Single Cell & Spatial Transcriptomics

Yan Asmann, PhD
Associate Professor of Biomedical Informatics
Department of Quantitative Health Sciences
Mayo Clinic, Jacksonville Florida

Outline

Lecture

- NanoString spatial omics platforms
- Case study in Non-Small Cell Lung Cancer (NSCLC)
- Introduction to pseudo-bulk scRNA analyses

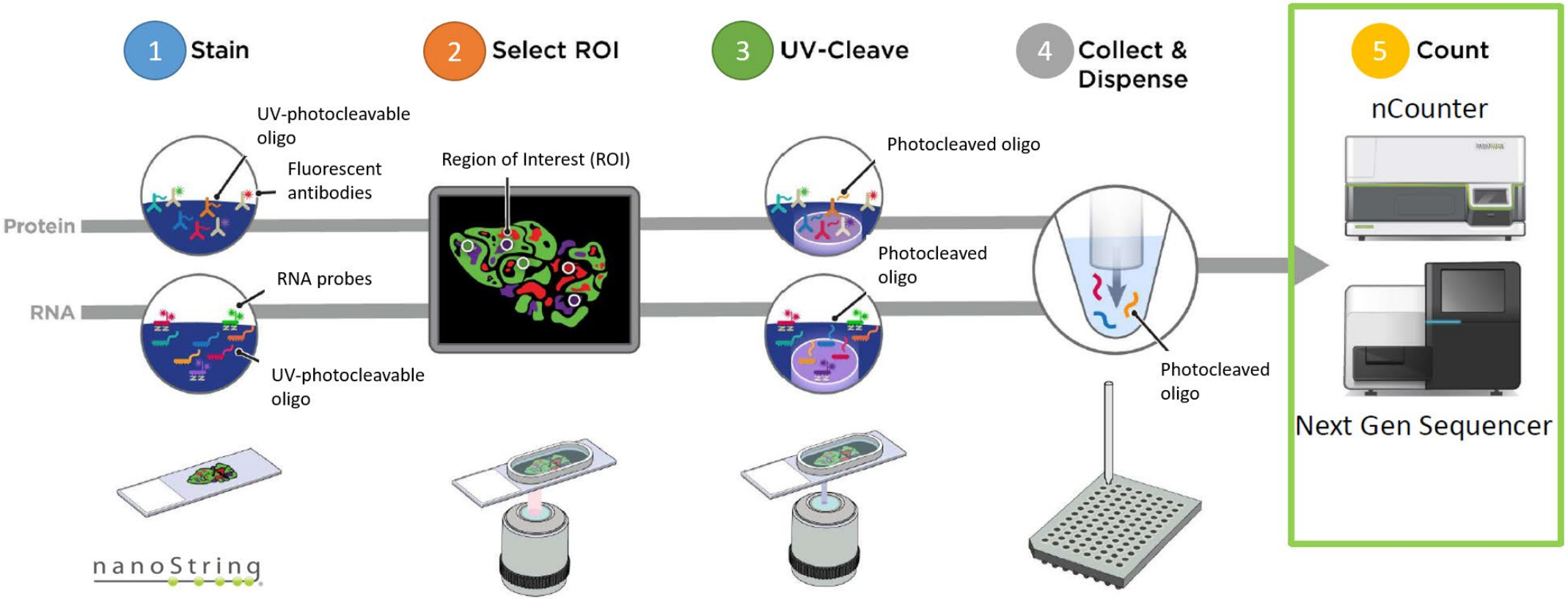
Lab (self-paced)

- Multi-sample scRNA analyses
- Pseudo-bulk scRNA analyses
- Lab material preparation: Zach Fogarty and Clark Ikezu

Introduction to NanoString Spatial Platforms

- **GeoMx** Digital Spatial Profiler (DSP): High resolution (10 microns)
- **CosMx** Spatial Molecular Imaging (SMI): True single cell resolution

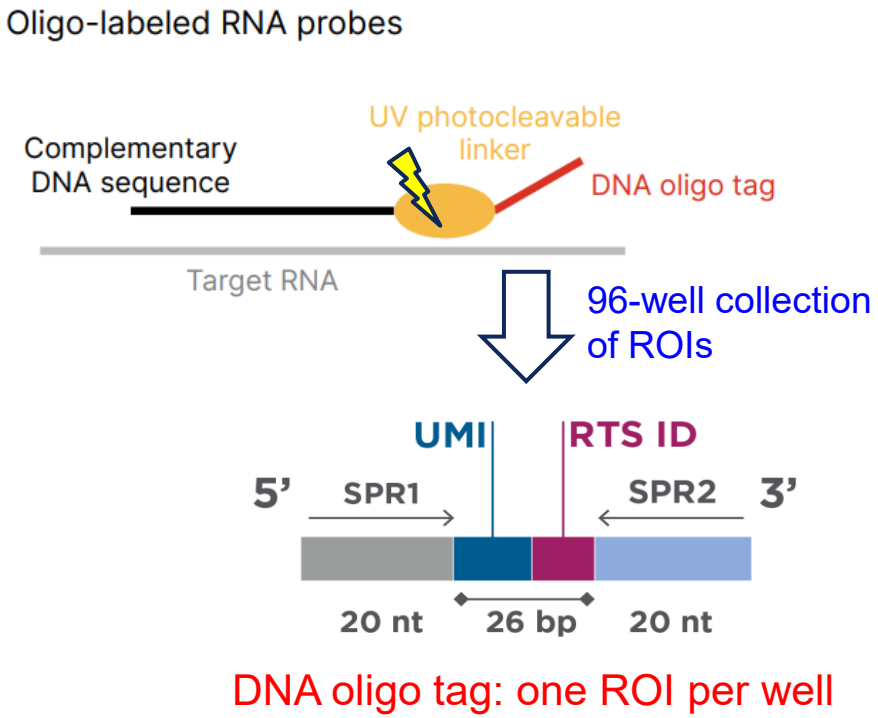
GeoMx Digital Spatial Profiler (DSP):



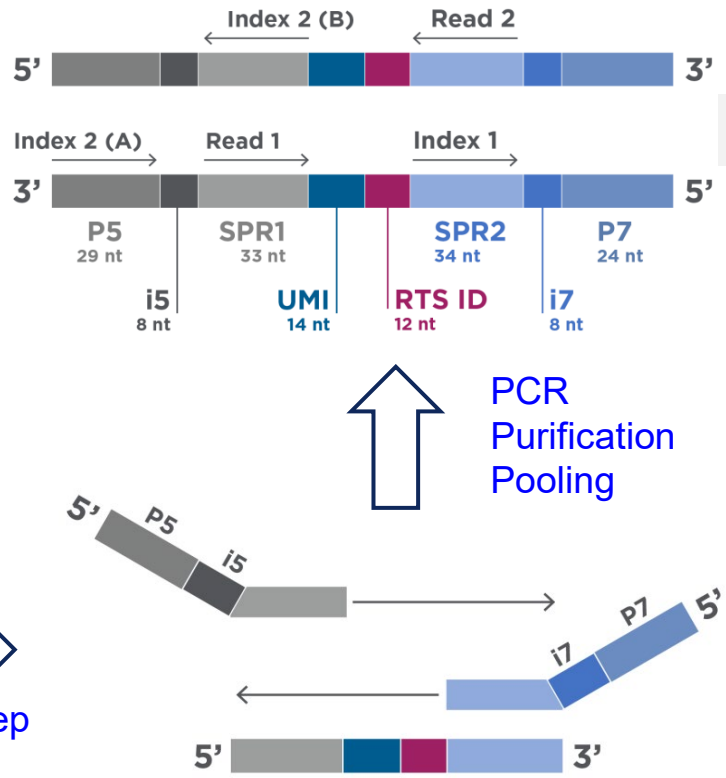
- Select distinct tissue compartments or cell types
- Use micromirrors to expose the ROI to UV
- Release barcodes for quantification

GeoMx Digital Spatial Profiler (DSP):

Full spatial ROI transcriptome: The GeoMx Human Whole Transcriptome Atlas (18,000+ protein-coding genes)



Sequencing library Prep

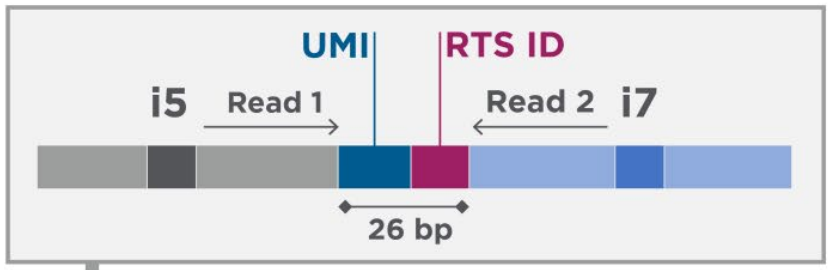


Sequencing library

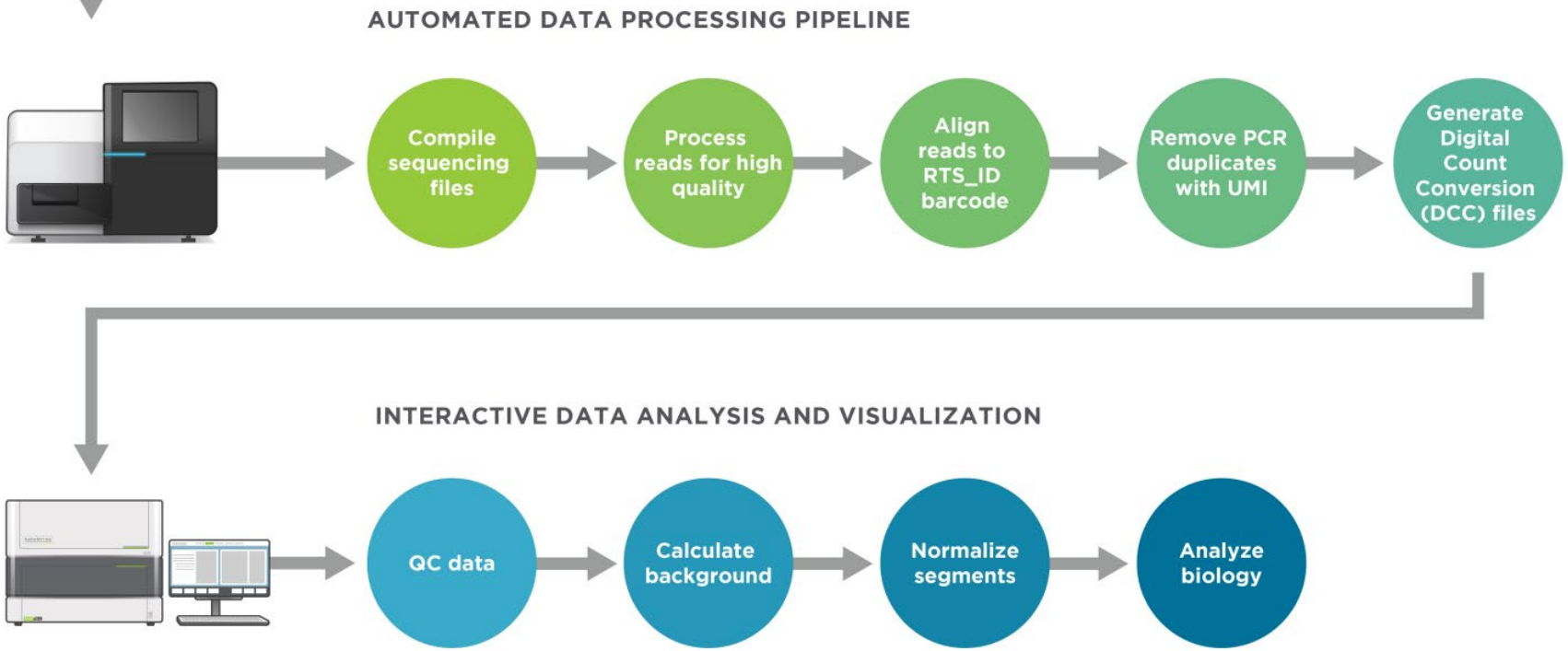
- Readout tag sequence identifier (RTS ID): biological targets
- Unique molecular identifier (UMI): identify PCR artifacts
- SPR: sequencing primers

- i5/i7: uniquely identifies ROI
- P5/P7: binding to Illumina flowcell
- P5/P7: PCR primers

