LUMIC: LATENT DIFFUSION FOR MULTIPLEXED IMAGES OF CELLS

Anonymous authors
Paper under double-blind review

ABSTRACT

The rapid advancement of high-content, single-cell technologies like robotic confocal microscopy with multiplexed dyes (morphological profiling) can be leveraged to reveal fundamental biology, ranging from microbial and abiotic stress to organ development. Specifically, heterogeneous cell systems can be perturbed genetically or with chemical treatments to allow for inference of causal mechanisms. An exciting strategy to navigate the high-dimensional space of possible perturbation and cell type combinations is to use generative models as priors to anticipate high-content outcomes in order to design informative experiments. Towards this goal, we present the Latent diffUsion for Multiplexed Images of Cells (LUMIC) framework that can generate high quality and high fidelity images of cells. LUMIC combines diffusion models with DINO (self-Distillation with NO labels), a vision-transformer based, self-supervised method that can be trained on images to learn feature embeddings, and HGraph, a hierarchical variational graph encoder-decoder to represent chemicals. To demonstrate the ability of LUMIC to generalize across cell lines and treatments, we apply it to two cell lines treated with chemicals and stained with three dyes from the JUMP Pilot dataset and a newly-generated in-house dataset of five cell lines treated with chemicals and stained with three dyes. To quantify prediction quality, we evaluate the DINO embeddings, Kernel Inception Distance (KID) score, and recovery of morphological feature distributions. LUMIC significantly outperforms previous methods and generates realistic out-of-sample images of cells across unseen compounds and cell types.

1 INTRODUCTION

High-content imaging assays have revolutionized the ability to observe and analyze the morphological impact of a wide variety of drugs on different cell types. For example, the Cell Painting assay uses six fluorescent dyes imaged across five different channels to capture cell phenotypes (Bray et al., 2016), and has already facilitated the identification of drug targets and mechanisms of action (Chandrasekaran et al., 2023). Although morphological profiling assays are cost-effective and straightforward methods that only require commonly available laboratory equipment, the number of possible cell type and chemical compound combinations is infeasible to explore experimentally.

At the same time, substantial progress has been made in generative machine learning, enabling conditional image, video, and text generation (Saharia et al. (2022); Ho et al. (2022); Achiam et al. (2023)). Deep generative models, such as Generative Adversarial Networks (GANs), normalizing flows, and denoising diffusion probabilistic models (DDPMs), have recently garnered attention for their ability to generate high-quality and fidelity samples (Goodfellow et al. (2020); Rezende & Mohamed (2015); (Ho et al., 2020)). Specifically, diffusion models have become popular because of their training stability, ease of guidance, and state-of-the-art performance on image generation tasks. Diffusion models outperform GANs and flow models without requiring specific architectural and optimization choices to prevent mode collapse and stable training (Ho et al., 2020).

With the rise of generative models and the curation of large biological datasets, it is now feasible and promising to train generative models on perturbation data for morphological assays. For example, the JUMP dataset contains 116k chemical perturbations (Chandrasekaran et al., 2023)–analogous to the massive datasets used to train generative models in the natural language and vision space. Specifically, the use of generative models can be used to simulate the interactions of various compound and cell line interactions to identify target morphologies and limit the large search space for experimentation.

While generative modeling has already been attempted on biological data for perturbation prediction, LUMIC is the first method, to our knowledge, to model chemical perturbation effects across multiple cell lines, filling a key gap in existing literature. The ability to perform cross-cell-line image generation allows modeling heterogeneous cellular responses of different cancer/organ strains/phenotypes to perturbations. This increased modeling capability is an important tool for optimizing early drug discovery processes (Caie et al., 2010) and better characterizing biological signaling pathways (Heinrich et al., 2023). Several previous studies have investigated the problem of generating cellular images. Mol2Image presents a flow-based model to generate only U2OS cells conditional on a graph neural network chemical embedding (Yang et al., 2021). CP2Image trains a GAN conditional on CellProfiler representations to generate images; however, using CellProfiler features lacks the flexibility of other methods to directly control the chemical compound used to treat a cell. The authors of Hussain et al. (2020) use DCGAN, a GAN in which the pooling layers are replaced by strided convolutions, to generate high-content microscopy images, but the scope of the study was limited, as it was trained and evaluated only on ten compounds. IMPA, one of the most similar methods to ours, adopts a conditional GAN to style transfer perturbations onto U2OS cells, treating cells as content and compounds as the style. PhenDiff adopts an image-to-image diffusion model to ”translate” from control images to perturbations by conditioning on a class embedding (Bourou et al., 2023); however, the use of chemical class labels prevents this method from being able to generate images of chemicals not seen during training. MophoDiff utilizes a StableDiffusion backbone to generate images of cells impacted by both chemical and genetic perturbations but was only evaluated on single cell line datasets (Navidi et al., 2024). We summarize the existing methodology and the gap that LUMIC fills in **Table 1**.

Model	**Architecture**	**Unseen Compound Generation**	**Multi-Cell Line Generation**
Mol2Image	Flow Model	✓	✗
CP2Image	GAN	✗	✗
DCGAN	GAN	✗	✗
IMPA	GAN	✓	✗
PhenDiff	Diffusion	✗	✗
MorphoDiff	Diffusion	✓	✗
LUMIC	Diffusion	✓	✓

Table 1: Overview of related work: LUMIC leverages a diffusion pipeline to generate combinations of unseen compound and cell line, a task that is unachievable with existing methods

We compare our method only against IMPA and PhenDiff, as these are the most relevant and recent papers that have publically available code and predict the morphological responses to perturbations.

LUMIC adopts a standard DDPM pipeline and is not only able to beat existing methods on single cell line generation tasks but also removes the limitation of single cell line generation, allowing for the controllable prediction of multiple cell line and chemical compound interactions. Specifically, our key innovation is to model how perturbing a specific type of cell (specified as an image of the control cell well) changes its morphology. Analogous to text-to-image approaches such as DALL-E 2 and Stable Diffusion, we use diffusion in the image embedding space (latent diffusion) to learn the context-specific effect of a compound (Ramesh et al. (2022); Rombach et al. (2022)). This allows us to predict how either seen or unseen compounds will affect either seen or unseen cell lines. To evaluate our approach of “transferring” perturbations onto a new cell line, we performed laboratory experiments to generate a new Cell Painting dataset featuring the same treatments across multiple cell lines. Biological and computational evaluations demonstrate that LUMIC can meaningfully predict the effects of chemical treatment for both unseen compounds and unseen cell types. In summary, LUMIC is able to: 1) Beat existing methods on single cell line perturbation generation tasks; and 2) Generate images of multiple cell lines after chemical treatment, including unseen cell lines and unseen chemical.

2 Methods

2.1 Datasets

The JUMP Pilot Target 1 dataset is a subset of the JUMP Cell Painting dataset and consists of 306 different chemical perturbations on two different cell lines (U2OS cells and A549 cells) at 2 different time points (24 and 48 hours).

We used only the 24 hour timepoint, resulting in 4 plates per cell line and 27,702 images total. Additionally, pairs of compounds within this dataset are known to target the same protein encoded by a given gene (Chandrasekaran et al., 2024). We split the data based on the gene that they target: a total of 30 compounds are held out, 16 compounds target genes not seen during training (8 genes), which we refer to as unseen genes, and 14 compounds target genes seen during training, which we refer to as seen genes. Three of the five fluorescent channels–nucleus (Hoechst; DNA), actin cytoskeleton/golgi/plasma membrane (phalloidin; AGP), and mitochondria (MitoTracker; Mito)–were stacked to form an RGB image for compatibility with the standard DINO architecture. We also generated a new experimental dataset, which we refer to as the "style transfer dataset". We plated 3T3 (Fibroblasts), A549, HEK293T, HeLa, and RPTE (Kidney) cells and stained them using the same stains as the JUMP Pilot dataset (DNA, AGP, Mito) in a CellPainting style assay for a total of 3,168 images (Bray et al., 2016). We randomly select 10 compounds to hold out across all 5 cell lines and hold out HeLa completely (except control images) during training to make up the test set.

Each cell line consists of a singular cell type, leading to a combined total of 6 unique cell types and approximately 360 unique compounds. More details on dataset implementations are in **Appendix A.3**

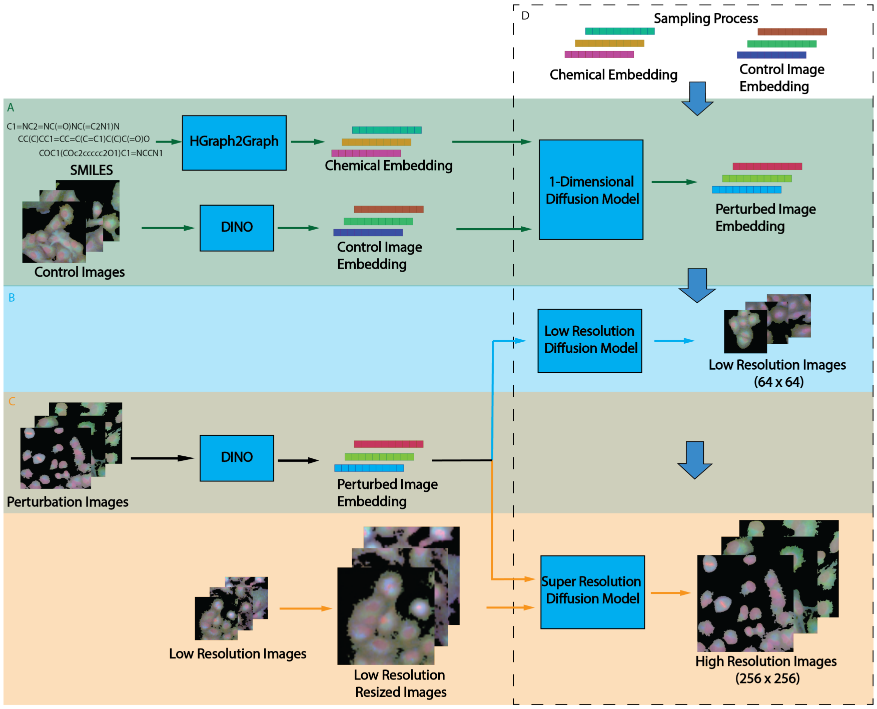
D
Sampling Process
Chemical Embedding
Control Image Embedding
A
C1=NC2=NC(=O)NC(=C2N1)N
CC(C)CC1=CC=C(C=C1)C(C)C(=O)O
COC1(COc2ccccc2O1)C1=NCCN1
SMILES
HGraph2Graph
Chemical Embedding
1-Dimensional Diffusion Model
Perturbed Image Embedding
Control Images
DINO
Control Image Embedding
B
Low Resolution Diffusion Model
Low Resolution Images (64 x 64)
C
Perturbation Images
DINO
Perturbed Image Embedding
Low Resolution Images
Low Resolution Resized Images
Super Resolution Diffusion Model
High Resolution Images (256 x 256)

