

# Metacells facilitate the analysis of single-cell multiomics data

Léonard Hérault<sup>1,2</sup>, 

leonard.herault@unil.ch

Aurélien Gabriel<sup>1,2</sup>, Mariia Bilous<sup>1,2</sup>, David Gfeller<sup>1,2</sup>

<sup>1</sup> Department of Oncology, Ludwig Institute for Cancer research, University of Lausanne

<sup>2</sup> Swiss Institute of Bioinformatics

## Introduction

- Single-cell multiomics: measurement of different modalities (e.g., ATAC, RNA, proteins) in the same cell (Fig.3A)
- Precise analysis of cell-type specific transcriptional regulation (Fig.3B).

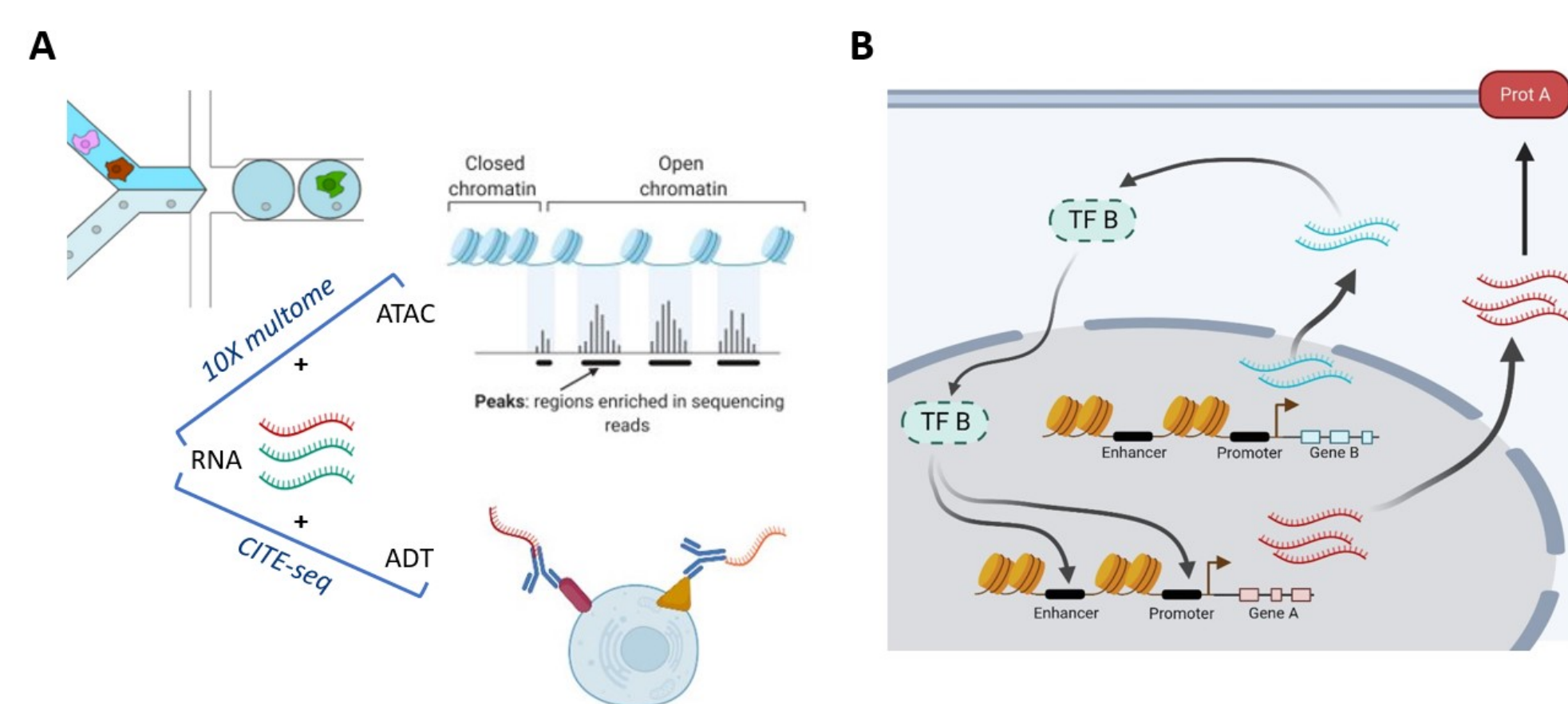


Figure 1: Single-cell multiomics (A) can be used to study cell-type specific transcriptional regulation (B)

- Limitations:** Large size, high sparsity of the data.
- Solution:** Merging highly similar cells in metacells, proposed for scRNA-seq (Baran et al. 2019).
- Aim:** extension of SuperCell (Bilous et al. 2022), to single-cell multiomics. (Fig.2B).

