JAMA Network Open

Original Investigation | Health Informatics

Use of Deep Learning to Develop and Analyze Computational Hematoxylin and Eosin Staining of Prostate Core Biopsy Images for Tumor Diagnosis

Aman Rana, MS; Alarice Lowe, MD; Marie Lithgow, MD; Katharine Horback, MD; Tyler Janovitz, MD; Annacarolina Da Silva, MD; Harrison Tsai, MD; Vignesh Shanmugam, MD; Akram Bayat, PhD; Pratik Shah, PhD

Abstract

IMPORTANCE Histopathological diagnoses of tumors from tissue biopsy after hematoxylin and eosin (H&E) dye staining is the criterion standard for oncological care, but H&E staining requires trained operators, dyes and reagents, and precious tissue samples that cannot be reused.

OBJECTIVES To use deep learning algorithms to develop models that perform accurate computational H&E staining of native nonstained prostate core biopsy images and to develop methods for interpretation of H&E staining deep learning models and analysis of computationally stained images by computer vision and clinical approaches.

DESIGN, SETTING, AND PARTICIPANTS This cross-sectional study used hundreds of thousands of native nonstained RGB (red, green, and blue channel) whole slide image (WSI) patches of prostate core tissue biopsies obtained from excess tissue material from prostate core biopsies performed in the course of routine clinical care between January 7, 2014, and January 7, 2017, at Brigham and Women's Hospital, Boston, Massachusetts. Biopsies were registered with their H&E-stained versions. Conditional generative adversarial neural networks (cGANs) that automate conversion of native nonstained RGB WSI to computational H&E-stained images were then trained. Deidentified whole slide images of prostate core biopsy and medical record data were transferred to Massachusetts Institute of Technology, Cambridge, for computational research. Results were shared with physicians for clinical evaluations. Data were analyzed from July 2018 to February 2019.

MAIN OUTCOMES AND MEASURES Methods for detailed computer vision image analytics, visualization of trained cGAN model outputs, and clinical evaluation of virtually stained images were developed. The main outcome was interpretable deep learning models and computational H&E-stained images that achieved high performance in these metrics.

RESULTS Among 38 patients who provided samples, single core biopsy images were extracted from each whole slide, resulting in 102 individual nonstained and H&E dye–stained image pairs that were compared with matched computationally stained and unstained images. Calculations showed high similarities between computationally and H&E dye–stained images, with a mean (SD) structural similarity index (SSIM) of 0.902 (0.026), Pearson correlation coefficient (PCC) of 0.962 (0.096), and peak signal to noise ratio (PSNR) of 22.821 (1.232) dB. A second cGAN performed accurate computational destaining of H&E-stained images back to their native nonstained form, with a mean (SD) SSIM of 0.900 (0.030), PCC of 0.963 (0.011), and PSNR of 25.646 (1.943) dB compared with native nonstained images. A single blind prospective study computed approximately 95% pixel-by-pixel overlap among prostate tumor annotations provided by 5 board certified pathologists on computationally stained images, compared with those on H&E dye–stained images. This study also used the first visualization and explanation of neural network kernel activation maps during H&E staining and destaining of RGB images by cGANs. High similarities between kernel activation maps of

(continued)

Key Points

Question Can deep learning systems perform hematoxylin and eosin (H&E) staining and destaining and are the virtual core biopsy samples generated by them as valid and interpretable as their real-life unstained and H&E dye–stained counterparts?

Findings In this cross-sectional study, deep learning models were trained using nonstained prostate core biopsy images to generate computationally H&E stained images, and core biopsy images were extracted from each whole slide, consisting of approximately 87 000 registered patch pairs of 1024 × 1024 × 3 pixels each. Comprehensive analyses of virtually stained images vs H&E dye–stained images confirmed successful computational staining.

Meaning The findings of this study suggest that whole slide nonstained microscopic images of prostate core biopsy, instead of tissue samples, could be integrated with deep learning algorithms to perform computational H&E staining and destaining for rapid and accurate tumor diagnosis.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

computationally and H&E-stained images (mean-squared errors <0.0005) provide additional mathematical and mechanistic validation of the staining system.

CONCLUSIONS AND RELEVANCE These findings suggest that computational H&E staining of native unlabeled RGB images of prostate core biopsy could reproduce Gleason grade tumor signatures that were easily assessed and validated by clinicians. Methods for benchmarking, visualization, and clinical validation of deep learning models and virtually H&E-stained images communicated in this study have wide applications in clinical informatics and oncology research. Clinical researchers may use these systems for early indications of possible abnormalities in native nonstained tissue biopsies prior to histopathological workflows.

JAMA Network Open. 2020;3(5):e205111. doi:10.1001/jamanetworkopen.2020.5111

Introduction

Cancer is the second leading cause of death in the US.[1] An estimated 164 690 US men were diagnosed with prostate cancer and 29 430 died of the disease in 2018.[1] While the survival rate for people with localized prostate cancer is more than 98%, it is reduced to 30% when the cancer spreads to other parts of the body, such as distant lymph nodes, bones, or other organs.[1] This reduction in survival rate may be prevented with early diagnosis. The current criterion standard for prostate cancer diagnosis uses dye staining of core biopsy tissue and subsequent microscopic histopathologic examination by trained pathologists.[2] Hematoxylin and eosin (H&E) is the most widely used dye staining method that leverages interactions of H&E dyes with tissues for visualization.[3] Every day, up to 3 million slides are stained with this technique. The total end-to-end processing time from slide scanning to automated staining is less than 10 minutes. However, tissue processing of H&E dye staining for paraffin sections can take between 7 to 26 hours.[4] Microscopic diagnosis of tumors using H&E dye–stained biopsy slides presents challenges, such as inconsistencies introduced during tissue preparation and staining and human errors, and it also requires significant processing time, imaging systems, and procedural costs.[5] Other key challenges include sampling time, which can limit the amount of tissue that can be stained owing to time and cost involved, resulting in evaluation of only three 4-µm sections of tissue to represent a 1-mm diameter core.[6] Irreversible dye staining of tissues leads to loss of precious biopsy samples that are no longer available for biomarker testing. Automated, low-cost, and rapid generative algorithms and methods that can convert native nonstained whole slide images (WSIs) to computationally H&E stained versions with high precision could be transformative by benefiting patients and physicians and by reducing errors and costs.

Whole-slide pathological images are approved by the US Food and Drug Administration[7] for cancer diagnosis and can rapidly be integrated into machine learning and artificial intelligence algorithms for automatic detection of cellular and morphological structures to tumors and virtual staining.[8] Studies testing operational feasibility and validation of results obtained by generative models and machine learning algorithms in controlled clinical trials or hospital studies for virtual staining of whole-slide pathology images do not exist, to our knowledge, precluding clinical deployment of these systems.

We previously communicated convolutional neural networks for learning associations between expert annotations of disease and fluorescent biomarkers manifested on RGB (red, green, and blue channel) images and their complementary nonfluorescent pixels found on standard white-light images.[9] Subsequently, we communicated conditional generative adversarial neural networks (cGANs) that accept native nonstained prostate core biopsy autofluorescence RGB WSIs and computationally stain them in a manner visually similar to H&E by learning hierarchical nonlinear mappings between image pairs before and after H&E dye staining.[10] In this study, we report several

novel mechanistic insights and methods to facilitate benchmarking and clinical and regulatory evaluations of computationally H&E stained images for oncological applications. Specifically, we trained high fidelity, explainable, and automated computational staining and destaining algorithms to learn mappings between naturally autofluorescent pixels[11] of nonstained cellular organelles and their stained counterparts. We also devised robust loss function for our machine learning algorithms to preserve tissue structure. Furthermore, we established that our virtual staining neural network models were generalizable to accurately stain previously unseen images acquired from patients and tumor grades not part of training data. We generated neural activation maps to provide the first instance of explainability and mechanisms used by cGANs models for virtual H&E staining and destaining and establish computer vision analytics to benchmark the quality of generated images. Finally, we evaluated computationally stained images for prostate tumor diagnoses with multiple pathologists for clinical evaluation (**Figure 1**).

By describing explainable algorithms and quantitative methods that can consistently, rapidly, and accurately perform computational staining and destaining of prostate biopsy RGB WSI, this study communicates a detailed method and process that may be useful to generate evidence for clinical and regulatory authentication of computationally H&E stained images. However, greater numbers of virtually stained H&E images sourced from larger pools of patients are needed for prospective evaluation of such models.

Methods

Partners Human Research Committee approved our study protocol for utilization of excess material from prostate core biopsies performed in the course of routine clinical care between January 7, 2014, and January 7, 2017, at Brigham and Women's Hospital, Boston, Massachusetts. Informed consent was waived because data were deidentified and samples were obtained as part of routine clinical care. Deidentified WSIs and electronic health record (EHR) data were transferred to the Massachusetts Institute of Technology for processing and analyses and was exempt from institutional review board review per the Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects. This study is reported following Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.

Study Population

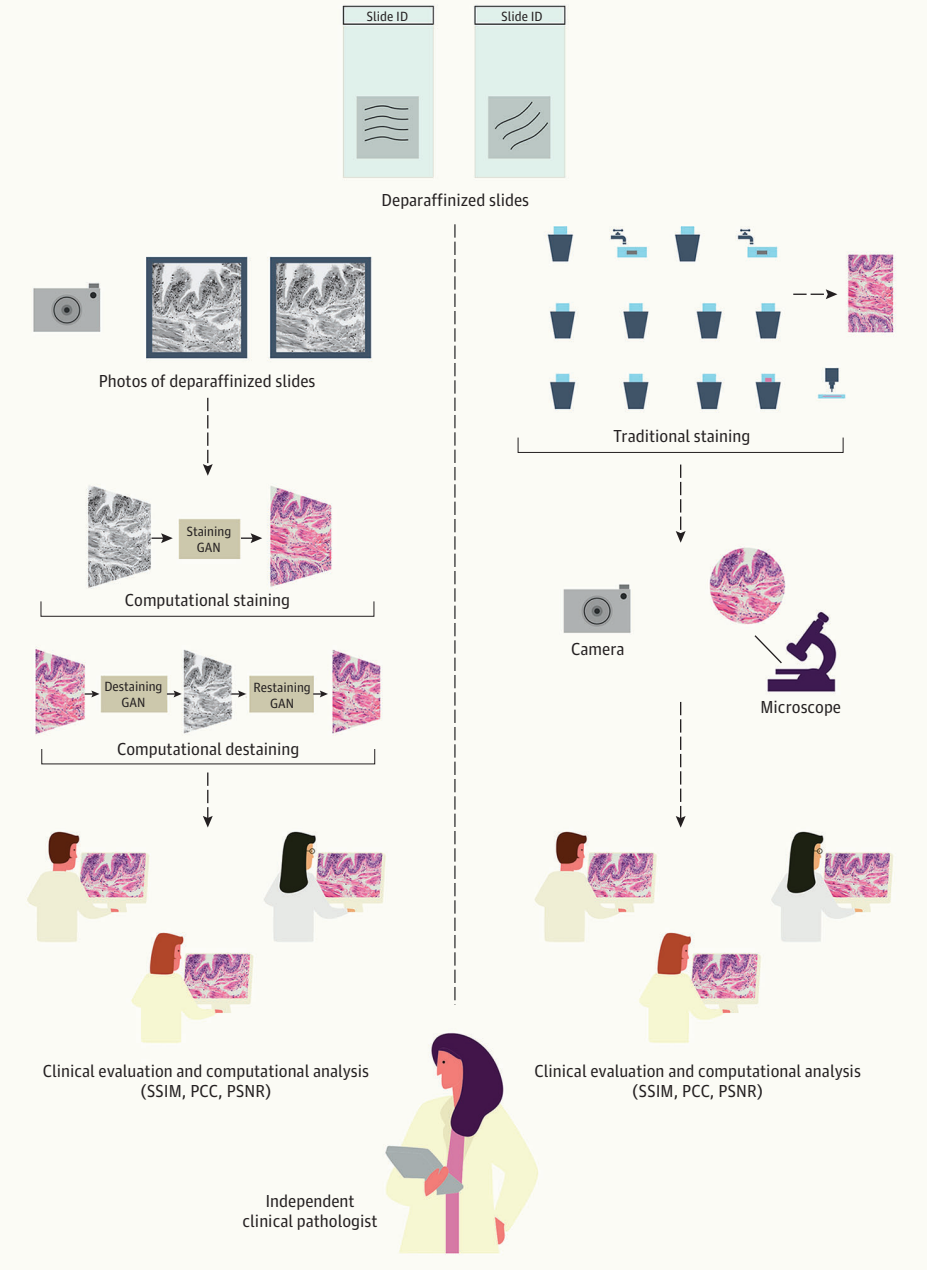
Thirty-eight men (mean [SD] age, 66.2 [8.9] years), including 32 white men (84%), 4 African American men (11%), 1 Hispanic/Latino man (3%), and 1 Dominican man (3%), provided 46 core biopsy samples (eTable 1 in the Supplement). Each biopsy sample contained 1 to 6 cores of tissue, and 0% to 100% of each tissue core contained prostatic adenocarcinoma of various Gleason grades. Individual prostate tissue needle core biopsy images from each whole slide image were extracted, which resulted in 102 high-resolution native nonstained and H&E dye–stained image pairs that were registered to form 102 RGB WSIs pair images.

Deparaffinized nonstained slides were scanned at 20 × magnification. Subsequently, slides were stained with H&E dye on a Dako autostainer (Agilent), and these stained slides were rescanned at 20 × magnification at Harvard Medical School Tissue Microarray and Imaging Core. Deidentified data in the form of nonstained and H&E dye–stained images at 20 × magnification were analyzed at Massachusetts Institute of Technology. Individual prostate tissue needle core biopsy images from each WSI were extracted, which resulted in high-resolution native nonstained and H&E dye–stained image pairs, which then were registered to form RGB WSI pair images. The RGB WSIs were too large to enter into deep learning networks; therefore, each image was cropped into multiple patches of 1024 × 1024 × 3 pixels, resulting in approximately 87 000 registered pair patches.

Training and Validation Data Sets

The numbers of image patches and their Gleason grade tumor grading used in this study for test data were 7019 Gleason grade 3 tumors, 6149 Gleason grade 4 tumors, and 270 Gleason grade 5 tumors (eTable 1 in the Supplement) and were sufficient to study the computational staining problem. Core biopsy images from a larger and diverse patient population from additional medical centers are currently being procured to improve generalizability of clinical findings reported in this study. The registered data set of images (WSI pairs) was divided into approximately 74 000 training and 13 500 validation image patches. Validation and training data sets were balanced to include images from healthy patients as well as patients with different grades of prostate tumors and of each tumor grade. More information about the data collection process and training and validation sample descriptions can be found in eAppendix 1 and eTable 1 in the Supplement.

Figure 1. Overview of the Staining Processes



Left, Computational staining and destaining of whole slide prostate core biopsy images with conditional generative adversarial neural networks (CGAN). Right, traditional staining with hematoxylin and eosin (H&E) dyes using physical prostate core tissue biopsy slides. PCC indicates Pearson correlation coeffecient; PSNR, peak signal to noise ratio; and SSIM, structural similarity index.

Machine Learning Model Architecture and Training

A cGAN pix2pix-based model was trained to learn distribution and mappings among registered images in the training data set.[12] The staining model accepts a native nonstained RGB WSI and generates computationally H&E stained RGB WSI. The destaining model reverses the process and generates computationally destained images from H&E dye–stained RGB WSI patches. A novel Pearson correlation coefficient (PCC) term was added to the cGAN loss function to improve the quality, enforce tissue structure preservation of computationally stained images, and help reduce the tiling artifacts in the computationally stained image. More details about loss function and technical implementation are presented in eAppendix 2 in the Supplement.

Image Evaluation Metrics

The computationally stained image patches generated by our model were compared with H&E dye patches to obtain a quantitative measure of the generated images. We used PCC, peak signal to noise ratio[13] (PSNR) and structural similarity index[14] (SSIM) to quantify similarities and differences between a given pair of images at a pixel level. The values of PCC and SSIM range from 0 to 1, with higher values indicating higher levels of similarity. Accurate values of PSNR for wireless transmission quality loss are considered to be between 20 dB to 25 dB,[15,16] and higher PSNR is better. The mean and total increase in pixel intensity after computationally staining and destaining were calculated by subtracting the mean pixel intensity of the second image from the first.

Clinical Validation of Computationally H&E Stained RGB WSI

Computationally stained patches from the test data set were used to reconstruct 13 RGB WSIs. Their corresponding RGB WSI H&E dye–stained images were used as ground truth examples and also labeled for tumors. A single-blind study was conducted for evaluation of all images for prostate cancer diagnosis. Four board-certified and trained expert pathologists provided detailed labels in the form of free-form outlines encompassing tumors, indicating tumor regions (with grade) and other atypical manifestations on computationally stained and corresponding H&E dye–stained images. In the first round, 2 randomly selected pathologists were provided computationally stained images while H&E dye–stained images were given to the other 2 raters. After a period of 4 weeks, the image sets were swapped among the pathologists, and another round of annotations were conducted. Pathologists annotated images in the form of free-hand drawing using the Sedeen Viewer (PathCore) on identical notebook computer screens (Dell Computers). By using different colors corresponding to each tumor grade, annotations were classified with tumor grade: Gleason grade 3, Gleason grade 4, or Gleason grade 5. A separate comments box was used to note other clinical observations and for anatomical features. The annotations and the associated labels (ie, Gleason grades 3, 4, or 5) were extracted from the XML files generated by Sedeen using the labels and annotations using Python code. The agreement was calculated in the form of intersection over union which measures the number of pixels on computationally stained and corresponding H&E dye–stained images that have common raters annotations divided by the total number of pixels present across both images.[17] An independent fifth clinical pathologist ratified corresponding computationally stained images. Accuracies and errors were calculated using pixel-by-pixel overlap in the labels. Color-coded error overlaid validation images were generated visualizing the true positives, false positives, and false negatives (eFigure 1 in the Supplement).

