

ST2HE: enhancing spatial transcriptomics interpretability via virtual staining for histological annotation

Zhentao Liu¹, Arun Das^{2,3}, Wen Meng^{2,4}, Yu-Chiao Chiu^{3,5,6}, Shou-Jiang Gao^{2,4}, Yufei Huang^{1,2,3,7,*}

¹Department of Electrical and Computer Engineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, United States

²Cancer Virology Program, UPMC Hillman Cancer Center, Pittsburgh, PA, United States

³Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

⁴Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

⁵Cancer Therapeutics Program, University of Pittsburgh Medical Center Hillman Cancer Center, Pittsburgh, PA, United States

⁶Pittsburgh Liver Research Center, University of Pittsburgh Medical Center and University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

⁷Department of Pharmaceutical Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

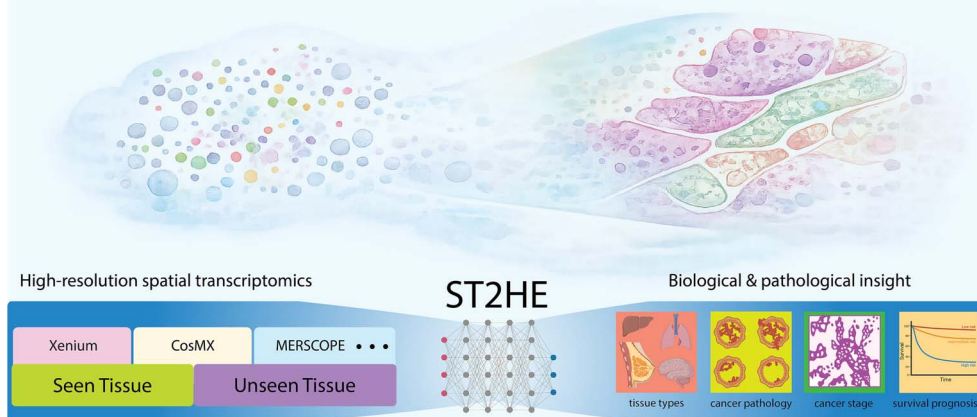
*Corresponding author. UPMC Hillman Cancer Center, 5051 Centre Avenue, Pittsburgh, PA 15213, United States. E-mail: yuh119@pitt.edu

Abstract

High-resolution spatial transcriptomics (HR-ST) technologies offer unprecedented insights into tissue architecture but lack standardized frameworks for histological annotation. We present ST2HE, a cross-platform generative framework that synthesizes virtual hematoxylin and eosin images directly from HR-ST data. ST2HE integrates nuclei morphology and spatial transcript coordinates using a one-step diffusion model, enabling histologically informative image generation across diverse tissue types and HR-ST platforms. Conditional and tissue-independent variants support both known and novel tissue contexts. Evaluations on breast cancer, non-small cell lung cancer, and Kaposi's sarcoma demonstrate ST2HE's ability to preserve morphological features and support downstream annotations of tissue histology and phenotype classification. Ablation studies reveal that larger context windows, balanced loss functions, and multi-colored transcript visualization enhance image fidelity. ST2HE bridges molecular and histological domains, enabling interpretable, scalable annotation of HR-ST data and advancing computational pathology.

Graphical Abstract

Enhancing interpretability of high-resolution spatial transcriptomics



Keywords spatial transcriptomics, interpretability, virtual histology, computational pathology, generative models

Received: February 10, 2026. **Revised:** May 4, 2026. **Accepted:** May 20, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Introduction

Spatial transcriptomics (ST) technologies are rapidly advancing, enabling gene expression mapping with preserved spatial context. High-resolution ST (HR-ST) techniques such as Xenium [1], NanoString CosMx SMI [2], and MERFISH [3] allow precise localization of transcripts across diverse tissue types. Despite these advances, no standardized system exists for annotating spatial structures and phenotypes in HR-ST data. Current strategies fall into two categories: (i) Manual expert annotation, which is accurate but subjective and difficult to scale. (ii) Computational approaches that cluster cells or analyse neighborhood composition but lack biologically meaningful labels [4–7], as no reference database links spatial transcriptomic patterns to known tissue structures. These limitations highlight the need for frameworks that integrate HR-ST data with high-quality histopathological annotations.

A central challenge is the absence of a reference linking gene expression patterns to histopathological structures. Subsequently, annotations such as tumor subtypes (e.g. ductal carcinoma in situ (DCIS)) cannot be reliably annotated without expert pathological examination (Fig. 1a). Thus, despite HR-ST's resolution, its interpretability remains constrained (Fig. 1a). By contrast, hematoxylin and eosin (H&E) staining remains the gold standard in histopathological examination [8], supported by decades of practice and rich literature [8–10]. Numerous foundation models trained on H&E images including convolutional neural networks [10], transformer-based architectures [11], and CLIP-based models [12–14] capture morphological and pathological features and have proven effective in tumor classification, tissue segmentation, and morphology recognition. Transferring this knowledge to HR-ST could significantly enhance tissue characterization by integrating molecular and histopathological context.

However, pairing H&E with HR-ST remains challenging. While Xenium supports same-slide H&E, tissue processing can damage the tissue, thus reducing fidelity. In many practical scenarios, researchers face a trade-off between H&E and immunofluorescence staining for downstream validation, particularly when tissue quantity is limited, RNA integrity requirements constrain staining order, or signal degradation from sequential staining reduces data quality. Although advanced multiplexed workflows such as cyclic immunofluorescence (e.g. CODEX) can sometimes enable both modalities on the same section, these approaches require specialized instrumentation and expertise that are not universally accessible. Even when both modalities are present, image registration is required to align the two imaging modalities — a process that is typically performed manually, limiting scalability for large cohort studies. While automated registration methods exist, they often yield suboptimal alignment quality, particularly in regions with tissue deformation or morphological distortion introduced during staining. For platforms that do not support same-slide H&E, the use of adjacent sections introduces additional biological variation. These issues collectively complicate the integration of histology with HR-ST data at scale. Recent advances in virtual staining have shown that generative models can generate realistic H&E images from label-free [15–20] or fluorescence-labeled modalities [21]. Yet, no method exists to directly translate HR-ST into H&E images, despite HR-ST's unique ability to capture molecular and spatial information beyond imaging.

To address this gap, we propose ST2HE, the first framework for generating virtual H&E directly from HR-ST data. Built on

Pix2Pix-Turbo [22], a one-step latent diffusion model for high-fidelity image-to-image translation (Fig. 1b), ST2HE takes 4',6-diamidino-2-phenylindole (DAPI) image overlaid with locations of spatial transcripts as inputs, where the image helps learn nuclear morphology and transcript locations to describe spatial gene expression (Fig. 1c). A key design innovation of ST2HE lies in how these spatial transcripts are encoded. Rather than assigning fixed colors to specific genes, each gene is assigned a unique color from a palette that is independently randomized per tile. This deliberate design choice decouples gene identity from color representation, enabling the model to learn from the spatial organization and density patterns of transcripts rather than gene-specific signals. Critically, this makes ST2HE independent of any specific gene panel, allowing it to generalize across different HR-ST platforms such as Xenium and NanoString CosMx that use distinct gene panels. Biological specificity is thus preserved through the spatial organization of transcripts within tissue morphology, rather than through gene-specific color signals. We trained ST2HE using Xenium data spanning five tissue types (Fig. 1d) and developed two complementary variants: ST2HE-CondGen, a tissue-conditioned model for training tissue types and ST2HE-UnCondGen, a tissue-independent model for unseen tissue contexts (Fig. 1e). Together, these two models provide flexibility for real-world applications and enable direct use of pretrained pathology models like CONCH [13] to annotate HR-ST samples. By bridging HR-ST with histopathology, ST2HE expands the utility of HR-ST for discovery, precision medicine, and translational research.

Results

ST2HE generates histologically accurate virtual H&E for xenium breast cancer samples

We evaluated ST2HE using a publicly available Xenium breast cancer dataset, which provides whole-tissue ST profiling of breast cancer samples (Section:Methods). Cell types were previously annotated by integrating scRNA-seq and H&E imaging, identifying major populations, including two tumor subtypes: DCIS and invasive carcinoma [23] (Fig. 2a). Although these subtypes occupy distinct regions in gene expression space, they remain challenging to annotate without histological context. Gene expression separation in embedding space does not constitute pathological annotation, as the subtype labels themselves were derived using H&E imaging in the original study. Furthermore, pathological annotations of tumor subtypes such as DCIS versus invasive carcinoma are fundamentally defined by tissue architecture and spatial organization rather than individual cell gene expression profiles. Single-cell typing alone cannot capture these tissue-level structural patterns. To illustrate this challenge, we simulated the real-world scenario where only HR-ST data is available by merging all tumor subtype labels into a single 'tumor cell' category and performing spatial niche clustering using k-means on neighborhood cell type composition vectors. Despite capturing spatial neighborhood organization, niche clustering assigned DCIS and invasive carcinoma cells to the same cluster (Fig. 2b), demonstrating that transcriptomic neighborhood analysis alone cannot distinguish these subtypes. This underscores the need for histopathological context to enable accurate annotation of HR-ST data, which ST2HE directly addresses.