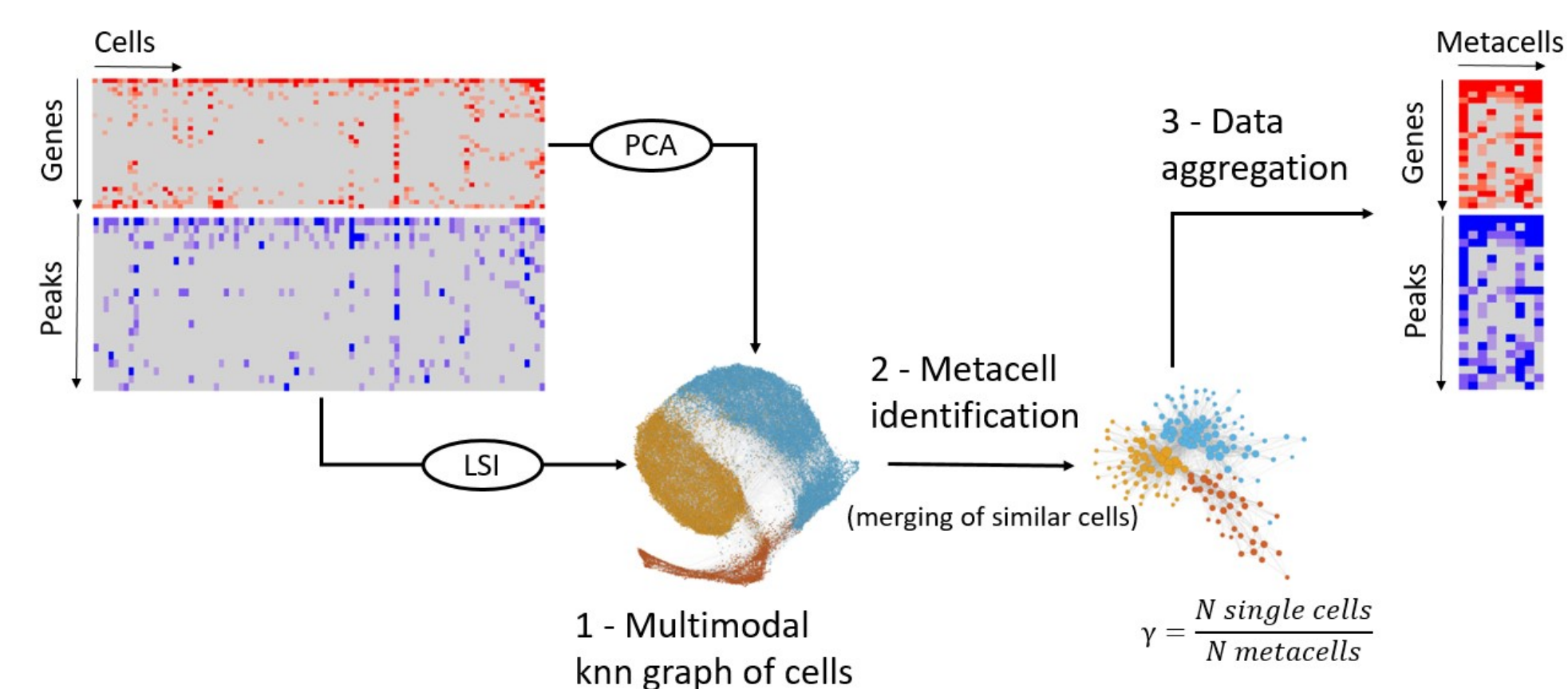


Figure 2: SuperCell workflow to identify multiomics metacells at a graining level  $\gamma$ . The multimodal knn graph is computed using the WNN method from Seurat (Hao et al. 2021).

## Benchmark

Multimodal version of SuperCell versus unimodal tools (Fig.3):

