GENETICALLY ENGINEERED NEURAL NETWORKS FOR PREDICTING PROSTATE CANCER PROGRESSION AFTER RADICAL PROSTATECTOMY

STEVEN R. POTTER, M. CRAIG MILLER, LESLIE A. MANGOLD, KERRIE A. JONES, JONATHAN I. EPSTEIN, ROBERT W. VELTRI, AND ALAN W. PARTIN

ABSTRACT

Objectives. To use pathologic, morphometric, DNA ploidy, and clinical data to develop and test a genetically engineered neural network (GENN) for the prediction of biochemical (prostate-specific antigen [PSA]) progression after radical prostatectomy in a select group of men with clinically localized prostate cancer.

Methods. Two hundred fourteen men who underwent anatomic radical retropubic prostatectomy for clinically localized prostate cancer were selected on the basis of adequate follow-up, pathologic criteria indicating an intermediate risk of progression, and availability of archival tissue. The median age was 58.9 years (range 40 to 87). Men with Gleason score 5 to 7 and clinical Stage T1b-T2c tumors were included. Follow-up was a median of 9.5 years. Three GENNs were developed using pathologic findings (Gleason score, extraprostatic extension, surgical margin status), age, quantitative nuclear grade (QNG), and DNA ploidy. These networks were developed using three randomly selected training (n = 136) and testing (n = 35) sets. Different variable subsets were compared for the ability to maximize prediction of progression. Both standard logistic regression and Cox regression analyses were used concurrently to calculate progression risk.

Results. Biochemical (PSA) progression occurred in 84 men (40%), with a median time to progression of 48 months (range 1 to 168). GENN models were trained using inputs consisting of (a) pathologic features and patient age; (b) QNG and DNA ploidy; and (c) all variables combined. These GENN models achieved an average accuracy of 74.4%, 63.1%, and 73.5%, respectively, for the prediction of progression in the training sets. In the testing sets, the three GENN models had an accuracy of 74.3%, 80.0%, and 78.1%, respectively.

Conclusions. The GENN models developed show promise in predicting progression in select groups of men after radical prostatectomy. Neural networks using QNG and DNA ploidy as input variables performed as well as networks using Gleason score and staging information. All GENN models were superior to logistic regression modeling and to Cox regression analysis in prediction of PSA progression. The development of models using improved input variables and imaging systems in larger, well-characterized patient groups with long-term follow-up is ongoing.
predict treatment outcomes. However, the variability and complexity of the data may exceed the capacity of standard modeling methods. Artificial neural networks (ANNs) attempt to simulate human decision-making using adaptation and inference parameters. ANNs can better define non-linear patterns between predictor variables and previously unknown outcomes than linear statistical models.

Validation of an ANN requires separate training and testing phases. In the training phase, the ANN “learns” the relationships of input and outcome and assigns weights to the input variables. Once these weights are formalized, the ANN is considered “trained.” The ANN must then be validated using a different data set. The term “genetic” in the phrase “genetically engineered neural network (GENN)” refers to a method of network development in which the network architecture is determined by the data presented to it. The GENN develops the relationships between input variables and outcome, selects for the “fittest” solutions, and ultimately “evolves” an optimal network. Use of ANNs in urologic oncology has shown promise.

Previously, we used logistic regression analysis to evaluate the ability of quantitative nuclear grade (QNG) and Gleason score to predict progression after RRP. We determined that QNG and Gleason score stratified patients into low, moderate, and high-risk groups for prostate cancer progression. In follow-up to that retrospective study, we now compare the ability of GENNs and logistic regression modeling to predict progression in a subset of RRP patients in whom accurate prediction is especially difficult.

**MATERIAL AND METHODS**

**PATIENTS**

A total of 214 men with prostatectomy Gleason score of 5 to 7 and clinical Stage T1b-T2c cancer were nonconsecutively selected from a cohort of more than 1800 RRP patients treated between 1982 and 1996 at one institution. The selection of these men was based on adequate follow-up (at least 5 years for patients without progression), complete clinical data, and the availability of archival tissue. All men underwent anatomic RRP. Men with seminal vesicle invasion or lymph node involvement discovered at surgery were excluded because of the known high risk of progression. Men who underwent adjuvant or neoadjuvant hormonal or radiation therapy were also excluded, as the natural history of prostate cancer in these men could not be ascertained. Most were treated before the availability of preoperative PSA testing. These 214 men formed the training and testing groups for the development and analysis of the three GENN models and had a minimum follow-up among patients without progression of 5 years (range 5 to 16). All preoperative clinical, pathologic, and postoperative data were gathered prospectively and are summarized in Table I.

