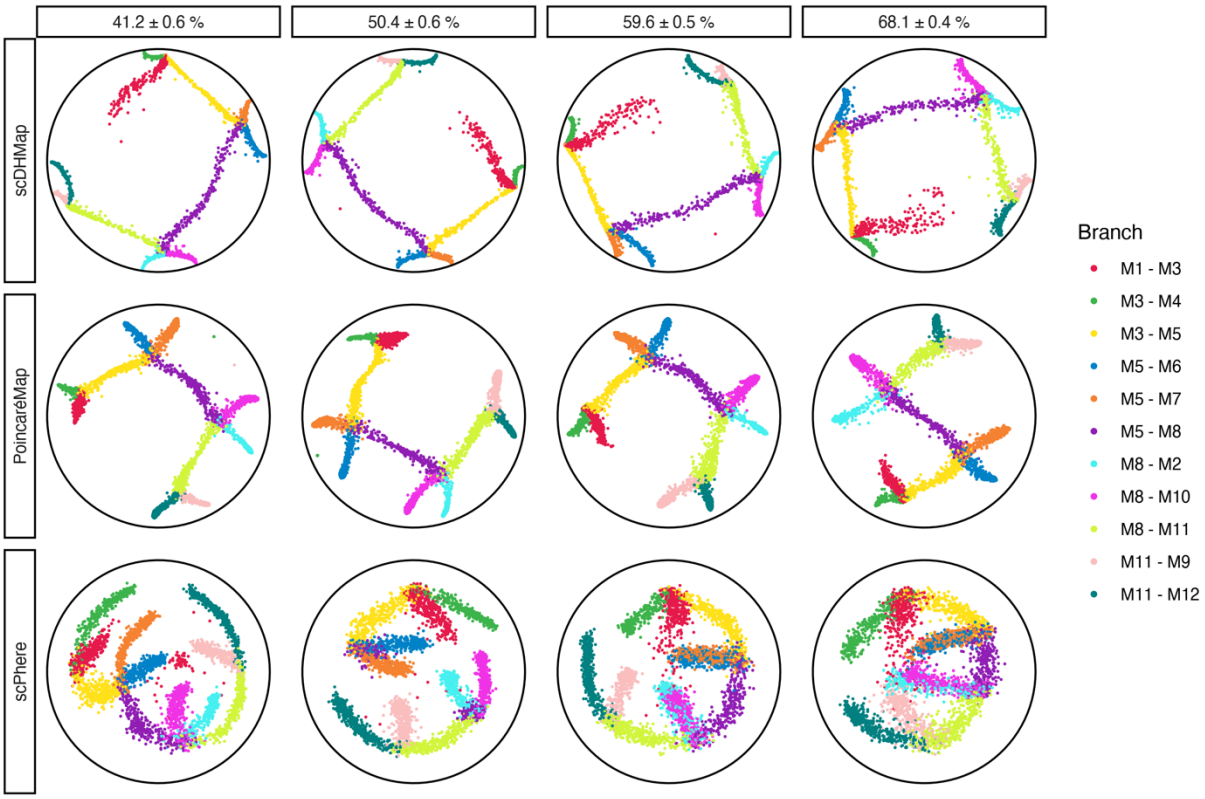
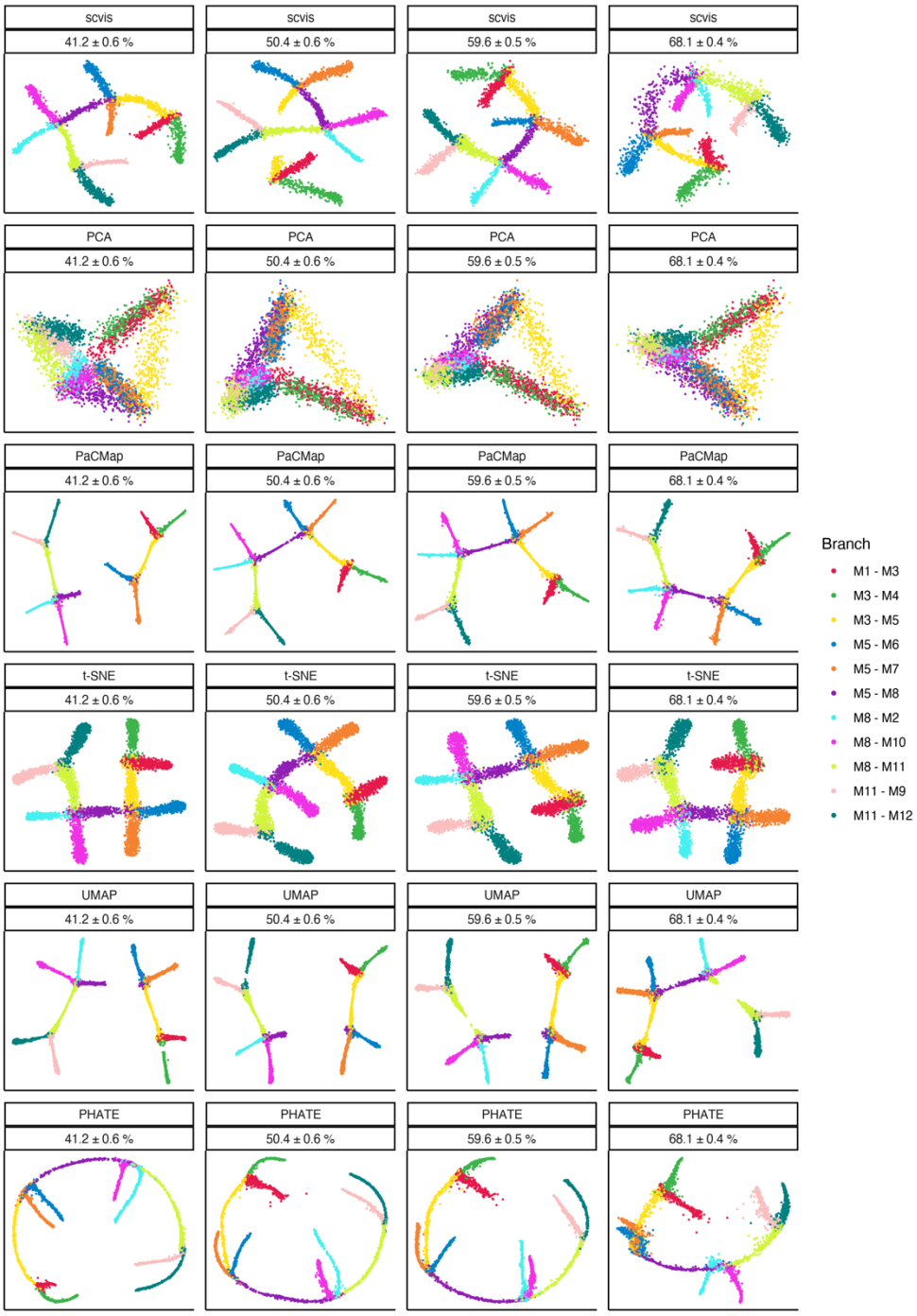
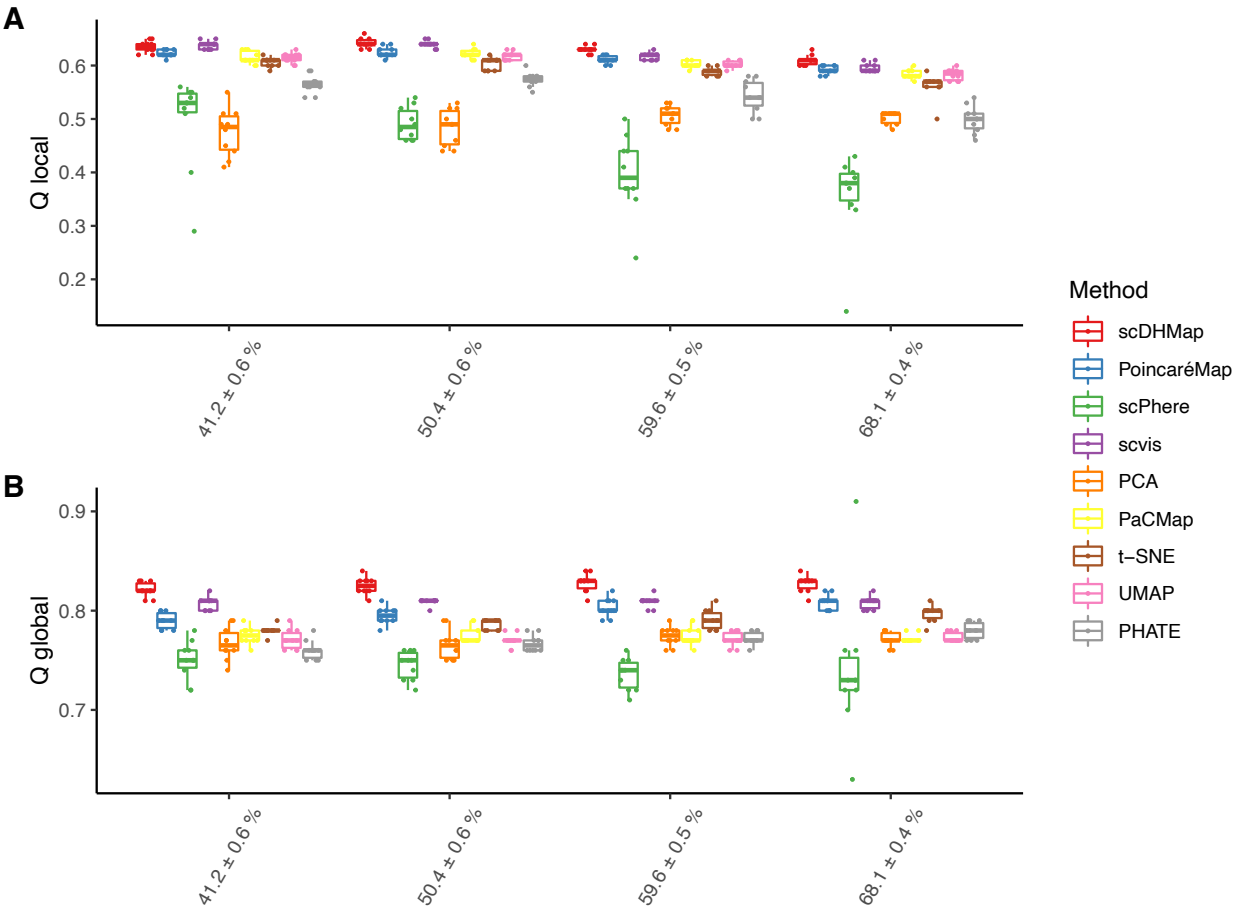
Supplementary Figures



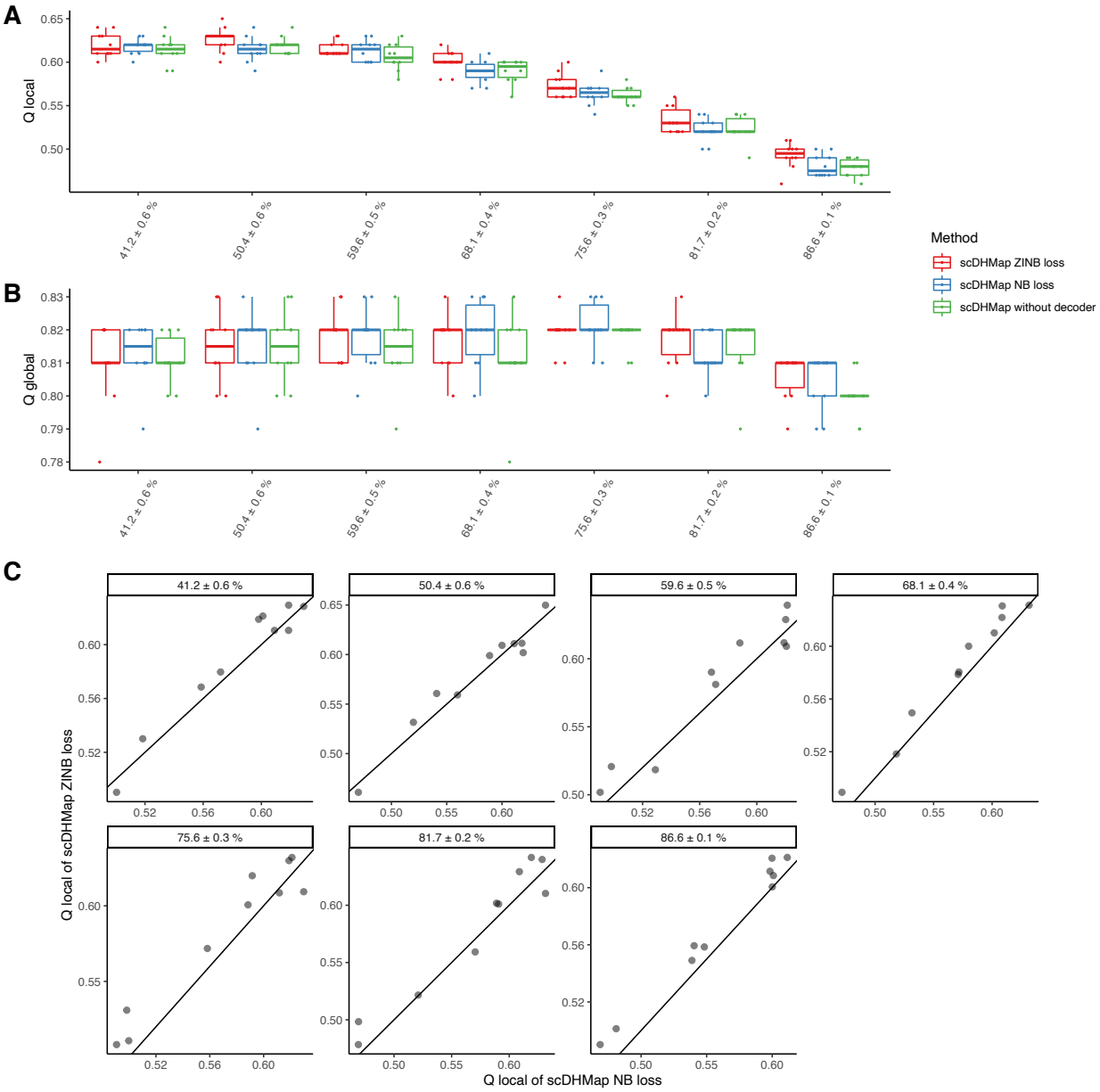
Supplementary Figure S1. 2-dimensional embeddings of different methods acting on hyperbolic space on simulated datasets with various dropout rates. Rows stand for methods, and columns stand for various dropout rates. Colors represent branches in the data.



