Permutation Enhances the Rigor in Genomics Data Analysis

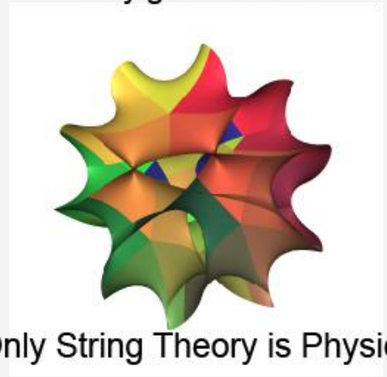
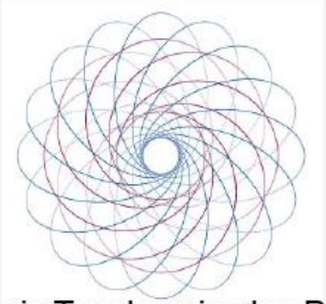
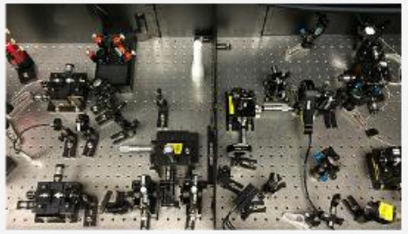
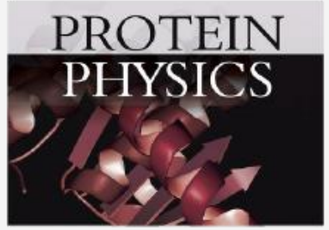
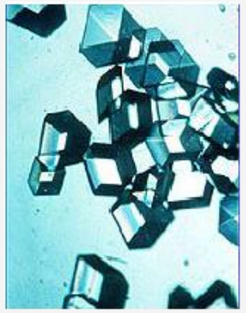
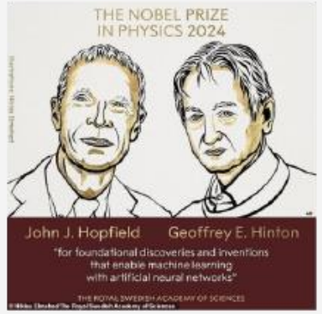
Jingyi Jessica Li
Professor and Program Head of Biostatistics
Fred Hutchinson Cancer Center

Affiliate Professor of Biostatistics
University of Washington

Junction of Statistics and Biology

Fred Hutch Cancer Center

Liberalism in scientific research

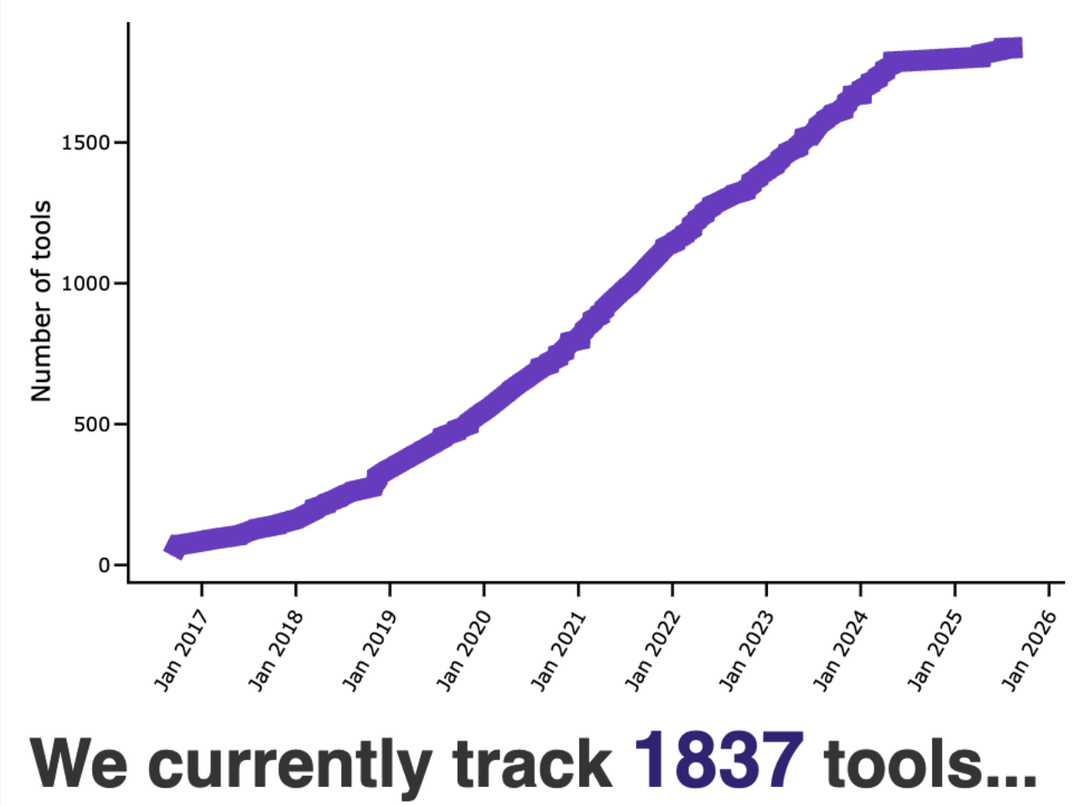
What is Physics?	Object Fundamentalism Only general laws	Object Neutralism Not necessarily general laws	Object Liberalism Anything related to reality
Method Fundamentalism Only theories are Physics	 Only String Theory is Physics	 Condensed Matter Physics is also Physics	 Algebraic Topology is also Physics
Method Neutralism Experiments can also be Physics	 Accelerator Data Analysis is also Physics	 Quantum Optics is also Physics	 Life Science is also Physics
Method Liberalism Anything that involves experiment is Physics	 Divination is also Physics	 Long Material is also Physics	 **Neural Network is also Physics**

Translated from https://www.163.com/dy/article/JE0VILQC05118OGM.html (https://www.zhihu.com/pin/1513735925213917184) by Dr. Pan Liu

Genomics is a “liberal” discipline

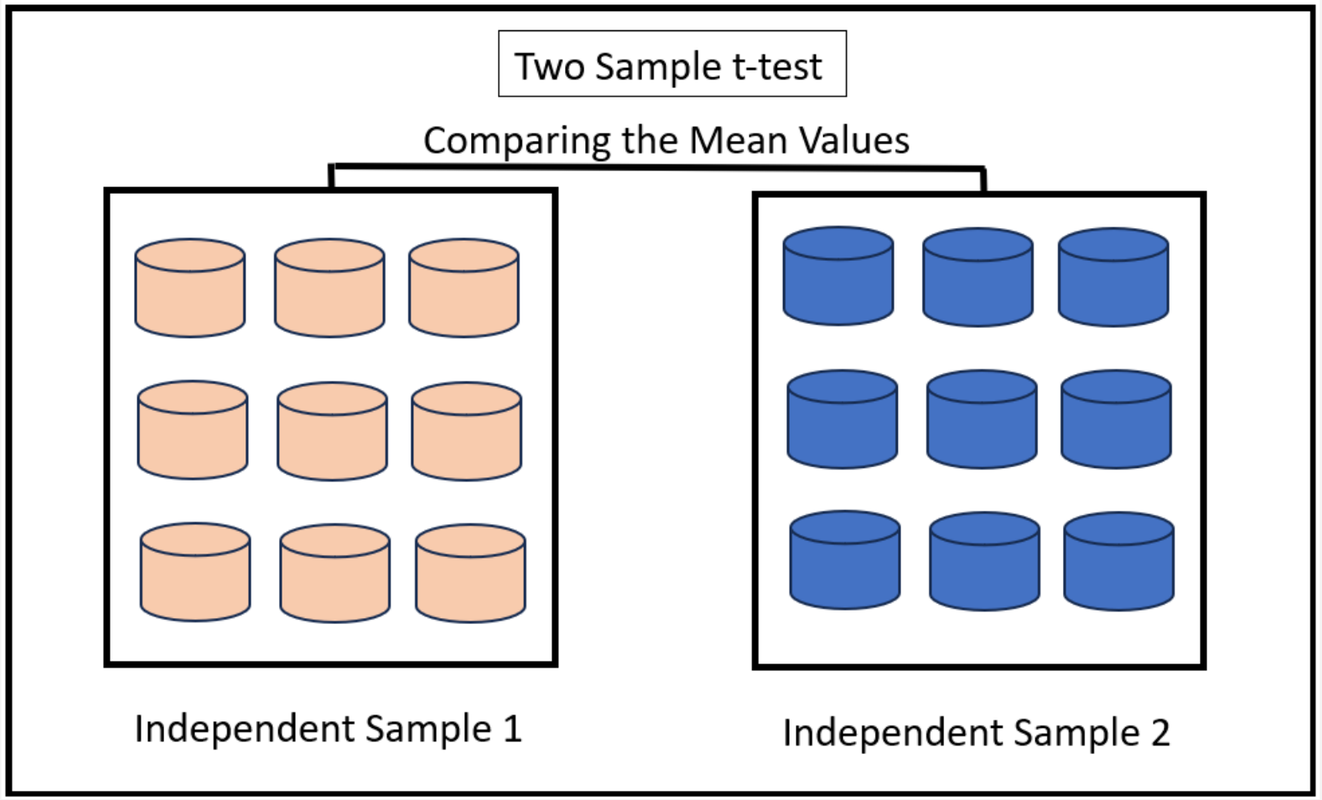
1. Interdisciplinary nature
2. Data-driven focus
3. Rapid evolution of methods
4. Flexible analytical approaches

Single-cell methods



https://www.scrna-tools.org/

Statistics ensures rigor in data analysis



Guinness Brewery

Dublin, Ireland

William Sealy Gosset, who developed the "*t*-statistic" and published it under the pseudonym of "Student"

Journal of the American Statistical Association ›
Volume 116, 2021 - Issue 536

Enter keywords, authors, DO

Submit an article | Journal homepage

83,587
Views

26
CrossRef citations to date

249
Altmetric

Listen

Reviews

What are the Most Important Statistical Ideas of the Past 50 Years?

Andrew Gelman & Aki Vehtari

Pages 2087-2097 | Received 30 Nov 2020, Accepted 23 May 2021, Published online: 08 Jul 2021

Cite this article https://doi.org/10.1080/01621459.2021.1938081 Check for updates

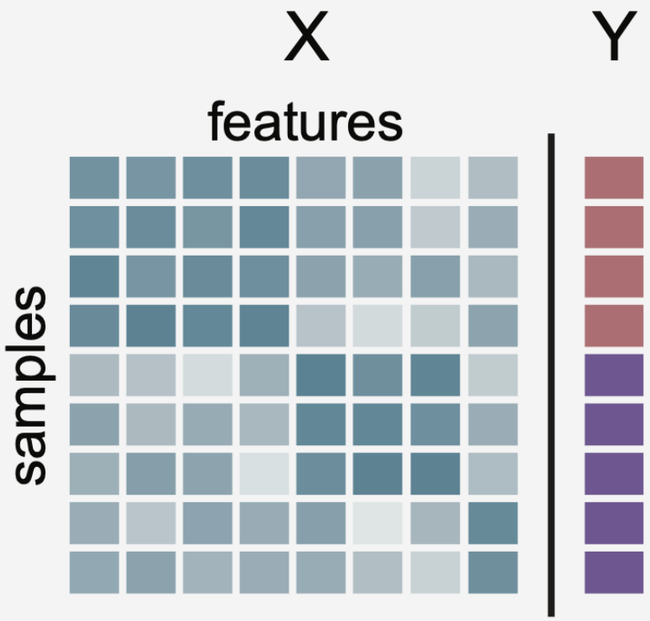
Bootstrapping and Simulation-Based Inference

"In **permutation** testing, resampled datasets are generated by breaking the (possible) dependency between the predictors and target by randomly shuffling the target values."

https://medium.com/bitgrit-data-science-publication/the-8-most-important-statistical-ideas-of-the-past-50-years-11220e46736f

How to permute data?

Supervised learning

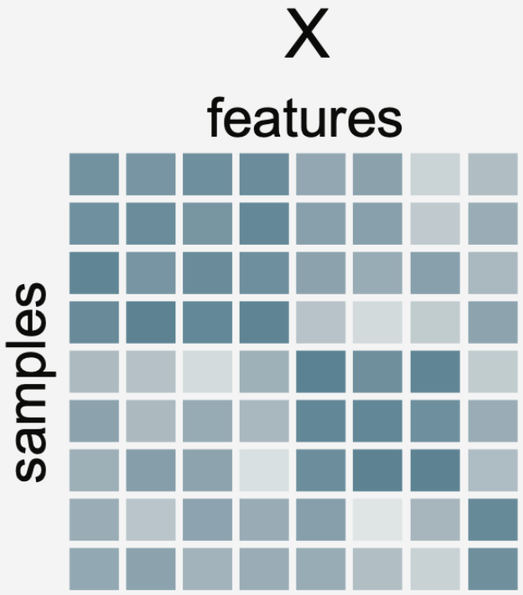


Bulk RNA-seq:

features = genes

Y = sample condition labels

Unsupervised learning

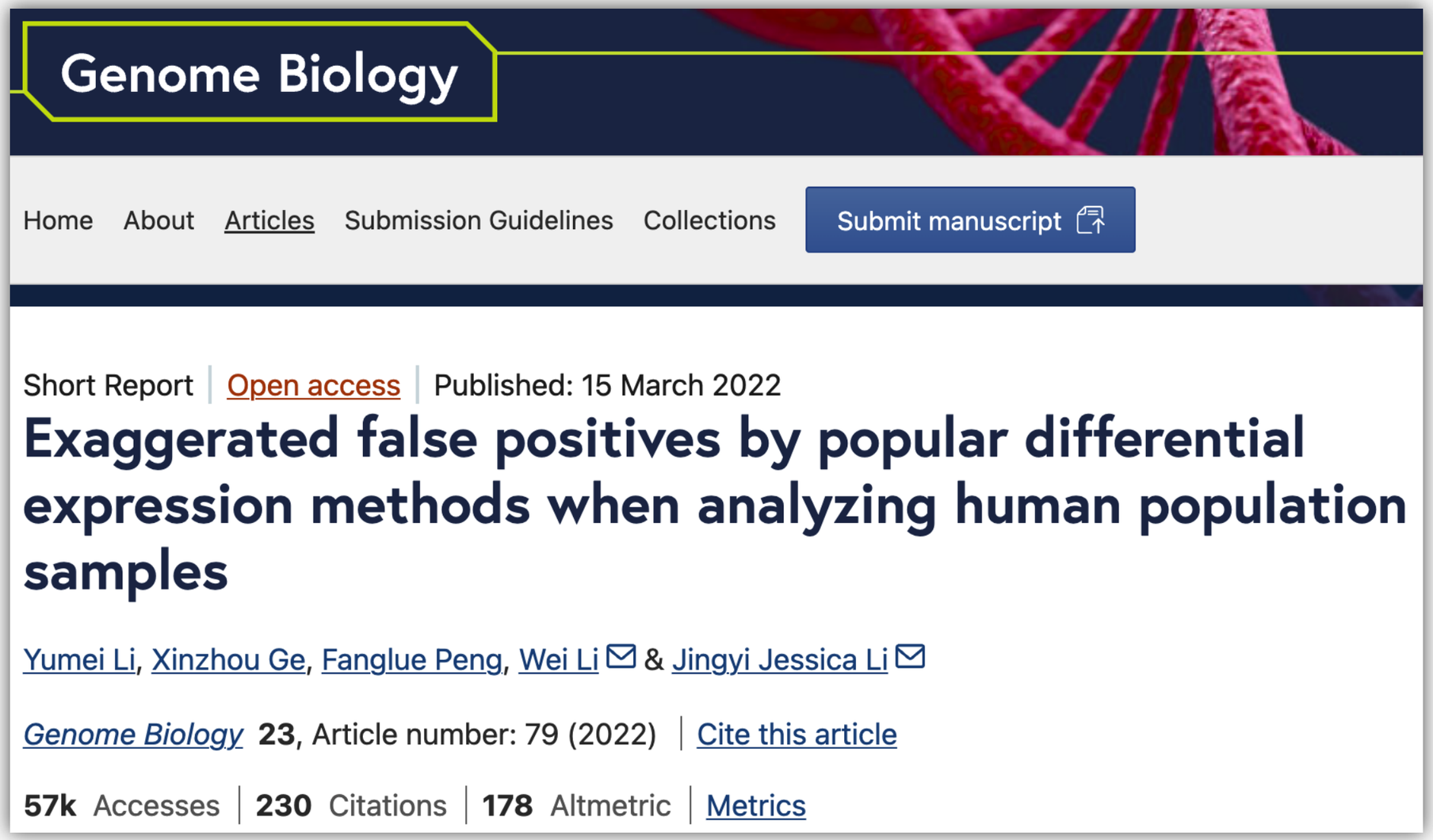


Single-cell RNA-seq:

samples = cells;

features = genes

Teaser: bulk RNA-seq DE analysis



Genome Biology

Home About Articles Submission Guidelines Collections Submit manuscript

Short Report | Open access | Published: 15 March 2022

Exaggerated false positives by popular differential expression methods when analyzing human population samples

Yumei Li, Xinzhou Ge, Fanglue Peng, Wei Li & Jingyi Jessica Li

Genome Biology **23**, Article number: 79 (2022) | Cite this article

57k Accesses | **230** Citations | **178** Altmetric | Metrics

Yumei Li
(Wei Li Lab→
Soochow U)

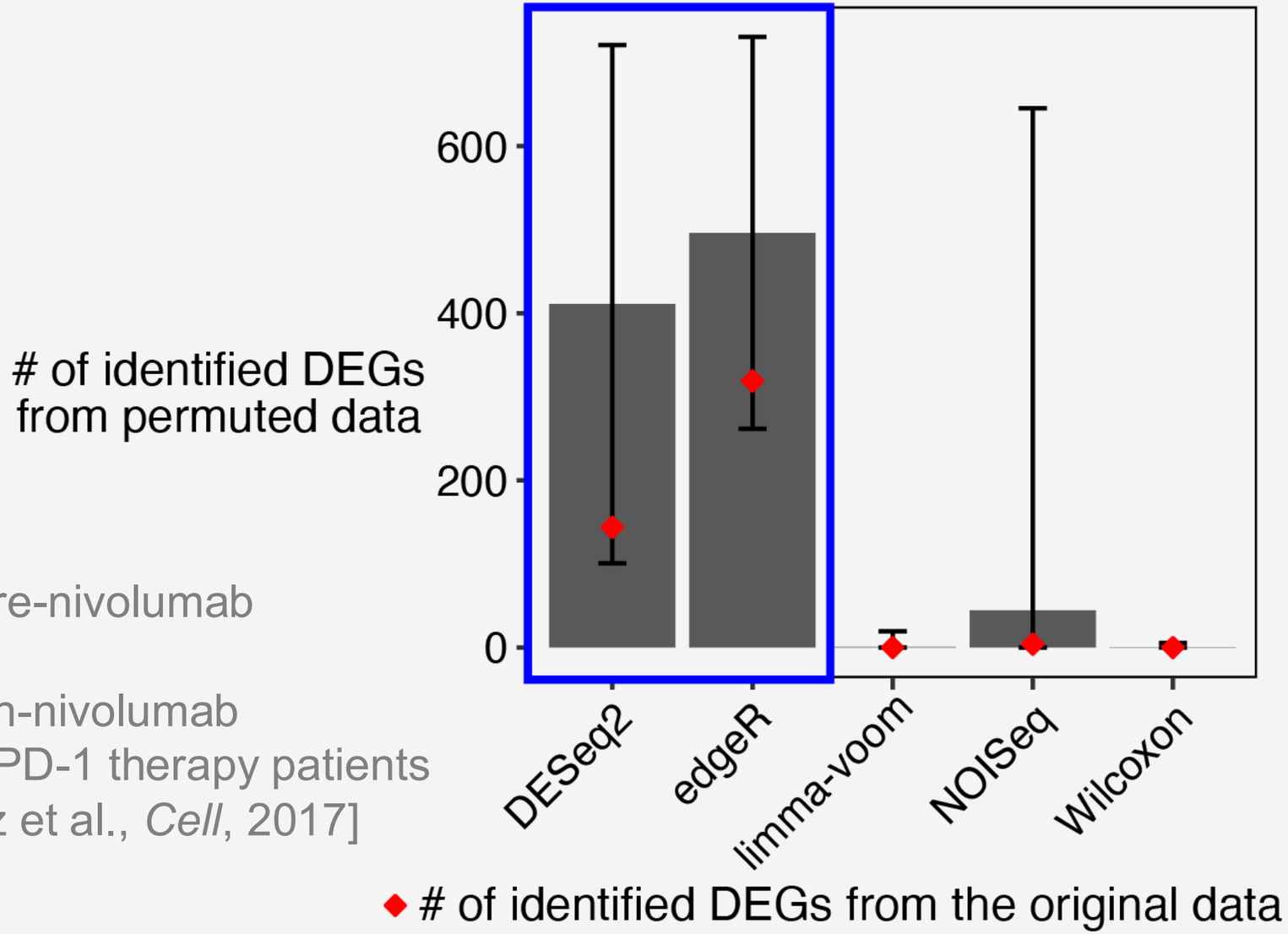
Xinzhou Ge
(JSB→
OregonState)

Wei Li
(UC Irvine)

X/Twitter: @jsb_ucla

Teaser: bulk RNA-seq DE analysis

Q: Why are many genes identified as DE genes from permuted data?



