Award Number: W81XWH-10-1-0090

TITLE: Validation of Biomarkers Predictive of Recurrence Following Prostatectomy

PRINCIPAL INVESTIGATOR: Dr. Carlos Moreno

CONTRACTING ORGANIZATION: Emory University Atlanta, GA  30322

REPORT DATE: April 2011

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

   X  Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.
Validation of Biomarkers Predictive of Recurrence Following Prostatectomy

Dr. Carlos S. Moreno
cmoreno@emory.edu

Emory University
Atlanta, GA 30322-1018

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

Approved for public release; distribution unlimited

Thus far we have made good progress on our goal to validate biomarkers of recurrence in prostate cancer. We have published a manuscript now in press at *The American Journal of Pathology* describing our set of biomarkers, and will be presenting these data orally at the American Society of Investigative Pathology annual meeting in Washington, DC on April 11, 2011. We have initiated collection of samples at the Atlanta VA Medical Center and begun tissue coring and RNA extraction. We have faced some administrative hurdles with getting the project started regarding IRB approvals and subcontract awards. However, now that all of these obstacles have been overcome, we expect rapid progress in the next year for data generation. In addition, the miRNA DASL platform has been discontinued by Illumina, and we are thus planning on transitioning the project to next generation sequencing methods, which are likely to be more accurate and comprehensive. Because of the delays in getting the project initiated, we are requesting a no cost one year extension to change the project from a two-year to a three-year project.

Prostate Cancer, Biomarkers, Genes, MicroRNAs, Microarrays

No security classification.

U

U

UU

UU

46
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>Body</td>
<td>5</td>
</tr>
<tr>
<td>Key Research Accomplishments</td>
<td>13</td>
</tr>
<tr>
<td>Reportable Outcomes</td>
<td>14</td>
</tr>
<tr>
<td>Conclusion</td>
<td>15</td>
</tr>
<tr>
<td>References</td>
<td>16</td>
</tr>
<tr>
<td>Appendices</td>
<td>18</td>
</tr>
</tbody>
</table>
Introduction

An important challenge in prostate cancer research is to develop effective predictors of tumor recurrence following surgery in order to determine whether immediate adjuvant therapy is warranted. To identify biomarkers predictive of biochemical recurrence, we isolated the RNA from 70 formalin-fixed paraffin-embedded (FFPE) radical prostatectomy specimens with known long term outcome to perform DASL expression profiling with a custom-designed panel of 522 prostate cancer relevant genes that we designed. We identified a panel of ten protein-coding genes and two miRNA genes (RAD23B, FBP1, TNFRSF1A, CCNG2, NOTCH3, ETV1, BID, SIM2, ANXA1, miR-519d, and miR-647) that could be used to separate patients with and without biochemical recurrence (p < 0.001), as well as for the subset of 42 Gleason score 7 patients (p < 0.001). We performed an independent validation analysis on 40 samples and found that the biomarker panel was also significant at prediction of biochemical recurrence for all cases (p = 0.013) and for a subset of 19 Gleason score 7 cases (p = 0.010), both of which were adjusted for relevant clinical information including T-stage, PSA and Gleason score. Importantly, these biomarkers could significantly predict clinical recurrence for Gleason score 7 patients. A manuscript describing these biomarkers is now in press at the American Journal of Pathology.¹

In March 2010, we received funding for a two-year Prostate Cancer IDEA Award to validate and improve this set of biomarkers on an independent, large set of patient samples. Over the past year, we have dealt with a large number of administrative hurdles to obtain the necessary approvals to proceed with the research funded by this award. Now that all of these hurdles have been overcome, we are beginning to implement our research plan in earnest. Because of the delays in the beginning of the project, we are requesting a no-cost extension of the project from a 2-year to a 3-year project.
Body

Custom Prostate DASL profiling

Funding for this IDEA Award was based on our DASL expression profiling data using our custom-designed prostate cancer panel and the Illumina DASL microRNA (miRNA) panel on 70 prostatectomy patient samples to identify biomarkers predictive of recurrence. In addition, an independent validation profiling experiment was performed on 40 additional samples. MiRNA probes were filtered to retain only those that were present on the miRNA microarrays used for both the training and validation sets, reducing the total number of probes examined to 403 miRNA probes. The training set included 29 cases with observed biochemical PSA recurrence (median time to recurrence = 19 months), and 41 cases censored, i.e., without observed recurrence during follow-up (median follow-up time = 83.0 months). A summary of the clinical characteristics of the training and validation sets of samples is provided in Table 1.

Integrated DASL biomarker analysis

After fitting a univariate Cox proportional hazard (PH) model for each individual probe using the training data, a set of 27 important probes were preselected based on an FDR threshold of 0.30. Next, to identify the optimal prediction score based on the preselected probes, we fit a lasso Cox proportional hazard (PH) model\(^2, 3\) first using the set of 25 preselected mRNA probes only, resulting in a panel of nine protein-coding genes shown in Table 2 (RAD23B, FBP1, TNFRSF1A, NOTCH3, ETV1, BID, SIM2, ANXA1, and BCL2). A final prediction model was then built to include the predictive score based on this panel of nine mRNA biomarkers as well as the relevant clinical biomarkers including T-stage, PSA and Gleason score, which could be used to predict recurrence following radical prostatectomy. Kaplan-Meier analysis (Figure 1A) demonstrated that these probes could significantly discriminate patients at higher and lower risk of recurrence by the log rank test (p < 0.001). We next applied the final predictive model developed on the training set to the validation set, a separate, independent DASL profiling experiment performed on a different day. Kaplan-Meier analysis (Figure 1B) on this validation set determined that the model could discriminate patients at higher and lower risk of recurrence (p = 0.010).

Subsequently, we repeated the above training procedure using the complete set of 27 preselected mRNA and miRNA probes, and we identified an optimal panel of ten mRNAs and two microRNAs (Table 3) and built a final prediction model for prostate cancer biochemical recurrence, which again included relevant clinical biomarkers. Kaplan-Meier analysis and the log-rank test determined that this panel could also significantly discriminate patients at higher and lower risk of recurrence both in the training set (p < 0.001, Figure 1C) and in the validation set (p = 0.013, Figure 1D).

Prediction of Cases with a Gleason Score 7

Prediction of recurrence for patients with a Gleason score 7 is particularly difficult. In order to address this issue, we applied the biomarker panels to the subset of cases in the training and validation sets that had a Gleason score 7. The prediction model based on the nine-mRNA panel was significant at discriminating biochemical recurrence in Gleason score 7 cases in both the training set (p < 0.001, Figure 2A) and the validation set (p = 0.027, Figure 2B). For the prediction model based on the combined panel of ten mRNAs and two miRNAs in Table 3, the predictive value was again significant for both the training set (p < 0.001, Figure 2C) and the validation set (p = 0.010, Figure 2D). A summary of the p-values for predicting biochemical recurrence is given in Table 4. In all cases, the prediction models that use one of the two gene biomarker panels plus clinical information outperforms the prediction model using only clinical information.

