

Supplementary methods

Modeling gene expression data by autoencoder-based framework

Regarding gene expression data, we adopting our previous study modeled it as drawn from negative binomial (NB) distribution by an autoencoder-based framework¹. Specifically, we learned it's d -dimensional features z through an encoder E , and then transformed z into the parameters of NB distribution by corresponding decoder (D_u and D_θ):

$$p(x|z) = NB(x; u_x, \theta_x) = NB(x, l_x; D_u(z), D_\theta(z)) \quad (1)$$

$$NB(x; u_x, \theta_x) = \frac{\Gamma(x + \theta_x)}{\Gamma(\theta_x)\Gamma(x + 1)} \left(\frac{u_x}{u_x + \theta_x}\right)^x \left(\frac{\theta_x}{\theta_x + u_x}\right)^{\theta_x} \quad (2)$$

where each dimension of u_x and θ_x indicates the mean and dispersion of NB distribution for each gene, and which is simultaneously inferred by D_u by using 'softmax' activation function at the last layer and D_θ , respectively. One-dimensional constant variable l_x calculated by the sum of read counts of all selected genes for each cell, serves as the cell-specific normalized factors.

The training objective of the model is to maximize the marginal likelihood of observed gene expression data, and the loss function is summarized as follows:

$$\log p(x|z) = E_{z \sim p(z|x, E)}(\log p(x|z; D_u, D_\theta)) \quad (3)$$

In this work, each neural network uses batch normalization, 'relu' is regarded as the activation function between two hidden layers, and the Adam optimizer with both a $1e^{-6}$ weight decay and $8e - 05$ learning rate is used to minimize the above loss function. In addition, the framework of the autoencoder is fixed as [2000, 1000, 50, 1000, 2000] because of 2,000 HVGs used to capture the inner structure of gene expression data.

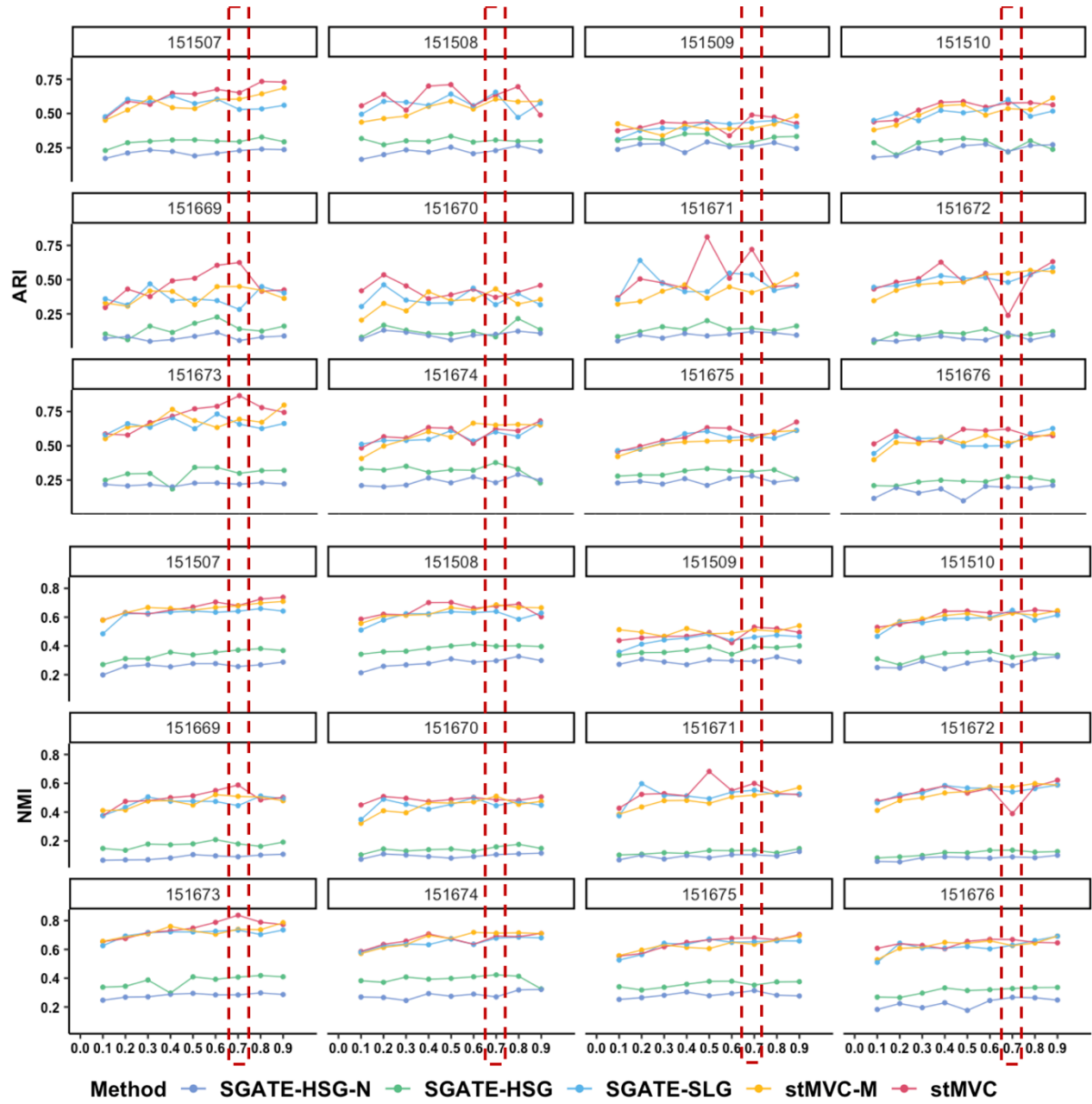
Evaluation of the quality of denoised gene expression data

