# Predicting Drug-Induced Transcriptional Responses Using Latent Diffusion Model

Chaewon Kim[a], Sunyong Yoo[a,b*]
[a] Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju, Republic of Korea
[b] R&D Center, MATILO AI Inc., Gwangju, Republic of Korea

[*]Correspondence: syyoo@jnu.ac.kr; Tel: +82-62-530-1761

## Abstract

Accurate prediction of drug-induced transcriptional responses is essential for drug discovery and precision medicine. Existing computational models, including encoder–decoder architectures and generative adversarial network-based approaches, achieve reasonable accuracy but often fail to account for gene–gene correlations and generalize to unseen conditions. Here, we present a latent diffusion model that combines a variational autoencoder (VAE) with a diffusion process. The VAE compresses gene expression (GE) profiles into a low-dimensional latent space, where the diffusion process learns the joint probability distribution of latent GE representations and their noisy intermediates. Learning these distributions allow the model to capture gene–gene correlations more effectively. Moreover, our model incorporates multiple perturbation conditions—including cell line, compound, dose, and time—to enhance generalization performance on unseen conditions. The reverse diffusion process is designed to predict both the mean and variance of the latent representations, which robustly enhances the fidelity of the generated GE profiles. The proposed model demonstrated the highest accuracy in reconstructing perturbed GE profiles compared to previous studies, achieving a root mean squared error (RMSE) of 1.340, a Pearson correlation coefficient of 0.832 and an $R^2$ score of 0.669. In addition, the proposed model demonstrated superior performance in preserving gene–gene correlation, as shown by correlation heatmaps, compared to existing approaches. To evaluate the biological relevance of generated transcriptional profiles, we conducted a half-maximal inhibitory concentration prediction task using the generated profiles as model inputs. Our model outperformed the baseline methods, achieving a RMSE of 1.335 and $R^2$ score of 0.819. In conclusion, we demonstrated the potential of diffusion-based generative models as reliable and versatile tools for modeling transcriptional responses, with implications for drug discovery and precision medicine applications.

**Keywords:** Transcriptional responses prediction; Latent diffusion model; Drug-induced gene expression; Gene–gene correlation;

# 1. Introduction

Drug-induced cellular response refers to the physiological and biochemical changes that cells undergo in reaction to external signals, such as drug treatment [1]. These cellular responses provide crucial insights for disease treatment, drug development, and precision medicine [2, 3]. Accurate prediction of cellular responses can help reduce the enormous costs associated with drug development, predict potential adverse effects in advance, and improve drug safety.

Transcriptome-level changes have been utilized as representative metrics to characterize cellular states and their directional changes [4, 5]. GE patterns at the transcriptomic level reveal how drugs activate or inhibit biological pathways [1, 6], enabling quantitative analysis of the complex characteristics of drug-induced cellular responses and gene−gene interactions [4, 7]. Recent advances in high-throughput screening (HTS) technology have enabled the rapid acquisition of large-scale transcriptomic response data [8], leading to the establishment of comprehensive transcriptomic databases such as the Library of Integrated Network-Based Cellular Signatures (LINCS) L1000 project [9]. However, it is impractical to experimentally obtain transcriptomic response data for a large number of drug–cell combinations [3]. Due to this limitation, there has been increasing effort to predict drug-perturbed GE using computational methods.

Previous studies have predicted drug-perturbed GE using various approaches such as encoder−decoder architectures and generative adversarial networks (GANs). Encoder–decoder-based approaches effectively predicted drug-perturbed GE using cell line features and chemical features such as molecular graphs and SMILES representations [10, 11]. In particular, PRnet [11] achieved remarkable performance in predicting drug-perturbed GE using unperturbed GE and functional-class fingerprints as input features. However, these models fail to sufficiently capture gene–gene correlations, as they were not designed to consider such covariation structures. Since gene–gene correlations reflect the biological pathways in cells that drive drug-induced cellular responses [12, 13], failing to capture these correlations undermines the biological validity of predictions. GAN-based approaches employ generator networks with strong modeling capabilities to generate predictions [14-16]. Nevertheless, most studies do not fully incorporate perturbation condition features, which limit their generalization ability. Furthermore, their capacity to capture gene–gene correlations remains limited and largely implicit, without explicitly modelling or evaluating such correlations.

To address these limitations, diffusion models have emerged as a promising approach [17-19]. Unlike encoder–decoder or GAN-based approaches, diffusion models learn the joint distribution of complex data by progressively denoising random noise through probabilistic sampling processes [20]. This distributional modeling allows them to capture the covariance structure among genes, thereby preserving gene–gene correlations. Recently, PertDiT [21] has demonstrated the feasibility of applying diffusion model to predict drug-perturbed GE by incorporating transformer-based denoisers. PertDiT also has achieved state-of-the-art (SOTA) performance using unperturbed GE and text embedding vector of drugs as condition features in denoiser. However, since PertDiT performs the diffusion process directly in the high-dimensional GE space, it incurs a significant computational cost and training instability. Moreover, PertDiT's denoiser predicts only the means of the noise distribution while fixing the variance to a constant value, which limits its ability to capture the full distributional structure

of transcriptomic responses. In contrast, subsequent research on diffusion models has demonstrated that learning both mean and variance generally improves the variational lower bound and facilitates log-likelihood optimization [22]. Log-likelihood serves as a key metric for evaluating generative model quality, and its optimization enables models to capture diverse modes of the data distribution more effectively [23]. This suggests that modeling both mean and variance in diffusion model denoisers is crucial for capturing the distributional structure of real data and enhancing generation quality.

