

Biomarkers, Genomics, Proteomics, and Gene Regulation

Protein-Coding and MicroRNA Biomarkers of Recurrence of Prostate Cancer Following Radical Prostatectomy

Qi Long,^{*†} Brent A. Johnson,^{*}
Adeboye O. Osunkoya,^{†‡§¶} Yu-Heng Lai,^{¶||}
Wei Zhou,^{†||**††} Mark Abramovitz,^{‡‡}
Mingjing Xia,[†] Mark B. Bouzyk,^{†,††}
Robert K. Nam,^{§§} Linda Sugar,^{§§}
Aleksandra Stanimirovic,^{§§} Daron J. Williams,[‡]
Brian R. Leyland-Jones,^{†**} Arun K. Seth,^{§§}
John A. Petros,^{†‡§¶**} and Carlos S. Moreno^{†||}

From the Departments of Biostatistics and Bioinformatics,^{*} Pathology and Laboratory Medicine,[‡] Urology,[§] Hematology and Medical Oncology,^{**} and Human Genetics,^{††} the Winship Cancer Institute,[†] and the Program in Genetics and Molecular Biology,^{||} Emory University, Atlanta, Georgia; the Atlanta VA Medical Center,[¶] Decatur, Georgia; the Virginia Mason Institute of Research,^{‡‡} Montreal, Quebec, Canada; and the Department of Laboratory Medicine and Pathobiology, University of Toronto and Department of Anatomic Pathology,^{§§} Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

An important challenge in prostate cancer research is to develop effective predictors of tumor recurrence following surgery 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 radical prostatectomy specimens with known long-term outcomes to perform DASL expression profiling with a custom panel that we designed of 522 prostate cancer-relevant genes. We identified a panel of 10 protein-coding genes and two miRNA genes (*RAD23B*, *FBP1*, *TNFRSF1A*, *CCNG2*, *NOTCH3*, *ETV1*, *BID*, *SIM2*, *LETMD1*, *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, prostate-specific antigen, 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. (*Am J Pathol* 2011, 179:46–54; DOI: 10.1016/j.ajpath.2011.03.008)

Prostate cancer remains the most common noncutaneous 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.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, to aid clinicians in deciding on the appropriate

Supported in part by an Emory University Research Committee grant and the PHS Grant UL1 RR025008 from the Clinical and Translational Science Award program, National Institutes of Health, National Center for Research Resources; NIH R01 CA106826 and DOD PCRP IDEA Award PC093328 (to C.S.M.); Emory Cancer Genomics Shared Resource and the Winship Cancer Center Support Grant (NIH P30CA138292); and the Canadian Cancer Society Research Institute (grant # 019038).

Accepted for publication March 3, 2011.

Supplemental material for this article can be found at <http://ajp.amjpathol.org> or doi:10.1016/j.ajpath.2011.03.008.

Current address of M.A., Departments of Pathology & Oncology, Jewish General Hospital, Montreal, QC, Canada.

Address reprint requests to Carlos S. Moreno, Ph.D., Department of Pathology & Laboratory Medicine, Winship Cancer Institute, Emory University, Whitehead Research Building, Rm. 105J, 615 Michael St., Atlanta, GA 30322. E-mail: cmoreno@emory.edu.

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 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 on 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.^{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 (miRNA) 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 prostate cancer cases were used (29 with biochemical recurrence and 41 controls without recurrence), 45 patients from Sunnybrook Health Science Center (Toronto, ON), and 25 patients from Emory University (Atlanta, GA). 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 between 1998 and 2006. The clinical data include multiple clinicopathological variables such as prostate-specific antigen (PSA) levels, histological grade (Gleason score), tumor stage (pathological 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 institutional review board-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, had documented long-term follow-up, and consented to participate or were included by institutional review board waiver. The cases were assigned prostate identification numbers to protect their identities. These patients did not receive neoadjuvant 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 Supplemental Table S1 (available 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 a very high percentage tumor content. H&E-stained slides were reviewed by a board-certified urologic pathologist (A.O.O.) 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,⁴ 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 a NanoDrop spectrophotometer (NanoDrop, 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), A_{260}/A_{280} ratio >1.8, and RPL13a C_T values less than 30 cycles were used for miRNA and DASL profiling.

