SPATIAL TRANSCRIPTOMICS-INFORMED HISTOPATHOLOGIC IMAGE GENERATION FOR MODELING THE SKIN TISSUE MICROENVIRONMENT

Vedhsai Thiriveedi
Nashua High School South
Nashua, NH 03062, USA
Email: vedhsait@gmail.com

Arnav Chaphalkar
Adlai E. Stevenson High School
Palatine, IL 60069, USA
Email: arnav.chaphalkar@gmail.com

Gokul Srinivasan
University of California, Los Angeles
Los Angeles, CA 90095, USA
Email: gokulsrin@gmail.com

Carly Miles
University of Michigan
Ann Arbor, MI 48109, USA
Email: carmiles@umich.edu

Zarif Azher
California Institute of Technology
Pasadena, CA 91125, USA
Email: zazher@caltech.edu

Anvith Kakkera, Fred Kolling IV, Laurent Perreard, Scott Palisoul, Louis J. Vaickus, Matthew Davis, Matthew LeBoeuf
Dartmouth College Geisel School of Medicine
Hanover, NH 03755, USA
Email: kakkeraanvith@gmail.com, fred.w.kolling.iv@dartmouth.edu, laurent.perreard@dartmouth.edu, scott.m.palisoul@hitchcock.org, louis.j.vaickus@hitchcock.org, matthew.j.davis@hitchcock.org, matthew.leboeuf@hitchcock.org

I-Chuang Liao, Minh-Khang Le, Keluo Yao, Joshua Levy[1]
Cedars Sinai Medical Center
Los Angeles, CA 90048, USA
Email: ichuangliao@gmail.com, *khang.le@cshs.org, keluo.yao@cshs.org, joshua.levy@csmc.edu*

Spatial transcriptomics (ST) has transformed the study of tissue biology by coupling genome-wide gene expression with spatial histologic context, but its high cost limits application to large patient cohorts. While considerable effort has focused on predicting molecular measurements from routine histology, comparatively little attention has been devoted to the reverse problem: generating biologically informed histopathologic images directly from spatial transcriptomic measurements. Such models could enable simulation of tissue microenvironments, facilitate mechanistic studies of tissue organization, and augment computational pathology workflows in data-limited settings. Here,

[1] To whom correspondence should be addressed.

© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

we present an initial evaluation of ST-informed histopathologic image generation in human skin using a conditional latent diffusion model conditioned on spatial gene expression, tissue architecture and cell type proportions. Across 39,579 Visium ST spots from sixteen patients, annotated for seven tissue architectures, generated images produced varied, realistic histologic representations that preserved key characteristics of the underlying tissue microenvironment. Quantitative analyses using CellViT demonstrated significant agreement in cellular abundance between real and generated tissues for epithelial, stromal, and immune cellular composition. We next evaluated whether these generated tissues improve downstream computational pathology tasks– i) ST inference, ii) tissue architecture classification. Although generative augmentation did not produce practically meaningful improvements in overall spatial gene expression prediction (ΔPCC = +0.0103), inference improved most in adnexal (ΔSpearman = +0.0175, p = 0.002) and dermal (ΔSpearman = +0.0110, p = 0.010) regions, while architecture classification (Δ macro-F1 = +0.0141) improved most for inflammatory (ΔF1 = +0.0376) and stromal (ΔF1 = +0.0360) tissue. These results suggest that generative modeling preferentially benefits specific tissue microenvironments rather than uniformly improving all prediction tasks, providing an initial benchmark for future ST-informed histopathologic image generation work. Future improvements in image realism and representation learning may further expand the utility of ST-informed generative pathology.

Keywords: Spatial transcriptomics; histopathologic image generation; conditional latent diffusion; tissue microenvironment; computational pathology; generative augmentation.

1. Introduction

The characterization of tissue architecture has long been central to understanding disease pathogenesis [1]. For over a century, routine hematoxylin and eosin (H&E) staining has enabled pathologists to assess tissue organization, cellular morphology, and disease processes directly from histologic sections. More recently, spatial transcriptomics (ST) has transformed tissue analysis by coupling genome-wide gene expression with preserved spatial context, allowing molecular programs to be studied within their native tissue microenvironment. On platforms such as 10x Genomics Visium, thousands of genes are simultaneously measured at regularly spaced locations across an H&E-stained tissue section, providing an unprecedented opportunity to study the relationships between molecular state and tissue morphology [2].

Despite its transformative potential, widespread adoption of ST remains limited by cost, assay complexity, and throughput. Consequently, considerable effort has focused on predicting molecular measurements directly from routine histology. Recent advances in computational pathology, particularly pathology foundation models pretrained on millions of histologic image patches [3–9], have substantially improved the prediction of spatial gene expression and tissue architecture from H&E images. These models leverage transferable visual representations that can be adapted to downstream prediction tasks while requiring only relatively small, labeled ST datasets.

Nevertheless, accurate inference across diverse downstream tasks ultimately depends on training datasets that adequately capture the breadth of histologic and microenvironmental variation present within a study cohort. Because spatial transcriptomic profiling remains costly and labor intensive, many datasets incompletely sample rare tissue architectures, cellular neighborhoods, specialized microenvironmental states, and underrepresented patient populations. Consequently, downstream

models may become biased toward well-represented tissue contexts while underperforming in biologically varied regions or across underrepresented patient subgroups.

While predicting molecular state from morphology has become an active area of research, the converse problem– generating histologic appearance directly from spatial molecular measurements– has received comparatively little attention [4,10–18]. Yet such models offer unique opportunities to investigate how gene expression gives rise to tissue architecture, simulate biologically plausible tissue microenvironments, evaluate the fidelity of learned molecular-to-morphologic relationships, and potentially augment limited ST datasets for downstream computational pathology tasks.

