

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Recurrence analysis on prostate cancer patients with Gleason score 7 using integrated histopathology whole-slide images and genomic data through deep neural networks

Jian Ren
Kubra Karagoz
Michael L. Gatzka
Eric A. Singer
Evita Sadimin
David J. Foran
Xin Qi

SPIE.

Jian Ren, Kubra Karagoz, Michael L. Gatzka, Eric A. Singer, Evita Sadimin, David J. Foran, Xin Qi, "Recurrence analysis on prostate cancer patients with Gleason score 7 using integrated histopathology whole-slide images and genomic data through deep neural networks," *J. Med. Imag.* 5(4), 047501 (2018), doi: 10.1117/1.JMI.5.4.047501.

Recurrence analysis on prostate cancer patients with Gleason score 7 using integrated histopathology whole-slide images and genomic data through deep neural networks

Jian Ren,^{a,†} Kubra Karagoz,^{b,†} Michael L. Gatz,^b Eric A. Singer,^c Evita Sadimin,^d David J. Foran,^d and Xin Qi^{d,*}

^aRutgers, the State University of New Jersey, Department of Electrical and Computer Engineering, Piscataway, New Jersey, United States

^bRutgers Cancer Institute of New Jersey, Department of Radiation Oncology, New Brunswick, New Jersey, United States

^cRutgers Cancer Institute of New Jersey, Section of Urologic Oncology, New Brunswick, New Jersey, United States

^dRutgers Cancer Institute of New Jersey, Department of Pathology and Laboratory Medicine, New Brunswick, New Jersey, United States

Abstract. Prostate cancer is the most common nonskin-related cancer, affecting one in seven men in the United States. Gleason score, a sum of the primary and secondary Gleason patterns, is one of the best predictors of prostate cancer outcomes. Recently, significant progress has been made in molecular subtyping prostate cancer through the use of genomic sequencing. It has been established that prostate cancer patients presented with a Gleason score 7 show heterogeneity in both disease recurrence and survival. We built a unified system using publicly available whole-slide images and genomic data of histopathology specimens through deep neural networks to identify a set of computational biomarkers. Using a survival model, the experimental results on the public prostate dataset showed that the computational biomarkers extracted by our approach had hazard ratio as 5.73 and *C*-index as 0.74, which were higher than standard clinical prognostic factors and other engineered image texture features. Collectively, the results of this study highlight the important role of neural network analysis of prostate cancer and the potential of such approaches in other precision medicine applications. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.5.4.047501](https://doi.org/10.1117/1.JMI.5.4.047501)]

Keywords: prostate cancer; Gleason score; deep neural networks; whole-slide images; genomic data.

Paper 18105RR received May 16, 2018; accepted for publication Oct. 23, 2018; published online Nov. 15, 2018.

1 Introduction

Prostate cancer remains the most common noncutaneous malignant tumor in the Western world accounting for approximately one in five of newly diagnosed tumors in men and resulting in an estimated 29,430 deaths in 2018.¹ In the United States, approximately one in seven men will be diagnosed with this disease.¹ Based on Gleason score, prostate specific antigen (PSA) value, tumor stage, age, and race, patients with prostate cancer are stratified into low-, intermediate-, and high-risk groups.^{2,3}

A strong predictor of survival among men with prostate cancer is the Gleason score rendered by a pathologist based upon a microscopic evaluation of a representative histopathology specimen.⁴ These scores are based solely upon morphology and structural patterns of the constituent cells and glands. Patients with Gleason score 6 or lower often undergo active surveillance as there is reduced risk of tumor progression for those patients compared to patients with score 7 or higher.^{5,6} Tumors that are assigned Gleason score 7 can be delineated into a primary region exhibiting a histopathology pattern graded as 4 and a secondary region exhibiting a histopathology pattern graded as 3. Such samples are referred to as Gleason 4 + 3 tumors, whereas the inverse pattern exhibiting a primary pattern of 3 and a secondary pattern of 4 would constitute a Gleason 3 + 4 tumors. Patients with Gleason 4 + 3 tumors

have an increased risk of recurrence and progression leading to an increased risk of prostate cancer-specific mortality when compared to those afflicted with Gleason 3 + 4 tumors.^{7–9} The literature clearly shows that predicting disease recurrence in a man with Gleason score 7 prostate cancer can have a significant impact on his disease management and survival.^{8–10}

Phenotypically, tumor regions with Gleason pattern 3 are composed of single glands with distinct size and shape whereas ones with Gleason pattern 4 exhibit large irregular cribriform glands or fused, ill-defined glands with poorly formed glandular lumina.^{11–13} In spite of established guidelines, Gleason grading remains a relatively subjective process that results in an ~30% grading discrepancy among the scores provided by pathologists.^{11–15} There have been many attempts to develop computer-aided Gleason grading methods and systems^{11,16–21} in order to introduce objective, reproducible criteria into the process of Gleason pattern quantification, and grading. One previous study has explored an integration of image features along with protein expression to predict recurrent prostate cancer.²² However, to date, there has been no study focused on utilizing patients' pathology images and genomic pathway analyses in combination to predict recurrence-free survival (RFS) for men with prostate cancer.

Microarray-based gene expression signatures have been used in various studies to identify cancer subtypes, determine the RFS of disease, and characterize response to specific therapies.²³ Multiple investigations have also shown that gene expression signatures can be used to analyze oncogenic pathways and

*Address all correspondence to: Xin Qi, E-mail: qixi@rutgers.edu

†Equal contribution.

these signatures have been used to identify differences between specific cancer types and tumor subtypes. Moreover, patterns of oncogenic pathway activity have been used to identify differences in underlying molecular mechanisms and have been shown to correlate with clinical outcomes of patients afflicted with specific cancers.^{24–27}

In recent years, whole-slide image (WSI) has been more widely used in histopathology diagnosis. With a fast development of deep learning, histopathology image analysis approaches have demonstrated significant advances in cellular segmentation^{28–31} and tissue classifications^{30–34} using convolutional neural networks (CNN). Some research groups reported their studies using histopathology WSI for many applications.^{35–37} Due to a giga-pixel size of a WSI's, it is often impractical to train the CNN using WSIs directly. Consequently, patch-based algorithms are widely applied in histopathology image analysis.^{34,38–43}

In this study, we developed a computational biomarker quantification system by integrating histopathology WSIs and genomic data into one deep neural network. In order to use the distribution of Gleason patterns on a WSI, we applied patches as inputs to the network. The patches were forwarded through a CNN to get the image features. Then based on the patches' spatial relationship, the image features were modeled using a recurrence neural network (RNN),⁴⁴ namely long short-term memory (LSTM).⁴⁵ The pathway scores calculated from the genomic data were forwarded to a multilayer perceptron (MLP) to get the genomic features. And the image and genomic features were integrated together to get the computational biomarkers. Moreover, we used RFS (months) since their initial treatment as the time-to-recurrence variable for a survival model.

