



OPEN ACCESS

EDITED BY

Lin Wan,
Chinese Academy of Sciences (CAS), China

REVIEWED BY

Suoqin Jin,
Wuhan University, China
Youtao Lu,
University of Pennsylvania, United States

*CORRESPONDENCE

Tallulah S. Andrews,
✉ tandrew6@uwyo.edu

RECEIVED 29 September 2025

REVISED 01 December 2025

ACCEPTED 08 December 2025

PUBLISHED 27 January 2026

CITATION

Tchatchoua Ngassam B, Niu H, Pang S,
Shydlouskaya V and Andrews TS (2026)
Applications of AI to single-cell and spatial
transcriptomics: current state-of-the-art and
challenges.
Front. Bioinform. 5:1715821.
doi: 10.3389/fbinf.2025.1715821

COPYRIGHT

© 2026 Tchatchoua Ngassam, Niu, Pang,
Shydlouskaya and Andrews. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

Applications of AI to single-cell and spatial transcriptomics: current state-of-the-art and challenges

Boris Tchatchoua Ngassam ¹, Huilin Niu ¹, Sunny Pang¹,
Valeryia Shydlouskaya ² and Tallulah S. Andrews ^{1,3*}

¹Department of Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada, ²Department of Microbiology and Immunology, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada, ³Department of Computer Science, University of Western Ontario, London, ON, Canada

Artificial intelligence (AI) has become a common tool for bioinformatics, with hundreds of methods published in recent years. Due to the training data demands of deep-learning algorithms, high-throughput single-cell and spatial transcriptomics is one of the most popular areas for these applications. Here we review how AI is being used for single-cell and spatial transcriptomics analysis, and how these approaches compare to alternative statistical or heuristic-based methods. We explored 10 common analysis tasks: dimensionality reduction, cross-dataset integration, data denoising, data augmentation, deconvolution, cell-cell interactions, transcriptional velocity, transcriptomic-chromatin accessibility integration, and integrating single-cell and spatial transcriptomics modalities. We highlight which algorithms are likely to be useful for discovery researchers, and which are not yet ready for general research use.

KEYWORDS

cell-cell interactions, cross-dataset integration, data denoising, deconvolution, dimensionality reduction, integrating single-cell and spatial transcriptomics modalities, transcriptional velocity

1 Introduction

Artificial intelligence (AI) has revolutionized the analysis of big data across many fields, including biomedical research, and is entering clinical practice, with over 1,000 algorithms and devices approved by the FDA (Health, 2025). While the predominant use of AI in clinical practice is in biomedical image analysis, in research, AI approaches have gained increasing popularity in bioinformatics, and especially single-cell and spatial transcriptomics (Ge et al., 2024; Erfanian et al., 2023; Zahedi et al., 2024; Molho et al., 2024; Ma and Xu, 2022). AI is often used synonymously or as a subtopic of the broader field of machine learning. Machine learning involves a computer or algorithm deriving at least some aspects of a model from observed or “training” data. This includes tasks as simple as estimating the slope and intercept of the best-fit line, or those as complex as labelling MRI images with specific pathological lesions. AI, or deep learning (DL) as we will refer to it, is a specific class of models based on neural networks (NN) with multiple interconnected layers of functions capable of learning complex, non-linear patterns within large-scale datasets.

Single-cell and spatial transcriptomics are especially amenable to DL due to the large number of observations, as most datasets consist of thousands to millions of individual cells and thousands to tens of thousands of transcripts (Svensson et al., 2018). State-of-the-art single-cell transcriptomics (scRNA-seq) experiments typically generate large-scale datasets composed of 20,000–500,000 individual cells from at least three samples from one or more conditions (Figure 1A). These data undergo quality control, normalization, dimensionality reduction, integration across samples or across modalities, then they are clustered and annotated with cell type labels based on the expression of characteristic genes (Heumos et al., 2023; Luecken and Theis, 2019; Andrews and Hemberg, 2018; Kiselev et al., 2019). Many of these tasks are classic machine learning problems which could potentially be performed by DL models. Spatial transcriptomics (ST) adds two additional layers of information: two-dimensional coordinates of each cell, which may soon to be three-dimensional (Schott et al., 2024), as well as one or more layers of histology (H&E) and/or immunofluorescent (IF) images of the tissue. ST comes in two main types: sequencing-based (Figure 1B) and imaging-based (Figure 1C). In imaging-based ST, transcripts are individually measured with single-molecule fluorescent *in situ* hybridization (Chen et al., 2015; He et al., 2022) (Figures 1B,C). Transcripts are aggregated at the level of individual cells by identifying nuclei and cell boundaries, referred to as tissue-segmentation or simply segmentation (Mitchel et al., 2025; Polański et al., 2024). In many cases, this single-cell resolution ST data is analyzed using the same tools developed for scRNAseq. For sequencing-based ST, tissue is placed on a slide covered in oligonucleotide spots which capture and tag transcripts with a spatial barcode. Resolution is determined by the size of each uniquely barcoded spot. In many cases, these spots will overlap more than 1 cell, thus requiring “deconvolution” to estimate the contribution of each cell to the transcripts captured by that spot (Stahl et al., 2016; Rodriques et al., 2019; Gaspard-Boulinc et al., 2025). For both approaches, but particularly for sequencing-based techniques, information from the matching images can be combined with transcriptomics to improve the identification of distinct anatomical regions either in parallel with or integrated into the ST analysis workflow (Williams et al., 2022; Pham et al., 2023; Zhao et al., 2021). Tissue segmentation and extraction of biologically relevant features from tissue imaging is dominated by DL algorithms (Chen et al., 2024; Stringer et al., 2021; Warren and Moustafa, 2023; Kuntz et al., 2021; Greenwald et al., 2022).

While these technologies have generated large amounts of high-dimensional datasets, the analysis of these data is challenged by a combination of biological complexity and technical noise. Biologically, cellular states exist along continuous trajectories—such as differentiation or activation—and exhibit high heterogeneity within and across tissues. Technically, the data is affected by low sensitivity, batch effects, ambient RNA contamination, and spatial blur in low-resolution spatial assays (Ge et al., 2024; Kiselev et al., 2019; Mitchel et al., 2025; Lähnemann et al., 2020; Young and Behjati, 2020; Svensson et al., 2017). These factors introduce spurious variation, obscure true biological signals, and complicate tasks such as clustering, integration, and cell–cell communication inference.

In recent years, DL has emerged as a novel approach to address the computational challenges of scRNA-seq and ST.

These methods excel at feature extraction and classification of high-dimensional, noisy data, thus making them well-suited for cell type annotation, multimodal data integration, and nonlinear dimensionality reduction (Erfanian et al., 2023; Karin et al., 2024; Sarker, 2021). DL methods can take advantage of GPU, parallel computing, and iterative optimization on batches of data to scale analyses to datasets of millions of observations; however, similar or better performance can also be achieved by optimizing classical statistical methods (Chockalingam et al., 2025). DL models are extremely flexible and can be combined to allow for the joint analysis of multiple data types such as integration of scRNA-seq and ST data, or imaging and transcriptomic data.

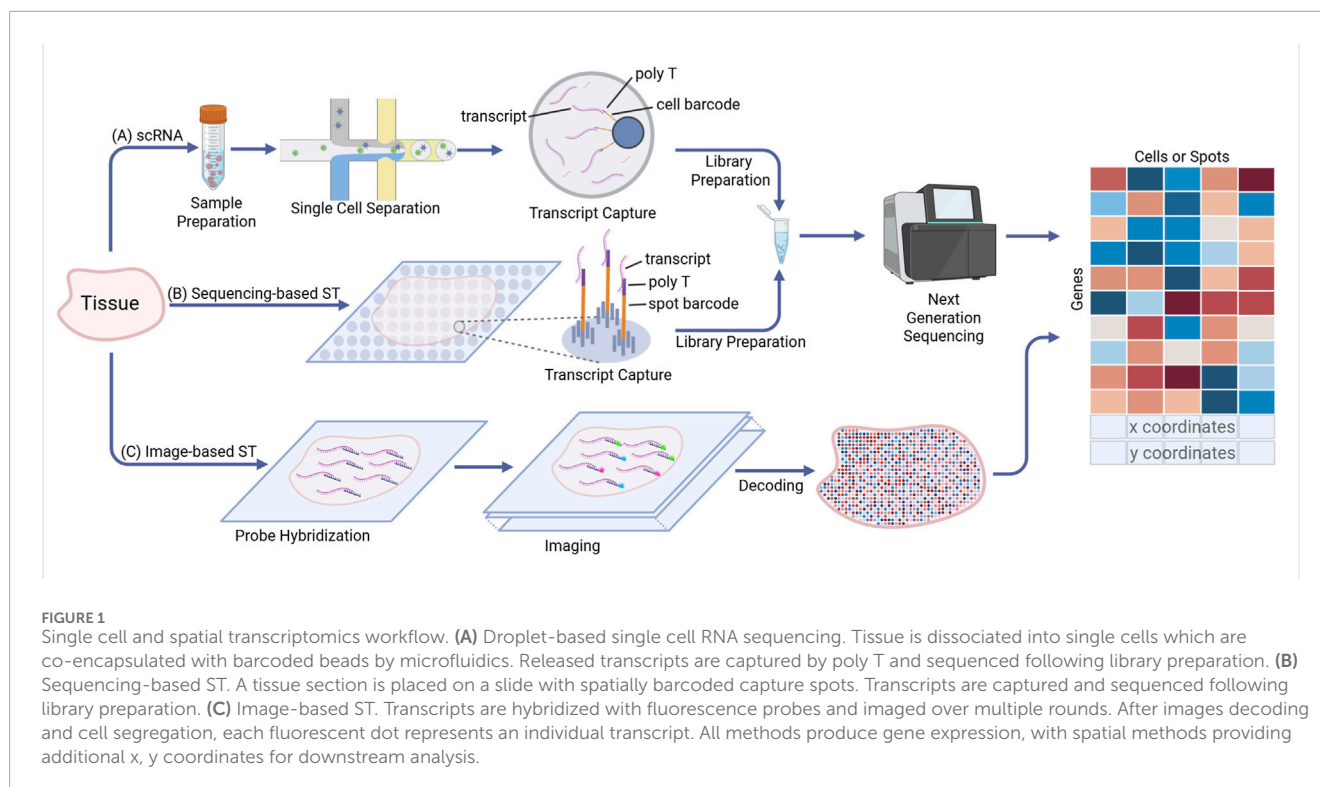
In recent years, there has been an explosion of methods developed for scRNA-seq and ST analysis using DL models (Table 1). Despite their growing number, only a few have achieved broad adoption in the research community. While existing reviews (Zahedi et al., 2024; Ma and Xu, 2022; Li Y. et al., 2022; Era et al., 2019; Luo et al., 2024; Wani et al., 2025) have primarily focused on the technical aspects of these models, their architecture, and training strategies, we focus instead on their performance in biological discovery research and on which, if any, of these tools have been shown to enhance accuracy, reproducibility, and sensitivity for biological discovery. As such, we first provide a brief overview of different model architectures, then discuss DL approaches to addressing specific bioinformatics analysis tasks, and their applicability to real-world discovery research. This will help biologically focused researchers understand when and how to use these methods and help bioinformaticians determine which tasks are appropriate for DL models and how to evaluate their design to ensure the resulting model is useful to the biomedical research community.

2 Common deep learning architecture

2.1 Convolutional neural networks (CNN)

Convolutional neural networks (CNN) were originally developed for structured data in the form of multiple arrays, such as images which are composed of pixel intensities in 2D arrays for each color channel (Lecun and Bengio, 1998). Their design is built around three core principles (Lecun and Bengio, 1998): (i) local receptive fields, which focus computation on neighboring input values to capture features such as edges and corners in images; (ii) shared weights, which enable the same filter to be applied across inputs, thereby reducing the number of parameters; and (iii) subsampling or pooling operations, which introduce robustness of outputs to distortions and shifts. Together, these principles allow CNNs to efficiently recognize local patterns and build hierarchical feature representations using fewer parameters than fully connected networks (Figure 2A). Due to these advantages, CNNs have become a popular architecture in fields such as computer vision, where extracting informative features from local patterns is crucial.

Although scRNA-seq lacks inherent spatial structure, gene expression data has been successfully adapted by restructuring it into an image-like format used by CNNs. A method called convolutional neural network for co-expression (CNNE) encodes



gene pair co-expression as 2D histograms, which serve as input “images” (Yuan and Bar-Joseph, 2019). This approach allows CNNs to learn complex, nonlinear gene-to-gene relationships directly from single-cell expression data. CNNs are particularly valuable for ST to extract morphological features from tissue sections that complement transcriptomics data. Methods such as SpaCell (Tan et al., 2020) combine pretrained CNN models with an autoencoder network to learn joint embeddings of histology and gene expression. Similarly, stLearn (Pham et al., 2023) leverages a pretrained CNN model to extract morphological features from histology images and integrates them with gene expression data to map spatial domains within tissue sections.

2.2 Autoencoders (AE)

Autoencoders (AE) are deep feed-forward neural networks fundamentally designed for unsupervised representation learning, where the goal is to learn lower-dimensional features of high-dimensional data. Structurally, an AE consists of an encoder network and a decoder network (Figure 2B). The encoder compresses input data (such as gene expression vector from a cell) into a lower-dimensional latent space, while retaining the most significant features. The decoder, which typically mirrors the architecture of the encoder, aims to reconstruct the high-dimensional input data from the learned low-dimensional representation. The entire network is trained to minimize the reconstruction error given as the mean squared error between input and reconstructed data. The resulting latent representations, also called embeddings, are particularly valuable as they serve as nonlinear counterparts to traditional linear dimensionality reduction techniques such as

Principal Component Analysis (PCA). While popular pipelines like Seurat (Stuart et al., 2019; Satija et al., 2015; Butler et al., 2018) use PCA and assume linear relationships among genes, AEs can capture complex nonlinear relationships inherent in scRNA-seq data. A key advantage of AEs lies in their flexibility to adapt the reconstruction objective based on the statistical properties of the data. For instance, loss functions can use negative binomial or zero-inflated negative binomial distributions, which are appropriate for single-cell and spatial transcriptomics data (BinTayyash et al., 2021; Svensson, 2020; Zhao et al., 2022) instead of standard statistics such as mean squared error (MSE), which assume Gaussian noise. This way, AE can incorporate probabilistic assumptions directly into the loss function by modeling the likelihood of an appropriate probability distribution. The model can then account for data-specific characteristics such as sparsity, overdispersion, and technical noise commonly observed in scRNA-seq data, hence learning more biologically meaningful representations that respect the underlying statistical structure of gene expression measurements.

In scRNA-seq analysis, Deep Count Autoencoders (DCA) leverage the flexibility of AE by modeling the output as the parameters of the zero-inflated negative binomial distribution (Eraslan et al., 2019), commonly used for RNA-seq counts (Svensson, 2020). Additionally, prior domain knowledge can be incorporated into an AE in a semi-supervised training manner as implemented by scDCC (Single Cell Deep Constrained Clustering) (Tian et al., 2021). scDCC integrates soft pairwise constraints derived from prior biological information (marker genes or cell type annotation) into the model's loss function. These constraints guide the model to group related cells and separate dissimilar ones during latent space optimization, effectively

TABLE 1 Deep learning-based tools in single-cell RNA sequencing and spatial transcriptomics.

Tool name	Task category	Modality	DL model	Key features/notes	Year	Code
SCANVI	Annotation/transfer	scRNA-seq, ST	Conditional VAE	Semi-supervised cell labeling. Extends scVI for annotation	2021	Code
BBKNN	Batch correction	scRNA-seq	Graph-KNN	Batch integration	2020	Code
BERMUDA	Batch correction	scRNA-seq	AE + clustering	Deep AE-based alignment	2019	Code
scArches	Batch correction/integration	scRNA-seq	VAE w/fine-tuning	Architectural surgery for model reuse	2021	Code
trVAE	Batch correction/integration	scRNA-seq	Conditional VAE	Domain transfer using adversarial training	2020	Code
scGEN	Batch correction/integration	scRNA-seq	VAE	Predicts perturbed gene states	2019	Code
Graphcomm	Cell-cell interaction	scRNA-seq	GAT	Integrates multimodal data for cell-cell communication	2025	Code
scSDNE	Cell-cell interaction	scRNA-seq	GNN + AE	Semi-supervised graph embedding integrating ligand-receptor and gene regulation data	2025	Code
DeepCCI	Cell-cell interaction	scRNA-seq	GCN + ResNet	Supervised cell-cell interaction network prediction using L-R pairs	2023	Code
scTenifoldXct	Cell-cell interaction	scRNA-seq	Neural network + semi-supervised, manifold alignment	Predicts cell-cell interactions and maps communication graphs using ligand-receptor gene embedding and manifold alignment	2023	Code
Spatialscope	Cell-cell interaction	Spatial transcriptomics	Deep generative model	Decomposes ST spots to single cells using generative models	2023	Code
CellFM	Cell type annotation	scRNA-seq	RetNet	Foundamental model for annotation	2025	Code
scAtlasVAE	Cell type annotation	scRNA-seq	VAE	Cross atlas comparison and transfer learning for cell subtype annotation	2024	Code
scGAA	Cell type annotation	scRNA-seq	Transformer	Combines horizontal and vertical attention mechanisms, does not require batch information	2024	Code
TOSICA	Cell type annotation	scRNA-seq	Transformer	Combine cell type marker genes and transformer attention layers	2023	Code
scBERT	Cell type annotation	scRNA-seq	Transformer	Pretrained on gene expression	2022	Code
SIMS	Classification	scRNA-seq	Transformer	Uses TabNet transformer for lable transfer from cell atlas	2024	Code
expiMap	Classification	scRNA-seq	AE w/pathway constraints	Maps cells to known pathways in a reference dataset	2023	Code
scDLC	Classification	scRNA-seq	LSTM + DNN	Sequential modeling for classification	2022	Code
SEDR	Clustering	Spatial transcriptomics	VGAE	Integrates latent of GE + spatial embedding	2024	Code

(Continued on the following page)

TABLE 1 (Continued) Deep learning-based tools in single-cell RNA sequencing and spatial transcriptomics.

