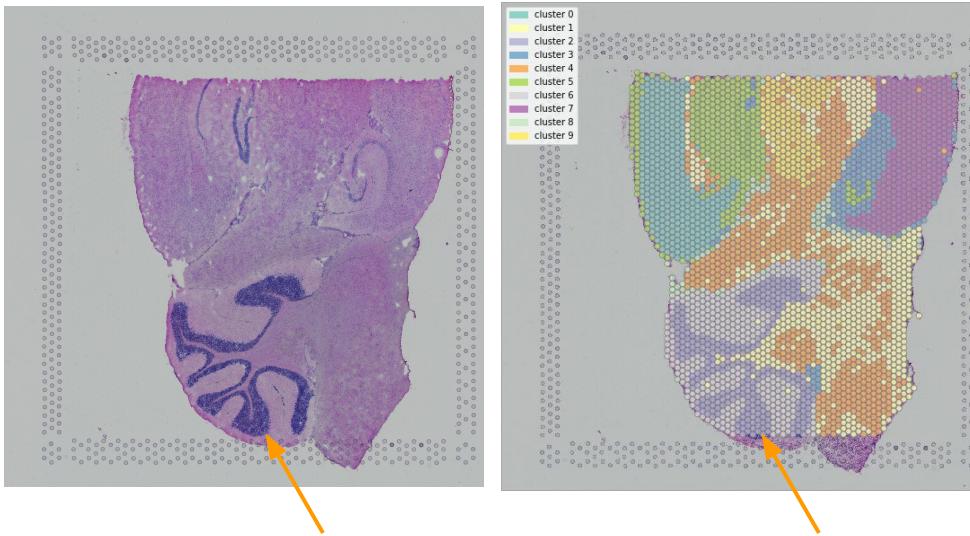
Example Analysis : Expression as a function of distance

- Say we cluster spatial data and find an interesting domain (e.g. cluster 2)



Example Analysis : Expression as a function of distance

- Say we cluster spatial data and find an interesting domain (e.g. cluster 2)

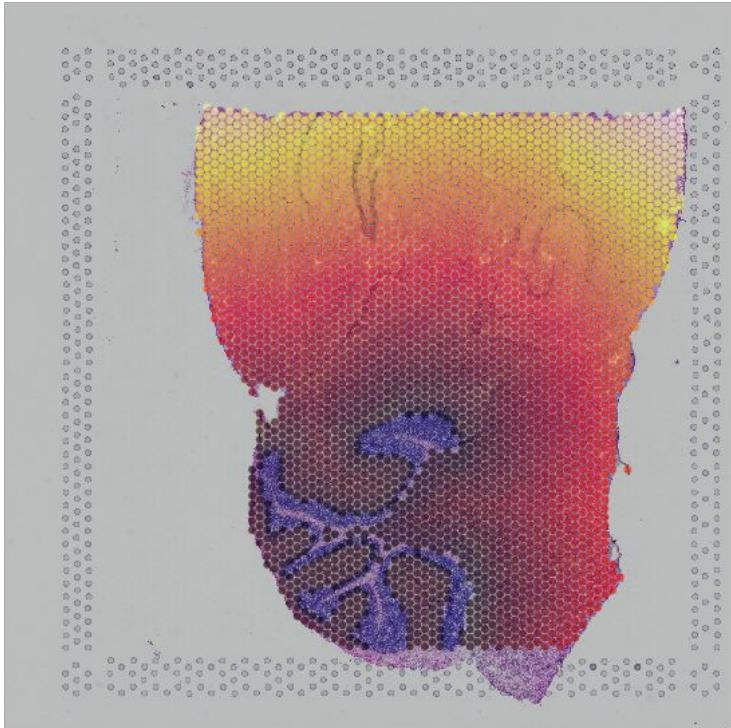


- Now, we may ask how expression varies as a function of the distance to this cluster?

