

METHODS

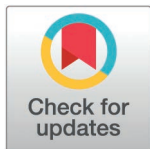
ScanNet: Single-cell annotation informed by transcriptional regulation Network via iterative heterogeneous graph learning

Yongyu Long¹, Wenhao Zhang¹, Lan Cao¹, Xiaobing Huang^{2*}, Ying Wang^{1*}

1 Department of Automation, National Institute for Data Science in Health and Medicine, State Key Laboratory of Mariculture Breeding, Xiamen Key Laboratory of Big Data Intelligent Analysis and Decision, Xiamen University, Xiamen, Fujian, China, **2** Department of Medical Oncology, Fuzhou First Hospital Affiliated with Fujian Medical University, Fuzhou, Fujian, China

☞ These authors are Equally contribution to this work.

* 13609522917@139.com (XH); wangying@xmu.edu.cn (YW)



Abstract

Accurate annotation of cell types in single-cell transcriptome sequencing (scRNA-seq) data is critical for understanding cellular identities. The transcriptional regulatory networks (TRNs), which map the regulatory relationships between transcription factors (TFs) and their target genes (TGs), capture the molecular dependencies underlying transcriptional programs. However, most existing cell type annotation methods do not fully exploit this regulatory information. Therefore, we introduce **ScanNet**, a Single cell annotation method informed by transcriptional regulation Network, to integrate prior knowledge of TRN into data of gene expression and capture the cell-type-specific characteristics underlying TRN mechanism. TRN can be naturally represented as heterogeneous graphs consisting of two regulatory elements, TFs and TGs connected by directed edges, thereby encoding the regulatory dependencies that shape transcriptional programs and ultimately determine cellular identity. To leverage this structure, ScanNet introduces an iterative heterogeneous graph convolutional framework that learns both local and global cellular embeddings through a dual-channel encoder. The Regulation-level Encoder applies iterative heterogeneous graph convolution to capture local TF-TG regulatory interactions within TRN, while the Expression-level Encoder learns global cellular transcriptional states. By integrating the multiple-view representations, ScanNet can accurately annotate cell types. Comprehensive evaluations across eight scRNA-seq datasets spanning different species, sample scales, and sequencing platforms demonstrate that ScanNet consistently outperforms ten state-of-the-art cell type annotation methods. By embedding prior TRN structures into a heterogeneous graph, ScanNet also achieves robust performance in cross-platform cell type annotation and in identifying novel cell types under constrained structural information. Moreover, the ScanNet framework can be flexibly transferred to single-cell ATAC-seq (scATAC-seq) data by mapping chromatin

OPEN ACCESS

Citation: Long Y, Zhang W, Cao L, Huang X, Wang Y (2026) ScanNet: Single-cell annotation informed by transcriptional regulation Network via iterative heterogeneous graph learning. PLoS Comput Biol 22(8): e1014602. <https://doi.org/10.1371/journal.pcbi.1014602>

Editor: Pingzhao Hu, Western University, CANADA

Received: October 14, 2025

Accepted: July 19, 2026

Published: August 6, 2026

Copyright: © 2026 Long et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The code used for model training and evaluation is available on GitHub (<https://github.com/Ying-Lab/ScanNet>).

Funding: This work was supported by National Natural Science Foundation of China (62573367 to YW), National Key Research and Development Program of

China (2025YFA0922603 to YW), and Fuzhou Inter-institutional Science and Technology Cooperation Project (2024-Y-018 to XH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

accessibility to gene level, where it achieves superior performance compared to existing annotation tools. Overall, ScanNet is a scalable, transferable, and mechanistically informed framework for accurate cell type annotation across diverse single-cell data modalities.

Author summary

Accurate cell type annotation is essential for deciphering cellular heterogeneity. Most existing annotating methods are based on gene expression profile from scRNA-seq data. However, behind the observed gene expression profile, gene regulation is the essential mechanism to drive the transcriptional progress, which offer more intrinsic biological insight. While most current annotation approaches don't incorporate regulatory information, limiting their capacity to capture the mechanistic basis of cell states. Therefore, we incorporate transcriptional regulatory networks (TRNs) as prior knowledge to assist more accurate cell type annotation. We introduce ScanNet, an iterative heterogenous graph learning framework that incorporates TRNs as both prior knowledge and structural constraints to capture biologically interpretable, cell-type-specific characteristics. In benchmark evaluations with ten baselines on multiple datasets, ScanNet shows (1) superior annotation performance on scRNA-seq data, (2) robust performance across seven sequencing platforms, (3) capability to identify novel cellular populations beyond existing annotations, and (4) effective transferability to scATAC-seq data.

Introduction

Single-cell RNA sequencing (scRNA-seq) techniques enable the measurement of gene expression profiles at single-cell resolution, providing a powerful tool for revealing cellular heterogeneity and transcriptional states [1–4]. This cellular resolution necessitates accurate cell type annotation to properly interpret cell-type-specific molecular mechanisms [5–7]. However, the inherently sparse, high drop-out rate and noisy nature of scRNA-seq data poses significant challenges to cell annotation.

Traditional manual approaches map marker genes with unsupervised clusters and annotate cells manually. While offering high biological interpretability, manual cell annotation is time-consuming, subjective, and poorly scalable, making it unreliable for large and complex single-cell datasets. To improve efficiency, computational methods start using similarity metrics to assess resemblance between query cells and reference datasets, automatically annotating the queried cells. For example, scmap [8] maps the cells with reference dataset based on options of Cosine similarity, Pearson correlation and Spearman correlation. CHETAH [9] constructs a hierarchical classification tree, where Spearman correlation and silhouette scores are used to assess

the similarity between cells. However, these similarity-based approaches are susceptible to batch effects and the quality of reference datasets, showing fluctuated performance.

With the development of machine learning and deep learning, more supervised methods have emerged to improve annotation accuracy. These methods typically rely on feature selection or feature mapping, which subsequently serve as the basis for classification. For example, scID [10] selects feature genes by employing Fisher's linear discriminant analysis to estimate gene weights; CaSTLe [11] and XGBoost [12] rely on tree-based ensemble methods; and scPred [13] employs a support vector machine to project gene expressions into a latent feature space. Furthermore, superCT [14] and ACTINN [15] utilize multi-layer neural networks to embed gene expressions, thereby reducing the need for feature selection; scGCN [16] constructs cell similarity graph and uses graph convolutional network (GCN) to facilitate cross-dataset cell annotation; scGraphformer [17] combines GCN with Transformer to uncover cellular heterogeneity.

However, due to individual variability and sequencing-related technical noise, the methods that rely solely on expression data still lack generalizability across datasets. Accordingly, biological networks are introduced as structural priors to uncover the biological essence and enhance model generalization. sigGCN [18] incorporates protein interaction network from String [19] by utilizing Chebyshev Polynomial-based variant of GCN to learn cell heterogeneity representations. Similarly, scGraph [20] integrates protein interaction networks by GraphSAGE [21]. scRGCL [22] utilizes a residual GCN and contrastive learning to optimize cell feature representation and enhance the discrimination of different cell types. While these methods attempt to capture molecular dependencies among genes, correlation-based interactions still fail to encapsulate causal regulatory relationships. The transcriptional regulatory network (TRN), in contrast, represents the directional relationships from transcription factors (TFs) to target genes (TGs) that drive gene expression, thereby describing the essential mechanisms of cellular heterogeneity.

Therefore, in this study, we propose **ScanNet**, a Single cell annotation method informed by transcriptional regulation Network via iterative heterogeneous graph learning. TRN can be naturally represented by a heterogeneous graph with two types of regulatory elements, where TFs modulate TG expressions by binding to promoter or enhancer regions, either activating or repressing transcription [23–27]. To leverage this structure, ScanNet introduces an iterative heterogeneous graph convolutional (iHGCN) framework that learns both local and global cellular embeddings through a dual-channel encoder consisting of a Regulation-level Encoder and an Expression-level Encoder. Specifically, the Regulation-level Encoder applies iHGCN to capture local TF-TG molecular interactions within cells and learn biologically-informed Regulation-level Embeddings that fusing regulatory information and gene expression. While the Expression-level Encoder learns global cellular transcriptional states from overall cellular expression by neural networks. By integrating these two complementary representations, ScanNet can accurately annotate cell types.

We conduct extensive experiments to demonstrate the superiority of our method over state-of-the-art baselines under four scenarios. Firstly, evaluations across eight scRNA-seq datasets that varied from species, sample scales, and sequencing platforms demonstrate the superiority of ScanNet over baseline methods in basic annotation task. By introducing TRN as basis structure and prior constraint, ScanNet achieves robust annotation performance in cross-platform datasets despite technical noise and batch effects. In addition, ScanNet can automatically identify previously unseen cell types, also referred as novel types through entropy-based uncertainty quantification in open-world scenario. Finally, the ScanNet framework can be flexibly transferred to single-cell ATAC-seq data by mapping chromatin accessibility to gene level, where it achieves superior performance compared to existing annotation tools. Overall, by integrating cellular transcriptional states with underlying regulatory mechanisms through iHGCN, ScanNet not only effectively captures the fundamental signatures of cell states but also deeply utilizes the causal linkages between transcriptional regulation and gene expression dynamics, providing regulatory perspective-based insights that advance our understanding of the mechanistic drivers of cellular heterogeneity.

Results

The framework of ScanNet

ScanNet is a new solution driven by both regulatory mechanisms and data characteristics for scRNA-seq data annotation (Fig 1a). ScanNet takes scRNA-seq data as input and reference TRN as basis regulatory structure to construct a heterogeneous graph that simultaneously models gene expression profiles and regulatory mechanisms. Specifically, the reference TRN is modeled as a heterogeneous graph by distinguishing TF and TG nodes and their regulatory interactions. The regulatory intensity is quantified as edge weights, with 1 assigned to activating regulation and -1 to repressive regulation. The TRN further guides the selection of relevant genes from the scRNA-seq data, yielding TF and TG expression features for model input (S1 Text). With the prior constraint, ScanNet is able to capture cellular heterogeneity from both molecular mechanisms and transcriptional state through a dual-channel encoder with iterative graph convolution. The dual-channel encoder comprises a Regulation-level Encoder, which is responsible to extract causal regulatory dependencies at the molecular level through iHGCN (Fig 1b), and an Expression-level Encoder that learns global transcriptional states at the cellular level. By embedding regulatory dependencies, ScanNet generates biologically-informed cellular representations that not only enhance annotation accuracy but also improve generalization and applicability to diverse annotation tasks, like cross-platform annotation, novel cell type discovery, and transferability to single-cell ATAC-seq (scATAC-seq).