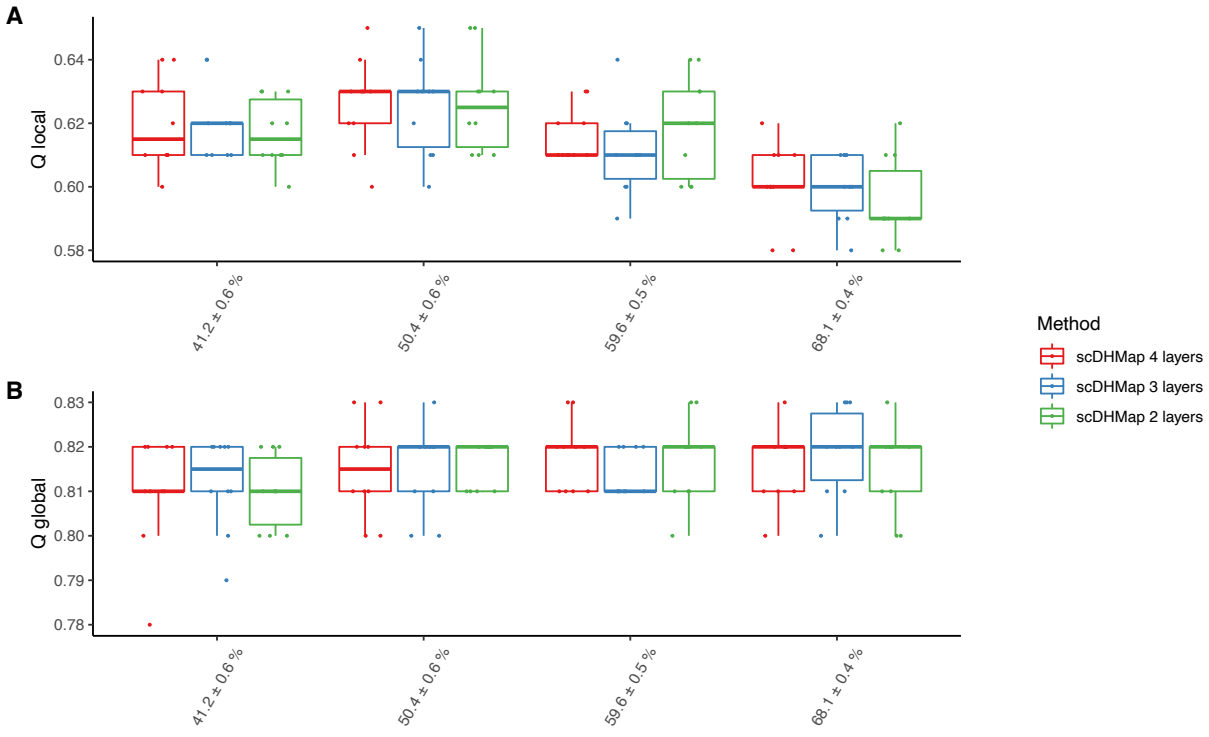
Supplementary Figure S2. 2-dimensional embeddings of different methods acting on Euclidean space on simulated datasets with various dropout rates. Rows stand for methods, and columns stand for various dropout rates. Colors represent branches in the data.



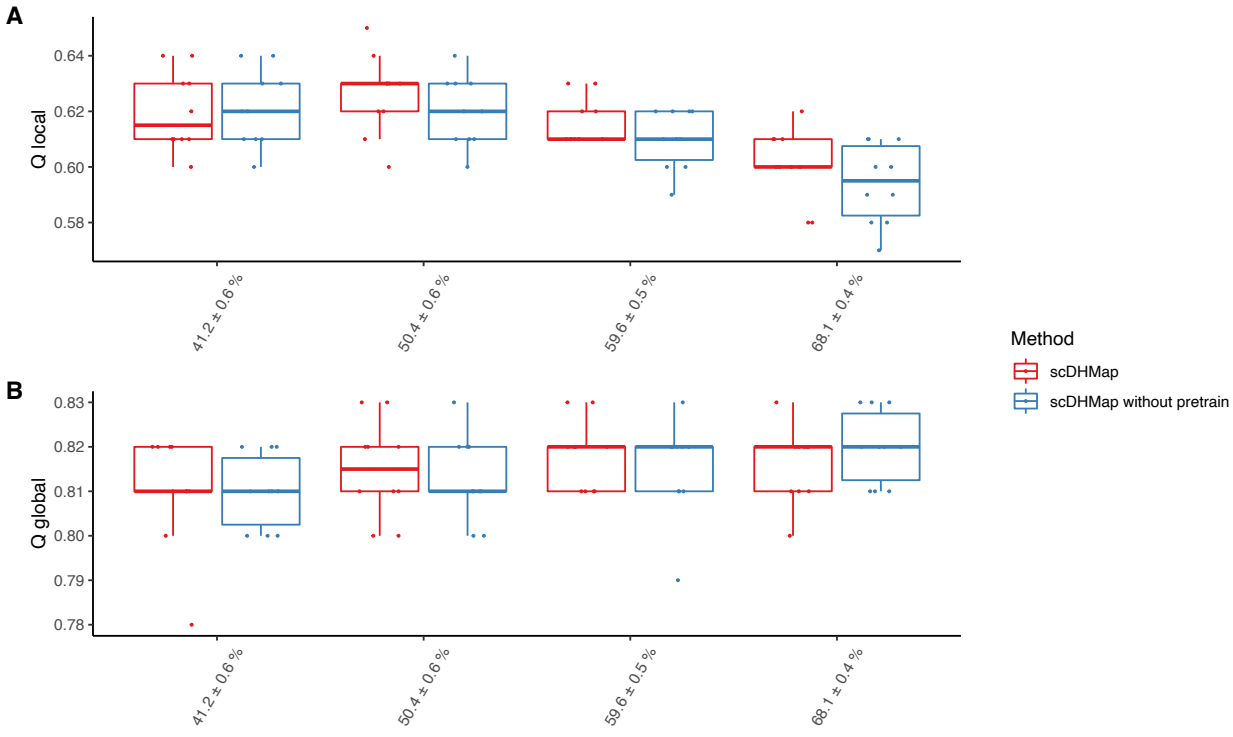
Supplementary Figure S3. Embedding quality metrics of different methods with 3 dimensions on simulated datasets with various dropout rates.



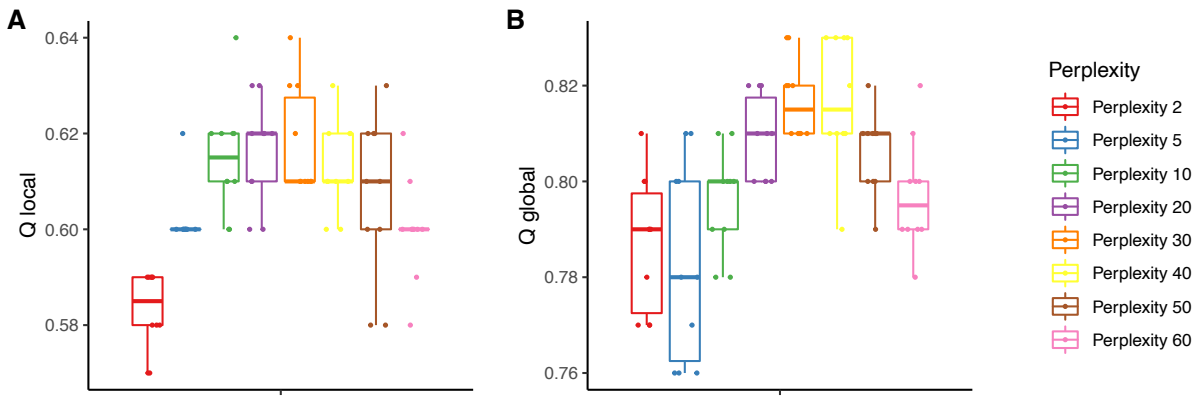
Supplementary Figure S4. (**A**-**B**) Embedding quality metrics of scDHMap with ZINB (zero-inflated negative binomial) loss, scDHMap with NB (negative binomial) loss, and scDHMap without the ZINB model-based decoder on simulated datasets with various dropout rates. (**C**) Scatter plots of Q locals of scDHMap with ZINB loss and NB loss on simulated datasets with various dropout rates. The diagonals are also displayed. Small jitters are added to avoid point overlapping.



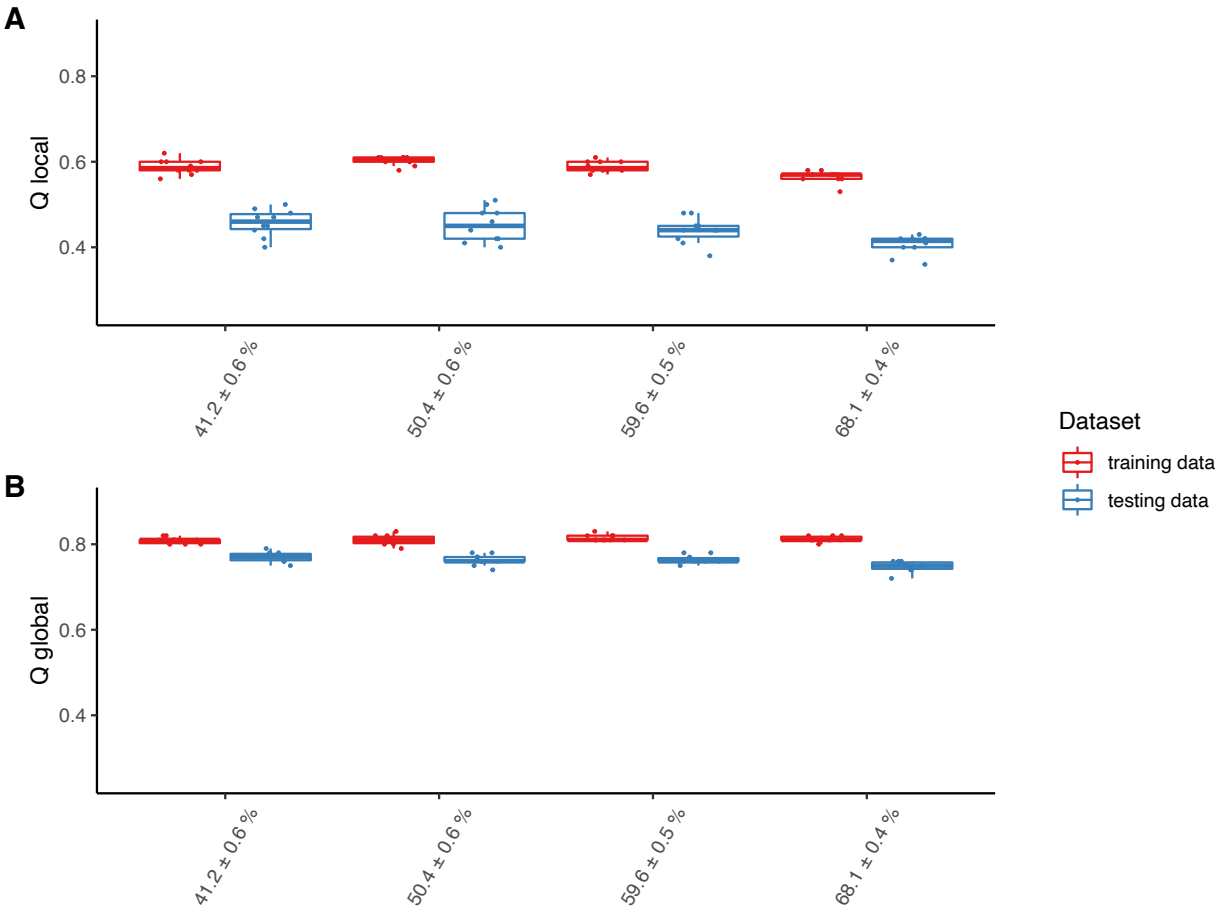
Supplementary Figure S5. Embedding quality metrics of scDHMap with different network architectures (4 hidden layers, 3 hidden layers, 2 hidden layers in encoder and decoder) on simulated datasets with various dropout rates.