Because breast cancer is a training tissue type, we used the conditional ST2HE to generate virtual H&E images (Fig. 2c), enabling the generation of more accurate breast cancer-specific morphology. This

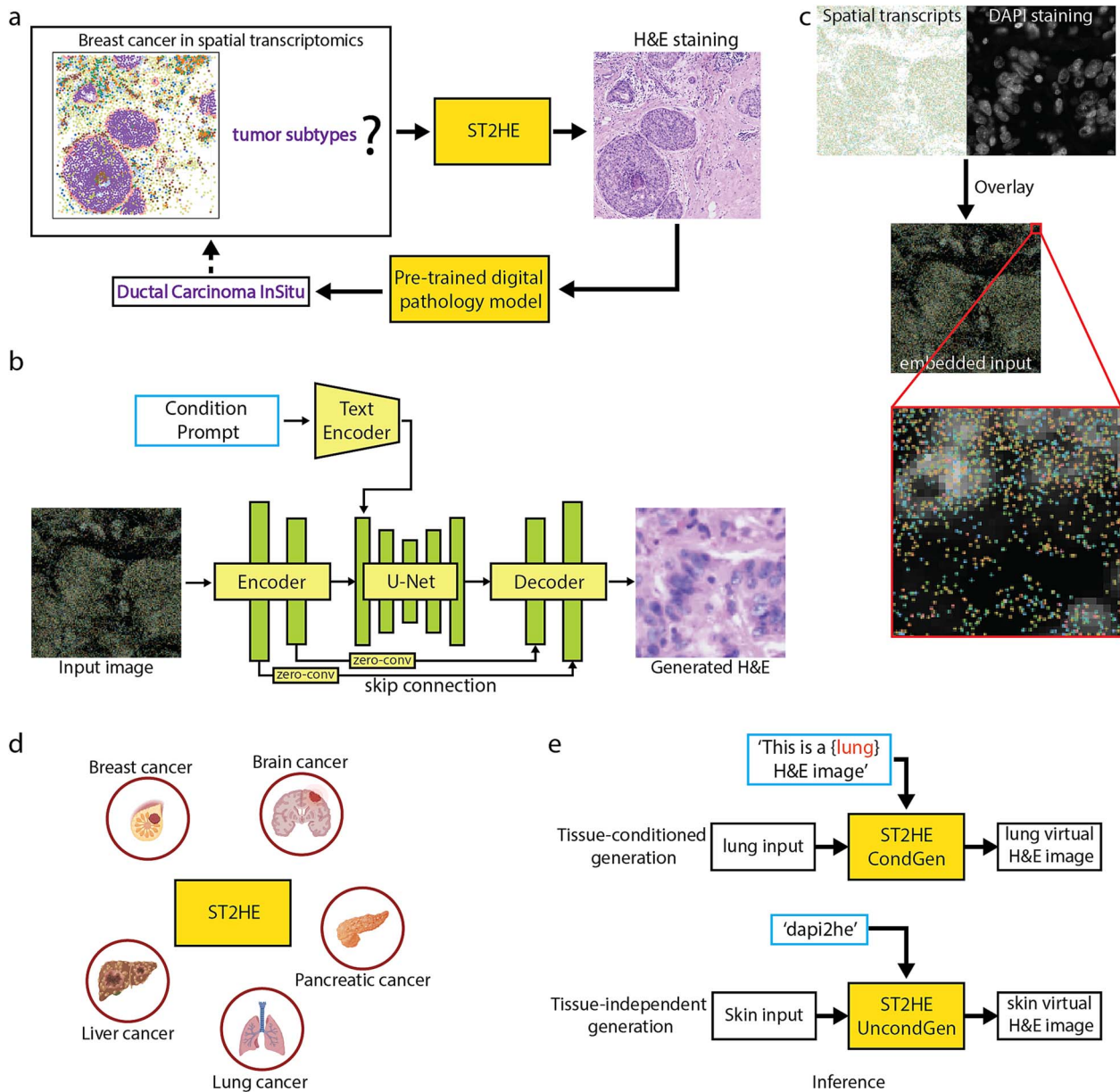


Figure 1 Overview of the ST2HE framework. (a) Challenge in annotating high-resolution spatial transcriptomics (HR-ST) data. Although HR-ST captures gene expression at single-cell resolution, tumor subtypes such as ductal carcinoma in situ (DCIS) remain difficult to assign directly from spatial profiles. In contrast, hematoxylin and eosin (H&E) staining is the clinical gold standard, supported by extensive digital pathology models. (b) ST2HE architecture for virtual H&E generation. Built on the Pix2Pix-Turbo backbone, the generator follows an encoder-U-net-decoder design. Text prompts are encoded via CLIP and injected to guide tissue-specific virtual staining, producing images that align transcriptomic inputs with histological features. (c) preparation of input tiles. DAPI images were sampled across tissues and overlaid with spatial transcript coordinates, represented as colored dots with randomized palettes to avoid fixed gene-color associations. These transcript-enhanced inputs form the basis for virtual H&E generation. (d) Training across multiple tissues. ST2HE was trained on xenium data from five tissue types, capturing diverse morphologies and gene panels to support cross-platform generalization. (e) Variants of ST2HE. The tissue-conditioned model (ST2HE-CondGen) generates virtual H&E guided by tissue-specific prompts (e.g. 'This is a lung H&E image'), while the tissue-independent model (ST2HE-UnCondGen) uses a generic prompt ('dapi2he') to enable generation for tissues not represented in training.

dataset includes paired, ground truth H&E staining, thus allowing evaluation of the generated H&E via pixel-wise similarity, structural comparison, and tissue-level assessment. We first assessed the generated image quality using mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). Across the dataset, ST2HE achieved an average MSE of 928.18, PSNR of 18.60 dB, and SSIM of 0.59. Stromal regions showed the highest structural similarity (SSIM = 0.60), followed by DCIS (SSIM = 0.58) and

invasive tumor regions (SSIM = 0.57), with all pairwise differences being statistically significant (Tukey's Honestly Significant Difference, $P < 0.05$). PSNR was also higher in stromal tissue compared with invasive tumor (PSNR = 18.67 versus 18.41 dB, $P = 0.0077$), while differences in MSE among tissue types were not significant ($P = 0.11$) (Fig. 2e). These results indicate that ST2HE best preserves structural features in stromal regions, with reduced performance in tumor areas, likely due to their complexity. We next evaluated sample-level

