Reconciling single-cell multiomics modalities using metacells

Léonard Hérault [1,2,] (iD)

✉ leonard.herault@unil.ch

Aurélie Gabriel[1,2] **Mariia Bilous**[1,2] **David Gfeller**[1,2]

[1] *Department of Oncology, Ludwig Institute for Cancer research, University of Lausanne*
[2] *Swiss Institute of Bioinformatics*

Introduction

The increased throughput of single-cell omics technologies allows researchers to study cell-type specific gene regulation at unprecedented resolution. This promise depends on the development of computational methods to cope with both the large size and the high sparsity of these data. Thus, we and others have developed approaches to identify metacells, disjoint and homogeneous groups of cells, in scRNA-seq data (Baran et al. 2019; Bilous et al. 2022). We are now extending our `SuperCell` method to identify metacells in single-cell multiomics (sc-multiomics) data that combines the measurement of different types of molecules (modalities) in the same single cell.

$\gamma = \dfrac{N \text{ single cells}}{N \text{ metacells}}$

Methods

For sc-multiomics, our new `SuperCell` workflow identifies metacells on a weighted Nearest Neighbor (wNN) graph (Hao et al. 2021). Single-cell raw data from each modality are aggregated per metacells that can be used for downstream analyses.

Datasets

- *10X multiome* data of peripheral blood mononuclear cells (PBMCs): scRNA-seq & ATAC-seq (10'000 cells). We analyzed the correlation between gene expression and activity.
- CITE-seq data from bone marrow (BM) cells: scRNA-seq & 25 protein levels (30'000 cells). We analyzed the correlation between gene expression and the corresponding protein level.