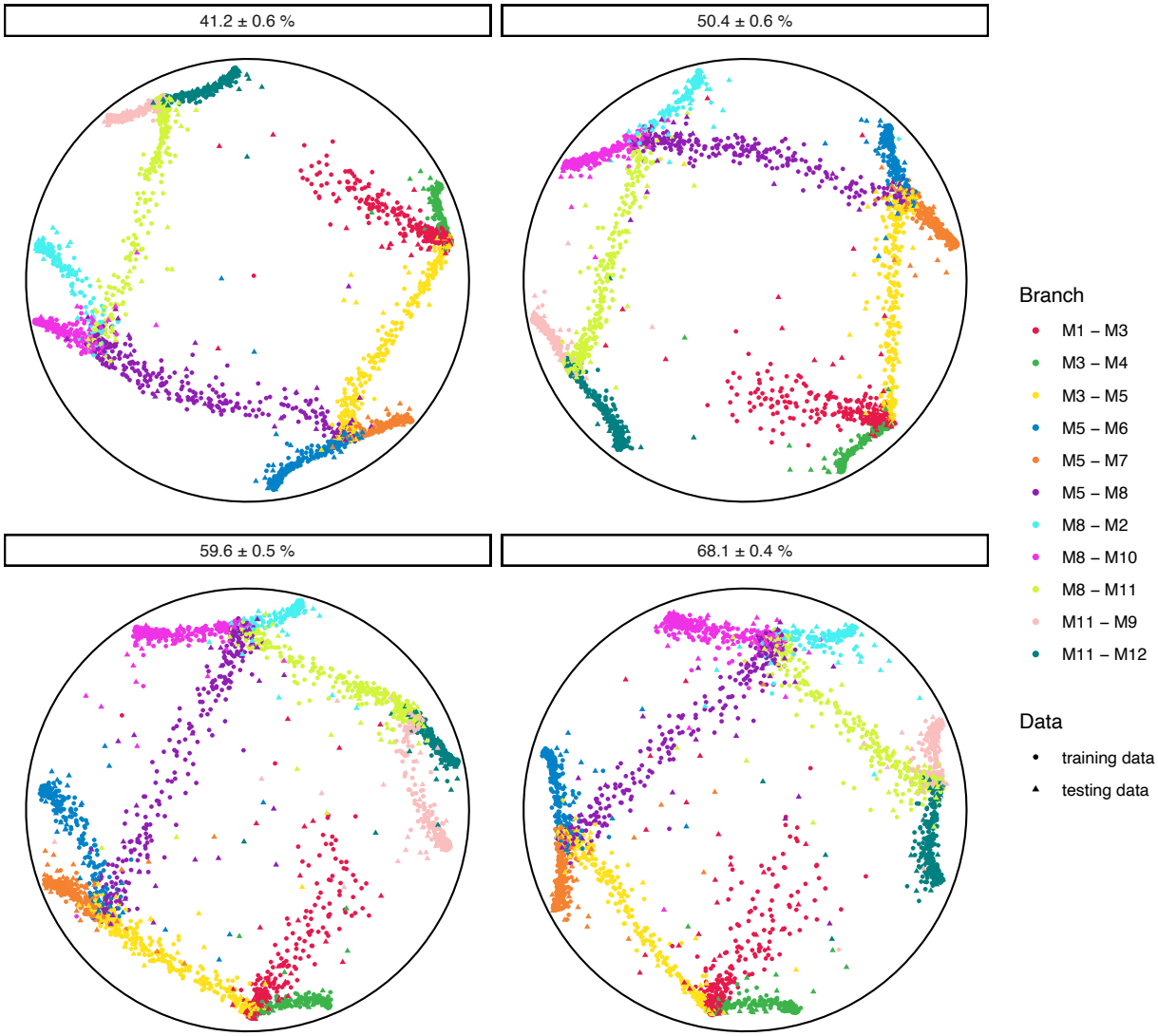
Supplementary Figure S6. Embedding quality metrics of scDHMap with and without pretrain on simulated datasets with various dropout rates.



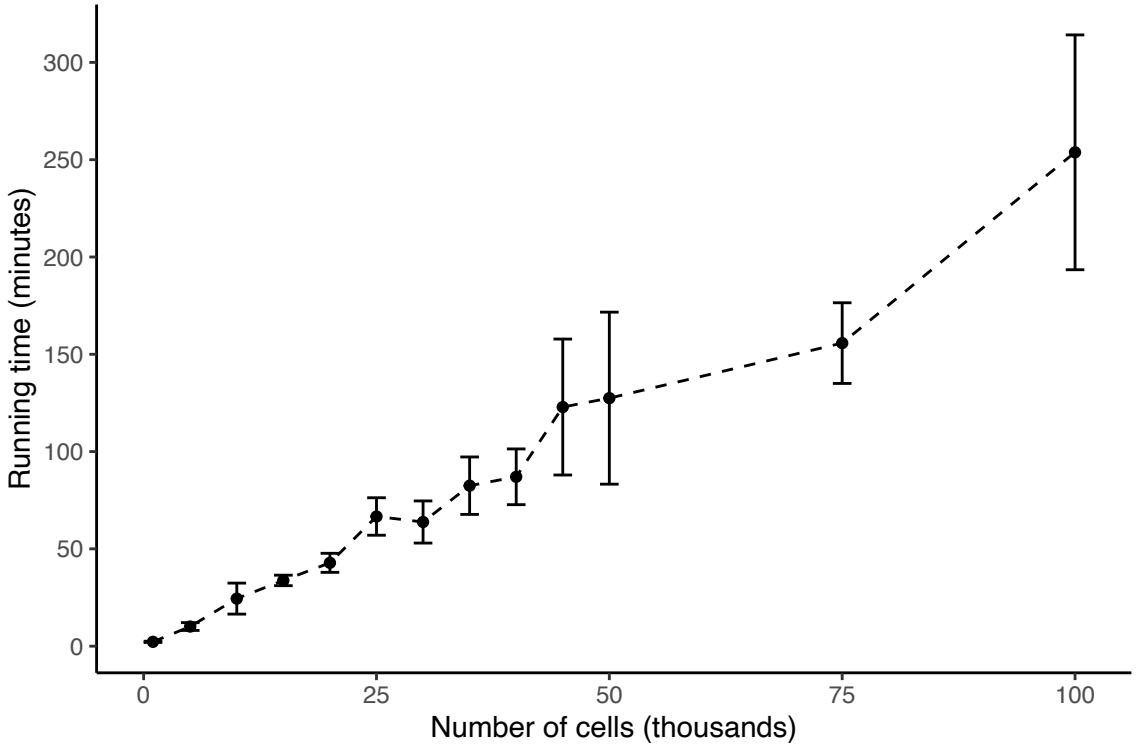
Supplementary Figure S7. Embedding quality metrics of scDHMap with different perplexities on simulated datasets of 59.6 ± 0.5% dropout rates.



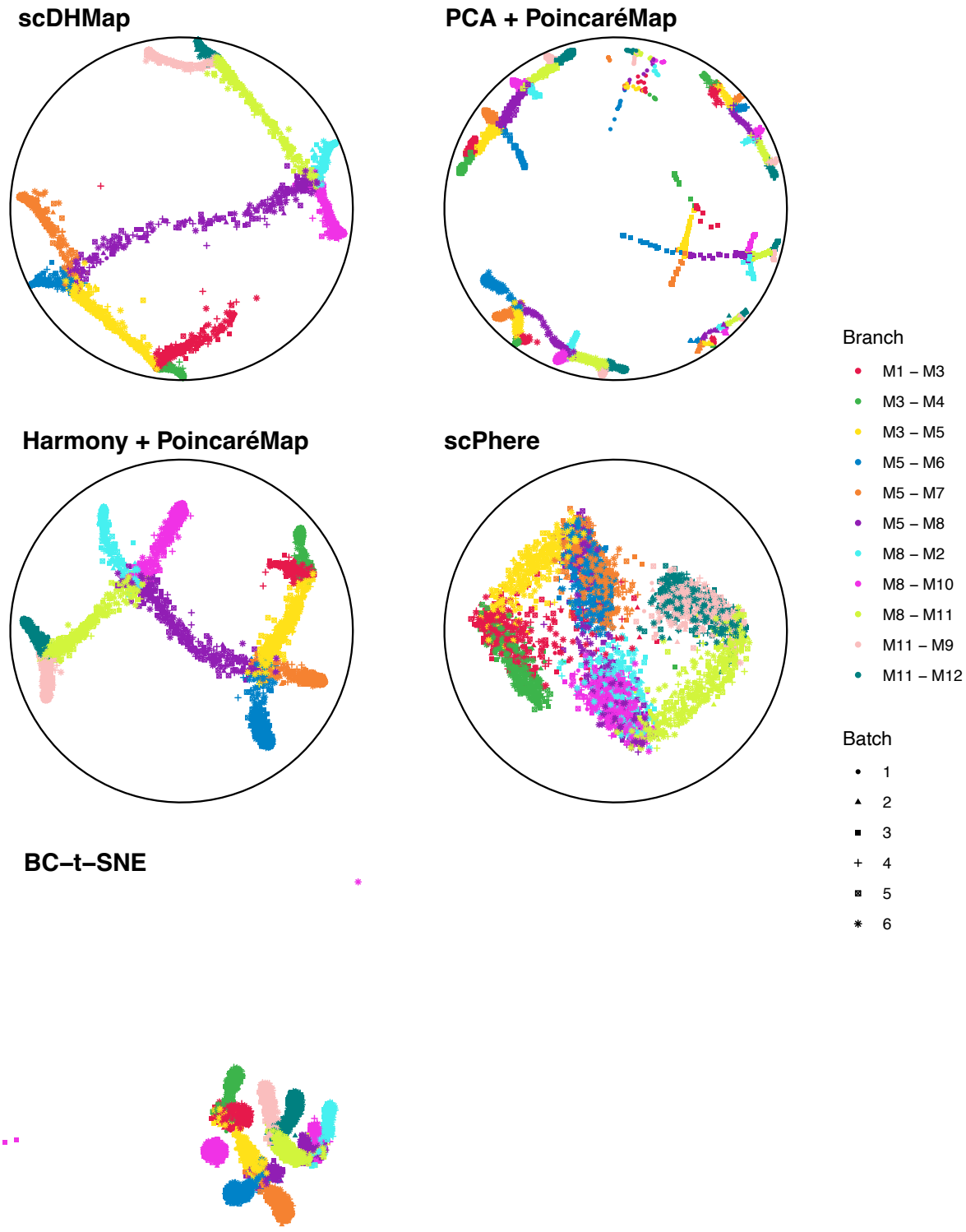
Supplementary Figure S8. Embedding quality metrics of scDHMap on training and testing sets of simulated datasets with various dropout rates. Each dataset is randomly divided into training and testing sets by the proportion of 90% and 10%. Embedding quality metrics are calculated on training and testing sets separately.



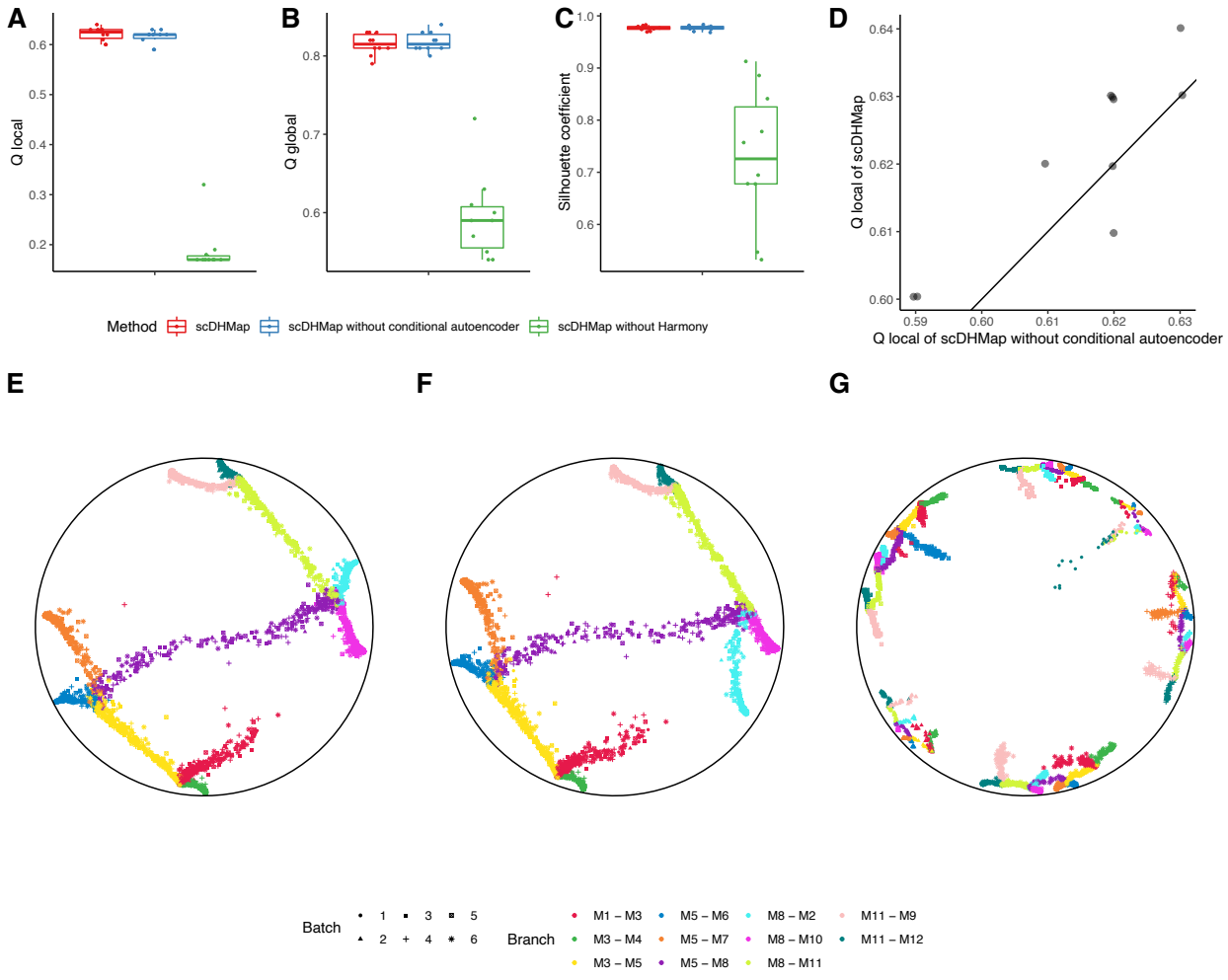
Supplementary Figure S9. 2-dimensional embeddings of scDHMap of training and testing simulated datasets with various dropout rates. Colors represent branches in the data, and dot shapes represent the training and testing datasets.



