SpotDiff: Spatial Gene Expression Imputation Diffusion with Single-Cell RNA Sequencing Data Integration

Tianyi Chen[1], Yunfei Zhang[2], Lianxin Xie[2], Wenjun Shen[3], Si Wu[2*], Hau-San Wong[1*]

[1]City University of Hong Kong
[2]South China University of Technology
[3]Shantou University Medical College
tychen.cs@gmail.com, {cszhangyunfei, cslianxin.xie}@mail.scut.edu.cn, wjshen@stu.edu.cn
cswusi@scut.edu.cn, cshswong@cityu.edu.hk

Abstract

The advent of Spatial Transcriptomics (ST) has revolutionized understanding of tissue architecture by creating high-resolution maps of gene expression patterns. However, the low capture rate of ST leads to significant sparsity. The aim of imputation is to recover biological signals by imputing the dropouts in ST data to approximate the true expression values. In this paper, we introduce a Spatial Gene Expression Imputation Diffusion model to facilitate ST data imputation, and our model is referred to as SpotDiff. Specifically, we incorporate a spot-gene prompt learning module to capture the association between spots and genes. Further, SpotDiff integrates single-cell RNA sequencing data to impute gene expression at each spot. The proposed approach is able to reduce the uncertainty in the imputation process, since the aggregation of multiple single-cell measurements yield a stable representation of the corresponding spot expression profile. Extensive experiments have been performed to demonstrate that SpotDiff outperforms existing imputation methods across multiple benchmarks in terms of yielding more accurate and biologically relevant gene expression profiles, particularly in highly sparse scenarios.

Introduction

The emergence of spatial transcriptomics (ST) has revolutionized our understanding of gene expression by revealing its spatial organization within tissues (Marx 2021; Yuan et al. 2024). This technology preserves the spatial context of gene expression data, offering crucial insights into cellular micro environments and tissue architecture. Nonetheless, the low capture rate of ST data frequently result in data sparsity, presenting substantial challenges to accurate imputation tasks (Li et al. 2024a; Qiao and Huang 2024). A significant challenge in these imputation tasks is the lack of true corresponding data, complicating the recovery of biological signals by imputing the dropouts in ST data to approximate the true gene expression profiles.

In recent years, generative models have made remarkable strides across various domains (Karras, Laine, and Aila 2019; Gao et al. 2023; Sargsyan et al. 2023). Models such as Generative Adversarial Networks (Goodfellow et al. 2014), Variational Autoencoders (Kingma and Welling 2013), Diffusion Models (Ho, Jain, and Abbeel 2020), Matrix Factorization (MF) (Lan et al. 2024) and K-Nearest Neighbors (KNN) (Luo 2022) have been applied to biological data imputation, aiming to generate realistic data that aligns with the target distribution. By learning the complex distributions of data, these generative models can partially recover missing gene expression values (Hu et al. 2021; Bergenstråhle et al. 2022; Wang et al. 2023). However, they can only replicate observed distributions and are unable to infer unobserved biological signals, which restricts their capacity to detect novel or rare biological phenomena. Further, several methods attempt to utilize single-cell RNA sequencing (scRNA-seq) data for imputation, but these approaches encounter two primary challenges: scRNA-seq data is highly sparse, making it an unreliable source for imputation; second, they often neglect the potential advantages of incorporating multi-modal data to enhance the imputation process. These methods typically rely on one modal data source and fail to fully exploit the complementary between different data types. In Figure 1, we briefly contrast the traditional models of imputation methods with the imputed multi-modal information aggregated by SpotDiff.

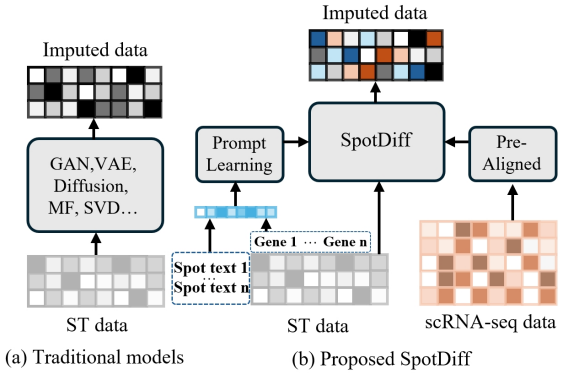

Figure 1: Schematic representation of the imputation process for ST data using (a) traditional models versus the (b) proposed SpotDiff approach.

In this paper, we propose a multi-modal conditional diffusion model (SpotDiff) for the imputation of ST data. To

*Corresponding authors.

accurately characterize multi-modal information within the raw ST data, we incorporate a spot-gene prompt learning module to capture the association between spots and genes. Further, SpotDiff integrates scRNA-seq data to impute gene expression at each spot, which reduces the uncertainty in the imputation process. We utilized the DiT framework to perform a progressively integrated fusion of conditions, resulting in the final denoised imputed results. Experiments demonstrate that SpotDiff significantly outperforms existing imputation methods across various ST imputation benchmarks at both single-cell and 10X resolutions, delivering more accurate and biologically relevant spatial gene expression profiles. SpotDiff provides a novel approach to addressing the challenges of ST data imputation and offers a powerful tool for biological downstream analyses. The main contributions of SpotDiff are as follows:

- We introduce a multi-modal conditional diffusion model that incorporate a spot-gene prompt learning module to capture the association between spots and genes.
- SpotDiff apply an integration strategy to mitigate the adverse effects of scRNA-seq data sparsity, effectively leveraging a substantial amount of data to impute gene expression at each spot.
- SpotDiff outperforms existing imputation techniques on multiple benchmarks, showcasing its superior performance in biological downstream analyses.