Dataset description and experimental design

We evaluate the performance of ScanNet against ten baseline methods across eight scRNA-seq datasets for the cell annotation task. Benchmark datasets include four human datasets, the paired multimodal PBMC dataset [28], Muraro dataset [29] derived from human pancreas, BMMC dataset [30] from bone marrow and Immune dataset from multi-tissue immune system [31], as well as four mouse datasets, BaronMouse dataset [32] from mouse pancreas, ChenMouse dataset [33] and AMB dataset [34] from mouse visual cortex, and TM dataset [35] from mouse musculus. For each benchmark dataset, cell-type labels are manually curated by domain experts and treated as the gold standard for a unified evaluation of automated annotation methods. Although datasets are derived from different species, the reference TRN [24] is unified which collects TF-TG regulations from multiple cell types and tissues, covering both human and mouse data. These datasets vary from tissues, sequencing platforms, species as well as the number of TFs, TGs, cells and cell types, as shown in Table 1. This diversity ensures a comprehensive assessment of performance across various biological contexts and technical challenges in cell classification. To mitigate the impact of extremely rare cell types on classification performance, scRNA-seq-based cell annotation experiments filter out cell types with fewer than 10 cells. The excluded rare cell types can subsequently serve as novel cell types to evaluate the model's novel type detection capability.

scRNA-seq data generated by different sequencing platforms exhibit systematic technical variations that pose challenges for cell annotation [36]. To evaluate the generalization of ScanNet across different sequencing platforms, another dataset (PBMCbench [37]) containing scRNA-seq data from seven sequencing platforms is introduced (S1 Table). Specifically, we design cross-validation experiments in which ScanNet is trained on one platform and tested on the rest six platforms, finally getting seven groups of experimental results. In addition, the batch-correction capability of ScanNet is evaluated by comparison with six baseline methods that support batch correction.

In a real-world scenario, the most models would assign the cells from previously unknown cell types to known types wrongly [38,39]. To evaluate ScanNet's annotation capability in open-world scenario, we apply the pre-trained ScanNet model (trained on known cell types only) to new test datasets containing unknown cell types, also referred as novel types. Specifically, rare cell types with fewer than 10 cells, which are removed from the BaronMouse and AMB datasets during the initial training phase (Table 1), are reintroduced into the test datasets as new test datasets to simulate scenarios encountering novel types.

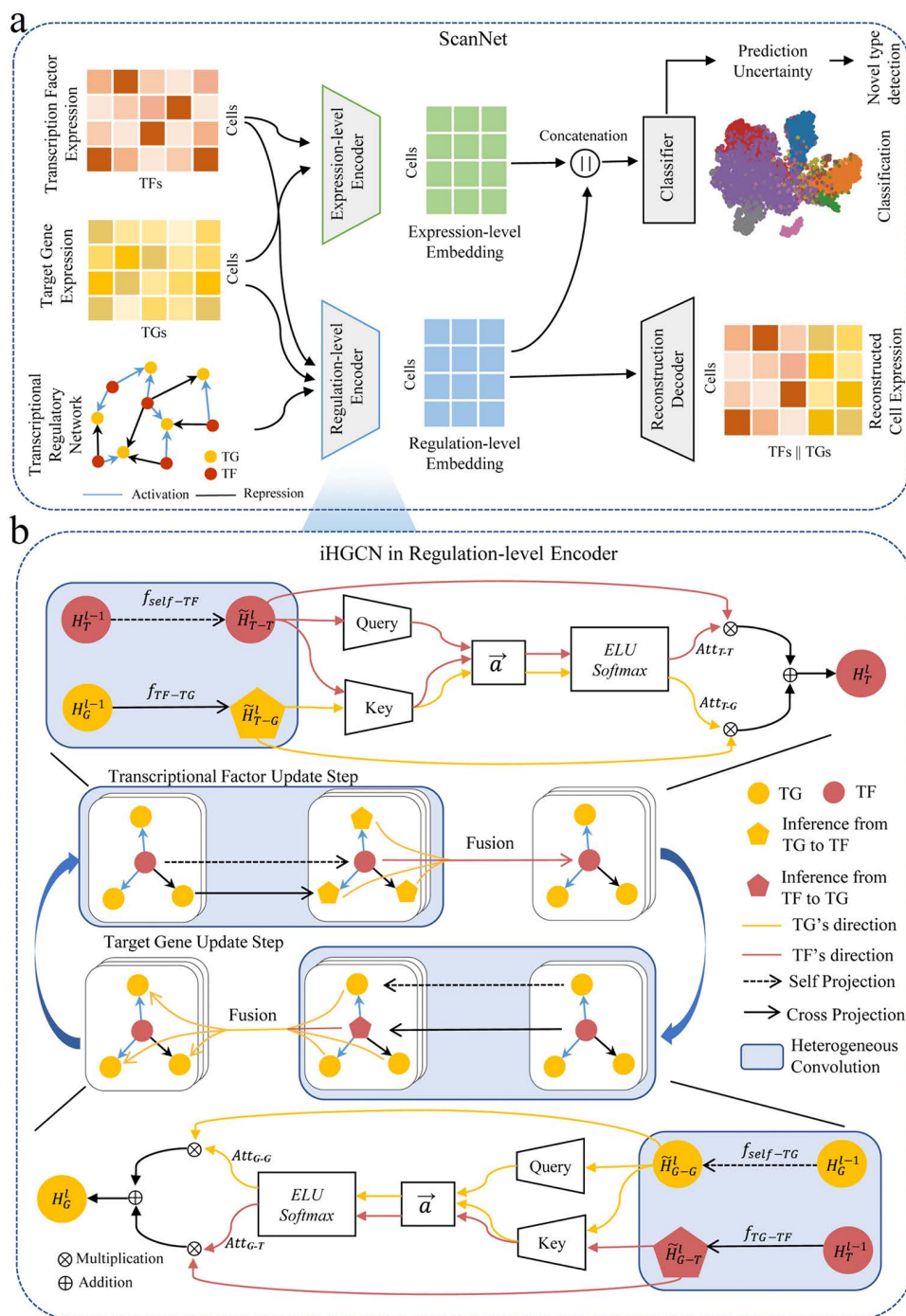


Fig 1. Framework of ScanNet. (a) Overall architecture of ScanNet. ScanNet takes the expressions of TF and TG, and a reference TRN as input. ScanNet consists of a dual-channel encoder, a Regulation-level Encoder and an Expression-level Encoder, a Reconstruction Decoder and a Classifier. (b) Details of the iHGCN in the Regulation-level Encoder. The iHGCN captures regulatory dependencies by alternately updating TF and TG representations through heterogeneous message passing on the TRN, involving TF update and TG update steps.

<https://doi.org/10.1371/journal.pcbi.1014602.g001>

Table 1. Description of the eight scRNA-seq benchmark datasets.

Dataset	Sequencing Platform	Specie	Tissue	TF	TG	Cell	Cell Type
PBMC	10X Chromium v3	Human	Peripheral blood mononuclear cell	266	684	9,631	19(19)
Muraro	CEL-seq2	Human	Pancreas	106	798	2,122	9(8)
BMMC	10X Chromium	Human	Bone marrow mononuclear cell	213	673	69,249	22(22)
Immune	10X-based integration	Human	Multi-tissue immune system	460	917	329,762	35(35)
AMB	SMART-Seq v4	Mouse	Primary mouse visual cortex	52	362	12,832	22(16)
TM	SMART-Seq2	Mouse	Whole mouse musculus	90	979	54,865	55(55)
BaronMouse	inDrop	Mouse	Pancreas	112	1,006	1,886	13(10)
ChenMouse	SNARE-seq	Mouse	Mouse cerebral cortex	245	657	9,190	22(22)

Note: the numbers outside the brackets indicate the total number of cell types; the numbers in brackets denote those with more than 10 cells.

<https://doi.org/10.1371/journal.pcbi.1014602.t001>

scATAC-seq provides insights into cellular heterogeneity by measuring chromatin accessibility, where the opening and closing of chromatin regions determine regulatory protein access and thereby influence gene expression initiation [40]. Given that scATAC-seq offers insights into deeper regulatory mechanisms through chromatin state profiling, we attempt to transfer the ScanNet framework to cell annotation on scATAC-seq data. Existing scATAC-seq classification tools [41] typically convert peak matrixes into gene score matrixes for downstream analysis, therefore we employ Signac [42] to transform chromatin accessibility signals into gene score matrixes that substitute for gene expression matrices in the ScanNet framework (S2 Text and S1 Fig). To evaluate the transferability of the ScanNet framework to scATAC-seq data, we use the scATAC-seq data from PBMC dataset [28]. Since this scATAC-seq data is paired with scRNA-seq data from the same PBMC dataset, we can compare the annotation performance of ScanNet across different modalities.

Given that benchmark datasets exhibit substantial class imbalance, we employ stratified sampling [43] to split each cell type into training and test sets at a 7:3 ratio, in which samples are drawn proportionally from each types to ensure that the overall sample accurately reflects the composition of the population. In addition, four evaluation metrics including Area Under the Precision-Recall Curve (AUPRC), F1, Precision and Recall, are employed to benchmark performance (S3 Text).

To extensively evaluate the performance of ScanNet in cell annotation, we benchmark it against ten baseline methods. These methods can be categorized into three groups: (i) reference mapping method (Seurat), (ii) deep learning-based methods (ACTINN [15], scAGN [44], scGCN [45], scGraphformer [17], scANVI [46]), and (iii) biological knowledge-informed methods (e.g., sigGCN [18], GEDFN [47], scGraph [20], and scRGCL [22]). Deep learning models are solely based on transcriptional expression profiles, while biological knowledge informed methods incorporate additional gene interaction networks as prior knowledge, ensuring that biological knowledge is explicitly embedded into cellular features. The protein-protein interaction (PPI) networks used by sigGCN, scGraph, and scRGCL are derived from the STRING database [19], whereas GEDFN uses a network obtained from the HINT database [48]. In contrast, ScanNet leverages transcriptional regulatory relationships between TFs and TGs encoded in a reference TRN as prior structural information. Among all compared methods, only scGCN, scAGN, Seurat, and scANVI adopt a semi-supervised learning paradigm, in which cell-type labels are transferred from a reference dataset to a query dataset for cell annotation. The remaining methods are fully supervised. The other benchmark methods are supervised learning models. All baseline models are implemented with their default hyperparameters.

ScanNet outperforms baselines in cell type annotation

We benchmark ScanNet against ten baseline methods on the eight benchmark datasets spanning different species, tissues and sizes. ScanNet performs better than all baselines in terms of F1 score and Precision (Fig 2a and S2 Table).

