

Bioimage informatics

SST-editing: *in silico* spatial transcriptomic editing at single-cell resolution

Jiqing Wu ^{1,*} and Viktor H. Koelzer ^{1,*}

¹Department of Pathology and Molecular Pathology, Computational and Translational Pathology Laboratory (CTP), University Hospital of Zurich, University of Zurich, Zurich, Switzerland

*Corresponding authors. Department of Pathology and Molecular Pathology, Computational and Translational Pathology Laboratory (CTP), University Hospital of Zurich, University of Zurich, Zurich, Switzerland. E-mails: Jiqing.Wu@usz.ch (J.W.) and Viktor.Koelzer@usz.ch (V.H.K.)

Associate Editor: Hanchuan Peng

Abstract

Motivation: Generative Adversarial Nets (GAN) achieve impressive performance for text-guided editing of natural images. However, a comparable utility of GAN remains understudied for spatial transcriptomics (ST) technologies with matched gene expression and biomedical image data.

Results: We propose *In Silico* Spatial Transcriptomic editing that enables gene expression-guided editing of immunofluorescence images. Using cell-level spatial transcriptomics data extracted from normal and tumor tissue slides, we train the approach under the framework of GAN (Inversion). To simulate cellular state transitions, we then feed edited gene expression levels to trained models. Compared to normal cellular images (ground truth), we successfully model the transition from tumor to normal tissue samples, as measured with quantifiable and interpretable cellular features.

Availability and implementation: <https://github.com/CTPLab/SST-editing>.

1 Introduction

Recently, Generative Adversarial Nets (GAN) (Goodfellow *et al.* 2014) and GAN Inversion (Xia *et al.* 2022) have achieved remarkable editing effects on generated (Kang *et al.* 2023) and real natural images (Patashnik *et al.* 2021). These methods mostly leverage encoded textual descriptions for desired image alterations. Typically, a pre-trained text encoder (Radford *et al.* 2021) is used to generate textual representations, which are then fed into the GAN (Inversion) model for text-guided editing of natural images.

In application to biomedicine, the utility of GAN holds great promise for *in silico* modeling of disease states. Taking bulk RNA expression as the input, a recent study (Carrillo-Perez *et al.* 2023) introduced RNA-GAN that showcased the generation of tissue image tiles. Further, GAN (Inversion)-enabled editing could aid the understanding of complex biological systems and enable low-cost and efficient cellular image manipulations. Trained with (immuno)fluorescence (IF) images that capture diverse cellular states, prior studies (Lamielle *et al.* 2023, Wu and Koelzer 2023) have demonstrated the editability of GAN and GAN Inversion on cellular images.

Without molecular data at fine granularity, the above GAN (Inversion)-based approaches are nonetheless unable to perform “text-guided” editing of biomedical images. Leveraging cutting-edge spatial transcriptomic (ST) (Moses and Pachter 2022) data, we propose *In Silico* Spatial Transcriptomic editing (SST-editing) using GAN (Inversion). We first train the approach with paired cell-level ST data, e.g.

mRNA transcripts as surrogates for the expression of specific genes, and IF biomarkers. By feeding edited gene expression data to the trained models, we shift tumor cell phenotypes toward the normal cell state and study the resulting editing effects on histology IF images using quantifiable and interpretable cellular features.

2 Materials and methods

2.1 Training data

Experiments are performed on the CosMx (He *et al.* 2022) hepatocellular carcinoma (HCC, liver tumor) dataset including both normal and tumor tissue slides. With a gigapixel spatial resolution, CosMx comprehensively documents a sparse 3D array of 1000-plex gene expression and associated multi-channel IF images (e.g. CD298/B2M, DAPI). In the appendix, we also report experiments on the Xenium (Janesick *et al.* 2022) lung tumor dataset.

2.1.1 Cellular expression

For the fine-grained analysis at single-cell resolution, the 1000-plex gene expression is represented as a 3D sparse array center-cropped on a detected cell. After summing the 3D array along spatial dimensions, we use the 1D table of gene expression data as the “text” input for training. To study editing effects, we algorithmically edit both the complete gene panel and the expression level of selected genes including *HLA-A* and *B2M*. This is motivated by the positive correlation between the expression level of these two genes and

CD298/B2M protein expression as measured by IF in the samples under study (please see also [Supplementary Fig. S2](#)).