GeoMx NGS Pipeline



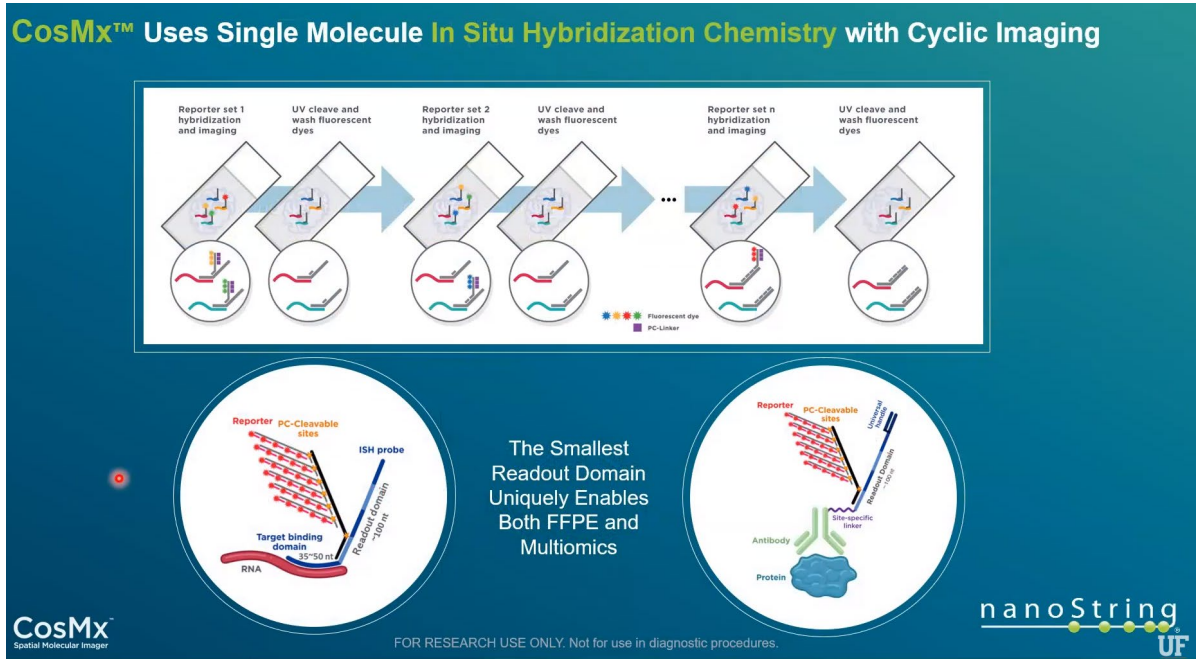
- A gene count table per ROI
- Compare ROIs of the same sample: e.g., Tumor adjacent to immune cells vs. others
- Compare similar ROIs between samples

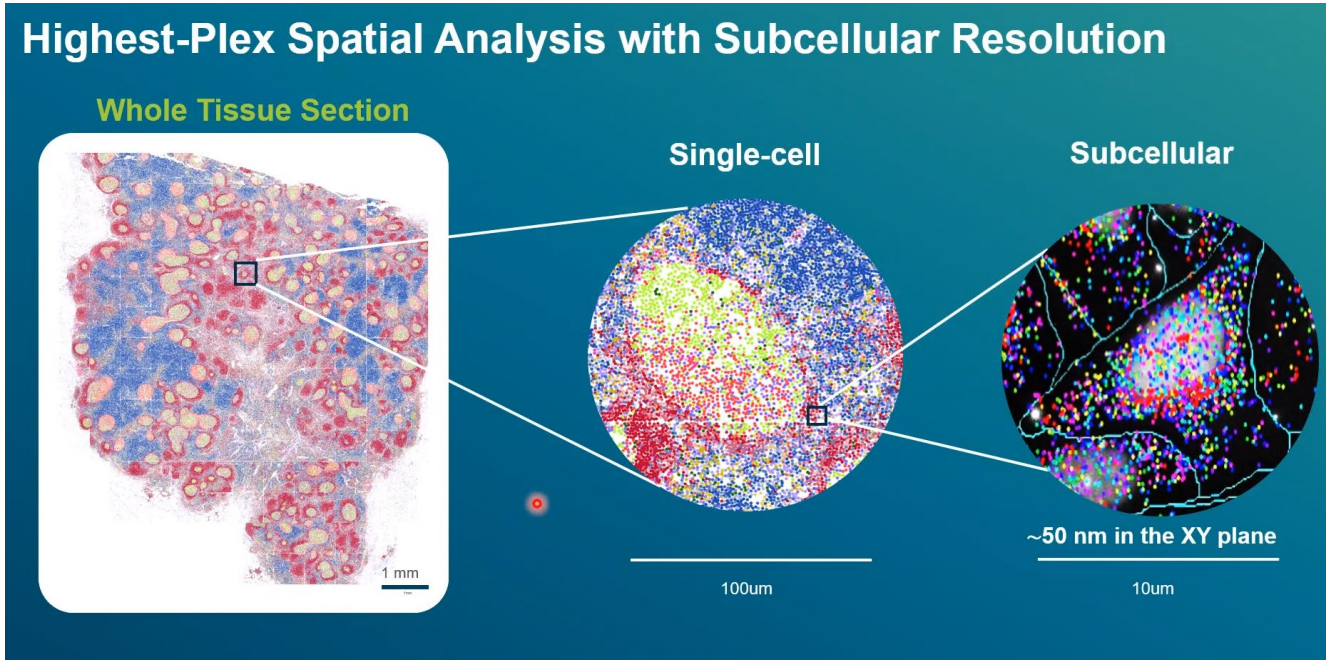


GeoMx Digital Spatial Profiler Summary

- Pseudo-spatial: via selection of ROIs
- Resolution: 10 microns
  - ✓ cell size: 10-20 microns in diameter
  - ✓ not conducive to single-cell analysis: need at least 20-300 cells per ROI for reliable quantification of transcripts
- Throughput: full transcriptome, plus more targeted panels
  - ✓ cancer transcriptome atlas (>1800 RNAs)
  - ✓ TCR profiling add-on

CosMx Spatial Molecular Imaging (SMI): True Single Cell Resolution, True Spatial





In Situ Chemistry: protein, RNA

Z-stack of multi-channel tissue images

3D

Subcellular Resolution

GeoMx *vs.* CosMx

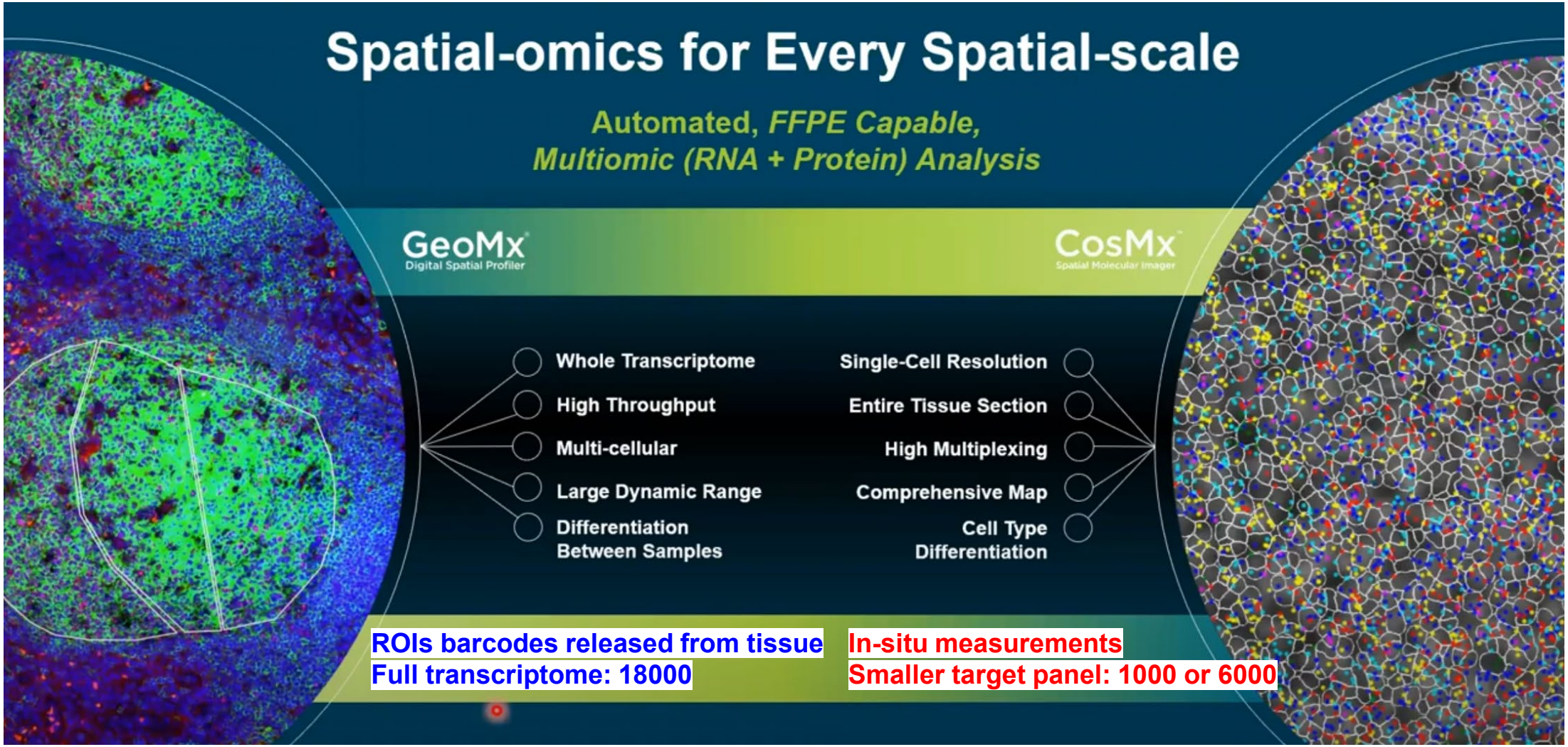
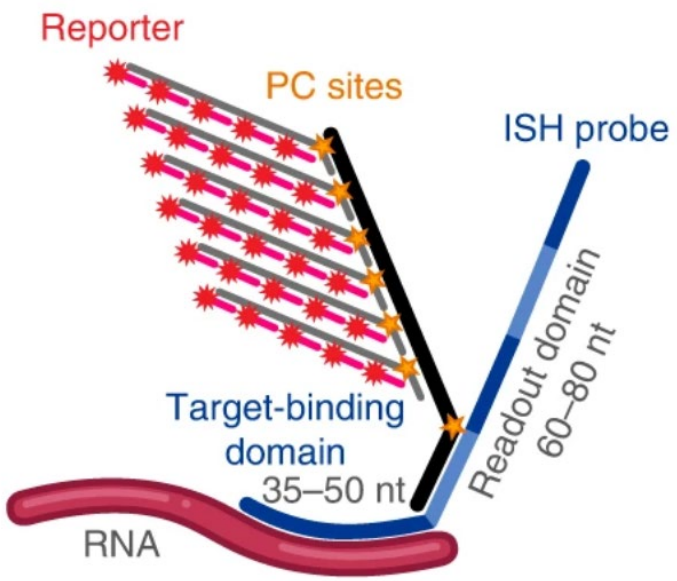