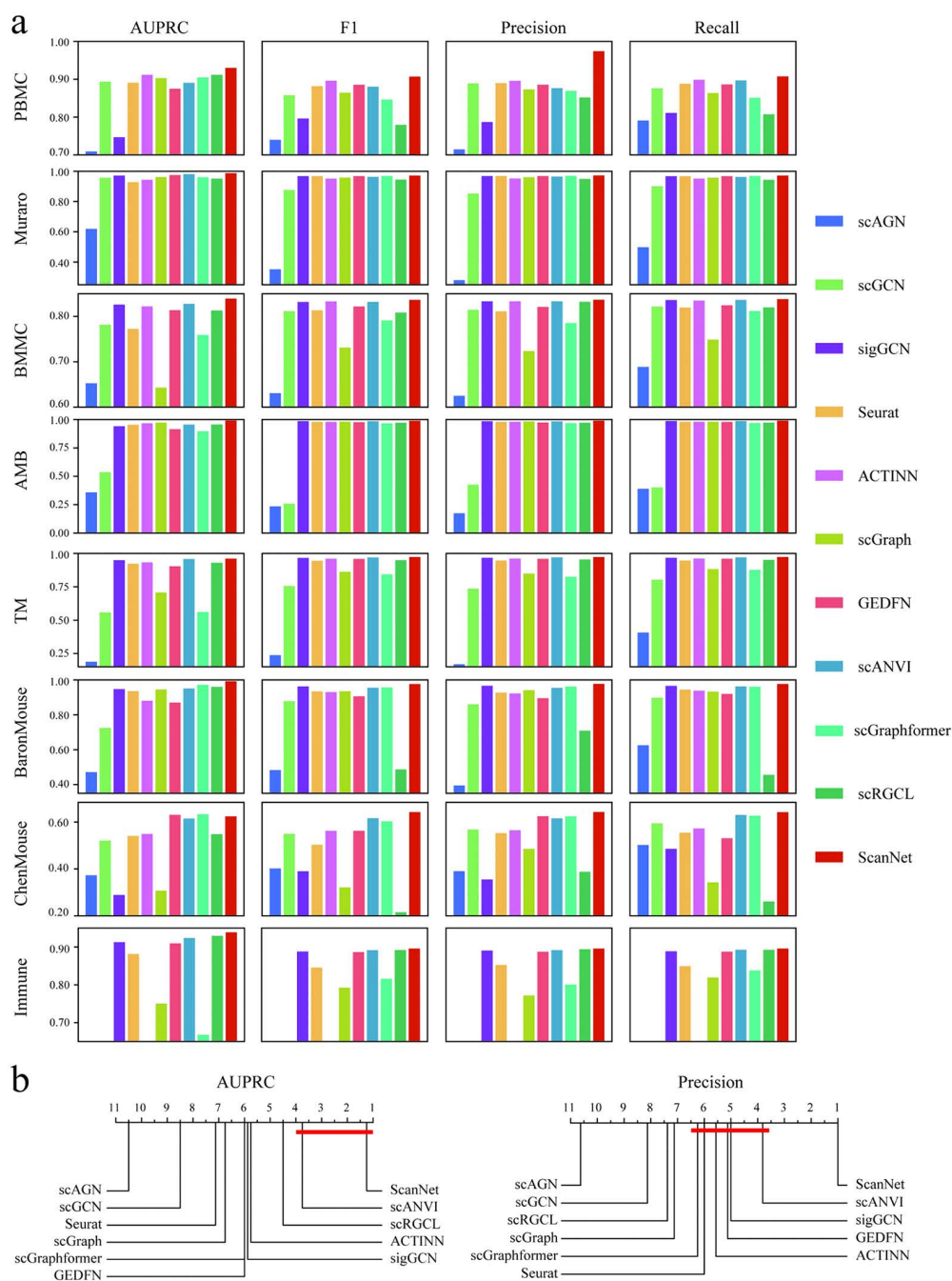


Fig 2. Performance evaluation of ScanNet on cell type annotation. (a) Comparison with ten methods across eight benchmark datasets on four evaluation metrics: AUPRC, F1 score, Precision, and Recall. **(b)** Critical difference diagrams for the two metrics: AUPRC and Precision. Methods connected by red are not significantly different in performance according to the Nemenyi test.

<https://doi.org/10.1371/journal.pcbi.1014602.g002>

ScanNet achieves an F1 score of 0.9762 and Precision of 0.9776 in BaronMouse dataset, outperforming the second-best baseline by 1.40% and 2.06% respectively. In PBMC dataset, the precision of ScanNet is 0.9742, while those for the best three baselines (ACTINN, Seurat and scGCN) are 0.8959, 0.8899 and 0.8891. For AUPRC and Recall, ScanNet

outperforms all baselines on all datasets. In AMB dataset, ScanNet is superior to the top two baseline methods, scGraph and ACTINN, with 1.83% and 2.41% improvement in terms of AUPRC. Both the ChenMouse and BMMC datasets are considered challenging due to the high similarity among cell types. Although ScanNet achieves a slightly lower AUPRC than scGraphformer in the ChenMouse dataset, it outperforms scGraphformer in terms of F1 score and Precision. Moreover, ScanNet achieves the best performance across all evaluation metrics on the BMMC dataset. These results indicate that ScanNet is also well suited for complex datasets.

In order to quantify the comparison with statistical rigor, we employed the Nemenyi test [49] and visualized the results using Critical Difference (CD) diagrams [50] at confidence level=0.95, following the procedure established by Zhang et al. [40]. The CD diagrams are designed to determine whether one algorithm achieves significantly better performance than others through statistical testing. In the CD diagram, the horizontal axis denotes the mean rank of each method across all datasets, and red lines connect methods whose differences are not statistically significant. ScanNet ranks first across all classification metrics (Figs 2b and S2). Although its AUPRC performance is comparable to that of scANVI, ScanNet achieves a markedly better average rank (1.28 vs. 3.85). Moreover, ScanNet significantly outperforms all competing methods, including scANVI, in term of Precision. Finally, although ScanNet does not achieve the highest AUPRC in the ChenMouse dataset, it exhibits statistically significant superiority over scGraphformer in overall AUPRC across datasets.

scRNA-seq datasets usually exhibit obvious imbalance in types (S3 Fig). The cell numbers in different cell types are illustrated by the bar heights spanning from 894 beta cells to 10 B cells in BaronMouse dataset (Fig 3). This uneven distribution

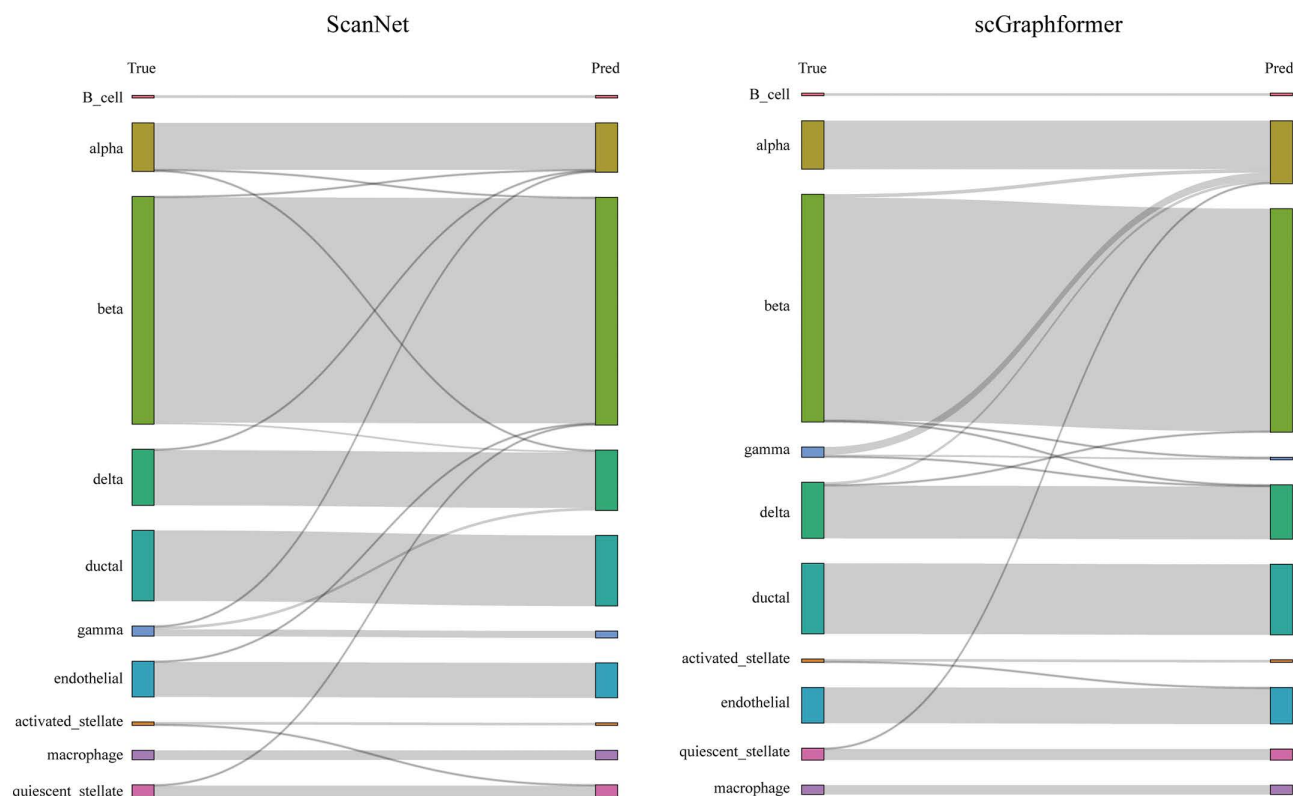


Fig 3. Comparison of cell type annotation performance between ScanNet and scGraphformer on the BaronMouse dataset. Sankey diagrams illustrate the correspondence between ground-truth (left) and predicted (right) cell type annotations on the test set. Each node represents a cell type, and the width of each connecting link reflects the number of cells assigned to the corresponding true-predicted pair. ScanNet demonstrates higher annotation accuracy with fewer cross-type misclassifications compared to scGraphformer.

<https://doi.org/10.1371/journal.pcbi.1014602.g003>

of sample numbers across cell types creates a typical scenario where model tends to learn the feature distribution of types with large sample size (e.g., beta cells) while neglecting the learning of minority class types (e.g., gamma cells). The Sankey diagrams (Fig 3) visualizes the correspondence between predicted and true cell types. We illustrate the results on BaronMouse dataset, which has the most pronounced cell type imbalance, and compare ScanNet with scGraphformer, the strongest baseline method. Nodes on the left represent ground-truth annotations, while nodes on the right denote model predictions. The width of each connecting link is proportional to the number of cells in each true-predicted type pair, which highlights that, for the minority gamma cell population, ScanNet demonstrates more direct correspondences with thicker Sankey flows between true and predicted labels, yielding a recall of 66.67%, whereas scGraphformer attains only 16.67%. Moreover, ScanNet attains a precision of 94.38% on alpha cells, surpassing scGraphformer's 77.03%. This result reveals that ScanNet achieves superior accuracy on minority cell types, demonstrating its enhanced ability to mitigate recognition bias toward majority classes while maintaining robust overall predictive performance.

In addition to benchmark datasets of varying complexity and class imbalance, we further evaluate ScanNet on a large-scale human immune atlas dataset containing 329,762 cells to assess its performance on large-scale atlases. Because large-scale single-cell datasets introduce significant computational challenges, particularly for methods that rely on constructing global cell-cell graphs, as the number of cells increases, the cost of nearest-neighbor search, graph construction, and message passing grows rapidly in both time and memory, making some methods difficult to execute reliably under our hardware constraints. Therefore, for the large-scale dataset, we report the results of representative baseline methods that can be successfully executed under our computational environment, including Seurat, scGraph, sigGCN, scANVI, scGraphformer, GEDFN, and scRGCL. ScanNet maintains strong annotation performance on this large-scale dataset, achieving competitive or superior results compared with representative baseline methods across all evaluation metrics (Fig 2), indicating that ScanNet not only achieves strong performance across datasets of varying biological complexity, but also scales effectively to large single-cell atlases while maintaining robust annotation accuracy. Importantly, the performance advantage of ScanNet remains consistent across scales, suggesting that the incorporation of transcriptional regulatory priors provides stable benefits that are not diminished by increasing data size.

All these results demonstrate ScanNet achieves consistent classification advantages over tissues, species, and sample distribution. By systematically integrating TRN with gene expression data, ScanNet captures both regulatory mechanisms and transcriptional outcomes, providing mechanistic insights into cellular heterogeneity.

ScanNet exhibits robust performance across different sequencing platforms

To evaluate ScanNet's robustness to technical noise led by different sequencing platforms, we conduct experiments on PBMCbench dataset [37], which contains scRNA-seq data from the same donor's peripheral blood mononuclear cells sequenced with seven different platforms.