Custom Prostate Cancer DASL Assay Pool

The DASL assay enables quantitation of expression of up to 1536 probes using RNA isolated from archived FFPE tumor tissue samples in a high throughput format.^{2,3} Data from multiple publicly available gene expression datasets,⁵⁻⁸ along with genes involved in prostate cancer progression based on current understanding of the disease,^{6,9} 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* fusion 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,¹⁰ Bibikova et al,¹¹ True et al,¹² Lapointe et al,⁷ and/or Singh et al.¹³ It also includes prognostic markers from Dhanasekaran et al⁵ and Yu et al,¹⁴ and genes associated with metastasis in Varambally et al.⁶ In addition, a number of genes known from other

studies to be critical in prostate cancer such as *NKX3.1*, *PTEN*, and the androgen receptor are all included in the panel. Other genes that play important roles in the Wnt, Hedgehog, TGF β , Notch, MAPK, and PI3K pathways are also present in this gene set. Finally, primer sets that detect chromosomal translocations in *ERG*,⁹ *ETV1*,¹⁵ and *ETV4*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 (Illumina, San Diego, CA). The optimal oligonucleotide sequence for each of the 1536 gene probes was determined using an oligonucleotide scoring algorithm. The oligonucleotide pool or DASL Assay Pool was synthesized by Illumina for use with the 96-well Universal Array Matrix.

The DASL 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 (two, three, or four) were hybridized on Universal Array Matrices, and scanned using the BeadStation 500 Instrument (Illumina). For miRNA DASL assays, the human miRNA v2 DASL panel (Illumina), which allows for the determination of expression of 1146 human miRNAs (>97% coverage of miRBase release 12), was used. These data are available in the Gene Expression Omnibus (GEO) database under accession number GSE26367.

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 (SD 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 and subsequently verified by a positive biopsy, and/or a positive imaging study [ProstaScint (EUSA Pharma, Langhorne, PA) or computed tomography scan] accompanied by a detectable postoperative PSA result and lack of evidence for metastases. Also, patients whose PSA levels 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 biochem-

ical recurrence. Specifically, we first fit a univariate Cox proportional hazard (PH) model for each individual oligonucleotide probe using the training dataset, and a set of important mRNA and miRNA probes were then preselected based on a false discovery rate threshold of 0.30. Next, to identify the optimal prediction score based on the preselected probes, we fit a lasso Cox PH model^{17,18} using the training dataset, where the tuning parameter for lasso was selected using a leave-one-out cross-validation technique.¹⁸ 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 dataset, 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 datasets.

Missing data are present in this study, in particular, for clinical recurrence, PSA, and T-stage data. We adopted the available-case approach¹⁹ 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 the *Materials and Methods* section) and the Illumina DASL microRNA 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

Table 1. A Summary of the Clinical Characteristics of the Training and Validation Sets of Patient Samples

	Training set (total)	Training set (no BCR)	Training set (BCR)	Validation set (total)	Validation set (no BCR)	Validation set (BCR)
Number cases	70	41	29	40	27	13
Clinical recurrence	8	0	8	11	0	11
No clinical recurrence	57	41	16	29	27	2
Median time F/U (months)	84	83	81	74	75	73
Median time to BCR (months)	19	NA	19	14	NA	14
Median time no BCR (months)	48	83	19	34.5	56	14
Gleason score (avg ± SD)	6.9 ± 0.6	6.7 ± 0.6	7.0 ± 0.6	7.0 ± 0.8	6.8 ± 0.7	7.4 ± 1
PSA (avg ± SD)	9.2 ± 5.4	8.7 ± 6.4	9.9 ± 3.8	12.7 ± 8.4	12.4 ± 9.9	13.1 ± 5.3
Age (avg ± SD)	61.9 ± 7.7	61.2 ± 7.7	62.9 ± 7.8	63.6 ± 8.4	63.5 ± 8.3	64 ± 8.9

BCR, biochemical recurrence; F/U, follow up; NA, not applicable; PSA, prostate specific antigen; SD, standard deviation.

months), and 41 cases censored, ie, 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 Supplemental Table S3 for the training set and Supplemental Table S4 for the validation set (available at <http://ajp.amjpathol.org>).

