Ultrafast and interpretable single-cell 3D genome analysis with Fast-Higashi

(SUPPLEMENTAL INFORMATION)

A Supplementary Figures

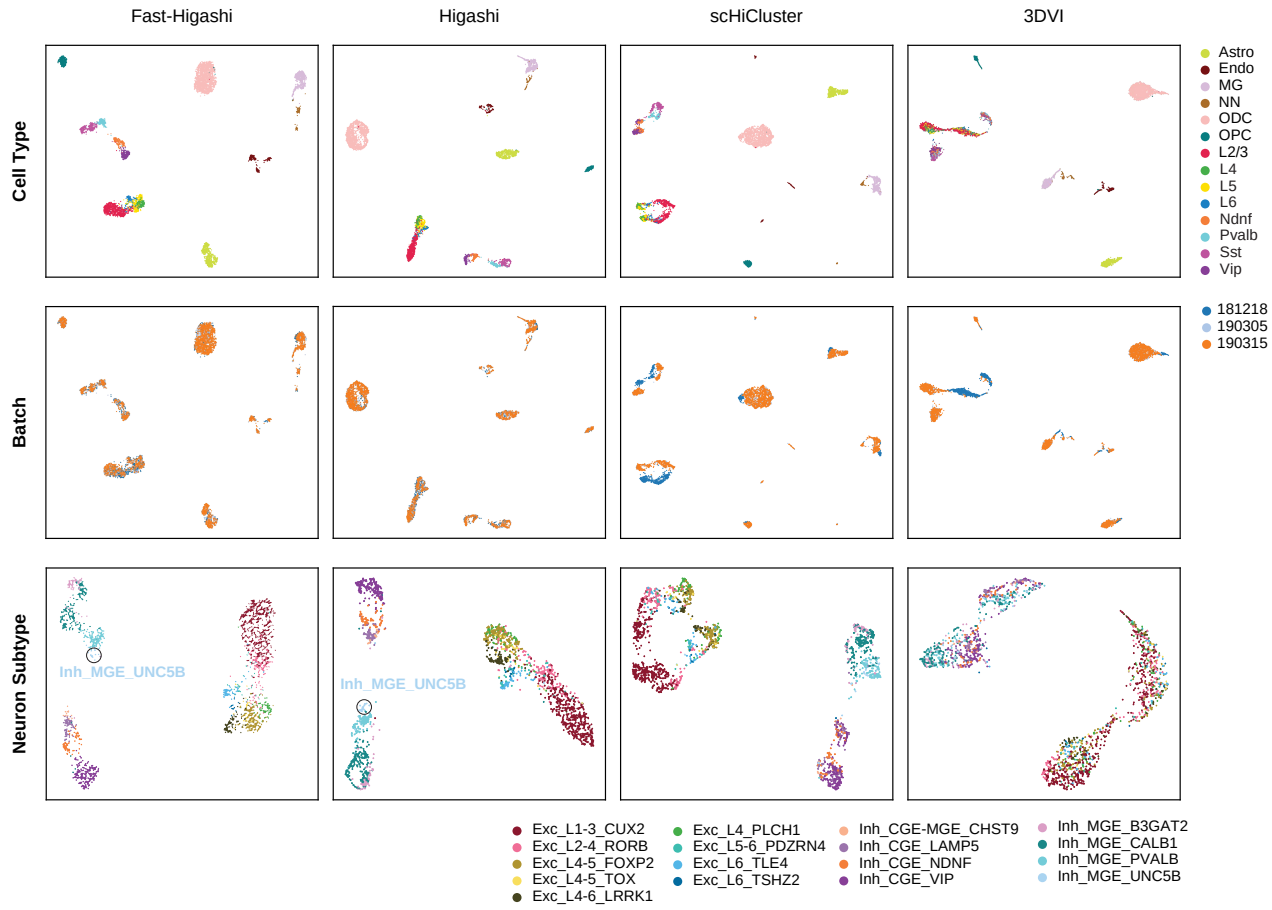

Figure S1: UMAP visualization of the embeddings from different scHi-C embedding methods for the Lee et al. (2019) dataset. Scatter plots are colored with the cell type information from the original data source, batch id, and the more refined cell type information from Luo et al. (2022) (from top to bottom).

