

scRDiT: GENERATING SINGLE-CELL RNA-SEQ DATA BY DIFFUSION TRANSFORMERS AND ACCELERATING SAMPLING

Shengze Dong, Zhuorui Cui, Ding Liu*,
School of Computer Science and Technology
Tiangong University
Tianjin, 300387
People's Republic of China

Jinzhi Lei†
School of Mathematical Sciences
Tiangong University
Tianjin, 300387
People's Republic of China

ABSTRACT

Motivation: Single-cell RNA sequencing (scRNA-seq) is a groundbreaking technology extensively utilized in biological research, facilitating the examination of gene expression at the individual cell level within a given tissue sample. While numerous tools have been developed for scRNA-seq data analysis, the challenge persists in capturing the distinct features of such data and replicating virtual datasets that share analogous statistical properties.

Results: Our study introduces a generative approach termed scRNA-seq Diffusion Transformer (scRDiT). This method generates virtual scRNA-seq data by leveraging a real dataset. The method is a neural network constructed based on Denoising Diffusion Probabilistic Models (DDPMs) and Diffusion Transformers (DiTs). This involves subjecting Gaussian noises to the real dataset through iterative noise-adding steps and ultimately restoring the noises to form scRNA-seq samples. This scheme allows us to learn data features from actual scRNA-seq samples during model training.

Our experiments, conducted on two distinct scRNA-seq datasets, demonstrate superior performance. Additionally, the model sampling process is expedited by incorporating Denoising Diffusion Implicit Models (DDIM). scRDiT presents a unified methodology empowering users to train neural network models with their unique scRNA-seq datasets, enabling the generation of numerous high-quality scRNA-seq samples.

Availability and implementation: <https://github.com/DongShengze/scRDiT>

Keywords Diffusion Models · Transformers · single-cell RNA-seq · Deep Learning

1 Introduction

Single-cell RNA sequencing (scRNA-seq) is a powerful technique that allows researchers to analyze the gene expression profiles of individual cells within a tissue or population. By capturing and sequencing the RNA molecules within each cell, scRNA-seq offers a comprehensive snapshot of the cellular composition and molecular heterogeneity of a sample[1]. This technology has significantly advanced biological research, facilitating an unprecedented exploration of cellular processes at the single-cell level. scRNA-seq has emerged as an indispensable asset in modern biology, shedding light on crucial aspects such as cellular diversity, gene regulation, and the mechanisms underlying various diseases [2, 3, 4, 5].

Since the advent of the first scRNA-seq study was published in 2009[6], a multitude of data analysis methods has emerged, addressing diverse aspects such as cell clustering[7, 8, 9, 10], rare cell type detection[11, 12, 13], differentially expressed gene identification[14, 15, 16], and trajectory inference[17, 18, 19, 20]. These methodologies are instrumental in uncovering specific patterns in scRNA-seq data. However, despite their value, a comprehensive understanding of the data features is still in its infancy, leaving researchers grappling with the challenge of formulating scRNA-seq data through simplistic rules.

*Corresponding author. E-mail: liuding@tiangong.edu.cn

†Corresponding author. E-mail: jzlei@tiangong.edu.cn

A prominent limitation of scRNA-seq data stems from the scarcity of available samples. The cost-intensive nature of single-cell RNA sequencing creates financial barriers, impeding widespread access. Additionally, ethical considerations and the limited availability of patients for specific diseases may curtail the clinical application of this technique. The resulting dearth of scRNA-seq samples compromises the overall depth of information and introduces uncertainty into downstream tasks, casting doubt on the reliability of obtained results[21].

To address these challenges, a promising avenue involves the development of computational methods to generate *in silico* datasets. Several simulators, such as scDesigns[22, 23], ZINB-WaVE[24], Splatter[25], SPARSim[26], scGAN[27], scVI[28], SymSim[29], have been designed to generate synthetic scRNA-seq datasets. These simulators employ various methodologies, including probability models[22, 24, 25, 26, 30, 23], kinetic models[29], and deep neural network models[28, 27]. scDesign2, for instance, is a transparent simulator capable of producing high-fidelity synthetic data with gene correlations[23]. It mimics the experimental scRNA-seq workflow steps, including cell isolation, RNA capture, library preparation, sequencing, and data analysis.

Deep learning techniques have been extensively applied in biology and medicine research[31, 32, 33]. In the realm of scRNA-seq, generative adversarial networks (GANs)[34] have been employed for data synthesis[35]. For instance, the conditional single-cell GAN (cscGAN) was developed to learn nonlinear gene-gene dependencies in various cell type samples, generating class-specific cellular scRNA-seq[27]. While GANs have been successful, their training process instability and susceptibility to model collapse pose challenges. Alternatively, Variational Auto-Encoder (VAE)[34] provides interpretability but may not guarantee high-quality scRNA-seq sample generation[36].

Denosing Diffusion Probabilistic Models (DDPM) present a compelling alternative, capable of fitting complex and nonlinear distributions[37]. Previous work attests to DDPM outperforms GAN in generative tasks[38]. While commonly applied in computer vision and natural language processing, DDPM’s success in bioinformatics applications[39], such as generating RNA tertiary structures[40], brain MRI scans[41], and biological sequences design[42], highlights its versatility. Previous works using diffusion models for generative tasks mostly use the UNet[43] neural network architecture as the skeleton[44]. scRDiT pioneered Diffusion Transformers (DiTs)[45], a method based on Vision Transformer (ViT)[46] that has been used in the field of computer vision, as the neural network architecture for diffusion models to synthesize scRNA-seq samples. The DiTs-based DDPM shows better performance than that of the UNet-based in this task. What’s more, we try zero-negation, a special data preprocessing method, which comprehensively considers the characteristics of diffusion model and scRNA-seq data, and further improves the quality of the synthesized samples.

