

Dr. Qin Ma
Professor, OSU



Dr. Dong Xu
Professor, MU

Ma, Q., et al. Deep learning shapes single-cell data analysis.
Nat Rev Mol Cell Biol (2022)

Ma, Q., et al. Harnessing the deep learning power of foundation models in single-cell omics. *Nat Rev Mol Cell Biol* (2024)

Graph representation learning of single-cell and spatial transcriptomics data

Qin Ma, PhD

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

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

Co-director of the Computational Health and Life Sciences
Translational Data Analytics Institute



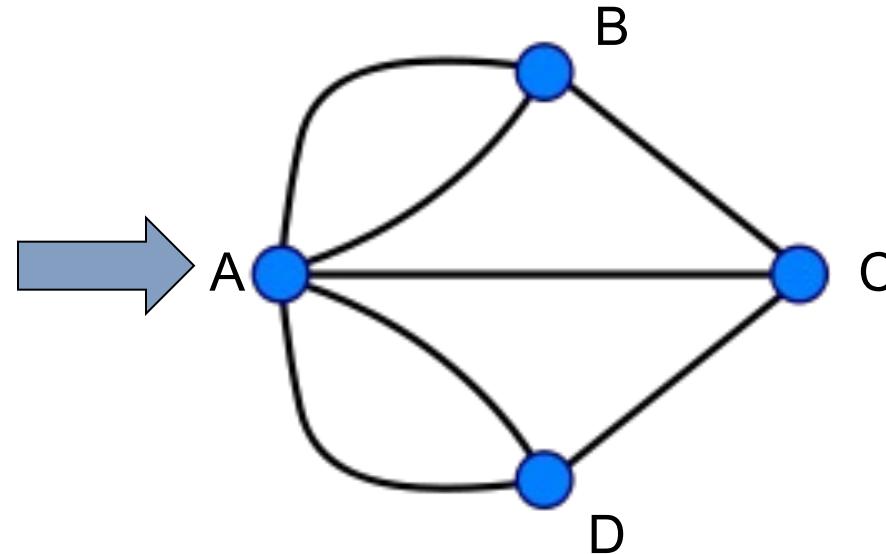
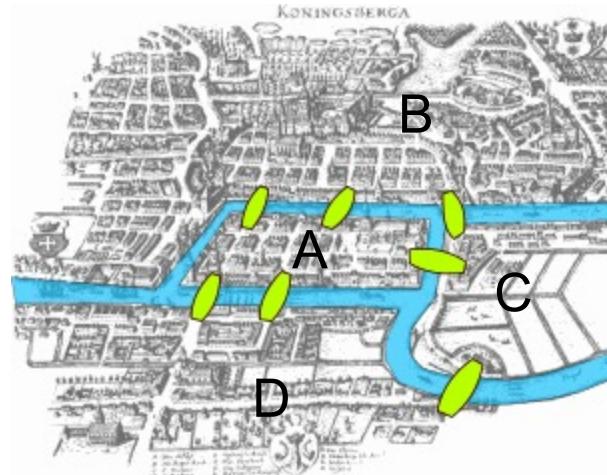
Why Graphs?

Systems biology – an approach that studies biological problems through studying interrelationships of all of the elements in a system rather than studying them one at a time

Leroy Hood, Institute for Systems Biology

300 years ago: Seven Bridges of Konigsberg

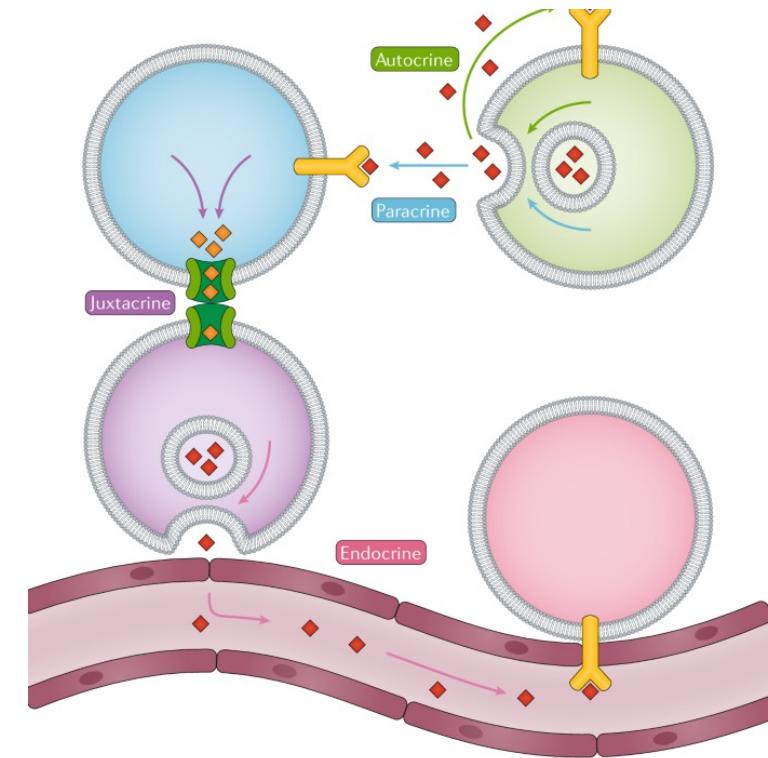
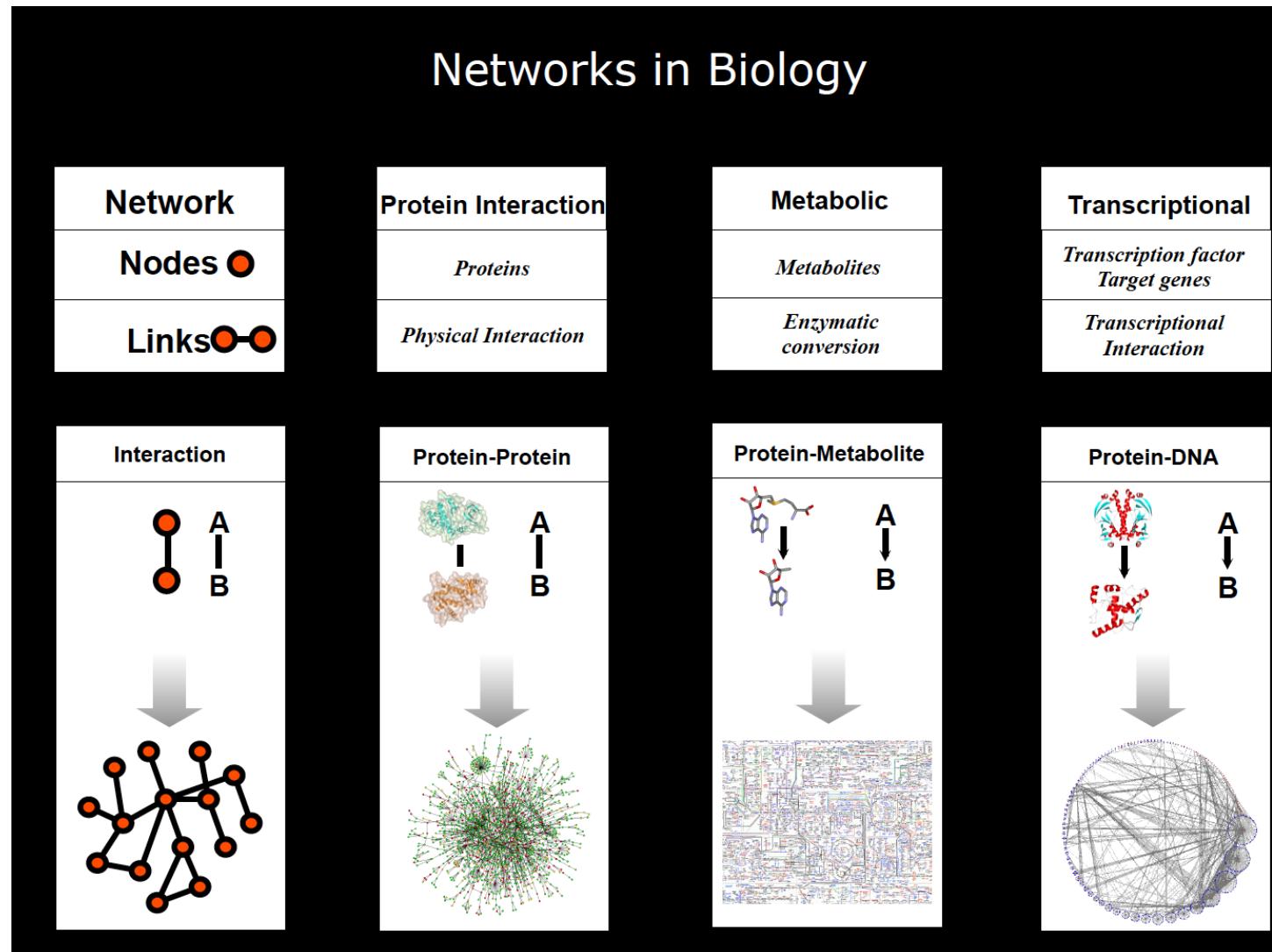
The problem was to find a walk through the city that would cross each bridge once and only once.



Its negative solution by Leonhard Euler in 1735 laid the foundations of graph theory.

Model: Euler tour \Leftrightarrow no nodes of odd degree

Recent 30 years: The Biological Networks



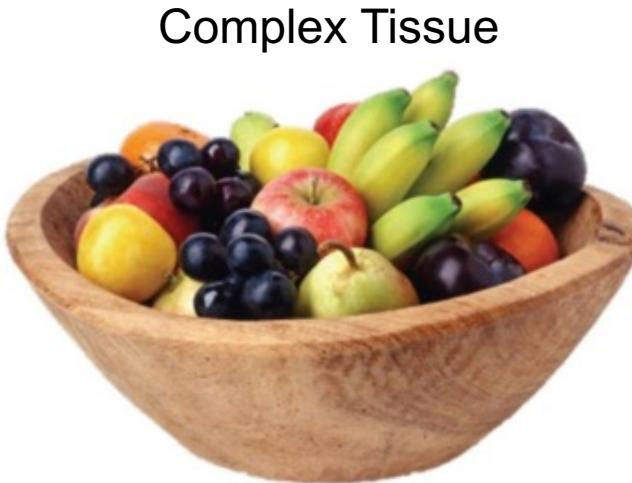
Cell-cell communications

Armingol, E., Officer, A., Harismendy, O. et al. *Nat Rev Genet* **22**, 71–88 (2021).

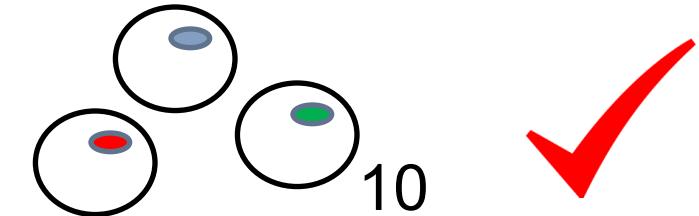
Systems biology questions

- What are the differentially expressed genes or gene modules between health and cancer patients?
- How do genetic variations or mutations in specific genes affect the behavior and function of biological systems?
- How do different signaling pathways interact and regulate each other to coordinate cellular responses to external stimuli?
- How to construct gene regulatory networks for different cell types or cell states?
- What molecular program drives the dynamic programs during cell developments and response to drugs?