Related Work

Early Biological Data Imputation Methods

Biological data often suffer from missing values due to experimental limitations and technical errors. Various imputation methods have been developed to address this issue, each tailored to the unique challenges posed by different types of biological data. Traditional statistical techniques, such as mean imputation, k-nearest neighbors, and principal component analysis, have been widely used due to their simplicity and ease of implementation (Little and Rubin 2019; Troyanskaya et al. 2001; Baghfalaki, Ganjali, and Berridge 2016). However, these methods often fail to capture the complex, nonlinear relationships inherent in biological systems. More advanced approaches, such as matrix factorization and Bayesian methods, have shown improved performance by leveraging the underlying structure of the data (Stekhoven and Bühlmann 2012; Chen et al. 2020; Ou-Yang et al. 2022; Gan, Vinci, and Allen 2022). Despite these advancements, there remains a need for more sophisticated imputation techniques that can accurately reconstruct missing values while preserving the biological relevance of the data.

Imputation with Generative Models

Generative models have emerged as powerful tools for imputation in various domains. These models, such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs) and Diffusion Models, can learn to generate realistic data samples by capturing the underlying data distribution (Goodfellow et al. 2014; Kingma and Welling 2013; Ho, Jain, and Abbeel 2020). In the context of data imputation, generative models offer a distinct advantage by leveraging their ability to model complex, high-dimensional data. Recent studies have demonstrated the efficacy of GANs and VAEs in imputing missing values in biological datasets, outperforming traditional imputation methods (Xu et al. 2020; Si et al. 2023; Zhu et al. 2024). For instance, Huang et al. introduced scGGAN, a Graph-based GAN imputation model that effectively handles missing data by integrating graph relationship and generating plausible values conditioned on the observed data (Huang et al. 2023). Similarly, deep learning-based approaches, such as Denoising Autoencoders, have shown promise in reconstructing missing values in high-throughput sequencing data (Inoue 2024). These advancements highlight the potential of generative models to revolutionize data imputation in biology, providing more accurate and biologically meaningful results.

Spatial Transcriptomics Imputation Methods

Spatial transcriptomics is a cutting-edge technology that enables the spatial mapping of gene expression within tissues, offering unprecedented insights into cellular heterogeneity and tissue architecture (Ståhl et al. 2016). However, the resolution of ST data is often limited by technical constraints, resulting in sparse and incomplete datasets. To address this challenge, various imputation methods have been developed to enhance the resolution and completeness of ST data. Traditional imputation techniques, such as spatial smoothing and interpolation, have been employed to fill in missing values by leveraging the spatial continuity of gene expression (Svensson, Teichmann, and Stegle 2018). More recently, spatially aware deep learning models have been proposed to improve imputation accuracy (Hu et al. 2021; Bergenstråhle et al. 2022). For example, SpaFormer leverages the power of transformers, a type of deep learning model designed for natural language processing, to impute ST data. By integrating the spatial location with positional encoding, SpaFormer can effectively capture and model the spatial relationships and gene expression patterns within tissues (Wen et al. 2023). stMCDI combines the strengths of diffusion models and graph neural networks to impute ST data. The masked conditional diffusion model facilitates the prediction of missing values by modeling the diffusion of gene expression signals across the tissue (Li et al. 2024b). Also, stDiff leverages single-cell RNA-seq data to inform the imputation of ST. This integrative approach allows for the preservation of single-cell resolution while enhancing the spatial context, leading to more accurate and biologically meaningful imputations (Li et al. 2024a).

Methodology

Problem Definition

The problem of ST data imputation can be defined as follows: Given an ST data matrix, the goal is to generate an imputed matrix that more closely approximates the true spatial gene expression in ST data. Let the ST data matrix be $\mathbf{X}_{st} \in \mathbb{R}^{c_1 \times g_{st}}$, where c_1 denotes the number of spots or cells (spots for 10X resolution, cells for single-cell resolution), and g_{st} denotes the number of genes in ST data.

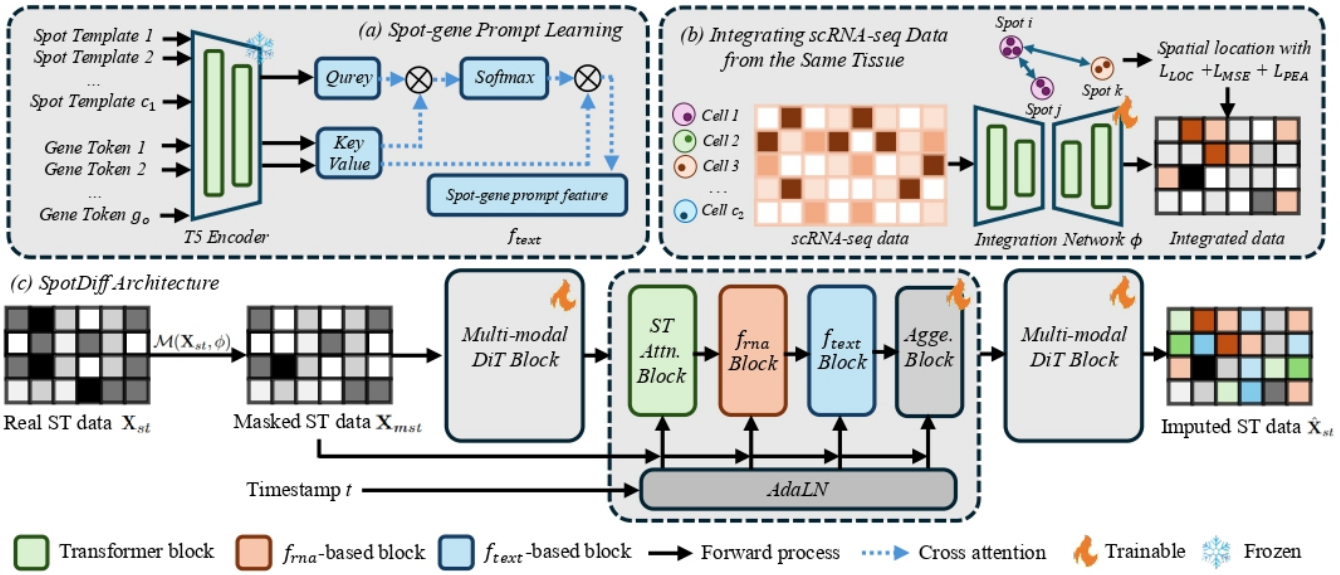

Figure 2: Overview of the SpotDiff architecture. The figure illustrates the architecture of the SpotDiff model, showcasing its three key components. (a) The spot-gene prompt learning module leverages a T5 encoder to enhance feature extraction, using spot templates and gene tokens to capture associations. (b) The proposed integration network reconstructs and integrates scRNA-seq data into the target space of masked ST data. (c) The SpotDiff architecture combines scRNA-seq data and textual modal data with masked ST data through integration network and cross-attention mechanisms.

