Biomarkers, Genomics, Proteomics, and Gene Regulation

A 12-Gene Expression Signature Is Associated with Aggressive Histological in Prostate Cancer

SEC14L1 *and* TCEB1 *Genes Are Potential Markers of Progression*

Laia Agell,*[†] Silvia Hernández,[‡] Lara Nonell,[§]
Marta Lorenzo,* Eulàlia Puigdecanet,[§]
Silvia de Muga,*[†] Nuria Juanpere,*[†]
Raquel Bermudo,[¶‖] Pedro L. Fernández,[¶**]
José A. Lorente,[†,††] Sergio Serrano,*[†] and
Josep Lloreta*[‡]

From the Department of Pathology,* Hospital del Mar-Mar Health Park, Barcelona; the Departments of Cell Biology, Morphological Sciences, and Surgery,[†] the Autonomous University of Barcelona, Barcelona; the Department of Health and Experimental Sciences,[‡] Pompeu Fabra University, Barcelona; the Microarray Analysis Service,[§] Hospital del Mar Research Intitute (IMIM), Barcelona; Clinic Hospital,[¶] August Pi i Sunyer Biomedical Research Institute (IDIBAPS) Biobank, Barcelona; the Department of Cell Biology,[‖] Barcelona Molecular Biology Institute, Higher Scienctific Research Council (CSIC), Barcelona; the Department of Pathology,** Clinic Hospital-IDIBAPS, University of Barcelona, Barcelona; and the Department of Urology,[††] Hospital del Mar-Mar Health Park-Hospital del Mar Research Intitute (IMIM), Barcelona, Spain

The main challenge for clinical management of prostate cancer is to distinguish tumors that will progress faster and will show a higher tendency to recur from the more indolent ones. We have compared expression profiles of 18 prostate cancer samples (seven with a Gleason score of 6, eight with a Gleason score of 7, and three with a Gleason score of ≥8) and five nonneoplastic prostate samples, using the Affymetrix Human Array GeneChip Exon 1.0 ST. Microarray analysis revealed 99 genes showing statistically significant differences among tumors with Gleason scores of 6, 7, and ≥8. In addition, mRNA expression of 29 selected genes was analyzed by real-time quantitative RT-PCR with microfluidic cards in an extended series of 30 prostate tumors. Of the 29 genes, 18 (62%) were independently confirmed in the extended series by quantitative RT-PCR: 14 were up-regulated and 4 were down-regulated in tumors with a higher Gleason score. Twelve of these genes were differentially expressed in tumors with a Glea-son score of 6 to 7 versus ≥8. Finally, IHC validation of the protein levels of two genes from the 12-gene signature (*SEC14L1* and *TCEB1*) showed strong protein expression levels of both genes, which were statistically associated with a high combined Gleason score, advanced stage, and prostate-specific antigen progression. This set of genes may contribute to a better understanding of the molecular basis of prostate cancer. *TCEB1* and *SELC14L1* are good candidate markers for predicting prognosis and progression of prostate cancer.   (Am J Pathol 2012, 181:1585–1594; http://dx.doi.org/10.1016/j.ajpath.2012.08.005)

Despite the high incidence of prostate cancer (PCa), only 30% of patients will be afflicted by tumor progression. Prostate cancer (PCa) PCa is the third most diagnosed cancer in Spain[1] and the second leading cause of cancer-related death in US men.[2] There have been numerous advances in basic research on PCa initiation and

progression, as well as clinical advances that have improved patient outcome,[3] but there are still challenges to face, such as identifying which relevant genes are altered and classifying patients into clearly defined high- and low-risk groups to improve their specific management.

The main screening technology that has revolutionized the diagnosis of PCa during the past 30 years is the detection of serum prostate-specific antigen (PSA). In fact, the widespread use of serum PSA has resulted in the identification of an increasing number of asymptomatic low-stage tumors in younger men. Although early diagnosis provides an opportunity for curative surgery, new recommendations[4] favor later and less frequent PSA screening tests, because many patients with clinically localized low-grade carcinomas may not require aggressive treatment and are candidates for active surveillance, because their tumors are relatively indolent. One of the goals of research in PCa is the identification of molecular markers for the early distinction between patients with more aggressive tumors, which will have a higher risk of progression, and, patients with apparently similar tumors, which will carry a much lower risk of progression.

Many molecular studies have shown that genetic alterations are important for prostate carcinogenesis, but few oncogenes or tumor suppressor genes have been consistently linked to prostate adenocarcinoma.[3,5] Microarray technology is a powerful tool for detecting differentially expressed genes and can be useful to search for new prognostic markers that can be translated into clinical practice. In the case of PCa, the microarray gene expression studies previously performed have compared the expression profiles between normal and tumoral prostate tissues[6–14] and between prostate tumor samples classified according to different clinicopathological features, such as metastatic versus organ-confined tumors[15] or the different Gleason score categories.[16–19]

There are previous studies[16–19] reporting differentially expressed genes in association with Gleason score. Singh et al[16] analyzed 52 prostate tumor samples and identified a gene expression signature of 29 genes associated with Gleason score. They have also developed a model that, using gene expression data alone, accurately predicted patient outcome after prostatectomy. Lapointe et al[17] reported a 52-gene expression signature in 62 primary prostate tumors in which two genes (AZGP1 and MUC1) were associated with a higher Gleason score. The expression of these genes was validated by immunohistochemistry (IHC), concluding that they were strong predictors of tumor recurrence. True et al[18] used laser microdissection of prostate tissue to isolate cancer cells from Gleason pattern 3, 4, and 5 foci. They identified an 86-gene profile that distinguished high- from low-grade carcinomas. Recently, Ross et al,[19] using tissue laser microdissection, reported 670 genes that were differentially expressed between Gleason scores 6 and 8. The main involved pathways were androgen receptor signaling, growth factor, and cytokine-mediated pathways.

Herein, we report a gene expression signature of 99 genes differentially expressed in tumors with Gleason scores of 6, 7, and ≥8. From these 99 genes, mRNA expression of 29 selected genes was validated by quantitative RT-PCR (RT-qPCR) in TaqMan low-density arrays (TLDAs), and 18 (62%) of the 29 genes were confirmed as differentially expressed. Subsequently, this signature was further refined to 12 genes that were differentially expressed in tumors with Gleason scores of 6 to 7 versus ≥8. As a result, a signature of PCa with aggressive histological characteristics was obtained. Furthermore, we analyzed the protein expression levels of two of these genes (SEC14L1 and TCEB1) as possible markers for tumor subtypes: high protein levels of both genes were correlated with a Gleason score of ≥8, advanced tumor stage, and PSA progression-free survival. Our results support the existence of an aggressive histological gene expression signature in PCa. TCEB1 and SELC14L1 emerge as new potential molecular markers of poor prognosis in PCa.