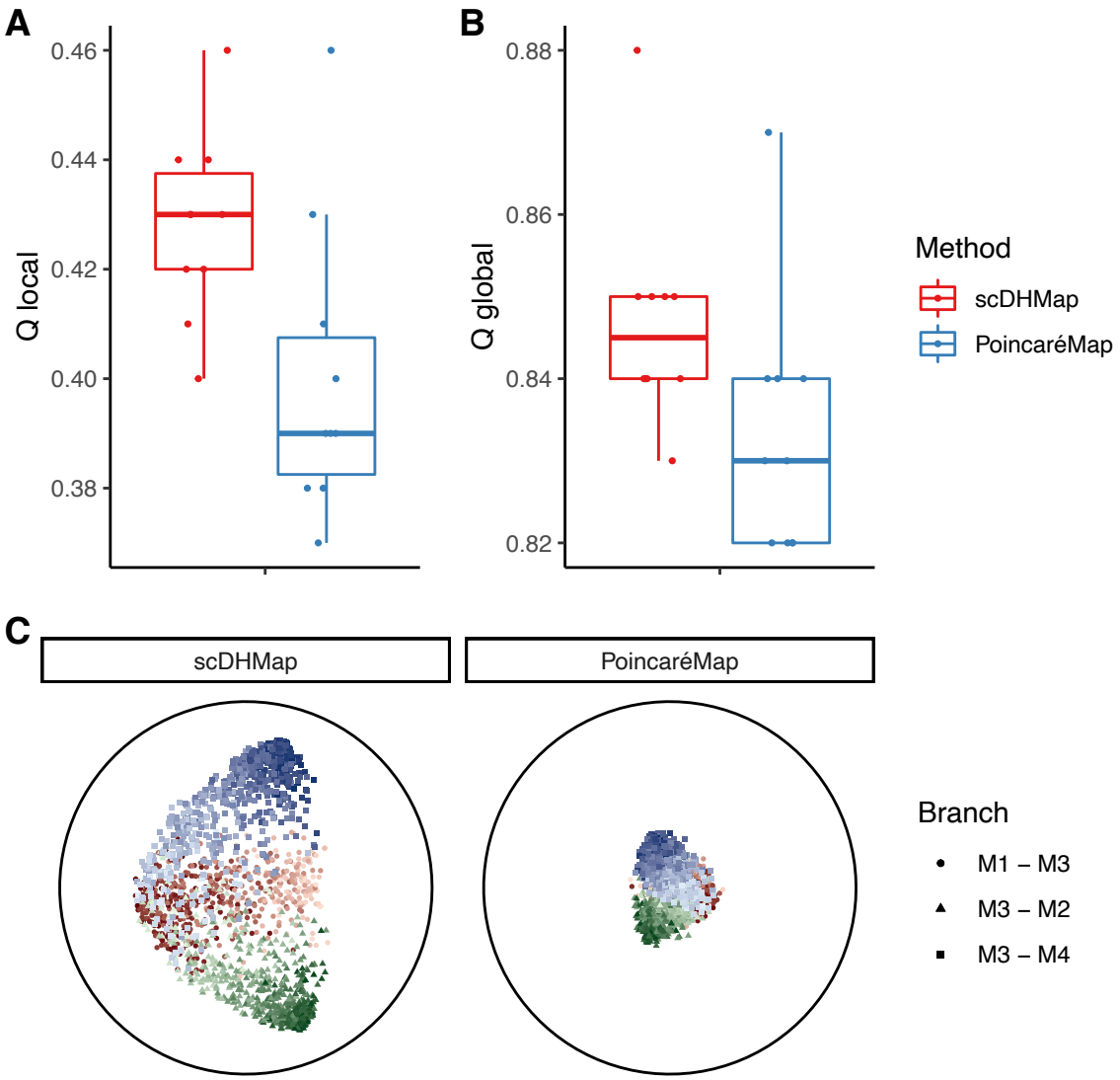
Supplementary Figure S10. Running time of scDHMap on simulated datasets with different numbers of cells. Each setting was repeated three times, average and standard deviation were plotted. Results were obtained on Intel(R) Core(TM) i9-10920X CPU @ 3.50GHz and NVIDIA(R) GeForce(TM) RTX 3090 GPU.



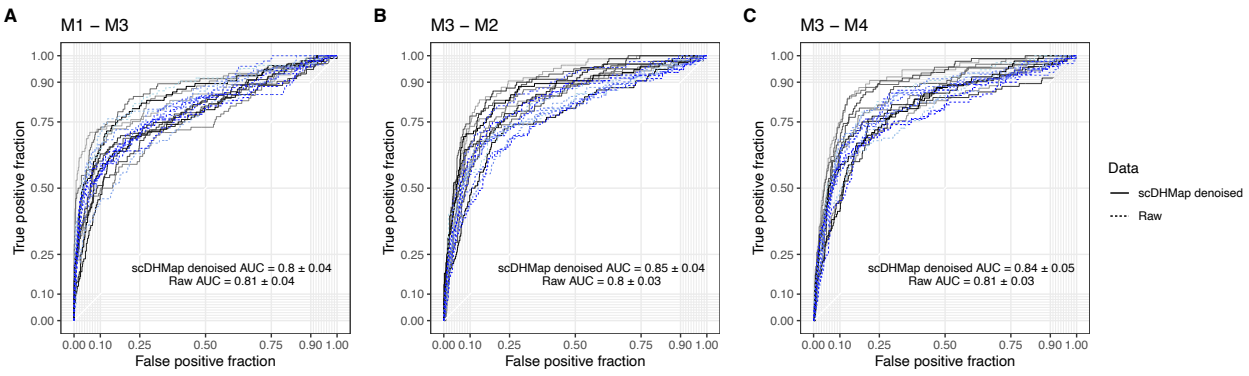
Supplementary Figure S11. 2-dimensional embeddings of different methods on simulated datasets with six different batches. Colors represent branches, and dot shapes represent batches.



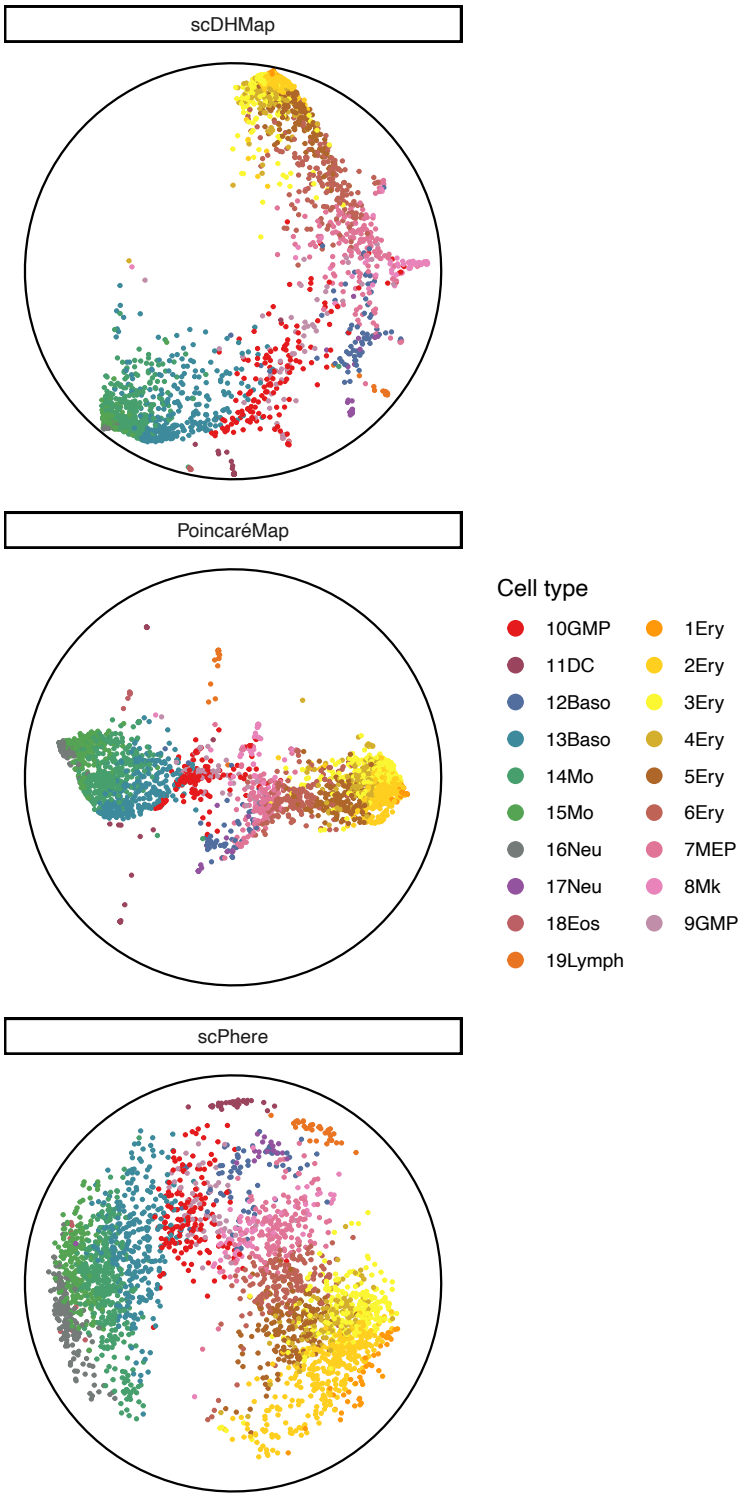
Supplementary Figure S12. Ablation study of batch alignment. (**A**-**B**) Embedding quality metrics of models on simulated datasets with six batches. Ten datasets were generated. (**C**) Silhouette coefficient (SIL) for quantifying the batch alignments. Larger values mean better alignments. (**D**) Scatter plot of Q locals of scDHMap and scDHMap without conditional autoencoder. The diagonal is also displayed. Small jitters are added to avoid point overlapping. (**E**) Embedding of scDHMap. (**F**) Embedding of scDHMap without conditional autoencoder. (**G**) Embedding of scDHMap without Harmony. Colors represent branches, and dot shapes represent batches (**E**-**G**).



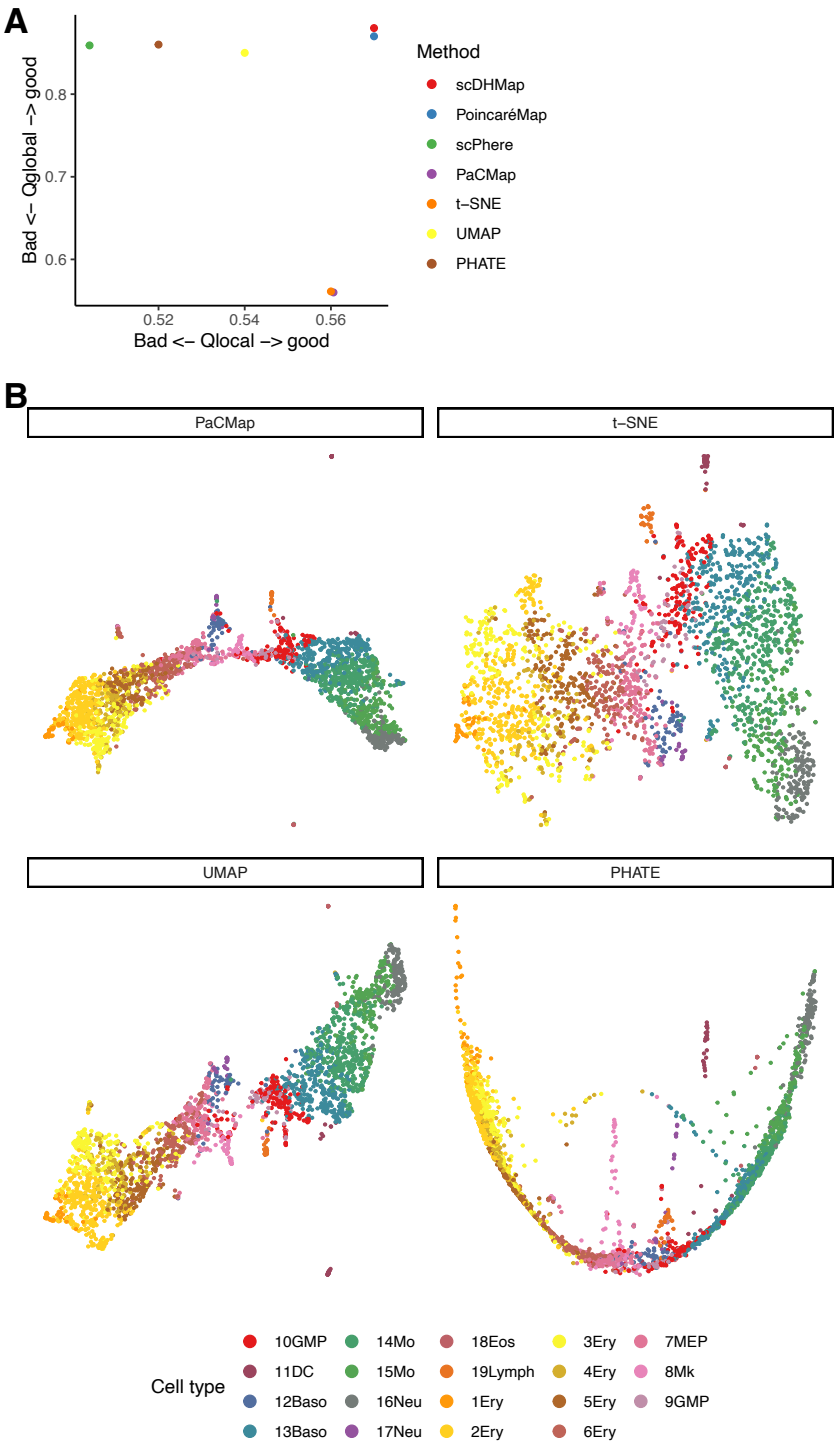
Supplementary Figure S13. Embedding qualities of scDHMap and PoincaréMap on simulated datasets with three trajectory branches (M1 – M3, M3 – M2, and M3 – M4), boxplot of (**A**) Q local scores and (**B**) Q global scores. (**C**) Plots of scDHMap and PoincaréMap embeddings, shapes represent branches, and colors from shallow to deep represent the trajectory steps (pseudotime) from start to end in each branch.