Similar to our previous studies^{1,2}, we adopted a Gini-index (GI) based measure to quantify the quality of denoised gene expression data by considering the information of marker genes, housekeeping genes, and cell clustering predicted by each computational method. Specifically, the average of each of both marker and housekeeping genes was calculated for each cluster; the GI for each of genes including marker and housekeeping gene was calculated by the 'gini' function from reldist package³; and then the difference between GI per each marker and the

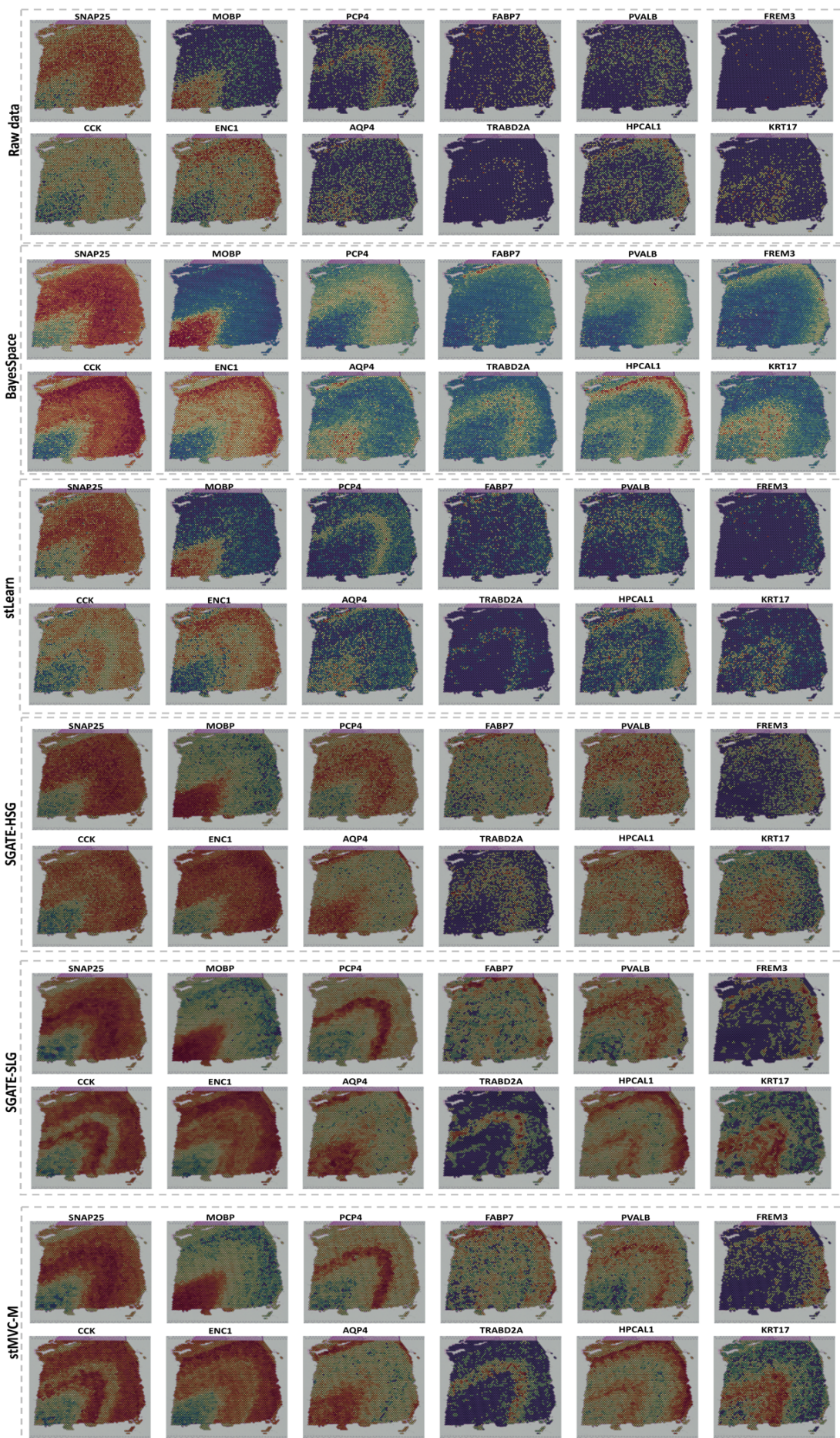
average of that of all housekeeping genes as the AGI to evaluate the specific level of each marker gene for a cluster $cluster_m$. The higher AGI score, the better the denoised data.