Materials and Methods

Tumor Samples and Patients

A total of 30 frozen and 43 formalin-fixed, paraffin-embedded (FFPE) prostate cancer samples were the subject of this study. From the 30 frozen samples, 29 were obtained from radical prostatectomy specimens and 1 was obtained from a cystoprostatectomy specimen with an incidentally detected tumor. From the 43 FFPE samples, 39 were prostatectomy specimens and 4 were needle biopsy specimens. Samples were collected from 2002 to 2010; 20 of them were obtained from the Parc de Salut MAR Biobank, and 10 were obtained from the Tumor Bank of the Hospital Clínic–Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. Samples were obtained following ethical and institutional protocols. Tissue fragments were embedded in optimal cutting temperature (OCT) medium (Tissue-Tek; Sakura Finetek, Torrence, CA), snap frozen, and stored at −80°C. We also included five frozen nontumor prostate samples as controls. The Gleason scores were re-evaluated by two genitourinary pathologists (N.J. and J.A.L.) who reviewed the whole prostate Gleason score, as well as the score in the frozen sample and in the homologous paraffin section, to ensure concordance between the three values. Detailed pathological and clinical data for all of the frozen specimens are provided in Table 1. Regarding the FFPE tissues in which the correlation between clinical variables and IHC expression was performed, the mean follow-up was 34.3 months (range: 11 to 101 months). Tumor progression was considered when PSA values were >0.4 ng/mL after prostatectomy. None of the cases had received preoperative or postoperative radiation or hormone therapy.

Total RNA Isolation

Microscopic examination of H&E-stained sections from frozen tissues was used to select the tumor area. All

Table 1.   Clinical-Pathological Features (Gleason Score, Tumor Stage, and PSA Progression-Free Survival) of the Samples Analyzed in the Microarray and in the RT-qPCR Studies

Tumor no.	Gleason score	Tumor stage	Progression	Microrray analysis	RT-qPCR analysis
1	3 + 3	pT2b	No	Yes	Yes
2	3 + 4	pT2a	No	Yes	Yes
3	3 + 4	pT3a	No	Yes	Yes
4	3 + 3	pT3a	No	Yes	Yes
5	3 + 3	pT3a	Yes	Yes	Yes
6	3 + 4	pT3a	No	Yes	Yes
7	5 + 4	pT2b	No	Yes	Yes
8	3 + 3	pT2a	No	Yes	Yes
9	3 + 3	pT2b	No	Yes	Yes
10	3 + 3	pT2b	No	Yes	Yes
11	4 + 3	pT3b	No	Yes	Yes
12	3 + 3	pT2b	No	Yes	Yes
13	4 + 5	pT3b	Yes	Yes	Yes
14	3 + 4	pT3b	No	Yes	Yes
15	5 + 4	pT3b	Yes	Yes	Yes
16	3 + 4	pT2a	No	Yes	Yes
17	3 + 4	pT2a	No	Yes	Yes
18	3 + 4	pT2a	No	Yes	Yes
19	5 + 5	pT3a	Yes	No	Yes
20	4 + 4	pT3b	No	No	Yes
21	4 + 5	pT3b	No	No	Yes
22	4 + 4	pT3b	Yes	No	Yes
23	3 + 3	pT2a	No	No	Yes
24	3 + 3	pT2b	No	No	Yes
25	3 + 3	pT3a	No	No	Yes
26	3 + 3	pT2b	No	No	Yes
27	3 + 3	pT3a	No	No	Yes
28	4 + 3	pT3a	No	No	Yes
29	3 + 4	pT3a	No	No	Yes
30	3 + 4	pT3a	No	No	Yes
31	Normal tissue	NA	No	Yes	Yes
32	Normal tissue	NA	No	Yes	Yes
33	Normal tissue	NA	No	Yes	Yes
34	Normal tissue	NA	No	Yes	Yes
35	Normal tissue	NA	No	Yes	Yes

NA, not applicable.

cases contained a minimum of 70% of tumor cells, with most of the cases higher than that figure and with a maximum of near 100%. Total RNA was extracted from the 30 frozen prostate tumor samples and 5 nontumor samples with Ultraspec (Biotecx Laboratories, Houston, TX) and an RNeasy Mini kit (Qiagen, Cathsworth, CA), from 10 to 15 sections (10 μm thick). Total RNA purity and quality were assessed with a NanoDrop ND-100 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). Only samples with good RNA integrity (RNA integrity number) were subsequently used in microarray experiments. From the 23 samples analyzed by microarray experiments, 18 were prostate tumors and 5 were normal prostate tissues. Thirteen tumors showed RNA integrity number values ≥7. The five prostate tumors and the five normal prostate tissues showed RNA integrity number values between 6.3 and 6.9.

Microarray Hybridization

A total of 23 frozen prostate samples were used for microarray analysis. Of these samples, 5 were normal prostate tissues and 18 were prostate tumor tissues. Prostate tumors were grouped according to Gleason score: 6 (n =

7), 7 (n = 8), and ≥8 (n = 3). Total RNA, 200 ng, from each sample was processed and hybridized to Affymetrix Human Array GeneChip Exon 1.0 ST (Affymetrix, Santa Clara, CA), according to the Affymetrix GeneChip Whole Transcript Sense Target Labeling Assay. After hybridization, the array was washed and stained in the Affymetrix GeneChip Fluidics Station 450. The stained array was scanned using an Affymetrix GeneChip Scanner 3000 7G, generating .CEL files for each array.

Gene Expression Profile Analysis

After quality control of raw data, they were background corrected, quantile normalized, and summarized to a logarithmic gene level by the robust multichip average,[20] obtaining a total of 18,708 transcript clusters. Core annotations were used to summarize data into transcript clusters. Normalized data were then filtered to avoid noise generated by nonexpressed transcript clusters. Only transcripts with a signal intensity higher than the median values in any of the groups were considered for further analysis, which led to 10,452 transcript clusters. Linear Models for Microarray,[21] a moderate t-statistics model, was used for detecting differentially expressed genes among the conditions in the study. We have used

the standard microarray analysis method, applying the false-discovery rate to correct for multiple comparisons,[22] and only genes with an adjusted $P < 0.05$ were considered significant. We have also performed Volcano plots for paired conditions (see Supplemental Figure S1 at *http://ajp.amjpathol.org*).