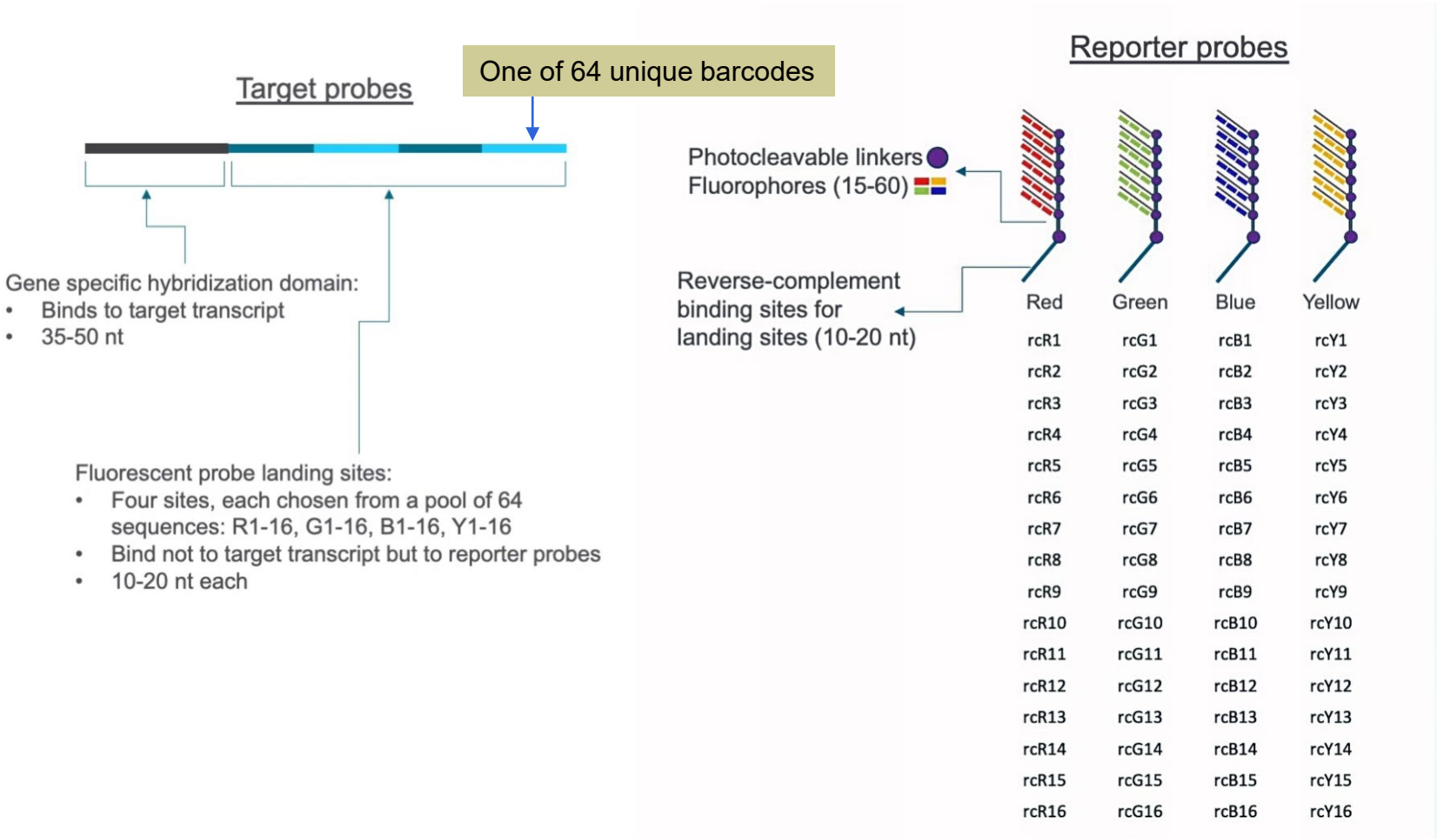
Figure from Nanostring

CosMx SMI chemistry and workflow



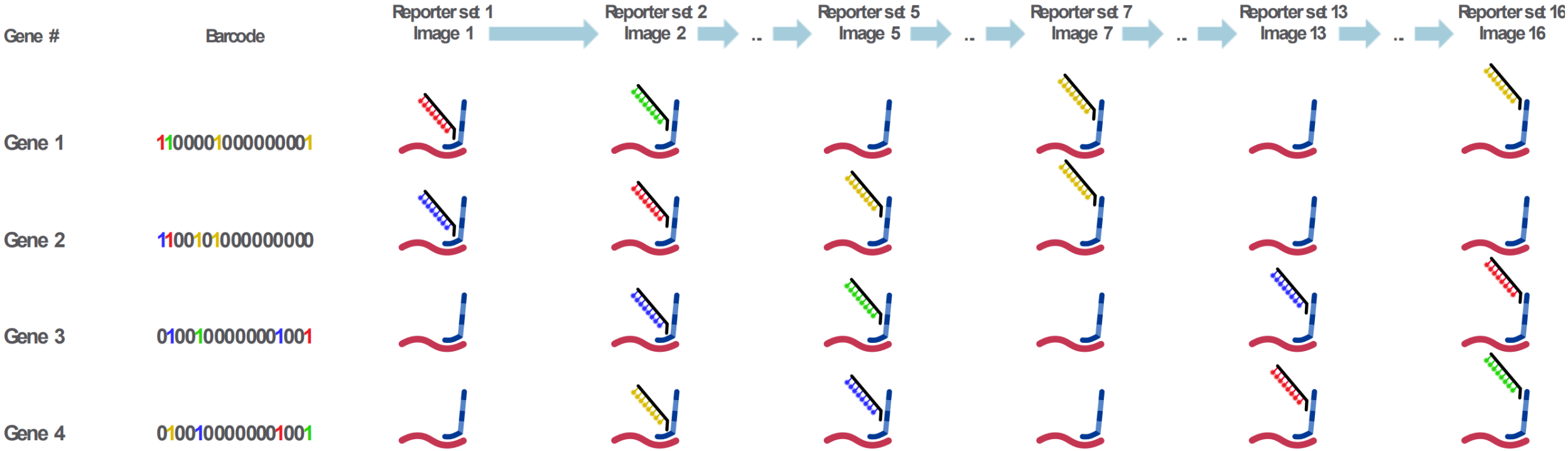
- **in situ hybridization (ISH) probes**
  - ✓ Target binding domain (30-50 nt)
  - ✓ Read out domain (60-80 nt)
  - ✓ 4 consecutive 10-20 nt sequences that are individual landing sites of 4 reporter probes
- **fluorescent readout probes called reporters**
  - ✓ photocleavable sites to remove reporters (☆)
  - ✓ The reporter landing domain still occupied after PC
- **detect RNAs in intact tissue**
  - ✓ 5 ISH probes are designed per gene to detect different regions of the RNA target,

CosMx SMI chemistry and workflow



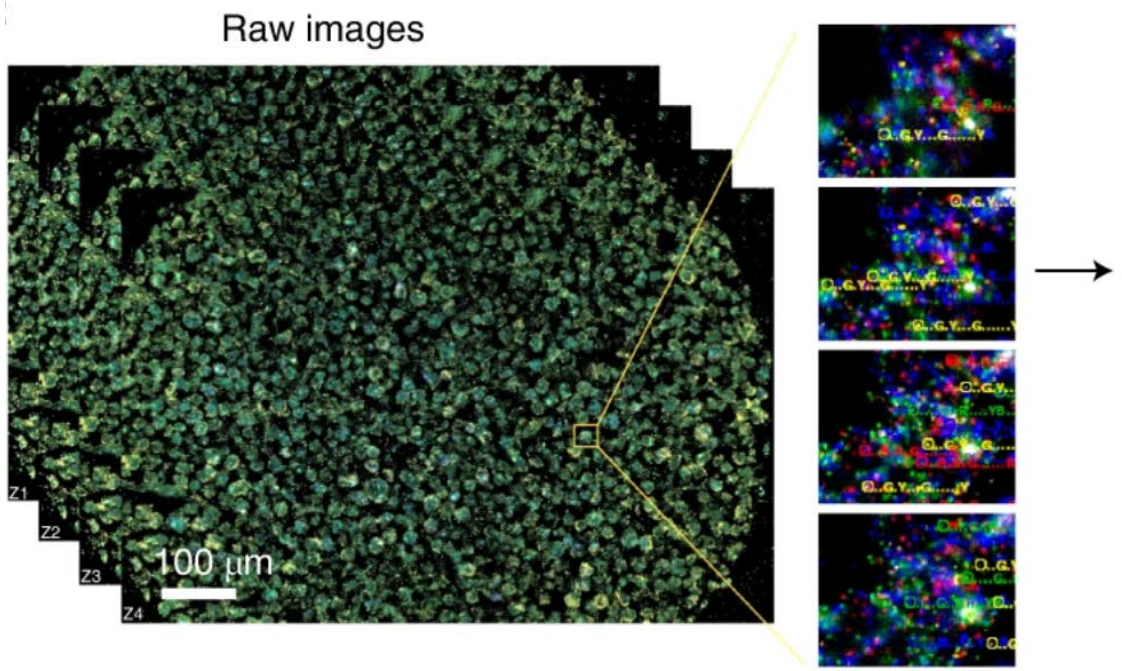
[Watch Nanostring University Video]

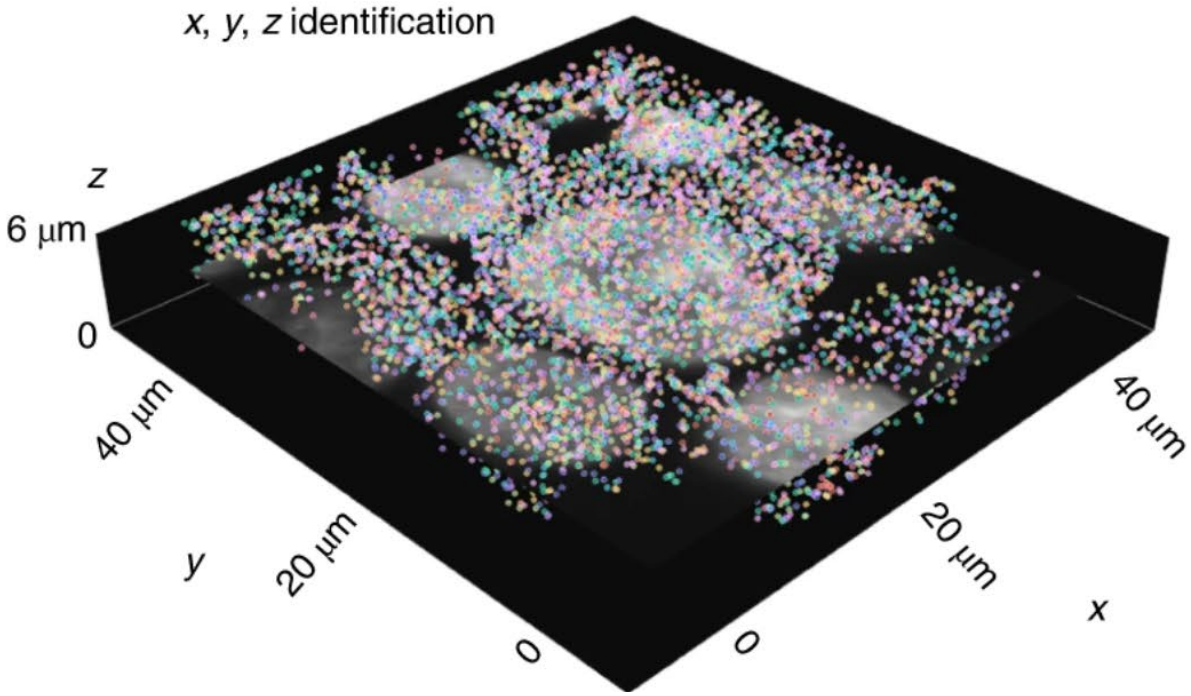
CosMx SMI chemistry and workflow



- Every gene has a 16-digit binary barcode (with four 1s, and 12 0s)
- with Hamming distance 4 (HD4): every barcode is separated by an HD of at least **four** from all other barcodes to maximally suppress RNA decoding error.
- Images will localize a gene in a cellular/subcellular location

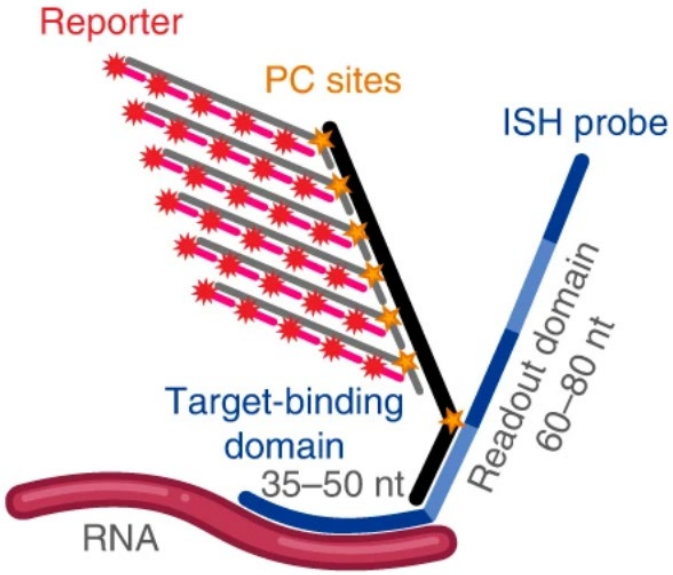
CosMx SMI: 3D RNA localization

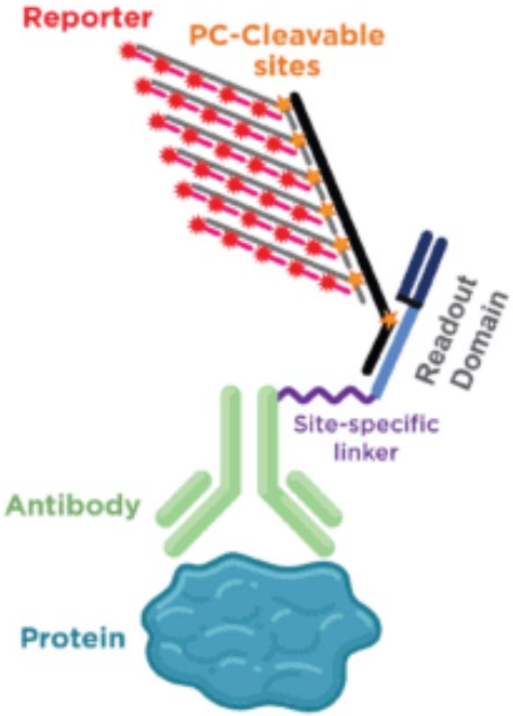




Optical Z-stacks: number of Z-stacks can be different between experiments.

