EasyChair Preprint

№ 10094

Predicting Papillary Renal Cell Carcinoma Prognosis Using Integrative Analysis of Histopathological Images and Genomic Data

Shaira Kee, Michael Aaron Sy, Samuel Border, Nicholas Lucarelli, Akshita Gupta, Pinaki Sarder, Marvin Masalunga and Myles Joshua Tan

May 12, 2023

Predicting Papillary Renal Cell Carcinoma Prognosis Using Integrative Analysis of Histopathological Images and Genomic Data

Shaira L. Kee[1[0000-0002-4330-4595]], Michael Aaron G. Sy[1,2[0000−0001−9281−7592]], Samuel P. Border[3,4[0000−0000−0000−0000]], Nicholas J. Lucarelli[3,4[0000−0000−0000−0000]], Akshita Gupta[4,5[0009−0002−7440−5003]], Pinaki Sarder[3,4,6[0000−0003−2450−5233]], Marvin C. Masalunga[7[0000-0002-7718-1300]], Myles Joshua T. Tan[1,2,4,6[0000−0002−1426−6526]]

[1] Department of Natural Sciences, University of St. La Salle, Bacolod, Philippines
[2] Department of Electronics Engineering, University of St. La Salle, Bacolod, Philippines
[3] J. Crayton Pruitt Family Department of Biomedical Engineering, Herbert Wertheim College of Engineering, University of Florida, Gainesville, FL, United States
[4] Division of Nephrology, Hypertension and Renal Transplantation – Quantitative Health Section, Department of Medicine, College of Medicine, University of Florida
[5] Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States
[6] Department of Electrical and Computer Engineering, Herbert Wertheim College of Engineering, University of Florida, Gainesville, FL, United States
mylesjoshua.tan@medicine.ufl.edu
[7] Medical Education Unit, School of Medicine, Southwestern University - PHINMA, Cebu, Philippines
mcmasalunga.swu@phinmaed.com

Abstract. Renal cell carcinoma (RCC) is a common malignant tumor of the adult kidney, with the papillary subtype (pRCC) as the second most frequent. There is a need to improve evaluative criteria for pRCC due to overlapping diagnostic characteristics in RCC subtypes. To create a better prognostic model for pRCC, we proposed an integration of morphologic and genomic features. Matched images and genomic data from The Cancer Genome Atlas were used. Image features were extracted using CellProfiler, and prognostic image features were selected using least absolute shrinkage and selection operator and support vector machine algorithms. Eigengene modules were identified using weighted gene co-expression network analysis. Risk groups based on prognostic features were significantly distinct ($p < 0.05$) according to Kaplan-Meier analysis and log-rank test results. We used two image features and nine eigengene modules to construct a model with the Random Survival Forest method, measuring 11-, 16-, and 20-month areas under the curve (AUC) of a time-dependent receiver operating curve. The integrative model (AUCs: 0.877, 0.769, and 0.811) outperformed models trained with eigengenes alone (AUCs: 0.793, 0.748, and 0.772) and morphological features alone (AUCs: 0.547, 0.497, 0.483). This suggests that an integrative prognostic model based on histopathological images and genomic features could significantly improve survival prediction for pRCC patients and assist in clinical decision-making.

Keywords: Renal cell carcinoma (RCC), prognostic model, genomic features, histopathological images, The Cancer Genome Atlas.

1 Introduction

In the past decade, localized renal cell carcinoma (RCC) has increased in prevalence and incidence substantially [1]. It is the most common malignant neoplasm arising from the adult kidney [2] and is responsible for ~95% of all cases [3]. RCCs are malignant tumors of the renal cortex displaying distinct clinical, morphologic, and genetic characterizations [4], [5]. Currently, there are 20 different RCC variants [6]. Classically, the Heidelberg classification system categorizes RCC into the following histologic subtypes: clear cell, papillary, chromophone, collecting duct, and unclassified RCC [7]. Papillary renal cell carcinoma (pRCC) is the second most commonly identified subtype of RCC (10%-15% of cases), following the clear cell subtype (ccRCC) [8]. pRCC is distinguished from ccRCC morphologically by the presence of basophilic or eosinophilic cells in a papillary or tubular form [9]. pRCC compared to ccRCC, has been reported with a greater male predominance [9]. Clear cell papillary renal cell carcinoma (ccpRCC) is a distinct histotype that progresses in a more indolent manner. [4]. pRCC tumors possess immunohistochemical and genetic profiles that are distinct from ccRCC and ccpRCC [9]. While frontline surgical extirpation of suspected malignant localized RCC remains the standard of care, recent increases in renal mass biopsies for risk stratification indicate a growing preference of both patients and surgeons for the characterization of tumor at the outset to inform treatment decisions. Precise pathologic information coupled with emerging molecular tools remains the best course of pretreatment risk stratification [3].

