A Pan-Cancer Patient-Derived Xenograft Histology Image Repository with Genomic and Pathologic Annotations Enables Deep Learning Analysis

Brian S. White[1], Xing Yi Woo[1,2], Soner Koc[3], Todd Sheridan[1], Steven B. Neuhauser[4], Shidan Wang[5], Yvonne A. Evrard[6], Li Chen[6], Ali Foroughi pour[1], John D. Landua[7], R. Jay Mashl[8], Sherri R. Davies[8], Bingliang Fang[9], Maria Gabriela Rosa[9], Kurt W. Evans[9], Matthew H. Bailey[10], Yeqing Chen[11], Min Xiao[11], Jill C. Rubinstein[1], Brian J. Sanderson[1], Michael W. Lloyd[4], Sergii Domanskyi[1], Lacey E. Dobrolecki[7], Maihi Fujita[12], Junya Fujimoto[9], Guanghua Xiao[5], Ryan C. Fields[8], Jacqueline L. Mudd[8], Xiaowei Xu[11], Melinda G. Hollingshead[13], Shahanawaz Jiwani[6], Saul Acevedo[3]; PDXNet Consortium, Brandi N. Davis-Dusenbery[3], Peter N. Robinson[1], Jeffrey A. Moscow[13], James H. Doroshow[13], Nicholas Mitsiades[7], Salma Kaochar[7], Chong-xian Pan[14], Luis G. Carvajal-Carmona[14], Alana L. Welm[12], Bryan E. Welm[12], Ramaswamy Govindan[8], Shunqiang Li[8], Michael A. Davies[9], Jack A. Roth[9], Funda Meric-Bernstam[9], Yang Xie[5], Meenhard Herlyn[11], Li Ding[8], Michael T. Lewis[7], Carol J. Bult[4], Dennis A. Dean II[3], and Jeffrey H. Chuang[1]

ABSTRACT

Patient-derived xenografts (PDX) model human intra- and intertumoral heterogeneity in the context of the intact tissue of immunocompromised mice. Histologic imaging via hematoxylin and eosin (H&E) staining is routinely performed on PDX samples, which could be harnessed for computational analysis. Prior studies of large clinical H&E image repositories have shown that deep learning analysis can identify intercellular and morphologic signals correlated with disease phenotype and therapeutic response. In this study, we developed an extensive, pan-cancer repository of >1,000 PDX and paired parental tumor H&E images. These images, curated from the PDX Development and Trial Centers Research Network Consortium, had a range of associated genomic and transcriptomic data, clinical metadata, pathologic assessments of cell composition, and, in several cases, detailed pathologic annotations of neoplastic, stromal, and necrotic regions. The amenability of these images to deep learning was highlighted through three applications: (i) development of a classifier for neoplastic, stromal, and necrotic regions; (ii) development of a predictor of xenograft-transplant lymphoproliferative disorder; and (iii) application of a published predictor of microsatellite instability. Together, this PDX Development and Trial Centers Research Network image repository provides a valuable resource for controlled digital pathology analysis, both for the evaluation of technical issues and for the development of computational image–based methods that make clinical predictions based on PDX treatment studies.

Significance: A pan-cancer repository of >1,000 patient-derived xenograft hematoxylin and eosin–stained images will facilitate cancer biology investigations through histopathologic analysis and contributes important model system data that expand existing human histology repositories.

Introduction

The high clinical failure rate of cancer therapies is often attributed to the lack of tumor heterogeneity in preclinical models (1). This concern has motivated the increased use of patient-derived xenografts (PDX), in which a fresh human tumor biopsy is implanted subcutaneously or orthotopically in the flank of an immunodeficient mouse. The level of immunodeficiency is impacted by mouse strain, including nude, nonobese diabetic (NOD), NOD/ severe combined immunodeficient (NOD/SCID), NOD/SCID/IL2 receptor gamma null (NSG), and SCID-beige mice (2). If the

[1]The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut. [2]Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [3]Velsera, Charlestown, Massachusetts. [4]The Jackson Laboratory, Bar Harbor, Maine. [5]University of Texas Southwestern Medical Center, Dallas, Texas. [6]Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland. [7]Baylor College of Medicine, Houston, Texas. [8]Washington University School of Medicine, St. Louis, Missouri. [9]University of Texas MD Anderson Cancer Center, Houston, Texas. [10]Simmons Center for Cancer Research, Brigham Young University, Provo, Utah. [11]The Wistar Institute, Philadelphia, Pennsylvania. [12]Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah. [13]National Cancer Institute, Bethesda, Maryland. [14]University of California, Davis, Davis, California.

PDXNet consortium list included in Supplementary Data.

B.S. White, X.Y. Woo, and S. Koc contributed equally to this article as co-first author.

Corresponding Author: Jeffrey H. Chuang, The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06032. E-mail: jeff.chuang@jax.org

implantation successfully establishes the model in the P_0 mouse, the tumor can be passaged in future generations (P_1, P_2, etc.). After being sufficiently expanded, the model may be used for preclinical drug trials, and these have successfully predicted therapeutic outcome in patients (2).

PDXs have been shown to recapitulate phenotypes of their human progenitors along other dimensions as well. For example, PDX models exhibit metastatic patterns similar to those of their progenitors (https://pdxportal.research.bcm.edu; bioRxiv 2023.02.15.528735; ref. 3). Furthermore, PDX mRNA expression profiles correlate with those of their progenitors (4, 5), and these can be stably maintained for at least 15 generations (6). Similar consistency between PDX and human has been demonstrated for copy number alterations across 21 passages (7) and for methylation profiles (8). Finally, PDXs retain the invasive histologic phenotype of their matched progenitor, as reflected in staining with hematoxylin and eosin (H&E; https://pdxportal.research.bcm.edu; bioRxiv 2023.02.15.528735; refs. 6, 8).

This latter finding raises the intriguing possibility that recent successes applying deep learning (DL) for whole-slide (WSI) analysis of H&E images in human clinical samples (9) will be applicable to PDXs, as well. DL-based analyses of human data are capable of predicting metastases (10), gene mutations (11) and expression (12), survival (13), cancer types (14), molecular (15), clinical (16), and histologic (17) tumor subtypes, and response to both chemotherapy (18) and immune checkpoint inhibitors (19). Much of the groundbreaking work in the field leverages convolutional neural networks (CNN), in which convolutional layers slide over two- or three-dimensional image patches to mathematically summarize them and are thus particularly suited to spatial processing (20). Other approaches have been applied to biomedical image analysis, including autoencoders, which learn a latent representation of the input that can optimally reconstruct it (21), and generative adversarial networks (GAN), in which two networks are trained simultaneously—one to generate images and a second to discriminate between those generated images and real images (22). Most recently, transformers have gained in popularity owing to their improved performance and ability to capture long-range dependencies (23). DL approaches can, in principle, assess the complexity of spatial interactions between cancer, stromal, and immune cells reflected in H&E images, thus moving beyond cancer cell-intrinsic, univariate gene biomarkers. For example, DL methods were able to associate tumor-infiltrating lymphocyte spatial structure with survival and differentiate malignant breast tumor samples from benign breast biopsies based on stromal signatures (24).

The explosive growth of DL-based applications in digital pathology has been made possible by extensive, public repositories of human H&E images, such as The Cancer Genome Atlas (TCGA), The Cancer Imaging Archive, and the Imaging Data Commons. No similar resource exists for PDX images. Here, we describe a large-scale repository of >1,000 PDX and >100 matched patient tumor H&E images, along with expansive genomic, transcriptomic, clinical, and pathologic annotations. The images were curated as part of the NCI's PDX Development and Trial Centers Research Network (PDXNet) program, aimed at collaborative model development and preclinical testing of targeted therapeutic agents. Thumbnails and clinical metadata of the images, along with the raw genomic and transcriptomic data, are available on the PDXNet portal (25). The raw images are hosted on the Seven Bridges Cancer Genomics Cloud (CGC; www.cancergenomicscloud.org; ref. 26), in which they are publicly accessible (https://cgc.sbgenomics.com/u/brian.white/pdxnet-image-repository). We present several use cases demonstrating the

ways in which the application of DL to PDX images in this repository can classify (i) neoplastic, stromal, and necrotic regions; (ii) xenograft-transplant lymphoproliferative disorder (XTLD) cases; and (iii) microsatellite instability (MSI) samples. The repository should facilitate monitoring of potential divergence between a PDX and its human progenitor during passaging (27) and exploration of questions important to PDX models, such as the impact of human cell turnover within the xenografts (28). We expect this PDXNet image repository to be valuable for controlled digital pathology analysis and development of novel computational methods based on spatial behaviors within patient-derived cancer tissues.

Materials and Methods

H&E image repository

Images were collected from Baylor College of Medicine (BCM), Huntsman Cancer Institute, MD Anderson Cancer Center (MDACC), The Wistar Institute (WISTAR), Washington University in St Louis (WUSTL), and The Jackson Laboratory (JAX). They are organized hierarchically according to these sites (Supplementary Figs. S1 and S2), as further described in Supplementary Methods.

The images and all associated transcriptomic and genomic data, metadata, and annotations are hosted on the Seven Bridges CGC (www.cancergenomicscloud.org; ref. 26), linked according to their respective identifiers (Supplementary Tables S1 and S2). The CGC colocalizes data and computational resources, including those with graphics processing units, in the cloud to enable efficient analysis. We have implemented the major components of our processing pipeline, namely, quality control and tissue mask generation with HistoQC and cell segmentation and phenotyping with HoVer-Net, as publicly available "apps"—effectively containers that can be automatically distributed across computational cloud resources for parallel processing of images. All analysis and processing for this study were performed on the CGC. Execution of HoVer-Net is computationally intensive, and it benefited from graphics processing units available on the CGC. All genomic and transcriptomic data are hosted on the PDXNet portal (25).