CosMx SMI chemistry and workflow: RNA vs. Protein

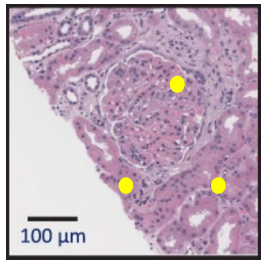




Protein Barcode Chemistry

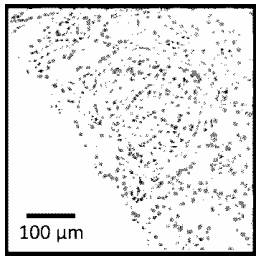
- Each antibody has a specific linker with a readout domain
- Protein chemistry is similar, but with a single readout domain for a single-color reporter to quantify a protein target

CosMx Spatial Molecular Imaging (SMI): Lab Workflow



FOV (Fields of View) Selection

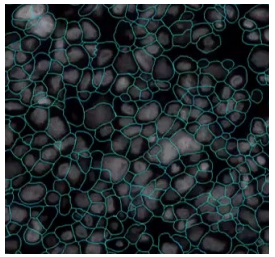
H&E Slide



Adjacent Slide

On selected FOVs

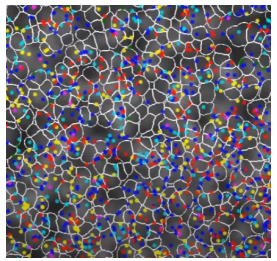
Protein Segmentation Markers



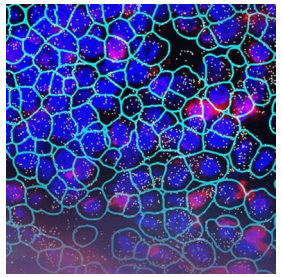
On selected FOVs

Panels:
Up to 108-Plex

64-Plex Protein Assay



1000-Plex RNA Assay



Panels:
Human 6K Discovery RNA
Human Immuno-Oncology
Mouse Neuroscience
Custom Assays …
Whole Transcriptome (>18K genes, 2025)

Measurements:

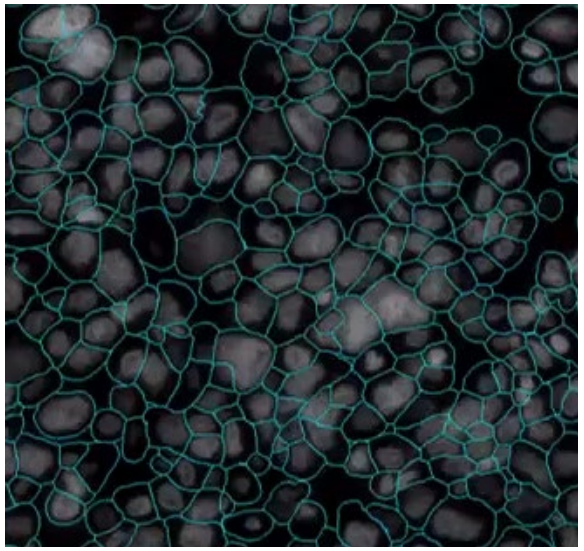
- Segmentation
- Cell typing
- RNA expression
- RNA cellular coordinates
- Niche or Neighborhood
- Cell/FOV level data

Data Size:
per FOV: ~2GB (10-300 FOV per project)

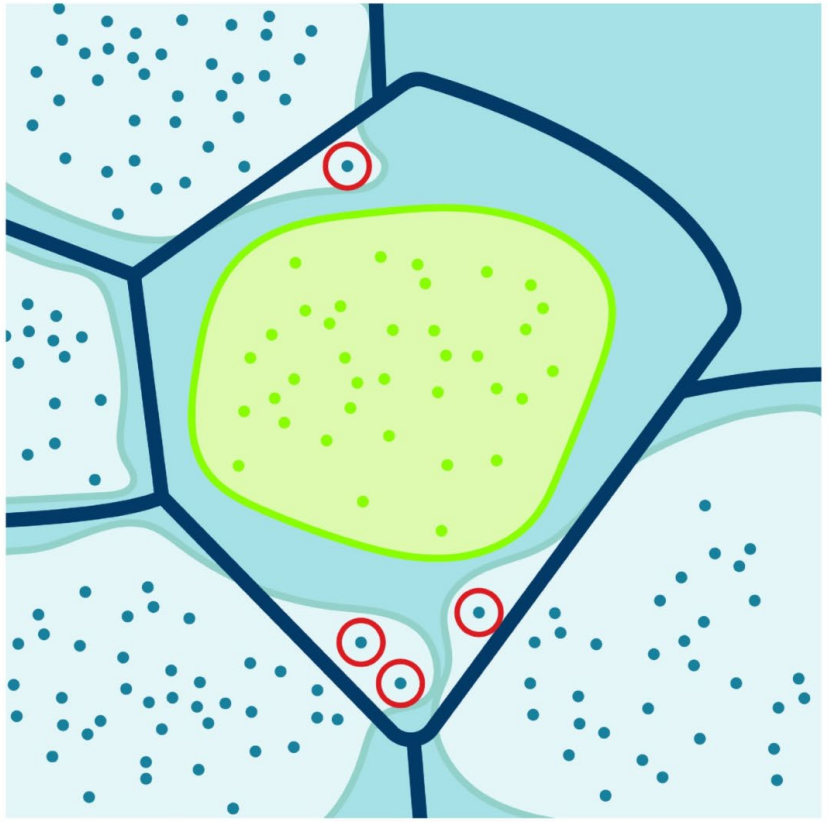
Single cells identified by segmentation

Accurate cell segmentation is challenging:

- Heterogeneous Shapes**:** nearly impossible to define mathematical shape models.
- Variation in Size and Shape**:** Unlike nuclei, the cytoplasm exhibits significant variations in shape and size.
- Weak Boundary Gradients**:** Cells that are in close proximity can have weak boundary gradients
- Makeshift Nature of Segmentation Approaches**:** dataset constraints, including differences in staining or imaging modality, artifacts in image capture, or morphological differences.



Single cells identified by segmentation



Tumor mRNA

T-cell mRNA

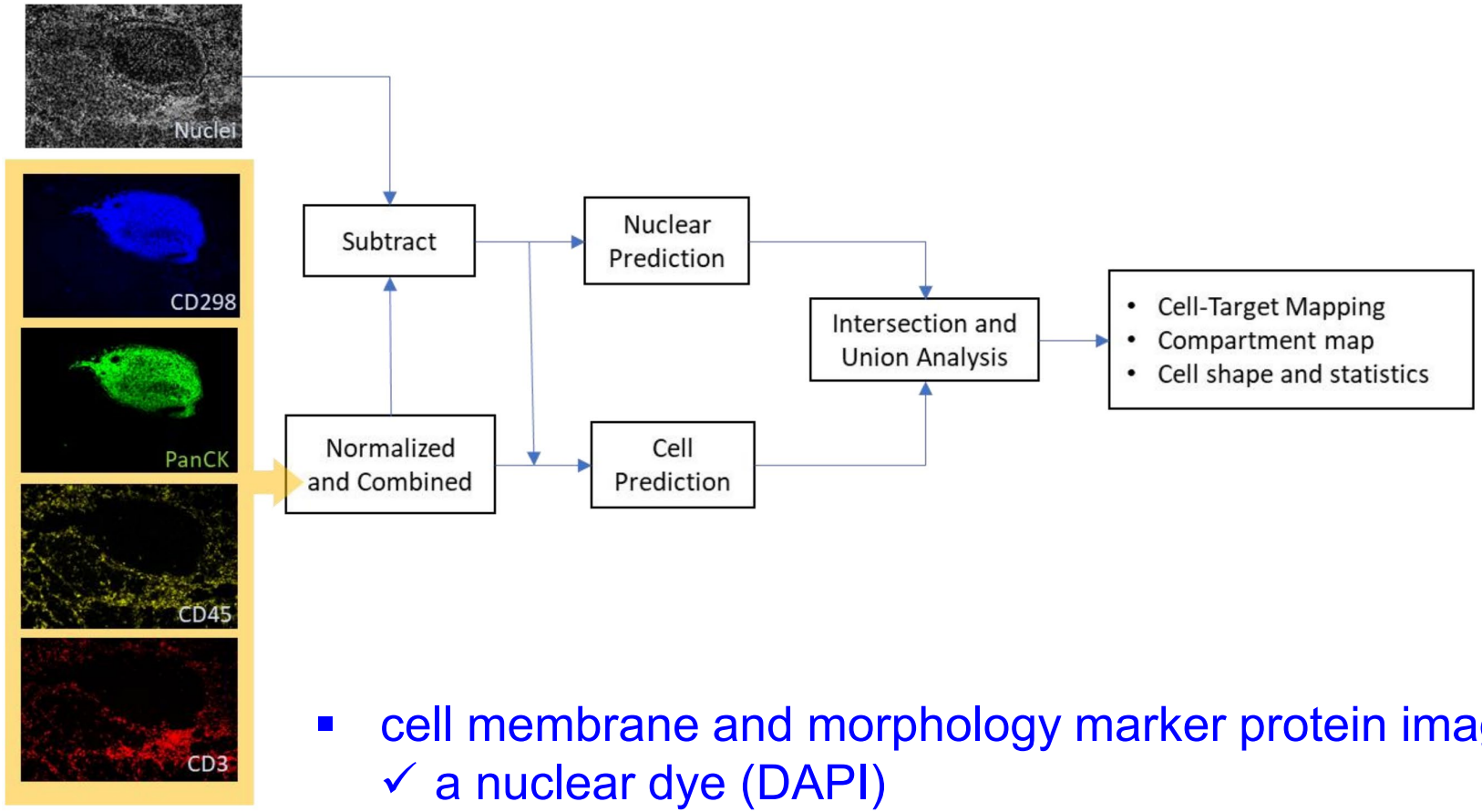
Segmentation boundaries

Tumor mRNA falsely attributed to T-cell after mild segmentation errors.

Consequence of minor segmentation error

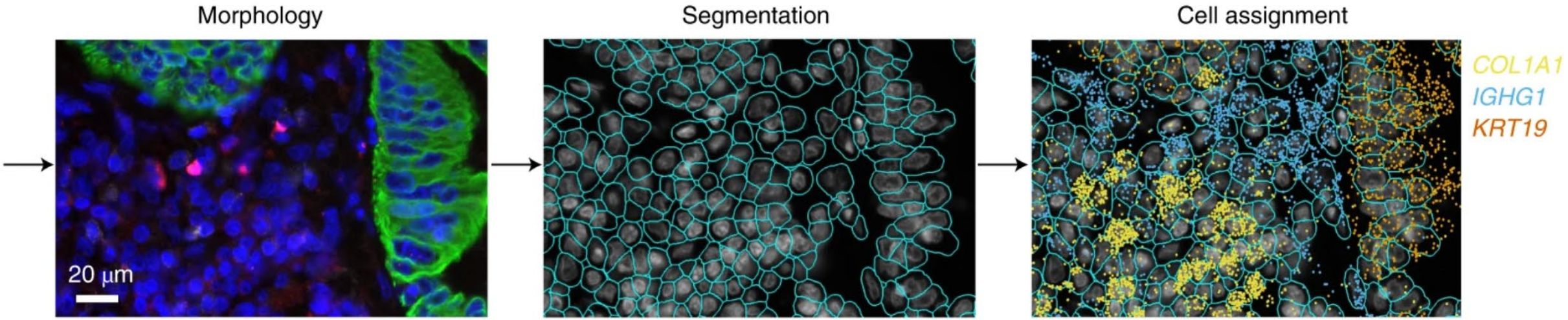
Precise cell segmentation is the most important parameter when determining data accuracy. An imager's ability to identify accurate cell boundaries to minimize segmentation errors provides the confidence to draw biologically impactful conclusions from your spatial data.

