

Single cell genomics II

Genomics Bio5488

Guoyan Zhao

Associate Professor of Genetics & Neurology

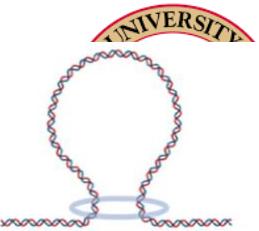




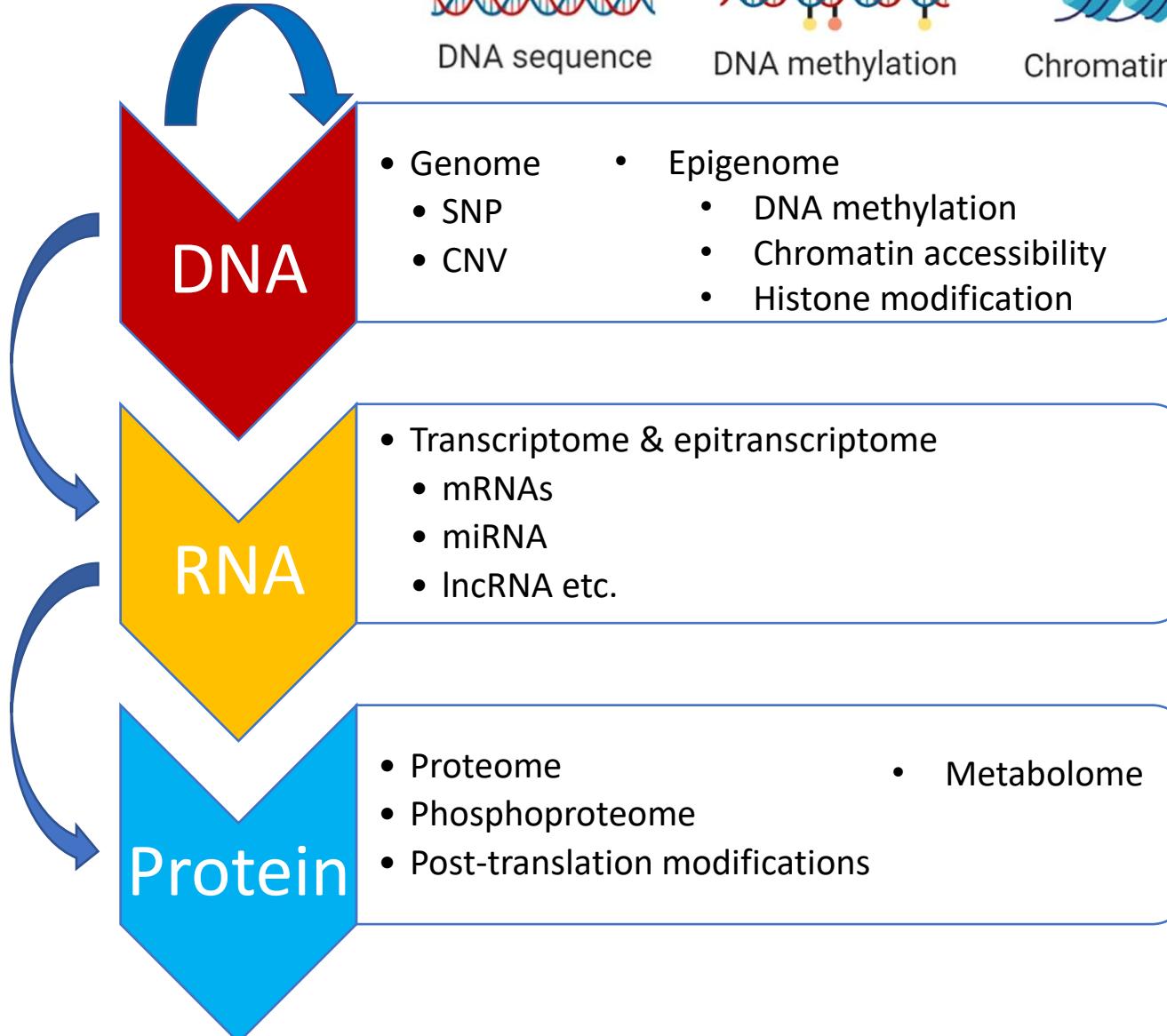
Single cell genomics

- Single cell genomics I
 - History of Single cell technology
 - Single cell RNA-sequencing (scRNA-seq) technology
 - Basic scRNA-seq data analysis workflow
 - Unlocking biological insights
 - Other unimodal single-cell technology
- Single cell genomics II
 - Single-cell multiomics
 - Transcriptome + Epigenome
 - Transcriptome + Protein
 - Transcriptome + CRISPR screening
 - Transcriptome + TCR/BCR
 - Transcriptome + Antigen specificity
 - Spatial genomics
 - Spatial transcriptomics
 - Spatial proteomics
 - Spatial multiomics
 - Spatial metabolomics
- Single cell genomics Lab
 - scRNA-seq data analysis

Unimodal single-cell genomics



Central dogma



Bulk version

RNA-seq
BS-seq
DNase-seq
ATAC-seq
MNase-seq
ChIP-seq
Cut&Tag
...



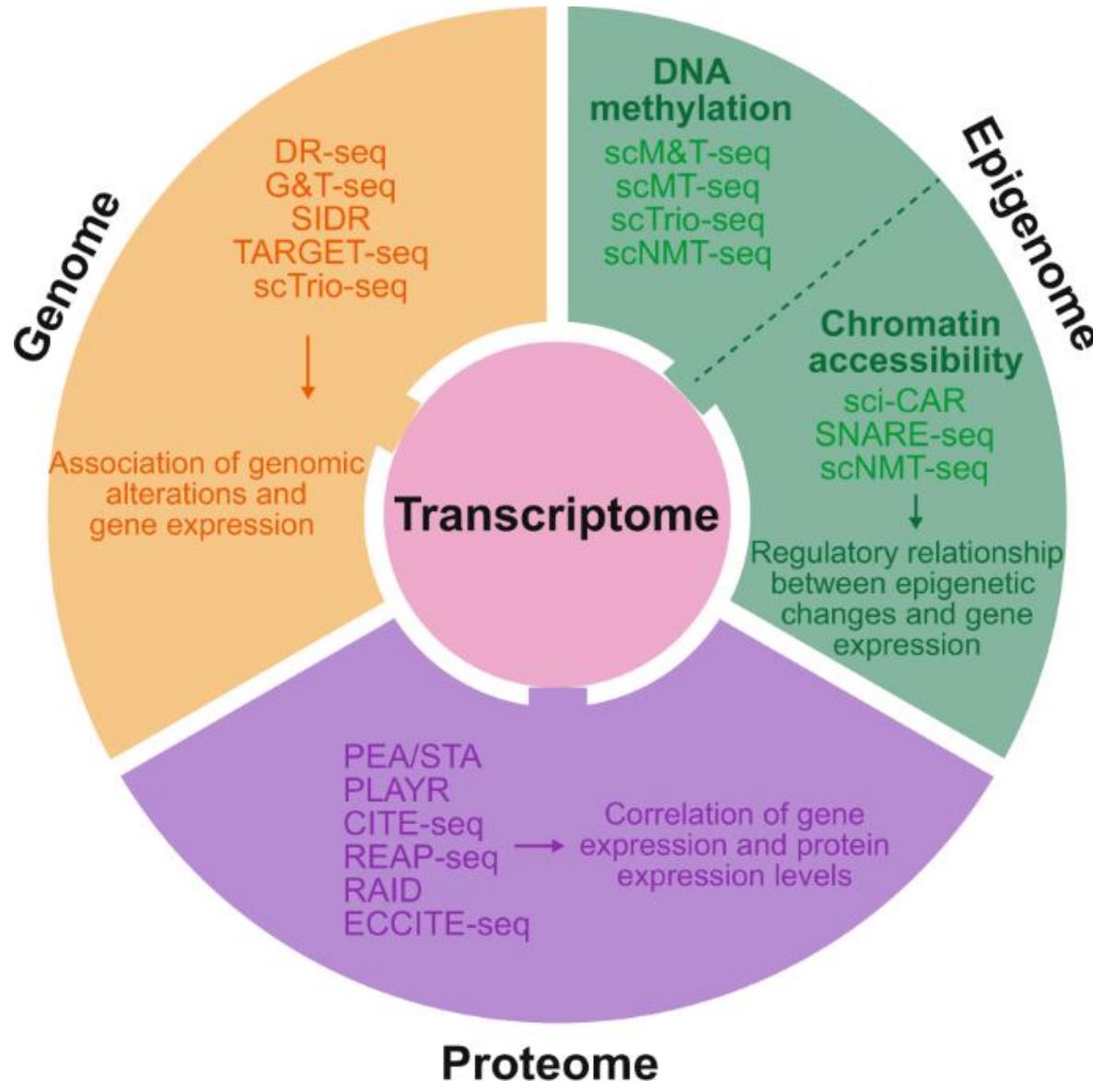
Single-cell version

Single-cell multi-omics

Combine ≥ 2 modalities in single-cell version from the same exact cell at the same time.

- 5mC
- 5hmC
- Open chromatin
- Nucleosome
- Nucleosome with modified histone
- Transcription factor
- RNA polymerase
- Cohesin

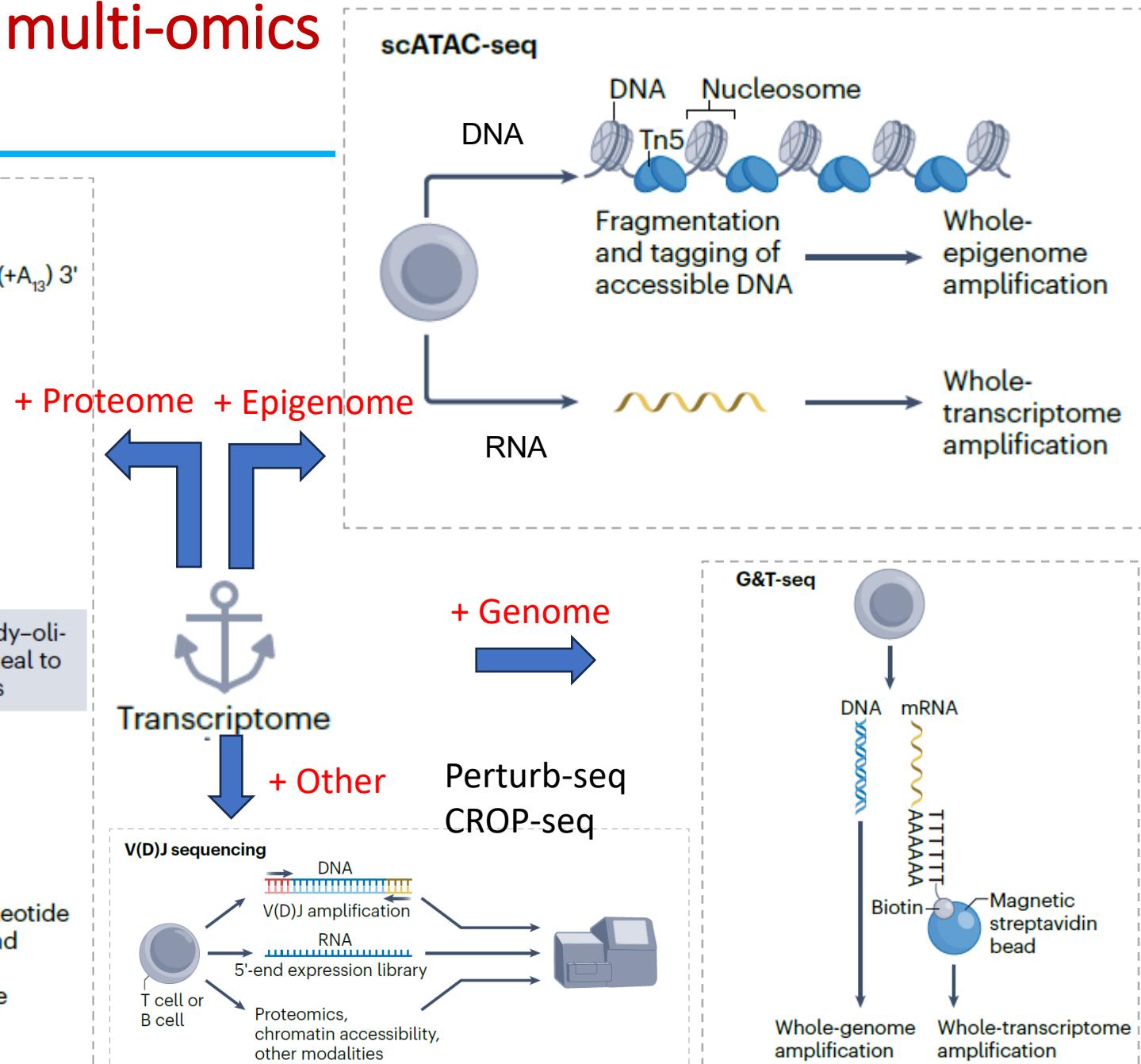
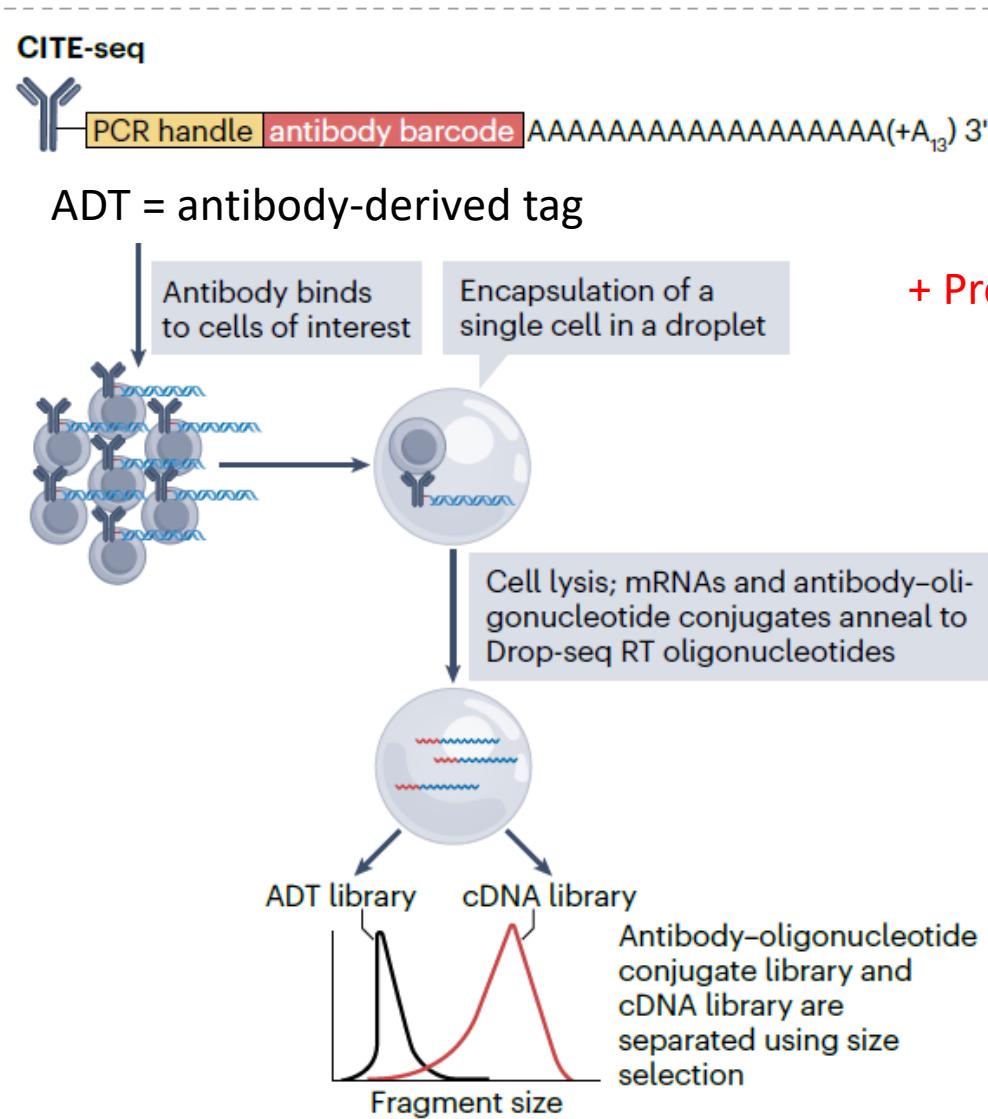
The landscape of multi-omics technologies



- optimization of throughput and resolution
- modality integration
- accuracy

- Single-cell multiomics: technologies and data analysis method, Lee et al., 2020 <https://www.nature.com/articles/s12276-020-0420-2>

Representative single-cell multi-omics technologies



Single cell tri-omics



- single-cell triple omics sequencing (scTrio-seq) :
 - genome, methylome and transcriptome (Hou et al. *Cell Res.* 2016)
- scNOMe-seq
 - nucleosome positioning, chromatin accessibility and DNA methylation (Pott et al., *eLife* , 2017)
- scNMT-seq and scNOMeRe-seq:
 - DNA methylation, chromatin accessibility and transcriptome (Clark et al. *Nat. Commun.*, 2018; Wang, Y. et al. *Nat. Commun.*, 2021).
- ATAC with select antigen profiling by sequencing (ASAP-seq) and DOGMA-seq
 - chromatin accessibility, gene expression and protein expression (Mimitou, E. P. et al. *Nat. Biotechnol.* 2021).

