

RESEARCH ARTICLE

Automated clear cell renal carcinoma grade classification with prognostic significance

Katherine Tian^{1,2} , Christopher A. Rubadue¹ , Douglas I. Lin¹, Mitko Veta³, Michael E. Pyle¹, Humayun Irshad¹, Yujing J. Heng ^{1,4*}

1 Department of Pathology, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA, United States of America, **2** The Harker School, San Jose, CA, United States of America, **3** Medical Image Analysis Group, Eindhoven University of Technology, Eindhoven, The Netherlands, **4** Cancer Research Institute, Beth Israel Deaconess Cancer Center, Boston, MA, United States of America

 These authors contributed equally to this work.

* yheng@bidmc.harvard.edu



Abstract

We developed an automated 2-tiered Fuhrman's grading system for clear cell renal cell carcinoma (ccRCC). Whole slide images (WSI) and clinical data were retrieved for 395 The Cancer Genome Atlas (TCGA) ccRCC cases. Pathologist 1 reviewed and selected regions of interests (ROIs). Nuclear segmentation was performed. Quantitative morphological, intensity, and texture features ($n = 72$) were extracted. Features associated with grade were identified by constructing a Lasso model using data from cases with concordant 2-tiered Fuhrman's grades between TCGA and Pathologist 1 (training set $n = 235$; held-out test set $n = 42$). Discordant cases ($n = 118$) were additionally reviewed by Pathologist 2. Cox proportional hazard model evaluated the prognostic efficacy of the predicted grades in an extended test set which was created by combining the test set and discordant cases ($n = 160$). The Lasso model consisted of 26 features and predicted grade with 84.6% sensitivity and 81.3% specificity in the test set. In the extended test set, predicted grade was significantly associated with overall survival after adjusting for age and gender (Hazard Ratio 2.05; 95% CI 1.21–3.47); manual grades were not prognostic. Future work can adapt our computational system to predict WHO/ISUP grades, and validating this system on other ccRCC cohorts.

OPEN ACCESS

Citation: Tian K, Rubadue CA, Lin DI, Veta M, Pyle ME, Irshad H, et al. (2019) Automated clear cell renal carcinoma grade classification with prognostic significance. PLoS ONE 14(10): e0222641. <https://doi.org/10.1371/journal.pone.0222641>

Editor: Mathieu Hatt, INSERM, FRANCE

Received: May 28, 2019

Accepted: September 4, 2019

Published: October 3, 2019

Copyright: © 2019 Tian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used in this study were in whole or in part based on the data generated by the TCGA Research Network: (<http://cancergenome.nih.gov/>). TCGA data are available at (<http://cancergenome.nih.gov/>). The codes for image analysis are available at: (https://github.com/humayun/NucleiMorphology_Fiji); (https://github.com/humayun/NucleiSeg_FeatureExt_C); (https://github.com/humayun/expath/blob/master/Classification/PreExpansion_ImageClassification.R).

Funding: YJH is supported by the Klarman Family Foundation. KT was supported by the BIDMC High

Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common malignant tumor of epithelial origin in the kidney [1]. For over 30 years, ccRCC was graded using the 4-tiered Fuhrman nuclear grading system which incorporates nuclear size, nucleolar prominence, and nuclear membrane irregularities. Diagnostic challenges can occur with the presence of other morphological features such as sarcomatoid or spindle cell pattern, when higher grade ccRCC show more eosinophilic staining in the cytoplasm, or other renal cancer histologic types (e.g. papillary RCC type1 and chromophobe RCC) exhibit clear cytoplasm [2,3]. The correct classification of ccRCC grade and stage is important for guiding clinical management, molecular-based

School Summer Research Program. At the time of this work, Drs. Lin and Irshad were both employees of BIDMC, and were not at commercial companies (i.e., Foundation Medicine for Lin and Figure-Eight Technologies Inc for Irshad). Those two commercial affiliations did not provide salaries for Lin and Irshad at the time of the study, and did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: DIL is a Senior Pathologist at Foundation Medicine. HI is a Scientist at Figure-Eight Technologies Inc. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

therapies, and prognosis [4,5]. Fuhrman grade is widely accepted as a prognostic factor despite mediocre inter-observer agreement [6,7]. To improve inter-observer agreement, simplified 2- or 3-tiered grading systems have been proposed. These simplified systems appear to retain prognostic ability similar to that of 4-tiered systems [8,9]. Recently, a new nuclear/nucleolar grading system, known as the World Health Organization (WHO)/International Society of Urological Pathology (ISUP) Grading Classification for RCC, was introduced [10].

Technological advances have enabled computational pathology to discover novel histomics features from whole slide images (WSIs) that may add diagnostic and/or prognostic information [11–13]. Computational pathology techniques can analyze cancer WSIs [14–16], including the detection of malignant RCC cells [17]. In this study, we developed an automated grading system to predict 2-tiered Fuhrman grade using ccRCC WSIs from The Cancer Genome Atlas (TCGA). Our specific aims were to establish a computational pipeline to extract nuclei histomics features, develop a model to predict 2-tiered ccRCC grade, and evaluate the prognostic efficacy of computer predicted grades.

Materials and methods

Cases and grade assignment

TCGA ccRCC clinical data, including Fuhrman's grade (accessed June 2017), and hematoxylin and eosin (H&E) WSIs were retrieved for 395 cases [18,19]. TCGA ccRCC cases were contributed by seven participating medical centers. The TCGA Fuhrman's grade for each case is the consensus of at least two pathologists from the case's medical center. In order to identify tumor areas on each diagnostic WSI (i.e., regions of interest (ROIs)) for this computational pathology study, Pathologist 1 reviewed each WSI, identified an average of five ROIs for each case (Fig 1), and assigned a Fuhrman grade of 1 to 4 for each ROI. The highest grade among all the ROIs was the designated grade. Thus, each patient had two assigned grades: "TCGA grade" and "Grade by Pathologist 1". TCGA and Pathologist 1 grades were re-stratified into the 2-tiered grading system: low (grades 1 and 2) and high (grades 3 and 4).

Image processing and nuclei segmentation

ROIs ($n = 1855$) from 395 WSIs were extracted and split into 2000 pixel by 2000 pixel patches (Fig 2). Nuclei segmentation was performed using Fiji (ImageJ, National Institutes of Health) [20] and using our previously published workflow [14]. H&E patches were converted from the Red, Green, and Blue (RGB) color space to the Hue, Saturation, and Value (HSV) color space (i.e., binary patches; Fig 3). A nonlinear mapping approach was applied as preprocessing to handle the variation across H&E staining inconsistency [21]. The nuclei segmentation method consists of two steps: adaptive thresholding in each HSV color channel to identify nuclei regions from the background, and marker controlled watershed-based nuclei segmentation to separate touching and overlapping nuclei. We further applied morphological operations to fine-tune the segmentation of nuclei. Extracted nuclei of area less than 200 pixels or greater than 2000 pixels were excluded to improve the specificity of nuclear detection [14].

2D histomics feature extraction

For each patch, 72 nuclei histomics features were extracted: nine morphological features, 15 intensity-based features, and 48 texture-based features. Morphological features describe the shape and size variation of nuclei. Intensity features (first order statistical features) describe the distribution of color variation in the nucleus. Three color channels were analyzed: lightness from HSV color space, lightness from Lab color space, and Hematoxylin channel from H&E

