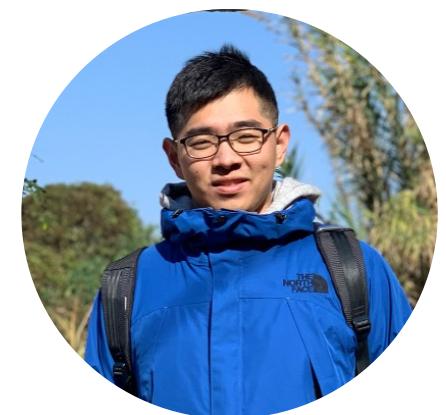


SSGG Short Course Series:  
Selective Introduction of Multi-Omics Analysis

## Lecture 4

# Multi-Omics Causal Mediation Analysis and Single-Cell Multi-Omics Analysis

April 20, 2023



Instructor: **Rick Chang**

# Outline

## Multi-Omics Causal Mediation Analysis

1. Causal mediation analysis
2. The difficulty of mediation analysis in omics data
3. Overview of high-dimensional mediation analysis
4. Penalization-based method: HIMA (lab session)

## Single-Cell Multi-Omics Analysis

1. Bulk vs. Single-Cell
2. Single-Cell multi-omics data and integration methods
3. Integrated analysis: Seurat 4.0 (lab session)

# Causation vs. Association

- Causal inference is an essential component for the discovery of disease mechanism.

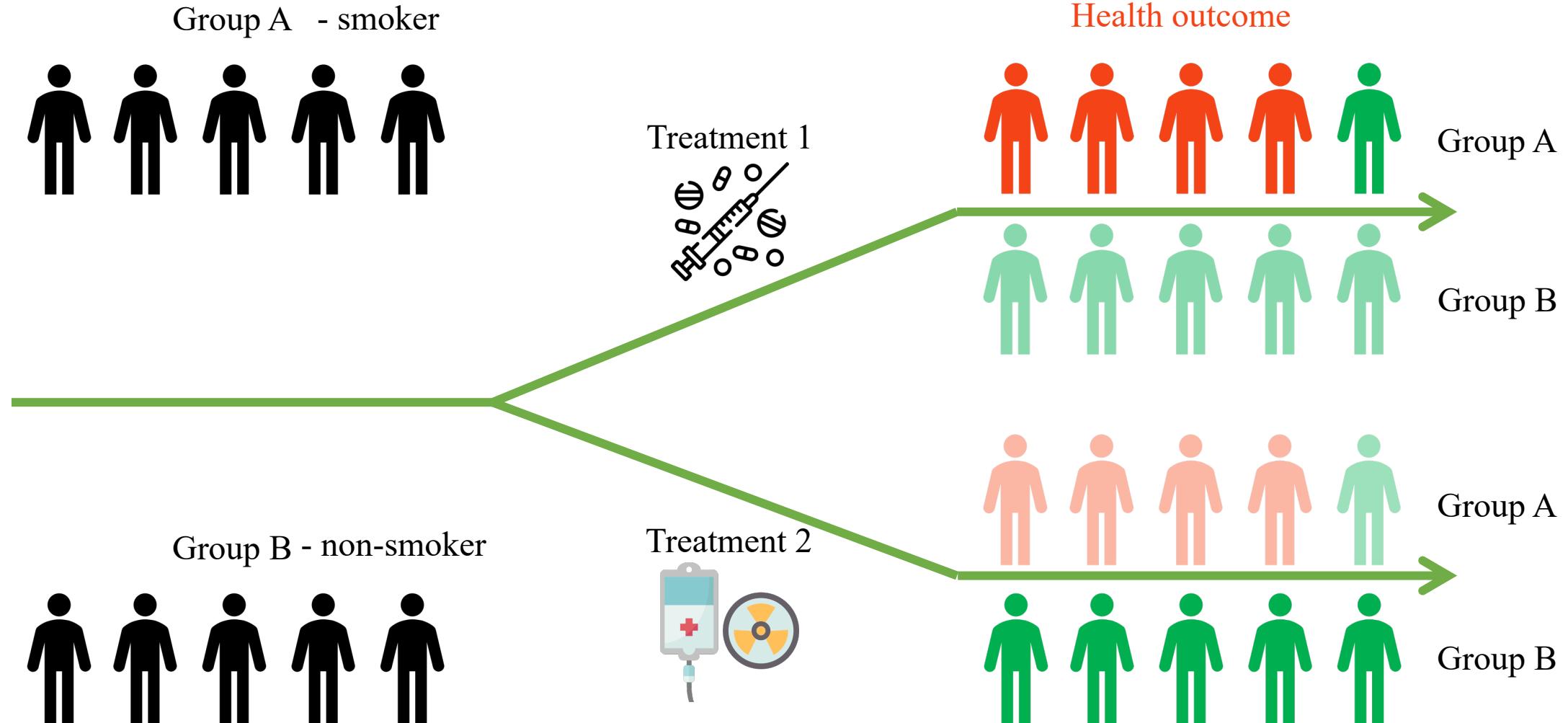


Exposure X ..... ? .....> Outcome Y



# Causation vs. Association

- Causation: what might have happened if X occurred
- Any unmeasured confounding variable?



# Mediation Effect

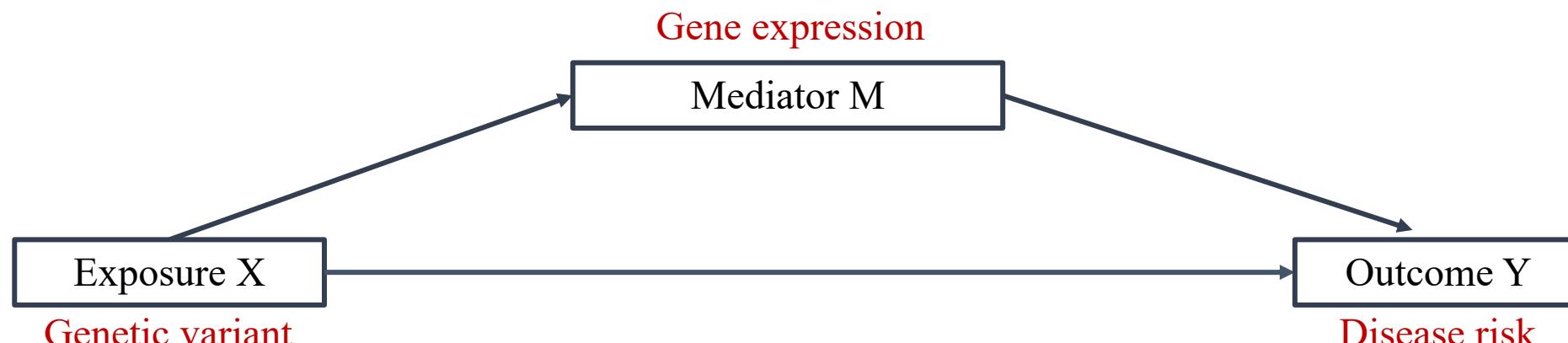
Mediation analysis: to further explore the mechanism behind the causation

Causal mediation effect:

- Exposure has a causal effect on the mediator
- Mediator has a causal effect on the outcome conditional on the exposure

Example

- Genetic variant leads different expression level which may increase the risk of disease.



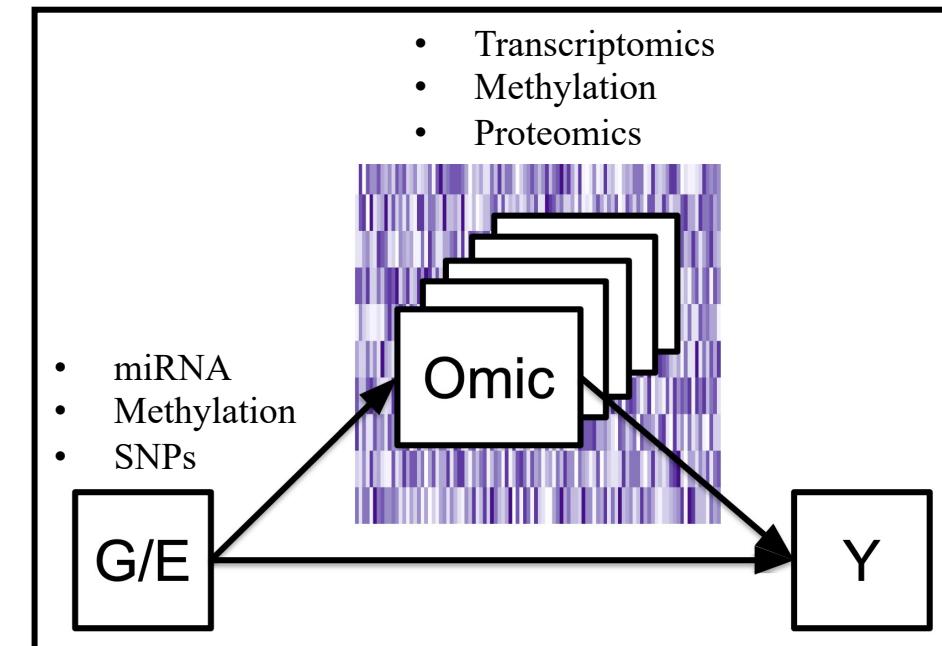
# Causal Mediation Analysis In Omics Studies

- Causal mediation analysis seeks to investigate the intermediate mechanism through an exposure on the outcome of interest.
- Rising interest in omics studies to identify the mechanism of molecular-level traits
  - E.g. DNA → RNA → Protein → outcome
- Mediation analysis in omics studies is challenging:
  - High-dimensional mediators → identifiability problem
  - Composite null hypothesis → weak power