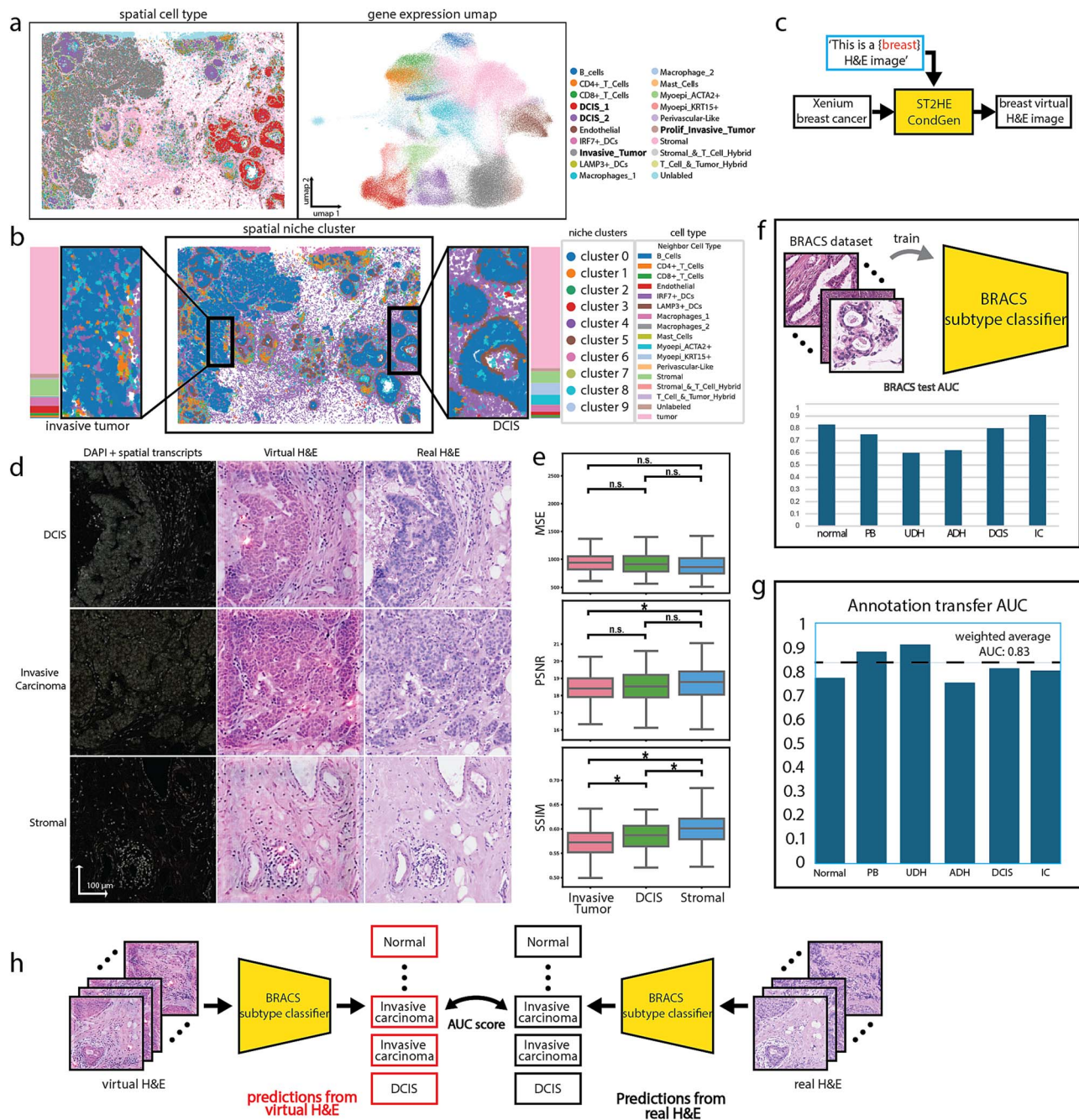


Figure 2 Application of ST2HE to breast cancer spatial transcriptomics. (a) Cell type visualization of a breast cancer sample. Left: Spatial cell type map; right: UMAP embedding of gene expression colored by annotated cell types. Tumor-associated populations are grouped into ductal carcinoma in situ (DCIS) and invasive carcinoma (invasive tumor), with additional immune and stromal populations highlighting tissue heterogeneity. (b) Spatial niche clustering based on neighbouring cell composition. Cells from DCIS and invasive carcinoma were assigned to the same cluster, showing that clustering without histopathology cannot distinguish these tumor subtypes. (c) Virtual H&E images of breast cancer tissue generated by ST2HE-CondGen, leveraging breast cancer as a training tissue type. (d) Structural fidelity of virtual H&E images. Left: DAPI with transcriptomic coordinates (input); middle: Virtual H&E generated by ST2HE; right: Real H&E. Nuclei positions, morphology, and tissue architecture are faithfully reconstructed. (e) Pixel-level similarity between virtual and real H&E images across tissue types. Boxplots summarize MSE, PSNR and SSIM. Differences among tissue types were assessed by ANOVA; significant comparisons are marked with * and non-significant with n.s. (f) Performance of a breast cancer pathology classifier on real H&E images from the BRACS dataset. Bar plots show AUC values for six diagnostic categories: Normal, pathological benign (PB), usual ductal hyperplasia (UDH), atypical ductal hyperplasia (ADH), DCIS, and invasive carcinoma. (g) Annotation transfer to virtual H&E images. Bar plots show AUC values for the same six categories, with performance comparable to real H&E, indicating preservation of pathological features. (h) Scheme of classification evaluation. A classifier fine-tuned on BRACS real H&E tiles was applied to virtual H&E images and predicted labels were compared with real H&E to assess annotation transfer.

perceptual similarity with the Fréchet Inception Distance (FID) and obtained a score of 83.1, indicating that the generated H&E capture key structural features. Unlike traditional virtual staining, our method relies solely on transcript locations and DAPI signals, making morphology reconstruction more challenging. Beyond quantitative metrics, we also analysed structural fidelity. Nucleus positions and shapes were well preserved, guided by the DAPI signals and cell morphology and cytoplasmic regions were faithfully reconstructed from ST signals. The spatial arrangement of colored transcript signals also helped the model learn morphology distributions, thus enhancing overall pathological patterns (Fig. 2d). Overall, ST2HE preserves key histological structures sufficient for downstream computational annotation tasks.

We then evaluated annotation capability by fine-tuning the CONCH's image encoder [13] using H&E tiles from the BRACS dataset [24], which contains annotations of multiple breast cancer subtypes. Representative hematoxylin and eosin tiles from the BRACS dataset used for training and validating the breast cancer subtype classifier are shown in Supplementary Fig. S1. The classifier achieved an AUC of 0.83 for normal, 0.75 for pathological benign, 0.6 for usual ductal hyperplasia, 0.62 for atypical ductal hyperplasia, 0.8 for DCIS, and 0.91 for invasive carcinoma (Fig. 2f). These results validate its use for annotation prediction. We next applied the classifier to virtual H&E images and compared the predicted labels with those obtained from real H&E images (Fig. 2h). The annotation transfer performance was quantified using AUC for each class, with the virtual H&E achieving 0.78 for normal, 0.89 for pathological benign, 0.92 for usual ductal hyperplasia, 0.76 for atypical ductal hyperplasia, 0.82 for DCIS and 0.81 for invasive carcinoma. With a weighted average AUC of 0.83 for real H&E images, this demonstrates strong agreement in predicted annotations (Fig. 2g), confirming ST2HE's utility for downstream breast cancer analysis. Spatial co-localization analysis further illustrates that CONCH-derived tumor heatmaps aligned with known cell-type distributions from the same tissue (Supplementary Fig. S1b).

ST2HE Enables Cross-Platform Virtual H&E Staining on NanoString CosMx data

To evaluate the cross-platform capability of our model, we tested ST2HE using the NanoString NSCLC dataset (see Methods). This dataset includes three representative lung samples (Lung5_rep1, Lung6, and Lung9_rep1) with HR-ST profiles annotated across multiple biological layers: cell types, niches, and tumor subtypes [25]. Tumor subtype annotations were performed by expert pathologists at the UPMC Hillman Cancer Center, using DAPI staining, spatial cell type distributions, and fluorescence-based imaging [26] (Fig. 3a). The primary challenge in this setting is that NanoString utilizes a gene panel distinct from Xenium plus the cross-platform batch noise, which is ideal for testing ST2HE's robustness across platforms. Additionally, since several tumor subtypes appear in a different sample, the model must preserve subtype-specific pathological features while mitigating sample-specific variability. The lack of paired ground truth H&E images in this dataset further complicates direct evaluation of the generated virtual H&E. Consequently, we assessed the generated image quality through downstream analyses and expert annotations.

Using the ST2HE-CondGen model (Fig. 3b), we generated virtual H&E images for the three NSCLC samples. The generated images exhibit morphological characteristics consistent with known NSCLC pathology features, as assessed by pathologist review of representative examples (Fig. 3c). For instance, the micropapillary subtype shows

epithelial clusters forming rounded papillary structures. All three NSCLC subtypes, i.e. micropapillary, solid, and complex acinar retain their distinguishing structural patterns, as confirmed by pathologist annotations [27] (Fig. 3c). To further assess the structural fidelity of the generated images, we embedded them using the CONCH model. Uniform Manifold Approximation and Projection (UMAP) visualization of CONCH embedding showed clear separation for micropapillary samples, while solid and complex acinar samples appeared more closely clustered but still visibly separated (Fig. 3d). This proximity may stem from batch effects, as both subtypes originate from the same lung sample (Lung9), possibly influencing their representation (Fig. 3d).

To annotate the generated images and compared the subtype predictions to the known annotations, we leveraged the full CONCH model, which employs both the image and text encoders in a contrastive framework. Unlike previous analyses where only image embeddings were used for visualization, here the model compares image embeddings with text embeddings of descriptive prompts (e.g. 'An image of {micropapillary tumor}') in a shared latent space and assigns the label with the closest match (Fig. 3e). Although initial classification results yielded AUC scores of 0.51 for complex acinar, 0.64 for solid, and 0.83 for micropapillary (Fig. 3f), UMAP visualizations have revealed that the generated H&E images preserved distinct subtype features. This discrepancy suggests a domain mismatch between the generated images and the data distribution used to train CONCH. To bridge this gap, we fine-tuned the CONCH model using additional generated H&E images from consecutive NanoString slides (see Methods). Representative virtual H&E tiles generated from consecutive NanoString NSCLC sections and used for fine-tuning the pathology classifier are shown in Supplementary Fig. S2. This realignment significantly improved classification performance, resulting in AUC scores of 0.72 for complex acinar, 0.74 for solid, and 0.93 for micropapillary (Fig. 3f). These improvements underscore the effectiveness of our synthetic H&E images in downstream pathology tasks.

To further evaluate the biological relevance of the generated H&E images, we used CONCH to identify tumor regions. We applied a sliding window approach (see Methods) to the classifier across image tiles, generating tumor subtype probability heatmaps from normalized prediction scores (Fig. 3g). These heatmaps showed strong alignment with ground truth annotations, providing additional validation. Notably, even after fine-tuning, the CONCH model continued to focus specifically on tumor regions, indicating that ST2HE learned generalized morphological patterns rather than overfitting to the generated images. Had the synthetic H&E images lacked biological fidelity, CONCH would not consistently attend to tumor regions. Thus, this result confirms both the biological integrity of the generated images and ST2HE's suitability for automated pathological phenotype detection. Overall, our findings demonstrate ST2HE's ability to generalize beyond its training technology platform, successfully producing virtual H&E images of NSCLC that support downstream pathological annotation. Furthermore, the improved performance achieved by fine-tuning the classifier with synthetic images further demonstrate the quality and biological relevance of the generated images, confirming ST2HE's utility in pathology annotations for HR-ST.