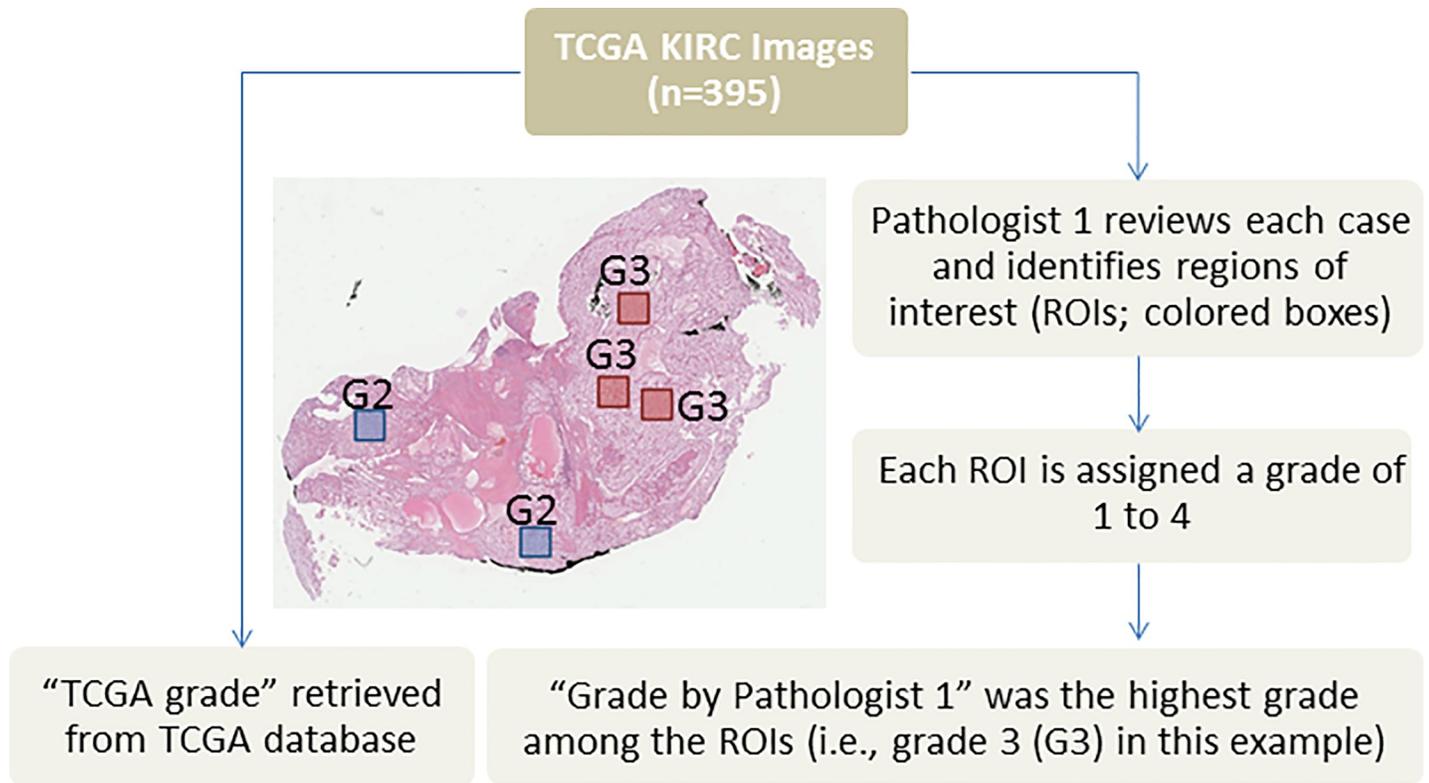


Fig 1. Schematic diagram showing how regions of interest (ROIs) were identified by Pathologist 1. Pathologist 1 identified ROIs and assigned a Fuhrman grade for each ROI. The highest grade among all ROIs was the “Grade by Pathologist 1”. Each case also had a “TCGA grade” retrieved from the TCGA database.

<https://doi.org/10.1371/journal.pone.0222641.g001>

color deconvolution [22]. Five first order statistical features were computed—mean, median, standard deviation, skewness, and kurtosis—for each of the three color channels, for a total of 15 intensity features. Texture features (second order statistical features) quantitatively describe

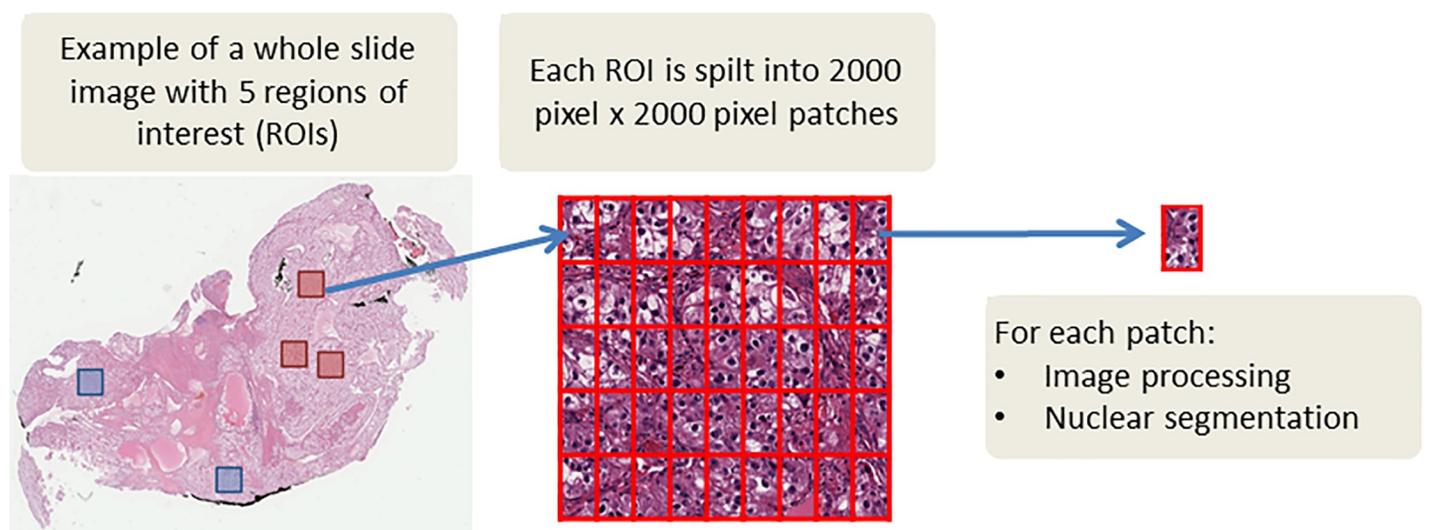


Fig 2. From whole slide image to patches for image processing and nuclei segmentation.

<https://doi.org/10.1371/journal.pone.0222641.g002>

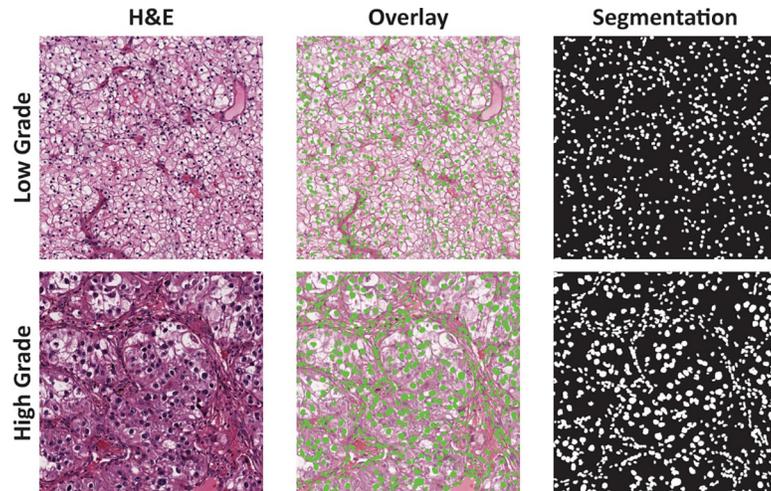


Fig 3. Examples of nuclei detection and segmentation in low and high grade clear cell renal cell carcinoma. The rightmost column shows computer-generated segmentation mask where cell nuclei are labelled white against a black background. The middle column shows the overlay of segmented nuclei (green spots) over each hematoxylin and eosin (H&E) patch.

<https://doi.org/10.1371/journal.pone.0222641.g003>

patterns and texture of pixel values. Two types of second order statistical features were computed: co-occurrence based features ($n = 8$) and run length based features ($n = 8$). Co-occurrence based features include correlation, cluster shade, cluster prominence, energy, entropy, Haralick correlation, inertia, and inverse difference moment [23]. Run length based features include gray-level non-uniformity, run-length non-uniformity, low and high gray-level run emphasis, short run low and high gray-level emphasis, and long run low and high gray-level emphasis [24]. Likewise, texture features were extracted from the three selected color channels, resulting in a total of 48 texture features. Feature formulas have been previously described [14,25].

Data summarization and selection of representative ROI

Data extracted at the patch level were summarized to the ROI level by calculating the median and median absolute deviation (MAD) (i.e., 144 summarized features). Some cases had multiple ROIs annotated with the highest grade. Thus, one ROI among the highest grade ROIs was selected to represent the case. To do so, the median of all ROIs with the highest grade was calculated, and the ROI with the smallest Euclidean-distance to the calculated median was chosen (Fig 4).

