Graph neural networks for single-cell omics data: a review of approaches and applications

Sijie Li [iD]‡, Heyang Hua [iD]‡, Shengquan Chen [iD]*

School of Mathematical Sciences and The Key Laboratory of Pure Mathematics and Combinatorics, Ministry of Education (LPMC), Nankai University, No. 94 Weijin Road, Nankai District, Tianjin 300071, China

*Corresponding author. School of Mathematical Sciences and The Key Laboratory of Pure Mathematics and Combinatorics, Ministry of Education (LPMC), Nankai University, No. 94 Weijin Road, Nankai District, Tianjin 300071, China. E-mail: chenshengquan@nankai.edu.cn

‡Sijie Li and Heyang Hua contributed equally to this work.

Abstract

Rapid advancement of sequencing technologies now allows for the utilization of precise signals at single-cell resolution in various omics studies. However, the massive volume, ultra-high dimensionality, and high sparsity nature of single-cell data have introduced substantial difficulties to traditional computational methods. The intricate non-Euclidean networks of intracellular and intercellular signaling molecules within single-cell datasets, coupled with the complex, multimodal structures arising from multi-omics joint analysis, pose significant challenges to conventional deep learning operations reliant on Euclidean geometries. Graph neural networks (GNNs) have extended deep learning to non-Euclidean data, allowing cells and their features in single-cell datasets to be modeled as nodes within a graph structure. GNNs have been successfully applied across a broad range of tasks in single-cell data analysis. In this survey, we systematically review 107 successful applications of GNNs and their six variants in various single-cell omics tasks. We begin by outlining the fundamental principles of GNNs and their six variants, followed by a systematic review of GNN-based models applied in single-cell epigenomics, transcriptomics, spatial transcriptomics, proteomics, and multi-omics. In each section dedicated to a specific omics type, we have summarized the publicly available single-cell datasets commonly utilized in the articles reviewed in that section, totaling 77 datasets. Finally, we summarize the potential shortcomings of current research and explore directions for future studies. We anticipate that this review will serve as a guiding resource for researchers to deepen the application of GNNs in single-cell omics.

Keywords: graph neutral networks; single-cell; epigenomics; transcriptomics; proteomics

Introduction

The advent of bulk sequencing technologies such as ribonucleic acid (RNA) sequencing (RNA-seq) [1, 2], assay of transposase accessible chromatin sequencing (ATAC-seq) [3, 4], and bisulfite sequencing (BS-seq) [5, 6] has significantly advanced the investigation of differential gene expression [7], detection of chromatin accessibility changes between different cell states [8], and interpretation of biological processes such as cellular development and aging through deoxyribonucleic acid (DNA) methylation, providing insights into the mechanisms and etiology of diseases like cancer and leukemia [8–10]. However, bulk sequencing technologies measure the average signals of cells in the whole tissue, thus failing to capture the heterogeneity among individual cells [11]. Rapid revolution in sequencing technology has facilitated the emergence of sequencing technologies at single-cell resolution, including single-cell RNA sequencing (scRNA-seq) [12], spatial resolved transcriptomics profiling at single-cell resolution [13–15], single-cell chromatin accessibility sequencing (scCAS) [16–20], single-cell DNA methylation (scDNAm) [21], single-cell Hi-C [22], single-cell proteomics sequencing [23, 24], and single-cell multi-omics sequencing [25–32].

Single-cell data provides a revolutionary scale and high resolution that bulk data lacked in the past, allowing for the characterization of cell populations at an unprecedented level, especially the heterogeneity of cell subtypes and rare cell types, as well as

various cell states [33, 34]. It has been widely utilized for clustering and classification to identify cell populations and reported to shed light on cell type heterogeneity through various downstream analyses [35–39], facilitating exploration of complex phenotypes and potential gene regulatory mechanisms. Single-cell omics data also enables the decipherment of differentiation trajectories [40], inference of cell–cell communications [41] and Granger causal relationship [42], elucidation of the diversity of cell phenotypes within the tumor microenvironment and inference of the tumor lineages [43, 44], play a crucial role in uncovering cancer biology.

Although single-cell data provides a high-resolution perspective for researching cell heterogeneity, modeling single-cell data presents significant challenges due to the unique characteristics of these omics datasets. Single-cell data is high-dimensional, sparse, and often noisy, making it difficult to directly apply traditional statistical and machine learning methods. Furthermore, the complex biological relationships between different types of omics data (e.g. gene expression, chromatin accessibility, DNA methylation) are difficult to model using traditional machine learning approaches. These challenges necessitate the development of novel computational approaches, one feasible method being to model single-cell data as graphs. Actually, both cells and features in single-cell data can be regarded as nodes, potentially connected by various types of links, such as cell–cell communications [45–47] and gene–gene regulatory interactions [48]. Such graph structure
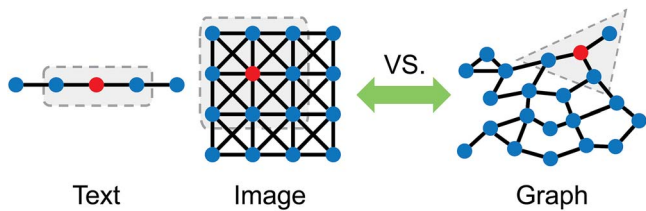
Figure 1. 1D and 2D data versus graph data. 1D data (e.g. text) is Euclidean, with a fixed sequence that allows for linear processing, such as sequential convolutions over the data. 2D data (e.g. images) is also Euclidean, with a grid-like structure that allows for spatial convolution operations to capture local features like edges and textures. Unlike one- and 2D data, graph data is non-Euclidean. It does not have a fixed structure or linear order, and its relationships are defined by the graph's topology.

alters the fundamental assumption of traditional machine learning algorithms that the random variables are independent [49]. However, the non-Euclidean structure of single-cell data, unlike the grid-like structure of 1D and 2D data (Fig. 1), poses challenges to deep learning operations such as convolution, which are traditionally designed for Euclidean space [50]. Emerging graph neural networks (GNNs) have extended deep learning to non-Euclidean structure data [51] and demonstrated promising potential for the exploration of single-cell data.

GNNs [52, 53] are end-to-end deep learning neural network models that have recently made remarkable progress. Several variants of GNNs, including graph convolutional networks (GCNs) [54], GraphSAGE [55], graph attention networks (GATs) [56], graph transformer networks (GTNs) [57], graph autoencoders (GAEs) [5, 8] and variational graph autoencoders (VGAEs) [58] have been extensively applied in the field of bioinformatics [59] and proven highly effective across various applications in diverse areas, such as disease prediction [60, 61] and drug discovery [62].

For single-cell omics analysis, GNN-based models can learn low-dimensional representations integrated with topological information from various graphs such as cell–cell [63], cell-gene [64], and gene–gene [65] graphs. GNN models have exhibited remarkable capabilities in a wide range of single-cell omics analysis, including preprocessing tasks such as enhancement, imputation, and batch effect correction of the single-cell data, as well as key tasks like cell clustering and classification to identify cell types [66–70]. Besides, GNN models also contribute to gaining valuable biological insights by modeling single-cell data for downstream analyses such as gene regulatory network reconstruction and intercellular communication. Notably, GNN-based approaches for single-cell omics have demonstrated superior performance compared to traditional methods, including those specifically designed for single-cell data analysis and general traditional machine learning techniques (Supplementary Text S1). Therefore, summarizing the application of GNN models in single-cell omics is essential for advancing single-cell data analysis. Lazaros et al. have provided a foundational reference for the application of GNN-based models in single-cell data [71]. However, they mainly focus on applications for scRNA-seq and spatial transcriptomics (ST) data, with limited discussion on single-cell epigenomics, such as scCAS data and scDNAm data. Additionally, their work did not comprehensively consider all main GNN variants, for example, graph transformer networks are not adequately covered in their study. Furthermore, dozens of novel GNN-based models have not been included, partially due to their timeliness.

Our comprehensive survey provides a concise overview of the principles behind various variants of GNN models and thoroughly explores their notable applications in single-cell omics analyses, keeping abreast of their latest related developments. Firstly, we elaborate on the principles of GNNs and introduce six variants of GNNs that play a significant role in single-cell omics analysis, including GCNs, GraphSAGE, GATs, GTNs, GAEs, and VGAEs. Then, we investigate their key applications across multiple modalities of single-cell data, including scCAS data, scDNAm data, scHi-C data, scRNA-seq data, spatially resolved transcriptomics (SRT) data, and single-cell multi-omics data. We review all the GNN models designed in >107 publications, focusing on their applications in various tasks such as data imputation, dimensionality reduction, cell type identification, cell type deconvolution, spatial domain identification, and multi-omics integration. In each omics-specific chapter, we have also summarized the publicly available single-cell datasets commonly used in the GNN-based articles reviewed in that section, with a total of 77 commonly used datasets collected throughout the manuscript. Finally, we elucidate the inherent challenges for each single-cell modality data and discuss the pros and cons of the GNN-based single-cell modeling methodologies. We anticipate that this survey will provide valuable guidance and reference for the future applications of GNNs and the development of computational methods for single-cell data.

Principles of graph neural networks: a brief overview

In this section, we provide a concise introduction to GNN and six commonly used GNN variants that we captured during our review process, enabling quick reference for readers. We recommend [51, 72] for further extensive review of GNNs.

A graph can be represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the set of nodes and edges in the graph, respectively. The adjacency matrix A of $\mathcal{G}$ is a $N_v \times N_v$ matrix, where $N_v = |\mathcal{V}|$ is the number of nodes. If there exists an edge from node v to node u, i.e. $e_{vu} \in \mathcal{E}$, then $\mathbf{A}(v, u)$ represents the weight of e_{vu}; otherwise $\mathbf{A}(v, u) = 0$. For undirected graphs, $\mathbf{A}(v, u) = 1$ if $e_{vu} \in \mathcal{E}$ and A is a symmetric matrix, which does not hold necessarily for directed graphs. A graph is homogeneous if it contains only one type of nodes and one type of edges; otherwise, it is classified as a heterogeneous graph. The edge features $\mathbf{e}_{vu}$ represents the attribute information of edge e_{vu}. Edge features can be discrete, such as binary values indicating the existence of an edge, or continuous, representing the strength of the relationship between nodes. Additionally, they can be both unidimensional, as previously mentioned, or multidimensional, capturing the complex relationships between nodes. We have summarized the frequently used symbols in this article and their corresponding meanings in Table 1 for conciseness.

The GNN model was first introduced by Gori et al. [52] and Scarselli et al. [53]. GNN aims to extend neural network approaches to handle various types of graph inputs, such as directed and undirected graphs. It learns the representation of each node v, denoted as the hidden state $\boldsymbol{h}_v \in \mathbb{R}^s$, leveraging the neighborhood of v. GNN mainly consists of the forward and backward process. The forward process involves a parametric function $f_{\boldsymbol{w}}$, known as the transition function, where $\boldsymbol{w}$ represents a set of parameters. $f_{\boldsymbol{w}}$ aggregates information from the edges connected to a node and its neighboring nodes to generate the hidden state of the node. However, obtaining the nodes' hidden states is not the ultimate aim. The goal of a GNN is not merely to generate the hidden states of nodes. These hidden states can be seen as feature representations of nodes, but their real significance lies in their application to practical tasks. To achieve

Table 1. Notations used in this study.

Notations	Descriptions
$\mathcal{G}$	A graph
$\mathcal{V}$	The set of nodes in $\mathcal{G}$
$\mathcal{E}$	The set of edges in $\mathcal{G}$
v	A node $v \in \mathcal{V}$
e_{vu}	Edge $e_{vu} \in \mathcal{E}$ between node v and node u
A	The adjacency matrix of $\mathcal{G}$
$\mathbf{A}^T$	The transpose of matrix A
N_v	The number of nodes in $\mathcal{G}$
N_e	The number of edges in $\mathcal{G}$
$\mathcal{N}(v)$	The neighborhood of node v, $\mathcal{N}(v) = \{ u \in \mathcal{V} \| (v, u) \in \mathcal{E} \}$
$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbf{y}_v$	The feature of node v
$\mathbf{y}_{co[v]}$	The features of the edges of node v
$\mathbf{e}_{vu}$	The features of the edge e_{vu}
$\mathbf{y}_{\mathcal{N}(v)}$	The features of the nodes in $\mathcal{N}(v)$
$\mathbf{h}_v$	The hidden state of node v
$\mathbf{h}_v^{(t)}$	The hidden state of node v at the t-th iteration
$\mathbf{h}_{\mathcal{N}(v)}$	The hidden states of the nodes in $\mathcal{N}(v)$
$\mathbf{h}_{\mathcal{N}(\sqsubseteq)}^{(t)}$	The hidden states of the nodes in $\mathcal{N}(v)$ at the t-th iteration
$\|\cdot\|$	The number of elements in a set
$\sigma(\cdot)$	The activation function
$\mathbf{I}_N$	The $N \times N$ identity matrix
$\mathbf{D}$	The degree matrix of $\mathcal{G}$, $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$
$\|$	The concatenation operator

specific tasks, such as node classification, the GNN uses a local output function $g_{\mathbf{w}}$, which is typically a learnable transformation function, to convert these hidden states into final output results, such as node labels. The states of nodes and outputs are iteratively computed as follows:

$$\mathbf{h}_v^{(t+1)} = f_{\mathbf{w}} \left(\mathbf{y}_v, \mathbf{y}_{co[v]}, \mathbf{h}_{\mathcal{N}(v)}^{(t)}, \mathbf{y}_{\mathcal{N}(v)} \right) \qquad (1)$$

$$\mathbf{o}_v^{(t)} = g_{\mathbf{w}} \left(\mathbf{h}_v^{(t)}, \mathbf{y}_v \right) \qquad (2)$$

where $\mathbf{o}_v^{(t)}$ denotes the outputs of the i-th iteration and $\mathbf{y}_{co[v]}$ represents the features of the edges of node v. Given a threshold ε_f and a norm $\|\cdot\|$, the aforementioned iteration stops when $\|\mathbf{h}_v^{(t)} - \mathbf{h}_v^{(t-1)}\| < \varepsilon_f$. GNN then calculates the gradient in the backward process to update the parameter set $\mathbf{w}$. GNN iteratively performs the forward and backward processes until a specified stopping criterion is reached.

