



# Development of a Histopathology Informatics Pipeline for Classification and Prediction of Clinical Outcomes in Subtypes of Renal Cell Carcinoma

Eliana Marostica<sup>1</sup>, Rebecca Barber<sup>1,2</sup>, Thomas Denize<sup>3</sup>, Isaac S. Kohane<sup>1</sup>, Sabina Signoretti<sup>3</sup>, Jeffrey A. Golden<sup>3,4</sup>, and Kun-Hsing Yu<sup>1,3</sup>

## ABSTRACT

**Purpose:** Histopathology evaluation is the gold standard for diagnosing clear cell (ccRCC), papillary, and chromophobe renal cell carcinoma (RCC). However, interrater variability has been reported, and the whole-slide histopathology images likely contain underutilized biological signals predictive of genomic profiles.

**Experimental Design:** To address this knowledge gap, we obtained whole-slide histopathology images and demographic, genomic, and clinical data from The Cancer Genome Atlas, the Clinical Proteomic Tumor Analysis Consortium, and Brigham and Women's Hospital (Boston, MA) to develop computational methods for integrating data analyses. Leveraging these large and diverse datasets, we developed fully automated convolutional neural networks to diagnose renal cancers and connect quantitative pathology patterns with patients' genomic profiles and prognoses.

**Results:** Our deep convolutional neural networks successfully detected malignancy (AUC in the independent validation cohort: 0.964–0.985), diagnosed RCC histologic subtypes (independent validation AUCs of the best models: 0.953–0.993), and predicted stage I ccRCC patients' survival outcomes (log-rank test  $P = 0.02$ ). Our machine learning approaches further identified histopathology image features indicative of copy-number alterations (AUC > 0.7 in multiple genes in patients with ccRCC) and tumor mutation burden.

**Conclusions:** Our results suggest that convolutional neural networks can extract histologic signals predictive of patients' diagnoses, prognoses, and genomic variations of clinical importance. Our approaches can systematically identify previously unknown relations among diverse data modalities.

## Introduction

Renal cell carcinoma (RCC) occurs in approximately 4.4 per 100,000 individuals globally (1). The International Agency for Research on Cancer reported 403,262 new cases of kidney cancer (2.2% of all cancers) and 175,098 deaths attributed to kidney cancer (1.8% of deaths attributed to any cancer) in 2018 (2). The three most common histologic subtypes of RCC are clear cell (ccRCC), papillary (pRCC), and chromophobe (chRCC; ref. 3). Histopathologic evaluation remains the state-of-the-art method for diagnosing and subtyping RCC, and therapeutic strategies tailored to the histologic and genetic subtypes of RCC are required to achieve optimal treatment outcomes (4–6). However, current reporting of RCC pathology is often incomplete and interrater disagreement has been reported previously (7). As an illustration, a previous study shows that the interrater agreement for RCC diagnosis is moderate ( $\kappa = 0.55$ ) even among

experienced pathologists (8). In addition, billions of pixels in a typical whole-slide histopathology image of RCC likely contain a plethora of biological signals not utilized by manual pathology evaluation.

Computer vision has emerged as one of the most promising fields for enhancing medicine with artificial intelligence methods (9, 10). Recent studies showed that deep learning algorithms are able to diagnose histologic classification with accuracies comparable with expert pathologists (11–15). Not only do these approaches show potential for improving diagnosis and prognosis for patients, but they may also help to reduce monotonous tasks for pathologists and free up their time for more complicated tasks (16). Previous studies also demonstrated that machine learning methods can predict patient prognosis and molecular subtypes of patients with non-small cell lung cancer based on hematoxylin and eosin (H&E)-stained images (14), indicating the potential of computer vision methods in extracting the previously overlooked signals in high-resolution histopathology images.

In this study, we developed informatics pipelines to connect RCC histopathology images with genomic information, clinical profiles, and biomarkers of response to immune checkpoint blockade. Specifically, we (i) developed an automated weak supervision framework to process whole-slide histopathology images and diagnose the subtypes of RCC, with the results validated in two independent cohorts; (ii) devised a machine learning-based prognostic model to predict the survival outcomes of patients with RCC using their histopathology images; and (iii) linked histopathology images with genomic data to uncover previously unknown molecular-morphologic connections, with a focus on genetic aberrations with known clinical significance. Our informatics framework successfully predicted individual patients' tumor mutation burden and prognoses, which could inform treatment decision making. The methods we developed are extensible to studying other tumor types and other types of multi-omics aberrations.

<sup>1</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts. <sup>2</sup>Department of Computer Science, Princeton University, Princeton, New Jersey. <sup>3</sup>Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts. <sup>4</sup>Cedars-Sinai Medical Center, Los Angeles, California.

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

**Corresponding Author:** Kun-Hsing Yu, Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Fourth Floor, Boston, MA 02115. Phone: 617-432-2144; Fax: 617-432-0693; E-mail: [kun-hsing\\_yu@hms.harvard.edu](mailto:kun-hsing_yu@hms.harvard.edu)

Clin Cancer Res 2021;27:2868–78

doi: 10.1158/1078-0432.CCR-20-4119

©2021 American Association for Cancer Research.

### Translational Relevance

Renal cell carcinoma causes more than 175,000 deaths per year worldwide, and histopathology assessment is critical for diagnosing this deadly cancer. However, visual histopathology evaluation suffers from interrater variability and cannot identify molecular aberrations or prognoses of the patients. In this study, we developed fully automated methods to characterize the histologic subtypes, genomic variations, and tumor mutation burden objectively using whole-slide histopathology images. We showed that our approaches successfully detected malignancy and identified histologic subtypes, with the results validated in two independent cohorts (AUC > 0.95). We further predicted copy-number alterations, tumor mutation burden, and patients' survival outcomes. Our computational approaches demonstrated the linkages between histopathology, genomic profiles, and patient prognosis, and our prediction models can enable personalized treatment selection.

## Materials and Methods

### Study design

This study was conducted to determine the capabilities of deep convolutional neural networks (DCNN) to detect diagnostic, prognostic, and biological signals of RCC tumors from histopathology slide images using supervised transfer learning. Whole-slide histopathology images (Tissue Slides) and clinical data for the three RCC subtypes (chRCC, ccRCC, and pRCC) were acquired from The Cancer Genome Atlas (TCGA; refs. 17–20, RRID:SCR\_003193). There were 231 slides from 103 patients with chRCC, 1,657 slides from 537 patients with ccRCC, and 475 slides from 288 patients with pRCC. The pathology diagnosis of each sample was obtained on the basis of the consensus of a panel of 5 to 11 renal pathologists (17–20). Several patients contributed both slides with normal tissue and slides with malignant tissue. In prognostic and genomic prediction tasks, only slides with malignant tissue were included. The first external validation dataset for ccRCC was obtained from the NCI Clinical Proteomic Tumor Analysis Consortium (CPTAC; ref. 21). This cohort consists of 782 ccRCC slides from 222 patients of both normal and malignant tissue. An additional independent dataset of 131 patients (41 pRCC, 59 ccRCC, and 31 chRCC) was collected from the Brigham and Women's Hospital Department of Pathology to further validate the generalizability of our subtype prediction algorithms. A Hamamatsu NanoZoomer S210 digital slide scanner was used to digitize the slides from the Brigham and Women's Hospital (Boston, MA). Written informed consent from the participants was obtained. This study is conducted in accordance with the Declaration of Helsinki and is performed after approval by the Harvard Medical School Institutional Review Board (IRB number: IRB20-0957).

