Automated Renal Cancer Grading Using Nuclear Pleomorphic Patterns

Purpose Nuclear pleomorphic patterns are essential for Fuhrman grading of clear cell renal cell carcinoma (ccRCC). Manual observation of renal histopathologic slides may lead to subjective and inconsistent assessment between pathologists. An automated, image-based system that classifies ccRCC slides by quantifying nuclear pleomorphic patterns in an objective and consistent interpretable fashion can aid pathologists in histopathologic assessment.

Methods In the current study, histopathologic tissue slides of 59 patients with ccRCC who underwent surgery at Singapore General Hospital were assembled retrospectively. An automated image classification pipeline detects and analyzes prominent nucleoli in ccRCC images to classify them as either low (Fuhrman grade 1 and 2) or high (Fuhrman grade 3 and 4). The pipeline uses machine learning and image pixel intensity–based feature extraction techniques for nuclear analysis. We trained classification systems that concurrently analyze different permutations of multiple prominent nucleoli image patches.

Results Given the parameters for feature combination and extraction, we present experimental results across various configurations for the classification of a given ccRCC histopathologic image. We also demonstrate that the image score used by the pipeline, termed fraction value, is correlated ($R = 0.59$) with an existing multigene assay–based scoring system that has previously been demonstrated to be a strong indicator of prognosis in patients with ccRCC.

Conclusion The current method provides an objective and fully automated way by which to process pathologic slides. The correlation study with a multigene assay–based scoring system also allows us to provide quantitative interpretation for already established nuclear pleomorphic patterns in ccRCC. This method can be extended to other cancers whose corresponding grading systems use nuclear pattern information.

Clin Cancer Inform. © 2018 by American Society of Clinical Oncology

Daniel Aitor Holdbrook

Malay Singh

Yukti Choudhury

Emarene Mationg Kalaw

Valerie Koh

Hui Shan Tan

Ravindran Kanesvaran

Puay Hoon Tan

John Yuen Shyi Peng

Min-Han Tan

Hwee Kuan Lee

Author affiliations and support information (if applicable) appear at the end of this article.

D.A.H. and M.S. contributed equally as co-first authors.

Corresponding author: Hwee Kuan Lee, PhD, Bioinformatics Institute, 30 Biopolis St, #07-01, Matrix, Singapore 138671; e-mail: leehk@bii.a-star.edu.sg.

INTRODUCTION

Nucleus morphology is an important consideration in the assessment of cancer malignancy in tissue biopsies. Atypical nuclei that appear enlarged, contain prominent nucleoli, and stain heavily—hyperchromasia—are significant diagnostic features. Alterations in nucleoli also act as distinguishing factors for the stages of cancer progression. As the nucleolus is the site of ribosome biogenesis, it increases in size in response to biosynthetic demand. Nucleolar hypertrophy is often observed in highly proliferative cancer cell lines.[1-3]

Pathologic assessment of cancer slides typically involves the identification and assessment of several morphologic features of the nucleus.[1,4-6] In particular, analysis of prominent nucleoli is often an important consideration for cancer prognosis and diagnosis. Indeed, changes to the features of the nucleolus have been reported to

provide an independent prognostic variable in 24 types of cancer.[4]

In clear cell renal cell carcinoma (ccRCC), the Fuhrman nuclear grade has been established as an important prognostic predictor for the survival of patients.[7,8] Adequately preserved renal tissues exhibit nuclei that are round and uniform, with finely granular, evenly distributed chromatin. In higher-grade renal tumors, nuclei are large or have bizarre-looking nucleoli. Sarcomatoid differentiation occurs in 5% of tumors and prognosis is often dismal.[9] Renal tumors with sarcomatoid differentiation display spindle-shaped cells with evident nuclear pleomorphism and cellular atypia.[9]

The reliability of using nuclear and nucleolar features for the estimation of initial pathologic changes has led to their repeated use in the development of automated techniques for diagnostic purposes.[10-23] Many of these studies

in ccRCC have focused on nuclei segmentation and classification according to Fuhrman grade.[14-18] It is suggested that the application of nuclear morphometry may further refine criteria for ccRCC grading.[14,24,25]

By using image features that are relevant to these clinical observations, we have developed an automated classification pipeline for histopathologic slides of patients with ccRCC. Here, we demonstrate that the results of the automated detection and subsequent classification of prominent nucleoli correlate with patient survival and with a clinically relevant, gene-based prognostic indicator. Thus, we provide the quantitative underpinning of the already established importance of nuclear patterns. Furthermore, we describe how improvements in classification accuracy may be obtained by accounting for intraslide nuclear heterogeneity by using concatenated feature vectors of individual prominent nucleoli.

METHODS

Overview

An automated image classification pipeline was developed to grade images that were extracted from ccRCC histopathologic whole-slide images (WSIs). This pipeline uses two separate prominent nucleoli detectors to create two sets of prominent nucleoli image patches from a given image. The pipeline then quantifies nuclear pleomorphic patterns by concatenating features extracted from multiple prominent nucleoli image patches. These concatenated features are then used to classify the two sets of prominent nucleoli image patches to generate a fraction value (FV) score—defined as the fraction of patches (features) that have been predicted to be high in a given detection set. FV score is computed for both sets of detected prominent nucleoli image patches. The pipeline at the final stage then uses these two FV scores to grade the given image as either high or low. The automated image classification pipeline is illustrated in Figure 1, starting from the patient (i) and ending with the final classification (x). The process of image data collection from WSIs is shown in the first three steps. The subsequent processing and classification of these images is detailed in steps iv to x.

ccRCC Data Set

The ccRCC data set was retrospectively assembled from a cohort of 59 patients with ccRCC. Additional details are provided in the Data Supplement.

Data Usage

A two-tier classification system was used to grade patient images—low (Fuhrman grade 1 and 2) and high (Fuhrman grade 3 and 4). Equivalent systems have been reported elsewhere.[14,26-28]

The division of patients and images into training and test sets is explained in the Data Supplement. In summary, 48 of 59 patients were used for training. Images from these patients were used to train a prominent nucleoli detector (Fig 1, iv) and a patch classifier (Fig 1, viii), with and without feature concatenation. Of the remaining 11 patients, six were used exclusively for testing the classification system, and five patients were excluded from the study. The test set was expanded to include unique images that were obtained from 14 patients in the training set. These extra images were used exclusively for testing.

