CONDITIONAL SINGLE-CELL RNA GENERATION: A DECODER MODEL AND A BENCHMARK OF GENERATIVE MODELS

Skander Karkar
Mediterranean Institute of Technology
Tunis, Tunisia
skander.karkar@medtech.tn

Olav De Clerck
Deeplife
Paris, France
olav.declerck@deeplife.co

ABSTRACT

Generative models for single-cell RNA data have advanced rapidly, yet their relative strengths, limitations, and biological faithfulness remain poorly understood. In particular, recent generative diffusion and transformer models have not been systematically compared to each other, and standard generation metrics may fail to capture biologically relevant behavior. In this work, we present a benchmark of recent generative models for conditional scRNA-seq generation, including a diffusion model (scDiffusion), a transformer model (C2S-Scale), a variational autoencoder (SCVI which we adapt for conditional generation) and a decoder-only transformer model that we propose and that in contrast to C2S preserves exact count information. We evaluate the four models using a diverse set of distributional, discriminability, and gene-level metrics, and propose the recovery of differentially expressed genes as a biologically meaningful downstream task. While all models generate visually and distributionally realistic cells, the diffusion model being the weakest, we find that no method reliably preserves the log-fold changes required for standard DGE analysis, revealing a systematic attenuation of biological signal.

1 INTRODUCTION

Neural networks have recently been developed for the analysis of scRNA data. Models span a wide range of architectures, including transformers (scBERT (Boiarsky et al. (2023)), GeneFormer (Theodoris et al. (2023)), scGPT (Li et al. (2024)), scFoundation (Wang et al. (2023)), UCE (Rosen et al. (2023)), scMulan (Zhu et al. (2024)), C2S-Scale (Levine et al. (2024); Rizvi et al. (2025)), scPRINT (Kalfon et al. (2024; 2025)), scMPT (Palayew et al. (2024))), GANs (scGAN (Marouf et al. (2020)), LSH-GAN (Lall et al. (2022))), VAEs (SCVI (Lopez et al. (2018)) and its variants such as totalVI (Gayoso et al. (2021)) and MultiVI (Ashuach et al. (2023))), GNNs (PDGrapher (Zhang et al. (2023)), GEARS (Wang et al. (2021)), scNET (Sheinin et al. (2025))), and diffusion models (scDiffusion (Luo et al. (2024)), scGND (Liu et al. (2024))).

While the performances of these models have been well evaluated on tasks such as cell type prediction and batch correction, their ability to generate biologically faithful synthetic cells remains less well understood. In particular, conditional generation, i.e. producing realistic transcriptomes given metadata such as tissue or disease, has been relatively under-explored, and existing models have largely been evaluated in isolation using different metrics. Indeed, looking at more recent generative models such as C2S-Scale (Rizvi et al. (2025)), scDiffusion (Luo et al. (2024)), scMulan (Zhu et al. (2024)) and scGND (Liu et al. (2024)), one finds that they have either not been compared to other generative models, or only to older GANs such as scGAN (Marouf et al. (2020)) and adapted older foundational models such as scGPT (Li et al. (2024)) that all perform poorly on generation tasks, and certainly not to each other. C2S-Scale is the largest model and has been trained on the most data, but it performs weakly on the (non-generative) benchmarks in (Zhang et al. (2025)). We aim to address this gap by providing a comparison of models of different types for conditional scRNA generation, using generative metrics and the downstream biological task of retrieving differentially expressed genes. In addition to a diffusion model (scDiffusion) and a transformer model (C2S-Scale), we adapt the popular variational autoencoder SCVI into a conditional VAE, and propose a

decoder-only transformer architecture, which we call scDecoder, which differs from C2S-Scale in that it uses exact count information instead of ranking genes by their counts.

