

Introduction to single-cell multi-omics and spatial transcriptomics and corresponding challenges

Qin Ma, PhD

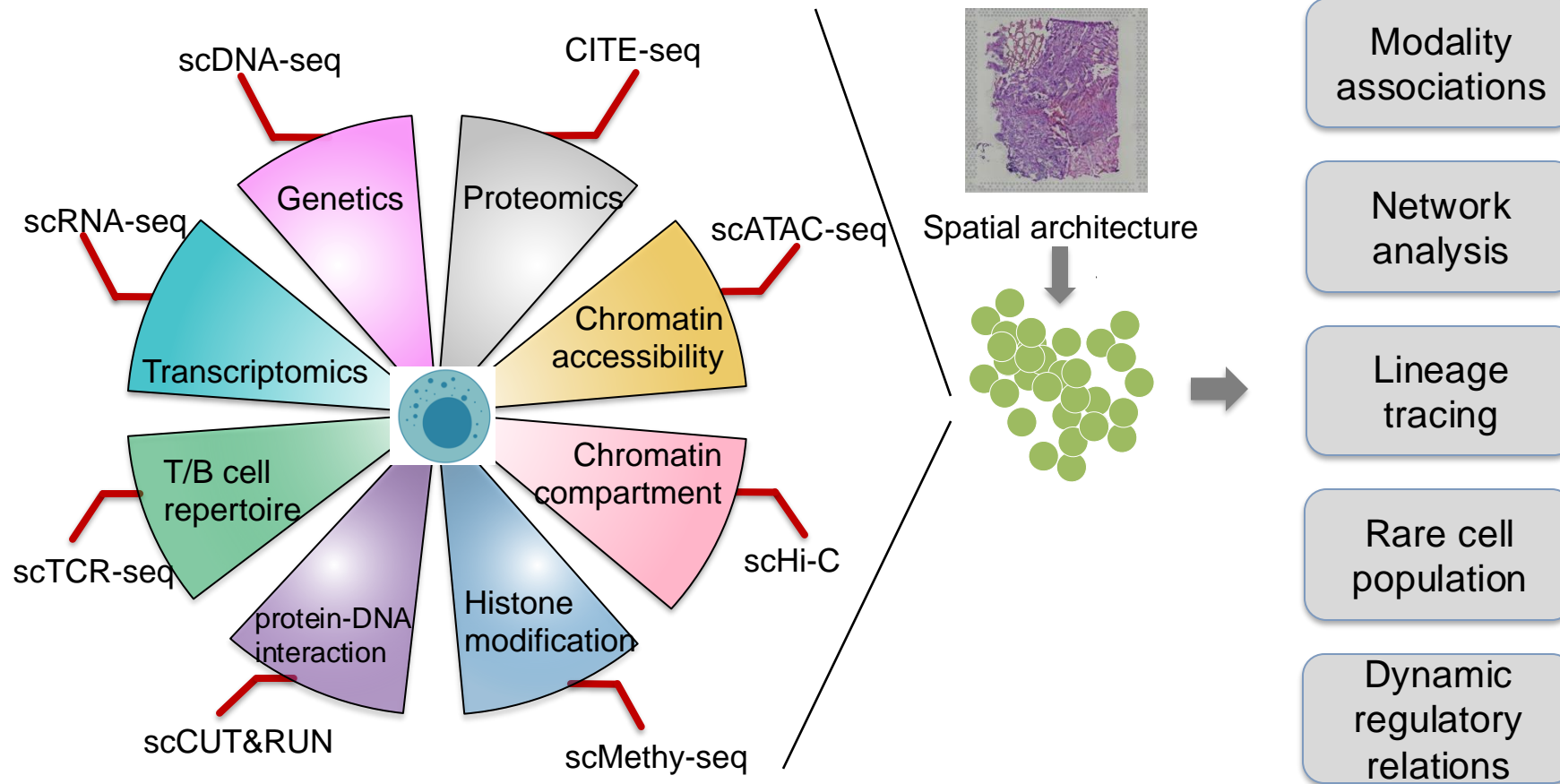
Professor and Section Chief of Computational Biology and Bioinformatics
Department of Biomedical Informatics

Leader of the Immuno-Oncology Informatics Group
Pelotonia Institute for Immuno-Oncology



THE OHIO STATE UNIVERSITY
WEXNER MEDICAL CENTER

scMulti-omics enhance and enable various biological analyses



Editorial | Published: 06 January 2020

Method of the Year 2019: Single-cell multimodal omics

[Nature Methods](#) 17, 1 (2020) | [Cite this article](#)

Editorial | [Published: 06 January 2021](#)

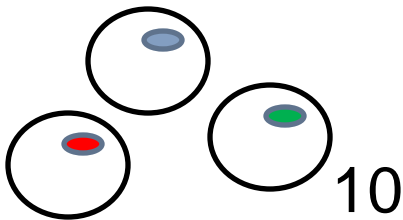
Method of the Year 2020: spatially resolved transcriptomics

[Nature Methods](#) 18, 1 (2021) | [Cite this article](#)

Understanding complexity using sequencing



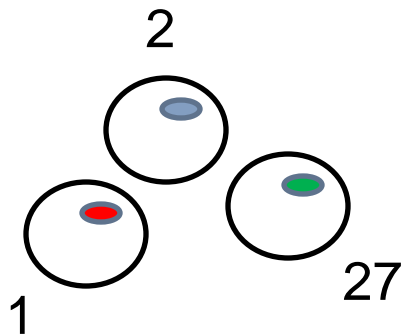
Bulk Genomics



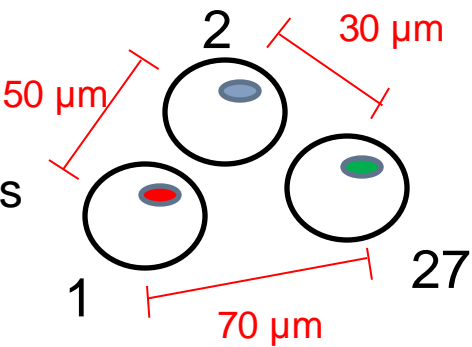
Complex Tissue



Single Cell Genomics


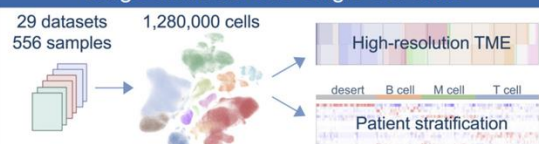
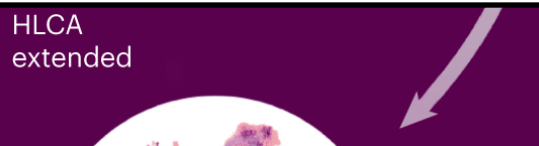