All images that were extracted from the slides by a trained pathologist had dimensions of 1360×1024 pixels, and were used without any color normalization or adjustment. An outline of the data preparation is illustrated in Figure 2.

Prominent Nucleoli Detection

Images that were generated from patients' WSIs (Fig 1, iv and v) were fed into the prominent nucleoli detector to extract image patches for subsequent analysis (Fig 1, v to x). The prominent nucleoli detection system for tissue slide images has been reported previously.[13] This system is composed of multiple cascades of classifiers originally proposed in Viola and Jones.[29] The two detectors—one trained on high ccRCC images and the other on low ccRCC images—provide the location of prominent nucleoli, demarcated as pixel locations, in the extracted images. The two-detector scheme ensures that there is no bias toward either high or low nucleoli. Additional details are provided in the Data Supplement.

Fig 1. Automated image classification pipeline. A biopsy is obtained from a patient (i) and is subsequently sectioned and stained to produce a histopathologic slide (ii). An image (iii) is then extracted from a histopathologic slide. Two detectors (iv), one trained on high and another on low clear cell renal cell carcinoma (ccRCC) images, are used to detect the location of prominent nucleoli in the image. (Step iv is also discussed in the Data Supplement in more detail.) These locations are used to create image patches (v) and, later, to generate individual patch features (vi). As an option, the individual patch features can be concatenated (vii) to generate a new feature space. The single or concatenated features are then classified as high (1) or low (0) (viii). Feature extraction is discussed separately in Figures 2C and 3. A fraction value (FV) score (ix) is assigned to a whole image on the basis of classification scores for all patch features obtained from the high and low detectors. An FV score is obtained for both the high and low detector results, and separately for individual and concatenated features. These FV scores are used for classifying the given extracted image as high or low (x). The figure demonstrates the preparation of the data set for three concatenated features in step vii. Step vii is omitted in the classification of features for individual patches. Figures 4 and 5 illustrate the performance of the pipeline. Cross-references to figures with additional information are indicated on the right.
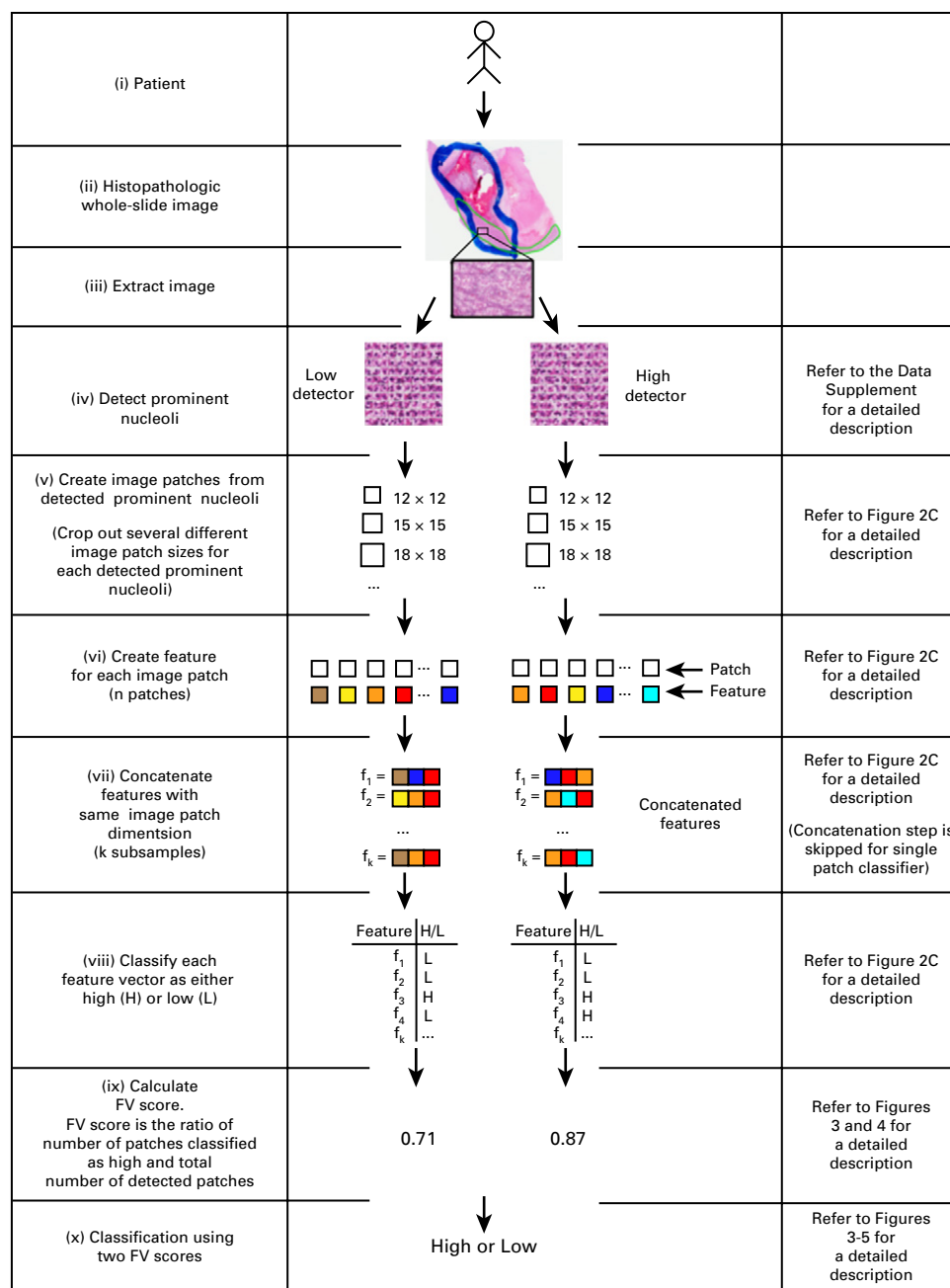
Feature Extraction

A range of image patch sizes (s × s pixels) were extracted from detected prominent nucleoli image locations (Fig 1, iv). Four types of features were created from these extracted image patches (Fig 1, vi): (1) histogram of polar gradient (HPG),[13] (2) enhanced histogram of polar gradient (EHPG),[13] (3) exclusive component analysis feature (XCA),[13] and (4) raw pixel intensity values.