ScanNet is evaluated by leave-six-out validation, that is trained on data from one sequencing platform and tested on the remaining six, yielding a total of seven results. Fig 4a presents AUPRC pairwise comparisons between ScanNet and six baseline models. Most scatter points lie above the diagonal, demonstrating that ScanNet's AUPRC outperforms that of most methods. In CD diagrams, ScanNet achieves the top overall ranking across all classification metrics, and exhibiting batch robustness comparable to that of the state-of-the-art method scANVI (Fig 4b).

Additionally, t-SNE visualizations compare cellular representations derived from the PCA-based embeddings of normalized expressions (Fig 4c, left panel) and ScanNet-generated embeddings (Fig 4c, right panel), with upper plots colored by cell type and lower plots by batch. In the expression-based visualization (left panel), different cell types show considerable overlap and mixing, making it difficult to distinguish clear cell-type boundaries. Moreover, when colored by sequencing platform (lower left), cells from the same platform tend to cluster together, indicating strong platform-driven batch effects. In contrast, the ScanNet embedding visualization (right panel) demonstrates markedly different patterns. When colored by cell type (upper right), ScanNet forms compact and well-separated clusters for each cell type, with clear boundaries between

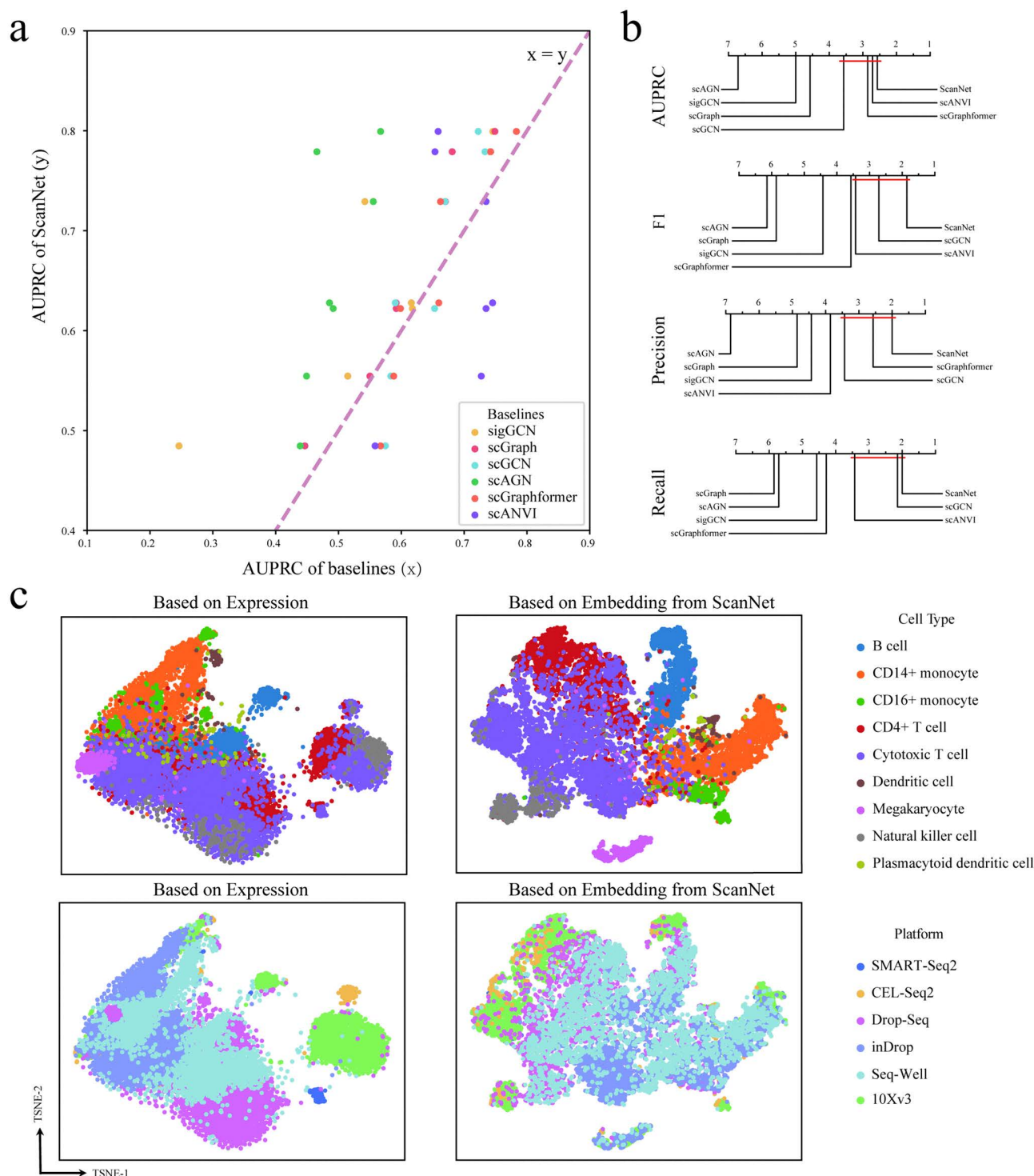


Fig 4. Evaluation of the generalization ability and robustness of ScanNet across different sequencing platforms. (a) Scatter plot comparing AUPRC scores of ScanNet with seven baselines across seven experimental settings. Each point corresponds to one comparison, with the x-axis representing baseline AUPRC and the y-axis representing ScanNet AUPRC. The dashed diagonal line ($x = y$) denotes performance parity. (b) Critical difference diagrams for four evaluation metrics (AUPRC, F1 score, Precision, and Recall). (c) t-SNE visualizations on the test set, colored by cell type

(upper panels) and sequencing protocol (lower panels). The left panels show PCA-based t-SNE embeddings derived from the normalized and scaled expression profiles of selected TFs and TGs, while the right panels are based on embeddings learned by ScanNet, showing improved separation of cell types and reduced batch effects across protocols.

<https://doi.org/10.1371/journal.pcbi.1014602.g004>

different cellular populations. Importantly, when colored by sequencing platform (lower right), ScanNet embeddings show reduced platform-driven clustering compared with the expression space, although some residual platform-associated structure remains in specific cell populations, such as cytotoxic T cells. The residual platform-specific structure is present to varying extents across methods, and that no method completely removes platform effects while simultaneously preserving clear cell-type separation (S4 Fig). Compared with other methods, ScanNet provides a more favorable balance between cell-type discrimination and cross-platform mixing. S3 Table summarizes the cross-platform classification metrics across seven experimental combinations and S4 Table records batch correction metrics of ScanNet and baselines.

By incorporating the underlying gene regulatory mechanisms and directed relationship from TFs to TGs, ScanNet captures the intrinsic regulatory processes that drive cellular identity, thereby mitigating platform-specific biases inherent to divergent sequencing technologies and exhibiting robust classification performance.

ScanNet can identify novel cell types

ScanNet leverages prediction uncertainty quantified by entropy to distinguish novel cell types adapting to real-world scenarios. ScanNet exhibits higher uncertainty (higher entropy) when encountering previously unseen cell types compared to the confident predictions (lower entropy) for known cell types.

In the experiment, cell types with fewer than 10 cells are removed from the training set and applied as entirely unknown novel cell types during testing. Unknown cell types (first row in each matrix) are predominantly classified as “Novel” rather than being assigned to known cell types (Fig 5a), indicating ScanNet demonstrates robust performance on both AMB and BaronMouse datasets, accurately predicting known cell types while effectively identifying novel cell types not encountered during training. Quantitatively, for the AMB dataset, ScanNet achieves 71.43% recall and 71.43% precision for novel cell detection. For the BaronMouse dataset, while precision is 66.67%, recall reaches an impressive 95.24%, highlighting ScanNet’s high sensitivity in identifying novel cells. In addition, we evaluate entropy-based novel cell-type identification for more cells, further demonstrating ScanNet can identify novel types with large number of cells (S4 Text and S5 Fig).

The entropy-based uncertainty quantification is effective and shows distinct distributions of entropy values in known and novel cell types (Fig 5b). In the AMB dataset, the mean and variance of known type distribution are 0.0168 and 0.0069, while these of novel type are 0.911 and 0.2155, implying that entropy of known types exhibit high prediction confidence. In contrast, novel cell types display broader distributions with higher entropy values, reflecting greater prediction uncertainty. This clear discrepancy validates entropy as an effective metric for distinguishing novel from known cell types based on model confidence.

These results demonstrate that ScanNet successfully identifies novel cell types through entropy-based uncertainty quantification, addressing a critical limitation of existing supervised models that misclassify unknown types into predefined categories. With high detection sensitivity, ScanNet provides practical functionality for real application scenarios where new cell populations continuously emerge, making it valuable for both exploratory research and clinical applications.

While ScanNet can identify cells that deviate from the reference distribution via entropy as “Novel”, such novel cells require further biological interpretation in practice. In real single-cell analysis pipelines, high-resolution annotation is typically achieved by combining automated predictions with marker-based validation. We therefore evaluate whether ScanNet can be integrated with marker-based tools to form a hybrid pipeline for improving the annotation of novel cells. Using scCATCH [51], a widely used marker-based tool, we adopt a stratified strategy in which ScanNet first partitions cells according to prediction confidence. High-confidence cells retain the labels predicted by ScanNet, whereas low-confidence