Pathologists use immunohistochemical staining to increase contrast between specific cell types in biopsy specimens for tumor evaluation [10]. With the growing digitization of Whole-Slide Images (WSIs), computational image analysis has shown great potential in diagnosis and discovery of new biomarkers for multiple types of cancers, such as breast [11], colon [12], and lung [13]. Morphological interpretation of histologic images is the basis of pathological evaluation. WSIs are a rich source of biological information as this level of resolution facilitates detailed assessment of the relationship of cancer cells with other cells and tumor microenvironment (TME), all of which are referred to as “hallmarks of cancer” [14]. Accurate and reproducible models can be made to assess prognosis through automatically quantifying morphological features. A pipeline is developed which automatically segments cancer images and generates quantitative features [15]. The majority of previous work in quantitative pathology has required laborious image annotation by skilled pathologists, which is prone to human error [16]. Automation using image analysis and ML methods has significantly corrected inconsistencies that result from histologic preparation [11], [12], as even expert human eyes have difficulty in distinguishing some granular image features [17].

In addition to histopathologic images, information on molecular alteration has also been widely adopted for predicting cancer clinical outcomes [18], [19]. Cancers are diverse, with varying genetics, phenotypes, outcomes and subtypes. In the past decades, molecular stratification of tumors using gene expression microarrays has

been considered an important field in cancer research [20], with an increase in application of integrative genomics or panomics approaches [21], [22]. This approach is used to identify biomarkers for stratification of patients into groups with different clinical outcomes. Solid tumors are heterogeneous tissues composed of a mixture of cancer and normal cells that further complicates the interpretation of their molecular profiles [23]. Cancer, immune, and stromal cells form an integral part of the TME; however, their admixture poses challenges for molecular assays, especially in large scale analyses [23]. Alternatively, Molecular profiles yield quantifiable results via computational and statistical inferences on big data. For instance, studies have shown that lymphocytic infiltration can be inferred from gene expression profiles [24], as well as cellularity from single-nucleotide polymorphism (SNP) information [25]. However, these approaches are indirect and strongly rely on statistical assumptions [26]. To leverage the richness of histopathological information and quantitative results of computational analyses, a systematic approach that integrates histopathology and genomics was developed. An image-based approach was used to increase the power of molecular assays and to complement them with each other to unveil prognostic features otherwise invisible in molecular data [2]. Recent studies have also highlighted the contribution of gene expression and morphologic phenotypes to cancer growth and progression [11], [27]. The integration extends recent approaches that only identified morphological features prediction of survival by image analysis [11]. One study ventured into an integrative approach for predicting ccRCC prognosis or time-to-event outcomes [2]. The study utilized both histopathologic images and eigengenes to predict patient outcome. Eigengenes are the summarized genetic module expression profile used to reduce a gene co-expression network involving thousands of genes [28]. The model developed from the study was able to generate risk indices that correlated strongly with survival of ccRCC patients, outperforming predictions based on morphologic features or eigengenes separately. Cheng et al. [2] showed significant correlation according to Bonferroni correction between image features and eigengenes, with results suggesting that low expression of some eigengene were related to poor prognostic outcome, implying impaired renal function; while high expression of some eigengene observed to coexpress in multiple types of cancers [29] indicated aggressiveness and was negatively related to patient prognosis [2]. The integration of multi-modal data for prognosis estimation has led to new insights into the influence of tissue genotype on phenotype [30], [31].

Random Survival Forest (RSF) is an ensemble tree-based machine learning (ML) model for the analysis of survival data. The method has been used in several studies showing superior predictive performance to traditional strategies in low and high-dimensional settings [32], [33]. RSF can capture complex relationships between predictors and survival without prior specification requirements. Risk prediction models play a vital role in personalized decision-making, especially for time-to-event outcomes of cancer patients [33]. Traditional prediction models often ignore the longitudinal nature of medical records, using only baseline information. The use of risk prediction models incorporates longitudinal information to produce updated survival predictions during follow-up check-ups. This allows for high accuracy in the automated prediction of cancer prognosis, showing significant promise in improving the quality of care where pathologists are scarce. This study aims to overcome the

difficulty of sub-classifying pRCC using morphological features alone and contribute to better prognosis methods for pRCC patients.

2 Materials and Methods

2.1 Dataset Overview

The pRCC patient samples used in the study include matched hematoxylin and eosin (H&E)-stained WSIs, transcriptome, somatic mutation, and clinical information. The patient data were acquired from the The Cancer Genome Atlas (TCGA) data portal of the National Cancer Institute Genomic Data Commons. Microscopic images (20× and 40× magnification) were obtained from the TCGA. Gene expression profiles with 20,531 Entrez-identified sequences were taken from Broad GDAC Firehose. Prior to preprocessing, the dataset included 287 pRCC cases. Patients with missing data were excluded from the study, leaving us with 277 patients.