The grade of the parent image, determined by a trained pathologist, served as the ground truth label for the extracted image patches. Additional details are provided in the Data Supplement.

Feature Concatenation

Feature vectors (Fig 1, vii) that were extracted from all image patches in a given image (Fig 1, vi) for HPG, EHPG, XCA, and LR were concatenated end to end. The process of feature concatenation is illustrated in Figure 2C. Additional details are provided in the Data Supplement.

Patch Classification

Concatenated features are used to classify patches (Fig 1, viii) as either high or low. Concatenated features generated in step vii were used to define 196 weak classifiers using
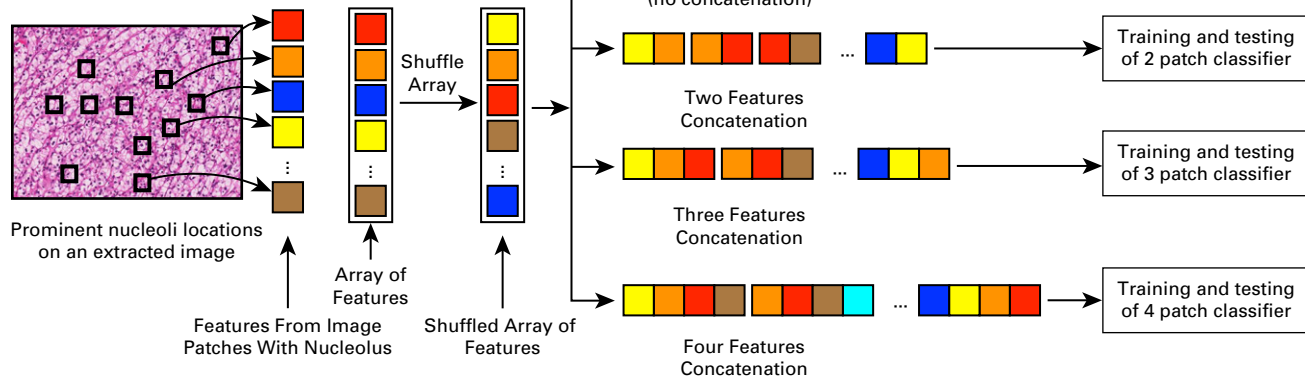
A

Fuhrman Grade	No. of images; total and average (over images) annotated prominent nucleoli; estimated total and average (over images) nuclei		
	Training weak classifiers/cascades	Boosting weak classifiers/detectors	Testing
1 and 2 (low)	25; 2,736, 109.44; 53,309, 2,132.36	25; 5,243, 113.97; 128,058, 2,783.87	10; 1,273, 127.3; 20,972, 2,097.2
3 and 4 (high)	25; 2,545, 106.04; 104,934, 4,372.25	25; 3,095, 123.8; 98,679, 3,947.16	10; 1,746, 174.6; 30,170, 3017
Total (without duplicates)	94; 10,067, 108.24; 283,090, 3,043.98		20; 3,019, 150.95; 51,142, 2,557.1
No. of patients (Total = 54)	48		20 (6 unseen)

B

i) High   ii) Low
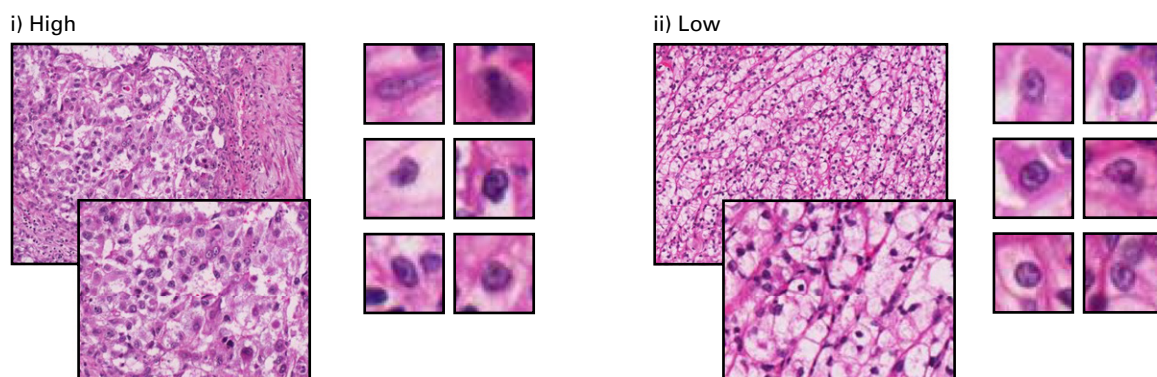
C

Only one image patch size and one type of color space is used to generate the array of features at a time for concatenation.

Prominent nucleoli locations on an extracted image

Features From Image Patches With Nucleolus

Array of Features

Shuffle Array

Shuffled Array of Features

Single Feature (no concatenation) → Training and testing of 1 patch classifier

Two Features Concatenation → Training and testing of 2 patch classifier

Three Features Concatenation → Training and testing of 3 patch classifier

Four Features Concatenation → Training and testing of 4 patch classifier

D

No. of feature vectors (fraction of all possible combinations)			
No concatenation: use of 1 patch's feature	Concatenation of 2 patches' features	Concatenation of 3 patches' features	Concatenation of 4 patches' features
2,500 (1.0)	2,500 (8×10^{-4})	2,500 (1×10^{-6})	2,500 (2×10^{-9})
–	5,000 (2×10^{-3})	5,000 (2×10^{-6})	5,000 (3×10^{-9})
–	10,000 (3×10^{-3})	10,000 (4×10^{-6})	10,000 (6×10^{-9})
–	20,000 (6×10^{-3})	20,000 (8×10^{-6})	20,000 (1×10^{-8})
–	40,000 (1×10^{-2})	40,000 (2×10^{-5})	40,000 (2×10^{-8})

support vector machines (SVMs; for HPG), logistic regression (for raw pixel intensities), and AdaBoost (for EHPG and XCA). These 196 weak classifiers were combined by AdaBoost[30] in the second stage of patch classifier training. Additional details are provided in the Data Supplement.

