The American Journal of Pathology, Vol. 181, No. 1, July 2012

http://dx.doi.org/10.1016/j.ajpath.2012.03.043

Biomarkers, Genomics, Proteomics, and Gene Regulation

Shared Gene Expression Alterations in Prostate Cancer and Histologically Benign Prostate from Patients with Prostate Cancer

Farhad Kosari,* John C. Cheville,† Cristiane M. Ida,† R. Jeffrey Karnes,‡ Alexey A. Leontovich,§ Thomas J. Sebo,† Sibel Erdogan,† Erika Rodriguez,¶ Stephen J. Murphy,* and George Vasmatzis*

From the Departments of Molecular Medicine, Laboratory Medicine and Pathology,† and Urology,‡ Mayo Clinic, Rochester, Minnesota; and the Departments of Health Sciences Research§ and Pathology,¶ Johns Hopkins University, Baltimore, Maryland*

Prostate cancer (PCa) field effect alterations provide important clues regarding the initiation of these tumors and suggest targets for prevention or biomarkers for early detection. However, biomarkers of PCa field effects that have passed independent validation are lacking, largely because these alterations are subtle and difficult to distinguish from unrelated small changes in gene expression. We hypothesized that shared expression alterations in PCa and benign prostates containing PCa (BPCs) would have a higher potential for independent validation than alterations identified in BPCs alone. Expression analyses were performed on 37 PCas and 36 unmatched BPCs and were contrasted with 28 benign prostates (BPs) from patients free of PCa. Most of the protein-coding genes and nonexonic RNAs selected according to the hypothesis were validated by quantitative RT-PCR in an independent set of 51 BPCs and BPs. A statistical model based on two markers distinguished BPCs from BPs in the RT-PCR set and in an external microarray (area under the curve = 0.84 and 0.90, respectively). In addition, genes with predominant expression in stroma were identified by expression profiling of pure stroma and epithelial cells. Pathway analysis identified dysregulated platelet-derived growth factor receptor signaling in BPC stroma. These results validate our approach for finding PCa field effect alterations and demonstrate a PCa transcriptome fingerprint in nonneoplastic cells in prostates containing cancer. ***(Am J Pathol 2012, 181:34–42; http://dx.doi.org/10.1016/j.ajpath.2012.03.043)***

The diagnosis of prostate cancer (PCa) is based mainly on needle biopsy of the prostate gland. However, needle biopsy has a 30% false-negative rate due to sampling error.[1,2] As a result, many of the approximately 800,000 men with a negative biopsy result in the United States each year undergo repeated biopsies, which can be frustrating for patients and urologists. For benign prostate needle biopsy specimens that lack atypical small acini or high-grade prostatic intraepithelial neoplasia, there is no additional information that can be gained by pathologic assessment. However, the possibility that these prostate glands harbor PCa is significant. Despite a lack of morphologic changes, there is a considerable body of evidence suggesting that molecular alterations associated with tumor in adjacent nonneoplastic cells, the so-called tumor field effect, can provide valuable clues regarding the presence of tumor. The prostatic tumor field effect was first reported 15 years ago based on subtle histologic changes in the tissue architecture and cytologic features in benign tissue adjacent to and at some distance from PCa.[3] Subsequent studies have documented tumor-associated molecular alterations in nonneoplastic tissue adjacent to PCa in resected specimens and, notably, in needle biopsy specimens.[4,5] Recently, several investigators have used microarrays to identify expression alterations associated with PCa field effects.[6–9] These profiles were often independent of PCa and included limited or no independent validation.

In this study, we sought to identify field effect alterations that were shared in benign prostates from patients with PCa (BPCs) and unmatched PCa. We posited that

Supported by a grant from Bernard and Edith Waterman Award in The Center of Individualized Medicine and by a Developmental Award from the Prostate Specialized Program of Research Excellence in Prostate Cancer at Mayo Clinic.

Accepted for publication March 6, 2012.

Supplemental material for this article can be found at *http://ajp.amjpathol.org* or at *http://dx.doi.org/10.1016/j.ajpath.2012.03.043*.

