Using Protein Expressions to Predict Survival in Clear Cell Renal Carcinoma

Hyung L. Kim, David Seligson, Xueli Liu, Nicolette Janzen, Matthew H. T. Bui, Hong Yu, Tao Shi, Robert A. Figlin, Steve Horvath, and Arie S. Belldegrun

Departments of 1 Urology, 2 Pathology and Laboratory Medicine, 3 Biostatistics, 4 Human Genetics and Biostatistics, and 5 Medicine and Urology, University of California School of Medicine, Los Angeles, California

ABSTRACT

Purpose: An accurate system for predicting survival for patients with solid tumors will allow for better patient selection for both established and novel therapies. We propose a staging system for clear cell variants of renal cell carcinoma (RCC) that includes molecular predictors and standard clinical predictors such as tumor-node-metastasis (TNM) stage, histological grade, and performance status (PS).

Experimental Design: A custom tissue array was constructed using clear cell RCC from 318 patients, representing all stages of localized and metastatic RCC, and immunohistochemically stained for molecular markers Ki67, p53, gelsolin, CA9, CA12, PTEN, EpCAM, and vimentin. We present a strategy for evaluating individual candidate markers for prognostic information and integrating informative markers into a multivariate prognostic system.

Results: The overall median follow-up and the median follow-up for surviving patients were 28 and 55 months, respectively. A prognostic model based primarily on molecular markers included metastasis status, p53, CA9, gelsolin, and vimentin as predictors and had high discriminatory power: its statistically validated concordance index (C-index) was found to be 0.75. A prognostic model based on a combination of clinical and molecular predictors included metastasis status, T stage, Eastern Cooperative Oncology Group PS, p53, CA9, and vimentin as predictors and had a C-index of 0.79, which was significantly higher (P < 0.05) than that of prognostic models based on grade alone (C = 0.65), TNM stage alone (C = 0.73), or the University of California Los Angeles integrated staging system (C = 0.76).

Conclusions: Protein expressions obtained using widely available technology can complement standard clinical predictors such as TNM stage, histological grade, and PS.

INTRODUCTION

Clear cell carcinomas of the kidney represent over 80% of renal cell carcinomas [RCCs (1)]. The treatment of choice for localized RCC is a nephrectomy; however, approximately 20% of all patients surgically treated with curative intent will ultimately experience disease recurrence (2, 3). Furthermore, approximately 30% of patients will present with metastatic disease. RCC is generally resistant to conventional oncologic therapies, and patients with systemic disease are treated with immunotherapy, which produces response rates of approximately 15–30% (4–7). As a result of the modest response rates, RCC has become a model for testing novel immunotherapeutic strategies. An enhanced ability to predict patient survival will allow for better selection of patients most likely to benefit from systemic therapies and for more accurate comparison of clinical trials based on varying inclusion criteria.

In general, prior attempts to predict patient survival have relied on traditional clinical parameters such as tumor stage and grade (3, 8). More recently, methods based on gene arrays, which screen for differential expression of thousands of genes, have identified large numbers of potential prognostic markers (9, 10). Our study, evaluating protein expression in a high-throughput tissue array, is a natural extension to the efforts for molecular staging. All of the markers examined in this study were selected based on previous reports linking the markers to the development of malignancies. Ki67 and p53 are related to poor prognosis in RCC (20, 21). Vimentin, an intermediate filament, has been shown to be an independent predictor of survival in some studies (14, 15).

Gelsolin, EpCAM (epithelial cell adhesion molecule), and vimentin may be involved in cell motility and cancer progression. Gelsolin is the most potent protein known that functions to sever actin during cell motility (16), and it has been described as a highly significant indicator of poor prognosis in non-small-cell lung cancer (17). EpCAM is widely expressed on the surface of many carcinomas (18, 19). Vimentin, an intermediate filament, has previously been identified as an independent predictor of poor prognosis in RCC (20, 21).

CA9, CA12, and PTEN (phosphatase and tensin homologue deleted from chromosome 10) are critical components of the hypoxia pathway, which allows enlarging tumors to adapt to an oxygen-poor microenvironment. In addition, overexpression of CA9 and CA12 is a direct consequence of a VHL mutation, found in over 75% of sporadic clear cell RCC (22, 23). De-
creased expression of CA9 has previously been shown to predict worse survival (24). PTEN is a phosphatase that regulates cellular migration, proliferation, and apoptosis (25, 26). Although PTEN mutation may be a rare event in RCC (27, 28), PTEN deletion has been shown to correlate with poor prognosis (27).