FV Score

Classification in step viii provides an individual prediction for all feature vectors (Fig 1, ix) in both the high and low prominent nucleoli detection sets. These predictions were then used to compute two FV scores for a given image. FV is defined as the proportion of concatenated features that are predicted to be high for a given detection set.

Image Classification

The final classification of the whole image (Fig 1, x) was performed by SVM with a radial basis function (C = 1,500; γ = 0.1) trained on feature vectors that consisted of only two FVs obtained in step ix. Additional details are provided in the Data Supplement.

RESULTS

Automated Detection and Classification Performance

A pipeline was created to demonstrate that the methods described previously[13,31,32] could be applied to a system of automated slide analysis. At the start of the pipeline, a pathologist was given a WSI from which he or she extracted a smaller subset of 1360 × 1024-pixel images. Manual extraction of these 1360 × 1024-pixel images is required to generate the training data set. For testing, this procedure is conceptually unnecessary. The pipeline then processes a given test image by using two separate detectors. The two detectors—one trained on high images and another on low images—were used to identify pixels that were at the center of prominent nucleoli in test images. Each pixel has an associated RankBoost[33] score, as described previously,[13] which represents the likelihood of it being the site of a prominent nucleolus. Those pixels with the highest RankBoost[33] scores were selected for additional assessment. Features of prominent nucleoli image patches were extracted and concatenated. These features were then classified as being from either a high or low image patch. Hence, for a given test image, two sets of prominent nucleoli image patches were detected, and combinations of image patches were classified as either high or low. FV scores were calculated on the basis of the fraction of all image patch combinations considered and classified as high in the individual detection sets. To determine the optimal number of prominent nucleoli image patches to include in a set in the final classification of a whole test image, test images were assessed separately by using the top 50, 100, 200, 400, and 800 detections from each detector. The final classification of the test image—as either high or low—was performed by an SVM using these two FV features.

In all cases, whether an individual or concatenated feature vectors was considered, the highest F- scores (0.78–0.83) were obtained for the top 400 nucleoli patch detections (Fig 3A). F-score is defined as the harmonic mean of precision and recall, and values range from 0 to 1. A higher F-score indicates a better binary classification performance for a given classifier. The F-scores for two, three, and four concatenated feature vectors demonstrated a common pattern of increasing from approximately 0.7 to more than 0.8 for the top 50 to 400 prominent nucleoli detections (Fig 3B). At more than 400 prominent nucleoli detections, the performance of the pipeline decreased consistently in all training

Fig 2. An outline of data preparation. (A) Table displays how 114 images were organized for training and testing. There were 94 images in the training set. Six images were duplicated between training sets for the weak classifiers and boosted classifier. Testing images for high and low were obtained from 20 different patients. Table also displays the number of patients used at each stage of the study and contains a summary of the number of annotated nucleoli in the images. (B) Example of a high (i) and low (ii) image. The full image (1360 × 1024 pixels) is shown, and an enlarged portion is shown below. Six representative annotated nucleoli are also shown for both high and low images. (C) Schematic representation of the concatenation of two feature vectors. Patches were centered on prominent nucleoli, with a variable border length that ranged from 12 to 30 pixels. Feature vectors extracted from the patches were concatenated end to end, which resulted in a vector with 2×, 3×, or 4× the original dimension, depending on the number of features concatenated together. Only features from the same image were concatenated together. (D) Table displaying the feature space explored in the data set after concatenating. For concatenated feature vectors, there are $^{2500}C_2$, $^{2500}C_3$, and $^{2500}C_4$ permutations for patches that were obtained from both high and low images. This means that the feature space may be more sparsely explored when more vectors are concatenated. Feature vectors for individual patches were shuffled and resampled during concatenation in order to increase the size of the feature space explored.
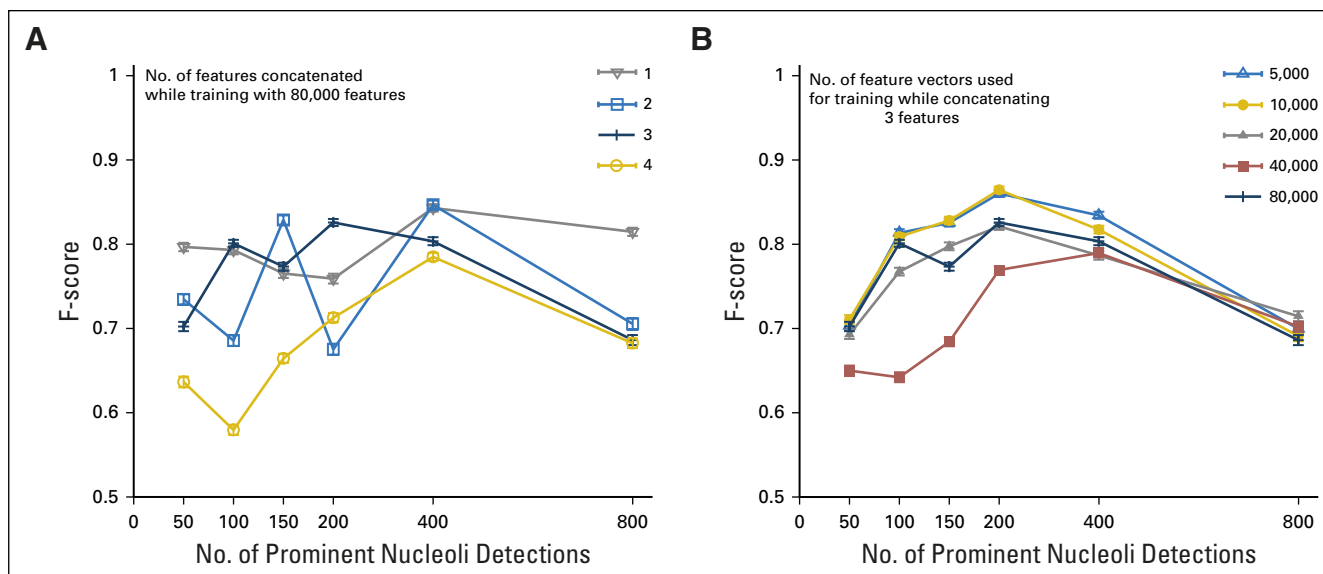
Fig 3. Performance of concatenated feature vectors for the classification of 20 individual patient images by using fraction value scores. (A and B) F-scores for individual and concatenated feature vectors using 80,000 feature vectors used for the training of concatenated patch classifiers (A), and three concatenated feature vectors with increasing numbers of feature vectors sampled for training (B). F-scores were obtained by leave 4-out (two high- and two low-labeled images) cross-validation using individual image fraction value scores obtained for high and low prominent nucleoli detections.

