

---

## Supplementary information

---

# Modeling gene regulatory networks using neural network architectures

---

In the format provided by the  
authors and unedited

Supplementary materials for  
**Modeling Gene Regulatory Networks Using Neural Network Architectures**

Hantao Shu<sup>1</sup>; Jingtian Zhou<sup>2,3</sup>; Qiuyu Lian<sup>4,5</sup>; Han Li<sup>1</sup>; Dan Zhao<sup>1</sup>; Jianyang Zeng<sup>1,\*</sup>; Jianzhu Ma<sup>6,\*</sup>

<sup>1</sup>Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 10084, China

<sup>2</sup>Genomic Analysis Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, United States

<sup>3</sup>Bioinformatics Program, University of California, San Diego, La Jolla, CA 92093, United States

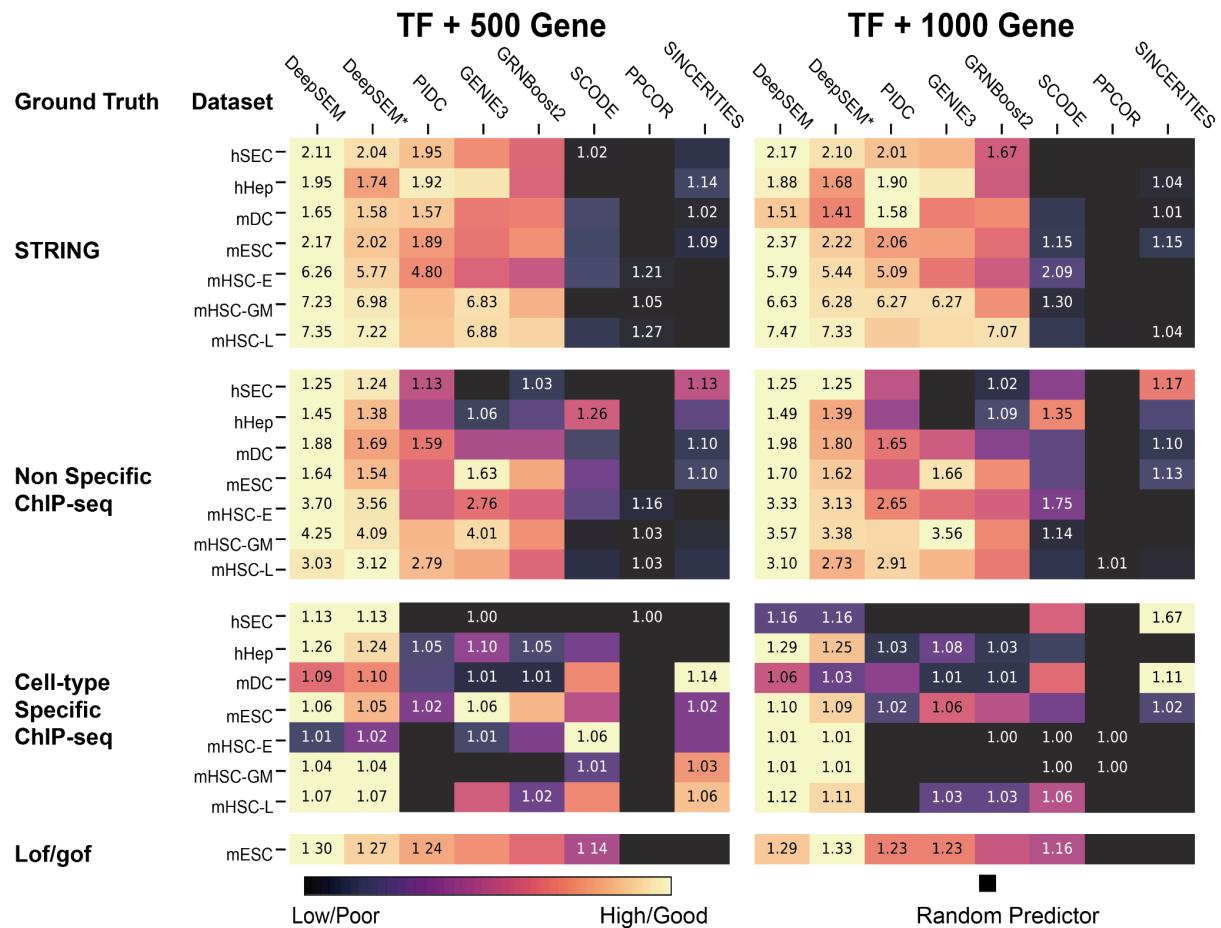
<sup>4</sup>UM-SJTU Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>5</sup>Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China

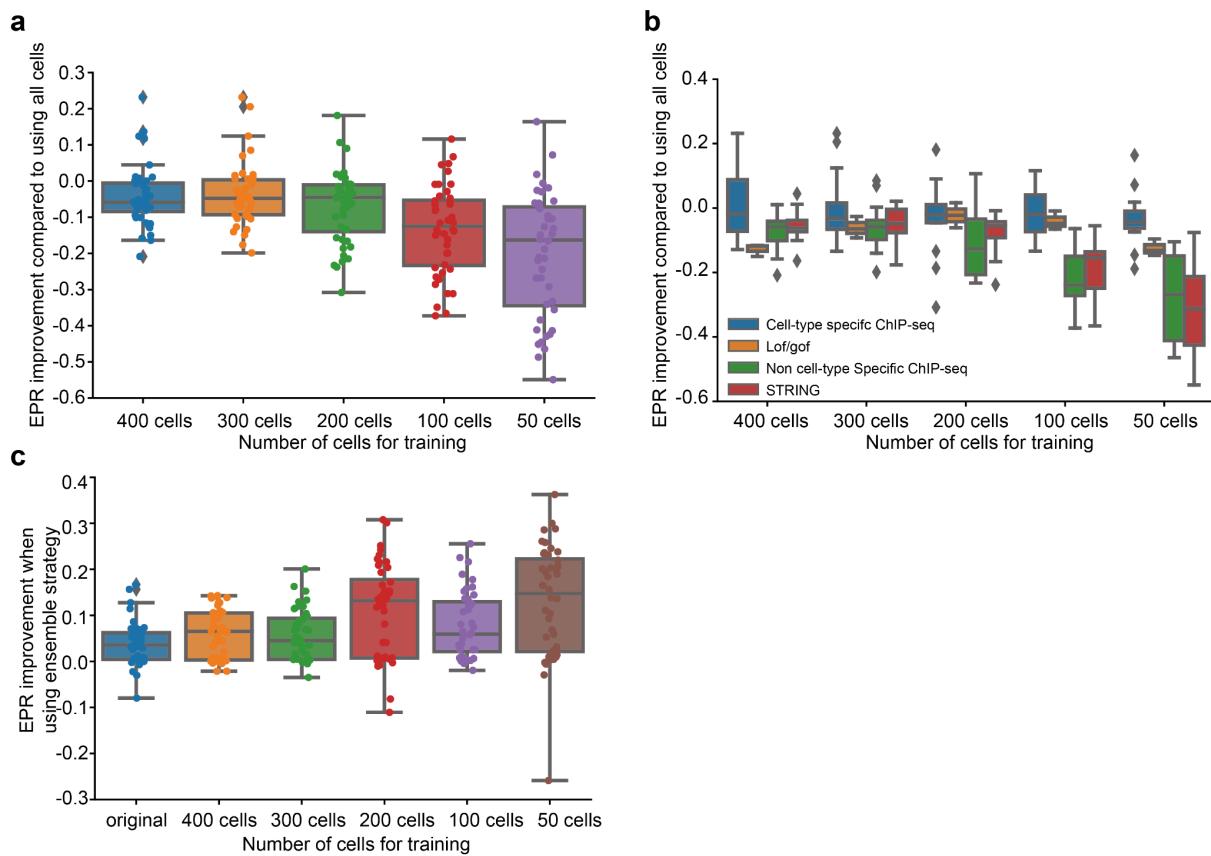
<sup>6</sup>Institute for Artificial Intelligence, Peking University, Beijing 100871, China

\* Correspondence: zengjy321@mail.tsinghua.edu.cn (J.Zeng), majianzhu@pku.edu.cn (J.M.)