Whole-slide pathology images, obtained at 20× magnification or higher, were processed into 1,000 × 1,000 pixel patches. Image patches from TCGA were randomly split into training, validation, and test sets to build and optimize machine learning models as well as objectively evaluate the model performance. To ensure objective evaluations of the models and prevent information leakage, all image patches belonging to a single patient were in only one of the training, validation, or testing sets. Approximately 60% of all patches were in the training set, 20% were in the validation set, and 20% were in the test set. The training and validation sets were used for training and hyperparameter optimization, respectively. Autonomous hyperparameter optimization was

performed using the Talos Python package, version 0.6.3 (Python; ref. 22; RRID:SCR\_008394). The test set was left untouched until the evaluation of the finalized DCNN models. In malignancy detection, the generalizability of the DCNN was further examined in the untouched CPTAC external validation cohort. The generalizability of the subtype prediction platform was validated by applying the same machine learning algorithm to the Brigham and Women's Hospital dataset and evaluated the results in the test set. E. Marostica maintained the blinded data and oversaw the evaluation of these tasks.

### Identification of images with malignant cells

To demonstrate the feasibility of a fully automated histopathology analytic pipeline, the first task we investigated was the identification of image patches with malignant cells. A weak supervision approach was taken in developing models for malignant region identification. In this approach, the slides characterized by an absence of malignant cells were labeled as benign. Slides adjacent to the cancerous tissue selected for multi-omics profiling by TCGA were labeled as malignant, but they may contain regions of benign cells. These machine learning models systematically compared the regions of these slides to identify signals indicative of malignancy. This approach can accommodate a large number of slides without detailed pixel-level or region-level annotations, which better accommodate the typical pathology slides collected from the current clinical workflow. Models were trained separately on three RCC subtypes. Three different DCNN architectures, VGG-16 (ref. 23; RRID:SCR\_016494), Inception-v3 (24), and ResNet-50 (25), were compared in each task, and the models were initialized using the weights from ImageNet (26). The processed image patches serve as the input to the DCNNs, and the models were implemented using the Keras application programming interface, version 2.1.2 in Python (RRID:SCR\_008394). Binary cross-entropy was used as the loss function. Autonomous hyperparameter optimization was performed to select the optimal batch size (32 or 64), learning rate multiplier (range: 0.999 to 1e<sup>-4</sup>), number of epochs (10 or 15), and the optimization algorithm (Adam or RMSProp) from the prespecified ranges. Talos' learning rate normalizer was used in all hyperparameter optimization runs herein. Therefore, the effective learning rate is the product of the reported learning rate multiplier and the default learning rate of the optimizer. For each model, an error analysis was conducted by visual evaluation of 100 image regions misclassified by the models.

### Renal cancer subtype classification

The same transfer learning approach was applied to the multi-class classification task of diagnosing the histologic subtypes of RCC. The hyperparameters of DCNNs were optimized by Talos. All models were trained for 15 epochs using a batch size of 32 and the categorical cross-entropy loss function. We excluded tumor samples with MiTF translocation-associated carcinomas ( $n = 7$  in ccRCC,  $n = 8$  in pRCC, and none in chRCC TCGA set) from the subtype classification, because these tumors were subsequently proven to belong to different disease entities.

### Identification of the relations between histopathology images and somatic genomic variations

We employed the DCNN frameworks to systematically identify the relations between copy-number alterations (CNA) of the renal cancer samples ( $n = 66$  patients with chRCC;  $n = 528$  patients with ccRCC;  $n = 288$  patients with pRCC) and the corresponding histopathology images. Clinically important genes were selected by a literature review (4, 5), and CNA prediction was framed as a binary

task to predict the presence of any number of alterations at a specific gene using the histopathology images. The CNA status of each sample was obtained from cBioPortal (refs. 27, 28; RRID:SCR\_014555). To better focus on regions occupied by tumor cells, image patches predicted to be benign by the malignancy classification model were removed from this analysis. Because genetic aberrations are correlated with the histology of RCC, we developed independent models for ccRCC, pRCC, and chRCC. This stratified approach controlled for histology when examining the relations between genetic profiles and computational pathology features. Two transfer learning approaches were employed to develop prediction models for this set of tasks: (i) a binary classification task for each gene and (ii) a single multi-task classification for all genes, which predict the presence of CNAs in multiple genes simultaneously. A total of eight genes (*FH*, *FLCN*, *MET*, *SDHB*, *SDHD*, *TSC1*, *TSC2*, *VHL*) were selected for the binary classification task since they have been linked to the tumorigenesis and prognoses of RCC (29). Additional genes (*EGFR*, *KRAS*, *MYC*, *BCL2*, *AKT2*, *TP53*, *RB1*, *PTEN*, *NF1*, *NF2*, *WT1*) were selected for the multi-task classification task for their relation to RCC and relatively high prevalence rates (29). All hyperparameters were optimized via automated hyperparameter search: a batch size of 32 was used in combination with the binary cross-entropy loss function when training for 15 epochs with RMSprop in all multi-task classification models. Image augmentation included horizontal flips, vertical flips, and rotations in the multiple of 90 degrees. Image pixels were rescaled within the range [0–1]. In both transfer learning approaches for this task, the performance was reported in the independent test set.

DCNNs were also employed to identify the associations between somatic genetic mutations of the renal cancer samples ( $n = 65$  patients with chRCC;  $n = 402$  patients with ccRCC;  $n = 276$  patients with pRCC) and the corresponding histopathology images. Genes were selected for the analysis if the frequency of mutation was greater than or equal to 5% in the dataset, and the mutation status of each sample was obtained from cBioPortal (refs. 27, 28; RRID:SCR\_014555). As with the CNA prediction task, image patches predicted to be benign by the malignancy classification model were removed. Mutation prediction was performed using multi-task classification, in which the DCNN aimed to predict the presence of mutations in the selected gene panel. Image augmentation, which included horizontal flips, vertical flips, and rotations in the multiple of 90 degrees, and 10-fold cross-validation were employed.

### Overall survival prediction of patients with renal cancer

A supervised multi-task logistic regression approach (30) was employed to predict the length of the overall survival of patients with renal cancer. Patients' overall survival, obtained from TCGA through cBioPortal (refs. 27, 28; RRID:SCR\_014555), were binned such that a patient with an event received a one-hot encoded label vector with a 1 corresponding to the bin of overall survival in months. A patient with censored overall survival was labeled similarly, except that each bin after the time of censorship was also labeled as 1 (Supplementary Fig. S1). The image patches served as the input to the DCNN models, and the vector of binary values for each patient was the prediction target for the machine learning models. The cross-entropy loss and accuracy functions were updated so that a bin prediction larger than the reported overall survival of a censored patient would not be penalized (30). In this study, two bins were determined by the midpoint of the overall survival in the training set.

Because of the small sample sizes of patients in the chRCC, pRCC, and CPTAC ccRCC cohorts ( $n = 53$ ,  $n = 172$ ,  $n = 100$  for stage I patients, respectively; with only 2, 12, and 2 patients having mortality

in these cohorts), the survival prediction task was performed in the TCGA ccRCC cohort (269 patients total; 45 patients with events) and 10-fold cross-validation was employed. Upsampling of uncensored data points was performed in the training set of each fold to facilitate the model training process. Autonomous hyperparameter optimization was performed separately within each fold. Visualization of Kaplan–Meier curves was performed using R 3.5.1 (ref. 31; RRID: SCR\_001905) with the survival (32, 33) and survminer (34) packages. Gradient-weighted class activation mapping (grad-CAM) visualizations were generated to identify the regions of the greatest importance for survival prediction using the keras-vis Python package (35, 36).

### Prediction of tumor mutation burden using histopathology images

We further extended our transfer learning approaches to predict patients' tumor mutation burden using their histopathology images. The tumor mutation count for each patient was obtained from cBioPortal (refs. 27, 28; RRID:SCR\_014555). There were 44 patients with chRCC, 302 patients with ccRCC, and 187 patients with pRCC with available data; our analysis focuses on patients with ccRCC due to sample size considerations (median = 52, range: 8–591). Image patches predicted as normal by our malignancy detection model were removed prior to training. Supervised transfer learning with DCNN models was used for this regression task. Image augmentation was performed using vertical and horizontal flips, random rotations within 90 degrees, random zooms within the range 0.8–1.2, and ZCA whitening with the default epsilon of  $1e^{-6}$ . Image pixel values were rescaled between 0 and 1. Mean squared error was utilized as the loss function and early stopping that monitored this quantity was incorporated into the training and hyperparameter optimization process. The predictions of each image patch were subsequently aggregated to make the patient-level prediction. The performance in the independent test set was reported.