Results

10X multiome dataset of PBMCs

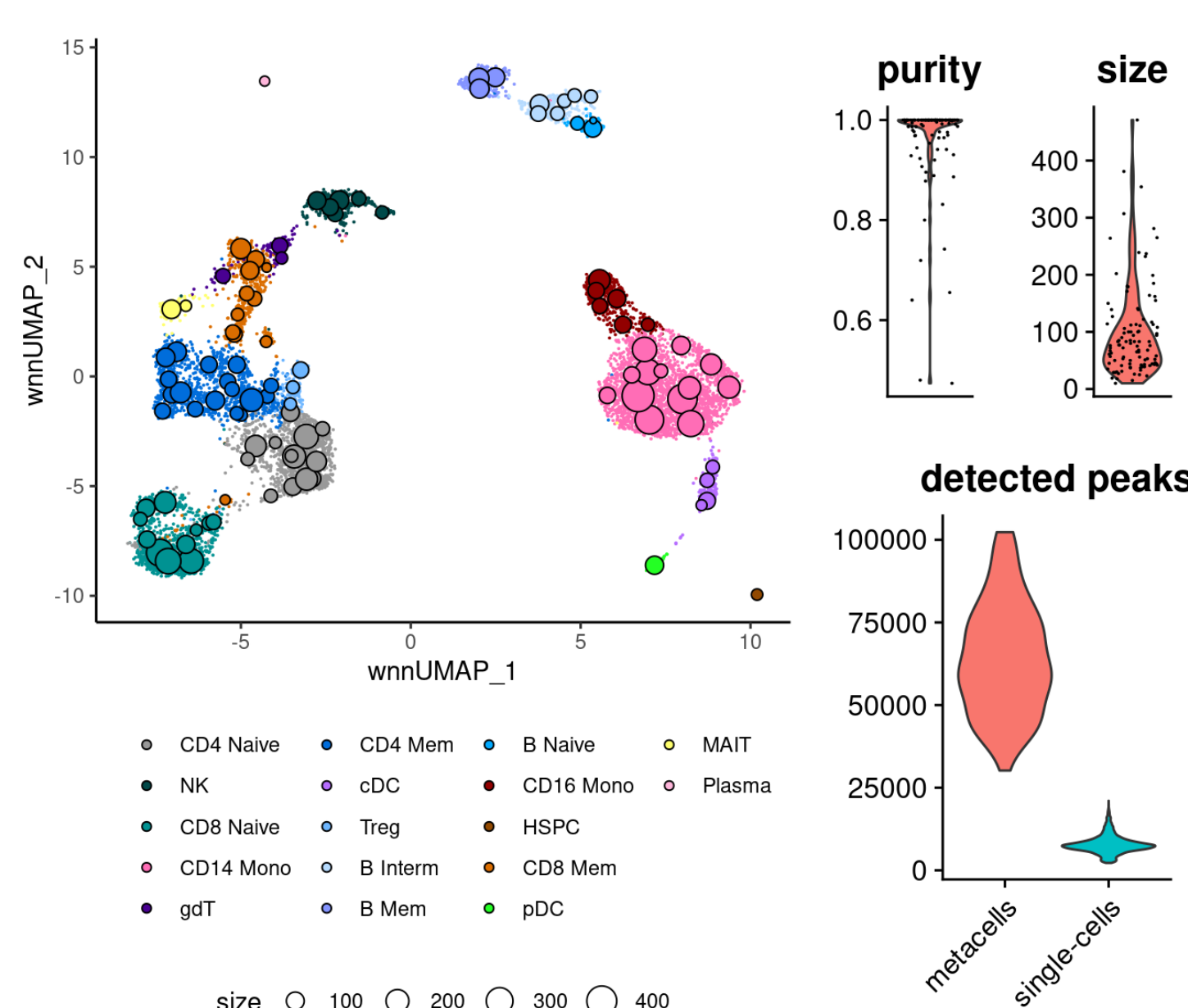`SuperCell` identified robust metacells in the multiomic