

No to Relativity: Total-mRNA-Aware Single Cell Analysis

Qiuyu Lian^{1,2}, Jianzhu Ma³, Kui Hua⁴, Ziqi Rong¹, Jinpu Cai¹, Xinzhu Jiang¹, Jieli Zhou¹, Jiaying Xu¹, Zhiyuan Zhuang¹, Li Fan⁵, Henry Zou⁵, Xiaoling Huang², Qi Yan⁶, Jin Gu⁷, Xuegong Zhang⁷, Hongbin Shen², John Marioni^{4,8,9}, Kong Chen⁵✉, Hongyi Xin^{1,2}✉

¹UM-SJTU Joint Institute, Shanghai Jiao Tong University, Shanghai, China. ²Department of Automation, Shanghai Jiao Tong University, Shanghai, China

³Institute for Artificial Intelligence, Peking University, Beijing, China. ⁴Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

⁵Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, USA. ⁶Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York, USA

⁷Bioinformatics Division, BNRIST and Department of Automation, Tsinghua University, Beijing, China. ⁸EMBL European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK

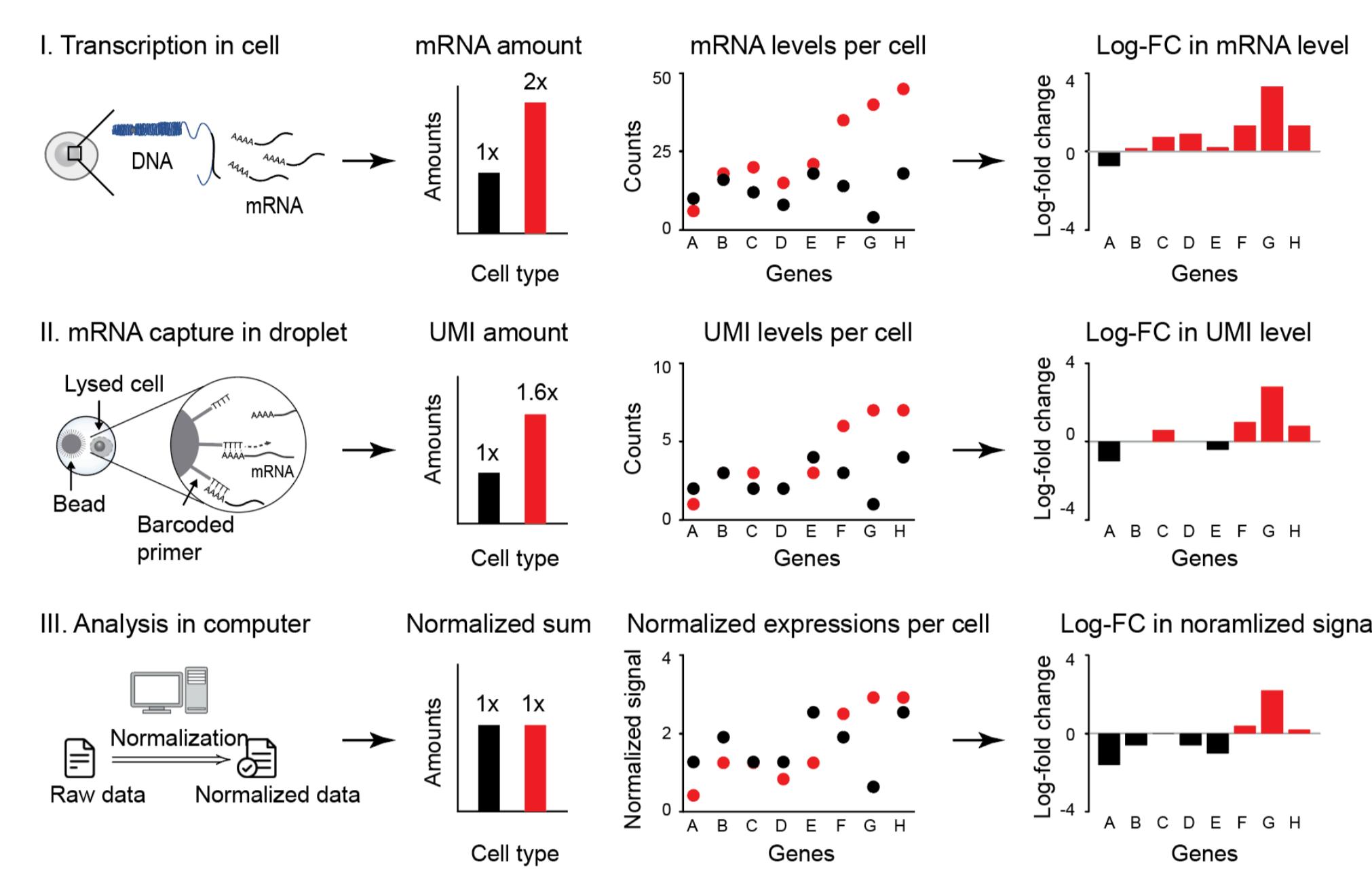
⁹Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK

Introduction

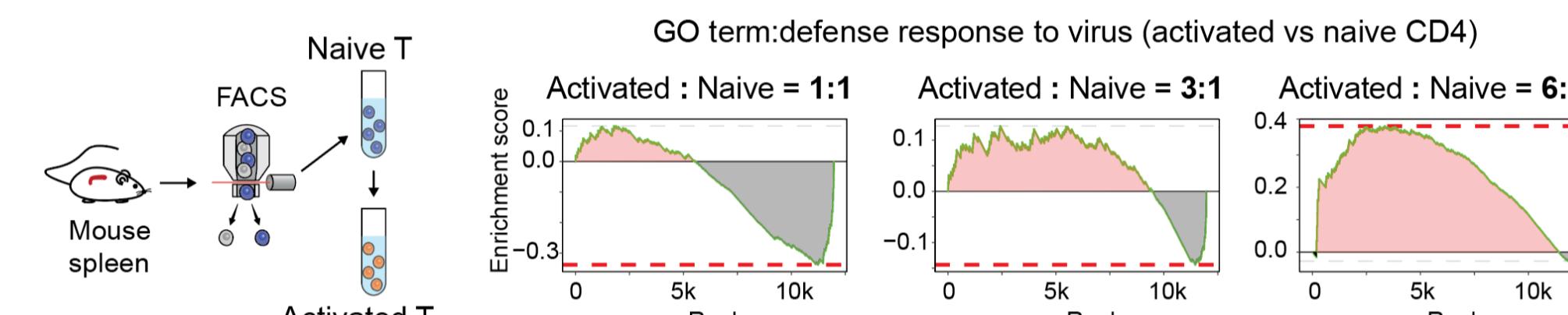
Missing total mRNA control in droplet-based scRNA-seq

- Controlling total mRNA content differences between cell populations is critical in comparative transcriptomic measurements.
- ERCC is not compatible with droplet-based scRNA-seq.

