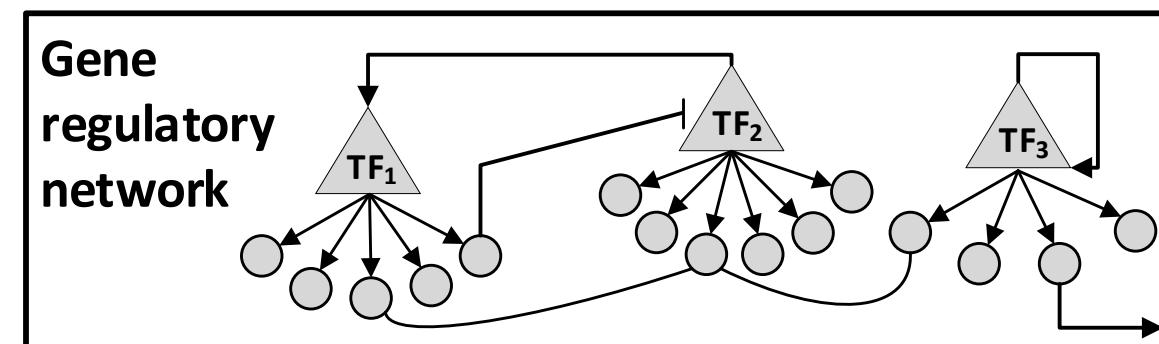
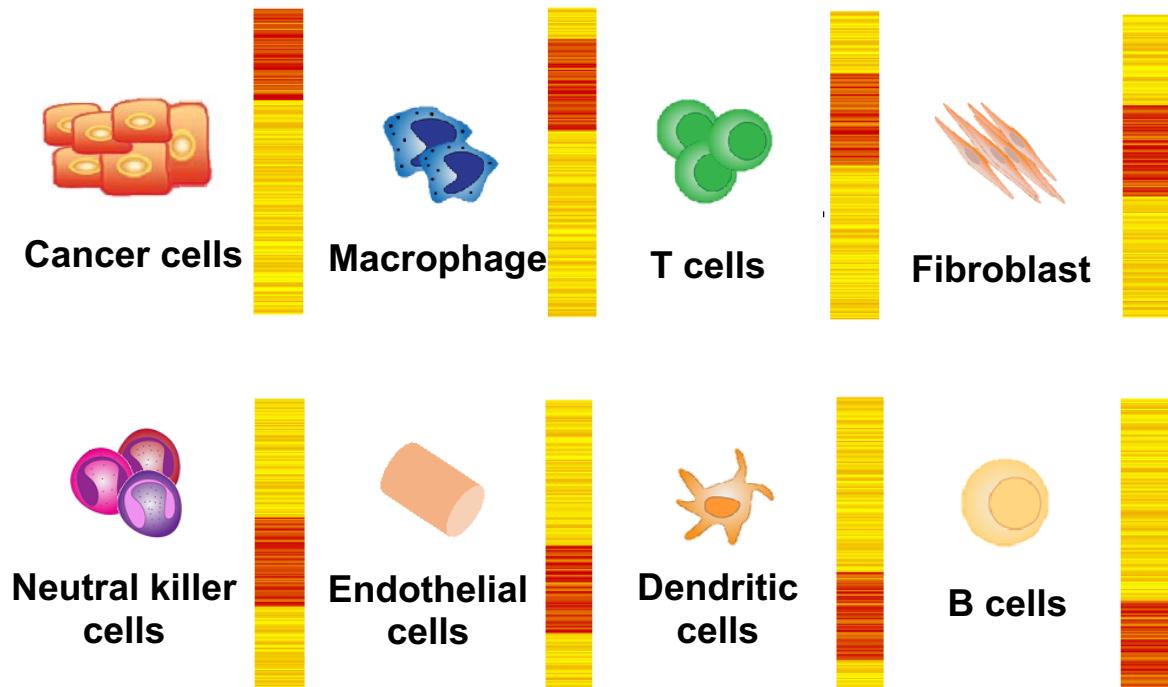
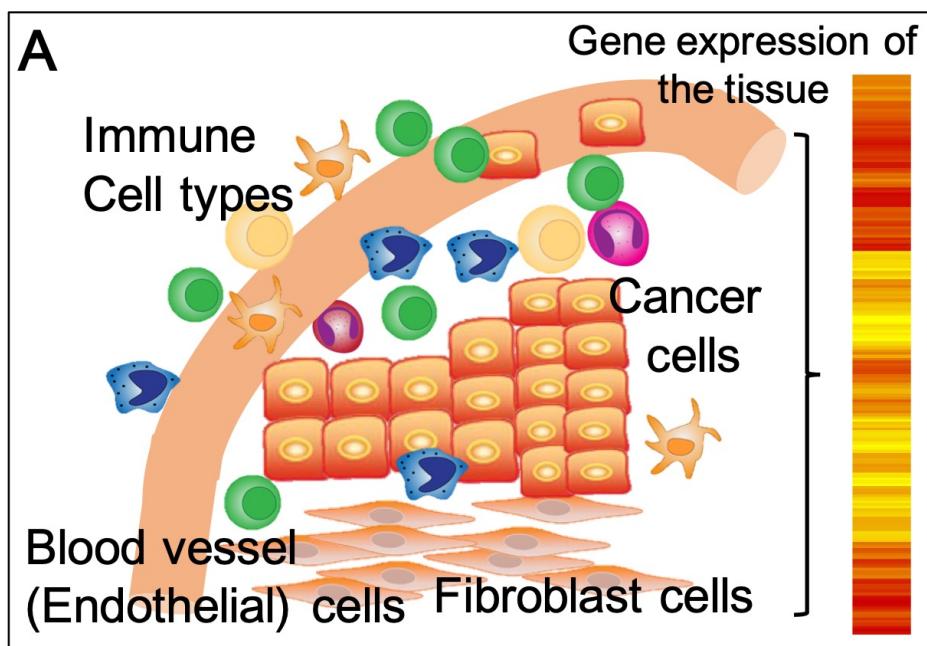
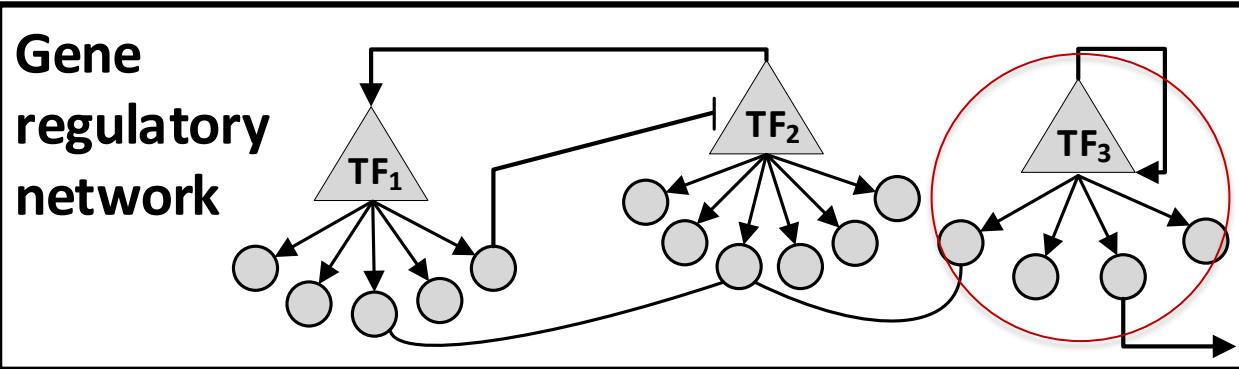


**Network analysis using single-cell multi-omics
Classical network analysis in integrating modalities**

Single-cell gene expression to gene regulation

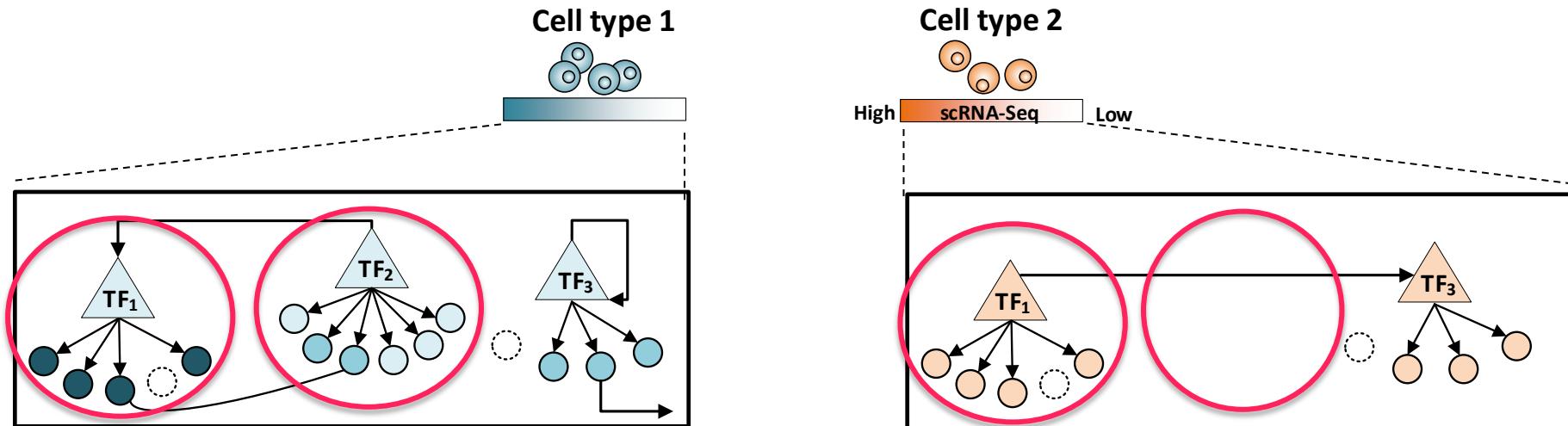


What is cell-type-specific regulon?



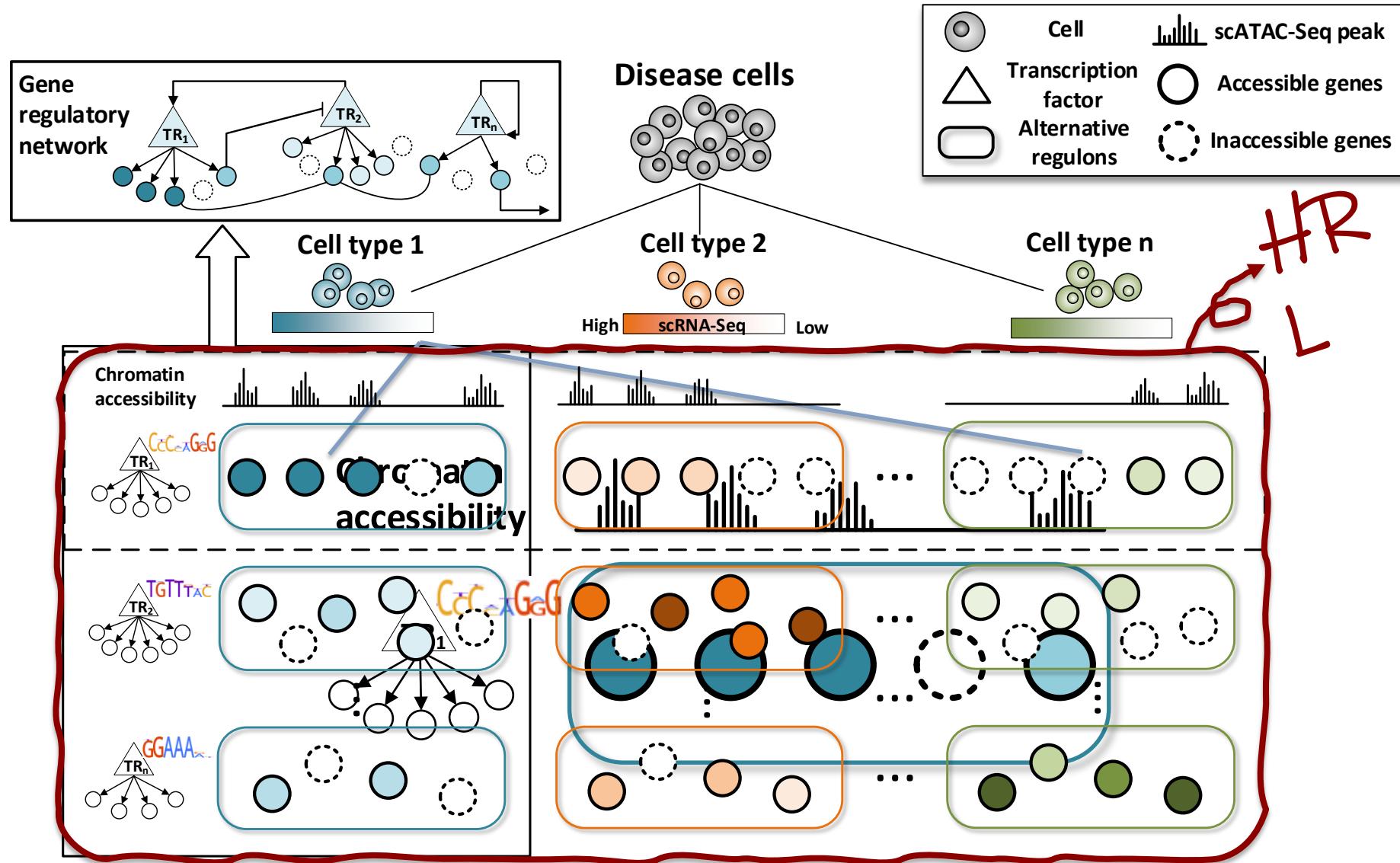
Regulon -- a maximal group of genes co-regulated by the same transcription factor (TF)

It is the fundamental unit in a GRN



Cell-type-specific regulon (CTSR) --
a regulon that is specifically active in a cell type

Heterogeneous Regulatory Landscape



CTSR can reveal unique regulatory mechanisms

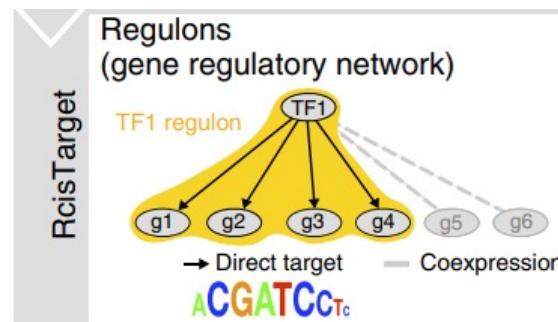
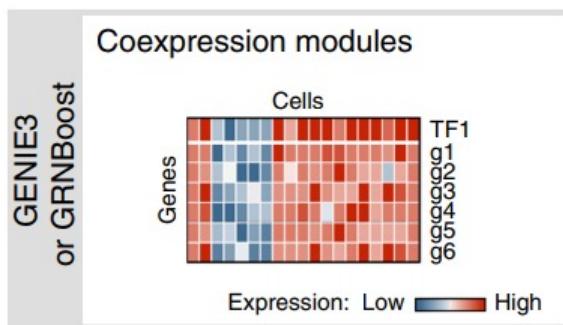
SCENIC: single-cell regulatory network inference and clustering

Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, VÂN ANH HUYNH-THU, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, Joost van den Oord, Zeynep Kalender Atak, Jasper Wouters & Stein Aerts 

Nature Methods 14, 1083–1086(2017) | [Cite this article](#)

23k Accesses | **213 Citations** | 71 Altmetric | [Metrics](#)