schemes. This is most likely because many of the prominent nucleoli detections greater than the top 400 may not contain an actual prominent nucleolus, and they act as noise that impedes the performance of the pipeline.

Performance of Concatenated Feature Vectors

Classification performance was compared for patch classifiers that were trained by using one (not concatenated), two, three, and four concatenated feature vectors. The same number of training samples (80,000) was used for each set of patch classifiers. Concatenation of feature vectors provided an improvement in F-scores, although not consistently, for two or four patch feature vectors (Fig 3A) over a single patch feature; however, for three patches, there was a marked improvement in F-scores compared with the single patch feature for the top 50, 100, and 150 prominent nucleoli detections. In addition, the F-score of patch classification did not improve significantly with additional data points in the concatenated feature vector training set (Fig 3B).

Pipeline performance was reduced substantially for four concatenated feature vectors, with F-scores consistently below 0.8. This was likely a result of the increase in the relative sparseness of the training data set, which was partially improved by additional sampling (Figs 2D and 3B). With the use of concatenated features, part

of the nuclear heterogeneity can be quantified, but this information may not be sufficient to fundamentally alter classification performance. Meanwhile, classification on the basis of a single patch feature omits altogether the information on pleomorphic patterns.

Increasing the Margin Between High and Low ccRCC Images

The objective for slide or medical image classification is to minimize the misdiagnosis rate; hence, the results were examined for instances in which the images would separate broadly into two groups when FVs from high and low detections were plotted on two-dimensional scatter plot (Fig 4). Intuitively, high images should cluster in an area of the plot with high FVs on both axes, and low images cluster toward the origin (Fig 4C). Images were prone to cluster in this way, particularly if feature vectors from prominent nucleoli were concatenated. This, perhaps, is indicative of how the concatenated patch classification systems may provide greater resolution between high and low images, as it gives greater importance to nuclear heterogeneity (pleomorphism). Indeed, we observed that the most distinct clusters were formed when three or four prominent nucleoli feature vectors were concatenated (Fig 4); however, for these studies, an increase in F-score over the single patch classification system was not always observed
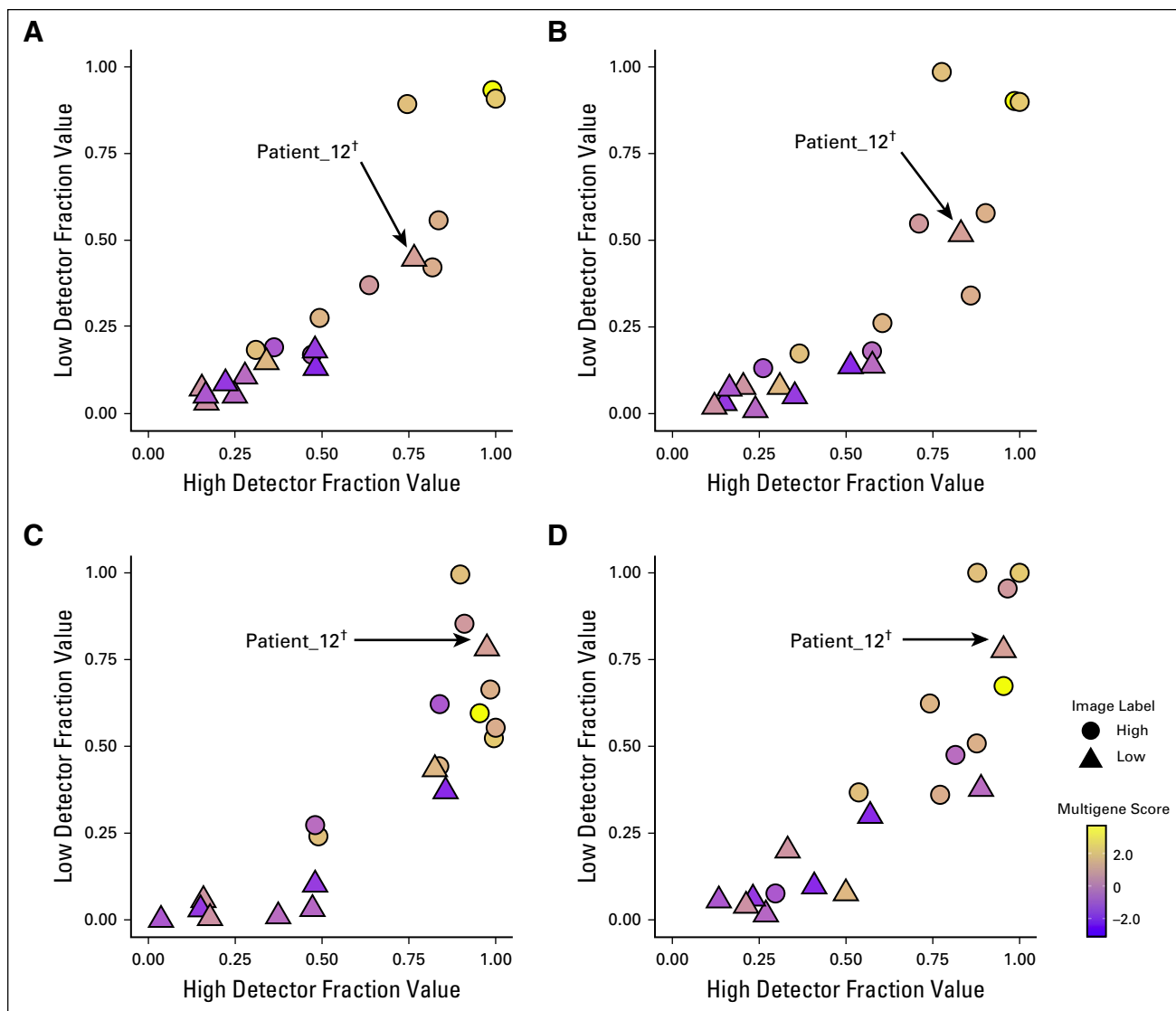
Fig 4. Fraction values (FVs) and multigene scores for individual patient images. FVs have been calculated for high and low prominent nucleoli detections. (A-D) Results are plotted for features created from a single patch (A), two concatenated patches (B), three concatenated patches (C), and four concatenated patches (D). FV is calculated as the proportion of individual or concatenated patches predicted to be high, and is calculated separately for the top 200 prominent nucleoli detections from the high and low detectors. For the concatenated feature vectors, a data set that consisted of 80,000 concatenated patch features, generated from 5,000 individual patches, was used to train the weak classifiers in each case. The multigene score is correlated with the FV for both high and low detectors—that is, lower multigene scores are observed for images close to the origin. The arrow points to the triangular point that corresponds to Patient 12. This patient's image was labeled as low, but it has high multigene and FV scores.

