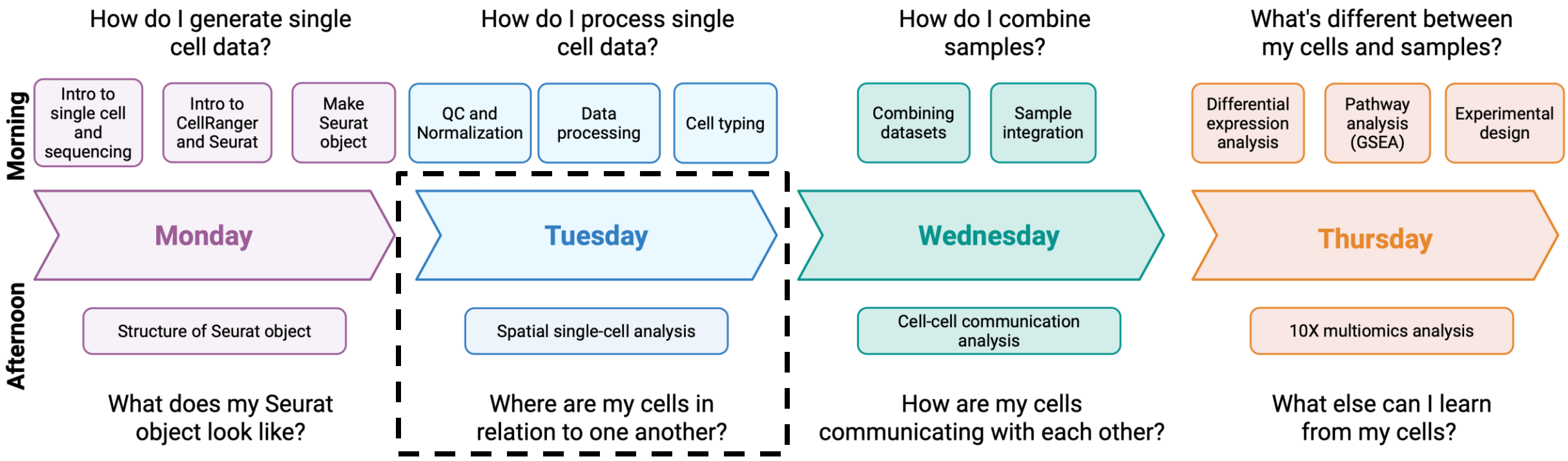
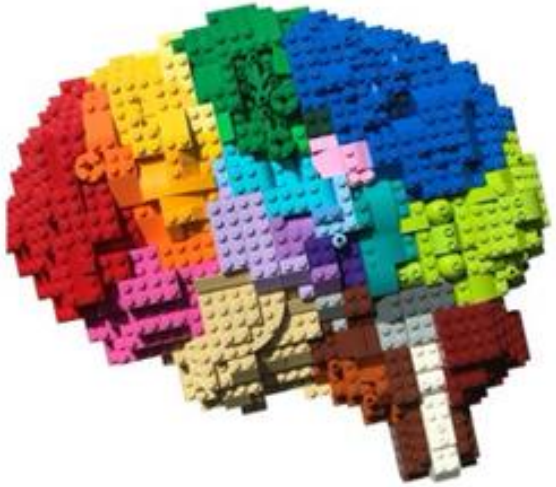
scRGOT Coder Upgrade: Spatial Transcriptomics

Corinne Strawser, PhD
Bioinformatics Scientist
Dr. Elaine Mardis & Dr. Katie Miller Labs
Institute for Genomic Medicine

NATIONWIDE CHILDREN'S
When your child needs a hospital, everything matters.


How do I generate single cell data?
How do I process single cell data?
How do I combine samples?
What's different between my cells and samples?
Morning
Intro to single cell and sequencing
Intro to CellRanger and Seurat
Make Seurat object
QC and Normalization
Data processing
Cell typing
Combining datasets
Sample integration
Differential expression analysis
Pathway analysis (GSEA)
Experimental design
Monday
Tuesday
Wednesday
Thursday
Afternoon
Structure of Seurat object
Spatial single-cell analysis
Cell-cell communication analysis
10X multiomics analysis
What does my Seurat object look like?
Where are my cells in relation to one another?
How are my cells communicating with each other?
What else can I learn from my cells?

Why spatial?



Complex Tissue

Networks of many cells each with individual gene expression patterns that regulate interactions and ultimately tissue function

