Table S1. The statistics of 15 real scRNA-seq datasets.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | dataset | organ | platform | cell types | cells | gene number |
| 1 | Human\_PBMC | Human PBMC | 10X | 8 | 4271 | 16653 |
| 2 | Human\_p | Human pancreas | inDrop | 14 | 3605 | 20125 |
| 3 | Human\_k | kidney | 10X | 11 | 5685 | 25215 |
| 4 | Mouse\_E | Mouse embryo stem cells | inDrop | 4 | 2717 | 24047 |
| 5 | Mouse\_k | Kidney | Drop-seq | 8 | 3660 | 23797 |
| 6 | Mouse\_h | Mouse brain | Drop-seq | 46 | 12089 | 23284 |
| **7** | Turtle\_b | Turtle brain | Drop-seq | 15 | 18664 | 23500 |
| 8 | Bach | Mammary Gland | 10x Genomic | 8 | 23184 | 19965 |
| 9 | Muraro | Pancreas | CEL-seq2 | 9 | 2122 | 19046 |
| 10 | Plasschaert | Trachea | inDrop | 8 | 6977 | 28205 |
| 11 | Pollen | Human tissues | SMARTer | 11 | 301 | 21721 |
| 12 | Qx\_Limb\_Muscle | Limb Muscle | 10X | 6 | 3909 | 23341 |
| 13 | QS\_Diaphragm | Diaphragm | Smart-seq2 | 5 | 870 | 23341 |
| 14 | QS\_Limb\_Muscle | Limb Muscle | Smart-seq2 | 6 | 1090 | 23341 |
| 15 | QS\_Lung | Lung | Smart-seq2 | 11 | 1676 | 23341 |

Table S2. Clustering performance comparison of the different clustering algorithms on 15 real scRNA-seq datasets measured by ARI.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | scAClc | scMGCA | scTAG | scDFN | scDeepCluster | scSCC |
| Human\_PBMC | **0.78131** | 0.6666 | 0.5882 | 0.7704 | 0.7056 | 0.7057 |
| Human\_p | **0.95501** | 0.9024 | 0.6954 | 0.6298 | 0.5867 | 0.7406 |
| Human\_k | 0.66187 | **0.7673** | 0.672 | 0.5758 | 0.464 | 0.5695 |
| Mouse\_E | **0.91077** | 0.692 | 0.6711 | 0.748 | 0.6607 | 0.7406 |
| Mouse\_k | **0.93488** | 0.7119 | 0.8511 | 0.6525 | 0.7381 | 0.7553 |
| Mouse\_h | **0.78798** | 0.4424 | 0.5799 | 0.6364 | 0.5314 | 0.343 |
| Turtle\_b | **0.89257** | 0.4858 | 0.559 | 0.4128 | 0.4653 | 0.7658 |
| Bach | 0.84288 | 0.8926 | 0.8852 | 0.8524 | 0.5284 | **0.9308** |
| Muraro | **0.9288** | 0.8843 | 0.7946 | 0.928 | 0.6372 | 0.9003 |
| Plasschaert | **0.91247** | 0.8215 | 0.5347 | 0.5337 | 0.2805 | 0.7465 |
| Pollen | 0.89519 | 0.8642 | 0.7538 | **0.9351** | 0.822 | 0.7317 |
| Qx\_Limb\_Muscle | 0.85804 | 0.809 | **0.9646** | 0.8011 | 0.4759 | 0.8105 |
| QS\_Diaphragm | 0.95383 | **0.9652** | 0.9153 | 0.958 | 0.6561 | 0.4645 |
| QS\_Limb\_Muscle | 0.91048 | 0.9751 | 0.9667 | **0.9779** | 0.6925 | 0.9559 |
| QS\_Lung | 0.61384 | 0.5834 | 0.6734 | 0.5961 | 0.4101 | **0.8304** |