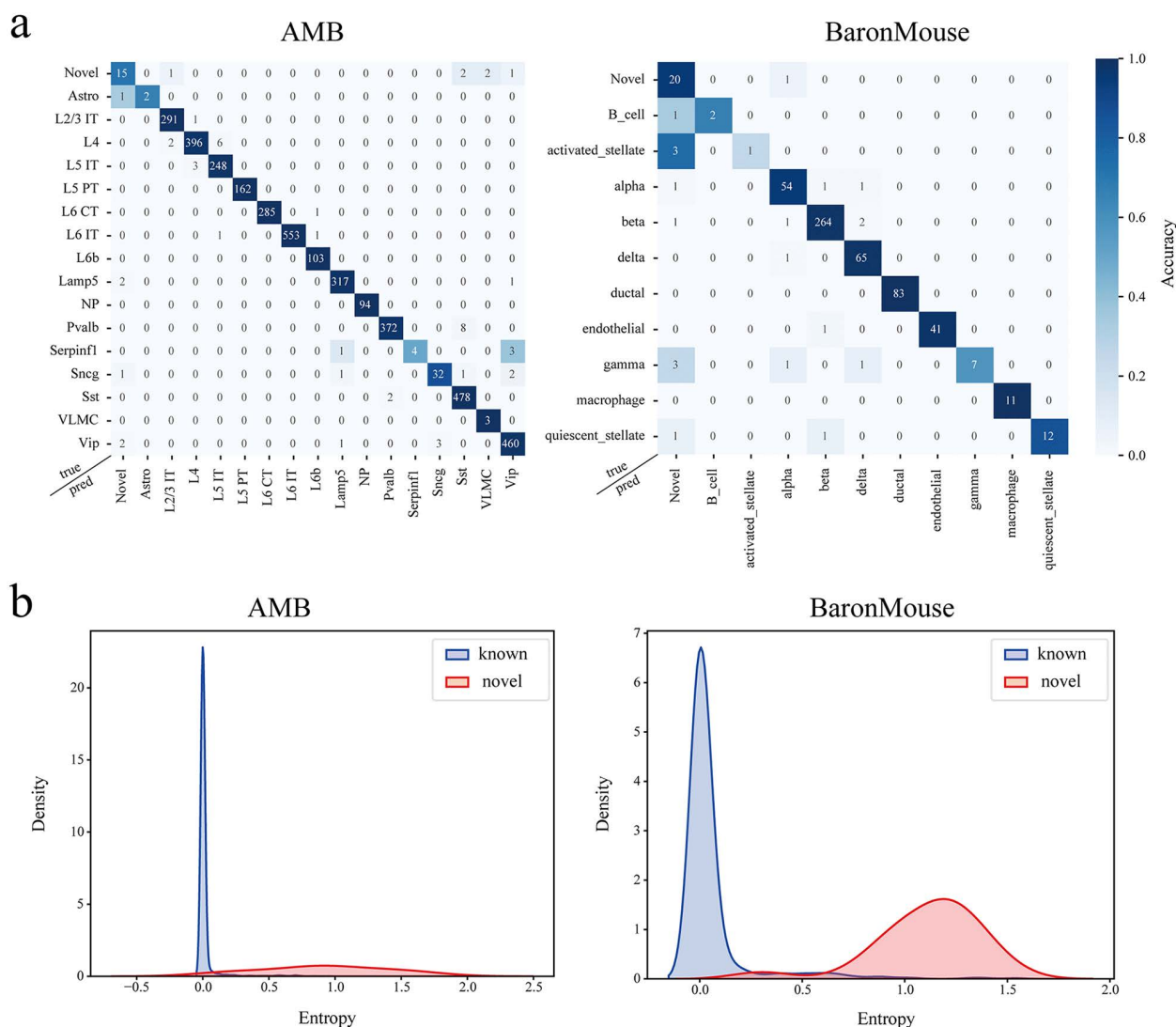


Fig 5. Identification of novel cell types by ScanNet. (a) Confusion matrices of ScanNet on the test set of AMB and BaronMouse datasets, showing that ScanNet can distinguish novel types from known cell types. (b) Entropy-based density distributions. The higher entropy values associated with novel cells demonstrate that prediction uncertainty can effectively separate unknown from known cell types.

<https://doi.org/10.1371/journal.pcbi.1014602.g005>

cells are further annotated with scCATCH (S6 Fig). Importantly, for these low-confidence cells, scCATCH is applied to clusters derived from ScanNet embeddings rather than to clusters derived from the original expression profiles, corresponding results are provided in S7 Fig. This hybrid pipeline enhances the resolution of novel cell types. While ScanNet identifies such cells as “Novel”, scCATCH refines them into biologically meaningful candidate lineages, with reduced lineage confusion compared with standalone scCATCH based on the original expression profiles (S7 Fig). These results demonstrate that ScanNet can serve as an effective front-end module in practical annotation pipelines.

ScanNet performs well with transferred scATAC-seq data

The framework of ScanNet can be transferred to other omics data, scATAC-seq data. To evaluate annotation performance on scATAC-seq data, annotation performance on scATAC-seq data is compared with Cellcano [41], a well-known cell classification

tool specifically designed for scATAC-seq, implemented here with default hyperparameters. ScanNet is applied to paired scRNA-seq data and scATAC-seq data separately. ScanNet performs superior on scATAC-seq data than Cellcano across all evaluation metrics demonstrating the flexible extension ability of ScanNet from scRNA-Seq to scATAC-Seq data (Fig 6a).

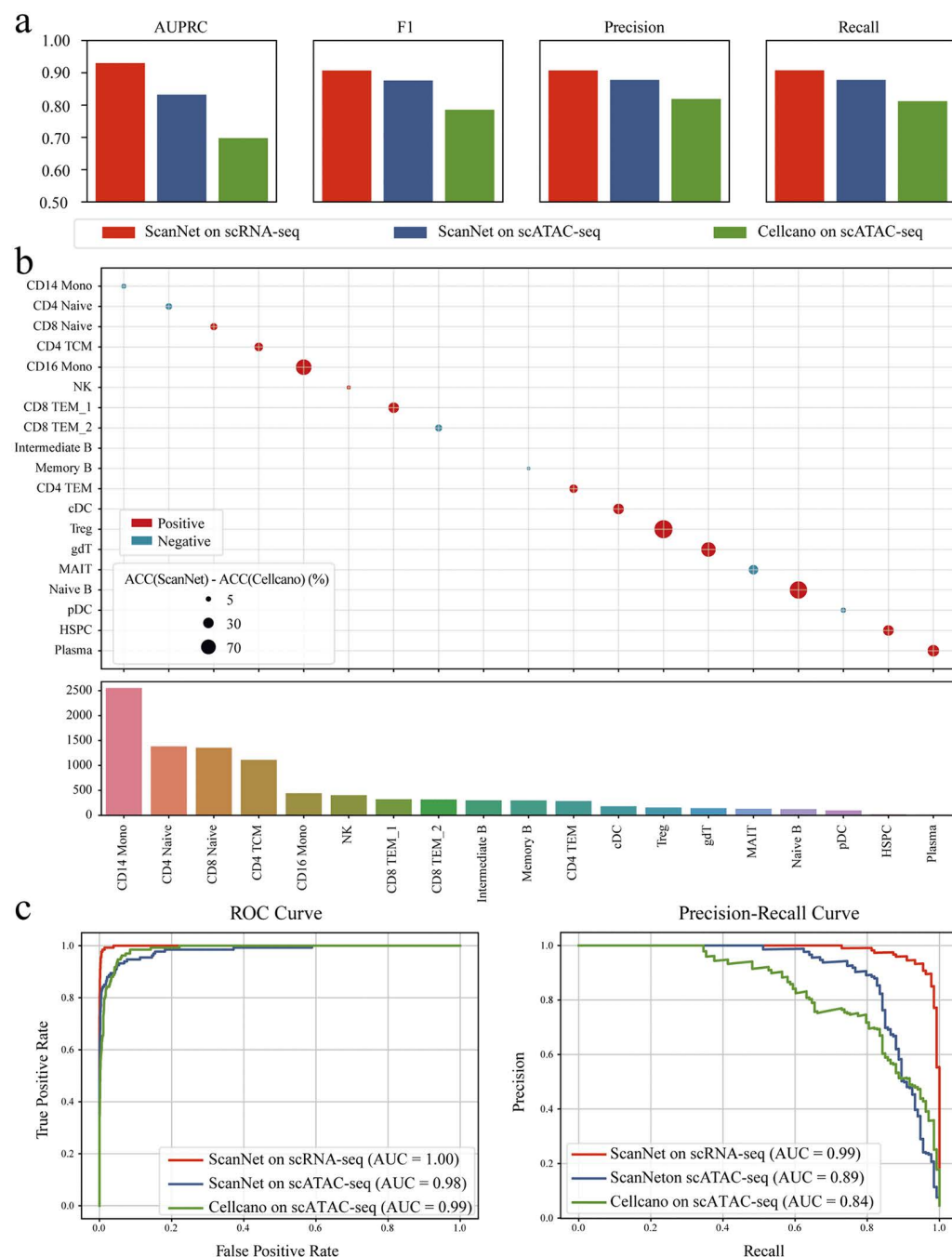


Fig 6. Annotation performance of ScanNet on scATAC-seq data. (a) Bar chart comparing annotation performance on the PBMC dataset. (b) Accuracy comparison between ScanNet and Cellcano across all cell types. The upper scatter plot depicts accuracy differences by cell type, where point size indicates the magnitude of the difference and color represents the performance direction (red: ScanNet superior; blue: Cellcano superior). The lower bar chart summarizes the number of cell types in the PBMC dataset. (c) The ROC and PRC of two models on the CD16 Mono type.

<https://doi.org/10.1371/journal.pcbi.1014602.g006>

We also evaluate ScanNet's performance on sample size imbalance across types for scATAC-Seq data. [Fig 6b](#) presents a detailed accuracy comparison between ScanNet and Cellcano across all cell types in scATAC-seq data. The upper scatter plot visualizes accuracy differences by cell type where point size reflects the magnitude of differences and color denotes whose performance is better. Specifically, red represents ScanNet superior, blue stands Cellcano superior. The lower bar chart displays the sample size (number of cells) for each cell type shown in the scatter plot above, with both subplots sharing a common horizontal axis for cell-type alignment. ScanNet outperforms Cellcano in 12 out of 19 cell types, with particularly substantial advantages in minority cell types such as Naive B, gdT, and Treg cells, where ScanNet's performance exceeds Cellcano by over 40–70%. In contrast, Cellcano shows marginal advantages only in 6 cell types, where the performance differences are relatively small (typically < 15%). However, Cellcano faces challenges when classifying cell types with limited sample sizes, often misclassifying them, as seen with CD16 Mono and Treg cells. In contrast, ScanNet demonstrates exceptional robustness to severe class imbalance, maintaining high performance across both abundant cell types and rare cell types (e.g., HSPC, Plasma with < 20 cells). ScanNet is also evaluated on scATAC-seq without paired scRNA-seq data and compared with Cellcano ([S5 Table](#)). ScanNet consistently outperforms Cellcano, indicating that ScanNet can generalize well to scATAC-seq datasets even when no matched scRNA-seq data are available, demonstrating its strong transferability and generalization across modalities.

To further exhibits the efficacy of ScanNet on minority cell types, we plot the receiver operating characteristic (ROC) and precision-recall curves (PRC) for the CD16 Mono population ([Fig 6c](#)). While ScanNet and Cellcano exhibit comparable AUROC performance (0.98 vs 0.99), ScanNet achieves a notable higher AUPRC (0.89 vs 0.84), highlighting its superior precision-recall trade-off in minority class annotation.

Although ScanNet's performance on scATAC-seq data is slightly lower than its scRNA-seq performance, it still substantially outperforms the scATAC-seq-specific method Cellcano, demonstrating ScanNet's effective cross-modal transferability and adaptability to diverse single-cell sequencing technologies.

Discussions

Cell type annotation is a critical and fundamental task in single-cell omics research for serving as a critical bridge connecting fundamental mechanism exploration with clinical translation. However, the inherent feature of scRNA-seq data inevitably posed significant challenges to this task. To address these challenges, we propose ScanNet, an innovative model engineered to integrate reference TRN and scRNA-seq data via iterative heterogeneous graph convolution framework for cell annotation.

Benchmark experiment on eight benchmark datasets demonstrates ScanNet's effectiveness in cell annotation tasks, where it outperforms existing methods across four metrics. Moreover, ScanNet exhibits strong recognition capabilities for minority cell types in imbalanced datasets. In batch generalization experiments, ScanNet effectively mitigates noise introduced by diverse sequencing platforms, as evidenced by t-SNE analysis showing that cell clustering is primarily driven by cell types rather than batch. To address the real application scenarios where unannotated datasets often contain previously unknown cell types, ScanNet accurately identifies novel classes in the simulated scenarios, demonstrating remarkable flexibility and generalization capabilities. Despite being originally designed for scRNA-seq data, ScanNet outperforms existing benchmark methods designed for scATAC-seq data, highlighting its versatility and robustness across diverse single-cell omics data. All these results indicate that ScanNet effectively integrates the causal relationships between transcriptional regulation and transcriptional expression, capturing cell heterogeneity through upstream regulatory mechanisms and achieving superior annotation performance.