Developing the machine learning model to predict grade

Cases with concordant 2-tiered grade by TCGA and Pathologist 1 ($n = 277$) were used to develop the automated 2-tiered grading system. Concordant cases were split into a training set ($n = 235$; 85%) and held-out test set ($n = 42$; 15%; Fig 5). The *sampling* package, R, was used to select the 42 patients in the held-out test set based on grade, age, gender, and stage, ensuring that they were representative of the concordant cases. Histomics features were z-scored. Seven machine learning classification methods were explored to classify ccRCC cases into either low or high grade using nuclei histomics features [26,27] (Fig 5). All methods achieved similar area under the receiver-operator characteristic curves (AUC ROC; S1 Table). Lasso regression was the top performing method with a built-in feature selection capability. Lasso regression is one

Data extracted at the patch level were combined to calculate the median and median absolute deviation for each feature for each ROI

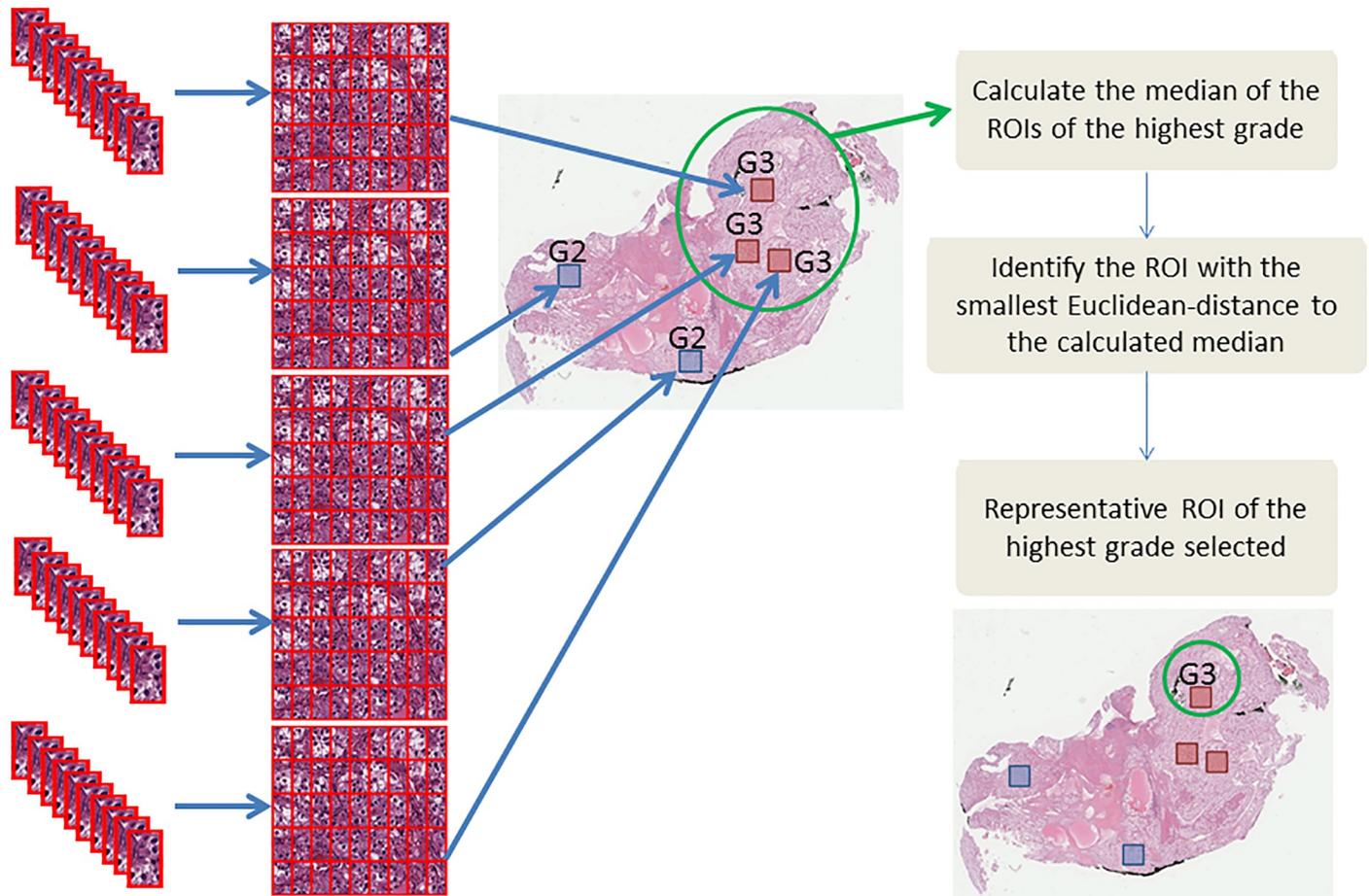


Fig 4. Data summarization and the selection of the representative region of interest (ROI).

<https://doi.org/10.1371/journal.pone.0222641.g004>

type of linear regression with L1 regularization. The Lasso procedure uses L1 regularization penalty, which has the effect of shrinking the regression weights of the least predictive features to 0, thereby creating simpler models that are less prone to overfitting [28]. In the Lasso model, a hyper parameter λ determines the amount of the L1 regularization penalty applied. We decided to move forward to use Lasso to build our final classification model because it is computationally efficient and more interpretable compared to other machine learning methods such as deep learning. Lasso regression and its optimal hyper parameter selected the final list of histomics features most associated with grade. We evaluated its performance on the held out test set.

Survival analyses

The Lasso model was applied to predict the grade of the previously held out test set ($n = 42$) and cases with discordant grades ($n = 118$). These 160 cases were combined to create an extended test set to evaluate the prognostic capability (i.e., overall survival [OS]) of our predicted grade using crude and adjusted Cox proportional hazard models. The adjusted Cox