Activation Maps

Input images containing Gleason grade 3, 4, and 5 signatures were entered into our trained computational staining network to visualize activation maps for each input image. Full-scale RGB WSIs at 20 × resolution were collected and constructed for 8 image data sets (ie, 4 pairs), each with 13 images: ground native nonstained, ground H&E dye–stained, predicted reconstructed computationally stained, and predicted reconstructed destained (also referred to as *predicted destained images*) (eAppendix 3 in the Supplement). A total of 448 unique patches in each of the 8

data sets with no overlap were created for each data set and set to size 1024 × 1024 × 3 (3 color channels). For each matching patch pair to be entered into computational staining or destaining models, we examined the grid linearly and isolated consolidated activation maps from layer 1 to layer 19 (eFigure 2 in the Supplement). These individual activation maps concatenated together to form a single image per layer of the model architecture. Examples of concatenated activation maps are presented in eFigure 3 and eFigure 4 in the Supplement. The normalized mean-square error (MSE) was calculated for all layers between the matching patch pairs (eAppendix 3 and eFigure 5 in the Supplement).

Results

Quantitative Evaluation of Computationally Stained and Destained Images

Computationally H&E stained WSIs were compared pixel-by-pixel to corresponding H&E dye-stained images (**Table 1**). We calculated SSIM and PSNR (in logarithmic dB), and PCC were used as quality measures of computationally stained images with H&E dye-stained images regarded as ground truth.[18,19] A mean (SD) SSIM of 0.902 (0.026), PCC of 0.962 (0.096), and PSNR of 22.821 (1.232) dB were calculated, indicating high accuracy of computational H&E staining of test images. High PCC accuracy scores (81.8% of patches with PCC ≥0.7 and 39.4% patches with PCC ≥0.8) indicate that computationally stained patches matched H&E dye-stained patches at a pixel level.

Comparison of each RGB color channel's pixel intensities (PXI) between native nonstained and computationally stained images (−42 PXI), and those between native nonstained and H&E dye-stained images (−44 PXI) show that computationally stained images had mean intensity difference of only 2 PXI (eTable 2 in the Supplement). Similar low differences between ground truth H&E dye and computationally stained images were observed after comparing individual color channels listed in eTable 3 in the Supplement: red channel (unstained vs computationally stained: −58 PXI; unstained vs H&E stained: −58 PXI; H&E vs computationally stained: 0 PXI), green channel (unstained vs computationally stained: 6 PXI, unstained vs H&E stained: −8 PXI, H&E vs computationally stained: 2 PXI), and blue channel (unstained vs computationally stained: −62 PXI, unstained vs H&E stained: −65 PXI, H&E vs computationally stained: 3 PXI).

Table 1. Comparison of Computationally Stained and Ground Truth H&E Dye-Stained Images, and of Computationally Destained vs Ground Truth Native Nonstained Images

	Computationally stained vs H&E stained			Computationally destained vs nonstained		
Image	PCC, %[a]	SSIM, %[b]	PSNR, db[c]	PCC, %[a]	SSIM, %[b]	PSNR, db[c]
1	95.0	86.0	20.563	95.1	85.3	23.486
2	95.2	89.1	22.387	96.5	89.5	25.706
3	94.9	86.0	20.683	96.4	86.6	24.871
4	95.7	92.9	22.870	96.8	93.6	27.469
5	96.0	94.7	24.838	97.0	94.9	21.194
6	95.5	91.4	22.903	95.7	91.4	25.285
7	96.0	88.1	22.486	93.8	86.5	21.863
8	96.8	93.1	24.132	95.9	92.7	26.164
9	97.8	89.0	23.411	96.8	87.4	24.165
10	95.6	91.3	23.177	96.7	92.7	27.359
11	97.2	90.7	23.945	98.4	89.9	26.792
12	97.5	90.2	23.200	96.3	89.9	24.957
13	96.5	89.9	22.074	97.0	90.0	26.082
Total, mean (SD)	96.1 (1.0)	90.2 (2.6)	22.821 (1.232)	96.3 (1.1)	90.0 (3.0)	25.646 (1.943)

Abbreviations: H&E hematoxylin and eosin; PCC, Pearson correlation coefficient; PSNR, peak signal to noise ratio; SSIM, structural similarity index.

[a] Pearson correlation coefficient of 1.0 indicates a perfect match.

[b] Structural similarity index of 1.0 indicates perfect match.

[c] Peak signal to noise ratio of 22 dB or more is considered high quality.

Prostate core biopsy H&E dye–stained images were computationally destained and compared with native nonstained images. Mean (SD) PCC was 0.900 (0.030), mean (SD) SSIM was 0.963 (0.011), and mean (SD) PSNR was 25.646 (1.1943) dB (Table 1), thus showing high similarities with native ground truth nonstained images. RGB pixel intensities between computationally destained and H&E dye–stained images (47 PXI), and native nonstained and H&E dye–stained images (44 PXI) also indicated that computationally destained and ground truth nonstained images only had 3 PXI difference in their overall intensities (eTable 2 in the Supplement). These results indicate high fidelity of learning, reproducing, and erasing of multichromatic information by computational H&E staining and destaining algorithms. Mean (SD) change in pixel intensities in the red and blue channels were higher compared with the green channel likely because H&E dye predominantly consists of blue and red or pink colors.

Analyses of Physician Annotations

Intersection over union indicating agreements or disagreements among pathologists examining the same set of images (intra–intersection over union) was calculated by pixel-by-pixel comparisons of their tumor and nontumor annotations (eTable 4 in the Supplement). Pathologists examining H&E dye–stained images had high mean (SD) intra–intersection over union agreement scores for diagnosing any tumors (0.81 [0.07]). Pathologists examining computationally H&E stained images also had high and comparable mean (SD) intra–intersection over union agreement scores for diagnosing any tumor (0.77 [0.08]). These results indicated high internal consistency in clinical diagnoses provided by each set of pathologists on their respective images. Furthermore, tumor diagnoses using computationally stained images were not associated with rater's sensitivity or specificity while detecting tumors.

Tumor labels provided by 2 sets of physicians in our single-blind study on ground truth H&E dye–stained images vs computationally stained images were then compared using inter–intersection over union agreement score metric[17] (**Table 2**). An overall inter–intersection over union score of 0.79 was calculated for any tumor diagnoses. The mean (SD) inter–intersection over union agreement score for Gleason grade 3 tumors was 0.70 (0.17) and 0.73 (0.15) for Gleason grade 4 labels. Gleason grade 5 tumors are rare, and we only had 1 example in validation data that was annotated, with an accuracy of 0.64 (Table 2). The mean (SD) inter–intersection over union agreement score for annotations of healthy areas in the tissue where no tumors were found on images was 0.90 (0.12)

Table 2. Intersection Over Union–Based Agreement Among Pathologists for Tumor Signatures Provided Using Computationally Stained Images Compared With Pathologists Using Ground Truth Hematoxylin And Eosin Dye–Stained Images

	Intersection over union[a]				
			Gleason grade		
Image	Any tumor	Healthy	3	4	5
1	0.90	0.96	0.90	NA	NA
2	0.86	0.55	NA	0.78	NA
3	NA	1.00	NA	NA	NA
4	0.92	0.89	0.76	NA	NA
5	0.52	0.90	NA	0.49	0.64
6	0.80	0.93	0.58	NA	NA
7	0.70	0.94	0.53	NA	NA
8	0.79	0.92	NA	0.77	NA
9	0.58	0.96	0.48	NA	NA
10	0.86	0.86	0.70	0.72	NA
11	0.92	0.99	0.92	NA	NA
12	NA	1.00	NA	NA	NA
13	0.93	0.78	NA	0.89	NA
Mean	0.79 (0.14)	0.90 (0.12)	0.70 (0.17)	0.73 (0.15)	0.64 (0)

Abbreviation: NA, not applicable

[a] Higher intersection over union score is better, with a score of 1.0 indicating perfect match of labels.

(Table 2). These results indicate that our trained machine learning models can accurately generate both tumor and nontumor signatures via computational H&E staining. Physician raters showed concordance and comparable sensitivity and specificity in diagnosis made using H&E dye–stained images compared with those made by using computationally stained images.

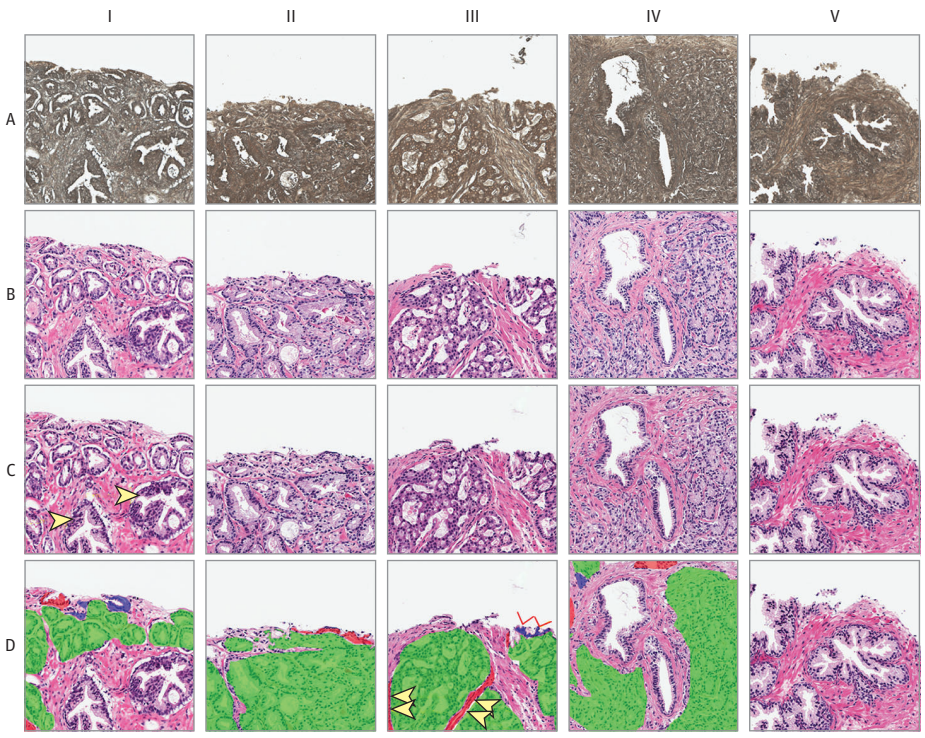
Clinical Evaluations of Computationally Stained Images

Figure 2 shows representative input nonstained image patches that had Gleason grade 3 or 4 tumors or were benign and their computational H&E staining and accuracy calculated using annotations by multiple physicians. It is evident that the computationally H&E stained patches represent tumor signatures with high accuracy, and pathologists were able to correctly identify tumors. Most observed disagreements between raters did not represent misidentification of glands as benign or malignant. Instead, they showed differences in rater annotation at borders of tumor labels, mainly due to differences in labeling style. eAppendix 4 and eFigure 1 in the Supplement provide detailed clinical evaluations and outcomes of individual patches and reconstructed RGB WSI computationally H&E stained images.

Comparison With Patient Records

Most of the diagnoses rendered using computationally stained images agreed with the corresponding initial clinical diagnosis reported in electronic health records (EHRs) (eTable 5 in the Supplement), supporting the validity of the generated images for tumor detection and diagnoses. Most of the samples showed identical tumor fractions and Gleason grading. None of the differences between EHRs and diagnoses based on computationally H&E stained images were clinically significant with regard to treatment decisions (eAppendix 5 in the Supplement). We were able to overturn originally reported results in patient records in 2 instances using computationally stained images (eFigure 1 and eAppendix 5 in the Supplement).

Figure 2. Representative Image Patches Generated by the Computational Staining Neural Network and Their Comparison With Corresponding Ground Truth Hematoxylin and Eosin (H&E) Dye–Stained Images



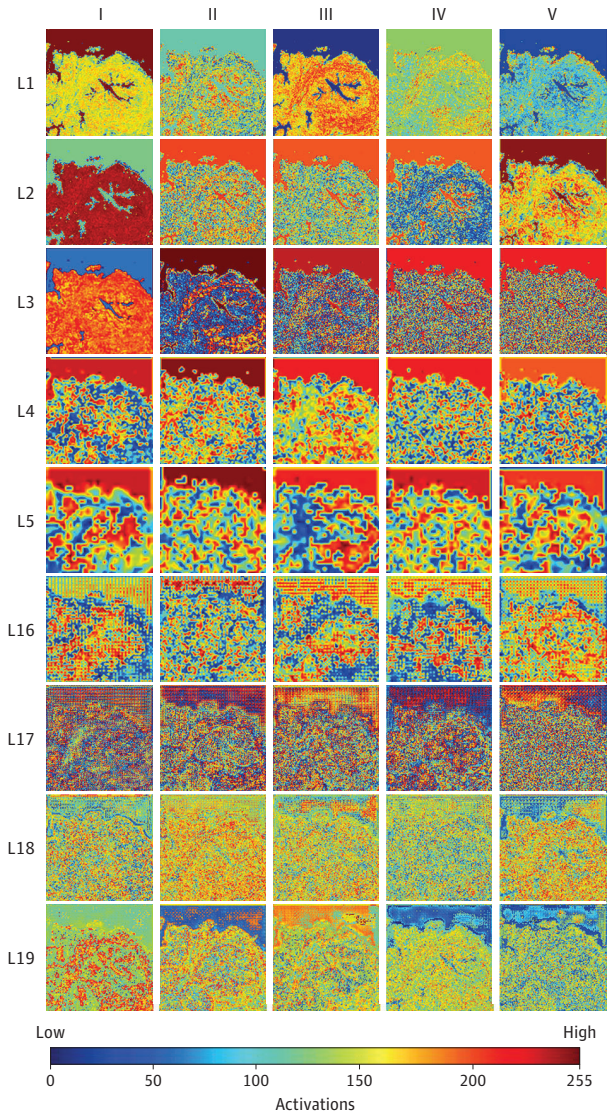
Row A, Deparaffinized native nonstained image patches entered into the neural network. Row B, Ground truth H&E dye–stained patches. Row C, computationally H&E stained patches generated by the neural network. Arrows in C-I indicate the 2 benign glands, all other glands represent tumors. Row D, shows computationally H&E stained patches overlaid with colors indicating agreements and disagreements between physician annotations on these images compared with ground truth H&E dye–stained images. Variation in labeling detail by annotators (arrows) are shown in D-III. Green indicates true positive; blue, false negative; and red, false positive.

Analysis and Explanation of Neural Network Activation Maps

Neural activation maps of trained staining and destaining cGAN models were analyzed after entering healthy or Gleason grade 3, 4, or 5 images patches (Figure 2). In this study, we did not use a classification approach to identify image features, but rather performed pixel-by-pixel visualization, explanation, and intensity ranking (>200 value) of various cGAN kernels to create an activation map of a particular nonstained image patch (healthy vs with a particular Gleason tumor grade) as it passes through each network layer while getting stained (eFigure 2 in the Supplement).

We demonstrate and compare presence of unique low- and high-level features in input images that activate neurons and feature maps in the cGAN generator network (**Figure 3**; eFigure 3 and eFigure 4 in the Supplement). For example, initial layers of the convolutional layers in the generator

Figure 3. Activation Maps of Kernels of Trained Generator Neural Network Model



Activation maps of kernels of trained generator neural network model layers after feeding a native nonstained prostate core biopsy image patch without tumor as it gets computationally hematoxylin and eosin–stained. Rows show top 5 activation maps from layers L1 to L5 and L16 to L19 arranged in decreasing order of their activations from left to right (columns I-V).

detected low-level features, such as tissue geometry, edges, corners, shapes, and a few changes in color (Figure 3; eFigure 3, and eFigure 4 in the Supplement). We observed well-demarcated boundaries between tissues and background and gross distinctions between glands and stroma are suggested (Figure 3) or are well defined (eFigure 3 in the Supplement). Kernels of initial layers of trained models thus help with differentiating tissue from background and morphological tasks to define higher order anatomical structures. The later convolutional layers leverage previously learned low level features and ability to differentiate tissue from background with fine-grained structures, such as anatomical arrangement of nuclei and tumor signatures (Figure 3; eFigure 3 and eFigure 4 in the Supplement). eFigure 4 in the Supplement shows additional examples of indeterminate atypical glands and tumors with edge or crush artifacts that are well preserved on the computationally generated images but were differentially designated as tumor or nontumor by raters. The activation maps of kernels of various generator neural network layers after feeding H&E dye–stained patch with Gleason grade 4 and 5 prostate tumors were demonstrated.

We compared kernel activation maps of all 448 validation image patches used to test our trained staining and destaining machine learning models with corresponding ground truth dye-stained and native nonstained images (eFigure 5 in the Supplement). The MSE was calculated by comparing activation maps generated by each of the 19 neural network layers in response to pairs of images being evaluated. The MSE was low for the first layer, increases for second layer, and then decreases for the remaining layers. These results, in unification with our detailed SSIM, PSNR, PCC, and physician validation, provide significant evidence of the high quality of computationally stained and destained images, with consequent high sensitivity and specificity in diagnosing tumors using them.

Discussion

Most surgical and medical treatments for cancer, including chemotherapy, endocrine therapy, and immunotherapy, are dictated by histopathologic examination and diagnosis. Increase in use of core biopsies for diagnosis, in place of larger surgical biopsies, has resulted in significant decrease in the volume of tumor available for performing an ever-increasing battery of biomarker testing for diagnostic, prognostic, and predictive information. In this cross-sectional study, computationally stained and destained images were evaluated by multiple image analytics and matched ground truth images with high similarity. The high quality of the computationally stained and destained images were comprehensively and stringently validated using pixel-based comparison (eg, MSE, PSNR), spatial structural comparison (eg, SSIM), and localized correlational comparison (eg, PCC), which revealed their macroscopic and microscopic suitability for clinical deployment.