We chose a Cox proportional-hazard regression model^{46,47} since it is commonly used in medical research for investigating associations between the survival time of patients and predictor variables.

2 Methods

In this section, we introduced our approach on building a unified system using WSI and genomic data through deep neural networks to quantify computational biomarkers, which were fed into a survival model for patients' recurrence analysis. Our methods consisted of four steps. First, the pathway activities of prostate cancer were quantified by pathway scores using RNA sequences. Second, the histopathology WSIs were preprocessed to obtain the region-of-interest as the image patches preparation. Third, the image patches and pathway scores were integrated into a unified system using the deep learning approach to extract computational biomarkers. Finally, we used the computational biomarkers in conjunction with clinical prognostic factors as the input of the survival model to calculate the disease recurrence ratios and probabilities. Figure 1 shows the overview of the pipeline of the whole study.

2.1 Experiment Dataset

In this study, we used publicly available prostate cancer data downloaded from the data portal of the Genomic Data Commons (GDC).⁴⁸ GDC is the largest public available data portal that includes image data from The Cancer Genome Atlas (TCGA),⁴⁹ genomic data, and clinical data. The TCGA barcode⁴⁸ is the primary identifier of GDC data acquisition

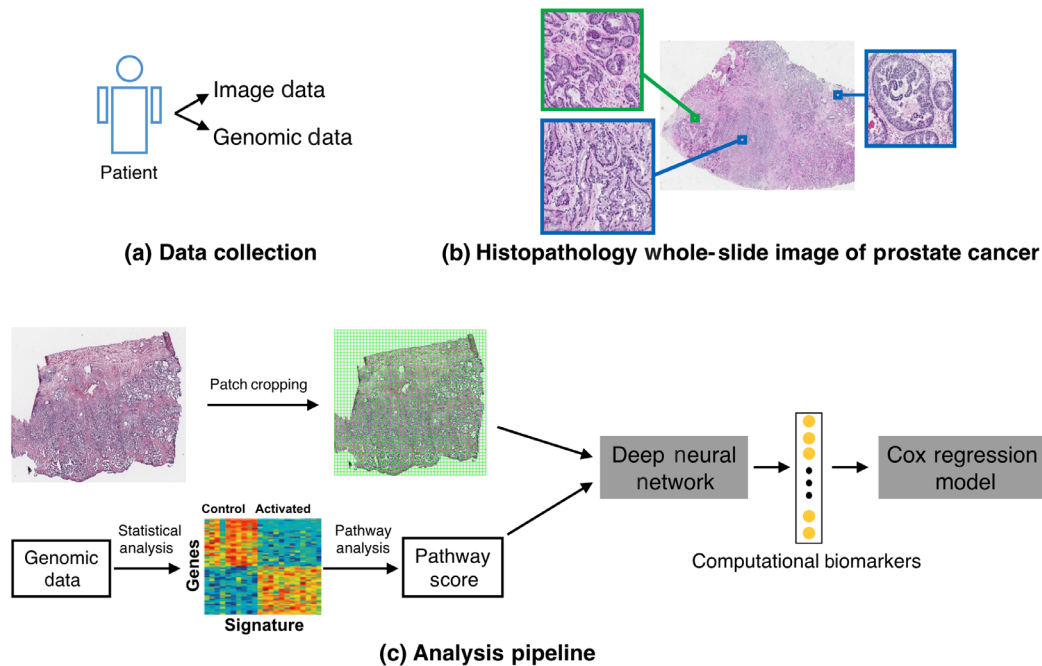


Fig. 1 An overview of the pipeline of our study using histopathology WSIs and genomic data for prostate cancer recurrence prediction for patients with Gleason score 7. (a) WSI images and genomic data were collected from patients with prostate cancer; (b) a prostate WSI exhibits different Gleason patterns. For example, a region in a green square has the Gleason pattern 3 while regions in blue squares have the Gleason pattern 4; (c) the pathway scores were quantified using RNA sequences. Patches of region of interests were automatically selected from WSIs. The image patches and pathway scores were integrated into deep neural networks to extract computational biomarkers, which were fed into a Cox regression model in conjunction with clinical prognostic factors for disease recurrence analysis.

protocol. For this study, in total, there were 43 Gleason 3 + 3, 146 Gleason 3 + 4, 101 Gleason 4 + 3, and 49 Gleason 4 + 4, which contain 1229, 4753, 2997, and 1597 patches, respectively. For the recurrence study of patients with Gleason 7, we used all the data from Gleason 6, 7, and 8 to train the networks to extract the computational biomarkers. In this way, the training data contained more images of Gleason patterns 3 and 4 compared to a training data if only use patients' data with Gleason 7 (3 + 4 or 4 + 3). For the recurrence study of patients with Gleason 7, the computational biomarkers of patients with Gleason 7 were fed into a survival model while the patients with other Gleason score were withheld.

The patients were randomly divided into the training set, validation set, and testing test with the ratio of 70%, 10%, and 20%; these groups were utilized for the recurrence analyses. In addition to the Gleason score, we compared the computational biomarkers quantified from the unified image and genomic data system with other clinical factors including patients' PSA, age, and tumor stage, which are publicly available from GDC data portal.

The WSI patches preparation was a two-step cropping-selection process. First, the image patches within each WSI were automatically cropped under 40 \times objective magnification with a patch size 4096 \times 4096. The patches were cropped with a stride as 4096 to avoid overlapping. We resized all the patches to the size of 256 \times 256 using Lanczos filtering.⁵⁰ Second, any specimens with insufficient tissue patches were automatically eliminated from the experiments due to the heterogeneous quality of the prostate WSIs. The patches with the tissue accounting for at least 20% of the whole area were selected.

2.2 Pathway Score Quantification from RNA Sequencing Data

To quantify pathway scores, we used the gene expression data, which were RNA (Illumina HiSeq) sequencing data from patients with Gleason score 7. The data are publicly available through GDC data portal. We preprocessed the RNA data by log transformed and median centered. A panel of previously published 265 experimentally derived gene expression signatures was applied to the entire cohort to identify patterns of oncogenic signaling in each tumor. To apply a given signature, the expression data were filtered to contain only those genes included in the given signature and the mean expression value of these genes was calculated to provide a score for each sample.^{25,26}

2.3 Computational Biomarker Extraction

In order to obtain computational biomarkers from the WSIs and genomic data, we built a unified feature quantification system using CNN to model WSI histopathology image patches and genomic data together. Furthermore, we leveraged the RNN to model the spatial relationship of the cropped patches within the WSI. The network architecture is shown in Fig. 2.

