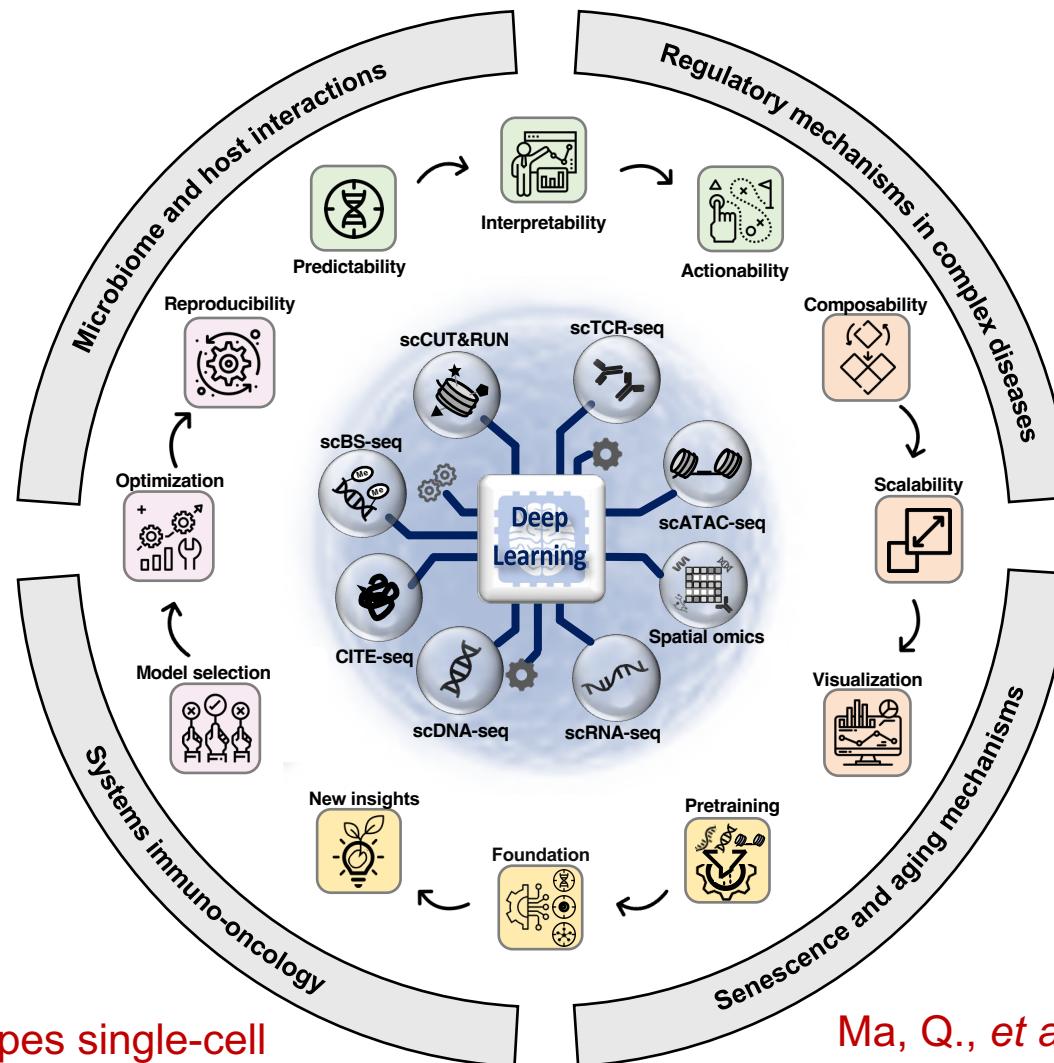


Single-cell multi-omics dataset acquisition, model training, and analysis

Yi Jiang
Graduate Research Assistant
Department of Biomedical Informatics
The Ohio State University
7/12/24

**Dataset acquisition, including methods that convert
the data to the required format of DeepMAPS**

Develop DL methods for new hypothesis generation



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)

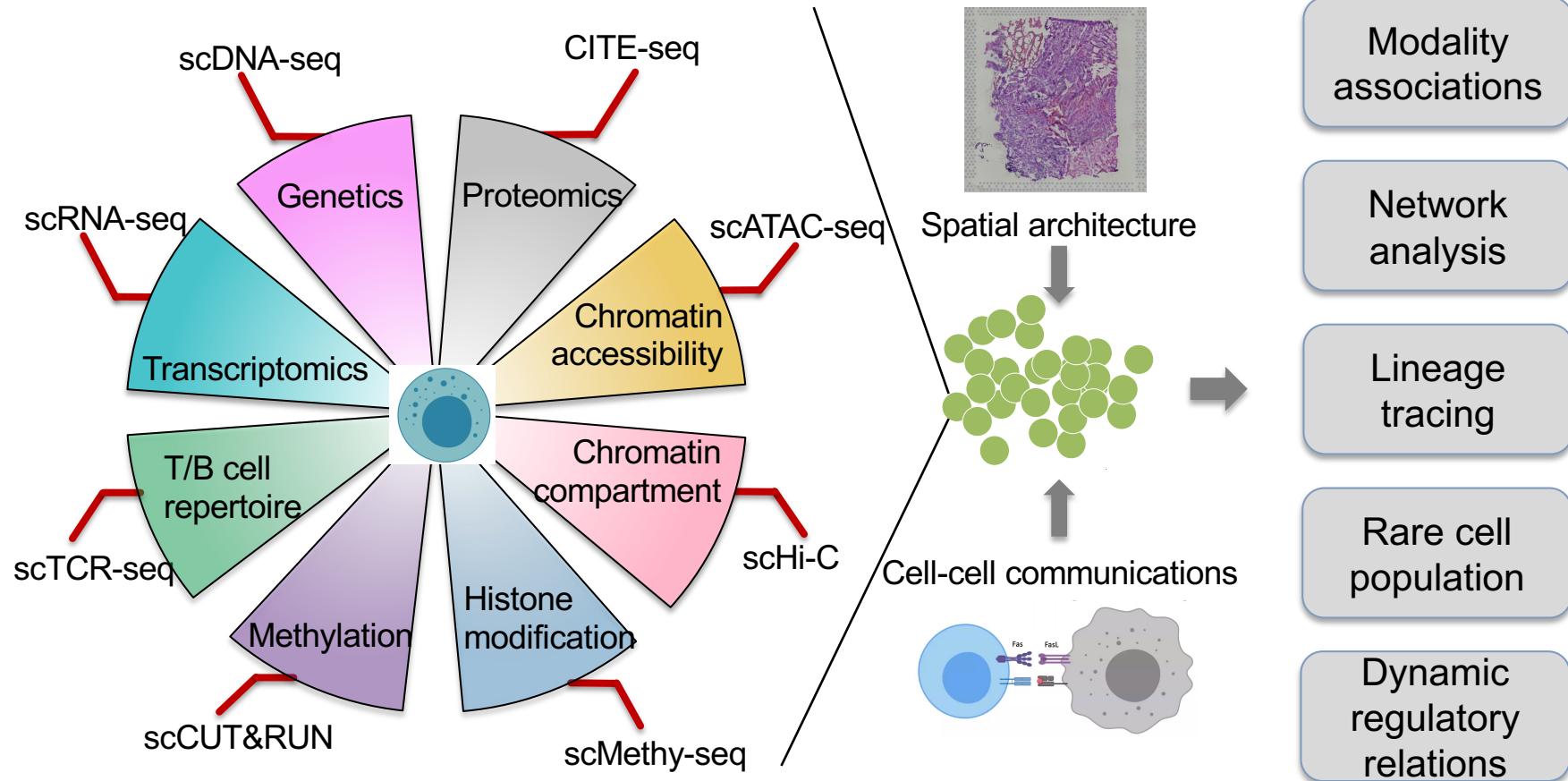


Dr. Qin Ma
Professor, OSU

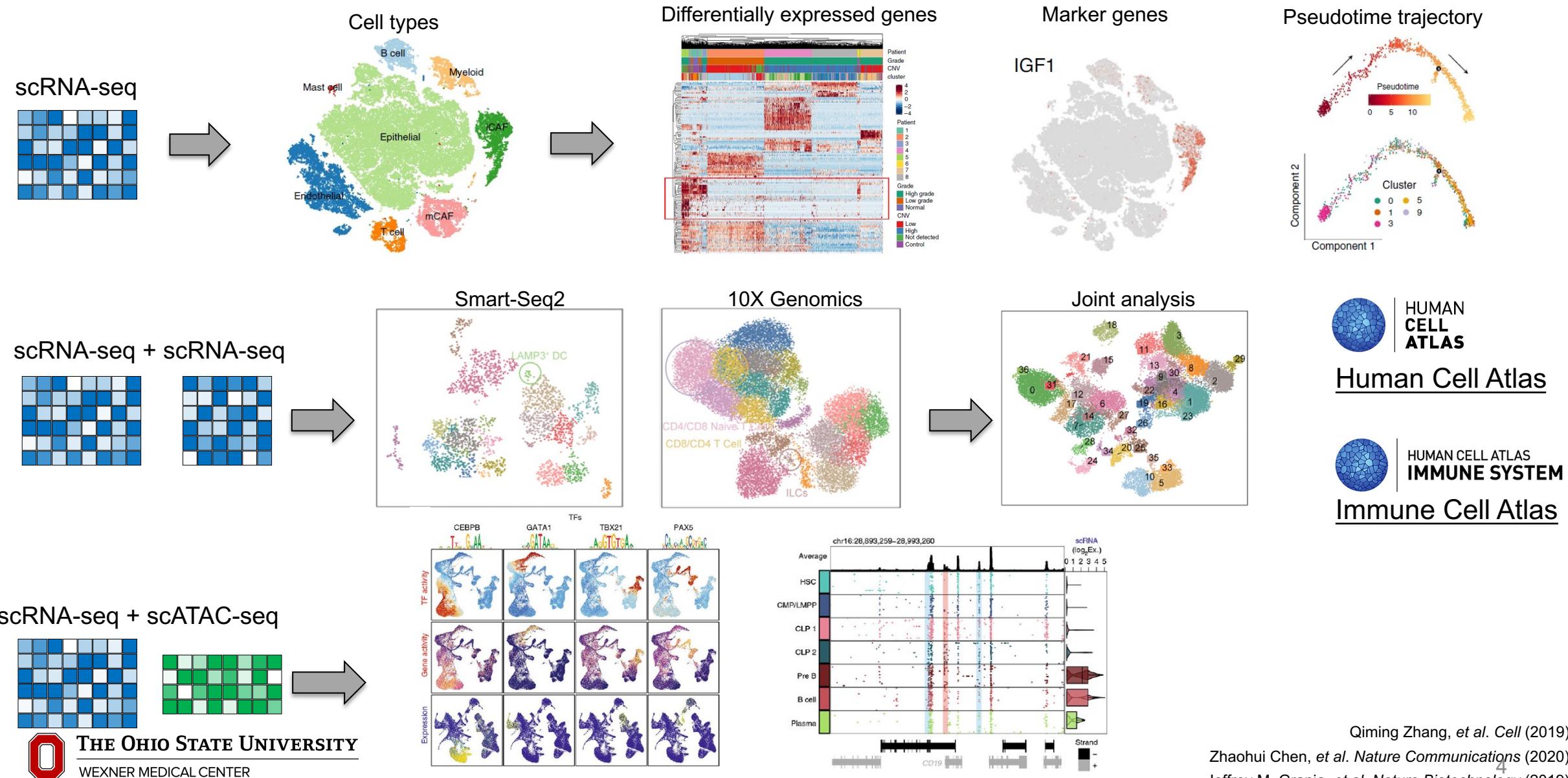


Dr. Dong Xu
Professor, MU

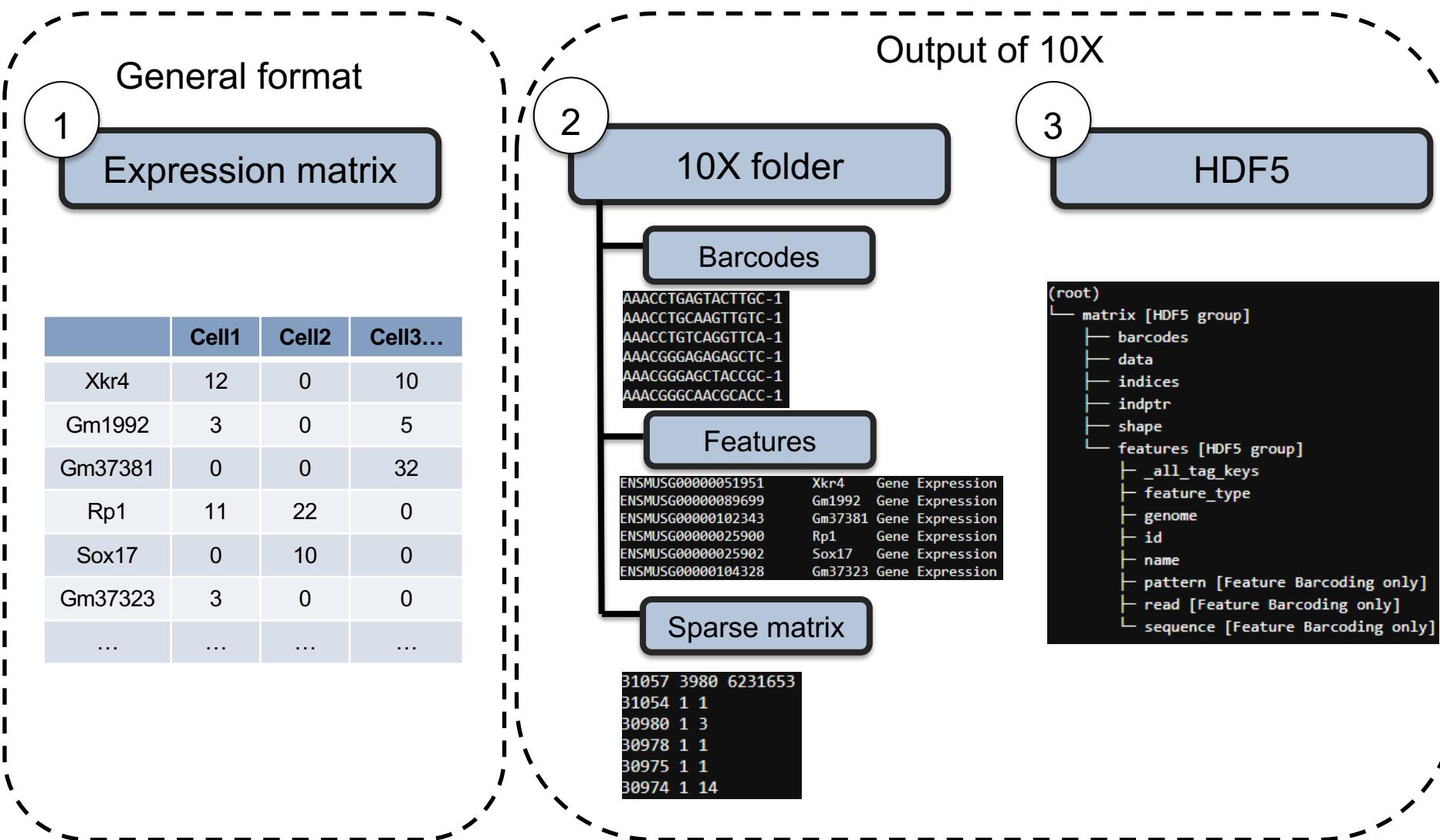
scMulti-omics enhance and enable various biological analyses