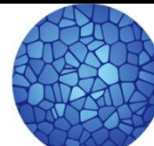


Spatial Transcriptomics








Big data in single-cell biology and spatial biology

Single-cell database

 Curated Cancer Cell Atlas Collected, annotated and analyzed cancer scRNA-seq datasets	<ul style="list-style-type: none">• Pan cancer• 2.3 M cells
Large-scale NSCLC single-cell atlas 29 datasets 556 samples 1,280,000 cells 	<ul style="list-style-type: none">• Lung cancer• 1.3 M cells
HLCA extended 	<ul style="list-style-type: none">• Lung cells• 2.4 M cells
 CELL × GENE DISCOVER	<ul style="list-style-type: none">• Human cells• 59.4 M
 HUMAN CELL ATLAS	<ul style="list-style-type: none">• Human cells• 50.1 M

Spatial database database

	<ul style="list-style-type: none">• All species• 61.3 M spots
 STomicsDB	<ul style="list-style-type: none">• 128 tissues• 221 datasets• 17 species
 SOAR	<ul style="list-style-type: none">• 2,785 samples• 304 datasets• 40 tissues
 ENABLE MEDICINE	<ul style="list-style-type: none">• 20k+ samples• 20+ disease• 200+ proteins
 AQUILA	<ul style="list-style-type: none">• 6.5K ROIs• 15.7 M Spots• 30 disease

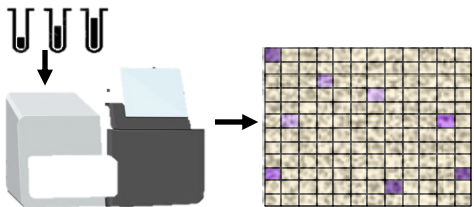
Three “V” define big data:

- High Volume
- High Velocity
- High Variety

 SenNet	 HuBMAP Human BioMolecular Atlas Program	 HTAN HUMAN TUMOR ATLAS NETWORK
---	--	--

Challenges in single-cell and spatial omics data: data quality and scalability

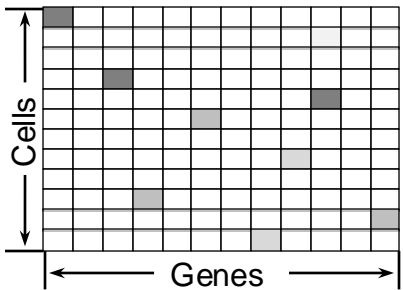
Technology noise



- Technology noise in single-cell and spatial omics data, such as amplification biases, batch effects, and droplet-based artifacts, can obscure true biological signals by introducing inconsistencies from sample preparation to data acquisition.

Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	✓

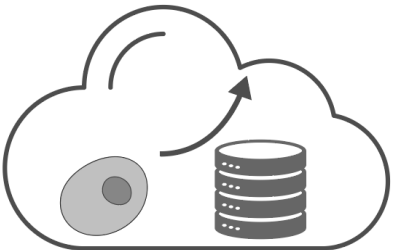
Data sparsity



- Data sparsity in single-cell and spatial omics refers to the phenomenon where many measurements (e.g., gene expression levels) contain a high proportion of zeros, indicating that no transcripts were detected for many genes in many cells.

Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	

Large scalability



- This scalability issue involves not only computational demands, such as storage, memory, and processing speed but also analysis challenges in managing the high dimensionality and complexity of the data to extract meaningful biological insights.

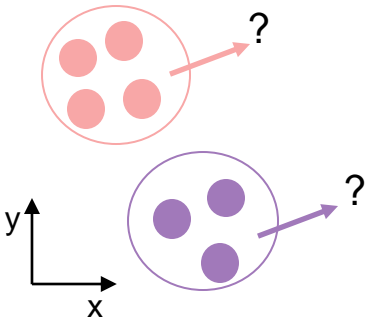
Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	✓

Challenges in single-cell and spatial omics data: limited annotations and data heterogeneity

Limited labeled data

Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	✓

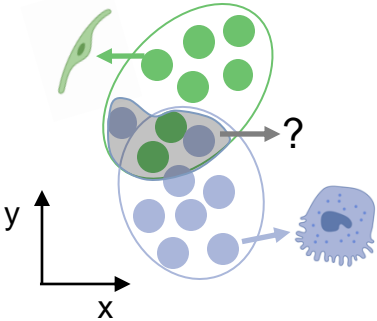
- Limited labeled data refers to the scarcity of annotated samples where specific cell types, states, or spatial features are identified and verified, typically through expert knowledge or additional experimental validation.



Heterogeneity

Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	✓

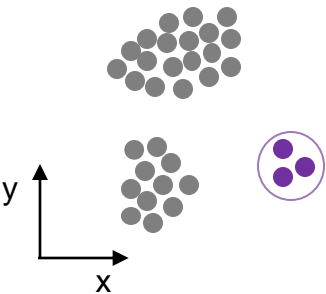
- Heterogeneity refers to the diverse and varied nature of biological samples at the cellular or subcellular level. This diversity can be due to differences in cell types, states, genetic expressions, or microenvironmental conditions within a tissue.



Rare population

Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	✓

- Rare cell populations are infrequent yet crucial groups of cells within larger populations. Their identification offers profound insights into disease mechanisms and therapeutic targets.



Challenges in single-cell and spatial omics data: integration and resolution gaps

Incompatible resolution and transcripts counts

- It refers to the trade-off where high-resolution data provides detailed spatial or cellular information but often captures fewer transcripts per region or cell. In contrast, lower-resolution data includes more transcript counts but less spatial or single-cell detail.

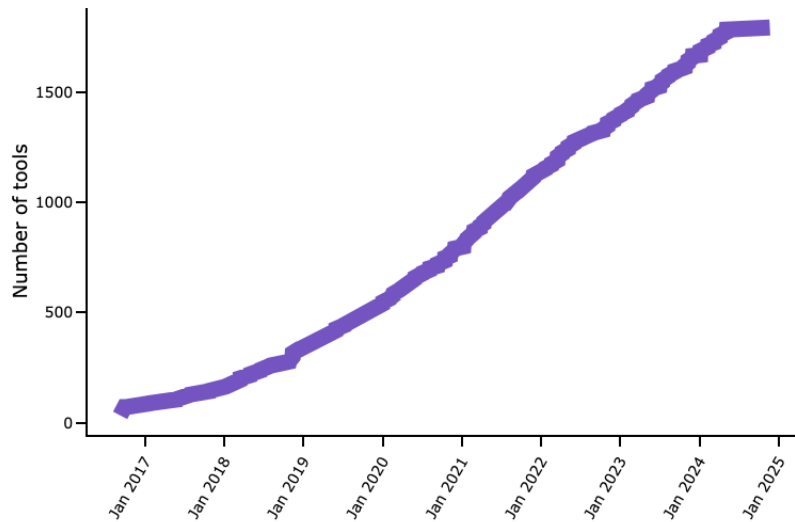
Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	✓