Despite its efficiency, DDPM faces challenges in sample generation speed compared to VAE and GAN. Denosing Diffusion Implicit Models (DDIM), a more efficient iterative variant with the same training process as DDPM[47], addresses this limitation.

This paper introduces scRNA-seq Diffusion Transformer (scRDiT), a DiTs-based approach that leverages DDIM to generate high-quality scRNA-seq samples, striking a balance between efficiency and reliability. Specialized models for different cell types exhibit similarity to real samples, and DDIM accelerates the sampling process without unacceptable degradation in performance, achieving speedups of $10\times$ to $20\times$. Dimensionality reduction analysis and comparison with real data distributions using various metrics showcase the method’s commendable results and improved synthesis efficiency.

2 Materials and methods

In this section, we first briefly introduce the fundamentals of denosing diffusion probabilistic models (DDPMs) and then elucidate the architecture of our denosing network. Next, we present a general training method for this network, which can be used on different datasets. Finally, we provide a concise overview of the sampling accelerating method from denosing diffusion implicit models (DDIMs).

2.1 Data prepare

The numerical framework of scRDiT consists of two steps: (1) model training and (2) synthetic data generation. scRDiT trains a diffusion model based on a real scRNA-seq dataset in the model training step. If the dataset contains more than one cell type, scRDiT divides the dataset into subsets, one per cell type, and fits a cell type-specific model to each subset.

Before model training, scRDiT screens out hypervariable genes through the coefficient of variation (cv) of gene expression levels. Consider gene i and the vector $\mathbf{y}_i$ of its expression values in each cell, the coefficient of variation cv_i

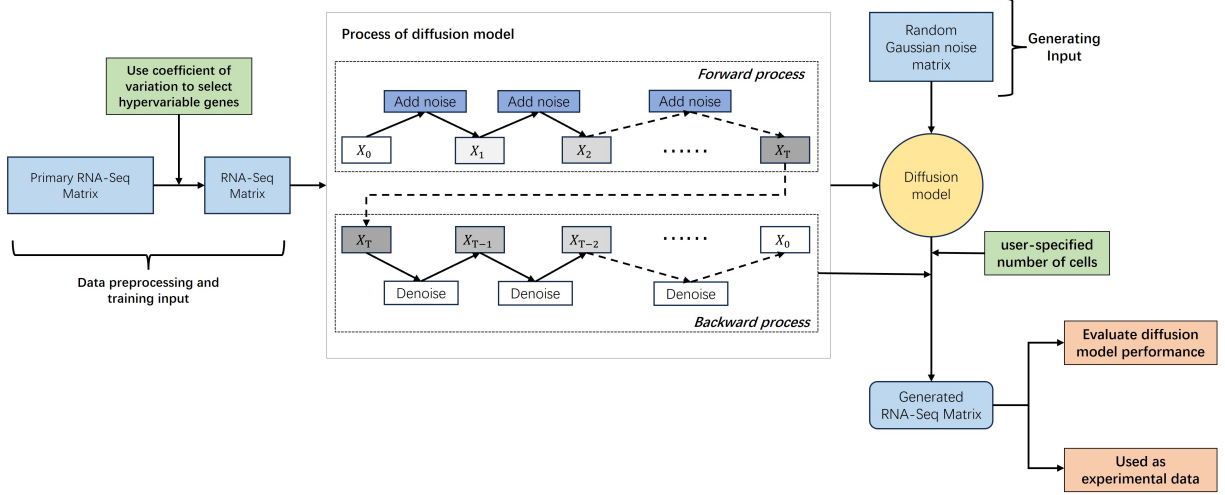


Figure 1: This figure shows how scRDiT process on scRNA-seq data. After data preprocessing, the real scRNA-seq data were fed into the model for training to fit the real data. After training, the user specifies the batch size of simulated data to be synthesized, and the model randomly generates Gaussian noise and reverts it to simulated scRNA-seq data. The generated simulated data can be used for downstream tasks or to test the reliability of the data.

for gene i is [48]:

$$cv_i = \frac{sd(\mathbf{y}_i)}{\text{mean}(\mathbf{y}_i)}, \quad (1)$$

where $sd(\mathbf{y}_i)$ represents the standard deviation of the components of vector $\mathbf{y}_i$, and $\text{mean}(\mathbf{y}_i)$ represents their mean. We arrange the coefficient of variation of all genes from large to small and select the largest 2000 genes as the gene set of model fitting. Filtering out the genes containing less information can reduce the dimensionality of the data and facilitate downstream analysis.

The denoising process of DDPMs has some limitations when it comes to generating specific discrete values. In particular, the generation of zero is crucial in scRNA-seq samples. When fitting directly to the raw data, it is possible for genes that we expect to be expressed as zero to fluctuate in a very small range around zero without any integer zero appearing in the synthetic samples. The issue can seriously impact the quality of the generated data.

To resolve this problem, we adopt a zero-negation, which change zero in the original data to a negative number n during training, so as to strengthen the characteristics of the data. (Practically, we take $n = -10$.) Finally, we truncate the data that were negative in the generated scRNA-seq sample to zero. Experiment results in section 3.2 show that this method significantly improves the reliability of the synthesized samples.