GNNs have evolved into several key variants designed to tackle different challenges in graph learning [54–58, 73, 74]. GCNs extend traditional CNNs to graph-structured data by aggregating information from neighboring nodes (Fig. 2a). GraphSAGE improves scalability by using a sampling-based aggregation approach, making it more efficient for large graphs (Fig. 2b and 2c). GATs introduce an attention mechanism that allows nodes to weigh the importance of their neighbors, enhancing the learning process (Fig. 2d). GTNs combine graph structures with transformer-based attention, enabling more flexible and dynamic learning of graph data. GAEs and VGAEs, based on autoencoder architectures, focus on unsupervised learning of graph representations, with VGAEs incorporating variational inference to model probabilistic graph structures (Fig. 2e). In the Supplementary Texts S2–S7, we provide a detailed illustration of these six key variants of GNNs, each of which introduces unique innovations to better capture

graph structures and improve performance on a wide range of applications. There are also other GNN models, such as message passing neural networks (MPNNs; Supplementary Text S8). We also provided a brief discussion on the difference between GATs and GTNs in the Supplementary Text S9.

Applications of graph neural network-based approaches for single-cell omics data analysis

Recent advancements in single-cell resolution technologies have revolutionized the profiling of chromatin accessibility, DNA methylation, and proteins, enhancing our understanding of the development, differentiation, aging, and various diseases and phenotypes at the genetic regulation and expression levels. Various derivatives of graph neural networks have been developed into hundreds of powerful bioinformatics computational methods for a diverse array of tasks in single-cell omics and single-cell multi-omics. In this section, we review the diverse applications of GNN-based computational methods in the analysis of single-cell omics data (Fig. 3). We also provided an overall introduction to the single-cell-processing pipeline for each application and the roles of GNNs in these pipelines in Supplementary Text S10.

Graph neural networks for single-cell epigenomics

GNN-based methods have shown significant potential in analyzing single-cell epigenomics, providing a deeper understanding of the genome's regulatory landscape. This section explores the applications of GNN-based methods in three fundamental areas of single-cell epigenomics: chromatin accessibility, DNA methylation, and chromosome conformation. These data types are often sparse, with each cell exhibiting distinct epigenetic marks that define its state. (i) Chromatin accessibility, which influences gene expression by determining the openness of chromatin regions, can be profiled at single-cell resolution using techniques like scATAC-seq, revealing key regulatory dynamics [75, 76]. (ii) DNA methylation, an important epigenetic modification involving methyl group addition to DNA, is essential for gene expression regulation and cell identity. scDNAm provides detailed information on methylation diversity and lineage relationships. (iii) Chromosome conformation refers to the 3D organization of the genome, which impacts gene regulation and genome stability. Single-cell Hi-C and related technologies allow researchers to capture spatial chromatin interactions, shedding light on the nuclear architecture. In the following sections, we will introduce various GNN-based methods and their applications in analyzing chromatin accessibility, DNA methylation, and chromosome conformation.

Chromatin accessibility

Cell type identification is a key task in the analysis of scCAS data, enabling the exploration of complex, cell-type-specific gene regulatory mechanisms and enhancing our understanding of cellular functions and their alterations in health and disease states [77, 78]. In terms of methodologies, cell type identification in scCAS data can be divided into two main approaches: transferring labels from well-annotated reference datasets from other omics and directly annotating cell types using well-labeled scCAS data as references. In the following sections, we discuss several GNN-based methods that exemplify these approaches.

For methods that transfer labels from well-annotated reference datasets from other omics to annotate scCAS data, Song *et al.*
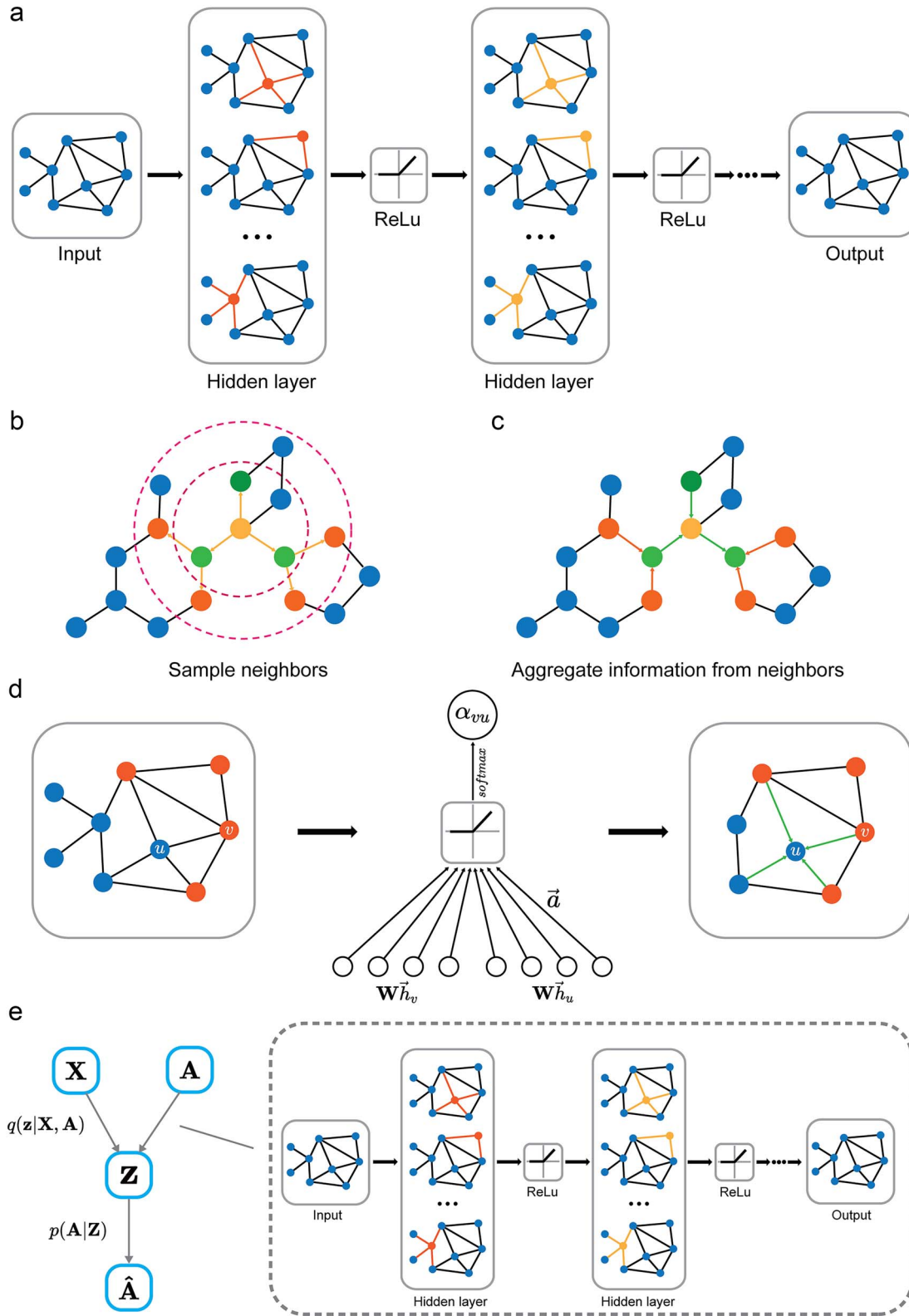
Figure 2. Overview of several GNN variants. (a) A typical pipeline of GCN, from input graph data and node features to feature extraction through convolutional layers, followed by non-linear activation and pooling for downstream tasks like node or graph classification. (b and c) An illustration of the information aggregation progress in GraphSAGE. In (b), the neighbor sampling process is shown, where a fixed-size subset of neighbors is selected for each node; In (c), these sampled neighbors' features are aggregated using an appropriate aggregation function to update the node representations. (d) An illustration of GATs, where attention coefficients are computed to weight the importance of each neighbor's features, allowing for adaptive aggregation based on the learned attention scores. (e) A pipeline of VGAEs, where graph data is encoded into latent space using a variational approach, followed by a decoder that reconstructs the graph structure, enabling unsupervised learning of graph representations.
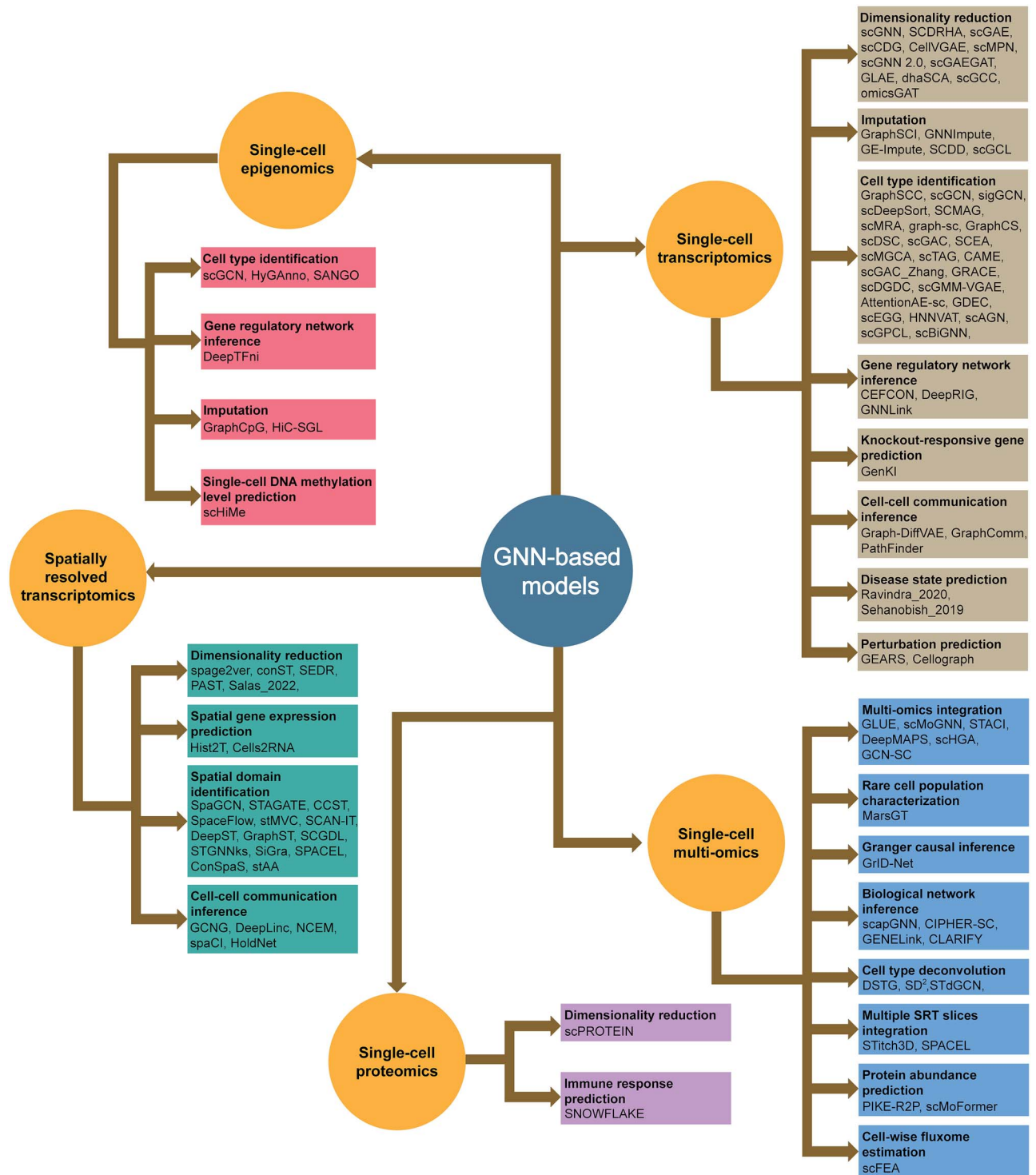
Figure 3. An overview of GNN-based models for various omics, categorized by the types of omics they process and the tasks they address.

introduced scGCN [79], which effectively transfers labels from annotated scRNA-seq data to scCAS data in a semi-supervised manner using the scRNA-seq dataset as a reference. scGCN first establishes graph structures within the query dataset and between the reference and query data. It then utilizes a three-layer modified GCN [69] to annotate the cell type labels of the scCAS dataset using the labeled scRNA-seq dataset. The GCNs take the adjacency matrix of a hybrid graph composed of the above two graph structures and the concatenated feature matrices of the reference and query datasets. scGCN is trained by computing the loss function from the predicted results on the reference dataset against its known labels.

