Supplemental Material for

Cross-species cell-type assignment of single-cell RNA-seq by a heterogeneous graph neural network

Xingyan Liu[1,2†], Qunlun Shen[1,2†] and Shihua Zhang[1,2,3,4*]

[1]NCMIS, CEMS, RCSDS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;
[2]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China;
[3]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China;
[4]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou 310024, China.

[†]These authors contributed equally to this work.

*To whom correspondence should be addressed. Tel/Fax: +86 01 82541360; Email: zsh@amss.ac.cn.

Supplemental Figures

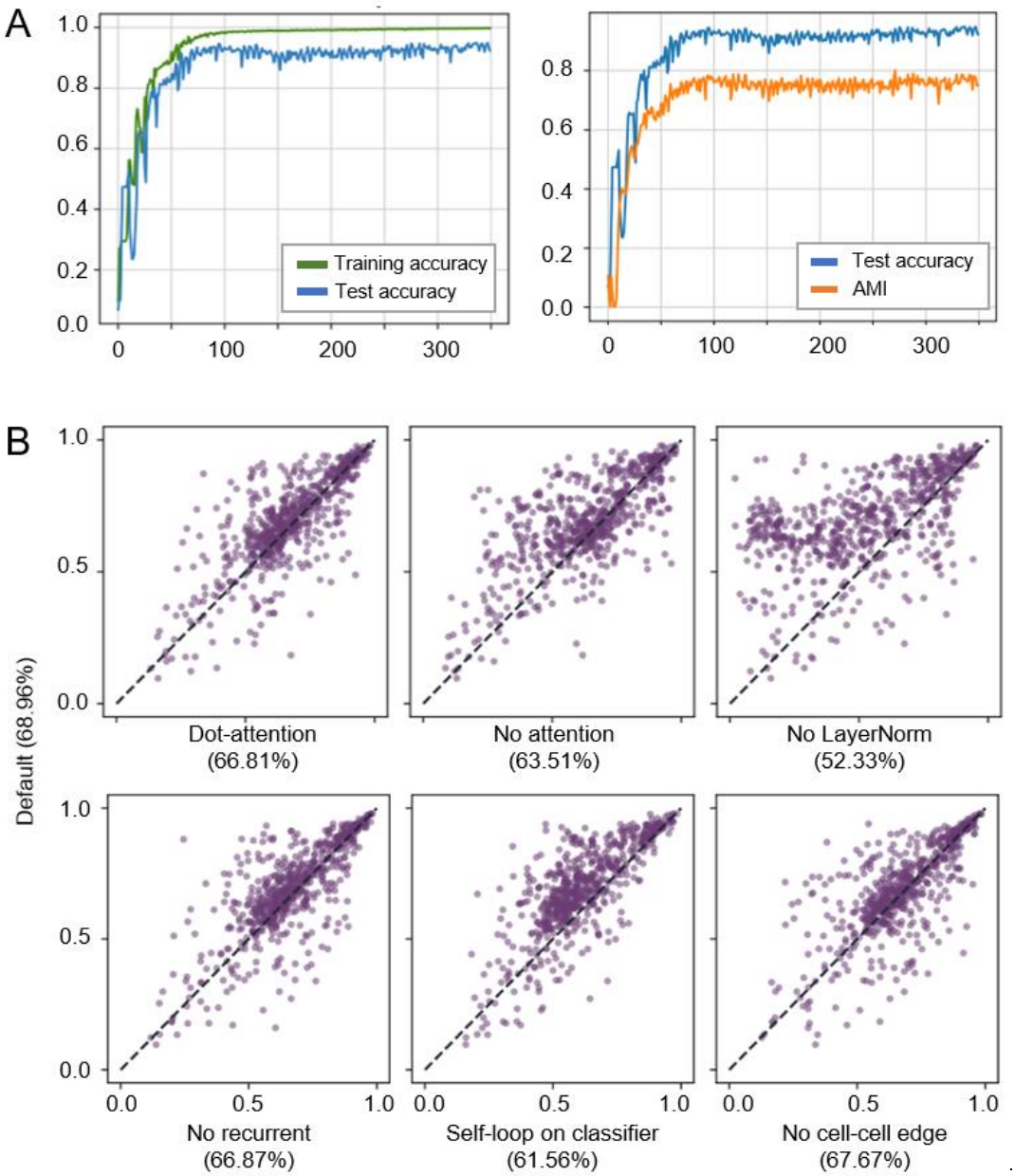


Figure S1. Performance evaluation during the training process and ablation experiments about six key factors of CAME. (**A**) (Left) Classification accuracy of training and test during the training process of CAME. (Right) The changes of test classification accuracy and the adjusted mutual information (AMI) between the predicted labels and pre-clustered ones of query cells during the training process of CAME. (**B**) Ablation experiments showing the importance of each part or key hyper-parameters adopted by CAME, where the values in parentheses represent the average accuracy across all the reference-query pairs.

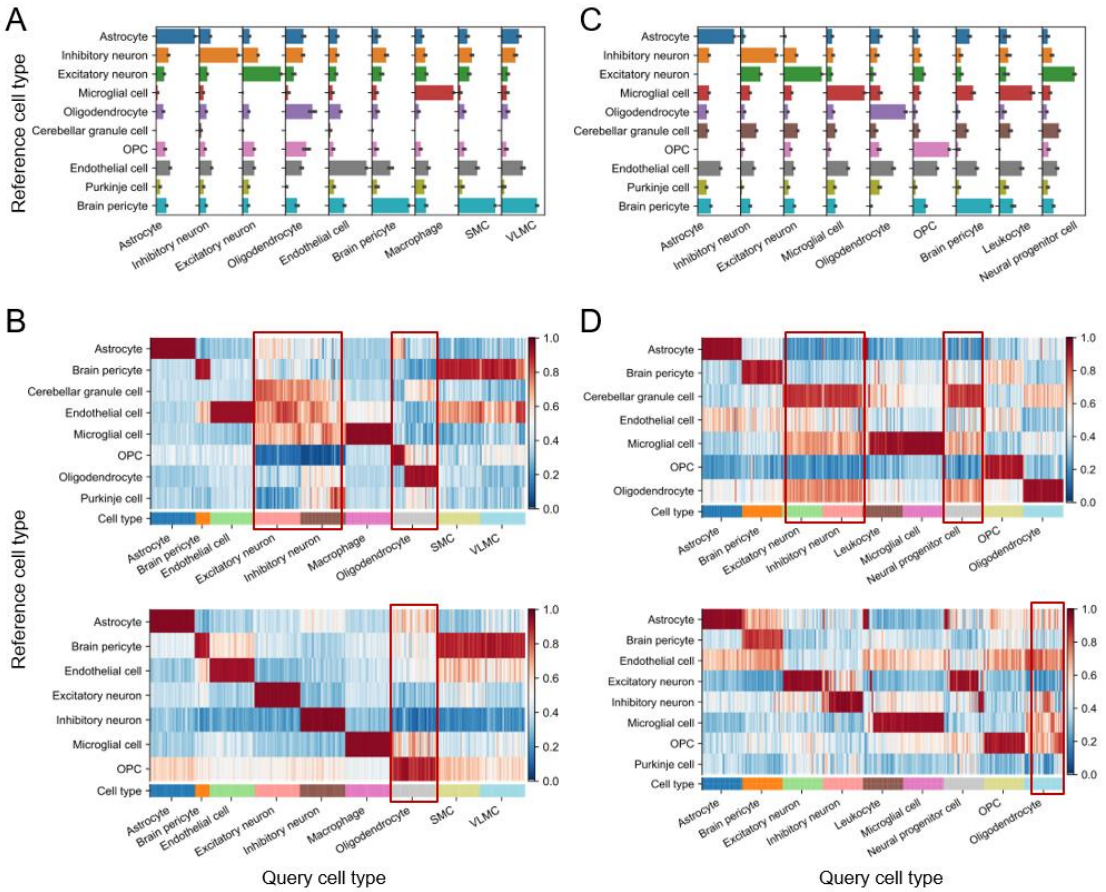
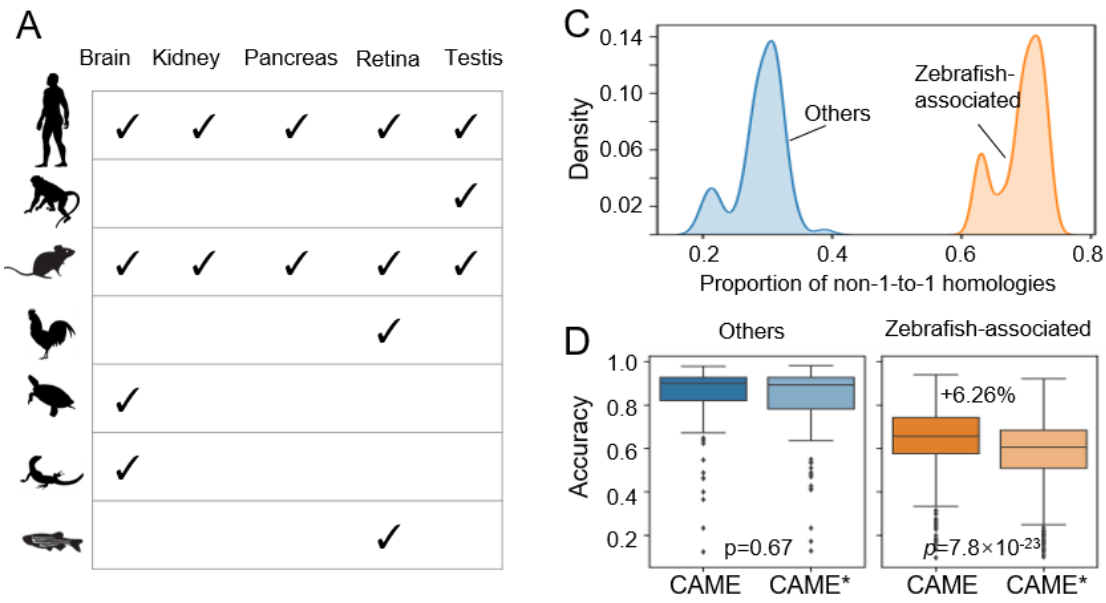


Figure S2. Multi-label cell-type prediction analysis. (**A, C**) Distribution of predicted probabilities of mouse brain cells (A) and turtle brain cells (C) with the human reference respectively, where each column shows a query cell type. The probabilities are calculated in a multi-label manner. (**B, D**) Reference type-removal experiments. Upper, both inhibitory and excitatory neurons were removed from the reference for two cases, and the neurons in the query data were predicted with much more uncertainty. Lower, predictions with the removal of oligodendrocytes from the reference for two cases, where most of the oligodendrocytes were assigned higher probability to its precursor type.