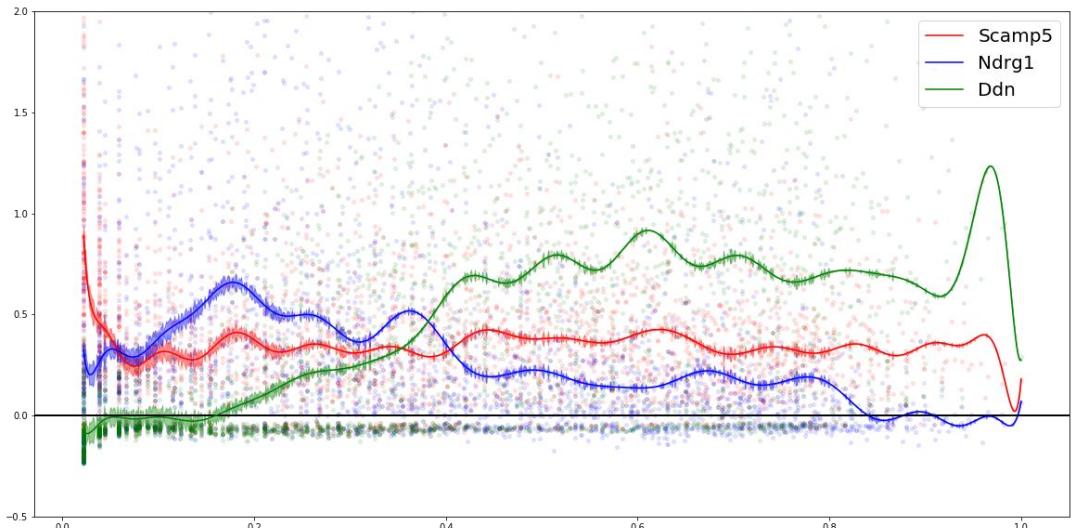


Example Analysis : Expression as a function of distance

P3

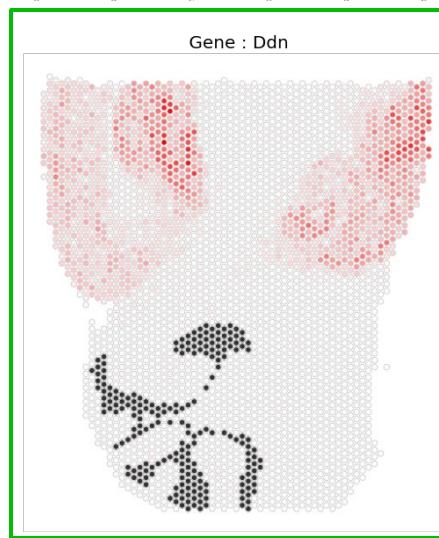
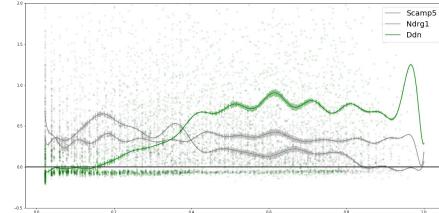
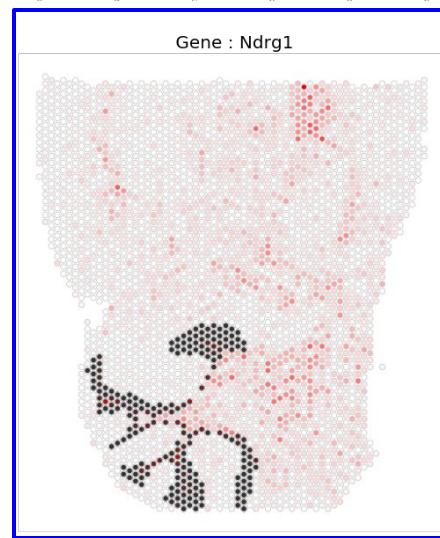
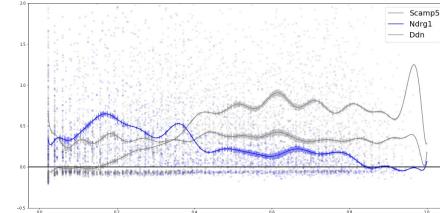
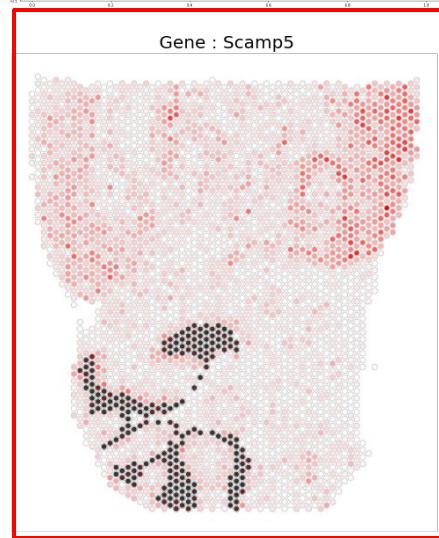
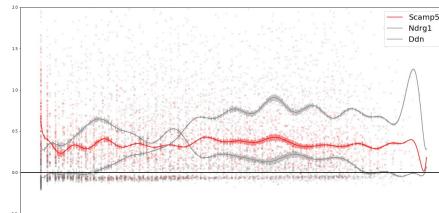