2.2 Diffusion models

The diffusion models are probabilistic models that generate target data samples from Gaussian noise by iterative denoising process. The diffusion model consists of two steps, the forward and reverse processes. The forward process is also called the diffusion process. Either the forward or backward process is a parameterized Markov chain, and the backward process can be used to generate data samples. The complete process of the diffusion model is shown at Figure 1.

Consider a real scRNA-seq sample $\mathbf{x}_0 \sim q(\mathbf{x}_0)$, where $q(\cdot)$ represents the distribution. The forward process iteratively adds Gaussian noise to the real sample. For each step t of the forward process, we add noise to the previous step $\mathbf{x}_{t-1}$ and get a new dataset $\mathbf{x}_t$. The dataset $\mathbf{x}_t$ can be represented as a weighted summation of $\mathbf{x}_{t-1}$ and a Gaussian noise $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Thus, the relationship between $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ is written as:

$$\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \mathbf{z}, \quad \text{where } \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (2)$$

The parameter α_t is defined as:

$$\alpha_t = 1 - \beta_t \quad (3)$$

where β_t is a predefined hyperparameter. The parameters $\{\beta_t\}_{1 \leq t \leq T}$ is an increase series so that

$$0 < \beta_1 < \beta_2 < \dots < \beta_T < 1.$$

From Eq.(2), $\mathbf{x}_t$ only depends on $\mathbf{x}_{t-1}$ at the previous time step. This process can be considered a Markov process:

$$\begin{aligned} q(\mathbf{x}_T|\mathbf{x}_0) &= \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}), \\ q(\mathbf{x}_t|\mathbf{x}_{t-1}) &= \mathcal{N}(\sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}) \end{aligned} \quad (4)$$

This process brings $\mathbf{x}_t$ closer and closer to pure noise as t increases. When $T \rightarrow \infty$, $\mathbf{x}_t$ becomes a Gaussian noise. The above iterative process yield the relationship between $\mathbf{x}_t$ and $\mathbf{x}_0$ as:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I}), \text{ where } \bar{\alpha}_t = \prod_{i=1}^t \alpha_i. \quad (5)$$

The reverse process of the diffusion model is the denoising process. We can restore the original distribution of $\mathbf{x}_0$ from the complete standard Gaussian distribution $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if we can obtain the reversed distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ step by step.

To obtain the reversed distribution, we use neural network $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ to fit the reverse process $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$. If $\mathbf{x}_0$ is known, the reverse process can be obtained through the Bayesian formula

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t; \mathbf{x}_0) = \mathcal{N}(\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{I}), \quad (6)$$

where $q(\mathbf{x}_{t-1}|\mathbf{x}_t; \mathbf{x}_0)$ means the conditional distribution $p(\mathbf{x}_{t-1}|\mathbf{x}_t)$ given the sample data $\mathbf{x}_0$, and

$$\tilde{\mu}_t = \frac{\sqrt{\bar{\alpha}_t}\beta_t}{1 - \bar{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_t)}{1 - \bar{\alpha}_t}\mathbf{x}_t, \text{ and } \tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}\beta_t.$$

Then, we can deduce

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mu_\theta(\mathbf{x}_t, t), \sigma_\theta(\mathbf{x}_t, t)), \quad (7)$$

where $\mu_\theta(\mathbf{x}_t, t)$ represents the predicted expectation expressed as

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right), \text{ and } \sigma_\theta(\mathbf{x}_t, t) = \sigma_t^2 \mathbf{I},$$

with $\epsilon_\theta(\mathbf{x}_t, t)$ the predicted noise perturbation.

If we ignore the variance of the distribution above and fit the mean with the neural network, we obtain the sampling process of the diffusion model as[37]:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}, \text{ where } \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (8)$$

Practically, we usually take $\sigma_t^2 = \beta_t$. Since t and $\mathbf{x}_t$ are known, we only need to fit $\epsilon_\theta(\mathbf{x}_t, t)$ with neural network.

2.3 Neural network architecture

We design a neural network architecture as the noise predictor with reference to the DiTs[45]. A scRNA-seq sample with Gaussian noise and a timestep t are fed into the model to obtain the predicted noise. For a noised sample with n genes as the input tensor of dimension $(n, 1)$, "Patchify" layer converts the tensor into a sequence of n/p tokens with convolutional layer, where p is a hyperparameter of patch size. Each token is a vector of length h , where hidden size h is also a hyperparameter that could be adjusted for different tasks. "Patchify" layer's output should be a tensor P of dimension $(n/p, h)$, and timestep embed layer will embed t into a tensor T of the same dimension. Pass T and P into the scalable Transformer unit with $N \times$ DiT blocks[45] and then to the final layer. The final output of the model is the predicted noise for timestep t .

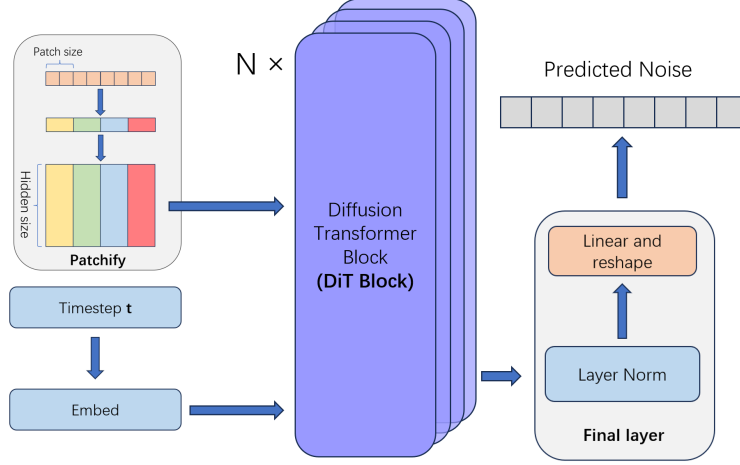


Figure 2: Illustration of the DiT architecture.

