https://doi.org/10.1093/nar/gkab775

scDeepSort: a pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network

Xin Shao [1,4,†], Haihong Yang [2,7,†], Xiang Zhuang [2], Jie Liao [1,4], Penghui Yang [1], Junyun Cheng [1], Xiaoyan Lu [1,5], Huajun Chen [2,3,7,*] and Xiaohui Fan [1,4,5,6,*]

[1] Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China, [2] College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China, [3] The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, China, [4] iMedicine Lab, Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Zhejiang University, Hangzhou 310058, China, [5] Innovation Center in Zhejiang University, State Key Laboratory of Component-Based Chinese Medicine, Hangzhou 310058, China, [6] Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310058, China and [7] Hangzhou Innovation Center, Zhejiang University, Hangzhou 310058, China

Received June 21, 2021; Revised August 04, 2021; Editorial Decision August 23, 2021; Accepted August 26, 2021

ABSTRACT

Advances in single-cell RNA sequencing (scRNA-seq) have furthered the simultaneous classification of thousands of cells in a single assay based on transcriptome profiling. In most analysis protocols, single-cell type annotation relies on marker genes or RNA-seq profiles, resulting in poor extrapolation. Still, the accurate cell-type annotation for single-cell transcriptomic data remains a great challenge. Here, we introduce scDeepSort (https://github.com/ZJUFanLab/scDeepSort), a pre-trained cell-type annotation tool for single-cell transcriptomics that uses a deep learning model with a weighted graph neural network (GNN). Using human and mouse scRNA-seq data resources, we demonstrate the high performance and robustness of scDeepSort in labeling 764 741 cells involving 56 human and 32 mouse tissues. Significantly, scDeepSort outperformed other known methods in annotating 76 external test datasets, reaching an 83.79% accuracy across 265 489 cells in humans and mice. Moreover, we demonstrate the universality of scDeepSort using more challenging datasets and using references from different scRNA-seq technology. Above all, scDeepSort is the first attempt to annotate cell types of scRNA-seq data with a pre-trained GNN model, which can realize the accurate cell-type annotation without additional references, i.e. markers or RNA-seq profiles.

INTRODUCTION

Recent advancements in single-cell RNA sequencing (scRNA-seq) that permit the identification of various cell types based on transcriptomics at single-cell resolution have facilitated our understanding of the heterogeneity of cellular phenotypes and their composition within complex tissues (1,2). In the data processing protocols of scRNA-seq experiments, cell-type annotation is a vital step for subsequent analysis (3–5). Cell type identification is commonly performed by mapping differentially expressed genes at the level of pre-computed clusters with prior knowledge of cell markers like SCSA (6) and scCATCH (7). Another cell-based annotation strategy tries to compare the similarities between a single cell and a reference database of bulk or single-cell RNA-seq profiles to determine potential cellular identities. Several methods including SingleR (8), CHETAH (9), scMap (10), scID (11), scPred (12), ACTINN (13), CellAssign (14), Garnett (15), SCINA (16), singleCellNet (17) and support vector machine (SVM) belong to this category as described in the recent review (18). Such methods rely heavily on references, severely limiting the extrapolation of these methods. Still, the accurate cell-type annotation for single-cell transcriptomic data remains a great challenge (19).

Fortunately, recent advances in deep learning have enabled major progress in the ability of artificial intelligence techniques to integrate big data, incorporate existing knowledge, and learn arbitrarily complex relationships (20,21). Given the state-of-the-art accuracy deep learning has achieved in numerous prediction tasks, it has been increasingly used in biological research (22) and

[*] To whom correspondence should be addressed. Tel: +86 571 88208596; Fax: +86 571 88208426; Email: fanxh@zju.edu.cn
Correspondence may also be addressed to Huajun Chen. Tel: +86 571 86704050; Fax: +86 571 86704050; Email: huajunsir@zju.edu.cn
[†] The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

biomedical applications (23,24). For example, Chaudhary *et al.* established a deep learning-based model using RNA-seq, miRNA-seq and methylation data of 360 hepatocellular carcinomas (HCC) patients to help predict patient survival (25). The graph neural networks (GNN), one of the commonly-used deep learning methods (26), is the connectionist model that captures the graph dependence through message passing between the graph nodes. Unlike standard neural networks, GNNs retain a state that represents information from its neighborhood with arbitrary depth, which has demonstrated ground-breaking performance on many learning tasks (27). Moreover, recent published large-scale scRNA-seq resources have provided the foundation for deep learning models that can execute challenging prediction tasks (28–30).

In this study, we designed a pre-trained cell-type annotation method called scDeepSort, based on a weighted GNN framework since cells and genes from the scRNA-seq data are the natural graph structure as genes are expressed by cells, which addresses this challenge (see Figure 1 for an overview). First, we prepared the most comprehensive single-cell transcriptomics atlases, namely a human cell landscape (HCL) (31) and a mouse cell atlas (MCA) (32), as the underlying training set (Supplementary Table S1). Second, we trained our scDeepSort mainly on HCL and MCA using a weighted GNN model. Last, we compared the performance of scDeepSort with other known methods on 76 external human and mouse test datasets. Moreover, we evaluated the performance of scDeepSort with three additional scRNA-seq atlases (33–35) and with the Tabula Muris (TM) dataset (36). The present results indicated that scDeepSort is a robust method that can help scientists realize the accurate cell-type annotation of scRNA-seq data without additional references, i.e., markers or RNA-seq profiles.

MATERIALS AND METHODS

Datasets

All scRNA-seq datasets were retrieved from several high-quality reports and the Gene Expression Omnibus (GEO), including human and mouse primary tissues, wherein unannotated cells were excluded and normal or healthy cells were included. The human cell landscape (HCL, https://figshare.com/articles/HCL_DGE_Data/7235471) provided data for 562 977 cells from 56 types of tissues and the mouse cell atlas (MCA, https://figshare.com/articles/MCA_DGE_Data/5435866) provided 201 764 cells involving 32 tissues. External test datasets used for comparing scDeepSort with other methods were freely available from public platforms detailed in Supplementary Table S2. The TM dataset were downloaded from the Figshare website (https://figshare.com/projects/Tabula_Muris_Transcriptomic_characterization_of_20_organs_and_tissues_from_Mus_musculus_at_single_cell_resolution/27733). The Loom (http://mousebrain.org), DropViz (http://dropviz.org) and oncoscape (https://oncoscape.v3.sttrcancer.org/atlas.gs.washington.edu.mouse.rna/landing) datasets were all downloaded from websites. The large human pancreas dataset (37) by Baron et al. was downloaded from GSE84133.

Data preprocessing

All scRNA-seq data were preprocessed using R (version 3.6.1). For the Zheng dataset, the raw count was processed in accordance with the pipeline detailed in the Satija Lab tutorial, using Seurat 3.0, wherein cells with more than 2500 or fewer than 200 unique features or with mitochondrial counts greater than 5% were filtered out. For other datasets, all cells in the datasets were included in the filtered matrices. Human and mouse gene symbols were revised in accordance with NCBI gene data (https://www.ncbi.nlm.nih.gov/gene/) updated on 10 January 2020, wherein unmatched genes and duplicated genes were removed. For all human and mouse datasets, the raw data were normalized via the global-scaling normalization method LogNormalize in preparation for running the subsequent scDeepSort pipeline and other methods.

scDeepSort algorithm

scDeepSort consists of three components: the embedding layer, weighted graph aggregator, and linear classifier layers. The embedding layer stores the representation of graph nodes and is frozen during training. The weighted graph aggregator layer inductively learns graph structure information, generating linear separable feature space for cells. In this layer, a modified version of the GraphSAGE (38) information processing framework was applied as the backbone GNN. The final linear classifier layer classifies the final cell state representation produced from the weighted graph aggregator layer into one of the predefined cell type categories.

Weighted cell-gene graph generation. To construct the weighted cell-gene graph, cells and genes were both treated as graph nodes and the gene expression for each cell was regarded as the weighted edge between cells and genes, constituting the embedding layer. First, the principal component analysis (PCA) was used to extract dense representations for gene nodes from the cell-gene data matrix, and cell node representations were then calculated by the weighted sum of gene node representations and the cell–gene data matrix. For an input single-cell data matrix $D \in \mathbb{R}^{m\times n}$ (m genes and n cells), PCA was applied to extract dense representations of a fixed-size dimension ($d = 400$) as initial representations because the performance of d = 200, 400 and 600 is very close as shown in Supplementary Figure S1. Considering that it takes the lower cumulative variance contribution rate as well as the longer training time for d = 200 and it takes the higher memory and GPU for d = 600, we finally set dimension as 400. A weighted sum of gene representations with single-cell data matrix D as input was used to obtain the cell representations with the same dimension d. By collecting gene and cell representations, a matrix $X \in \mathbb{R}^{(m+n)\times d}$ was constructed as the initial node embeddings. Second, a weighted adjacency matrix $A \in \mathbb{R}^{(m+n)^2}$ was generated from the input single-cell data matrix D, in which the gene expression (>0) was directly regarded as the weights of edges between cells and genes.