Hierarchical cluster analysis was also performed to see how data aggregated and to generate heat maps. All data analysis was performed in R version 2.11.1 (R-project Foundation, Auckland, New Zealand) with packages aroma.affymetrix, Biobase, Affy, limma, and genefilter. Functional analysis was performed with Ingenuity Pathway Analysis software version 9.0 (Ingenuity Systems, Inc, Redwood, CA) and GSEA software (Gene Set Enrichment Analysis, Cambridge, MA).[23] The data discussed herein have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus[24] and are accessible through Gene Expression Omnibus Series accession number GSE30521 (*http://www.ncbi.nlm.nih.gov/geo*).

To validate the concordance between our gene signature results and those of previous studies, we performed a series of comparative tests. For each referred list from previous articles, the total number of genes in the respective platform was considered. A sample of the published genes was then randomly selected, and the number of genes that were coincident with a similarly random selection of 99 genes from our platform was assessed; this procedure was repeated 1 million times (see Supplemental Table S1 at *http://ajp.amjpathol.org*). The resulting percentage indicates the effect of random concordance.

Real-Time RT-qPCR Analysis

Twenty-nine genes were selected for expression validation through RT-qPCR in the TLDA (Applied Biosystems, Foster City, CA). In addition to the 18 tumor samples previously analyzed in the Affymetrix Human Array GeneChip Exon 1.0 ST, 12 new samples from additional prostate tumors were included in this analysis. All these cases ($n = 30$) were grouped according to their Gleason score as follows: 6 ($n = 12$), 7 ($n = 11$), and ≥ 8 ($n = 7$). Custom-designed TLDAs contained primers and probes for 29 genes (see Supplemental Table S2 at *http://ajp.amjpathol.org*). We selected 29 of the initial 99 genes for the TLDA validation based on our previous reports on PCa[25,26] and by different functional criteria. Some belonged to the phosphatidylinositol 3-kinase–AKT signaling pathway or to the Ras family, other genes were involved in cell cycle control or DNA repair, another group of genes was located at chromosome 8 in a region reported to be amplified in 40% of high-grade prostate cancer tumors,[27] and, finally, other genes were reported in previous gene expression analyses on prostate cancer.[16–19] TLDA was configured for the analysis of 32-gene sets in triplicate, using an ABI PRISM 7900 HT instrument (Applied Biosystems). A total of 100 μL of reaction mixture with 50 μL of cDNA template (1000 ng) and an equal volume of TaqMan universal master mix (Applied Biosystems) were added to each loading port of TLDA. Thermal cycler conditions were as follows: 2 minutes at 50°C, 10 minutes at 94.5°C, and then 30 seconds at 97°C and 1 minute at 59.7°C for 40 cycles. The C_T was automatically given by an SDS 2.1 software package (Applied Biosystems). Relative quantification values were determined using the following equation: Relative Quantification = $2^{-\Delta\Delta C_T}$.

Average C_T values were obtained using the SDS 2.1 software. The relative expression level of each target gene was displayed as follows: $\Delta C_T = C_{Tref} - C_{Ttarget}$. *GAPDH* and *B2M* were used as endogenous control genes; according to a previous study, *HPRT1* was also included.[28] Normalization was performed using the geometric mean of the three housekeeping genes,[29] and gene expression was validated by an analysis of variance test.

IHC of SEC14L1 and TCEB1 in Prostate Tumors

IHC staining for SEC14L1 and TCEB1 was performed with SEC14L1 antibody (Sigma-Aldrich, St Louis, MO) and TCEB1 antibody (ProteinTech Group, Inc., Chicago, IL), respectively. SELC14L1 antibody was used at 1:50 dilution, and TCEB1 antibody was used at 1:25 dilution, after antigen retrieval with citrate buffer (pH 9) in autoclave.

Forty-three new independent samples not used in the previous mRNA expression analysis were tested for SEC14L1 and TCEB1 protein immunostaining (15 tumors with a Gleason score of 6, 17 tumors with a Gleason score of 7, and 11 tumors with a Gleason score of ≥ 8). Each antibody was detected in both cytoplasm and nucleus. The results were graded, considering separately cytoplasm and nuclear immunostaining, as 0 (negative), 1 (weak), 2 (moderate), and 3 (strong). The score (histoscore) for each of them was the sum of the product of the staining intensity and the corresponding tumor percentage (Histoscore = [$1 \times$ (%1 + Cells)] + [$2 \times$ (%2 + Cells)] + [$3 \times$ (%3 + Cells)]). For this study, the global tumor histoscore was obtained from the addition of the nuclear and cytoplasmic histoscores.

Statistical Analysis

Categorical variables are presented as frequencies and percentages, and quantitative variables are presented as median and range. The receiver operating characteristic curve was obtained to quantify the discrimination power and to determine the optimal cutoff points for *SEC14L1* and *TCEB1* histoscore values with respect to Gleason score (≥ 180 and ≥ 125), tumor stage (≥ 180 and ≥ 215), and progression (≥ 210 and ≥ 125), respectively. The Fisher's exact test was used to assess the relationship between two categorical variables. $P > 0.05$ was considered statistically significant. Statistical analysis was performed using the SPSS statistical package, version 15.0 (SPSS Inc., Chicago, IL). The relationship with PSA progression-free survival was analyzed using the Kaplan-Meier (log-rank) test in 42 patients (one patient was lost to follow-up). For PSA progression-free survival analysis, patients were censored at their last clinical follow-up ap-