space of PBMCs. (Fig.1).
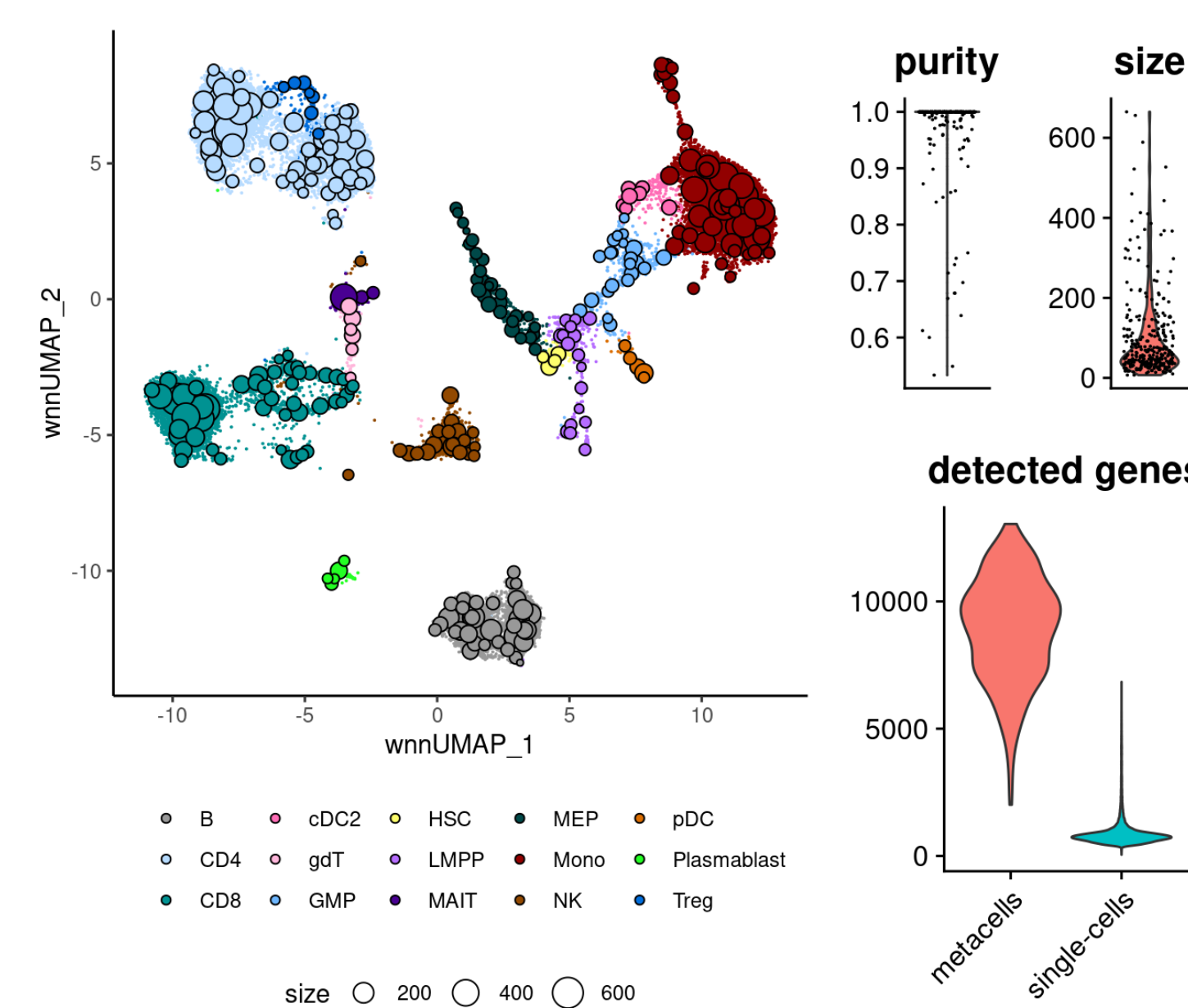
Figure 1: identified metacells in the multimomic space of PBMCs.