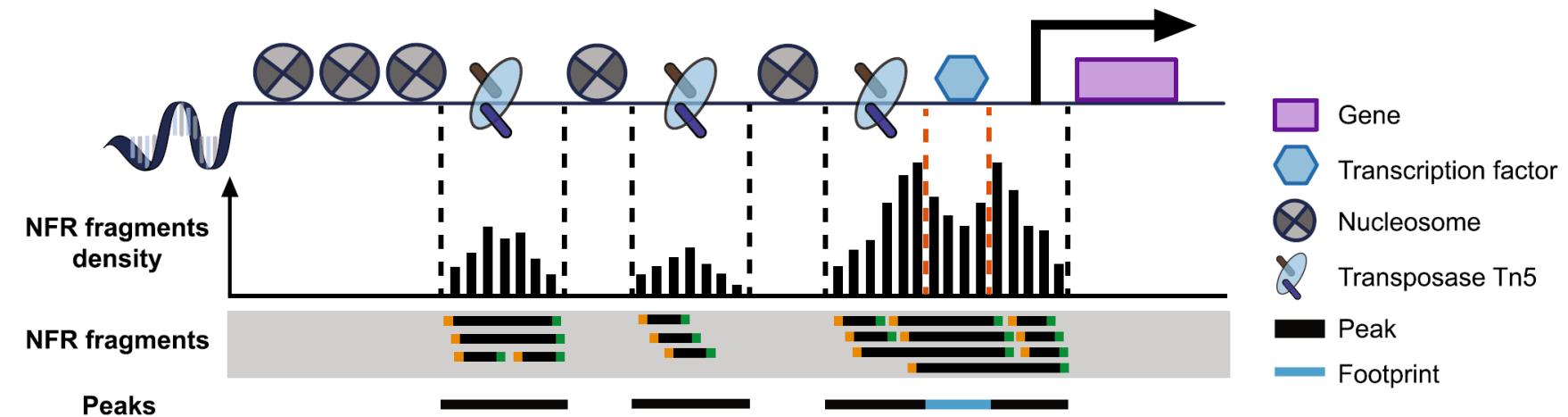
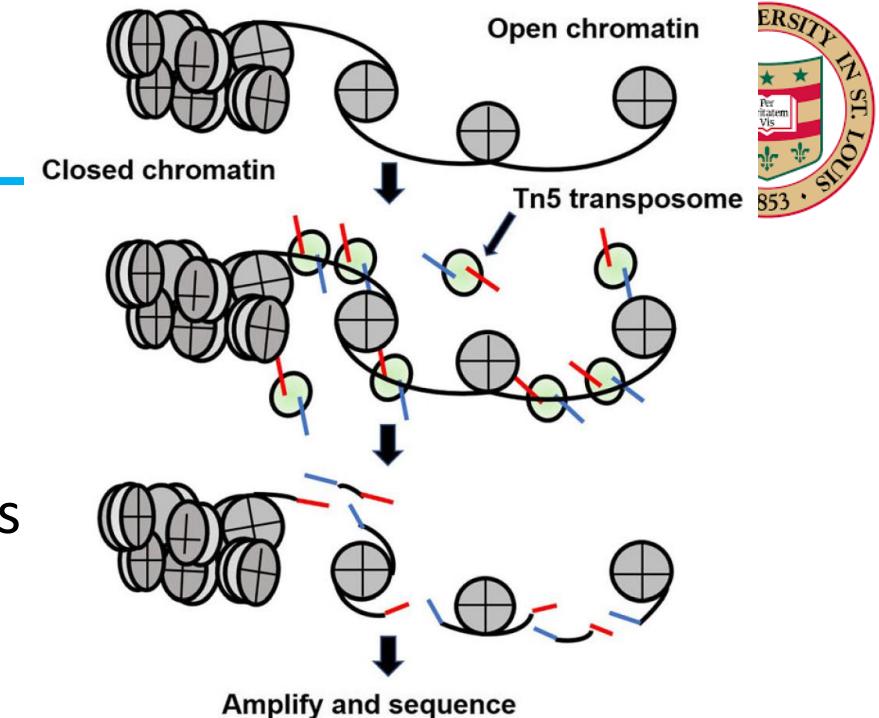


Advantages of single-cell Multi-omics

- More subpopulations may be identified, as different technologies may pick up different types of variations.
- Link the genome and its epigenetic regulation to gene and protein expression at the single-cell level.
 - How does variation in DNA sequence effect the epigenome?
 - How does variation in chromatin state affect gene expression?
 - Infer cause-effect relationships and, thus, the mechanisms behind a known phenotype.

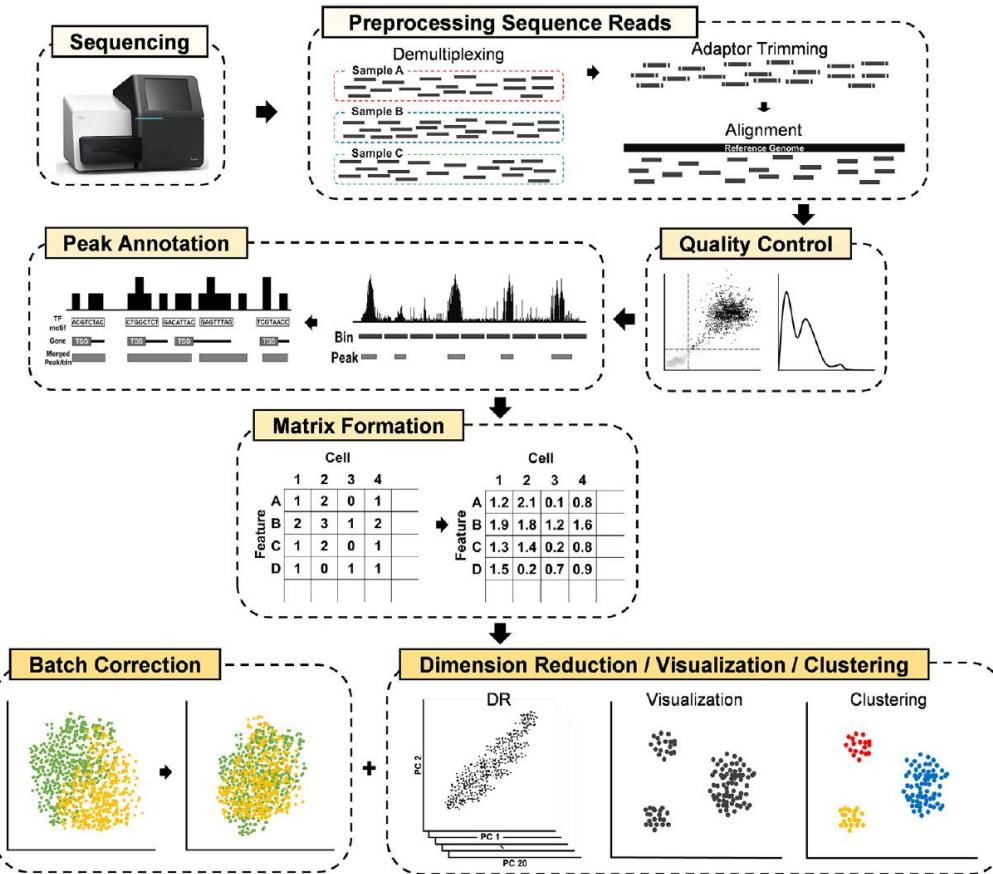
Single Cell ATAC-seq

- ATAC-seq: Assay for transposase-accessible chromatin using sequencing
- Uses a hyperactive Tn5 transposase to insert sequencing adaptors into accessible chromatin regions
- Measuring chromatin accessibility = potential regulatory sequences

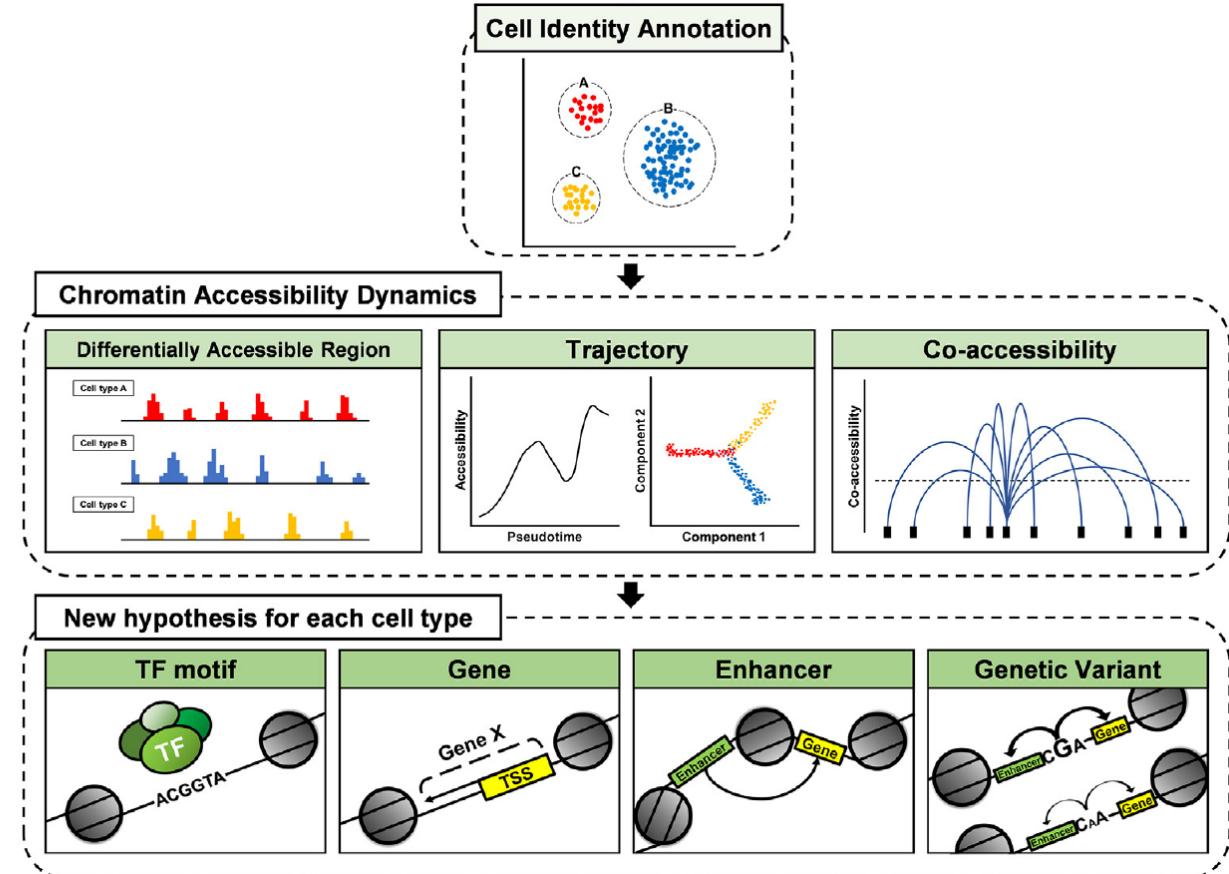


Single Cell ATAC-seq data analysis

Pre-processing



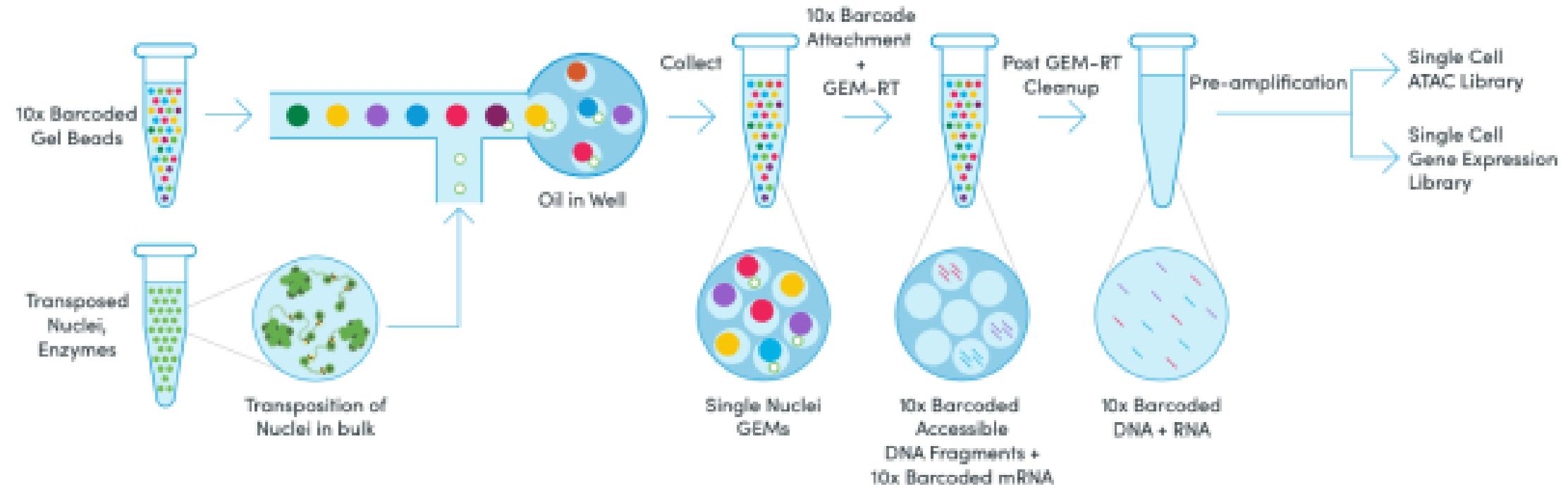
Downstream Analysis



Single Cell Multiome ATAC + Gene Expression



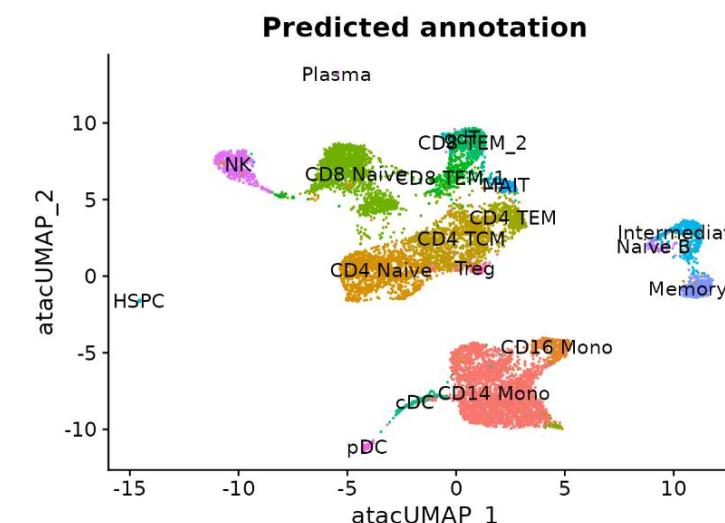
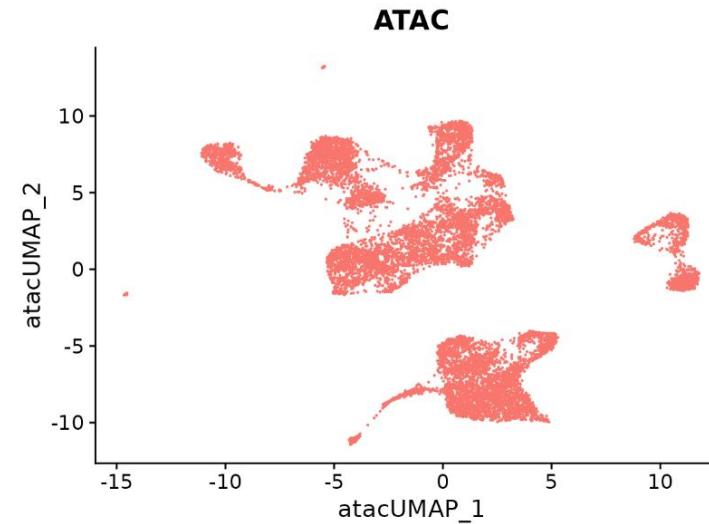
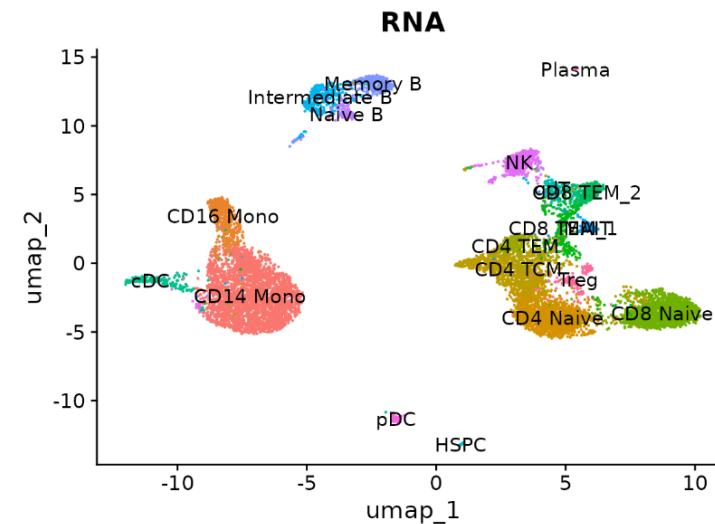
- Performs direct measurements of 3' gene expression and chromatin accessibility from the same cell.