Aggregating process. To inductively learn graph structure information, we followed a graph neural network

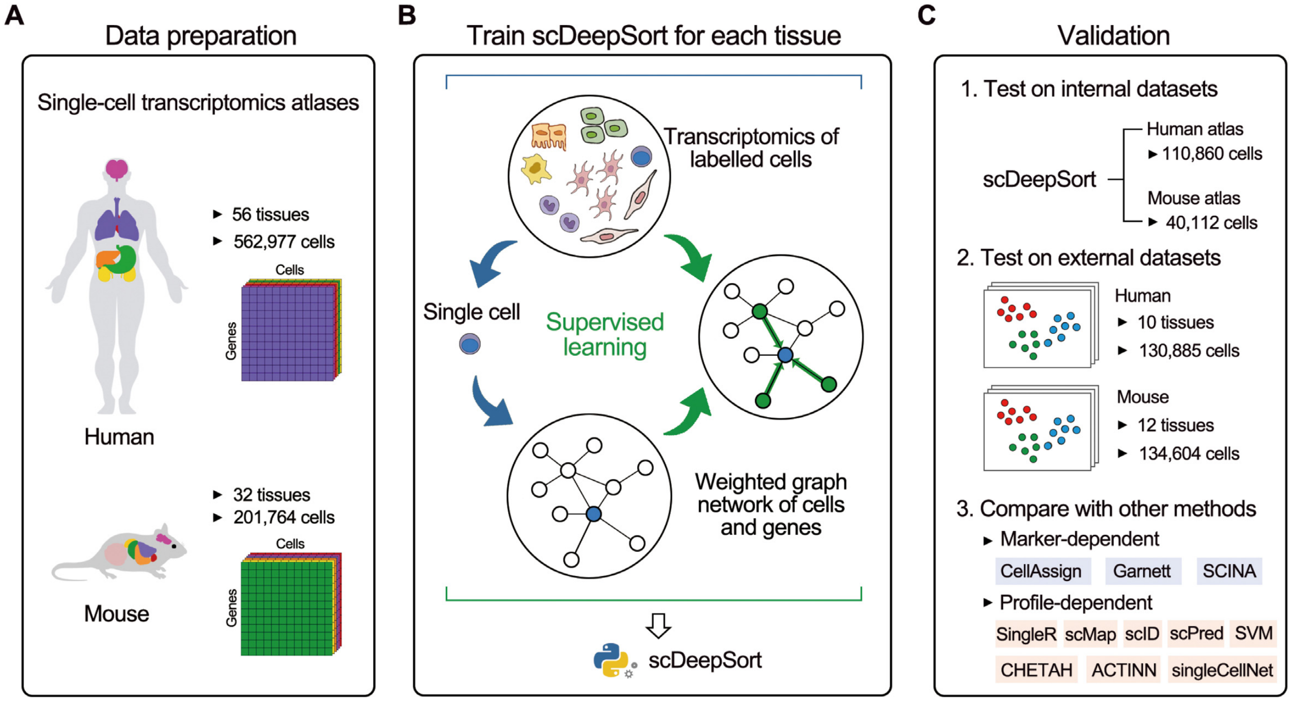

Figure 1. General conceptual framework and validation of scDeepSort. (**A**) Human and mouse single-cell transcriptomics atlases were curated from HCL and MCA as the underlying data for training scDeepSort. The human and mouse atlases include 562,977 cells from 56 tissues and 201,764 cells from 32 tissues, respectively. (**B**) For each cell, a graph network was constructed of this cell, its genes and neighboring cells for supervised learning scDeepSort with known cell labels from the transcriptomic atlases for each tissue. (**C**) Internal human and mouse atlases datasets and external test datasets of single-cell transcriptomics data involving multiple tissues were employed to test the performance of scDeepSort. Markers- and profiles-dependent annotation methods (CellAssign, Garnett, SingleR, scMap, ACTINN, CHETAH, scID, scPred, SCINA, singleCellNet and SVM) were compared with scDeepSort on human and mouse external test datasets.

framework called GraphSAGE. The essential processes of GraphSAGE are sampling a batch of 500 nodes with their neighbors and aggregating graph neighborhood to generate node representations for each node. However, we proposed a new weighted graph aggregator layer to replace the aggregator of GraphSAGE. Let h_i^k (a 200-dimensional vector in our experiments) represents the embedding of node i in the kth layer. Our weighted graph aggregator layer can be summarized as:

$$h_i^k = \sigma\left(W^{k-1} AGG\left(h_i^{k-1}, h_{N(i)}^{k-1}\right) + b^{k-1}\right)$$

where $N(i)$ is the set of one-hop neighbors of node i. The output of the aggregate function AGG is then transformed to the target dimension by a linear transformation shared among all nodes, followed by a non-linear activation function σ called Rectified Linear Unit (ReLU). In practice, we set $k = 1$ (k is the number of the weighted graph aggregator layer) because the performance of $k = 1$ and 2 are very close, but $k = 2$ requires longer training time and more memory during training, as shown in Supplementary Figure S1. The aggregate function AGG contains two new techniques. The first technique is called the weighted adjacency matrix normalization. The main reasons for applying normalization to the weighted adjacency matrix are twofold. Gene expression varies a lot across different kinds of cells. For single cell, the expression level and pattern of different genes can also vary. Thus, we normalize weighted adjacency matrix A as following:

$$a_{ij} \leftarrow d_i \times \frac{a_{ij}}{\sum_{j\in N(i)} a_{ij}}$$

where a_{ij}, the weight of an edge from node j to node i, is the element of A, and d_i denotes the indegree (the number of connected genes) of cell node i. The second technique is the learnable sharing confidence. Due to batch effects and missing value issues, we proposed to add learnable parameters to each edge as a confidence matrix while leveraging the context of one-hop neighborhood of nodes in a weighted graph. For a gene node j, we proposed a learnable sharing parameter β_j as the confidence value for the edges that interact with node j. Another learnable parameter α as the confidence value of the self-loop edge for each cell. Its value will be shared among cells since we may encounter new cells in test datasets. Therefore, the overall formulation of gathering neighborhood information given each sub-graph of cell node i is stated below:

$$h_i^k = \sigma\left(W^{k-1}\frac{\alpha h_i^{k-1} + \sum_{j\in N(i)}\beta_j a_{ij} h_j^{k-1}}{1+|N(i)|} + b^{k-1}\right)$$

A linear classifier layer. The weighted graph aggregator layer produces a latent feature space for the graph. To classify the final cell state representation into one of the pre-defined cell-type categories, we extract cell node representations and feed them into a linear classifier layer.

$$\hat{y}_i = softmax\left(Wh_i^k + b\right)$$

Cross entropy loss was then used to measure the difference between the predicted class distribution and the labels. Therefore, the objective function can be written as:

$$\hat{\theta} = argmin_\theta - \sum_{c=1}^{C} y_c log\, \hat{y}_c$$

θ represents all trainable parameters in our method, including the weight W^{k-1} and bias b^{k-1} of our weighted Graph Neural Network, as well as the learnable confidence of α and β_j for the self-loop edge of each cell and the edges that interact with gene node j, respectively. We train our model with the above objective function using a stochastic gradient descent method called Adam, with default hyper parameters except for the learning rate of 0.001 and the weight decay rate of 0.0005 until convergence or after 500 epochs. The training process is conducted in a mini-batch manner with batch size of 500.

To annotate the new dataset, new cells will be connected to the gene node of the trained cell–gene graph and the gene expression of new cells will be added to the edge between the new cell node and the gene node as the weight. According to the connected gene node representation, the cell representation of the test cells will be calculated by the weighted sum of gene node representations and the cell–gene data matrix. Using the aggregator function consisting of the learned representation of the connected gene nodes, the weights, the learned confidence, and the raw representation of each test cell, a new representation of the test cell will be calculated, which will go through the linear classification layer to generate a predicted cell label for each test cell.

scDeepSort performance evaluation on internal datasets

For each cell type, cells numbering at least more than 5‰ of the total cells in each tissue were included and randomly divided into training and test sets, ensuring that the ratio of training and test cells was set to 8:2, 7:3, 6:4 and 5:5 with five replicates for each alternative split.

Performance comparison with other methods on external test datasets

CellMatch, MCA or HCL were used as the reference datasets for reference-dependent methods. To compare the performance of scDeepSort with other methods on annotating cell types of single-cell transcriptomics data, only the cell types that existed in both cell marker database (CellMatch) and RNA-seq profiles (MCA and HCL) were selected to construct the test datasets. Considering the few samples of main cell types (e.g. alpha, beta and delta cells) in the HCL pancreas dataset, another large human pancreas dataset by Baron *et al.* (37) was integrated with the HCL pancreas dataset as the reference for profile-dependent methods.

For scDeepSort, all cells from a particular tissue (Supplementary Table S1) were used to train the GNN-based deep learning model for cell-type prediction on the test cells that originated from the same tissue.