Gene activity and expression for key Transcription Factors TFs of lymphoid cells (TCF7) and myeloid cells (SPI1) appeared more correlated at the metacell level (Fig.2).
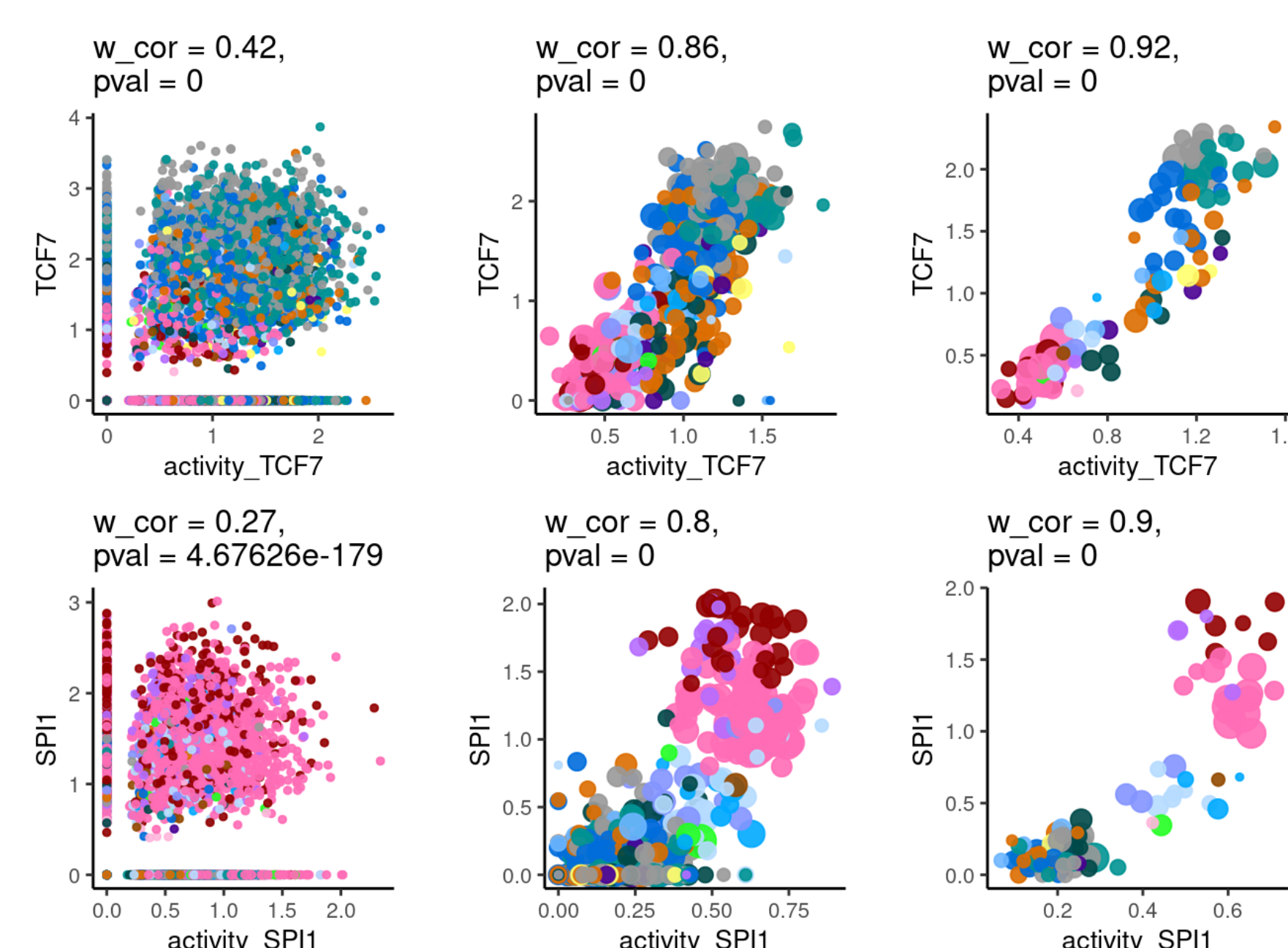
Figure 2: gene expression-activity correlation for key TFs of immune cells. Single-cell level, gamma 20 and gamma 50 from left to right.