B

	Human	Macaque	Mouse	Chick	Turtle	Lizard	Zebrafish
Human		19717	16622	12299	13345	12291	9427
Macaque	84.96%		15922				
Mouse	69.63%	69.72%		12436	13517	12406	9548
Chick	75.38%		74.19%				8944
Turtle	71.82%		63.71%			12898	
Lizard	71.81%		67.82%		70.16%		
Zebrafish	45.36%		44.88%	51.70%			

Figure S3. Performance of CAME for cross-species cell-type assignment. (**A**) Overview of datasets. (**B**) Proportion (lower triangular) and number (upper triangular) of homologous genes between each pair of species. (**C**) Density plot of the proportions of the non-one-to-one homologous genes in HIGs (**Methods**) for two groups of dataset pairs. (**D**) Box plots showing the cell-type assignment performance of CAME and CAME* (using only one-to-one homologous genes) based on zebrafish-excluded and zebrafish-associated dataset pairs.

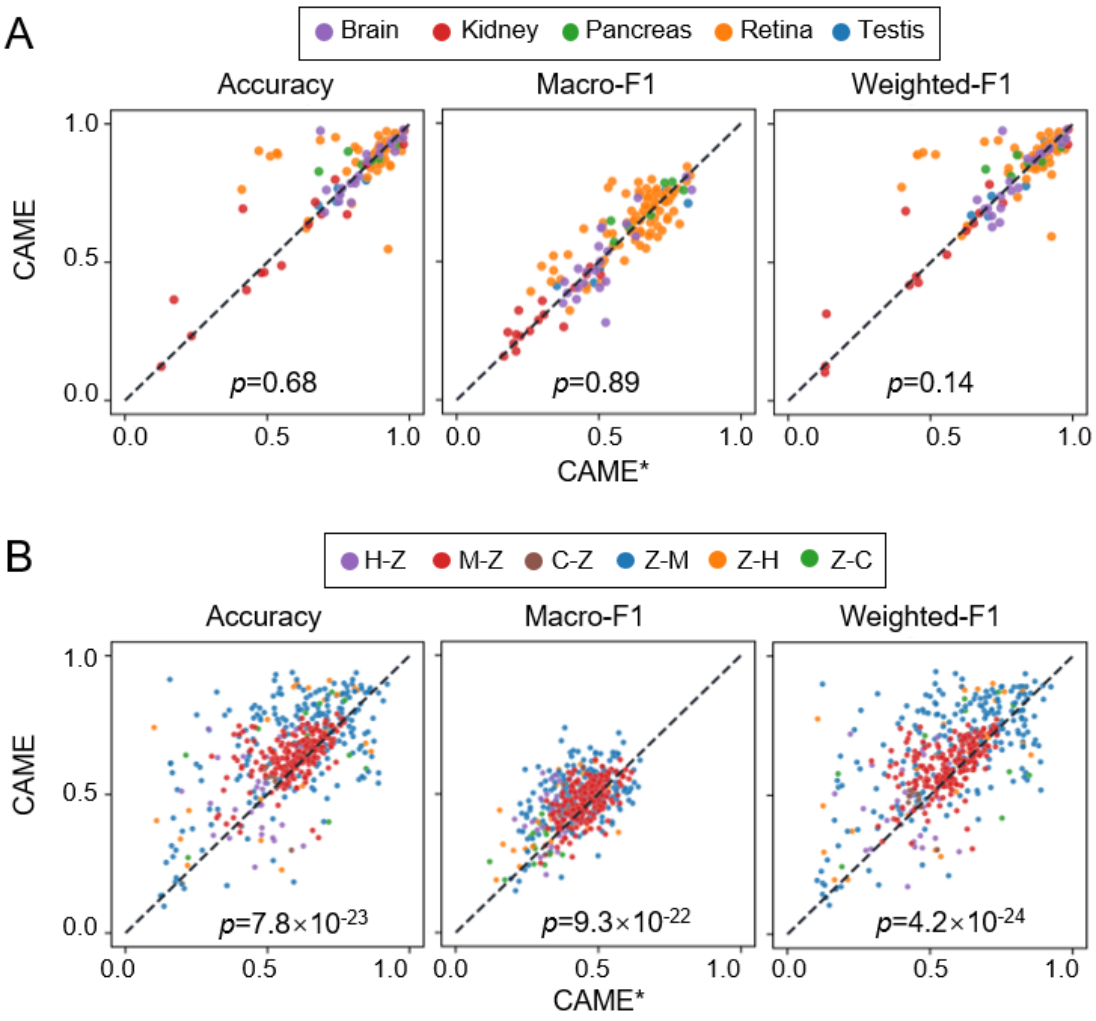


Figure S4. Statistics of gene homologies and comparison between CAME and CAME*. (**A**) Performance comparison between CAME and CAME* in terms of cross-species cell-type assignment for zebrafish-excluded dataset pairs. (**B**) Performance comparison between CAME and CAME* in terms of cross-species cell-type assignment for zebrafish-associated dataset pairs, where “X-Y” indicates the label transfer from species X to Y. C: chick, H: human, M: mouse, Z: zebrafish.

A

Method	Index	Brain	Kidney	Pancreas	Retina	Testis
SAMap	mean	0.0595	0.2212	0.1377	0.0644	0.0944
	median	0.0299	0.1455	0.1485	0.0182	0.0561
Scmap	mean	0.6884	0.4705	0.7278	0.6101	0.7457
	median	0.7139	0.489	0.7739	0.7003	0.758
ItClust	mean	0.5742	0.3036	0.397	0.6437	0.6181
	median	0.6338	0.2243	0.0693	0.7158	0.6852
SCN	mean	0.6466	0.44	0.5611	0.6138	0.7028
	median	0.6714	0.5262	0.5962	0.5974	0.7029
Cell BLAST	mean	0.6638	0.307	0.8059	0.6423	0.6434
	median	0.6808	0.288	0.848	0.6381	0.6894
SciBet	mean	0.8275	0.5173	0.618	0.7258	0.749
	median	0.8486	0.48	0.6139	0.7251	0.7345
Seurat	mean	0.72	0.4891	0.7321	0.8211	0.7505
	median	0.7391	0.471	0.7396	0.8454	0.7457
CAME*	mean	0.8352	0.5767	0.8544	0.8719	0.7984
	median	0.8176	0.5972	0.8636	0.9099	0.7978
CAME	**mean**	**0.8413**	**0.593**	**0.8846**	**0.8936**	**0.8022**
	median	**0.8401**	**0.6549**	**0.8749**	**0.915**	**0.7821**

B

Method	Reference	Chick	Human	Mouse	Zebrafish		
	Query	Zebrafish			Chick	Human	Mouse
SAMap	mean	0.2373	0.0874	0.0898	0.0513	0.0608	0.075
	median	0.1588	0.0679	0.0762	0.0185	0.0015	0.0069
Scmap	mean	0.2105	0.1631	0.3542	0.1496	0.2971	0.1924
	median	0.2086	0.1605	0.3512	0.0977	0.2739	0.1387
ItClust	mean	0.4754	0.4595	0.4395	0.2487	0.1352	0.2505
	median	0.4755	0.4482	0.4743	0.1439	0.0693	0.1587
SCN	mean	0.2687	0.1862	0.2791	0.3217	0.3357	0.3374
	median	0.2653	0.1808	0.2827	0.34	0.3877	0.3477
Cell BLAST	mean	0.5496	0.3457	0.3691	0.5731	0.3622	0.465
	median	0.5438	0.3495	0.3807	0.5479	0.3524	0.4663
SciBet	mean	0.428	0.4137	0.4799	0.5234	0.3847	0.4329
	median	0.4163	0.425	0.4807	0.5262	0.3882	0.4286
Seurat	mean	0.3234	0.3736	0.5533	0.492	0.5167	0.5511
	median	0.2848	0.3688	0.5657	0.5024	0.5376	0.548
CAME*	mean	0.5525	0.4726	0.5937	0.6201	0.5605	0.5935
	median	0.5326	0.48	0.6049	0.6621	0.6293	0.6268
CAME	mean	**0.5679**	**0.5044**	**0.6385**	**0.6685**	**0.6466**	**0.6808**
	median	**0.5918**	**0.5153**	**0.6405**	**0.6767**	**0.6872**	**0.7307**

Figure S5. Performance comparison of classification accuracy between CAME and other benchmarking methods. (**A**) Mean and median classification accuracies of CAME and other benchmarking methods on 139 pairs of cross-species scRNA-seq datasets (zebrafish excluded). Each column corresponds to a specific tissue. (**B**) Mean and median classification accuracies of CAME and baseline methods on the cross-species scRNA-seq data associated with zebrafish. Each column corresponds to different species pairing with each other respectively.