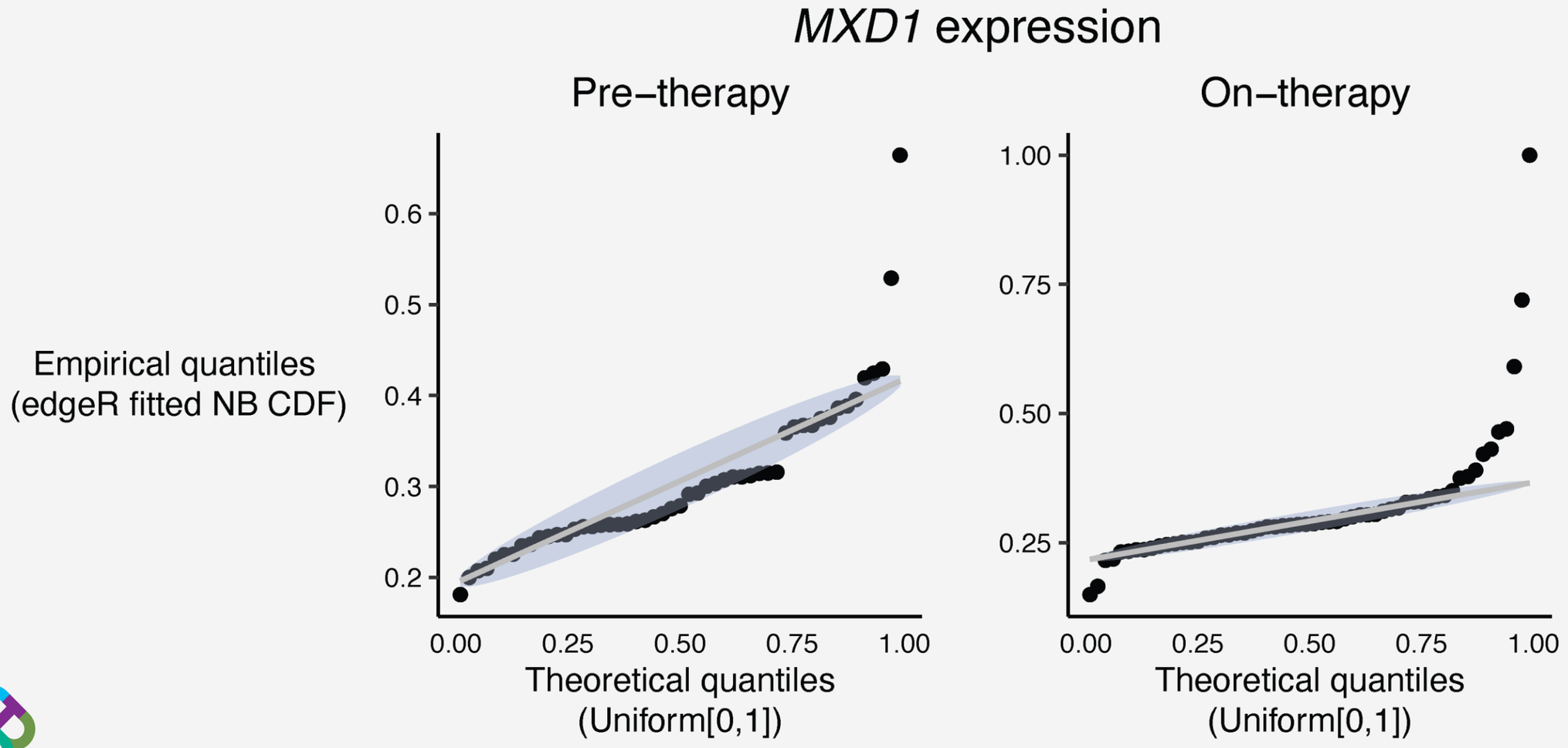
51 pre-nivolumab
vs.
58 on-nivolumab
anti-PD-1 therapy patients
[Riaz et al., *Cell*, 2017]

[Li*, Ge* et al., *Genome Biology*, 2022]

Teaser: bulk RNA-seq DE analysis

Q: Why are many genes identified as DE genes from permuted data?

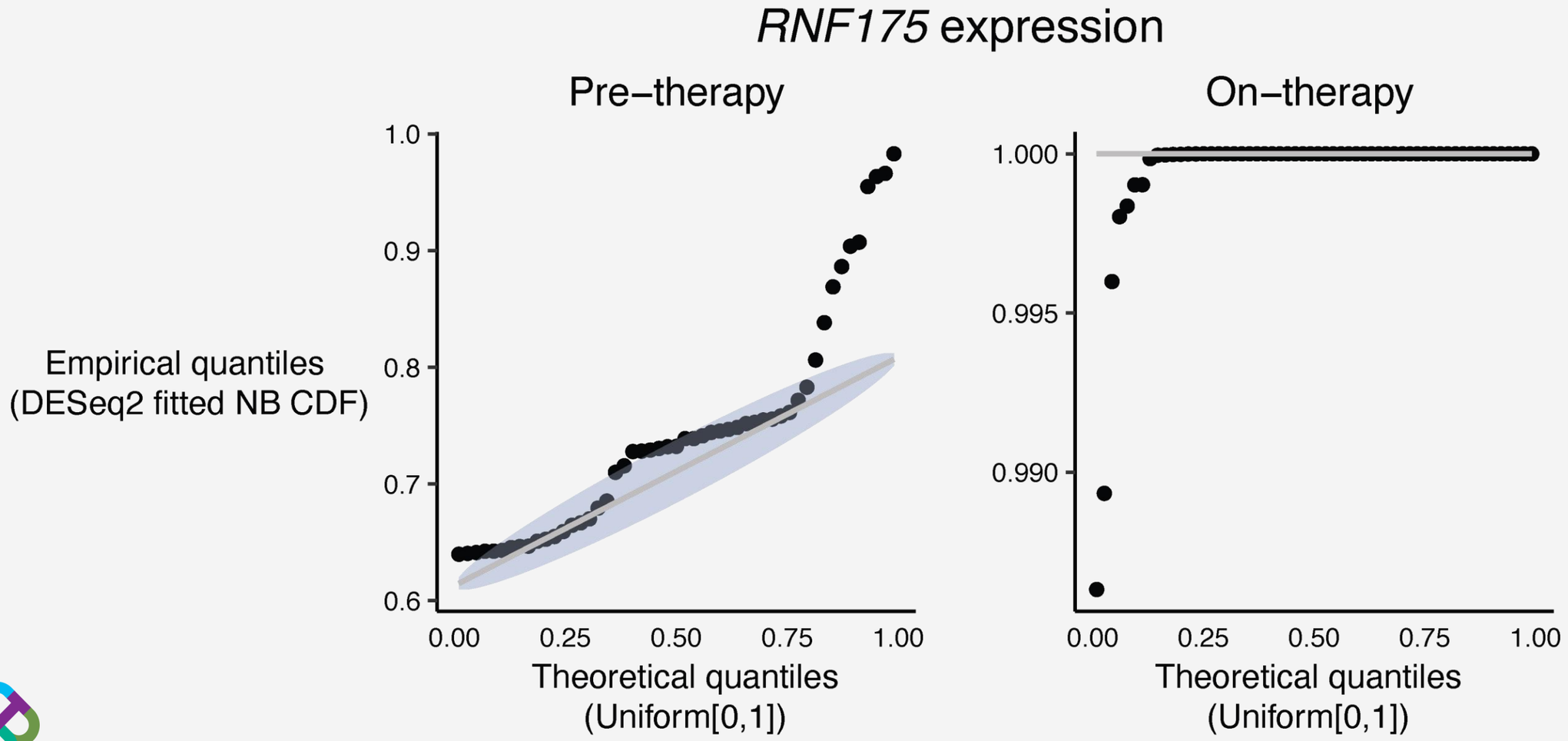
A: The negative binomial assumption does not hold on this dataset.



Teaser: bulk RNA-seq DE analysis

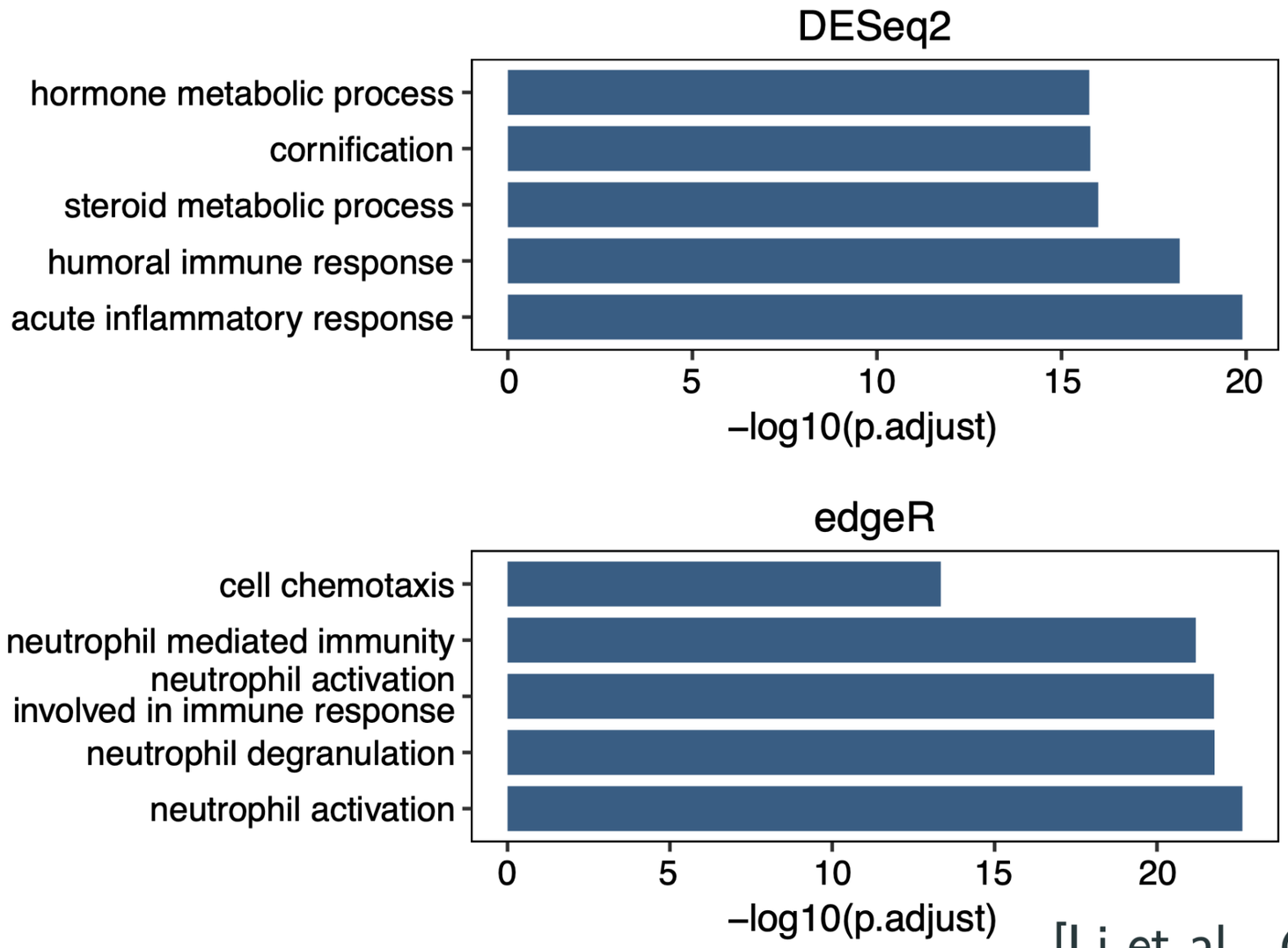
Q: Why are many genes identified as DE genes from permuted data?

A: The negative binomial assumption does not hold on this dataset.



Teaser: bulk RNA-seq DE analysis

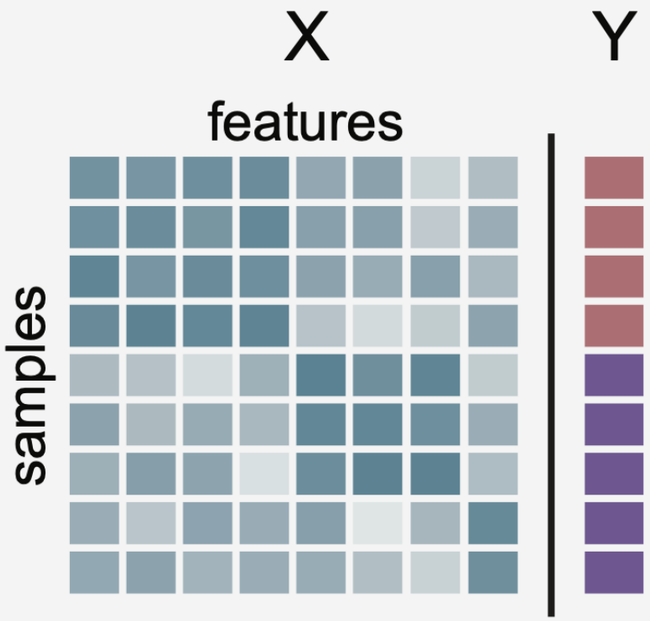
False discoveries may mislead scientific conclusions



[Li et al., *Genome Biology*, 2022]

How to permute data?

Supervised learning

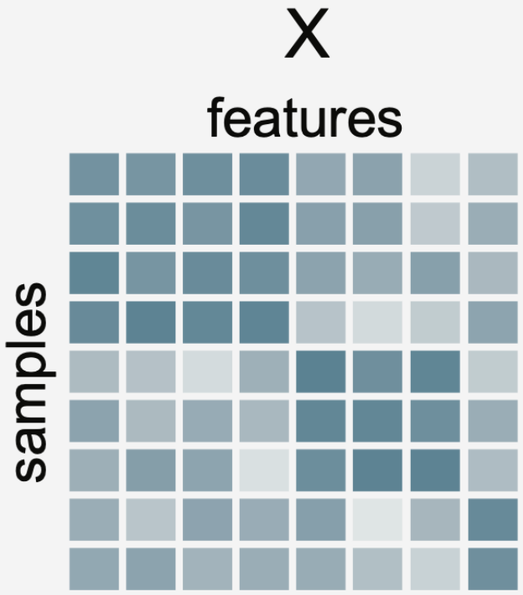


Bulk RNA-seq:

features = genes

Y = sample condition labels

Unsupervised learning



Single-cell RNA-seq:

samples = cells;

features = genes

Two examples where permutation helps

1. Single-cell data **visualization**

Statistical method scDEED for detecting dubious 2D single-cell embeddings and optimizing t-SNE and UMAP hyperparameters

Lucy Xia, Christy Lee & Jingyi Jessica Li

Nature Communications **15**, Article number: 1753 (2024) | Cite this article

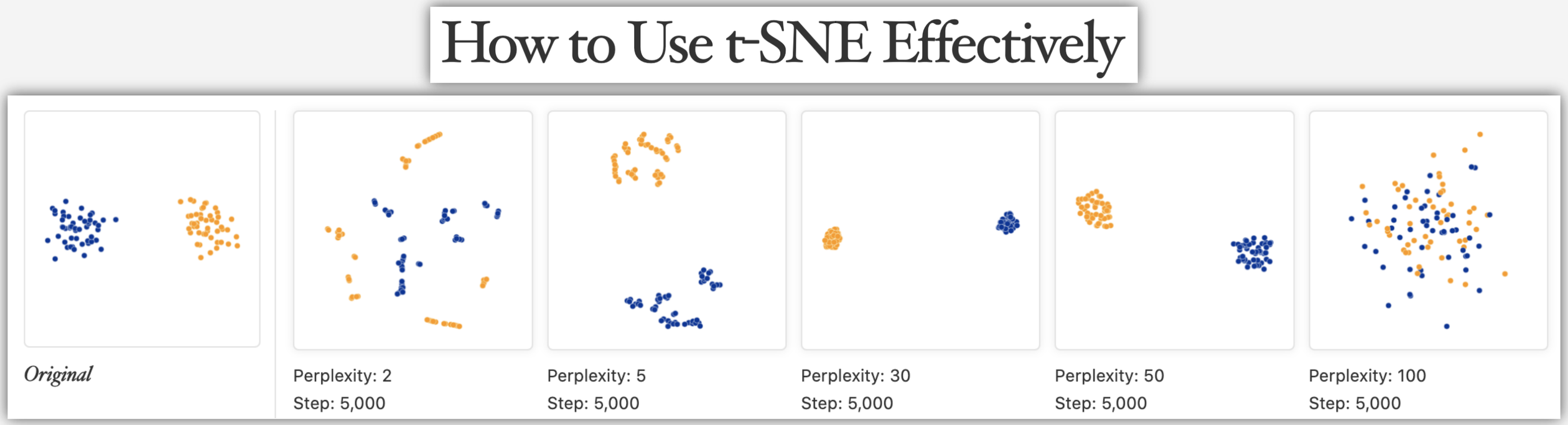
2. Aggregating single cells into **metacells**

mcRigor: a statistical method to enhance the rigor of metacell partitioning in single-cell data analysis

Pan Liu & Jingyi Jessica Li

Nature Communications **16**, Article number: 8602 (2025) | Cite this article

Example 1: dubious t-SNE/UMAP embeddings?



- **Hyperparameters** really matter
- **Distances between clusters** might not mean anything
- …

Source: https://distill.pub/2016/misread-tsne/

Example 1: dubious t-SNE/UMAP embeddings?

nature methods

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature methods › technology features › article

Technology Feature | Published: 24 May 2024

Seeing data as t-SNE and UMAP do

Vivien Marx

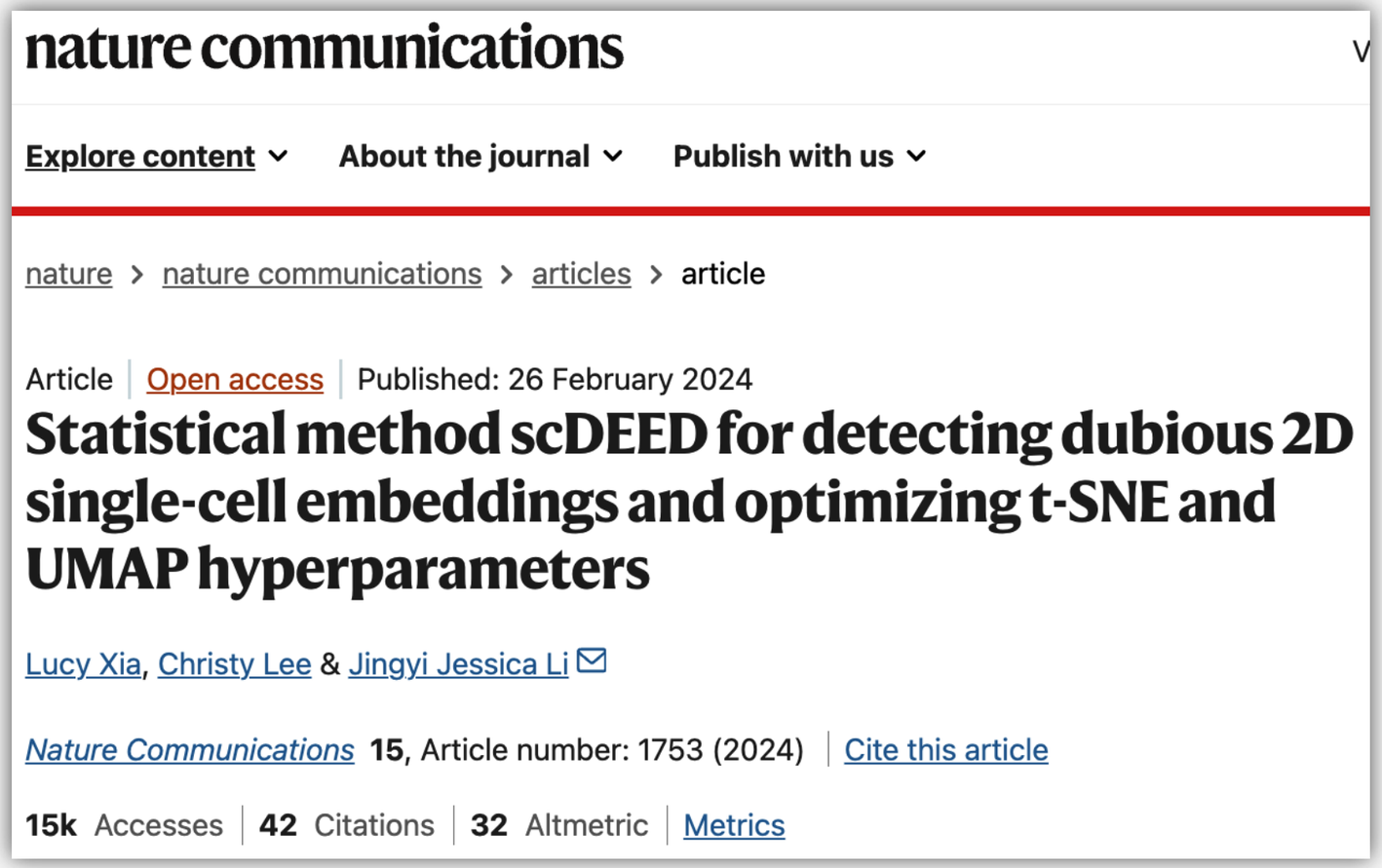
Nature Methods **21**, 930–933 (2024) | Cite this article

Dimension reduction helps to visualize high-dimensional datasets. These tools should be used thoughtfully and with tuned parameters. Sometimes, these methods take a second thought.

Example 1: dubious t-SNE/UMAP embeddings?

Q: Is a cell's embedding dubious or trustworthy?