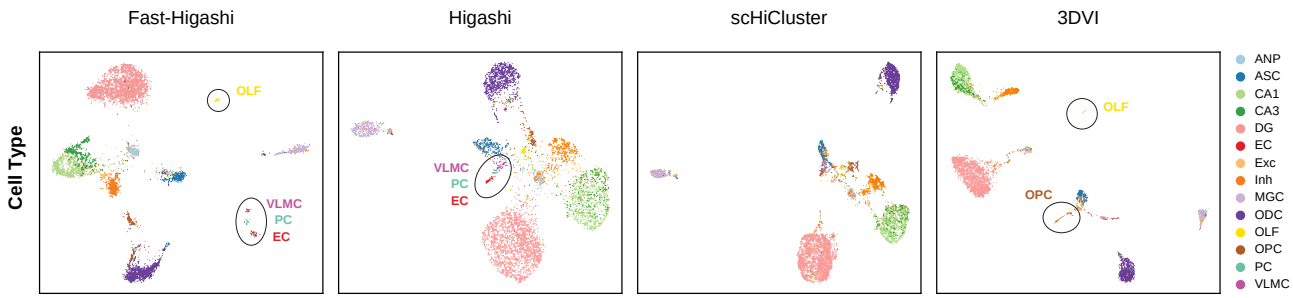

Figure S2: UMAP visualization of the embeddings from different scHi-C embedding methods for the Liu et al. (2021) dataset.

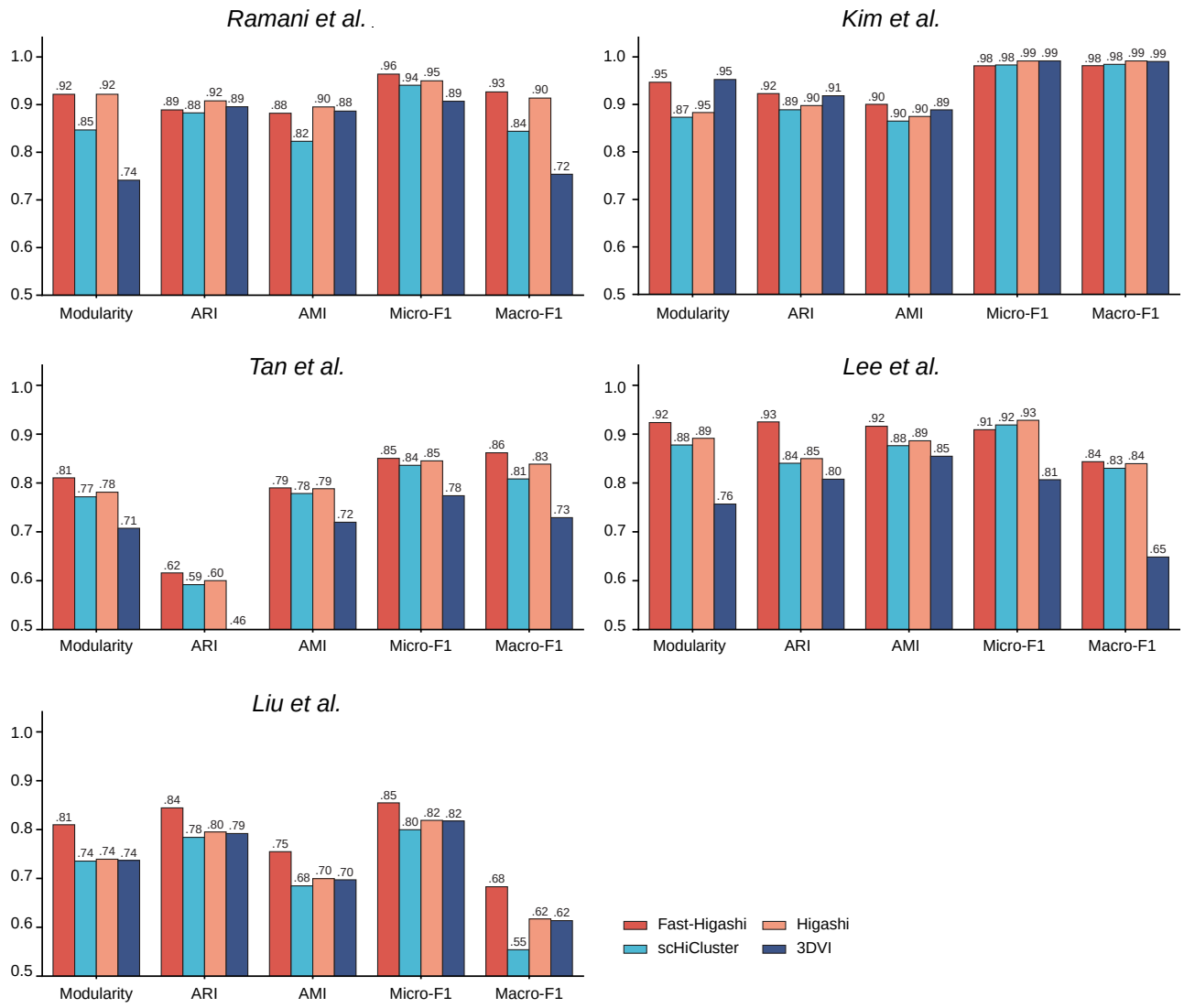

Figure S3: Quantitative evaluation for different scHi-C embeddings methods across different scHi-C datasets. The evaluation metrics include the modularity scores, similarity scores (adjusted rand index and adjusted mutual information) between the Louvain clustering results and the reference cell type label, and measurement of prediction accuracy (Micro-F1 and Macro-F1) between the predicted cell type and the reference cell type.