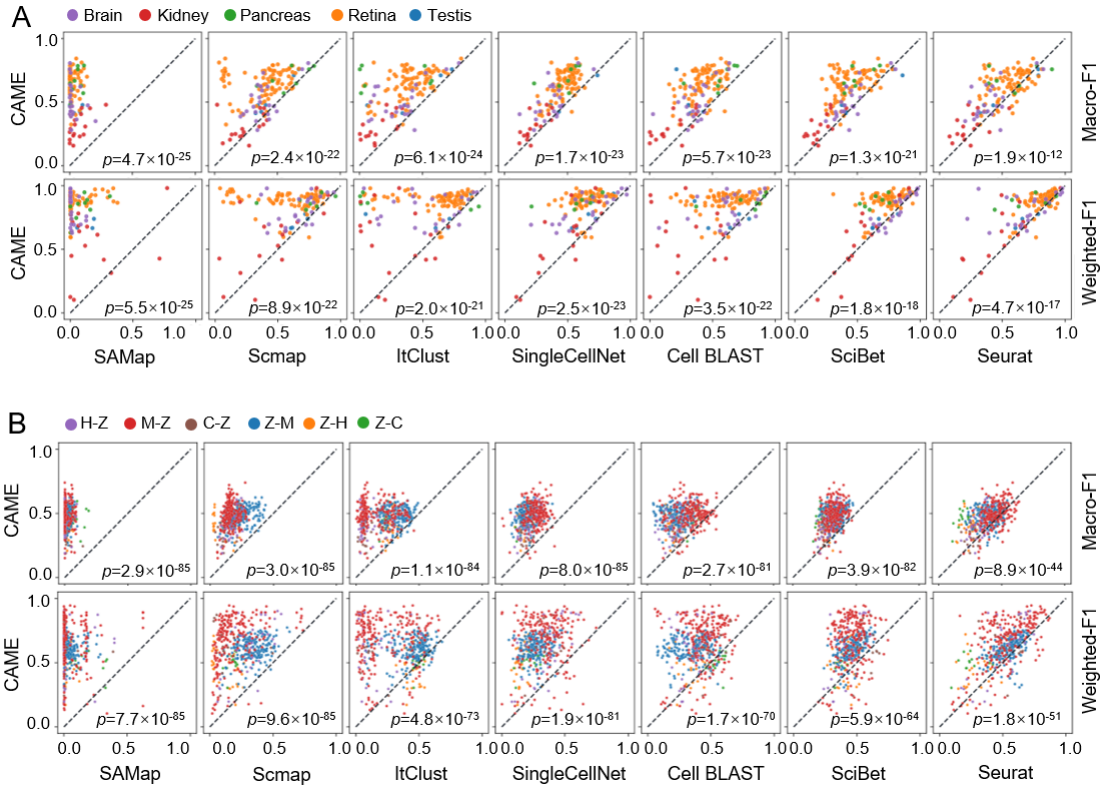


Figure S6. Performance comparison of cell-type classification of CAME and six benchmarking methods. (**A**) Pairwise comparison of the cell-typing performance of CAME and other six benchmarking methods on 139 pairs of cross-species scRNA-seq datasets in terms of macro-F1 and weighted-F1 (zebrafish-excluded**)**. (**B**) Pairwise comparison of the cell-typing performance of CAME and other six benchmarking methods on 510 pairs of cross-species scRNA-seq datasets in terms of macro-F1 and weighted-F1 (zebrafish-associated). The notation “X-Y” indicates that X is the reference and Y is the query. H: Human, M: Mouse, C: Chick, Z: Zebrafish.

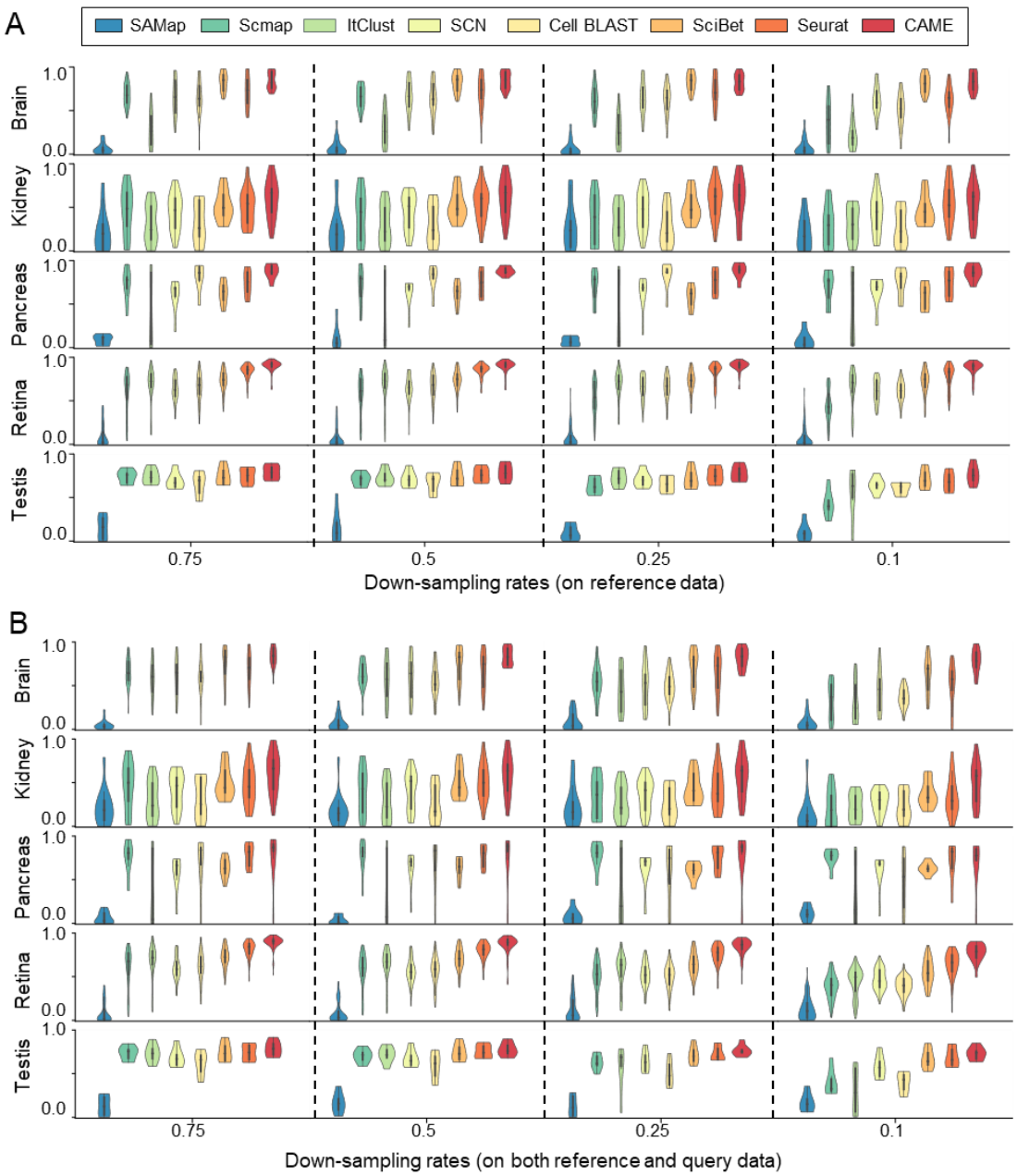


Figure S7. Robustness to inconsistent and insufficient sequencing depths (zebrafish-excluded). (**A**) Violin plots of classification accuracies of CAME and six benchmarking methods with different down-sampling rates (0.75, 0.5, 0.25, 0.1) for read counts on the reference data. (**B**) Violin plots of classification accuracies of CAME and six benchmarking methods with different down-sampling rates (0.75, 0.5, 0.25, 0.1) on both reference and query data.

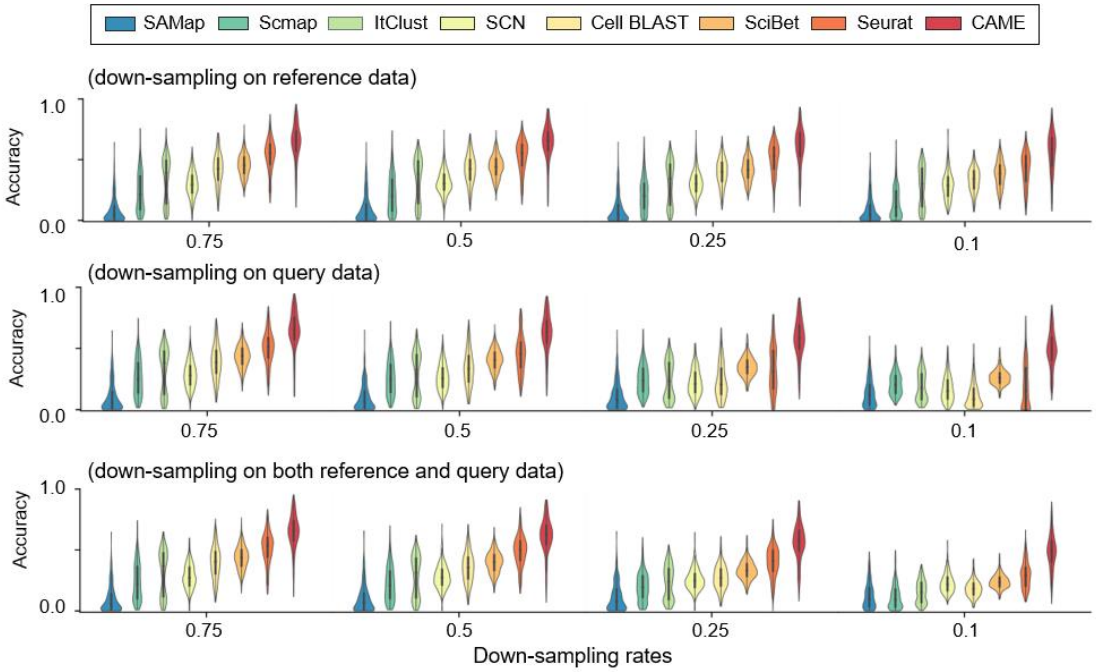


Figure S8. Robustness to inconsistent and insufficient sequencing depths (zebrafish-associated). Violin plots of classification accuracies of CAME and six benchmarking methods with different down-sampling rates (0.75, 0.5, 0.25, 0.1) for read counts on reference, query, and both datasets respectively.