ST2HE generalizes for unseen tissue types in Kaposi's sarcoma

To evaluate the model's ability to generalize to an entirely unseen tissue type, we used the tissue-independent ST2HE-UnCondGen model

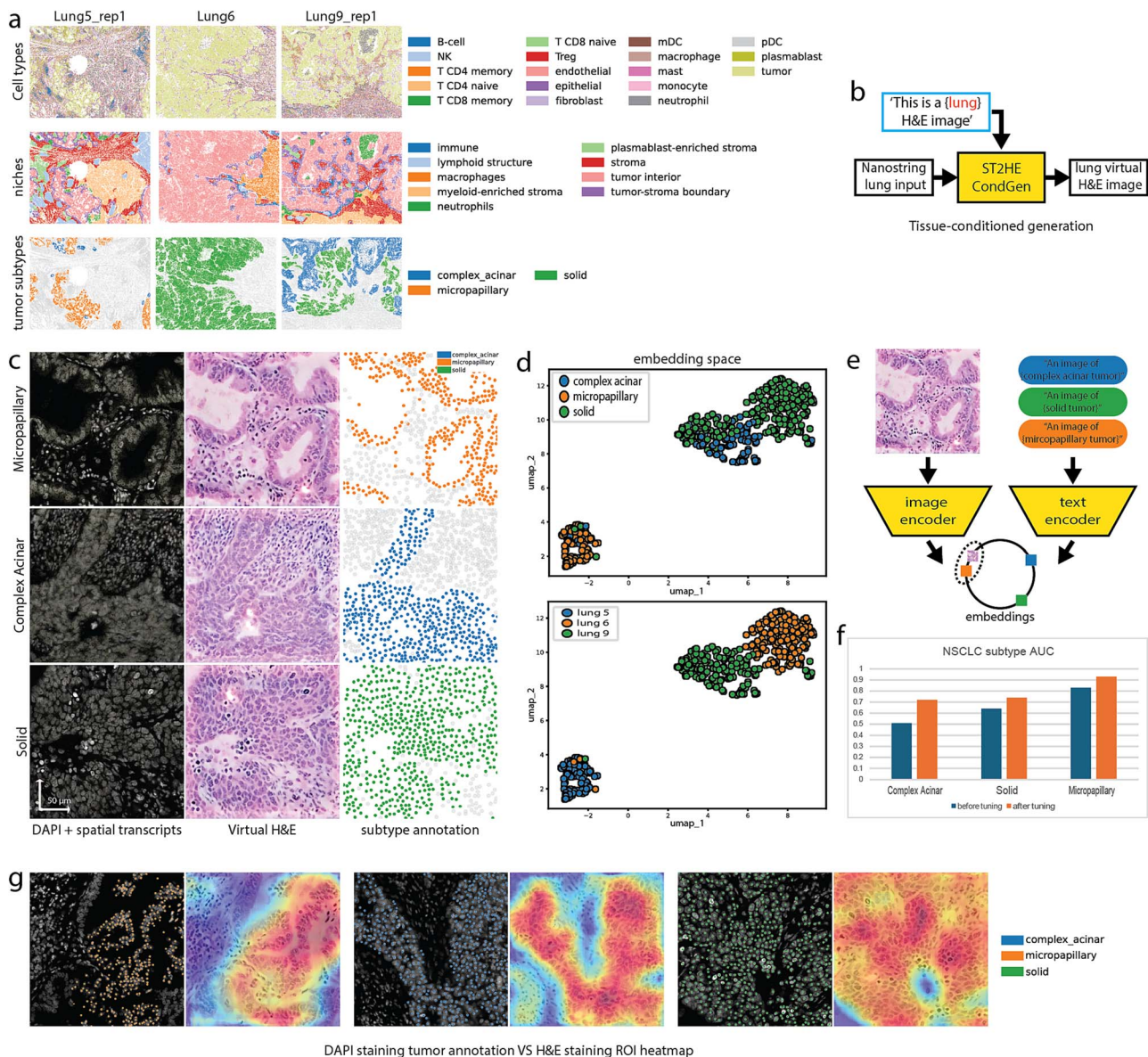


Figure 3 Cross-platform application of ST2HE to non-small cell lung cancer. (a) Spatial transcriptomics of three NanoString lung samples (Lung5_rep1, Lung6, Lung9_rep1). Cell types (top), spatial niches (middle), and tumor subtypes (bottom) were annotated by expert pathologists using DAPI and fluorescence images. Subtypes include complex acinar, micropapillary, and solid. (b) Tissue-conditioned generation of virtual H&E images using ST2HE-CondGen guided by lung-specific prompts. (c) Representative examples of three tumor subtypes. Left: DAPI with spatial transcripts; middle: Virtual H&E generated by ST2HE; right: Pathologist subtype annotations. Characteristic features such as papillary structures in micropapillary subtype are preserved. (d) UMAP embedding of CONCH image features from virtual H&E images. Top: Subtype clusters colored by tumor subtype (complex acinar, micropapillary, solid). Bottom: The same embedding colored by sample identity (Lung5, Lung6, Lung9). (e) Schematic of zero-shot classification using CONCH. Virtual H&E images are embedded by the image encoder, while class-specific prompts (e.g. 'An image of micropapillary tumor') are embedded by the text encoder. Both are projected into a shared contrastive space, and the image is assigned the label corresponding to the nearest text embedding. (f) NSCLC subtype classification performance. Initial AUC scores show variability across subtypes but fine-tuning CONCH with synthetic H&E images improves accuracy, particularly for complex acinar and solid subtypes. (g) Tumor region heatmaps from subtype probability scores. For each subtype, sliding-window predictions from the CONCH classifier are visualized as heatmaps overlaid on DAPI and virtual H&E images. Examples are shown for micropapillary, complex acinar, and solid subtypes, where predicted tumor regions align with ground truth annotations, confirming the biological relevance of ST2HE-generated H&E images.

trained on Xenium datasets containing skin tissues (Fig. 4a). With a generic prompt '{dapi2he}', we generated virtual H&E images for our in-house Xenium dataset of Kaposi's sarcoma (KS), a highly inflammatory tumor caused by infection with Kaposi's sarcoma-associated herpesvirus (KSHV) [28] (Fig. 4b). Whole-slide hematoxylin and eosin scans of the three Kaposi's sarcoma tissue microarrays used for evaluation are provided in Supplementary Fig. S3. The goal of this

experiment was to assess whether ST2HE could generate histologically meaningful images for KS, a cancer type absent from the training set, and capture features sufficient to distinguish between different KS stages.

Our in-house KS dataset includes three slides of tissue arrays, each comprising 16, 16, 11 tissues cores of cutaneous KS lesions, respectively, derived from different KS patients (Fig. 4b). Each core is