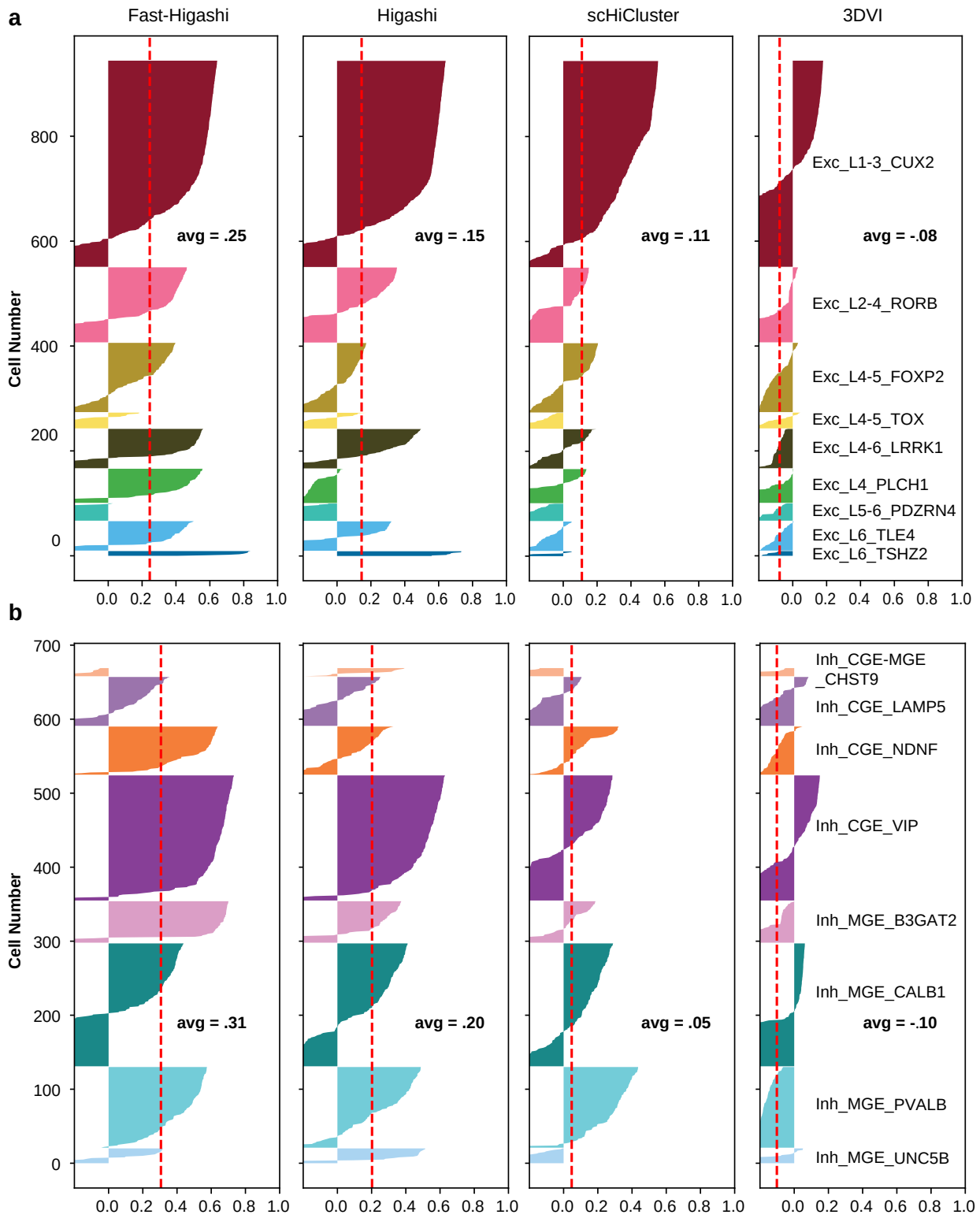

Figure S4: Quality of the embeddings for the neuron cells in the Lee et al. (2019) dataset measured as silhouette coefficients for neuron subtypes.