- Better accuracy (high purity, compact metacells in both modalities)
- Faster

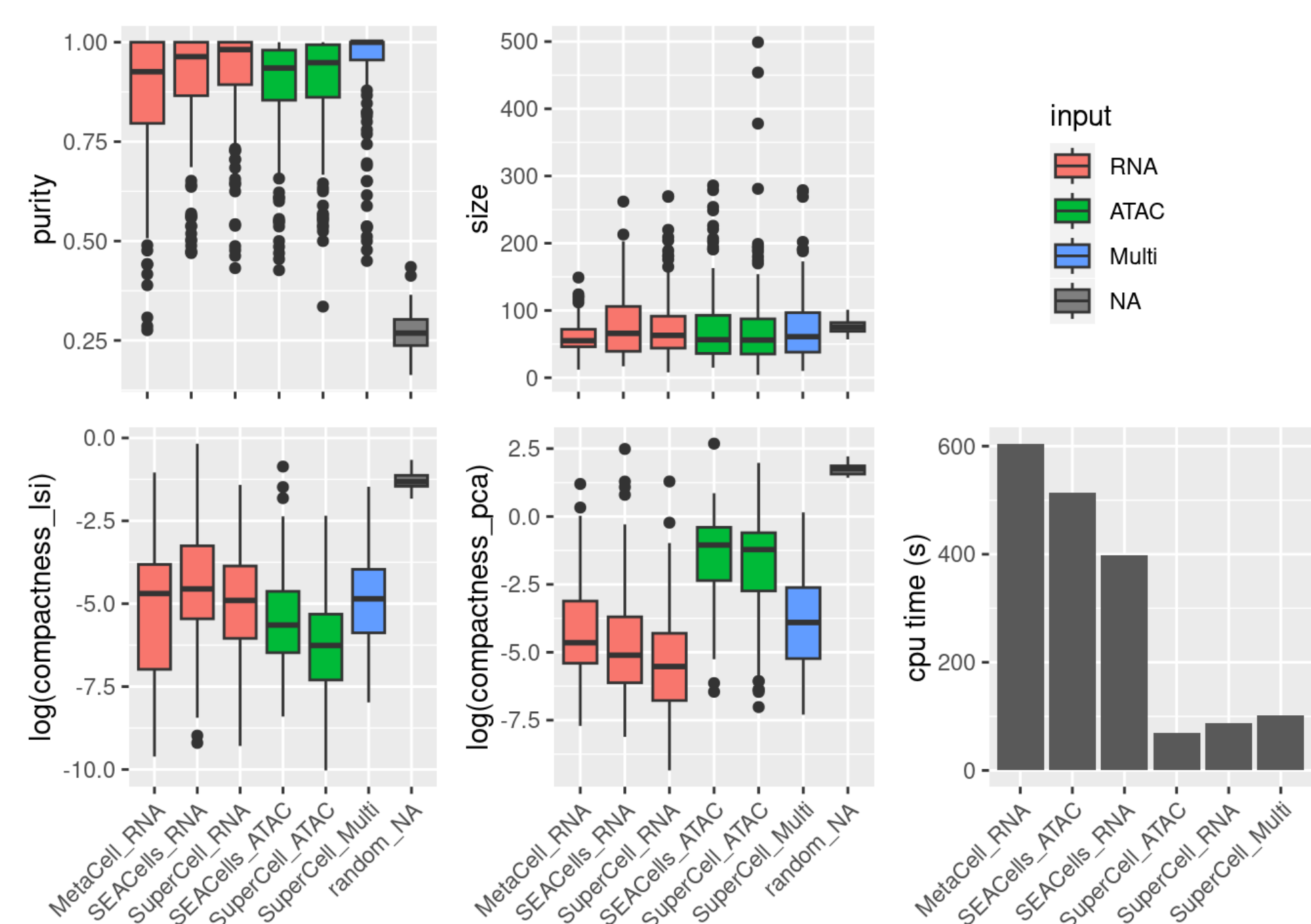


Figure 3: Benchmark of metacell tools on a 10x multiome (RNA + ATAC) dataset of PBMC, graining level  $\gamma=75$ . Tools: new version of SuperCell, SEACells (Persad et al. 2022), MetaCell2 (Ben-Kiki et al. 2021).

## SuperCell Analyses

### 10X multiome dataset of PBMCs

SuperCell identifies robust metacells in the PBMC multiomic space (Fig.4).