PARP5
SLPI
PLAT
OPRK1
EPB41L4B
NAPG
FAM174A
EIF1
NIPA2
LPGAT1
CAMK1D
PSMA7
KLK12
KCTD12
WRB
MS4A8B
C1orf58
PARP1
MED14
ACVR2A
PPM2C
NCOA3
RNF19A
BTBD10
CNIH4
UBE2W
C1D
SECISBP2
SPTLC2
WDYHV1
KRAS
ZHX1
C17orf75
ABI1
PRKDC
OSBPL8
RAD21
ZNF706
CASK
STRBP
SUMO2
CFL1
YWHAZ
TCEB1
PCMT1
RAB2A
HNRNPA0
COMMD2
POLR2K
DDX27
SEC14L1
TMEM189–UBE2V1
TBCB
MAPRE1
KPNA2
THOC4
RAE1
PIGU
LOC552889
CDKN1B
RHEB
DCK
H6PD
HS3ST3A1
UFSP1
PRRX2
SOX17
ZNF703
FBLN2
SEPN1
S100A3
OLFML3
CCND2
C11orf54
C7orf23
C16orf70
RPS14
ELF2
TCEB3
ATP8B1
RBL2
LIFR
RNLS
SFXN2
PTEN
AZGP1
MYBPC1
C3orf25
CHRNA2
ACADL
D0XDC1
FLJ41603
SIDT1
SNTB2
AK3L1
EPHX2
RPL13
DPP4
HGD

Figure 1. Heat map showing the 99 differentially expressed genes for the three analyzed conditions (prostate carcinomas with combined Gleason scores of 6, 7, and ≥8).

pointment or when an increase in serum PSA >0.4 ng/mL was detected.

Results

Gene Expression Profile Associated with Prostate Cancer

In the gene expression microarray experiments performed on 18 prostate tumor samples and 5 normal prostate samples, we identified an initial set of 3380 genes differentially expressed between prostate cancer and normal prostate tissue (see Supplemental Table S3 at *http://ajp.amjpathol.org*).

Herein, we mainly concentrated on comparing tumor samples, classified according to Gleason score. We identified a cohort of 99 differentially expressed genes that could distinguish between tumors with a combined Gleason score of 6 (3 + 3), tumors with a Gleason score of 7 (3 + 4 or 4 + 3), and the more aggressive tumors (those with a Gleason score of ≥8) (Figure 1; see also Supplemental Table S4 at *http://ajp.amjpathol.org*). The analysis of these genes, using Ingenuity Pathways Analysis software, showed six top canonical pathways: i) phosphatidylinositol 3-kinase–AKT (*CDKN1B* and *YWHAZ*), ii) estrogen receptor (*PRKDC* and *NCOA3*), iii) glioma signaling (*CAMK1D*), iv) DNA double-stranded break repair by nonhomologous end joining (*PARP1* and *PRKDC*), v) cell cycle [G$_1$/S checkpoint regulation (*CDKN1B*)], and vi) granzyme B signaling (*PARP1* and *PRKDC*). Calculation was performed according to either ratio (number of genes from the data set mapping to a given canonical pathway/total number of genes mapping to the same canonical pathway) or significance. The probability of random concordance with signatures in

Table 2. *t*-Test and ANOVA of the 12-Gene Signature

Type of test	AZPG1	DPP4	HGD	MYBPC1	PARP1	PRKDC	RNF19A	SEC14L1	SLP1	TCEB1	YWHAY	ZNF706
t-Test												
3 + 3 vs 4 + 5	0.02	0.01	0.00	0.00	0.00	0.02	0.01	0.04	0.00	0.04	0.02	0.014
3 + 3 vs 3 + 4	0.78	0.19	0.38	0.22	0.90	0.56	0.06	0.13	0.06	0.08	0.32	0.072
3 + 4 vs 4 + 5	0.02	0.04	0.04	0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.05	0.00
ANOVA	0.01	0.00	0.00	0.00	0.01	0.03	0.000	0.00	0.00	0.00	0.03	0.00

ANOVA, analysis of variance.

previous studies is shown in Supplemental Table S1 (available at *http://ajp.amjpathol.org*).

Gene Validation by RT-qPCR Analysis and Prostate Cancer Aggressive Histological Signature

We selected 29 of the initial 99 genes for the TLDA validation based on our previous reports on PCa[25,26] and by different functional criteria. Some genes belonged to the phosphatidylinositol 3-kinase–AKT signaling pathway or to the Ras family, others were involved in cell cycle control or DNA repair, another group of genes was located at chromosome 8 in a region reported to be amplified in 40% of high-grade prostate cancer tumors,[27] and, finally, several genes were reported in previous gene expression analyses on prostate cancer.[16–19] In this RT-qPCR validation analysis, we used a group of 30 cases composed of the 18 tumor samples previously analyzed in the microarray study, plus a second set of 12 new prostate tumor samples that we had available. The reason for including the initial cases in the validation set was the relatively few new high-quality samples available with adequate follow-up. Although it would have been better to have a complete new validation set, it was also true that including approximately 40% of new cases ensured that the mixture was different. On the other hand, the mean 3-year follow-up was based on the fact that, for practical purposes, we considered PSA progression-free survival the most reliable end point. This parameter was used in patient management, it was likely to detect earlier progression of the disease, and it allowed including many

samples. On the other hand, a 10-year overall survival requirement would strongly limit the availability of cases.

Of the 29 genes, 18 (62%) were validated (see Supplemental Table S2 at *http://ajp.amjpathol.org*). We performed scatter plots with regression lines for each of the 29 genes validated by RT-qPCR and for the 18 samples with both microarray and RT-qPCR data. Each plot contained Spearman ρ and *P* values (see Supplemental Figure S2 at *http://ajp.amjpathol.org*).

These 18 validated genes were as follows: *AZGP1, CAMK1D, CDKN1B, DPP4, HGD, KCTD12, MYBPC1, NCOA3, PARP1, PPM2C, PRKDC, RNF19A, SEC14L1, SLPI, SUMO2, TCEB1, YWHAZ,* and *ZNF706*. All of them had significant *P* values by the analysis of variance test. We performed a series of comparisons (Gleason score 6 versus ≥8, Gleason score 7 versus ≥8, and Gleason score 6 versus 7), with the purpose of selecting those genes that could distinguish cases with a Gleason score of ≥8 from the rest (Table 2). By using this approach, we narrowed the number of genes distinguishing low-grade (≤7) from high-grade (≥8) tumors down to 12. From this signature of 12 genes, 4 were down-regulated (*AZGP1, DPP4, HGD,* and *MYBPC1*) and 8 were up-regulated (*PARP1, PRKDC, RNF19A, SEC14L1, SLPI, TCEB1, YWHAZ,* and *ZNF706*) in tumors with a Gleason score of ≥8 (Table 2 and Figure 2). Regarding the down-regulated genes, the level of down-regulation ranged from
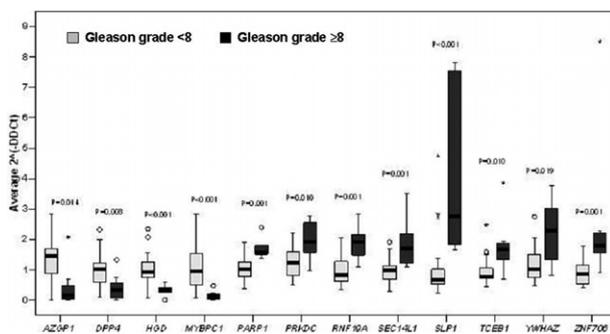