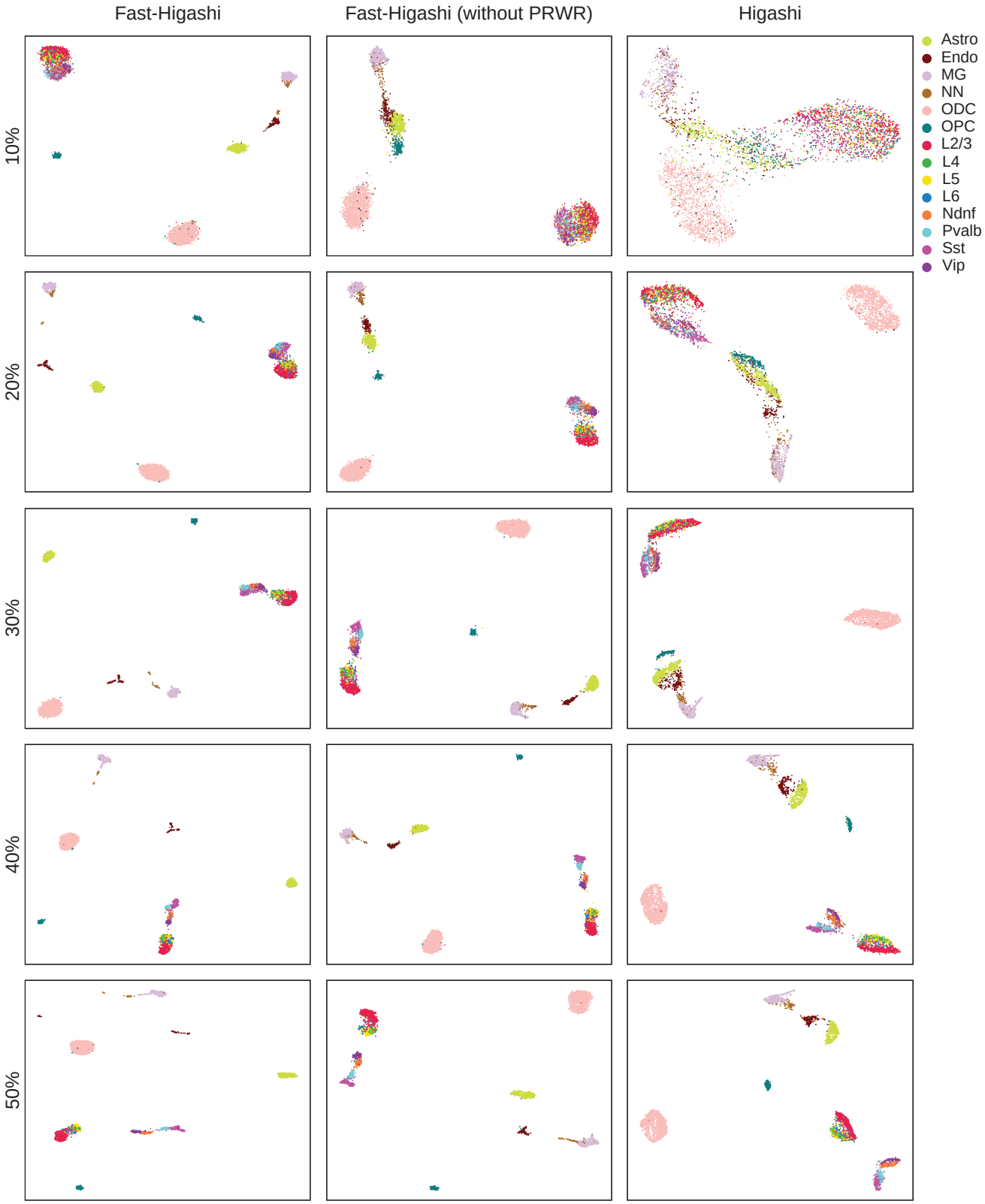

Figure S5: UMAP visualization of embeddings generated by Higashi, Fast-Higashi, and Fast-Higashi without PRWR for the Lee et al. dataset under different downsample ratio. Scatter plots are colored with the reference cell type. See also Fig. S6.

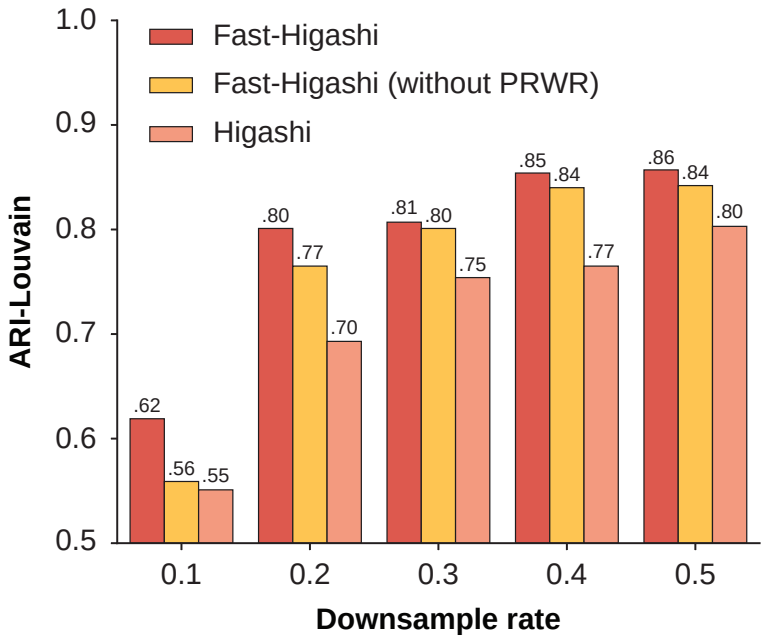

Figure S6: Quantitative evaluation based on adjusted rand index (ARI) scores of the Louvain clustering results for embeddings generated by Higashi, Fast-Higashi, and Fast-Higashi without PRWR. The evaluation was carried out on the Lee et al. dataset with different downsample ratio. Related to Fig. S5.