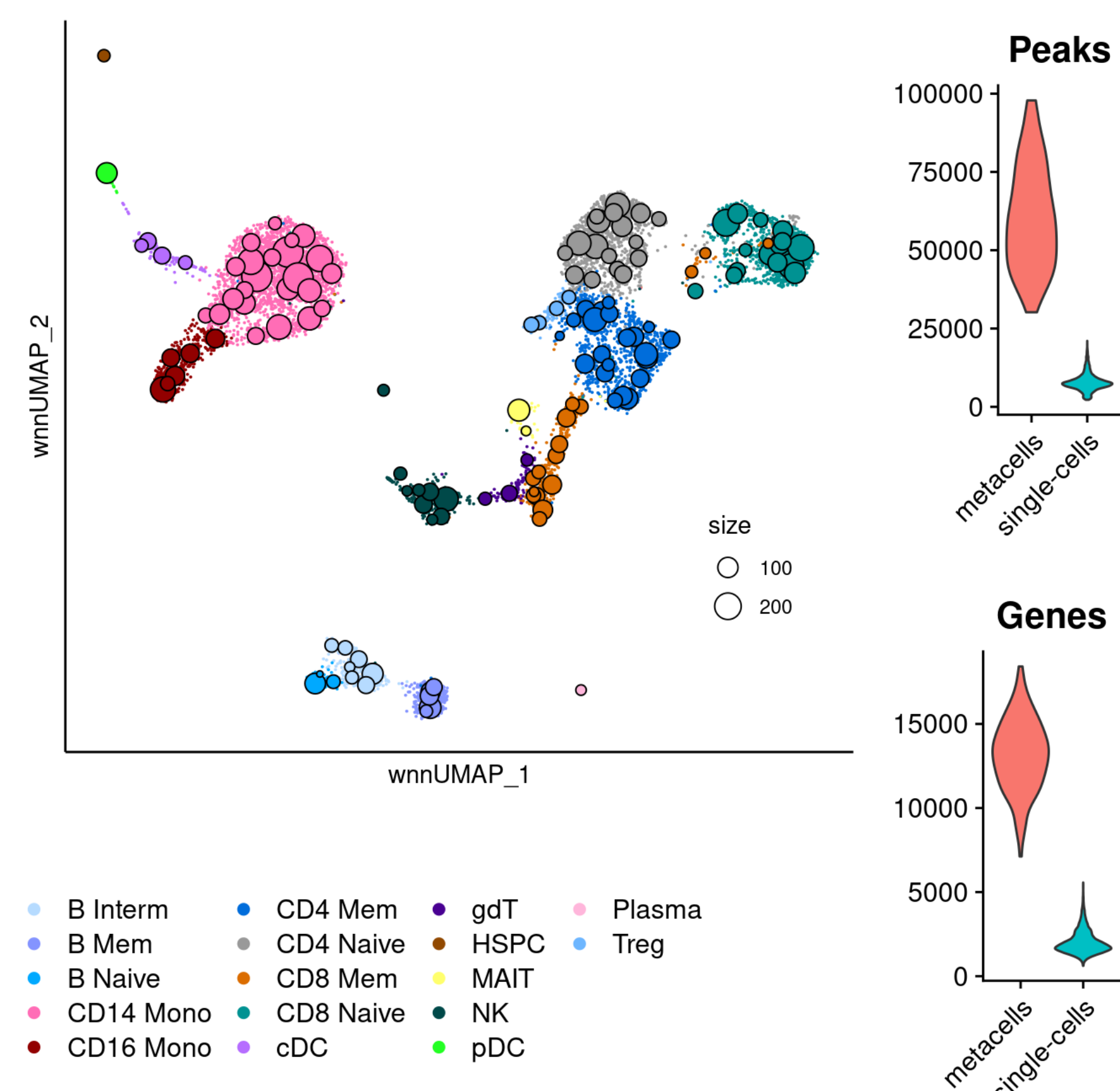


Figure 4: identified metacells in the multimomic space of PBMCs, graining level  $\gamma=75$

Gene accessibility and expression appear more correlated at the metacell level (Fig.5).