Integration gaps among complex multi-modalities

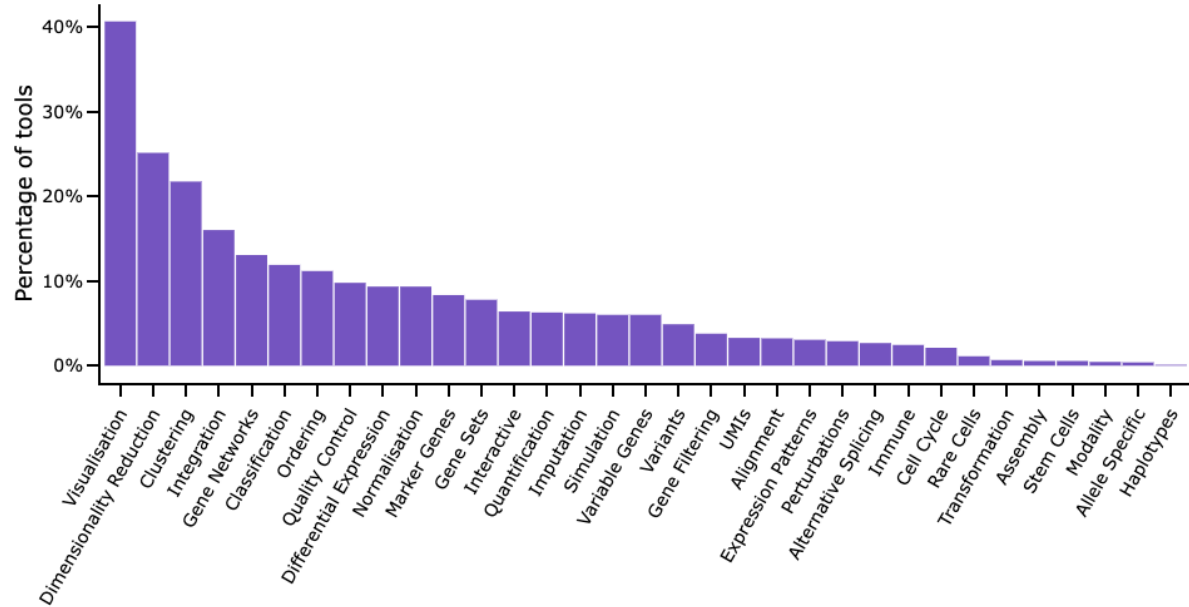
- It refers to the challenges in effectively combining data from different omics (like genomics, transcriptomics, and metabolomics) or from different technological platforms. These gaps can make it difficult to construct a comprehensive view of cellular function and organization.

Single-cell data	Spatial data	
	Single-cell level	Multi-cellular level
✓	✓	✓

A lot of tools have been developed to address one particular or some challenges above



We currently track **1796** tools...

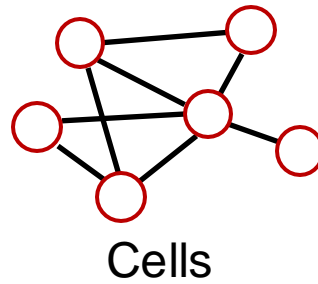


...in over **30** categories

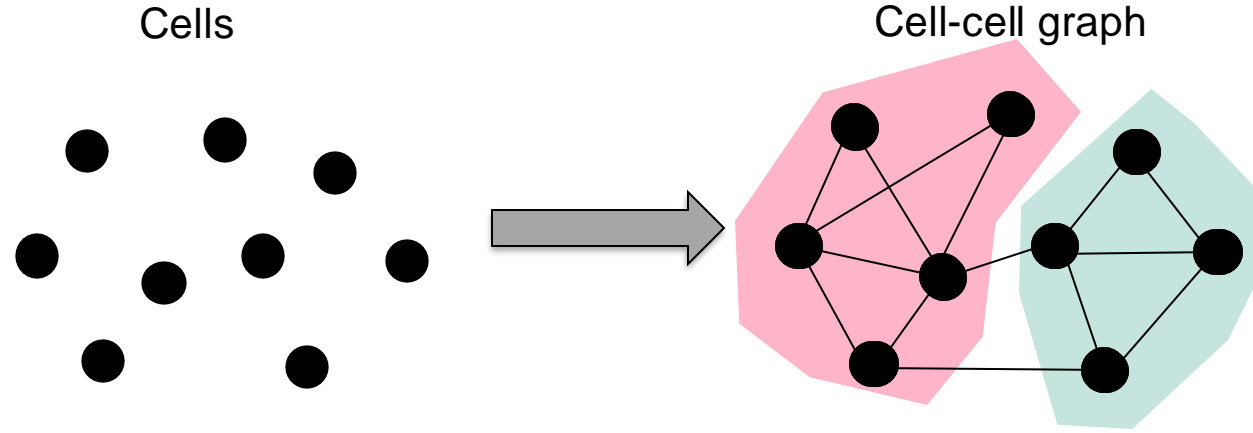
graph representation learning in our group

Minimal graph representation of cellular heterogeneity:

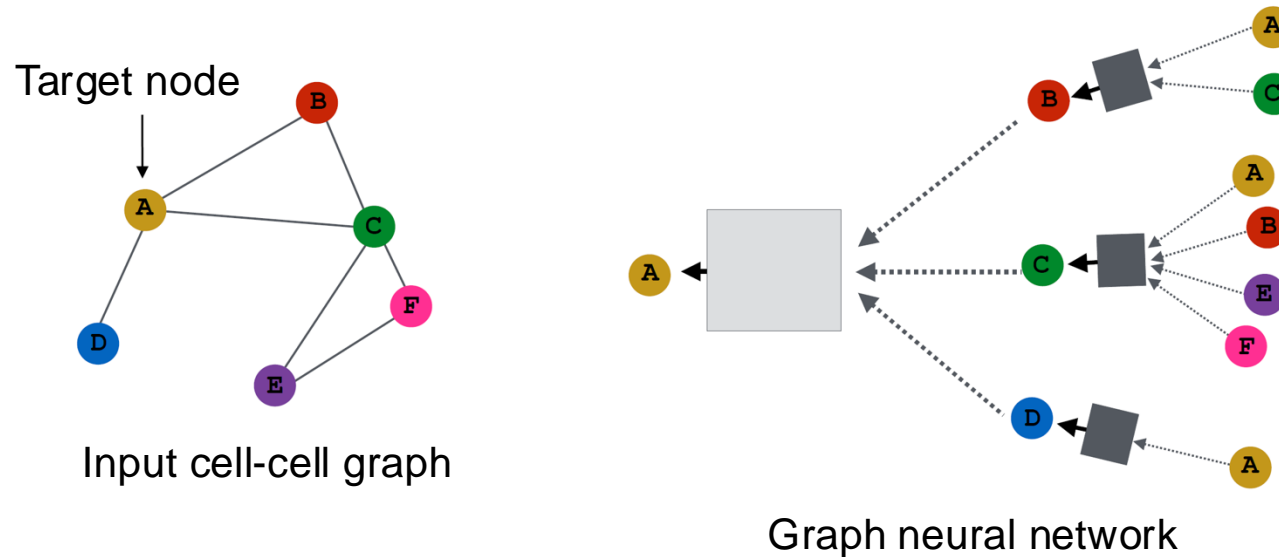
cell graph



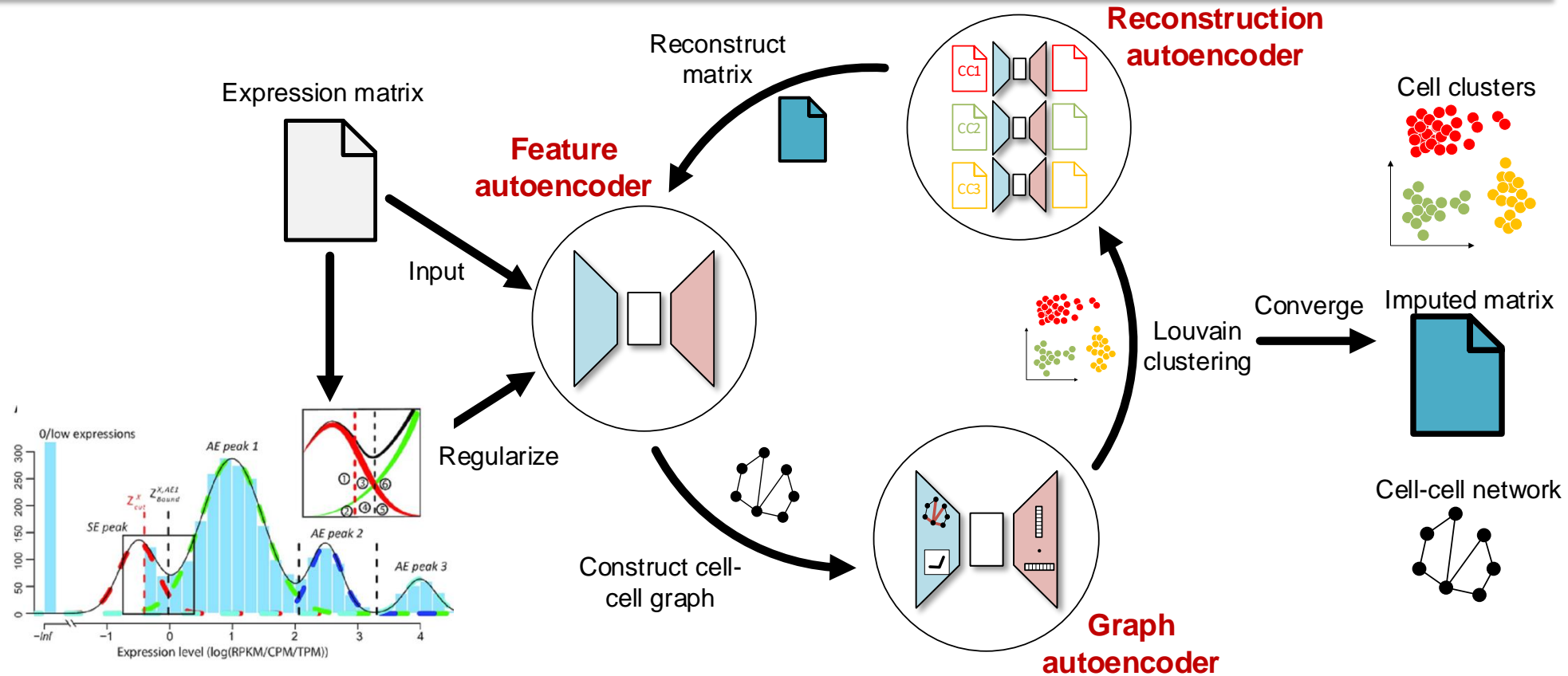
Graph Neural Network for gene imputation and cell clustering



However, traditional statistical models and bioinformatic methods consider **cells are independent** and do not reflect the **influence of cell-cell relations**.



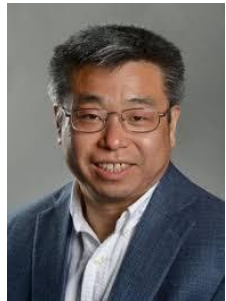
scGNN: single-cell Graph Neural Network for gene imputation and cell clustering



Dr. Chi Zhang
Asso Prof, OHSU



Dr. Anru Zhang
Asso Prof, Duke



Dr. Dong Xu
Professor, MU



Dr. Juexin Wang
Assis Prof, IU

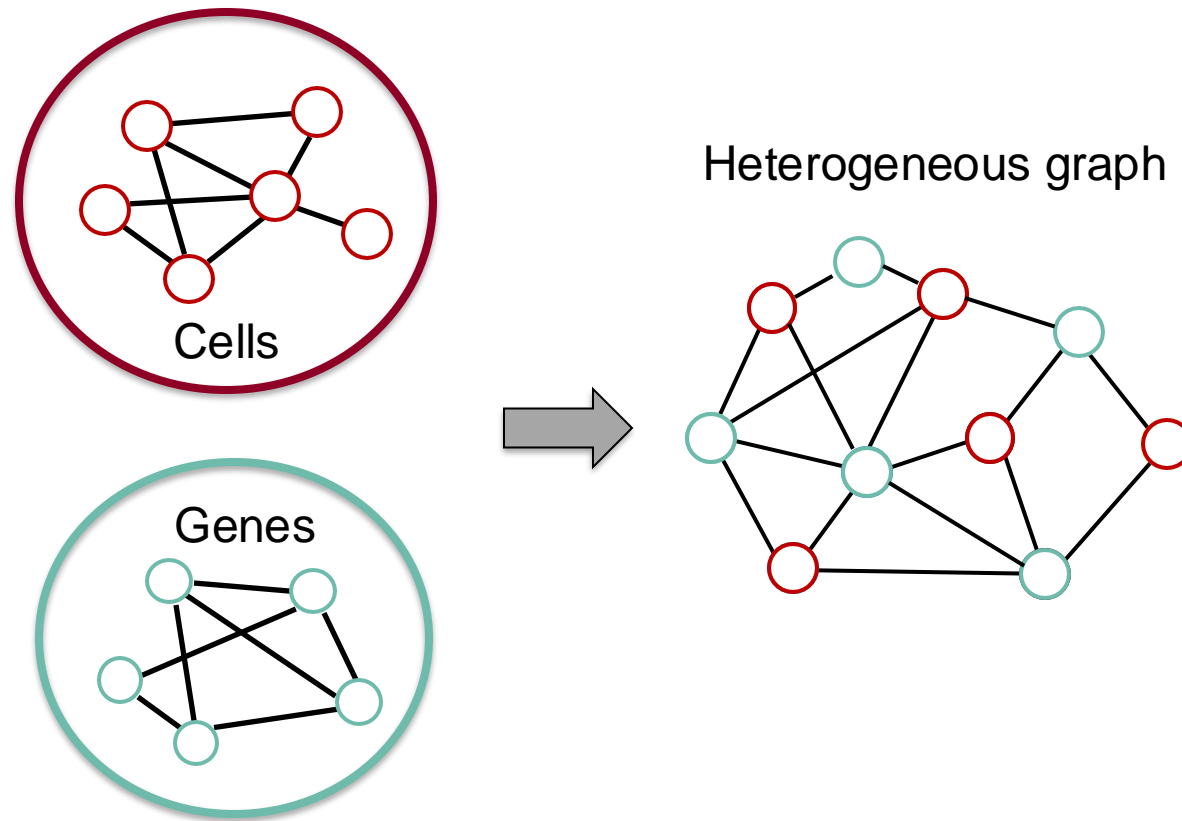
- The first graph neural network for scRNA-seq data;
- Consider neighbor effects in characterizing cell-cell relations;
- Simultaneously perform cell clustering and gene imputation;
- Achieved better performance than existing tools;
- scGNN has been cited ~520 times since 2021.