$$AGI = gini(gene_i) - \frac{1}{|G_{hk}|} (\sum_{gene_j \in G_{hk}} gini(gene_j)) \quad (4)$$

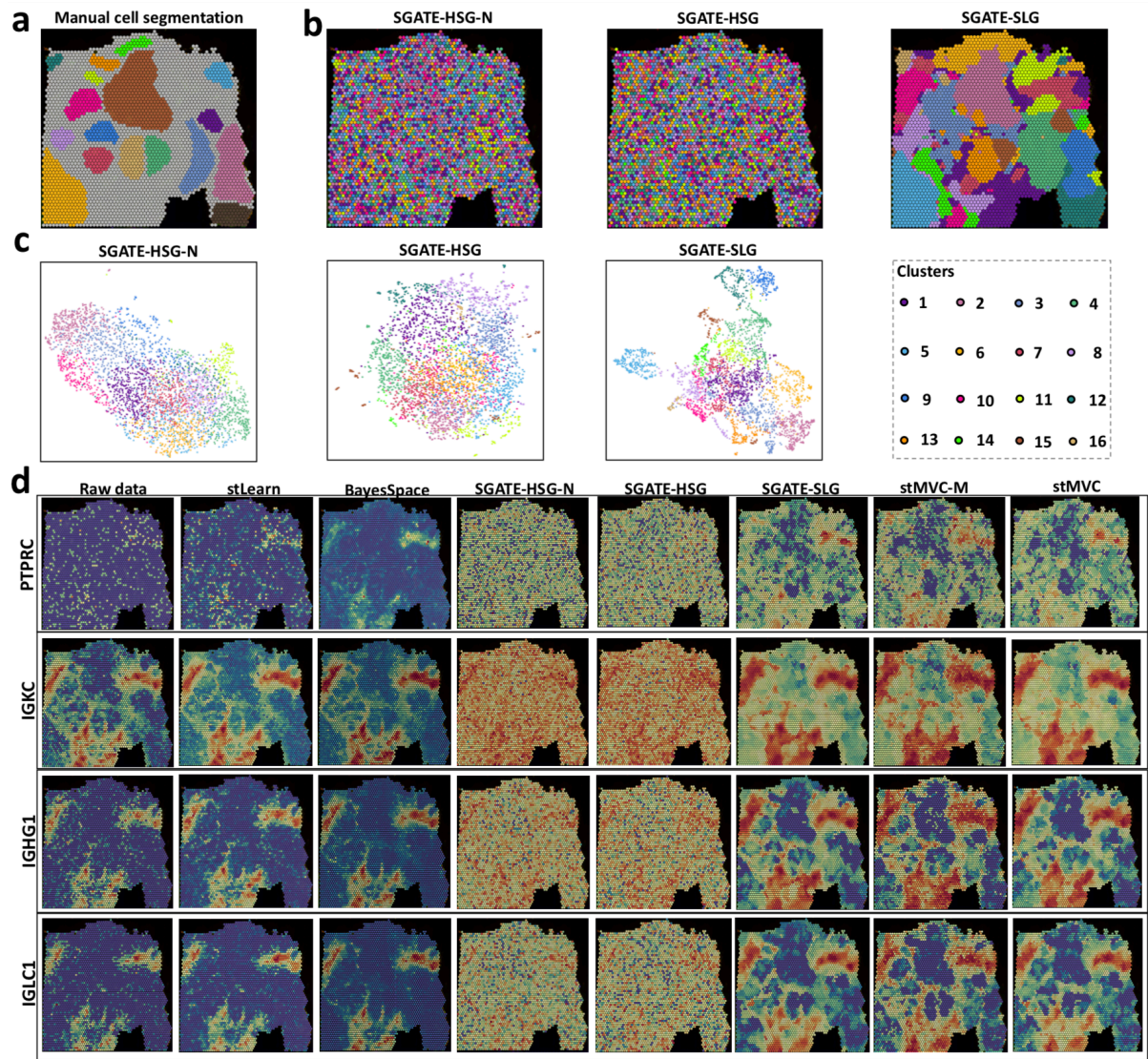
Where G_{hk} indicates the gene set of the housekeeping genes downloaded from the previous studies ⁴, and $|G_{hk}|$ is the number of housekeeping genes.



