Northwestern Medicine®

The New Space Race in Genomics

2023 ICSA Midwest Chapter & NIC-ASA Joint Fall Meeting

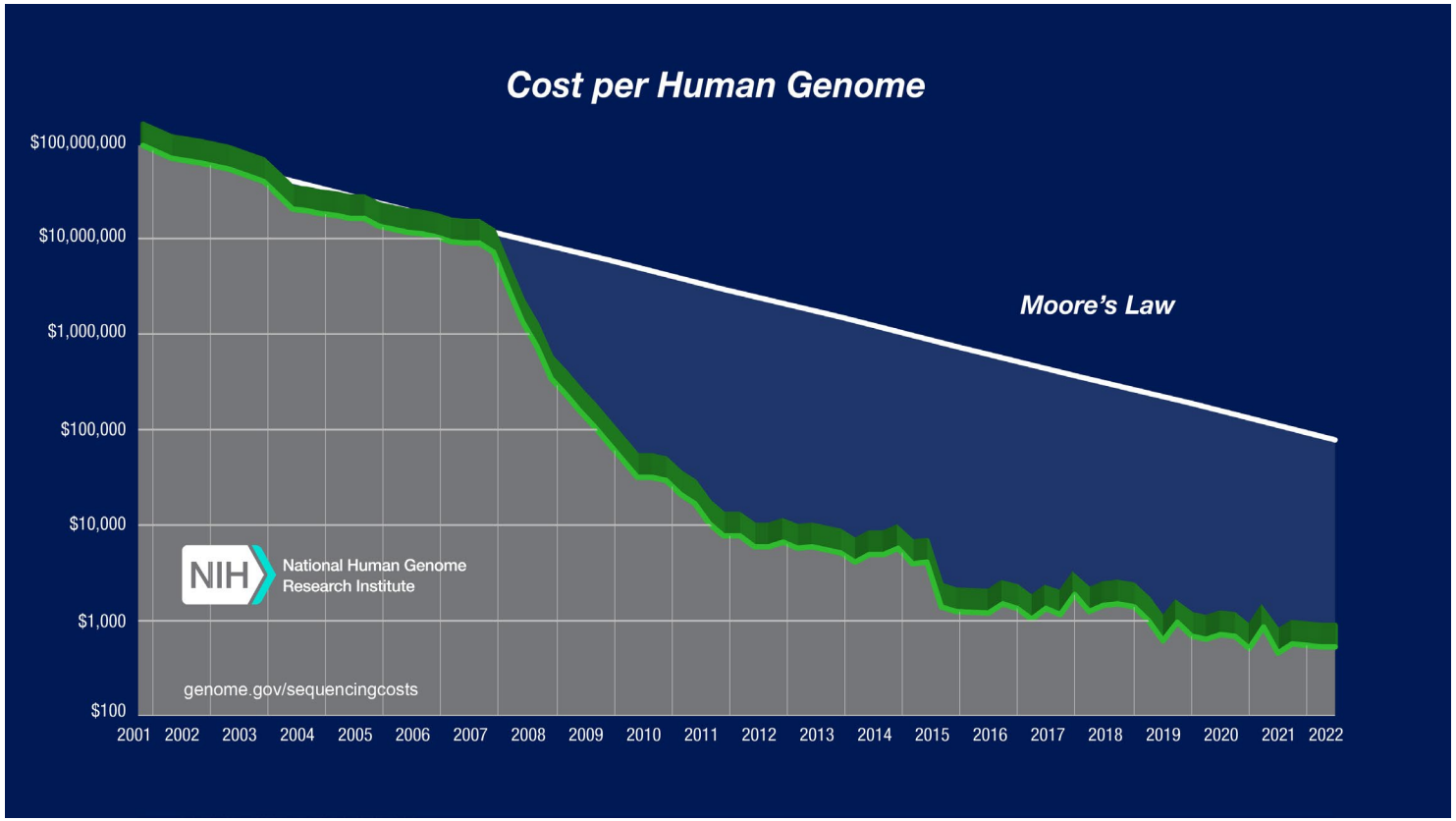
Oct. 13, 2023

Xinkun Wang, Ph.D., Director