(Figs 3 and 4), because of the consistent misclassification of between two and four patient images. This error can be attributed, in part, to the reduction in the number of categories for image labels—from Fuhrman's four categories to the two categories with FVs. Images that were near the decision boundary of high versus low were difficult to categorize. Indeed, this was evident when the FV scores were compared with a multigene-based expression assay (multigene score), which provides a continuous scale—that is, noncategorical—measure of prognosis.[26] Whereas it can be observed that the multigene score correlates with FV scores (Fig 5A), there are images labeled as high that have low FVs and low multigene scores, and vice versa. The correlation between multigene and FV scores likely relies on the same region of the tissue that is undergoing: (1) harvest for RNA extraction for multigene score evaluation; and (2) image collection for FV score calculation.

Fig 5. Correlation between multigene score and fraction value (FV). (A) A plot of the average FV with respect to the multigene score for 58 patients (62 regions) on manually annotated images. (B) Kaplan-Meier survival plot for the 58 patients used in the study. Shaded areas demonstrate the regions within the 95% CI. Individual patients who had both high- and low-labeled images were considered to be only in the high group in the survival analysis. (C-F) The survival analysis is presented for the 20 test patients, where the high and low labels were predicted by using pipelines that were trained with a single patch (C), two concatenated patches (D), three concatenated patches (E), and four concatenated patches (F). †The death of patient 12 (C-F). This patient's image was labeled as low, but it has high multigene and FV scores. The last stage of the pipeline—that is, the support vector machine—predicted this image to be high. P values displayed in the figure are the result of a Log-rank test.
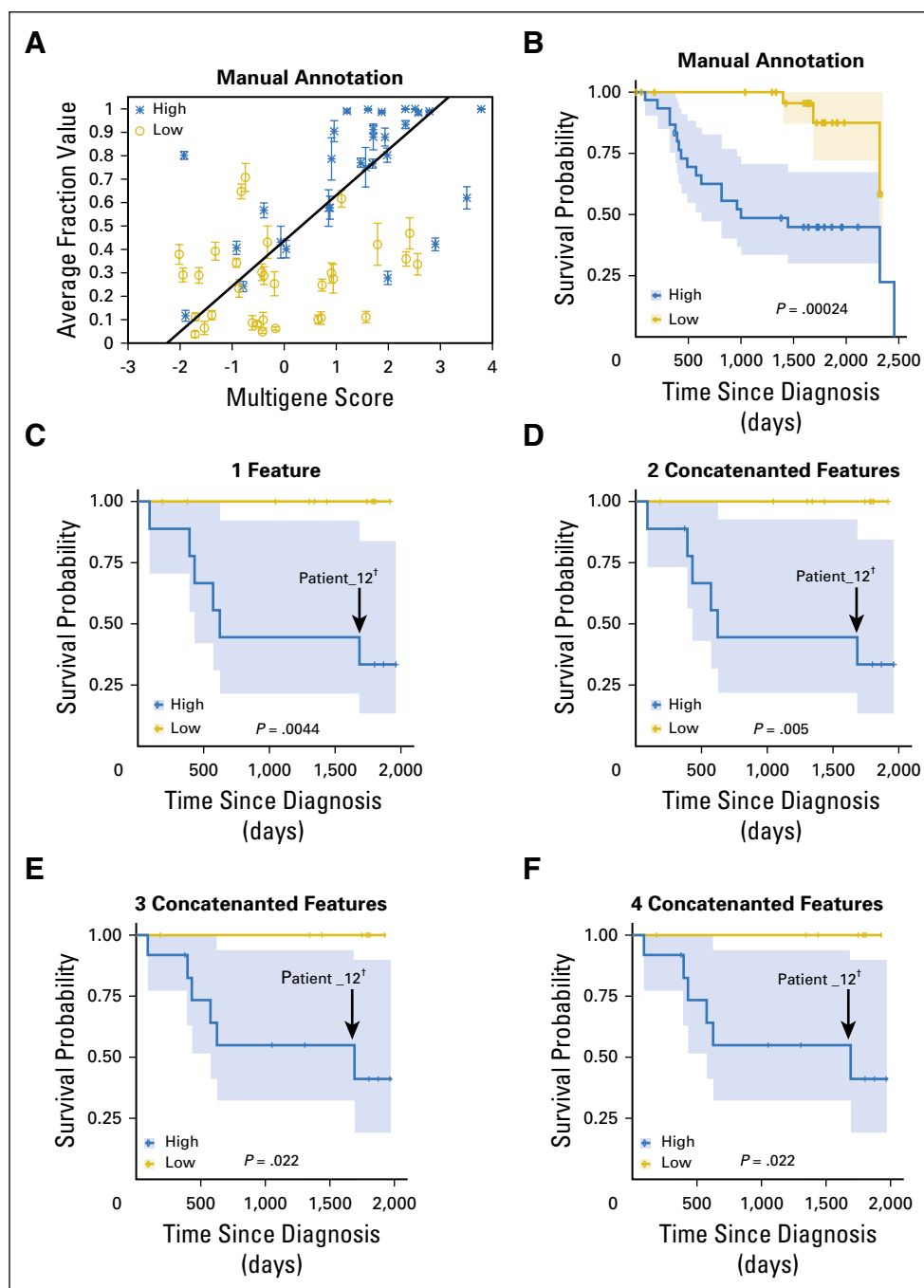
Score and Survival