Regional annotation

A board-certified pathologist (TS) annotated tumor, stromal, and necrotic regions. Regions were labeled using QuPath (version 0.3.0-rc1). More specifically, a region comprised of at least 50% of a particular cell type when viewed at $20\times$ resolution was annotated with that cell type.

Whole-exome sequencing–derived homologous recombination deficiency, tumor mutational burden, and microsatellite instability annotations

Homologous recombination deficiency (HRD), tumor mutational burden (TMB), and MSI annotations were downloaded from the PDXNet portal on September 12, 2023 and were generated from whole-exome sequencing (WES) data, as previously described (25) and further detailed in Supplementary Methods.

HistoQC

Prior to downstream analysis (e.g., with HoVer-Net or Inception v3), H&E image quality control (QC) and tissue mask generation were performed using HistoQC (29), a tool for artifact detection in digital pathology slides. Per-dataset configuration and QC results are provided in Supplementary Tables S3 and S4 and in Supplementary Methods.

HoVer-Net

Cell nuclei were segmented and phenotyped as neoplastic, necrotic, connective, inflammatory, nonneoplastic epithelial, and unlabeled using HoVer-Net (30). HoVer-Net is a CNN that exploits information encoded within the vertical and horizontal distances of nuclear pixels relative to their centers of mass to simultaneously segment and classify nuclei. Here, HoVer-Net–predicted inflammatory cells are reported as immune. Connective cells are generally reported as stromal, except in correlating with pathologist assessment, in which stromal cells are taken to be those predicted by HoVer-Net as either connective or inflammatory. HoVer-Net was run using the pretrained PanNuke model.

Histology-based Digital Staining

As an alternative to HoVer-Net, the Mask-RCNN Histology-based Digital (HD) Staining (HD-Staining; 31) was applied to segment and phenotype nuclei within a single lung H&E image. HD-Staining has proven effective in classifying a rich set of cell types, including a tripartite separation of immune cells into macrophages, red blood cells, and a broader immune category. Tissue segmentation was first performed using Otsu thresholding followed with morphologic dilation and erosions (32). A 256×256 pixel window was slid over the $20\times$ H&E image with a step size of 226 pixels. Cell nuclei were then simultaneously segmented and classified as tumor, stroma, lymphocyte, red blood cell, macrophage, and karyorrhexis (31). Only the nuclei with centroids located within the central 226×226 pixel area were kept to minimize the edge effect.

Inception v3

A total of 2,048 features of 512×512 pixel tiles at $20\times$ resolution (0.50 μm per pixel) were computed as the outputs of the average pooling layer of the Inception v3 CNN pretrained on ImageNet, as previously described (33).

Labeling tiles with tissue, region, and HoVer-Net information

Tissue masks, regional annotations, and HoVer-Net output were all summarized at the level of 512×512 pixel, nonoverlapping tiles at $20\times$ magnification, so as to conform to those outputs by Inception v3 here and in our previous studies (33). All images were provided at $20\times$ magnification, except those contributed by WUSTL and WISTAR, which were provided at $40\times$ magnification. Regardless, derived masks, annotations, Inception features, and HoVer-Net results were scaled to $20\times$ magnification, in which they were overlapped as described in Supplementary Methods.

Whole-genome duplication calculation

Whole-genome duplication (WGD) status for each PDX sample was inferred from WES data. Sequenza was used to calculate the allele-specific copy number, purity, and ploidy from matched tumor–normal samples. A permutation test was implemented for significance of being WGD given an allelic specific copy number profile and different ploidy and P value cut-offs were applied to call WGD (P value ≤ 0.001 with ploidy ≤ 3; P value ≤ 0.05 with ploidy $=4$; or all samples with ploidy ≥ 5).

Correlation with pathologic assessment

Fractions of phenotyped cell types predicted by HoVer-Net were correlated with those provided through pathologic assessment for neoplastic, stromal, and necrotic cells. Unlike in other analyses, in particular prediction of tissue regions, stromal cells were here considered those predicted to be either inflammatory/immune or connective cells by HoVer-Net. We assessed tumor types with at least seven PDX samples and performed the correlation only over PDX (and not human progenitor) samples. We estimated the fraction of cells of a given (pheno)type predicted by HoVer-Net as the ratio of the total area of that cell type within the slide to the total tissue area of the slide, both properly scaled (e.g., from the $1.5\times$ magnification applied by HistoQC in generating the tissue mask and as described above). We report weighted Pearson correlations to account for the statistical nonindependence of multiple slides from the same patient. In particular, we assign to each slide a weight proportional to the inverse of the number of slides from that patient and use these as input to cov.wt in R.

Tissue region prediction

Tissue regions were predicted at the tile level using a random forest classifier trained on proportion and counts of each HoVer-Net–predicted cell type and the total number of cells and evaluated using fivefold cross-validation with the ranger R package. Tiles were filtered to ensure at least 50% overlap with the tissue mask or at least 50% overlap with one of the annotated regions. Folds were defined at the patient level, independently within each site [BCM, MDACC, and Patient-Derived Models Repository (PDMR)], and stratified by diagnosis [breast cancer, lung adenocarcinoma, and squamous cell lung cancer (LUSC)] using createFolds in the caret R package. In total, slides from 41 patients were used for training and evaluation.

During training, tiles were weighted so as to give equal weight across diagnoses, across sites within a diagnosis, across human progenitors within a site/diagnosis pair, and across tiles within a human progenitor. Weighting was performed independently for each of the five folds (i.e., weights assigned to all but the one held-out fold) as well as across the entire dataset. Model training was performed independently for each fold and also for the entire dataset using ranger with the weights passed to the case.weights argument and with probability = TRUE, mtry = NULL, and importance = "impurity". Held-out tiles were predicted using the corresponding model (i.e., trained on the four non held-out folds), and a confusion matrix was generated comparing them with their pathologist annotations. Furthermore, accuracy, precision, recall, and specificity of the predictions relative to the pathologist ground truth were calculated using the yardstick R package. Precision, recall, and specificity were calculated with estimator = "macro_weighted."

HoVer-Net–based XTLD prediction

XTLD status was predicted at the tile level using a random forest classifier trained on the proportion of each HoVer-Net–predicted cell type, and not individual cell type or total cell counts, and evaluated using fivefold cross-validation with the ranger R package, similar to as described in "Tissue region prediction." Specifically, slides were filtered so as to be derived from PDX (and not human progenitor) samples diagnosed with XTLD or LUSC. To reduce technical variability, within each site, slides were filtered so as to be digitized by the most frequently used scanner at that site. The scanner was identified by extracting the tiff.ImageDescription property field from the WSI using OpenSlide in python and then parsing its identifier from the 'ScanScope ID =' string within the image description. These were filtered to retain only those with >99% overlap with the tissue mask. Only one slide was retained per human progenitor, favoring XTLD slides over LUSC slides from the same human progenitor and then retaining that with the most tiles. Folds and tile weights were defined using the same procedure as

described in "Tissue region prediction," except relative to sites PDMR and MDACC and to diagnoses XTLD and LUSC. Training and evaluation of a random forest model was also performed as described in the "Tissue region prediction," with the exception that only cell type proportions and not counts were used as features. Furthermore, this was treated as a "weakly supervised" problem, in which tiles were labeled according to the diagnosis of the entire slide (i.e., XTLD or LUSC). Reported variable importance was accessed from the variable.importance field of the model returned by ranger function and trained on the entire dataset (rather than on individual folds). ROC curves and AUCs were calculated using the pROC R library by first generating a "roc" object using the roc function, by plotting the curve using ggroc, and by calculating the AUC using auc.

Inception v3–based XTLD prediction

XTLD status was predicted at the tile level using a random forest classifier trained on the 2,048 tile-level features computed by Inception v3 and evaluated using fivefold cross-validation with the ranger R package, exactly as described in the "HoVer-Net–based XTLD prediction," except as noted here. In particular, to ensure a fair comparison across methods, the same folds used to train and evaluate the HoVer-Net–based XTLD classifier were used in training and evaluating the Inception v3–based XTLD classifier. The same tile filtering and weighting procedure was used as described in "HoVer-Net–based XTLD prediction," though the included tiles and their weights differed.

The Inception v3–based model trained on all (MDACC and PDMR) XTLD and LUSC data was applied to the held-out JAX XTLD and LUSC images. This was done in a blinded fashion, with the labels only revealed after predictions were made. Tiles from JAX images were filtered to retain only those with >99% overlap with the tissue mask. Representative tiles were selected as those with high (or low) value for the Inception v3 feature with highest variable importance, feature 1895. This was done by ordering the tiles by increasing feature 1895 value, retaining only one tile per human progenitor, and selecting the five tiles with largest value. A similar procedure was repeated to select tiles with low values.

MSI prediction

MSI status was predicted using nine pretrained models downloaded from https://zenodo.org/record/5151502. As previously described (34), these models were trained by excluding one of nine datasets—DACHS, DUSSEL, MECC, QUASAR, RAINBOW, TCGA, UMM, YORKSHIRE, or MUNICH. We report either predictions for each model independently or as the mean prediction across all models, as indicated. Our python prediction code (MSIpred-revised.py) is based on that provided by the authors of the original manuscript at https://github.com/KatherLab/preProcessing.git. Briefly, we processed WSI images at 20× resolution into 512 × 512 pixel tiles without overlap. Individual tiles passing quality control (Canny edge detection as implemented in python's OpenCV library) were normalized by applying the Macenko method (35) and the template image (https://github.com/jnkather/DeepHistology/blob/master/subroutines_normalization/Ref.png) originally used in training the models. Normalized tiles were resized to 224 × 224 pixels using INTER_CUBIC interpolation in OpenCV and passed to the models for prediction.