A: Examine the cell's neighbors before and after embedding



nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature communications > articles > article

Article | Open access | Published: 26 February 2024

Statistical method scDEED for detecting dubious 2D single-cell embeddings and optimizing t-SNE and UMAP hyperparameters

Lucy Xia, Christy Lee & Jingyi Jessica Li

Nature Communications **15**, Article number: 1753 (2024) | Cite this article

15k Accesses | **42** Citations | **32** Altmetric | Metrics

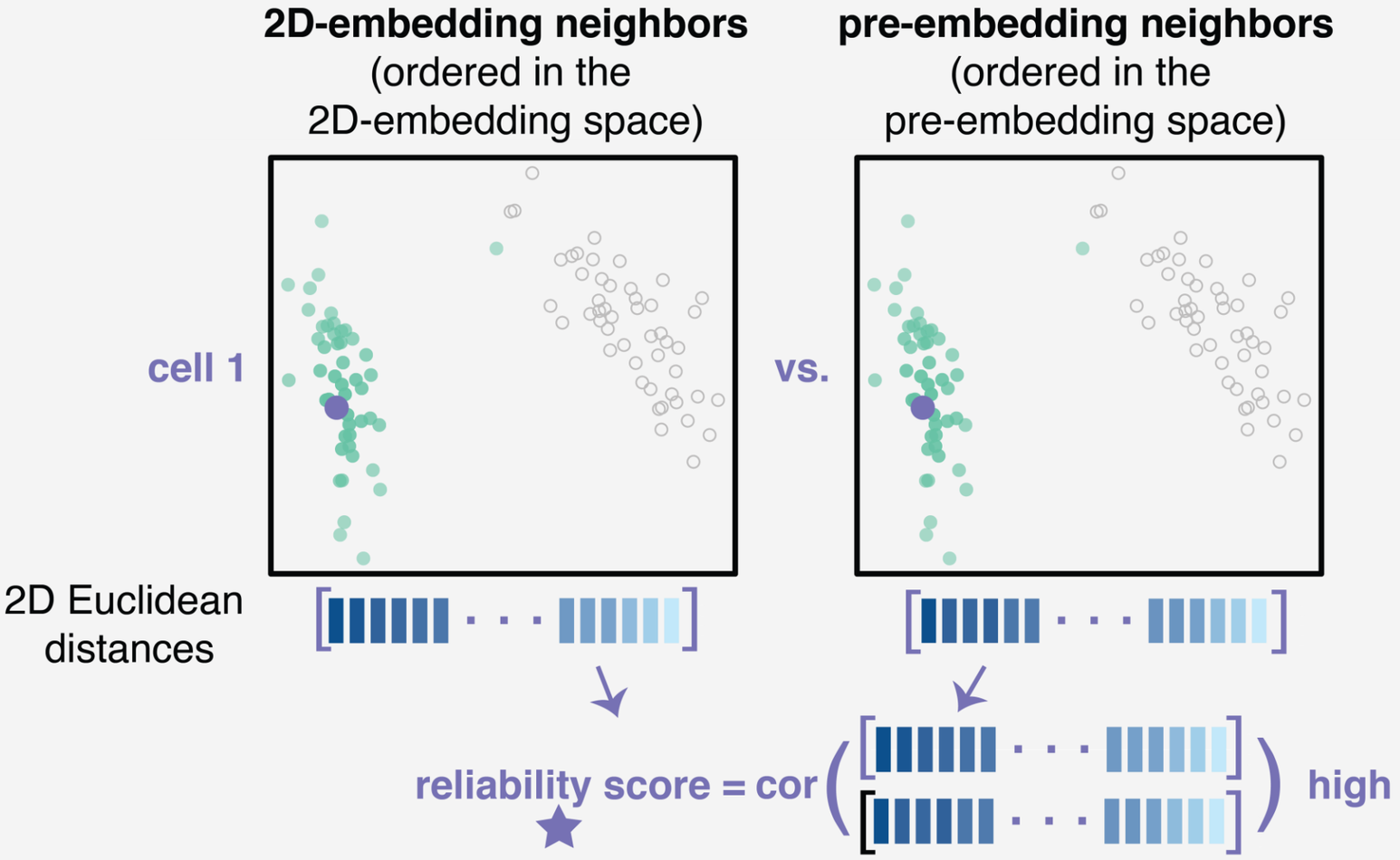
Lucy Xia
(HKUST)

Christy Lee
(JSB)

Example 1: dubious t-SNE/UMAP embeddings?

scDEED intuition

A trustworthy cell embedding



pre-embedding neighbors

defined in the PC space

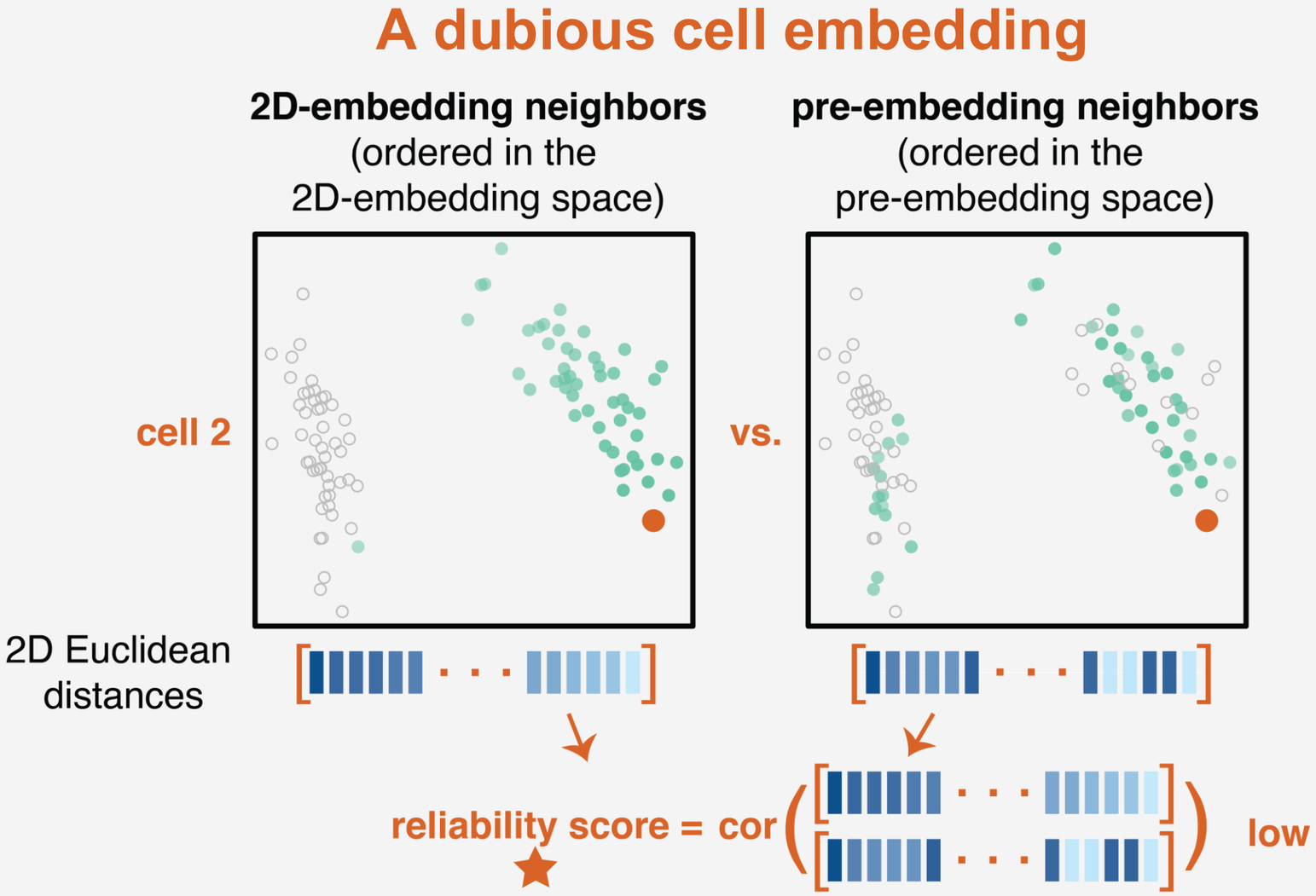
neighborhood size

= half of all cells

“mid-range distances”

Example 1: dubious t-SNE/UMAP embeddings?

scDEED intuition



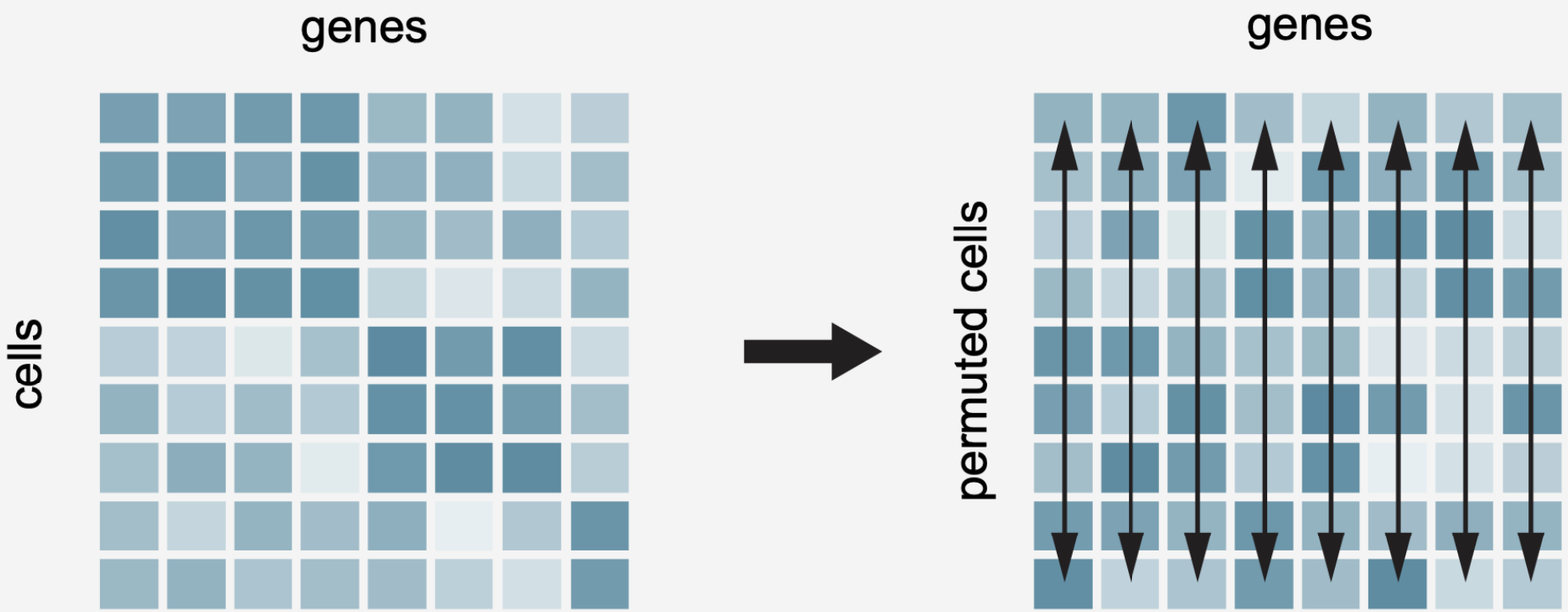
Example 1: dubious t-SNE/UMAP embeddings?

Q: What is the null hypothesis?

A: A cell's neighbors are random after embedding.

Q: How to obtain such a case?

A: Permutation.



within-gene permutation

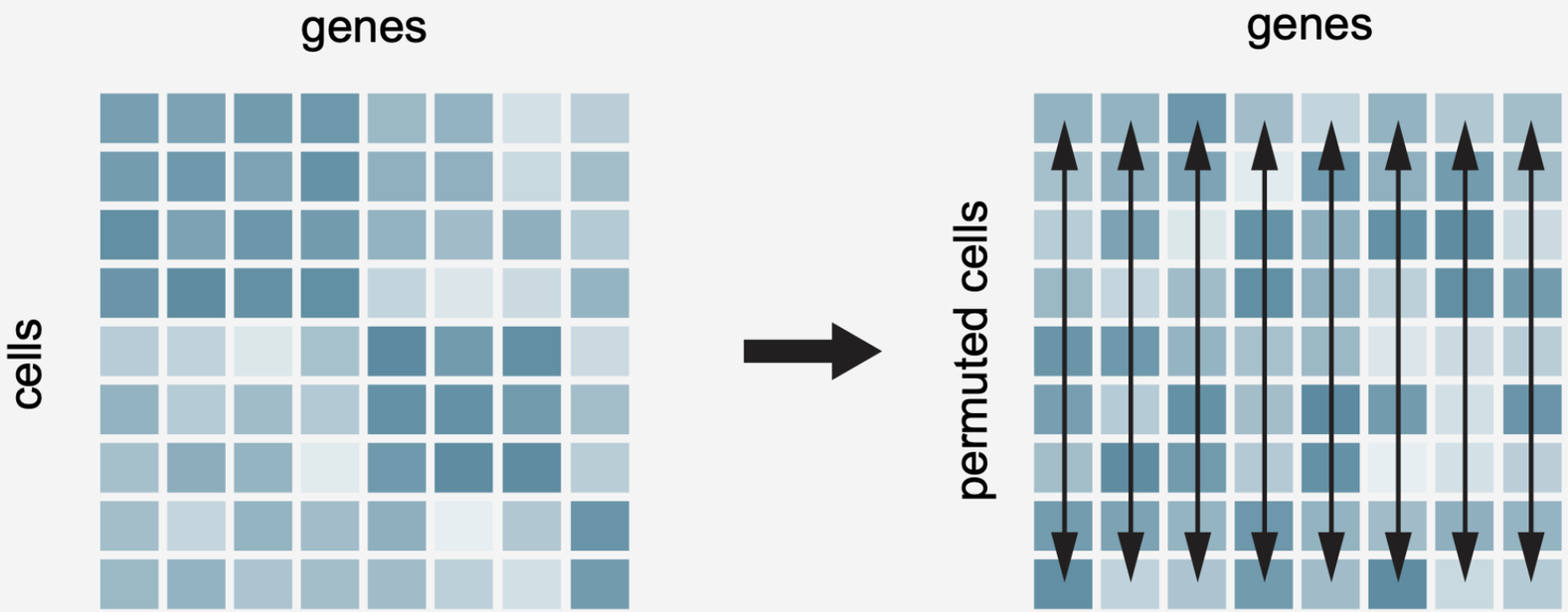
Example 1: dubious t-SNE/UMAP embeddings?

Q: What is preserved by within-gene permutation?

A: Every gene's distribution.

Q: What is not preserved?

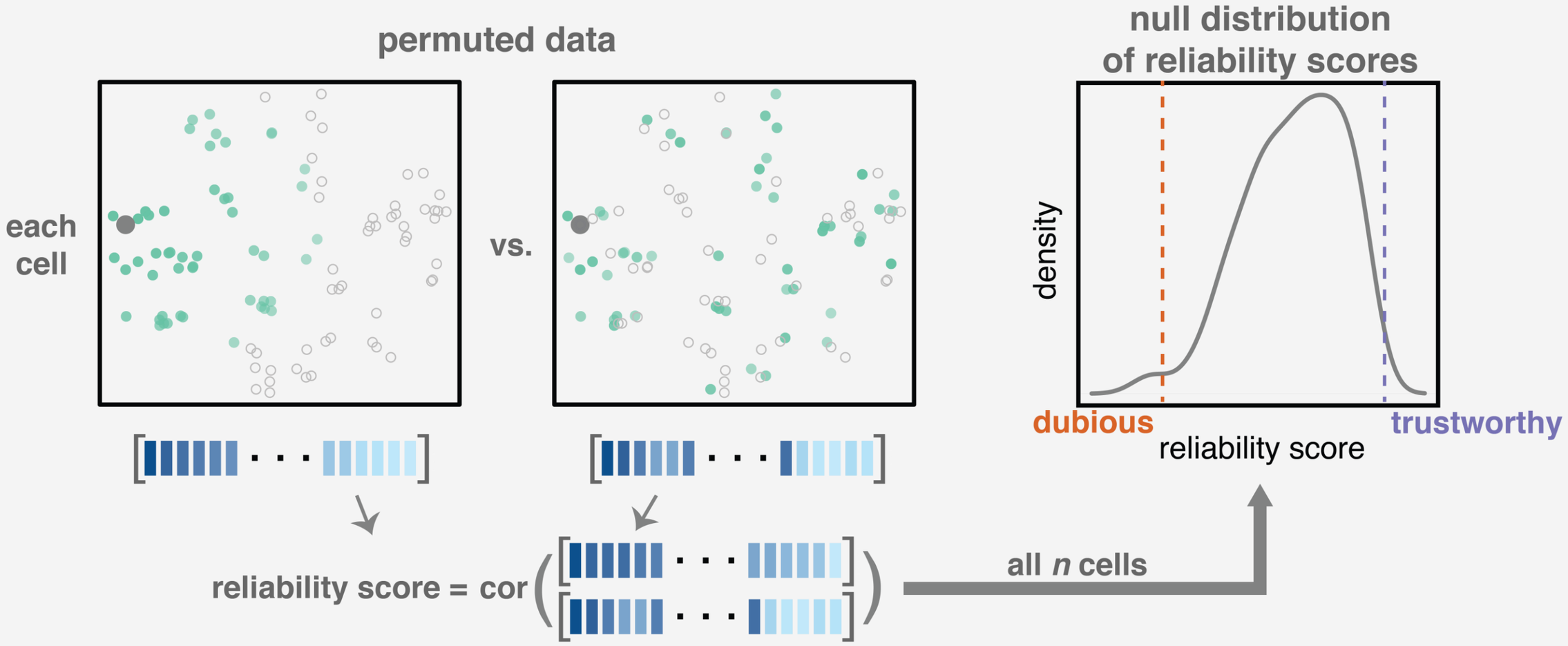
A: Gene-gene correlations and cell-cell relationships.



within-gene permutation

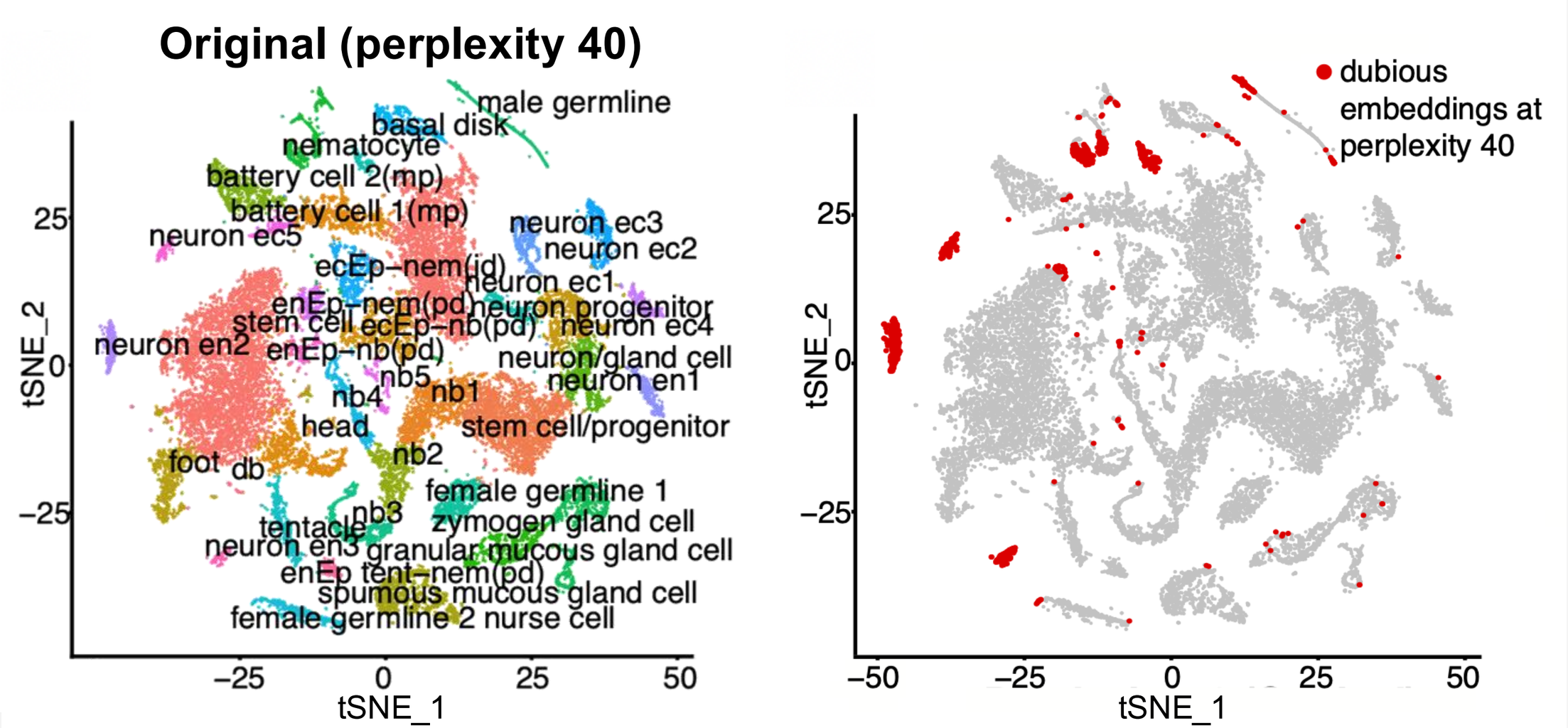
Example 1: dubious t-SNE/UMAP embeddings?

Permuted cells are exchangeable → A cell's neighbors are random



Example 1: dubious t-SNE/UMAP embeddings?