2 MODEL

Since C2S-Scale takes as input a ranked list of genes without their exact expression levels, a linear regression between the log-count and log-rank is learned on the data to recover exact counts after generation. As in the C2S paper, we find that this regression captures around 80% of the variation in gene counts. However, we would like to compare this model to a similar decoder-only transformer that is trained on exact count values. We therefore propose a model that is a decoder transformer, called scDecoder in the following. After trainable embedding and position embedding layers for the gene and metadata tokens, the transformer is made up of 4 successive transformer blocks containing 4 heads of causal multi-head self-attention with an embedding dimension of 64. To reduce the input length, the 2000 most highly variable genes are selected, which is a very common preprocessing step (Li et al. (2024)), and the counts are normalized in each cell to sum to 1000. Since the model requires integer counts, they are then rounded up to avoid creating false zeros. Each gene name is a token and is repeated as many times as its count in the input sequence to the model, which is quite similar to the UCE model which samples with replacement 1024 genes in proportion to their counts. The metadata information such as cell type disease and sample ID is appended to the start of the sequence. A metadata term is set to 'unknown' with a 5% probability. So the input sequence looks like: $(\text{[start-token]}, \text{liver}, \text{cancer}, \text{[gene-start-token]}, A, A, A, B, D, D, \text{[end-token]})$, for a liver cell with cancer where gene A appears 3 times, gene B once, gene D twice, and gene C does not appear. The model is trained with a next-token prediction loss, similarly to C2S-Scale, and in contrast to models such as GeneFormer, which predict randomly masked tokens. This allows for natural conditional generation based on metadata tokens. Since the genes are not ordered based on expression levels as in C2S-Scale and can be shuffled in the input sequence, this also allows for generating the rest of a transcriptome when knowing the expression levels of some genes.

3 EXPERIMENTS

3.1 DATASET AND BASELINES

Our inflammatory bowel disease dataset comprises gene expression profiles from 594,959 human cells, spanning 58,676 genes across 13 gastrointestinal tissues. The dataset, unified from 14 different datasets in the Gene Expression Omnibus (GEO) database (R et al. (2002); T et al. (2013)) [1], includes only human-derived cells and was preprocessed using 10x Genomics Cell Ranger v7.1.0 (Zheng et al. (2017)). The dataset contains cells associated with nine different colon diseases. We focus on the tissue and disease columns for conditional generation.

We compare our model, which we call scDecoder, to two recent state-of-the-art generative models for scRNA data, C2S-Scale (Rizvi et al. (2025)) and scDiffusion (Luo et al. (2024)). We also compare to the popular SCVI model (Lopez et al. (2018)), which is a reference model for embedding scRNA data, and being a VAE, can be made conditional. It is the smallest model of the four, while C2S is by far the largest: scDecoder has 8 millions parameters, we finetune the 160 million parameter version of C2S-Scale because as in (Zhang et al. (2025)), we find that the C2S models with 160, 410 and 1 billion parameters perform similarly, scDiffusion has 13.5 million parameters, including the 8.5 million parameters of the autoencoder, and finally SCVI has 1.4 million parameters.

3.2 GENERATIONS METRICS

We report in Table 1 generation metrics evaluated on Crohn, healthy, and all cells. No single model dominates all metrics. scDiffusion achieves the lowest Earth Mover's Distance (EMD) on the combined dataset, but performs poorly on most other metrics. scDecoder has the lowest logistic-regression accuracy (LR ACC), suggesting that its generated cells are hardest to distinguish from real

[1] The datasets are GSE182270, GSE202052, GSE214695, GSE221987, GSE225199, GSE114374, GSE116222, GSE134649, GSE144469, GSE148837, GSE156776, GSE150516, GSE157477, GSE164985

ones, and also the highest top-genes overlap. C2S-Scale, SCVI ans scDecoder perform comparably on the other metrics. UMAPs, QQ-plots and correlation plots are in Figure 2 in the Appendix.

Table 1: Metrics between generated and real IBD cells. EMD: Earth mover's distance (Wasserstein W_1 distance). LR ACC: the training accuracy of a logistic regression trained to distinguish real cells from generated cells. MSE MEAN: mean squared error between the average gene expression levels across real and generated cells. MSE ZP: mean squared error between the proportion of cells in which a gene is not expressed across real and generated cells. scFID: a generation metric imitating the FID for image generation using scGPT as in C2S-Scale Rizvi et al. (2025). Top genes $\%$: the percentage of the 20 most expressed genes in the real data that are among the 20 most expressed in the generated data. Crohn refers to cells from the mucosa of the ascending colon with Crohn's disease and Healthy refers to healthy cells from the mucosa of the ascending colon. Results are averaged with standard deviations over 5 training then generation runs.