For CellAssign, external test datasets were first transformed as SingleCellExperiment objects with a normalized matrix. The CellMatch database containing tissue-specific cell markers was then used as a reference. All other parameters in CellAssign were kept as default.

For Garnett, marker genes from the CellMatch database were extracted and checked to train classifiers for each test dataset. The parameter of the number of unknown type cells was set as 50 by default during classification. Then, the trained classifiers were used to classify the cells for each test dataset.

For SingleR, external test datasets were annotated based on reference datasets (HCL and MCA) with default parameters, wherein the method was set to 'single'.

For scMap and CHETAH, external test datasets were transformed into SingleCellExperiment objects and annotated based on reference datasets (HCL and MCA) with default parameters.

For ACTINN, external test datasets and reference datasets from human and mouse cell atlases were transformed into .h5 objects and reference datasets (HCL and MCA) were used to annotate test datasets with default parameters.

For scID, reference datasets (HCL and MCA) were used to construct markers by the method of MAST in Seurat. Then, reference datasets and markers were used to annotate external test datasets with default parameters.

For SCINA, the CellMatch database containing tissue-specific cell markers was used as a reference to annotate external test datasets with default parameters.

For scPred, reference datasets (HCL and MCA) were transformed into scPred objects and trained with default parameters, which were then used to annotate external test datasets.

For singleCellNet, reference datasets (HCL and MCA) were trained with default parameters and used to annotate external test datasets with default parameters (nrand = 50 in 'get_cate').

For SVM, reference datasets (HCL and MCA) were trained with default parameters (the scikit-learn python package) over the same cell representations as the ones applied in scDeepSort and used to annotate external test datasets.

Testing scRNA-seq datasets using different training sets

To test MCA datasets and mouse external test datasets using TM datasets as the training set, seven TM datasets having the common tissues with MCA datasets (i.e., bladder, bone marrow, kidney, liver, lung, mammary gland and spleen) were used.

To test mouse external test datasets using TM datasets as the training set, 11 TM datasets having the common tissues with mouse external test datasets (i.e., bone marrow, kidney, liver, lung, and spleen) were used.

To test TM datasets using MCA datasets as the training set, seven MCA datasets having the common tissues with TM datasets (i.e., bladder, bone marrow, kidney, liver, lung, mammary gland, and spleen) were used.

Table 1. Comparison of scDeepSort with other methods

Methods	Property			Performance			
	Marker-dependent	Profile-dependent	Unsure cells	95% CI of mean *F*1 score	95% CI of mean MCC	Accuracy (265 489 cells)	Ratio of unsure cells
scDeepSort	-	√	√	0.47–0.68	0.48–0.69	83.79%	0.32%
CellAssign	√	-	-	0.13–0.36	0.14–0.39	13.36%	-
Garnett	√	-	√	0.07–0.18	0.03–0.24	17.38%	39.09%
SingleR	-	√	-	0.39–0.59	0.36–0.56	65.99%	-
scMap-cell	-	√	√	0.17–0.35	0.23–0.42	26.01%	66.20%
scMap-cluster	-	√	√	0.03–0.22	0.17–0.49	13.84%	84.15%
ACTINN	-	√	-	0.42–0.63	0.43–0.63	76.05%	-
CHETAH	-	√	√	0.33–0.54	0.32–0.54	63.45%	23.74%
scID	-	√	√	0.16–0.31	0.11–0.27	21.31%	9.00%
SCINA	√	-	√	0.27–0.48	0.24–0.46	39.34%	27.10%
scPred	-	√	√	0.18–0.40	0.22–0.48	42.16%	48.75%
singleCellNet	-	√	-	0.42–0.61	0.45–0.63	78.24%	-
SVM	-	√	-	0.18–0.36	0.14–0.41	46.41%	-

CI, confidence interval. MCC, Matthews correlation coefficient.

Accuracy, F1 score and MCC evaluation

For the accuracy in Figures 4 and 5, it is defined as the percentage of consistent cells with the same cell type, as in the literature, for each external test dataset and each tissue. For the accuracy in Table 1, it is the ratio of consistent cells from all tissues involving 265 489 cells. To calculate the *F*1 score and Matthews correlation coefficient (MCC), test datasets containing at least two cell types were selected, which generated 16 external test datasets across nine tissues for human and 11 external test datasets across six tissues for mouse. For each cell type in each test dataset, we calculated a *F*1 score and MCC by following equations:

$$F1\text{ score } = \frac{2*\mathrm{TP}}{2*\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

$$\mathrm{MCC} = \frac{\mathrm{TP}*\mathrm{TN}-\mathrm{FP}*\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})*(\mathrm{TP}+\mathrm{FN})*(\mathrm{TN}+\mathrm{FP})*(\mathrm{TN}+\mathrm{FN})}}$$

where TP, FP, FN, and TN are short for the true positives, the false positives, the false negatives, and the true negatives, respectively.

We calculated an *F*1 and MCC for each cell type within each tissue, and then aggregated them across cell types among all test datasets to obtain the 95% CI of mean *F*1 score and mean MCC (Table 1), while we calculated a mean *F*1 and MCC of all cell types for each test dataset and each tissue (Figures 4 and 5).

Statistics

R (version 3.6.1) and GraphPad Prism 8.0.1 were used for the statistical analysis.

RESULTS

General description of scDeepSort

In brief, we applied a supervised deep learning model based on a weighted GNN framework to build the scDeepSort model with underlying data of human and mouse single-cell transcriptomics atlases. First, dense representations, namely vectors, for cells and genes were obtained with dimensionality reduction methods to initialize fixed-size node embeddings, since the single-cell transcriptomics data are usually sparse matrices. Principal component analysis (PCA) was used to extract dense representations for gene nodes from the cell–gene data matrix and cell node representations were then calculated by the weighted sum of gene node representations and the cell–gene data matrix. Then, an undirected and weighted graph containing cell nodes and gene nodes was constructed from an adjacent weighted matrix by taking the gene expression as the weighted edges between cells and genes to model the intrinsic geometric information, which constitutes scDeepSort's first embedding layer (Figure 2).

In detail, the scDeepSort model's architecture consists of three components: the embedding layer, the weighted graph aggregator layer, and the linear classifier layer (Figure 2). The embedding layer, namely the input layer, generate vectors for graph nodes, i.e. cells and genes, which are initialized as previously described and are frozen during training. In the weighted graph aggregator layer, GraphSAGE (38) was applied as the backbone GNN framework and heavily modified in some aspects as detailed in Methods. The weighted graph aggregator layer gathers information about the neighborhood and itself during training for each subgraph, which produced a new representation for each node. The final linear classifier layer, namely the output layer, classifies the new cell representation produced from the weighted graph aggregator layer into one of the predefined cell type categories. For the cell node, once a label is predicted by the linear classifier layer, the loss between this prediction and the correct label is computed and then used to update the parameters of three layers until convergence.

Performance and robustness on internal datasets

In this study, a total of 562 977 human cells from 56 tissues and 201 764 mouse cells from 32 tissues which were collected from recent published HCL and MCA were curated to construct single-cell transcriptomic atlases (Supplementary Table S1). For each cell type, cells that were from cell types representing <0.5% of the cells were dropped. For each tissue in the human and mouse atlases, cells of various