2.1.2 Cellular image

Given the striking difference demonstrated by the expression level (pixel intensity) of CD298/B2M and cellular morphology (DAPI) between normal and tumor cells, we take the combination of these two markers as the “image” modality. Consequently, the two-channel image is a 3D array center-cropped on the same cell with paired gene expression input.

2.1.3 Cell subtype

Apart from utilizing the entire cell population, we run an ablative investigation on cell subtypes of interest. Based on the clustering annotation provided in the dataset, we select normal cells [Hep 1, 3, 4, 5, 6 in [Fig. 1i.1](#) (left)] and HCC cells [Tumor 1 in [Fig. 1i.1](#) (right)] to analyze the editing effects guided by expert pathologist review.

2.2 SST-editing

In reference to normal (or tumor) cellular images as the ground truth, we model the transition from tumor cellular images to normal ones (or vice versa) by SST-editing. To this end, we develop our approach upon the state-of-the-art StyleGAN2 ([Karras et al. 2020](#)) architecture.

2.2.1 Step 1 (GAN training)

Instead of using convoluted latent (textual) representations for natural image manipulation, we propose to directly utilize the gene expression data for cellular image manipulation. Concretely, this is done by feeding the gene expression table to a mapping network [[Fig. 1a](#) (left)] for controlling cellular features. Taking the discriminator D as the adversary, we train the generator G to output high-quality IF cellular images. To learn distinctive features of normal and tumor cells, both networks are trained with the adversarial loss \mathcal{L}_{adc} conditioned on binary cell labels [normal (0) and tumor (1)]. Together with the R_1 regulation \mathcal{L}_{R_1} and path length regulation loss $\mathcal{L}_{\text{path}}$ ([Karras et al. 2020](#)), we have the objective:

$$\min_G ((\max_D \mathcal{L}_{\text{adc}}) + \lambda_{R_1} \mathcal{L}_{R_1} + \lambda_{\text{path}} \mathcal{L}_{\text{path}}), \quad (1)$$

where λ_{R_1} and λ_{path} are hyperparameters. To manipulate **generated (Gen)** cellular images, we then feed edited gene expression data to the GAN model.

2.2.2 Step 2 (GAN Inversion training)

Going beyond generated cellular images, we extend the analysis of editing effects to real (reconstructed) cellular images. This is carried out by GAN Inversion (auto-encoder), in which we reuse the trained generator G as the decoder and aim to faithfully reconstruct the input cellular image. To strike a balance between editability and reconstruction quality, we address this dual challenge by using a pre-trained lightweight CLIP ([Shariatnia 2021](#)) encoder E_c and a “pixel2style2pixel” ([Richardson et al. 2021](#)) encoder E_p [[Fig. 1a](#) (right)], respectively. Then, we apply the contrastive $\mathcal{L}_{\text{moco}}$, l_2 reconstruction \mathcal{L}_2 and perceptually learned \mathcal{L}_{PIP} loss suggested in ([Alaluf et al. 2021](#)), while the ID similarity loss used in the same study tailored for facial images is excluded. As a whole, we have the objective:

$$\min_{E_c, E_p} (\lambda_1 \mathcal{L}_{\text{moco}} + \lambda_2 \mathcal{L}_2 + \lambda_3 \mathcal{L}_{\text{PIP}}), \quad (2)$$

where $\lambda_{1,2,3}$ are the hyperparameters. After training with paired gene expression data and real cellular images, we feed the edited gene expression data to manipulate the **reconstructed (Rec)** cellular images.

GAN and GAN Inversion training were run for 800 k iterations with a batch size of 16 for GAN and 8 for GAN Inversion, respectively.

