Computational Approaches to Dissect Admixed Transcriptome Data

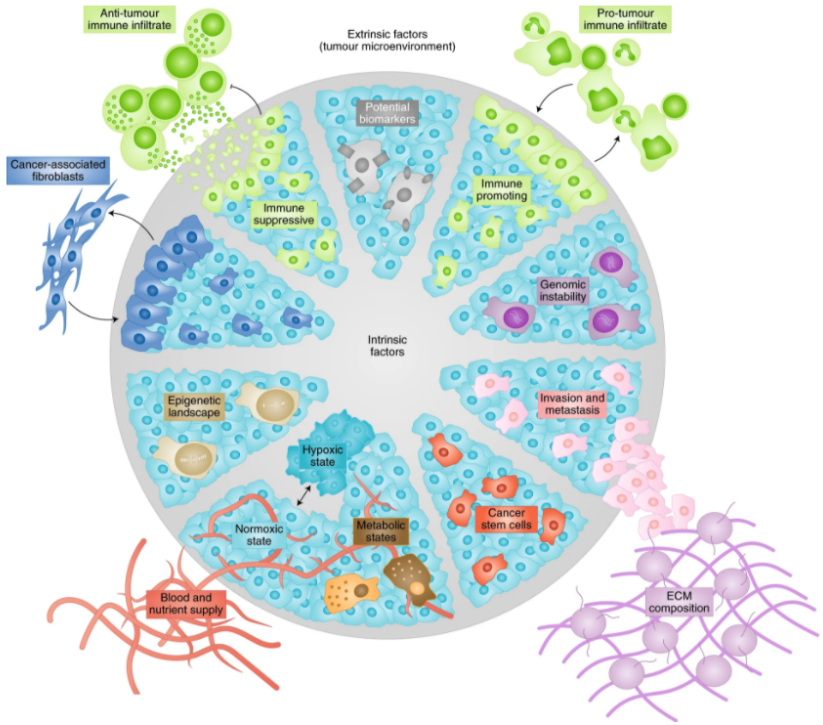
Hao Feng

Assistant Professor
Department of Population and Quantitative Health Sciences
Case Western Reserve University

hxf155@case.edu
https://hfenglab.org

Heterogeneous mixture

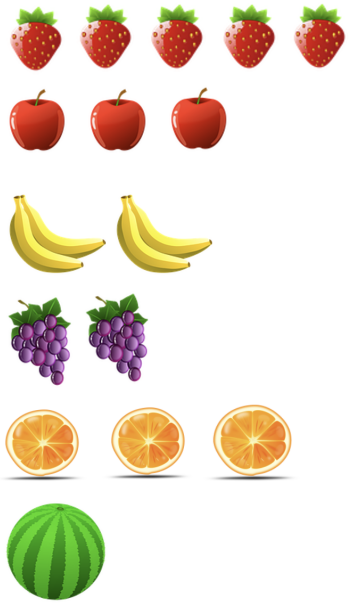
- Human tissues have diverse cell types/states.
- Traditional RNA-seq ("bulk" RNA-seq) can measure **averaged signal** across millions of cells.
- Single-cell RNA-seq (scRNA-seq) give us the first data-driven approach to study the **heterogeneous tissue** at single-cell level.



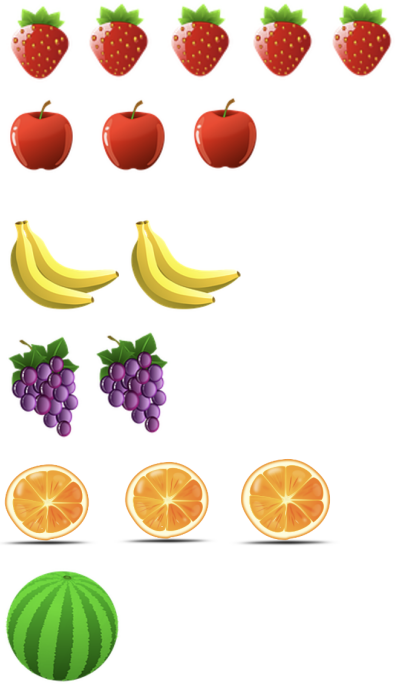
Lawson et al. Nature. https://www.nature.com/articles/s41556-018-0236-7

Bulk RNA-seq

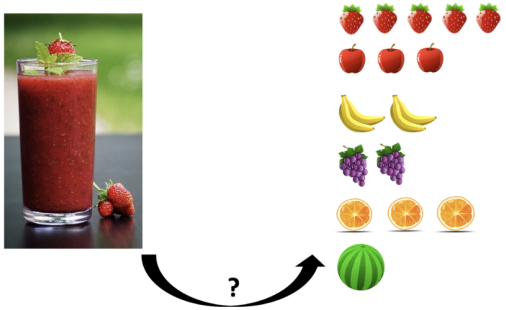
Single-cell RNA-seq



?



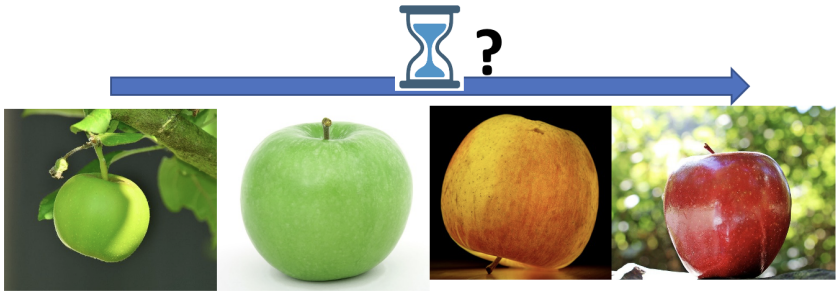
Deconvolution and beyond



Smoothie A use:

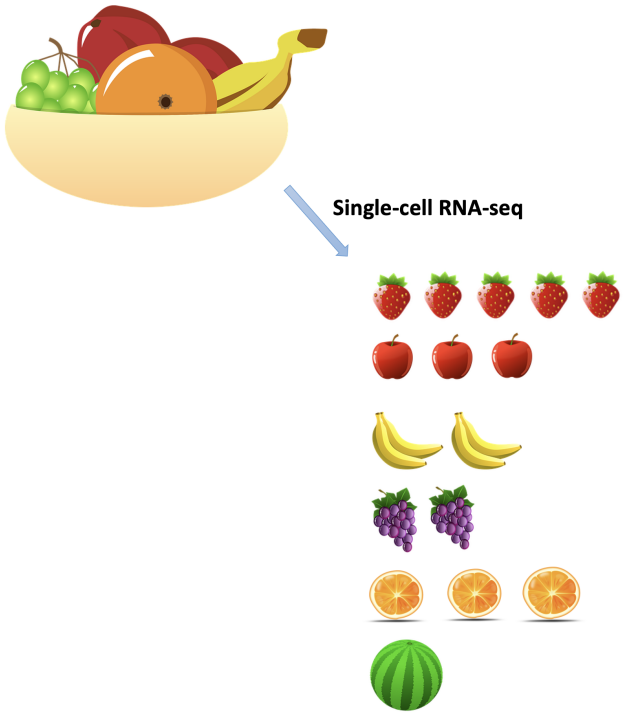
?

Smoothie B use:



1. A hybrid neural network model for single-cell RNA-seq cell type prediction.
2. A statistical method to recover individual-specific and cell-type-specific transcriptome reference panels.

1. A hybrid neural network model for single-cell RNA-seq cell type prediction.



scRNA-seq data

```
> Baron_counts[1:20,1:9]
        human2_lib1.final_cell_0001 human2_lib1.final_cell_0002 human2_lib1.final_cell_0003 human2_lib1.final_cell_0004 human2_lib1.final_cell_0005 human2_lib1.final_cell_0006
SST                            3476                        3340                           0                        2962                        3367                        3008
INS                              24                           5                           6                           6                           6                           7
GCG                               8                          11                        1995                           8                           4                          10
REG1A                             1                           0                           0                           0                           0                           0
PPY                               1                           0                           0                           0                           1                           2
TTR                              10                           1                         273                          13                           4                           1
IAPP                              3                           0                           2                           3                           1                           0
REG3A                             0                           2                           0                           0                           0                           0
PRSS2                             0                           0                           0                           0                           0                           0
CTRB2                             0                           0                           0                           0                           0                           0
REG1B                             0                           0                           0                           0                           0                           0
SPINK1                            0                           0                           3                           0                           0                           0
SERPINA1                         89                          26                         362                          20                           5                          13
SERPINA3                          0                           0                           0                           0                           0                           0
EEF1A1                           44                          21                          66                          52                          43                          30
OLFM4                             0                           0                           0                           1                           0                           0
GNAS                             71                          15                         103                          90                          67                          49
FTL                              14                           5                          24                           5                          12                           3
CTRB1                             0                           0                           0                           0                           0                           0
TMSB4X                            7                           5                           5                           6                           4                           5
```

scRNA-seq data analysis questions

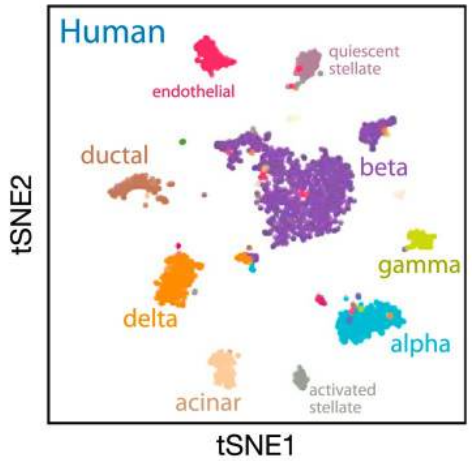
- **Data preprocessing**
  - Normalization
  - Batch effect correction
  - Imputation
- **Data analyses**
  - Cell clustering
  - Cell type identification
  - Differential expression
  - Pseudo-time construction
  - Rare cell type discovery;
  - alternative splicing; allele specific expression
  - RNA velocity
- **Visualization**
  - *t*-SNE, UMAP

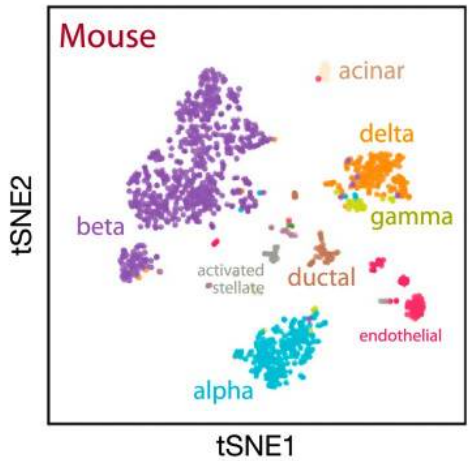
scRNA-seq data analysis questions

- **Data preprocessing**
  - Normalization
  - Batch effect correction
  - Imputation
- **Data analyses**
  - Cell clustering
  - **Cell type identification**
  - Differential expression
  - Pseudo-time construction
  - Rare cell type discovery;
  - alternative splicing; allele specific expression
  - RNA velocity
- **Visualization**
  - *t*-SNE, UMAP

Cell type identification

- Sequencing output of scRNA-seq is anonymous in terms of cell identities.
- Annotating the cells is a key task in scRNA-seq data analysis.





Baron et al. Cell Systems. doi: 10.1016/j.cels.2016.08.011

Cell type identification methods

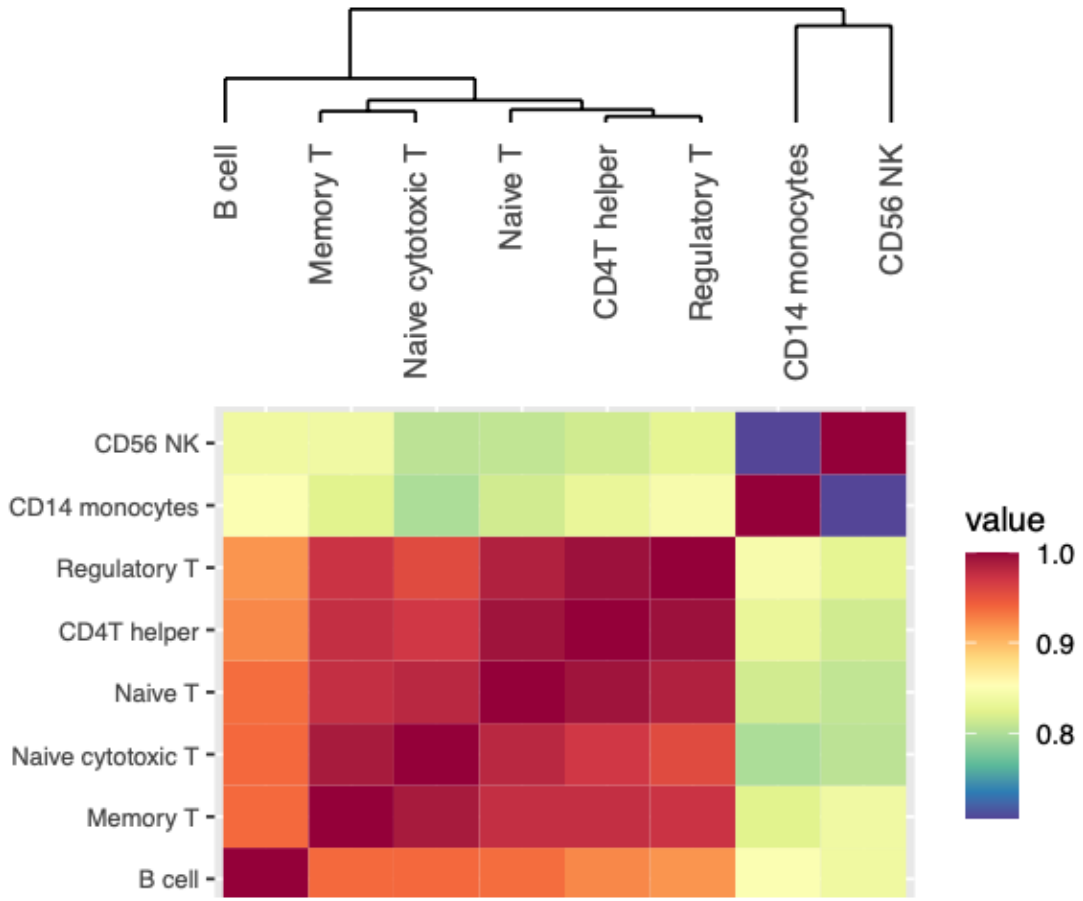
- Two-step approach: Clustering (unsupervised) + labeling.
  - Seurat, SC3, TSCAN, etc...
  - Laborious, time consuming, not best projection, rely on marker gene heavily.
- One-step approach: supervised labeling.
  - scmap, CHETAH, CellAssign, etc...
  - (1) marker-based, (2) correlation-based, and (3) tree structure based.
  - Not suitable for novel cell type discovery.