To better simulate the missing parts of ST data due to various external factors, we apply random masking strategy $\mathcal{M}(\cdot, \phi)$ to the ST data. Following the settings in references (Li et al. 2024a,b), the masking threshold ϕ is set to 0.3, obtaining $\mathbf{X}_{mst} = \mathcal{M}(\mathbf{X}_{st}, \phi)$. Thus, our task can be represented as mapping the sparse matrix $\mathbf{X}_{mst}$ to the imputed matrix $\hat{\mathbf{X}}_{st}$. Additionally, SpotDiff incorporates the influence of scRNA-seq data on the imputation task, hence we also define the corresponding scRNA-seq matrix, denoted as $\mathbf{X}_{rna} \in \mathbb{R}^{c_2 \times g_{rna}}$, where c_2 represents the number of cells and g_{rna} the number of genes. We compute the overlapping genes between the normalized and log-transformed $\mathbf{X}_{mst}$ and $\mathbf{X}_{rna}$, resulting in the updated masked ST data $\mathbf{X}_{mst} \in \mathbb{R}^{c_1 \times g_o}$ and scRNA-seq data $\mathbf{X}_{rna} \in \mathbb{R}^{c_2 \times g_o}$, where g_o denotes the number of overlapping genes between the two datasets. The final imputation process can be succinctly expressed as $\hat{\mathbf{X}}_{st} = \psi(\mathbf{X}_{mst})$, where ψ represents our imputation model, and $\hat{\mathbf{X}}_{st}$ denotes the imputed ST data. Figure 2 shows the architecture of the proposed SpotDiff.

Spot-gene Prompt Learning

For the ST data imputation task, it is essential to establish a textual description and association of spots and genes to enhance the text-modal representations. We begin by constructing prompts for each masked ST spot data $\mathbf{X}_{mst}$ and designing a template $\mathcal{T}$ for each spot:

$$\mathcal{T} = \begin{array}{l}\text{'The spot has highest expression with } [gene]\ [count], \ldots \\ [gene]\ [count] \text{ and non-zero lower expressed } [gene] \\ [count], \ldots, [gene]\ [count]\text{'}.\end{array} \tag{1}$$

We fill in the names of the n_1 highest expressed genes and the n_2 lowest expressed genes other than non-zero genes of each spot and their corresponding expression values into the above template to form a descriptive statement of the spot T_{spot}. The template we design enables the model to generate a coherent narrative about the gene expression landscape of each spot, facilitating textual-modal understanding of spatial gene associations. By utilizing the T5 text encoder, we convert spot descriptions T_{spot} into embeddings f_{spot}. This approach not only highlights high-expressing genes but also incorporates lower-expressing genes, providing a comprehensive overview of gene expression within the tissue. To further refine the model ability to discern local gene influences, we introduce gene name tokens T_{gene} that represent the individual genes mentioned in the spot description. Encoding these tokens allows us to capture localized expression patterns, ensuring that the model considers both global spot characteristics and localized gene associations. By inputting these into the T5 encoder, we obtain gene token embeddings f_{gene}.

We then design a spot-gene cross-attention mechanism to consider the associations between local gene tokens and the global spot description. By treating the global spot description as the query and the gene token embeddings as keys and values, the model effectively integrates information from both sources. The attention module can be formally expressed as follows:

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) * V. \tag{2}$$

Correspondingly, we have the operation of the spot-gene

cross-attention module expressed as follows:

$$f_{text} = \text{Attn}(f_{spot}, f_{gene}, f_{gene}), \tag{3}$$

where f_{text} denotes the output of the cross-attention module, and is also used as the condition for SpotDiff. The corresponding graphical description is shown in Figure 2 (a).

Integrating scRNA-seq Data from the Same Tissue

To fully recover the ST data and perform meaningful imputation, relying solely on raw ST data and corresponding spot-gene description is insufficient. Generative models can capture only the original data distribution, using the learned target domain distribution to impute the missing values does not yield meaningful imputation. Therefore, we propose integrating scRNA-seq data from the same tissue to impute gene expression at each spot, facilitating a more comprehensive imputation of ST data.

ST data are sparse and lack true counterparts, making imputation challenging. Therefore, we propose integrating scRNA-seq data from the same tissue to accurately impute gene expression at each spot. The original scRNA-seq data can provide richer information in terms of sample size and expression levels. By integrating a large volume of scRNA-seq data, we can reduce the uncertainty associated with individual scRNA-seq measurements. Aggregating multiple single-cell measurements yields a stable representation of the corresponding spot expression profile. This approach ensures that, when the quantity and expression levels of scRNA-seq data are sufficiently high, we can obtain more reliable expression information to facilitate the imputation task. For the integration of scRNA-seq data, we propose an integration network ω, designed to align and integrate scRNA-seq data into the raw masked ST data. This process can be expressed as $\widetilde{\mathbf{X}}_{st} = \omega(\mathbf{X}_{rna})$. The integrated network comprises a stacked 6-layer Unet-based architecture.