However, scGCN preprocesses scCAS data and converts peak features into gene activity, resulting in a training process that lacks informative peak-level features. HyGAnno [80] employs annotated scRNA-seq data as reference and trains a parallel VGAE model that incorporates both gene-level and peak-level features to transfer cell type labels to scCAS data. It demonstrated improved performance in cell type annotation across multiple datasets compared to scGCN.

In addition to methods that use other omics to annotate scCAS data, there are also methods like SANGO [81] that directly utilize well-labeled scCAS data to annotate cell types in new scCAS datasets. Initially, SANGO performs one-hot encoding of the DNA sequences corresponding to peaks in both the query scCAS dataset and the well-labeled reference scCAS dataset. It then uses a channel attention CNN to learn low-dimensional embeddings for cells from these encoded sequences. These learned embeddings, representing both the reference and query datasets, are subsequently fed into a GTN. The GTN is fine-tuned using the known cell type annotations from the reference dataset, which allows it to predict and annotate cell types in the query dataset. In this way, SANGO effectively performs cell type annotation for scCAS data.

Deoxyribonucleic acid methylation

DNA methylation is a key epigenetic modification involved in regulating gene expression and influencing development and disease. However, single-cell DNA methylation data often suffers from significant sparsity due to the limited DNA available from individual cells, leading to numerous missing values. To address this challenge, methylome imputation aims to infer these missing values, thereby providing a more complete understanding of methylation patterns and epigenetic regulation.

GraphCpG [82] addresses the challenge of methylome imputation by transforming the DNA methylation matrix into an undirected bipartite graph, where nodes represent either cells or loci, and edges depict the methylation state of a locus within a cell. The model first constructs this bipartite graph from the methylation data, with edges indicating known methylation states. Using a GCN-based model, GraphCpG learns embeddings for both cell and locus nodes by aggregating information from their neighbors, capturing relationships between similar cells and loci. These embeddings are then fed into a decoder to perform link prediction, effectively imputing missing methylation values by estimating the likelihood of connections between cells and loci. By leveraging the spatial and structural information inherent in the data, GraphCpG proves to be more effective than traditional statistical methods for methylome imputation.

Chromosome conformation

Chromosome conformation analysis at the single-cell level is challenging due to the high sparsity and complexity of the data, particularly with single-cell Hi-C (scHi-C), which aims to capture chromosomal interactions in individual cells. Compared to bulk Hi-C data, scHi-C data exhibits significantly higher sparsity, prompting the development of computational methods to impute missing chromosomal interactions.

To address the challenge of missing values, HiC-SGL [83] models scHi-C data as a set of graphs, with each cell represented by a graph whose nodes are chromosomal segments (bins) and edges represent interactions between these bins. HiC-SGL utilizes a cell-encoder to capture embeddings that reflect global features of the cell graph, and incorporates a transformer within the GNN to extract subgraph representations. The decoder predicts the existence probability of each edge within the cell graph. Furthermore, HiC-SGL has demonstrated the capacity of these cell embeddings to characterize cell heterogeneity through k-means clustering.

Beyond imputing chromosomal interactions, an important aspect of understanding chromosome conformation is to predict base-pair-specific DNA methylation levels, which are crucial for epigenetic regulation analysis. To address this problem, Zhu *et al.* proposed scHiMe [84], which predicts base-pair specific methylation levels in individual cells using scHi-C data and DNA nucleotide sequences. scHiMe incorporates the metacell strategy and applies a GTN-based model on the promoter-promoter spatial interaction network to predict DNA methylation levels.

In summary, GNN-based approaches have proven effective in single-cell epigenomics for tasks such as cell type identification, methylome imputation, and chromosomal interaction analysis. The graph structures of each method are summarized in Supplementary Table S1. Future work should aim to optimize these models for scalability and make their results more biologically interpretable. Additionally, GNNs could be explored for other tasks in single-cell epigenomics research, such as trajectory inference and integrating various single-cell omics data types, further broadening their impact on understanding complex cellular mechanisms.

Graph neural networks for single-cell transcriptomics

Unachievable with bulk RNA sequencing techniques, scRNA-seq technologies provide granular insights into the transcriptome landscapes at the single-cell level across diverse species, samples, and tissues. scRNA-seq data has clearer semantic meaning, with each feature directly corresponding to the expression level of a gene. Compared to single-cell epigenomics data, scRNA-seq data is typically less sparse and has lower dimensionality, making it more manageable for computational models. The vast amount of single-cell transcriptomic data has spurred an explosive growth in computational tools designed to process and analyze these data (Table 2, Supplementary Text S11), owing to inherent challenges such as high noise levels, significant sparsity, and frequent dropouts [38].

Imputation

Generally, cells with identical functions often exhibit similar features, which inspired us to recover dropouts in gene expression profiles by leveraging similar cells. GNN-based methods for scRNA-seq data imputation can be categorized based on their different strategies for constructing cellular relationships and enhancing imputation through graph learning.

One common approach is to leverage cellular similarity metrics to construct a graph that represents relationships between cells, such as using k-nearest neighbors (kNN), Pearson correlation, or Spearman correlation to create similarity graphs [85–87, 130]. For example, GraphSCI [85] leverages the Pearson correlation coefficient to construct a gene co-expression graph from raw gene expression data. It then utilizes GCN to encode expression levels and reconstruct the co-expression network. Concurrently, an autoencoder is employed to sample the encoded embeddings, thus imputing the gene expression matrix.

Another approach of methods enhances imputation performance by incorporating dynamic weighting of cellular relationships. GNNImpute [88] incorporates a graph attention mechanism to dynamically assign different weights to similar cells on a kNN graph, which is constructed based on principal component analysis of the preprocessed gene expression matrix.

Dimensionality reduction

Dimensionality reduction is a critical task in single-cell transcriptomics data analysis, as scRNA-seq data is often high-dimensional and noisy, making downstream analyses challenging. GNN-based methods have been applied to dimensionality reduction, offering an effective way to learn low-dimensional representations of cells. To better organize the numerous GNN-based models for

Table 2. Details of GNN-based methods for single-cell transcriptomics.

Algorithm name	Model	Graph	Year	Task type	Reference
GraphSCI	GCNs	Gene–gene	2021	Imputation	[85]
GE-Impute	GNNs	Cell–cell	2022	Imputation	[86]
SCDD	GCNs	Cell–cell	2022	Imputation	[87]
GNNImpute	GATs	Cell–cell	2021	Imputation	[88]
scGCL	GCNs	Cell–cell	2023	Imputation	[89]
scGAEGAT	GAEs, GATs	Cell–cell	2022	Imputation, clustering	[90]
scGNN 2.0	GAEs, GATs	Cell–cell	2022	Imputation, clustering	[91]
scMPN	VGAEs	Cell–cell	2024	Imputation, clustering	[92]
scDGAE	GAEs, GCNs, GATs	Cell–cell	2023	Imputation, clustering	[93]
scGNN	GAEs, GCNs	Cell–cell	2021	Learn low-dimensional embedding	[63]
scGAE	GAEs, GATs	Cell–cell	2021	Learn low-dimensional embedding	[94]
GLAE	GAEs, GCNs	Cell–cell	2022	Learn low-dimensional embedding	[95]
scCDG	GAEs, GCNs	Cell–cell	2021	Learn low-dimensional embedding	[96]
scGCC	GATs	Cell–cell	2023	Learn low-dimensional embedding	[97]
SCDRHA	GAEs, GATs	Cell–cell	2021	Learn low-dimensional embedding	[98]
omicsGAT	GATs	Cell–cell	2022	Learn low-dimensional embedding	[99]
CellVGAE	VGAEs, GATs	Cell–cell	2021	Learn low-dimensional embedding	[100]
dhaSCA	GCNs	Cell–cell	2023	Learn low-dimensional embedding	[101]
scDGDC	GCNs	Cell–cell	2023	Learn low-dimensional embedding	[102]
AttentionAE-sc	GAEs, GCNs	Cell–cell	2023	Learn low-dimensional embedding	[103]
scDSC	GNNs	Cell–cell	2022	Learn low-dimensional embedding	[104]
SCEA	GAEs	Cell–cell	2023	Learn low-dimensional embedding	[105]
scTAG	GAEs	Cell–cell	2022	Learn low-dimensional embedding	[106]
scGCN	GCNs	Cell–cell	2021	Label transfer	[79]
SCMAG	GCNs	Cell–cell	2021	Cell type annotation	[107]
GraphCS	GNNs	Cell–cell	2022	Cell type annotation	[108]
graph-sc	GAEs, GCNs	Gene-cell	2021	Cell type annotation	[109]
HNNVAT	GCNs	Cell–cell	2023	Cell type annotation	[110]
sigGCN	GCNs	Gene–gene	2021	Cell type annotation	[111]
scBiGNN	GATs	Gene–gene and cell–cell	2023	Cell type annotation	[112]
CAME	GCNs	Cell–cell, cell-gene, and gene–gene	2022	Cell type annotation	[113]
scDeepSort	GraphSAGE	Cell-gene	2021	Cell type annotation	[114]
scMRA	GCNs	Prototype-prototype, prototype-cell, and cell–cell	2021	Cell type annotation	[115]
scAGN	GATs	Cell–cell	2023	Cell type annotation	[116]
GraphSCC	GCNs	Cell–cell	2021	Clustering	[117]
scGPCL	GNNs	Cell-gene	2023	Clustering	[118]
scMGCA	GAEs	Cell–cell	2023	Clustering	[119]
scGAC	GAEs, GATs	Cell–cell	2022	Clustering	[120]
scEGG	GAEs, GATs	Cell–cell	2024	Clustering	[121]
GDEC	GCNs	Gene–gene	2023	Clustering	[122]
GRACE	GAEs	Cell–cell	2023	Clustering	[123]
scASGC	GCNs	Cell–cell	2023	Clustering	[124]
scGMM-VGAE	VGAEs, GCNs	Cell–cell	2023	Clustering	[125]
CEFCON	GATs	Gene–gene	2023	Gene regulatory network inference	[126]
DeepRIG	GAEs, GCNs	Gene–gene	2023	Gene regulatory network inference	[127]
GNNLink	GCNs	Gene–gene	2023	Gene regulatory network inference	[128]
GenKI	VGAEs	Gene–gene	2023	KO-responsive gene identification	[129]
PathFinder	GTNs	Gene–gene	2024	Cell–cell communication inference	[47]
Graph-DiffVAE	VGAEs, GCNs	Cell–cell	2020	Cell–cell communication inference	[131]
GraphComm	GATs	Protein–protein	2023	Cell–cell communication inference	[132]
Ravindra_2020	GATs	Cell–cell	2020	Disease state prediction	[60]
Sehanobish_2021	GATs	Cell–cell	2021	Disease state prediction	[192]
GEARS	GNNs	Gene–gene	2023	Single-cell perturbation prediction	[193]
Cellograph	GCNs	Cell–cell	2024	Single-cell perturbation prediction	[194]

dimensionality reduction, we classify them based on their core architecture and methodology.

GAE-based methods employ autoencoders with graph structures to learn low-dimensional embeddings that capture the relationships between cells. These methods typically start by constructing a cell graph, followed by utilizing an autoencoder framework to derive meaningful representations [63, 94–96]. scGNN [63] is a representative GAE-based model that employs four stacked autoencoders for scRNA-seq analysis. It begins by learning feature embeddings, constructs a cell–cell kNN graph, prunes it, and applies a GAE to derive low-dimensional representations. These representations are then used for clustering,

followed by iterative gene expression reconstruction until clustering stabilizes. Finally, an imputation autoencoder reconstructs the gene expression matrix.

Cell type identification

Identifying cell types from scRNA-seq data is a crucial application of GNNs, and a diverse set of methods has been developed, each taking a different approach to leveraging graph structures and neural network architectures. These methods can be organized into distinct categories based on how they define cell relationships, the types of data they incorporate, and their underlying learning strategies.

The success of GNN-based cell type identification relies heavily on how the graph is constructed to represent relationships within scRNA-seq data. Different approaches have been developed based on cell similarity, gene–gene interactions, or combining heterogeneous data. The most common strategy for GNN-based cell type identification is constructing a cell similarity graph, where nodes represent cells and edges capture relationships based on similarity metrics [107, 108, 117]. For instance, GraphSCC [117] leverages a denoising autoencoder along with a GCN applied to a kNN-based cell graph, integrating dual self-supervised learning to enhance the quality of cell clustering.

Cell–cell communication inference

Cell-to-cell communication is crucial for understanding interactions in multi-cellular systems, coordination of biological processes, and changes in different conditions like development or disease. Below, we introduce three GNN-based methods that have been developed to infer cell–cell interactions.

Graph-DiffVAE [131] begins by constructing an undirected and unweighted cell–cell graph represented by a binary adjacency matrix based on Pearson correlation coefficients between cells. Subsequently, this adjacency matrix, combined with cell features is fed into a VGAE-based model to predict relationships between cells of various types, aiding in the understanding of cell differentiation. Moreover, GraphComm [132] combines an annotated ligand-receptor database and utilizes a GAT to predict cell–cell communication interactions for scRNA-seq data. PathFinder [47] employs a divide-and-conquer approach to tackle this complexity. By partitioning multi-cellular intra- and inter-cellular signaling networks into smaller, more manageable signaling paths, PathFinder uses a GTN-based framework to deduce the interactions both within and between cells.