Our proposed method

NeuCA: a **Neu**ral network-based **C**ell **A**nnotation

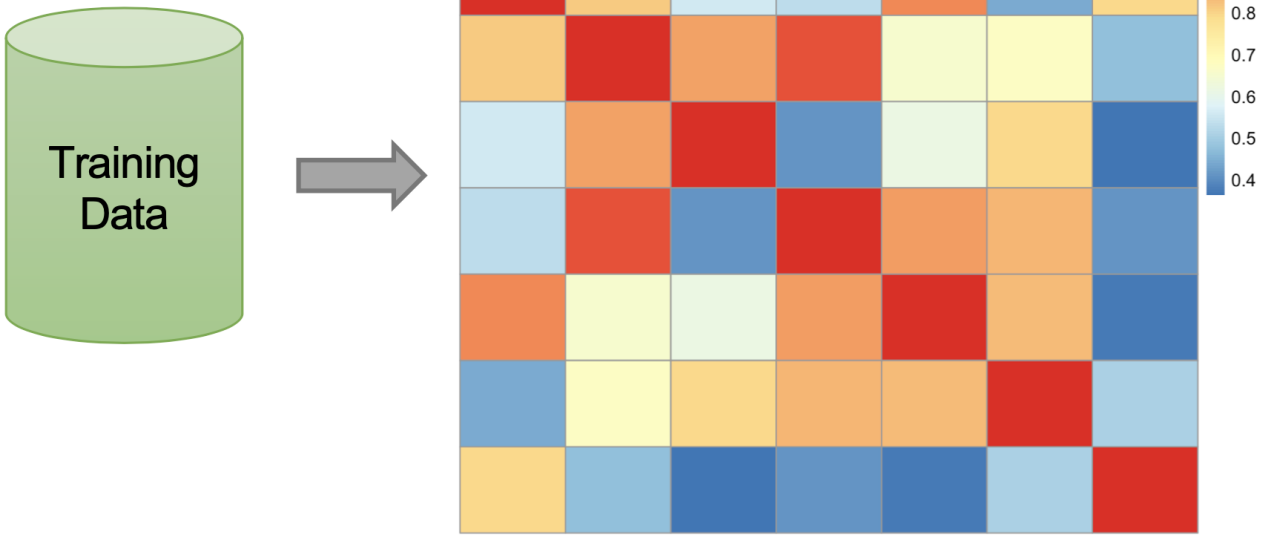
- Exhaustive: Annotate all cells.
- Reliable: Leverage on massive amount of well-studied existing scRNA data.
- Flexible: Adopt different prediction strageties, depending on correlation.

Correlations of cell types



High-correlation cell types pose major challenges.

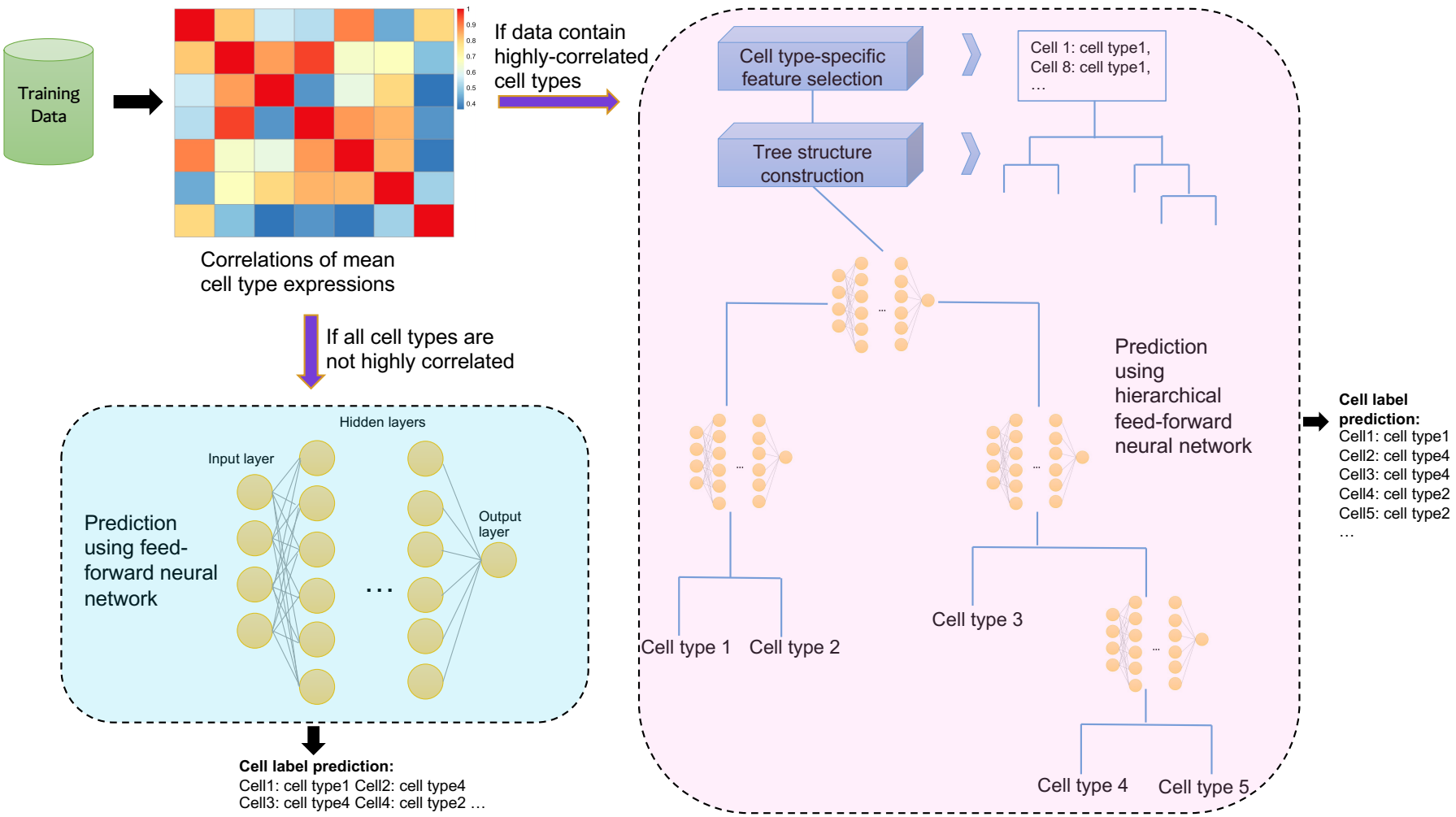
Tackle the issue of high-correlation



Correlations of mean cell type expressions

How about adopting a flexible approach?

Method: NeuCA



Wires inside the cogs

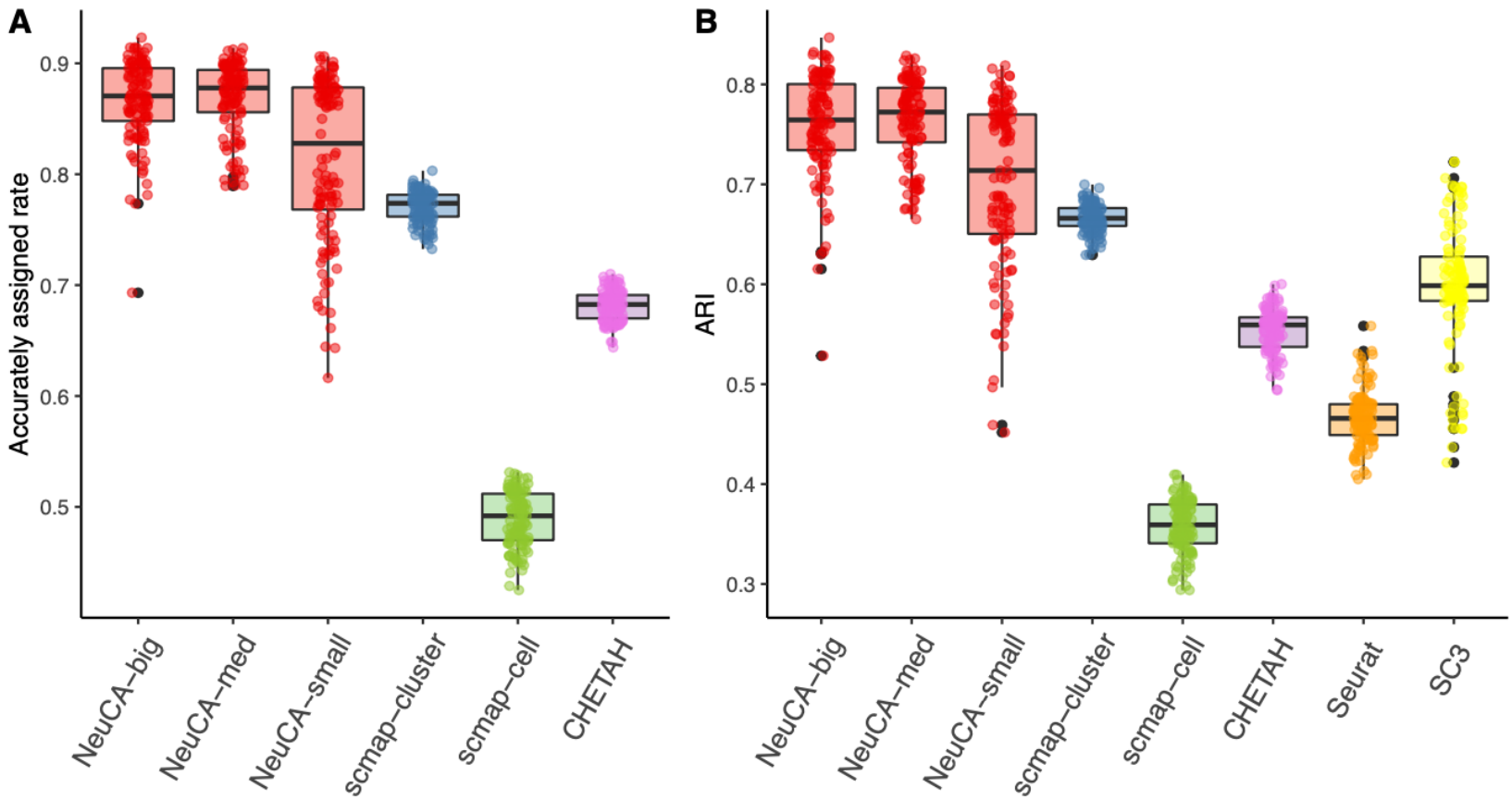
- Route 1: feed-forward neural-network
  - feature selection.
  - 3 model sizes: big, medium, small. (256 to 64 units/nodes)
  - activation function: Rectified Linear Unit (ReLU).
  - output: Softmax.
  - categorical cross-entropy loss.
- Route 2: marker-guided hierarchical neural-network
  - feature selection (gene-specific sensitivities).
  - cell type hierarchical tree.
  - hierarchical neural-network tree.
  - similar model sizes as Route 1.

Simulation

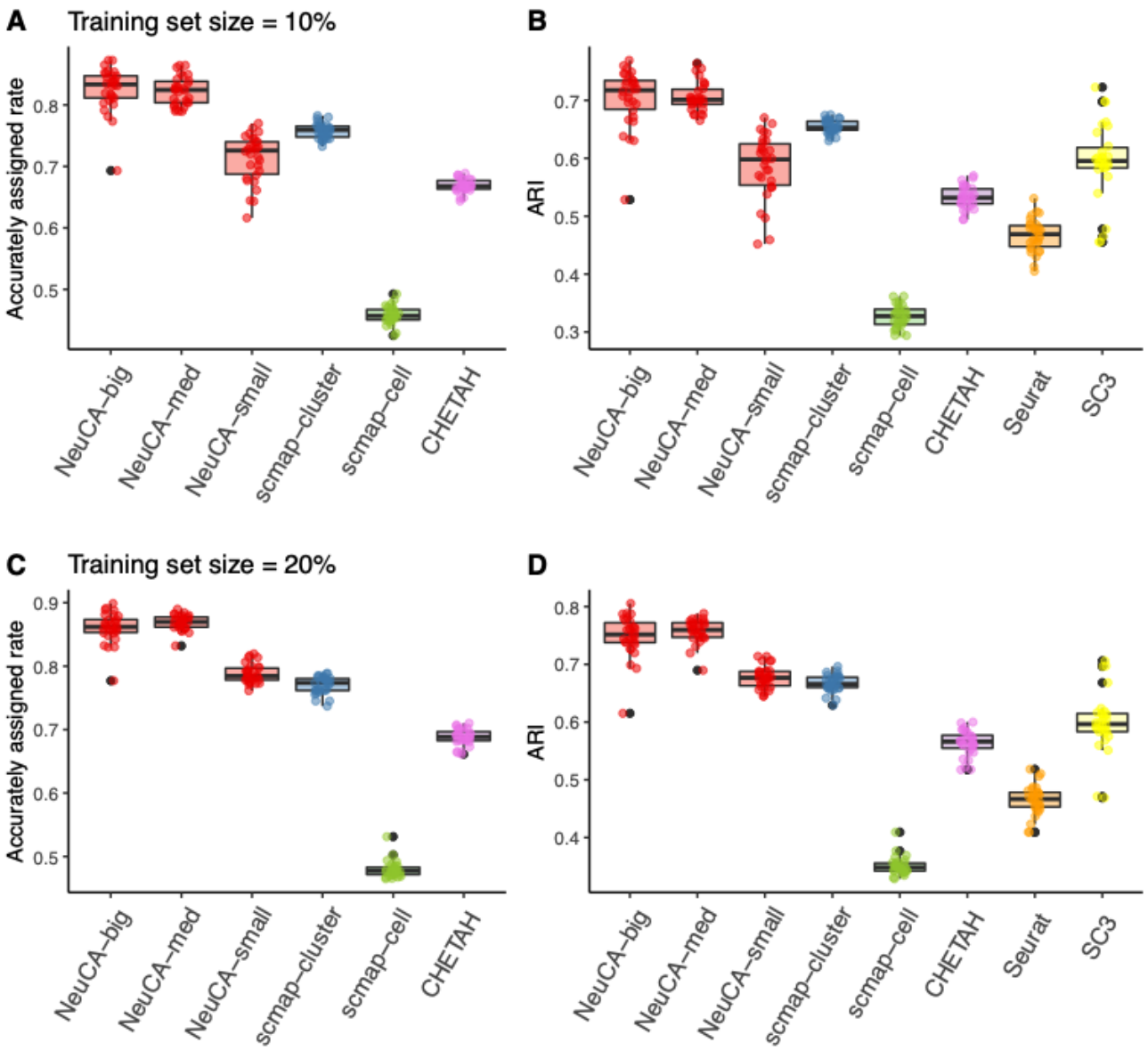
Real-data based simulation

- 10X PBMC scRNA-seq data.
- 80 Monte Carlo simulations are conducted and aggregated.
- Training set proportion ranging from 10% to 80%.