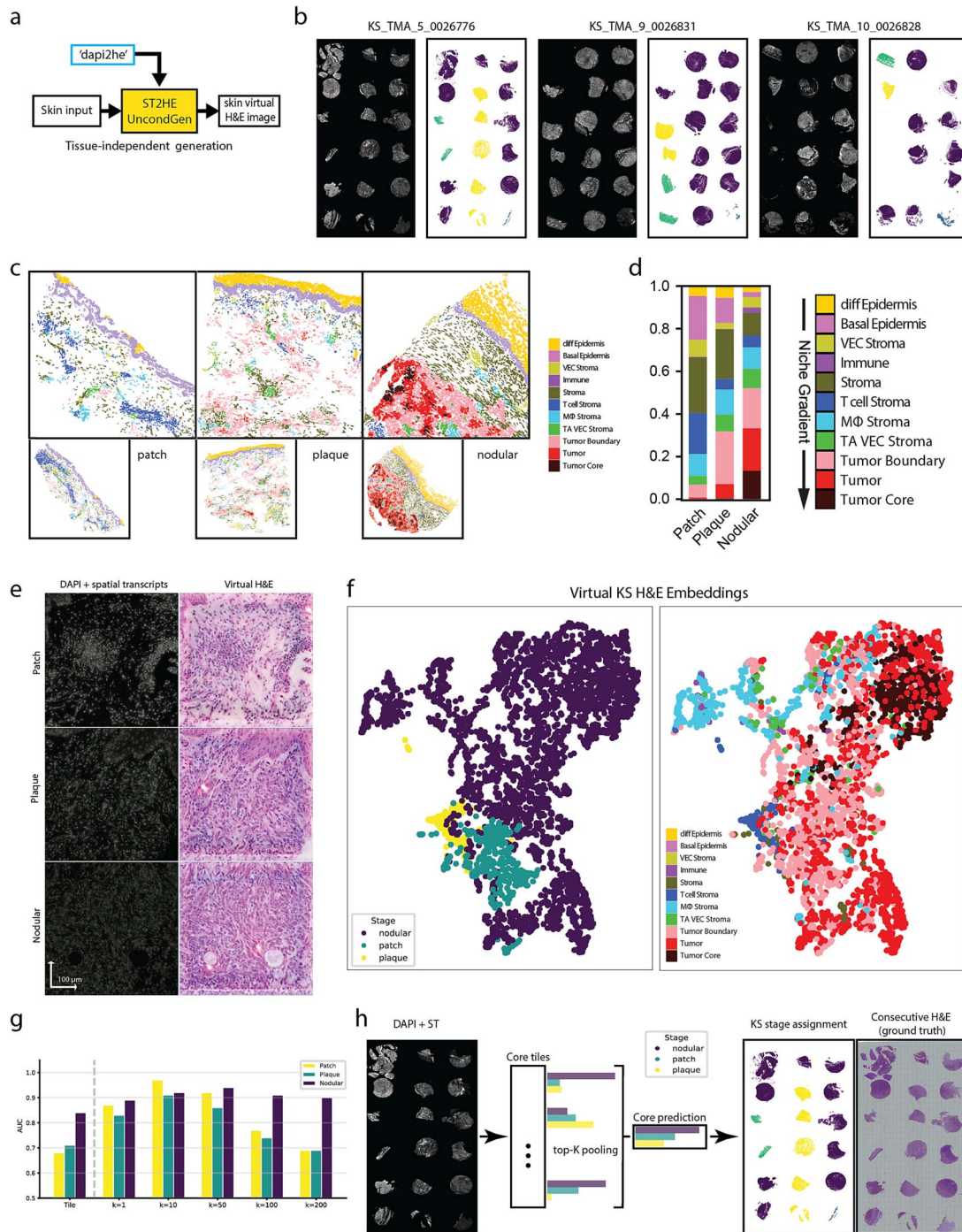


Figure 4 Generalization of ST2HE to Kaposi's sarcoma, an unseen tissue type. (a) Tissue-independent generation with ST2HE-UnCondGen. A generic prompt ('dapi2he') was used to create virtual H&E images of KS, a cancer absent from the training set. (b) Representative DAPI inputs (left) and corresponding virtual H&E images (right) generated by ST2HE-UnCondGen from three tissue microarray slides (KS_TMA_5, KS_TMA_9, KS_TMA_10), comprising 16, 16, and 11 tissue cores respectively. Each core is outlined and colored by KS stage (Patch, Plaque, or Nodular). (c) Spatial niche composition across KS stages. Eleven distinct niches, including epidermal, immune, stromal, and tumor-associated regions, were identified from spatial transcriptomics data. (d) Niche composition gradient across KS progression, showing a shift from epidermal/stromal niches in early patch lesions toward tumor and tumor core niches in nodular lesions. (e) Generated virtual H&E images reflect stage-specific morphology. Patch lesions show sparse spindle cells, plaque lesions exhibit increased density with irregular clustering, and nodular lesions display tightly packed spindle cells forming consolidated structures. (f) UMAP embeddings of virtual H&E tiles. Left: Clustering by KS stage. Right: Clustering by spatial niche, showing heterogeneity in nodular lesions and enrichment for tumor-associated niches. (g) KS stage classification performance. Bar plots show AUC for patch, plaque, and nodular stages. Tile-level prediction performance is moderate but core-level performance improves markedly with top-K pooling. (h) Schematic of the top-K pooling strategy for core-level classification. Tile-level predictions from virtual H&E images are ranked by confidence and the most informative tiles are aggregated to generate a core-level prediction. This approach improves stage classification performance compared with tile-level evaluation.

annotated with one of three clinically defined KS stages (Patch, Plaque, or Nodular) based on the H&E stains from adjacent slides [29]. The dataset captures rich spatial transcriptomics features. Using spatial neighboring cell type composition analysis, we defined eleven distinct tissue niches including differentiated and basal epidermis, immune niche, various stromal compartments such as vascular endothelial cells (VEC), transit-amplifying VEC, macrophages ($M\Phi$), and T-cell stroma, and tumor niches including tumor boundary, tumor, and tumor core (Fig. 4c). Analysis of tissues across KS stages revealed a progression associated shift in niche composition, transitioning from predominantly epidermal and stromal niches toward tumor-associated niches as KS advances from Patch to Nodular stage (Fig. 4d). Notably, Nodular cores also contain niches of Patch and Plaque stages, indicating pathology patterns in later stages overlap with those in earlier stages (Fig. 4c&d). This overlap may complicate the ability of a predictive model to accurately infer the KS stage.

We first examined the generated H&E images of the cores from these slides. Interestingly, visual inspection revealed that ST2HE captured general morphological differences across KS stages (Fig. 4e). Particularly, in the Patch cores, we observed sparsely distributed spindle cells, whereas the Plaque cores showed moderately increased spindle cell density with more irregular clustering and the Nodular cores exhibited highly cellular regions with tightly packed spindle cells forming more consolidated structures (Fig. 4e). These observations align with pathological characteristics of KS progression [29–31]. Although the ST2HE never encountered skin tissue during training, the generated images appeared to reflect structurally meaningful differences across disease stages.

To quantitatively evaluate these features, we embedded the generated H&E tiles using the CONCH image encoder and visualized the resulting latent space using UMAP (Fig. 4f). The embeddings revealed stage-associated clustering, with the Nodular cores forming a dominant, compact region and the Patch cores being separated more distinctly from the rest. The Plaque stage exhibits partial overlap with both (Fig. 4f, left panel). When colored by spatial niche, early-stage Patch regions predominantly aligned with the expected early-stage niches, showing relatively homogeneous composition. In contrast, late-stage Nodular regions exhibited extensive heterogeneity, encompassing nearly all spatial niches, while being heavily enriched for late-stage niches such as tumor core and tumor boundary (Fig. 4f, right panel). These patterns suggest that the model captured biologically meaningful differences reflective of KS stage and tumor architecture, despite it has never seen any KS data during training. Correlation of spindle packing scores with angiogenesis and inflammation gene modules confirmed a significant morphology-molecular relationship across KS stages (Supplementary Fig. S3b, S3c). The spindle packing score is a tissue-level architectural measure derived from Frangi vesselness filtering of virtual H&E images, where higher scores reflect progressive spindle cell consolidation from sparse structures in Patch to dense masses in Nodular stage. Together with the breast cancer co-localization analysis (Supplementary Fig. S1b), these results support tissue-level biological consistency of the generated images, though they do not establish single-cell-level structural fidelity.

We next tested whether these virtual H&E images could support downstream classification tasks. A stage-specific classifier trained on real H&E tiles from consecutive sections was applied to the generated virtual H&E tiles. As expected, tile-level classification performance was moderate, with AUC scores of 0.68 for Patch, 0.71 for Plaque, and 0.84 for Nodular (Fig. 4g). These suboptimal tile-level results are

expected for two reasons. First, the model was trained without any skin tissue, making it challenging to generate accurate H&E features for skin-derived KS samples. Second, and more fundamentally, stage annotations in our dataset are assigned at the core level, meaning all tiles from a given core carry the same stage label regardless of their local biological content. As shown in Fig. 4d, KS cores contain heterogeneous mixtures of niches where late-stage Nodular cores frequently include early-stage epidermal and stromal niches whose tiles are incorrectly labeled as Nodular at the tile level. This inherent label noise naturally limits tile-level classification performance. To address this, we adopted a top-K pooling strategy (Fig. 4h), which aggregates predictions from the most confidently predicted tiles within each core. Since tiles representing the dominant stage morphology of the core tend to receive higher confidence scores, top-K pooling effectively filters out mislabeled tiles from mixed-stage niches. We systematically evaluated performance across a range of K values ($K = 1, 10, 50, 100, 200$), with results presented in Fig. 4g. Performance peaks at small to medium K values and degrades at larger K as tiles from mixed-stage niches are increasingly included, consistent with the niche heterogeneity (Fig. 4d). These results demonstrate that while individual tiles may be noisier, ST2HE preserves essential morphological cues across multiple tiles to enable accurate core-level inference. In summary, these results underscore ST2HE's capacity to generalize to previously unseen tissue types. Even though the model was trained without any skin data, it was able to generate virtual H&E images for KS that reflect relevant biological patterns and support downstream classification. These findings indicate that spatial transcriptomics features learned from other tissues can encode broadly applicable histological signals, demonstrating the potential of virtual staining for use in diverse and novel tissue settings.

Larger context, balanced loss, and multi-colored transcripts improve virtual H&E generation

A core innovation of ST2HE is integrating DAPI staining with transcript locations to generate virtual H&E images. To assess how input configurations affect output quality, we conducted ablation studies using the Xenium breast cancer data. We first examined the tile size, which determines local context. Models trained with 512×512 tiles outperformed those with 256×256 (AUC of 0.81 versus 0.73; Table 1), demonstrating that larger context improves morphological fidelity and classification accuracy. Even larger tiles may offer further improvements, but memory constraints prevented us from exploring them in this study.

Next, we evaluated ST2HE's loss weighting, which balances contributions from global-context features and local structural details. Experiments showed that models prioritizing detailed morphology consistently performed better across tasks. Specifically, when the global-context loss was assigned to a higher weight (0.3, denoted as *hg*), performance declined compared to the default setting (0.1), with an average AUC of 0.72 versus 0.80 (Table 1). These results suggest that preserving fine-grained structural details is more critical for reliable interpretation and annotation than emphasizing global style.

We further compared conditional and non-conditional training across both breast cancer and NSCLC datasets. For breast cancer, ST2HE-CondGen slightly outperformed ST2HE-UnCondGen for downstream subtype classification (AUC 0.83 versus 0.81; Table 1).

Table 1 Classification AUC performance of different ST2HE model variants across histopathological categories.

	ST_256_hg	MT_256	MT_256_hg	ST_512	MT_512	MT_512_cond
Normal	0.55	0.7	0.69	0.89	0.89	0.78
PB	0.73	0.77	0.81	0.91	0.90	0.89
UDH	0.75	0.86	0.80	0.92	0.91	0.92
ADH	0.83	0.82	0.80	0.78	0.85	0.76
DCIS	0.79	0.86	0.81	0.79	0.73	0.82
IC	0.63	0.74	0.72	0.76	0.80	0.81
avg	0.67	0.77	0.76	0.80	0.81	0.83

Values represent classification AUC for each subtype; bold values indicate the highest AUC in each row (the best-performing variant for that category). ‘ST’ and ‘MT’ indicate single-tissue and multi-tissue training, respectively, the subscripts (256, 512) denote tile size, ‘_hg’ indicates high weight of global context loss, and ‘_cond’ indicates the use of conditional generation.