Evaluation by trained pathologists showed tumorous and healthy tissues were morphologically well represented most of the computationally stained images with high accuracy. The glands and stroma of benign prostatic tissue and carcinoma were identifiable, showing preserved architectural features (ie, location and shape of the glands), defined gland/stromal interface, and cytological characteristics (ie, location and appearance of the nuclei and nucleoli, if present). Most of the differences in annotations were observed either on the tumor/nontumor interface or boundary or the biopsy boundary. This can be attributed to the labeling style of individual raters.[17] Previous studies have reported that human readers show substantial variability and lower performance compared with computer algorithms in terms of tumor segmentations.[8,20] A similar limitation of using a human reader panel to establish a reference standard for evaluation of computer algorithms may have affected this study. In validation images, presence of morphologically ambiguous glands, a known histopathological dilemma that clinically requires additional testing for confident diagnosis, also led to differing labels among raters, as they were asked to categorize each gland as benign or malignant without assistance from supplemental studies. In most cases, these ambiguous cases were well represented in the computationally stained images but led to labeling differences owing to the ambiguity of these regions of interest. Small difference calculated by PSNR, SSIM, and PCC,

independent of the human raters, may also be in part due to registration differences in small, out-of-focus areas during whole-slide imaging. Input image pairs (ie, nonstained and H&E stained) used for training in our work were corrected for differences in field of view, illumination, and focal planes, but they may still have had minor variances. However, these small variances in computationally stained images had no effects on overall clinical assessments. Color variations in digital slides may have been due to differences in staining reagents, thickness of tissue sections, and staining protocols, which can negatively affect clinical diagnoses. We report minimal color variation across our computationally stained H&E images, as seen by their uniform overall RGB and individual RGB channel intensity values, which often matched training images. Physician raters in the study did not report difficulty in reading colors of nuclei, glands, cells, and tumors in computationally stained images, which was ratified by an additional independent pathologist. Thus, the trained neural network model reproduces a consistent and normalized color hue from the vast training data set that does not affect clinical decision-making from computational images. The subsequent absence of false-positive errors in healthy tissue cores of patients illustrates the fine grain reproduction of our computationally stained and destained images. We were also pleased to find high concordances between diagnoses made using the computationally stained images in this study and the patient's EHR. In fact, we found 2 instances in which the diagnoses made using computationally stained images overturned the initial findings in the EHR. In both cases, additional laboratory tests and clinical examinations were performed to confirm our findings. These results demonstrated that raters and the tumor diagnosis performed using computationally stained WSI used in our study matched or exceeded the initial microscopic diagnosis performed using H&E-stained tissue slides after prostate biopsy extraction.

Virtual staining of histopathological slide images has been reported using approaches with signals that require long detection times,[21] dye staining of nonstained specimens prior to imaging,[22] laser illumination and excitation with specific wavelengths,[23] sparse sampling, and poor depth resolution.[22,24] Previous virtual staining studies have performed limited analytics[25] to benchmark the quality of their virtually stained images. Most previous studies did not perform pixel-level comparisons with ground truth images and used small numbers of nonblinded raters who used coarse annotations without tumor gradations.[26,27] While other studies have reported no clinical validation and benchmarking of their results.[28-30] Similarly, previous deep learning research for virtual staining has used specialized illumination sources and did not report robust validation studies on mechanisms to establish computer vision or diagnostic utility of generated images.[8] Bayramoglu et al[31] virtually stained lung tissue slides multispectral images with a cGAN and achieved an SSIM of 0.3873, but they performed no clinical validation. Bulingame et al[32] used cGAN to convert H&E-stained pancreas slide RGB images to immunofluorescence images and achieve an SSIM of 0.883, but they also did not report clinical validation of generated images. Two studies by Rivenson et al[33,34] used a fluorescence scope with specialized ultraviolet filters to capture various tissue biopsy images and virtually H&E stain them using a neural network. Results and findings communicated in our study differ from previous deep learning based virtual staining studies in several key aspects. As examples, a wide field fluorescence microscope to image tissue[33,34] vs the nonfluorescent mode of the Food and Drug Administration–approved and widely available automated slide scanning system to capture images used in our study. A single pathologist compared anatomical features among virtually stained images using coarse labels, and pixel-level comparisons between tumor labels on virtual and ground truth images or concordance with EHR of patients were not conducted to calculate true- and false-positive occurrences of tumor diagnoses reported in that study.[33,34] Computational destaining of tissue images and stringent image analytics, such as PSNR or PCC, to benchmark quality of virtually stained images have not been reported in previous deep learning based studies.[31-34] Analysis or visualization of key neural network kernels and image features that get activated during the staining process have not been investigated, thus precluding mechanistic insights or mathematical validation of previous findings reported in literature.[31-34]

In this study, we evaluated trained neural network models that computationally stained native unlabeled RGB images of prostate core biopsy (acquired without band pass filters or specialized

hardware) with anatomical features of prostate and reproduce cancer tumor signatures with high accuracies. Computational pixel-by-pixel analysis and comparisons using PSNR, SSIM, and PCC demonstrated high similarities between our computationally stained images and their H&E dye–stained counterparts. Pixel-by-pixel changes in RGB color channels after computational staining and destaining by neural networks matched corresponding changes in RGB intensity when native nonstained images were H&E dye–stained in pathology laboratories vice versa. Detailed clinical validation in a single blind study found high interrater and intrarater agreements, calculated by pixel-by-pixel analyses of tumor labels provided by multiple board certified and trained physicians. Computationally stained images thus accurately represented healthy tissue as well as tumors of different Gleason grades, which were easily detected by human visual perception. Clinical diagnoses made using computationally stained images in our study were consistent with tumor diagnoses reported in EHRs. We investigated layers of generator neural networks and calculated activation of kernels during staining of different prostate tumor grades and benign tissue signatures to visualize and explain the process of computational H&E staining and destaining. Activation maps of our trained neural network models during computational staining or destaining of test images were highly similar to H&E dye–stained or native nonstained images. Thus, by visualizing and comparing activation feature maps of kernels of trained models, this work also presents the first explainable deep neural network framework for computationally H&E staining or destaining of native RGB images, to our knowledge.

Limitations

There were a few limitations of this study. The validation process for tumor diagnoses and Gleason grading of computationally H&E stained images can be affected by interobserver variability. For example, despite using a large rater panel in a single-blind study, tumor regions annotated by pathologists on WSIs are often coarse and may contain nonrelevant tissue that increases disagreements. Additional fine-grained image annotation tools are needed for precise validation of results generated by computational staining algorithms. The amount of data or number of patients used in this study was not exhaustive for clinical trials or other regulatory evaluations. The numbers of images used were found to be sufficient for accurate H&E staining, and adding additional images could result in modest improvements. This study described detailed methods that could be used to interpret deep learning systems and virtually H&E-stained images derived from them by computer vision analytics, and our findings may be useful to clinical and regulatory science researchers in the field. Because Gleason grade 5 tumors are quite rare, only 1 WSI was evaluated in the validation data that was annotated with intersection over union accuracy of 0.64. More nuanced diagnostic validation requiring evaluation of tissue beyond typical H&E staining (eg, evaluation requiring immunohistochemical staining) is not addressed in this study and is a possible development area of this work. The clinical outcomes from this study are limited to the evaluation of prostate core biopsies as a representative tissue type, but our methods and approach should generalize to other tissue biopsy evaluations. Application to other tumor types within core biopsies or to resection specimens of prostate cancer or other conditions will be evaluated in future work.

Conclusions

This cross-sectional study communicates methods and processes that may be useful for additional research and validation of computational H&E staining deep learning models and images generated by them. Adoption of these systems may reduce the time and effort required for manual staining and slide preparation, and more importantly, enable the preservation of precious tissue samples which could be used in a targeted fashion for biomarker evaluation. Greater numbers of virtually stained H&E images sourced from larger pools of patients are needed before prospective clinical evaluation of models described in this study can begin.[35-37]

ARTICLE INFORMATION

Accepted for Publication: February 28, 2020.

Published: May 20, 2020. doi:10.1001/jamanetworkopen.2020.5111

Corresponding Author: Pratik Shah, PhD, Massachusetts Institute of Technology, Program in Media Arts and Sciences, 20 Ames St, Cambridge, MA 02139 (pratiks@mit.edu).

Author Affiliations: Program in Media Arts and Sciences, Massachusetts Institute of Technology, Cambridge (Rana, Bayat, Shah); Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts (Lowe, Horback, Janovitz, Da Silva, Tsai, Shanmugam); Department of Pathology, Stanford University Medical Center, Stanford, California (Lowe); Boston University School of Medicine, VA Boston Healthcare, West Roxbury, Massachusetts (Lithgow).

Author Contributions: Drs Shah and Lowe had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Mr Rana and Dr Lowe contributed equally to this study.

Concept and design: Rana, Lowe, Shah.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Rana, Lowe, Shah.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Rana, Bayat, Shah.

Obtained funding: Shah.

Administrative, technical, or material support: Rana, Lowe, Lithgow, Janovitz, Da Silva, Shanmugam, Bayat.

Supervision: Lowe, Shah.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was supported by the Program in Media, Arts and Sciences at Massachusetts Institute of Technology and Department of Pathology at Brigham and Women's Hospital.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: Keya Larrel Oliver provided technical assistance for the study by procuring and coordinating samples and images from histopathological cores. Hyung-Jin Yoon, PhD assisted in analyzing results, literature review, and minor edits to figures and tables. They were not compensated for these contributions.

REFERENCES

1. American Cancer Society. Key statistics for prostate cancer: prostate cancer facts. Accessed April 10, 2020. https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html

2. Ravery V, Dominique S, Panhard X, Toublanc M, Boccon-Gibod L, Boccon-Gibod L. The 20-core prostate biopsy protocol—a new gold standard? *J Urol*. 2008;179(2):504-507. doi:10.1016/j.juro.2007.09.033

3. Feldman AT, Wolfe D. Tissue processing and hematoxylin and eosin staining. In: *Histopathology*. Springer; 2014:31-43. doi:10.1007/978-1-4939-1050-2_3

4. Leica Biosystem. An introduction to specimen processing. Accessed April 10, 2020. https://www.leicabiosystems.com/knowledge-pathway/an-introduction-to-specimen-processing/

5. McCann MT, Ozolek JA, Castro CA, Parvin B, Kovacevic J. Automated histology analysis: opportunities for signal processing. *IEEE Signal Process Mag*. 2015;32:78-87. doi:10.1109/MSP.2014.2346443

6. Goldstein AS, Huang J, Guo C, Garraway IP, Witte ON. Identification of a cell of origin for human prostate cancer. *Science*. 2010;329(5991):568-571. doi:10.1126/science.1189992

7. Abels E, Pantanowitz L. Current state of the regulatory trajectory for whole slide imaging devices in the USA. *J Pathol Inform*. 2017;8:23. doi:10.4103/jpi.jpi_11_17

8. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019;20(5):e253-e261. doi:10.1016/S1470-2045(19)30154-8

9. Yauney G, Angelino K, Edlund D, Shah P. Convolutional neural network for combined classification of fluorescent biomarkers and expert annotations using white light images. Abstract presented at: 2017 IEEE: 17th International Conference on Bioinformatics and Bioengineering; October 23, 2017; Washington, DC. doi:10.1109/BIBE.2017.00-37

10. Rana A, Yaunery G, Lowe A, Shah P. Computational histological staining and destaining of prostate core biopsy RGB images with generative adversarial neural networks. Abstract presented at: 2018 IEEE: 17th International Conference on Machine Learning and Applications; December 17-20, 2018; Orlando, FL.

11. Monici M. Cell and tissue autofluorescence research and diagnostic applications. *Biotechnol Annu Rev*. 2005; 11:227-256. doi:10.1016/S1387-2656(05)11007-2

12. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition; July 21-26, 2017; Honolulu, HI. doi:10.1109/CVPR.2017.632

13. National Instruments. Peak signal-to-noise ratio as an image quality metric. Updated March 5, 2019. Accessed April 10, 2020. https://www.ni.com/en-us/innovations/white-papers/11/peak-signal-to-noise-ratio-as-an-image-quality-metric.html

14. Zhang L, Zhang L, Mou X, Zhang D. FSIM: a feature similarity index for image quality assessment. *IEEE Trans Image Process*. 2011;20(8):2378-2386. doi:10.1109/TIP.2011.2109730

15. Thomos N, Boulgouris NV, Strintzis MG. Optimized transmission of JPEG2000 streams over wireless channels. *IEEE Trans Image Process*. 2006;15(1):54-67. doi:10.1109/TIP.2005.860338

16. Xiangjun L, Jianfei C. *Robust Transmission of JPEG2000 Encoded Images Over Packet Loss Channels*. ICME; 2007:947-950.

17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1): 159-174. doi:10.2307/2529310

18. Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. Paper presented at: 20th International Conference on Pattern Recognition; August 23-26, 2010; Istanbul, Turkey.

19. Potetz B, Lee TS. Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *J Opt Soc Am A Opt Image Sci Vis*. 2003;20(7):1292-1303. doi:10.1364/JOSAA.20.001292

20. Senaras C, Niazi MKK, Lozanski G, Gurcan MN. DeepFocus: detection of out-of-focus regions in whole slide digital images using deep learning. *PLoS One*. 2018;13(10):e0205387. doi:10.1371/journal.pone.0205387

21. Tu H, Liu Y, Turchinovich D, et al. Stain-free histopathology by programmable supercontinuum pulses. *Nat Photonics*. 2016;10(8):534-540. doi:10.1038/nphoton.2016.94

22. Fereidouni F, Harmany ZT, Tian M, et al. Microscopy with ultraviolet surface excitation for rapid slide-free histology. *Nat Biomed Eng*. 2017;1(12):957-966. doi:10.1038/s41551-017-0165-y

23. Tao YK, Shen D, Sheikine Y, et al. Assessment of breast pathologies using nonlinear microscopy. *Proc Natl Acad Sci U S A*. 2014;111(43):15304-15309. doi:10.1073/pnas.1416955111

24. Mertz J. Optical sectioning microscopy with planar or structured illumination. *Nat Methods*. 2011;8(10): 811-819. doi:10.1038/nmeth.1709

25. Lahiani A, Klaiman E, Grimm O. Enabling histopathological annotations on immunofluorescent images through virtualization of hematoxylin and eosin. *J Pathol Inform*. 2018;9:1. doi:10.4103/jpi.jpi_61_17

26. Orringer DA, Pandian B, Niknafs YS, et al. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nat Biomed Eng*. 2017;1(2):0027. doi:10.1038/s41551-016-0027

27. Petersen D, Mavarani L, Niedieker D, et al. Virtual staining of colon cancer tissue by label-free Raman micro-spectroscopy. *Analyst*. 2017;142(8):1207-1215. doi:10.1039/C6AN02072K

28. Bautista PA, Abe T, Yamaguchi M, Yagi Y, Ohyama N. Digital staining of pathological tissue specimens using spectral transmittance. Medical Imaging 2005. *Image Processing*. 2005;5747:1892-1903.

29. Bautista PA, Abe T, Yamaguchi M, Yagi Y, Ohyama N. Digital staining for multispectral images of pathological tissue specimens based on combined classification of spectral transmittance. *Comput Med Imaging Graph*. 2005; 29(8):649-657. doi:10.1016/j.compmedimag.2005.09.003

30. Bini J, Spain J, Nehal K, Hazelwood V, DiMarzio C, Rajadhyaksha M. Confocal mosaicing microscopy of basal-cell carcinomas ex vivo: progress in digital staining to simulate histology-like appearance. Paper presented at: Advanced Biomedical and Clinical Diagnostic Systems IX; February 21, 2011; San Francisco, CA.

31. Bayramoglu N, Kaakinen M, Eklund L, Heikkila J. Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks. Paper presented at: 2017 IEEE International Conference on Computer Vision; October 22-29, 2017; Venice, Italy.

32. Burlingame EA, Margolin AA, Gray JW, Chang YH. SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks. Paper presented at: SPIE Medical Imaging; February 10-15, 2018; Houston, TX.

33. Rivenson Y, Wang H, Wei Z, et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng*. 2019;3(6):466-477. doi:10.1038/s41551-019-0362-y

34. Rivenson Y, Wang H, Wei Z, et al. Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue. *Nat Biomed Eng*. 2019;3(6):466-477. doi:10.1038/s41551-019-0362-y

35. US Food and Drug Administration. Digital health. Accessed April 10, 2020. https://www.fda.gov/medical-devices/digital-health

36. US Food and Drug Administration. Digital health innovation action plan. Accessed April 10, 2020. https://www.fda.gov/media/106331/download

37. Shah P, Kendall F, Khozin S, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit Med*. 2019;2(1):69. doi:10.1038/s41746-019-0148-3

SUPPLEMENT.

eAppendix 1. Data Collection and Image Registration Process
eTable 1. Data Distribution
eAppendix 2. Loss Function
eAppendix 3. Interpretation
eFigure 1. Color Coded Overlaid Validation Images
eFigure 2. Visualization and Explanation of Computational Hematoxylin and Eosin Staining Process by Custom Autoencoder Neural Network
eFigure 3. Activation Maps of Kernels of Trained Generator Neural Network Model Layers
eFigure 4. Activation Maps of Kernels of Various Generator Neural Network Layers After Entering Hematoxylin And Eosin Dye–Stained Patch With Gleason Grades 4 And 5 Prostate Tumor
eFigure 5. Comparison of Mean Squared Errors Between Kernel Activation Maps of Pairs of 448 Validation Image Patches Generated by the Trained Neural Network Models
eTable 2. Mean Pixel Intensity Following Computational Staining and Destaining
eTable 3. Change in Mean Pixel Intensity in Red, Green, and Blue Channels per Image
eTable 4. Intrarater Agreement Calculated on Dye Stained and Computationally Stained Images
eAppendix 4. Evaluation of the Activation Maps of Trained Deep Neural Network
eTable 5. Comparison of Tumor Grades Between Original Expert Microscopic Diagnosis
eAppendix 5. Comparison With Patient Records
eReferences

Supplementary Online Content

Rana A, Lowe A, Lithgow M, et al. Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. *JAMA Netw Open*. 2020;3(5):e205111. doi:10.1001/jamanetworkopen.2020.5111

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Data Collection and Image Registration Process

Data collection, transfer and processing of whole slide images: Thirty-eight patients (mean age 66.2 years) consisting of White, African American, Hispanic/Latino, and Asian men provided forty-six core biopsy samples. Of these, nine patients had known prostate cancer diagnosis and were undergoing active surveillance. Eighteen patients underwent subsequent prostatectomy and the remaining were either healthy or undergoing prostate cancer treatment at Brigham and Women's Hospital. Each biopsy sample contained one to six cores of tissue. Zero to 100% of each tissue core contained prostatic adenocarcinoma of various Gleason grades. Samples were enriched for higher-grade tumors (Gleason grade 4 and 5). Forty-six non-stained and corresponding H&E dye stained RWSI were collected from 38 patients and imaged at 20x magnification. Briefly, prostate core biopsy specimens were immediately fixed in 10% formalin, paraffin embedded, cut into 4-micron thick sections and placed on standard glass slides that were placed in archival storage at room temperature. Deparaffinized non-stained slides were scanned with the Aperio ScanScope XT system (Leica Biosystems, Buffalo Grove, IL) at 20x magnification. Subsequently, the slides were stained with H&E dye on the Agilent Dako Autostainer (Agilent, Santa Clara, CA), and these stained slides were re-scanned on the Aperio ScanScope XT at 20x magnification at Harvard Medical School Tissue Microarray & Imaging Core. Deidentified data in the form of non-stained and H&E dye stained images at 20x magnification were analyzed at Massachusetts Institute of Technology. Individual prostate tissue needle core biopsy images from each whole slide image were extracted. Extracted core images were horizontally or vertically rotated to reduce non-tissue pixels. This resulted in 102 high-resolution native non-stained and H&E dye stained image pairs.