Figure 1: Architecture of LUMIC in which the image generation of different cell lines and chemical compounds is broken up into 3 separate models: A) Perturbed image embedding generation using a 1-dimensional diffusion model with chemical conditioning information and control image embeddings as inputs (green). B) Low resolution image generation using diffusion model with perturbed image embeddings as inputs (blue). C) High resolution image generation using diffusion model with perturbed image embeddings and resized low resolution images as inputs (orange). D) Sampling process in which chemical information and control image embeddings are passed into the 1d-diffusion model, low resolution diffusion model, and high resolution diffusion model with all intermediate outputs passed into the following model in order to generate the interaction of a desired cell line and compound from scratch.

2.2 LUMIC: Latent Diffusion for Multiplexed Images of Cells

LUMIC uses conditional diffusion in the latent image space to predict the embedding of the perturbed cell from the embeddings of the unperturbed cell and the small molecule. To do this, we need a way of embedding both control and perturbed images; a way to embed small molecules; a network to predict perturbed image embedding; and a way to predict a high-resolution perturbed image from its embedding in latent space. LUMIC achieves this by leveraging pretrained chemical and visual encoders, an image embedding diffusion model, a low-resolution image diffusion model, and a high-resolution image diffusion model. Background details on DDPM are explained in **Appendix A.1**.

2.2.1 Embedding Control and Perturbed Images with DINO

To learn image feature embeddings that capture cell line information, we trained a DINO (self-Distillation with NO labels) model, which is a vision-transformer that uses a self-supervised loss (Caron et al., 2021). DINO effectively learns representations of cellular morphology directly from CellPainting images (Doron et al., 2023) and outperforms other self-supervised learning methods, including SymCLR and MAE, as well as computer vision based feature embeddings from CellProfiler in downstream biological tasks (Kim et al., 2023). We trained our DINO model on $\sim$27,000 images from the JUMP Pilot subset and $\sim$3,000 images for our style transfer dataset. More details on DINO training is included in **Appendix A.4**.

2.2.2 Embedding Small Molecules with HGraph

To encode chemical information, we trained an HGraph model, which is a hierarchical graph encoder-decoder that utilizes structural motifs as building blocks to encode SMILES (a way to represent compounds using ASCII strings) (Jin et al., 2020), on all 306 chemical compounds from the JUMP Pilot dataset, 61 compounds from our style transfer dataset, and 250k compounds randomly sampled from the ZINC dataset by Kusner et al. (2017). During training of LUMIC, the SMILES are passed into HGraph, and a 128-dimensional vector is sampled from the latent space.

2.2.3 Image Embedding Diffusion Model

To generate image representations (DINO) with diffusion, we modified the standard U-Net architecture introduced in Ronneberger et al. (2015) to use 1-dimensional convolutions over the image embedding (a vector) instead of 2-dimensional convolutions over an image (a matrix). Given a control image of a cell line, we first take a 256 x 256 random crop of the image and encode it into a 384-dimensional feature embedding using DINO. Then a compound (SMILES) is encoded using HGraph and a chemical latent is sampled. The model then outputs the image embedding of the targeted interaction (that cell type treated with that compound) as seen in **Fig. 1 A**. This diffusion model employs linear attention to learn conditional information. We trained this model using the Adam optimizer with a learning rate of 5e-4 and a batch size of 64 for 24 hours, totaling 75,000 steps (Kingma, 2014).

2.2.4 Low-Resolution Image Diffusion Model

The low resolution diffusion model takes the visual embeddings (from DINO) and decodes them into their respective image. The model takes random (256 x 256) crops of images, encodes them, and learns to generate the (64 x 64) low resolution image based on the embedding (**Fig.1 B**). This model follows the efficient U-Net architecture proposed in (Saharia et al., 2022) and uses cross-attention at each layer to capture the conditioning information. We trained this model using the Adam optimizer with a learning rate of 5e-5 and a batch size of 64 for 24 hours totaling 150,000 steps.

2.2.5 Super-Resolution Image Diffusion Model

The super resolution diffusion model takes the visual embeddings and the low resolution (64 x 64) image and generates the corresponding high resolution (256 x 256) image. The model is trained by taking 256 x 256 random crops and inputting the low resolution (resizing the 256 x 256 crop to 64 x 64 and then resizing it back to 256 x 256) version as well as the encoded version (passing the 256 x 256 crop into DINO) as shown in **Fig. 1 C**. The model architecture is consistent with the efficient U-Net architecture used in (Saharia et al., 2022) with linear attention. We trained this model using the Adam optimizer with a learning rate of 5e-5 and a batch size of 33 for ten days, totaling 1,000,000 steps across 3 A40 GPUs.

2.3 Evaluation

We use the Kernel Inception Distance (KID), which is the squared maximum mean discrepancy (MMD) between the distributions of image representations from the Inception V3 network, to determine the quality of the generated images (Bińkowski et al., 2018). KID is a commonly used metric when there are low samples to evaluate the quality of the real and generated images. The MMD is a statistical measure used to quantify the distance between two distributions (Gretton et al., 2012). Smaller KIDs (closer to 0) are better, indicating smaller distance between true and generated image distributions. All classifier model training was done using scikit-learn and a validation set of 20% (of actual data) was held out during training, and we refer to the validation set as ”Real” in all tables (Pedregosa et al., 2011). We report the balanced accuracy score for all classification results.

3 Results

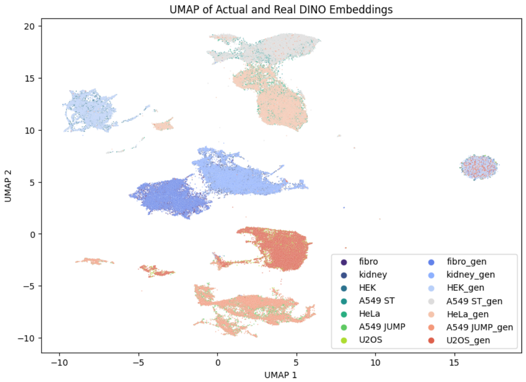

Figure 2: UMAP of real and generated DINO embeddings overlaid and grouped by cell type

	KID
LUMIC	**0.015**
PhenDiff	0.331
IMPA	0.173

Table 2: Comparison of LUMIC Against Other Methods by Calculating the KID of Unconditional U2OS generation using Random Crops during Training

	Control Images		Unseen Compounds	
	Real	Generated	Real	Generated
MLP	0.925	0.916	0.929	0.918
KNN	0.873	0.877	0.912	0.915

Table 3: Accuracy of Different Machine Learning Classifiers on Cell Type Prediction using DINO Embeddings

	Real DINO Embedding	Generated DINO Embedding
A549 Seen	0.274	0.243
A549 Unseen	0.331	0.305
U2OS Seen	0.214	0.202
U2OS Unseen	0.275	0.269

Table 4: KNN Accuracy of Gene Classification for the JUMP Dataset Averaged over the Set of Test Compounds

3.1 LUMIC Generates Realistic Image Embeddings that Preserve Cell Type and Treatment Semantics

To qualitatively evaluate the DINO image embeddings, we used UMAP to visualize the real and generated DINO image embeddings colored by cell type and compounds, where the real images are random 256x256 crops (as following the training procedure) and the generated images are generated 256x256 crops using the entire sampling pipeline (**Fig. 2**). Both the real and generated DINO embeddings cluster well by cell type, and the generated embeddings align with their respective ground truth clusters, suggesting that the real and generated embeddings contain corresponding cell type information. The isolated cluster (right most) contains a mixture of cell types and consist largely of cell-free black backgrounds present because of how we randomly cropped images.

The real and generated image embeddings also reflect differences among chemical treatments **Fig. 3**). The fibroblast cells (**Fig. 3A** and **3B**) separate less clearly than the HeLa cells (**Fig. 3C** and **3D**), possibly reflecting cell type differences in the magnitude of morphology change induced by chemical treatment. Regardless, treatment differences among cells are apparent in both cases. In particular, images of HeLa cells treated with a given compound tend to cluster with each other and apart from images of cells treated with a different compound. This phenomenon is also seen in the JUMP datasets. A549 cells (**Appendix Fig. A1**) show relatively less separation among treatments compared with U2OS cells (**Appendix Fig. A2**), which have almost layer-like subclusters. Nevertheless, the UMAPs of the generated embedding reflect the shape and distribution of their intended cell type and compound. This indicates that both the real and generated DINO image embeddings reflect differences in cell type and chemical treatment for both seen and unseen cell types and chemical compounds.

To quantitatively evaluate whether the generated embeddings reflect cell type semantics, we trained MultiLayer Perceptron (MLP) and k-Nearest-Neighbor (KNN) models to predict cell type from DINO embedding. We reasoned that if a classifier could accurately identify the cell type of a generated embedding, then LUMIC is correctly obeying the cell type conditioning. Reassuringly, the performance of the MLP and the KNN models is very similar on the real validation image embeddings and the set of generated embeddings of each type (**Table 3**).

We next used a similar strategy to evaluate whether embeddings generated by LUMIC preserve treatment semantics. The MLP and KNN classifiers trained to identify the held out treatments from embeddings of real cells were still able to identify the held out treatment of generated image embeddings (**Table 5**). The small drop in classification performance between real and generated embeddings indicates that the embeddings generated by LUMIC do indeed reflect the differences among images from different treatments, even for treatments not seen during LUMIC training. This is a much more difficult task as different compounds may cause little or similar changes in morphology, resulting in slightly lower accuracies for both the real and generated embeddings compared to the cell line classification task.

