Science Advances
AAAS

Supplementary Materials for

Generative adversarial networks accurately reconstruct pan-cancer histology from pathologic, genomic, and radiographic latent features

Frederick M. Howard *et al.*

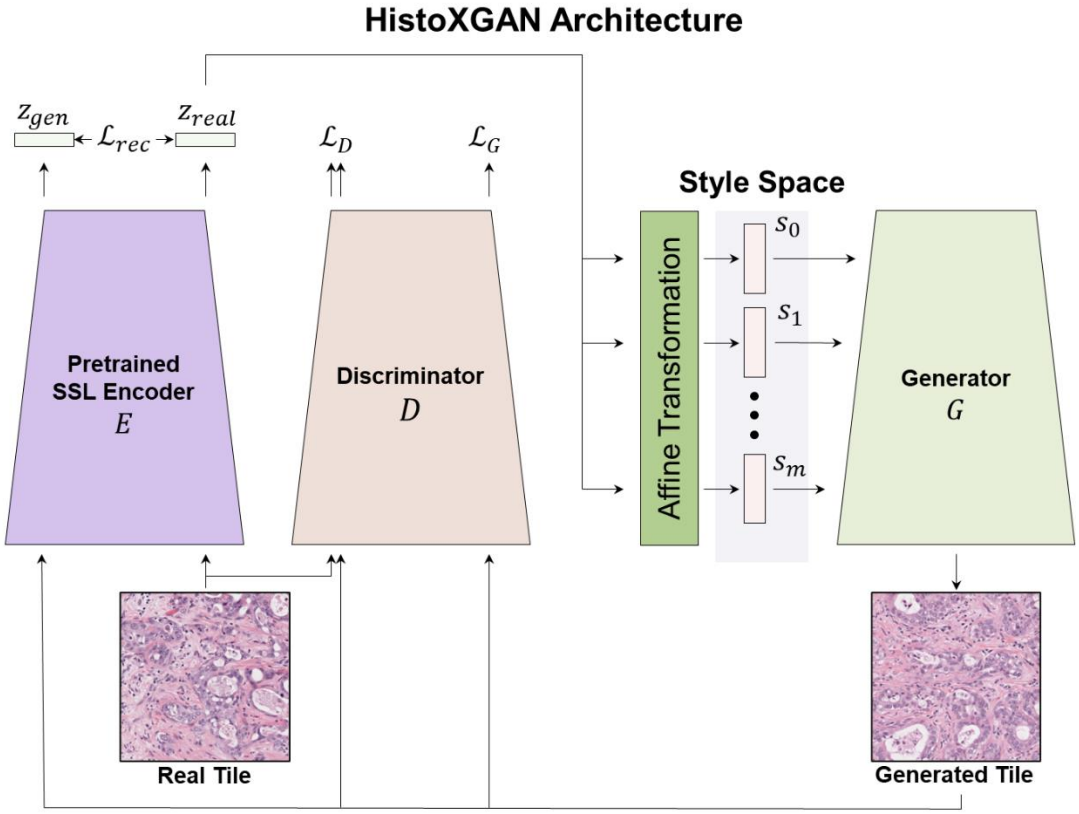
Corresponding author: Frederick M. Howard, frederick.howard@uchospitals.edu;
Alexander T. Pearson, alexander.pearson@uchicagomedicine.org

Sci. Adv. **10**, eadq0856 (2024)
DOI: 10.1126/sciadv.adq0856

The PDF file includes:

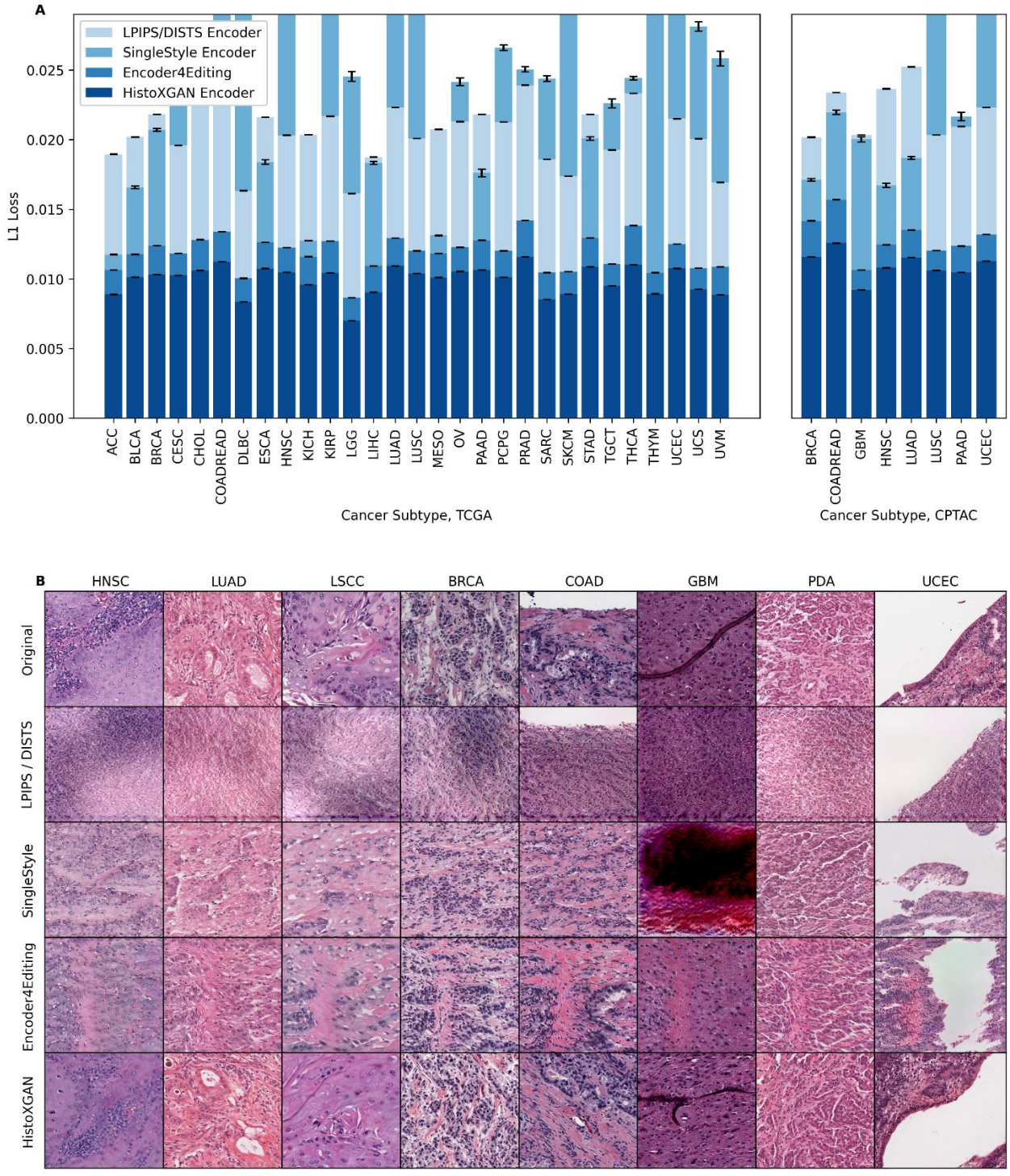
Other Supplementary Material for this manuscript includes the following:

Fig. S1.



Overview of Generator Architecture and Study Analysis. The HistoXGAN architecture is based on StyleGAN2, with simultaneous training of a generator *G* and discriminator *D*. During training, latent image features (z) are extracted using a pretrained self-supervised feature encoder *E*. These latent features are transformed into style vectors (s_0, …, s_n) which are used to recreate a pathology image tile. An additional L1 loss $\mathcal{L}_{rec}$ is added for the comparison of the encoded features from the reconstructed pathology tile z_{gen} and the input tile z_{real}.

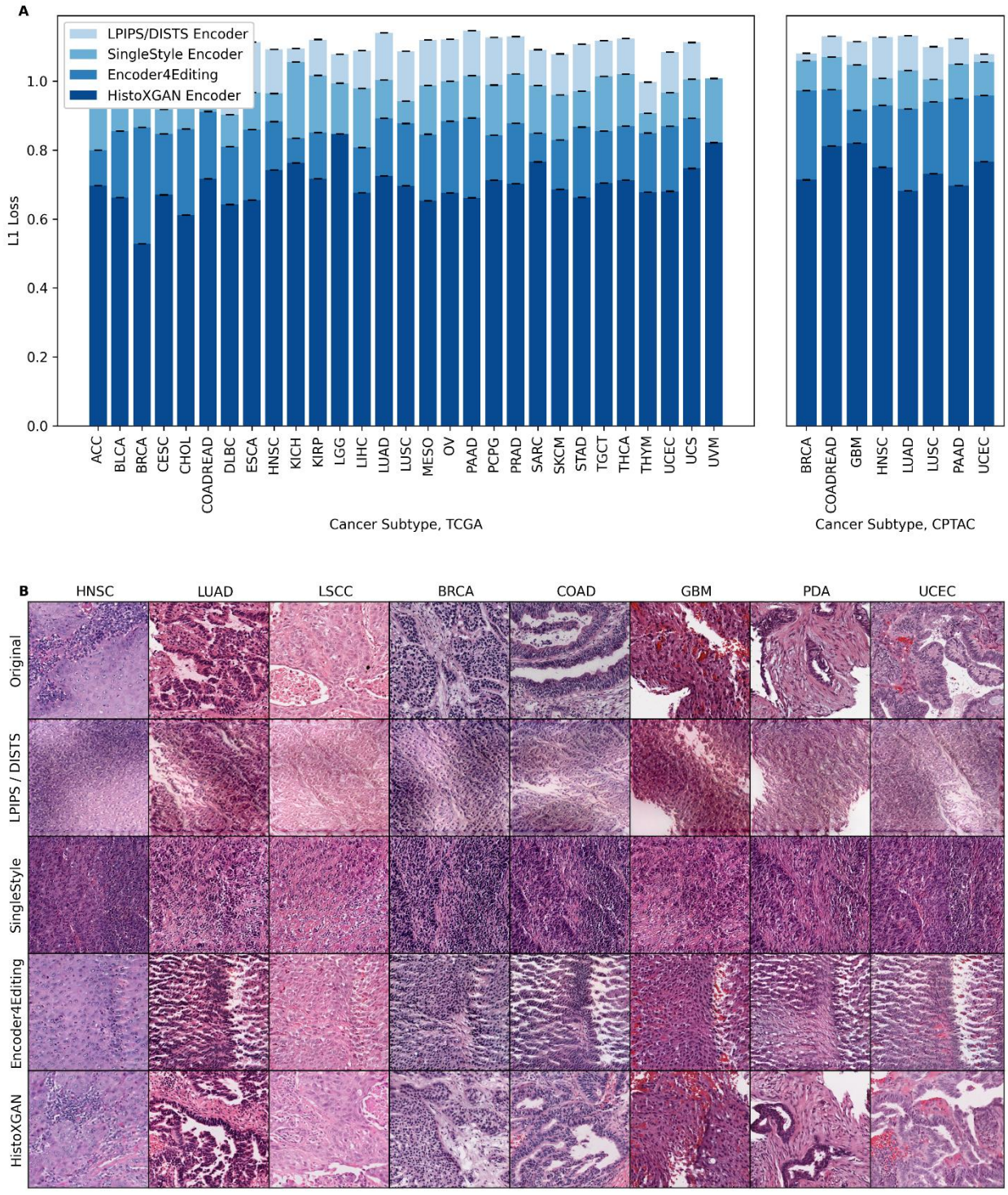
Fig. S2.