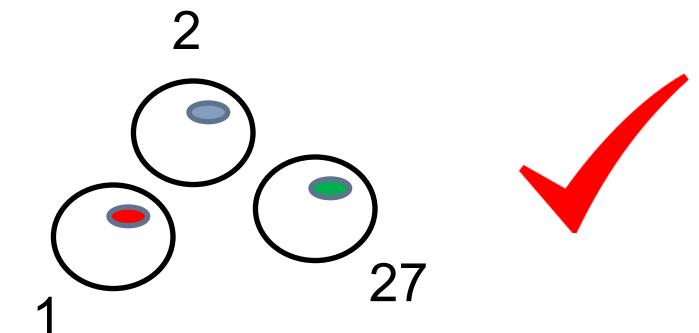
Understanding complexity using sequencing



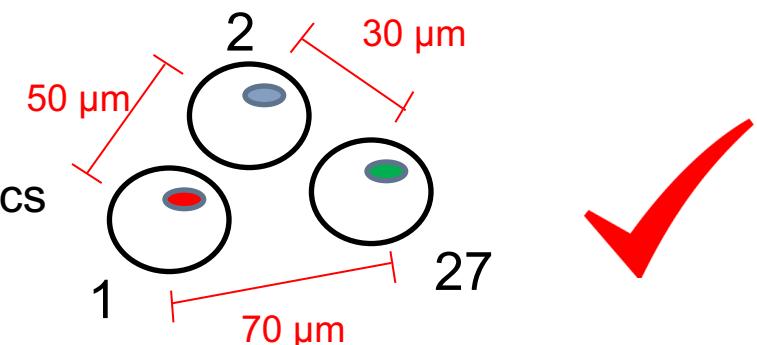
Bulk Genomics



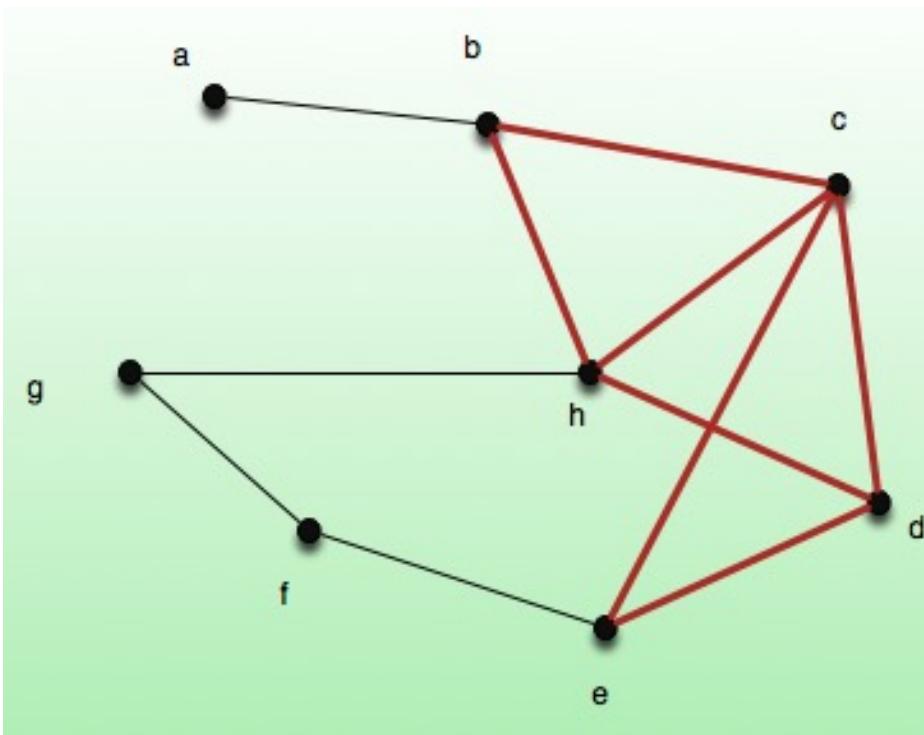
Single Cell Genomics



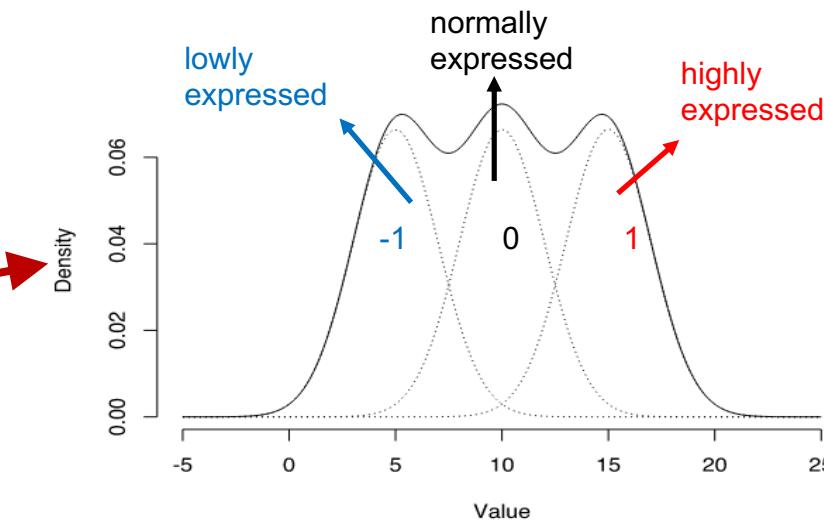
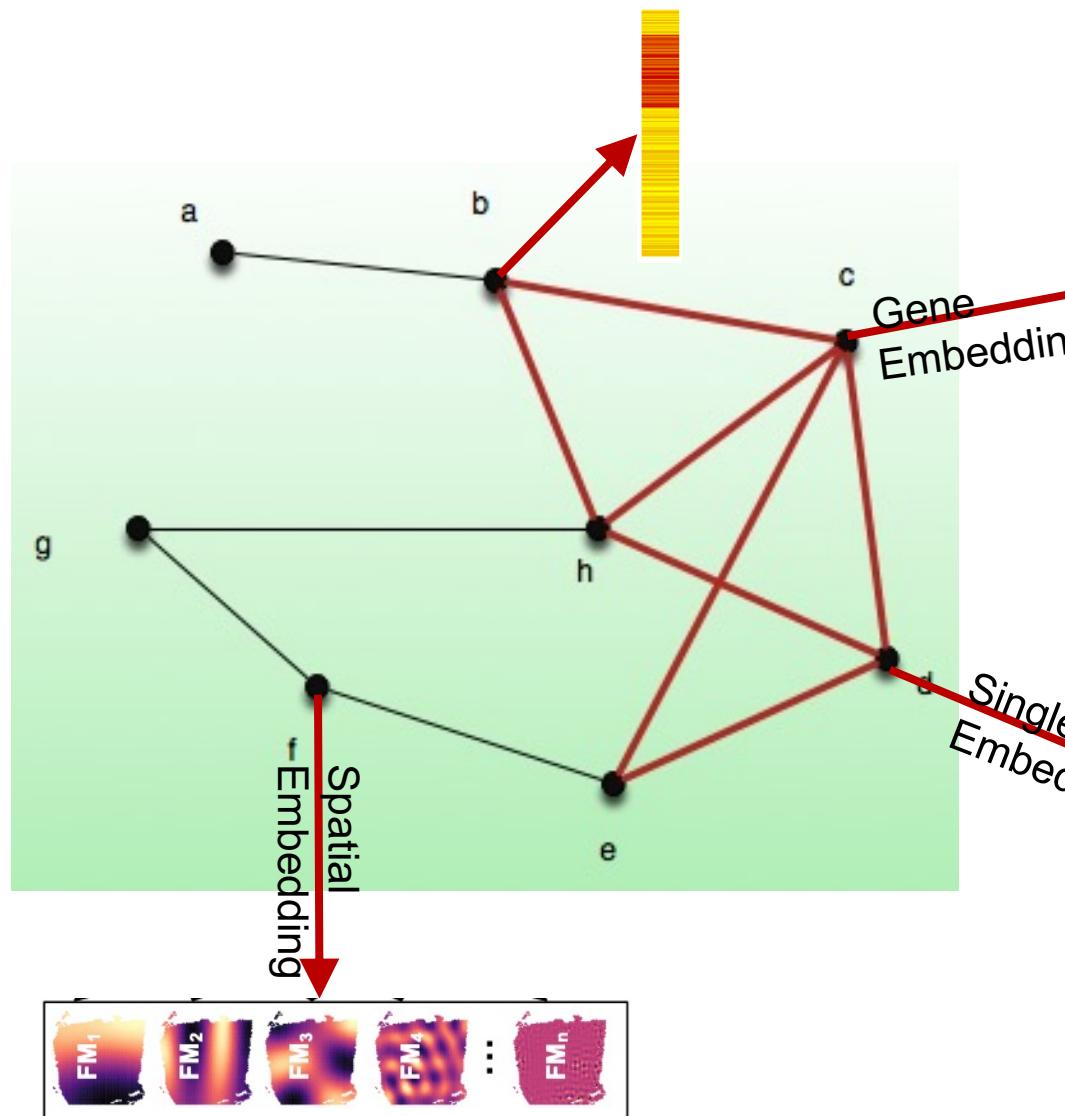
Spatial Transcriptomics



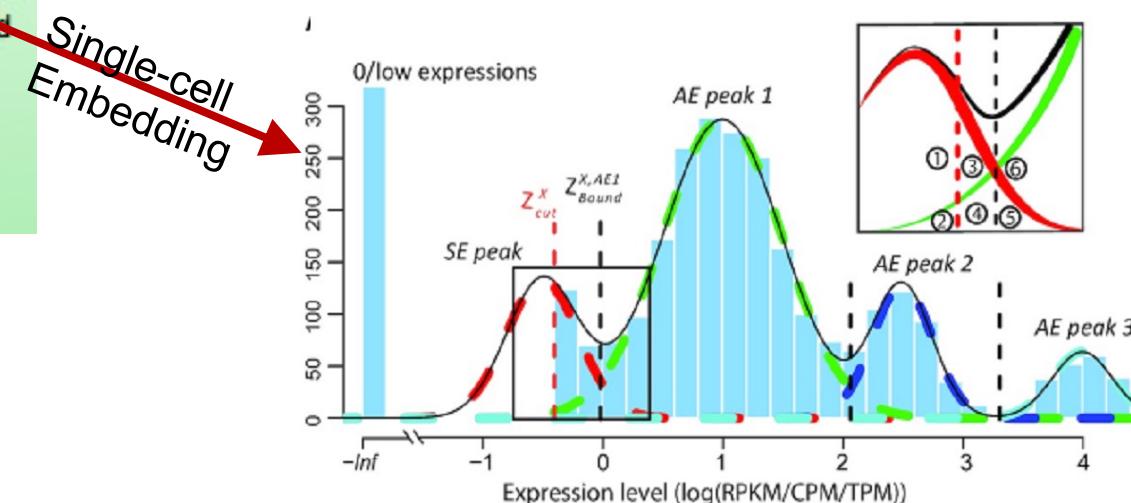
The minimal graph representation system of molecular mechanism: gene module/network



Graph representation learning: gene module detection



Li, Guojun, et al. "QUBIC: a qualitative biclustering algorithm for analyses of gene expression data." *Nucleic acids research* 37.15 (2009): e101-e101.