2.2 Histopathological Image Processing

Image Preprocessing and Image Feature Extraction. WSIs were first chopped into equal-sized patches (1000×1000×3) using Openslide. The resulting image patches were then assessed for tissue content using pixel intensity statistics. The generated sub-images were then filtered according to quality. Sub-images with >50% white background were excluded. To do so, a range value for the whole dataset was determined, where images from one randomly chosen patient were grouped according to the cell coverage of their image patches. The mean and standard deviation (SD) of the pixel value of each image were calculated, and were averaged to determine the pixel value range for each group. Then, 20 sub-images were randomly selected for the next step to eliminate sample selection bias and reduce computing load. Since there were varying sources of the WSIs, the images were color normalized to prepare for downstream analysis. Macenko color normalization [34] was used. Then,, H&E stains were separated from the original images to facilitate nuclear extraction for downstream feature calculation. Quantitative performance evaluation of nuclei segmentation was performed using manually annotated subset of image patches.

CellProfiler version 4.2.1 [35], [36] software was used to extract image features from each sub-image. Individual cell segmentation was conducted using builtin modules in CellProfiler and included Otsu Thresholding followed by morphological postprocessing. Here, we measured object intensity, object size, object shape, image granularity, and image texture. The measurement of object size and shape includes features such as area, Zernike shape, perimeter, formfactor, solidity, Euler's number, and orientation. The Zernike shape features comprise a set of 30 shape characteristics that are derived from Zernike polynomials ranging from order 0 to order 9. Finally, we extracted 448 image features from each sub-image and calculated the mean value of the 20 representative sub-images for each patient.

Feature Elimination. To obtain prognosis-related features, an R implementation of the support vector machine's recursive feature elimination (SVM-RFE) and Lasso Regression algorithms were employed to filter the prognostic image features most correlated with pRCC prognosis. A 5-fold cross validation in both the SVM-RFE and

Lasso Regression algorithms were applied. A total of 441 image features were filtered using the SVM-RFE and Lasso Regression. Feature reduction pipeline was built using the glmnet and caret libraries. Data for this model was split using a 70:30 ratio. The image features were also filtered using the e1071 multiple SVM-RFE (mSVM-RFE) [37] as adapted from https://tinyurl.com/3ceynuh9. SVM-RFE is a ML method with a backward feature SVM. Feature filtration using SVM-RFE was performed using a method performed by Li et al. [35]. The SVM-RFE was used to rank the pathological image features in a descending order of significance, iteratively eliminating the minimum features and training the model with the remaining image features until all features are removed. The maximal cross-validated accuracy was adopted as the evaluation index to select the optimal feature subset related to the prognosis. Finally, intersection of the optimal subset of features from the SVM-RFE model and LASSO regression was used to obtain the most relevant pathological features to the prognosis.

2.3 Gene Coexpression Analysis

The profiles of mRNA expression for the pRCC tumors in TCGA were transformed from Illumina HiSeq 2000 RNA-seq V2 read counts to normalized transcripts per million (TPM). TPM fulfills the invariant average criterion that RPKM does not account for. By definition, TPM and RPKM are proportional, thus are closely related as shown by this equation:

$$\text{TPM} = 10^6 \times \frac{RPKM}{Sum(RPKM)} \quad (1)$$

Expression data were then scaled with the natural logarithm operation [38]:

$$X_{\text{input}} = \log(X_{\text{original}} + 1) \quad (2)$$

where X_{original} was the genetic data or the non-negative RNA sequencing expression values (Illumina Hi-Seq RNA-seq v2 RSEM normalized), and X_{input} was the input covariate vector for the coexpression network analysis.

WGCNA package [39] was used to cluster genes into coexpressed modules and each module was summarized as an eigengene using the protocol described in the study [39]. This approach allows for substantial improvement in statistical power and for more focus to be placed on important biological processes and genetic variations related to coexpressed gene modules. We adopted code from https://tinyurl.com/4psu2wrw. The generated module eigengenes were then correlated with prognostic image features. Biological relevance of the module eigengenes were obtained through Metascape (https://metascape.org/).

2.4 Risk Categorization

Each prognostic feature was correlated with survival status to determine the direction of relationship. For each prognostic feature, we divided the patients into two groups (low and high-risk groups) where the median of each prognostic feature was used as a cut-off point. Depending on the relationship and feature values relative to the median, prognostic features of each patient were categorized as high or low risks. Finally, we assigned a patient's risk level based on the dominant risk level among their features.

2.5 Prognosis Prediction Model

RSF algorithm was used to construct an integrative prognostic model. The method will be implemented using the R package randomForestSRC, tailoring based on the different ensemble parameters that affect building of survival trees. The number of splits in each candidate variable can reduce computation time compared to testing all possible split points for each covariate. A 7:3 ratio and 10-fold cross validation was used during model development. The RF model estimated survival risk for each patient and determined their risk scores, which was plotted on a survival curve. As predicted by the learned model, we compared 11-month survival differences between the two groups using Kaplan-Meier (KM) analysis and log-rank test. KM analysis was used for patient stratification, while the log-rank test calculated the *p*-value, where $p < 0.05$ is considered significant.