<https://aertslab.org/>



Cell

A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain

Cell

Toward Minimal Residual Disease-Directed Therapy in Melanoma

Cell Reports

Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas

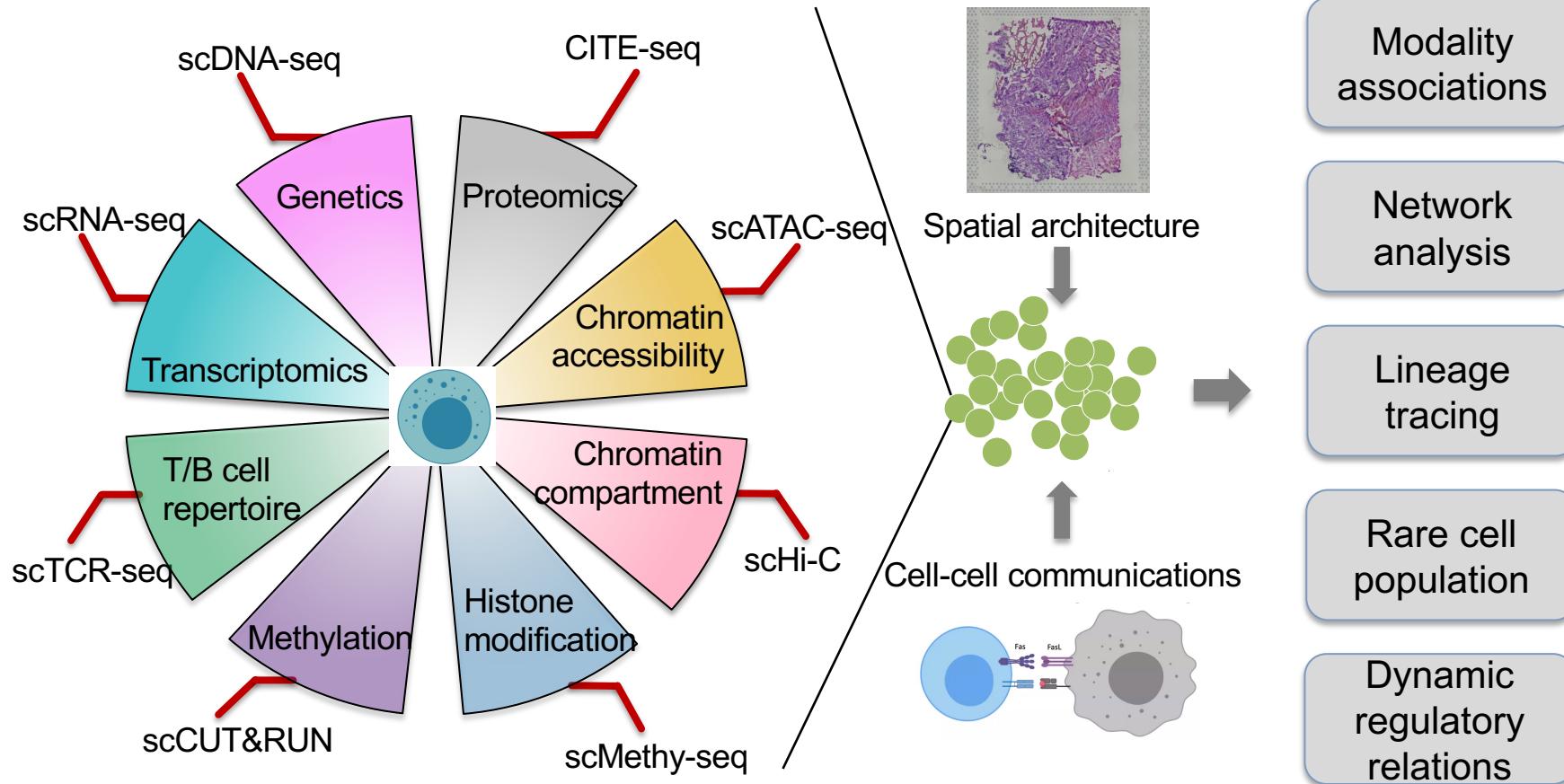
nature neuroscience

RESOURCE

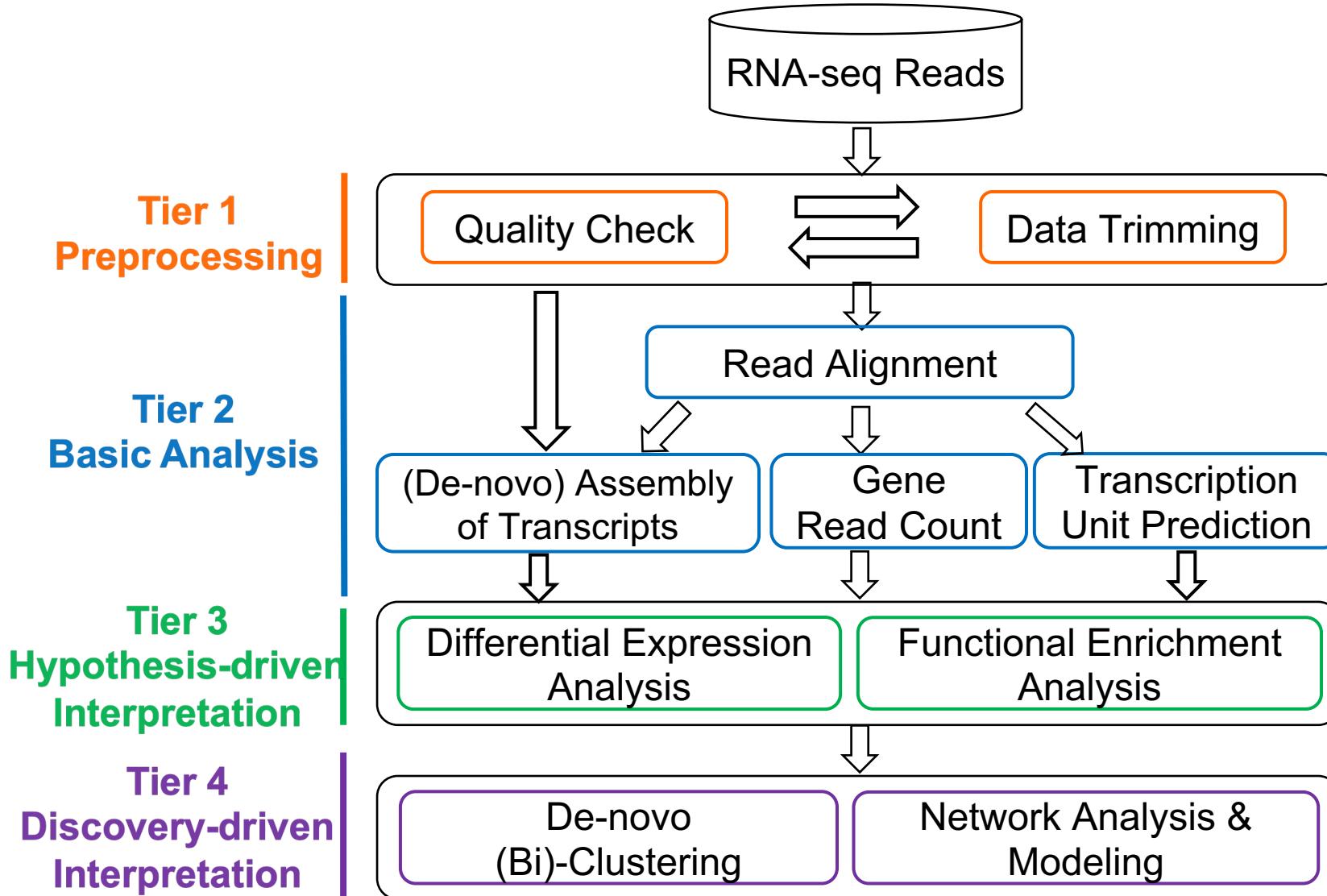
<https://doi.org/10.1038/s41593-019-0393-4>

A single-cell atlas of mouse brain macrophages reveals unique transcriptional identities shaped by ontogeny and tissue environment

scMulti-omics enhance and enable various biological analyses



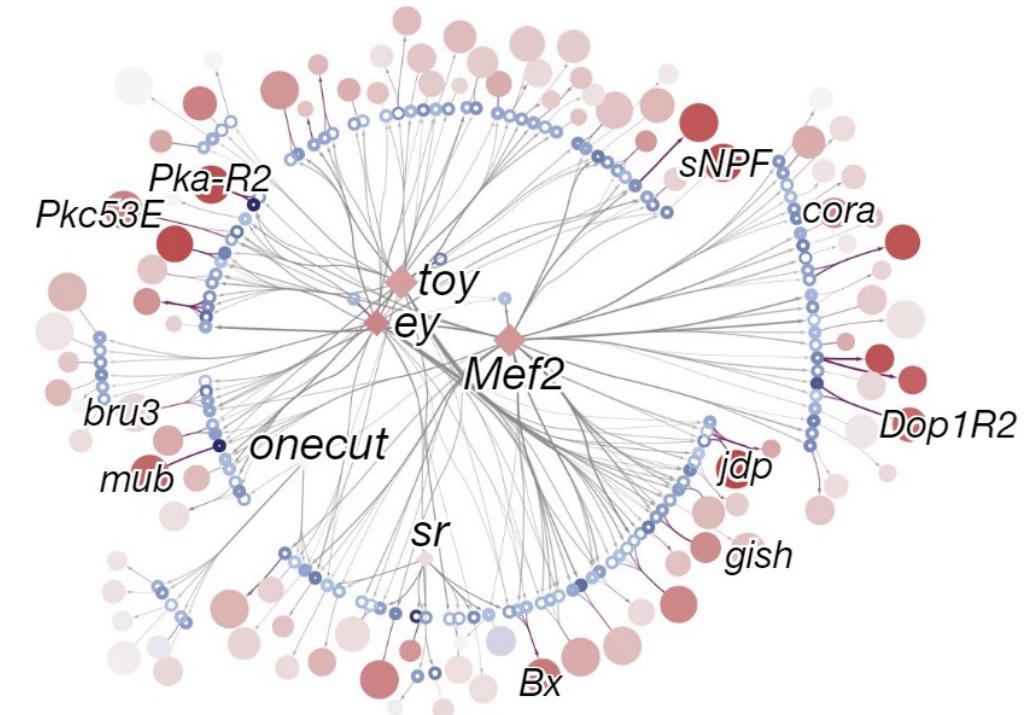
An RNA-seq Data Analysis Pipeline with Critical Thinking



Identifying cell-specific regulatory relations

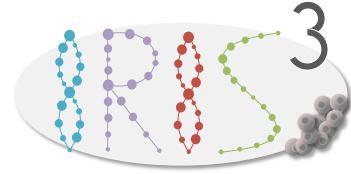
Terminologies and definitions

- **Cistrome**
 - -> TF-peaks
- **Differentially accessible region (DAR)**
 - -> Similar to differentially expressed genes
- **Core-set region/DAR/enhancer**
 - -> cell type specific and constant accessible
- **Enhancer-gene regulatory network (eGRN)**
 - -> TF-enhancer/motif-gene network
- **Dynamic enhancer**
 - -> Enhancers with dynamic accessibility change
- **Switching enhancer**
 - -> Change of TF binding cite in different cell types
- **Cell-type-specific regulation**
 - -> Specifically exist/high in one/few cell types



IRIS3: Integrated Cell-type-specific Regulon Inference Server using Single-cell RNA-Seq data

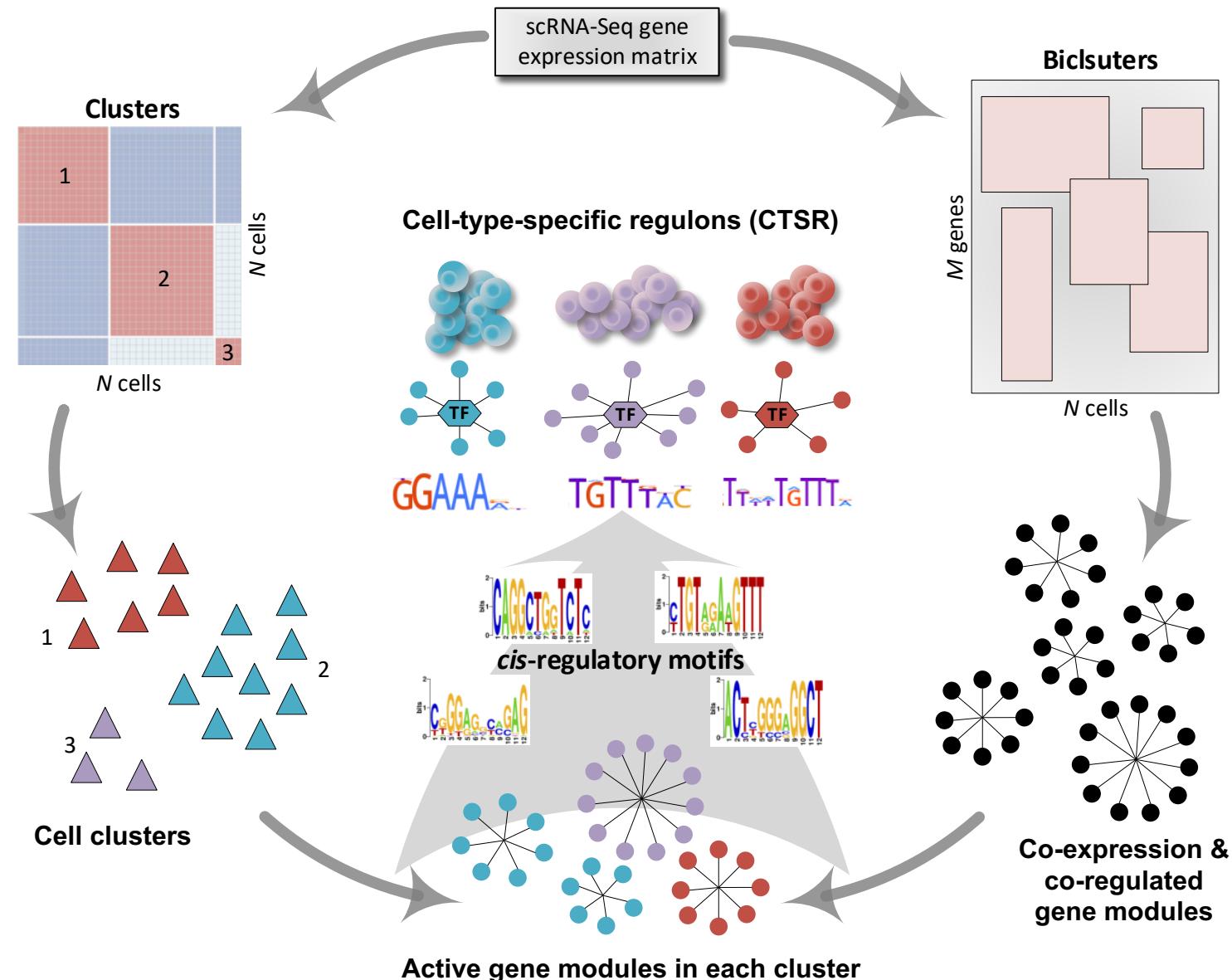
THE OHIO STATE UNIVERSITY
WEXNER MEDICAL CENTER



- First-of-its-kind webserver
- Multi-functional
 - Cell clustering
 - DEG analysis
 - Functional enrichment
 - De novo motif finding
 - Linking to databases
 - Trajectory analysis
 - etc.
- Better performance than SCENIC