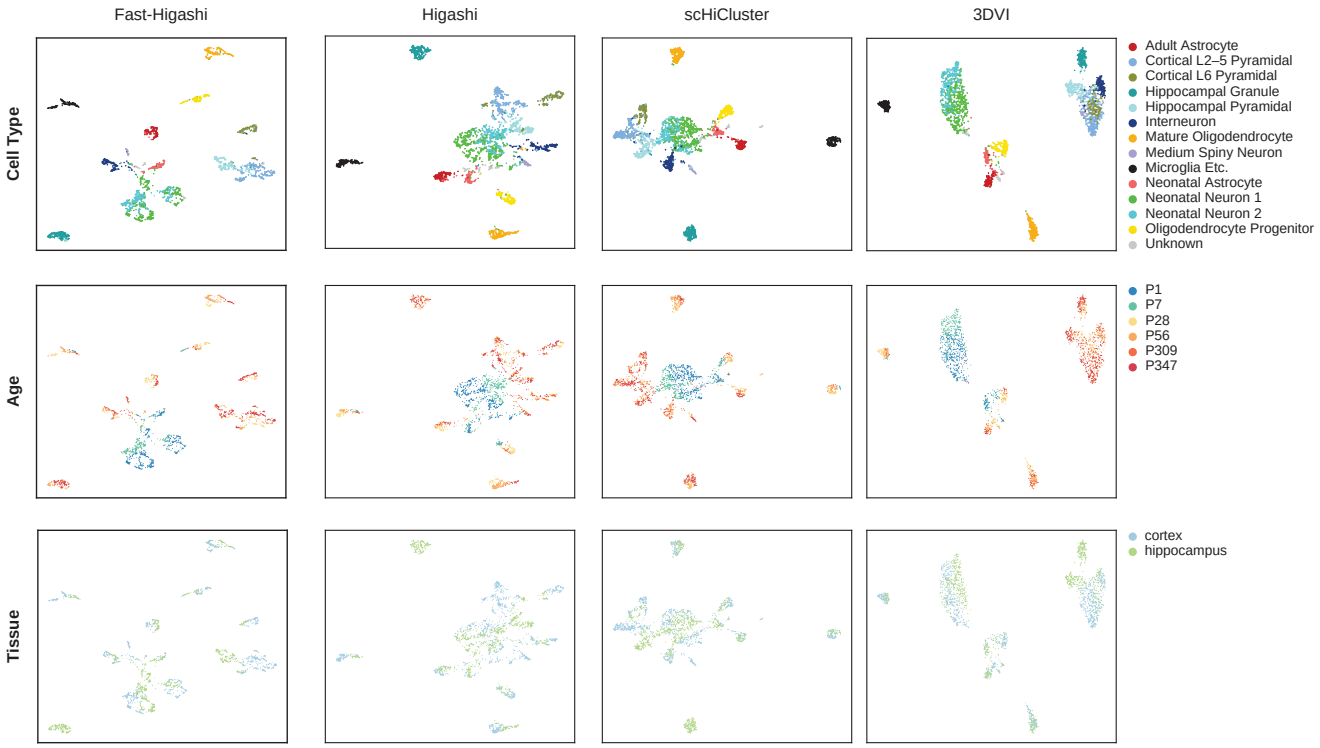

Figure S7: UMAP visualization of the embeddings from different scHi-C embedding methods for the Tan et al. (2021) dataset. Scatter plots are colored with the cell type information, the ages of the mouse, and the tissue origin of the cells (from top to bottom).