A framework capable of generating biologically faithful histopathologic images from spatial molecular measurements offers one potential solution to these challenges. Rather than merely increasing the number of training examples, generated tissues could expand the diversity of biologically plausible microenvironmental configurations available during model development, improving representation of tissue contexts that are difficult or impractical to collect experimentally [19]. Recent advances in conditional diffusion modeling provide a promising framework for generating histopathologic images directly from spatial molecular measurements [4,10–13]. Such models have demonstrated remarkable success in natural and biomedical imaging by synthesizing realistic images while preserving structured conditioning information. Whether these approaches can faithfully reconstruct biologically meaningful tissue architecture and microenvironmental organization from spatial transcriptomic measurements, however, remains largely unexplored. Furthermore, it remains unclear whether generated tissues generated in this manner can improve downstream computational pathology tasks or simply produce visually convincing images that fail to transfer biologically relevant information.

Here, we evaluate spatial transcriptomics-informed histopathologic image generation using a conditional latent diffusion model trained on paired H&E and Visium data from human skin. We assess whether generated histology faithfully recapitulates tissue architecture and cellular composition, determine whether generated images improve downstream spatial gene expression inference and tissue architecture prediction, and characterize the biological contexts in which generative modeling is most effective. Together, this work provides an initial framework for evaluating histopathologic image generation from spatial transcriptomics.

2. Methods

2.1. *Study Cohort, Spatial Transcriptomic Profiling, and Tissue Annotation*

2.1.1. *Patient Cohort and Spatial Transcriptomic Profiling*

Sixteen normal human skin specimens were collected from the Department of Dermatology at Dartmouth Hitchcock Medical Center as bisected Burow's triangles routinely discarded during Mohs micrographic surgery [20]. The specimens were derived from histologically normal skin adjacent to the surgical repair site and would otherwise have been discarded. This study was reviewed by the Dartmouth Health Human Research Protection Program (STUDY02001286) and determined to be exempt under Category 4. Participants ranged in age from 19 to 85 years (mean ± SD, 60.4 ± 18.5 years) and included 8 females and 8 males. Specimens were obtained from chronically sun-exposed anatomic sites, including the cheek (n = 7), forehead (n = 3), nasal dorsum (n = 3), pre-auricular region (n = 2), and postauricular region (n = 1). Histopathologic assessment demonstrated a range of

chronic photodamage (4 mild, 2 moderate, and 10 severe solar elastosis found by pathologic review). Tissues were formalin-fixed and paraffin-embedded (FFPE), sectioned, transferred to Visium slides using the 10x Genomics CytAssist workflow, and stained with hematoxylin and eosin (H&E) using an automated tissue stainer. Whole-slide images were acquired at 40x magnification using a Leica GT450 whole-slide scanner. Spatial transcriptomic profiling was subsequently performed using the 10x Genomics Visium platform, which captures transcriptomes from 55-µm-diameter spatial capture spots distributed across the tissue section.

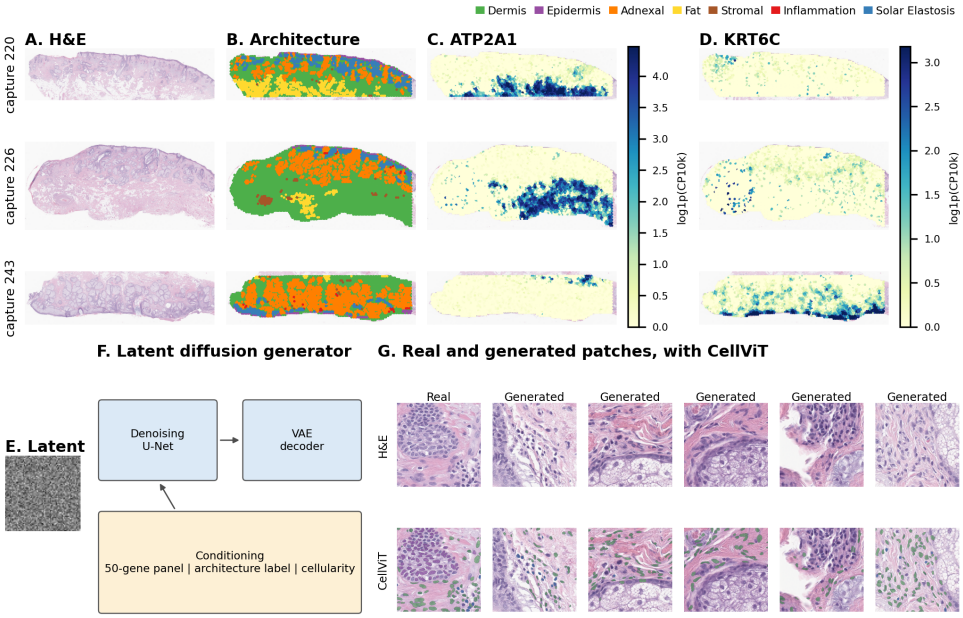


Figure 1. Study cohort and overview of the spatial transcriptomics-conditioned histopathology generation framework. (A) Specific H&E WSI examples from the human skin cohort. **(B)** Expert tissue architecture annotations identifying the seven histologic compartments used throughout the study. **(C–D)** Representative Visium ST expression maps for ATP2A1 and KRT6C. **(E)** Histopathology patches extracted from individual Visium spots serve as inputs for training. **(F)** Overview of the latent diffusion framework, in which image generation is conditioned on a 50-gene expression panel, tissue architecture label, and cellularity before decoding into H&E patches. **(G)** Representative real and independently generated H&E patches from the same biological conditioning profile with corresponding CellViT-derived cell segmentations.

2.1.2. *Tissue Architecture Annotation*

Whole-slide H&E images were manually annotated by expert dermatopathologists to delineate major histologic compartments represented within the skin, including the dermis, epidermis, adnexal structures (e.g., hair follicles, sebaceous and eccrine glands), adipose tissue, stroma, inflammation, and solar elastosis. Tissue architecture labels were assigned to Visium spots based on their spatial overlap with the annotated regions.