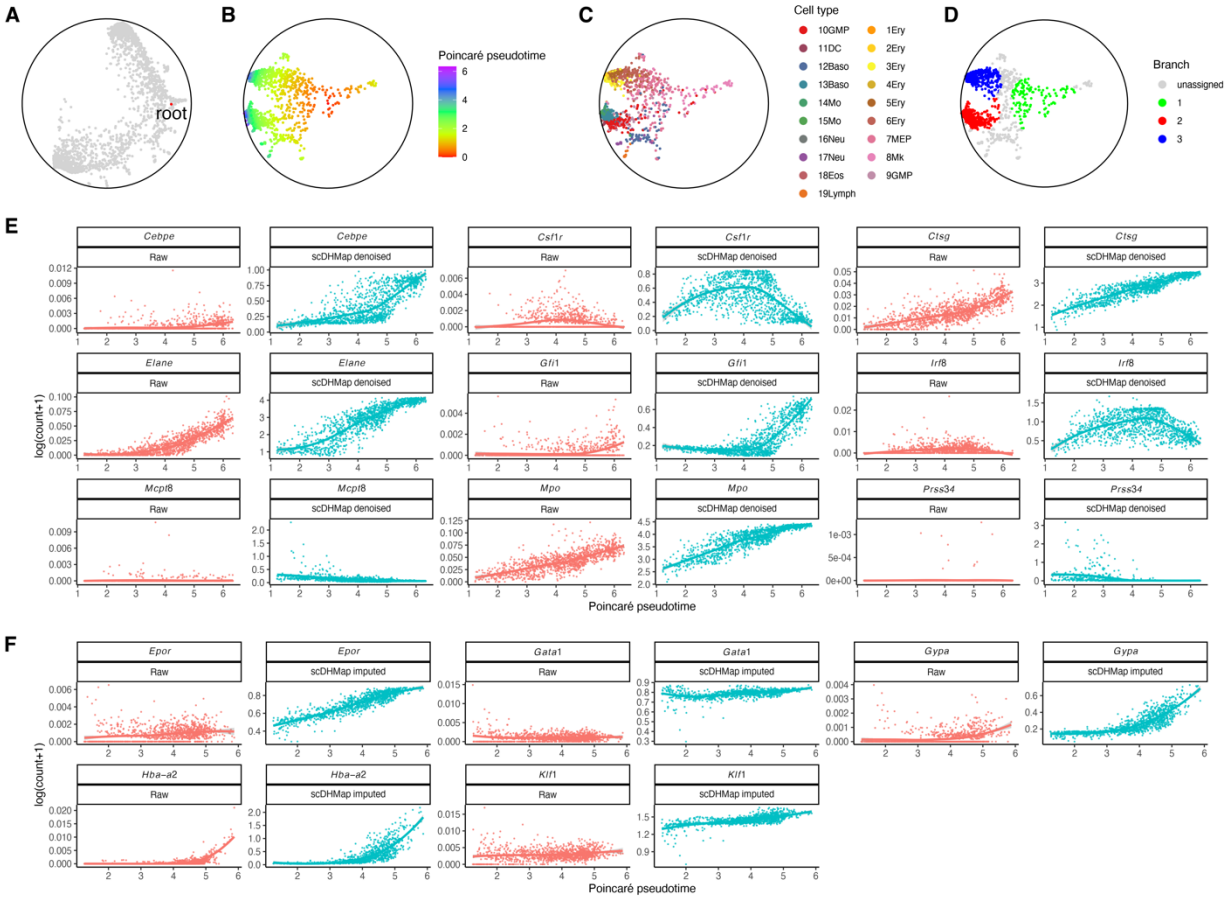
Supplementary Figure S14. AUC (area under the curve) plots of trajectory differential expression (DE) analysis on simulated datasets with three branches using scDHMap denoised and raw counts. True trajectory orders are used for the DE test.



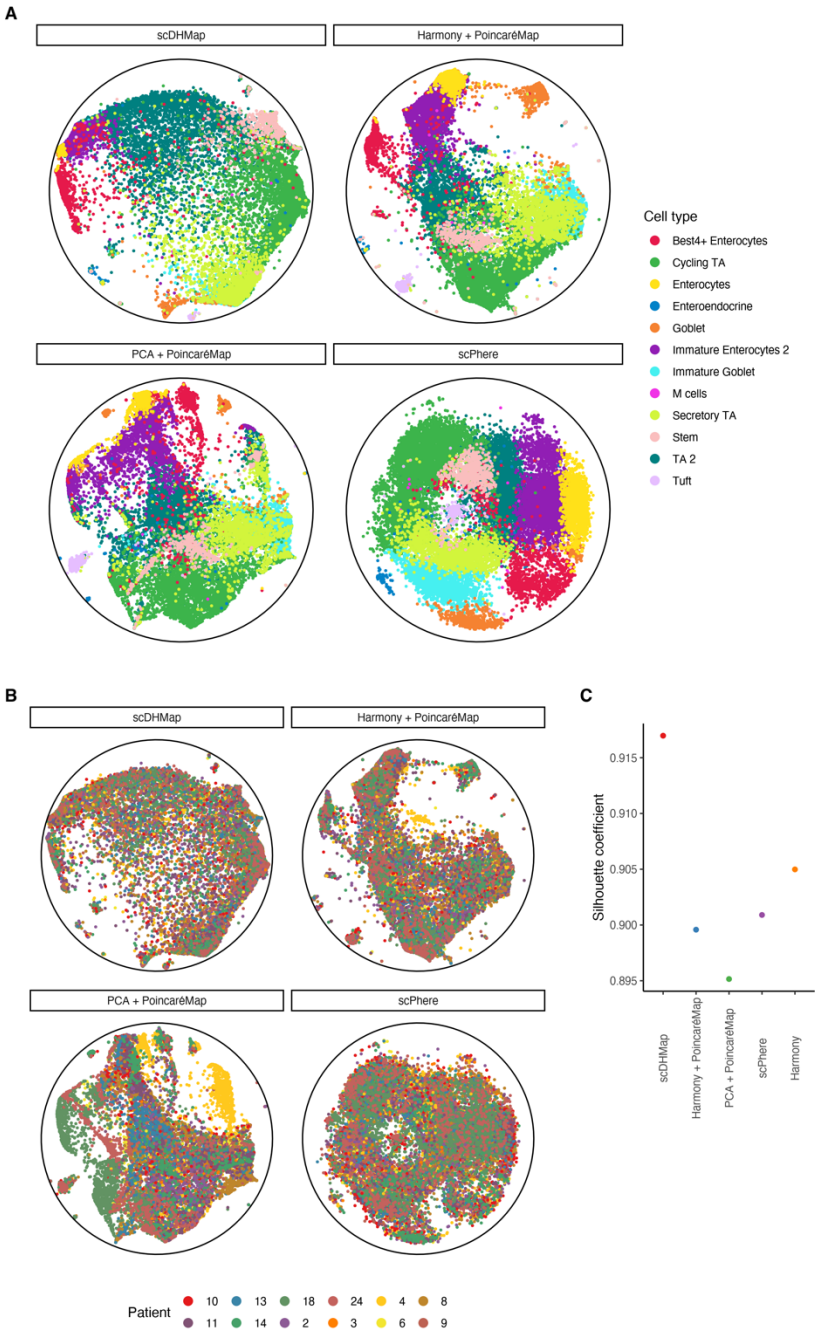
Supplementary Figure S15. Embeddings of scDHMap, PoincaréMap, and scPhere on Paul cells. Colors represent different cell types.



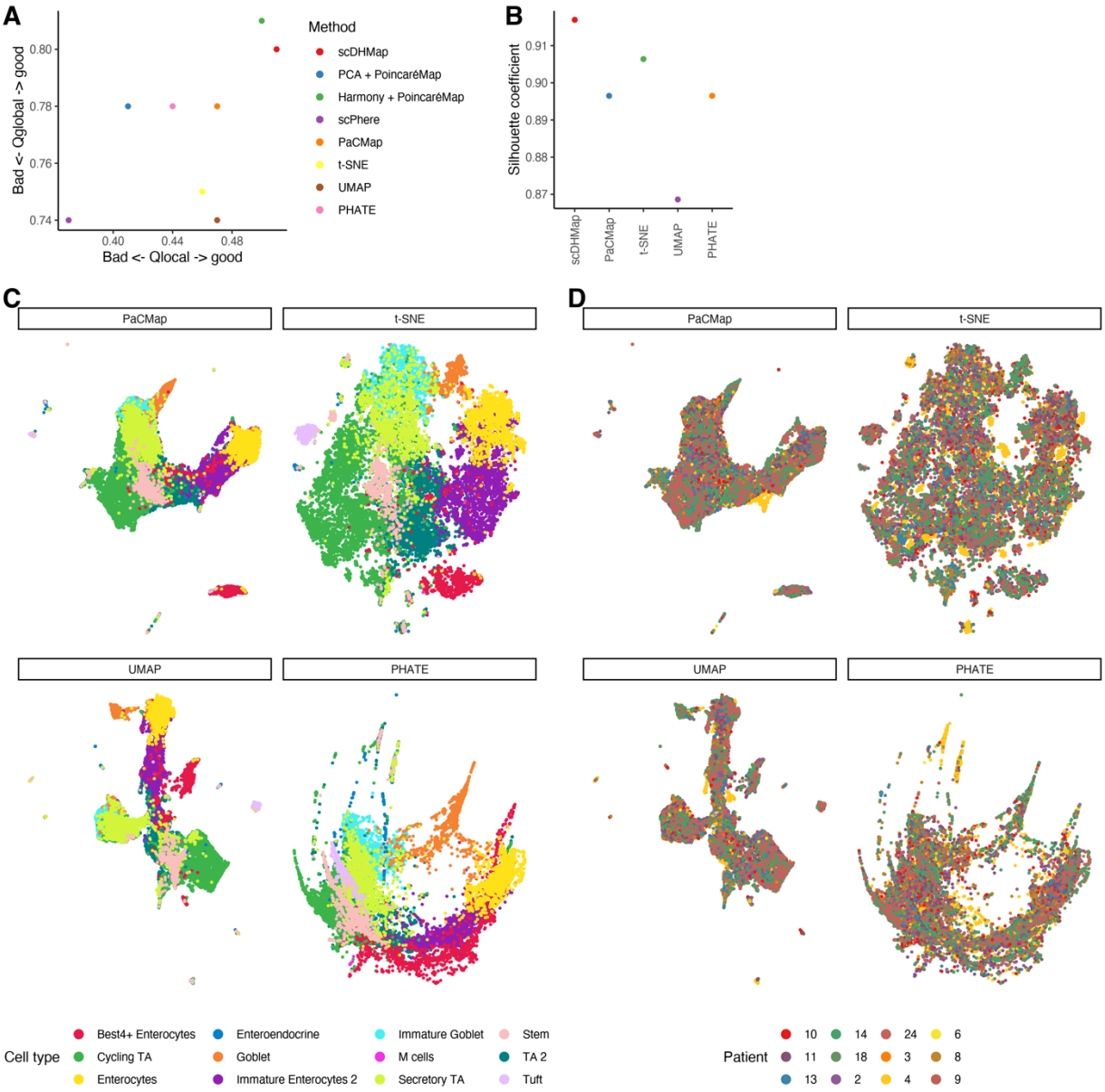
Supplementary Figure S16. (**A**) Embedding qualities of different embedding methods on Paul cells. (**B**) Embeddings of PaCMap, *t*-SNE, UMAP and scPhere of Paul cells. Colors represent different cell types.



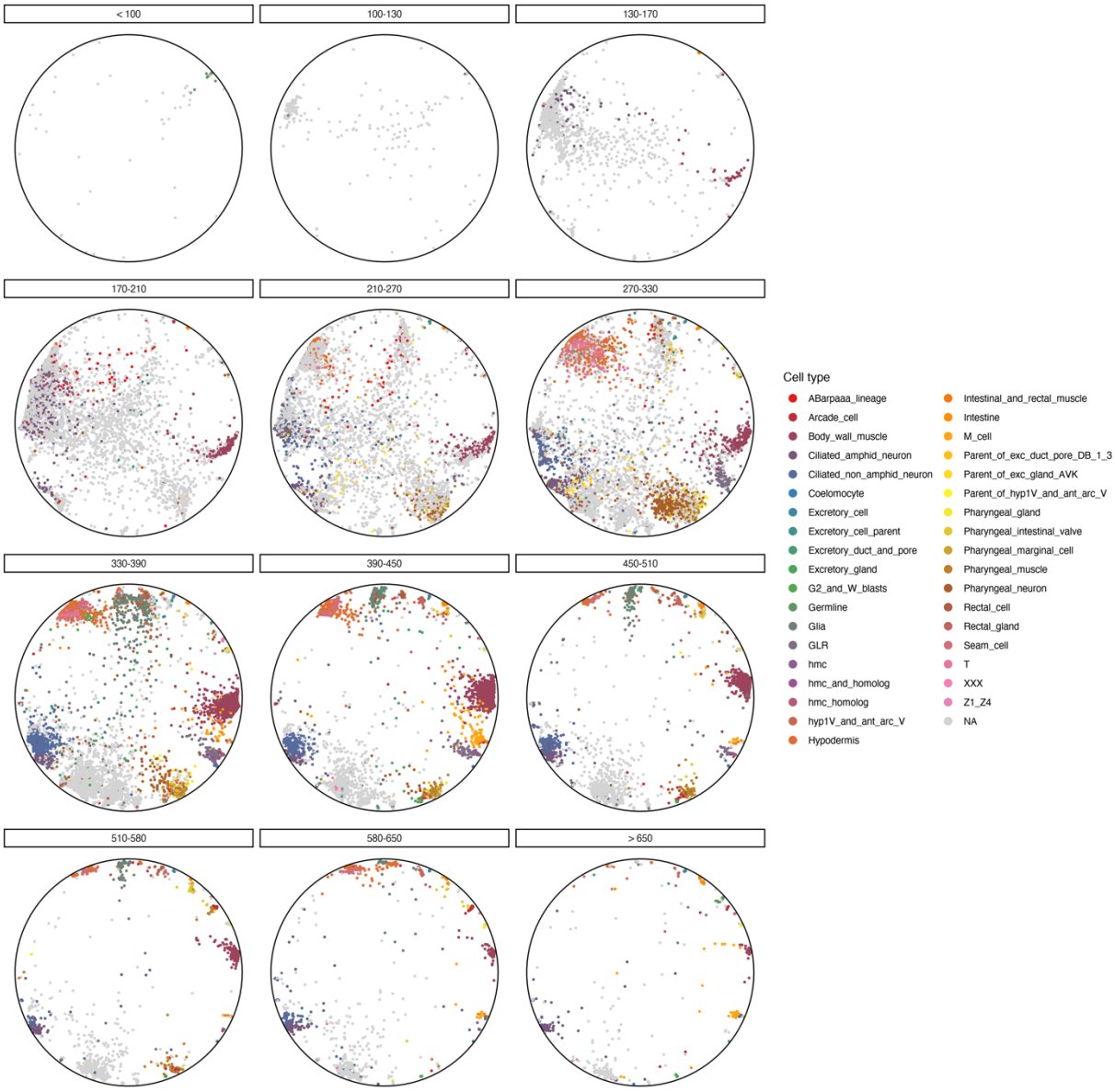
Supplementary Figure S17. (**A**) The position of the root cell in the scDHMap'e embedding of Paul cells. (**B**) Transform the Poincaré origin to the root cell, and Poincaré pseudotime is defined as the geodesic distance between each cell and the root cell. (**C**) Cell types in the transformed embedding. (**D**) The embedding is divided into two main branches and one short trunk based on the Poincaré pseudotime of scDHMap. (**E**) Marker genes against Poincaré pseudotime plotting in branch 2. (**F**) Marker genes against Poincaré pseudotime plotting in branch 3. Trend lines are smoothed by the LOESS regression (**E**, **F**).