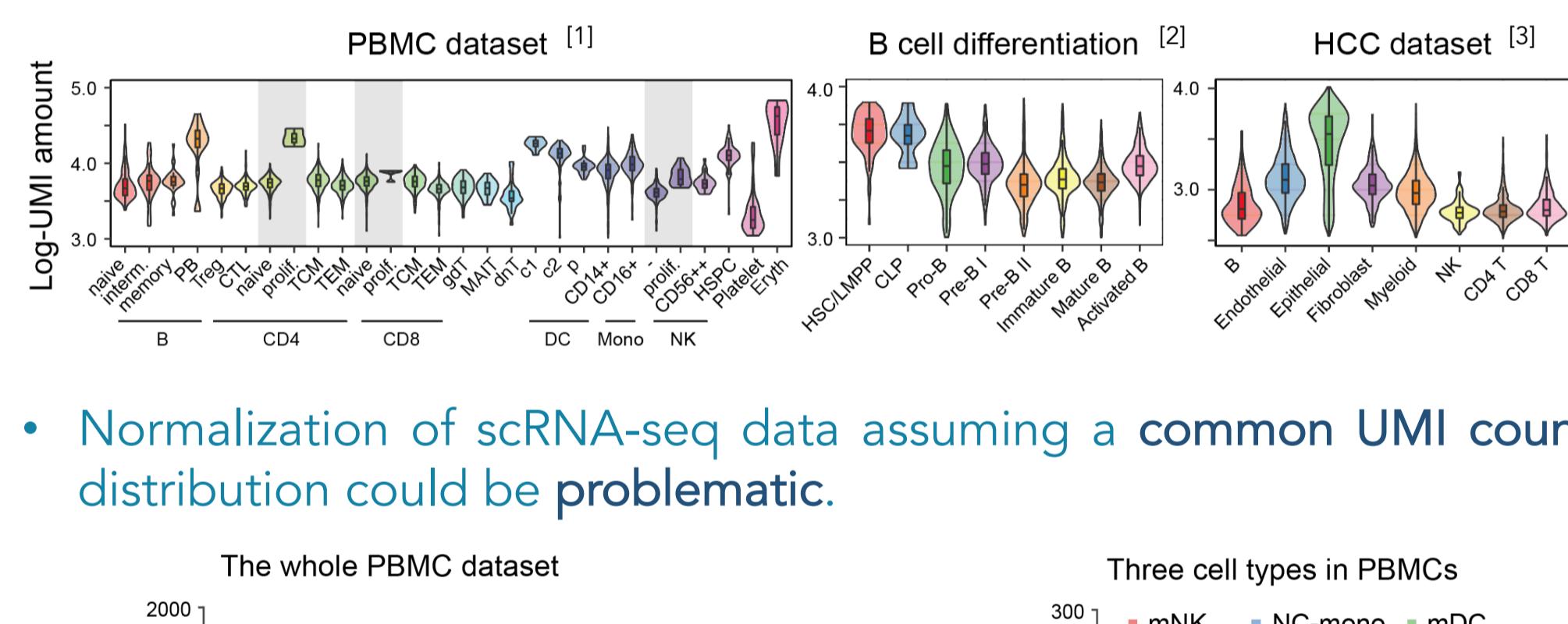
a) Ignoring total mRNA differences between cell types misleads DE analysis.



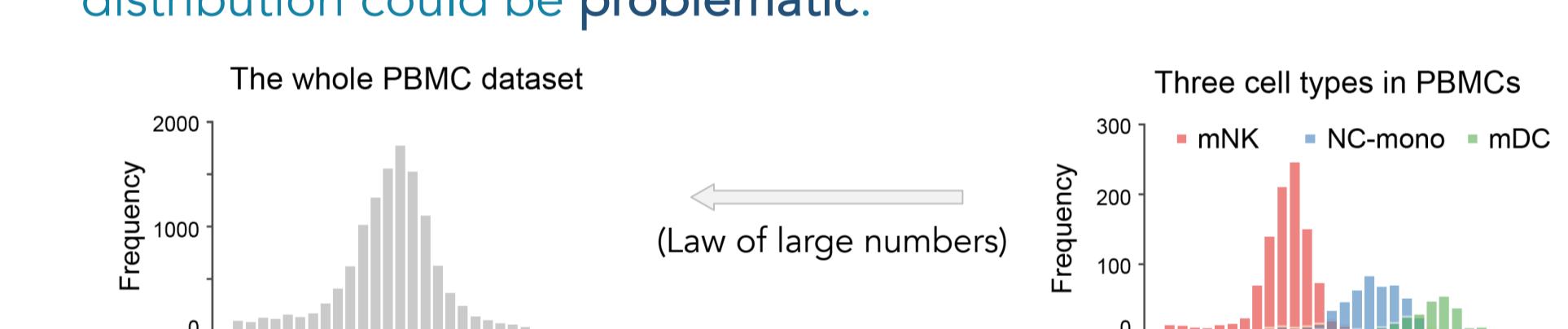
b) GSEA results could vary greatly under different total mRNA ratio assumptions.



Ubiquitous total mRNA differences by cell types

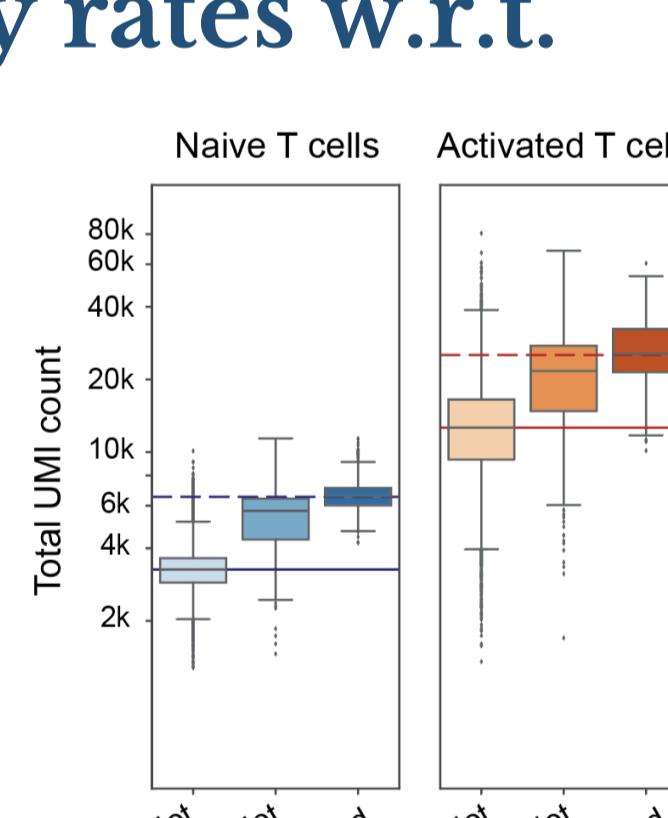


- Normalization of scRNA-seq data assuming a common UMI count distribution could be problematic.