Integrated DASL Biomarker Analysis

After fitting a univariate Cox proportional hazard model for each individual probe using the training data, a set of 27 important probes were preselected based on a false discovery rate threshold of 0.30 (see Supplemental 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 model^{17,18} 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 10 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. 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 10 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 ($P = 0.010$,

Table 2. Nine-Gene Predictor of Prostate Cancer Recurrence Following Surgery

Symbol	Description	Coefficient	References
<i>RAD23B</i>	RAD23 homolog B	0.152	20, 21
<i>FBP1</i>	Fructose-1,6-bisphosphatase 1	0.310	7, 10, 33
<i>TNFRSF1A</i>	Tumor necrosis factor receptor superfamily, member 1A	-0.560	23, 24
<i>NOTCH3</i>	Notch homolog 3	0.426	25, 26
<i>ETV1</i>	Ets variant gene 1 (ETV1)	0.157	9, 15
<i>BID</i>	BH3 interacting domain death agonist (BID)	0.248	27, 28
<i>SIM2</i>	Single-minded homolog 2	0.043	29-32
<i>ANXA1</i>	Annexin A1	-0.185	33-36
<i>BCL2</i>	B-cell CLL/lymphoma 2	0.028	37-38

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 indicate a negative association with recurrence.

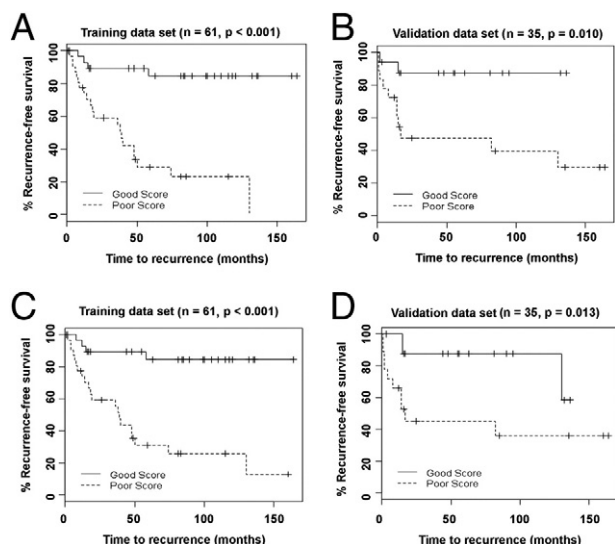


Figure 1. Prediction of biochemical recurrence in all prostate cancer patients using two biomarker panels, adjusted for clinical information. **A:** Kaplan-Meier analysis of the training set patients that were separated based on the mRNA panel described in Table 2. **B:** Kaplan-Meier analysis on the 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.