Analysis of clinical recurrence

Although most patients who have clinical recurrence following prostatectomy also have
biochemical recurrence, there is a significant population of patients with biochemical recurrence who do not have clinically significant recurrences observed during their follow-ups. To evaluate our biomarker panel of biochemical recurrence for predicting the clinical recurrence, we tested the prediction model based on the combined mRNA/miRNA panel in the same training and validation samples using their clinical recurrence outcome data. Unfortunately, clinical recurrence data was lacking on some of the samples, and the total number of samples used in the training set was reduced. In the training data, the combined mRNA/miRNA panel was highly significant for predicting clinical recurrence in all patients (p=0.002) as well as in the subset of patients with a Gleason score 7 (p=0.004); in the validation data, it was also significant for predicting recurrence in patients with a Gleason score 7 (p=0.023) and trended towards significance in all patients (p=0.078). A summary of the p-values for predicting clinical recurrence is given in Table 5. In all cases, the prediction model that uses the combined mRNA and miRNA panel plus the clinical information, again, outperforms the prediction model that uses only the clinical information.

We also performed an analysis to construct a predictive set of biomarkers based on the clinical recurrence data instead of biochemical recurrence. Only three probes passed the initial preselection step for the univariate Cox PH modeling, all corresponding to the ETV1 gene, which is likely due to the considerably fewer number of clinical recurrences in the training set as well as the smaller total sample size. Furthermore, the prediction model built on this set of gene biomarkers did not perform as well as the models built on biochemical recurrence (data not shown).

**American Journal of Pathology Manuscript**

A manuscript describing these results was submitted to the *American Journal of Pathology* in September, 2010, a revision was submitted in December, 2010, and the paper was accepted for publication in March, 2011. This manuscript is attached as Appendix 1.
### Table 1: A summary of the clinical characteristics of the training and validation sets of patient samples. (BCR = Biochemical Recurrence, F/U = follow up, PSA = prostate specific antigen, SD = standard deviation).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Coefficient</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD23B</td>
<td>RAD23 homolog B</td>
<td>0.152</td>
<td>4, 5</td>
</tr>
<tr>
<td>FBP1</td>
<td>Fructose-1,6-bisphosphatase 1</td>
<td>0.310</td>
<td>6-8</td>
</tr>
<tr>
<td></td>
<td>Tumor Necrosis Factor Receptor Superfamily,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>Member 1A</td>
<td>-0.560</td>
<td>9, 10</td>
</tr>
<tr>
<td>NOTCH3</td>
<td>Notch homolog 3</td>
<td>0.426</td>
<td>11, 12</td>
</tr>
<tr>
<td>ETV1</td>
<td>Ets Variant Gene 1 (ETV1)</td>
<td>0.157</td>
<td>13, 14</td>
</tr>
<tr>
<td>BID</td>
<td>BH3 Interacting Domain Death Agonist (BID)</td>
<td>0.248</td>
<td>15, 16</td>
</tr>
<tr>
<td>SIM2</td>
<td>Single-Minded Homolog 2</td>
<td>0.043</td>
<td>17-20</td>
</tr>
<tr>
<td>ANXA1</td>
<td>Annexin A1</td>
<td>-0.185</td>
<td>21-24</td>
</tr>
<tr>
<td>BCL2</td>
<td>B-cell CLL/lymphoma 2</td>
<td>0.028</td>
<td>25, 26</td>
</tr>
</tbody>
</table>

### Table 2: Nine-gene predictor of prostate cancer recurrence following surgery. Coefficient is derived from the lasso Cox proportion hazards model and was used for computing the predictive score. Positive coefficients indicate a positive association with recurrence, and negative coefficients a negative association with recurrence.
Table 3: Twelve-gene predictor of prostate cancer recurrence following surgery using ten mRNAs and two microRNAs. Coefficient is derived from the lasso Cox proportion hazards model and was used for computing the predictive score. Positive coefficients indicate a positive association with recurrence, and negative coefficients a negative association with recurrence.
### Table 4: Summary of p-values (Logrank test) of prediction of biochemical recurrence on training and validation sets for the entire dataset and the subset of Gleason score 7 cases using two biomarker panels, all of which are adjusted for T-stage, PSA, and Gleason score, or using clinical information only. Significant p-values are indicated in bold.

<table>
<thead>
<tr>
<th>Training Set</th>
<th>mRNA panel</th>
<th>Combined mRNA/miRNA panel</th>
<th>Clinical Information Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cases (n=61)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.096</td>
</tr>
<tr>
<td>Gleason score 7 (n=42)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.641</td>
</tr>
<tr>
<td>Validation Set</td>
<td>mRNA panel</td>
<td>Combined mRNA/miRNA panel</td>
<td>Clinical Information Only</td>
</tr>
<tr>
<td>All Cases (n=35)</td>
<td>0.010</td>
<td><strong>0.013</strong></td>
<td><strong>0.020</strong></td>
</tr>
<tr>
<td>Gleason score 7 (n=19)</td>
<td>0.027</td>
<td><strong>0.010</strong></td>
<td><strong>0.028</strong></td>
</tr>
</tbody>
</table>

### Table 5: Summary of p-values (Logrank test) of prediction of clinical recurrence on training and validation sets for the entire dataset and the subset of Gleason score 7 cases using the combined and mRNA/miRNA panel, all of which are adjusted for T-stage, PSA, and Gleason score, or using clinical information only. Significant p-values are indicated in bold.

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Combined mRNA/miRNA panel</th>
<th>Clinical Information Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cases (n=56)</td>
<td><strong>0.002</strong></td>
<td>0.262</td>
</tr>
<tr>
<td>Gleason score 7 (n=37)</td>
<td><strong>0.004</strong></td>
<td>0.136</td>
</tr>
<tr>
<td>Validation Set</td>
<td>Combined mRNA/miRNA panel</td>
<td>Clinical Information Only</td>
</tr>
<tr>
<td>All Cases (n=35)</td>
<td>0.078</td>
<td>0.193</td>
</tr>
<tr>
<td>Gleason score 7 (n=19)</td>
<td><strong>0.023</strong></td>
<td>0.080</td>
</tr>
</tbody>
</table>
Figure 1: Prediction of biochemical recurrence in all prostate cancer patients using two biomarker panels, adjusted for clinical information. (A) Kaplan-Meier analysis of the training set patients that were separated based on the mRNA panel described in Table 2. (B) Kaplan-Meier analysis on the validation cases using the mRNA panel. (C) Kaplan-Meier analysis of the training set using the combined mRNA and miRNA panel described in Table 3. (D) Kaplan-Meier analysis of the validation set using the combined mRNA and miRNA panel.
Figure 2: Prediction of biochemical recurrence in prostate cancer patients with a Gleason score 7 using two biomarker panels, adjusted for clinical information. (A) Kaplan-Meier analysis of the training set of Gleason score 7 cases using the mRNA panel described in Table 2. (B) Kaplan-Meier analysis of the Gleason score 7 cases in the validation set using the mRNA panel. (C) Kaplan-Meier analysis of the Gleason score 7 cases in the training set using the combined mRNA and miRNA panel described in Table 3. (D) Kaplan-Meier analysis of the Gleason score 7 cases in the validation set using the combined mRNA and miRNA panel.
Administrative hurdles