Image registration and processing: Deparaffinized single core images (henceforth called as non-stained images) and subsequent H&E dye stained single core images of the same biopsy (henceforth called as H&E dye stained images) were registered using Photoshop CC software (Adobe Systems, San Jose, CA) and corrected for variances.[1,2] Tissue shearing during the staining procedure resulted in regions that could not be registered that were cropped and discarded.

eTable 1. Data Distribution

Patient	Age	Demographics	Tumor	Core (train, test)	Training Patches	Test-G3 Patches	Test-G4 Patches	Test-G5 Patches
1	84	Dominican: Hispanic	G3	6,1	2023	21	-	-
2	59	Unavailable: White	G4, G5 (G4>G5)	2,2	1977	-	-	-
3	61	Unavailable: White	B9, G3, G4 (G3>G4)	3	1068	-	-	-
4	71	European: White	G3	5	1244	-	-	-
5	71	Unavailable: White	G3, G4 (G3>G4)	2	2119	-	-	-
6	71	Unavailable: White	G4	2	1602	-	-	-
7	72	Unavailable: White	G4	3,4	1608	-	292	-
8	72	European: White	B9, G3, G4 (B9>>G3/G4)	3	173	-	-	-
9	59	American: White	B9	1,3	1825	-	-	-
10	73	American: White	G3	4	82	-	-	-
11	60	Unavailable: White	G3	2,2	3522	-	-	-
12	68	Unavailable: White	G3, G4, G5, (G4> G3/G5)	2	1556	-	-	-
13	53	Unavailable: White	G3, G4	2,5	2930	300	744	-
14	51	Unavailable: Black /African American	B9, G4	6	3369	-	-	-
15	72	Unavailable: White	G3, G4	1,3,4	3297	2274	1738	-
16	84	Unavailable: White	G4, G5 (G4>G5)	3,2	5083	-	220	270
17	77	Declined: White	G4	5	4094	-	-	-
18	56	Filipino: White	G3, G4 (G4> G3)	2,6	1161	738	1133	-
19	64	American: White	B9, G3, G4 (G3>G4), Rare G5	2,4	543	505	1351	-
20	63	Unavailable: White	G4, G5 (G4 >G5)	2	716	-	-	-

21	67	Black/African American	G3, G4 (G4 >G3)	6	3549	-	-	-
Patient	**Age**	**Demographics**	**Tumor**	**Core (train, test)**	**Training Patches**	**Test-G3 Patches**	**Test-G4 patches**	**Test-G5 patches**
23	52	American: White	G3, G4 (G3>>G4)	5,6	5714	-	-	-
24	79	American: White	G3, G4, G5, (G4> G3/G5)	2	1923	-	-	-
25	62	Unavailable: White	G3, G4 (G3>>G4)	3	1661	-	-	-
26	63	Unavailable: White	G3, G4 (G3>>G4)	2	2725	-	-	-
27	77	Black/African American	G3, G4 (G3>G4)	5,6	3586	-	-	-
28	53	American: White	G4, G5 (G4>G5)	2	1331	-	-	-
29	75	American: White	G3	1,2	905	1351	-	-
30	62	Unavailable: White	G3	2	2025	-	-	-
31	65	White/ Middle Eastern	G3	3	2151	-	-	-
32	67	African: White	G3	1, 2	1005	920	-	-
33	68	Dominican: Other	G4, G5 (G4 > G5)	2	947	-	-	-
34	73	American: White	G4	2,6	1857	-	536	-
35	64	American: White	B9, G3	3	1148	-	-	-
36	76	Declined: White	G4, G3 (G4 >G3)	2	810	-	-	-
37	54	Black / African American	G3	1,2	1151	831	-	-
38	67	Other: White	G3	1	870	-	-	-
Total	**-**	**-**	**-**	**-**	**78,322**	**7,019**	**6,146**	**270**

eAppendix 2. Loss Function

Consider I_u and I_s respectively represent the native non-stained and H&E stained image patches in the training dataset. The generator takes in I_u as the input and generates I_{cs}, the corresponding computationally stained image patch, as the output. The discriminator analyses the output image, I_{cs}, and predicts the probability that I_{cs}, is real (from the training dataset) or fake (output from generator). The loss function consisted of the cGAN loss,[3] a L1 component and a PCC factor between I_s and I_{cs}. The loss equation was:

$$\mathfrak{L}_{cGAN}(G,D) = E_{x,y}[logD(x,y)] + \alpha E_{x,y}[\log\left(1 - D\big(x, G(x,z)\big)\right)]$$

$$\mathfrak{L}_{cGAN}(G,D) = E_{x,y,z}[\|y - G(x,z)\|_1]$$

$$\mathfrak{L}_{PCC}(G) = E_{x,y,z}[PCC\big(y, G(x,z)\big)]$$

The final loss function is:

$$G^* = arg\min_G \max_D \mathfrak{L}_{cGAN}(G,D) + \lambda\mathfrak{L}_{L1}(G) + \gamma\mathfrak{L}_{PCC}(G)$$

Where *x* is the input image, *y* is the target image and *z* is the random noise, added as dropout in our work. $\mathfrak{L}_{cGAN}(G,D)$ is the $cGAN$ loss function, $\mathfrak{L}_{L1}(G)$ is the $L1$ loss between the output of the generator and the target image, and $\mathfrak{L}_{PCC}(G)$is the proposed term that calculated the Pearson's correlation coefficient between the generator output and target image. α=1, λ=100 and γ =10 gave best results. After training, the model accepted unseen native non-stained image patches and generated computationally H&E stained images patches.

Technical implementation of the CGAN model: The discriminator was trained after every single training step for the generator. Both networks were trained for 10 epochs each using Adam optimization[12], and a batch size of one on a NVIDIA GeForce 1080 TI GPU (NVIDIA, Santa Clara, CA) with 12 GB of VRAM and CUDA acceleration to speed up training. One epoch of training took approximately 16 GPU hours. The patches were randomly flipped and dropped out to prevent overfitting and increase generalization capability of the model.

eAppendix 3. Interpretation

Methods: Non-overlapping patches were cropped from these four sets of images, resulting in four sets of around 2000 patches (of size 1024x1024 pixels each). Patches not containing any tissue were discarded, resulting in 448 patches for each set. These four sets of patches were fed into trained staining and destaining models. Nonstained and destained patches were fed into computational staining model, while computationally and H&E stained patches were fed into the computational destaining model. The activation maps generated after each layer block (lrelu – conv2d - batch norm) were saved for further analysis. Activation maps for selected patches were analyzed to understand the transformation of input images as they passed through the generator neural network layer, and to identify which convolutional kernels are activated. Activation maps generated by the top five most activated kernels were extracted, ranked and visualized by heatmaps. Activation maps for each layer were ordered in descending order (ranked by numbers of pixels of intensity greater than 200). Top five activation maps (heatmap) from each layer are shown in Figure 3 (benign input patch) and eFigure 3 (Gleason grade 3 tumor patch). Heatmaps were the standard 'jet' heatmaps plotted using matplotlib library in python.

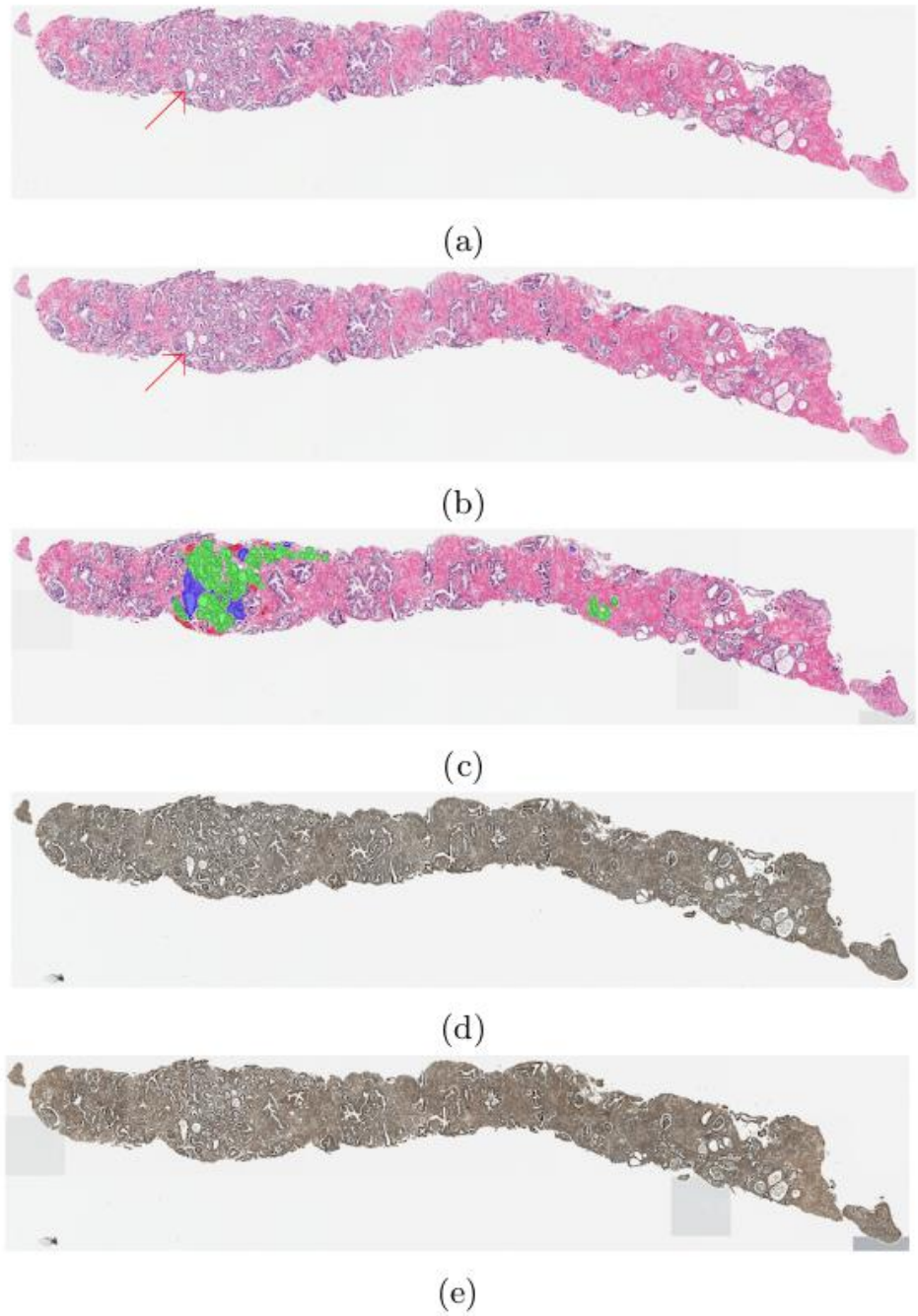
The activation maps, for an input image, for each layer were rescaled to 0-1 and resized to a standard size (128x128 pixels) and concatenated into a grid (16 by K grid, where k is variable and is equal to number of patches divided by 16). This process was done for every one of the 448 patches for all four image sets. The activation maps for the matching patches (unstained-destained and computationally stained-H&E stained input patch pairs) were compared and the mean squared error (MSE) was calculated. The MSE plot for the two sets of matching input patches can be seen in eFigure 5.

Results: The top five activation maps (heatmaps) obtained from the first five layers (rows I-V) and the last four layers (rows 16-19) of the generator (excluding the input and output layers) are presented in Figure 3 and eFigure3 and 4. The first five rows represent the first 5 layers of the generator and the last 4 rows represent the last four layers of the generator. It is evident from the activation heatmaps that the network does a good job of extracting the tissue from the background (eFigure 3, and 4: L1-I and L1-III). eFigure 3 and 4: L1-II, L1-III and L2-II, L2-IV and L2-V indicated that the model learnt to recognize some features in the tissue. As the image passed through the layers, we can see localized high activations (bright red and orange color), indicating region of interest learnt by the trained generator. While the initial layers learn low level features like tissue, background, circles and simple patterns, deeper layers learn to recognize high-level patterns (using a combination of the knowledge accumulated by the previous layers). It is difficult to understand these visualizations and they look like noise to the human eye. Layers 16 through 19 show the activation heatmap from the decoder side of the generator (Figure 3 and efigures 3 and 4). These layers try to decode the encoded information back to the original size (1024x1024 output), while in the process computationally staining the image."

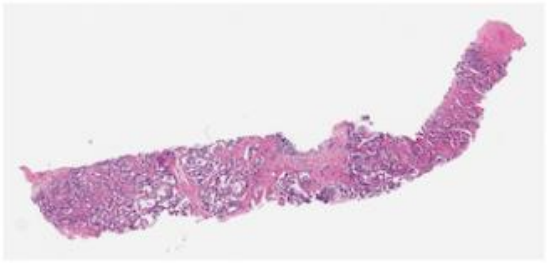
In eFigure 5 the blue lines represent the flow of MSE value through the layers for all 448 patch pairs unstained-destained (left) and H&E stained - computationally stained (right). The green and orange lines are the first and the third quartile values. We can see the variance between the MSE values at all the layers for the 448 patches. The variance is higher for the encoder layers and is less for the decoder layers. This the case for both unstained-destained plot and H&E stained - computationally stained plot. Peaks were observed at layers 3, 10 and 17.

eFigure 1. Color Coded Overlaid Validation Images

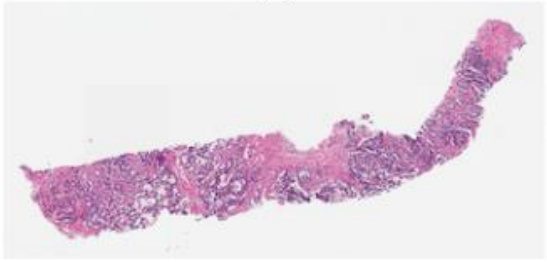
In Figures 1- 13, each figure contains (a) Ground truth Hematoxylin and Eosin (H&E) dye stained RGB Whole Slide Images (RWSI); (b) Computationally H&E stained RWSI; (c) Computationally H&E stained RWSI overlaid with colors representing comparisons between true positive (green), false positive (red) and false negative (blue) of IOU of healthy and tumor annotations provided by five physicians on ground truth H&E dye and computationally H&E stained prostate core biopsy images; (d) Ground truth native non-stained RWSI image; (e) Computationally destained RWSI.



eFigure 1.1 (a) Ground truth H&E dye stained RWSI with arrow highlighting atypical gland indeterminate for malignancy; (b) Computationally H&E stained RWSI with arrow showing atypical gland with preserved morphology as shown on image in panel (a).