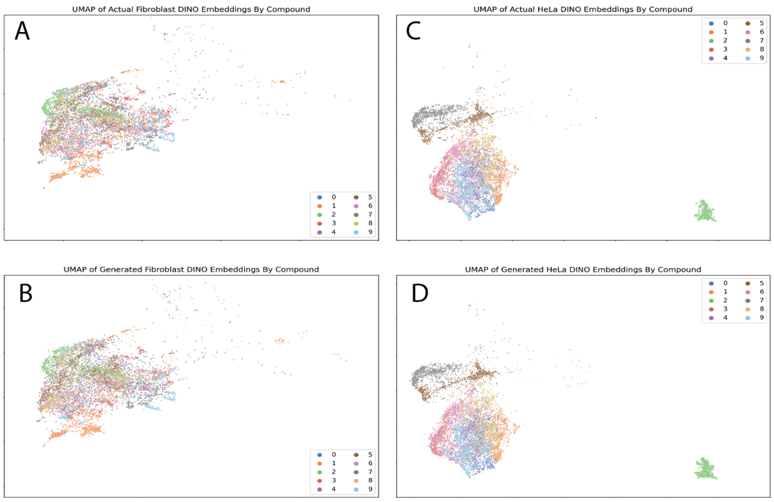

Figure 3: UMAPs of real and generated embeddings grouped by compound (same UMAP as **Fig. 2** but subset by cell type and recolored by compounds). (A) UMAP of the DINO embeddings of the real Fibroblast (seen cell line) images in the test set (unseen compounds) grouped by compound. (B) UMAP of the generated DINO embeddings for Fibroblasts (seen cell line) and test set compounds (unseen compounds) grouped by compound. (C) UMAP of the DINO embeddings of the real HeLa images (unseen cell line) in the test set (unseen compound) grouped by compound. (D) UMAP of the generated DINO embeddings for HeLa (unseen cell line) and test set compounds (unseen compounds) grouped by compound.

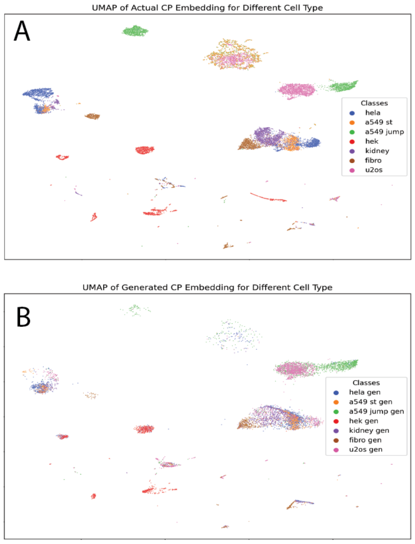

Figure 4: (A) UMAP of the extracted CellProfiler features from real images grouped by cell types. (B) UMAP of the extracted CellProfiler features from generated images grouped by cell type.

To further evaluate the quality of the generated embeddings, we trained a KNN classifier to identify the target gene of the compound used in each treatment (recall that the JUMP Pilot dataset profiled a set of compounds with a total of 146 known target genes). We trained the classifier on the DINO embeddings of real images, and then ran this classifier on image embeddings generated by LUMIC to quantify how well the embeddings retained signatures of the gene targeted by the compound. A high classifier score would indicate that the embeddings indeed do preserve the differences in the genes targeted caused by different compounds; the classifier performance on real embeddings represents an upper bound on generation performance. As seen in **Table 4**, the classifier performance on real and generated embeddings is similar, indicating that the model does indeed generate embeddings that distinguish among drug treatments. Of note, the accuracy of the model on actual images, while seemingly low, is similar to the low mean average precision values in a comparable evaluation for the same dataset, further exemplifying the difficulty of this task (Chandrasekaran et al., 2024). Nevertheless, these results indicate that we are able to meaningfully encode the targeted gene in the generated embeddings, capturing the underlying biology of the interactions.

3.2 LUMIC Generates More Realistic Morphology Images than Previous Approaches

We compared LUMIC against previous approaches for morphological responses to chemical perturbations. Since previous approaches cannot generate images for unseen cell types and/or molecules, we chose to evaluate unconditional

	MLP		KNN	
	Real	Generated	Real	Generated
A549	0.859	0.721	0.741	0.729
Kidney	0.810	0.696	0.651	0.647
Fibroblast	0.724	0.609	0.571	0.574
HEK293T	0.554	0.470	0.406	0.409
HeLa	0.869	0.724	0.729	0.735

Table 5: Accuracy of different classifiers on test set compound prediction using real and generated DINO embeddings.

	Train	Validation	Generated
MLP	0.720	0.735	0.472
KNN	0.527	0.507	0.349

Table 6: Accuracy of different classifiers on predicting cell type using CellProfiler Features

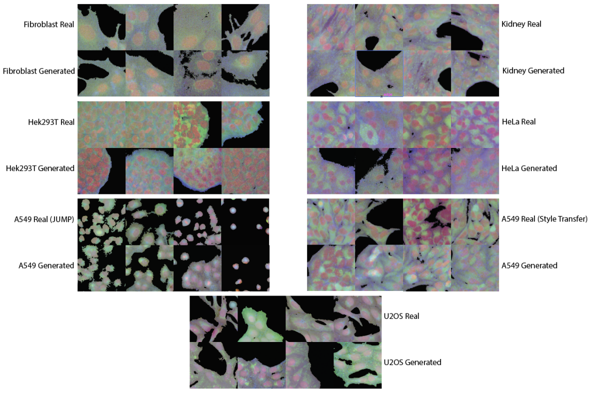

Figure 5: Real and generated samples of different cell types and chemical compounds from the test set. High-resolution images are included in the appendix as **Fig. A5**.

single cell line generation after training on random crops. We calculated the KID between the real and generated images, where the smaller the distance the more realistic the generated images (Bińkowski et al., 2018). LUMIC outperforms the other methods, achieving a significantly lower KID (**Table 2**). Thus, our model is able to capture accurate morphology and growth patterns with fidelity that surpasses the existing state of the art methods for single cell type generation.

3.3 Generated Images from Unseen Cell Types and Treatments Outperform Baseline Model

Through visual inspection of **Fig. 5 and Appendix Section A.6**, the generated images contain many morphological similarities with their intended cell type. For quantitatively evaluations, we assessed whether LUMIC can meaningfully predict images from unseen cell types and unseen treatments by comparing it with a baseline. We reasoned that, in order to be useful, the model predictions for a given cell type must be more similar to real images of the same cell type than real images of a different cell type. Similarly, for a cell type/treatment combination, the generated images must be more similar to the real cell type/treatment combination than to untreated control cell images of the same cell type. This baseline is not trivial, because it requires the generative model to preserve both the overall characteristics of the morphology and semantics of a particular cell line and/or treatment. Reassuringly, LUMIC beats this baseline for image generation conditional on cell type by a large margin with the corresponding class KID having a mean of 0.016 and a standard deviation of 0.010 and the different cell type comparisons achieving a mean KID of 0.101 and a standard deviation of 0.062 (full table in **Appendix Table 7**). This indicates that LUMIC both generates realistic images and obeys the semantics of cell type conditions. We next assessed LUMIC images generated from specified cell type/treatment combinations. We investigated three prediction tasks of increasing difficulty: (1) the cell type is in the training data but the treatments are held out (seen cell type, unseen treatment); (2) all treatments of a given cell type are held out, but the treatments are observed for other cell types (unseen cell type, seen treatment); and (3) a cell type/treatment combination is predicted when neither the cell type nor the treatment have been observed during training (unseen cell type, unseen treatment). For the JUMP dataset, we also investigated the difference between treatments with compounds targeting seen or unseen genes.

Remarkably, LUMIC outperforms the baseline across all three prediction tasks. Specifically, for the JUMP dataset, where compounds are held out that impact seen and unseen genes, LUMIC not only beats the baseline comparison against control images by an average of 0.017 and a standard deviation of 0.006 (full table in **Appendix Table 8**), but even performs better on compounds that impact unseen genes, suggesting the ability to remain accurate on out of distribution chemical perturbations. When looking at the seen cell lines in the style transfer dataset, LUMIC is able to predict effects of unseen compounds on multiple seen cell types (full table in **Appendix Table 9**), suggesting its applicability in downstream use cases to expand the scope of existing screens by increasing the number of compounds being explored. Finally, when we held out all HeLa cell treatments, LUMIC was able to predict both seen and unseen compounds on an unseen cell line better than baseline. The ability to accurately perform predictions on an unseen cell line provided seen compounds suggests that LUMIC could aid in understanding the different responses of drugs across distinct cell lines, mimicking the experimental setup in Heinrich et al. (2023) to identify the optimal cell line for screening without experimental limitations in order to best classify the MOA and bioactivity of a compound. Moreover, LUMIC shows promise in being able to create entirely new screens provided only control images, which would result in further understanding of cellular molecular pathways as well as effectively modeling the heterogeneity in cellular responses to chemical perturbations for early stage pharmaceutical discovery (Heinrich et al., 2023; Caie et al., 2010). Notably, the generation quality for predictions of the unseen cell type was worse, with the within class KID being much closer to baseline KID values than in the easier generation tasks with an average difference of 0.003 and a standard deviation of 0.002 (full table in **Appendix Table 10**).