Enrichment of microsatellite stable (MSS) labels near the low end of sorted median values over tiles (of mean predicted MSI probabilities over the nine models) was calculated using the fgsea package in R. In particular, we invoked fgseaSimple with parameters nperm = 1,000,

gseaParam = 0, and scoreType = "neg." Setting gseaParam = 0 has the effect of considering only the ordering of the values in calculating the enrichment score, rather than their magnitudes. Setting scoreType = "neg" has the effect of considering enrichment of MSS samples near the low end of predicted MSI probabilities (i.e., near the high end of predicted MSS probabilities).

Data availability

Raw H&E images, metadata, HistoQC output, HoVer-Net output, and Inception features are hosted on the Seven Bridges CGC (www.cancergenomicscloud.org) at https://cgc.sbgenomics.com/u/brian.white/pdxnet-image-repository. Scripts in the GitHub repository https://github.com/TheJacksonLaboratory/pdxnet-image-analysis-aacr2022 were used to generate the figures (Supplementary Table S5) and tables (Supplementary Table S6). Genomic, transcriptomic, and clinical metadata hosted on the PDXNet portal may be accessed at https://portal.pdxnetwork.org/, with data download described under About > Data Access. All other raw data are available upon request from the corresponding author.

Results

A pan-cancer, multi-institutional repository links histology images to clinical annotations, pathologic assessments, and genomic and transcriptomic data

The PDXNet repository consists of 1,094 H&E images from 351 PDX models (Supplementary Tables S7 and S8; see Supplementary Methods). They represent 37 cancer types, including colon adenocarcinoma (242 PDX images from 58 models), pancreatic ductal adenocarcinoma (PDAC; 119 PDX images from 41 models), breast cancer (a category encompassing ductal carcinoma *in situ*, lobular carcinoma *in situ*, invasive breast carcinoma, invasive lobular carcinoma, and breast cancer not otherwise specified; 91 PDX images from 74 models), lung adenocarcinoma (78 PDX images from 22 models), LUSC (76 PDX images from 23 models), and skin cutaneous melanoma (SKCM; 58 PDX images from 24 models; **Fig. 1A**; Supplementary Fig. S3; Supplementary Tables S9 and S10).

Images were generated at five PDX Development and Trial Centers within PDXNet—BCM, Huntsman Cancer Institute, MDACC, WISTAR, and WUSTL—as well as at the NCI PDMR and JAX. Metadata were collected from these sites and aggregated from existing portals hosted at BCM (https://pdxportal.research.bcm.edu; bioRxiv 2023.02.15.528735) and PDMR (https://pdmr.cancer.gov/). These metadata include sample type (PDX or human progenitor), mouse strain, engraftment site (e.g., subcutaneous or mammary fat pad), diagnosis, primary cancer site, sex, age, race, and ethnicity (Supplementary Table S7). Background PDX strains are predominantly NSG ($n = 221$ models), occasionally SCID-beige ($n = 40$), and rarely others ($n = 2$). Most PDX images are from early passages, with those from P_0, $n = 210$; P_1, $n = 389$; and P_2, $n = 214$ predominating over those from later passages ($>P_2$, $n = 194$; **Fig. 1B**).

The repository facilitates analyses both between PDX models and their human progenitors and across data modalities. It pairs images from 139 PDX models with those from 128 unique progenitors (**Fig. 1A**; Supplementary Table S11). Of these, cancer diagnoses with images from at least four unique, paired progenitors include breast cancer, colon adenocarcinoma, LUSC, lung adenocarcinoma, non-rhabdomyosarcoma soft tissue sarcoma (a category encompassing noninfantile fibrosarcoma, gastrointestinal stromal tumor, non-uterine leiomyosarcoma, malignant peripheral nerve sheath tumor,
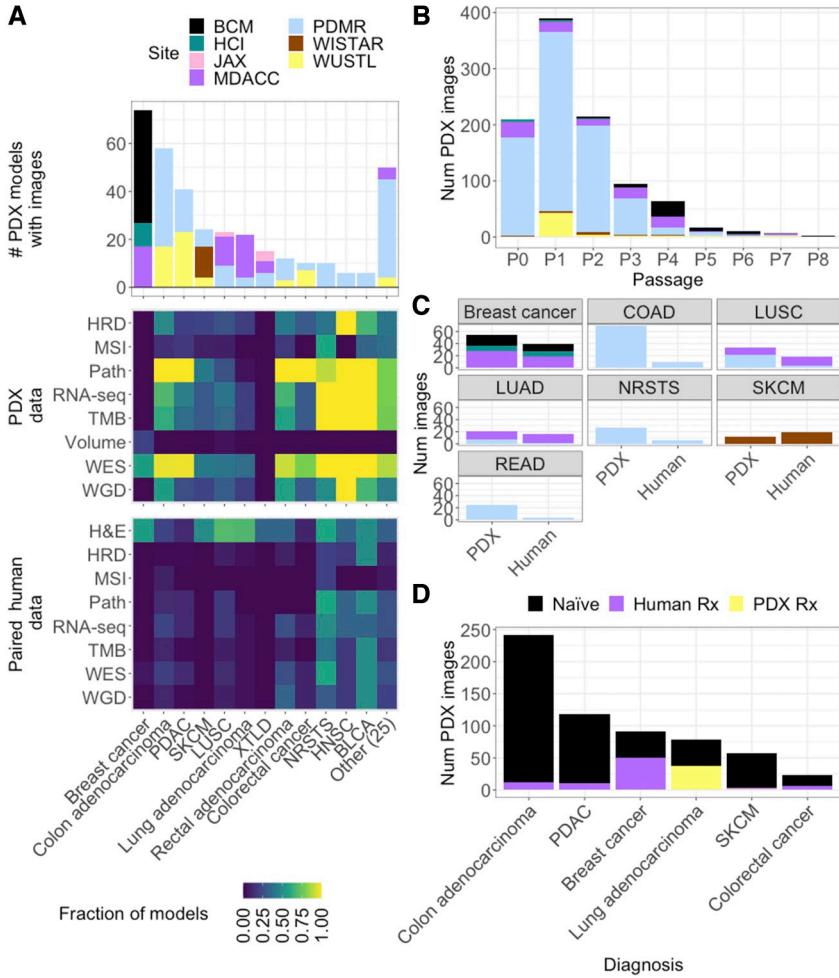
Figure 1.
PDXNet image repository captures heterogeneity across tumor types, passages, and treatment status and pairs human and PDX histology images with pathologic assessment and genomic and transcriptomic data. **A,** Top, distribution of the number of PDX models with H&E images by cancer diagnosis. Fraction of PDX models of a given diagnosis paired with HRD and MSI status, slide-level pathologic assessment (Path), WGD predictions, and RNA-seq, Middle, TMB, tumor volume, and WES data. Bottom, fraction of PDX models with paired human data of the indicated modality. **B,** Distribution of PDX images by passage. **C,** Counts of images for those derived from human progenitor/PDX pairs, restricted to diagnoses with at least four such progenitors. **B** and **C,** Bar charts are colored according to the contributing site as indicated in **A**. **D,** Distribution of PDX images across diagnoses with at least two treated PDXs or at least two treated progenitors. BLCA, bladder cancer; HNSC, head and neck squamous carcinoma; NRSTS, nonrhabdomyosarcoma soft tissue sarcoma; Rx, treatment.

and synovial sarcoma), SKCM, and rectal adenocarcinoma (**Fig. 1C**). The repository includes images from mice that were treated therapeutically, whose human progenitor was treated therapeutically, or that are completely treatment-naïve. Diagnoses having at least two treated PDXs or PDXs derived from at least two treated progenitors include colon adenocarcinoma, PDAC, breast cancer, lung adenocarcinoma, LUSC, SKCM, and colorectal cancer (**Fig. 1D**).

H&E images are linked to transcriptomic and genomic data and derived results, hosted on the PDXNet portal (25): 151 models are characterized by RNA sequencing (RNA-seq) and 250 by WES; progenitors from 48 models are profiled by RNA-seq and from 51 models by WES (**Fig. 1A**; Supplementary Tables S10 and S11). PDX models and progenitors from PDMR are annotated with HRD and MSI statuses, as well as TMB, derived from WES data (see Materials and Methods; **Fig. 1A**; Supplementary Tables S10 and S11). Additionally, most images include pathologic assessment of tumor stage ($n = 890$) and slide-level proportions of cancer, stromal, and necrotic regions ($n = 866$; **Fig. 1A**; Supplementary Table S7). Tumor volume was profiled longitudinally following therapeutic treatment for a limited number of models (**Fig. 1A**; Supplementary Table S10).