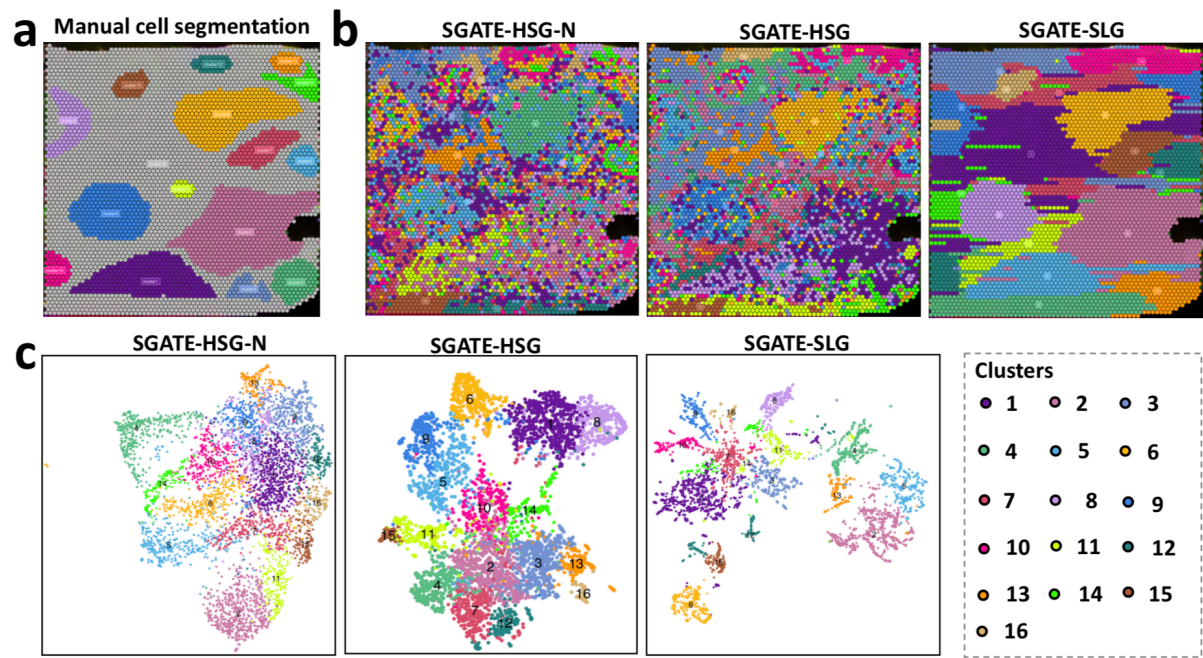
Supplementary Fig.1. Evaluation of the proportion of labels used to train the stMVC, stMVC-M, and all SGATE-based single-view models by clustering accuracy including ARI and NMI, which were the comparison of the predicted cell clusters by the Louvain algorithm with the annotation, on the 12 slices of DLPFC dataset. Each color indicates one method. The X-axis indicates the proportion of labels for the training of the model.