Activation of the output layer was visualized using the keras-vis Python package (36) to improve the interpretability of this DCNN. In this approach, we first performed activation maximization with an image patch as the seed input, and the output from this approach was converted to the “Jet” color mapping in the matplotlib package. The mean activation across the color channels was taken and overlaid with the original image patch. Since activation maximization generates large pixel values that indicate objects of interest in the images, this approach highlights regions that activate the decision neuron to a greater extent for the seeded image patches.

### Statistical methods

AUC and 95% confidence intervals (CI) with 2,000 bootstrap replicates were calculated using the pROC package (37) in R using default settings. In the multi-class classification tasks, the Hand and Till multi-class AUCs were calculated (37, 38). Differences between predicted longer-term and shorter-term survivors' Kaplan–Meier curves were compared using the log-rank test provided by the survival package (32, 33) in R ( $\alpha = 0.05$ ). Tumor mutation counts (median = 52, range: 8–591) were log transformed and subsequently normalized to values between 0 and 1 based on the observed values in the training set prior to training and hyperparameter optimization.

### Data and materials availability

TCGA imaging and clinical data are available through TCGA Genomic Data Commons portal (<https://portal.gdc.cancer.gov/>) and the CPTAC data are available at the CPTAC Data Portal (<https://cptac-data-portal.georgetown.edu/cptacPublic/>). Additional

TCGA genomic and clinical data are available on cBioPortal (<https://www.cbioportal.org/>, RRID:SCR\_014555; TCGA, RRID:SCR\_003193, PanCancer Atlas). Our trained models could be found at [https://github.com/hms-dbmi/rcc\\_pathology](https://github.com/hms-dbmi/rcc_pathology).

## Results

### Patient cohorts

We obtained a total of 2,363 whole-slide histopathology images from TCGA, which represent 537 patients with ccRCC, 288 patients with pRCC, and 103 patients with chRCC. We acquired an additional 782 whole-slide images of 222 patients with ccRCC from the CPTAC Clear Cell Renal Cell Carcinoma cohort as the first independent test set, and we further collected and digitized slides from 131 patients at Brigham and Women's Hospital to validate the generalizability of our models. Patient characteristics are shown in **Table 1**. These datasets serve as the foundation for our prediction tasks for cancer subtype diagnosis, prognosis, genetic variation, and tumor mutation burden.

### Transfer learning identified malignancy and diagnosed histologic subtypes

We first developed transfer learning approaches using three DCNN architectures (ResNet-50, VGG-16, and InceptionV3) to identify histopathology image regions containing malignant cells. We used extensive hyperparameter tuning to optimize the hyperparameters governing the model training process, and we evaluated our finalized models in the untouched test sets. For VGG-16, the optimal learning rate multiplier of 0.001 was used in all three RCC histologic subtypes. RMSprop was used in chRCC, while Adam was used in ccRCC and pRCC. Inception-v3 trained with a learning rate multiplier of 0.001, using Adam in chRCC and ccRCC and RMSprop in pRCC. The optimal learning rate multiplier in this task was 0.9 for ResNet-50,

using Adam in chRCC and RMSprop in ccRCC and pRCC. The performance of the finalized VGG-16, Inception-v3, and ResNet-50 models was compared based on the AUC in the test set.

ResNet-50 achieved the highest AUC in all three histologic subtypes (**Fig. 1A–C**). The AUCs in the test set for predicting malignant versus benign regions for chRCC, ccRCC, and pRCC using VGG-16 were 0.931 (95% CI, 0.926–0.936), 0.954 (95% CI, 0.953–0.956), and 0.954 (95% CI, 0.951–0.957), respectively. Using Inception-v3, the AUCs were higher for chRCC, ccRCC, and pRCC: 0.925 (95% CI, 0.920–0.931), 0.971 (95% CI, 0.969–0.972), and 0.977 (95% CI, 0.975–0.980). ResNet-50 had the best performance, with an AUC of 0.953 (95% CI, 0.949–0.957) for chRCC, 0.983 (95% CI, 0.982–0.983) for ccRCC, and 0.991 (95% CI, 0.990–0.992) for pRCC. Error analyses revealed that slide quality issues are the top reasons for misclassification (Supplementary Table S1).

To facilitate comparison with prior literature that reported slide-level predictions, we aggregated these predictions for each slide, and the AUCs increased as expected. The slide-aggregated AUCs in the test set for pRCC were 0.997 for VGG-16, 0.998 for Inception-v3, and 1.00 for ResNet-50. In chRCC, slide-aggregated AUCs were 0.987 for VGG-16, 0.980 for Inception-v3, and 0.990 for ResNet-50. In ccRCC, slide-aggregated AUCs were also higher (0.997 for VGG-16, 0.999 for Inception-v3, and 0.9998 for ResNet-50). We validated the ccRCC results in the independent validation set from CPTAC and achieved a region prediction AUC of 0.894 (95% CI, 0.892–0.896) with VGG-16, 0.937 (95% CI, 0.936–0.938) with Inception-v3, and 0.918 (95% CI, 0.916–0.919) with ResNet-50 (**Fig. 1D**). The slide-aggregated AUCs for CPTAC were 0.964 for VGG-16, 0.985 for Inception-v3, and 0.970 for ResNet-50. These results indicate the generalizability of our trained models.

To classify the three most prevalent histologic subtypes of renal cancer, we employed similar transfer learning approaches to establish DCNN models that classified these subtypes. The optimized VGG-16 model used a learning rate multiplier of 0.01 with the Adam optimizer, while Inception-v3 used a learning rate multiplier of 0.01 with the RMSprop optimization algorithm. ResNet-50 used a learning rate multiplier of 0.9 paired with the Adam optimizer. The multi-class AUC in the test set for this multi-class classification task was 0.926 with VGG-16, 0.897 with Inception-v3, and 0.953 with ResNet-50 (**Fig. 2A–C**). We further validated our subtype classification algorithms using a second independent dataset from the Brigham and Women's Hospital (Boston, MA). The multi-class classification AUC was 0.897 with VGG-16, 0.782 with Inception-v3, and 0.993 with ResNet-50 (**Fig. 2D**). Using ResNet-50, the one-versus-all AUC in pRCC was 0.990 (95% CI, 0.988–0.992), 0.991 in ccRCC (95% CI, 0.989–0.993), and 0.996 in chRCC (95% CI, 0.995–0.997).