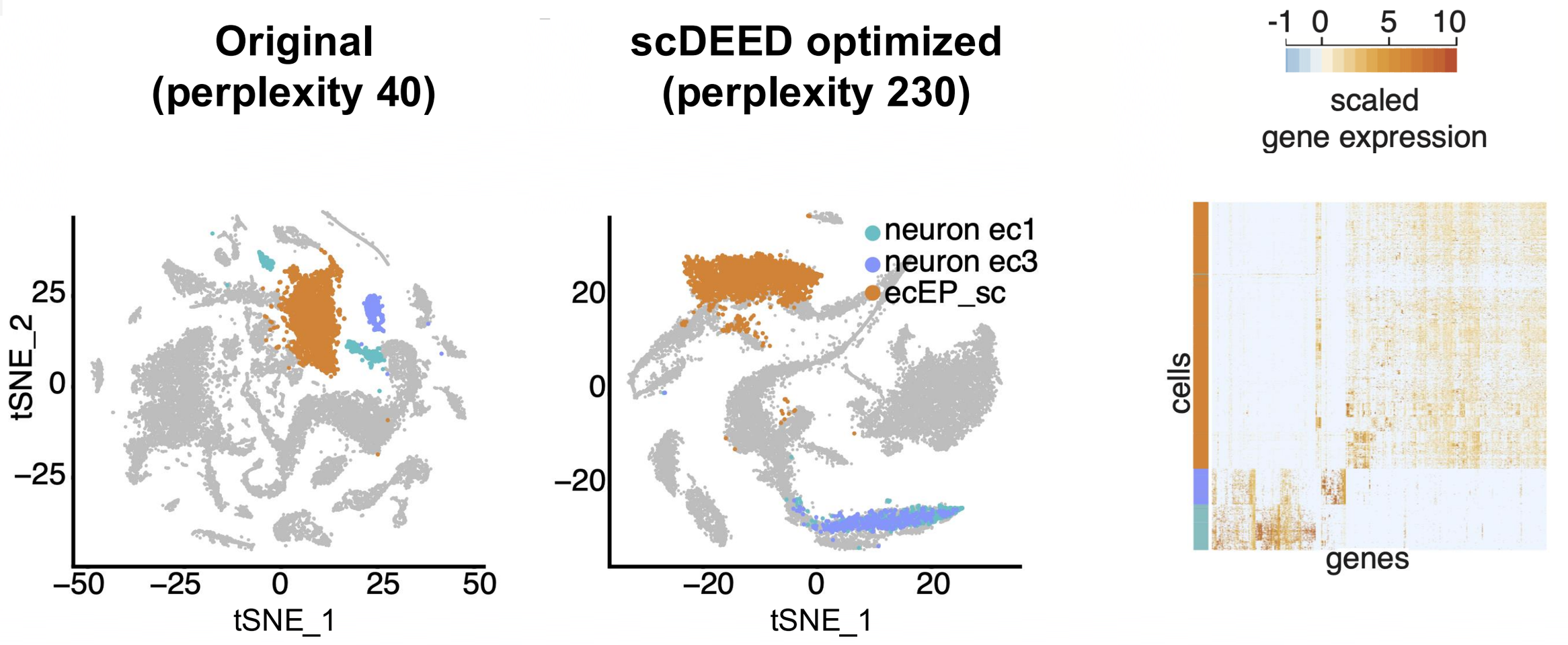
scDEED detects dubious embeddings



Hydra single-cell RNA-seq data [Siebert et al., *Science*, 2019]

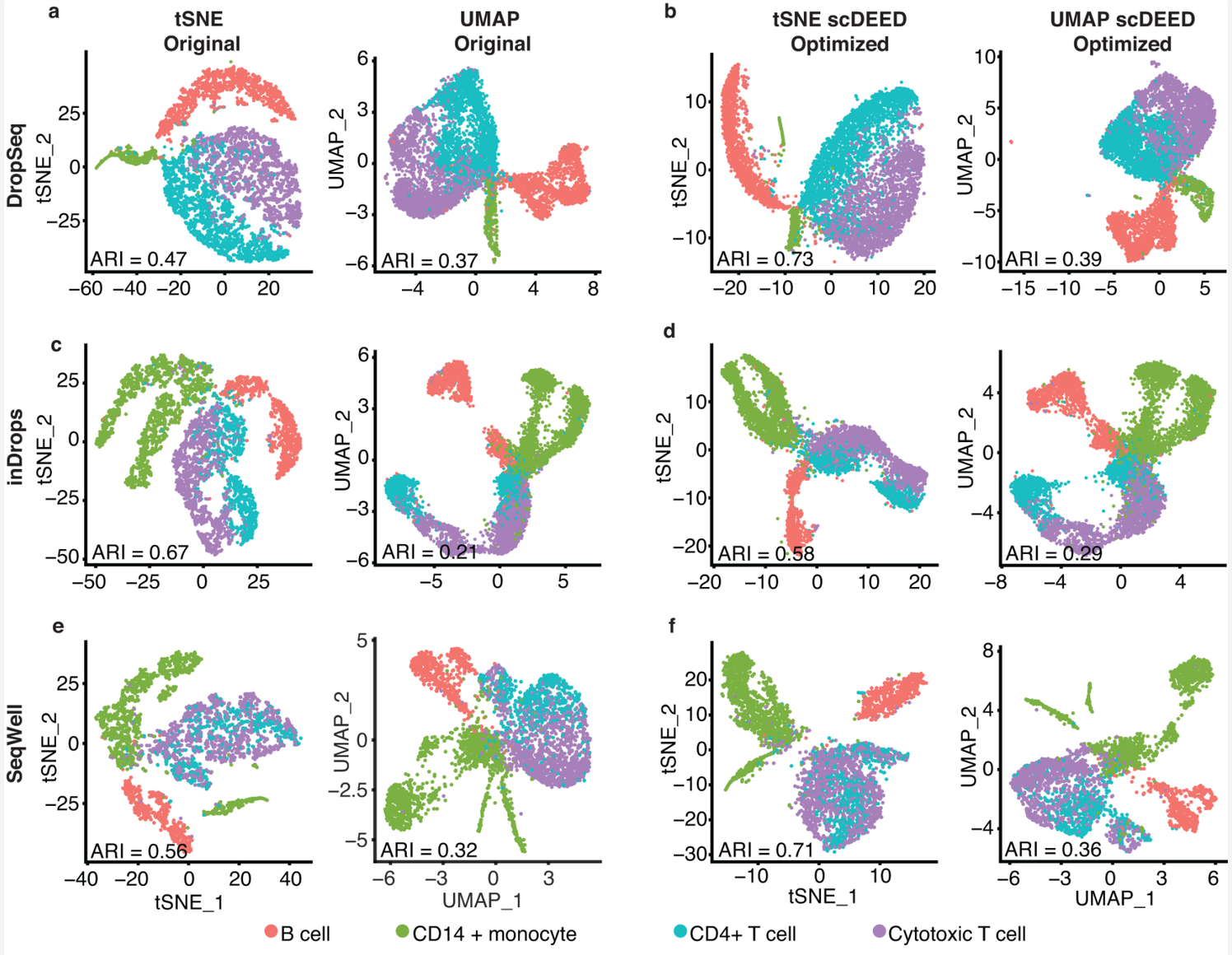
Example 1: dubious t-SNE/UMAP embeddings?

scDEED optimizes hyperparameters by minimizing dubious embeddings



Example 1: dubious t-SNE/UMAP embeddings?

scDEED enhances the consistency between t-SNE and UMAP



Two examples where permutation helps

1. Single-cell data **visualization**

Statistical method scDEED for detecting dubious 2D single-cell embeddings and optimizing t-SNE and UMAP hyperparameters

Lucy Xia, Christy Lee & Jingyi Jessica Li

Nature Communications **15**, Article number: 1753 (2024) | Cite this article

2. Aggregating single cells into **metacells**

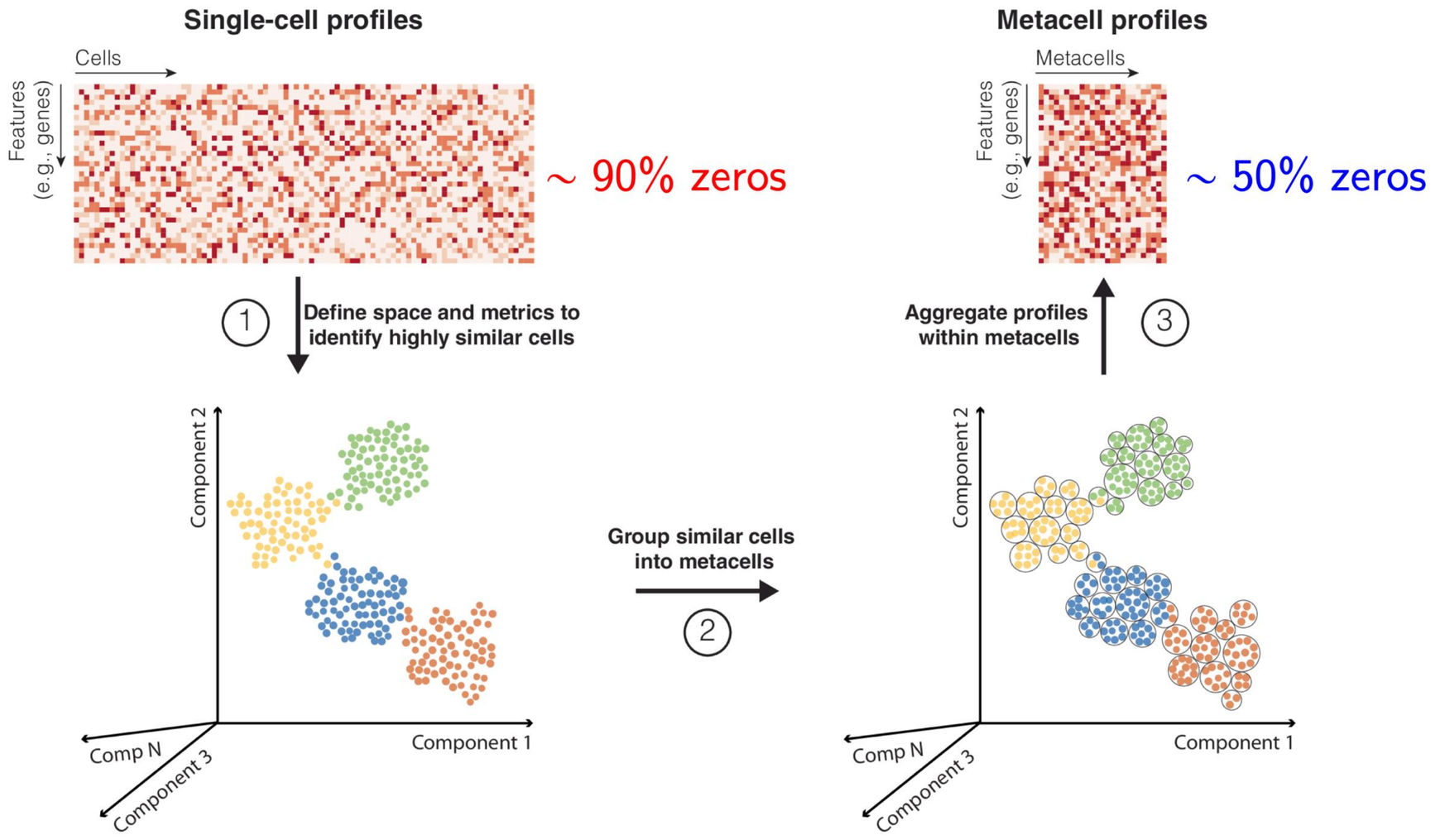
mcRigor: a statistical method to enhance the rigor of metacell partitioning in single-cell data analysis

Pan Liu & Jingyi Jessica Li

Nature Communications **16**, Article number: 8602 (2025) | Cite this article

Example 2: aggregating single cells into metacells

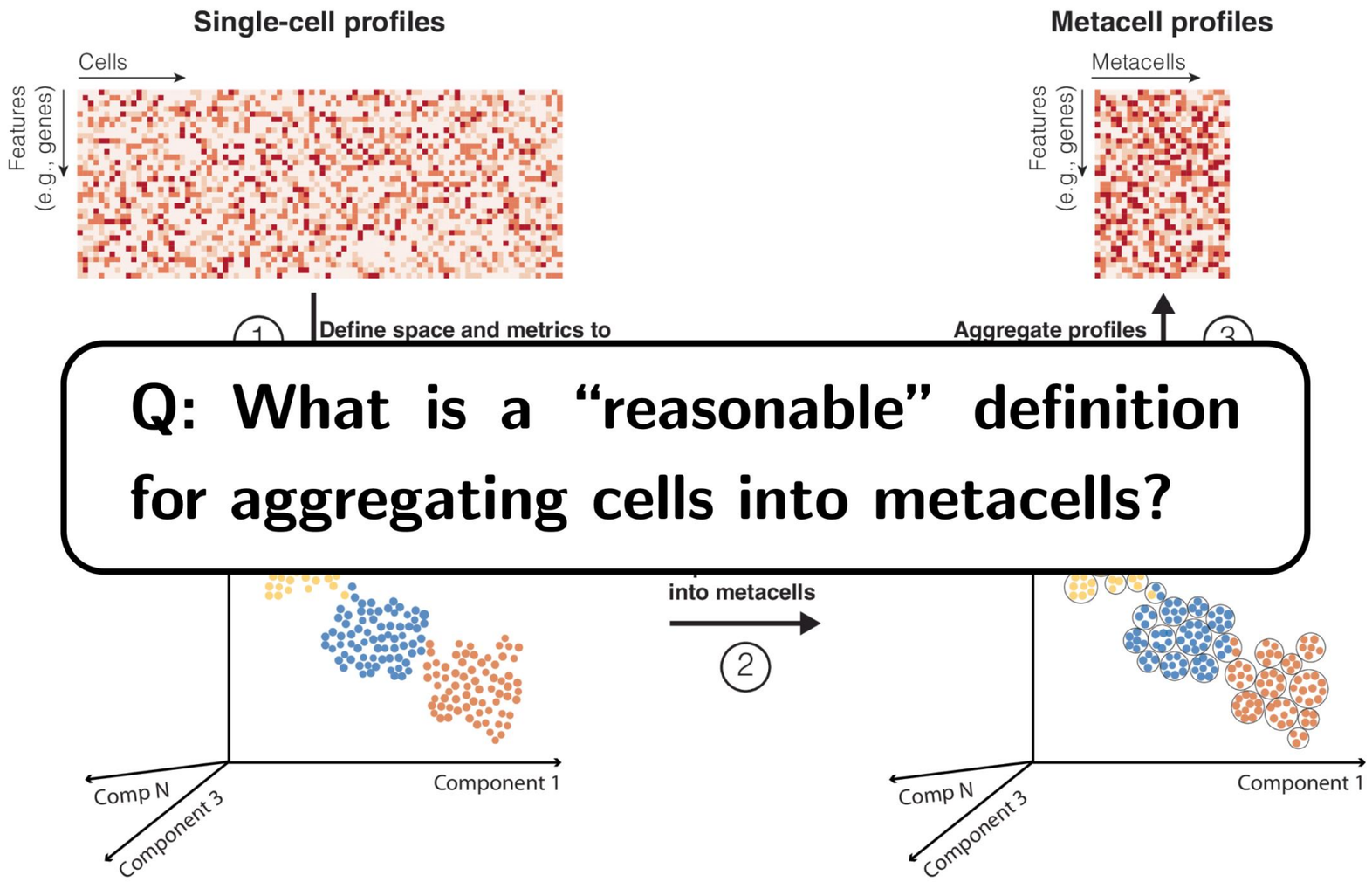
Metacell: a heuristic solution to the sparsity issue in single-cell data



Bilous, M., et al. "Building and analyzing metacells in single-cell genomics data." Molecular Systems Biology (2024): 1-23.

Example 2: aggregating single cells into metacells

Metacell: a heuristic solution to the sparsity issue in single-cell data



Q: What is a "reasonable" definition for aggregating cells into metacells?

Bilous, Mariia, et al. "Building and analyzing metacells in single-cell genomics data." Molecular Systems Biology (2024): 1-23.

Example 2: aggregating single cells into metacells

Metacell methods and applications

Metacell Methods

Method | Open access | Published: 11 October 2019

MetaCell: analysis of single-cell RNA-seq data using *K*-nn graph partitions

Yael Baran, Akhiad Bercovich, Arnau Sebe-Pedros, Yaniv Lubling, Amir Giladi, Elad Chomsky, Zohar Meir, Michael Hoichman, Aviezer Lifshitz & Amos Tanay

Genome Biology **20**, Article number: 206 (2019) | Cite this article

40k Accesses | **163** Citations | **46** Altmetric | Metrics

MetaCell

Method | Open access | Published: 19 April 2022

Metacell-2: a divide-and-conquer metacell algorithm for scalable scRNA-seq analysis

Oren Ben-Kiki, Akhiad Bercovich, Aviezer Lifshitz & Amos Tanay

Genome Biology **23**, Article number: 100 (2022) | Cite this article

10k Accesses | **19** Citations | **27** Altmetric | Metrics

MetaCell-2

Research article | Open access | Published: 13 August 2022

Metacells untangle large and complex single-cell transcriptome networks

Mariia Bilous, Loc Tran, Chiara Cianciaruso, Aurélie Gabriel, Hugo Michel, Santiago J. Carmona, Mikael J. Pittet & David Gfeller

BMC Bioinformatics **23**, Article number: 336 (2022) | Cite this article

7314 Accesses | **15** Citations | **39** Altmetric | Metrics

SuperCell

Article | Open access | Published: 27 March 2023

SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data

Sitara Persad, Zi-Ning Choo, Christine Dien, Noor Sohail, Ignas Masilionis, Ronan Chaligné, Tal Nawy, Chrysothemis C. Brown, Roshan Sharma, Itsik Pe'er, Manu Setty & Dana Pe'er

Nature Biotechnology **41**, 1746–1757 (2023) | Cite this article

46k Accesses | **27** Citations | **116** Altmetric | Metrics

SEACells

Metacell Applications

Resource | Published: 18 June 2018

Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis

Amir Giladi, Franziska Paul, Yoni Herzog, Yaniv Lubling, Assaf Weiner, Ido Yofe, Diego Jaitin, Nina Cabezas-Wallscheid, Regine Dress, Florent Ginhoux, Andreas Trumpp, Amos Tanay & Ido Amit

Nature Cell Biology **20**, 836–846 (2018) | Cite this article

25k Accesses | **224** Citations | **60** Altmetric | Metrics

Letter | Published: 18 July 2018

Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells

Chamutal Bornstein, Shir Nevo, Amir Giladi, Noam Kadouri, Marie Pouzolles, François Gerbe, Eyal David, Alice Machado, Anna Chuprin, Beáta Tóth, Ori Goldberg, Shalev Itzkovitz, Naomi Taylor, Philippe Jay, Valérie S. Zimmermann, Jakub Abramson & Ido Amit

Nature **559**, 622–626 (2018) | Cite this article

24k Accesses | **198** Citations | **74** Altmetric | Metrics

Article | Open access | Published: 24 March 2021

NASH limits anti-tumour surveillance in immunotherapy-treated HCC

Dominik Pfister, Nicolás Gonzalo Núñez, Roser Pinyol, Olivier Govaere, Matthias Pinter, Marta Szydlowska, Revant Gupta, Mengjie Qiu, Aleksandra Deczkowska, Assaf Weiner, Florian Müller, Ankit Sinha, Ekaterina Friebel, Thomas Engleitner, Daniela Lenggenhager, Anja Moncsek, Danijela Heide, Kristin Stirm, Jan Kosla, Eleni Kotsiliti, Valentina Leone, Michael Dudek, Suhail Yousuf, Donato Inverso, ... Mathias Heikenwalder

+ Show authors

Nature **592**, 450–456 (2021) | Cite this article

114k Accesses | **671** Citations | **250** Altmetric | Metrics

Resource | Open access | Published: 23 December 2021

Temporal single-cell tracing reveals clonal revival and expansion of precursor exhausted T cells during anti-PD-1 therapy in lung cancer

Baolin Liu, Xueda Hu, Kaichao Feng, Ranran Gao, Zhiqiang Xue, Sujie Zhang, Yuanyuan Zhang, Emily Corse, Yi Hu, Weidong Han & Zemin Zhang

Nature Cancer **3**, 108–121 (2022) | Cite this article

68k Accesses | **160** Citations | **67** Altmetric | Metrics

Example 2: aggregating single cells into metacells

Metacell methods and applications

Metacell Methods

Method | Open access | Published: 11 October 2019

MetaCell: analysis of single-cell RNA-seq data using *K*-nn graph partitions

Yael Baran, Akhiad Bercovich, Arnau Sebe-Pedros, Yaniv Lubling, Amir Giladi, Elad Chomsky, Zohar Meir, Michael Hoichman, Aviezer Lifshitz & Amos Tanay

Genome Biology **20**, Article number: 206 (2019) | Cite this article

40k Accesses | **163** Citations | **46** Altmetric | Metrics

Method | Open access | Published: 19 April 2022

Metacell-2: a divide-and-conquer metacell algorithm for scalable scRNA-seq analysis