Pathologist cell-type estimates allow the evaluation of nuclei segmentation and classification

To further characterize the images, we applied HoVer-Net, a nuclei segmentation and classification method trained on *human* H&E images (30), to the *PDX* images. HoVer-Net classifies nuclei as neoplastic, stromal ("connective"), necrotic, immune ("inflammatory"), nonneoplastic epithelial, or other. We first observed that tumor-predicted nuclei were larger in PDX samples with WGD than in samples with diploid genomes (Supplementary Fig. S4), consistent with prior studies of TCGA H&E images (36). Next, to approximate a pathologist's estimate of tissue content, we derived a *slide*-level estimate of neoplastic, stromal, and necrotic tissue content by aggregating predicted *nuclei*-level area within each nuclei class (see Materials and Methods). HoVer-Net–based slide-level estimates generally correlated well for neoplastic (median *weighted* Pearson correlation $r = 0.54$), stromal ($r = 0.54$), and necrotic ($r = 0.70$) cells across diagnoses (colon adenocarcinoma, colorectal cancer, and PDAC) and datasets (PDMR and WUSTL; **Fig. 2A–D**; Supplementary Fig. S5A–S5C). Collectively, these analyses suggest that nuclei segmentation and classification methods trained on human H&E images are applicable to PDX H&E images, as well.
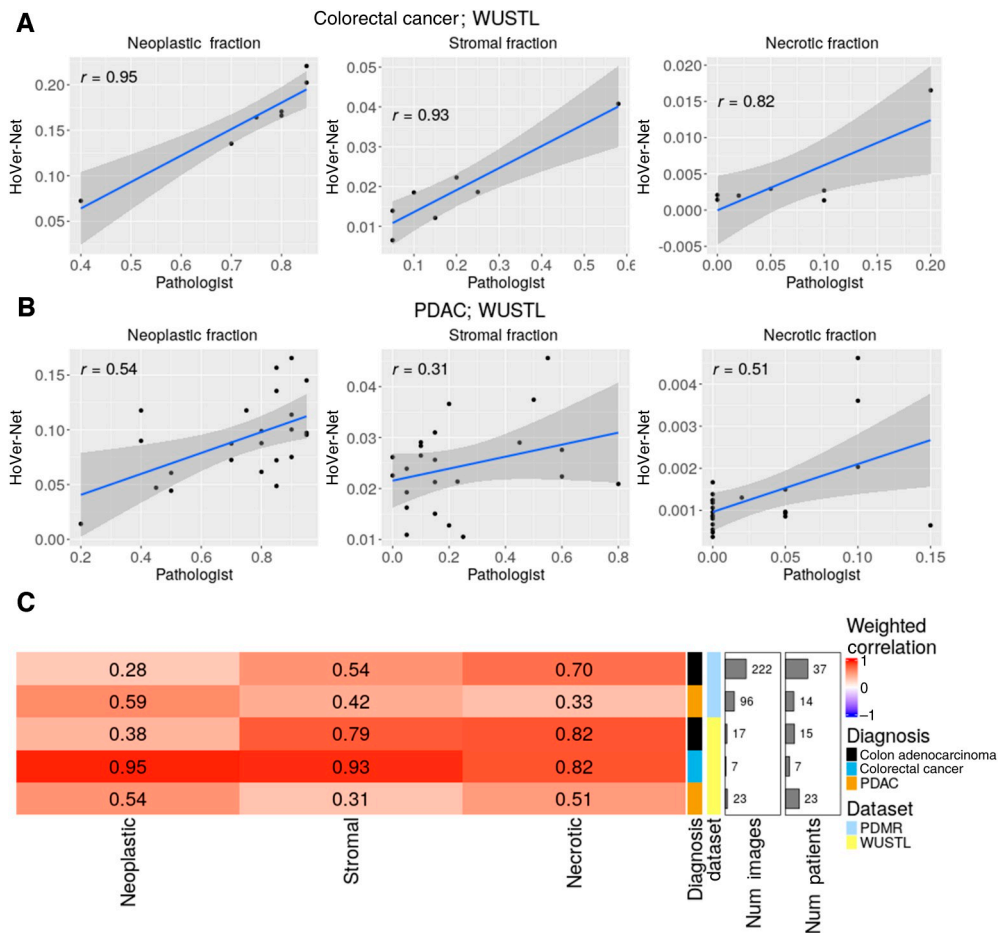
Figure 2.
Cell segmentation and classification trained on human data correlate with pathologist estimates of cell-type proportions in PDX samples. **A** and **B,** Correlation of pathologist (*x*-axis) and HoVer-Net–based slide-level estimates (*y*-axis) for neoplastic (left), stromal (middle), and necrotic (right) cell proportions in colorectal cancer (**A**) and PDAC (**B**) from WUSTL. Each point corresponds to an individual image and PDX (i.e., no replication). **C,** Weighted Pearson correlations of pathologist and HoVer-Net–based estimates of neoplastic, stromal, and necrotic cell proportions in colon adenocarcinoma, colorectal cancer, and PDAC in the PDMR and WUSTL datasets. All PDX strains are NSG. Colorectal cancer not otherwise specified, for example, as colon adenocarcinoma or rectal adenocarcinoma. Correlations are computed over images, inversely weighted by the number of H&E images from each patient (see Materials and Methods).

Pathologist pixel-level annotations enable training of a regional classifier

To spatially contextualize slide-level correlations between predicted and assessed cell type proportions, a board-certified pathologist manually annotated cancer epithelium ("tumor"), stroma, and necrotic regions within H&E images of LUSC ($n = 15$), lung adenocarcinoma ($n = 18$), and breast cancer ($n = 8$; **Fig. 3A–F**). We labeled tiles according to their intersection with these annotations and found that, as expected, tumor tiles were predominantly composed of cells classified as neoplastic in the lung datasets (Supplementary Fig. S6). Consistently, the spatial organization of predicted tumor cells closely resembled pathologist annotations (**Fig. 3A**). Breast cancer exhibited a different pattern, with most tumor tiles composed of a mix of neoplastic and stromal cells (Supplementary Fig. S6). Lung stromal tiles harbored predicted neoplastic, immune, and stromal cells, whereas breast stromal tiles were predominantly composed of stromal cells. Finally, necrotic tiles had levels of predicted necrotic cells in