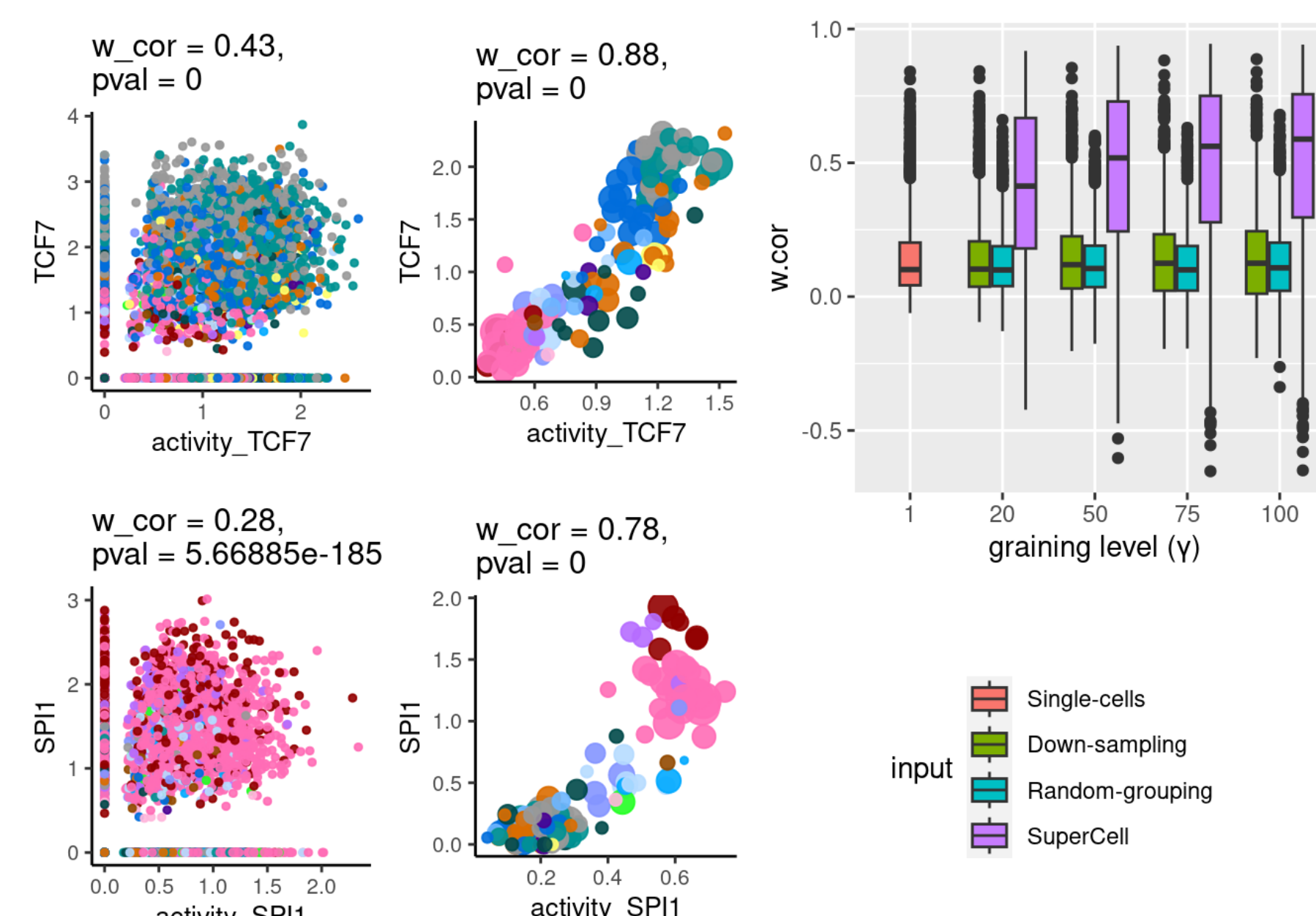


Figure 5: **A.** gene accessibility - gene expression correlation for TCF7 and SPI1. Left: Single-cell level, right: metacell  $\gamma=75$ . Same color legend as in Fig.4. **B.** Same correlations for the 2000 highly variable genes (on RNA) with increasing  $\gamma$ .

Correlation between TF expression and corresponding motif accessibility also becomes clearer using metacells (Fig.6).

