Spatial Transcriptomics-Informed Histopathologic Image Generation for Modeling the Skin Tissue Microenvironment

Vedhsai Thiriveedi[1], Joshua J. Levy PhD[2]

[1] Nashua High School South [2] Department of Pathology, Dartmouth Hitchcock Medical Center and Cedars Sinai Medical Center

ABSTRACT

- ST couples genome-wide expression with histologic context, but high cost limits application to large cohorts.
- Effort has focused on predicting molecules from histology. We evaluate the reverse: generating histology from ST measurement.
- A conditional latent diffusion model is trained on 39,579 Visium spots from 16 patients across 7 annotated tissue architectures.
- Generated tissue agrees significantly with real tissue on epithelial, stromal and immune abundance, and augmentation helps specific microenvironments rather than uniformly.

INTRODUCTION

- H&E staining has been the basis of histologic diagnosis for over a century. ST platforms such as 10x Genomics Visium measure gene expression at regularly spaced locations while preserving spatial position.
- ST profiling is costly and labor intensive, so datasets incompletely sample rare tissue architectures and underrepresented patient populations.
- Downstream models trained on them become biased toward well represented tissue contexts and underperform in biologically varied regions.
- Within ST, conditional diffusion has addressed denoising, super resolution and virtual transcriptomics, not generation of the histology itself.
- Goal: determine whether ST measurement can condition image generation in a way that carries biologically relevant information, and whether it improves downstream computational pathology.

METHODS

- Cohort: 16 normal skin specimens collected as bisected Burow's triangles, routinely discarded during Mohs micrographic surgery. 8 female and 8 male, ages 19 to 85 (mean 60.4), all chronically sun exposed sites.
- FFPE tissue transferred to Visium slides by 10x Genomics CytAssist; slides scanned at 40x on a Leica GT450, with expert dermatopathologist annotation of all seven architectures.
- Generator: conditional latent diffusion. Conditioning is a numeric vector, not a text prompt: a 50-gene highly variable panel, the architecture label, and cellularity.
- Raw log1p CP10k expression passes through a learned Transformer encoder; conditioning enters the denoising U-Net through AdaLN-Zero with cross attention. Generation runs in the latent space of a frozen VAE.
- Evaluation: CellViT segments nuclei on real and generated patches; epithelial, stromal and immune abundance compared per spot by Spearman correlation.
- Perturbation arm modifies one conditioning input while all others are held constant, measured by cosine similarity to the matched real patch.
- Downstream: gene expression prediction and architecture classification, on a frozen Phikon foundation model and a ResNet from scratch. Pretrain on generated, then fine tune on real.

Critical methodology: captures were partitioned at the slide level, not the spot level, so each capture was assigned exclusively to training (25,438 spots), validation (4,029) or test (10,112). This prevents leakage of shared tissue-level features, a common pitfall in digital pathology that artificially inflates accuracy. Every comparison was pre-registered before running and all arms were reported.

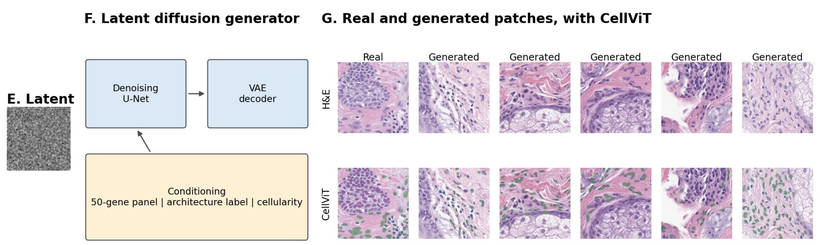


Figure 1: Generator schematic. Latent noise, the denoising U-Net and VAE decoder, and one real patch alongside five independently generated patches from identical conditioning, each with CellViT segmentation.

RESULTS

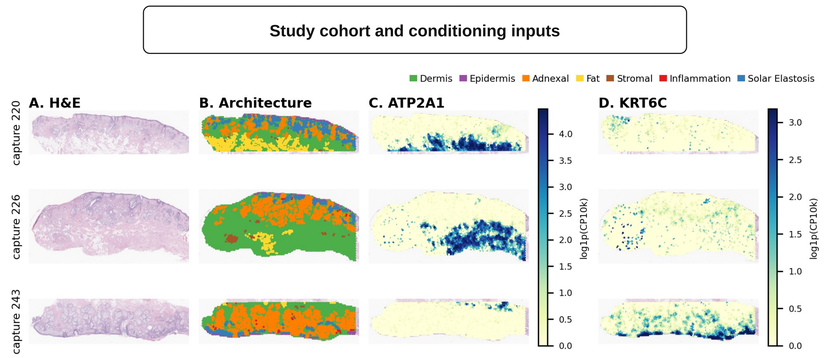


Figure 2: Study cohort. Whole slide H&E, expert tissue architecture annotation, and Visium expression maps for ATP2A1 and KRT6C across the three held-out test captures.