3.4 Intensity Features Show Largest Difference Between Real and Generated Images

Cell morphology images from the Cell Painting dyes are often analyzed using hand-crafted features, such as those from the CellProfiler pipeline. These features capture various aspects of cell and nuclear shape and size, as well as dye intensity. To further investigate how well our generated images reflect biological properties of cells, we ran CellProfiler to segment cells and extract features. We plot a UMAP of the real and generated CellProfiler features and show that they do not match up as well as the real and generated DINO embeddings as seen in **Fig. 4A** and **4B**. Specifically, the bottom most green cluster contains A549 JUMP compounds that are present in the actual embeddings but not in the generated ones and the noticeable blue cluster of HeLa on the right side, which is again present in the actual embedding UMAP but not on the generated ones. This suggests that the biological features extracted from the generated images do not accurately match those extracted from the actual images. We then removed highly correlated features (> 0.9) and features with low variance (< 0.01). We trained MLP and KNN models to classify cell type from CellProfiler features of real cells, and evaluated the classifier on CellProfiler features of generated cell images (**Table 6**). The classifier was much less accurate at identifying the cell types of generated images from their CellProfiler features, indicating some sort of distribution shift.

To further evaluate which features differ the most between real and generated images, we trained a random forest (per cell type and per seen/unseen group when applicable) with 100 decision trees to classify real and generated images and observe high accuracy with a mean of 0.853 and a standard deviation of 0.075, suggesting that there are apparent differences between the two (full table in **Appendix Table 11**). We then used SHapley Additive exPlanations (SHAP) to identify the most important features per class (Lundberg, 2017). All of the most important features as well as the majority of remaining features after filtering are either "Intensity" or "AreaShape" metrics, with "Intensity" features reflecting the overall distribution of the intensities for the images and the "AreaShape" measurements being calculated after manual intensity based thresholding to identify the different cellular compartments. To identify the group of features that differ most between the real and generated images, we perform an ablation by removing the features with "Intensity" or "AreaShape". We then retrain the RF classifier to distinguish between real and fake images. The set of features that leads to a larger decrease in accuracy compared to the baseline features represent the more impactful group of features in distinguishing between real and generated images. For all but one of the classes, removing "Intensity" features causes the larger decreases in accuracy, suggesting that the main difference between real and generated images are driven by the overall intensities being generated more than the shapes of the cells. However, since the main features that allow the RF to distinguish between real vs generated suggest a difference in the overall intensity distribution between the real and generated images, it may be difficult to accurately calculate CellProfiler features using manual thresholding. Thus, this analysis suggests that there is room for future improvement by better calibrating the distribution of generated intensity values.

3.5 Conclusion

In this study, we demonstrate the effectiveness of LUMIC in predicting cellular morphologies of seen cell types and unseen compounds, unseen cell types and seen compounds, and unseen cell type/unseen compounds combinations. This capability promises to accelerate understanding and characterize the impact of compounds across cell lines/types as well as streamline the early drug discovery process. Inspired by the idea of "style transfer", we treat chemical perturbations as a "style" that can be transferred across various different cell lines. To do this, we first generating an image embedding of the interaction based on the control image and a chemical embedding, before decoding the image embedding into a high resolution image. We validate our findings both computationally and biologically, further indicating the promise of deep generative models in facilitating drug discovery and improved cellular understanding.

Meaningfulness Statement

Explain what you consider a "meaningful representation of life" and how your work contributes to this direction. This section does not count towards the page limit.

A "meaningful representation of life" contains accurate biological meaning, such as cell type and transcriptional state, while maintaining biological variation. LUMIC provides a framework to be able to generate meaningful representations of different cells, a fundamental building block of life, that retain their intended biological information (cell type and perturbational state). Moreover, LUMIC also is able to model the heterogeneity among cells and their perturbed states in both the generated representations as well as generated images by modeling not just the variations of a singular cell but also modeling the growth patterns of a given cell line.

Author Contributions

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors. Do not include author contributions in the anonymous submission.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper. Do not include acknowledgments in the anonymous submission.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. *arXiv preprint arXiv:1801.01401*, 2018.

Anis Bourou, Thomas Boyer, Kévin Daupin, Véronique Dubreuil, Aurélie De Thonel, Valérie Mezger, and Auguste Genovesio. PhenDiff: Revealing invisible phenotypes with conditional diffusion models. *arXiv preprint arXiv:2312.08290*, 2023.

Mark-Anthony Bray, Shantanu Singh, Han Han, Chadwick T Davis, Blake Borgeson, Cathy Hartland, Maria Kost-Alimova, Sigrun M Gustafsdottir, Christopher C Gibson, and Anne E Carpenter. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature protocols*, 11(9):1757–1774, 2016.

Peter D Caie, Rebecca E Walls, Alexandra Ingleston-Orme, Sandeep Daya, Tom Houslay, Rob Eagle, Mark E Roberts, and Neil O Carragher. High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Molecular cancer therapeutics*, 9(6):1913–1926, 2010.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

Srinivas Niranj Chandrasekaran, Jeanelle Ackerman, Eric Alix, D Michael Ando, John Arevalo, Melissa Bennion, Nicolas Boisseau, Adriana Borowa, Justin D Boyd, Laurent Brino, et al. Jump cell painting dataset: morphological impact of 136,000 chemical and genetic perturbations. *BioRxiv*, pp. 2023–03, 2023.

Srinivas Niranj Chandrasekaran, Beth A Cimini, Amy Goodale, Lisa Miller, Maria Kost-Alimova, Nasim Jamali, John G Doench, Briana Fritchman, Adam Skepner, Michelle Melanson, et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nature Methods*, pp. 1–8, 2024.

Michael Doron, Théo Moutakanni, Zitong S Chen, Nikita Moshkov, Mathilde Caron, Hugo Touvron, Piotr Bojanowski, Wolfgang M Pernice, and Juan C Caicedo. Unbiased single-cell morphology with self-supervised vision transformers. *bioRxiv*, pp. 2023–06, 2023.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

Louise Heinrich, Karl Kumbier, Li Li, Steven J Altschuler, and Lani F Wu. Selection of optimal cell lines for high-content phenotypic screening. *ACS chemical biology*, 18(4):679–685, 2023.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet, et al. Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*, 2022.

Shaista Hussain, Ayesha Anees, Ankit Das, Binh P Nguyen, Mardiana Marzuki, Shuping Lin, Graham Wright, and Amit Singhal. High-content image generation for drug discovery using generative adversarial networks. *Neural Networks*, 132:353–363, 2020.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *International conference on machine learning*, pp. 4839–4848. PMLR, 2020.

Vladislav Kim, Nikolaos Adaloglou, Marc Osterland, Flavio M Morelli, Marah Halawa, Tim König, David Gnutt, and Paula A Marin Zapata. Self-supervision advances morphological profiling by unlocking powerful image representations. *BioRxiv*, pp. 2023–04, 2023.

Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Matt J Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar variational autoencoder. In *International conference on machine learning*, pp. 1945–1954. PMLR, 2017.

Scott Lundberg. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*, 2017.

Zeinab Navidi, Jun Ma, Esteban A Miglietta, Le Liu, Anne E Carpenter, Beth A Cimini, Benjamin Haibe-Kains, and Bo Wang. Morphodiff: Cellular morphology painting with diffusion models. *bioRxiv*, pp. 2024–12, 2024.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with CLIP latents, 2022.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

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*, pp. 234–241. Springer, 2015.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494, 2022.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

David R Stirling, Madison J Swain-Bowden, Alice M Lucas, Anne E Carpenter, Beth A Cimini, and Allen Goodman. CellProfiler 4: improvements in speed, utility and usability. *BMC bioinformatics*, 22:1–11, 2021.

Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 18(1):100–106, 2021.

Karren Yang, Samuel Goldman, Wengong Jin, Alex X Lu, Regina Barzilay, Tommi Jaakkola, and Caroline Uhler. Mol2Image: improved conditional flow models for molecule to image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6688–6698, 2021.

A Appendix

A.1 Denoising Diffusion Probablistic Model Background

During training, a data sample x_0 is slowly corrupted through a T step forward Markov chain to create noised samples $x_1, x_2, \dots, x_T$. Crucially, if the noise follows a Gaussian distribution, then since the sum of Gaussians is also a Gaussian, the noised sample x_t at timestep t can be computed efficiently by

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon \tag{1}$$

where $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$ and α_t is the scaling factor controlled by a variance scheduler, and $\epsilon \sim \mathcal{N}(0, I)$ is Gaussian noise. Then, to learn the reverse process, a neural network ϵ_θ parameterized by θ can be trained to predict the noise at each timestep using an L_2 loss where t is the discrete uniform distribution between 1 and T:

$$L(\theta) = \mathbb{E}_{t, x_0, \epsilon}[||\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t)||^2 \tag{2}$$

Finally, after the model has been trained, the reverse process can be used to generate new samples by sampling $x_T \sim \mathcal{N}(0, 1)$ and iteratively denoising from T to 1 using the following equation where $z \sim \mathcal{N}(0, I)$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}}} \epsilon_\theta(x_t, t)) + \sigma_t z \tag{3}$$

To accelerate inference by generating high quality samples in fewer steps, the Denoising Diffusion Implicit Model (DDIM)(Song et al., 2020) can be used instead to sample at a subsequence of the original time schedule, $\tau = [\tau_1, \tau_2, \dots, \tau_S]$ where $S < T$, and is formulated as

$$x_{\tau_{i-1}} = \sqrt{\alpha_{\tau_{i-1}}} \underbrace{\left(\frac{x_{\tau_i} - \sqrt{1 - \alpha_{\tau_i}} \cdot \epsilon_\theta(x_{\tau_i})}{\sqrt{\alpha_{\tau_i}}} \right)}_{\text{`` predicted } x_0\text{''}} + \underbrace{\sqrt{1 - \alpha_{\tau_{i-1}} - \sigma_{\tau_i}^2} \cdot \epsilon_\theta(x_{\tau_i})}_{\text{``direction pointing to } x_t\text{''}} + \underbrace{\sigma_{\tau_i} \epsilon_{\tau_i}}_{\text{random noise}} \tag{4}$$

and

$$\sigma_{\tau_i} = \eta \sqrt{\frac{1 - \alpha_{\tau_{i-1}}}{1 - \alpha_{\tau_i}}} \sqrt{1 - \frac{\alpha_{\tau_i}}{\alpha_{\tau_{i-1}}}} \tag{5}$$

where $\eta \in [0, 1]$ is a hyperparameter that interpolates between 0 to make sampling deterministic and 1 for standard denoising diffusion probabilistic model(DDPM) sampling (Eq. 3).