Table S3. Clustering performance comparison of the different clustering algorithms on 15 real scRNA-seq datasets measured by NMI.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | scAClc | scMGCA | scTAG | scDFN | scDeepCluster | scSCC |
| Human\_PBMC | **0.79888** | 0.7473 | 0.6538 | 0.7697 | 0.7612 | 0.7466 |
| Human\_p | **0.91085** | 0.8514 | 0.7525 | 0.816 | 0.7723 | 0.7177 |
| Human\_k | 0.77477 | **0.8196** | 0.7802 | 0.74 | 0.6919 | 0.7641 |
| Mouse\_E | **0.91536** | 0.6898 | 0.6877 | 0.7295 | 0.748 | 0.7177 |
| Mouse\_k | **0.91531** | 0.7949 | 0.8328 | 0.7679 | 0.7962 | 0.8363 |
| Mouse\_h | **0.78103** | 0.7266 | 0.7382 | 0.7582 | 0.7488 | 0.7219 |
| Turtle\_b | **0.86275** | 0.7064 | 0.6972 | 0.7175 | 0.7411 | 0.7843 |
| Bach | 0.84154 | 0.8338 | 0.8514 | 0.8505 | 0.728 | **0.8966** |
| Muraro | 0.88469 | 0.8358 | 0.8047 | **0.8924** | 0.7905 | 0.8563 |
| Plasschaert | **0.86181** | 0.7602 | 0.577 | 0.6836 | 0.5452 | 0.7528 |
| Pollen | **0.93393** | 0.9166 | 0.8606 | 0.9299 | 0.902 | 0.8067 |
| Qx\_Limb\_Muscle | 0.92651 | 0.8677 | **0.947** | 0.8728 | 0.7721 | 0.8874 |
| QS\_Diaphragm | 0.91259 | 0.9392 | 0.8987 | **0.9448** | 0.8027 | 0.6494 |
| QS\_Limb\_Muscle | 0.85995 | 0.9568 | 0.9448 | **0.9614** | 0.8153 | 0.919 |
| QS\_Lung | 0.76288 | 0.765 | 0.7996 | 0.7626 | 0.697 | **0.8215** |

Table S4. The number of cells on the Mouse\_h dataset and the correspondence between real cell type and cell number.

|  |  |  |
| --- | --- | --- |
| Cell type | Cluster number | Cell number |
| Glu1 | 0 | 13 |
| Glu2 | 1 | 21 |
| Glu3 | 2 | 13 |
| Glu4 | 3 | 131 |
| Glu5 | 4 | 200 |
| Glu6 | 5 | 51 |
| Glu7 | 6 | 211 |
| Glu8 | 7 | 50 |
| Glu9 | 8 | 35 |
| Glu10 | 9 | 15 |
| Glu11 | 10 | 42 |
| Glu12 | 11 | 24 |
| Glu13 | 12 | 24 |
| Glu14 | 13 | 51 |
| Glu15 | 14 | 25 |
| GABA1 | 15 | 23 |
| GABA2 | 16 | 27 |
| GABA3 | 17 | 97 |
| GABA4 | 18 | 18 |
| GABA5 | 19 | 73 |
| GABA6 | 20 | 49 |
| GABA7 | 21 | 17 |
| GABA8 | 22 | 402 |
| GABA9 | 23 | 71 |
| GABA10 | 24 | 23 |
| GABA11 | 25 | 35 |
| GABA12 | 26 | 62 |
| GABA13 | 27 | 165 |
| GABA14 | 28 | 71 |
| GABA15 | 29 | 112 |
| GABA16 | 30 | 66 |
| GABA17 | 31 | 50 |
| GABA18 | 32 | 31 |
| Hista | 33 | 17 |
| POPC | 34 | 51 |
| OPC | 35 | 1741 |
| IMO | 36 | 151 |
| MO | 37 | 3541 |
| Astro | 38 | 1148 |
| Ependy | 39 | 413 |
| Tany | 40 | 609 |
| Epith1 | 41 | 818 |
| Epith2 | 42 | 379 |
| Micro | 43 | 724 |
| Macro | 44 | 167 |