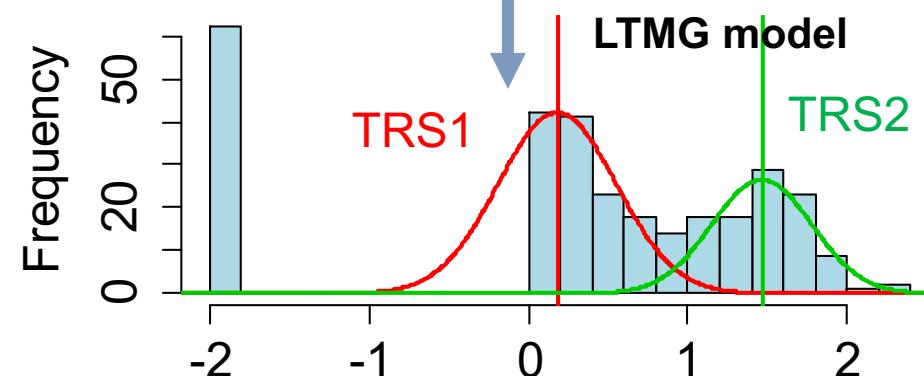
<https://bmbi.bmi.osumc.edu/iris3/>

Anjun Ma, et al, Nucleic Acids Research, (2020)

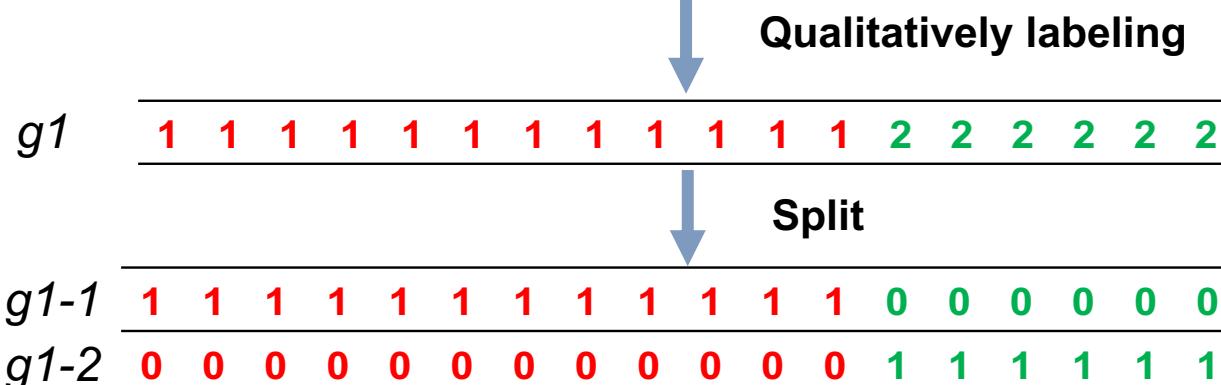


Qualitative representation of gene expression

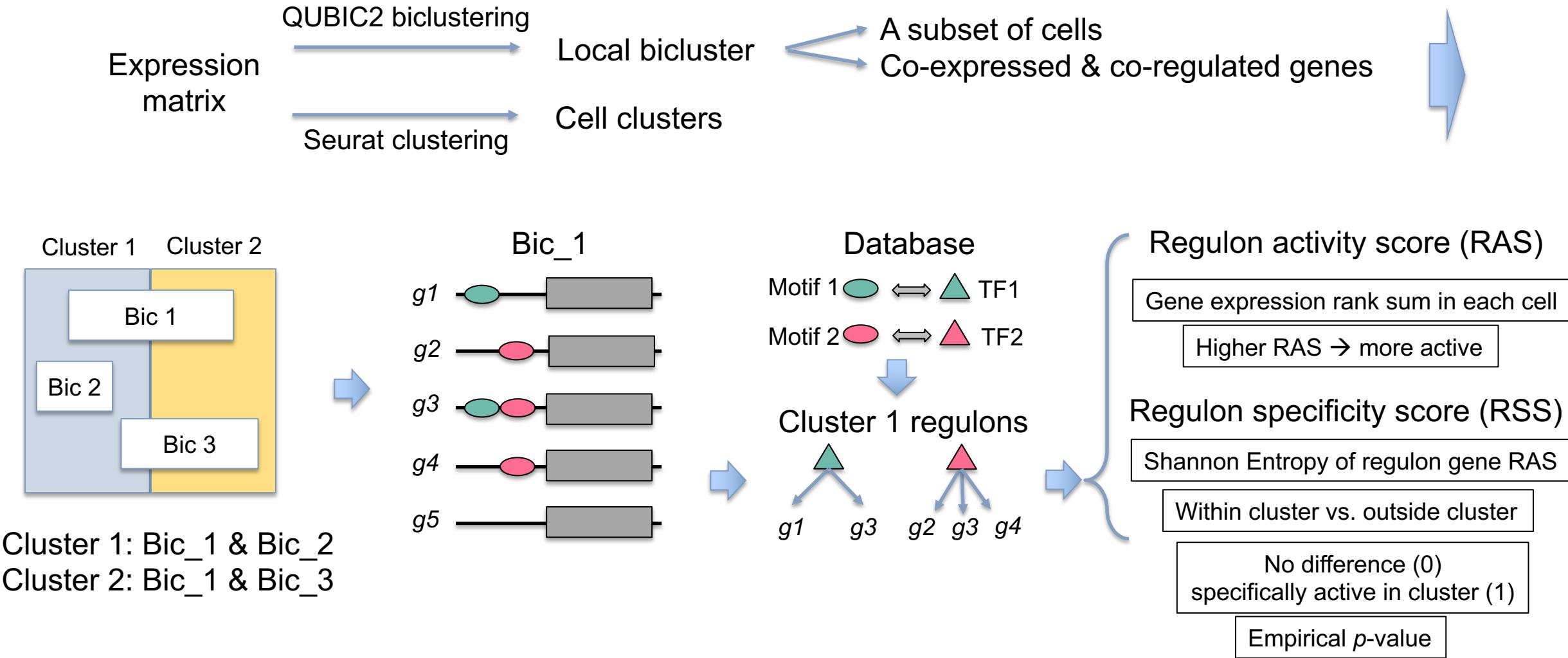
g1 [0.42 | 0.87 | 0.70 | 1.03 | 1.25 | 2.04 | 1.45 | 1.95 | 7.13 | 5.48 | 7.71 | 0.95 | 9.15 | 2.81 | 0.77 | 4.02 | 3.12 | 2.68]



The representation assumed that if a gene receives K possible transcription regulatory signals (TRSs) in all the cells, its expression profile follows a mixture of K left truncated normal distributions.



'1': the gene received a specific TRS under the current condition
 '0' : the gene did not receive the TRS under the current condition



A regulon with significant RSS is a CTSR

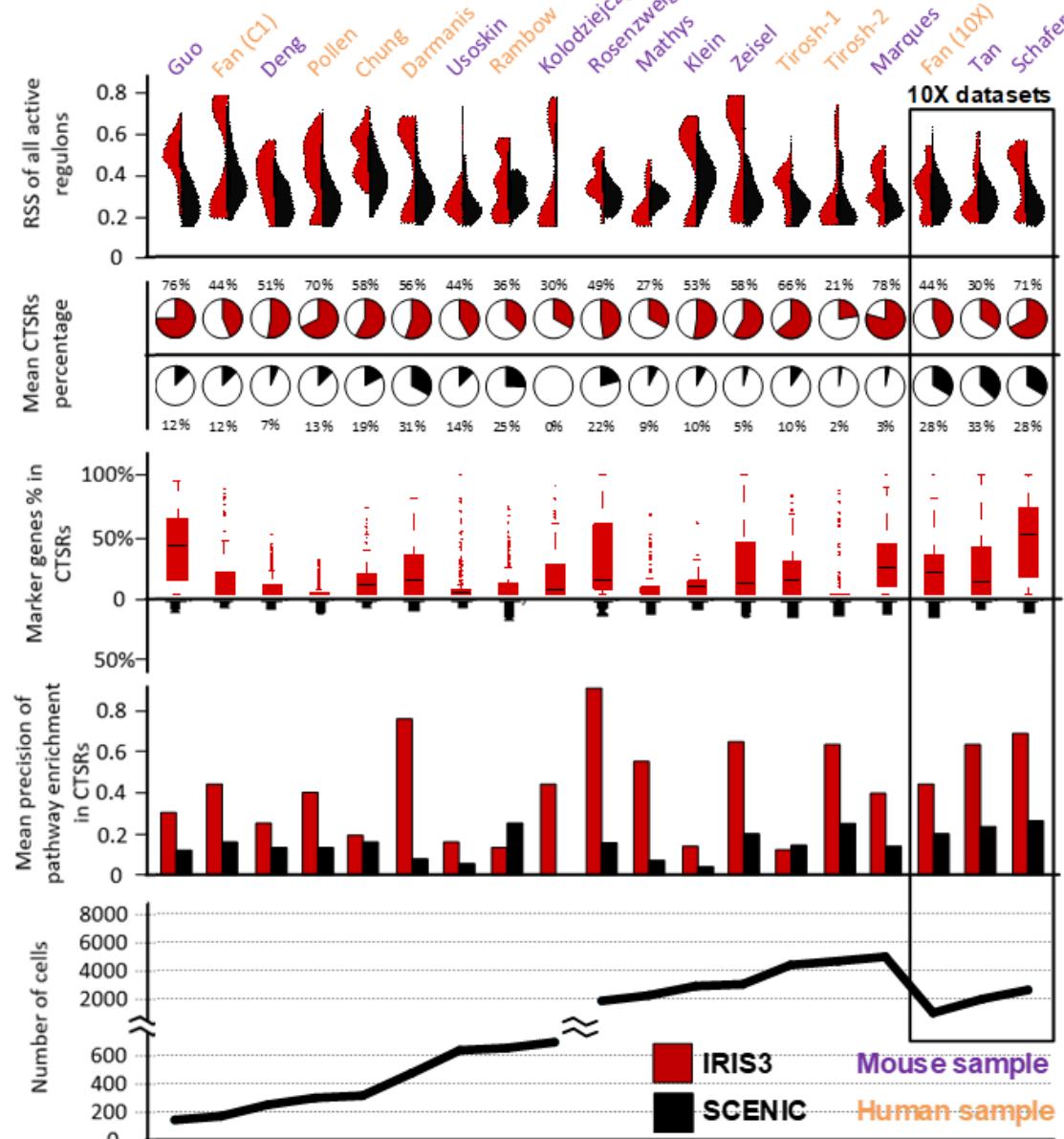
Evaluation: IRIS3 outperformed SCENIC in CTSR identification

19 scRNA-Seq benchmark data

- 100 – 5000 cells
- Both mouse and human
- Both Smart-Seq and 10X

Four criteria

- RSS significance
- Number of CTSRs
- Differentially expressed gene coverage
- Pathway enrichment



IRIS3 is a multi-functional scRNA-Seq server

Regulon specificity score: 0.83854466 Number of genes: 90 Number of differentially expressed genes: 68

Differentially expressed gene	Gene Symbol	Copy	Ensembl ID	Copy	Gene UMAP plot
★	Adamts4		ENSMUSG00000006403		Display
★	Apod		ENSMUSG00000022548		Display
★	Aspa		ENSMUSG00000020774		Display
★	Bpgm		ENSMUSG00000038871		Display

TF: KLF4

Motif name	Motif logo	Motif details	Motif comparison
CT1-R1-Motif-1		Open	Open

[Heatmap](#)

[Send gene list to Enrichr](#)

[ATAC-seq peak enrichment](#)

[Additional TAD covered genes](#)

[Regulon UMAP plot](#)

[Trajectory plot](#)

Avg_logFC	Pct.1	Pct.2	Adjusted p-value
3.7275	1	0.607	0
3.7221	0.984	0.182	0
3.6925	0.989	0.112	0
3.4366	0.967	0.265	0
3.4212	0.946	0.168	0
3.311	0.999	0.168	0
3.3043	0.787	0.119	0
3.2569	0.994	0.14	0
3.2478	0.999	0.397	0
3.2473	0.971	0.136	0

Previous 1 2 3 4 5 ... 10 Next

Trans

escri

Ch

SOX

OLI

SUZ

WT

SM

Regulon Plot Colored by cell clusters

[Download\(PDF\)](#)

Cell cluster: 1_oligodendrocytes, 4_microglia, 6_endothelial_mural, 7_astrocytes_ependymal, 5_interneurons, 3_pyramidal_SS, 2_pyramidal_CA1

Regulon Plot Colored by CT1-R1 Score

[Download\(PDF\)](#)

Regulon Score: max=191, median=75%Q:77, 25%Q:0, min=0

UMAP Plot Colored by Cell Clusters

[Download\(PDF\)](#)

Cell cluster: 1_oligodendrocytes(820), 2_pyramidal_CA1(399), 3_pyramidal_SS(399), 4_microglia(98), 5_interneurons(290), 6_endothelial_mural(235), 7_astrocytes_ependymal(224)

UMAP Plot Colored by CT1-R1 Score

[Download\(PDF\)](#)

Regulon score: max=150, median=75%Q:77, 25%Q:0, min=0

Optimal Alignment

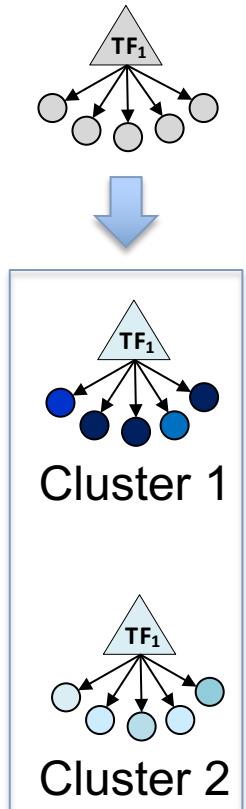
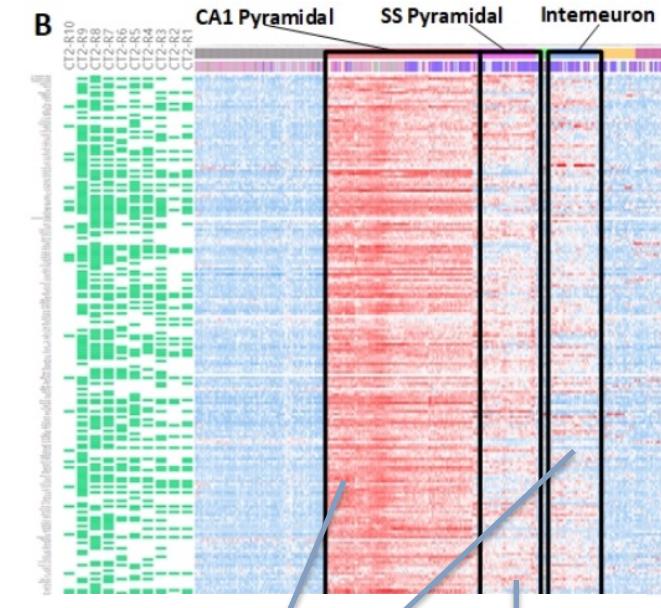
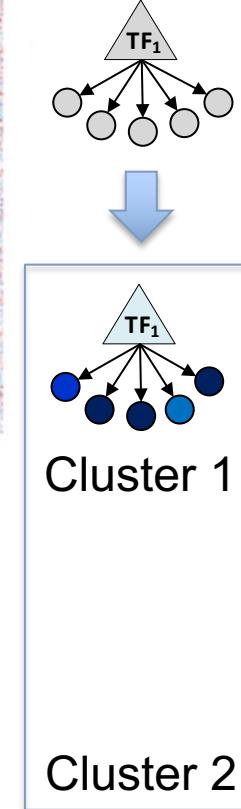
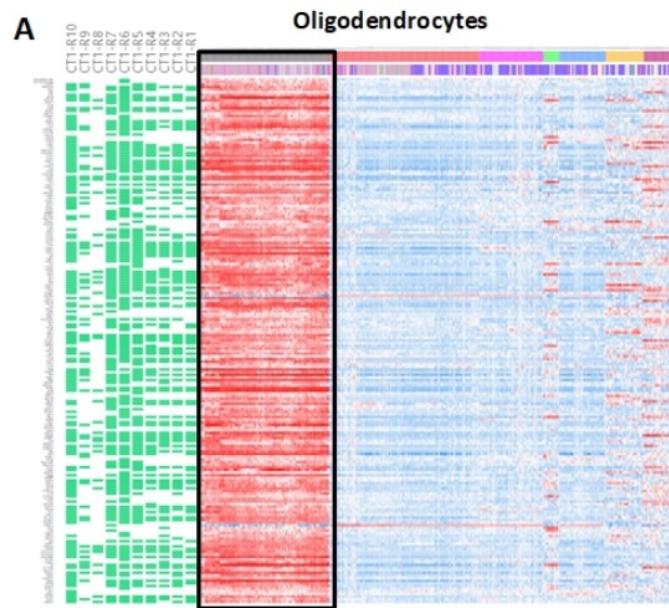
format