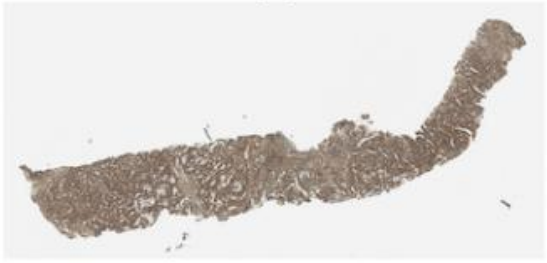


(a)

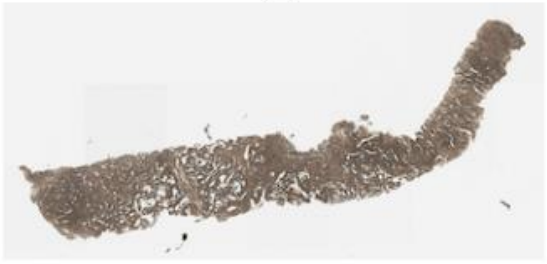


(b)

(c)

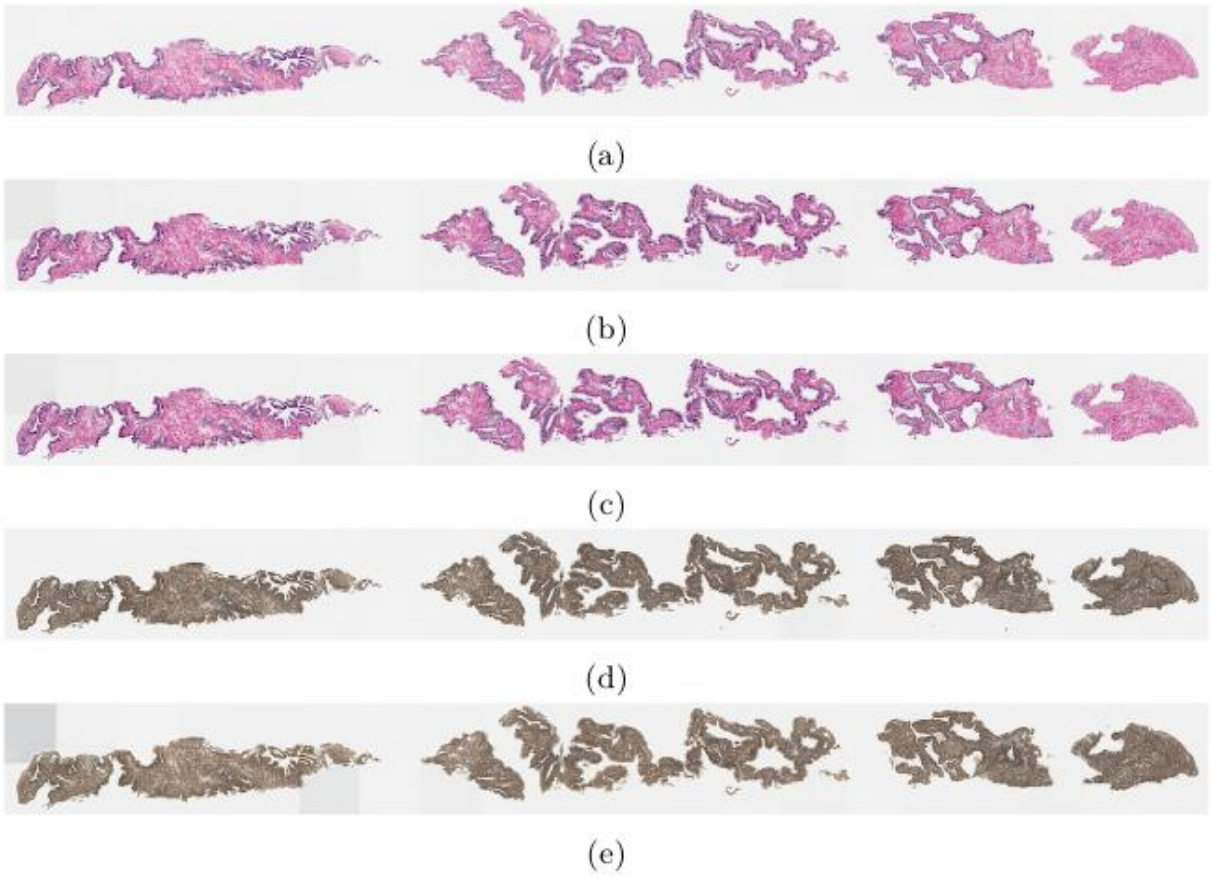


(d)

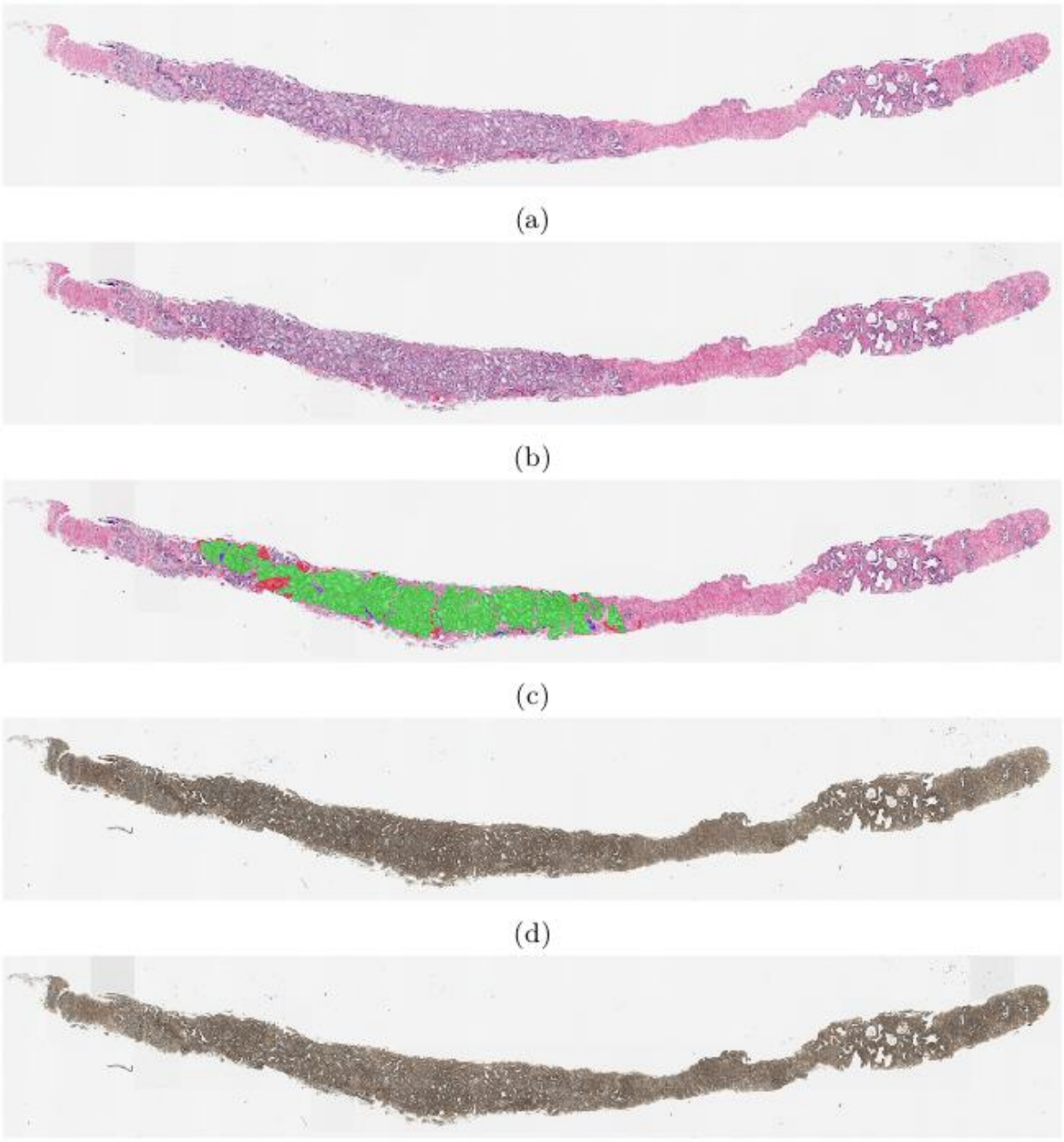


(e)

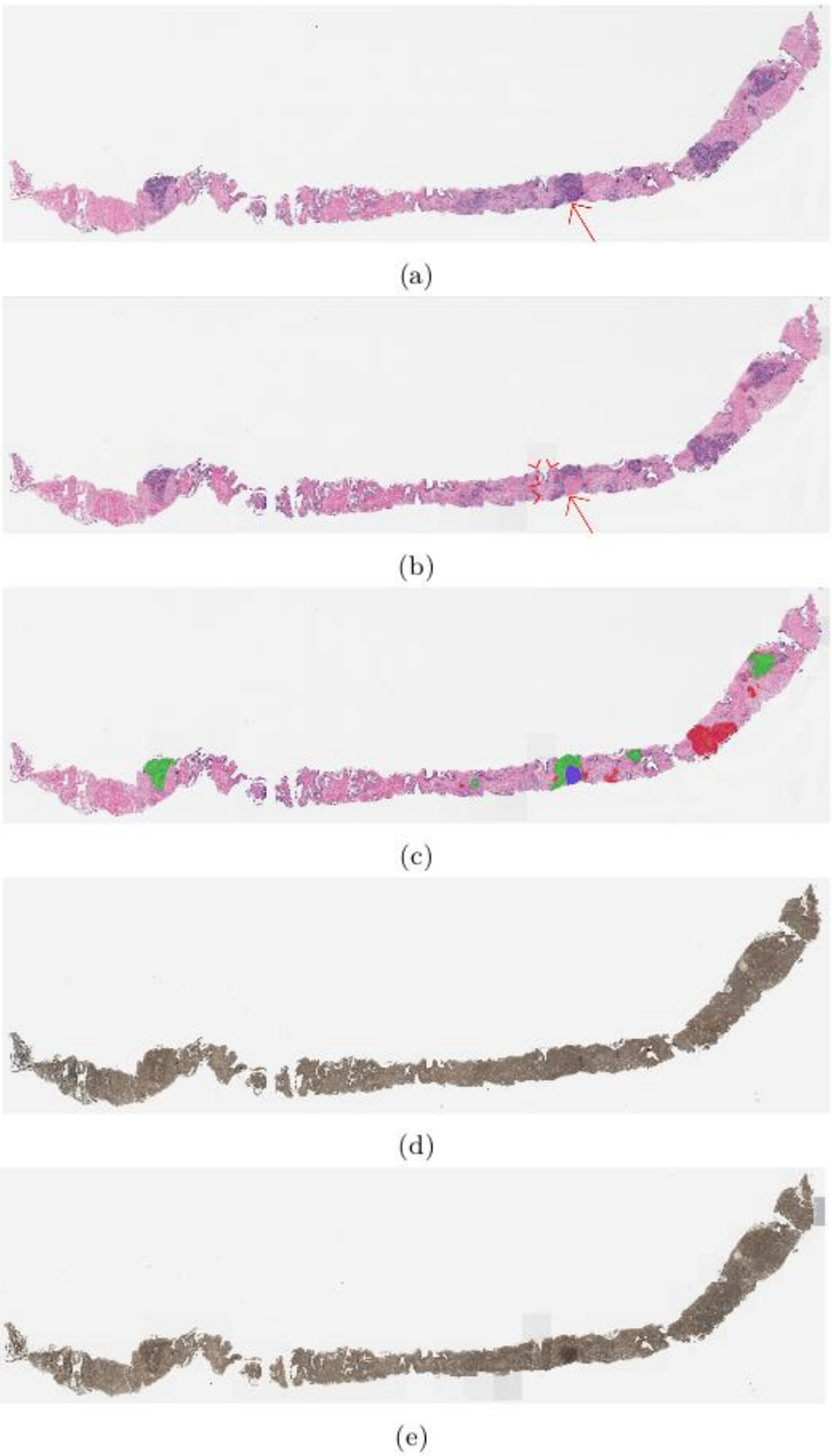
eFigure 1.2



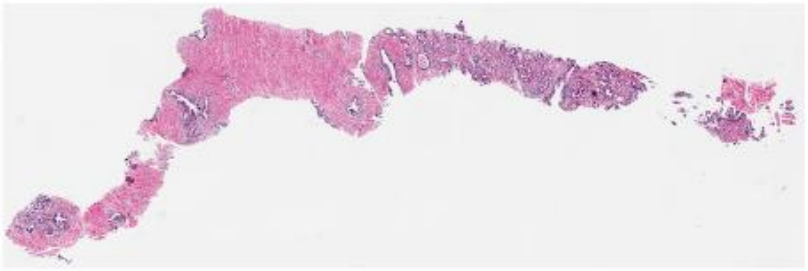
eFigure 1.3



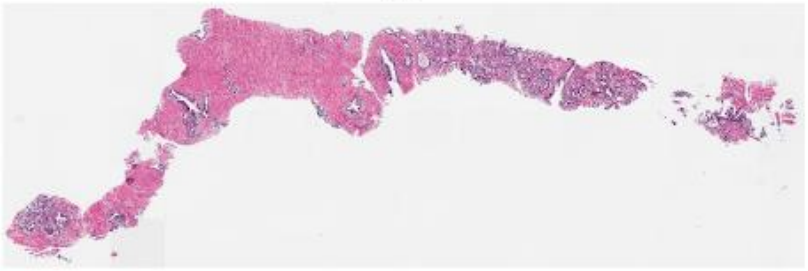
eFigure 1.4



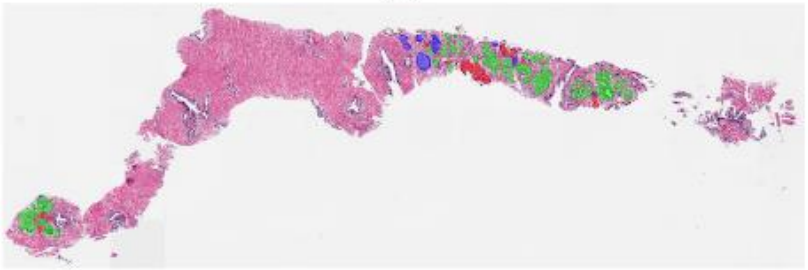
eFigure 1.5 (a) Ground truth H&E dye stained RWSI showing comedo necrosis (arrow); (b) Computationally H&E stained RWSI where malignant glands are well preserved (arrowheads), comedo necrosis is not evident (arrow).



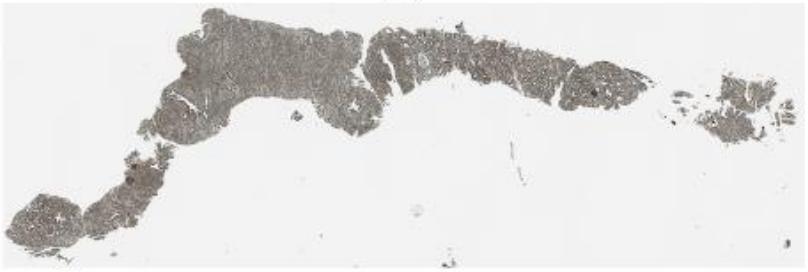
(a)



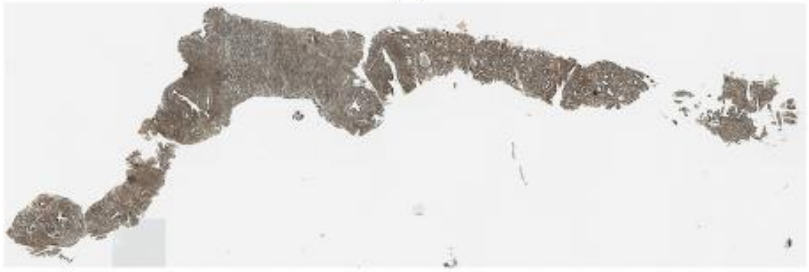
(b)



(c)

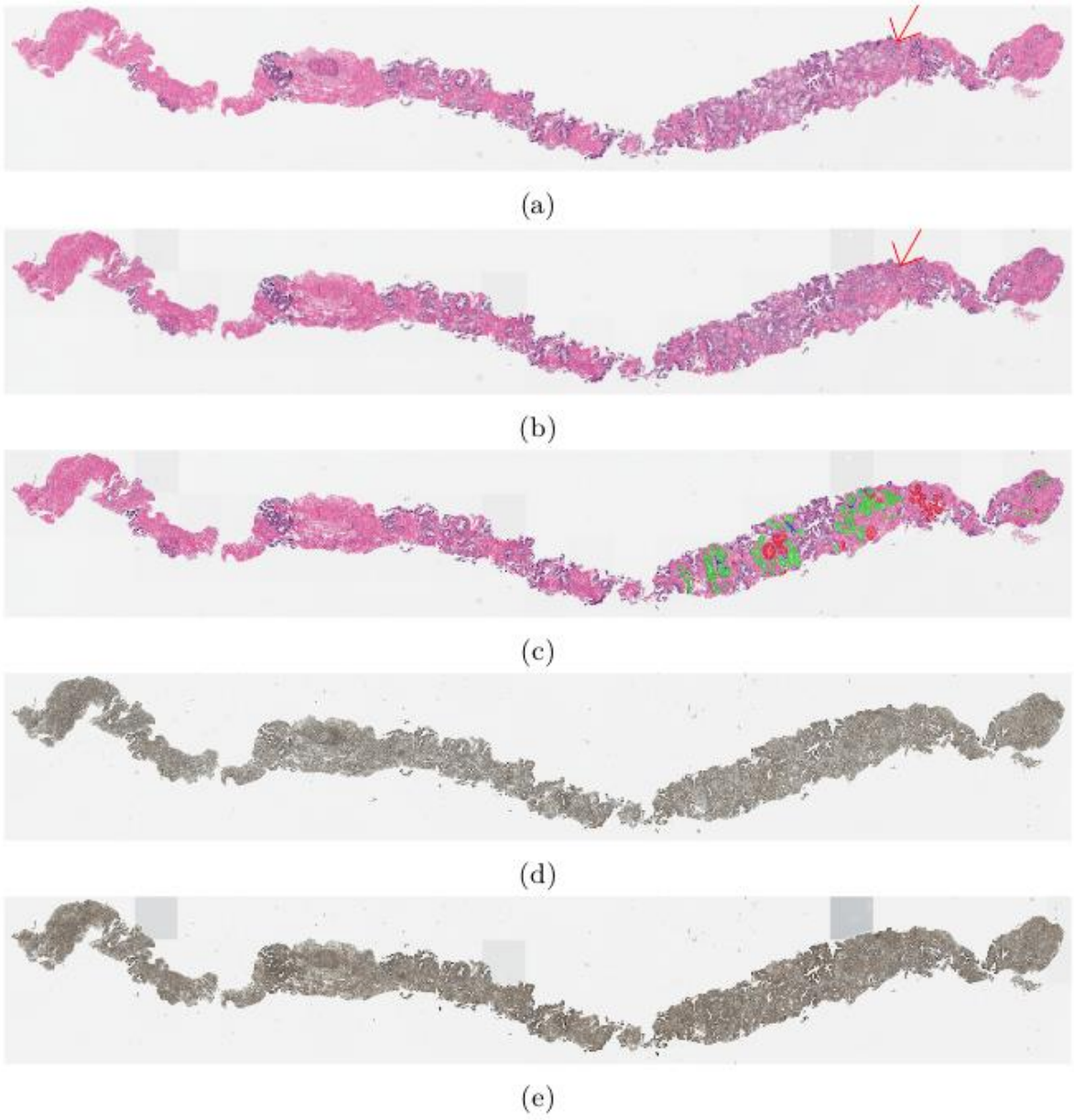


(d)