Figure 2. Box plot for the 12-gene signature, comparing the RT-qPCR gene expression levels in prostate tumors with a combined Gleason score of ≤7 versus ≥8. The 12 genes are displayed on the *x* axis, and the respective expression levels are displayed on the *y* axis. **Asterisk**, extreme outlier value; **circle**, outlier value.
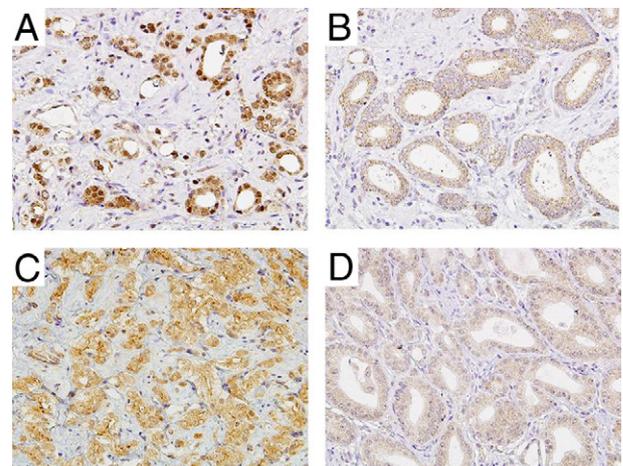

Figure 3. The IHC expression of *SEC14L1* and *TCEB1*. Original magnification, ×200. **A:** Combined Gleason score of 3 + 4 (+5) PCa, showing moderate to strong *SEC14L1* expression (histoscore = 310). **B:** Combined Gleason score of 3 + 3 tumor, showing weak *SEC14L1* expression (histoscore = 110). **C:** Combined Gleason score of 4 + 5 (+3) PCa, showing strong *TCEB1* expression (histoscore = 255). **D:** Gleason score of 3 + 3 tumor, showing weak *TCEB1* expression (histoscore = 120).
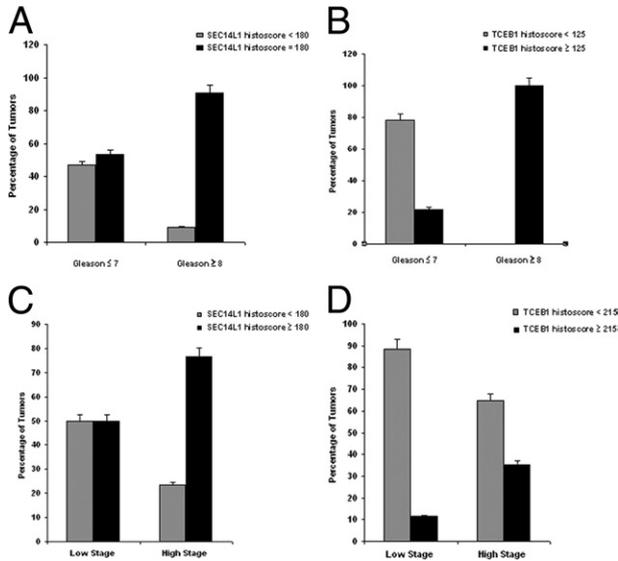
Figure 4. *SEC14L1* and *TCEB1* IHC expression according to combined Gleason score (**A** and **B**) and pathological stage (**C** and **D**). **A:** A total of 53% (17/32) of Gleason score ≤7 and 91% (10/11) of Gleason score ≥8 tumors have strong *SEC14L1* expression. **B:** A total of 22% (7/32) of Gleason score ≤7 tumors and the 11 (100%) tumors with a Gleason score ≥8 have strong *TCEB1* expression. **C:** A total of 50% (13/26) of pT2 or less and 76.5% (13/17) of pT3 or more tumors have strong *SEC14L1* expression. **D:** A total of 11.5% (3/26) of pT2 or less and 35% (6/17) of pT3 or more tumors have strong *TCEB1* expression.

0.14 for *MYBPC1* to 0.416 for *DPP4*. Among the up-regulated genes, the degree of up-regulation ranged from 1.6 for *PARP1* to 5.5 for *SLPI* (Figure 2).

IHC Validation of SECL14L1 *and* TCEB1

From the final 12-gene signature associated with high-grade tumors, we selected two genes, *SEC14L1* and *TCEB1*, to investigate the relationship between their protein expression levels, assessed by IHC, and clinical-pathological parameters (combined Gleason score, pathological stage, and PSA progression-free survival). For this purpose, we used an independent set of 43 primary prostate tumors. The two genes were selected because they showed up-regulation in the RT-qPCR study, clear separation between Gleason groups, and a relatively narrow range of expression.

The evaluation of protein expression was based on the histoscore calculation (nucleus + cytoplasm), which ranged from 100 to 375 for *SEC14L1* and from 5 to 280 for *TCEB1*. Different cutoff levels were used for the correlation with combined Gleason score, stage, and PSA progression-free survival (see *Materials and Methods*). An *SEC14L1* histoscore of ≥180 was statistically associated with a Gleason score of ≥8 (*P* = 0.03, Fisher's exact test) (Figure 3); thus, 17 (53%) of 32 Gleason score ≤7 versus 10 (91%) of 11 Gleason score ≥8 tumors showed high SEC14L1 protein levels (Figure 4). A *TCEB1* histoscore of ≥125 was also statistically associated with a high Glea-son score (*P* = 0.003, Fisher's exact test) (Figure 3); 7 (22%) of the 32 Gleason score ≤7 versus all of the 11 (100%) Gleason score ≥8 tumors showed high TCEB1 protein levels. Furthermore, there were no statistical dif-

ferences between tumors with combined Gleason scores 6 and 7 in the histoscore levels of *SEC14L1* (*P* = 0.076, Fisher's exact test) and *TCEB1* (*P* = 0.941, Fisher's exact test). These results were in concordance with RT-qPCR analysis results.