In addition, we conduct ablation experiments of two embeddings and regulatory graph structure. The results confirm that ScanNet benefits from the complementary contributions of Expression-level and Regulation-level embeddings, and that the reference TRN contributes to classification accuracy ([S8 Fig](#) and [S6 Table](#)). We also further add a loss balancing and hyperparameter sensitivity analysis to examine the optimization dynamics of the MSE reconstruction loss and CE

classification loss ([S9 Fig](#)), as well as the effects of different loss-weight settings ([S7 Table](#)) and L2 regularization coefficients ([S8 Table](#)). The results show that the two loss terms contribute complementarily during training and that ScanNet maintains stable performance across different weight ratios and regularization strengths. Besides, to investigate whether the learned attentions in ScanNet capture cell-type-specific regulatory differences, we reconstruct TF-TG interactions at the single-cell level ([S5 Text](#)). The reconstructed regulatory representations exhibit clear cell-type-specific organization ([S10 Fig](#)), which suggests that ScanNet has the potential to uncover cell-type-specific regulatory dependencies and provides a promising direction for future interpretability studies. We also compare ScanNet with ACTINN trained on all genes and find that ScanNet achieves better overall performance ([S9 Table](#)). Using more TFs and TGs yields no significant classification improvement ([S10 Table](#)). These results support using only TF and TG features in ScanNet.

ScanNet introduces the reference TRN network as priori biological knowledge, incorporating the regulatory logic and molecular mechanisms governing the gene expression, thereby enabling accurate cell annotation. ScanNet deploys a dual-channel encoder to dissect cellular heterogeneity by concurrently capturing transcriptional regulatory states and resultant gene expression. Specially, Regulation-level Encoder employs iterative heterogeneous node encoding optimization to capture the causal relationship between regulatory state information and expression outcomes by heterogeneous graph convolution, while the Expression-level Encoder extracts regulatory outcome features from scRNA-seq data within cells. This design allows for the extraction of fine-grained regulatory information at the gene level while simultaneously integrating broader contextual information at the cell level. Through dual-channel encoder, ScanNet constructs a unified and biologically informed cell representation that captures not only the overall transcriptional state but also the underlying regulatory mechanisms. The ablation results provide empirical evidence for the effectiveness of ScanNet's dual-channel design and indicate that incorporating the TRN as prior information leads to improved performance.

Certainly, while ScanNet exhibits excellent performance in four annotation tasks, promising avenues for enhancement emerge from its design and experimental insights. First, a uniform TRN structure is applied to diverse cell types in ScanNet, without deeper exploration of cell-type-specific regulatory interactions that is critical for unraveling transcriptional divergence across biological contexts. Second, its capacity to resolve regulatory relationships and cell-type specificity could be strengthened by integrating multi-omics data, which would refine its ability to capture spatial and epigenetic determinants of cell identity. In future work, we aim to integrate finer-grained regulatory information, cell-type-specific regulatory relationship, with scRNA-seq and scATAC-seq data, thereby providing a more efficient, robust, and scalable solution for single-cell data analysis. This enhanced framework will not only dissect the causal regulatory circuits underlying cell identity specification but also empower researchers to collaboratively generate more valuable insights.

Materials and method

The dual-channel encoder

Transcriptional regulation orchestrates gene expression [26], establishing a causal link from TF to TG. With this prior knowledge, ScanNet deploys a dual-channel encoder, consists of a Regulation-level Encoder and an Expression-level Encoder, to dissect cellular heterogeneity from regulatory perspective and expression perspective respectively. The Regulation-level Encoder employs iHGCN to alternately update TF and TG representations through heterogeneous message passing on the TRN, combining self-projection, cross-projection, and attention-based fusion and learns cellular regulatory states, encoded into Regulation-level Embeddings. The Expression-level Encoder utilizes fully connected neural network to extract global cellular transcriptional states and regulatory outcome features inside cells from scRNA-seq data into Expression-level Embeddings, while neglecting inner regulation interactions within cells. The Regulation-level Encoder solely capture local features, due to learning filters' localization property of GCN. Therefore, through the dual-channel encoder, ScanNet constructs a unified and biologically informed cell representation that captures not only the overall transcriptional state but also the underlying regulatory mechanisms. Comprehensive definitions of all variables, matrices, and operators used from data preprocessing through model training are provided in [S11 Table](#).

Iterative heterogeneous fusing encoding regulatory process in Regulation-level Encoder

iHGCN-based Regulation-level Encoder capture the regulatory dependencies by iteratively learning embeddings of TFs and TGs, which aggregates information from TGs to update TF representations, and vice versa. This bidirectional message passing enables Regulation-level Encoder to facilitate precise localization of cell states. The Regulation-level Encoder employs two iHGCN layers to integrate TF-TG regulatory relationships encoded in the TRN into the embeddings of TF and TG nodes, followed by a pooling layer to derive the Regulation-level Embedding for each cell. Taking the “Transcriptional Factor Update Step” in Fig 1b as an example, a detailed explanation of the iHGCN layer in Regulation-level Encoder is provided as follows. The iHGCN layer consists of two steps: heterogeneous feature extraction from TF-TG regulatory relationships, and attention-driven TF-TG representation fusion for cellular embedding generation.

Heterogeneous feature extraction from TF-TG regulatory relationships

The iHGCN applies heterogeneous convolution to learn TF-TG regulatory associations from the TRN. In order to eliminate the influence of both TF and TG node degree on information transmission, row normalization is employed to the adjacency matrix, ensuring that the information from each node is appropriately represented during the aggregation.

$$\mathbf{D}^T = \text{diag} \left(\sum_{j=1}^G |\tilde{\mathbf{A}}_{ij}^{T-G}| \right) \in R^{T \times T}, \quad (1)$$

$$\mathbf{D}^G = \text{diag} \left(\sum_{i=1}^T |\tilde{\mathbf{A}}_{ji}^{G-T}| \right) \in R^{G \times G}, \quad (2)$$

$$\tilde{\mathbf{A}}_n^{T-G} = (\mathbf{D}^T)^{-1/2} \tilde{\mathbf{A}}^{T-G} (\mathbf{D}^G)^{-1/2} \in R^{T \times G}, \quad (3)$$

$$\tilde{\mathbf{A}}_n^{G-T} = (\mathbf{D}^G)^{-1/2} \tilde{\mathbf{A}}^{G-T} (\mathbf{D}^T)^{-1/2} \in R^{G \times T}, \quad (4)$$

where $\tilde{\mathbf{A}}^{G-T}$ is the transpose of $\tilde{\mathbf{A}}^{T-G}$, $|\cdot|$ denotes absolute values, and $\sum$ represents the row-wise sum. $\mathbf{D}^T$ is the diagonal degree matrix of $\tilde{\mathbf{A}}^{T-G}$, where each diagonal entry records the number of TGs regulated by the corresponding TF. Likewise, $\mathbf{D}^G$ quantifies the number of TFs that regulate each TG.

The functions $f_{\text{self-TF}}$ and $f_{\text{TF-TG}}$ are designed to extract intrinsic features from TF nodes and regulatory information from TG nodes, respectively:

$$\begin{cases} \tilde{\mathbf{H}}_{T-T}^l = f_{\text{self-TF}}(\mathbf{H}_T^{l-1}) = \mathbf{H}_T^{l-1} \mathbf{W}_{\text{self-TF}}^l \\ \tilde{\mathbf{H}}_{T-G}^l = f_{\text{TF-TG}}(\mathbf{H}_G^{l-1}) = \tilde{\mathbf{A}}_n^{T-G} \mathbf{H}_G^{l-1} \mathbf{W}_{T-G}^l \end{cases}, \quad (5)$$

where $f_{\text{self-TF}}$ focuses on capturing the inherent characteristics of TFs, while $f_{\text{TF-TG}}$ uses graph convolution to integrate regulatory interaction information from neighboring TG nodes to infer TFs' potential features. Specifically, $\mathbf{H}_T^{l-1} \in R^{T \times d_{l-1}}$ and $\mathbf{H}_G^{l-1} \in R^{G \times d_{l-1}}$ are the input feature of TF and TG of the l^{th} layer (the output of $(l-1)^{\text{th}}$ layer), $\mathbf{W}_{T-G}^l \in R^{d_{l-1} \times d_l}$ is the weight matrix of aggregating information from TG, $\mathbf{W}_{\text{self-TF}}^l \in R^{d_{l-1} \times d_l}$ is the projection matrix to project $\mathbf{H}_T^{l-1}$ into the same eigenspace as $\tilde{\mathbf{H}}_{T-G}^l \in R^{T \times d_l}$.

However, the embedding $\tilde{\mathbf{H}}_{T-G}^I$ based on TRN does not retain the feature information of the TF itself, potentially limiting the completeness of the feature representation. To overcome this limitation, iHGCN introduces an attention mechanism to integrate the embedding vector of TF itself ($\tilde{\mathbf{H}}_{T-T}^I$) and the inferred embedding vector of TF from TG ($\tilde{\mathbf{H}}_{T-G}^I$), thereby generating a more comprehensive and informative embedding for TF.

Attention-driven TF-TG representation fusion and cellular embedding generation: As depicted in Fig 1b, the attention mechanism maps a set of queries and a set of key-value pairs to an output. In practice, we pack together queries, keys and values into three matrixes $\mathbf{Q}$, $\mathbf{K}$ and $\mathbf{V}$ respectively. Then attention mechanism can be formulated as: $\text{softmax}(f_a(\mathbf{Q}, \mathbf{K})) \mathbf{V}$, where f_a is the attention function such as dot-product [52] or neural network [53]. Here, embedding vectors $\tilde{\mathbf{H}}_{T-G}^I$ and $\tilde{\mathbf{H}}_{T-T}^I$ are both as values matrix $\mathbf{V}$. We define a weight matrix $\mathbf{W}_k \in R^{d_l \times d_a}$ to map them into keys matrix $\mathbf{K}$, and define a weight matrix $\mathbf{W}_q \in R^{d_l \times d_a}$ to map $\tilde{\mathbf{H}}_{T-T}^I$ into the query matrix $\mathbf{Q}$, where d_a is the hidden layer dimension. Formally,

$$\mathbf{K}_{T-T} = \tilde{\mathbf{H}}_{T-T}^I \mathbf{W}_k, \quad (6)$$

$$\mathbf{K}_{T-G} = \tilde{\mathbf{H}}_{T-G}^I \mathbf{W}_k, \quad (7)$$

$$\mathbf{Q}_T = \tilde{\mathbf{H}}_{T-T}^I \mathbf{W}_q, \quad (8)$$

Note that mapping $\tilde{\mathbf{H}}_{T-T}^I$ as the query is the key to achieve personalized importance estimation for each TF node. The attention function is implemented as follows:

$$\mathbf{e}_{T-T} = \text{ELU}([\mathbf{K}_{T-T} \parallel \mathbf{Q}_T] \mathbf{w}_a^T), \quad (9)$$

$$\mathbf{e}_{T-G} = \text{ELU}([\mathbf{K}_{T-G} \parallel \mathbf{Q}_T] \mathbf{w}_a^T), \quad (10)$$

where $\parallel$ denotes the row-wise concatenation operation, $\mathbf{w}_a^T \in R^{2d_a \times 1}$ is the parameter vector and ELU is the activation function. The i^{th} element of $\mathbf{e}_{T-T}$ denotes the self-contribution weight of the i^{th} TF, whereas the i^{th} element of $\mathbf{e}_{T-G}$ represents the regulation aggregated importance vector of its neighbors (TGs) on the i^{th} TF. Then, the normalized attention coefficients are computed with the *softmax* function to update the representation of TF nodes:

$$[\mathbf{Att}_{T-T} \parallel \mathbf{Att}_{T-G}] = \text{softmax}([\mathbf{e}_{T-T} \parallel \mathbf{e}_{T-G}]), \quad (11)$$

$$\mathbf{H}_T^I = \mathbf{Att}_{T-T} \odot \tilde{\mathbf{H}}_{T-T}^I + \mathbf{Att}_{T-G} \odot \tilde{\mathbf{H}}_{T-G}^I, \quad (12)$$

where $\mathbf{H}_T^I$ is the i^{th} iHGCN-updated TF embeddings, $\mathbf{Att}_{T-T}$ represents the normalized self-information attention weight for TFs, $\mathbf{Att}_{T-G}$ represents the normalized weight of regulatory information from TGs, and $\odot$ is hadamard operator. The i^{th} iHGCN-updatpled TG embeddings $\mathbf{H}_G^I$ are obtained in an analogous manner.