2.4 Training

Here, we show the training method of DiT. DiT requires multiple training epochs in order to achieve convergence. The number of required epochs is determined according to the dataset’s characteristics.

At the beginning of each epoch, we first draw a sample $x_0 \sim q(x_0)$ from the dataset. Randomly pick a time step t from $\{1, 2, \dots, T\}$ following

$$t \sim \text{Uniform}(\{1, \dots, T\}).$$

Next, we add Gaussian noises of time step t to the sample x_0 as:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon \quad (9)$$

where $\epsilon \sim \mathcal{N}(0, I)$, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. Then, we pass x_t and t to DiT as parameters. DiT outputs the predicted noise ϵ_θ through the given x_t and t . Use ϵ_θ and the true Gaussian noise ϵ to calculate the loss $\mathcal{L}$ function as[37]:

$$\mathcal{L} = \mathbb{E}_{t, x_0, \epsilon} \left[\|\epsilon - \epsilon_\theta(\sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon, t)\|^2 \right]. \quad (10)$$

At the end of each epoch, apply the backpropagation to calculate the gradient and update the weights of DiT. Repeat the steps above until the model converges.

2.5 Sampling acceleration

It is well-known that the denoising diffusion probabilistic model (DDPM) has a major drawback compared to most other generative models, which is its slow sampling speed. This is because DDPM typically requires 1000 denoising operations for each sampling process, resulting in a high time cost when using DDPM to generate sample data on low-performance machines. In the field of image generation, a common solution to this problem is to use the denoising diffusion implicit model (DDIM), a much faster sampling method. We applied this method to the task of generating scRNA-seq data and obtained satisfactory results.

The DDIM sampling method removes the Markov assumption of DDPM by enabling us to sample using a subset of time steps $\{1, 2, \dots, T\}$. This means that we can directly predict the outcome after several denoising steps following a particular denoising step. For instance, if the model starts denoising at $T = 1000$, a DDIM-based model can predict the outcome after 10 denoising steps, including $t = 990, 980, \dots, 10, 1$. This results in a ten-fold acceleration of the sampling process.

To be able to accelerate the sampling process, we need to change the sampling method of DDPM to:

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \left(\frac{x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_\theta}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2}\epsilon_\theta + \sigma_t\epsilon_t, \quad (11)$$

$$\sigma_t = \eta \sqrt{\frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}} \sqrt{1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}}, \quad \eta \in [0, 1], \quad (12)$$

where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is independent of $\mathbf{x}_t$, and ϵ_θ is the predicted noise of $\mathbf{x}_t$ and time step t . [47] Here, η is a hyper-parameter to adjust the acceleration efficiency. Particularly, when $\eta = 1$, DDIM is equivalent to DDPM.

To apply DDIM to accelerate the sampling process, we select a sub-sequence τ of sampling steps, where:

$$\tau = \{\tau_1, \tau_2, \dots, \tau_i, \dots\} \subset \{1, \dots, T\},$$

and formulate the sampling method (11) and (12) as:

$$\mathbf{x}_{\tau_{i-1}} = \sqrt{\bar{\alpha}_{\tau_{i-1}}} \left(\frac{\mathbf{x}_{\tau_i} - \sqrt{1 - \bar{\alpha}_{\tau_i}} \epsilon_\theta}{\sqrt{\bar{\alpha}_{\tau_i}}} \right) + \sqrt{1 - \bar{\alpha}_{\tau_{i-1}} - \sigma_{\tau_i}^2} \epsilon_\theta + \sigma_{\tau_i} \epsilon_{\tau_i}, \quad (13)$$

$$\sigma_{\tau_i} = \eta \sqrt{\frac{1 - \bar{\alpha}_{\tau_{i-1}}}{1 - \bar{\alpha}_{\tau_i}}} \sqrt{1 - \frac{\bar{\alpha}_{\tau_i}}{\bar{\alpha}_{\tau_{i-1}}}}, \quad \eta \in [0, 1]. \quad (14)$$

It's important to keep in mind that using the loss function $\mathcal{L}$ in the form of (10) during the training of neural networks, such as DiTs, as noise predictors, can lead to direct use of the training results in the sampling process of DDIM. This means that we can use the same neural network model, which was trained for DDPM, to speed up the sampling process by DDIM without having to retrain it [47].

3 Results

In implementing the DiT model structure as a noise predictor, we trained multiple models to generate scRNA-seq samples of the datasets. In addition, we tested different neural networks and conducted many comparisons to demonstrate the superiority of DiT architecture in this task. Using zero-negation preprocessed data as the training set (section 2.1), we enhance the adaptation of the diffusion model to scRNA-seq data. Two sampling methods, Denoising Diffusion Probabilistic Models (DDPM) and Denoising Diffusion Implicit Models (DDIM), were employed, and comprehensive comparative experiments were conducted. Notably, DDIM demonstrated effective acceleration of the sampling process. Furthermore, we tested various acceleration rates and compared the generated sample results from the models. The findings indicate that DDIM's sampling method is faster than that of DDPM and maintains high-quality samples. Importantly, the control over the model's performance loss remains within an acceptable range.

