**Novel three gene molecular signature for prostate cancer recurrence**

Raymond B. Nagle\(^{a,b,1}\), Ubaradka G. Sathyanarayana\(^{a,1}\), Janice P. Riley\(^{a}\), Jackie Neri\(^{b}\), Bruce Dalkin\(^{b}\), Amit Algotar\(^{b}\), Ned C. Haubein\(^{d}\), G. Scott Lett\(^{d}\), Phillip Miller\(^{e}\) & Gary A. Pestano\(^{e}\)

\(^{a}\)Ventana Medical Systems, Inc., Tucson, AZ 85755, USA; \(^{b}\)University of Arizona \(^{c}\)Cancer Center and \(^{d}\)Department of Urology, School of Medicine, AZ 85724, USA; and \(^{e}\)The BioAnalytics Group, LLC, Jamesburg, NJ, USA.

\(^{1}\)First two authors contributed equally to this work.

**Correspondence**
Raymond B. Nagle
University of Arizona Cancer Center, 1515 N Cambell Ave., Tucson, AZ 85724, USA.
E-mail: rnagle@email.arizona.edu

**Abstract**
The goal of this study was to identify potentially useful low complexity molecular signatures for prostate cancer recurrence. Twenty-eight primary carcinomas from FFPE cases with or without prostate cancer recurrence at least five years post-surgery were selected for multiplexed mRNA analysis by DASL assay. Marked over-expression of \textit{WNT5A} and \textit{TK1} was noted in recurrent prostate cancer samples, whereas up-regulation of \textit{GAS1} was associated with non-recurrence of prostate cancer. Significantly, this pattern of expression was consistently observed for these genes and also the AUC was 0.846 for this panel in predicting individual’s likelihood of recurrence. The qRT-PCR assay confirmed the expression data for the 3 genes obtained on DASL assay. Further, Spearman Rank Correlation Coefficient analysis indicated that the two experimental data showed significant correlation. We have identified a three gene signature, characterized by over-expression of \textit{WNT5A} and \textit{TK1}, and down-regulation of \textit{GAS1}, that may predict the likelihood of prostate cancer recurrence.

**Keywords:** Prostate cancer, gene expression, tissue diagnostic, biomarker

**Abbreviations used:** FFPE, formalin fixed paraffin embedded; IHC, immuno-histochemistry; DASL, cDNA-mediated annealing, selection, extension, and ligation; qRT-PCR, quantitative Real-Time PCR; LOOCV, leave one out cross validation, AUC Area Under the receiver operating characteristic Curve.

**Introduction**
Prostate cancer is the most common visceral neoplasm in men. In 2013, it is estimated that prostate cancer alone will account for 238,590 new cases and will account for 10% of all cancer deaths in men \(^{1}\). The incidence increases with age and the routine availability of serum prostate specific antigen (PSA) testing has dramatically increased the number of aging men having the diagnosis. In most men, the disease is slowly progressive, but a significant number progress to metastatic disease, which in time becomes androgen-independent. The choice of appropriate treatment is usually dependent on the age of the patient and the stage of the prostate cancer. This decision is complicated by the lack of available accurate methods to pre-surgically determine the clinical stage and the biologic potential of a given patient.

An important clinical question is how aggressively to treat such patients with localized prostate cancer. Further, current treatment considerations are based on clinical stage, biopsy-determined Gleason grade and prospective serum PSA levels. However, these prognostic indicators do not accurately predict clinical outcome for individual patients. Hence, critical understanding of the molecular abnormalities that define those tumors at high risk for relapse is needed to help identify more precise molecular markers.

Unlike many tumor types, specific patterns of oncogene expression have not been consistently identified in prostate cancer progression, although a number of
candidate genes and pathways likely to be important in individual cases have been identified. A number of groups have attempted to examine prostate cancer progression by comparing gene expression of primary carcinomas to normal prostate (see catalogue at http://www.oncomine.org). Because of differences in technique as well as the true biologic heterogeneity seen in prostate cancer, these studies have reported thousands of candidate genes, but shared only moderate consensus. Nevertheless, a few genes have emerged, including hepsin (HPN), α-methylacyl-CoA racemase (AMACR), and enhancer of Zeste homolog 2 (EZH2) which have been shown experimentally to have probable roles in prostate carcinogenesis. Most recently, Tomlins et al. used novel bioinformatics approaches and gene expression methods to identify fusion of the androgen-regulated transmembrane protease, serine 2 (TMPRSS2) with members of the erythroblast transformation specific (ETS) DNA transcription factors family. This fusion appears commonly in prostate cancer and has been shown to be prevalent in more aggressive tumors. A number of studies have shown distinct classes of tumors separable by their gene expression which may relate to the known clinical heterogeneity. A number of gene expression studies have been performed looking for gene dysregulation in metastatic versus primary prostate cancer.

Another factor impacting clinical utility of the various proposed panels is the fact that most samples availability for validation exist only as formalin-fixed paraffin-embedded (FFPE) tissues. In contrast, many of the cDNA microarray studies conducted to date typically have used snap frozen tissues. The ability to perform and analyze gene expression in FFPE tissues will greatly accelerate research by correlating already available clinical information such as histological grade and clinical stage with gene specific signatures. In this study, we used the array-based DASL® (cDNA mediated annealing, selection, extension and ligation) assay and have compared the expression of 502 known cancer related genes in a cohort of men who have progressive disease exhibiting recurrence following prostatectomy.

Availability of reliable molecular markers in predicting the clinical course of disease is a critical need in the management of prostate cancer. It has become evident that multiplex assays that simultaneously detect multiple genetic abnormalities are needed to detect the modulation of genes involved in the development of malignancy. Early detection of these genetic changes may help to prevent cancer, or aid in prognosis and therapy. The goal of this study was to utilize highly multiplexed biomarker assays based on mRNA recovered from FFPE tissues to identify low complexity molecular signatures that may predict prostate cancer recurrence.

**Materials and Methods**

**Samples**

Patients were treated at University Hospital (Tucson, AZ) by radical prostatectomy following positive transrectal biopsy and negative bone scan. They presented with either abnormal digital rectal exam, or elevated serum PSA (>0.4 ng/ml) with normal DRE but subsequent positive sextant biopsy. All prostatectomy specimens were inked on the surface, fixed overnight in 10% neutral buffered formalin, and embedded in paraffin blocks. Four micron H&E stained sections were used to determine Gleason sum scores, tumor volume, location, and pathologic stage. Only men with a minimum 6-year follow-up were included in the study. Recurrence was defined as return of serum PSA greater than 0.3 ng/ml. Fourteen recurrent and fourteen non-recurrent patients were selected for gene expression studies (Table 1). All studies were approved by the University of Arizona IRB committee. Representative areas of tumor and adjacent normal (Figure 1) were selected by a pathologist using the H&E stained slides from each patient. Representative photographs of H&E stained sections of recurrent (R) and non-recurrent (NR) prostate cancer samples before taking punches used in this study were taken (data not shown). A Beecher punch was used to manually retrieve cores (1.0 mm diameter, 2-5 mm length) from FFPE blocks into RNase free eppendorf tubes for RNA isolation. The coring tool was dipped in xylene and flamed...
at a Bunsen burner between patient samples to prevent RNA carry over.

**RNA Isolation**

Total RNA was isolated from the FFPE cores using the High Pure RNA Paraffin kit (Roche, IN). Total RNA was quantified by UV spectroscopy using the NanoDrop-1000 (NanoDrop Technologies Inc., DE). Agilent BioAnalyzer™ traces were also conducted using the RNA Nano 6000 Series II Lab Chip (Agilent Inc.) to assess overall RNA quality.