### DCNNs revealed the relations between cell morphology and CNAs

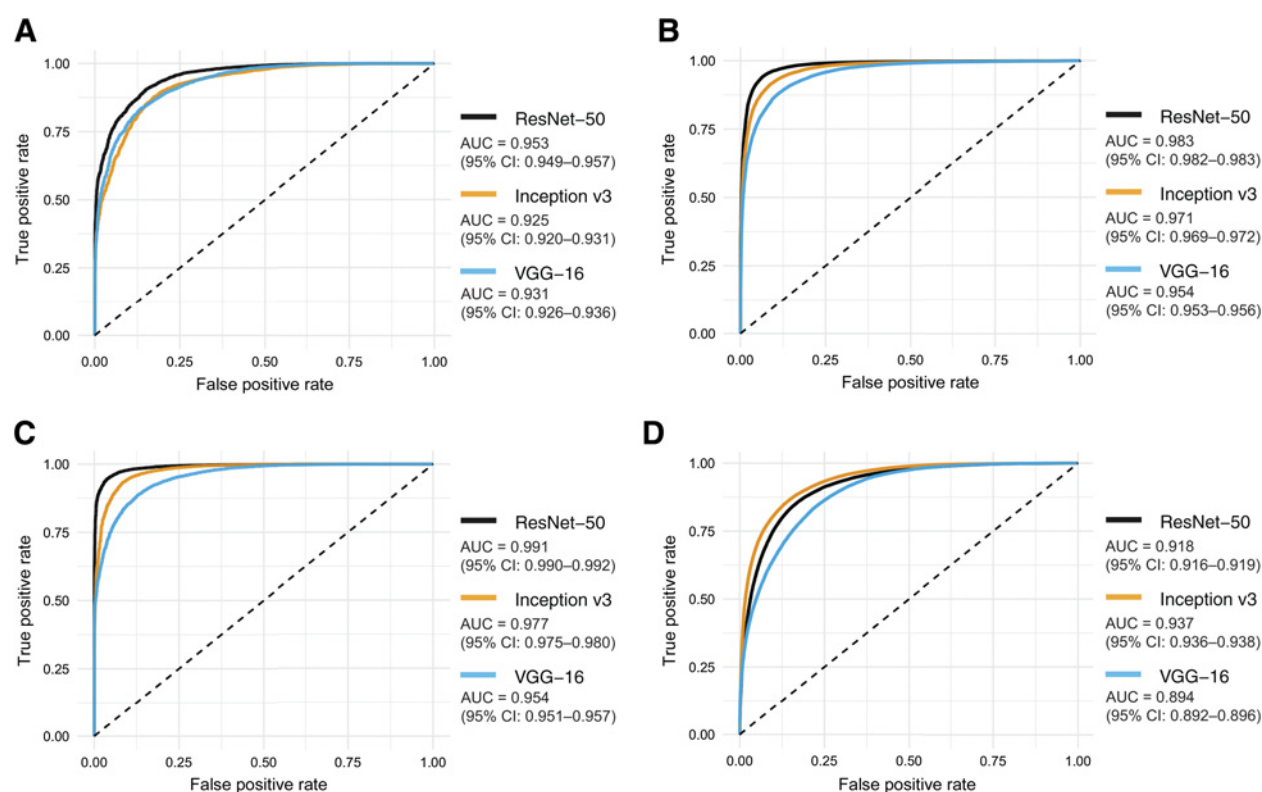
We investigated the linkages between renal cancer histopathology patterns and genomic aberrations by employing DCNNs to connect (i) cell morphology and CNAs and (ii) cell morphology and somatic mutations. We developed specific models for ccRCC, pRCC, and chRCC separately to control for histology types. When predicting CNA status, we selected clinically important genes and leveraged DCNNs to associate histopathology findings with these genetic variations. We employed two approaches to this task. The first approach trained individual binary classification models to predict the presence of CNAs in each gene (Supplementary Fig. S2). The second approach leveraged a multi-task classification design to predict all genetic

**Table 1.** Patient characteristics of TCGA and CPTAC Renal Cell Carcinoma cohorts. Stage denotes the clinical stage of the patient.

Number of patients	chRCC (TCGA) N = 103	pRCC (TCGA) N = 288	ccRCC (TCGA) N = 537	ccRCC (CPTAC) N = 222
Age	51.7 ± 13.8 years	61.5 ± 12.0 years	60.6 ± 12.2 years	60.2 ± 11.7 years
Sex				
Not available	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Male	58 (56.3%)	212 (73.6%)	346 (64.4%)	153 (68.9%)
Female	45 (43.7%)	76 (26.4%)	191 (35.6%)	69 (31.1%)
Race				
Not available	2 (1.9%)	15 (5.2%)	7 (1.3%)	112 (50.5%)
Native American or Alaska Native	0 (0.0%)	2 (0.7%)	0 (0.0%)	0 (0.0%)
Asian	2 (1.9%)	6 (2.1%)	8 (1.5%)	2 (0.9%)
Black	11 (10.7%)	60 (20.8%)	56 (10.4%)	3 (1.4%)
White	88 (85.4%)	205 (71.2%)	466 (86.8%)	105 (47.3%)
Clinical stage				
Not available	0 (0.0%)	27 (9.4%)	0 (0.0%)	17 (7.7%)
Discrepancy	0 (0.0%)	2 (0.7%)	3 (0.6%)	0 (0.0%)
Stage I	45 (43.7%)	171 (59.4%)	269 (50.1%)	104 (46.8%)
Stage II	32 (31.1%)	21 (7.3%)	57 (10.6%)	22 (9.9%)
Stage III	19 (18.4%)	52 (18.1%)	125 (23.3%)	58 (26.1%)

Abbreviations: CPTAC, Clinical Proteomic Tumor Analysis Consortium; TCGA, The Cancer Genome Atlas.





**Figure 1.**

Convolutional neural networks detected malignancy in whole-slide histopathology images of RCC. ROC curves of VGG-16, Inception-v3, and ResNet-50-based models for malignancy identification in TCGA chRCC (A), ccRCC (B), and prRCC (C) cohorts. External validation was performed in an independent ccRCC cohort from CPTAC (D). ResNet-50 achieved the highest performance in each TCGA RCC subtype, while Inception-v3 had slightly better performance in the CPTAC cohort.

variations simultaneously. In general, the multi-task classification model performed better on individual CNAs than the simple binary classification approach (Fig. 3). In ccRCC, the classification performance by ResNet-50 on KRAS CNA prediction was the highest (test set AUC = 0.724). Other genes with moderate signals include *WT1* (AUC = 0.721), *EGFR* (AUC = 0.717), and *VHL* (AUC = 0.705; Fig. 3A). Inception-v3 achieved similar performance in KRAS CNA prediction (AUC = 0.722) and *VHL* CNA prediction (AUC = 0.712; Fig. 3B). VGG-16 CNA prediction performance was less than 0.7 in all genes included in this task (Fig. 3C). The grad-CAM visualizations showed that regions of tumor clusters were of high importance in CNA prediction (Fig. 3D). These findings are consistent with the fact that the *VHL* gene is implicated in cancer cell proliferation and motility in RCCs (39). The task of predicting somatic mutations proved more difficult for DCNNs. In prRCC, the AUCs ranged from 0.419 (*KMT2C*; VGG-16) to 0.684 (*MET*; Inception-v3) for somatic mutations with known clinical relevance (Supplementary Fig. S3). We further conducted prediction for 9p deletion (27, 28), the deletion of *CDKN2A* (27, 28), and whole-genome doubling (40), and we found weak quantitative histopathology signals predicting these genomic variations (Supplementary Fig. S4).