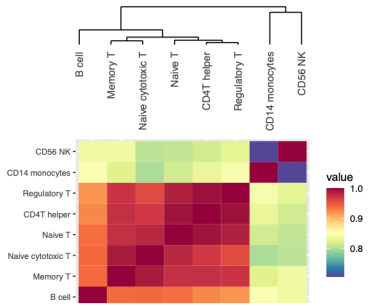
Simulation results: real-data based



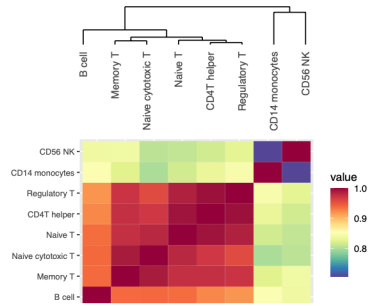
Real-data based simulation

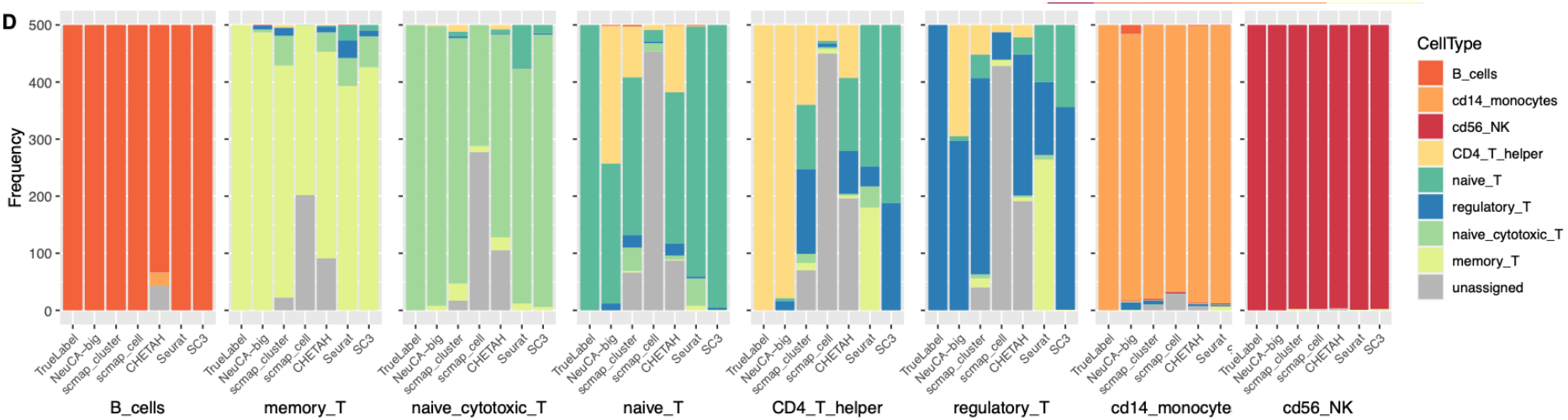


Real-data based simulation



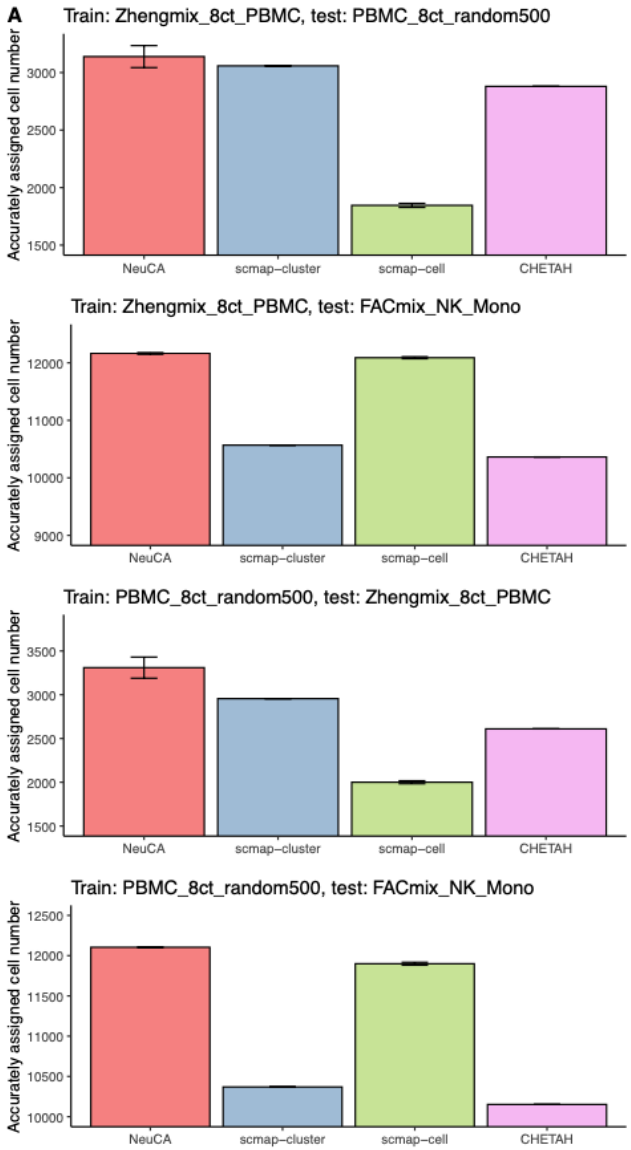
Real-data based simulation



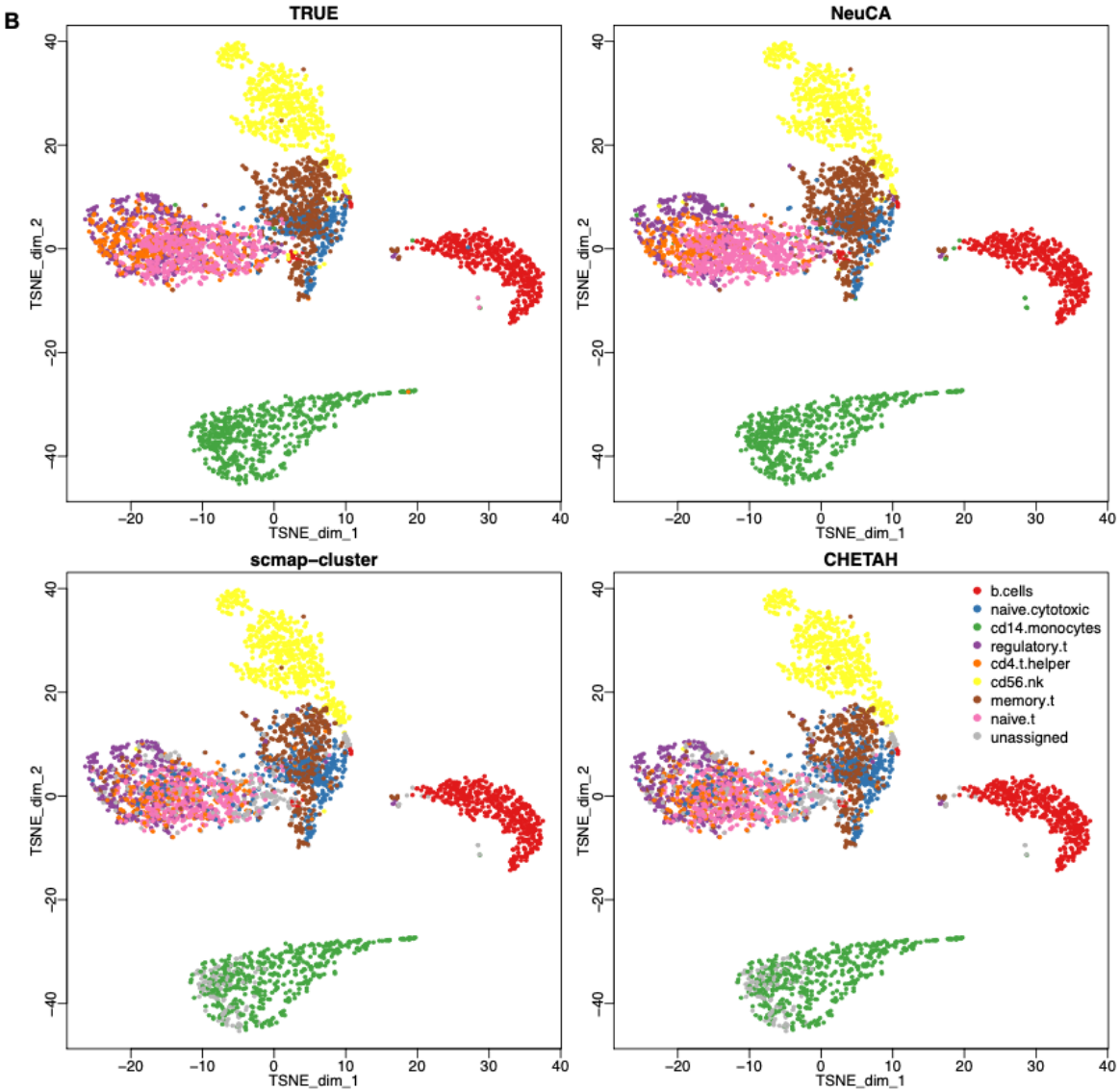


Real data results

Various PBMC datasets



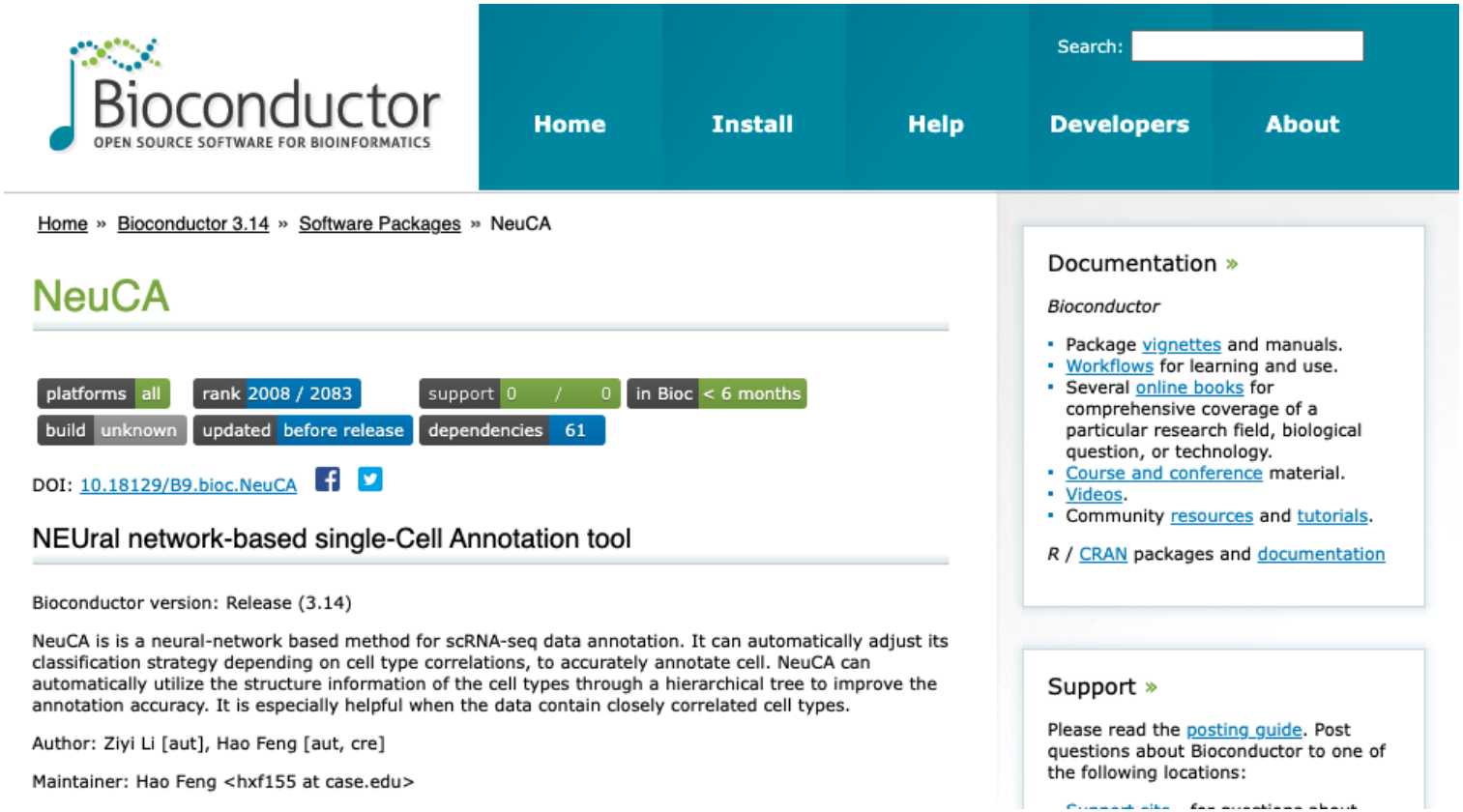
Real data results: a t-SNE visualization



Real data results: pancreas data

Software

On Bioconductor: https://bioconductor.org/packages/NeuCA



Software

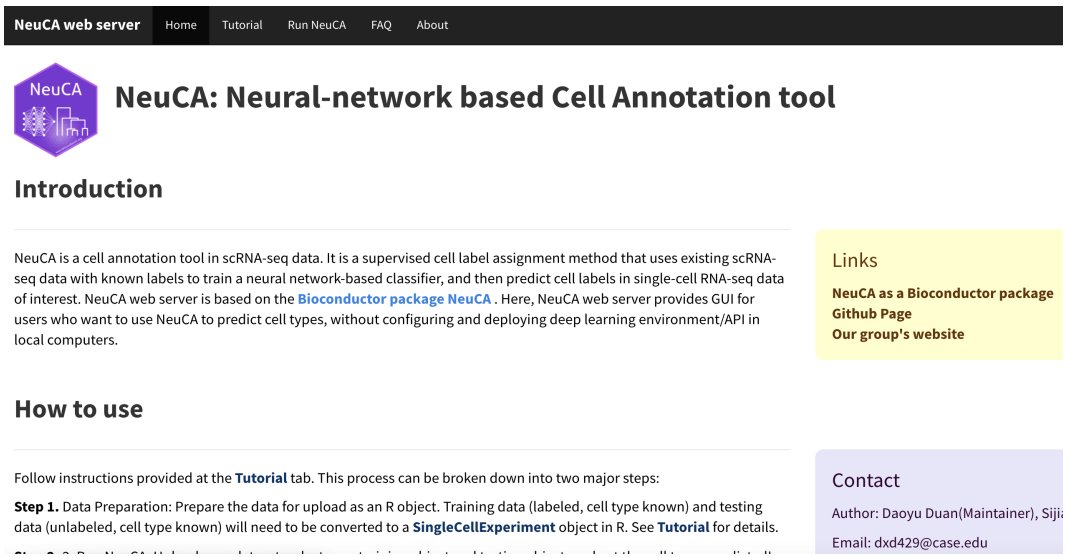
On Bioconductor: https://bioconductor.org/packages/NeuCA

Usage

```
NeuCA(train, test, model.size = "big", verbose = FALSE)
```

NeuCA web server

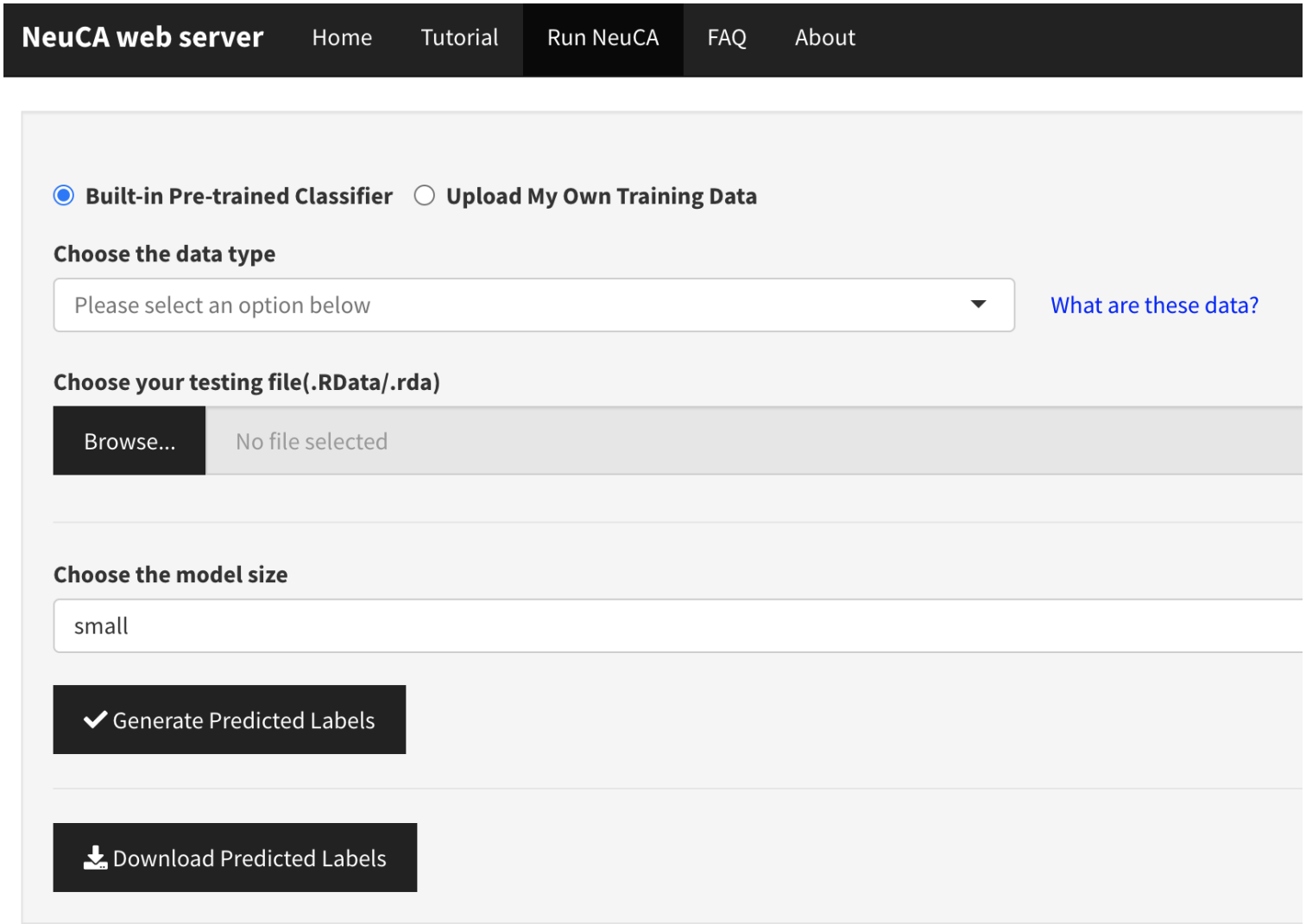
R Shiny App: https://statbioinfo.shinyapps.io/NeuCA