To enforce conditioning information $\boldsymbol{c}$, such as class labels or latent representations, classifier-free guidance can be used (Ho & Salimans, 2022). This is particularly useful in predicting the outcomes of perturbation experiments, where conditional generation allows for fine-tune control over the desired cell line and chemical interaction by steering the model towards the desired output. The key idea is that given paired data (x_0, c), the model learns both a conditional and an unconditional model by randomly dropping out the label during training. Specifically, at time step t, the learned noise is defined by

$$\epsilon_{guided}(x_t, t, c) = \lambda \epsilon_\theta(x_t, t, c) + (1 - \lambda)(\epsilon_\theta(x_t, t)) \tag{6}$$

where $\epsilon_\theta(x_t, t)$ is the noise prediction without conditioning information, $\epsilon_\theta(x_t, t, c)$ is the noise prediction with conditioning, and λ is the guidance weight.

A.2 Training Details

All three diffusion models were trained independently. During training of the diffusion models, random crops were resampled up to 50 times if the percentage of black pixels exceeded 30% of the entire image. We used random crops instead of single-cell crops in order to capture how the cells grow together and to capture how a chemical may impact this growth. Moreover, if the growth pattern of a cell and compound interaction is very sparse, we want our generated images to reflect this. We did not crops centered around a single cell because our generated dataset was very confluent (many overlapping cells), and it was not possible to accurately do so. Then, the images were randomly flipped horizontally (p = 0.5), channels were randomly dropped out (p = 0.2), and finally images were padded by 50 before being passed into DINO. All models were trained on single NVidia A40 single precision GPUs (unless otherwise mentioned) until the visual quality of the images reached a sufficient level. All diffusion models were also trained using exponential moving average (EMA), averaging over the past ten training weights, which has shown to improve sampling quality(Nichol & Dhariwal, 2021). The EMA model was also used during sampling. A cosine noise schedule with 1,000 timesteps was used during training of all diffusion models, while a linear noise schedule with 1,000 timesteps was used for the low resolution noise schedule for the super-resolution model. For all 3 models, DDIM sampling ($\eta = 0$) was used with 250 steps, which was where the decrease in the KID between the different sampling steps became less drastic. Approximately 1,000 images were generated for each class using the entire generative modeling pipeline (generating the embedding, decoding that embedding into the 64 x 64 image, and generating the 256 x 256 image from the generated embedding and low resolution generated image shown in the vertical dashed box in **Fig. 1 D**).

A.3 Experimental Details for Style Transfer Dataset

To generate the data, multiple cell lines were treated with the same panel of compounds. Each compound was dissolved in DMSO at a concentration of 2 mM and distributed into an Echo Qualified 384-Well Polypropylene Microplates (Beckman Coulter, 001-14615). Using an Echo 655 Liquid Handler, 250 nL of each compound was dispensed into 384-well PhenoPlates (Revvity, 6057302) for a final concentration of 10 µM at a final volume of 50 µL/well.

3T3 (Fibroblasts), A549, HEK293T, HeLa, and RPTE (Kidney) cells were grown to 90% confluency in DMEM/F12 (ThermoFisher Scientific, 11320033) supplemented with 10% FBS (ThermoFisher Scientific, A5256701) in T75 flasks. Media was then aspirated and cells were washed with 10 mL of PBS (ThermoFisher Scientific, 10010023) before lifting with 1 mL of 0.25% Trypsin-EDTA (ThermoFisher Scientific, 25200056). After 5 minutes of incubation, 10 mL of the 10% FBS DMEM/F12 was added into each flask to remove cells. Cells were spun down at 300 x g for 5 minutes to pellet and resuspended in fresh 10% FBS DMEM/F12 before counting by hemocytometer. Cells were diluted to a final concentration of 120,000 cells/mL and 50 µL was dispensed into each well for a final seeding density of 6,000 cells/well.

Plates were incubated at $37°C$ for 24 hours before fixing and staining using the same stains as the JUMP Pilot data (DNA, AGP, Mito) in a CellPainting style assay (Bray et al., 2016). We used a Yokogawa CQ1 High-Content Imaging system to image multiple fields of view for each well, for a total of 3,168 images. We randomly select 10 compounds to hold out across all 5 cell lines and hold out HeLa completely (except control images) during training to make up the test set.

For both datasets, the images were normalized using sklearn's Quantile Transformation with respect to the controls (Pedregosa et al., 2011). For the style transfer dataset, Cellpose was used to identify the nuclei of the cells and create a binary mask (Stringer et al., 2021), and CellProfiler was used to segment the images to remove any residual dye in the background (Stirling et al., 2021).

A.4 DINO Training Details

During training, we used additional pre-processing steps and transformations including taking random 338x338 crops and resizing them to 224x224 and randomly dropping out channels (Doron et al., 2023). During sampling, we padded images by 50 pixels on each side to help DINO focus on the content rather than gaps between cells, since some of the cells cultured for the style transfer dataset were partially to fully confluent, as shown by visualizing the attention maps.

	A549	HEK293T	Fibroblast	Kidney	HeLa
A549	**0.013**	0.214	0.082	0.078	0.043
HEK293T	0.256	**0.007**	0.134	0.109	0.022
Fibroblast	0.082	0.094	**0.011**	0.067	0.064
Kidney	0.053	0.161	0.035	**0.016**	0.064
HeLa	0.049	0.204	0.101	0.109	**0.035**

Table 7: KID of real vs. generated cell lines. Rows represent real images from each cell line; columns are generated images. Lower KID is better. Lower on-diagonal KID and higher off-diagonal KID indicate that LUMIC beats the baseline.

A.5 Tables

Class	Corresponding Class	Control Images of Same Cell Type
U2OS Seen Gene	**0.029**	0.040
U2OS Unseen Gene	**0.029**	0.054
A549 Seen Gene	**0.016**	0.028
A549 Unseen Gene	**0.014**	0.035

Table 8: Average KID across seen and unseen target genes on seen cell lines from the JUMP dataset.

Cell Line	Corresponding Class	Control Images of Same Cell Type
A549	**0.051**	0.113
Fibroblast	**0.025**	0.107
Kidney	**0.017**	0.035
HEK293T	**0.009**	0.045

Table 9: Average KID of unseen compounds on various seen (A549, Fibroblast, Kidney, HEK293T) cell lines.

Compounds	Corresponding Class	Control Images of Same Cell Type
Seen Compounds	**0.075**	0.079
Unseen Compounds	**0.074**	0.075

Table 10: Average KID of seen and unseen compounds on HeLa (unseen cell lines).

	RF Accuracy	RF Accuracy after removing features containing ”Intensity”	RF Accuracy after removing features containing ”AreaShape”
A549 Seen	0.96	0.95 (0.01)	**0.92 (0.04)**
A549 Unseen	0.89	**0.82 (0.07)**	0.89 (0.00)
U2OS Seen	0.85	**0.76 (0.09)**	0.84 (0.01)
U2OS Unseen	0.82	**0.75 (0.07)**	0.79 (0.03)
A549 (Style Transfer)	0.81	**0.76 (0.05)**	0.77 (0.04)
HEK293T	0.69	**0.65 (0.04)**	0.69 (0.00)
Kidney	0.86	**0.77 (0.09)**	0.85 (0.01)
Fibroblast	0.8	**0.73 (0.07)**	0.8 (0.00)
HeLa Seen	0.92	**0.89 (0.03)**	0.91 (0.01)
HeLa Unseen	0.93	0.9 (0.03)	0.9 (0.03)

Table 11: The RF Accuracy and Number of Features after Feature Selection to Classify Between Real and Generated CellProfiler Features for each distinct combination of cell line + seen/unseen compounds

A.6 Supplemental Figures

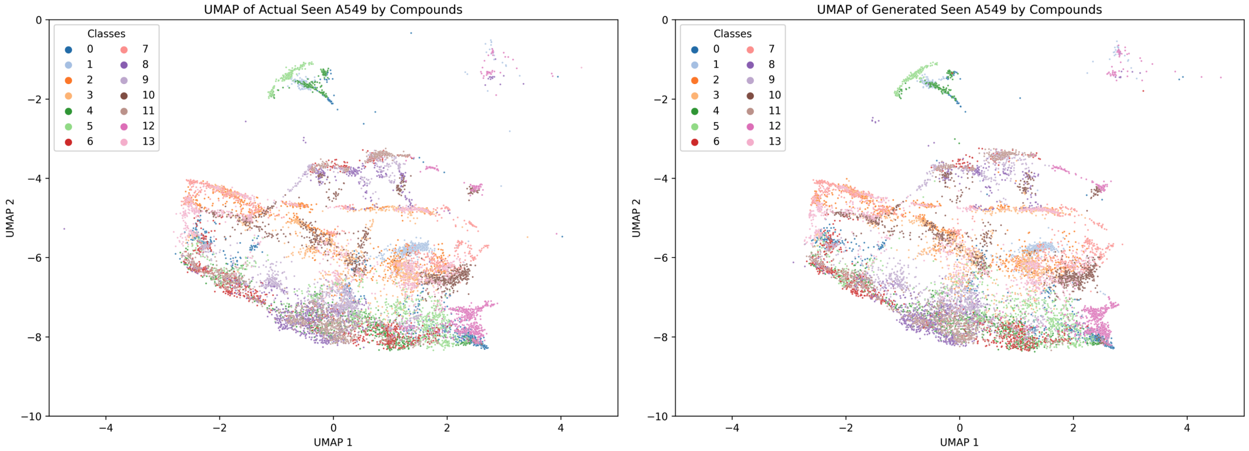

Figure A1: UMAP of real (left) and generated (right) DINO embeddings for the test set compounds that target seen genes on A549 from the JUMP dataset