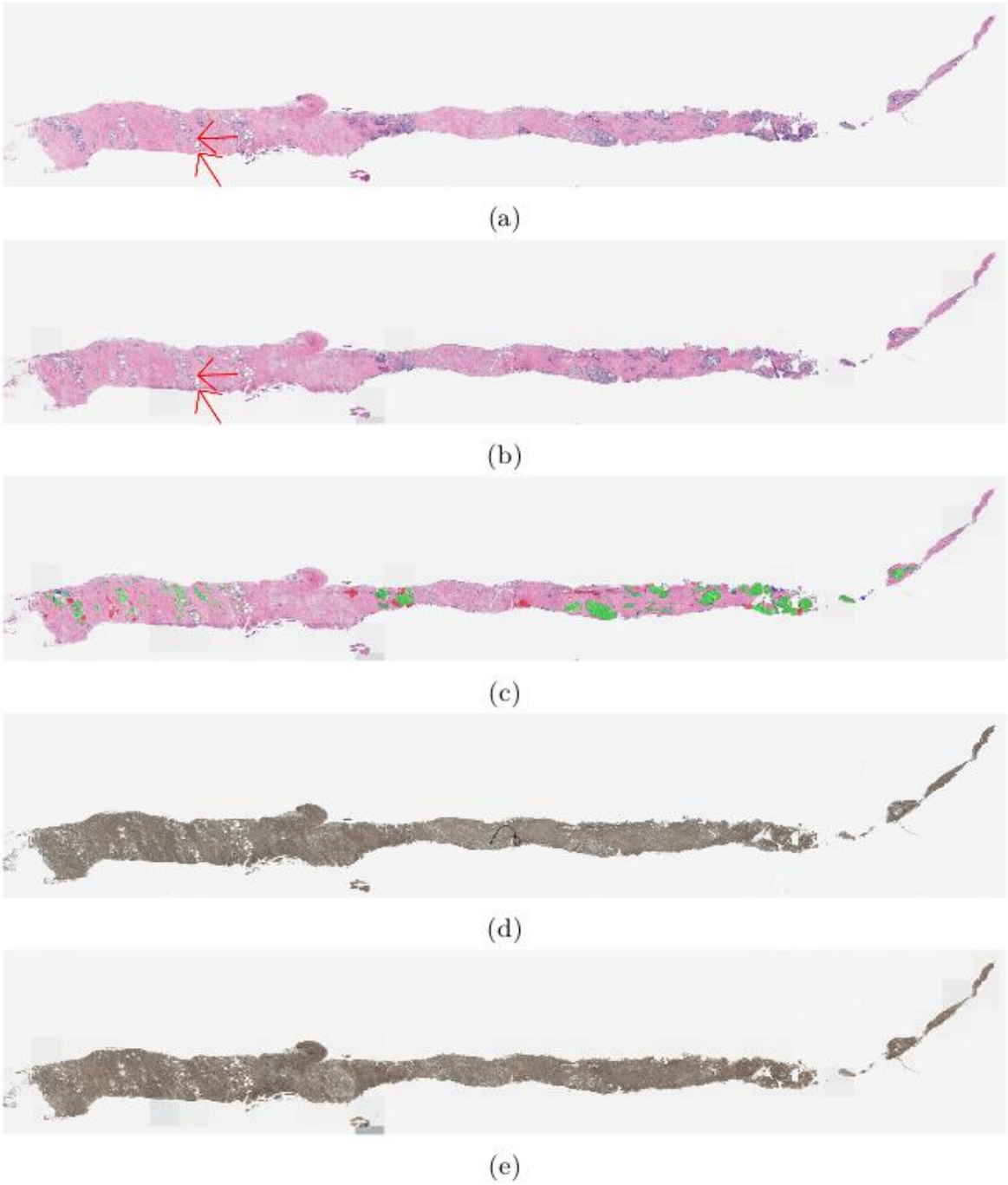


(e)

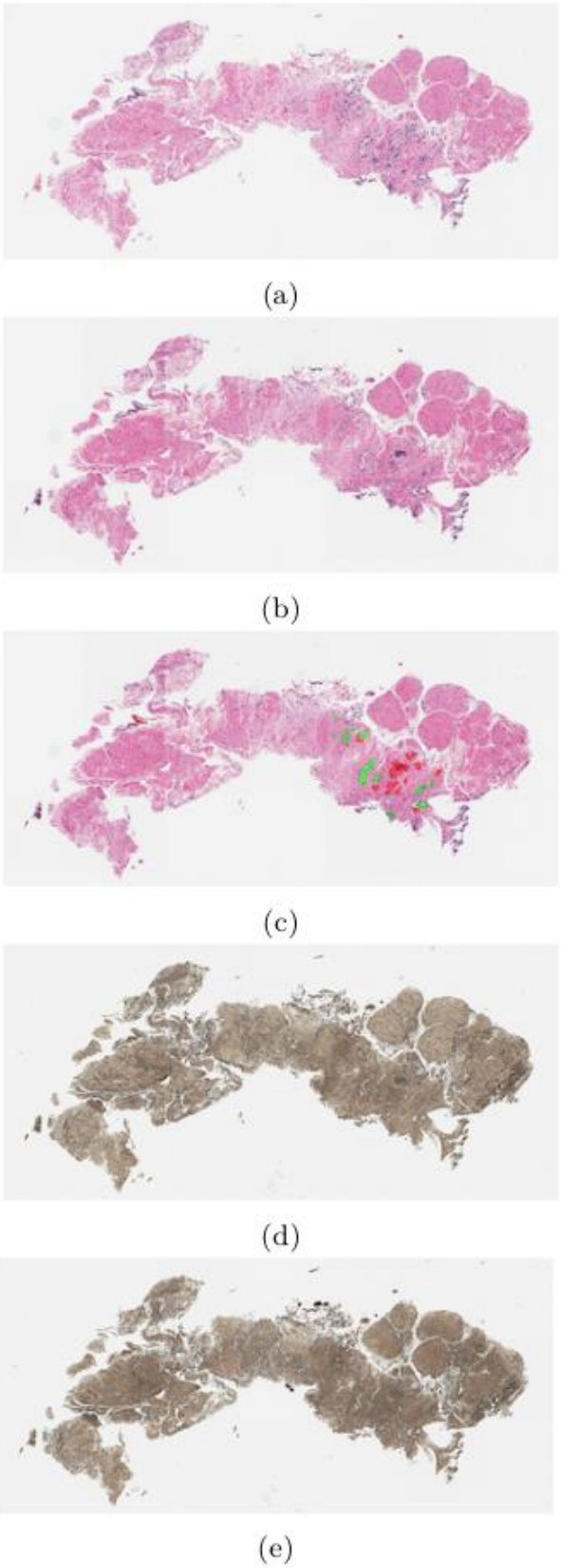
eFigure 1.6



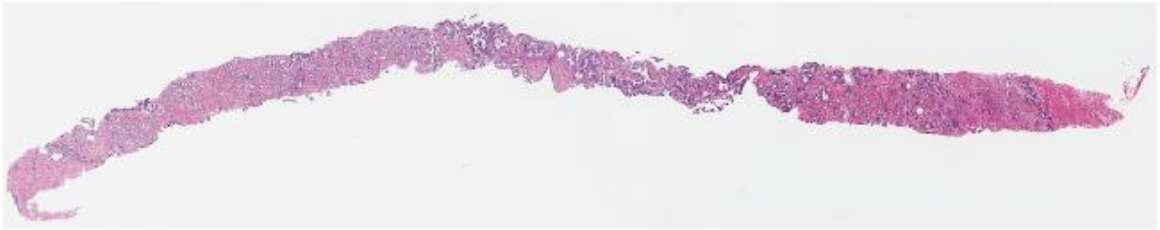
eFigure 1.7 (a) Ground truth H&E dye stained RWSI with arrow showing tumor; (b) Computationally H&E stained RWSI with arrow showing tumor.



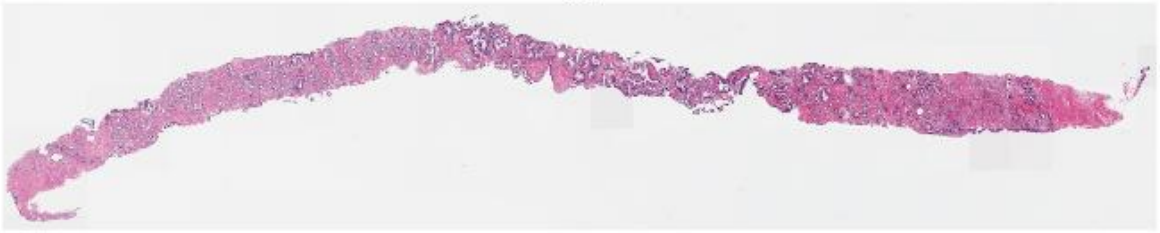
eFigure 1.8 (a) Ground truth H&E dye stained RWSI with arrows indicating rare poorly formed glands; (b) Computationally H&E stained RWSI with arrows indicating rare poorly formed glands.



eFigure 1.9



(a)



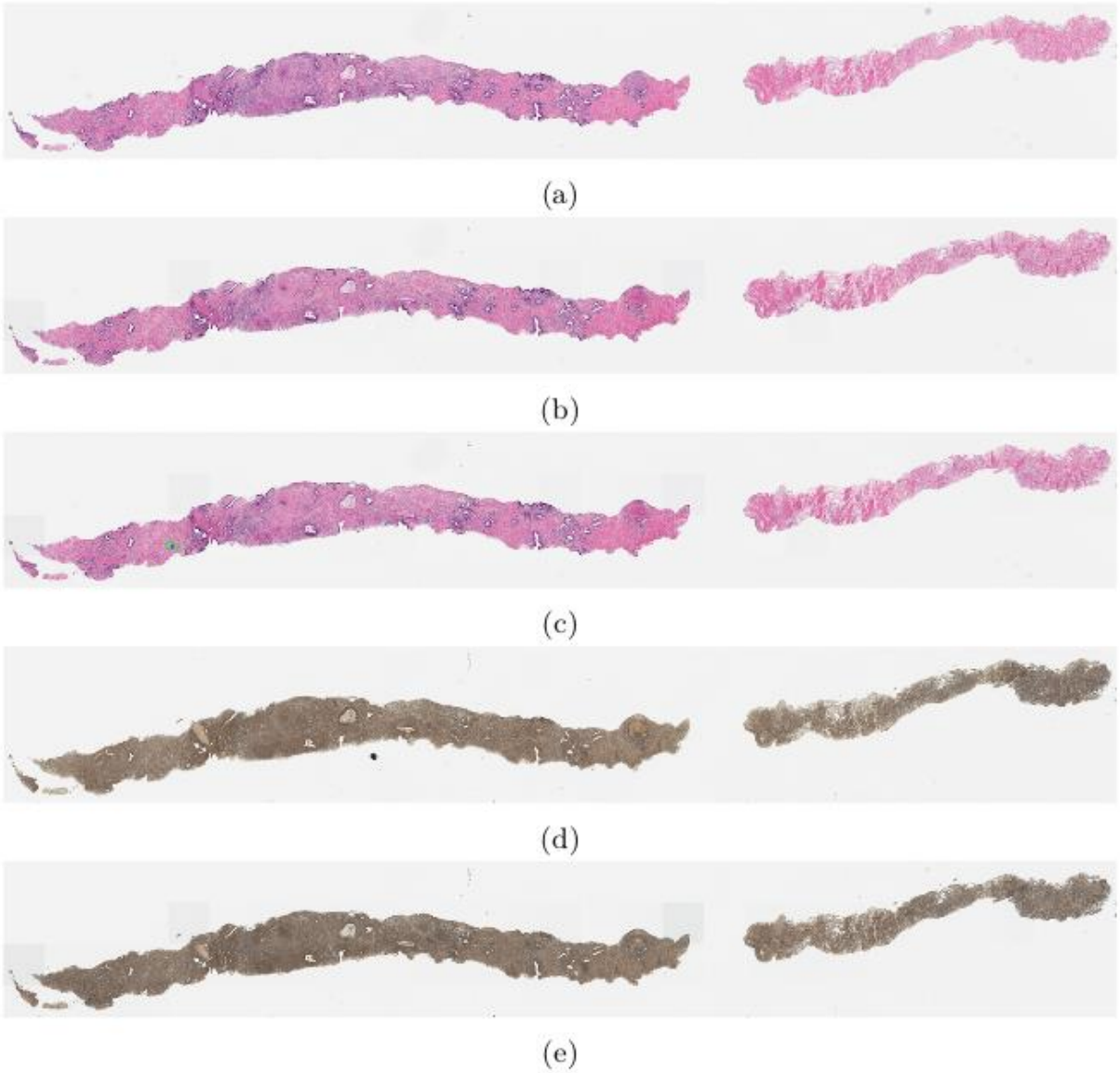
(b)

(c)

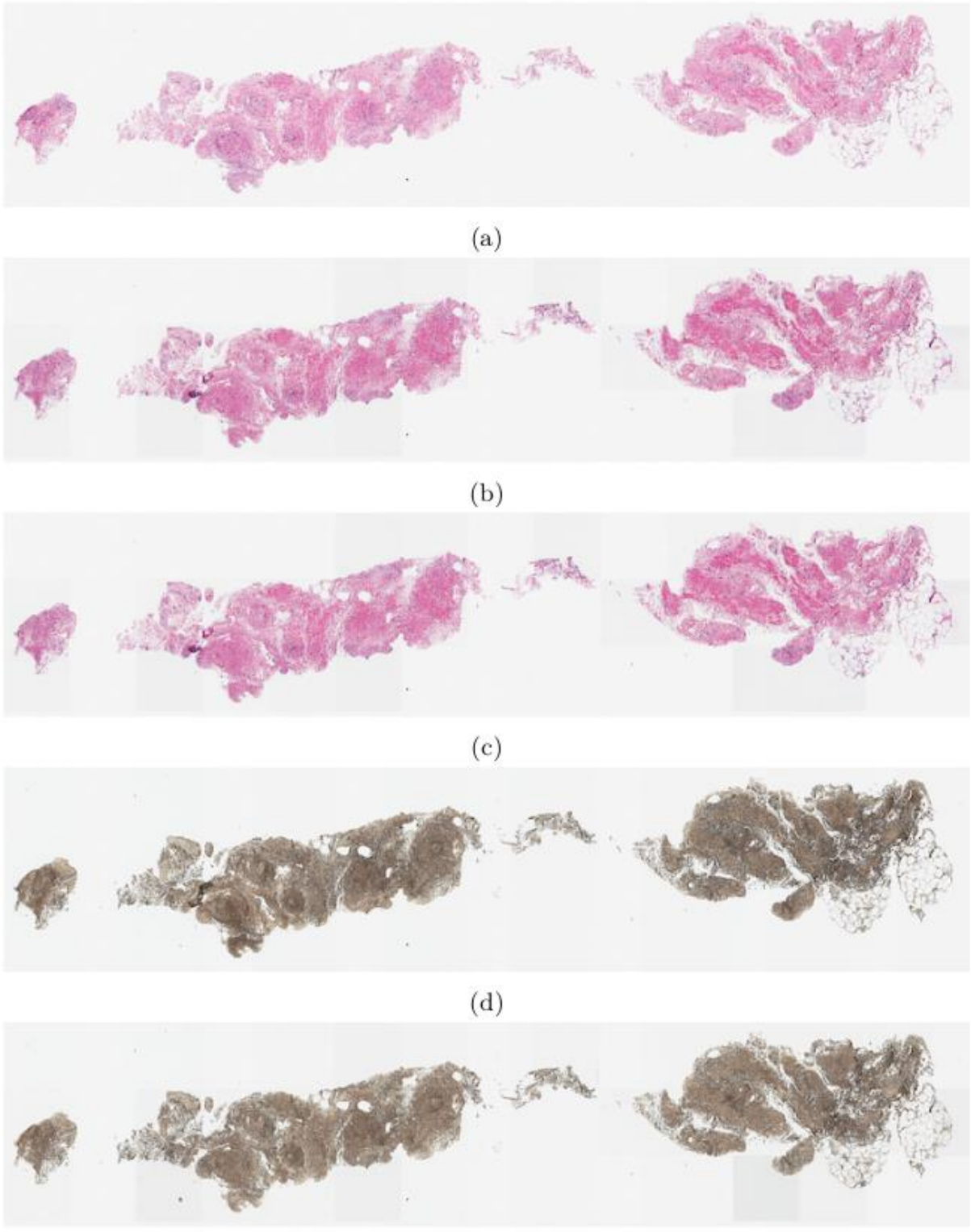
(d)

(e)