Analysis of this correlation for the most variable genes at the transcriptomic level (top 2000) showed the same trend (Fig.3).

Figure 3: Gene expression-activity correlation for 2000 genes at different gammas. The dataset was subsampled at each gamma for comparison.

CITE-seq dataset of BM cells

`SuperCell` identified robust metacell in the CITE-seq space of BM cells (Fig.4).

Figure 4: Identified metacell in the CITE-seq space of BM cells

Metacell analysis increased the correlation between RNA and protein levels of key markers of HSCs (CD34) and CD8 T cells (CD8A, Fig.5).
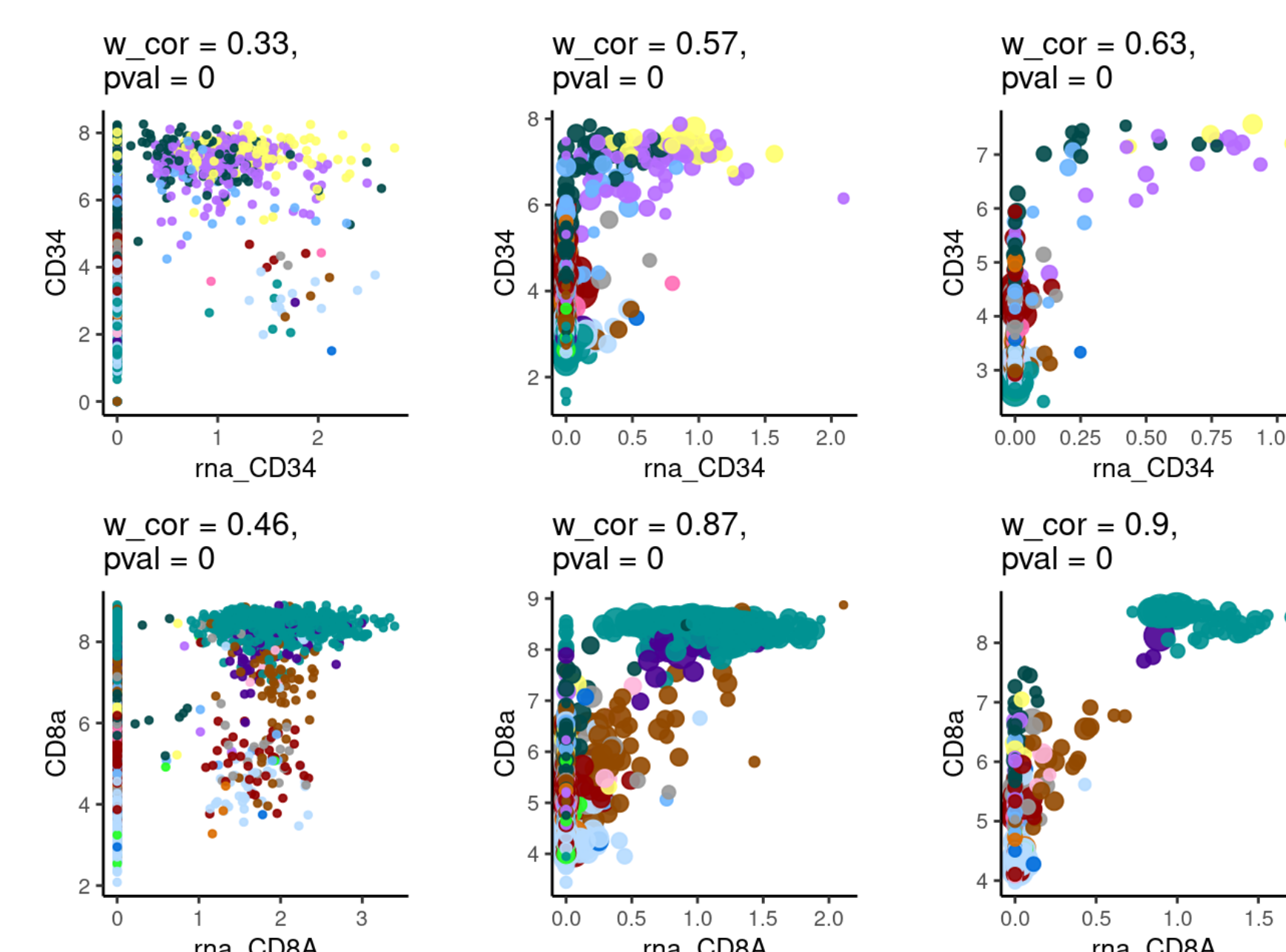
Figure 5: RNA protein correlations for key markers of BM cells. Single-cell level, gamma 20 and gamma 50 from left to right.