Northwestern University Sequencing (NUSeq) Core

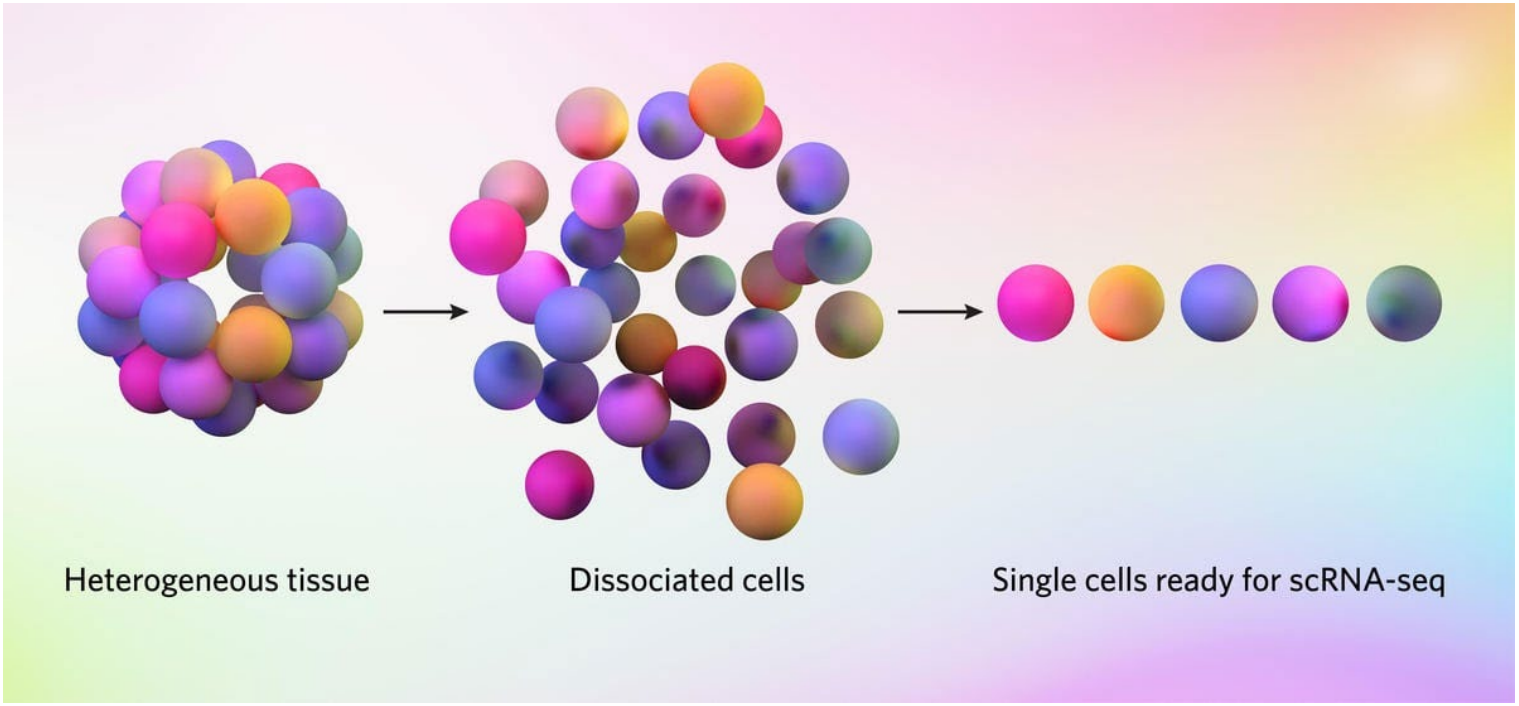
Recent Genomics Technology Trends

- More data at lower cost: sequencing technologies