Tool name	Task category	Modality	DL model	Key features/notes	Year	Code
SiGra	Clustering	Spatial transcriptomics	Graph transformer	Integrates multichannel images + expression	2023	Code
SpaGCN	Clustering	Spatial transcriptomics	GCN	Uses spatial coordinates + histology + GE	2021	Code
scDCC	Clustering	scRNA-seq	AE	Semi-supervised with pairwise constraints	2021	Code
scVAE	Clustering	scRNA-seq	VAE	Use VAE to learn low dimensional representation to facilitate accurate clustering	2020	Code
scDeepCluster	Clustering	scRNA-seq	AE	Unsupervised clustering with deep autoencoder	2019	Code
GSI	Clustering	Spatial transcriptomics	VAE	Integrates image + GE + spatial coordinates to improve clustering	2025	Code
Deep scSTAR	Clustering/Annotation/Embedding	scRNA-seq	DAE + MLP + MTL	Denosing autoencoder with supervised MLP in latent space	2025	Code
SAUCIE	Clustering/Batch correction	scRNA-seq	AE	Use maximal mean discrepancy penalty to match distributions of batches	2019	Code
STAGATE	Clustering/batch correction	Spatial transcriptomics	GAT	Adaptive graph attention on spots	2022	Code
SPADE	Clustering/deconvolution	Spatial transcriptomics	spaGCN + Lasso regression	H&E img + GE for clustering, then uses ref scRNAseq data for domain deconvolution	2024	Code
SPACEL	Clustering/deconvolution	Spatial transcriptomics	VAE + GCN	Self-supervised local clustering + simulation	2023	Code
SpaCell	Clustering/embedding	ST + histology	AE + CNN	AE model for embeddings and CNN for classification	2020	Code
scResolve	Deconvolution	Spatial transcriptomics	Transformer + VAE	Reference-free, integrate cell segmentation of histology image	2024	Code
UniCell Deconvolve	Deconvolution	Spatial transcriptomics + bulk RNA seq	Deep feedforward network	Foundamental model	2023	Code
DAISM-DNNXMBD	Deconvolution	Bulk RNA seq	DNN	Train 1 DNN for each cell type	2022	Code
Tangram	Deconvolution	Spatial transcriptomics	Custom model	Custom probabilistic model + gradient descent optimization + backpropagation	2021	Code
DSTG	Deconvolution	Spatial transcriptomics	CCA + MNN + GCN	Graph reconstruction	2021	Code
Scaden	Deconvolution	Bulk RNA seq	DNN	Ensemble of three best-performing DNN	2020	Code

(Continued on the following page)

TABLE 1 (Continued) Deep learning-based tools in single-cell RNA sequencing and spatial transcriptomics.

Tool name	Task category	Modality	DL model	Key features/notes	Year	Code
scAR	Denoising	scRNA-seq	VAE	Ambient RNA denoising	2022	Code
DCA	Denoising/imputation	scRNA-seq	Autoencoder	Deep count autoencoder (NB/ZINB)	2018	Code
STGNNks	Embedding	Spatial transcriptomics	GAE	Graph-based clustering	2023	Code
scSemiProfiler	Embedding	scRNA-seq	VAE-GAN + active learning	Learns cell states <i>via</i> active bulk supervision	2023	Code
scGNN	Embedding/clustering	scRNA-seq	GNN	Graph-based denoising, clustering, embedding	2021	Code
scVI	Embedding/imputation/integration	scRNA-seq, ST	VAE	Probabilistic latent space, batch correction	2018	Code
scGFT	Generation	scRNA-seq	GAN, VAE, GFT	Generate synthetic scRNA seq data that reflects natural biological variability	2025	Code
STAGE	Generation	Spatial transcriptomics	AE	Data generation	2024	Code
scCross	Generation	scRNA-seq	VAE + GAN + MNN	Cross-domain latent space used for simulation	2024	Code
SRTsim	Generation	Spatial transcriptomics	Empirical sim	Simulates spot-based ST data	2023	Code
cscGAN	Generation	scRNA-seq	Conditional GAN	Cell type aware generator	2020	Code
scIGANs	Imputation	scRNA-seq	GAN	Conditional GAN	2020	Code
DeepImpute	Imputation	scRNA-seq	DNN	Imputation using sub-neural network modules	2019	Code
autoCell	Imputation/feature extraction	scRNA-seq	Graph-enhanced VAE	Uses VAE and GNN	2023	Code
scGPT	Integration	scRNA-seq	Transformer	Foundational model	2024	Code
SpatialGLUE	Integration	ST + proteinmics + epigenomics	AE + graph fusion	Integrate the different omics modalities with spatial information	2024	Code
MultiVI	Integration	scRNA + ATAC	VAE	Joint ATAC–RNA modeling	2023	Code
SCALEX	Integration	scRNAseq	Encoder + GAN	Use feature links to preserve biological variation	2021	Code
spaVAE	Low-dimensional space	Spatial transcriptomics	VAE	NB model based VAE, combining Gaussian process prior and Gaussian prior	2024	Code
COVET	Low-dimensional space	scRNA-seq, ST	ENVI + CVAE	Encode the covariance of gene expression between neighboring cells joint latent space	2024	Code
scMODAL	Multimodal integration	scRNA-seq + ATAC	Multimodal AE + GAN	Use feature links to align cell embeddings	2025	Code
scMVP	Multimodal integration	scRNA-seq + scATAC-seq	Multi-view VAE	Handles paired multi-omics, encodes ATAC with attention, integrates views for embedding and clustering.	2022	Code

(Continued on the following page)

TABLE 1 (Continued) Deep learning-based tools in single-cell RNA sequencing and spatial transcriptomics.

Tool name	Task category	Modality	DL model	Key features/notes	Year	Code
GLUE	Multimodal integration	scRNA-seq, scATAC-seq, snmC-seq	VAE per modality + graph linking features	Models regulatory feature interaction across modalities; scalable large-dataset integration.	2022	Code
totalVI	Multimodal integration	scRNA-seq + protein	VAE	Probabilistic multimodal model	2021	Code
Cobolt	Multimodal integration	scRNA-seq + scATAC-seq (and others)	Multimodal VAE	Integrates joint and single-modality datasets.	2021	Code
scButterfly	Multimodal integration	scRNAseq + ATAC	U-net + AE	Image-guided gene embedding	2024	Code
Monae	Multimodal integration	scRNAseq + ATAC	AE + Contranstive learning	Modality-specific auto-encoders	2024	Code
MIDAS	Multimodal integration/Batch correction/Embedding	scRNAseq + ATAC + ADT (proteomics)	AE	Self supervised modality alignment, transfer learning	2024	Code
BIDCell	Self-supervised learning	Spatial transcriptomics	AE + biologically-informed loss	Learns spatial gene-region relationships	2023	Code
STAGNN	Spatial clustering	Spatial transcriptomics	GAT	Graph attention network (GAT) and the time series model informer	2024	Code
TransformerST	Spatial domain clustering	Spatial transcriptomics	ViT + adaptive graph transformer	Uses H7E image features and GE in self-attention transformer	2024	Code
DeepST	Spatial domain detection	Spatial transcriptomics	Multi-stage deep learning using DNN, VGAE	Image + gene-based spatial clustering	2022	Code
GIST	ST integration	Spatial transcriptomics	CNN + graph transformer	Uses GE + cell type-informative paired tissue images e.g., IF	2022	Code
DeepVelo	Trajectory/RNA velocity embedding	scRNA-seq	GCN + DNN	Models gene- and cell-specific transcriptional kinetics	2024	Code
VeloVI	Trajectory/RNA velocity embedding	scRNA-seq	VAE	Learns gene-specific kinetics, provides uncertainty quantification for velocities; flexible for time-dependent transcription rates.	2023	Code
cellDancer	Trajectory/RNA velocity embedding	scRNA-seq	DNN	Predicts cell- and gene-specific transcription, splicing and degradation rates	2023	Code
VeloVAE	Trajectory/RNA velocity embedding	scRNA-seq	VAE	Extends velocity modelling with VAE framework to capture kinetic variability.	2022	Code
LatentVelo	Trajectory/RNA velocity embedding	scRNA-seq	VAE/latent emberdding model	Learns latent representation for velocity; enables batch correction and dynamics embedding.	2022	Code
VeloAE	Trajectory/RNA velocity embedding	scRNA-seq	AE	Embeds velocity information (spliced/unspliced) for better dynamic modelling.	2021	Code

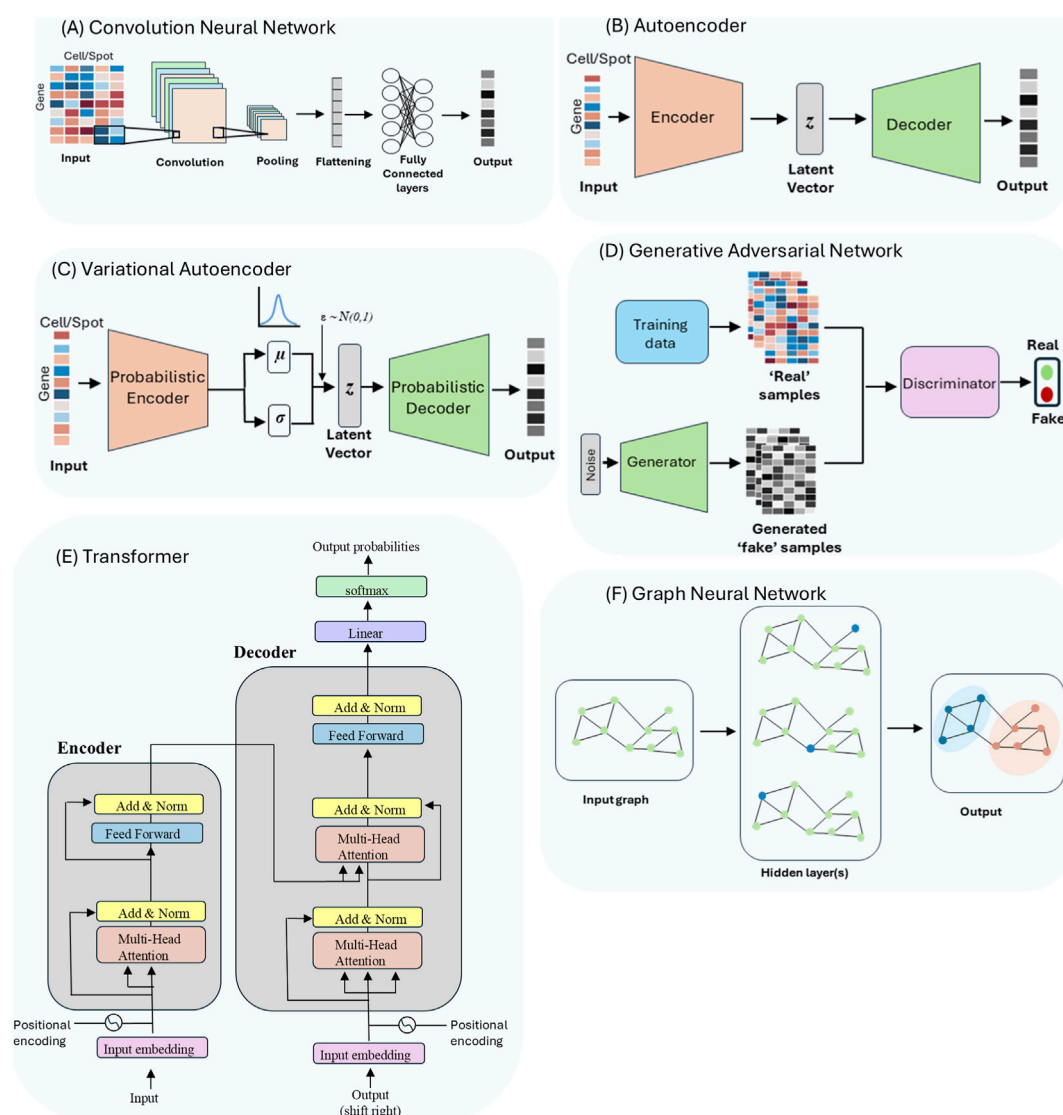


FIGURE 2

Deep learning architectures commonly applied to single-cell and spatial transcriptomics. **(A)** Convolutional neural network (CNN): extracts local spatial patterns from image-like inputs (e.g., cell/spot \times gene maps) via convolution–pooling stacks. **(B)** Autoencoder (AE): learns a low-dimensional latent vector (z) that reconstructs the input, enabling denoising and feature learning. **(C)** Variational autoencoder (VAE): probabilistic AE that learns a distribution over (z) (parameterized by (μ, σ)) and samples ($z + \epsilon \sim N(0,1)$) for generative modeling. **(D)** Generative adversarial network (GAN): a generator synthesizes expression profiles from noise while a discriminator distinguishes real from generated samples. **(E)** Transformer tokenizes inputs and applies positional embeddings with stacked self-attention and feed-forward blocks in an encoder to produce task-specific outputs. **(F)** Graph neural network (GNN): propagates information over a cell/spot graph to model neighborhood structure and produce node-level outputs.

shaping the embedding to reflect domain knowledge. This approach improves clustering accuracy and biological relevance, especially in complex or noisy datasets, showcasing autoencoders as versatile frameworks for single-cell data analysis.

2.3 Variational autoencoders

Variational autoencoders (VAEs) are a probabilistic extension of standard AEs, designed to improve representation learning and generative modeling by incorporating principles of Bayesian inference to learn a distribution over a latent (lower-dimensional)

space. This probabilistic formulation addresses a key limitation of AEs: their deterministic latent space, which often results in discontinuous or overfitted representations that generalize poorly to unseen data and lack support for structured sampling (Kingma and Welling, 2022; Doersch, 2021; Kingma and Welling, 2019; Rezende et al., 2014). Despite their architectural similarity, VAEs differ fundamentally in that they encode each input to the parameters of a probability distribution (usually Gaussian) from which a latent variable is sampled (Figure 2C). The decoder reconstructs the input data from this latent representation. This formulation enables VAEs to learn smooth, continuous, and structured latent representations by optimizing a joint loss function

composed of a reconstruction term and a Kullback-Leibler (KL) divergence term, which regularizes the approximate posterior distribution to be close to the prior distribution. The key advantage of VAEs lies in their ability to model data uncertainty and support generative capabilities through a probabilistic latent space. This is particularly valuable for scRNA-seq, where modeling sparsity, overdispersion and technical noise is essential (Svensson, 2020).

Models such as scVI (Lopez et al., 2018) (Single-Cell Variational Inference) build upon the VAE framework to model scRNA-seq count data using a negative binomial likelihood, while simultaneously correcting for batch effects. Similarly, totalVI (Gayoso et al., 2021) extends the VAE architecture to jointly model RNA and protein data from CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing), enabling multimodal inference (Stoeckius et al., 2017). Concretely, totalVI places a logistic-normal prior on a shared cell-level latent representation that parameterizes modality-specific likelihoods by using a negative binomial RNA counts and a negative-binomial mixture for proteins, respectively. In ST, SpaVAE (Tian et al., 2024) incorporates spatial coordinates *via* a Gaussian process prior on the latent space that is indexed by the spot coordinates while keeping some latent dimensions under the standard gaussian prior to capture non-spatial spot variations. In general, VAEs are flexible in that different likelihoods can be used and latent priors can also be customized to encode known structure in the data such as spatial information and batch effects.

2.4 Generative adversarial networks (GANs)

Instead of learning to reconstruct what already exists, GANs learn by deception (Goodfellow et al., 2014). They consist of a generator, which creates synthetic data from random noise, and a discriminator, which attempts to distinguish between real and generated samples (Figure 2D). Through adversarial training, the generator improves its ability to produce realistic outputs, while the discriminator becomes more adept at detecting “fake” or synthetic data. This dynamic results in a generator that can synthesize high-quality, biologically plausible gene expression profiles.

In scRNA-seq, cscGAN/scGAN (Marouf et al., 2020) learns to generate cell type conditioned expression profiles that preserve gene–gene dependencies, supporting augmentation of rare populations and improving downstream classification and clustering. scIGAN (Xu et al., 2020) frames imputation as generation, using an adversarial loss (often combined with count-aware objectives) to recover missing values while retaining biological variability in different cell types. Adversarial alignment has also been used for batch/platform correction. For instance, iMAP (Wang D. et al., 2021) couples an autoencoder backbone with a GAN discriminator that removes batch signal from the latent space, enabling cross-platform integration of tumor microenvironment datasets while preserving cell-state structure.

GANs are widely used in digital pathology for histology image generation and translation, demonstrating strong capability on imaging. However, in ST there is still no widely adopted, end-to-end GAN framework that jointly models histology images, gene expression, and spatial coordinates. Challenges such as training instability, mode collapse, and lack of biological interpretability

make it difficult to ensure that generated spatial gene expression patterns reflect true biological variation rather than technical artifacts. As a result, GANs are not standard components of ST analysis pipelines, where AE, VAEs, GNNs, and transformers currently dominate.

2.5 Transformer

Transformers are deep learning models originally developed for natural language processing (NLP) with an encoder-decoder architecture composed of self-attention layers (Vaswani et al., 2023) (Figure 2E). Although they are similar to AEs in design, they differ in several aspects. The encoder and decoder can be trained and used individually, as seen in models used by BERT and GPT respectively (Yenduri et al., 2023; Devlin et al., 2019). The self-attention layers dynamically integrate each input element with all elements within the same input sequence, capturing contextual relationships. Additionally, the encoder is not constrained by a low-dimensional latent space, and the decoder is usually trained to autoregressively generate a target sequence rather than reconstruct the input (Vaswani et al., 2023; Xiong et al., 2025). These properties have made transformers the backbone of modern foundational models, which are pretrained on large and heterogeneous datasets and then adapted to a wide range of downstream tasks with minimal supervision.