Spots colored by distance do cluster 2



Gene expression as a function of the distance
to cluster 2

Example Analysis : Expression as a function of distance

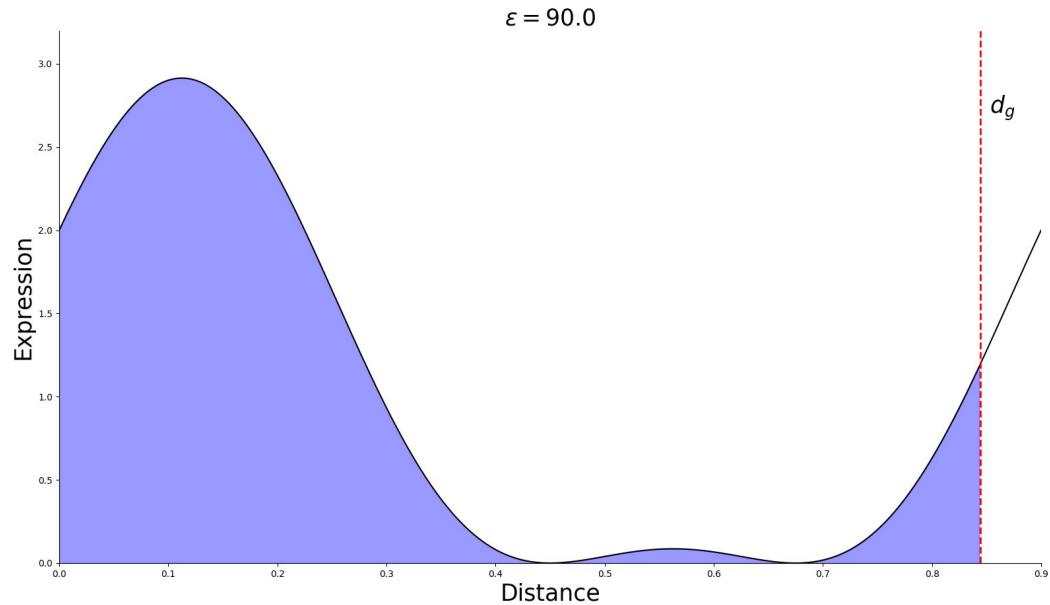
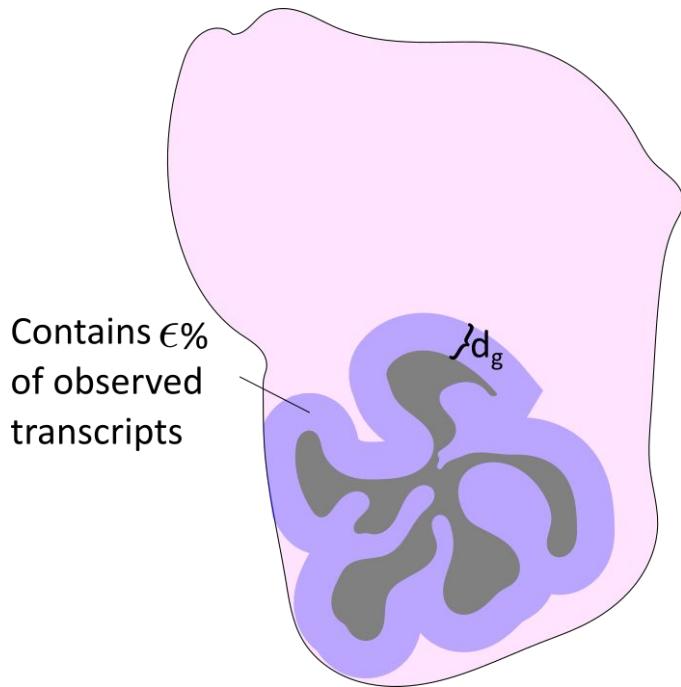