scMulti-omics analysis refines knowledge of cell heterogeneity



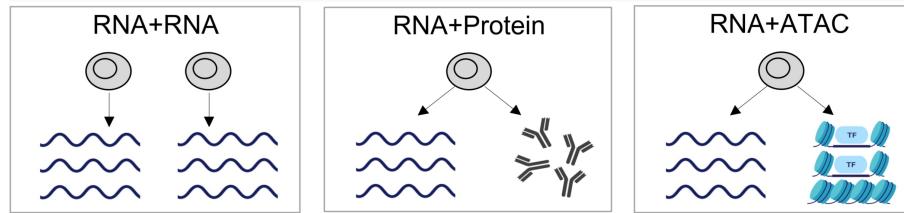
Single-cell common data format



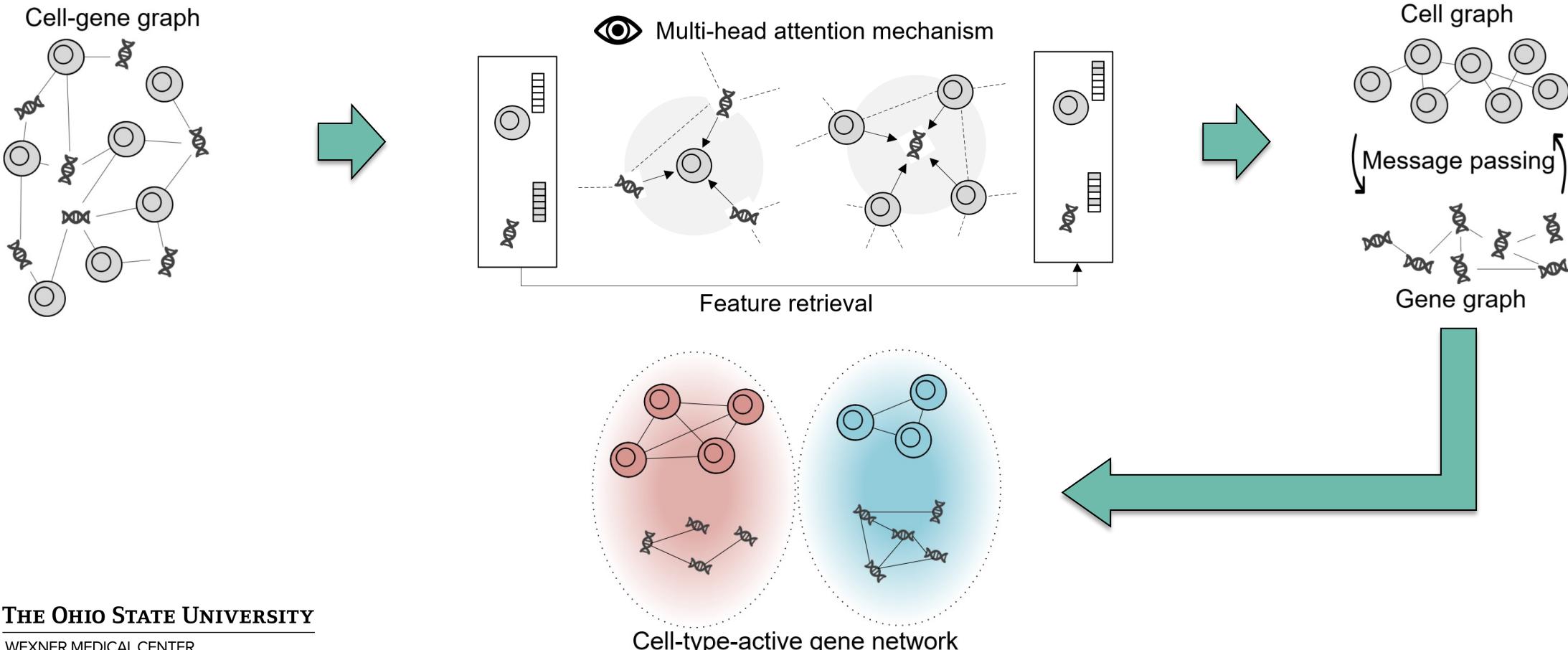
HDF5 detail can be found at: https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/output/molecule_info