A

		Brain	Kidney	Pancreas	Retina	Testis
hvg1000+deg50	mean	0.8386	0.5939	0.8159	0.9038	0.8002
	median	0.8118	0.6278	0.8563	0.9127	0.7898
hvg1000+deg100	mean	0.8414	0.6506	0.7896	0.8981	0.7492
	median	0.8280	0.6647	0.8331	0.9127	0.7620
hvg2000+deg100	mean	0.8361	0.5883	0.7866	0.8906	0.7787
	median	0.8243	0.6279	0.8092	0.9058	0.7723
hvg3000+deg50	mean	0.8323	0.5474	0.8169	0.9028	0.7904
	median	0.8306	0.6210	0.8570	0.9108	0.7933
hvg3000+deg100	mean	0.8345	0.5721	0.8187	0.8901	0.7628
	median	0.8310	0.5637	0.8414	0.9090	0.7629
Seurat hvg2000 +deg50	mean	0.8331	0.5429	0.8297	0.9031	0.7644
	median	0.8047	0.5772	0.8640	0.9169	0.7847
256 hidden units	mean	0.8529	0.5572	0.8345	0.9097	0.7805
	median	0.8457	0.5931	0.8496	0.9221	0.8053
512 hidden units	mean	0.8479	0.6078	0.8624	0.9151	0.8139
	median	0.8616	0.6806	0.8802	0.9207	0.8080
1024 hidden units	mean	0.8398	0.5255	0.8498	0.9004	0.7766
	median	0.8445	0.5576	0.8705	0.9146	0.8041
three hidden layers	mean	0.8377	0.5678	0.8295	0.9003	0.7803
	median	0.8565	0.6326	0.8553	0.9167	0.7808
default	mean	0.8413	0.5930	0.8315	0.8936	0.8022
	median	0.8401	0.6549	0.8590	0.9150	0.7821

B

	Reference	Chick	Human	Mouse	Zebrafish		
	Query	Zebrafish			Chick	Human	Mouse
hvg1000+deg50	mean	0.5917	0.5288	0.6295	0.5665	0.6719	0.6734
	median	0.5879	0.5562	0.6354	0.6335	0.7124	0.7408
hvg1000+deg100	mean	0.5742	0.5383	0.6399	0.5889	0.6634	0.6567
	median	0.5879	0.5377	0.6539	0.6508	0.7143	0.7297
hvg2000+deg100	mean	0.5816	0.5506	0.6399	0.6049	0.6543	0.6554
	median	0.5687	0.5405	0.6444	0.6432	0.6872	0.7094
hvg3000+deg50	mean	0.5671	0.4934	0.6453	0.6324	0.6662	0.6630
	median	0.5781	0.4921	0.6502	0.6603	0.7063	0.7157
hvg3000+deg100	mean	0.5860	0.4910	0.6300	0.5795	0.6326	0.6501
	median	0.5868	0.5202	0.6368	0.6583	0.6925	0.7044
Seurat hvg2000+deg50	mean	0.5554	0.5250	0.6255	0.6729	0.6989	0.6856
	median	0.5718	0.5225	0.6332	0.6678	0.7389	0.7228
256 units	mean	0.6039	0.5477	0.6476	0.6401	0.7003	0.7084
	median	0.5978	0.5770	0.6528	0.6715	0.7446	0.7606
512 units	mean	0.5928	0.5947	0.6558	0.6402	0.6241	0.7168
	median	0.5777	0.6114	0.6553	0.7132	0.7238	0.7971
1024 units	mean	0.5912	0.5515	0.6424	0.7000	0.5940	0.6945
	median	0.5816	0.5427	0.6411	0.7457	0.6349	0.7717
three hidden layers	mean	0.5895	0.5273	0.6449	0.6044	0.6362	0.6464
	median	0.5883	0.5432	0.6504	0.6576	0.6387	0.7105
default	mean	0.5679	0.5044	0.6385	0.6685	0.6466	0.6808
	median	0.5918	0.5153	0.6405	0.6767	0.6873	0.7307

Figure S9. Performance comparison of classification accuracy of CAME under different gene selection strategies and hyperparameters. (A) Mean and median classification accuracies of CAME of CAME under different gene selection strategies and hyperparameters on 139 pairs of cross-species scRNA-seq datasets (zebrafish excluded). Each column corresponds to a specific tissue. (B) Mean and median classification accuracies of CAME under different gene selection strategies and hyperparameters on the cross-species scRNA-seq data associated with zebrafish. Each column corresponds to different species pairing with each other respectively.

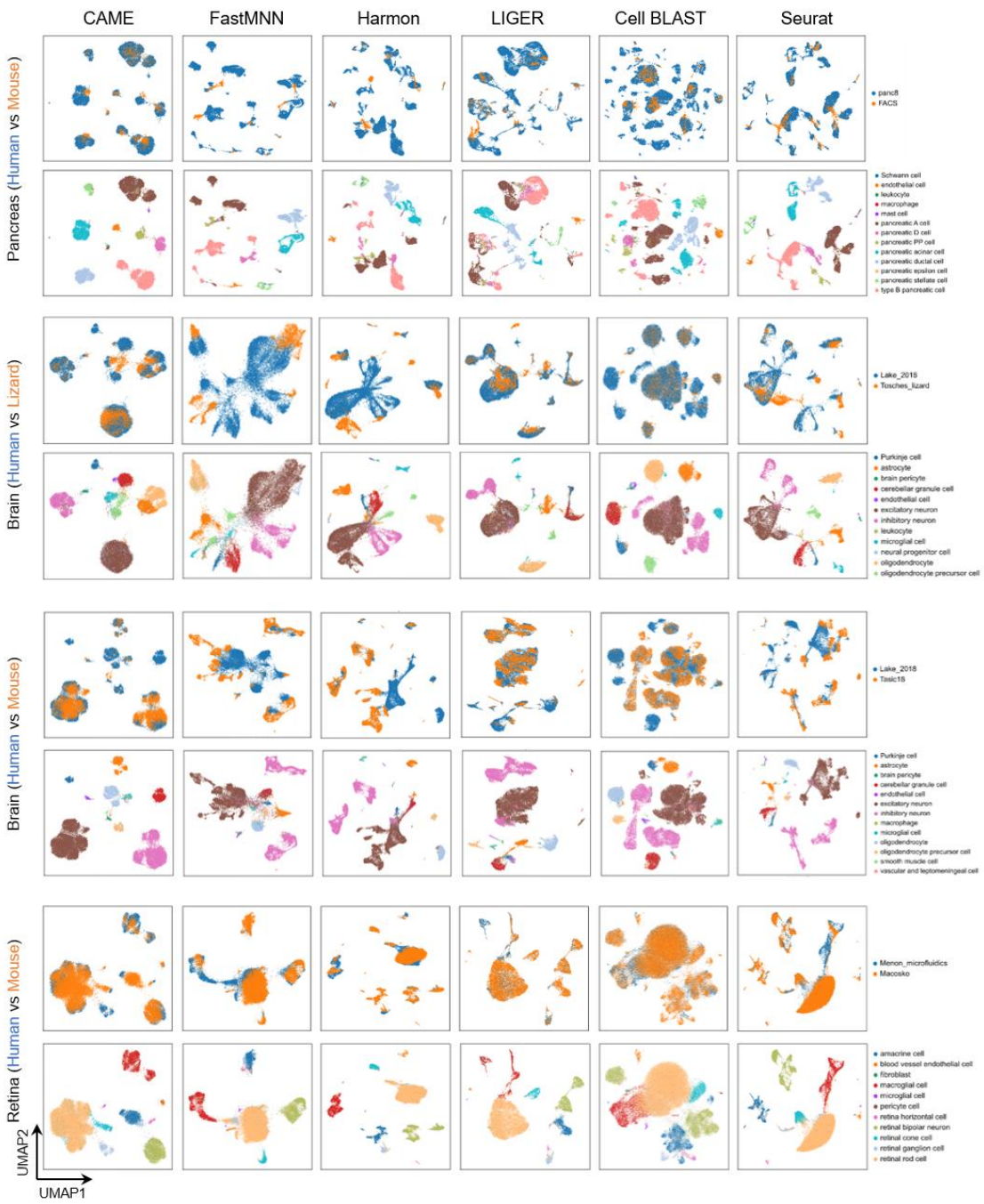


Figure S10. Cross-species alignment of homologous cell types by CAME and other five methods. UMAP plots of cell embeddings based on CAME and five typical data integration methods including FastMNN, Harmony, LIGER, Cell BLAST, and Seurat on four pairs of datasets, including human and mouse pancreas data, ("panc8" and "FACS"), human and lizard brains ("Lake_2018" and "Tosches_lizard"), human and mouse brains ("Lake_2018" and "Tasic18"), human and mouse retinal cells ("Menon_microfluidics" and "Macosko"). Cells are colored by their dataset identities (the first row) or cell types (the second row) respectively.