■ ■ ■ Example Analysis : Expression as a function of distance

- Can also ask : “*within which distance (d_g) from cluster 2 is a certain amount ($\varepsilon %$) of transcripts are contained*”

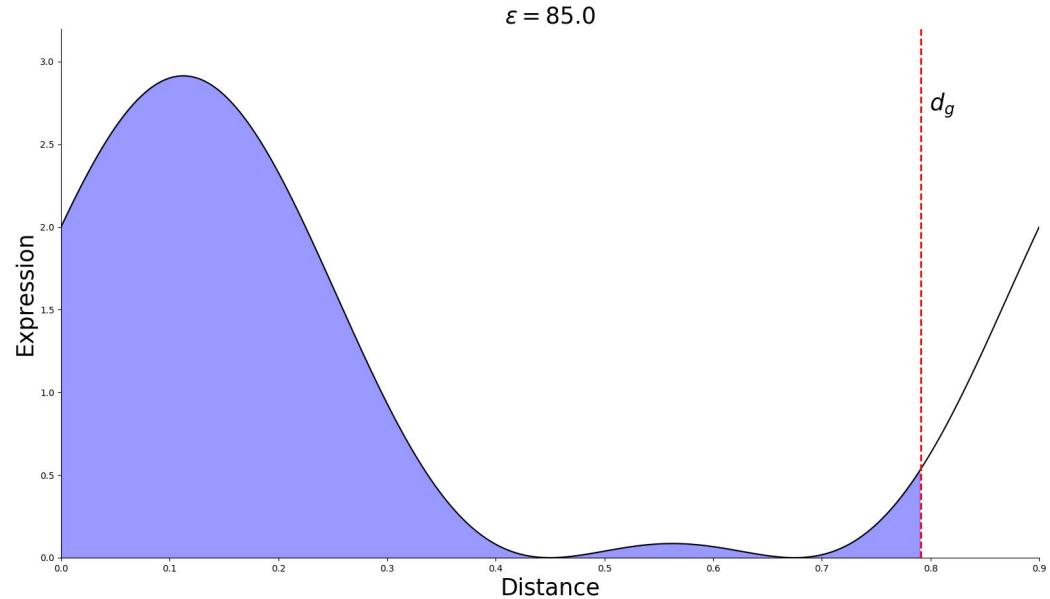
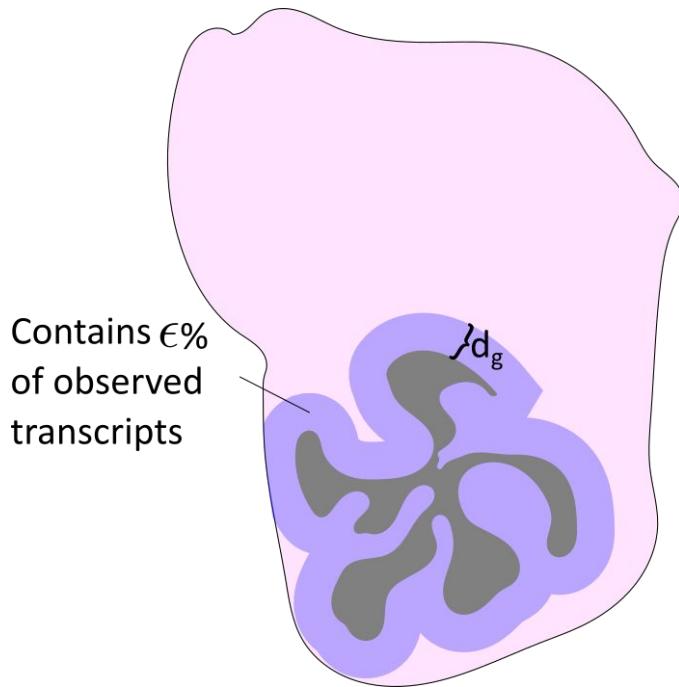
Example Analysis : Expression as a function of distance