Metric	Cells	scDecoder	scDiffusion	C2S-Scale	Cond-SCVI	CFGen
EMD (↓)	Crohn	201 ± 1.43	**183** ± 0.65	193 ± 0.89	184 ± 0.78	201 ± 1.31
	Healthy	196 ± 1.28	230 ± 0.82	189 ± 0.66	**171** ± 0.15	197 ± 2.04
	All	199 ± 1.77	**171** ± 1.21	188 ± 1.11	178 ± 0.73	199 ± 2.16
LR ACC % (↓)	Crohn	**88.7** ± 2.44	96.0 ± 1.11	97.7 ± 0.94	98.3 ± 0.51	94.8 ± 1.29
	Healthy	**88.6** ± 2.43	92.8 ± 1.34	94.8 ± 1.24	97.7 ± 0.74	93.7 ± 1.05
	All	**82.2** ± 3.01	91.5 ± 2.04	94.2 ± 1.89	96.6 ± 0.77	91.4 ± 2.34
MSE MEAN (↓)	Crohn	**0.49** ± 0.01	1.71 ± 0.02	0.60 ± 0.02	0.94 ± 0.01	**0.49** ± 0.01
	Healthy	0.36 ± 0.01	1.22 ± 0.02	0.20 ± 0.01	0.48 ± 0.01	**0.18** ± 0.01
	All	0.29 ± 0.01	2.36 ± 0.02	0.28 ± 0.01	0.58 ± 0.02	**0.23** ± 0.01
MSE ZP (↓)	Crohn	7×10^{-4} ± 0	0.30 ± 0.04	1×10^{-3} ± 0	$\mathbf{6}\times10^{-4}$ ± 0	1×10^{-3} ± 0
	Healthy	$\mathbf{3}\times10^{-4}$ ± 0	0.31 ± 0.04	2×10^{-3} ± 0	$\mathbf{3}\times10^{-4}$ ± 0	5×10^{-4} ± 0
	All	$\mathbf{3}\times10^{-4}$ ± 0	0.31 ± 0.05	1×10^{-3} ± 0	$\mathbf{3}\times10^{-4}$ ± 0	7×10^{-4} ± 0
scFID (↓)	Crohn	0.015 ± 0	0.31 ± 0.02	0.015 ± 0	**0.013** ± 0	0.017 ± 0
	Healthy	**0.011** ± 0	0.31 ± 0.01	0.017 ± 0	0.015 ± 0	0.018 ± 0
	All	$\mathbf{7}\times10^{-3}$ ± 0	0.30 ± 0.01	0.011 ± 0	0.011 ± 0	0.014 ± 0
Top genes % (↑)	Crohn	**79.8** ± 0.4	24.0 ± 1.9	54.6 ± 2.0	70.6 ± 0.6	60 ± 0
	Healthy	**79.8** ± 0.4	22.1 ± 1.5	55.4 ± 1.9	69.6 ± 0.6	60 ± 0
	All	**80.4** ± 0.4	25.0 ± 0.1	56.8 ± 0.7	70.2 ± 0.5	60 ± 0

3.3 Differential gene expression

Differential gene expression (DGE) analysis aims at identifying genes that are expressed at statistically significantly different levels (either up-regulated i.e over-expressed or down-regulated i.e under-expressed) between two conditions, for example between healthy cells and cells with a given disease or between cells before and after treatment. DGE is used for understanding disease mechanisms and identifying potential biomarkers and drug targets. Since DGE is an important tool used by biostatisticians, an important downstream task for conditional generative models of single-cell data is to check whether the DGE analyzes between two conditions on real and generated data are close, i.e if the same up-regulated and down-regulated genes are found. DGE analysis is a statistical test on the log of the ratio between the expression levels of a gene in two conditions. A cutoff is applied to the log-fold change (LFC) and a cutoff to the (adjusted) p-value to select the up-regulated and down-regulated genes. Before performing differential gene analysis, the data is typically pseudo-bulked. Pseudo-bulking means aggregating (summing or taking the mean) the counts of each gene across cells from the same sample (i.e from the same organ and cell-type from the same patient). In the data, each individual observation comes with a sample ID that allows this aggregation, which serves to reduce noise and dropout zeros. We therefore condition the models on this sample ID metadata term as well, allowing them to generate cells from different fictitious samples which can be pseudo-bulked before differential gene analysis on the generated data.