Reconstruction Accuracy in Training and Validation Datasets for RetCCL Encoders. We compare reconstruction accuracy from the real and reconstructed images for HistoXGAN and other architectures for embedding images in GAN latent space. For comparison, we use encoders designed to recreate images from a StyleGAN2 model trained identically to the HistoXGAN

model. The Learned Perceptual Image Patch Similarity (LPIPS) / Deep Image Structure and Texture Similarity (DISTS) encoder uses an equal ratio of LPIPS / DISTS loss between the real and reconstructed images to train the encoder. The Single Layer and Encoder4Editing encoders are trained to minimize L1 loss between RetCCL feature vector of the real and reconstructed images. (**A**) HistoXGAN provides more accurate reconstruction of RetCCL features across the TCGA dataset used for GAN training (n = 8,120) and CPTAC validation (n = 1,328) dataset, achieving an average of 30% improvement in L1 loss over the Encoder4Editing encodings in the validation dataset. (**B**) HistoXGAN reconstructed images consistently provided more accurate representations of features from the input image across cancer types in the CPTAC validation dataset.

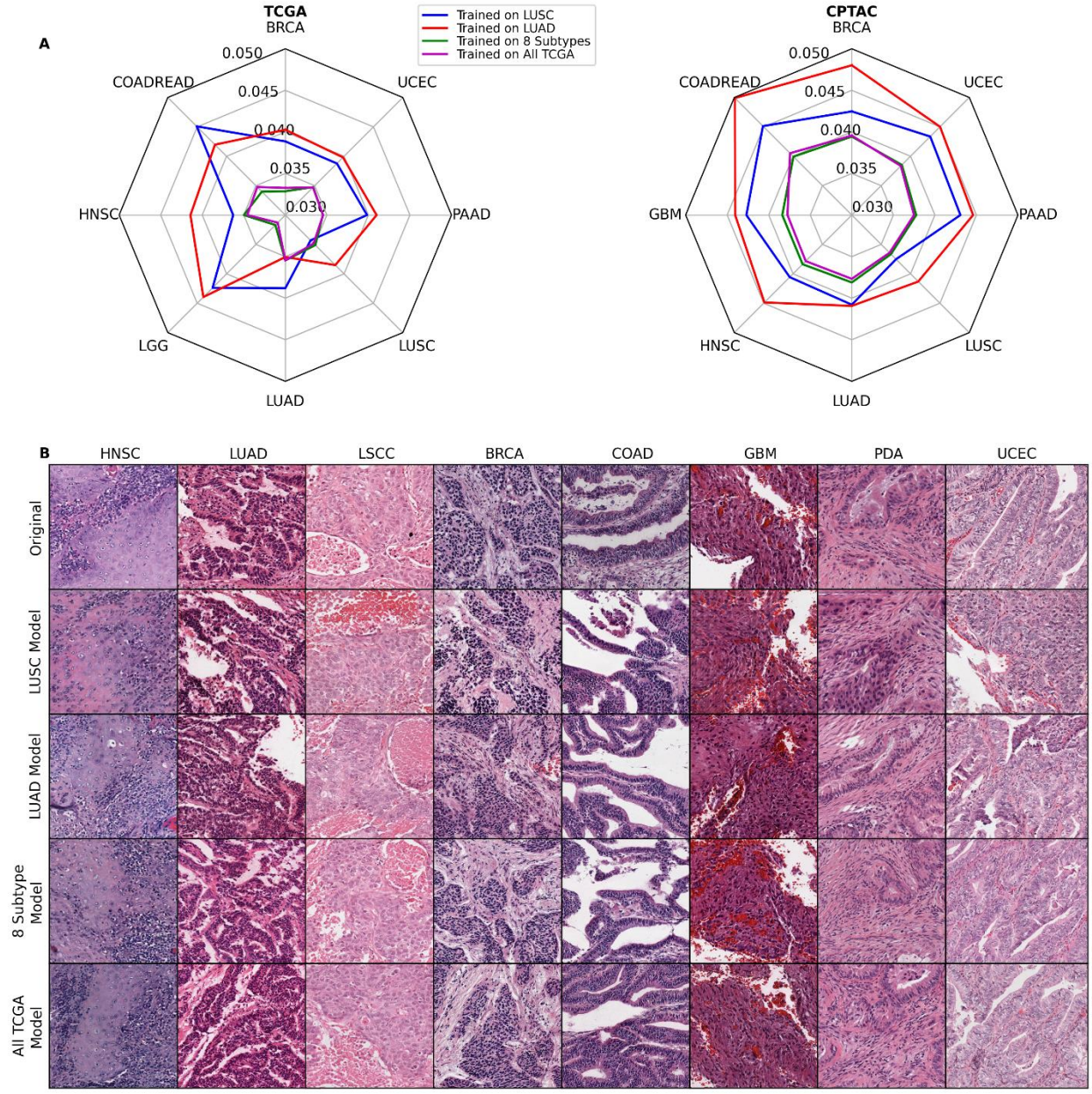
Fig. S3.



Reconstruction Accuracy in Training and Validation Datasets for UNI Encoders. We compare reconstruction accuracy from the real and reconstructed images for HistoXGAN and other architectures for embedding images in GAN latent space. Feature extraction was done with the UNI encoder – notably, this encoder was trained on a dataset completely external to TCGA / CPTAC, highlighting the generalizability of HistoXGAN. For comparison, we use encoders

designed to recreate images from a StyleGAN2 model trained identically to the HistoXGAN model. The Learned Perceptual Image Patch Similarity (LPIPS) / Deep Image Structure and Texture Similarity (DISTS) encoder uses an equal ratio of LPIPS / DISTS loss between the real and reconstructed images to train the encoder. The Single Layer and Encoder4Editing encoders are trained to minimize L1 loss between UNI feature vectors of the real and reconstructed images. (**A**) HistoXGAN provides more accurate reconstruction of UNI features across the TCGA dataset used for GAN training (n = 8,120) and CPTAC validation (n = 1,328) dataset, achieving an average of 28% improvement in L1 loss over the Encoder4Editing encodings in the validation dataset. (**B**) HistoXGAN reconstructed images consistently provided more accurate representations of features from the input image across cancer types in the CPTAC validation dataset.

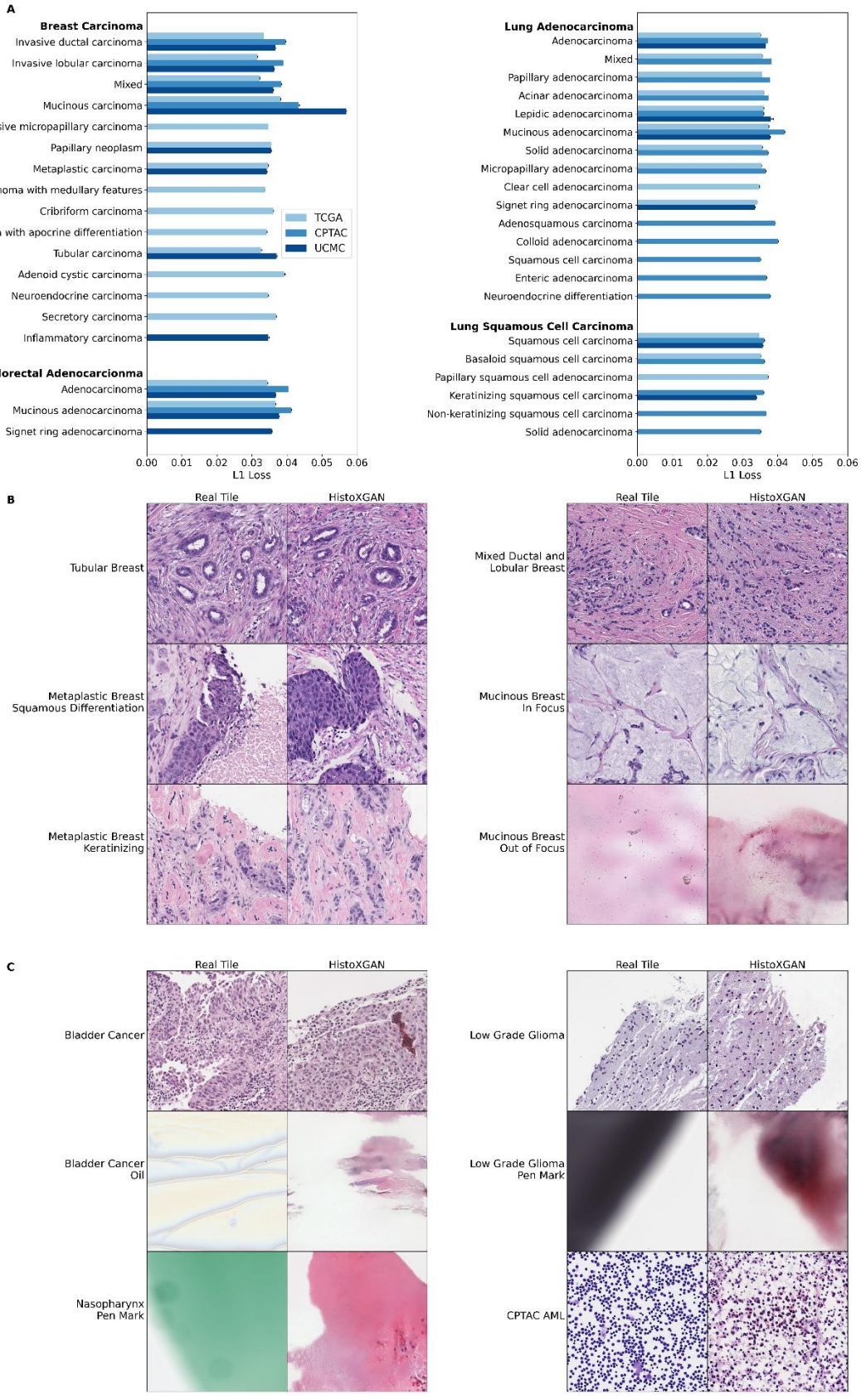
Fig. S4.