The Regulation-level Encoder employs two iHGCN layers to learn molecular-level embeddings for TFs and TGs ($\mathbf{H}_T^2$, $\mathbf{H}_G^2$), and then aggregates them via a *Read_Out* function to produce cells' Regulation-level Embeddings:

$$\mathbf{H}^R = \text{Read_Out}([\mathbf{H}_T^2 \parallel \mathbf{H}_G^2]), \quad (13)$$

where *Read_Out* refers to the max-pooling layer, and $\mathbf{H}^R \in \mathbb{R}^{N \times d_o}$ represents the Regulation-level Embedding of cells, in which d_o is the output dimension of the Regulation-level Encoder.

Gene expression encoding in Expression-level Encoder

The Expression Encoder is designed to capture the holistic cellular state by modeling transcriptional profiles, thereby generating comprehensive global representations. This global perspective compensates for the limitation inherent to the Regulation Encoder, which focuses on gene-level regulatory relationship and consequently lacks a unified view of the entire cell system. Specifically, the Expression-level Encoder employs a simple MLP to capture the overall cellular state:

$$\mathbf{H}^E = \text{Relu}(\text{Relu}(\mathbf{X}\mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2), \quad (14)$$

where $\mathbf{X} = [\mathbf{X}_T \parallel \mathbf{X}_G]$ is the concatenated expression of TFs and TGs, representing gene expression within cells, $\mathbf{H}^E \in \mathbb{R}^{N \times d_o}$ is the Expression-level Embedding vector where d_o is the output dimension of Expression-level Encoder same as that of Regulation-level Encoder, and *Relu* is the activation function.

The decoder to reconstruct gene expression and the classifier to annotate cell type

The Reconstruction Decoder is designed to reconstruct downstream cell expressions from upstream Regulation-level Embeddings, ensuring that Regulation-level Embeddings accurately capture the regulatory state of cells. Thereby, the Reconstruction Decoder is designed to rebuild the expression of TFs and TGs from $\mathbf{H}^R$:

$$\hat{\mathbf{X}} = [\hat{\mathbf{X}}_T \parallel \hat{\mathbf{X}}_G] = \mathbf{H}^R \mathbf{W}_{re} + \mathbf{b}_{re}, \quad (15)$$

where $\mathbf{W}_{re} \in \mathbb{R}^{d_o \times (T+G)}$ is the weight matrix, and $\mathbf{b}_{re} \in \mathbb{R}^{(T+G)}$ is the bias vector. As a result, $\hat{\mathbf{X}} \in \mathbb{R}^{N \times (T+G)}$ has the same dimension as the concatenated input $\mathbf{X} = [\mathbf{X}_T \parallel \mathbf{X}_G]$, which $T + G$ is the number of all TF nodes and TG nodes in TRN.

It is worth noting that the reconstruction objective is applied only to the gene expression matrix, rather than the TRN. In our framework, the TRN is incorporated as a shared prior across all cells to guide information propagation between transcription factors and target genes. Since this prior is identical for all cells, introducing a reconstruction loss on the TRN would implicitly enforce the model to recover the same graph structure across samples. However, the primary objective of cell type annotation is to capture cell-specific transcriptional variations. Therefore, constraining the model to reconstruct a global and invariant network structure may reduce its ability to learn discriminative representations. For this reason, we restrict the reconstruction objective to the gene expression matrix, which directly reflects cell-level heterogeneity, while the TRN is used as structural prior knowledge rather than a reconstruction target.

Since Regulation-level Embeddings $\mathbf{H}^R$ and Expression-level Embeddings $\mathbf{H}^E$ characterize cell features from regulatory mechanisms and expression outcomes, the Classifier is trained on the concatenation of these two embeddings to achieve accurate cell annotation:

$$\mathbf{z} = [\mathbf{H}^R \parallel \mathbf{H}^E] \mathbf{W}_{cl} + \mathbf{b}_{cl}, \quad (16)$$

$$\hat{\mathbf{Y}} = \text{softmax}(\mathbf{z}), \quad (17)$$

where $\mathbf{W}_{cl} \in \mathbb{R}^{2d_o \times C}$ is the weight matrix in which C is the number of cell types, $\mathbf{b}_{cl} \in \mathbb{R}^C$ is the bias vector, and $\hat{\mathbf{Y}} \in \mathbb{R}^{N \times C}$ is the predicted probabilities matrix.

The optimization of loss function

The loss function comprises three key components: the reconstruction loss of expression data $\mathcal{L}_{MSE}$, the classification loss of cell typing $\mathcal{L}_{CE}$ and L2 regularization, as given in Eq (18)–(20) respectively,

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (\mathbf{X}_i - \hat{\mathbf{X}}_i)^2, \quad (18)$$

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C \mathbf{Y}_{ij} \log(\hat{\mathbf{Y}}_{ij}), \quad (19)$$

$$\mathcal{L} = \mathcal{L}_{MSE} + \mathcal{L}_{CE} + \lambda \sum w^2, \quad (20)$$

$\mathcal{L}_{MSE}$ is the reconstruction loss between normalized $\mathbf{X} = [\mathbf{X}_T \parallel \mathbf{X}_G]$ and reconstructed expression data $\hat{\mathbf{X}} = [\hat{\mathbf{X}}_T \parallel \hat{\mathbf{X}}_G]$ to minimize the loss of regulatory information that occurs when converting embeddings from gene to cell within Regulation-level Encoder; $\mathcal{L}_{CE}$ is the classification loss of ground truth labels $\mathbf{Y}$ and predicted probabilities $\hat{\mathbf{Y}}$ to enhance classification performance; and $\lambda \sum w^2$ is L2 regularization applied to all parameters ensuring a more stable training process and improved robustness. By minimizing $\mathcal{L}$ with the SGD optimizer, ScanNet preserves gene regulatory information to the greatest extent while simultaneously enhancing cell type prediction through the inferred cell potential embedding.

Recognize novel types by prediction uncertainty

In real application scenarios, the emergence of previous unknown cell types, also referred as novel types, is frequently encountered. In other words, real cell data not only contain known types trained during training process but may also include novel cell types. Therefore, prediction uncertainty is introduced as a quantitative metric of output confidence, while keeping the model architecture unchanged. Prediction uncertainty is quantified via the entropy of the predicted probability distribution derived from classifier outputs, where higher entropy indicates a more uniform distribution, reflecting greater uncertainty in predictions, and lower entropy implies greater confidence.

To emulate real application scenarios, the experimental design is as follows:

$$\mathbf{X}'_{test} = \mathbf{X}_{test} \cup \mathbf{X}_{novel}, \quad (21)$$

$$\hat{\mathbf{Y}}_{test} = \text{ScanNet}(\mathbf{X}'_{test}, \mathbf{A}^{T-G}), \quad (22)$$

where $\mathbf{X}_{novel}$ represents cells of novel types, not encountered in training, and is integrated into the test set $\mathbf{X}_{test}$ to create a new test set $\mathbf{X}'_{test}$ containing known and novel cell types, then $\mathbf{X}'_{test}$ is fed into the pre-trained ScanNet to obtain the probability distribution $\hat{\mathbf{Y}}_{test}$ over predefined classes. We hypothesize that prediction uncertainty, quantified via entropy, will be higher for cells of novel types compared to known types. Hence, prediction uncertainty is calculated for each cell and cells with high uncertainty are identified as novel types. Mathematically, given the probability distribution $\hat{\mathbf{Y}}_{test}$ from the Classifier, the prediction uncertainty is calculated as Eq (23):

$$\mathbf{E}_i = -\sum_{j=1}^C \hat{\mathbf{Y}}_{ij} \log(\hat{\mathbf{Y}}_{ij}), \quad (23)$$

where $\hat{Y}_{ij}$ denotes the predicted probability that the i^{th} cell belongs to the j^{th} class in $\hat{Y}_{test}$, C represents the number of predefined classes during training and E_i is the prediction uncertainty of i^{th} cell.

Then, the decision is made by comparing prediction uncertainty E_i with the threshold t :

$$\hat{y}_i = \begin{cases} C_{novel} & \text{if } E_i > t \\ \operatorname{argmax}_j (\hat{Y}_{ij}) & \text{if } E_i \leq t \end{cases} \quad (24)$$

When the prediction uncertainty exceeds t , the probability distribution is more dispersed, reflecting less confidence on the prediction, and such cells are assigned as novel C_{novel} . Conversely, if the prediction uncertainty is less than t , ScanNet show high confidence on the prediction and the cell is assigned to the class with the highest probability. In this manuscript, the default value of threshold t is set to 0.75. C_{novel} indicates that the cell belongs to novel types not encountered during prior training.

Code availability

The code used for model training and evaluation is available on GitHub (<https://github.com/Ying-Lab/ScanNet>).

Supporting information

S1 Table. Description of the PBMCbench dataset.

(DOCX)

S2 Table. Detailed performance metrics for cell annotation comparison.

(DOCX)

S3 Table. Cross-platform performance comparison.

(DOCX)

S4 Table. Batch correction comparison.

(DOCX)

S5 Table. Annotation performance on scATAC-seq data.

(DOCX)

S6 Table. Ablation analysis of embedding and graph in ScanNet.

(DOCX)

S7 Table. ScanNet performance under different loss weighting schemes.

(DOCX)

S8 Table. ScanNet performance under different L2 regularization coefficient.

(DOCX)

S9 Table. Performance comparison between ScanNet and ACTINN (all genes).

(DOCX)

S10 Table. Robustness of ScanNet to variations in TF and TG sizes.

(DOCX)

S11 Table. Description of variables and operators.

(DOCX)

S1 Fig. Schematic diagram of transferring ScanNet to scATAC-seq data.

(DOCX)

S2 Fig. Critical difference diagrams for F1 and Recall.

(DOCX)

S3 Fig. Quantity statistics of the PBMC dataset.

(DOCX)

S4 Fig. T-SNE visualizations on the test set of all baselines.

(DOCX)

S5 Fig. Accuracy confusion matrix in detecting novel types for more cells.

(DOCX)

S6 Fig. Integration of ScanNet with marker gene-based annotation workflow.

(DOCX)

S7 Fig. Annotation of novel type cells using the integrated ScanNet-scCATCH workflow and scCATCH alone.

(DOCX)

S8 Fig. Boxplots of embedding and graph ablation in ScanNet.

(DOCX)

S9 Fig. Training dynamics of MSE and CE losses on the Muraro dataset.

(DOCX)

S10 Fig. Clustering of cell types based on reconstructed TRN.

(DOCX)

S1 Text. Data preprocessing.

(DOCX)

S2 Text. Experimental design of ScanNet transferred to scATAC-seq.

(DOCX)

S3 Text. Performance metrics.

(DOCX)

S4 Text. ScanNet can identify novel types for more cells.