When this project was initiated, we anticipated that the IRB protocols already in place would be sufficient for conducting the described research project. However, review by the DOD determined that the IRB protocols were too general and not sufficiently specific. Consequently, we submitted new, specific IRB protocols at Emory University, the Emory/Atlanta VA Medical Center, and at Sunnybrook Research Centre at the University of Toronto. While we obtained IRB approval letters fairly quickly at Emory and the VA, approval from Sunnybrook took much longer. The difficulty hinged on award of the subcontract, since the Sunnybrook IRB would not issue approval without funding, and Emory would not issue the subcontract without the IRB approval. We were stuck in a Catch-22 for several months before we finally obtained Sunnybrook IRB approval in November, 2010 and permission to begin work at Sunnybrook was obtained in December, 2010. Unfortunately, it took another two months before all of the legal issues could be resolved, including the fact that there is no HIPAA law in Canada, so that the subcontract could be awarded to Sunnybrook by Emory University. Now that all of the administrative hurdles have been overcome, we expect to begin receiving samples from Toronto in the very near future.

Progress at Emory University

In the meantime, once approval was obtained from the DOD on August 13, 2010 to commence work on the project at Emory, we began immediately to work on identifying samples that could use in our validation study. We identified 150 cases at the VA hospital between 1990-2000 that could potentially be used for this project. We were able to locate slides and formalin-fixed paraffin embedded (FFPE) blocks for 100 of those cases, and Dr. Oyesiku identified regions of cancer and benign tissue in slides for each of them. These samples were then submitted for processing to obtain 1 mm tissue cores. We anticipate that we will have RNA ready for WG-DASL analysis in the next few weeks. We are planning to examine additional cases at Emory and the VA between 2000-2003 to identify samples for use in Aim 2.

Platform issues

In our initial proposal, we planned to use the Illumina miRNA DASL platform for analysis of miRNA biomarker expression levels. Since submission of our application for initial review, Illumina has discontinued this platform. Consequently, we are evaluated several different options for comprehensive analysis of miRNA levels. These options include TaqMan Low Density Arrays, Affymetrix miRNA arrays, High-Throughput Genomics Quantitative Nuclease Protection Assays, Illumina HiSeq sequencing, Ion Torrent sequencing, and Nanostring sequencing. There are strengths and weaknesses to each of these options that we are currently scrutinizing. However, we are likely to use a 48-plex multiplexing method of next generation sequencing on the Illumina HiSeq platform. This should give us deep coverage and high data quality at an acceptable cost. We plan on performing pilot studies to determine whether we can obtain acceptable data using FFPE-derived RNA shortly.
Key Research Accomplishments

Our key research accomplishments are summarized below:

- Identified a set of 10 mRNAs and 2 miRNAs predictive of recurrence following prostatectomy.

- Published a manuscript describing these biomarkers of recurrence.

- Obtained all necessary IRB and DOD approvals to commence work.

- Identified 100 prostate cancer cases at the Atlanta/VA for the validation study.

- Marked tumor and benign areas to enable coring of FFPE blocks.

- Initiated RNA extraction of the first 100 cases for WG-DASL analysis.
## Reportable Outcomes

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grant Award Received</td>
<td>March 25, 2010</td>
</tr>
<tr>
<td>New IRB Protocol Requested from DOD</td>
<td>May 20, 2010</td>
</tr>
<tr>
<td>New IRB Protocol Approved by Emory University</td>
<td>June 20, 2010</td>
</tr>
<tr>
<td>Approval received from VA Research Committee to use VA samples</td>
<td>August 10, 2010</td>
</tr>
<tr>
<td>Permission to begin research at Emory obtained from DOD</td>
<td>August 13, 2010</td>
</tr>
<tr>
<td>Manuscript describing initial biomarkers submitted for publication</td>
<td>September 24, 2010</td>
</tr>
<tr>
<td>Initial set of 100 samples identified at the Emory/VA</td>
<td>October 12, 2010</td>
</tr>
<tr>
<td>Abstract submitted to American Society for Investigative Pathology (ASIP) 2011 meeting in Washington, DC</td>
<td>November 5, 2010</td>
</tr>
<tr>
<td>New IRB Protocol Approved by U. Toronto/Sunnybrook</td>
<td>November 15, 2010</td>
</tr>
<tr>
<td>Permission to begin research at U. Toronto/Sunnybrook obtained from DOD</td>
<td>December 21, 2010</td>
</tr>
<tr>
<td>Revised manuscript describing biomarkers submitted for publication</td>
<td>December 31, 2010</td>
</tr>
<tr>
<td>Abstract for ASIP 2011 meeting selected for oral presentation</td>
<td>February 7, 2011</td>
</tr>
<tr>
<td>Tissue blocks pulled and slides marked for 100 samples at the VA</td>
<td>February 18, 2010</td>
</tr>
<tr>
<td>Subaward Contract agreed between Emory and U. Toronto/Sunnybrook</td>
<td>February 28, 2011</td>
</tr>
<tr>
<td>Manuscript describing initial biomarkers accepted for publication in AJP</td>
<td>March 3, 2011</td>
</tr>
<tr>
<td>Tissue coring from 100 VA FFPE tissue blocks initiated</td>
<td>March 15, 2011</td>
</tr>
</tbody>
</table>
Conclusion

Thus far we have made good progress on our goal to validate biomarkers of recurrence in prostate cancer. We have published a manuscript now in press at *The American Journal of Pathology* describing our set of biomarkers\(^1\), and will be presenting these data orally at the American Society of Investigative Pathology annual meeting in Washington, DC on April 11, 2011. We have initiated collection of samples at the Atlanta VA Medical Center and begun tissue coring and RNA extraction. We have faced some administrative hurdles with getting the project started regarding IRB approvals and subcontract awards. However, now that all of these obstacles have been overcome, we expect rapid progress in the next year for data generation. In addition, the miRNA DASL platform has been discontinued by Illumina, and we are thus planning on transitioning the project to next generation sequencing methods, which are likely to be more accurate and comprehensive. Because of the delays in getting the project initiated, we are requesting a no cost one year extension to change the project from a two-year to a three-year project.
References

9. Thorburn A: Death receptor-induced cell killing, Cell Signal 2004, 16:139-144


Appendices

See attached manuscript by Long et al\textsuperscript{1}.
Protein-coding and MicroRNA Biomarkers of Recurrence of Prostate Cancer Following Radical Prostatectomy

Qi Long1,4, Brent A. Johnson1, Adeboye O. Osunkoya3,4,5,6, Yu-Heng Lai2,3, Wei Zhou2,3,4,7,9, Mark Abramovitz8,11, Mingjing Xia4, Mark B. Bouzyk4,9, Robert K. Nam10, Linda Sugar10, Aleksandra Stanimirovic10, Brian R. Leyland-Jones4,7, Arun K. Seth10, John A. Petros3,4,5,6,7, and Carlos S. Moreno2,3,4