3.1 Training results

We applied the neural network architecture in section 2.3 to train two models for generating fibroblast and malignant scRNA-seq data, respectively. We employed scRDiT to assess its effectiveness on a tumor cell dataset (GSE103322), specifically consisting of 5902 cells derived from 19 patients with oral cavity head and neck squamous cell carcinoma [49]. The dataset was partitioned into two subsets: 2215 malignant tumor cells and 1422 fibroblasts. Usually, different epochs are required for different datasets due to the variation in data features. For the models trained with different epochs, we generated a set of virtual data through DDPM's sampling method and compared the similarity between the generated data and real data. Various measurements were used to check the sample quality, including the KL divergence, Wasserstein distance, and maximum mean discrepancy (MMD). The results are shown in Table 3.1 and Figure 3-b.

We use t-SNE dimensionality reduction [50] to reduce real and generated data samples to two dimensions and make scatter plots. The fitting process for different kinds of cells are shown in Figure 3-a. The result shows that as the number of training epochs increases, the data distribution of the generated samples becomes closer to the real samples.

Moreover, we used two data sets for model training to validate the method's generality. The first comprised 407 immune cells from the GSE103322 dataset [49], while the second incorporated a dataset (GSE181297) of 1936 keloid cells [51]. The results are also shown in Table 3.1 and Figure 3-b.

The trained model is robust in generating sampling data. We generated 5k scRNA-seq samples for each category with the best-performing models. The results of t-SNE dimensionality reduction for all samples are shown in Figure 3-c. The real data and the generated samples show good consistency, which implies the robustness and stability of our method.

3.2 Comparison

The previous work mostly use UNet [43] architecture as the noise predictor [44] for DDPMs. In order to verify the superiority of DiT [45] network architecture, we tested different neural network architectures to synthesize scRNA-seq samples. The performance comparison in Table 3.2 shows that DiT is able to generate the most plausible samples.

Table 1: Sample qualities of different cell types. The models are trained with different training epochs. The qualities are measured with different methods: KL divergence, Wasserstein distance, and MMD. (Bolded numbers for the best results.)

single-cell RNA-seq type	fibroblast				malignant			
epoch	200	400	600	800	400	800	1200	1600
KL Divergence	4.570	4.096	4.016	3.971	3.798	3.117	3.648	3.179
Wasserstein Distance	0.052	0.047	0.052	0.026	0.061	0.045	0.023	0.007
MMD	0.128	0.066	0.084	0.022	0.077	0.070	0.055	0.046
single-cell RNA-seq type	keloid cells				immune cells			
epoch	200	400	600	800	400	800	1200	1600
KL Divergence	0.149	0.145	0.143	0.151	6.455	7.208	7.405	7.357
Wasserstein Distance	0.011	0.008	0.008	0.005	0.120	0.023	0.015	0.043
MMD	0.047	0.036	0.034	0.015	0.116	0.071	0.029	0.041

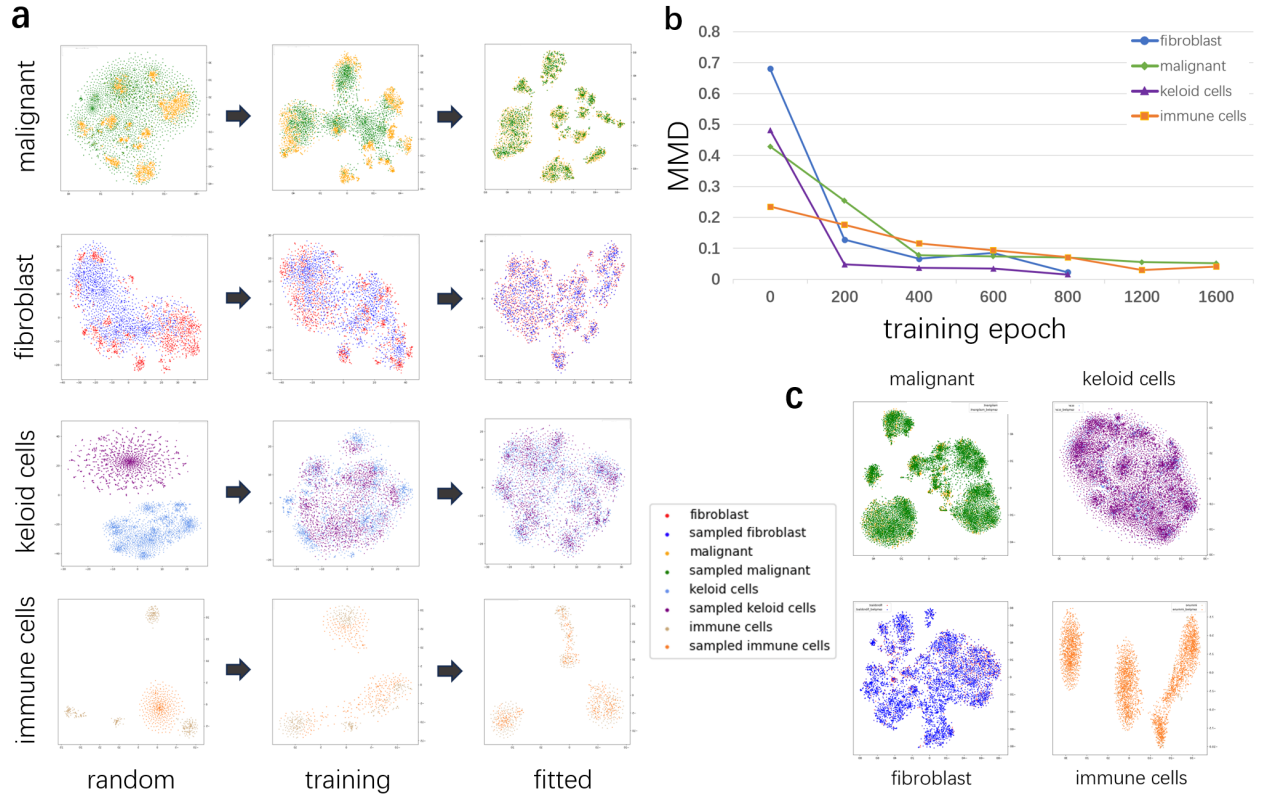


Figure 3: scRDiT generating scRNA-seq data. **a.** The fitting process of the model. t-SNE plots of real data and generated scRNA-seq samples from beginning of training to the final performance. **b.** Dependence of MMD with the increase of training epochs. **c.** t-SNE plots of virtual scRNA-seq samples and real data.