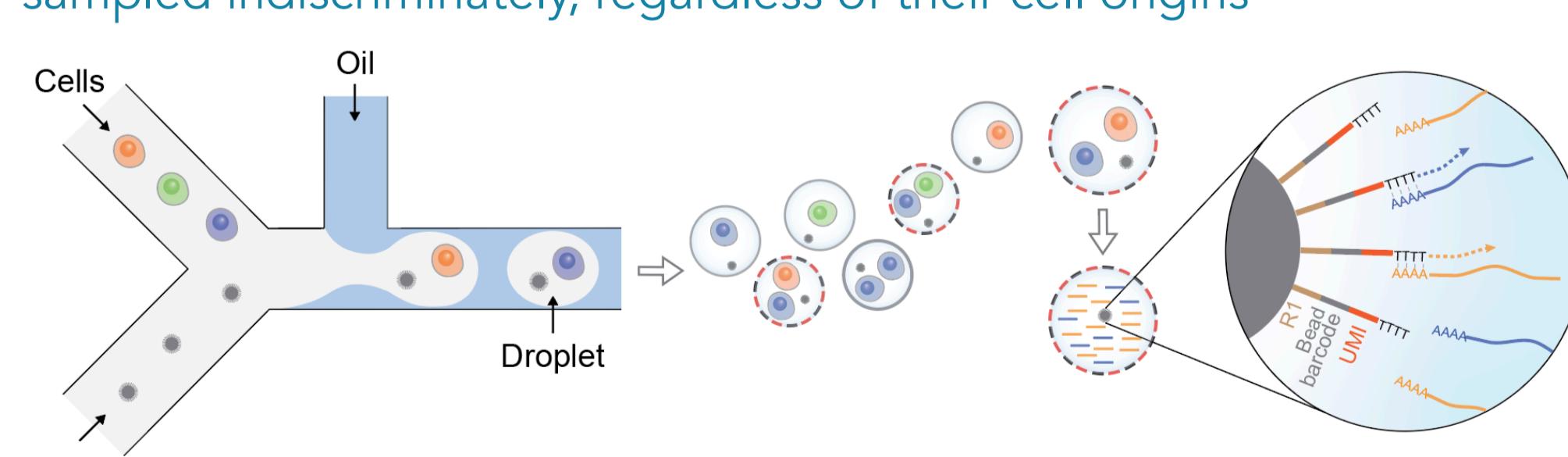
Decaying mRNA recovery rates w.r.t. larger cellular contents

- The UMI amount ratio tends to underestimate total-mRNA ratio
 - Greater UMIs indicate larger mRNA contents in droplets.
 - mRNA recovery rate decreases with total mRNA content increasing.



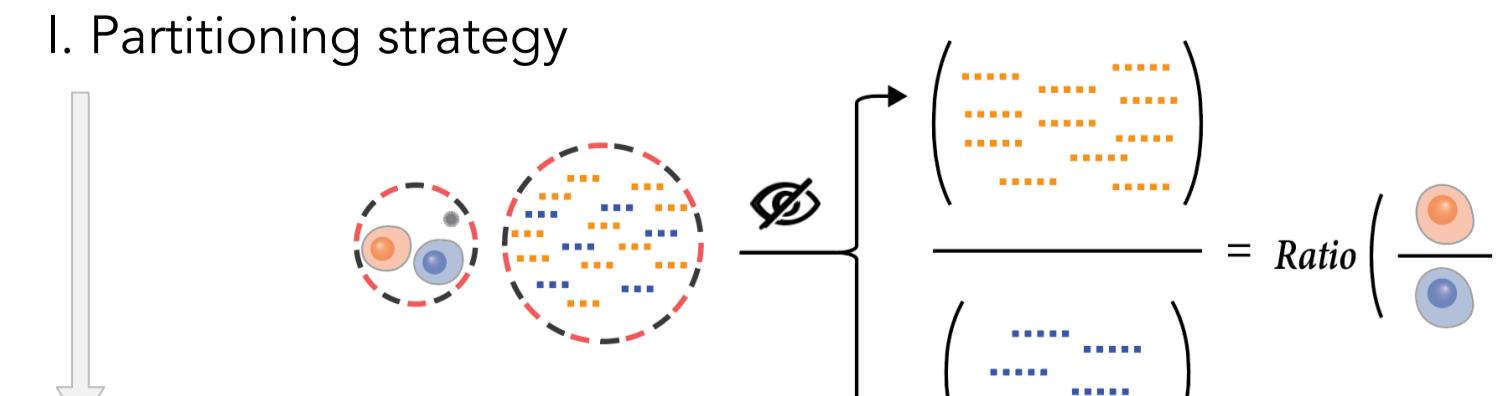
Hypothesis & Idea

Hypothesis: mRNA molecules in a heterotypic doublets are randomly sampled indiscriminately, regardless of their cell origins

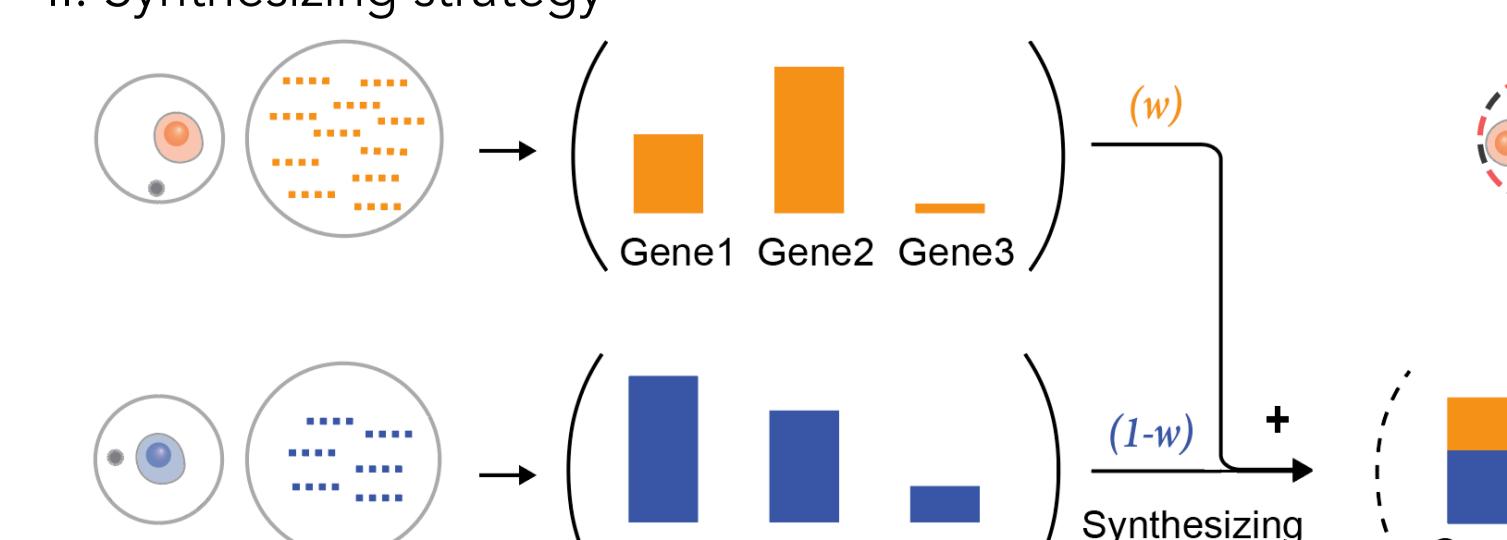


Key idea: TOMAS infers the total mRNA ratio between a pair of cell groups by deconvoluting their respective heterotypic doublets.