With the explosive growth of available genomic and proteomic data, the number of markers that have been correlated with prognosis is continuously increasing. We present a strategy for evaluating individual candidate markers for prognostic information and integrating informative markers into a multivariate prognostic system. Levels of marker expression were used to develop two prognostic models for predicting disease-specific survival (DSS) in patients with clear cell RCC. The accuracy of these models was compared with traditional, clinical parameters used for predicting survival.

MATERIALS AND METHODS

Patients. The patient cohort included 318 patients treated with a partial or radical nephrectomy for clear cell RCC between 1989 and 1999. After approval by the institutional review board (KCP 99-233), immunohistochemical studies were performed, and clinical data from an established kidney cancer database were reviewed. Patients were staged using radiographic studies and postoperative pathological data, according to the 1997 tumor-node-metastasis (TNM) criteria proposed by the American Joint Committee on Cancer (29). Performance status (PS) was determined using the Eastern Cooperative Oncology Group (ECOG) PS scale (30). Patients with M stage = 1 or N stage > 0 were considered to have metastatic disease. Tumor grade was categorized using Fuhrman grade (31). The patients were also categorized according to the University of California Los Angeles integrated staging system [UISS (8)].

After surgery, patients were evaluated for disease recurrence by physical examination, liver function tests, chest X-ray, and computerized tomography of the abdomen/pelvis every 6–12 months. The primary outcome of interest was DSS. Patients with metastatic RCC were treated with cytoreductive nephrectomy (n = 155) followed by interleukin-2-based immunotherapy (n = 116) and then evaluated radiologically in 3 months to determine the response to therapy. A positive response was defined as a complete or partial response.

Tissue Array Construction. Archival tumor specimens from the cohort of 318 patients were obtained from the Department of Pathology. All tumors were reviewed to confirm clear cell histology. Three core tissue biopsies, each 0.6 mm in diameter, were taken from selected, morphologically representative regions of each paraffin-embedded renal tumor and precisely arrayed using a custom-built instrument as described previously (32).

Immunohistochemistry. Array sections were deparaffinized and immunostained using either the DAKO Envision Plus (DAKO, Carpinteria, CA) or Vector ABC Elite (Vector Laboratories, Burlingame, CA) staining systems as described previously (33). Antigen retrieval involved pepsin digestion for 10 min for EpCAM staining, treatment in a pressure cooker for 3 min for Ki67, and heat treatment for 25 min for all other markers. Ki67 staining was performed with mouse monoclonal antibody MIB-1 (DAKO) at 0.5 μg/ml. Gelsolin staining was performed with monoclonal mouse antibody GS-2C4 (Sigma, St. Louis, MO) at 3.8 μg/ml. Immunostaining for PTEN was performed using rabbit polyclonal antibody PN37 (Zymed, South San Francisco, CA) at 2 μg/ml. Immunostaining for EpCAM was performed using monoclonal mouse antibody KS1/4BD (PharMingen, San Diego, CA) at 20 μg/ml. Immunostaining for p53 was performed with monoclonal mouse antibody DO–7 (DAKO) at a 1:100 dilution. CA9 staining was performed with monoclonal mouse antibody M75 (a gift from Dr. Eric Stanbridge, University of California at Irvine, Irvine, CA) at a 1:25,000 dilution of stock. CA12 staining was performed with a rabbit polyclonal antibody (a gift from Dr. Michael Lerman, National Cancer Institute, Bethesda, MD) at a 1:450 dilution. Vimentin staining was performed with mouse monoclonal antibody V9 (DAKO) at a 1:1000 dilution.

Immunostaining was scored by recording the percentage of cells staining and scoring the area of maximum staining on a 4-point scale, with 0 representing no staining and 3 representing the highest staining. All three cores that were arrayed per tumor were scored and pooled to yield the mean, median, minimum, and maximum values for the three cores.

Statistical Analysis. Length of follow-up was the time from nephrectomy to date of death or last contact. The end point of interest was DSS. Kaplan-Meier curves were generated to visualize survival rates, and we used Cox proportional hazards models to relate DSS to molecular and clinical predictors. The Schoenfeld residual test was used to evaluate the proportional hazard assumptions. All computations were performed with the R statistical software using the Design, Hmisc, Rpart, and RandomForest libraries. P < 0.05 was considered significant, and P > 0.10 was the criterion when performing Cox regression backward step-down variable deletions.