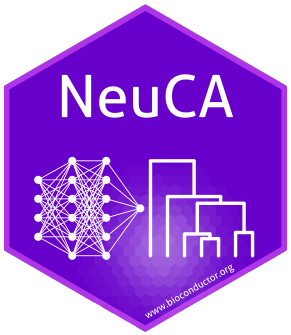
(Human) Molar
(Human) Choroid Plexus
(Human) Healthy Lung
(Human) Aging Skin
(Human) Fetal Maternal Decidual
(Human) Muscle
(Human) Bronchoalveolar from COVID-19 Patients
(Human) Adult Retina
(Human) Fetal Gut
(Mouse) Enteric
(Mouse) Hippocampus
(Mouse) Spinal Cord

NeuCA web server

R Shiny App: https://statbioinfo.shinyapps.io/NeuCA

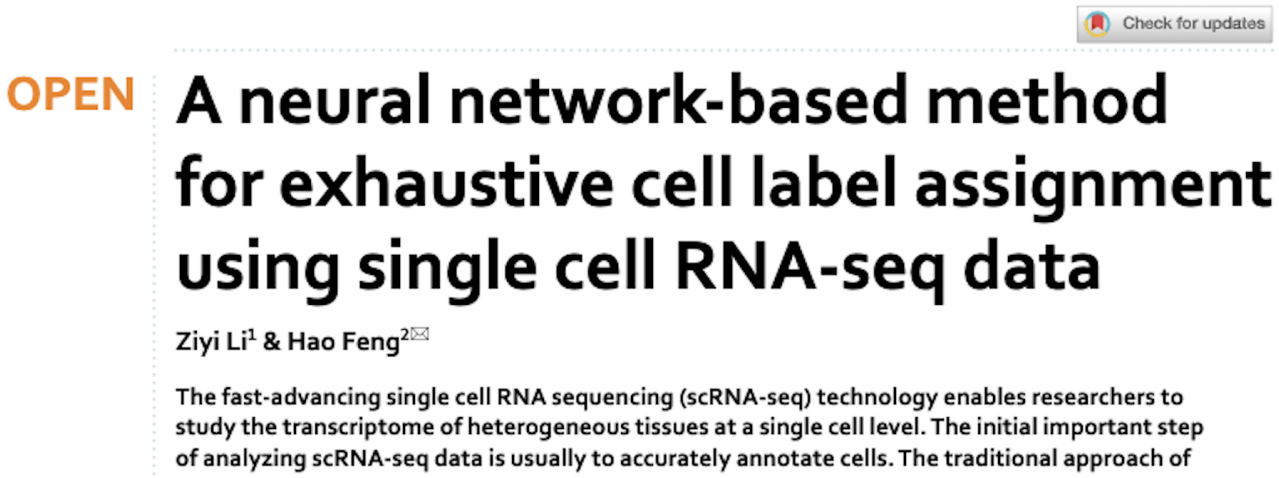


NeuCA summary



- One-step supervised learning method for cell label assignment.
- A neural-network based classifier.
- Flexible: adopt different approaches depending on correlation level.
- Perform well even with low amount of training set.
- High accuracy.
- R/BioC package and web server with GUI (free!).

Publication



Check for updates

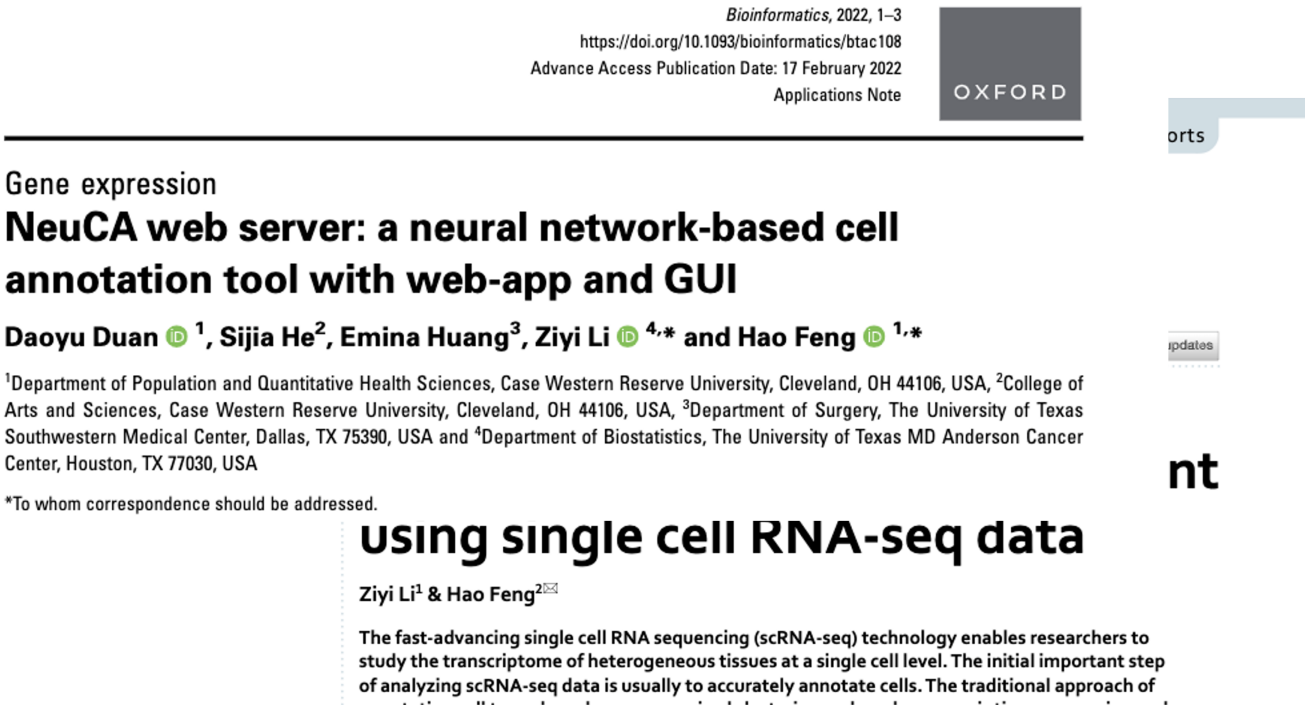
OPEN

A neural network-based method for exhaustive cell label assignment using single cell RNA-seq data

Ziyi Li[1] & Hao Feng[2]✉

The fast-advancing single cell RNA sequencing (scRNA-seq) technology enables researchers to study the transcriptome of heterogeneous tissues at a single cell level. The initial important step of analyzing scRNA-seq data is usually to accurately annotate cells. The traditional approach of

Publication



Bioinformatics, 2022, 1–3
https://doi.org/10.1093/bioinformatics/btac108
Advance Access Publication Date: 17 February 2022
Applications Note

OXFORD

Gene expression

NeuCA web server: a neural network-based cell annotation tool with web-app and GUI

Daoyu Duan [1], Sijia He[2], Emina Huang[3], Ziyi Li [4,*] and Hao Feng [1,*]

[1]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106, USA, [2]College of Arts and Sciences, Case Western Reserve University, Cleveland, OH 44106, USA, [3]Department of Surgery, The University of Texas Southwestern Medical Center, Dallas, TX 75390, USA and [4]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

*To whom correspondence should be addressed.

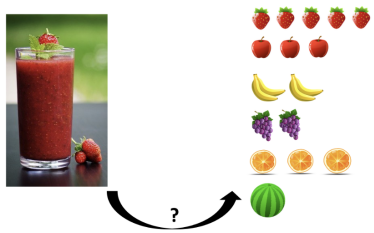
using single cell RNA-seq data

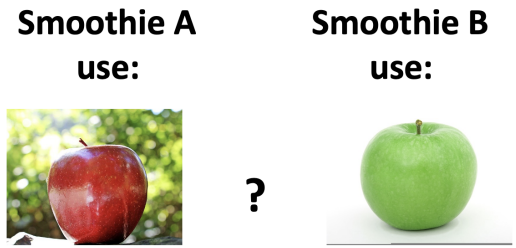
Ziyi Li[1] & Hao Feng[2]✉

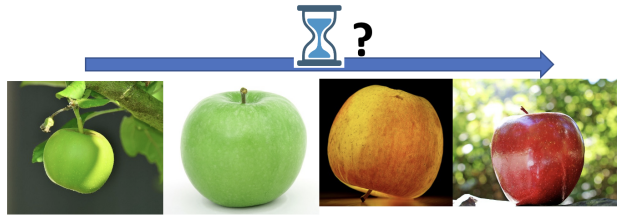
The fast-advancing single cell RNA sequencing (scRNA-seq) technology enables researchers to study the transcriptome of heterogeneous tissues at a single cell level. The initial important step of analyzing scRNA-seq data is usually to accurately annotate cells. The traditional approach of

doi.org/10.1093/bioinformatics/btac108
doi.org/10.1038/s41598-021-04473-4

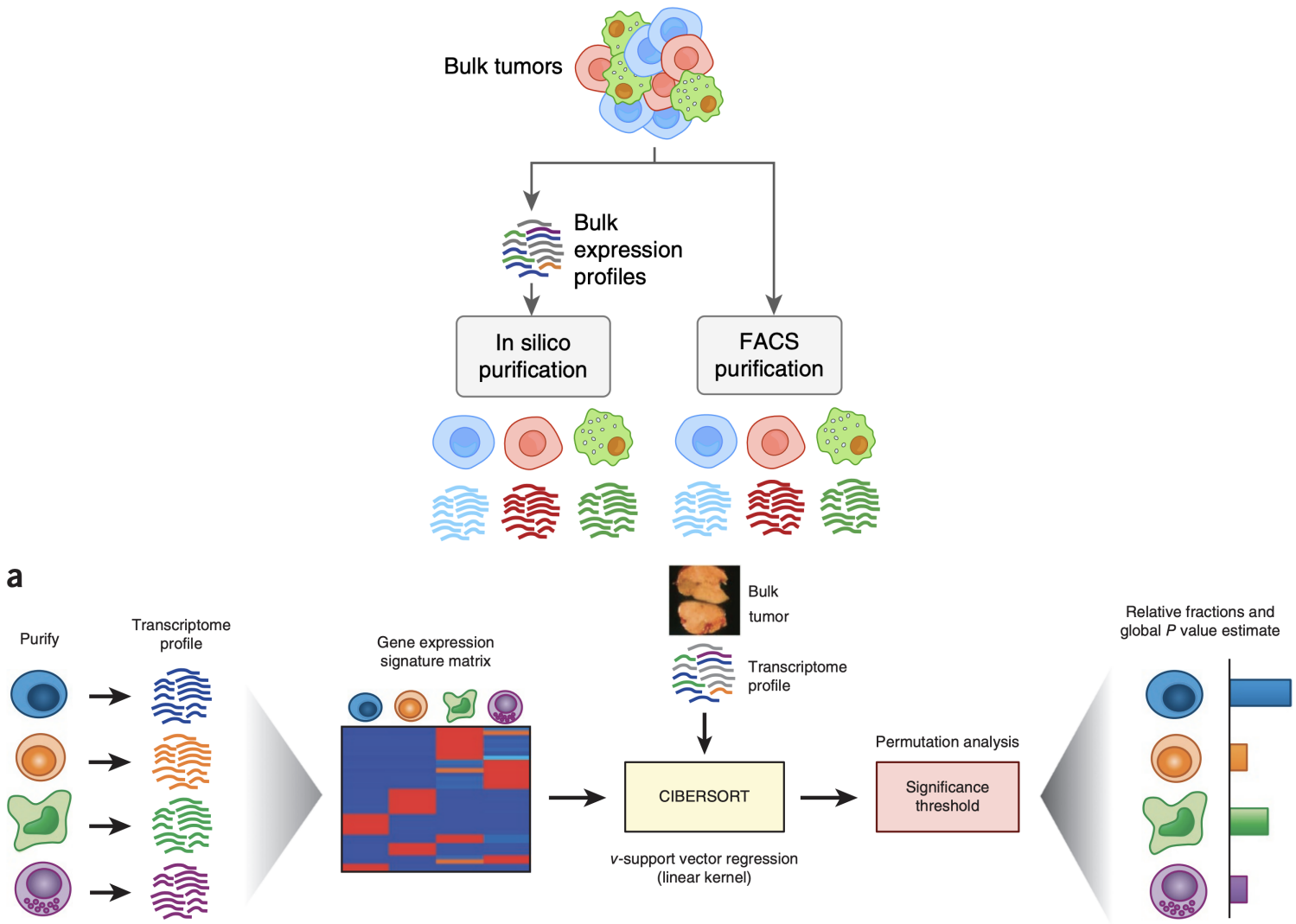
2. A statistical method to recover individual-specific and cell-type-specific transcriptome reference panels.