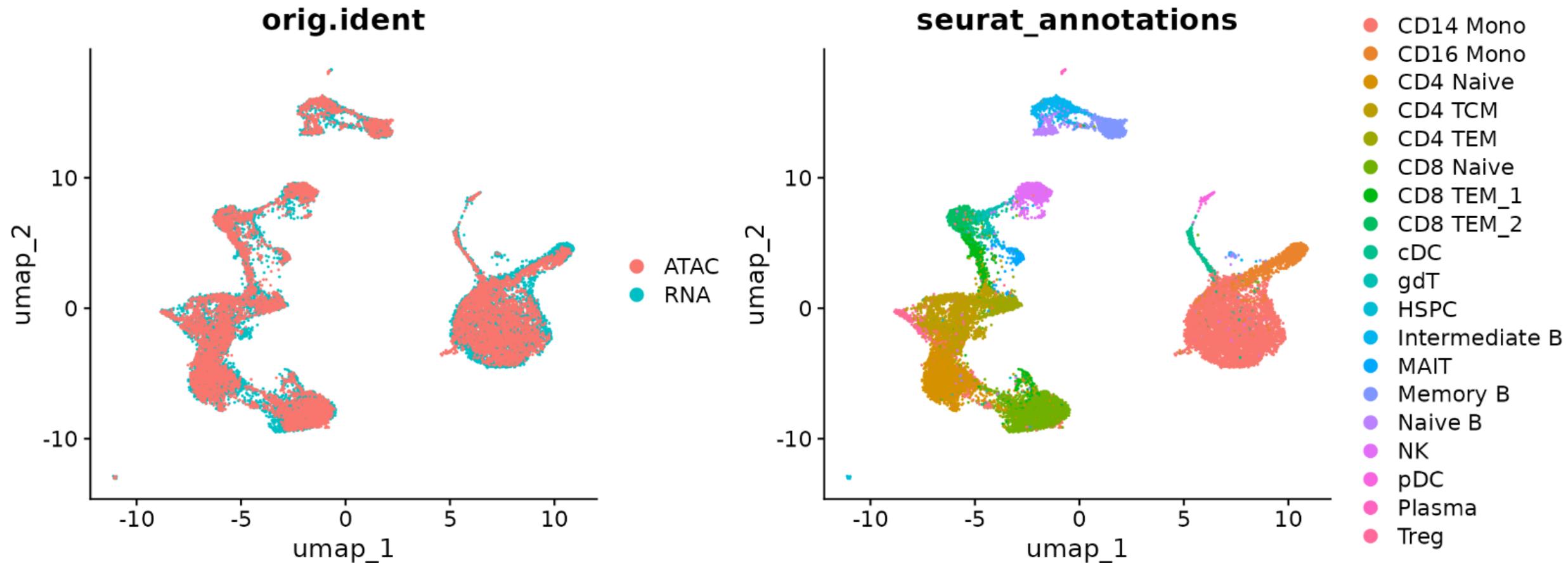
Single Cell Multiome – data analysis



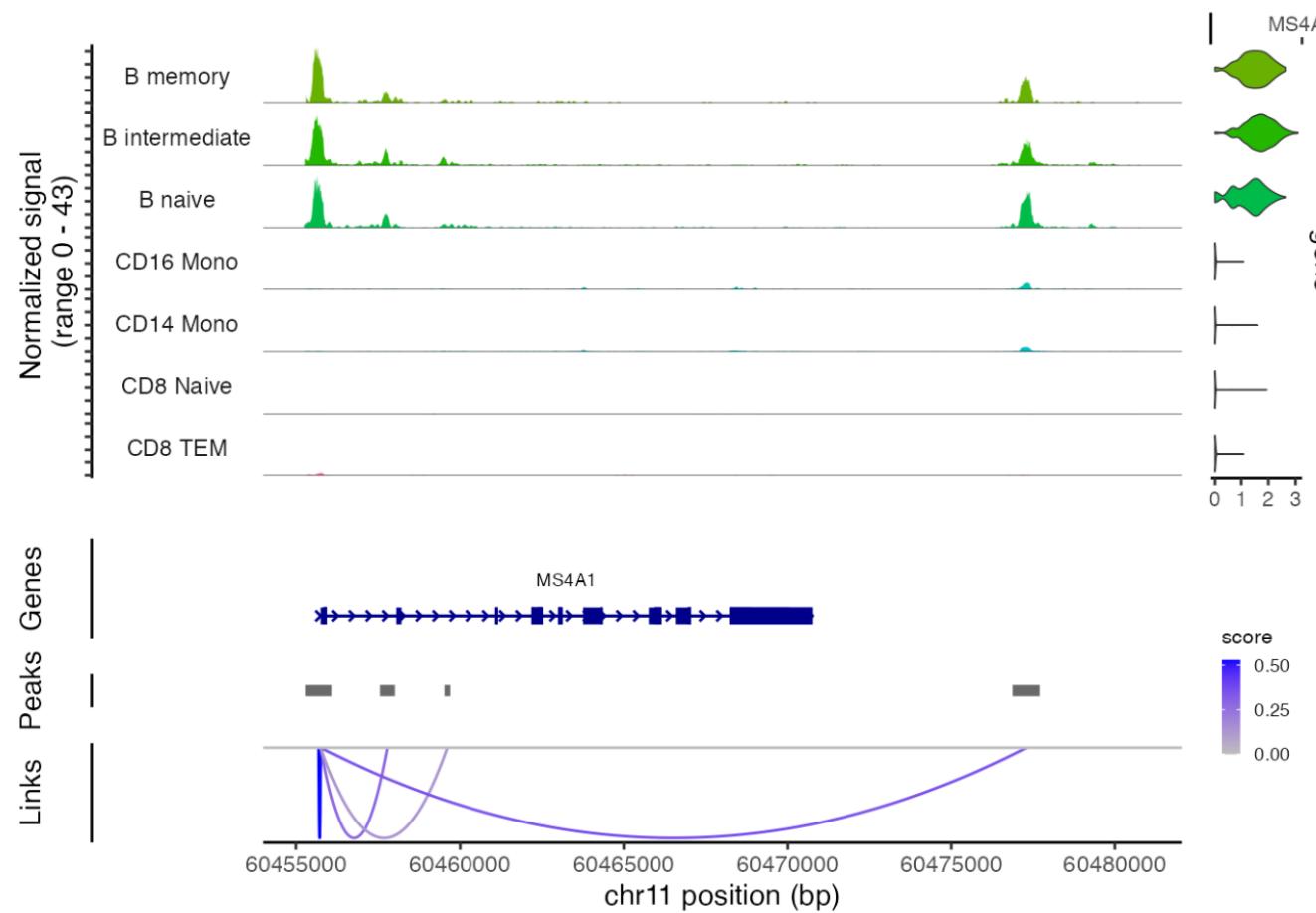
Annotate scATAC-seq cells via label transfer



Co-embedding scRNA-seq and scATAC-seq datasets



Linking peaks to genes

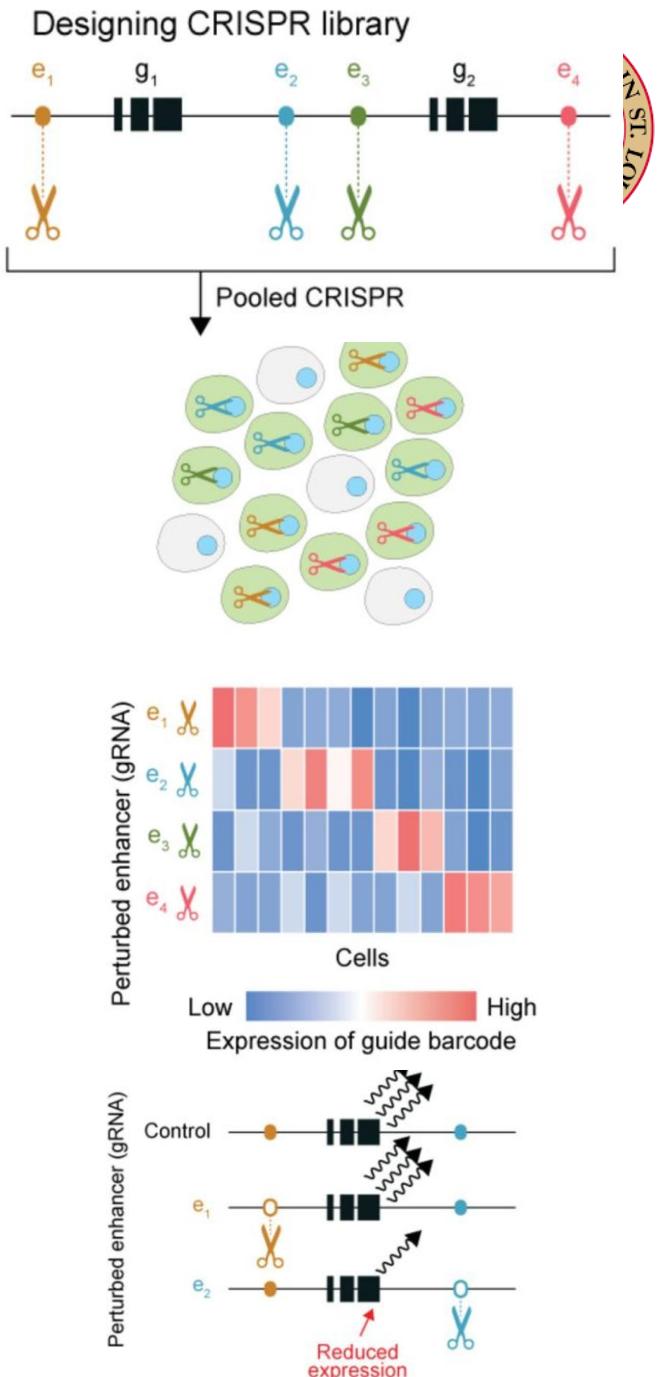


Application of scATAC-seq:

- **Gene Regulatory Network (GRN) Inference:** Linking transcription factor (TF) activity - target gene expression.
- **Functional Element Annotation:** Identifying enhancers and mapping them to target genes.
- **Disease Mechanism Analysis:** Identifying pathogenic mechanisms by linking regulatory element accessibility to gene expression changes.

CRISPR + single cell genomics

- Use CRISPR/Cas9 technology to knockout/edit genes or non-coding regions (enhancers)
 - infect pools of cells with viral constructs containing guide RNAs that target specific areas of the genome.
 - Use scRNA-seq to profile the transcriptome of each cell + the specific guide RNAs that were transduced
 - linking gene expression changes with the factor being manipulated
- Perturb-seq, CROP-seq
- provide cause-effect information

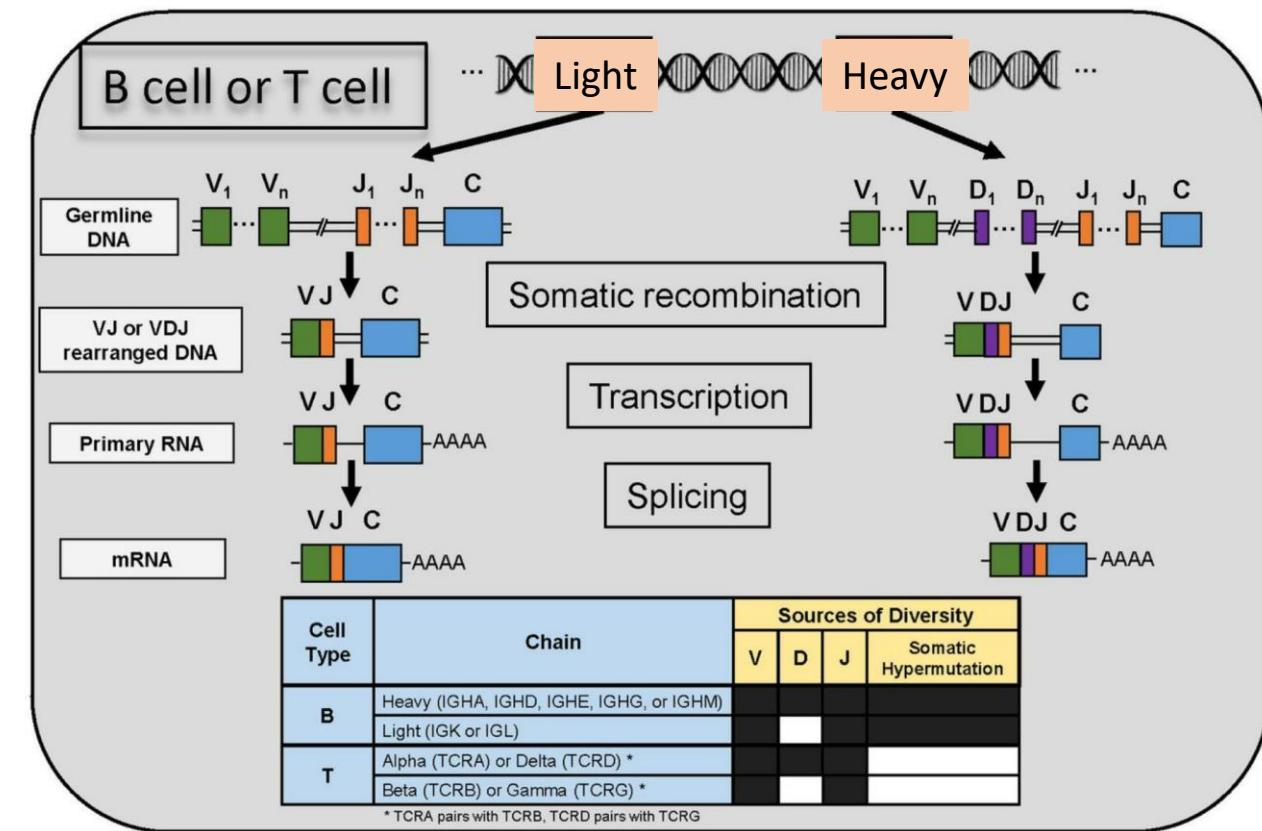


V(D)J recombination and Paired TCR sequences



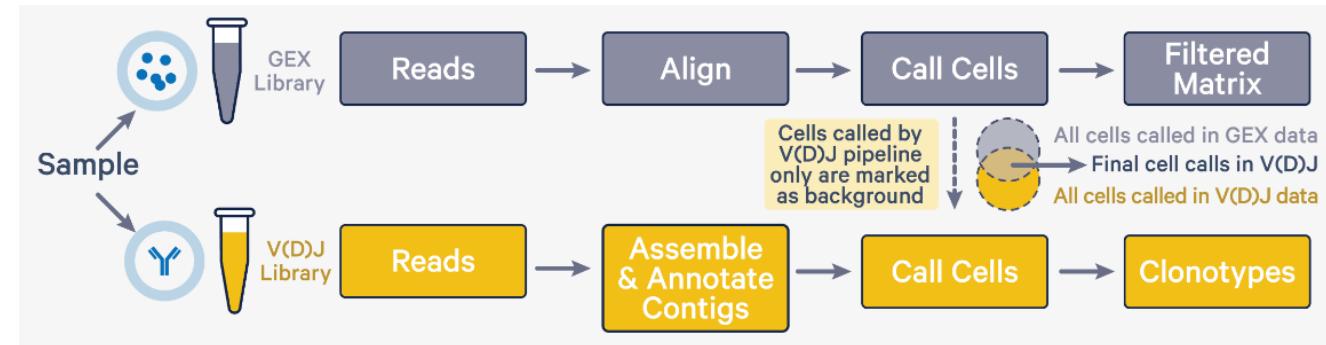
- Partition cells into GEMs. All cDNA generated in a single GEM share a common [10x Barcode](#).
- Perform enrichment PCR targeting the 5' end to the C-region
- followed by enzymatic fragmentation results in a pool of molecules originating from the same transcript.
- The molecules carry the same [10x Barcode](#) and [UMI](#) sequences, but with different insert lengths, resulting in different sequence start points.
- The diversity of start points gives complete coverage of the targeted portion of each V(D)J transcript, which is typically ~650 bp.