Address reprint requests to Farhad Kosari, Ph.D., John C. Cheville, M.D., or George Vasmatzis, Ph.D., Mayo Clinic, 200 First St. SW, Rochester, MN 55905. E-mail: kosari.farhad@mayo.edu, cheville.john@mayo.edu, or vasm@mayo.edu.

this strategy would identify expression changes with a better possibility of independent validation than changes identified in BPCs alone. In addition, gene expression changes specific to benign prostate tissues, such as changes related to hyperplasia, atrophy, postatrophic hyperplasia, etc, would be excluded. Although the definition of field effect varies among investigators, we defined the presence of alterations in the entire peripheral zone of the prostate gland as field effect instead of changes restricted to benign glands just adjacent to cancer. In addition, this study focused on high-grade PCa (Gleason score ≥7), as these tumors are much more likely to be clinically significant and more detrimental to the patient if missed on the needle biopsy specimen. Significant gene expression alterations in BPCs compared with in benign prostate glands from patients free of PCa who had their prostates resected during cystoprostatectomy for bladder cancer (BPs) were identified and validated. These expression alterations are referred to as markers in this article, and they refer to genes, pseudogenes, and transcribed non-exonic sequences. A logistic regression model was developed and tested in two independent data sets to determine the presence of PCa based on field effect alterations in BPCs. We also studied dysregulated pathways in BPC stroma by gene set enrichment analysis. The results of this study indicate the presence of a wide field effect in PCa and have important implications in developing assays aimed at improving the diagnosis and management of patients at risk for PCa.

Materials and Methods

Patient Samples

BP samples were from cystoprostatectomy procedures performed at Mayo Clinic (Rochester, MN) between January 1, 2007, and December 31, 2011, on patients with bladder cancer. Patients identified with incidental PCa in their surgical records were excluded. BPC and PCa samples were obtained from the Mayo Clinic Specialized Program of Research Excellence in Prostate Cancer tumor bank. All patients with PCa and BPCs underwent radical prostatectomy, and none had received preoperative hormone therapy, chemotherapy, or radiotherapy. PCa samples were from patients with a tumor Gleason score ≥7 and were independent of BPC samples. The discovery step by microarray expression profiling included BPs (n = 28), BPCs (n = 36), and PCa (n = 37). In the BPC group, there were 16 samples from patients with a Gleason score of 6 and 20 samples from patients with a Gleason score ≥8. Quantitative RT-PCR (RT-qPCR) experiments included BPs (n = 15) and BPCs (n = 23) of the discovery samples for confirmation and an independent set of BPs (n = 18) and BPCs (n = 33) for validation. The BPC validation set included 15 samples from patients with a Gleason score of 6 and 18 samples from patients with a Gleason score ≥8. Finally, 16 independent benign prostates from radical prostatectomy and cystoprostatectomy operations were used in the laser capture microdissection (LCM) expression profiling to identify stroma-related genes in pathway analysis.

Collection and Processing of Bulk Samples

Prostate tissue specimens were surgically excised and submitted for surgical pathology analysis. All the specimens were evaluated rapidly after removal from the patient. In each case, most of the peripheral zone was examined by rapid frozen section, and after diagnostic material was obtained, sections were collected for research. Fresh cancer and benign tissue specimens (both up to 2 cm in greatest dimension) were flash frozen in liquid nitrogen. In cases of PCa, the benign prostatic tissue was obtained from the peripheral zone on the opposite side. At the time of collection, the distance from main tumor mass to benign ranged from 0.5 to 2 cm. Collection of BP samples included the peripheral zone of prostates only. For all the samples, an H&E-stained section was prepared from the snap-frozen tissue to ensure that the tissue was appropriately identified and of adequate quality. A careful pathologic examination excluded cystoprostatectomy cases with lesions suspicious of incidental prostate or bladder cancers on the tissue slides. In cases where incidental microscopic foci of cancer were present in BPCs, these foci were cut out of the frozen tissue block by manual macrodissection. Similarly, PCa samples were enriched for the tumor content by macrodissecting nonneoplastic tissues. Samples with significant inflammation and high-grade prostatic intraepithelial neoplasia were excluded. Frozen tissues were sectioned and collected in a cryostat and kept frozen (−80°C) until RNA isolation. H&E-stained slides from before and after every 10 sections and at the end of sectioning were inspected to ensure that no tumor cells contaminated BP and BPC samples.