**Quantitative Real-Time PCR (qRT-PCR)**

qRT-PCR was performed on an ABI 7500 (SDS v1.4; Applied Biosystems Inc., CA) to qualify samples as potentially useful for DASL® (ILLUMINA Corp., CA) gene expression analysis. This assay was conducted by measuring the expression of housekeeping gene RPL13a (GenBank accession # NM_012423.2) with detection by SYBR Green (SYBR® Green PCR master mix, ABI). Primers used were: forward, 5’ GTACGCTGTGAAGGCATCAA 3’, and reverse, 5’ GTTGGTGTTCATCCGCTTG 3’; amplicon size, 90 bp. Each reaction contained 25 µL of SYBR Green PCR Master Mix (ABI), 1 µL of cDNA template, and 250 nM each forward and reverse primer in a total reaction volume of 50 µL. All assays were done in triplicate in Micro Amp optical 96-well reaction plates (ABI) closed with Micro Amp optical adhesive covers (ABI). The PCR consisted of an initial enzyme activation step at 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec, 60°C for 1 minute. To access the final product a dissociation curve was generated using a ramp from 60 to 95°C (ABI). To validate the DASL assay data, qRT-PCR assay was performed on the test samples based on the manufacturer’s instructions with TaqMan™ gene expression assays (ABI) for the following genes: GAS1, TK1 and WNT5A (assay IDs: Hs00266715_s1, Hs00177406_ml and Hs00180103_ml). The assay that interrogated the sequence closest to the target sequence in the Illumina platform was chosen (Table 2). Relative quantification of the expression level of each transcript in each sample was calculated using the Delta-Delta CT method in the ABI 7500 system software. Normal prostate RNA was used as the calibrator and human β-actin (ACTB) gene was used as the endogenous control.

**cDNA Synthesis and DASL Expression Analysis**

Illumina's DASL is a gene assay designed to generate expression profiles from RNA such as that derived from FFPE tissue. The DASL assay was used with the standard DASL Human Cancer Panel (Table 3) from Illumina (www.illumina.com) and with the Universal-16 Bead Chip. The assay was performed according to standard Illumina protocols (Illumina Bead Station DASL System Manual). Briefly, the human cancer panel from Illumina comprises a pool of selected probe groups for 502 unique cancer gene mRNAs, each mRNA being targeted in three locations by three separate probes (Table 2). For each sample, input quantity for the reaction was normalized to 200 ng (40 ng/µl). This was converted into cDNA using biotinylated random nonamers, oligo-deoxynthymidine 18 primers and Illumina-supplied reagents according to manufacturer's instructions. The resulting biotinylated cDNA was annealed to assay oligonucleotides and bound to streptavidin-conjugated paramagnetic particles to select cDNA/oligo complexes. After oligo hybridization, mis-hybridized and non-hybridized oligos were washed away, while bound oligos were extended and ligated to generate templates to be subsequently amplified with shared PCR primers. The fluorescent-labeled complementary strand was hybridized as per standard protocols to Universal DASL 16x1 Bead Chip. Universal-16 Bead Chip platform is composed of 16 individual arrays and for each sample three technical replicates were performed. After hybridization, the arrays were scanned using the Illumina Bead Array Reader 500 system. Intensity data extractions and processing was performed with the Bead Studio Gene Expression Module (Illumina, GX version 3).

**Data Analyses**

Clinical parameters, including age, follow-up time, presenting PSA and Gleason score were evaluated with Student's t-tests to assess differences in the means between non-recurrent and recurrent subjects. Fisher's exact test was used to detect differences between proportions of T-score. Statistical significance was assessed at p<0.05. These were done using Stata 10 statistical software (StataCorp IC, College Station, TX).
DASL data analyses, including analysis for differential gene expression, clustering using rank invariant normalization and heat map generation, were all conducted with algorithms in Bead Studio (Illumina Corp.). The heat map used a log (base2) transformation and mean signal subtraction for each gene’s un-normalized signal data. In addition, the average signal intensities recorded for the individual prostate samples were analyzed with the nonparametric Mann-Whitney U-test, which tests whether two data sets come from different distributions. The U-test provided statistical significance levels which quantified the differences in expression levels between non-recurrent and recurrent sample sets for WNT5A, TK1, and GAS1.

Logistic regression analysis was used to develop models that predict the probability of recurrence for individual patients based on their expressed levels of WNT5A, TK1, and GAS1. A commonly-used statistic for evaluating the predictions of such models is the area under the receiver operating characteristic (ROC) curve constructed from the results. The AUC represents the probability that a randomly selected recurrent patient will have a higher logistic model score than a randomly selected non-recurrent patient. Two cross validation methods were used to estimate the AUC; leave one out cross validation (LOOCV) and 6-fold cross-validation. Both methods partition the samples into a training set (used to calibrate the logistic model parameters) and a test set, from which the AUC is determined. Due to the small number of samples, bootstrap re-sampling was used to improve the AUC estimates, using 100 randomly selected test cases. In the case of LOOCV, each sample was tested against the model trained on all of the other samples, and the results were combined to construct a single ROC curve.

The Spearman Rank Correlation coefficient was used for the evaluation of the correlation between the qPCR and Illumina data for each of the target genes (GAS1, TK1, and WNT5A). All P values <0.05 were considered statistically significant. We computed the Spearman rank correlation coefficient as a measure of correlation. The coefficient approaches -1 for strong negative correlation, 0 for no correlation, and +1 for strong positive correlation. We used the student’s t-distribution to determine whether the computed Spearman coefficient was significantly different than 0, our null hypothesis indicating that there is no correlation between the methods.

Results

Prostate Cancer Staging and Recurrence

The clinical parameters for the individuals were subjected to statistical analysis to determine whether there were significant differences between recurrent and non-recurrent sample groups (Table 1). Differences among continuous variables (age, follow-up time, presenting PSA and Gleason score) between non-recurrent and recurrent samples were not statistically significant (Table 4). However, the proportion of the subjects having stage T2 was statistically significantly higher in non-recurrent as compared to the recurrent subjects (Table 4). The proportion of subjects having stage T3 was statistically significantly lower in non-recurrent as compared to recurrent subjects (Table 4). These data support a relationship between prostate cancer aggressiveness and the likelihood of cancer recurrence.

Recovery of mRNA from FFPE

Prior to multiplexed analyses in the DASL assay, we conducted surrogate assays to assess the quality of RNA isolated from the archived FFPE sample blocks. Total RNA yields ranged from 1.6 µg to 13.6 µg. Quality assessments included qRT-PCR analysis and BioAnalyzer traces. qRT-PCR showed that RNA isolated from all of the 28 FFPE samples had CT values in the range of 19 to 25 (data not shown), and were acceptable for analysis by the DASL assay. Dissociation curve analysis also yielded a single peak indicating good quality RNA (data not shown). We did not observe the significant presence of smaller fragments that would have indicated degradation. In addition, all samples pre-qualified by qRT-PCR were of acceptable quality by BioAnalyzer assessment (data not shown). These measures of either the single control gene expression or overall RNAs, did not indicate unacceptable levels of degradation in any of the archived samples. Further, we
did not note any correlation between the age of the blocks and our ability to extract RNA for these analyses.

**DASL Gene Expression Analyses**

Of the 502 genes analyzed in the Cancer DASL assay pool (DAP), RNA message was detectable for 367 of these genes for all samples. A complete list of all genes is provided in the Table 3. The freshly isolated RNA samples from the breast cancer control showed an expression pattern that resulted in these samples clustering as a group that was distinct from the prostate group (Figure 2A). In addition, the breast cancer cell line, MCF 7 expressed a profile that distinguished this line from normal cells. These data confirm the expected relationships for breast and prostate cancer [1-2], as well as for the MCF-7 cell line [3] and normal specimens [4] and demonstrated the suitability of this assay for further analyses of the prostate cancer samples.

No clear molecular signature for prostate cancer recurrence was determined with unsupervised cluster analyses on all samples for all genes (Figure 2A). Instead, the determination of signatures was dependent on the stringency of detection obtained for specific samples (Figure 2B). Nine samples with a low negative control signals (<300), defined as low binding to irrelevant probes, and comparable to that obtained with the background signals from the freshly isolated RNA samples, were easily sorted into recurrent or non-recurrent groups (Figure 2C).