Graph neural networks for spatially resolved transcriptomics

Previously, array-based ST technologies could only measure gene expression at the spot level, which allowed for the estimation of cell-type compositions in each spot but could not distinguish individual cell types within a spot. Recent technology advances in SRT have facilitated the capture of cellular spatial relative localization histological imaging, and gene expression profiles at single-cell and even sub-cellular resolutions, providing deeper insights into complex disease mechanisms [67, 133, 134]. This modality provides gene expression profiles along with the spatial coordinates of cells, which requires GNN-based models to process two types of data simultaneously and learn the correspondence between spatial structure and gene expression. Consequently, the graph structure must encode not only gene expression similarity

but also spatial neighborhood information [135]. In the following sections, we will explore how GNN-based models can be applied to SRT data to reveal intricate spatial interactions and enhance biological insights (Table 3, Supplementary Text S12).

Spatial gene expression prediction

Spatial gene expression prediction aims to estimate gene expression levels from spatial information, such as histological images, enabling researchers to bridge the gap between visual tissue morphology and molecular data. This task is particularly valuable for enhancing our understanding of the spatial organization.

Hist2ST [136] is a method designed to predict gene expression directly from histology images. Hist2ST begins by segmenting the histology image into patches and extracting 2D vision features from these patches. Subsequently, Hist2ST learns spot features with global spatial dependencies through a transformer module and constructs a spot kNN graph based on Euclidean distance, which is then fed into a GraphSAGE-based GNN to incorporate local spatial dependencies. Additionally, Hist2ST utilizes a ZINB model to capture the features learned and output the predicted gene-by-spot expression matrix.

Dimensionality reduction

In SRT analysis, dimensionality reduction remains a crucial step. By projecting the data into a lower-dimensional space, researchers can better capture spatial heterogeneity and facilitate downstream analyses that are specific to spatial data, such as identifying spatial domains, inferring spatial patterns, and correcting batch effects. Below, we present several GNN-based methods that employ dimensionality reduction for ST, highlighting their different strategies and advantages.

One common approach involves constructing cell or spot graphs based solely on spatial coordinates [138, 139]. spage2vec [138] employs GraphSAGE to encode low-dimensional representations from a kNN graph constructed using spatial information. The learned representations characterize spatial transcriptomic heterogeneity and facilitate spatial domain identification through the Leiden clustering algorithm.

Spatial domain identification

Spatial domain identification in SRT aims to delineate spatially coherent domains within tissues based on gene expression, spatial localization, and other available data modalities [160–162]. By identifying these spatial domains, researchers can better understand how cellular environments influence gene expression and how different tissue regions contribute to overall function or pathology. Below, we introduce several GNN-based methods that have been developed for spatial domain identification, categorized by their graph construction techniques, learning paradigms, and data integration strategies [133, 145–152].

The most basic approach for identifying spatial domains involves constructing graphs that represent spatial dependencies among spots or cells. SpaGCN [133] constructs an undirected weighted graph that represents spatial dependencies from SRT data. Each node represents a spot or cell and the edge weights between two nodes are determined by their level of relatedness. SpaGCN aggregates spatial localization, histology information, and gene expression via a GCN, employs iterative clustering to identify spatial domains, and detects spatially variable genes enriched in each domain.

Table 3. Details of GNN-based methods for SRT.

Algorithm name	Model	Graph	Year	Task type	Reference
Hist2ST	GraphSAGE	Spot-spot	2022	Spatial gene expression prediction	[136]
Cells2RNA	GATs	Cell–cell	2023	Spatial gene expression prediction	[137]
spage2vec	GraphSAGE	Spot-spot	2020	Learn low-dimensional embedding	[138]
Salas_2022	GraphSAGE	Spot-spot	2022	Learn low-dimensional embedding	[139]
conST	VGAEs	Cell–cell or spot-spot	2022	Learn low-dimensional embedding	[140]
PAST	VGAEs	Spot-spot	2023	Learn low-dimensional embedding	[141]
SEDR	VGAEs, GCNs	Cell–cell or spot-spot	2024	Learn low-dimensional embedding	[142]
CCST	GCNs	Cell–cell or spot-spot	2022	Clustering	[143]
SCGDL	GCNs	Spot-spot	2023	Clustering	[144]
STAGATE	GAEs, GATs	Cell–cell or spot-spot	2022	Spatial domain identification	[145]
stMVC	GATs	Cell–cell or spot-spot	2022	Spatial domain identification	[146]
GraphST	GCNs	Spot-spot	2023	Spatial domain identification	[147]
SpaGCN	GCNs	Cell–cell or spot-spot	2021	Spatial domain identification	[133]
SpaceFlow	GCNs	Cell–cell or spot-spot	2022	Spatial domain identification	[148]
SCAN-IT	GCNs	Cell–cell or spot-spot	2022	Spatial domain identification	[149]
DeepST	VGAEs	Cell–cell or spot-spot	2022	Spatial domain identification	[150]
SiGra	GTNs	Cell–cell or spot-spot	2023	Spatial domain identification	[151]
SPACEL	GCNs	Cell–cell or spot-spot	2023	Spatial domain identification	[152]
STGNNks	GCNs	Spot-spot	2023	Spatial domain identification	[153]
ConSpaS	GAEs	Spot-spot	2023	Spatial domain identification	[154]
stAA	VGAEs	Cell–cell or spot-spot	2024	Spatial domain identification	[155]
GCNG	GCNs	Cell–cell	2020	Ligand-receptor interaction inference	[156]
spaCI	GATs	Cell–cell	2022	Ligand-receptor interaction inference	[157]
DeepLinc	VGAEs, GCNs	Cell–cell	2022	Cell–cell communication inference	[41]
HoldNet	GATs, GCNs	Cell–cell	2023	Cell–cell communication inference	[158]
NCEM	GCNs, GATs	Cell–cell	2022	Cell–cell communication inference	[159]

Cell–cell communication inference

Cell–cell communications can be categorized into intracellular and intercellular interactions [41, 156–158]. GCNG [156], a supervised GCN-based approach, is specifically designed for delineating intercellular gene interactions, particularly ligand-receptor communication, from single-cell SRT data. Taking the spatial cell location and gene expression as input, GCNG also constructs a cell–cell graph by computing the Euclidean distance between cells utilizing their spatial coordinates. Instead of using kNN, GCNG selects neighboring cells according to a predefined threshold of distance to construct the cell neighborhood graph.

Graph neural networks for single-cell proteomics

Proteins are the direct executors of cellular activities. Despite insights from scRNA-seq data about gene expression, the correlation between mRNA and protein levels often lacks consistency [163], which may lead to inaccuracies when inferring protein abundance solely based on scRNA-seq profiles. Recent advancements in single-cell proteomics based on mass spectrometry and next-generation sequencing methods have significantly enhanced our understanding of cellular heterogeneity, cell function, and disease mechanisms [163–166], yielding thrilling insights. Proteomic data typically has clear semantic meaning, as the features represent actual protein quantities. However, due to current technological constraints, single-cell proteomics data often has lower dimensionality, with only dozens to hundreds of protein features measurable per cell, requiring GNNs must be adapted to handle these issues.

scPROTEIN [167] treats each cell in the single-cell proteomic data abundance matrix as a node, calculates the Pearson correlation coefficient between abundance features to establish edges, and thus constructs an undirected and unweighted graph. scPROTEIN learns cell embeddings from cell-by-protein matrix based on GCN [54], and has demonstrated the successful application of these embeddings in various downstream analyses including cell clustering, classification, and batch correction.

SNOWFLAKE [168] integrates single-cell proteomics, morphological, and structural data, modeling single-cell neighboring information from imaging data as single-cell spatial graphs. Based on a GNN-based framework, it forecasts immune reactions and assesses the distinctiveness of tissue microenvironments. The graph structures for each method are presented in Supplementary Table S2.

Proteins do not function in isolation; instead, they interact within complex networks, engaging in signaling pathways, forming protein complexes, and affecting cellular activities in tandem. Therefore, these protein–protein interactions are naturally modeled as networks, providing an ideal setting for leveraging GNN-based frameworks to capture such intricate relationships. Future applications of GNN-based models could include the development of integrative frameworks to combine proteomic data with other omics, enhancing our understanding of protein-level regulation, or employing GNNs to predict protein function and interaction in different cell states, which will be introduced in the next section. The potential to advance single-cell proteomic research using GNNs remains vast, warranting further exploration and innovation.

Graph neural networks for single-cell multi-omics

Single-cell multi-omics methods allow for the measurement of different molecular aspects from individual cells, including the genome, epigenome, transcriptome, proteome, or metabolome. While this modality combines information from multiple data

Table 4. Details of GNN-based methods for single-cell multi-omics.

Algorithm name	Model	Graph	Year	Task type	Reference
GLUE	VGAEs	Feature-feature	2022	Integration	[169]
GCN-SC	GCNs	Cell–cell	2023	Integration	[170]
STitch3D	GATs	Spot-spot	2023	Multiple SRT slices integration	[171]
scMoGNN	GCNs	Cell-feature	2022	Modality prediction, modality matching	[172]
STACI	VGAEs	Cell–cell	2022	Learn joint low-dimensional embedding	[173]
scHGA	GNNs	Cell-gene-loci	2023	Learn joint low-dimensional embedding	[174]
MarsGT	GTNs	Cell-gene and cell-peak	2024	Rare cell population identification	[175]
DeepMAPs	GAEs, GTNs	Cell-gene	2023	Integration and cell-type-specific biological network inference	[64]
GrID-Net	GNNs	Cell–cell	2022	Granger causal inference	[176]
scapGNN	VGAEs	Cell–cell or gene–gene	2023	Active pathway inference and gene module identification	[177]
CIPHER-SC	GCNs	Disease-gene	2020	Disease-gene association inference	[178]
GENELink	GATs	Gene–gene	2022	Gene regulatory network inference	[179]
CLARIFY	GAEs	Cell–cell and gene–gene	2023	Cell–cell interaction inference	[180]
SD^2	GCNs	Spot-spot	2022	ST deconvolution	[181]
DSTG	GCNs	Spot-spot	2021	Cell type deconvolution	[182]
STdGCN	GCNs	Spot-spot	2023	Cell type deconvolution	[183]
SPACEL	GCNs	Cell–cell or spot-spot	2023	Spatial domain identification	[152]
PIKE-R2P	GNNs	Protein–protein	2021	Protein abundance prediction	[184]
scMoFormer	GTNs	Cell-gene, gene-protein, gene–gene, and protein–protein	2023	Protein abundance prediction	[185]
scFEA	GNNs	Metabolite-metabolic module	2021	Cell-wise fluxome prediction	[186]

types, it also introduces complexities due to the heterogeneity of data types. Each omics layer has its own feature space and scale, making integration challenging. Therefore, GNNs applied to multi-omics data must be designed to integrate heterogeneous data types into a unified representation. The integrated analysis of these layers has revealed complex interactions across multi-omics, providing invaluable insights into cellular biology [66, 75] (Table 4, Supplementary Text S13).

Multi-omics integration

Single-cell multi-omics integration methods align cell states across various omics layers to provide a unified view of cellular heterogeneity. Recent advancements in GNN-based models for multi-omics integration focus on learning unified cell representations by constructing heterogeneous or multi-relational graphs that combine diverse single-cell data types, such as transcriptomics, epigenomics, and spatial data.

A common approach for multi-omics integration is aligning different omics modalities through advanced embedding strategies [64, 169, 172]. For example, GLUE [169] utilizes a VGAE-based model to integrate single-cell transcriptomics, and single-cell epigenomics (scCAS and scDNAm) data. GLUE initially applies omics-specific VAEs to yield low-dimensional cell representations for each of the three omics. To ensure that the learned cell embeddings from different omics layers have uniform semantics, GLUE incorporates a guidance graph based on prior knowledge of regulatory interactions, with nodes representing features from each omics layer and edges representing the sign and credibility of the regulatory interaction. Finally, GLUE employs an omics discriminator to align the cell embeddings across different omics, thus integrating the data from all three omics.

Rare cell population characterization

Rare cell populations play a critical role in complex disease processes, such as cancer progression and immune response. The ability to accurately characterize these rare populations provides valuable insights into disease mechanisms and potential therapeutic targets. MarsGT [175] is a notable method that addresses the challenge of rare cell population characterization by integrating paired scRNA-seq and scATAC-seq data into a heterogeneous graph structure. In this graph, nodes represent cells, genes, and peaks, while edges capture the interactions between cells and genes or cells and peaks. By modeling these relationships, MarsGT uses a probability-based heterogeneous graph transformer framework to discern cell clusters, focusing especially on rare cell populations that may be underrepresented in typical clustering approaches. MarsGT further predicts cluster-specific peak-gene interactions and infers enhancer gene regulatory networks (eGRNs), which play a vital role in understanding gene regulation within rare cell populations. This comprehensive integration of multiple data types and advanced modeling of interactions allows MarsGT to effectively identify these elusive yet biologically significant cell clusters.