2.3.1 Modeling histopathology image patches and genomic data

In order to combine the image information along with the genomic data, we used the patches and pathway scores as the input to the network. We forwarded the pathway scores into

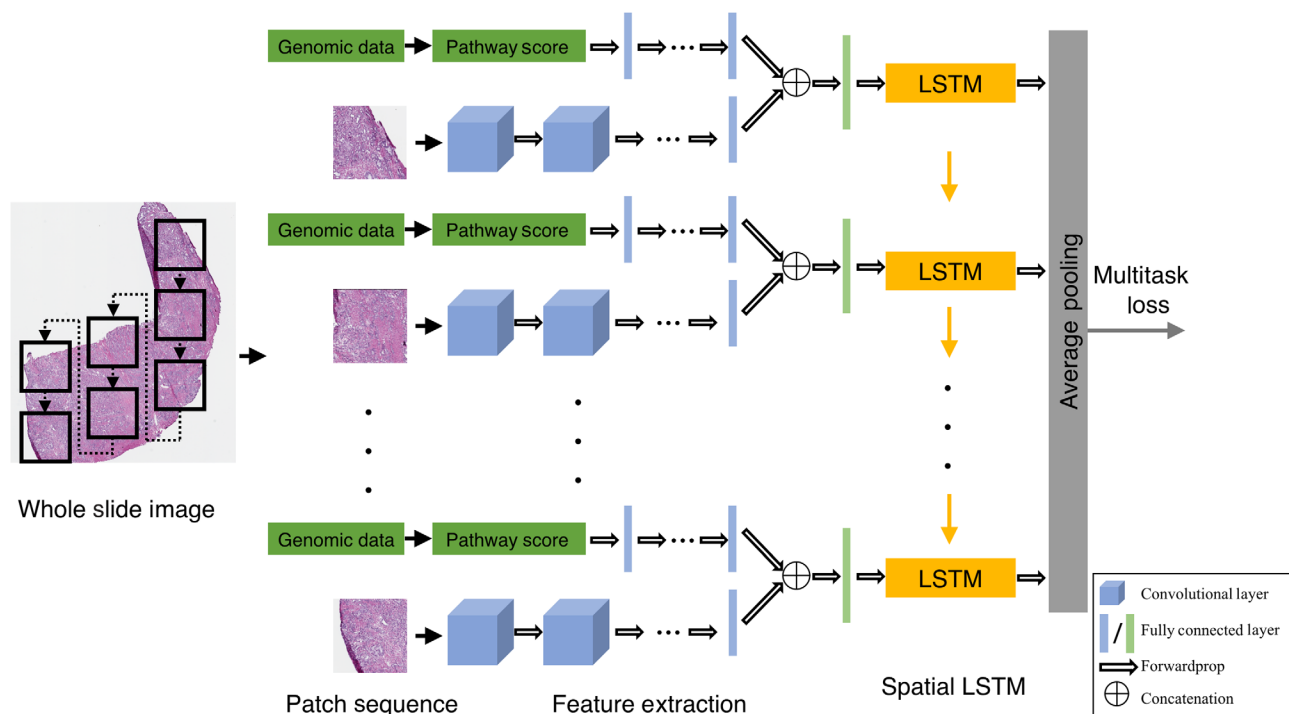


Fig. 2 Network architecture for extracting computational biomarkers from the WSI and genomic data. We used seven LSTM cells in the network. The calculated pathway scores from the genomic data were forwarded into an MLP that contains three FC layers. The last layer of the MLP was connected with the features extracted from the image patches to serve as the input for the LSTM after an FC layer. On top of the LSTM, we utilized an average pooling layer.

an MLP that includes three fully connected (FC) layers, with 1024, 512, and 256 hidden units, respectively. The genomic features were the output of the last FC layer. Meanwhile, we incorporated the AlexNet⁵¹ to extract the features from image patches. We concatenated the genomic features obtained from the pathway scores with the image features from the second to the last layer of the AlexNet. The concatenated features served as the input to an FC layer before LSTM.

Due to the giga-pixel WSI's, we considered an integrity of the whole tissue regions on a single WSI instead of using the individual patches to quantify image features as shown in previous studies.^{39,41} The spatial relationship of the adjacent patches was modeled as an image sequence. We adopted a type of RNN,⁴⁴ LSTM,⁴⁵ to model the features extracted from the image patches and genomic data given LSTM has shown its successes among various applications including speech recognition,^{52,53} language translation models,⁵⁴ image captioning,⁵⁵ and video classification.⁵⁶ Compared with the traditional RNN that has vanishing and exploding gradients problems, LSTM is more effectively in sequence modeling by incorporating memory cells with several gates to obtain long-range dependencies.

More formally, for the input feature sequence (x_1, x_2, \dots, x_T) that x_i represents the activations from the CNN of the i 'th patch, we used LSTM to compute the output sequence (y_1, y_2, \dots, y_T) , where the layer of LSTM was computed recursively from $t = 1$ to $t = T$ following the equations:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \quad (4)$$

$$h_t = o_t \tanh(c_t), \quad (5)$$

where h_t is the hidden vector, i_t , c_t , f_t , and o_t represent the activation vectors of the input gate, memory cell, forget gate, and output gate, respectively. W terms denote the weight matrices connecting different units, b terms denote the bias vectors, and σ is the logistic sigmoid function. The memory cell c_t has the inputs of the weighted sum of the current inputs and the previous memory cell unit c_{t-1} , which could learn when to forget the old information and when to consider the new information. The output gate o_t controls the propagation of information to the following step. The visualization of the LSTM cell is shown in Fig. 3.