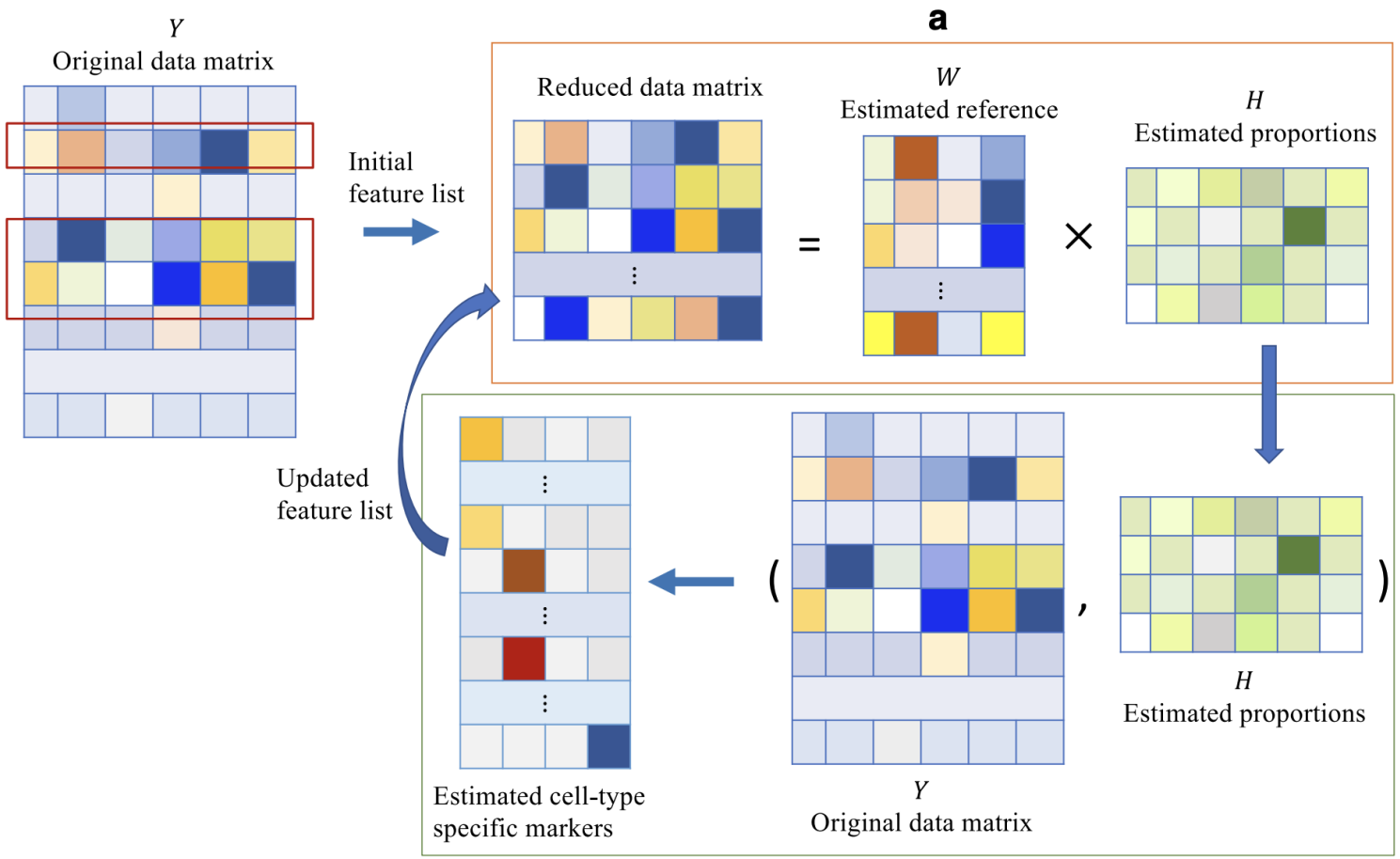


Cell composition of complex tissues



Newman et al. Nat Biotechnol. 2019; Newman et al. Nat Methods. 2015

Cell composition of complex tissues

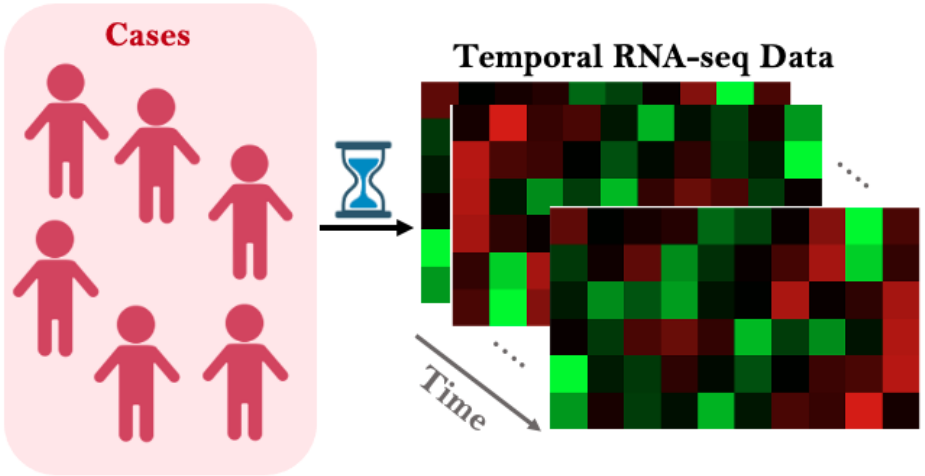


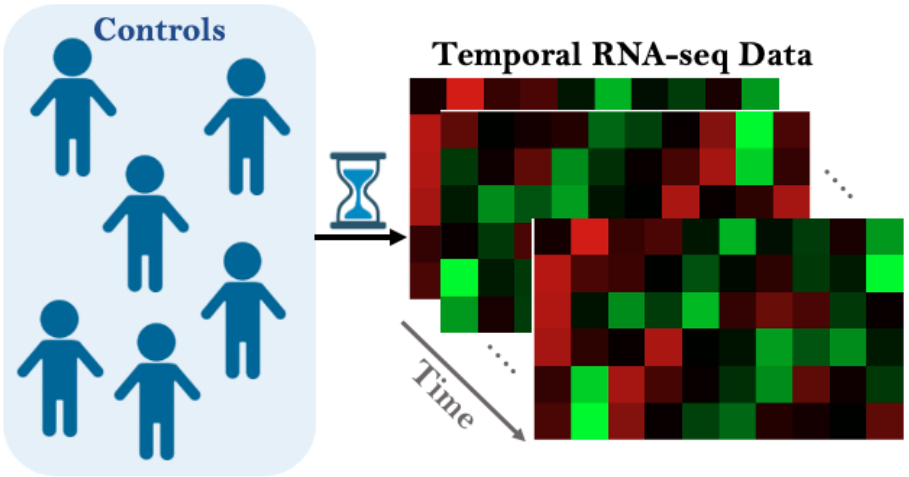
Li et al. Genome Biology 2019

Existing problems

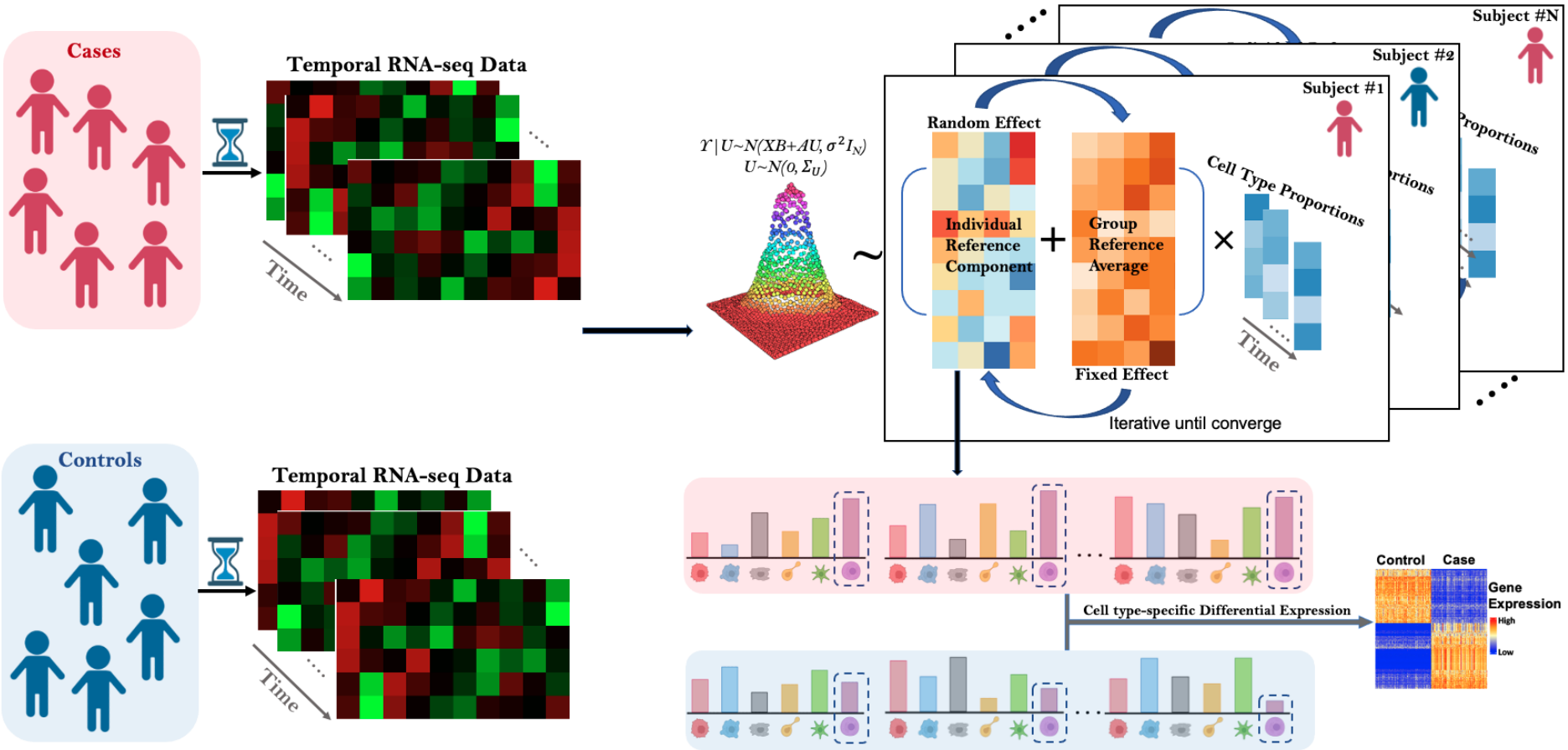
- Lack of individual-level reference profile.
- Repeatedly measured samples over the same individual are not optimally used.
- Bias in subsequent differential expression testing.

Data structure

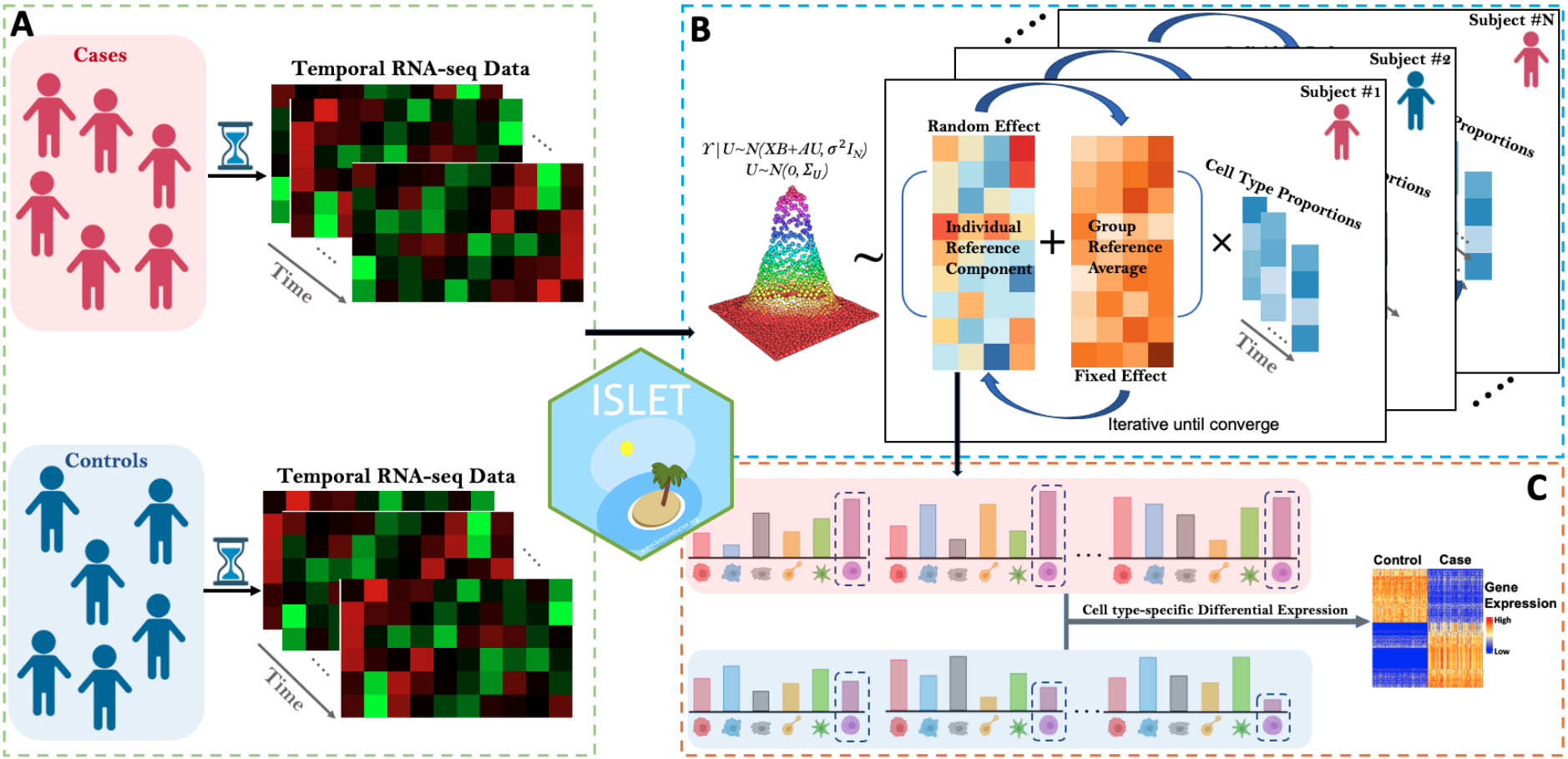




Proposed solution



ISLET


A
Cases
Temporal RNA-seq Data
Time
Controls
Temporal RNA-seq Data
Time
ISLET
B
$Y|U \sim N(XB+AU, \sigma^2 I_N)$
$U \sim N(0, \Sigma_U)$
Random Effect
Individual Reference Component
Group Reference Average
Fixed Effect
Cell Type Proportions
Time
Iterative until converge
Subject #1
Subject #2
Subject #N
C
Cell type-specific Differential Expression
Control
Case
Gene Expression
High
Low

ISLET

ISLET: **I**ndividual-**S**pecific Ce**L**l Typ**E** Referencing **T**ool

Data: Gene expression data from bulk RNA-seq

For one specific gene:

- Subject is index by j, where $j \in 1, 2, ...J$.
- For each subject j, there are T_j longitudinal observations.
- y_{jt}: the observed gene expression for subject j at time t.

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T_1} \\ \vdots \\ y_{J1} \\ y_{J2} \\ \vdots \\ y_{JT_J} \end{pmatrix}_{N \times 1}$$

Here, $N = \sum_{j=1}^{J} T_j$.

Other known input

- Number of cell types: K.
- Cell type proportions: θ_{jT_jk}, naturally $\sum_{k=1}^{K} \theta_{jT_jk} = 1$.
- Binary scalar z_j to indicate the subject's disease status: (e.x. cancer vs. normal).