# Some Applications

General setting: One exposure → multiple mediators → one outcome

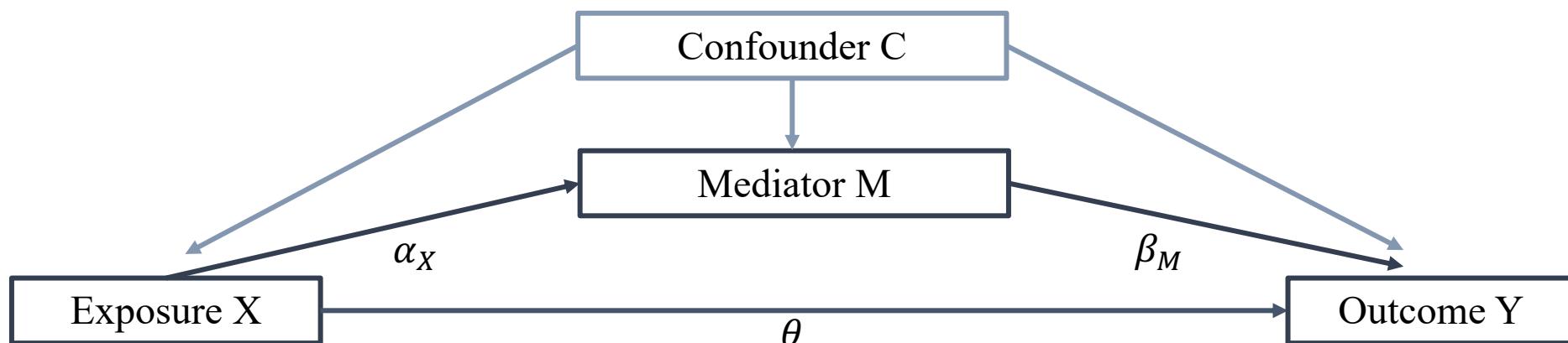
- Environment → DNA Methylation → Outcome
  - E.g. Normative Aging Study ([Bind et al., 2014 Epigenetics](#); [Zhang et al., 2016 Bioinformatics](#); [Liu et al., 2022 JASA](#))  
Prostate Cancer ([Dai et al., 2022 JASA](#))  
Atherosclerosis ([Song et al., 2020 Biometrics](#); [Clark-Boucher et al., 2023 medRxiv](#))
  - Also known as **epigenome-wide mediation analysis**
- microRNAs → Gene expression → Outcome
  - E.g. Glioblastoma ([Huang et al., 2014 AOAS](#);  
[Huang and Pan, 2016 Biometrics](#))  
Brain cancer ([Loh et al., 2020 Biometrics](#))
- Others
  - E.g. Air Pollution ([Inoue et al., 2020 JASA](#))  
Neuroimaging ([Chén et al., 2018 Curr. Environ. Health Rep.](#);  
[Zhao et al., 2021 CSDA](#))



# Causal Mediation Model

Two linear regressions method proposed by Baron and Kenny, 1986 J Pers Soc Psychol

- (Model  $X \rightarrow M$ )  $M = C\alpha_C + X\alpha_X + \epsilon_M$
- (Model  $M \rightarrow Y$ )  $Y = C\beta_C + X\theta + M\beta_M + \epsilon_Y$ , where  $\epsilon_Y \sim N(0, \sigma_Y^2)$  and  $\epsilon_M \sim N(0, \sigma_M^2)$
- Since  $Y = C\beta_C + X\theta + M\beta_M + \epsilon_Y$   
 $= C\beta_C + X\theta + (C\alpha_C + X\alpha_X + \epsilon_M)\beta_M + \epsilon_Y$   
 $= C\beta_C + X\theta + C\alpha_C\beta_M + X\alpha_X\beta_M + \epsilon_Y^*$
- Direct effect is  $\theta$ , and indirect effect (mediation effect) can be expressed as  $\alpha_X\beta_M$
- Total effect  $\gamma = \theta + \alpha_X\beta_M$



# High-Dimensional Mediation Analysis

- Challenge 1: High-dimensional mediators ( $M_1, \dots, M_p$ )

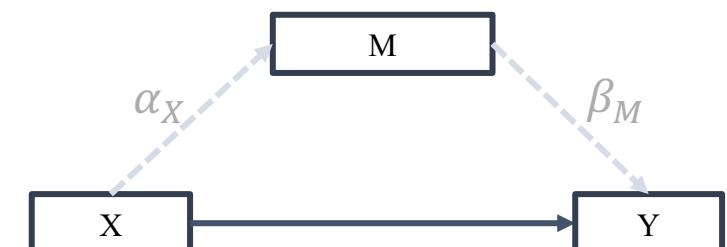
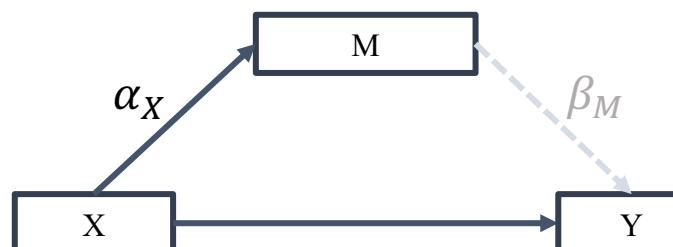
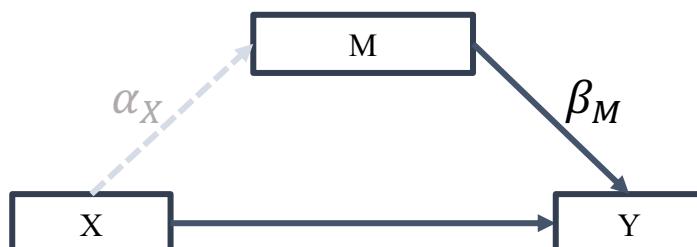
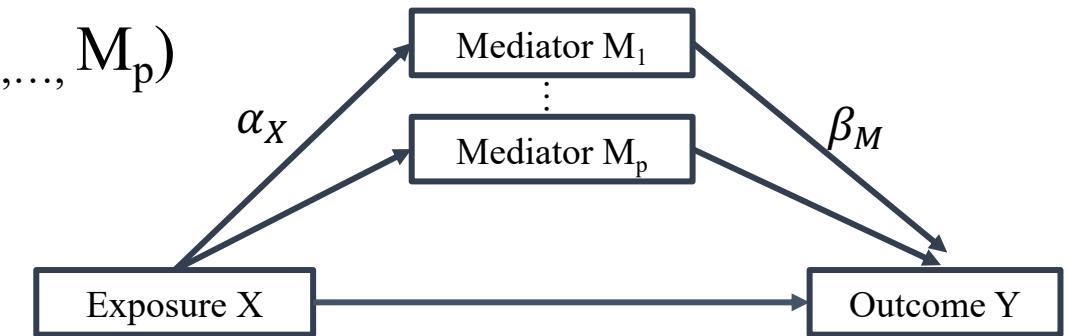
- (Model  $M \rightarrow Y$ )  $Y = C\beta_C + X\theta + \sum_j M_j \beta_{M,j} + \epsilon_Y$

- When the number of mediators ( $p$ ) is much greater than the sample size ( $N$ ),  $\beta_{M,j}$  are not estimable.

- Identification assumptions could be easily violated after dimension reduction (Huang and Pan, 2016 Biometrics)

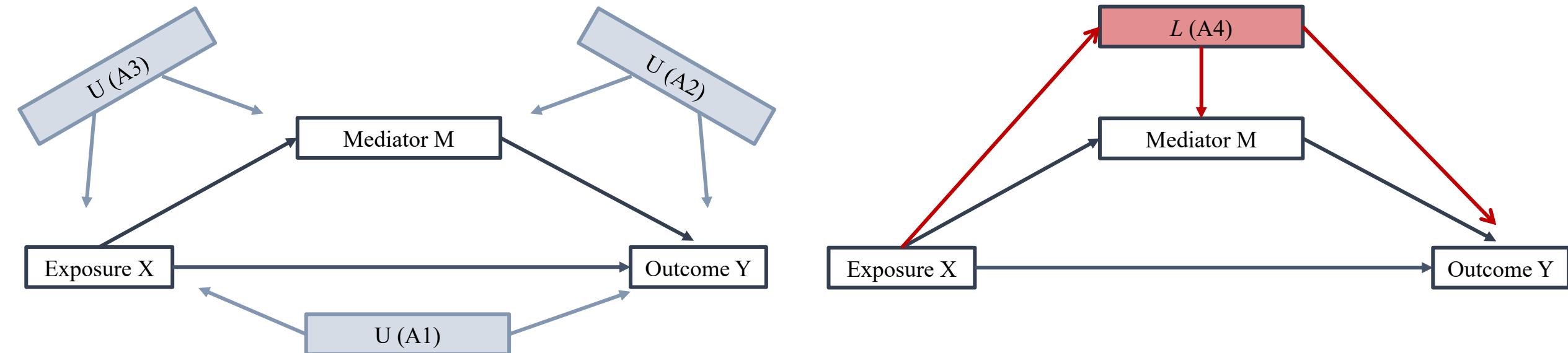
- Challenge 2: Composite null hypothesis ( $H_0: \alpha_X \beta_M = 0$ )

- Traditional hypothesis tests are underpowered for testing composite null hypothesis. (Liu et al., 2022 JASA)  
E.g. Sobel's test and joint significant test (MaxP)



# Four Identification Assumptions

- [A1]  $Y(x) \perp\!\!\!\perp X|C$ : no unmeasured confounding for the association of Y and X
- [A2]  $Y(x, m) \perp\!\!\!\perp M|(X, C)$ : no unmeasured confounding for the association of Y and M given X
- [A3]  $M(x) \perp\!\!\!\perp X|C$ : no unmeasured confounding for the association of M and X
- [A4]  $Y(x, m) \perp\!\!\!\perp M(x^*)|C$ : no X-induced confounder for the M-Y association  
(cross-world assumption)



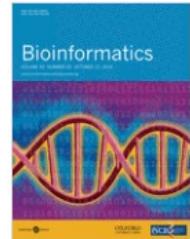
# Overview Of High-Dimensional Mediation Analysis

| Methods   | Test Statistics   | Null Distribution  |
|---|---|--|
| correlation-based method  | $P_{\max}$  | permutation  |
| Huang-Pan method  | marginal and component-wise ME based on PCA                                     | Monte Carlo (normal-based or bootstrapping)  |
| causal inference test (CIT)   | $P_{\max}$  | permutation  |
| direction of mediation  | PCA-based   | bootstrapping  |
| MCP-subset  | $P_{\max}$  | screening followed by multiple comparison procedure  |
| MCP-subset based on Westfall-Young  | $P_{\max}$  | screening followed by multiple comparison procedure  |
| MCP-subset based on multivariate  | $P_{\max}$  | screening followed by multiple comparison procedure  |
| HDMA<br>gHMA <sup>#</sup>   | $P_{\max}$<br>ACAT combining gHMA-L and gHMA-NL                                 | screening followed by debiased estimation<br>screening followed by multiple comparison procedure |
| global test + ScreenMin <sup>#</sup>  | $P_{\min}$ followed by $P_{\max}$   | screening followed by multiple comparison procedure  |
| <b>Second category:</b> Mediation methods accounting for the composite nature of the null             |   |  |
| Methods   | Test Statistics   | Null Distribution  |
| JTV-comp <sup>#</sup>   | mixture of multiple-mediator based $P$ value without estimating the proportions | composite null   |
| JT-comp   | mixture of single-mediator based $P$ value without estimating the proportions   | composite null   |
| DACT  | mixture of single-mediator based $P$ value with estimated proportion            | composite null   |
| JS-mixture  | mixture of single-mediator based $P$ value with estimated proportion            | composite null   |
| <b>Third category:</b> Penalization-based mediation regression methods and Bayesian mediation methods |   |  |
| Methods   | Prior Effects Assumptions   | Optimization Procedure   |
| pathway Lasso   | penalization based method   | ADMM   |
| HIMA  | $P_{\max}$  | screening followed by minimax concave penalty estimation   |
| BAMA  | spike-and-slab prior  | MCMC   |
| BAMA with joint priors  | Gaussian mixture prior and, product threshold Gaussian prior                    | MCMC   |
| BAMA with joint priors considering correlation among mediators  | the Potts prior and logistic normal prior                                       | MCMC   |