I. Partitioning strategy



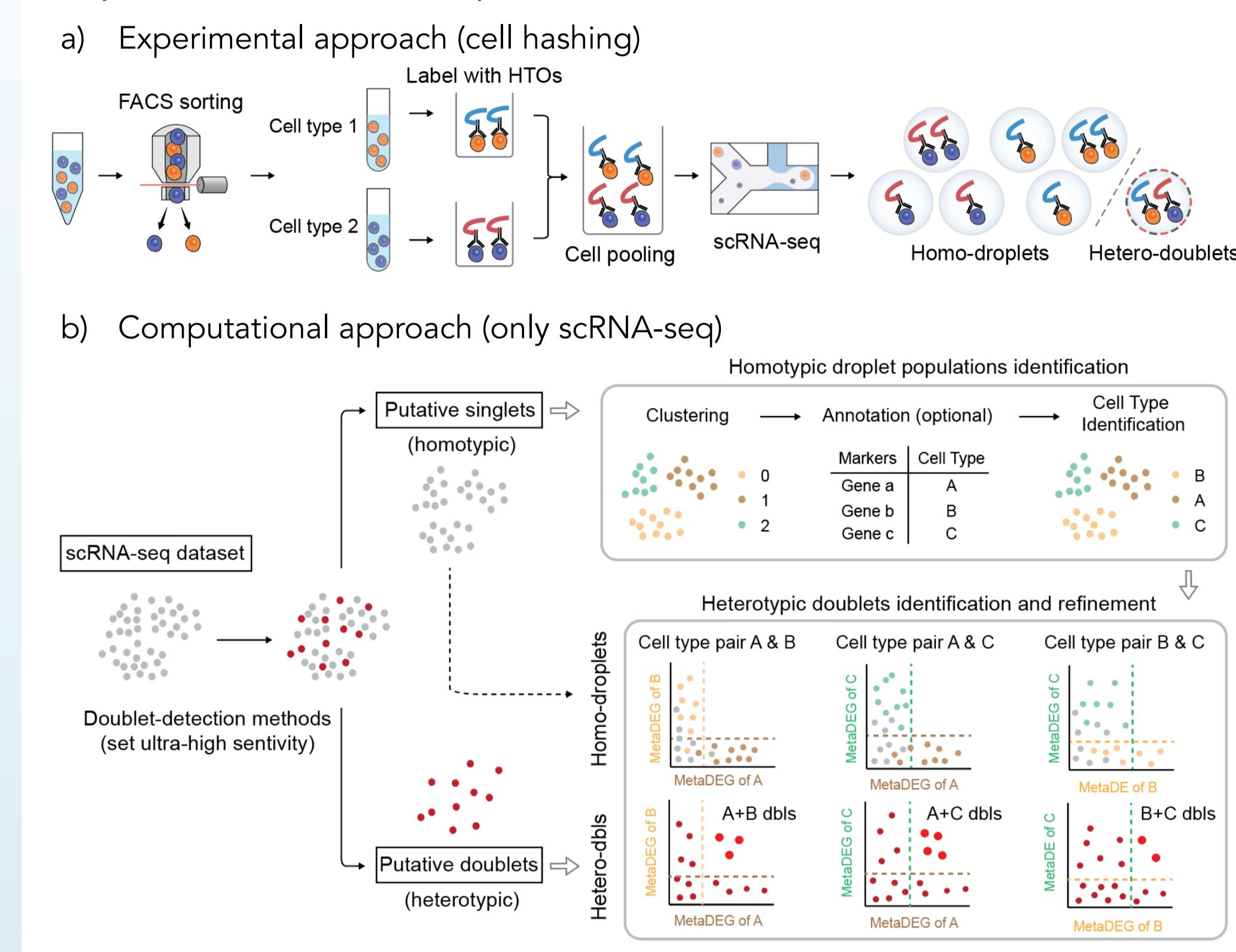
II. Synthesizing strategy



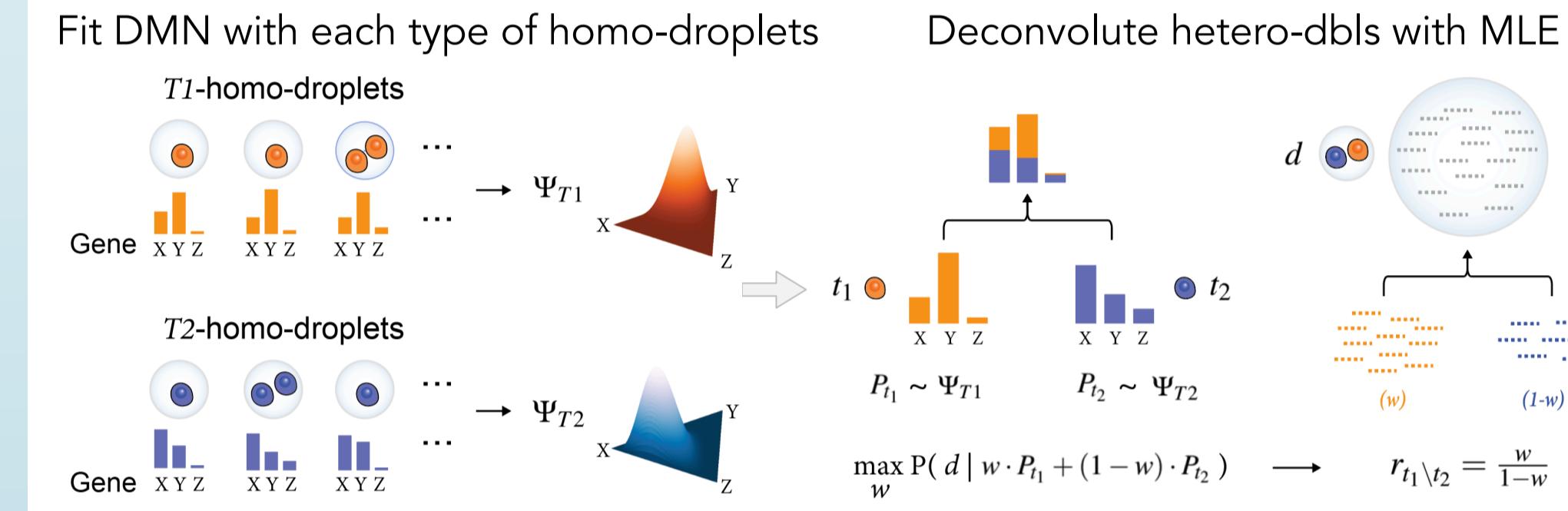
Method

TOTal-mRNA-Aware ScRNA-seq analysis (TOMAS)

Step1: Identify homo-droplets and hetero-doublets.