Negative control oligonucleotides targeted 27 random sequences that do not appear in the human genome (Illumina Product Guide 2006/7). The mean signal of these probes defines the system background. The standard deviation of signal on these probes defines the noise. This is a comprehensive measurement of background, representing the imaging system background as well as any signal resulting from non-specific binding of dye or cross-hybridization. The Bead Studio™ application uses the signals and signal standard deviation of these probes to establish gene expression detection limits. Using this criteria to select samples for analyses resulted in a smaller signature of 33 genes (Table 5) identified as significantly (detection p value \( \leq 0.001 \)) differentially expressed between the two groups of prostate cancer and could be used to classify those that recurrent or not. A heat map showing the relative expression of these genes is shown in Figure 3A.

**Low Complexity Gene Signature for Prostate Cancer Recurrence**

Marked over-expression of the proto-oncogene, WNT5A [5], and TK1, was noted in samples in which prostate cancer later recurred (Figure 3A). The expression of GAS1 was most associated with non-recurrence of prostate cancer (Figure 3A). These three genes showed clear differential expression between the recurrent and non-recurrent patient groups in a subset of nine samples (Figure 3B). WNT5A and TK1 expression was increased in the recurrent compared to the non-recurrent cases (Figure 3B). In contrast, GAS1 expression was noticeably increased in the non-recurrent as compared to the recurrent cases (Figure 3B). All 28 samples for recurrent and non-recurrent patient groups were subjected to qRT-PCR analysis using ABI TaqMan™ assay to validate the data obtained on DASL assay for the three target genes. The qRT-PCR assay confirmed the expression data for all the 3 genes obtained on DASL assay. Further, to assess the correlation between the qPCR and Illumina data for all 28 samples and for the target genes GAS1, TK1, and WNT5, we computed the Spearman Rank Correlation coefficient as a measure of correlation. We found significant correlation between the data from the two methods (Table 6).

To assess the significance of the differential expression noted for these genes we assessed the larger sample set. One outlier sample, from the non-recurrence group, (patient # 61) showed high background signal and was also unresponsive across all genes, that is, neither increases nor decreases were noted for any of the genes assessed. The Mann-Whitney U-test, which measures the confidence that two data sets come from separate distributions, indicated that the recurrent and non-recurrent samples for WNT5A and GAS1 showed differences that were statistically significant at the level of p<0.05. The differential expression between non-recurring and
differential expression between non-recurring and recurring for TK1 was significant at p<0.01 (Table 7). The very high significance for TK1 can be understood by striking correlation between the expression of TK1 and recurrence. Thus, for this sample set, the distribution of expression levels for non-recurrent and recurrent samples was different for each of the three genes.

Although the previous tests demonstrated separate recurring and non-recurring distributions for WNT5A, GAS1, and TK1, these distributions do overlap and their ability to reliably predict recurrence is a separate question, which was assessed using logistic regression modeling. A logistic regression model was fit to the entire set of 27 samples, and an ROC curve was constructed to evaluate how well the model fit the data. An area under the ROC curve of 0.846 was achieved for the three gene panel, as shown in Figure 4. This compares favorably with an AUC of 0.758 for the gene panel (SPINK1, PCA3, GOLPH2, and TMPRSS2: ERG) recently identified by Laxman et al. and 0.508 for the PSA serum test.

Finally, the ability of the model to predict recurrence for samples not included in the model training set was assessed. Due to the limited number of samples, both bootstrapping and LOOCV were employed. An AUC of 0.734 was found using a bootstrapping approach, and an AUC of 0.690 was found using the LOOCV technique. For comparison, Laxman et al. calculated an AUC of 0.736 for their panel of genes using the leave one out method.

Discussion

This study was carried out utilizing highly multiplexed biomarker assays based on mRNA recovered from widely available archival FFPE tissues. We specifically focused on the identification of low complexity molecular signatures that have the potential to predict prostate cancer recurrence and that may be utilized in routine clinical pathology practice. We selected and compared a cohort of prostatectomies from men with indolent disease to cohort of men who have progressive disease exhibiting recurrence following prostatectomy. There were no significant differences between the two cohorts among various clinical parameters compared except in the tumor stages and cancer recurrence. A higher number of patients having stage T3 were in the recurrent group than in the non-recurrent group (Table 4). However, this may not be a strong predictive factor of cancer recurrence since there were a number of patients with high tumor stage in the non-recurrent group as well, and two of the non-recurrent cases had obturator lymph node metastasis (T4a) at the time of original surgery (Table 1). This is consistent with previous reports where selected genes were better predictors of recurrence independent of tumor grade or stage. These findings emphasize the continuing need to not only identify molecular signatures but to correlate these with clinical parameters to help develop cancer management or treatment options that may be easily integrated into current clinical practice.

The RNA isolated from FFPE samples were of good quality as judged by BioAnalyzer traces and the SYBR Green assays. These qualifications are in agreement with studies previously conducted with RNA isolated from FFPE samples. In our study, we identified a 33 gene signature that may be used to classify recurrent and non-recurrent cancer specimens (Table 5). A subset of 9 samples was sorted into recurrent and non-recurrent groups in this manner (Figure 2C). DASL assay on all the 28 samples was validated by qRT-PCR. The correlation between the two were significant as shown by Spearman Rank Correlation Coefficient analysis (Table 6). Of particular interest was the finding that prostate cancer recurred in samples with marked over-expression of genes implicated in cell proliferation/differentiation, and in DNA synthesis; WNT5A and TK1, respectively. Samples for which prostate cancer had not recurred post at least five years of follow-up had significant over-expression of the putative tumor suppressor, GAS1. In addition, our analysis of the ability of these three genes to predict the likelihood of the prostate cancer recurrence in the larger patient cohort (27 cases) showed (Figure 5) that this panel performs comparably to other recently identified panels, and significantly better than the PSA serum test.
The role of WNT5A in prostate cancer is controversial. Our results for the WNT5A gene are consistent with previously published data including that reported by Fan et al., where a sixteen gene expression signature correlated with Gleason score and relapse in prostate cancer. WNT5A was among the up-regulated genes in recurrent tumors in this panel. In another study, Glinsky et al. reported 218 gene signatures that may distinguish recurrent and non-recurrent prostate cancers. WNT5A was again identified as over-expressed gene in recurrent tumors. Yamamoto et al. have shown that WNT5A abnormal expression was correlated with high Gleason scores and biochemical relapse of prostate cancer which is consistent with our data wherein we show association of overexpression with recurrent prostate cancer. Contrary reports on the role of WNT5A overexpression in prostate cancer is evident as shown by Khaja et al. where they have shown that elevated level of WNT5A protein is associated with better outcome. Our data adds further clarity and is in agreement with others as mentioned above. Notably, there are 19 WNT genes so far identified in humans. The WNTs are lipid-modified secreted glycoproteins that regulate diverse biologic functions including roles in developmental patterning, cell proliferation, differentiation, cell polarity, and morphogenetic movement. WNT5A protein has been shown to influence transcription by effecting histone methylation, increasing cell migration, induce endothelial proliferation, and increase expression of certain metalloproteinases. WNT5A has been demonstrated in a number of human neoplasms including melanoma, breast cancer, and colon cancer. Wang et al. have recently shown that WNT5A gene expression was increased greater than 50-fold in prostate carcinoma cell lines as compared to normal. Recently, it has been reported that prostate cancer cells attain the capacity to proliferate under androgen-depleted condition by activating the WNT5A/BMP-6 loop that involves bone stromal cells. There by leading to resistance to androgen deprivation via protein kinase C (PKC)/NFkB (p52) signalling. The data presented in this study suggests a role for WNT5A in prostate tumor progression.