Juexin Wang*, Anjun Ma*, et al., scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature Communications*. (2021)

Wan, Changlin, et al. "LTMG: a novel statistical modeling of transcriptional expression states in single-cell RNA-Seq data." *Nucleic acids research* 47.18 (2019): e111-e111.

Minimal graph representation of molecular and cell biology:

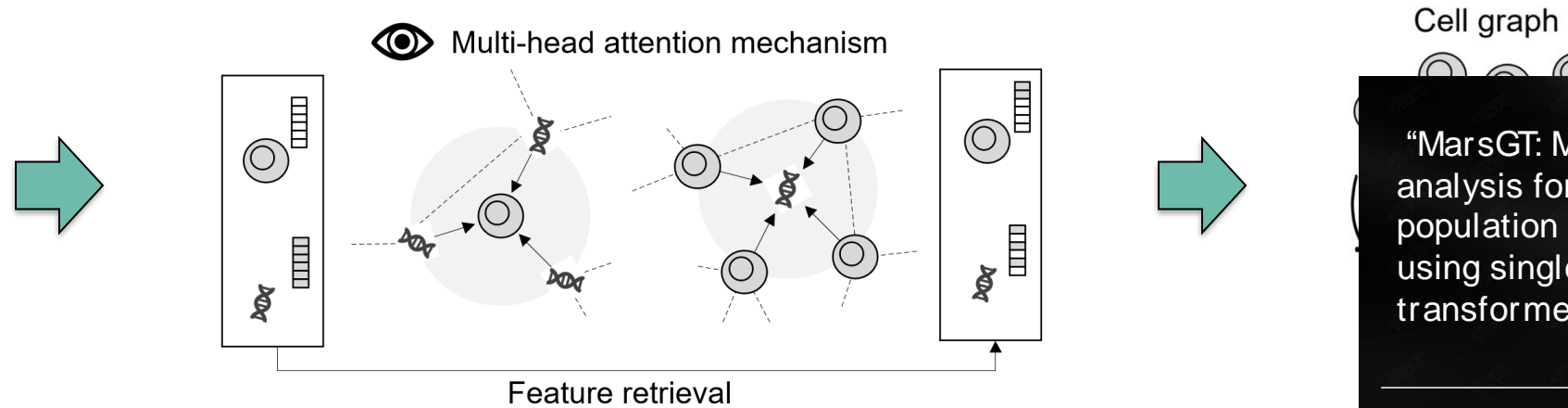
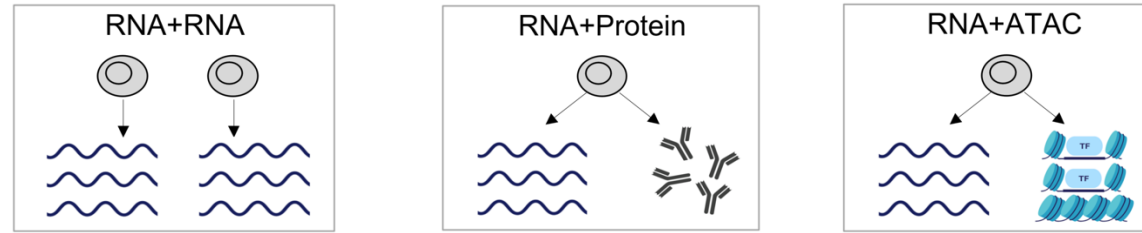
cell-gene graph



DeepMAPS: Deep learning based Multi-omics Analysis platform for Single cells

"Single-cell biological network inference using a heterogeneous graph transformer."

Chosen by *Nature Communications* as **1 of the 50 Best Papers** recently published in *Cancer Research*!



Cell graph

"MarsGT: Multi-omics analysis for rare population inference using single-cell graph transformer"

Chosen by *Nature Communications* as **1 of the 50 Best Papers** recently published in *Biotechnology and Methods*