UNet models require much more training epochs to fit the real data than DiT models (Figure4-a) . It is worth mentioning that each training epoch and each sampling process of UNet models is much slower than that of DiT models.

Most methods that apply deep learning to generate scRNA-seq samples obtain results that have problems in the simulation of coefficient of variation and zero proportion. Preprocessing the training data using zero-negation, introduced in section2.1, can effectively solve this problem. The results of this method on different type of samples are shown in Figure4-b.

Table 2: Performance of different neural networks on the same scRNA-seq datasets. scRDiT-zn using special preprocessing method, zero-negation, on the training set and obtained the best results on each type of samples.

Model	MMD			
	fibroblast	malignant	keloid cells	immune cells
UNet	0.039	0.066	0.036	0.054
scRDiT	0.065	0.053	0.028	0.061
scRDiT-zn	0.022	0.046	0.015	0.030

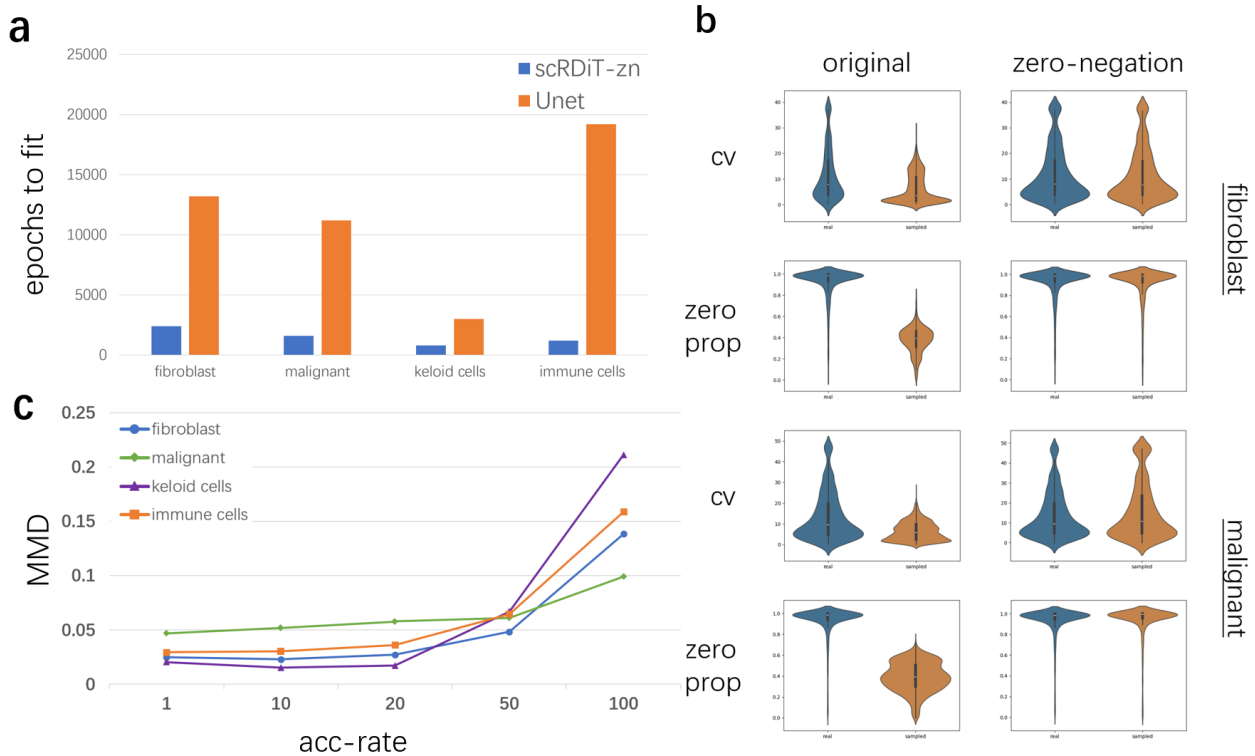


Figure 4: Comparison with scRDiT and UNet. **a**. Number of epochs to fit real data based on DiT and UNet. **b**. Violin plots for the coefficient of variation (cv) and zero proportion (zero prop) of scRNA-seq data for the two methods. **c**. Effect of different acceleration rates on the quality of the synthesized samples.

3.3 Sampling acceleration

To delve deeper into the acceleration method outlined in Section 2.5, we implemented the DDIM sampling approach in the sampling processes. We specifically opted for various time-step sub-sequences to attain different acceleration rates. The qualities of the sampled data are detailed in Table 3.3. Additionally, the variation of the models' performance (measured with MMD between real data and sampled data) under different acceleration rates is shown in Figure 4-c. The outcomes indicate that the quality of the sampled data remains consistent, while the sampling speeds experience a noticeable increase. It is possible to greatly speed up the generation of scRNA-seq samples using the DDIM sampling method. However, increasing acceleration beyond a certain point can have a negative impact on the quality of generated samples. In our experiments, we found that an acceleration rate of 10 to 20 times strikes a balance between time consumption and sample quality.