- Can also ask : “within which distance (d_g) from cluster 2 is a certain amount (ϵ %) of transcripts are contained”



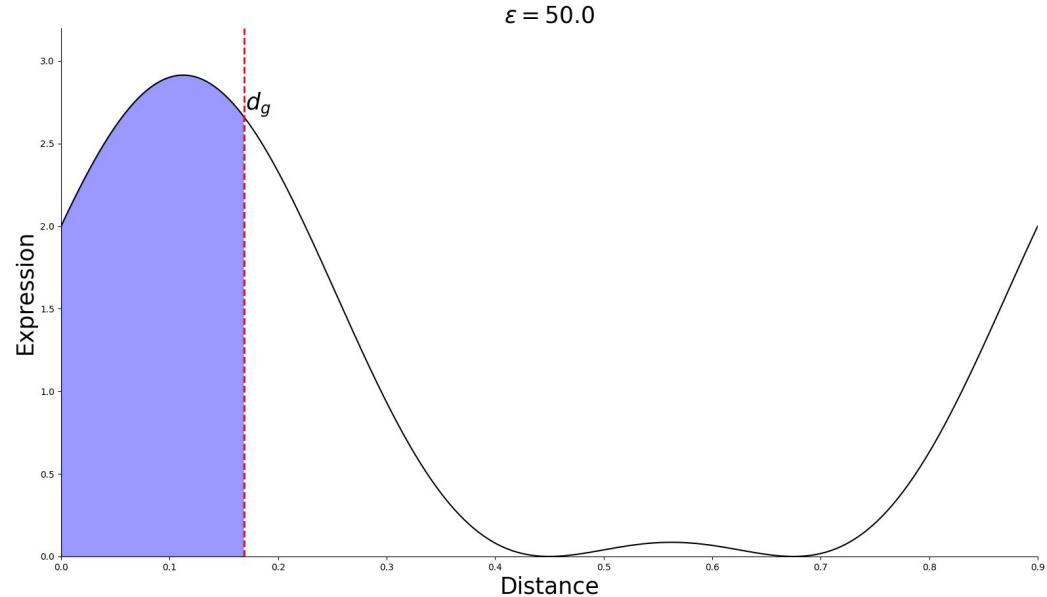
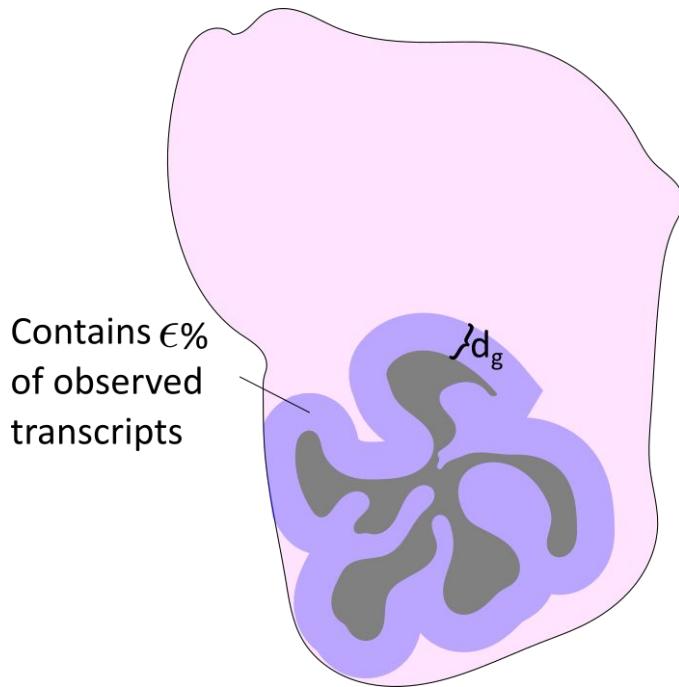
Example Analysis : Expression as a function of distance