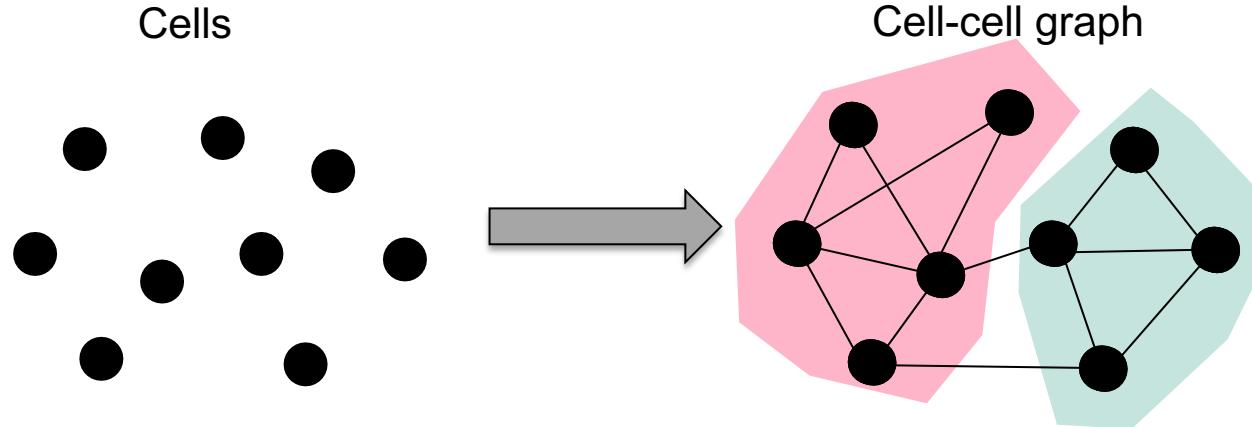


Chang, Yuzhou, et al. "SpaGFT is a graph Fourier transform for tissue module identification from spatially resolved transcriptomics." *bioRxiv* (2022): 2022-12.

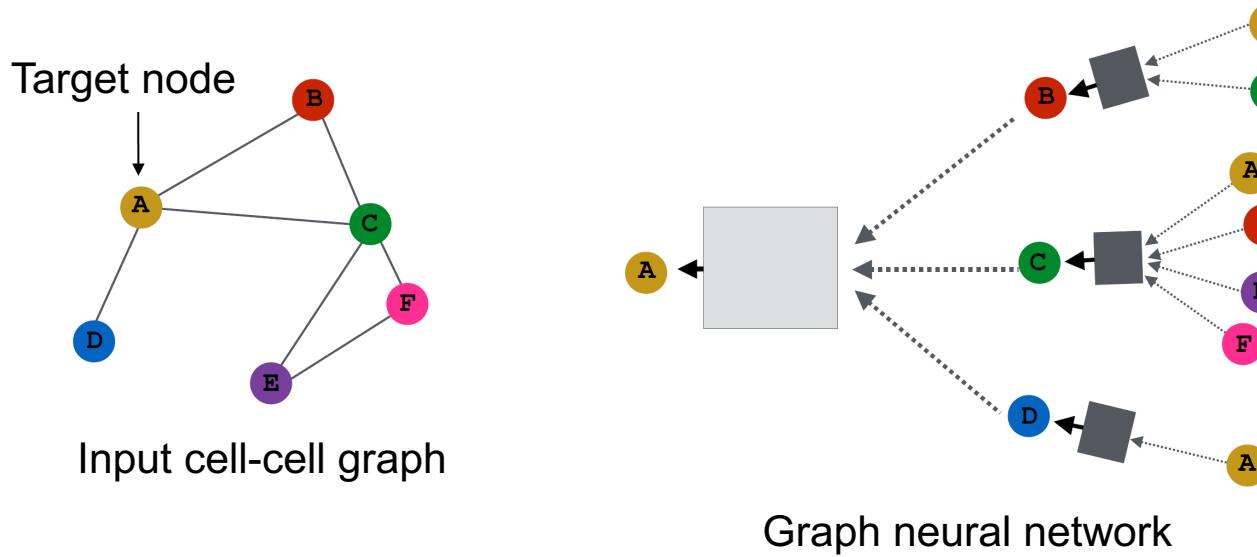
Wan, Changlin, et al. "LTMG: a novel statistical modeling of transcriptional expression states in single-cell RNA-Seq data." *Nucleic acids research* 47.18 (2019): e111-e111.

The minimal graph representation system of cellular heterogeneity: cell network

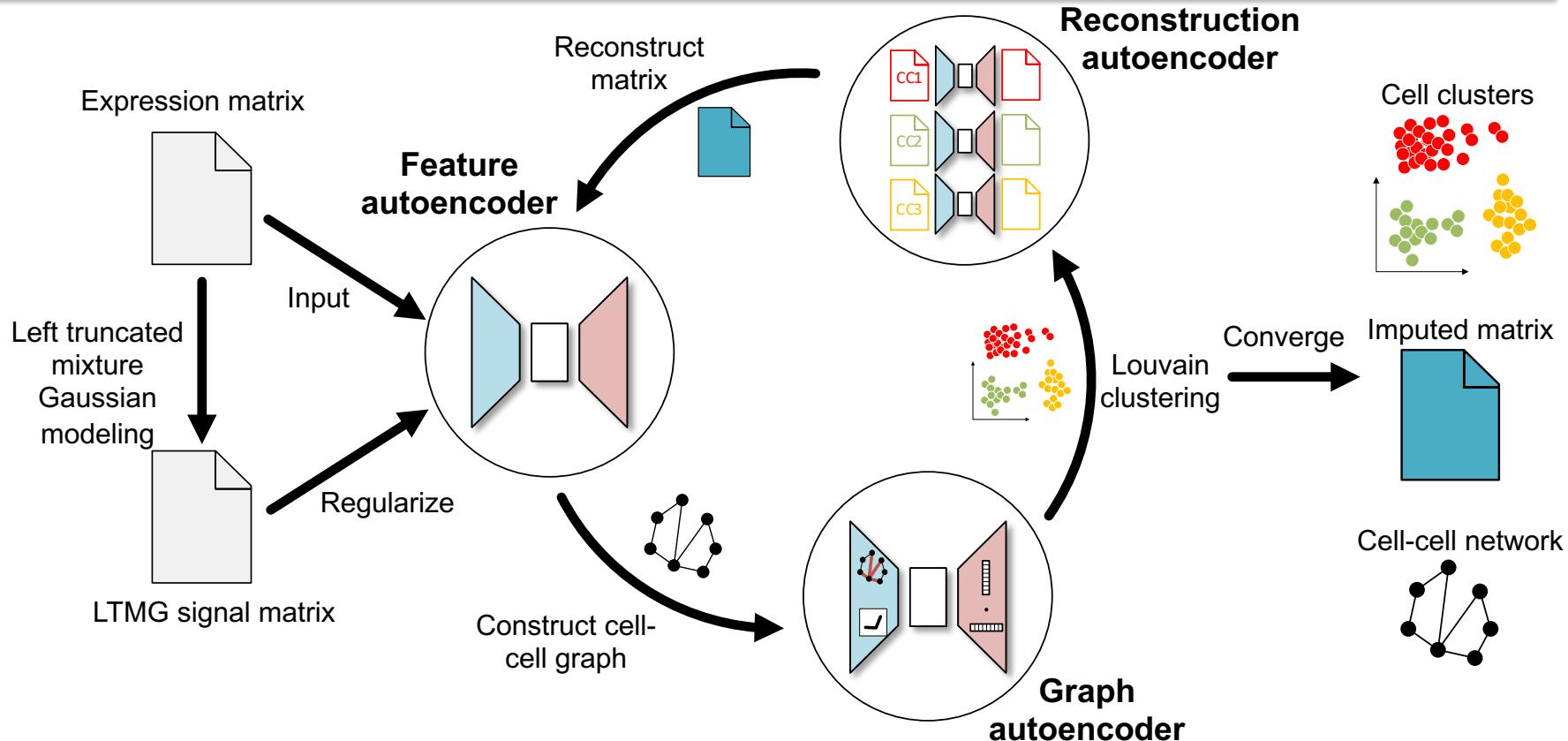
Graph Neural Network for gene imputation and cell clustering



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. Dong Xu
Professor, MU

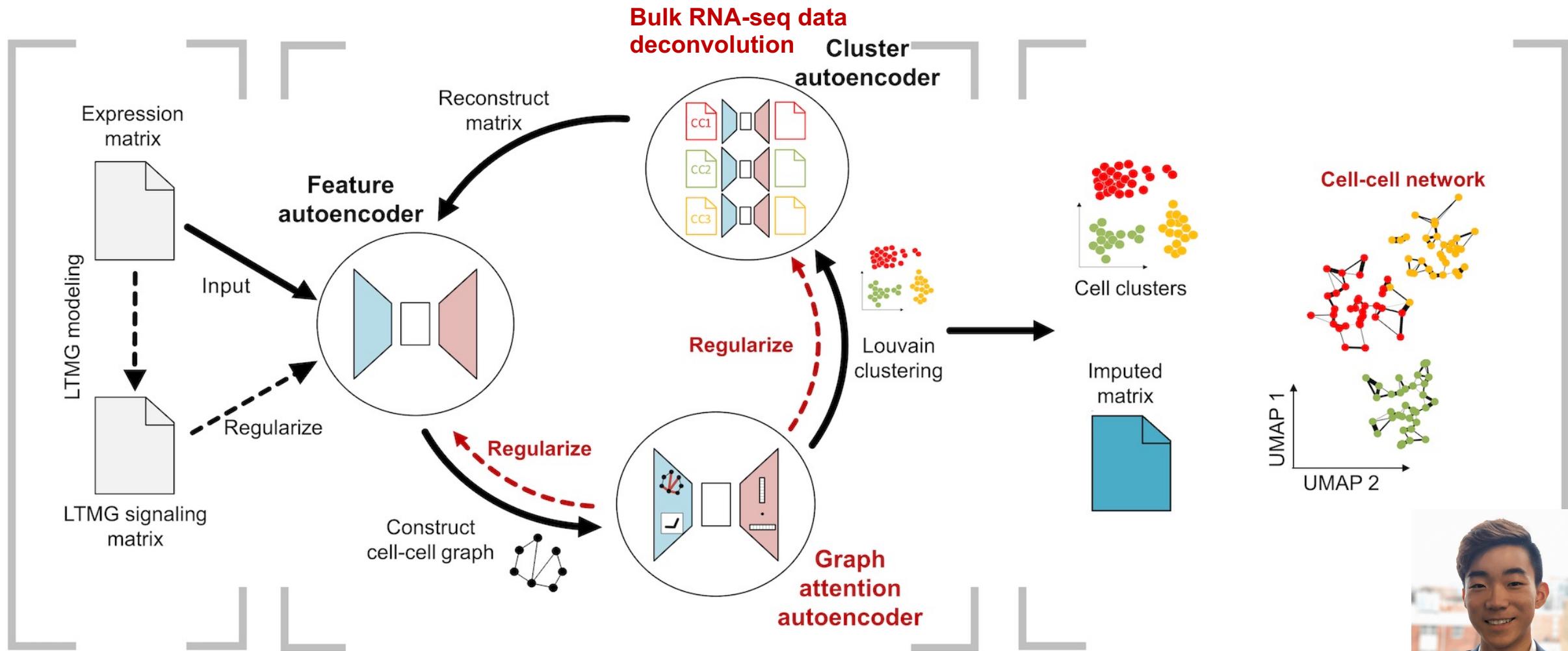


Dr. Juexin Wang
Assis Prof, IU

- 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 ~400 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)

scGNN2.0: a graph neural network tool for imputation and clustering of single-cell RNA-Seq data