1Department of Biostatistics and Bioinformatics, Emory University
2Program in Genetics and Molecular Biology, Emory University,
3Department of Pathology and Laboratory Medicine, Emory University,
4Winship Cancer Institute,
5Urology Department, Emory University,
6Atlanta VA Medical Center, Decatur, Georgia.
7Department of Hematology and Medical Oncology, Emory University,
8VM Institute of Research, Montreal, Quebec,
9Department of Human Genetics, Emory University,
10Department of Laboratory Medicine and Pathobiology, University of Toronto, and Department of Anatomic Pathology, Sunnybrook Health Sciences Centre
11Current address: Departments of Pathology & Oncology, Jewish General Hospital, 3755 Cote Ste Catherine, Montreal, QC, H3T 1E2; Canada

Corresponding Author:
Carlos S. Moreno
Associate Professor
Department of Pathology & Laboratory Medicine
Winship Cancer Institute
Emory University
Whitehead Research Building, Rm. 105J
615 Michael St.
Atlanta GA, 30322

(404) 712-2809 (Ph)
(404) 727-8538 (Fax)
cmoreno@emory.edu

Key Words: Prostate Cancer, Recurrence, Biomarkers

Running Title: Biomarkers of Recurrence in Prostate Cancer
ABSTRACT

An important challenge in prostate cancer research is to develop effective predictors of tumor recurrence following surgery in order to determine whether immediate adjuvant therapy is warranted. To identify biomarkers predictive of biochemical recurrence, we isolated the RNA from 70 formalin-fixed paraffin-embedded (FFPE) radical prostatectomy specimens with known long term outcome to perform DASL expression profiling with a custom-designed panel of 522 prostate cancer relevant genes that we designed. We identified a panel of ten protein-coding genes and two miRNA genes (RAD23B, FBP1, TNFRSF1A, CCNG2, NOTCH3, ETV1, BID, SIM2, ANXA1, miR-519d, and miR-647) that could be used to separate patients with and without biochemical recurrence (p < 0.001), as well as for the subset of 42 Gleason score 7 patients (p < 0.001). We performed an independent validation analysis on 40 samples and found that the biomarker panel was also significant at prediction of recurrence for all cases (p = 0.013) and for a subset of 19 Gleason score 7 cases (p = 0.010), both of which were adjusted for relevant clinical information including T-stage, PSA and Gleason score. Importantly, these biomarkers could significantly predict clinical recurrence for Gleason score 7 patients. These biomarkers may increase the accuracy of prognostication following radical prostatectomy using formalin-fixed specimens.
INTRODUCTION

Prostate cancer remains the most common non-cutaneous cancer diagnosed for U.S. males, and ranks second among tumor site-specific mortality, with estimates for 2009 at over 192,000 new cases and 27,000 deaths\textsuperscript{1}. The majority of patients with prostate cancer are clinically asymptomatic with early-stage, organ-confined disease, and in fact, more than 50\% of men who reach the age of 80 develop clinically insignificant prostate cancer. However, a subpopulation of prostate cancer patients progress to highly aggressive, androgen-independent metastatic disease, which is inevitably fatal. One of the important challenges in current prostate cancer research is to develop effective methods to determine whether a patient is likely to progress to aggressive, metastatic disease in order to aid clinicians in deciding on the appropriate course of treatment. Biomarker assays that could predict progression and metastasis for prostate cancer patients would be of great utility in aiding clinical management of this large patient population. An important challenge in prostate cancer research is to develop effective predictors of tumor recurrence following surgery in order to determine whether immediate adjuvant therapy is warranted. Thus, biomarkers that could predict the likelihood of success for surgical therapies would be of great clinical significance.

In the past few years, enormous progress has been made in developing technologies to exploit formalin-fixed paraffin-embedded (FFPE) tumor tissue samples for gene expression analysis. The DASL (cDNA-mediated Annealing, Selection, extension and Ligation) assay is a unique expression profiling platform that is based upon massively multiplexed RT-PCR applied in a microarray format, that allows for the determination of expression of RNA isolated from 96 FFPE tumor tissue samples in a high throughput format\textsuperscript{2, 3}.

Here, we have identified biomarkers predictive of recurrence by expression profiling archived FFPE tumor samples using both a custom panel of prostate cancer associated mRNA genes and a panel
of microRNA genes. These biomarkers were developed on a training set of 70 patients (29 with biochemical recurrence and 41 controls) and validated on an independent set of 40 samples (13 with biochemical recurrence and 27 controls) and were able to significantly discriminate between patients with and without biochemical recurrence following radical prostatectomy. Moreover, these biomarkers were able to discriminate biochemical recurrence in patients with Gleason score 7, for whom outcome is particularly difficult to predict.
MATERIALS AND METHODS

Patient Samples

In the initial training set, 70 cases were used (29 with biochemical recurrence and 41 controls), 45 patients from Sunnybrook Health Science Center (Toronto, ON), and 25 patients from Emory University. The 45 cases of paraffin-embedded tissue samples from Toronto were drawn from men who underwent radical prostatectomy as the sole treatment for clinically localized prostate cancer (PCa) between 1998 and 2006. The clinical data includes multiple clinicopathologic variables such as prostate specific antigen (PSA) levels, histologic grade (Gleason score), tumor stage (pathologic stage category for example; organ confined, pT2; or with extra-prostatic extension, pT3a; or with seminal vesicle invasion, pT3b), and biochemical recurrence rates. For the cases from Emory University, both the training set (25 cases) and validation set (40 cases) FFPE samples were also selected from a screen of over a thousand patients through an IRB-approved retrospective study at Emory University of men who had undergone radical prostatectomy between 1990 and 1994. Those who were included met specific inclusion criteria, had available tissue specimens, documented long term follow-up and consented to participate or were included by IRB waiver. The cases were assigned prostate ID numbers to protect their identities. These patients did not receive neo-adjuvant or concomitant hormonal therapy. Their demographic, treatment and long-term clinical outcome data have been collected and recorded in an electronic database. Clinical data recorded include PSA measurements, radiological studies and findings, clinical findings, tissue biopsies and additional therapies that the subjects may have received. Clinical data associated with the samples used in this study are given in Supplementary Table S1 (see Supplemental Table S1 at http://ajp.amjpathol.org).

RNA Preparation
Tissue cores (1 mm) were used for RNA preparation rather than sections because of the heterogeneity of samples and the opportunity for obtaining cores with very high percentage tumor content. H&E stained slides were reviewed by a board certified urologic pathologist (AOO) to identify regions of cancer to select corresponding areas for cutting of cores from paraffin blocks. Total RNA was prepared at the Emory Biomarker Service Center from FFPE cores as previously described, using the Ambion Recoverall MagMax methodology in 96-well format on a MagMax 96 Liquid Handler Robot (Life Technologies, Carlsbad, CA). FFPE RNA was quantitated using an Nanodrop spectrophotometer (Wilmington, DE), and tested for RNA integrity and quality by Taqman analysis of the RPL13a ribosomal protein on a HT7900 real-time PCR instrument (Applied Biosystems, Foster City, CA). Samples with sufficient yield (>500 ng), A260/A280 ratio > 1.8 and RPL13a CT values less than 30 cycles were used for miRNA and DASL profiling.