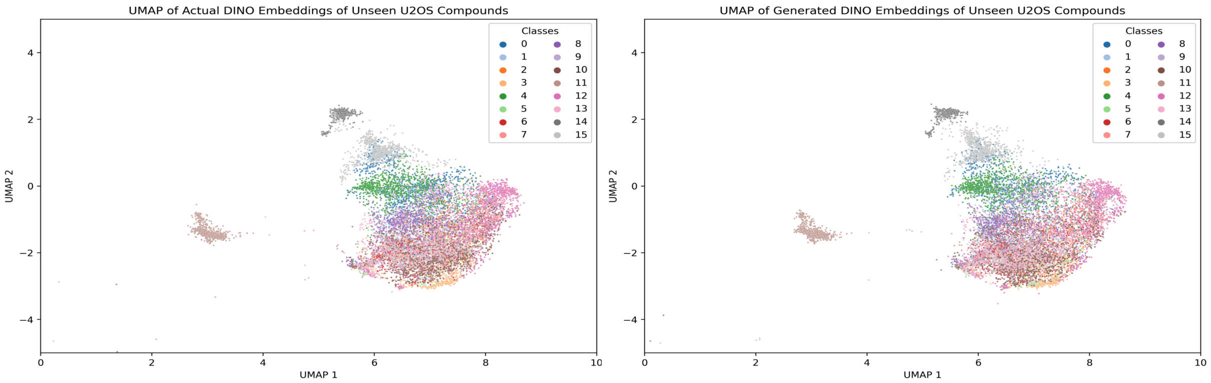

Figure A2: UMAP of real (left) and generated (right) DINO embeddings for the test set compounds that target unseen genes on U2OS from the JUMP dataset

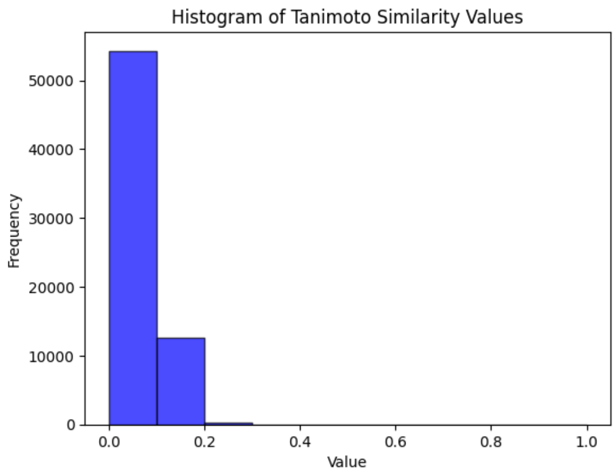

Figure A3: Histogram of the Tanimoto similarity for all possible unique pairs of the compounds in the JUMP pilot 1 dataset and Style Transfer Dataset

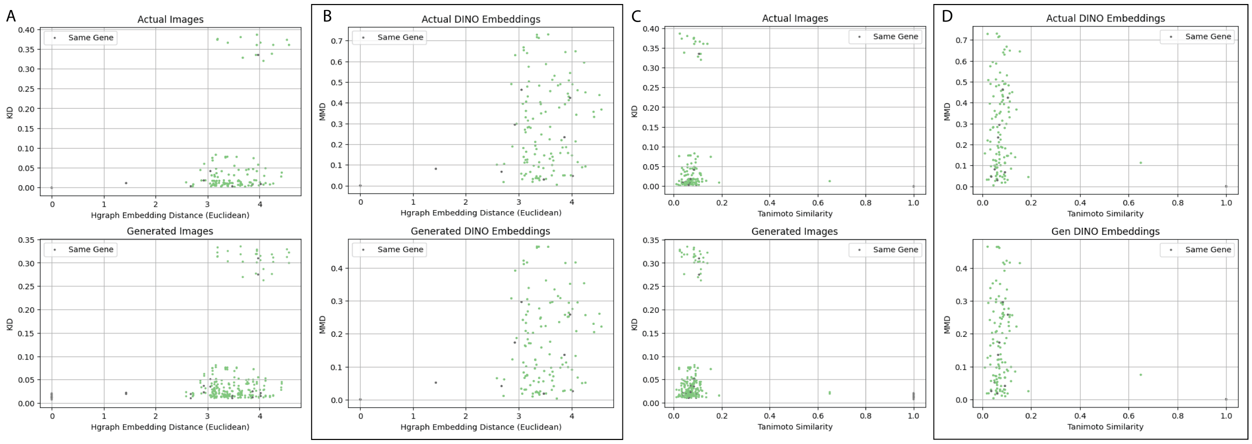

Figure A4: Scatter plots for A549 Unseen Genes. (A) KID against HGraph embedding distance, with real images on the top and the generated images on the bottom. (B) MMD against HGraph embedding distance, with real embeddings on the top and the generated embeddings on the bottom. (C) KID against the Tanimoto similarity with the real images on the top and the generated images on the bottom. (D) MMD against the Tanimoto similarity with the real embeddings on the top and the generated embeddings on the bottom.

A.7 Generated Samples

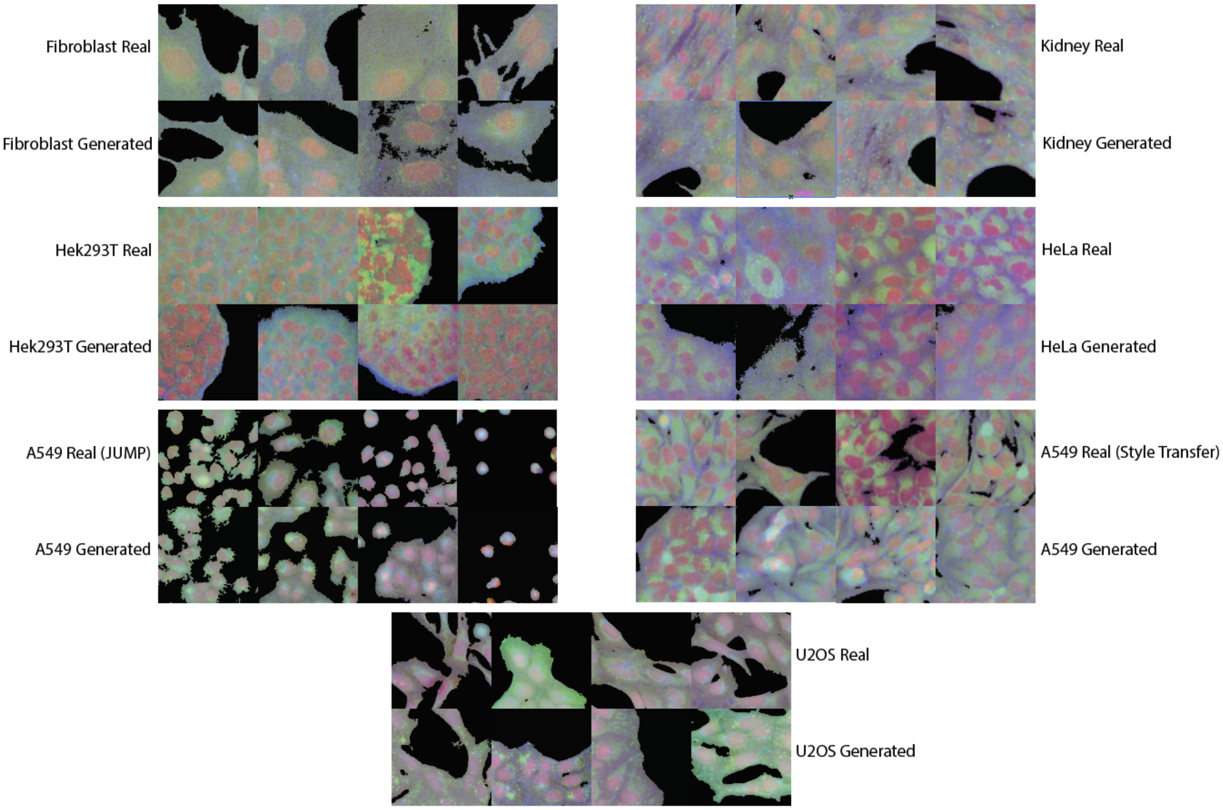

Figure A5: Real and Generated Images for All Cell Types (counting A549 as 2 different cell types for each appearance in both datasets)

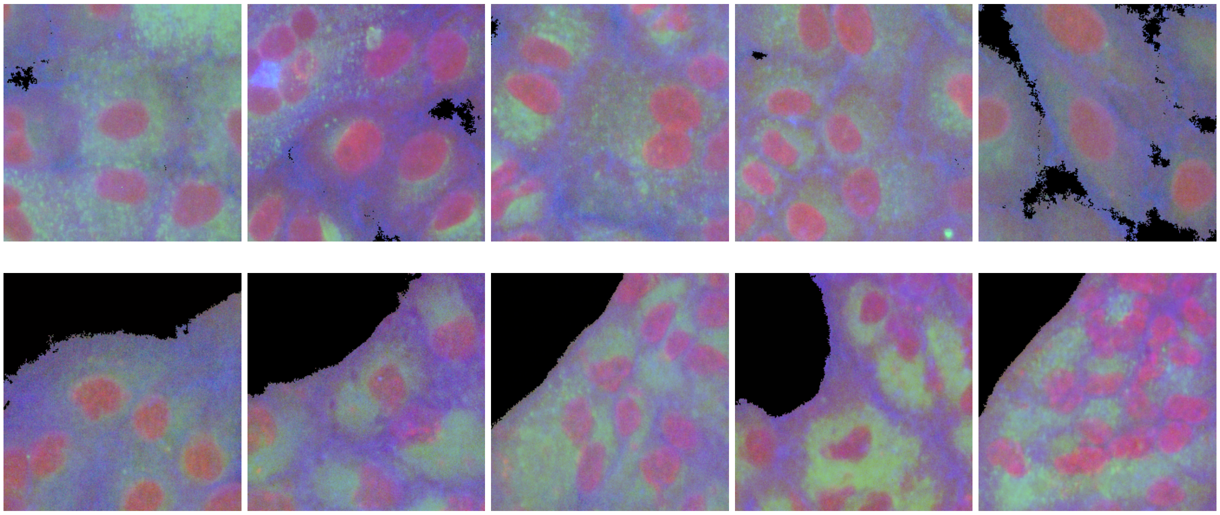

Figure A6: Actual (top row) and generated (bottom row) images of HeLa cells (unseen cell line) treated with propylthiouracil (unseen compound) from the Style Transfer Dataset