Single cells identified by segmentation



- cell membrane and morphology marker protein images
  - ✓ a nuclear dye (DAPI)
  - ✓ Protein markers: Membrane (CD298), epithelial cells (PanCK), and T cells (CD3), …
- machine-learning augmented cell segmentation (Cellpose neural network models)
- transcript-based segmentation refinement

Single cells identified by segmentation



Most genes/transcripts have a cell assignment

Data

- A table per FOV: genes, counts, cell ID, annotation
- A Seurat object: for analytics similar to those of single cell RNA-Seq (cell typing, U-map, differential expression, etc.)
- Nanostring AtoMx Pipeline: inspection of images, QC etc.

CosMx Spatial Omics: an example in NSCLC

Category	Total*	Lung 5_rep 3	Lung 5_rep 4	Lung 5_rep 5	Lung 6	Lung 9_rep 1	Lung 9_rep 2	Lung 12	Lung 13
Tissue type	FFPE human lung	FFPE human lung	FFPE human lung	FFPE human lung	FFPE human lung	FFPE human lung	FFPE human lung	FFPE human lung	FFPE human lung
Panel	980 plex	980 plex	980 plex	980 plex	980 plex	980 plex	980 plex	980 plex	980 plex
Number of slides analyzed	8	1	1	1	1	1	1	1	1
Total tissue area analyzed (μm^3)	753,480,217	97,014,620	97,014,620	97,014,620	97,014,620	64,676,413	145,521,930	90,546,979	64,676,413
Number of Field of Views (FOVs)	233	30	30	30	30	20	45	28	20
Total number of cells	800,327	100,292	106,660	100,264	93,795	91,972	150,504	73,997	82,843
% Cells passed QC (≥ 20 transcripts)	96.1	94.7	99.3	97.6	96.2	95.3	92.8	96.6	98.1
Number of cells analyzed	769,114	94,977	105,903	97,898	90,193	87,677	139,713	71,489	81,286
Transcripts detected	262,649,897	36,505,900	42,342,772	31,583,902	35,952,059	26,404,493	33,597,576	26,074,273	30,188,922
% of transcripts assigned to cells	79.2	81.7	83.3	82.1	80.9	85.7	71.2	69.1	77.7
Cellular transcripts/μm^3	0.446	0.523	0.59	0.456	0.423	0.456	0.28	0.393	0.547
Mean transcripts/cell	260	297	331	259	310	246	159	244	283
Mean negatives/cell/target	0.0429	0.0323	0.0611	0.0283	0.0238	0.0415	0.0412	0.076	0.0463
Genes detected	850	874	763	865	849	805	724	629	736
% Genes detected	88.5	91.0	79.5	90.1	88.4	83.9	75.4	65.5	76.7
Mean false call/cell/target	0.0092	0.0058	0.0096	0.0076	0.0100	0.0105	0.0083	0.0090	0.0144

Data Size:

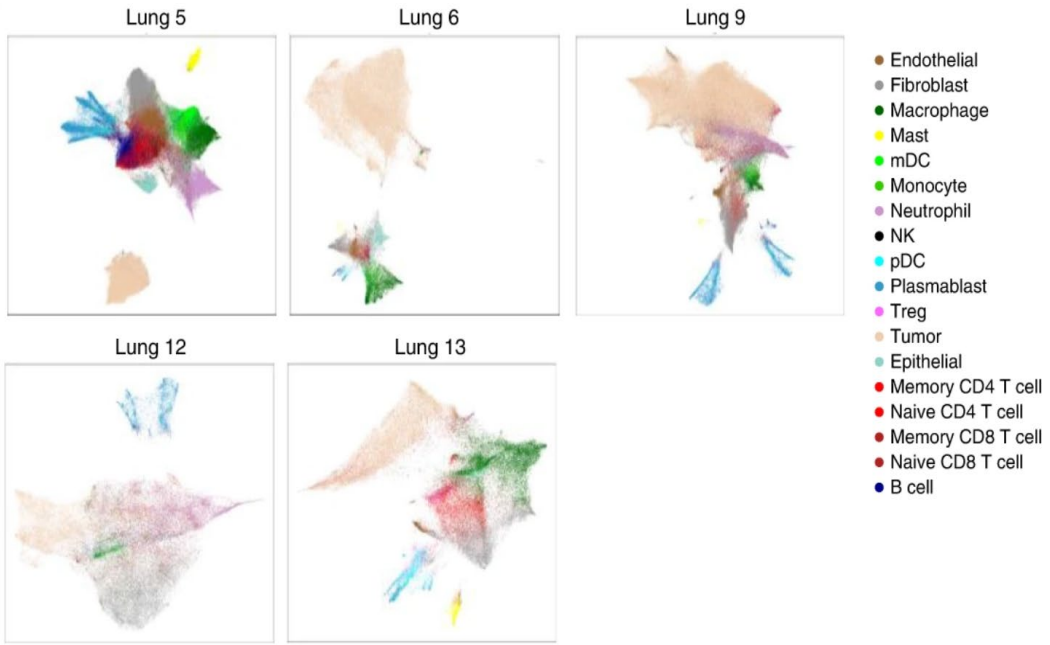
per FOV: ~2GB (233 FOVs: 466 GB)

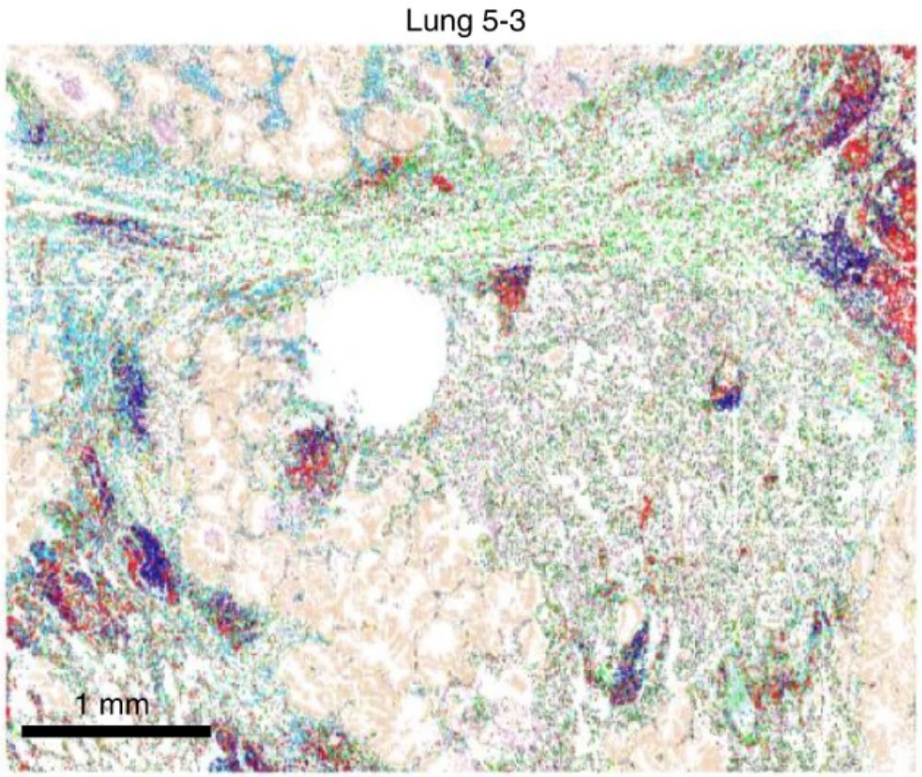
5 patients, 8 samples

He S, et al. Nat Biotechnology. 2022, 40(12):1794-1806.

Cell Typing:

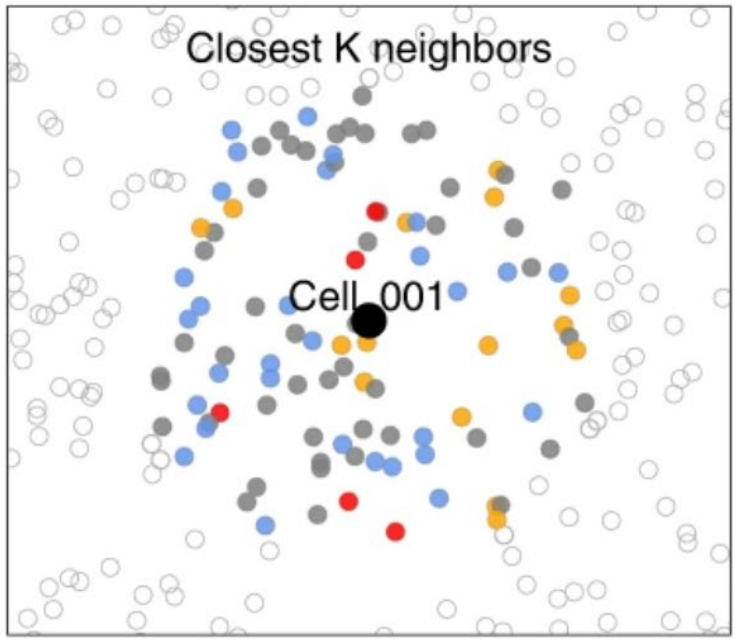
- Similar to scRNA-Seq's cell typing
- Challenges: 1000 RNA targets only, may not be sufficient for cell typing of all cells
  - ✓ 1000 RNA targets were partially optimized for cell typing
  - ✓ *de novo* clustering without assigning cell types?
- Cell typing **+ spatial information**





B/T Cells co-localization

cell neighborhoods: from spatial information



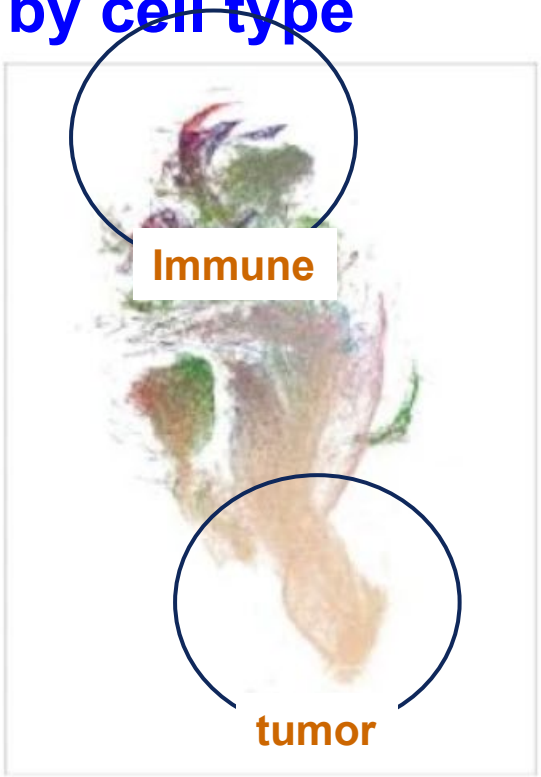
Neighborhood matrix:
number in closest K neighbors

Cell 001	5	15	30	50
Cell 002	...	...	...	...

For **every cell**, the nearest K neighbors are identified, and a summary of those neighbors is recorded (e.g., abundance of each cell type or # of cells in each cell type) → matrix of cells and **neighborhood characteristics (project-specific, e.g., average gene expression profiles within specific cell types)**.