2.1.3. *Data Preprocessing and Dataset Partitioning*

Following quality control and removal of non-tissue regions, tissue-overlapping Visium spots were paired with corresponding H&E image patches centered on each capture location. Gene expression profiles were log1p-normalized and restricted to the 2,000 most highly variable genes (HVGs) for downstream prediction tasks. To prevent information leakage, Visium captures were partitioned at the slide level such that each capture was assigned exclusively to the training, validation, or test cohort. This resulted in 25,438 tissue-overlapping spots from 10 slides for training, 4,029 spots from

3 slides for validation, and 10,112 spots from 3 slides for testing, with five-fold sample-level cross-validation as appropriate.

2.2. *Spatial Transcriptomics-Informed Histopathologic Generation*

2.2.1. *Model Architecture*

The image generator leverages a previously developed conditional latent diffusion framework to synthesize a histopathology image patch for each Visium spot from its associated biological measurements [21]. Images are represented in a compact latent space using a pretrained pathology variational autoencoder (VAE), which remains fixed throughout training. A latent-space U-Net iteratively predicts denoising updates conditioned on spot-level expression profiles and additional conditioning inputs (**2.2.2**) through adaptive layer normalization (AdaLN-Zero) and cross-attention blocks, allowing conditioning information to influence every stage of the generative process.

2.2.2. *Conditioning Inputs*

Each Visium spot is conditioned using three complementary sources of biological information: (i) an expression panel of the 50 most highly variable genes (HVGs), selected from manual experimentation evaluated using the approach from **2.3.2**, (ii) a tissue architecture label corresponding to one of seven histologic classes (dermis, epidermis, adnexal structures, adipose tissue, stroma, inflammation, and solar elastosis), and (iii) automated measurements of the local cellular composition– the latter of which was derived from automated nucleus segmentation and classification using CellViT [22], as described in Section **2.3**. Cell type information was incorporated either as: (i) cell-type-specific abundance/proportion (epithelial, stromal, immune, and other nuclei), (ii) a single total cell abundance measurement, or (iii) omitted entirely. These conditioning strategies were compared to evaluate the contribution of increasingly detailed tissue microenvironment information to image generation.

2.2.3. *Model Training and Sampling*

The conditional generator was optimized using a rectified-flow objective that directly models the transformation from noise to the target latent representation [23]. During inference, latent representations were iteratively sampled according to the learned flow and subsequently decoded into multiple generated H&E image patches using the fixed pretrained VAE decoder. The generation hyperparameters, guidance scale and interval [24], were selected from a sweep over scales 3.0, 4.0, 5.0, and 6.0 and intervals full 0.0 to 1.0 versus mid-noise 0.1 to 0.7. The configurations were evaluated on 960 patches using the cell-type correlation strategy introduced in Section **2.3.2** to determine the optimal guidance scale of 3.0 and guidance interval of 0.1-0.7.

2.3. *Tissue Microenvironment Quantification for Assessing Quality of Generated Images*

Because the generator is conditioned on spot-level molecular and tissue microenvironment representations rather than explicit pixel-level tissue morphology, multiple histologically distinct H&E patches may faithfully represent the same underlying biological state. For example, identical gene expression profiles and cellular composition may produce different, plausible local arrangements of epithelial, stromal, inflammatory, and adnexal structures. Thus, differences in spot-

level cellular composition, instead of direct pixel-wise or perceptual measures, were used to compare real and generated images for their faithfulness to the local microenvironment.

2.3.1. *Cell Segmentation, Classification, and Spot-Level Tissue Microenvironment Features*

Automated nucleus segmentation and classification were performed on both real and generated H&E image patches using CellViT. Individual nuclei were detected within each spot-centered image patch and classified as epithelial, stromal, immune, or other cell types. Cell-level predictions were aggregated to generate spot-level measurements of tissue microenvironment composition. Summary metrics included total cellular abundance together with the fractions and count of epithelial, stromal, and immune cells within each Visium spot.

2.3.2. *Tissue Microenvironment Concordance*

Image generation from identical conditioning inputs was expected to produce similar TME features. Spearman rank correlation was used to quantify agreement in CellViT-derived epithelial, stromal, and immune cell composition from real and generated images across Visium spots. Concordance analyses were additionally stratified by tissue architecture to determine whether performance differed across histologic contexts. The stability of these associations were also evaluated across repeated generations from identical conditioning inputs.

2.3.3. *Conditional Perturbation Experiments*

Image generation from altered conditioning inputs was anticipated to result in the expected changes to TME features [25]. Individual conditioning inputs, including tissue architecture labels, cell-type composition, and gene-expression programs associated with inflammatory, stromal, epidermal, and adnexal tissue states, were selectively perturbed while holding the Visium spot location fixed. Generated patches were subsequently assessed for differences in CellViT-derived cell proportions to evaluate whether the resulting morphology shifted in the expected direction. The findings were further supported by calculating the cosine distance between foundation model (Phikon; **2.4.1**) derived image features between real images and patches generated with identical and perturbed conditioning inputs.

2.4. *Assessing the Impact of ST-Informed Generation on Downstream Prediction Tasks*

2.4.1. *Downstream Tasks and Prediction Models*

To evaluate the utility of ST-informed generative augmentation, we considered two spot-level prediction tasks: (i) prediction of the expression of 2,000 highly variable genes and (ii) classification of each Visium spot into one of seven tissue architecture classes. Image representations were extracted using either a frozen pathology foundation model (Phikon), which produced a 768-dimensional feature vector for each real or generated H&E patch [6,9], or a ResNet convolutional neural network trained from scratch on the study cohort. Image features were passed to a multilayer perceptron consisting of a 512-unit hidden layer (dropout = 0.3) followed by a task-specific output layer. For gene expression prediction, the output layer predicted the expression of 2,000 highly variable genes trained with the mean squared error loss on pseudo-log transformed expression [26]. For tissue architecture classification, a separate output layer predicted one of seven tissue architecture classes trained with cross-entropy and focal losses. Models were optimized using AdamW (learning

rate = 0.001, batch size = 128) for up to 50 epochs with early stopping based on validation performance and coarse hyperparameter scans.