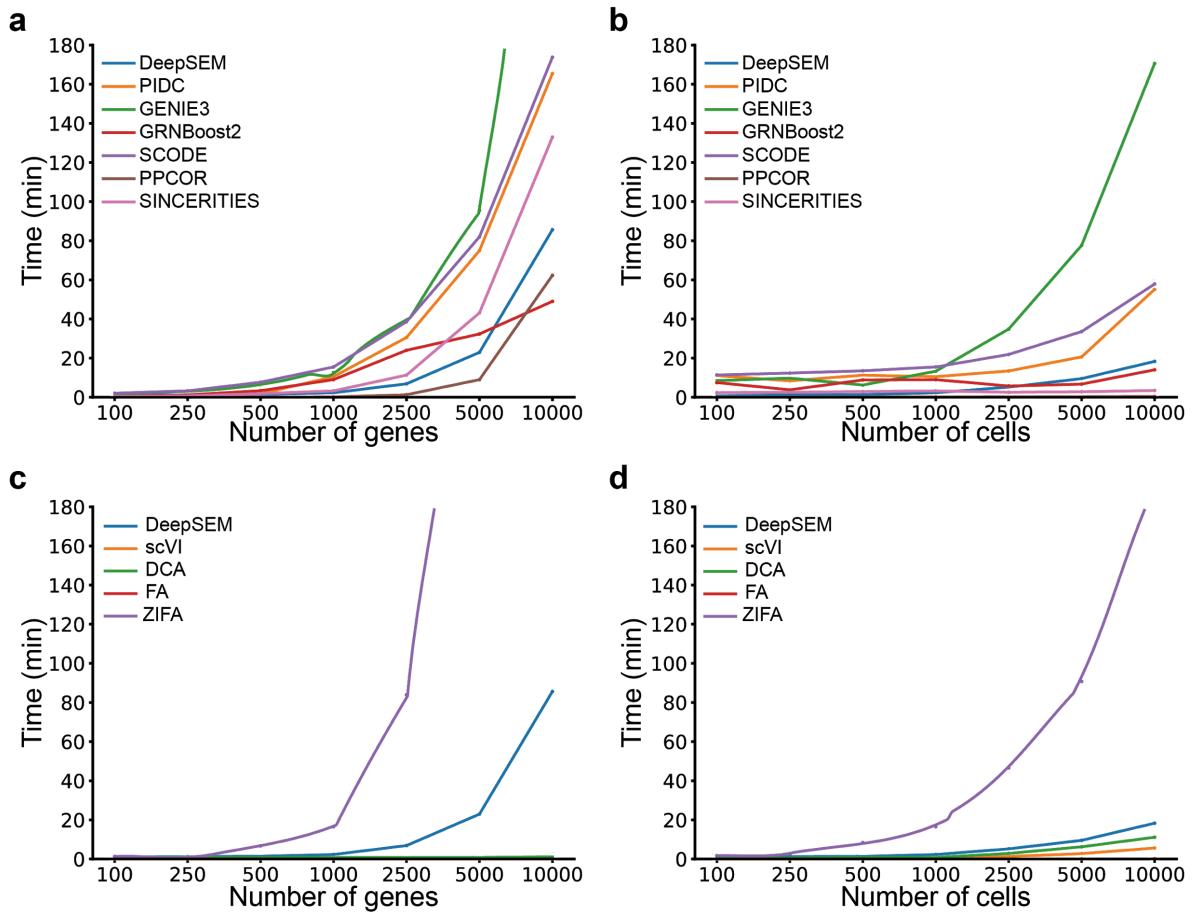
## Supplementary Figures and Tables



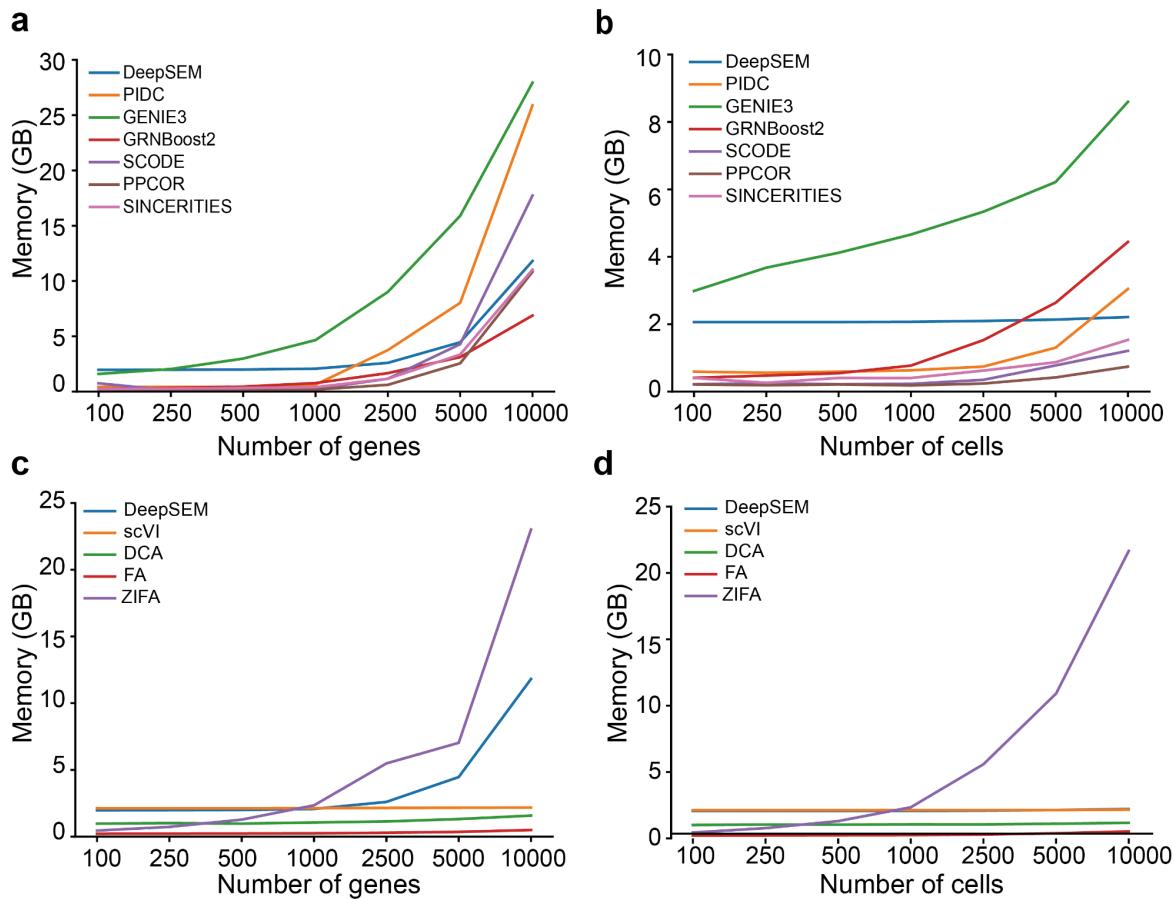
**Supplementary Figure 1. Summary of GRN inference results for all the methods in terms of AUPRC ratio.** The performance of GRN inference on seven different datasets with four distinct ground truth sources with 500 (left) and 1000 (right) most-varying genes and all varying TFs evaluated by median AUPRC ratio value over ten repeats. For each dataset, the color is scaled between 0 and 1 using the min-max scale (ignoring AUPRC ratio values less than 1), and black squares denote the performance worse than random predictors. DeepSEM\* denotes the DeepSEM without using ensemble learning. AUPRC ratio is defined as the odds ratio of the area under the precision-recall curve (AUPRC) between the model and the random predictions.



**Supplementary Figure 2: Performance of DeepSEM with limited training cell.** **a-b**, The performance improvement of DeepSEM with the limited number of training cells compared to the original dataset in terms of EPR. **a**, The overall performance improvement for all datasets. **b**, The performance improvement for each type of ground truth. **c**, The performance improvement rate of ensemble strategy for DeepSEM with the limited number of training cells. The Box plots display the median (line), the interquartile range (box), and the whiskers (extending 1.5 times the interquartile range).

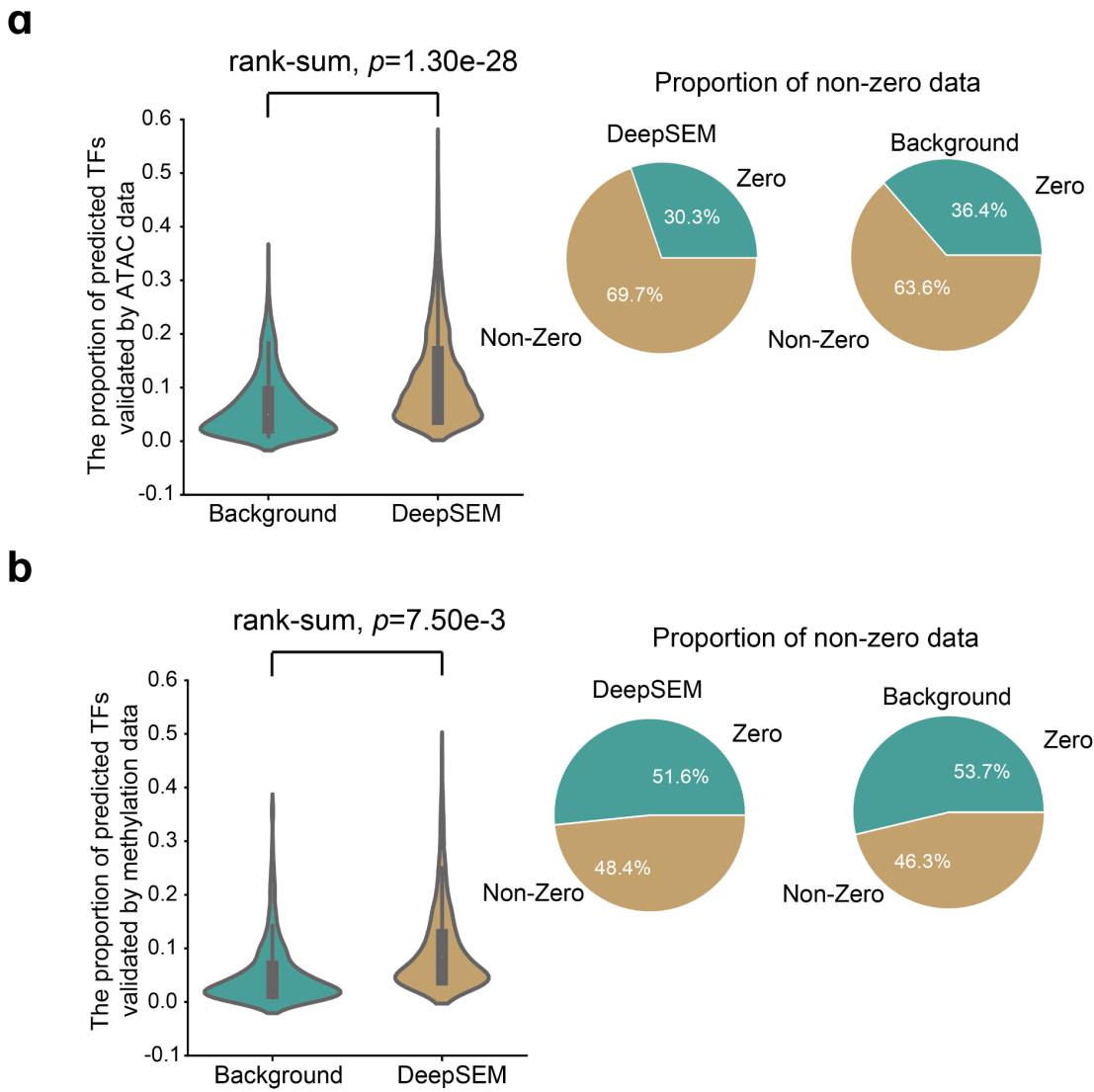


**Supplementary Figure 3. Running time of different methods on the simulated datasets.** Algorithms were tested on a machine with one Intel Xeon CPU E5-2630 CPU with 256 GB RAM, and one GeForce GTX 1080 Ti GPU with 12 GB RAM. Algorithms whose running time is longer than three hours were excluded in the figure. **a**, Running time of training DeepSEM and other GRN inference methods on a simulated dataset with 1000 cells when the number of genes for each cell increased. **b**, Running time of training DeepSEM and other GRN inference methods on a simulated dataset with 1000 genes for each cell when the number of cells increased. **c**, Running time of training DeepSEM and other embedding methods on a simulation dataset with 1000 cells when the number of genes for each cell increased. **d**, Running time of training DeepSEM and other embedding methods on a simulation dataset with 1000 genes for each cell when the number of cells increased.