During the training phase, we explicitly consider the positional information of each ST spot. By analyzing the distances between spots, we can infer that nearby spots likely share complementary expression profiles, while those farther apart should exhibit distinct expression patterns. The spatial awareness is crucial for the model to generalize and accurately impute missing values. To enforce this spatial context, we design a loss term that encourages complementary expressions for nearby spots while ensuring that the expressions of distant spots are differentiated. This spatially-informed training strategy enhances the model capability to generate biologically meaningful imputed data. We have the loss term $\mathcal{L}_{LOC}$ as follows:

$$\begin{aligned}\mathcal{L}_{LOC} = \mathbb{E}_{\omega}^{i,j,k}\Big[&\left\|\mathbf{X}_{mst_i} - \mathbf{X}_{mst_j}\right\|_2^2 \odot \exp\left(\Delta_{i,j}\right) \\ &+ \left\|\mathbf{X}_{mst_i} - \mathbf{X}_{mst_k}\right\|_2^2 \odot \exp\left(\Delta_{i,k}\right)\Big],\end{aligned} \tag{4}$$

where $\Delta_{i,j} = \frac{d(i,j)}{\sigma}$ and $\Delta_{i,k} = \frac{d(i,k)}{\sigma}$ represent the normalized distance between spots i and j, k, with i and j denoting two nearby spots and k representing a distant spot. Here, $d(i,j)$ and $d(i,k)$ are the distances between the respective spots, and σ is the distance decay coefficient. $\mathcal{L}_{loc}$ ensures complementary expressions for nearby spots and pulls apart the expressions of distant spots. At the same time, to make the integrated target consistent with the ST data distribution, we also apply MSE loss $\mathcal{L}_{MSE} = \mathbb{E}_{\omega}\|\widetilde{\mathbf{X}}_{st} - \mathbf{X}_{mst}\|_2^2$ in the integration network ω. Further, to ensure that ω can capture the correlation between integrated and true gene expression profiles for each gene across all samples, we apply the gene-wise Pearson loss:

$$\mathcal{L}_{PEA} = \mathbb{E}_{\omega}[1 - \frac{1}{n}\sum_{j=1}^{n}\frac{\text{Cov}(\widetilde{\mathbf{X}}_{st}, \mathbf{X}_{mst})}{\sigma_{\widetilde{\mathbf{X}}_{st}}\sigma_{\mathbf{X}_{mst}}}], \tag{5}$$

where $\text{Cov}(\cdot,\cdot)$ denotes the covariance and σ represents the standard deviation. The integration network aligns the two types of data, making subsequent similarity calculations more reasonable, while leveraging the characteristics of integration to the scRNA-seq data according to the target domain ST. After obtaining the integrated ST data $\widetilde{\mathbf{X}}_{st}$, we utilize cosine similarity to identify the nearest spot:

$$\text{CosSim}(\mathbf{X}_{mst}, \widetilde{\mathbf{X}}_{st}) = \frac{\mathbf{X}_{mst} \odot \widetilde{\mathbf{X}}_{st}}{\|\mathbf{X}_{mst}\|\|\widetilde{\mathbf{X}}_{st}\|}. \tag{6}$$

For each ST data point, we select the top $\mathcal{N}$ nearest scRNA-seq data points based on their similarity. We then perform a weighted fusion based on normalized similarity to obtain the final nearest scRNA-seq data:

$$\mathbf{X}_{rna}^{pair} = \sum_{n=1}^{\mathcal{N}} w_n \widetilde{\mathbf{X}}_{st_n}, \tag{7}$$

where the weights w_n for $n = 1, \ldots, \mathcal{N}$ are computed based on the normalized similarities. In this manner, each ST data point is augmented with the nearest scRNA-seq data through a weighted fusion of the most similar scRNA-seq data points. We designed a linear layer to reshape the integrated data $\mathbf{X}_{rna}^{pair}$ to obtain integrated scRNA-seq features $f_{rna} = \text{Linear}(\mathbf{X}_{rna}^{pair})$ so that they can be inserted into the imputation model ψ as conditions. The corresponding graphical description is shown in Figure 2 (b).

SpotDiff: Spatial Gene Expression Diffusion

Our final model is based on the diffusion model framework of DiT. The diffusion model operates as a Markov chain that progressively denoises data. Specifically, the forward process gradually adds Gaussian noise to the data, while the reverse process aims to reconstruct the original data by removing this noise. Given the masked ST data $\widetilde{\mathbf{X}}_{mst}$, diffusion model ψ seeks to generate high-fidelity imputed ST data $\hat{\mathbf{X}}_{st}$. The diffusion process can be expressed as:

$$\mathbf{X}_t = \sqrt{\bar{\alpha}_t}\mathbf{X}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{8}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ is noise sampled from the standard normal distribution, $\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$ in which α_s is a variance schedule. The reverse process is parameterized by a neural

Methods	osmFISH				FISH				STARmap			
	PCC↑	SSIM↑	RMSE↓	COSSIM↑	PCC↑	SSIM↑	RMSE↓	COSSIM↑	PCC↑	SSIM↑	RMSE↓	COSSIM↑
gimVI	0.2011	0.0876	1.3071	0.3845	0.4984	0.3254	1.0045	0.8266	0.1854	0.0774	1.3102	0.3485
Tangram	0.2124	0.1120	1.3158	0.4011	0.4622	0.3785	1.0121	0.8260	0.2254	0.1254	1.2451	0.3815
GraphST	0.2354	0.1211	1.2741	0.4658	0.3432	0.1136	1.1262	0.8014	0.1754	0.0662	1.3220	0.3154
SpaFormer	0.3020	0.2421	1.1003	0.5721	0.5112	0.4325	0.8654	0.8654	0.2454	0.1316	1.2237	0.4657
stMDCI	0.2954	0.2444	1.1021	0.5845	0.5876	0.5012	0.7985	0.8741	0.2874	0.1884	1.1878	0.4951
stDiff	0.2775	0.2293	1.1902	0.5545	0.4109	0.3532	1.0084	0.8221	0.1754	0.0556	1.3500	0.2215
STEM	0.2745	0.2284	1.2015	0.5412	0.3874	0.2483	1.0842	0.6657	0.1984	0.0945	1.3025	0.3548
Ours	**0.3721**	**0.3021**	**1.0442**	**0.6615**	**0.6215**	**0.5564**	**0.7548**	**0.8845**	**0.3215**	**0.2254**	**1.1584**	**0.5512**