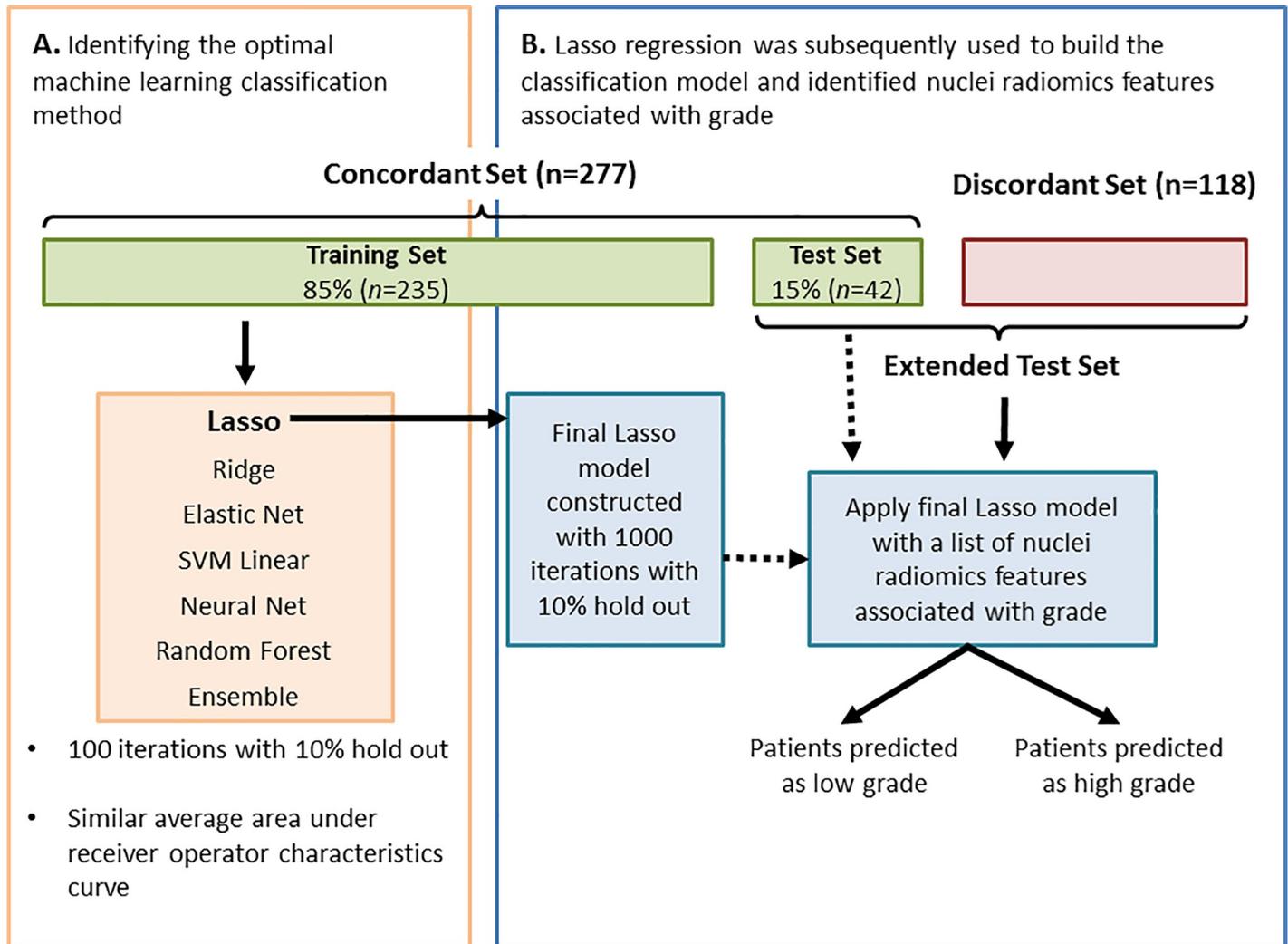


Fig 5. A summary of the workflow used to develop the 2-tiered clear cell renal cell carcinoma (ccRCC) grade classification. Seven machine learning classification methods were evaluated to determine the optimal method to develop a robust classification model for ccRCC using cases from the Training Set (A). Lasso regression produced an average area under the receiver operator characteristic curve of 0.84 and identified nuclei histomics features associated with ccRCC grade. The Test Set was used to evaluate the performance of the final model; and grades were predicted in the Extended Test Set (B).

<https://doi.org/10.1371/journal.pone.0222641.g005>

models include patient age, gender, and cancer stage. TCGA treatment information was missing from 69% of the cases and thus was not included in the adjusted Cox models. Kaplan-Meier curves were plotted to visualize differences between the curves (survival package, R) [29].

Additional pathological review for discordant cases

The grades provided by TCGA may be assessed from ROIs other than the representative ROIs selected in our study. To obtain a fairer comparison between manual and predicted grades among the discordant cases, the representative ROIs were additionally reviewed by Pathologist 2.

Statistical analyses

Confusion matrices determined the concordance of the 2-tiered and 4-tiered grades between two raters [27]. Inter-rater reliability among three raters was evaluated using Fleiss' kappa.

Boxplots were created using ggplot2 version 2.2.1. Comparisons between the nine morphological features with 2-tiered and 4-tiered grading were done using Mann-Whitney U or Kruskal Wallis test, respectively. All tests of statistical significance were two-sided. Statistical significance was achieved when p -value was <0.05 or when the false discovery rate (FDR) was <0.05 . All analyses were conducted using R version 3.4.0.

Results

The majority of TCGA ccRCC cases were white males. Most participants were between the ages of 50 to 69 and had stage I disease (Table 1). The agreement of 4-tiered grading between TCGA and Pathologist 1 was poor (frequency of agreement = 0.47, Cohen’s kappa = 0.20; S1A Fig). When the grading was stratified into 2-tiers, 277 out of 395 cases were concordant (frequency of agreement = 0.70, Cohen’s kappa = 0.41; S1B Fig). Most of the discordant cases were assigned high grade by TCGA and low grade by Pathologist 1.

Computer extracted morphological features reflected the variation of ccRCC nuclei as observed by pathologists. Nuclei size (i.e., area, perimeter, and spherical perimeter and radius)

Table 1. Demographic table of the 395 The Cancer Genome Atlas (TCGA) clear cell renal cell carcinoma cases with 2-tiered histological grade (low and high). Note that the TCGA grade for each patient in the discordant set is the opposite grade assigned by Pathologist 1.

	Total <i>n</i> (%)	Concordant Cases		Discordant cases (Grades by TCGA)		Discordant cases (Grades by Pathologist 1)	
		Low <i>n</i> (%)	High <i>n</i> (%)	Low <i>n</i> (%)	High <i>n</i> (%)	Low <i>n</i> (%)	High <i>n</i> (%)
Cases, <i>n</i>	395 (100)	162 (58.5)	115 (41.5)	28 (23.7)	90 (76.3)	90 (76.3)	28 (23.7)
Age group, <i>n</i>							
<50	80 (20.3)	36 (22.2)	22 (19.1)	4 (14.3)	18 (20.0)	18 (20.0)	4 (14.3)
50–59	106 (26.8)	50 (30.9)	29 (25.2)	6 (21.4)	21 (23.3)	21 (23.3)	6 (21.4)
60–69	109 (27.6)	37 (22.8)	33 (28.7)	10 (35.7)	29 (32.2)	29 (32.2)	10 (35.7)
70–79	82 (20.8)	31 (19.1)	26 (22.6)	8 (28.6)	17 (18.9)	17 (18.9)	8 (28.6)
>80	18 (4.6)	8 (4.9)	5 (4.3)	0 (0.0)	5 (5.6)	5 (5.6)	0 (0.0)
Gender, <i>n</i>							
Female	130 (32.9)	67 (41.4)	28 (24.3)	10 (35.7)	25 (27.8)	25 (27.8)	10 (35.7)
Male	265 (67.1)	95 (58.6)	87 (75.7)	18 (64.3)	65 (72.2)	65 (72.2)	18 (64.3)
Race, <i>n</i>							
Asian	7 (1.8)	3 (1.9)	2 (1.7)	0 (0.0)	2 (2.2)	2 (2.2)	0 (0.0)
Black	33 (8.4)	13 (8.0)	10 (8.7)	2 (7.1)	8 (8.9)	8 (8.9)	2 (7.1)
White	349 (88.4)	142 (87.7)	102 (88.7)	25 (89.3)	80 (88.9)	80 (88.9)	25 (89.3)
Not reported	6 (1.5)	4 (2.5)	1 (0.9)	1 (3.6)	0 (0.0)	0 (0.0)	1 (3.6)
Stage, <i>n</i>							
Stage I	207 (52.4)	121 (74.7)	35 (30.4)	13 (46.4)	38 (42.2)	38 (42.2)	13 (46.4)
Stage II	44 (11.1)	17 (10.5)	15 (13.0)	5 (17.9)	7 (7.8)	7 (7.8)	5 (17.9)
Stage III	92 (23.3)	18 (11.1)	36 (31.3)	8 (28.6)	30 (33.3)	30 (33.3)	8 (28.6)
Stage IV	52 (13.2)	6 (3.7)	29 (25.2)	2 (7.1)	15 (16.7)	15 (16.7)	2 (7.1)
Type of Treatment, <i>n</i>							
Chemotherapy	7 (1.8)	3 (1.9)	4 (3.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Immunotherapy	6 (1.5)	2 (1.2)	2 (1.7)	0 (0.0)	2 (2.2)	2 (2.2)	0 (0.0)
Molecular therapy	79 (20.0)	31 (19.1)	24 (20.9)	5 (17.9)	19 (21.1)	19 (21.1)	5 (17.9)
Radiation	10 (2.5)	2 (1.2)	4 (3.5)	2 (7.1)	2 (2.2)	2 (2.2)	2 (7.1)
Mixed therapy	21 (5.3)	4 (2.5)	10 (8.7)	2 (7.1)	5 (5.6)	5 (5.6)	2 (7.1)
Unknown	272 (68.9)	120 (74.1)	71 (61.7)	19 (67.9)	62 (68.9)	62 (68.9)	19 (67.9)

<https://doi.org/10.1371/journal.pone.0222641.t001>

and shape (i.e., roundness, elongation, flatness and major axis of ellipse fit) were significantly larger and less spherical in higher grades (FDR<0.05; [S2 Fig](#) and [S2 Table](#)).