V(D)J Recombination, the somatic recombination of Variable (V), Diversity (D), and Joining (J) gene sequences

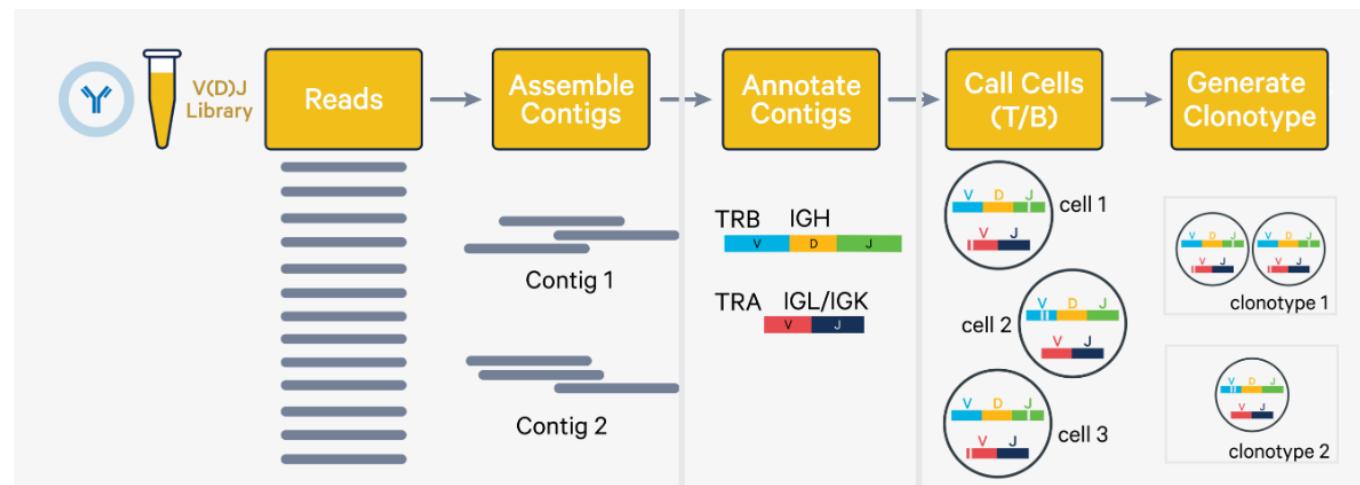


Single Cell Immune Profiling – data analysis

- Cell calling



- V(D)J contig annotation



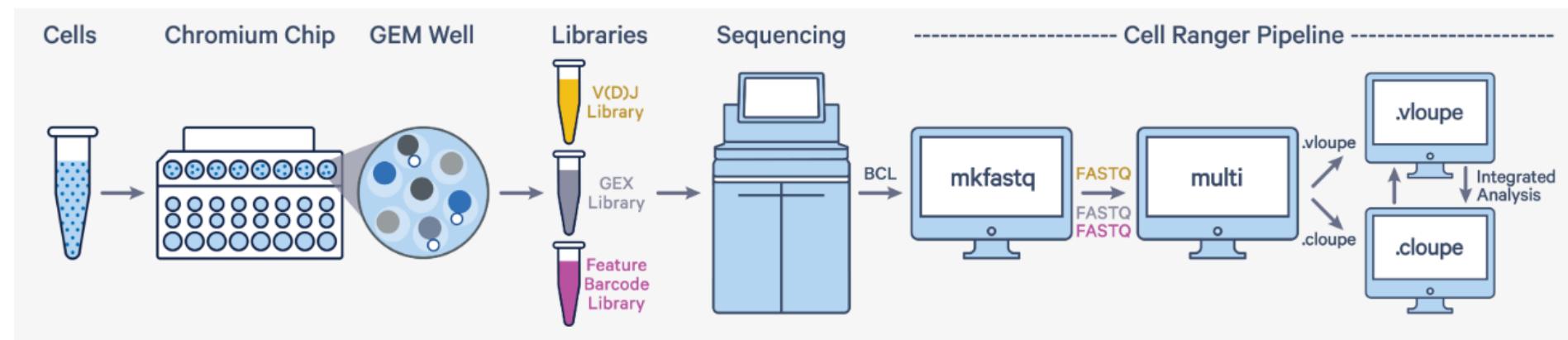
TRA pair with TRB
TRD pair with TRG

- Determine clone expansion

Single Cell Immune Profiling – 5' Chromium Next GEM Single Cell Immune Profiling Solution



- Simultaneous analysis of following libraries at single cell resolution for the same set of cells:
 - V(D)J transcripts and clonotypes for B or T cells
 - 5' Single Cell Gene Expression
 - Cell surface proteins (Antibody Capture)
 - Barcode Enabled Antigen Mapping (BEAM™) (Antigen Capture)
 - CRISPR Guide Capture



Commercially Single Cell Multiome technology



Transcriptome + Epigenome
Transcriptome + Protein
Transcriptome + CRISPR screening
Transcriptome + TCR/BCR
Transcriptome + Antigen specificity



Transcriptome + Protein
Transcriptome + Epigenome
Transcriptome + TCR/BCR

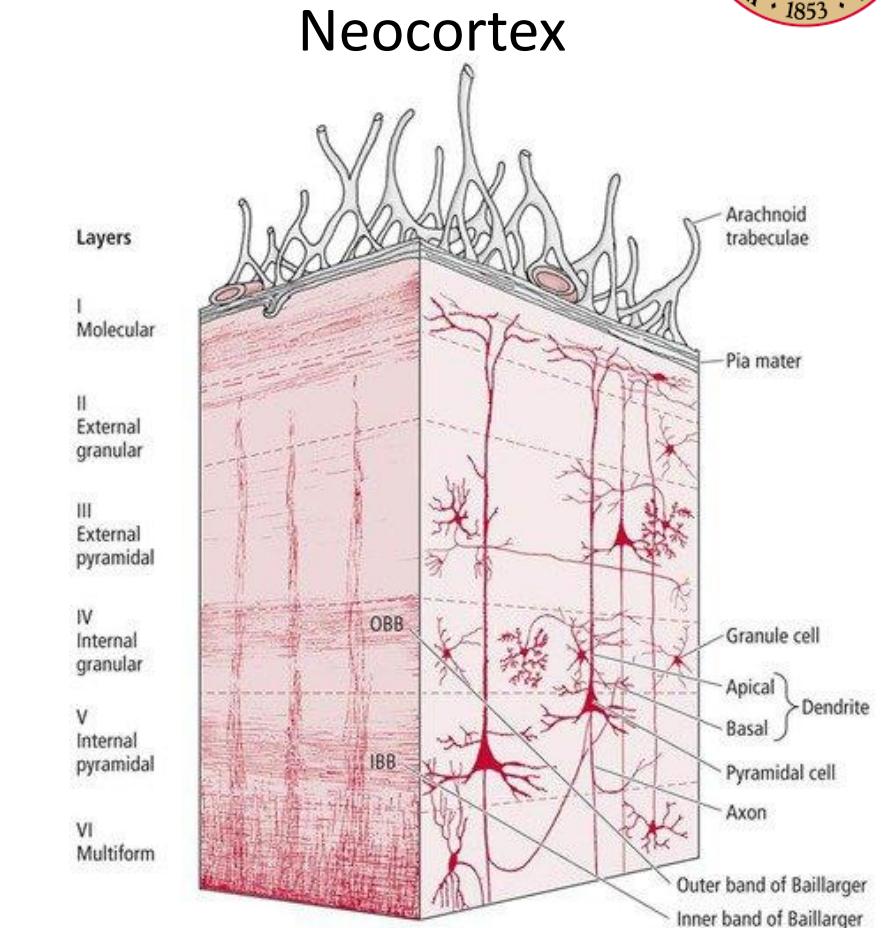


Tapestri

DNA + Protein

Spatial genomics

- Above-described methods lose the spatial context.
- The position of any given cell, relative to its neighbors and non-cellular structures, can provide helpful information for defining cellular phenotype, cell state, and ultimately cell and tissue function.
- Provide spatially resolved, high-dimensional assessment of transcripts, proteins, or metabolites



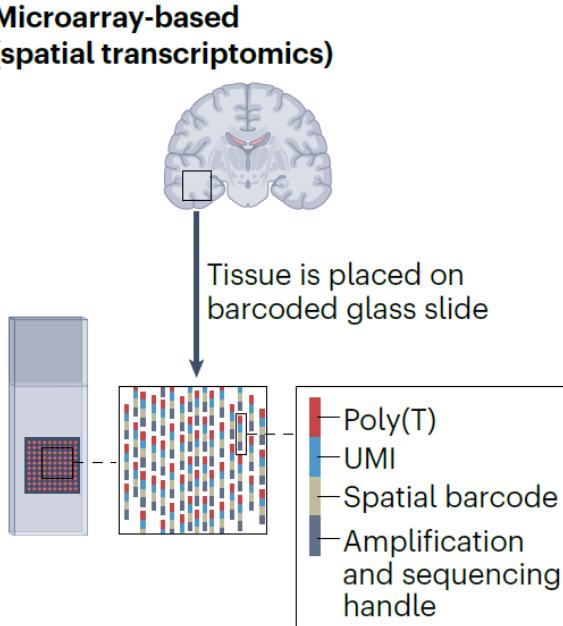
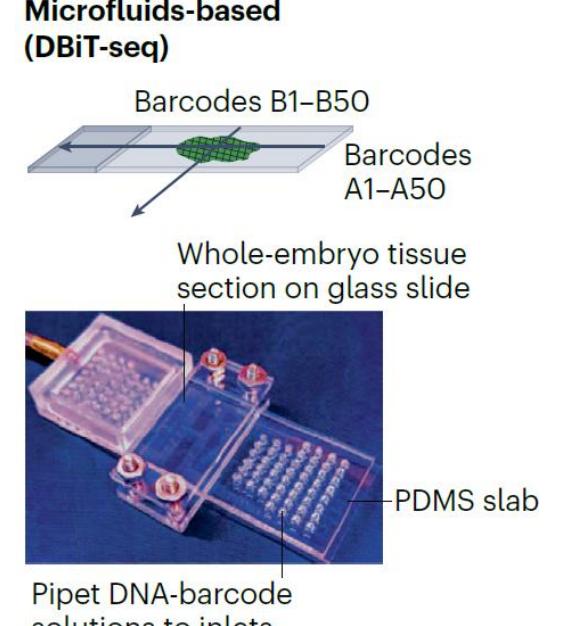
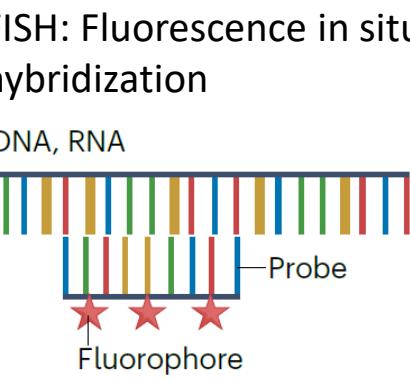
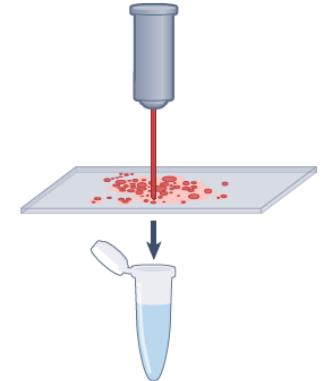
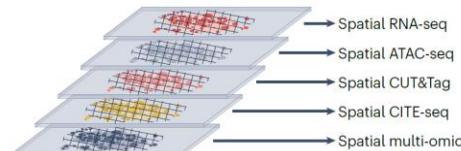
The cerebral cortex forms six layers with each layer having distinct cell types.

Diverse forms of spatial molecular technologies (not necessary at single cell resolution)

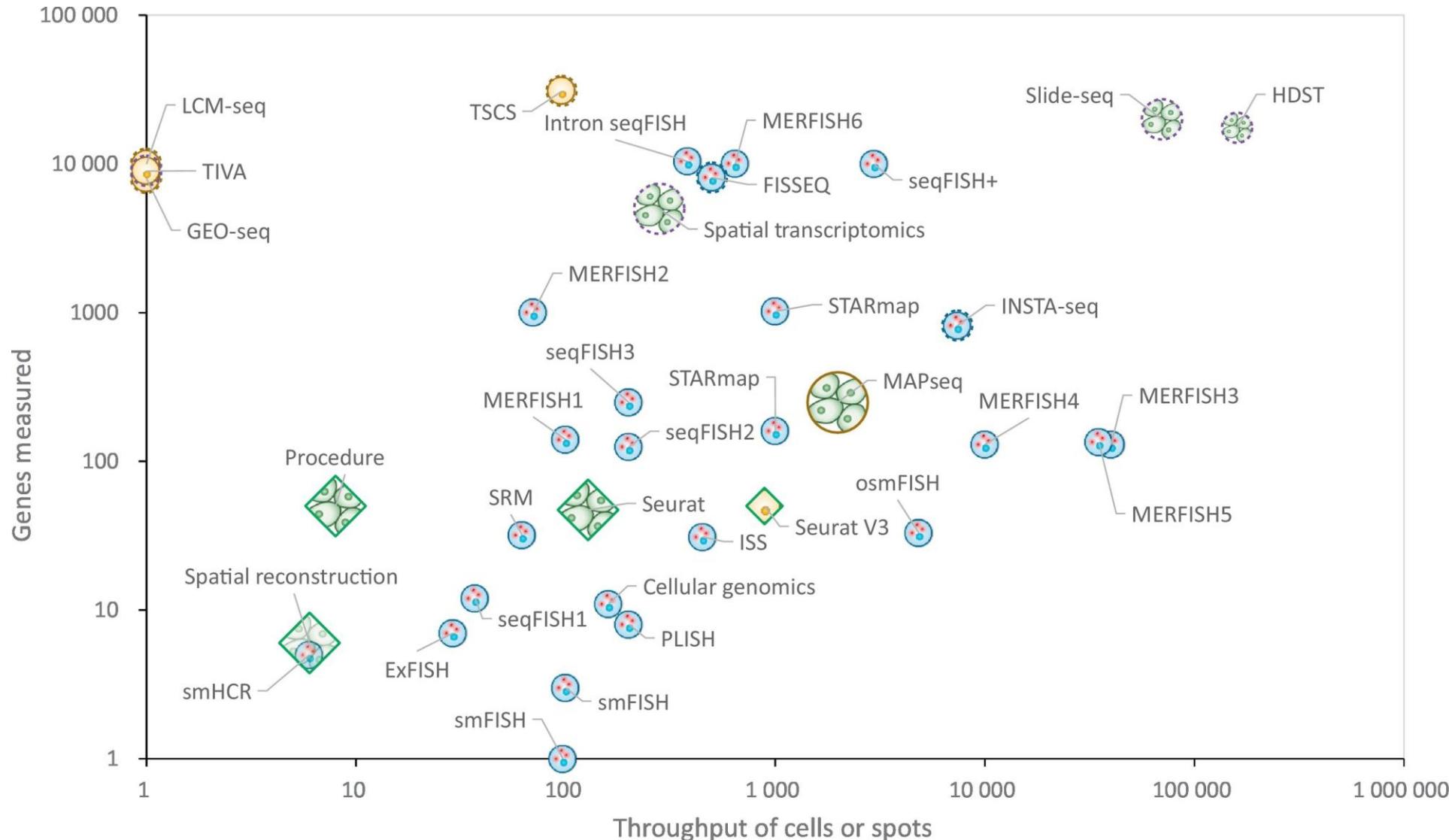


- Resolution
 - few cells per observation to subcellular resolution
- Multiplexing
 - dozens of features to genome-wide expression profiles
- Modality
 - transcriptomics
 - proteomics
 - metabolomics
- Technology
 - Imaging-based technologies
 - Sequencing-based technologies
- often with an associated high-content image of the captured tissue

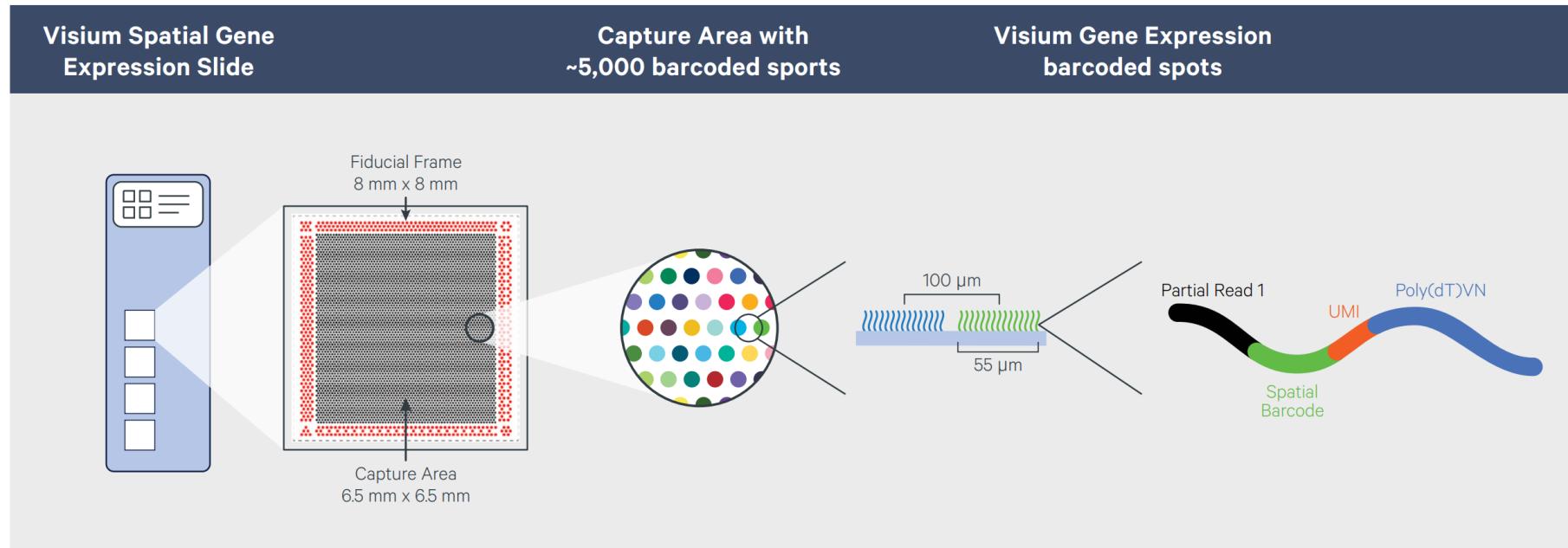
Spatial transcriptomics technologies

	Sequencing-based	Imaging-based	LCM-based	
Microarray-based (spatial transcriptomics)	 <p>Tissue is placed on barcoded glass slide</p> <ul style="list-style-type: none"> Poly(T) UMI Spatial barcode Amplification and sequencing handle 			
Microfluidics-based (DBiT-seq)		 <p>Barcodes B1-B50</p> <p>Barcodes A1-A50</p> <p>Whole-embryo tissue section on glass slide</p> <p>PDMS slab</p> <p>Pipet DNA-barcode solutions to inlets</p>		
FISH: Fluorescence in situ hybridization		 <p>DNA, RNA</p> <p>Probe</p> <p>Fluorophore</p>		
LCM-based (Geo-seq)				
Benefits	High spatial resolution, little specialized equipment, unbiased	Co-mapping capability, high resolution, high genes/pixel	High capture efficiency, subcellular resolution	Preservation of tissue morphology, quick, high resolution
Limitations	Low capture efficiency, low resolution compared to FISH 10x Visium (55 µm) Slide-seqV2 (10 µm) 10x Visium HD (2 µm) Stereo-seq (0.5 µm)	Near single-cell resolution, tissue size is limited	Readout limited to targeted genes, marker gene count, transcript length MERFISH (Vizigen) seqFISH+ CosMx SMI 10x Xenium	Costly, sample quality is a limitation co-profiling of mRNAs, microRNAs, DNA methylation and protein expression
	 <p>Spatial RNA-seq</p> <p>Spatial ATAC-seq</p> <p>Spatial CUT&Tag</p> <p>Spatial CITE-seq</p> <p>Spatial multi-omics</p>			