Methods	MERFISH				10x_BA				10x_HBC			
	PCC↑	SSIM↑	RMSE↓	COSSIM↑	PCC↑	SSIM↑	RMSE↓	COSSIM↑	PCC↑	SSIM↑	RMSE↓	COSSIM↑
gimVI	0.2365	0.1245	1.3548	0.3025	0.1845	0.1121	1.2675	0.4571	0.1837	0.0532	1.2730	0.4161
Tangram	0.3421	0.2355	1.1335	0.4412	0.1847	0.0994	1.2841	0.4325	0.1829	0.0530	1.2736	0.4150
GraphST	0.2157	0.1025	1.3848	0.2215	0.1896	0.1021	1.2654	0.3845	0.1663	0.0044	1.2868	0.3938
SpaFormer	0.3325	0.2154	1.1669	0.4021	0.2401	0.1521	1.2330	0.4754	0.2100	0.1125	1.2665	0.4165
stMDCI	0.3412	0.2215	1.1284	0.4251	0.2311	0.1422	1.2366	0.4654	0.2244	0.1551	1.2886	0.4189
stDiff	0.2451	0.1185	1.2749	0.3256	0.1954	0.0658	1.2548	0.5142	0.2085	0.1201	1.2899	0.4088
STEM	0.1854	0.0854	1.4551	0.1884	0.1896	0.1021	1.2654	0.4687	0.1832	0.1685	1.2733	0.4156
Ours	**0.3784**	**0.3548**	**1.0215**	**0.4851**	**0.2559**	**0.2215**	**1.1025**	**0.6651**	**0.2345**	**0.2025**	**1.1554**	**0.5012**

Table 1: Quantitative results with competing methods in osmFISH, FISH, STARmap, MERFISH, 10x_BA and 10x_HBC. Best results are **boldfaced** and second best results are underlined.

network ϵ_ψ that predicts the added noise, and is trained to minimize the denoising objective function:

$$\mathcal{L}_{DIFF} = \mathbb{E}_{\psi}^{\epsilon,t}\left[\left\|\epsilon - \epsilon_\psi(\mathbf{X}_t, t, f_{text}, f_{rna}, \widetilde{\mathbf{X}}_{mst})\right\|_2^2\right]. \quad (9)$$

The conditioning information includes both the masked ST data and the textual features f_{text} derived from the spot-gene prompt learning module. During the training phase, the model is trained on the noisy data $\mathbf{X}_t$ at various timestamps t to predict the noise ϵ. In the integration network ω, we minimize the combined loss functions to ensure accurate spatial integration and reconstruction of the ST data. For the imputation network ψ, the loss function is designed to capture discrepancies between imputed and raw ST data. The two networks are trained sequentially: first, the integration network ω is trained, followed by the diffusion network ψ. The formal expressions for minimizing the loss terms are as follows:

$$\begin{aligned} &\min_{\omega} \quad \mathcal{L}_{MSE} + \lambda_1\mathcal{L}_{LOC} + \lambda_2\mathcal{L}_{PEA}, \\ &\min_{\psi} \quad \mathcal{L}_{DIFF}, \end{aligned} \quad (10)$$

where λ_1 and λ_2 denote the loss weight parameter. In summary, SpotDiff integrates multi-modal information, including masked ST data, spot-gene textual data and integrated scRNA-seq data, to achieve high-fidelity imputation of ST data. By leveraging the strengths of each data modality, SpotDiff demonstrates robust performance across diverse biological contexts.

Experiments

Datasets and Pre-processing

In this study, we utilized six distinct datasets to evaluate the performance of imputation methods. They are osmFISH (Codeluppi et al. 2018), FISH (Li et al. 2022), STARmap (Wang et al. 2018), MERFISH (Li et al. 2022), 10x_BA (Long et al. 2023) and 10x_HBC (Long et al. 2023). These datasets encompass various tissues and organ types, each with corresponding ST and scRNA-seq data. The details of datasets are summarized in Table 2. To ensure data prepossessing consistency, we applied a uniform data prepossessing pipeline (Li et al. 2022) to all datasets. This included log transformation and normalization of the gene expression data. The overlapped genes were used as the basis for the ST imputation training tasks. We further identified overlapping genes between the ST and scRNA-seq data for each dataset. The integration of scRNA-seq data with ST data allowed us to leverage the high-resolution single-cell information to enhance the ST imputation accuracy.

Data Protocols (Tissue)	Raw ST Data		scRNA-seq Data		Overlapped
	Spots/Cells	Genes	Cells	Genes	Genes
osmFISH (Codeluppi et al. 2018) (Somatosensory cortex)	3,405	33	5,613	30,527	33
FISH (Li et al. 2022) (Embryo)	3,039	84	1,297	8,924	84
STARmap (Wang et al. 2018) (Primary Visual Cortex)	1,549	1,020	14,249	34,041	996
MERFISH (Li et al. 2022) (Primary Visual Cortex)	2,399	268	14,249	34,041	265
10x_BA (Long et al. 2023) (Brain Anterior)	2,695	32,285	116,921	22,764	1,099
10x_HBC (Long et al. 2023) (Human Breast Cancer)	3,798	3,6601	46,080	5,000	921

Table 2: Summary of imputation datasets with data protocol (tissue), raw ST data, scRNA-seq data and overlapped genes.

Comparison Methods

We compare several state-of-the-art methods for ST imputation, specifically gimVI (Lopez et al. 2019), Tangram (Biancalani et al. 2021), GraphST (Long et al. 2023), SpaFormer (Wen et al. 2023), stMDCI (Li et al. 2024b), stDiff (Li et al.