Among diffusion model variants, latent diffusion models (LDMs) offer particular advantages for high-dimensional data. LDMs perform the diffusion process in a compressed latent space rather than the original data space, significantly reducing computational costs while maintaining generation quality [24]. This latent space representation enables more efficient learning of complex data distributions and provides better control over the generation process. For the LINCS L1000, which involves approximately 1,000 landmark genes with complex correlation structures, the dimensional reduction inherent in LDMs can help capture essential biological patterns while avoiding the computational burden of operating in the full GE space.

Therefore, this study proposes an approach for predicting drug-perturbed GE based on LDM. This approach encodes GE into a low-dimensional latent space using VAE, enabling stable learning of complex GE patterns while improving computational efficiency. Furthermore, by predicting both mean and variance of latent vectors during the reverse diffusion process, the model generates more robust and high-fidelity GE. To demonstrate the biological relevance of generated GE, we performed the half-maximal inhibitory concentration ($IC_{50}$) predictions using GE generated as different models, including the proposed LDM. The objectives of this study are as follows: (i) to generate GE that faithfully reproduces the gene−gene correlation structure of real data, and (ii) to achieve high generalization performance under unseen perturbation conditions, including novel cell lines and drugs not present in the training data.

# 2. Methods

## 2.1 Model Architecture

This study employs an LDM framework to generate drug-perturbed GE. The overall architecture is presented in Fig. 1. The model consists of two main components, a VAE and a diffusion model. Initially, the VAE is pretrained on perturbed GE to obtain a stable latent representation. The pretrained VAE encoder then maps perturbed GE onto latent space, where the diffusion model performs training and inference. During the reverse diffusion process, the denoiser model predicts both the mean and variance of the latent distribution, thereby enabling more robust GE generation. Furthermore, diffusion models learn to approximate complex multivariate distributions through iterative denoising process, allowing them to capture gene–gene correlation more comprehensively compared to deterministic models. Finally, by integrating diverse perturbation conditions during training, the model learns a condition-aware latent representation which enhances generalization performance on unseen conditions not present in the training data.

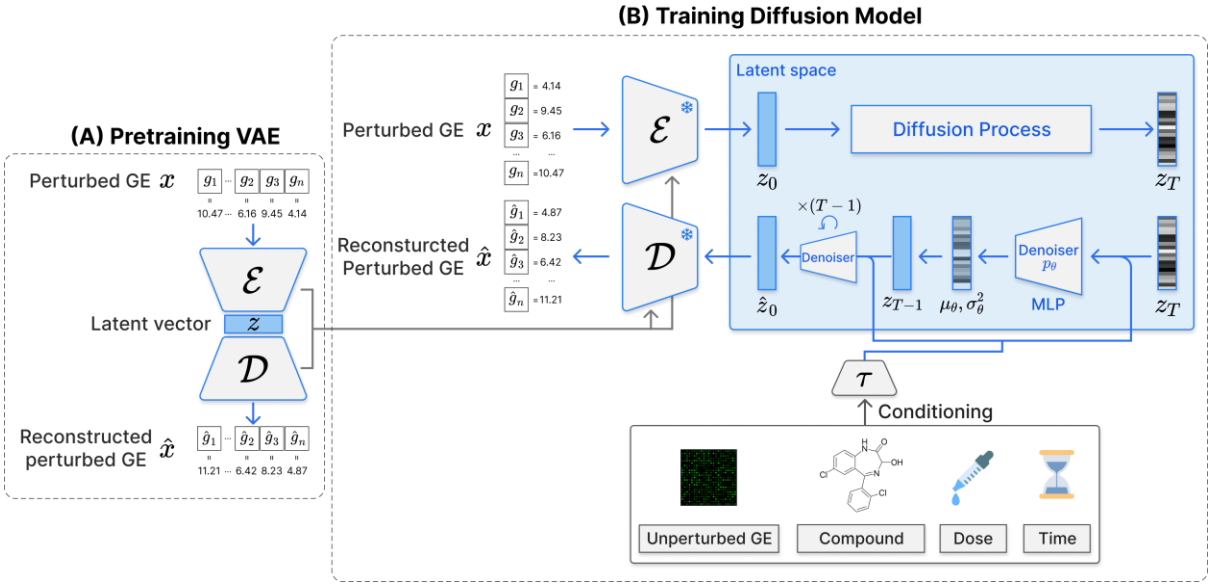


**Fig. 1.** Overview of model architecture. **A.** VAE is pretrained on perturbed GE profiles. The encoder maps perturbed GE onto a latent space, from which the decoder can accurately reconstruct the original input. **B.** Diffusion model is trained in latent space provided by the pretrained VAE. The schematic illustrates both the forward and reverse diffusion process, where the denoiser predicts the mean and variance of latent representations at each timestep. Four conditioning features are incorporated in reverse diffusion process: unperturbed GE, compound structure, treatment dose, and treatment time.