2.4.2. *ST-Informed Generative Augmentation*

Standard image augmentation was replaced or supplemented with ST-informed generative augmentation, in which multiple generated H&E patches were generated for each training Visium spot while preserving its underlying conditioning. We compared augmentation strategies that varied (i) the number of generated patches generated per spot, (ii) the generated-to-real mixing ratio, (iii) the incorporation of conventional image augmentation, (iv) filtering of generated patches based on tissue microenvironment concordance (**2.3**), (v) targeted augmentation of underrepresented tissue architectures or difficult-to-predict genes, and (vi) model pretraining on generated images prior to fine-tuning on both real and generated images. For each downstream prediction task, only the best-performing ST-informed *generative* augmentation strategy, identified through validation set evaluation (**2.4.3-4**), was carried forward to the primary and stratified analyses presented in the *Results*. Based on these comparisons, the optimal augmentation strategy for each downstream task generated four stochastic generated patches for each real training spot, yielding 101,752 generated image patches for downstream model training.

2.4.3. *Evaluation of Spatial Gene Expression Prediction*

Prediction performance was evaluated using the Pearson correlation coefficient (PCC) and Spearman correlations between predicted and observed expression for each of the 2,000 highly variable genes (HVGs), which were selected by ranking genes according to Moran's I [27]and averaging ranks across tissue sections. Performance was summarized across all HVGs and within a difficult-to-predict gene set selected by baseline prediction performance (ie., 500 genes with lowest PCC; no generative augmentation). Results were reported across all Visium spots and after stratification by tissue architecture. Changes in PCC and Spearman are denoted ΔPCC/Spearman. Performance across approaches was compared using median ΔSpearman across genes and Wilcoxon signed-rank tests on per-gene associations.

2.4.4. *Evaluation of Tissue Architecture Classification*

Tissue architecture classification performance was evaluated by predicting one of seven tissue architecture classes for each Visium spot. Performance was quantified using the macro-averaged F1 score and per-class F1 scores, enabling assessment of architecture-specific performance. Spot-level classification performance was compared between models trained using conventional image augmentation and ST-informed generative augmentation.

2.4.5. *Comparison by Model Pretraining*

We hypothesized that ST-informed generative augmentation would preferentially benefit models trained from scratch by increasing the variety of biologically plausible training examples available for representation learning. In contrast, pathology foundation models have already learned robust visual representations through large-scale pretraining and were therefore expected to derive comparatively smaller gains from augmentation. Improvements in spatial gene expression prediction and tissue architecture classification were evaluated separately for each prediction model.

2.5. *Biological Interpretation of Downstream Prediction Effects*

To identify where ST-informed generative augmentation provided the greatest benefit, augmentation-related changes in downstream prediction performance were examined at both the tissue architecture and gene levels. Architecture-specific analyses quantified the effect of ST-informed generative augmentation on both spatial gene expression prediction and tissue architecture classification within each of the seven tissue architectures. Gene-level analyses ranked genes according to the improvement in prediction performance achieved with ST-informed generative augmentation for the HVG and difficult-to-predict gene sets. Finally, the genes with greatest improvements were annotated using manually curated gene sets, drawn from Gene Ontology and Reactome sets, to identify gene programs preferentially benefiting from generative augmentation.

2.6. *Primary Modeling Contributions*

The primary contributions of this study, in comparison to the prior art, are: (i) the finetuning of an image generator to human skin with conditioning that combines ST-conditioning using a 50-gene panel, architecture conditioning, and spot-level cell-type conditioning, (ii) automated compositional CellViT-based assessment framework for measuring concordance between real and generated tissue, (iii) the conditional perturbation experiments, and (iv) comparison of generative augmentation strategies for downstream prediction, with architecture- and gene-level analyses.

3. Results

3.1. *ST-Informed Generation Recapitulates the Tissue Microenvironment Across Diverse Histologic Contexts*

3.1.1. *Correlation Between Measured and Generated Cell-Type Abundance*

Representative examples suggested that ST-informed generation faithfully preserved tissue microenvironment composition across diverse histologic contexts and responded appropriately to controlled perturbations of the biological conditioning signals (**Fig. 2**). These observations were quantitatively evaluated by comparing CellViT-derived epithelial, stromal, and immune-cell abundance between real and generated tissue across all Visium spots (**Fig. 3**).

Generated histology demonstrated concordance with the measured tissue microenvironment, with epithelial (Spearman $\rho = 0.16$, $p<0.001$), stromal ($\rho = 0.48$, $p<0.001$), and immune-cell abundance ($\rho = 0.37$, $p<0.001$) all exhibiting significant positive correlations between real and generated tissue (see **Fig. 3A-C** for correlations by abundance). Architecture-stratified analyses revealed greater variability across tissue contexts, with the strongest agreement generally observed for cell types enriched within a given architecture. Notable examples included stromal cells in fat ($\rho = 0.72$, 95% CI: 0.68–0.77), epithelial cells in stromal regions ($\rho = 0.48$, 95% CI: 0.31–0.63), immune cells in dermis ($\rho = 0.46$, 95% CI: 0.43–0.48), and stromal cells in inflammatory regions ($\rho = 0.54$, 95% CI: 0.18–0.80) (**Fig. 3D**).