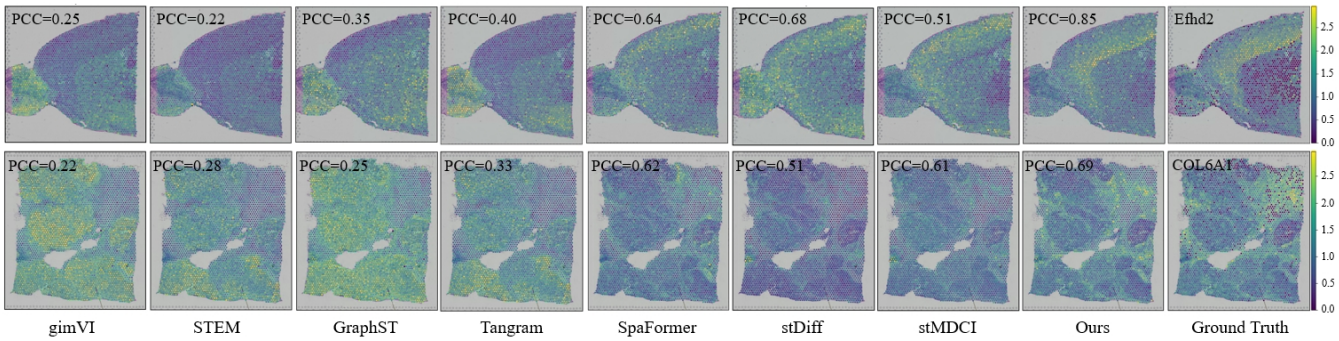

Figure 3: Visualization of marker gene recovery for 10x_BA and 10x_HBC. The images illustrate the imputation results of marker genes obtained using various models, compared to the ground truth.

2024a) and STEM (Hao et al. 2024). These methods have been selected based on their recent advancements and relevance in the high-related research field. In experiments, to ensure a more comprehensive and fair comparison, we selected five metrics to quantitatively analyze the final results: Pearson Correlation Coefficient (PCC), Structural Similarity Index (SSIM), Root Mean Square Error (RMSE), and Cosine Similarity (COSSIM).

Quantitative Results Table 1 provides the quantitative results, where we measure the performance using several metrics. The results highlight the superiority of our method across various datasets. Our method consistently outperforms the competing methods across all six datasets in each metric, demonstrating the effectiveness of our approach. This highlights the importance of the innovative conditions we have introduced for the imputation task. Notably, in the osmFISH dataset, our method achieves a PCC of 0.3721, which significantly surpasses the second-best method, SpaFormer with a PCC of 0.3020. These results underscore the significant advancements our method brings to ST imputation. Similarly, for the FISH dataset, our method attains a PCC of 0.6215, outperforming the second-best method, stMDCI which has a PCC of 0.5876. In other datasets, our method surpasses other imputation methods in all quantitative indicators.

Recovery of Marker Genes Further, we investigate whether the imputation model can restore marker genes in a biologically meaningful way. We randomly masking marker genes and utilizing imputation models to impute the marker genes based on the remaining data. As shown in Figure 3, we present marker gene reference images for 10x_BA and 10x_HBC. SpotDiff excels in capturing the precise gene expression and spatial information, whereas other methods display a coarser grasp of gene expression patterns. Models such as gimVI, STEM, GraphST, and Tangram exhibit biases in gene expression prediction, while SpaFormer, stDiff, and stMDCI tend to overestimate gene expression, resulting in generally higher count values. These findings highlight the robustness of our method in recovering marker gene expression counts across various datasets.

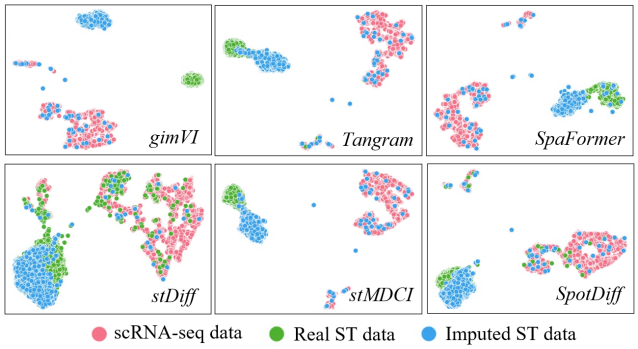

Figure 4: UMAP plots of scRNA-seq data, raw ST data and imputed ST data.

Methods	f_{rna}	f_{text}	osmFISH	FISH	STARmap	MERFISH
Baseline			0.2987	0.5124	0.2315	0.3025
		✓	0.3215	0.5651	0.2655	0.3315
	✓		0.3612	0.6055	0.3154	0.3644
	✓	✓	0.3721	0.6215	0.3215	0.3784

Table 3: Ablation study for different combinations of f_{rna} and f_{text} across various datasets.

Overall Distribution Matching To visually illustrate the overall distribution matching between the imputed and raw ST data, we generated UMAP plot in the STARmap dataset for scRNA-seq data, real ST data, and imputed ST data. Figure 4 reveals a gap between the imputed results of gimVI, Tangram, SpaFormer, and stMDCI compared to the real data. In contrast, stDiff and SpotDiff exhibit results that are more in line with expectations. However, stDiff still shows some distribution shift for the primary ST distribution when compared to SpotDiff. Therefore, SpotDiff demonstrates a superior ability to accurately capture the overall distribution of the existing raw ST data compared to other methods.

Ablation Study

We present an ablation study to evaluate the impact of components in SpotDiff, as summarized in Table 3. Our focus is on assessing the effects of incorporating f_{rna} and f_{text}

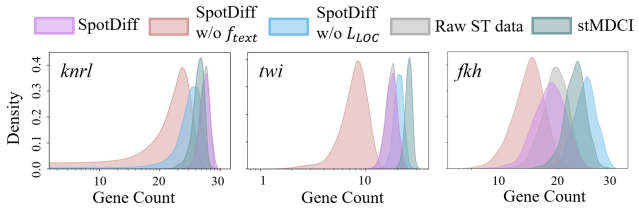

Figure 5: Distribution visualization of gene expression for 'knrl', 'twi', and 'fkh' in the FISH dataset.

features, along with analyzing the influence of various loss functions. The baseline model does not utilize these features, serves as a reference point for comparison. The results indicate that the inclusion of f_{rna} significantly enhances the imputation performance across all datasets. This improvement highlights the importance of leveraging scRNA-seq data, which captures detailed gene expression profiles at single-cell resolution. Further, the addition of f_{text} features also contributes to performance gains. These textual features provide contextual insights that enrich the model understanding of gene associations.