This advantage was consistent in the NSCLC dataset, where ST2HE-CondGen achieved AUC scores of 0.51, 0.64, and 0.83 for complex acinar, solid, and micropapillary subtypes respectively, compared to 0.53, 0.59, and 0.80 for ST2HE-UnCondGen (Supplementary Fig. S2). Across both datasets, tissue conditioning provides consistent if modest performance gains, suggesting that the tissue prompt helps refine morphological generation when label information is available. However, ST2HE-CondGen depends on the availability of tissue-type labels, which are not always practical to obtain. To balance accuracy and flexibility, we retained both versions in our pipeline, applying ST2HE-CondGen when labels are available and ST2HE-UnCondGen when they are not.

An important design choice in ST2HE is transcript dot size, controlled by parameter *s* in the Matplotlib scatter function. We systematically evaluated *s* from 0.1 to 0.9, measuring LPIPS, L2, CLIP similarity, and clean FID. L2 and CLIP similarity peaked at *s* = 0.1, clean FID favored *s* = 0.3, and classification AUC peaked at *s* = 0.5 (Fig. 5a-c). Larger dot sizes caused transcript overlap and blurrier morphology. Balancing these outcomes, we selected *s* = 0.3 as the optimal dot size.

We compared three visualization strategies: multi-colored transcripts (each gene assigned a unique color), mono-colored transcripts (all genes shown in a single color), and a DAPI-only input with no transcript visualization. Multi-colored transcripts consistently performed the best (AUC of 0.83 versus 0.75 for mono-colored and 0.64 for DAPI-only inputs; Fig. 5d). This result indicates the multi-colored approach provides a rich visual cue that helps the model distinguish spatial gene expression patterns and associate them with morphological context. In contrast, mono-colored inputs led the model to confuse transcript dots with DAPI nuclear signals, producing scattered, small cell-like artifacts rather than coherent tissue morphology (Fig. 5e). DAPI-only inputs lacked transcript context and produced poor image quality (Fig. 5e). These findings support multi-colored transcript visualization as the default configuration.

Finally, we evaluated the robustness of ST2HE under varying input conditions that reflect real-world challenges in cross-platform and cross-tissue applications, including differences in tissue complexity, gene panel composition, and transcript positional uncertainty. These analyses were conducted on the Xenium breast cancer dataset where ground truth H&E is available for quantitative assessment, and provide indirect support for cross-platform generalization. Notably, downstream classification AUC remained stable across 25%–100% gene retention, only dropping when transcripts were removed entirely, while pixel-wise metrics paradoxically improved with fewer genes. We discuss the implications of this finding in the

Discussion. We note that additional factors including DAPI image quality, segmentation differences, platform chemistry, and residual gene-panel effects may still influence cross-platform performance and represent important directions for future investigation. Results are presented in Supplementary Fig. S4.

Discussion

We developed ST2HE, a generative framework that synthesizes histologically realistic virtual H&E images directly from HR-ST data. By leveraging DAPI morphology and gene-level spatial context, ST2HE bridges the gap between transcriptomics and histopathology. Across datasets including breast cancer, NSCLC, and KS, generated images preserved key morphological features and supported downstream tasks such as subtype classification and stage prediction, demonstrating the feasibility of virtual H&E for spatial biology and computational pathology.

A major strength of ST2HE is its flexibility across platforms and tissues, enabled by dual conditional and generalization modes. In cross-platform and unseen-tissue evaluations, ST2HE produced images capturing tissue-level morphological patterns while retaining predictive accuracy with pretrained models like CONCH, underscoring the fidelity of integrating transcriptional signals with generative modeling.

The practical utility of ST2HE is greatest in scenarios where high-quality paired H&E is unavailable or impractical to obtain. We propose the following use-case framework: (i) When tissue quantity is limited and researchers must prioritize between H&E and immunofluorescence, ST2HE removes this constraint by generating virtual H&E directly from HR-ST data, freeing remaining tissue for immunofluorescence or other downstream assays. (ii) When H&E is missing or of low quality, ST2HE provides a direct substitute, enabling histological annotation from HR-ST data alone without additional tissue sectioning or staining. (iii) When the HR-ST platform does not support same-slide H&E (e.g. NanoString CosMx, MERFISH), ST2HE enables histological interpretation directly from HR-ST data, bypassing the need for adjacent section registration and its associated biological variation. (iv) When large-scale cohort processing is required, ST2HE eliminates the need for manual image registration between H&E and HR-ST modalities, enabling scalable automated histological annotation across large datasets.

ST2HE also provides interpretability and modularity. Ablation studies showed that smaller transcript dots and multi-colored encodings improved structural extraction and preserved spatial context. Overlaying transcript locations on DAPI images enhanced interpretability,

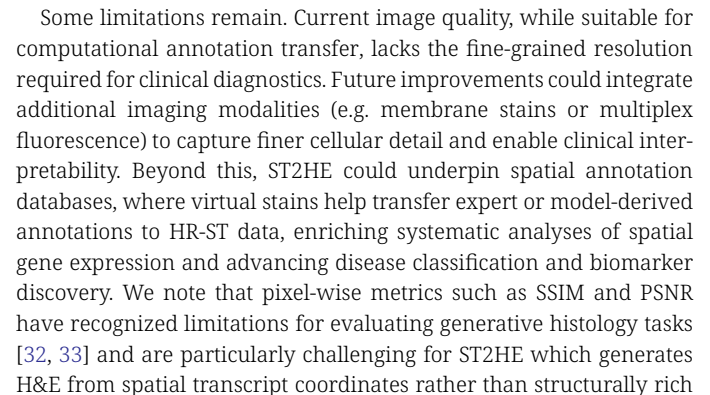


Figure 5 Ablation study of context size, training loss, and transcript color scheme on virtual H&E generation. (a) Validation performance of different transcript dot sizes ($s = 0.1\text{--}0.9$) using LPIPS, L2, CLIP similarity, and clean FID, capturing perceptual, pixel-level, semantic, and realism-based quality. (b) Weighted average ROC AUC across dot sizes. Models performed best at intermediate dot sizes ($s = 0.3\text{--}0.7$). (c) Representative virtual H&E images generated using different dot sizes. Larger dot sizes caused transcript overlap and blurrier morphology, while smaller sizes preserved nuclear features but reduced perceptual realism. Real H&E is shown for comparison. (d) Weighted average classification AUC for different transcript visualization strategies. Multi-colored transcript inputs outperformed mono-colored and DAPI-only inputs, demonstrating the importance of gene-specific color information. (e) Representative outputs under different input settings. Real H&E (left) is compared with virtual H&E generated by multi-colored transcripts, mono-colored transcripts, and DAPI-only inputs. Black boxes highlight regions where multi-colored inputs preserve structural fidelity and pathology-relevant features better than the alternatives.

imaging modalities. Downstream annotation transfer AUC of 0.83 is therefore the more relevant quality criterion for ST2HE's intended use. The contribution of DAPI image quality differences between platforms to cross-platform performance remains an open question. Unlike transcript density and panel composition, which can be systematically evaluated through controlled experiments, disentangling DAPI quality effects would require controlled cross-platform imaging experiments that are beyond the scope of the current study. We identify this as an important direction for future work in defining the generalization boundary of ST2HE.

Materials and methods

Data preparation

Raw spatial transcriptomics data and their corresponding tissue images were initially curated and processed to create suitable inputs for model training. The workflow involved three key steps: (i) aligning DAPI and H&E images through image registration, (ii) filtering and overlaying transcript coordinates onto DAPI images, and (iii) generating transcript-enhanced DAPI/H&E patch pairs. These processed pairs are the input and output for model training. We utilized five Xenium formalin-fixed paraffin-embedded (FFPE) datasets from 10× Genomics: (i) preview FFPE Human Breast Sample, (ii) Human Lung Cancer with Human Immuno-Oncology Profiling Panel and Custom Add-on, (iii) FFPE Human Pancreatic Ductal Adenocarcinoma with Human Immuno-Oncology Profiling Panel, (iv) FFPE Human Brain Cancer with Human Immuno-Oncology Profiling Panel and Custom Add-on, and (v) Human Liver with the Xenium Human Multi-Tissue and Cancer Panel, as Xenium provides both high-resolution DAPI images with transcript coordinates and matched post-Xenium H&E images. This pairing makes them uniquely suited for supervised learning, since ground-truth stained images are available for direct comparison.

Image registration

Xenium DAPI and H&E images were aligned using a 3×3 transformation matrix provided with each dataset, resampled to $0.2125 \mu\text{m}/\text{pixel}$ via bicubic interpolation, and warped using linear perspective transformation to achieve pixel-level correspondence.

Model training data generation

To prepare image tiles for model training, we first selected representative regions across the tissue. Farthest point sampling was applied to segmented cell centroids to identify spatially diverse index cells. For each selected cell, a 512×512 pixel window was extracted from the DAPI image and the corresponding post-Xenium H&E image, provided the tile contained sufficient content. Tiles with over 90% black or near-white pixels (grayscale 0 or > 230) were discarded. DAPI tiles were normalized and saved as grayscale PNGs, and H&E tiles were saved in RGB format. Transcriptomic coordinates with quality scores above 20 were retained for overlay. Transcripts were plotted on the DAPI background as colored dots, with each gene assigned a unique color from a hue-sorted palette. The palette was independently randomized per tile to prevent color-gene correspondence across tiles, and dot sizes were varied to improve model robustness. Axes, labels, and spines were removed to produce clean overlays. The resulting transcript-enhanced DAPI images, comprising 2000 tiles per tissue type, were

exported and split into training and validation sets for model development and evaluation.