Since it is a sequential task to train LSTM, patches from a WSI were formed by a specific routine. As shown in Fig. 2, we used center coordinates of patches to remark the location of each patch. The sequence of patches within a single WSI was arranged from right up patch down to lower left one, which was illustrated by the dotted lines with black arrows on an example of a WSI on Fig. 2. In this way, it allowed us to consider both unique characteristics of each patch and fine-grained variations among patches within a single WSI. For each tumor WSI, the patches and the pathway scores were fed into the network to get features and then incorporated into the LSTM recursively. In addition, the average pooling layer was applied on top of

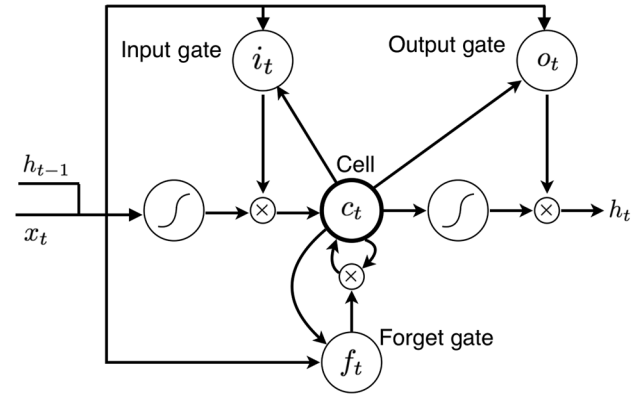


Fig. 3 The visualization of an LSTM cell.

the network to get the computational biomarkers for the WSI and the genomic data. The number of hidden units for each LSTM was 1024. During the training process, we applied the multitask loss and assigned the primary pattern and the Gleason score for the WSIs and genomic data.

2.3.2 Multitask loss function

For the TCGA prostate WSIs, the primary Gleason pattern, the secondary Gleason pattern, and the sum of both patterns (i.e., Gleason score) were publicly available from GDC data portal. To model the variations among Gleason patterns, we utilized the multitask loss to enable the network to learn as much information about the Gleason pattern distributions from the patches of a WSI as possible. Therefore, we gave the primary pattern and the sum score as labels for each patch along with the pathway score and use the following multitask loss function:

$$\mathcal{L}_{\text{multitask}} = - \sum_{i=0}^N \mathbf{t}_i^p \cdot \log \hat{\mathbf{t}}_i^p - \sum_{i=0}^N \mathbf{t}_i^s \cdot \log \hat{\mathbf{t}}_i^s, \quad (6)$$

where for the i 'th input sample within the batch of N samples, \mathbf{t}_i^p and \mathbf{t}_i^s are the one-hot encoding of the Gleason grading for the primary pattern and the sum score, respectively, $\hat{\mathbf{t}}_i^p$ and $\hat{\mathbf{t}}_i^s$ are the predicted grading of the model.

2.4 Survival Model

In conjunction with clinical prognostic factors including the primary and secondary Gleason patterns, PSA, age, and tumor stage, computational biomarkers were fed into a Cox regression model^{46,47} for studying patient's RFS. In our study, we used RFS (months) as the time variable for a survival model. For high dimensional data, only those with Wald test^{57,58} p -value < 0.05 were selected and used in conjunction with clinical prognostic factors as input variables for the Cox regression model.

One of the most popular regression techniques for survival analysis is Cox proportional hazards regression, which is used to relate several risk factors or exposures, considered simultaneously, to assess differences in overall survival. In a Cox proportional hazards regression model, the measure of effect is the hazard ratio, which is the risk of failure (i.e., here is the risk or probability of the recurrence of the disease), given that the participant has survived up to a specific time. Given the patients (t_i, l_i, X_i) , where $i = 1, 2, \dots, N$, we have the t_i as the patient's recurrence time for individual i ; l_i as the label of the censored

data that equals 1 if the recurrence occurred at that time and 0 if the patient has been censored; X_i as the vector of covariates of the selected image features and clinical factors. The hazard function is the nonparametric part of the Cox proportional hazards regression function corresponding to

$$H(X_i, l_i, t_i) = H_0(t) \exp \sum_{j=1}^p x_{ij} \beta_j. \quad (7)$$

Here, x_{ij} is the computational biomarkers j for patient i , where $j = 1, 2, \dots, p$ and β_j is the Cox regression parameter for each patient. Here, H_0 is the baseline hazard function. The hazard ratio is derived from $HR(X_i) = \frac{H(X_i, l_i, t_i)}{H_0}$, representing the relative risk of instant failure for patients having the predictive value X_i compared to the ones having the baseline values. Here, d_i is weighting parameters for each patient:

$$HR(X_i) = \sum_i^N d_i \left\{ X_i \beta_i - \log \left[\sum_j^p I(t_j - t_i) \exp(X_i \beta_i) \right] \right\}. \quad (8)$$

In the study, we assessed the computational biomarkers in conjunction with other clinical prognostic factors by their recurrence hazard ratios and concordance indices (C -index).^{59,60} The hazard ratio and C -index both are global indices for validating the predictive ability of prognostic features of a given survival model. Under a given survival model, higher values mean that prognostic features predict higher risks and probabilities of survival for higher observed survival times. In our study, we examined RFS; the higher the hazard ratio and C -index, the greater the likelihood of disease recurrence.

3 Experiments and Results

In this section, we validated our approach on a publicly available prostate cancer dataset from the GDC data portal. The experimental results showed that the computational biomarkers discovered by the proposed method were effective for recurrence correlation for patients with Gleason score 7.

3.1 Implementation Details

The training process of our network included two steps. We first trained the CNN using minibatch Stochastic gradient descent with batch size as 32, momentum as 0.9, and weight-decay as 5×10^{-5} . The initial learning rate was 10^{-3} and annealed by 0.1 after every 10,000 iterations. We trained the CNN for total of 50,000 iterations until the loss converge. Then, we utilized the genomic data to train the MLP to extract the genomic features and used image and genomic features to train LSTM. We kept the same momentum, weight-decay, and learning rate except, we annealed the learning rate by 0.1 after every 2000 iterations and trained the network for a total of 5000 iterations. The implementation was based on Caffe toolbox.⁶¹

3.2 Pathway Analysis

Multiple studies have shown that gene expression signatures reflect the activation status of oncogenic pathways irrespective of specific mutations driving signaling.²⁴⁻²⁶ Thus, we examined genomic-based patterns of oncogenic pathway activity in prostate cancer patients with Gleason score 7 using a panel of previously published 265 gene expression signatures.