RNA Isolation and Microarray Expression Profiling of Bulk Samples

Total RNA from each sample was isolated using standard kits (Qiagen Inc., Valencia, CA). DNase treatment was performed on the RNA purification column to remove genomic DNA contamination (RNase-free DNase set; Qiagen Inc.). RNA quantity and quality were measured by spectrophotometry and by Agilent 2100 bioanalyzer (Agilent Technologies Inc., Santa Clara, CA). There was no difference in the RNA quality assessed by the RNA integrity number between the BP and BPC samples (see Supplemental Figure S1 at *http://ajp.amjpathol.org*). Labeling and hybridization to the Affymetrix U133Plus2 chips (Affymetrix Corp., Santa Clara, CA) followed standard protocols and used 1.0 to 1.2 μg of total RNA.

Bioinformatics and Statistical Analysis

The microarray signal intensity (.cel) files from Mayo samples were normalized and processed by the gcrma package in R version 2.11.1 (*http://cran.at.r-project.org*). As an additional measure of quality control for contamination of

prostate tumor cells in the BPC, expression levels of ERG, ETV1, ETV4, and SPINK1 were compared between BPCs and BPs. Together, these four genes are overexpressed in most prostate tumors. We did not see any evidence of overexpression of these genes in any of the BPC samples (data not shown).

Sammon Mapping

Clustering analysis used Sammon mapping in the "MASS" library in R. Distance measures were calculated by the "manhattan" method in the "stat" library.

Selection of Up-Regulated Genes

Two metrics used for gene selection were signal-to-noise ratio (SNR) and *P* values by *t*-tests. SNRs were calculated as $SNR = (\mu_1 - \mu_2)/(\sigma_1 + \sigma_2)$, where μ's were mean expression values and σ's were a maximum of $0.2 \times \mu$ and SD.[10] We also required that the average expression in samples overexpressing a gene had >3.5 $\log_2$ intensity. $\log_2$ expression intensities for the gcrma-normalized data ranged from 1 to 16.5. Based on our experience with RT-qPCR, gene expression intensities <3.5 were not reliable and often were not detected. Seven genes with significant q values for multiple comparisons between BPCs and BPs ($q < 0.05$) and *AMACR* ($q = 0.06$) based on the literature[11,12] were selected for validation by RT-PCR.

Permutation Analysis

The central objective of the analysis was to determine whether the results obtained by using the BPC samples could also be obtained by combining and randomly shuffling the BP and BPC samples. We examined the overlap between overexpressed genes compared with BPs in BPCs and in PCa. In the list of 270 to 285 unique genes (350 probe sets) with the highest SNRs comparing BPCs and PCa with BPs, there were 19 overlapping genes and expressed sequence tags (21 probe sets) with significant *P* values ($P < 0.05$). To determine the likelihood of this overlap, we assigned a randomly selected 28 of the BP and BPC samples into one group (group 1) and the remaining 36 into another group (group 2) and identified the overlap in 350 probe sets with the highest SNRs and significant ($P < 0.05$) overexpression in group 2 and PCa compared with group 1. This process was repeated 1000 times, and each time the number of overlapping probe sets was recorded to generate a histogram.

Selection of Down-Regulated Genes

Probe sets ($n = 50$) with the lowest SNRs that were significant ($P \leq 0.05$) in BPCs and PCa compared with BPs were considered. These probe sets were also significant in BPCs after multiple comparison correction ($q < 10^{-5}$). Five probe sets corresponding to misspliced or nonexonic RNA were selected for RT-PCR analysis.

Confirmation and Validation by RT-qPCR

Total RNA (125 to 500 ng) was used in reverse transcription using SuperScript III (Invitrogen, Carlsbad, CA). Affymetrix probe set target sequences were used to design primers on the Primer3 website (*http://frodo.wi.mit.edu/primer3*, last accessed February 1, 2011). All qPCRs were performed in duplicates using a Dynamic Array integrated fluidic circuit (Fluidigm Corp., San Francisco, CA). Each primer set was tested by a 4-point standard curve. Five genes, including ACTB, were used to normalize the data. The other four normalizer genes (DUS2L, EIF2B1, LYK5, and NUDC) were identified based on small standard deviations in the BP and BPC samples in the microarray data. Table 1 includes the primer sequences. Reported values were calculated as $reference_x$