Popular methods

\*Lab session

# High-Dimensional Mediation Analysis (HIMA)



Volume 32, Issue 20  
October 2016

*Bioinformatics*, 32(20), 2016, 3150–3154  
doi: 10.1093/bioinformatics/btw351  
Advance Access Publication Date: 29 June 2016  
Original Paper

Genetics and population analysis

## Estimating and testing high-dimensional mediation effects in epigenetic studies

Haixiang Zhang<sup>1</sup>, Yinan Zheng<sup>2</sup>, Zhou Zhang<sup>2</sup>, Tao Gao<sup>2</sup>, Brian Joyce<sup>2</sup>, Grace Yoon<sup>3</sup>, Wei Zhang<sup>2</sup>, Joel Schwartz<sup>4</sup>, Allan Just<sup>5</sup>, Elena Colicino<sup>4</sup>, Pantel Vokonas<sup>6</sup>, Lihui Zhao<sup>2</sup>, Jinchi Lv<sup>7</sup>, Andrea Baccarelli<sup>4</sup>, Lifang Hou<sup>2</sup> and Lei Liu<sup>2,\*</sup>

<sup>1</sup>Center for Applied Mathematics, Tianjin University, Tianjin 300072, China, <sup>2</sup>Department of Preventive Medicine,

<sup>3</sup>Department of Statistics, Northwestern University, Chicago, IL 60611, USA, <sup>4</sup>Department of Environmental Health, Harvard University, Boston, MA 02115, USA, <sup>5</sup>Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, <sup>6</sup>Veterans Affairs Boston Healthcare System and Boston University School of Medicine, VA Normative Aging Study, Boston, MA 02118, USA and <sup>7</sup>Data Sciences and Operations Department, University of Southern California, Los Angeles, CA 90089, USA

\*To whom correspondence should be addressed.

Associate Editor: Oliver Stegle

Received on December 31, 2015; revised on May 5, 2016; accepted on May 24, 2016

- The most user-friendly tools for high-dimensional mediation analysis

Supported mediator types:

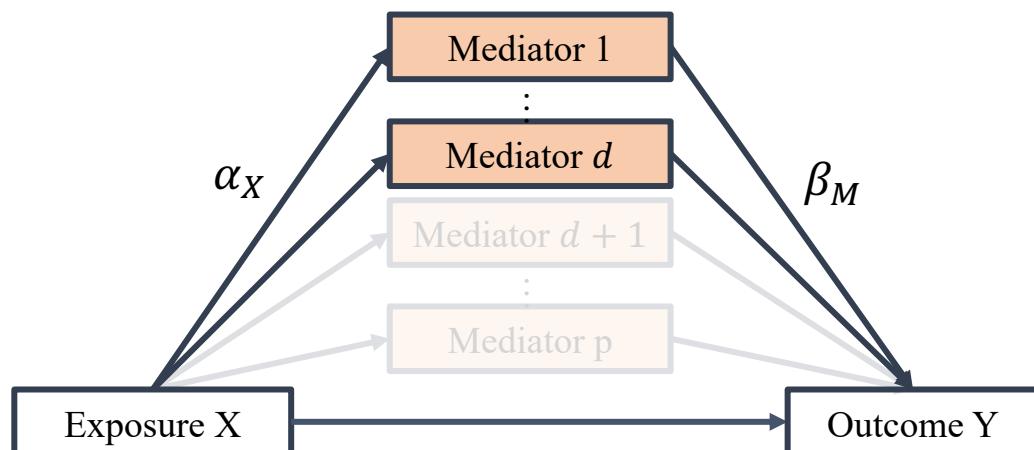
- Epigenetics
- Transcriptomics
- Proteomics
- Metabolomics
- Microbiome

Supported outcomes:

- Continuous
- Binary
- Count
- Survival

# High-Dimensional Mediation Analysis (HIMA)

- HIMA assumes that the true mediators are sparse and applied **Sure independence screening** and **penalty regression** to reduce the dimensionality.
- Workflow
  1. **Sure independence screening** to identify those mediators with large absolute  $\beta_k$
  2. Penalty regression: **Minimax concave penalty** for variable selection
  3. **Joint significance test (MaxP)** of mediator effect (p-value of  $\alpha_X$  and  $\beta_M$ )  
 $T_{MaxP} = \max(p_\alpha, p_\beta)$
  4. Control the family wise error rate (Bonferroni)



(Model  $X \rightarrow M$ )

- $M_1 = C\alpha_{C,1} + X\alpha_{X,1} + \epsilon_{M1}$   
⋮

- $M_d = C\alpha_{C,d} + X\alpha_{X,d} + \epsilon_{Md}$

(Model  $M \rightarrow Y$ )

- $Y = C\beta_C + X\theta + M_1\beta_{M,1} + \cdots + M_d\beta_{M,d} + \epsilon_Y$

# TCGA Glioblastoma Multiforme

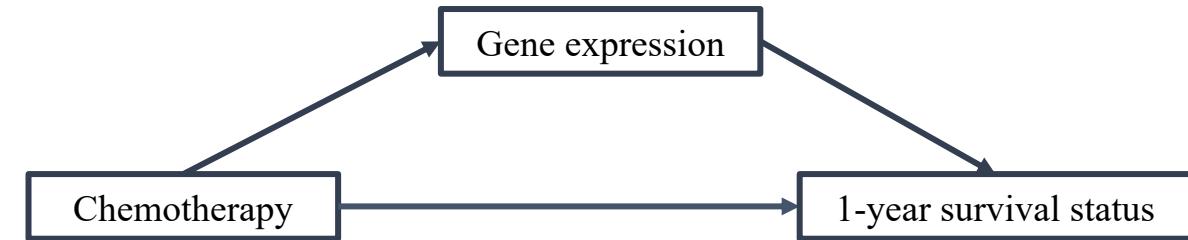
- 469 patients of glioblastoma multiforme have complete genomic data on gene expression (UNC AgilentG4502A-07) archived in The Cancer Genome Atlas (TCGA).
- **Chemotherapy** have been reported to be associated with **survival** of cancer patients.
- Hypothesis: chemotherapy affects survival outcome mainly through its influence on gene expression levels

Exposure (X): chemotherapy (Yes/No)

Mediator (M): gene expression (17450 genes)

Outcome (Y): dichotomous 1-year survival status

Confounders (C): Age, Gender



|        | alpha      | beta       | gamma<br>(Total effect) | alpha*beta<br>(Mediation effect) | % total effect | Bonferroni.p | BH.FDR    |
|--------|------------|------------|-------------------------|----------------------------------|----------------|--------------|-----------|
| DHRS12 | 0.2270880  | -0.3282940 | 1.517597                | -0.0745516                       | -4.9124789     | 0.0202180    | 0.0067547 |
| NDUFA7 | 0.1575399  | 1.0372184  | 1.517597                | 0.1634033                        | 10.7672359     | 0.0086816    | 0.0024884 |
| OR52R1 | -0.1427211 | 0.0156191  | 1.517597                | -0.0022292                       | -0.1468885     | 0.3949076    | 0.0987269 |
| PIGS   | -0.1480009 | 0.5642783  | 1.517597                | -0.0835137                       | -5.5030204     | 0.0202642    | 0.0067547 |

To understand the biological mechanism across multi-omics, you can also try different exposure.

E.g. methylation and miRNA

# Outline

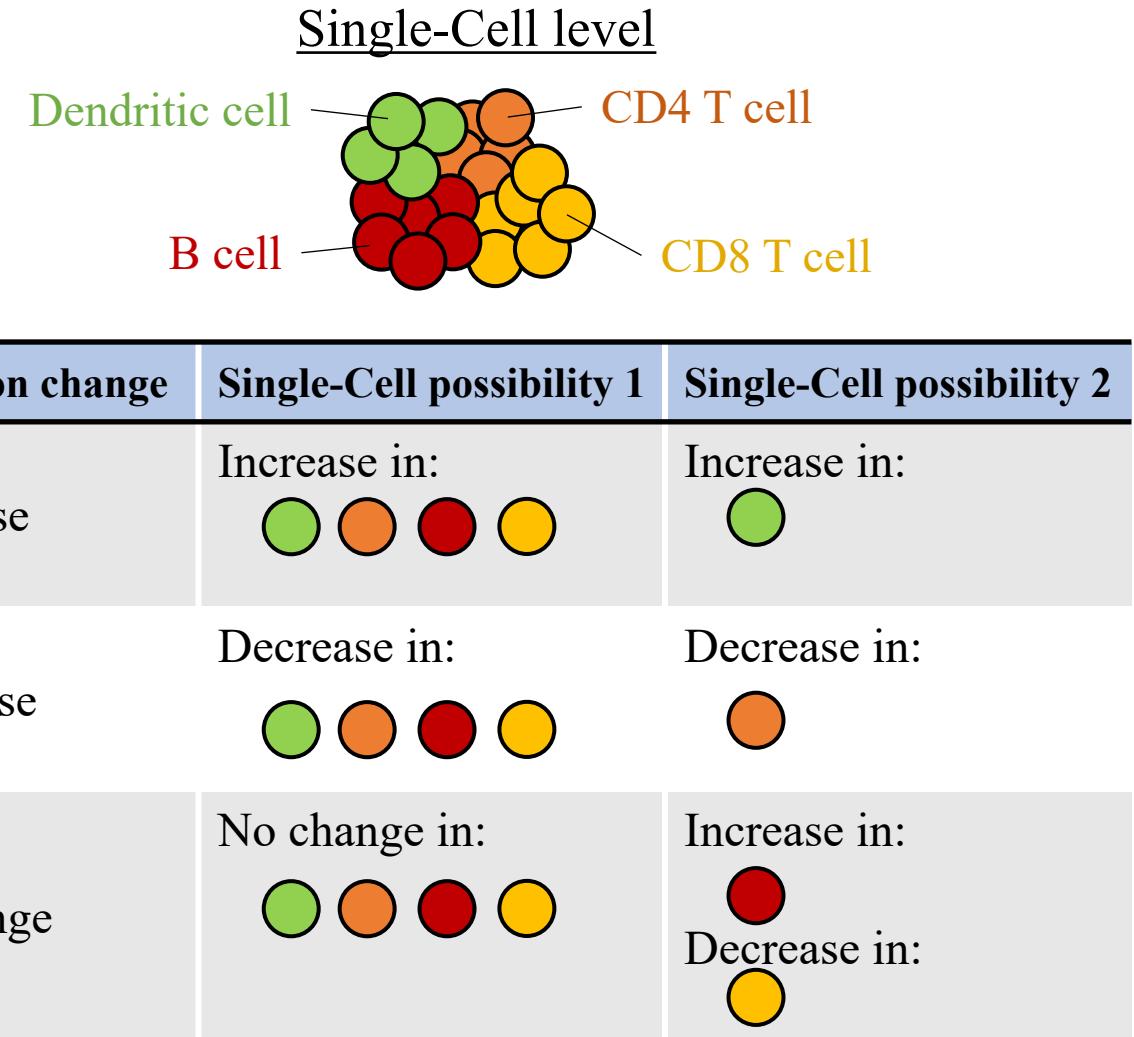
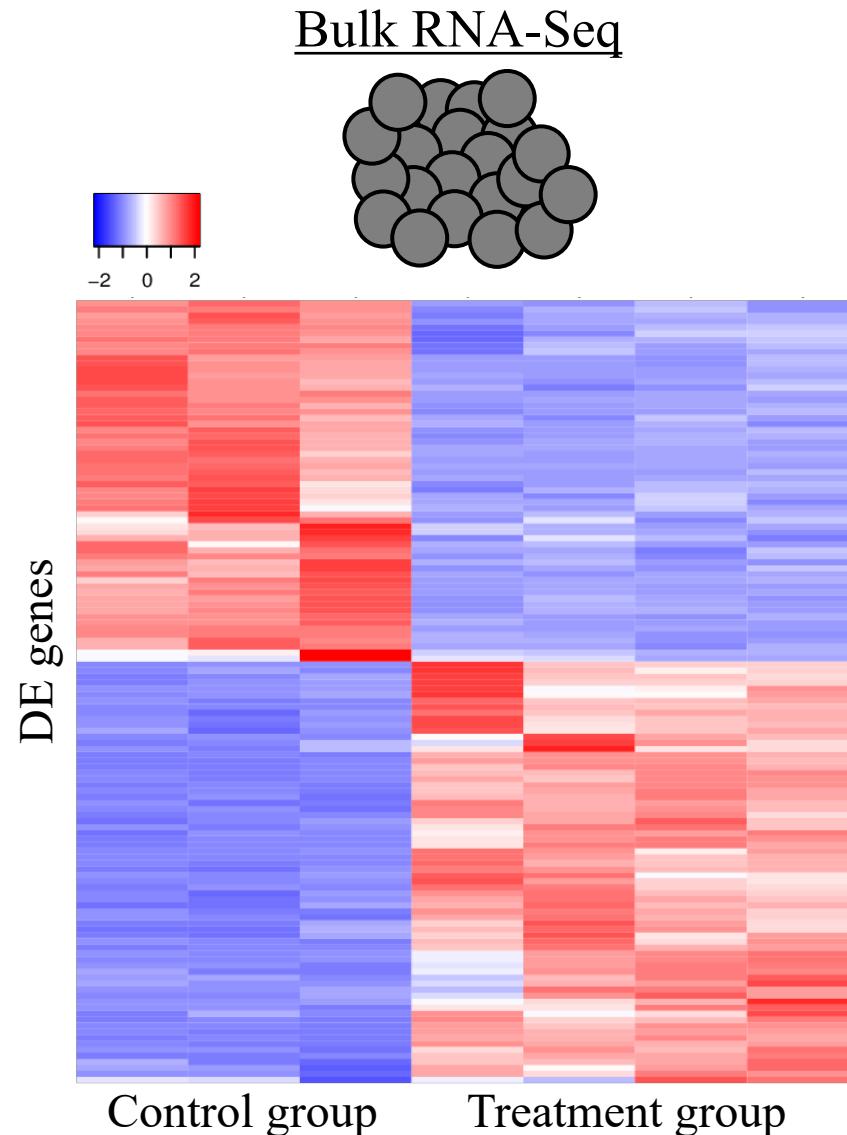
## Multi-Omics Causal Mediation Analysis