Men were followed up with serum PSA measurements at 3-month intervals for 1 year, at 6-month intervals for an additional year, and yearly thereafter (after PSA testing became available in 1987). An annual interview and digital rectal examination were performed. Biochemical recurrence was defined as a postoperative serum PSA greater than 0.2 ng/mL. No patient received radiation or hormonal therapy before biochemical disease recurrence.

**IMAGE DATA ACQUISITION**

Representative sequential 5-μm-thick sections were cut from archival, formalin-fixed, paraffin-embedded tissue. Alternating sections were stained with hematoxylin-eosin and Feulgen reagents and areas of cancer marked. Approximately 150 nuclei from each tumor were analyzed. Forty-one nuclear morphometric descriptors were measured for each image, including 11 DNA content, 22 markovian texture, and 8 nuclear shape features.

**NEURAL NETWORK ANALYSIS**

All data were analyzed using NeuroGenetic Optimizer software, version 2.6 (BioComp Systems, Redmond, Wash), which builds predictive models using genetic algorithms. Input variables included prostatectomy pathologic findings (Gleason score and extraprostatic extension and surgical margin status), age, DNA ploidy, and QNG (the variance of 41 different nuclear morphometric descriptors). These variables were classified as nominal (extraprostatic extension and margin status), categorical (Gleason score and DNA ploidy), or continuous (age and nuclear morphometric descriptors).

Using pathologic findings and age (model 1), QNG and DNA ploidy (model 2), or a combination of all variables (model 3), we constructed three randomly selected training and testing sets balanced for the number of patients with (n = 84) and without (n = 87) progression in our cohort. The training sets consisted of 80% of the balanced sample; the testing sets used the remaining 20% of the balanced sample. The same three training and testing sets were used for network analysis and logistic regression modeling. To avoid network overfitting, each network was limited to a maximum of 200 training iterations.

**STATISTICAL ANALYSIS**

All data were analyzed with Stata version 5.0 statistical analysis software (Stata, College Station, Tex). Logistic regression

---

**TABLE I. Summary of demographic and clinical data in 214 men presenting with clinically localized prostate cancer**

<table>
<thead>
<tr>
<th>Average age (yr)</th>
<th>58.9 ± 6.4 (40–87)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average follow-up time (yr)</td>
<td>7.8 ± 3.9 (1–16)</td>
</tr>
<tr>
<td>Average time to progression (yr)</td>
<td>4.5 ± 3.5 (1–14)</td>
</tr>
<tr>
<td>Average follow-up</td>
<td>9.9 ± 2.7 (5–16)</td>
</tr>
<tr>
<td>(nonprogression) (yr)</td>
<td></td>
</tr>
<tr>
<td>Clinical stage (n)</td>
<td></td>
</tr>
<tr>
<td>T1b–T1c</td>
<td>6 (3)</td>
</tr>
<tr>
<td>T2a</td>
<td>72 (33)</td>
</tr>
<tr>
<td>T2b</td>
<td>113 (53)</td>
</tr>
<tr>
<td>T2c</td>
<td>23 (11)</td>
</tr>
<tr>
<td>Prostatectomy Gleason scores (n)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>50 (23)</td>
</tr>
<tr>
<td>6</td>
<td>75 (35)</td>
</tr>
<tr>
<td>7</td>
<td>89 (42)</td>
</tr>
</tbody>
</table>

Numbers in parentheses for clinical stage and Gleason scores are percentages, all others are the range.

Data are presented as the average ± standard deviation, unless otherwise noted.
analysis was used to evaluate the accuracy of the various GENNs. The outcome variable was biochemical progression. Receiver operating characteristic curves and the areas under the curves were calculated for each of the GENN models, as were sensitivity, specificity, and accuracy. Accuracy was defined as the overall percentage of cases that were correctly classified. Kaplan-Meier analysis was performed using the average results of model 3. The actuarial curve significance was determined using the log-rank test of equality and the Wilcoxon-Gehan test.