Biological network inference

Biological networks are key to unraveling cellular heterogeneity and interpreting cellular mechanisms. Below, we discuss several GNN-based approaches for biological network inference, emphasizing their unique contributions and methods for mitigating limitations in previous techniques.

Han *et al.* proposed scapGNN [177] to infer gene-cell, gene–gene, and cell–cell association networks. scapGNN aims to infer gene-cell, gene–gene, and cell–cell association networks by leveraging

both scRNA-seq and scCAS data. It starts by converting single-cell data into pathway activity score matrices to characterize cellular heterogeneity, allowing for a detailed examination of cellular states. Starting from gene expression of scRNA-seq data or gene activity matrices from scCAS data, scapGNN learns low-dimensional embeddings of cells and genes using a deep neural network autoencoder. It constructs gene correlation and cell correlation networks separately from the gene-cell matrix using Pearson correlation coefficients, then learns gene–gene and cell–cell association networks with VGAEs, incorporating gene and cell embeddings separately. Using the random walk with restart algorithm, it infers a weighted gene-cell association network and generates a pathway-by-cell activity matrix, which aids in clustering and identifying gene modules that are crucial for depicting specific cell phenotypes. Finally, scapGNN integrates inferred networks from both two omics using Brown's method [187].

Cell type deconvolution

Cell type deconvolution is an essential task in ST, aiming to estimate the proportions of different cell types present within each spatial spot. Below, we describe several GNN-based methods that utilize multi-omics data to tackle this task effectively.

DSTG [182] utilizes a GNN to consider the previously overlooked topological information between spots to deconvolute SRT data. It constructs pseudo-ST data using annotated scRNA-seq data, subsequently applying canonical correlation analysis to conduct dimensionality reduction on both pseudo and real SRT data. DSTG then connects pseudo and real SRT data into a spot neighborhood graph, linking spots considered as mutual k-nearest neighbors. It inputs the link graph along with variable gene expression matrices from both pseudo and real SRT data into a GCN to predict cell type proportions at each spot through semi-supervised learning by calculating cross-entropy between the actual cell proportions and the predicted compositions of the pseudo-ST data.

In summary, GNN-based methods have emerged as powerful tools for analyzing single-cell omics data, significantly advancing our understanding of genetic regulation and expression. As shown in Fig. 4, GNN-based frameworks, such as GCNs, GATs, and GAEs, have been widely utilized in the analysis of single-cell omics data. Additionally, a variety of GNN-based models have been successfully developed to tackle a broad spectrum of tasks in single-cell omics, including cell type identification, dimensionality reduction, and biological network inference. These advancements hold great promise for unlocking deeper insights into complex biological systems, supporting the development of more precise diagnostic and therapeutic strategies. We have summarized the principal framework for applying GNNs to single-cell omics in Supplementary Text S14 for insightful guidance. We also collected a total of 77 GNN-based methods commonly used publicly available datasets in Supplementary Tables S3–S7 to provide readers with practical resources for their own studies according to the type of omics. We explain the collection of these data in Supplementary Text S15.

Discussion: challenges and opportunities

Throughout over a hundred computational methods for various single-cell omics data reviewed in this survey, many of them model the single-cell resolution data as cell–cell or gene–gene graphs to characterize the data via graph structures, which are then fed into subsequent GNN-based networks. When constructing these graphs, it is widely adopted to calculate the Euclidean distance or assess other similarity metrics between cells and then establish the cell neighborhood graphs via kNN with varying values of k, thresholds of similarity metrics, or other criteria to select neighbors. This provides a solid reference and opens potential avenues for further refinement.

Incorporating external information into GNN-based models has become an essential strategy to enhance biological relevance and interpretability. For example, incorporating protein–protein interaction networks, gene regulatory networks, and prior biological knowledge of ligand-receptor interactions provides valuable context, enabling more accurate modeling of complex biological systems. These external sources help GNNs capture functional relationships that might not be evident from the data alone, such as regulatory interactions or intercellular communication pathways.

GNNs, by their nature, have high computational demands, especially as the number of nodes and edges increases, which presents a significant challenge in terms of scalability and efficiency. For example, for single-cell transcriptomics, where each cell or gene is treated as a node, the sheer scale of data often leads to excessive memory requirements and prolonged training times. This makes the selection of input features a challenging and critical step, as processing the full transcriptome could quickly become impractical for most models. To mitigate these challenges, feature selection techniques are often employed to reduce the dimensionality of the data. Selecting only highly variable genes or highly expressed genes can significantly cut down the model size, making GNNs more tractable while retaining essential biological information. However, this approach also presents a trade-off: reducing the data to manage computational feasibility may result in the loss of information critical for capturing nuanced biological behaviors [4, 188]. The inherent scalability limitations of GNNs necessitate careful optimization of these trade-offs to ensure that computational feasibility is maintained without compromising model performance or biological insight.

Another challenge of GNNs is their tendency to aggregate node information based primarily on local neighborhoods, making it challenging to capture global structural information, which is essential for tasks requiring long-range dependencies, such as inferring gene regulatory networks or understanding cellular differentiation. A common solution to this issue is increasing the depth of the network, allowing for the aggregation of information over larger node neighborhoods. However, this solution introduces more challenges. On the one hand, as the network depth increases, the computational complexity grows exponentially, making deep GNNs difficult to train on large-scale single-cell data. On the other hand, increasing the number of layers exacerbates the issue of over-smoothing, where node representations become overly similar, leading to a loss of discriminative ability. In single-cell transcriptomics, for instance, over-smoothing could result in different cell types or states being indistinguishable, thereby hindering the model's ability to accurately annotate cells or predict their responses. Therefore, capturing global structural information and avoiding over-smoothing are in tension, making this another trade-off that requires careful optimization to ensure the model can effectively capture global dependencies without compromising the ability to distinguish between important biological features.

With the emergence of foundation models [189–191] pretrained on large single-cell transcriptomics datasets, there is a growing opportunity for GNN-based models to leverage these models' generalization capabilities. These pre-trained GNNs, having learned from complex biological data, capture patterns
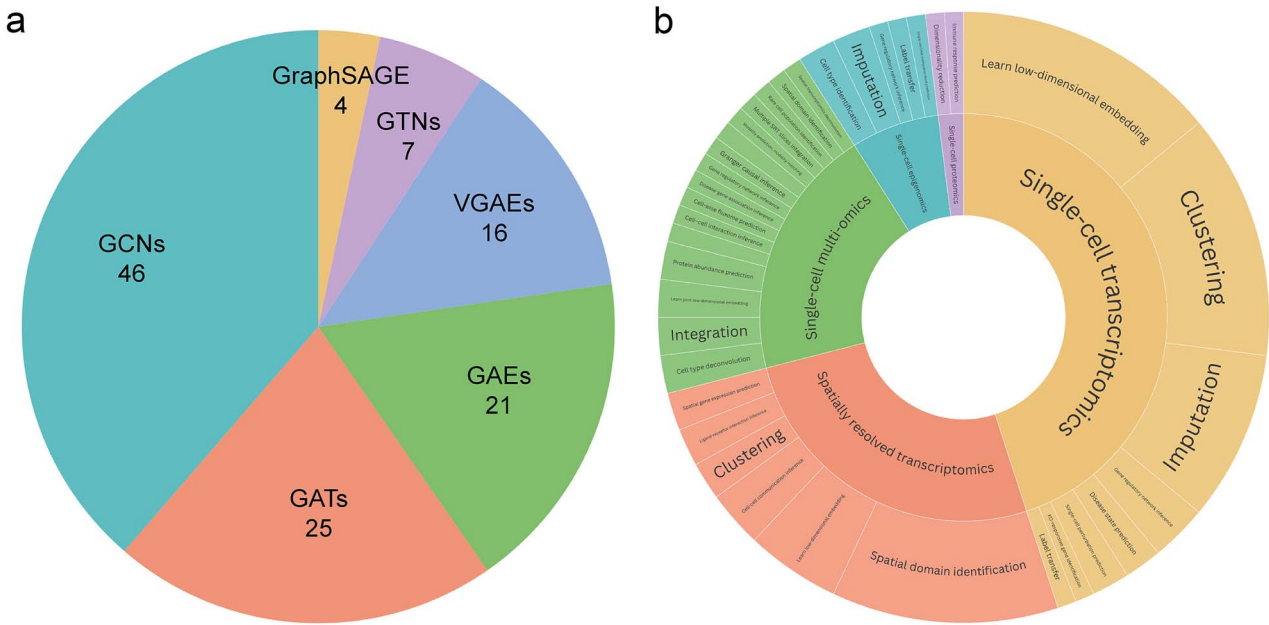
Figure 4. Distribution and growth of GNN-based models. (a) The pie chart shows the proportion of six different GNN variants in the review. (b) The sunburst chart illustrates the successful application of GNNbased models across various tasks in multiple single-cell omics data analysis.

in gene expression and regulatory mechanisms across diverse conditions, offering a strong starting point for downstream tasks like cell type classification or gene network inference. Fine-tuning these pre-trained GNN models on smaller, task-specific datasets can significantly boost performance, especially in scenarios with limited labeled data.

Furthermore, with advancements in various technologies, such as high-resolution imaging and single-cell proteomics, data with subcellular-level resolution has become increasingly available. This creates a significant opportunity for GNN-based frameworks to manage and analyze such detailed datasets. Moreover, the integration of subcellular data with other types of omics data, such as transcriptomics and epigenomics, could enable GNNs to develop more comprehensive models of cellular behavior. This approach could lead to breakthroughs in understanding complex diseases and developing personalized medicine strategies, paving the way for more innovative and accurate single-cell data analyses.

Conclusion

Our survey begins by introducing the basic structure of GNNs and their six primary variants, followed by a detailed exploration of how GNN-based models are applied across different single-cell omics, highlighting specific implementation methods and approaches. Our review has documented 107 successful applications of GNN-based frameworks in a wide array of tasks involving single-cell epigenomics, transcriptomics, ST, proteomics, and multi-omics. These methods have demonstrated significant utility in understanding cellular heterogeneity, spatial domain identification, biological network inference, cell–cell communication, Granger causal inference, cell type deconvolution, multiple SRT slices integration, and the integration of multi-omics data. Besides, in each omics section, we also comprehensively summarize the publicly available single-cell data of the corresponding omics that are widely used by the GNN-based approaches reviewed in that section. We anticipate that the extensive collection of datasets we have compiled will provide readers with valuable resources for their own research.

Additionally, we have highlighted several key considerations in graph construction, feature selection, integration of external information, and the use of foundation models. These findings suggest new potential directions for future experimentation, including enhancing biological interpretability and improving computational scalability. We anticipate that this survey will serve as a resource for researchers aiming to deepen their understanding of deep learning applications in single-cell analysis, providing both practical guidelines for adapting GNNs to single-cell omics and valuable insights for the broader development of computational methods in the field.

Key Points

- This review introduces six key variants of Graph neural networks (GNNs)—GCNs, GraphSAGE, GATs, GTNs, GAEs, and VGAEs—and their roles in single-cell omics analysis.
- This study systematically reviews 107 applications and commonly used publicly available benchmark datasets of GNN-based models in single-cell epigenomics, transcriptomics, spatial transcriptomics, proteomics, and multi-omics for various computational tasks.
- We discuss various graph construction strategies including cell–cell, gene–gene, and heterogeneous graphs, tailored to address the complexity of biological relationships.
- We highlight the successful application of GNN-based frameworks and outline potential avenues for future research in applying GNNs to single-cell omics.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflicts of interest: None declared.

Funding

Data availability

No data was used in this study.

References

1. Nagalakshmi U, Wang Z, Waern K. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008;**320**:1344–9. https://doi.org/10.1126/science.1158441.

2. Mortazavi A, Williams BA, McCue K. *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;**5**:621–8. https://doi.org/10.1038/nmeth.1226.

3. Buenrostro JD, Giresi PG, Zaba LC. *et al.* Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;**10**:1213–8. https://doi.org/10.1038/nmeth.2688.

4. Buenrostro JD, Wu B, Chang HY. *et al.* ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 2015;**109**:21.29. 21–9. https://doi.org/10.1002/0471142727.mb2129s109.

5. Frommer M, McDonald LE, Millar DS. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 1992;**89**:1827–31. https://doi.org/10.1073/pnas.89.5.1827.

6. Susan JC, Harrison J, Paul CL. *et al.* High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 1994;**22**:2990–7. https://doi.org/10.1093/nar/22.15.2990.

7. Stark R, Grzelak M, Hadfield J. RNA sequencing: The teenage years. *Nat Rev Genet* 2019;**20**:631–56. https://doi.org/10.1038/s41576-019-0150-2.

8. Corces MR, Buenrostro JD, Wu B. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* 2016;**48**:1193–203. https://doi.org/10.1038/ng.3646.

9. Li X, Wang C-Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci* 2021;**13**:36. https://doi.org/10.1038/s41368-021-00146-0.

10. Park Y, Wu H. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 2016;**32**:1446–53. https://doi.org/10.1093/bioinformatics/btw026.

11. Li Y, Ma L, Wu D. *et al.* Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. *Brief Bioinform* 2021;**22**. https://doi.org/10.1093/bib/bbab024.

12. Picelli S, Faridani OR, Björklund ÅK. *et al.* Full-length RNA-seq from single cells using smart-seq2. *Nat Protoc* 2014;**9**:171–81. https://doi.org/10.1038/nprot.2014.006.