Dependency of Reconstruction Accuracy on Choice of Training Dataset. We compared accuracy of reconstruction of CTransPath features across four HistoXGAN models, with identical training parameters aside from choice of training dataset. One model was trained exclusively on TCGA-LUSC, one was trained exclusively on TCGA-LUAD, one was trained on all eight cancer subtypes listed above, and one was trained across twenty-nine tumor types in TCGA as previously described. (**A**) Performance across eight cancer types in TCGA / CPTAC are shown. Models trained on single cancer subtypes had worse performance than models trained on multiple cancer types, although minimal difference in performance was seen between the eight and twenty-nine subtype models. The TCGA-LUSC model outperformed TCGA-LUAD model in all squamous tumor types (including the TCGA-HNSC, CPTAC-HNSC, and CPTAC-LSCC datasets); the TCGA-LUAD model outperformed in some adenocarcinoma subsets

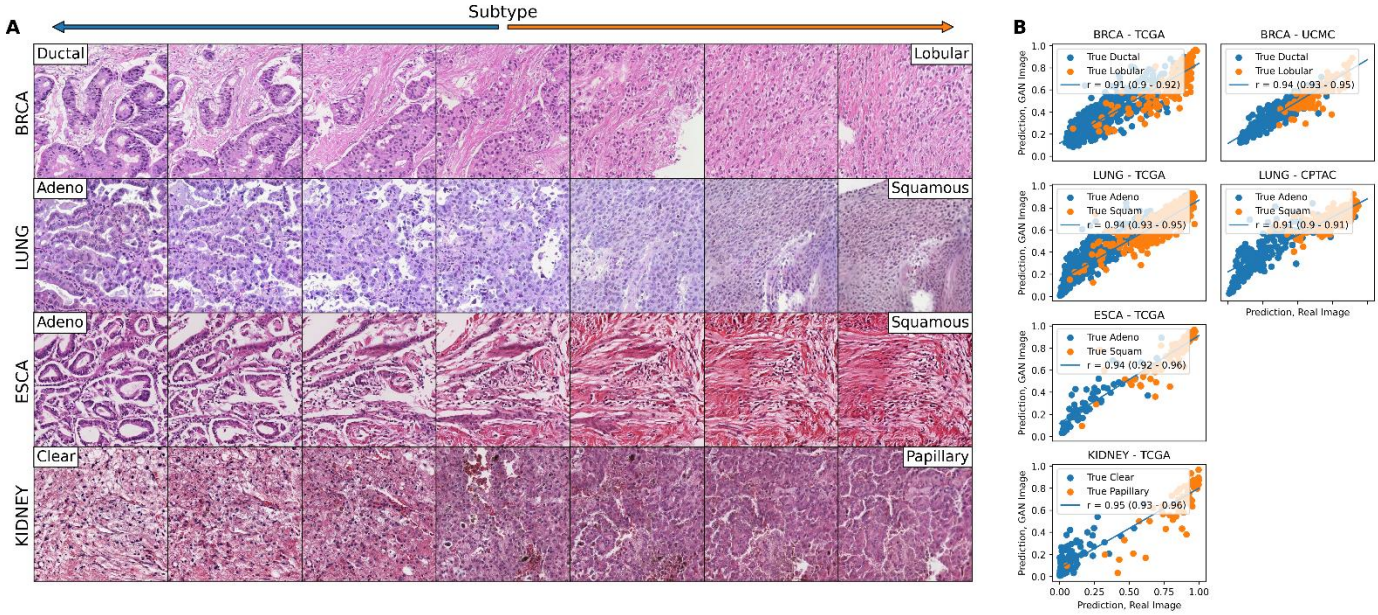
(including TCGA-COADREAD) – but overall had worse reconstruction accuracy in external datasets. (**B**) Example reconstructions from CPTAC image tiles are shown for the four models. Some of the TCGA-LUSC reconstructions have less evident gland formation than the source images, such as the reconstructions from CPTAC-PDA and CPTAC-UCEC, which may be due to lack of gland formation in image tiles available during training in the squamous dataset.

Fig. S5.


A
Breast Carcinoma
Invasive ductal carcinoma
Invasive lobular carcinoma
Mixed
Mucinous carcinoma
Invasive micropapillary carcinoma
Papillary neoplasm
Metaplastic carcinoma
Invasive carcinoma with medullary features
Cribriform carcinoma
Carcinoma with apocrine differentiation
Tubular carcinoma
Adenoid cystic carcinoma
Neuroendocrine carcinoma
Secretory carcinoma
Inflammatory carcinoma
Colorectal Adenocarcinonma
Adenocarcinoma
Mucinous adenocarcinoma
Signet ring adenocarcinoma
TCGA
CPTAC
UCMC
0.00 0.01 0.02 0.03 0.04 0.05 0.06
L1 Loss
Lung Adenocarcinoma
Adenocarcinoma
Mixed
Papillary adenocarcinoma
Acinar adenocarcinoma
Lepidic adenocarcinoma
Mucinous adenocarcinoma
Solid adenocarcinoma
Micropapillary adenocarcinoma
Clear cell adenocarcinoma
Signet ring adenocarcinoma
Adenosquamous carcinoma
Colloid adenocarcinoma
Squamous cell carcinoma
Enteric adenocarcinoma
Neuroendocrine differentiation
Lung Squamous Cell Carcinoma
Squamous cell carcinoma
Basaloid squamous cell carcinoma
Papillary squamous cell adenocarcinoma
Keratinizing squamous cell carcinoma
Non-keratinizing squamous cell carcinoma
Solid adenocarcinoma
0.00 0.01 0.02 0.03 0.04 0.05 0.06
L1 Loss
B
Real Tile
HistoXGAN
Tubular Breast
Metaplastic Breast Squamous Differentiation
Metaplastic Breast Keratinizing
Mixed Ductal and Lobular Breast
Mucinous Breast In Focus
Mucinous Breast Out of Focus
C
Bladder Cancer
Bladder Cancer Oil
Nasopharynx Pen Mark
Low Grade Glioma
Low Grade Glioma Pen Mark
CPTAC AML

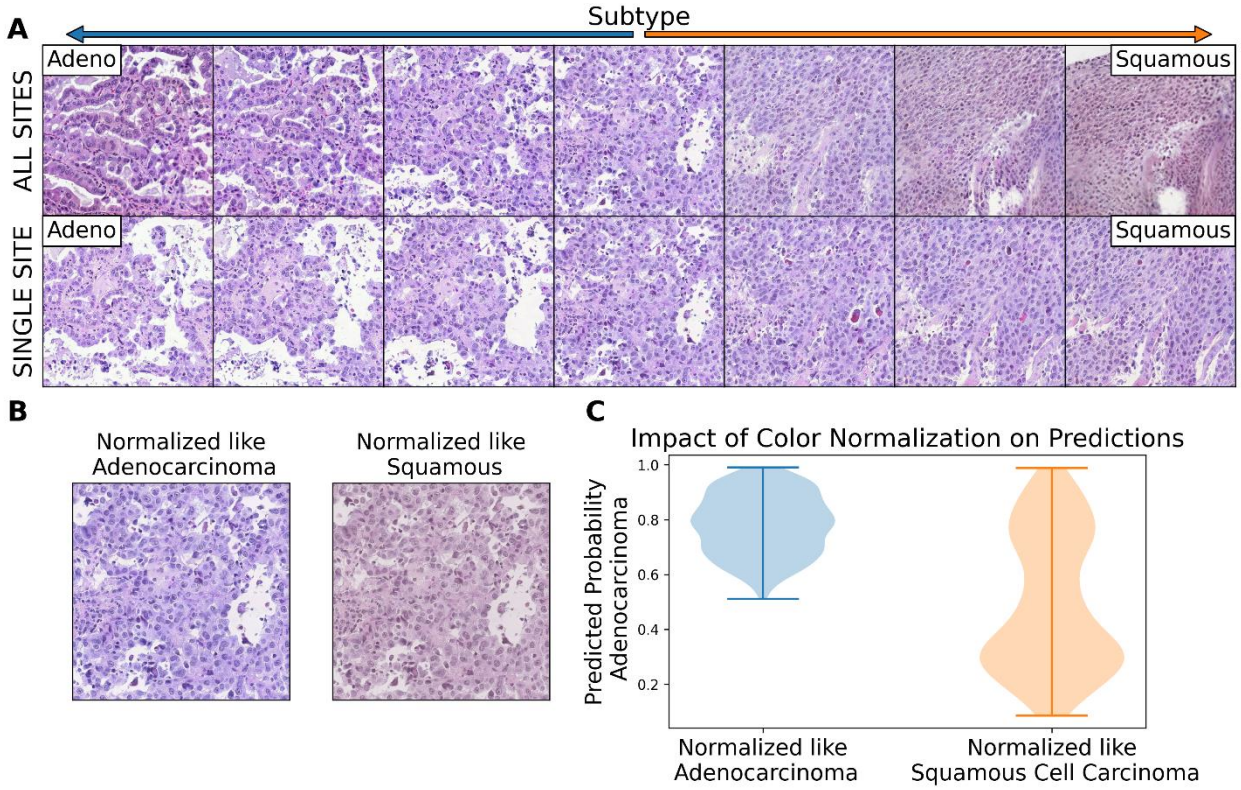
Evaluation of Reconstruction Accuracy Across Histologic Subtypes and Rare Cancer Types. We compared reconstruction accuracy from the real and reconstructed images for HistoXGAN across histologic subtypes and rare cancer types, with the results illustrated here with the CTransPath encoder. (**A**) Reconstruction accuracy is listed for the breast cancer, lung squamous cell carcinoma, lung adenocarcinoma, and colorectal adenocarcinoma datasets from the training TCGA cohort and the validation CPTAC and University of Chicago Medical Center (UCMC) datasets. As illustrated, performance was robust across histologic subtypes that were both common in training and rare or absent in training – although the worst numeric performance was seen for reconstruction of mucinous tumors in breast, lung, and colon adenocarcinomas. (**B**) As shown, HistoXGAN can reconstruct tumors with features illustrative of uncommon breast cancer subtypes, with source image tiles taken from the UCMC validation cohort. Of note, the worst validation performance was seen in the UCMC mucinous breast cancers – two slides of mucinous breast tumor were available in this cohort, one of which was out of focus, as illustrated – explaining this poor performance. (**C**) Among 176 unique OncoTree diagnoses, HistoXGAN demonstrated the worst reconstruction accuracy for one slide each of nasopharyngeal cancer (mean loss 0.087), bladder urothelial carcinoma (mean loss 0.060), and low-grade glioma not otherwise specified (mean loss 0.058). Visualization of these three cases demonstrated that the low reconstruction accuracy was largely due to artifacts such as pen markings and oil in the slide images. Conversely, in the CPTAC AML cohort (mean loss 0.066), the high reconstruction loss was due to errors in reconstruction, with excessive stellate cytoplasm generated along with stroma more reminiscent of a solid tumor, demonstrating that this approach is not extensible to reconstruction of blood smear images.