Dr. Dong Xu
Professor, MU

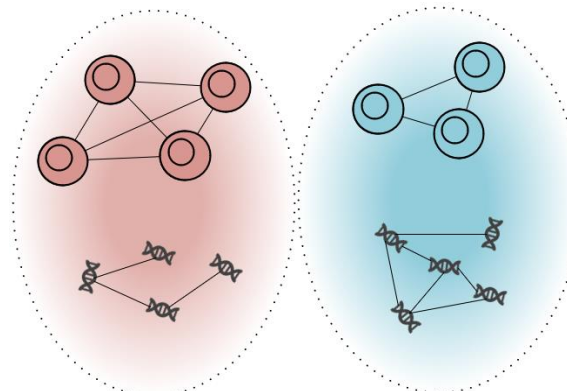


Dr. Anjun Ma
Assis Prof, OSU



Dr. Juexin Wang
Assis Prof, IU

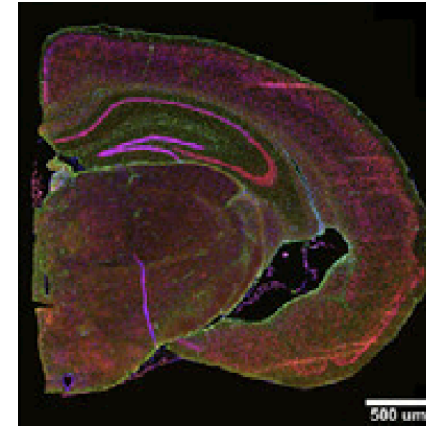
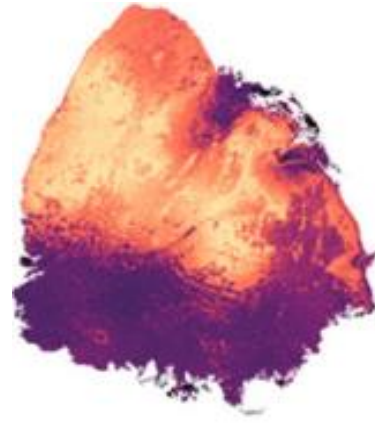
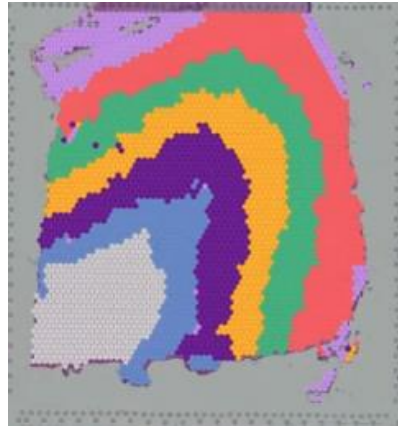
Ma, Anjun, et al. "Single-cell biological network inference using a heterogeneous graph transformer." *Nature Communications* 14.1 (2023): 964.



Cell-type-active gene network

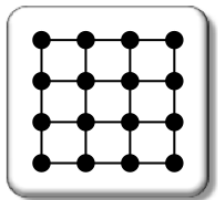


Wang, Xiaoying, et al. "Multi-omics analysis for **rare population** inference using single-cell graph transformer." *Nature Communications* 15, 338 (2024).



Minimal graph representation of tissue biology:

spatial cell/spot graph



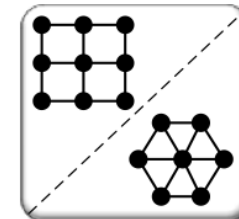
Pixel graph

Subcellular components
Node = subcellular pixel
Edge = pixel-pixel distance



Cell graph

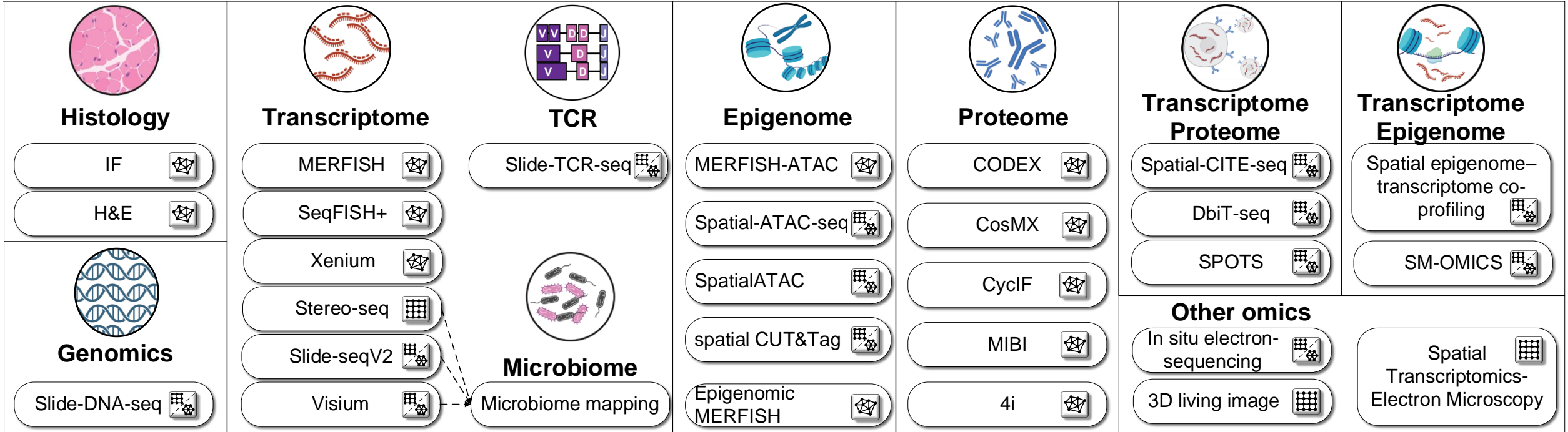
Cellular components
Node = cell
Edge = cell-cell distance