Edison Haocheng Gu, et al., scGNN 2.0: a graph neural network tool for imputation and clustering of single-cell RNA-Seq data, *Bioinformatics*, Volume 38, Issue 23, 1 December 2022



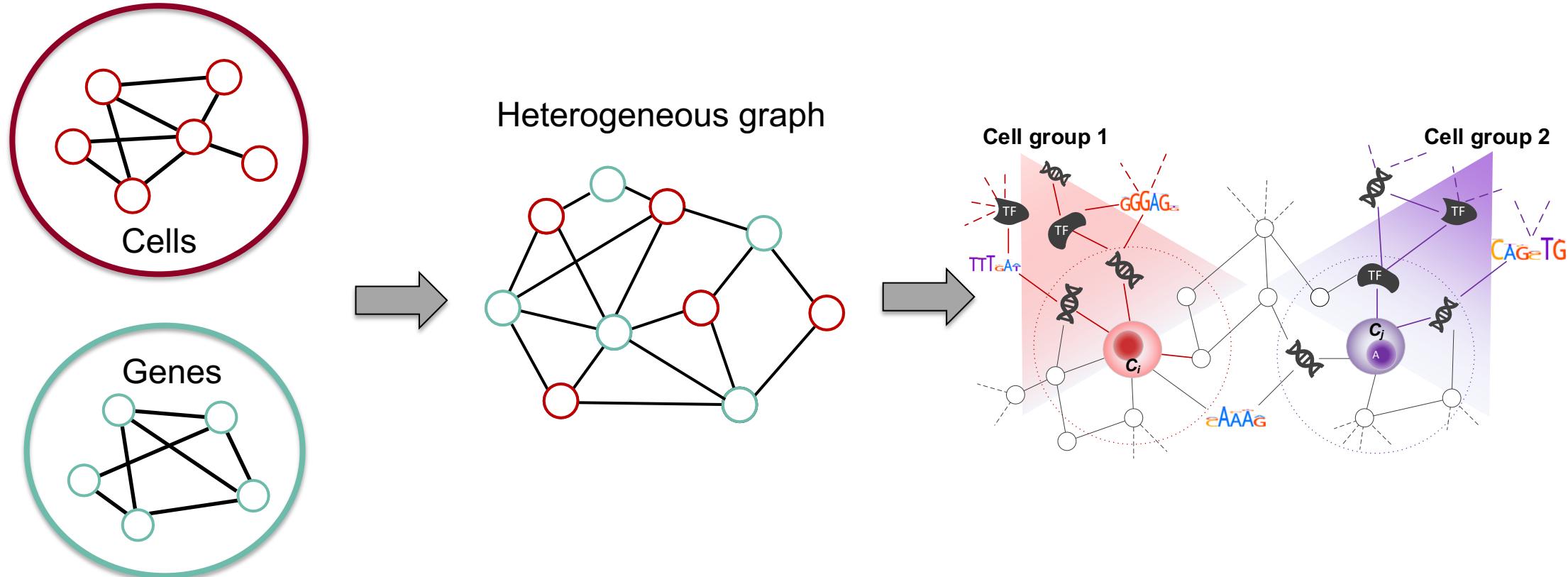
Edison Gu

Challenge in analyzing scMulti-omics data

Current methods have limited power to understand the cross-talk between cells and different modalities.

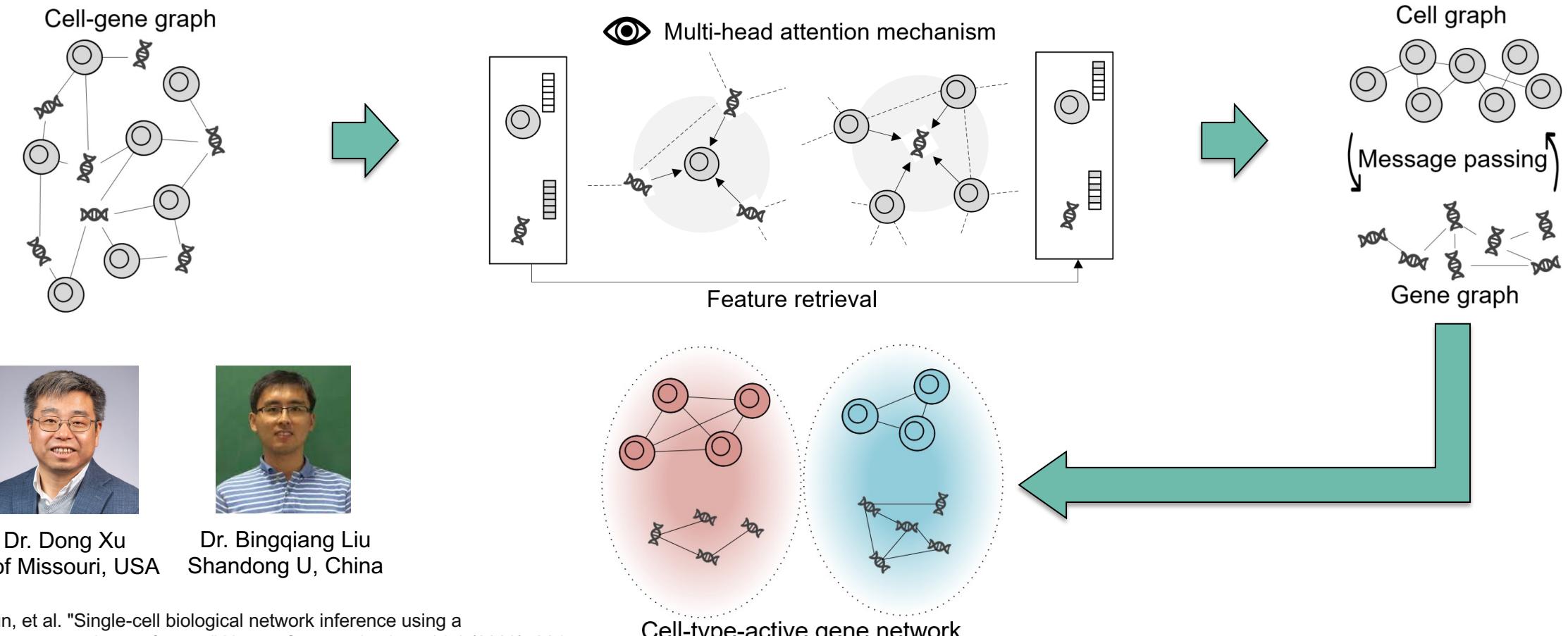
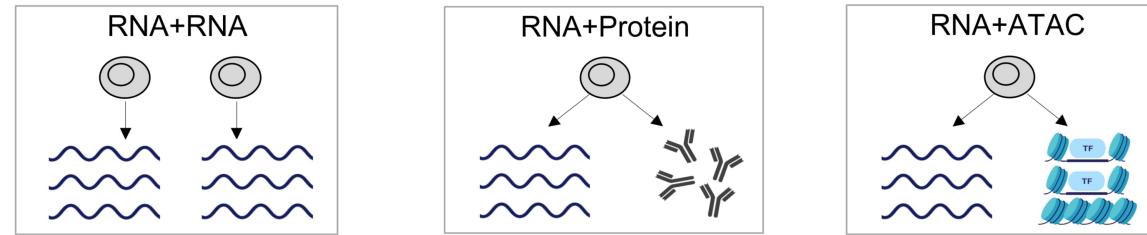
- inferring the underlying biological networks of diverse cell types
- elucidating the response of these networks to external stimuli in a specific cell type
- discovering the molecular programs that drive transitions from one cell type to another

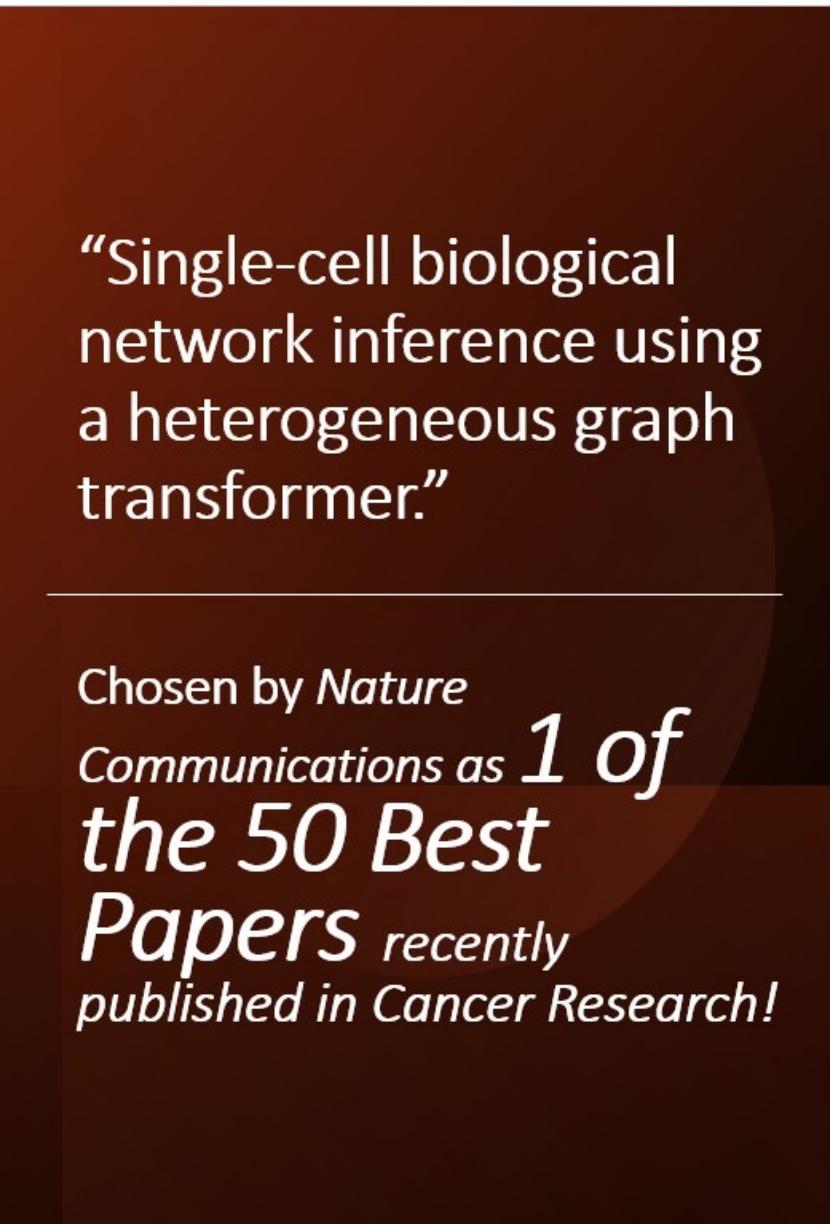
Graph neural network + heterogeneous graph