The predictive accuracy of various Cox regression models was quantified by calculating the concordance index (C-index), which provides the area under the receiver operating characteristics curve for censored data (34, 35). A C-index of 0.5 indicates that outcomes are completely random, whereas a C-index of 1 indicates that the model is a perfect predictor. To protect against overfitting during stepwise regression, we used a bootstrap procedure as implemented in the “validate” function of the Design library (34), which allowed for computation of an unbiased estimate of the C-index. We used 500 bootstrap samples.

To test whether the difference in statistical accuracy between non-nested Cox regression models is significant, we used the rcorrp.cens function in the Design library (34). This computes U-statistics for testing whether the predictions of one model are more concordant than those of another model, extending the C-index. Specifically, the fraction of pairs was determined for which one model correctly selects the patient with the longer DSS when the competing model did not. Before calculating the C-index or using the rcorrp.cens function, TNM staging system (T stage, N stage, and M stage) and UISS (T stage, N stage, M stage, ECOG PS, and grade) were fitted to our data.

To account for potential overfitting resulting from selection of the staining criterion and cutoff used to stratify DSS, we used a prevalence method described in Liu et al. (36) Briefly, the prevalence method is carried out in three steps for each marker. First, an intercept only Cox regression model is fit to the
survival times to arrive at deviance residuals, which are normalized transforms of martingale residuals. Second, random regression forest predictors are fit where the outcome is the deviance residual and the covariates are the pooled marker stainings (37). Random regression forest predictors work well in the setting of many covariates but relatively few observations. Third, the random forest construction is used to define a tumor marker hazard score M as the out-of-bag estimate of the deviance residual. The hazard score M is a real number.

Liu et al. (36) have shown that when M is used as a covariate in a Cox regression model involving the same survival outcome that was used to construct it, it does not lead to an inflation of the false positive rate of the one-sided Wald test that tests whether coefficient is positive. Therefore, we used the corresponding M for each marker as the bias-corrected marker covariate in all univariate and multivariate Cox regression analyses in computing the C-index and in the rcorr.cens function.

The Cox proportional hazards regression analysis was used to construct two predictive models. One model was primarily based on molecular markers (marker model). A second predictive model was based on a combination of molecular markers and clinical variables (clinical/marker model). We showed previously (24) that CA9 is an independent predictor for survival in patients with metastatic RCC, but not localized RCC. Therefore, an interaction term for CA9 and metastasis status (Met*CA9) was used.

To visualize the relationship between clinical predictors and DSS, we constructed a nomogram for a Cox model that only contains significant terms and dichotomized marker stainings. Although dichotomized marker values were used, only markers shown to contain prognostic information using the prevalidation method were included in the nomogram. This nomogram should be considered as an approximation of the validated models described above. In particular, the nomogram is not used for model comparisons and inferences as described above. Rather, the nomogram is a descriptive tool that needs further validation with new and independent data.

For each molecular marker, the optimal cutoff for the staining scores to stratify DSS was determined using the default settings of the recursive partitioning function (RPART) in the freely available R statistical software. To avoid cutoffs that are overly sensitive to assay conditions, the optimal cutoffs were confirmed to be robust before they were used to construct the prognostic nomograms. Ki67 was considered positive if there was >15% nuclear staining at any intensity in the core with median staining, and p53 was positive if there was >15% staining in the core with maximum staining. Gelsolin was positive if there was any level of cytoplasmic staining in any core. PTEN was positive if there was >50% cytoplasmic staining in the core with lowest staining. EpCAM was considered positive if there was any staining in any core. CA9 and CA12 were considered positive if there was 100% staining in all cores and the mean percentage staining in the three cores was >80%, respectively. Vimentin was considered positive if the mean of the maximum staining intensity score in the three cores was <2.