2.2.3 Step 3 (Gene expression-guided editing)

Due to the lack of one-on-one correspondence between individual cells on the normal and tumor slides, clear guidance for individually editing the gene expression of each cell is not available in the experimental setup. Therefore, we propose to collectively edit the gene expression of paired cells by matching the gene data distribution [i.e. sample covariance matrix (SCM) ([Wu and Koelzer 2022](#))] of one cell population to another. We achieved this by scaling the eigenvalues and rotating the eigenbasis of SCM. For $i = 0, 1$, consider the SCM $\frac{1}{n_i} \mathbf{G}_i^T \mathbf{G}_i = \mathbf{O}_i \boldsymbol{\lambda}_i \mathbf{O}_i^T$, where \mathbf{G}_i is the collection of n_i p -plex gene expression from the normal or tumor slide, \mathbf{O}_i is the $p \times p$ eigenbasis and $\boldsymbol{\lambda}_i$ is the $p \times p$ (sorted) diagonal eigenvalues obtained by eigenvalue decomposition. For the collection of \mathbf{G}_i , we apply the linear transformation $\mathbf{G}'_i = \mathbf{G}_i \mathbf{O}_i (\sqrt{\boldsymbol{\lambda}_{(i+1)\%2}} / \sqrt{\boldsymbol{\lambda}_i}) \mathbf{O}_{(i+1)\%2}^T$ such that for the edited gene collection \mathbf{G}'_i it holds $\frac{1}{n_i} \mathbf{G}'_i^T \mathbf{G}'_i = \frac{1}{n_{(i+1)\%2}} \mathbf{G}_{(i+1)\%2}^T \mathbf{G}_{(i+1)\%2}$. Due to the computational fluctuation of smaller eigenvalues and the dominant effect of the leading eigenvalue, it suffices to scale the largest eigenvalue in our experiments. By keeping the selected group of gene expression values unchanged during the linear transformation, we narrow down the editing process to the gene of interest.

For all implementation details including, but not limited to, the above key algorithmic steps, we refer interested readers to the curated GitHub repository <https://github.com/CTPLab/SST-editing>.

3 Results

Overall, we achieve three levels of experimental settings that correspond to eight conditions: Analysis of (i) generated (GAN) and reconstructed (GAN Inversion) cellular images; (ii) (selected) gene expression values before and after in silico editing; (iii) the total cell population and clinically relevant cell subtypes (“Hepatocytes”).

3.1 GAN (Inversion) evaluation

We determine the optimal GAN model using Fréchet Inception Distance (d_{FID}) ([Heusel et al. 2017](#)), which effectively measures the statistical distance between the feature distribution of cellular image collections. Further, we report Peak Signal-to-Noise Ratio and Structural Similarity Index Measure to benchmark the GAN Inversion model. As such, the best models are determined at 550 k (GAN, [Supplementary Fig. S4](#)) and 700 k (GAN Inversion, [Supplementary Fig. S5](#)) iterations, see [Fig. 1i.0, i.2](#) for more image visualizations.

3.2 Editing effect evaluation

3.2.1 Cellular quantification

Before editing: blue plots of [Fig. 1q.0, q.1](#) illustrate two groups of genes with the highest expression difference between normal hepatocytes and HCC, e.g. genes with low