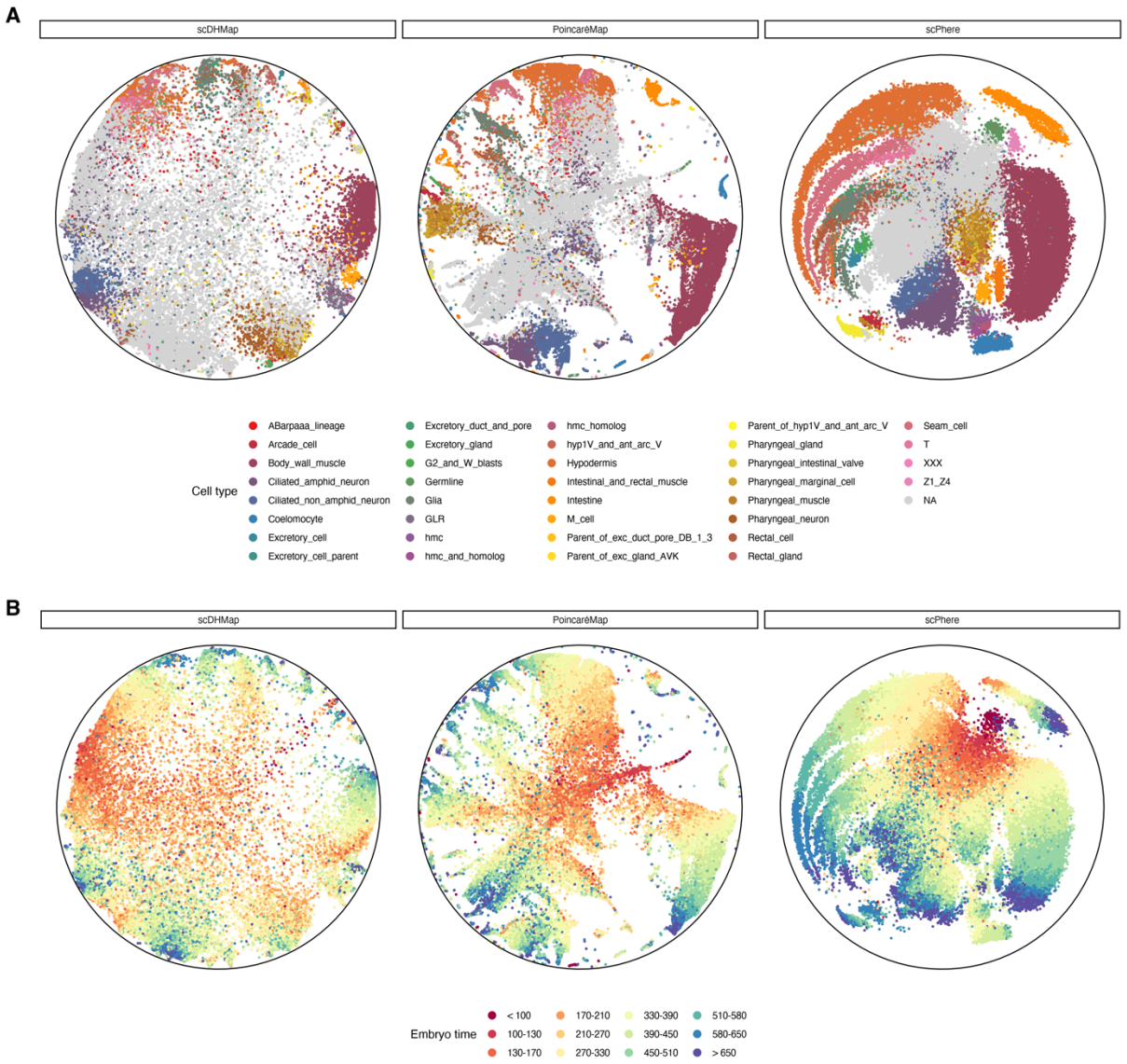
Supplementary Figure S18. (**A**) Embeddings of scDHMap, Harmony + PoincaréMap, PCA + PoincaréMap, and scPhere on the colon epithelial cells. Harmony + PoincaréMap uses Harmony corrected PCs as input, PCA + PoincaréMap uses PCs as input. Colors represent cell types. (**B**) Embeddings against different patient IDs. Colors represent patient IDs. (**C**) Silhouette coefficient of scDHMap, Harmony + PoincaréMap, PCA + PoincaréMap, scPhere, and Harmony for batch alignment.



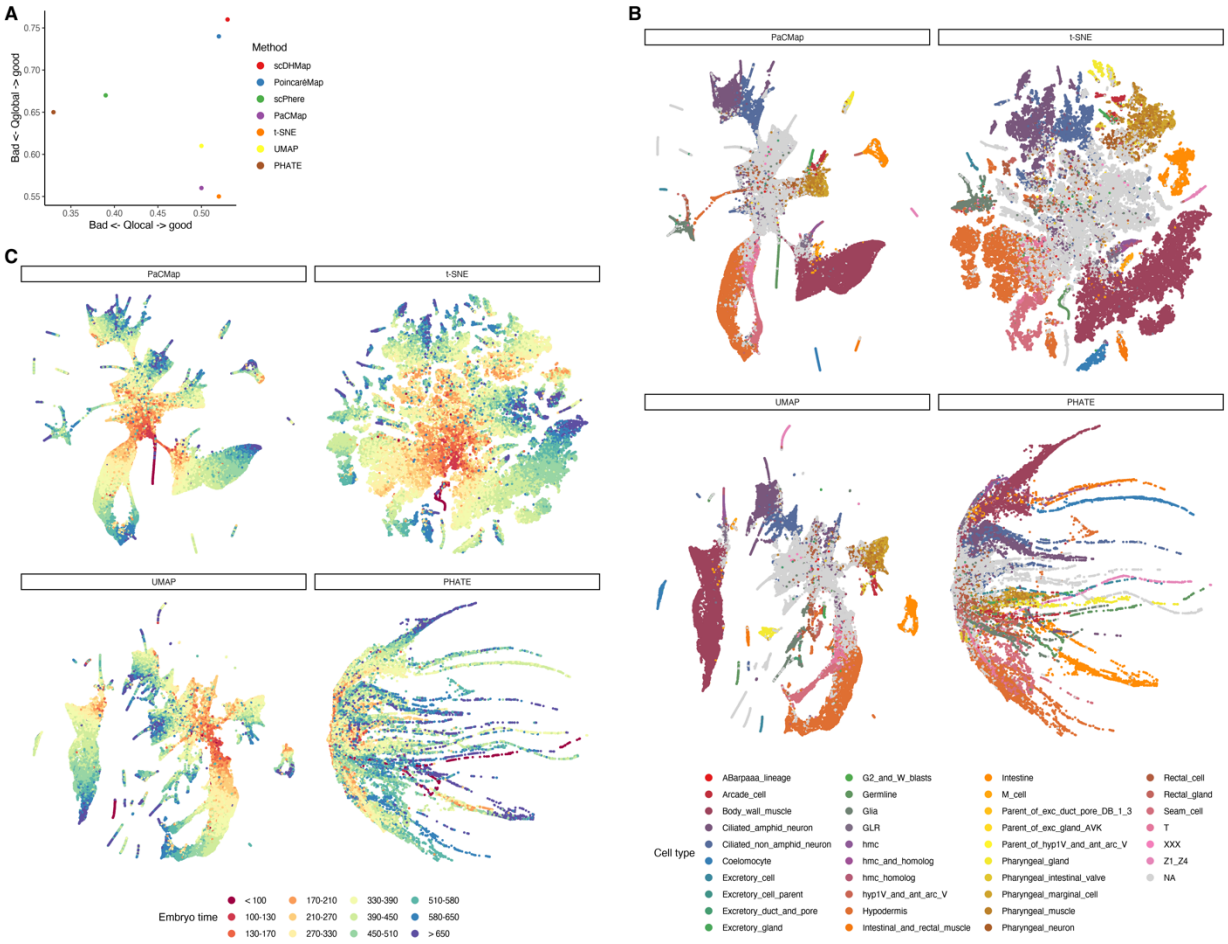
Supplementary Figure S19. (**A**) Embedding qualities of different methods on the colon epithelial cells. (**B**) Silhouette coefficient of scDHMap, PaCMap, *t*-SNE, UMAP, and PHATE for batch alignment. All methods used Harmony corrected 50 PCs as the input. (**C**) Embeddings of PaCMap, *t*-SNE, UMAP, and PHATE of the colon epithelial cells. Colors represent cell types. (**D**) Embeddings against different patient IDs. Colors represent patient IDs.



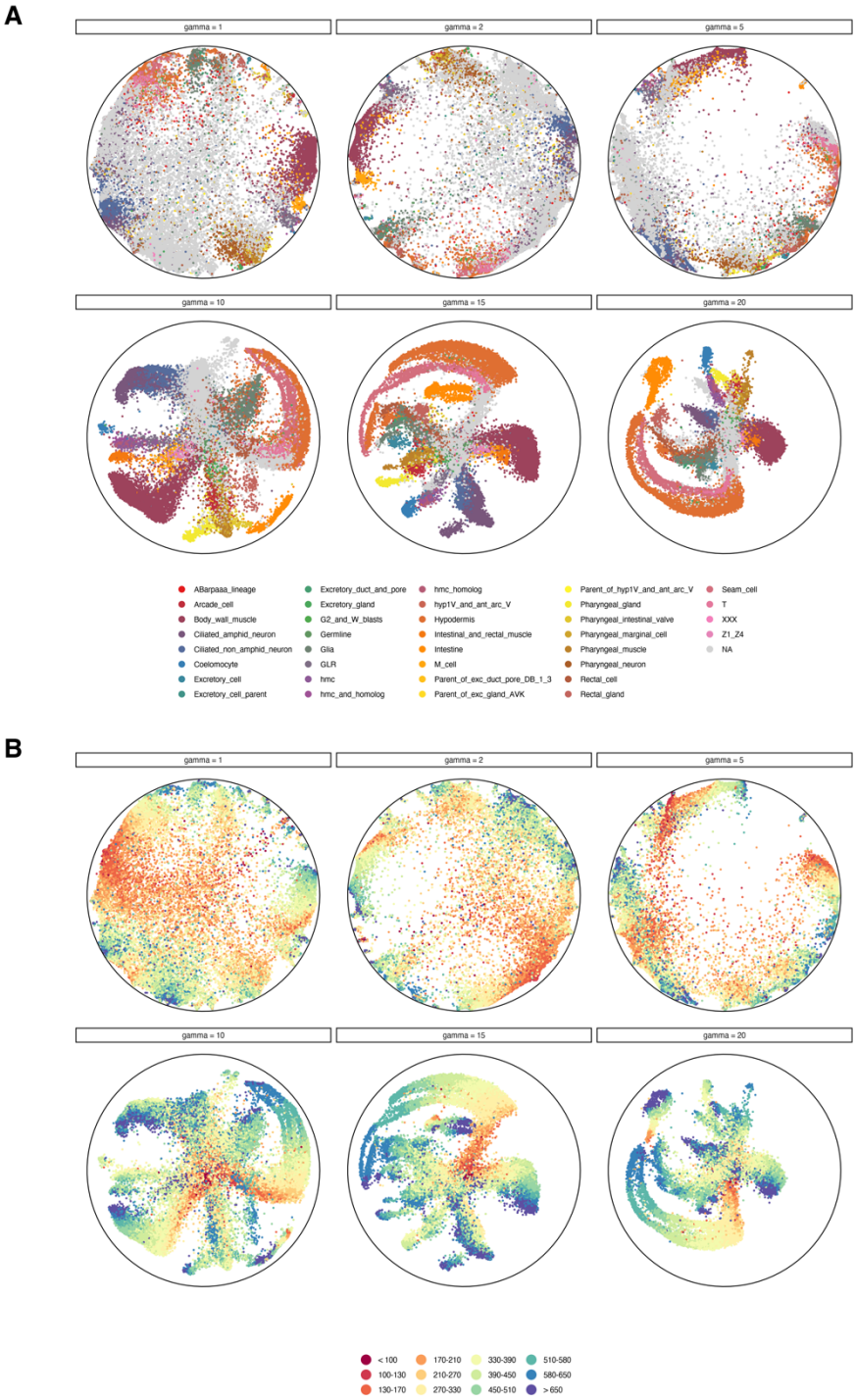
Supplementary Figure S20. scDHMap embedding of the *C. elegans* embryonic cells with each panel representing one embryonic time bin. Colors represent cell types.