## 2.2 Data Preprocessing

This study used the LINCS L1000 dataset, which provides GE data from genetic (shRNA, cDNA) and chemical (small molecules) perturbations across various cell lines, compounds, time points, and dose conditions [9]. The complete dataset contains GE for 978 landmark genes and 11,350 additional genes. While the expression of 978 landmark genes was experimentally measured, the 11,350 additional genes were computationally imputed using a transformation matrix [1]. To ensure data reliability, this study used only landmark gene data.

The dataset is available in three versions: phase I, II, and beta, with each version categorized into five levels according to the data processing pipeline. We used level 3 quantile-normalized GE profiles from phase I. The following data cleaning procedures were applied: (i) removal of compounds appearing fewer than 5 times in the entire datasets; (ii) removal of compounds with invalid SMILES strings that could not be successfully parsed by RDKit [25]; (iii) pairing of unperturbed and perturbed observations. The final dataset comprised 836,841 observations across 82 cell lines and 17,766 compounds, covering diverse treatment dose and time conditions.

## 2.3 Variational Autoencoder Pretraining

Perturbed GE is represented as a 978-dimensional vector. Training a diffusion model directly on this high-dimensional space is computationally expensive and may lead to unstable

optimization [24]. To address this challenge, we employed a VAE to map GE profiles onto a lower-dimensional latent space, where the diffusion model is subsequently trained.

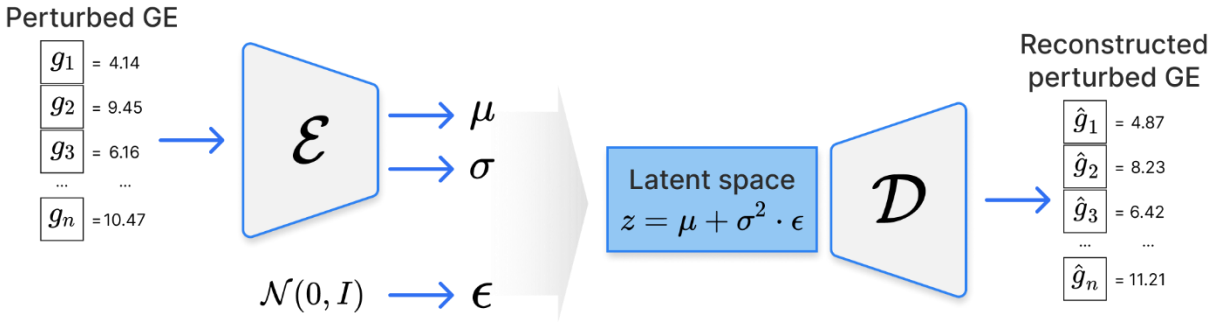


**Fig. 2.** Structure of variational autoencoder. The encoder maps perturbed GE profiles onto a latent space parameterized by mean and variance. The decoder reconstructs the original profiles from latent samples obtained via the reparameterization trick.

The VAE follows an encoder-decoder architecture [26]. Given an input GE vector $\mathbf{x}$, the encoder approximates the posterior distribution of the latent representation $\mathbf{z}$, denoted as $q_\phi(\mathbf{z}|\mathbf{x})$, by predicting its mean $\mu_{VAE}$ and log-variance $\log \sigma^2_{VAE}$. The decoder reconstructs the original GE profile from the latent representation by modeling the conditional distribution $p_\psi(\mathbf{x}|\mathbf{z})$.

The VAE is trained by maximizing the evidence lower bound (ELBO) [26]:

$$\text{ELBO} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\psi(\mathbf{x}|\mathbf{z})\right] - \mathrm{D}_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})), \tag{1}$$

where $p(\mathbf{z})$ is the prior distribution, assumed to be a standard Gaussian $\mathcal{N}(0, \mathbf{I})$.

In practice, the VAE loss function is:

$$\mathcal{L}_{\text{VAE}} = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2 + k \cdot \frac{1}{N}\sum_{i=1}^{N}(\mu_{VAE,i}^2 + \sigma_{VAE,i}^2 - \log \sigma_{VAE,i}^2 - 1), \tag{2}$$

where the first term denotes the reconstruction loss, which is measured in terms of the mean squared error. $N$ means the number of samples. The second term is the Kullback-Leibler divergence that regularizes the latent distribution toward the prior [27]. $\mu_{VAE,i}$ and $\sigma^2_{VAE,i}$ denote the mean and variance of latent representations of $i^{\text{th}}$ sample, respectively. $k$ is the KL term weight which set to 1 in our experiments.

## 2.4 Diffusion Model Training

GE follows a high-dimensional and complex distribution that reflects nonlinear interactions among numerous genes. Therefore, we adopted a diffusion model, which has been shown to outperform existing generative approaches in terms of generation quality and diversity [20]. A

diffusion model represents complex data distributions through a Markovian sequence of Gaussian noising and denoising steps, allowing for flexible and stable modeling of multimodal biological data. This probabilistic formulation can enhance the biological plausibility and consistency of the predicted GE, as it naturally captures stochastic variation inherent in GE [28].