excess of those of other tiles in both tumor types, despite the additional presence of a larger proportion of neoplastic and immune cells in lung cancer and an appreciable proportion of stromal cells in breast cancer. These apparent differences between breast and lung sample cellular composition may be partially explained by experimental and technical artifacts—in particular, we observed lighter hematoxylin staining in the breast H&E images and a concomitant incidence of unsegmented, incorrectly segmented, and misclassified nuclei relative to lung images (Supplementary Fig. S7A and S7B). To begin to address true cellular heterogeneity within the stromal and necrotic regions, we applied a second nuclei segmentation and phenotyping approach, HD-Staining (31), to a PDX lung sample (**Fig. 3B**). The spatial pattern of HD-Staining phenotypes was broadly similar to that predicted by HoVer-Net—though HD-Staining's prediction of macrophages within one necrotic area was more fine-grained than the immune prediction of HoVer-Net and the two disagreed about the immune content within one stromal region.
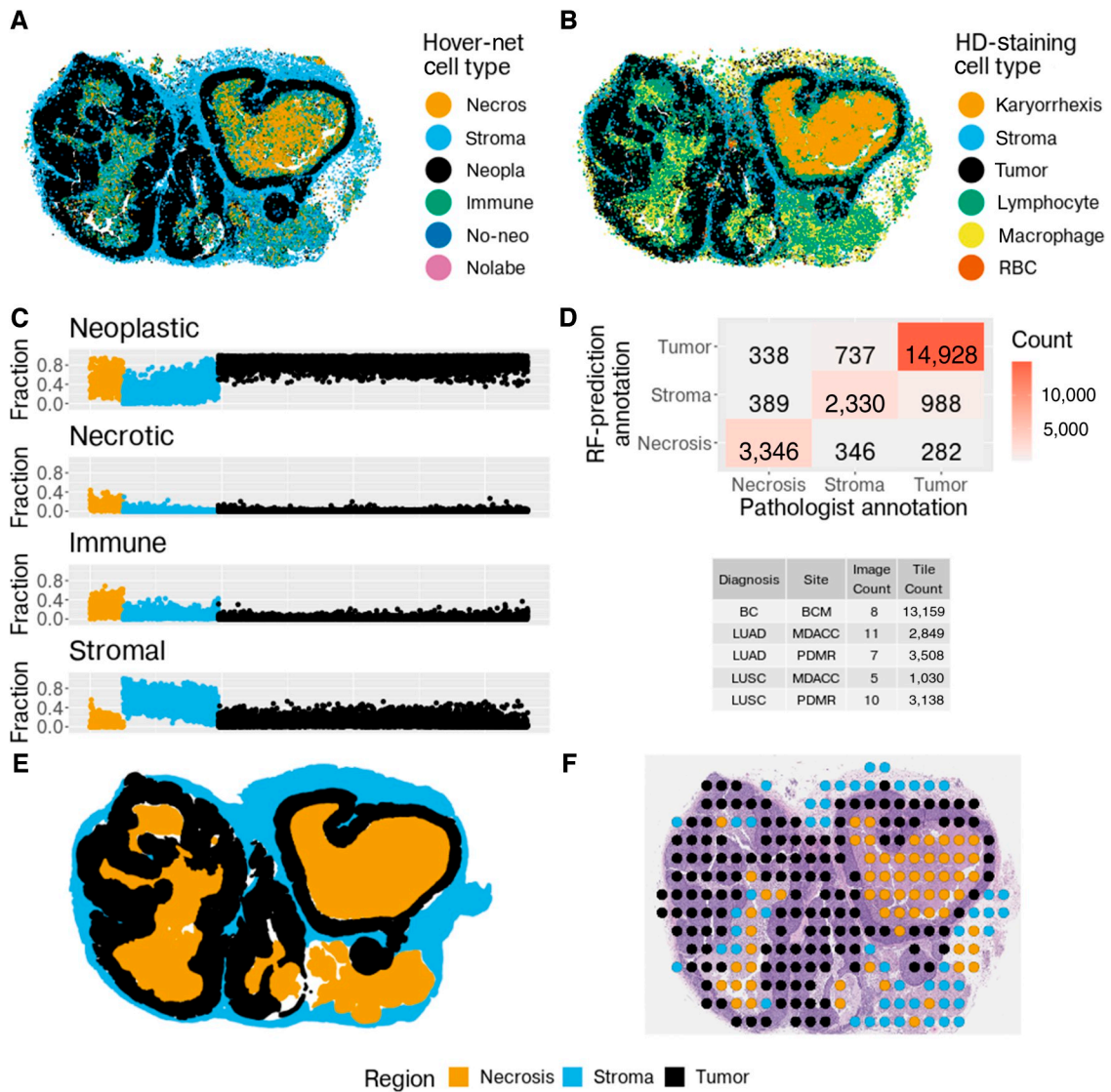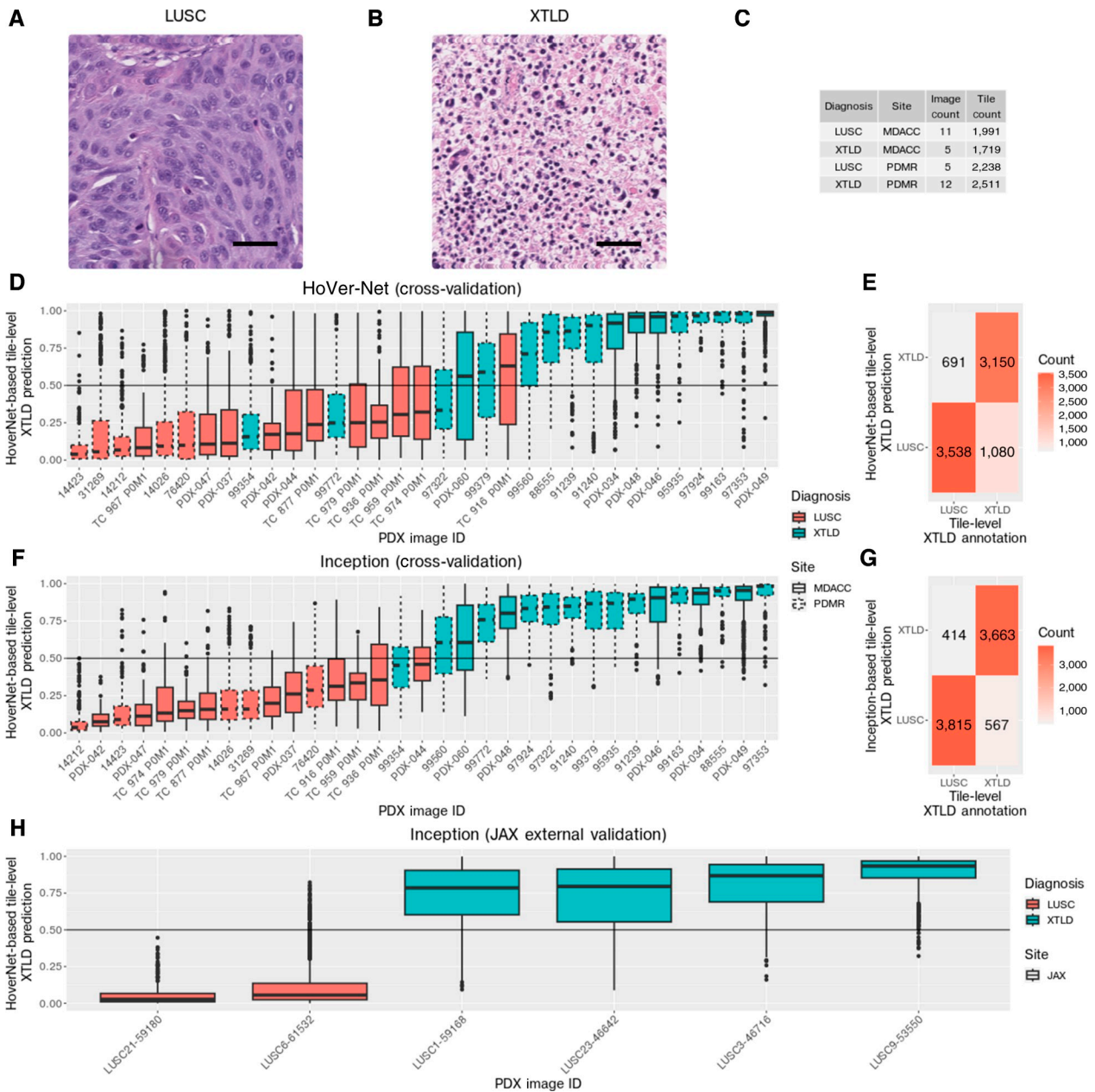
Figure 3.
Regional classifier predicts cancer epithelium ("tumor"), stroma, and necrotic areas based on nuclei classification. **A** and **B,** Cell phenotypes predicted by HoVer-Net (**A**) or HD-Staining (**B**) within a PDX lung tumor. **C,** Labeled tiles (orange, necrosis; blue, stroma; black, tumor) represented according to the cellular fraction that is predicted by HoVer-Net to be neoplastic, necrotic, immune, or stromal. **D,** Top, confusion matrix comparing pathologist-provided tile labels with corresponding prediction from the random forest (RF) HoVer-Net classifier. Results shown for held-out set following fivefold cross-validation, with folds defined according to the patient/progenitor (see Materials and Methods). Bottom, image and tile count used for cross-validation stratified by diagnosis and site. **E,** Pathologist regional annotations. **F,** Random forest HoVer-Net–based classifier predictions of regions (within 512 × 512 pixel tiles). necros, necrotic; neopla, neoplastic; nolabe, unlabeled; no-neo, nonneoplastic epithelial; RBC, red blood cell.

The localization of particular cell types within morphologic regions has clinical implications. For example, tumor–stromal interactions have been associated with disease progression and chemotherapy resistance in breast cancer (37). Although regional annotations facilitate correlative studies with clinical phenotypes, the manual burden in generating them is high. In our studies, a pathologist expended approximately 20 hours per sample demarcating high-resolution tumor, stromal, and necrotic regions within it, in detail (see Materials and Methods).

To alleviate this annotation burden, we instead sought to predict tumor, stroma, and necrotic regions within H&E images. We first profiled each tile according to the fraction (**Fig. 3C**) and count (Supplementary Fig. S8A–S8D) of each HoVer-Net–predicted cell type within it. Consistent with our previous results, we observed that these tile-level cell-type profiles were region- and cancer type–specific. We used the profiles as features and applied fivefold cross-validation at the level of patients/progenitors to train a random forest that probabilistically predicted regions using the features. Prediction of pathologist annotations was strong (**Fig. 3D**; accuracy and macroweighted precision, recall, and specificity in the range 0.87–0.89), and predictions correctly recapitulated irregular region boundaries (**Fig. 3E** and **F**). Furthermore, prediction performance

Figure 4.
XTLD can be computationally predicted from H&E images using segmented and classified cells or DL features. **A** and **B,** H&E images of LUSC (**A**) and XTLD (**B**). Each image from a distinct human progenitor. Images at 20× resolution. Scale bars, 50 μm. **C,** Image and tile count used for cross-validation stratified by diagnosis and site. **D** and **E,** HoVer-Net–based classifier performance. **D,** Distribution of tile-level XTLD probability predicted by random forest trained using HoVer-Net–derived features, according to the PDX image (*x*-axis), from an LUSC (red) or XTLD (blue) sample in the MDACC (solid line) or PDMR (dashed line) dataset. Results shown for held-out set following fivefold cross-validation, with folds defined according to the patient/progenitor (see Materials and Methods). **E,** Confusion matrix showing tile-level concordance between HoVer-Net–based classification (rows) and annotation (columns). **F** and **G,** Inception-based classifier performance, as in **D** and **E,** respectively, but with predictions from a random forest trained using Inception features. **H,** Distribution of tile-level XTLD probability predicted by random forest trained using Inception features, according to the PDX image (*x*-axis), from LUSC (red) or XTLD (blue) samples in JAX external validation dataset. Each image is from a distinct human progenitor. Samples are ordered by median XTLD prediction probability over tiles in **D**, **F**, and **H**. All PDXs are NSG mice.

was consistently high for all regions across cancer types, datasets, and individual samples. In particular, the median probability of a tile being assigned to the correct region was >50% for each of the four samples in the lung PDMR dataset. This was also the case for 8 out of 10 lung stromal regions and 11 of 12 lung tumor regions in the MDACC dataset, as well as for 6 of 7 breast tumor regions and 4

of 5 breast necrosis regions in the BCM dataset (Supplementary Fig. S9A–S9C). Significantly, the classifier achieved this strong, pan-tissue classification performance despite the experimental variability observed in the H&E images (Supplementary Fig. S7A and S7B).

An H&E-based classifier identifies XTLD

During development of PDX models in immunocompromised hosts, such as NSG mice, proliferation of atypical human lymphocytes at the implantation site can overtake or limit the growth of the human tumor cells. Such XTLDs (predominantly B-cell type lymphomas) are frequent, occurring at rates that exceed 10% in some transplanted cancer types (38), with variability across cancer types (39). Cases of XTLD masquerading as a bona fide xenograft can waste significant effort in xenograft treatment studies. XTLD can be detected during a QC assessment of the generated model using IHC directed at human pan-immune (CD45) or B-cell (CD20) markers. It has also been detected and predicted based on expression differences between Epstein–Barr virus–associated lymphoma samples and nonlymphoma samples (40). Nevertheless, for laboratories that do not routinely perform IHC during a QC step and for legacy models without such evaluations, XTLD remains a concern.

We therefore sought to develop an XTLD classifier applicable to H&E images generated during a standard pathologic assessment of NSG PDX models. We reasoned that the higher ratio of epithelial tumor cells to small lymphocytes in LUSC (**Fig. 4A**) relative to XTLD (**Fig. 4B**) would be detectable via cell segmentation and classification. Indeed, we observed that the proportion of segmented immune cells was elevated in XTLD samples relative to LUSC samples in the MDACC ($P = 4.6 \times 10^{-4}$) and PDMR ($P = 0.08$) datasets (Supplementary Fig. S10A). Furthermore, the median cell cross-sectional area of LUSC samples was higher than XTLD samples in MDACC and for three of five LUSC samples in PDMR (Supplementary Fig. S10B).