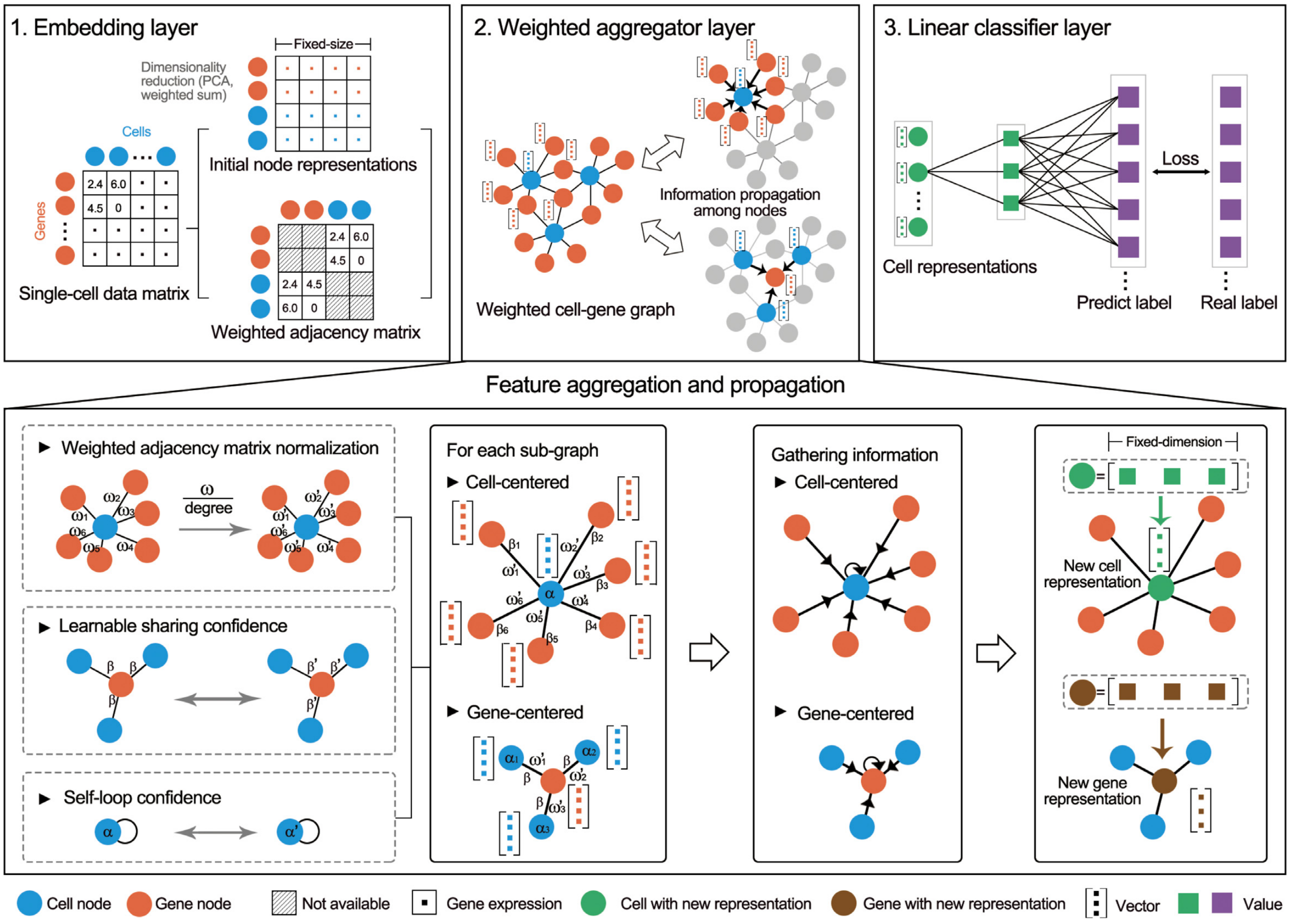

Figure 2. The weighted GNN algorithm of scDeepSort. The algorithm consists of embedding, weighted graph aggregator and linear classifier layers. The embedding layer stores the graph node representations and is frozen during training, wherein dimensionality reduction methods (PCA and weighted sums) were used to generate the initial fixed-size node representations, and the gene expression for each cell was regarded as the weighted edge between cells and genes forming a weighted adjacent matrix. In the weighted aggregator layer, a self-loop confidence for each cell node and a learnable sharing confidence for each gene node were incorporated into the weighted cell–gene graph. For each node (i.e. cell-centered and gene-centered) subgraph, weighted edges were normalized and the node itself and its neighborhood were then gathered to generate a new cell node representation during aggregation. The linear classifier layer categorizes the final cell state representation as a predefined cell type.

types were randomly divided into training and test sets, ensuring that the ratio of training and test cells was set to 8:2 for each cell type. In total, this process generated 443 566 training cells and 110 860 test cells from the HCL, 160 519 training cells and 40 112 test cells from the MCA. After supervised learning on training sets, we evaluated the scDeepSort's performance on HCL and MCA test sets with five replicates.

Across 56 human tissues and 32 mouse tissues, scDeepSort annotated the most cells, with the mean accuracy ranging from 85.20% to 99.11% for HCL and from 73.22% to 98.14% for MCA, respectively (Figure 3A). Among the common tissues (i.e., bladder, bone marrow, fetal lung, fetal brain, fetal intestine, fetal liver, fetal stomach, kidney, liver, lung, muscle, pancreas, peripheral blood, placenta, spleen, and stomach) between HCL and MCA, the accuracies between them were much similar across 16 tissues except for the fetal brain (Supplementary Figure S2).

Moreover, we have validated the robustness of the pipeline by running the experiments for each alternative split (8:2, 7:3, 6:4 and 5:5 of the training set and the test set). As shown in Figure 3B, we selected the splits of 8:2, 7:3, 6:4 and 5:5 to test the HCL and MCA test sets for each tissue with five replicates. Concordantly, these four kinds of splits showed close accuracies across 56 human tissues and 32 mouse tissues.

To get a better idea of what role the reference has in the performance, we have mapped the TM cells to the MCA annotation, which generated seven datasets having the common tissues with MCA (i.e., bladder, bone marrow, kidney, liver, lung, mammary gland, and spleen). Then we compared the performance of different training sets (TM or MCA) on the annotation of cell types, i.e. testing TM using MCA as the training set and testing MCA using TM as the training set (Figure 3C). Although both of them accurately annotated most cells across seven tissues (79% and 76% accuracies, respectively), accuracies of testing TM using MCA as the training set were slightly higher than that of testing MCA using TM as the training set except on the spleen dataset. We observed that there are 18 401 cells in TM datasets and 21 651 cells in MCA datasets involving these common tissues, which

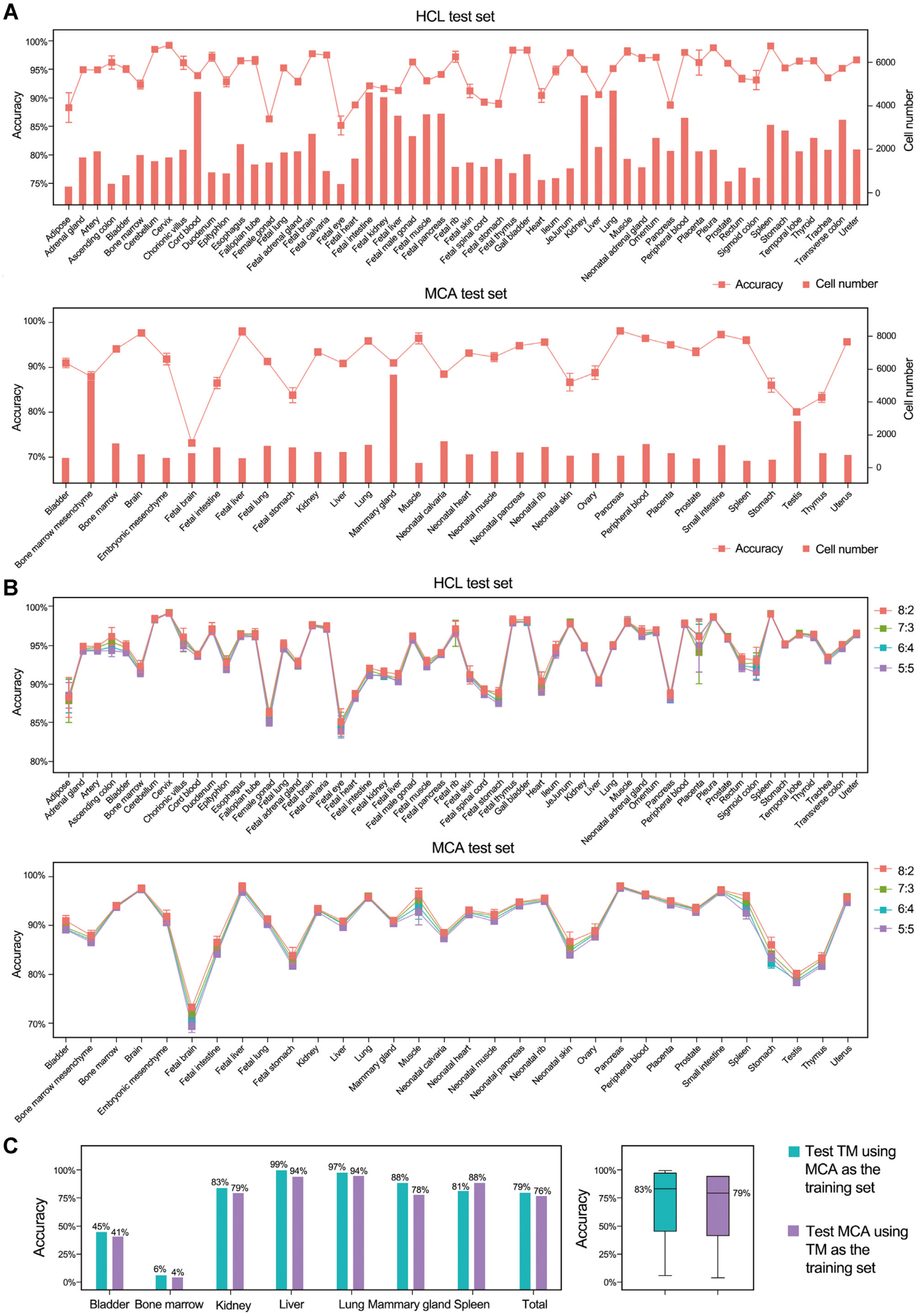

Figure 3. Performance and robustness of scDeepSort on internal test datasets and performance of different references on annotation of cell types. (**A**) Accuracy of scDeepSort in annotating the HCL test set across 56 tissues (line chart) and the MCA test set across 32 tissues with five replicates by the 8:2 split of the training set and the test set. The bar chart shows the number of test cells for each tissue. (**B**) Accuracy of scDeepSort in annotating the HCL and MCA test set by the split of 8:2, 7:3, 6:4 and 5:5 with five replicates for each alternative split. The data of the accuracy represents the mean ± SD. (**C**) Performance of different references on the annotation of cell types. The accuracy for each dataset was labelled for bar plot (left) and the median accuracy across seven datasets was labelled beside the box (right).