Fig. S6.



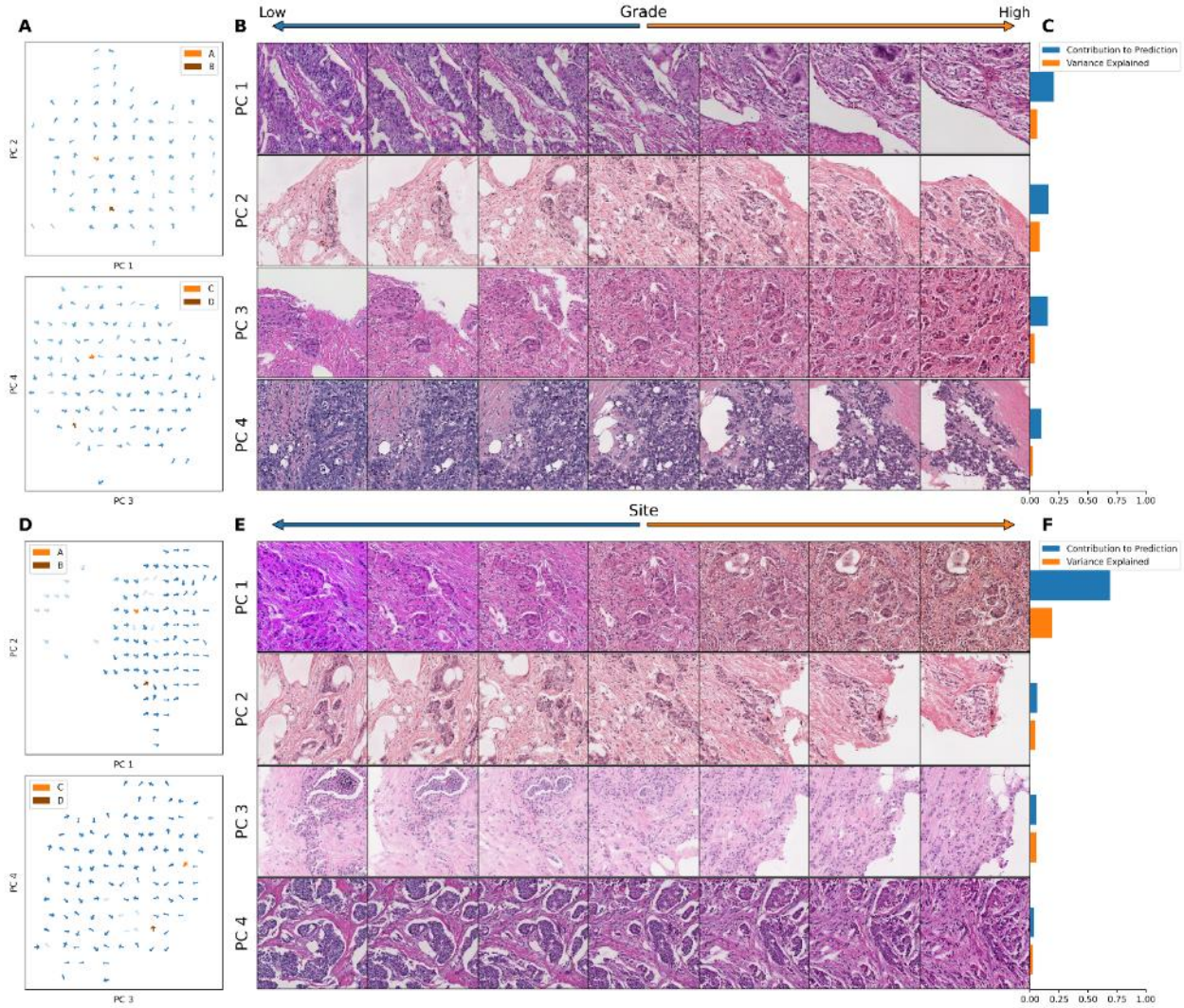
Perceptual Consistency of Histologic Subtype in Reconstructed Images Across Cancer Types. (**A**) Illustration of transition between tumor histologic subtypes across a single image from four cancer types. A vector differentiating the two most frequent subtype classes is derived from the coefficients of a logistic regression predicting subtype from CTransPath features. This vector is subtracted from the base image to visualize the first class and added to the base image to visualize the second class. (**B**) Correlation between predictions of histologic subtype from real and reconstructed tiles, averaged per patient, across cancer types, demonstrating a high perceptual similarity of the histologic subtype of the real and generated images. For the TCGA datasets, a deep learning model was trained to predict subtype from real tiles for each cancer type using three-fold cross validation. The correlation between predictions for real / generated images is aggregated for the three held out validation sets. For the CPTAC and University of Chicago Medical Center (UCMC) validation, a deep learning model trained across the entire corresponding TCGA dataset was used to generate predictions. True pathologist determined subtype is indicated by the color of each datapoint.

Fig. S7.



Quantifying Contributions of Batch Effect to Model Predictions. Transitions between readily identifiable histologic features such as tumor subtype and grade with HistoXGAN occasionally demonstrate changes in staining or other visual characteristics of the image patch. We thus evaluate if these changes represent confounding related to HistoXGAN or are due to batch effects present in the dataset used to fit the transition. (**A**) Transitions between lung adenocarcinoma and squamous cell carcinoma demonstrate a visual color shift when fit on cases from all sites within TCGA. There is minimal color shift when the transition is fit using cases from a single site (Asterand Bioscience). (**B**) To further test this hypothesis, we evaluate the impact of stain color normalization using the standard Reinhard algorithm to normalize image tiles to appear similar to HistoXGAN generated tiles after transition to adenocarcinoma (i.e. the top left image in **A**) and transition to squamous cell carcinoma (i.e. the top right image in **A**). A sample tile post-normalization is shown. (**C**) Predicted probability of adenocarcinoma was calculated after color normalizing tiles ‘like adenocarcinoma’ and ‘like squamous cell carcinoma’ using a model trained on TCGA lung squamous / adenocarcinoma aggregated from all sites, demonstrating that model predictions can vary significantly with changes in stain color.

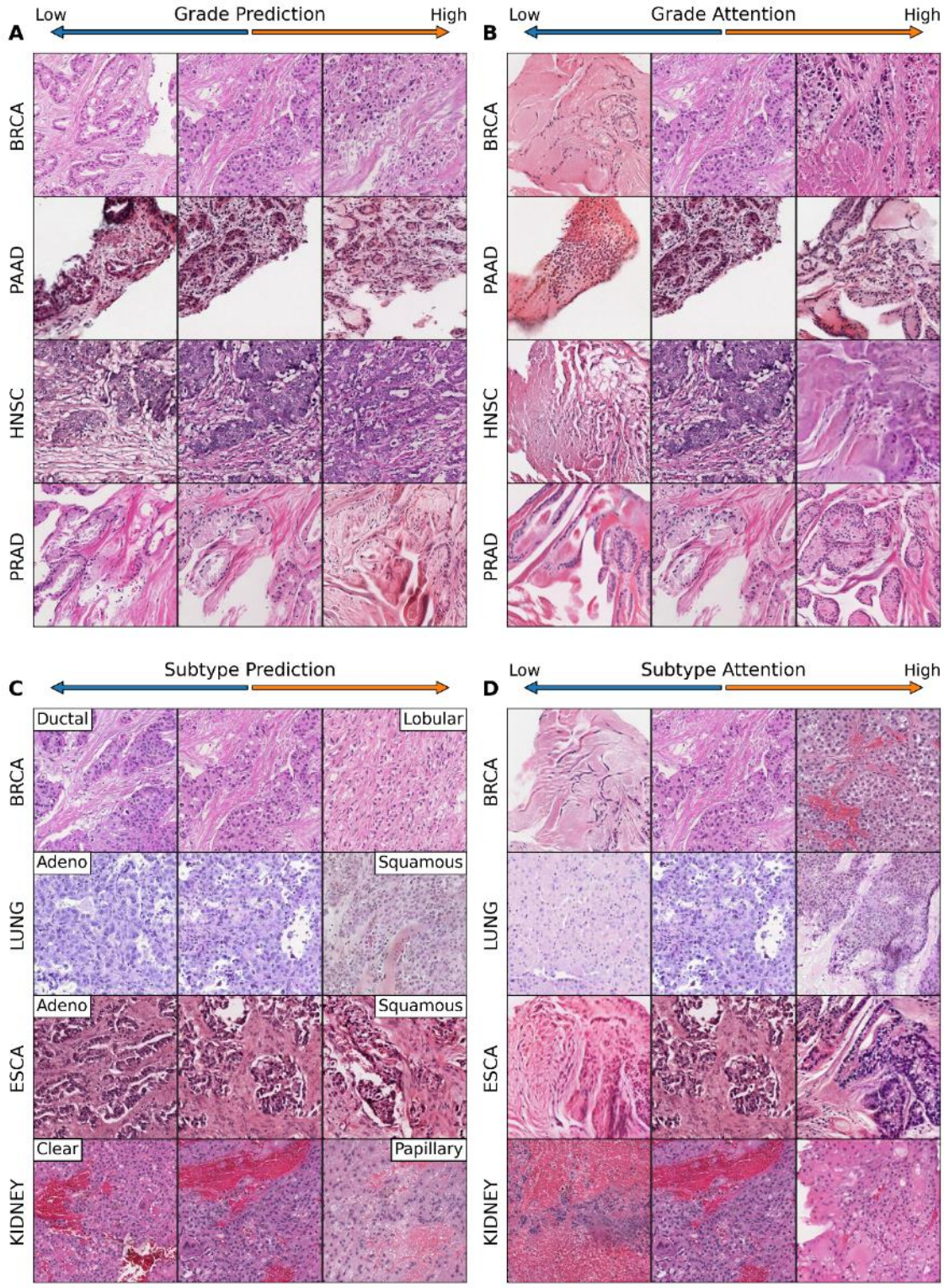
Fig. S8.