Table S5. Clustering result comparison of scAClc in different module ((i) only HVGs; (ii) only RFGs; (iii) without the dual feature selection method; (iv) dual feature selection method.) measured by ARI.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | only HVGs | only RFGs | without both | dual feature selection |
| Human\_PBMC | 0.7344 | 0.7141 | 0.5766 | 0.7813 |
| Human\_p | 0.9554 | 0.8934 | 0.8973 | 0.9550 |
| Human\_k | 0.7194 | 0.6166 | 0.6778 | 0.6618 |
| Mouse\_E | 0.8075 | 0.9123 | 0.8278 | 0.9107 |
| Mouse\_k | 0.8141 | 0.7929 | 0.8034 | 0.9348 |
| Mouse\_h | 0.6557 | 0.7733 | 0.8511 | 0.7879 |
| Turtle\_b | 0.6935 | 0.7796 | 0.7663 | 0.8925 |
| Bach | 0.8419 | 0.8506 | 0.7724 | 0.8428 |
| Muraro | 0.9328 | 0.9331 | 0.7290 | 0.9288 |
| Plasschaert | 0.7544 | 0.7651 | 0.9361 | 0.9124 |
| Pollen | 0.8567 | 0.8656 | 0.3177 | 0.8951 |
| Qx\_Limb\_Muscle | 0.9933 | 0.9858 | 0.9914 | 0.8580 |
| QS\_Diaphragm | 0.9822 | 0.9519 | 0.5874 | 0.9538 |
| QS\_Limb\_Muscle | 0.7026 | 0.8068 | 0.6400 | 0.9104 |
| QS\_Lung | 0.9011 | 0.5886 | 0.5775 | 0.6138 |
| Average | **0.8230** | **0.8153** | **0.7301** | **0.8560** |

Table S6. Clustering result comparison of scAClc in different module ((i) only HVGs; (ii) only RFGs; (iii) without the dual feature selection method; (iv) dual feature selection method.) measured by NMI.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | only HVGs | only RFGs | without both | dual feature selection |
| Human\_PBMC | 0.7699 | 0.7656 | 0.7018 | 0.7988 |
| Human\_p | 0.9127 | 0.8702 | 0.8777 | 0.9108 |
| Human\_k | 0.8096 | 0.7489 | 0.7875 | 0.7747 |
| Mouse\_E | 0.8271 | 0.916 | 0.8614 | 0.9153 |
| Mouse\_k | 0.8698 | 0.8484 | 0.8600 | 0.9153 |
| Mouse\_h | 0.6904 | 0.7652 | 0.7962 | 0.7810 |
| Turtle\_b | 0.7428 | 0.8305 | 0.7590 | 0.8627 |
| Bach | 0.8522 | 0.8424 | 0.7794 | 0.8415 |
| Muraro | 0.8902 | 0.8904 | 0.8282 | 0.8846 |
| Plasschaert | 0.7550 | 0.7714 | 0.8713 | 0.8618 |
| Pollen | 0.9232 | 0.9078 | 0.5801 | 0.9339 |
| Qx\_Limb\_Muscle | 0.9871 | 0.9770 | 0.9840 | 0.9265 |
| QS\_Diaphragm | 0.9686 | 0.9067 | 0.8032 | 0.9125 |
| QS\_Limb\_Muscle | 0.7666 | 0.7538 | 0.7853 | 0.8599 |
| QS\_Lung | 0.8521 | 0.7290 | 0.7720 | 0.7628 |
| Average | **0.8412** | **0.8349** | **0.8032** | **0.8628** |