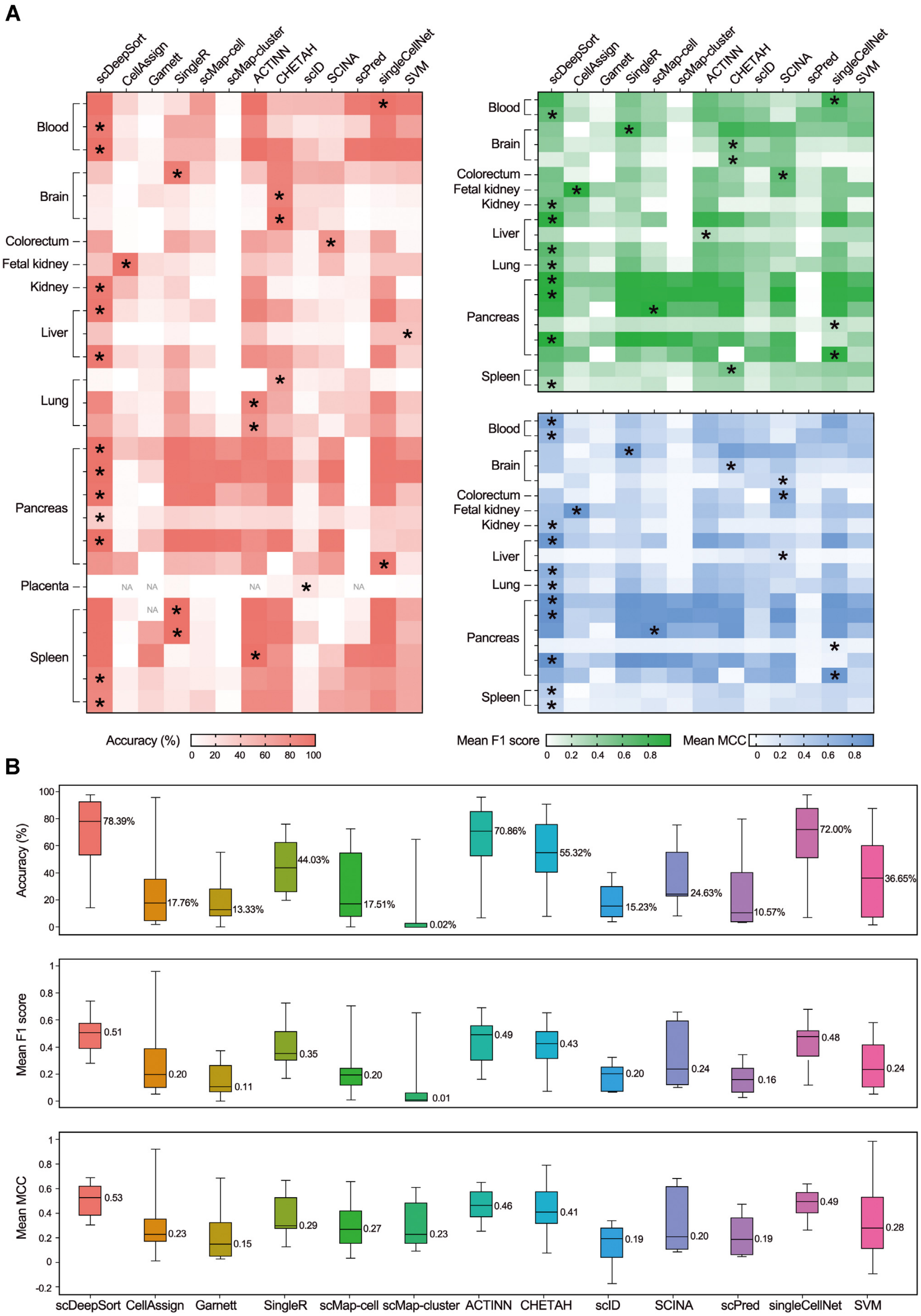

Figure 4. Performance comparison on human external test datasets. (**A**) Heatmaps of accuracies for scDeepSort, CellAssign, Garnett, SingleR, scMap-cell, scMap-cluster, ACTINN, CHETAH, scID, SCINA, scPred, singleCellNet and SVM on 27 external test datasets across 10 tissues and mean $F1$ scores and mean MCCs on 20 external test datasets across nine tissues. The asterisk represents the top-ranked method for each test dataset. NA, not available. (**B**) Boxplots summarized the maximal, minimal, median and quantile accuracies, mean $F1$ scores and mean MCCs for different methods among different tissues. The median value is labelled beside the corresponding box. MCC, Matthews correlation coefficient.

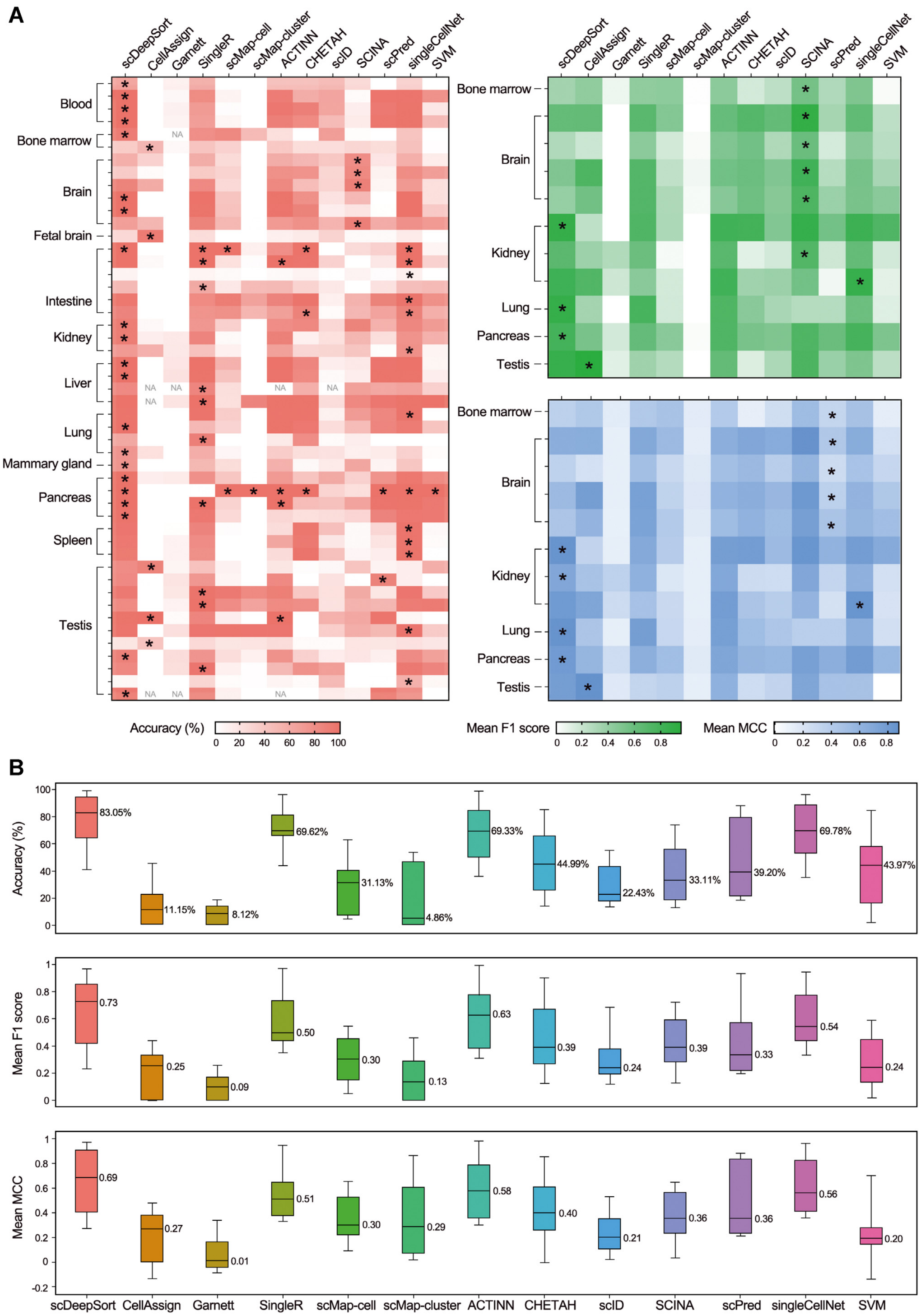

Figure 5. Performance comparison on mouse external test datasets. (**A**) Heatmaps of accuracies for scDeepSort, CellAssign, Garnett, SingleR, scMap-cell, scMap-cluster, ACTINN, CHETAH, scID, SCINA, scPred, singleCellNet and SVM on 49 external test datasets across 18 tissues and mean *F*1 scores and mean MCCs on 11 external test datasets across six tissues. The asterisk represents the top-ranked method for each test dataset. NA, not available. (**B**) Boxplots summarized the maximal, minimal, median and quantile accuracies, mean *F*1 scores and mean MCCs for different methods among different tissues. The median value is labelled beside the corresponding box. MCC, Matthews correlation coefficient.

might account for the slight superiority of testing TM using MCA as the training set because MCA has a larger sample size than TM.