Transformers have driven significant advances in modeling sequential data in domains like natural language processing (Wu et al., 2025), time-series analysis (Wen et al., 2023), and DNA (Avsec et al., 2021) and protein sequences (Rives et al., 2021), for which they were originally designed. Transcriptomics data is inherently non-sequential and requires the encoding of gene expression values into token-like embeddings, analogous to tokens in NLP, which transformers can process. Current approaches vary in how they represent expression levels, each with distinct advantages and limitations. One approach is ordering, where genes are ranked by transcript abundance within a cell and treated as an ordered sequence of tokens, with each gene assigned a learned embedding (Levine et al., 2024), as implemented by tGPT (Shen et al., 2023), iSEEK (Shen et al., 2022), GeneMamba (Qi et al., 2025), and Geneformer (Theodoris et al., 2023). While this method captures relative patterns and is more robust to technical noise and batch effects (Shen et al., 2023; Qi et al., 2025), quantitative expression information is lost during data transformation (Levine et al., 2024), resulting in reduced data resolution. A second approach is bin-based discretization, where gene counts are grouped into predefined bin sizes, each with an assigned learnable embedding (Yang et al., 2022; Cui et al., 2024). Although the absolute scale of expression is preserved and sequence modeling is simplified, fine-grained biological signal is lost, particularly for genes with subtle but functionally relevant expression differences, which can be sensitive to bin boundaries and potentially affect downstream analysis. Alternatively, the value projection strategy avoids discretization altogether by directly mapping gene expression values to a learnable embedding, which is combined with a gene-specific embedding (Hao et al., 2024a; Zeng et al., 2025), resulting in a transformer input token. This retains the full resolution of the original data and avoids artifacts due to discretization.

In ST, transformers' ability to take multimodal input and model long range dependencies offers distinct advantages over other methods (Xu P. et al., 2023; Hao et al., 2024b; Wen et al., 2024). In contrast to local neighborhood-based approaches such as GNN or clustering algorithms, that focus on immediate spatial proximity, transformers can capture global spatial relationships across tissue sections through self-attention.

2.6 Graph neural networks

Graph Neural Networks (GNNs) are deep learning models designed to operate on graph-structured data, where entities are represented as nodes and their relationships as edges (Figure 2F). Unlike architectures that treat samples as independent vectors, GNNs iteratively update node representations by aggregating information from their neighbors, making them well suited to capture community structure, dependencies, and spatial organization. This is particularly relevant for single-cell and spatial transcriptomics, where cells can be connected by transcriptional similarity, gene co-expression networks, or spatial spots by physical adjacency.

A key strength of GNNs is that they operate directly on graphs while integrating with other deep models, which improves representation learning for biological data. Graph Convolutional Networks (GCNs) extend convolution to cell-cell graphs and enable semi-supervised label transfer. scGCN (Song et al., 2021) builds a hybrid graph that links reference and query datasets through mutual-nearest-neighbor connections in a shared low-dimensional space and augments it with within-query neighbors. A GCN then propagates labels across this graph using variable-gene features, aligning matched cells and flagging unlabeled cells. In ST, SpaGCN (Hu et al., 2021) constructs a weighted spatial graph that combines spot proximity, histology image features and gene expression similarity and then uses a GCN to learn spot representations for tissue domain detection.

Beyond CNNs, GNNs have been incorporated into standard and variational AE frameworks to enable representation learning guided by transcriptomic similarity and spatial proximity. Models such as GVAE (Graph Variational Autoencoders) (Simonovsky and Komodakis, 2018) integrate GNNs with VAEs, leveraging the generative capacity of VAEs together with graph-based regularization. In scRNA-seq, graph-sc (Ciortan and Defrance, 2022) uses a graph autoencoder framework to learn low-dimensional embeddings used for clustering, while scGNN (Wang J. et al., 2021) extends this approach by reconstructing both gene expression and cell similarity graph structures. More recently, self-attention has been incorporated into GNN, giving rise to Graph Attention Networks (GATs) that learn edge-specific weights during neighborhood aggregation instead of averaging contributions equally from all neighbors as in GCNs (Veličković et al., 2018). STAGATE (Dong and Zhang, 2022) adapts this approach with a graph-attention autoencoder on the spatial neighbor network, where self-attention layers learns edge-specific weights normalized with softmax which are then used to update spot specific representations. In contrast, GraphST (Long et al., 2023) employs a GNN encoder with contrastive learning on the spatial graph, encouraging nearby neighbors map to similar representations and forcing distant spots

to map to dissimilar ones. This contrastive formulation yields representations that are more robust to noise and batch effects, thereby improving domain separation as well as downstream clustering.

2.7 Hybrid models

Recent advances in deep learning for single-cell and spatial transcriptomics have led to the development of hybrid models that combine the strengths of multiple architectures to address complex, multimodal challenges. These models integrate components from different frameworks such as VAEs, GANs, GNNs, and Transformers to capture diverse aspects of biological data, including nonlinear dependencies, spatial structure, temporal dynamics, and multimodal relationships. Unlike monolithic architecture, hybrid models are designed to be modular and flexible, enabling tailored solutions for specific biological questions.

One common hybrid design combines VAEs and GANs, leveraging the probabilistic latent space of the VAE for structured representation learning and the adversarial refinement of the GAN for improved sample generation. iMAP (Wang D. et al., 2021) (AE + GAN) exemplifies this approach by using a GAN to align latent spaces across batches.

Another combination integrates GNNs with VAEs (different from GVAE), where the GNN captures spatial or transcriptional neighborhood information, and the VAE provides a probabilistic and generative framework. For instance, scGNN (Wang J. et al., 2021) combines graph-based message passing with autoencoding to jointly reconstruct gene expression and preserve cell-cell similarity.

More recently, hybrid models have incorporated transformers and GNNs, merging global attention with local graph structure. STAGATE (Dong and Zhang, 2022) uses a GAT to model spatial dependencies, effectively combining the neighborhood aggregation of GNNs with the weighted feature integration of attention. This allows the model to identify both local tissue domains and long-range functional relationships. These hybrid approaches demonstrate that the future of deep learning in genomics lies not in isolated architecture, but in strategic integration, where each component addresses a specific biological or technical challenge. By combining the generative power of VAEs, the spatial awareness of GNNs, the global context of transformers, and the realism of GANs, hybrid models offer a more comprehensive and interpretable framework for analyzing the complexity of single-cell and spatial data.

3 Applications of DL to scRNA-seq and ST analysis tasks

Most methods utilize unsupervised models, which do not require any "ground truth" or predetermined labels for the training data. This enables these methods to be trained on each individual experiment, customizing the model for each application. Alternatively, DL models can be pretrained on hundreds to thousands of datasets of a similar type to create a generalizable 'foundation' model (Chen et al., 2024; Heimberg et al., 2025).

For example, the UNI foundation model of pathology images was trained on over 100,000 individual images (Chen et al., 2024), whereas stLearn (Pham et al., 2023) and scVI (Lopez et al., 2018) retrain their NNs to extract dataset-specific features. In contrast, supervised models require training data with a known ground truth answer for the specific task it is designed to perform. Most often, these models involve classification, such as stDeepSort, which was trained on various reference datasets to annotate cell types in single-cell data (Shao et al., 2021), or Cellpose, trained to recognize and segment cells based on thousands of manually labelled training images (Stringer et al., 2021).

The most common use of DL when analyzing high dimensional data, such as scRNA-seq and ST, is to learn a lower dimensional embedding space, conceptually similar to principal component (PCA) space but without the assumptions and constraints. This embedding space can then be used for a variety of tasks either within the DL framework or extracted and used in standard statistical analysis as a replacement for PCA. Here we will discuss the main approaches to generating DL embeddings and their application for scRNA-seq and ST data.

3.1 Dimensionality reduction, clustering, and spatial domain identification

Clustering is one of the most fundamental analytical tasks in scRNA-seq and ST as it enables researchers to uncover distinct cellular populations and tissue substructures in an unsupervised, unbiased manner. Due to the, high-dimensional nature of scRNA-seq and ST data, clustering is always performed on a lower dimensional representation of the data (Figures 3A,B). Conventionally, this is PCA space (Luecken and Theis, 2019; Kiselev et al., 2019; Butler et al., 2018; Wolf et al., 2018), which is used to generate a cell-cell similarity graph, to which community detection algorithms such as Louvain (Blondel et al., 2008) or Leiden (Traag et al., 2019) clustering are applied. However, PCA assumes the lower dimensions to be linear and orthogonal and requires input data to be approximately normally distributed, thus requires pre-processing and normalization prior to use with scRNA-seq and ST data. To overcome these limitations, autoencoders (AEs/VAEs) and transformers can be used, and their learned lower dimensional embedding can be substituted for normalization and PCA in the conventional clustering pipeline. These approaches preserve the unsupervised and unbiased nature of the analysis while relaxing the assumptions and constraints required by PCA.

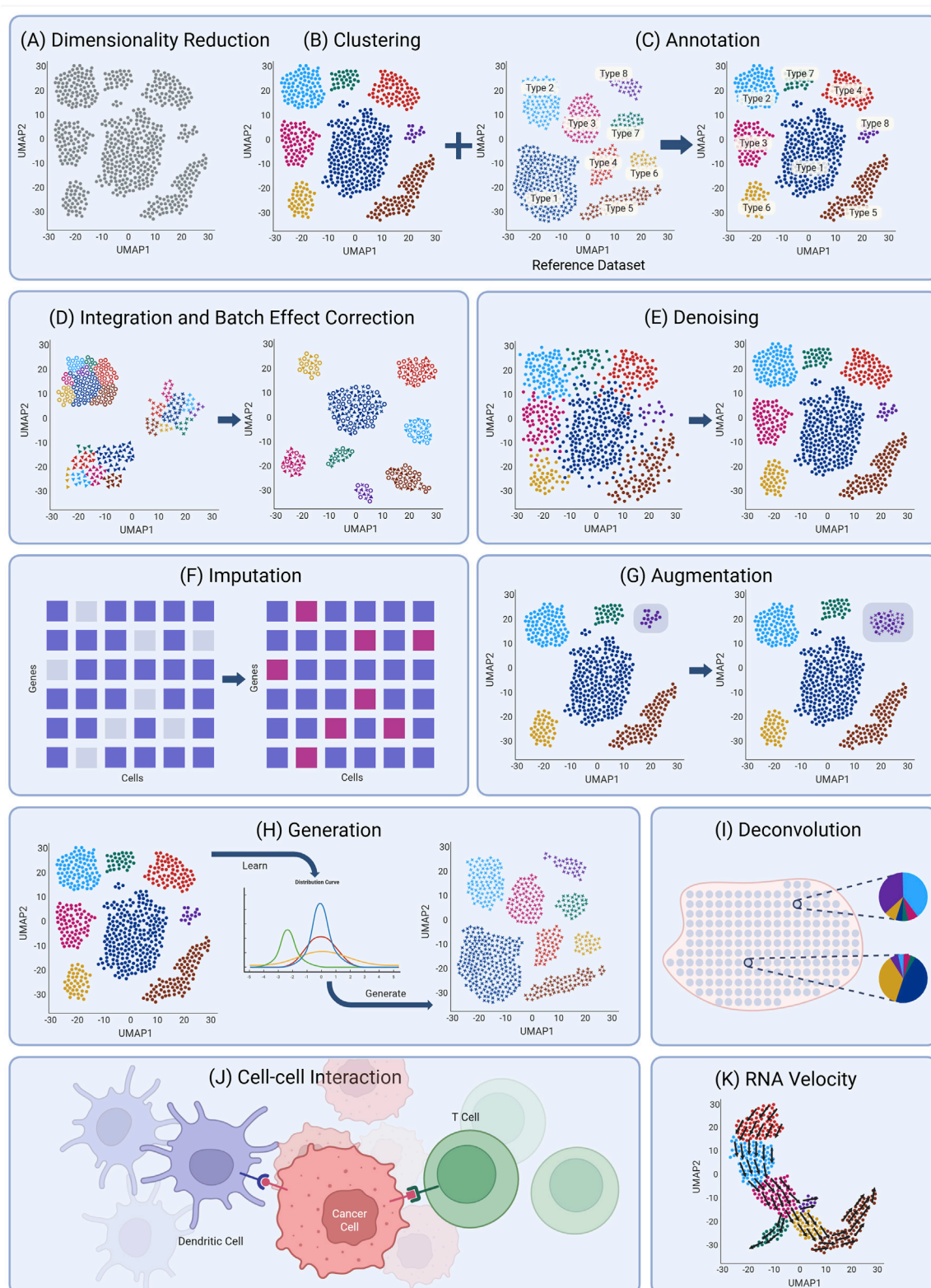
For scRNA-seq, a common approach is to use a VAE as implemented in scVI (Luecken and Theis, 2019; Kiselev et al., 2019; Lopez et al., 2018; Wolf et al., 2018), which incorporates a negative binomial distribution in the cost-function to model raw scRNA-seq data. Unlike most DL methods, scVI is widely used in biological analysis and is a foundation for other methods including scArches (Lotfollahi et al., 2022) and scANVI (Xu et al., 2021). In independent benchmarks, scVI embeddings are found to perform similarly to classical PCA for identification of cell types (Liang et al., 2024; Li and Quon, 2019). Other DL clustering methods for scRNA-seq include scDCC (Tian et al., 2021) and scDeepCluster (Tian et al., 2019). ScDeepCluster uses an AE architecture with a decoder that generates parameters of a zero-inflated negative binomial which

is used to calculate a probabilistic loss function for scRNAseq data. scDCC extends scDeepCluster by incorporating soft pairwise constraints (e.g., must-link/cannot-link pairs derived from marker genes or protein expression) into the loss function, allowing prior biological knowledge to guide the clustering process. The method demonstrated good performance on both small (thousands of cells) and large (tens of thousands of cells) datasets, where even a few thousand constraints representing a small fraction of possible cell pairs enhanced clustering performance based on quantitative scores (e.g., Adjusted Rand Index) and more meaningful clusters than scDeepCluster, especially in difficult cases like the worm neuron dataset. However, scDCC performed similarly to state-of-the-art non-DL methods in their in-house benchmark. Whereas scDeepCluster marginally outperformed rival methods but did not compare to Louvain/Leiden clustering. Benchmarking of clustering performance is challenging due to the lack of truly orthogonal ground truth; however, these results suggest that there is no need for non-linear DL dimensionality reduction for cell type identification in scRNA-seq. In terms of applicability to biological discovery, scVI and scANVI have been used in multiple studies for dataset integration and embedding, demonstrating their utility (Salcher et al., 2022; Lindeboom et al., 2024; Yang LX. et al., 2025).

In addition to the above methods, which train a model on one specific dataset, foundation models trained on hundreds of datasets are increasingly common in scRNA-seq. Pre-trained models, such as scGPT (Cui et al., 2024) or SCimilarity (Heimberg et al., 2025) project data onto a common lower-dimensional space which could be used for clustering and novel cell type discovery. Additionally, this lower dimensional data can also be used for automatic annotation, which we will discuss further in the next section, as this space can be biased towards the most frequent cell types and miss rare cell types (Cui et al., 2024). scAtlasVAE took a foundation model approach to specifically examining T-cell heterogeneity and was able to characterize novel T-cell phenotypes when used in an unsupervised manner, identifying 18 unique and reproducible T-cell states (Xue et al., 2025).

DL approaches are also common for ST clustering due to the ease of incorporating image and/or spatial information into such models compared to the standard clustering pipeline. GCNs can incorporate spatial information by linking adjacent cells/spots into a spatial-proximity graph, leading to their use in methods such as SpaGCN (Hu et al., 2021), STAGATE (Dong and Zhang, 2022), GraphST (Long et al., 2023), SiGra (Tang et al., 2023), and DeepST (Xu et al., 2022). Similar to scRNA-seq, benchmarking studies find that DL approaches perform similarly to non-DL methods that also incorporate spatial information (Yuan et al., 2024; Hu et al., 2024a), but outperform methods that do not incorporate spatial information.

Image information is typically incorporated into ST clustering using a separate image-focused AE/VAE or GNN, which learns salient image features from individual image patches associated with the gene expression spots. These are then integrated with gene-expression features to obtain a combined embedding for each tissue spot. Although deep learning is commonly used to extract complex, high-level image features in ST clustering, some methods use non-DL approaches to integrate spatial context through hand-crafted image features. For instance, Squidpy (Palla et al., 2022) computes interpretable morphological features—such as summary statistics

**FIGURE 3**

Deep learning application in single cell transcriptomics and spatial transcriptomics. **(A)** Dimensionality reduction. High-dimensional data is projected into low-dimensional space (e.g., UMAP). **(B)** Cells are clustered into distinct groups represented by different colors. **(C)** Automatic annotation of cell clusters using a reference dataset. **(D)** Integration and batch correction across different batches. **(E)** Data is denoised to recover true signal. **(F)** Data imputation to infer missing gene expression. Grey blocks (left) represent missing values, and pink blocks (right) represent imputed values. **(G)** Synthetic cells are generated to enrich rare cell type (light grey shading). **(H)** A new dataset is generated by learning distribution parameters from a reference

(Continued)

FIGURE 3 (Continued)
 dataset. **(I)** Each spatial transcriptomics spot is resolved into cell type fractions. **(J)** Cell-cell interactions between different cell types (e.g., dendritic cells, cancer cells, and T cells) are modeled through ligand-receptor signaling to infer intercellular communication. **(K)** Directional RNA velocity vectors are projected onto a UMAP to infer cell state transitions and lineage trajectories.