Bulk RNA-seq

tissue-average gene expression

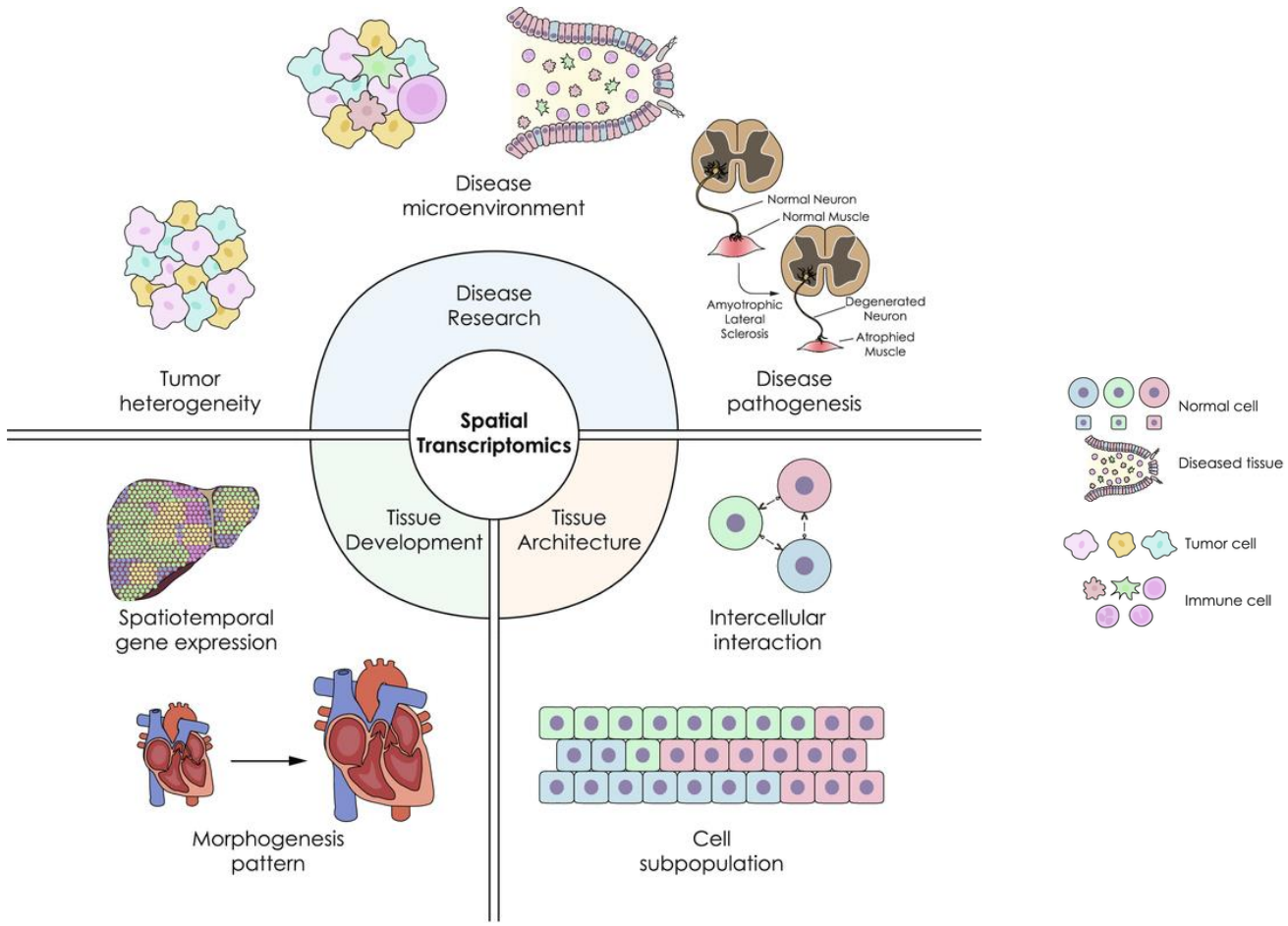
Single-cell RNA-seq

Cell-specific gene expression

Spatial Transcriptomics

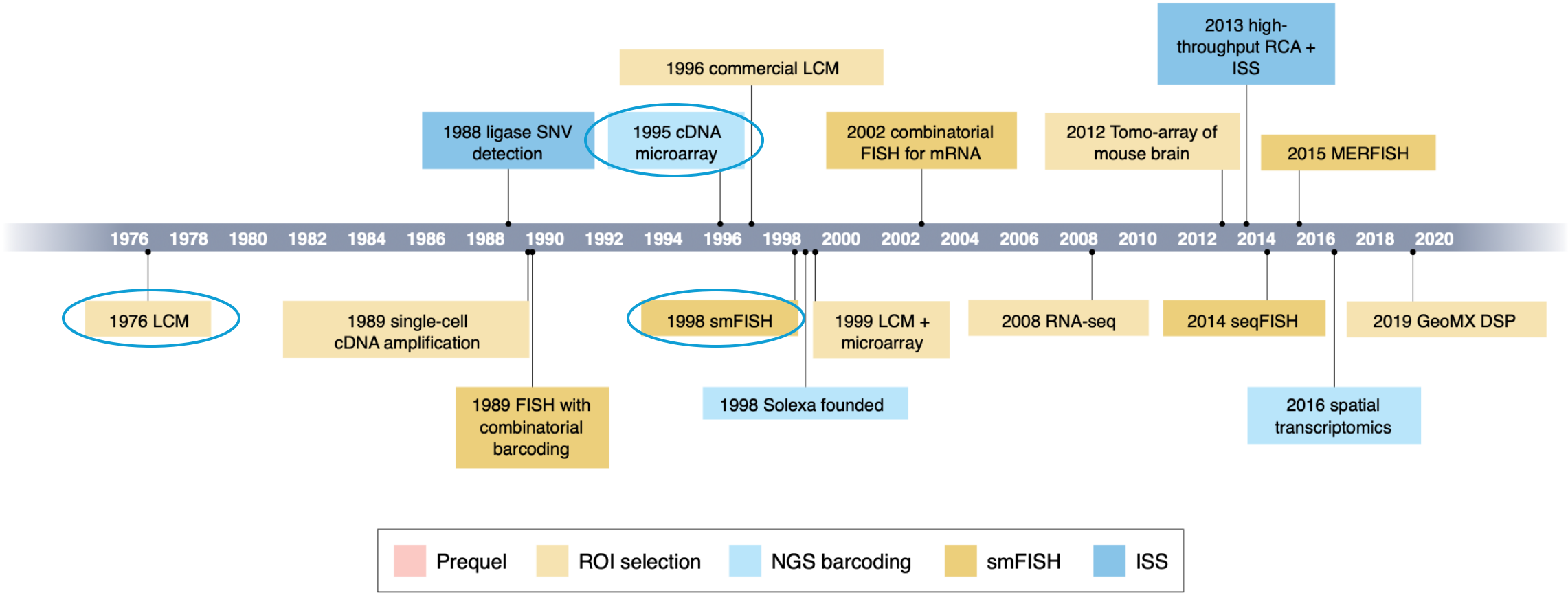
region-, cell-, or subcellular-specific gene expression with spatial context

Spatial transcriptomics has broad applications



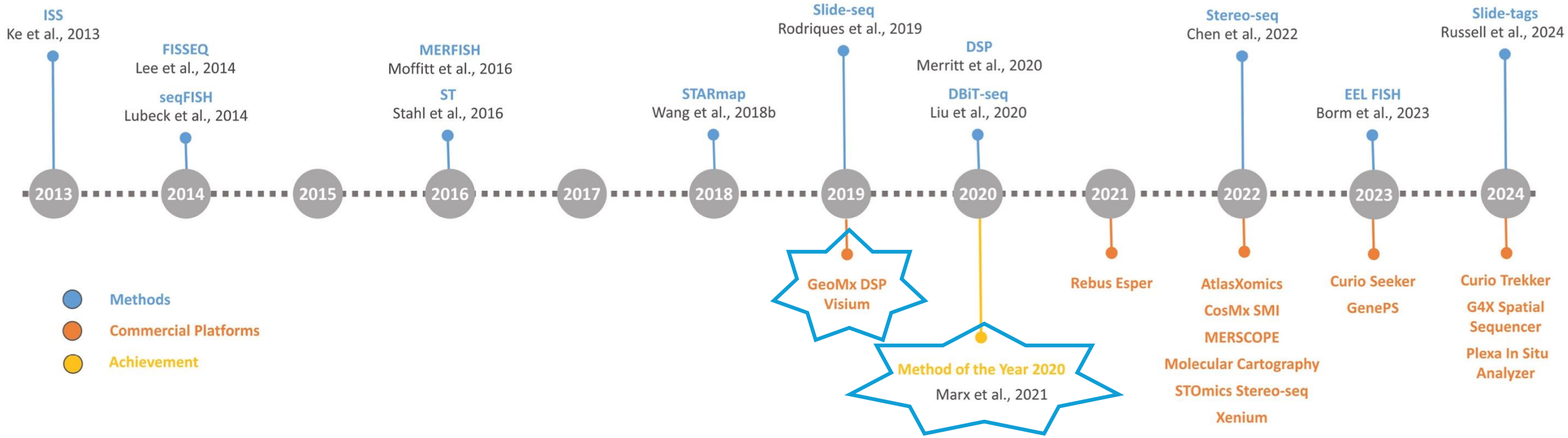
Du, Jun & Yang et al. (2023). Advances in spatial transcriptomics and related data analysis strategies. *Journal of Translational Medicine*. https://doi.org/10.1186/s12967-023-04150-2.