Performance comparison of scDeepSort with other methods

We collected human and mouse scRNA-seq data manually from high-quality studies, forming 76 external test datasets (Supplementary Table S2) to comprehensively compare the performance of scDeepSort with other single-cell annotating methods (Table 1). For the marker-dependent methods (CellAssign, Garnett and SCINA), CellMatch (7) was used as the reference for annotation. CellMatch is a comprehensive and tissue-specific cellular taxonomy reference database providing a panel of 20,792 human and mouse marker genes involved with 184 tissue types and 353 cell types. For profile-dependent methods (SingleR, scMap, ACTINN, CHETAH, scID, scPred, singleCellNet and SVM), the HCL and MCA were used as references, wherein scDeepSort was systematically trained based on the same references (i.e. 562 977 human cells from 56 tissues and 201 764 mouse cells from 32 tissues). Considering the insufficient sample size of some cell types (e.g. alpha, beta and delta cells) in the HCL pancreas dataset, another large human pancreas dataset (37) by Baron *et al.* was integrated with the HCL pancreas dataset to pre-train scDeepSort. To ensure that all methods are able to predict the right cell identity, only cell types and subtypes that were recorded in both cell marker database (CellMatch) and RNA-seq profiles (HCL and MCA) were selected to establish external test datasets, which generated a total of 265 489 test cells for human and mouse (Supplementary Table S2).

Generally, scDeepSort accurately annotated most cell types and subtypes with an 83.79% accuracy across 265 489 human and mouse cells, outperforming all other methods (singleCellNet, 78.24%; ACTINN, 76.05%; SingleR, 65.99%; CHETAH, 63.45%; SVM, 46.41%; scPred, 42.16%; SCINA, 39.34%; scID, 21.31%; scMap-cell, 26.01%; Garnett, 17.38%; scMap-cluster, 13.84%; CellAssign, 13.36%, Table 1). Also, scDeepSort obtained the highest mean *F*1 score (0.47–0.68) and mean MCC ranging from 0.48 to 0.69 (Table 1).

Specifically, there are 27 human external test datasets containing a total of 130 885 cells involving 10 tissues: blood, brain, colorectum, fetal kidney, kidney, liver, lung, pancreas, placenta and spleen. By mapping the predicted cell label with the real one (Supplementary Table S3), we systematically compared the accuracy, mean F1 score, and mean Matthews correlation coefficient (MCC) of scDeepSort with other methods for each test dataset (Supplementary Table S4, Figure 4A), wherein datasets containing at least two cell types or subtypes were selected to calculate the F1 score and MCC. For each tissue, the result shows that scDeepSort was superior to other methods given the median accuracy (78.39%), mean *F*1 score (0.51) and mean MCC (0.53), followed by singleCellNet (72.00% median accuracy), ACTINN (70.86% median accuracy), CHETAH (55.32% median accuracy), SingleR (44.03% median accuracy), SVM (36.65% median accuracy) and SCINA (24.63% median accuracy) (Figure 4B). For other methods (i.e. CellAssign, Garnett, scMap, scID and scPred), they exhibited significant inadequacies in annotating human test datasets.

In the same manner, we also evaluated the performance of scDeepSort and other methods on annotating 49 external test datasets of mouse cells, which includes 134 604 cells from 12 tissues: blood, bone marrow, brain, fetal brain, intestine, kidney, liver, lung, mammary gland, pancreas, spleen and testis. Concordantly, scDeepSort showed an excellent performance among different datasets (Figure 5A), which also exhibited the highest median accuracy (83.05%) followed by singleCellNet (69.78%), SingleR (69.62%), ACTINN (69.33%), CHETAH (44.99%) and SVM (43.97%) among different tissues as shown in Figure 5B. However, Garnett and scMap-cluster seemed unable to distinguish the most cells' identities exhibiting a relatively poor performance on annotating mouse test datasets considering the accuracy, mean *F*1 score, and MCC (Supplementary Table S4, Figure 5B). The reason might be that Garnett need collect sufficient training samples from test datasets by using cell markers when training the classifier, and the low accuracy for Garnett may have been caused by too few training samples at the root of a cell type hierarchy leading to the missing of some cell types in the trained classifier. As for scMap-cluster, it uses the median gene expression of each cell cluster for similarity comparison with test cells, which might introduce some uncertainties when projecting test cells to reference cell clusters. Moreover, a selection-bias effect specific to the implementation could be the cause as well.

In short, the present results showed that scDeepSort can accurately predict most test cells with the top-ranked performance, wherein the performance of some methods is also quite well such as SingleR, singleCellNet, ACTINN and CHETAH. Indeed, some methods exhibit an unappealing performance in our benchmarking analysis, especially for marker-dependent ones, i.e. CellAssign, Garnett, and SCINA. The relatively poor performance of marker-dependent methods may be largely caused by the markers, as these methods heavily rely on the quality and integrality of selected marker genes. On the other hand, it is a common practice to use markers to annotate the cell type for pre-computed clusters rather than single cells. It has been reported that some cluster-based marker-dependent methods (e.g., scCATCH and SCSA) show decent performance in predicting the cell types of clusters (6,7). Unlike the cluster-based methods, CellAssign, Garnett and SCINA predict the cell type for each cell instead of the cluster. As a consequence, the poor performance of CellAssign, Garnett, and SCINA might result from the poor extrapolation of markers on cell-based marker-dependent cell typing methods.

Although CellAssign, scmap, scID and SCINA performed worse than other methods across 265 489 cells, it's worth to point out that they exhibited significant performance on some individual test datasets. For example, CellAssign obtained the highest accuracy among all methods over the human fetal kidney dataset and the mouse fetal brain dataset, etc., while scID is the top-ranked method over the human placenta dataset. On annotating the second mouse pancreas dataset, both scMap-cell and scMap-cluster perfectly classified the most cells' identities. For

SCINA, as shown in Figure 5, it exhibited the highest performance over the most mouse brain datasets.

Moreover, to demonstrate the universality of scDeepSort, we first used it to annotate three more challenging atlases (Supplementary Figure S3). For DropViz, almost all cells involving astrocytes, microglia/macrophage, and oligodendrocytes were assigned into the correct category. For the oncoscape dataset, scDeepSort accurately assigned neurons and neuronal progenitor cells as the label for most neuron-related cells (i.e. inhibitory neurons, inhibitory interneurons, inhibitory neuron progenitor, postmitotic premature neurons, sensory neurons, granule neurons and neuronal progenitor cell) since these subtypes of neurons hardly exist in the training set of MCA. It is noted that some excitatory neurons were predicted as ganglion cells. In fact, the ganglion cell, namely the retinal neurons, is classified by its location in the tissue, while the excitatory neuron is classified by the function. Therefore, it is hard to compare the relation between them. For the loom dataset, more than half of neuron-related cells (neurons, granule neurons and hippocampus neurons) were predicted as neural progenitor cells and postmitotic neurons, while near half neurons were predicted as the unsure type. For these more challenging atlases, scDeepSort can accurately predict the main cell type for most cells except the partial neurons in the Loom dataset.

In addition, we mapped 49 external mouse test datasets to the TM datasets, wherein 11 external test datasets having the common tissues and the corresponding cell types with TM datasets were included. Concordantly, the results showed that TM-based models also accurately annotated the most test cells among 11 datasets reaching a 95% accuracy (Supplementary Figure S4), suggesting the universality of our weighted GNN-based deep learning model in cell-type classification on scRNA-seq data with Drop-Seq considering the multiple scRNA-seq technologies (39).

It is known that an important feature of cell typing methods is to be able to assign a cell into an unsure category considering the unseen cell types in the training set. Like most methods including Garnett, scMap, CHETAH, scID, SCINA and scPred, scDeepSort also has this ability, while CellAssign, SingleR, ACTINN and singleCellNet do not. Among 265 489 test cells, scDeepSort assigned 0.32% cells as the unsure type, while the ratio of unsure cells seems too high for other methods, i.e. scMap-cluster (84.15%), scMap-cell (66.20%), scPred (48.75%) and Garnett (39.09%), which might also account for the unappealing outcomes of these methods to a certain extent.