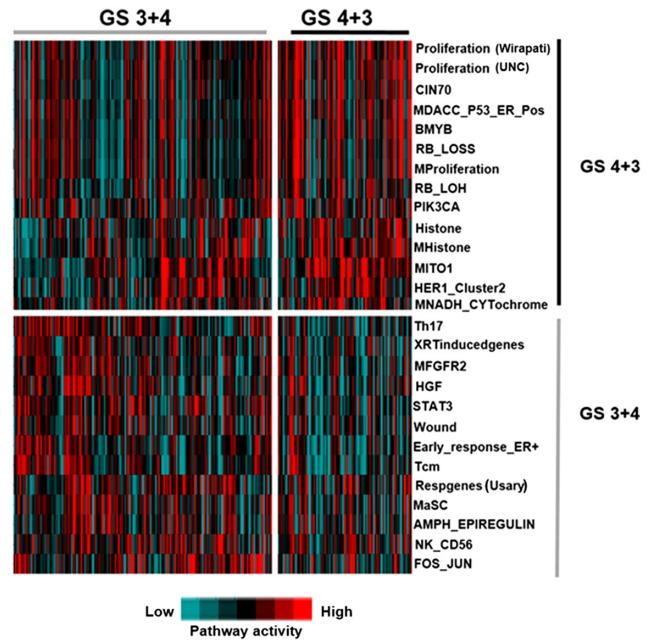


Fig. 4 Differential patterns of pathway activity in Gleason score 3 + 4 and 4 + 3 prostate tumors. Comparative analysis of Gleason score 4 + 3 ($n = 101$) and Gleason score 3 + 4 ($n = 146$) tumors identified 27 significantly altered signaling pathways (t -test, $p < 0.01$) as defined by mRNA-based gene expression signature scores. Tumors with a Gleason score 4 + 3 showed higher proliferation, BMYB, RB-LOH, and histone modification signature scores while tumors with a Gleason score 3 + 4 showed higher levels of immune system related pathway signatures including Th17 cells, Tcm, and STAT3.

In order to qualitatively assess unique patterns of pathway activity that define the 4 + 3 and 3 + 4 subset of Gleason score seven tumors, pathway signatures in each group, using all tumors across the entire cohort (i.e., training, test, and validation tumors) were assessed by a Student's two tailed t -test. Significant pathway scores were clustered using Cluster 3.0⁶² and visualized by Java TreeView.⁶³ Quantitative assessment of patterns of pathway activity of Gleason score 4 + 3 and 3 + 4 subgroups is shown in Fig. 4, which displayed a heatmap identifying 27 differentially expressed signatures ($p < 0.01$). Of these, we determined that 14 signatures including three unique proliferation signatures (Wirapati,⁶⁴ UNC,⁶⁵ and murine proliferation⁶⁵) as well as several proliferation-associated signatures predicative of BMYB,⁶⁶ RB-LOSS,⁶⁷ PIK3CA,⁶⁸ and HER1⁶⁹ signaling were significantly higher in patients with Gleason score 4 + 3. We further determined that 13 signatures were upregulated in Gleason 3 + 4 patients including immune systems signatures associated with Th17 cells,⁷⁰ Tcm,⁷⁰ NK-CD56,⁷⁰ HGF,⁷¹ and STAT3²⁶ signaling. Consistent with our findings, many studies report⁷²⁻⁷⁴ that Gleason 3 + 4 tumors have a better prognosis than Gleason 4 + 3 tumors, which would correlate with relatively higher levels of proliferation as well as lower levels of immune-related signaling evident in Gleason 4 + 3 tumors compared to Gleason 3 + 4 samples.

3.3 Integrated Recurrence Analysis in Conjunction with Clinical Factors

3.3.1 Image data on recurrence analysis

For the integrated recurrence analysis using a survival model, we first conducted the experiments where only the WSIs of

Table 1 Recurrence hazard ratios and corresponding *C*-indices of clinical prognostic factors and different image features from various image quantification methods. The results are obtained by using image features quantified from the WSIs. LBP, HOG, and SURF are the texture methods. CNN-LSTM is using the image features obtained from CNN with LSTM while CNN-Only is using the image features obtained from CNN without considering patches' spatial relation on a WSI.

Methods	Primary pattern	Secondary pattern	PSA	Age	Tumor stage	Image features	<i>C</i> -index
LBP	1.05	0.94	0.85	1.00	1.03	1.05	0.68
HOG	1.04	0.94	0.85	1.00	1.03	1.05	0.64
SURF	1.07	0.97	0.86	1.00	1.03	1.05	0.61
CNN-Only	1.11	1.12	0.80	1.00	1.17	2.44	0.70
CNN-LSTM	1.70	1.06	0.80	0.99	1.26	5.06	0.71

tissue slides were used. Thus, the networks were trained without integrating the genomic features. This setting of experiment is denoted as CNN-LSTM. We also considered the setting that only CNN was applied on the image patches without considering their spatial relation on a WSI and the image features were extracted from the second to the last layer of AlexNet. The setting is denoted as CNN-Only. To compare the effectiveness of the feature extraction from the images, we applied three texture feature methods including SURF,⁷⁵ HOG,⁷⁶ and LBP⁷⁷ on the WSIs to obtain image features. The image features were concatenated with clinical prognostic factors as multivariate inputs of the Cox regression model. During each iteration, each image feature in conjunction with clinical factors was fed into the Cox regression model to calculate the corresponding hazard ratios and *C*-indices. The survival model implementation was based on an R survival package.⁷⁸

The maximum hazard ratios of recurrence of image features in conjunction with clinical factors are shown in Table 1. Within our study, we used the disease RFS times as the time variable in the Cox regression model, the higher values of hazard ratio and *C*-index of the features indicated that the image features had the higher correlations with the disease recurrence and progression. From the result of using texture features, there were no significance differences among LBP, HOG, and SURF for recurrence ratios. CNN-LSTM analysis determined that image features identified by computational image analysis outperformed other texture features and CNN-Only with higher hazard ratio and *C*-index. When conjunction with CNN-LSMT, the primary pattern still showed greater hazard ratio and *C*-index relative to those identified using other methods.

3.3.2 Image and genomic data on recurrence analysis

Before integrating image features and pathway scores, we first analyzed the correlation between them. Because the number of image features and the number of pathway scores were different, to calculate their correlation coefficients, we randomly chose the same number of image features paired with the same number of pathway scores and repeated the process *N* times until all image features had been paired. Here, the image features included features quantified from texture methods (LBP, HOG, and SURF) and CNN-LSTM. Using a *t*-test on correlation coefficients, the mean and standard deviation of *p*-values is shown in Table 2. Because *p*-value >0.05, there was no significant correlations between image features and pathway scores. This showed that the two types of data provided complementary information for

prostate cancer diagnosis and prognosis. It was reasonable to integrate image and genomic data together for predicting patients' recurrence.

Then, we showed the experimental results by combining image features obtained from WSIs and the genomic features obtained from the pathway scores. We utilized all 265 gene expression signatures integrated with image data to identify the computational biomarkers as shown in Fig. 2. The setting was denoted as CNN-LSTM + PS. We also considered the setting where LSTM was deactivated when obtained the biomarkers from image and genomic data. We denoted the approach as CNN-Only + PS. The methods using texture features obtained from WSIs together with pathway scores for the recurrence analysis were denoted as LBP-PS, HOG-PS, and SURF-PS. We also considered only using pathway scores with clinical factors together as the input of the Cox regression model and denote it as PS. The maximum hazard ratios of the computational biomarkers from WSIs and pathway scores in conjunction with clinical factors are shown in Table 3.