### 2.4.1 Forward diffusion

Given the latent representation $\mathbf{z}_0$ obtained from the pretrained VAE encoder, Gaussian noise is gradually injected over $T$ timesteps. At each timestep $t$, the noisy latent representation is sampled according to the following distribution:

$$q(\mathbf{z}_t|\mathbf{z}_{t-1}) = \mathcal{N}\left(\mathbf{z}_t; \sqrt{1-\beta_t}\mathbf{z}_{t-1}, \beta_t \cdot \mathbf{I}\right), \tag{3}$$

where $\beta_t \in (0,1)$ denotes the noise scale parameter at timestep $t$. We employed a linear noise scheduler:

$$\beta_t = \beta_{\min} + \frac{t-1}{T-1}(\beta_{\max} - \beta_{\min}), \tag{4}$$

where $\beta_{\min}$ and $\beta_{\max}$ specify the starting and ending noise levels of the linear schedule that set to $1 \times 10^{-5}$ and 0.01, respectively.

Equivalently, the forward process can be expressed in closed form as:

$$q(\mathbf{z}_t|\mathbf{z}_0) = \mathcal{N}\left(\mathbf{z}_t; \sqrt{\bar{\alpha}_t}\mathbf{z}_0, (1-\bar{\alpha}_t)\mathbf{I}\right), \tag{5}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \Pi_{s=1}^{t}\alpha_s$. This forward process generates a sequence of increasingly noisy latent representations that approaches a standard Gaussian distribution as $t \to T$.

### 2.4.2 Reverse diffusion

The reverse diffusion process progressively denoises the latent representations by modeling the reverse conditional distribution $q(\mathbf{z}_{t-1}|\mathbf{z}_t)$. Since this distribution is intractable without knowledge of the entire data distribution, we approximate it with a parameterized neural network:

$$p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{c}) = \mathcal{N}\left(\mu_\theta(\mathbf{z}_t, t, \mathbf{c}), \Sigma_\theta(\mathbf{z}_t, t, \mathbf{c})\right), \tag{6}$$

where the condition vector $\mathbf{c}$ encodes basal GE, compound structure, treatment dose, and treatment time.

Basal GE profiles for each cell line are embedded using a multilayer perceptron (MLP). Compound features are extracted by embedding molecular SMILES strings with Molformer [29], a transformer-based molecular representation model with rotary position embeddings. Treatment dose and time are embedded via independently parameterized MLPs. While dose effects have been extensively studied due to their strong influence on GE changes, treatment

time has received less attention despite its biological importance. Since cellular metabolism and signaling pathways evolve temporally, treatment time represents a critical factor influencing GE [30]. Therefore, we explicitly incorporate time features. The resulting embeddings are concatenated and compressed into the condition vector $\mathbf{c}$ through the MLP $\tau$.

The denoising network takes three inputs: the noisy latent variable $\mathbf{z}_t$, embedding vector of timestep $t$, and condition vector $\mathbf{c}$. It predicts both the mean $\mu_\theta$ and interpolation vector $s \in [0,1]^d$ for log variance estimation where $d$ means latent dimensionality. To ensure training stability, we avoid direct prediction of the variance. Instead, we adopt the interpolation method from the improved diffusion model approach [22], using $s$ to compute the covariance matrix $\Sigma_\theta$ for $\mathbf{z}_{t-1}$ as follows:

$$\Sigma_\theta(\mathbf{z}_t, t, \mathbf{c}) = \exp(s \cdot \log(\beta_t) + (1-s)\log\tilde{\beta}. \tag{7}$$

This interpolation occurs between an upper bound, the forward process variance $\beta_t\mathbf{I}$, and a lower bound, the posterior variance $\tilde{\beta}_t\mathbf{I}$, where $\tilde{\beta} = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. By constraining the predicted variance within this stable range, this technique prevents the numerical issues associated with learning the variance directly.

At each reverse time step, $\mathbf{z}_{t-1}$ is sampled from the predicted Gaussian distribution that is composed with $\mu_\theta$ and $\log\Sigma_\theta$. After $T$ denoising steps, the final latent representation $\hat{\mathbf{z}}_0$ is decoded by the pretrained VAE decoder to reconstruct the perturbed GE profile $\hat{\mathbf{x}}$.

The diffusion model is trained to predict both the mean and log-variance of the reverse distribution $p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{c})$. Given the true posterior from the forward process, we employ the negative log-likelihood (NLL) of a Gaussian distribution. For a predicted Gaussian distribution with mean $\mu_\theta$ and log-variance $\log\Sigma_\theta$, the negative log-likelihood of the target $\mathbf{z}_{t-1}$ is:

$$\mathcal{L}_{var} = \frac{1}{2}\mathbb{E}\left[\frac{1}{\Sigma_\theta} \odot (\mathbf{z}_{t-1} - \mu_\theta)^2 + \log\Sigma_\theta\right], \tag{8}$$

where $\odot$ denotes element-wise multiplication.

This objective can be interpreted as the sum of two complementary components. The first is a precision weighted mean squared error, which penalizes prediction errors more strongly when the model assigns low variance, thereby encouraging accurate reconstructions under confident predictions. The second is a variance regularization term, represented by $\log\Sigma_\theta$, which discourages collapse to near-zero variance, thereby preventing overconfident estimates.