Spatial transcriptomics is not “new”…

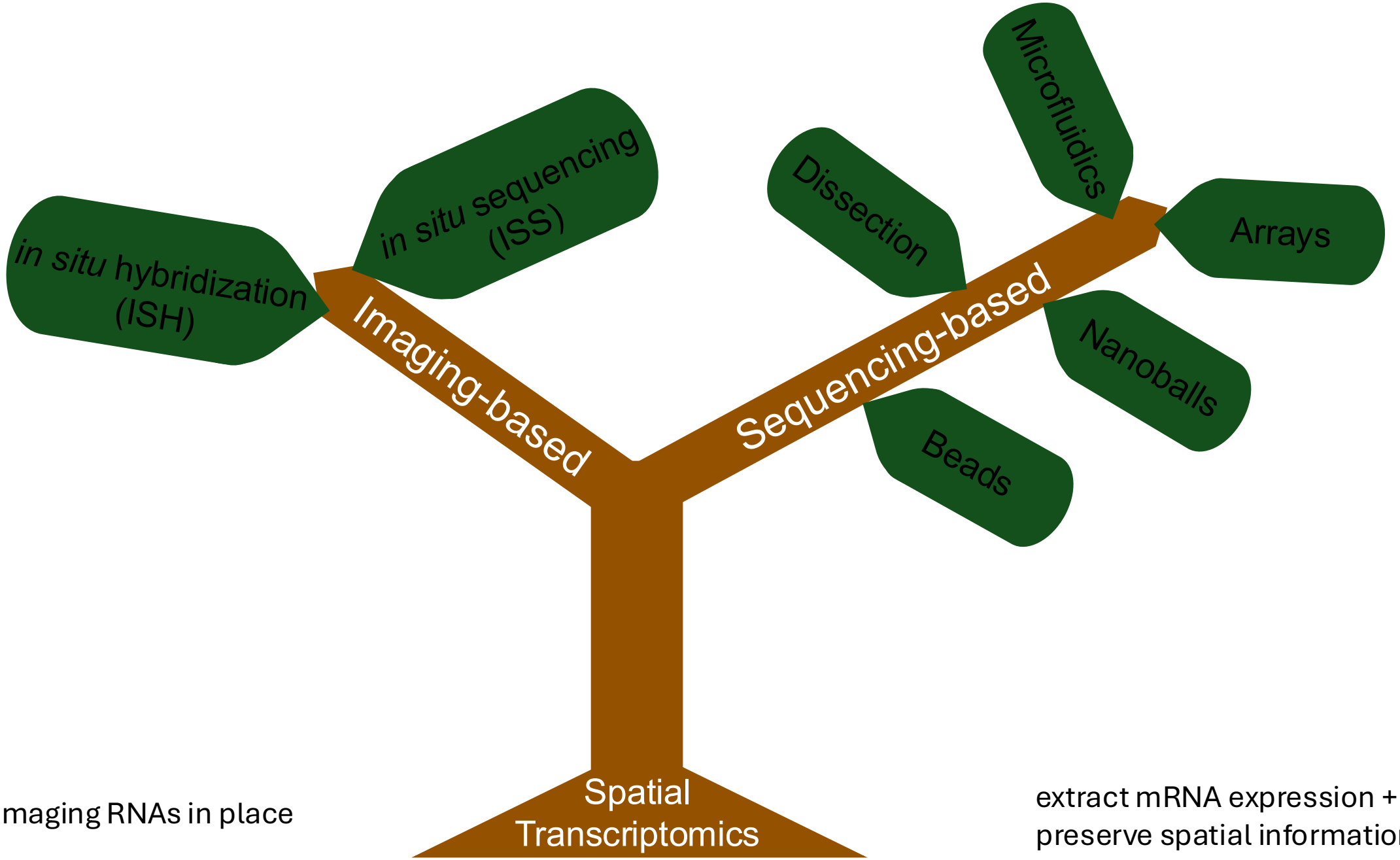


Moses, L., & Pachter, L. (2022). Museum of spatial transcriptomics. *Nature methods*. https://doi.org/10.1038/s41592-022-01409-2

…but it's expanding!



Valihrach, L., et al (2024). A practical guide to spatial transcriptomics. *Molecular aspects of medicine*. https://doi.org/10.1016/j.mam.2024.101276