The minimal graph representation system of molecular program for cellular heterogeneity: cell-gene network

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





nature reviews genetics

<https://doi.org/10.1038/s41576-023-00618-5>

Review article

Check for updates

Gene regulatory network inference in the era of single-cell multi-omics

Pau Badia-i-Mompel ¹, Lorna Wessels ^{1,2}, Sophia Müller-Dott ¹, Rémi Trimbour ^{1,3}, Ricardo O. Ramirez Flores ¹, Ricard Argelaguet ⁴ & Julio Saez-Rodriguez ¹

Table 1 | Existing tools for inference of gene regulatory networks from multi-omics data

Tool ^a	Possible inputs	Type of multimodal data	Type of modelling	Type of interactions	Statistical framework	Default motif database/motif matcher	Default upstream/downstream distance cutoffs	Language	Refs.
ANANSE	Groups, contrasts	Unpaired	Linear	Weighted	Frequentist	CIS-BP/GimmeMotifs	100 kb/100 kb	Python	84
CellOracle	Groups, trajectories	Unpaired	Linear	Signed, weighted	Frequentist or Bayesian	CIS-BP/GimmeMotifs	500 kb/500 kb	Python	76
DC3	Groups	Unpaired	Linear	Binary	Frequentist	Undefined/HOMER	Based on Hi-C	Python	88
DeepMAPS	Groups	Paired or integrated	Linear	Weighted	Frequentist	JASPAR/PWMScan	150 kb/150 kb or exon	Python	62
Dictys	Groups, trajectories	Unpaired/paired or integrated	Linear	Signed, weighted	Frequentist	HOCOMOCO/HOMER	500 kb/500 kb		81
DIRECT-NET	Groups	Paired or integrated	Non-linear	Binary	Frequentist	JASPAR/MOODs	250 kb/250 kb		72
FigR	Groups	Paired or integrated	Linear	Signed, weighted	Frequentist	CIS-BP/MOODs	50 kb/50 kb		63
GLUE	Groups	Paired or integrated	Non-linear	Weighted	Frequentist	JASPAR/cisTarget	150 kb/150 kb		64

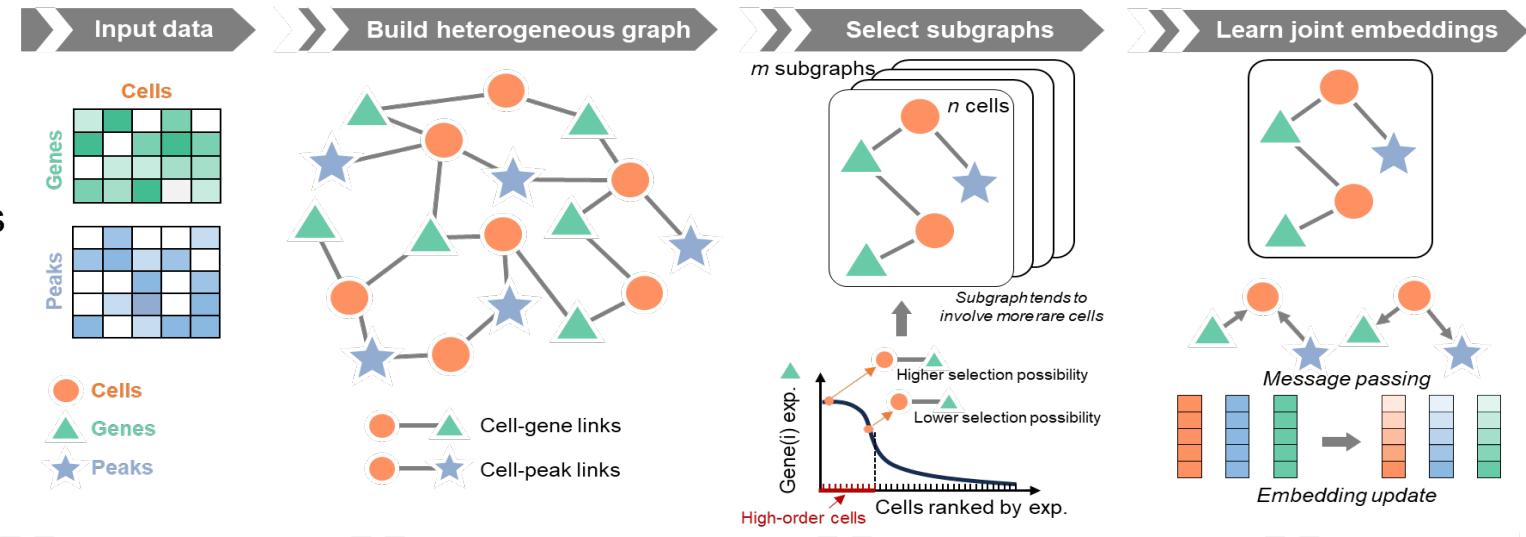


Dr. Anjun Ma
Assistant Prof. OSU

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

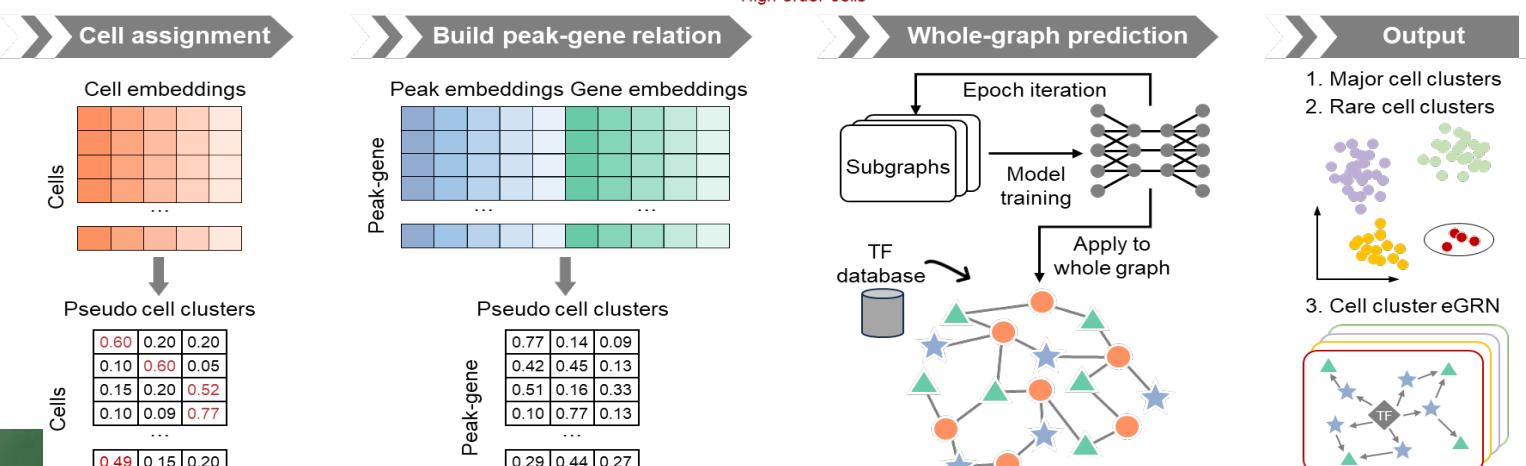
Challenge:

- High false positives when inferring rare populations
- Limited performance with complex samples
- Inability to concurrently identify major and rare cell types
- Compromised accuracy with ultra-rare cell types



Solution:

- Multipartite heterogeneous graph transformer
- Probability-based subgraph-sampling
- Regulatory information (eGRN)



Application:

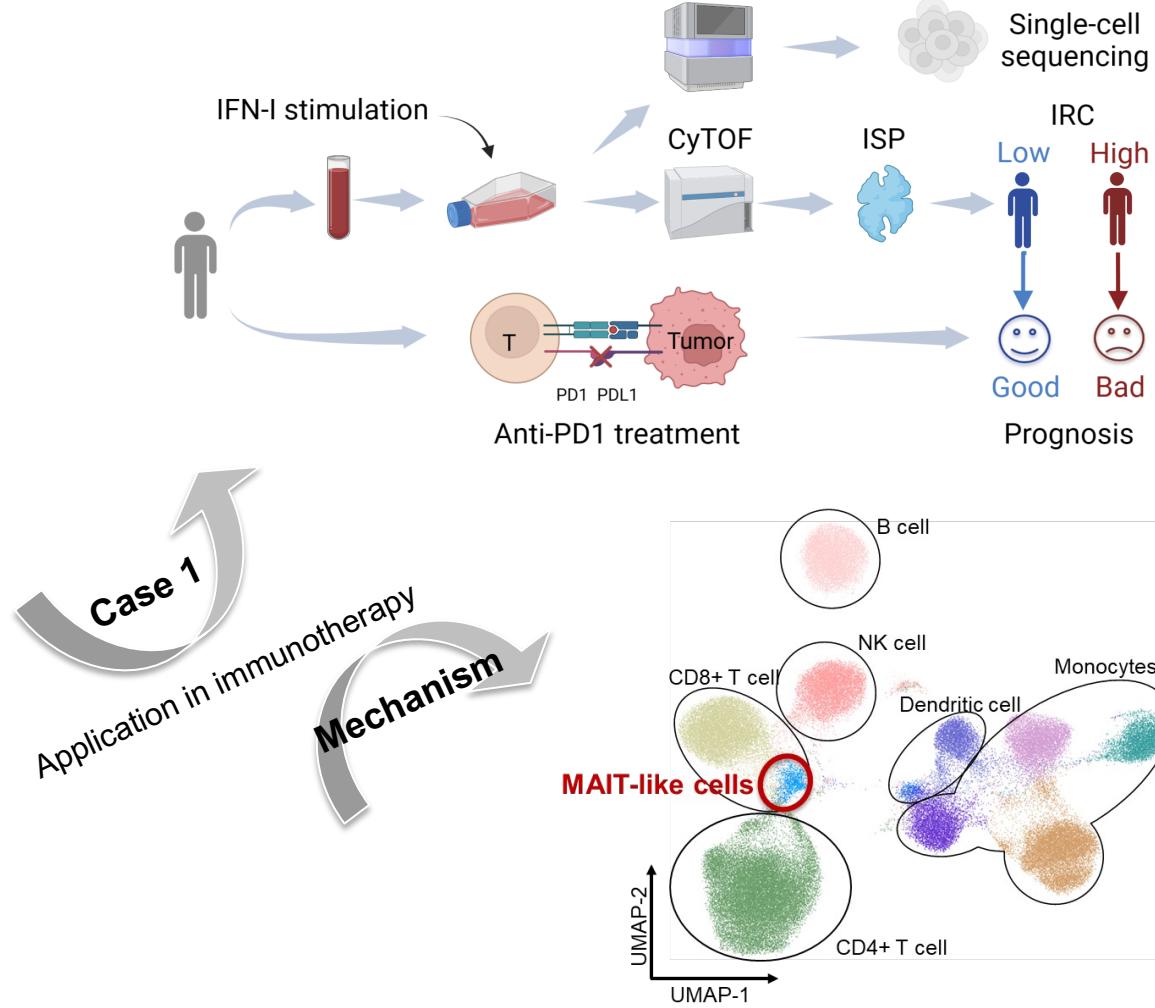
- Identify misidentified rare cell types
- Detect intermediate cell state
- Reveal the mechanism of immunotherapy