Regarding pathological tumor stage, an *SEC14L1* his-toscore of ≥180 was statistically associated with high (≥pT3) tumor stage (*P* = 0.03, Pearson's χ^2 test). Also, a *TCEB1* histoscore of ≥215 was statistically associated with a high tumor stage (*P* = 0.04, Fisher's exact test).

Finally, a Kaplan-Meier analysis for PSA progression-free survival, performed on 42 patients, showed that there was a statistical association between PSA progression-free survival and histoscore levels of *SEC14L1* ≥210 (log-rank test, *P* < 0.006) (Figure 5A) and of *TCEB1* ≥125 (log-rank test, *P* = 0.001) (Figure 5B). We also analyzed the relationship between the combination of the two an-tibodies and PSA progression-free survival. The results were statistically significant (log-rank test, *P* = 0.003), with the best survival in patients with low histoscore val-ues and the worst survival in those with high values for the two antibodies. Patients with high *TCEB1* and low *SEC14L1* levels had intermediate survival values (Figure 5C). We did not include the combination of low *TCEB1* and high *SEC14L1* levels because there was only one case in this category.

Discussion

Gene expression profiling by microarray and RT-qPCR techniques has been a useful tool to classify tumors at the molecular level. Its application may be helpful in improv-ing diagnosis, prognosis, and patient stratification.[30] The discovery of new therapeutic targets and new means for customizing therapy, specific to patient profiles, is a key objective in the management of PCa. Several previous
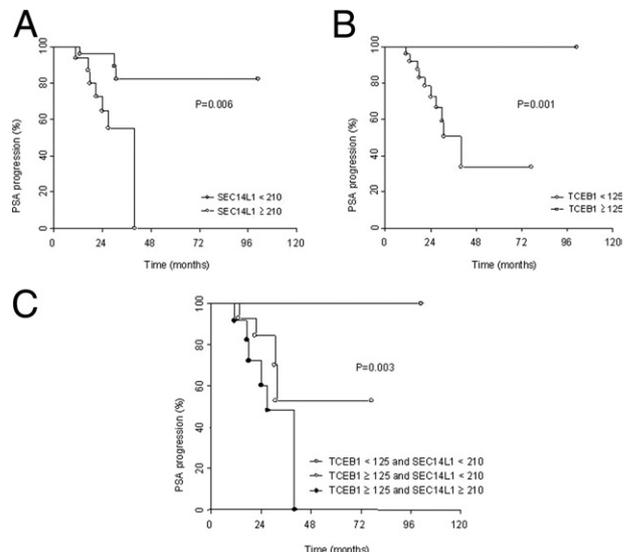

Figure 5. PSA progression-free survival (Kaplan-Meier) plots for *SEC14L1* (**A**), *TCEB1* (**B**), and the combination of *SEC14L1* plus *TCEB1* (**C**). Follow-up (in months) is shown on the *x* axis, and cumulative PSA progression-free survival is shown on the *y* axis.

reports on gene expression microarrays in prostate tumors have been published. Some articles have compared normal prostate with prostate tumor tissues and have found different gene signatures associated with PCa.[6–14] On the other hand, some studies have investigated the gene expression profiles associated with the different clinicopathological prostate tumor categories, such as organ-confined versus metastatic tumors,[15] or different Gleason score groups.[16–19]

The main goal of our approach has been to identify new molecular predictors of prostate tumor behavior and progression. The present study, although based on a few cases, identifies a new prostate cancer signature, with a 12-gene expression profile associated with aggressive histological characteristics. Two genes, *SEC14L1* and *TCEB1*, validated by RT-qPCR and IHC, emerge as potential molecular markers for prostate cancer progression and prognosis.

As a secondary goal, we have compared tumor with normal prostate samples, and we have reported 3380 differentially expressed genes. Our results show that some of the top 100 differentially expressed genes have been previously reported. For example, Bermudo et al[13] reported 26 validated genes by RT-qPCR in tumor versus normal samples. From these genes, we have four in common: *ROR2*, *LAMB3*, *CX3CL1*, and *TACSTD1*. The *TACDST1* gene was reported by Welsh et al[7] and by Luo et al,[12] with whom we also share the *KRT14* gene. Finally, we also share the *ZNF185*, *CSRP1*, and *TRIM29* genes with Vanaja et al.[14] Thus, the concordance with previous expression-profiling studies shows that our method replicates some of their results. Other genes among the first 400 have also been well documented in the literature, such as *AMACR*[7,12,13,31] and *KRT15*.[12,13]

We have mainly focused our study on the comparison between the microarray profiles of the three different Gleason score categories. We have initially detected 99 differentially expressed genes. From these genes, 29 were selected for assessment by RT-qPCR analysis and 62% of them were validated. The validation index in our study is similar to those in other previous reports, and illustrates the need for verifying the results of microarray assays by complementary techniques, such as RT-qPCR and IHC. The lack of validation of some of the genes, as well as the relatively low concordance among the different microarray studies, could be because of the heterogeneity of the prostate samples. A major concern of microarray studies on prostate cancer is sample heterogeneity, particularly regarding the proportion of tumor cells versus normal glands and stroma. Although most of our samples had a proportion of tumor cells >70%, we cannot exclude this factor as the reason for the lack of validation in some genes. Another explanation could be the presence of sequence differences between the probes used for the RT-qPCR and microarray analyses.

Several reports have investigated the expression profiles associated with PCa. Different authors have identified expression signatures with 29, 52, 86, and 670 genes, respectively, that were statistically associated with the Gleason score.[16–19] When comparing our results

with previous literature, our study shares nine genes with True et al[18]: *KCTD12*, *YWHAZ*, *RAB2*, *SEC14L1*, *TCEB1*, *MYBPC1*, *HGD*, *AZGP1*, and *DPP4*. From these genes, eight were validated in our RT-qPCR study and one was not (*RAB2*). In addition, *AZGP1* is the only gene that we have in common with Lapointe et al.[17] They validated this gene by IHC, and they found that strong expression of *AZGP1* was associated with decreased risk of recurrence (*P* = 0.0008), independent of tumor grade, stage, and preoperative PSA levels. On the other hand, Ross et al[19] reported three genes (*TCEB1*, *KCTD12*, and *PPM2C*) that were also present in the signature of True et al[18] and in our own set of validated genes. Considering our global list of 99 genes, we have two other genes in common with Singh et al,[16] *CCND2* and *RPL13*. *CCND2* was not validated in our RT-qPCR study, and *RPL13* was not included among the 29 genes selected for RT-qPCR validation. Finally, there are three genes in common between Singh et al[16] and Lapointe et al[17]: *SPARC*, *BGN*, and *COL1A2*. However, subsequent studies,[18,19] including our own, have not found these genes in their signatures.