ST2HE model for virtual H&E staining

ST2HE was built based on the Pix2Pix-Turbo model, a one-step image-to-image translation model designed for paired image synthesis. Pix2Pix-Turbo builds upon the traditional Pix2Pix framework by integrating a single-step diffusion-based generator with adversarial learning objectives, significantly reducing inference time while maintaining image quality. Additionally, it incorporated text-based conditional control to guide the generation of different tissue types, utilizing a CLIP-based text-image alignment loss.

The ST2HE generator is based on a pre-trained one-step diffusion model, which has been adapted for paired image-to-image translation. It employs Low-Rank Adaptation to efficiently fine-tune the model for virtual staining while reducing computational overhead. The network structure consolidates encoder, U-Net, and decoder components into an end-to-end trainable model with skip connections via zero-convolution layers to preserve fine image details (Fig. 1c). Text embeddings were processed through a CLIP encoder and integrated into the generator to modulate the output according to specified tissue types. The ST2HE discriminator is based on CLIP backbone, following the Vision-Aided GAN approach [34]. Since CLIP provides both text and image embeddings, the image embeddings are used for adversarial training, ensuring that the generated images align semantically with real stained images. This enables the model to enforce realism beyond tile-based discrimination, making it more suitable for text-guided virtual staining. We note that ST2HE does not introduce modifications to the Pix2Pix-Turbo architecture or loss functions. Its contributions are methodological, centered on a principled framework for bridging spatial transcriptomics and computational pathology through novel input encoding and training strategies tailored to HR-ST data.

Mathematically, the loss function of the Pix2Pix-Turbo model consists of three terms, the adversarial loss in equation (1), the reconstruction loss in equation (2), and the CLIP loss in equation (3):

Adversarial loss

$$\mathcal{L}_{GAN} = E_y [\log D(E_{img}(y))] + E_x [\log (1 - D(E_{img}(G(x, t))))] \quad (1)$$

where G is the generator, D is the discriminator operating on CLIP image embeddings $E_{img}(\cdot)$, x is the input unstained image, y is the corresponding H&E-stained image, and t is the text conditioning input.

Reconstruction loss

$$\mathcal{L}_{rec} = E_{x,y} [|y - G(x, t)|_2] + LPIPS(G(x, t), y) \quad (2)$$

This term ensures that the generated images remain structurally similar to the ground truth using both L2 and LPIPS perceptual loss.

CLIP loss

$$\mathcal{L}_{CLIP} = -E_{x,t} [\text{cosine_similarity}(E_{img}(G(x, t)), E_{txt}(t))] \quad (3)$$

where $E_{txt}(\cdot)$ represents the CLIP text embeddings, ensuring semantic alignment between the generated image and the specified tissue type.

The total loss function is given by:

$$G^* = \arg \min_G \max_D \mathcal{L}_{rec} + \lambda_{GAN} \mathcal{L}_{GAN} + \lambda_{CLIP} \mathcal{L}_{CLIP} \quad (4)$$

where λ_{GAN} and λ_{CLIP} are weighting factors that balance adversarial learning, pixel-level reconstruction, and text-image alignment. The detailed architecture of the ST2HE generator, conditional control, discriminator, and loss components is provided in [Supplementary Table S1](#).

Model training

Training the ST2HE model leverages paired DAPI–H&E patches and, optionally, tissue-specific prompts to guide the model. The network was trained to minimize the loss function equation (4) so that it preserved fine-grained morphological features, maintained global tissue context, and ensured semantic consistency with the target tissue type.

We also implemented a JSON configuration system to manage training prompts. In the conditional setting, each training input was paired with a tissue-specific prompt (e.g. ‘This is a {tissue} H&E image’). For the unconditional setting, all samples shared a generic prompt (“dapi2he”). These text prompts were embedded via CLIP and passed to the model for condition generation on the desired tissue type.

Additional implementation details, training environment specifications, and optimization settings are described in [Supplementary Note 1](#).

Downstream analysis

To evaluate the utility of the generated virtual H&E tiles, we assessed their suitability for downstream histopathological analyses across three distinct datasets representing different tissue types and pathological contexts. Specifically, we tested whether the generated images could support tumor subtype annotations using pathology foundation models.

Pathology Foundation Model

In this work, we leveraged the CONCH (Contrastively Aligned Visual-Language Pretraining) foundation model to enable a broad spectrum of downstream histopathological analyses. CONCH, pretrained on extensive H&E slide datasets with paired textual annotations, learns generalizable visual-language representations that are well-suited for interpreting diverse pathological features. Its rich embeddings capture essential aspects of tissue architecture, cellular morphology, and disease-specific cues, making it a powerful starting point for diagnostic modeling.

Breast Cancer Subtype Classification

Generated H&E tiles were first produced from DAPI images overlaid with spatial transcript coordinates, following the same input preparation used during model training. These virtual H&E tiles served as input to the downstream classification workflow. For subtype classification, the CONCH image encoder was fine-tuned using the BRACS dataset, and a single fully connected layer was appended to the encoder output to predict diagnostic subtypes, including benign, atypical hyperplasia, DCIS, and invasive carcinoma. This approach allows the model to leverage the fine-grained morphological information captured in the generated H&E tiles for accurate subtype annotation.

NSCLC Subtype Classification and Tumor Segmentation

Virtual H&E tiles were generated from DAPI images with spatial transcript overlays using the same input preparation as in training. The NanoString NSCLC dataset provides multiple consecutive DAPI and transcript stacks per sample. For testing, we used the center

stack, while the surrounding consecutive stacks were used to generate virtual H&E images for model fine-tuning. Pathologist-provided subtype annotations were embedded as textual prompts, and the CONCH image encoder was fine-tuned on the generated H&E tiles paired with these text prompts. Subtype classification for the test stack was then performed by computing the similarity between generated H&E embeddings and candidate textual prompts representing each NSCLC subtype.

To visualize tumor localization, we applied a 64×64 kernel sliding over the generated H&E images with a stride of 16 pixels. Each kernel patch was scored by its embedding similarity to the tumor text prompt. The resulting scores were stacked and averaged across overlapping kernels to produce a normalized heatmap, which was then smoothed using a Gaussian filter to generate a continuous representation of tumor probability.

KS stage tile and core classification

for the KS dataset, lesions were classified into Patch, Plaque, and Nodular stages, reflecting progressive angio-proliferative changes. Real H&E images from three consecutive slides per case, paired with pathologist annotations, were used to fine-tune the CONCH image encoder with an appended fully connected classification layer. For tile-level predictions, each H&E tile was classified individually, capturing local morphological features such as vascular patterns and spindle cell density. For core-level predictions, we aggregated tile-level scores using top-K pooling, assigning the KS stage based on the most confident tiles within a core. To evaluate the robustness of this approach, we systematically varied K across a range of values (K = 1, 10, 50, 100, 200) and report core-level AUC for each K value in [Fig. 4g](#).

Evaluation data

To assess the generalization and effectiveness of our approach, we curated evaluation datasets spanning three distinct pathological contexts.

Breast Cancer

Evaluation was conducted on the Preview FFPE Human Breast Sample 1, a Stage IIB HER2-positive breast cancer tissue containing both DCIS and invasive carcinoma. The sample was analysed using the Xenium In Situ platform on a $5 \mu\text{m}$ FFPE section and co-registered with post-Xenium H&E and immunofluorescence (IF) images to support spatially resolved tumor and microenvironment analysis. Cell type annotations for this dataset were adopted directly from Janesick et al. [23], and were not generated in this study. Analyses relying on these labels therefore inherit the assumptions and limitations of the original annotation framework.

Non-Small Cell Lung Cancer

Evaluation utilized NanoString slides lung5, lung6, and lung9, which represent multiple NSCLC tumor subtypes. Detailed pathological review identified complex acinar, micropapillary, and solid subtypes. Subtype annotations were provided with the assistance of a pathologist. These samples enabled robust testing of the model’s ability to capture morphological diversity and spatial transcriptomic variation within lung tumors.

Kaposi's Sarcoma

Evaluation was performed using an in-house dataset generated with the Xenium platform (10× Genomics) on three slides of tissue microarrays, each comprising of 16, 16, 11 tissues cores of cutaneous KS lesions annotated with corresponding KS stages (Patch, Plaque, and Nodular). The dataset was generated using the Xenium Human Skin Gene Expression Panel Kit and a custom designed panel including a total of 306 genes targeting key major skin and immune cell markers, inflammatory and angiogenic regulators, tumor-associated genes, and representative KSHV latent and lytic genes.

Supplementary materials include extended methodological details, architecture specifications, and additional figures supporting the analyses presented in this study.

Key Points

- Introduces ST2HE, a generative framework for interpreting spatial transcriptomics via virtual histology
- Enables platform- and tissue-agnostic analysis of high-resolution spatial transcriptomics data
- Bridges molecular and morphological information to support histological annotation
- Facilitates downstream tissue and phenotype analysis using computational pathology tools

Acknowledgements

This research was supported in part by the University of Pittsburgh Center for Research Computing and Data, RRID:SCR_022735. Specifically, this work used the HTC cluster, which is supported by S10OD028483.

Author contributions

Z.L. conceived the study, developed the ST2HE framework, performed data acquisition, conducted analyses and visualization, and wrote the original draft. A.D. contributed to method development and data acquisition. W.M. performed data acquisition. Y.-C.C. assisted with result analysis and interpretation, contributed to writing (review and editing) and provided funding and resources. S.-J.G. supervised the data acquisition, contributed to writing (review and editing), and provided funding and resources. Y.H. conceived and supervised the study, contributed to method development, writing (original draft, review and editing), and provided funding and resources. All authors approved the final version of the manuscript.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflict of interest

The authors declare no competing interests.