Dr. Bingqiang Liu
Shandong U, China

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

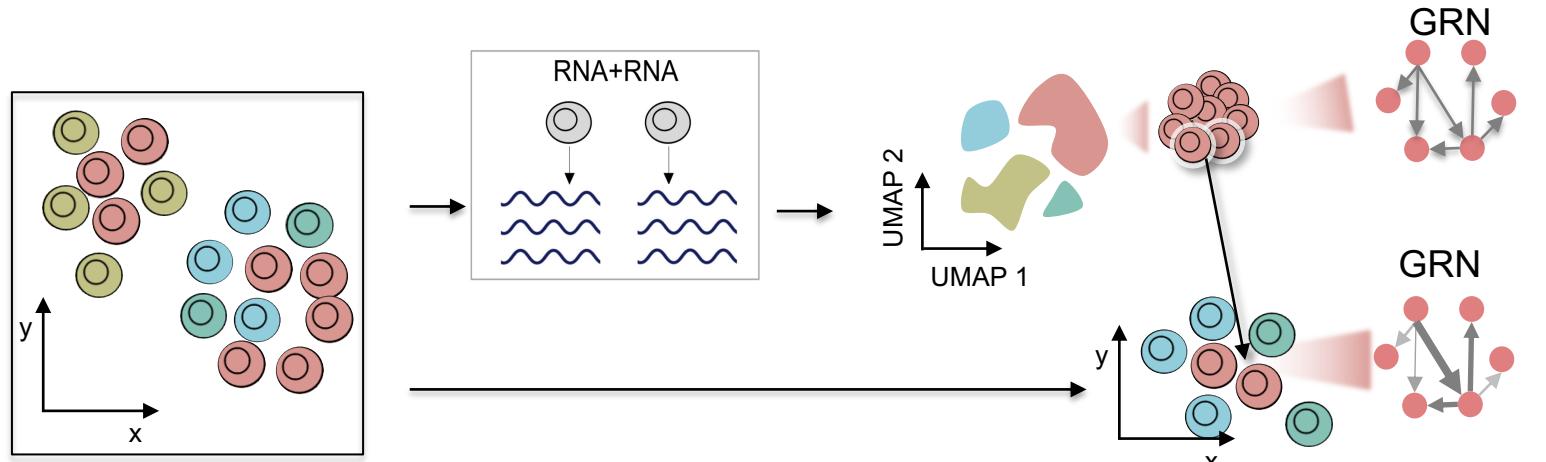
MarsGT is 1 of the 50 best papers in Biotechnology and Methods



Dr. Xiaoying Wang
Postdoc, OSU

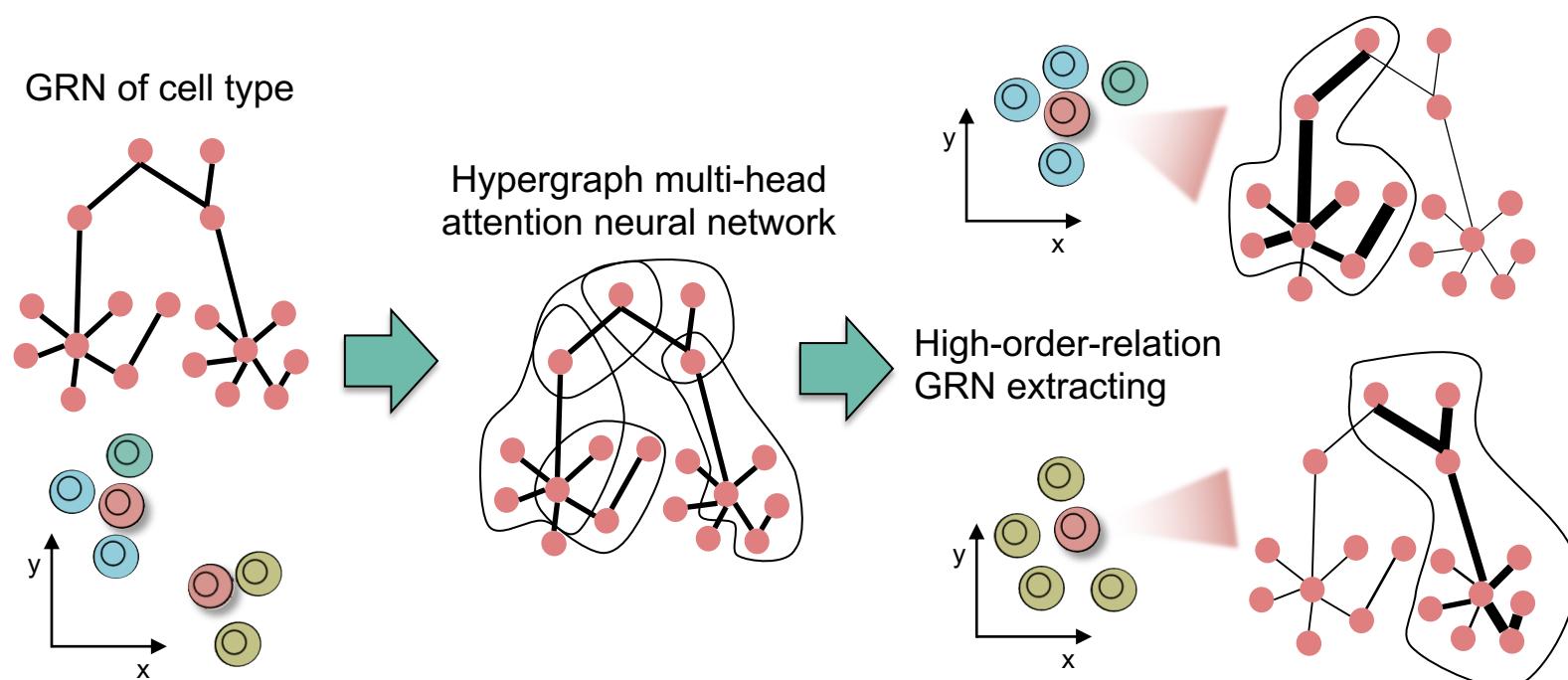
Wang, Xiaoying, et al. "Multi-omics analysis for rare population inference using single-cell graph transformer." *Nature Communications* 15, 338 (2024).
[Application guidance blog link: https://go.nature.com/3v5UiZD](https://go.nature.com/3v5UiZD)

HyperGRN: Elucidate high-order-relation Gene Regulatory Networks in the Spatial Context

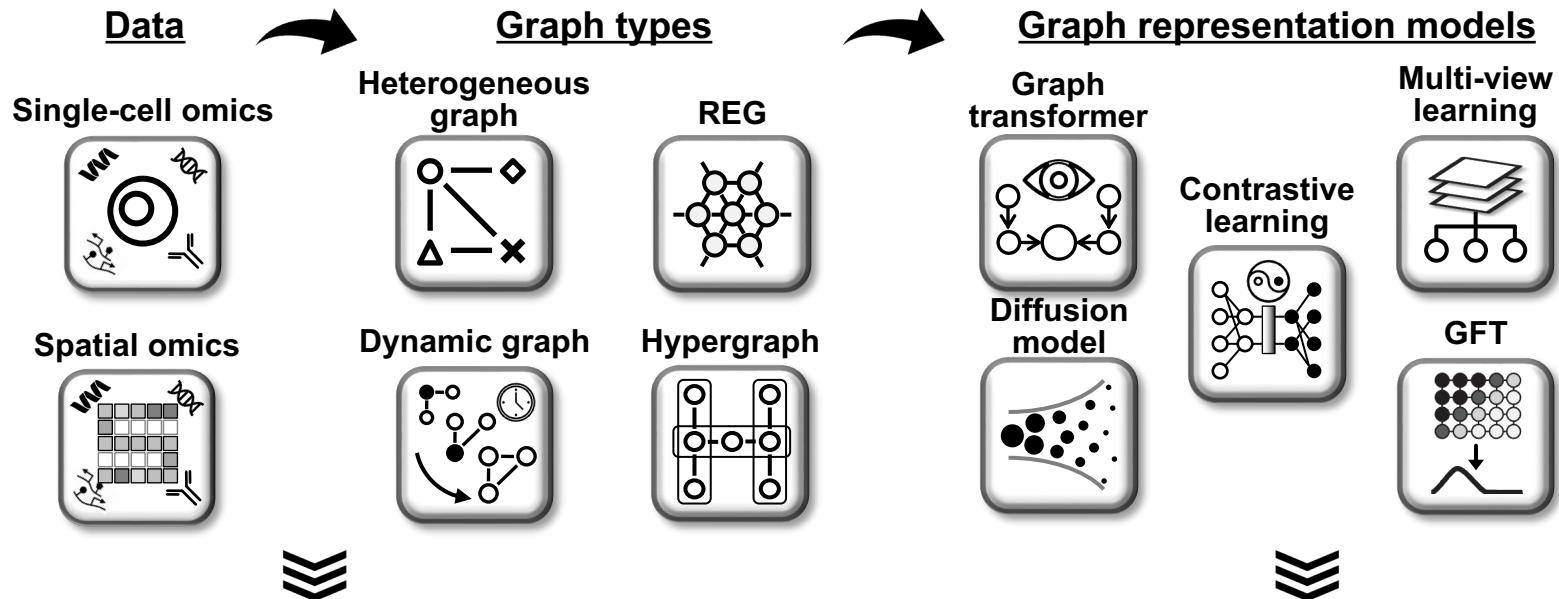


Even from the same cell type, cells may have diverse high-order-relation gene regulatory networks due to heterogeneous cellular neighborhoods in spatial context

- We aim to use a hypergraph multi-head attention neural network to extract high-order-relation GRNs
- Explore the effects of microenvironments on the heterogeneity of cells



Yi Jiang, GRA
BMBL, Ohio State University
10-minute talk with the title
HyperGRN: Elucidate high-order-relation Gene Regulatory Networks in the Spatial Context
at the technical session, Saturday 23rd, 11:15 a.m. to 12:30 p.m.