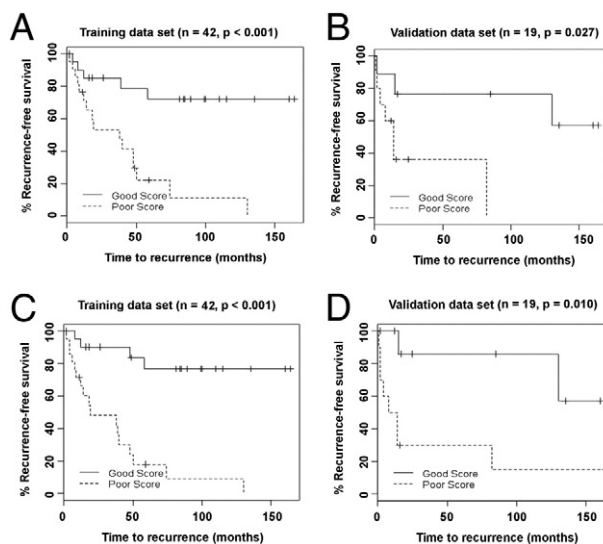


Figure 2. Prediction of biochemical recurrence in prostate cancer patients with a Gleason score 7 using two biomarker panels, adjusted for clinical information. **A:** Kaplan-Meier analysis of the training set of Gleason score 7 cases using the mRNA panel described in Table 2. **B:** Kaplan-Meier analysis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

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 were 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 toward 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.

Table 3. Twelve-Gene Predictor of Prostate Cancer Recurrence Following Surgery Using 10 mRNAs and Two microRNAs

Symbol	Description	Coefficient	References
<i>RAD23B</i>	RAD23 homolog B	0.070	20, 21
<i>FBP1</i>	Fructose-1,6-bisphosphatase 1	0.251	7, 10, 22
<i>TNFRSF1A</i>	Tumor necrosis factor receptor superfamily, member 1A	-0.588	23, 24
<i>CCNG2</i>	Cyclin G2	0.008	39-41
<i>hsa-miR-647</i>	hsa-miR-647	-0.318	
<i>LETMD1</i>	LETM1 domain containing 1	0.063	42-44, 45
<i>NOTCH3</i>	Notch homolog 3	0.367	25, 26
<i>ETV1</i>	ETS variant gene 1 (ETV1)	0.179	9, 15
<i>hsa-miR-519 days</i>	hsa-miR-519 days	0.551	46
<i>BID</i>	BH3 interacting domain death agonist (BID)	0.128	27, 28
<i>SIM2</i>	Single-minded homolog 2	0.124	29-32
<i>ANXA1</i>	Annexin A1	-0.143	33-36

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 (Log-Rank Test) of Prediction of Biochemical Recurrence on Training and Validation Sets

Dataset	Combined		Clinical information only
	mRNA panel	mRNA/miRNA panel	
Training			
All cases (<i>n</i> = 61)	<0.001	<0.001	0.096
Gleason score 7 (<i>n</i> = 42)	<0.001	<0.001	0.641
Validation			
All cases (<i>n</i> = 35)	0.010	0.013	0.020
Gleason score 7 (<i>n</i> = 19)	0.027	0.010	0.028

Prediction of biochemical recurrence for the entire dataset and the subset of Gleason score 7 cases was made using two biomarker panels adjusted for T-stage, PSA, and Gleason score, or using clinical information only. Significant *P* values are indicated in bold.

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.¹¹ There was no overlap between our panel of 10 mRNA and two miRNA biomarkers described here and the previously described 16-gene panel even though 10 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 10 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^{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 preselection step is similar to the recently proposed sure independence screening methods,^{47,48} which have been shown to achieve better performance in the presence of high-dimensional data for survival analysis compared to regularized regression without a preselection step.⁴⁹

Two other recent studies have used DASL profiling to prostate cancer, but have not detected any signature that improved on clinical models in validation sets.^{50,51} Although 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 12-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 *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 *miR-519d* other than it may be associated with obesity.⁴⁶ NOTCH3 is one of four Notch

Table 5. Summary of *P* Values (Log-Rank Test) of Prediction of Clinical Recurrence on Training and Validation Sets

Dataset	Combined mRNA/ miRNA panel	Clinical information only
Training		
All cases (<i>n</i> = 56)	0.002	0.262
Gleason score 7 (<i>n</i> = 37)	0.004	0.136
Validation		
All cases (<i>n</i> = 35)	0.078	0.193
Gleason score 7 (<i>n</i> = 19)	0.023	0.080

Prediction of clinical recurrence for the entire dataset and the subset of Gleason score 7 cases was made using the combined mRNA/miRNA panel, adjusted for T-stage, PSA, and Gleason score, or using clinical information only. Significant *P* values are indicated in bold.