Logistic regression analysis was performed concurrently on the same three randomly selected training and testing sets using the same combinations of input variables. A multivariate significance stringency of \( P < 0.25 \) was used for backward stepwise logistic regression analysis. Again, receiver operating characteristic curves and the areas under the curves were calculated for each model, as were sensitivity, specificity, and accuracy. The Cox proportional hazards model was performed on the training and testing set output of model 3.

**RESULTS**

Of the 149 tumors (70%) with extraprostatic spread at pathologic staging, 66 (31%) also had positive margins. The remaining 65 tumors (30%) were organ confined. During a median follow-up of 9.5 years, 84 men (40%) developed biochemical progression within a median of 4 years (range 1 to 14). In the biochemical progression-free men (n = 130), 75% of the tumors had prostatectomy Gleason scores of 5 or 6; of the men with biochemical progression (n = 84), 67% had a prostatectomy Gleason score of 7.

The three GENN models achieved an average accuracy of 74.4%, 63.1%, and 73.5% for predicting progression in the training sets. The testing sets produced an average accuracy of 74.3%, 80.0%, and 78.1% (Table II). The use of QNG and DNA ploidy alone as input variables (model 2) had a lower sensitivity and higher specificity than the use of pathologic results and patient age (model 1). The training and testing sets were analyzed concurrently by logistic regression and Cox proportional hazards modeling (Table III). Logistic regression analysis maximized performance in the training sets, and the GENN models maximized performance in the testing sets. For the testing set, Cox analysis yielded a sensitivity of only 39%, a specificity of 67%, and an accuracy of 53% (Table III).

Kaplan-Meier analysis, performed on the average outputs of model 3 for the entire patient sample, allowed stratification of tumors into four biochemical recurrence risk groups (Fig. 1). The log-rank test of equality was used to calculate the significance levels for the differences between the risk groups (\( P \) value between groups I and II = 0.092; between groups II and III <0.0001; and between groups III and IV = 0.0113).

**COMMENT**

Although PSA testing has revolutionized the early detection of prostate cancer, PSA levels alone have a limited ability to predict progression. Prediction is especially problematic in men with clinically organ-confined cancer who, at surgery, have tumors with a Gleason score of 5 to 7 and negative seminal vesicles and lymph nodes.\(^{14}\)

We developed and tested ANNs and compared them with the results of logistic regression analysis in a selected cohort of men at intermediate risk of cancer progression and with a lengthy follow-up. Our findings suggest that GENNs are useful in progression prediction and may aid in clinical deci-
sion-making and the rational design of clinical trials. All GENN testing set models were superior to logistic regression modeling in predicting progression. Progression prediction using a Cox regression model was also inferior to ANN performance. Development of three different GENN models allowed comparison of different input variables.

The use of ANNs in predicting outcome after surgery shows promise, but some limitations are apparent. Currently, a pathologist and imaging technician are required to select cancer nuclei for QNG determination. The utility of QNG (models 2 and 3) was reduced by the limitations of the nuclear imaging system used. Analysis with current state-of-the-art systems is ongoing and will likely improve the contribution of QNG in these models.

Because of limitations on patient numbers necessitated by our desire for a lengthy follow-up and intermediate progression risk, we did not construct a separate set of previously unstudied patients to serve as a validation cohort. This does not invalidate the comparison of GENN and logistic regression analysis results. Because the testing set patients were not used to adjust the input weights...
in our networks, the testing set results are useful in assessing these networks as tools for predicting progression. The collection of a validation patient cohort is underway.

The absence of PSA values as input variables, necessary because the length of follow-up achieved meant that most men had undergone RRP before the PSA era, was potentially limiting. However, new input variables, such as PSA or other serologic, immunohistochemical, or molecular markers, can be incorporated into GENNs with relative ease and are likely to increase the predictive value. Few of these men had Stage T1c lesions, and the development of predictive models using a more representative percentage of nonpalpable cancers is ongoing.

CONCLUSIONS

The application of ANNs in progression prediction shows promise in men at intermediate risk of progression in whom prediction has historically been most inaccurate. GENN creation is a logical step in the development of progression modeling. Networks were developed with high sensitivity and specificity for the prediction of prostate cancer progression in a group of men with long-term prospective follow-up after RRP. Advances in nuclear imaging systems and input variable selection promise further improvements. The development of these improved models in larger, well-characterized patient groups with long-term follow-up is ongoing. Further development of GENNs will provide improved prognostication after radical prostatectomy, allowing early and appropriate evaluation of investigational adjuvant therapies.

REFERENCES