Our finding of a significant association of TK1 in prostate cancer recurrence is also consistent with published data. However, in this case, the data is limited to a study of breast tumors where it was reported that total TK levels in breast tumors were significantly higher in patients who subsequently had recurrence compared with levels in those who did not. Recently, TK1 has been reported as universal malignancy marker as it is over expressed in a variety of carcinoma tissues including prostate cancer. Authors also show that in non-malignant tissues from cancer patients, weak expression is seen indicating that these may be precancerous lesions. In true normal cases (from individuals without cancer), TK1 is not expressed. Also, they suggest that it would be interesting to see whether those patients with TK1 expression in precancerous lesions, showed worse prognosis or later relapse of cancer. Our data supports this thought demonstrating that TK1 overexpression is associated with recurrence of Prostate cancer. The human thymidine kinase gene is located on chromosome 17 at gene locus 17q25.2-q25.3. Thymidine kinase catalyzes the phosphorylation of thymidine to deoxythymidine monophosphate, and is expressed in the G1/S boundary of the cell cycle and is an essential gene for DNA replication. It is also reported that serum TK1 protein assays play an important role in the detection of early stages of breast and prostate cancer.

In contrast to the up-regulation of WNT5A and TK1 in the cases for which recurrence was later observed, we noted that the expression of GAS1 was most associated with non-recurrence of prostate cancer (Figure 3B). The current study is the first to document the over-expression of GAS1 in indolent prostate carcinomas. GAS1’s known function on growth repression however indicates that it may play a role in retarding cell proliferation and could contribute to the slow progression seen in the majority of prostate carcinomas. GAS1 is located on chromosome 9 and encodes a 45 kDa glycoprophathydinositol (GPI)-linked protein. It was subsequently shown to arrest the cell cycle when over expressed. Lee et al. and others have suggested that GAS1 acts by attenuation of sonic hedgehog
signaling. It is increased in neurons and neuroblastoma cells that are committed to die by apoptosis. GAS1 blocks entry to S phase and prevents cycling of normal and transformed cells. Using the rat castration model, GAS1 has been shown to be up-regulated in secretory epithelium of the ventral prostate undergoing apoptosis. In the context of this study, the increased expression of GAS1 in the non-recurrent cases could result in suppression of proliferation or increased apoptosis.

In summary, we have identified a gene signature, characterized by over-expression of \textit{WNT5A} and TK1, and down-regulation of GAS1, that may predict the likelihood of prostate cancer recurrence. This novel three gene signature was capable of distinguishing recurrent and non-recurrent prostate cancers in surgical specimens removed at least five years prior to follow-up. This small gene panel may be amenable to the development of RT/PCR or IHC tests once clinical utility in prognosis and therapy is established. Further studies will test this hypothesis on larger cohorts of post-prostatectomy patients and also will include studies on expression of \textit{WNT5A} receptors like Frizzled 5 through which it acts (the SO2 genes DASL panel we tested did not have Frizzled 5).

\textbf{Acknowledgments}

We would like to acknowledge Dr. Anne Cress of the University of Arizona for critical review of the manuscript. Tim Gilmartin and Chad Geringer of Illumina Corp. and James Averil of the University of Arizona provided technical assistance with the DASL assay and prostate sample recovery, respectively. We would like to thank TACMASR (Tissue Acquisition and Cellular/Molecular Analysis Shared Resource), at University of Arizona for providing tissue samples. This work was supported by grants from National Institute of Health, NIH (PO1 CA 56666) and Arizona Biomedical Research Commission, ABRC (ABRC-0713) to Raymond B. Nagle; also by Arizona Cancer Center (AZCC) core grant CA23074.

\textbf{Table 1} : Clinical and pathological data of non-recurrent and recurrent prostate cancers

\begin{tabular}{|c|c|c|c|c|c|c|}
\hline

A. Non-recurrent & & & & & & \\
Patient # & Age & Presenting & Last & Follow-up & T-score & Gleason Score \\
& & PSA & PSA & Time (yrs.) & & \\
4 & 81 & 6.0 & <0.4 & 10.7 & T2c & 3+3=6/10 \\
17 & 57 & 4.4 & <0.4 & 7.11 & T2c & 3+3=6/10 \\
22 & 66 & 22.0 & <0.4 & 13.10 & T3a & 4+5=9/10 \\
23 & 77 & 7.0 & <0.04 & 13.4 & T3a & 4+5=9/10 \\
56 & 80 & 3.5 & <0.04 & 12.0 & T2c & 3+3=6/10 \\
57 & 85 & 8.0 & <0.4 & 12.2 & T3c & 3+3=6/10 \\
58 & 76 & 7.0 & <0.4 & 10.2 & T2c & 3+3=6/10 \\
59 & 80 & 14.0 & <0.04 & 10.9 & T4a & 3+3=6/10 \\
60 & 77 & 12.8 & <0.4 & 10.8 & T4a & 3+3=6/10 \\
61 & 76 & 5.6 & <0.04 & 9.2 & T2c & 3+3=6/10 \\
62 & 78 & 11.3 & <0.4 & 8.3 & T3a & 4+3=7/10 \\
63 & 64 & 23.0 & <0.04 & 7.8 & T2c & 4+4=8/10 \\
64 & 64 & 6.1 & <0.04 & 7.0 & T2c & 3+3=6/10 \\
65 & 72 & 8.2 & <0.04 & 7.8 & T3b & 3+3=6/10 \\
\hline

B. Recurrent & & & & & & \\
Patient # & Age & Presenting & Last & Lag time from & Follow-up & Gleason \\
& & PSA & PSA & surgery to & Time (yrs.) & Score \\
& & & & recurrence (yrs.) & & \\
28 & 79 & 10.1 & 0.23 & 5.4 & 8.4 & T3a & 4+3=7/10 \\
29 & 74 & 7.4 & 1.9 & 9.5 & 13.3 & T3a & 4+3=7/10 \\
30 & 61 & 48.9 & 1 & 0.5 & 8.2 & T4a & 4+3=7/10 \\
31 & 87 & 7.4 & 361.5 & 7.9 & 13.5 & T3a & 4+3=7/10 \\
34 & 87 & 5.8 & 826.14 & 5.0 & 7.5 & T3a & 3+3=6/10 \\
\hline
\end{tabular}
<table>
<thead>
<tr>
<th>Patient #</th>
<th>Age</th>
<th>Presenting PSA</th>
<th>Last PSA</th>
<th>Lag time from surgery to recurrence (yrs.)</th>
<th>Follow-up Time (yrs.)</th>
<th>T-score</th>
<th>Gleason Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>79</td>
<td>3.6</td>
<td>28.53</td>
<td>6.10</td>
<td>8.6</td>
<td>T3c</td>
<td>5+4=9/10</td>
</tr>
<tr>
<td>38</td>
<td>77</td>
<td>4.9</td>
<td>2.89</td>
<td>11.9</td>
<td>13.7</td>
<td>T3a</td>
<td>4+3=7/10</td>
</tr>
<tr>
<td>39</td>
<td>87</td>
<td>12.5</td>
<td>2.41</td>
<td>12.3</td>
<td>13.11</td>
<td>T3c</td>
<td>5+5=10/10</td>
</tr>
<tr>
<td>44</td>
<td>77</td>
<td>154</td>
<td>3893</td>
<td>6.2</td>
<td>8.0</td>
<td>T4a</td>
<td>4+4=8/10</td>
</tr>
<tr>
<td>46</td>
<td>72</td>
<td>14.5</td>
<td>1.8</td>
<td>2.3</td>
<td>7.3</td>
<td>T3c</td>
<td>5+4=8/10</td>
</tr>
<tr>
<td>50</td>
<td>73</td>
<td>13.4</td>
<td>2.68</td>
<td>6.4</td>
<td>8.2</td>
<td>T3c</td>
<td>4+3=7/10</td>
</tr>
<tr>
<td>51</td>
<td>84</td>
<td>3.9</td>
<td>14</td>
<td>0.11</td>
<td>8.9</td>
<td>T4a</td>
<td>5+5=10/10</td>
</tr>
<tr>
<td>52</td>
<td>71</td>
<td>4.5</td>
<td>21.6</td>
<td>0.7</td>
<td>7.4</td>
<td>T3c</td>
<td>3+3=6/10</td>
</tr>
</tbody>
</table>