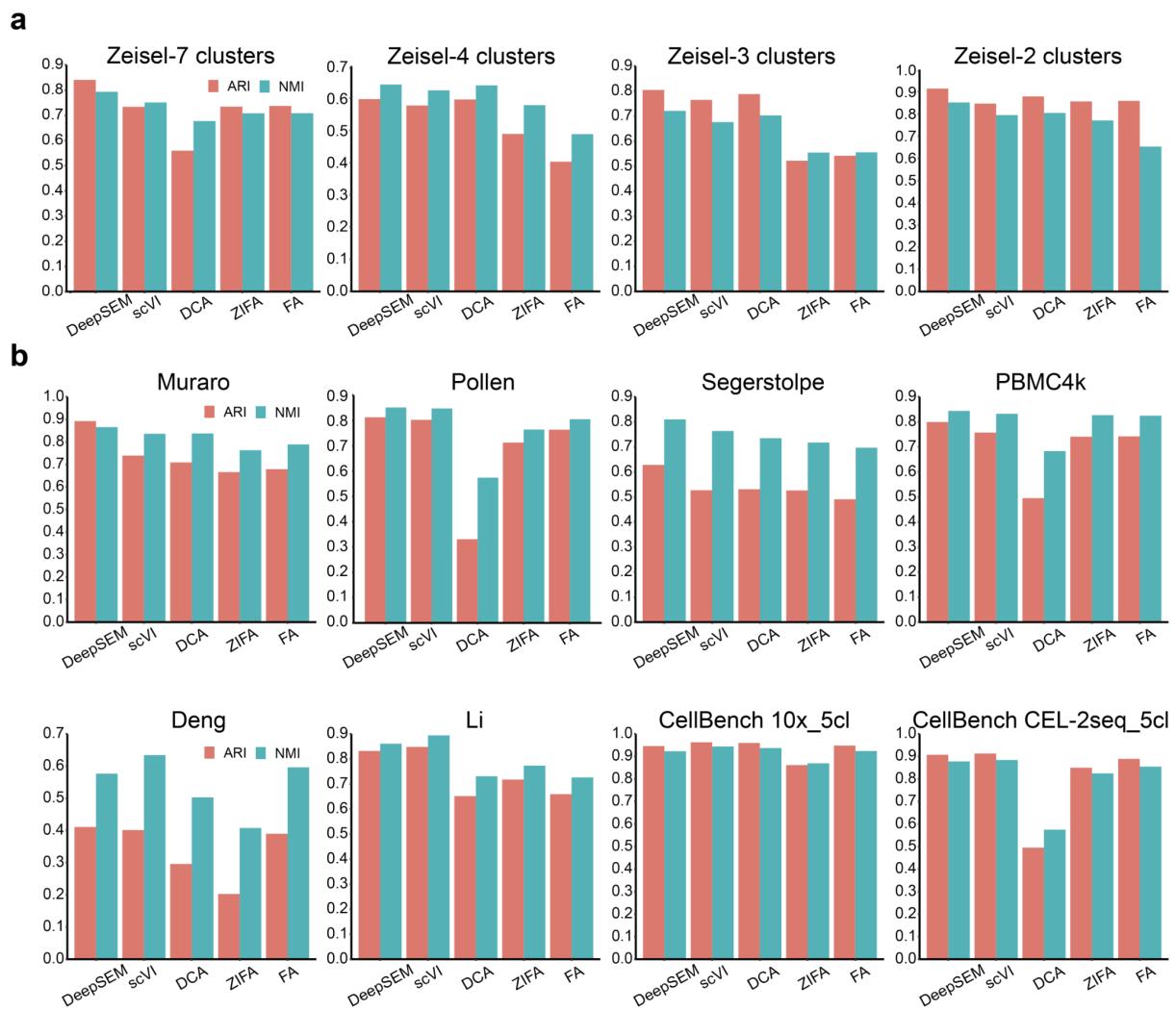


**Supplementary Figure 4: Memory cost of different methods on the simulated datasets.**

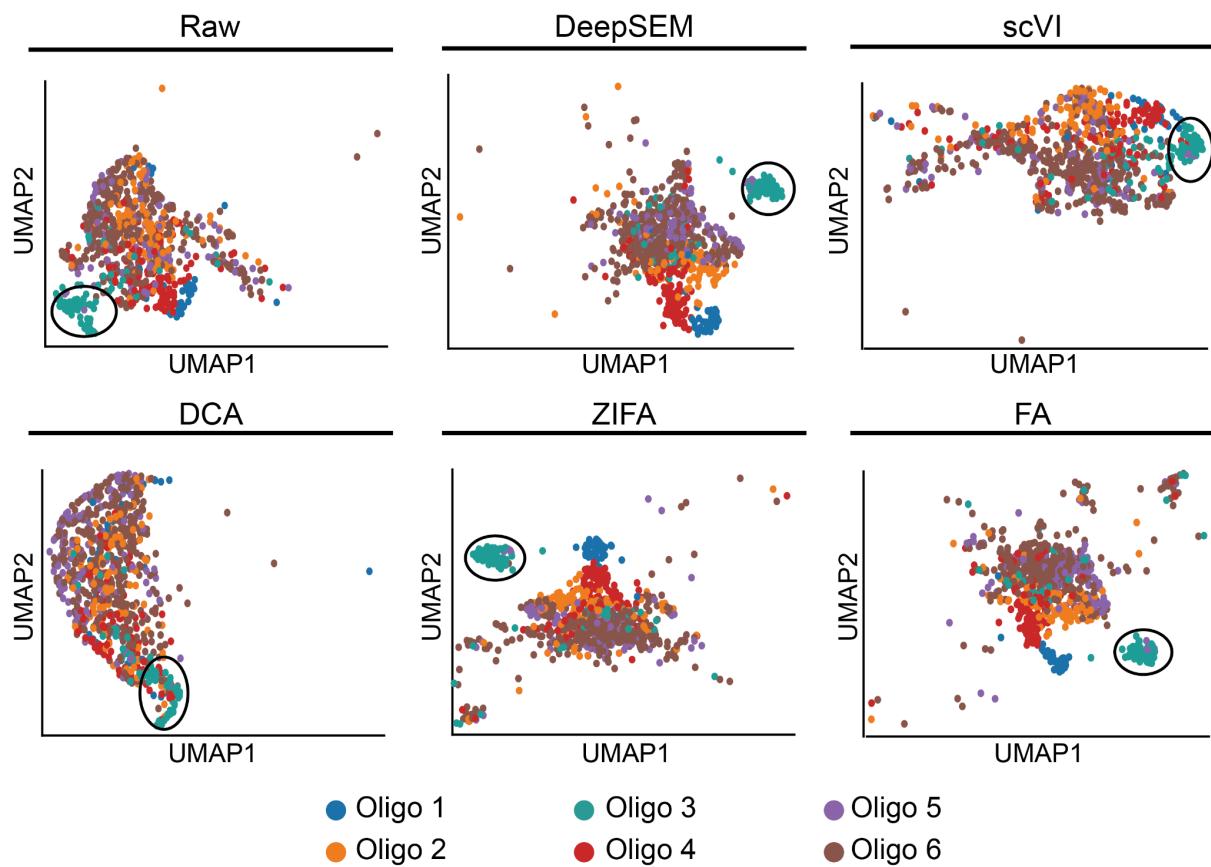
Algorithms were tested on a machine with one Intel Xeon CPU E5-2630 CPU with 256 GB RAM, and one GeForce GTX 1080 Ti GPU with 12 GB RAM. **a**, Memory cost for training DeepSEM and other GRN inference methods on a simulated dataset with 1000 cells when the number of genes for each cell increased. **b**, Memory cost for training DeepSEM and other GRN inference methods on a simulated dataset with 1000 genes for each cell when the number of cells increased. **c**, Memory cost of training DeepSEM and other embedding methods on a simulation dataset with 1000 cells when the number of genes for each cell increased. **d**, Memory cost of training DeepSEM and other embedding methods on a simulation dataset with 1000 genes for each cell when the number of cells increased.



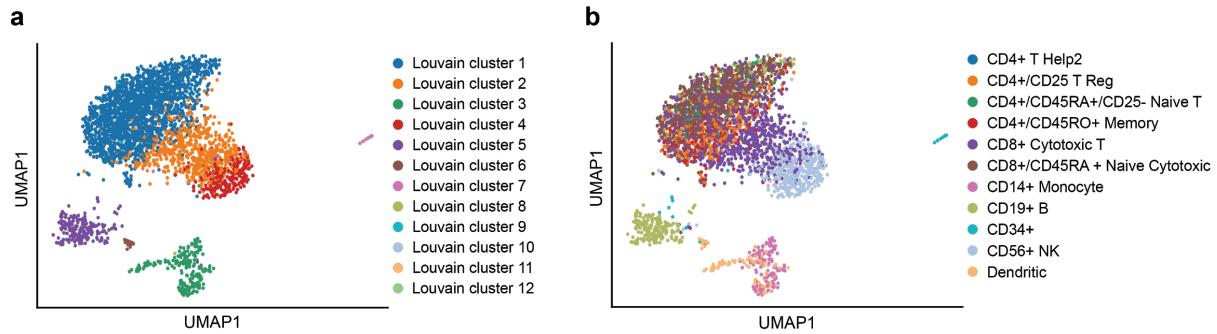
**Supplementary Figure 5. GRNs identified by DeepSEM are supported by single-cell epigenetic data.** For each cell type specific gene (panel **a**: n=1755, panel **b**: n=897) the proportion of the DeepSEM predicted TFs (Top 5% among all TFs) whose motifs were in ATAC peaks (**a**), or DMR regions (**b**) at +/- 200kbp of its transcription start sites (TSS) compared with background (random selection). Genes without any regulators whose motifs were in ATAC peaks (**a**) or DMR regions (**b**) at +/200kbp of its TSS are removed. We ran tests five times and reported the average p-values of one-sided Wilcoxon test. Left: The violin plot of non-zero data. Right: The proportion of non-zero data.