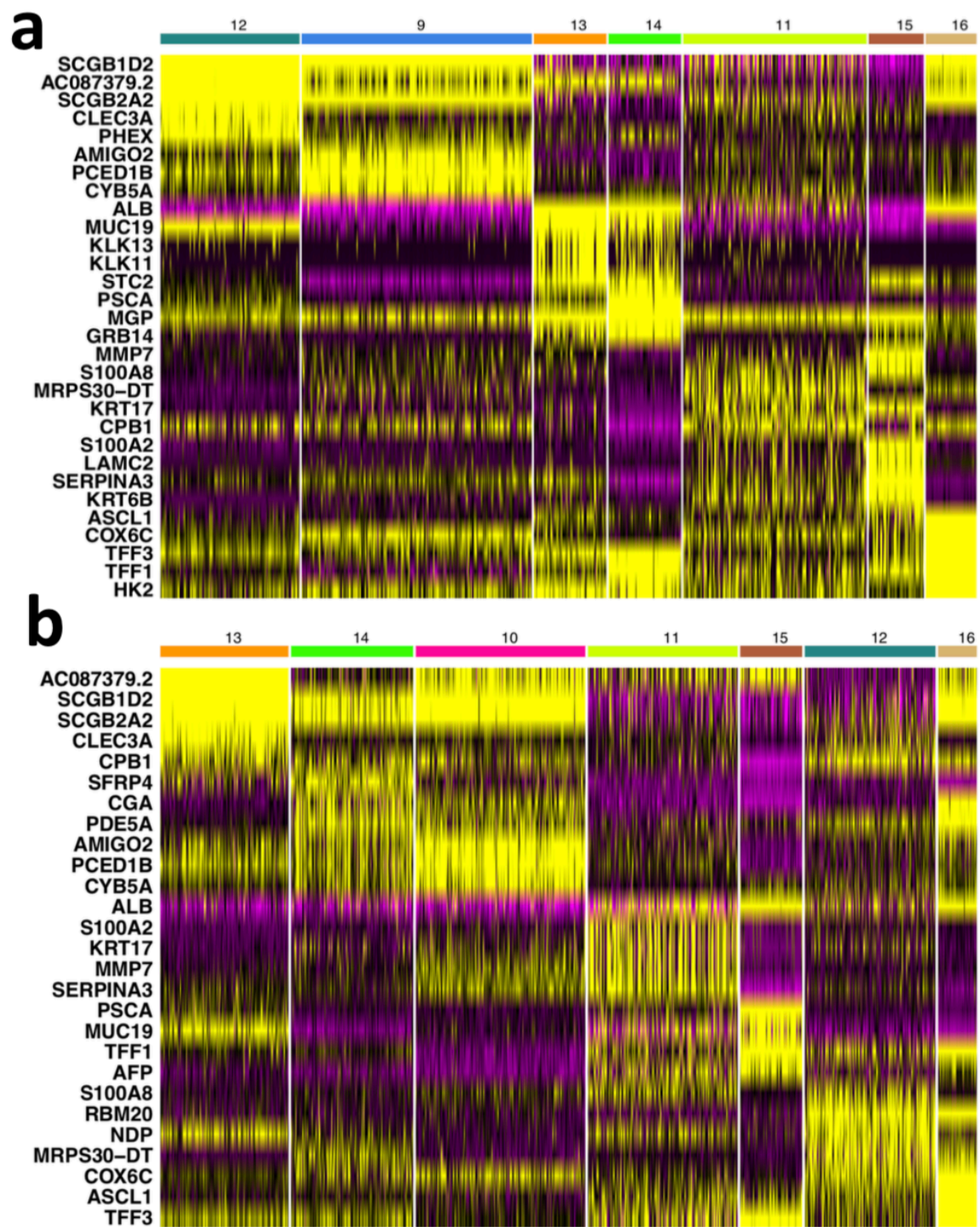
Supplementary Fig.2. Spatial expression of layer-specific genes: SNAP25, MOBP, PCP4, FABP7, PVALB, FREM3, CCK, ENC1, AQP4, TRABD2A, HPCA1, and KRT17 for slice 151673 data denoised by BayesSapce, stLearn, SGATE-HSG, SGATE-SLG, and stMVC-M, respectively, as well as raw data as a comparison