Table 1. Primers Used in RT-qPCR Experiments

Probe set	Marker	Forward primer	Reverse primer	AmL
Target genes				
201291_s_at	*TOP2A*	5′-AGATTCTGGACCAACCTTCAAC-3′	5′-GCCTGCAGAGTTCATCTTTCTT-3′	83
204750_s_at	*DSC2*	5′-CGCGATCTTAATATTTGCCAGT-3′	5′-TTTCTCGGCATCTAGTTTGGAG-3′	76
209631_s_at	*GPR37*	5′-AACAAATAAATCTGACCCAACCAA-3′	5′-ATACGCCGTGAAATGTCCACT-3′	82
204324_s_at	*GOLIM4*	5′-TGTGATGTTGGAAAGCTCATTG-3′	5′-AACAAAGAACACCTGGGAACTG-3′	81
216867_s_at	*PDGFA*	5′-TCGGGAGAACAAAGAGACAGTG-3′	5′-TACTGCTTCACCGAGTGCTACA-3′	77
213668_s_at	*SOX4*	5′-ACTTCGAGTTCCCGGACTACTG-3′	5′-CAGGTTGGAGATGCTGGACTC-3′	83
228729_at	*CCNB1*	5′-AATGGTGAATGGACACCAACTC-3′	5′-ATTCTTAGCCAGGTGCTGCATA-3′	87
209426_s_at	*AMACR*	5′-CACGTGAAACAGAGTGATTGGT-3′	5′-TGGAATGTGCTTAGAGGGAGAT-3′	75
210469_at	*DLG5*	5′-GAGAAGCCCGCACTTTCTACAT-3′	5′-TGCAATCTGAACACCTGACTTG-3′	76
222018_at	*NACA*	5′-CCTTTGTTCCTTGACTCCCTCT-3′	5′-TGGAATGAGGTTCCTTAATTGG-3′	91
226848_at	*NR2C2*	5′-CAGATGTGTTCCCTTCACTCTTG-3′	5′-CCTCTGTTGATGAATTTCCAGGT-3′	84
227751_at	*PDCD5*	5′-GGGAGAAAGGCTGAATCTGTTG-3′	5′-AAAGGGTGGGAATGGAGTCA-3′	70
1556026_at	*IDSP1*	5′-GGTGGATTTGAGAAGATGTGGA-3′	5′-ACACTACAGCATTCAGGGTTCC-3′	84
Normalization genes				
47105_at	*DUS2L*	5′-CCATCAGCCTAGAGCATGGAC-3′	5′-CCTGTCACTCAGATCCACCAA-3′	74
201632_at	*EIF2B1*	5′-GAATCTGCATCTGGCTTCCA-3′	5′-CCTTCCTCGTGTTCTCTTGGTA-3′	87
52169_at	*STRADA*	5′-GCGACCAGCCTCATTCTATTTA-3′	5′-GGCAGCTTACTACTTGCCCTTT-3′	82
210574_s_at	*NUDC*	5′-AGATGATGTATGACCAGCGACA-3′	5′-AATCTCCTGTTTCTTCTGTTCGTC-3′	71
213867_x_at	*ACTB*	5′-TCCTCTCCCAAGTCCACACA-3′	5′-GCACGAAGGCTCATCATTCA-3′	129

AmL, amplicon length.

– $marker_{g,x}$ + 35, where $reference_x$ is the average of the five housekeeping genes in sample *x* and $marker_{g,x}$ is the raw expression for marker g in sample *x*.

Logistic Regression

RT-qPCR data on the confirmation and validation sets were used to build and test a logistic regression model. To minimize overfitting, model building was limited to two markers. The first parameter to enter the model in the confirmation set was NACA, followed by CCNB1. These two markers produced the highest area under the curve (AUC) in the validation set also. For additional testing, we used the public data set of Wang et al[13] (Gene Expression Omnibus record GSE17951). Expression data were extracted from the signal intensity files (.cel files) by the gcrma package in R. The model was first built in the Mayo Clinic microarray data set, and the model coefficients for the two markers were applied to the data set of Wang et al[13] (GSE17951). In this data set, benign samples with 0% tumor from patients with PCa according to the sample information file were categorized as BPCs, and organ donor samples were categorized as BPs. The study by Wang et al focused on the expression profiles of prostate stroma adjacent to tumor. By inspecting the prostate-specific antigen expression levels, we identified seven samples in this data set with low epithelial content (see Supplemental Figure S2 at *http://ajp.amjpathol.org*). The logistic regression model was tested with and without these samples.