Identifying Orthogonal Features that Contribute to Model Predictions. Tile based weakly supervised models were trained to predict grade and contributing site for TCGA-BRCA (n = 943). The gradient with respect to a prediction of these outcomes (grade low / high; site A / B) was calculated for the average feature vector across each slide in the dataset. Principal component analysis was applied to these gradients, and components were sorted by the magnitude of difference of the component between gradients toward each outcome class, i.e. the strength of the contribution of the component to the prediction. (**A, D**) Gradients to predict high grade / site B were projected into spaces of pairs of principal components. Magnitude of the gradient is indicated by arrow thickness. Whereas the component that led to a stronger prediction of high / low grade was highly variable, for nearly all images application of the first principal component led to a stronger prediction for site B (i.e. all arrows pointing to the right in the first plot in figure D). The tile with the maximum gradient for each component is selected for plotting (labeled A - D for principal components 1 – 4). (**B, E**) The top four principles components for grade and site were used to perturb images. The first principal component in site clearly caused a shift in the stain color of the image. (**C, F**) The relative contribution of the component to the

gradient and the proportion of variance explained by the component are listed. Approximately 70% of the gradient towards a higher prediction could be explained by the first principal component for site, illustrating the site model is relatively simple and based nearly entirely on stain color difference.

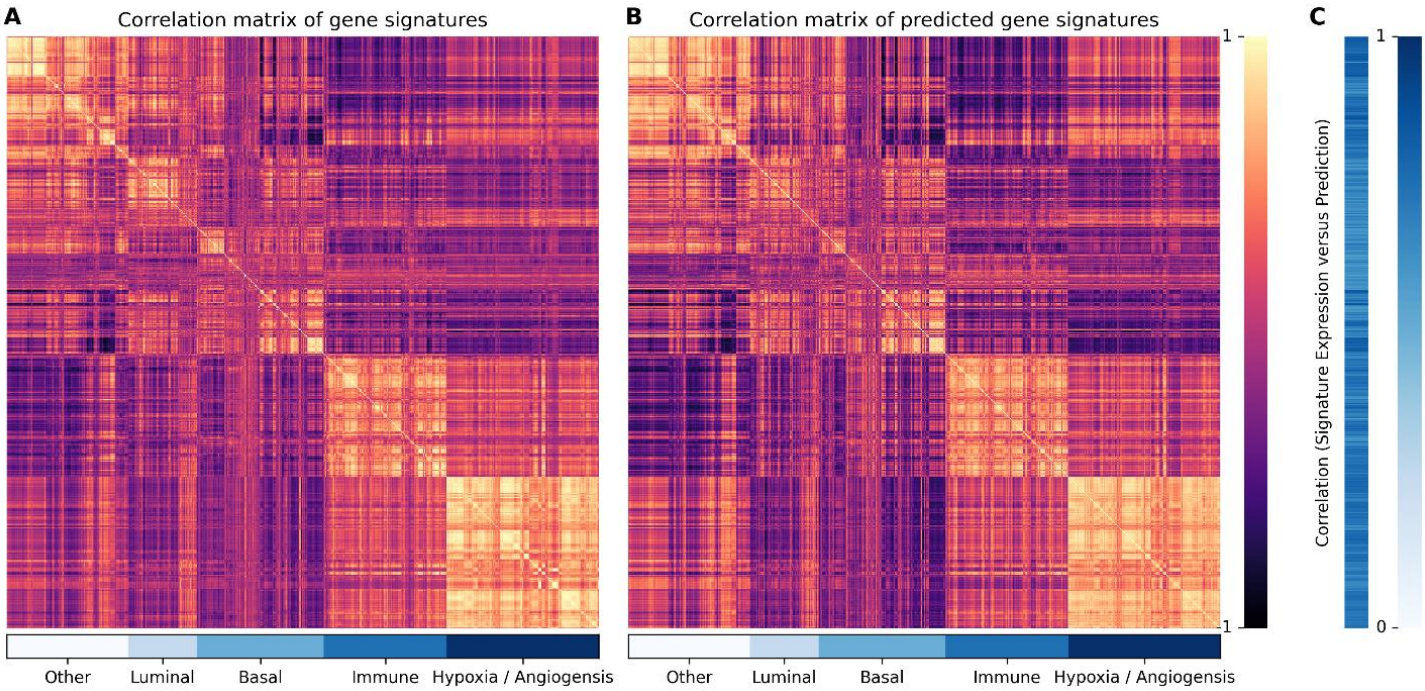
Fig. S9.



Interpretability of Attention-based / Multiple Instance Learning Models. We demonstrate that this HistoXGAN approach can be used to understand attention-based multiple instance learning (MIL) approaches or other multi-step models. Attention-MIL models were trained to

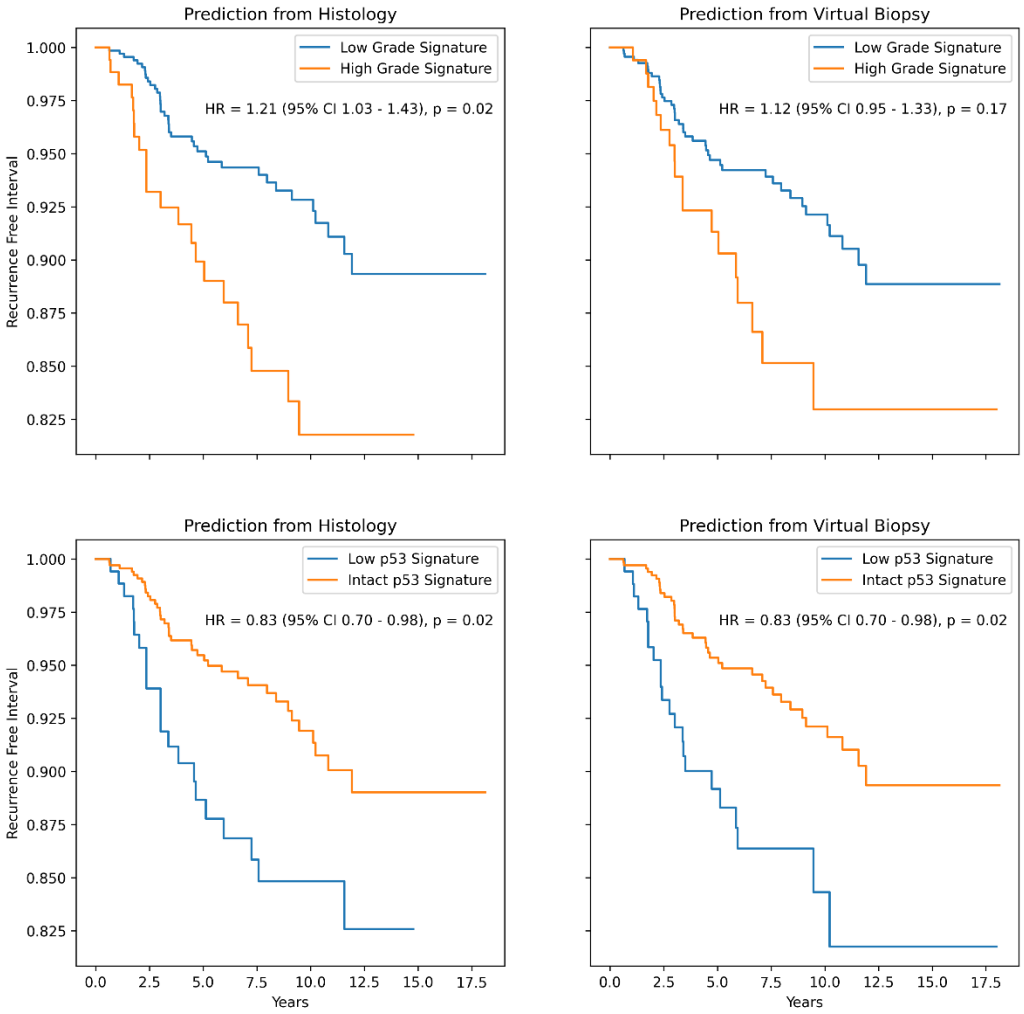
predict grade and histologic subtype across multiple cancers in TCGA. Image tiles were generated with HistoXGAN, and gradient descent was used to perturb the image features to increase the predicted likelihood of grade / subtype class or increase attention towards the image tile. Grade and subtype prediction demonstrated expected histologic findings, and tiles demonstrated a more fibrous and acellular appearance as attention decreased.

Fig. S10.