**Custom Prostate Cancer DASL Assay Pool (DAP)**

The DASL assay enables quantitation of expression of up to 1,536 probes using RNA isolated from archived FFPE tumor tissue samples in a high throughput format. Data from multiple publicly available gene expression datasets, along with genes involved in prostate cancer progression based on current understanding of the disease, were distilled to develop a highly predictive set of 522 genes for use in the DASL assay. Due to specific probe design considerations, this panel had three probes for 497 genes, two probes for 20 genes, and a single probe for five genes, two of which were specific to TMPRSS2-ERG and TMPRSS2-ETV1 fusions transcripts. The unique combination of genes was optimized for performance in the DASL assay using stringent criteria that predicts excellent performance of the primer sets. The panel includes genes found to be correlated with Gleason score in Liu et al, Bibikova et al, True et al, LaPointe et al, and/or Singh et al.
It also includes prognostic markers from Dhanasekaran et al\textsuperscript{5} and Yu et al\textsuperscript{14}, and genes associated with metastasis in Varambally et al\textsuperscript{6}. In addition, a number of genes known from other studies to be critical in prostate cancer such as \textit{NKK3.1}, \textit{PTEN}, and the androgen receptor are all included in the panel. Other genes that play important roles in the Wnt, Hedgehog, TGF\textbeta, Notch, MAPK and PI3K pathways are also present in this gene set. Finally, primer sets that detect chromosomal translocations in \textit{ERG} \textsuperscript{9}, \textit{ETV1} \textsuperscript{15}, and \textit{ETV4} \textsuperscript{16} are also included in this panel. The custom prostate cancer panel list of 522 candidate genes (see Supplemental Table S2 at http://ajp.amjpathol.org) was submitted to Illumina for synthesis. The optimal oligonucleotide sequence for each of the 1,536 gene probes was determined using an oligonucleotide scoring algorithm. The oligonucleotide pool or DASL Assay Pool (DAP) was synthesized by Illumina for use with the 96-well Universal Array Matrix (UAM).

\textbf{The DASL (cDNA-mediated Annealing, Selection, extension and Ligation) assay}

The DASL assay was performed with our 522-gene custom designed human prostate cancer panel using 200 ng of input RNA at the Emory Biomarker Service Center, Emory University according to the manufacturer’s protocols. Samples, including technical replicates (2, 3 or 4) were hybridized on UAMs, and scanned using the BeadStation 500 Instrument (Illumina Inc.). For miRNA DASL assays, the human miRNA v2 DASL panel (Illumina, Inc.), which allows for the determination of expression of 1,146 human miRNAs (> 97\% coverage of miRBase release 12) was used. These data are available at GEO under accession number GSE26367.

\textbf{Data Analysis}

DASL fluorescent intensities were interpreted in GenomeStudio, quantile normalized, and exported for meta-analysis. Average signal intensity, genes detected (p-value = 0.01), background, and
noise (standard deviation of background) were analyzed for trends by plate, row, and column. The two endpoints of interest were postoperative biochemical recurrence, defined as two detectable PSA readings (>0.2 ng/ml), and clinical recurrence, defined as evidence of local or metastatic disease. The primary outcome of interest was time to biochemical recurrence following surgery. A local recurrence was defined as recurrence of cancer in the prostatic bed that was detected by either a palpable nodule on digital rectal examination (DRE) and subsequently verified by a positive biopsy, and/or a positive imaging study (Prostascint or CT scan) accompanied by a detectable postoperative PSA result and lack of evidence for metastases. Also, patients whose PSA level decreased following adjuvant pelvic radiation therapy for elevated postoperative PSA were considered as local recurrence cases. A recurrence with metastases was defined as a positive imaging study indicating presence of a tumor outside of the prostatic bed.

To identify important biomarkers and build and evaluate prediction models for prostate cancer recurrence, we adopted the following strategy. In the training step, the prediction model was built based on the time to biochemical recurrence. Specifically, we first fit a univariate Cox proportional hazard (PH) model for each individual oligonucleotide probe using the training data set, and a set of important mRNA and miRNA probes were then preselected based on a false discovery rate (FDR) threshold of 0.30. Next, to identify the optimal prediction score based on the preselected probes, we fit a lasso Cox PH model using the training data set, where the tuning parameter for lasso was selected using a leave-one-out cross-validation technique. The lasso Cox PH model was fitted first using the set of preselected mRNA probes only and then using the complete set of preselected mRNA and miRNA probes, resulting in an optimal mRNA panel and an optimal combined mRNA/miRNA panel, respectively. Based on each biomarker panel, a final prediction model for recurrence was built to also incorporate relevant clinical biomarkers, namely, T-stage, PSA and Gleason score, through
fitting Cox PH models. For comparison, we also built a prediction model using only clinical information, namely, T-stage, PSA and Gleason score, through fitting a Cox PH model.

To evaluate and validate the final prediction models obtained from the training phase, 79 samples from 40 patients were used and replicate samples from the same patient were again averaged to generate a single average signal for each patient. Each prediction model from the training phase was used to generate a predictive score for each subject in the validation data set, and subjects were subsequently divided into high and low scoring groups using the median predictive score. Kaplan Meier analysis was performed to compare the time to biochemical recurrence, between high (poor score) and low (good score) risk groups, and the statistical significance was determined using the log-rank test. Similarly, we also evaluated the final model that uses the combined mRNA/miRNA panel for predicting time to clinical recurrence in both training and validation data sets.

Missing data are present in this study, in particular, for clinical recurrence, PSA and T-stage data. We adopted the available-case approach in our analyses and the sample sizes used in each step of building and evaluating prediction models may be less than the total sample size.
RESULTS

Custom Prostate DASL profiling

We performed DASL expression profiling with our custom-designed prostate cancer panel (see Materials and Methods section) and the Illumina DASL microRNA (miRNA) panel on 70 prostatectomy patient samples to identify biomarkers predictive of recurrence. An independent validation profiling experiment was performed on 40 additional samples. MiRNA probes were filtered to retain only those that were present on the miRNA microarrays used for both the training and validation sets, reducing the total number of probes examined to 403 miRNA probes. The training set included 29 cases with observed biochemical PSA recurrence (median time to recurrence = 19 months), and 41 cases censored, i.e., without observed recurrence during follow-up (median follow-up time = 83.0 months). A summary of the clinical characteristics of the training and validation sets of samples is provided in Table 1. The complete dataset for the combined mRNA and miRNA data are provided in Supplementary Table S3 for the training set and Supplementary Table S4 for the validation set (see http://ajp.amjpathol.org).