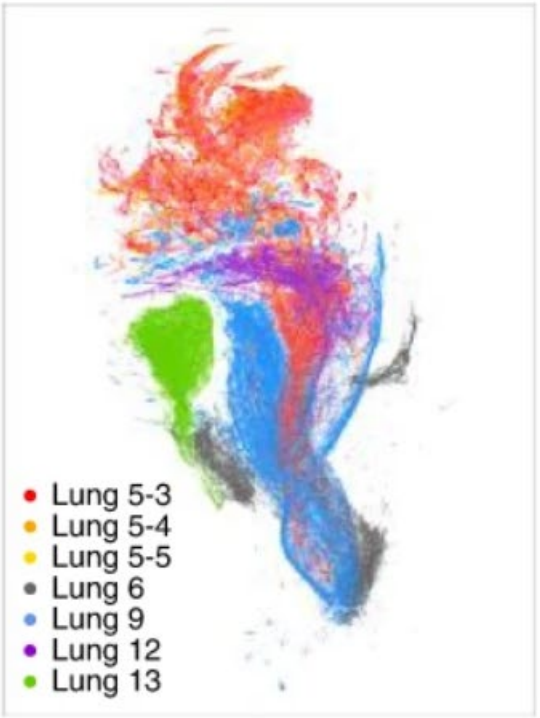
UMAP illustration of the neighborhood matrix: of all cells, all FOVs, across all samples (769,114 cells)

by cell type



- Endothelial
- Fibroblast
- Macrophage
- Mast
- mDC
- Monocyte
- Neutrophil
- NK
- pDC
- Plasmablast
- Treg
- Tumor
- Epithelial
- Memory CD4 T cell
- Naive CD4 T cell
- Memory CD8 T cell
- Naive CD8 T cell
- B cell

by sample

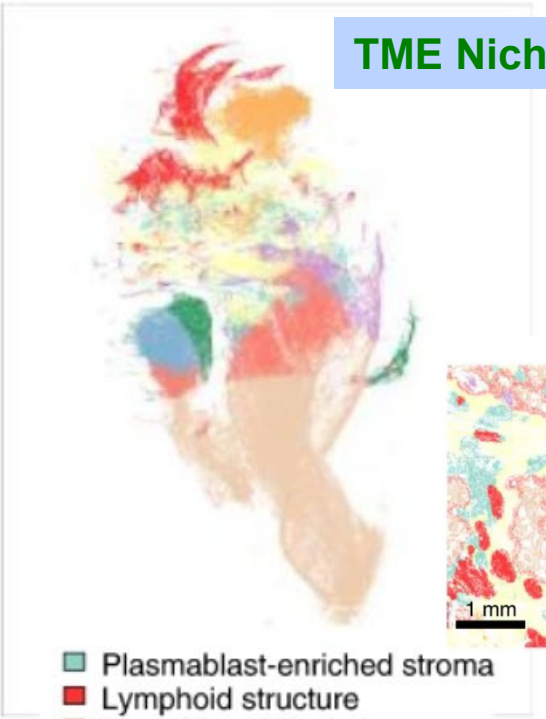


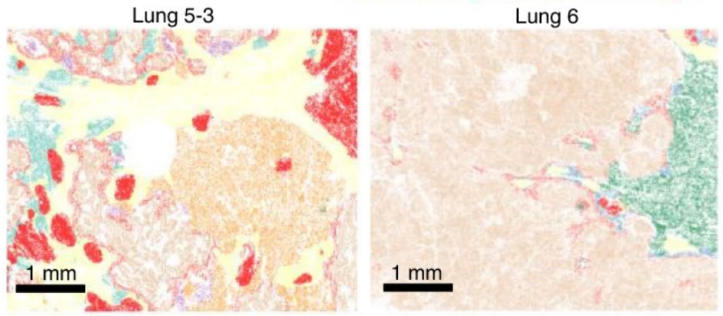
- Lung 5-3
- Lung 5-4
- Lung 5-5
- Lung 6
- Lung 9
- Lung 12
- Lung 13

Neighborhoods unique to or shared by samples

by neighborhood clusters

TME Niches

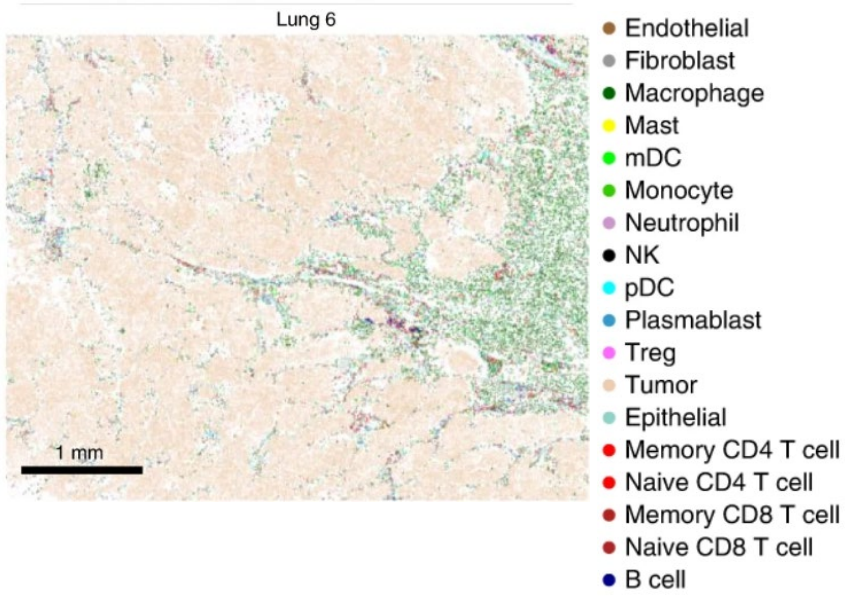




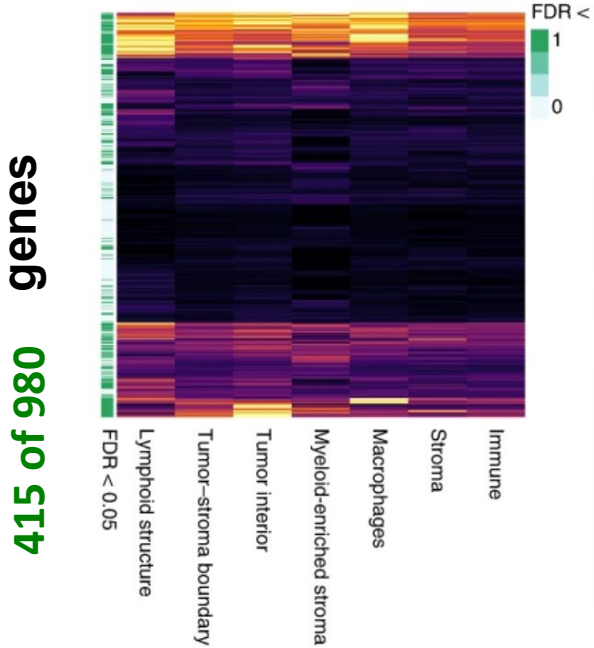
- Plasmablast-enriched stroma
- Lymphoid structure
- Myeloid-enriched stroma
- Tumor–stroma boundary
- Tumor interior
- Stroma
- Neutrophils
- Macrophages
- Immune-enriched

9 clusters

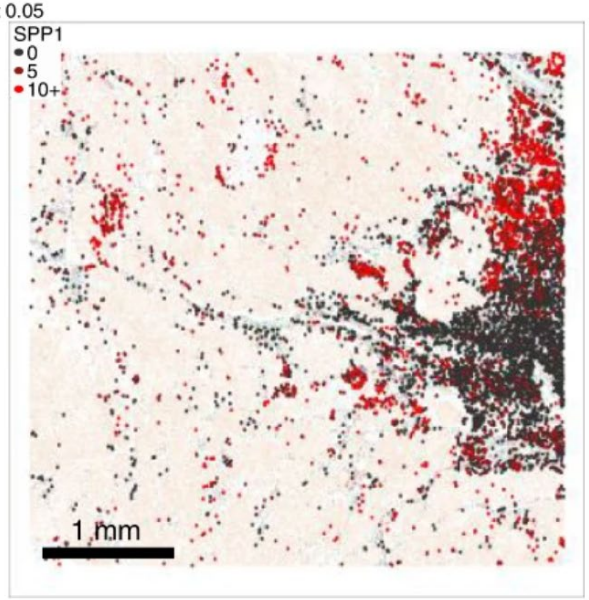
Biological questions that can be answered by neighborhoods or niches



Mean gene expression of macrophages in each niche



SPP1 ($p = 5 \times 10^{-61}$): two sub-types of macrophages

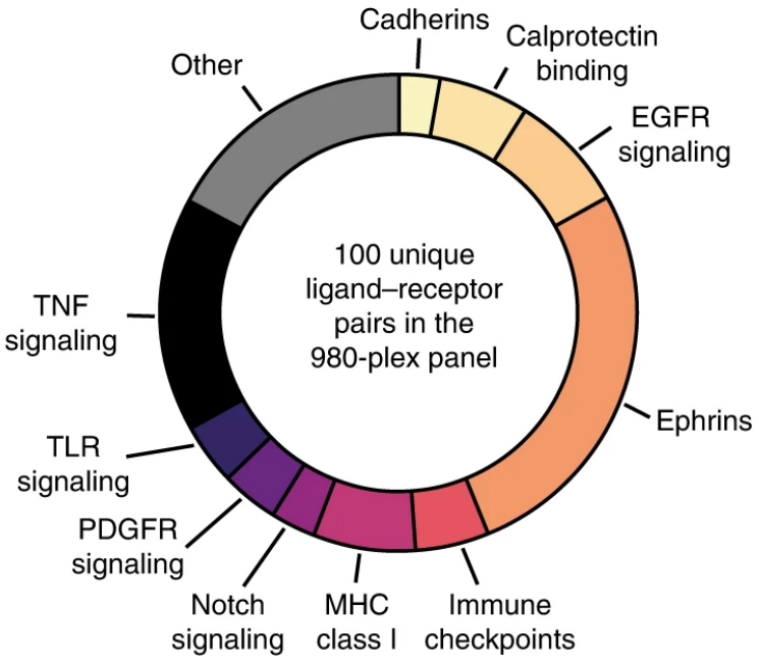


SPP1: mediates macrophage polarization and upregulate PD-L1 expression

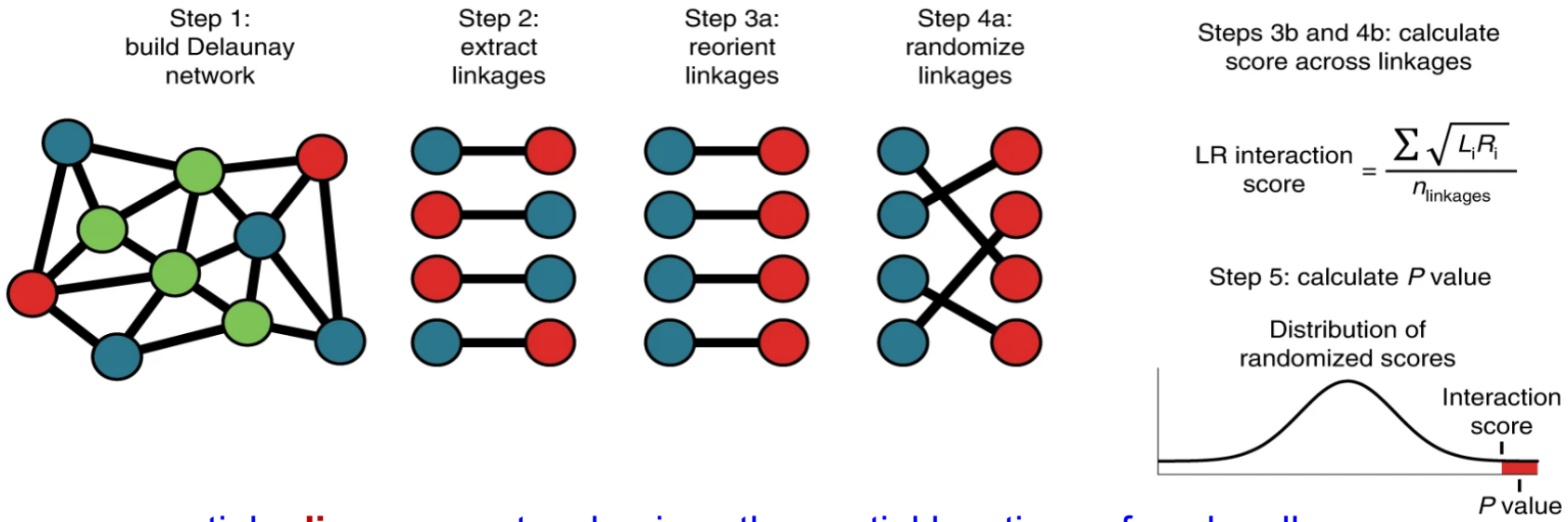
Changes of gene expression in macrophages between niches in Lung 6

Does a cell type change expression in response to neighbors?