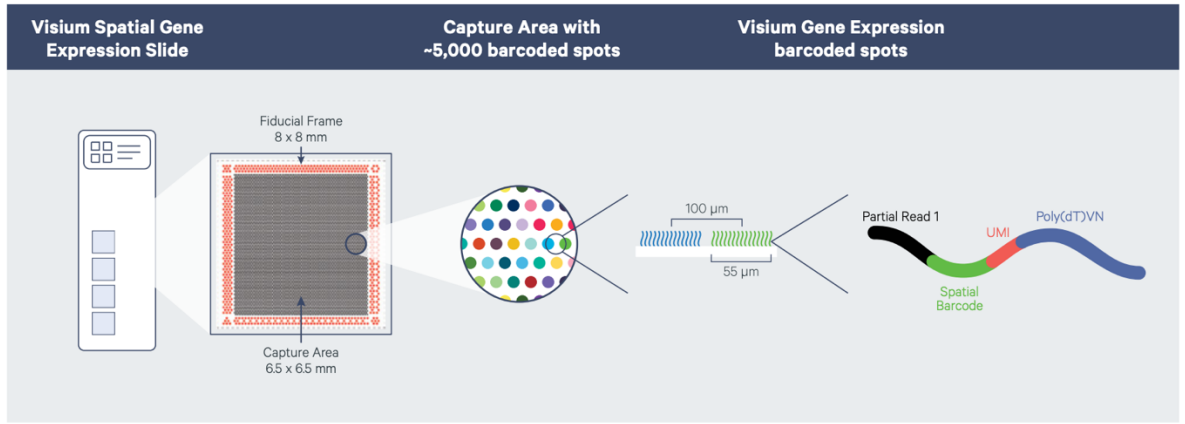


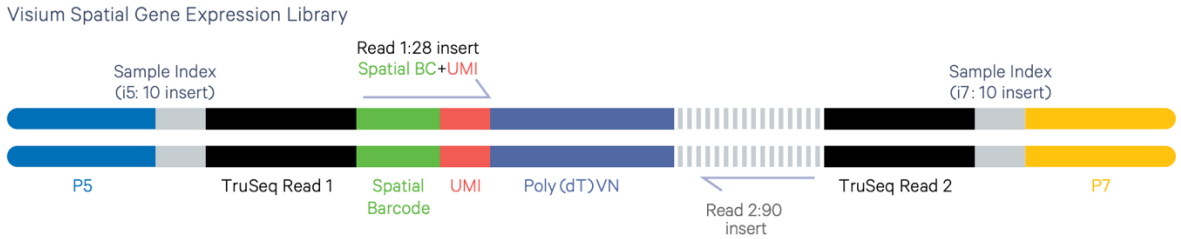
imaging RNAs in place

extract mRNA expression + preserve spatial information

10X Genomics Visium

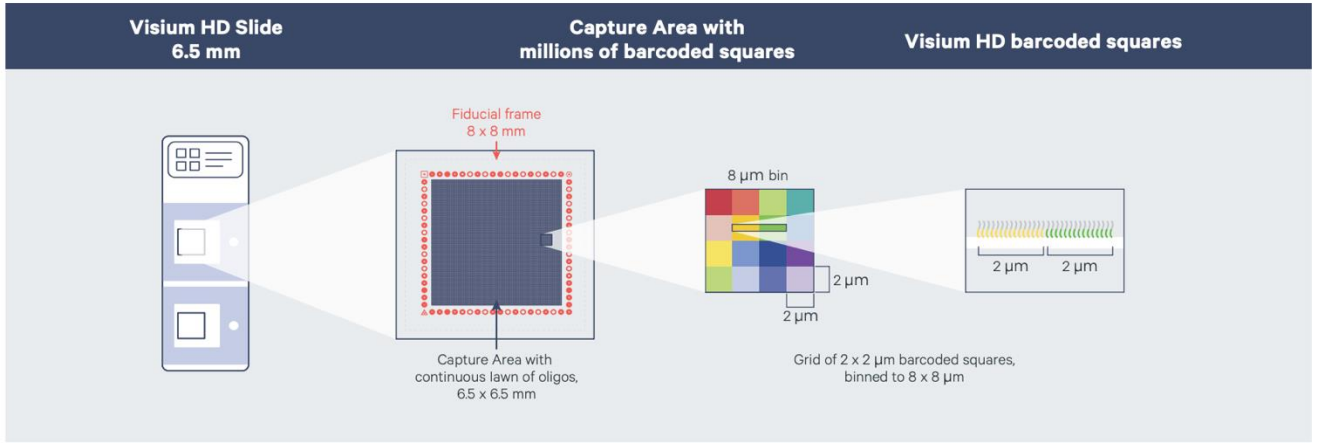
- Visium slides has two or four capture areas of either 6.5 x 6.5 mm or 11 x 11 mm.
- ~5,000 capture spots (6.5 x 6.5 mm capture area)
- ~14,000 barcoded spots (11 x 11 mm capture area)
- Each spot has millions of oligonucleotides with:
  - poly(dT) sequence for the capture of polyadenylated molecules
  - UMI used for the identification of duplicate molecules that arise during the library preparation and sequencing process
  - Spatial Barcode, which is shared by all oligonucleotides within each individual gene expression capture spot

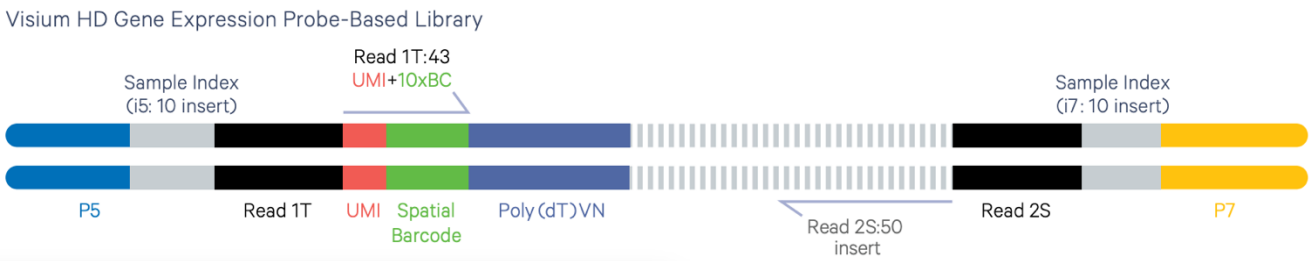




10X Genomics Visium HD

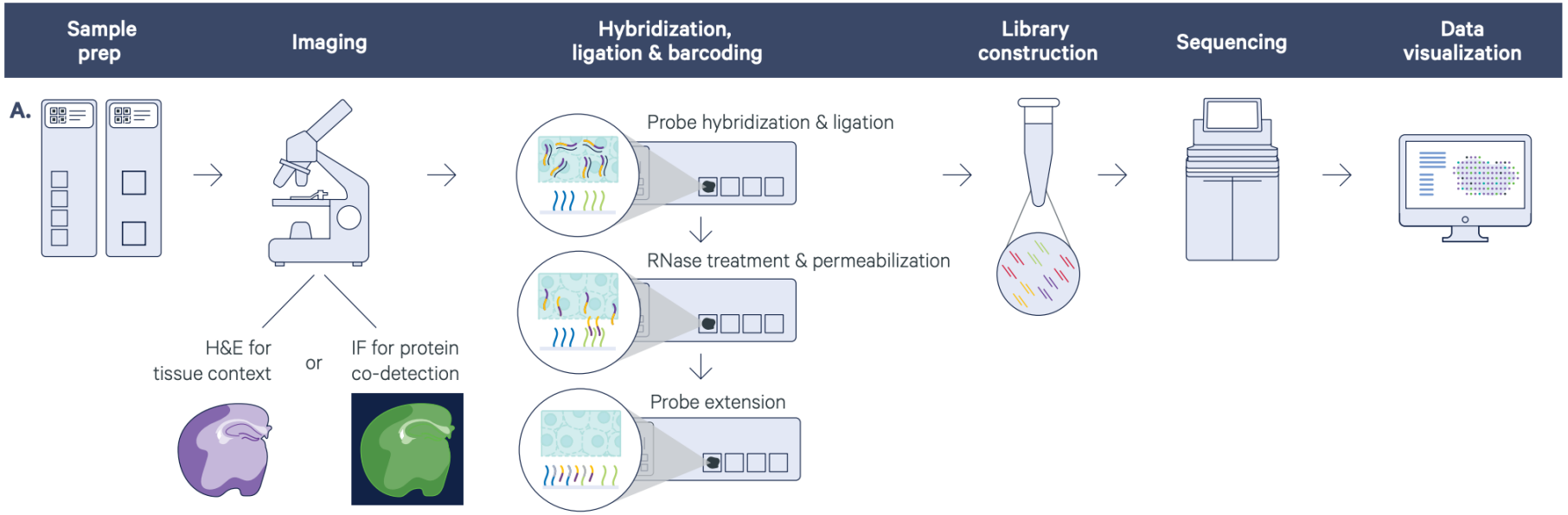
- The HD slide has two capture areas, each 6.5 x 6.5 mm
- ~ 11 million 2 x 2 µm barcoded squares without gaps
- Oligonucleotides each have
  - poly(dT) sequence for capture of the ligation product
  - UMI for the identification of duplicate molecules that arise during the library preparation and sequencing process
  - Spatial Barcode