family receptors in humans, and Notch signaling has been shown to be important for prostate cancer cell growth, migration, and invasion,^{25,26} as well as normal prostate development.^{52,53} FBP1 is expressed in the prostate and is involved in gluconeogenesis.²² 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,⁵⁴ and the fact that FBP1 was overexpressed in independent microarray analyses of prostate cancers,^{7,10} 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,^{9,15} and has been used in clinical models of recurrence following prostatectomy.⁵⁵ BID is a proapoptotic protein that binds to BCL2 and potentiates apoptotic responses on cleavage in response to tumor necrosis factor α (TNF α) and other death receptors.^{27,28} SIM2 was identified as a potential biomarker of prostate cancer in 2002²⁹ and later independently confirmed by Halvorsen et al³⁰ and Arredouani et al.³¹ SIM2 functions as a transcription factor that represses the proapoptotic gene *BNIP3*.⁵⁶ RAD23B plays a critical role in DNA damage recognition and nucleotide excision repair,²⁰ as well as inhibiting MDM2-mediated degradation of the p53 tumor suppressor.²¹ *LETMD1* (also known as *HCCR*) is an oncogene that is induced by Wnt⁴² and PI3K/AKT signaling,⁴³ inhibits p53 function,⁴⁴ and is a biomarker for hepatocellular⁵⁷ and breast⁴⁵ cancers. Cyclin G2 is an atypical cyclin that is induced by DNA damage³⁹ in a p53-independent manner, as well as by PI3K/AKT/FOXO signals,⁴⁰ and induces p53-dependent cell cycle arrest.⁴¹

The three genes in the predictive biomarker panel negatively associated with recurrence were *miR-647*, the TNF α receptor (*TNFRSF1A*), and annexin A1 (*ANXA1*). Although little is known about *miR-647*, *TNFRSF1A* (also known as *TNFR1*) mediates proapoptotic responses to TNF α ligand.^{23,24} Annexin A1 expression is reduced in early onset prostate cancer³³ and high-grade prostatic intraepithelial neoplasia.³⁴ *ANXA1* plays important roles in vesicle trafficking and reduced *ANXA1* promotes EMT and metastasis,³⁵ and up-regulates autocrine IL-6 signaling.³⁶ 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, although 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 whether 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.

Acknowledgments

We thank Juliana DaSilva, Malini Krishnamoorthy, Benjamin Barwick, and the Winship Cancer Genomics Shared Resource.

References

1. American Cancer Society: Cancer Facts & Figures 2009. Atlanta, American Cancer Society, 2009
2. Bibikova M, Talantov D, Chudin E, Yeakley JM, Chen J, Doucet D, Wickham E, Atkins D, Barker D, Chee M, Wang Y, Fan JB: Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am J Pathol* 2004, 165:1799–1807
3. Fan JB, Yeakley JM, Bibikova M, Chudin E, Wickham E, Chen J, Doucet D, Rigault P, Zhang B, Shen R, McBride C, Li HR, Fu XD, Oliphant A, Barker DL, Chee MS: A versatile assay for high-throughput gene expression profiling on universal array matrices. *Genome Res* 2004, 14:878–885
4. Abramovitz M, Ordanic-Kodani M, Wang Y, Li Z, Catzavelos C, Bouzyk M, Sledge GW Jr., Moreno CS, Leyland-Jones B: Optimization of RNA extraction from FFPE tissues for expression profiling in the DASL assay. *Biotechniques* 2008, 44:417–423
5. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM: Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001, 412:822–826
6. Varambally S, Yu J, Laxman B, Rhodes DR, Mehra R, Tomlins SA, Shah RB, Chandran U, Monzon FA, Becich MJ, Wei JT, Pienta KJ, Ghosh D, Rubin MA, Chinnaiyan AM: Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* 2005, 8:393–406
7. Lapointe J, Li C, Higgins JP, Van De Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, Ekman P, DeMarzo AM, Tibshirani R, Botstein D, Brown PO, Brooks JD, Pollack JR: Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* 2004, 101:811–816
8. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL: Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res* 2002, 62:4499–4506
9. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM: Recurrent fusion of *TMPRSS2* and *ETS* transcription factor genes in prostate cancer. *Science* 2005, 310:644–648
10. Liu P, Ramachandran S, Ali-Seyed M, Scharer CD, Laycock N, Dalton WB, Williams H, Karanam S, Datta MW, Jaye D, Moreno CS: Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells. *Cancer Res* 2006, 66:4011–4019
11. Bibikova M, Chudin E, Arsanjani A, Zhou L, Garcia EW, Modder J, Kostelec M, Barker D, Downs T, Fan JB, Wang-Rodriguez J: Expression signatures that correlated with Gleason score and relapse in prostate cancer. *Genomics* 2007, 89:666–672
12. True L, Coleman I, Hawley S, Huang CY, Gifford D, Coleman R, Beer TM, Gelmann E, Datta M, Mostaghel E, Knudsen B, Lange P, Vessella R, Lin D, Hood L, Nelson PS: A molecular correlate to the Gleason grading system for prostate adenocarcinoma. *Proc Natl Acad Sci U S A* 2006, 103:10991–10996

13. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Golub TR, Sellers WR: Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 2002, 1:203–209
14. Yu YP, Landsittel D, Jing L, Nelson J, Ren B, Liu L, McDonald C, Thomas R, Dhir R, Finkelstein S, Michalopoulos G, Becich M, Luo JH: Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J Clin Oncol* 2004, 22:2790–2799
15. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, Yu J, Wang L, Montie JE, Rubin MA, Pienta KJ, Roulston D, Shah RB, Varambally S, Mehra R, Chinnaiyan AM: Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 2007, 448:595–599
16. Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, Helgeson BE, Cao X, Wei JT, Rubin MA, Shah RB, Chinnaiyan AM: TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res* 2006, 66:3396–3400
17. Tibshirani R: The lasso method for variable selection in the Cox model. *Stat Med* 1997, 16:385–395
18. Goeman JJ: L1 penalized estimation in the Cox proportional hazards model. *Biom J* 2010, 52:70–84
19. Little RJA, Rubin DB: *Statistical Analysis with Missing Data*. New York, Wiley, 2002
20. Thoma BS, Vasquez KM: Critical DNA damage recognition functions of XPC-hHR23B and XPA-RPA in nucleotide excision repair. *Mol Carcinog* 2003, 38:1–13
21. Brignone C, Bradley KE, Kisselev AF, Grossman SR: A post-ubiquitination role for MDM2 and hHR23A in the p53 degradation pathway. *Oncogene* 2004, 23:4121–4129
22. Yanez AJ, Nualart F, Droppelmann C, Bertinat R, Brito M, Concha II, Slebe JC: Broad expression of fructose-1,6-bisphosphatase and phosphoenolpyruvate carboxykinase provide evidence for gluconeogenesis in human tissues other than liver and kidney. *J Cell Physiol* 2003, 197:189–197
23. Hsu H, Xiong J, Goeddel DV: The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. *Cell* 1995, 81:495–504
24. Thorburn A: Death receptor-induced cell killing. *Cell Signal* 2004, 16:139–144
25. Wang Z, Li Y, Ahmad A, Banerjee S, Azmi AS, Kong D, Wojewoda C, Miele L, Sarkar FH: Down-regulation of Notch-1 is associated with Akt and FoxM1 in inducing cell growth inhibition and apoptosis in prostate cancer cells. *J Cell Biochem* 2011, 112:78–88
26. Wang Z, Li Y, Banerjee S, Kong D, Ahmad A, Nogueira V, Hay N, Sarkar FH: Down-regulation of Notch-1 and Jagged-1 inhibits prostate cancer cell growth, migration and invasion, and induces apoptosis via inactivation of Akt, mTOR, and NF-kappaB signaling pathways. *J Cell Biochem* 2010, 109:726–736
27. Kulik G, Carson JP, Vomastek T, Overman K, Gooch BD, Srinivasula S, Alnemri E, Nunez G, Weber MJ: Tumor necrosis factor alpha induces BID cleavage and bypasses antiapoptotic signals in prostate cancer LNCaP cells. *Cancer Res* 2001, 61:2713–2719
28. Luo X, Budihardjo I, Zou H, Slaughter C, Wang X: Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell* 1998, 94:481–490
29. Deyoung MP, Scheurle D, Damania H, Zylberberg C, Narayanan R: Down's syndrome-associated single minded gene as a novel tumor marker. *Anticancer Res* 2002, 22:3149–3157