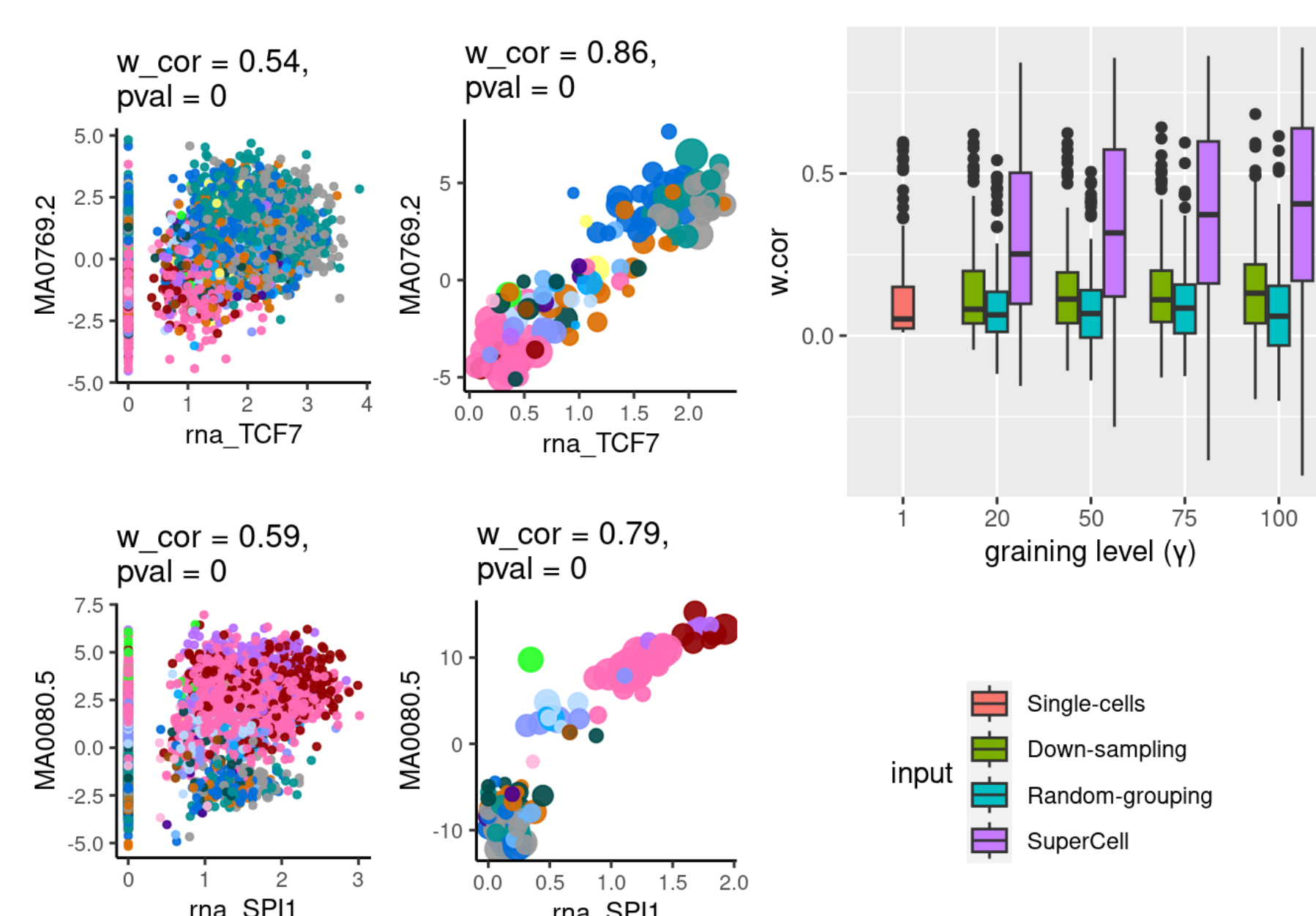


Figure 6: **A.** TF expression (RNA) - motif accessibility (ATAC) correlation for TCF7 and SPI1. Left: single-cell, right: metacells,  $\gamma=75$ . Same color legend as in Fig.4. **B.** Same correlations for 200 TFs (with  $\text{cor} > 0.01$  in single-cells) with increasing  $\gamma$ .

### CITE-seq atlas of 160'000 PBMCs

- Identification of metacells per samples ( $\gamma=20$ )
- Correction of the batch effect at the metacell level (Fig.7A&B).
- This workflow runs on a standard laptop (Fig.7C).

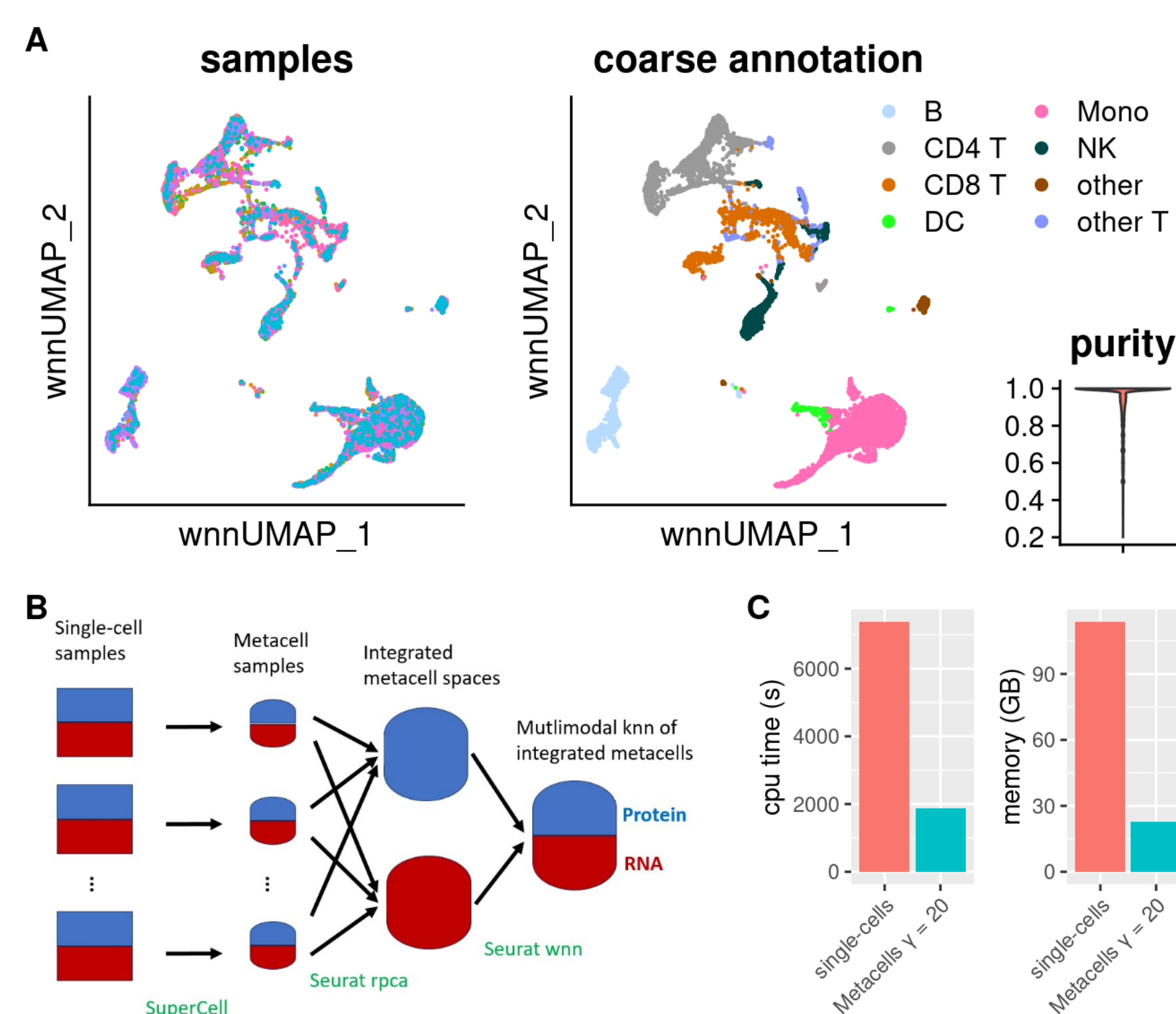


Figure 7: **A.** UMAP visualizations of 8'000 multiomic metacells, metacell purities (original annotation from (Hao et al. 2021)). **B.** Metacell workflow. **C.** Computational resources used by Single-cell and metacell workflows.