Expression Profiling of the Prostate Stroma and Epithelia

LCM was used to collect 11 epithelial cell populations from cystoprostatectomy operations and 5 stromal samples (4 matching epithelial collections and 1 from a radical prostatectomy operation). Close to 5000 LCM pulses were used for each sample. Total RNA was isolated using the PicoPure kit (Arcturus Corp., Mountain View, CA). The quality and quantity of RNA from the LCM samples were controlled by a qPCR assay based on the ratio of concentration of 3′ to middle transcript of β-actin. Details of this procedure and for sectioning and processing of slides, LCM, and linear amplification are described previously.[14] Briefly, this protocol required 2 ng of starting total RNA and used a slightly modified version of the Affymetrix small sample preparation protocol (Affymetrix Corp.). Biotin-labeled amplified cDNA samples were hybridized to the U133Plus2 chips according to the standard Affymetrix protocols.

Pathway Analysis

We aimed to identify enriched gene ontology (GO) categories in BPC stroma. Genes differentially expressed in BPC stroma compared with BP stroma were identified in a two-step process. The first step selected probe sets that were significantly up- or down-regulated in BPCs compared with in BPs after adjusting for multiple comparisons[15] ($q < 0.05$) in the bulk microarray data. The second step selected probe sets with more prominent expression in prostate stroma than in prostate epithelia cells in the LCM data set. Probe sets that were significantly higher in stroma ($q < 0.05$) with a fold change ≥2 were selected. Steps 1 and 2 had 278 common probe sets representing 218 unique genes.

Expression values of these probe sets in the bulk data set were input to the Matisse software package (Matisse Software Inc., Mountain View, CA) to identify enriched GO categories. The DEGAS (Dysregulated Gene Set Analysis via Subnetworks) algorithm was used to build a network of genes dysregulated in BPCs.[16] Other than the gene expression data, DEGAS requires a protein interaction network. We used the 2007 human functional linkage network (Human_Interactome_May.sif).[17] Probe sets/genes with twofold differential expression compared with BPs in a subset of 31 or more BPCs were selected to build a network (I = 5 in DEGAS). Identified GO categories in the network were tested by the TANGO module in the Matisse software package for statistical significance after correcting for multiple comparisons.

Results

No Evidence of Systematic Differences in the Expression Profiles of Benign Samples from Patients with and without PCa

Significant disparities in ischemia time or tissue processing between benign samples from patients with (BPCs) and without (BPs) PCa can confound the influence of PCa field effect on gene expression. Such disparities can be readily detected by a clustering analysis, such as Sammon mappings. We selected the most variable probe sets in the expression profiles of BPCs and BPs. Sammon mapping by these probe sets did not separate BPs from BPCs (Figure 1A). The two groups were intermixed and did not separate into discrete groups. Changing the number of selected probe sets did not alter the patterns. We also examined whether there were major influences of ischemia in differential gene expression profiles. Two publications by Lin et al[18] and Dash et al[19] describe the

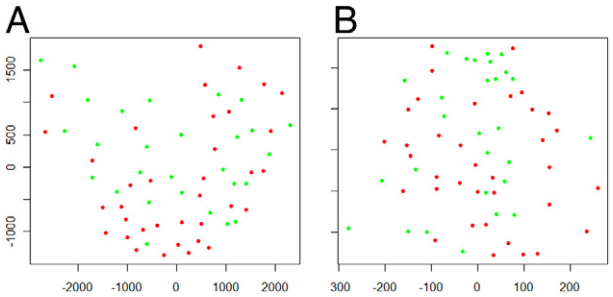

Figure 1. BPs and BPCs do not show expression differences due to disparities in tissue processing or ischemia. **A:** Sammon map of the most variable probe sets (n = 350) in BPs (green dots) and BPCs (red dots) does not separate these two groups. **B:** Sammon map that uses the most variable probe sets corresponding to the ischemia-related genes by Lin et al[18] does not separate the BP and BPC samples.