Algorithms

Data + Graph types + Graph representation models = Pattern prediction

**It is a way of thinking:
how to solve significant biology problems using a computational way**

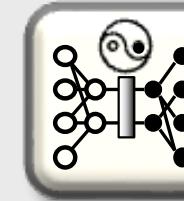
Single-cell omics



Heterogeneous graph



Graph transformer Contrastive learning



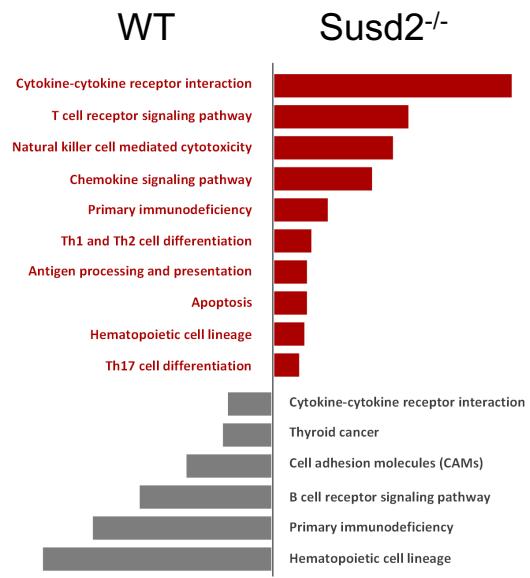
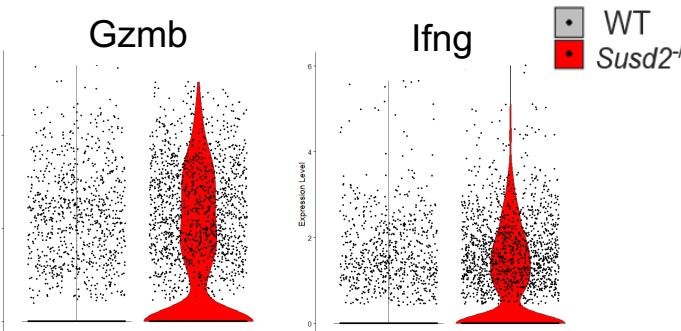
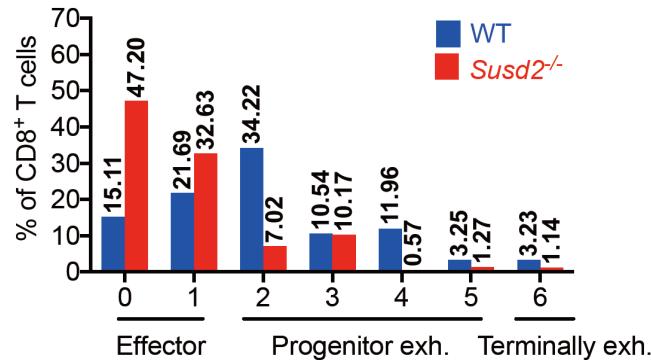
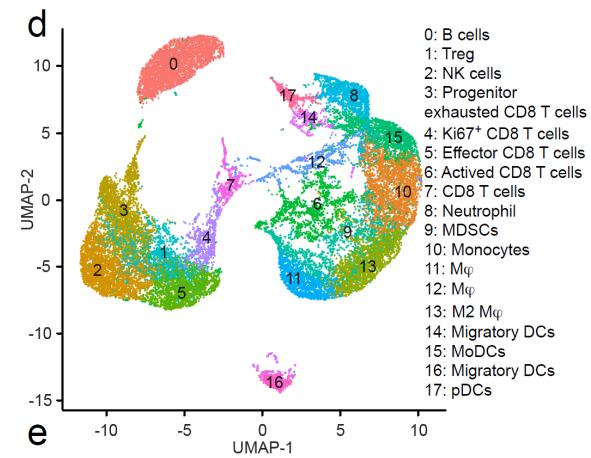
Single-cell regulation



Dr. Haitao Wen
Assoc. Prof, OSU

Bao Zhao, et al. Sushi domain containing 2 suppresses CD8+ T cell antitumor immunity by targeting IL-2 receptor signaling. *Nature Immunology*. 2022

18 cell clusters
(16k Susd2^{-/-} cells, 10k WT cells)

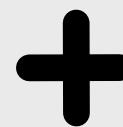
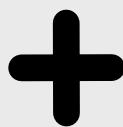


R01AI162779 Targeting immune inhibitory molecule SUSD2 to reverse immunosuppression.

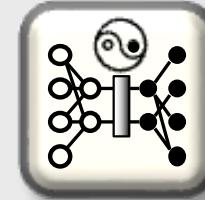
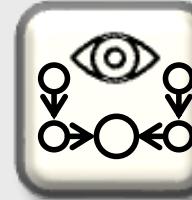
Single-cell omics



Heterogeneous graph



Graph transformer Contrastive learning



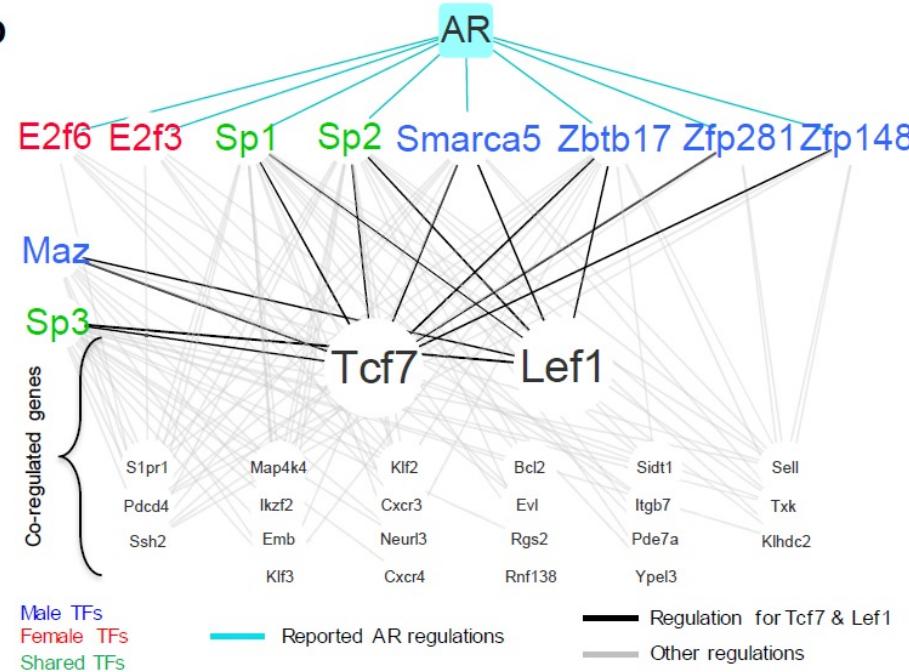
Single-cell regulation

Dr. Zihai Li
Professor and Founding
Director of PIIO



Hynmwoo Kwon, et al. Androgen conspires with the CD8+ T cell exhaustion program and contributes to sex bias in cancer. *Science Immunology*. 2022

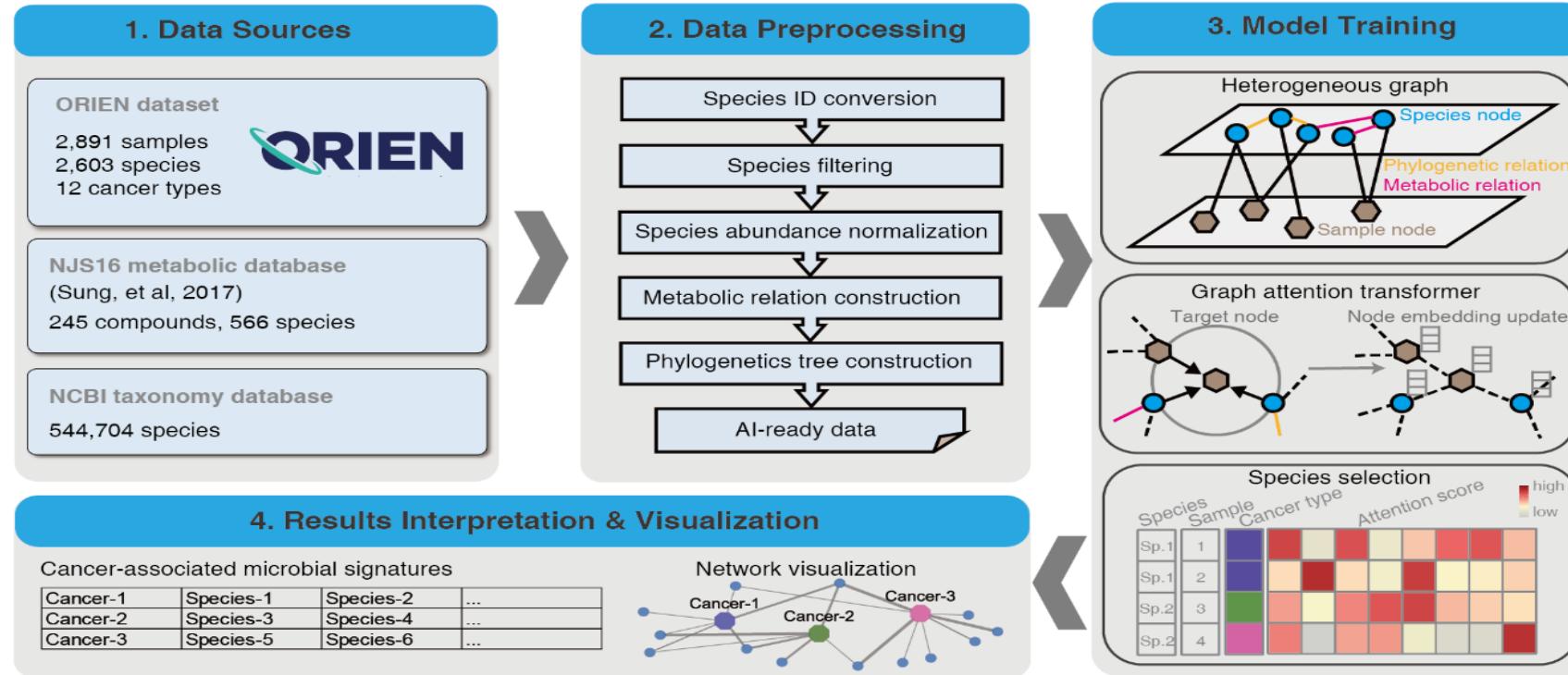
b



- Unique TFs and regulatory mechanisms have been identified in male and female.
- These TFs showed a central regulation of Tcf7 which is highly related to CD8+ T cell exhaustion.
- All male-biased or shared transcription factors except Sp3 and *Maz* have ChIP-confirmed androgen elements
- IRIS3 opens a gate to identify cell-type-specific and conditional-specific regulons.