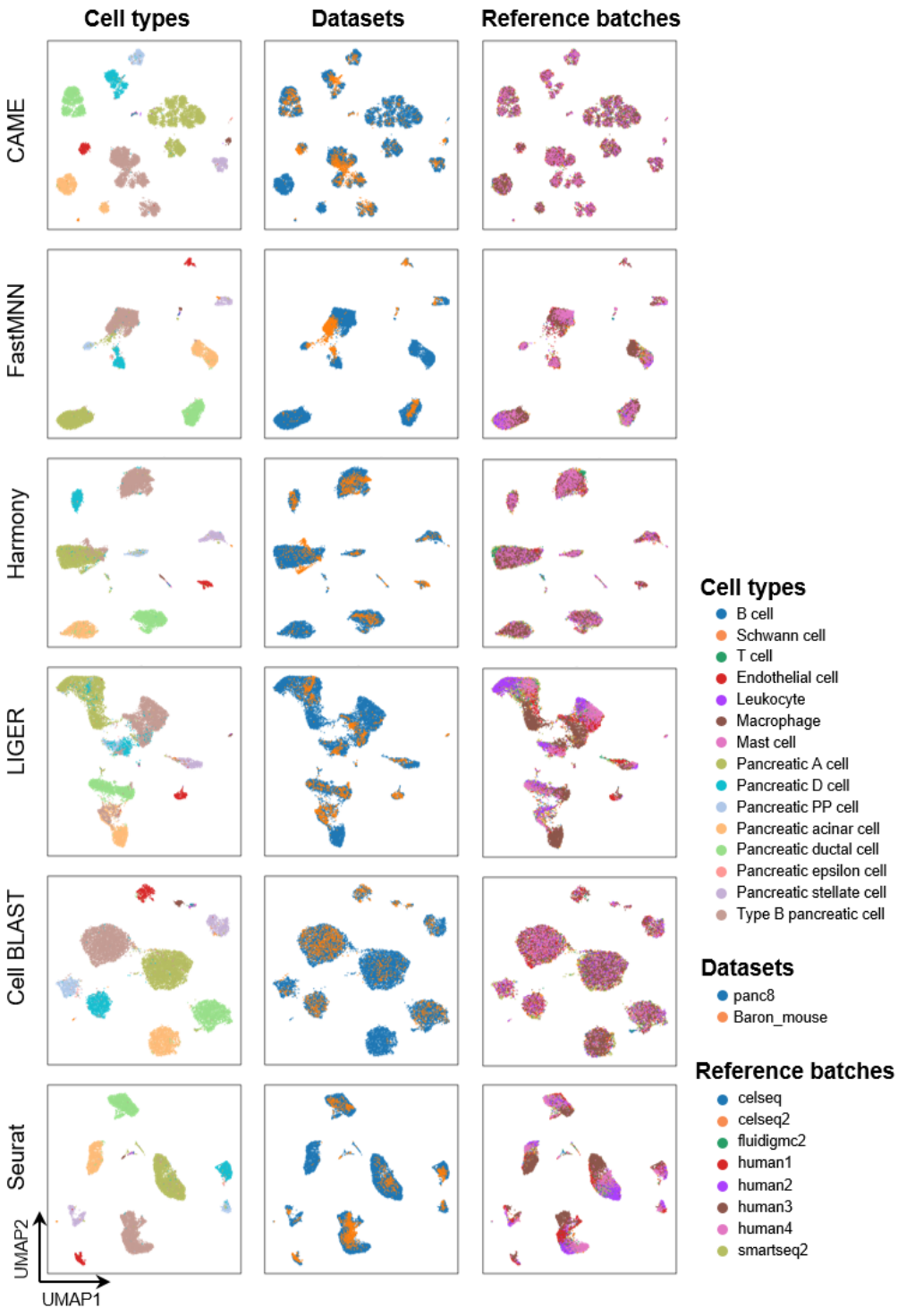


Figure S11. Robustness to multi-reference batch effects with given batch labels. UMAP plots of cell embeddings based on the output of CAME and five typical data integration methods including FastMNN, Harmony, LIGER, Cell BLAST, and Seurat. The reference datasets are the human pancreatic scRNA-seq data from eight batches by five different platforms and the query is from mouse pancreas cells. Cells are colored by their cell types (the first column) and dataset identities (the second column) respectively. The UMAP plots on the third column show the reference datasets, colored by batch identities.

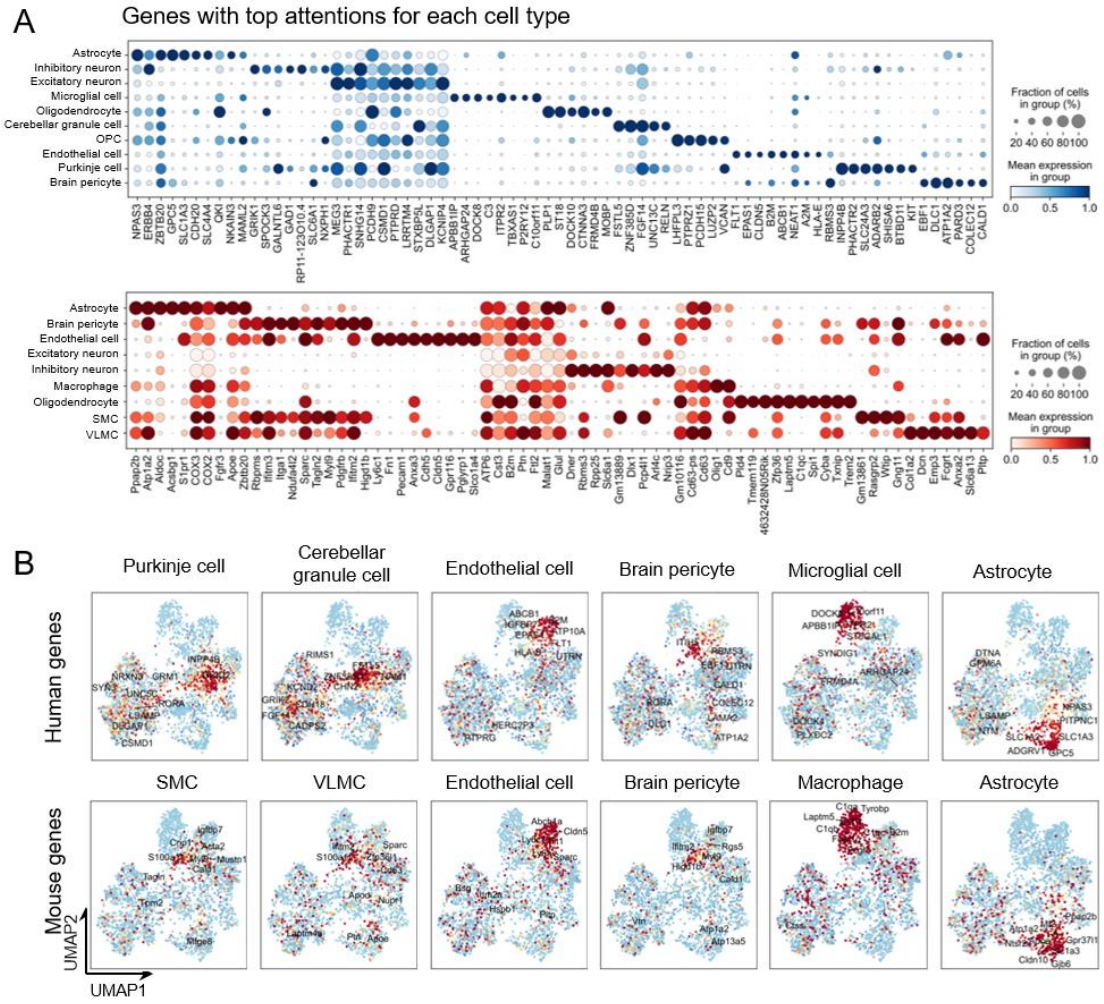


Figure S12. Genes of top attentions and gene embeddings of human and mouse brains. (**A**) Dot plots showing the expressions of genes with top attentions in the corresponding cell types. (**B**) UMAP plots of genes showing the average expression patterns (z-scored across cell-types for each gene) of the major types for human and mouse brain, where each point represents a gene and the color of each scatter is scaled by the expression level of that cell type in each gene. SMC is short for “smooth muscle cell”, and VLMC is short for “vascular and leptomeningeal cell”.