are still rapidly evolving



Recent Genomics Technology Trends

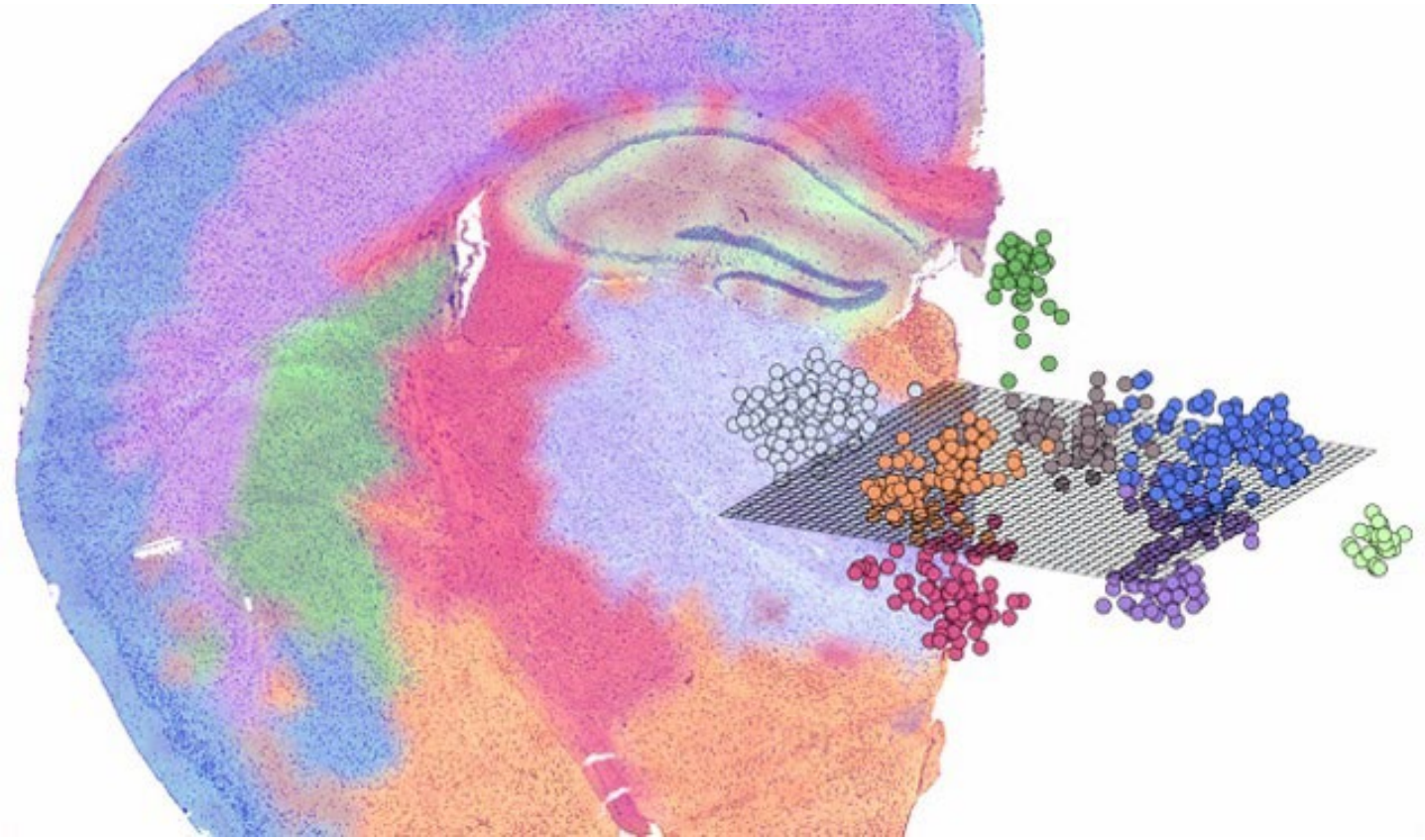
- Higher resolution: from bulk tissue/cell population to individual cells