*Patient had an elevated PSA 0.4 in January of 2005, 6 yrs after the surgery.*

**Table 2:** Illumina and ABI probe details for the 3 genes used in this study

<table>
<thead>
<tr>
<th>Illumina Accession no.</th>
<th>Gene symbol</th>
<th>Illumina start</th>
<th>Illumina probe sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>2051</td>
<td>GCCGATTGCCTTAGAGGGAACCGCTTAAATGTTTTGGATAAGTT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1534</td>
<td>TGGGACAGATAGAAGGGATGGTTGGGGATACTTCCCAAAACTTTT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1370</td>
<td>GTGGAGAGGGCGAGTGTCAGTGCAGTATGAAAT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1273</td>
<td>CTGGTTGATGTTTCACAGGAACACAGCATCTTCCCATAGAGAT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>161</td>
<td>AGGTTGATGAGAGGTCCGTCGCTCCAGAGTCTTGGATAAG</td>
</tr>
<tr>
<td>NM_003392.2</td>
<td>WNT5A</td>
<td>2948</td>
<td>CACCTGGTCCCCTTGTGTTGAGGAAGATATACAGAATAG</td>
</tr>
<tr>
<td>NM_003392.2</td>
<td>WNT5A</td>
<td>804</td>
<td>CCAATTTTCTCATCCTCCGAGTGGTGAATGAG</td>
</tr>
<tr>
<td>NM_003392.2</td>
<td>WNT5A</td>
<td>597</td>
<td>GGAGGAGAAGGCGAGTACATACAGAATACCTAAAGACCCC</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1534</td>
<td>TGGGACAGATAGAAGGGATGGTTGGGGATACTTCCCAAAACTTTT</td>
</tr>
</tbody>
</table>

**Note:** Illumina probes were used on DASL platform for expression analysis and ABI probes were used for qRT-PCR on the same set of samples. The ABI Assay ID for GAS1, Hs00266715_s1; for TK1, Hs00177406_ml; and for WNT5A, Hs00180103_ml.

**Table 3 DASL™ CANCER PANEL**

<table>
<thead>
<tr>
<th>ABCB1</th>
<th>BRC2</th>
<th>CHEK1</th>
<th>E2F2</th>
<th>FGF2</th>
<th>HIF1A</th>
<th>KIT</th>
<th>MMP1</th>
<th>NUMA1</th>
<th>RAD23A</th>
<th>SPP1</th>
<th>TNFSF6</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCB2</td>
<td>BTX</td>
<td>COL11A1</td>
<td>E2F3</td>
<td>FGF3</td>
<td>HLF</td>
<td>KRAS2</td>
<td>MMP10</td>
<td>OGG1</td>
<td>RAD50</td>
<td>SRC</td>
<td>TNFSF8</td>
</tr>
<tr>
<td>ABCG2</td>
<td>CASP10</td>
<td>COL1A1</td>
<td>E2F5</td>
<td>FGF5</td>
<td>HMMR</td>
<td>L1CAM</td>
<td>MMP14</td>
<td>OSM</td>
<td>RAD51</td>
<td>STAT1</td>
<td>TOP1</td>
</tr>
<tr>
<td>ABL1</td>
<td>CASP2</td>
<td>COL4A3</td>
<td>EGF</td>
<td>FGFR</td>
<td>HOXA9</td>
<td>LAF4</td>
<td>MMP2</td>
<td>PBX1</td>
<td>RAD52</td>
<td>STAT3</td>
<td>TOP2A</td>
</tr>
<tr>
<td>ADPR1</td>
<td>CASP3</td>
<td>COMT</td>
<td>EGFR</td>
<td>FGFR3</td>
<td>HRAS</td>
<td>LAMB1</td>
<td>MMP3</td>
<td>PCNA</td>
<td>RAD54A</td>
<td>STAT5B</td>
<td>TP5</td>
</tr>
<tr>
<td>AHR</td>
<td>CASPS</td>
<td>COPEB</td>
<td>EGR1</td>
<td>FGFR4</td>
<td>ICAM1</td>
<td>LCK</td>
<td>MMP7</td>
<td>PCTK1</td>
<td>RAD54L</td>
<td>STK11</td>
<td>TP7</td>
</tr>
<tr>
<td>AIM2</td>
<td>CAV1</td>
<td>CREBBP</td>
<td>ELK1</td>
<td>FGFR5</td>
<td>IFNG</td>
<td>LCN2</td>
<td>MMP9</td>
<td>PDGA</td>
<td>RAF1</td>
<td>SYK</td>
<td>TPR</td>
</tr>
<tr>
<td>AKT1</td>
<td>CBFA2T1</td>
<td>CRK</td>
<td>ELK3</td>
<td>FGFR1</td>
<td>IFNGR1</td>
<td>LIF</td>
<td>MOS</td>
<td>PFGRA</td>
<td>RALB</td>
<td>TAL1</td>
<td>TRAF3</td>
</tr>
<tr>
<td>AKT2</td>
<td>CRKL</td>
<td>E2F2</td>
<td>FGFR2</td>
<td>IFNGR2</td>
<td>LG1</td>
<td>MIP</td>
<td>PLDGA</td>
<td>RAN</td>
<td>TCF7L2</td>
<td>TRAF4</td>
<td></td>
</tr>
<tr>
<td>ALK</td>
<td>CBLB</td>
<td>CSFR1</td>
<td>FGFR3</td>
<td>IFG1</td>
<td>LG3</td>
<td>MRE11A</td>
<td>PDGFRA</td>
<td>RAPA1</td>
<td>TDFG1</td>
<td>TSC1</td>
<td></td>
</tr>
<tr>
<td>ALOX12</td>
<td>CCNA2</td>
<td>CSF2</td>
<td>ENC1</td>
<td>FGFR4</td>
<td>IFG1R</td>
<td>LG4</td>
<td>MSF</td>
<td>PAPI1GDS</td>
<td>TEK</td>
<td>TSC2</td>
<td></td>
</tr>
<tr>
<td>APAF1</td>
<td>CCNC</td>
<td>CSF3</td>
<td>EPHA1</td>
<td>FG2</td>
<td>LMO1</td>
<td>MSH2</td>
<td>PGR</td>
<td>RAP2A</td>
<td>TERT</td>
<td>TSG101</td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>CCND1</td>
<td>CSFR3</td>
<td>EPHB4</td>
<td>FHIT</td>
<td>IFG2R</td>
<td>LMO2</td>
<td>MSH3</td>
<td>PIK3CA</td>
<td>RARA</td>
<td>TFAP2C</td>
<td>TMY5</td>
</tr>
<tr>
<td>AR</td>
<td>CCND2</td>
<td>CSK</td>
<td>EPO</td>
<td>FLU1</td>
<td>IGFBP1</td>
<td>LTA</td>
<td>MSH6</td>
<td>PIM1</td>
<td>RAR8</td>
<td>TFDP1</td>
<td>TYR3</td>
</tr>
<tr>
<td>ARAF1</td>
<td>CCND3</td>
<td>CSPG2</td>
<td>EP1S5</td>
<td>FLT1</td>
<td>IGFBP2</td>
<td>LYN</td>
<td>MST1R</td>
<td>PLA2G2A</td>
<td>RASA1</td>
<td>TFE3</td>
<td>VAV1</td>
</tr>
<tr>
<td>AREG</td>
<td>CCNE1</td>
<td>CTGF</td>
<td>EP5S</td>
<td>FLT3</td>
<td>IGFBP3</td>
<td>MAD</td>
<td>MTA</td>
<td>PLAG1</td>
<td>RB1</td>
<td>TFF1</td>
<td>VAV2</td>
</tr>
<tr>
<td>ARHA</td>
<td>CENH</td>
<td>CTNNA1</td>
<td>ERBB2</td>
<td>FLT4</td>
<td>IGFBP5</td>
<td>MAD2</td>
<td>MTHFR</td>
<td>PLAT</td>
<td>RBBP1</td>
<td>TFBP</td>
<td>VGB</td>
</tr>
<tr>
<td>ARHD1</td>
<td>CD34</td>
<td>CTNNB1</td>
<td>ERBB3</td>
<td>FOLR1</td>
<td>IGFBP6</td>
<td>MAD4</td>
<td>MUC1</td>
<td>PLaur</td>
<td>RBBP2</td>
<td>TFRC</td>
<td>VEGF</td>
</tr>
<tr>
<td>ARHH</td>
<td>CD44</td>
<td>CTSD</td>
<td>ERBB4</td>
<td>FOS</td>
<td>IL1</td>
<td>MAF</td>
<td>MX1</td>
<td>PLG</td>
<td>RBBP5</td>
<td>TGFA</td>
<td>VEGFB</td>
</tr>
<tr>
<td>ARHI</td>
<td>CD59</td>
<td>CTSL</td>
<td>ERCC1</td>
<td>FOSB</td>
<td>IL12A</td>
<td>MALT1</td>
<td>MYB</td>
<td>PML</td>
<td>RBBP6</td>
<td>TGFB1</td>
<td>VHL</td>
</tr>
<tr>
<td>ARNT</td>
<td>CD9</td>
<td>CUL2</td>
<td>ERCC2</td>
<td>FOSL2</td>
<td>IL12B</td>
<td>MAP3K8</td>
<td>MYBL2</td>
<td>PMS1</td>
<td>RBL2</td>
<td>TGFB2</td>
<td>VIL2</td>
</tr>
<tr>
<td>ATF1</td>
<td>CDC2</td>
<td>CXCL9</td>
<td>ERCC3</td>
<td>FRAP1</td>
<td>IL13</td>
<td>MAPK10</td>
<td>MYC</td>
<td>PNU1L1</td>
<td>RECQL</td>
<td>TGF8B</td>
<td>WEE1</td>
</tr>
</tbody>
</table>