3 Results

3.1 Patient Characteristics

277 pRCC patients (72 female and 205 male) were included. Histopathological images, mRNA expression data, and clinicopathological information were downloaded from TCGA, Broad GDAC Firehose, and cBioportal.

Table 1. Demographic and clinical characteristics of pRCC patients.

Characteristics		Total (*n* = 277)	Train (*n* = 194)	Test (*n* = 83)
Gender	Male	205	141	64
	Female	72	53	19
Events	Alive	235	164	71
	Dead	42	30	12

3.2 Prognosis-related Image Features and Co-expression Gene Module Selection

441 image features were used for the data dimension algorithms used, specifically LASSO and mSVM-RFE. The optimal subset of features determined by the feature elimination of the mSVM-RFE algorithm obtained six features while LASSO regression identified 20. We then found the intersection of the results of the two algorithms to obtain two pRCC prognostic image features (Image Granularity feature and Zernike shape feature). To identify the prognostic co-expression gene modules, WGCNA was applied to evaluate the relationship between the two prognostic image features and eigengene modules. The most significant positive association with the

Zernike feature was the dark red module, while it was the turquoise module for the Image Granularity feature. Examples of selected histopathological sub-images in both high-risk and low-risk groups are presented in Fig 1.

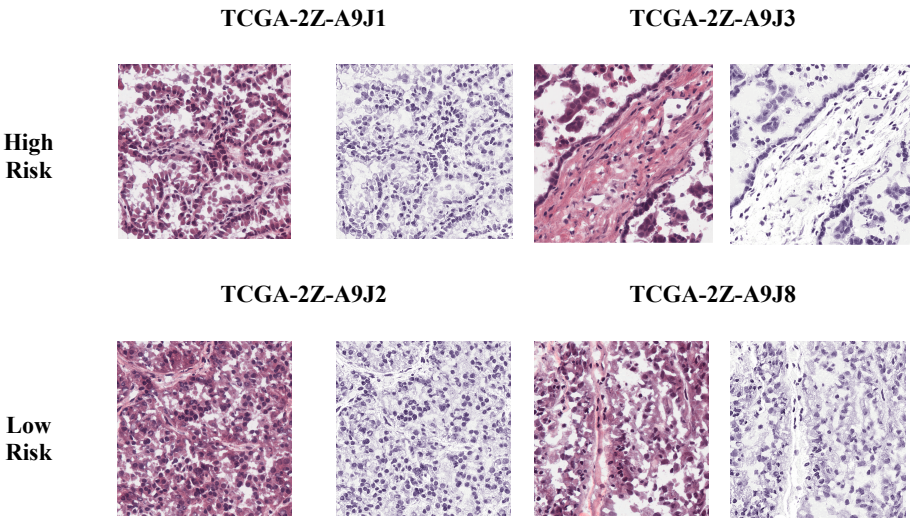

Fig. 1. H&E and eosin-stained histopathological sub-images in high- and low-risk groups

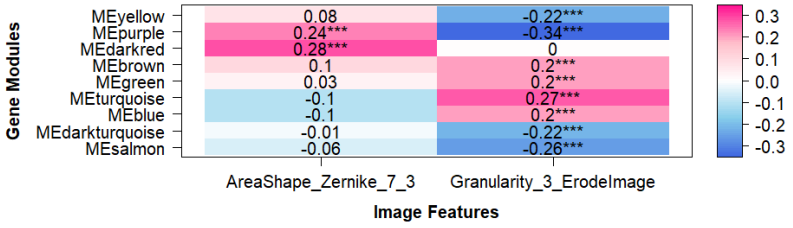

Fig. 2. Significant associations with the image features from nine modules (salmon, dark turquoise, blue, turquoise, green, brown, dark red, purple, and yellow).

3.3 Enrichment Analysis of the Key Gene Modules

There were 471 genes in the purple module, 147 genes in the dark red module, and 3500 genes in the turquoise module, with other significant genes shown in Fig. 2. These indicate that there are significant intrinsic associations among the biological functioning of the selected genes in each module. Most genes were enriched in the biological processes such as metabolic process, response to stimulus, biological regulation, developmental process, localization, growth, immune system response, signaling, and cellular process. More specific associations include NRF2 pathway, glucuronidation, cellular response to DNA damage, and VEGFA-VEGFR2 signaling pathway.

3.4 Construction and Evaluation of the Integrative Prognostic Model

The pRCC patients were randomly divided into a training (n =194) and test set (n = 83). 30 trees were used in the integrative model. The time-dependent ROC curve can better demonstrate the model's predictive ability over time, as it incorporates both survival state and survival time in the results. In the test set shown in Fig. 4, the 11-, 16-, and 20-month AUCs were 0.877, 0.769, and 0.811 respectively. The predictive accuracy of the test set remained at a good level, especially at the 11-month discrimination. The integrative model performed significantly better than the standalone models as seen in Fig. 5, with 11-, 16-, and 20-month AUCs of 0.793, 0.748, and 0.772 respectively for the Genomic prognostic model training set; and 0.547, 0.497, and 0.483 respectively for the Phenotypic prognostic model training set. The test sets of both standalone models performed poorly as well.