Source: https://www.the-scientist.com/university/single-cell-sequencing-in-a-nutshell-71048

Recent Genomics Technology Trends

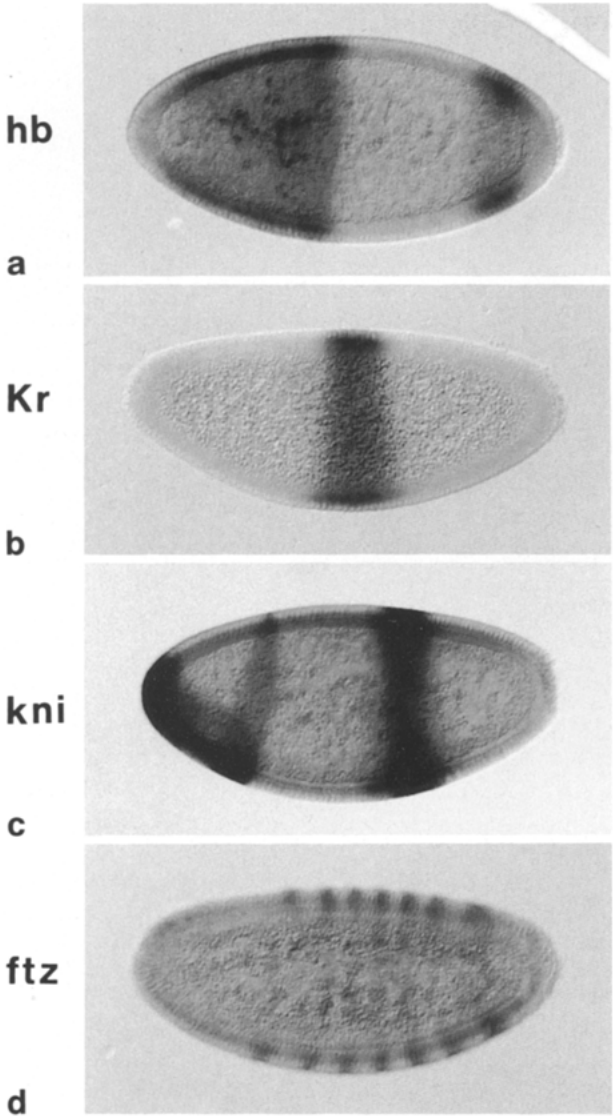
- Race into space: Spatial context matters



Soruce: https://www.science.org/content/webinar/mining-transcriptome-using-spatial-transcriptomics-comprehensive-2d-3d-visualization-all-mrnas-tissue-sections

Spatial Transcriptomics

- Not a brand new technology: started from 1970s, but at a much lower throughput – one gene at a time
- The current surge in spatial transcriptomics is due to the demand to simultaneously quantify a large number of genes, from hundreds, thousands, to the entire genome