Validating Predicted Gene Signatures as a Method of Assessing the Biologic Similarity of Histologic Images. Deep learning models were trained with three-fold cross validation to predict 775 clinically relevant gene expression signatures from digital histology in the TCGA-BRCA cohort, and predictions in the held-out test set were compared to the true values of gene expression from RNA sequencing. (**A**) Correlation matrix (illustrating Pearson correlation coefficient) of gene signatures in the TCGA cohort, demonstrating that the 775 signatures yield several orthogonal groups of highly correlated signatures, illustrating some degree of redundancy – especially in the immune / hypoxia signatures. (**B**) Correlation matrix of the gene signatures predicted from digital histology in the pooled held-out cohorts, demonstrating a highly similar pattern of signature grouping to the true gene expression signatures. (**C**) Pearson correlation between the gene signatures calculated from RNA sequencing, and the predicted signatures in the pooled held-out cohorts. Clinically relevant gene expression signatures could be accurately predicted from histology, with an average correlation coefficient between real and predicted signature of 0.45.

Fig. S11.



MRI Virtual Biopsy Facilitates Application of Prognostic AI Histology Models. Deep learning models were trained to predict clinically relevant gene expression signatures, including a histologic grade signature (PMID 27861902) and a p53 expression module (PMID 20335537). High histologic grade signature and intact p53 expression – both defined by a top 20% quantile and predicted from whole slide tumor images from the aforementioned models – were associated with significant improvements in freedom from recurrence. When these same deep learning models were applied to reconstructed histology from a virtual biopsy, similar prognostic value was seen, although there was only a trend towards significance in those with predicted high tumor grade signature.

Table S1. Reconstruction Accuracy in Training and Validation Datasets for RetCCL Encoders.

We compare reconstruction accuracy from the real and reconstructed images for HistoXGAN and other architectures for embedding images in GAN latent space. For comparison, we use encoders designed to recreate images from a StyleGAN2 model trained identically to the HistoXGAN model. The Learned Perceptual Image Patch Similarity (LPIPS) / Deep Image Structure and Texture Similarity (DISTS) encoder uses an equal ratio of LPIPS / DISTS loss between the real and reconstructed images to train the encoder. The Single Layer and Encoder4Editing encoders are trained to minimize L1 loss between RetCCL feature vector of the real and reconstructed images.

Source		n	n Tiles	LPIP / DISTS L1 Loss	Single Layer L1 Loss	Encoder4Editing L1 Loss	HistoXGAN L1 Loss
TCGA	**ACC**	56	78789	0.019 (0.004)	0.012 (0.005)	0.011 (0.002)	0.009 (0.002)
TCGA	**BLCA**	378	342463	0.020 (0.004)	0.017 (0.051)	0.012 (0.002)	0.010 (0.002)
TCGA	**BRCA**	943	454985	0.022 (0.004)	0.021 (0.068)	0.012 (0.002)	0.010 (0.002)
TCGA	**CESC**	267	165088	0.020 (0.004)	0.026 (0.076)	0.012 (0.002)	0.010 (0.002)
TCGA	**CHOL**	38	51414	0.023 (0.004)	0.028 (0.092)	0.013 (0.003)	0.011 (0.002)
TCGA	**COADREAD**	428	195493	0.023 (0.004)	0.085 (0.283)	0.013 (0.002)	0.011 (0.002)
TCGA	**DLBC**	43	32073	0.016 (0.004)	0.089 (0.306)	0.010 (0.003)	0.008 (0.002)
TCGA	**ESCA**	147	99625	0.022 (0.004)	0.018 (0.048)	0.013 (0.002)	0.011 (0.002)
TCGA	**HNSC**	401	166061	0.020 (0.004)	0.131 (0.408)	0.012 (0.002)	0.010 (0.002)
TCGA	**KICH**	101	105461	0.020 (0.003)	0.013 (0.003)	0.012 (0.002)	0.010 (0.002)
TCGA	**KIRP**	270	234740	0.022 (0.004)	0.044 (0.169)	0.013 (0.003)	0.010 (0.002)
TCGA	**LGG**	464	155579	0.016 (0.004)	0.025 (0.136)	0.009 (0.002)	0.007 (0.002)
TCGA	**LIHC**	359	331769	0.019 (0.004)	0.018 (0.062)	0.011 (0.003)	0.009 (0.002)
TCGA	**LUAD**	467	335499	0.022 (0.004)	0.037 (0.179)	0.013 (0.002)	0.011 (0.002)
TCGA	**LUSC**	474	370542	0.020 (0.004)	0.031 (0.114)	0.012 (0.002)	0.010 (0.002)
TCGA	**MESO**	73	42242	0.021 (0.004)	0.013 (0.006)	0.012 (0.002)	0.010 (0.002)
TCGA	**OV**	104	108620	0.021 (0.004)	0.024 (0.092)	0.012 (0.002)	0.011 (0.002)
TCGA	**PAAD**	168	119144	0.022 (0.004)	0.018 (0.090)	0.013 (0.002)	0.011 (0.002)
TCGA	**PCPG**	173	207803	0.021 (0.004)	0.027 (0.089)	0.012 (0.002)	0.010 (0.002)
TCGA	**PRAD**	394	202439	0.024 (0.004)	0.025 (0.083)	0.014 (0.003)	0.012 (0.002)
TCGA	**SARC**	250	326262	0.019 (0.003)	0.024 (0.113)	0.010 (0.002)	0.009 (0.002)
TCGA	**SKCM**	418	322527	0.017 (0.004)	0.045 (0.160)	0.011 (0.002)	0.009 (0.002)
TCGA	**STAD**	371	283734	0.022 (0.004)	0.020 (0.068)	0.013 (0.003)	0.011 (0.002)
TCGA	**TGCT**	129	109128	0.019 (0.003)	0.023 (0.107)	0.011 (0.002)	0.009 (0.002)
TCGA	**THCA**	480	279362	0.023 (0.004)	0.024 (0.066)	0.014 (0.003)	0.011 (0.002)
TCGA	**THYM**	114	157626	0.017 (0.005)	0.412 (0.791)	0.010 (0.003)	0.009 (0.003)
TCGA	**UCEC**	477	351784	0.022 (0.004)	0.059 (0.168)	0.013 (0.002)	0.011 (0.002)
TCGA	**UCS**	53	67853	0.020 (0.004)	0.028 (0.088)	0.011 (0.002)	0.009 (0.002)
TCGA	**UVM**	80	35766	0.017 (0.004)	0.026 (0.097)	0.011 (0.003)	0.009 (0.002)
CPTAC	**BRCA**	105	56318	0.024 (0.005)	0.017 (0.023)	0.014 (0.003)	0.012 (0.002)
CPTAC	**COADREAD**	104	76915	0.025 (0.004)	0.022 (0.045)	0.016 (0.003)	0.013 (0.002)
CPTAC	**GBM**	177	143858	0.019 (0.004)	0.020 (0.073)	0.011 (0.002)	0.009 (0.002)
CPTAC	**HNSC**	108	35064	0.020 (0.004)	0.017 (0.030)	0.012 (0.002)	0.011 (0.002)

CPTAC	**LUAD**	221	425473	0.023 (0.004)	0.019 (0.066)	0.014 (0.002)	0.012 (0.002)
CPTAC	**LUSC**	202	359394	0.020 (0.003)	0.053 (0.178)	0.012 (0.002)	0.011 (0.002)
CPTAC	**PAAD**	164	99093	0.021 (0.004)	0.022 (0.095)	0.012 (0.003)	0.010 (0.002)
CPTAC	**UCEC**	247	184767	0.022 (0.004)	0.042 (0.150)	0.013 (0.002)	0.011 (0.002)

Table S2. Reconstruction Accuracy in Training and Validation Datasets for UNI Encoders.

We compare reconstruction accuracy from the real and reconstructed images for HistoXGAN and other architectures for embedding images in GAN latent space. For comparison, we use encoders designed to recreate images from a StyleGAN2 model trained identically to the HistoXGAN model. The Learned Perceptual Image Patch Similarity (LPIPS) / Deep Image Structure and Texture Similarity (DISTS) encoder uses an equal ratio of LPIPS / DISTS loss between the real and reconstructed images to train the encoder. The Single Layer and Encoder4Editing encoders are trained to minimize L1 loss between UNI feature vectors of the real and reconstructed images.