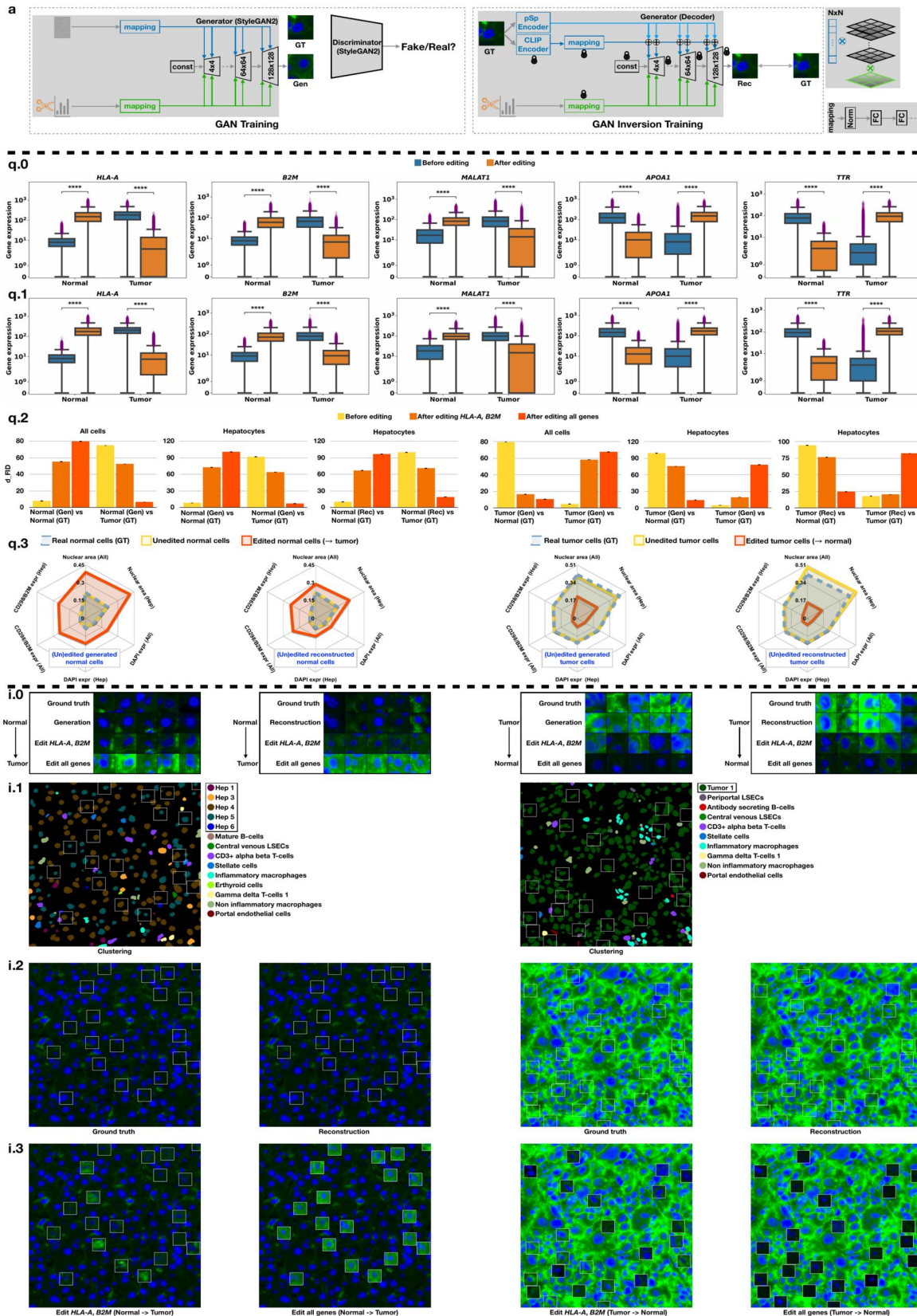


Figure 1. (a) Model illustrations. (q) Numerical quantification of CosMx results. (q.0 and q.1) The comparison of gene expression shifts on all the cells (q.0) and the “Hepatocyte” (Hep) subtypes (q.1) of the normal and tumor slide. Here, **** means $P \leq .0001$. (q.2) For the generated (Gen, GAN) and reconstructed (Rec, GAN Inversion) cells, the d_{FID} comparison of cellular state transitions *w.r.t.* all cells and Hep subtypes. We randomly repeat the d_{FID} computation four times and report the mean and standard deviation. (q.3) The editing effect comparison of interpretable cellular features for all the cells and hepatocytes (Hep). (i) Visual interpretation of CosMx results. (i.0) The image gallery of cellular state transitions for Gen and Rec cells. Here, we present the transition that occurred in the DAPI (blue) and CD298/B2M (green) channels. (i.1) The visualization of cell subtypes on a region of interest extracted from the normal and tumor slide. (i.2) The randomly sampled ground truth and reconstructed cellular images within the bounding boxes. (i.3) The morphological transitions of these cellular images guided by edited gene expression levels

expression in normal and upregulation in HCC (*HLA-A*, *B2M*, *MALAT1*) and genes downregulated in HCC compared to normal (*APOA1*, *TTR*). After editing: by shifting the gene expression distribution from one cellular state to another, we witness a meaningful change in the expression of edited genes when comparing paired populations. The quantitative comparison presents a highly significant shift in normal (or tumor) gene expression patterns toward the tumor (or normal) spectrum for the total cell population (Fig. 1q.0) and hepatocyte subtypes (Fig. 1q.1, ~80% of the total cells). After feeding the edited gene expression profiles to the GAN (Inversion) model, we measure the manipulation effects on tumor (or normal) cellular images in comparison to ground truth normal (or tumor) cells, respectively. To support the analysis of generated (Gen) and reconstructed (Rec) cellular images with interpretable morphometric features, we perform the quantification using d_{FID} and cellular features such as nuclear area and CD298/B2M expression level.