Modeling, including deep learning models and automated hyperparameter tuning

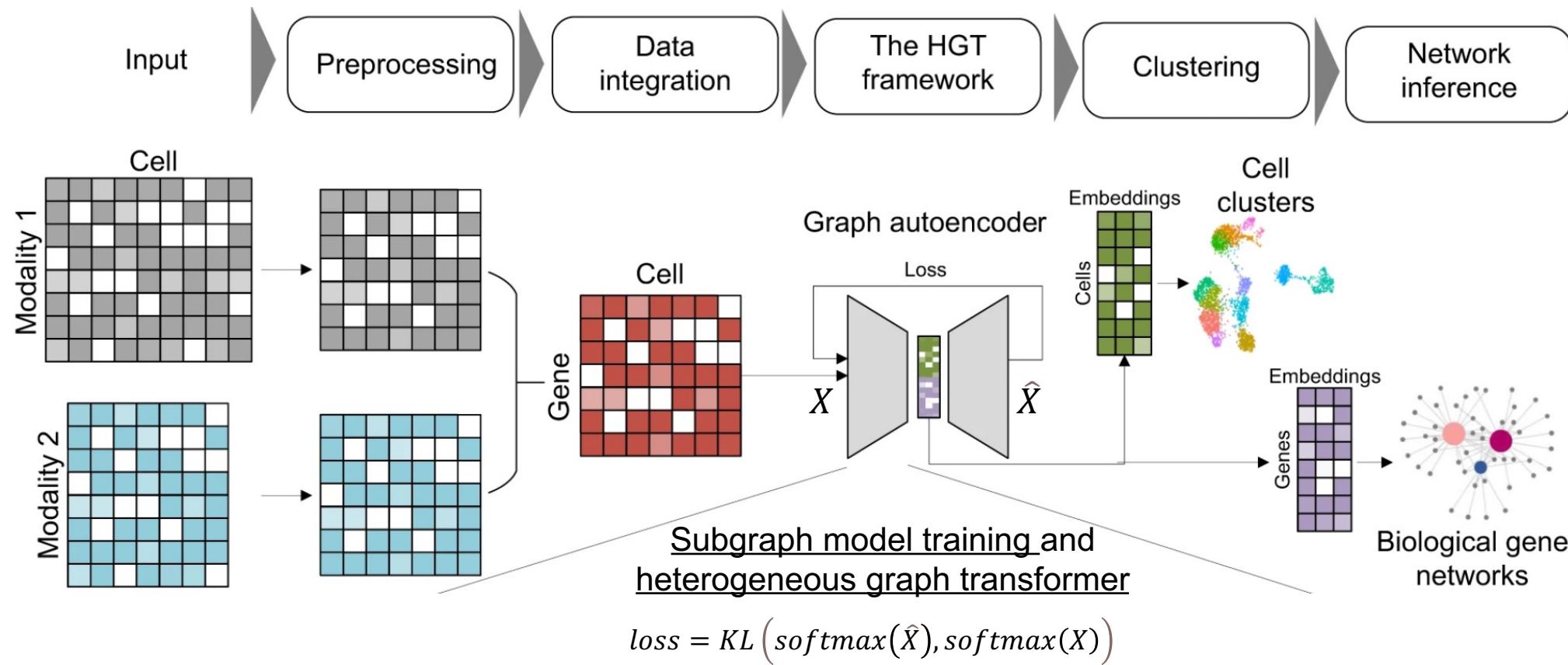
Overview of DeepMAPS framework



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



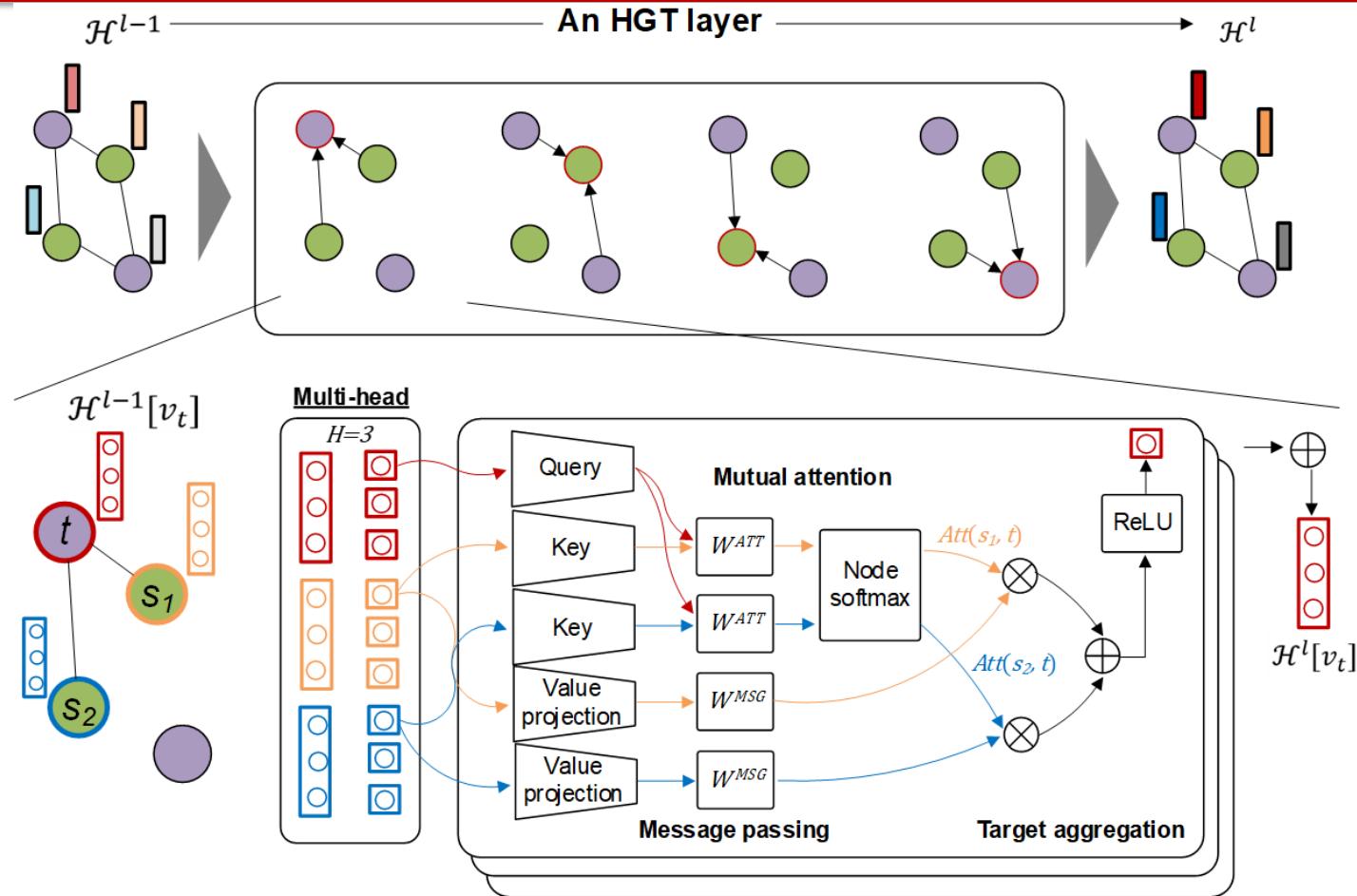
Overview of DeepMAPS framework



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

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

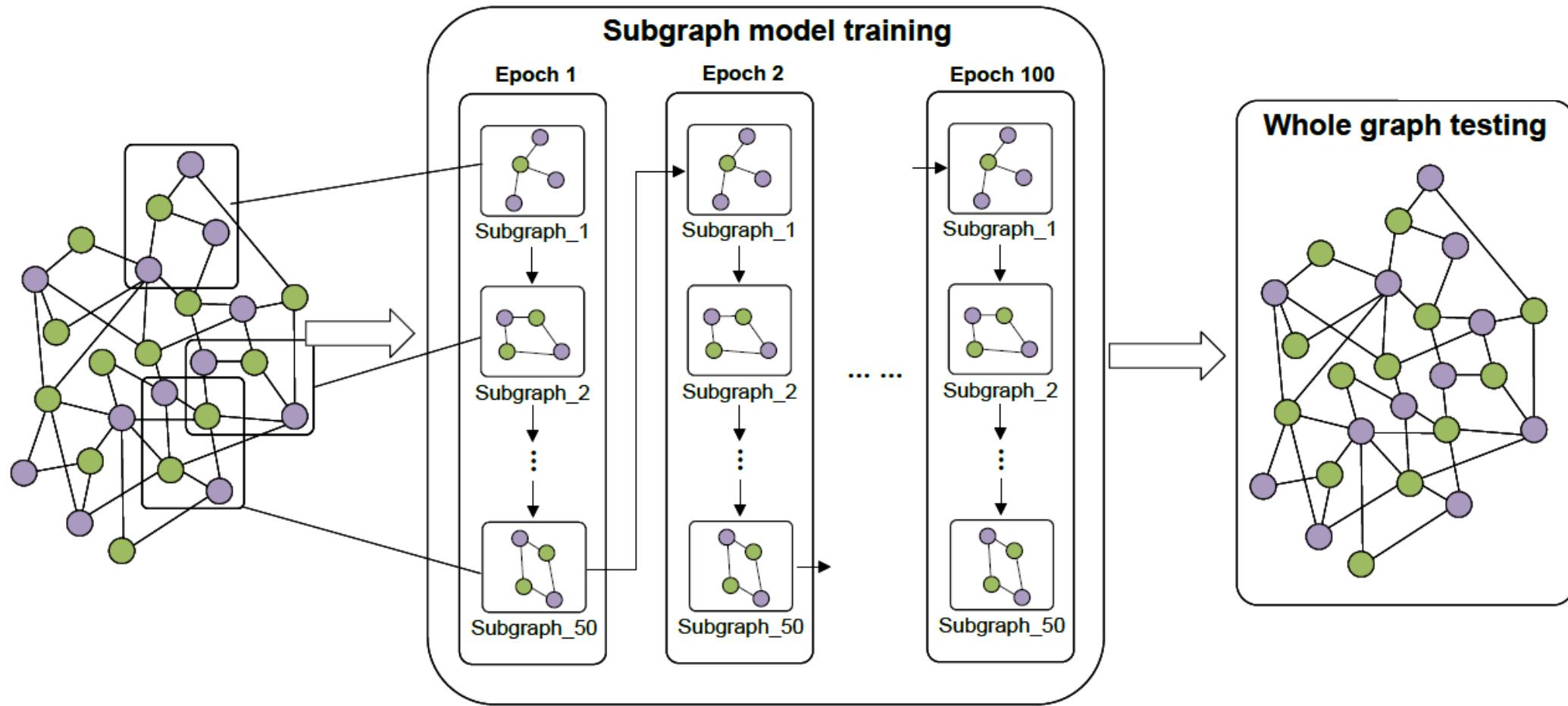
Detailed architecture of heterogeneous graph transformer



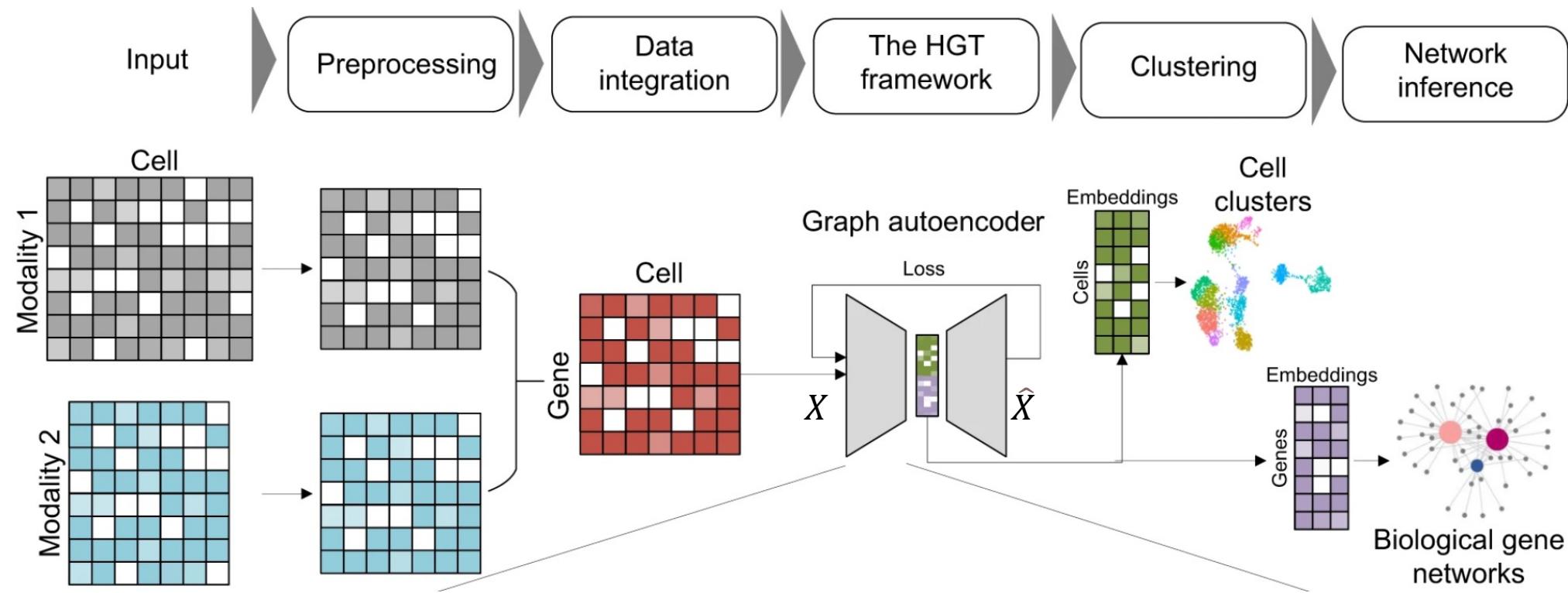
Advantages of heterogeneous graph transformer with multi-head mechanism:

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

Subgraph training for large dataset



Overview of DeepMAPS framework

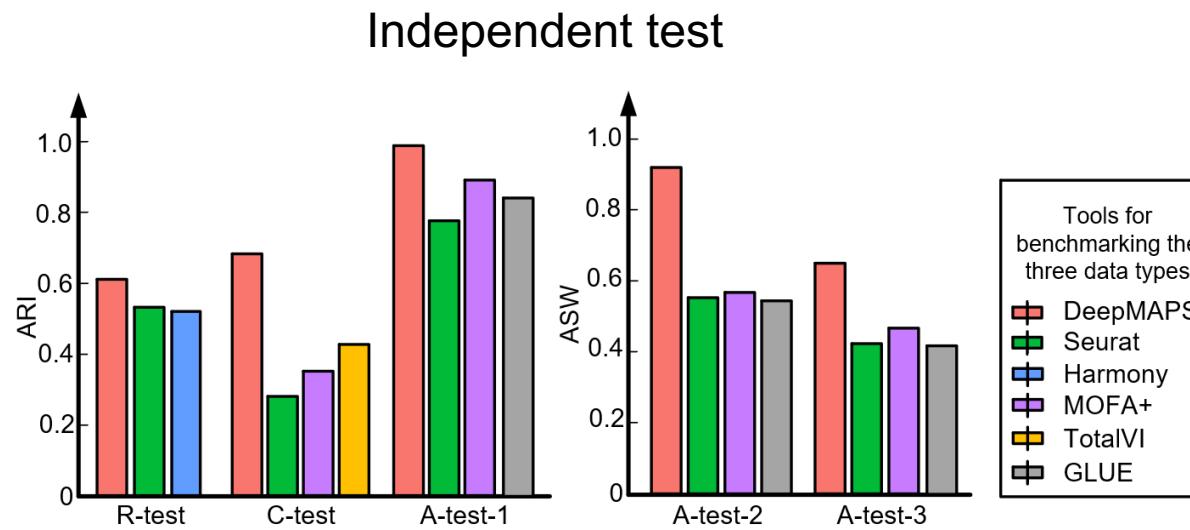
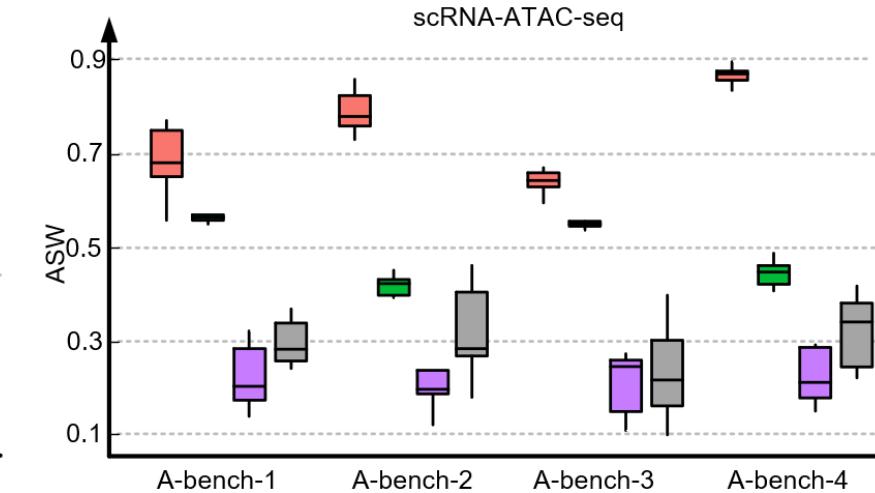
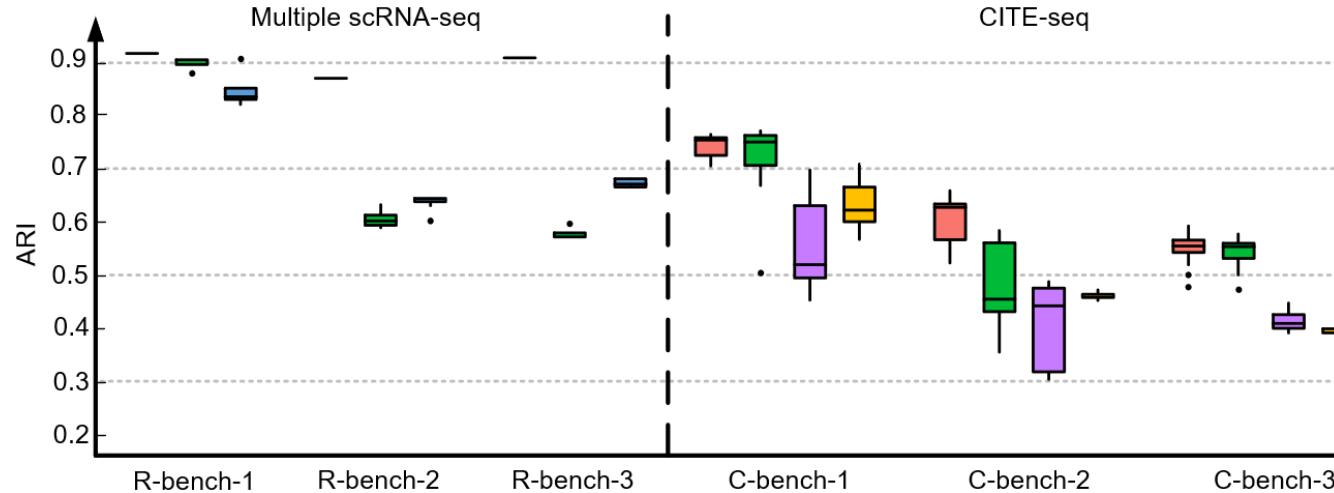


We used a Steiner Forest Problem (SFP) model to select genes that highly contribute to cell cluster characterization and construct cell cluster-active gene association networks.