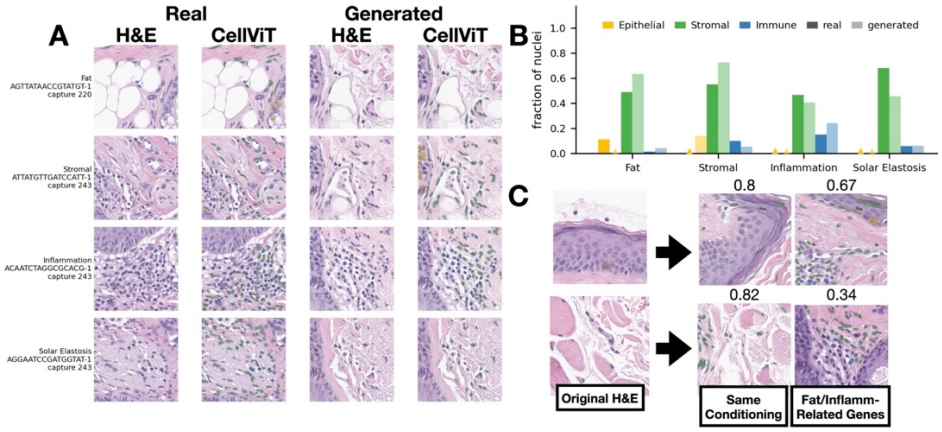


Figure 2. ST-informed generation preserves biologically meaningful tissue microenvironments and enables controlled perturbation of biological conditioning signals. (A) Representative real and ST-informed generated H&E patches from five Visium spots spanning diverse tissue architectures, with corresponding CellViT-derived cell segmentations illustrating preservation of local cellular composition despite stochastic variation in morphology. **(B)** Comparison of CellViT-derived epithelial, stromal, and immune-cell fractions between representative real and generated patches, for the same set of patches. **(C)** Representative perturbation experiments in which individual biological conditioning signals were selectively modified while all other inputs were held constant. Cosine similarities to the original H&E images are displayed above each generated example, demonstrating higher cosine similarity when same conditioning inputs are used versus perturbing genes/architectures to include more fat and inflammation.

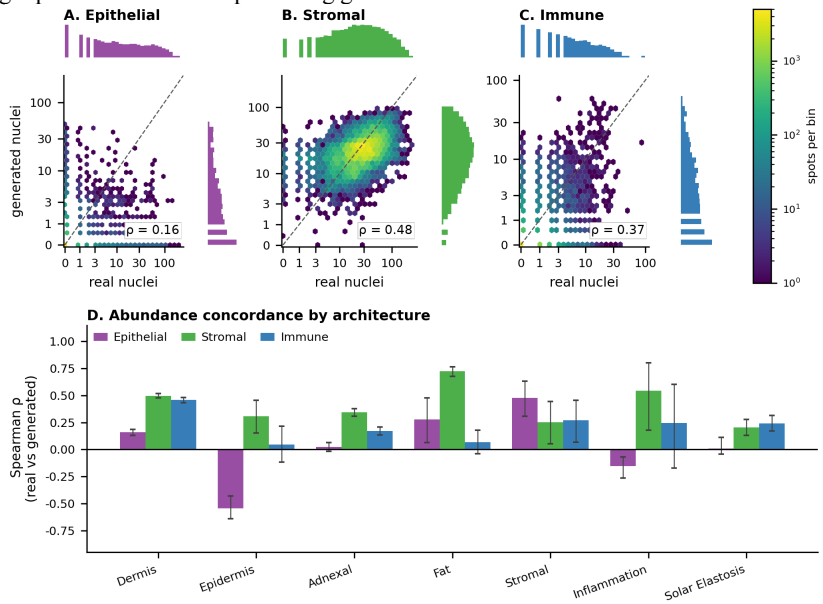


Figure 3. Generated histology quantitatively recapitulates the tissue microenvironment. (A-C) Hexbin density plots comparing CellViT-derived cell abundances between real and generated H&E patches across all Visium spots for **(A)** epithelial, **(B)** stromal, and **(C)** immune-cell populations. Color indicates the number of observations within each hexagonal bin, dashed lines indicate the line of identity, and inset Spearman correlation coefficients quantify concordance between real and generated tissue. **(D)** Architecture-stratified Spearman correlations between real and generated cell abundances. Correlations are shown separately for epithelial, stromal, and immune-cell populations within each tissue architecture, with error bars indicating 95% confidence intervals across stochastic generations.

3.1.2. *Changes in Cell-Type Fractions and Image Features Following Perturbation of Biological Conditioning Signals*

To evaluate whether the model learned biologically meaningful relationships between conditioning signals and tissue morphology, we perturbed individual biological inputs while holding all other conditioning variables constant and report on examples of resulting shifts in cell proportions in Fig. 2. As expected, perturbations produced corresponding changes in CellViT-derived epithelial, stromal, and immune-cell fractions.

For representative epidermal image patches, perturbing gene-expression programs toward stromal and inflammatory programs reduced the similarity between generated and corresponding real image patches (mean cosine similarity = 0.62 and 0.64, respectively) relative to images generated under the original biological conditioning (mean cosine similarity = 0.794). Changing the tissue architecture label produced a comparatively smaller reduction in similarity (0.771 and 0.703 for stromal and inflammatory labels, respectively).

3.2. *ST-Informed Generative Augmentation Provides Context-Dependent Improvements in Downstream Prediction Tasks*

3.2.1. *Spatial Gene Expression Prediction*

Across 2,000 HVGs, ST-informed generative augmentation produced modest but consistent improvements in spatial gene expression prediction when applied to a frozen pathology foundation model (Phikon) (**Fig. 4A**). Pretraining on generated images followed by fine-tuning using real and generated images increased the median Pearson correlation coefficient (PCC) by +0.0103 across all HVGs and by +0.0160 for the predefined difficult-to-predict gene set. Although overall gains were modest, improvements were concentrated within specific tissue architectures (**Fig. 4A, C**). The largest improvements were observed in the adnexal compartment (ΔSpearman = +0.0175, p = 0.002) and dermis (ΔSpearman = +0.0110, p = 0.010), with smaller improvements in stromal tissue (ΔSpearman = +0.0080), inflammation (ΔSpearman = +0.0075), and epidermis (ΔSpearman = +0.0058), whereas solar elastosis and adipose tissue showed little or no benefit (**Fig. 4A,C**). Correspondingly, the genes exhibiting the greatest improvement were enriched for dermal extracellular matrix and stromal biology (e.g., ELN, CILP) as well as perivascular and endothelial programs (e.g., RGS5, PLVAP, and LYVE1) (**Fig. 5**).