(mean, standard deviation), histogram-based quantiles, or textural properties (contrast, homogeneity) derived from co-occurrence matrices—for each spatial spot directly from the histology image. Similarly, SpaGCN (Hu et al., 2021) integrates image information by mapping each spatial spot to its corresponding location in the H&E image, calculating a smoothed mean RGB color value from a local pixel neighborhood, and then combining these values into a single weighted feature that reflects tissue patterns. Whereas those which use AE/VAE extracted images, gain a significant benefit from the image features, but most of the performance is driven by the gene-expression information (Tang et al., 2023; Li B. et al., 2024).

All of these methods have been demonstrated to reproduce known anatomy, but none have demonstrated a capability to identify novel, biologically meaningful structures, due to limitations in validation and ground truth availability. Thus, these approaches should be considered validated as a supplement to aid anatomical annotation by an expert. However, their capacity for novel discovery remains unknown.

Overall, AE and VAE methods for scRNAseq perform comparably to PCA and may be good alternatives when working with very large datasets. In particular, scVI has proven strong performance in many studies. For ST, DL approaches are a necessity when integrating image information into lower dimensional embeddings. GraphST is currently the best performing DL method for ST spatial domain identification.

3.2 Automatic annotation

Increasingly, scRNA-seq clustering is being supplemented with direct algorithmic annotation of cells with their cell type identity (Luecken and Theis, 2019) (Figure 3C). Comparing novel cells to existing annotated scRNA-seq dataset enables the inference of cell type identity through simple guilt-by-association approaches, and many early methods simply used standard similarity metrics or standard machine-learning algorithms such as support vector machines or random forests while achieving reasonably accurate results (Kiselev et al., 2018; Abdelaal et al., 2019). However, these methods tended to perform poorly on fine-scale classification of subtypes or cell-states.

DL models are highly amenable to supervised classification tasks such as cell type annotation, and, once trained, are highly efficient and scalable to millions of novel data points (Cheng et al., 2023a). Thus, dozens of novel DL models have been developed for this task using a variety of architectures, including GPT-4 and scBERT - large language models which use marker genes to annotate cells using the scientific literature (Yang et al., 2022; Hou and Ji, 2024); scGAA and TOSICA - attention-based transformer models which compare novel cells to narrow reference datasets (Chen J. et al., 2023); and pre-trained foundation models, such as scGPT (Cui et al., 2024) or CellFM (Zeng et al., 2025).

Most of these methods achieve annotation accuracies of ~80–90%; however, in many cases, benchmarking is performed by splitting individual datasets into training and test sets, which is biased in favor of good model performance. This is because there are no systematic batch effects between the training and test data, as would be present in a real use case when these models are applied to completely novel scRNA-seq dataset (Yang et al., 2022; Cui et al., 2024; Zeng et al., 2025; Cheng et al., 2023a). Only scGPT was tested on a left-out data partition, achieving good results (accuracy >85%) for 70% of cell-types; however, performance rapidly declined as the difference between query and reference datasets increased, with fewer than 50% of cell types achieving good performance when the query dataset originated from an unseen disease state (Cui et al., 2024). Many of these methods are so recent that no independent benchmarking is available. However, in previous independent benchmarks, DL models outperformed many non-DL annotation algorithms but did not outperform a support vector machine trained on the same reference data (Kiselev et al., 2018; Chen J. et al., 2023). In these independent benchmarks, performance was found to rapidly degrade for DL models when reference data does not exactly match the query data, in agreement with the results shown for scGPT. However, DL models do show promise in their ability to accurately distinguish similar cell subtypes when provided sufficient training data (Zeng et al., 2025).

In discovery research, automatic annotation is typically used simply as a first pass, which is then manually checked and refined. Thus, even imperfect results from automatic annotation can still be useful to guide and accelerate annotation efforts (Clarke et al., 2021). Algorithms that assign a confidence score to annotations are most useful, since novel cell types may be discovered where automatic annotation has low confidence (Chen J. et al., 2023; Ergen et al., 2024). DL models naturally provide quantitative scores for annotation confidence, enhancing their utility in this use-case. In addition, as scRNA-seq resources continue to grow, approaches such as foundation models may be more easily expanded or fine-tuned to incorporate new training data compared to approaches based on traditional statistics. Thus, researchers should either use the method with training data most similar to their own, or if that is unknown we recommend scGPT for human data due to its extensive benchmarking so users can accurately assess how confident they should be in the results.

3.3 Integration and batch effect correction

Transcriptomic experiments often include multiple biological replicates which may be collected across multiple experimental batches, individuals, tissues, or different platforms, leading to various non-biological variations known as batch effects (Figure 3D). These technical artifacts cause identical cell types from

different batches to appear distinct (Luecken et al., 2022; Chazarra-Gil et al., 2021; Tran et al., 2020). Early batch effect correction approaches, such as Combat (Johnson et al., 2007), used statistical regression to remove batch covariates. However, these methods tend to remove important biological variation unless it is specified as *a priori* within the model. To circumvent this, the next-generation of methods used techniques such as canonical correlation analysis or mutual nearest neighbors to identify shared biological variation across batches to preserve, while removing factors of variation ascribed to batch effects (Butler et al., 2018; Haghverdi et al., 2018; Hie et al., 2024). The current state-of-the-art non-DL integration method is Harmony (Korsunsky et al., 2019), which uses an iterative clustering then correction approach and is consistently among the top-performing methods in recent benchmarks (Tran et al., 2020; Antonsson and Melsted, 2024).

DL approaches to data integration modify the AE/VAE approach, as described above, to learn a ‘joint’ embedding space that captures biological groups while mixing different technical batches. A common approach to this modification is the use of adversarial learning, which penalizes the model for embeddings that leave batches separate (Hrovatin et al., 2024). Methods using this approach, such as scVI (Lopez et al., 2018), scANVI (Xu et al., 2021), and SAUCIE (Amodio et al., 2019), are not constrained by the linearity assumptions required by many non-DL methods, thus potentially enabling more efficient batch effect removal. An alternative approach uses conditional AE/VAEs which include the batch label in the joint embedding; data is then integrated by treating the batch effect as a linear transformation in the lower-dimensional space and projecting all batches onto a single reference sample or reference dataset. Prominent methods using this approach include scGen (Lotfollah et al., 2019) and scArches (Lotfollahi et al., 2022). Foundation models, such as scGPT, can also be fine-tuned to create project-specific joint embeddings. The extensive pre-training of such models includes ignoring batch effects and emphasizing conserved biology.

Despite theoretical advantages of DL methods for batch integration, they have often struggled in benchmarking studies, rarely matching the performance of Harmony (Luecken et al., 2022; Korsunsky et al., 2019; Lee et al., 2023). One potential cause of their poor performance is a tendency to over-correct and remove biological information, particularly when batches have substantially different cell type proportions (Luecken et al., 2022; Hrovatin et al., 2024). This can be mitigated by explicitly modeling cell types to ensure their preservation, as can be done for scGen and scANVI; however, since the goal of integration is usually to merge samples prior to clustering and cell type annotation, such an approach is generally limited to meta-analyses and atlasing projects.

While scRNA-seq integration can be achieved even with linear models, DL methods have been more successful when integrating multi-omics data, i.e., joint scRNA-seq and single-cell ATAC-seq (Lee et al., 2023). DL models excel at projecting different data types, such as multiome data, into similar embedding spaces, facilitating their integration (see section 3.9). This capability is further enhanced when combined with graph-based representations, which model cells as nodes and similarities or spatial relationships as edges. Graph structures enable the propagation of information across neighboring cells,

effectively capturing local dependencies, preserving topology, and improving the alignment of biological states across datasets. This is particularly valuable for integrating spatial transcriptomics data or enforcing structural continuity multiple slides of the same tissue (Khan et al., 2025; Zhang C. et al., 2024). Similar to single-slide clustering performance, the top two methods for ST integration are a Bayesian statistical approach, (Li and Zhou, 2022), and a DL approach, (Long et al., 2023; Hu et al., 2024a).

While some DL methods are competitive with state-of-the-art non-DL approaches for dataset integration, there is no clear advantage to using DL for these tasks. Scalability is often cited as the main advantage of DL integration, there are several highly scalable non-DL approaches as well, including Harmony. Two non-DL approaches are consistently among top-performers in independent benchmarks: Harmony and scMerge (Luecken et al., 2022; Tran et al., 2020; Antonsson and Melsted, 2024; Lin et al., 2019). When integrating experimental replicates containing identical cell type frequencies Harmony is recommended, however, if samples contain some non-overlapping cell type scMerge is preferable (Tran et al., 2020). For atlasing and meta-analyses it can be more optimal to utilize scANVI if cell type labels are available for the respective datasets (Luecken et al., 2022). For ST data, these scRNAseq methods can be used when data is aggregated at the cell or spot level; however spatial information is lost and this often results in poor spatial contiguity of integrated clusters. For spatially contiguous ST data the Bayesian-statistics based BASS algorithm has been shown to be the best option (Hu et al., 2024). However, altering observed data can only result in a loss of information, thus integration should only be used when inspection of the data indicates substantial batch effects are present.

3.4 Denoising and imputation

Denoising and imputation are two closely related but conceptually distinct tasks in single-cell transcriptomics. Denoising refers to the reduction of technical noise such as amplification bias, batch effects, or stochastic dropout while preserving the true biological signal (Figure 3E). The goal is not to “fill in” missing values, but to refine observed expression levels to better reflect underlying biology. In contrast, imputation explicitly aims to predict unobserved or missing values, such as zero counts, that are likely due to technical dropout rather than true biological absence (Figure 3F). While both processes can result in modified gene expression matrices, their objectives differ: denoising aims to improve signal-to-noise ratios, while imputation attempts to recover missing information. Despite this distinction, the terms are often used inconsistently in the scRNA-seq and ST literature. Many methods described as “imputation tools” (e.g., MAGIC (van Dijk et al., 2018), scImpute (Li and Li, 2018)) perform what is effectively denoising, as they smooth expression values without necessarily distinguishing between true zeros and dropouts.

Denoising data was one of the first applications of DL models (Vincent et al., 2008). AE models have been used to denoise many types of data in various contexts; in the biomedical field,

(Gondara, 2016), (Su et al., 2015), and many-omics dataset (Eraslan et al., 2019; Lal et al., 2021; Webel et al., 2024). Due to the low input material in single-cell assays, there are many missing values, and sampling- or RNA-capture-related noise is high relative to the true biological signals. Hence, many DL algorithms have been developed to denoise scRNA-seq and ST data.

One of the first and most used approaches is deep-count autoencoder (DCA) (Eraslan et al., 2019). DCA modified the traditional AE architecture to output parameters of a statistical distribution for each input gene, rather than a single predicted value. Multiple distributions are available, including negative binomial and zero-inflated negative binomial for RNA-seq data. This alteration allows DCA to account for uncertainty in the input data and biological stochasticity. Another popular method, scVI, takes a similar approach (Lopez et al., 2018). Many other model designs have been explored, including CNNs (Zhang W. et al., 2024), gene partitioning and sub-networks (Arisdakessian et al., 2019), GCNs (Huang et al., 2023), and contrastive learning (Xu et al., 2020; Shi et al., 2023). Application of these methods to biological datasets can improve the interpretability of the data; for instance, DCA increased *CD3E* expression from 80% to 99.9% in T cells and recovered *ITGAX* expression consistent with NK biology.

Only DCA, scVI, and DeepImpute have been independently benchmarked alongside non-DL denoising and imputation methods (Cheng et al., 2023b; Andrews and Hemberg, 2019; Hou et al., 2020; Huang et al., 2025). These benchmarks find conflicting results, reflecting differences in testing datasets and specific tasks used to evaluate performance. When evaluated on their ability to recover corrupted expression values or improve accuracy of automatic cell type annotation, DL denoising methods performed well, similar to other imputation and denoising methods. For unsupervised clustering and pseudotime analysis, results range from modest improvement to worse performance than the raw data, depending on the specific dataset and analysis pipeline. Whereas for gene-gene correlations, differential expression, cell type markers, and cell-cell interactions, all benchmarks find that denoising introduced a significant number of false-positive results. Hence, for scRNA-seq data, denoising remains controversial and rarely used in discovery research.

For ST data, integration with scRNA-seq is more common than direct denoising of ST data alone, which is discussed later in this manuscript. However, some methods do exist to directly denoise ST data using GNNs (Tang et al., 2023; Duan et al., 2024). Benchmarking of these methods is more limited, but SiGra is shown to increase the number of differentially expressed genes - though the extent to which these are false positives is not explored - and to improve distinctiveness of clustering. Whereas Impeller (Duan et al., 2024) is only shown to recover masked expression values.

Overall, it is not recommended to perform denoising or imputation except to enhance the sensitivity of clustering analysis, and caution must be exercised in the interpretation of results to avoid false-positives. Integration across experiments or modalities is likely a more useful task and more reliable approach for increasing statistical power by increasing the number of samples in discovery research.

3.5 Data generation and augmentation

Deep learning has increasingly been leveraged for data generation and augmentation in scRNA-seq and ST to address limitations posed by small sample sizes, rare cell types, and costly experimental procedures. Data augmentation in scRNA-seq and ST analysis is used differently than in machine learning and typically refers to the computational creation of additional data points, and adding them - 'augmenting' - to the original measured data (Figure 3G). In contrast, we will use 'data generation' to describe methods which create data either for the purposes of simulating data for benchmarking, or to generate data of a different modality-e.g., predict scRNA-seq from bulk RNA-seq.

In scRNA-seq, VAEs-based models like scVI and scVAE (Li and Li, 2018) can be used to generate synthetic cells that preserve the statistical properties and cellular identities of the original cell (Figure 3H). Generative models such as cscGAN (Xu P. et al., 2023) and scGFT (Vincent et al., 2008) have demonstrated the ability to generate realistic synthetic cells that preserve intrinsic gene expression profiles of the original data. Current state-of-the-art clustering and trajectory analysis algorithms, such as maximum modularity or minimum spanning trees, can be biased with respect to the number of cells, leading to poor performance when datasets include rare cell types. Selective generation and augmentation using cscGAN or scGFT can rebalance datasets, which were shown to improve clustering and trajectory inference performance to correctly identify rare cell types and accurately resolve trajectory branches. However, similar to denoising, data augmentation involves artificially amplifying the power of statistical tests, thus, are likely to result in inflated type-1 errors if used for differential expression, though this has not yet been tested.

In spatial transcriptomics, data generation is typically used for denoising purposes (Hu et al., 2021; Tang et al., 2023; Pratama et al., 2025). For instance, SiGra, discussed previously, replaces observed data with generated data to perform its denoising. Similarly, the STAGE model focuses more on accurate data generation but uses that generated data to recover and denoise down-sampled data as well as to impute between sequential ST slices (Li et al., 2024b). Both methods integrate spatial embeddings with gene expression features using autoencoders and other representation learning approaches to learn a feature space, from which new samples can be drawn and decoded into new expression data. SiGra uses both gene expression and features from matching histology, whereas STAGE uses gene expression only. Compared to single-cell RNA-seq, there are currently relatively few methods dedicated specifically to data generation and augmentation in ST. While emerging techniques focus on integrating image features, spatial coordinates, and gene expression for augmentation, these models only generate gene expression data, not matching image data, thus lacking the ability to fully generate ST data.

Similar to imputation, there is substantial risk of increasing Type-I errors when augmenting datasets with synthetically generated data. Thus, such approaches must be used with care. For data augmentation, the main utility is in facilitating detection of rare cell types or smoothing out cell density along developmental trajectories to better align data with the limitations and assumptions of the analytical tools for clustering and trajectory analysis. The only other use for data generation is for benchmarking algorithms,

however, most DL generative algorithms lack the fine-scale control required to design specific ground-truth cases for that type of testing thus this area is still dominated by small-scale statistical simulation methods often custom designed for a specific benchmarking task.

3.6 Deconvolution

In transcriptomics, deconvolution is the decomposition of bulk expression data into cell type proportions or cell type specific expression (Im and Kim, 2023) (Figure 3I). Deconvolution is typically applied to bulk RNA-seq or low-resolution ST where each spot typically contains multiple cells. Methods for bulk RNA-seq deconvolution can be broadly grouped into statistical approaches: (Chu et al., 2022; Peng et al., 2019; Wang et al., 2019):enrichment-based methods (Aran et al., 2017; Yoshihara et al., 2013) and machine learning models (Newman et al., 2015; Newman et al., 2019). With the emergence of deep learning, at least 13 DL-based deconvolution tools have been developed for bulk RNA seq using a scRNA-seq reference (Lomas Redondo et al., 2025). These methods are typically based on multilayer perceptrons (MLPs), autoencoders, or transformers, and are trained to reconstruct cell type proportions from mixed bulk expression profiles. Scaden (Menden et al., 2020) was one of the first deep learning tools in this area. It uses an ensemble strategy that combines three deep neural networks with different numbers of layers, activation functions, and dropout settings to improve generalization. DAISM-DNN^{XMBD} (also called Aginome-XMU) instead trains a separate deep neural model for each cell type to predict proportions (Lin et al., 2022).

Bulk deconvolution methods are typically benchmarked by comparing their predictions against cell type proportions derived from *in vitro* experiments or from *in silico* bulk samples generated using single-cell RNA-seq data. Both Scaden and DAISM-DNN^{XMBD} have been independently benchmarked among the top-performing methods, with Scaden suffering high false-positive rates (Tran et al., 2023) and DAISM performing well in both coarse-grain and fine-grain deconvolution (White et al., 2024). This demonstrates that deep learning provides a strong alternative to traditional approaches. Newer methods may outperform DAISM, but this cannot be established until a systematic benchmark study has been performed that includes the other DL-based deconvolution tools.

Overall, bulk RNA-seq deconvolution enables researchers to reduce experimental costs while still gaining insight into the tumor or tissue microenvironment. However, the performance of DL deconvolution methods requires high quality training dataset and is prone to poor generalization (Wolfram-Schauerte et al., 2025). Most researchers still rely on traditional deconvolution approaches, and only a few studies have utilized DL-based tools for deconvolution (Chen et al., 2025; Codino et al., 2025; D'Sa et al., 2025).