To validate the concordance between our results and previous studies, it is advisable to exclude an effect of random concordance. For this purpose, we have performed a series of comparative tests. Thus, for each referred list from previous articles, the total number of genes in the respective platform was considered. A sample of the published genes was then randomly selected, and the number of genes that were coincident with a similarly random selection of 99 genes from our platform was assessed (see Supplemental Table S1 at http://ajp.amjpathol.org). The percentage indicates the concordance obtained by performing this strategy of random resampling up to 1,000,000 times. The probability of sharing the same genes in the different signatures, as a result of a random coincident event, is low compared with most of the previous studies, with the exception of three genes in common with the article by Ross et al,[19] a fact that could be explained by the many genes contained in their signature.

By performing comparisons two by two (Gleason score of 6 versus ≥8, Gleason score of 7 versus ≥8, and Gleason score of 6 versus 7), we have selected those genes that distinguished tumors with Gleason scores of ≤7 and ≥8, thus refining a gene signature with 12 differentially expressed genes. From these 12 genes, 4 were down-regulated (*AZGP1*, *DPP4*, *HGD*, and *MYBPC1*) and 8 were up-regulated (*PARP1*, *PRKDC*, *RNF19A*, *SEC14L1*, *SLPI*, *TCEB1*, *YWHAZ*, and *ZNF706*). These genes could be markers of an aggressive phenotype. Several validated genes that we found associated with Gleason score in our analysis have been previously linked to different human neoplasms, including prostate cancer. For example, *PARP1* is involved in the regulation of various important cellular processes, such as differentiation, proliferation, and tumor transformation. The use of inhibitors of *PARP1* is a recent promising therapy in breast and prostate cancers.[32,33] The *SLPI* gene has been involved in the secretory machinery of PSA in prostate carcinoma cells.[34] *DPP4* is secreted by the normal prostate and has inhibited the malignant phenotype of

prostate cancer cells by blocking the basic fibroblast growth factor signaling pathway.[35] *AZGP1* is associated with a decreased risk of prostate cancer recurrence.[18] Some authors have reported an association between the loss of *AZGP1* expression and recurrence of prostate cancer.[36,37] Our results are in keeping with these studies, because we found *AZGP1* to be down-regulated in high-grade tumors. Finally, 3 of the 12 genes in our refined set (*RNF19A*, *TCEB1*, and *ZNF706*) are located in 8q21-23, a region amplified in >40% of primary prostate cancers and associated with higher histological grades.[27]

For the present study, we selected two genes, *SEC14L1* and *TCEB1*, to be assessed by IHC because of their up-regulation in the RT-qPCR study, clear-cut separation between Gleason groups, and a relatively narrow range of expression. To our knowledge, there are no previous IHC studies on clinical PCa samples for any of these genes.

SEC14L1 has not been investigated in PCa. Its protein belongs to the SEC14 cytosolic factor family, and its role in intracellular transport has been previously analyzed.[38] Our results indicate that strong immunostaining of *SEC14L1* (histoscore levels ≥180) is associated with a Gleason score of ≥8 and advanced tumor stage. Moreover, histoscore levels of ≥210 are inversely associated with PSA progression-free survival.

On the other hand, there are some studies on *TCEB1* in PCa. Interestingly, *TCEB1* is located at chromosome region 8q21.11; the gain of the long arm of chromosome 8 (8q) is one of the most commonly recurrent findings in advanced prostate tumors, and it is associated with poor prognosis.[39,40] Moreover, Porkka et al[41] have shown by fluorescence in situ hybridization analysis that 23% of hormone-independent prostate tumors had *TCEB1* amplification, whereas none of the hormone-dependent tumors did, and that amplification of *TCEB1* was associated with advanced androgen-independent prostate cancer.[41] Finally, Jalava et al[42] have shown that *TCEB1* promotes invasion in prostate cancer cells. Our analysis has revealed that the *TCEB1* gene is also up-regulated in tumors with a Gleason score of ≥8, advanced tumor stage, and PSA progression-free survival.

Interestingly, the PSA progression-free survival analysis, using a combination of both antibodies, is also statistically significant and has the added value of better patient stratification.

In conclusion, the present study reveals global gene expression differences that are sufficiently robust to distinguish tumors with Gleason scores of 6, 7, and ≥8. In addition, these results show that there is a 12-gene signature associated with aggressive tumor histological characteristics. Protein levels of two genes in this 12-gene signature, *SEC14L1* and *TCEB1*, have been identified as good candidate predictors of progression. The role of the remaining genes in this signature in the pathogenesis of prostate cancer remains to be elucidated. It will be interesting to observe patients with a Gleason score of ≤7 and high SEC14L1 and TCEB1 protein levels, to assess if they could be at a higher risk of tumor progression.