To assess whether augmentation preferentially benefits models lacking large-scale visual pretraining, we repeated these experiments using a ResNet trained entirely from scratch. In contrast to the modest gains observed with Phikon, ST-informed generative augmentation produced substantially larger improvements, increasing prediction performance by ΔSpearman = +0.051 on the held-out test set (**Fig. 4A, C**), indicating that generated images provide the greatest benefit when task-specific visual representations must be learned directly from the available training data.

3.2.2. *Tissue Architecture Classification*

ST-informed generative augmentation similarly improved tissue architecture classification (**Fig. 4B**). Using the frozen Phikon encoder, augmentation increased macro-F1 by +0.0141 (p = 0.0098, Wilcoxon signed-rank test) relative to conventional training. Performance gains were not uniform across tissue architectures, with the largest improvements observed for inflammation (ΔF1 = +0.0376), stroma (ΔF1 = +0.0360), and dermis (ΔF1 = +0.0017) (**Fig. 4B,D**), whereas more abundant tissue architectures exhibited comparatively smaller changes. The benefits of augmentation were substantially larger for models trained from scratch. Incorporating ST-informed generated images alongside conventional augmentation increased macro-F1 by +0.1146 on the held-out test set (**Fig. 4B,D**), representing an approximately eight-fold larger improvement than observed for the pretrained foundation model.

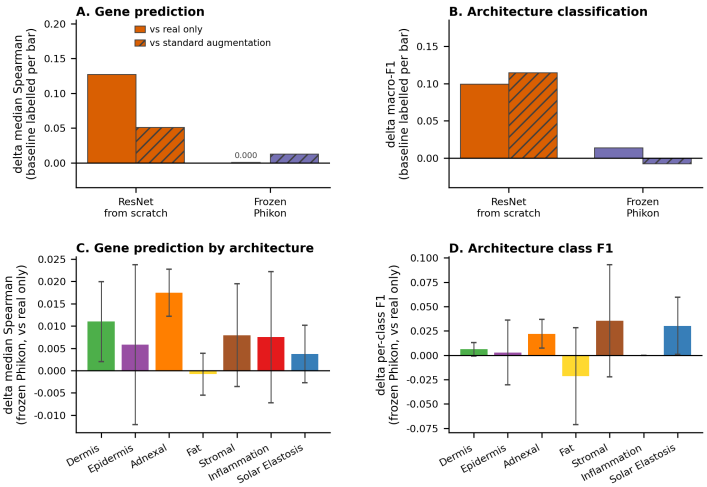


Figure 4. ST-informed generative augmentation improves downstream prediction in a task- and model-dependent manner. **(A)** Improvement in spatial gene expression prediction (ΔSpearman) following ST-informed generative augmentation for a ResNet trained from scratch and a frozen pathology foundation model (Phikon). Improvements are shown relative to training on real images alone and, for the scratch-trained model, relative to conventional image augmentation. **(B)** Improvement in tissue architecture classification following ST-informed generative augmentation for both downstream models. **(C)** Architecture-stratified improvements in spatial gene expression prediction (ΔSpearman) for the frozen Phikon model relative to training on real images alone. **(D)** Architecture-stratified improvements in tissue architecture classification for the frozen Phikon model relative to training on real images alone. Error bars denote 95% confidence intervals.

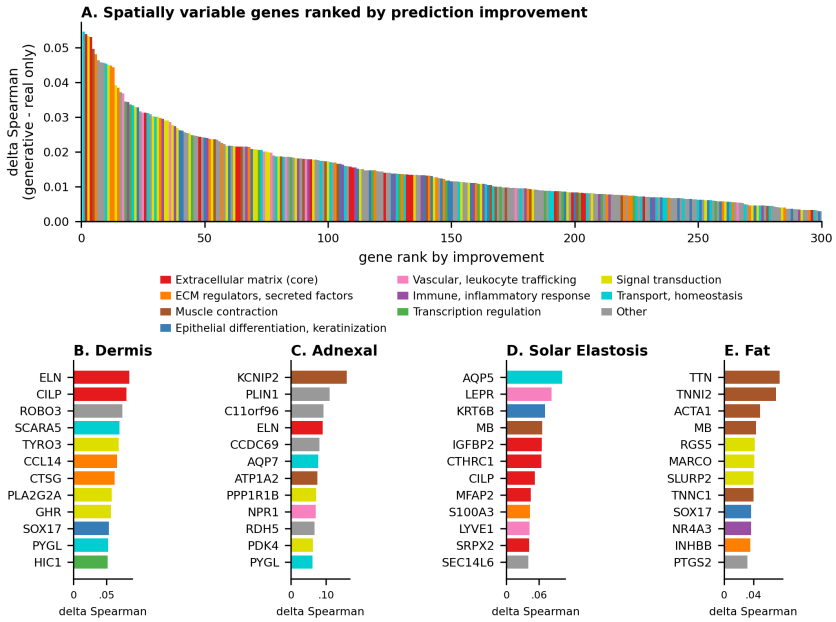


Figure 5. ST-informed generative augmentation preferentially improves biologically coherent molecular programs within specialized tissue microenvironments. **(A)** Spatially variable genes ranked by the improvement in prediction performance (ΔSpearman) following ST-informed generative augmentation relative to training on real images alone. Genes are colored according to manually curated gene set categories. **(B–E)** Representative architecture-specific rankings of the highest-improving genes.

4. Discussion

The development of multimodal approaches in computational pathology increasingly relies on large collections of histopathologic images paired with molecular measurements that capture the breadth of histologic and molecular variation across patient populations. However, assembling cohorts that adequately represent this variation remains challenging, particularly for rare tissue states and disease subtypes, especially for less characterized patient populations. Generative modeling, particularly

latent diffusion models, has emerged as a promising strategy for expanding training datasets through the generation of realistic histopathology. Beyond producing visually convincing images, these approaches have shown promise for mitigating out-of-distribution shifts and improving the robustness, generalizability, and fairness of downstream predictive models [11,28,29]. Nevertheless, ensuring that generated tissue faithfully reflects the underlying local tissue microenvironment remains an important challenge, motivating the incorporation of spatial transcriptomic information, which capture aspects of the tissue microenvironment, into generative frameworks.