The KM curve of the overall unweighted data showed a significantly different survival rate between the risk groups (p = 0.00028). Also, the test results of the KM analysis with weights from Random Forest demonstrated that the survival rate of low-risk score patients was significantly better than that of high-risk score patients (p = 0.0018).

4 Discussion and Conclusion

Our study was able to identify two significant image features with the prognosis of pRCC through feature elimination, including a Zernike shape and a Granularity image. We conclude that the texture and morphology of pathological images may be correlated to pRCC prognosis. Apart from predicting prognosis, the variation in the pathological image features may incur different cell arrangements which may cause variations in the invasion of various potential tumors. We were able to conclude that histopathological image features have a certain ability to predict patient survival, and the combination of genomics and clinical data could further improve the prognosis prediction of pRCC. In the enrichment analysis, associations with the NRF2 pathway and metabolic processes were observed. Recent studies found that NRF2 indeed exhibits an aberrant activation in cancer [40]. Evidence shows that NRF2/KEAP1 signaling pathway is associated with cancer cell proliferation and tumorigenesis through metabolic reprogramming. This correlates with the established notion that RCC is known as a metabolic disease [41] as seen from the diverse array of metabolic defects and perturbations occurring as a result of the genetics driving the tumors. Similarly, study of the genetic information in pRCC denotes significant relation to the metabolic process.In the enrichment analysis, associations with the NRF2 pathway and metabolic processes were observed. Recent studies found that NRF2 indeed exhibits an aberrant activation in cancer [40]. Evidence shows that NRF2/KEAP1 signaling pathway is associated with cancer cell proliferation and tumorigenesis through metabolic reprogramming. This correlates with the established notion that RCC is known as a metabolic disease [41] as seen from the diverse array of metabolic defects and perturbations occurring as a result of the genetics driving the tumors. Similarly, study of the genetic information in pRCC denotes significant relation to the metabolic process.

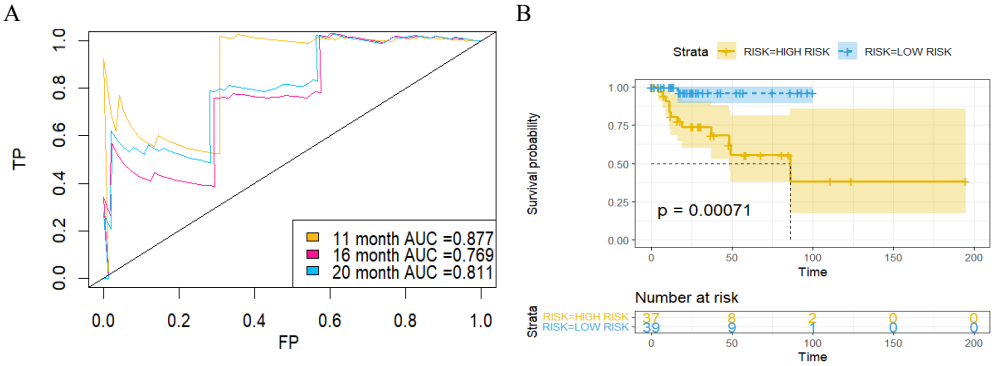

Fig. 3. (A) Test Group Survival Analysis AUC for Genomic and Phenotypic Data. (B) KM Curve of Genomic and Phenotypic Data with weight from Random Forest

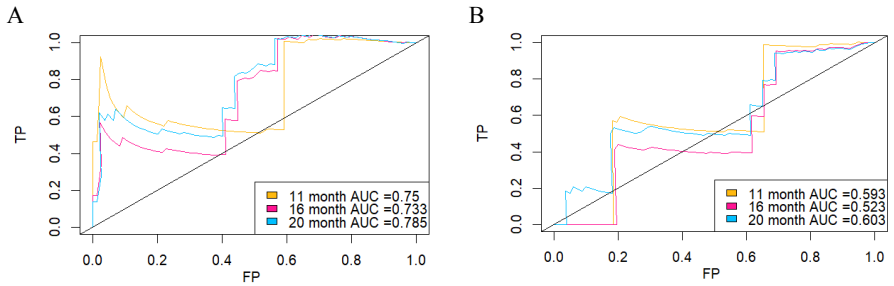

Fig. 4. (A) Test Group Survival Analysis AUC for Genomic Data. (B) Test Group Survival AUC for Phenotypic Data.