Table S7. Clustering result comparison of scAClc in different module ((i) standard contrastive learning; (ii) without contrastive learning; (iii) anchor-centered contrastive learning;) measured by ARI.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | standard | without | anchor-centered |
| Human\_PBMC | 0.7649 | 0.7787 | 0.7813 |
| Human\_p | 0.9545 | 0.9577 | 0.9550 |
| Human\_k | 0.6310 | 0.5178 | 0.6618 |
| Mouse\_E | 0.9123 | 0.9170 | 0.9107 |
| Mouse\_k | 0.8159 | 0.9389 | 0.9348 |
| Mouse\_h | 0.8269 | 0.5490 | 0.7879 |
| Turtle\_b | 0.7640 | 0.8651 | 0.8925 |
| Bach | 0.8176 | 0.6332 | 0.8428 |
| Muraro | 0.9309 | 0.9288 | 0.9288 |
| Plasschaert | 0.7872 | 0.9110 | 0.9124 |
| Pollen | 0.7756 | 0.8951 | 0.8951 |
| Qx\_Limb\_Muscle | 0.9917 | 0.9903 | 0.8580 |
| QS\_Diaphragm | 0.9783 | 0.9574 | 0.9538 |
| QS\_Limb\_Muscle | 0.9211 | 0.9171 | 0.9104 |
| QS\_Lung | 0.7414 | 0.5235 | 0.6138 |
| Average | **0.8409** | **0.8187** | **0.8560** |

Table S8. Clustering result comparison of scAClc in different module ((i) standard contrastive learning; (ii) without contrastive learning; (iii) anchor-centered contrastive learning;) measured by NMI

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | standard | without | anchor-centered |
| Human\_PBMC | 0.7877 | 0.7966 | 0.7988 |
| Human\_p | 0.9095 | 0.9138 | 0.9108 |
| Human\_k | 0.7468 | 0.6869 | 0.7747 |
| Mouse\_E | 0.9173 | 0.9260 | 0.9153 |
| Mouse\_k | 0.8712 | 0.9189 | 0.9153 |
| Mouse\_h | 0.8028 | 0.6551 | 0.7810 |
| Turtle\_b | 0.8270 | 0.8345 | 0.8627 |
| Bach | 0.8165 | 0.7088 | 0.8415 |
| Muraro | 0.8864 | 0.8846 | 0.8846 |
| Plasschaert | 0.7786 | 0.8594 | 0.8618 |
| Pollen | 0.8960 | 0.9339 | 0.9339 |
| Qx\_Limb\_Muscle | 0.9845 | 0.9817 | 0.9265 |
| QS\_Diaphragm | 0.9528 | 0.9188 | 0.9125 |
| QS\_Limb\_Muscle | 0.8641 | 0.8654 | 0.8599 |
| QS\_Lung | 0.7810 | 0.7508 | 0.7628 |
| Average | **0.8548** | **0.8423** | **0.8628** |

Table S9. Clustering result comparison of scAClc in different parameters ((i) The dual feature selection method screens 2000 genes respectively; (ii) The dual feature selection method screens 3000 genes respectively; (iii) The dual feature selection method screens 4000 genes respectively;) measured by ARI.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | 2000 | 3000 | 4000 |
| Human\_PBMC | 0.7544 | 0.7813 | 0.8089 |
| Human\_p | 0.9559 | 0.9550 | 0.8939 |
| Human\_k | 0.6413 | 0.6618 | 0.6805 |
| Mouse\_E | 0.9111 | 0.9107 | 0.8222 |
| Mouse\_k | 0.8129 | 0.9348 | 0.8588 |
| Mouse\_h | 0.7854 | 0.7879 | 0.7875 |
| Turtle\_b | 0.7682 | 0.8925 | 0.7717 |
| Bach | 0.5772 | 0.8428 | 0.6132 |
| Muraro | 0.9316 | 0.9288 | 0.7345 |
| Plasschaert | 0.9126 | 0.9124 | 0.5971 |
| Pollen | 0.7686 | 0.8951 | 0.5967 |
| Qx\_Limb\_Muscle | 0.9851 | 0.8580 | 0.9913 |
| QS\_Diaphragm | 0.5962 | 0.9538 | 0.9682 |
| QS\_Limb\_Muscle | 0.9273 | 0.9104 | 0.8408 |
| QS\_Lung | 0.5149 | 0.6138 | 0.7511 |
| Average | **0.7895** | **0.8560** | **0.7811** |