Throughput, sensitivity and resolution



Sequencing-based technologies - Visium



- enables whole transcriptome analysis of entire tissue sections
- protein co-detection
- Not single-cell resolution
 - Each barcoded spot captures the transcripts from 1-10 cells.
- relatively low sensitivity
- high cost
- labor intensive process



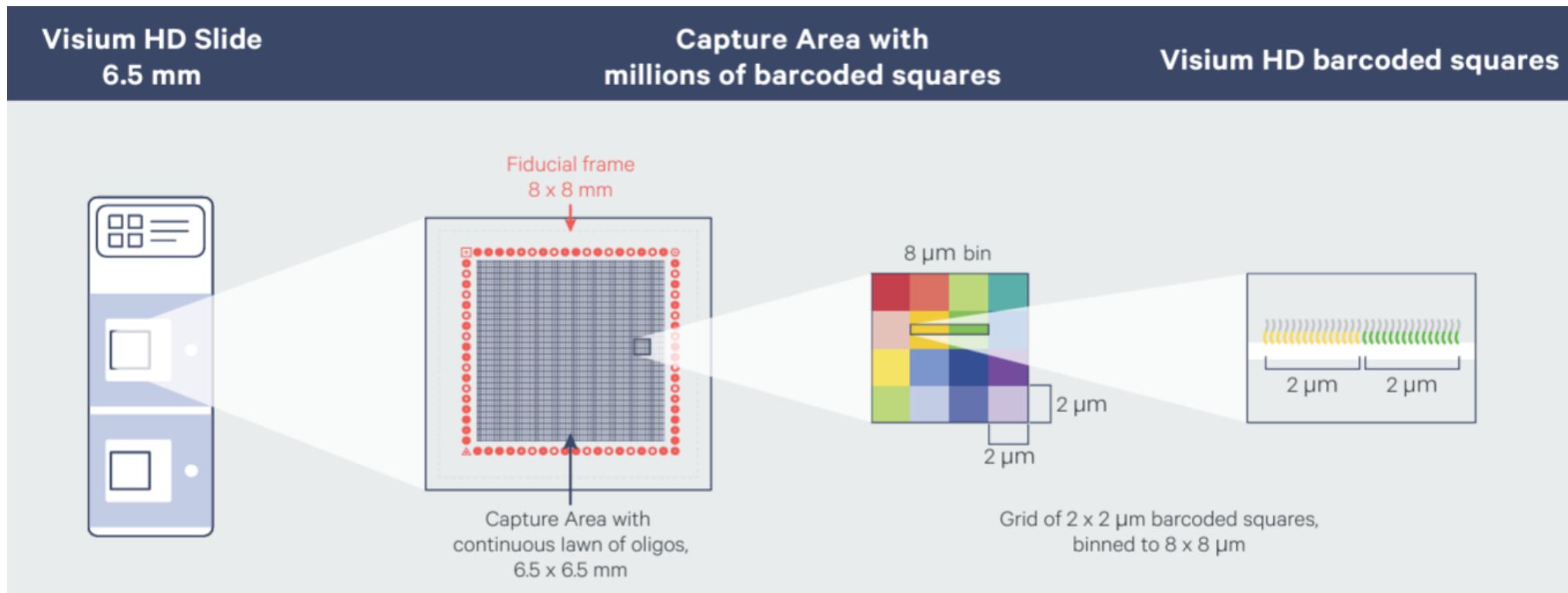
Visium Gene and Protein Expression co-detection

- Enables protein and whole transcriptome RNA mapped together in a single experiment from a single tissue section



- a pre-validated, 35-plex antibody panel optimized for use on human FFPE tissues.

Sequencing-based technologies - Visium HD



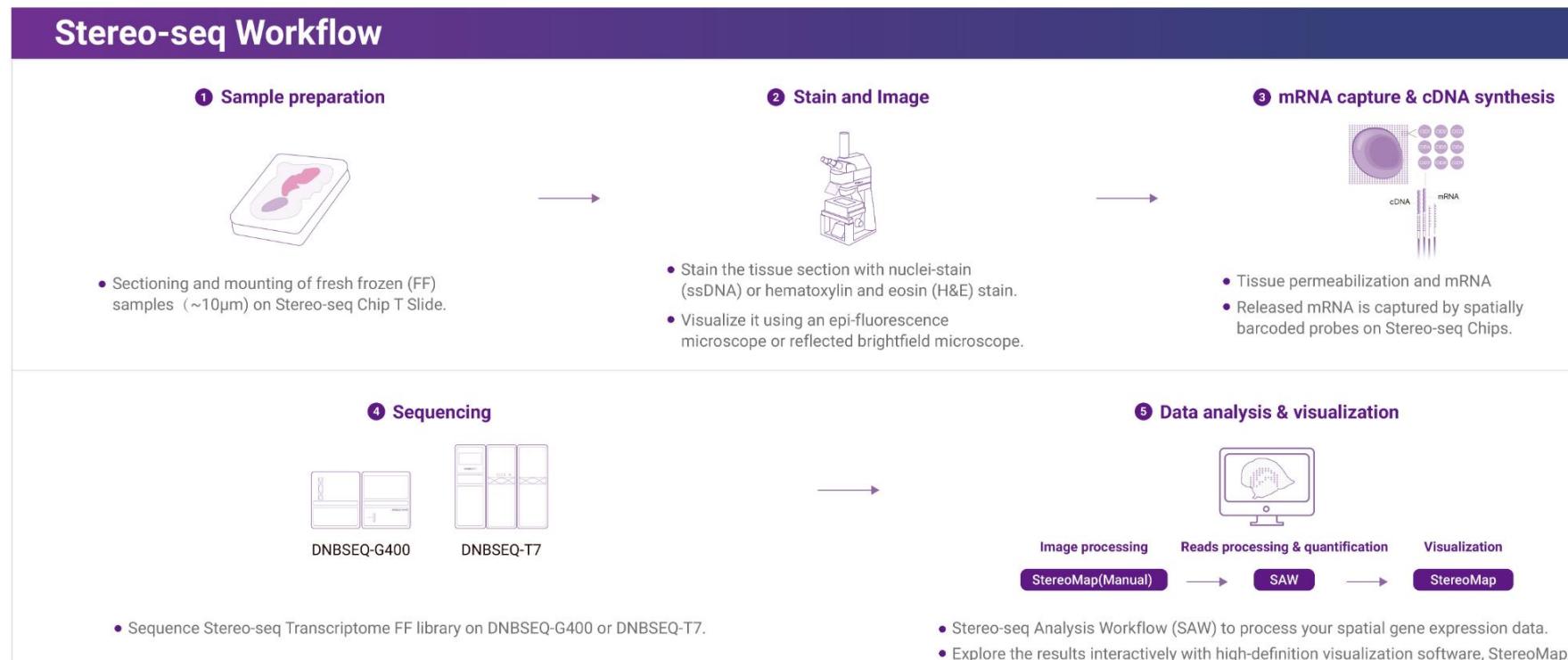
- uses a panel of predesigned probes to target 18,085 human genes or 19,405 mouse genes
- ~11 million $2 \times 2 \mu\text{m}$ barcoded squares without gaps.
- data output at $2 \mu\text{m}$, as well as multiple bin sizes.
- $8 \times 8 \mu\text{m}$ bin: the recommended starting point for visualization and analysis.
- a recommended minimum depth of 275 million read pairs for Capture Areas covered fully by tissue

Sequencing-based technologies - Stereo-seq

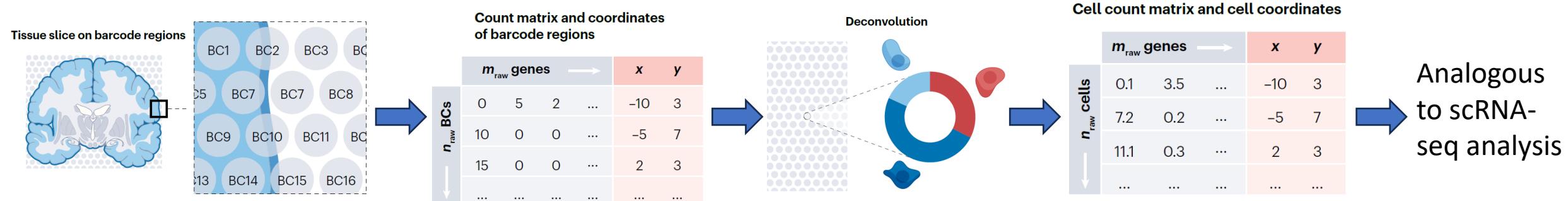
STomics



- Stereo-seq v1.3, with poly(dT)-based capture method
- enables unbiased transcriptome exploration
- ideal for studying diverse biological systems
- Resolution of 0.5 μ m, require higher sequencing depth



Sequencing-based technology - data analysis



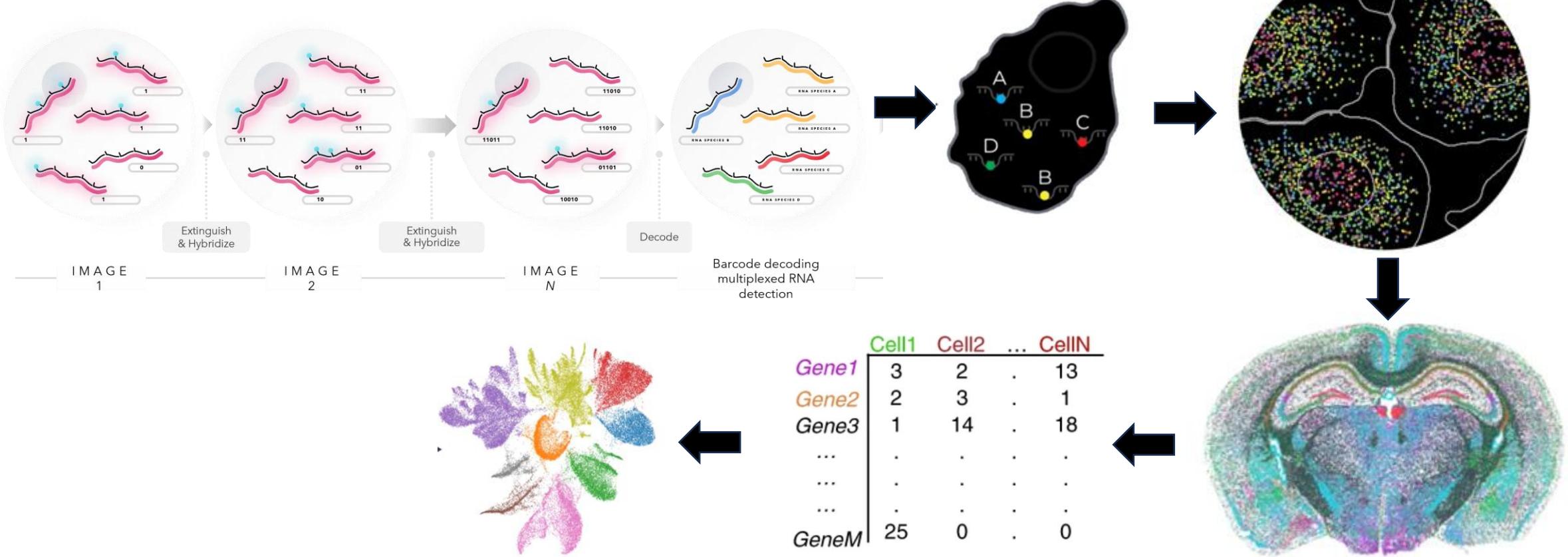
- do not capture single-cell resolution, the gene expression profile of spots reflects cell-type composition rather than distinct cell types.
- estimates the cell-type composition per spot based on the gene expression profile of the cell populations in a single-cell-resolved reference: Cell2location, SpatialDWLS, RCTD

Imaging-based technologies - MERSCOPE

VIZgen



- MERFISH: multiplexed error-robust fluorescence *in situ* hybridization, Chen et al., 2015 at Xiaowei Zhuang lab.
- fluorescent-labeled probes + combinatorial barcode + sequential hybridization and imaging



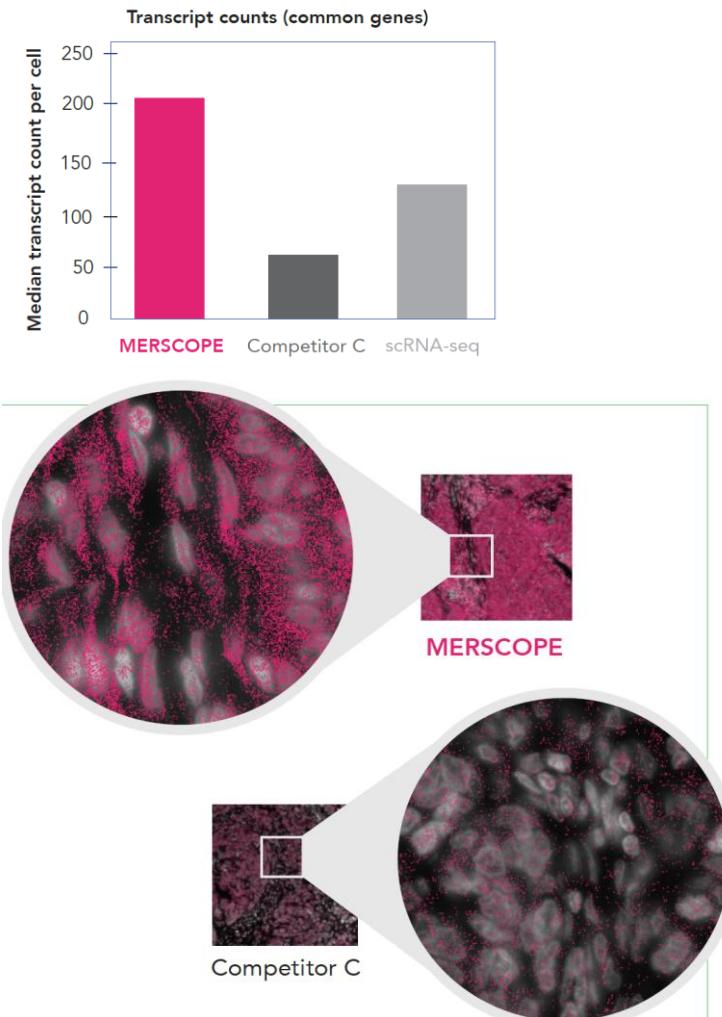


Establishing Data Quality Standards

Key Quality Metric	Sensitivity	Specificity	Information Density	Effective Multiplexing Capacity
Central Question	How many transcripts of each targeted RNA species are detected?	How many transcripts are identified correctly?	How many transcripts can the technology identify in a cellular volume?	How many different genes can the technology accurately profile at once?
Biological Significance	Many biologically relevant transcripts are expressed at low copy numbers.	Lower noise allows more transcriptomic variations to be discovered and stand out above the noise.	Characterizing subtle variations between individual cells requires sufficient information about each cell.	Measuring more biomarkers better characterizes the complexity inherent to a biological system.