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

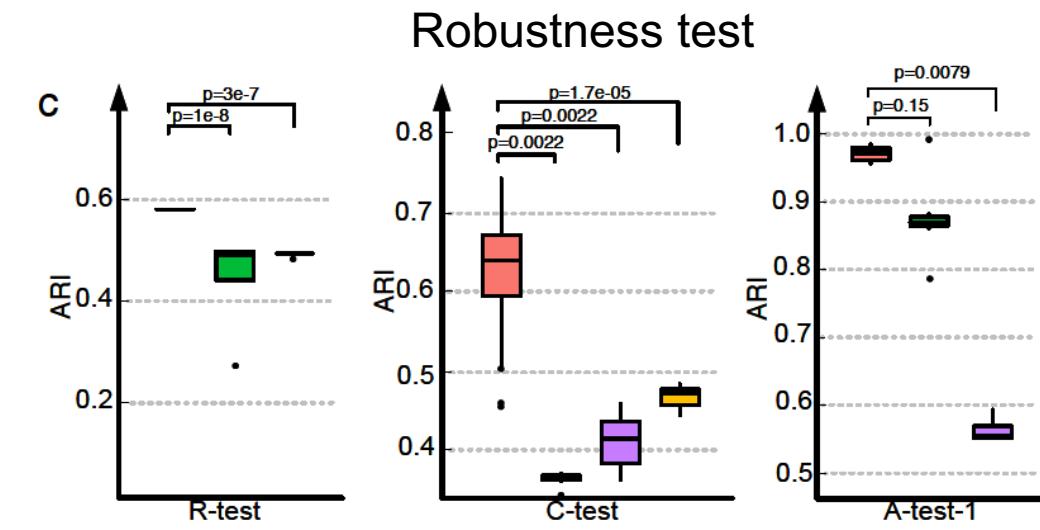
$$s_{ij|q} = \sum_k b_{qk}^A \cdot r_{ik|j}$$

DeepMAPS benchmarking: better performance in cell clustering

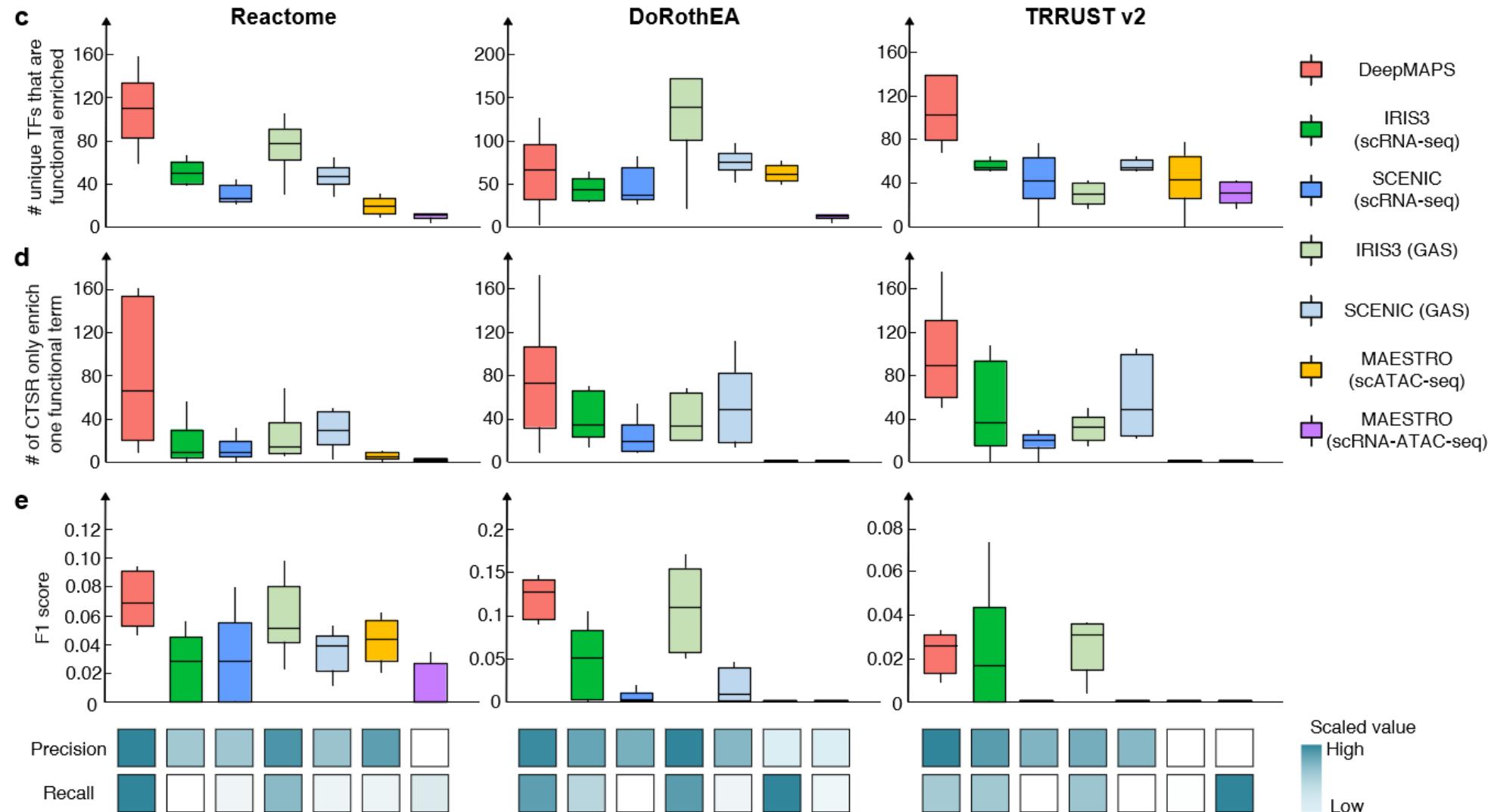


Tools for benchmarking the three data types

- DeepMAPS
- Seurat
- Harmony
- MOFA+
- TotalVI
- GLUE



DeepMAPS benchmarking: better performance in GRN inference



DeepMAPS Tutorial

Tutorial Objectives

- Get to know the DeepMAPS website: main page
- Start a new job
- Access example data
- Result tour for
 - matched scRNA-seq and scATAC-seq data
 - CITE-seq data
 - multiple scRNA-seq data
- User workspace
- Technical supports

Main page navigation

New project

Browse public projects

DeepMAPS EXPLORE TUTORIAL NEWS ABOUT

We are working intensively to upgrade DeepMAPS (0.22.0) workflow. Please let us know if there are any bugs, issues, or suggestions via qin.ma@osumc.edu.

Search project ID

LOGIN

Welcome to DeepMAPS

Deep learning based Multi-omics Analysis Platform for Single cells

New project

Create a project from your data

View Examples

Browse showcase datasets

View workspace

Manage data and results

Start a new project or load example data

View example results

Access users' workspace. All your previous jobs will show up here. Login required.

Cite us: Ma, A. et al. (2021) Biological network inference from single-cell multi-omics data using heterogeneous graph transformer. bioRxiv. ↗

1.1 Matched scRNA-seq and scATAC-seq data submission

SCRNA-SEQ SCRNA-SEQ AND SCATAC-SEQ CITE-SEQ

Data upload

Feature barcode matrix
Support format: *.h5, *.hdf5

scATAC-seq fragments (optional)
Support format: *fragments.tsv.gz

scATAC-seq fragments index (optional)
Support format: *fragments.tsv.gz.tbi

Metadata (optional)
Support format: *.csv, *.txt

RNA velocity matrix (optional)
Support format: *.csv, *.txt

EXAMPLE

Project information

Project title

Species

Project description (optional)

E-mail

Optional: The running time could take a few hours if the dataset is large. Hence, we strongly recommend you to leave your email, and you will be notified by email when the job is done.

Used for scATAC-seq peak annotation

A h5 or hdf5 file. (Required)

User defined cell meta-information (optional)

An RNA velocity matrix. Can be generated via CellRank using scRNA-seq fastq file. Without this file, cell clustering and only performed via Seurat. (optional)

Load example (10X Human lymphoma 14k cells)

Download data: Human lymphoma 14k cells (h5)

Download data: Human lymphoma 14k cells (velocity matrix)

Download data: Human lymphoma 14k cells (fragments)

Download data: Human lymphoma 14k cells (fragments index)

START

For more information, please check:
<https://www.10xgenomics.com/resources/datasets/fresh-frozen-lymph-node-with-b-cell-lymphoma-14-k-sorted-nuclei-1-standard-2-0-0>

1.1 Matched scRNA-seq and scATAC-seq data submission

SCRNA-SEQ SCRNA-SEQ AND SCATAC-SEQ CITE-SEQ

Data upload

- Feature barcode matrix
Support format: *.h5, *.hdf5
0 files
- scATAC-seq fragments (optional)
Support format: *fragments.tsv.gz
0 files
- scATAC-seq fragments index (optional)
Support format: *fragments.tsv.gz.tbi
0 files
- Metadata (optional)
Support format: *.csv, *.txt
0 files
- RNA velocity matrix (optional)
Support format: *.csv, *.txt
0 files

EXAMPLE

Project information

Project title

Species

Project description (optional)

0 / 100

Now hours if the dataset is large. Hence, we strongly recommend you to leave your email, and you will be

Once everything is ready, let's get started!

Give a name to your project

Please select the right sample species. (required)

Take some notes for the project. E.g., library preparation, related paper, experiments, etc. (optional)

We recommend you leave your email. Some steps may take a while. A notice will send to you once the job is done.

1.2 CITE-seq data submission

SCRNA-SEQ SCRNA-SEQ AND SCATAC-SEQ CITE-SEQ

Data upload

Feature barcode matrix
Support format: *.h5, *.hdf5 0 files

Metadata (optional)
Support format: *.csv, *.txt 0 files

EXAMPLE ← Load example (CITE-seq PBMC & Lung 3.8k cells)
Download data: (CITE-seq PBMC & Lung)

Project information

Project title
Species

Project description (optional)

E-mail

Optional: The running time could takes a few hours if the dataset is large. Hence, we strongly recommend you to leave your email, and you will be notified by email when the job is done.

START

1.3 Multiple scRNA-seq data submission

SCRNA-SEQ SCRNA-SEQ AND SCATAC-SEQ CITE-SEQ

Data upload

Gene expression matrix 1
Support format: *.csv, *.txt, *.h5 0 files

Gene expression matrix 2
Support format: *.csv, *.txt, *.h5 0 files

ADD A ROW **REMOVE A ROW**

Metadata (optional)
Support format: *.csv, *.txt 0 files

EXAMPLE

Project information

Project title

Species

Project description (optional)

E-mail

Optional: The running time could takes a few hours if the dataset is large. Hence, we strongly recommend you to leave your email, and you will be notified by email when the job is done.

START

Add 2 or more scRNA-seq files

For more datasets, use add a row button.

Load example (Human IFNB PBMCs, 2800 cells)

Download gene expression matrix: Human IFNB (sample 1)

Download gene expression matrix: Human IFNB (sample 2)

Download metadata file: Human IFNB

1.4 Waiting for uploading

E-mail

Optional: The running time could takes a few hours if the dataset is large. Hence, we strongly recommend you to leave your email, and you will be notified by email when the job is done.

START



SINGLE SCRNA-SEQ MULTIPLE SCRNA-SEQ SCRNA-SEQ AND SCATAC-SEQ CITE-SEQ

Data upload

Feature barcode matrix
lymph_node_matrix.h5

Support format: *.h5, *.hdf5
scATAC-seq fragments (optional)
lymph_node_nts.tsv.gz

Support format: *fragments.tsv.gz
scATAC-seq fragments index (optional)
lymph_node.tsv.gz.tbi

Support format: *fragments.tsv.gz.tbi

Metadata (optional)

Support format: *.csv, *.tsv
RNA velocity matrix (optional)
lymph_node_riv.csv.gz

Support format: *.csv, *.tsv

EXAMPLE

Project information

Project title
test

Species
Human

Project description (optional)
testtesttesttesttesttest

E-mail
anjun.ma@osumc.edu

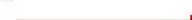
Optional: The running time could takes a few hours if the dataset is large. Hence, we strongly recommend you to leave your email, and you will be notified by email when the job is done.

START

Loading .

CANCEL

Upload progress: 1%



Loading

CANCEL

Upload progress: 2%

Once the uploading process is complete, you will go to the “Quality Control” page. If you leave your email, you will be notified.

2. How to get access to the example result

View Examples

Browse showcase datasets

→

Public datasets

Filters:

Species
Human

Results:

Human IFNB-Stimulated and Control PBMCs

Type Multiple scRNA-seq dataset
Species: Human
Tissue: blood
Number of cells: 2741
Update on: 2022-01-16
Creator: admin

BROWSE

Human PBMC and lung tumor leukocytes

Type CITE-seq
Species: Human
Tissue: PBMC & Lung
Number of cells: 3485
Update on: 2022-01-16
Creator: admin

BROWSE

Fresh Frozen Lymph Node with B-cell Lymphoma (14k sorted nuclei)

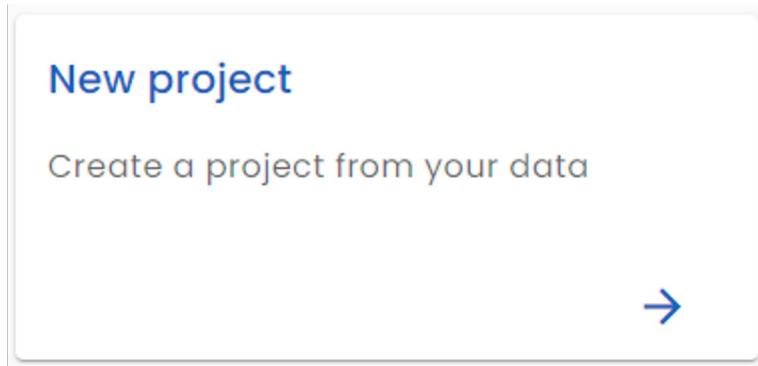
Type scRNA-seq and scATAC-seq
Species: Human
Tissue: lymph
Number of cells: 13510
Update on: 2022-01-16
Creator: admin

BROWSE

You have two ways to navigate our example results

The first one is to click the “View Example” button on the main page, and choose one of your interest.

2. How to get access to the example result



The second one is to in the new project, you can load the example data and start.

SCRNA-SEQ SCRNA-SEQ AND SCATAC-SEQ CITE-SEQ

Data upload

Feature barcode matrix
Support format: *.h5, *.hdf5 0 files

scATAC-seq fragments (optional)
Support format: *fragments.tsv.gz 0 files

scATAC-seq fragments index (optional)
Support format: *fragments.tsv.gz.tbi 0 files

Metadata (optional)
Support format: *.csv, *.txt 0 files

RNA velocity matrix (optional)
Support format: *.csv, *.txt 0 files

Load example (10X Human lymphoma 14k cells)

[Download data: Human lymphoma 14k cells \(h5\)](#)

[Download data: Human lymphoma 14k cells \(velocity matrix\)](#)

[Download data: Human lymphoma 14k cells \(fragments\)](#)

[Download data: Human lymphoma 14k cells \(fragments index\)](#)

Project description (optional)

E-mail

Optional: The running time could takes a few hours if the dataset is large. Hence, we strongly recommend you to leave your email, and you will be notified by email when the job is done.

START

3. Example data illustration

Next, we will go through the result pages using the matched scRNA-seq and scATAC-seq example data.