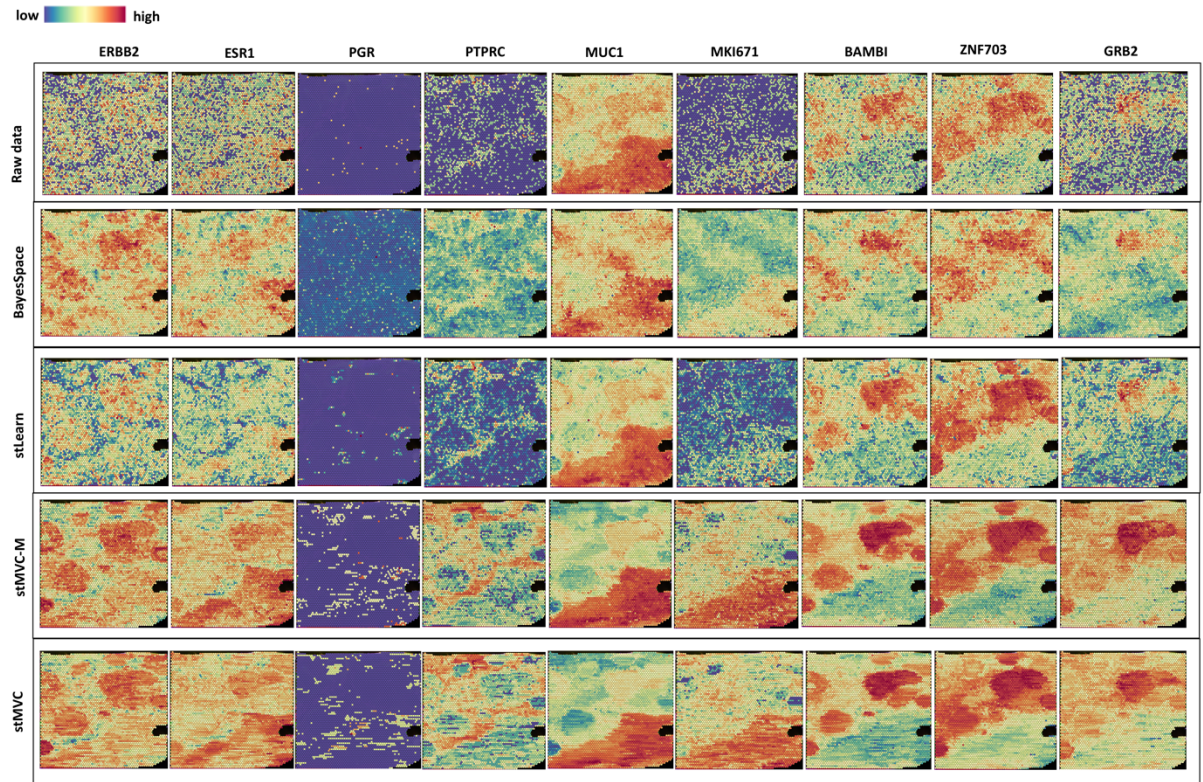
Supplementary Fig. 3. Method comparison on the OEAD sample. **a** Manual cell segmentation with 18 regions. Each color indicates one region. **b** Spatial clustering by SGATE-based single-view models: SGATE-HSG-N, SGATE-HSG, and SGATE-SLG. **c** UMAP visualization of the latent features by SGATE-HSG-N, SGATE-HSG, and SGATE-SLG. Each cluster is indicated by one color. **d** Spatial expression of genes for immune-related markers: PTPRC, IGKC, IGHG1, IGLC1 for the data denoised by BayesSpace, stLearn, SGATE-HSG, SGATE-SLG, stMVC-M, and stMVC, respectively, as well as raw data as a comparison