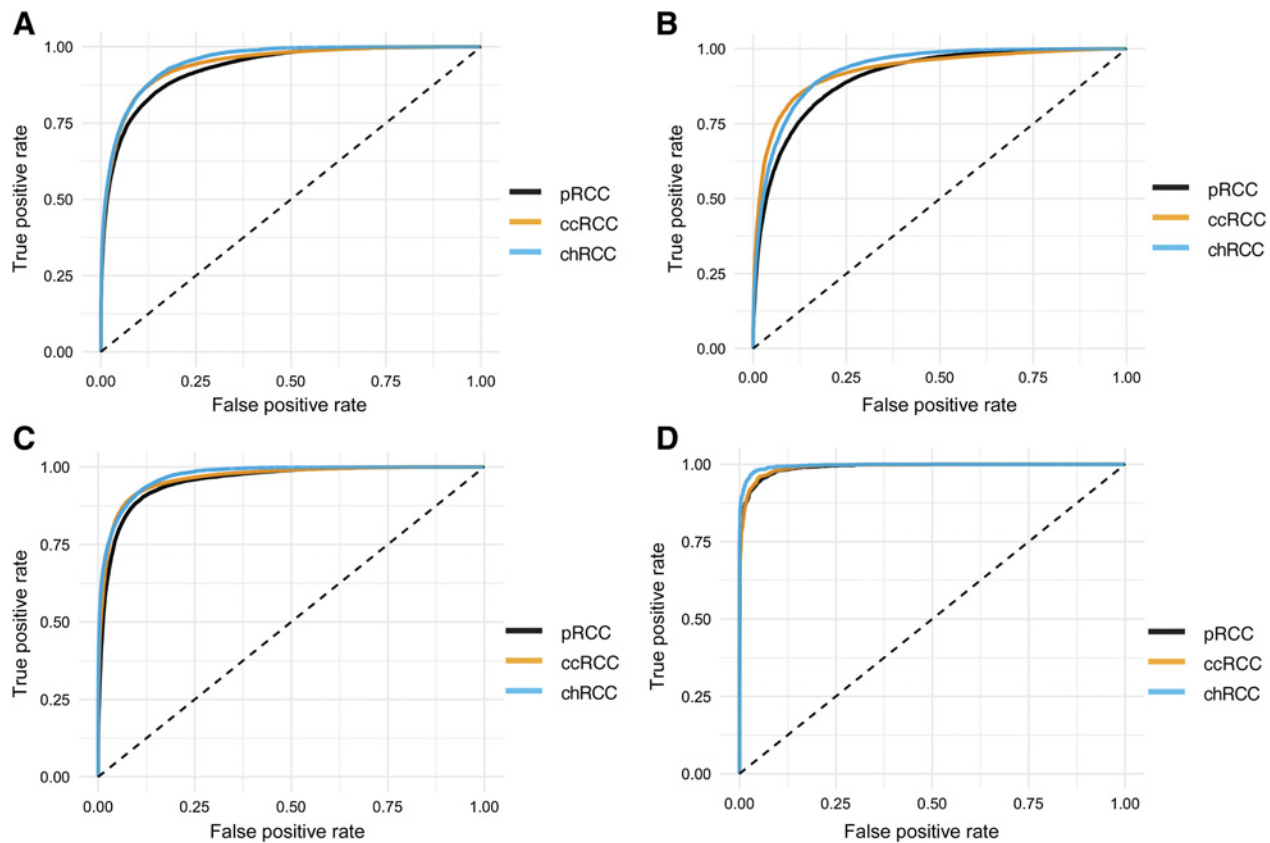
#### DCNNs predicted patients' overall survival outcomes using histopathology images

We further trained DCNN models to predict the overall survival of patients with renal cancer from the whole-slide H&E-stained histo-

pathology images. For each patient in the study cohort, we obtained the number of days from the initial diagnosis to mortality or the last day of follow-up. We employed multi-task logistic regression (30) survival models to directly connect histopathology images with the survival outcomes of each patient. On evaluation, the ResNet-50 model was able to differentiate longer-term survivors from shorter-term survivors among patients with stage I ccRCC (Fig. 4A; log-rank test  $P = 0.02$ ,  $n = 269$ ). grad-CAMs showed that regions occupied by tumor cell clusters received higher weights (Fig. 4B and C). Autonomous hyperparameter optimization within each fold of cross-validation found an optimal learning rate modifier of 0.01 to 0.001 for the Adam optimizer, depending on the fold. The ResNet-50 model performed better than the VGG-16 and Inception-v3-based architectures (Supplementary Fig. S5). These results suggest that the histopathology characteristics indicative of overall survival may be more subtle and require more advanced neural networks to extract the relevant signals.

#### DCNNs predicted the tumor mutation burden of the samples

Immune checkpoint inhibitors demonstrated significant survival benefits in patients with advanced RCC (41) and tumor mutation burden is a widely used biomarker for response to immune checkpoint blockade (42). However, it was unknown whether histopathology images contain signals predictive of tumor mutation burden. We leveraged a DCNN framework to predict patients' tumor mutation count using the histopathology images of their tumors, and we focused



**Figure 2.**

Convolutional neural networks successfully diagnosed patients with different subtypes of renal cell cancer. ROC curves for binary multi-class classification of chRCC, ccRCC, and pRCC using VGG-16 (A), Inception-v3 (B), and ResNet-50 (C) are shown. The multi-class area under the ROC curve for predicting histologic subtype was 0.926 for VGG-16, 0.897 for Inception-v3, and 0.953 for ResNet-50 (37, 38). The ResNet-50 algorithm was further validated in an additional independent validation set from Brigham and Women's Hospital (D), achieving a multi-class AUC of 0.993 (pRCC AUC = 0.990, 95% CI, 0.988–0.992; ccRCC AUC = 0.991, 95% CI, 0.989–0.993; chRCC AUC = 0.996, 95% CI, 0.995–0.997).

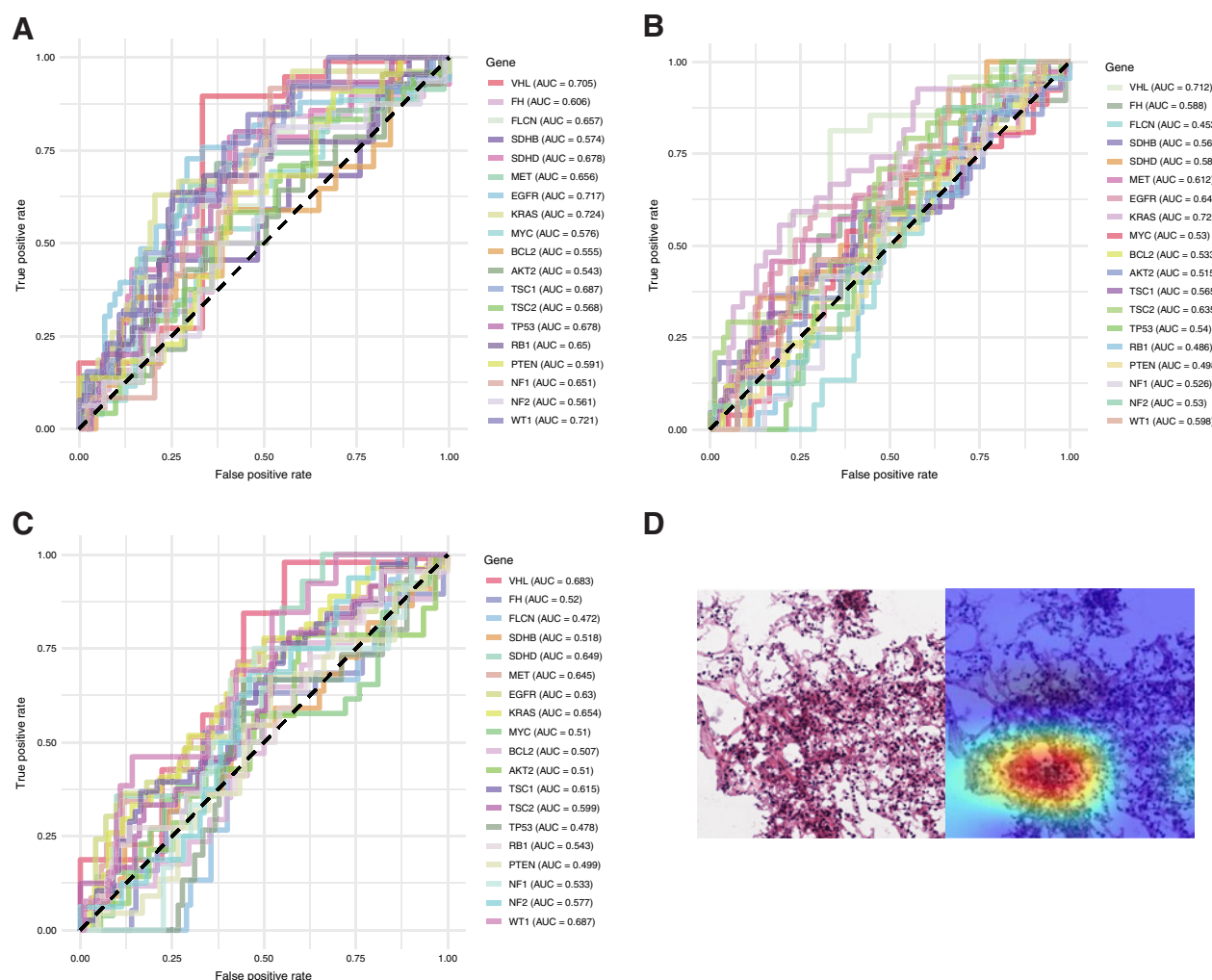
our analyses on the patients with ccRCC because the number of patients with available tumor mutation count in other subtypes was limited. Autonomous hyperparameter optimization selected a learning rate modifier of 3 with the Adam optimizer for ResNet-50. Using the early stopping mechanism to monitor the validation loss, the training process typically completed within 5 epochs. In the held-out test set, the Spearman correlation coefficient between true and predicted values was 0.419 (Spearman correlation test  $P = 0.0003$ ,  $n = 71$ ; Fig. 5A). Activation maximization visualizations of the output layer highlighted regions containing tumor cells with few immune cells (Fig. 5B and C). These results demonstrated that H&E-stained histopathology images of ccRCC tumor tissue contain previously unrecognized signals indicative of tumor mutation burden.