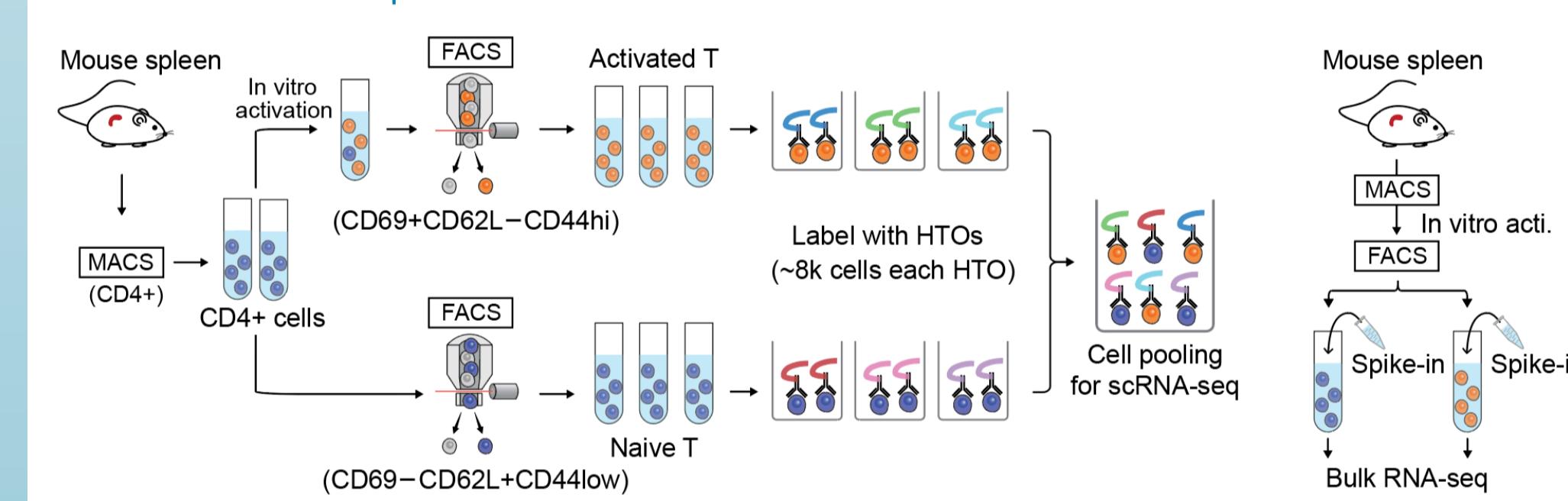
Step2: Fit homo-droplets with Dirichlet-Multinomial (DMN) model and deconvolute hetero-doublets via synthesis.