3.2.2 Cellular interpretation

When decreasing the leading and highly overexpressed genes *HLA-A* in tumor cells and *B2M* for the CosMx HCC dataset, we model the quantifiable transition of tumor cells (large nuclei, cellular atypia, variation in nuclear size) toward normal cells. Even more striking morphological effects become evident when editing the expression of all genes, in terms of clearly decreasing d_{FID} scores (Fig. 1q.2), decreasing average *B2M* marker expression and reduced average nuclear area (Fig. 1q.3) as derived from the CellPose (Stringer *et al.* 2021) method. Complementary to these quantification results, normal liver cells acquire a remarkably malignant appearance (Fig. 1i.0, i.3) when driven by edited genes of interest (increase in nuclear area and overexpression of *B2M*, Fig. 1q.3) toward the tumor spectrum. Importantly, the *in silico* editing of *HLA-A* and *B2M* indeed correlates to the emergence or disappearance of CD298/B2M protein expression as captured by the IF imaging (green channel), supporting the reliability of biological interpretations by the proposed approach.

4 Conclusions

In the biomedical context, this study sheds light on the algorithmic editability of ST data. By editing tumor gene expression profiles toward the normal spectrum, we achieved the reversal of tumor cellular images to normal ones, as measured with quantifiable and interpretable cellular features. The SST-editing approach, which exhibits low ethical, legal and regulatory risks in the simulated intervention of human biological material, thus provides a new perspective to model pathological processes in real-life clinical tissue samples.

Acknowledgements

This study is supported by core funding of the University of Zurich to the Computational and Translational Pathology Lab led by V.H.K. at the Department of Pathology and Molecular Pathology, University Hospital and University of Zurich.

Author contributions

Jiqing Wu and Viktor H. Koelzer conceived the research idea. Jiqing Wu implemented the algorithm and carried out the

experiments. Jiqing Wu and Viktor H. Koelzer analyzed the results. Jiqing Wu and Viktor H. Koelzer drafted the manuscript. Viktor H. Koelzer supervised the project.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

J.W. declares no competing interests. V.H.K. reports being an invited speaker for Sharing Progress in Cancer Care (SPCC) and Indica Labs; advisory board of Takeda; sponsored research agreements with Roche and IAG, all unrelated to the current study. V.H.K. is a participant in several patent applications on the assessment of cancer immunotherapy biomarkers by digital pathology, a patent application on multimodal deep learning for the prediction of recurrence risk in cancer patients, and a patent application on predicting the efficacy of cancer treatment using deep learning.

Funding

None declared.