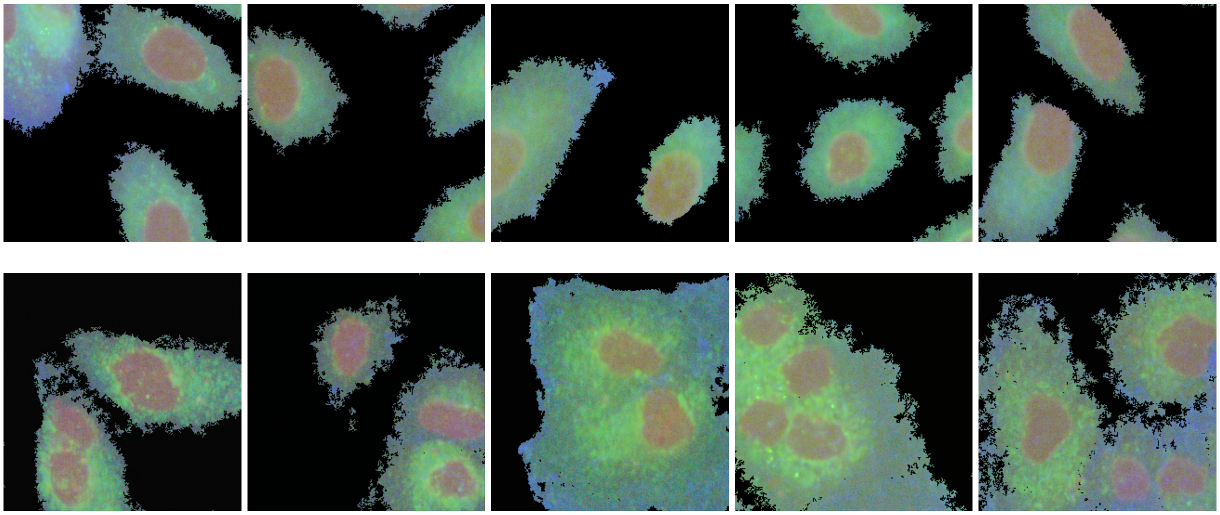

Figure A7: Actual (top row) and generated (bottom row) images of HeLa cells (unseen cell line) treated with methotrexate (seen compound) from the Style Transfer Dataset

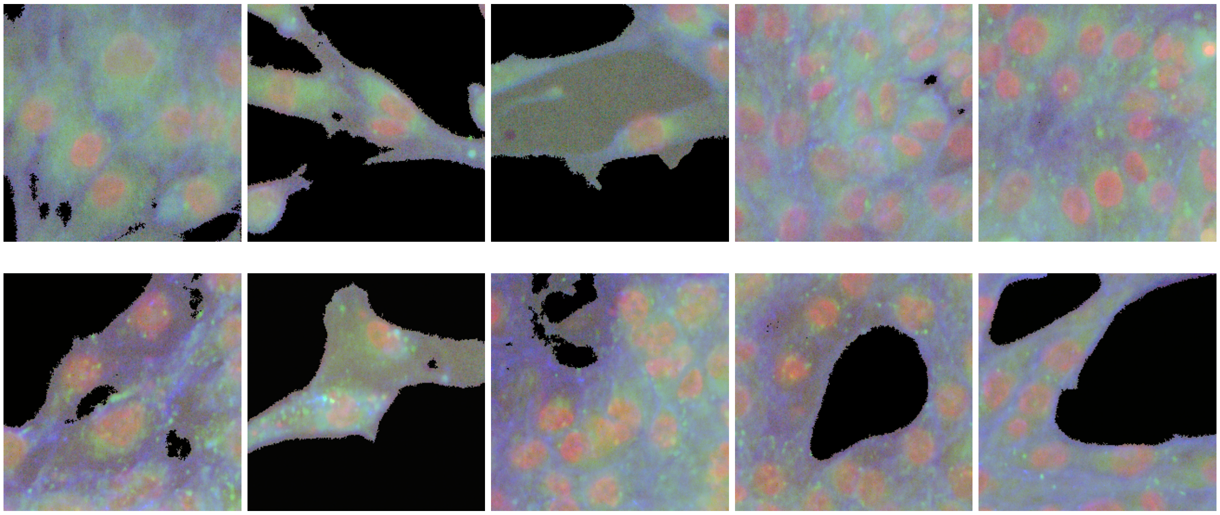

Figure A8: Actual (top row) and generated (bottom row) images of RPTE cells (seen cell line) treated with diclofenac (unseen compound) from the Style Transfer Dataset

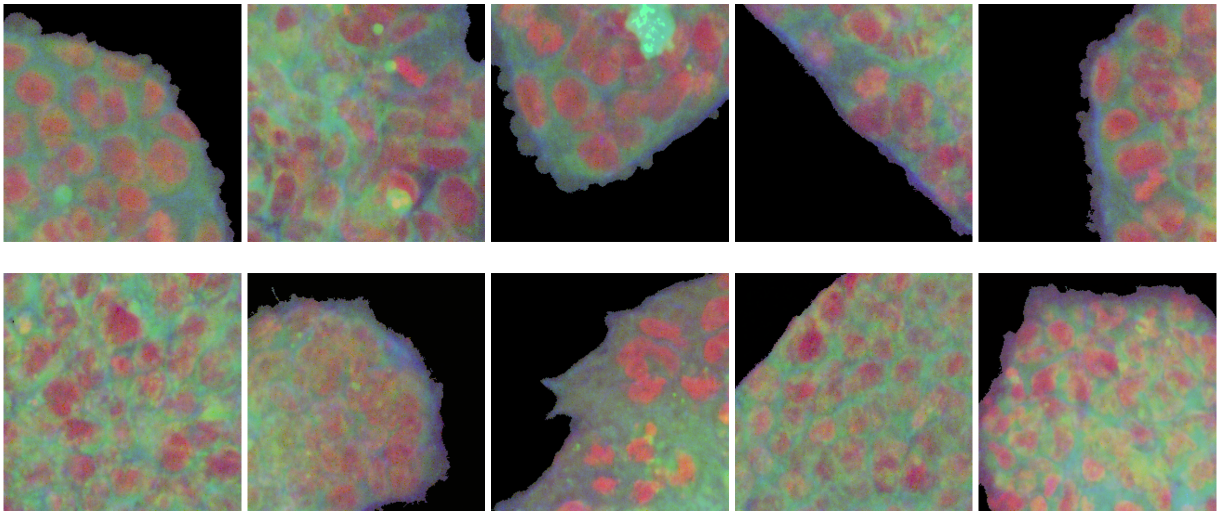

Figure A9: Actual (top row) and generated (bottom row) images of HEK293T cells (seen cell line) treated with sulfamethoxazole (unseen compound) from the Style Transfer Dataset

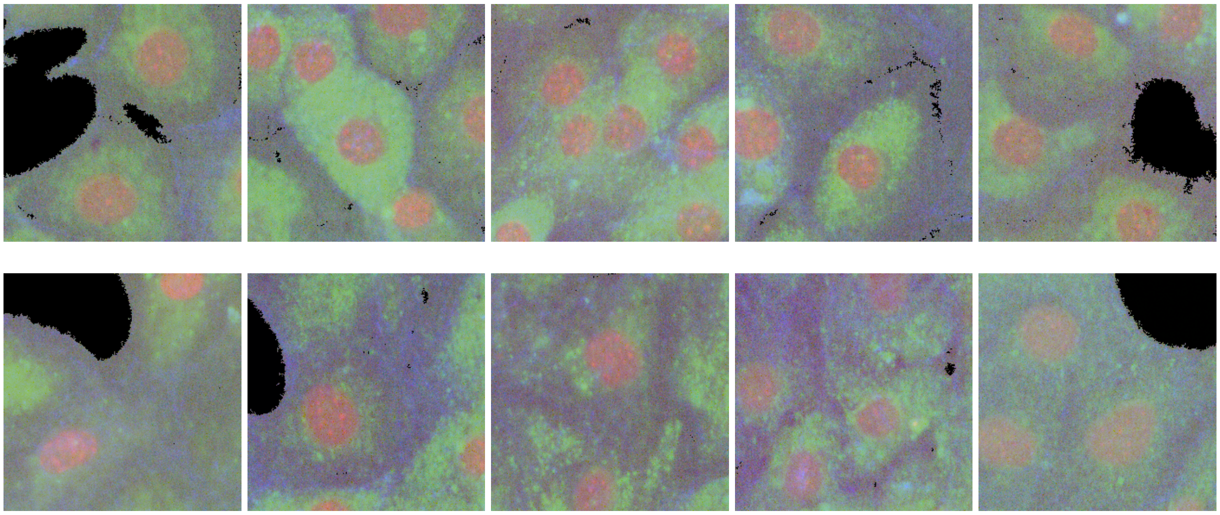

Figure A10: Actual (top row) and generated (bottom row) images of 3T3 cells (seen cell line) treated with chlorpromazine (unseen compound) from the Style Transfer Dataset

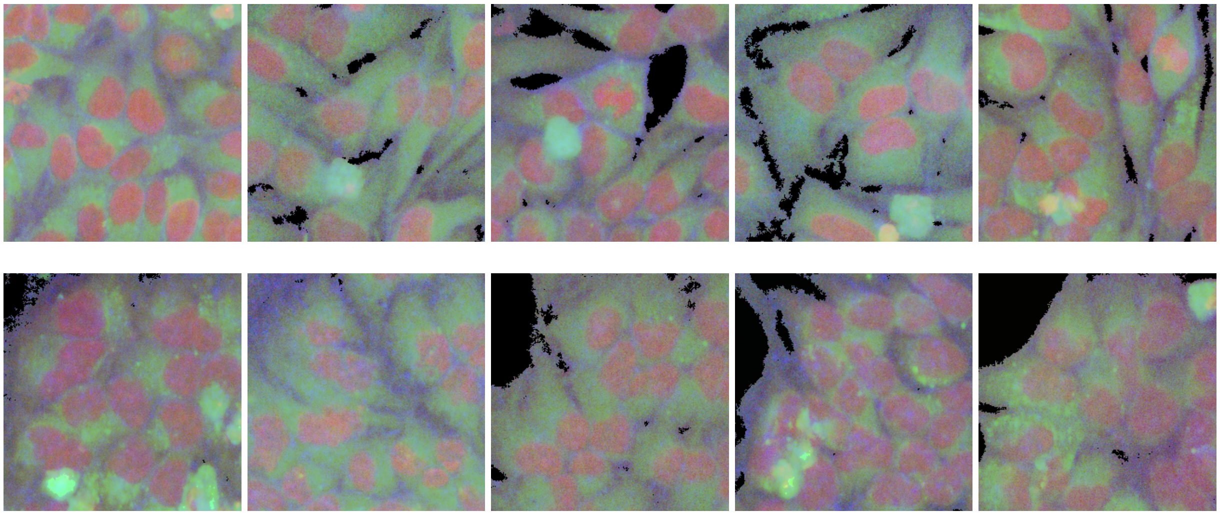

Figure A11: Actual (top row) and generated (bottom row) images of A549 cells (seen cell line) treated with busulfan (unseen compound) from the Style Transfer Dataset