13. Chen KH, Boettiger AN, Moffitt JR. *et al.* Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;**348**:aaa6090. https://doi.org/10.1126/science.aaa6090.

14. Wang X, Allen WE, Wright MA. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;**361**:eaat5691. https://doi.org/10.1126/science.aat5691.

15. Yuan Z, Pan W, Zhao X. *et al.* SODB facilitates comprehensive exploration of spatial omics data. *Nat Methods* 2023;**20**:387–99. https://doi.org/10.1038/s41592-023-01773-7.

16. Satpathy AT, Granja JM, Yost KE. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* 2019;**37**:925–36. https://doi.org/10.1038/s41587-019-0206-z.

17. Domcke S, Hill AJ, Daza RM. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* 2020;**370**:eaba7612. https://doi.org/10.1126/science.aba7612.

18. Zhang K, Hocker JD, Miller M. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* 2021;**184**:5985–6001. e5919. https://doi.org/10.1016/j.cell.2021.10.024.

19. Gao Z, Jiang R, Chen S. OpenAnnotateApi: Python and R packages to efficiently annotate and analyze chromatin accessibility of genomic regions. *Bioinform adv* 2024;**4**. https://doi.org/10.1093/bioadv/vbae055.

20. Gao Z, Chen X, Li Z. *et al.* scEpiTools: A database to comprehensively interrogate analytic tools for single-cell epigenomic data. *J Genet Genomics* 2024;**51**:462–5. https://doi.org/10.1016/j.jgg.2023.09.011.

21. Clark SJ, Smallwood SA, Lee HJ. *et al.* Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc* 2017;**12**:534–47. https://doi.org/10.1038/nprot.2016.187.

22. Nagano T, Lubling Y, Stevens TJ. *et al.* Single-cell hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013;**502**:59–64. https://doi.org/10.1038/nature12593.

23. Dou M, Clair G, Tsai C-F. *et al.* High-throughput single cell proteomics enabled by multiplex isobaric labeling in a nanodroplet sample preparation platform. *Anal Chem* 2019;**91**:13119–27. https://doi.org/10.1021/acs.analchem.9b03349.

24. Woo J, Williams SM, Markillie LM. *et al.* High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nat Commun* 2021;**12**:6246. https://doi.org/10.1038/s41467-021-26514-2.

25. Stoeckius M, Hafemeister C, Stephenson W. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017;**14**:865–8. https://doi.org/10.1038/nmeth.4380.

26. Cao J, Cusanovich DA, Ramani V. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 2018;**361**:1380–5. https://doi.org/10.1126/science.aau0730.

27. Ma S, Zhang B, LaFave LM. *et al.* Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* 2020;**183**:1103–1116. e1120. https://doi.org/10.1016/j.cell.2020.09.056.

28. Plongthongkum N, Diep D, Chen S. *et al.* Scalable dual-omics profiling with single-nucleus chromatin accessibility and mRNA expression sequencing 2 (SNARE-seq2). *Nat Protoc* 2021;**16**:4992–5029. https://doi.org/10.1038/s41596-021-00507-3.

29. Fiskin E, Lareau CA, Ludwig LS. *et al.* Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. *Nat Biotechnol* 2022;**40**:374–81. https://doi.org/10.1038/s41587-021-01065-5.

30. Chen AF, Parks B, Kathiria AS. *et al.* NEAT-seq: Simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat Methods* 2022;**19**:547–53. https://doi.org/10.1038/s41592-022-01461-y.

31. Angermueller C, Clark SJ, Lee HJ. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* 2016;**13**:229–32. https://doi.org/10.1038/nmeth.3728.

32. Clark SJ, Argelaguet R, Kapourani C-A. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and

transcription in single cells. *Nat Commun* 2018;**9**:781. https://doi.org/10.1038/s41467-018-03149-4.

33. Yuan G-C, Cai L, Elowitz M. *et al.* Challenges and emerging directions in single-cell analysis. *Genome Biol* 2017;**18**:1–8.

34. Sun F, Li H, Sun D. *et al.* Single-cell omics: Experimental workflow, data analyses and applications. *Sci China Life Sci* 2025;**68**: 5–102. https://doi.org/10.1007/s11427-023-2561-0.

35. Chen X, Chen S, Song S. *et al.* Cell type annotation of single-cell chromatin accessibility data via supervised Bayesian embedding. *Nat Mach Intell* 2022;**4**:116–26. https://doi.org/10.1038/s42256-021-00432-w.

36. Tang S, Cui X, Wang R. *et al.* scCASE: Accurate and interpretable enhancement for single-cell chromatin accessibility sequencing data. *Nat Commun* 2024;**15**:1629. https://doi.org/10.1038/s41467-024-46045-w.

37. Li S, Li Y, Sun Y. *et al.* EpiCarousel: Memory-and time-efficient identification of metacells for atlas-level single-cell chromatin accessibility data. *Bioinformatics* 2024;**40**:btae191. https://doi.org/10.1093/bioinformatics/btae191.

38. Hua H, Long W, Pan Y. *et al.* scCrab: A reference-guided cancer cell identification method based on Bayesian neural networks. Interdiscip Sci Comput. *Life Sci* 2024;**17**:12–26. https://doi.org/10.1007/s12539-024-00655-6.

39. Chen S, Wang R, Long W. *et al.* ASTER: Accurately estimating the number of cell types in single-cell chromatin accessibility data. *Bioinformatics* 2023;**39**:btac842. https://doi.org/10.1093/bioinformatics/btac842.

40. Jin S, MacLean AL, Peng T. *et al.* scEpath: Energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 2018;**34**: 2077–86. https://doi.org/10.1093/bioinformatics/bty058.

41. Li R, Yang X. De novo reconstruction of cell interaction landscapes from single-cell spatial transcriptome data with DeepLinc. *Genome Biol* 2022;**23**:124. https://doi.org/10.1186/s13059-022-02692-0.

42. Wu AP, Singh R, Berger B. Granger causal inference on DAGs identifies genomic loci regulating transcription. In: *International Conference on Learning Representations*, 2022.

43. Suva ML, Tirosh I. Single-cell RNA sequencing in cancer: Lessons learned and emerging challenges. *Mol Cell* 2019;**75**: 7–12. https://doi.org/10.1016/j.molcel.2019.05.003.

44. Lim B, Lin Y, Navin N. Advancing cancer research and medicine with single-cell genomics. *Cancer Cell* 2020;**37**:456–70. https://doi.org/10.1016/j.ccell.2020.03.008.

45. Wang H, Shin D-G. CGCom: A framework for inferring cell-cell communication based on graph neural network. *bioRxiv* 2023;**2023**:2011. 2010.566642.

46. Vasighizaker A, Hora S, Zeng R. *et al.* SEGCECO: Subgraph embedding of gene expression matrix for prediction of CEll-cell COmmunication. *Brief Bioinform* 2024;**25**:bbae160. https://doi.org/10.1093/bib/bbae160.

47. Feng J, Song H, Province M. *et al.* PathFinder: A novel graph transformer model to infer multi-cell intra-and inter-cellular signaling pathways and communications. *Front Cell Neurosci* 2024;**18**:1369242. https://doi.org/10.3389/fncel.2024.1369242.

48. Aibar S, González-Blas CB, Moerman T. *et al.* SCENIC: Single-cell regulatory network inference and clustering. *Nat Methods* 2017;**14**:1083–6. https://doi.org/10.1038/nmeth.4463.

49. L'Heureux A, Grolinger K, Elyamany HF. *et al.* Machine learning with big data: Challenges and approaches. *IEEE Access* 2017;**5**: 7776–97. https://doi.org/10.1109/ACCESS.2017.2696365.

50. Bronstein MM, Bruna J, LeCun Y. *et al.* Geometric deep learning: Going beyond euclidean data. *IEEE Signal Process Mag* 2017;**34**: 18–42. https://doi.org/10.1109/MSP.2017.2693418.

51. Wu Z, Pan S, Chen F. *et al.* A comprehensive survey on graph neural networks. *IEEE Trans Neural Networks Learn Syst* 2020;**32**: 4–24. https://doi.org/10.1109/TNNLS.2020.2978386.

52. Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains. In: *Proceedings. 2005 IEEE international joint conference on neural networks, 2005,* pp. 729–34. New York City, US: IEEE, 2005.

53. Scarselli F, Gori M, Tsoi AC. *et al.* The graph neural network model. *IEEE Trans Neural Netw* 2009;**20**:61–80. https://doi.org/10.1109/TNN.2008.2005605.

54. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations*, 2017.

55. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: Guyon I, Luxburg UV, Bengio S. *et al.*, (eds.), *Advances in Neural Information Processing Systems*. 57 Morehouse Lane, Red Hook, NY 12571, USA: Curran Associates, Inc., 2017.

56. Velickovic P, Cucurull G, Casanova A. *et al.* Graph attention networks. *Stat* 2017;**1050**:10–48550.

57. Yun S, Jeong M, Kim R. *et al.* Graph transformer networks. In: Wallach H, Larochelle H, Beygelzimer A. *et al.*, (eds.), *Advances in Neural Information Processing Systems*. 57 Morehouse Lane, Red Hook, NY 12571, USA: Curran Associates, Inc., 2019.

58. Kipf TN, Welling M. Variational graph auto-encoders. *arXiv preprint arXiv* 1611.07308 2016.

59. Zhang X-M, Liang L, Liu L. *et al.* Graph neural networks and their current applications in bioinformatics. *Front Genet* 2021;**12**:690049. https://doi.org/10.3389/fgene.2021.690049.

60. Ravindra N, Sehanobish A, Pappalardo JL. *et al.* Disease state prediction from single-cell data using graph attention networks. In: *ACM Conference on Health, Inference, and Learning*, pp. 121–30. New York, NY, USA: Association for Computing Machinery, 2020.

61. Chereda H, Bleckmann A, Menck K. *et al.* Explaining decisions of graph convolutional neural networks: Patient-specific molecular subnetworks responsible for metastasis prediction in breast cancer. *Genome Med* 2021;**13**:1–16. https://doi.org/10.1186/s13073-021-00845-7.

62. Zhang Z, Chen L, Zhong F. *et al.* Graph neural network approaches for drug-target interactions. *Curr Opin Struct Biol* 2022;**73**:102327. https://doi.org/10.1016/j.sbi.2021.102327.

63. Wang J, Ma A, Chang Y. *et al.* scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat Commun* 2021;**12**:1882. https://doi.org/10.1038/s41467-021-22197-x.

64. Ma A, Wang X, Li J. *et al.* Single-cell biological network inference using a heterogeneous graph transformer. *Nat Commun* 2023;**14**:964. https://doi.org/10.1038/s41467-023-36559-0.

65. Li H, Sun Y, Hong H. *et al.* Inferring transcription factor regulatory networks from single-cell ATAC-seq data based on graph neural networks. *Nat Mach Intell* 2022;**4**:389–400. https://doi.org/10.1038/s42256-022-00469-5.

66. Vandereyken K, Sifrim A, Thienpont B. *et al.* Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* 2023;**24**:494–515. https://doi.org/10.1038/s41576-023-00580-2.

67. Heumos L, Schaar AC, Lance C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 2023;**24**:550–72. https://doi.org/10.1038/s41576-023-00586-w.

68. Shema E, Bernstein BE, Buenrostro JD. Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat Genet* 2019;**51**:19–25. https://doi.org/10.1038/s41588-018-0290-x.

69. Schwartzman O, Tanay A. Single-cell epigenomics: Techniques and emerging applications. *Nat Rev Genet* 2015;**16**:716–26. https://doi.org/10.1038/nrg3980.

70. Kelsey G, Stegle O, Reik W. Single-cell epigenomics: Recording the past and predicting the future. *Science* 2017;**358**:69–75. https://doi.org/10.1126/science.aan6826.

71. Lazaros K, Koumadorakis DE, Vlamos P. *et al.* Graph neural network approaches for single-cell data: A recent overview. *Neural Comput Appl* 2024;**36**:1–25. https://doi.org/10.1007/s00521-024-09662-6.

72. Zhou J, Cui G, Hu S. *et al.* Graph neural networks: A review of methods and applications. *AI Open* 2020;**1**:57–81. https://doi.org/10.1016/j.aiopen.2021.01.001.

73. Bruna J, Zaremba W, Szlam A. *et al.* Spectral networks and locally connected networks on graphs. In: *2nd International Conference on Learning Representations, ICLR 2014*, 2014.

74. Hammond DK, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. *Appl Comput Harmon Anal* 2011;**30**:129–50. https://doi.org/10.1016/j.acha.2010.04.005.

75. Cao Y, Zhao X, Tang S. *et al.* scButterfly: A versatile single-cell cross-modality translation method via dual-aligned variational autoencoders. *Nat Commun* 2024;**15**:2973. https://doi.org/10.1038/s41467-024-47418-x.

76. Cui X, Chen X, Li Z. *et al.* Discrete latent embedding of single-cell chromatin accessibility sequencing data for uncovering cell heterogeneity. *Nature Computational Science* 2024;**4**:1–14. https://doi.org/10.1038/s43588-024-00625-4.