Oren Ben-Kiki, Akhiad Bercovich, Aviezer Lifshitz & Amos Tanay

Genome Biology **23**, Article number: 100 (2022) | Cite this article

10k Accesses | **19** Citations | **27** Altmetric | Metrics

Research article | Open access | Published: 13 August 2022

Metacells untangle large and complex single-cell transcriptome

Mariia Bilous, Loc Tran, Ch
Pittet & David Gfeller

BMC Bioinformatics **23**, Artic

7314 Accesses | **15** Citations | **39** Altmetric | Metrics

Article | Open access | Published: 27 March 2023

SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data

Sitara Persad, Zi-Ning Choo, Christine Dien, Noor Sohail, Ignas Masilionis, Ronan Chaligné, Tal Nawy, Chrysothemis C. Brown, Roshan Sharma, Itsik Pe'er, Manu Setty & Dana Pe'er

Nature Biotechnology **41**, 1746–1757 (2023) | Cite this article

46k Accesses | **27** Citations | **116** Altmetric | Metrics

No consensus on metacell definition

Metacell Applications

Resource | Published: 18 June 2018

Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis

Amir Giladi, Franziska Paul, Yoni Herzog, Yaniv Lubling, Assaf Weiner, Ido Yofe, Diego Jaitin, Nina Cabezas-Wallscheid, Regine Dress, Florent Ginhoux, Andreas Trumpp, Amos Tanay & Ido Amit

Nature Cell Biology **20**, 836–846 (2018) | Cite this article

25k Accesses | **224** Citations | **60** Altmetric | Metrics

Letter | Published: 18 July 2018

Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells

Chamutal Bornstein, Shir Nevo, Amir Giladi, Noam Kadouri, Marie Pouzolles, François Gerbe, Eyal David, Alice Machado, Anna Chuprin, Beáta Tóth, Ori Goldberg, Shalev Itzkovitz, Naomi Taylor, Philippe Jay, Valérie S. Zimmermann, Jakub Abramson & Ido Amit

Nature **559**, 622–626 (2018) | Cite this article

24k Accesses | **198** Citations | **74** Altmetric | Metrics

Article | Open access | Published: 24 March 2021

NASH limits anti-tumour surveillance in immunotherapy-treated HCC

er Govaere, Matthias Pinter, Marta Szydlowska,
Weiner, Florian Müller, Ankit Sinha, Ekaterina
oncsek, Danijela Heide, Kristin Stirm, Jan Kosla,
uf, Donato Inverso, ... Mathias Heikenwalder

Nature **592**, 450–456 (2021) | Cite this article

114k Accesses | **671** Citations | **250** Altmetric | Metrics

Resource | Open access | Published: 23 December 2021

Temporal single-cell tracing reveals clonal revival and expansion of precursor exhausted T cells during anti-PD-1 therapy in lung cancer

Baolin Liu, Xueda Hu, Kaichao Feng, Ranran Gao, Zhiqiang Xue, Sujie Zhang, Yuanyuan Zhang, Emily Corse, Yi Hu, Weidong Han & Zemin Zhang

Nature Cancer **3**, 108–121 (2022) | Cite this article

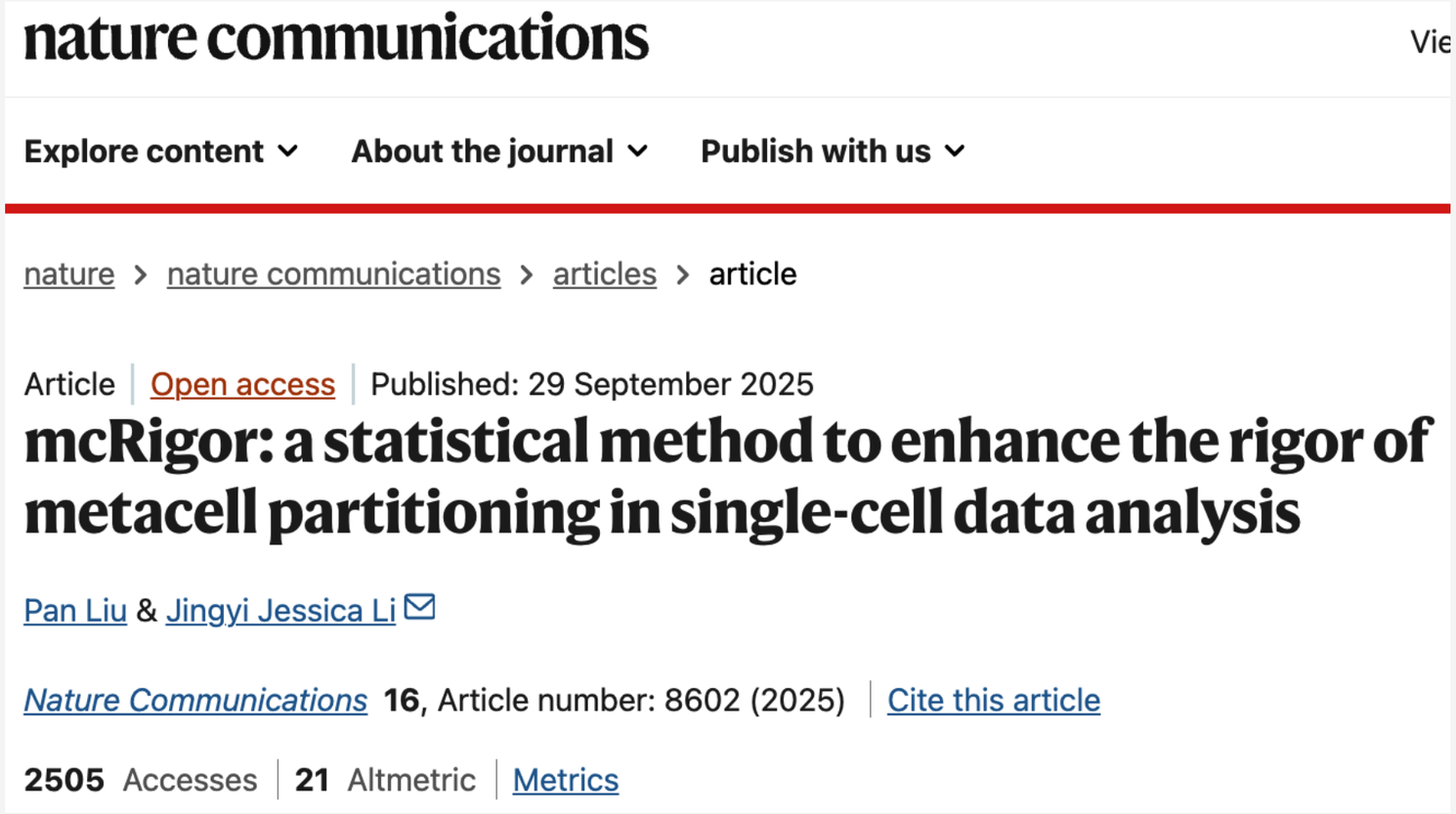
68k Accesses | **160** Citations | **67** Altmetric | Metrics

Example 2: aggregating single cells into metacells

Q: How to define a “metacell”?

Q: How to detect dubious metacells?

Q: How to optimize metacell partitioning?



nature communications

Vie

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature communications > articles > article

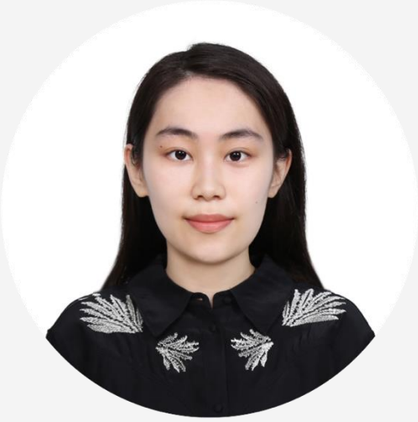
Article | Open access | Published: 29 September 2025

mcRigor: a statistical method to enhance the rigor of metacell partitioning in single-cell data analysis

Pan Liu & Jingyi Jessica Li

Nature Communications **16**, Article number: 8602 (2025) | Cite this article

2505 Accesses | **21** Altmetric | Metrics



Pan Liu

(JSB)

Accepted by *RECOMB* 2025

Example 2: aggregating single cells into metacells

Q: How to define a “metacell”?

The first publication that proposed the **“metacell”** concept

Method | Open access | Published: 11 October 2019

MetaCell: analysis of single-cell RNA-seq data using *K*-nn graph partitions

Yael Baran, Akhiad Bercovich, Arnau Sebe-Pedros, Yaniv Lubling, Amir Giladi, Elad Chomsky, Zohar Meir, Michael Hoichman, Aviezer Lifshitz & Amos Tanay

Genome Biology **20**, Article number: 206 (2019) | Cite this article

40k Accesses | **163** Citations | **46** Altmetric | Metrics

“A homogeneous collection of single-cell profiles that could have been resampled from the same original cell.”

Example 2: aggregating single cells into metacells

Q: How to define a “metacell”?

The first publication that proposed the **“metacell”** concept

Method | Open access | Published: 11 October 2019

MetaCell: analysis of single-cell RNA-seq data using *K*-nn graph partitions

Yael Baran, Akhiad Bercovich, Arnau Sebe-Pedros, Yaniv Lubling, Amir Giladi, Elad Chomsky, Zohar Meir, Michael Hoichman, Aviezer Lifshitz & Amos Tanay

Genome Biology **20**, Article number: 206 (2019) | Cite this article

40k Accesses | **163** Citations | **46** Altmetric | Metrics

“A homogeneous collection of single-cell profiles that could have been resampled from the same original cell.”

$\Longrightarrow$ Variation within a metacell is attributed exclusively to measurement (technical) error

Example 2: aggregating single cells into metacells

Q: How to define “measurement error”?

Perspective | Published: 24 May 2021

Separating measurement and expression models clarifies confusion in single-cell RNA sequencing analysis

Abhishek Sarkar & Matthew Stephens

Nature Genetics **53**, 770–777 (2021) | Cite this article

Expression model: distribution of true expression levels

Measurement model: distribution of observed counts | true expression levels

Example 2: aggregating single cells into metacells

A statistical definition of "metacell"

"A homogeneous collection of single-cell profiles that could have been resampled from the same original cell."

$\Longrightarrow$ Variation within a metacell is attributed exclusively to measurement error

Two-layer observation model: Cell (observation) $i = 1, \ldots, n$
Feature $j = 1, \ldots, p$

Expression model: $\boldsymbol{\lambda}_i \sim \mathcal{F}(\cdot|\mathbf{x}_i)$

Measurement model: $y_{ij} \sim \mathcal{G}(y_{i+}\lambda_{ij})$

$$\Downarrow$$

Statistical definition: A **metacell** is a group of single cells that share the same $\boldsymbol{\lambda}$

Example 2: aggregating single cells into metacells

A statistical definition of "metacell"

"A homogeneous collection of single-cell profiles that could have been resampled from the same original cell."

$\Longrightarrow$ Variation within a metacell is attributed exclusively to measurement error

Two-layer observation model: Cell (observation) $i = 1, \ldots, n$

Feature $j = 1, \ldots, p$

Expression model: $\boldsymbol{\lambda}_i \sim \mathcal{F}(\cdot|\mathbf{x}_i)$

Measurement model: $y_{ij} \sim \mathcal{G}(y_{i+}\lambda_{ij})$

$$\Downarrow$$

Statistical definition: A **metacell** is a group of single cells that share the same $\boldsymbol{\lambda}$

$$\Downarrow$$

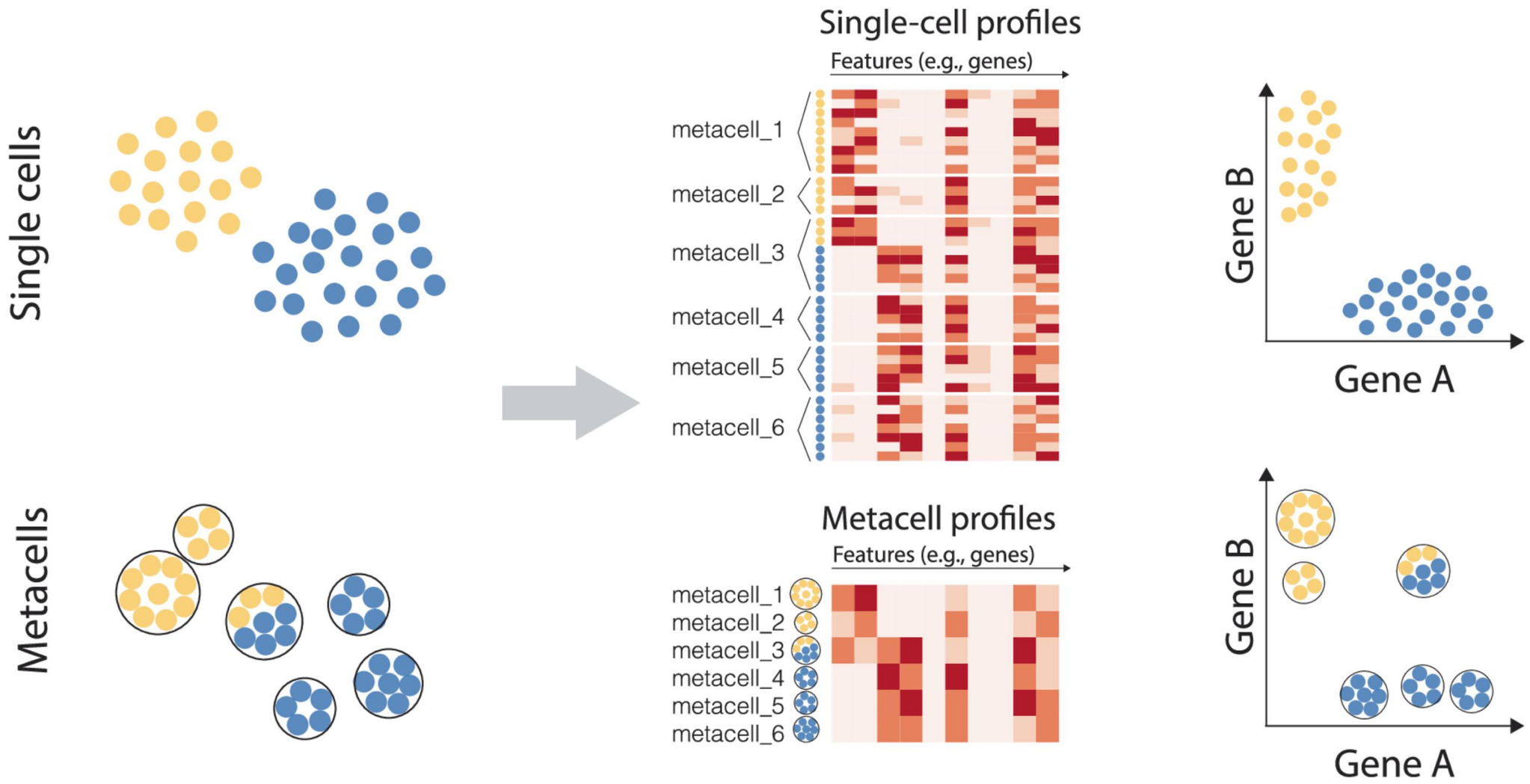
Satisfying this definition?

Yes: **trustworthy metacells** No: **dubious metacells**

A statistical problem

Example 2: aggregating single cells into metacells

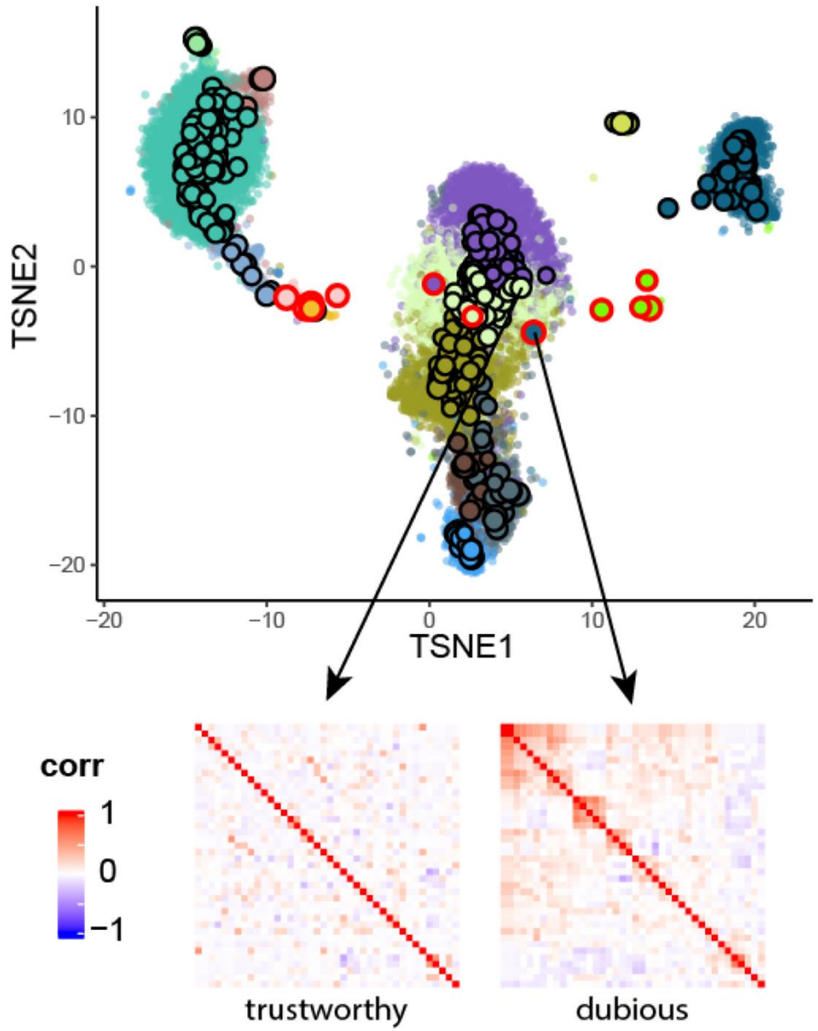
Dubious metacells can bias analysis



Bilous, M., et al. "Building and analyzing metacells in single-cell genomics data." Molecular Systems Biology (2024): 1-23.

Example 2: aggregating single cells into metacells

Our proposal: mcRigor

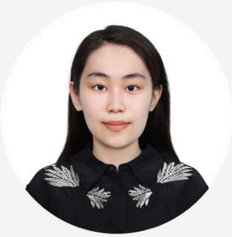


Goals: a statistical criterion to

- Identify dubious metacells consisting of single cells from different cell states
- Nominate the top-performing metacell method and optimize its hyperparameter

$$\text{granularity level } \gamma = \frac{\#\text{single cells}}{\#\text{metacells}}$$

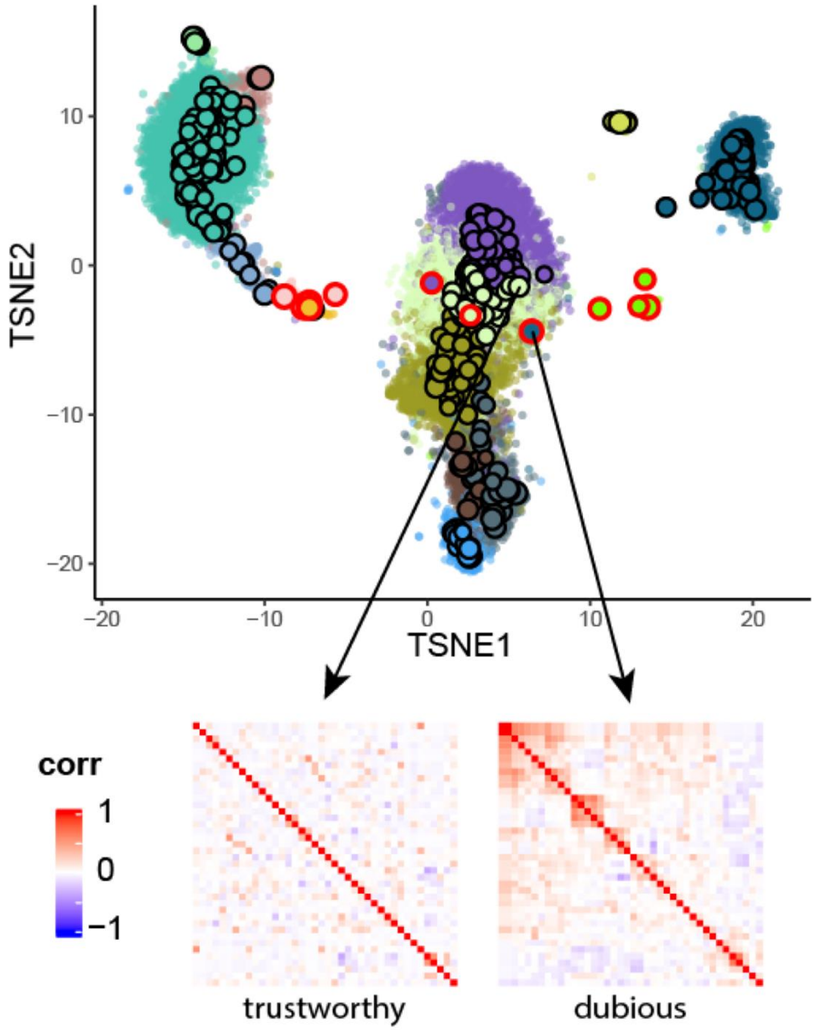
in a data-specific way



Pan Liu
(JSB)

Example 2: aggregating single cells into metacells

Our proposal: mcRigor



Goals: a statistical criterion to

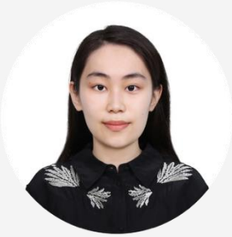
- Identify dubious metacells consisting of single cells from different cell states
- Nominate the top-performing metacell method and optimize its hyperparameter

$$\text{granularity level } \gamma = \frac{\#\text{single cells}}{\#\text{metacells}}$$

in a data-specific way

Intuition:

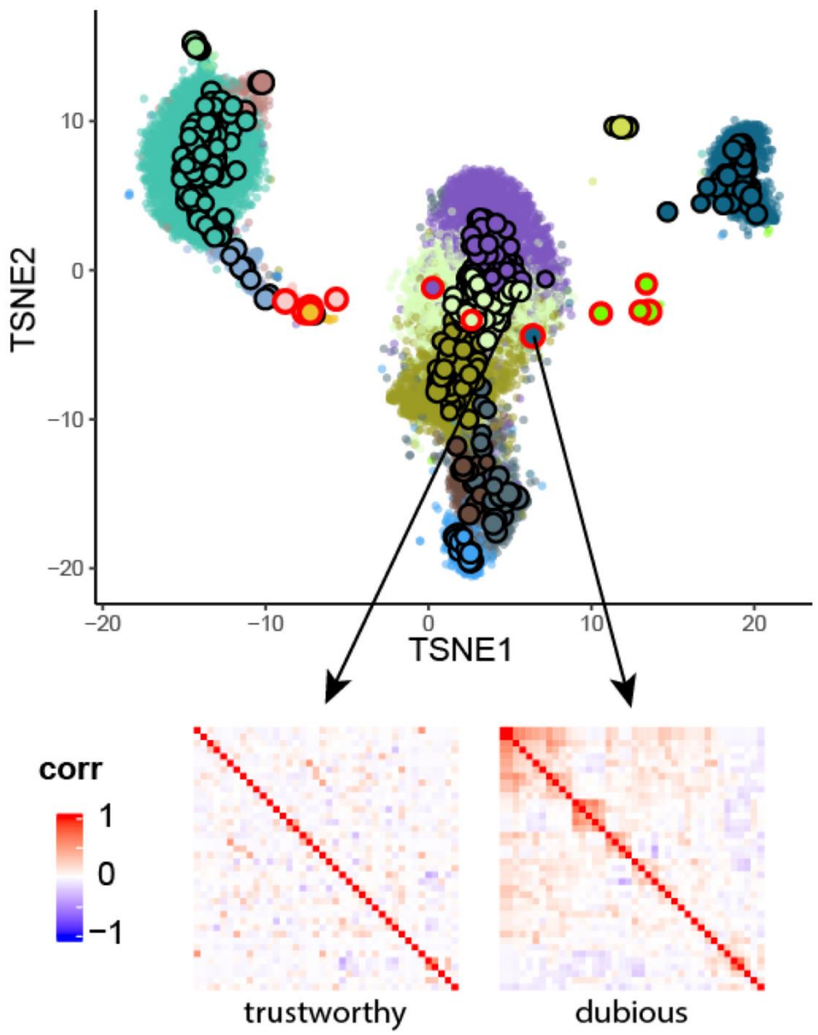
- Within a trustworthy metacell, features are approximately uncorrelated



Pan Liu
(JSB)

Example 2: aggregating single cells into metacells

Our proposal: mcRigor



Goals: a statistical criterion to

- Identify dubious metacells consisting of single cells from different cell states
- Nominate the top-performing metacell method and optimize its hyperparameter

$$\text{granularity level } \gamma = \frac{\#\text{single cells}}{\#\text{metacells}}$$

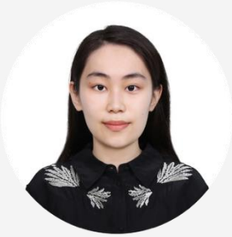
in a data-specific way

Intuition:

- Within a trustworthy metacell, features are approximately uncorrelated

Strategy:

- Per-metacell statistic **"mcDiv"**
- Cell-library-size-preserved permutation

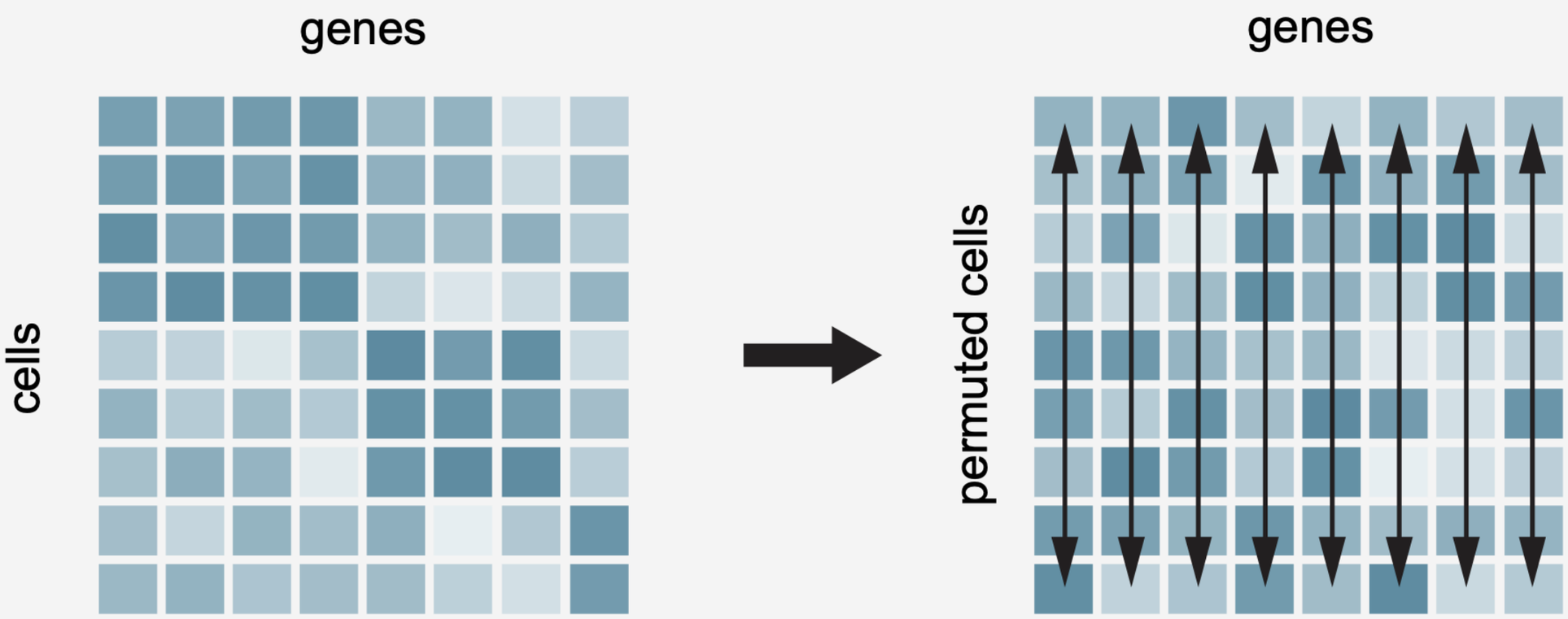


Pan Liu
(JSB)

Example 2: aggregating single cells into metacells

Q: Is within-gene permutation enough?

A: Genes become uncorrelated, but cell library sizes are gone.



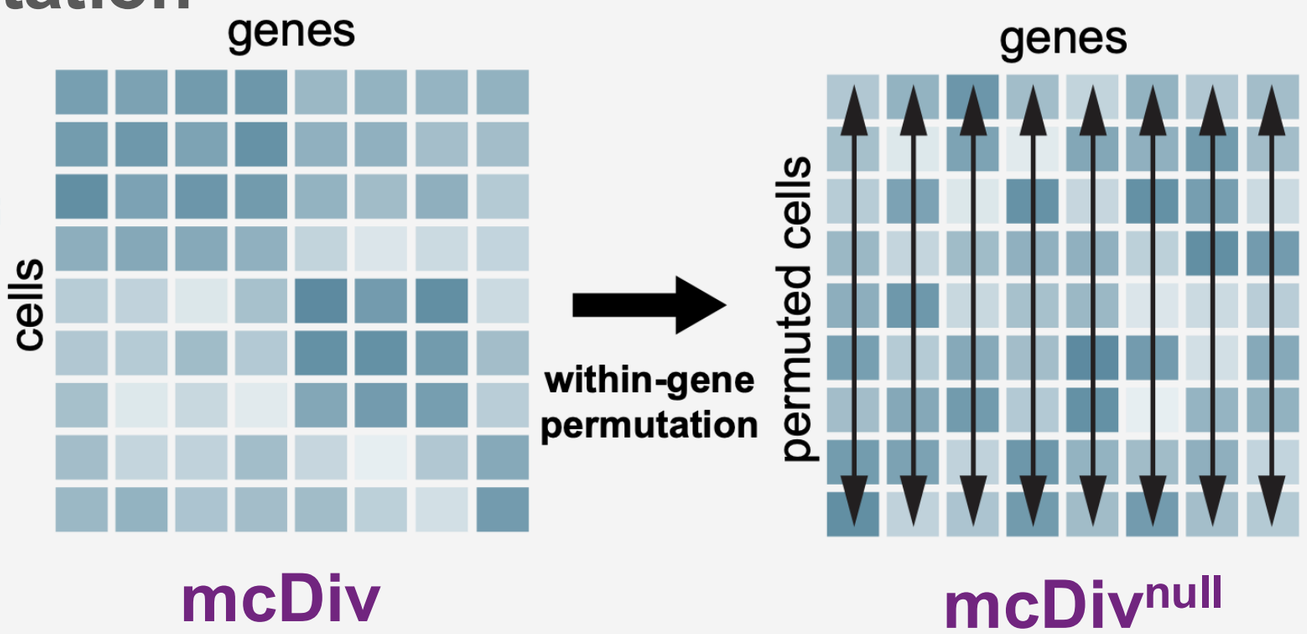
within-gene permutation

Example 2: aggregating single cells into metacells

Double permutation

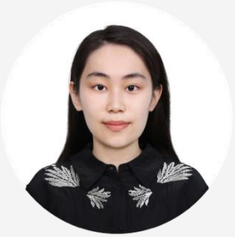
Within-gene permutation:

- **preserves genes marginal distributions**
- **removes gene correlations**
- **removes cell library sizes**



mcDiv **mcDiv^{null}**

X



Pan Liu
(JSB)

Example 2: aggregating single cells into metacells

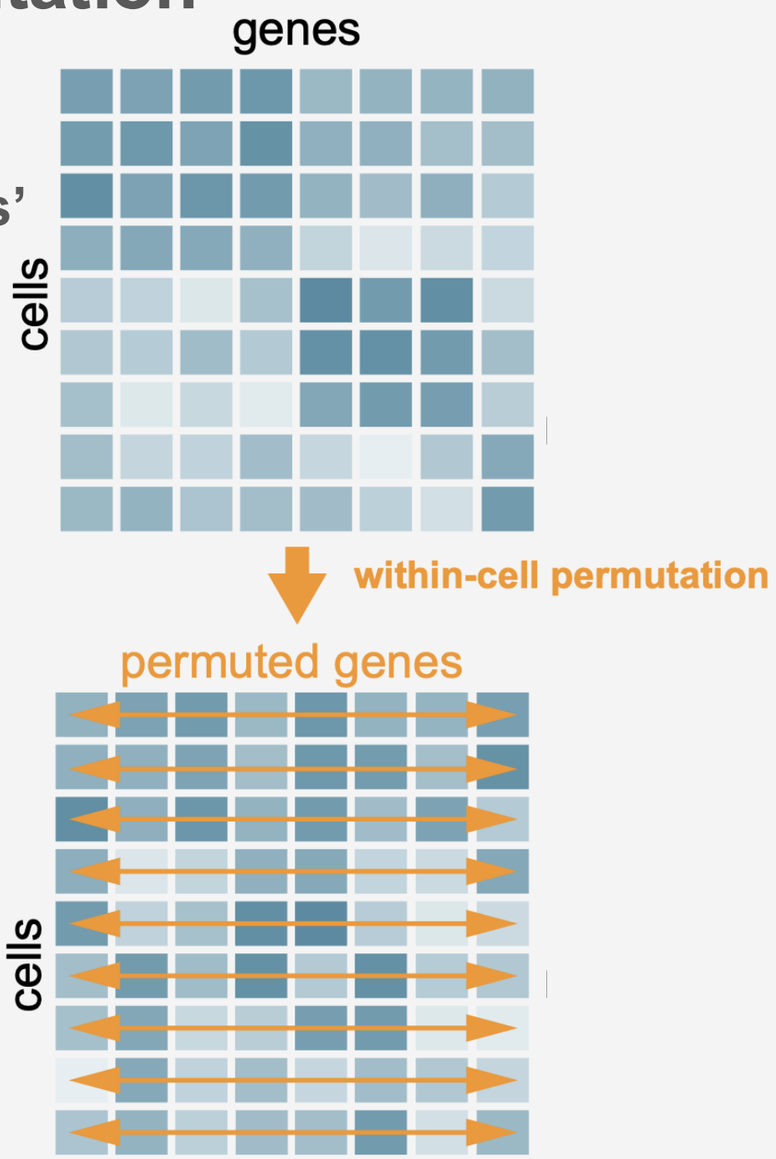
Double permutation

Within-gene permutation:

- preserves genes' marginal distributions
- removes gene correlations
- removes cell library sizes

Within-cell permutation:

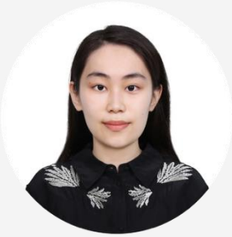
- preserves cell library sizes
- removes genes' marginal distributions



mcDiv

However, the genes are now different …

mcDiv^{null}



Pan Liu
(JSB)

Example 2: aggregating single cells into metacells

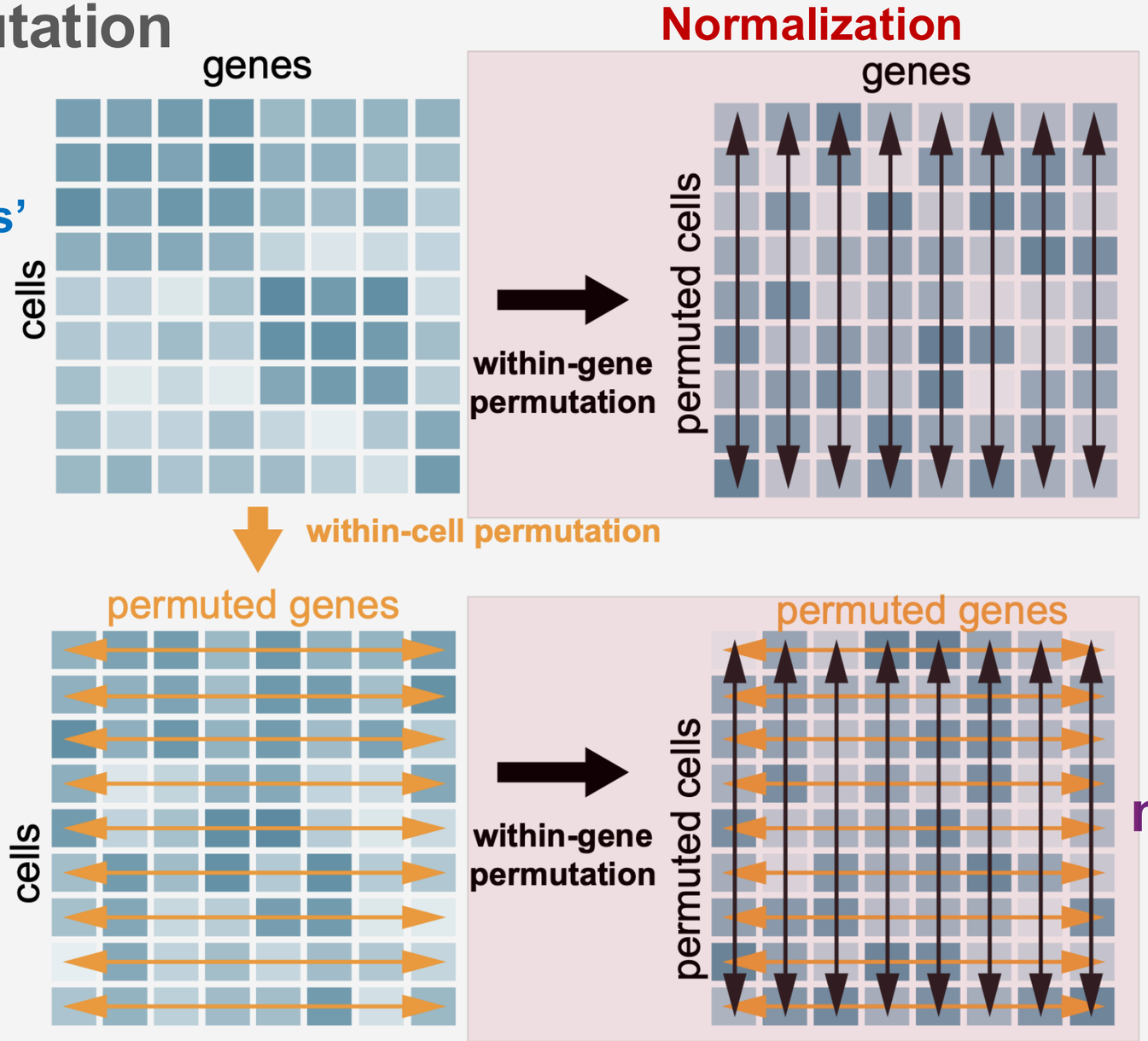
Double permutation

Within-gene permutation:

- **preserves genes' marginal distributions**
- removes gene correlations
- removes cell library sizes

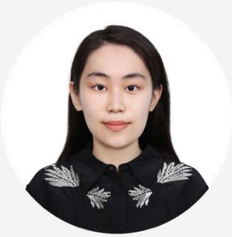
Within-cell permutation:

- **preserves cell library sizes**
- removes genes' marginal distributions



mcDiv

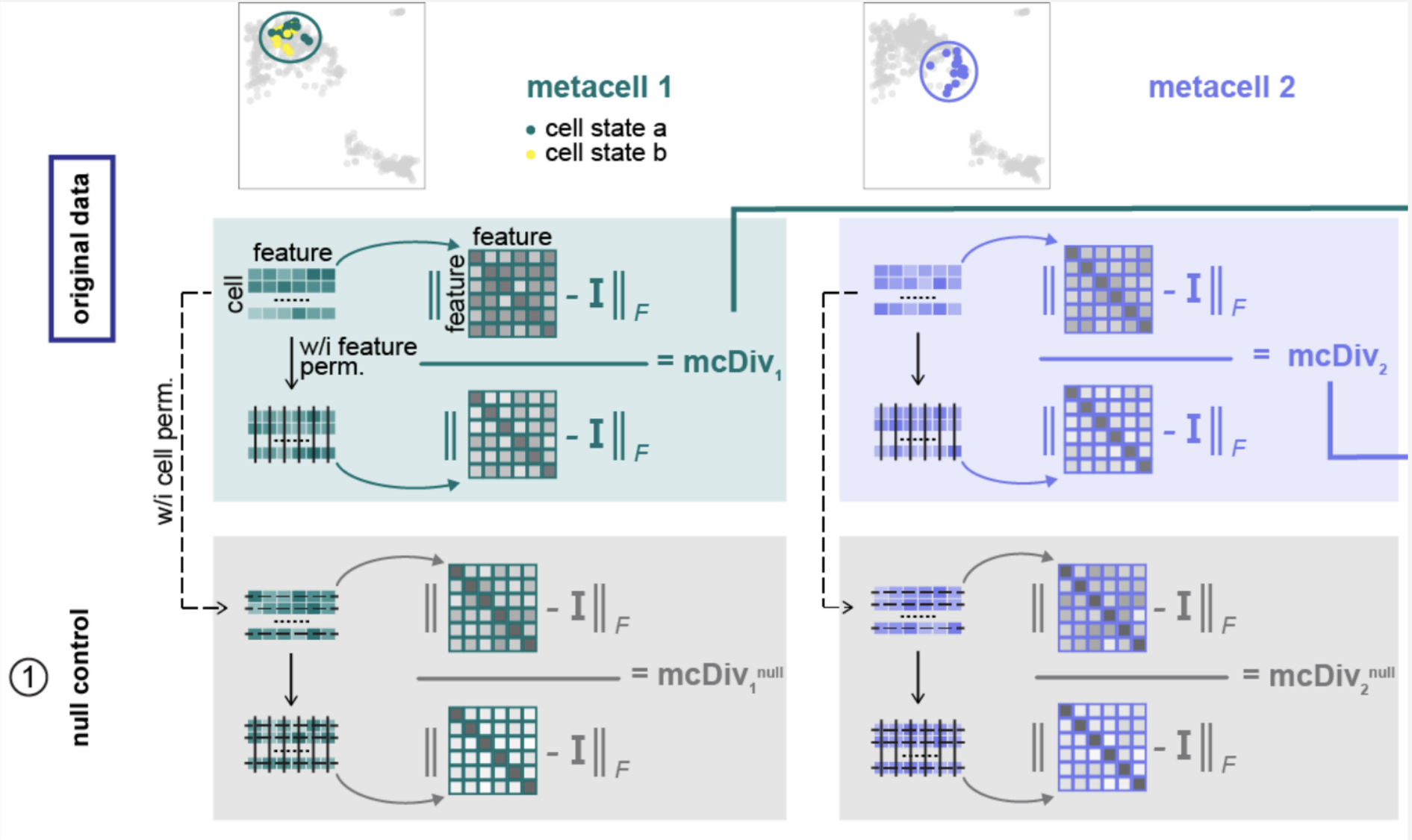
mcDiv^{null}



Pan Liu
(JSB)