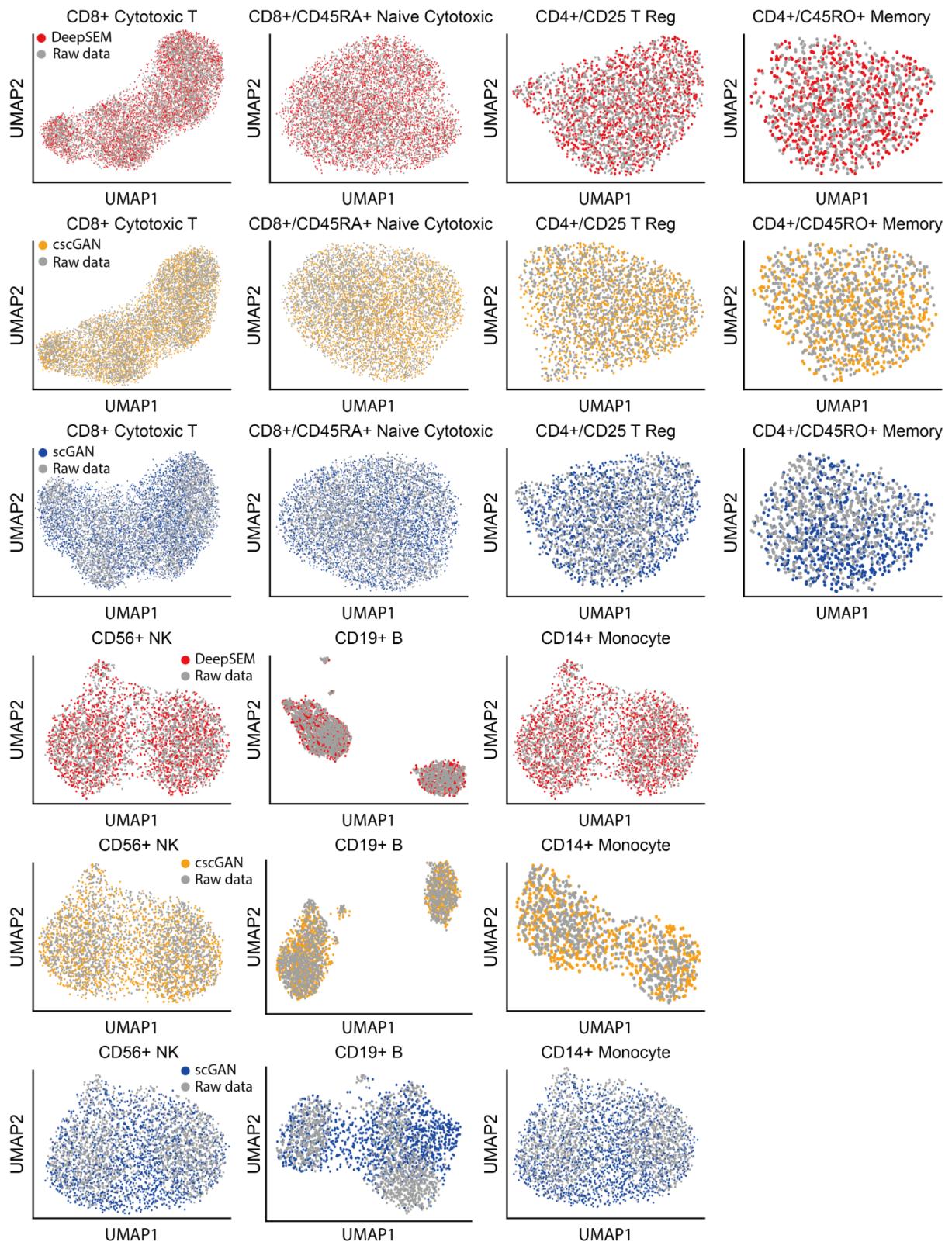
**Supplementary Figure 6. Clustering performance of DeepSEM on nine different datasets compared to scVI, DCA, ZIFA, and FA in terms of ARI and NMI.** **a**, Clustering performance on the Zeisel dataset<sup>1</sup>. The ground truths of Zeisel-7 clusters were the same as the cell type annotation used in scVI<sup>2</sup>. The ground truths of Zeisel-4 clusters, Zeisel-3 clusters, and Zeisel-2 clusters were identified by hierarchical clustering algorithms as in Zeisel et al.<sup>1</sup>. **b**, Clustering performance on the other eight datasets. ARI denotes the adjusted rand index (higher is better) and NMI denotes the normalized mutual information (higher is better).



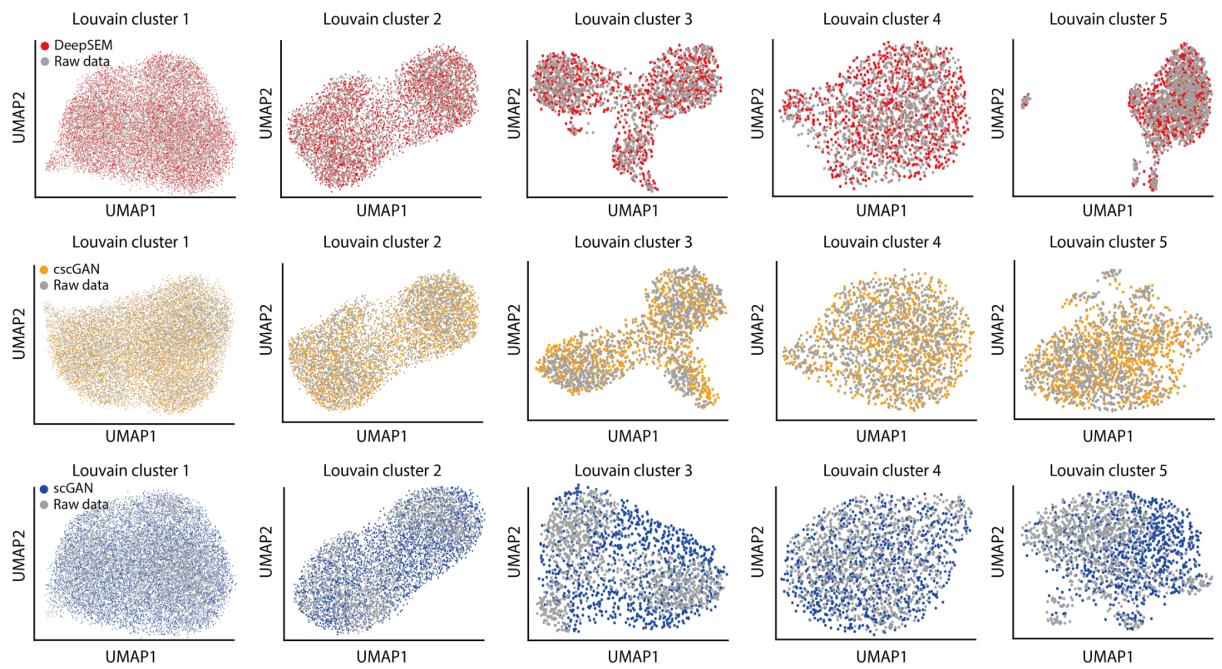
**Supplementary Figure 7. UMAP visualization of DeepSEM on oligodendrocytes cells on the Zeisel dataset.** Oligos 1-6 represent six cell subtypes of oligodendrocytes identified by Zeisel et al.<sup>1</sup>. Zeisel et al.<sup>1</sup> indicated that these subtypes represent different maturation stages of oligodendrocytes cells except Oligo 3 which may present as a distinct cellular state specific for this tissue (highlighted in black circles). More specifically, the maturation of oligodendrocytes cells may start from subtype Oligo 1 (oligodendrocyte precursor cells), followed with Oligo 4 (immature), Oligo 2 (pre-myelinating), Oligo 5 (myelinating) and Oligo 6 (post-myelination). The UMAP visualization of DeepSEM can reveal the differentiation trajectory of oligodendrocytes cells and can clearly distinguish Oligo 3 from others which were consistent with the analysis performed by Zeisel et al.<sup>1</sup>.



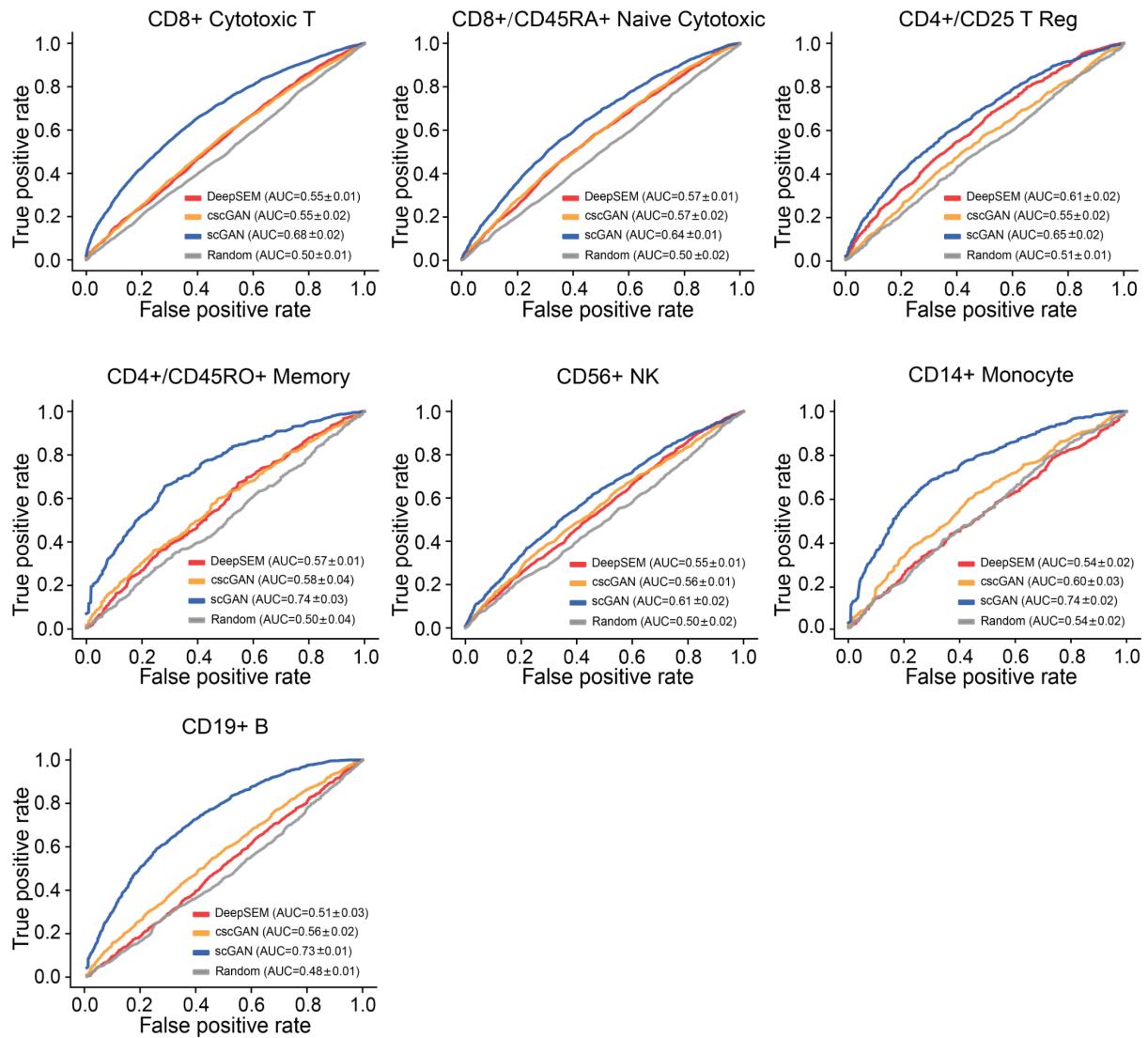
**Supplementary Figure 8. UMAP embedding of the PBMC dataset with different types of annotations. a, UMAP embedding annotated by the Louvain algorithm<sup>3</sup>; b, UMAP embedding annotated by Zheng et al.<sup>4</sup>.**