Lasso classification model

The final Lasso model with the optimal λ at 0.0101 had an average ROC AUC of 0.84. The model predicted 2-tiered ccRCC grade with 83.3% accuracy (95% confidence interval (CI) 0.69–0.93), 84.6% sensitivity, 81.3% specificity, 18.8% false positive rate, and 15.4% false negative rate in the test set. The agreement between predicted and manual grades was good (frequency of agreement = 0.83, Cohen's kappa = 0.65). The 18 unique histomics features associated with ccRCC 2-tiered grade are in [Table 2](#).

Prognostic efficacy of predicted grades

There were 65 death events out of 160 cases in the extended test set. Cases predicted as high grade had significantly poorer OS compared to low grade ([Fig 6](#)). The association between predicted grade and OS was significant in the crude analysis (hazard ratio (HR) 2.07; 95% CI 1.25–3.43) and after adjusting for age and gender (HR 2.05; 95% CI 1.21–3.47). The association was attenuated when stage was included in the model (HR 1.66; 95% CI 0.97–2.83).

Comparing predicted grade with TCGA and Pathologist 1

Among the concordant cases, 2-tiered manual grades were significantly associated with OS ([Fig 7A](#); [Table 3](#)). Predicted grade for concordant cases were not evaluated as the majority of the concordant cases were part of the training set used to build the Lasso model. Within the discordant cases, neither grade provided by TCGA nor Pathologist 1 was associated with OS ([Fig 7B and 7C](#)). Predicted grade was significantly associated with OS (crude model HR 2.01; 95% CI 1.14–3.54) and when adjusted for age and gender (HR 2.31; 95% CI 1.26–4.24). The association of predicted grade and OS among the discordant cases was attenuated when adjusted stage was included in the model (HR 1.83; 95% CI 0.98–3.41; [Fig 7D](#); [Table 3](#)).

Additional pathological review for discordant cases

There was no effective agreement between TCGA, Pathologist 1, and Pathologist 2 among the discordant cases (4-tiered grading: Fleiss' kappa = -0.23; 2-tiered grading: Fleiss' kappa = -0.33). When comparing between TCGA and Pathologist 2, there was no effective agreement (4-tiered grading: frequency of agreement = 0.33, Cohen's kappa = -0.14; 2-tiered grading: frequency of agreement = 0.39, Cohen's kappa = -0.19). Despite assessing the same representative ROIs, the agreement between Pathologist 1 and Pathologist 2 was poor for 4-tiered grading (frequency of agreement = 0.48, Cohen's kappa = 0.11) and slightly improved for 2-tiered grading (frequency of agreement = 0.61, Cohen's kappa = 0.20). Discordant cases between Pathologist 1 and Pathologist 2 were more likely to be assigned as high grade by Pathologist 2. Contingency tables between TCGA, Pathologist 1, and Pathologist 2 are in [S3 Table](#).

Grades assigned by Pathologist 2 were not associated with OS ([Table 3](#)). Further analyses were explored to determine if the incorporation of manual grade by Pathologist 2 may improve prognostic efficacy. The grades for discordant cases were re-assigned as low or high by using the most frequent grade among TCGA, Pathologist 1, and Pathologist 2, and among Pathologist 1, Pathologist 2, and the predicted grade (i.e., integrating manual and computer). Re-assigned grades were not associated with OS ($p>0.05$; [S4 Table](#)). Next, these cases were further divided into cases that did and did not agree between Pathologist 1 and Pathologist 2. Manual grades were not associated with OS in cases that did and did not agree between

Table 2. Nuclear histomics features associated with 2-tiered ccRCC grade selected in the final Lasso classification model (18 unique features; 26 total features).

Feature	Type	Biological Relevance	Color Space	Summary Function	Coefficient
Elongation	Morphology	Nuclear pleomorphism, nuclear shape (irregular)	-	MAD	-1.51E-01
Minor axis of the Ellipse Fit	Morphology	Nuclear pleomorphism, nuclear shape (irregular)	-	Median	-1.25E+00
Flatness	Morphology	Nuclear shape (irregular)	-	MAD	-4.20E-16
Kurtosis	Intensity	Uneven distribution of nucleus staining	HSV	Median	5.38E-03
Skewness	Intensity	Uneven distribution of nucleus staining	H&E	MAD	-2.22E-01
			HSV	Median	-2.87E-01
			Lab	MAD	-1.10E-01
Correlation	Texture	Granularity of chromatin (a)	HSV	MAD	-2.48E-02
Haralick Correlation	Texture	Granularity of chromatin (a)	H&E	Median	-2.24E-01
Energy	Texture	Granularity of chromatin (b)	Lab	Median	-9.36E-01
				MAD	-3.74E-01
Inverse difference moment	Texture	Granularity of chromatin (b)	H&E	Median	-4.81E-01
				MAD	-4.19E-02
			HSV	Median	1.67E-01
				MAD	-1.19E-01
Inertia	Texture	Granularity of chromatin (b)	H&E	Median	-1.77E-01
			Lab	Median	-9.72E-02
Entropy	Texture	Granularity of chromatin (c)	HSV	Median	8.97E-03
Low gray-level run emphasis	Texture	Granularity of chromatin (c)	H&E	MAD	-7.52E-02
Long run high gray-level emphasis	Texture	Granularity of chromatin (c)	HSV	MAD	1.50E-01
Long run low gray-level emphasis	Texture	Granularity of chromatin (c)	H&E	MAD	-7.71E-16
Short run high gray-level emphasis	Texture	Granularity of chromatin (c)	HSV	MAD	4.02E-02
Short run low gray-level emphasis	Texture	Granularity of chromatin (c)	H&E	MAD	-1.28E-15
Gray level non-uniformity	Texture	Granularity of chromatin (d)	HSV	Median	-4.26E-01
			Lab	MAD	2.45E+00
High gray-level run emphasis	Texture	Granularity of chromatin (d)	HSV	MAD	6.21E-03

a) Correlation is a co-occurrence based texture feature, describing roughness and repeated direction inside the nuclei.

b) Co-occurrence based texture feature, describing roughness inside the nuclei.

c) Run-length matrix based texture feature, describing randomness of gray-level distribution.

d) Run-length matrix based texture feature, describing coarseness inside nuclei.

MAD: median absolute deviation; Lab: Lab color space; HSV: hue-saturation-value color space; H&E: Hematoxylin and Eosin color space. Median and MAD were used to summarize the data extracted at the patch level to the region of interest (ROI) level.

<https://doi.org/10.1371/journal.pone.0222641.t002>

Pathologist 1 and Pathologist 2 ($p > 0.05$; Table 4). Predicted grade was only associated with OS in cases that agreed between Pathologist 1 and Pathologist 2 (Table 4). S1 File contains the manual and predicted grades of these ccRCC cases.

Discussion

This study utilized the large and diverse TCGA ccRCC dataset to extract quantitative histomics features from ROIs and applied a Lasso regression model to develop an automated 2-tiered grading system using 18 unique features (26 total features) which achieved an ROC AUC of 0.84. Using discordant cases as an independent validation set, our data-driven system stratified ccRCC cases into low and high grades that were significantly associated with OS. The

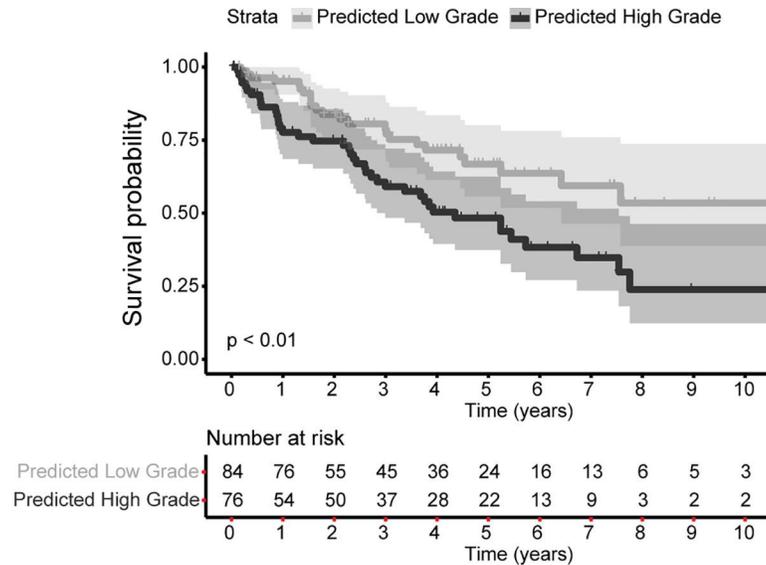


Fig 6. Prognostic efficacy of predicted grades. Cases predicted as high grade have significantly poorer overall survival rates compared to cases predicted as low grade in the extended test set (hazard ratio 2.07, 95% confidence interval of 1.25–3.43, $p < 0.01$; 65 death events among 160 cases). The shaded areas reflect the 95% confidence interval for high or low grade.