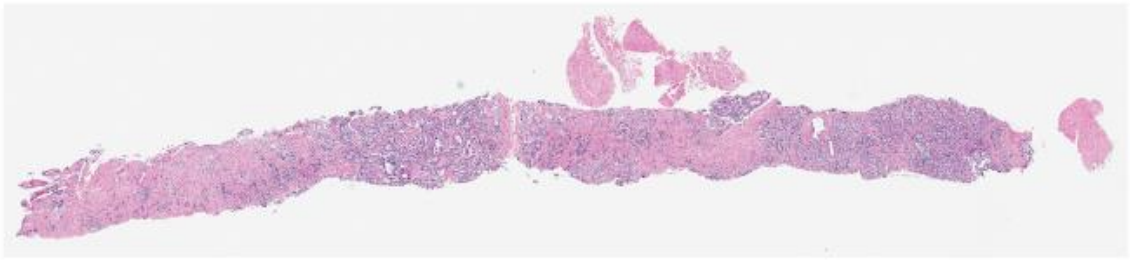
eFigure 1.10



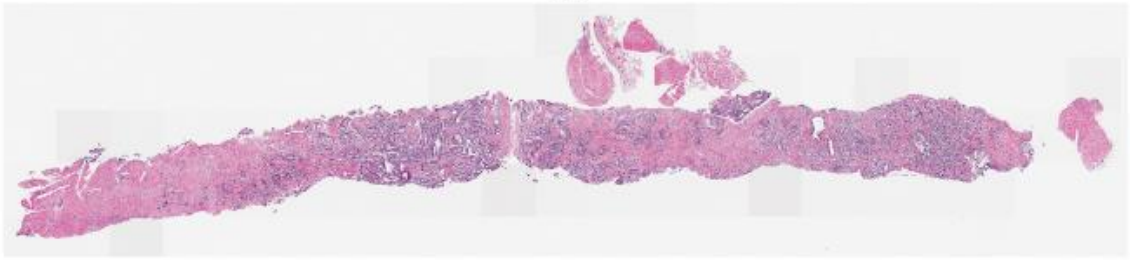
eFigure 1.11



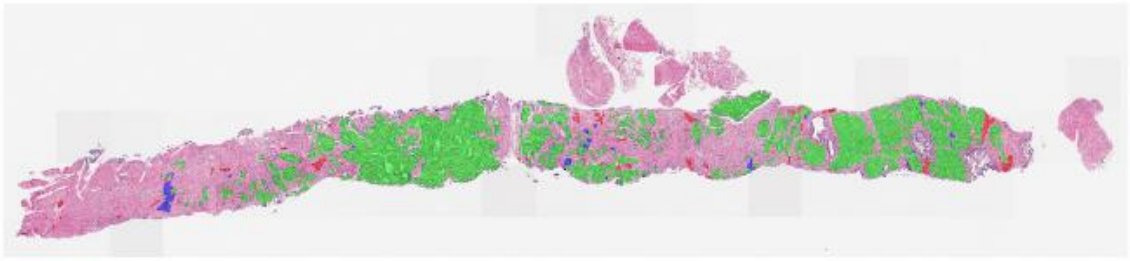
eFigure 1.12



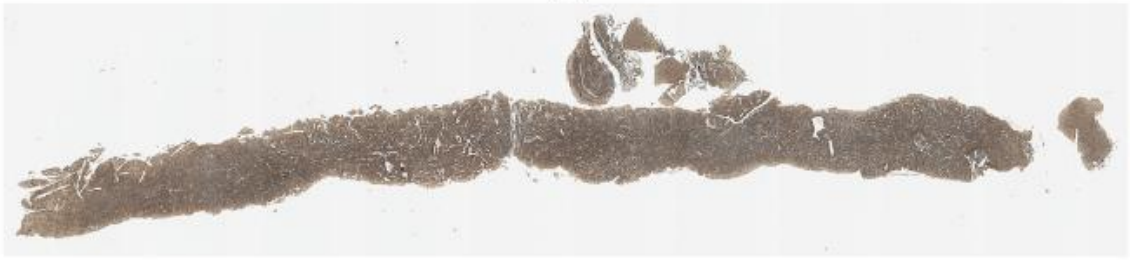
(a)



(b)



(c)



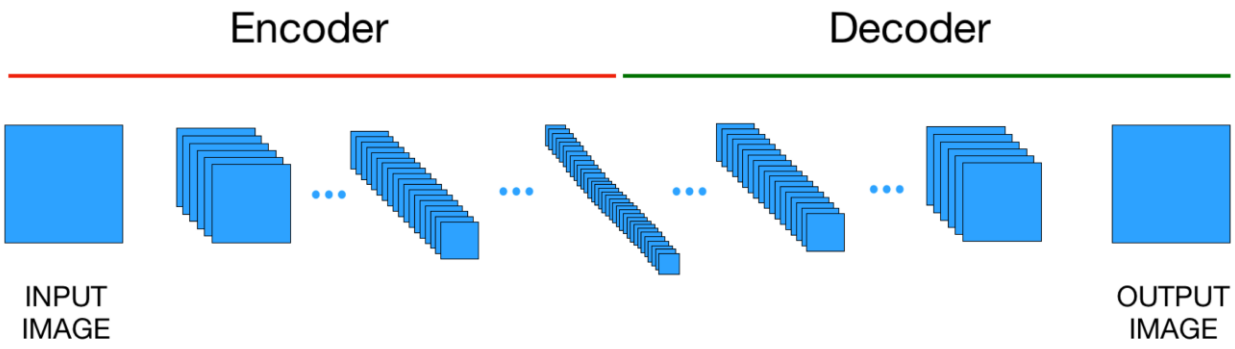
(d)

(e)

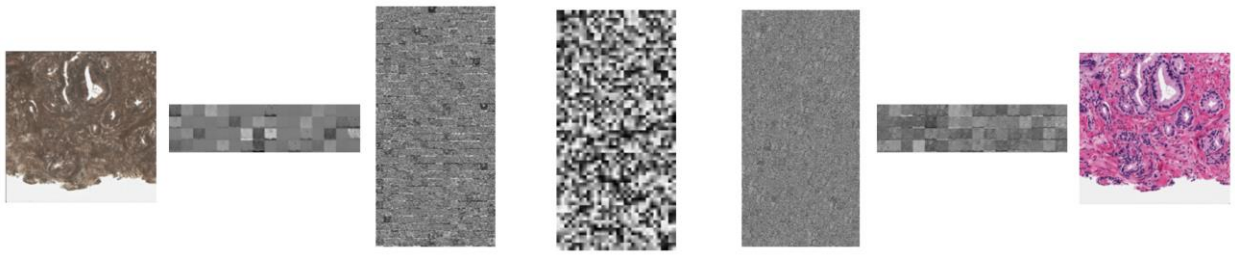
eFigure 1.13

eFigure 2. Visualization and Explanation of Computational Hematoxylin and Eosin Staining Process by Custom Autoencoder Neural Network

Panel (a) Processing of native non-stained prostate core biopsy images as various layers of the encoder and decoder neural networks computationally stain them. The blue boxes represent hidden activation layers of the neural network. Panel (b) A single input native non-stained patch and representative concatenated activation maps from the corresponding hidden layers in panel (a) of kernels of the decoder neural network as it flows through them, are shown.



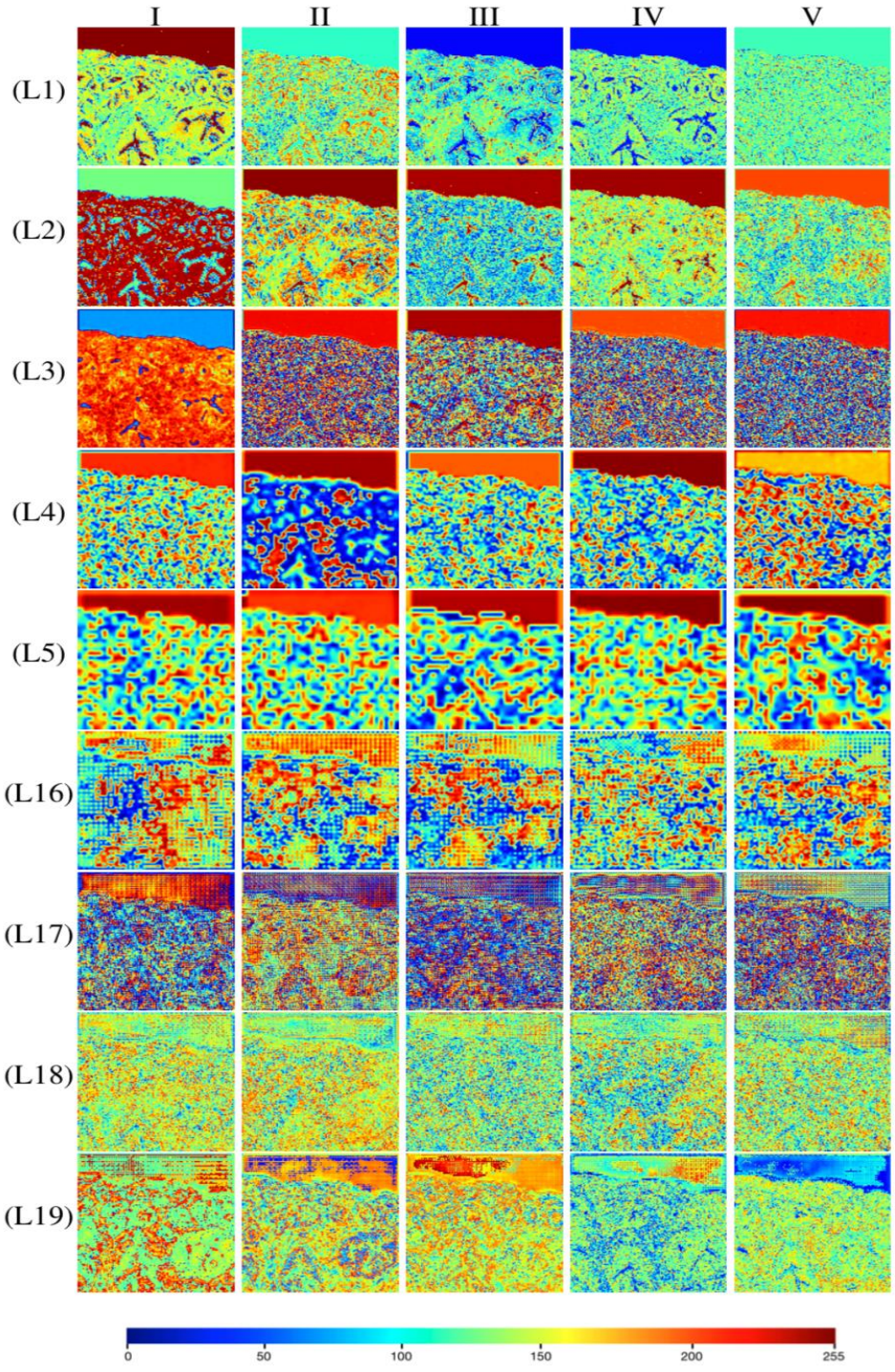
(a)



(b)

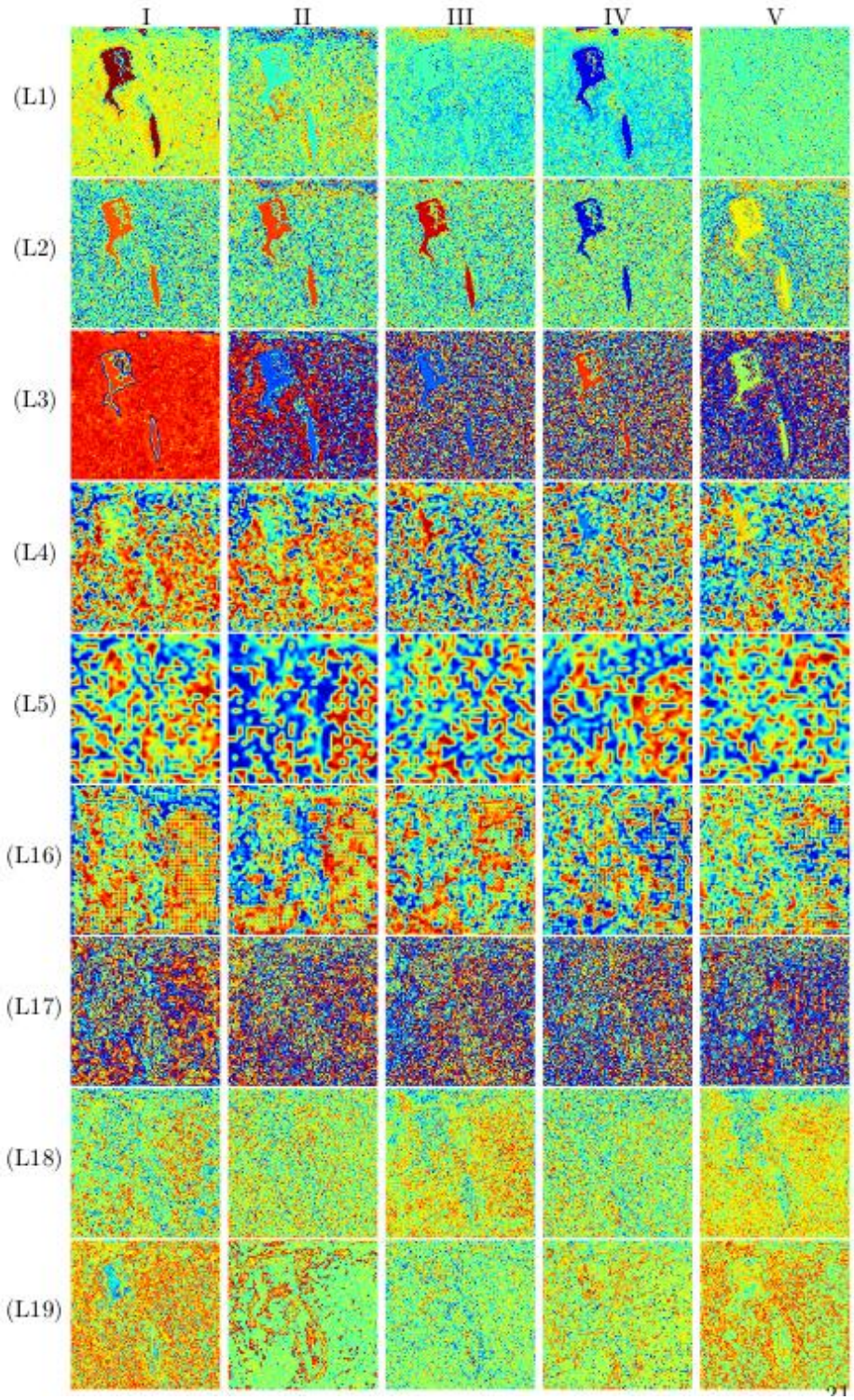
eFigure 3. Activation Maps of Kernels of Trained Generator Neural Network Model Layers

Activation maps of kernels of trained generator neural network model layers after feeding a native non-stained prostate core biopsy image patch with Gleason grade 3 tumor as it gets computationally Hematoxylin and Eosin stained. Rows show top five activation maps from layers L1 - L5 and L16 - L19 arranged in decreasing order of their activations from left to right (columns I-V).

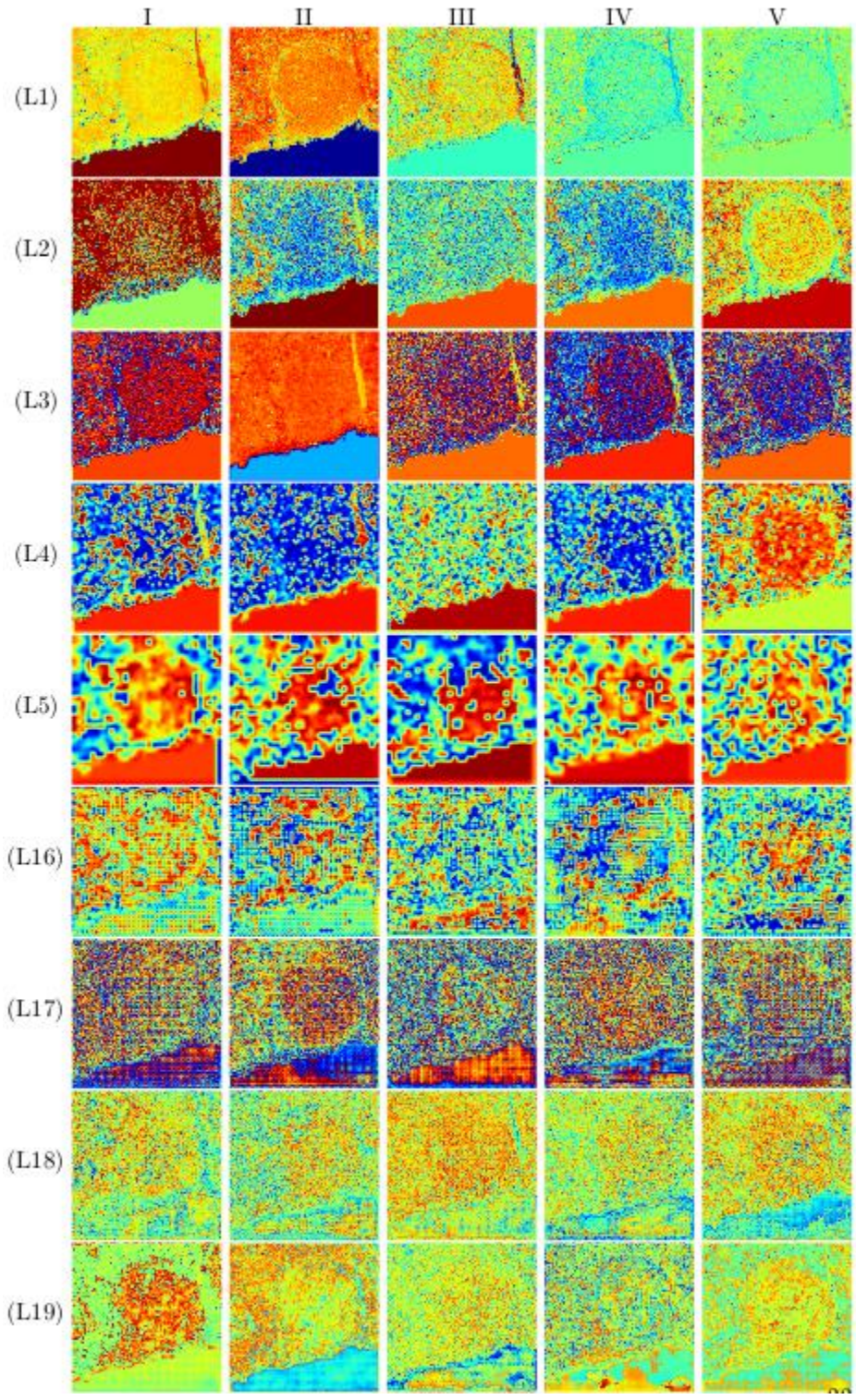


eFigure 4. Activation Maps of Kernels of Various Generator Neural Network Layers After Entering Hematoxylin And Eosin Dye–Stained Patch With Gleason Grades 4 And 5 Prostate Tumor

Panel (a) and (b) show Gleason grade 4 and 5 respectively. Rows show top five activation maps from layers L1 - L5 and L16 - L19 arranged in decreasing order of their activation from left to right (columns I-V).