The target $\mathbf{z}_{t-1}$ is obtained from the forward process. Given $\mathbf{z}_0$ and sampled noise $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$, we compute $\mathbf{z}_t$ and then derive the corresponding $\mathbf{z}_{t-1}$ using the reparameterization:

$$\mathbf{z}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}z_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{z}_t + \sqrt{\tilde{\beta}_t}\epsilon_{t-1}. \tag{9}$$

The overall training objective is:

$$\mathcal{L}_{total} = \mathbb{E}_{t,z_0,\epsilon}[\mathcal{L}_{var}(t)], \tag{10}$$

where timesteps are sampled uniformly, $t \sim \mathcal{U}(1, T)$. This variational loss drives the model to learn both accurate mean predictions which are essential for stable and high-fidelity generation.

## 2.5 Performance evaluation

### 2.5.1 Variational Autoencoder Evaluation

The performance of the VAE is critical for the stable and efficient training of the diffusion model. Since the VAE's decoder is used to reconstruct the final predicted GE profile $\hat{\mathbf{x}}$ from the denoised latent representation $\hat{\mathbf{z}}_0$, its reconstruction fidelity determines the practical upper bound for the overall predictive performance of our framework. Therefore, we evaluated the reconstruction quality of VAE using Pearson correlation coefficient (PCC) and $R^2$ score between the input and reconstructed GE vectors.

To ensure an accurate assessment of generalization, we employed a compound-based split guided by the SMILES string of drugs, thereby preventing any overlap of drugs across the three subsets. To achieve a robust and unbiased evaluation, we first extracted 20% of entire dataset for the fixed test set. Then, we performed five-fold cross-validation on the remaining 80% data pool. As a result, the test set remained constant for all five folds, while the 80% pool was iteratively divided into distinct training and validation sets for model training.

### 2.5.2 Diffusion Model Evaluation

The generation performance of diffusion model was evaluated using the root mean squared error (RMSE), PCC and $R^2$ score by comparing the predicted GE profiles with ground truth (GT), which means the perturbed GE profiles data from LINCS L1000. This predictive accuracy was benchmarked against the existing baseline models, PRnet [11] and PertDiT [21].

To strictly evaluate the model's generalization ability under unseen perturbation conditions, we adopted a compound-based data splitting strategy consistent with that described in Section 2.5.1. The dataset was divided into training, validation, and test sets at a 6:2:2 ratio, using the SMILES strings of drugs to ensure that no compound appeared in more than one subset. This evaluation protocol enables an objective assessment of the model's robustness and its capacity to predict transcriptomic responses for previously unseen drugs.

### 2.5.3 Gene–gene Expression Capturing Evaluation

An important aspect of transcriptomic modeling is the ability to reproduce the correlation structure among genes, as gene–gene interactions underlie regulatory networks and biological processes. To assess this capability, we compared the predicted correlation patterns from the proposed model and the baseline models against the GT, which is defined as the correlation structure observed in the perturbed GE profiles data from LINCS L1000.

For qualitative assessment, we first identified the set of top 30 most strongly correlated genes ranked by absolute correlation magnitude as determined by GT. We then visualized the pairwise

correlations of selected genes using heatmaps. Furthermore, we examined the overlap among the sets of these highly correlated genes, as identified by the proposed model, the baseline models, and the GT.

#### 2.5.4 Biological Relevance Evaluation

To further explore the practical utility of our model in drug discovery, we evaluated whether the generated GE profiles could be leveraged to improve the prediction of drug sensitivity. Specifically, we focused on the $IC_{50}$ as a representative pharmacological endpoint.

We obtained $IC_{50}$ values, which correspond to specific compound, cell line, dose, and time conditions from the GDSC2 [31] database by downloading the fitted dose–response descriptions. For conditions with multiple $IC_{50}$ entries corresponding to the same compound, cell line, dose, and time, duplicates were removed. We retained only those $IC_{50}$ measurements corresponding to cell lines with available basal GE profiles in the LINCS L1000 dataset, resulting in a total of 3,457 $IC_{50}$ values. For each of these conditions, perturbed GE profiles were generated using either baseline models or the proposed model.

To predict $IC_{50}$ values, we constructed a downstream regression model, utilizing a MLP architecture. The input to this model consisted of a concatenated vector of the generated perturbed GE and the corresponding compound embeddings extracted by Molformer [29]. The perturbed GE profiles were normalized using a standard scaler fitted exclusively on the training data, which was then consistently applied to the validation and test sets to prevent data leakage. The $IC_{50}$ prediction dataset was split randomly into training, validation, and test sets in an 8:1:1 ratio. The regression model was trained by minimizing the mean squared error (MSE) loss function. This setup allowed us to quantify the increase in predictive performance achieved when using GE profiles generated by our model compared to the baseline models.

# 3. Results

## 3.1 Model performance and generalization ability

We first assessed the performance of the VAE, which serves as the foundation for constructing a biologically meaningful latent space. The VAE demonstrated highly stable reconstruction of perturbed GE profiles, achieving a PCC of $0.956 \pm 0.000$ and an $R^2$ score of $0.914 \pm 0.001$ between the input and reconstructed expression vectors. These results indicate that the VAE effectively captures the intrinsic structure of transcriptomic data. Consequently, the learned latent representation provides a robust basis for subsequent diffusion training, minimizing information loss and ensuring the reliability of generated GE from the diffusion model.