Interactions between Tumor and T-cells

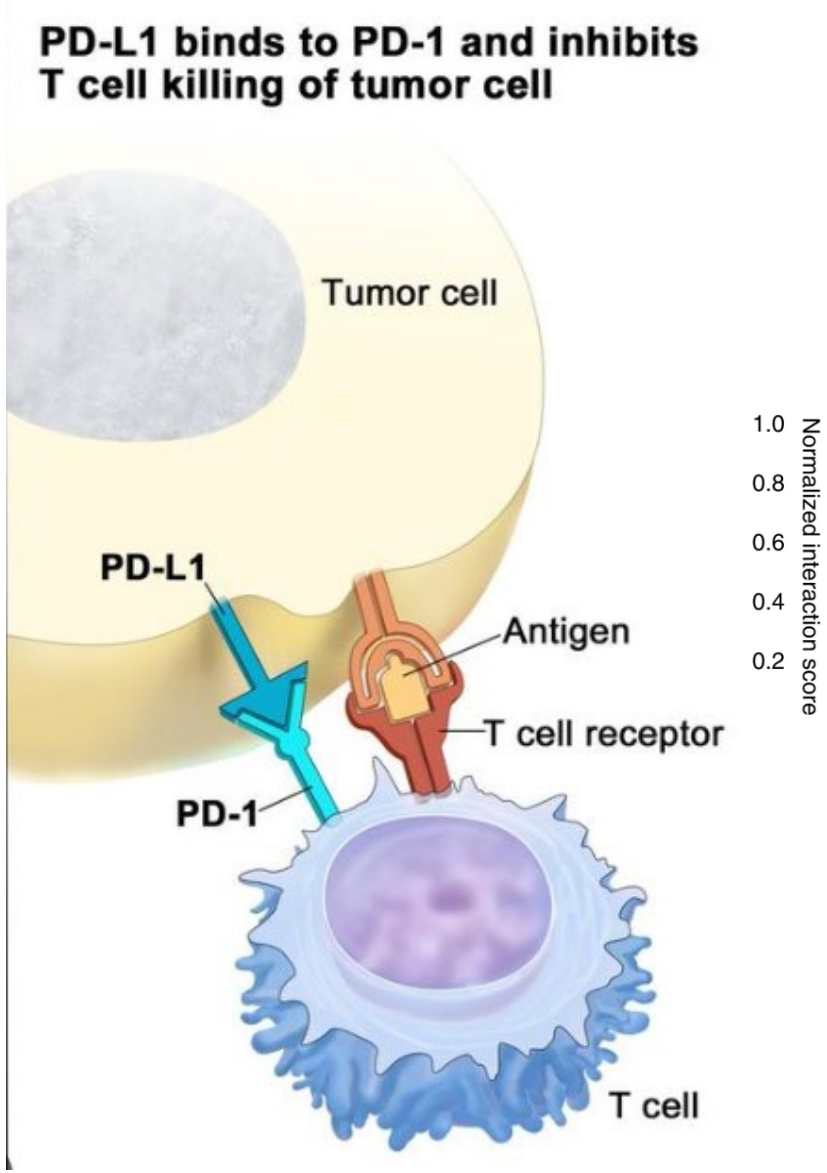


Distribution of ligand-receptor pairs between 980 RNA targets: including many tumor-immune interface pairs

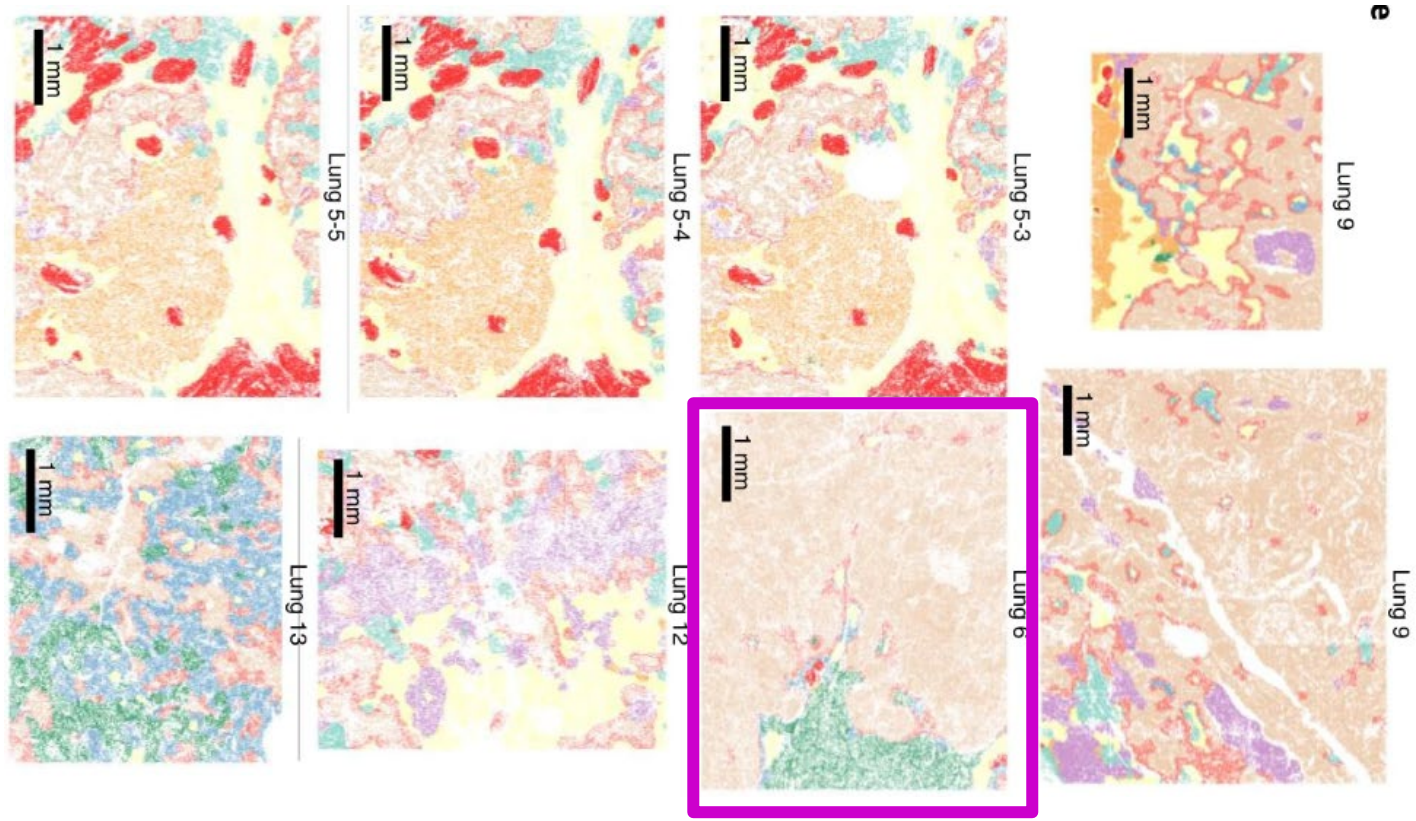


- a spatial **adjacency** network, given the spatial locations of each cell.
- For each pair of adjacent tumor -T cells: **an interaction score** using the geometric mean of their ligand and receptor expression, respectively.
- LR interaction score was calculated across all adjacent cell pairs for 100 distinct LR pairs
- An average score was calculated for each LR pair.
- Each average score was tested to determine whether it was enriched by the spatial arrangement of cells within the adjacency matrix: producing a null distribution of simulated average scores calculated using randomized adjacency networks.

Interactions between Tumor and T-cells



PD-L1/PD-1 (CD274/PDCD1): Low interaction in Lung 6, higher in Lungs 5, 13, 9, 12



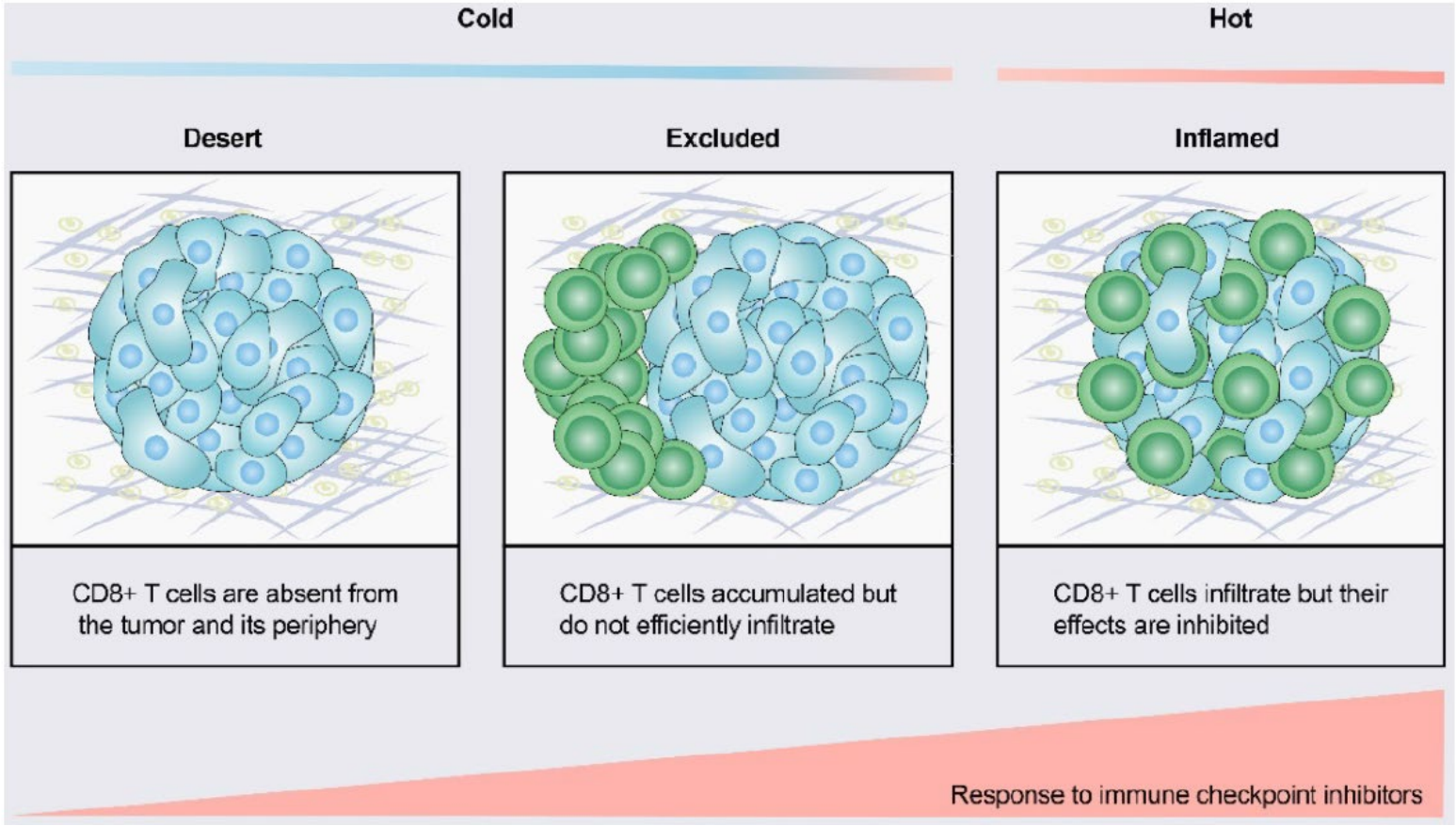
CosMx SMI Summary

- Spatial at true single cell resolution
- Biggest challenge: cell segmentation (foundation of all analyses)
- Throughput:
  - ✓ Current: ~1000 RNA
  - ✓ Future: full transcriptome (>18K genes, 2025)

Clinical Application of spatial transcriptomics

- Mulholland EJ, et al., *Redefining clinical practice through spatial profiling: a revolution in tissue analysis*. Ann R Coll Surg Engl. 2024;106(4):305-312.
- Zhang L, et al., *Clinical and translational values of spatial transcriptomics*. Signal Transduct Target Ther. 2022;7(1):111
- Hu W, et al. *Spatial transcriptomics in human biomedical research and clinical application.* Curr Med 2, 6 (2023).

Clinical Application of spatial transcriptomics



Immunotherapy

- Activate patient's own immune system
- Hot tumors are more likely to respond to immunotherapy
- e.g., lung 6 vs. others

Tumor immune phenotypes and immunotherapy outcome

Multi-Sample and Pseudo Bulk Analyses of scRNA data (Lab)

Murphy, A.E., Skene, N.G. *A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis.* *Nat Commun* **13**, 7851 (2022).