Bulk RNA-seq deconvolution tools can be used for ST data, but additional improvements in performance may be achieved by incorporating the spatial information. Many ST deconvolution methods use non-DL approaches such as numerical optimization (Dong and Yuan, 2021), or probabilistic models (Kleshchevnikov et al., 2022). Several DL-based deconvolution methods not only estimate the cell type fractions but can also estimate the number of cells per spot, generate gene

expression for each deconvolved cell, or estimate individual cell locations (Gaspard-Boulinec et al., 2025).

Reference-based DL deconvolution methods use three general strategies: supervised-learning, similarity-based integration, and foundation models. Supervised-learning creates synthetic ST spots by combining scRNA-seq data and use this as ground truth to train a neural network to predict cell type fractions from the aggregated expression profile (Lund et al., 2022; Bae et al., 2022; Zhan et al., 2025; Xu H. et al., 2023; Mañanes et al., 2024). Similarity-based integration methods embed scRNA-seq and ST data into a shared space through graph construction (Long et al., 2023; Ding et al., 2024; Song and Su, 2021; Li and Luo, 2024; Yin et al., 2024; Zhang et al., 2023), autoencoders (Liao et al., 2022; Hao et al., 2024; Coleman et al., 2023; Li H. et al., 2022), or optimization (Biancalani et al., 2021) to match ST spots to scRNA-seq cell types based on similarity or distance measures. In some methods, pseudo-spots are generated to aid embedding (Ding et al., 2024; Song and Su, 2021; Li and Luo, 2024; Yin et al., 2024; Zhang et al., 2023; Li H. et al., 2022). UniCell Deconvolve (UCD) is the only foundation model trained for deconvolution (Charytonowicz et al., 2023). It is a feedforward neural network trained on over 840 cell types from 899 single cell datasets. UCD uses transfer learning to adapt the foundation model to specific context where users have an option to input a contextualized reference profile to fine-tune a regression model using UCD base embedding. UCD outperformed other methods on synthetic mixtures from its own training data, but had only average performance on out-of-sample tests unless it was fine-tuned on the relevant datasets (Charytonowicz et al., 2023). An alternative approach is taken by scResolve, which imputes pixel-level gene expression which is combined with cell-segmentation of the respective histology image to infer single-cell resolution expression (Chen H. et al., 2023). This enables reference-free deconvolution and potentially novel cell type discovery.

Due to the wide variety of spatial deconvolution tools, no systematic benchmark study has yet been conducted across all methods, and most DL-based approaches have not been benchmarked. Benchmarking is especially challenging in ST deconvolution since ground truth is not available; instead, simulated ST datasets generated from scRNA-seq are typically used. Tangram (Biancalani et al., 2021) and DSTG (Song and Su, 2021) have been benchmarked in multiple independent studies alongside non-DL methods (Li et al., 2023; Chen J. et al., 2022; Yan and Sun, 2023; Li B. et al., 2022). While Tangram was shown to be superior in predicting the spatial distribution of transcripts in one study, both Tangram and DTSG generally ranked within the top third of approaches benchmarked. However, the top three performing methods overall were non-DL approaches. DL methods have the advantage of integrating multimodal data, such as histology images, which may provide additional information such as cell morphology to aid deconvolution.

For discovery focused researchers cell2Location (Kleshchevnikov et al., 2022) and SpatialDWLS (Dong and Yuan, 2021) remain top choices for deconvolution when reliable reference single-cell datasets are available. Tangram is an acceptable alternative, and scResolve is the only method capable of deconvolution when no reference single-cell data is available.

3.7 Cell-cell interactions

A key goal of single-cell RNAseq was to identify interactions between different cell types which would normally be obscured in bulk tissue samples. Many heuristic methods have been developed for this task, including CellChat (Jin et al., 2021), CellPhonedb (Efremova et al., 2020), SingleCellSignalR (Cabello-Aguilar et al., 2020), and NicheNet (Browaeys et al., 2020), which use databases of ligand-receptor (LR) pairs and calculate a co-expression score of each pair between pairs of cell types. Some of these have been expanded to account for spatial location, for use with spatial transcriptomics (Efremova et al., 2020; Dimitrov et al., 2024). Currently, there are only a few DL approaches to inferring these interactions in single cell data and none for spatial transcriptomics.

DeepCCI (Yang et al., 2023a) integrates ResNet and a GCN model to infer cell-cell interactions with a common decoding layer. This decoding layer is trained using consensus interactions obtained from the heuristic methods. As a result, in their in-house benchmarking DeepCCI identifies the same interaction as multiple heuristic methods though may have fewer false-positive results than any of the heuristic methods used alone. It is unclear whether DeepCCI gains anything from the DL components, as opposed to their in-house consensus of the heuristic models used to train it.

An advantage of DL approaches is the ability to integrate multiple data sources; this is utilized by GraphComm (So et al., 2025) to integrate pathway annotations in addition to direct LR interactions into a prior interaction probability between each LR pair. Coexpression of LR pairs is calculated and is integrated with the prior using a graph attention network. The embedding contains both cell types and LR genes and is used to generate LR pairwise scores and cell type x cell type scores by multiplying the respective embeddings. Alternatively, ScTenifoldXct (Yang Y. et al., 2023) and scSDNE (Jia et al., 2025) first infer gene-gene dependencies either using a DL model (scSDNE) or a regression model (ScTenifoldXct), which is combined with a LR coexpression score which is then used to generate a gene embedding space using a graph-autoencoder architecture. Cell-cell interactions are inferred from proximity of LR pairs in the gene embedding space. ScSDNE and ScTenifoldXct have the advantage of using semi-supervised learning, whereas GraphComm relies on database-derived LR interactions to train their embedding space. Limited in-house benchmarking is available for these, but they perform similarly to heuristic methods, with GraphComm seeming to have higher sensitivity, whereas scSDNE and ScTenifoldXct are more conservative, performing similar to a consensus of heuristic methods.

Cell-cell interaction inference remains challenging, primarily due to the lack of any true gold-standard benchmarks. In many cases, methods are benchmarked using spatial transcriptomics data, as distant cells are unlikely to interact, but this cannot provide individual LR interaction information, or with very small sets of manually curated interactions. This is particularly problematic for DL algorithms due to their reliance on training data to optimize the models. Typically, researchers use multiple LR algorithms and use some kind of consensus as evidenced by the popularity of the LIANA package (Dimitrov et al., 2024). The natural ability of DL to integrate multiple types of data may be an advantage here, as significant amounts of perturbation data are available which

could potentially be used to augment cell-cell interaction inference. However, currently there is little evidence due to lack of gold-standard datasets to favour any specific method over any other.

3.8 Combining single-cell and spatial transcriptomics

ST and scRNA-seq are complementary techniques; scRNA-seq accurately assesses the entire transcriptome for each individual cell but it loses all spatial information, whereas in ST spatial information is preserved but either data is not at single-cell resolution and/or does not capture the entire transcriptome. As a result, many methods have been developed to combine scRNA-seq and ST using different approaches. SIMO uses optimal-transport to align single cells to ST based on only RNAseq or both RNA and ATACseq modalities (Yang P. et al., 2025). Alternatively CellTrek (Wei et al., 2022) uses mutual-nearest-neighbour integration combined with random forests to predict spatial location of individual cells from proximity within the integrated embedding space. In in-house benchmarking CellTrek performed well on simulated ST data but was not compared to DL alternatives.

One of the first and most established models is Tangram, which learns a mapping between scRNA-seq and ST that optimizes the spatially correlation between mapped and observed gene expression (Biancalani et al., 2021). The authors demonstrate its effectiveness in recapitulating known expression patterns across cortical layers. In independent benchmarks, Tangram out-performs other methods for recovering downsampled gene expression values but shows modest performance at predicting cell type composition of ST data (Li B. et al., 2022). However, notably neither the original publication nor independent benchmarks assessed potential for generation of false-positive results. Generative DL models can predict scRNA-seq profiles from ST data based on a reference scRNA-seq dataset. For example, SpatialScope uses a probabilistic DL model to predict cell type composition of individual ST spots and to decompose gene expression by cell type, and then uses a generative DL model to create scRNA-seq for individual cells based on the decomposed profiles (Wan et al., 2023). In contrast, stImpute predicts gene expression for unmeasured genes in imaging-based ST using a joint AE embedding and GNN, based on known gene-gene relationships (Zeng et al., 2024).

Prediction of additional data modalities or higher resolution data from cheaper, lower resolution experimental protocols is a popular use-case for DL method development. ScSemiProfiler predicts scRNA-seq from bulk RNA, which has the advantage of being able to predict cell type specific differences in expression which is not possible with non-generative deconvolution methods (Wang et al., 2024). Using matched bulk and scRNA-seq data from COVID-19 patients, the authors were able to show their method could capture individual difference beyond what was present in the training data. However, they did not evaluate whether scSemiProfiler's cells would lead to the same biological conclusions on the effect of COVID-19 as the original scRNA-seq. Thus, it remains unclear if this approach is viable for discovery research.

Lastly, over a dozen algorithms have been published that predict ST expression data from histology images. Histology images are

plentiful and easily collected, whereas ST is relatively rare and expensive; therefore, accurate prediction of the latter from the former would be very valuable. However, performance of all current methods is relatively poor with correlations between predicted gene expression and true measured gene expression below 0.2 for most genes (Wang et al., 2025). While performance is best for genes with strong spatial patterning, correlations remain below 0.5 in nearly all cases, still far below an accuracy that would be useful for discovery research. Such methods may improve as ST experimental platforms improve, though it is also possible that much of gene expression does not manifest as any visible difference in histology images, thus placing a hard limit on the maximum accuracy of these methods. The most likely limitation of current models, however, is the availability of ST training data with high quality matching histology images as most publicly available data only release a compressed low-resolution image.

Overall, discovery researchers are recommended to choose methods which project single-cells onto ST data rather than any generative approaches, such as SIMO or CellTrek, and to use multiple different methods to ensure conclusions are robust to the approach chosen. While generative DL approaches are promising for converting between transcriptomic technologies, there is insufficient benchmarking in real-world use cases to know whether these methods lead to false or misleading conclusions.

3.9 Integrating multiomic data

ST data can be considered multiomic in that images and spatial coordinates can be treated as another layer of data to be integrated. However, more often multiomic data refers specifically to single-cell data where both mRNA is captured and sequenced and DNA is captured either for direct DNA sequencing or most often for ATAC assays, which measure open chromatin across the genome (Mimitou et al., 2021; Cao et al., 2018; Reyes et al., 2019). While first developed for single cells, equivalent assays have been developed for spatially-resolved assays (Jiang et al., 2023; Guo et al., 2025; Deng et al., 2022). However, currently only simultaneous single-cell RNA-seq and ATAC-seq has been developed into a simple off-the-shelf platform, thus is by far the most used multiome technique.

Popular methods for single-cell multiome (scMultiome) data integration and analysis include ArchR (Granja et al., 2021), Signac (Stuart et al., 2021), and MOFA (Argelaguet et al., 2020) which perform data normalization, dimensionality reduction, and clustering. Signac and ArchR in addition identify correlated open-chromatin peaks and nearby gene-expression which can be used to infer gene-regulatory networks. These approaches are all statistical approaches, with ArchR and Signac both using latent semantic indexing for data embedding, and MOFA using a Bayesian probabilistic model for joint factor analysis.

DL approaches have several advantages for multiomic data integration. They can innately align different input data such that ATAC peaks do not have to be assigned to genes prior to integration. They can be regularized to learn comparable representations for different modalities from the data rather than using heuristic normalization strategies. Finally, the architecture can be data-type invariant allowing the same structure to be used for many

different data modalities. The general structure of DL multiome methods starts with modality-specific AEs or VAEs then combines the modality-specific embeddings into a single representation (Ashuach et al., 2023; Gong et al., 2021; Li G. et al., 2022; Cao and Gao, 2022).

MultiVI (Ashuach et al., 2023) uses this approach to expand the scVI architecture to multiome data by penalizing the model for divergent representations for the same cell in different modalities then using the average representation for each cell. This enables efficient integration of paired and unpaired datasets since unpaired data simply uses the single representation value. Cobolt (Gong et al., 2021) has a very similar architecture but uses a Dirichlet prior and reconstructs the original matrices rather than using the decoder to estimate the original distribution. scMVP (Li G. et al., 2022) has the same overall architecture but uses self-attention and mask-attention encoders for each modality and simply concatenates the latent spaces for the joint embedding. Whereas GLUE (Cao and Gao, 2022) uses heuristic methods to infer ATAC-peak to RNA-gene associations which are used as knowledge graph as an additional decoder output from the concatenated multiomic latent space of their AE.

In multiple independent benchmarks (Xiao et al., 2024; Liu et al., 2025; Hu et al., 2024b; Fu et al., 2025), Seurat's weighted nearest network (WNN) consistently output performs other integration methods in perfectly matched RNA + ATAC data, whereas MultiVI is consistently optimal for partially overlapping datasets. In contrast, GLUE is the best performer when ATAC and RNA datasets are from separate samples. Notably, these results were simply for the level of integration of the lower dimensional embedding, i.e., the mixing of ATAC and RNA modalities while preserving or enhancing cell type identities. One benchmark (Hu et al., 2024b) evaluated modality prediction, and while MultiVI was a top performer, all methods had relatively poor performance (correlation <0.4) generally due to overestimation for genes upregulated in a particular group of cells, this is in line with other benchmarking of imputation methods where data smoothing typically inflates signals resulting in false-positives (Andrews and Hemberg, 2019).

Overall MultiVI and GLUE are both established methods with strong performance in benchmarks and would be good choices especially for projects with not completely overlapping scMultiome data. Heuristic methods, particularly Seurat's WNN method, are good choices for perfectly matched datasets but are inadequate for non-overlapping datasets. Imputation is still unreliable and should not be used for statistical analyses, though may be useful for identifying trends for independent validation. While DL algorithms have been developed for integration and imputation of scMultiome, inference of gene-regulatory networks which is often the main goal of Multiome studies has not yet been addressed with DL methods and may be an opportunity for future method development.

In an independent benchmark on curated datasets, scJoint, MultiVI and GLUE were top performing methods for integrated cell type identification in scMultiome data (Xiao et al., 2024). However others find high variability in performance dataset to dataset and that MultiVI was particularly sensitive and either were among top performers or worse performers depending on the dataset in question (Lee et al., 2023).

For spatial multiome, many of the above single-cell methods would be applicable; however, when spatial data includes contiguous homogeneous regions, it is often beneficial to incorporate spatial information as we noted above. Currently, the only method that integrates spatial location for spatial multiome data is SpatialGlue (Long et al., 2024). This method encodes spatial information as a graph linking spatially proximal cells or spots and uses an AE structure to learn a joint embedding space. To integrate RNA and ATAC data, separate GCN encoders combine the spatial graphs with modality-specific similarity graphs. These encodings are combined with an attention head to generate a single embedding space across both spatial modalities. In-house benchmarking on datasets with known anatomical regions showed good performance compared to non-spatial statistical or DL models. In agreement with ST vs. scRNA-seq data analysis, significant improvements in identifying spatial regions can be achieved by incorporating physical proximity, and DL models are more easily adapted to include this information than statistical methods.

3.10 RNA velocity

While scRNA-seq provides a snapshot of transcriptional states, RNA velocity methods have become increasingly valuable tools for investigating cell trajectories (Shima and mura, 2025; Bergen et al., 2021; Ge et al., 2025). Although new, several computational approaches now exist that leverage the relative abundances of spliced and unspliced mRNA to quantify transcriptional dynamics. Early ordinary differential equation (ODE)-based approaches like velocityto assumed specific cells were near steady-state, whereas scVelo relaxed this assumption through maximum-likelihood inference (La Manno et al., 2018; Bergen et al., 2020). More recent approaches incorporate additional molecular information, such as chromatin accessibility and protein expression, thereby refining trajectory inference and interpretability (Luo et al., 2025).

Recently, DL-based RNA velocity models have emerged to better capture nonlinear transcriptional dynamics and complex cellular transitions (Ge et al., 2025; Luo et al., 2025; Gayoso et al., 2024). VeloAE employs an autoencoder architecture to learn denoised, low-dimensional representations of RNA velocity (Qiao and Huang, 2021). VeloVAE and VeloVI employ VAE frameworks to infer RNA velocity and jointly quantify uncertainty (Gayoso et al., 2024; Gu et al., 2022). VeloVAE models a shared developmental timeline across all cells by learning latent time and cell-state representations, enabling explicit modelling of cell-fate branching and differentiation pathways. Conversely, VeloVI fits gene-specific dynamical models by leveraging information across cells, offering robust and reliable uncertainty estimates for RNA velocity at both gene and cell levels. DeepVelo integrates a graph convolutional network with a VAE to model gene- and cell-specific transcriptional kinetics, improving accuracy across heterogeneous cell populations (Chen Z. et al., 2022; Cui et al., 2023). LatentVelo and cellDancer both utilize neural architectures; LatentVelo embeds cell states and velocities into a latent space, while cellDancer employs gene-specific networks that aggregate local neighborhood information to infer cell- and gene-level kinetics (Li et al., 2024c; Farrell et al., 2023).