For scDeepSort, it is mainly pre-trained based on GPU consuming a small percentage of the resource during training (Supplementary Table S5), i.e. about 8% memory (∼5G) and 5% video memory (∼2.4G), wherein training time is within minutes as the examples in Supplementary Figure S4. Considering that other methods are dependent on CPU, a server without GPU was selected to fairly evaluate the computation time for all methods on annotating external test datasets by varying the number of cells and genes for profile-dependent methods and varying the number of cell types and markers for marker-dependent methods (Supplementary Table S5). For CellAssign, SingleR, scMap, ACTINN, CHETAH, scID and SCINA, they train and predict simultaneously, the computation time depends on the training time. Although Garnett, scPred and singleCellNet train and predict separately, the computation time for prediction is very short despite the large-scale test datasets (Supplementary Table S5). Obviously, the computation time for profile-dependent methods (i.e. SingleR, scMap-cell, ACTINN, CHETAH, scID, scPred and singleCellNet) is positively correlated with the size of the training set, while the computation time for marker-dependent methods (i.e. CellAssign, Garnett, and SCINA) is influenced by the number of markers and cell types. Compared to other methods, the pre-training is an advantage of scDeepSort which saves computation time. Therefore, the computation time of scDeepSort lies in the prediction as it needs construction of cell–gene graph network and adds the new test cells into the graph network. As shown in Supplementary Table S5, the computation time of scDeepSort over six test datasets is within minutes enabling the fast annotation of scRNA-seq datasets.

DISCUSSION

In this study, we developed a scalable cell-type annotation tool for single-cell transcriptomics data by using a pre-trained deep learning model with a GNN model for the first time. From human and mouse scRNA-seq datasets, scDeepSort was able to be able to annotate most cells under the context of a specific organ, significantly outperforming known methods and the classical machine learning method (SVM). Actually, what makes deep learning different from standard machine learning is not the count of layers but how the information being processed and transformed over times (40,41). As one of the most famous deep learning methods, the advantage of GNN lies in the ability of artificial intelligence techniques to integrate big data, incorporate existing knowledge, and learn arbitrarily complex relationships, which has achieved the state-of-the-art accuracy on numerous prediction tasks in the computer science (26,27).

In fact, the accuracy of our designed weighted GNN-based scDeepSort improves 13% and 6% in predicting cell types and subtypes for internal datasets and external human pancreas datasets, respectively, compared to the traditional GNN-based deep learning model (Supplementary Table S6). Moreover, on three more challenging single-cell atlases (33–35), scDeepSort also accurately predicts the main cell type (Supplementary Figure S3) for most cells except the partial neurons in the Loom dataset (35). The present results indicated the superiority of our weighted GNN-based deep learning model in processing big data like high-throughput scRNA-seq data and in prediction.

An important feature of cell typing methods is to be able to assign a cell into an unsure category considering the unseen cell types in the training set. Like most methods, scDeepSort is allowed to predict the unsure type. Besides the external test datasets, a negative control was included revalidating the predicted unsure type by scDeepSort using samples from different tissues (Supplementary Figure S5). Nevertheless, a good cell-type annotation method is to accurately identify the cell type rather than assigning an unsure type in practice. Limited training datasets would directly influence the cell-type annotation, e.g. generating the unsure type, especially for these cell subtypes without sufficient training data. Indeed, some rare or intermediary cell types without sufficient training data were dropped causing the inability to annotate them, which is a common limitation for all single-cell annotation methods. However, increasing scRNA-seq studies will enable the expansion and perfection of atlases across the two species, especially for the rare or intermediary cell types. Future integration of HCL, MCA, and other known large-scale single-cell atlases containing the rare or intermediary cell types with sufficient training data will effectively address this issue. Hence, we have provided an extra function on GitHub for users to train a new model using our proposed weighted GNN method with their own reference datasets containing the rare or intermediary cell types of interest, given the possible rare or intermediary cell types which are not present in our pre-trained models. Under this scenario, users are able to annotate the rare or intermediary cell types on a new dataset.

Interestingly, it seems easy to map cells across the human pancreas as reported in the recent benchmark analysis (18), wherein the widely-used human pancreas datasets (37,42–44) were tested. Concordantly, SVM and most methods also achieved good performance on the human pancreas datasets in our study. However, SVM and some methods indeed exhibited an unappealing performance on the enlarged datasets and tissues. It is noted that it is the first time that the most comprehensive benchmarking datasets with 265 489 cells involving 22 tissues in humans and mice were introduced to compare the performance of different cell-typing methods generating the unbiased conclusion. Above all, we have systematically and comprehensively demonstrated the ground-breaking and robust performances of our proposed weighted GNN model on the internal and external benchmarking datasets. The key point of our work is that the top-ranked performance for the accurate cell-type annotation, which is the most important part in the application scenario of scRNA-seq (45).

In addition, we tried to provide the interpretability of our weighed GNN models using the GNNExplainer (46), a model-agnostic approach for providing interpretable explanations for predictions of GNN-based models, to generate possible explanations. Taking the macrophage of human liver datasets as the example, we used GNNExplainer to identify a compact subgraph structure and a small subset of node features that have a crucial role in GNN's prediction as shown in Supplementary Figure S6A below, which identifies the important graph structure and 59 important node features for the macrophage. Among the 59 node features, most of them (95%) have a significantly high expression (Z score > 1.5) across macrophages (Supplementary Figure S6B), wherein six node features are macrophages' markers (32,47,48), i.e., TMSB4X, RPL10, RPL11, S100A4, CD52 and GZMA (Supplementary Figure S6C). The results of GNNExplainer indicated that the highly expressed node features including marker genes might play a more important role in the information aggregation and propagation in the constructed cell-gene graph during iterations enabling learning better cell representations, which significantly improve the performance of the GNN model.

Despite of the great success on many fields, deep learning models, are commonly believed as black boxes. It is well-known that it is easy to interpret the machine learning methods as most of them are shallow. However, the interpretability is the common failing and difficulty for the majority of deep learning models, which is also a great challenge in the field of computer science (27,49). Also, there still lack of efficient methods for the interpretability of GNN models, although the GNNExplainer can partly generate the possible explanations for them. Future improvements are supposed to be made in the science community on the biological interpretation of the GNN models, which will substantially facilitate the development of biological research and biomedical applications.

DATA AVAILABILITY

No new data was generated for this study. All data used in this study is publicly available as previously described. scDeepSort is available as a python package (https://github.com/ZJUFanLab/scDeepSort) and the source code and results of comparison with other methods are available at github (https://github.com/ZJUFanLab/scDeepSort_performance_comparison).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by Alibaba Cloud.

FUNDING

National Natural Science Foundation of China [81973701, 91846204/U19B2027]; Natural Science Foundation of Zhejiang Province [LZ20H290002]; National Youth Top-notch Talent Support Program [W02070098]. Funding for open access charge: National Natural Science Foundation of China [81973701]; Natural Science Foundation of Zhejiang Province [LZ20H290002].

Conflict of interest statement. None declared.

REFERENCES

1. Macosko,E.Z., Basu,A., Satija,R., Nemesh,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A.R., Kamitaki,N., Martersteck,E.M. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
2. Klein,A.M., Mazutis,L., Akartuna,I., Tallapragada,N., Veres,A., Li,V., Peshkin,L., Weitz,D.A. and Kirschner,M.W. (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, **161**, 1187–1201.
3. Shao,X., Lu,X., Liao,J., Chen,H. and Fan,X. (2020) New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein Cell*, **11**, 866–880.
4. Liao,J., Lu,X., Shao,X., Zhu,L. and Fan,X. (2021) Uncovering an organ's molecular architecture at single-cell resolution by spatially resolved transcriptomics. *Trends Biotechnol.*, **39**, 43–58.
5. Shao,L., Xue,R., Lu,X., Liao,J., Shao,X. and Fan,X. (2021) Identify differential genes and cell subclusters from time-series scRNA-seq data using scTITANS. *Comput. Struct. Biotechnol. J.*, **19**, 4132–4141.
6. Cao,Y., Wang,X. and Peng,G. (2020) SCSA: a cell type annotation tool for single-cell RNA-seq data. *Front Genet*, **11**, 490.
7. Shao,X., Liao,J., Lu,X., Xue,R., Ai,N. and Fan,X. (2020) scCATCH: automatic annotation on cell types of clusters from single-cell RNA sequencing data. *iScience*, **23**, 100882.