1. Causal mediation analysis
2. The difficulty of mediation analysis in omics data
3. Overview of high-dimensional mediation analysis
4. Penalization-based method: HIMA (lab session)

## Single-cell Multi-Omics Analysis

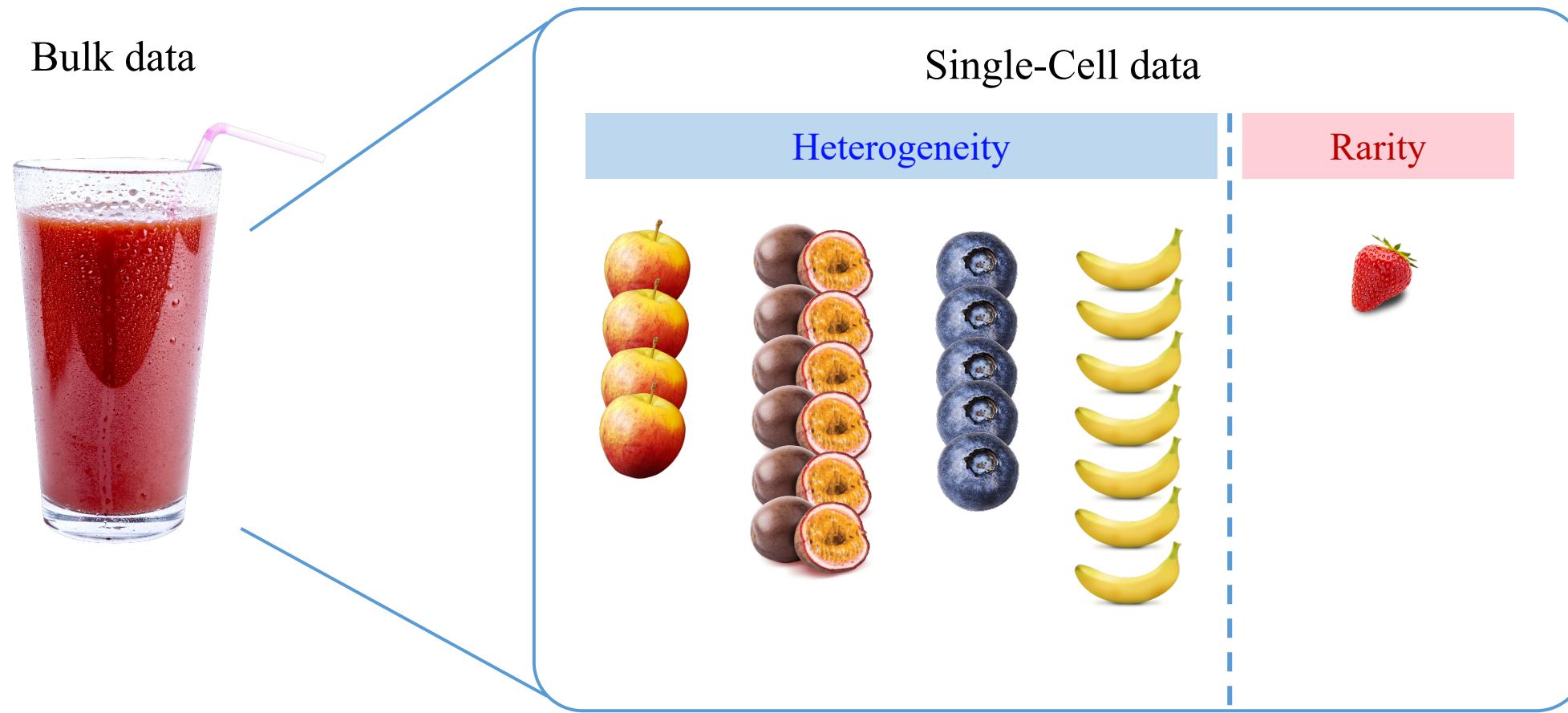
1. Bulk vs. Single-Cell
2. Single-Cell multi-omics data and integration methods
3. Integrated analysis: Seurat 4.0 (lab session)

# From Bulk To Single-Cell



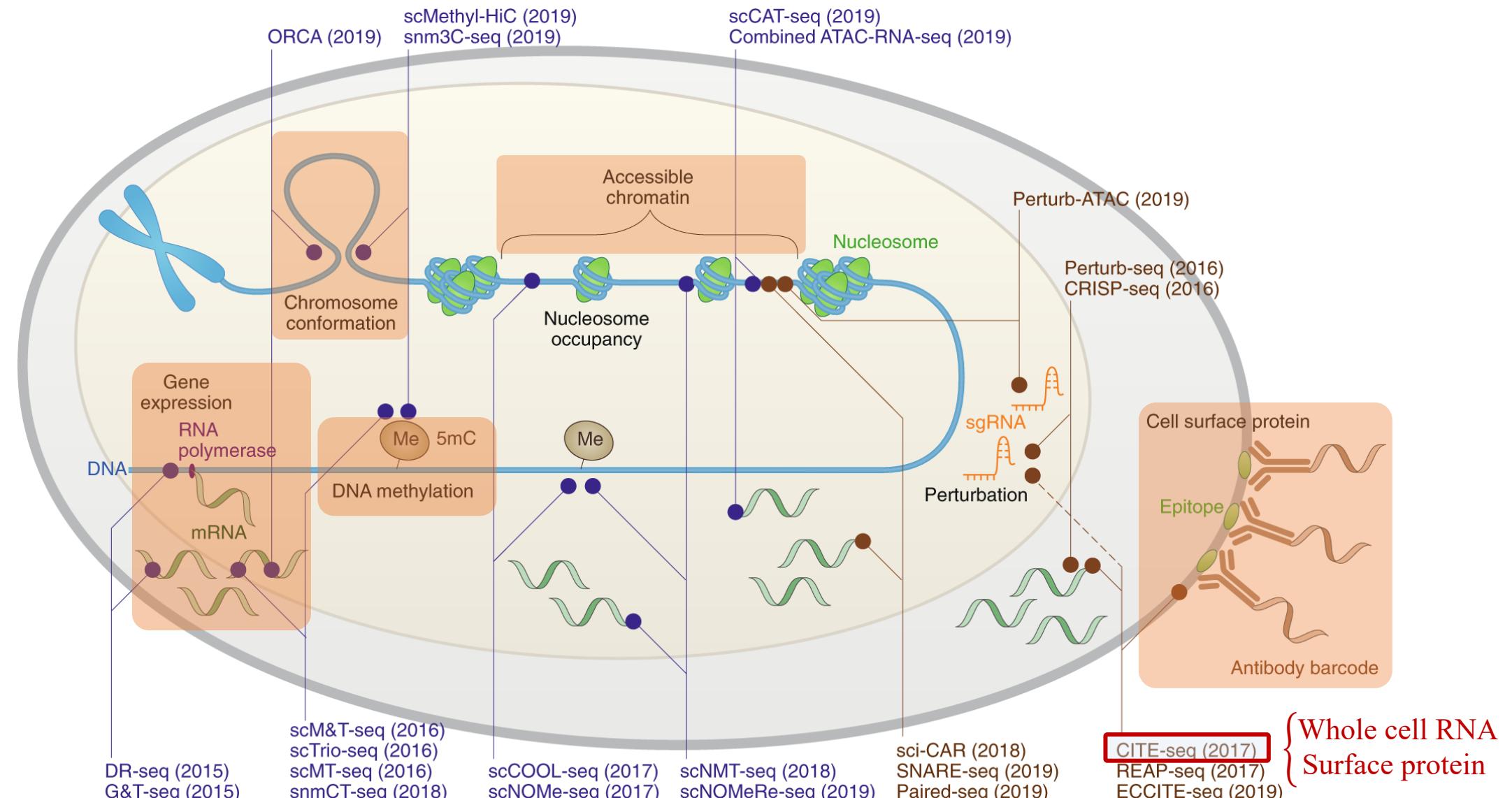
# Heterogeneity And Rarity

- Understanding disease mechanism is challenging because of heterogeneity and rarity of target cells.
- E.g. HIV cells < 0.1% (Collora et al., 2022 Immunity)



Analogy from Ya-Chi Ho's talk

# Single-Cell Multi-Omics Data



# Single-Cell (Multi-)Omics Methods

- CITE-Seq and 10X Multiome are the two frequently used methods in single-cell multi-omics.