In this study, FV score was used to detect whether an image obtained from WSI is high or low. To demonstrate the robustness of this classification system, a small-scale, single-patch—that is, not concatenated—study was performed by using a total of 29 patients who were separated into independent training and test sets: 19 patients—nine high and 10 low—for training and

10 patients—five high and five low—patients for testing (Data Supplement). Here, instead of using the prominent nucleoli detectors, we used manual annotations to determine the locations of prominent nucleoli. A simple threshold (≥ 0.5 for high and < 0.5 for low) was used to grade the images on the basis of FV scores. This simple measure correctly classified nine of 10 high images and nine of 10 low images (Data Supplement).

FV score focuses exclusively on the features of stained nuclei. To determine whether these image features had a molecular and/or genetic underpinning, we compared the FV score with the multigene score. Multigene scores were obtained for 62 regions from 58 patients in the current study. Multigene scores and WSIs were collected in the same laboratory. This ensured that the site of image extraction coincided with the site of tissue harvest for the multigene assay. A five-fold cross-validated testing of the single-patch classifier resulted in a correlation ($R = 0.59$) between the average FV predicted for the images from a region in the slide and the multigene score for the corresponding tissue in the same region (Fig 5A). This high degree of correlation further validates the automated image-based classification pipeline, and provides a molecular link to nuclear-based image grading. The multigene score has already been demonstrated to be highly correlated with patient prognosis,[26] and its correlation with FV score further validates the use of this measure in grading histopathologic images.

Apart from the above FV-based image classification, we assessed the survival of patients in this cohort in which the groupings were decided by patients' histopathologic grading—high and low as assigned by a trained pathologist. Kaplan-Meier modeling was used for the survival analysis, and a Log-rank test was used to determine statistical significance between high and low groups. Over the 6-year period, a lower survival probability was observed for patients who were determined to be in the high group at diagnosis ($P = .00024$; Fig 5B), which indicates that the internal labeling system is a significant indicator of prognosis.

Additional validation of the FV score as a prognostic indicator was achieved by performing survival analysis on the predicted high and low labels from test images from the pipeline in which FV was calculated by using detected prominent nucleoli (Fig 5C-F). In all cases, regardless of whether the pipeline used a single or concatenated feature, no patient who was predicted to be low had died at the censor date. Of particular note, extracted images for one of the patients, patient 12, were all labeled as low by the pathologist. The image classification system classified this patient's image as high, along with assigning high FV scores. Patient 12 had died by the end of the study (Fig 5C-F). This implies that the

pipeline can correctly identify ccRCC grading. In addition, we observed that Patient 12 had elevated multigene scores (Fig 4).

DISCUSSION

Automated screening of patient data will enable the large-scale public health surveys currently considered to be too costly and time consuming. To tackle this problem, we developed a pipeline to automatically detect prominent nucleoli that can be used subsequently to classify WSI obtained from patients with ccRCC. The scoring system—FV —used by the pipeline correlates with clinical measures, such as the multigene score and patient survival statistics, and can therefore be used either as a cost-effective method for the rapid assessment of prognosis or to enhance the diagnostic power of existing methods; however, to demonstrate the robustness required for routine clinical applications, it may be necessary to re-evaluate the pipeline by using a larger, more heterogeneous patient data set. These limitations are discussed in the Data Supplement.

The Fuhrman grading system predominantly focuses on the morphology of the nucleus and the existence of prominent nucleoli. We attempted to capture some of this information and provide a quantitative assessment of patient histopathologic images. Critically, we have demonstrated that, by focusing on prominent nucleoli, we are able to achieve high predictive power, which indicates that this feature and the features of the surrounding nucleus play a dominant role in cancer grading. The concatenation of feature vectors of individual nucleoli patches allowed the pipeline to assess the intraslide nuclear heterogeneity for use in classifications. This resulted in small gains in accuracy, and much of the predictive power clearly results from assessing nucleoli individually. Thus, the relative incidence of nuclear irregularity (atypia) in high- or low-scoring images—and not simply its presence—seems to be the determining factor for cancer grade. Therefore, improvements in classification can be achieved by incorporating this information.

AUTHOR CONTRIBUTIONS

Conception and design: Daniel Aitor Holdbrook, Malay Singh, Min-Han Tan, Hwee Kuan Lee
Administrative support: Daniel Aitor Holdbrook, Malay Singh, Hwee Kuan Lee
Provision of study materials or patients: John Yuen Shyi Peng, Puay Hoon Tan
Collection and assembly of data: Daniel Aitor Holdbrook, Malay Singh, Yukti Choudhury, Emarene Mationg Kalaw, Valerie Koh, Hui Shan Tan, Ravindran Kanesvaran, John Yuen Shyi Peng, Puay Hoon Tan, Hwee Kuan Lee
Data analysis and interpretation: Daniel Aitor Holdbrook, Malay Singh, Emarene Mationg Kalaw, Yukti Choudhury, Hwee Kuan Lee
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

Affiliations

Daniel Aitor Holdbrook, **Malay Singh**, **Emarene Mationg Kalaw,** and **Hwee Kuan Lee**, Bioinformatics Institute; **Malay Singh** and **Hwee Kuan Lee**, National University of Singapore; **Yukti Choudhury** and **Min-Han Tan**, Lucence Diagnostics; **Yukti Choudhury** and **Min-Han Tan**, Institute of Bioengineering and Nanotechnology; **Valerie Koh**, **Puay Hoon Tan**, and **John Yuen Shyi Peng**, Singapore General Hospital; **Hui Shan Tan**, **Ravindran Kanesvaran**, and **Min-Han Tan**, National Cancer Center Singapore; and **Hwee Kuan Lee**, Institute for Infocomm Research, Singapore.