77. Jia Y, Li S, Jiang R. *et al.* Accurate annotation for differentiating and imbalanced cell types in single-cell chromatin accessibility data. *IEEE/ACM Trans Comput Biol Bioinform* 2024;**21**:461–71. https://doi.org/10.1109/TCBB.2024.3372970.

78. Li S, Tang S, Wang Y. *et al.* Accurate cell type annotation for single-cell chromatin accessibility data via contrastive learning and reference guidance. *Quant Biol* 2024;**12**:85–99. https://doi.org/10.1002/qub2.33.

79. Song Q, Su J, Zhang W. scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. *Nat Commun* 2021;**12**:3826. https://doi.org/10.1038/s41467-021-24172-y.

80. Zhang W, Cui Y, Liu B. *et al.* HyGAnno: Hybrid graph neural network–based cell type annotation for single-cell ATAC sequencing data. *Brief Bioinform* 2024;**25**:bbae152. https://doi.org/10.1093/bib/bbae152.

81. Zeng Y, Luo M, Shangguan N. *et al.* Deciphering cell types by integrating scATAC-seq data with genome sequences. *Nat Comput Sci* 2024;**4**:1–14. https://doi.org/10.1038/s43588-024-00622-7.

82. Deng Y, Tang J, Zhang J. *et al.* GraphCpG: Imputation of single-cell methylomes based on locus-aware neighboring subgraphs. *Bioinformatics* 2023;**39**:btad533. https://doi.org/10.1093/bioinformatics/btad533.

83. Zheng J, Yang Y, Dai Z. Subgraph extraction and graph representation learning for single cell hi-C imputation and clustering. *Brief Bioinform* 2024;**25**:bbad379. https://doi.org/10.1093/bib/bbad379.

84. Zhu H, Liu T, Wang Z. scHiMe: Predicting single-cell DNA methylation levels based on single-cell hi-C data. *Brief Bioinform* 2023;**24**:bbad223. https://doi.org/10.1093/bib/bbad223.

85. Rao J, Zhou X, Lu Y. *et al.* Imputing single-cell RNA-seq data by combining graph convolution and autoencoder neural networks. *iScience* 2021;**24**:102393. https://doi.org/10.1016/j.isci.2021.102393.

86. Wu X, Zhou Y. GE-impute: Graph embedding-based imputation for single-cell RNA-seq data. *Brief Bioinform* 2022;**23**:bbac313. https://doi.org/10.1093/bib/bbac313.

87. Liu J, Pan Y, Ruan Z. *et al.* SCDD: A novel single-cell RNA-seq imputation method with diffusion and denoising. *Brief Bioinform* 2022;**23**:bbac398. https://doi.org/10.1093/bib/bbac398.

88. Xu C, Cai L, Gao J. An efficient scRNA-seq dropout imputation method using graph attention network. *BMC Bioinformatics* 2021;**22**:1–18. https://doi.org/10.1186/s12859-021-04493-x.

89. Xiong Z, Luo J, Shi W. *et al.* scGCL: An imputation method for scRNA-seq data based on graph contrastive learning. *Bioinformatics* 2023;**39**. https://doi.org/10.1093/bioinformatics/btad098.

90. Feng X, Fang F, Long H. *et al.* Single-cell RNA-seq data analysis using graph autoencoders and graph attention networks. *Front Genet* 2022;**13**:1003711. https://doi.org/10.3389/fgene.2022.1003711.

91. Gu H, Cheng H, Ma A. *et al.* scGNN 2.0: A graph neural network tool for imputation and clustering of single-cell RNA-Seq data. *Bioinformatics* 2022;**38**:5322–5. https://doi.org/10.1093/bioinformatics/btac684.

92. Feng X, Xiu Y-H, Long H-X. *et al.* Advancing single-cell RNA-seq data analysis through the fusion of multi-layer perceptron and graph neural network. *Brief Bioinform* 2024;**25**:bbad481. https://doi.org/10.1093/bib/bbad481.

93. Feng X, Zhang H, Lin H. *et al.* Single-cell RNA-seq data analysis based on directed graph neural network. *Methods* 2023;**211**:48–60. https://doi.org/10.1016/j.ymeth.2023.02.008.

94. Luo Z, Xu C, Zhang Z. *et al.* A topology-preserving dimensionality reduction method for single-cell RNA-seq data using graph autoencoder. *Sci Rep* 2021;**11**:20028. https://doi.org/10.1038/s41598-021-99003-7.

95. Shan Y, Yang J, Li X. *et al.* GLAE: A graph-learnable auto-encoder for single-cell RNA-seq analysis. *Inform Sci* 2023;**621**:88–103. https://doi.org/10.1016/j.ins.2022.11.049.

96. Wang H-Y, Zhao J-P, Su Y-S. *et al.* scCDG: A method based on DAE and GCN for scRNA-seq data analysis. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**:3685–94.

97. Tian S-W, Ni J-C, Wang Y-T. *et al.* Scgcc: Graph contrastive clustering with neighborhood augmentations for scrna-seq data analysis. *IEEE J Biomed Health Inform* 2023;**27**:6133. https://doi.org/10.1109/JBHI.2023.3319551.

98. Zhao J, Wang N, Wang H. *et al.* SCDRHA: A scRNA-Seq data dimensionality reduction algorithm based on hierarchical autoencoder. *Front Genet* 2021;**12**:733906. https://doi.org/10.3389/fgene.2021.733906.

99. Baul S, Ahmed KT, Filipek J. *et al.* omicsGAT: Graph attention network for cancer subtype analyses. *Int J Mol Sci* 2022;**23**:10220. https://doi.org/10.3390/ijms231810220.

100. Buterez D, Bica I, Tariq I. *et al.* CellVGAE: An unsupervised scRNA-seq analysis workflow with graph attention networks. *Bioinformatics* 2022;**38**:1277–86. https://doi.org/10.1093/bioinformatics/btab804.

101. Ji C, Yu N, Wang Y. *et al.* An end-to-end deep hybrid autoencoder based method for single-cell RNA-Seq data analysis. *IEEE/ACM Trans Comput Biol Bioinform* 2023;**20**:3889. https://doi.org/10.1109/TCBB.2023.3328029.

102. Wang L, Li W, Xie W. *et al.* Dual-GCN-based deep clustering with triplet contrast for ScRNA-seq data analysis. *Comput Biol Chem* 2023;**106**:107924. https://doi.org/10.1016/j.compbiolchem.2023.107924.

103. Li S, Guo H, Zhang S. *et al.* Attention-based deep clustering method for scRNA-seq cell type identification. *PLoS Comput Biol* 2023;**19**:e1011641. https://doi.org/10.1371/journal.pcbi.1011641.

104. Gan Y, Huang X, Zou G. *et al.* Deep structural clustering for single-cell RNA-seq data jointly through autoencoder and graph neural network. *Brief Bioinform* 2022;**23**:bbac018. https://doi.org/10.1093/bib/bbac018.

105. Abadi SAR, Laghaee SP, Koohi S. An optimized graph-based structure for single-cell RNA-seq cell-type classification based on non-linear dimension reduction. *BMC Genomics* 2023;**24**:227. https://doi.org/10.1186/s12864-023-09344-y.

106. Yu Z, Lu Y, Wang Y. *et al.* Zinb-based graph embedding autoencoder for single-cell rna-seq interpretations. In: *Proceedings of the AAAI conference on artificial intelligence*, pp. 4671–9. Palo Alto, California USA: AAAI Press, 2022.

107. Peng H, Fan W, Fang C. *et al.* SCMAG: A Semisupervised single-cell clustering method based on matrix aggregation graph convolutional neural network. *Comput Math Methods Med* 2021;**2021**:6842752. https://doi.org/10.1155/2021/6842752.

108. Zeng Y, Wei Z, Pan Z. *et al.* A robust and scalable graph neural network for accurate single-cell classification. *Brief Bioinform* 2022;**23**:bbab570. https://doi.org/10.1093/bib/bbab570.

109. Ciortan M, Defrance M. GNN-based embedding for clustering scRNA-seq data. *Bioinformatics* 2022;**38**:1037–44. https://doi.org/10.1093/bioinformatics/btab787.

110. Wang K, Li Z, You Z-H. *et al.* Adversarial dense graph convolutional networks for single-cell classification. *Bioinformatics* 2023;**39**:btad043. https://doi.org/10.1093/bioinformatics/btad043.

111. Wang T, Bai J, Nabavi S Single-cell classification using graph convolutional networks. *BMC Bioinformatics* 2021;**22**:364. https://doi.org/10.1186/s12859-021-04278-2

112. Yang R, Dai W, Li C. *et al.* scBiGNN: Bilevel graph representation learning for cell type classification from single-cell RNA sequencing data. *NeurIPS 2023 AI for Science Workshop* 2023.

113. Liu X, Shen Q, Zhang S. Cross-species cell-type assignment from single-cell RNA-seq data by a heterogeneous graph neural network. *Genome Res* 2023;**33**:96–111. https://doi.org/10.1101/gr.276868.122.

114. Shao X, Yang H, Zhuang X. *et al.* scDeepSort: A pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network. *Nucleic Acids Res* 2021;**49**:e122–2. https://doi.org/10.1093/nar/gkab775.

115. Yuan M, Chen L, Deng M. scMRA: A robust deep learning method to annotate scRNA-seq data with multiple reference datasets. *Bioinformatics* 2022;**38**:738–45. https://doi.org/10.1093/bioinformatics/btab700.

116. Bhadani R, Chen Z, An L. Attention-based graph neural network for label propagation in single-cell omics. *Genes* 2023;**14**:506. https://doi.org/10.3390/genes14020506.

117. Zeng Y, Zhou X, Rao J. *et al.* Accurately clustering single-cell RNA-seq data by capturing structural relations between cells through graph convolutional network. In: *IEEE International Conference on Bioinformatics and Biomedicine*. New York City, US: IEEE, 2020, 519–22.

118. Lee J, Kim S, Hyun D. *et al.* Deep single-cell RNA-seq data clustering with graph prototypical contrastive learning. *Bioinformatics* 2023;**39**:btad342. https://doi.org/10.1093/bioinformatics/btad342.

119. Yu Z, Su Y, Lu Y. *et al.* Topological identification and interpretation for single-cell gene regulation elucidation across multiple platforms using scMGCA. *Nat Commun* 2023;**14**:400. https://doi.org/10.1038/s41467-023-36134-7.

120. Cheng Y, Ma X. scGAC: A graph attentional architecture for clustering single-cell RNA-seq data. *Bioinformatics* 2022;**38**:2187–93. https://doi.org/10.1093/bioinformatics/btac099.

121. Hu D, Guan R, Liang K. *et al.* scEGG: An exogenous gene-guided clustering method for single-cell transcriptomic data. *Brief Bioinform* 2024;**25**. https://doi.org/10.1093/bib/bbae483.

122. Wang YM, Sun Y, Wang B. *et al.* Transfer learning for clustering single-cell RNA-seq data crossing-species and batch, case on uterine fibroids. *Brief Bioinform* 2024;**25**:bbad426. https://doi.org/10.1093/bib/bbad426.

123. Ha JS, Jeong H. GRACE: Graph autoencoder based single-cell clustering through ensemble similarity learning. *PloS One* 2023;**18**:e0284527. https://doi.org/10.1371/journal.pone.0284527.

124. Wang S, Zhang Y, Zhang Y. *et al.* scASGC: An adaptive simplified graph convolution model for clustering single-cell RNA-seq data. *Comput Biol Med* 2023;**163**:107152. https://doi.org/10.1016/j.compbiomed.2023.107152.

125. Lin E, Liu B, Lac L. *et al.* scGMM-VGAE: A Gaussian mixture model-based variational graph autoencoder algorithm for clustering single-cell RNA-seq data. *Mach Learn: Sci Technol* 2023;**4**:035013. https://doi.org/10.1088/2632-2153/acd7c3.

126. Wang P, Wen X, Li H. *et al.* Deciphering driver regulators of cell fate decisions from single-cell transcriptomics data with CEFCON. *Nat Commun* 2023;**14**:8459. https://doi.org/10.1038/s41467-023-44103-3.

127. Wang J, Chen Y, Zou Q. Inferring gene regulatory network from single-cell transcriptomes with graph autoencoder model. *PLoS Genet* 2023;**19**:e1010942. https://doi.org/10.1371/journal.pgen.1010942.

128. Mao G, Pang Z, Zuo K. *et al.* Predicting gene regulatory links from single-cell RNA-seq data using graph neural networks. *Brief Bioinform* 2023;**24**:bbad414. https://doi.org/10.1093/bib/bbad414.

129. Yang Y, Li G, Zhong Y. *et al.* Gene knockout inference with variational graph autoencoder learning single-cell gene regulatory networks. *Nucleic Acids Res* 2023;**51**:6578–92. https://doi.org/10.1093/nar/gkad450.

130. Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–64. New York, NY, USA: Association for Computing Machinery, 2016.

131. Bica I, Andrés-Terré H, Cvejic A. *et al.* Unsupervised generative and graph representation learning for modelling cell differentiation. *Sci Rep* 2020;**10**:9790. https://doi.org/10.1038/s41598-020-66166-8.

132. So E, Hayat S, Nair SK. *et al.* GraphComm: A graph-based deep learning method to predict cell-cell communication in single-cell RNAseq data. *bioRxiv* 2024;**2023**:2004.2026.538432.