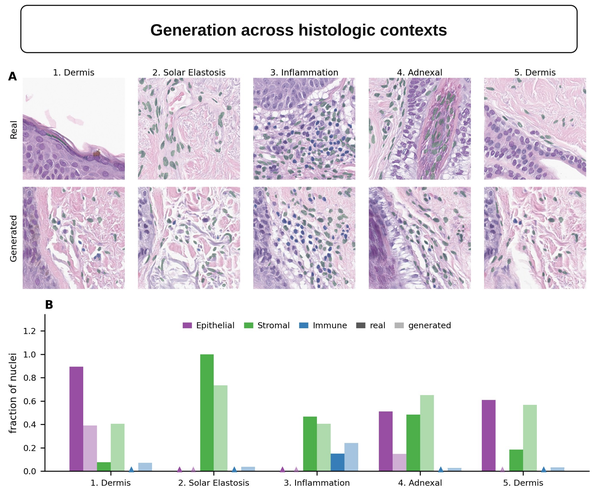


Figure 3: Real and generated patches from five Visium spots spanning different tissue architectures, with CellViT segmentation and the corresponding cell fraction comparison.

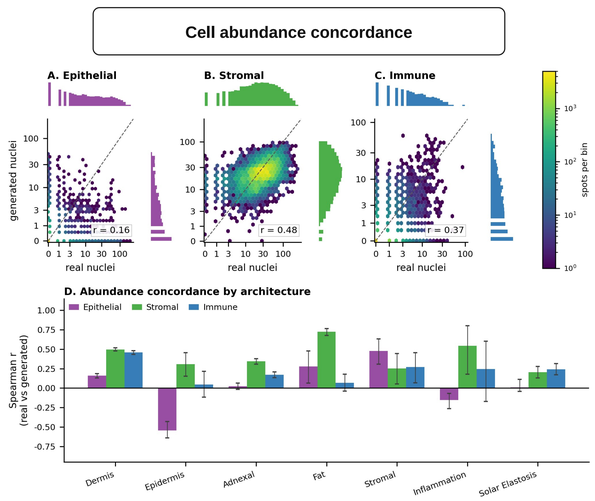
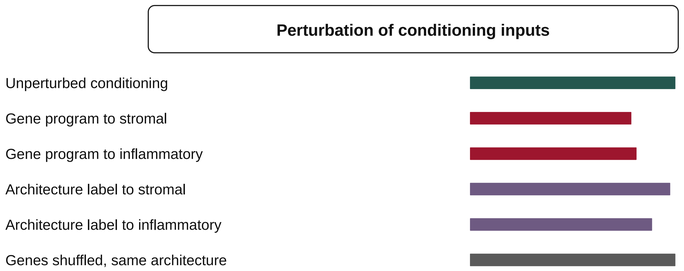


Figure 4: Hexbin density of CellViT cell abundance per Visium spot for epithelial, stromal and immune populations, with Spearman concordance stratified by tissue architecture in panel D.

RESULTS

- Pooled across all test spots: epithelial 0.16, stromal 0.48, immune 0.37, all significant at $p < 0.001$.
- Stromal composition generalizes best. It is the most abundant lineage, and agreement is strongest where a lineage is enriched, reaching 0.72 for stromal cells in fat.
- Within epidermis, epithelial concordance is negative at minus 0.535. CellViT is PanNuke trained and calls epidermal keratinocytes neoplastic, a class excluded here, so 67 of 155 epidermal test spots read zero real epithelial content. An evaluation artifact, not a generator failure.



Mean cosine similarity to the matched real patch across 155 epidermal test spots. Two real patches from the same architecture reach 0.78, from different architectures 0.50. Retargeting the gene program drives the output below the same-architecture floor, so the generator is using the molecular signal.

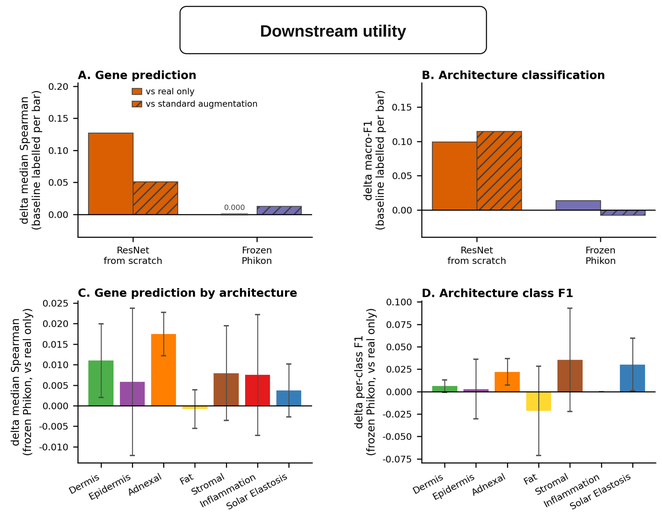


Figure 5: Change in spatial gene expression prediction and tissue architecture classification following ST-informed generative augmentation, overall and stratified by architecture.

CONCLUSION

- Principal Finding: spatial transcriptomics can condition histopathologic image generation, and perturbation confirms the generator uses the molecular signal.
- Downstream Impact: overall gene prediction gain is small at +0.0103, concentrating in adnexal (+0.0175) and dermis (+0.0110). A ResNet trained from scratch gains +0.051 and +0.1146 macro-F1, eight times the frozen foundation model.
- Interpretation: generated tissue benefits the microenvironments where real training examples are scarce, and benefits most where visual representations are not already learned.
- Bottleneck: generated patches stay separable from real by every encoder tested, with AUC at or near 1.0, locating the limit at decoder realism rather than the conditioning.
- Future Directions: whole slide simulation, tissue digital twins, larger conditioning panels, and joint optimization of the generative and downstream models. Manuscript submitted to PSB 2027, Precision Medicine session.