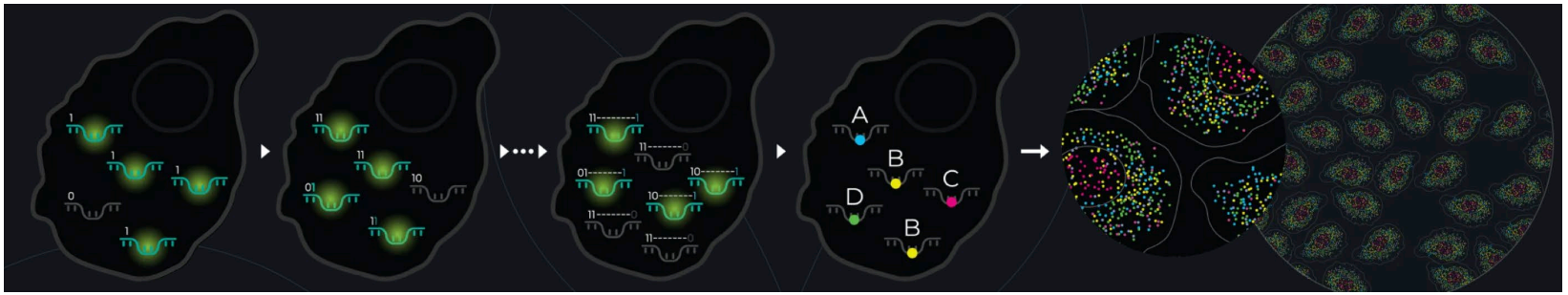
Localization of specific RNAs in Drosophila embryos.

Chromosoma (Berl) (1989) 98:81-85

Key Spatial Transcriptomics Technologies

Four major approaches

- Highly multiplexed single molecule fluorescent in situ hybridization (smFISH)
  - Augmentation of the traditional in situ hybridization approach
  - Example: MERFISH (Vizgen)

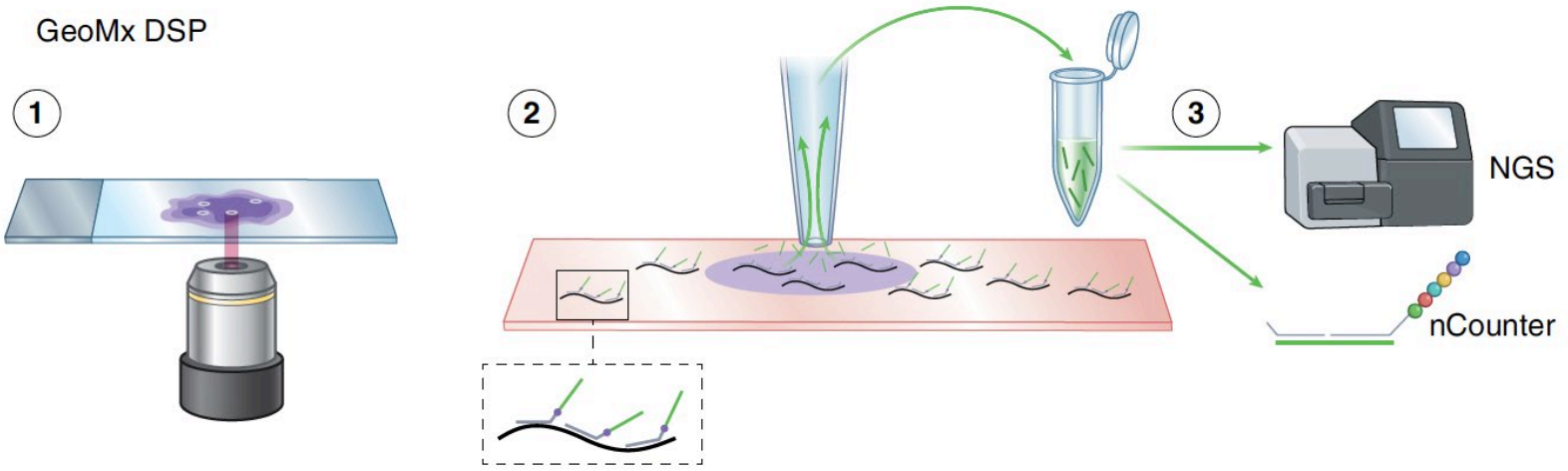


Source: Vizgen

Key Spatial Transcriptomics Technologies

Four major approaches (2)

- Selection of region of interest
  - Microdissection or capture of ROI
  - Example: Nanostring GeoMx Digital Spatial Profiler