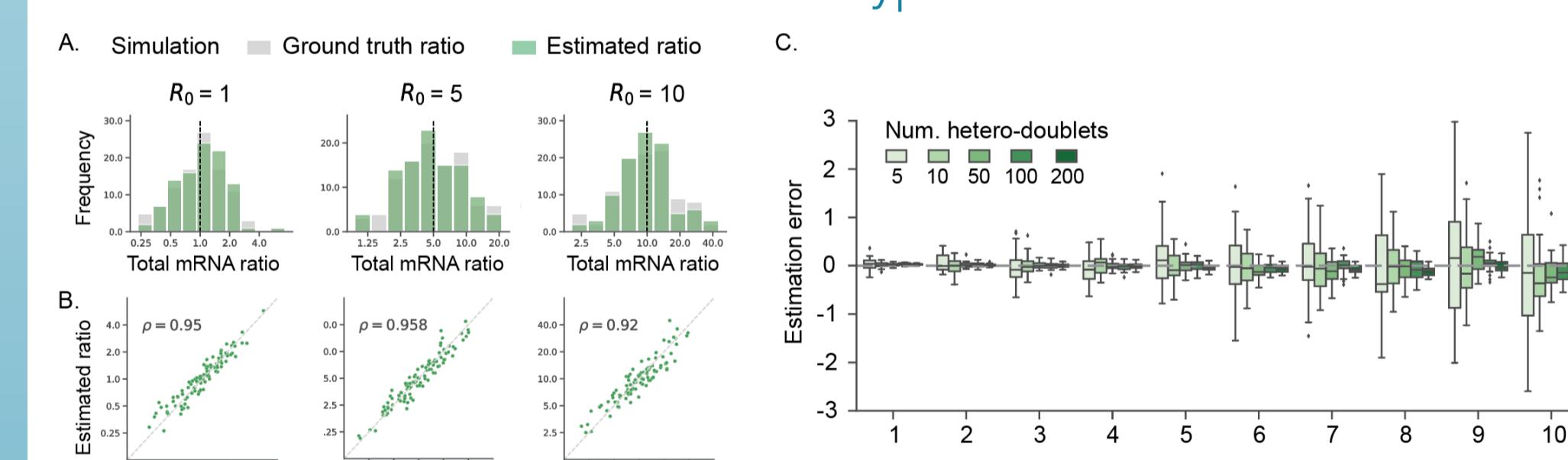
Results

Total-mRNA bias correction in T cell activation DE and GSEA analyses

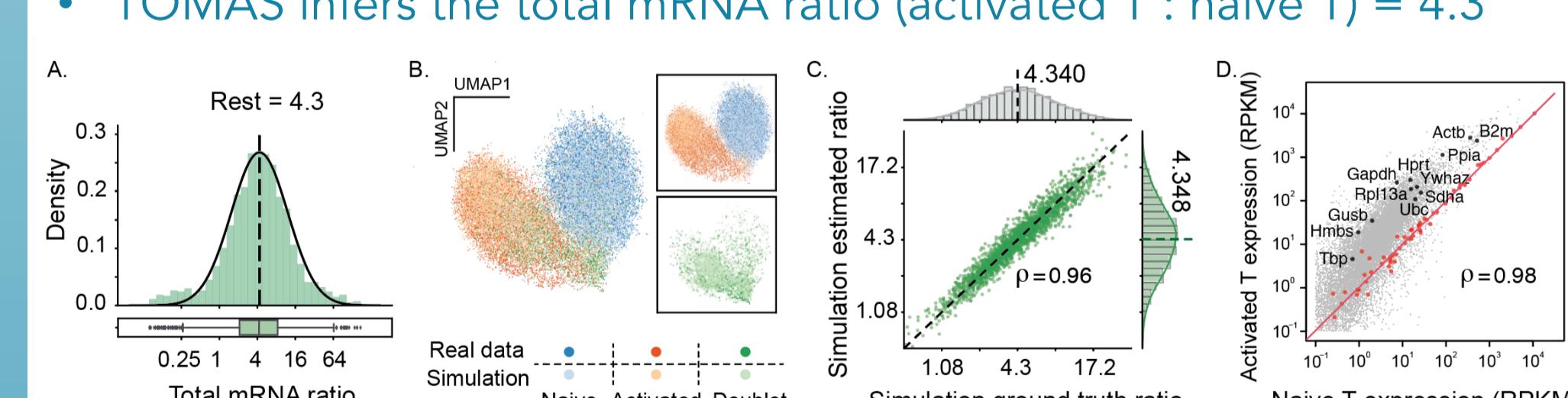
• CD4 T cell activation measured by scRNA-seq with cell hashing and bulk RNA-seq with ERCC



• Simulation datasets demonstrate TOMAS's accuracy in estimating the total mRNA ratio between cell types

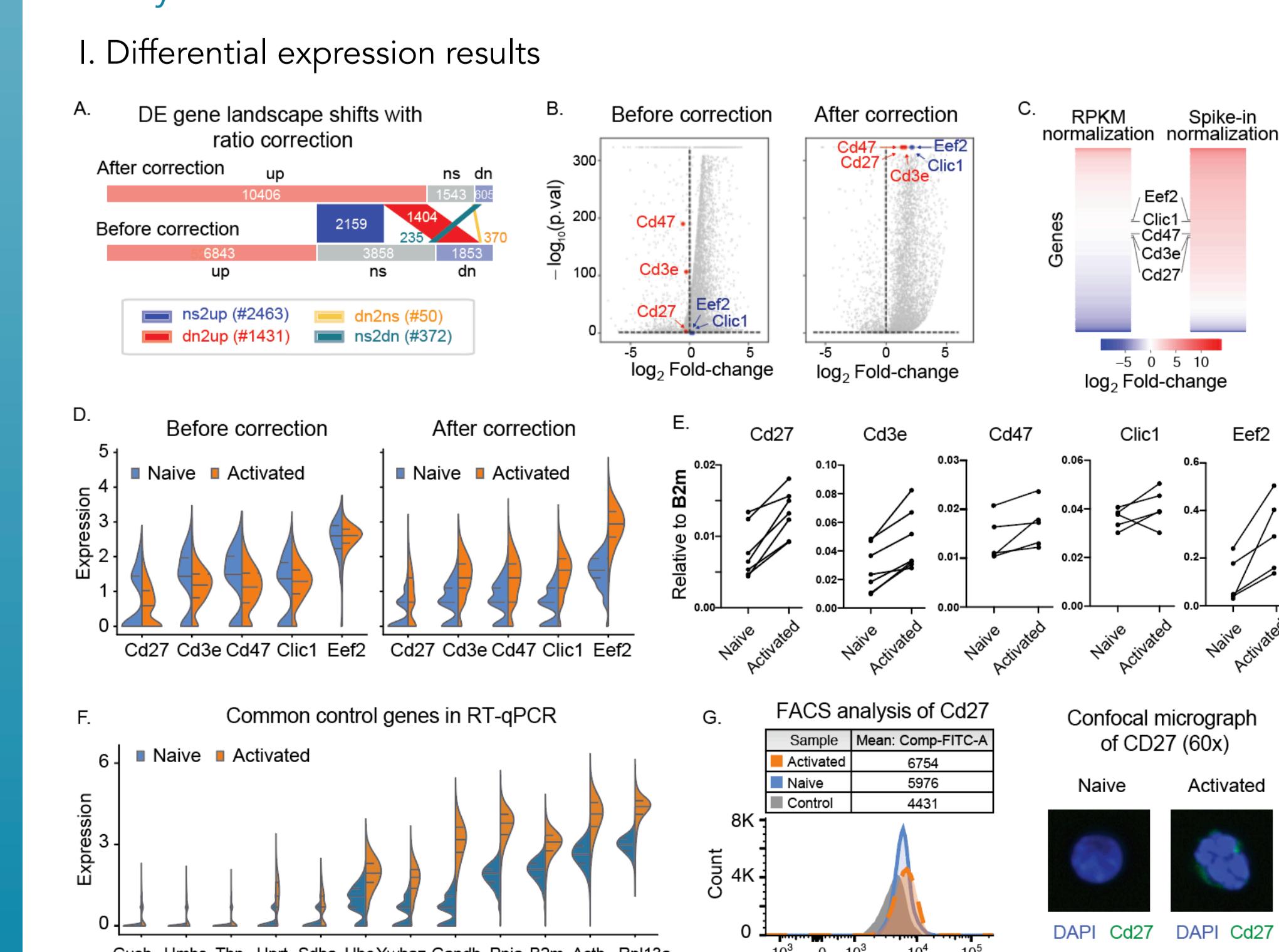


• TOMAS infers the total mRNA ratio (activated T : naïve T) = 4.3

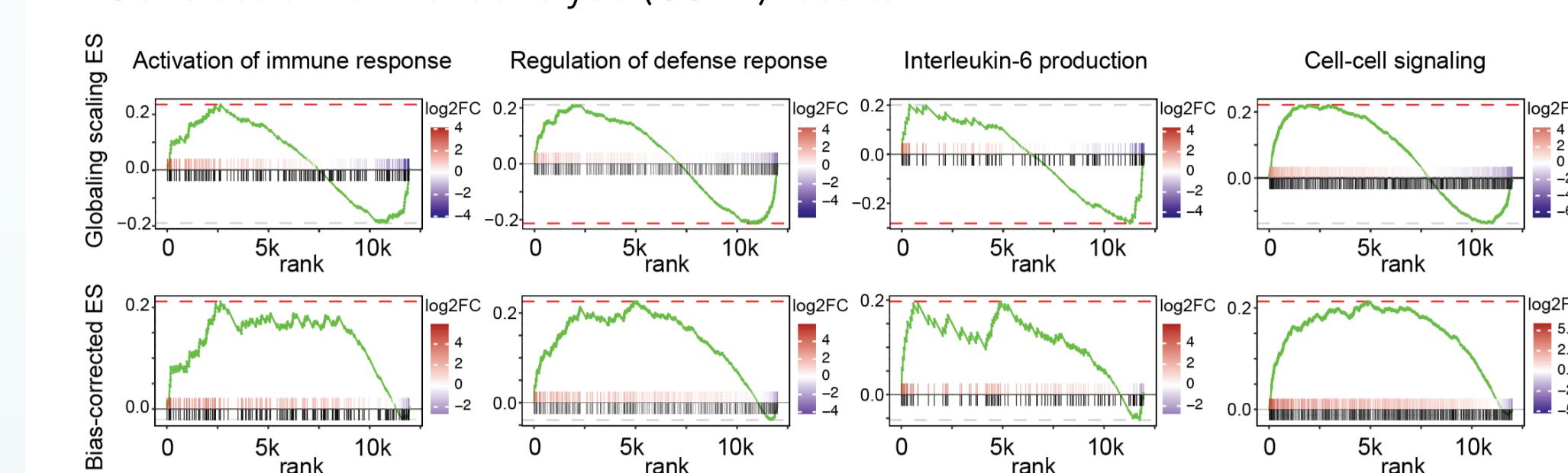


• Total-mRNA bias correction leads to meaningful downstream analyses of naïve and activated CD4+ T cells