Zimmerman, K.D., Espeland, M.A. & Langefeld, C.D. *A practical solution to pseudoreplication bias in single-cell studies.* Nat Commun **12**, 738 (2021).

Squair, J.W., Gautier, M., Kathe, C. *et al. Confronting false discoveries in single-cell differential expression*. *Nat Commun* **12**, 5692 (2021).

Why Pseudo Bulk: *pseudo-replication bias in single cell transcriptome analyses*

- Cells from the same individual share common genetic and environmental backgrounds and are **not statistically independent**; therefore, they are subsamples or **pseudo-replicates**

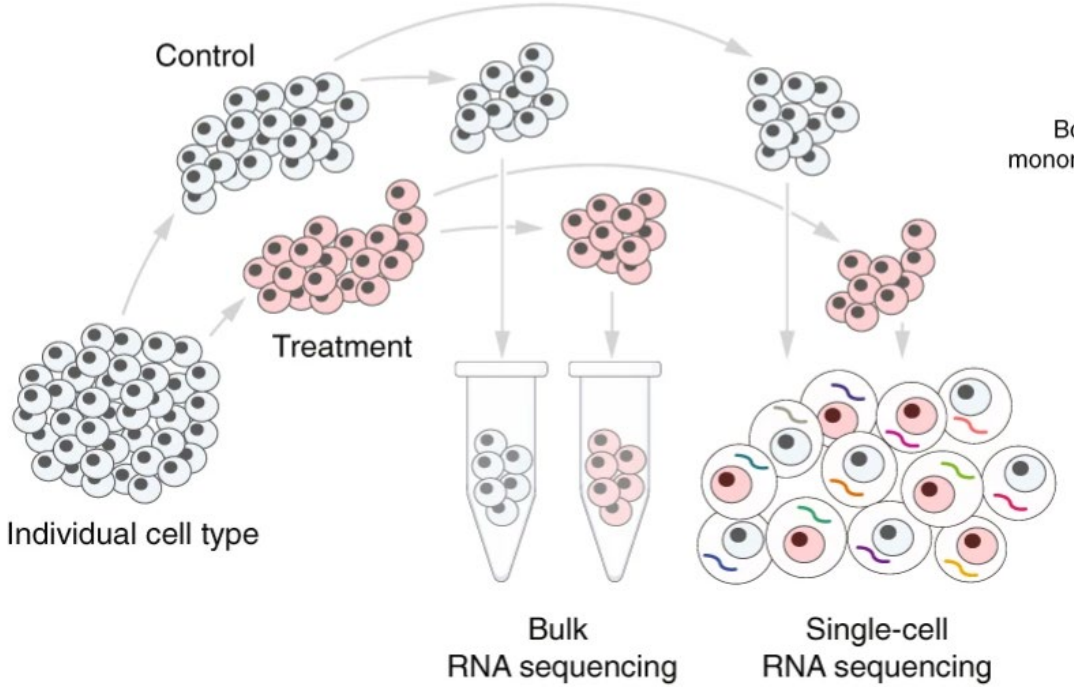
- Thus, single-cell data have a hierarchical structure that many current single-cell methods do not address, leading to biased inference, highly inflated type 1 error rates, and reduced robustness and reproducibility.

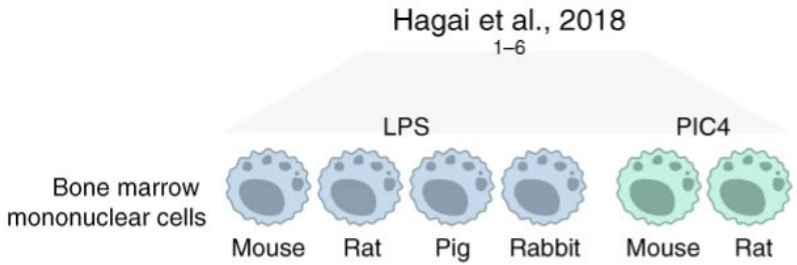
- Example: differential express between two groups of cells
  - ✓ 2000 cells *vs.* 5000 cells (e.g., tumor cells between two group of patients, or tumors between different regions)
  - ✓ LARGE sample size (2000 vs. 5000) → extremely SMALL p values (inflated)

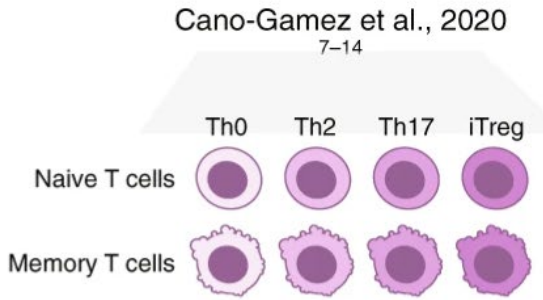
- **Pseudo Bulk**: A pseudo-bulk sample is formed by aggregating the expression values from a group of cells from the same individual (counter-intuitive?)

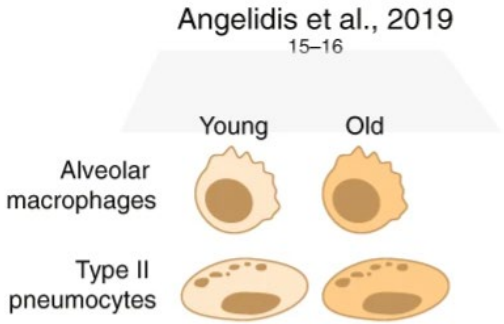
“Gold Standard” Data Sets:

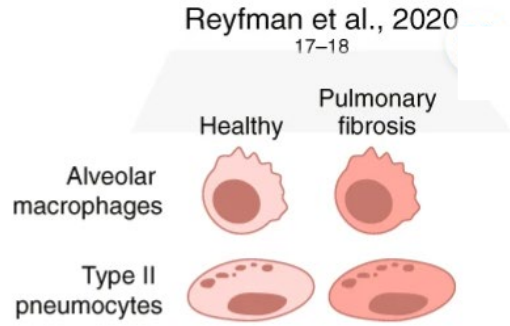
matched bulk and scRNA-seq performed on the same population of purified cells, exposed to the same perturbations, and sequenced in the same laboratories.







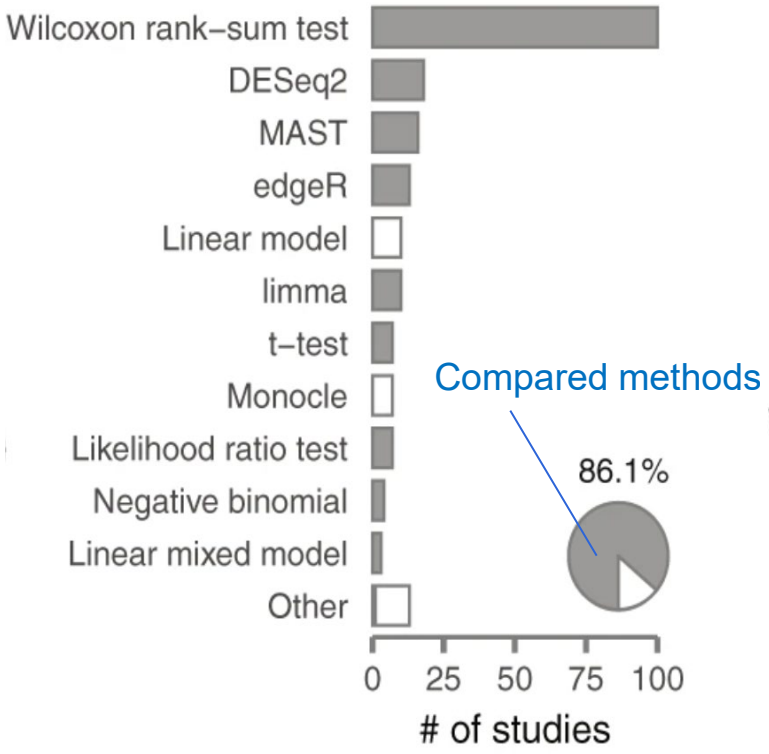


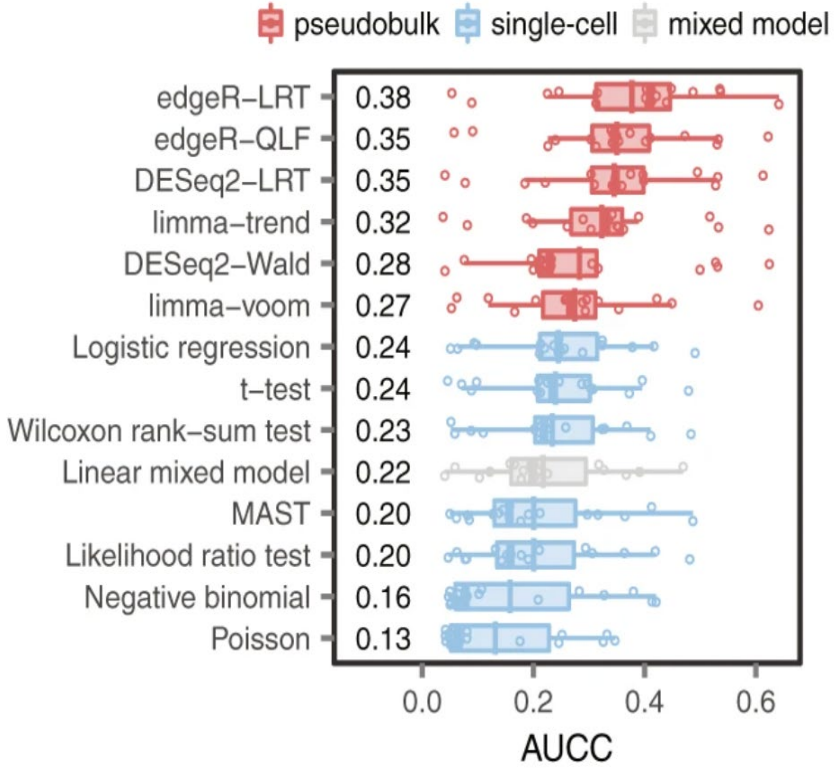


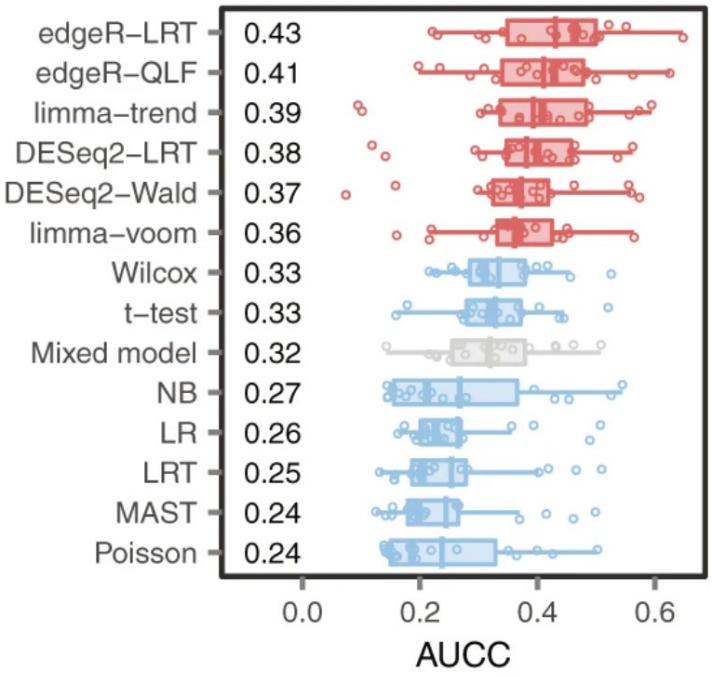
18 data sets from 4 publications

Squair, J.W., Gautier, M., Kathe, C. *et al. Confronting false discoveries in single-cell differential expression*. *Nat Commun* **12**, 5692 (2021).

Comparison of fourteen DE methods







AUCC of GO term enrichment

The most frequently used DE methods from 500 scRNA papers

Area under the concordance curve (AUCC): concordance between bulk and scRNA DE measurements

Alternatives to Pseudo-bulk Approaches

" … that pseudo-bulk aggregation methods are conservative and underpowered relative to mixed models. To compute differential expression within a specific cell type across treatment groups, we propose applying **generalized linear mixed models** with a random effect for individual …"

*Zimmerman, K.D., et al., A practical solution to pseudoreplication bias in single-cell studies. Nat Commun **12**, 738 (2021).*

NOTE: a 2022 Nature Communications paper supports the superior performance of pseudo-bulk approaches