REFERENCES

1. Montanaro L, Treré D, Derenzini M: Nucleolus, ribosomes, and cancer. Am J Pathol 173:301-310, 2008

2. Stępiński D: Nucleolus-derived mediators in oncogenic stress response and activation of p53-dependent pathways. Histochem Cell Biol 146:119-139, 2016

3. Derenzini M, Trerè D, Pession A, et al: Nucleolar size indicates the rapidity of cell proliferation in cancer tissues. J Pathol 191:181-186, 2000

4. Derenzini M, Montanaro L, Treré D: What the nucleolus says to a tumour pathologist. Histopathology 54:753-762, 2009

5. Rashid F, Ul Haque A: Frequencies of different nuclear morphological features in prostate adenocarcinoma. Ann Diagn Pathol 15:414-421, 2011

6. Derenzini M, Trerè D, Pession A, et al: Nucleolar function and size in cancer cells. Am J Pathol 152:1291-1297, 1998

7. Fuhrman SA, Lasky LC, Limas C: Prognostic significance of morphologic parameters in renal cell carcinoma. Am J Surg Pathol 6:655-663, 1982

8. Erdoğan F, Demirel A, Polat O: Prognostic significance of morphologic parameters in renal cell carcinoma. Int J Clin Pract 58:333-336, 2004

9. Shuch B, Bratslavsky G, Linehan WM, et al: Sarcomatoid renal cell carcinoma: A comprehensive review of the biology and current treatment strategies. Oncologist 17:46-54, 2012

10. Street WN, Wolberg WH, Mangasarian OL: Nuclear feature extraction for breast tumor diagnosis. Biomed Image Process Biomed Vis 1905:861-870, 1993

11. Makarov D V, Marlow C, Epstein JI, et al: Using nuclear morphometry to predict the need for treatment among men with low grade, low stage prostate cancer enrolled in a program of expectant management with curative intent. Prostate 68:183-189, 2008

12. Farhane Z, Bonnier F, Casey A, et al: Cellular discrimination using in vitro Raman micro spectroscopy: The role of the nucleolus. Analyst (Lond) 140:5908-5919, 2015

13. Yap CK, Kalaw EM, Singh M, et al: Automated image based prominent nucleoli detection. J Pathol Inform 6:39, 2015

14. Yeh F.-C., Parwani A. V., Pantanowitz L, et al: Automated grading of renal cell carcinoma using whole slide imaging. J Pathol Inform 5:23, 2014

15. Schüffler PJ, Fuchs TJ, Ong CS, et al: TMARKER: A free software toolkit for histopathological cell counting and staining estimation. J Path Inform 4:, 2013

16. Yeh F-C, Ye Q, Hitchens TK, et al: Mapping stain distribution in pathology slides using whole slide imaging. J Path Inform 5:1, 2014

17. Kruk M, Kurek J, Osowski S, et al: Ensemble of classifiers and wavelet transformation for improved recognition of Fuhrman grading in clear-cell renal carcinoma. Biocybern Biomed Eng 37:357-364, 2017

18. Kruk M, Osowski S, Kozlowski W, et al.: Computer-assisted Fuhrman grading system for the analysis of clear-cell renal carcinoma: A pilot study. Przegl Elektrotechniczny 89:268-270, 2013

19. Fuchs TJ, Buhmann JM: Computational pathology: Challenges and promises for tissue analysis. Comput Med Imaging Graph 35:515-530, 2011

20. Fuchs TJ, Wild PJ, Moch H, et al: Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. Med Image Comput Comput Assist Interv 11:1-8, 2008.

21. Greenspan H, van Ginneken B, Summers RM: Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Trans Med Imaging 35:1153-1159, 2016

22. Madabhushi A, Lee G: Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Imag Anal 33:170-175, 2016.

23. Janowczyk A, Madabhushi A: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform 7:29, 2016

24. Pound CR, Partin AW, Epstein JI, et al: Nuclear morphometry accurately predicts recurrence in clinically localized renal cell carcinoma. Urology 42:243-248, 1993

25. Lohse CM, Blute ML, Zincke H, et al: Comparison of standardized and nonstandardized nuclear grade of renal cell carcinoma to predict outcome among 2,042 patients. Am J Clin Pathol 118:877-886, 2002

26. Choudhury Y, Wei X, Chu Y-H, et al: A multigene assay identifying distinct prognostic subtypes of clear cell renal cell carcinoma with differential response to tyrosine kinase inhibition. Eur Urol 67:17-20, 2015

27. Sun M, Lughezzani G, Jeldres C, et al: A proposal for reclassification of the Fuhrman grading system in patients with clear cell renal cell carcinoma. Eur Urol 56:775-781, 2009

28. Zisman A, Pantuck AJ, Dorey F, et al: Improved prognostication of renal cell carcinoma using an integrated staging system. J Clin Oncol 19:1649-1657, 2001

29. Viola P, Jones MJ: Robust real-time face detection. Int J Comput Vis 57:137-154, 2004

30. Freund Y, Schapire RE: A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119-139, 1997

31. Singh M, Zeng Z, Kalaw EM, et al: A study of nuclei classification methods in histopathological images, in Innovation in Medicine and Healthcare 2017. New York, NY, Springer, 2017, pp 78-88

32. Singh M, Kalaw EM, Giron DM, et al: Gland segmentation in prostate histopathological images. J Med Imaging (Bellingham) 4:027501, 2017

33. Freund Y, Iyer R, Schapire RE, et al: An efficient boosting algorithm for combining preferences. J Mach Learn Res 4:933-969, 2004