Source		n	n Tiles	LPIP / DISTS L1 Loss	Single Layer L1 Loss	Encoder4Editing L1 Loss	HistoXGAN L1 Loss
TCGA	**ACC**	56	78789	1.107 (0.067)	0.998 (0.089)	0.800 (0.093)	0.697 (0.092)
TCGA	**BLCA**	378	342463	1.103 (0.076)	0.971 (0.082)	0.855 (0.089)	0.662 (0.112)
TCGA	**BRCA**	943	454985	1.133 (0.058)	0.983 (0.076)	0.866 (0.094)	0.528 (0.087)
TCGA	**CESC**	267	165088	1.041 (0.082)	0.918 (0.090)	0.847 (0.091)	0.670 (0.101)
TCGA	**CHOL**	38	51414	1.131 (0.055)	0.981 (0.076)	0.861 (0.081)	0.611 (0.091)
TCGA	**COADREAD**	428	195493	1.110 (0.061)	0.983 (0.078)	0.912 (0.095)	0.717 (0.120)
TCGA	**DLBC**	43	32073	1.021 (0.098)	0.903 (0.090)	0.810 (0.106)	0.642 (0.105)
TCGA	**ESCA**	147	99625	1.113 (0.061)	0.967 (0.083)	0.860 (0.083)	0.655 (0.102)
TCGA	**HNSC**	401	166061	1.092 (0.077)	0.964 (0.096)	0.883 (0.102)	0.742 (0.131)
TCGA	**KICH**	101	105461	1.094 (0.077)	1.056 (0.068)	0.834 (0.089)	0.763 (0.099)
TCGA	**KIRP**	270	234740	1.121 (0.072)	1.017 (0.075)	0.851 (0.092)	0.717 (0.103)
TCGA	**LGG**	464	155579	1.078 (0.090)	0.994 (0.088)	0.789 (0.103)	0.846 (0.109)
TCGA	**LIHC**	359	331769	1.089 (0.076)	0.978 (0.092)	0.807 (0.091)	0.676 (0.099)
TCGA	**LUAD**	467	335499	1.140 (0.062)	1.003 (0.077)	0.892 (0.091)	0.725 (0.123)
TCGA	**LUSC**	474	370542	1.087 (0.073)	0.942 (0.087)	0.877 (0.091)	0.696 (0.111)
TCGA	**MESO**	73	42242	1.119 (0.066)	0.988 (0.083)	0.846 (0.087)	0.653 (0.096)
TCGA	**OV**	104	108620	1.122 (0.061)	0.999 (0.088)	0.884 (0.095)	0.676 (0.103)
TCGA	**PAAD**	168	119144	1.146 (0.057)	1.016 (0.068)	0.893 (0.081)	0.661 (0.109)
TCGA	**PCPG**	173	207803	1.127 (0.061)	0.988 (0.084)	0.844 (0.106)	0.713 (0.105)
TCGA	**PRAD**	394	202439	1.129 (0.060)	1.021 (0.065)	0.877 (0.085)	0.703 (0.110)
TCGA	**SARC**	250	326262	1.091 (0.082)	0.987 (0.092)	0.849 (0.093)	0.765 (0.112)
TCGA	**SKCM**	418	322527	1.079 (0.080)	0.959 (0.090)	0.829 (0.098)	0.686 (0.116)
TCGA	**STAD**	371	283734	1.107 (0.066)	0.971 (0.085)	0.867 (0.097)	0.662 (0.117)
TCGA	**TGCT**	129	109128	1.117 (0.059)	1.014 (0.067)	0.855 (0.103)	0.704 (0.093)
TCGA	**THCA**	480	279362	1.124 (0.054)	1.020 (0.077)	0.870 (0.096)	0.713 (0.112)
TCGA	**THYM**	114	157626	0.997 (0.115)	0.907 (0.119)	0.849 (0.120)	0.678 (0.110)
TCGA	**UCEC**	477	351784	1.084 (0.064)	0.966 (0.078)	0.869 (0.089)	0.680 (0.091)
TCGA	**UCS**	53	67853	1.112 (0.069)	1.005 (0.072)	0.892 (0.089)	0.747 (0.112)
TCGA	**UVM**	80	35766	1.008 (0.077)	1.008 (0.111)	0.817 (0.135)	0.821 (0.162)
CPTAC	**BRCA**	105	56318	1.128 (0.082)	1.059 (0.081)	0.973 (0.098)	0.714 (0.134)
CPTAC	**COADREAD**	104	76915	1.131 (0.068)	1.070 (0.079)	0.976 (0.089)	0.812 (0.131)
CPTAC	**GBM**	177	143858	1.100 (0.080)	1.047 (0.084)	0.916 (0.096)	0.820 (0.117)
CPTAC	**HNSC**	108	35064	1.081 (0.071)	1.008 (0.074)	0.930 (0.079)	0.750 (0.115)
CPTAC	**LUAD**	221	425473	1.131 (0.055)	1.031 (0.058)	0.920 (0.078)	0.682 (0.113)

CPTAC	**LUSC**	202	359394	1.115 (0.074)	1.005 (0.082)	0.939 (0.093)	0.731 (0.127)
CPTAC	**PAAD**	164	99093	1.124 (0.069)	1.049 (0.073)	0.949 (0.088)	0.697 (0.130)
CPTAC	**UCEC**	247	184767	1.078 (0.068)	1.055 (0.088)	0.958 (0.098)	0.766 (0.128)

Table S3. Dependency of Reconstruction Accuracy on Choice of Training Dataset.

The accuracy of reconstruction of CTransPath features across four HistoXGAN models was compared, with identical training parameters aside from choice of training dataset. One model was trained exclusively on TCGA-LUSC, one was trained exclusively on TCGA-LUAD, one was trained on all eight cancer subtypes listed above, and one was trained across twenty-nine tumor types in TCGA as previously described.

Source		n	n Tiles	LUSC Model L1 Loss	LUAD Model L1 Loss	8 Subtype Model L1 Loss	All TCGA Model L1 Loss
TCGA	**ACC**	56	78789	0.040 (0.006)	0.042 (0.006)	0.038 (0.005)	0.035 (0.004)
TCGA	**BLCA**	378	342463	0.037 (0.006)	0.039 (0.006)	0.035 (0.005)	0.034 (0.004)
TCGA	**BRCA**	943	454985	0.039 (0.007)	0.040 (0.007)	0.033 (0.004)	0.033 (0.004)
TCGA	**CESC**	267	165088	0.036 (0.006)	0.040 (0.006)	0.035 (0.005)	0.034 (0.004)
TCGA	**CHOL**	38	51414	0.038 (0.006)	0.039 (0.005)	0.035 (0.005)	0.034 (0.004)
TCGA	**COADREAD**	428	195493	0.045 (0.012)	0.042 (0.007)	0.034 (0.004)	0.035 (0.004)
TCGA	**DLBC**	43	32073	0.037 (0.006)	0.039 (0.008)	0.035 (0.006)	0.032 (0.005)
TCGA	**ESCA**	147	99625	0.038 (0.007)	0.040 (0.007)	0.036 (0.005)	0.035 (0.005)
TCGA	**HNSC**	401	166061	0.036 (0.005)	0.041 (0.006)	0.035 (0.004)	0.035 (0.004)
TCGA	**KICH**	101	105461	0.049 (0.008)	0.052 (0.010)	0.047 (0.007)	0.036 (0.005)
TCGA	**KIRP**	270	234740	0.044 (0.008)	0.043 (0.008)	0.042 (0.007)	0.036 (0.005)
TCGA	**LGG**	464	155579	0.042 (0.006)	0.044 (0.007)	0.032 (0.005)	0.031 (0.005)
TCGA	**LIHC**	359	331769	0.041 (0.007)	0.043 (0.007)	0.039 (0.006)	0.033 (0.004)
TCGA	**LUAD**	467	335499	0.039 (0.006)	0.035 (0.004)	0.035 (0.004)	0.035 (0.004)
TCGA	**LUSC**	474	370542	0.034 (0.004)	0.039 (0.005)	0.035 (0.004)	0.035 (0.004)
TCGA	**MESO**	73	42242	0.039 (0.006)	0.040 (0.006)	0.036 (0.005)	0.034 (0.004)
TCGA	**OV**	104	108620	0.039 (0.006)	0.039 (0.006)	0.036 (0.005)	0.035 (0.005)
TCGA	**PAAD**	168	119144	0.040 (0.007)	0.041 (0.008)	0.035 (0.005)	0.035 (0.005)
TCGA	**PCPG**	173	207803	0.039 (0.005)	0.040 (0.005)	0.037 (0.005)	0.035 (0.004)
TCGA	**PRAD**	394	202439	0.044 (0.008)	0.043 (0.008)	0.038 (0.006)	0.034 (0.004)
TCGA	**SARC**	250	326262	0.040 (0.008)	0.044 (0.009)	0.037 (0.006)	0.033 (0.005)
TCGA	**SKCM**	418	322527	0.037 (0.005)	0.039 (0.005)	0.036 (0.005)	0.034 (0.004)
TCGA	**STAD**	371	283734	0.041 (0.008)	0.041 (0.008)	0.036 (0.006)	0.035 (0.005)
TCGA	**TGCT**	129	109128	0.039 (0.005)	0.039 (0.006)	0.037 (0.005)	0.034 (0.005)
TCGA	**THCA**	480	279362	0.047 (0.009)	0.046 (0.009)	0.042 (0.008)	0.035 (0.005)
TCGA	**THYM**	114	157626	0.037 (0.006)	0.040 (0.007)	0.036 (0.006)	0.033 (0.006)
TCGA	**UCEC**	477	351784	0.039 (0.005)	0.040 (0.006)	0.035 (0.004)	0.035 (0.004)
TCGA	**UCS**	53	67853	0.039 (0.006)	0.040 (0.005)	0.036 (0.005)	0.035 (0.005)
TCGA	**UVM**	80	35766	0.037 (0.006)	0.041 (0.006)	0.038 (0.006)	0.033 (0.005)
CPTAC	**BRCA**	105	56318	0.042 (0.006)	0.048 (0.007)	0.039 (0.005)	0.040 (0.005)
CPTAC	**COADREAD**	104	76915	0.045 (0.007)	0.050 (0.008)	0.040 (0.005)	0.041 (0.005)
CPTAC	**GBM**	177	143858	0.043 (0.006)	0.044 (0.007)	0.038 (0.006)	0.038 (0.006)
CPTAC	**HNSC**	108	35064	0.041 (0.007)	0.045 (0.006)	0.038 (0.005)	0.038 (0.005)
CPTAC	**LUAD**	221	425473	0.041 (0.006)	0.041 (0.005)	0.038 (0.004)	0.038 (0.004)
CPTAC	**LUSC**	202	359394	0.038 (0.005)	0.041 (0.005)	0.037 (0.004)	0.036 (0.004)
CPTAC	**PAAD**	164	99093	0.043 (0.008)	0.045 (0.008)	0.038 (0.006)	0.037 (0.006)

CPTAC	**UCEC**	247	184767	0.043 (0.007)	0.045 (0.007)	0.039 (0.006)	0.038 (0.006)

Table S4. Homogenous Loss for Models Predicting Tissue Source Site.

Tile based weakly supervised models were trained to predict site four datasets within TCGA. The gradient with respect to site prediction was calculated for the average feature vector across each slide in the dataset. Principal component analysis was applied to these gradients, and components were sorted both by the magnitude of difference of the component between gradients toward each outcome class, i.e. the strength of the contribution of the component to the prediction. This process was repeated when applying Reinhard and CycleGAN normalization. Predictions are dominated by a single uniform principal component without normalization and with Reinhard normalization, but this effect is largely mitigated by CycleGAN normalization.

		Without Normalization		Reinhard Normalized		CycleGAN Normalized	
Source	n	Contribution (Std)	Variance Explained	Contribution (Std)	Variance Explained	Contribution (Std)	Variance Explained
BRCA	943	0.691 (0.937)	0.189	0.616 (0.562)	0.28	0.160 (0.830)	0.068
BLCA	378	0.686 (0.711)	0.28	0.402 (1.711)	0.381	0.258 (1.586)	0.277
COADREAD	428	0.566 (0.713)	0.166	0.315 (0.636)	0.098	0.186 (0.402)	0.052
ESCA	147	0.207 (0.323)	0.067	0.164 (0.637)	0.169	0.442 (0.580)	0.118

Table S5. Correlation of Annotated Histologic Features with *PIK3CA* mutational status.