For instance, generating 2k scRNA-seq samples using an NVIDIA RTX4060 GPU device typically takes about 50 minutes without acceleration, while it only takes less than 5 minutes with 10 $\times$ acceleration. This means that by using the accelerated sampling method, we can generate many samples at an acceptable time cost on a low-performance device while ensuring time cost on a low-performance device while ensuring high sample quality. This partly overcomes the problem that the diffusion model is slower to do generative tasks than other generative models such as GAN.

Furthermore, we found that using non-equidistant time steps in sub-sequences for sampling can further accelerate the process while maintaining high sample quality.

Table 3: The similarity between the generated scRNA-seq samples and the real samples under different acceleration rates.

scRNA-seq type	fibroblast					malignant				
Acc-Rate	1	10	20	50	100	1	10	20	50	100
KL Divergnece	4.057	3.971	3.988	3.835	3.294	3.179	3.137	3.117	3.052	2.960
Wasserstein Distance	0.037	0.026	0.026	0.028	0.072	0.007	0.007	0.007	0.012	0.029
MMD	0.024	0.022	0.027	0.048	0.138	0.046	0.051	0.057	0.061	0.099

scRNA-seq type	keloid cells					immune cells				
Acc-Rate	1	10	20	50	100	1	10	20	50	100
KL Divergnece	0.157	0.151	0.149	0.132	0.127	7.405	7.097	7.230	6.995	6.629
Wasserstein Distance	0.006	0.005	0.006	0.012	0.020	0.015	0.012	0.015	0.029	0.039
MMD	0.020	0.015	0.017	0.066	0.211	0.029	0.030	0.036	0.064	0.158

4 Summary

In this paper, we propose a novel approach for generating scRNA-seq samples, utilizing the diffusion model. This method provides a viable solution to the challenge of limited available scRNA-seq samples. Employing the DiTs structure as a reference, we trained neural networks to serve as noise predictors, and achieved better results than several other neural networks. Our model effectively reconstructs simulated scRNA-seq samples from Gaussian noise. To enhance synthesis efficiency and address the prolonged sampling processes of DDPM, we leverage DDIM. In summary, our method demonstrates superior performance in generating synthetic scRNA-seq samples akin to the datasets we utilized, exhibiting excellent stability and speed.

Acknowledgments

DL is supported by the Tianjin Natural Science Foundation of China (20JCYBJC00500), the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (2018KJ217). JL is supported by the National Natural Science Foundation of China (12331018)

References

- [1] Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 10(1):57–63, 2009.
- [2] Samuel Marguerat and Jürg Bähler. RNA-seq: from technology to biology. *Cell. Mol. Life Sci.*, 67:569–579, 2010.
- [3] Shihao Shen, Juwon Park, Zhi-xiang Lu, Lan Lin, Michael D Henry, Ying Nian Wu, Qing Zhou, and Yi Xing. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA*, 111(51):E5593–E5601, 2014.
- [4] Davis J McCarthy, Yunshun Chen, and Gordon K Smyth. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.*, 40(10):4288–4297, 2012.
- [5] Zhiyu Peng, Yanbing Cheng, Bertrand Chin-Ming Tan, Lin Kang, Zhijian Tian, Yuankun Zhu, Wenwei Zhang, Yu Liang, Xueda Hu, Xuemei Tan, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.*, 30(3):253–260, 2012.
- [6] Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, Kaiqin Lao, and M Azim Surani. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods*, 6(5):377–382, 2009.
- [7] Vladimir Yu Kiselev, Kristina Kirschner, Michael T Schaub, Tallulah Andrews, Andrew Yiu, Tamir Chandra, Kedar N Natarajan, Wolf Reik, Mauricio Barahona, Anthony R Green, and Martin Hemberg. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods*, 14(5):483–486, May 2017.