The Impact of Key Components In Figure 5, we compare the distribution of the top three marker genes 'knr', 'twi' and 'fkh' in the FISH dataset. The proposed spot-gene prompt learning focuses on capturing the impact of significant genes on imputation. By disabling f_{text}, SpotDiff fails to capture the expression of significant genes effectively, with intensity and abundance not reaching the levels of the raw ST data. The peak regions of gene expression also do not match the raw ST data. Further, we demonstrated the importance of spot positional information in ST for imputation. Disabling $\mathcal{L}_{LOC}$ results in shifts and instability in gene expression counts, with several instances of bimodal expression observed in 'knrl' and 'twi'. We also plotted the results of the main comparison method stMDCI. SpotDiff provides more precise capture of expressed genes, showing the closest matching with the raw ST data distribution.

Integration Performance We further visualized the actual effects of the integrated network using the MERFISH dataset. In the mouse cerebral cortex, glutamatergic neurons are characterized by specific layered spatial locations (Biancalani et al. 2021). The spots include subtypes such as 'L2/3 IT', 'L4', 'L5 PT', and 'L6 CT'. In Figure 6, we display these subtypes, highlighting their distinct spatial positions (marked by red dashed boxes). Meanwhile, the integrated network ω in SpotDiff demonstrates the marker gene integration effects for each of the four subtypes. It is evident that SpotDiff captures more spatial information about the marker genes within the subtypes compared to the gold standard integration model Tangram, showing superior expression in the designated spatial locations.

Cell Population Identification

We utilize ST data with well-annotated cell types from Mop (Zhang et al. 2021) to perform further clustering analyses. The known cell types serve as a reference for clustering,

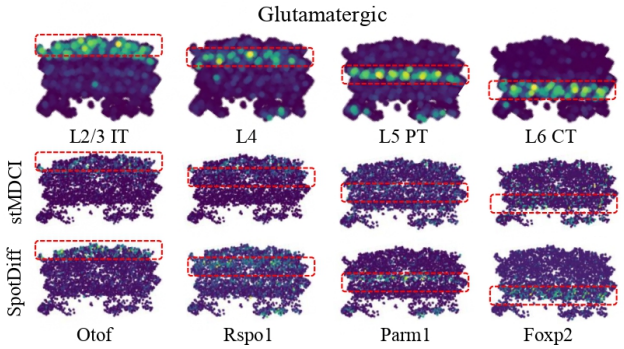

Figure 6: Integration Results in the MERFISH dataset.

Metrics	Raw ST data	gimVI	Tangram	stDiff	stMDCI	SpotDiff
ARI↑	0.6404	0.6832	0.6732	0.7254	0.7022	**0.7336**
AMI↑	0.8073	0.8038	0.7936	0.8345	0.7784	**0.8440**
NMI↑	0.8097	0.8074	0.7972	0.8361	0.7841	**0.8561**

Table 4: Clustering metrics for raw ST data and imputed results in the Mop dataset using comparison methods.

and we employed three clustering metrics (ARI, AMI and NMI) to evaluate the performance. The clustering results from the raw ST data is obtained using Leiden clustering (Traag, Waltman, and Van Eck 2019), serve as a baseline for comparison. In Table 4, the results of gimVI and Tangram show slight improvements over the raw ST data. In contrast, stDiff, stMDCI, and SpotDiff significantly outperform the baseline, with SpotDiff achieving the highest performance, particularly excelling in cell population discovery.

Conclusion

In this paper, we address the critical need for biologically meaningful ST imputation by incorporating the multi-modal biological textual information. We further utilize an integration network for scRNA-seq data to impute gene expression at each spot, which significantly reduce the uncertainty in the imputation process. We demonstrate that the associations between spots and genes within the ST data are crucial for improving the imputation performance. Experiments indicate that SpotDiff outperforms existing imputation methods in quantitative results, while also achieving optimal recovery of marker genes and matching the overall ST data distribution. Additional biological analyses further validate the superiority of SpotDiff in cell population discovery.

Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities (Project No. 2024ZYGXZR077), in part by the National Natural Science Foundation of China (Project No. 62072189), in part by the GuangDong Basic and Applied Basic Research Foundation (Project No. 2024A1515011437, 2023A1515030154), and in part by TCL Science and Technology Innovation Fund (Project No. 20231752).

References

Baghfalaki, T.; Ganjali, M.; and Berridge, D. 2016. Missing value imputation for RNA-sequencing data using statistical models: a comparative study. *Journal of Statistical Theory and Applications*, 15(3): 221–236.

Bergenstråhle, L.; He, B.; Bergenstråhle, J.; Abalo, X.; Mirzazadeh, R.; Thrane, K.; Ji, A. L.; Andersson, A.; Larsson, L.; Stakenborg, N.; et al. 2022. Super-resolved spatial transcriptomics by deep data fusion. *Nature biotechnology*, 40(4): 476–479.

Biancalani, T.; Scalia, G.; Buffoni, L.; Avasthi, R.; Lu, Z.; Sanger, A.; Tokcan, N.; Vanderburg, C. R.; Segerstolpe, Å.; Zhang, M.; et al. 2021. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nature methods*, 18(11): 1352–1362.

Chen, C.; Wu, C.; Wu, L.; Wang, X.; Deng, M.; and Xi, R. 2020. scRMD: imputation for single cell RNA-seq data via robust matrix decomposition. *Bioinformatics*, 36(10): 3156–3161.

Codeluppi, S.; Borm, L. E.; Zeisel, A.; La Manno, G.; van Lunteren, J. A.; Svensson, C. I.; and Linnarsson, S. 2018. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature methods*, 15(11): 932–935.

Gan, L.; Vinci, G.; and Allen, G. I. 2022. Correlation Imputation for single-cell RNA-seq. *Journal of Computational Biology*, 29(5): 465–482.