Three steps are included:

- 3.1. data quality control,
- 3.2. data integration and cell clustering,
- 3.3. network construction

DeepMAPS provides real-time parameter changing and result observations.

3.1 Quality control

Project summary
Project information
ID: lymphoma_14k
Create time: 2022-01-15
Creator: Admin
Species: Human
Uploaded files: ["lymph_node_lymphoma_14k_filtered_feature_bc_matrix.h5"]

Expand to see project information

Select data for preprocessing

Original and preprocessed data statistics

Choose parameters

Move mouse to one figure will show figure illustration and download options

Download PNG
Download JPG
Download file (CSV)

1 Quality control
Project summary

2 Cell clustering

Step indicator
3 Network construction

SCRNA-SEQ ASSAY SCATAC-SEQ ASSAY

Original data statistics
Cells: 8770
Genes: 21477
Average expression level: 0.1172
Zero expression percentage: 0.0595

Current data statistics
Cells: 8770
Genes: 21477
Average expression level: 0.1172
Zero expression percentage: 0.0595

Preprocessing
Max counts per cell: 20000
Min cells per gene: 0.001
Mitochondrial counts ratio: 0.1
Ribosomal counts ratio: 0.5
 Remove outlier cells
Normalization method: LogNormalize
UNLOCK PARAMETERS
LOAD

Number of total read counts per cell

Number of expressed genes per cell

Number of read counts per cell

Number of cells among each expressed gene

Preprocessing results and general data statistic visualizations

Top variable genes

Gene	Residual variance	Standard deviation	Mean	Max
CEMIP	13.90	4.96	0.04	6.05
FTL	10.84	17.29	0.49	5.99
CCSER1	9.70	4.02	0.17	5.07
HDAC9	9.33	5.37	0.23	5.49
SEMA3A	7.86	0.75	0.04	4.98
SLC8A1	6.79	23.13	0.44	5.67
TFEC	6.74	6.43	0.28	4.80
RIMS1	6.61	2.35	0.18	5.06
CCL4L2	6.45	3.00	0.23	5.34
A1136456.1	6.39	1.59	0.13	5.10

Rows per page: 15 1-15 of 2000

3.2.1 Cell clustering – Data integration and cell clustering

Integrative clustering

Integrative clustering

Method

Velocity weighted method

Clustering resolution

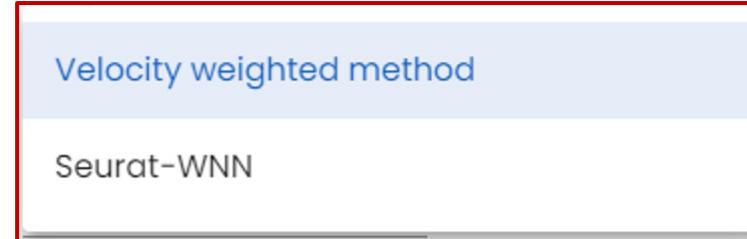
0.5

CALCULATE

Cell identity

Cell labeling

Cell selection



Users can choose to use either the velocity weighted integration method in DeepMAPS or the weighted nearest neighbor (WNN) method in Seurat 4.0

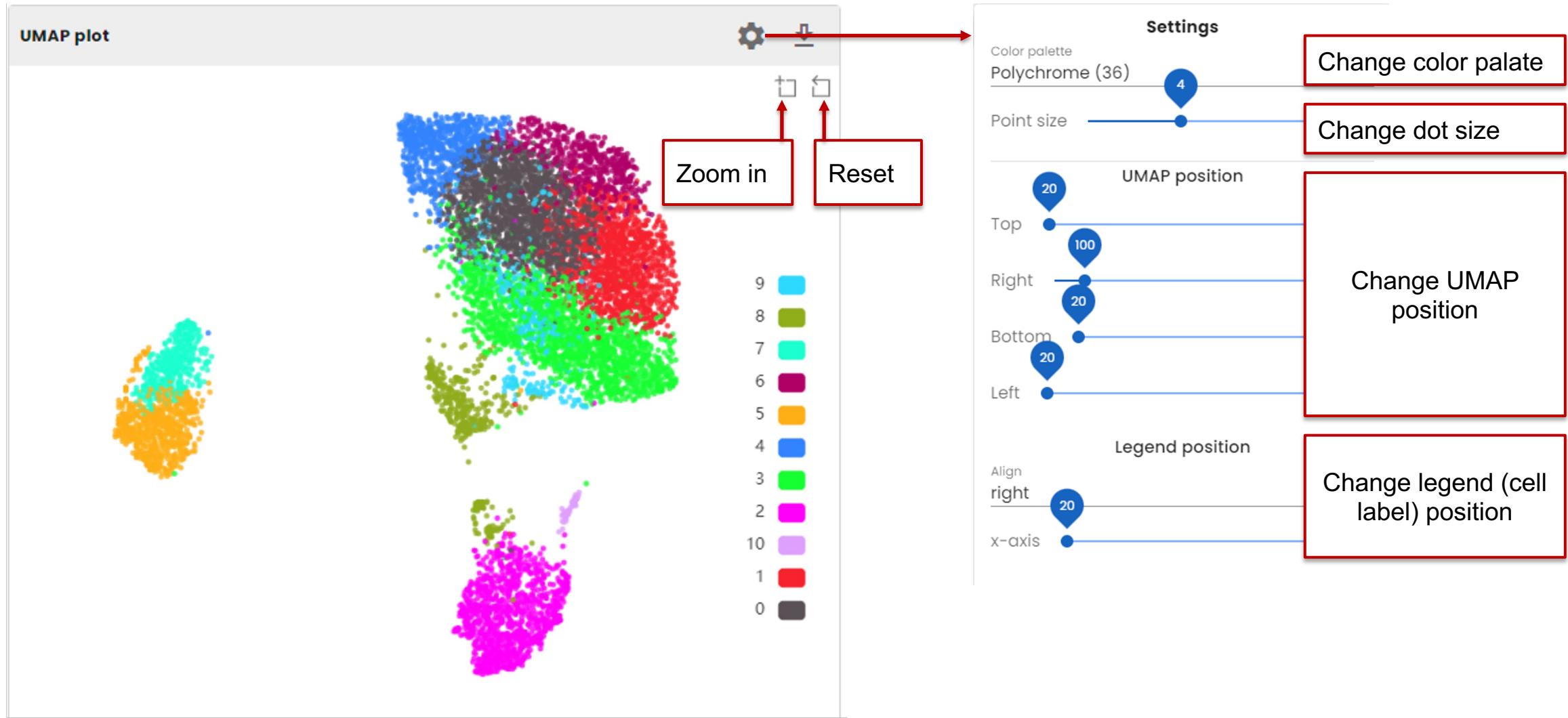
Note that, a velocity matrix is needed as input for using the velocity weighted integration. If you do not provide such velocity data, please use WNN instead.

You will then asked to select number of principle components to be used for data integration (for WNN only), and the clustering resolution in Louvain clustering method (for both)

Click “calculate” button to continue.

An email will be sent to you once the clustering step is done.

3.2.3 Cell clustering – UMAP



3.2.2 Cell clustering – Change active cell identity

If you have uploaded a metadata file, you can switch the active cell identity to observe different cell labels on the UMAP, e.g., treatment, sample id, male/female. By default, the clustering results will be selected.

Integrative clustering

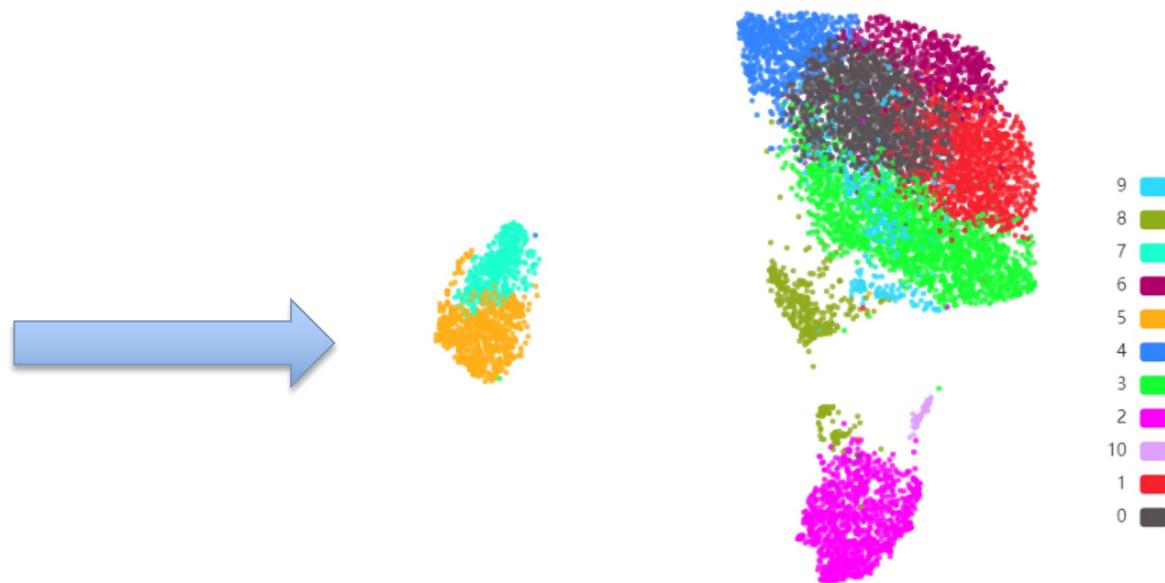
Cell identity

Active cell identity

Select identity

Cell labeling

Cell selection



3.2.4 Cell clustering – Rename existing cell labels

Cell labeling

Rename clusters

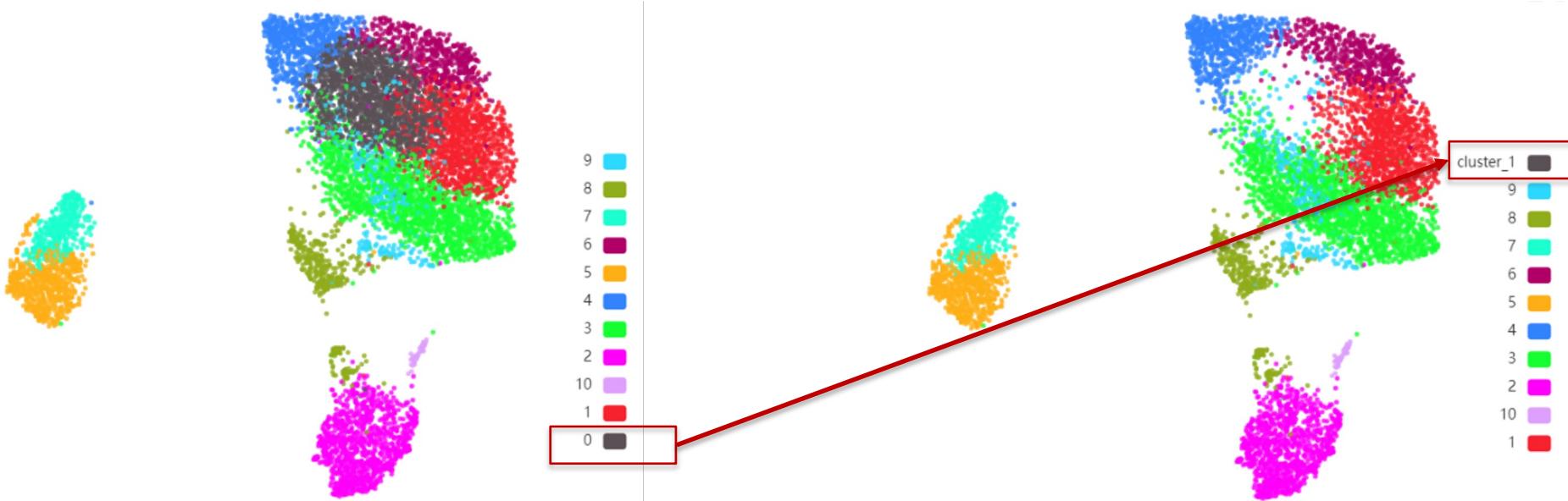
Old cluster name
0

New cluster name
cluster_1

RENAME

Once you decided the name of each predicted cell cluster, you can use the “Rename clusters” function to manually annotate the name. Here as an example, cluster 0 is renamed as “cluster_1”.

Please remember to change to the right active cell identity before rename clusters.



3.2.5 Cell clustering – Cluster merge and create new cell labels

Custom cell labeling

Step 1: Create cell filters

1. Choose a label type → Category

2. Choose clusters you want to merge → Cluster

Add t → ADD CLUSTER FILTER

Gene → ADD GENE FILTER

Direction >

Threshold

Filters applied:

in seurat_clusters: 7

in seurat_clusters: 5

For reselection, remove all rules first → REMOVE ALL

Delete rule if you select it wrong

3. Give a name to your new label group → new_type_2

4. Give a new cell type label → test_type1

ASSIGN CELLS → 5. Confirm and see result

You can merge your cells by different rules (multiple clusters) or specific gene expression. Create a new label group and name cell types in your way!

Cell identity

Active cell identity

Select identity

- seurat_clusters
- cell_type
- new_type
- new_type_2**

The figure illustrates the process of merging cell clusters and creating new cell labels using the Seurat 'Custom cell labeling' tool. The interface is divided into two main steps: Step 1 for creating cell filters and Step 2 for assigning cells to new labels.

Step 1: Create cell filters

This step involves defining filtering rules. A red box highlights the 'Category' dropdown, which is set to 'Cluster'. Another red box highlights the 'Cluster' dropdown, which is set to 'test_type1'. A third red box highlights the 'Add t' button, which has triggered the 'ADD CLUSTER FILTER' button. Below these, a 'Gene' dropdown is shown, along with 'Direction' and 'Threshold' fields. A 'Filters applied:' section shows two rules: 'in seurat_clusters: 7' and 'in seurat_clusters: 5'. A 'REMOVE ALL' button is available for reselection. A red box highlights the 'Delete rule if you select it wrong' button.