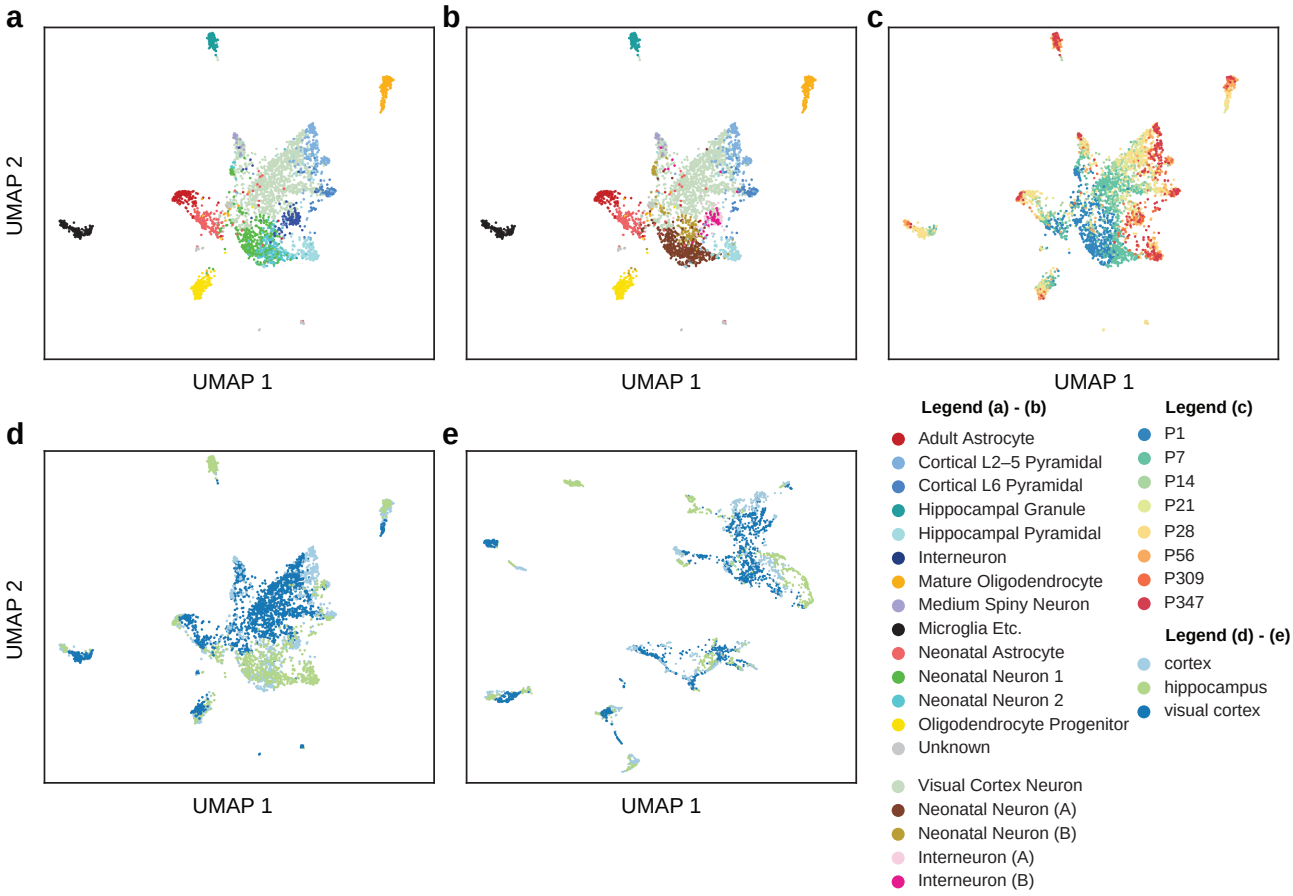

Figure S8: a. b. c. d. UMAP visualization of the joint embeddings from scHiCluster for visual cortex, hippocampus, and cortex tissues from the Tan et al. (2021) dataset. Colors reflect the original cell type information from Tan et al. (2021), the refined cell type from this work, the ages of the mouse, and the tissue of origin of the cells, respectively. **e.** UMAP visualization of Fast-Higashi joint embeddings for visual cortex, hippocampus, and cortex tissues from the Tan et al. (2021) dataset (colored by the tissue of the cells).