We conduct DGE analysis on the real data to extract the up- and down-regulated genes in Crohn's disease. We find that the models replicate the general trends observed in the data on the differentially expressed genes, but that they attenuate the differences in expressions (Figure 1), making no genes appear in the DGE analysis on the generated data for the common values of p-value and LFC cutoff

used on real RNA data. In Figure 1, SCVI outperforms the other models by having more distinct gene expressions between the two conditions. The same trends are observed if we simply plot the mean expression per gene across cells without pseudo-bulking (Figure 4 in the Appendix). A linear regression of the log fold change of the generated gene counts against the log fold change of the real gene counts has a slope of 0.36 for SCVI, 0.24 for C2S and 0.2 for scDecoder, confirming that the magnitude of the change in expression is reduced by these models.

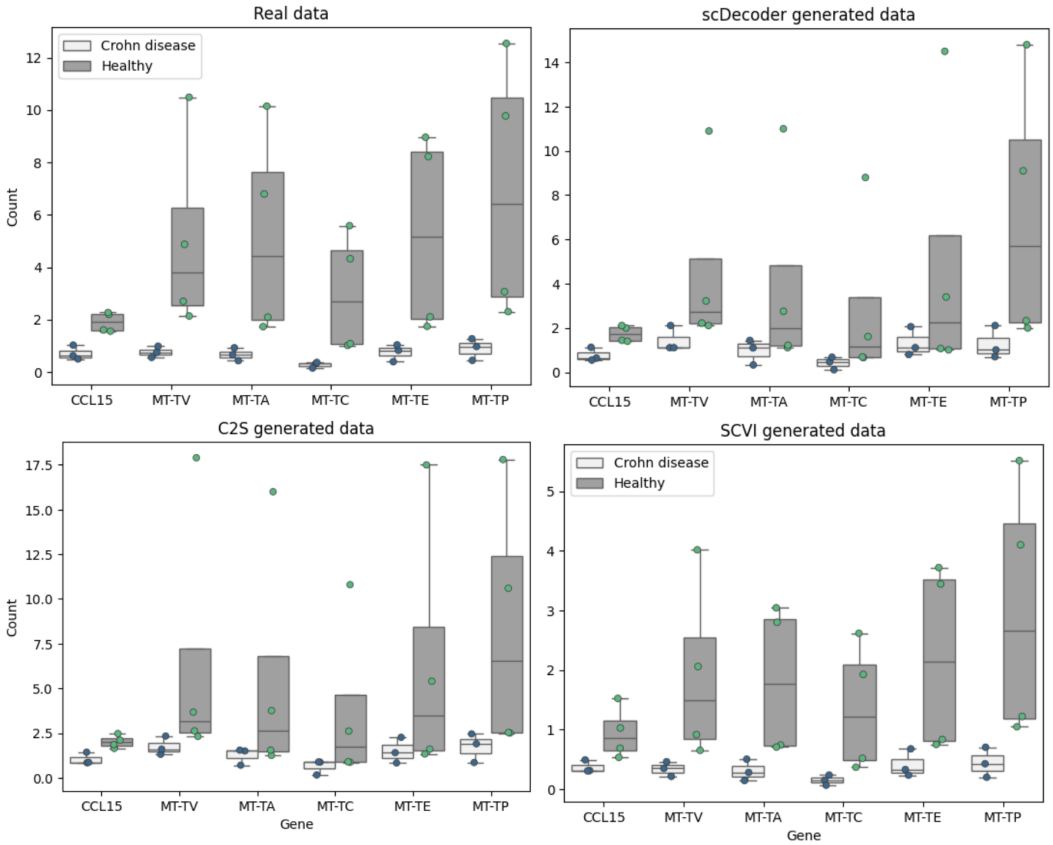

Figure 1: Box-plots of gene expressions in real and generated data of the genes that are down-regulated in the real data after pseudo-bulking.