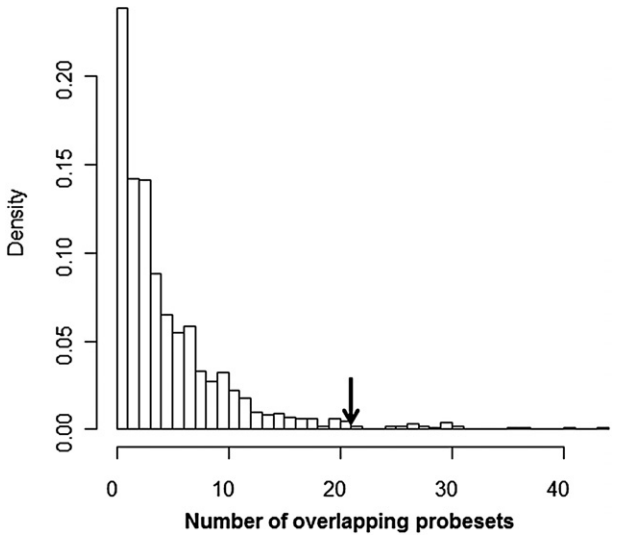

Figure 2. Permutation analysis indicates significant overlap between BPCs and PCa in overexpressed probe sets with the highest SNRs compared with BPs. The histogram was generated by random shuffling of BPC and BP class labels (see *Materials and Methods*). The **arrow** points to the number of overlapping probe sets with correct BP and BPC labels.

significant gene expression alterations due to ischemia during and after prostate operations, respectively. A Sammon map that used the most variable or all probe sets corresponding to the significant genes by Lin et al (Figure 2 in the referenced article) did not separate BPCs from BPs (Figure 1B; also see Supplemental Figure S3A at *http://ajp.amjpathol.org*). Similarly, probe sets corresponding to the genes reported by Dash et al did not separate the two groups (see Supplemental Figure S3, B and C, at *http://ajp.amjpathol.org*). Of the 293 probe sets corresponding to the 98 ischemia genes in both studies, only 7 (~2%) were overexpressed in BPCs compared with BPs ($q < 0.05$). A comparable number of ischemia genes/probe sets were down-regulated in BPCs compared with in BPs. This indicates a small and balanced change in ischemia-related genes in BPs and BPCs.

Evidence of Cancer from the Most Up-Regulated Markers in the Benign Prostate Glands of Patients with PCa

We examined gene expression alterations that overlapped between BPCs and PCa. SNR, a metric designed to identify genes with higher changes between two groups than within two groups,[10] was used to select genes. BPCs and PCa were each compared with BPs, and genes were ranked by SNR. Within the top 270 to 285 genes (350 probe sets) in the two comparisons, there were 21 overlapping probe sets with a significant ($P < 0.05$) increase in BPCs and PCa compared with BPs (see Supplemental Table S1 at *http://ajp.amjpathol.org*[15]). To determine the significance of this overlap, we performed 1000 permutations of BP and BPC sample labels and each time recorded the number of overlapping probe sets using the same criteria (see *Materials and Methods*). Figure 2 is a histogram of the number of overlapping probe sets found. The median and mean numbers of overlapping probe sets were three and five, respectively. Interestingly, an overlap of 21 probe sets was observed in the top 97th percentile, indicating that expression differences between BPs and BPCs were most likely related to their categorization.

It is important to note that the BPCs and PCa were from independent samples and not from matched normal tumor pairs. Therefore, the significant overlaps in the top overexpressed genes between the two categories cannot be attributed to a common background between tumor and adjacent benign prostate tissue samples. Also, it is noteworthy that based on collection of the tissue and review of the H&E slides, collection of BPC in most cases was at a significant distance from tumor, ranging from 1 to 2 cm. However, we cannot entirely exclude the possibility that some tumor may have been close to the collected tissue in the three-dimensional space of the prostate gland.