Step 2: Assign cells to new label

This step involves assigning cells to a new label group. A red box highlights the 'new_type_2' input field, and another red box highlights the 'test_type1' input field. An 'ASSIGN CELLS' button is at the bottom. A red box highlights the '5. Confirm and see result' button.

Visualizations and Results

To the right of the filter interface, there is a scatter plot showing cell clusters colored by their assigned labels. A legend on the right side of the plot maps colors to labels from 0 to 10. A red circle highlights a cluster of cells. Below the plot, a red box highlights the 'test_type1' label in a legend. A sidebar titled 'Cell identity' and 'Active cell identity' shows the current state of cell assignments, with 'new_type_2' highlighted in blue.

3.2.5 Cell clustering – Cluster merge and create new cell labels

Custom cell labeling

Step 1: Create cell filters

Category: Cluster

Gene: Direction > Threshold

ADD CLUSTER FILTER

ADD GENE FILTER

Filters applied:

- in seurat_clusters: 6
- in seurat_clusters: 4
- in seurat_clusters: 0

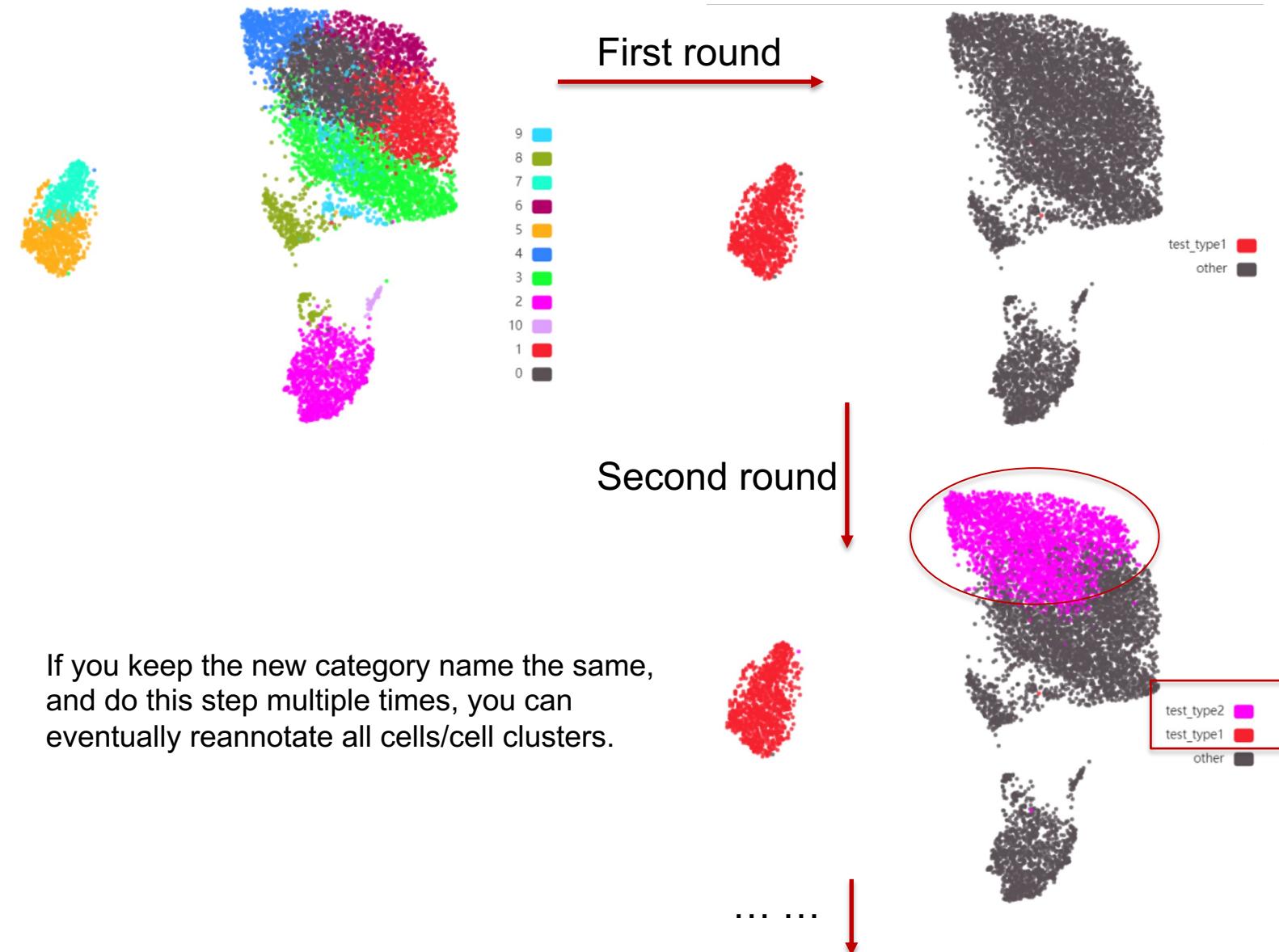
REMOVE ALL

Step 2: Assign cells to new label

Set new category: new_type_2

Add new label: test_type2

ASSIGN CELLS



3.2.6 Cell clustering – Subset cells/cell clusters

Cell selection

Cell selection

Category

new_type_2

Cluster

test_type1,

test_type2

ADD CLUSTER FILTER

Gene

Direction

>

Threshold

ADD GENE FILTER

Selections applied:

new_type_2: test_type2



new_type_2: test_type1

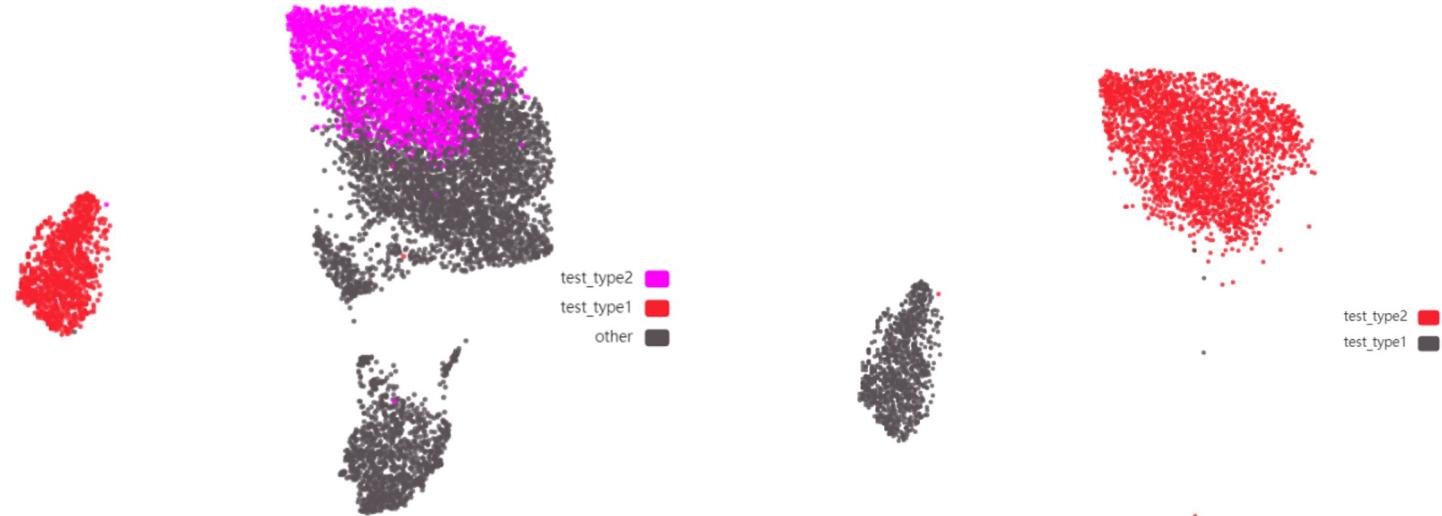


REMOVE ALL

SUBSET CELLS

RESET

To remove any unwanted cells or perform cell sub-clustering, you can use the “Cell selection” function. Similar as the previous one to set rules and subset.



3.2.7 Cell clustering – DEG analysis

The screenshot shows a user interface for performing differential gene expression (DEG) analysis. At the top, there are dropdown menus for 'Group 1' (set to '0') and 'Group 2' (set to '1 (+9 others)'). A 'Direction' dropdown is set to 'all', which is highlighted with a red arrow and a red box. Below these are input fields for 'p-value threshold' (0.05), 'min cell percentage' (0.1), and 'logFC threshold' (0.25). A 'CALCULATE' button is present. To the right of the direction dropdown is a vertical stack of three buttons: 'up', 'down', and 'all', with 'all' being the active option (highlighted with a blue background).

Below the search bar is a table with columns: Gene, pct_1, pct_2, logFC ↓, and adjusted p-value. The table lists the following genes:

Gene	pct_1	pct_2	logFC ↓	adjusted p-value
AGAP1	0.615	0.29	1.1754	7.9405e-158
GNLY	0.508	0.317	1.0507	5.4977e-54
MEAK7	0.367	0.177	0.9993	2.1002e-69
RNF213	0.974	0.875	0.9046	7.704e-227
AC139720.1	0.415	0.229	0.8641	7.5721e-56
CADM1	0.694	0.387	0.8623	1.535e-115
GZMK	0.431	0.223	0.8523	2.2284e-69
TIAM1	0.453	0.263	0.8464	1.7529e-56
BCL11B	0.793	0.498	0.8412	6.6515e-125

At the bottom, there are pagination controls for 'Rows per page' (set to 10) and '1-10 of 863'.

Whether you want both positive and negative DEGs, or just one direction.

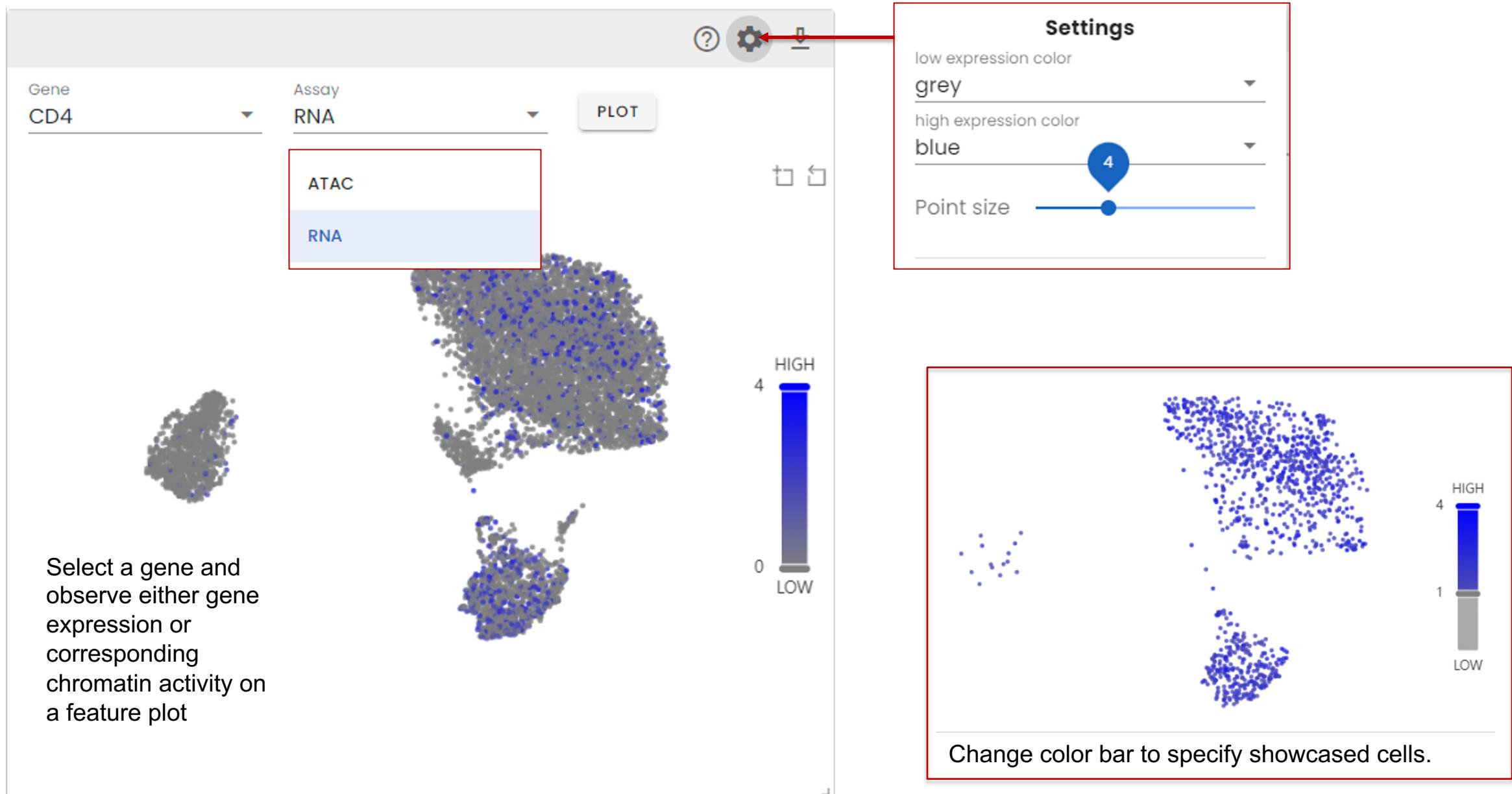
DEG analysis uses Wilcoxon rank test. You need to select the two groups to be compared, either a normal comparison (1 vs all) or a 1 vs 1 comparison. Note that, switch to the right active cell identity first, otherwise you may not see the right cell labels.