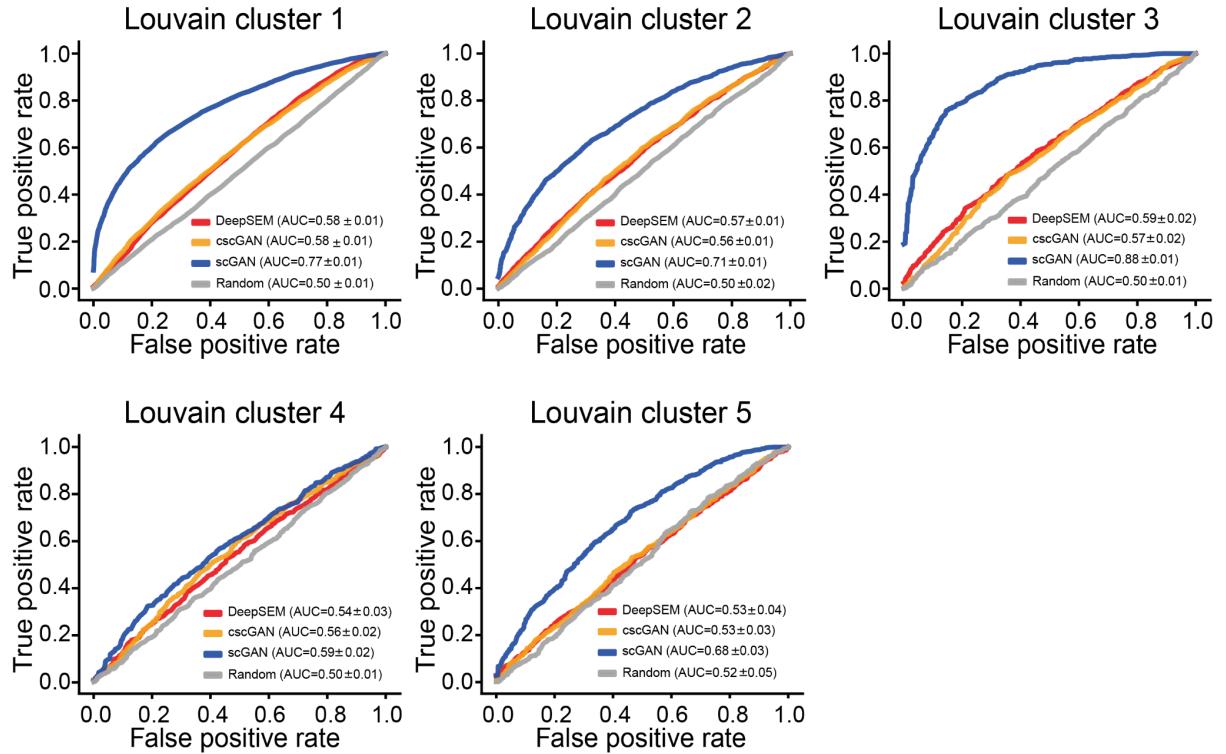
**Supplementary Figure 9. UMAP visualization of scRNA-seq simulations with cell types annotated by Zheng et al.<sup>4</sup>.**



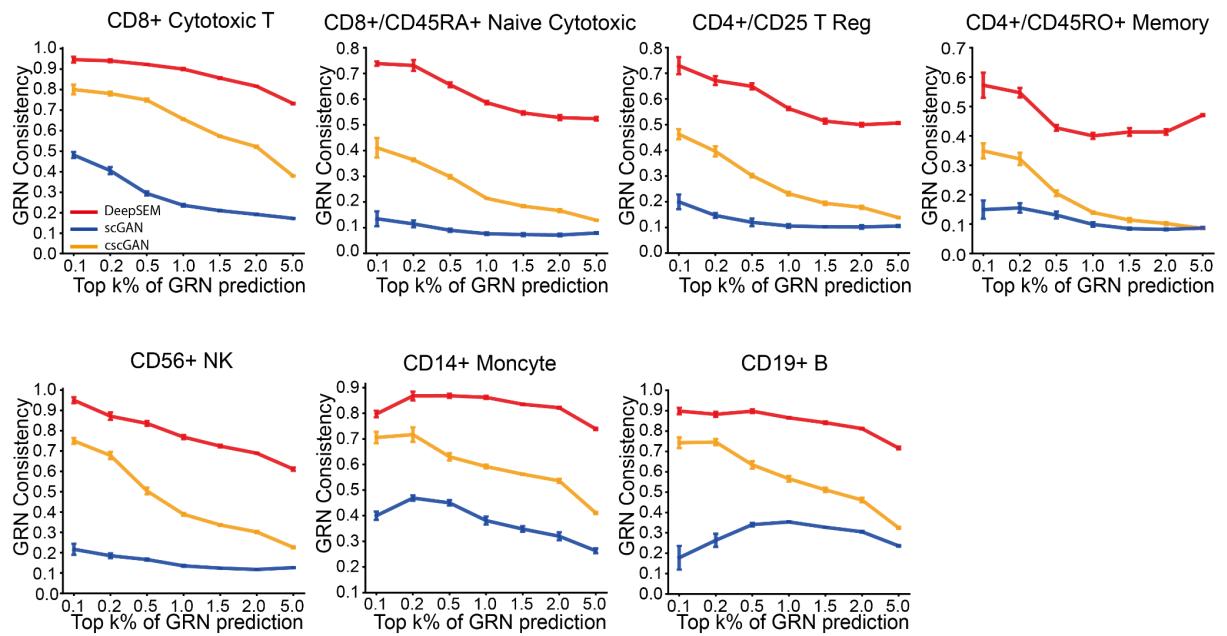
**Supplementary Figure 10. UMAP visualization of scRNA-seq simulations with cell types annotated by the Louvain algorithm<sup>3</sup>.**



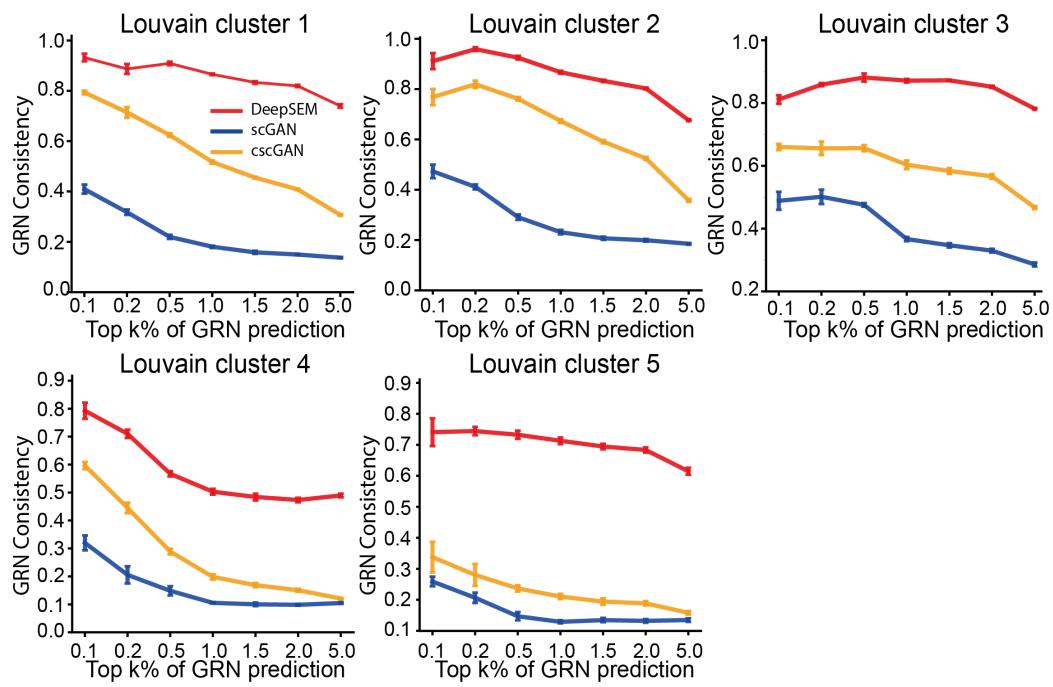
**Supplementary Figure 11. Performance on discriminating simulated cells from real ones on seven datasets annotated by Zheng et al.<sup>4</sup>.** ROC curves and corresponding AUC scores (means  $\pm$  SD, lower is better) were adopted to quantify the performance of random forest prediction with a five-fold cross validation.