30. Halvorsen OJ, Rostad K, Oyan AM, Puntervoll H, Bo TH, Stordrange L, Olsen S, Haukaas SA, Hood L, Jonassen I, Kalland KH, Akslen LA: Increased expression of SIM2-s protein is a novel marker of aggressive prostate cancer. *Clin Cancer Res* 2007, 13:892–897
31. Arredouani MS, Lu B, Bhasin M, Eljanne M, Yue W, Mosquera JM, Bubleky GJ, Li V, Rubin MA, Libermann TA, Sanda MG: Identification of the transcription factor single-minded homologue 2 as a potential biomarker and immunotherapy target in prostate cancer. *Clin Cancer Res* 2009, 15:5794–5802
32. DeYoung MP, Tress M, Narayanan R: Identification of Down's syndrome critical locus gene SIM2-s as a drug therapy target for solid tumors. *Proc Natl Acad Sci U S A* 2003, 100:4760–4765
33. Paweletz CP, Ornstein DK, Roth MJ, Bichsel VE, Gillespie JW, Calvert VS, Vocke CD, Hewitt SM, Duray PH, Herring J, Wang QH, Hu N, Linehan WM, Taylor PR, Liotta LA, Emmert-Buck MR, Petricoin EF, 3rd: Loss of annexin 1 correlates with early onset of tumorigenesis in esophageal and prostate carcinoma. *Cancer Res* 2000, 60:6293–6297
34. Patton KT, Chen HM, Joseph L, Yang XJ: Decreased annexin I expression in prostatic adenocarcinoma and in high-grade prostatic intraepithelial neoplasia. *Histopathology* 2005, 47:597–601
35. Maschler S, Gebeshuber CA, Wiedemann EM, Alacakaptan M, Schreiber M, Custic I, Beug H: Annexin A1 attenuates EMT and metastatic potential in breast cancer. *EMBO Mol Med* 2010, 2:401–414
36. Inokuchi J, Lau A, Tyson DR, Ornstein DK: Loss of annexin A1 disrupts normal prostate glandular structure by inducing autocrine IL-6 signaling. *Carcinogenesis* 2009, 30:1082–1088
37. Alnemri ES, Robertson NM, Fernandes TF, Croce CM, Litwack G: Overexpressed full-length human BCL2 extends the survival of baculovirus-infected Sf9 insect cells. *Proc Natl Acad Sci U S A* 1992, 89:7295–7299
38. Wagner AJ, Small MB, Hay N: Myc-mediated apoptosis is blocked by ectopic expression of Bcl-2. *Mol Cell Biol* 1993, 13:2432–2440
39. Bates S, Rowan S, Vousden KH: Characterisation of human cyclin G1 and G2: dNA damage inducible genes. *Oncogene* 1996, 13:1103–1109
40. Martinez-Gac L, Marques M, Garcia Z, Campanero MR, Carrera AC: Control of cyclin G2 mRNA expression by forkhead transcription factors: novel mechanism for cell cycle control by phosphoinositide 3-kinase and forkhead. *Mol Cell Biol* 2004, 24:2181–2189
41. Arachchige Don AS, Dallapiazza RF, Bennis DA, Brake T, Cowan CE, Horne MC: Cyclin G2 is a centrosome-associated nucleocytoplasmic shuttling protein that influences microtubule stability and induces a p53-dependent cell cycle arrest. *Exp Cell Res* 2006, 312:4181–4204
42. Cho GW, Kim MH, Kim SH, Ha SA, Kim HK, Kim S, Kim JW: TCF/beta-catenin plays an important role in HCCR-1 oncogene expression. *BMC Mol Biol* 2009, 10:42
43. Cho GW, Shin SM, Namkoong H, Kim HK, Ha SA, Hur SY, Kim TE, Chai YG, Kim JW: The phosphatidylinositol 3-kinase/Akt pathway regulates the HCCR-1 oncogene expression. *Gene* 2006, 384:18–26