SALL4_MOUSE.H11MO.0.A

Optimal Alignment: GGGCTGCGC

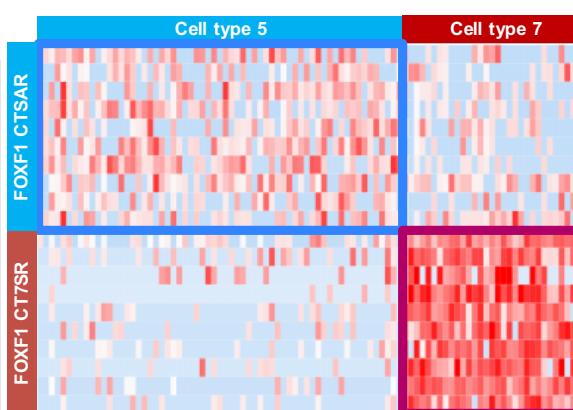
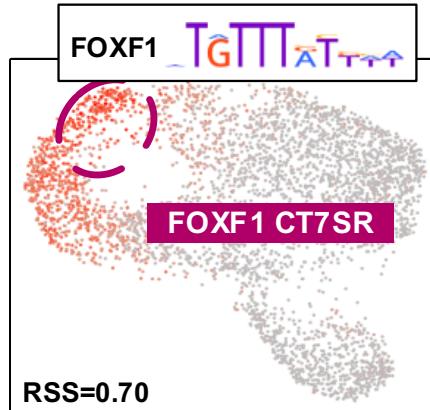
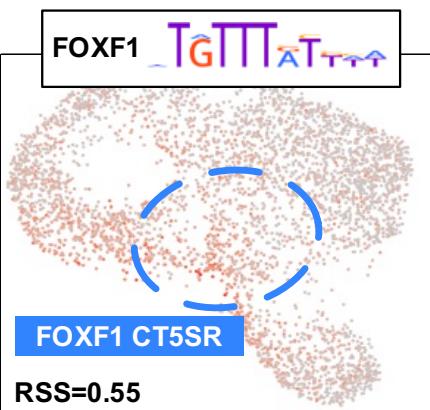
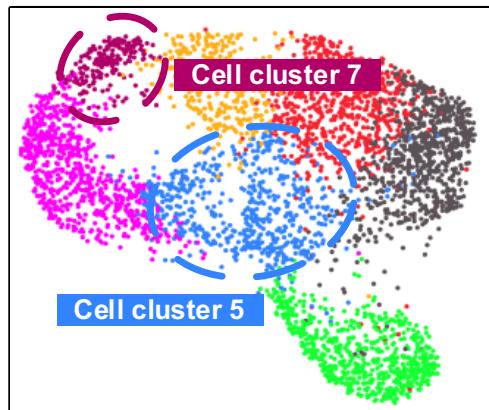
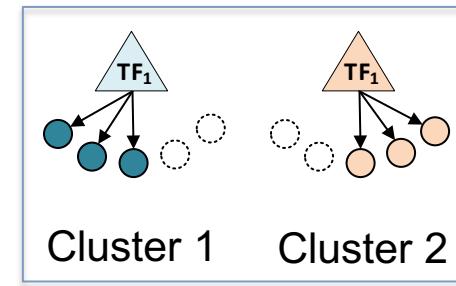
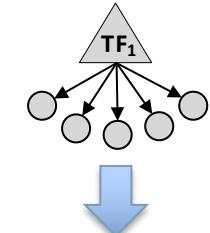
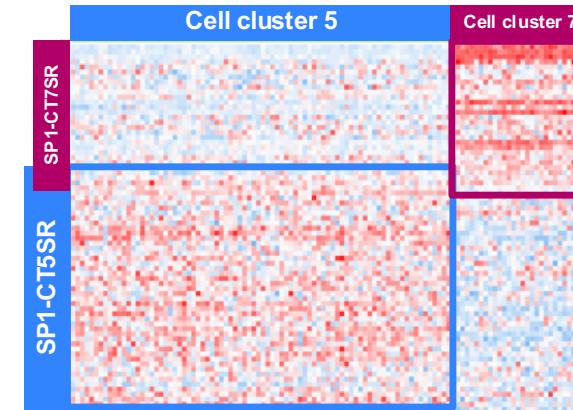
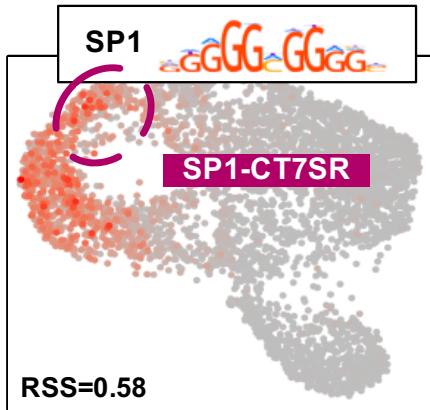
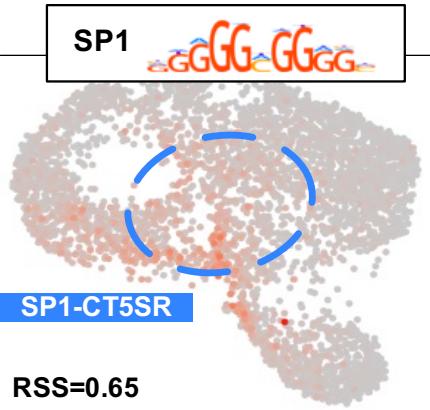
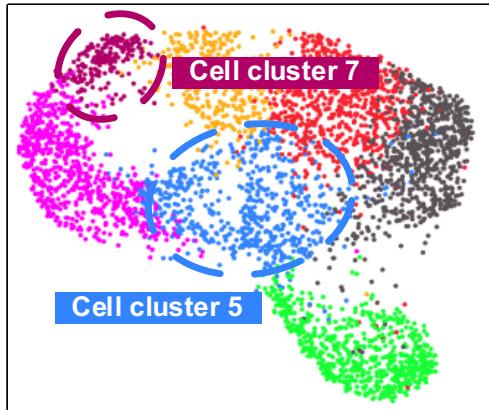
format: GGGCTGCGC

IRIS3 can identify CTSRs differentially activated in neuron cells



Dr. Phillip Popovich, Chair
Department of Neuroscience
Ohio State University

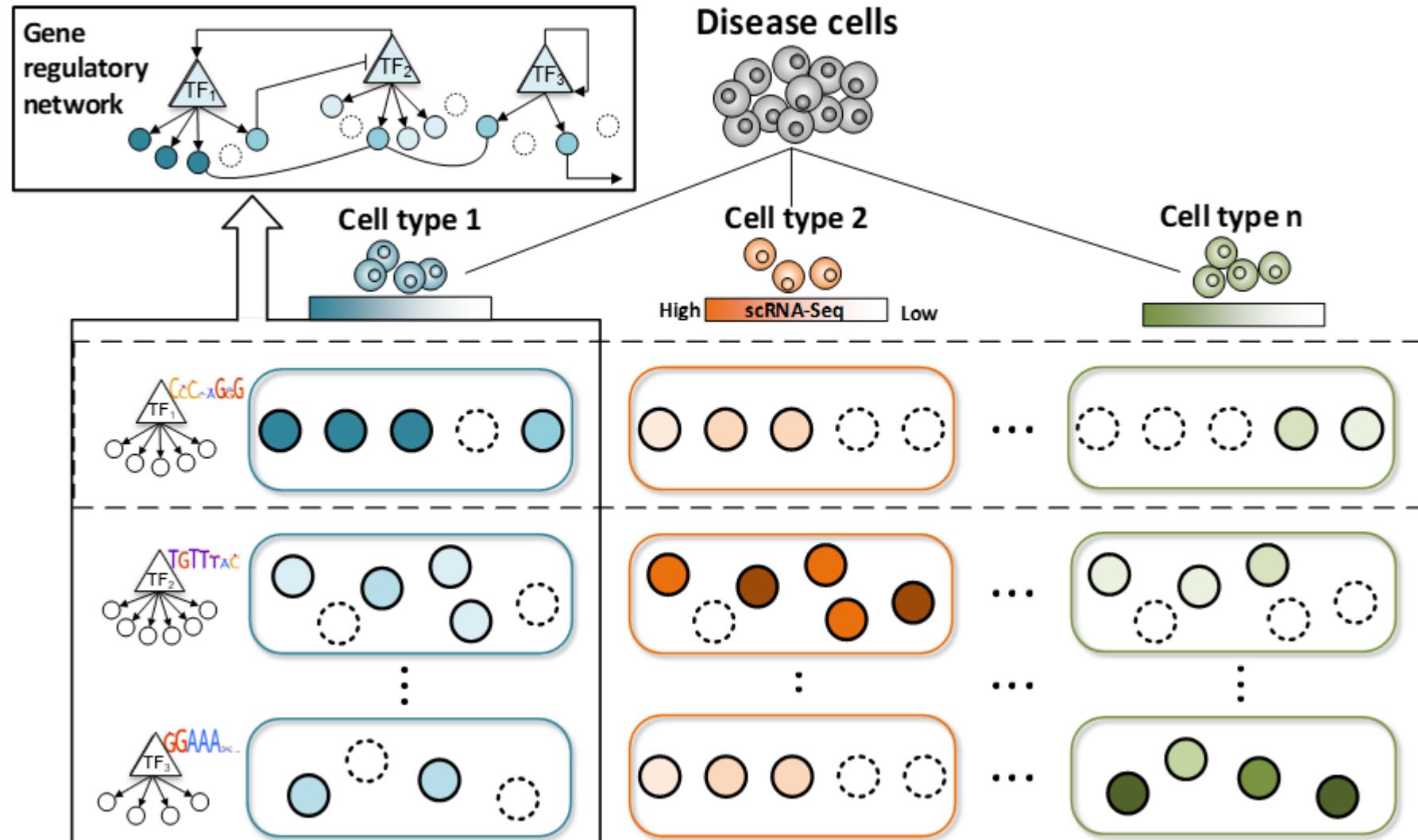
IRIS3 can identify alternative CTSR patterns in immuno-oncology