Spot graph

Multicellular components
Node = spot
Edge = spot-spot distance

Spatial omics features provide insights at different resolution

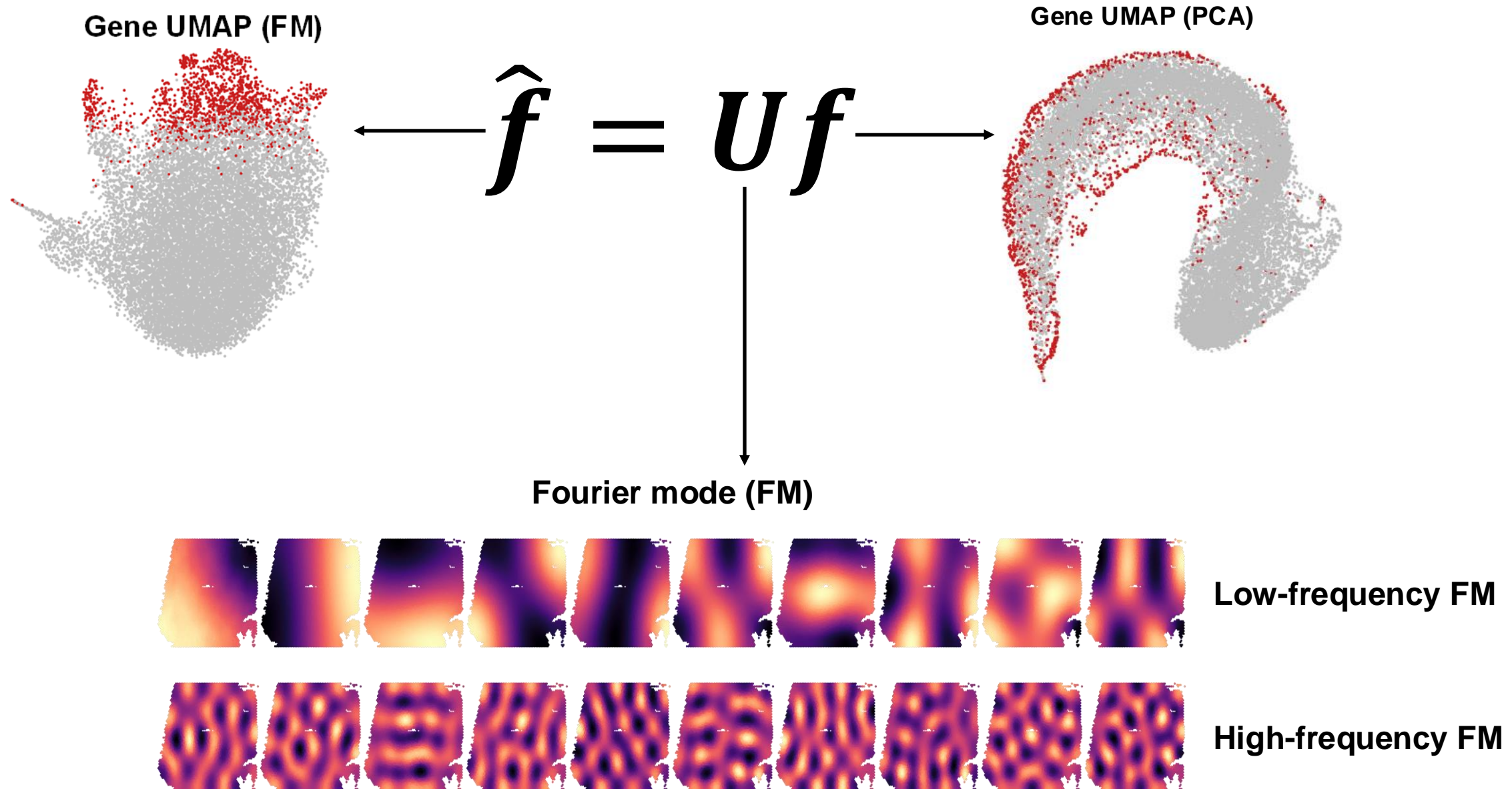


Dr. Garry Nolan
Professor, Stanford

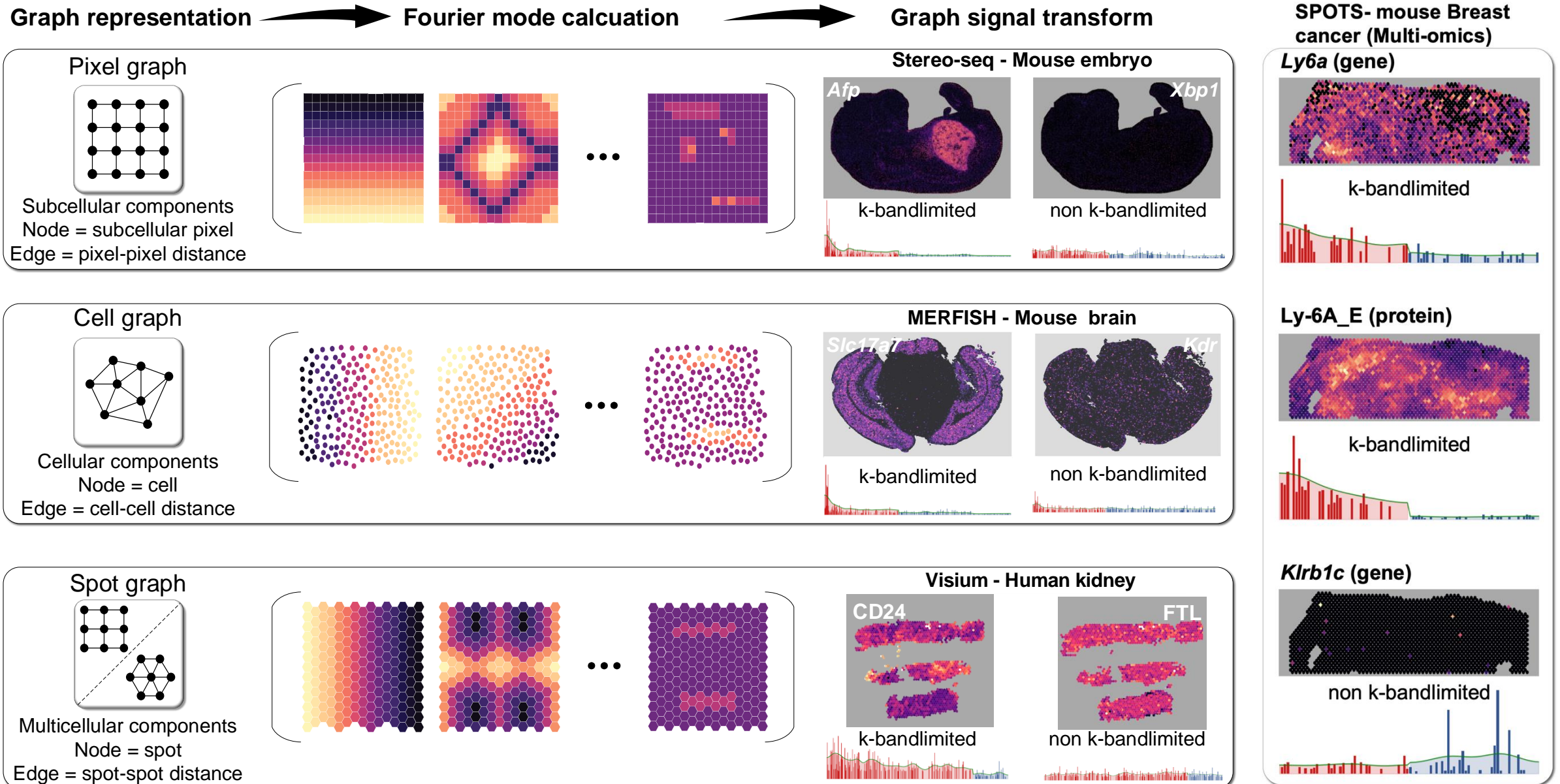


Dr. Sizun Jiang
Assis Prof, Harvard

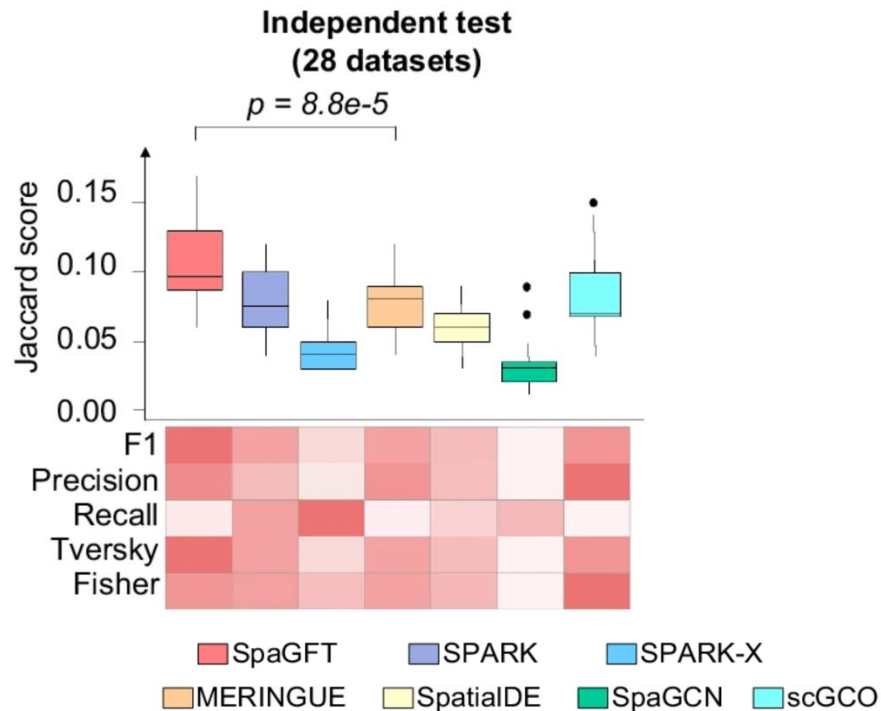
Graph Fourier Transform for spatial transcriptome data



SpaGFT: Spatial omics feature representation using graph Fourier transform



Two applications: spatially variable gene (SVG) and feature signal enhancement / noise removal

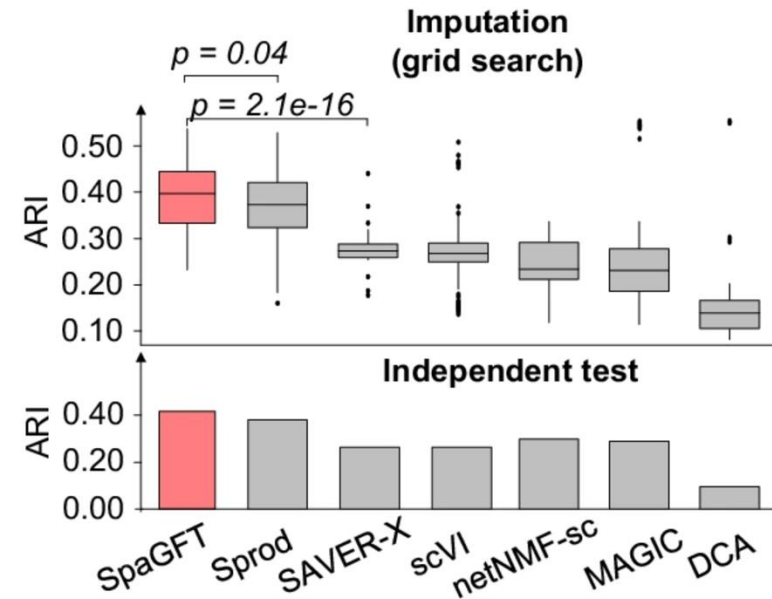


Benchmarking data:

- **28** human/mouse brain datasets
- Across **3** spatial transcriptomics platforms
- **458** curated ground truth SVG

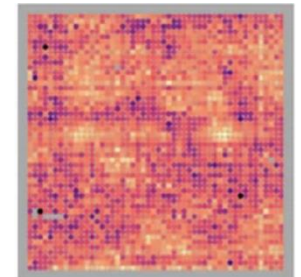
Conclusion:

SpaGFT Outperformed than other **6** tools

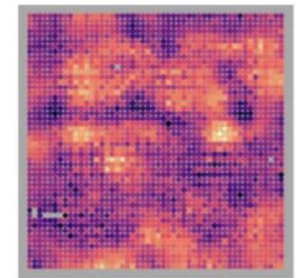


Based on the benchmarking, SpaGFT outperformed the other **6** imputation tools, including one tool specific for spatial denoising tool (i.e., Sprod).

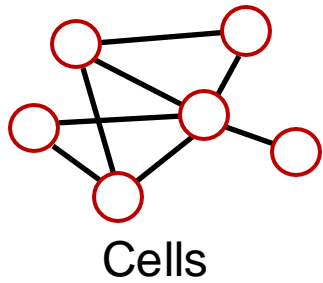
Demonstration:
CD19 protein
intensity on human
tonsil