Integrated DASL biomarker analysis

After fitting a univariate Cox proportional hazard (PH) model for each individual probe using the training data, a set of 27 important probes were preselected based on an FDR threshold of 0.30 (see Supplementary Table S5 at http://ajp.amjpathol.org). Next, to identify the optimal prediction score based on the preselected probes, we fit a lasso Cox proportional hazard (PH) model first using the set of 25 preselected mRNA probes only, resulting in a panel of nine protein-coding genes shown in Table 2 (RAD23B, FBP1, TNFRSF1A, NOTCH3, ETV1, BID, SIM2, ANXA1, and BCL2). A final prediction model was then built to include the predictive score based on this panel of nine mRNA
biomarkers as well as the relevant clinical biomarkers including T-stage, PSA and Gleason score, which could be used to predict recurrence following radical prostatectomy. Kaplan-Meier analysis (Figure 1A) demonstrated that these probes could significantly discriminate patients at higher and lower risk of recurrence by the log rank test (p < 0.001). We next applied the final predictive model developed on the training set to the validation set, a separate, independent DASL profiling experiment performed on a different day. Kaplan-Meier analysis (Figure 1B) on this validation set determined that the model could discriminate patients at higher and lower risk of recurrence (p = 0.010).

Subsequently, we repeated the above training procedure using the complete set of 27 preselected mRNA and miRNA probes, and we identified an optimal panel of ten mRNAs and two microRNAs (Table 3) and built a final prediction model for prostate cancer biochemical recurrence, which again included relevant clinical biomarkers. Kaplan-Meier analysis and the log-rank test determined that this panel could also significantly discriminate patients at higher and lower risk of recurrence both in the training set (p < 0.001, Figure 1C) and in the validation set (p = 0.013, Figure 1D).

**Prediction of Cases with a Gleason Score 7**

Prediction of recurrence for patients with a Gleason score 7 is particularly difficult. In order to address this issue, we applied the biomarker panels to the subset of cases in the training and validation sets that had a Gleason score 7. The prediction model based on the nine-mRNA panel was significant at discriminating biochemical recurrence in Gleason score 7 cases in both the training set (p < 0.001, Figure 2A) and the validation set (p = 0.027, Figure 2B). For the prediction model based on the combined panel of ten mRNAs and two miRNAs in Table 3, the predictive value was again significant
for both the training set ($p = < 0.001$, Figure 2C) and the validation set ($p = 0.010$, Figure 2D). A summary of the p-values for predicting biochemical recurrence is given in Table 4. In all cases, the prediction models that use one of the two gene biomarker panels plus clinical information outperforms the prediction model using only clinical information.

**Analysis of clinical recurrence**

Although most patients who have clinical recurrence following prostatectomy also have biochemical recurrence, there is a significant population of patients with biochemical recurrence who do not have clinically significant recurrences observed during their follow-ups. To evaluate our biomarker panel of biochemical recurrence for predicting the clinical recurrence, we tested the prediction model based on the combined mRNA/miRNA panel in the same training and validation samples using their clinical recurrence outcome data. Unfortunately, clinical recurrence data was lacking on some of the samples, and the total number of samples used in the training set was reduced. In the training data, the combined mRNA/miRNA panel was highly significant for predicting clinical recurrence in all patients ($p=0.002$) as well as in the subset of patients with a Gleason score 7 ($p=0.004$); in the validation data, it was also significant for predicting recurrence in patients with a Gleason score 7 ($p=0.023$) and trended towards significance in all patients ($p=0.078$). A summary of the p-values for predicting clinical recurrence is given in Table 5. In all cases, the prediction model that uses the combined mRNA and miRNA panel plus the clinical information, again, outperforms the prediction model that uses only the clinical information.

We also performed an analysis to construct a predictive set of biomarkers based on the clinical recurrence data instead of biochemical recurrence. Only three probes passed the initial preselection step for the univariate Cox PH modeling, all corresponding to the $ETV1$ gene, which is likely due to
the considerably fewer number of clinical recurrences in the training set as well as the smaller total sample size. Furthermore, the prediction model built on this set of gene biomarkers did not perform as well as the models built on biochemical recurrence (data not shown).
DISCUSSION

In the past few years, enormous progress has been made in developing technologies to exploit FFPE tumor tissue samples for gene expression and proteomic analysis. The use of FFPE tissues as a starting material is attractive because this approach should make biomarkers identified in this way much easier to translate into widespread clinical practice. DASL profiling makes it possible to define gene sets using FFPE prostate cancer tissues that could have potential prognostic and predictive value. For example, the DASL assay has been used recently to identify a 16-gene set that correlates with prostate cancer relapse\textsuperscript{11}. There was no overlap between our panel of ten mRNA and two miRNA biomarkers described here and the previously described 16-gene panel even though ten of the genes in the 16-gene panel previously reported were included in our 522 custom prostate DASL panel. When we analyzed the performance of the probes corresponding to those ten mRNAs in our dataset, we found that they were not able to significantly discriminate patients at higher or lower risk of recurrence. In this previous study, the gene signature selection and prediction model building were performed in separate steps and the signature selection was based on the correlation between the gene expression and Gleason score rather than between the gene expression and time to biochemical recurrence; our analytic approach overcomes these limitations. Specifically, our approach of building (training) prediction models takes advantage of recent advancement in regularized regression models for survival outcomes\textsuperscript{17,18}; regularized regression models can achieve simultaneous feature selection and model estimation and avoid model overfitting, leading to better prediction performance. Our use of a pre-selection step is similar to the recently proposed sure independence screening methods\textsuperscript{20,21}, which have been shown to achieve better performance in the presence of high-dimensional data for survival analysis compared to regularized regression without a pre-selection step\textsuperscript{22}.

Two other recent studies have employed DASL profiling to prostate cancer, but not detected
any signature that improved upon clinical models in validation sets\textsuperscript{23,24}. While these studies used large cohorts with long-term follow-up, they examined different panels of mRNA transcripts and did not include probes corresponding to miRNA genes. Moreover, these earlier studies suggested that tumor heterogeneity may play an important role in confounding signature identification. For our study of prostatectomy specimens, we identified the most prominent tumor lesion, and used a tissue core sample from that region to minimize stromal contributions and tumor heterogeneity.