References

1. Cayuela A, Rodriguez-Dominguez S, Martin VE, Candau BR: Recent changes in prostate cancer mortality in Spain: trends analysis from 1991 to 2005. Actas Urol Esp 2008, 32:184–189
2. Kopper L, Timar J: Genomics of prostate cancer: is there anything to "translate"? Pathol Oncol Res 2005, 11:197–203
3. Shen MM, Abate-Shen C: Molecular genetics of prostate cancer: new prospects for old challenges. Genes Dev 2010, 24:1967–2000
4. Wolf AM, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, Volk RJ, Brooks DD, Dash C, Guessous I, Andrews K, DeSantis C, Smith RA; American Cancer Society Prostate Cancer Advisory Committee: American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin 2010, 60:70–98
5. Dong JT: Prevalent mutations in prostate cancer. J Cell Biochem 2006, 97:433–447
6. Magee JA, Araki T, Patil S, Ehrig T, True L, Humphrey PA, Catalona WJ, Watson MA, Milbrandt J: Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res 2001, 61:5692–5696
7. Welsh JB, Sapinoso LM, Su AI, Kern SG, Wang-Rodriguez J, Moskaluk CA, Frierson HF Jr, Hampton GM: Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res 2001, 61:5974–5978
8. Chaib H, Cockrell EK, Rubin MA, Macoska JA: Profiling and verification of gene expression patterns in normal and malignant human prostate tissues by cDNA microarray analysis. Neoplasia 2001, 3:43–52
9. Bull JH, Ellison G, Patel A, Muir G, Walker M, Underwood M, Khan F, Paskins L: Identification of potential diagnostic markers of prostate cancer and prostatic intraepithelial neoplasia using cDNA microarray. Br J Cancer 2001, 84:1512–1519
10. Chetcuti A, Margan S, Mann S, Russell P, Handelsman D, Rogers J, Dong Q: Identification of differentially expressed genes in organ-confined prostate cancer by gene expression array. Prostate 2001, 47:132–140
11. 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
12. Luo JH, Yu YP, Cieply K, Lin F, Deflavia P, Dhir R, Finkelstein S, Michalopoulos G, Becich M: Gene expression analysis of prostate cancers. Mol Carcinog 2002, 33:25–35
13. Bermudo R, Abia D, Ferrer B, Nayach I, Benguria A, Zaballos A, del Rey J, Miào R, Campo E, Martínez-A C, Ortiz AR, Fernández PL, Thomson TM: Co-regulation analysis of closely linked genes identifies a highly recurrent gain on chromosome 17q25.3 in prostate cancer. BMC Cancer 2008, 8:315
14. Vanaja DK, Cheville JC, Iturria SJ, Young CY: Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res 2003, 63:3877–3882
15. 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
16. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP: Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 2002, 1:203–209
17. 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 P, 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
18. 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, Melson PS: A molecular correlate to the gleason grading system for prostate adenocarcinoma. Proc Natl Acad Sci U S A 2006, 103:10991–10996

19. Ross AE, Marchionni L, Vuica-Ross M, Cheadle C, Fan J, Berman DM, Schaeffer EM: Gene expression pathways of high grade localized prostate cancer. Prostate 2011, doi: 10.1002/pros.21373

20. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 2003, 31:e15

21. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004, 3:Article3

22. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995, 57:289–300

23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005, 102:15545–15550

24. Edgar R, Domrachev M, Lash AE: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002, 30:207–210

25. Agell L, Hernandez S, Salido M, de Muga S, Juanpere N, Arumi-Uria M, Menendez S, Lorenzo M, Lorente JA, Serrano S, Lloreta J: PI3K signaling pathway is activated by PIK3CA mRNA overexpression and copy gain in prostate tumors, but PIK3CA, BRAF, KRAS and AKT1 mutations are infrequent events. Mod Pathol 2011, 24:443–452

26. de Muga S, Hernandez S, Agell L, Salido M, Juanpere N, Lorenzo M, Lorente JA, Serrano S, Lloreta J: Molecular alterations of EGFR and PTEN in prostate cancer: association with high-grade and advanced-stage carcinomas. Mod Pathol 2010, 23:703–712

27. Tsuchiya N, Slezak JM, Lieber MM, Bergstralh EJ, Jenkins RB: Clinical significance of alterations of chromosome 8 detected by fluorescence in situ hybridization analysis in pathologic organ-confined prostate cancer. Genes Chromosomes Cancer 2002, 34:363–371

28. Ohl F, Jung M, Xu C, Stephan C, Rabien A, Burkhardt M, Nitsche A, Kristiansen G, Loening SA, Radonic A, Jung K: Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization? J Mol Med 2005, 83:1014–1024

29. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002, 3:RESEARCH0034

30. Hoheisel JD: Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet 2006, 7:200–210

31. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM: Integrative molecular concept modeling of prostate cancer progression. Nat Genet 2007, 39:41–51

32. Dong Y, Bey EA, Li LS, Kabbani W, Yan J, Xie XJ, Hsieh JT, Gao J, Boothman DA: Prostate cancer radiosensitization through poly(ADP-ribose) polymerase-1 hyperactivation. Cancer Res 2010, 70:8088–8096

33. Haffner MC, De Marzo AM, Meeker AK, Nelson WG, Yegnasubramanian S: Transcription-induced DNA double strand breaks: both an oncogenic force and potential therapeutic target? Clin Cancer Res 2011, 17:3858–3864

34. Catz SD: Characterization of Rab27a and JFC1 as constituents of the secretory machinery of prostate-specific antigen in prostate carcinoma cells. Methods Enzymol 2008, 438:25–40

35. Wesley UV, McGroarty M, Homoyouni A: Dipeptidyl peptidase inhibits malignant phenotype of prostate cancer cells by blocking basic fibroblast growth factor signaling pathway. Cancer Res 2005, 65: 1325–1334

36. Henshall SM, Horvath LG, Quinn DI, Eggleton SA, Grygiel JJ, Stricker PD, Biankin AV, Kench JG, Sutherland RL: Zinc-alpha2-glycoprotein expression as a predictor of metastatic prostate cancer following radical prostatectomy. J Natl Cancer Inst 2006, 98: 1420–1424

37. Yip PY, Kench JG, Rasiah KK, Benito RP, Lee CS, Stricker PD, Henshall SM, Sutherland RL, Horvath LG: Low AZGP1 expression predicts for recurrence in margin-positive, localized prostate cancer. Prostate 2011, 71:1638–1645

38. Zhao S, Xu C, Qian H, Lv L, Ji C, Chen C, Zhao X, Zheng D, Gu S, Xie Y, Mao Y: Cellular retinaldehyde-binding protein-like (CRALBPL), a novel human Sec14p-like gene that is upregulated in human hepatocellular carcinomas, may be used as a marker for human hepatocellular carcinomas. DNA Cell Biol 2008, 27:159–163

39. Ribeiro FR, Jeronimo C, Henrique R, Fonseca D, Oliveira J, Lothe RA, Teixeira MR: 8q Gain is an independent predictor of poor survival in diagnostic needle biopsies from prostate cancer suspects. Clin Cancer Res 2006, 12:3961–3970

40. Saramaki O, Visakorpi T: Chromosomal aberrations in prostate cancer. Front Biosci 2007, 12:3287–3301

41. Porkka K, Saramaki O, Tanner M, Visakorpi T: Amplification and overexpression of elongin C gene discovered in prostate cancer by cDNA microarrays. Lab Invest 2002, 82:629–637

42. Jalava SE, Porkka KP, Rauhala HE, Isotalo J, Tammela TL, Visakorpi T: TCEB1 promotes invasion of prostate cancer cells. Int J Cancer 2009, 124:95–102