- Can also ask : “within which distance (d_g) from cluster 2 is a certain amount (ϵ %) of transcripts are contained”



Example Analysis : Expression as a function of distance

- Can also ask : “within which distance (d_g) from cluster 2 is a certain amount (ϵ %) of transcripts are contained”

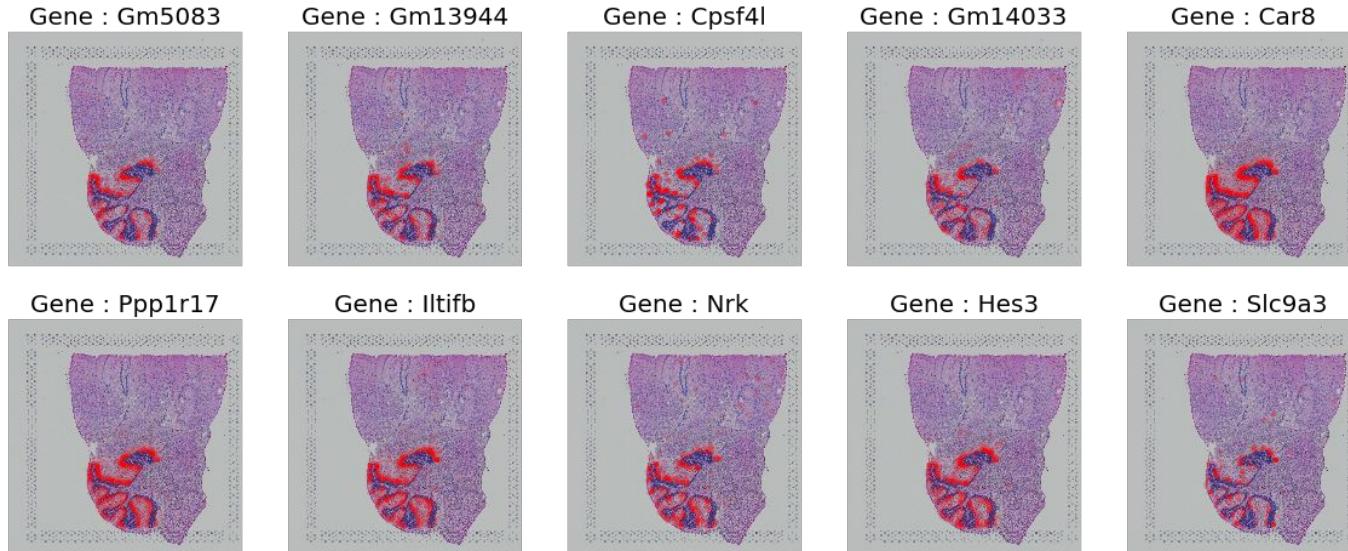


■ ■ ■ Example Analysis : Expression as a function of distance

- Alternatively : “which genes has $\varepsilon\%$ within the shortest distance (d_g) from cluster 2”

Example Analysis : Expression as a function of distance

- Alternatively : “which genes has $\epsilon\%$ within the shortest distance (d_g) from cluster 2”



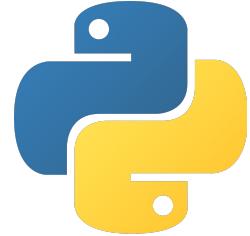


Exercises



Exercise Session

- Aims:
 - Getting familiar with spatial data
 - Overview - concept focused
- Written to be “independent” of lectures
 - Might experience some redundancy
 - but... *repetitio est mater studiorum*
- Three Parts
 - Part 1 - “*Getting Comfy with Spatial Data*”
 - Orienting, Inspecting and visualizing spatial data
 - Basic analysis workflow
 - Part 2 - “*Integrating Single Cell and Spatial RNA-Seq Data*”
 - Working with mapped data
 - Downstream analysis
 - Part 3 - “*Digging deeper into spatial analysis*”
 - Spatial gene set enrichment
 - Expression as a function of distance



■ ■ || Questions



Thank you for the attention