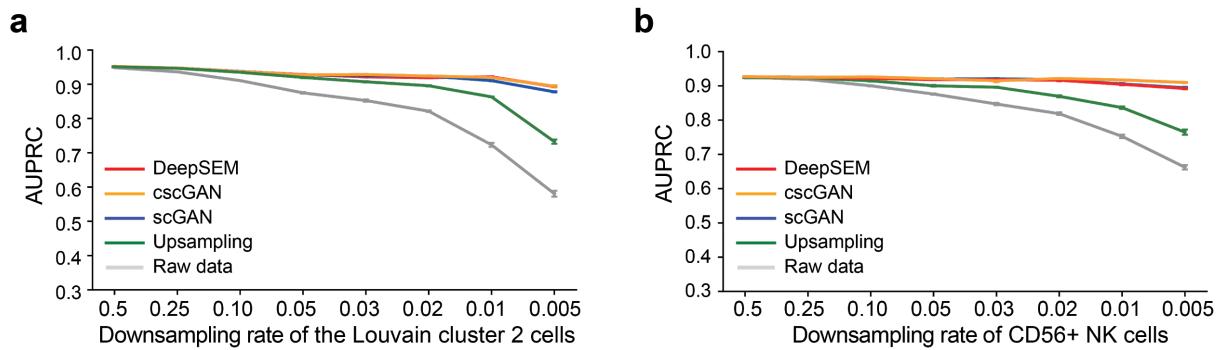
**Supplementary Figure 12. Performance for discriminating simulated cells from real ones on five datasets annotated by the Louvain algorithm<sup>3</sup>.** ROC curves and corresponding AUC scores (means ± SD, lower is better) were adopted to quantify the performance of random forest prediction with a five-fold cross validation.



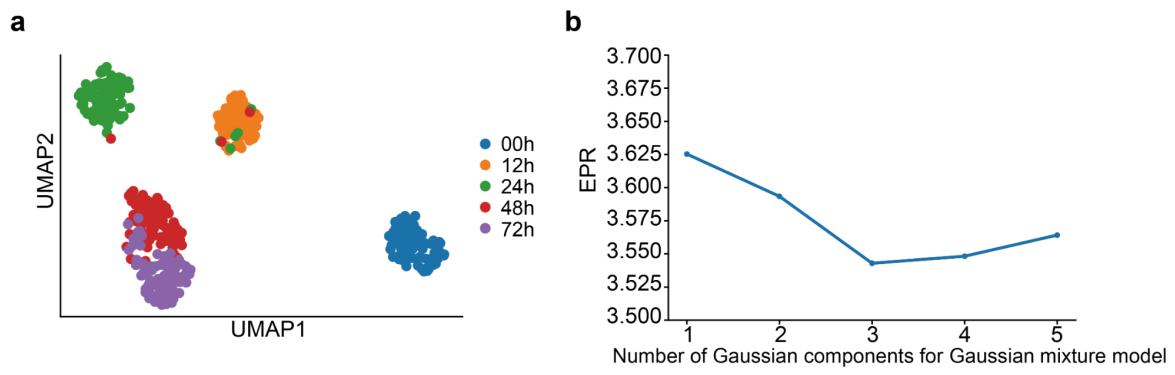
**Supplementary Figure 13. GRN consistency (means  $\pm$  95%CI, higher is better) of cells simulated by different methods on seven datasets annotated by Zheng et al.<sup>4</sup>.**



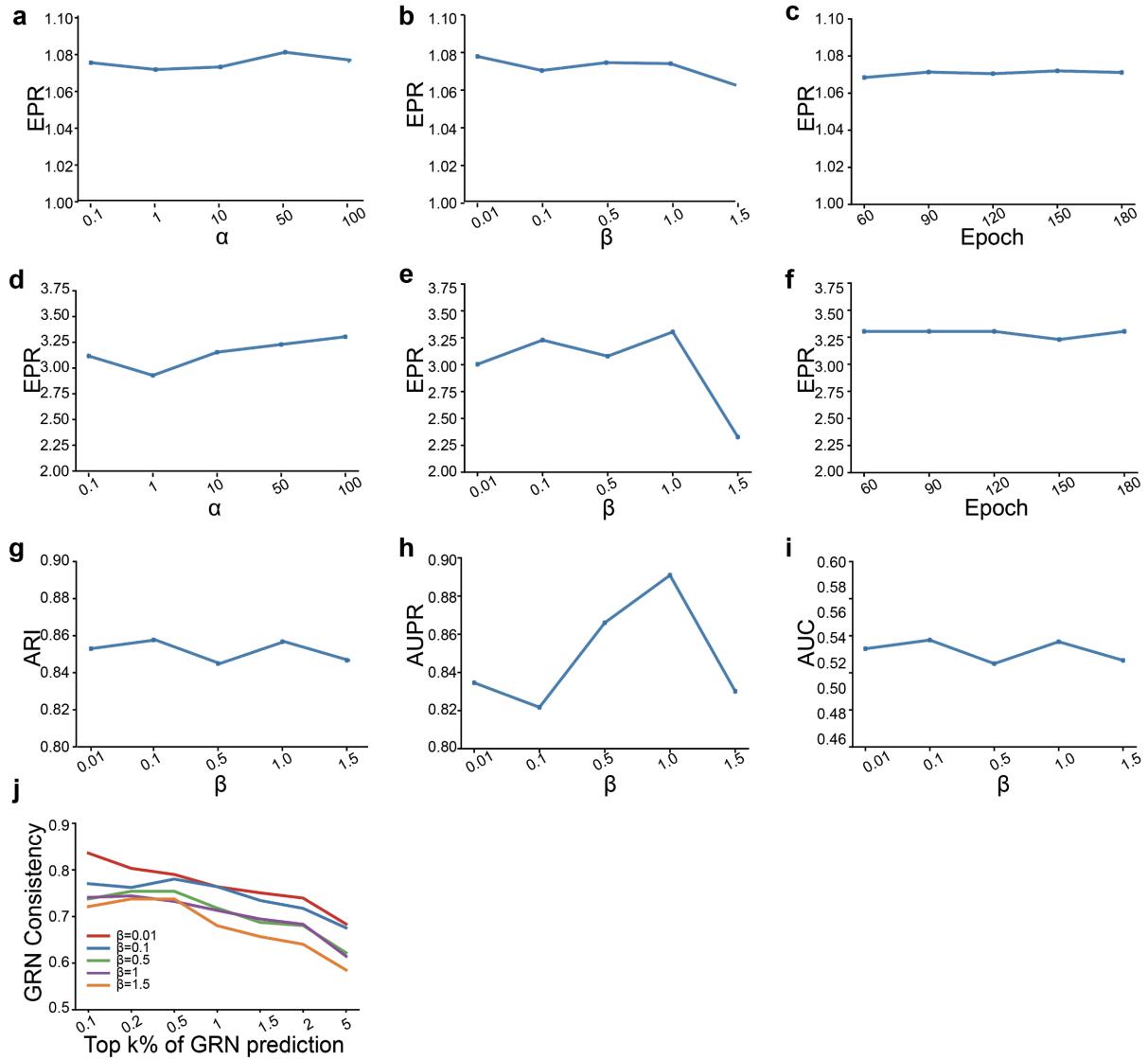
**Supplementary Figure 14. GRN consistency (means $\pm$ 95%CI, higher is better) of cells simulated by different methods on five datasets annotated by the Louvain algorithm<sup>3</sup>.**



**Supplementary Figure 15. Prediction performance of cell type classification with respect to different downsampling rates.** Prediction performance was measured by AUPRC scores (means $\pm$ 95%CI) using a random forest classifier trained to discriminate (a) Louvain cluster 2 cells annotated by the Louvain algorithm and (b) CD56+ NK cells annotated by Zheng et al.<sup>4</sup> from other cells.



**Supplementary Figure 16. Performance of DeepSEM with different Gaussian components for Gaussian mixture model.** **a**, UMAP plot of mESC dataset<sup>5</sup> colored by the different time points. **b**, EPR value of DeepSEM on mESC dataset<sup>5</sup> by using varying TFs and top 500 varying genes as features and STRING as the ground truth.



**Supplementary Figure 17. Hyper-parameter sweeping for DeepSEM.** **a-c**, The influence of individual hyper-parameters (i.e.,  $\alpha$ ,  $\beta$ , and the number of epochs) on the cell-type non-specific gene regulatory network inference task. Performance was evaluated in terms of ERP on the mHSC-L dataset<sup>6</sup> using varying TFs and top 500 varying genes as features and the non-specific ChIP-seq data as the ground truth. **d-f**, The influence of individual hyper-parameters (i.e.,  $\alpha$ ,  $\beta$ , and the number of epochs) on the performance of the cell-type specific gene regulatory network inference task. Performance was evaluated in terms of ERP on the mHSC-L dataset<sup>6</sup> by using varying TFs and top 500 varying genes as features and the cell-type specific ChIP-seq data as ground truth. **g**, The influence of hyper-parameter  $\beta$  on the clustering task. Performance was evaluated by ARI on the Zeisel dataset<sup>1</sup>. **h-j**, The influence of hyper-parameter  $\beta$  on the performance of the scRNA-seq simulation task. **h**, The influence of hyper-parameter  $\beta$  on improving cell type classification performance. Performance was evaluated in terms of AUPR on discriminating cells annotated as Louvain cluster 2 from others with only 0.5% Louvain cluster 2 as the training data. **i**, The performance of discriminating real cells from those simulated by DeepSEM on cells annotated as Louvain cluster 4, evaluated in terms of AUC score. **j**, The influence of hyper-parameter  $\beta$  in terms of GRN consistency evaluated on Louvain cluster 4.