Sensitivity

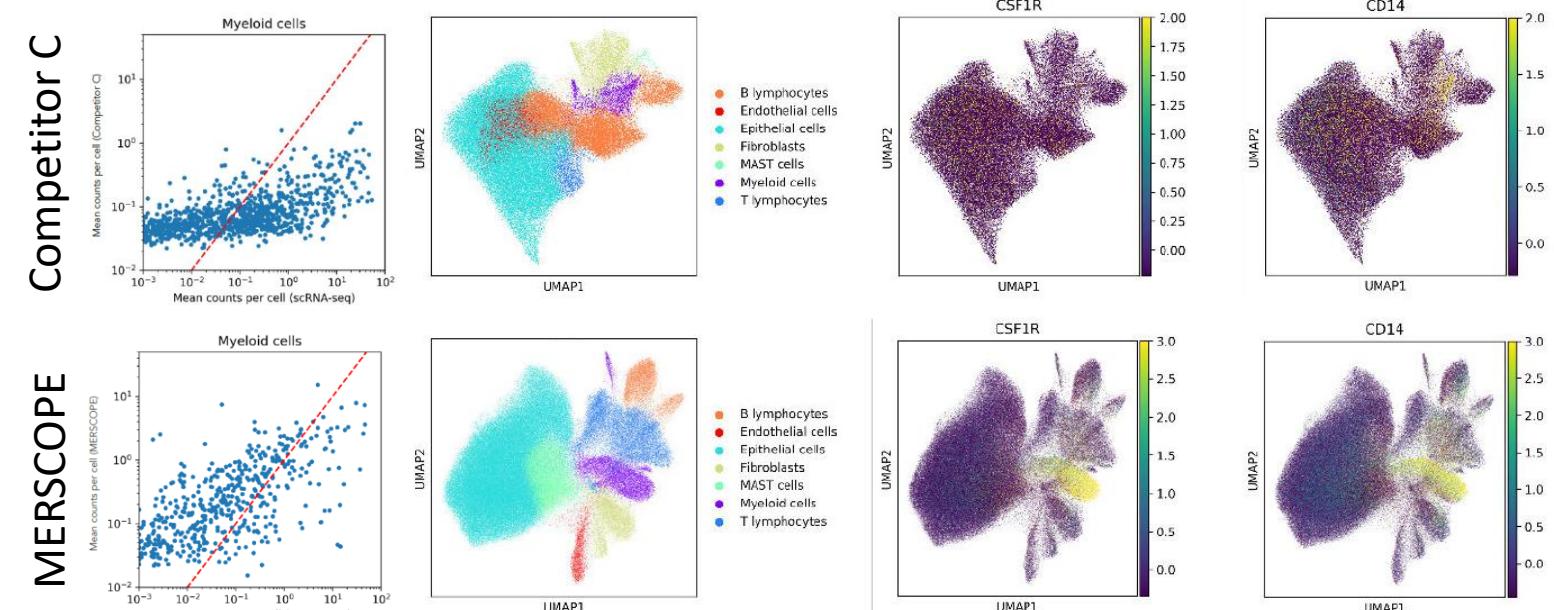
- detection efficiency when compared to the total number of expressed transcripts.



Specificity

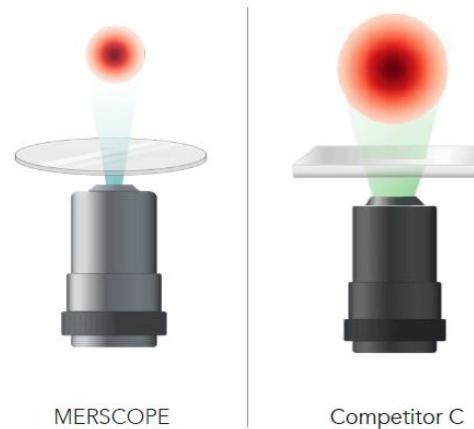
The fraction of reported transcripts that correspond to a true transcript within the biological sample.

- errors from autofluorescence in the tissue
 - Incomplete probe binding
 - molecular crowding.

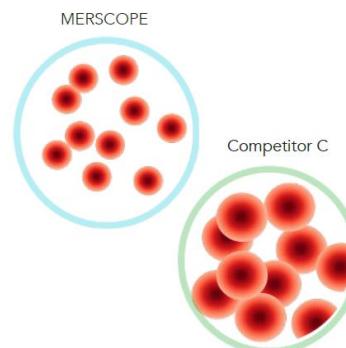


Information Density

The amount of information that can be measured within a given tissue volume or area



MERSCOPE optics are optimized for single molecule detection.



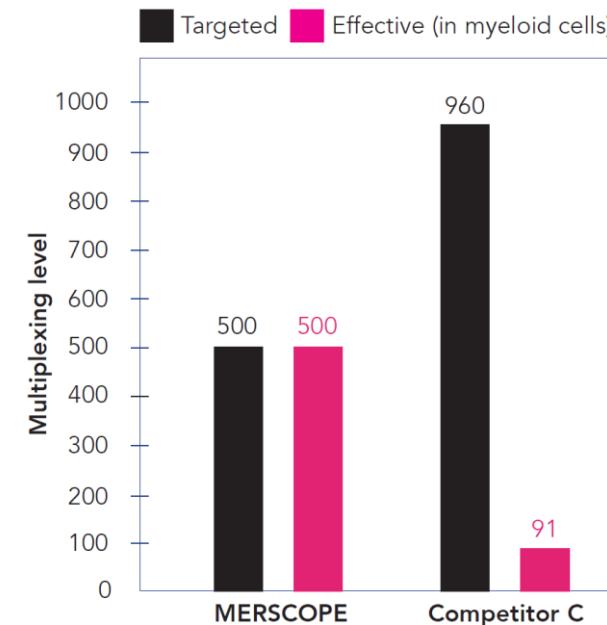
Effective Multiplexing Capacity



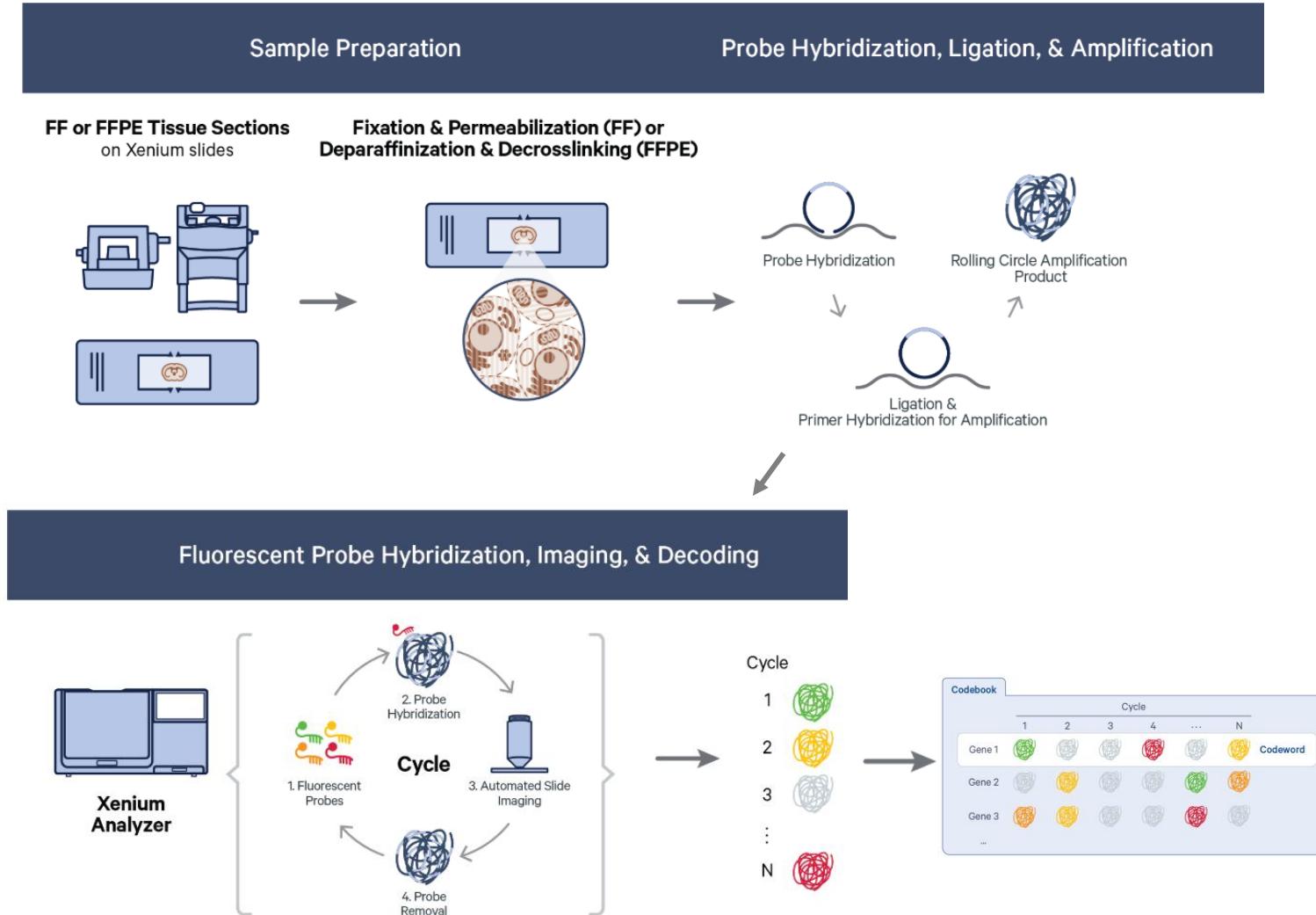
The number of RNAs a spatial genomics technology can detect in a single experiment.

- true multiplexing capacity determined by
 - targeted multiplexing capacity
 - the noise floor (ambient background signal)

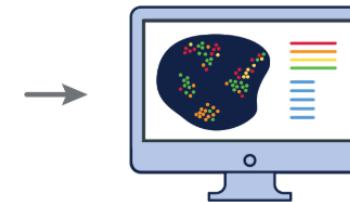
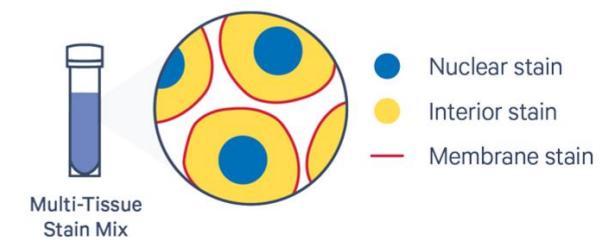
effective vs targeted multiplexing capacity



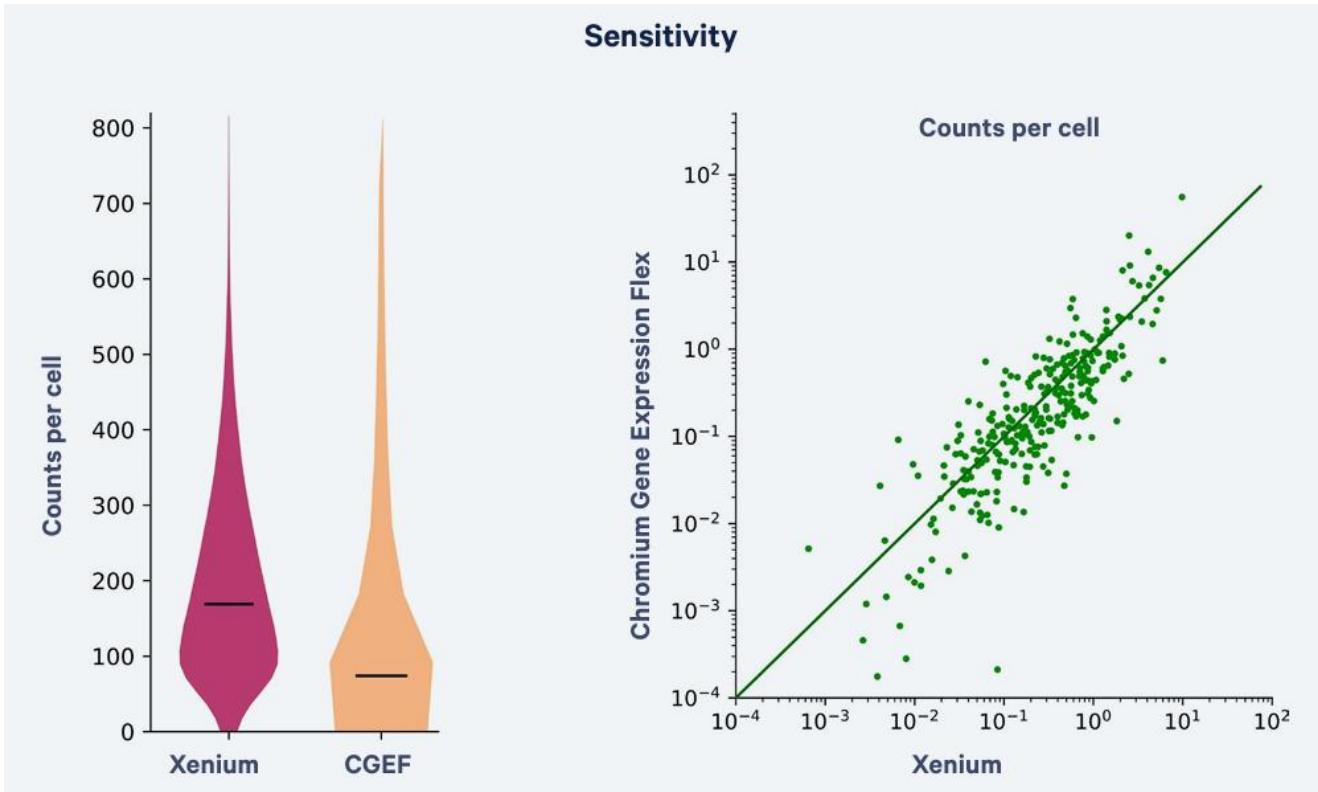
Imaging-based technologies – Xenium



- up to 5,000 genes
 - 472 mm² of tissue per run
 - multimodal segmentation

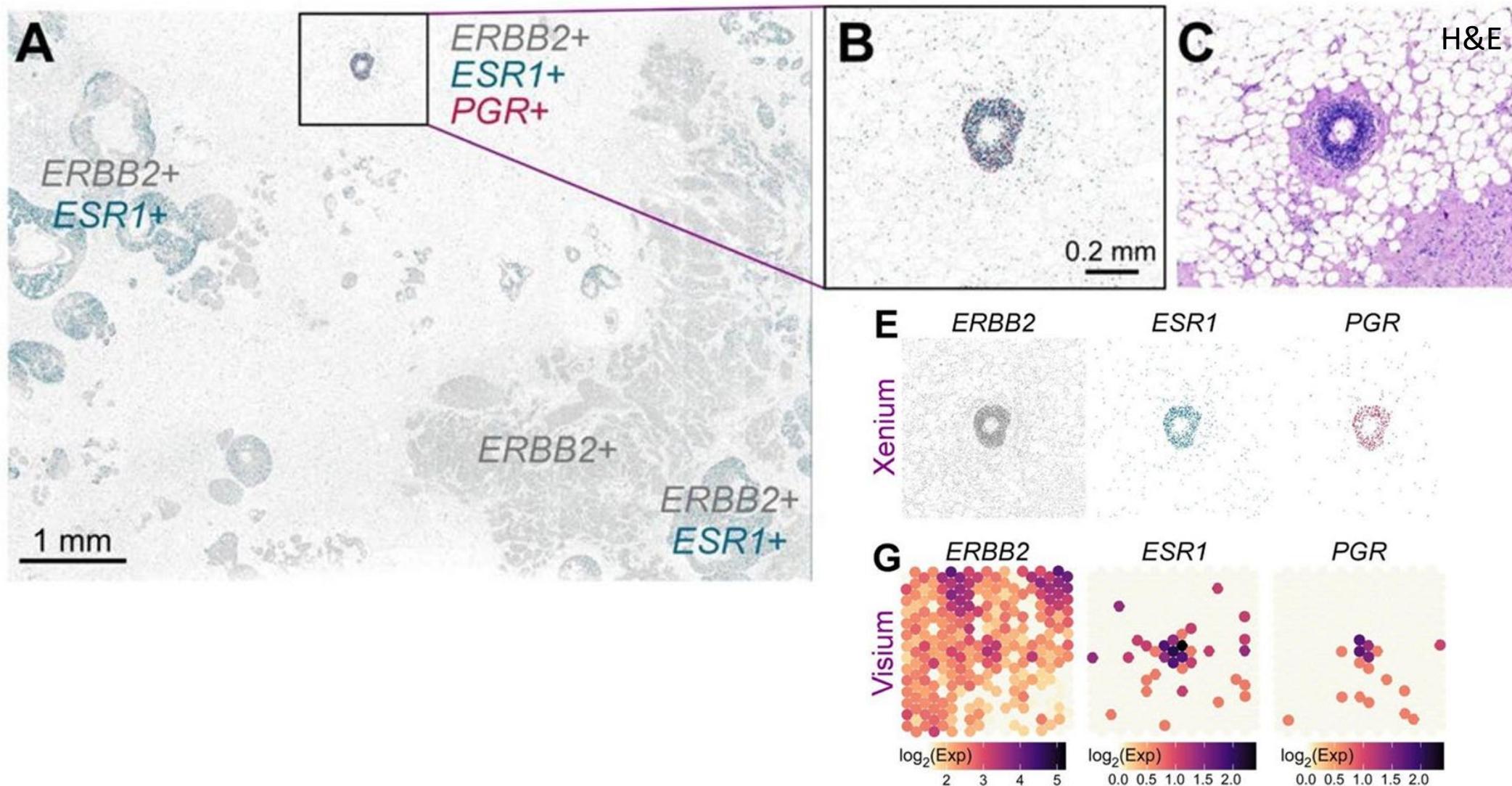


Sensitivity



Xenium was 1.4x more sensitive than 10x Genomics' most sensitive single cell assay, Chromium Gene Expression Flex

Xenium vs. Visium



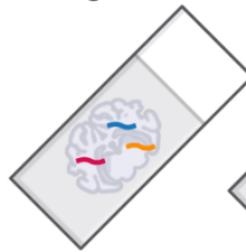
Imaging-based technologies - CosMx™ SMI



1

SAMPLE PREPARATION

Permeabilize,
fix, retrieve
targets



Hybridization of RNA
specific probes
and antibodies



Flow cell
assembly



2

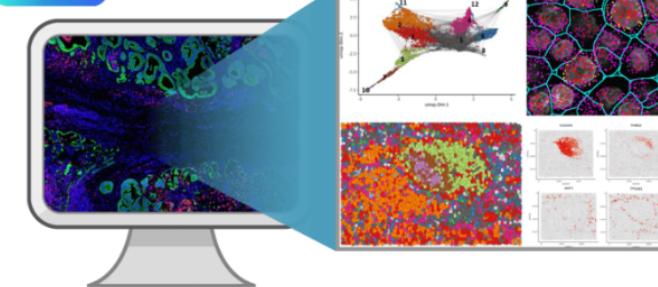
INTEGRATED READOUT



Robust *in situ* hybridization
chemistry and readout

3

INTERACTIVE DATA ANALYSIS

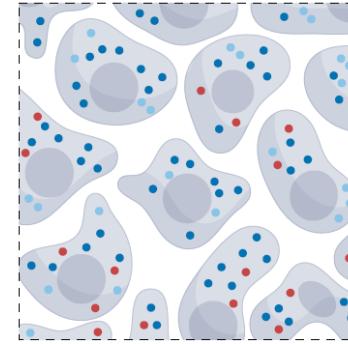
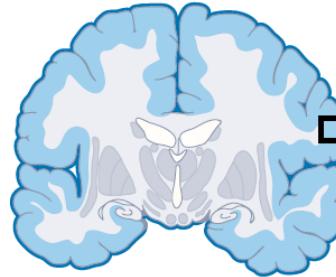


Cloud-based scalable computing
and storage with interactive data viewer

- cellular and subcellular resolution.
- up to 6,000 RNA
- 64 validated proteins

Imaging-based technologies - data analysis

Segmentation mask with transcript locations



Coordinates and counts of transcript

	x	y	Count	Cell
gene1	-10	3	15	1
gene2	-15	4	2	1
gene1	2	3	5	2
gene2	4	2	10	2
gene1	-12	10	3	3
gene2	-14	8	1	3
...