Table S10. Clustering result comparison of scAClc in different parameters ((i) The dual feature selection method screens 2000 genes respectively; (ii) The dual feature selection method screens 3000 genes respectively; (iii) The dual feature selection method screens 4000 genes respectively;) measured by NMI.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | 2000 | 3000 | 4000 |
| Human\_PBMC | 0.7877 | 0.7988 | 0.8036 |
| Human\_p | 0.9121 | 0.9108 | 0.8719 |
| Human\_k | 0.7672 | 0.7747 | 0.7979 |
| Mouse\_E | 0.9134 | 0.9153 | 0.8505 |
| Mouse\_k | 0.8634 | 0.9153 | 0.8763 |
| Mouse\_h | 0.7736 | 0.7810 | 0.7825 |
| Turtle\_b | 0.7495 | 0.8627 | 0.7388 |
| Bach | 0.7404 | 0.8415 | 0.7073 |
| Muraro | 0.8882 | 0.8846 | 0.8326 |
| Plasschaert | 0.8596 | 0.8618 | 0.6915 |
| Pollen | 0.8834 | 0.9339 | 0.8039 |
| Qx\_Limb\_Muscle | 0.9747 | 0.9265 | 0.9844 |
| QS\_Diaphragm | 0.7923 | 0.9125 | 0.9339 |
| QS\_Limb\_Muscle | 0.8740 | 0.8599 | 0.8313 |
| QS\_Lung | 0.7581 | 0.7628 | 0.7895 |
| Average | **0.8358** | **0.8628** | **0.8197** |

Table S11. Clustering result comparison of scAClc in different parameters ((i) The contrastive loss weight ranges from 0.1 to 0.06; (ii) The contrastive loss weight ranges from 0.12 to 0.08; (iii) The contrastive loss weight ranges from 0.14 to 0.1;) measured by ARI.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | 0.1-0.06 | 0.12-0.08 | 0.14-0.1 |
| Human\_PBMC | 0.7745 | 0.7813 | 0.7791 |
| Human\_p | 0.9564 | 0.9550 | 0.9559 |
| Human\_k | 0.5714 | 0.6618 | 0.6629 |
| Mouse\_E | 0.9150 | 0.9107 | 0.9106 |
| Mouse\_k | 0.9413 | 0.9348 | 0.9349 |
| Mouse\_h | 0.8309 | 0.7879 | 0.7753 |
| Turtle\_b | 0.8310 | 0.8925 | 0.7951 |
| Bach | 0.6340 | 0.8428 | 0.8537 |
| Muraro | 0.9272 | 0.9288 | 0.7288 |
| Plasschaert | 0.9115 | 0.9124 | 0.6161 |
| Pollen | 0.8951 | 0.8951 | 0.8951 |
| Qx\_Limb\_Muscle | 0.8552 | 0.8580 | 0.9897 |
| QS\_Diaphragm | 0.9538 | 0.9538 | 0.5857 |
| QS\_Limb\_Muscle | 0.9025 | 0.9104 | 0.9096 |
| QS\_Lung | 0.7411 | 0.6138 | 0.6162 |
| Average | **0.8427** | **0.8560** | **0.8006** |