**Supplementary Table 1. Summary of ground truth GRN networks used in the GRN predictions.**

Ground Truth	Cell type	Source	Download link
<b>STRING</b>	mESC, mDC, mHSC-E, mHSC-GM, mHSC-L, hESC, hHep	STRING <sup>7</sup>	<a href="https://string-db.org/">https://string-db.org/</a>
<b>Cell-type Non specific ChIP-seq</b>	mESC, mDC, mHSC-E, mHSC-GM, mHSC-L	TRRUST <sup>8</sup> RegNetwork <sup>9</sup>	<a href="https://www.grnpedia.org/trrust/">https://www.grnpedia.org/trrust/</a> <a href="http://www.regnetworkweb.org/">http://www.regnetworkweb.org/</a>
	hESC, hHep	TRRUST <sup>8</sup> RegNetwork <sup>9</sup> DoRothEA <sup>10</sup>	<a href="https://www.grnpedia.org/trrust/">https://www.grnpedia.org/trrust/</a> <a href="http://www.regnetworkweb.org/">http://www.regnetworkweb.org/</a> <a href="https://saezlab.github.io/dorothea/index.html">https://saezlab.github.io/dorothea/index.html</a>
<b>Cell type specific ChIP-seq</b>	mHSC-E, mHSC-GM, mHSC-L, mDC	ChIP-Atlas <sup>11</sup>	<a href="https://chip-atlas.org/peak_browser">https://chip-atlas.org/peak_browser</a>
	mESC	ChIP-Atlas <sup>11</sup> ESCAPE <sup>12</sup>	<a href="https://chip-atlas.org/peak_browser">https://chip-atlas.org/peak_browser</a> <a href="http://www.maayanlab.net/ESCAPE/download.php">http://www.maayanlab.net/ESCAPE/download.php</a>
	hESC, hHep	ChIP-Atlas <sup>11</sup> ChEA <sup>13</sup>	<a href="https://chip-atlas.org/peak_browser">https://chip-atlas.org/peak_browser</a> <a href="https://maayanlab.cloud/Harmonizome/dataset/CHEA+Transcription+Factor+Targets">https://maayanlab.cloud/Harmonizome/dataset/CHEA+Transcription+Factor+Targets</a>
<b>lof/gof</b>	mESC	ESCAPE <sup>12</sup>	<a href="http://www.maayanlab.net/ESCAPE/download.php">http://www.maayanlab.net/ESCAPE/download.php</a>
<b>scATAC-seq</b>	Mouse cortex	Fang et al. <sup>14</sup>	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126724">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126724</a>
<b>snmC-seq (DMR)</b>	Mouse cortex	Luo et al. <sup>15</sup>	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSEGSE97179">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSEGSE97179</a> Supplimentary Tables in Luo et al. <sup>15</sup> .

**Supplementary Table 2. Summary of scRNA-seq datasets used in GRN prediction.**

Dataset	# of cells	# of genes	GEO
<b>Human embryonic stem cells (hESC)<sup>16</sup></b>	759	17735	GSE75748
<b>Human mature hepatocytes (hHep)<sup>17</sup></b>	426	11515	GSE81252
<b>Mouse dendritic cells (mDC)<sup>18</sup></b>	384	7371	GSE48968
<b>Mouse embryonic stem cells (mESC)<sup>5</sup></b>	422	18385	GSE98664
<b>Erythroid lineages mouse hematopoietic stem cells (mHSC-E)<sup>6</sup></b>	1072	4762	GSE81682
<b>Lymphoid lineages mouse hematopoietic stem cells (mHSC-L)<sup>6</sup></b>	848	4762	GSE81682
<b>Granulocyte-macrophage lineages mouse hematopoietic stem cells (mHSC-GM)<sup>9</sup></b>	890	4762	GSE81682
<b>Mouse cortex VISp (L2/3 IT, L4, L5 IT, L5 PT, L6 IT, L6 CT)<sup>19</sup></b>	6456	31301	GSE115746

**Supplementary Table 3: Summary statistics for the GRN prediction result on mouse cortex dataset.**

Cell Type	# of TF	# of target gene	# of gene regulation detected	# of cell type specific gene regulation	# of gene regulation supported by other type of data
L2/3 IT	487	2762	55724	24510	4789
L4			48746	21922	2861
L5 IT			60111	27611	2026
L5 PT			58957	28806	1131
L6 IT			56283	26161	2948
L6 CT			56777	26276	2444

**Supplementary Table 4. Top 12 cell-type specific TFs identified by DeepSEM and validated by epigenetic data on cell type marker genes *Rorb* and *Syt6*.**

<i>Rorb</i>	<i>Syt6</i>
<i>Mef2c</i>	<i>Tbr1</i>
<i>Pknox2</i>	<i>Stat1</i>
<i>Nr2f1</i>	<i>Yy1</i>
<i>Pou3f2</i>	<i>NfiA</i>
<i>Sp4</i>	<i>Nr2f1</i>
<i>Srebf2</i>	<i>Klf16</i>
<i>Zbtb18</i>	<i>Jun</i>
<i>Smadd3</i>	<i>Ctcf</i>
<i>Pbx1</i>	<i>Thap1</i>
<i>Sp1</i>	<i>Tef</i>
<i>Rora</i>	<i>Sp3</i>
<i>Tbr1</i>	<i>Foxp2</i>

**Supplementary Table 5. Summary of datasets used in embedding visualization and clustering.**

Dataset	# of cells	# of cell types	Download link	Cells exclusion (# of cells)
<b>Zeisel<sup>1</sup></b>	3005	7	<a href="http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60361">http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60361</a>	0
<b>Muraro<sup>20</sup></b>	2126	8	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85241">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85241</a>	Unclear cells(3) Epsilon cells(4)
<b>Segerstolpe<sup>21</sup></b>	2147	9	<a href="https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5061/">https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5061/</a>	Low quality cells and unannotated cells (1348) Epsilon cells(7) Mast cells(5) MHC class II cells(5)
<b>Pollen<sup>22</sup></b>	301	11	<a href="https://www.ncbi.nlm.nih.gov/sra?term=SRP041736">https://www.ncbi.nlm.nih.gov/sra?term=SRP041736</a>	0
<b>PBMC4k<sup>4</sup></b>	3844	8	<a href="https://cf.10xgenomics.com/samples/cell-exp/2.1.0/pbmc4k/pbmc4k_web_summary.html">https://cf.10xgenomics.com/samples/cell-exp/2.1.0/pbmc4k/pbmc4k_web_summary.html</a>	Others and unannotated cells(186)
<b>Li<sup>23</sup></b>	561	9	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81861">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81861</a>	0
<b>Deng<sup>24</sup></b>	268	6	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45719">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45719</a>	0
<b>CellBench10x5cl<sup>25</sup></b>	3918	5	<a href="https://github.com/LuyiTian/sc_mixology">https://github.com/LuyiTian/sc_mixology</a>	0
<b>CellBench CEL-2 seq5cl<sup>25</sup></b>	909	5	<a href="https://github.com/LuyiTian/sc_mixology">https://github.com/LuyiTian/sc_mixology</a>	0

**Supplementary Table 6. KNN neighborhood consistency between raw UMI counts and UMAP embedding provided by DeepSEM and other baseline methods.**

	DeepSEM	scVI	DCA	ZIFA	FA
<b>R (larger is better)</b>	0.2099	0.1705	0.2144	0.0814	0.0661
<b>EMD (lower is better )</b>	0.3184	0.3629	0.3566	0.2779	0.2890
<b>Knn preservation (larger is better)</b>	98.1662	98.1376	98.2658	98.1335	98.1313

**Supplementary Table 7: Embedding performance of DeepSEM on seven cell types compared with other methods in terms of average membership weight among cells.**

	DeepSEM	scVI	DCA	FA	ZIFA
<b>Astrocytes/Ependymal</b>	0.428	0.278	0.270	0.315	0.323
<b>Endothelial/Mural</b>	0.378	0.319	0.218	0.311	0.327
<b>Interneurons</b>	0.308	0.300	0.288	0.285	0.294
<b>Microglia</b>	0.753	0.572	0.235	0.183	0.185
<b>Oligodendrocytes</b>	0.276	0.236	0.207	0.268	0.274
<b>Pyramidal CA1</b>	0.258	0.210	0.198	0.264	0.269
<b>Pyramidal SS</b>	0.365	0.285	0.272	0.350	0.348
<b>Average</b>	0.395	0.314	0.241	0.283	0.289

**Supplementary Table 8. Summary of default hyper-parameters for DeepSEM.**

Hyper-parameters	Cell-type Non-specific GRN prediction (STRING/ Non-specific ChIP-seq)	Cell-type specific GRN prediction (SpecificChIP-seq and lof/gof)	Embedding & Simulation
$\alpha$	100	10	10
$\beta$	1	0.1	1
$\epsilon$	0.0002		
# of epochs	120		
# of layers	Encoder: 3 layers; Decoder: 3 layers		
# of neural of each layers	128		
Activation function	Tanh		
Learning rate	MLPs:1e-4; W: 2e-5		

**Supplementary Table 9. Summary of cell types identified by Zheng et al.<sup>4</sup> on the PBMC dataset used in scRNA-seq data simulation.**

Annotated cell type	# of cells	Download Link
<b>CD8+ Cytotoxic T</b>	20773	<a href="https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/fresh_68k_pbmc_donor_a">https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/fresh_68k_pbmc_donor_a</a>
<b>CD8+/CD45RA+ Naive Cytotoxic</b>	16666	
<b>CD56+ NK</b>	8776	
<b>CD4+/CD25 T Reg</b>	6187	
<b>CD19+ B</b>	5908	
<b>CD4+/CD45RO+ Memory</b>	3061	
<b>CD14+ Monocyte</b>	2862	
<b>Dendritic</b>	2099	
<b>CD4+/CD45RA+/CD25- Naive T</b>	1873	
<b>CD34+</b>	277	
<b>CD4+ T Helper2</b>	97	

**Supplementary Table 10. Summary of clusters identified by the Louvain algorithm<sup>3</sup> on the PBMC dataset used in scRNA-seq data simulation.**

Cluster	# of cells	Marker genes
1	39931	<i>RPS3A,RPL3,LDHB</i>
2	14321	<i>CCL5,NKG7,GZMA</i>
3	4667	<i>FTL,TYROBP,CST3</i>
4	4625	<i>NKG7,GNLY,CST7</i>
5	4092	<i>CD74,HLA-DRA,HLA-DRB1</i>
6	362	<i>CD74,CYBA,GPX1</i>
7	334	<i>GPX1,OAZ1,TAGLN2</i>
8	77	<i>CCL5,NKG7,GNLY</i>
9	65	<i>RPL3,RPLP1,LTB</i>
10	48	<i>RPL3,MALAT1,RPS3A</i>
11	39	<i>B2M,IL32,VIM</i>
12	18	<i>RPS3A,RPL3,TMSB10</i>

## Supplementary Sections

### **Supplementary Section 1. TF and gene selection used in the GRN inference benchmark.**

Following the BEELINE benchmark<sup>26</sup>, for each dataset, we removed those genes that were expressed in fewer than 10% of cells and used the general additive model implemented in the 'gam' package (<https://CRAN.R-project.org/package=gam>) in R to find the high variable genes with p-values corrected by Bonferroni correction. Only those TFs with corrected p-values of variance less than 0.01 and the top 500/1000 varying genes were considered as features in the evaluation of all methods.