The mixed-effect model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_0^2 \boldsymbol{I})$$

$$\boldsymbol{X} = \begin{pmatrix} \theta_{111} & \theta_{112} & \dots & \theta_{11K} & z_1\theta_{111} & z_1\theta_{112} & \dots & z_1\theta_{11K} \\ \theta_{121} & \theta_{122} & \dots & \theta_{12K} & z_1\theta_{121} & z_1\theta_{122} & \dots & z_1\theta_{12K} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{1T_11} & \theta_{1T_12} & \dots & \theta_{1T_1K} & z_1\theta_{1T_11} & z_1\theta_{1T_12} & \dots & z_1\theta_{1T_1K} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{J11} & \theta_{J12} & \dots & \theta_{J1K} & z_J\theta_{J11} & z_J\theta_{J12} & \dots & z_J\theta_{J1K} \\ \theta_{J21} & \theta_{J22} & \dots & \theta_{J2K} & z_J\theta_{J21} & z_J\theta_{J22} & \dots & z_J\theta_{J2K} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{JT_J1} & \theta_{JT_J2} & \dots & \theta_{JT_JK} & z_J\theta_{JT_J1} & z_J\theta_{JT_J2} & \dots & z_J\theta_{JT_JK} \end{pmatrix}_{N \times 2K}$$

Design matrix for the random effect

$$\mathbf{a}_{jk} := \begin{pmatrix} \theta_{j_1 k} \\ \theta_{j_2 k} \\ \vdots \\ \theta_{j T_j k} \end{pmatrix}$$

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{a}_{11} & 0 & 0 & 0 & \dots & \mathbf{a}_{1K} & 0 & 0 & 0 \\ 0 & \mathbf{a}_{21} & 0 & 0 & \dots & 0 & \mathbf{a}_{2K} & 0 & 0 \\ 0 & 0 & \ddots & 0 & \dots & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{a}_{J1} & \dots & 0 & 0 & 0 & \mathbf{a}_{JK} \end{pmatrix}_{N \times Q}$$

$Q = JK$

Fixed and random effects

$$\boldsymbol{\beta} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_K \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}_{2K \times 1}$$

$$\boldsymbol{u} = \begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{J1} \\ u_{12} \\ u_{22} \\ \vdots \\ u_{J2} \\ \vdots \\ u_{1K} \\ u_{2K} \\ \vdots \\ u_{JK} \end{pmatrix}_{Q \times 1}$$

Towards EM algorithm setup

$$\boldsymbol{w} = (\boldsymbol{y}, \boldsymbol{u}) := (\boldsymbol{w}_{obs}, \boldsymbol{w}_{mis}) \quad \boldsymbol{w}_{obs} := \boldsymbol{y} \quad \boldsymbol{w}_{mis} := \boldsymbol{u}$$

$$\boldsymbol{w}_{obs} | \boldsymbol{w}_{mis} = \boldsymbol{y} | \boldsymbol{u} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{u}, \sigma_0^2 \boldsymbol{I})$$

$$\boldsymbol{w}_{mis} = \boldsymbol{u} \sim N(\boldsymbol{0}, \Sigma_U)$$

$$\Sigma_U = \begin{pmatrix} \sigma_1^2 \boldsymbol{I}_J & 0 & \dots & 0 \\ 0 & \sigma_2^2 \boldsymbol{I}_J & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \sigma_K^2 \boldsymbol{I}_J \end{pmatrix}$$

Towards EM algorithm setup

$$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} \tag{1}$$

$$COV(\boldsymbol{y}) = \boldsymbol{A}\Sigma_U\boldsymbol{A}^T + \sigma_0^2\boldsymbol{I} \equiv \boldsymbol{V} \tag{2}$$

$$COV(\boldsymbol{y}, \boldsymbol{u}) = \boldsymbol{A}\Sigma_U \tag{3}$$

Ready for EM algorithm

So we have:

$$\begin{pmatrix} \boldsymbol{w}_{obs} \\ \boldsymbol{w}_{mis} \end{pmatrix} = N\left[\begin{pmatrix} \boldsymbol{X\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V} & \boldsymbol{A}\Sigma_U \\ \Sigma_U^T \boldsymbol{A}^T & \Sigma_U \end{pmatrix} \right]$$

Ready for EM algorithm

So we have:

$$\begin{pmatrix} \boldsymbol{w}_{obs} \\ \boldsymbol{w}_{mis} \end{pmatrix} = N\left[\begin{pmatrix} \boldsymbol{X\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V} & \boldsymbol{A\Sigma_U} \\ \Sigma_U^T \boldsymbol{A}^T & \Sigma_U \end{pmatrix} \right]$$

Theorem

If $X = (X_1, X_2)$, *and* $X \sim N[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}]$, *then* $[X_1|X_2] \sim N(\mu_{1|2}, \Sigma_{1|2})$, *where* $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$ *and* $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Thus, we have:

$$[\boldsymbol{w}_{mis}|\boldsymbol{w}_{obs}] = [\boldsymbol{u}|\boldsymbol{y}] \sim N(\boldsymbol{\mu}_p, \Sigma_p)$$

E-step

Define: $\boldsymbol{s} = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}$
Then we have:

$$E[\boldsymbol{u}|\boldsymbol{w}_{obs} = \boldsymbol{y}] = \Sigma_U^T \boldsymbol{A}^T \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$E\left[\boldsymbol{s}^T\boldsymbol{s}|\boldsymbol{w}_{obs} = \boldsymbol{y}\right] = tr(\boldsymbol{A}\Sigma_p\boldsymbol{A}^T) + (\boldsymbol{A}\boldsymbol{\mu}_p + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})^T(\boldsymbol{A}\boldsymbol{\mu}_p + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})$$

$$E\left[\boldsymbol{u}_k^T\boldsymbol{u}_k|\boldsymbol{w}_{obs} = \boldsymbol{y}\right] = tr(\Sigma_{p_k}) + \boldsymbol{\mu}_{p_k}^T\boldsymbol{\mu}_{p_k}$$

M-step

$$\hat{\boldsymbol{\beta}}^{(t+1)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{A}E_{\eta(t)}(\boldsymbol{u}^{(t)}))$$

$$\hat{\sigma}_0^{2(t+1)} = \frac{E_{\eta(t)}\left[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{A}\boldsymbol{u})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{A}\boldsymbol{u})\right]}{N}$$

$$\hat{\sigma}_k^{2(t+1)} = \frac{E_{\eta(t)}(\boldsymbol{\mu}_k^T\boldsymbol{\mu}_k)}{J}$$

Simulation setup

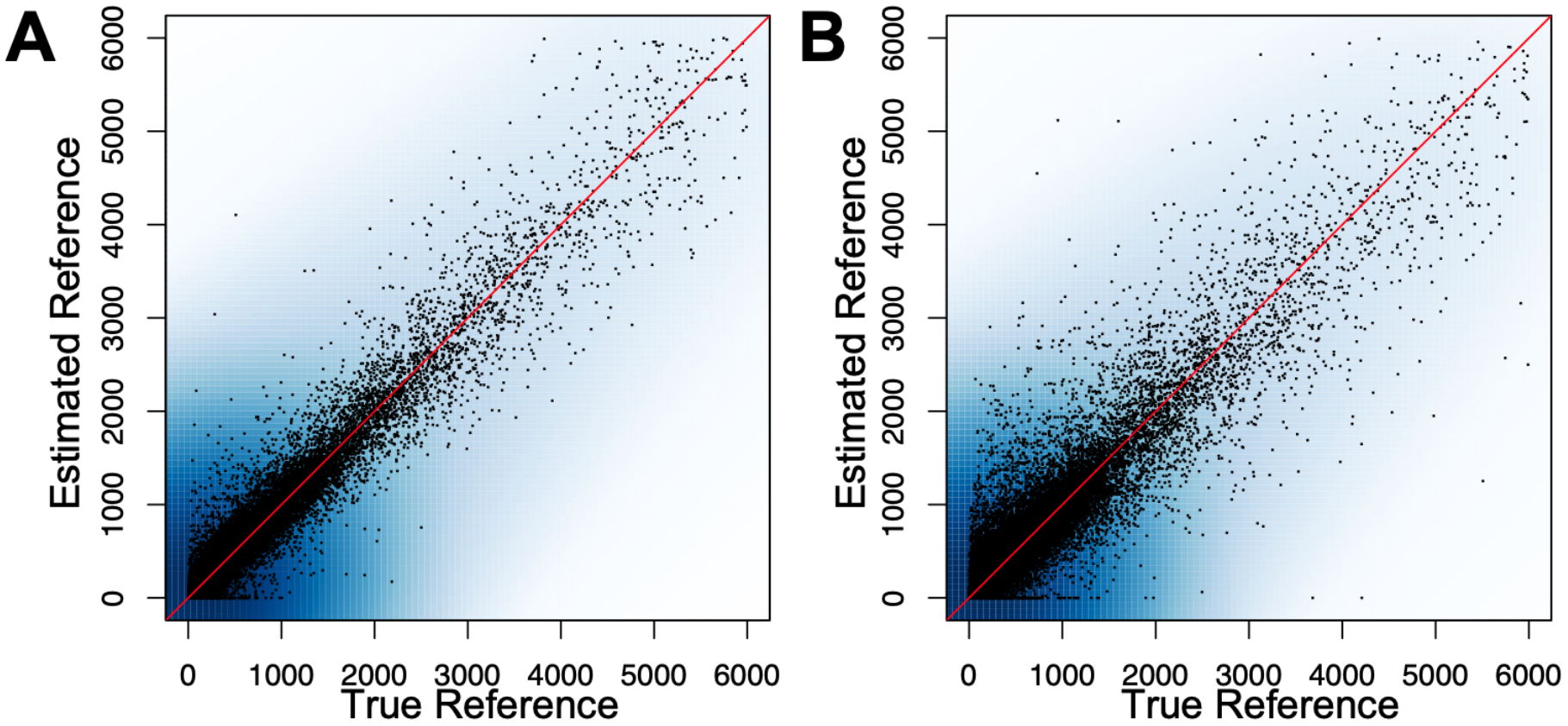
- N = 50, 100, 150, 200
- LFC = 0(null), 0.5, 0.75, 1.0, 1.25, 1.5.
- 10% or 0%(null) csDEG.
- 6 cell types
- Reference panel generated from real bulk cell line.
- Proportions from Dirichlet with parameters from scRNA-seq data.
- Gamma-Poisson for observed counts.

Simulation procedure

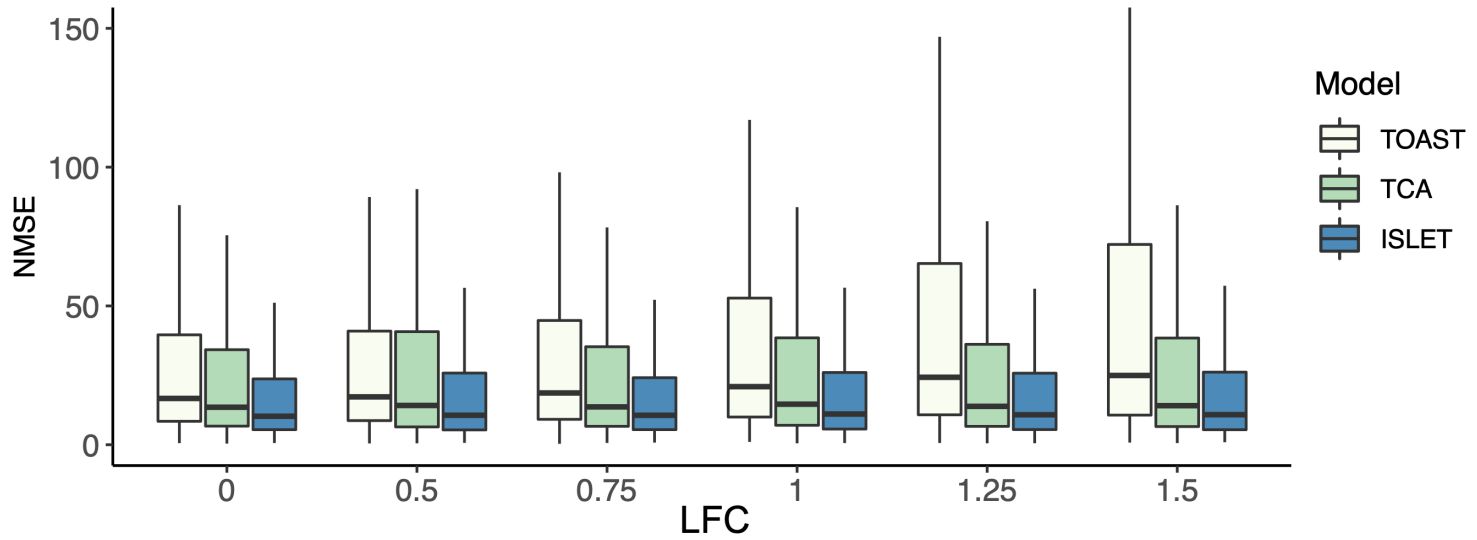
1. $$\boldsymbol{\mu}_{g,K\times 1} \sim MVN(\hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}}_m)$$
$$\boldsymbol{\phi}_{g,K\times 1} \sim MVN(\hat{\boldsymbol{d}}, \hat{\boldsymbol{\Sigma}}_d)$$
2. $$\boldsymbol{M}_{G\times K} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_G]^T; \boldsymbol{\Phi}_{G\times K} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_G]^T$$
3. $$\boldsymbol{X}^i_{G\times K} \sim Gamma\{shape = \frac{1}{\exp(\boldsymbol{\Phi})}, scale = \exp(\boldsymbol{M}) \cdot \exp(\boldsymbol{\Phi})\}$$
4. $$\boldsymbol{\theta}_{it} \sim Dir(\boldsymbol{\alpha})$$
5. $$\boldsymbol{r}_{it} = \boldsymbol{X}^i \boldsymbol{\theta}_{it}$$
$$\boldsymbol{y_{it}} | \boldsymbol{r}_{it} \sim Poisson(\boldsymbol{r}_{it})$$