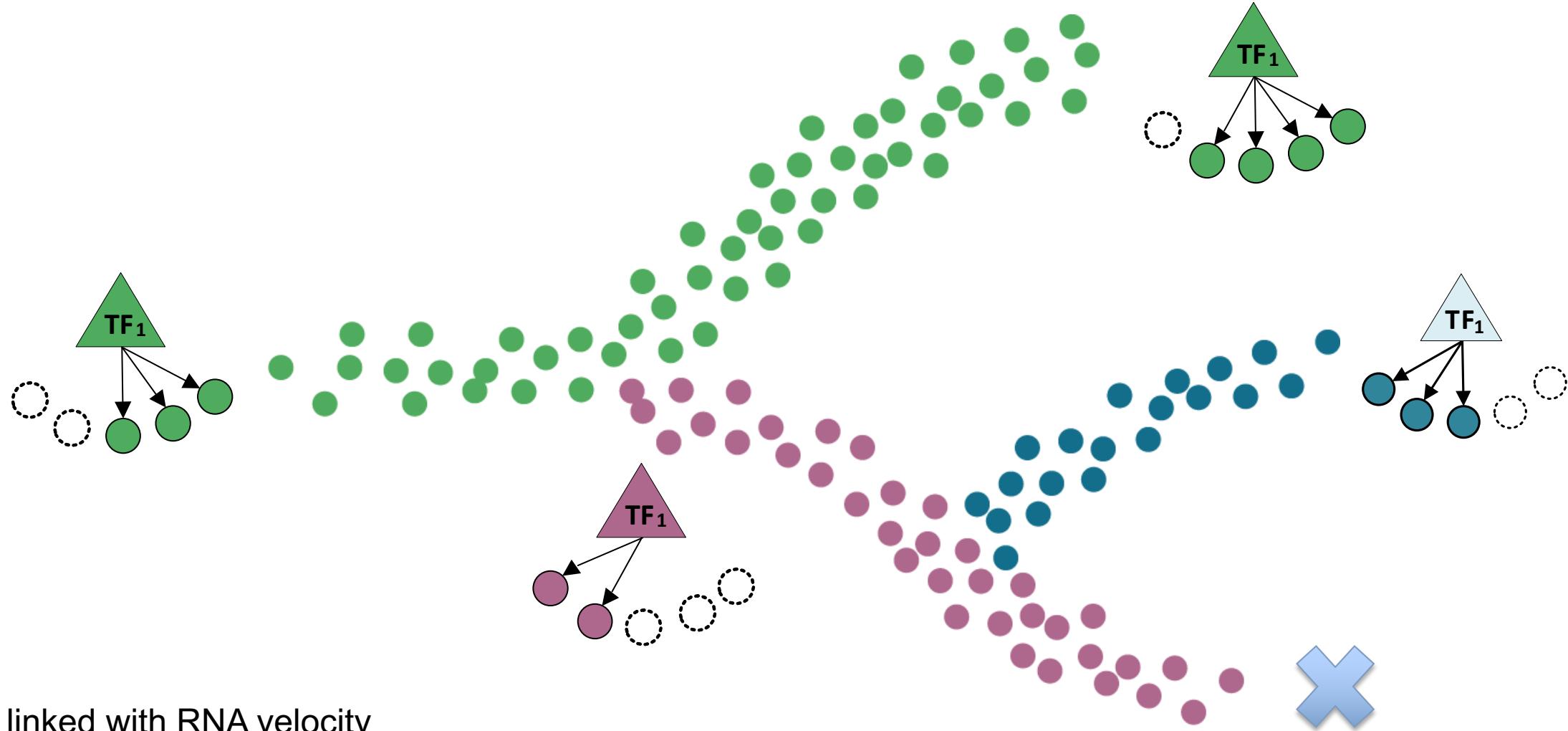
An immuno-informatics tool

Dr. Hazem Ghoneim
Molecular Biology and Cancer Genetics
Ohio State University

TF regulatory landscape



Dynamic TF regulations

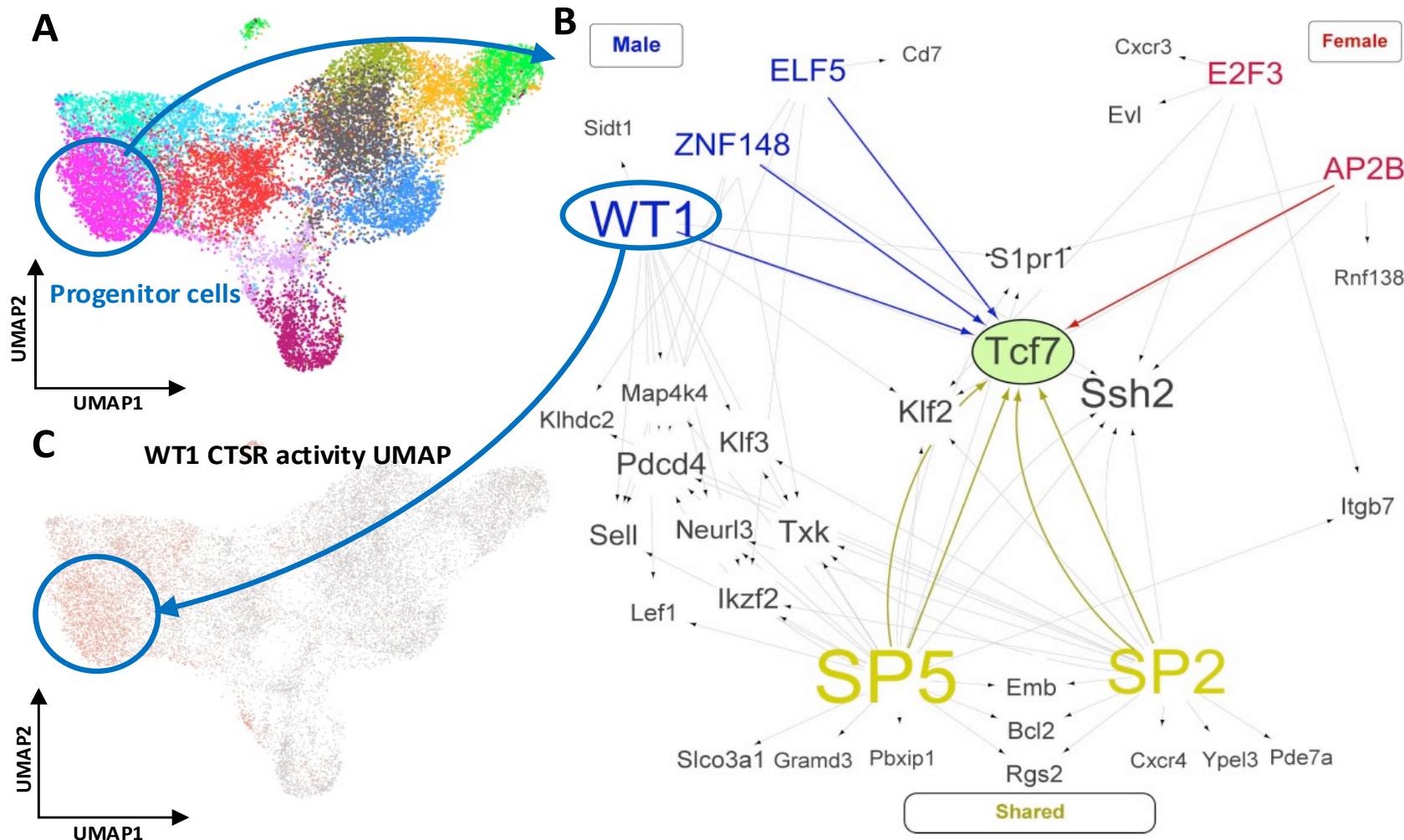


Can be linked with RNA velocity

IRIS3 elucidates the sex-biased regulatory mechanisms in progenitor exhausted T cells in bladder cancer



Dr. Zihai Li, Director
Pelotonia Institute for Immuno-Oncology
Ohio State University



Modeling with heterogeneous graph transformer Transformer model

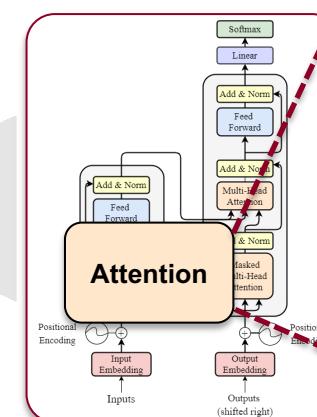
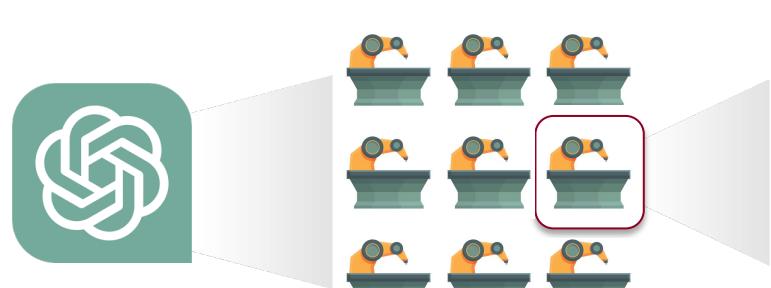
Understand transformer

Transformer is a powerful deep learning algorithm, widely applied to many fields:

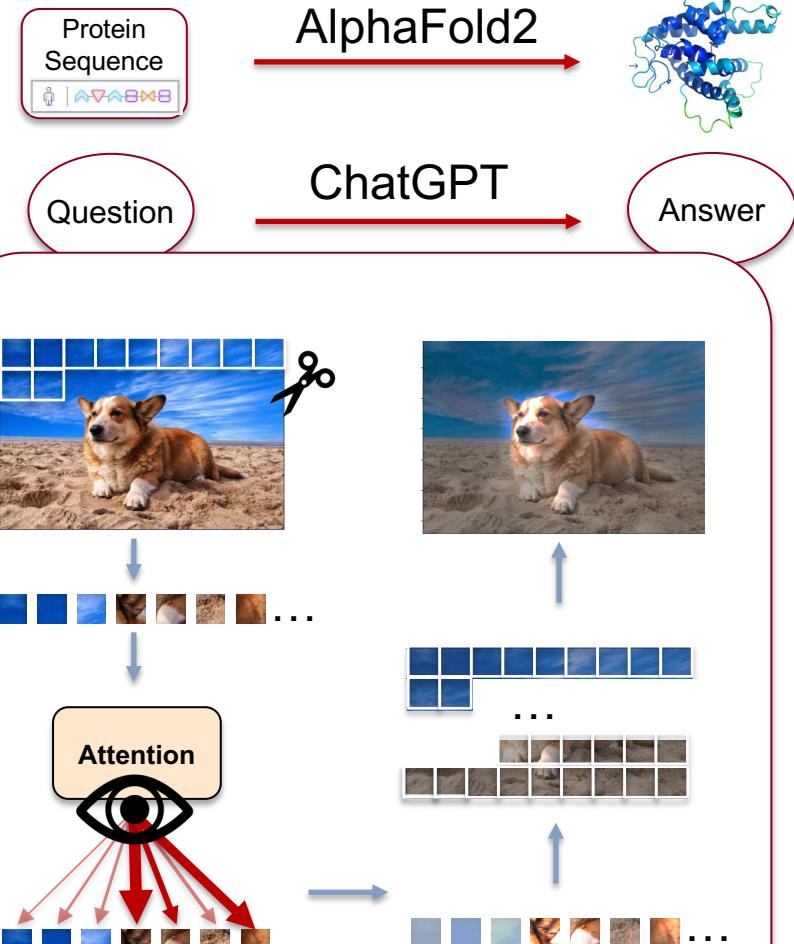
- Natural Language Processing, like Question Answering (ChatGPT), Protein Structure Prediction (AlphaFold2)
- Audio/Speech, like Recognition and Detection (SpeechT5)
- Computer vision, like Medical Image Segmentation (UNETR)
- ...

So, what is the Transformer?

- If ChatGPT is a factory, the Transformer is its machine. Many Transformers are stacked together to drive production.



Transformer-based model



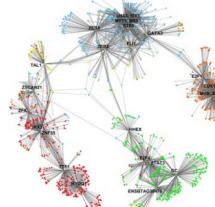
Various graph transformer model

Graph Transformer

Big data → Knowledge

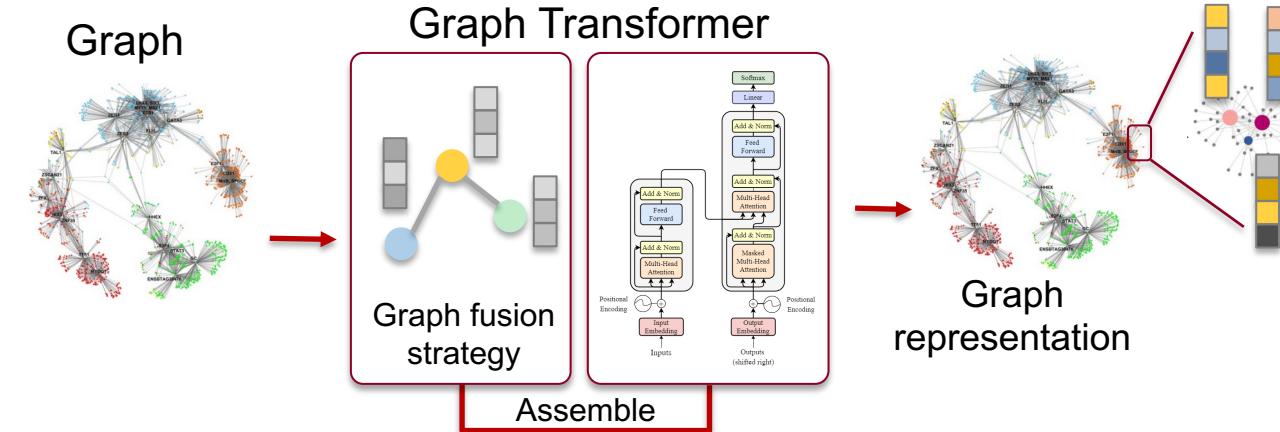


Cell-cell similarity networks



Gene interaction networks

Graph Transformer brings the power of Transformers to these **complex, heterogenous** and **large-scale** graph data



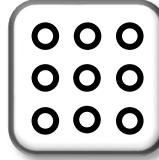
Why Graph?

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

A stronger applicative power with large model,
pre-trained model,
foundation model,
multimodal,
transfer learning...

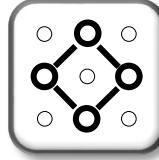
Graph transformer for single-cell omics

Data representation



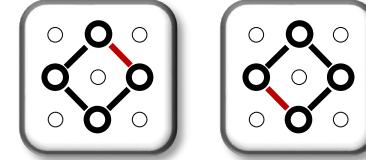
Summarizing, describing, & visualizing features

Factual prediction



Pattern recognition and forecasting

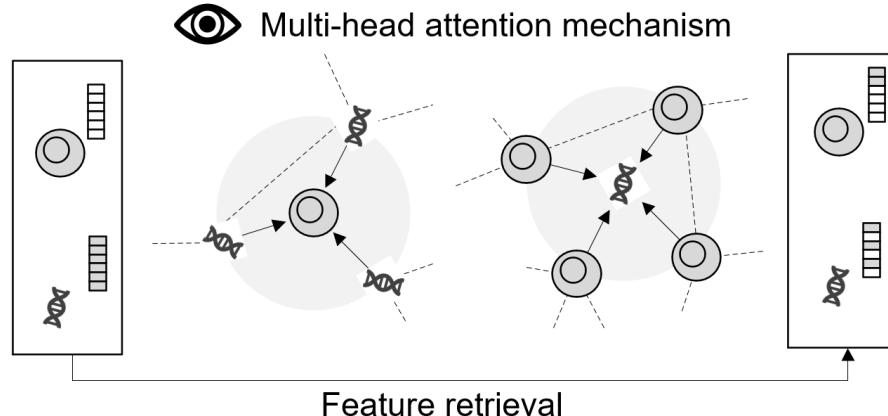
Causal inference



Understanding and reasoning

Question:

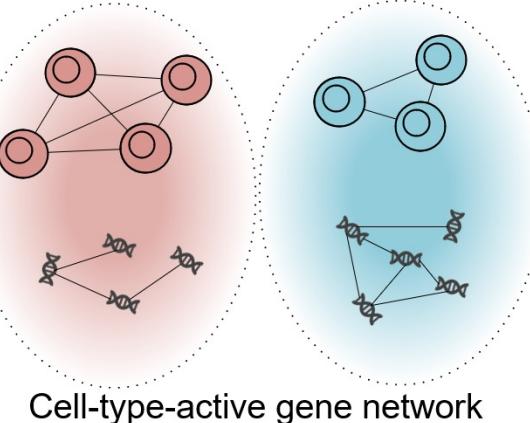
What happened?
What is the occurrence of Y in X?



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

Question:

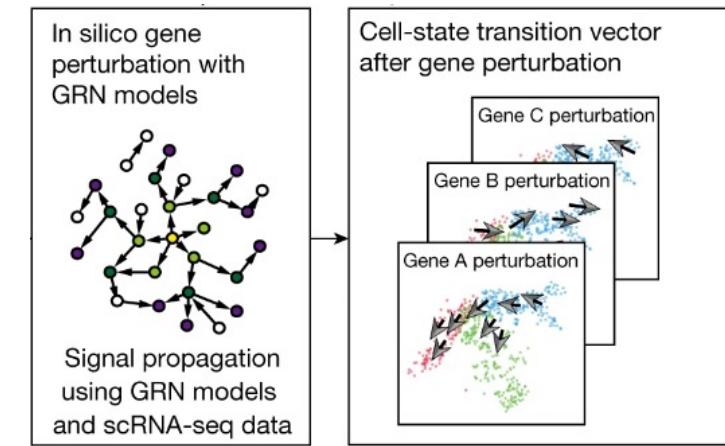
What will happen?
Are X more likely to have Y?



Cell-type-active gene network

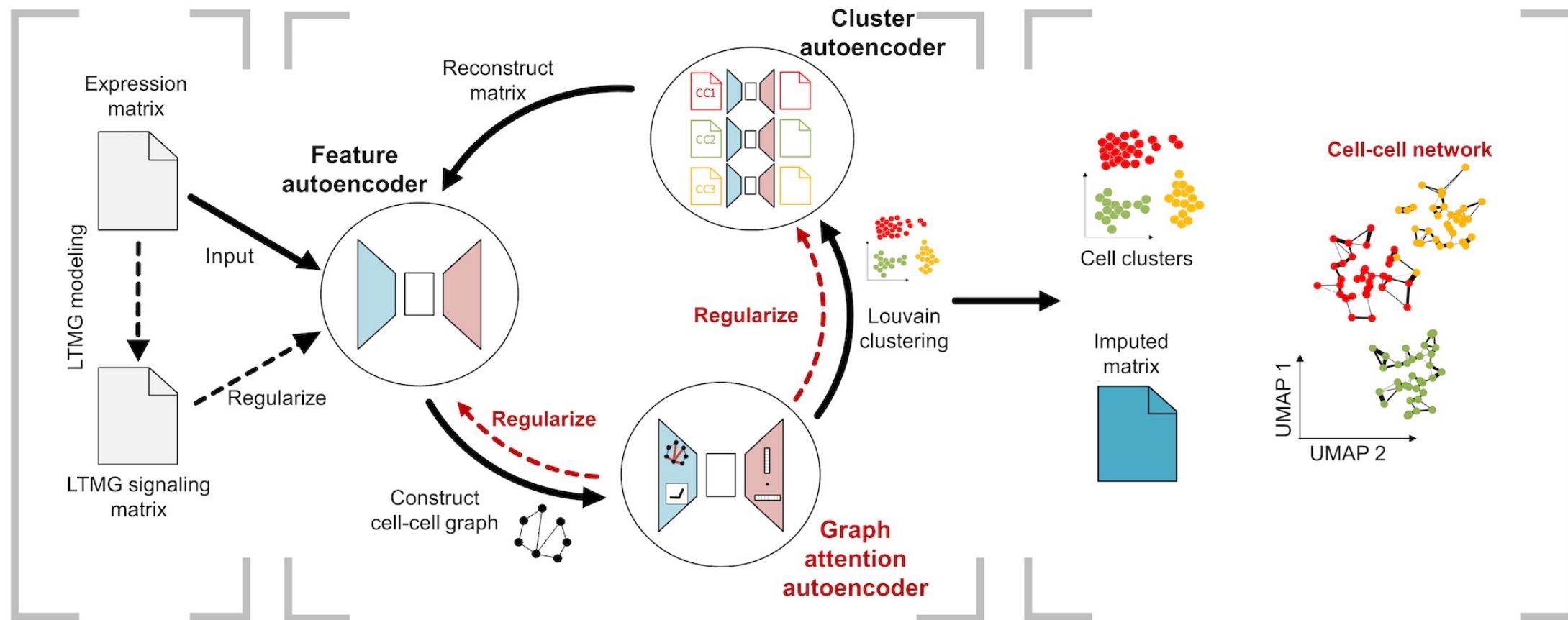
Question:

What will happen if ...?
If we change X, how would it change Y?