### **Supplementary Section 2. Brief introduction and running details about GRN inference algorithms used in this study.**

In this study, we compared DeepSEM with the following six different GRN inference algorithms, including PIDC<sup>27</sup>, GENIE3<sup>28</sup>, GRNBoost2<sup>29</sup>, SCODE<sup>30</sup>, PPCOR<sup>31</sup>, and SINCERITIES<sup>32</sup>.

1. PIDC. PIDC infers the gene regulatory network using partial information decomposition (PID) between genes. With consideration of high order information, PIDC identifies regulatory networks between genes based on the unique contribution to pairwise mutual information (MI) combined with information about the local network context of each gene.
2. GENIE3. GENIE3 infers the gene regulatory network through decomposing the problem into the expression regression problems and using tree-based ensemble methods to predict expression of each target gene from the information of other

genes. For each target gene, the importance of the feature genes is used as the weights of their interactions.

3. GRNBoost2. As a faster alternative to GENIE3, GRNBoost2 uses a stochastic gradient boosting machine regression with early stopping to infer target gene expression. GRNBoost2 uses the same strategy to obtain the gene regulatory network as in GENIE3.
4. PPCOR. PRCOR predicts the gene regulatory network based on the partial and semi-partial correlations between genes. Following the BEELINE benchmark<sup>26</sup>, we took the absolute value of the semi-partial correlation between two genes as their gene interaction score.
5. SCODE. SCODE is designed to reconstruct the gene regulatory network from the scRNA-seq dataset. The main technique employed in SCODE is the linear ordinary differentiation equation (ODE) model and an efficient optimization algorithm performed on differentiating cells.
6. SINCERITIES. SINCERITIES infers the gene regulatory network using the regularized linear regression on time-stamped transcriptional data. Using the change of gene expression of TFs to predict the gene expression of the target gene in the next window shift, the gene regulatory network can be obtained based on the weights of regularized linear regression.

All the code of these baselines can be downloaded from the BEELINE project <https://github.com/Murali-group/Beeline>. We used the same hyper-parameter as in Pratapa et al.<sup>26</sup>.

### **Supplementary Section 3. Scalability of DeepSEM**

To assess the scalability of DeepSEM, we also evaluated the running time and memory costing of DeepSEM on the dataset simulated by Splatter<sup>33</sup> with different numbers of genes and cells included in the model and compared it to other GRN prediction and single-cell modelling algorithms. DeepSEM was the second fastest algorithm among all the GRN prediction algorithms and only slower than PPCOR<sup>31</sup>. For the single cell modelling algorithm, DeepSEM is relatively slow compared to other VAE models such scVI<sup>2</sup> due to the inverse operation of the GRN layer (**Supplementary Fig. 3**). However, the running time is on the same scale and the problem can be easily alleviated by including more computational resources. Simultaneously, although DeepSEM costs less than 5GB memory for most situations, the memory cost is tolerable for most computers (**Supplementary Fig. 4**).

### **Supplementary Section 4. Summary Statistics of GRN prediction result on mouse cortex dataset.**

We provided summary statistics of the GRN prediction result on the mouse cortex dataset. In order to determine the threshold of GRN edge importance. We ran DeepSEM for 10 times with default hyper-parameters on the preprocessed mouse cortex dataset and ensemble the prediction result by average the absolute weight of  $W$ . Next, we selected GRN whose identified importance was significantly larger than others ( $W > \mu + 2\sigma$ , p-values  $<0.0215$  for t-test) for each cell type, where  $\mu$  and  $\sigma$  denote mean values and standard values for the predicted GRN edge important metrics. Next, for each target gene, we further selected at most top 10% TFs as the identified GRN by DeepSEM. The results are shown in **Supplementary Table 3**.

**Supplementary Section 5. Brief introduction about the cell embedding algorithm and its running details in this study.**

1. scVI<sup>2</sup>. scVI is a Bayesian variational inference framework which has been widely used in fundamental analysis tasks. The latent space can be used for dimension reduction and visualization. We downloaded the source code of scVI from <https://github.com/YosefLab/scvi-tools> and implemented it with default hyper-parameters. As recommended, we applied scVI on raw count data.
2. DCA<sup>34</sup>. DCA is a denoising auto-encoder model with a zero-inflated negative binomial (ZINB) loss function. The bottleneck layer can be used for dimension reduction and visualization. We downloaded the source code of DCA from <https://github.com/theislab/dca> and implemented it with default hyper-parameters. As recommended, we applied DCA on raw count data.
3. ZIFA<sup>35</sup>. Based on the factor analysis framework, ZIFA is a latent variable model with an additional zero-inflation modulation layer. Using the EM algorithm for optimization, ZIFA can reduce the dimension of scRNA-seq data and further visualize the scRNA-seq profiles. We downloaded the source code of ZIFA from <https://github.com/epierson9/ZIFA> and implemented it with default hyper parameters. We assigned the dimension of the latent variables to 50. We applied ZIFA to log transformed data.
4. Factor analysis (FA)<sup>36</sup>. Factor analysis is a kind of general linear model to extract maximum common variance from all variables. We used an implementation from the scikit-learn package (<https://scikit-learn.org/>). We assigned the dimension of the latent variables to 50, which was the same as in ZIFA. We also applied FA to log transformed data.

## **Supplementary Section 6. Brief introduction about scRNA-seq simulation algorithm and running details in this study.**

We considered the following baseline methods in this study:

1. scGAN<sup>37</sup>. scGAN uses the wasserstein generative adversarial network framework (WGAN)<sup>38</sup> to generate a scRNA-seq profile. The discriminative network is trained to spot the difference between real and simulated scRNA-seq profiles and the generated network is trained to simulate realistic scRNA-seq profiles. scGAN uses the normalized count data as the training data and then generates the normalized count data. In this study, scGAN was trained on the cell-type specific scRNA-seq profile. The code of scGRN was downloaded from <https://github.com/imsb-uke/scGAN> and we used the default hyper-parameters.
2. cscGAN<sup>37</sup>. Similar to scGAN, cscGAN uses the Wasserstein Generative Adversarial Network (WGAN)<sup>38</sup> framework to generate scRAN-seq profiles. However, by leveraging conditional information about cell types and learning information from other cell types, cscGAN generates more realistic profiles compared to scGAN. In this study, cscGAN was trained on the whole training dataset. We downloaded its source code from <https://github.com/imsb-uke/scGAN> and used the default hyper-parameters.

## **Supplementary Section 7. Hyper-parameter sweeping.**

First, we showed that DeepSEM is robust to the change of hyper-parameters in most cases (**Supplementary Fig. 17**). For GRN prediction, we swept the key hyper-parameters including  $\alpha, \beta$ , and the number of epochs by leaving one cell line out and accessed the average EPR improvement during the GRN inference evaluation. We provided two examples of hyper-parameter sweeping in **Supplementary Fig. 17a-f**. The embedding visualization and clustering results showed that DeepSEM is robust to the changes of hyper-parameters.

Hyper-parameter  $\beta$  plays a more important role than  $\alpha$  and the number of epochs in terms of their influence on the performance of DeepSEM. An example of sweeping  $\beta$  on the Zeisel dataset is shown in **Supplementary Fig. 17g**.

For the scRNA-seq simulation, we found that hyper-parameter  $\beta$  is important to the scRNA-seq simulation performance. More specifically, if we assign  $\beta$  with a relatively small value (for example  $\beta = 0, 0.1$ ), the simulated scRNA-seq data will be more similar to the real cells but less diverse. However, assigning a large value to  $\beta$  (for example  $\beta = 1.5$ ) may make the convergence of DeepSEM unstable. Several examples are shown in **Supplementary Fig. 17h-j**. Overall, we used the vanilla VAE ( $\beta = 1$ ) in all visualization clustering and simulation experiments. We also provide a default hyper-parameter setting for GRN inference, which works well on most datasets. We summarize the default hyper-parameters used in this study in **Supplementary Table 8**.

## References

1. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
2. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
3. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
4. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
5. Hayashi, T. *et al.* Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* **9**, 619 (2018).
6. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood, The Journal of the American Society of Hematology* **128**, e20–e31 (2016).
7. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
8. Han, H. *et al.* TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* **5**, 11432 (2015).
9. Liu, Z.-P., Wu, C., Miao, H. & Wu, H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* **2015**, (2015).
10. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).

11. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, (2018).
12. Xu, H. *et al.* ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. *Database* **2013**, bat045 (2013).
13. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
14. Fang, R. *et al.* Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* **12**, 1337 (2021).
15. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
16. Chu, L.-F. *et al.* Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17**, 173 (2016).
17. Camp, J. G. *et al.* Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533–538 (2017).
18. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
19. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
20. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst* **3**, 385–394.e3 (2016).
21. Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24**, 593–607 (2016).
22. Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058 (2014).
23. Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates

- cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
24. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
  25. Tian, L. *et al.* Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nat. Methods* **16**, 479–487 (2019).
  26. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154 (2020).
  27. Chan, T. E., Stumpf, M. P. H. & Babtie, A. C. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Syst* **5**, 251–267.e3 (2017).
  28. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, (2010).
  29. Moerman, T. *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).
  30. Matsumoto, H. *et al.* SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33**, 2314–2321 (2017).
  31. Kim, S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Commun Stat Appl Methods* **22**, 665–674 (2015).
  32. Papili Gao, N., Ud-Dean, S. M. M., Gandrillon, O. & Gunawan, R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics* **34**, 258–266 (2018).
  33. Zappia, L., Phipson, B. & Oshlack, A. Splatter: simulation of single-cell RNA sequencing data. *Genome Biol.* **18**, 174 (2017).
  34. Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq

- denoising using a deep count autoencoder. *Nat. Commun.* **10**, 390 (2019).
35. Pierson, E. & Yau, C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol.* **16**, 241 (2015).
  36. Jolliffe, I. T. Principal Component Analysis and Factor Analysis. in *Principal Component Analysis* (ed. Jolliffe, I. T.) 115–128 (Springer New York, 1986).
  37. Marouf, M. et al. Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. *Nat. Commun.* **11**, 166 (2020).
  38. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. C. Improved Training of Wasserstein GANs. in *Advances in Neural Information Processing Systems 30* (eds. Guyon, I. et al.) 5767–5777 (Curran Associates, Inc., 2017).