**Table 2:** Illumina and ABI probe details for the 3 genes used in this study

<table>
<thead>
<tr>
<th>Illumina Accession no.</th>
<th>Gene symbol</th>
<th>Illumina start</th>
<th>Illumina probe sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>2051</td>
<td>GCCGATTGCCTTAGAGGGAACCGCTTAAATGTTTTGGATAAGTT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1534</td>
<td>TGGGACAGATAGAAGGGATGGTTGGGGATACTTCCCAAAACTTTT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1370</td>
<td>GTGGAGAGGGCGAGTGTCAGTGCAGTATGAAAT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1273</td>
<td>CTGGTTGATGTTTCACAGGAACACAGCATCTTCCCATAGAGAT</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>161</td>
<td>AGGTTGATGAGAGGTCCGTCGCTCCAGAGTCTTGGATAAG</td>
</tr>
<tr>
<td>NM_003392.2</td>
<td>WNT5A</td>
<td>2948</td>
<td>CACCTGGTCCCCTTGTGTTGAGGAAGATATACAGAATAG</td>
</tr>
<tr>
<td>NM_003392.2</td>
<td>WNT5A</td>
<td>804</td>
<td>CCAATTTTCTCATCCTCCGAGTGGTGAATGAG</td>
</tr>
<tr>
<td>NM_003392.2</td>
<td>WNT5A</td>
<td>597</td>
<td>GGAGGAGAAGGCGAGTACATACAGAATACCTAAAGACCCC</td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>1534</td>
<td>TGGGACAGATAGAAGGGATGGTTGGGGATACTTCCCAAAACTTTT</td>
</tr>
</tbody>
</table>

**Note:** Illumina probes were used on DASL platform for expression analysis and ABI probes were used for qRT-PCR on the same set of samples. The ABI Assay ID for GAS1, Hs00266715_s1; for TK1, Hs00177406_ml; and for WNT5A, Hs00180103_ml.
### Table 1: 502 genes selected from 10 publicly available gene lists.

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
<th>Gene 3</th>
<th>Gene 4</th>
<th>Gene 5</th>
<th>Gene 6</th>
<th>Gene 7</th>
<th>Gene 8</th>
<th>Gene 9</th>
<th>Gene 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATM</td>
<td>CDC25A</td>
<td>CYP1A1</td>
<td>ERC C4</td>
<td>FRZB</td>
<td>IL1A</td>
<td>MAPK14</td>
<td>MYCL1</td>
<td>PPARD</td>
<td>REL</td>
</tr>
<tr>
<td>AXL</td>
<td>CDC25B</td>
<td>CYP1B1</td>
<td>ERC C5</td>
<td>FVT1</td>
<td>IL1B</td>
<td>MAS1</td>
<td>MYCL2</td>
<td>PPARG</td>
<td>RELA</td>
</tr>
<tr>
<td>BAD</td>
<td>CDC25C</td>
<td>DAB2</td>
<td>ERC C6</td>
<td>FYN</td>
<td>IL1RN</td>
<td>MATK</td>
<td>MYCN</td>
<td>PPP2R1B</td>
<td>RET</td>
</tr>
<tr>
<td>BAG1</td>
<td>CDH1</td>
<td>DAP3</td>
<td>ERG</td>
<td>FZD7</td>
<td>I12</td>
<td>MBD2</td>
<td>NAT2</td>
<td>PRCC</td>
<td>RIPK1</td>
</tr>
<tr>
<td>BAK1</td>
<td>CDH11</td>
<td>DPK1</td>
<td>ESR1</td>
<td>G2P1</td>
<td>I13</td>
<td>MCAM</td>
<td>NBS1</td>
<td>PRKARI A1</td>
<td>RLF</td>
</tr>
<tr>
<td>BARD1</td>
<td>CGD13</td>
<td>DCC</td>
<td>ETS1</td>
<td>GADD45A</td>
<td>I14</td>
<td>MCC</td>
<td>NEO1</td>
<td>PRKR</td>
<td>ROS1</td>
</tr>
<tr>
<td>BCL2</td>
<td>CDK10</td>
<td>DCN</td>
<td>ETS2</td>
<td>GAS1</td>
<td>I16</td>
<td>MCF2</td>
<td>NF1</td>
<td>PTCCH</td>
<td>RRAS</td>
</tr>
<tr>
<td>BCL2A1</td>
<td>CDK2</td>
<td>DDB2</td>
<td>ETV1</td>
<td>GAS1</td>
<td>I18</td>
<td>MCL1</td>
<td>NFKB1</td>
<td>PTCH2</td>
<td>RRF</td>
</tr>
<tr>
<td>BCL2L1</td>
<td>CDK4</td>
<td>DDI7T3</td>
<td>ET6</td>
<td>GFI1</td>
<td>I1K</td>
<td>MDM4</td>
<td>NFKB2</td>
<td>PTEN</td>
<td>SEMA3F</td>
</tr>
<tr>
<td>BCL3</td>
<td>CDK6</td>
<td>DDX6</td>
<td>EV11</td>
<td>GLI</td>
<td>IING1</td>
<td>MDS1</td>
<td>NFKBIA</td>
<td>PTEN</td>
<td>SEMA3F</td>
</tr>
<tr>
<td>BCL6</td>
<td>CDK7</td>
<td>DEK</td>
<td>EV12A</td>
<td>GLI2</td>
<td>INHA</td>
<td>MEL</td>
<td>NGFR</td>
<td>PTGS2</td>
<td>SH3BP2</td>
</tr>
<tr>
<td>BCR</td>
<td>CDK9</td>
<td>DCKT</td>
<td>EXT1</td>
<td>GL13</td>
<td>IRF1</td>
<td>MEN1</td>
<td>NOS3</td>
<td>PTHLH</td>
<td>SHH</td>
</tr>
<tr>
<td>BIRC2</td>
<td>CDKN1A</td>
<td>DLC1</td>
<td>EXT2</td>
<td>GML</td>
<td>ITOB1</td>
<td>MET</td>
<td>NOTCH1</td>
<td>PTK2</td>
<td>SIAH1</td>
</tr>
<tr>
<td>BIRC3</td>
<td>CDKN1B</td>
<td>DLEU1</td>
<td>FANCA</td>
<td>GRB2</td>
<td>ITGB4</td>
<td>MLF1</td>
<td>NOTCH2</td>
<td>PTK7</td>
<td>SKI</td>
</tr>
<tr>
<td>BIRC5</td>
<td>CDKN2A</td>
<td>DG3</td>
<td>FANCG</td>
<td>GRB7</td>
<td>JAK2</td>
<td>MLF2</td>
<td>NOTCH4</td>
<td>PTPRF</td>
<td>SKIL</td>
</tr>
<tr>
<td>BLM</td>
<td>CDKN2B</td>
<td>DMBT1</td>
<td>FAT</td>
<td>GRPR</td>
<td>JUN</td>
<td>MLH1</td>
<td>NQO1</td>
<td>PTPRG</td>
<td>SMARC5</td>
</tr>
<tr>
<td>BM1</td>
<td>CDKN2C</td>
<td>DSP</td>
<td>FER</td>
<td>GSTP1</td>
<td>JUNB</td>
<td>MILL</td>
<td>NRAS</td>
<td>PTPRH</td>
<td>SMARC5</td>
</tr>
<tr>
<td>BMP4</td>
<td>CDKN2D</td>
<td>DTR</td>
<td>FES</td>
<td>HCK</td>
<td>JUND</td>
<td>MLLT3</td>
<td>NTRK1</td>
<td>PURA</td>
<td>SOD1</td>
</tr>
<tr>
<td>BRAF</td>
<td>CEACAM1</td>
<td>DVL3</td>
<td>FGF1</td>
<td>HDAC1</td>
<td>KAI1</td>
<td>MLLT4</td>
<td>NTRK2</td>
<td>PXN</td>
<td>SPARC</td>
</tr>
<tr>
<td>BRCA1</td>
<td>CEBPA</td>
<td>E2F1</td>
<td>FGF12</td>
<td>HDGF</td>
<td>KAD</td>
<td>MLLT6</td>
<td>NTRK3</td>
<td>QARS</td>
<td>SPI1</td>
</tr>
</tbody>
</table>