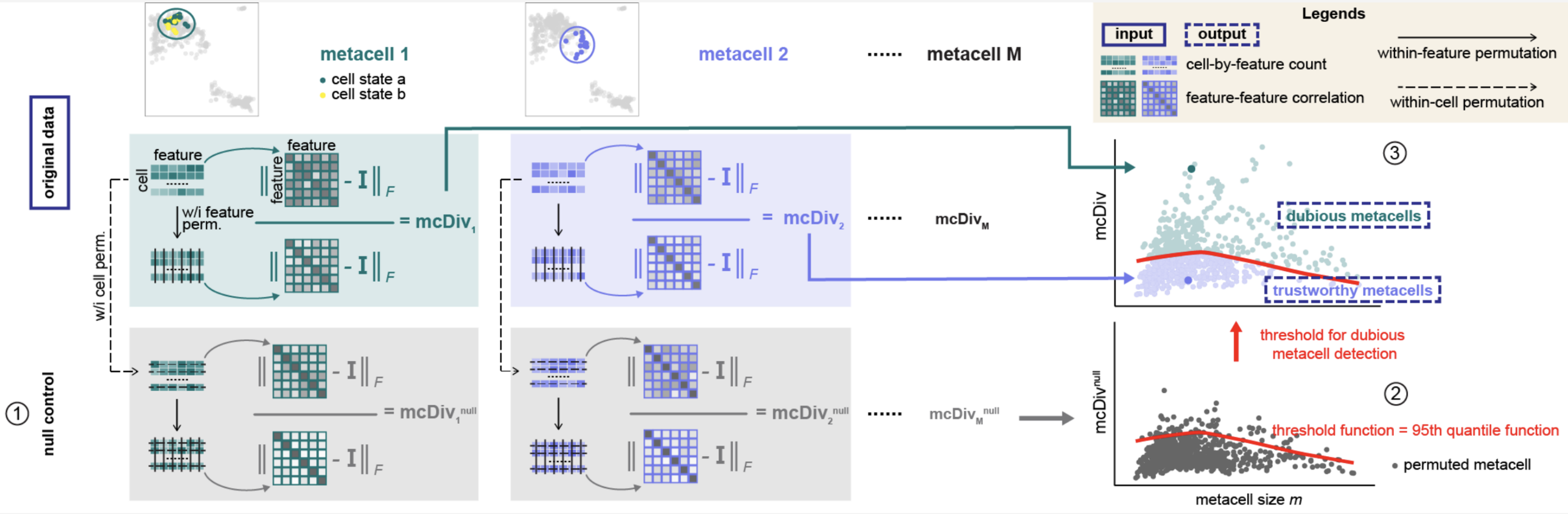
Example 2: aggregating single cells into metacells

mcRigor function 1: detecting dubious metacells



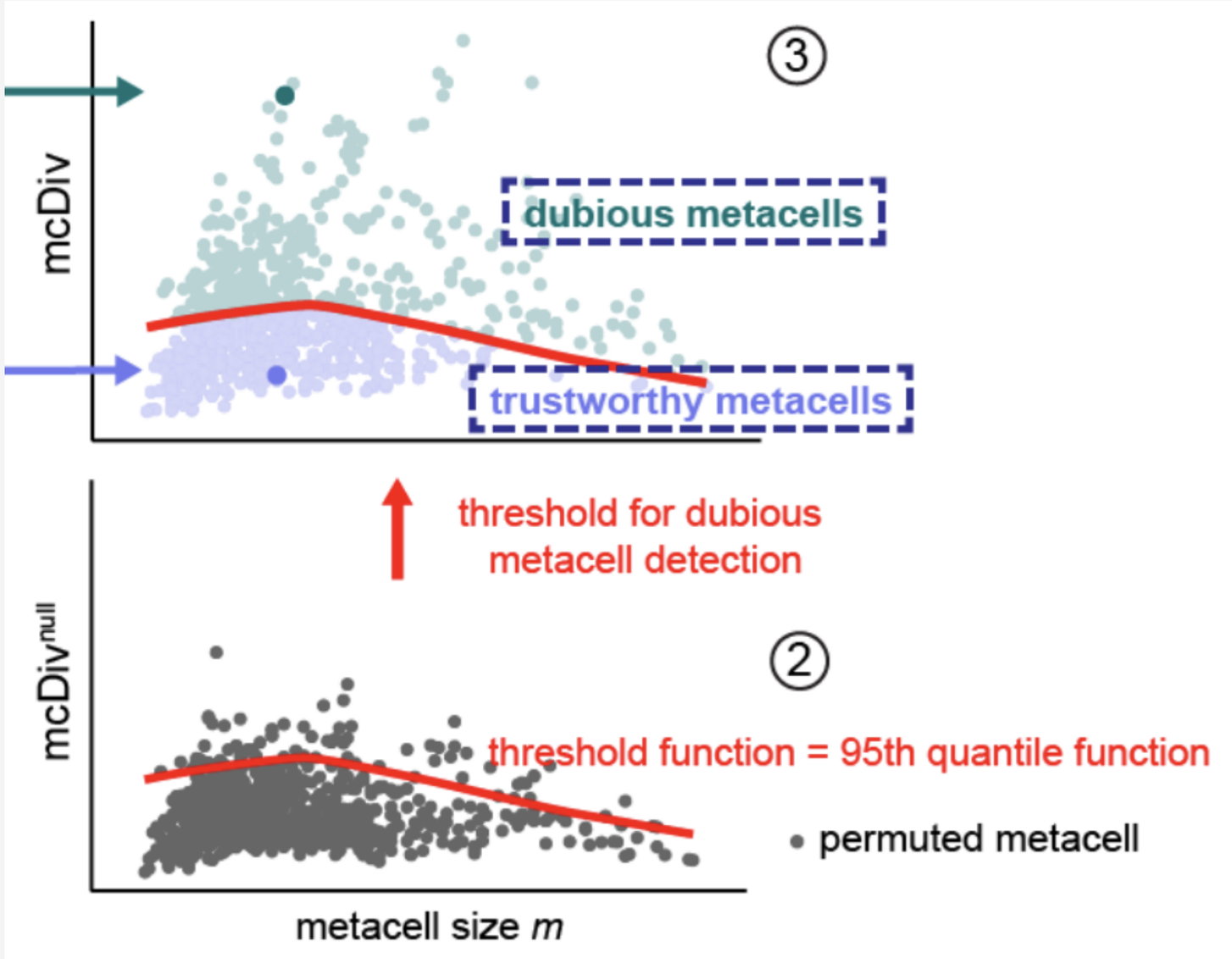
Example 2: aggregating single cells into metacells

mcRigor function 1: detecting dubious metacells



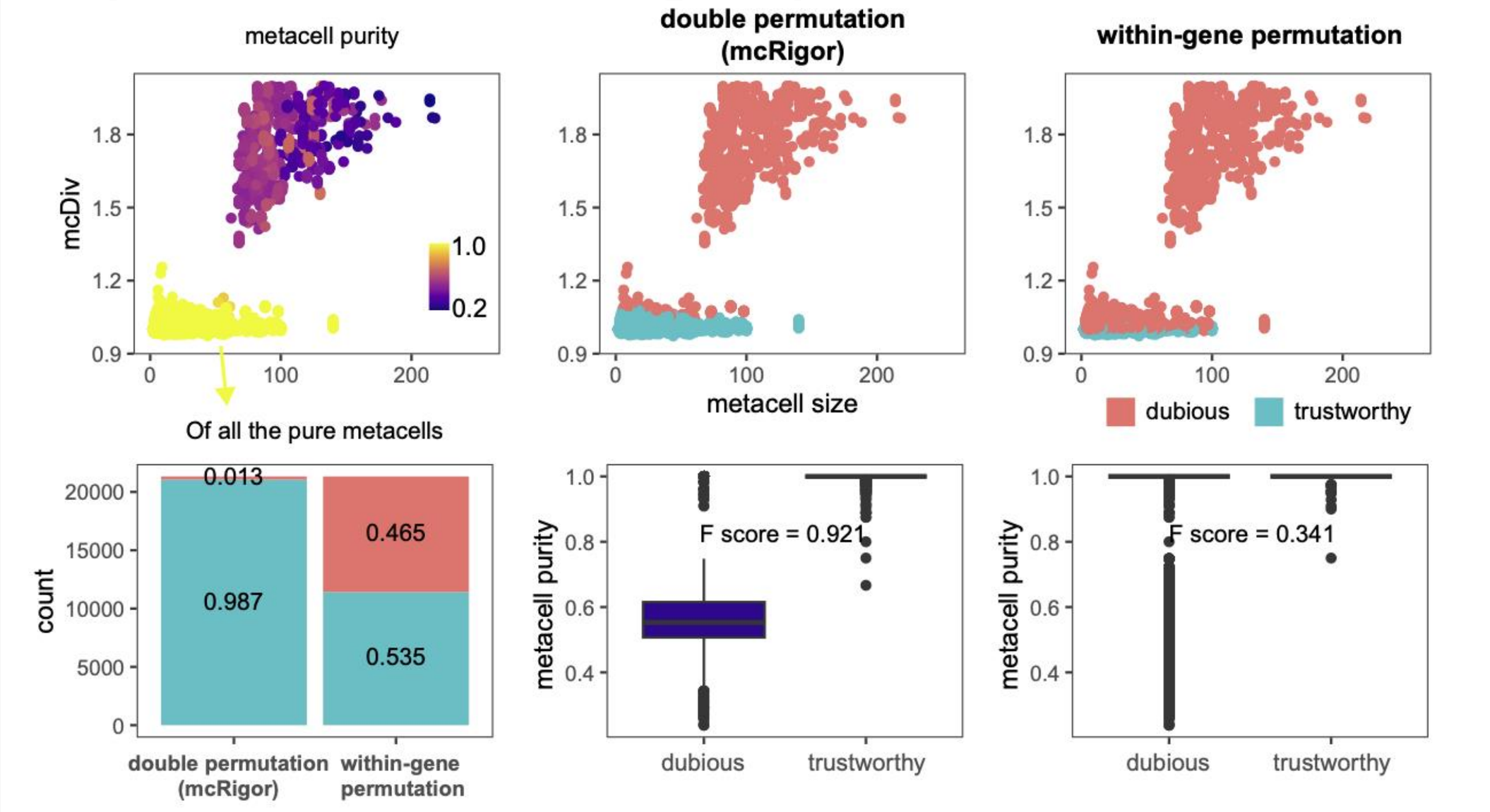
Example 2: aggregating single cells into metacells

mcRigor function 1: detecting dubious metacells



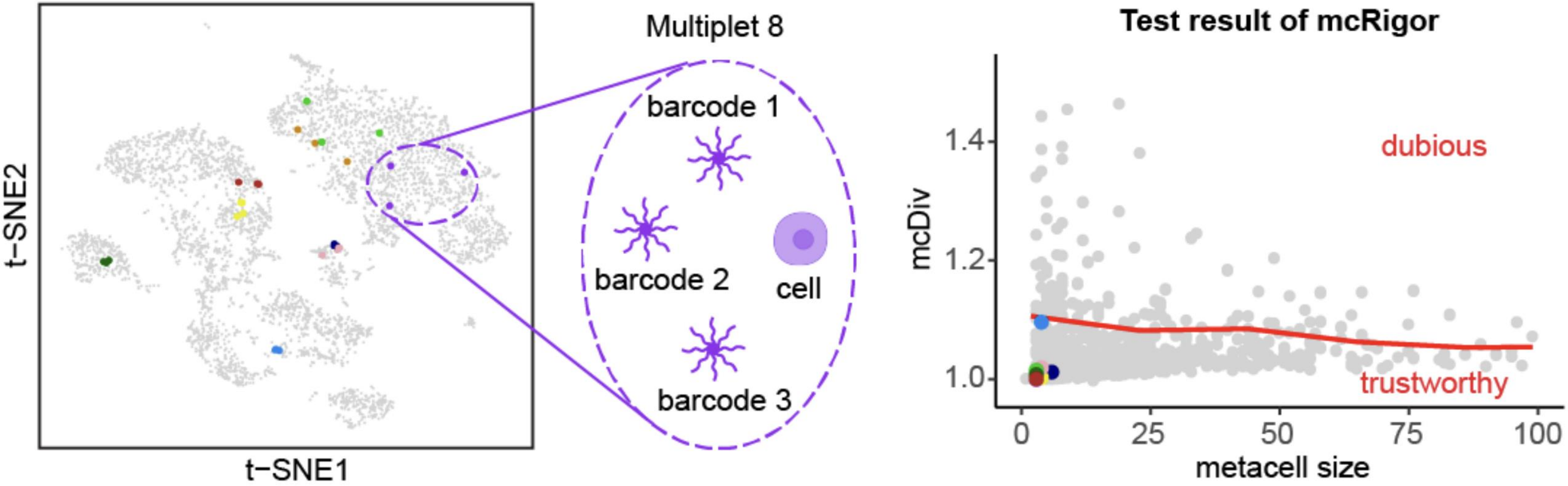
Example 2: aggregating single cells into metacells

Double permutation vs. within-gene permutation



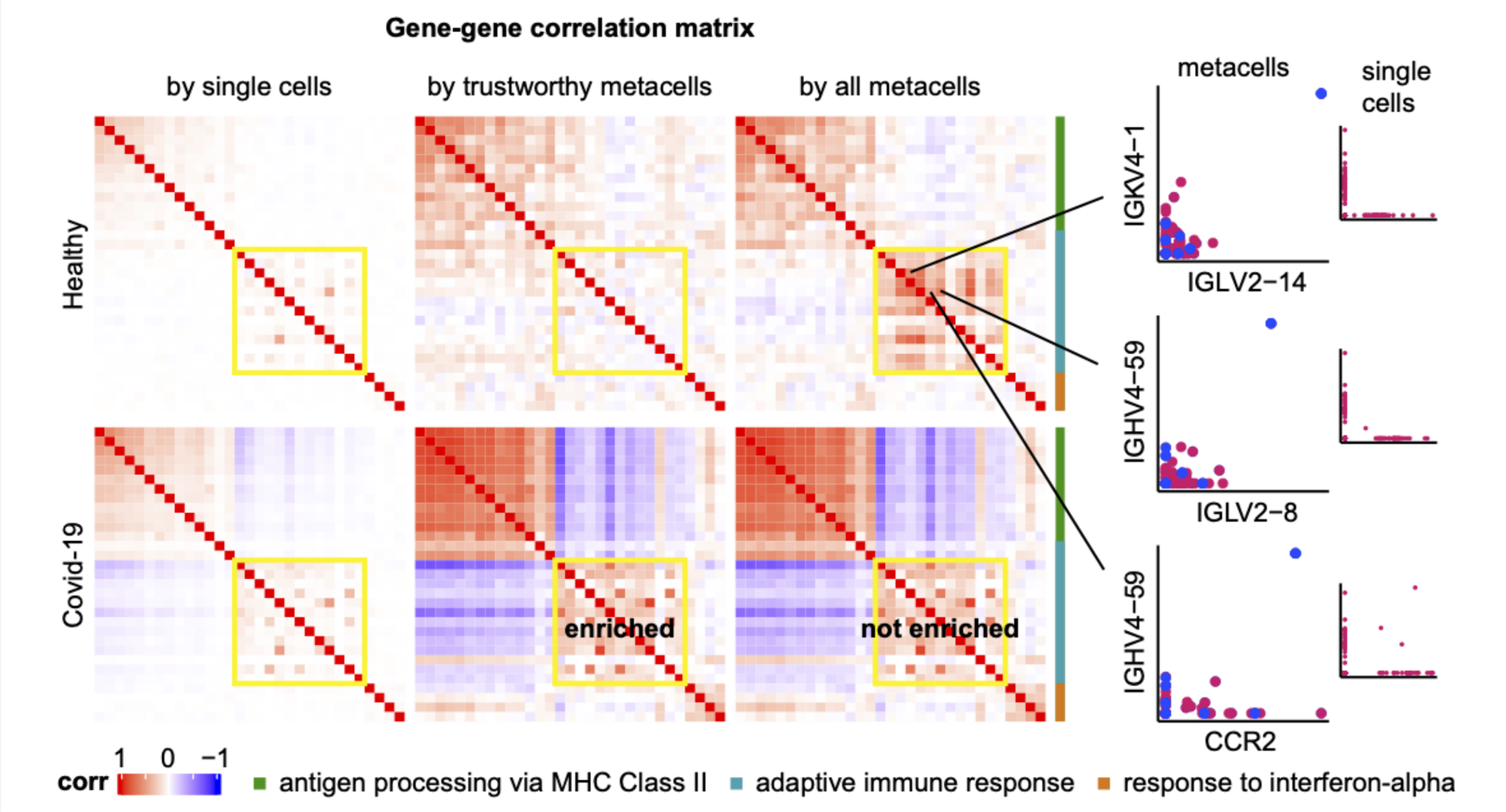
Example 2: aggregating single cells into metacells

Test of mcRigor on barcode multiplets



Example 2: aggregating single cells into metacells

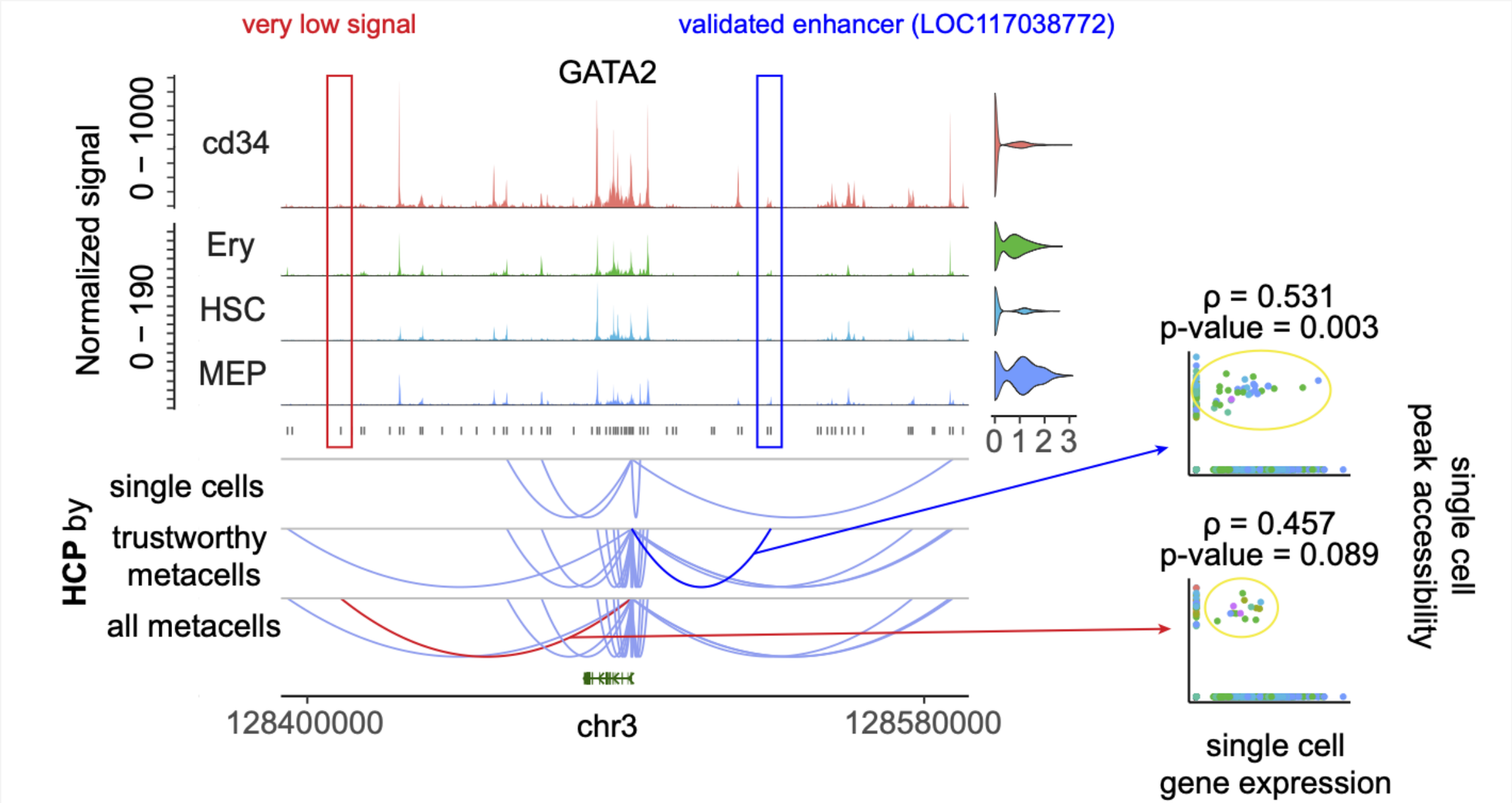
mcRigor improves co-expression by removing dubious metacells



SuperCell: *Bilous, M., et al. "Metacells untangle large and complex single-cell transcriptome networks." BMC Bioinformatics 23 (2022): 336.*
Data: *Wilk, A. J., et al. "A single-cell atlas of the peripheral immune response in patients with severe covid-19." Nature Medicine 26 (7): 1070-1076*