In our twelve-gene predictive biomarker panel, nine of the genes are positively associated with recurrence, and three are negatively associated with recurrence. The nine genes positively associated with recurrence included \textit{miR-519d}, Notch homolog 3 (Notch3), Fructose-1,6-bisphosphatase 1 (FBP1), ETS variant gene 1 (ETV1), BH3 interacting domain death agonist (BID), Single-Minded homolog 2 (SIM2), RAD23 homolog B (RAD23B), LETM1 domain containing 1 (LETMD1), and Cyclin G2 (CCNG2). Little is known about \textit{miR-519d} other than it may be associated with obesity\textsuperscript{25}. NOTCH3 is one of four Notch family receptors in humans, and Notch signaling has been shown to be important for prostate cancer cell growth, migration, and invasion\textsuperscript{26,27} as well as normal prostate development\textsuperscript{28,29}. FBP1 is expressed in the prostate and is involved in gluconeogenesis\textsuperscript{30}. The identification of this metabolic enzyme as a biomarker of recurrence is initially surprising, but given the recent identification of isocitrate dehydrogenase 1 (IDH1) mutations in glioblastoma\textsuperscript{31}, and the fact that FBP1 was overexpressed in independent microarray analyses of prostate cancers\textsuperscript{7,32}, the potential of FBP1 as a biomarker should not be underestimated. ETV1 is well established as one of the commonly recurrent translocations found in prostate cancers\textsuperscript{9,15}, and has been used in clinical models of recurrence following prostatectomy\textsuperscript{33}. BID is a pro-apoptotic protein that binds to BCL2 and potentiates apoptotic responses upon cleavage in response to tumor necrosis factor alpha (TNF\(\alpha\)) and other death receptors\textsuperscript{34,35}. SIM2 was identified as a potential biomarker of prostate cancer in 2002\textsuperscript{36}.  

and later independently confirmed by Halvorsen et al\textsuperscript{37} and Arredouani et al\textsuperscript{38}. SIM2 functions as a transcription factor that represses the proapoptotic gene BNIP3\textsuperscript{39}. RAD23B plays a critical role in DNA damage recognition and nucleotide excision repair\textsuperscript{40}, as well as inhibiting MDM2-mediated degradation of the p53 tumor suppressor\textsuperscript{41}. LETMD1 (also known as HCCR) is an oncogene that is induced by Wnt\textsuperscript{42} and PI3K/AKT signaling\textsuperscript{43}, inhibits p53 function\textsuperscript{44}, and is a biomarker for hepatocellular\textsuperscript{45} and breast\textsuperscript{46} cancers. Cyclin G2 is an atypical cyclin that is induced by DNA damage\textsuperscript{47} in a p53-independent manner, as well as by PI3K/AKT/FOXO signals\textsuperscript{48}, and induces p53-dependent cell cycle arrest\textsuperscript{49}.

The three genes in the predictive biomarker panel negatively associated with recurrence were miR-647, the TNFα receptor (TNFRSF1A), and annexin A1 (ANXA1). While little is known about miR-647, TNFRSF1A (also known as TNFR1) mediates pro-apoptotic responses to TNFα ligand\textsuperscript{50,51}. Annexin A1 expression is reduced in early onset prostate cancer\textsuperscript{52} and high-grade prostatic intraepithelial neoplasia\textsuperscript{53}. ANXA1 plays important roles in vesicle trafficking and reduced ANXA1 promotes EMT and metastasis\textsuperscript{54}, and upregulates autocrine IL-6 signaling\textsuperscript{55}. Thus, as a whole, this panel of biomarkers appears to reflect changes in DNA stability, PI3K signaling, p53 activity, apoptosis, and differentiation consistent with more aggressive disease.

Although this study goes beyond a pilot study, enhanced by selection of samples from multiple institutions, the number of specimens tested is still relatively small. Re-analysis of our data using only the Emory samples for the training set did not identify any significant probes, likely due to the substantially smaller sample size. Thus, while the performance of our panel of biomarkers is significant, even for Gleason score 7 patients, future studies beyond the scope of this work will be necessary to perform independent validation on much larger sample sets with greater statistical power. Moreover, it is now feasible to perform DASL assays on virtually the entire genome, in an assay that
queries 24,526 transcripts derived from the RefSeq database. Future studies will test combined mRNA and miRNA biomarker panels, and query the entire genome to determine if other biomarker panels can achieve even greater success in prediction of biochemical and clinical recurrence of prostate cancer. Planned larger scale validation studies will determine whether these biomarkers are predictive for Gleason score 7 cases, and their utility at predicting clinical as well as biochemical recurrence.
REFERENCES

1. ACS: Cancer Facts and Figures 2009. Edited by Atlanta, American Cancer Society, 2009, p. 300


18. Goeman JJ: L1 penalized estimation in the Cox proportional hazards model, Biom J 2010, 52:70-84


22. Zhao SD, Li Y: Principled sure independence screening for Cox models with ultra-high-dimensional covariates, Harvard University Biostatistics Working Paper Series 2010,


39. Farrall AL, Whitelaw ML: The HIF1alpha-inducible pro-cell death gene BNIP3 is a novel target of SIM2s repression through cross-talk on the hypoxia response element, Oncogene 2009, 28:3671-3680


51. Thorburn A: Death receptor-induced cell killing, Cell Signal 2004, 16:139-144


ACKNOWLEDGEMENTS

The authors would like to thank Juliana DaSilva, Malini Krishnamoorthy, and Benjamin Barwick. This work was supported in part by the Emory Biomarker Service Center. Grant funding support came from NIH R01 CA 106826 and DOD PCRP IDEA Award PC093328. This research is funded in part by the Canadian Cancer Society Research Institute (grant # 019038).
### TABLES

<table>
<thead>
<tr>
<th>Number Cases</th>
<th>Clinical Recurrence</th>
<th>No Clinical Recurrence</th>
<th>Median Time F/U (months)</th>
<th>Median Time to BCR (months)</th>
<th>Median Time no BCR (months)</th>
<th>Gleason Score (Avg +/- SD)</th>
<th>PSA (Avg +/- SD)</th>
<th>Age (Avg +/- SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set (Total)</td>
<td>70</td>
<td>8</td>
<td>57</td>
<td>84</td>
<td>48</td>
<td>6.9 +/- 0.6</td>
<td>9.2 +/- 5.4</td>
<td>61.9 +/- 7.7</td>
</tr>
<tr>
<td>Training Set (No BCR)</td>
<td>41</td>
<td>0</td>
<td>41</td>
<td>83</td>
<td>83</td>
<td>6.7 +/- 0.6</td>
<td>8.7 +/- 6.4</td>
<td>61.2 +/- 7.7</td>
</tr>
<tr>
<td>Training Set (BCR)</td>
<td>29</td>
<td>8</td>
<td>16</td>
<td>81</td>
<td>19</td>
<td>7.0 +/- 0.6</td>
<td>9.9 +/- 3.8</td>
<td>62.9 +/- 7.8</td>
</tr>
<tr>
<td>Validation Set (Total)</td>
<td>40</td>
<td>11</td>
<td>29</td>
<td>74</td>
<td>34.5</td>
<td>7.0 +/- 0.8</td>
<td>12.7 +/- 8.4</td>
<td>63.6 +/- 8.4</td>
</tr>
<tr>
<td>Validation Set (No BCR)</td>
<td>27</td>
<td>0</td>
<td>27</td>
<td>75</td>
<td>56</td>
<td>6.8 +/- 0.7</td>
<td>12.4 +/- 9.9</td>
<td>63.5 +/- 8.3</td>
</tr>
<tr>
<td>Validation Set (BCR)</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>73</td>
<td>14</td>
<td>7.4 +/- 1</td>
<td>13.1 +/- 5.3</td>
<td>64 +/- 8.9</td>
</tr>
</tbody>
</table>