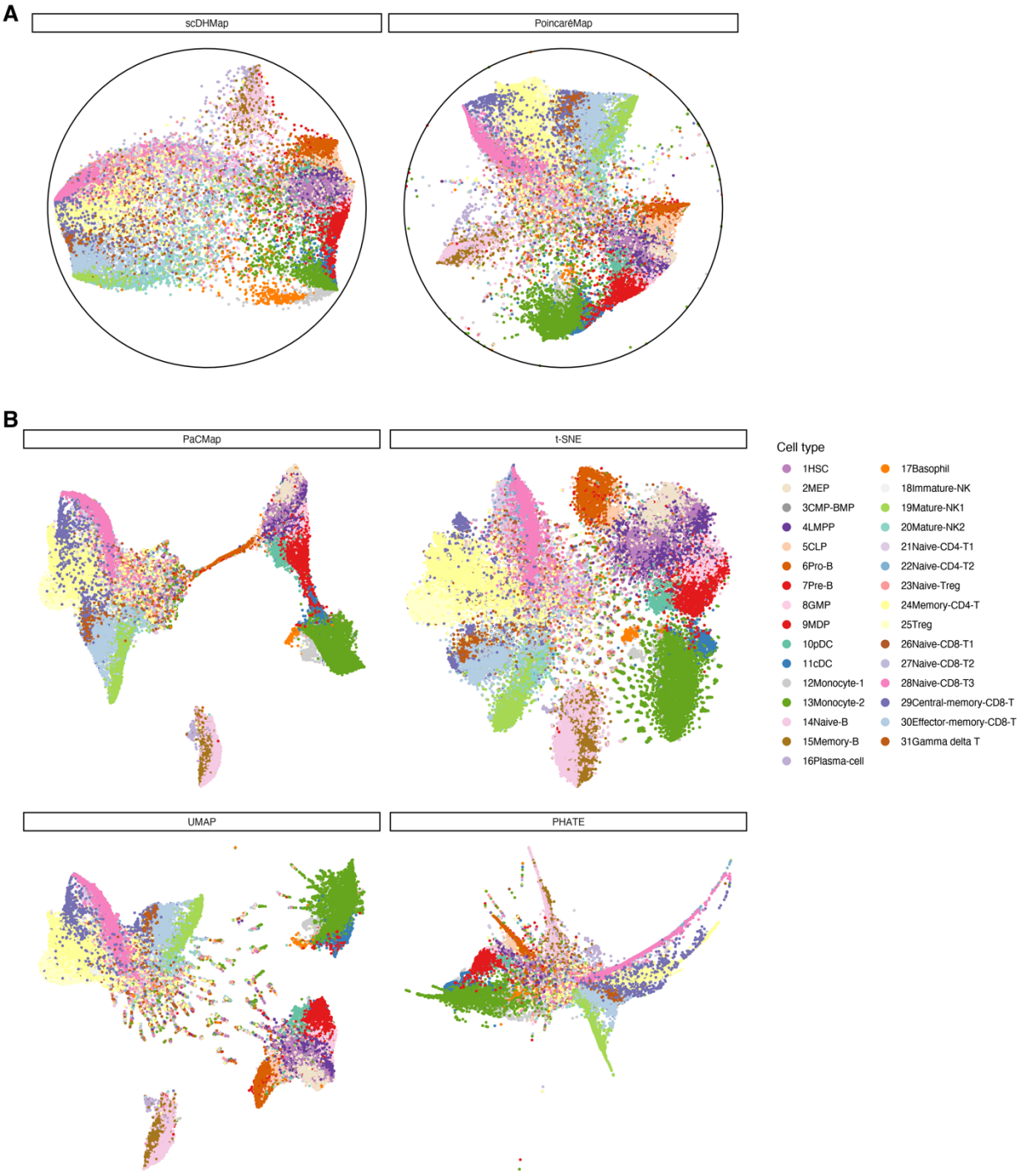
Supplementary Figure S21. Embeddings of scDHMap, PoincaréMap, and scPhere on the *C. elegans* embryonic cells. (**A**) Colors represent cell types. (**B**) Colors represent embryonic time.



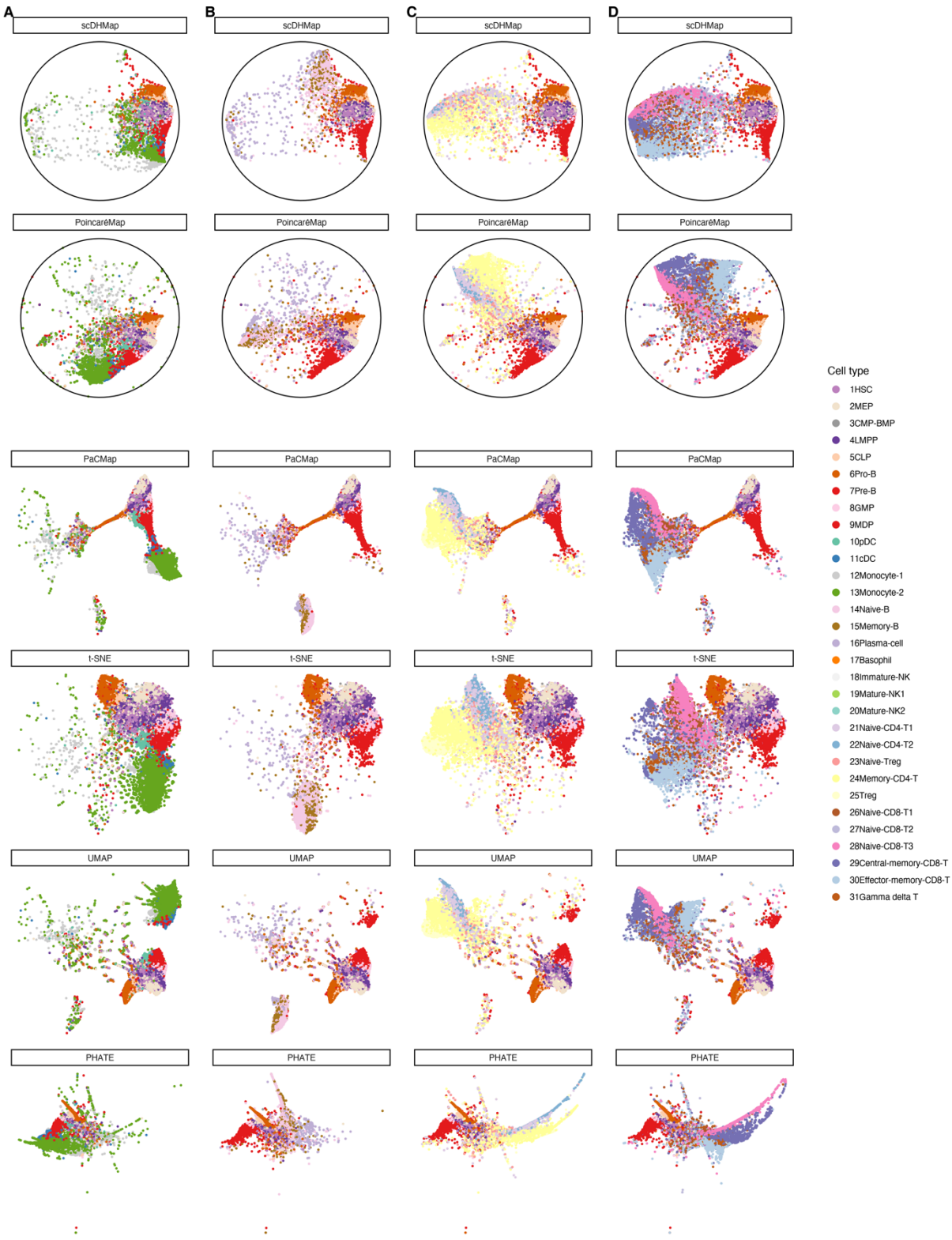
Supplementary Figure S22. (**A**) Embedding qualities of different methods on the *C. elegans* embryonic cells. (**B**, **C**) Embeddings of PaCMap, *t*-SNE, UMAP, and PHATE of the *C. elegans* embryonic cells. Colors represent cell types (**B**). Colors represent embryonic time (**C**).



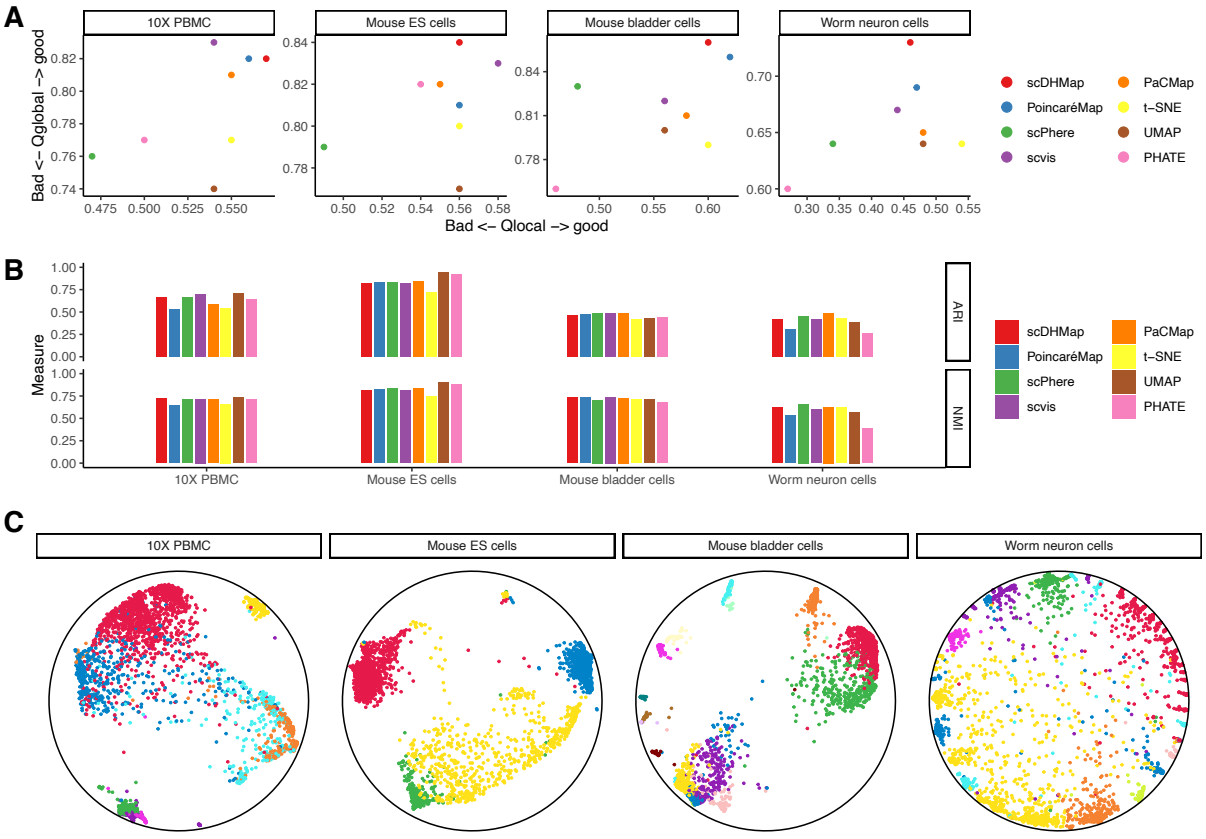
Supplementary Figure S23. Embeddings of scDHMap with different gamma values of the Cauchy kernel on the *C. elegans* embryonic cells. Colors represent cell types (**A**). Colors represent embryonic time (**B**).



Supplementary Figure S24. Embeddings of different methods on Satpathy's scATAC-seq data. Colors represent cell types.


A
B
C
D
scDHMap
PoincaréMap
PaCMap
t-SNE
UMAP
PHATE
Cell type
1HSC
2MEP
3CMP-BMP
4LMPP
5CLP
6Pro-B
7Pre-B
8GMP
9MDP
10pDC
11cDC
12Monocyte-1
13Monocyte-2
14Naive-B
15Memory-B
16Plasma-cell
17Basophil
18Immature-NK
19Mature-NK1
20Mature-NK2
21Naive-CD4-T1
22Naive-CD4-T2
23Naive-Treg
24Memory-CD4-T
25Treg
26Naive-CD8-T1
27Naive-CD8-T2
28Naive-CD8-T3
29Central-memory-CD8-T
30Effector-memory-CD8-T
31Gamma delta T

Supplementary Figure S25. Embeddings of specific cell types of different methods on Satpath's scATAC-seq data. Colors represent different cell types. (**A**) Progenitor cells (cluster 1 – 9) and myeloid cells (cluster 10 – 13). (**B**) Progenitor cells and B cells (cluster 14 – 16). (**C**) Progenitor cells and $CD4^+$ T cells (cluster 21 – 25). (**D**) Progenitor cells and $CD8^+$ T cells (cluster 26 – 31). HSC, hematopoietic stem cell; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte-erythroid progenitor; BMP, basophil-mast cell progenitor.



Supplementary Figure S26. Embedding of different methods on four real scRNA-seq data with diverse cell types. (**A**) Embedding quality metrics. (**B**) K-means clustering performance (we used hyperbolic k-means on hyperbolic embeddings) quantified by adjusted Rand index (ARI) and normalized mutual information (NMI). (**C**) Embedding of scDHMap, colors represent cell types.