Table 1 Characteristics of patients with clear cell RCC (N = 318)

<table>
<thead>
<tr>
<th>Patient characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>60</td>
</tr>
<tr>
<td>Median</td>
<td>61</td>
</tr>
<tr>
<td>Range</td>
<td>27–88</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>215</td>
</tr>
<tr>
<td>Female</td>
<td>103</td>
</tr>
<tr>
<td>Mean tumor size (cm)</td>
<td>7.4</td>
</tr>
<tr>
<td>Metastasis</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>155</td>
</tr>
<tr>
<td>No</td>
<td>163</td>
</tr>
<tr>
<td>1997 TNM stage</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>90</td>
</tr>
<tr>
<td>II</td>
<td>19</td>
</tr>
<tr>
<td>III</td>
<td>54</td>
</tr>
<tr>
<td>IV</td>
<td>155</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>156</td>
</tr>
<tr>
<td>3</td>
<td>113</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>ECOG PS</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>113</td>
</tr>
<tr>
<td>1</td>
<td>191</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>IL2-based therapy*</td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>38</td>
</tr>
<tr>
<td>No response</td>
<td>78</td>
</tr>
<tr>
<td>Deaths</td>
<td>186</td>
</tr>
<tr>
<td>Median follow-up (mo)</td>
<td></td>
</tr>
<tr>
<td>All patients</td>
<td>28</td>
</tr>
<tr>
<td>Surviving patients</td>
<td>55</td>
</tr>
<tr>
<td>UISS</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>128</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

* IL2, interleukin-2; response includes complete and partial responses.

Below, the use of dichotomized staining scores is discussed in more detail.

RESULTS

The patient characteristics are summarized in Table 1. The mean age is 60 years, and the male:female ratio is approximately 2:1. The mean tumor size is 7.4 cm, and 155 (49%) patients presented with metastatic disease. The overall median follow-up is 28 months; 186 patients have died, and the median follow-up for surviving patients is 55 months. UISS is a prognostic model that predicts DSS based on the interaction of TNM stage, Fuhrman grade, and ECOG PS (8). The most common UISS categories were group 4 (n = 128; 40%) and group 1 (n = 72; 23%). In a single variable analysis, increased immunohistochemical staining for Ki67, p53, vimentin, and gelsolin correlated with worse survival, whereas the inverse was true for CA9, PTEN, CA12, and EpCAM, with decreased staining correlating with worse survival. However, only gelsolin, p53, CA9, Ki67, and vimentin were statistically significant predictors of DSS in a univariate analysis (Table 2, left).
All markers were combined in a multivariable Cox regression analysis that included metastasis status as a covariate (Met) and as an interaction term for CA9. (Table 2, center) Met, gelsolin, p53, and Met*CA9 remained significant predictors of survival and were used to create a prognostic model (marker model). Covariates included in the marker model were retained by a backward stepwise elimination.

Using a similar approach, a prognostic model was constructed using a combination of clinical variables and marker data (clinical/marker model). In a multivariable Cox analysis, when each of the markers were controlled for T stage, Met, ECOG PS, and grade, we found that CA9, vimentin, and p53 were statistically significant predictors independent of the clinical variables. (Table 2, right). These three markers were combined in a multivariable Cox analysis that included Met, T stage, ECOG PS, and grade. Covariates retained in a Cox step-down analysis (Table 3) were used to construct a second prognostic model (clinical/marker model).

The predictive ability of various models was quantified by calculating the C-index, which is the area under the receiver operating characteristics curve adapted for survival data. (Table 4). The C-index was corrected for overfitting by a bootstrap procedure. A C-index of 0.5 indicates that outcomes are completely random, whereas a C-index of 1 indicates that the model is a perfect predictor. When using clinical variables alone, TNM stage and UISS had a C-index of 0.73 and 0.75, respectively. The marker model and the clinical/marker model had a C-index of 0.75 and 0.79, respectively.

Although the C-index allows various models to be ranked...
according to accuracy, it cannot be used for hypothesis testing. Therefore, we performed a test for concordance to hypothesis test and obtain P values. When the various models were compared for concordance, the clinical/marker model was significantly better than TNM stage or UISS (P < 0.001 and P = 0.038, respectively; Table 5). The differences between the marker model versus TNM stage (P = 0.152) and marker model versus UISS (P = 0.143) were not statistically significant.

The factors included in the marker model (Met, gelsolin, p53, Met*CA9, and vimentin) were used to construct a prognostic nomogram. (Fig. 1A). Fig. 1B illustrates the stratification of the survival curves based on total points assigned by the nomogram, and P values comparing adjacent survival curves are indicated. These P values should be considered as descriptive measures of curve separation. Similarly, factors included in the clinical/molecular model (Met, T stage, ECOG PS, Met*CA9, p53, and vimentin) were used to construct a second nomogram (Fig. 2A). Fig. 2B illustrates the stratification of the survival curves based on total points assigned by the second nomogram.