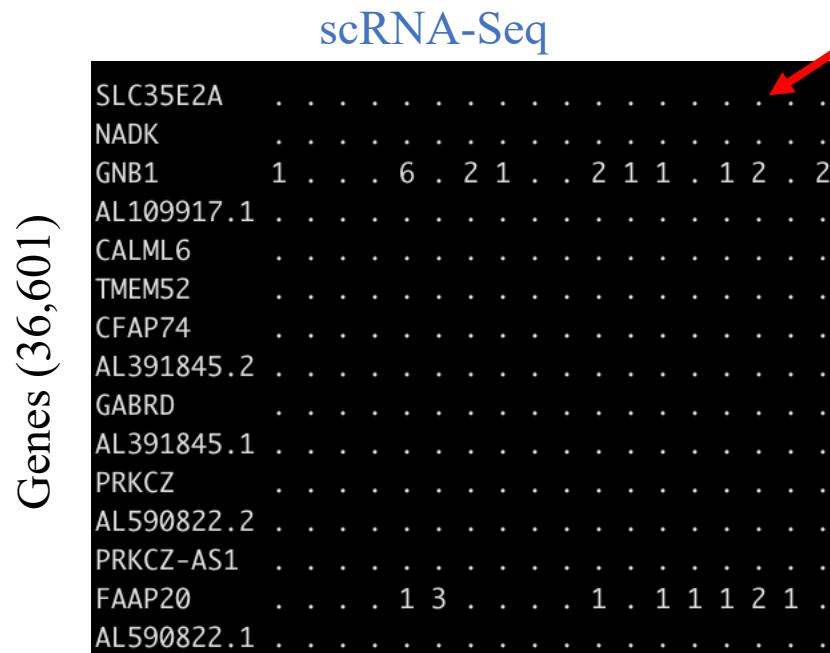
|                     | Epigenome               | Transcriptome |                | Proteome                    |
|---------------------|-------------------------|---------------|----------------|-----------------------------|
|                     | Chromatin accessibility | Nuclear RNA   | Whole cell RNA | Protein abundance           |
| scRNA-Seq           |                         |               | ✓              |                             |
| snRNA-Seq           |                         | ✓             |                |                             |
| scATAC-Seq          | ✓                       |               |                |                             |
| <b>CITE-Seq</b>     |                         |               | ✓              | ✓ (surface)                 |
| ASAP-Seq            | ✓                       |               |                | ✓ (surface + intracellular) |
| <b>10X Multiome</b> | ✓                       | ✓             |                |                             |
| inCITE-Seq          |                         | ✓             |                | ✓ (intranuclear)            |
| DOGMA-Seq           | ✓                       |               | ✓              | ✓(surface + intracellular)  |

- DOGMA = CITE-Seq + 10X Multiome

# CITE-Seq

Example: CITE-Seq = scRNA-Seq + ADT

- scRNA: gene expression (Transcriptome)
- ADT: surface protein (Proteome)



Zero expression because of "technical" dropouts

Antibody-Derived Tags (ADTs)

|             |     |     |     |     |     |     |     |     |     |     |     |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| CD86-A0006  | 2   | .   | 1   | 1   | .   | .   | 2   | 1   | 1   | 2   | 1   |
| CD274-A0007 | 23  | 20  | 10  | 29  | 24  | 43  | 18  | 69  | 25  | 13  | 58  |
| CD270-A0020 | 39  | 45  | 39  | 40  | 41  | 31  | 30  | 27  | 33  | 29  | 29  |
| CD155-A0023 | 7   | 6   | 4   | 13  | 4   | 4   | 1   | 4   | 4   | 8   | 6   |
| CD112-A0024 | 9   | 21  | 4   | 33  | 13  | 5   | 16  | 10  | 6   | 12  | 16  |
| CD47-A0026  | 166 | 140 | 118 | 66  | 82  | 102 | 113 | 107 | 146 | 57  | 125 |
| CD48-A0029  | 73  | 66  | 54  | 33  | 136 | 127 | 101 | 55  | 111 | 65  | 84  |
| CD40-A0031  | 2   | 1   | 4   | 8   | 3   | 5   | 3   | 2   | 3   | 5   | 5   |
| CD154-A0032 | 23  | 40  | 19  | 92  | 20  | 54  | 27  | 28  | 24  | 29  | 35  |
| CD52-A0033  | 55  | 33  | 18  | 31  | 18  | 40  | 33  | 34  | 33  | 18  | 37  |
| CD3-A0034   | 7   | 1   | 2   | 3   | 2   | 5   | 4   | 3   | 6   | 2   | 2   |
| CD8-A0046   | 2   | .   | 2   | 142 | .   | 14  | 64  | .   | 309 | 140 | 1   |
| CD56-A0047  | 6   | 5   | 2   | 8   | 9   | 2   | 5   | 5   | 5   | 10  | 6   |
| CD19-A0050  | 1   | .   | 1   | .   | .   | .   | .   | 1   | .   | .   | .   |
| CD33-A0052  | 4   | 4   | 1   | 1   | .   | 1   | 5   | 2   | 1   | 5   | 2   |

Noisy but broad

✓ Large number of genes

✗ High rate of false negative

Targeted but narrow

✓ Doesn't have dropout problem

✗ Limited number of antibodies

# Data Integration Strategy for Matched and Unmatched Data

Table 1 | Methods for matched data analysis

| Tool       | Data type | Model                                      | Additional notes   | Documentation   | Ref.          |
|------------|-----------|--|--|---|---------------|
| BREM-SC    | T + P     | Early integration, probabilistic modelling | This method models the observed data by multinomial distributions and assumes data from both modalities to be generated in a cluster-specific manner   | <a href="https://github.com/tarot0410/BREMSC">https://github.com/tarot0410/BREMSC</a>       | <sup>35</sup> |
| scAI       | T + C     | Early integration, latent space modelling  | scAI iteratively updates a regularized matrix factorization model to obtain an optimal common cell-loading matrix across two modalities  | <a href="https://github.com/sqjin/scAI">https://github.com/sqjin/scAI</a>                   | <sup>36</sup> |
| MOFA+      | T + C     | Early integration, latent space modelling  | MOFA and MOFA+ were built on the framework of group factor analysis but extend the model to enable the integration of different data types (count versus binary)   | <a href="https://github.com/bioFAM/MOFA2">https://github.com/bioFAM/MOFA2</a>               | <sup>38</sup> |
| TotalVI    | T + P     | Early integration, latent space modelling  | This method uses a variational autoencoder framework built on scVI. In this method, the protein measurements are modelled with a negative binomial mixture distribution to account for background reads                    | <a href="https://github.com/YosefLab/scVI-tools">https://github.com/YosefLab/scVI-tools</a> | <sup>39</sup> |
| CiteFuse   | T + P     | Late integration, latent space modelling   | The similarity measurement for protein data is based on a proportionality coefficient and the similarity measurement for RNA data is constructed with the Pearson correlation  | <a href="https://github.com/SydneyBioX/CiteFuse">https://github.com/SydneyBioX/CiteFuse</a> | <sup>42</sup> |
| Seurat 4.0 | T + P     | Late integration, latent space modelling   | Computes a weighted average cell affinity matrix from modality-specific affinity matrices. The weights are computed to reflect the predictive information within a cell's local neighbourhood defined within each modality | <a href="https://github.com/satijalab/seurat">https://github.com/satijalab/seurat</a>       | <sup>40</sup> |

BREM-SC, Bayesian random effects mixture model-single cell; C, chromatin accessibility; MOFA, multi-omics factor analysis; P, proteome; scAI, single-cell aggregation and integration; scVI, single-cell variational inference; T, transcriptome.

Table 2 | Methods for unmatched data analysis

| Strategy        | Tool        | Data type    | Feature matching | Algorithm  | Additional notes  | Documentation   | Ref.          |
|-----------------|-------------|--------------|------------------|--|---|---|---------------|
| Group matching  | Stereoscope | T + ST       | R                | Deconvolution  | This method assumes negative binomial distributions of genes and tolerates differential gene capture efficiencies between two technologies  | <a href="https://github.com/almaan/stereoscope">https://github.com/almaan/stereoscope</a>                                       | <sup>53</sup> |
|                 | MAESTRO     | T + C        | R                | CCA + MNN  | This method implements ChIP-seq data-based TF enrichment score calculators to define core TFs in each cell-type cluster   | <a href="https://github.com/liulab-dfcI/MAESTRO">https://github.com/liulab-dfcI/MAESTRO</a>                                     | <sup>49</sup> |
| Common features | STvEA       | MI + ET      | R                | MNN  | This method also provides a framework to transfer cell-type annotations from one modality to another  | <a href="https://github.com/CamaraLab/STvEA">https://github.com/CamaraLab/STvEA</a>   | <sup>54</sup> |
|                 | Clonealign  | T + D        | R                | Variational Bayes  | This method assumes correlation between DNA copy number and gene expression within the same region  | <a href="https://github.com/kieranrcampbell/clonealign">https://github.com/kieranrcampbell/clonealign</a>                       | <sup>56</sup> |
| Aligning spaces | Seurat 3.0  | T + C        | R                | CCA + SNN  | This method identifies anchor cells between datasets based on SNN across modalities; these anchor cells serve as a bridge for matching  | <a href="https://github.com/satijalab/seurat">https://github.com/satijalab/seurat</a>   | <sup>57</sup> |
|                 | LIGER       | T + M, T + C | R                | iNMF   | The relative contribution of dataset-specific factors and shared factors is determined by a hyperparameter $\lambda$ , which can be used to fine-tune the integration results                       | <a href="https://github.com/welch-lab/liger">https://github.com/welch-lab/liger</a>   | <sup>58</sup> |
| Aligning spaces | MAGAN       | MI + T       | R                | GAN  | This method identifies cell-to-cell correspondence by adding a loss function defined by similarity of cell matching; such loss function requires at least some shared features between two datasets | <a href="https://github.com/KrishnaswamyLab/MAGAN">https://github.com/KrishnaswamyLab/MAGAN</a>                                 | <sup>60</sup> |
|                 | MATCHER     | T + C        | NR               | Manifold alignment   | This method assumes 1D structure (pseudotime) with a pre-specified direction  | <a href="https://github.com/jw156605/MATCHER">https://github.com/jw156605/MATCHER</a>   | <sup>61</sup> |
| UnionCom        | MMD-MA      | T + M        | NR               | MMD  | In addition to the MMD loss, the loss function also has a distortion loss and a penalty to ensure the dimensionality and orthogonality of each projection   | <a href="https://bitbucket.org/noblelab/2019_mmd_wabi/src/master/">https://bitbucket.org/noblelab/2019_mmd_wabi/src/master/</a> | <sup>62</sup> |
|                 | UnionCom    | T + M        | NR               | GUMA   | The algorithm generalizes the GUMA method to achieve soft matching between datasets, enabling matching with different numbers of cells  | <a href="https://github.com/caokai1073/UnionCom">https://github.com/caokai1073/UnionCom</a>                                     | <sup>63</sup> |
| SCOT            | T + C       | NR           | GWOT             | A late integration method in which a similarity matrix is constructed by each modality separately, after which probabilistic transportation between datasets is achieved by GWOT | <a href="https://github.com/rsinghlab/SCOT">https://github.com/rsinghlab/SCOT</a>   |   | <sup>64</sup> |