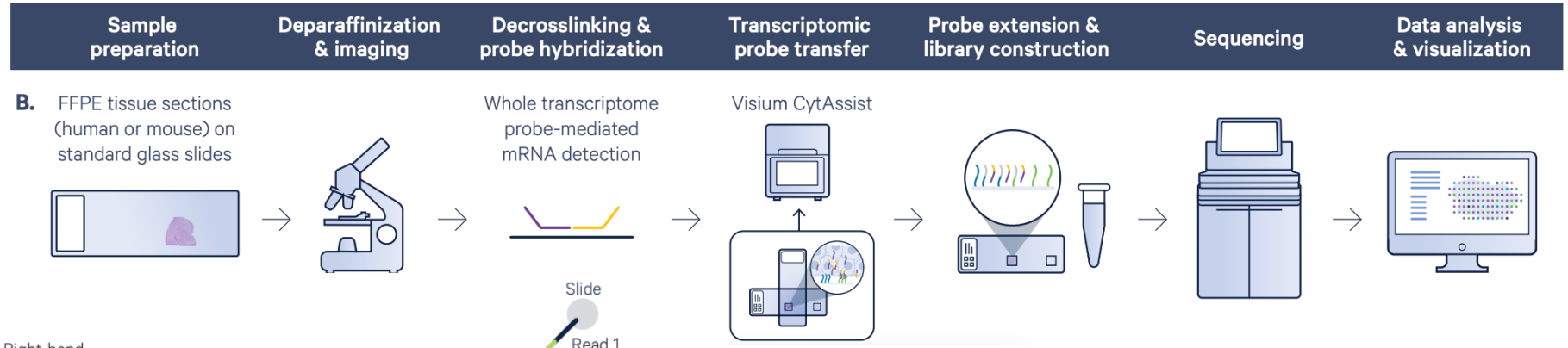


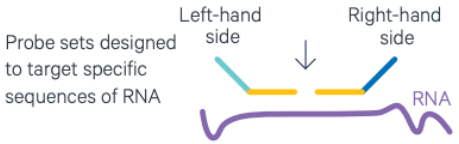
10X Genomics Visium Workflows

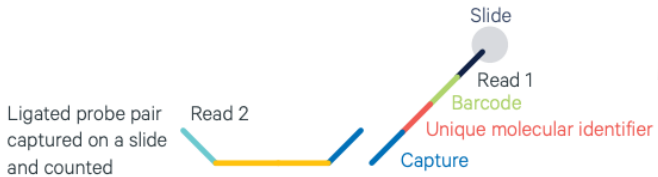
poly-A Visium Workflow



Probe-based Visium & Visium HD workflow







10x GENOMICS®

10X Genomics Visium and Visium HD

PRO

- Very well documented
- Microscopy image
- FFPE, fresh frozen, and fixed frozen (HD and LD v2) or fresh frozen (LD v1)
- Protein co-detection
- Customizable
- Any species (LD v1)

CON

- U-Pick: subcellular (HD) or multi-cellular (LD)
- Probe-based (HD & LD v2) – *Visium HD 3' coming soon*
- CytAssist required (HD & LD v2)
- Mouse or human (HD & LD v2)

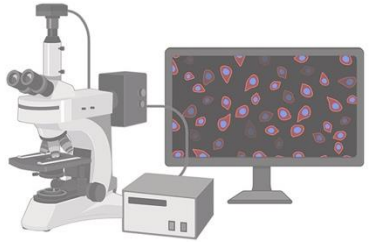
Spatial Transcriptomics Analysis

Raw-data processing

Pre- processing

Downstream analysis

Imaging-based



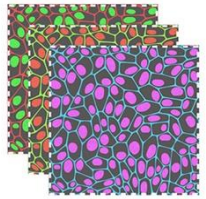
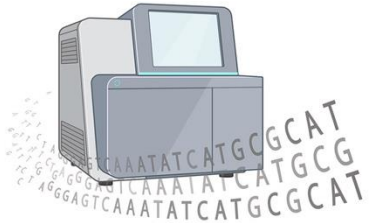
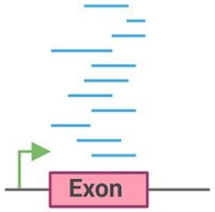


Image registration
Decoding
Segmentation

Sequencing-based





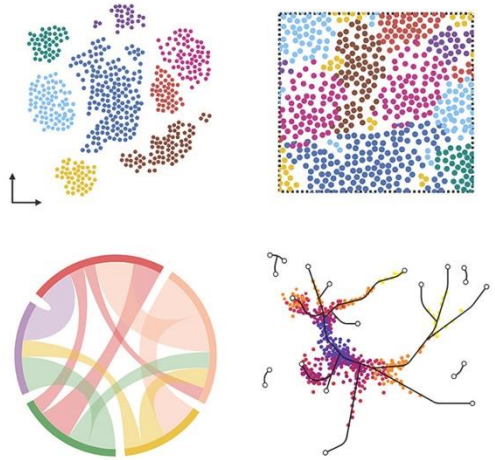
Trimming
Alignment
UMI correction

Cells / Spots

Genes

28	77	65	44	74	80	59	7	45	44	96
98	99	54	43	27	83	81	65	28	41	19
59	59	11	96	50	16	79	45	21	52	70
81	46	43	95	89	68	35	45	32	38	7
15	37	88	90	65	95	10	40	85	97	63
56	29	44	6	4	82	46	25	90	18	71
62	42	68	36	7	71	69	4	93	58	60
20	10	60	75	95	17	67	93	95	45	59
67	58	83	71	32	27	86	8	46	38	79
51	64	11	47	66	19	68	99	29	6	64
39	71	69	79	57	7	91	30	28	19	26
26	55	15	94	94	95	79	29	65	65	84
78	53	43	21	35	39	71	25	83	45	65
89	9	62	22	8	78	96	72	80	18	8
38	38	45	36	36	88	72	82	88	43	87

Counting
Quality control



Normalization
Integration
Imputation
Dimensionality reduction
Clustering
Deconvolution
Multimodal analysis
Annotation & characterization
Spatially variable genes
Spatial domains
Spatial interaction analysis
Spatial trajectory inference