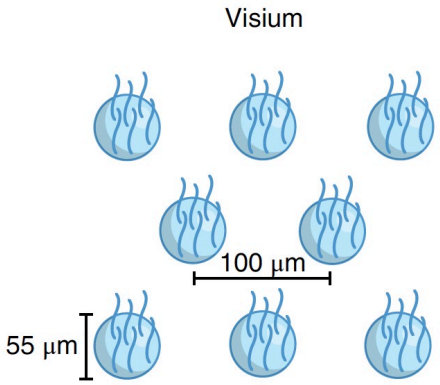


Nature Methods volume 19, pages534–546 (2022)

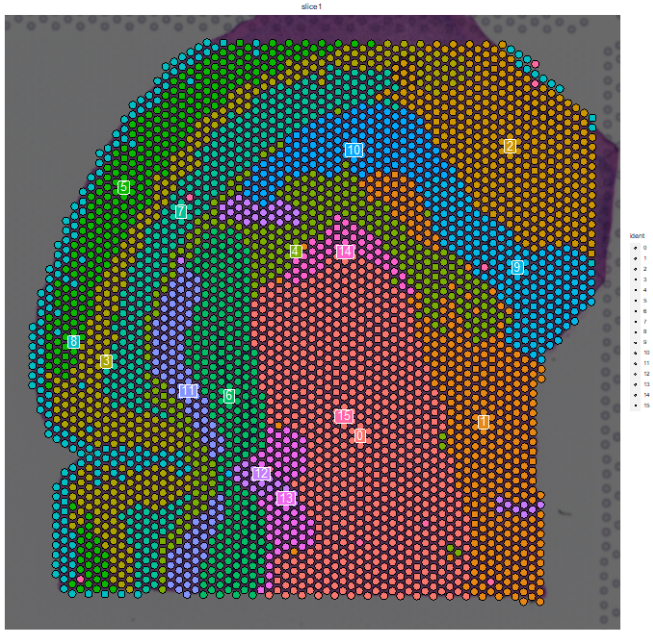
Key Spatial Transcriptomics Technologies

Four major approaches (3)

- Spatial barcoding followed by sequencing
  - Each spatial region is indexed by specific barcode
  - Example: Visium from 10x Genomics



Nature Methods volume 19, pages534–546 (2022)

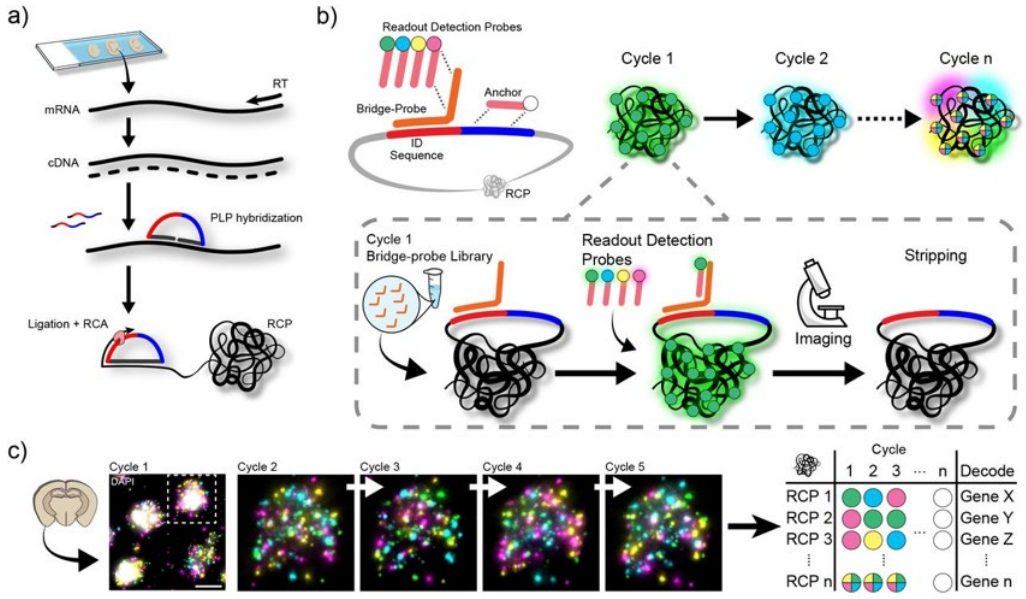


NUSeq data

Key Spatial Transcriptomics Technologies

Four major approaches (4)