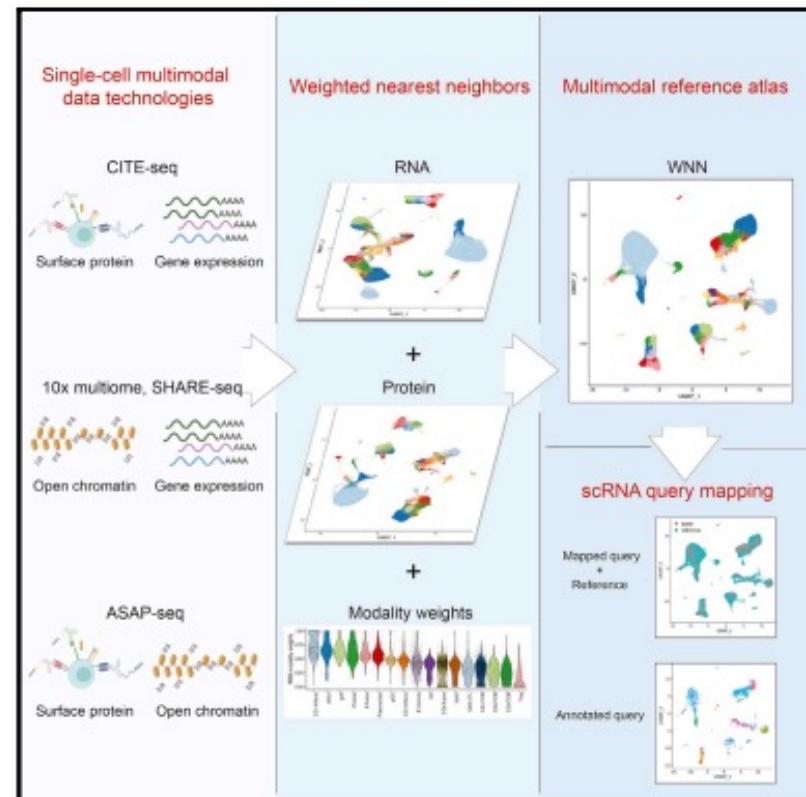
C, chromatin accessibility; CCA, canonical correlation analysis; ChIP-seq, chromatin immunoprecipitation followed by sequencing; D, DNA; ET, simultaneous epitope and transcriptome; GAN, generative adversarial networks; GUMA, generalized unsupervised manifold alignment; GWOT, Gromov–Wasserstein optimal transport; iNMF, integrative non-negative matrix factorization; M, methylome; MI, multiplexed immunohistochemistry; MMD, maximum mean discrepancy; MNN, mutual nearest neighbours; NR, not required; R, required; SNN, shared nearest neighbours; ST, spatial transcriptome; T, transcriptome; TF, transcription factor.

# Seurat 4.0

Cell

## Integrated analysis of multimodal single-cell data

### Graphical abstract



### Resource

#### Authors

Yuhan Hao, Stephanie Hao,  
Erica Andersen-Nissen, ....,  
Raphael Gottardo, Peter Smibert,  
Rahul Satija

#### Correspondence

rsatija@nygenome.org (R.S.),  
smibertp@gmail.com (P.S.)

#### In brief

A framework that allows for the integration of multiple data types using single cells is applied to understand distinct immune cell states, previously unidentified immune populations, and to interpret immune responses to vaccinations.

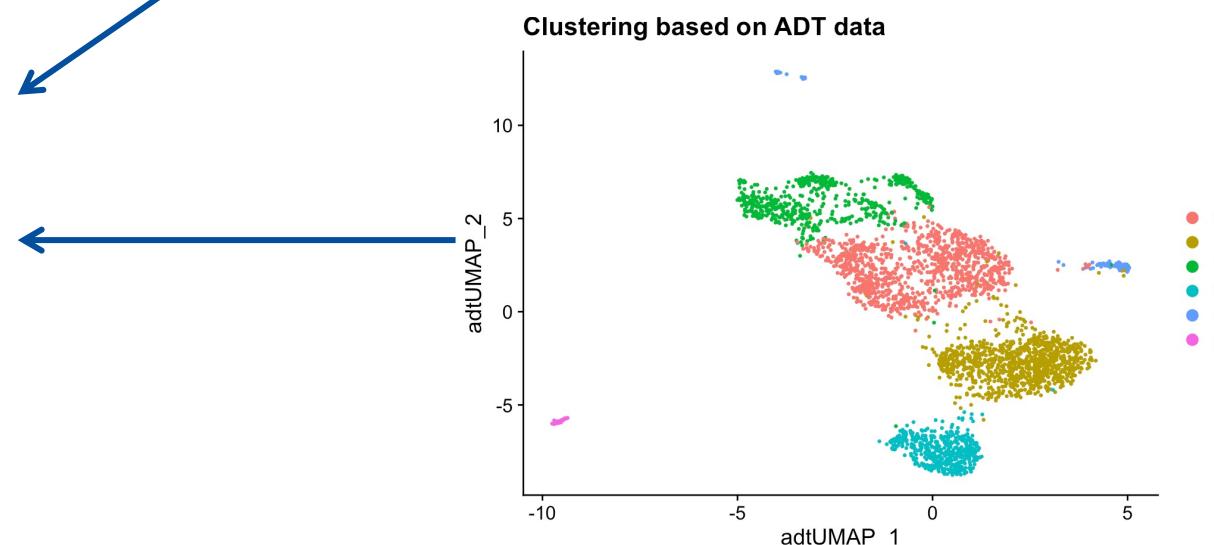
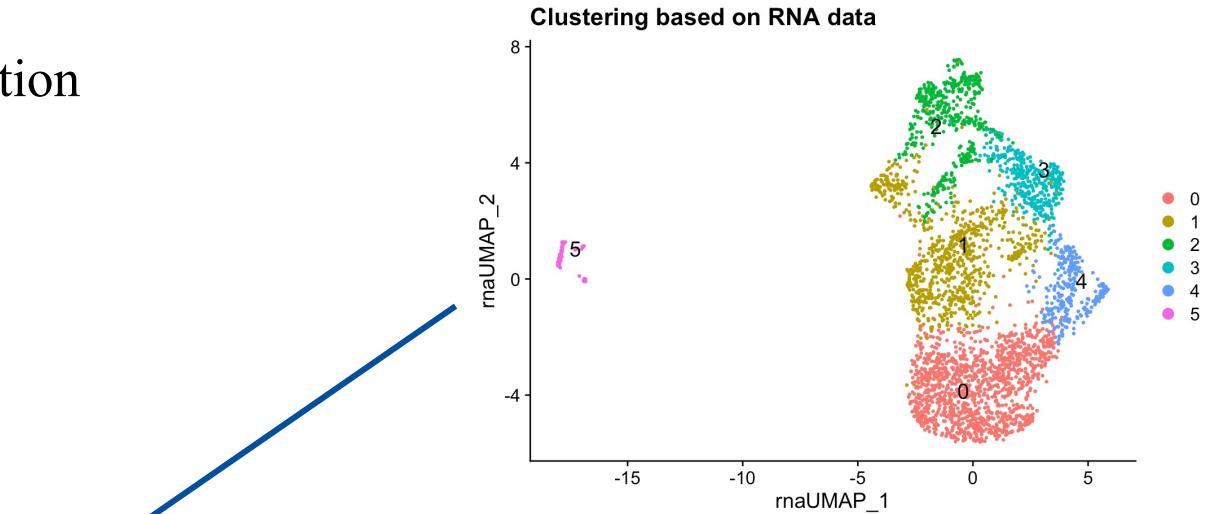
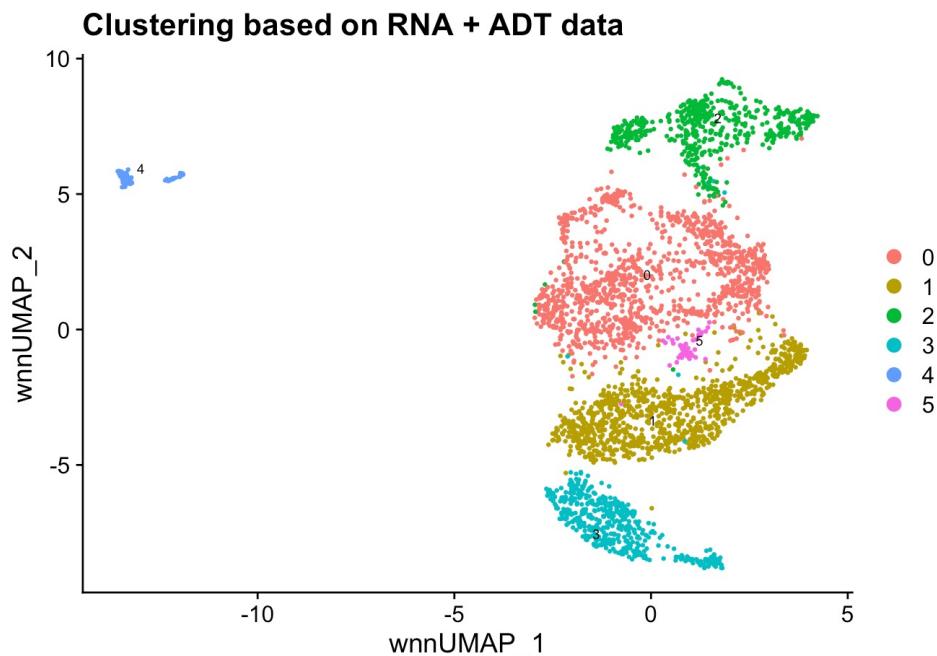
- Introduce to WNN
- Hands-on CITE-Seq analysis
- Omics vs. Multi-Omics analysis