Despite the associations between XTLD and both immune cell proportion and cell size, the orderings induced by these two metrics imperfectly segregate XTLD and LUSC samples. Hence, we attempted a more general classification at the image tile level by representing each tile according to its cell classification–derived proportions of neoplastic, stromal, necrotic, and immune cells, i.e., excluding the count-based features used in the regional classifier. We used these tile-level representations and fivefold cross-validation to train a random forest to predict XTLD status, similar to our approach above in predicting region labels (**Fig. 4C–E**). Tile-level prediction performance was strong (**Fig. 4E**), with the distribution of tile-level XTLD probabilities skewed toward one for tiles from XTLD images and toward zero for those from LUSC images (Supplementary Fig. S11A) and with a tile-level AUC of 0.87 (Supplementary Fig. S11B). We defined a sample-level prediction probability as the median prediction probability over tiles in the sample. The resulting predictions were correct (i.e., >0.5) for 29 of 33 samples (**Fig. 4D**) and sample-level AUCs (Supplementary Fig. S11C), precision, recall, specificity, and accuracy (Supplementary Fig. S11D) were between 0.82 and 0.94. As expected, immune proportions were more important than neoplastic and necrotic cell proportions and of similar importance to stromal cell proportions in predicting XTLD (Supplementary Fig. S12).

To improve upon sample-level prediction, we next sought to generalize our classifier beyond its four biologically inspired, but constrained, feature inputs. For this purpose, we extracted 2,048 DL features from each tile using an Inception v3 network pretrained on ImageNet (14, 41). These features were then input to a random forest classifier and the classifier was trained using fivefold cross-validation and the same assignment of tiles to folds as employed for the HoVer-Net–based classifier. The Inception-based classifier improved performance over the HoVer-Net–based model (**Fig. 4F** and **G**). At the tile-level, it had a more pronounced skew of predicted probabilities toward their correct assignments (Supplementary Fig. S11E) and a higher tile-level AUC (0.96; Supplementary Fig. S11F) than the HoVer-Net–based classifier. At the sample-level, the Inception-based classifier showed improved AUC (>0.99; Supplementary Fig. S11G), precision, recall, specificity, and accuracy (0.94–1.00; Supplementary Fig. S11H). In aggregate, it correctly predicted 32 of 33 samples (**Fig. 4F**). In particular, the only sample misclassified by the Inception-based model (99354) was also misclassified by the HoVer-Net–based model, whereas the three additional samples misclassified by the HoVer-Net–based model (99772, 97322, and TC 916 P0M1) were correctly classified by the Inception-based model. Additionally, the Inception-based model perfectly classified six samples in an external validation dataset (**Fig. 4H**).

To interpret the information encoded by Inception that drives classification performance (33), we examined feature 1895, the feature with highest variable importance in the random forest model (Supplementary Fig. S13A and S13B). We observed that this single feature segregated tiles in XTLD cases (high feature value) from those in LUSC cases (low feature value; Supplementary Fig. S13C). A montage of tiles suggest that feature 1895 encodes cellular density and/or cell size (or something correlated with these)—those with high feature 1895 value (**Fig. 5A**) are dense in (small, round) lymphocytes or necrotic regions, whereas those with low value (**Fig. 5B**) have fewer, large cells and admixed stroma.

A pretrained H&E-based classifier identifies MSI in human progenitor and PDX model samples

To assess the applicability of our repository as an external validation set for evaluating published models, we applied a suite of pretrained MSI predictors (34) to colon adenocarcinoma human progenitor (**Fig. 6A**; Supplementary Fig. S14) and PDX (**Fig. 6B**; Supplementary Fig. S15) samples. The authors collected nine datasets of H&E images and trained nine different CNN (ResNet)-based models by excluding one of the datasets. We calculated the mean predicted MSI probability over these nine models and then ordered the samples based on the median value over tiles of their mean predicted MSI probabilities. MSS samples generally have lower median values than MSI samples in both human progenitor (**Fig. 6A**; sample-level AUC = 1; enrichment $P = 0.13$) and PDX (**Fig. 6B**; sample-level AUC = 0.93; enrichment $P = 2.00 \times 10^{-3}$) samples. Although MSI and MSS samples are interleaved near the high end of the predicted value spectrum, the concentration of MSS samples at the low end is consistent with the authors' intent that their models can be used as prescreening tools to rule out MSI. Furthermore, we find concordance between human progenitor and PDX sample predictions. For example, the two human progenitor samples with high MSS predictions (i.e., with lowest median MSI predictions), 944381 and 782815, have corresponding PDX samples within the quarter of predicted most stable PDX samples. Conversely, the highest predicted MSI human progenitor, 947758, has a corresponding PDX sample amongst the quarter predicted highest MSI PDX samples. Nevertheless, the impact of dataset heterogeneity is reflected in the divergent predictions across models that were trained on different, but overlapping, subsets of data (Supplementary Figs. S14 and S15).

Figure 5.
Inception encodes morphologic features useful in predicting XTLD. **A,** Tiles with large values of Inception feature 1895. The tile with the largest value is shown for each of four samples (noted above tile) with highest values, all of which are XTLD cases. **B,** Tiles with small values of Inception feature 1895. The tile with the smallest value is shown for each of four samples (noted above tile) with smallest values, all of which are LUSC cases. All PDXs are NSG mice. Images at 20× resolution. Scale bars, 50 μm.

A   Tiles with largest values of Inception feature 1895

LUSC1-59168   PDX-034   LUSC9-53550   PDX-060

B   Tiles with small values of Inception feature 1895

LUSC21-59180   14026   PDX-044   PDX-037
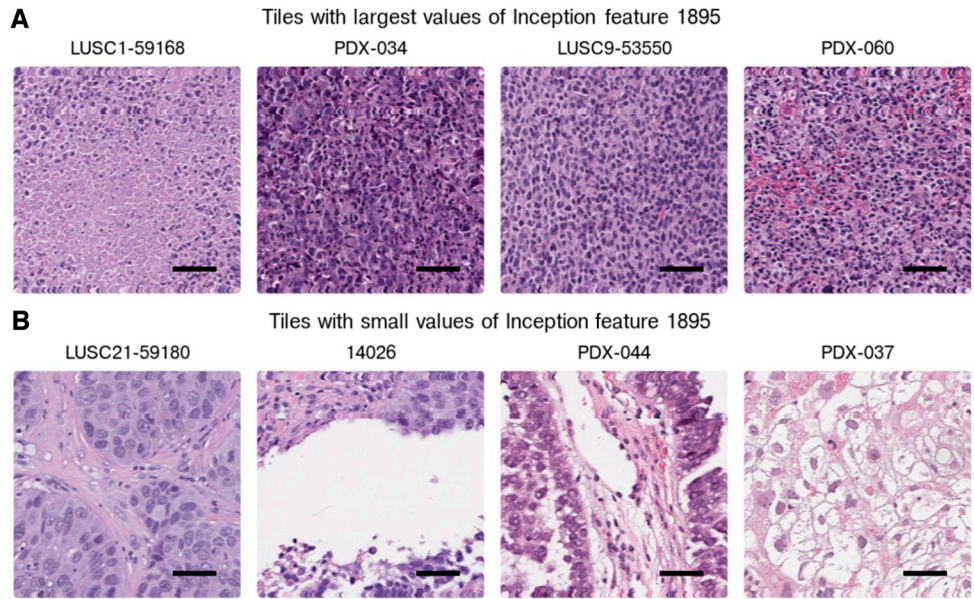
Discussion

We presented a repository of >1,000 H&E images of PDX samples and their human progenitors. All images are clinically annotated, many have paired expression and genomic data, and several have associated neoplastic, stromal, and necrotic regions manually segmented by a board-certified pathologist. To the best of our knowledge, this is the first, large-scale, publicly accessible set of PDX-derived images. As a demonstration of the utility of this resource, we provided three use cases: (i) development of a tumor-associated region classifier; (ii) development of a classifier of XTLD—the unintended outgrowth of human lymphocytes, rather than the epithelial tumor cells, at the xenograft implantation site; and (iii) application of a published, pretrained MSI classifier (34).
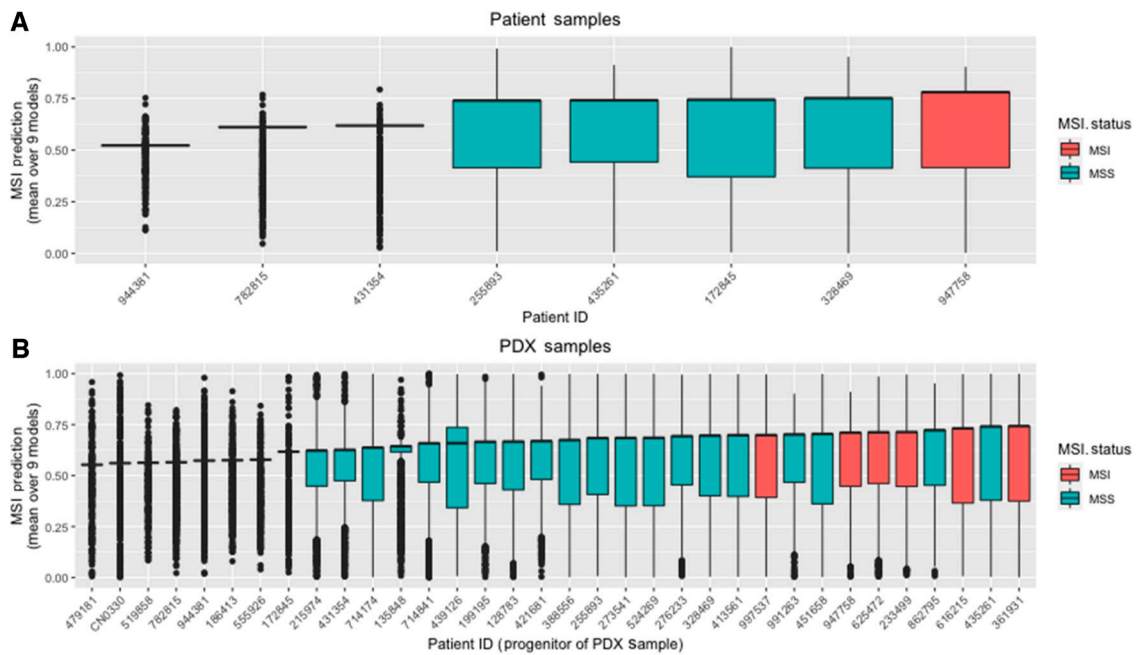
Figure 6.
MSI can be computationally predicted from PDX samples using a human-trained model. Distribution of tile-level MSI probability predicted using a published model (34) from human progenitor (**A**) or PDX (**B**) model H&E images derived from MSI (red) or MSS (blue) samples. All progenitor samples in **A** have corresponding derived PDX samples in **B**. Samples ordered by median (mean over models) MSI prediction over tiles.