Compared with other clinical factors, using pathway scores alone achieved equivalent hazard ratio. For the texture methods, the recurrence hazard ratios were equivalent to the ones without pathway scores. Using CNN-LSTM + PS, the Gleason primary pattern and computational biomarkers showed the increased recurrence ratios compared to the ones without pathway scores. In addition, the Gleason primary pattern and computational biomarkers showed the highest recurrence ratios compared to other clinical factors. The result showed that CNN-LSTM-PS outperformed other methods in prostate cancer recurrence analysis due to its highest recurrence hazard ratio.

Furthermore, we show the *C*-index of the clinical factors and computational features under the Cox regression model for prostate cancer recurrence probability prediction in the last column

Table 2 Correlation analysis of image features and pathways scores using a test-test on their correlation coefficients.

Image features	Mean of <i>p</i> -value	Standard deviation of <i>p</i> -value
LBP	0.50	0.29
HOG	0.49	0.30
SURF	0.43	0.30
CNN-LSTM	0.48	0.29

Table 3 Recurrence hazard ratios and corresponding *C*-indices of clinical prognostic factors and computational biomarkers under a Cox regression model using different image feature quantification methods along with the genomic data. Given the genomic data, we show the results using image features with pathway scores (PS). Here, LBP + PS, HOG + PS, SURF + PS, CNN-Only + PS, and CNN-LSTM + PS are image features quantified from LBP, HOG, SURF, CNN-Only, and CNN-LSTM methods with PS.

Methods	Primary pattern	Secondary pattern	PSA	Age	Tumor stage	Biomarkers	<i>C</i> -index for biomarkers
PS	0.95	0.98	0.86	1.00	1.04	1.02	0.65
LBP + PS	1.04	1.00	0.87	1.00	1.02	1.08	0.69
HOG + PS	1.04	1.00	0.87	1.00	1.02	1.08	0.65
SURF + PS	1.07	1.00	0.86	1.00	1.03	1.07	0.62
CNN-Only + PS	1.13	1.11	0.80	1.00	1.17	2.58	0.71
CNN-LSTM + PS	2.56	0.63	0.66	1.01	1.05	5.73	0.74
<i>C</i> -index for clinical factors	0.61	0.59	0.66	0.55	0.53	—	—

of Table 1 and the last row and column of Table 3. As a global index for validating the predictive ability of a survival model, in our study, the *C*-index was equivalent to a rank correlation of the risk of a recurrence of disease. High values mean that the model predicts higher probabilities of recurrence for higher observed recurrence times. From the clinical results, PSA showed higher *C*-index values, which were correlated to a higher recurrence prediction probability compared to other clinical factors. Interestingly, texture features on WSIs or pathway scores individually showed an equivalent recurrence probability.

4 Discussion

From the experimental results, our proposed method achieved the highest recurrence hazard ratio and the strongest *C*-index related to prostate cancer recurrence probability compared to other clinical prognostic factors and methods. It demonstrated that the approach was beneficial for recurrence analysis on the patients with Gleason score 7. The unified WSIs and genomic data analysis through the proposed networks could be applied to other prostate cancer risk group such as Gleason 6⁷⁹⁻⁸¹ or other cancer recurrence analysis, such as breast cancer.⁸²

Pathway analysis, albeit with the caveat of a small sample size, identified 27 differentially expressed pathway activities in tumors with Gleason score 3 + 4 and 4 + 3. Thus, these signatures could be utilized to differentiate patients with Gleason score 7 as two subgroups, which corresponds with a favorable or unfavorable prognosis.⁸³ The recurrence analysis (Table 3) using pathway scores alone did not show an advantage over other clinical prognostic factors. The integration of pathway score with WSIs achieved the best recurrence prediction on patients with Gleason score 7. The comparison indicated that using the pathway scores directly had a limited contribution in recurrence prediction on patients with Gleason score 7. However, the embedded genomic features obtained through MLP were more effective for prostate cancer recurrence analysis.

There are other clinical factors for prostate cancer prognosis besides those used in the study, such as patients' race. Because in the study, <2% men were Asian or African, 30% were Caucasian, and the rest were unknown, we excluded patients' race factor in the recurrence analysis. Other clinical factors, such as American joint committee on cancer metastasis stage, neoplasm disease stage codes, and so on, were not available for all the patients in the GDC prostate cancer datasets.

The prostate cancer datasets were acquired from various institutions and each institution may have different scanners or WSI scanning protocols. Thus, there was color heterogeneity among the prostate cancer WSIs. In the study, we did not adopt color normalization^{84,85} on the randomly selected testing set because it was not feasible to determine the reference image from the training set for color normalization. When we apply the approach to a new dataset, we could fine-tune the network based on the training data from that dataset. Given the limited size of the public prostate dataset, the results achieved from our experiments were preliminary. In order to further validate the generalizability of our approach on a wider population of prostate cancer patients, we will collect more prostate images from local institutions to perform extensive experiments.

5 Conclusion

In this study, we performed recurrence analyses for prostate cancer patients with Gleason score 7 integrating histopathology WSIs and genomic data. The image features and genomic features were obtained using CNN and MLP, respectively. The combination of the features was modeled using LSTM to get the computational biomarkers. Experimental results utilizing on publicly available prostate cancer dataset showed that the computational biomarkers extracted by our approach were more closely correlated with patients recurrence risk compared to standard clinical prognostic factors and engineered image texture features. The results of our study suggest that these approaches could be utilized to predict recurrence and progression for patients with prostate cancer.

Disclosures

Dr. Singer is the principal investigator on an investigator-initiated clinical trial that is funded by Astellas/Medivation (NCT02885649).⁸⁶ The other authors declare that they have no competing interests. The public prostate GDC data have ethics approval with the NIH/NCI, which can be accessible from Ref. 48.

Acknowledgments

This research was funded, in part, by grants from NIH contracts 4R01LM009239-08, 4R01CA161375-05, and 1UG3CA225021-01, and P30CA072720.