Using the latent representations learned by the VAE, we next evaluated the performance of the diffusion model in reconstructing perturbed GE profiles under unseen perturbation conditions. As shown in Table 1, the diffusion model achieved the lowest reconstruction error and the highest correlation with the GT profiles compared to the other methods. Our model recorded an RMSE of 1.340, a PCC of 0.832, and an $R^2$ score of 0.669. These values were slightly higher than those obtained by PertDiT, which had an RMSE of 1.370, a PCC of 0.830, and an $R^2$ score of 0.665. Compared with PRnet, which showed an RMSE of 1.726, a PCC of

0.751, and an $R^2$ score of 0.437, our model exhibited a marked improvement in both prediction accuracy and GE pattern consistency.

Because the model takes molecular feature vectors derived from SMILES representations as input, it can generate GE profiles for previously unseen compounds based on their chemical structures. This capability enables the model to infer plausible transcriptional responses without requiring prior exposure to a given compound during training.

These results collectively demonstrate that the proposed diffusion model can robustly reconstruct GE patterns and generalize beyond the compounds observed during training.

**Table 1.** GE reconstruction performance of baseline models and proposed model.

| Model | RMSE | PCC | $R^2$ score |
|---|---|---|---|
| **PRnet** [11] | 1.726 | 0.751 | 0.437 |
| **PertDiT** [21] | 1.370 | 0.830 | 0.665 |
| **Proposed model** | **1.340** | **0.832** | **0.669** |

## 3.2 Gene–Gene Correlation Capturing

The preservation of the underlying gene–gene correlation structure is critical, as these interactions are fundamental to biological processes and regulatory networks. The proposed diffusion-based model demonstrated a substantial advantage over the baseline models in faithfully reproducing this structure across both global and local metrics.

The qualitative assessment of predicted correlation patterns revealed superior fidelity of the proposed model. We visualized pairwise correlations among the top 30 genes ranked by absolute magnitude in the GT. These heatmaps illustrated that the proposed model reproduced the overall correlation structures more faithfully than baseline models (Fig. 3).

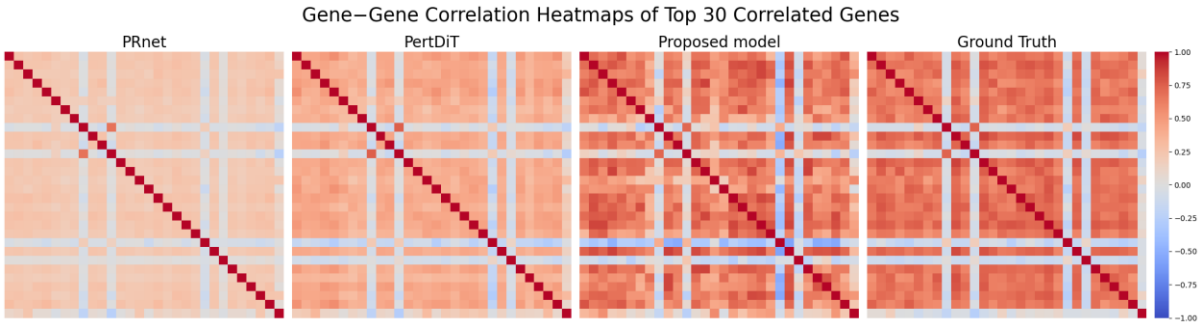


**Fig. 3.** Gene–gene correlation heatmaps of top 30 correlated genes. Heatmaps are shown for PRnet, PertDiT, the proposed model, and the GT. The top 30 genes were selected separately in GT, and pairwise correlation patterns among these genes are compared.

Crucially, the sets of highly correlated genes identified by the proposed model showed significantly stronger structural agreement with the GT. Specifically, 12 of the top 30 genes identified by the proposed model overlapped with those found in the GT, whereas PRnet only achieved an overlap of 3 genes with the GT and 2 genes for PertDiT. These results highlight the proposed model's superior ability to preserve gene–gene correlations.

These results demonstrate that the proposed diffusion-based model not only achieves superior reconstruction accuracy but also robustly preserves the underlying gene–gene correlation structure. This robust preservation is crucial for generating biologically faithful transcriptomic profiles.

### 3.3 Potential applications in drug discovery

We evaluated whether the GE profiles generated by the proposed model could improve downstream pharmacological predictions, specifically $IC_{50}$ estimation. As shown in Table 2, the $IC_{50}$ regression model trained on GE profiles generated by the proposed model achieved the best predictive performance, with an RMSE of 1.335, and an $R^2$ score of 0.819. In comparison, the models using profiles from PertDiT and PRnet obtained RMSE values of 1.405 and 1.591, and $R^2$ score of 0.800 and 0.743, respectively. This improvement indicates that our model preserves biologically relevant gene–gene correlations that influence drug sensitivity [32], thereby enhancing the predictive power of downstream pharmacological tasks.

Beyond these quantitative improvements, the ability to predict $IC_{50}$ directly from computationally generated transcriptional responses highlights the model's potential for virtual screening and drug repurposing. Unlike experimental perturbation assays, which are time-consuming and costly, our generative approach provides a scalable framework for estimating cellular responses across untested compound–cell line combinations. The generated GE profiles can serve as reliable substitutes for experimentally measured perturbation data, significantly reducing the time and cost required for large-scale drug evaluation. These findings demonstrate the potential of the proposed model as a practical tool for accelerating drug discovery and precision medicine.