Cell identity

Active cell identity

Select identity
hgt_cluster

3.2.8 Cell clustering – Gene feature plot



3.2.9 Cell clustering – Violin plot



View gene violin plot. If you have multiple metadata, you can choose to split visualization by one type. We follow the same logic as used in Seurat.

3.2.10 Cell clustering – Functional enrichment

Gene set enrichment analysis

Tool: GSEA MSigDB data...▼ Direction: all CALCULATE

pathway	Adjusted p-value	NES ↓	si
DEURIG_T_CELL_PROLYMPHOCYTIC_LEUKEMIA_DN	0.7061	2.8736	5
GRAHAM_CML_DIVIDING_VS_NORMAL QUIESCENT_UP	0.5972	2.4825	13
LEE_DIFFERENTIATING_T_LYMPHOCYTE	0.7113	2.4601	5
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	0.5972	2.3336	12
PUJANA_BRCA2_PCC_NETWORK	0.6096	2.1762	3
BOHN_PRIMARY_IMMUNODEFICIENCY_SYNDROM_UP	0.6096	2.1026	8
KANG_DOXORUBICIN_RESISTANCE_UP	0.5972	2.0821	9
GRAHAM_CML QUIESCENT_VS_NORMAL QUIESCENT_UP	0.6096	2.0709	7
FUJII_YBX1_TARGETS_DN	0.6316	2.0455	19

Rows per page: 10 1-10 of 4918 < >

Hallmark gene sets (H)

Positional gene sets (C1)

Curated gene sets (C2)

Regulatory target gene sets (C3)

Computational gene sets (C4)

Ontology gene sets (C5)

Oncogenic signature gene sets (C6)

3.2.11 Cell clustering – Coverage plot



The same coverage function used in Seurat and Signac. More details can be found here.
https://satijalab.org/signac/articles/pbmc_vignette.html#plotting-genomic-regions-1

3.3 Network construction

Gene Regulatory Network

CALCULATE

1. Select cell cluster
0

Plot regulons network:

2. Select regulons
SRF (+4 others)

Add top N by centrality

Add top N by logFC

Selections applied:

TF	Centrality	logFC	Remove
SRF	0.1168	0.2796	trash
NR3C1	0.1303	0.1832	trash
SREBF1	0.1284	0.1045	trash
NFIA	0.1448	0.0965	trash
JUND	0.1249	0.0917	trash

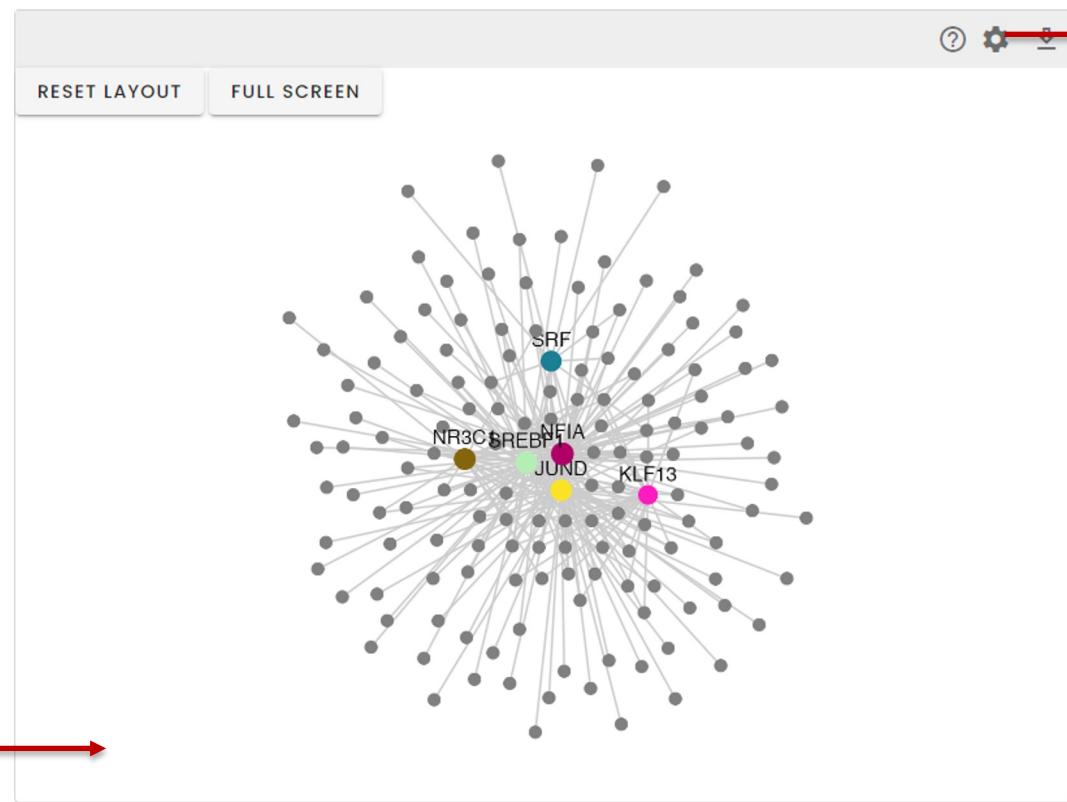
Rows per page: 10 1-5 of 5 < >

PLOT REMOVE ALL

Search TF/genes of interest

TF Target genes

DeepMAPS provides network construction via Cytoscape. Choose a cluster, select TFs, and plot network.



3.3 Network construction

Regulons						
TF	Centrality	# genes	RAS	logFC ↓	adj.p-value	Details
SRF	0.1168	13	0.1331	0.2796	0.1036	<button>Show</button>
NR3C1	0.1303	18	0.1072	0.1832	1	<button>Show</button>
SREBF1	0.1284	45	0.0598	0.1045	1	<button>Show</button>
NFIA	0.1448	69	0.1112	0.0965	0.0027	<button>Show</button>
JUND	0.1249	51	0.0743	0.0917	1	<button>Show</button>
SOX2	0.1206	47	0.0985	0.0829	1	<button>Show</button>
MAFK	0.1376	53	0.0583	0.0687	1	<button>Show</button>
MYB	0.1485	54	0.0946	0.0643	1	<button>Show</button>
MEF2B	0.3114	67	0.1194	0.0509	1	<button>Show</button>
CDX2	0.2179	92	0.2048	0.0454	1	<button>Show</button>

Rows per page: 10 < >

For each cell cluster, we provide a full regulon list named by TFs. For more details of how centrality, RAS, and logFC is calculated, please check DeepMAPS manuscript.

3.3 Network construction

TF	Centrality	# genes	RAS	logFC ↓	adj.p-value	Details
SRF	0.1168	13	0.1331	0.2796	0.1036	HIDE
NR3C1	0.1303	18	0.1072	0.1832	1	SHOW

HIDE
SHOW

Regulon genes

Number of genes: 13

BATF UNIPROT GENECARDS

TPM4 UNIPROT GENECARDS

RNF149 UNIPROT GENECARDS

TOX UNIPROT GENECARDS

STIM1 UNIPROT GENECARDS

PSTPIP1 UNIPROT GENECARDS

TRAF3 UNIPROT GENECARDS

LYST UNIPROT GENECARDS

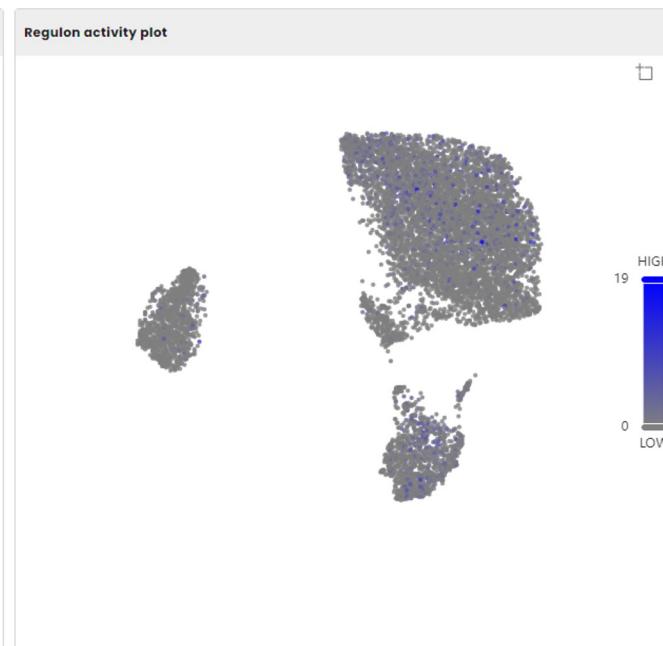
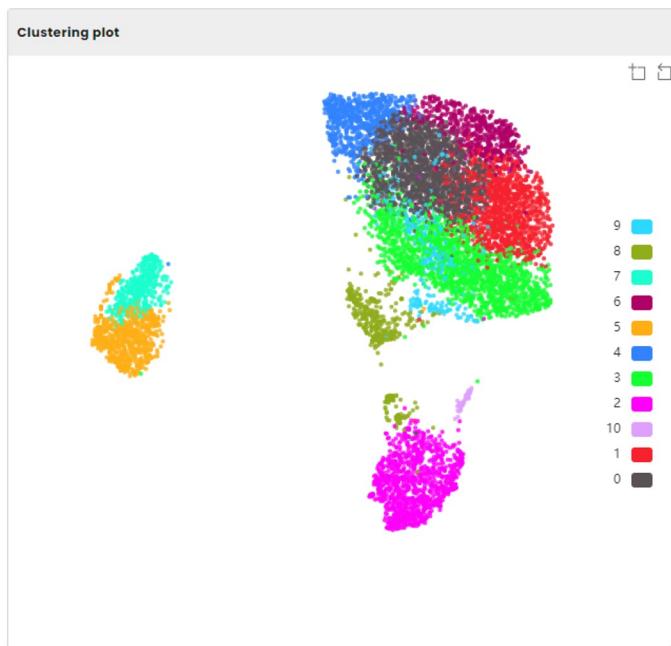
EVL UNIPROT GENECARDS

ARID5A UNIPROT GENECARDS

CFL1 UNIPROT GENECARDS

IGF2R UNIPROT GENECARDS

We can also check all regulated genes in a TF-regulon and regulon activity UMAP.



CITE-seq and multiple scRNA-seq integrative analysis examples are similar to the scRNA-seq and scATAC-seq, with minor differences. We will add more functions in the future updates.

(Network construction currently is only provided for scRNA-seq and scATAC-seq analysis)

4. User workspace

Once you register and login your account, you will see a workspace that record all history jobs.

Profile



Email: flykun0620@gmail.com
Name: Cankun Wang

Projects

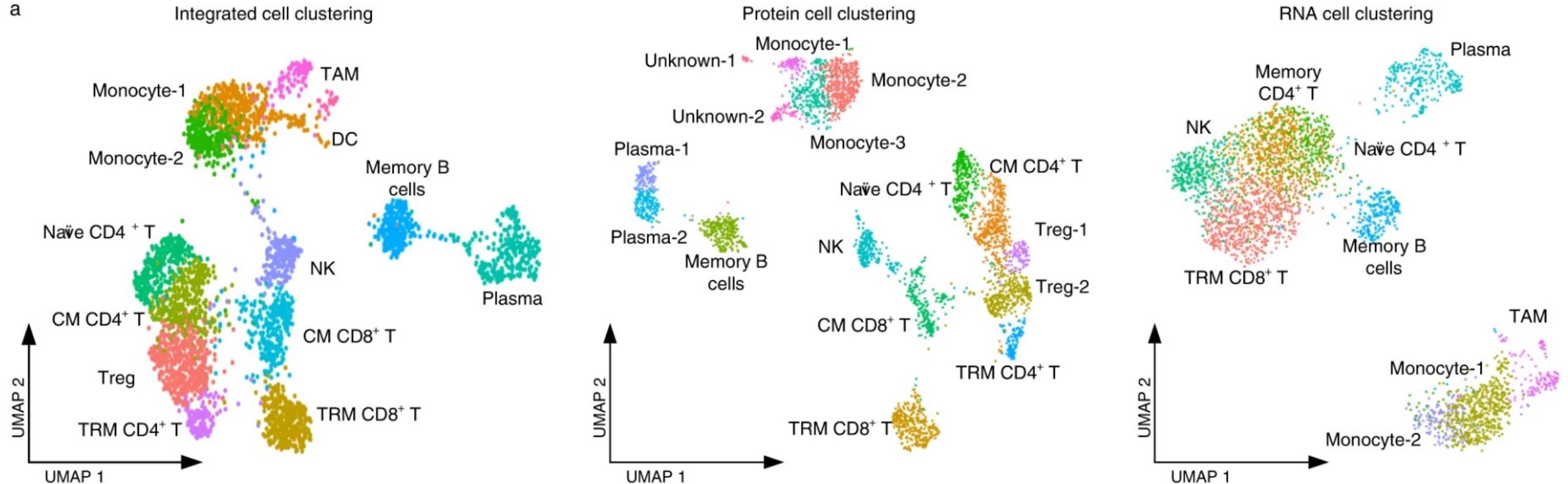
Job ID	Title	description	Created at	Type	Action
1642388187073	TEST PROJECT 1	TEST	2022-01-17	multiple-rna	OPEN
1642398270344	TEST PROJECT 2		2022-01-17	multiome	OPEN
lymphoma_14k	10X Human lymphoma 14k cells		2022-01-17	multiome	OPEN

Applications:

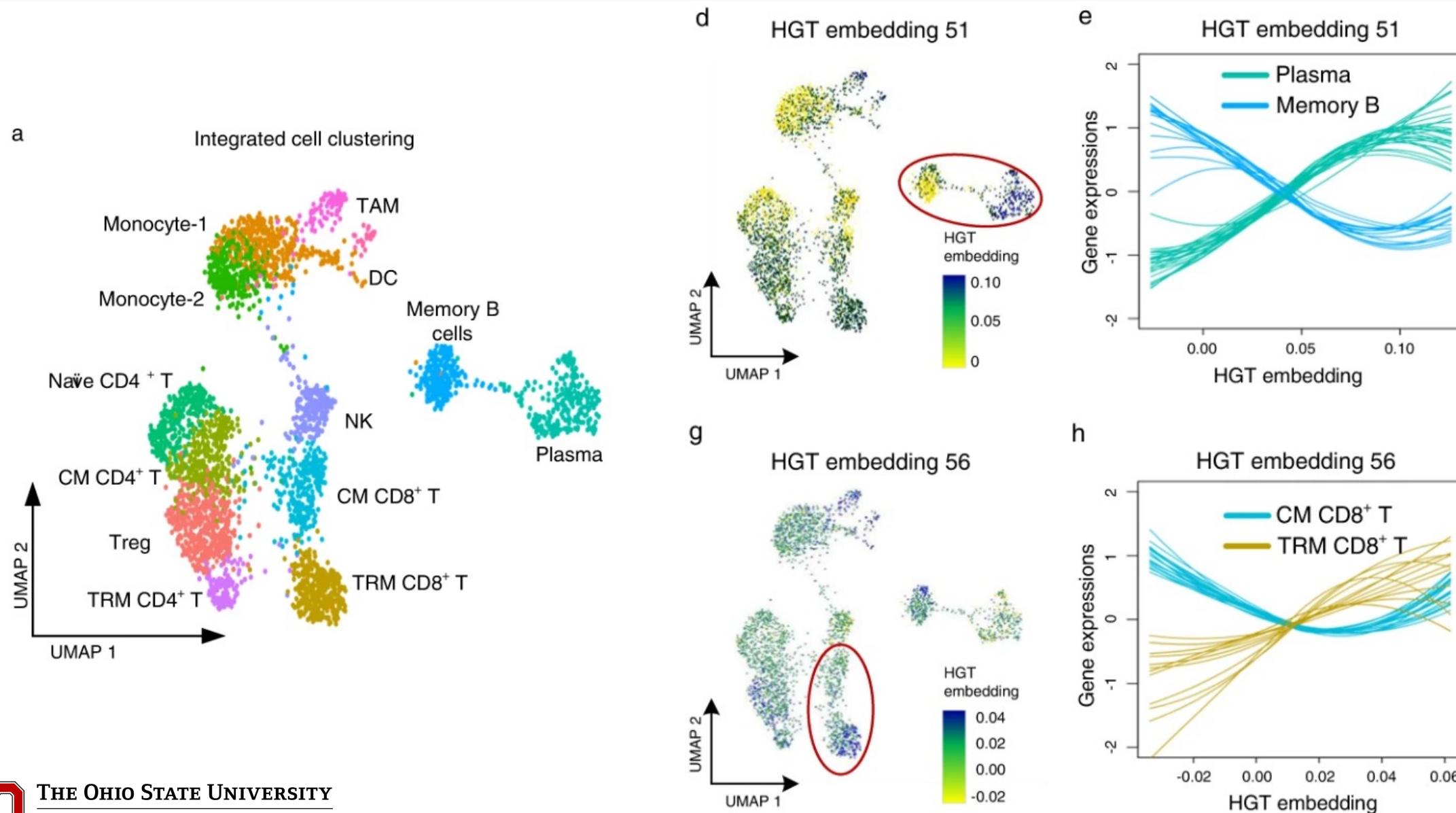
- **Analysis of human PBMC and lung tumor leukocytes with CITE-seq data**
- **Analysis of human IFNB-stimulated and control PBMCs with multiple scRNA-seq data**

Analysis of human PBMC and lung tumor leukocytes with CITE-seq data

Case study: lung tumor PBMC CITE-seq data

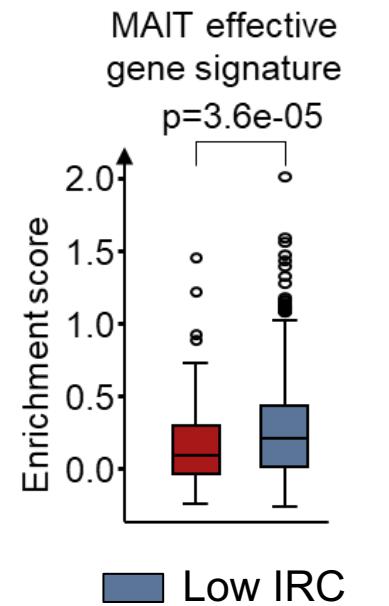
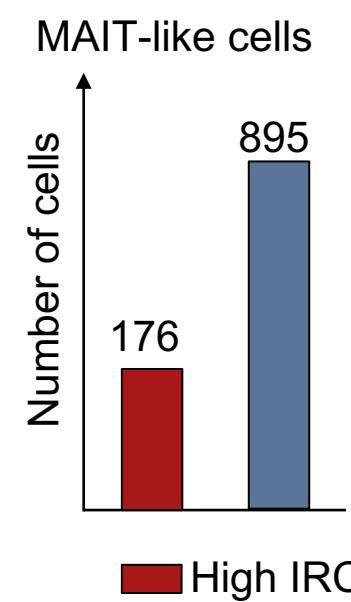
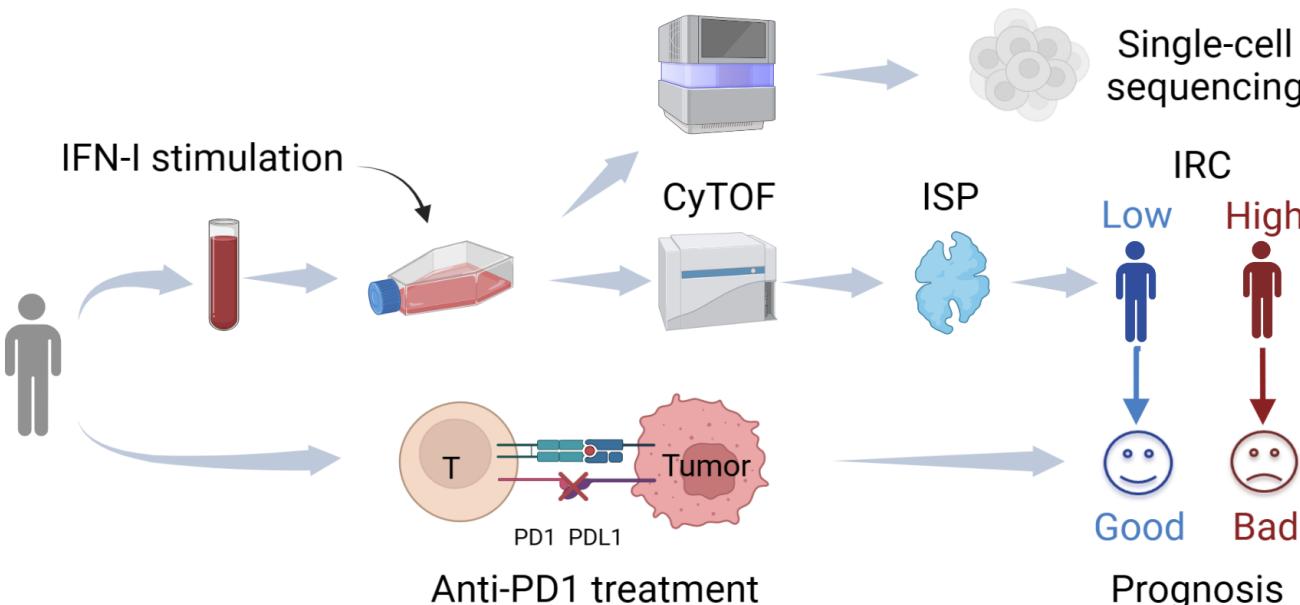


Case study: lung tumor PBMC CITE-seq data



Analysis of human IFNB-stimulated and control PBMCs with multiple scRNA-seq data

Different Interferon-I response capacity (IRC) related with MAIT effective



High IRC: MAIT/Total=176/44020=0.40%
Low IRC: MAIT/Total=895/30771=2.9%

Data information (Eight patients):

- Four high IRC with bad prognosis after PD1-PDL1 blocking therapy.
- Four Low IRC with good prognosis after PD1-PDL1 blocking therapy.

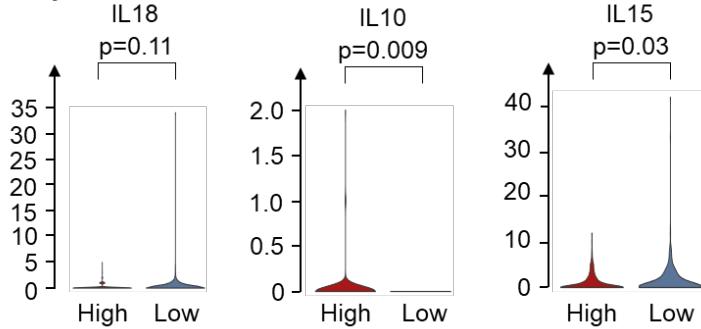
Results:

- The effective gene signature active score in MAIT-like cells is higher in low IRC patients than that in high IRC patients.

MarsGT infers the potential mechanism in sample with differential prognosis

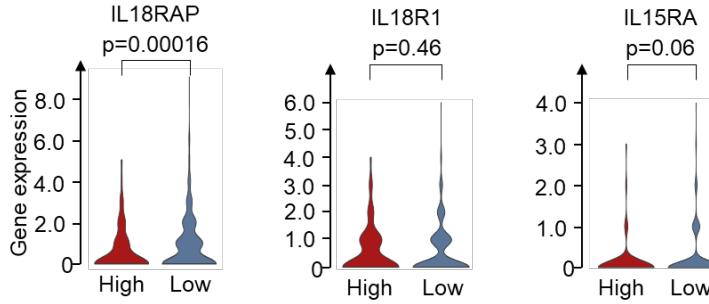
①

Cytokines in DC cells



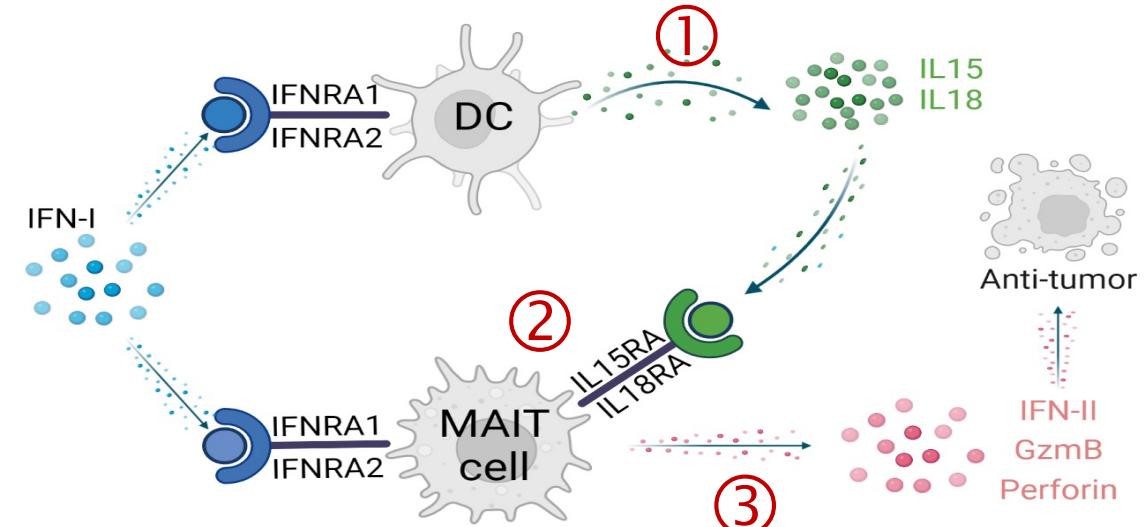
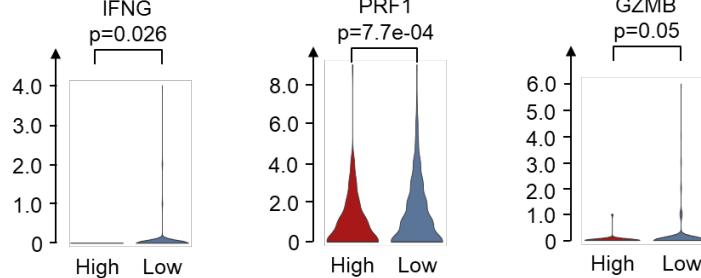
②

Receptors in MAIT-like cells



③

Secreted factors in MAIT-like cells



The transcription factor of GZMB, IFNG RPF1

- GZMB: STAT1, FOS (Low IRC)
- IFNG: NFKB1/2(Low IRC)
- RPF1: JUN (both high and low IRC)

We infer the potential mechanism based on the observed. The critical transcription factors are validated by the ChIP-seq data in the public domain, **we would like to have a chance to validate our prediction by experiment.**



THE OHIO STATE UNIVERSITY

WEXNER MEDICAL CENTER



THANK YOU

Acknowledgement

Please cite:

Ma, A., Wang, X., Li, J. et al. Single-cell biological network inference using a heterogeneous graph transformer. Nat Commun 14, 964 (2023). <https://doi.org/10.1038/s41467-023-36559-0>

Funding:

R01-GM131399, R35-GM126985, and U54AG075931 from the National Institute of General Medical Sciences of the National Institutes of Health.
NSF1945971 from the National Science Foundation.