Previously reported annotations for epithelial, nuclear, and mitotic grade, as well as for necrosis, fibrous foci, and inflammation were compared between cases with / without *PIK3CA* mutations. Adjusted odds ratio was computed for the association of each feature with *PIK3CA* (association listed for high tubule formation score, high nuclear pleomorphism score, high mitosis score, and present necrosis, fibrosis, and inflammation). Additionally, to determine how many annotated cases would be needed to identify these correlations, we calculated these adjusted odds ratios for subsets of 50, 100, 200, and 400, 800 cases of the whole cohort, with these subsets selected at random and repeated 100-fold. The median odds ratio over these 100 iterations is listed below.

Feature	Adjusted Odds Ratio (95% CI)	z-stat	p-value	Median adjusted odds ratio (95% CI) computed with 100 iterations from subset of cases				
				50 cases	100 cases	200 cases	400 cases	800 cases
Decreased Tubule Formation	0.6 (0.44 - 0.82)	-3.22	0.001	0.61 (0.16 - 2.28)	0.64 (0.24 - 1.67)	0.62 (0.30 - 1.27)	0.61 (0.38 - 1.00)	0.60 (0.43 - 0.84)
Nuclear Pleomorphism	0.61 (0.43 - 0.87)	-2.77	0.006	0.55 (0.07 - 4.48)	0.61 (0.20 - 1.83)	0.59 (0.27 - 1.25)	0.60 (0.35 - 1.03)	0.62 (0.42 - 0.90)
Mitosis	0.86 (0.58 - 1.28)	-0.75	0.45	1.0 (0.14 - 7.35)	0.82 (0.23 - 2.99)	0.90 (0.38 - 2.12)	0.88 (0.48 - 1.62)	0.86 (0.56 - 1.32)
Necrosis	0.58 (0.4 - 0.84)	-2.9	0.004	0.57 (0.08 - 4.27)	0.57 (0.17 - 1.89)	0.59 (0.26 - 1.34)	0.59 (0.32 - 1.09)	0.59 (0.39 - 0.88)
Fibrous Focus	1.08 (0.79 - 1.47)	0.49	0.62	1.12 (0.23 - 5.52)	1.06 (0.35 - 3.26)	1.04 (0.48 - 2.25)	1.07 (0.66 - 1.73)	1.09 (0.77 - 1.53)
Inflammation	1.02 (0.73 - 1.43)	0.1	0.92	1.15 (0.21 - 6.46)	1.08 (0.35 - 3.36)	1.03 (0.50 - 2.14)	0.98 (0.58 - 1.66)	1.01 (0.69 - 1.48)

Table S6: Correlation of Annotated Histologic Features with Homologous Recombination Deficiency.

Previously reported annotations for epithelial, nuclear, and mitotic grade, as well as for necrosis, fibrous foci, and inflammation were compared between cases with high / low HRD scores. Adjusted odds ratio was computed for the association of each feature with HRD status (association listed for high tubule formation score, high nuclear pleomorphism score, high mitosis score, and present necrosis, fibrosis, and inflammation). Additionally, to determine how many annotated cases would be needed to identify these correlations, we calculated these adjusted odds ratios for subsets of 50, 100, 200, and 400, 800 cases of the whole cohort, with these subsets selected at random and repeated 100-fold. The median odds ratio over these 100 iterations is listed below.

Feature	Adjusted Odds Ratio (95% CI)	z-stat	p-value	Median adjusted odds ratio (95% CI) computed with 100 iterations from subset of cases				
				50 cases	100 cases	200 cases	400 cases	800 cases
Decreased Tubule Formation	1.65 (1.0 - 2.73)	1.95	0.05	1.84 (0.19 - 17.89)	1.82 (0.47 - 7.11)	1.63 (0.55 - 4.90)	1.62 (0.77 - 3.39)	1.64 (0.97 - 2.75)
Nuclear Pleomorphism	2.69 (1.64 - 4.42)	3.91	9.38E-05	5.25 (0.26 - 106.46)	2.84 (0.57 - 14.07)	2.79 (1.09 - 7.12)	2.80 (1.39 - 5.64)	2.68 (1.61 - 4.48)
Mitosis	3.29 (2.11 - 5.13)	5.26	1.45E-07	4.34 (0.17 - 109.34)	3.86 (0.95 - 15.63)	3.58 (1.46 - 8.82)	3.35 (1.66 - 6.77)	3.29 (2.07 - 5.23)
Necrosis	2.45 (1.61 - 3.74)	4.16	3.23E-05	3.39 (0.41 - 28.15)	2.77 (0.74 - 10.43)	2.46 (0.95 - 6.34)	2.41 (1.31 - 4.41)	2.45 (1.59 - 3.78)
Fibrous Focus	1.0 (0.66 - 1.5)	-0.02	0.98	1.00 (0.0 - inf)	0.97 (0.25 - 3.75)	0.95 (0.39 - 2.35)	0.98 (0.53 - 1.81)	1.00 (0.65 - 1.53)
Inflammation	0.97 (0.64 - 1.47)	-0.13	0.89	0.84 (0.11 - 6.60)	0.95 (0.26 - 3.53)	1.03 (0.43 - 2.45)	0.94 (0.48 - 1.86)	0.98 (0.64 - 1.50)

Table S7. Demographics of Included Patients from All Analyzed Cohorts. Of note, demographic details were not collected / are not available form the UCMC OncoTree cohort. The TCGA cohort includes the n = 8120 training cohort along with cases from TCGA-KIRC used in a validation experiment.

Variable		TCGA (n = 8213)	CPTAC (n = 1415)	UCMC MRI Cohort (n = 934)
Age	Age at Diagnosis, mean (SD)	59.0 (14.7)	62.3 (11.2)	55.7 (12.0)
	Missing Age	48	313	102
Gender	Male	3973	379	0
	Female	4240	220	832
	Missing	0	816	102
Race	White	6061	410	495
	Black	748	39	274
	Asian	608	39	50
	American Indian / Alaska Native	21	2	1
	Native Hawaiian / Pacific Islander	10	0	2
	Missing / Not Reported	765	922	112
Ethnicity	Hispanic / Latino	302	406	28
	Not Hispanic / Latino	6223	30	799
	Missing / Not Reported	1688	979	107

Table S8. Additional Demographics of Included Patients from University of Chicago with Dynamic Contrast Enhanced Magnetic Resonance Imaging and Digital Histology.

Variable		Missing	Overall (n = 934)
Age at Diagnosis, mean (SD)		102	55.7 (12.0)
Histologic Subtype, n (%)	**Invasive ductal carcinoma**	50	664 (75.1)
	Invasive lobular carcinoma		116 (13.1)
	Other		104 (11.8)
Grade, n (%)	**1**	54	100 (11.4)
	2		457 (51.9)
	3		323 (36.7)
Estrogen Receptor Status, n (%)	**Negative**	48	162 (18.3)
	Positive		724 (81.7)
Progesterone Receptor Status, n (%)	**Negative**	54	261 (29.7)
	Positive		619 (70.3)
HER2 Status, n (%)	**Negative**	76	749 (87.3)
	Positive		109 (12.7)
Pathologic Stage Group, n (%)	**1**	402	278 (52.3)
	2		166 (31.2)
	3		87 (16.4)
	4		1 (0.2)

Table S9. Feature Classes Extracted from MRI Images.

Class	Number of Features	Description
Shape-Based	14	Describe the size and shape of the region of interest using the segmentation mask and do not depend on gray-scale values. Examples include sphericity and compactness.
First Order Statistics	**3,348**	Describe the distribution of voxel intensities within the region of interest. Examples include mean, minimum, and maximum gray-scale values.
Gray Level Co-occurrence Matrix	**4,462**	Describe the distribution of co-occurrence of gray-scale values at a given offset distance. Examples include joint average and correlation.
Gray Level Run Length Matrix	**2,960**	Quantify runs of consecutive pixels with the same gray level value. Examples include short run emphasis and long run emphasis.
Gray Level Size Zone Matrix	**2,960**	Describe the number of zones of voxels with the same gray-scale values. Examples include small area emphasis and large area emphasis.
Neighboring Gray Tone Difference Matrix	**45**	Describe the difference between the gray-scale values of a voxel and its neighbors up to a specified distance. Examples include coarseness and busyness.
Gray Level Dependence Matrix	**2,590**	Describe the number of voxels that are dependent on the center voxel. Two voxels are dependent if their gray-scale value differences are below a specified threshold.

Data S1. Accuracy of Image Generation with HistoXGAN Across Rare Histologic Subtypes of Breast, Colon, and Lung Cancer.

L1 loss is listed for CTransPath features from the reconstructed versus source image, categorized by histologic subtype of breast, colon, and lung cancer in the TCGA (training), CPTAC (validation), and University of Chicago Medical Center (UCMC, validation) datasets.

Data S2. Accuracy of Image Generation with HistoXGAN Across OncoTree Diagnoses.

A dataset of 786 cases with 176 unique OncoTree diagnosis codes curated from University of Chicago was used for this analysis.

Data S3. Accuracy of Image Generation with HistoXGAN Across Specimen Source Sites.

A dataset of 786 cases with annotations for specimen source curated from University of Chicago was used for this analysis.

Data S4. Accuracy of Prediction of Clinically Relevant Breast Cancer Gene Signatures in the TCGA Cohort.

Predictions were made for n = 938 cases using three-fold cross validation, and predictions in the held-out cohort were pooled for assessment of accuracy.

Data S5. Accuracy of Prediction of Clinically Relevant Breast Cancer Gene Signatures from Histology Generated from MRI, and Prediction Directly from MRI. (separate file)

Predictions were made for n = 934 cases from University of Chicago with matched histology and MRI. An encoder was trained with five-fold cross validation to translate MRI radiomic features into a histology feature vector, which was then reconstructed to a tile image for prediction of signatures. Histology feature vectors were generated for cases in the held-out test set for each fold of cross validation. Similarly, predictions were made from MRI features directly, using the same cross folds to train simple logistic regression models to predict gene signatures from MRI radiomic features and make predictions in the held-out test sets.