Within the spatial transcriptomics literature, diffusion-based methods have primarily focused on denoising, super-resolution, and virtual transcriptomics [10,11,30–34]. In contrast, our study investigated whether spatial transcriptomic measurements can be leveraged to generate biologically faithful histopathology that improves downstream computational pathology tasks. Our findings demonstrate that incorporating spatially resolved molecular information into histopathology generation preserves key features of the tissue microenvironment while providing measurable improvements in spatial gene expression prediction and tissue architecture classification. Although these improvements were modest overall, they were concentrated within tissue architectures characterized by distinct molecular programs and typically represented by a smaller number of Visium spots/image patches, suggesting that ST-informed generation may prove valuable for specific tissue contexts. The improvements in downstream modeling were particularly notable for models trained from scratch, suggesting that ST-informed augmentation is most beneficial when visual representations have not already been learned through large-scale pretraining.

Looking ahead, ST-informed image generation has the potential to enable several emerging applications in computational pathology. One promising direction is the development of tissue "digital twins," or virtual representations of patient specimens that preserve key histologic and molecular characteristics for simulation, hypothesis generation, and algorithm development [35–39]. Such models could facilitate privacy-preserving data sharing across institutions while expanding access to biologically representative tissue for training and validation of computational pathology methods. Although the present study focused on patch-level image generation, extending these approaches to whole-slide simulation represents an important next step. Achieving this goal will require methods capable of modeling spatial transcriptomic patterns [40] across larger tissue regions while maintaining continuity between adjacent image patches, these approaches may build on previously developed spatially aware stitching approaches [41]. Future studies should also investigate alternative conditioning strategies, including larger gene panels, cell-cell or higher order interaction features, multimodal molecular measurements, and other forms of spatial context. Extending generative approaches to higher-resolution platforms such as Visium HD, which approaches single-cell resolution, and single-cell Xenium profiling, presents the next logical extension for this work.

This study has several limitations. First, the analysis was performed using histologically normal skin tissue adjacent to cutaneous malignancies rather than tumor tissue itself. Although these specimens exhibit biologically relevant field effects associated with photocarcinogenesis, they do not capture the architectural and molecular complexity of tumors. For example, tumor adjacent skin may carry a higher inflammatory burden than that found from a healthy participant or distant sites in the same patient. Future studies should evaluate ST-informed generation across varied tumor types and disease contexts, including skin biopsies evaluating autoimmune/inflammatory conditions and,

separately, containing tumor. Second, the study was conducted using data from a single institution and may therefore reflect institution-specific tissue processing, staining, imaging, and spatial transcriptomic characteristics. Third, biological conditioning was limited to a predefined panel of 50 highly variable genes selected to provide a concise representation of the local molecular state. Genes with heightened lineage specificity do not necessarily possess these characteristics. Alternative gene selection strategies or larger gene panels may further improve image generation. Fourth, evaluation of image realism relied primarily on CellViT-derived estimates of tissue microenvironment composition. While these provide an objective assessment of cellular composition, future studies will supplement algorithmic associations with independent, blinded expert pathologic review as CellViT was trained on tissue histology and thus evaluations were not fully independent of the conditioning signals. In addition, automated cell and tissue assignment generally performs well in nonneoplastic tissue, and evaluation in tumor-containing tissue may provide a more challenging setting in which performance differences are more apparent. Finally, although ST-informed augmentation consistently improved downstream prediction relative to non-augmented training, the predictiveness of virtual spatial transcriptomics remained below previously reported performance. Because our objective was to quantify the incremental value of image generation rather than maximize predictive performance, future work jointly optimizing both the generative and downstream prediction models may yield substantially larger improvements.

5. Acknowledgements

The authors acknowledge support from the National Institutes of Health (R37CA299857, P30GM149408 and P20GM130454), Cedars-Sinai Medical Center, the Cedars Cancer Developmental Funds (Team Science and Molecular Twins Awards), the Dartmouth Cancer Center (5P30CA023108-47), philanthropic support from Dartmouth Health Pathology, and the EDIT AI Program. Large language models (ChatGPT, OpenAI; Apple Intelligence) were used to assist with language editing, manuscript organization, and code development. All scientific content, analyses, interpretations, and final manuscript revisions were reviewed and verified by the authors, who take full responsibility for the accuracy and integrity of the work.