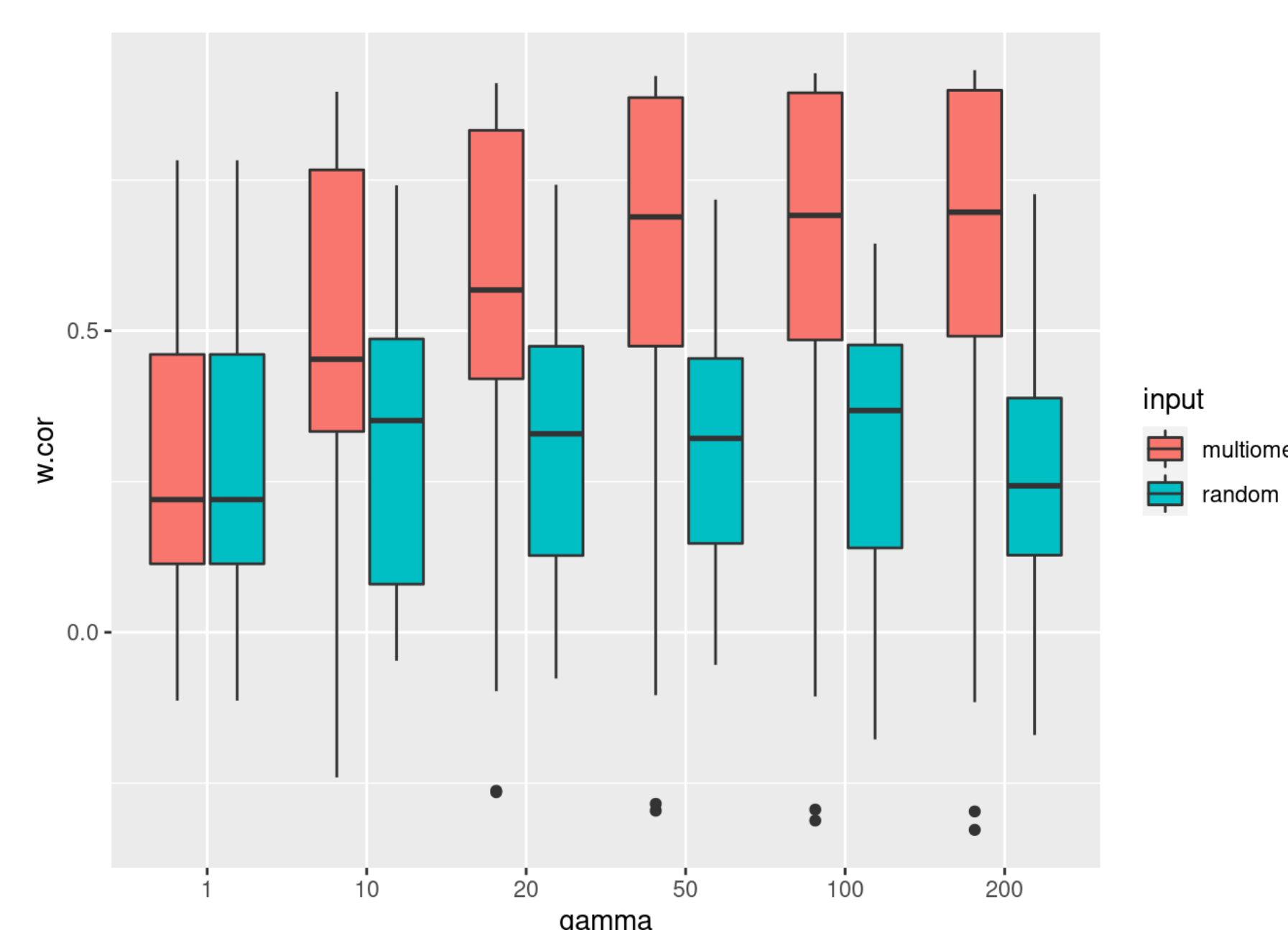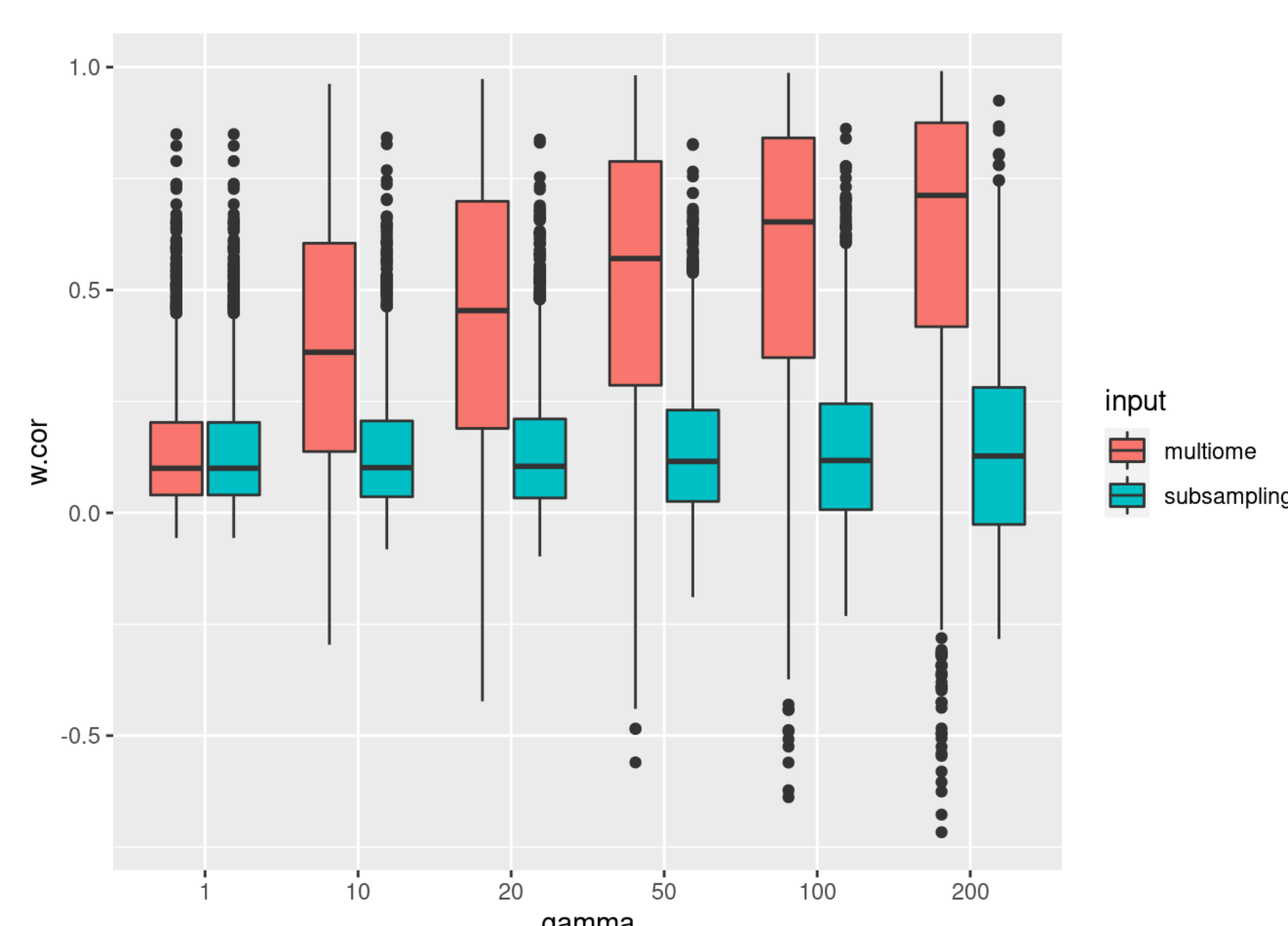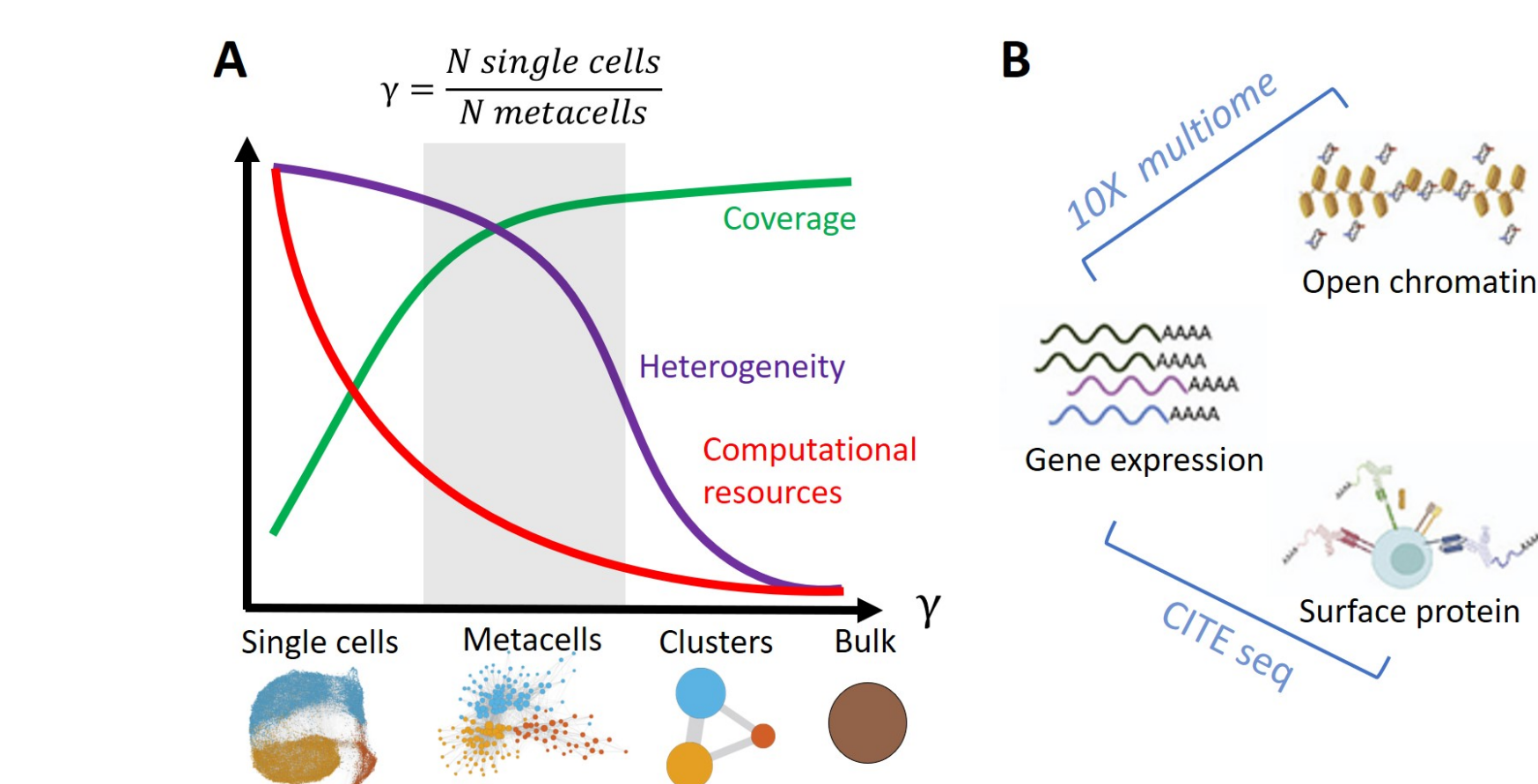RNA-protein correlation analysis for the 25 gene-protein pairs exhibited the same trend (Fig.6).