**Table 2.** Comparison of $IC_{50}$ prediction performance using perturbed GE profiles generated by baseline models and the proposed model.

| Model | RMSE | $R^2$ score |
|---|---|---|
| **PRnet** [11] | 1.591 | 0.743 |
| **PertDiT** [21] | 1.405 | 0.800 |
| **Proposed model** | **1.335** | **0.819** |

## 4. Discussion

In this study, we developed an LDM framework for predicting drug-perturbed GE profiles.

The proposed model demonstrated strong generalization ability under unseen perturbation conditions, robustly captured gene–gene correlation structures and improved the prediction of pharmacological responses such as $IC_{50}$ when compared with existing approaches such as PRnet and PertDiT. These findings underscore the potential of diffusion-based generative modeling as a effective tool for advancing transcriptomic prediction and drug discovery.

Our results contribute to a growing body of research applying generative deep learning models to systems biology. Previous encoder–decoder or GAN-based approaches have shown promise in modeling drug perturbations but often failed to fully capture gene–gene dependencies, limiting their biological interpretability. By contrast, diffusion models explicitly learn the joint distribution of GE vectors through a probabilistic denoising process, which enables the recovery of gene–gene interactions. This highlights the advantage of the diffusion paradigm in faithfully preserving transcriptomic structure.

Importantly, the ability of the proposed model to enhance $IC_{50}$ prediction demonstrates its utility for downstream pharmacological applications. While profiles sampled by baseline models provided some predictive value, our diffusion-based model yielded more accurate and stable $IC_{50}$ estimates, suggesting that biologically consistent gene–gene correlations translate into improved performance in real-world drug sensitivity tasks. This finding aligns with recent studies emphasizing the importance of accurate representation of cellular states for pharmacogenomic prediction.

Despite these strengths, several limitations should be acknowledged. First, the study relied on LINCS L1000 landmark genes, which, although widely used, capture only a subset of the transcriptome; extending this framework to whole-transcriptome data would broaden its applicability. Second, while our model incorporated basal GE, compound features, dose, and time, other cellular context factors such as epigenetic modifications, pathway activity, or microenvironmental cues were not considered. Incorporating such multimodal information may further enhance predictive performance. Third, although our evaluation demonstrated generalization to unseen compounds, the scope of external validation remains limited to GDSC2 $IC_{50}$ data; broader pharmacogenomic benchmarks and prospective experimental validation are needed to confirm translational utility.

In summary, our work demonstrates that diffusion models can overcome key limitations of previous approaches by providing biologically faithful and generalizable predictions of perturbed GE. At the same time, it opens new directions for integrating generative models into pharmacogenomic pipelines, while highlighting areas where further methodological and experimental refinement are required.

# 5. Conclusions

In this work, we introduced a latent diffusion model framework for predicting drug-perturbed GE. By combining a variational autoencoder for latent space construction with a diffusion process that jointly models mean and variance, the proposed approach achieved stable learning and robust generalization. Our results demonstrate that the model not only reproduces the gene–gene correlation structure of real data more faithfully than prior methods such as PRnet

and PertDiT, but also provides biologically consistent representations that enhance downstream applications such as $IC_{50}$ prediction.

These findings highlight the potential of diffusion-based generative models as versatile tools for drug discovery and precision medicine. By enabling more accurate modeling of cellular responses to chemical perturbations, the proposed framework may facilitate improved drug sensitivity prediction, inform mechanism-of-action studies, and support the prioritization of candidate compounds. Future extensions incorporating multimodal cellular features and broader experimental validation could further strengthen the translational impact of this approach.

## Competing interests

No competing interest is declared.

## Data availability statement

The L1000 was downloaded from the Gene Expression Omnibus with the accession number (GSE92742). The $IC_{50}$ data was downloaded from the GDSC2 website (https://www.cancerrxgene.org/). All codes for the manuscript are available at GitHub website (https://github.com/bmil-jnu/Perturbed-GE-LDM).

## Author contributions statement

C.K. and S.Y. designed research; C.K. conducted research; C.K. developed, trained and evaluated the models; C.K. conducted the experiments; C.K. and S.Y. wrote and reviewed the manuscript.

## Funding

This research was supported by a grant from Ministry of Food and Drug Safety (RS-2024-00332003, RS-2025-02215961) and a grant from the National Research Foundation of Korea funded by the Ministry of Science and ICT (RS-2025-16063391).

## References

1. Herwig R. Computational modeling of drug response with applications to neuroscience. *Dialogues Clin Neurosci* 2014;**16**:465–77. https://doi.org/10.31887/DCNS.2014.16.4/rherwig
2. Ma Y, Ding Z, Qian Y *et al.* Predicting cancer drug response by proteomic profiling. *Clin Cancer Res* 2006;**12**:4583–9. https://doi.org/10.1158/1078-0432.CCR-06-0290

3. Zhao J, Zhang XS, Zhang S. Predicting cooperative drug effects through the quantitative cellular profiling of response to individual drugs. *CPT Pharmacometrics Syst Pharmacol* 2014;**3**:e102. https://doi.org/10.1038/psp.2013.79
4. Szalai B, Veres DV. Application of perturbation gene expression profiles in drug discovery-from mechanism of action to quantitative modelling. *Front Syst Biol* 2023;**3** https://doi.org/ARTN 1126044