## Discussion

In this study, we established a generalizable informatics framework for predicting the subtypes, prognoses, and immunotherapy responses of patients with renal cancer using their digital whole-slide histopathology images, and we validated our results in independent test sets. We first created a pipeline using transfer learning to identify cancerous regions from histopathology slide images, classified the three major

histologic subtypes of RCC (AUC = 0.953), and validated our algorithms in two independent datasets from Brigham and Women's Hospital and CPTAC, respectively (AUC of the best model = 0.99). We further developed prediction models to identify patients' survival outcomes, CNA profiles, and tumor mutation burden. These results demonstrated that whole-slide histopathology images contain previously untapped biological signals useful for prognosis identification, molecular subtype characterization, and treatment response prediction. Our approaches have the potential to augment the current histopathology evaluation paradigm by providing the predicted patient prognoses and immunotherapy responses before the treatment is administered.

Leveraging state-of-the-art machine learning algorithms, computational pathology methods can extract useful signals from digital whole-slide histopathology images. Previous studies showed that DCNN models can diagnose several cancer types using histopathology images, with performance comparable with pathologists (11, 12, 15, 43). These deep learning methods have achieved AUCs as high as 0.994 in lymph node metastases of women with breast cancer (11), and 0.97 in non-small cell lung cancer (12). Our malignancy identification model for RCC showed comparable performance. It is worth noting that our weak supervision approach only requires slide-level labels without detailed region-level or pixel-level segmentation. Because the majority

**Figure 3.**

Multi-task classification of CNAs in ccRCC reveals cell morphology-genomic relations. ROC curves and gene-specific AUC values for the test sets using ResNet-50 (A), Inception-v3 (B), and VGG-16 (C) are shown. (D) grad-CAM visualization of a correctly predicted *VHL* CNA using Inception-v3 shows areas of tumor cells considered to be of high importance. Red indicates areas of high attention paid by the model for this prediction, while dark blue areas indicate areas of low attention.

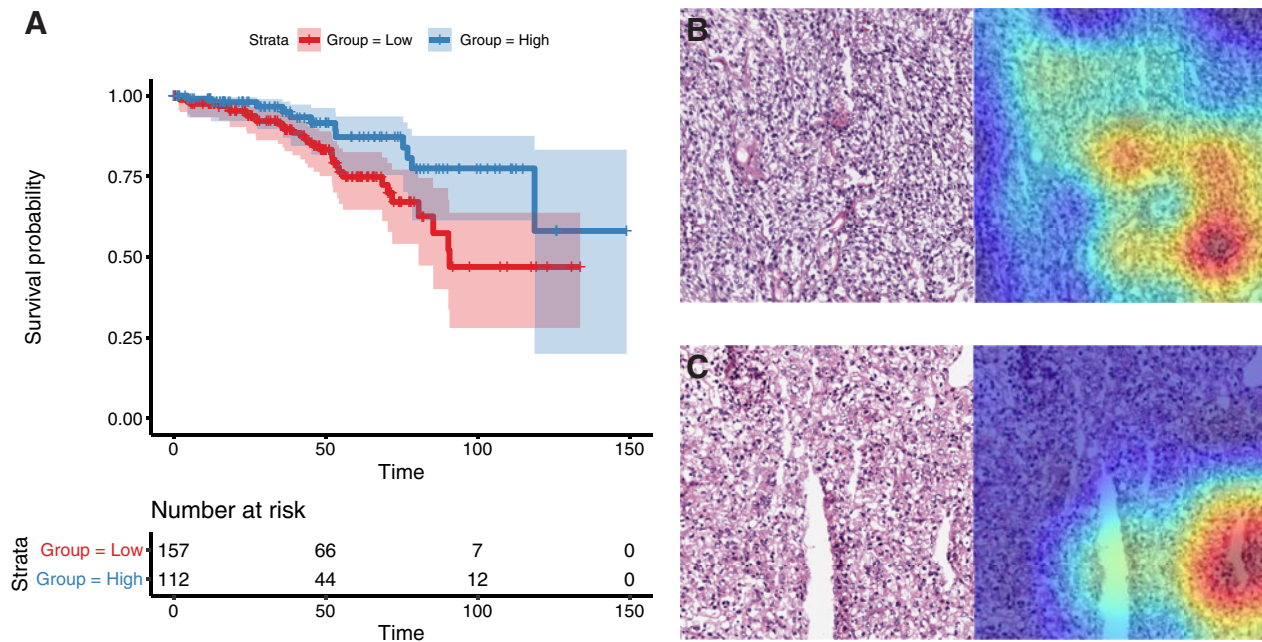
of histopathology slides in cancer centers do not come with region-level or pixel-level annotations, our approaches are directly applicable to clinical use cases, without additional efforts in manual segmentations. Since the diagnosis of each tumor sample in TCGA dataset is reviewed by a panel of renal pathologists, the data and labels gave the machine learning models room to grow beyond the capabilities of a single pathologist. In addition, our transfer learning approaches better optimized the parameters of the diagnostic classification model and achieved improved performance. A previous study focusing on the identification of cancerous regions of ccRCC and chRCC achieved a slide-aggregated AUC of 0.98 for ccRCC and 0.95 for chRCC (44). We achieved region-wise AUCs of 0.98 and 0.95 for ccRCC and chRCC (slide-aggregated AUC 0.9998 and 0.99), while also achieving an AUC of 0.991 in pRCC (slide-aggregated AUC of 1) and validated our results in an independent ccRCC dataset. We further developed a subtype classifier with a higher microaverage AUC compared with previous studies (0.95 vs. 0.91; ref. 44), with the results validated in two independent cohorts, one from a nationwide

research consortium and the other collected at our affiliated hospital.

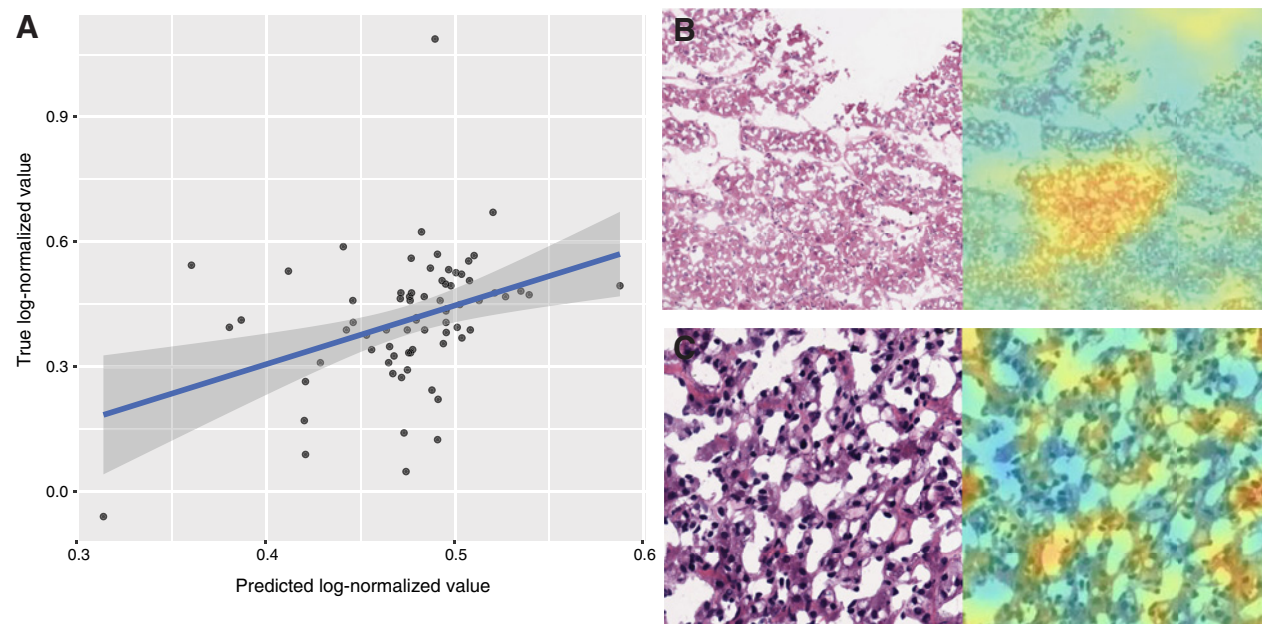
Furthermore, our analyses successfully connected high-resolution histopathology images with genomic and clinical profiles and revealed their unexpected relations. We extended the transfer learning paradigm to predict patients' survival outcomes and to discover molecular-morphologic connections in RCC. The success of our adapted multi-task logistic regression approach established a robust method for identifying overall survival outcomes of individual patients with RCC and shed light on previously overlooked prognostic patterns. These prognostic signals are directly informed by the data, without manual feature extraction or substantial human guidance. Robust prognostic prediction can guide personalized treatment plans and enable advance care planning (45).