Regarding benchmarking, the accuracy and stability of these methods remain variable across datasets (Bergen et al., 2021; Luo et al., 2025; Gayoso et al., 2024; Gorin et al., 2022). Though deep learning approaches often perform better on complex datasets, no single method excels in both accuracy and stability (Shima and mura, 2025; Gayoso et al., 2024). Accuracy measures how closely predicted velocities align with known or expected biological trajectories. However, benchmarking remains limited due to limited ground truths, thus relying on indirect metrics based on velocity cosine similarity and agreement with known lineages (Bergen et al., 2021; Luo et al., 2025; Gayoso et al., 2024). Although most methods displayed locally consistent velocities between neighboring cells, most fail to reliably infer true cell-state transitions, particularly in complex or branching trajectories (Luo et al., 2025; Qiao and Huang, 2021; Gorin et al., 2022; Ancheta et al., 2024). In addition, discrepancies between methods remain common, primarily due to differences in model assumptions and datasets used (Bergen et al., 2021; Luo et al., 2025; Gayoso et al., 2024; Ancheta et al., 2024). Downsampling had the greatest impact on ground-truth recovery, while inter-method consistency remained stable. (Shima and mura, 2025; Luo et al., 2025; Ancheta et al., 2024). Notably, DeepVelo, scVelo, VeloVI, and velocityto often showed higher agreement among themselves, but none stood out in either accuracy or consistency across datasets.

In discovery contexts, current RNA velocity approaches should be interpreted cautiously when resolving complex cell-state transitions (Bergen et al., 2021; Gorin et al., 2022). Methods like VeloVI and LatentVelo offer higher accuracy and stability in specific contexts, but none are universally dependable (Luo et al., 2025; Gayoso et al., 2024). Using multiple RNA velocity methods in combination can mitigate individual biases, while integrating multi-omic or lineage-tracing datasets can help correct technical biases by providing more reliable validation (Shima and mura, 2025; Bergen et al., 2020; Mao et al., 2025). As the field of RNA velocity advances, deep learning methods will become more robust, capturing transcriptional kinetics from diverse datasets and reducing dependence on traditional ODE assumptions.

4 Conclusion

A plethora of algorithms and software packages have been produced using DL to solve many common problems in scRNA-seq and ST analysis. However, the performance of these models has been variable, with only the top models being competitive with state-of-the-art non-DL alternatives. There is no evidence that DL is inherently more accurate than non-DL algorithms, nor is it inherently more scalable when compared to optimized non-DL approaches. While DL can remove the linearity assumptions that constrain alternative approaches, there is little evidence that this provides a substantial benefit. The advantage of DL algorithms is their flexibility in handling a wide range of data types, which enables simple approaches for combining different data modalities, while graph-based models can be easily used to incorporate a spatial dimension. In addition, generative DL can enable novel approaches, mainly the prediction of one data modality from another, that are

not easily amenable to non-DL models. However, it remains to be proven that such algorithms can reach sufficient precision for their use in discovery research.

Author contributions

BT: Data curation, Visualization, Writing – original draft, Writing – review and editing. HN: Visualization, Writing – original draft, Writing – review and editing. SP: Writing – review and editing, Data curation. VS: Data curation, Writing – review and editing. TA: Conceptualization, Funding acquisition, Writing – original draft, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by NSERC Discovery grant (#03419-2023).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M. J. T., et al. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol.* 20 (1), 194. doi:10.1186/s13059-019-1795-z
- Amodio, M., van Dijk, D., Srinivasan, K., Chen, W. S., Mohsen, H., Moon, K. R., et al. (2019). Exploring single-cell data with deep multitasking neural networks. *Nat. Methods* 16 (11), 1139–1145. doi:10.1038/s41592-019-0576-7
- Ancheta, S., Dorman, L., Treut, G. L., Gurung, A., Royer, L. A., Granados, A., et al. (2024). Challenges and progress in RNA velocity: comparative analysis across multiple biological contexts. *Biorxiv* 29. doi:10.1101/2024.06.25.600667
- Andrews, T. S., and Hemberg, M. (2018). Identifying cell populations with scRNASeq. *Mol. Asp. Med.* 59, 114–122. doi:10.1016/j.mam.2017.07.002
- Andrews, T. S., and Hemberg, M. (2019). False signals induced by single-cell imputation. *F1000Research* 7, 1740. doi:10.12688/f1000research.16613.2
- Antonsson, S. E., and Melsted, P. (2024). Batch correction methods used in single cell RNA-sequencing analyses are often poorly calibrated. *Biorxiv* 21. doi:10.1101/2024.03.19.585562
- Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1
- Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., et al. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* 21 (1), 111. doi:10.1186/s13059-020-02015-1
- Arisdakessian, C., Poirion, O., Yunits, B., Zhu, X., and Garmire, L. X. (2019). DeepImpute: an accurate, fast, and scalable deep neural network method to impute single-cell RNA-seq data. *Genome Biol.* 20 (1), 211. doi:10.1186/s13059-019-1837-6
- Asuach, T., Gabitto, M. I., Koodli, R. V., Saldi, G. A., Jordan, M. I., and Yosef, N. (2023). MultiVI: deep generative model for the integration of multimodal data. *Nat. Methods* 20 (8), 1222–1231. doi:10.1038/s41592-023-01909-9
- Avsec, Z., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., et al. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* 18 (10), 1196–1203. doi:10.1038/s41592-021-01252-x
- Bae, S., Na, K. J., Koh, J., Lee, D. S., Choi, H., and Kim, Y. T. (2022). CellDART: cell type inference by domain adaptation of single-cell and spatial transcriptomic data. *Nucleic Acids Res.* 50 (10), e57. doi:10.1093/nar/gkac084
- Bergen, V., Lange, M., Peidli, S., Wolf, F. A., and Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* 38 (12), 1408–1414. doi:10.1038/s41587-020-0591-3
- Bergen, V., Soldatov, R. A., Kharchenko, P. V., and Theis, F. J. (2021). RNA velocity—current challenges and future perspectives. *Mol. Syst. Biol.* 17 (8), e10282. doi:10.15252/msb.202110282
- Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., et al. (2021). Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* 18 (11), 1352–1362. doi:10.1038/s41592-021-01264-7
- BinTayyash, N., Georgaka, S., John, S. T., Ahmed, S., Boukouvalas, A., Hensman, J., et al. (2021). Non-parametric modelling of temporal and spatial counts data from RNA-seq experiments. *Bioinformatics* 37 (21), 3788–3795. doi:10.1093/bioinformatics/btab486
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008 (10), P10008. doi:10.1088/1742-5468/2008/10/P10008
- Browaeys, R., Saelens, W., and Saey, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* 17 (2), 159–162. doi:10.1038/s41592-019-0667-5
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36 (5), 411–420. doi:10.1038/nbt.4096
- Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., and Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* 48 (10), e55. doi:10.1093/nar/gkaa183
- Cao, Z. J., and Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* 40 (10), 1458–1466. doi:10.1038/s41587-022-01284-4
- Cao, J., Cusanovich, D. A., Ramani, V., Aghamirzaie, D., Pliner, H. A., Hill, A. J., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 361 (6409), 1380–1385. doi:10.1126/science.aau0730
- Charytonowicz, D., Brody, R., and Sebra, R. (2023). Interpretable and context-free deconvolution of multi-scale whole transcriptomic data with UniCell deconvolve. *Nat. Commun.* 14 (1), 1350. doi:10.1038/s41467-023-36961-8

The author TA declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Generative AI was used to improve grammar, spelling, and wording of the text, as well as generating preliminary descriptions of AI architectures (section 2) which were manually refined.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Chazarra-Gil, R., van Dongen, S., Kiselev, V. Y., and Hemberg, M. (2021). Flexible comparison of batch correction methods for single-cell RNA-seq using BatchBench. *Nucleic Acids Res.* 49 (7), e42. doi:10.1093/nar/gkab004
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348 (6233), aaa6090. doi:10.1126/science.aaa6090
- Chen, J., Liu, W., Luo, T., Yu, Z., Jiang, M., Wen, J., et al. (2022a). A comprehensive comparison on celltype composition inference for spatial transcriptomics data. *Brief. Bioinform.* 23 (4), bbac245. doi:10.1093/bib/bbac245
- Chen, Z., King, W. C., Hwang, A., Gerstein, M., and Zhang, J. (2022b). DeepVelo: Single-cell transcriptomic deep velocity field learning with neural ordinary differential equations. *Sci. Adv.* 8 (48), eabq3745. doi:10.1126/sciadv.abq3745
- Chen, J., Xu, H., Tao, W., Chen, Z., Zhao, Y., and Han, J. D. J. (2023a). Transformer for one stop interpretable cell type annotation. *Nat. Commun.* 14 (1), 223. doi:10.1038/s41467-023-35923-4
- Chen, H., Lee, Y. J., Ovando, J. A., Rosas, L., Rojas, M., Mora, A. L., et al. (2023b). scResolve: recovering single cell expression profiles from multi-cellular spatial transcriptomics. *bioRxiv*, 2023.12.18.572269. doi:10.1101/2023.12.18.572269
- Chen, R. J., Ding, T., Lu, M. Y., Williamson, D. F. K., Jaume, G., Song, A. H., et al. (2024). Towards a general-purpose foundation model for computational pathology. *Nat. Med.* 30 (3), 850–862. doi:10.1038/s41591-024-02857-3
- Chen, M., Liu, J., Liang, G., Liu, Q., Li, S., and Yang, Y. (2025). Cross-species and cross-platform analysis reveals the application value of Guinea pig retina in myopia research at single-cell resolution. *Exp. Eye Res.* 259, 110558. doi:10.1016/j.exer.2025.110558
- Cheng, Y., Fan, X., Zhang, J., and Li, Y. (2023a). A scalable sparse neural network framework for rare cell type annotation of single-cell transcriptome data. *Commun. Biol.* 6 (1), 545. doi:10.1038/s42003-023-04928-6
- Cheng, Y., Ma, X., Yuan, L., Sun, Z., and Wang, P. (2023b). Evaluating imputation methods for single-cell RNA-seq data. *BMC Bioinforma.* 24 (1), 302. doi:10.1186/s12859-023-05417-7
- Chockalingam, S. P., Aluru, M., and Aluru, S. (2025). SCMENT: scalable and memory efficient integration of large-scale single-cell RNA-sequencing data. *Bioinformatics* 41 (2), btaf057. doi:10.1093/bioinformatics/btaf057
- Chu, T., Wang, Z., Peér, D., and Danko, C. G. (2022). Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat. Cancer* 3 (4), 505–517. doi:10.1038/s43018-022-00356-3
- Ciortan, M., and Defrance, M. (2022). GNN-based embedding for clustering scRNA-seq data. *Bioinformatics* 38 (4), 1037–1044. doi:10.1093/bioinformatics/btab787
- Clarke, Z. A., Andrews, T. S., Atif, J., Pouyababar, D., Innes, B. T., MacParland, S. A., et al. (2021). Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat. Protoc.* 16 (6), 2749–2764. doi:10.1038/s41596-021-00534-0
- Codino, A., Spagnoletti, L., Olobardi, C., Cuomo, A., Santos-Rosa, H., Palomba, M., et al. (2025). METTL9 sustains vertebrate neural development primarily via non-catalytic functions. *Nat. Commun.* 16 (1), 7051. doi:10.1038/s41467-025-62414-5
- Coleman, K., Hu, J., Schroeder, A., Lee, E. B., and Li, M. (2023). SpaDecon: celltype deconvolution in spatial transcriptomics with semi-supervised learning. *Commun. Biol.* 6 (1), 378. doi:10.1038/s42003-023-04761-x
- Cui, H., Maan, H., Taylor, M. D., and Wang, B. (2023). DeepVelo: deep learning extends RNA velocity to multi-lineage systems with cell-specific kinetics. *Biorxiv* 30. doi:10.1101/2022.04.03.486877
- Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., et al. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* 21 (8), 1470–1480. doi:10.1038/s41592-024-02201-0
- Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., Xiao, Y., et al. (2022). Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* 609 (7926), 375–383. doi:10.1038/s41586-022-05094-1
- Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Arxiv*. doi:10.48550/arXiv.1810.04805
- Dimitrov, D., Schäfer, P. S. L., Farr, E., Rodriguez-Mier, P., Lobentanzer, S., Badia-I-Mompel, P., et al. (2024). LIANA+ provides an all-in-one framework for cell–cell communication inference. *Nat. Cell Biol.* 26 (9), 1613–1622. doi:10.1038/s41556-024-01469-w
- Ding, J., Li, L., Lu, Q., Venegas, J., Wang, Y., Wu, L., et al. (2024). SpatialCTD: a large-scale tumor microenvironment spatial transcriptomic dataset to evaluate cell type deconvolution for immuno-oncology. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* 31 (9), 871–885. doi:10.1089/cmb.2024.0532
- Doersch, C. (2021). Tutorial on variational autoencoders. *Arxiv*. Available online at: <http://arxiv.org/abs/1606.05908> (Accessed July 11, 2023).
- Dong, R., and Yuan, G. C. (2021). SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol.* 22 (1), 145. doi:10.1186/s13059-021-02362-7
- Dong, K., and Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* 13 (1), 1739. doi:10.1038/s41467-022-29439-6
- Duan, Z., Riffle, D., Li, R., Liu, J., Min, M. R., and Zhang, J. (2024). Impeller: a path-based heterogeneous graph learning method for spatial transcriptomic data imputation. *Bioinformatics* 40 (6), btac339. doi:10.1093/bioinformatics/btac339
- D'Sa, K., Choi, M. L., Wagen, A. Z., Setó-Salvia, N., Kopach, O., Evans, J. R., et al. (2025). Astrocytic RNA editing regulates the host immune response to alpha-synuclein. *Sci. Adv.* 11 (15), eadp8504. doi:10.1126/sciadv.adp8504
- Efremova, M., Vento-Tormo, M., Teichmann, S. A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* 15 (4), 1484–1506. doi:10.1038/s41596-020-0292-x
- Eraslan, G., Avsec, Ž., Gagneur, J., and Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* 20 (7), 389–403. doi:10.1038/s41576-019-0122-6
- Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S., and Theis, F. J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* 10 (1), 390. doi:10.1038/s41467-018-07931-2
- Erfanian, N., Heydari, A. A., Feriz, A. M., Iañez, P., Derakhshani, A., Ghasemigol, M., et al. (2023). Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed. Pharmacother.* 165, 115077. doi:10.1016/j.biopha.2023.115077
- Ergen, C., Xing, G., Xu, C., Kim, M., Jayasuriya, M., McGeever, E., et al. (2024). Consensus prediction of cell type labels in single-cell data with popV. *Nat. Genet.* 56 (12), 2731–2738. doi:10.1038/s41588-024-01993-3
- Farrell, S., Mani, M., and Goyal, S. (2023). Inferring single-cell transcriptomic dynamics with structured latent gene expression dynamics. *Cell Rep. Methods* 3 (9), 100581. doi:10.1016/j.crmeth.2023.100581
- Fu, S., Wang, S., Si, D., Li, G., Gao, Y., and Liu, Q. (2025). Benchmarking single-cell multi-modal data integrations. *Nat. Methods* 22, 1–12. doi:10.1038/s41592-025-02737-9
- Gaspard-Boulinc, L. C., Gortana, L., Walter, T., Barillot, E., and Cavalli, F. M. G. (2025). Celltype deconvolution methods for spatial transcriptomics. *Nat. Rev. Genet.* 26, 1–19. doi:10.1038/s41576-025-00845-y
- Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., et al. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat. Methods* 18 (3), 272–282. doi:10.1038/s41592-020-01050-x
- Gayoso, A., Weiler, P., Lotfollahi, M., Klein, D., Hong, J., Streets, A., et al. (2024). Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells. *Nat. Methods* 21 (1), 50–59. doi:10.1038/s41592-023-01994-w
- Ge, S., Sun, S., Xu, H., Cheng, Q., and Ren, Z. (2024). Deep learning in single-cell and spatial transcriptomics data analysis: advances and challenges from a data science perspective. *Arxiv*. doi:10.48550/arXiv.2412.03614
- Ge, M., Miao, J., Qi, J., Zhou, X., and Lin, Z. (2025). TIVelo: RNA velocity estimation leveraging cluster-level trajectory inference. *Nat. Commun.* 16 (1), 6258. doi:10.1038/s41467-025-61628-x
- Gondara, L. (2016). “Medical image denoising using convolutional denoising autoencoders,” in *2016 IEEE 16th international conference on data mining workshops (ICDMW)*, IEEE: New York, NY. 241–246. doi:10.1109/ICDMW.2016.0041
- Gong, B., Zhou, Y., and Purdom, E. (2021). Cobolt: integrative analysis of multimodal single-cell sequencing data. *Genome Biol.* 22, 351. doi:10.1186/s13059-021-02556-z
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Arxiv*. doi:10.48550/arXiv.1406.2661
- Gorin, G., Fang, M., Chari, T., and Pachter, L. (2022). RNA velocity unraveled. *PLOS Comput. Biol.* 18 (9), e1010492. doi:10.1371/journal.pcbi.1010492
- Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., et al. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* 53 (3), 403–411. doi:10.1038/s41588-021-00790-6
- Greenwald, N. F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., et al. (2022). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* 40 (4), 555–565. doi:10.1038/s41587-021-01094-0
- Gu, Y., Blaauw, D., and Welch, J. D. (2022). Bayesian inference of RNA velocity from multi-lineage single-cell data. *Biorxiv*. doi:10.1101/2022.07.08.499381
- Guo, P., Mao, L., Chen, Y., Lee, C. N., Cardilla, A., Li, M., et al. (2025). Multiplexed spatial mapping of chromatin features, transcriptome and proteins in tissues. *Nat. Methods* 22 (3), 520–529. doi:10.1038/s41592-024-02576-0
- Haghverdi, L., Lun, A. T. L., Morgan, M. D., and Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36 (5), 421–427. doi:10.1038/nbt.4091
- Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X., et al. (2024a). Large-scale foundation model on single-cell transcriptomics. *Nat. Methods* 21 (8), 1481–1491. doi:10.1038/s41592-024-02305-7
- Hao, M., Bian, H., Yan, N., Chen, Y., Wei, L., and Zhang, X. (2024b). GeST: towards building a generative pretrained transformer for learning cellular spatial context.

Available online at: <https://openreview.net/forum?id=8e9KpZyksc> (Accessed August 26, 2025).