- In Situ Sequencing
  - Directly sequence transcripts on the spot
  - Example: Xenium from 10x Genomics



Gyllborg *et al,* Nucleic Acids Research (2020)

Key Parameters

- Resolution: from 50 µm to 10 nm
- Number of genes detected: from a few hundred to the entire genome
- Detection efficiency: from 5% to close to 100%
  - Molecular capture and optical crowding are limiting factors

Data Format

- Gene-by-location matrix

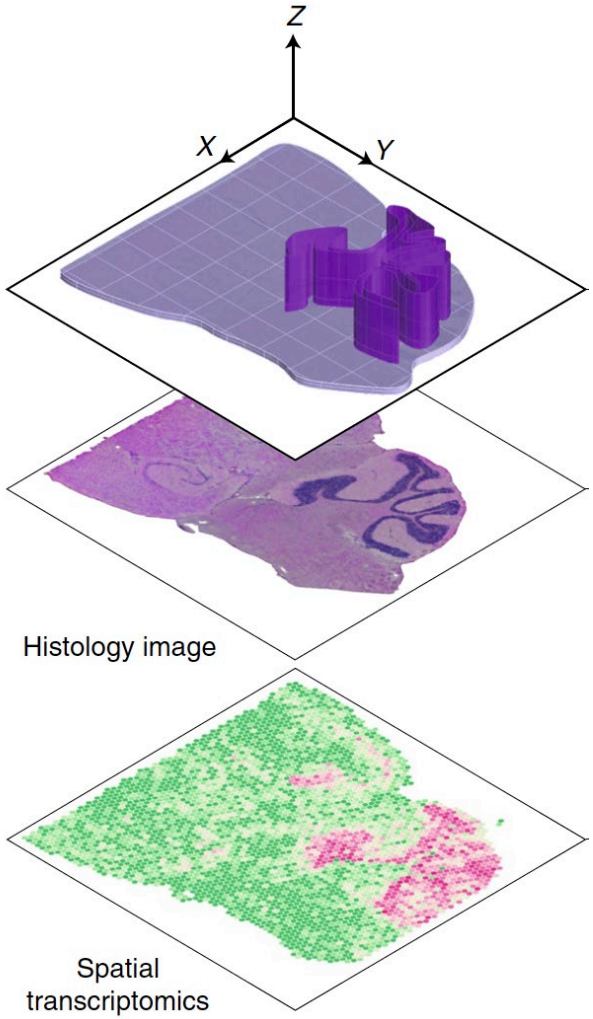
	Spatial coordinate 1	*Spatial coordinate 2*	*…*	*Spatial coordinate m*
Gene 1	18	0	…	3
Gene 2	2	3	…	5
…	…	…	…	…
Gene n	0	8	…	20

Spatial Transcriptomics Data Distributions

- Commonly used data distributions
  - Negative binomial distribution, especially zero-inflated negative binomial distribution
  - Poisson distribution
  - Gaussian distribution
  - Gamma distribution
  - Spatial point process models
  - Others
  - Can be platform or even gene dependent

