**Table 2b:** Human single-cell data sources used by different DL algorithms

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Title | Algorithm | Cell origin | # Cells | Data Sources | Reference |
| 68k PBMCs | DCA  SAVER-X  LATE, scVAE, scDeepCluster, scCapsNet, scDGN | Blood | 68,579 | 10X Single Cell Gene Expression Datasets |  |
| Human pluripotent | DCA | hESCs | 1,876 | GSE102176 | [122] |
| CITE-seq | SAVER-X | Cord blood mononuclear cells | 8,005 | GSE100866 | [123] |
| Midbrain and Dopaminergic Neuron Development | SAVER-X | Brain/ embryo ventral midbrain cells | 1,977 | GSE76381 | [124] |
| HCA | SAVER-X | Immune cell, Human Cell Atlas | 500,000 | HCA data portal |  |
| Breast tumor | SAVER-X | Immune cell in tumor micro-environment | 45,000 | GSE114725 | [125] |
| 293T cells | DeepImpute,  iMAP | Embryonic kidney | 13,480 | 10X Single Cell Gene Expression Datasets |  |
| Jurkat | DeepImpute,  iMAP | Blood/ lymphocyte | 3,200 | 10X Single Cell Gene Expression Datasets |  |
| ESC, Time-course | scGAN | ESC | 350, 758 | GSE75748 | [126] |
| Baron-Hum-1 | scGMAI, VASC | Pancreatic islets | 1,937 | GSM2230757 | [127] |
| Baron-Hum-2 | scGMAI, VASC | Pancreatic islets | 1,724 | GSM2230758 | [127] |
| Camp | scGMAI, VASC | Liver cells | 303 | GSE96981 | [128] |
| CEL-seq2 | PERMUDA, DESC | Pancreas/Islets of Langerhans |  | GSE85241 | [129] |
| Darmanis | scGMAI, sclGAN, VASC | Brain/cortex | 466 | GSE67835 | [130] |
| Tirosh-brain | Dhaka, scvis | Oligodendroglioma | >4800 | GSE70630 | [131] |
| Patel | Dhaka | Primary glioblastoma cells | 875 | GSE57872 | [132] |
| Li | scGMAI, VASC | Blood | 561 | GSE146974 | [61] |
| Tirosh-skin | scvis | melanoma | 4645 | GSE72056 | [67] |
| xenograft 3, and 4 | Dhaka | Breast tumor | ~250 | EGAS00001002170 | [133] |
| Petropoulos | VASC/netAE | Human embryos | 1,529 | E-MTAB-3929 |  |
| Pollen | scGMAI, VASC |  | 348 | SRP041736 | [134] |
| Xin | scGMAI, VASC | Pancreatic cells (α-, β-, δ-) | 1,600 | GSE81608 | [135] |
| Yan | scGMAI, VASC | embryonic stem cells | 124 | GSE36552 | [136] |
| PBMC3k | VASC, scVI | Blood | 2,700 | SRP073767 | [96] |
| CyTOF, Dengue | SAUCIE | Dengue infection | 11 M, ~42 antibodies | Cytobank, 82023 | [14] |
| CyTOF, ccRCC | SAUCIE | Immunue profile of 73 ccRCC patients | 3.5 M, ~40 antibodies | Cytobank: 875 | [137] |
| CyTOF, breast | SAUCIE | 3 patients |  | Flow Repository: FR-FCM-ZYJP | [125] |
| Chung, BC | DigitalDLSorter | Breast tumor | 515 | GSE75688 | [88] |
| Li, CRC | DigitalDLSorter | Colorectal cancer | 2,591 | GSE81861 | [89] |
| Pancreatic datasets | scDGN | Pancreas | 14693 | SeuratData |  |
| Kang, PBMC | scGen | PBMC stimulated by INF-b | ~15,000 | GSE96583 | [116] |
|  |  |  |  |  |  |