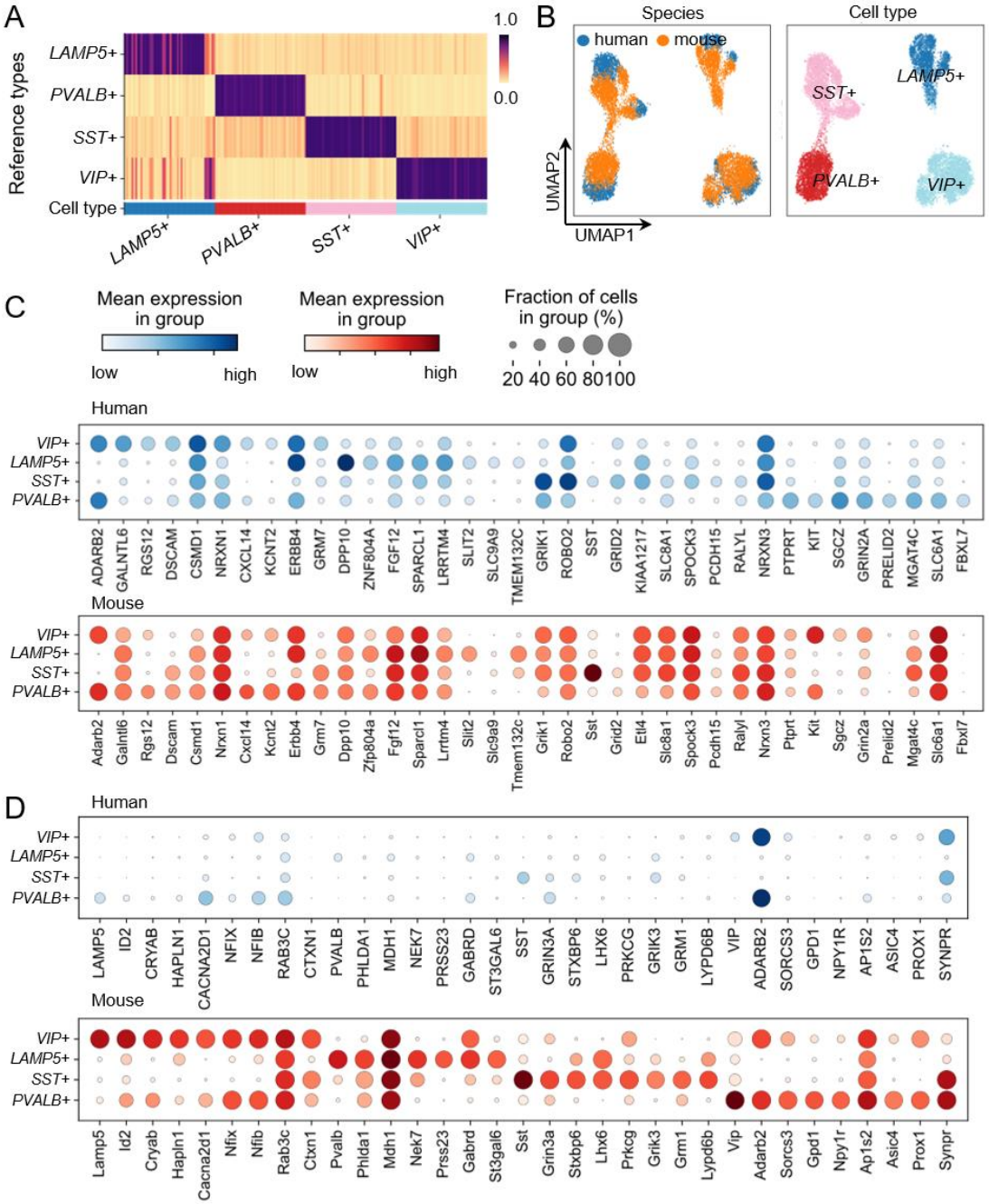


Figure S13. Comparisons of the four interneuron subtypes between human and mouse brains. (**A**) The predicted cell-type probabilities for each cell (each column) in the scRNA-seq data of the mouse brain, taking the gene expressions of the human brain as the reference (each row shows a cell type in human data). (**B**) The UMAP plots of cell embeddings by CAME, colored by datasets (left) or cell type (right). (**C**) The gene expression profiles of each inhibitory neural subtype in humans (upper, blue dots) or mice (lower, red dots), where the genes were detected as the top DEGs of each human interneuron subtype. (**D**) Similar to (C), the top DEGs of each mouse interneuron subtype are shown.

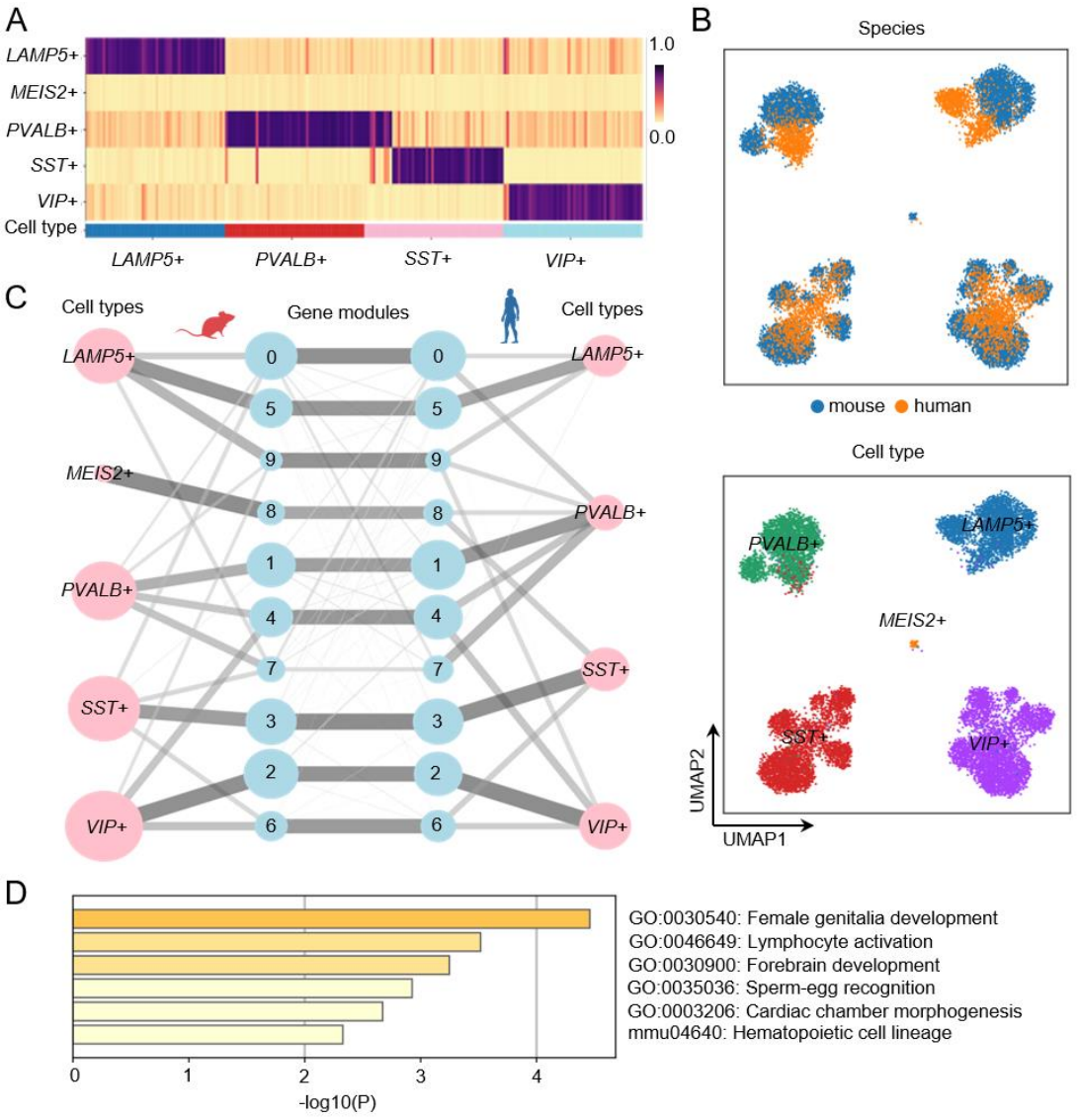


Figure S14. Explorations of the Meis2 inhibitory cell which is mouse-specific in cell-type transferring. (**A**) The predicted cell-type probabilities for each cell (each column) in the scRNA-seq data of the human cortex, taking the gene expressions of the human brain as the reference (each row shows a cell type in human data). (**B**) The UMAP plots of cell embeddings by CAME, colored by datasets (left) or cell type (right). (**C**) Abstracted graph of the heterogeneous cell-gene graph. Each node represents a cell type (pink) or a gene module (light blue). The size of a node is scaled by the number of single cells in that type or the number of genes in that gene module. The width of an edge is scaled by either the normalized mean expression levels of a cell type in the connected gene module or the conservancy of inter-species gene modules based on the gene embeddings learned by CAME. (**D**) Enrichment analysis of gene module 8 which is related to mouse-specific Meis2 cells.

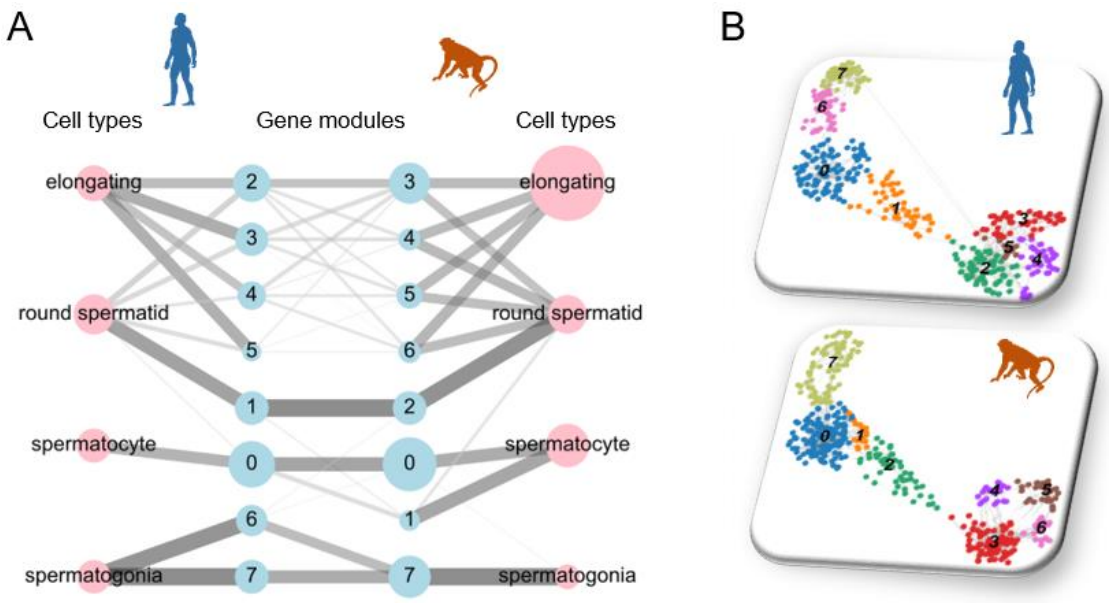


Figure S15. Separate module extraction based on gene embeddings by CAME on human and macaque testicular data respectively. (**A**) The abstracted heterogeneous cell-gene graph. (**B**) Gene modules detected by separate module extraction of genes from humans (above) and macaques (below). Non-one-to-one homologous gene pairs and species-specific genes were included.