Kamimoto, Kenji, et al. "Dissecting cell identity via network inference and in silico gene perturbation." *Nature* (2023): 1-10.

scGNN: in-house graph transformer for cellular heterogeneity



Dr. Dong Xu
Professor, MU

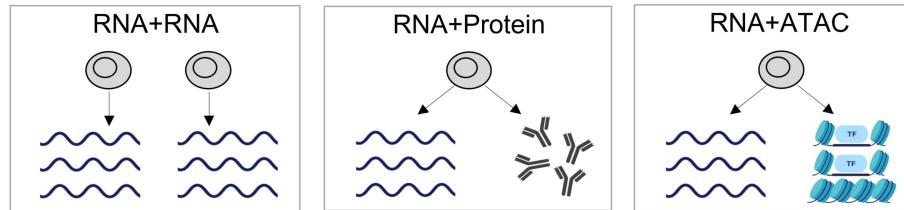


Dr. Juexin Wang
Assis Prof, IU

- 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 260 times in two years.

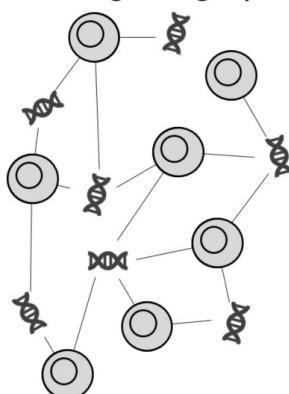
Juexin Wang*, Anjun Ma*, et al., scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature Communications*. (2021)

DeepMAPS: in-house graph transformer for single-cell gene network

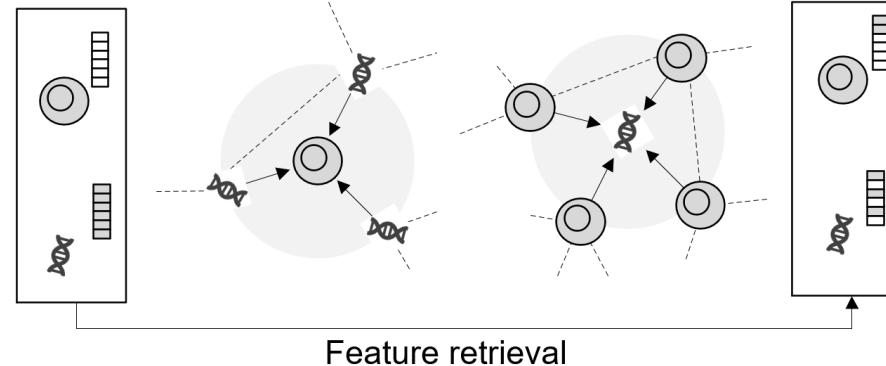


- Learn the **joint embedding** of cells and genes
- Calculate “**attention**” scores (importance) b/w cells and genes
- Learn cell clusters and gene clusters simultaneously

Cell-gene graph

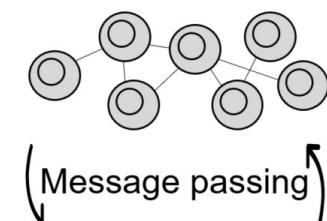


Multi-head attention mechanism

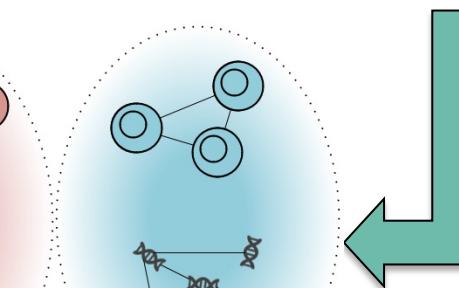


Feature retrieval

Cell graph



Gene graph



Cell-type-active gene network

Article | Open Access | Published: 21 February 2023

Single-cell biological network inference using a heterogeneous graph transformer

Anjun Ma, Xiaoying Wang, Jingxian Li, Cankun Wang, Tong Xiao, Yuntao Liu, Hao Cheng, Juixin Wang, Yang Li, Yuzhou Chang, Jinpu Li, Duolin Wang, Yuexu Jiang, Li Su, Gang Xin, Shaopeng Gu, Zihai Li, Bingqiang Liu✉, Dong Xu✉ & Qin Ma✉

Nature Communications 14, Article number: 964 (2023) | Cite this article

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

MarsGT: in-house graph transformer for rare population and biomarker

- scMulti-omics has the potential to better characterize rare cell populations, yet no existing tools
- MarsGT offers biological insights and suggests potential strategies for early detection and therapeutic intervention of disease.

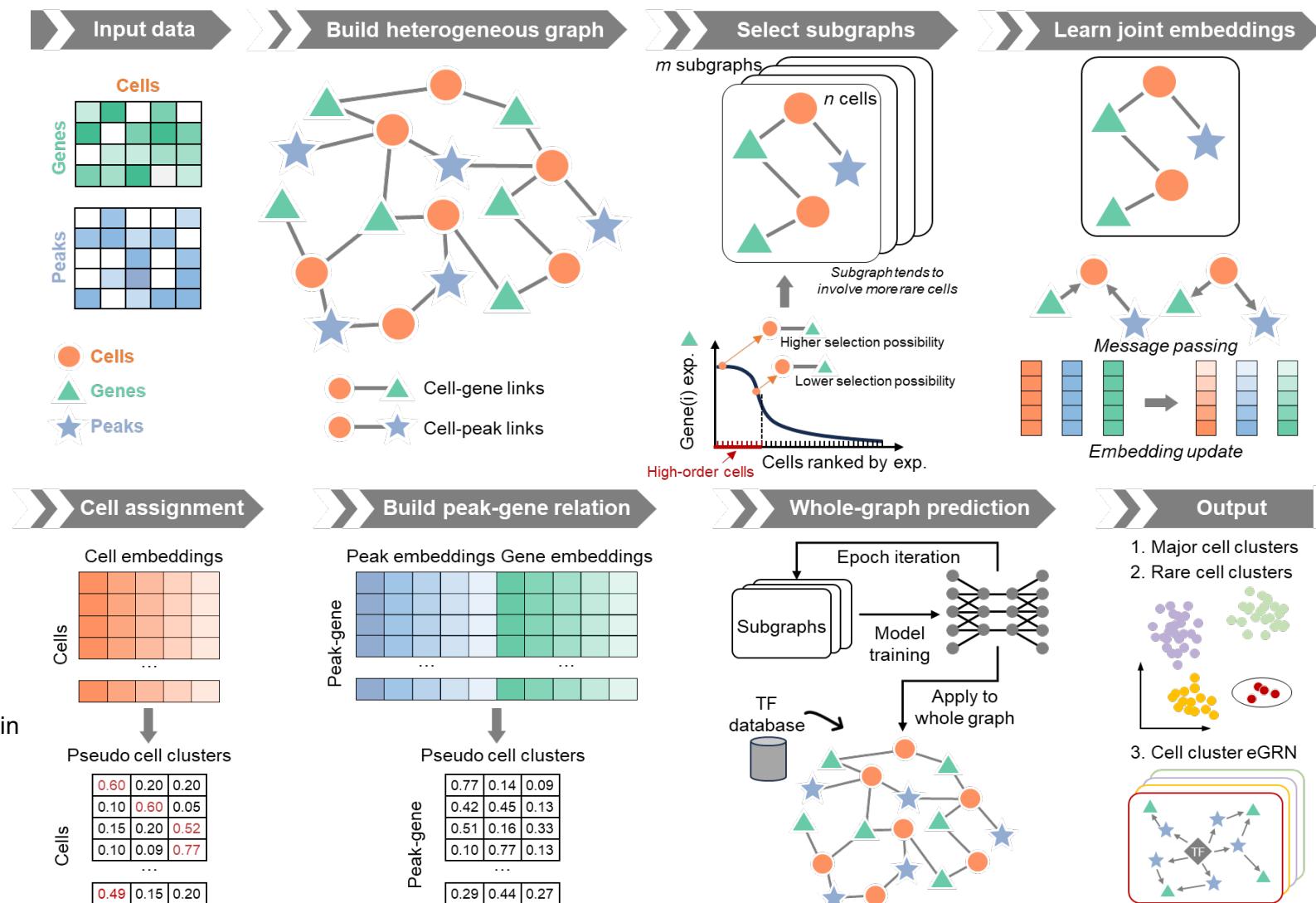


Xiaoying Wang
OSU



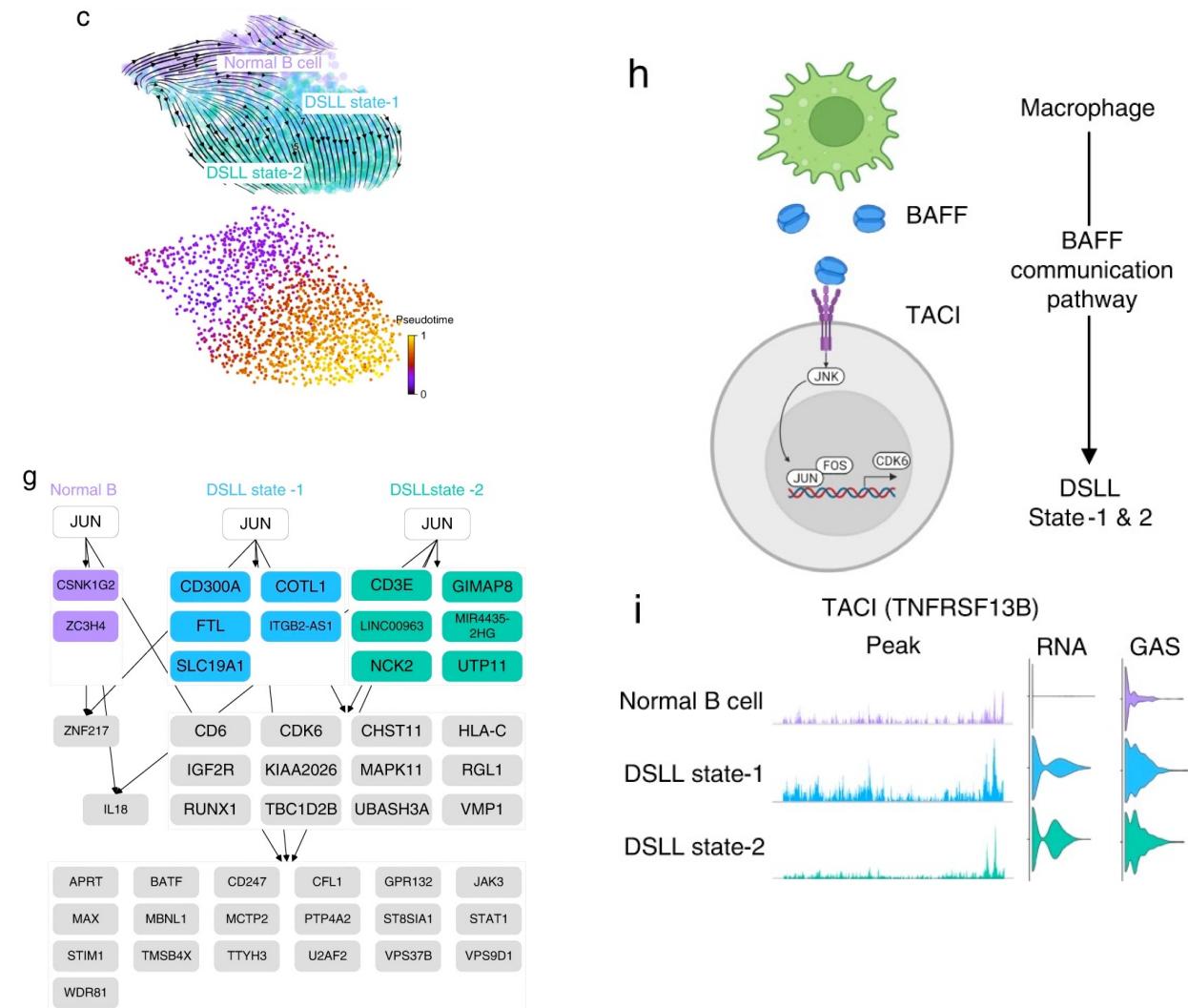
Dr. Gang Xin
OSU

Unpublished results



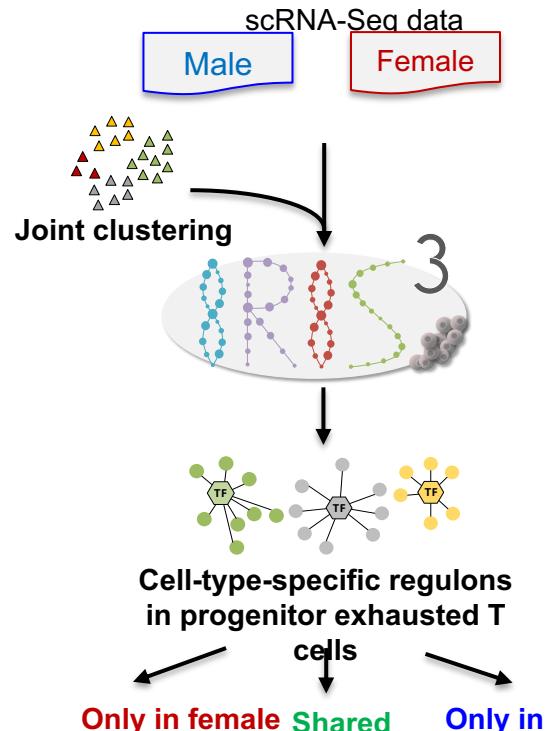
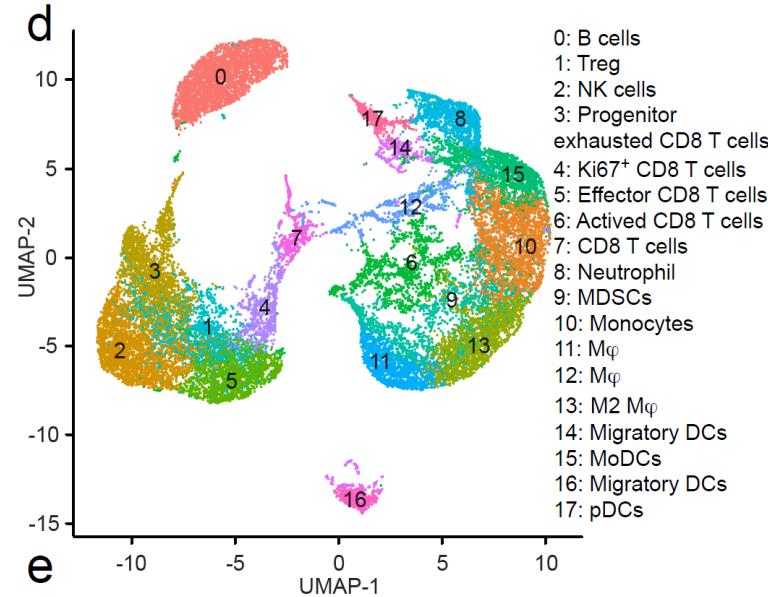
Use graph transformer to understand target cells in antitumor immunity

- Finding tumor cell targets through regulatory networks also helps you identify mechanisms of resistance.
- DSLL (lymph node biopsy) scRNA/ATAC-Seq in a single patient.
- DeepMAPS HGT identified coordinated cell identity and GRN patterns, revealing a developmentally related grouping of two tumor-cell states.
- While TACI (a good tumor target) signaling was present in both B cell states, the second state revealed more TACI-independent survival mechanisms.
- TACI presents as an encouraging target, but single cell GRN inference reveals a path to antigen escape in tumor cells.