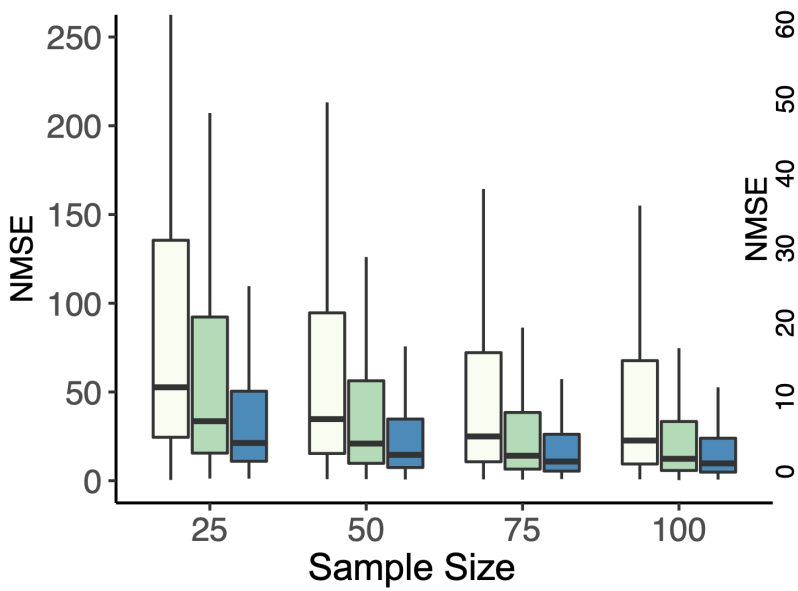
Simulation results

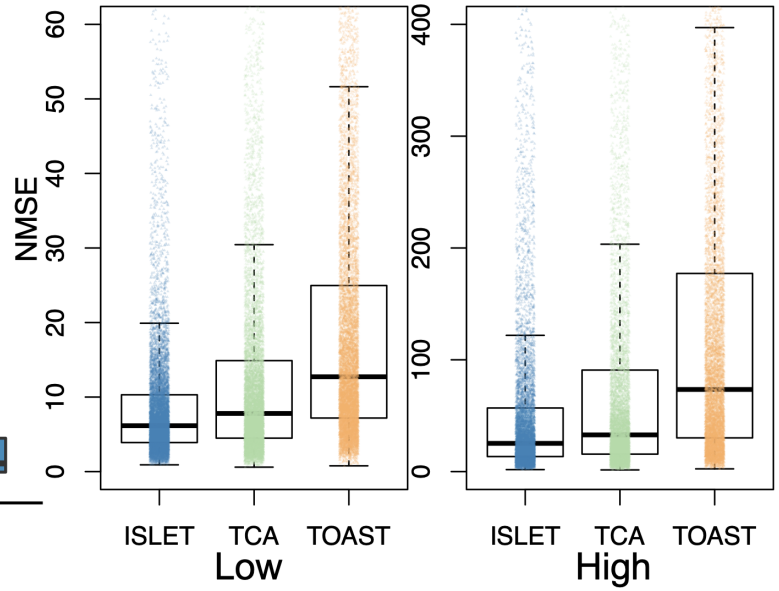
Individual reference panel recovery



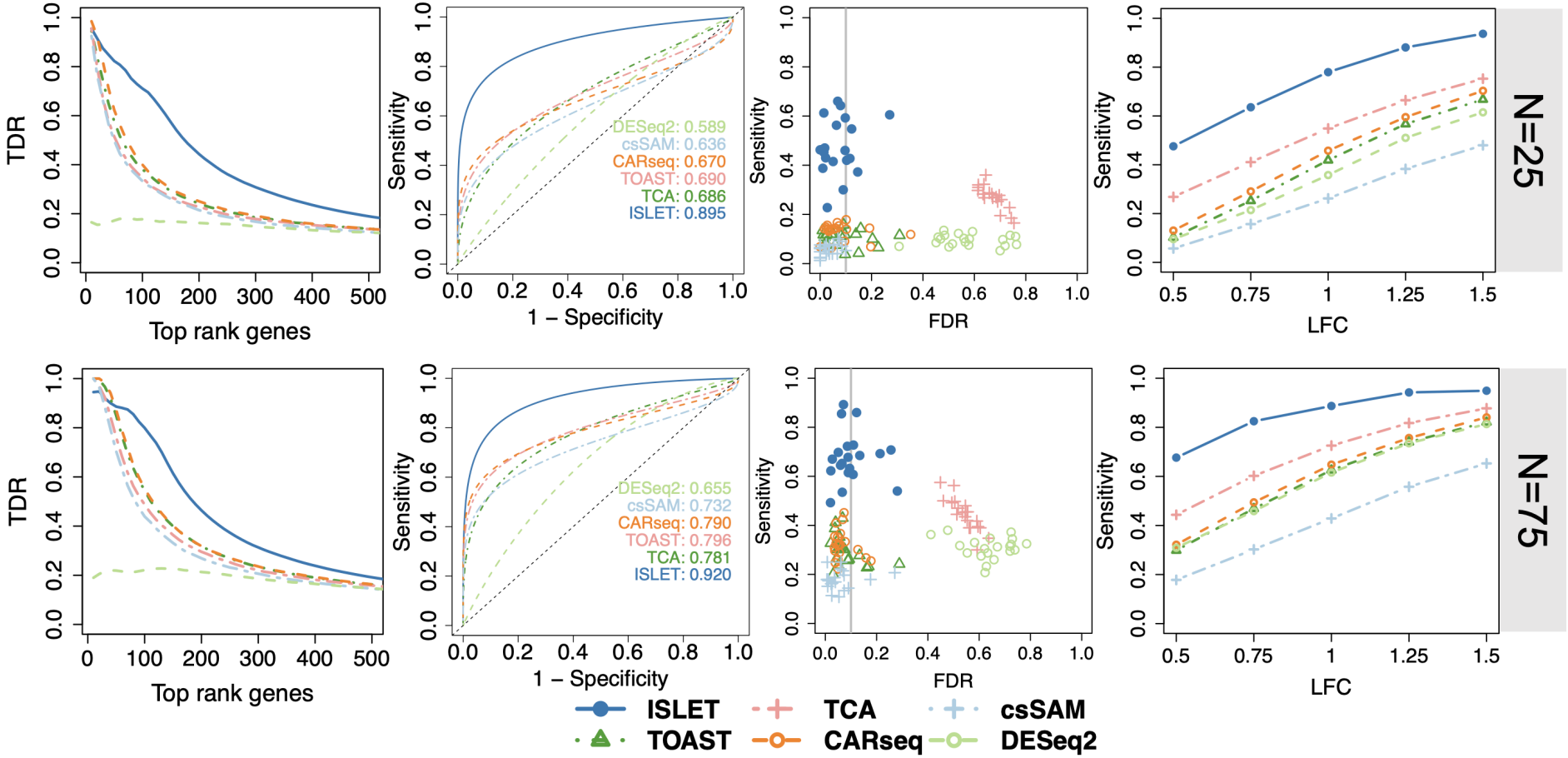
Individual reference panel recovery



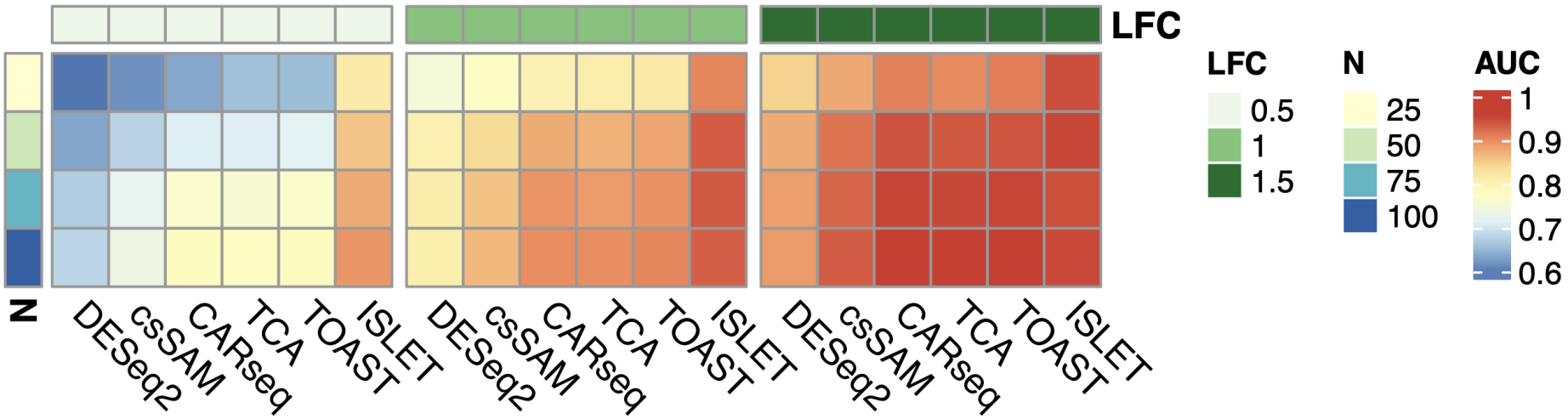




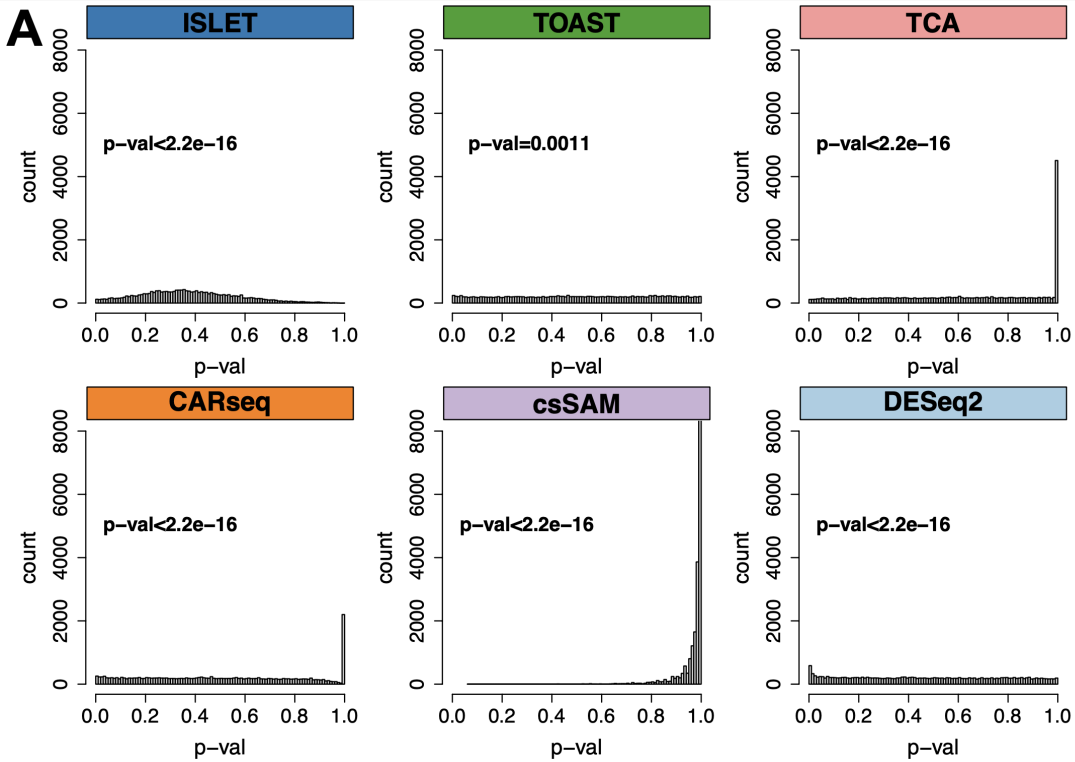
Identifying cell-type-specific DE genes

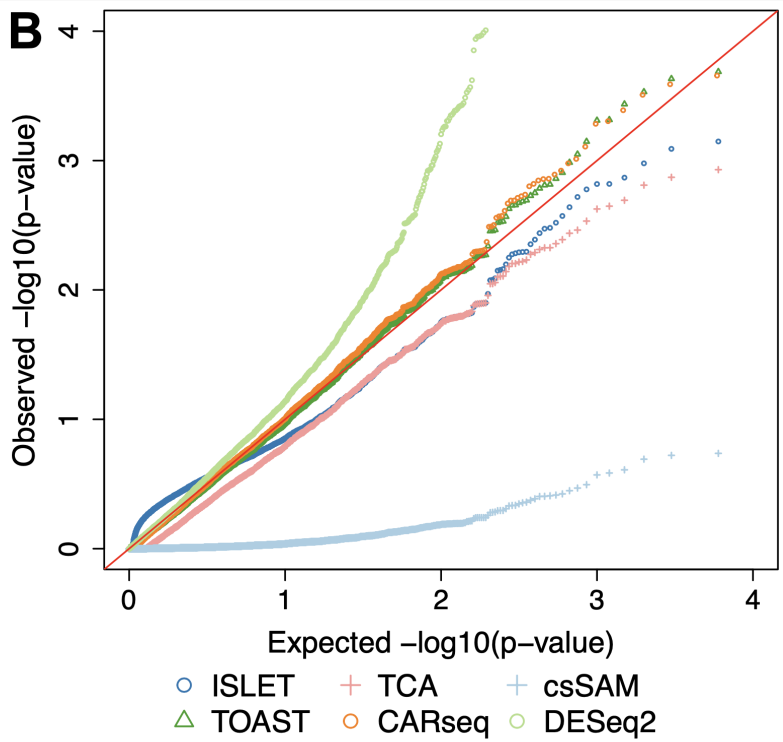


Identifying cell-type-specific DE genes



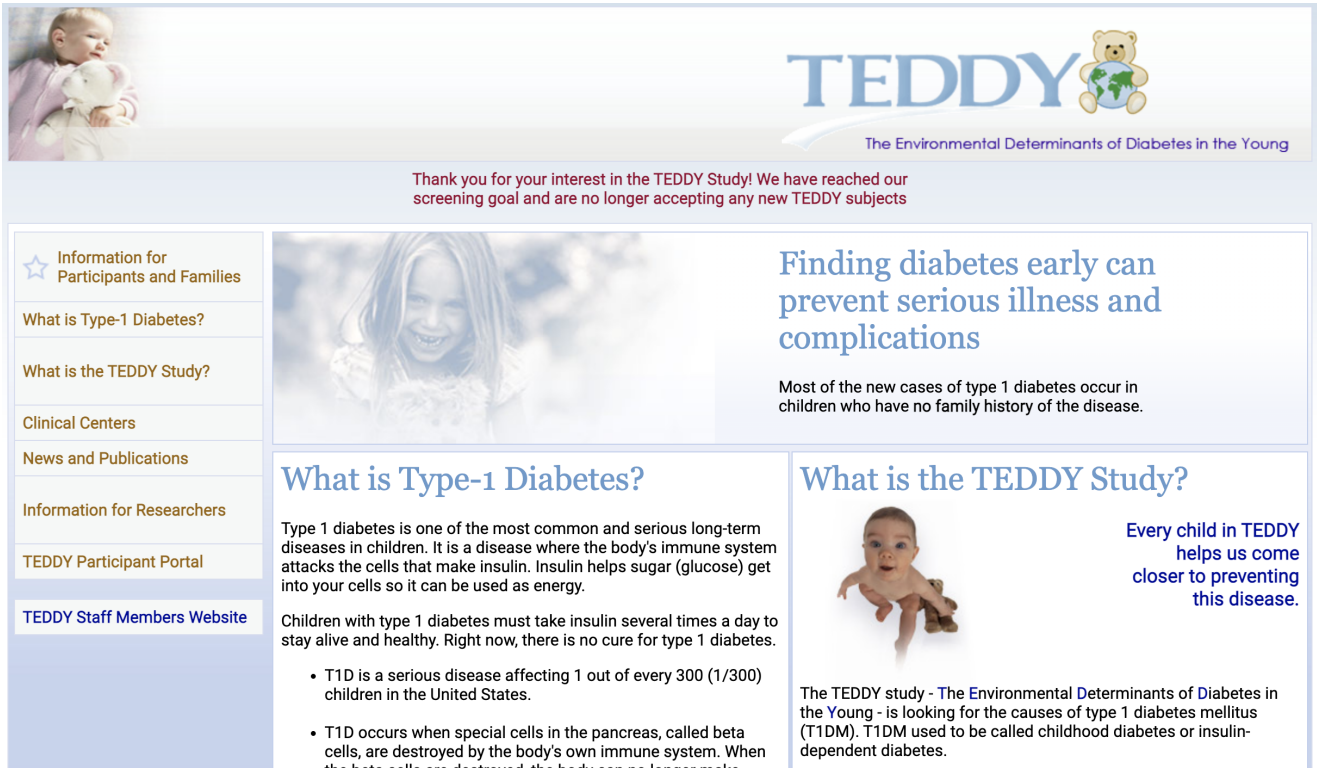
Type I error under the null



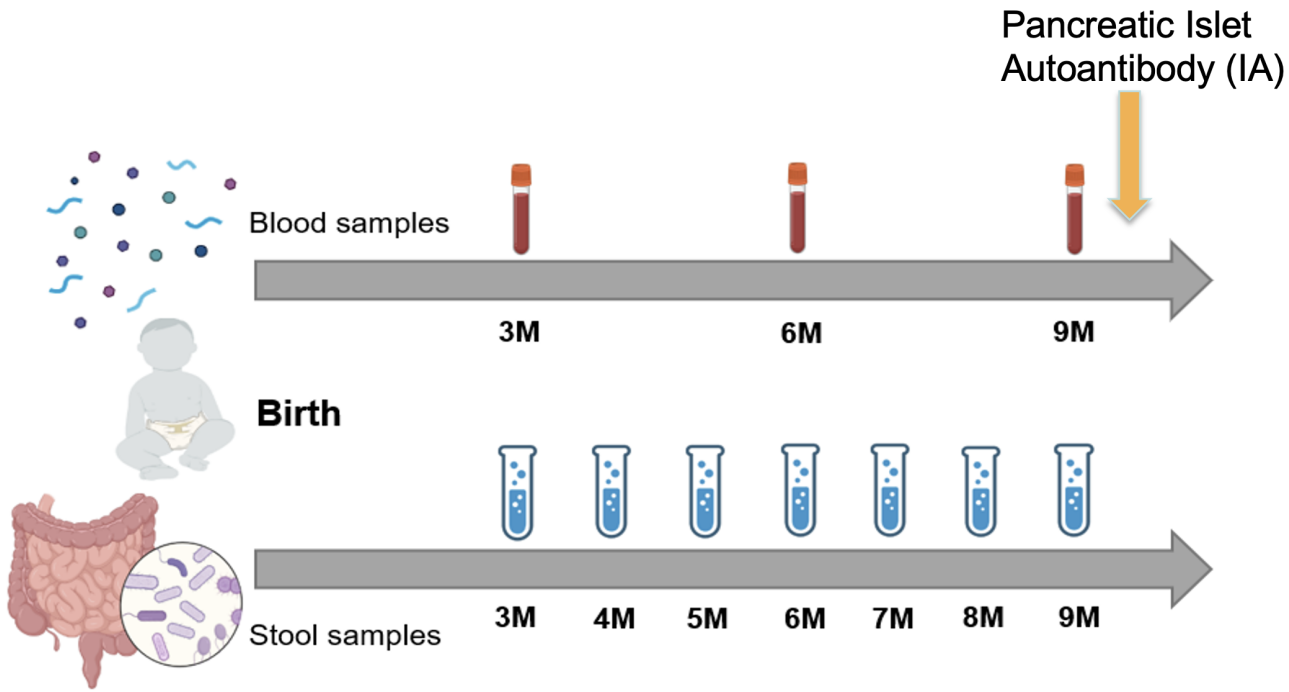


Real data: TEDDY

The Environmental Determinants of Diabetes in the Young

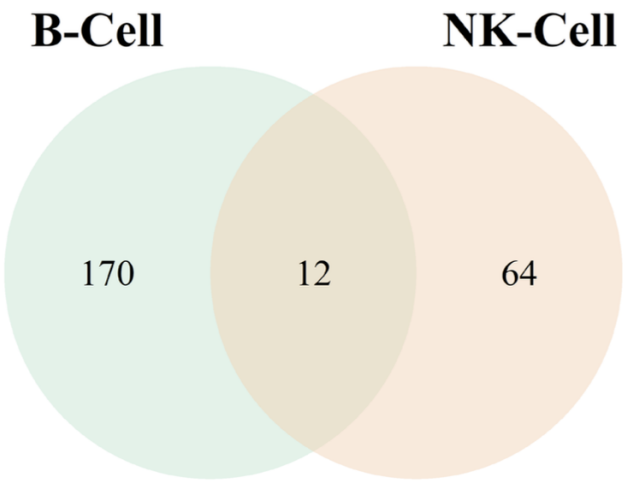


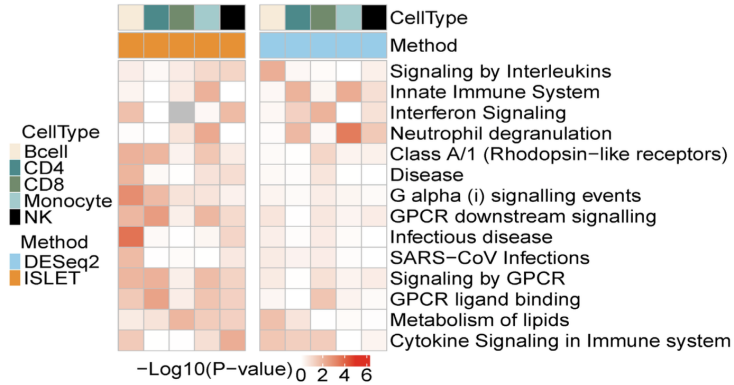
TEDDY Data



Six clinical centers in four countries: U.S., Finland, Germany, and Sweden.
Prospective cohort: 8,676 high-risk infants were enrolled from birth and followed every 3 months, for blood sample collection and islet autoantibody (IAbs) measurement.
Developing IA (cases) started at 9 months with a plateau between 1-2 years of age.

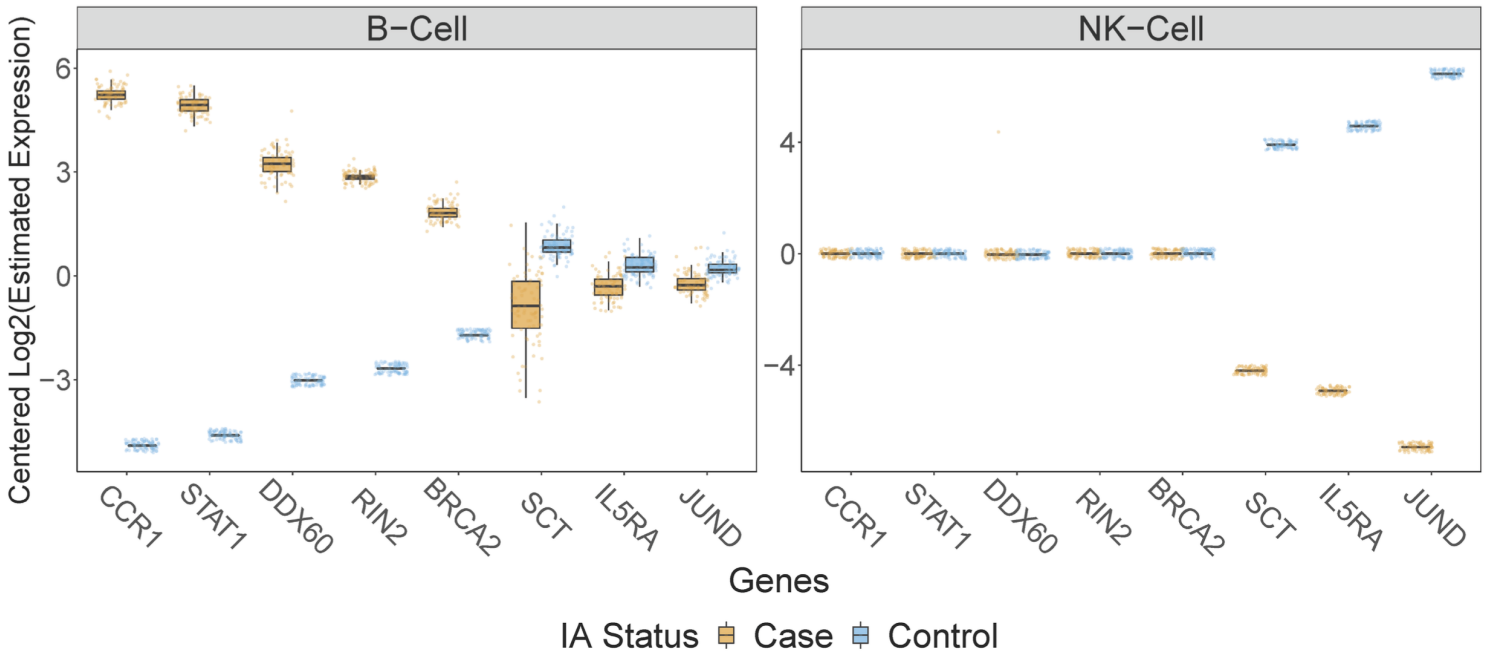
Real data: age-independent effect





IA-signatures strongly enriched in B cells and NK cells transcripts, kinases, and TFs.

Real data: age-independent effect

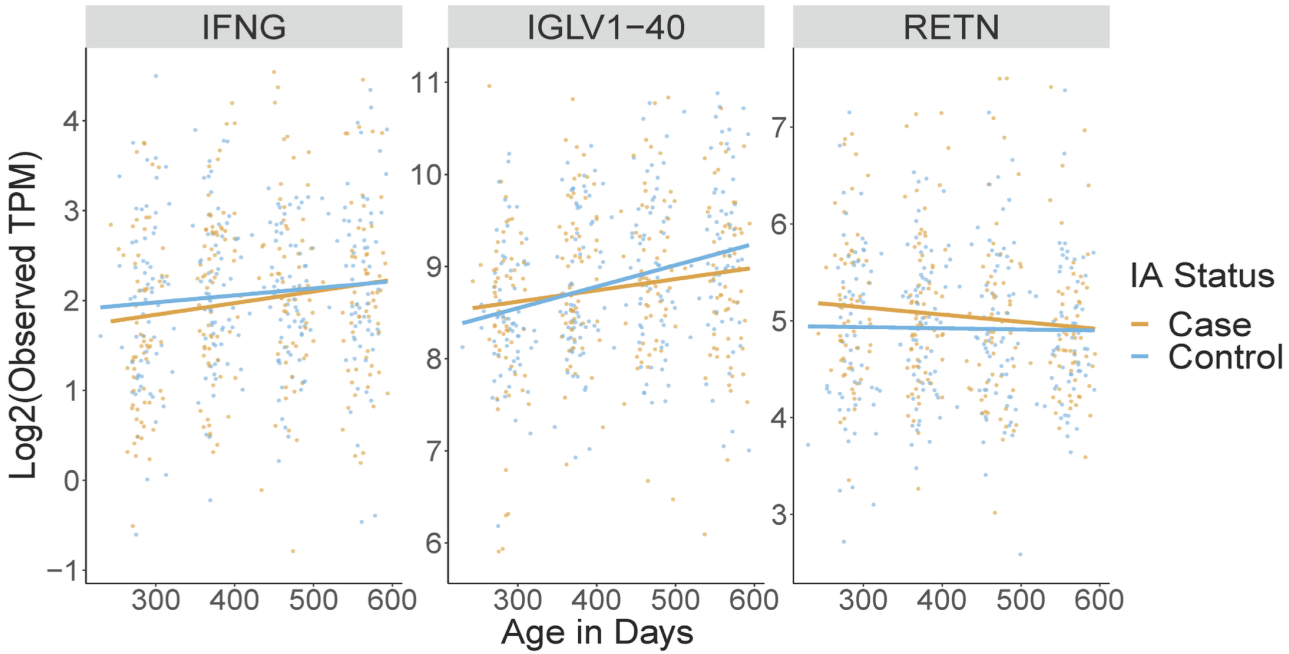


IFN-β-inducible genes in vitro and IFN-indicible transcriptional signatures in human PBMCs were also found prior to the development of autoantibodies (BABY DIET).

CCR1 (chemokine receptor): increased risk of IA in TEDDY among participants who experienced virus infection.

SCT, IL5RA, JUND: regulation of immune response.

Real data: change rate



RETN: Diabetes study found SNPs associated with plasma resistin and glucose levels.
IGLV1-40: neutralizing activity related to protective antibody responses in infants.
More significant and less ambiguous than TEDDY microarray.

ISLET available on Bioconductor

https://bioconductor.org/packages/ISLET/

Individual-specific and cell-type-specific deconvolution using ISLET

Hao Feng[1*] ***and Qian Li***[2]

[1]Department of Population and Quantitative Health Sciences, Case Western Reserve University
[2]Department of Biostatistics, St. Jude Children's Research Hospital

[*]hxf155@case.edu

9 September 2022

Abstract

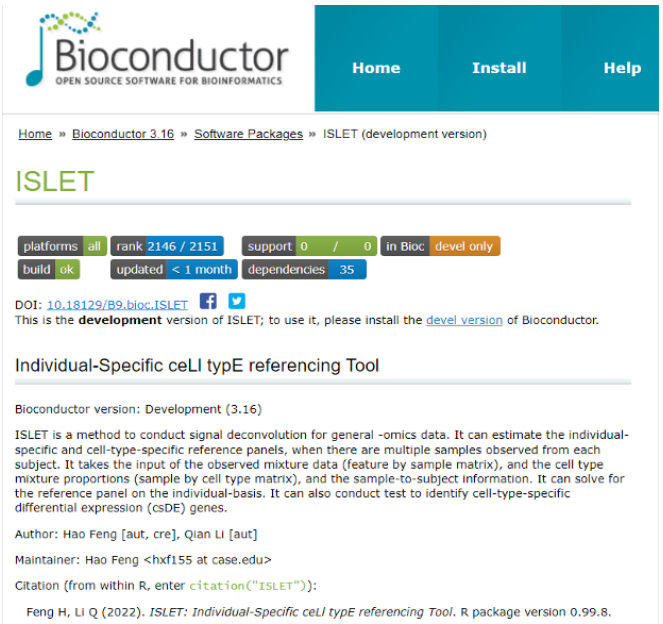
This vignette introduces the usage of the Bioconductor package ISLET (Individual-Specific ceLl typE referencing Tool). Complementary to classic deconvolution algorithms, ISLET can take cell type proportions as input, and infer the individual-specific and cell-type-specific reference panels. ISLET also offers functions to detect cell-type specific differential expression (csDE) genes. Additionally, it can test for csDE genes change rate difference between two groups, given an additional covariate of time points or age. ISLET is based on rigorous statistical framework of Expectation–Maximization(EM) algorithm, and has parallel computing embedded to provide superior computational performance.