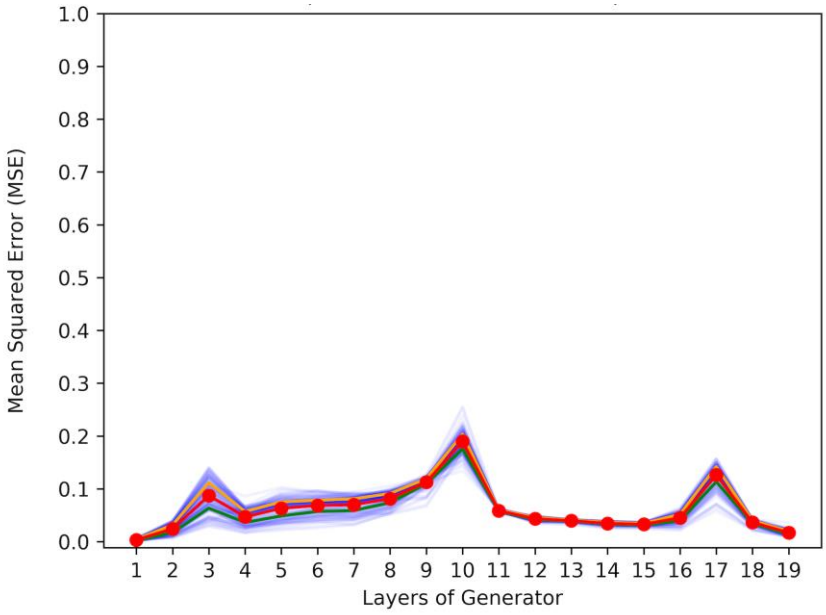
(a)



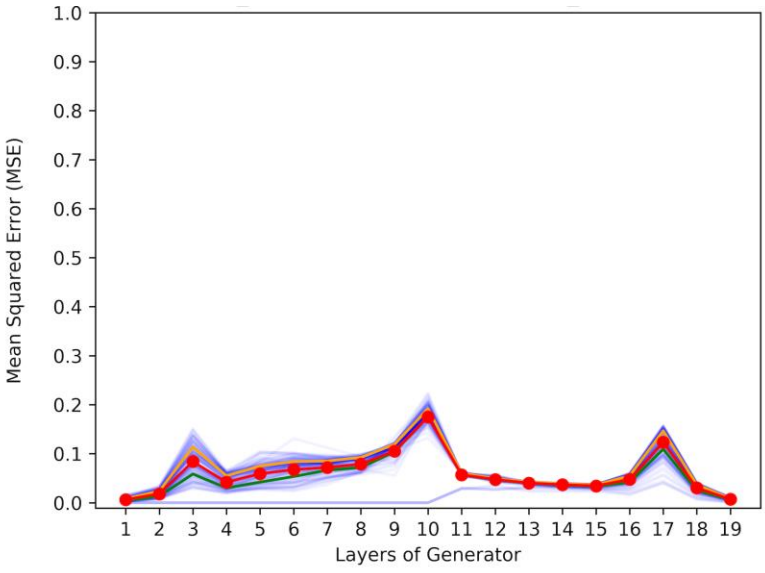
(b)

eFigure 5. Comparison of Mean Squared Errors Between Kernel Activation Maps of Pairs of 448 Validation Image Patches Generated by the Trained Neural Network Models

(a) MSE of ground truth native non-stained and corresponding computationally destained input patch activation maps generated by the trained computational staining model; (b) MSE of computationally Hematoxylin and Eosin (H&E) stained – and corresponding ground truth H&E dye stained matching input patch activation maps generated by the trained computational destaining model. Blue lines represent the MSE values for each of the 448 input pairs. Red curve represents average MSE value at each layer of the generator for all input pairs. The green and orange curves represent the first and third quartile MSE values for all input patch pairs. Lower MSE indicates more accuracy between the activation maps being compared.



(a)



(b)

eTable 2. Mean Pixel Intensity Following Computational Staining and Destaining

Average pixel intensity differences following computational staining and destaining: Difference between Native non-stained and computationally stained (U_C); native non-stained and H&E dye stained (U_H); ground truth H&E dye stained and computationally stained (H*_C); H&E dye stained and computationally destained (H_D); H&E dye stained and native non-stained (H_U); computationally destained and ground truth native non-stained (D_U*). All values are in pixel intensities (0 to 255) calculated by subtracting the 2nd from 1st image. Positive values indicate decrease in average pixel intensities and negative values indicate gain. Values have been rounded to nearest integer. H is H&E dye stained image, C is computationally stained image, D is computationally destained image and U is native non-stained image. Ground truth images are indicated with '*' to facilitate comparisons with computational images when necessary.

Image	Computational staining			Computational destaining		
	U_C (±std)	U_H(±std)	H*_C(±std)	H_D(±std)	H_U(±std)	D_U*(±std)
1	-42	-43	1	45	43	-3
2	-30	-26	-4	37	26	-11
3	-39	-47	8	40	47	1
4	-48	-49	1	48	49	1
5	-48	-44	-3	44	44	0
6	-32	-34	2	44	34	-10
7	-19	-21	1	32	21	-12
8	-53	-53	0	58	53	-5
9	-45	-48	3	56	48	-8
10	-42	-40	-2	44	40	-4
11	-43	-42	-2	48	42	-6
12	-56	-66	10	62	66	4
13	-50	-58	8	56	58	2
MEAN	**-42±10**	**-44±12**	**2±4**	**47±8**	**44±12**	**-3±5**

eTable 3. Change in Mean Pixel Intensity in Red, Green, and Blue Channels per Image

ST1: Change in average pixel intensity (Red channels) per image: Native non-stained to computationally stained image (U_C), Native non-stained to Hematoxylin and Eosin (H&E) dye stained image (U_H), H&E dye stained to computationally restained (H_C). All values are in pixel intensities (0 to 255) calculated by subtracting the 2nd image from 1st image. Positive values indicate drop in average pixel intensity and negative values indicate gain in pixel intensity. H is H&E dye stained image; C is computationally stained images and U is native non-stained image. Values have been rounded to nearest integer.

Image	U_C(± std)	U_H (±std)	H_C(±std)
1	-66	-66	0
2	-38	-36	-3
3	-51	-59	7
4	-58	-57	0
5	-60	-53	-7
6	-60	-59	-1
7	-48	-50	2
8	-68	-67	-1
9	-62	-61	-1
10	-55	-49	-6
11	-57	-54	-3
12	-75	-81	6
13	-59	-62	3
MEAN	**-58±9**	**-58±10**	**0±4**

ST2: Change in average pixel intensity (Green channels) per image: Native non-stained to computationally stained image (U_C), Native non-stained to Hematoxylin and Eosin (H&E) dye stained image (U_H), H&E dye stained to computationally restained (H_C). All values are in pixel intensities (0 to 255) calculated by subtracting the 2nd image from 1st image. Positive values indicate drop in average pixel intensity and negative values indicate gain in pixel intensity. H is H&E dye stained image; C is computationally stained images and U is native non-stained image. Values have been rounded to the nearest integer.

Image	U_C (±std)	U_H (±std)	H_C (±std)
1	-2	-2	0
2	6	11	-6
3	-2	-12	10
4	-15	-16	1
5	-14	-11	-3
6	10	6	4
7	24	24	0
8	-23	-22	-1
9	-11	-15	4
10	-5	-4	-1
11	-6	-4	-3
12	-23	-35	12
13	-19	-29	11
MEAN	**-6±13**	**-8±16**	**2±5**

ST3: Change in average pixel intensity (Blue channels) per image: Native non-stained to computationally stained image (U_C), Native non-stained to Hematoxylin and Eosin (H&E) dye stained image (U_H), H&E dye stained to computationally restained (H_C). All values are in pixel intensities (0 to 255) calculated by subtracting the 2nd image from 1st image. Positive values indicate drop in average pixel intensity and negative values indicate gain in pixel intensity. H is H&E dye stained image; C is computationally stained images and U is native non-stained image. Values have been rounded to nearest integer.

Image	U_C (±std)	U_H (±std)	H_C (±std)
1	-58	-60	2
2	-58	-55	-4
3	-62	-70	7
4	-71	-74	3
5	-70	-69	-1
6	-46	-49	3
7	-34	-35	2
8	-70	-71	1
9	-64	-68	4
10	-66	-67	1
11	-67	-68	1
12	-72	-82	11
13	-73	-83	10
MEAN	**-62±11**	**-65±13**	**3±4**

eTable 4. Intrarater Agreement Calculated on Dye Stained and Computationally Stained Images

Intra-rater agreement calculated on Hematoxylin and Eosin (H&E) dye stained RGB Whole Slide Images (RWSI) and computationally stained RWSI using Intersection over Union (IoU). `-' indicates that the tumor (or tumor grade) label was not provided by the pathologists. Higher IoU score is better with a score of 1.0 representing perfect match of labels.

Image	H&E dye RWSI (±std)	Computationally stained RWSI (±std)
1	0.83	0.59
2	0.85	0.75
3	-	-
4	0.73	0.75
5	0.84	0.70
6	0.78	0.89
7	0.76	0.78
8	0.71	0.73
9	0.79	0.80
10	0.76	0.78
11	0.97	0.87
12	-	-
13	0.82	0.83
MEAN	**0.81±0.07**	**0.77±0.08**

Statistical analysis using t-test (p value~0.28>0.005) for the intra-IOU values shows no significant difference between the mean of intra-IOU of H&E stained (mean~0.81, std~0.072) and Computationally stained images (mean~0.77, std~0.08).

eAppendix 4. Evaluation of the Activation Maps of Trained Deep Neural Network

Clinical evaluations of computationally stained images: Figure 2 in the manuscript shows representative input non-stained image patches in row (a) that had Gleason grade 3 (columns I, II) or 4 (columns III, IV) tumors or were benign (column V), and their computational H&E staining (row c) and accuracy calculated using annotations by multiple physicians (row d). Tissue morphology in computationally stained patches (row c) matches closely with H&E dye stained patches (row b). Patch c-I successfully generated a benign area along with tumor signature (as indicated by arrows) and confirmed in row d-I. Computationally stained patches (row c) retain appearance of benign and malignant glands and stroma seen in H&E dye stained patches (row b). Patch b-III also contains edge/crush artifact (arrowheads) that is preserved in computationally stained image (row c-III). Same patches are shown (row d) with color-coded areas of agreement and disagreement between the labels provided on H&E dye stained images and computationally stained RWSI. It is evident that the computationally H&E stained patches represent tumor signatures with high accuracy and pathologists are able to correctly identify tumor. Majority of observed disagreements between raters did not represent misidentification of glands as benign or malignant. Instead, they show differences in rater annotation at borders of tumor labels, mainly due to differences in labeling style with some raters providing course labels and others annotating detailed labels (row d-III, arrows), or biopsy edges, as some raters chose to score partial/crushed glands at the periphery of samples and others did not (row d-III, arrowheads). Reconstructed computationally stained images shown in eFigure 1.1.b and 1.2.b (used for validation of the trained neural network) morphologically represented benign and malignant glands and stroma well enough to be consistently identified by pathologists (eFigure 1.1.c and 1.2.c) when compared with corresponding H&E dye stained images (eFigure 1.1.a and 1.2.a). A vast majority of tumor also showed annotator agreement. In some instances, “atypical” glands that were morphologically indeterminate for malignancy led to interpretative discrepancies however showed preserved morphology in the computationally stained images (e.g. arrows in eFigure 1.1.a and 1.1.b). Ground truth non-stained (eFigure 1.1.d and 1.2.d) and corresponding computationally destained images (eFigure 1.1.e and 1.2.e) are also shown for comparison. eFigure 1.4.b, 1.6.b and 1.7.b show the most reported areas of disagreement many of which are attributed to atypical glands that were hard to categorize on both images but were well represented on the computer-generated images (eFigure 1.4.c, 1.6.c and 1.7.c). eFigure 1.5 shows the uncommon Gleason pattern 5 tumors with comedo necrosis (eFigure 1.5.a, arrow). The morphology of the tumor glands is well maintained (eFigure 1.5.b, arrowheads), but the comedo necrosis is not visualized eFigure 1.5.b, arrow). The dye-stained image in eFigure 1.7.a contains an infrequently encountered scenario (indicated by an arrow), the presence of rare malignant glands that are not well visualized on the computationally stained image (eFigure 1.7.b, arrow). Despite this altered appearance, there was no impact on clinical diagnosis as the blinded reviewers scored these areas as tumor. Some glands are poorly formed on both the dye stained and the computationally stained image (eFigure 1.8.a, 1.8.b, arrows), leading to disagreement between raters, even though the computationally stained image were identical to dye stained image. Images shown in eFigure 1.9 presented a challenging labeling exercise where tumor cell cytoplasm was very pale and did not show significant contrast to the background stroma in the dye-stained image (eFigure 1.9.a). This cytoplasmic pallor was also well preserved in the computationally stained image (eFigure 1.9.b). Despite this, appearance of the nuclei and the slight difference in cytoplasmic texture made the tumor identifiable in both images (eFigure 1.9.c and 1.9.d). The computationally stained images shown in eFigure 1.10.b, 1.11.b, 1.13.b were well represented. Majority of the disagreement in these images arose due to tumor/non-tumor boundary and biopsy edge issues. Validation images in eFigure 1.11.b and 1.13.b illustrated additional high-quality examples of preserved morphology generated by the computationally staining algorithm, which confirmed accurate matching with dye stained images in benign conditions. Non-necrotizing granulomas, marked chronic inflammation, reactive stromal changes and proteinaceous debris were all morphologically identifiable in the computational stained images (eFigure 1.11.c). Pathologists unanimously scored the matched H&E dye stained and computationally stained images shown in eFigure 1.3 and 1.12 as benign.

eTable 5. Comparison of Tumor Grades Between Original Expert Microscopic Diagnosis

Comparison of tumor grades between original expert microscopic diagnosis [as reported in the Electronic Health Records (EHR)], using the Hematoxylin and Eosin (H&E) stained glass slide and the diagnosis of the computationally stained image. *Agreements confirmed upon re-review of the microscopic slide and additional supportive studies; † Not clinically significant within the context of the patients known tumor.

Image	Initial diagnosis after biopsy	Diagnosis using computationally stained image
1	40% grade 3 tumor in core	40% grade 3 tumor in core
2	50% grade 3 tumor in core	90% grade 3 tumor in core*,†
3	Benign core	Benign core
4	50% grade 3 tumor in core	50% grade 3 tumor (majority) with
		traces of grade 4 tumor†
5	50% grade 4 and 5 tumor on core	50% grade 4 and 5 tumor on core
	(G4 > G5)	(G4 < G5)†
6	40% grade 3 and 4 tumor in core	40% grade 3 and 4 tumor in core
	(G3 >> G4)	(G3 >> G4)†
7	40% grade 3 tumor in core	40% grade 3 and 4 tumor in core
		(G4 >> G3)†
8	40% grade 3 and 4 tumor in core	40% grade 3 and 4 tumor in core
9	20% grade 3 tumor in core	20% grade 3 tumor in core
10	90% grade 3 and 4 tumor in core	90% grade 3 and 4 tumor in core
11	Healthy core	Tiny focus of grade 3 tumor in core*,†
12	Healthy core	Healthy core
13	90% grade 4 tumor in core	90% grade 3 and 4 tumor in core
		(G3 << G4)†

eAppendix 5. Comparison With Patient Records

After expert re-review of the original slides and additional evaluation by immunohistochemistry, the original EHR diagnosis was overturned in two cases, resulting in two additional cases of agreement. Pathologists reviewing computer-generated core 11 were able to better identify the presence of rare glands of Gleason grade 3 tumors than those who had rendered the original EHR diagnosis of benign (eFigure 1.11 marked blue/green in the supplement). Microscopic re-review of the original glass slide confirmed that it indeed had a tiny focus of grade 3 tumor that was overlooked at the time of the original diagnosis. Subsequent immunohistochemical analysis revealed the absence of basal cells around the glands in question, confirming the diagnosis of carcinoma made during this study and revealing the diagnosis conferred on the computationally generated images to be correct. eFigure 1.2 was the only study biopsy that showed a significant difference in tumor fraction, as this study reported 50% tumor fraction and the original EHR report was 90%. Re-review of the original glass slide again showed this study fraction to be more accurate than the original diagnosis (eFigure 1.2). Otherwise, the tumor fraction identified in all the computationally generated images approximated the fraction reported in the EHR for all images as evident from eTable 5. None of the differences between EHR and computationally generated H&E diagnosis were clinically significant with regard to treatment decisions. A difference in grade of tumor was identified in a minor component of computationally stained images (eFigure 1.4, 1.7 and 1.13). The small foci of higher or lower grade tumor identified in computationally stained images (eFigure 1.4.c, 1.7.c, and 1.13.c), which were not reported at the time of original diagnosis, comprised a very small fraction of tumor volume. These were often associated with diagnostically indeterminate questions (e.g. whether a gland represented a rare focus of grade 4 tumor or if it was tangential sectioning of grade 3 tumor), and were not clinically significant in the context of the patient's known tumor at the time of original EHR reported diagnosis.

eReferences

1. Torr PH, Zisserman A. MLESAC: A new robust estimator with application to estimating image geometry. Computer vision and image understanding. Apr 1, 2000;78(1):138-56.
2. Lowe DG. Distinctive image features from scale-invariant keypoints. International journal of computer vision. Nov 1,2004;60(2):91-110.
3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 33. 1977; 159–74.