References

1. Burgess, D. J. Spatial transcriptomics coming of age. *Nature Reviews Genetics* **20**, 317–317 (2019).
2. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat Methods* **19**, 534–546 (2022).
3. Hemker, K. *et al.* Towards Spatial Transcriptomics-driven Pathology Foundation Models. Preprint at https://doi.org/10.48550/arXiv.2602.14177 (2026).
4. Chen, W. *et al.* A visual–omics foundation model to bridge histopathology with spatial transcriptomics. *Nat Methods* **22**, 1568–1582 (2025).
5. Chen, R. J. *et al.* Towards a general-purpose foundation model for computational pathology. *Nat Med* **30**, 850–862 (2024).
6. Filiot, A., Jacob, P., Kain, A. M. & Saillard, C. Phikon-v2, A large and public feature extractor for biomarker prediction. Preprint at https://doi.org/10.48550/arXiv.2409.09173 (2024).
7. Lu, M. Y. *et al.* A visual-language foundation model for computational pathology. *Nature Medicine* **30**, 863–874 (2024).
8. Zimmermann, E. *et al.* Virchow2: Scaling Self-Supervised Mixed Magnification Models in Pathology. Preprint at https://doi.org/10.48550/arXiv.2408.00738 (2024).
9. Filiot, A. *et al.* Scaling self-supervised learning for histopathology with masked image modeling. *MedRxiv* 2023–07 (2023).
10. Cui, Y. *et al.* DiffusionST: a deep generative diffusion model-based framework for enhancing spatial transcriptomics data quality and identifying spatial domains. *Briefings in Bioinformatics* **26**, bbaf390 (2025).
11. Srinivasan, G. *et al.* Abstract 2439: Spatial transcriptomics-informed latent diffusion models generate histopathological images reflective of tumor microenvironment. *Cancer Res* **85**, 2439 (2025).
12. Yellapragada, S. *et al.* PathLDM: Text conditioned Latent Diffusion Model for Histopathology. *IEEE Winter Conf Appl Comput Vis* **2024**, 5170–5179 (2024).
13. DiPalma, J. Deep Learning for Fine-Grained Digital Histopathology Image Analysis. *Computer Science Technical Reports* https://digitalcommons.dartmouth.edu/cs_tr/386 (2025).
14. Shulman, E. D. *et al.* AI-Driven Spatial Transcriptomics Unlocks Large-Scale Breast Cancer Biomarker Discovery from Histopathology. 2024.10.16.618609 Preprint at https://doi.org/10.1101/2024.10.16.618609 (2025).
15. Fatemi, M. *et al.* Inferring spatial transcriptomics markers from whole slide images to characterize metastasis-related spatial heterogeneity of colorectal tumors: A pilot study. *Journal of Pathology Informatics* 100308 (2023) doi:10.1016/j.jpi.2023.100308.
16. Zeng, Y. *et al.* Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. *Briefings in Bioinformatics* **23**, bbac297 (2022).
17. Huang, T., Liu, T., Babadi, M., Ying, R. & Jin, W. STPath: a generative foundation model for integrating spatial transcriptomics and whole-slide images. *NPJ Digital Medicine* **8**, 659 (2025).
18. Wang, Y. *et al.* FmH2ST: foundation model-based spatial transcriptomics generation from histological images. *Nucleic Acids Research* **53**, gkaf865 (2025).
19. Ktena, I. *et al.* Generative models improve fairness of medical classifiers under distribution shifts. *Nat Med* **30**, 1166–1173 (2024).
20. Srinivasan, G. *et al.* Potential to Enhance Large Scale Molecular Assessments of Skin Photoaging through Virtual Inference of Spatial Transcriptomics from Routine Staining. *Pac Symp Biocomput* **29**, 477–491 (2024).

21. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. in *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* 10674–10685 (ieee, 2022).
22. Hörst, F. *et al.* CellViT: Vision Transformers for precise cell segmentation and classification. *Medical Image Analysis* **94**, 103143 (2024).
23. Liu, X., Gong, C. & Liu, Q. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. Preprint at https://doi.org/10.48550/arXiv.2209.03003 (2022).
24. Ho, J. & Salimans, T. Classifier-Free Diffusion Guidance. Preprint at https://doi.org/10.48550/arXiv.2207.12598 (2022).
25. Žigutytė, L. *et al.* Counterfactual diffusion models for interpretable morphology-based explanations of artificial intelligence models in pathology. *bioRxiv* 2024–10 (2025).
26. He, B. *et al.* Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* **4**, 827–834 (2020).
27. Moran, P. A. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
28. Ho, M. M., Dubey, S., Chong, Y., Knudsen, B. & Tasdizen, T. F2fldm: Latent diffusion models with histopathology pre-trained embeddings for unpaired frozen section to ffpe translation. in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* 4382–4391 (IEEE, 2025).
29. Alfasly, S., Uegami, W., Hoq, M. E., Alabtah, G. & Tizhoosh, H. Semantic and visual crop-guided diffusion models for heterogeneous tissue synthesis in histopathology. *Advances in Neural Information Processing Systems* **38**, 125564–125611 (2026).
30. Sun, P. *et al.* STsim: An Efficient Diffusion-Based Model for Generating Spot Images from Gene Expression in Spatial Transcriptomics. in *2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 4083–4088 (2025). doi:10.1109/BIBM66473.2025.11356203.
31. Vlachas, P. R., Nonchev, K., Koelzer, V. H. & Rätsch, G. Transcriptomics-Conditioned Virtual Tissue Synthesis via Diffusion Transformers. *bioRxiv* 2026–05 (2026).
32. Jiao, S. *et al.* DiffuST: a latent diffusion model for spatial transcriptomics denoising. *IEEE Transactions on Computational Biology and Bioinformatics* https://ieeexplore.ieee.org/abstract/document/11508149/ (2026).
33. Wang, X., Price, S. J. & Li, C. C3-Diff: Super-resolving Spatial Transcriptomics via Cross-modal Cross-content Contrastive Diffusion Modelling. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 5463–5473 (2026).
34. Zhu, S., Zhu, Y., Tao, M. & Qiu, P. Diffusion generative modeling for spatially resolved gene expression inference from histology images. in *International Conference on Learning Representations* vol. 2025 19666–19690 (2025).
35. Olawade, D. B. *et al.* Digital twins in oncology: From predictive modelling to personalised treatment strategies. *Critical Reviews in Oncology/Hematology* 105171 (2026).
36. Silva, A. & Vale, N. Digital twins in personalized medicine: bridging innovation and clinical reality. *Journal of personalized medicine* **15**, 503 (2025).
37. Khoshfekr Rudsari, H. *et al.* Digital twins in healthcare: a comprehensive review and future directions. *Frontiers in digital health* **7**, 1633539 (2025).
38. Moore, J. H. *et al.* SynTwin: A graph-based approach for predicting clinical outcomes using digital twins derived from synthetic patients. in *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* vol. 29 96 (NIH Public Access, 2024).
39. Moingeon, P., Chenel, M., Rousseau, C., Voisin, E. & Guedj, M. Virtual patients, digital twins and causal disease models: Paving the ground for in silico clinical trials. *Drug discovery today* **28**, 103605 (2023).

40. Zhu, J., Shang, L. & Zhou, X. SRTsim: spatial pattern preserving simulations for spatially resolved transcriptomics. *Genome Biol* **24**, 39 (2023).
41. Harb, R., Pock, T. & Müller, H. Diffusion-based generation of Histopathological Whole Slide Images at a Gigapixel scale. in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* 5119–5128 (2024). doi:10.1109/WACV57701.2024.00505.