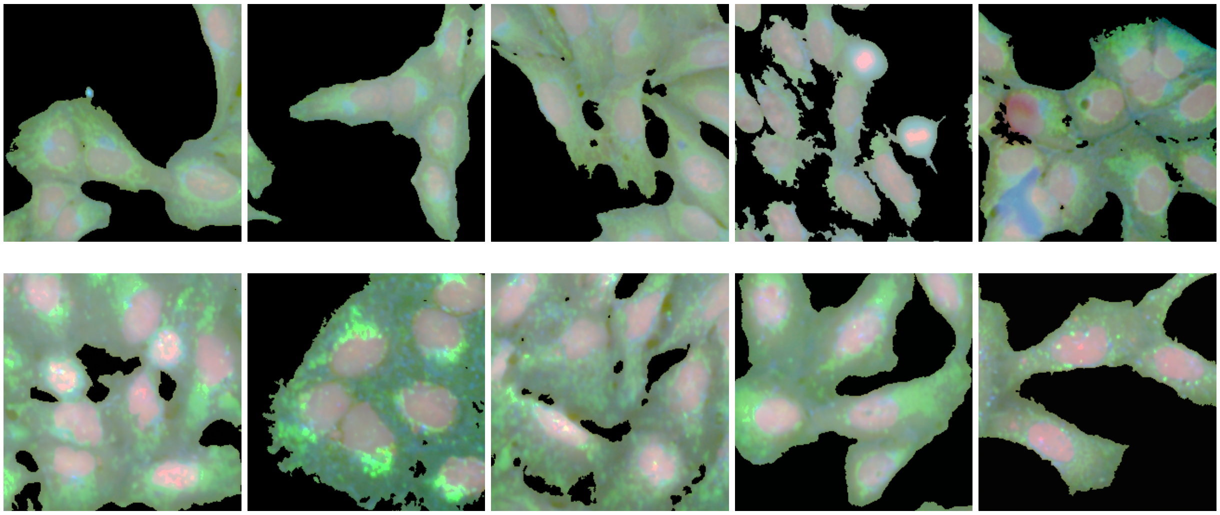

Figure A12: Actual (top row) and generated (bottom row) images of U2OS cells (seen cell line) treated with L-Pyroglutamic acid (unseen compound), which targets an unseen gene from the JUMP Pilot 1 Dataset

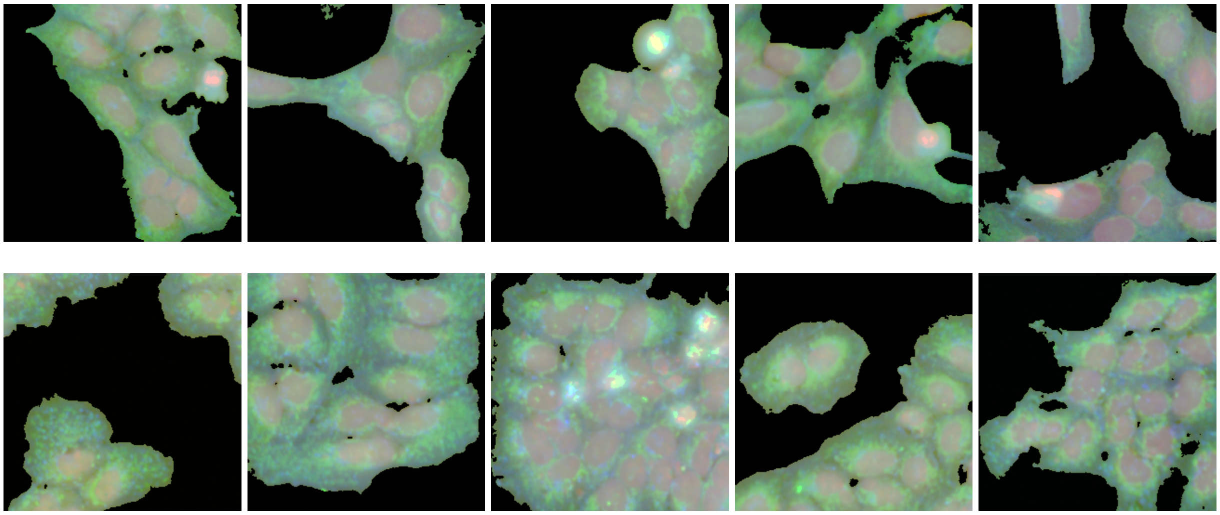

Figure A13: Actual (top row) and generated (bottom row) images of U2OS cells (seen cell line) treated with spebrutinib (unseen compound), which targets a seen gene from the JUMP Pilot 1 Dataset

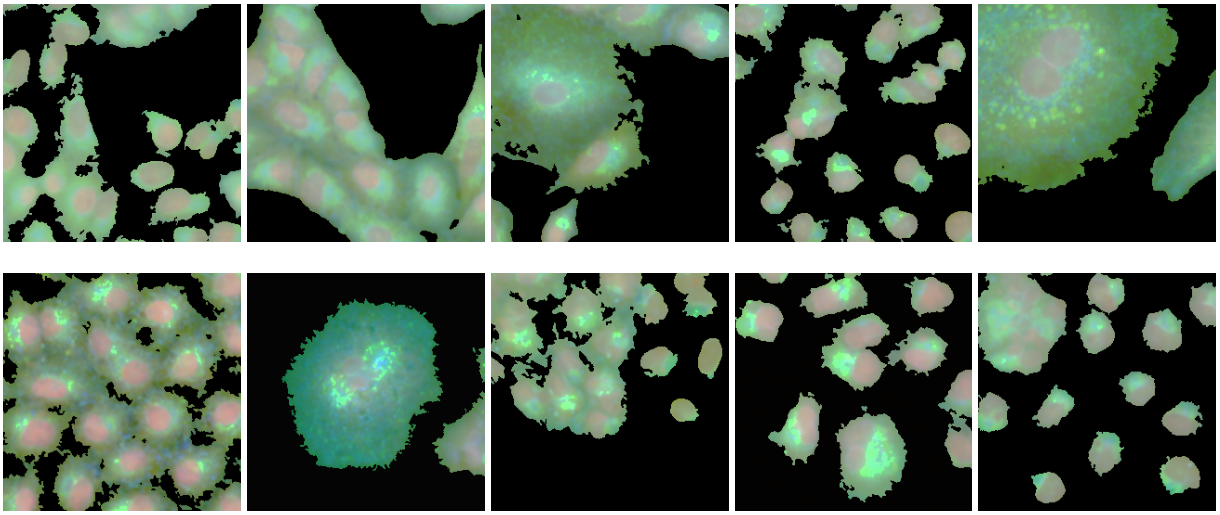

Figure A14: Actual (top row) and generated (bottom row) images of A549 cells (seen cell line) treated with pentostatin (unseen compound), which targets an unseen gene from the JUMP Pilot 1 Dataset

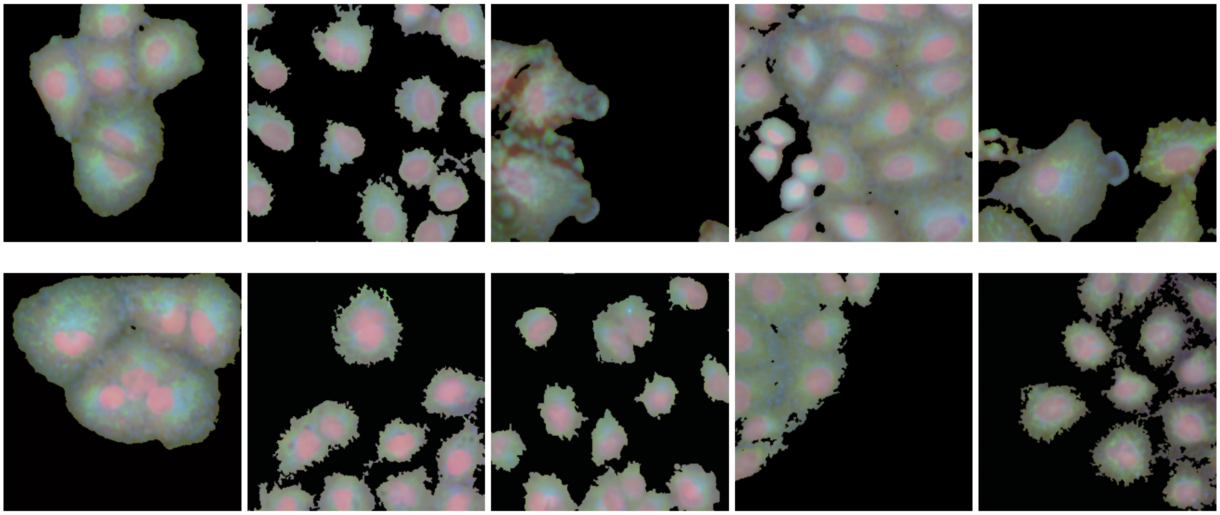

Figure A15: Actual (top row) and generated (bottom row) images of A549 cells (seen cell line) treated with dexamethasone (unseen compound), which targets a seen gene from the JUMP Pilot 1 Dataset