Funding

This study was supported by grants from the National Institutes of Health (U01CA279618 and R21GM155774 to Y. Huang; CA096512,

CA284554, CA278812, CA291244 and CA124332 to S.-J. Gao; and R35GM154967 to Y.-C. Chiu) and UPMC Hillman Cancer Center Startup Funds to S.-J. Gao and Y. Huang, and in part by award P30CA047904.

Data availability

The Xenium datasets for training and evaluating the ST2HE model are available from the Xenium data resource webpage <https://www.10xgenomics.com/datasets>, the NanoString CosMx SMI NSCLC data are available from <https://nanosttring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/nsclc-ffpe-dataset/>, and the breast carcinoma H&E dataset BRACS is available from <https://www.bracs.icar.cnr.it/>. Xenium KS data is deposited to Zenodo from <https://zenodo.org/records/17611373>.

Code availability

All custom code, including the ST2HE model weights and scripts for input data preparation, is publicly available at <https://github.com/Huang-AI4Medicine-Lab/ST2HE>. A DOI-minted archival release is available at <https://zenodo.org/records/20413722>. The repository is fully documented and includes scripts reproducing all analyses and figure panels. Package dependencies for training and inference will be provided to ensure exact reproducibility.

References

1. 10x Genomics. Xenium In Situ: High-Resolution Subcellular Spatial Gene Expression. Pleasanton, CA, USA: 10x Genomics, 2025. <https://www.10xgenomics.com/products/xenium-in-situ>
2. NanoString Technologies. High-Plex Spatial Profiling with the CosMx Spatial Molecular Imager. Seattle, WA, USA: NanoString Technologies, Inc., 2025. <https://www.nanostring.com/wp-content/uploads/2022/10/MK2318-CosMx-SMI-White-Paper.pdf>
3. Moffitt JR, Bambah-Mukku D, Eichhorn SW *et al*. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 2018;**362**:eaau5324. <https://doi.org/10.1126/science.aau5324>
4. Hu J, Li X, Coleman K *et al*. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* 2021;**18**:1342–51. <https://doi.org/10.1038/s41592-021-01255-8>
5. 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. <https://doi.org/10.1038/s41587-022-01483-z>
6. Zhao E, Stone MR, Ren X *et al*. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* 2021;**39**:1375–84. <https://doi.org/10.1038/s41587-021-00935-2>
7. Goltsev Y, Samusik N, Kennedy-Darling J *et al*. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* 2018;**174**:968–981.e15. <https://doi.org/10.1016/j.cell.2018.07.010>
8. Kiernan JA. Histological and Histochemical Methods: Theory and Practice. Bloxham, UK: Scion, 2015.
9. Guo JL, Lopez DM, Mascharak S *et al*. Hematoxylin and eosin architecture uncovers clinically divergent niches in pancreatic cancer. *Tissue Eng Part A* 2024;**30**:605–13. <https://doi.org/10.1089/ten.tea.2024.0039>
10. Coudray N, Ocampo PS, Sakellaropoulos T *et al*. Classification and mutation prediction from non-small cell lung cancer histopathology

- images using deep learning. *Nat Med* 2018;**24**:1559–67. <https://doi.org/10.1038/s41591-018-0177-5>
11. Lu MY, Williamson DFK, Chen TY *et al*. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021;**5**:555–70. <https://doi.org/10.1038/s41551-020-00682-w>
 12. Zhang S, Xu Y, Usuyama N *et al*. A Multimodal Biomedical Foundation Model Trained from Fifteen Million Image–Text Pairs. *NEJM AI* 2025;**2**:A10a2400640. <https://doi.org/10.1056/A10a2400640>
 13. Lu MY, Chen B, Williamson DFK *et al*. A visual-language foundation model for computational pathology. *Nat Med* 2024;**30**:863–74. <https://doi.org/10.1038/s41591-024-02856-4>
 14. Li J, Li D, Xiong C *et al*. BLIP: Bootstrapping language-image pretraining for unified vision-language understanding and generation. In: Chaudhuri K, Jegelka S, Song L, Szepesvári C, Niu G, Sabato S (eds.), *Proceedings of the 39th International Conference on Machine Learning*, vol. 162 of *Proceedings of Machine Learning Research*, Baltimore, MD, USA: PMLR, pp. 12888–900. 2022.
 15. Loo J, Robbins M, McNeil C *et al*. Autofluorescence virtual staining system for H&E Histology and multiplex immunofluorescence applied to Immuno-oncology biomarkers in lung cancer. *Cancer Res Commun* 2025;**5**:54–65. <https://doi.org/10.1158/2767-9764.CRC-24-0327>
 16. Autofluorescence Image Reconstruction and Virtual Staining for In-Vivo Optical Biopsy | IEEE Journals & Magazine | IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/9359782>.
 17. Christiansen EM, Yang SJ, Ando DM *et al*. In silico Labeling: predicting fluorescent labels in Unlabeled images. *Cell* 2018;**173**:792–803.e19. <https://doi.org/10.1016/j.cell.2018.03.040>
 18. McNeil C, Wong PF, Sridhar N *et al*. An end-to-end platform for digital pathology using hyperspectral autofluorescence microscopy and deep learning-based virtual histology. *Mod Pathol* 2024;**37**:100377. <https://doi.org/10.1016/j.modpat.2023.100377>
 19. Zhu R, He H, Chen Y *et al*. Deep learning for rapid virtual H&E staining of label-free glioma tissue from hyperspectral images. *Comput Biol Med* 2024;**180**:108958. <https://doi.org/10.1016/j.combiomed.2024.108958>
 20. Bayramoglu N, Kaakinen M, Eklund L *et al*. Towards Virtual H&E Staining of Hyperspectral Lung Histology Images Using Conditional Generative Adversarial Networks. In: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy: IEEE, pp. 64–71. 2017. <https://doi.org/10.1109/ICCVW.2017.15>
 21. Nadarajan G, Hope T, Wang D *et al*. Automated multi-class ground-truth labeling of H&E images for deep learning using multiplexed fluorescence microscopy. In: Tomaszewski JE, Ward AD (eds.), *Medical Imaging 2019: Digital Pathology*, Bellingham, WA, USA: SPIE, Vol. 10956, pp. 102–11. 2019.
 22. Parmar G, Park T, Narasimhan S *et al*. One-step image translation with text-to-image models. arXiv preprint arXiv:2403.12036, 2024. <https://doi.org/10.48550/arXiv.2403.12036>
 23. Janesick A, Shelansky R, Gottscho AD *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>
 24. Brancati N, Anniciello AM, Pati P *et al*. BRACS: a dataset for BReAst carcinoma subtyping in H&E Histology Images. *Database* 2022;**2022**:baac093.
 25. Monkman J, Taheri T, Ebrahimi Warkiani M *et al*. High-plex and high-throughput digital spatial profiling of non-small-cell lung cancer (NSCLC). *Cancers* 2020;**12**:3551. <https://doi.org/10.3390/cancers12123551>
 26. Das A, Hasib MM, Liu Z *et al*. Abstract 7394: delineating spatial expression signatures of lung adenocarcinoma subtypes from spatial single-cell transcriptomics using graph neural networks. *Cancer Res* 2024;**84**:7394. <https://doi.org/10.1158/1538-7445.AM2024-7394>
 27. Travis WD, Brambilla E, Nicholson AG *et al*. The 2015 World Health Organization classification of lung Tumors. *J Thorac Oncol* 2015;**10**:1243–60. <https://doi.org/10.1097/JTO.0000000000000630>
 28. Chang Y, Cesarman E, Pessin MS *et al*. Identification of herpesvirus-like DNA sequences in AIDS-Associated Kaposi's sarcoma. *Science* 1994;**266**:1865–9. <https://doi.org/10.1126/science.7997879>
 29. Meng W, Das A, Sinha H *et al*. Spatial single-cell atlas reveals KSHV-driven broad cellular reprogramming, progenitor expansion, immune and vascular remodeling in Kaposi's sarcoma. *bioRxiv* 2025:2025.09.01.673567. <https://doi.org/10.1101/2025.09.01.673567>
 30. Buonaguro FM, Tornesello ML, Buonaguro L *et al*. Kaposi's sarcoma: aetiopathogenesis, histology and clinical features. *J Eur Acad Dermatol Venereol* 2003;**17**:138–54. <https://doi.org/10.1046/j.1468-3083.2003.00670.x>
 31. Antman K, Chang Y. Kaposi's sarcoma. *N Engl J Med* 2000;**342**:1027–38. <https://doi.org/10.1056/NEJM200004063421407>
 32. Latonen L, Koivukoski S, Khan U *et al*. Virtual staining for histology by deep learning. *Trends Biotechnol* 2024;**42**:1177–91. <https://doi.org/10.1016/j.tibtech.2024.02.009>
 33. Hussain MA, Waris MA, Akram MU *et al*. VISGAB: virtual staining-driven GAN benchmarking for optimizing skin tissue histology. *Sci Rep* 2025;**15**:42430. <https://doi.org/10.1038/s41598-025-26493-0>
 34. Kumari N, Zhang R, Shechtman E *et al*. Ensembling Off-the-Shelf Models for GAN Training. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, pp. 10651–62. 2022.