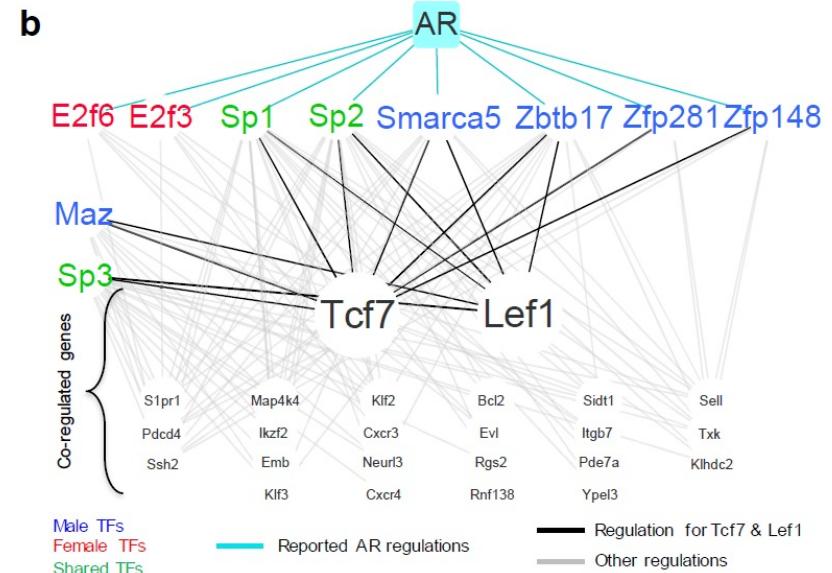


Use graph transformer to understand effector cells in antitumor immunity

Antitumor immunity signature gene identification in Susd2^{-/-} CD8+ T cells



Unique sex-biased regulatory mechanisms of CD8+ T cell exhaustion program



 Sushi domain containing 2 suppresses CD8+ T cell antitumor immunity by targeting IL-2 receptor signaling.
Nature Immunology. 2022

Androgen conspires with the CD8+ T cell exhaustion program and contributes to sex bias in cancer.
Science Immunology. 2022

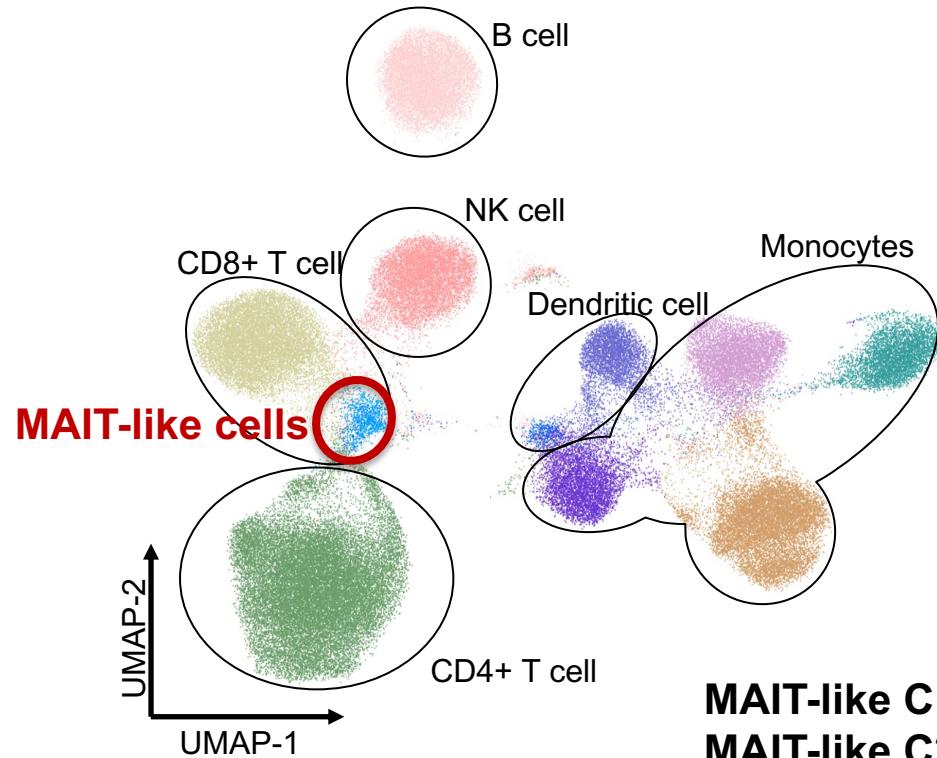


NCI-R01: Targeting immune inhibitory molecule SUSD2 to reverse immunosuppression. (2021-2026)

NCI-R01: Targeting GRP94-TGF-beta pathway for cancer immunotherapy. (2021-2026)

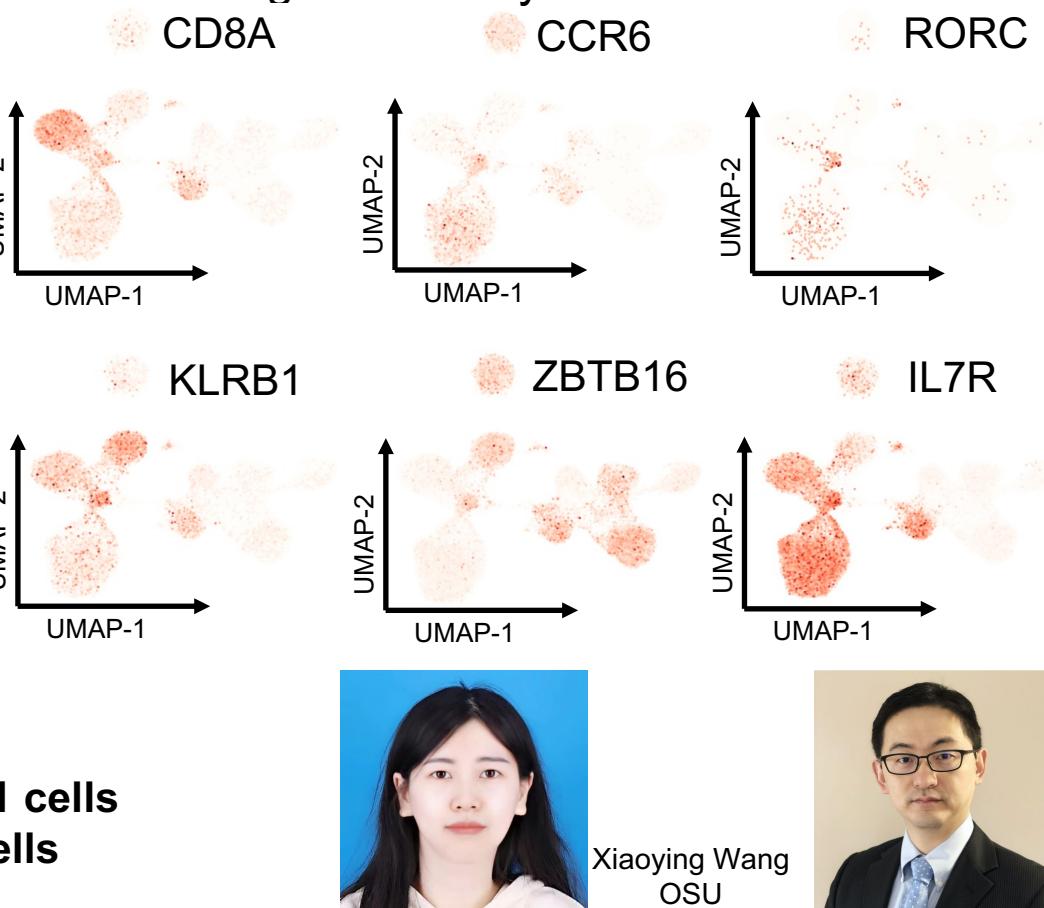
Use graph transformer to infer the mechanism for different survival of PD1-blocking immunotherapy

- In human melanoma data, it identified a rare MAIT-like population impacted by a high IFN-I response and revealed the mechanism of immunotherapy.
- MarsGT offers biological insights and suggests potential strategies for early detection and therapeutic intervention of disease.

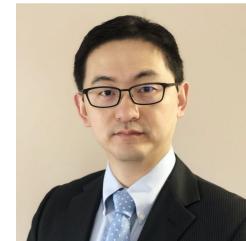


Unpublished results

MAIT-like C1: 1,071 cells
MAIT-like C2: 67 cells



Xiaoying Wang
OSU

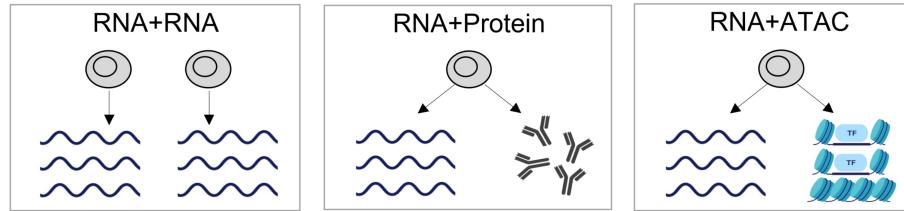


Dr. Gang Xin
OSU

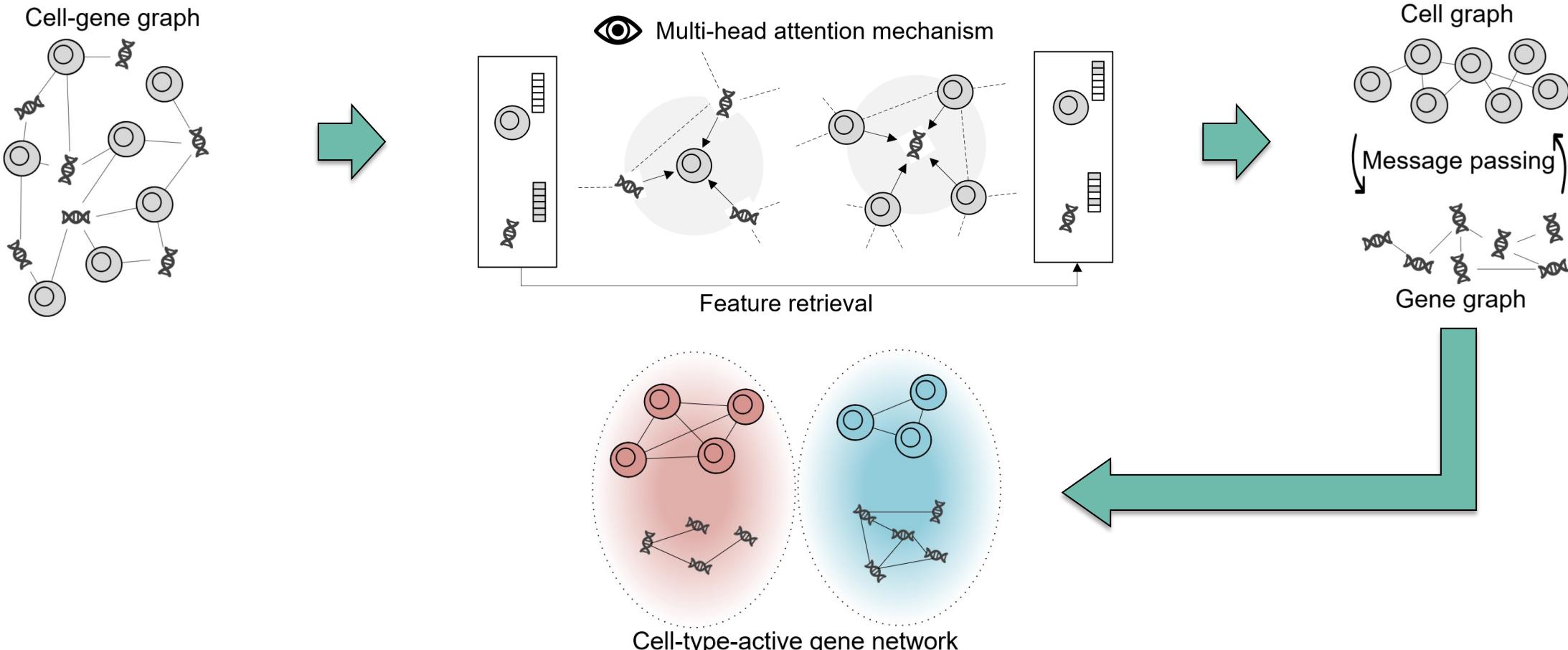
- Melanoma patients: scRNA-seq and scATAC-seq: ~100,000 cells
- 4 high IFN-I response samples, 4 low IFN-I response samples, 2 normal controls

Heterogenous graph transformer in modeling heterogeneous graph

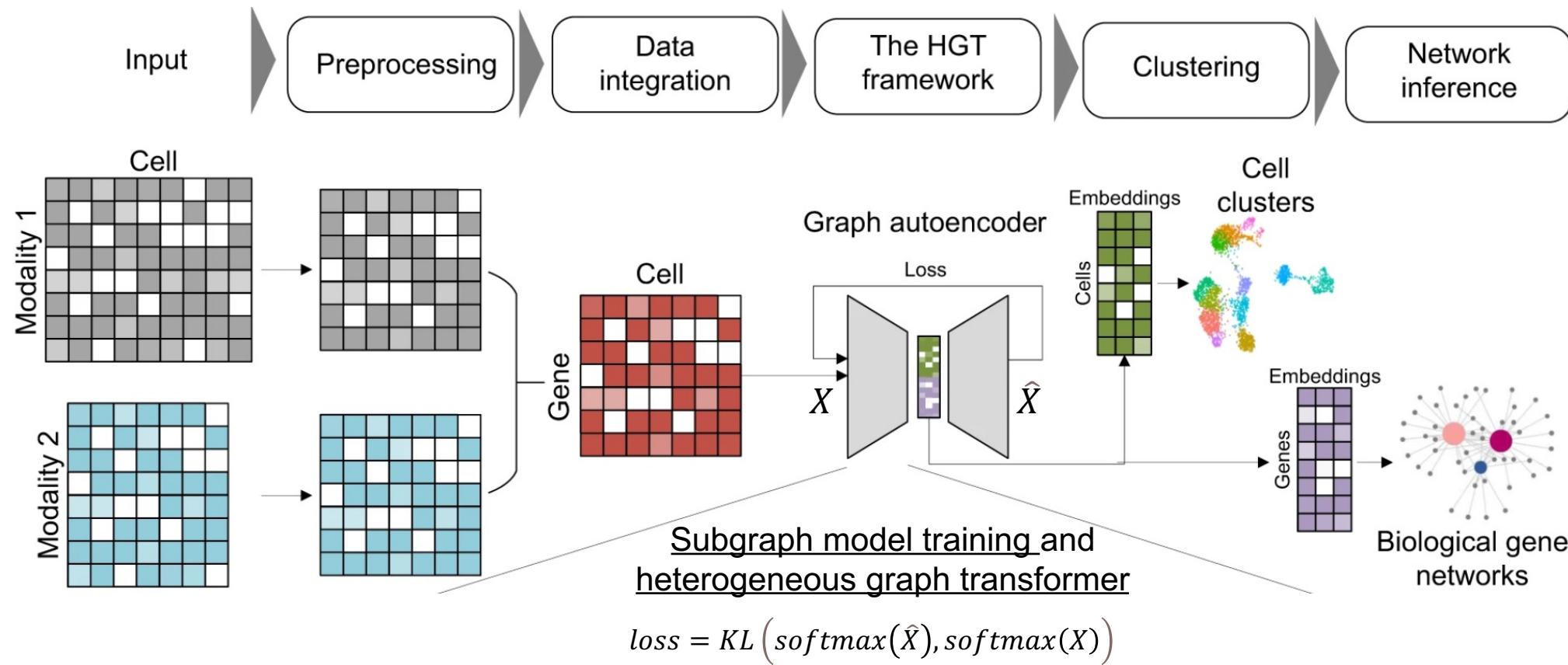
Overview of DeepMAPS framework



- Learn the **joint embedding** of cells and genes
- Calculate “**attention**” scores (importance) b/w cells and genes
- Learn cell clusters and gene clusters simultaneously