4 Discussion and Conclusion

Compared to previous work, we focus on generative metrics and find that scDiffusion is the weakest of the models tested. It should also be noted that the smallest model, SCVI, performs competitively on most metrics and the best on DGE retrieval. This could be due to the fact that it models gene counts using a ZINB (zero inflated negative binomial), perhaps indicating that probabilistic modeling of scRNA data is useful, compared to only relying on ever larger models. However for all models, we find that even though the up-regulated genes in the real data are indeed more expressed in the generated data, the precision of the models is not yet good enough for these genes to be reliably retrieved after statistical significance analysis. Future research directions could include exploring how the performance of our model and of conditional SCVI scales with more data and more parameters, initializing gene embeddings to the ESM (Lin et al. (2023)) embeddings, exploring generalization to unseen conditions using LLM embeddings of metadata terms and improving the retrieval of up-regulated and down-regulated genes from generated data. The LFC cutoff and p-value used on generated data could be seen as hyperparameters that are fitted together with the model to achieve a desired precision and recall. This together with generalization to unseen conditions could provide indications for disease markers and targets before conducting real-life experiments.

REFERENCES

Tal Ashuach, Mariano I. Gabitto, Rohan V. Koodli, Giuseppe-Antonio Saldi, Michael I. Jordan, and Nir Yosef. Multivi: deep generative model for the integration of multimodal data. *Nature Methods*, 20:1222—-1231, 2023.

Rebecca Boiarsky, Nalini Singh, Alejandro Buendia, Gad Getz, and David Sontag. A deep dive into single-cell rna sequencing foundation models. *bioRxiv*, 2023. doi: 10.1101/2023.10.19.563100. URL https://www.biorxiv.org/content/10.1101/2023.10.19.563100v1.

Adam Gayoso, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L. Nazor, Aaron Streets, and Nir Yosef. Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nature Methods*, 18:272—282, 2021.

Jérémie Kalfon, Jules Samaran, Gabriel Peyré, and Laura Cantini. scprint: pre-training on 50 million cells allows robust gene network predictions, 2024. URL bioRxiv.

Jérémie Kalfon, Gabriel Peyré, and Laura Cantini. scprint-2: Towards the next-generation of cell foundation models and benchmarks, 2025. URL bioRxiv.

Snehalika Lall, Sumanta Ray, and Sanghamitra Bandyopadhyay. Lsh-gan enables in-silico generation of cells for small sample high dimensional scrna-seq data. *Communications Biology*, 2022.

Daniel Levine, Syed Asad Rizvi, Sacha Levy, Nazreen Pallikkavaliyaveetil, David Zhang, Xingyu Chen, Sina Ghadermarzi, Ruiming Wu, Zihe Zheng, Ivan Vrkic, Anna Zhong, Daphne Raskin, Insu Han, Antonio Henrique De Oliveira Fonseca, Josue Ortega Caro, Amin Karbasi, Rahul M. Dhodapkar, and David Van Dijk. Cell2sentence: teaching large language models the language of biology. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Zhuoqi Li, Yixuan Li, Xiaoyuan Zhang, Xiaoman Zhang, and Xiaoliang Xie. scgpt: Towards building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, 21, 2024.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://www.science.org/doi/abs/10.1126/science.ade2574.

Yu-Chen Liu, Anqi Zou, Simon Liang Lu, Jou-Hsuan Lee, Juexin Wang, and Chao Zhang. scgnd: Graph neural diffusion model enhances single-cell rna-seq analysis, 2024. URL bioRxiv.

Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058, 2018. doi: 10.1038/s41592-018-0229-2. URL https://www.nature.com/articles/s41592-018-0229-2.