- [8] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M 3rd Mauck, Yuhao Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902, Jun 2019.
- [9] Yuqi Tan and Patrick Cahan. SingleCellNet: A computational tool to classify single cell RNA-seq data across platforms and across species. *Cell Syst.*, 9(2):207–213, Aug 2019.
- [10] Hannah A Pliner, Jay Shendure, and Cole Trapnell. Supervised classification enables rapid annotation of cell atlases. *Nat. Methods*, 16(10):983–986, Oct 2019.
- [11] Nelson Johansen and Gerald Quon. scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. *Genome Biol.*, 20(1):166, Aug 2019.
- [12] Daphne Tsoucas and Guo-Cheng Yuan. GiniClust2: a cluster-aware, weighted ensemble clustering method for cell-type detection. *Genome Biol.*, 19(1):58, May 2018.
- [13] Aashi Jindal, Prashant Gupta, Jayadeva, and Debarka Sengupta. Discovery of rare cells from voluminous single cell expression data. *Nat. Commun.*, 9(1):4719, Nov 2018.
- [14] Dongyuan Song and Jingyi Jessica Li. PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data. *Genome Biol.*, 22(1):124, Apr 2021.
- [15] Peter V Kharchenko, Lev Silberstein, and David T Scadden. Bayesian approach to single-cell differential expression analysis. *Nat. Methods*, 11(7):740–742, Jul 2014.
- [16] Koen Van den Berge, Hector Roux de Bézieux, Kelly Street, Wouter Saelens, Robrecht Cannoodt, Yvan Saeys, Sandrine Dudoit, and Lieven Clement. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.*, 11(1):1201, Mar 2020.
- [17] Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.*, 32(4):381–386, Apr 2014.
- [18] Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A Pliner, and Cole Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*, 14(10):979–982, Oct 2017.
- [19] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1):477, Jun 2018.
- [20] Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M Ibrahim, Andrew J Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J Steemers, Cole Trapnell, and Jay Shendure. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745):496–502, Feb 2019.
- [21] Katherine S Button, John PA Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma SJ Robinson, and Marcus R Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.*, 14(5):365–376, 2013.
- [22] Wei Vivian Li and Jingyi Jessica Li. A statistical simulator scDesign for rational scRNA-seq experimental design. *Bioinformatics*, 35(15):i41–i50, 2019.
- [23] Tanyi Sun, Dongyang Song, Wei Vivian Li, and Jingyi Jessica Li. scDesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. *Genome Biol.*, 22:163, 2022.
- [24] Davide Risso, Fanny Perraudeau, Svetlana Gribkova, Sandrine Dudoit, and Jean-Philippe Vert. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.*, 9:284, 2018.
- [25] Luke Zappia, Belinda Phipson, and Alicia Oshlack. Splatter: simulation of single-cell RNA sequencing data. *Genome Biol.*, 18:174, 2017.
- [26] Giacomo Baruzzo, Ilaria Patuzzi, and Barbara Di Camillo. SPARSim single cell: a count data simulator for scRNA-seq data. *Bioinformatics*, 36(5):1468–1475, 2020.
- [27] Mohamed Marouf, Pierre Machart, Vikas Bansal, Christoph Kilian, Daniel S Magruder, Christian F Krebs, and Stefan Bonn. Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. *Nat. Commun.*, 11(1):166, 2020.
- [28] Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nat. Methods*, 15:1053–1058, 2018.
- [29] Xiuwei Zhang, Chenling Xu, and Nir Yosef. Simulating multiple faceted variability in single cell RNA sequencing. *Nat. Commun.*, 19:2611, 2019.

- [30] Keegan D. Korthauer, Li-Fang Chu, Michael A. Newton, Yuan Li, James Thomson, Ron Stewart, and Christina Kendzioriski. A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol.*, 17:222, 2016.
- [31] Sarah Webb et al. Deep learning for biology. *Nature*, 554(7693):555–557, 2018.
- [32] Shuting Jin, Xiangxiang Zeng, Feng Xia, Wei Huang, and Xiangrong Liu. Application of deep learning methods in biological networks. *Brief. Bioinformatics*, 22(2):1902–1917, 2021.
- [33] Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, et al. Opportunities and obstacles for deep learning in biology and medicine. *J. Roy. Soc. Interface*, 15(141):20170387, 2018.
- [34] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Commun. ACM*, 63(11):139–144, 2020.
- [35] Yinglun Wang, Qiurui Chen, Hongwei Shao, Rongxin Zhang, and Han Shen. Generating bulk RNA-Seq gene expression data based on generative deep learning models and utilizing it for data augmentation. *Comput. Biol. Med.*, 169:107828, 2024.
- [36] Valentine Svensson, Adam Gayoso, Nir Yosef, and Lior Pachter. Interpretable factor models of single-cell RNA-seq via variational autoencoders. *Bioinformatics*, 36(11):3418–3421, 2020.
- [37] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.*, 33:6840–6851, 2020.
- [38] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Adv. Neural Inf. Process. Syst.*, 34:8780–8794, 2021.
- [39] Zhiye Guo, Jian Liu, Yanli Wang, Mengrui Chen, Duolin Wang, Dong Xu, and Jianlin Cheng. Diffusion models in bioinformatics: A new wave of deep learning revolution in action. *arXiv*, page 2302.10907, 2023.
- [40] Mihir Bafna, Vikranth Keerthipati, Subhash Kanaparthi, and Ruochi Zhang. DiffRNAFold: Generating RNA tertiary structures with latent space diffusion. In *Deep Generative Models for Health Workshop NeurIPS 2023*, 2023.
- [41] Benjamin L. Kidder. Advanced image generation for cancer using diffusion models. *bioRxiv*, page 2023.08.18.553859, 2023.
- [42] Pavel Avdeyev, Chenlai Shi, Yuhao Tan, Kseniia Dudnyk, and Jian Zhou. Dirichlet diffusion score model for biological sequence generation. *arXiv*, page 2305.10699, 2023.
- [43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [44] Yixuan Wang, Shuangyin Li, Shimin DI, and Lei Chen. Single-cell RNA-seq synthesis with latent diffusion model. *arXiv*, page 2312.14220, 2023.
- [45] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4195–4205, 2023.
- [46] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [47] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv*, page 2010.02502, 2020.
- [48] George F Reed, Freyja Lynn, and Bruce D Meade. Use of coefficient of variation in assessing variability of quantitative assays. *Clin. Vaccine Immunol.*, 9(6):1235–1239, 2002.
- [49] Sidharth V Puram, Itay Tirosh, Anuraag S Parikh, Anoop P Patel, Keren Yizhak, Shawn Gillespie, Christopher Rodman, Christina L Luo, Edmund A Mroz, Kevin S Emerick, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, 171(7):1611–1624, 2017.
- [50] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9(11), 2008.
- [51] Joonho Shim, Se Jin Oh, Eunhye Yeo, Ji Hye Park, Jai Hee Bae, Seok-Hyung Kim, Dongyoun Lee, and Jong Hee Lee. Integrated analysis of single-cell and spatial transcriptomics in keloids: highlights on fibrovascular interactions in keloid pathogenesis. *J. Invest. Dermatol.*, 142(8):2128–2139, 2022.