133. Hu J, Li X, Coleman K. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* 2021;**18**:1342–51. https://doi.org/10.1038/s41592-021-01255-8.

134. Yuan Z. MENDER: Fast and scalable tissue structure identification in spatial omics data. *Nat Commun* 2024;**15**:207. https://doi.org/10.1038/s41467-023-44367-9.

135. Shen R, Cheng M, Wang W. *et al*. Graph domain adaptation–based framework for gene expression enhancement and cell type identification in large-scale spatially resolved transcriptomics. *Brief Bioinform* 2024;**25**:bbae576. https://doi.org/10.1093/bib/bbae576.

136. Zeng Y, Wei Z, Yu W. *et al*. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Brief Bioinform* 2022;**23**:bbac297. https://doi.org/10.1093/bib/bbac297.

137. Fatemi MY, Lu Y, Sharma C. *et al*. Feasibility of inferring spatial Transcriptomics from single-cell histological patterns for studying colon cancer tumor heterogeneity. *medRxiv* 2023;**2023**. 2010. 2009.23296701.

138. Partel G, Wählby C. Spage2vec: Unsupervised representation of localized spatial gene expression signatures. *FEBS J* 2021;**288**: 1859–70. https://doi.org/10.1111/febs.15572.

139. Marco Salas S, Yuan X, Sylven C. *et al*. De novo spatiotemporal modelling of cell-type signatures in the developmental human heart using graph convolutional neural networks. *PLoS Comput Biol* 2022;**18**:e1010366. https://doi.org/10.1371/journal.pcbi.1010366.

140. Zong Y, Yu T, Wang X. *et al*. conST: An interpretable multimodal contrastive learning framework for spatial transcriptomics. *bioRxiv* 2022;**2022**:2001. 2014.476408.

141. Li Z, Chen X, Zhang X. *et al*. Latent feature extraction with a prior-based self-attention framework for spatial transcriptomics. *Genome Res* 2023;**33**:1757–73. https://doi.org/10.1101/gr.277891.123.

142. Xu H, Fu H, Long Y. *et al*. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med* 2024;**16**:12. https://doi.org/10.1186/s13073-024-01283-x.

143. Li J, Chen S, Pan X. *et al*. Cell clustering for spatial transcriptomics data with graph neural networks. *Nat Comput Sci* 2022;**2**: 399–408. https://doi.org/10.1038/s43588-022-00266-5.

144. Liu T, Fang Z-Y, Li X. *et al*. Graph deep learning enabled spatial domains identification for spatial transcriptomics. *Brief Bioinform* 2023;**24**:bbad146. https://doi.org/10.1093/bib/bbad146.

145. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun* 2022;**13**:1739. https://doi.org/10.1038/s41467-022-29439-6.

146. Zuo C, Zhang Y, Cao C. *et al*. Elucidating tumor heterogeneity from spatially resolved transcriptomics data by multiview graph collaborative learning. *Nat Commun* 2022;**13**:5962. https://doi.org/10.1038/s41467-022-33619-9.

147. Long Y, Ang KS, Li M. *et al*. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat Commun* 2023;**14**:1155. https://doi.org/10.1038/s41467-023-36796-3.

148. Ren H, Walker BL, Cang Z. *et al*. Identifying multicellular spatiotemporal organization of cells with SpaceFlow. *Nat Commun* 2022;**13**:4076. https://doi.org/10.1038/s41467-022-31739-w.

149. Cang Z, Ning X, Nie A. *et al*. SCAN-IT: Domain segmentation of spatial transcriptomics images by graph neural network. In: *BMVC: Proceedings of the British Machine Vision Conference*. South Road, Durham, DH1 3LE, UK: BMVA Press, 2021.

150. Xu C, Jin X, Wei S. *et al*. DeepST: Identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res* 2022;**50**:e131–1. https://doi.org/10.1093/nar/gkac901.

151. Tang Z, Li Z, Hou T. *et al*. SiGra: Single-cell spatial elucidation through an image-augmented graph transformer. *Nat Commun* 2023;**14**:5618. https://doi.org/10.1038/s41467-023-41437-w.

152. Xu H, Wang S, Fang M. *et al*. SPACEL: Deep learning-based characterization of spatial transcriptome architectures. *Nat Commun* 2023;**14**:7603. https://doi.org/10.1038/s41467-023-43220-3.

153. Peng L, He X, Peng X. *et al*. Stgnnks: Identifying cell types in spatial transcriptomics data based on graph neural network, denoising auto-encoder, and k-sums clustering. *Comput Biol Med* 2023;**166**:107440. https://doi.org/10.1016/j.compbiomed.2023.107440.

154. Wu S, Qiu Y, Cheng X. ConSpaS: A contrastive learning framework for identifying spatial domains by integrating local and global similarities. *Brief Bioinform* 2023;**24**:bbad395. https://doi.org/10.1093/bib/bbad395.

155. Fang Z, Liu T, Zheng R. *et al*. stAA: Adversarial graph autoencoder for spatial clustering task of spatially resolved transcriptomics. *Brief Bioinform* 2024;**25**:bbad500. https://doi.org/10.1093/bib/bbad500.

156. Yuan Y, Bar-Joseph Z. GCNG: Graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol* 2020;**21**:1–16. https://doi.org/10.1186/s13059-020-02214-w.

157. Tang Z, Zhang T, Yang B. *et al*. spaCI: Deciphering spatial cellular communications through adaptive graph model. *Brief Bioinform* 2023;**24**:bbac563. https://doi.org/10.1093/bib/bbac563.

158. Li H, Ma T, Hao M. *et al*. Decoding functional cell–cell communication events by multi-view graph learning on spatial transcriptomics. *Brief Bioinform* 2023;**24**:bbad359. https://doi.org/10.1093/bib/bbad359.

159. Fischer DS, Schaar AC, Theis FJ. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat Biotechnol* 2023;**41**:332–6. https://doi.org/10.1038/s41587-022-01467-z.

160. Ruan Z, Zhou W, Liu H. *et al*. Precise detection of cell-type-specific domains in spatial transcriptomics. Cell rep. *Methods* 2024;**4**:100841. https://doi.org/10.1016/j.crmeth.2024.100841.

161. Jiang R, Li Z, Jia Y. *et al*. SINFONIA: Scalable identification of spatially variable genes for deciphering spatial domains. *Cells* 2023;**12**:604. https://doi.org/10.3390/cells12040604.

162. Yuan Z, Li Y, Shi M. *et al*. SOTIP is a versatile method for microenvironment modeling with spatial omics data. *Nat Commun* 2022;**13**:7330. https://doi.org/10.1038/s41467-022-34867-5.

163. Vistain LF, Tay S. Single-cell proteomics. *Trends Biochem Sci* 2021;**46**:661–72. https://doi.org/10.1016/j.tibs.2021.01.013.

164. Bennett HM, Stephenson W, Rose CM. *et al*. Single-cell proteomics enabled by next-generation sequencing or mass spectrometry. *Nat Methods* 2023;**20**:363–74. https://doi.org/10.1038/s41592-023-01791-5.

165. Ahmad R, Budnik B. A review of the current state of single-cell proteomics and future perspective. *Anal Bioanal Chem* 2023;**415**: 6889–99. https://doi.org/10.1007/s00216-023-04759-8.

166. Song R, Wang X, Zhang J. *et al*. GATDE: A graph attention network with diffusion-enhanced protein-protein interaction for cancer classification. *Methods* 2024;**231**:70–7. https://doi.org/10.1016/j.ymeth.2024.09.003.

167. Li W, Yang F, Wang F. *et al*. scPROTEIN: A versatile deep graph contrastive learning framework for single-cell proteomics embedding. *Nat Methods* 2024;**21**:623–34. https://doi.org/10.1038/s41592-024-02214-9.

168. Hu T, Allam M, Kaushik V. *et al*. Spatial Morphoproteomic features predict uniqueness of immune microar-

chitectures and responses in lymphoid follicles. *bioRxiv* 2024:2024.2001.2005.574186.

169. Cao Z-J, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol* 2022;**40**:1458–66. https://doi.org/10.1038/s41587-022-01284-4.

170. Gao H, Zhang B, Liu L. *et al.* A universal framework for single-cell multi-omics data integration with graph convolutional networks. *Brief Bioinform* 2023;**24**:bbad081. https://doi.org/10.1093/bib/bbad081.

171. Wang G, Zhao J, Yan Y. *et al.* Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with deep neural networks. *Nat Mach Intell* 2023;**5**:1200–13. https://doi.org/10.1038/s42256-023-00734-1.

172. Wen H, Ding J, Jin W. *et al.* Graph neural networks for multimodal single-cell data integration. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4153–63. New York, NY, USA: Association for Computing Machinery, 2022.

173. Zhang X, Wang X, Shivashankar G. *et al.* Graph-based autoencoder integrates spatial transcriptomics with chromatin images and identifies joint biomarkers for Alzheimer's disease. *Nat Commun* 2022;**13**:7480. https://doi.org/10.1038/s41467-022-35233-1.

174. Liu Y, Zhang J, Wang S. *et al.* A heterogeneous graph cross-omics attention model for single-cell representation learning. In: *IEEE International Conference on Bioinformatics and Biomedicine*. New York City, US: IEEE, 2022, 270–5.

175. Wang X, Duan M, Li J. *et al.* MarsGT: Multi-omics analysis for rare population inference using single-cell graph transformer. *Nat Commun* 2024;**15**:338. https://doi.org/10.1038/s41467-023-44570-8.

176. Singh R, Wu AP, Berger B. Granger causal inference on DAGs identifies genomic loci regulating transcription. *arXiv preprint arXiv* 2210.10168 2022.

177. Han X, Wang B, Situ C. *et al.* scapGNN: A graph neural network–based framework for active pathway and gene module inference from single-cell multi-omics data. *PLoS Biol* 2023;**21**:e3002369. https://doi.org/10.1371/journal.pbio.3002369.

178. Zhang Y, Chen L, Li S. CIPHER-SC: Disease-gene association inference using graph convolution on a context-aware network with single-cell data. *IEEE/ACM Trans Comput Biol Bioinform* 2020;**19**:819–29. https://doi.org/10.1109/TCBB.2020.3017547.

179. Chen G, Liu Z-P. Graph attention network for link prediction of gene regulations from single-cell RNA-sequencing data. *Bioinformatics* 2022;**38**:4522–9. https://doi.org/10.1093/bioinformatics/btac559.

180. Bafna M, Li H, Zhang X. CLARIFY: Cell–cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *Bioinformatics* 2023;**39**:i484–93. https://doi.org/10.1093/bioinformatics/btad269.

181. Li H, Li H, Zhou J. *et al.* SD2: Spatially resolved transcriptomics deconvolution through integration of dropout and spatial information. *Bioinformatics* 2022;**38**:4878–84. https://doi.org/10.1093/bioinformatics/btac605.

182. Song Q, Su J. DSTG: Deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Brief Bioinform* 2021;**22**:bbaa414. https://doi.org/10.1093/bib/bbaa414.

183. Li Y, Luo Y. Current protocols in molecular biology. *Genome Biol* 2024;**25**:206. https://doi.org/10.1186/s13059-024-03353-0.

184. Dai X, Xu F, Wang S. *et al.* PIKE-R2P: Protein–protein interaction network-based knowledge embedding with graph neural network for single-cell RNA to protein prediction. *BMC Bioinformatics* 2021;**22**:139. https://doi.org/10.1186/s12859-021-04022-w.

185. Tang W, Wen H, Liu R. *et al.* Single-cell multimodal prediction via transformers. In: *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 2422–31. New York, NY, USA: Association for Computing Machinery, 2023.

186. Alghamdi N, Chang W, Dang P. *et al.* A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. *Genome Res* 2021;**31**:1867–84. https://doi.org/10.1101/gr.271205.120.

187. Brown MB. 400: A method for combining non-independent, one-sided tests of significance. *Biometrics* 1975;**31**:987–92. https://doi.org/10.2307/2529826.

188. COS S, Vargas J, Montano AP. A survey of dimensionality reduction techniques. *arXiv preprint arXiv* 1403.2877 2014.

189. Theodoris CV, Xiao L, Chopra A. *et al.* Transfer learning enables predictions in network biology. *Nature* 2023;**618**:616–24. https://doi.org/10.1038/s41586-023-06139-9.

190. Hao M, Gong J, Zeng X. *et al.* Large-scale foundation model on single-cell transcriptomics. *Nat Methods* 2024;**21**:1481–91.

191. Feng S, Li S, Chen L. *et al.* Unveiling potential threats: Backdoor attacks in single-cell pre-trained models. *Cell Discov* 2024;**10**:122. https://doi.org/10.1038/s41421-024-00753-1.

192. Sehanobish A, Ravindra N, van Dijk D. Gaining insight into sars-cov-2 infection and COVID-19 severity using self-supervised edge features and graph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4864–73. Palo Alto, California USA: AAAI Press, 2021.

193. Roohani Y, Huang K, Leskovec J. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nat Biotechnol* 2024;**42**:927–35. https://doi.org/10.1038/s41587-023-01905-6.

194. Shahir JA, Stanley N, Purvis JE. Cellograph: A semi-supervised approach to analyzing multi-condition single-cell RNA-sequencing data using graph neural networks. *BMC Bioinformatics* 2024;**25**:25. https://doi.org/10.1186/s12859-024-05641-9.