44. Ko J, Lee YH, Hwang SY, Lee YS, Shin SM, Hwang JH, Kim J, Kim YW, Jang SW, Ryoo ZY, Kim IK, Namkoong SE, Kim JW: Identification and differential expression of novel human cervical cancer oncogene HCCR-2 in human cancers and its involvement in p53 stabilization. *Oncogene* 2003, 22:4679–4689
45. Jung SS, Park HS, Lee IJ, Namkoong H, Shin SM, Cho GW, Ha SA, Park YG, Lee YS, Ko J, Kim JW: The HCCR oncogene as a biomarker for human breast cancer. *Clin Cancer Res* 2005, 11:7700–7708
46. Martinelli R, Nardelli C, Pilone V, Buonomo T, Liguori R, Castano I, Buono P, Masone S, Persico G, Forestieri P, Pastore L, Sacchetti L: miR-519d overexpression is associated with human obesity. *Obesity (Silver Spring)* 2010, 18:2170–2176
47. Fan J, Lv J: Sure independence screening for ultrahigh dimensional feature space. *J R Stat Soc Ser B Stat Methodol* 2008, 70:849–911
48. Fan J, Song R: Sure independence screening in generalized linear models and NP-dimensionality. *Ann Stat* 2010, 38:3567–3604
49. Zhao SD, Li Y: Principled sure independence screening for Cox models with ultra-high-dimensional covariates. *Harvard University Biostatistics Working Paper Series* 2010. Boston, Harvard University Department of Biostatistics, 2010
50. Sboner A, Demichelis F, Calza S, Pawitan Y, Setlur SR, Hoshida Y, Perner S, Adami HO, Fall K, Mucci LA, Kantoff PW, Stampfer M, Andersson SO, Varenhorst E, Johansson JE, Gerstein MB, Golub TR, Rubin MA, Andren O: Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med Genomics* 2010, 3:8
51. Nakagawa T, Kollmeyer TM, Morlan BW, Anderson SK, Bergstralh EJ, Davis BJ, Asmann YW, Klee GG, Ballman KV, Jenkins RB: A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS ONE* 2008, 3:e2318
52. Leong KG, Gao WQ: The Notch pathway in prostate development and cancer. *Differentiation* 2008, 76:699–716

53. Wang XD, Leow CC, Zha J, Tang Z, Modrusan Z, Radtke F, Aguet M, de Sauvage FJ, Gao WQ: Notch signaling is required for normal prostatic epithelial cell proliferation and differentiation. *Dev Biol* 2006, 290:66–80
54. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr., Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW: An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008, 321:1807–1812
55. Cheville JC, Karnes RJ, Therneau TM, Kosari F, Munz JM, Tillmans L, Basal E, Rangel LJ, Bergstralh E, Kovtun IV, Savci-Heijink CD, Klee EW, Vasmatazis G: Gene panel model predictive of outcome in men at highrisk of systemic progression and death from prostate cancer after radical retropubic prostatectomy. *J Clin Oncol* 2008, 26:3930–3936
56. 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
57. Yoon SK, Lim NK, Ha SA, Park YG, Choi JY, Chung KW, Sun HS, Choi MJ, Chung J, Wands JR, Kim JW: The human cervical cancer oncogene protein is a biomarker for human hepatocellular carcinoma. *Cancer Res* 2004, 64:5434–5441