<https://doi.org/10.1371/journal.pone.0222641.g006>

prognostic efficacy of predicted grades in the discordant cases outperformed the manual grades assessed by TCGA, Pathologist 1, and Pathologist 2. This proof-of-concept study demonstrated the potential of computational pathology to predict ccRCC grades via a more objective and quantitative pipeline, as well as addressed the issue of grade disagreement commonly encountered between pathologists.

The grading of ccRCC is highly challenging and subjective, but the accurate assignment of ccRCC grade is important for clinical care and follow-up. Research groups, specifically Yeh and colleagues [30], Kruk and colleagues [31], and Holdbrook and colleagues [32], have been actively developing computational pathology systems to provide objectivity and/or automate ccRCC grading. Each computational system is highly unique with differences in image processing, feature extraction, classification method, and predicting 2 or 4-tiered grades. We utilized an unbiased data-driven approach where we extracted a set of high dimensional nuclear features ($n = 144$), and used Lasso, a machine learning-based method, to build our final predictive model. This is different from Yeh et al. [30] who only evaluated 1 feature (i.e., maximum nuclei size) to predict 2-tiered grade, Kruk et al. [31] who pre-selected features (out of 31 features) prior to building the final model to predict 4-tiered grade, and Holdbrook et al. who used up to 4 concatenate feature vectors to calculate fraction value scores prior to classification into low or high grade [32]. In addition, our Lasso regression allowed us to identify the 18 unique histomics biomarkers in our final predictive model while the features in the models by Kruk et al [31] and Holdbrook et al. [32] are unknown. Our 18 features provided information about the nucleus, the uneven distribution of nucleus staining, and the granularity of chromatin and nucleoli, highlighting that the addition of computer textual and intensity-related features to traditional pathology morphological features can improve the ability to predict ccRCC grade. We and Holdbrook et al [32] demonstrated that our predicted grade had prognostic significance whilst the studies by Kruk et al [31] and Yeh et al [30] did not report if their grade was associated with prognosis. Lastly, our system was trained using a much larger and more diverse dataset of 277 cases from seven TCGA participating institutions, and we validated our

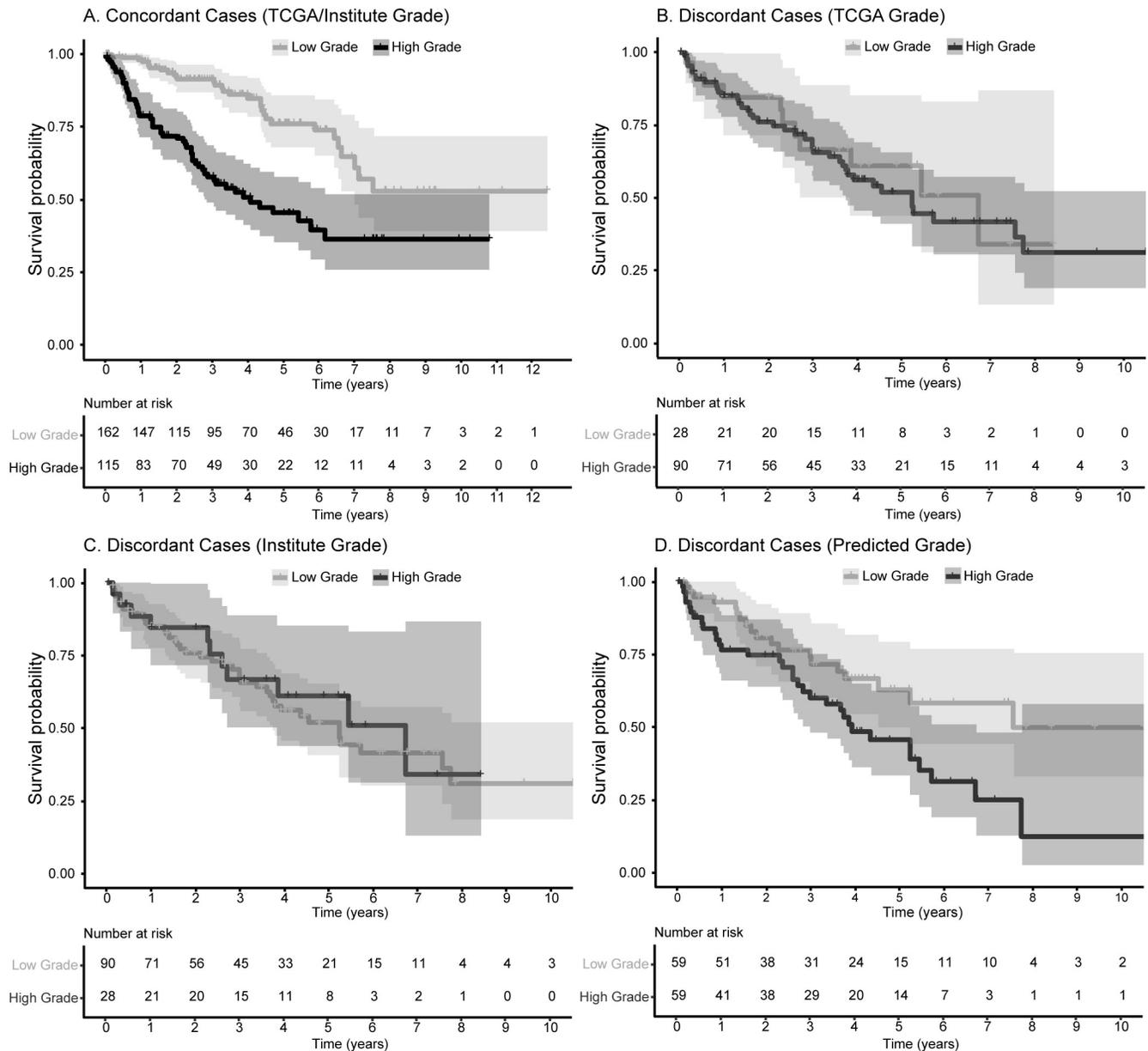


Fig 7. Kaplan-Meier curves comparing manual and predicted grades with overall survival in concordant and discordant cases. Grades assigned by TCGA/Pathologist 1 were significantly associated with overall survival within the concordant cases (A). In the discordant set, neither grades assigned by TCGA (B) nor Pathologist 1 (C) were associated with overall survival while predicted grade remained significantly prognostic (D). Please refer to Table 3 for hazard ratios and 95% confidence intervals for each analysis. The shaded areas reflect the 95% confidence interval for high or low grade.

<https://doi.org/10.1371/journal.pone.0222641.g007>

system using 160 cases. This is in contrast to those three studies which used small numbers for training ($n = 38$ to 70) and validation ($n = 6$ to 62), and obtaining their cases from a single institution. Collectively, our work and others are substantial efforts to improve ccRCC grading. Each computational method will require further refinement and validation before their clinical utility can be determined.

Each TCGA grade is the consensus of at least two pathologists. One reason for grade disagreement between TCGA and Pathologist 1 can be explained by TCGA pathologists assessing different ROIs than the representative ROIs selected in our study. However, even when

Table 3. The association of manual or computer predicted 2-tiered grade with overall survival in the concordant and discordant cases.