Cell count matrix and cell coordinates

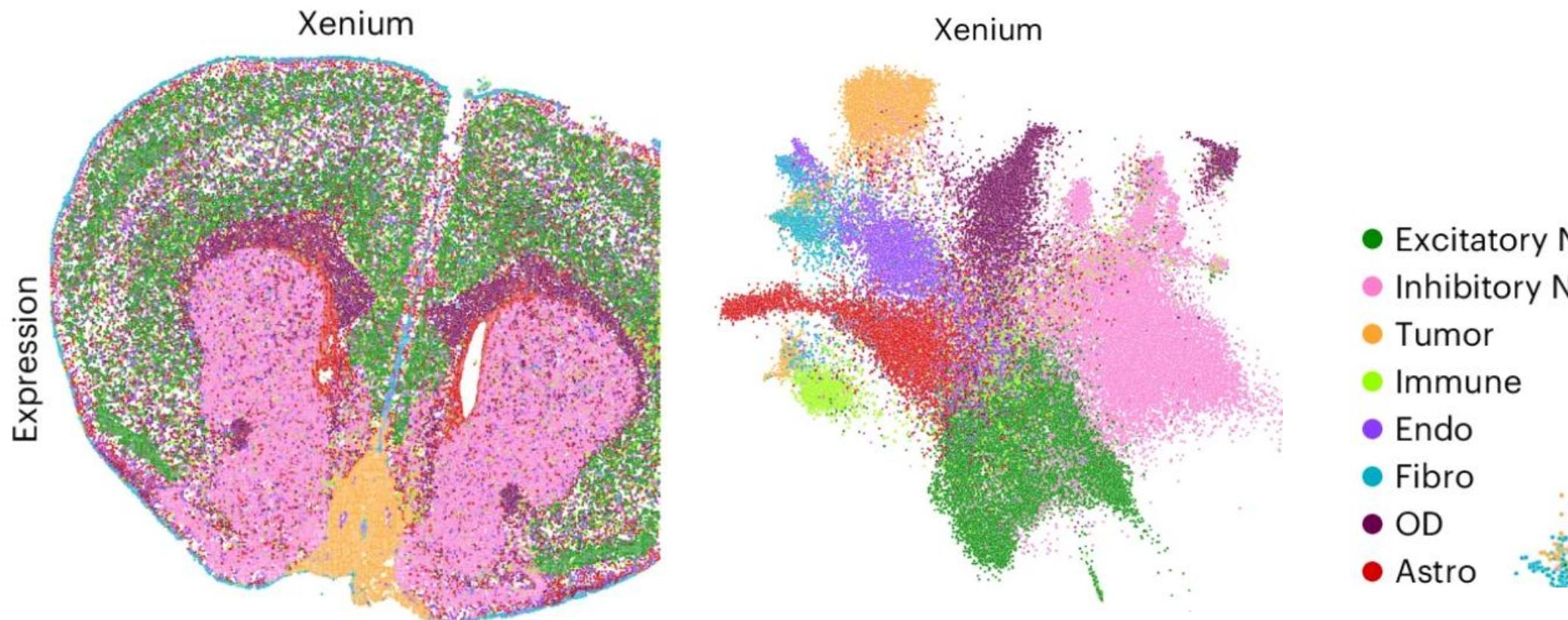
	m_{raw} genes →	x	y
↓ n_{raw} cells	15	2	...
5	10	...	3 2
3	1	...	-13 9
...

Analogous
to scRNA-
seq analysis

- predefined set of transcripts
- Cell segmentation
- similar to scRNA-seq data analysis (Squidpy, Giotto, Seurat or SpatialExperiment)
- Imputation of the whole transcriptome (measured in standard scRNA-seq) in a spatially resolved manner (Tangram outperform other imputation methods)



Spatial mapping and spatial variation detection



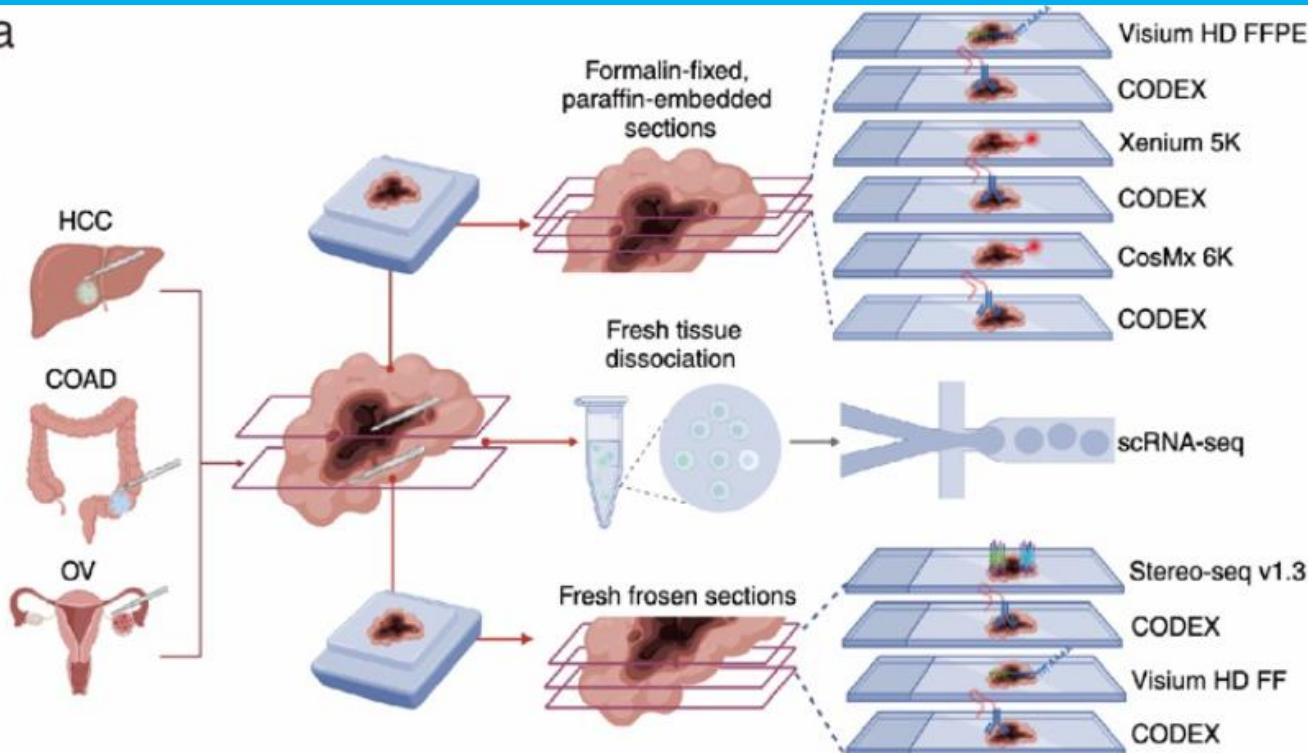
Spatial multi-omics



- Visium CytAssist Gene and Protein Expression up to 35 antibodies
- MERSCOPE up to 6 proteins
- CosMx™ SMI up to 6,000 RNA and 64 validated proteins

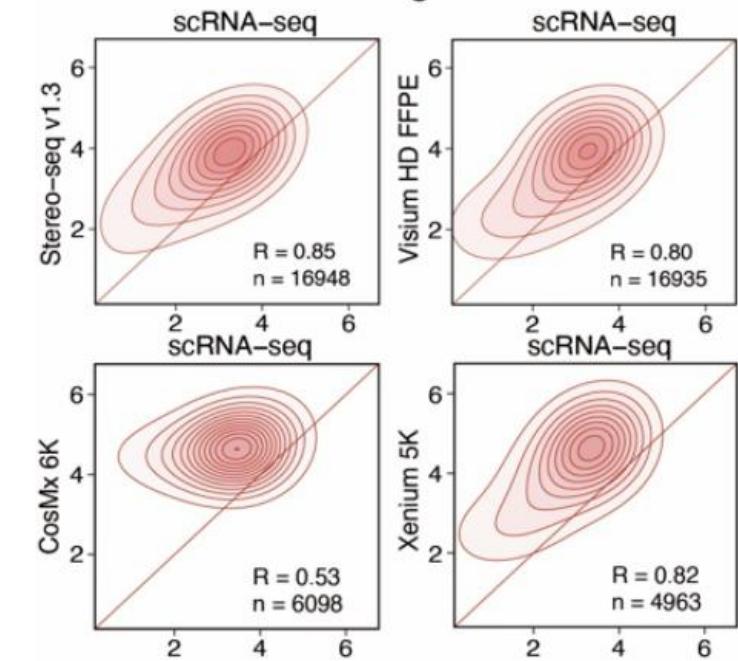
Which technology is better?

a



- **Sensitivity - Xenium 5K**
- molecular capture efficiency
 - Stereo-seq v1.3, Visium HD FFPE, and Xenium 5K showed high correlations with scRNA-seq
 - CosMx 6K displaying a skewed pattern

- high-throughput gene capture capacity (>5,000 genes)
- subcellular resolution ($\leq 2\mu\text{m}$)
- wide adoption (commercialized)





Which technology is better? – sequencing-based (sST) vs. imaging-based (iST)

- Background noise
 - iST platforms: Xenium 5K < CosMx 6K
 - sST platforms: Stereo-seq v1.3 and Visium HD FFPE
 - Both exhibited transcript diffusion beyond tissue boundaries
 - better diffusion control for Visium HD FFPE > Stereo-seq v1.3
- spatial concordance with CODEX
 - Platform and sample-dependent variations were observed, with Xenium 5K performing better in COAD, Visium HD FFPE showing better concordance in HCC, and CosMx 6K and Stereo-seq v1.3 performing better for certain cell types in ovary.
 - **iST platforms exhibited better concordance with CODEX than sST platforms**
- Cell segmentation
 - a critical step for ST platforms with subcellular resolution, heavily impacting downstream analyses.
 - **higher accuracy for iST platforms using DAPI images**
 - **iST platforms had a higher proportion of transcripts confined within the segmented cells**
 - Xenium 5K showed better separation of mutually exclusive cell-type-specific marker gene pairs after segmentation.
- **iST technologies demonstrated better clustering quality than sST platforms, highlighting the advantage of their higher spatial resolution**
- Xenium 5K showed a higher proportion of cells consistently annotated as the same cell type across tools, indicating robust annotation reliability. Xenium 5K exhibited the most distinct marker gene expression patterns, facilitating more accurate cell type annotations.



Which technology should I use? - What You Have

- Biological question – spatial info is critical to answer your biological question
- Species – Visium (many species) vs GeoMx DSP (human and mouse only, morphology markers)
- Sample number & size
 - GeoMx – many small tissues or 1 large tissue
 - Visium – 2 or 4 tissues,
 - 2 cm X 3 cm with Stereo-seq
 - 6.5 X 6.5mm with 10X visium
- Tissue type – Fresh frozen vs FFPE
- RNA quality
- Budget
 - Xenium: 2 slides at \$16,650 (5K panel) + \$1915 (advanced segmentation) = \$18,565
 - MERSCOPE: \$15,510 + \$2,160 each additional sample imaging kit
- Accessibility



Which technology should I use? - What You Need

- mRNA capture efficiency (Imaging vs. sequencing based)
- Spatial resolution
- Number of genes profiled (WTA vs Panels, need custom genes?)
- Protein panels (Available vs no available, # of proteins)
- Imaging area (0.65X0.65cm² vs 2X3cm²)
- Sensitivity & specificity

Platform	Species	Tissue Type	# of genes	# of Proteins	Resolution	Capture efficiency	Imaging Area
10X Visium	Any with V1 kit	Fresh & FFPE	WTA	35	55um	Low	0.65X0.65cm, 1.1X1.1cm
GeoMx DSP	Human & mouse	Fresh & FFPE	WTA	570	50um	Low	3.6X1.4cm (select ROI)
VisiumHD	Human & mouse	FFPE	WTA	N/A	Single cell	Low	0.65X0.65cm
Stereo-Seq	Any	Fresh	WTA	N/A	Single cell	Low	1X1cm, 2X3cm
MERSCOPE	Human & mouse	Fresh & FFPE	1000	6	Single mol	High	1 X 1 cm ² , Ultra 3.0x 3.0 cm ²
Xenium	Human & mouse	Fresh & FFPE	5000	N/A	Single mol	High	1 X 2 cm (select FOV)
CosMX SMI	Human & mouse	Fresh & FFPE	6000	64	Single cell	High	2 X 1.5 cm (select FOV)



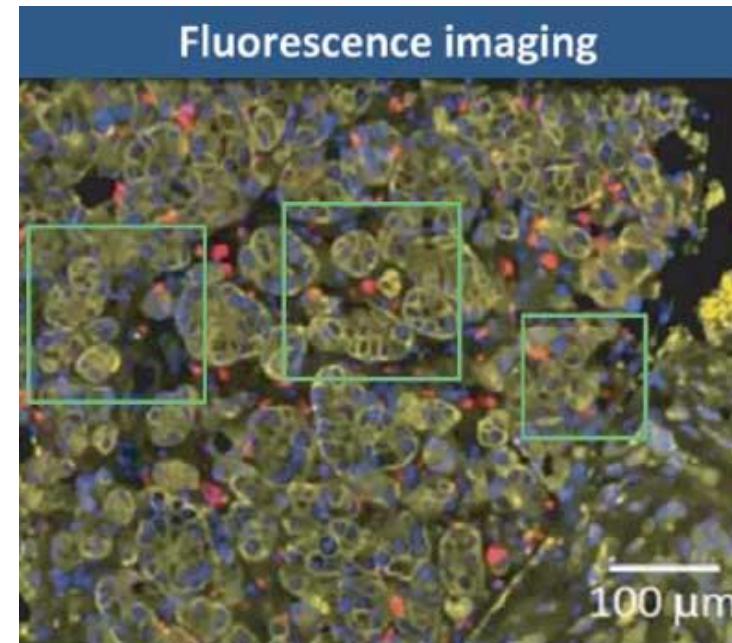
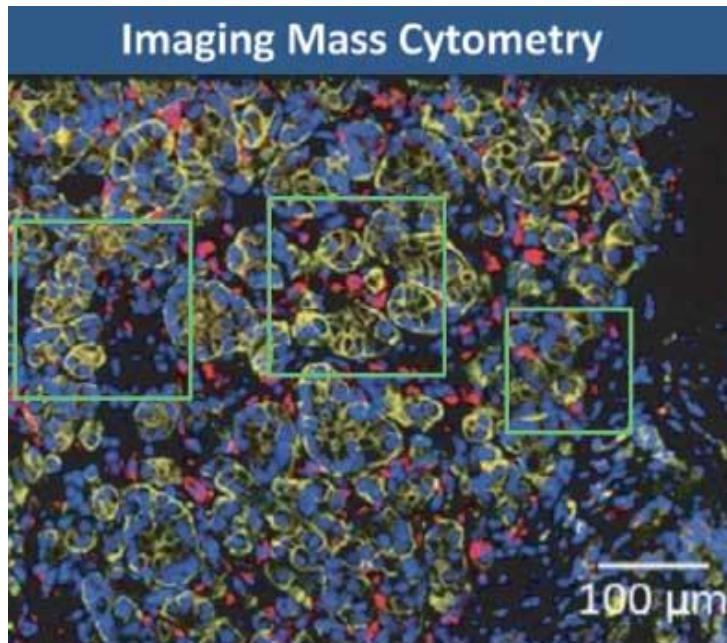
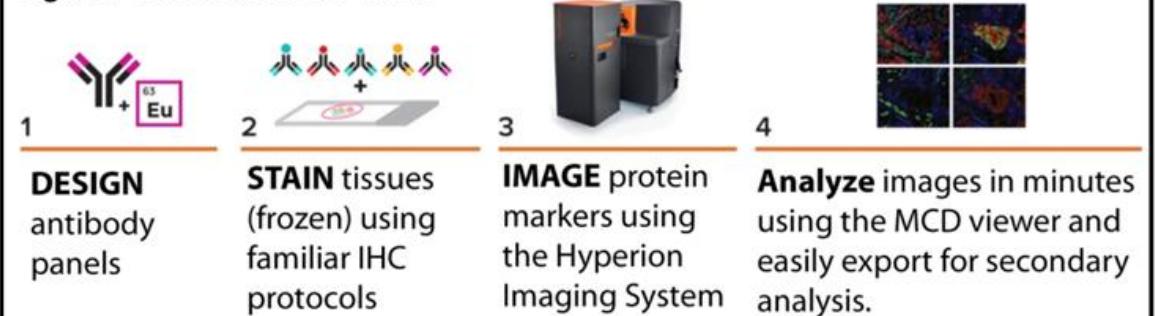
- Proteins are the main factors that shape cellular structure, enzymatic activity, cell–cell communication, and other essential physiological functions.
- Highly multiplexed images of specimens such as tissue and organ slices to understand their protein composition and spatial organization
- immunohistochemistry-based methods
 - imaging mass cytometry (IMC)
 - PhenoCycler/co-detection by indexing (CODEX)
 - iterative bleaching extends multiplexity (IBEX)
 - cyclic immunofluorescence (cyclIF)
 - multiplexed ion beam imaging (MIBI)
- deep visual proteomics (DVP)
 - laser dissected and individual dissociated cells are analyzed by mass spectrometry in such a way that their spatial context information is retained to create spatial protein maps.
 - not limited by the number of available antibodies
 - substantially greater proteome coverage

Spatial Proteomics - Imaging Mass Cytometry (IMC)



- Use metal-tagged antibodies (40+ proteins)
- Single-cell resolution
- No autofluorescence
- Protein and RNA co-detection
- Cons: small imaging area

Figure 1. Workflow for IMC.