Supplementary Table 1. Comparison of dimensionality reduction methods

Method	Low distortion of hierarchical structures	Local structural preservation	Global structural preservation	Efficient on large datasets	Batch correction
scDHMap	✓	✓	✓	✓	✓
PoincaréMap	✓	✓	✓		
scPhere	✓			✓	✓
scvis		✓	✓	✓	
PaCMap		✓	✓	✓	
***t*-SNE**		✓		✓	
UMAP		✓		✓	
PHATE		✓	✓	✓	
Diffusion Map		✓	✓	✓	

Supplementary Note 1. Summary of published embedding methods of single-cell genomics data

t-SNE (van der Maaten and Hinton 2008) and UMAP (McInnes et al. 2018) are two popular dimensionality reduction methods that have been widely used in single-cell analysis. The main idea of *t*-SNE and UMAP is to retain the distribution of pairwise distances in high dimensions when reducing to low-dimensional Euclidean space. These two methods are mainly focused on local structure preservation (e.g., the perplexity in *t*-SNE and the number of nearest neighbors in UMAP); therefore, they do not guarantee that the global hierarchical structure will be retained. Recently, a manifold learning method PaCMap has been proposed to improve the preservation of the global structure, which optimizes the low-dimensional embedding using a combining of local neighbor pairs, mid-near pairs, and further pairs (Wang et al. 2021). However, PaCMap is not designed for revealing complex hierarchical tree structures either. Diffusion map has been developed for defining differentiation trajectories, which uses kernel width for tackling uncertainties and dropout events (Haghverdi et al. 2015). However, diffusion map is not optimized for visualizing complex hierarchical trees; and more importantly, diffusion map typically encodes high dimensional structures in higher dimensions, which makes it not amenable to visualization. Graph abstractions (PAGA) provides a high-level graph representation of the scRNA-seq data, in which nodes represent partitions or clusters and edges represent the relationships between these nodes (Wolf et al. 2019). PAGA does not provide an embedding of cellular resolution. PHATE is a dimensionality reduction framework that combines both local similarities by using kernel function and global relationships by using diffusion map and generates the low-dimensional visualization via metric MDS (Moon et al. 2019). PHATE has been illustrated to recover hierarchical structures but is also affected by the distortion in the low-dimensional Euclidean space. Monocle applies reversed graph embedding to reduce the data to a tree-like topology in a low-dimensional Euclidean space (Qiu et al. 2017), but also undermines the visualization of complex trees. Meanwhile, scVI (Lopez et al. 2018) and SAUCIE (Amodio et al. 2019) are deep learning approaches that use autoencoders to reduce dimensionalities. Autoencoders can identify major structural patterns such as clusters in the single-cell data(Tian et al. 2019), but there is no guarantee for preserving the pairwise high dimensional similarities. Scvis combines the autoencoder with the *t*-SNE regularization to preserve the pairwise similarities, making the model performs dimensionality reduction and structure preservation simultaneously (Ding et al. 2018). But scvis is not designed for visualizing the complex hierarchy either due to the distortion in the low-dimensional Euclidean space.

Supplementary Note 2. Embedding quality criteria

Lee et al. (Lee and Verleysen 2010) proposed a scale-independent quality criterion for quantifying the embedding qualities. Let $X = \{x_i\}_{i=1}^{N}$ be a high-dimensional dataset of *N* samples and $Y = \{y_i\}_{i=1}^{N}$ be a 2-dimensional representation of this dataset. Let δ_{ij} denotes the distance from x_i to x_j in the high-dimensional space and d_{ij} denotes the distance from y_i to y_j in the low-dimensional space, and $\delta_{ij} = \delta_{ji}$ and $d_{ij} = d_{ji}$. The high ($R = \{\rho_{ij}\}_{1\le i,j\le N}$) and low-dimensional ranks ($V = \{r_{ij}\}_{1\le i,j\le N}$) between the points can be calculated by the distances:

$$\rho_{ij} = \left|\left\{k: \delta_{ik} < \delta_{ij} \text{ or } \left(\delta_{ik} = \delta_{ij} \text{ and } 1 \le k < j \le N\right)\right\}\right|$$
$$r_{ij} = \left|\left\{k: d_{ik} < d_{ij} \text{ or } \left((d_{ik} = d_{ij} \text{ and } 1 \le k < j \le N\right)\right\}\right|$$

where $|\cdot|$ denotes the cardinality of a set. According to this definition, reflexive ranks are set to zero, and non-reflexive ranks are $\{K = 1, \dots, N-1\}$.

A co-ranking matrix $\boldsymbol{Q} = \{q_{kl}\}_{1\le k,l\le N-1}$ is defined as

$$q_{ij} = \left|\left\{(i,j): \rho_{ij} = k \text{ and } r_{ij} = l\right\}\right|$$

The co-ranking matrix stores the information about how ranks of distances are preserved in a given low-dimensional embedding. The co-ranking matrix could be computed straightforwardly(Lee and Verleysen 2010) and used to compute the scale-independent quality criteria Q_{NX} for the dimensionality reduction for a given value of $K = 1, \dots, N-1$

$$Q_{NX}(K) = \frac{1}{KN} \sum_{(k,l)\in \mathbb{UL}_{\mathbb{K}}} q_{kl}$$

where $\mathbb{UL}_{\mathbb{K}} = \{1, \dots, K\} \times \{1, \dots, K\}$ is the upper left corner of the co-ranking matrix. $Q_{NX}(K) \in [0,1]$ accesses the overall embedding quality. Essentially, it measures the preservation of K neighborhoods. A perfect embedding has $Q_{NX}(K) = 1$ for every $K = 1, \dots, N-1$.

The left part of the $Q_{NX}(K)$ curve reflects the local preservation quality and the right part of $Q_{NX}(K)$ curve reflects the global preservation quality. To improve the readability, it can be divided into two scalar quality criteria Q local and Q global focusing separately on the local and global preservation of the embedding

$$Q_{local} = \frac{1}{K_{max}} \sum_{K=1}^{K_{max}} Q_{NX}(K)$$

and

$$Q_{global} = \frac{1}{N - K_{max}} \sum_{K=K_{max}}^{N-1} Q_{NX}(K)$$

where K_{max} defines the split of the Q_{NX} curve and is automatically computed as

$$K_{max} = \arg\max_{K}\left(Q_{NX}(K) - \frac{K}{N-1}\right)$$

The quantities of Q local and Q global range from 0 to 1 represent bad to good.

Supplementary Note 3. Trajectory differential expression test

Raw counts and scDHMap denoised counts are used in tradeSeq (Van den Berge et al. 2020) for trajectory differential expression (DE) test. The estimated mean parameters of the ZINB decoder of scDHMap are used as denoised counts. We round the mean parameters to integers. The parameter for tradeSeq is nknots = 4.

Supplementary Note 4. Branch assignment of Paul cells by the Poincaré pseudotime

The origin of the Poincaré embedding of scDHMap on Paul cells is first translated to the predefined root. The Poincaré pseudotime is defined as the Poincaré distance between a point to the root. Next, we use the Poincaré pseudotime (pt) to divide the embedding into three branches (Haghverdi et al. 2016).

First, we pick the root cell r_1, and we identify a tip r_2 which maximizes the Poincaré distance to r_1. We then identify the second tip r_3 that maximizes the $\mathrm{pt}(r_2, r_3)$, which means the Poincaré distance between r_2 and r_3. In brief:

$$r_2 = \arg\max_x \mathrm{pt}(r_1, x)$$
$$r_3 = \arg\max_x \mathrm{pt}(r_2, x)$$

Now we can perform a pseudotime ordering which orders cells according to the Poincaré distance to the tip of a branch (either r_1, r_2 and r_3). The ordering on every two branches will correlate with each other only on the third branch and anti-correlate on the two branches themselves. We will use this property to find a cutoff x for each branch. To separate branch 1, we first do three independent ordering $O1$, $O2$, $O3$ with assigning r_1, r_2 and r_3 as the root of ordering correspondingly. Then, Kendall-tau correlations are used to build a measure of concordance between the O_2 and O_3 ordering from s_1 until x and their anti-concordance for the rest of cells:

$$K_{2,3}(x) = \mathrm{Kendal.tau}\big(O2(r_1\colon x), O3(r_1\colon x)\big)$$
$$-\mathrm{Kendal.tau}(O2(x+1\colon end), O3(x+1\colon end))$$

Finally, the cutoff x of branch 1 is found by

$$x_{O1} = \arg\max_x (K_{2,3}(x) - K_{2,3}(x-1))$$

Other branches follow the similar procedure. Finally, there usually remain sets of cells cannot be assigned to a single branch. We denote these cells as unassigned.

Supplementary Note 5. Description of scRNA-seq datasets with different cell types

10X PBMC dataset (4K PBMCs from a healthy donor) is provided by 10X scRNA-seq platform (Zheng et al. 2017), which profiles the transcriptome of the peripheral blood mononuclear cells (PBMCs) from a healthy donor. The total number of cells is about 4000. PBMC 4k data is downloaded from the website of 10X genomics (https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.1.0/pbmc4k). We download filtered gene/cell matrix and cell labels identified by graph-based clustering (method description: https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/output/analysis).

Mouse embryonic stem cells dataset (Klein et al. 2015) is downloaded from GSE65525 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65525). The transcriptomes are profiled by the droplet-microfluidic approach for parallel barcoding. Allon M.K. *et.al* analyzed the heterogenous onset of differentiation of mouse embryonic stem cells after leukemia inhibitory factor (LIF) withdrawal. We download the read counts matrices of mouse ES cells sample 1, mouse ES cells LIF- 2days, mouse ES cells LIF- 4days and mouse ES cells LIF- 7days, and put all cells together. The labels of cells are defined as the intervals after LIF withdrawal.

Mouse bladder cells dataset of the Mouse Cell Atlas project (Han et al. 2018) is provided by the authors (https://figshare.com/s/865e694ad06d5857db4b). We download the batch gene removed raw digital expression matrix of all 400,000 single cells sorted by tissues and the table of cell assignments. The authors identified the cell types and described the method in (Han et al. 2018). From the raw count matrix, we select the cells from bladder tissue.

Worm neuron cells dataset is profiled by sci-RNA-seq (single-cell combinatorial indexing RNA sequencing) (Cao et al. 2017). The authors profiled about 50,000 cells from the nematode *Caenorhabditis elegans* at the L2 larval stage and identified the cell types (http://atlas.gs.washington.edu/worm-rna/docs/). We select the subset of the neural cells and removed the cells with the label of “Unclassified neurons”. As a result, we get 4186 neural cells.

Supplementary Note 6. Clustering evaluation metrics

Clustering performance is evaluated by two metrics (**Supplementary Figure S26**): adjusted Rand index(Rand 1971) (ARI) and normalized mutual information(Strehl and Ghosh 2003) (NMI), which compare the concordance between predicted labels and the ground-truth labels. The two metrics are implemented by the scikit-learn package(Pedregosa et al. 2011).

Rand index(Rand 1971) is a simple measure of agreement between two cluster assignments U and V. The Adjust Rand Index (ARI) corrects for the lack of a constant value of the Rand index when the cluster assignments are selected randomly(Hubert and Arabie 1985). The ARI is calculated by four quantities. Specifically, we define a, the number of pairs of two objects in the same group in both U and V; b, the number of pairs of two objects in different groups in both U and V; c, the number of pairs of two objects in the same group in U but in different groups in V; and d, the number of pairs of two objects in different groups in U but in the same group in V. The ARI is formally defined as

$$\mathrm{ARI} = \frac{\binom{n}{2}(a+d) - [(a+b)(a+c)+(c+d)(b+d)]}{\binom{n}{2}^2 - [(a+b)(a+c)+(c+d)(b+d)]}$$

Given the two clustering assignments U and V on a set of n data points, which have C_U and C_V clusters, respectively, NMI is defined as mutual information between U and V divided by the entropy of U and V. Specifically,

$$\mathrm{NMI} = \frac{\sum_{p=1}^{C_U}\sum_{q=1}^{C_V}\left|U_p \cap V_q\right| \log\frac{n\left|U_p \cap V_q\right|}{\left|U_p\right| \times \left|V_q\right|}}{\max\left(-\sum_{p=1}^{C_U}\left|U_p\right|\log\frac{\left|U_p\right|}{n}, -\sum_{q=1}^{C_V}\left|V_q\right|\log\frac{\left|V_q\right|}{n}\right)}$$

References

Amodio M, van Dijk D, Srinivasan K, Chen WS, Mohsen H, Moon KR, Campbell A, Zhao Y, Wang X, Venkataswamy M et al. 2019. Exploring single-cell data with deep multitasking neural networks. *Nat Methods* doi:10.1038/s41592-019-0576-7.

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**: 661-667.

Ding J, Condon A, Shah SP. 2018. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nat Commun* **9**: 2002.

Haghverdi L, Buettner F, Theis FJ. 2015. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**: 2989-2998.

Haghverdi L, Buttner M, Wolf FA, Buettner F, Theis FJ. 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**: 845-848.

Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F et al. 2018. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**: 1091-1107 e1017.

Hubert L, Arabie P. 1985. Comparing partitions. *Journal of Classification* **2**: 193-218.

Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**: 1187-1201.

Lee JA, Verleysen M. 2010. Scale-independent quality criteria for dimensionality reduction. *Pattern Recognition Letters* **31**: 2248-2257.

Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. 2018. Deep generative modeling for single-cell transcriptomics. *Nat Methods* **15**: 1053-1058.

McInnes L, Healy J, Melville J. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:180203426*.

Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, Yim K, Elzen AVD, Hirn MJ, Coifman RR et al. 2019. Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol* **37**: 1482-1492.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al. 2011. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**: 2825--2830.

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. 2017. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**: 979-982.

Rand WM. 1971. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association* **66**: 846-850.

Strehl A, Ghosh J. 2003. Cluster ensembles --- a knowledge reuse framework for combining multiple partitions. *J Mach Learn Res* **3**: 583-617.

Tian T, Wan J, Song Q, Wei Z. 2019. Clustering single-cell RNA-seq data with a model-based deep learning approach. *Nature Machine Intelligence* **1**: 191-198.

Van den Berge K, Roux de Bezieux H, Street K, Saelens W, Cannoodt R, Saeys Y, Dudoit S, Clement L. 2020. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**: 1201.

van der Maaten L, Hinton G. 2008. Visualizing Data using t-SNE. *J Mach Learn Res* **9**: 2579-2605.
Wang Y, Huang H, Rudin C, Shaposhnik Y. 2021. Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMAP, and PaCMAP for Data Visualization. *J Mach Learn Res* **22**: 1-73.
Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Gottgens B, Rajewsky N, Simon L, Theis FJ. 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**: 59.
Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**: 14049.