Package

ISLET 0.99.8

Contents

1 Install and help
- 1.1 Install ISLET
- 1.2 How to get help



ISLET summary

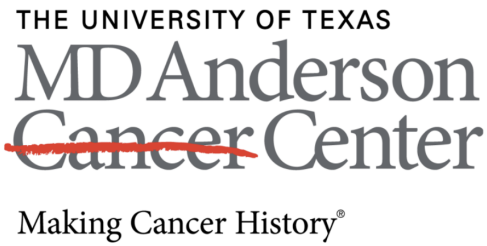
- First method to recover individual-specific and cell-type-specific reference panel.
- ISLET: an EM algorithm based deconvolution and testing framework.
- Optimize the usage of repeatedly measured samples within the same subject.
- Accurate, robust and powerful performance.
- A R/Bioconductor package.
- ISLET successfully identified gene signatures in B, NK and CD4+ T cells, prior to the onset of pancreatic β-cell autoantibody.

Acknowledgement

- Daoyu Duan
- Leslie Meng

- Qian Li

❖ IRG-16-186-21



- Ziyi Li

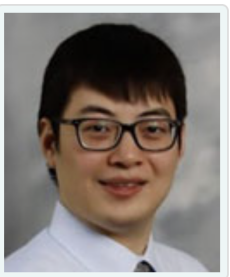
Monday, March 20 | 10:30 am – 12:15 pm
T2 | Cell-type-aware Differential Analysis for Bulk Transcriptome Data

Instructor:
Hao Feng, Department of Population and Quantitative Health Sciences, Case Western Reserve University

Course Description:

The real-world clinical tissue samples are composed of diverse cell types. Recent statistical method advances in signal decomposition have enabled transcriptome studies at cell type resolution. In this tutorial, we will cover common software packages for cell type aware analysis in bulk transcriptome data. First, we will briefly discuss signal deconvolution, to introduce the basics in estimating pure cell type reference profiles and estimating cell type proportions. Next, we will mainly focus on widely adopted tools to conduct cell-type-specific Differential Expression Genes (csDEG) analysis. Novel methods and research progress in this domain will also be covered. This tutorial will have both the lecture component and the hands-on coding practice. We will provide R code implementation for methods introduced in this tutorial.

Statistical/Programming Knowledge Required:
Basic R programming.



Hao "Harry" Feng, is an Assistant Professor in the Department of Population and Quantitative Health Sciences at Case Western Reserve University. His research focuses on the development and the application of statistical bioinformatics methods to better understand high-throughput -omics data, especially in epigenomics. He proposed several novel statistical models in epigenomics data modeling, single-cell data analysis, cell-free DNA methylation and signal deconvolution. He developed a number of open-source software tools that are available on R-CRAN and Bioconductor, with > 24,000 downloads annually. He received his Ph.D. in Biostatistics and Bioinformatics from Emory University

Decomposing Admixed Genomics Data: Cell-type-aware Analysis Methodology Advances

Chair & Organizer: Hao Feng, Case Western Reserve University

Speakers:

Aaron Newman, Stanford University

Stephanie Hicks, Johns Hopkins Bloomberg School of Public Health

Wenyi Wang, The University of Texas MD Anderson Cancer Center

Rafael Irizarry, Dana-Farber Cancer Institute, Harvard T.H. Chan School of Public Health.