The repository will enable others to explore key biological questions, many specifically relevant to or suited to analysis in PDXs: (i) What morphologic features of human tumors are recapitulated in PDXs? (ii) Stromal cells and cancer-associated fibroblasts, in particular, have been implicated in disease largely through their highly varied mediation of tumor/immune cell interactions (42). Do they retain this or some other function in immunocompromised mice? How are answers to these impacted by passaging, with early passages still harboring human immune and stromal cells and with murine cancer-associated fibroblasts replacing their human counterpart over time and passages (28)? Addressing these would be facilitated by placing immune and stromal cells in the context of the tumor core, periphery, or exterior using our regional classifier. (iii) Automated grading of a variety of epithelial cancers, such as breast (43) and colon (44) can be achieved through DL-based analysis of H&E images. Is tumor grade similarly imprinted on PDX sample morphology? If so, it should be possible to use the H&E images in our repository to predict their associated tumor stage and/or grade annotations. (iv) Chemotherapy-treated breast samples exhibit morphologic differences relative to treatment-naïve samples, including enlarged tumor cell nuclei (45). Are such differences observed across tumor types and in PDX models? (v) Genetic mutations (11) and gene expression (12) have been predicted from H&E using DL approaches. The (bulk) WES and RNA-seq data paired with histology images in our repository would provide labels for transcriptome-wide studies in PDX samples. What gene mutations and expression patterns can be predicted? Do they validate published results (12)? Are the associated, predictive morphologic signatures preserved across tumor types? (vi) We have previously observed mouse strain–specific tumor growth rates, which were characterized by different cellular densities and phenotypic composition using HoVer-Net (46). Specifically, the smallest tumors showed high contrast in stromal density at the tumor periphery versus the tumor central region, whereas the largest tumors were observed to show higher degree of stromal infiltration. Do related cellular or morphologic features correlate with response to drug treatment, as reflected in the repository's tumor volume data?

H&E image color variability observed across institutions and laboratories (47) may be caused by differences between stain batches and manufacturers, staining and fixation protocols, and tissue thickness (48), as well as differences across imaging parameters, including scanner models and image magnification. These site-specific effects adversely impact downstream DL-based classification (49) and confound prediction of survival, mutations, and tumor stage (50). The resulting lack of generalizability is a significant barrier to clinical deployment of these approaches (51). Several classes of color normalization attempt to address this issue: (i) template color matching maps summary statistics or histograms of RGB values in a reference image to those in a reference image (52); (ii) color deconvolution represents each of the H&E dyes as a "stain vector" and substitutes a reference stain vector for the corresponding target vector (35); and (iii) generative adversarial network approaches transfer strain distributions from a reference dataset (rather than single image) to a target dataset (53). Our repository has considerable variability across the dimensions impacting image color, with six sites contributing data generated at two resolutions (20× and 40×) from different scanner models across multiple years. As such, it offers ample opportunity to compare color normalization approaches that may be combined with phenotypic and treatment response data for PDXs.

Our region and XTLD classifiers were developed primarily using fivefold cross-validation. As such, they may be limited in their generalizability across datasets, including relative to the stain variability mentioned above. We have partially mitigated this concern through a small, external validation of the XTLD classifier. Regardless, these proof-of-principle classifiers demonstrate the methodologic domain of our repository. Furthermore, our application of pretrained MSI models demonstrates its relevance for external validation of published and third-party tools. Indeed, we showed the impact of dataset heterogeneity on these models, in which results were sensitive to their training datasets despite the fact that any two of these datasets included seven of eight cohorts in common.

The region and XTLD classifiers were trained on features engineered from cells segmented and classified by one particular algorithm—HoVer-Net. As a demonstration of the general applicability of our annotated images, we applied a second approach, HD-Staining, to a lung sample (31). Intriguingly, we observed significant differences between HoVer-Net– and HD-Staining–predicted cell types within stromal and necrotic regions annotated by our pathologist. These region-level annotations make possible a more systematic comparison of HoVer-Net, HD-Staining, and related methods across other tumor types. This remains for future work.

The H&E images presented here, alongside their genomic, transcriptomic, clinical, and pathologic data, complement existing large-scale human repositories, including TCGA, TCIA, GTEx, and CAMELYON17 (10) while uniquely contributing PDX samples. As such, they expand the heterogeneity of publicly accessible histology data available for technical studies (e.g., color normalization and batch correction) while also enabling exploration of the phenotypic effect of cellular interactions and morphology within a model system specifically intended to capture them.

Authors' Disclosures

Immunomedics, Inflection Biosciences, Karyopharm Therapeutics, Loxo Oncology, Mersana Therapeutics, OnCusp Therapeutics, Puma Biotechnology Inc., Sanofi, Silverback Therapeutics, Spectrum Pharmaceuticals, Theratechnologies, Zentalis and Dava Oncology; grants from Aileron Therapeutics, Inc., AstraZeneca, Bayer Healthcare Pharmaceuticals, Calithera Biosciences Inc., Curis, Inc., CytomX Therapeutics Inc., Daiichi Sankyo Co., Ltd., Debiopharm International, eFFEC-TOR Therapeutics, Genentech, Inc., Guardant Health, Inc., KLUS Pharma, Takeda Pharmaceuticals, Novartis, Puma Biotechnology, Inc., and Taiho Pharmaceutical Co.; and other support from European Organisation for Research and Treatment of Cancer, European Society for Medical Oncology, Cholangiocarcinoma Foundation, and Dava Oncology outside the submitted work. Y. Xie reports grants from NIH and Cancer Prevention and Research Institute of Texas (CPRIT) during the conduct of the study; grants from NIH and CPRIT outside the submitted work; and being a cofounder of the Adjuvant Genomics, Inc. M.T. Lewis reports grants from the NCI during the conduct of the study, as well as being a founder and limited partner in StemMed Ltd., founder, manager, and general partner in StemMed Holdings LLC, and founder and equity holder in Tvardi Therapeutics Inc. No disclosures were reported by the other authors.

Authors' Contributions

B.S. White: Conceptualization, data curation, software, investigation, methodology, writing–original draft, writing–review and editing. **X. Woo:** Conceptualization, data curation, software, investigation, methodology, writing–original draft, writing–review and editing. **S. Koc:** Data curation, software, methodology, writing–review and editing. **T. Sheridan:** Validation, investigation, writing–review and editing. **S.B. Neuhauser:** Resources, data curation. **S. Wang:** Software, investigation, methodology, writing–original draft, writing–review and editing. **Y.A. Evrard:** Conceptualization, resources, writing–original draft, writing–review and editing. **L. Chen:** Resources, writing–original draft, writing–review and editing. **A. Foroughi pour:** Conceptualization, investigation, writing–original draft, writing–review and editing. **J.D. Landua:** Resources, Data curation. **R.J. Mashl:** Resources, data curation. **S.R. Davies:** Resources, data curation. **B. Fang:** Resources, data curation. **M. Raso:** Resources, data curation. **K.W. Evans:** Resources, data curation. **M.H. Bailey:** Resources, data curation. **Y. Chen:** Resources, data curation. **M. Xiao:** Resources, data curation. **J. Rubinstein:** Conceptualization. **B.J. Sanderson:** Resources, data curation. **M.W. Lloyd:** Data curation. **S. Domanskyi:** Data curation. **L.E. Dobrolecki:** Resources, data curation. **M. Fujita:** Resources, data curation. **J. Fujimoto:** Resources, data curation. **G. Xiao:** Resources, data curation. **R.C. Fields:** Resources, data curation. **J.L. Mudd:** Resources, data curation. **X. Xu:** Resources, data curation. **M.G. Hollingshead:** Resources. **S. Jiwani:** Resources, data curation. **S. Acevedo:** Data curation, software. **PDXNet Consortium:** Resources, funding acquisition. **B.N. Davis-Dusenbery:** Funding acquisition. **P.N. Robinson:** Conceptualization, funding acquisition, project administration. **J.A. Moscow:** Conceptualization, project administration. **J.H. Doroshow:** Conceptualization, resources, funding acquisition. **N. Mitsiades:** Resources, funding acquisition. **S. Kaochar:** Funding acquisition. **C.-x. Pan:** Funding acquisition. **L.G. Carvajal-Carmona:** Resources, funding acquisition. **A.L. Welm:** Resources, funding acquisition. **B.E. Welm:** Resources, funding acquisition. **R. Govindan:** Funding acquisition. **S. Li:** Funding acquisition. **M.A. Davies:** Conceptualization, funding acquisition. **J.A. Roth:** Conceptualization, funding acquisition, writing–review and editing. **F. Meric-Bernstam:** Conceptualization, supervision, funding acquisition, writing–review and editing. **Y. Xie:** Conceptualization, resources, supervision, funding acquisition. **M. Herlyn:** Conceptualization, resources, funding acquisition. **L. Ding:** Conceptualization, resources, funding acquisition, writing–original draft, writing–review and editing. **M.T. Lewis:** Conceptualization, resources, funding acquisition, writing–original draft, writing–review and editing. **C.J. Bult:** Conceptualization, resources, supervision, funding acquisition. **D.A. Dean:** Conceptualization, supervision, funding acquisition, writing–original draft, writing–review and editing. **J.H. Chuang:** Conceptualization, resources, supervision, funding acquisition, writing–original draft, writing–review and editing.