**Table 1:** A summary of the clinical characteristics of the training and validation sets of patient samples. (BCR = Biochemical Recurrence, F/U = follow up, PSA = prostate specific antigen, SD = standard deviation).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Coefficient</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD23B</td>
<td>RAD23 homolog B</td>
<td>0.152</td>
<td>40, 41</td>
</tr>
<tr>
<td>FBP1</td>
<td>Fructose-1,6-bisphosphatase 1</td>
<td>0.310</td>
<td>7, 30, 32</td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>Tumor Necrosis Factor Receptor Superfamily, Member 1A</td>
<td>-0.560</td>
<td>50, 51</td>
</tr>
<tr>
<td>NOTCH3</td>
<td>Notch homolog 3</td>
<td>0.426</td>
<td>26, 27</td>
</tr>
<tr>
<td>ETV1</td>
<td>Ets Variant Gene 1 (ETV1)</td>
<td>0.157</td>
<td>9, 15</td>
</tr>
<tr>
<td>BID</td>
<td>BH3 Interacting Domain Death Agonist (BID)</td>
<td>0.248</td>
<td>34, 35</td>
</tr>
<tr>
<td>SIM2</td>
<td>Single-Minded Homolog 2</td>
<td>0.043</td>
<td>36-38, 56</td>
</tr>
<tr>
<td>ANXA1</td>
<td>Annexin A1</td>
<td>-0.185</td>
<td>52-55</td>
</tr>
<tr>
<td>BCL2</td>
<td>B-cell CLL/lymphoma 2</td>
<td>0.028</td>
<td>53, 58</td>
</tr>
</tbody>
</table>

**Table 2:** Nine-gene predictor of prostate cancer recurrence following surgery. Coefficient is derived from the lasso Cox proportion hazards model and was used for computing the predictive score.
Positive coefficients indicate a positive association with recurrence, and negative coefficients a negative association with recurrence.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Coefficient</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD23B</td>
<td>RAD23 homolog B</td>
<td>0.070</td>
<td>40, 41</td>
</tr>
<tr>
<td>FBP1</td>
<td>Fructose-1,6-bisphosphatase 1</td>
<td>0.251</td>
<td>7, 30, 32</td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>Tumor necrosis factor receptor superfamily, member 1A</td>
<td>-0.588</td>
<td>50, 51</td>
</tr>
<tr>
<td>CCNG2</td>
<td>Cyclin G2</td>
<td>0.008</td>
<td>47-49</td>
</tr>
<tr>
<td>hsa-miR-647</td>
<td>hsa-miR-647</td>
<td>-0.318</td>
<td></td>
</tr>
<tr>
<td>LETMD1</td>
<td>LETM1 domain containing 1</td>
<td>0.063</td>
<td>42-44, 46</td>
</tr>
<tr>
<td>NOTCH3</td>
<td>Notch homolog 3</td>
<td>0.367</td>
<td>26, 27</td>
</tr>
<tr>
<td>ETV1</td>
<td>ETS variant gene 1 (ETV1)</td>
<td>0.179</td>
<td>9, 15</td>
</tr>
<tr>
<td>hsa-miR-519d</td>
<td>hsa-miR-519d</td>
<td>0.551</td>
<td>25</td>
</tr>
<tr>
<td>BID</td>
<td>BH3 interacting domain death agonist (BID)</td>
<td>0.128</td>
<td>34, 45</td>
</tr>
<tr>
<td>SIM2</td>
<td>Single-minded homolog 2</td>
<td>0.124</td>
<td>36-38, 56</td>
</tr>
<tr>
<td>ANXA1</td>
<td>Annexin A1</td>
<td>-0.143</td>
<td>52-55</td>
</tr>
</tbody>
</table>

**Table 3:** Twelve-gene predictor of prostate cancer recurrence following surgery using ten mRNAs and two microRNAs. Coefficient is derived from the lasso Cox proportion hazards model and was used for computing the predictive score. Positive coefficients indicate a positive association with recurrence, and negative coefficients a negative association with recurrence.
Table 4: Summary of p-values (Logrank test) of prediction of biochemical recurrence on training and validation sets for the entire dataset and the subset of Gleason score 7 cases using two biomarker panels, all of which are adjusted for T-stage, PSA, and Gleason score, or using clinical information only. Significant p-values are indicated in bold.

<table>
<thead>
<tr>
<th>Training Set</th>
<th>mRNA panel</th>
<th>Combined mRNA/miRNA panel</th>
<th>Clinical Information Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cases (n=61)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.096</td>
</tr>
<tr>
<td>Gleason score 7 (n=42)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.641</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation Set</th>
<th>mRNA panel</th>
<th>Combined mRNA/miRNA panel</th>
<th>Clinical Information Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cases (n=35)</td>
<td>0.010</td>
<td>0.013</td>
<td>0.020</td>
</tr>
<tr>
<td>Gleason score 7 (n=19)</td>
<td>0.027</td>
<td>0.010</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Table 5: Summary of p-values (Logrank test) of prediction of clinical recurrence on training and validation sets for the entire dataset and the subset of Gleason score 7 cases using the combined and mRNA/miRNA panel, all of which are adjusted for T-stage, PSA, and Gleason score, or using clinical information only. Significant p-values are indicated in bold.

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Combined mRNA/miRNA panel</th>
<th>Clinical Information Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cases (n=56)</td>
<td>0.002</td>
<td>0.262</td>
</tr>
<tr>
<td>Gleason score 7 (n=37)</td>
<td>0.004</td>
<td>0.136</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation Set</th>
<th>Combined mRNA/miRNA panel</th>
<th>Clinical Information Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Cases (n=35)</td>
<td>0.078</td>
<td>0.193</td>
</tr>
<tr>
<td>Gleason score 7 (n=19)</td>
<td>0.023</td>
<td>0.080</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1: Prediction of biochemical recurrence in all prostate cancer patients using two biomarker panels, adjusted for clinical information. (A) Kaplan-Meier analysis of the training set patients that were separated based on the mRNA panel described in Table 2. (B) Kaplan-Meier analysis on the validation cases using the mRNA panel. (C) Kaplan-Meier analysis of the training set using the combined mRNA and miRNA panel described in Table 3. (D) Kaplan-Meier analysis of the validation set using the combined mRNA and miRNA panel.

Figure 2: Prediction of biochemical recurrence in prostate cancer patients with a Gleason score 7 using two biomarker panels, adjusted for clinical information. (A) Kaplan-Meier analysis of the training set of Gleason score 7 cases using the mRNA panel described in Table 2. (B) Kaplan-Meier analysis of the Gleason score 7 cases in the validation set using the mRNA panel. (C) Kaplan-Meier analysis of the Gleason score 7 cases in the training set using the combined mRNA and miRNA panel described in Table 3. (D) Kaplan-Meier analysis of the Gleason score 7 cases in the validation set using the combined mRNA and miRNA panel.