8. Aran,D., Looney,A.P., Liu,L., Wu,E., Fong,V., Hsu,A., Chak,S., Naikawadi,R.P., Wolters,P.J., Abate,A.R. *et al.* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.*, **20**, 163–172.
9. de Kanter,J.K., Lijnzaad,P., Candelli,T., Margaritis,T. and Holstege,F.C.P. (2019) CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing. *Nucleic Acids Res.*, **47**, e95.
10. Kiselev,V.Y., Yiu,A. and Hemberg,M. (2018) scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods*, **15**, 359–362.
11. Boufea,K., Seth,S. and Batada,N.N. (2020) scID uses discriminant analysis to identify transcriptionally equivalent cell types across single-cell RNA-Seq data with batch effect. *iScience*, **23**, 100914.
12. Alquicira-Hernandez,J., Sathe,A., Ji,H.P., Nguyen,Q. and Powell,J.E. (2019) scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.*, **20**, 264.
13. Ma,F. and Pellegrini,M. (2020) ACTINN: automated identification of cell types in single cell RNA sequencing. *Bioinformatics*, **36**, 533–538.
14. Zhang,A.W., O'Flanagan,C., Chavez,E.A., Lim,J.L.P., Ceglia,N., McPherson,A., Wiens,M., Walters,P., Chan,T., Hewitson,B. *et al.* (2019) Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat. Methods*, **16**, 1007–1015.
15. Pliner,H.A., Shendure,J. and Trapnell,C. (2019) Supervised classification enables rapid annotation of cell atlases. *Nat. Methods*, **16**, 983–986.
16. Zhang,Z., Luo,D., Zhong,X., Choi,J.H., Ma,Y., Wang,S., Mahrt,E., Guo,W., Stawiski,E.W., Modrusan,Z. *et al.* (2019) SCINA: a semi-supervised subtyping algorithm of single cells and bulk samples. *Genes*, **10**, 531.
17. Tan,Y. and Cahan,P. (2019) SingleCellNet: a computational tool to classify single cell RNA-Seq data across platforms and across species. *Cell Syst.*, **9**, 207–213.
18. Abdelaal,T., Michielsen,L., Cats,D., Hoogduin,D., Mei,H., Reinders,M.J.T. and Mahfouz,A. (2019) A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol.*, **20**, 194.
19. Lahnemann,D., Koster,J., Szczurek,E., McCarthy,D.J., Hicks,S.C., Robinson,M.D., Vallejos,C.A., Campbell,K.R., Beerenwinkel,N., Mahfouz,A. *et al.* (2020) Eleven grand challenges in single-cell data science. *Genome Biol.*, **21**, 31.
20. Gibney,E. (2015) DeepMind algorithm beats people at classic video games. *Nature*, **518**, 465–466.
21. Silver,D., Huang,A., Maddison,C.J., Guez,A., Sifre,L., van den Driessche,G., Schrittwieser,J., Antonoglou,I., Panneershelvam,V., Lanctot,M. *et al.* (2016) Mastering the game of Go with deep neural networks and tree search. *Nature*, **529**, 484–489.
22. Zhang,X., Peng,X., Han,C., Zhu,W., Wei,L., Zhang,Y., Wang,Y., Zhang,X., Tang,H., Zhang,J. *et al.* (2019) A unified deep-learning network to accurately segment insulin granules of different animal models imaged under different electron microscopy methodologies. *Protein Cell*, **10**, 306–311.
23. Wainberg,M., Merico,D., Delong,A. and Frey,B.J. (2018) Deep learning in biomedicine. *Nat. Biotechnol.*, **36**, 829–838.
24. Lv,H., Dao,F.Y., Zhang,D., Guan,Z.X., Yang,H., Su,W., Liu,M.L., Ding,H., Chen,W. and Lin,H. (2020) iDNA-MS: an integrated computational tool for detecting DNA modification sites in multiple genomes. *iScience*, **23**, 100991.
25. Chaudhary,K., Poirion,O.B., Lu,L. and Garmire,L.X. (2018) Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin. Cancer Res.*, **24**, 1248–1259.
26. Wu,Z., Pan,S., Chen,F., Long,G., Zhang,C. and Yu,P.S. (2021) A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.*, **32**, 4–24.
27. Jie,Z.A., Gc,A., Sh,A., Zz,A., Cheng,Y.B., Zl,A., Lw,C., Cl,C. and Ms,A.J.A.O. (2020) Graph neural networks: a review of methods and applications. *AI Open*, **1**, 57–81.
28. Alavi,A., Ruffalo,M., Parvangada,A., Huang,Z. and Bar-Joseph,Z. (2018) A web server for comparative analysis of single-cell RNA-seq data. *Nat. Commun.*, **9**, 4768.
29. Johansen,N. and Quon,G. (2019) scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. *Genome Biol.*, **20**, 166.
30. Wang,J., Ma,A., Chang,Y., Gong,J., Jiang,Y., Qi,R., Wang,C., Fu,H., Ma,Q. and Xu,D. (2021) scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat. Commun.*, **12**, 1882.
31. Han,X., Zhou,Z., Fei,L., Sun,H., Wang,R., Chen,Y., Chen,H., Wang,J., Tang,H., Ge,W. *et al.* (2020) Construction of a human cell landscape at single-cell level. *Nature*, **581**, 303–309.
32. Han,X., Wang,R., Zhou,Y., Fei,L., Sun,H., Lai,S., Saadatpour,A., Zhou,Z., Chen,H., Ye,F. *et al.* (2018) Mapping the mouse cell atlas by microwell-Seq. *Cell*, **173**, 1307.
33. Saunders,A., Macosko,E.Z., Wysoker,A., Goldman,M., Krienen,F.M., de Rivera,H., Bien,E., Baum,M., Bortolin,L., Wang,S. *et al.* (2018) Molecular diversity and specializations among the cells of the adult mouse brain. *Cell*, **174**, 1015–1030.
34. Cao,J., Spielmann,M., Qiu,X., Huang,X., Ibrahim,D.M., Hill,A.J., Zhang,F., Mundlos,S., Christiansen,L., Steemers,F.J. *et al.* (2019) The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, **566**, 496–502.
35. La Manno,G., Siletti,K., Furlan,A., Gyllborg,D., Vinsland,E., Mossi Albiach,A., MattssonLangseth,C., Khven,I., Lederer,A.R., Dratva,L.M. *et al.* (2021) Molecular architecture of the developing mouse brain. *Nature*, **596**, 92–96.
36. Tabula,Muris Consortium, Overall,coordination, Logistical,coordination, Organ,collection and processing, Library,preparation and sequencing, Computational,data analysis, Cell,type annotation, Writing,group, Supplemental,text writing group and Principal,investigators (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, **562**, 367–372.
37. Baron,M., Veres,A., Wolock,S.L., Faust,A.L., Gaujoux,R., Vetere,A., Ryu,J.H., Wagner,B.K., Shen-Orr,S.S., Klein,A.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**, 346–360.
38. Hamilton,W.L., Ying,R. and Leskovec,J. (2017) Inductive representation learning on large graphs. *Adv. Neural Inf. Process. Syst.*, **30**, 1025–1035.
39. Chen,W., Zhao,Y., Chen,X., Yang,Z., Xu,X., Bi,Y., Chen,V., Li,J., Choi,H., Ernest,B. *et al.* (2020) A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. *Nat. Biotechnol.*, https://doi.org/10.1038/s41587-020-00748-9.
40. Bini,S.A. (2018) Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J. Arthroplasty*, **33**, 2358–2361.
41. Zou,J., Huss,M., Abid,A., Mohammadi,P., Torkamani,A. and Telenti,A. (2019) A primer on deep learning in genomics. *Nat. Genet.*, **51**, 12–18.
42. Xin,Y., Kim,J., Okamoto,H., Ni,M., Wei,Y., Adler,C., Murphy,A.J., Yancopoulos,G.D., Lin,C. and Gromada,J. (2016) RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab.*, **24**, 608–615.
43. Segerstolpe,A., Palasantza,A., Eliasson,P., Andersson,E.M., Andreasson,A.C., Sun,X., Picelli,S., Sabirsh,A., Clausen,M., Bjursell,M.K. *et al.* (2016) Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.*, **24**, 593–607.
44. Muraro,M.J., Dharmadhikari,G., Grun,D., Groen,N., Dielen,T., Jansen,E., van Gurp,L., Engelse,M.A., Carlotti,F., de Koning,E.J. *et al.* (2016) A single-cell transcriptome atlas of the human pancreas. *Cell Syst.*, **3**, 385–394.
45. Pasquini,G., Rojo Arias,J.E., Schafer,P. and Busskamp,V. (2021) Automated methods for cell type annotation on scRNA-seq data. *Comput. Struct. Biotechnol. J.*, **19**, 961–969.
46. Ying,R., Bourgeois,D., You,J.X., Zitnik,M. and Leskovec,J. (2019) GNNExplainer: generating explanations for graph neural networks. *Adv. Neur. In*, **32**, 9240.
47. Medaglia,C., Giladi,A., Stoler-Barak,L., De Giovanni,M., Salame,T.M., Biram,A., David,E., Li,H., Iannacone,M., Shulman,Z. *et al.* (2017) Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science*, **358**, 1622–1626.
48. Jaitin,D.A., Kenigsberg,E., Keren-Shaul,H., Elefant,N., Paul,F., Zaretsky,I., Mildner,A., Cohen,N., Jung,S., Tanay,A. *et al.* (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, **343**, 776–779.
49. Miotto,R., Wang,F., Wang,S., Jiang,X. and Dudley,J.T. (2018) Deep learning for healthcare: review, opportunities and challenges. *Brief. Bioinform.*, **19**, 1236–1246.