### Weighted Nearest Neighbor Analysis

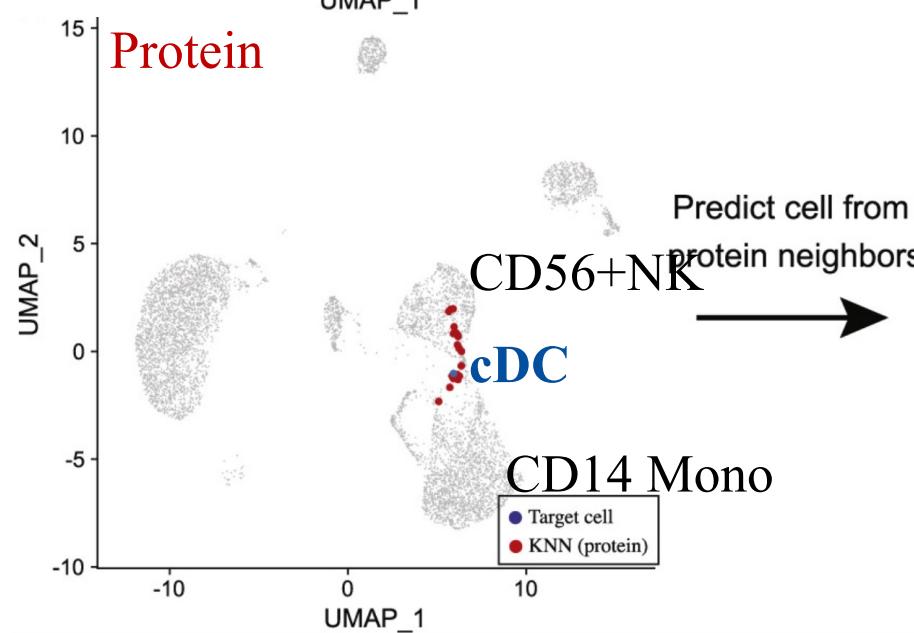
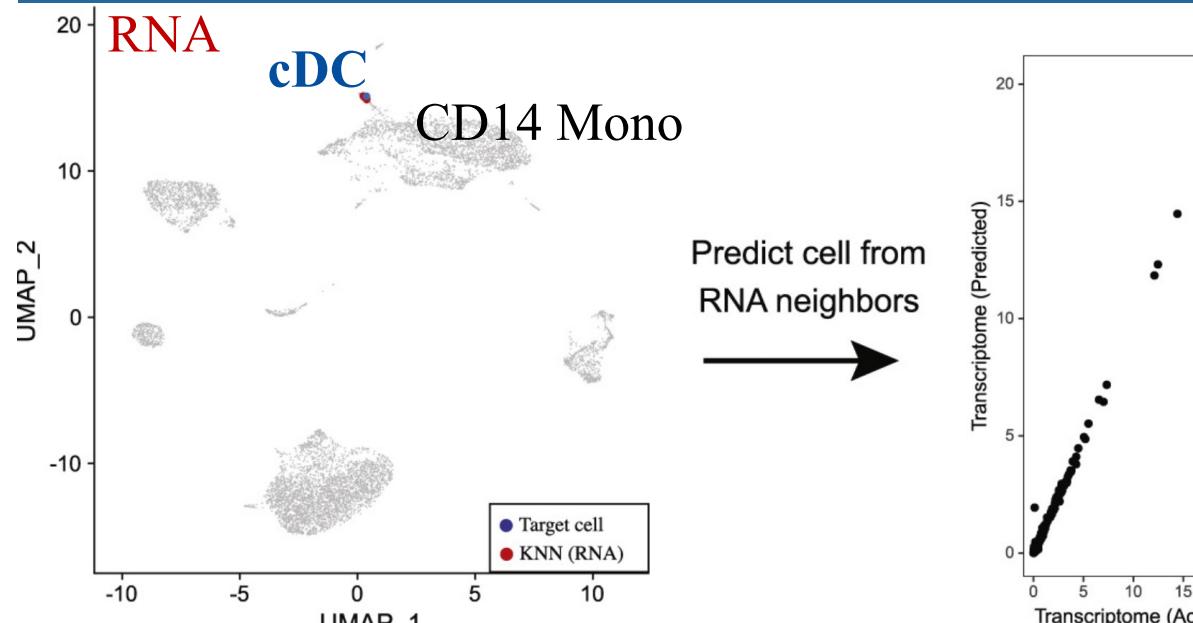
- Parallel integration analysis
- An unsupervised framework to learn the relative utility of each data type in each cell, enabling an integrative analysis of multiple modalities.

# Weighted Nearest Neighbor Analysis (WNN)

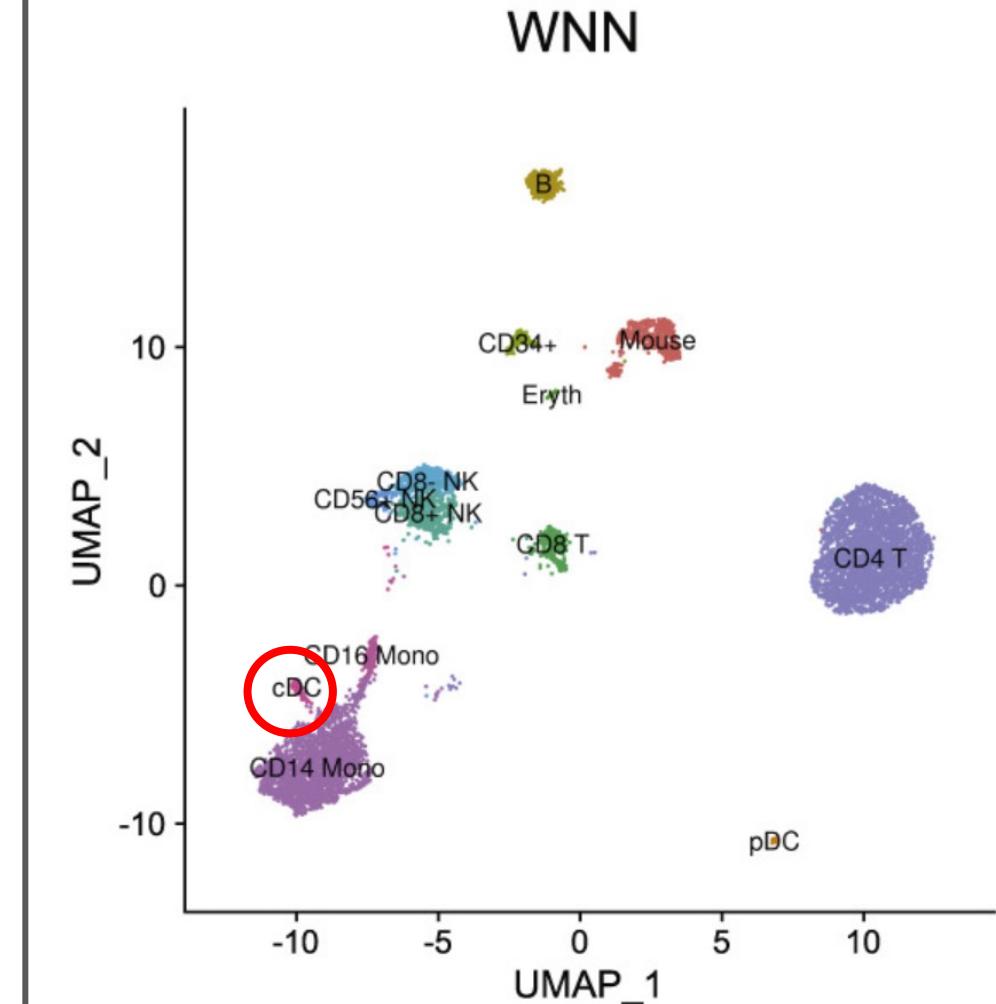
1. Constructing independent k-nearest neighbor (KNN) graphs for each modality
2. Performing within and across-modality prediction
3. Calculating cell-specific modality weights
4. Calculating a WNN graph