Major Goals

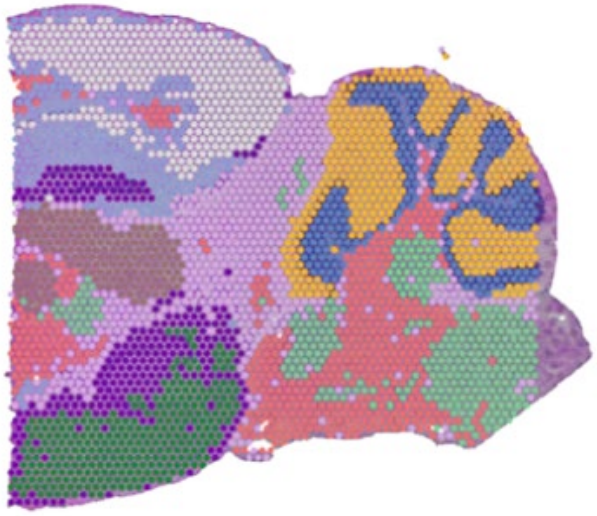
- Identification of spatially variable genes
  - Spatial location
  - Histology image
  - Gene expression profile



Nature Methods volume 18, pages1342–1351 (2021)

Major Goals

- Identification of spatial domains



Spatial domains

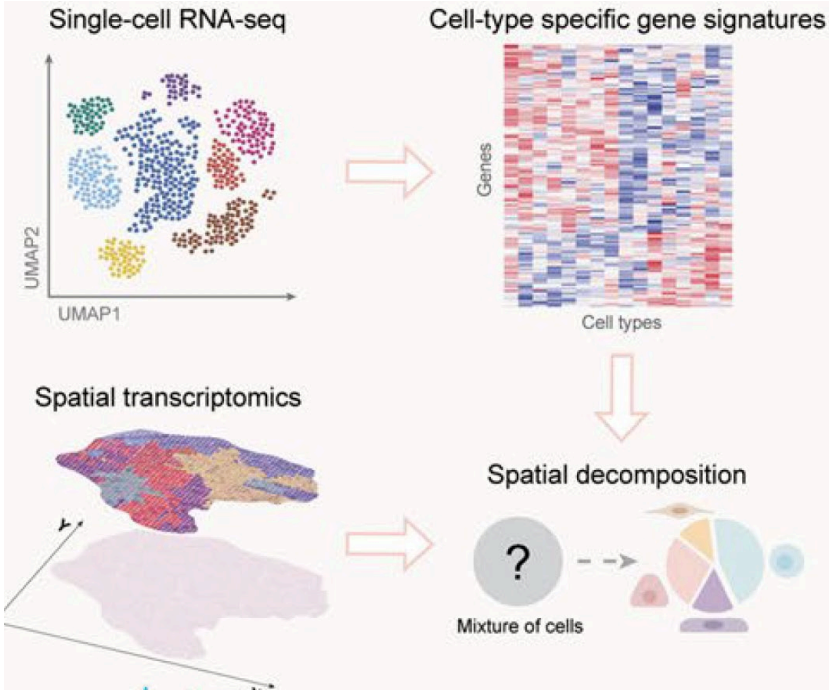
Nature Methods volume 18, pages1342–1351 (2021)

Major Goals

- Cell type identification (if spatial resolution can achieve single cell)
  - Based on dimensionality reduction, with the assumption that cells of the same type are similar to each other in terms of gene expression and therefore cluster together

Major Goals

- Spatial decomposition: if spatial resolution larger than a cell, determination of what cell types are in each capture region, i.e., cellular deconvolution



Genome Biology volume 23, Article number: 83 (2022)

Other Goals

- Gene imputation for missing genes
  - Often based on the use of single cell RNA-seq data
- Predicting spatial location of cells from single cell RNA-seq data
- Inference of cell-cell interactions

Challenges

- Inherently spatial, and gene expression can vary significantly from one location to another in the same tissue
- Signal sparsity: many gene expressions are missed with signal drop-out
  - Low capture efficiency
- High dimensional: large number of genes
- Low resolution for whole transcriptome platforms (this will improve over time)
- How to integrate spatial transcriptomics data with other data, including single cell sequencing, other –omics, and pathology (clinical) data

Northwestern Medicine®

Thank you! Questions?