Gao, H.; Zhang, X.; Gao, X.; Li, F.; and Han, H. 2023. ICoT-GAN: Integrated convolutional transformer GAN for rolling bearings fault diagnosis under limited data condition. *IEEE Transactions on Instrumentation and Measurement*, 72: 1–14.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Proc. Neural Information Processing Systems*.

Hao, M.; Luo, E.; Chen, Y.; Wu, Y.; Li, C.; Chen, S.; Gao, H.; Bian, H.; Gu, J.; Wei, L.; et al. 2024. STEM enables mapping of single-cell and spatial transcriptomics data with transfer learning. *Communications Biology*, 7(1): 56.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Hu, J.; Li, X.; Coleman, K.; Schroeder, A.; Ma, N.; Irwin, D. J.; Lee, E. B.; Shinohara, R. T.; and Li, M. 2021. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods*, 18(11): 1342–1351.

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. *Briefings in bioinformatics*, 24(2): bbad040.

Inoue, Y. 2024. scVGAE: A Novel Approach using ZINB-Based Variational Graph Autoencoder for Single-Cell RNA-Seq Imputation. *arXiv preprint arXiv:2403.08959*.

Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Kingma, D. P.; and Welling, M. 2013. Auto-encoding variational Bayes. In *arXiv:1312.6114*.

Lan, W.; Chen, J.; Liu, M.; Chen, Q.; Liu, J.; Wang, J.; and Chen, Y.-P. P. 2024. Deep imputation bi-stochastic graph regularized matrix factorization for clustering single-cell RNA-sequencing data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Li, B.; Zhang, W.; Guo, C.; Xu, H.; Li, L.; Fang, M.; Hu, Y.; Zhang, X.; Yao, X.; Tang, M.; et al. 2022. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nature methods*, 19(6): 662–670.

Li, K.; Li, J.; Tao, Y.; and Wang, F. 2024a. stDiff: a diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Briefings in Bioinformatics*, 25(3): bbae171.

Li, X.; Min, W.; Wang, S.; Wang, C.; and Xu, T. 2024b. stMCDI: Masked Conditional Diffusion Model with Graph Neural Network for Spatial Transcriptomics Data Imputation. *arXiv preprint arXiv:2403.10863*.

Little, R. J.; and Rubin, D. B. 2019. *Statistical analysis with missing data*, volume 793. John Wiley & Sons.

Long, Y.; Ang, K. S.; Li, M.; Chong, K. L. K.; Sethi, R.; Zhong, C.; Xu, H.; Ong, Z.; Sachaphibulkij, K.; Chen, A.; et al. 2023. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications*, 14(1): 1155.

Lopez, R.; Nazaret, A.; Langevin, M.; Samaran, J.; Regier, J.; Jordan, M. I.; and Yosef, N. 2019. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *arXiv preprint arXiv:1905.02269*.

Luo, Y. 2022. Evaluating the state of the art in missing data imputation for clinical data. *Briefings in Bioinformatics*, 23(1): bbab489.

Marx, V. 2021. Method of the Year: spatially resolved transcriptomics. *Nature methods*, 18(1): 9–14.

Ou-Yang, L.; Lu, F.; Zhang, Z.-C.; and Wu, M. 2022. Matrix factorization for biomedical link prediction and scRNA-seq data imputation: an empirical survey. *Briefings in Bioinformatics*, 23(1): bbab479.

Qiao, C.; and Huang, Y. 2024. Reliable imputation of spatial transcriptomes with uncertainty estimation and spatial regularization. *Patterns*, 5(8): 101021.

Sargsyan, A.; Navasardyan, S.; Xu, X.; and Shi, H. 2023. Mi-gan: A simple baseline for image inpainting on mobile devices. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7335–7345.

Si, T.; Hopkins, Z.; Yanev, J.; Hou, J.; and Gong, H. 2023. A novel f-divergence based generative adversarial imputation method for scRNA-seq data analysis. *Plos one*, 18(11): e0292792.

Ståhl, P. L.; Salmén, F.; Vickovic, S.; Lundmark, A.; Navarro, J. F.; Magnusson, J.; Giacomello, S.; Asp, M.; Westholm, J. O.; Huss, M.; et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294): 78–82.

Stekhoven, D. J.; and Bühlmann, P. 2012. MissForest—nonparametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1): 112–118.

Svensson, V.; Teichmann, S. A.; and Stegle, O. 2018. SpatialDE: identification of spatially variable genes. *Nature methods*, 15(5): 343–346.

Traag, V. A.; Waltman, L.; and Van Eck, N. J. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1): 1–12.

Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; and Altman, R. B. 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6): 520–525.

Wang, T.; Zhao, H.; Xu, Y.; Wang, Y.; Shang, X.; Peng, J.; and Xiao, B. 2023. scMultiGAN: cell-specific imputation for single-cell transcriptomes with multiple deep generative adversarial networks. *Briefings in Bioinformatics*, 24(6): bbad384.

Wang, X.; Allen, W. E.; Wright, M. A.; Sylwestrak, E. L.; Samusik, N.; Vesuna, S.; Evans, K.; Liu, C.; Ramakrishnan, C.; Liu, J.; et al. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, 361(6400): eaat5691.

Wen, H.; Tang, W.; Jin, W.; Ding, J.; Liu, R.; Shi, F.; Xie, Y.; and Tang, J. 2023. Single cells are spatial tokens: Transformers for spatial transcriptomic data imputation. *arXiv preprint arXiv:2302.03038*.

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 research*, 48(15): e85–e85.

Yuan, Z.; Zhao, F.; Lin, S.; Zhao, Y.; Yao, J.; Cui, Y.; Zhang, X.-Y.; and Zhao, Y. 2024. Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nature Methods*, 1–11.

Zhang, M.; Eichhorn, S. W.; Zingg, B.; Yao, Z.; Cotter, K.; Zeng, H.; Dong, H.; and Zhuang, X. 2021. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature*, 598(7879): 137–143.

Zhu, X.; Meng, S.; Li, G.; Wang, J.; and Peng, X. 2024. AGImpute: imputation of scRNA-seq data based on a hybrid GAN with dropouts identification. *Bioinformatics*, 40(2): btae068.