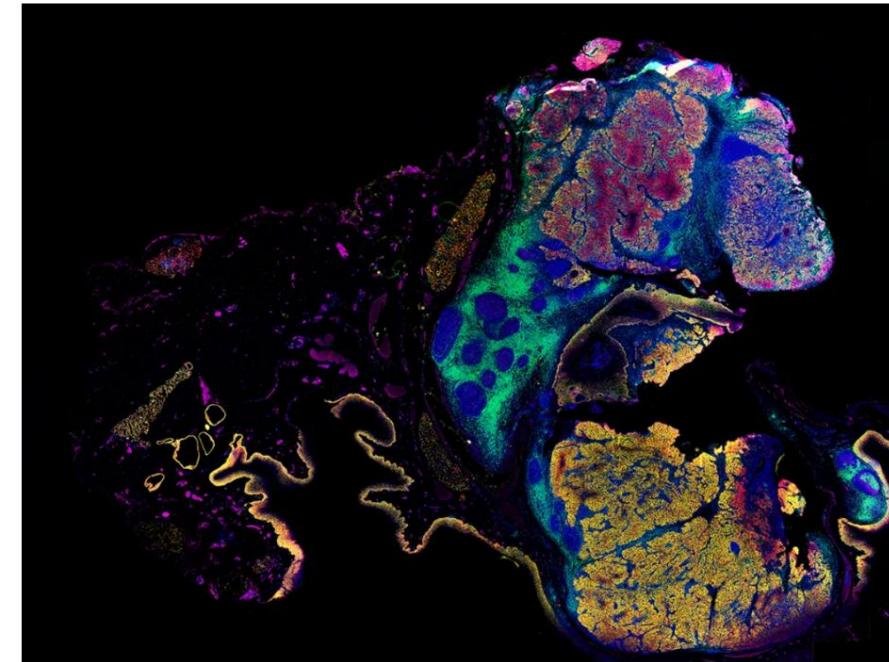
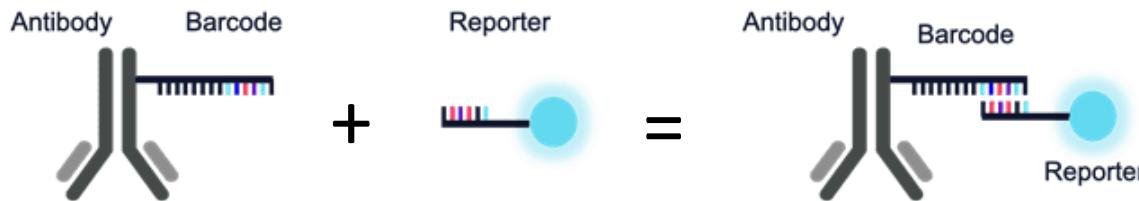


CD68
pan-cytokeratin

Spatial Proteomics - PhenoCycler (formerly CODEX®)

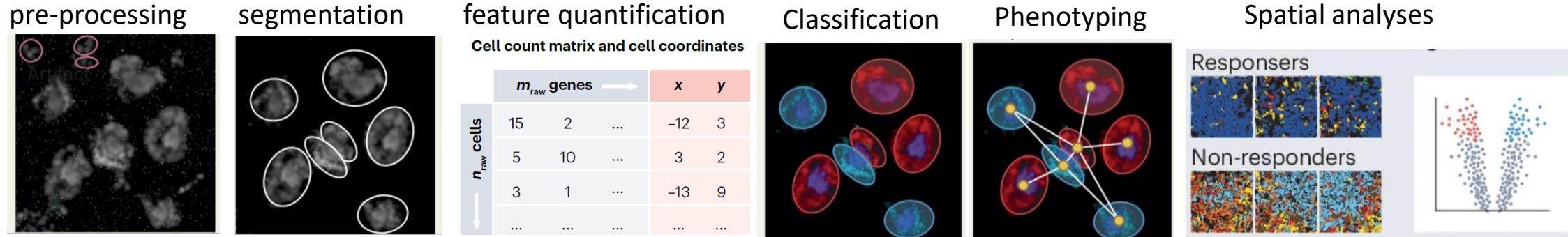


- integrating automated fluidics and iterative imaging
- Use barcoded antibody (100+ proteins)
- single-cell resolution
- imaging 1 million cells in 10 minutes
- Whole-Slide Imaging



World's-first whole-slide, 100+ plex publication on head and neck cancer with curated panel design based on cancer hallmarks

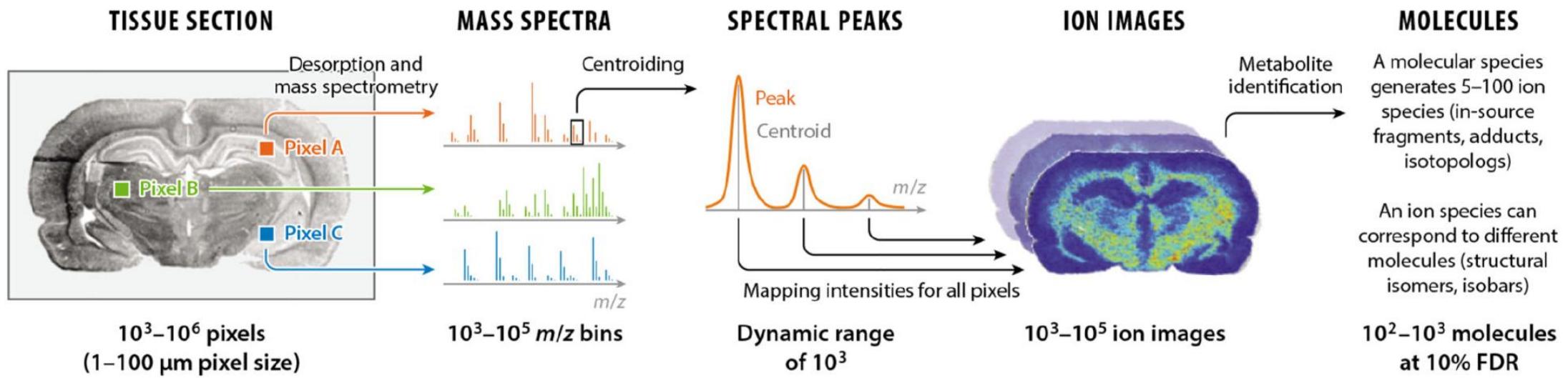
Spatial Proteomics data analysis



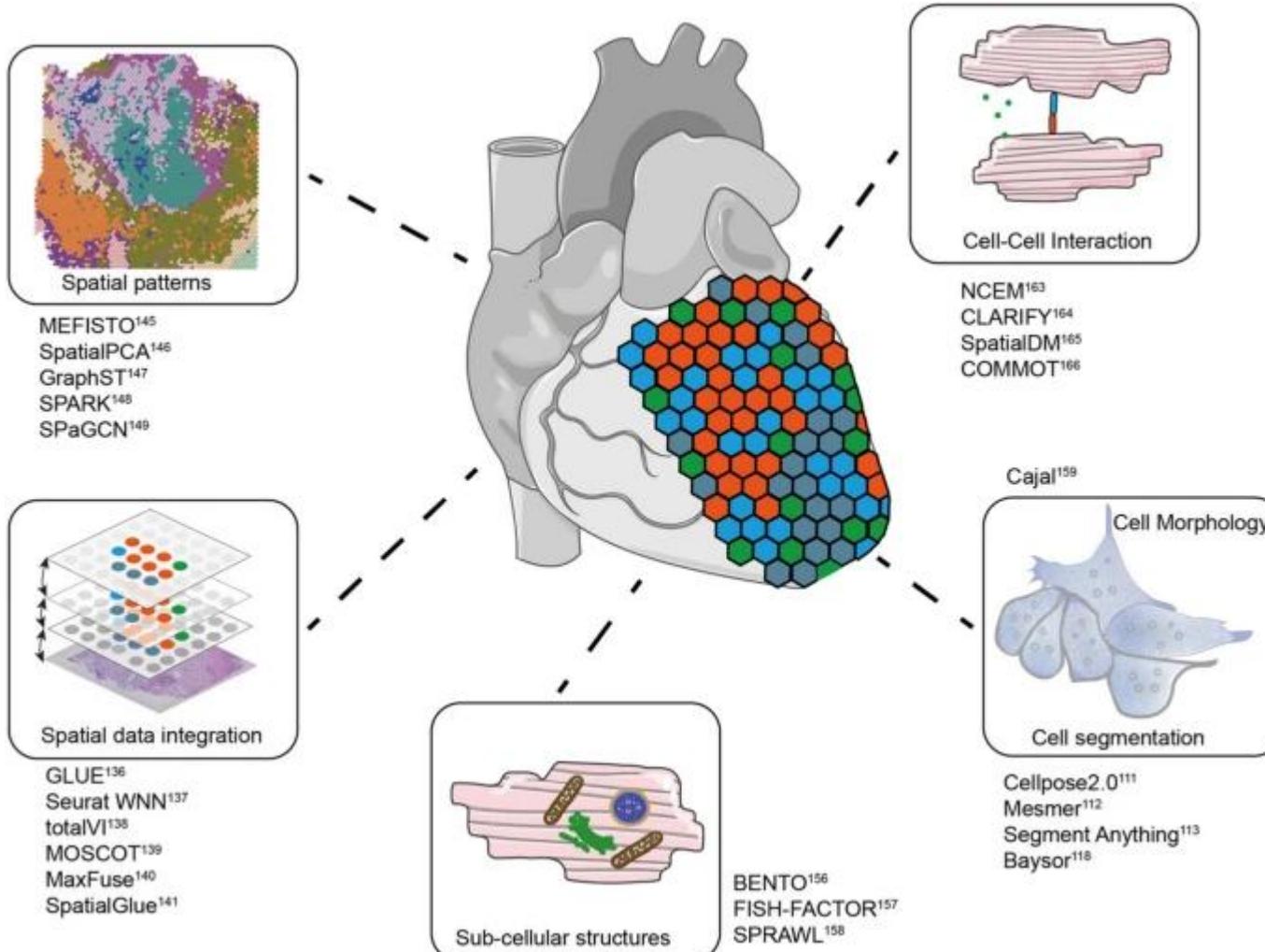
- Image pre-processing: image registration and correction of experimental and imaging artifacts, including channel crosstalk, background, noise etc.. (technology- specific)
- Cell segmentation: facilitated by recent advances in AI algorithms, primarily trained on large, manually curated datasets
- Feature quantification: the mean intensity of each protein, cell size, circularity, location
- Cell classification: normalization, filtering for lineage-based proteins, gating, clustering, supervised machine learning or probabilistic modeling
- Cell phenotyping: annotate cells based on marker proteins (e.g. neurons, astrocytes).
- Spatial analysis: cell-to-cell interactions, define microenvironments, segment microanatomical structures and stratify patients for clinical insights

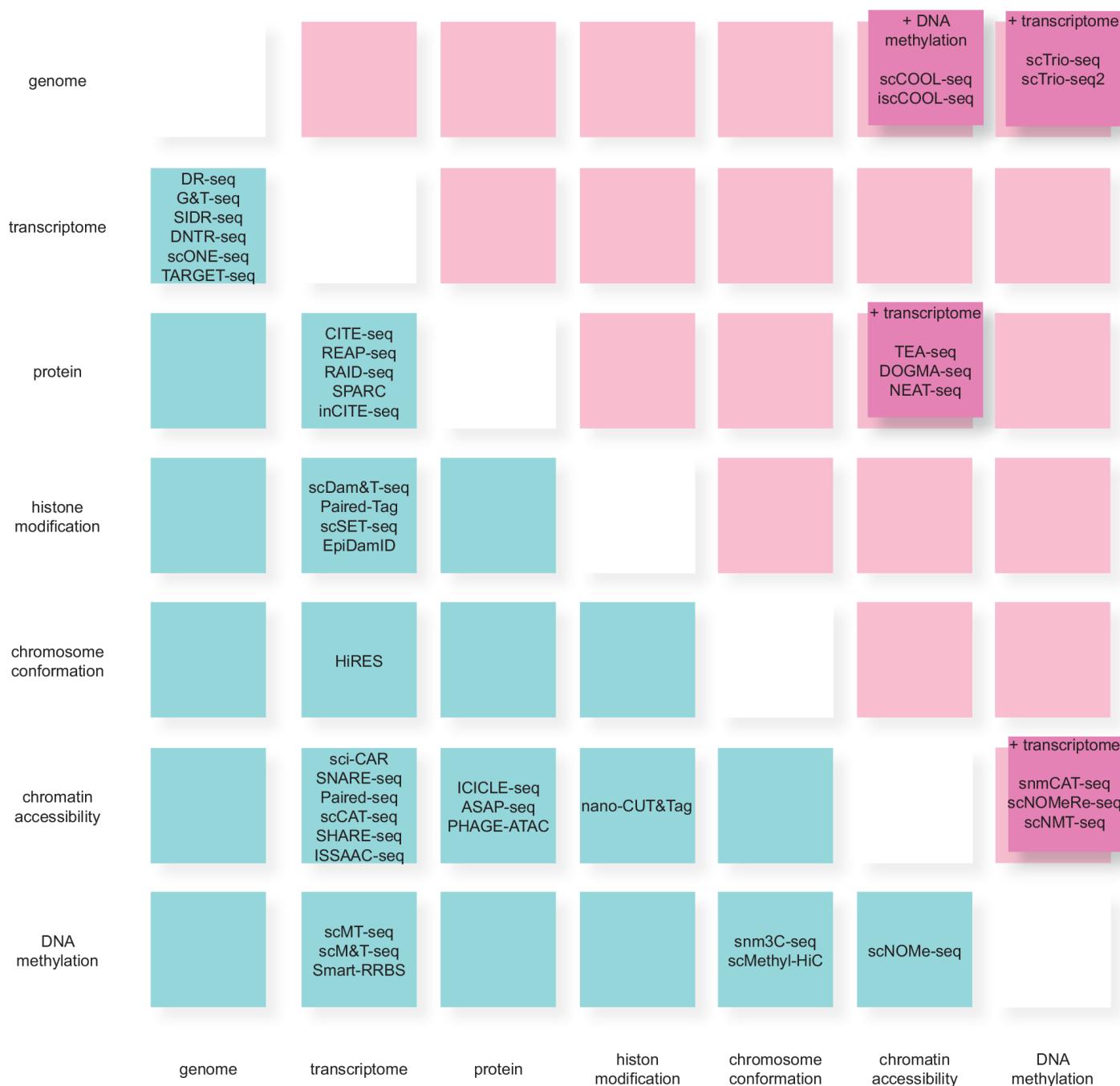
Spatial metabolomics

- enabled localizing metabolites, lipids, and drugs in tissue sections
- has reached 5–10 μm resolution
- Single-cell metabolomics just started



Spatial -omics – specific data analysis





single-cell multimodal omics

blue box : dual omics technology
magenta box: ≥ 3 omics modality

Advances in single-cell omics and multiomics for high-resolution molecular profiling, Lim et al., 2024 ,
<https://www.nature.com/articles/s12276-024-01186-2>



Single cell genomics

- Single cell genomics I
 - History of Single cell technology
 - Single cell RNA-sequencing (scRNA-seq) technology
 - Basic scRNA-seq data analysis workflow
 - Unlocking biological insights
 - Other unimodal single-cell technology
- Single cell genomics II
 - Single-cell multiomics
 - Transcriptome + Epigenome
 - Transcriptome + Protein
 - Transcriptome + CRISPR screening
 - Transcriptome + TCR/BCR
 - Transcriptome + Antigen specificity
 - Spatial genomics
 - Spatial transcriptomics
 - Spatial proteomics
 - Spatial multiomics
 - Spatial metabolomics
- Single cell genomics Lab
 - scRNA-seq data analysis

All the single cell technologies at your disposal