Valihrach, L., Zucha, D., Abaffy, P., & Kubista, M. (2024). A practical guide to spatial transcriptomics. *Molecular aspects of medicine*, *97*, 101276. https://doi.org/10.1016/j.mam.2024.101276

spaceranger mkfastq

mkfastq sample sheet

run_spaceranger_mkfastq.sh | samplesheet.csv

igm > projects > 241014_GSL-LPRV-4138_Visium > fastq > samplesheet.csv > data

```
Lane,Sample,Index,
*,NP-20-0025854_C1_Pre_Gene,SI-TS-E6,
*,NP-20-0043174_D4_Post_Gene,SI-TS-A5,
*,NP-20-0025854_C1_Pre_Protein,SI-NT-E1,
*,NP-20-0043174_D4_Post_Protein,SI-NT-F1,
*,NP-19-0035324_A1_Gene,SI-TS-D7,
*,NP-19-0042829_B3_Gene,SI-TS-D8,
*,NP-19-0035324_A1_Protein,SI-NT-C1,
*,NP-19-0042829_B3_Protein,SI-NT-D1,
```

mkfastq script

run_spaceranger_mkfastq.sh | samplesheet.csv

igm > projects > 241014_GSL-LPRV-4138_Visium > fastq > run_spaceranger_mkfastq.sh

```
#$ -q all.q
#$ -j y
#$ -cwd
#$ -S /bin/bash
#$ -terse
#$ -pe smp 16
#$ -N cellranger_mkfastq

module load spaceranger_3.0.1

spaceranger mkfastq \
--samplesheet=samplesheet.csv \
--run=/igm/runs/IGM_Seq10/241018_A00498_0918_BHH27NDRX5 >& bclconv.out
```

* note you'll have to change the header if you're using Slurm as your job scheduler (I'm using SGE here).

spaceranger count samplesheet

NP-19-0035324_A1.csv × $ NP-19-0035324_A1.sh ●

igm > projects > 241014_GSL-LPRV-4138_Visium > counts > NP-19-0035324_A1 > NP-19-0035324_A1.csv > data

```
1  fastqs,sample,library_type
2  /igm/projects/241014_GSL-LPRV-4138_Visium/fastq/HH27NDRX5/outs/fastq_path/HH27NDRX5,NP-19-0035324_A1_Gene,Gene Expression
3  /igm/projects/241014_GSL-LPRV-4138_Visium/fastq/HH27NDRX5/outs/fastq_path/HH27NDRX5,NP-19-0035324_A1_Protein,Antibody Capture
```

spaceranger count script

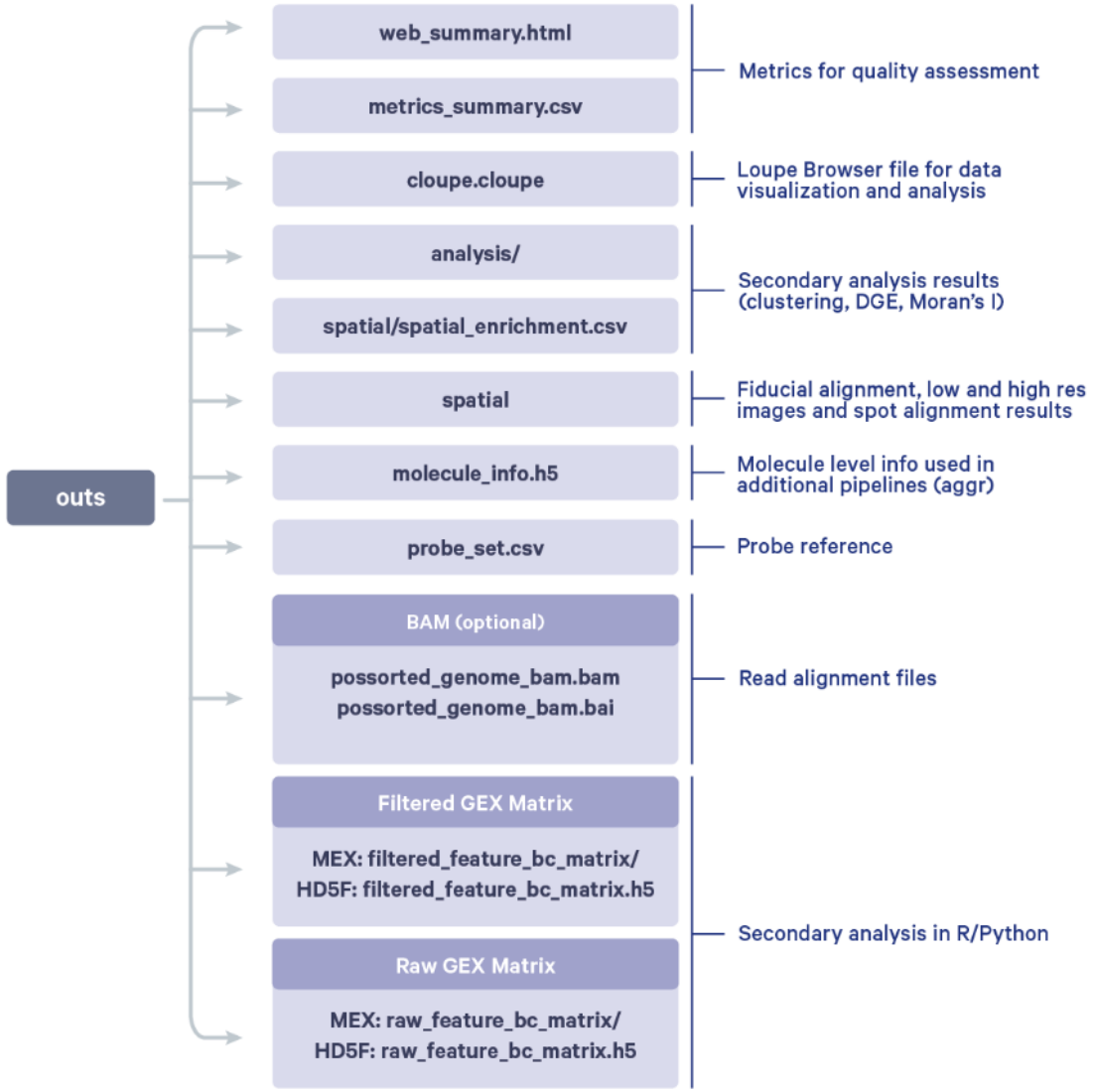
NP-19-0035324_A1.csv $ NP-19-0035324_A1.sh

igm > projects > 241014_GSL-LPRV-4138_Visium > counts > NP-19-0035324_A1 > $ NP-19-0035324_A1.sh