- Hao, M., Luo, E., Chen, Y., Wu, Y., Li, C., Chen, S., et al. (2024c). STEM enables mapping of single-cell and spatial transcriptomics data with transfer learning. *Commun. Biol.* 7 (1), 56. doi:10.1038/s42003-023-05640-1
- He, S., Bhatt, R., Brown, C., Brown, E. A., Buhr, D. L., Chantranuvattana, K., et al. (2022). High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* 40 (12), 1794–1806. doi:10.1038/s41587-022-01483-z
- Health, C. for D. (2025). *Artificial intelligence-enabled medical devices*. Silver Spring, Maryland: FDA. Available online at: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices> (Accessed September 9, 2025).
- Heimberg, G., Kuo, T., DePianto, D. J., Salem, O., Heigl, T., Diamant, N., et al. (2025). A cell atlas foundation model for scalable search of similar human cells. *Nature* 638 (8052), 1085–1094. doi:10.1038/s41586-024-08411-y
- Heumos, L., Schaar, A. C., Lance, C., Litnitskaya, A., Drost, F., Zappia, L., et al. (2023). Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* 24 (8), 550–572. doi:10.1038/s41576-023-00586-w
- Hie, B. L., Kim, S., Rando, T. A., Bryson, B., and Berger, B. (2024). Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nat. Protoc.* 19 (8), 2283–2297. doi:10.1038/s41596-024-00991-3
- Hou, W., and Ji, Z. (2024). Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. *Nat. Methods* 21 (8), 1462–1465. doi:10.1038/s41592-024-02235-4
- Hou, W., Ji, Z., Ji, H., and Hicks, S. C. (2020). A systematic evaluation of single-cell RNA-sequencing imputation methods. *Genome Biol.* 21 (1), 218. doi:10.1186/s13059-020-02132-x
- Hrovatin, K., Moinfar, A. A., Zappia, L., Lapuerta, A. T., Lengerich, B., Kellis, M., et al. (2024). Integrating single-cell RNA-seq datasets with substantial batch effects. *Biorxiv* 10, 2023.11.03.565463. doi:10.1101/2023.11.03.565463
- Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D. J., et al. (2021). SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* 18 (11), 1342–1351. doi:10.1038/s41592-021-01255-8
- Hu, Y., Xie, M., Li, Y., Rao, M., Shen, W., Luo, C., et al. (2024a). Benchmarking clustering, alignment, and integration methods for spatial transcriptomics. *Genome Biol.* 25 (1), 212. doi:10.1186/s13059-024-03361-0
- Hu, Y., Wan, S., Luo, Y., Li, Y., Wu, T., Deng, W., et al. (2024b). Benchmarking algorithms for single-cell multi-omics prediction and integration. *Nat. Methods* 21 (11), 2182–2194. doi:10.1038/s41592-024-02429-w
- Huang, Z., Wang, J., Lu, X., Mohd Zain, A., and Yu, G. (2023). scGGAN: single-cell RNA-seq imputation by graph-based generative adversarial network. *Brief. Bioinform* 24 (2), bbad040. doi:10.1093/bib/bbad040
- Huang, J., Chow, A. C. M., Tang, N. L. S., and Yam, S. C. P. (2025). An in-depth benchmark framework for evaluating single cell RNA-seq dropout imputation methods and the development of an improved algorithm afMF. *Clin. Transl. Med.* 15 (4), e70283. doi:10.1002/ctm2.70283
- Im, Y., and Kim, Y. (2023). A comprehensive overview of RNA deconvolution methods and their application. *Mol. Cells* 46 (2), 99–105. doi:10.14348/molcells.2023.2178
- Jia, C., Wang, H., Zhao, J., Xia, J., and Zheng, C. (2025). scSDNE: a semi-supervised method for inferring cell-cell interactions based on graph embedding. *PLOS Comput. Biol.* 21 (5), e1013027. doi:10.1371/journal.pcbi.1013027
- Jiang, F., Zhou, X., Qian, Y., Zhu, M., Wang, L., Li, Z., et al. (2023). Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nat. Methods* 20 (7), 1048–1057. doi:10.1038/s41592-023-01884-1
- Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., et al. (2021). Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* 12 (1), 1088. doi:10.1038/s41467-021-21246-9
- Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8 (1), 118–127. doi:10.1093/biostatistics/kxj037
- Karin, J., Mintz, R., Raveh, B., and Nitzan, M. (2024). Interpreting single-cell and spatial omics data using deep neural network training dynamics. *Nat. Comput. Sci.* 4 (12), 941–954. doi:10.1038/s43588-024-00721-5
- Khan, M., Arslanturk, S., and Draghi, S. (2025). A comprehensive review of spatial transcriptomics data alignment and integration. *Nucleic Acids Res.* 53 (12), gkaf536. doi:10.1093/nar/gkaf536
- Kingma, D. P., and Welling, M. (2019). An introduction to variational autoencoders. *Found. Trends Mach. Learn.* 12 (4), 307–392. doi:10.1561/22000000056
- Kingma, D. P., and Welling, M. (2022). Auto-encoding variational bayes. *Arxiv*. doi:10.48550/arXiv.1312.6114
- Kiselev, V. Y., Yiu, A., and Hemberg, M. (2018). Scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* 15 (5), 359–362. doi:10.1038/nmeth.4644
- Kiselev, V. Y., Andrews, T. S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* 20 (5), 273–282. doi:10.1038/s41576-018-0088-9
- Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H. W., Li, T., et al. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* 40 (5), 661–671. doi:10.1038/s41587-021-01139-4
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019). Fast, sensitive, and accurate integration of single cell data with Harmony. *Nat. Methods* 16 (12), 1289–1296. doi:10.1038/s41592-019-0619-0
- Kuntz, S., Kriehoff-Henning, E., Kather, J. N., Jutzi, T., Höhn, J., Kiehl, L., et al. (2021). Gastrointestinal cancer classification and prognostication from histology using deep learning: systematic review. *Eur. J. Cancer* 155, 200–215. doi:10.1016/j.ejca.2021.07.012
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., et al. (2018). RNA velocity of single cells. *Nature* 560 (7719), 494–498. doi:10.1038/s41586-018-0414-6
- Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21 (1), 31. doi:10.1186/s13059-020-1926-6
- Lal, A., Chiang, Z. D., Yakovenko, N., Duarte, F. M., Israeli, J., and Buenrostro, J. D. (2021). Deep learning-based enhancement of epigenomics data with AtacWorks. *Nat. Commun.* 12 (1), 1507. doi:10.1038/s41467-021-21765-5
- Lecun, Y., and Bengio, Y. (1998). “Convolutional networks for images, speech, and time series,” in *The handbook of brain theory and neural networks*. Cambridge, MA: MIT Press, 255–258. doi:10.5555/303568.303704
- Lee, M. Y. Y., Kaestner, K. H., and Li, M. (2023). Benchmarking algorithms for joint integration of unpaired and paired single-cell RNA-seq and ATAC-seq data. *Genome Biol.* 24 (1), 244. doi:10.1186/s13059-023-03073-x
- Levine, D., Rizvi, S. A., Lévy, S., Pallikkavaliyeetil, N., Zhang, D., Chen, X., et al. (2024). Cell2Sentence: teaching large Language models the Language of biology. *bioRxiv*, 2023.09.11.557287. doi:10.1101/2023.09.11.557287
- Li, W. V., and Li, J. J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* 9 (1), 997. doi:10.1038/s41467-018-03405-7
- Li, Y., and Luo, Y. (2024). STdGCN: spatial transcriptomic celltype deconvolution using graph convolutional networks. *Genome Biol.* 25 (1), 206. doi:10.1186/s13059-024-03353-0
- Li, R., and Quon, G. (2019). scBFA: modeling detection patterns to mitigate technical noise in large-scale single-cell genomics data. *Genome Biol.* 20 (1), 193. doi:10.1186/s13059-019-1806-0
- Li, Z., and Zhou, X. (2022). BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. *Genome Biol.* 23 (1), 168. doi:10.1186/s13059-022-02734-7
- Li, Y., Stanojevic, S., and Garmire, L. X. (2022a). Emerging artificial intelligence applications in Spatial Transcriptomics analysis. *Comput. Struct. Biotechnol. J.* 20, 2895–2908. doi:10.1016/j.csbj.2022.05.056
- Li, H., Li, H., Zhou, J., and Gao, X. (2022b). SD2: spatially resolved transcriptomics deconvolution through integration of dropout and spatial information. *Bioinforma. Oxf. Engl.* 38 (21), 4878–4884. doi:10.1093/bioinformatics/btac605
- Li, B., Zhang, W., Guo, C., Xu, H., Li, L., Fang, M., et al. (2022c). Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* 19 (6), 662–670. doi:10.1038/s41592-022-01480-9
- Li, G., Fu, S., Wang, S., Zhu, C., Duan, B., Tang, C., et al. (2022d). A deep generative model for multi-view profiling of single-cell RNA-seq and ATAC-seq data. *Genome Biol.* 23, 20. doi:10.1186/s13059-021-02595-6
- Li, H., Zhou, J., Li, Z., Chen, S., Liao, X., Zhang, B., et al. (2023). A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nat. Commun.* 14 (1), 1548. doi:10.1038/s41467-023-37168-7
- Li, B., Karami, M., Junayed, M. S., and Nabavi, S. (2024a). Multi-modal spatial clustering for spatial transcriptomics utilizing high-resolution histology images. *Arxiv*, 3469–3474. doi:10.48550/arXiv.2411.02534
- Li, S., Gai, K., Dong, K., Zhang, Y., and Zhang, S. (2024b). High-density generation of spatial transcriptomics with STAGE. *Nucleic Acids Res.* 52 (9), 4843–4856. doi:10.1093/nar/gkac294
- Li, S., Zhang, P., Chen, W., Ye, L., Brannan, K. W., Le, N. T., et al. (2024c). A relay velocity model infers cell-dependent RNA velocity. *Nat. Biotechnol.* 42 (1), 99–108. doi:10.1038/s41587-023-01728-5
- Liang, X., Cao, L., Chen, H., Wang, L., Wang, Y., Fu, L., et al. (2024). A critical assessment of clustering algorithms to improve cell clustering and identification in single-cell transcriptome study. *Brief. Bioinform* 25 (1), bbad497. doi:10.1093/bib/bbad497
- Liao, J., Qian, J., Fang, Y., Chen, Z., Zhuang, X., Zhang, N., et al. (2022). De novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution. *Nat. Commun.* 13 (1), 6498. doi:10.1038/s41467-022-34271-z

- Lin, Y., Ghazanfar, S., Wang, K. Y. X., Gagnon-Bartsch, J. A., Lo, K. K., Su, X., et al. (2019). scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. *Proc. Natl. Acad. Sci. U. S. A.* 116 (20), 9775–9784. doi:10.1073/pnas.1820061116
- Lin, Y., Li, H., Xiao, X., Zhang, L., Wang, K., Zhao, J., et al. (2022). DAISM-DNNXMBD: highly accurate cell type proportion estimation with *in silico* data augmentation and deep neural networks. *Patterns* 3 (3), 100440. doi:10.1016/j.patter.2022.100440
- Lindeboom, R. G. H., Worlock, K. B., Dratva, L. M., Yoshida, M., Scobie, D., Wagstaffe, H. R., et al. (2024). Human SARS-CoV-2 challenge uncovers local and systemic response dynamics. *Nature* 631 (8019), 189–198. doi:10.1038/s41586-024-07575-x
- Liu, C., Ding, S., Kim, H. J., Long, S., Xiao, D., Ghazanfar, S., et al. (2025). Multitask benchmarking of single-cell multimodal omics integration methods. *Nat. Methods* 22, 1–12. doi:10.1038/s41592-025-02856-3
- Lomas Redondo, A., Sánchez Velázquez, J. M., García, T. Á. J., and Sánchez-Arévalo Lobo, V. J. (2025). Deep learning based deconvolution methods: a systematic review. *Comput. Struct. Biotechnol. J.* 27, 2544–2565. doi:10.1016/j.csbj.2025.05.038
- Long, Y., Ang, K. S., Li, M., Chong, K. L. K., Sethi, R., Zhong, C., et al. (2023). Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* 14 (1), 1155. doi:10.1038/s41467-023-36796-3
- Long, Y., Ang, K. S., Sethi, R., Liao, S., Heng, Y., van Olst, L., et al. (2024). Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nat. Methods* 21 (9), 1658–1667. doi:10.1038/s41592-024-02316-4
- Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat. Methods* 15 (12), 1053–1058. doi:10.1038/s41592-018-0229-2
- Lotfollahi, M., Wolf, F. A., and Theis, F. J. (2019). scGen predicts single-cell perturbation responses. *Nat. Methods* 16 (8), 715–721. doi:10.1038/s41592-019-0494-8
- Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Büttner, M., Wagenstetter, M., et al. (2022). Mapping single-cell data to reference atlases by transfer learning. *Nat. Biotechnol.* 40 (1), 121–130. doi:10.1038/s41587-021-01001-7
- Luecken, M. D., and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15 (6), e8746. doi:10.15252/msb.20188746
- Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* 19 (1), 41–50. doi:10.1038/s41592-021-01336-8
- Lund, J. B., Lindberg, E. L., Maatz, H., Pottbaecker, F., Hübner, N., and Lippert, C. (2022). AntiSpodge: a neural-network-based RNA-profile deconvolution pipeline designed for spatial transcriptomics. *Nar. Genomics Bioinforma.* 4 (4), lqac073. doi:10.1093/nargab/lqac073
- Luo, Y., Ren, J., Yang, Q., Zhou, Y., You, Z., and Li, Q. (2025). Benchmarking RNA velocity methods across 17 independent studies. *bioRxiv*. doi:10.1101/2025.08.02.668272
- Luo, J., Fu, J., Lu, Z., and Tu, J. (2024). Deep learning in integrating spatial transcriptomics with other modalities. *Brief. Bioinform.* 26 (1), bbaf719. doi:10.1093/bib/bbae719
- Ma, Q., and Xu, D. (2022). Deep learning shapes single-cell data analysis. *Nat. Rev. Mol. Cell Biol.* 23 (5), 303–304. doi:10.1038/s41580-022-00466-x
- Mañanes, D., Rivero-García, I., Relano, C., Torres, M., Sancho, D., Jimenez-Carretero, D., et al. (2024). SpatialDDL: an R package to deconvolute spatial transcriptomics data using neural networks. *Bioinformatics* 40 (2), btac072. doi:10.1093/bioinformatics/btac072
- Mao, S., Zhang, C., Chen, R., Tang, S., Fan, X., and Hu, J. (2025). Cell lineage tracing: methods, applications, and challenges. *Quant. Biol.* 13 (4), e70006. doi:10.1002/qub.270006
- Marouf, M., Machart, P., Bansal, V., Kilian, C., Magruder, D. S., Krebs, C. F., et al. (2020). Realistic *in silico* generation and augmentation of single-cell RNA-seq data using generative adversarial networks. *Nat. Commun.* 11 (1), 166. doi:10.1038/s41467-019-14018-z
- Menden, K., Marouf, M., Oller, S., Dalmia, A., Magruder, D. S., Kloiber, K., et al. (2020). Deep learning-based cell composition analysis from tissue expression profiles. *Sci. Adv.* 6 (30), eaba2619. doi:10.1126/sciadv.aba2619
- Mimitou, E. P., Lareau, C. A., Chen, K. Y., Zorzetto-Fernandes, A. L., Hao, Y., Takeshima, Y., et al. (2021). Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* 39 (10), 1246–1258. doi:10.1038/s41587-021-00927-2
- Mitchel, J., Gao, T., Cole, E., Petukhov, V., and Kharchenko, P. V. (2025). Impact of segmentation errors in analysis of spatial transcriptomics data. *Biorxiv* 3. doi:10.1101/2025.01.02.631135
- Molho, D., Ding, J., Tang, W., Li, Z., Wen, H., Wang, Y., et al. (2024). Deep learning in single-cell analysis. *ACM Trans. Intell. Syst. Technol.* 15 (3), 40:1–40:62. doi:10.1145/3641284
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337
- Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37 (7), 773–782. doi:10.1038/s41587-019-0114-2
- Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., et al. (2022). Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* 19 (2), 171–178. doi:10.1038/s41592-021-01358-2
- Peng, X. L., Moffitt, R. A., Torphy, R. J., Volmar, K. E., and Yeh, J. J. (2019). De novo compartment deconvolution and weight estimation of tumor samples using DECODER. *Nat. Commun.* 10 (1), 4729. doi:10.1038/s41467-019-12517-7
- Pham, D., Tan, X., Balderson, B., Xu, J., Grice, L. F., Yoon, S., et al. (2023). Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. *Nat. Commun.* 14 (1), 7739. doi:10.1038/s41467-023-43120-6
- Polański, K., Bartolomé-Casado, R., Sarropoulos, I., Xu, C., England, N., Jahnsen, F. L., et al. (2024). Bin2cell reconstructs cells from high resolution Visium HD data. *Bioinformatics* 40, btac546. doi:10.1093/bioinformatics/btac546
- Pratama, R., Hilton, J., Cherry, J. M., and Song, G. (2025). Gene spatial integration: enhancing spatial transcriptomics analysis via deep learning and batch effect mitigation. *Bioinformatics* 41 (6), btac350. doi:10.1093/bioinformatics/btac350
- Qi, C., Fang, H., Hu, T., Jiang, S., and Zhi, W. (2025). Bidirectional Mamba for single-cell data: efficient context learning with biological fidelity. *Arxiv*. doi:10.48550/arXiv.2504.16956
- Qiao, C., and Huang, Y. (2021). Representation learning of RNA velocity reveals robust cell transitions. *Proc. Natl. Acad. Sci. U. S. A.* 118 (49), e2105859118. doi:10.1073/pnas.2105859118
- Reyes, M., Billman, K., Hacohen, N., and Blainey, P. C. (2019). Simultaneous profiling of gene expression and chromatin accessibility in single cells. *Adv. Biosyst.* 3 (11), 1900065. doi:10.1002/adbi.201900065
- Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. *Arxiv*. doi:10.48550/arXiv.1401.4082
- Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* 118 (15), e2016239118. doi:10.1073/pnas.2016239118
- Rodrigues, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., et al. (2019). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363 (6434), 1463–1467. doi:10.1126/science.aaw1219
- Salcher, S., Sturm, G., Horvath, L., Untergasser, G., Kuempers, C., Fotakis, G., et al. (2022). High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. *Cancer Cell* 40 (12), 1503–1520.e8. doi:10.1016/j.ccell.2022.10.008
- Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *Sn Comput. Sci.* 2 (6), 420. doi:10.1007/s42979-021-00815-1
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression. *Nat. Biotechnol.* 33 (5), 495–502. doi:10.1038/nbt.3192
- Schott, M., León-Periñán, D., Splendiani, E., Strenger, L., Licha, J. R., Pentimalli, T. M., et al. (2024). Open-ST: High-resolution spatial transcriptomics in 3D. *Cell* 187 (15), 3953–3972.e26. doi:10.1016/j.cell.2024.05.055
- Shao, X., Yang, H., Zhuang, X., Liao, J., Yang, P., Cheng, J., et al. (2021). scDeepSort: a pre-trained celltype annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network. *Nucleic Acids Res.* 49 (21), e122. doi:10.1093/nar/gkab775
- Shen, H., Shen, X., Feng, M., Wu, D., Zhang, C., Yang, Y., et al. (2022). A universal approach for integrating super large-scale single-cell transcriptomes by exploring gene rankings. *Brief. Bioinform.* 23 (2), bbab573. doi:10.1093/bib/bbab573
- Shen, H., Liu, J., Hu, J., Shen, X., Zhang, C., Wu, D., et al. (2023). Generative pretraining from large-scale transcriptomes for single-cell deciphering. *iScience* 26 (5), 106536. doi:10.1016/j.isci.2023.106536
- Shi, Y., Wan, J., Zhang, X., and Yin, Y. (2023). CL-Impute: a contrastive learning-based imputation for dropout single-cell RNA-seq data. *Comput. Biol. Med.* 164, 107263. doi:10.1016/j.combiomed.2023.107263
- Shimamura, T. (2025). RNA velocity and beyond: current advances in modeling single-cell transcriptional dynamics. *Allergol. Int.* 74 (4), 525–533. doi:10.1016/j.alit.2025.08.005
- Simonovsky, M., and Komodakis, N. (2018). GraphVAE: towards generation of small graphs using variational autoencoders. *Arxiv*. doi:10.48550/arXiv.1802.03480
- So, E., Hayat, S., Nair, S. K., Wang, B., and Haibe-Kains, B. (2025). GraphComm predicts cell cell communication using a graph based deep learning method in single cell RNA sequencing data. *Sci. Rep.* 15 (1), 36914. doi:10.1038/s41598-025-20812-1