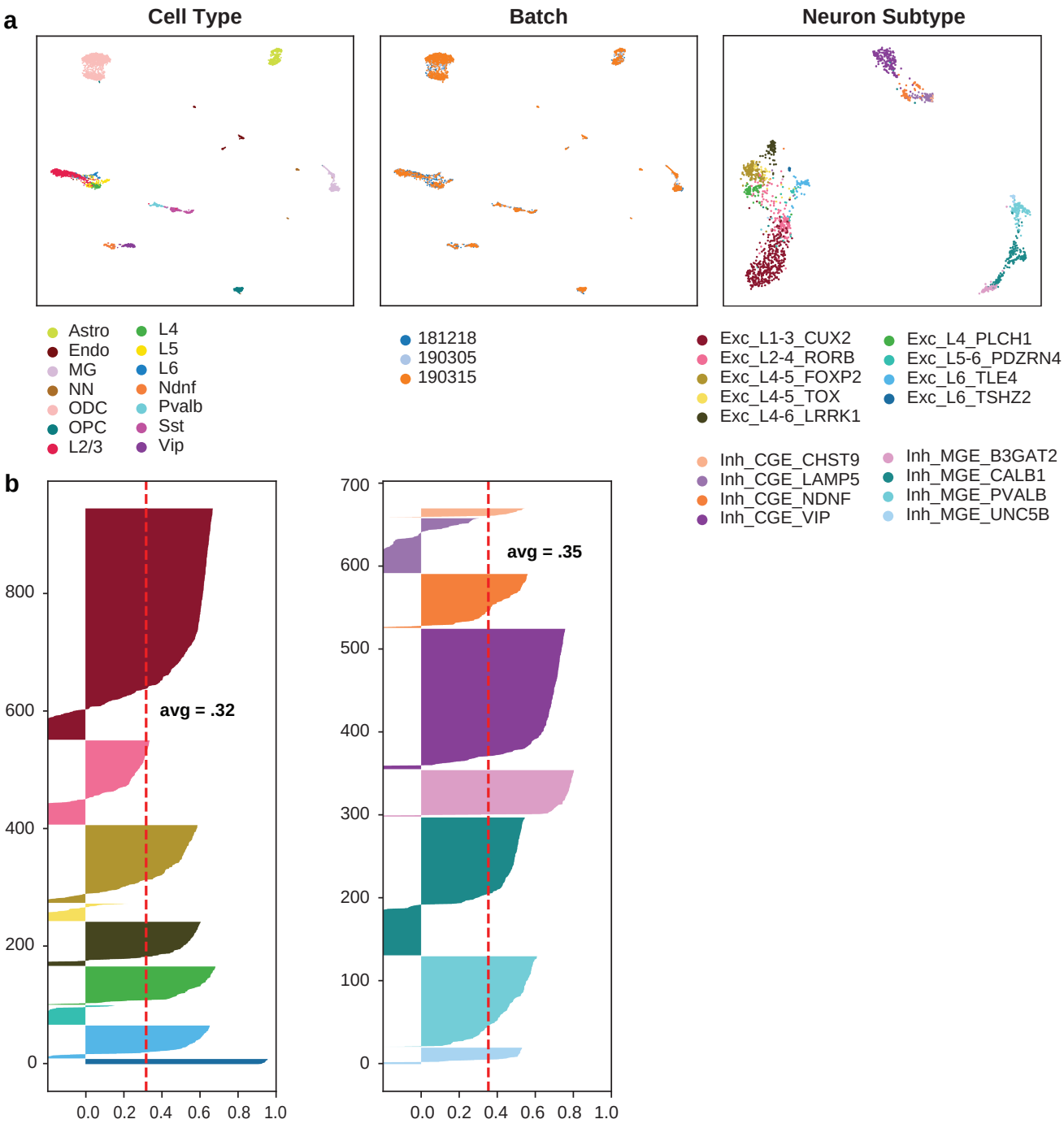

Figure S9: Embeddings from the Fast-Higashi initialized Higashi model on the Lee et al. (2019) dataset. **a.** UMAP visualization of embeddings. Scatter plots are colored with cell type information from the original data source, the more refined cell type information from Luo et al. (2022), and the batch id. **b.** Quality of the embeddings for the neuron cells in the Lee et al. (2019) dataset measured as silhouette coefficients for neuron subtypes.

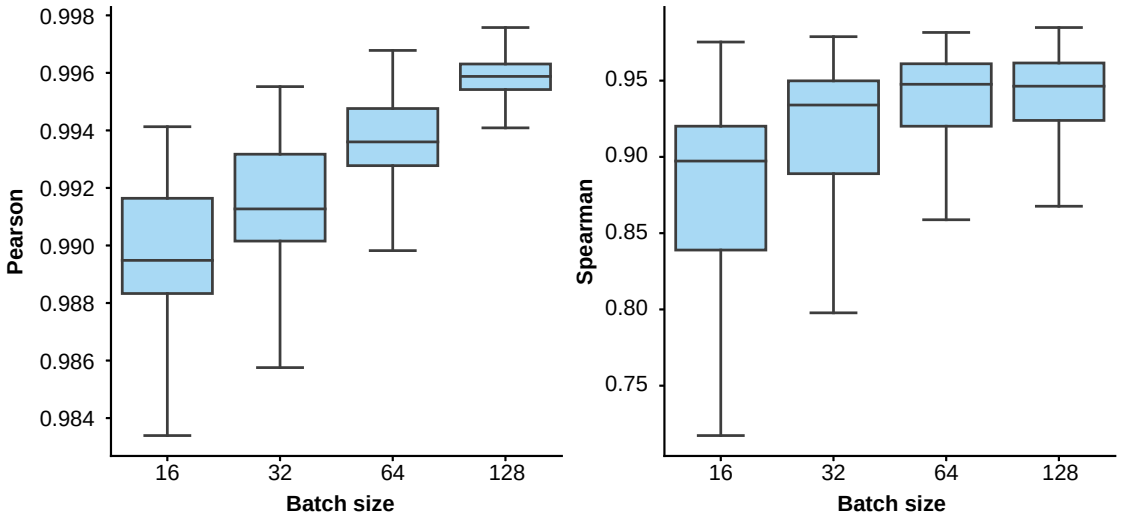

Figure S10: Similarity between the imputed results from the Partial RWR and the standard RWR with varying batch sizes.