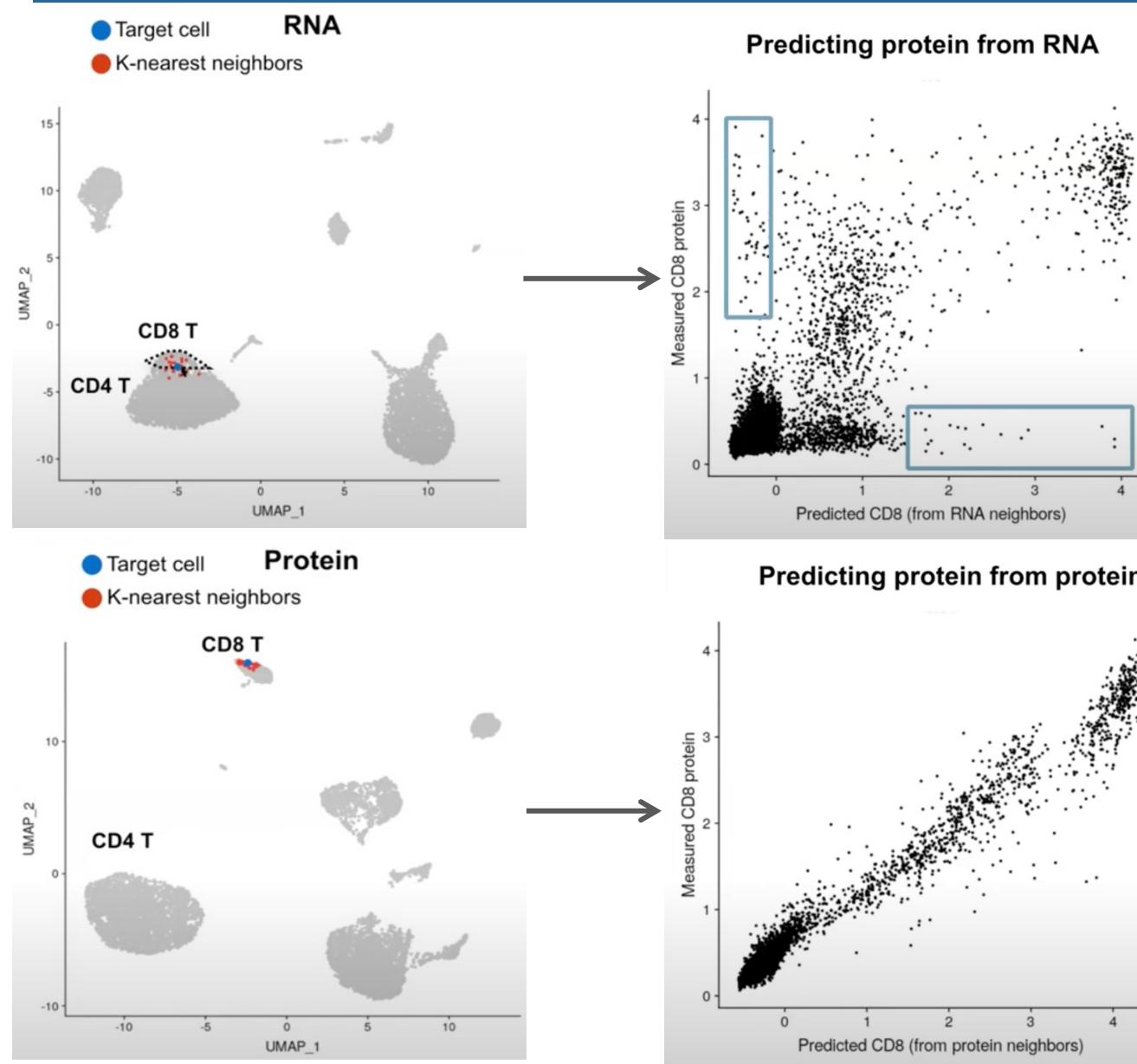
# Case I: Protein Is Worse Than RNA



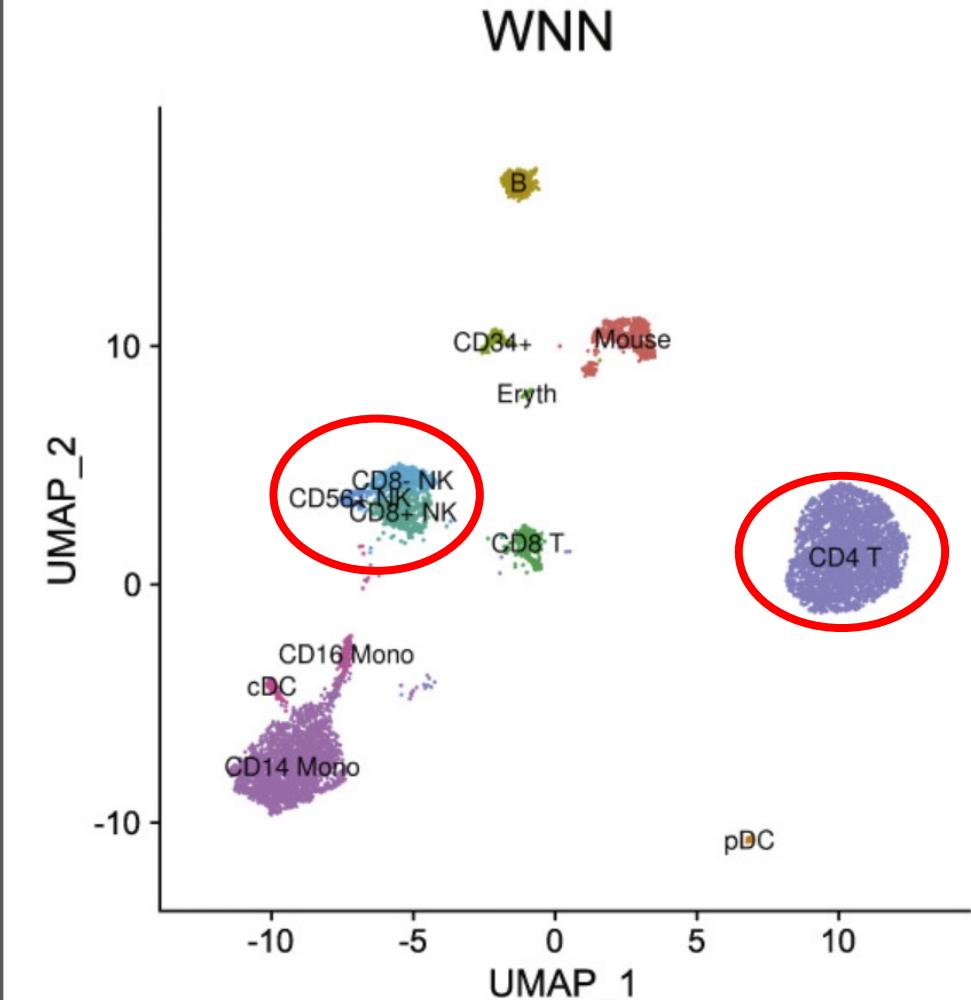
Target cell: cDC



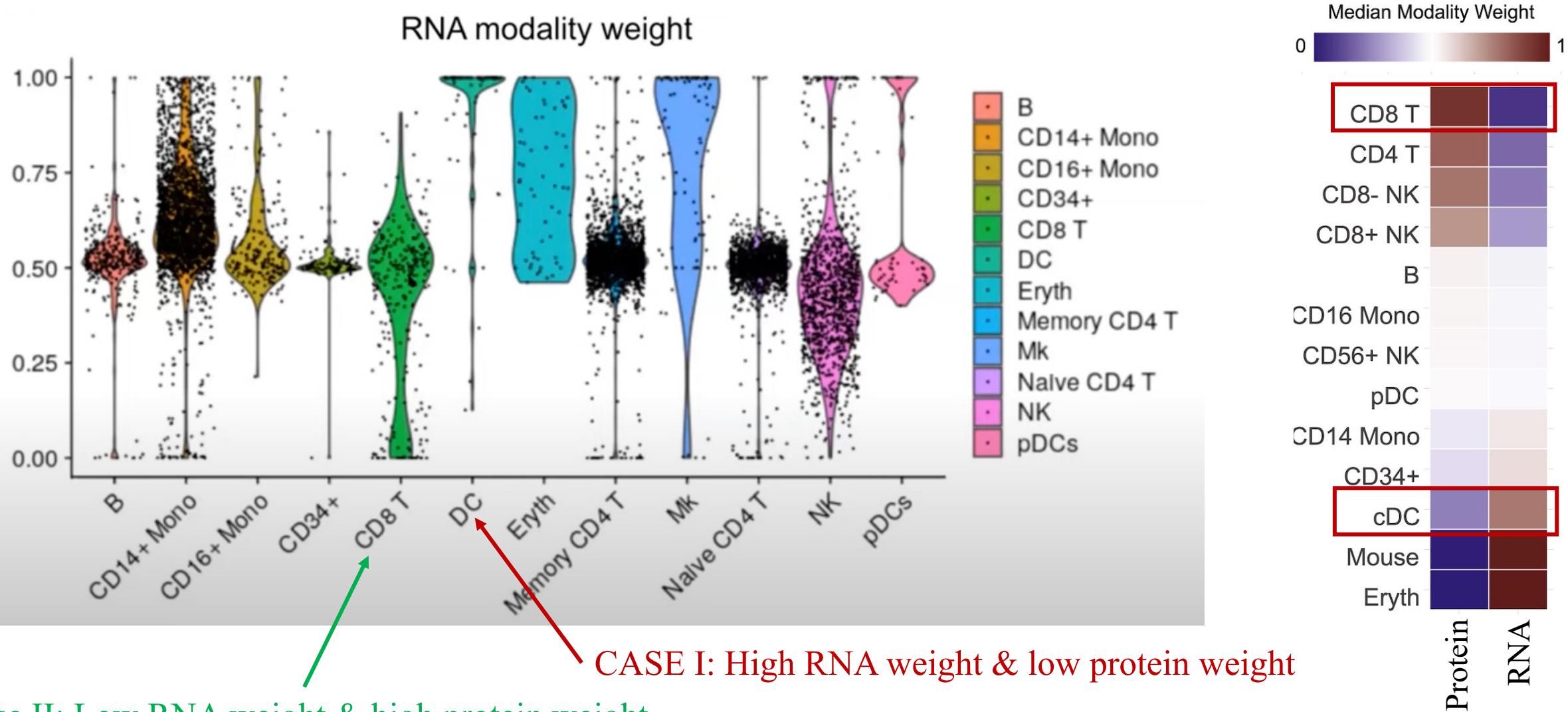
# Case II: Protein Is better Than RNA



Target cell: **CD8 T**



# Modality Weight From WNN



# References For Causal Mediation Analysis

- Bind et al. (2014). Air pollution and gene-specific methylation in the Normative Aging Study: association, effect modification, and mediation analysis. *Epigenetics*, 9(3), 448-458.
- Zhang et al. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics*, 32(20), 3150-3154.
- Liu et al. (2022). Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. *Journal of the American Statistical Association*, 117(537), 67-81.
- Dai et al. (2022). A multiple-testing procedure for high-dimensional mediation hypotheses. *Journal of the American Statistical Association*, 117(537), 198-213.
- Song et al. (2020). Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. *Biometrics*, 76(3), 700-710.
- Clark-Boucher et al. (2023). Methods for Mediation Analysis with High-Dimensional DNA Methylation Data: Possible Choices and Comparison. *medRxiv*, 2023-02.
- Huang et al. (2014). Joint analysis of SNP and gene expression data in genetic association studies of complex diseases. *The Annals of Applied Statistics* 8, 352–376.
- Huang and Pan (2016). Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*, 72(2), 402-413.
- Loh et al. (2022). Nonlinear mediation analysis with high-dimensional mediators whose causal structure is unknown. *Biometrics*, 78(1), 46-59.
- Liu et al. (2022). Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. *Journal of the American Statistical Association*, 117(537), 67-81.

# References For Causal Mediation Analysis

- Inoue et al. (2020). Air pollution and adverse pregnancy and birth outcomes: mediation analysis using metabolomic profiles. *Current environmental health reports*, 7, 231-242.
- Chén et al. (2018). High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, 19(2), 121-136.
- Zhao et al. (2020). Sparse principal component based high-dimensional mediation analysis. *Computational statistics & data analysis*, 142, 106835.
- Fairchild and MacKinnon (2009). A general model for testing mediation and moderation effects. *Prevention science*, 10, 87-99.
- Baron and Kenny (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173.
- VanderWeele and Vansteelandt (2014). Mediation analysis with multiple mediators. *Epidemiologic methods*, 2(1), 95-115.
- Zeng et al. (2021). Statistical methods for mediation analysis in the era of high-throughput genomics: current successes and future challenges. *Computational and structural biotechnology journal*, 19, 3209-3224.
- Dugourd et al. (2021). Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *Molecular systems biology*, 17(1), e9730.
- Qin et al. (2019). Identifying multi-omics causers and causal pathways for complex traits. *Frontiers in genetics*, 10, 110.
- Kelly et al. (2022). A review of causal discovery methods for molecular network analysis. *Molecular Genetics & Genomic Medicine*, 10(10), e2055.

# References For Single-Cell Multi-Omics

- Collora et al. (2022). Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones. *Immunity*, 55(6), 1013-1031.
- Pan and Jia (2021). Application of single-cell multi-omics in dissecting cancer cell plasticity and tumor heterogeneity. *Frontiers in Molecular Biosciences*, 8, 757024.
- Liu et al. (2019). Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nature communications*, 10(1), 470.
- Zhu et al. (2020). Single-cell multimodal omics: the power of many. *Nature methods*, 17(1), 11-14.
- Miao et al. (2021). Multi-omics integration in the age of million single-cell data. *Nature Reviews Nephrology*, 17(11), 710-724.
- Hao et al. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13), 3573-3587.
- Adossa et al. (2021). Computational strategies for single-cell multi-omics integration. *Computational and Structural Biotechnology Journal*, 19, 2588-2596.
- Ma et al. (2020). Integrative methods and practical challenges for single-cell multi-omics. *Trends in biotechnology*, 38(9), 1007-1022.