- Song, Q., and Su, J. (2021). DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Brief. Bioinform.* 22 (5), bbaa414. doi:10.1093/bib/bbaa414
- Song, Q., Su, J., and Zhang, W. (2021). scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. *Nat. Commun.* 12 (1), 3826. doi:10.1038/s41467-021-24172-y
- Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353 (6294), 78–82. doi:10.1126/science.aaf2403
- Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., et al. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14 (9), 865–868. doi:10.1038/nmeth.4380
- Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* 18 (1), 100–106. doi:10.1038/s41592-020-01018-x
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., et al. (2019). Comprehensive integration of single-cell data. *Cell* 177 (7), 1888–1902.e21. doi:10.1016/j.cell.2019.05.031
- Stuart, T., Srivastava, A., Madad, S., Lareau, C. A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. *Nat. Methods* 18 (11), 1333–1341. doi:10.1038/s41592-021-01282-5
- Su, H., Xing, F., Kong, X., Xie, Y., Zhang, S., and Yang, L. (2015). “Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders,” in *Medical image computing and computer-assisted intervention – MICCAI 2015*. Editors N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Springer, Cham: Springer), 383–390. doi:10.1007/978-3-319-24574-4_46
- Svensson, V. (2020). Droplet scRNA-seq is not zero-inflated. *Nat. Biotechnol.* 38 (2), 147–150. doi:10.1038/s41587-019-0379-5
- Svensson, V., Natarajan, K. N., Ly, L. H., Miragaia, R. J., Labalette, C., Macaulay, I. C., et al. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14 (4), 381–387. doi:10.1038/nmeth.4220
- Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13 (4), 599–604. doi:10.1038/nprot.2017.149
- Tan, X., Su, A., Tran, M., and Nguyen, Q. (2020). SpaCell: integrating tissue morphology and spatial gene expression to predict disease cells. *Bioinformatics* 36 (7), 2293–2294. doi:10.1093/bioinformatics/btz914
- Tang, Z., Li, Z., Hou, T., Zhang, T., Yang, B., Su, J., et al. (2023). SiGra: single-cell spatial elucidation through an image-augmented graph transformer. *Nat. Commun.* 14 (1), 5618. doi:10.1038/s41467-023-41437-w
- Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., et al. (2023). Transfer learning enables predictions in network biology. *Nature* 618 (7965), 616–624. doi:10.1038/s41586-023-06139-9
- Tian, T., Wan, J., Song, Q., and Wei, Z. (2019). Clustering single-cell RNA-seq data with a model-based deep learning approach. *Nat. Mach. Intell.* 1 (4), 191–198. doi:10.1038/s42256-019-0037-0
- Tian, T., Zhang, J., Lin, X., Wei, Z., and Hakonarson, H. (2021). Model-based deep embedding for constrained clustering analysis of single cell RNA-seq data. *Nat. Commun.* 12 (1), 1873. doi:10.1038/s41467-021-22008-3
- Tian, T., Zhang, J., Lin, X., Wei, Z., and Hakonarson, H. (2024). Dependency-aware deep generative models for multitasking analysis of spatial omics data. *Nat. Methods* 21 (8), 1501–1513. doi:10.1038/s41592-024-02257-y
- Traag, V. A., Waltman, L., and van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9 (1), 5233. doi:10.1038/s41598-019-41695-z
- Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M., et al. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* 21 (1), 12. doi:10.1186/s13059-019-1850-9
- Tran, K. A., Addala, V., Johnston, R. L., Lovell, D., Bradley, A., Koufariotis, L. T., et al. (2023). Performance of tumour microenvironment deconvolution methods in breast cancer using single-cell simulated bulk mixtures. *Nat. Commun.* 14 (1), 5758. doi:10.1038/s41467-023-41385-5
- van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174 (3), 716–729.e27. doi:10.1016/j.cell.2018.05.061
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2023). Attention is all you need. *Arxiv*. doi:10.48550/arXiv.1706.03762
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph attention networks. *6th International Conference on Learning Representations*. Appleton, WI: ICLR 2018. doi:10.48550/arXiv.1710.10903
- Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P. A. (2008). “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on machine learning - ICML '08* (New York, NY: ACM Press), 1096–1103. doi:10.1145/1390156.1390294
- Wan, X., Xiao, J., Tam, S. S. T., Cai, M., Sugimura, R., Wang, Y., et al. (2023). Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat. Commun.* 14 (1), 7848. doi:10.1038/s41467-023-43629-w
- Wang, X., Park, J., Susztak, K., Zhang, N. R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* 10 (1), 380. doi:10.1038/s41467-018-08023-x
- Wang, D., Hou, S., Zhang, L., Wang, X., Liu, B., and Zhang, Z. (2021a). iMAP: integration of multiple single-cell datasets by adversarial paired transfer networks. *Genome Biol.* 22 (1), 63. doi:10.1186/s13059-021-02280-8
- Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., et al. (2021b). scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat. Commun.* 12 (1), 1882. doi:10.1038/s41467-021-22197-x
- Wang, J., Fonseca, G. J., and Ding, J. (2024). scSemiProfiler: advancing large-scale single-cell studies through semi-profiling with deep generative models and active learning. *Nat. Commun.* 15 (1), 5989. doi:10.1038/s41467-024-50150-1
- Wang, C., Chan, A. S., Fu, X., Ghazanfar, S., Kim, J., Patrick, E., et al. (2025). Benchmarking the translational potential of spatial gene expression prediction from histology. *Nat. Commun.* 16 (1), 1544. doi:10.1038/s41467-025-56618-y
- Wani, S. A., Khan, S. A., and Quadri, S. (2025). Application of deep learning for single cell Multi-Omics: a state-of-the-art review. *Arch. Comput. Methods Eng.* 32 (5), 2987–3029. doi:10.1007/s11831-025-10230-x
- Warren, S. L., and Moustafa, A. A. (2023). Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: a systematic review. *J. Neuroimaging* 33 (1), 5–18. doi:10.1111/jon.13063
- Webel, H., Niu, L., Nielsen, A. B., Locard-Paulet, M., Mann, M., Jensen, L. J., et al. (2024). Imputation of label-free quantitative mass spectrometry-based proteomics data using self-supervised deep learning. *Nat. Commun.* 15 (1), 5405. doi:10.1038/s41467-024-48711-5
- Wei, R., He, S., Bai, S., Sei, E., Hu, M., Thompson, A., et al. (2022). Spatial charting of single-cell transcriptomes in tissues. *Nat. Biotechnol.* 40 (8), 1190–1199. doi:10.1038/s41587-022-01233-1
- Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., et al. (2023). Transformers in time series: a survey. *Arxiv*. doi:10.48550/arXiv.2202.07125
- Wen, H., Tang, W., Jin, W., Ding, J., Liu, R., Dai, X., et al. (2024). Single cells are spatial tokens: transformers for spatial transcriptomic data imputation. *Arxiv*. doi:10.48550/arXiv.2302.03038
- White, B. S., de Reyniès, A., Newman, A. M., Waterfall, J. J., Lamb, A., Petitprez, F., et al. (2024). Community assessment of methods to deconvolve cellular composition from bulk gene expression. *Nat. Commun.* 15 (1), 7362. doi:10.1038/s41467-024-50618-0
- Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., and Haque, A. (2022). An introduction to spatial transcriptomics for biomedical research. *Genome Med.* 14 (1), 68. doi:10.1186/s13073-022-01075-1
- Wolf, F. A., Angerer, P., and Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19 (1), 15. doi:10.1186/s13059-017-1382-0
- Wolfram-Schauerte, M., Vogel, T., Tuoken, H., Fäth Savitski, M., Simon, E., and Nieselt, K. (2025). Approaching the holistic transcriptome—convolution and deconvolution in transcriptomics. *Brief. Bioinform.* 26 (4), bbaf388. doi:10.1093/bib/bbaf388
- Wu, T., Wang, Y., and Quach, N. (2025). Advancements in natural language processing: exploring transformer-based architectures for text understanding. *Arxiv*. doi:10.48550/arXiv.2503.20227
- Xiao, C., Chen, Y., Meng, Q., Wei, L., and Zhang, X. (2024). Benchmarking multi-omics integration algorithms across single-cell RNA and ATAC data. *Brief. Bioinform.* 25 (2), bbae095. doi:10.1093/bib/bbae095
- Xiong, J., Liu, G., Huang, L., Wu, C., Wu, T., Mu, Y., et al. (2025). Autoregressive Models in Vision: a Survey. *Arxiv*. doi:10.48550/arXiv.2411.05902
- Xu, Y., Zhang, Z., You, L., Liu, J., Fan, Z., and Zhou, X. (2020). scIGANs: single-cell RNA-seq imputation using generative adversarial networks. *Nucleic Acids Res.* 48 (15), e85. doi:10.1093/nar/gkaa506
- Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M. I., and Yosef, N. (2021). Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* 17 (1), e9620. doi:10.15252/msb.20209620
- Xu, C., Jin, X., Wei, S., Wang, P., Luo, M., Xu, Z., et al. (2022). DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res.* 50 (22), e131. doi:10.1093/nar/gkac901
- Xu, P., Zhu, X., and Clifton, D. A. (2023a). Multimodal Learning with Transformers: a Survey. *Arxiv* 45, 12113–12132. doi:10.48550/arXiv.2206.06488
- Xu, H., Wang, S., Fang, M., Luo, S., Chen, C., Wan, S., et al. (2023b). SPACEL: deep learning-based characterization of spatial transcriptome architectures. *Nat. Commun.* 14 (1), 7603. doi:10.1038/s41467-023-43220-3
- Xue, Z., Wu, L., Tian, R., Gao, B., Zhao, Y., He, B., et al. (2025). Integrative mapping of human CD8+ T cells in inflammation and cancer. *Nat. Methods* 22 (2), 435–445. doi:10.1038/s41592-024-02530-0

- Yan, L., and Sun, X. (2023). Benchmarking and integration of methods for deconvoluting spatial transcriptomic data. *Bioinformatics* 39 (1), btac805. doi:10.1093/bioinformatics/btac805
- Yang, F., Wang, W., Wang, F., Fang, Y., Tang, D., Huang, J., et al. (2022). scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat. Mach. Intell.* 4 (10), 852–866. doi:10.1038/s42256-022-00534-z
- Yang, W., Wang, P., Luo, M., Cai, Y., Xu, C., Xue, G., et al. (2023a). DeepCCI: a deep learning framework for identifying cell–cell interactions from single-cell RNA sequencing data. *Bioinformatics* 39 (10), btad596. doi:10.1093/bioinformatics/btad596
- Yang, Y., Li, G., Zhong, Y., Xu, Q., Lin, Y. T., Roman-Vicharra, C., et al. (2023b). scTenifoldXct: a semi-supervised method for predicting cell–cell interactions and mapping cellular communication graphs. *Cell Syst.* 14 (4), 302–311.e4. doi:10.1016/j.cels.2023.01.004
- Yang, L. X., Qi, C., Lu, S., Ye, X. S., Merikhian, P., Zhang, D. Y., et al. (2025a). Alleviation of liver fibrosis by inhibiting a non-canonical ATF4-regulated enhancer program in hepatic stellate cells. *Nat. Commun.* 16 (1), 524. doi:10.1038/s41467-024-55738-1
- Yang, P., Jin, K., Yao, Y., Jin, L., Shao, X., Li, C., et al. (2025b). Spatial integration of multi-omics single-cell data with SIMO. *Nat. Commun.* 16 (1), 1265. doi:10.1038/s41467-025-56523-4
- Yenduri, G., Ramalingam, M., Chemmalar Selvi, G., Supriya, Y., Gautam, S., Praveen Kumar, R. M., et al. (2023). Generative pre-trained transformer: a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *Arxiv*. doi:10.48550/arXiv.2305.10435
- Yin, W., Wan, Y., and Zhou, Y. (2024). SpatialcoGCN: deconvolution and spatial information-aware simulation of spatial transcriptomics data via deep graph co-embedding. *Brief. Bioinform* 25 (3), bbae130. doi:10.1093/bib/bbae130
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612
- Young, M. D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* 9 (12), giaa151. doi:10.1093/gigascience/giaa151
- Yuan, Y., and Bar-Joseph, Z. (2019). Deep learning for inferring gene relationships from single-cell expression data. *Proc. Natl. Acad. Sci. U. S. A.* 116, 27151–27158. doi:10.1073/pnas.1911536116
- Yuan, Z., Zhao, F., Lin, S., Zhao, Y., Yao, J., Cui, Y., et al. (2024). Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nat. Methods* 21 (4), 712–722. doi:10.1038/s41592-024-02215-8
- Zahedi, R., Ghamsari, R., Argha, A., Macphillamy, C., Beheshti, A., Alizadehsani, R., et al. (2024). Deep learning in spatially resolved transcriptomics: a comprehensive technical view. *Brief. Bioinform* 25 (2), bbae082. doi:10.1093/bib/bbae082
- Zeng, Y., Song, Y., Zhang, C., Li, H., Zhao, Y., Yu, W., et al. (2024). Imputing spatial transcriptomics through gene network constructed from protein language model. *Commun. Biol.* 7 (1), 1271. doi:10.1038/s42003-024-06964-2
- Zeng, Y., Xie, J., Shangguan, N., Wei, Z., Li, W., Su, Y., et al. (2025). CellFM: a large-scale foundation model pre-trained on transcriptomics of 100 million human cells. *Nat. Commun.* 16 (1), 4679. doi:10.1038/s41467-025-59926-5
- Zhan, Y., Zhang, Y., Hu, Z., Wang, Y., Zhu, Z., Du, S., et al. (2025). LETSmix: a spatially informed and learning-based domain adaptation method for celltype deconvolution in spatial transcriptomics. *Genome Med.* 17 (1), 16. doi:10.1186/s13073-025-01442-8
- Zhang, T., Zhang, Z., Li, L., Dong, B., Wang, G., and Zhang, D. (2023). GTAD: a graph-based approach for cell spatial composition inference from integrated scRNA-seq and ST-seq data. *Brief. Bioinform* 25 (1), bbad469. doi:10.1093/bib/bbad469
- Zhang, C., Liu, L., Zhang, Y., Li, M., Fang, S., Kang, Q., et al. (2024a). spatiAlign: an unsupervised contrastive learning model for data integration of spatially resolved transcriptomics. *GigaScience* 13, giae042. doi:10.1093/gigascience/giae042
- Zhang, W., Huckaby, B., Talburt, J., Weissman, S., and Yang, M. Q. (2024b). cnnImpute: missing value recovery for single cell RNA sequencing data. *Sci. Rep.* 14 (1), 3946. doi:10.1038/s41598-024-53998-x
- Zhao, E., Stone, M. R., Ren, X., Guenthoer, J., Smythe, K. S., Pulliam, T., et al. (2021). Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* 39 (11), 1375–1384. doi:10.1038/s41587-021-00935-2
- Zhao, P., Zhu, J., Ma, Y., and Zhou, X. (2022). Modeling zero inflation is not necessary for spatial transcriptomics. *Genome Biol.* 23 (1), 118. doi:10.1186/s13059-022-02684-0