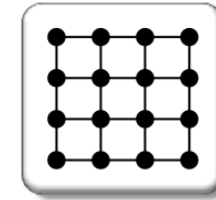
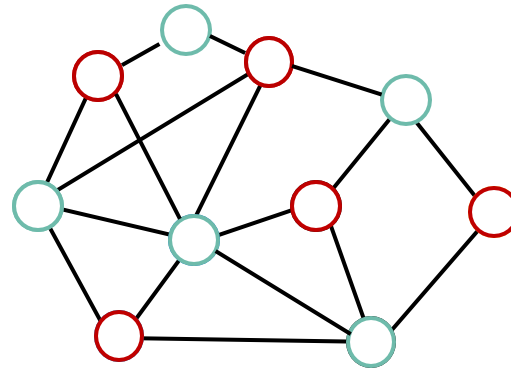
Before



After

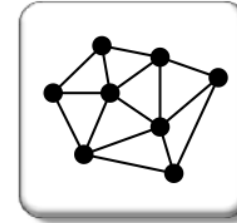


Heterogeneous graph



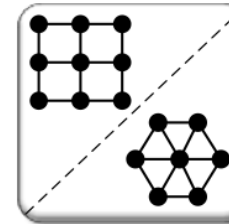
Subcellular components
Node = subcellular pixel
Edge = pixel-pixel distance

Pixel graph



Cellular components
Node = cell
Edge = cell-cell distance

Cell graph



Multicellular components
Node = spot
Edge = spot-spot distance

Spot graph



Acknowledgement

All BMBL lab members!

All PIIO/IOIG members!

Collaborators @ OSU

Dr. Anjun Ma

Dr. Dongjun Chung

Dr. Zihai Li

Dr. Haitao Wen

Collaborators @ Duke

Dr. Anru Zhang

Collaborators @ Stanford

Dr. Garry Nolan

Collaborators @ U of Missouri

Dr. Dong Xu

Collaborators @ Indiana U

Dr. Juexin Wang

Dr. Michael Eadon

Collaborators @ OHSU

Dr. Chi Zhang

Collaborators @ Harvard

Dr. Sizun Jiang

Collaborators @ Heidelberg

Dr. Julio Saez-Rodriguez



PI/MPI (° Complete)

U24CA252977°

R01GM131399°

NSF1945971°

NSF1546869°

R01DK138504

R01GM152585

R21HG012482

Data analysis core PI

U54AG075931

P01CA267108

P01AI177687



<https://u.osu.edu/bmb/>



@QinMaBMBL



Qin.Ma@osumc.edu