	Manual Grade			Computer Predicted Grade		
	Hazard Ratio	(95% CI)	p-value	Hazard Ratio	(95% CI)	p-value
A. Concordant cases between TCGA and Pathologist 1 (85 events out of 277 cases)						
Model A: Crude	3.12	(2.00, 4.86)	<0.01	NA	NA	NA
Model B: Adjusted for Age and Gender	3.00	(1.91, 4.71)	<0.01	NA	NA	NA
Model C: Adjusted for Age, Gender, and Stage	1.59	(0.99, 2.57)	0.06	NA	NA	NA
B. Discordant cases between TCGA and Pathologist 1 (52 events out of 118 cases)						
Grade assigned by TCGA						
Model A: Crude	1.16	(0.59, 2.25)	0.67	2.01	(1.14, 3.54)	0.02
Model B: Adjusted for Age and Gender	1.09	(0.56, 2.13)	0.80	2.31	(1.26, 4.24)	< 0.01
Model C: Adjusted for Age, Gender, and Stage	1.08	(0.55, 2.11)	0.83	1.83	(0.98, 3.41)	0.06
Grade assigned by Pathologist 1						
Model A: Crude	0.86	(0.44, 1.68)	0.67	NA	NA	NA
Model B: Adjusted for Age and Gender	0.92	(0.47, 1.79)	0.80	NA	NA	NA
Model C: Adjusted for Age, Gender, and Stage	0.93	(0.47, 1.82)	0.83	NA	NA	NA
Grade assigned by Pathologist 2						
Model A: Crude	1.09	(0.63, 1.89)	0.75	NA	NA	NA
Model B: Adjusted for Age and Gender	1.21	(0.70, 2.10)	0.50	NA	NA	NA
Model C: Adjusted for Age, Gender, and Stage	1.15	(0.66, 2.00)	0.62	NA	NA	NA

Confidence Interval, CI

<https://doi.org/10.1371/journal.pone.0222641.t003>

reviewing the same ROIs for discordant cases, there was very poor agreement between Pathologist 1 and Pathologist 2, reiterating the challenges of ccRCC grading. These discordant cases could be more diagnostically challenging or ambiguous. Since manual grades for concordant cases were significantly associated with OS, it could be argued that concordant cases were diagnostically less challenging where the tumors were overwhelmingly of a low or high grade, and

Table 4. The association of manual or computer predicted 2-tiered grade with overall survival in 118 discordant cases.

	Manual Grade			Computer Predicted Grade		
	Hazard Ratio	(95% CI)	p-value	Hazard Ratio	(95% CI)	p-value
A. Cases with identical grades between Pathologist 1 and Pathologist 2 (31 events out of 76 cases)						
Model A: Crude	0.99	(0.44, 2.23)	0.99	2.05	(1.00, 4.21)	0.05
Model B: Adjusted for Age and Gender	1.04	(0.46, 2.34)	0.93	2.42	(1.13, 5.20)	0.02
Model C: Adjusted for Age, Gender, and Stage	1.18	(0.52, 2.68)	0.69	1.89	(0.87, 4.12)	0.11
B. Cases with different grades between Pathologist 1 and Pathologist 2 (21 events out of 46 cases)						
Grade assigned by Pathologist 1						
Model A: Crude	0.64	(0.19, 2.20)	0.48	2.49	(0.83, 7.45)	0.10
Model B: Adjusted for Age and Gender	0.63	(0.18, 2.17)	0.46	2.49	(0.72, 7.28)	0.16
Model C: Adjusted for Age, Gender, and Stage	0.59	(0.17, 2.03)	0.40	2.03	(0.62, 6.66)	0.24
Grade assigned by Pathologist 2						
Model A: Crude	1.56	(0.46, 5.31)	0.48	NA	NA	NA
Model B: Adjusted for Age and Gender	1.58	(0.46, 5.44)	0.46	NA	NA	NA
Model C: Adjusted for Age, Gender, and Stage	1.70	(0.49, 5.89)	0.40	NA	NA	NA

Confidence Interval, CI

<https://doi.org/10.1371/journal.pone.0222641.t004>

that our model was trained using more homogeneous ROIs. Predicted grades for discordant cases were significantly associated with survival, in contrast to manual assessments or using the most frequent manual grade. Therefore, our automated system has the ability to diagnose a range of ccRCC cases with consistency and objectivity. In practical application, such computational system could be useful as a tool to provide a second-opinion in diagnostically ambiguous cases for pathologists.

Our study has some limitations. We did not use the WHO/ISUP grading system because the TCGA participating medical centers used the Fuhrman's system. However, since our computer system was constructed based on computer extracted nuclear features, it can be adapted to predict WHO/ISUP grades which also utilize nuclei/nucleoli features in the future. There are inherent limitations of reviewing cases using WSIs. Accurate grading may be hindered by the quality of WSIs and the lack of the Z-axis [33]. Our study reviewed diagnostic WSIs and analyzed manually selected ROIs that may not be representative of the entire tumor. For future work, automating ROI detection and grade prediction will allow the review of multiple tumor sections more efficiently. Lastly, our nuclei segmentation relied on conventional image analysis techniques. While qualitative evaluation of the segmentation results revealed that our image processing pipeline produced reasonably good results, the nuclei segmentation may not be optimal in more challenging cases. A solution is to employ deep learning based techniques to improve nuclei segmentation in future studies [30,34,35].

Conclusions

We developed an automated 2-tiered Fuhrman's grading system with prognostic significance. Our system demonstrated the potential of computational pathology to improve the reproducibility in the diagnosis and grading of ccRCC, and to aid the clinical management of ccRCC patients. Future work may include adapting our computational system to predict WHO/ISUP grades; validating our system on other ccRCC cohorts; using deep learning methods to detect ROIs, segment nuclei and predict grade; and exploring whether histomics features can predict prognosis independently of grade. This work is one step toward developing an artificial intelligence system for diagnostic pathology.

Supporting information

S1 Table. The average area under the receiver-operator characteristic curves (AUC ROC) for each machine learning method using the training set after 100 iterations of random 10% hold out. These methods were implemented using the glmnet and caret packages in R. (DOCX)

S2 Table. The association of the nine nuclear morphological features with Fuhrman's grade in the 277 concordant cases (using the 2-tiered grading system). (DOCX)

S3 Table. Contingency tables of Fuhrman's grade between TCGA and Pathologist 1 with Pathologist 2 among the 118 discordant cases. (DOCX)

S4 Table. The association of re-assigned grades for discordant cases. (DOCX)

S1 Fig. Agreement between TCGA and Pathologist 1. (A) The agreement between the TCGA and Pathologist 1 using the 4-tiered grading was poor (frequency of agreement = 0.47, Cohen's kappa = 0.20). (B) The agreement improved to moderate when using the 2-tiered

grading system (frequency of agreement = 0.70, Cohen's kappa = 0.41).
(PDF)

S2 Fig. These plots contain the 277 cases that were concordant between TCGA and Pathologist 1 using the 2-tiered grading system. Nine morphological features (**1**: Area, **2**: Roundness, **3**: Elongation, **4**: Flatness, **5**: Perimeter, **6**: Equivalent Spherical Perimeter, **7**: Equivalent Spherical Radius, **8**: Minor Axis of the Ellipse Fit, and **9**: Nuclei Major Axis of the Ellipse Fit) were plotted with their median and median absolute deviation values. Each feature and measure was stratified either by the 4-tiered (**A and B**) or 2-tiered (**C and D**) grades assigned by Pathologist 1, respectively.
(PDF)

S1 File. Patient data with manual and predicted grades.
(CSV)

Acknowledgments

The data used in this study were in whole or in part based on the data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

Author Contributions

Conceptualization: Humayun Irshad, Yujing J. Heng.

Data curation: Christopher A. Rubadue, Douglas I. Lin, Humayun Irshad.

Formal analysis: Katherine Tian, Yujing J. Heng.

Investigation: Katherine Tian, Christopher A. Rubadue, Douglas I. Lin, Mitko Veta, Humayun Irshad, Yujing J. Heng.

Methodology: Katherine Tian, Mitko Veta, Humayun Irshad, Yujing J. Heng.

Project administration: Yujing J. Heng.

Resources: Michael E. Pyle, Humayun Irshad, Yujing J. Heng.

Supervision: Humayun Irshad, Yujing J. Heng.

Writing – original draft: Katherine Tian, Christopher A. Rubadue, Douglas I. Lin, Michael E. Pyle, Humayun Irshad, Yujing J. Heng.

Writing – review & editing: Katherine Tian, Christopher A. Rubadue, Douglas I. Lin, Mitko Veta, Michael E. Pyle, Humayun Irshad, Yujing J. Heng.