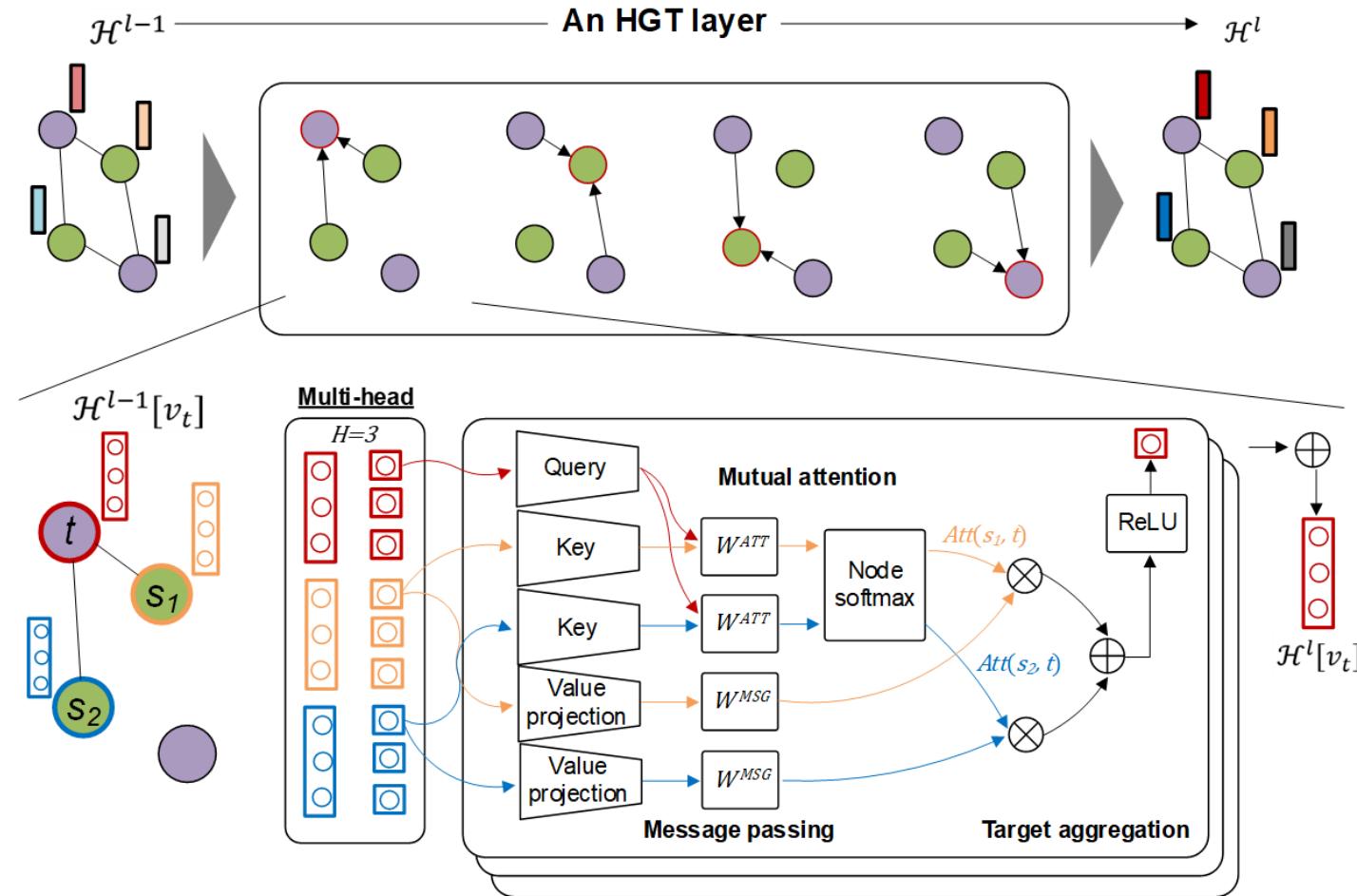
Overview of DeepMAPS framework



Note: ATAC peaks are annotated to corresponding genes based on distance-based regulatory potential score introduced in MAESTRO.

$$w_{ik} = \begin{cases} 0, & d_{ik} > 150\text{kb} \text{ or peak } k \text{ located in any nearby genes} \\ \frac{1}{Length(exon)}, & \text{peak } k \text{ located at the exon regions of the gene } j \\ 2^{-\frac{d_{ik}}{d_0}}, & \text{else} \end{cases}$$

Detailed architecture of heterogeneous graph transformer

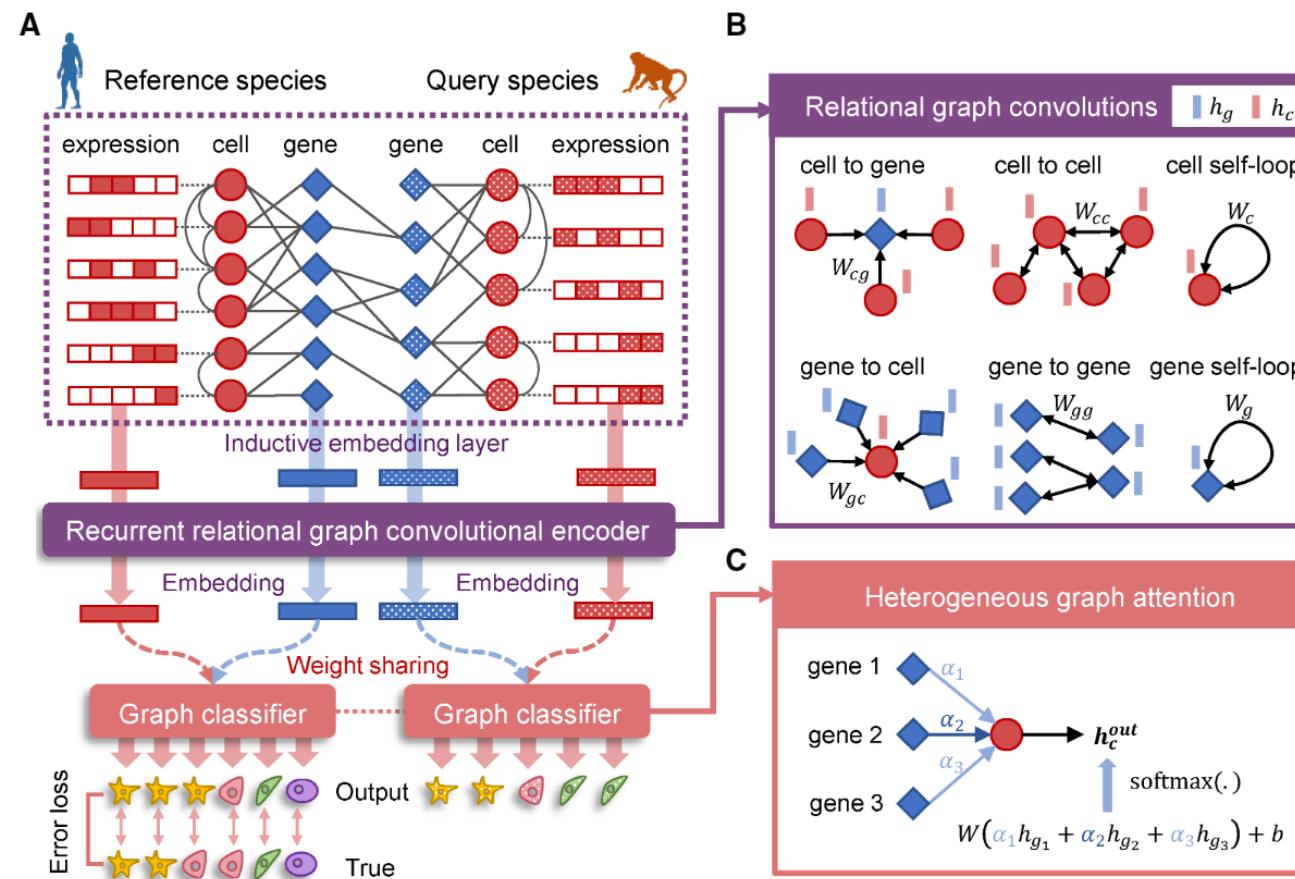


Advantages of heterogeneous graph transformer with multi-head mechanism:

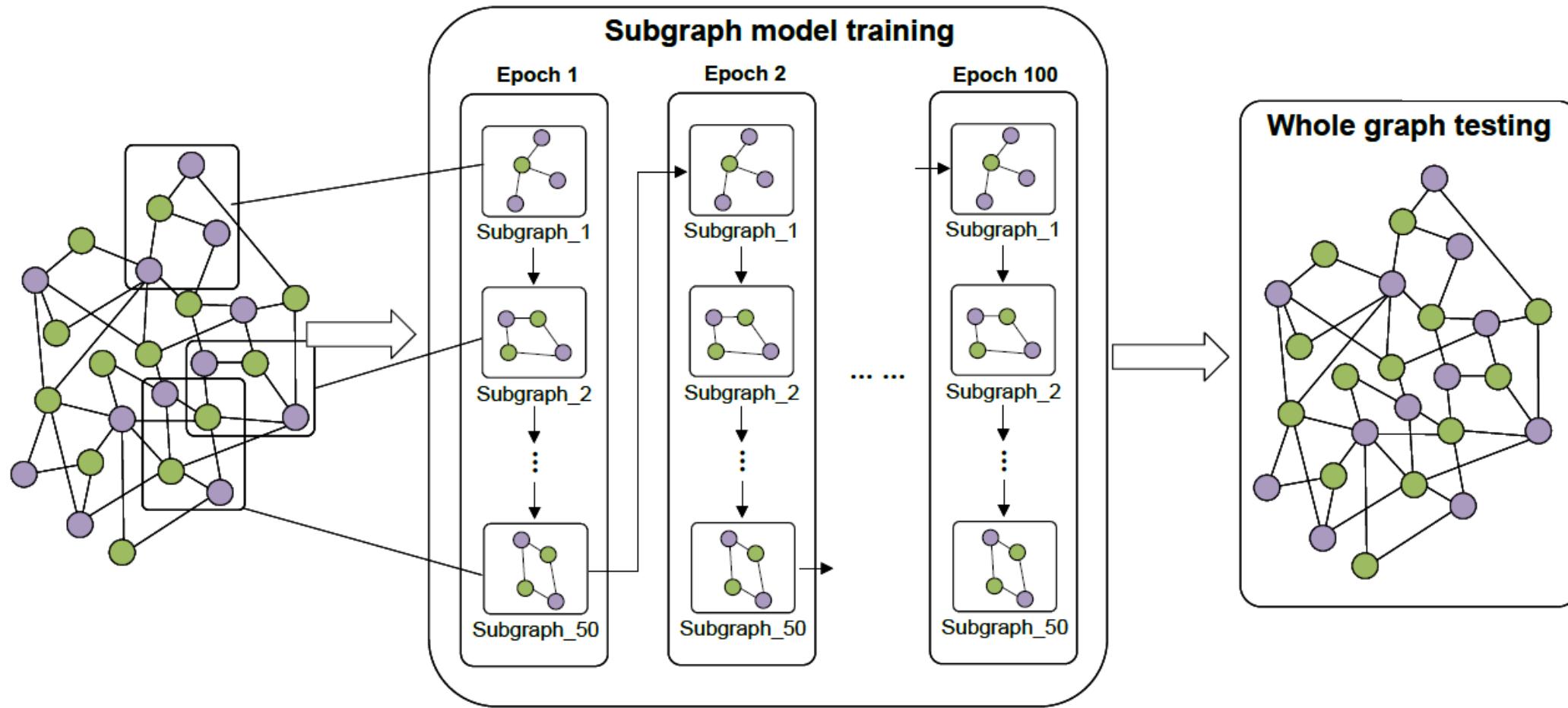
- (1) Anti-noise, (2) increase discrepancies and similarity, (3) computing efficient, (4) interpretable

Cross-species cell-type assignment from single-cell RNA-seq data by a heterogeneous graph neural network

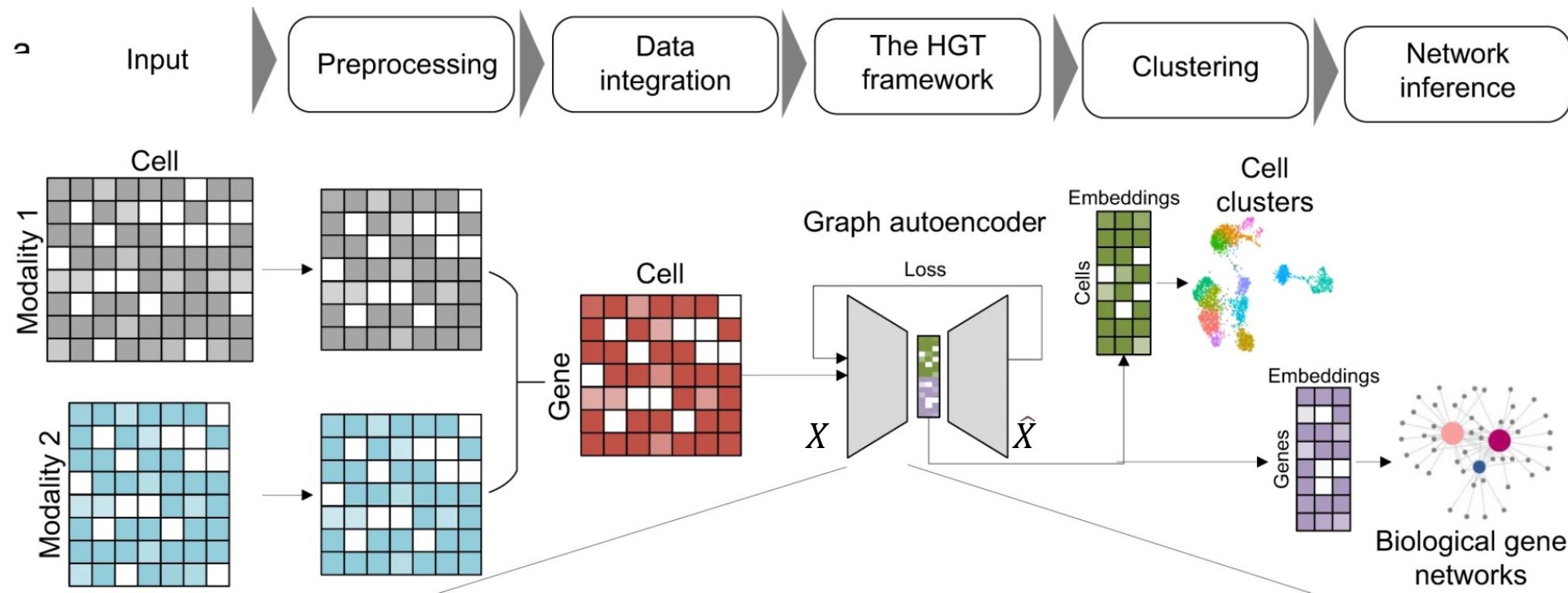
Xingyan Liu,^{1,2,5} Qunlun Shen,^{1,2,5} and Shihua Zhang^{1,2,3,4}



Subgraph training for large dataset



Overview of DeepMAPS framework



We used a Steiner Forest Problem (SFP) model to select genes that highly contribute to cell cluster characterization and construct cell cluster-active gene association networks.

For GRN construction:
design a regulatory intensity (RI) score

$$s_{ij|q} = \sum_k b_{qk}^A \cdot r_{ik|j}$$