Characterizing the genomic composition of a tumor can help determine optimized treatment plans for patients to improve their clinical outcomes. Previous studies suggested that CNAs were predictive of the survival and qualitative histology findings in



**Figure 4.**

Histopathology image patterns predicted the survival outcomes of patients with stage I ccRCC. Our convolutional neural networks coupled with multi-task logistic regression (30) distinguished longer-term survivors from shorter-term survivors using 10-fold cross-validation with ResNet-50 (**A**; log-rank test  $P = 0.02$ ,  $n = 269$ ). Original image patches and grad-CAM visualization of a correctly predicted shorter-term survivor (**B**) and a longer-term survivor (**C**) show areas of importance for survival prediction. Red indicates areas of high attention paid by the model for this prediction, while dark blue areas indicate areas of low attention.

**Figure 5.**

Deep convolutional neural networks illuminated trends in patients' ccRCC tumor mutation burden using histopathology images, suggesting the connection between histopathology image features and response to immune checkpoint inhibitors. **A**, True mutation count values were log transformed and normalized prior to model training. The median of patch-level predictions was aggregated for each patient. The Spearman correlation test reported a correlation coefficient of 0.419 ( $P = 0.0003$ ,  $n = 71$ ). **B**, Activation visualization of a histopathology image from a sample with a low tumor mutation count of 19. The ReLU backpropagation modifier was used to generate this visualization. **C**, Activation visualization of an image from a tumor sample with a relatively high mutation count of 80. Red regions suggest areas of high importance for higher mutation count values.



patients with ccRCC (46). Our studies demonstrated that CNAs in several genes, including *VHL*, *EGFR*, and *KRAS*, can affect quantitative histopathology patterns. These previously unknown molecular-morphologic connections can be systematically identified by fully automated computational procedures. We further identified weak signals in predicting whole-genome doubling. It is worth noting that the copy-number variations of several genes (e.g., *SDHB*, *MYC*, *BCL2*, *AKT2*, *PTEN*, and *NF2*) cannot be readily detected by our methods, which could be due to the subtle or minimal morphologic impact of these genes. Our analyses complement previous analyses on identifying mutations in single genes (47, 48) and demonstrated that a multi-task approach can better leverage the intrinsic relations among the molecular aberrations of different genes to arrive at improved predictions. Furthermore, our DCNN models predicted tumor mutation counts using histopathology image features, thereby revealing the connections between morphologic signals and patients who may benefit from immune checkpoint inhibitors or other novel treatments. The identified relations between histopathology and genomic variations of RCC tumors can facilitate personalized treatments in resource-limited settings where high-throughput genomic sequencing is not readily available. In addition, our prediction models for different genetic classification and prognostic tasks demonstrated the flexibility of neural network models to learn the connections of histopathology patterns and various clinical outcomes of interest. With the rapid development of novel treatments for RCC, our algorithms can expedite the identification of patient subgroups harboring the molecular and histopathology profiles relevant to the new treatment strategies.

We further compared the performance of three neural network architectures for RCC histopathology slide analysis. Results showed that ResNet-50 generally achieved better performance than VGG-16 and Inception-v3 across most tasks. Because of its high complexity, the ResNet-50 model was more likely to overfit the training data than the other two DCNN architectures. However, we overcame this obstacle by performing autonomous hyperparameter optimization to refine the models before touching any data in the test sets. ResNet-50 performed substantially better than VGG-16 and Inception-v3 in the multi-task classification of CNA and the tumor mutation burden task, suggesting its power to detect previously untapped biological signals in the histopathology images.

One limitation of the study is that even the large datasets we employed in this study may not capture the full spectrum of morphology heterogeneity in RCC. Samples with ambiguous morphologic patterns are not included in the two consortium studies, and the second validation set from our hospital may not fully represent the diverse morphology in different populations. Further studies are needed to evaluate the performance of our approaches in tumors with atypical histology manifestations and those belonging to rare molecular subtypes, such as sarcomatoid, rhabdoid, and pleomorphic histopathology, as well as those with MiTF translocation (49). Another limitation is that not all patients have received immune checkpoint blockade in our study cohorts and the tumor mutation counts were used as a proxy for immunotherapy response in our prediction model. Future research with clinically defined responses to immune checkpoint inhibitors could better ascertain the connections between histopathology image patterns and treatment responses. Moreover, although grad-CAMs visualize regions of importance that the models focus on, they are limited in resolution. Additional research on DCNN interpretability will help to elucidate the morphologic features related to the outcomes of interest.

Overall, our study developed automated methods to predict the subtypes, prognoses, and genomic aberrations of patients with RCC using histopathology images. Accurate diagnoses and successful prognostic predictions can guide clinical decision making, improve patients' outcomes, and reduce the cost of cancer management. Our approaches integrate information from multiple modalities, including imaging, multi-omics, and clinical data, and they are extensible to the histopathology evaluation of other complex diseases.

## Authors' Disclosures

E. Marostica reports that Harvard Medical School received the Blavatnik Center for Computational Biomedicine Award, the Schlager Family Award for Early Stage Digital Health Innovations from Brigham and Women's Hospital, and the Innovation Discovery Grant from Partners HealthCare during the conduct of the study. R. Barber reports grants from Harvard Medical School, Brigham and Women's Hospital, and Partners HealthCare during the conduct of the study. T. Denize reports grants from Harvard Medical School, Brigham and Women's Hospital, and Partners HealthCare during the conduct of the study. S. Signoretti reports grants from Exelixis and Bristol-Myers Squibb; personal fees from Merck, CRISPR Therapeutics, NCI, and AACR; and grants and personal fees from AstraZeneca outside the submitted work. In addition, S. Signoretti has a patent for Biogenex with royalties paid. K.-H. Yu reports grants from Harvard Medical School, Brigham and Women's Hospital, and Partners HealthCare during the conduct of the study; in addition, K.-H. Yu has a patent for U.S. Patent 10,832,406 issued and licensed to Harvard University. The authors thank the AWS Cloud Credits for Research, Microsoft Azure for Research Award, Google Cloud Platform research credit program, the NVIDIA GPU Grant Program, and the Extreme Science and Engineering Discovery Environment (XSEDE) at the Pittsburgh Supercomputing Center (allocation TG-BCS180016) for their computational supports. No disclosures were reported by the other authors.