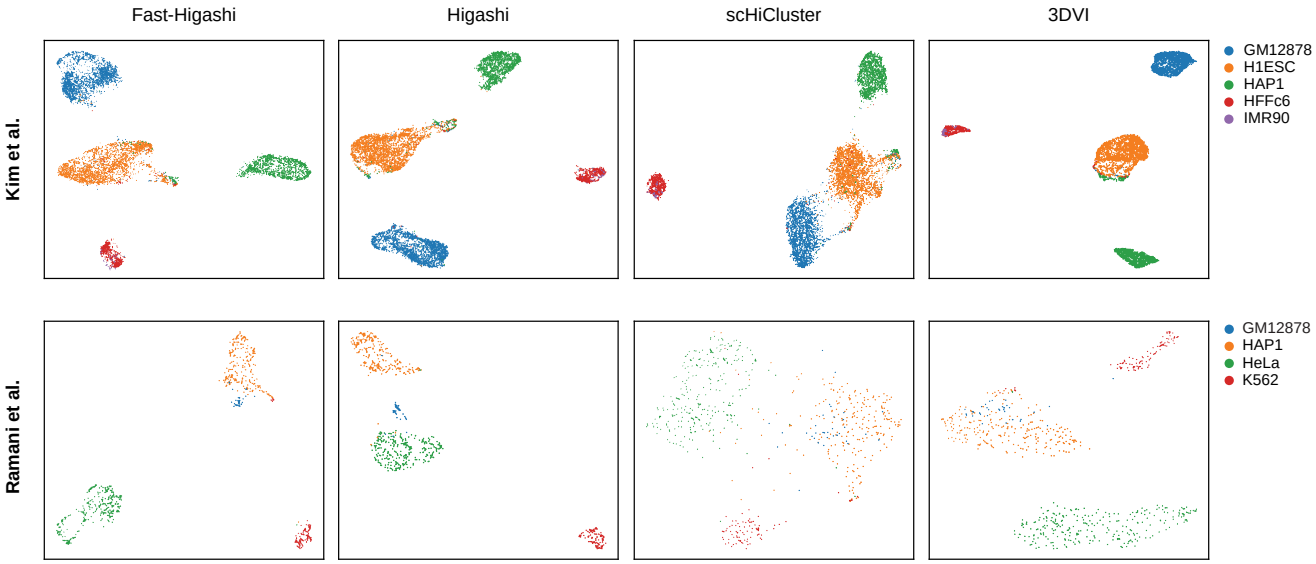

Figure S11: UMAP visualization of the embeddings from different scHi-C embedding methods for two sci-Hi-C datasets: the Ramani et al. (2017) dataset and the Kim et al. (2020) dataset.

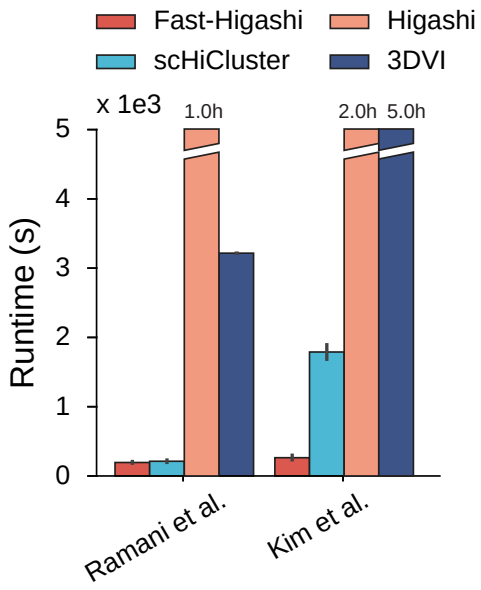

Figure S12: Runtime of different scHi-C embedding methods on two sci-Hi-C datasets.

References

H.-J. Kim, G. G. Yardımcı, G. Bonora, V. Ramani, J. Liu, R. Qiu, C. Lee, J. Hesson, C. B. Ware, J. Shendure, et al. Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. *PLoS Computational Biology*, 16(9):e1008173, 2020.

D.-S. Lee, C. Luo, J. Zhou, S. Chandran, A. Rivkin, A. Bartlett, J. R. Nery, C. Fitzpatrick, C. O'Connor, J. R. Dixon, et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nature Methods*, 16(10):999–1006, 2019.

H. Liu, J. Zhou, W. Tian, C. Luo, A. Bartlett, A. Aldridge, J. Lucero, J. K. Osteen, J. R. Nery, H. Chen, et al. DNA methylation atlas of the mouse brain at single-cell resolution. *Nature*, 598(7879):120–128, 2021.

C. Luo, H. Liu, F. Xie, E. J. Armand, K. Siletti, T. E. Bakken, R. Fang, W. I. Doyle, T. Stuart, R. D. Hodge, et al. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genomics*, 2(3):100107, 2022.

V. Ramani, X. Deng, R. Qiu, K. L. Gunderson, F. J. Steemers, C. M. Disteche, W. S. Noble, Z. Duan, and J. Shendure. Massively multiplex single-cell Hi-C. *Nature Methods*, 14(3):263, 2017.

L. Tan, W. Ma, H. Wu, Y. Zheng, D. Xing, R. Chen, X. Li, N. Daley, K. Deisseroth, and X. S. Xie. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell*, 184(3):741–758, 2021.