Example 2: aggregating single cells into metacells

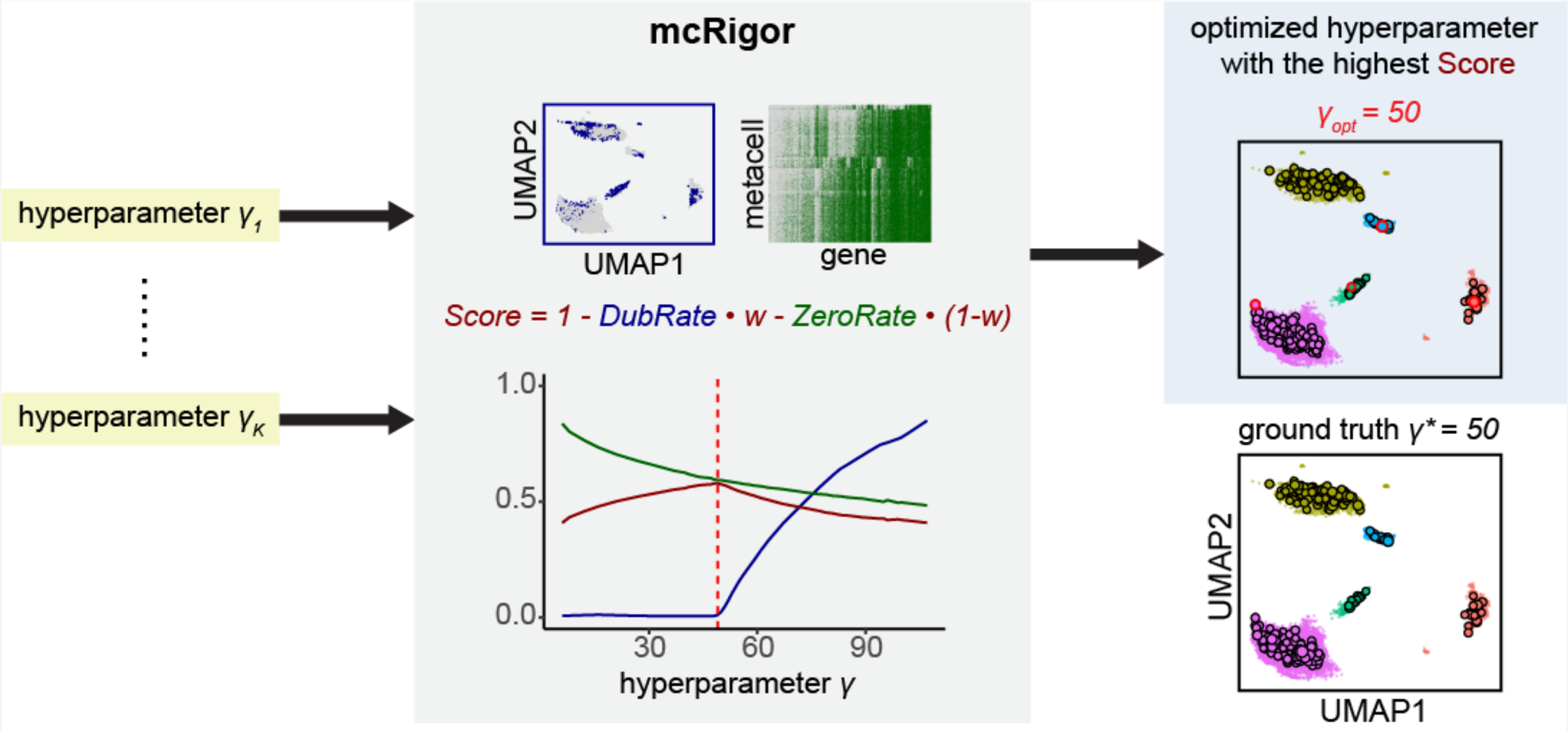
mcRigor improves the reliability of gene regulatory inference



SEACells: *Persad, S., et al. "SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data." Nat Biotechnol 41 (2023).*

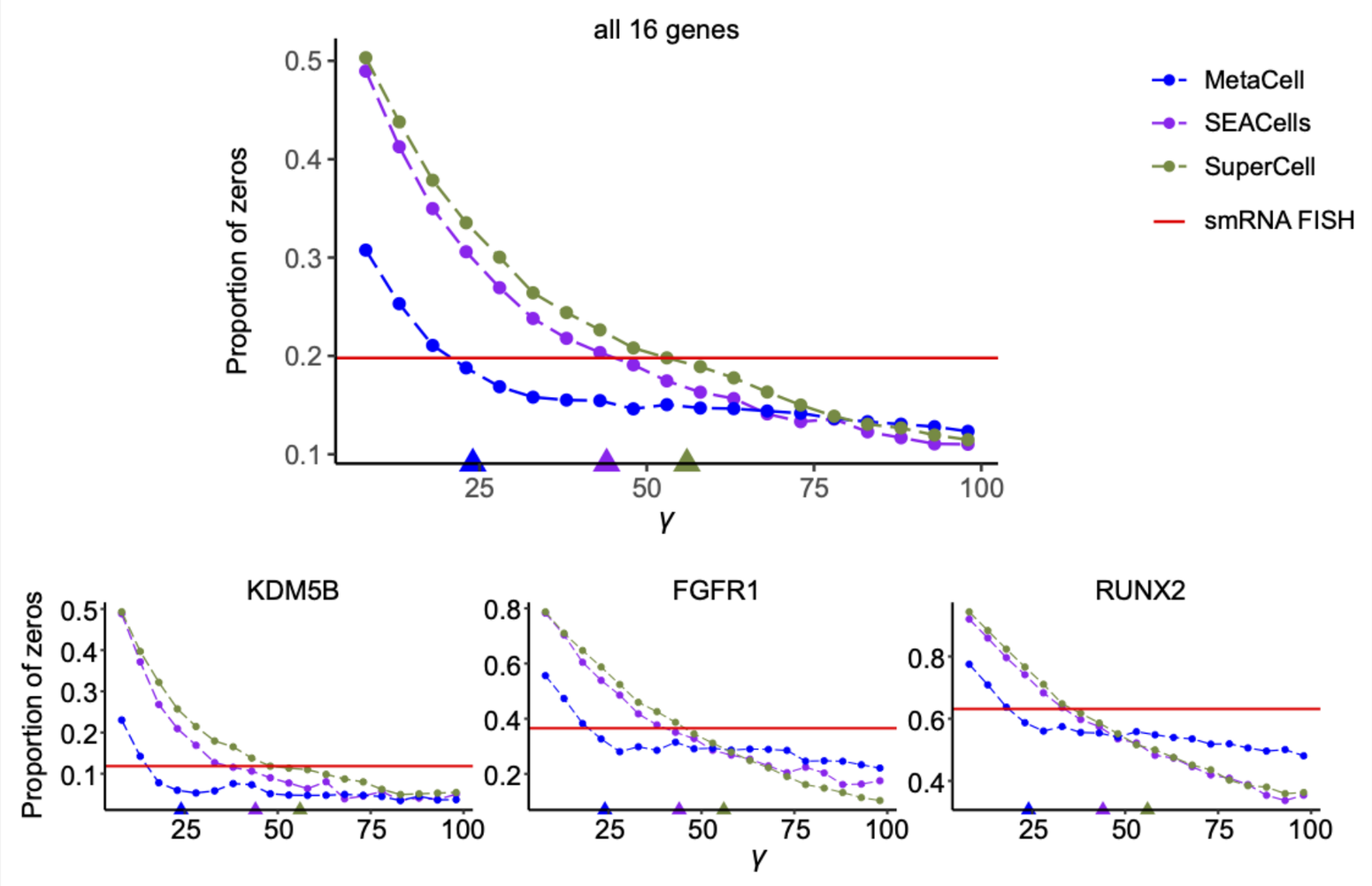
Example 2: aggregating single cells into metacells

mcRigor function 2: optimizing granularity level γ (hyperparameter)



Example 2: aggregating single cells into metacells

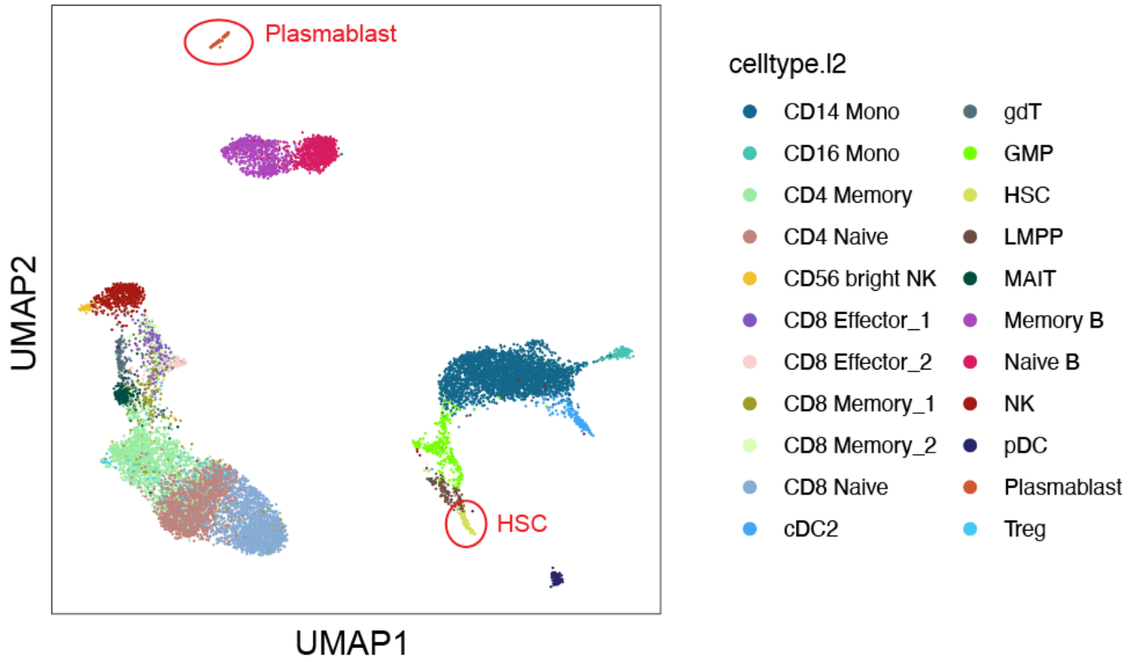
mcRigor helps distinguish biological zeros from technical zeros



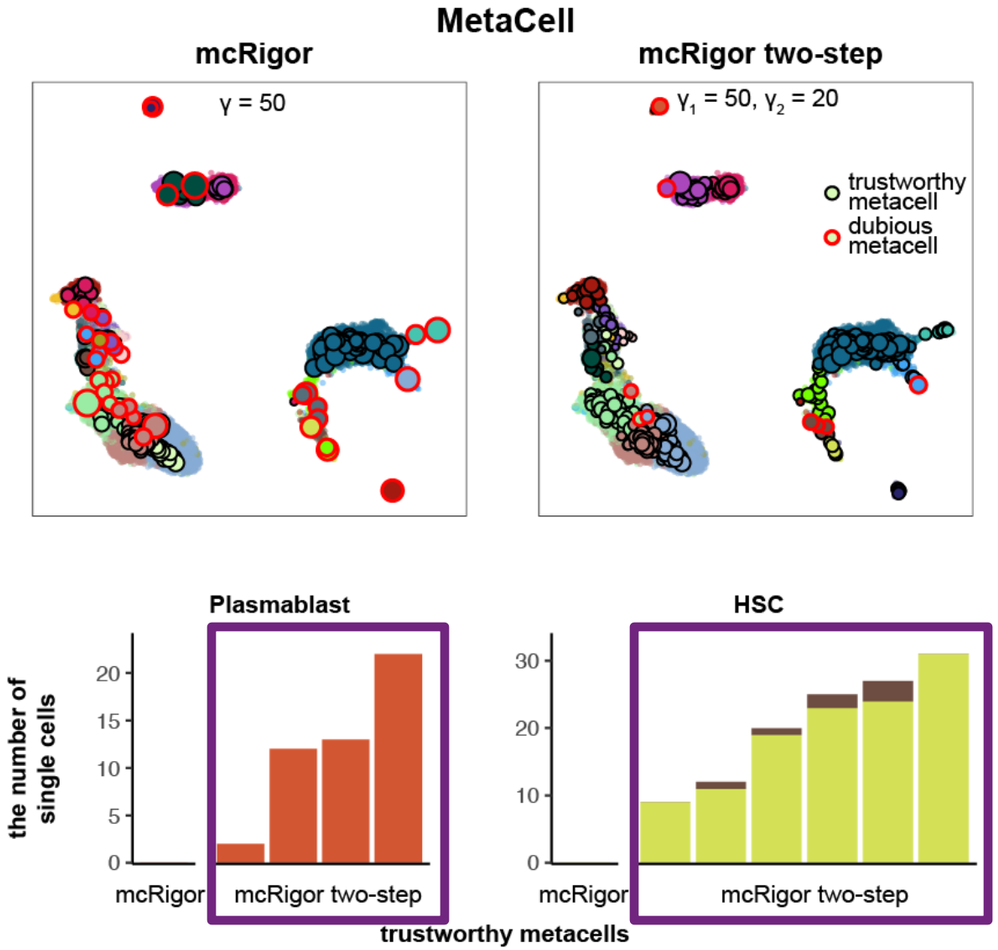
smFISH data: *Torre, E., H. et al. "Rare cell detection by single-cell RNA sequencing as guided by single-molecule RNA FISH." Cell Systems 6 (2), 171–179 (2018).*

Example 2: aggregating single cells into metacells

mcRigor two-step better preserves and reveals rare cell types



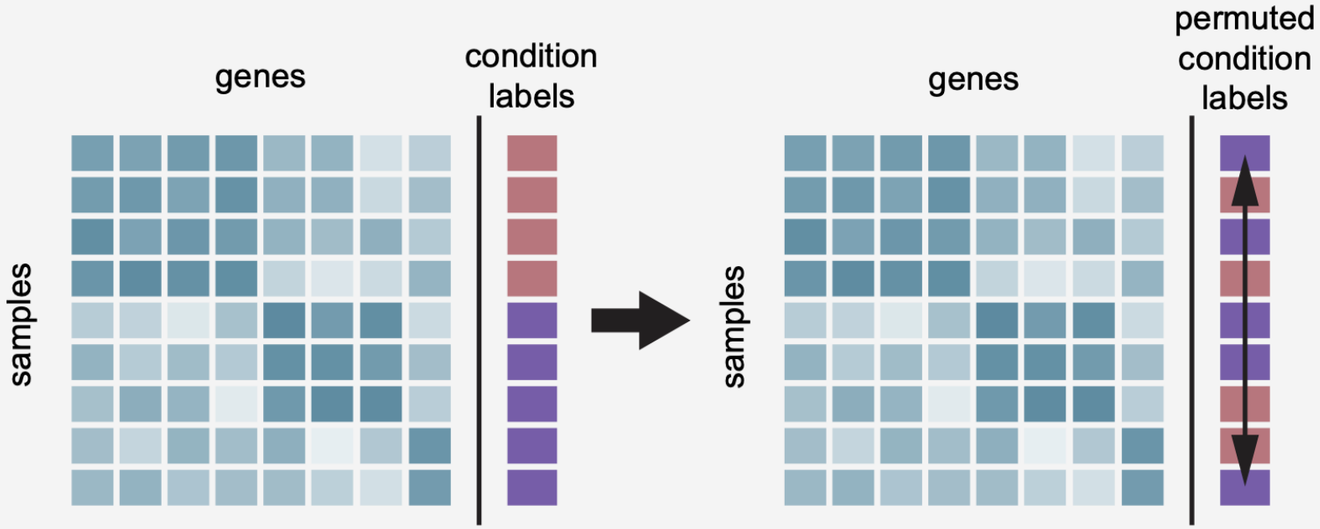
13408 total cells
106 Plasmablast cells (0.8%), 144 HSC cells (1.1%)



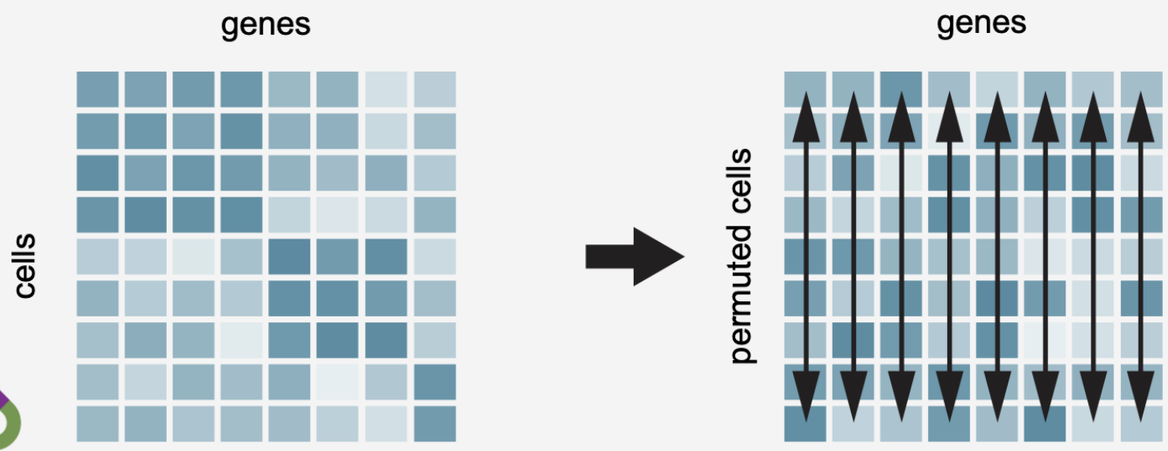
MetaCell: *Baran, Y., et al. "MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions." Genome Biology (2019).*
Data: *Stuart, T., et al. "Comprehensive integration of single-cell data." Cell (2019).*

Summary

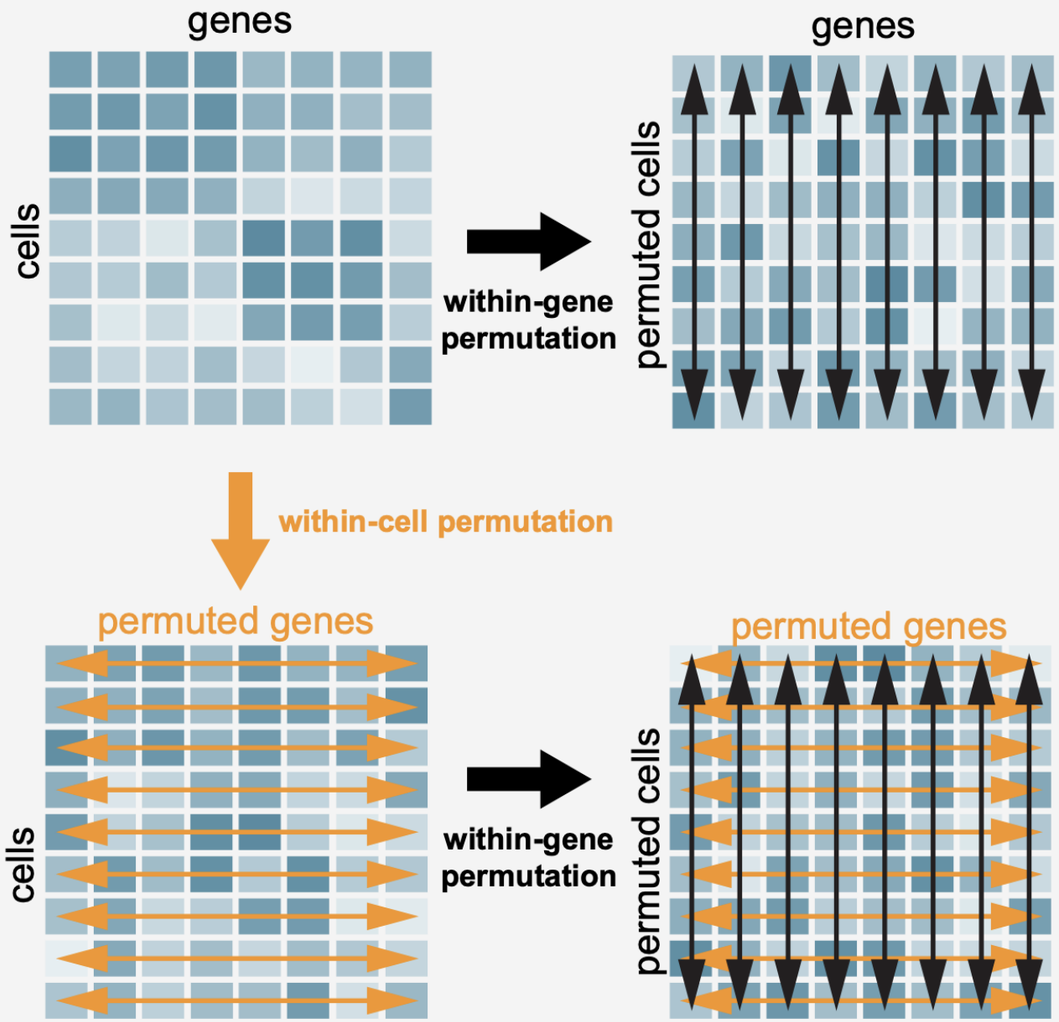
1. condition-label permutation **Bulk DE**



2. within-gene permutation **scDEED**



3. double permutation **mcRigor**



Acknowledgements

Bulk DE analysis

Yumei Li
(Wei Li Lab→ Soochow U)

Xinzhou Ge
(JSB→ OregonState)

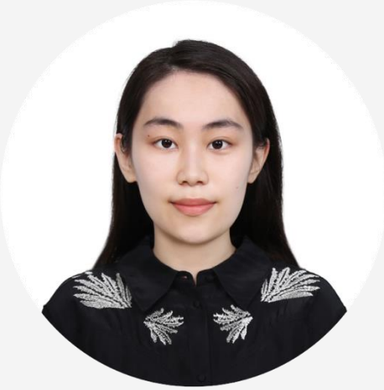
Wei Li
(UC Irvine)

scDEED

Lucy Xia
(HKUST)

Christy Lee

mcRigor



Pan Liu

https://jsb-lab.org/