Moreover, glucuronidation also suggests an important relation to the genetic prognostic features of pRCC patients. UDP-glucuronosyltransferases (UGT) encodes enzymes that regulate the glucuronidation pathway in humans. UGTs are important metabolic enzymes responsible for approximately 40-70% of endo and xenobiotic reactions [42], which include anti-cancer drugs. UGT enzymes are highly expressed in metabolic tissues such as the liver, intestine, and kidney, consistent with their role in facilitating the elimination of certain metabolites; however, the expression of most UGT members also extends to many other organs and blood cells. Generally, UGTs are anchored to the luminal side of the endoplasmic reticulum (ER), explaining why glucuronidation reactions generally occur in the lumen of the ER [43]. Moreover, UGTs, along with other drug-metabolizing enzymes and transporters, can participate in the inactivation of xenobiotics. Drug inactivation by UGTs is emerging as an important mechanism of drug resistance in cancer [43].

Typically, kidney damage is prone to trigger ER stress. In the kidney, ER stress and unfolded protein response (UPR) participates in acute and chronic histological damages, as it is linked to the molecular basis of progression of Chronic kidney diseases; however, contradictorily, it also promotes cellular adaptation and nephroprotection [44]. UPR pathway is launched by the ER, as it is involved in

maintaining ER homeostasis. Dysregulation of the UPR pathway is linked to kidney disease symptoms. Experimental models have also revealed that disruption of the UPR causes podocyte injury and albuminuria as a mouse grows older [45]. This may suggest why pRCC predominantly occurs in the older generation.

Overall, the model deepened the cognition about the genomic and histopathological image information of pRCC, which could potentially aid in clinical decision-making and treatment of the disease. Moreover, further exploration of biological mechanisms of the histopathological image features and the integration of other data sources, such as clinical data, can lead to a more comprehensive understanding of the disease and improve patient outcomes. Future studies can also explore the generalizability and reproducibility of the model on larger and more diverse patient cohorts to validate its clinical utility.

References

1. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA. Cancer J. Clin.*, vol. 70, no. 1, pp. 7–30, 2020, doi: 10.3322/caac.21590.
2. J. Cheng *et al.*, "Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis," *Cancer Res.*, vol. 77, no. 21, pp. e91–e100, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0313.
3. P. Filippou, B. Shuch, and S. P. Psutka, "Advances in the Characterization of Clear Cell Papillary Renal Cell Carcinoma: Identifying the Sheep in Wolf's Clothing," *Eur. Urol.*, vol. 79, no. 4, pp. 478–479, Apr. 2021, doi: 10.1016/j.eururo.2021.01.023.
4. D. M. Morlote, S. Harada, D. Batista, J. Gordetsky, and S. Rais-Bahrami, "Clear cell papillary renal cell carcinoma: molecular profile and virtual karyotype," *Hum. Pathol.*, vol. 91, pp. 52–60, Sep. 2019, doi: 10.1016/j.humpath.2019.05.011.
5. J. Rysz, B. Franczyk, J. Ławiński, and A. Gluba-Brzózka, "Characteristics of Clear Cell Papillary Renal Cell Carcinoma (ccpRCC)," *Int. J. Mol. Sci.*, vol. 23, no. 1, p. 151, Dec. 2021, doi: 10.3390/ijms23010151.
6. B. Shuch *et al.*, "Understanding Pathologic Variants of Renal Cell Carcinoma: Distilling Therapeutic Opportunities from Biologic Complexity," *Eur. Urol.*, vol. 67, no. 1, pp. 85–97, Jan. 2015, doi: 10.1016/j.eururo.2014.04.029.
7. G. Kovacs *et al.*, "The Heidelberg classification of renal cell tumours," *J. Pathol.*, vol. 183, no. 2, pp. 131–133, Oct. 1997, doi: 10.1002/(SICI)1096-9896(199710)183:2<131::AID-PATH931>3.0.CO;2-G.
8. M. Akhtar, I. A. Al-Bozom, and T. Al Hussain, "Papillary Renal Cell Carcinoma (PRCC): An Update," *Adv. Anat. Pathol.*, vol. 26, no. 2, pp. 124–132, Mar. 2019, doi: 10.1097/PAP.0000000000000220.
9. N. Mendhiratta, P. Muraki, A. E. Sisk, and B. Shuch, "Papillary renal cell carcinoma: Review," *Urol. Oncol. Semin. Orig. Investig.*, vol. 39, no. 6, pp. 327–337, Jun. 2021, doi: 10.1016/j.urolonc.2021.04.013.
10. I. Clark and M. S. Torbenson, "Immunohistochemistry and Special Stains in Medical Liver Pathology," *Adv. Anat. Pathol.*, vol. 24, no. 2, pp. 99–109, Mar. 2017, doi: 10.1097/PAP.0000000000000139.
11. A. H. Beck *et al.*, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci. Transl. Med.*, vol. 3, no. 108, p. 108ra113, Nov. 2011, doi: 10.1126/scitranslmed.3002564.
12. T. Gultekin, C. F. Koyuncu, C. Sokmensuer, and C. Gunduz-Demir, "Two-Tier Tissue Decomposition for Histopathological Image Representation and Classification," *IEEE Trans. Med. Imaging*, vol. 34, no. 1, pp. 275–283, Jan. 2015, doi:

10.1109/TMI.2014.2354373.
13. K.-H. Yu *et al.*, “Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features,” *Nat. Commun.*, vol. 7, p. 12474, Aug. 2016, doi: 10.1038/ncomms12474.
14. D. Hanahan and R. A. Weinberg, “Hallmarks of cancer: the next generation,” *Cell*, vol. 144, no. 5, pp. 646–674, Mar. 2011, doi: 10.1016/j.cell.2011.02.013.
15. H. Z. Al-Lahham, R. S. Alomari, H. Hiary, and V. Chaudhary, “Automating proliferation rate estimation from Ki-67 histology images,” in *Medical Imaging 2012: Computer-Aided Diagnosis*, Feb. 2012, vol. 8315, pp. 669–675. doi: 10.1117/12.911009.
16. L. Mulrane, E. Rexhepaj, S. Penney, J. J. Callanan, and W. M. Gallagher, “Automated image analysis in histopathology: a valuable tool in medical diagnostics,” *Expert Rev. Mol. Diagn.*, vol. 8, no. 6, pp. 707–725, Nov. 2008, doi: 10.1586/14737159.8.6.707.
17. J. M. Bartlett *et al.*, “Evaluating HER2 amplification and overexpression in breast cancer,” *J. Pathol.*, vol. 195, no. 4, pp. 422–428, Nov. 2001, doi: 10.1002/path.971.
18. S. Gulati *et al.*, “Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers,” *Eur. Urol.*, vol. 66, no. 5, pp. 936–948, Nov. 2014, doi: 10.1016/j.eururo.2014.06.053.
19. P. Maroto and B. Rini, “Molecular biomarkers in advanced renal cell carcinoma,” *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 20, no. 8, pp. 2060–2071, Apr. 2014, doi: 10.1158/1078-0432.CCR-13-1351.
20. A.-C. Haury, P. Gestraud, and J.-P. Vert, “The Influence of Feature Selection Methods on Accuracy, Stability and Interpretability of Molecular Signatures,” *PLoS ONE*, vol. 6, no. 12, p. e28210, Dec. 2011, doi: 10.1371/journal.pone.0028210.
21. R. R. Bastien *et al.*, “PAM50 Breast Cancer Subtyping by RT-qPCR and Concordance with Standard Clinical Molecular Markers,” *BMC Med. Genomics*, vol. 5, p. 44, Oct. 2012, doi: 10.1186/1755-8794-5-44.
22. S. He *et al.*, “Aurora kinase A induces miR-17-92 cluster through regulation of E2F1 transcription factor,” *Cell. Mol. Life Sci. CMLS*, vol. 67, no. 12, pp. 2069–2076, Jun. 2010, doi: 10.1007/s00018-010-0340-8.
23. Y. Yuan *et al.*, “Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling,” *Sci. Transl. Med.*, vol. 4, no. 157, p. 157ra143, Oct. 2012, doi: 10.1126/scitranslmed.3004330.
24. A. Calabrò *et al.*, “Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer,” *Breast Cancer Res. Treat.*, vol. 116, no. 1, pp. 69–77, Jul. 2009, doi: 10.1007/s10549-008-0105-3.
25. G. Assié, T. LaFramboise, P. Platzer, J. Bertherat, C. A. Stratakis, and C. Eng, “SNP Arrays in Heterogeneous Tissue: Highly Accurate Collection of Both Germline and Somatic Genetic Information from Unpaired Single Tumor Samples,” *Am. J. Hum. Genet.*, vol. 82, no. 4, pp. 903–915, Apr. 2008, doi: 10.1016/j.ajhg.2008.01.012.
26. P. Neuvial, H. Bengtsson, and T. P. Speed, “Statistical Analysis of Single Nucleotide Polymorphism Microarrays in Cancer Studies,” in *Handbook of Statistical Bioinformatics*, H. H.-S. Lu, B. Schölkopf, and H. Zhao, Eds. Berlin, Heidelberg: Springer, 2011, pp. 225–255. doi: 10.1007/978-3-642-16345-6_11.
27. E.-Y. Oh *et al.*, “Extensive rewiring of epithelial-stromal co-expression networks in breast cancer,” *Genome Biol.*, vol. 16, p. 128, Jun. 2015, doi: 10.1186/s13059-015-0675-4.
28. P. Langfelder and S. Horvath, “Eigengene networks for studying the relationships between co-expression modules,” *BMC Syst. Biol.*, vol. 1, p. 54, Nov. 2007, doi: 10.1186/1752-0509-1-54.
29. G. Zhang, S. Xu, Z. Yuan, and L. Shen, “ Weighted Gene Coexpression Network Analysis Identifies Specific Modules and Hub Genes Related to Major Depression ,” *Neuropsychiatr. Dis. Treat.*, vol. 16, pp. 703–713, Mar. 2020, doi:

10.2147/NDT.S244452.
30. R. Colen *et al.*, “NCI Workshop Report: Clinical and Computational Requirements for Correlating Imaging Phenotypes with Genomics Signatures,” *Transl. Oncol.*, vol. 7, no. 5, pp. 556–569, Oct. 2014, doi: 10.1016/j.tranon.2014.07.007.
31. F. C. Martins *et al.*, “Combined image and genomic analysis of high-grade serous ovarian cancer reveals PTEN loss as a common driver event and prognostic classifier,” *Genome Biol.*, vol. 15, no. 12, p. 526, Dec. 2014, doi: 10.1186/s13059-014-0526-8.
32. U. B. Mogensen, H. Ishwaran, and T. A. Gerds, “Evaluating Random Forests for Survival Analysis using Prediction Error Curves,” *J. Stat. Softw.*, vol. 50, no. 11, pp. 1–23, Sep. 2012, doi: 10.18637/jss.v050.i11.
33. K. L. Pickett, K. Suresh, K. R. Campbell, S. Davis, and E. Juarez-Colunga, “Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker,” *BMC Med. Res. Methodol.*, vol. 21, p. 216, Oct. 2021, doi: 10.1186/s12874-021-01375-x.
34. M. Macenko *et al.*, “A method for normalizing histology slides for quantitative analysis,” in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Boston, MA, USA, Jun. 2009, pp. 1107–1110. doi: 10.1109/ISBI.2009.5193250.
35. H. Li, L. Chen, H. Zeng, Q. Liao, J. Ji, and X. Ma, “Integrative Analysis of Histopathological Images and Genomic Data in Colon Adenocarcinoma,” *Front. Oncol.*, vol. 11, 2021, Accessed: Aug. 09, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fonc.2021.636451
36. K. Soliman, “CellProfiler: Novel Automated Image Segmentation Procedure for Super-Resolution Microscopy,” *Biol. Proced. Online*, vol. 17, no. 1, p. 11, Aug. 2015, doi: 10.1186/s12575-015-0023-9.
37. K.-B. Duan, J. C. Rajapakse, H. Wang, and F. Azuaje, “Multiple SVM-RFE for gene selection in cancer classification with expression data,” *IEEE Trans. Nanobioscience*, vol. 4, no. 3, pp. 228–234, Sep. 2005, doi: 10.1109/tnb.2005.853657.
38. Z. Huang *et al.*, “Deep learning-based cancer survival prognosis from RNA-seq data: approaches and evaluations,” *BMC Med. Genomics*, vol. 13, no. Suppl 5, p. 41, Apr. 2020, doi: 10.1186/s12920-020-0686-1.
39. P. Langfelder and S. Horvath, “WGCNA: an R package for weighted correlation network analysis,” *BMC Bioinformatics*, vol. 9, p. 559, Dec. 2008, doi: 10.1186/1471-2105-9-559.
40. M.-Y. Song, D.-Y. Lee, K.-S. Chun, and E.-H. Kim, “The Role of NRF2/KEAP1 Signaling Pathway in Cancer Metabolism,” *Int. J. Mol. Sci.*, vol. 22, no. 9, p. 4376, Apr. 2021, doi: 10.3390/ijms22094376.
41. W. K. Rathmell, J. C. Rathmell, and W. M. Linehan, “Metabolic Pathways in Kidney Cancer: Current Therapies and Future Directions,” *J. Clin. Oncol.*, vol. 36, no. 36, pp. 3540–3546, Dec. 2018, doi: 10.1200/JCO.2018.79.2309.
42. E. C. C. Mano, A. L. Scott, and K. M. Honorio, “UDP-glucuronosyltransferases: Structure, Function and Drug Design Studies,” *Curr. Med. Chem.*, vol. 25, no. 27, pp. 3247–3255, 2018, doi: 10.2174/0929867325666180226111311.
43. E. P. Allain, M. Rouleau, E. Lévesque, and C. Guillemette, “Emerging roles for UDP-glucuronosyltransferases in drug resistance and cancer progression,” *Br. J. Cancer*, vol. 122, no. 9, Art. no. 9, Apr. 2020, doi: 10.1038/s41416-019-0722-0.
44. M. Gallazzini and N. Pallet, “Endoplasmic reticulum stress and kidney dysfunction,” *Biol. Cell*, vol. 110, no. 9, pp. 205–216, 2018, doi: 10.1111/boc.201800019.
45. A. V. Cybulsky, “Endoplasmic reticulum stress, the unfolded protein response and autophagy in kidney diseases,” *Nat. Rev. Nephrol.*, vol. 13, no. 11, pp. 681–696, Nov. 2017, doi: 10.1038/nrneph.2017.129.