DISCUSSION

To our knowledge, this is the first study of any solid tumor to demonstrate that prognostic models based primarily on protein expression profiles can perform at least as well as TNM stage, histological grade, or PS. An accurate system for predicting survival is useful for patient counseling, planning follow-up, and selecting patients for additional treatment. For clinically localized RCC, patients at high risk for recurrence can be selected for adjuvant therapy trials. For metastatic RCC, low- and moderate-risk patients are good candidates for standard immunotherapy. However, high-risk patients may be selected for immediate enrollment into clinical trials. In the setting of a clinical trial, having accurate prognostic information helps ensure a consistent population of study patients, which will facilitate interpretation of outcomes.

The predictive accuracy of our marker model for RCC was comparable with UISS, which is a model combining standard clinical predictors; and the clinical/marker model was significantly more accurate than UISS. Molecular information may
reduce some of the subjectivity involved in current clinical methods for predicting survival. Although models based on gene expression profiles have been described (9, 38), the technology for gene expression profiling is costly and not yet widely available. The hope is that protein expression profiling can be more rapidly incorporated into a clinical setting using resources available at many centers. Immunohistochemical staining is already widely used for pathological analysis of breast and colon cancer tissue. However, careful multicenter clinical studies involving standardized marker staining protocols should be used to validate our findings.

Our study shows that accurate models for molecular staging of a solid tumor can be developed using a very limited number of markers. We demonstrate the use of a statistical technique for constructing prognostic models and comparing them with established systems for predicting survival. This approach can be applied to building and evaluating predictive models for other solid tumors. In this study, the various models were ranked based on predictive accuracy as measured by the concordance index. The accuracy of the predictive models was also compared by examining the concordance of the models for combinations of patient pairs. Using this approach, both the clinical/marker model and the marker model were significantly more accurate than histological grade, and the clinical/marker model was significantly more accurate than grade, TNM stage, and UISS.

Overfitting occurs when random patterns in the data are incorporated into a prognostic model as meaningful information, resulting in what appears to be a more accurate model. However, any apparent improvement in model performance due to overfitting is lost when the model is applied to an independent dataset. The complexity of the training dataset used to build the

![Fig. 2 A](image) Nomogram for predicting DSS survival using clinical data and molecular markers. See legend for Fig. 1A. B, Kaplan-Meier survival curves based on stratification using the total points assigned by the nomogram. See legend for Fig. 1C.
model correlates with the risk of overfitting. In our study, every step of the modeling process was validated and adjusted for overfitting. Therefore, when the various models are ranked according to predictive accuracy or compared to obtain P values, the significance of our results is not overstated. This helps to ensure that our models will perform similarly in an external dataset as it did in our dataset.

When evaluating a marker by immunohistochemistry, it is common to dichotomize the staining data using a single cutoff to determine positive and negative staining. Dichotomizing the staining simplifies the staining analysis and enhances interobserver reproducibility. However, when performing the statistical analysis, the selection of cutoff criterion can introduce bias and result in overfitting. Therefore, we describe a prevalidation technique for each marker that takes into account all measures of staining (such as intensity and percentage of staining), corrects for overfitting, and quantifies the true prognostic information contained in the staining. This prevalidation procedure leads to unbiased P values and C-indices. Only markers that were significant predictors of survival on prevalidation were dichotomized and used to construct nomograms. Finally, to avoid cutoffs that are overly sensitive to the staining conditions, only robust cutoffs were used.

We propose two nomograms that can be used to predict DSS. One nomogram is based on metastasis status and three molecular markers (CA9, p53, and vimentin). The second nomogram combines clinical and molecular variables (Met, T stage, ECOG PS, CA9, p53, and vimentin). By including metastasis status in the nomogram, we found that the nomogram accurately predicts DSS in both patients with localized RCC and those with metastatic RCC. Both nomograms can be used to calculate 2- and 4-year DSS rates as well as median survival. Although these nomograms are useful for visualizing our predictive models, they need to be tested on independent patient populations before being applied to patient care.

REFERENCES

(UICC) and the American Joint Committee on Cancer (AJCC). Cancer (Phila) 1997;80:992–3.


Using Protein Expressions to Predict Survival in Clear Cell Renal Carcinoma

Hyung L. Kim, David Seligson, Xueli Liu, et al.


Updated version  Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/10/16/5464

Cited articles  This article cites 30 articles, 5 of which you can access for free at: http://clincancerres.aacrjournals.org/content/10/16/5464.full.html#ref-list-1

Citing articles  This article has been cited by 17 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/10/16/5464.full.html#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.