A

Species1	Species2	Set	Amacrine cell	Blood vessel endothelial cell	Macroglial cell	Microglial cell	Retina horizontal cell	Retinal bipolar neuron	Retinal cone cell	Retinal ganglion cell	Retinal rod cell	Pericyte cell
Zebrafish	Human	common1v1	**1**	**17**	**6**	**57**	**4**	**1**	**0**	**4**	**3**	
		common1	4	34	16	75	9	3	11	15	17	
		common2	3	29	16	75	9	3	9	13	11	
		private1	117	188	152	249	132	66	129	121	47	
		private2	27	102	124	162	91	52	82	67	87	
Zebrafish	Mouse	common1v1	**12**	**12**	**9**	**25**	**3**	**1**	**4**	**10**	**5**	**5**
		common1	19	24	28	44	10	5	16	28	20	10
		common2	18	22	26	43	9	5	12	24	14	10
		private1	117	188	152	249	132	66	129	121	47	150
		private2	66	163	139	190	126	68	23	126	82	150
Zebrafish	Chick	common1v1	**10**		**7**			**2**	**11**	**5**	**10**	
		common1	21		18			3	31	27	20	
		common2	19		17			3	27	21	16	
		private1	117		152			66	129	121	47	
		private2	65		133			90	127	117	86	
Human	Mouse	common1v1	**10**	**28**	**32**	**46**	**18**	**14**	**4**	**12**	**35**	
		common1	10	32	34	53	18	14	5	13	35	
		common2	10	32	34	51	18	14	5	13	35	
		private1	27	102	124	162	91	52	82	67	87	
		private2	66	163	139	190	126	68	23	126	82	

B

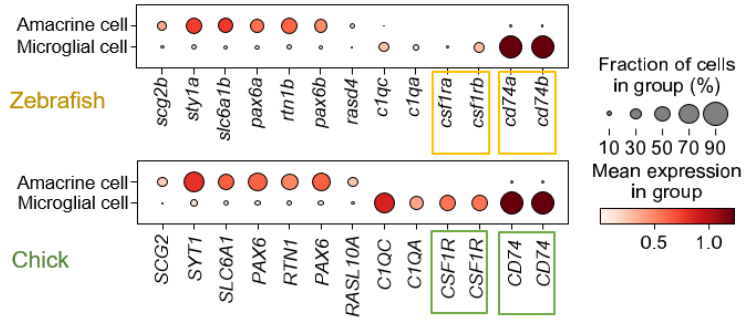
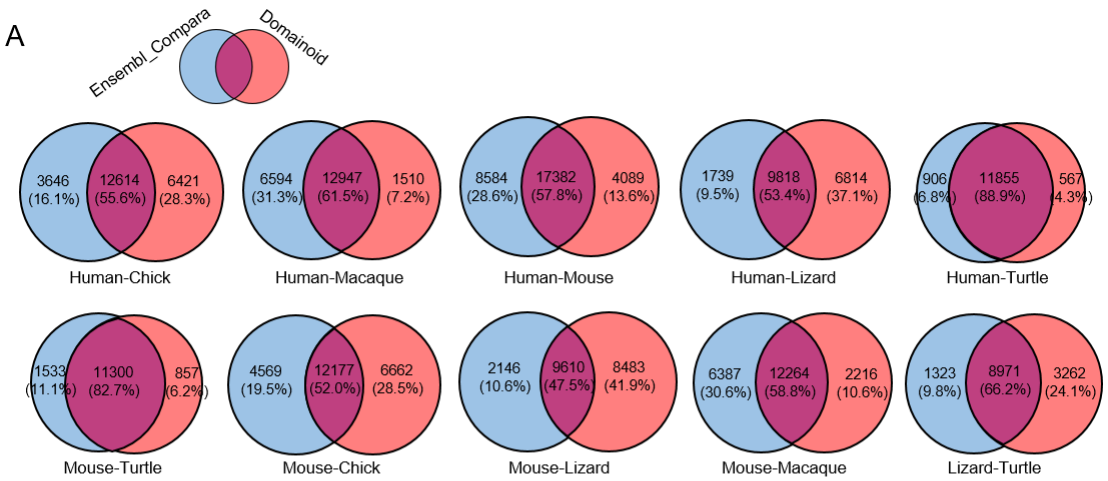


Figure S16. Statistics of common and private DEGs and gene expression comparison between zebrafish and chick. (**A**) Table of the number of cross-species overlapped DEGs in homologous cell types. The blank means that the cell type was not detected or annotated in either dataset. “common1v1”: commonly DEGs that are one-to-one homologous in the species pair. “common1”, “common2”: common DEGs that are homologous in the species pair. “private1”, “private2”: DEGs that are only occurred in species1 or species2, respectively. (**B**) Dot plot showing the common DEGs or canonical markers of amacrine cells and microglial cells in zebrafish and chick retinas respectively. Note that the significant marker *CSF1R* and *CD74* of chick microglial cells have more than one homology in zebrafish.



B

		H-M/M-H	H-C/C-H	H-E/E-H	H-L/L-H	H-T/T-H	M-C/C-M	M-L/L-M	M-T/T-M	M-E/E-M	T-L/L-T
Ensembl_Compara	Mean	0.7310	0.8061	0.9118	0.8162	0.8363	0.9059	0.8298	0.8079	0.7515	0.9491
	Median	0.8606	0.8133	0.9118	0.8162	0.8363	0.9025	0.8129	0.8090	0.7515	0.9491
Domainoid	Mean	0.7300	0.8354	0.8688	0.8181	0.8438	0.9045	0.8220	0.7938	0.7439	0.9639
	Median	0.8791	0.8309	0.8688	0.8181	0.8438	0.9077	0.8096	0.7462	0.7439	0.9639

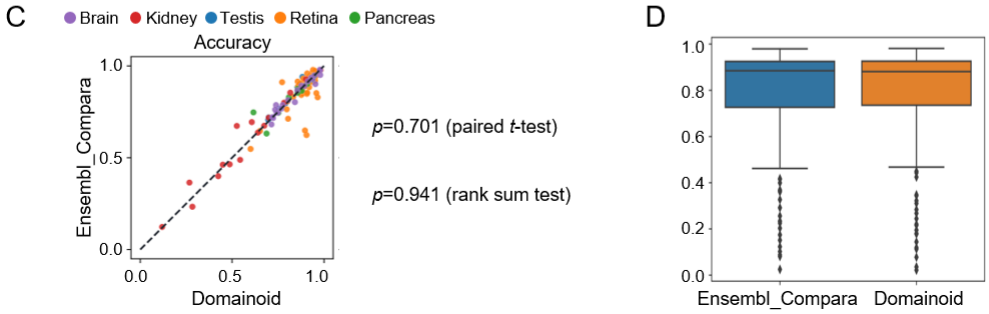


Figure S17. Performance comparison of CAME with homologous gene relationships inferred by Ensembl_Compara and Domainoid. (A) Venn plot showing the intersection and difference set of the homologous genes inferred by Ensembl_Compara and Domainoid. **(B)** Mean and median classification accuracies of CAME with homologous gene relationships inferred by Ensembl_Compara and Domainoid on cross-species scRNA-seq datasets. The notation "X-Y" indicates that X is the reference and Y is the query. H: Human, M: Mouse, C: Chick, T: Turtle, L: Lizard, E: Macaque. **(C)** Performance comparison of CAME in terms of cell-typing accuracy where each point represents a pair of cross-species datasets and is colored by different tissues. **(D)** Performance comparison of CAME in terms of cell-typing accuracy with a boxplot.