References

- R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: Cancer J. Clin.* **68**(1), 7–30 (2018).
- M. J. Zelefsky et al., "Five-year biochemical outcome and toxicity with transperineal ct-planned permanent i-125 prostate implantation for patients with localized prostate cancer," *Int. J. Radiat. Oncol. Biol. Phys.* **47**, 1261–1266 (2000).
- A. V. D'amico et al., "Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer," *J. Am. Med. Assoc.* **280**(11), 969–974 (1998).
- D. F. Gleason and G. T. Mellinger, "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging," *J. Urol.* **167**(2), 953–958 (2002).
- H. Carter, "Active surveillance for prostate cancer: an underutilized opportunity for reducing harm," *J. Natl. Cancer Inst. Monogr.* **2012**(45), 175–183 (2012).
- S. Ip et al., "An evidence review of active surveillance in men with localized prostate cancer," *Evid. Rep. Technol. Assess.* (204), 1–341 (2011).
- J. L. Wright et al., "Differences in prostate cancer outcomes between cases with Gleason 4 + 3 and Gleason 3 + 4 tumors in a population-based cohort," *J. Urol.* **182**(6), 2702–2707 (2009).
- A. Amin, A. Partin, and J. I. Epstein, "Gleason score 7 prostate cancer on needle biopsy: relation of primary pattern 3 or 4 to pathological stage and progression after radical prostatectomy," *J. Urol.* **186**(4), 1286–1290 (2011).
- M. J. Burdick et al., "Comparison of biochemical relapse-free survival between primary Gleason score 3 and primary Gleason score 4 for biopsy Gleason score 7 prostate cancer," *Int. J. Radiat. Oncol. Biol. Phys.* **73**(5), 1439–1445 (2009).
- J. R. Stark et al., "Gleason score and lethal prostate cancer: does 3 + 4 = 4 + 3?" *J. Clin. Oncol.* **27**(21), 3459–3464 (2009).
- S. Doyle et al., "Automated grading of prostate cancer using architectural and textural image features," in *4th IEEE Int. Symp. on Biomedical Imaging: from Nano to Macro, ISBI 2007*, IEEE, pp. 1284–1287 (2007).
- D. V. Makarov et al., "Updated nomogram to predict pathologic stage of prostate cancer given prostate-specific antigen level, clinical stage, and biopsy Gleason score (Partin tables) based on cases from 2000 to 2005," *J. Urol.* **69**, 1095–1101 (2007).
- A. Matoso and J. I. Epstein, "Grading of prostate cancer: past, present and future," *Curr. Urol. Rep.* **17**, 25 (2016).
- P. A. Rodriguez-Urrego et al., "Interobserver and intraobserver reproducibility in digital and routine microscopic assessment of prostate needle biopsies," *Hum. Pathol.* **42**, 68–74 (2011).
- P. C. Walsh, "The Gleason grading system: a complete guide for pathologists and clinicians," *J. Urol.* **189**, 1173 (2013).
- P.-W. Huang and C.-H. Lee, "Automatic classification for pathological prostate images based on fractal analysis," *IEEE Trans. Med. Imaging* **28**(7), 1037–1050 (2009).
- S. Doyle et al., "A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans. Biomed. Eng.* **59**(5), 1205–1218 (2012).
- A. Tabesh et al., "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imaging* **26**(10), 1366–1378 (2007).
- P. Khurd et al., "Computer-aided Gleason grading of prostate cancer histopathological images using texton forests," in *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, IEEE, pp. 636–639, (2010).
- J. Ren et al., "Computer aided analysis of prostate histopathology images Gleason grading especially for Gleason score 7," in *37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp. 3013–3016 (2015).
- J. Ren et al., "Computer aided analysis of prostate histopathology images to support a refined Gleason grading system," *Proc. SPIE* **10133**, 101331V (2017).
- G. Lee et al., "Supervised multi-view canonical correlation analysis (SMVCCA): integrating histologic and proteomic features for predicting recurrent prostate cancer," *IEEE Trans. Med. Imaging* **34**(1), 284–297 (2015).
- C. Sotiriou and M. J. Piccart, "Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care?" *Nat. Rev. Cancer* **7**(7), 545–553 (2007).
- M. L. Gatzka et al., "A pathway-based classification of human breast cancer," *Proc. Natl. Acad. Sci. U. S. A.* **107**(15), 6994–6999 (2010).
- A. H. Bild et al., "Oncogenic pathway signatures in human cancers as a guide to targeted therapies," *Nature* **439**(7074), 353–357 (2006).
- M. L. Gatzka et al., "An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer," *Nat. Genet.* **46**(10), 1051–1059 (2014).
- J. Ren et al., "Differentiation among prostate cancer patients with Gleason score of 7 using histopathology image and genomic data," *Proc. SPIE* **10579**, 1057904 (2018).
- X. Pan et al., "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing* **229**, 88–99 (2017).
- S. Naik et al., "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *5th IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, ISBI 2008*, IEEE, pp. 284–287 (2008).
- J. Xu et al., "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing* **191**, 214–223 (2016).
- N. Ing et al., "Semantic segmentation for prostate cancer grading by convolutional neural networks," *Proc. SPIE* **10581**, 105811B (2018).
- L. Hou et al., "Automatic histopathology image analysis with CNNs," in *Scientific Data Summit (NYSDDS)*, New York, IEEE, pp. 1–6 (2016).
- A. Cruz-Roa et al., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," *Proc. SPIE* **9041**, 904103 (2014).
- L. Hou et al., "Patch-based convolutional neural network for whole slide tissue image classification," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2424–2433 (2016).
- S. Kothari et al., "Pathology imaging informatics for quantitative analysis of whole-slide images," *J. Am. Med. Inf. Assoc.* **20**(6), 1099–1108 (2013).
- V. Roullier et al., "Multi-resolution graph-based analysis of histopathological whole slide images: application to mitotic cell extraction and visualization," *Comput. Med. Imaging Graphics* **35**(7), 603–615 (2011).
- R. Toth et al., "Histostitcher: an informatics software platform for reconstructing whole-mount prostate histology using the extensible imaging platform framework," *J. Pathol. Inf.* **5**(1), 1–9 (2014).
- H. Wang et al., "Novel image markers for non-small cell lung cancer classification and survival prediction," *BMC Bioinf.* **15**(1), 310 (2014).
- J. Yao et al., "Imaging biomarker discovery for lung cancer survival prediction," in *Int. Conf. on Medical Image Computing Computer-Assisted Intervention*, Springer, pp. 649–657 (2016).
- K.-H. Yu et al., "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features," *Nat. Commun.* **7**, 1–10 (2016).
- X. Zhu, J. Yao, and J. Huang, "Deep convolutional neural network for survival analysis with pathological images," in *IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, IEEE, pp. 544–547 (2016).
- X. Zhu et al., "Lung cancer survival prediction from pathological images and genetic data—an integration study," in *IEEE 13th Int. Symp. on Biomedical Imaging (ISBI)*, IEEE, 1173–1176 (2016).
- J. Ren et al., "Adversarial domain adaptation for classification of prostate histopathology whole-slide images," arXiv:1806.01357 (2018).
- Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Networks* **5**(2), 157–166 (1994).
- S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.* **9**(8), 1735–1780 (1997).
- D. R. Cox and D. Oakes, *Analysis of Survival Data*, Vol. **21**, CRC Press, Boca Raton, Florida (1984).
- D. G. Kleinbaum and M. Klein, *Survival Analysis*, Vol. **3**, Springer-Verlag, New York (2010).
- R. L. Grossman et al., "Toward a shared vision for cancer genomic data," *N. Engl. J. Med.* **375**(12), 1109–1112 (2016).
- C. Kandath et al., "Mutational landscape and significance across 12 major cancer types," *Nature* **502**(7471), 333–339 (2013).
- C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.* **18**(8), 1016–1022 (1979).

51. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012).
52. A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *IEEE Int. Conf. on Acoustics, speech and signal processing (ICASSP)*, IEEE, pp. 6645–6649 (2013).
53. A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proc. of the 31st Int. Conf. on Machine Learning (ICML-14)*, pp. 1764–1772 (2014).
54. I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems*, pp. 3104–3112 (2014).
55. J. Donahue et al., "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2625–2634 (2015).
56. Z. Wu et al., "Modeling spatial-temporal clues in a hybrid deep learning framework for video classification," in *Proc. of the 23rd ACM Int. Conf. on Multimedia*, ACM, pp. 461–470 (2015).
57. J. D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey (1994).
58. R. Davidson and J. G. MacKinnon, *Econometric Theory and Methods*, Oxford University Press, Oxford, England, United Kingdom (2004).
59. T. A. Gerds et al., "Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring," *Stat. Med.* **32**(13), 2173–2184 (2013).
60. M. Wolbers et al., "Concordance for prognostic models with competing risks," *Biostatistics* **15**(3), 526–539 (2014).
61. Y. Jia et al., "Caffe: convolutional architecture for fast feature embedding," in *Proc. of the 22nd ACM Int. Conf. Multimedia*, ACM, pp. 675–678 (2014).
62. M. B. Eisen et al., "Cluster analysis and display of genome-wide expression patterns," *Proc. Natl. Acad. Sci. U. S. A.* **95**(25), 14863–14868 (1998).
63. Y. Zhou and J. Liu, "Ava: visual analysis of gene expression microarray data," *Bioinformatics* **19**(2), 293–294 (2003).
64. P. Wirapati et al., "Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures," *Breast Cancer Res.* **10**(4), R65 (2008).
65. C. Fan et al., "Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures," *BMC Med. Genomics* **4**(1), 3 (2011).
66. A. Thorner et al., "In vitro and in vivo analysis of B-Myb in basal-like breast cancer," *Oncogene* **28**(5), 742–751 (2009).
67. J. I. Herschkowitz et al., "The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas," *Breast Cancer Res.* **10**(5), R75 (2008).
68. J. E. Huttli et al., "Oncogenic PI3K mutations lead to NF- κ B-dependent cytokine expression following growth factor deprivation," *Cancer Res.* **72**, 3260–3269 (2012).
69. K. A. Hoadley et al., "EGFR associated expression profiles vary with breast tumor subtype," *BMC Genomics* **8**(1), 258 (2007).
70. G. Bindea et al., "Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer," *Immunity* **39**(4), 782–795 (2013).
71. P. Casbas-Hernandez et al., "Role of HGF in epithelial-stromal cell interactions during progression from benign breast disease to ductal carcinoma in situ," *Breast Cancer Res.* **15**(5), R82 (2013).
72. G. Lee et al., "Nuclear shape and architecture in benign fields predict biochemical recurrence in prostate cancer patients following radical prostatectomy: preliminary findings," *Eur. Urol. Focus* **3**, 457–466 (2016).
73. P. Leo et al., "Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images," *J. Med. Imaging* **3**(4), 047502 (2016).
74. A. Madabhushi et al., "Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data," *Comput. Med. Imaging Graphics* **35**(7–8), 506–514 (2011).
75. H. Bay, "Surf: speeded up robust features," *Comput. Vision Image Understanding* **110**(3), 346–359 (2008).
76. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, CVPR 2005*, IEEE, Vol. **1**, pp. 886–893 (2005).
77. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002).
78. M. S. Schröder et al., "survcomp: an r/bioconductor package for performance assessment and comparison of survival models," *Bioinformatics* **27**(22), 3206–3208 (2011).
79. Y.-S. Ha et al., "Increased incidence of pathologically nonorgan confined prostate cancer in African-American men eligible for active surveillance," *Urology* **81**(4), 831–836 (2013).
80. H. B. Carter et al., "Gleason score 6 adenocarcinoma: should it be labeled as cancer?" *J. Clin. Oncol.* **30**(35), 4294–4296 (2012).
81. E. A. Singer et al., "Active surveillance for prostate cancer: past, present and future," *Curr. Opin. Oncol.* **24**(3), 243–250 (2012).
82. A. H. Beck et al., "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci. Transl. Med.* **3**(108), 108ra113 (2011).
83. A. C. Raldow et al., "Risk group and death from prostate cancer: implications for active surveillance in men with favorable intermediate-risk prostate cancer," *JAMA Oncol.* **1**(3), 334–340 (2015).
84. A. M. Khan et al., "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Trans. Biomed. Eng.* **61**(6), 1729–1738 (2014).
85. X. Li and K. N. Plataniotis, "A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics," *IEEE Trans. Biomed. Eng.* **62**(7), 1862–1873 (2015).
86. E. Singer, "A phase 0 study of the blockade of androgens in renal cell carcinoma using Enzalutamide (BARE)," <http://clinj.org/clinical-trials/index?show=trial&p=081604> (2018).

Jian Ren is a PhD candidate of electrical and computer engineering at Rutgers University. He received his BS degree in electronic science and technology from the University of Science and Technology of China in 2014.

David J. Foran is CIO and executive director of computational imaging and biomedical informatics at Rutgers Cancer Institute of New Jersey. He also serves as professor and chief medical informatics officer of pathology, Laboratory Medicine and Radiology at Rutgers-Robert Wood Johnson Medical School. His research focuses on the design, development, and implementation of new approaches in computer-assisted diagnostics, medical imaging, and precision medicine for resolving challenging clinical problems in pathology, radiology, and oncology.

Xin Qi is assistant professor of Department of Pathology and Laboratory Medicine at Rutgers Cancer Institute of New Jersey. She also serves as adjunct research assistant professor of radiology at Robert Wood Johnson University Hospital. She received her BS degree in precision instrument and opto-electronics engineering at Tianjin University and her master's and PhD in biomedical engineering at Case Western Reserve University.

Biographies for the other authors are not available.