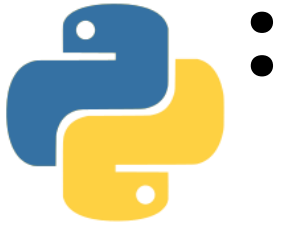
```
#$ -j y
#$ -cwd
#$ -S /bin/bash
#$ -N spaceranger_count
#$ -q all.q
#$ -pe smp 16
#$ -V

# set up
dir="/igm/projects/241014_GSL-LPRV-4138_Visium/counts/NP-19-0035324_A1"
id="NP-19-0035324_A1_Pre"
libraries="/igm/projects/241014_GSL-LPRV-4138_Visium/counts/NP-19-0035324_A1/NP-19-0035324_A1.csv"
image="/igm/projects/240725_GSL-LPRV-4007/images/H&E/NP-19-0035324_A1_Pre_transformed.jpeg"
cytaimage="/igm/projects/240725_GSL-LPRV-4007/images/assay_CAVG10321_2024-07-18_11-19-22_V53A24-007_1721316809_CytAssist/
CAVG10321_2024-07-18_11-33-29_2024-07-18_11-19-22_V53A24-007_A_NP-19-0035324_A1.tif"
alignment="/igm/projects/240725_GSL-LPRV-4007/images/assay_CAVG10321_2024-07-18_11-19-22_V53A24-007_1721316809_CytAssist/
final_alignment_np_19_0035324_a1.json"
transcriptome="/igm/home/cnh008/ref/C134_M032/refdata-gex-GRCh38-2020-A-C134"
probes="/igm/home/cnh008/ref/C134_M032/Visium_Human_Transcriptome_Probe_Set_v2.0_GRCh38-2020_M032_C134_IDT_custom_probes.csv"
protein="/igm/apps/10X_chromium/spaceranger-3.0.0/feature_refs/Visium_Human_Immune_Cell_Profiling_Panel_v1.0.csv"

# script
cd ${dir}
module purge
module load spaceranger_3.0.1

spaceranger count \
    --id=${id} \
    --transcriptome=${transcriptome} \
    --libraries=${libraries} \
    --probe-set=${probes} \
    --image=${image} \
    --cytaimage=${cytaimage} \
    --create-bam=true \
    --loupe-alignment=${alignment} \
    --feature-ref=${protein} \
    --localcores=16 \
    --localmem=64
```

* note you'll have to change the header if you're using Slurm as your job scheduler (I'm using SGE here).

Raw data processing:

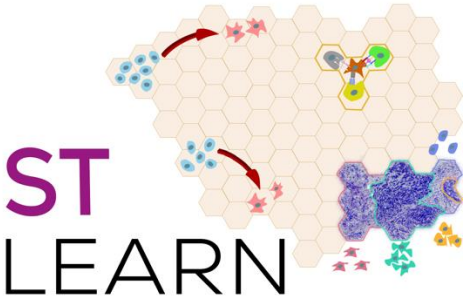


Downstream analysis: choose your own adventure!

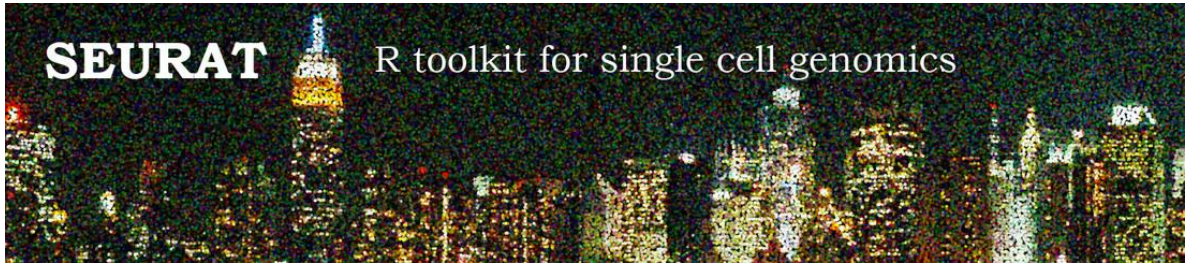


:

Giotto



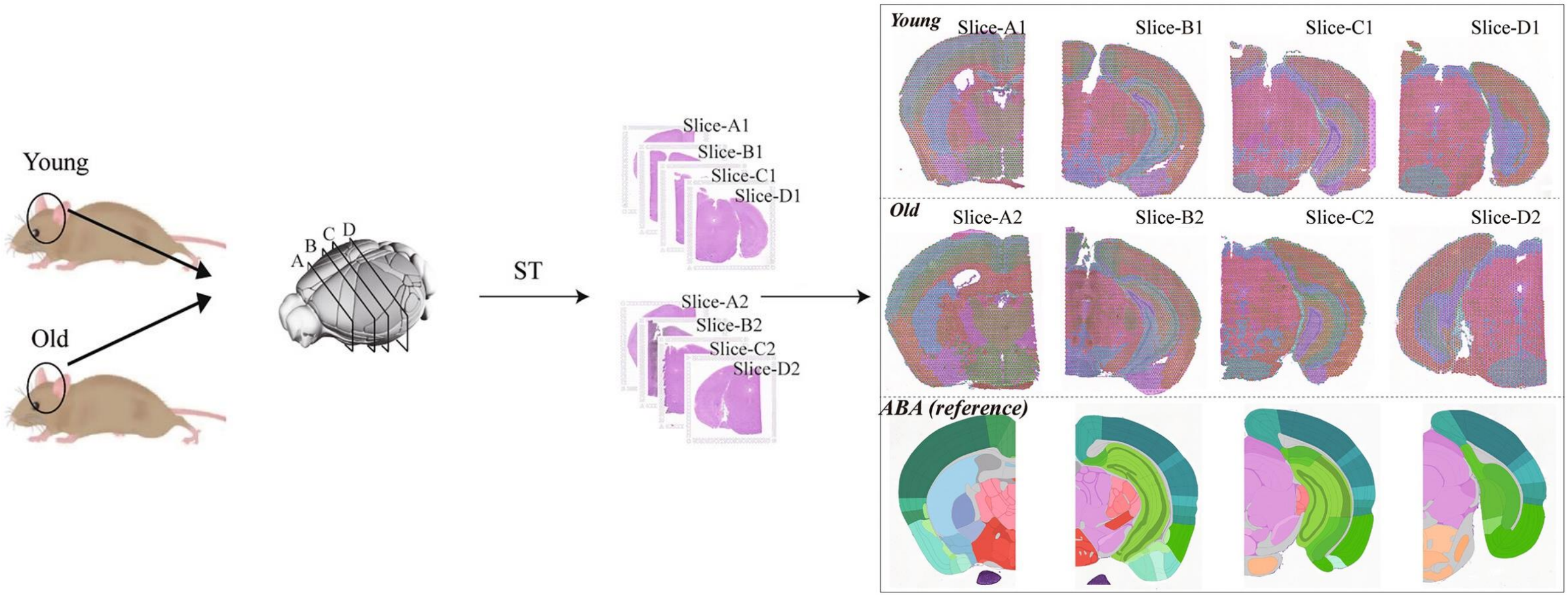
:



Giotto

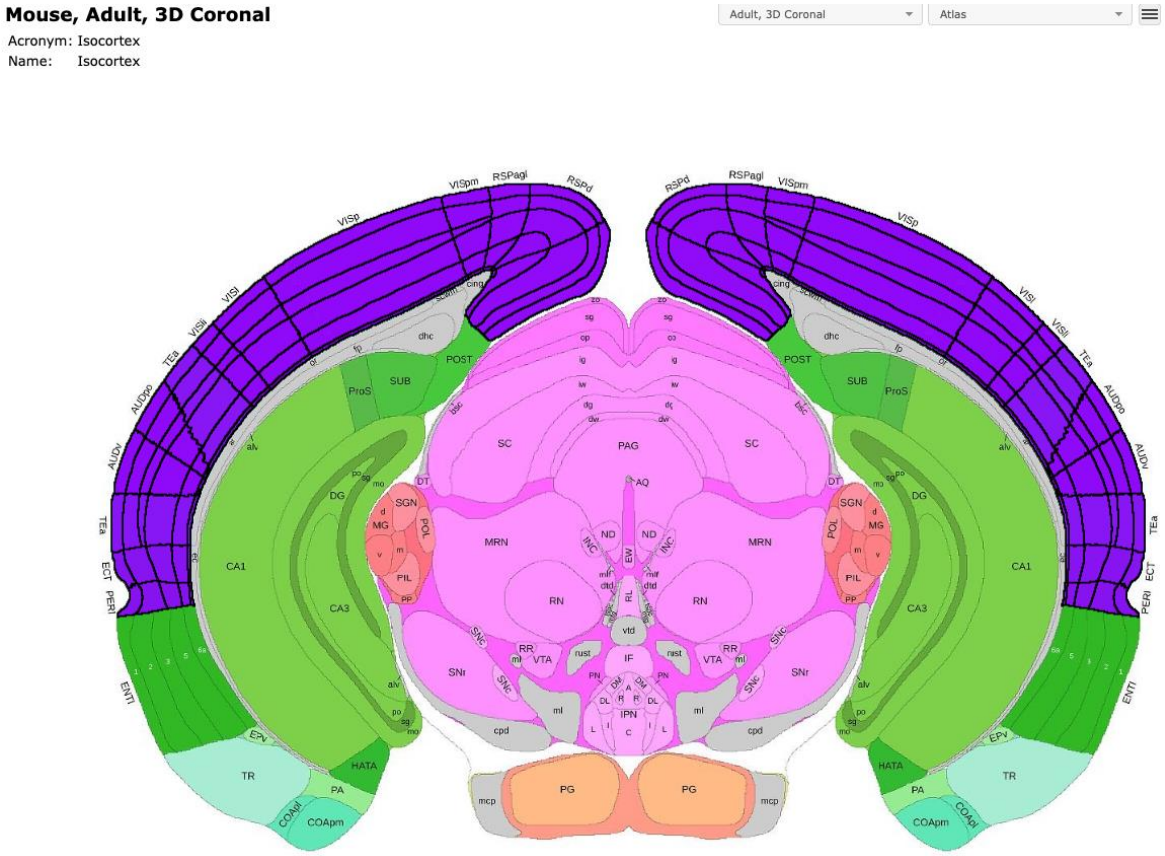
(formerly STUtility)

Today's data: Spatial transcriptomic map of aging mouse brain



Wu and Tu et al. 2024. Spatially resolved transcriptome of the aging mouse brain. Aging Cell. https://doi.org/10.1111/acel.14109

Regions of Interest (ROI)

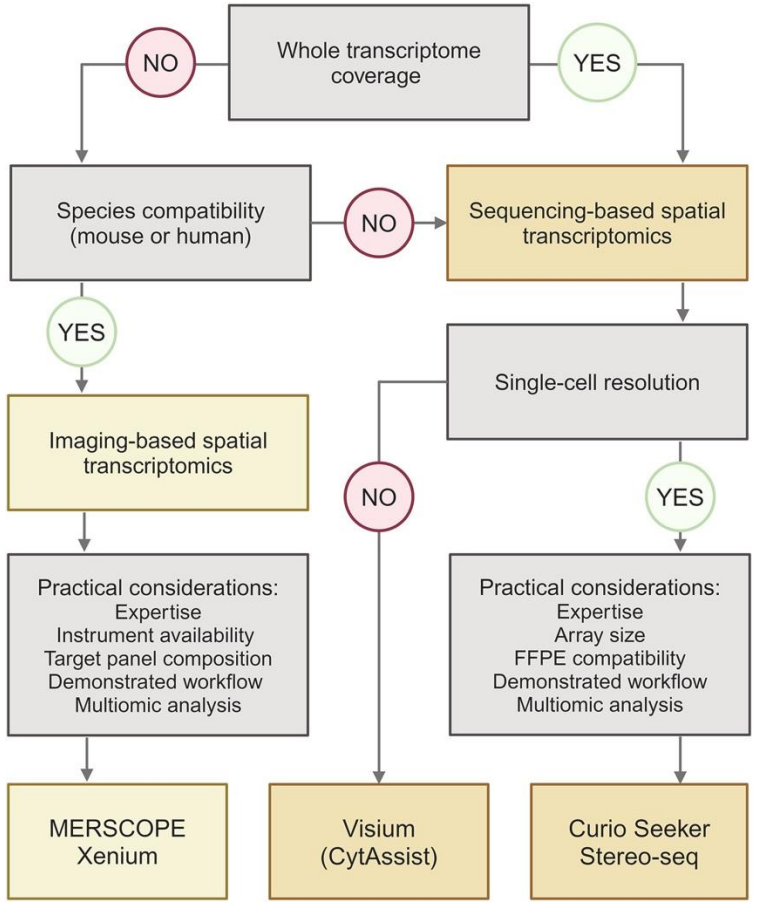


Other spatial -omics

- **DNA sequencing**: slide-DNA-seq *(Zhao et al., 2022)*
- **Epigenomics:** ATAC-seq (*Deng et al., 2022*) and CUT&TAG (*Deng et al., 2022)* both available with FlowGel assay by AtlasXomics
- **Proteomics:** CODEX *(Goltsev et al., 2018)*, Immuno-SABER *(Saka et al., 2019)*, scDVP *(Rosenberger et al., 2023)*, RIBOmap *(Zeng et al., 2023)*, CITE-seq *(Liu et al., 2023)*, Stereo-CITE-seq *(Liao et al., 2023)*
- **Metabolomics**: MSI *(Unsihuay et al., 2021)* and scSpaMet *(Hu et al., 2023)*
- **Temporally-resolved**: intMEMOIR *(Chow et al., 2021)*, Spatial iTracer *(He et al., 2022)*, Space-TREX *(Ratz et al., 2022)*
  - *Enable lineage tracing using CRISPR-Cas9 scarring or lentiviral barcoding!*

Experimental Considerations

1. Experimental design
   - Hypothesis generation or testing?
2. Target (gene or cell) detectability
   - Rare cells will be hard to identify with a low-res technology
   - Transcripts with low expression may be difficult to profile with a technology that has low capture efficiency
3. Tissue type and species
4. Availability of reference data
   - For cell deconvolution
5. Tissue section type and quality



Valihrach, L., Zucha, D., Abaffy, P., & Kubista, M. (2024). A practical guide to spatial transcriptomics. *Molecular aspects of medicine*, *97*, 101276. https://doi.org/10.1016/j.mam.2024.101276

What challenges remain?

- Compatibility with wider range of tissues
- Optimize data storage and processing speed
- Optimize spatial resolution and capture efficiency
- Reduce signal diffusion
- ST-specific normalization and batch correction
- Registration across slices
- Integration with multi-omics
- Increased throughput, reduced cost, greater accessibility