R01CA262069 Targeting GRP94-TGF-beta pathway for cancer immunotherapy

MEGA: A Bioinformatics Tool for Identifying Intratumoral Microbes from the ORIEN Dataset



Wang, CK, et al. *Cancer Research Communications*. 2024



Qin Ma, PhD



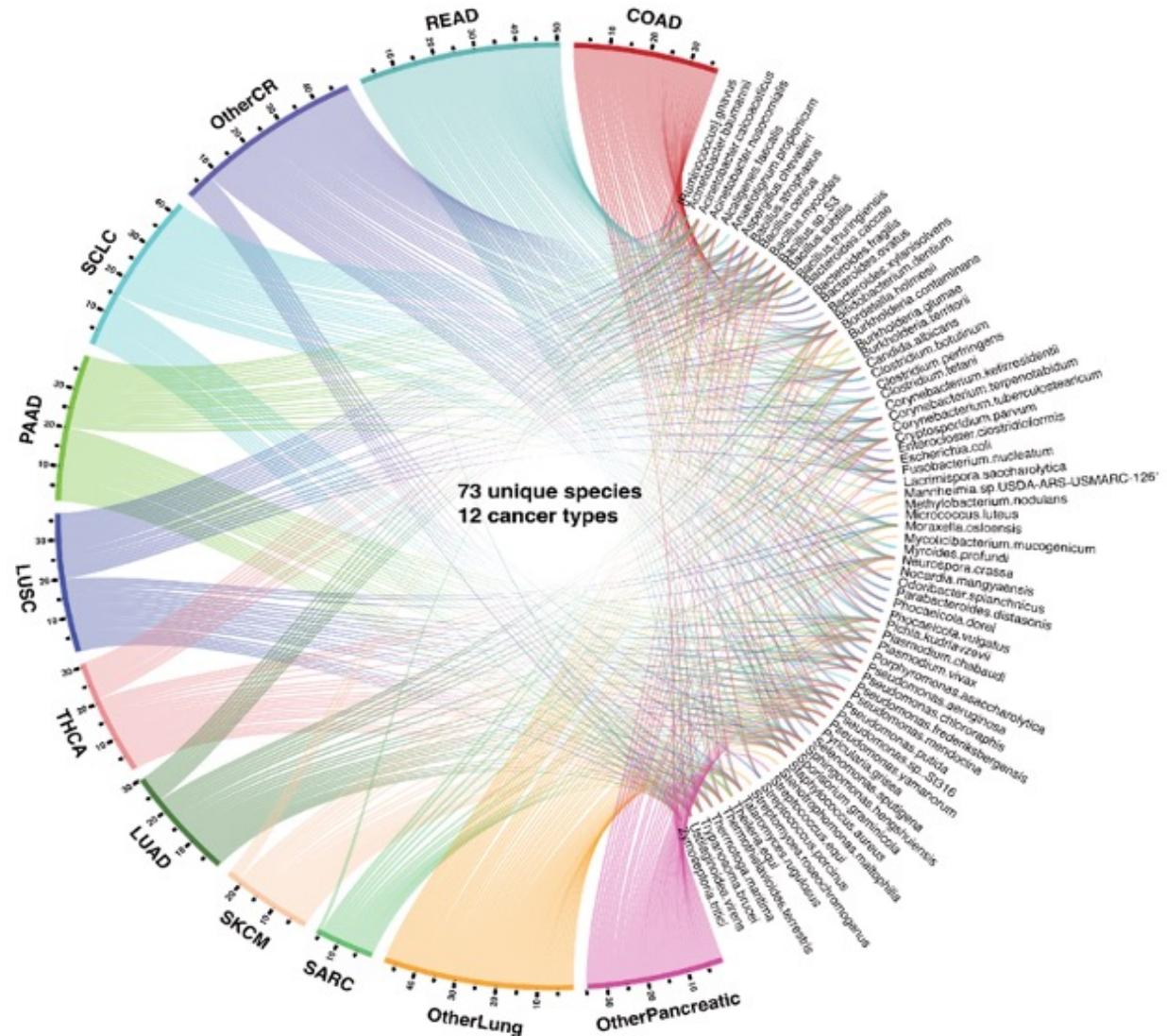
Daniel Spakowicz, PhD

- A deep learning-based Python package, **Microbial Heterogeneous Graph Attention (MEGA)**, for identifying cancer-associated intratumoral microbes.
- Applied MEGA in Oncology Research Information Exchange Network (ORIEN) datasets.
- Enables multiple functionalities for association interpretations and visualizations.

MEGA: A Bioinformatics Tool for Identifying Intratumoral Microbes from the ORIEN Dataset

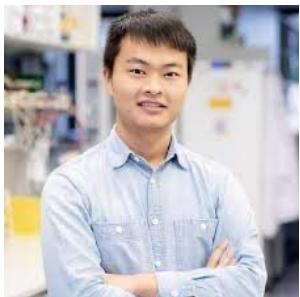
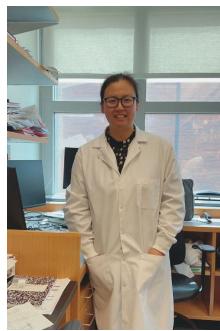
- MEGA enabled the identification of unique microbial communities comprising 73 species across 12 cancer types.
- 15 species were shared across all 12 cancer types.
- 8 species were uniquely shared among COAD, READ, OtherCR.

COAD (colon adenocarcinoma), LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma), OtherCR (other colorectal cancer types not specified), OtherLung (other lung cancer types not specified), OtherPancreatic (other pancreatic cancer types not specified), PAAD (pancreatic adenocarcinoma), READ (rectum adenocarcinoma), SARC (sarcoma), SCLC (small cell lung cancer), SKCM (skin cutaneous melanoma), and THCA (thyroid carcinoma).

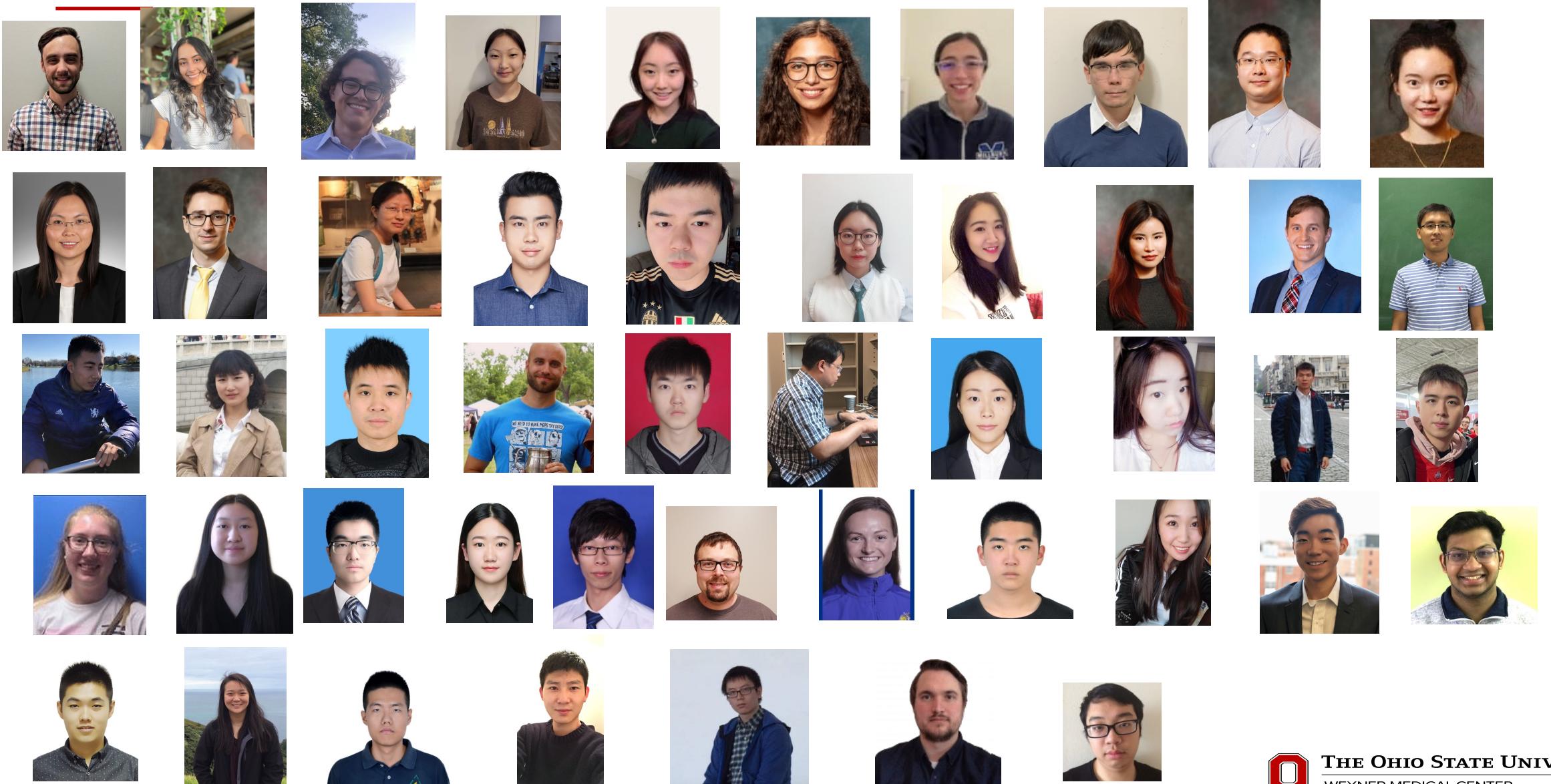


Circos plot representation of the distribution of identified species and cancer types

Acknowledgement



Acknowledgement - BMBL Alumni



Acknowledgement

BMBL @ OSU

All lab members!

Collaborators @ U of Missouri

Dr. Dong Xu

Collaborators @ OSU

Dr. Anjun Ma

Dr. Zihai Li

Collaborators @ Indiana U

Dr. Chi Zhang

Dr. Juixin Wang

Collaborators @ SDU

Dr. Bingqiang Liu

Collaborators @ Duke

Dr. Anru Zhang



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



@QinMaBMBL



Qin.Ma@osumc.edu