## Authors' Contributions

**E. Marostica:** Data curation, software, formal analysis, validation, investigation, visualization, methodology, writing—original draft, writing—review and editing. **R. Barber:** Data curation, software, investigation. **T. Denize:** Resources, visualization, writing—review and editing. **I.S. Kohane:** Writing—review and editing. **S. Signoretti:** Resources, writing—review and editing. **J.A. Golden:** Writing—review and editing. **K.-H. Yu:** Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing—original draft, project administration, writing—review and editing.

## Acknowledgments

We thank Mr. Alexander Bruce for his assistance with slide scanning at the Digital Imaging Facility, Department of Pathology, Brigham and Women's Hospital; Eliezer Van Allen and Arjun Manrai for their feedback on this project; and Samantha Lemos, Susan Marone, and Nichole Parker for their administrative support. We thank the AWS Cloud Credits for Research, Microsoft Azure for Research Award, Google Cloud Platform research credit program, the NVIDIA GPU Grant Program, and the Extreme Science and Engineering Discovery Environment (XSEDE) at the Pittsburgh Supercomputing Center (allocation TG-BCS180016) for their computational support. K.-H. Yu is partly supported by the Partners' Innovation Discovery Grant, the Schlager Family Award for Early Stage Digital Health Innovations, and the Blavatnik Center for Computational Biomedicine Award. This work was conducted with support from the Digital Imaging Facility, Department of Pathology, Brigham and Women's Hospital, Boston, MA and with financial contributions from Brigham and Women's Hospital.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 19, 2020; revised January 25, 2021; accepted March 10, 2021; published first March 15, 2021.

## References

- Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, et al. Epidemiology of renal cell carcinoma. *Eur Urol* 2019;75:74–84.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- Lopez-Beltran A, Carrasco JC, Cheng L, Scarpelli M, Kirkali Z, Montironi R. 2009 update on the classification of renal epithelial tumors in adults: updated classification of renal tumors. *Int J Urol* 2009;16:432–43.
- Linehan WM, Marston Linehan W, Srinivasan R, Schmidt LS. The genetic basis of kidney cancer: a metabolic disease. *Nat Rev Urol* 2010;7:277–85.
- Linehan WM. Genetic basis of kidney cancer: role of genomics for the development of disease-based therapeutics. *Genome Res* 2012;22:2089–100.
- Prasad SR, Humphrey PA, Catena JR, Narra VR, Srigley JR, Cortez AD, et al. Common and uncommon histologic subtypes of renal cell carcinoma: imaging spectrum with pathologic correlation. *Radiographics* 2006;26:1795–806.
- Shuch B, Pantuck AJ, Pouliot F, Finley DS, Said JW, Beldegrun AS, et al. Quality of pathological reporting for renal cell cancer: implications for systemic therapy, prognostication and surveillance. *BJU Int* 2011;108:343–8.
- Shuch B, Hofmann JN, Merino MJ, Nix JW, Vourganti S, Linehan WM, et al. Pathologic validation of renal cell carcinoma histology in the Surveillance, Epidemiology, and End Results program. *Urol Oncol* 2014;32:23.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
- Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018;2:719–31.
- Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karsssemeijer 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.
- Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24:1559–67.
- Yu K-H, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016;7:12474.
- Yu K-H, Berry GJ, Rubin DL, Ré C, Altman RB, Snyder M. Association of omics features with histopathology patterns in lung adenocarcinoma. *Cell Syst* 2017;5:620–7.
- Yu K-H, Wang F, Berry GJ, Ré C, Altman RB, Snyder M, et al. Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks. *J Am Med Inform Assoc* 2020;27:757–69.
- Hipp J, Flotte T, Monaco J, Cheng J, Madabhushi A, Yagi Y, et al. Computer aided diagnostic tools aim to empower rather than replace pathologists: lessons learned from computational chess. *J Pathol Inform* 2011;2:25.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 2013;499:43–9.
- Cancer Genome Atlas Research Network, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, et al. Comprehensive molecular characterization of papillary renal-cell carcinoma. *N Engl J Med* 2016;374:135–45.
- Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, Shen H, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* 2014;26:319–30.
- Yu K-H, Fitzpatrick MR, Pappas L, Chan W, Kung J, Snyder M. Omics Analysis System for Precision Oncology (OASISPRO): a web-based omics analysis tool for clinical phenotype prediction. *Bioinformatics* 2018;34:319–20.
- CPTAC-CCRCC - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. Available from: <https://wiki.cancerimagingarchive.net/display/Public/CPTAC-CCRCC>.
- autonomio talos [Computer software]; 2019. Available from: <https://github.com/autonomio/talos>.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition Available from: <http://arxiv.org/abs/1409.1556>.
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas (NV): Computer Vision Foundation; 2016. p. 2818–26. Available from: [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2016/html/Szegedy\\_Rethinking\\_the\\_Inception\\_CVPR\\_2016\\_paper.html](https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html).
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas (NV): Computer Vision Foundation; 2016. p. 770–8. Available from: [https://openaccess.thecvf.com/content\\_cvpr\\_2016/html/He\\_Deep\\_Residual\\_Learning\\_CVPR\\_2016\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html).
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015;115:211–52.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401–4.
- Schmidt LS, Srinivasan R, Marston Linehan W. The genetic basis of kidney cancer and implications for targeted therapies. *Renal cell carcinoma*. Boston: Springer; 2012. p. 3–25.
- Yu C-N, Greiner R, Lin H-C, Baracos V. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ, editors. *Advances in neural information processing systems* 24. Red Hook, NY: Curran Associates, Inc.; 2011. p. 1845–53.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: <https://www.R-project.org/>.
- Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. *Statistics for biology and health*. New York: Springer-Verlag; 2000.
- Therneau TM. A package for survival analysis in R. *Comprehensive R Archive Network (CRAN)*; 2015. Available from: <https://CRAN.R-project.org/package=survival>.
- Kassambara A. survminer: drawing survival curves using “ggplot2”. *Comprehensive R Archive Network (CRAN)*; 2018. Available from: <https://CRAN.R-project.org/package=survminer>.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice (Italy): Computer Vision Foundation; 2017. p. 618–26. Available from: [https://openaccess.thecvf.com/content\\_iccv\\_2017/html/Selvaraju\\_Grad-CAM\\_Visual\\_Explanations\\_ICCV\\_2017\\_paper.html](https://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html).
- Kotikalapudi R. keras-vis. Github. Available from: <https://github.com/raghakot/keras-vis>.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
- Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach Learn* 2001;45:171–86.
- Baldewijns MM, van Vlodrop IJH, Vermeulen PB, Soetekouw PMMB, van Engeland M, de Bruine AP. VHL and HIF signalling in renal cell carcinogenesis. *J Pathol* 2010;221:125–38.
- Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 2018;173:400–16.
- Carlo MI, Voss MH, Motzer RJ. Checkpoint inhibitors and other novel immunotherapies for advanced renal cell carcinoma. *Nat Rev Urol* 2016;13:420–31.
- Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer* 2019;19:133–50.
- Yu K-H, Hu V, Wang F, Matulonis UA, Mutter GL, Golden JA, et al. Deciphering serous ovarian carcinoma histopathology and platinum response by convolutional neural networks. *BMC Med* 2020;18:236.
- Tabibu S, Vinod PK, Jawahar CV. Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning. *Sci Rep* 2019;9:10509.

45. Kim P, Daly JM, Berry-Stoelzle MA, Schmidt ME, Michaels LC, Dorr DA, et al. Prognostic indices for advance care planning in primary care: a scoping review. *J Am Board Fam Med* 2020;33:322–38.
46. Pan C-C, Yeh Y-C, Wang Y-C, Chang Y-H. Differential expression analysis of clear cell renal cell carcinomas in THE CANCER GENOME ATLAS distinguishes an aggressive subset enriched with chromosome 7 and 12 gains. *Histopathology* 2020;76:950–68.
47. Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nature Cancer* 2020;1:800–10.
48. Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer* 2020;1:789–99.
49. Argani P. MiT family translocation renal cell carcinoma. *Semin Diagn Pathol* 2015;32:103–13.