RNA-protein correlation in the CITE-seq atlas is increased with metacells (Fig.8).

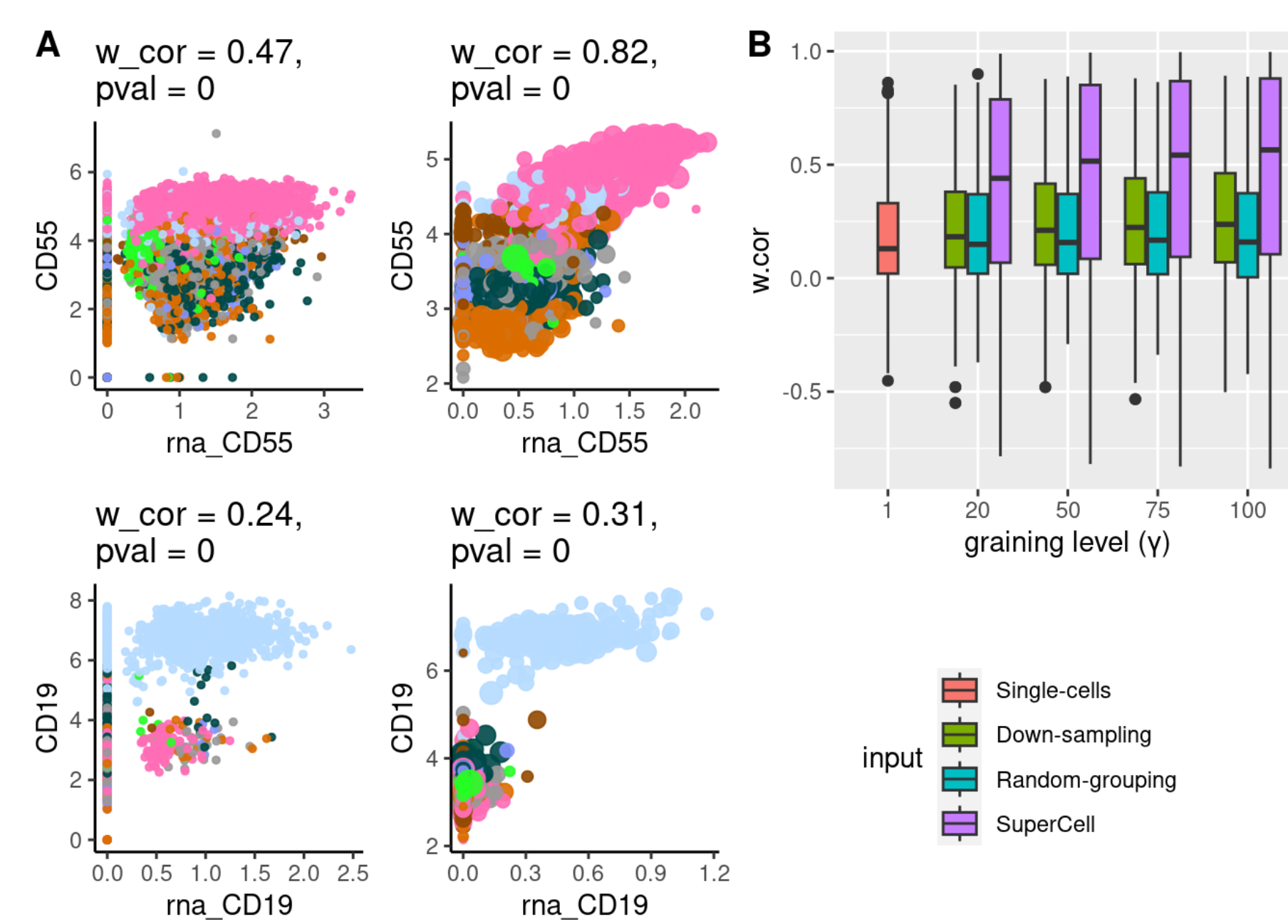
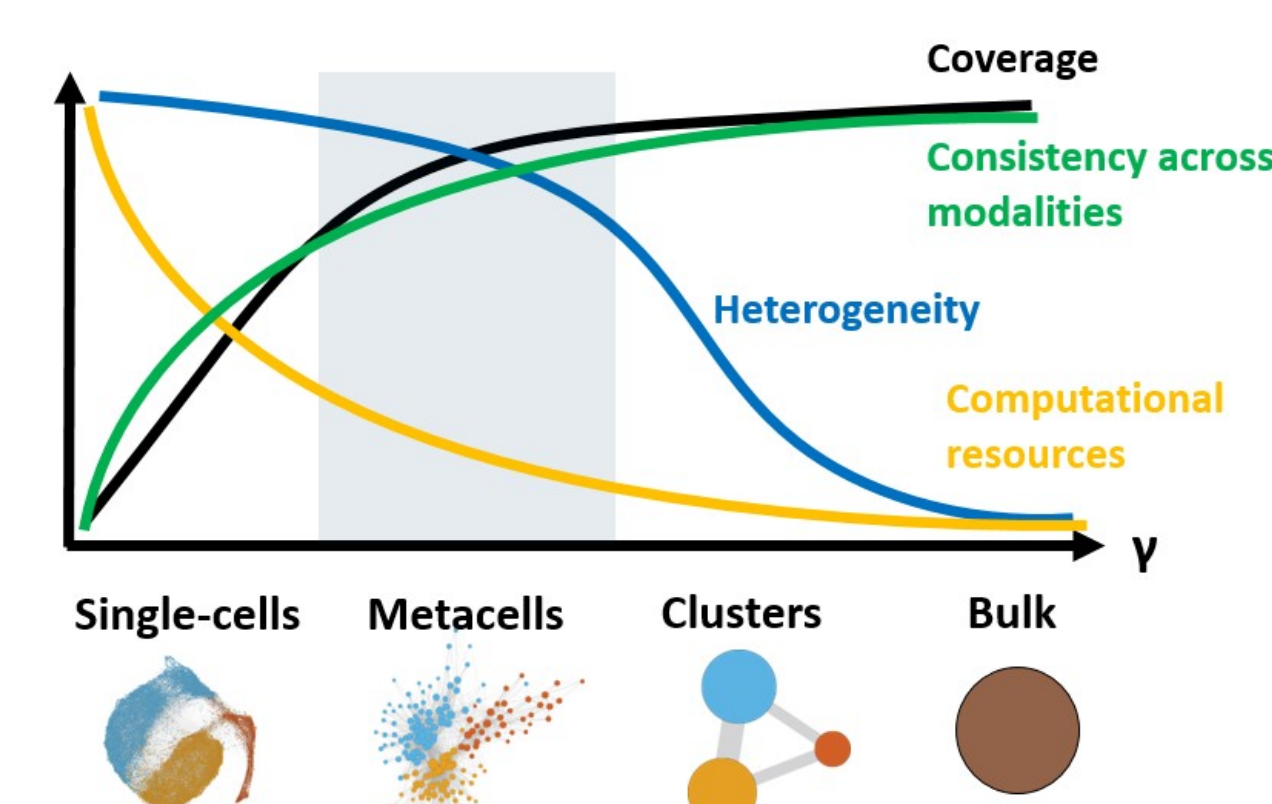