I. Differential expression results

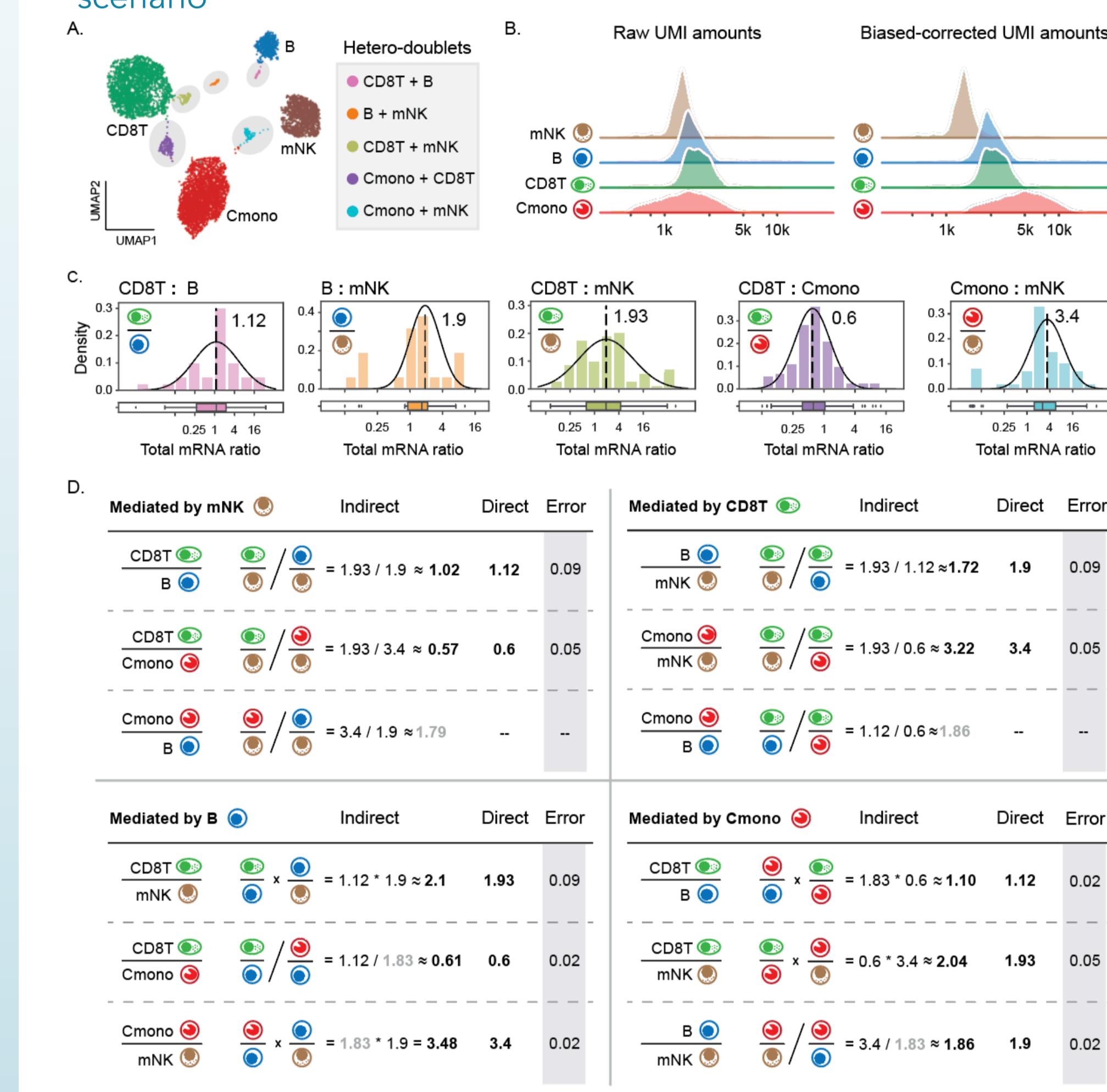


II. Gene set enrichment analysis (GSEA) results

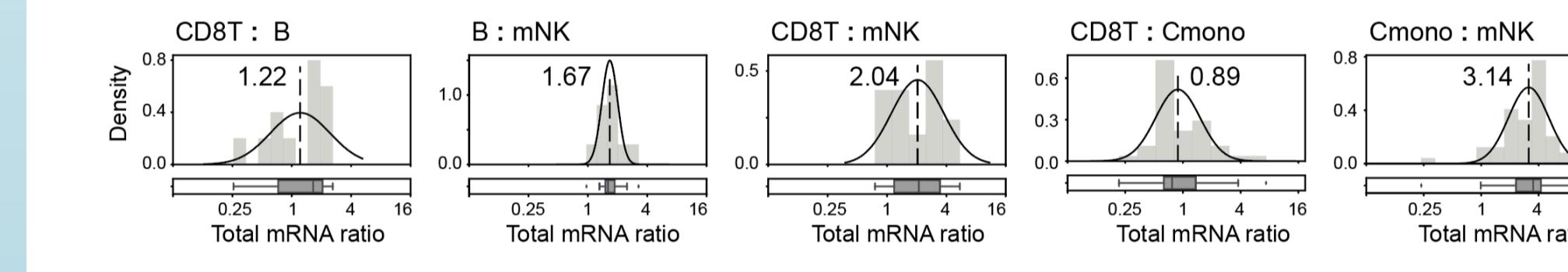


Total-mRNA ratio inference in PBMCs

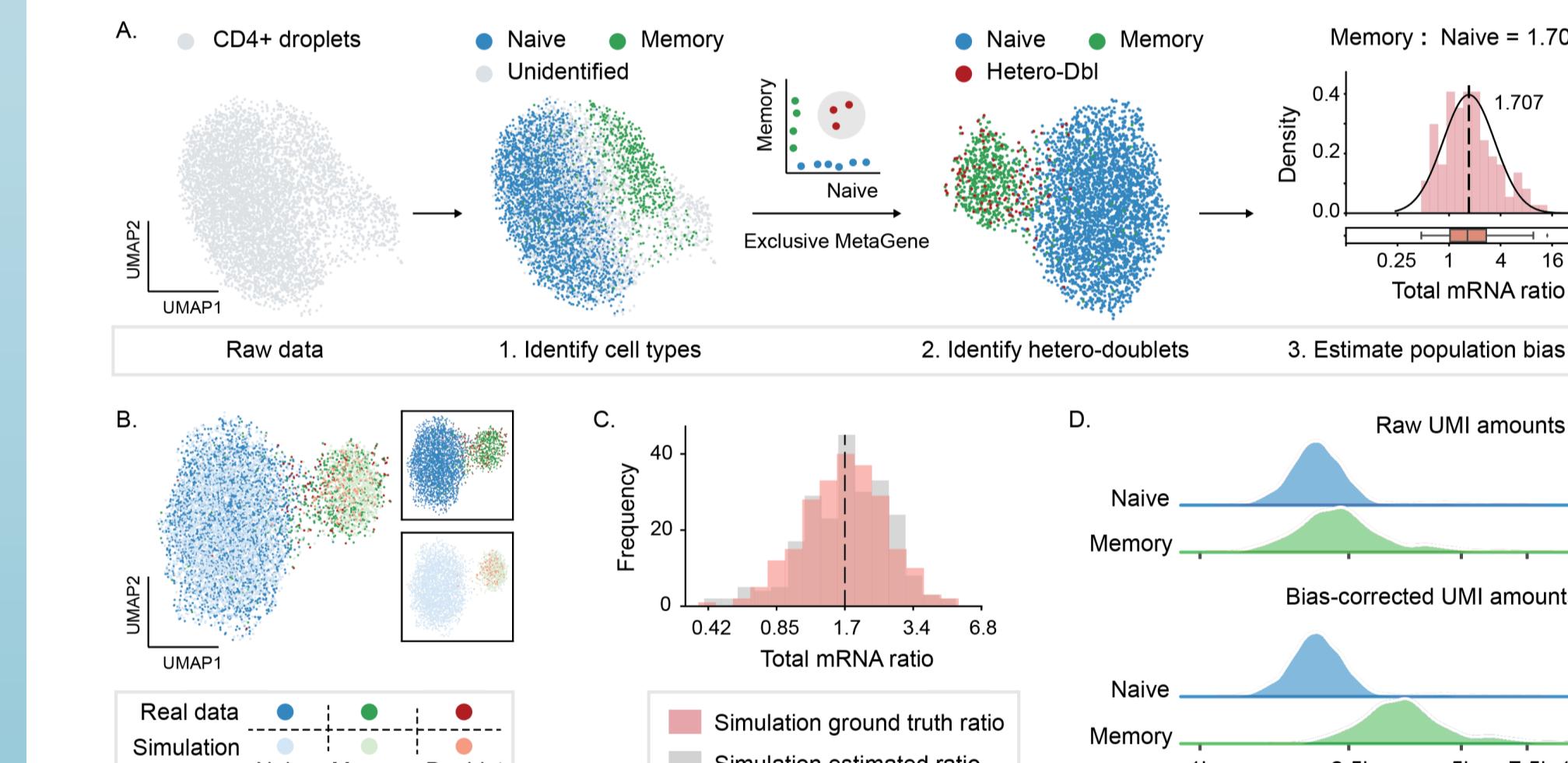
• TOMAS-inferred total mRNA ratios are transitive in multi-cell-type scenario



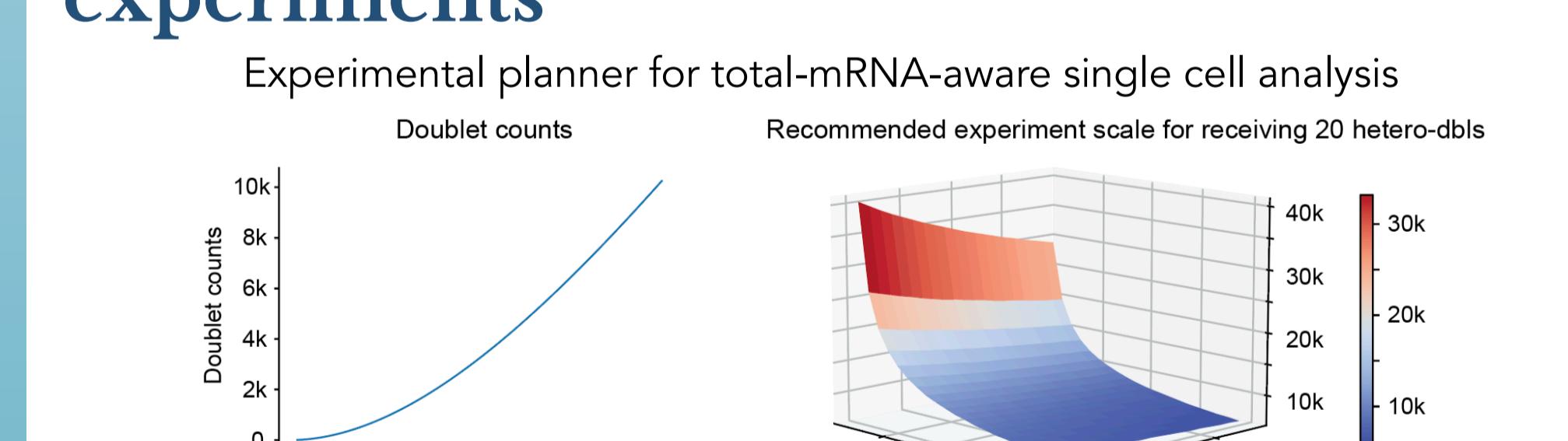
• Total mRNA ratios inferred using only scRNA-seq are consistent



• Dissection of CD4 T cell subtypes with only scRNA-seq



Towards large-scale scRNA-seq experiments



Conclusion

- We present TOMAS, a computational framework that derives total mRNA content ratios between cell populations via deconvoluting their heterotypic doublets.
- TOMAS empowers accurate scRNA-seq analysis for datasets with large total mRNA content differences between cell types.
- TOMAS paves the way for future large-scale, low per-cell cost, doublet-rich droplet-based scRNA-seq experiments, which improves the sensitivity and the statistical power in detecting and studying rare cell types.

Availability & Ref.

Availability

- Doc: <https://tomas.readthedocs.io/en/latest/>
- Software: <https://github.com/QiuyuLian/TOMAS>

Reference

- Xin, H. et al. Gmm-demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. *Genome Biol.* 21, 1–35 (2020).
- Zhang, Y. et al. Elucidating minimal residual disease of paediatric b-cell acute lymphoblastic leukaemia by single-cell analysis. *Nat. Cell Biol.* 1–11 (2022).
- Lasic, B. et al. Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat. Communications.* 11, 1–15 (2020).