### Figure 1: Representative Areas of tumor and adjacent normal tissues as seen on H&E stained slides. (A) Prostate sample showing normal (N) and cancer regions (Ca). (B) and (C), paraffin block and H&E stained slide respectively showing the punches made from the cancer and adjacent tissue.

### Figure 2: Expression analyses. (A) Cluster analysis using rank invariant normalization for all evaluable genes (367) and all samples (24 prostate tests and 4 control breast specimens namely CTRL1-MCF7, CTRL2-Breast/MCF7, CTRL3-Breast 1 and CTRL4-Breast 2). The control breast cancer samples (freshly isolated RNA) clustered separately from the prostate cancer samples. Correlation (1- r) values are displayed on the axis. (B) Negative control sample plots show a significant number of RNA samples with signal >300, indicative of high test sample binding to irrelevant probe. Red dotted line indicates signal at 300 and below the red line are the samples with low background signal. Red circles and green circles indicate 9 samples and 4 control samples with low background signals respectively. (C) Cluster analysis only for samples with low background binding (p value for detection <0.05)
Figure 3. Heat map and expression analyses for subset of samples. A, supervised cluster heat map showing a smaller signature of 33 genes which are significantly differentially expressed between the two groups of prostate cancer. Scale showing average signal intensities is on right. B, the 3 gene (WNT5A, TK1 and GAS1) differential expression between recurrent (n = 4) and non-recurrent (n = 5) groups for 9 samples. The average signal intensity between recurrent and non-recurrent groups for WNT5A: 2861.29 and 338.35; for TK1: 2156.17 and 752.25; and for GAS1 130.52 and 2387.13.

Figure 4. ROC curve showing the performance of a logistic regression model that includes WNT5A, GAS1, and TK1 and was fit to the entire set of 27 samples. The area under the curve is 0.846, which indicates the model fits the data very well. Bootstrap resampling was used to improve the AUC estimates, using 100 randomly selected test cases. Vertical axis (Y-axis) indicates true positive rate (sensitivity) i.e., scoring of recurrent samples as recurrent; horizontal axis (X-axis) indicates false positive rate (1-specificity) i.e., scoring of non-recurrent samples as recurrent.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Non-recurrent (N = 15)</th>
<th>Recurrent (N = 15)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age, yrs (SD)</td>
<td>73.8 (8.0)</td>
<td>75.8 (7.0)</td>
<td>0.48</td>
</tr>
<tr>
<td>Mean follow up time, months (SD)</td>
<td>121.0 (26.6)</td>
<td>114.3 (32.2)</td>
<td>0.55</td>
</tr>
<tr>
<td>Mean presenting PSA, ng/ml (SD)</td>
<td>9.9 (6.1)</td>
<td>20.4 (38.6)</td>
<td>0.33</td>
</tr>
<tr>
<td>Mean Gleason score (SD)</td>
<td>6.7 (1.2)</td>
<td>7.6 (1.2)</td>
<td>0.10</td>
</tr>
<tr>
<td>T-score, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>7 (50.4)</td>
<td>0 (0)</td>
<td>0.002*</td>
</tr>
<tr>
<td>T3</td>
<td>5 (35.7)</td>
<td>12 (80)</td>
<td>0.02*</td>
</tr>
<tr>
<td>T4</td>
<td>2 (14.3)</td>
<td>3 (20)</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Table 4: Comparison of different clinico-pathological parameters between non-recurrent and recurrent prostate cancer samples

*statistically significant at p<0.05
Table 5: The 33 gene signature between non-recurrent and recurrent human prostate cancers