Figure 6: RNA-Protein correlation for all gene-protein pairs. Random metacells were defined at each gamma for comparison.
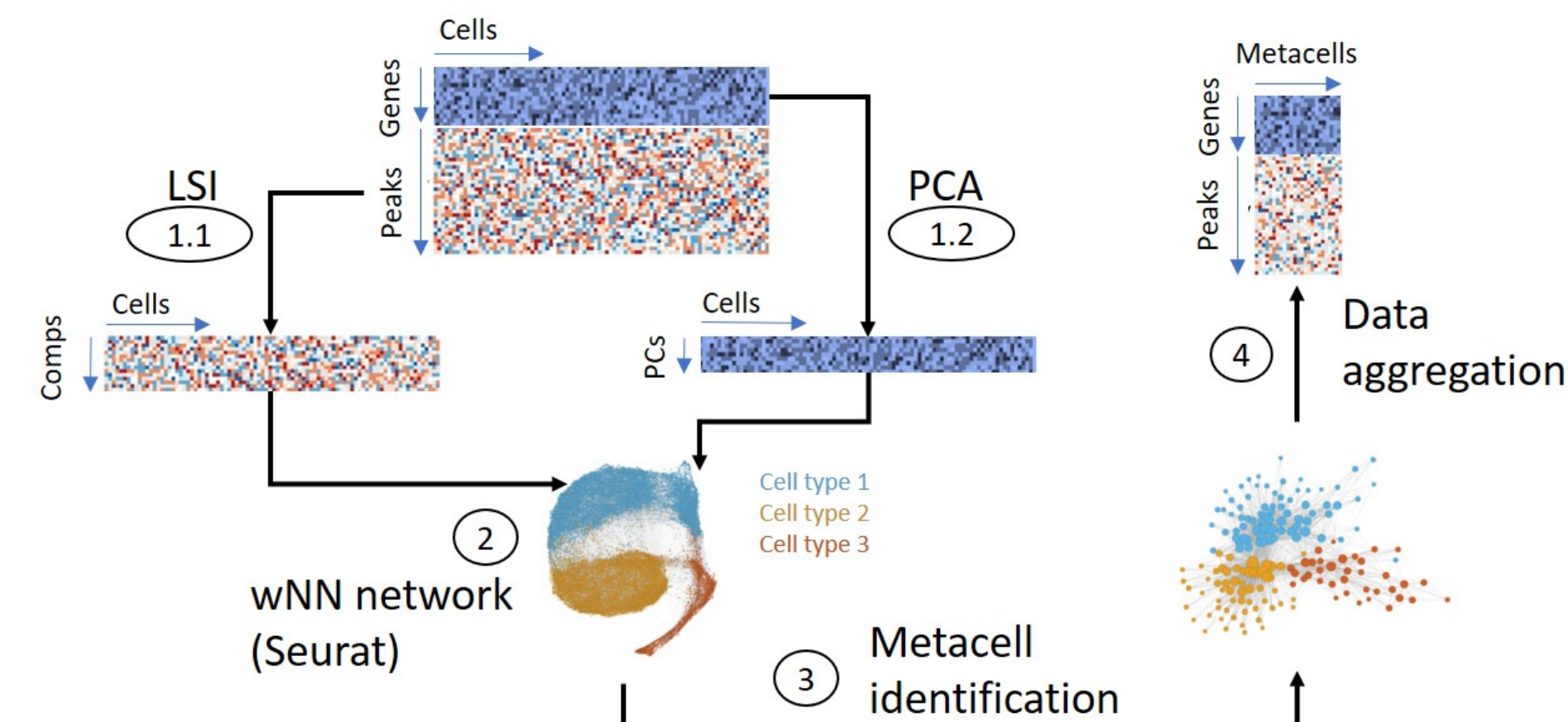
References

Baran, Yael, Akhiad Bercovich, Arnau Sebe-Pedros, Yaniv Lubling, Amir Giladi, Elad Chomsky, Zohar Meir, Michael Hoichman, 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/s13059-019-1812-2.
Bilous, Mariia, Loc Tran, Chiara Cianciaruso, Aurélie 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/s12859-022-04861-1.
Hao, Yuhan, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shiwei Zheng, Andrew Butler, Maddie J. Lee, et al. 2021. "Integrated Analysis of Multimodal Single-Cell Data." *Cell* 184 (13): 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.