Data availability

The CosMx platform provides two comprehensive spatial arrays of 1000-plex gene expression [~0.2 billion gene expression counts for the normal liver slide and 0.5 billion for the hepatocellular carcinoma (HCC) slide], which correspond to 340 and 464 k cells, respectively (<https://nanosttring.com/wp-content/uploads/2023/01/LiverPublicDataRelease.html>). Similarly, the Xenium dataset offers two large-scale spatial expression maps of 392-plex pre-designed and custom target genes [~24 million and 67 million total counts for the healthy lung slide and invasive adenocarcinoma (IAC) slide individually], along with 300 k (https://cf.10xgenomics.com/samples/xenium/1.3.0/Xenium_Preview_Human_Non_diseased_Lung_With_Add_on_FFPE/Xenium_Preview_Human_Non_diseased_Lung_With_Add_on_FFPE_analysis_summary.html) and 530 k (https://cf.10xgenomics.com/samples/xenium/1.3.0/Xenium_Preview_Human_Lung_Cancer_With_Add_on_2_FFPE/Xenium_Preview_Human_Lung_Cancer_With_Add_on_2_FFPE_analysis_summary.html) cells that are detected from the two slides. CosMx: The multiplexed fluorescence imaging (MFI) data of healthy liver tissue are available via <https://smi-public.objects.liquidweb.services/NormalLiverFiles.zip>, the MFI data of cancer liver tissue are available via <https://smi-public.objects.liquidweb.services/LiverCancerFiles.zip>, the ST data are available via <https://smi-public.objects.liquidweb.services/LiverDataReleaseTileDB.zip>. Xenium: The ST data and whole slide image (WSI) of healthy lung tissue are available via https://cf.10xgenomics.com/samples/xenium/1.3.0/Xenium_Preview_Human_Non_diseased_Lung_With_Add_on_FFPE/Xenium_Preview_Human_Non_diseased_Lung_With_Add_on_FFPE_outs.zip, the ST data and WSI of cancer lung tissue are available via https://s3-us-west-2.amazonaws.com/10x.files/samples/xenium/1.3.0/Xenium_Preview_Human_Lung_Cancer_With_Add_on_2_FFPE/Xenium_Preview_Human_Lung_Cancer_With_Add_on_2_FFPE_outs.zip. All the above links are free and publicly available to users, without further login or registration requirements.

References

- Alaluf Y, Patashnik O, Cohen-Or D. Restyle: A residual-based StyleGAN encoder via iterative refinement. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, virtual, 2021, 6711–20.
- Carrillo-Perez F, Pizurica M, Ozawa MG *et al.* Synthetic whole-slide image tile generation with gene expression profile-infused deep generative models. *Cell Rep Methods* 2023;3:100534.
- Goodfellow I, Pouget-Abadie J, Mirza M *et al.* Generative adversarial nets. In: *Advances in Neural Information Processing Systems 27*, Montreal, Quebec, Canada, 2014, 2672–80.
- He S, Bhatt R, Brown C *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol* 2022;40:1794–806.
- Heusel M, Ramsauer H, Unterthiner T *et al.* GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: *Advances in Neural Information Processing Systems 30*, Long beach, CA, USA, 2017.
- Janesick A, Shelansky R, Gottscho A *et al.* High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat Commun* 2023;14:8353. <https://doi.org/10.1038/s41467-023-43458-x>.
- Kang M, Zhu J-Y, Zhang R *et al.* Scaling up GANs for text-to-image synthesis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023, 10124–34.
- Karras T, Laine S, Aittala M *et al.* Analyzing and improving the image quality of StyleGAN. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, virtual, 2020, 8110–9.
- Lamiable A, Champetier T, Leonardi F *et al.* Revealing invisible cell phenotypes with conditional generative modeling. *Nat Commun* 2023;14:6386.
- Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods* 2022;19:534–46.
- Patashnik O, Wu Z, Shechtman E *et al.* Styleclip: text-driven manipulation of StyleGAN imagery. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, virtual, 2021, 2085–94.
- Radford A, Kim JW, Hallacy C *et al.* Learning transferable visual models from natural language supervision. In: *International Conference on Machine Learning*. PMLR, virtual, 2021, 8748–63.
- Richardson E, Alaluf Y, Patashnik O *et al.* Encoding in style: a StyleGAN encoder for image-to-image translation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, virtual, 2021, 2287–96.
- Shariatnia MM. Simple clip. 06, 2022. <https://doi.org/10.5281/zenodo.6845731>.
- Stringer C, Wang T, Michaelos M *et al.* Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* 2021;18:100–6.
- Wu J, Koelzer V. Sorted eigenvalue comparison d_{Eig} : a simple alternative to d_{FID} . In: *NeurIPS 2022 Workshop on Distribution Shifts: Connecting Methods and Applications*, New Orleans, LA, USA, 2022.
- Wu J, Koelzer VH. GILEA: GAN inversion-enabled latent eigenvalue analysis for phenome profiling and editing. bioRxiv, 2023, preprint: not peer reviewed. <https://doi.org/10.1101/2023.02.10.528026>.
- Xia W, Zhang Y, Yang Y *et al.* GAN inversion: a survey. *IEEE Trans Pattern Anal Mach Intell* 2023;45:3121–38.