10.3389/fsysb.2023.1126044
5. Pham TH, Qiu Y, Zeng JC *et al.* A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to covid-19 drug repurposing. *Nat Mach Intell* 2021;**3** https://doi.org/10.1038/s42256-020-00285-9
6. Pak M, Lee SS, Sung I *et al.* Improved drug response prediction by drug target data integration via network-based profiling. *Briefings in Bioinformatics* 2023;**24** https://doi.org/10.1093/bib/bbad034
7. Bang D, Koo B, Kim S. Transfer learning of condition-specific perturbation in gene interactions improves drug response prediction. *Bioinformatics* 2024;**40**:i130–i39. https://doi.org/10.1093/bioinformatics/btae249
8. Pereira DA, Williams JA. Origin and evolution of high throughput screening. *Brit J Pharmacol* 2007;**152**:53–61. https://doi.org/10.1038/sj.bjp.0707373
9. Subramanian A, Narayan R, Corsello SM *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–52.e17. https://doi.org/10.1016/j.cell.2017.10.049
10. Hetzel L, Böhm S, Kilbertus N *et al.* Predicting cellular responses to novel drug perturbations at a single-cell resolution. *Adv Neur In* 2022
11. Qi X, Zhao L, Tian C *et al.* Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nature Communications* 2024;**15**:9256. https://doi.org/10.1038/s41467-024-53457-1
12. Kim S, Bae S, Piao Y *et al.* Graph convolutional network for drug response prediction using gene expression data. *Mathematics-Basel* 2021;**9** https://doi.org/ARTN 772

10.3390/math9070772
13. Shu HT, Zhou JT, Lian QY *et al.* Modeling gene regulatory networks using neural network architectures. *Nat Comput Sci* 2021;**1**:491–501. https://doi.org/10.1038/s43588-021-00099-8
14. Targonski C, Bender MR, Shealy BT *et al.* Cellular state transformations using deep learning for precision medicine applications. *Patterns* 2020;**1** https://doi.org/ARTN 100087

10.1016/j.patter.2020.100087
15. Russkikh N, Antonets D, Shtokalo D *et al.* Style transfer with variational autoencoders is a promising approach to rna-seq data harmonization and analysis. *Bioinformatics* 2020;**36**:5076–85. https://doi.org/10.1093/bioinformatics/btaa624
16. Wei XJ, Dong JY, Wang F. Scpregan, a deep generative model for predicting the response of single-cell expression to perturbation. *Bioinformatics* 2022;**38**:3377–84. https://doi.org/10.1093/bioinformatics/btac357
17. He S, Zhu Y, Tavakol DN *et al.* Squidiff: Predicting cellular development and responses to perturbations using a diffusion model. 2024 https://doi.org/10.1101/2024.11.16.623974
18. Luo E, Hao M, Wei L *et al.* Scdiffusion: Conditional generation of high-quality single-cell data using diffusion model. *Bioinformatics* 2024;**40**:btae518. https://doi.org/10.1093/bioinformatics/btae518

19. Zhang J, Liu Z, Wang Y *et al.* Subgdiff: A subgraph diffusion model to improve molecular representation learning. *Advances in Neural Information Processing Systems* 2024;**37**:29620–56.
20. Ho J, Jain A, Abbeel P Denoising diffusion probabilistic models. Curran Associates, Inc. 6840–51.
21. Hu Q, Chen Z, Gu J. Predicting drug-perturbed transcriptional responses using multi-conditional diffusion transformer. *Quantitative Biology* 2026;**14**:e70016.
22. Nichol A, Dhariwal P. Improved denoising diffusion probabilistic models. 2021 https://doi.org/10.48550/arXiv.2102.09672
23. Razavi A, van den Oord A, Vinyals O Generating diverse high-fidelity images with vq-vae-2. Curran Associates, Inc.
24. Rombach R, Blattmann A, Lorenz D *et al.* High-resolution image synthesis with latent diffusion models. *Proc Cvpr Ieee* 2022:10674–85. https://doi.org/10.1109/Cvpr52688.2022.01042
25. Rdkit: Open-source cheminformatics software. https://www.rdkit.org (02 Oct, date last accessed).
26. Kingma DP, Welling M. Auto-encoding variational bayes. *arXiv preprint arXiv:13126114* 2013
27. Kullback S, Leibler RA. On information and sufficiency. *The annals of mathematical statistics* 1951;**22**:79–86.
28. Sadria M, Layton A. Scvaeder: Integrating deep diffusion models and variational autoencoders for single-cell transcriptomics analysis. *Genome Biol* 2025;**26**:64. https://doi.org/10.1186/s13059-025-03519-4
29. Ross J, Belgodere B, Chenthamarakshan V *et al.* Large-scale chemical language representations capture molecular structure and properties. 2022 https://doi.org/10.48550/arXiv.2106.09553
30. Canzler S, Schor J, Busch W *et al.* Prospects and challenges of multi-omics data integration in toxicology. *Archives of Toxicology* 2020;**94**:371–88. https://doi.org/10.1007/s00204-020-02656-y
31. Yang W, Soares J, Greninger P *et al.* Genomics of drug sensitivity in cancer (gdsc): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;**41**:D955–61. https://doi.org/10.1093/nar/gks1111
32. Ahmed KT, Park S, Jiang Q *et al.* Network-based drug sensitivity prediction. *BMC Med Genomics* 2020;**13**:193. https://doi.org/10.1186/s12920-020-00829-3