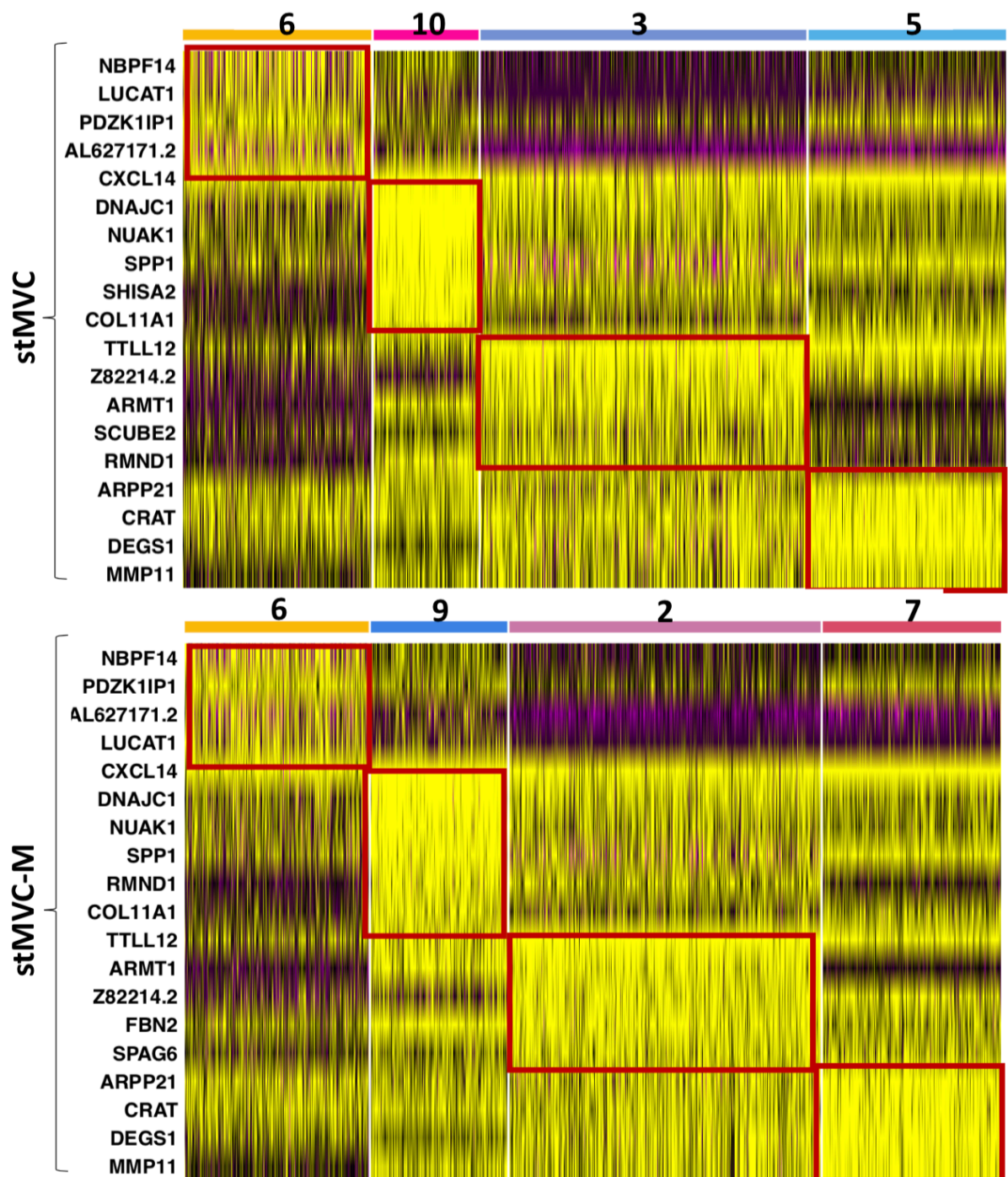
Supplementary Fig.4. Method comparison on the IDC sample. **a** Manual cell segmentation with 16 regions. Each color indicates one region. **b** Spatial clustering by SGATE-based single-view models: SGATE-HSG-N, SGATE-HSG, and SGATE-SLG. **c** UMAP visualization of the latent features by SGATE-HSG-N, SGATE-HSG, and SGATE-SLG, respectively. Each cluster is indicated by one color.



