A Meta-learning based Graph-Hierarchical Clustering Method for Single Cell RNA-Seq Data
Supplementary Materials

A. Supplementary Formulas

Adjusted Rand Index (ARI) is defined between ground truth clusters Y and predicted clusters Υ as follows :

$$ARI = \frac{C_n^2(a+d)-(a+b)(a+c)-(c+d)(b+d)}{C_n^2-(a+b)(a+c)-(c+d)(b+d)} \qquad (S1)$$

Where a denotes number of cells that belong to the same cluster in both ground truth and predicted clusters. d denotes number of cells who are in different clusters in both ground truth clusters and predicted clusters. b depicts number of cells belonging to same clusters in Y but different in Υ and c is defined vice versa.

Then, considered N as total number of cells, $|Y|$ as the number of ground truth clusters and $|\Upsilon|$ as the number of predicted clusters respectively, Normalized Mutual Information (NMI) is defined as:

$$NMI = \frac{\Sigma_{i=1}^{|Y|}\Sigma_{j=1}^{|\Upsilon|}\frac{|Y_i \cap \Upsilon_j|}{N}\log(\frac{N|Y_i \cap \Upsilon_j|}{|Y_i||\Upsilon_j|})}{mean(-\Sigma_{i=1}^{|Y|}\frac{|Y_i|}{N}\log\frac{|Y_i|}{N},\ -\Sigma_{j=1}^{|\Upsilon|}\frac{|\Upsilon_j|}{N}\log\frac{|\Upsilon_j|}{N})} \qquad (S2)$$

Finally, Clustering Accuracy (CA) is calculated to represent matching accuracy from the best match allocated from Y to Υ.

$$CA = \max_m \frac{\Sigma_{i=1}^{N} I(y_i = m(v_i))}{N} \qquad (S3)$$

Here $y_i \in Y, v_i \in \Upsilon$, and $I(\cdot)$ denotes the indicator function, and $m(\cdot)$ indicates the function that allocates best mapping relationships between ground truth labels y_i and calculated clusters v_i.

B. Supplementary Tables

TABLE S1. 15 BASIC SINGLE CELL TRANSCRIPTIONAL DATASETS

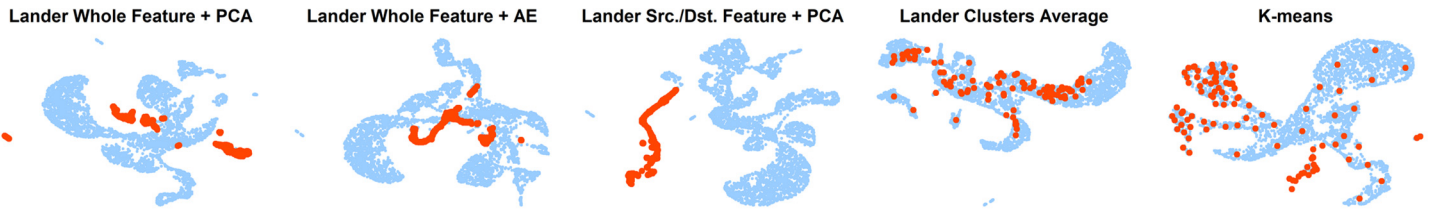| Dataset	Number of Cells	Number of Clusters	Ref.
Baron Human	8569	14	[1]
Baron Mouse	1886	13	[1]
Biase	56	4	[2]
Darmanis	466	9	[3]
Deng	268	6	[4]
Goolam	124	5	[5]
Klein	2717	4	[6]
Li	561	9	[7]
PBMC 68K	68579	11	[8]
Romanov	2881	7	[9]
Segerstolpe	3514	15	[10]
Shekhar Mouse Retina	27499	19	[11]
Tasic	1679	18	[12]
Xin	1600	8	[13]
Zeisel	3005	9	[14]