References

1. Amin MB, Amin MB, Tamboli P, Javidan J, Stricker H, de-Peralta Venturina M, et al. Prognostic impact of histologic subtyping of adult renal epithelial neoplasms: an experience of 405 cases. *Am J Surg Pathol.* 2002; 26: 281–291. <https://doi.org/10.1097/00000478-200203000-00001> PMID: 11859199
2. Goyal R, Gersbach E, Yang XJ, Rohan SM. Differential diagnosis of renal tumors with clear cytoplasm: Clinical relevance of renal tumor subclassification in the era of targeted therapies and personalized medicine. *Arch Pathol Lab Med.* 2013; 137: 467–480. <https://doi.org/10.5858/arpa.2012-0085-RA> PMID: 23544936
3. de Peralta-Venturina M, Moch H, Amin M, Tamboli P, Hailemariam S, Mihatsch M, et al. Sarcomatoid differentiation in renal cell carcinoma: a study of 101 cases. *Am J Surg Pathol.* 2001; 25: 275–284. <https://doi.org/10.1097/00000478-200103000-00001> PMID: 11224597
4. Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. *Lancet.* 2009; 373: 1119–1132. [https://doi.org/10.1016/S0140-6736\(09\)60229-4](https://doi.org/10.1016/S0140-6736(09)60229-4) PMID: 19269025

5. Foster K, Prowse a, van den Berg a, Fleming S, Hulsbeek MM, Crossey P a, et al. Somatic mutations of the von Hippel-Lindau disease tumour suppressor gene in non-familial clear cell renal carcinoma. *Hum Mol Genet.* 1994; 3: 2169–73. <https://doi.org/10.1093/hmg/3.12.2169> PMID: 7881415
6. Delahunt B. Advances and controversies in grading and staging of renal cell carcinoma. *Mod Pathol.* 2009; 22: S24–S36. <https://doi.org/10.1038/modpathol.2008.183> PMID: 19494851
7. Lang H, Lindner V, de Fromont M, Molinié V, Letourneux H, Meyer N, et al. Multicenter determination of optimal interobserver agreement using the Fuhrman grading system for renal cell carcinoma. *Cancer.* 2005; 103: 625–629. <https://doi.org/10.1002/cncr.20812> PMID: 15611969
8. Hong SK, Jeong CW, Park JH, Kim HS, Kwak C, Choe G, et al. Application of simplified Fuhrman grading system in clear-cell renal cell carcinoma. *BJU Int.* 2011; 107: 409–415. <https://doi.org/10.1111/j.1464-410X.2010.09561.x> PMID: 20804473
9. Rioux-Leclercq N, Karakiewicz PI, Trinh QD, Ficarra V, Cindolo L, De La Taille A, et al. Prognostic ability of simplified nuclear grading of renal cell carcinoma. *Cancer.* 2007; 109: 868–874. <https://doi.org/10.1002/cncr.22463> PMID: 17262800
10. Delahunt B, Cheville JC, Martignoni G, Humphrey PA, Magi-Galluzzi C, McKenney J, et al. The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. *Am J Surg Pathol.* 2013; 37: 1490–1504. <https://doi.org/10.1097/PAS.0b013e318299f0fb> PMID: 24025520
11. Louis DN, Feldman M, Carter AB, Dighe AS, Pfeifer JD, Bry L, et al. Computational pathology: A path ahead. *Arch Pathol Lab Med.* 2016; 140: 41–50. <https://doi.org/10.5858/arpa.2015-0093-SA> PMID: 26098131
12. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med.* 2011; 3: 108ra113. <https://doi.org/10.1126/scitranslmed.3002564> PMID: 22072638
13. Ehteshami Bejnordi B, Veta M, van Diest PJ, 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: 2199–2210. PMID: PMC5820737. <https://doi.org/10.1001/jama.2017.14585> PMID: 29234806
14. Dong F, Irshad H, Oh EY, Lerwill MF, Brachtel EF, Jones NC, et al. Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. *PLoS One.* 2014; 9: e114885. <https://doi.org/10.1371/journal.pone.0114885> PMID: 25490766
15. Tabesh A, Teverovskiy M, Pang HY, Kumar VP, Verbel D, Kotsianti A, et al. Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Trans Med Imaging.* 2007; 26: 1366–1378. <https://doi.org/10.1109/TMI.2007.898536> PMID: 17948727
16. 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. <https://doi.org/10.1038/ncomms12474> PMID: 27527408
17. Schüffler PJ, Fuchs TJ, Ong CS, Roth V, Buhmann JM. Computational TMA analysis and cell nucleus classification of renal cell carcinoma. In: Goesele M, Roth S, Kuijper A, Schiele B, Schindler K, editors. *Pattern Recognition DAGM 2010 Lecture Notes in Computer Science*, vol 6376. Berlin, Heidelberg: Springer; 2010. pp. 202–211. https://doi.org/10.1007/978-3-642-15986-2_21
18. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature.* 2013; 499: 43–49. <https://doi.org/10.1038/nature12222> PMID: 23792563
19. Gutman D, Cobb J, Somanna D. Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. *J Am Med Inf Assoc.* 2013; 20: 1091–1098. <https://doi.org/10.1136/amiajnl-2012-001469> PMID: 23893318
20. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: An open source platform for biological image analysis. *Nat Methods.* 2012; 9: 676–682. <https://doi.org/10.1038/nmeth.2019> PMID: 22743772
21. Khan AM, Rajpoot N, Treanor D, Magee D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans Biomed Eng.* 2014; 61: 1729–1738. <https://doi.org/10.1109/TBME.2014.2303294> PMID: 24845283
22. Ruifrok AC, Katz RL, Johnston DA. Comparison of quantification of histochemical staining by hue-saturation-intensity (HSI) transformation and color-deconvolution. *Appl Immunohistochem Mol Morphol.* 2003; 11: 85–91. <https://doi.org/10.1097/00129039-200303000-00014> PMID: 12610362
23. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern.* 1973; SMC-3: 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
24. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process.* 1975; 4: 172–179. [https://doi.org/10.1016/S0146-664X\(75\)80008-6](https://doi.org/10.1016/S0146-664X(75)80008-6)

25. Irshad H, Veillard A, Roux L, Racoceanu D. Methods for nuclei detection, segmentation, and classification in digital histopathology: A review-current status and future potential. *IEEE Rev Biomed Eng.* 2014; 7: 97–114. <https://doi.org/10.1109/RBME.2013.2295804> PMID: 24802905
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010; 33: 1–22. <https://doi.org/10.1359/JBMR.0301229> PMID: 20808728
27. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* 2008; 28: 1–26. <https://doi.org/10.18637/jss.v028.i07>
28. Tibshirani R. Regression shrinkage and selection via the lasso: A retrospective. *J R Stat Soc Ser B Stat Methodol.* 2011; 73: 273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
29. Therneau T, Grambsch P. Modeling survival data: extending the Cox model. *Technometrics.* 2002; 44: 85–86. <https://doi.org/10.1198/tech.2002.s656>
30. Yeh FC, Parwani A, Pantanowitz L, Ho C. Automated grading of renal cell carcinoma using whole slide imaging. *J Pathol Inform.* 2014; 5: 23. <https://doi.org/10.4103/2153-3539.137726> PMID: 25191622
31. Kruk M, Kurek J, Osowski S, Koktysz R, Swiderski B, Markiewicz T. Ensemble of classifiers and wavelet transformation for improved recognition of Fuhrman grading in clear-cell renal carcinoma. *Biocybern Biomed Eng.* 2017; 37: 357–364. <https://doi.org/10.1016/j.bbe.2017.04.005>
32. Holdbrook DA, Singh M, Choudhury Y, Kalaw EM, Koh V, Tan HS, et al. Automated Renal Cancer Grading Using Nuclear Pleomorphic Patterns. *JCO Clin Cancer Informatics.* 2018; 1–12. <https://doi.org/10.1200/cci.17.00100> PMID: 30652593
33. Heng YJ, Lester SC, Tse GMK, Factor RE, Allison KH, Collins LC, et al. The molecular basis of breast cancer pathological phenotypes. *J Pathol.* 2017; 241: 375–391. <https://doi.org/10.1002/path.4847> PMID: 27861902
34. Hatipoglu N, Bilgin G. Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships. *Med Biol Eng Comput.* 2017; 55: 1829–1848. <https://doi.org/10.1007/s11517-017-1630-1> PMID: 28247185
35. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans Med Imaging.* 2017; 36: 1550–1560. <https://doi.org/10.1109/TMI.2017.2677499> PMID: 28287963