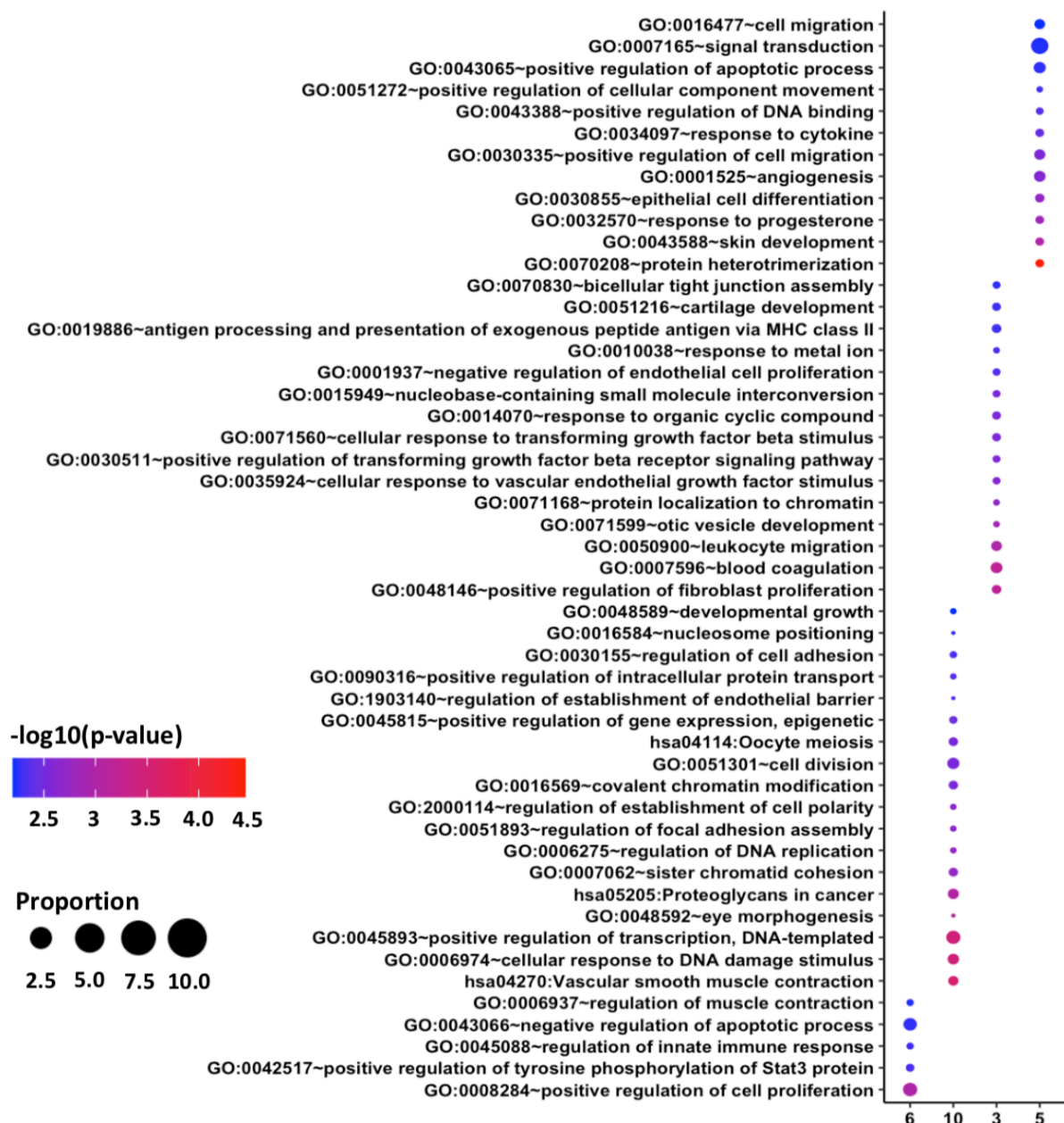
Supplementary Fig.5. Heatmap of the gene expression of SVGs identified from the clusters enriched in the region of both CIS and BH by stMVC (**a**) and stMVC-M (**b**).



Supplementary Fig.6. Spatial expression of genes for indicative markers: ERBB2, ESR1, PGR, immune genes: PTPRC, and tumor progression genes: BAMBI, MUC1, MKI67, ZNF703, and GRB2 for the data denoised by BayesSapce, stLearn, SGATE-HSG, SGATE-SLG, stMVC-M, and stMVC, respectively, as well as raw data as a comparison.



Supplementary Fig.7. Heatmap of the gene expression of SVGs from four stMVC clusters of stMVC and stMVC-M. The clusters 6, 10, 3, and 5 of stMVC are consistent with the clusters 6, 9, 2, and 7 of stMVC-M.



Supplementary Fig.8. Gene function enrichment analysis of SVGs in each of four domains from stMVC by DAVID. Dot size and color indicate the percentage and p – value of each function in each domain, respectively.

Reference

- 1 Zuo, C., Dai, H. & Chen, L. Deep cross-omics cycle attention model for joint analysis of single-cell multi-omics data. *Bioinformatics* (2021).
- 2 Zuo, C. & Chen, L. Deep-joint-learning analysis model of single cell transcriptome and open chromatin accessibility data. *Brief Bioinform* **22**, bbaa287 (2021).
- 3 Mark S. Handcock & Morris, M. *Relative Distribution Methods in the Social Sciences*. (Springer, 1999).
- 4 Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet* **29**, 569-574 (2013).