Acknowledgments

Note

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

References

1. Tentler JJ, Tan AC, Weekes CD, Jimeno A, Leong S, Pitts TM, et al. Patient-derived tumour xenografts as models for oncology drug development. Nat Rev Clin Oncol 2012;9:338–50.
2. Pompili L, Porru M, Caruso C, Biroccio A, Leonetti C. Patient-derived xenografts: a relevant preclinical model for drug development. J Exp Clin Cancer Res 2016;35:189.
3. DeRose YS, Wang G, Lin Y-C, Bernard PS, Buys SS, Ebbert MTW, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. Nat Med 2011;17:1514–20.
4. Petrillo LA, Wolf DM, Kapoun AM, Wang NJ, Barczak A, Xiao Y, et al. Xenografts faithfully recapitulate breast cancer-specific gene expression patterns of parent primary breast tumors. Breast Cancer Res Treat 2012;135:913–22.
5. Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, et al. Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. Cell Rep 2013;4:1116–30.
6. Zhang X, Claerhout S, Prat A, Dobrolecki LE, Petrovic I, Lai Q, et al. A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. Cancer Res 2013;73:4885–97.
7. Woo XY, Giordano J, Srivastava A, Zhao Z-M, Lloyd MW, de Bruijn R, et al. Conservation of copy number profiles during engraftment and passaging of patient-derived cancer xenografts. Nat Genet 2021;53:86–99.
8. Guillen KP, Fujita M, Butterfield AJ, Scherer SD, Bailey MH, Chu Z, et al. A human breast cancer-derived xenograft and organoid platform for drug discovery and precision oncology. Nat Cancer 2022;3:232–50.
9. Srinidhi CL, Ciga O, Martel AL. Deep neural network models for computational histopathology: a survey. Med Image Anal 2021;67:101813.
10. Bejnordi BE, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017;318:2199–210.
11. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018;24:1559–67.
12. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun 2020;11:3877.
13. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med 2019;16:e1002730.
14. Noorbakhsh J, Farahmand S, Foroughi Pour A, Namburi S, Caruana D, Rimm D, et al. Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. Nat Commun 2020;11:6367.
15. Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut 2021;70:544–54.
16. Rawat RR, Ortega I, Roy P, Sha F, Shibata D, Ruderman D, et al. Deep learned tissue "fingerprints" classify breast cancers by ER/PR/Her2 status from H&E images. Sci Rep 2020;10:7275.
17. Li J, Mi W, Guo Y, Ren X, Fu H, Zhang T, et al. Artificial intelligence for histological subtype classification of breast cancer: combining multi-scale

feature maps and the recurrent attention model. Histopathology 2022;80: 836–46.

18. Farahmand S, Fernandez AI, Ahmed FS, Rimm DL, Chuang JH, Reisenbichler E, et al. Deep learning trained on hematoxylin and eosin tumor region of Interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. Mod Pathol 2022;35:44–51.

19. Johannet P, Coudray N, Donnelly DM, Jour G, Illa-Bochaca I, Xia Y, et al. Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma. Clin Cancer Res 2021;27:131–40.

20. Chan H-P, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. Adv Exp Med Biol 2020;1213:3–21.

21. Burlingame EA, McDonnell M, Schau GF, Thibault G, Lanciault C, Morgan T, et al. SHIFT: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning. Sci Rep 2020;10:17507.

22. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. Med Image Anal 2019;58:101552.

23. Wagner SJ, Reisenbüchler D, West NP, Niehues JM, Zhu J, Foersch S, et al. Transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study. Cancer Cell 2023;41:1650–61.e4.

24. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. Cell Rep 2018;23:181–93.e7.

25. Koc S, Lloyd MW, Grover JW, Xiao N, Seepo S, Subramanian SL, et al. PDXNet portal: patient-derived Xenograft model, data, workflow and tool discovery. NAR Cancer 2022;4:zcac014.

26. Lau JW, Lehnert E, Sethi A, Malhotra R, Kaushik G, Onder Z, et al. The cancer genomics cloud: collaborative, reproducible, and democratized—a new paradigm in large-scale computational Research. Cancer Res 2017;77:e3–6.

27. Ben-David U, Ha G, Tseng Y-Y, Greenwald NF, Oh C, Shih J, et al. Patient-derived xenografts undergo mouse-specific tumor evolution. Nat Genet 2017; 49:1567–75.

28. Blomme A, Van Simaeys G, Doumont G, Costanza B, Bellier J, Otaka Y, et al. Murine stroma adopts a human-like metabolic phenotype in the PDX model of colorectal cancer and liver metastases. Oncogene 2018;37:1237–50.

29. Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: an open-source quality control tool for digital pathology slides. JCO Clin Cancer Inform 2019;3:1–7.

30. Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, et al. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal 2019;58:101563.

31. Wang S, Rong R, Yang DM, Fujimoto J, Yan S, Cai L, et al. Computational staining of pathology images to study the tumor microenvironment in lung cancer. Cancer Res 2020;80:2056–66.

32. Wang S, Chen A, Yang L, Cai L, Xie Y, Fujimoto J, et al. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. Sci Rep 2018;8:10393.

33. Foroughi Pour A, White BS, Park J, Sheridan TB, Chuang JH. Deep learning features encode interpretable morphologies within histological images. Sci Rep 2022;12:9428.

34. Echle A, Ghaffari Laleh N, Quirke P, Grabsch HI, Muti HS, Saldanha OL, et al. Artificial intelligence for detection of microsatellite instability in colorectal cancer—a multicentric analysis of a pre-screening tool for clinical application. ESMO Open 2022;7:100400.

35. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2009 Jun 28–Jul 01; Boston, MA [Internet]. IEEE; 2009. Available from: http://ieeexplore.ieee.org/document/5193250/.

36. Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer 2020;1:800–10.

37. Plava J, Cihova M, Burikova M, Matuskova M, Kucerova L, Miklikova S. Recent advances in understanding tumor stroma-mediated chemoresistance in breast cancer. Mol Cancer 2019;18:67.

38. Butler KA, Hou X, Becker MA, Zanfagnin V, Enderica-Gonzalez S, Visscher D, et al. Prevention of human lymphoproliferative tumor formation in ovarian cancer patient-derived xenografts. Neoplasia 2017;19:628–36.

39. Zhang L, Liu Y, Wang X, Tang Z, Li S, Hu Y, et al. The extent of inflammatory infiltration in primary cancer tissues is associated with lymphomagenesis in immunodeficient mice. Sci Rep 2015;5:9447.

40. Woo XY, Srivastava A, Graber JH, Yadav V, Sarsani VK, Simons A, et al. Genomic data analysis workflows for tumors from patient-derived xenografts (PDXs): challenges and guidelines. BMC Med Genomics 2019;12:92.

41. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV [Internet]. IEEE; 2016. Available from: http://ieeexplore.ieee.org/document/7780677/.

42. Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer 2021;20:131.

43. Bejnordi BE, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. Mod Pathol 2018;31:1502–12.

44. Vuong TLT, Lee D, Kwak JT, Kim K. Multi-task deep learning for colon cancer grading. 2020 International Conference on Electronics, Information, and Communication (ICEIC); 2020 Jan 19–22; Barcelona, Spain [Internet]. IEEE; 2020. Available from: https://ieeexplore.ieee.org/document/9051305/.

45. Rasbridge SA, Gillett CE, Seymour AM, Patel K, Richards MA, Rubens RD, et al. The effects of chemotherapy on morphology, cellular proliferation, apoptosis and oncoprotein expression in primary breast carcinoma. Br J Cancer 1994;70:335–41.

46. Sargent JK, Warner MA, Low BE, Schott WH, Hoffert T, Coleman D, et al. Genetically diverse mouse platform to xenograft cancer cells. Dis Model Mech 2022;15:dmm049457.

47. Nam S, Chong Y, Jung CK, Kwak T-Y, Lee JY, Park J, et al. Introduction to digital pathology and computer-aided pathology. J Pathol Transl Med 2020;54: 125–34.

48. Veta M, Pluim JPW, van Diest PJ, Viergever MA. Breast cancer histopathology image analysis: a review. IEEE Trans Biomed Eng 2014;61:1400–11.

49. Salvi M, Michielli N, Molinari F. Stain Color Adaptive Normalization (SCAN) algorithm: separation and standardization of histological stains in digital pathology. Comput Methods Programs Biomed 2020;193:105506.

50. Howard FM, Dolezal J, Kochanny S, Schulte J, Chen H, Heij L, et al. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. Nat Commun 2021;12:4423.

51. van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. Nat Med 2021;27:775–84.

52. Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Comput Graph Appl 2001;21:34–41.

53. Bentaieb A, Hamarneh G. Adversarial stain transfer for histopathology image analysis. IEEE Trans Med Imaging 2018;37:792–802.