Table S12. Clustering result comparison of scAClc in different parameters ((i) The contrastive loss weight ranges from 0.1 to 0.06; (ii) The contrastive loss weight ranges from 0.12 to 0.08; (iii) The contrastive loss weight ranges from 0.14 to 0.1;) measured by NMI.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | 0.1-0.06 | 0.12-0.08 | 0.14-0.1 |
| Human\_PBMC | 0.7954 | 0.7988 | 0.7949 |
| Human\_p | 0.9122 | 0.9108 | 0.9110 |
| Human\_k | 0.7251 | 0.7747 | 0.7737 |
| Mouse\_E | 0.9222 | 0.9153 | 0.9150 |
| Mouse\_k | 0.9217 | 0.9153 | 0.9165 |
| Mouse\_h | 0.8055 | 0.7810 | 0.7634 |
| Turtle\_b | 0.8017 | 0.8627 | 0.8249 |
| Bach | 0.7104 | 0.8415 | 0.8446 |
| Muraro | 0.8835 | 0.8846 | 0.8277 |
| Plasschaert | 0.8595 | 0.8618 | 0.6991 |
| Pollen | 0.9339 | 0.9339 | 0.9339 |
| Qx\_Limb\_Muscle | 0.9283 | 0.9265 | 0.9824 |
| QS\_Diaphragm | 0.9125 | 0.9125 | 0.7686 |
| QS\_Limb\_Muscle | 0.8518 | 0.8599 | 0.8618 |
| QS\_Lung | 0.7788 | 0.7628 | 0.7645 |
| Average | **0.8495** | **0.8628** | **0.8388** |

**Reference**

[1] G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu et al., “Massively parallel digital transcriptional profiling of single cells,” Nature communications, vol. 8, no. 1, p. 14049, 2017.

[2] M. Baron, A. Veres, S. L. Wolock, A. L. Faust, R. Gaujoux, A. Vetere, J. H. Ryu, B. K. Wagner, S. S. Shen-Orr, A. M. Klein 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, 2016.

[3] M. D. Young, T. J. Mitchell, F. A. Vieira Braga, M. G. Tran, B. J. Stewart, J. R. Ferdinand, G. Collord, R. A. Botting, D.-M. Popescu, K. W. Loudon et al., “Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors,” science, vol. 361, no. 6402, pp. 594–599, 2018.

[4] A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, and M. W. Kirschner, “Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells,” Cell, vol. 161, no. 5, pp. 1187–1201, 2015.

[5] R. Chen, X. Wu, L. Jiang, and Y. Zhang, “Single-cell rna-seq reveals hypothalamic cell diversity,” Cell reports, vol. 18, no. 13, pp. 3227–3241, 2017.

[6] M. A. Tosches, T. M. Yamawaki, R. K. Naumann, A. A. Jacobi, G. Tushev, and G. Laurent, “Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles,” Science, vol. 360, no. 6391, pp. 881– 888, 2018.

[7] X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye et al., “Mapping the mouse cell atlas by microwell-seq,” Cell, vol. 172, no. 5, pp. 1091–1107, 2018.

[8] M. J. Muraro, G. Dharmadhikari, D. Grün, N. Groen, T. Dielen, E. Jansen, L. Van Gurp, M. A. Engelse, F. Carlotti, E. J. De Koning et al., “A single-cell transcriptome atlas of the human pancreas,” Cell systems, vol. 3, no. 4, pp. 385–394, 2016.

[9] A. A. Pollen, T. J. Nowakowski, J. Shuga, X. Wang, A. A. Leyrat, J. H. Lui, N. Li, L. Szpankowski, B. Fowler, P. Chen et al., “Low-coverage single-cell mrna sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex,” Nature biotechnology, vol. 32, no. 10, pp. 1053– 1058, 2014.

[10] N. Schaum, J. Karkanias, N. F. Neff, A. P. May, S. R. Quake, T. Wyss-Coray, S. Darmanis, J. Batson, O.Botvinnik, M. B. Chen et al., “Single-cell transcriptomics of 20 mouse organs creates a tabula muris: The tabula muris consortium,” Nature, vol. 562, no. 7727, p. 367, 2018.

[11] B. Tasic, V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, T. Dolbeare et al., “Adult mouse cortical cell taxonomy revealed by single cell transcriptomics,” Nature neuroscience, vol. 19, no. 2, pp. 335– 346, 2016.