C. Supplementary Figures

Fig. S1. UMAP Visualization on Different Clustering Center Types from Ablation Experiments. Here light blue spots are cells and red spots are clustering centers.
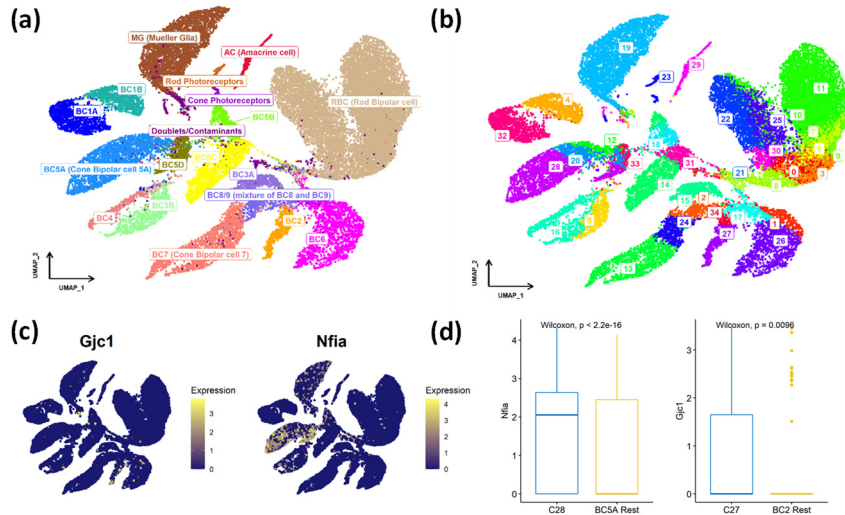

Fig. S2. Cell Subtype Detection on Shekhar Mouse Retina by MeHi-SCC. (a) Ground Truth Cell Types of Shekhar Mouse Retina, (b) Clustering Results Given by MeHi-SCC on Shekhar Mouse Retina, (c) Two Potential Marker Genes Visualized on Shekhar Mouse Retina, (d) Significance Tests for Nfia and Gjc1.

D. References

[1] M. Baron *et al.*, "A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure," *Cell Systems,* vol. 3, no. 4, pp. 346-360.e4, 2016.

[2] F. H. Biase, X. Cao, and S. Zhong, "Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing," (in eng), *Genome Res,* vol. 24, no. 11, pp. 1787-96, Nov 2014.

[3] S. Darmanis *et al.*, "A survey of human brain transcriptome diversity at the single cell level," *Proceedings of the National Academy of Sciences,* vol. 112, no. 23, pp. 7285-7290, 2015/06/09 2015.

[4] Q. Deng, D. Ramsköld, B. Reinius, and R. Sandberg, "Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells," *Science,* vol. 343, no. 6167, pp. 193-196, 2014/01/10 2014.

[5] M. Goolam *et al.*, "Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos," *Cell,* vol. 165, no. 1, pp. 61-74, 2016.

[6] Allon M. Klein *et al.*, "Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells," *Cell,* vol. 161, no. 5, pp. 1187-1201, 2015/05/21/ 2015.

[7] H. Li *et al.*, "Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors," *Nature Genetics,* vol. 49, no. 5, pp. 708-718, 2017/05/01 2017.

[8] G. X. Y. Zheng *et al.*, "Massively parallel digital transcriptional profiling of single cells," *Nature Communications,* vol. 8, no. 1, p. 14049, 2017/01/16 2017.

[9] R. A. Romanov *et al.*, "Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes," *Nature Neuroscience,* vol. 20, no. 2, pp. 176-188, 2017/02/01 2017.

[10] Å. Segerstolpe *et al.*, "Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes," *Cell Metabolism,* vol. 24, no. 4, pp. 593-607, 2016/10/11/ 2016.

[11] K. Shekhar *et al.*, "Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics," *Cell,* vol. 166, no. 5, pp. 1308-1323.e30, 2016/08/25/ 2016.

[12] B. Tasic *et al.*, "Adult mouse cortical cell taxonomy revealed by single cell transcriptomics," *Nature Neuroscience,* vol. 19, no. 2, pp. 335-346, 2016/02/01 2016.

[13] Y. Xin *et al.*, "RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes," *Cell Metabolism,* vol. 24, no. 4, pp. 608-615, 2016/10/11/ 2016.

[14] A. Zeisel *et al.*, "Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq," *Science,* vol. 347, no. 6226, pp. 1138-1142, 2015/03/06 2015.