<table>
<thead>
<tr>
<th>ACCESSION</th>
<th>SYMBOL</th>
<th>Full Name</th>
<th>Functional Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_033379.2</td>
<td>CDC2</td>
<td>cell division cycle 2</td>
<td></td>
</tr>
<tr>
<td>NM_002048.1</td>
<td>GAS1</td>
<td>growth arrest-specific 1</td>
<td></td>
</tr>
<tr>
<td>NM_005263.1</td>
<td>GFI1</td>
<td>growth factor independent 1</td>
<td>Cell cycle</td>
</tr>
<tr>
<td>NM_017579.1</td>
<td>DMBT1</td>
<td>deleted in malignant brain tumors</td>
<td></td>
</tr>
<tr>
<td>NM_000758.2</td>
<td>CSF2</td>
<td>colony stimulating factor 2</td>
<td></td>
</tr>
<tr>
<td>NM_000575.3</td>
<td>IL1A</td>
<td>interleukin 1; alpha</td>
<td>Immune response</td>
</tr>
<tr>
<td>NM_012485.1</td>
<td>HMRR</td>
<td>hyaluronan-mediated motility receptor</td>
<td></td>
</tr>
<tr>
<td>NM_000059.1</td>
<td>BRCA2</td>
<td>breast cancer 2; early onset</td>
<td></td>
</tr>
<tr>
<td>NM_005427.1</td>
<td>TP73</td>
<td>tumor protein p73</td>
<td></td>
</tr>
<tr>
<td>NM_000057.1</td>
<td>BLM</td>
<td>Bloom syndrome</td>
<td></td>
</tr>
<tr>
<td>NM_01951.2</td>
<td>E2FS</td>
<td>E2F transcription factor 5</td>
<td></td>
</tr>
<tr>
<td>NM_002315.1</td>
<td>LMO1</td>
<td>LIM domain only 1</td>
<td></td>
</tr>
<tr>
<td>NM_000135.1</td>
<td>FANCA</td>
<td>Fanconi anemia; complementation group A</td>
<td>DNA repair</td>
</tr>
<tr>
<td>NM_00251.1</td>
<td>MSH2</td>
<td>mutS homolog 2; colon cancer; nonpolyposis type 1</td>
<td>Anti-apoptosis</td>
</tr>
<tr>
<td>NM_00499.2</td>
<td>CYP1A1</td>
<td>cytochrome P450; family 1</td>
<td></td>
</tr>
<tr>
<td>NM_00358.1</td>
<td>TK1</td>
<td>thymidine kinase 1; soluble</td>
<td>Energy pathways</td>
</tr>
<tr>
<td>NM_000015.1</td>
<td>PLG</td>
<td>plasminogen</td>
<td></td>
</tr>
<tr>
<td>NM_000301.1</td>
<td>MMP3</td>
<td>matrix metalloproteinase 3</td>
<td></td>
</tr>
<tr>
<td>NM_002422.1</td>
<td>TGF1</td>
<td>teratocarcinoma-derived growth factor 1</td>
<td>Proliferation</td>
</tr>
<tr>
<td>NM_002809.1</td>
<td>CDC25C</td>
<td>cell division cycle 25C</td>
<td></td>
</tr>
<tr>
<td>NM_00697.1</td>
<td>ALOX12</td>
<td>arachidonate 12-lipoxygenase</td>
<td></td>
</tr>
<tr>
<td>NM_003392.2</td>
<td>WNT5A</td>
<td>wingless-type MMTV integration site family; member</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>NM_001274.2</td>
<td>CHEK1</td>
<td>CHK1 checkpoint homolog</td>
<td></td>
</tr>
<tr>
<td>NM_00219.2</td>
<td>INHA</td>
<td>inhibit; alpha</td>
<td></td>
</tr>
<tr>
<td>NM_03316.3</td>
<td>FGF8</td>
<td>fibroblast growth factor 8</td>
<td></td>
</tr>
<tr>
<td>NM_00430.3</td>
<td>ALK</td>
<td>anaplastic lymphoma kinase</td>
<td></td>
</tr>
<tr>
<td>NM_004119.1</td>
<td>FLT3</td>
<td>fms-related tyrosine kinase 3</td>
<td></td>
</tr>
<tr>
<td>NM_005372.1</td>
<td>MOS</td>
<td>v-mos Moloney murine sarcoma viral oncogene homolog</td>
<td></td>
</tr>
<tr>
<td>NM_198255.1</td>
<td>TERT</td>
<td>telomerase reverse transcriptase</td>
<td>telomere maintenance</td>
</tr>
<tr>
<td>NM_005376.2</td>
<td>MYCL1</td>
<td>v-myc myelocytomatosis viral oncogene homolog 1; ung carcinoma derived (avian)</td>
<td>Transcriptional control</td>
</tr>
<tr>
<td>NM_005378.3</td>
<td>MYCN</td>
<td>v-myc myelocytomatosis viral related oncogene; neuroblastoma derived (avian)</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Spearman rank correlation coefficient analysis of DASL vs. qRT-PCR data

<table>
<thead>
<tr>
<th>GASI</th>
<th>TK1</th>
<th>WNTSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>0.64423</td>
<td>0.50781</td>
<td>0.39923</td>
</tr>
<tr>
<td>4.29493</td>
<td>3.00572</td>
<td>2.22032</td>
</tr>
<tr>
<td>0.00022</td>
<td>0.00580</td>
<td>0.03532</td>
</tr>
</tbody>
</table>

Table 7: Average signal intensity for 3 gene panel in prostate cancer specimens

<table>
<thead>
<tr>
<th>Samples</th>
<th>GASI</th>
<th>TK1</th>
<th>WNTSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMA #28-R</td>
<td>159.1049</td>
<td>1814.738</td>
<td>1186.885</td>
</tr>
<tr>
<td>TMA #29-R</td>
<td>1482.696</td>
<td>1597.475</td>
<td>1545.549</td>
</tr>
<tr>
<td>TMA #30-R</td>
<td>243.8143</td>
<td>1935.002</td>
<td>433.7272</td>
</tr>
<tr>
<td>TMA #31-R</td>
<td>1803.755</td>
<td>1676.184</td>
<td>433.7272</td>
</tr>
<tr>
<td>TMA #34-R</td>
<td>1692.796</td>
<td>2294.229</td>
<td>2554.909</td>
</tr>
<tr>
<td>TMA #36-R</td>
<td>309.4907</td>
<td>2605.582</td>
<td>2667.039</td>
</tr>
<tr>
<td>TMA #38-R</td>
<td>940.9758</td>
<td>1861.906</td>
<td>1386.754</td>
</tr>
<tr>
<td>TMA #39-R</td>
<td>775.5903</td>
<td>711.236</td>
<td>7255.156</td>
</tr>
<tr>
<td>TMA #44-R</td>
<td>-217.1194</td>
<td>2294.229</td>
<td>2554.909</td>
</tr>
<tr>
<td>TMA #46-R</td>
<td>690.5883</td>
<td>1399.041</td>
<td>3222.52</td>
</tr>
<tr>
<td>TMA #48-R</td>
<td>940.9758</td>
<td>1861.906</td>
<td>1386.754</td>
</tr>
<tr>
<td>TMA #50-R</td>
<td>774.949</td>
<td>1110.019</td>
<td>460.8703</td>
</tr>
<tr>
<td>TMA #51-R</td>
<td>-119.9042</td>
<td>2289.141</td>
<td>3255.715</td>
</tr>
<tr>
<td>Samples</td>
<td>GAS1</td>
<td>TK1</td>
<td>WNT5A</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>TMA #52-R</td>
<td>-443.0978</td>
<td>2008.738</td>
<td>2767.253</td>
</tr>
<tr>
<td>TMA #4-R</td>
<td>692.9727</td>
<td>1802.924</td>
<td>1712.075</td>
</tr>
<tr>
<td>TMA #17-NR</td>
<td>4039.194</td>
<td>1087.417</td>
<td>369.8866</td>
</tr>
<tr>
<td>TMA #22-R</td>
<td>1183.308</td>
<td>697.892</td>
<td>723.1877</td>
</tr>
<tr>
<td>TMA #23-NR</td>
<td>1806.918</td>
<td>902.691</td>
<td>1204.131</td>
</tr>
<tr>
<td>TMA #56-NR</td>
<td>2331.08</td>
<td>950.781</td>
<td>90.11768</td>
</tr>
<tr>
<td>TMA #57-NR</td>
<td>1967.901</td>
<td>1969.589</td>
<td>628.6682</td>
</tr>
<tr>
<td>TMA #58-NR</td>
<td>3469.622</td>
<td>905.8434</td>
<td>-6.722403</td>
</tr>
<tr>
<td>TMA #59-NR</td>
<td>730.9319</td>
<td>1750.327</td>
<td>1844.76</td>
</tr>
<tr>
<td>TMA #60-NR</td>
<td>1741.388</td>
<td>278.0175</td>
<td>461.7899</td>
</tr>
<tr>
<td>TMA #62-NR</td>
<td>-722.001</td>
<td>834.6077</td>
<td>2086.022</td>
</tr>
<tr>
<td>TMA #63-NR</td>
<td>2572.344</td>
<td>933.8293</td>
<td>536.2645</td>
</tr>
<tr>
<td>TMA #64-NR</td>
<td>66.54094</td>
<td>875.1831</td>
<td>-707.029</td>
</tr>
<tr>
<td>TMA #65-NR</td>
<td>1755.69</td>
<td>68.09814</td>
<td>-132.0587</td>
</tr>
</tbody>
</table>

References
31. Iozzo RV, Eichstetter I, Danielson KG. Aberrant expression of the