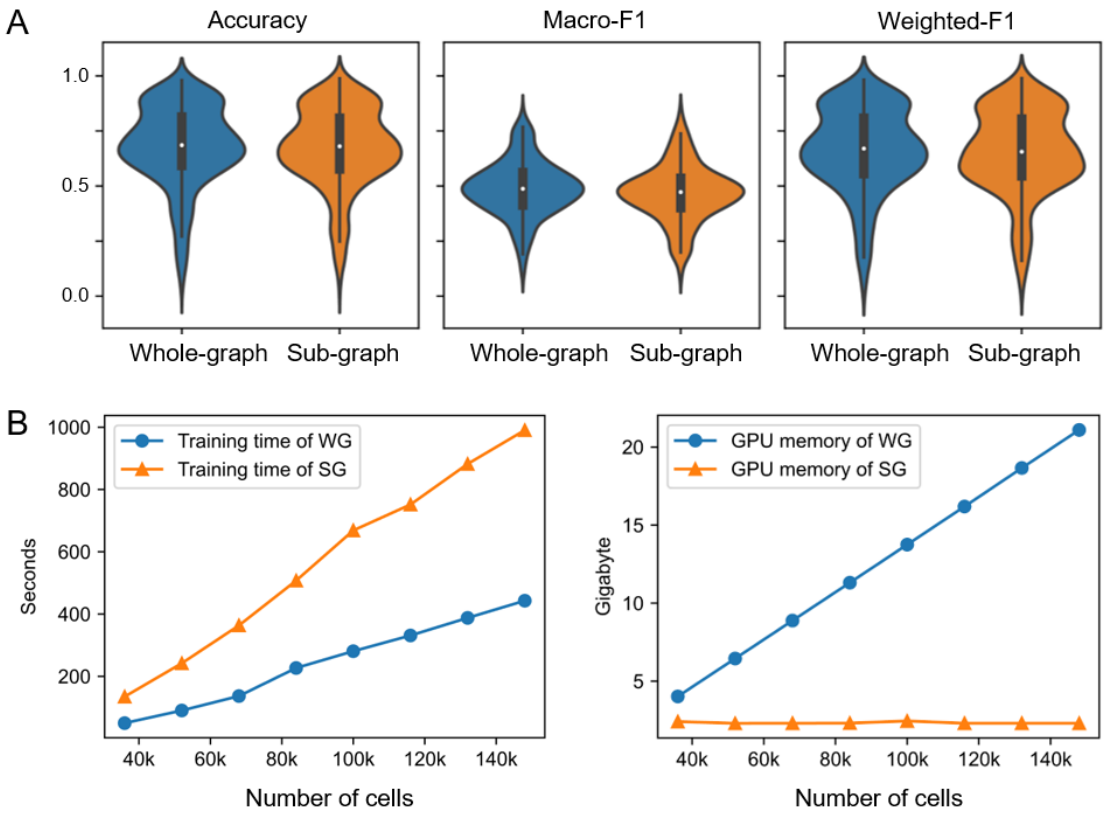


Figure S18. Performance comparison for model training on the whole graph (WG) and sub-graphs (SG) (i.e., mini-batches). The reference data and query data are up-sampled from the human and mouse pancreas dataset (the number of cells in reference and query is equal and the number of total cells is ranging from 20000 to 600000). To make a fair comparison, the genes were calculated from raw datasets. (**A**) Cell-typing performance comparison between the WG and SG training strategies based on 649 pairs of cross-species datasets in terms of Accuracy, Macro-F1, and Weighted-F1 scores. (**B**) The training time and peak graphical memory usage are showed for a different number of total cells.

Supplemental Tables

Table S1. scRNA-seq datasets used for benchmarking studies.

Dataset	Organism	Organ	Platform	Number of cells
Baron_human [1]	Homo sapiens	Pancreas	inDrop	8,569
Baron_mouse [1]	Mus musculus	Pancreas	inDrop	1,886
panc8 (SeuratData)	Homo sapiens	Pancreas	mixed	14,890
FACS [2]	Mus musculus	Pancreas	Smart-seq2	1,328
testis_human [3]	Homo sapiens	Testis	Drop-seq	13,837
testis_monkey [3]	Macaca fascicularis	Testis	Drop-seq	21,574
testis_mouse [3]	Mus musculus	Testis	Drop-seq	34,633
Lake_2018 [4]	Homo sapiens	Brain	snDrop-seq	35,289
Tasic18 [5]	Mus musculus	Brain	Smart-seq2	22,614
Campbell [6]	Mus musculus	Brain	Drop-seq	20,921
Chen [7]	Mus musculus	Brain	Drop-seq	12,089
Quake_Smart-seq2_Brain _Non-Myeloid [2]	Mus musculus	Brain Non-Myeloid	Smart-seq2	3,401
Tosches_lizard [8]	Pohona vitticeps	Brain	Drop-seq	4,187
Tosches_turtle [8]	Trachemys scripta elegans	Brain	Drop-seq	18,664
Adam [9]	Mus musculus	Kidney	Drop-seq	3,660
Karaiskos_mouse [10]	Mus musculus	Kidney	Drop-seq	12,954
Park [11]	Mus musculus	Kidney	10x	43,745
Quake_10x_Kidney [2]	Mus musculus	Kidney	10x	2,781
Wu_human [12]	Homo sapiens	Kidney	10x_snRNA-seq	4,298
Young [13]	Homo sapiens	Kidney	10x	5,685
Hochane [14]	Homo sapiens	Kidney	10x	14,606
Alemany_Kidney [15]	Danio rerio	Kidney	ScarTrace	4,415
Menon_microfluidics [16]	Homo sapiens	Retina	microfluidics	20,091
Menon_seqwell [16]	Homo sapiens	Retina	Seq-Well	3,014
Macosko [17]	Mus musculus	Retina	Drop-seq	44,808
mouse_NMDA_3hr [18]	Mus musculus	Retina	10x	7,442
mouse_NMDA_6hr [18]	Mus musculus	Retina	10x	6,202
mouse_NMDA_12hr [18]	Mus musculus	Retina	10x	4,225
mouse_NMDA_24hr [18]	Mus musculus	Retina	10x	4,194
mouse_NMDA_36hr [18]	Mus musculus	Retina	10x	1,573
mouse_NMDA_48hr [18]	Mus musculus	Retina	10x	5,189
mouse_NMDA_72hr [18]	Mus musculus	Retina	10x	2,591
mouse_NMDA_P60 [18]	Mus musculus	Retina	10x	6,889
mouse_LD_0hr [18]	Mus musculus	Retina	10x	11,383

mouse_LD_4hr [18]	Mus musculus	Retina	10x	3,094
mouse_LD_10hr [18]	Mus musculus	Retina	10x	6,371
mouse_LD_24hr [18]	Mus musculus	Retina	10x	4,756
mouse_LD_P60 [18]	Mus musculus	Retina	10x	15,256
zebrafish_LD_4hr [18]	Danio rerio	Retina	10x	3,475
zebrafish_LD_10hr [18]	Danio rerio	Retina	10x	5,226
zebrafish_LD_20hr [18]	Danio rerio	Retina	10x	6,784
zebrafish_LD_36hr [18]	Danio rerio	Retina	10x	10,183
zebrafish_LD_Adult [18]	Danio rerio	Retina	10x	19,485
zebrafish_NMDA_4hr [18]	Danio rerio	Retina	10x	7,387
zebrafish_NMDA_10hr [18]	Danio rerio	Retina	10x	4,727
zebrafish_NMDA_20hr [18]	Danio rerio	Retina	10x	4,603
zebrafish_NMDA_36hr [18]	Danio rerio	Retina	10x	4,034
zebrafish_NMDA_Adult [18]	Danio rerio	Retina	10x	19,485
zebrafish_TNFa_10hr [18]	Danio rerio	Retina	10x	9,769
zebrafish_TNFa_20hr [18]	Danio rerio	Retina	10x	5,086
zebrafish_TNFa_36hr [18]	Danio rerio	Retina	10x	6,269
zebrafish_TNFa_72hr [18]	Danio rerio	Retina	10x	7,931
zebrafish_TNFa_Adult [18]	Danio rerio	Retina	10x	19,485
chick_P10 [18]	Gallus gallus domesticus	Retina	10x	13,819

Table S2. Top 10 enriched GO terms for each of the gene modules extracted from scRNA-seq data of human and mouse brains. All the GO terms are filtered by the Benjamini-Hochberg adjusted *p*-value < 0.1 (see the Excel table).

Table S3. Top 10 enriched GO terms for each of the gene modules extracted from scRNA-seq data of human and macaque testis. All the GO terms are filtered by the Benjamini-Hochberg adjusted *p*-value < 0.1 (see the Excel table).

Reference

1. Baron, M. et al. (2016) A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. **Cell Syst** 3 (4), 346-360.e4.

2. Schaum, N. et al. (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. **Nature** 562 (7727), 367-372.

3. Shami, A.N. et al. (2020) Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. **Dev Cell** 54(4):529-547.e12.

4. Lake, B.B. et al. (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. **Nat Biotechnol** 36 (1), 70-80.

5. Tasic, B. et al. (2018) Shared and distinct transcriptomic cell types across neocortical areas. **Nature** 563 (7729), 72-78.

6. Campbell, J.N. et al. (2017) A molecular census of arcuate hypothalamus and median eminence cell types. **Nat Neurosci** 20 (3), 484-496.

7. Chen, R. et al. (2017) Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. **Cell Rep** 18 (13), 3227-3241.

8. Tosches, M.A. et al. (2018) Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. **Science** 360 (6391), 881-888.

9. Adam, M. et al. (2017) Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. **Development** 144 (19), 3625-3632.

10. Karaiskos, N. et al. (2018) A Single-Cell Transcriptome Atlas of the Mouse Glomerulus. **J Am Soc Nephrol** 29 (8), 2060-2068.

11. Park, J. et al. (2018) Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. **Science** 360 (6390), 758-763.

12. Wu, H. et al. (2018) Comparative Analysis and Refinement of Human PSC-Derived Kidney Organoid Differentiation with Single-Cell Transcriptomics. **Cell Stem Cell** 23 (6), 869-881.e8.

13. Young, M.D. et al. (2018) Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. **Science** 361 (6402), 594-599.

14. Hochane, M. et al. (2019) Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. **PLoS Biol** 17 (2), e3000152.

15. Alemany, A. et al. (2018) Whole-organism clone tracing using single-cell sequencing. **Nature** 556 (7699), 108-112.

16. Menon, M. et al. (2019) Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. **Nat Commun** 10 (1), 4902.

17. Macosko, E.Z. et al. (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. **Cell** 161 (5), 1202-1214.

18. Hoang, T. et al. (2020) Gene regulatory networks controlling vertebrate retinal regeneration. **Science** 370 (6519).