Figure 8: **A.** RNA-protein correlation for CD55 and CD19, same color legend as in Fig.7. **B.** RNA-Protein correlation for 213 gene-protein pairs with increasing  $\gamma$

## Conclusion



## Perspectives

- Downstream analyses of sc-multiomics:
  - gene regulatory network inference
  - multiomic velocity
- (Semi)-supervised identification of metacells
- Metaspots for spatial-omics

## References

Baran, Yael, Akhmad Berovich, Arnau Sebe-Pedros, Yaniv Lubling, Amir Giladi, Elad Chomsky, Zohar Meir, Michael Hirschman, Aviezer Lifshitz, and Amos Tanay. 2019. "MetaCell: Analysis of Single-Cell RNA-Seq Data Using K-NN Graph Partitions." *Genome Biology* 20 (1): 206. <https://doi.org/10.1186/s13059-019-1812-3>.  
Ben-Kiki, Oren, Akhmad Berovich, Aviezer Lifshitz, and Amos Tanay. 2021. "A Divide and Conquer Metacell Algorithm for Scalable scRNA-Seq Analysis." Preprint. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btab2334>.  
Bilous, Mariia, Loc Tran, Chiara Candian, Aurélien Gabriel, Hugo Michel, Santiago J. Carmona, Mikael J. Pittet, and David Gfeller. 2022. "Metacells Untangle Large and Complex Single-Cell Transcriptome Networks." *BMC Bioinformatics* 23 (1): 336. <https://doi.org/10.1186/s12859-022-04860-5>.  
Hao, Yuhao, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shihui Zheng, Andrew Butler, Maddie J. Lee, et al. 2021. "Integrated Analysis of Multimodal Single-Cell Data." *Cell* 184 (13): 3773–3887.e28. <https://doi.org/10.1016/j.cell.2021.06.048>.  
Persad, Shara, Zi-Ning Choo, Christine Dien, Ignas Masilionis, Ronan Chaligné, Tal Navey, Chrysostomos C. Brown, Itzik Peter, Manu Setty, and Dana Peter. 2022. "SEACells: Inference of Transcriptional and Epigenomic Cellular States from Single-Cell Genomics Data." *bioRxiv*. <https://doi.org/10.1101/2022.06.02.486748>.