(DOCX)

S5 Text. Reconstruction of cell-specific TF-TG regulatory networks.

(DOCX)

Acknowledgments

Shaorong Fang and Tianfu Wu from the Information and Network Center of Xiamen University are acknowledged for their help with high-performance computing (HPC).

Author contributions

Conceptualization: Yongyu Long, Wenhao Zhang, Lan Cao.

Data curation: Yongyu Long.

Formal analysis: Yongyu Long, Wenhao Zhang.

Funding acquisition: Xiaobing Huang, Ying Wang.

Investigation: Yongyu Long, Lan Cao.

Methodology: Yongyu Long.

Project administration: Wenhao Zhang, Lan Cao.

Software: Yongyu Long.

Supervision: Xiaobing Huang, Ying Wang.

Validation: Yongyu Long.

Visualization: Wenhao Zhang, Lan Cao.

Writing – original draft: Yongyu Long.

Writing – review & editing: Wenhao Zhang, Lan Cao.

References

1. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*. 2017;14(11):1083–6. <https://doi.org/10.1038/nmeth.4463> PMID: 28991892
2. Fiers MWEJ, Minnoye L, Aibar S, Bravo González-Blas C, Kalender Atak Z, Aerts S. Mapping gene regulatory networks from single-cell omics data. *Brief Funct Genomics*. 2018;17(4):246–54. <https://doi.org/10.1093/bfpg/elx046> PMID: 29342231
3. Wang L, Yu P, Zhou B, Song J, Li Z, Zhang M, et al. Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. *Nat Cell Biol*. 2020;22(1):108–19. <https://doi.org/10.1038/s41556-019-0446-7> PMID: 31915373
4. Lu Y, Chen JJ, Mu L, Xue Q, Wu Y, Wu P-H, et al. High-throughput secretomic analysis of single cells to assess functional cellular heterogeneity. *Anal Chem*. 2013;85(4):2548–56. <https://doi.org/10.1021/ac400082e> PMID: 23339603
5. Abdelaal T, Michielsen L, Cats D, Hoogduin D, Mei H, Reinders MJT, et al. A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol*. 2019;20(1):194. <https://doi.org/10.1186/s13059-019-1795-z> PMID: 31500660
6. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473(7345):43–9. <https://doi.org/10.1038/nature09906> PMID: 21441907
7. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res*. 2015;25(10):1491–8. <https://doi.org/10.1101/gr.190595.115> PMID: 26430159
8. Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. *Nat Methods*. 2018;15(5):359–62. <https://doi.org/10.1038/nmeth.4644> PMID: 29608555
9. de Kanter JK, Lijnzaad P, Candelli T, Margaritis T, Holstege FCP. CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing. *Nucleic Acids Res*. 2019;47(16):e95. <https://doi.org/10.1093/nar/gkz543> PMID: 31226206
10. Boufe A, Seth S, Batada NN. scID Uses Discriminant Analysis to Identify Transcriptionally Equivalent Cell Types across Single-Cell RNA-Seq Data with Batch Effect. *iScience*. 2020;23(3):100914. <https://doi.org/10.1016/j.isci.2020.100914> PMID: 32151972
11. Lieberman Y, Rokach L, Shay T. CaSTLe - Classification of single cells by transfer learning: Harnessing the power of publicly available single cell RNA sequencing experiments to annotate new experiments. *PLoS One*. 2018;13(10):e0205499. <https://doi.org/10.1371/journal.pone.0205499> PMID: 30304022
12. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*; 2016.
13. Alquicira-Hernandez J, Sathe A, Ji HP, Nguyen Q, Powell JE. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol*. 2019;20(1):264. <https://doi.org/10.1186/s13059-019-1862-5> PMID: 31829268
14. Xie P, Gao M, Wang C, Zhang J, Noel P, Yang C, et al. SuperCT: a supervised-learning framework for enhanced characterization of single-cell transcriptomic profiles. *Nucleic Acids Res*. 2019;47(8):e48. <https://doi.org/10.1093/nar/gkz116> PMID: 30799483
15. Ma F, Pellegrini M. ACTINN: automated identification of cell types in single cell RNA sequencing. *Bioinformatics*. 2020;36(2):533–8. <https://doi.org/10.1093/bioinformatics/btz592> PMID: 31359028
16. Song Q, Su J, Zhang W. scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. *Nat Commun*. 2021;12(1):3826. <https://doi.org/10.1038/s41467-021-24172-y> PMID: 34158507
17. Fan X, Liu J, Yang Y, Gu C, Han Y, Wu B, et al. scGraphformer: unveiling cellular heterogeneity and interactions in scRNA-seq data using a scalable graph transformer network. *Commun Biol*. 2024;7(1):1463. <https://doi.org/10.1038/s42003-024-07154-w> PMID: 39511415

18. Wang T, Bai J, Nabavi S. Single-cell classification using graph convolutional networks. *BMC Bioinformatics*. 2021;22(1):364. <https://doi.org/10.1186/s12859-021-04278-2> PMID: 34238220
19. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13. <https://doi.org/10.1093/nar/gky1131> PMID: 30476243
20. Yin Q, Liu Q, Fu Z, Zeng W, Zhang B, Zhang X, et al. scGraph: a graph neural network-based approach to automatically identify cell types. *Bioinformatics*. 2022;38(11):2996–3003. <https://doi.org/10.1093/bioinformatics/btac199> PMID: 35394015
21. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. *Advances in neural information processing systems*. 2017;30.
22. Yuan L, Sun S, Jiang Y, Zhang Q, Ye L, Zheng C-H, et al. scRGCL: a cell type annotation method for single-cell RNA-seq data using residual graph convolutional neural network with contrastive learning. *Brief Bioinform*. 2024;26(1):bbae662. <https://doi.org/10.1093/bib/bbae662> PMID: 39708840
23. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol*. 2004;14(3):283–91. <https://doi.org/10.1016/j.sbi.2004.05.004> PMID: 15193307
24. Müller-Dott S, Tsirvouli E, Vazquez M, Ramirez Flores RO, Badia-I-Mompel P, Fallegger R, et al. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Res*. 2023;51(20):10934–49. <https://doi.org/10.1093/nar/gkad841> PMID: 37843125
25. Iacono G, Massoni-Badosa R, Heyn H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol*. 2019;20(1):110. <https://doi.org/10.1186/s13059-019-1713-4> PMID: 31159854
26. Cha J, Lee I. Single-cell network biology for resolving cellular heterogeneity in human diseases. *Exp Mol Med*. 2020;52(11):1798–808. <https://doi.org/10.1038/s12276-020-00528-0> PMID: 33244151
27. Bonzanni N, Garg A, Feenstra KA, Schütte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013;29(13):i80–8. <https://doi.org/10.1093/bioinformatics/btt243> PMID: 23813012
28. Cao Y, Zhao X, Tang S, Jiang Q, Li S, Li S, et al. scButterfly: a versatile single-cell cross-modality translation method via dual-aligned variational autoencoders. *Nat Commun*. 2024;15(1):2973. <https://doi.org/10.1038/s41467-024-47418-x> PMID: 38582890
29. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst*. 2016;3(4):385–394.e3. <https://doi.org/10.1016/j.cels.2016.09.002> PMID: 27693023
30. Luecken M, Burkhardt D, Cannoodt R, Lance C, Agrawal A, Aliee H, et al. A sandbox for prediction and integration of DNA, RNA, and protein data in single cells. *Proc Thirty*. 2020:1–13.
31. Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*. 2022;376(6594):eabl5197. <https://doi.org/10.1126/science.abl5197> PMID: 35549406
32. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst*. 2016;3(4):346–360.e4. <https://doi.org/10.1016/j.cels.2016.08.011> PMID: 27667365
33. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol*. 2019;37(12):1452–7. <https://doi.org/10.1038/s41587-019-0290-0> PMID: 31611697
34. Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*. 2018;563(7729):72–8. <https://doi.org/10.1038/s41586-018-0654-5> PMID: 30382198
35. Schaum N, Karkanas J, Neff NF, May AP, Quake SR, Wyss-Coray T, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris: The Tabula Muris Consortium. *Nature*. 2018;562(7727):367.
36. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med*. 2018;50(8):1–14. <https://doi.org/10.1038/s12276-018-0071-8> PMID: 30089861
37. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol*. 2020;38(6):737–46. <https://doi.org/10.1038/s41587-020-0465-8> PMID: 32341560
38. Du X, Gozum G, Ming Y, Li Y. SIREN: Shaping Representations for Detecting Out-Of-Distribution Objects. In: *Advances in Neural Information Processing Systems 35*, 2022. 20434–49. <https://doi.org/10.52202/068431-1486>
39. Yang J, Wang P, Zou D, Zhou Z, Ding K, Peng W, et al. OpenOOD: Benchmarking Generalized Out-Of-Distribution Detection. In: *Advances in Neural Information Processing Systems 35*, 2022. 32598–611. <https://doi.org/10.52202/068431-2362>
40. Zhang W, Jiang R, Chen S, Wang Y. scIBD: a self-supervised iterative-optimizing model for boosting the detection of heterotypic doublets in single-cell chromatin accessibility data. *Genome Biol*. 2023;24(1):225. <https://doi.org/10.1186/s13059-023-03072-y> PMID: 37814314
41. Ma W, Lu J, Wu H. Cellcano: supervised cell type identification for single cell ATAC-seq data. *Nat Commun*. 2023;14(1):1864. <https://doi.org/10.1038/s41467-023-37439-3> PMID: 37012226
42. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. *Nat Methods*. 2021;18(11):1333–41. <https://doi.org/10.1038/s41592-021-01282-5> PMID: 34725479
43. Liberty E, Lang K, Shmakov K. Stratified sampling meets machine learning. *International conference on machine learning*; 2016:
44. Bhadani R, Chen Z, An L. Attention-Based Graph Neural Network for Label Propagation in Single-Cell Omics. *Genes (Basel)*. 2023;14(2):506. <https://doi.org/10.3390/genes14020506> PMID: 36833434

45. Song Q, Su J, Zhang W. scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. *Nat Commun*. 2021;12(1):3826. <https://doi.org/10.1038/s41467-021-24172-y> PMID: [34158507](#)
46. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol*. 2021;17(1):e9620. <https://doi.org/10.15252/msb.20209620> PMID: [33491336](#)
47. Kong Y, Yu T. A graph-embedded deep feedforward network for disease outcome classification and feature selection using gene expression data. *Bioinformatics*. 2018;34(21):3727–37. <https://doi.org/10.1093/bioinformatics/bty429> PMID: [29850911](#)
48. Das J, Yu H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol*. 2012;6:92. <https://doi.org/10.1186/1752-0509-6-92> PMID: [22846459](#)
49. Liu Y, Chen W. A SAS macro for testing differences among three or more independent groups using Kruskal-Wallis and Nemenyi tests. *J Huazhong Univ Sci Technolog Med Sci*. 2012;32(1):130–4. <https://doi.org/10.1007/s11596-012-0023-9> PMID: [22282259](#)
50. Demšar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*. 2006;7(Jan):1–30.
51. Shao X, Liao J, Lu X, Xue R, Ai N, Fan X. scCATCH: Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data. *iScience*. 2020;23(3):100882. <https://doi.org/10.1016/j.isci.2020.100882> PMID: [32062421](#)
52. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Advances in neural information processing systems*. 2017;30.
53. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *arXiv preprint arXiv:1710.10903*. 2017.