Erpai Luo, Minsheng Hao, Lei Wei, and Xuegong Zhang. scdiffusion: conditional generation of high-quality single-cell data using diffusion model. *Bioinformatics*, 2024.

Maher Marouf, Paul Machart, Vikas Bansal, Christoph Kilian, Daniel S. Magruder, Christof F. Krebs, and Stefan Bonn. Realistic in silico generation and augmentation of single-cell rna-seq data using generative adversarial networks. *Nature Communications*, 11(1):1–12, 2020. doi: 10.1038/s41467-020-17721-6. URL https://www.nature.com/articles/s41467-020-17721-6.

Steven Palayew, BO WANG, and Gary D. Bader. scMPT: towards applying large language models to complement single-cell foundation models, 2024. URL https://openreview.net/forum?id=nUpM7egYFd.

Edgar R, Domrachev M, and Lash AE. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res.*, 2002.

Syed Asad Rizvi, Daniel Levine, Aakash Patel, Shiyang Zhang, Eric Wang, Sizhuang He, David Zhang, Cerise Tang, Zhuoyang Lyu, Rayyan Darji, Chang Li, Emily Sun, David Jeong, Lawrence Zhao, Jennifer Kwan, David Braun, Brian Hafler, Jeffrey Ishizuka, Rahul M. Dhodapkar, Hattie Chung, Shekoofeh Azizi, Bryan Perozzi, and David van Dijk. Scaling large language models for next-generation single-cell analysis. *bioRxiv*, 2025.

Yanay Rosen, Yusuf Roohani, Ayush Agarwal, Leon Samotorčan, Tabula Sapiens Consortium, Stephen R. Quake, and Jure Leskovec. Universal cell embeddings: A foundation model for cell biology. *bioRxiv*, 2023. doi: 10.1101/2023.11.28.568918. URL https://www.biorxiv.org/content/early/2023/11/29/2023.11.28.568918.

Ron Sheinin, Roded Sharan, and Asaf Madi. scnet: learning context-specific gene and cell embeddings by integrating single-cell gene expression data with protein–protein interactions. *Nature Methods*, 2025.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, and Soboleva A. Ncbi geo: archive for functional genomics data sets–update. *Nucleic Acids Res.*, 2013.

Christina V. Theodoris, Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helene Mantineo, Elizabeth M. Brydon, Zexian Zeng, X. Shirley Liu, and Patrick T. Ellinor. Transfer learning enables predictions in network biology. *Nature*, 618, 2023.

Jian Wang, Jian Ma, Jian Chen, Ming Zhang, Yixuan Li, and Xiaoliang Xie. Large-scale foundation model on single-cell transcriptomics. *bioRxiv*, 2023.

Xinzhi Wang, Lijia He, Sarah M. Goggin, Assieh Saadatpour, Shoudan Li, and Winston Timp. Gears: Predicting transcriptional outcomes of gene perturbations in single cells with graph neural networks. *Genome Biology*, 22(1):1–23, 2021. doi: 10.1186/s13059-021-02489-5. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02489-5.

Fan Zhang, Tianyu Liu, Zhihong Zhu, Hao Wu, Haixin Wang, Donghao Zhou, Yefeng Zheng, Kun Wang, Xian Wu, and Pheng-Ann Heng. Cellverse: Do large language models really understand cell biology? In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2025. URL https://openreview.net/forum?id=FpxIViTl3C.

Xiaoman Zhang, Xiaoyuan Zhang, Zhuoqi Li, Yixuan Li, and Xiaoliang Xie. Pdgrapher: A graph neural network framework for predicting drug responses in single cells. *bioRxiv*, 2023. doi: 10.1101/2023.05.15.540855. URL https://www.biorxiv.org/content/10.1101/2023.05.15.540855v1.

Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, Tobias D. Wheeler, Geoff P. McDermott, Junjie Zhu, Mark T. Gregory, Joe Shuga, Luz Montesclaros, Jason G. Underwood, Donald A. Masquelier, Stefanie Y. Nishimura, Michael Schnall-Levin, Paul W. Wyatt, Christopher M. Hindson, Rajiv Bharadwaj, Alexander Wong, Kevin D. Ness, Lan W. Beppu, H. Joachim Deeg, Christopher McFarland, Keith R. Loeb, William J. Valente, Nolan G. Ericson, Emily A. Stevens, Jerald P. Radich, Tarjei S. Mikkelsen, Benjamin J. Hindson, and Jason H. Bielas. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, 8, 2017.

Kun Zhu, Yixuan Li, Zhuoqi Li, and Xiaoliang Xie. scmulan: A multitask generative pre-trained language model for single-cell omics. *bioRxiv*, 2024. doi: 10.1101/2024.01.25.577152. URL https://www.biorxiv.org/content/10.1101/2024.01.25.577152v1.

A APPENDIX

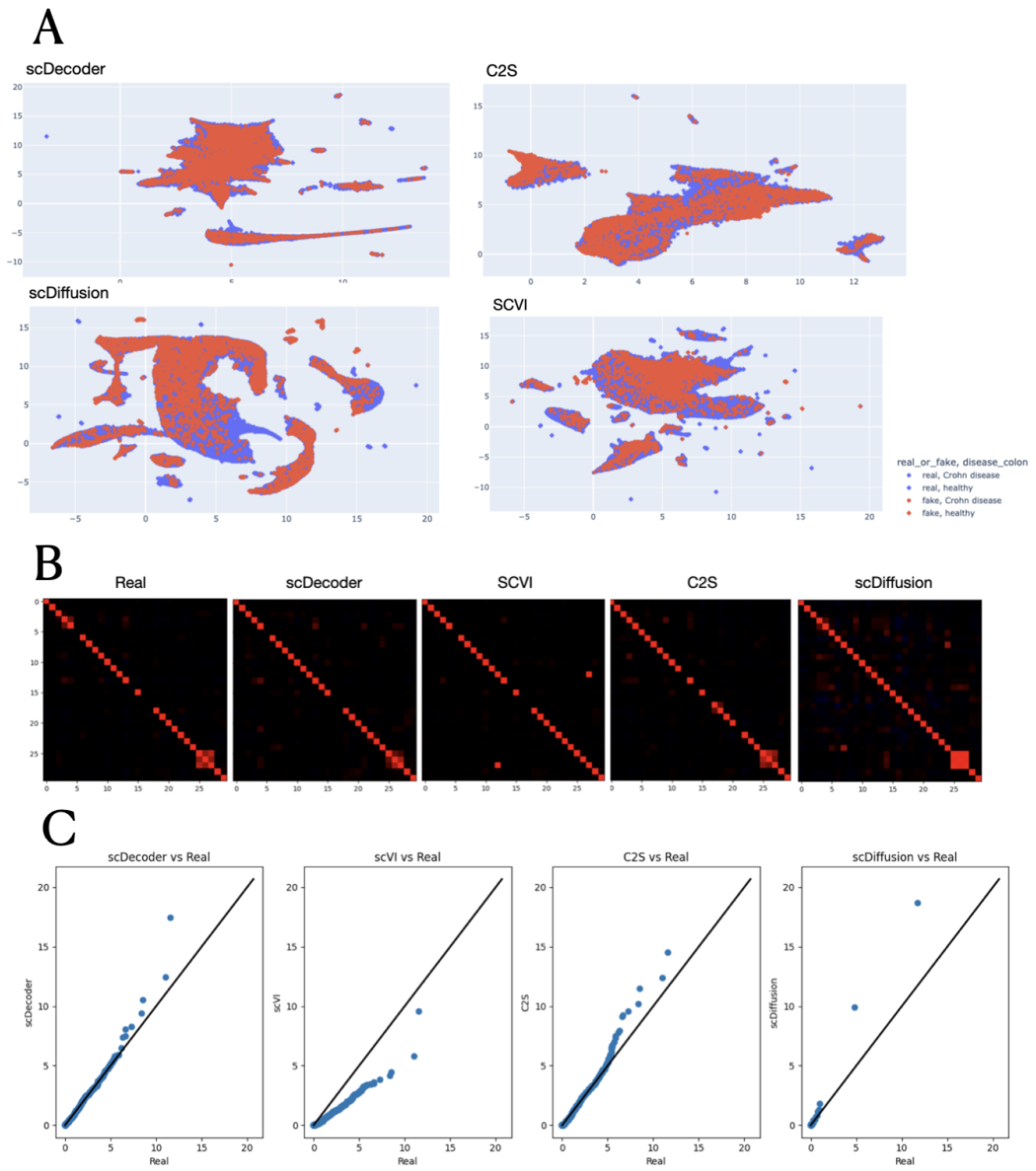

Figure 2: A: UMAPs of real and generated data of the four models. B: Correlation plots between the expression levels of 30 genes in real and generated data. C: QQ-plots of the mean gene counts of the generated data against the mean gene counts of the real data for all four models.

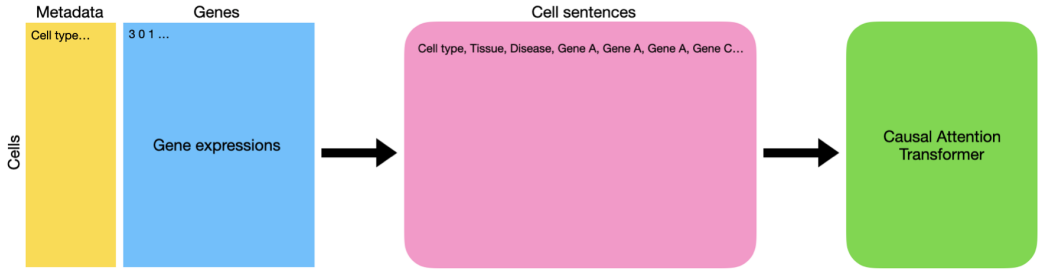

Figure 3: Model input.

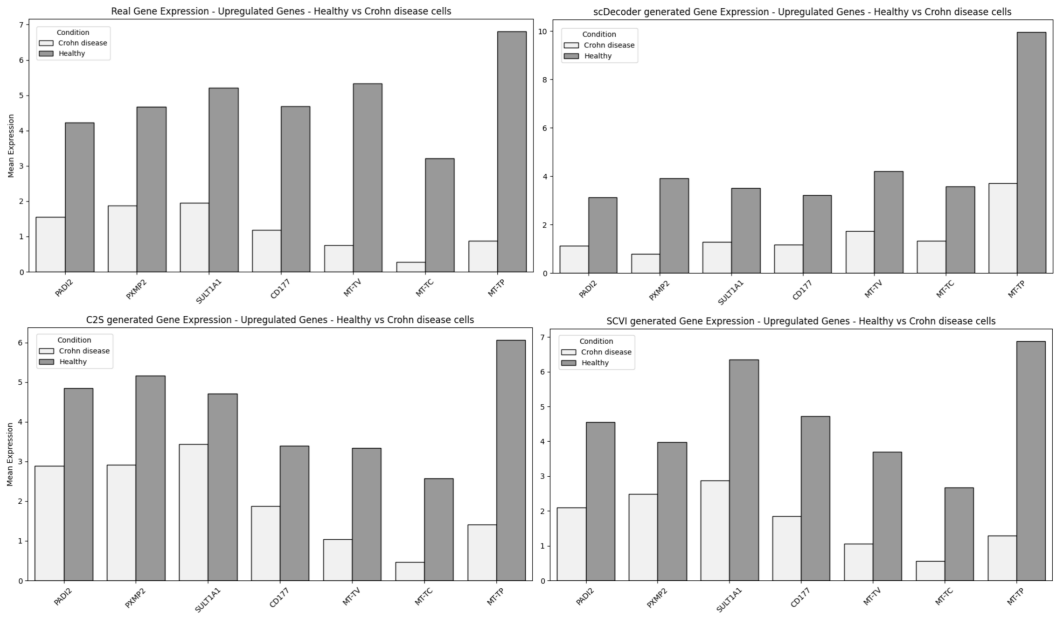

Figure 4: Mean expression of a selection of seven genes in both conditions in real and generated data.