Expression changes in BPCs were generally smaller in magnitude than in PCa, as shown in Figure 3 for SOX4, CCNB1, and GPR37. The subtle but detectable expression changes were manifestations of the tumor field effect in the benign prostate glands. From the list of up-regulated candidates, seven known genes with the highest SNRs and $q \leq 0.05$ (TOP2A, DSC2, GPR37, GOLIM4, PDGFA, SOX4, and CCNB1) were selected for validation by RT-qPCR. We also included AMACR in the validation experiments as previous studies indicated altered expression of AMACR by the PCa field effect.[11,12]

Down-Regulated Markers in BPC Include a High Percentage of Probe Sets Containing Nonexonic Sequences

The 50 most down-regulated probe sets in BPCs compared with in BPs, based on the SNR, were selected (see Supplemental Table S2 at *http://ajp.amjpathol.org*). Each

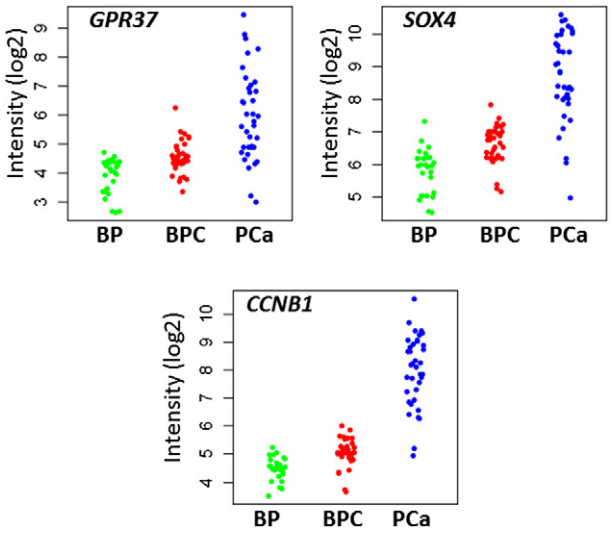

Figure 3. Expression changes compared with BPs in BPCs are smaller than in PCa. Microarray data are depicted for three representative genes (*GPR37*, *SOX4*, and *CCNB1*).

Table 2. RT-qPCR Confirmation and Validation Data

Marker	p-g-cnf*	p-g-val*	Confirmed†	Trend‡
Up-regulated markers				
GPR37	**0.018**	**0.03**	Yes	Yes
SOX4	**0.003**	**0.03**	Yes	Yes
CCNB1	**0.002**	**0.007**	Yes	Yes
GOLIM4	0.173	**0.052**	No	Yes
TOP2A	**0.018**	0.267	Yes	Yes
AMACR	0.093	0.1	No	Yes
DSC2	0.398	0.476	No	Yes
PDGFA	0.255	0.716	No	No
Down-regulated markers				
DLG5	**< 0.001**	**0.004**	Yes	Yes
NACA	**< 0.001**	**0.003**	Yes	Yes
NR2C2	**< 0.001**	**0.02**	Yes	Yes
PDCD5	**< 0.001**	0.063	Yes	Yes
IDSP1	**< 0.001**	0.254	Yes	Yes

Significant values are shown in bold ($P \leq 0.05$). PDCD5 was marginal ($P = 0.063$). All but one marker (PDGFA) had the expected trend based on the microarray data in the confirmation and validation sets.

*p-g-cnf and p-g-val are *t*-test *P* values with the null hypothesis that expression in BPCs is not higher than that in BPs in the confirmation and validation sets, respectively.

†Confirmed indicates whether the selected marker was significant in the confirmation set.

‡Trend indicates whether BPC average expression was higher than that of BP for up-regulated markers and lower than that of BP for down-regulated markers.

§ p-ls-cnf and p-ls-val are *t*-test *P* values with the null hypothesis that expression in BPCs is not less than that in BPs in the confirmation and validation sets, respectively.

of these probe sets was down-regulated in BPCs by at least twofold, and this down-regulation was statistically significant after adjusting for multiple comparisons by false discovery rate[15] ($q < 5 \times 10^{-5}$). More than 70% of the probe sets were also down-regulated in PCa compared with in BPs ($P \leq 0.05$). BLAT (BLAST-like alignment tool) searches in the RefSeq database revealed that >50% of the probe sets (26 of 50) had target sequences containing misspliced or nonexonic sequences. In contrast, the up-regulated list (see Supplemental Table S1 at *http://ajp.amjpathol.org*) contained <25% of such probe sets. The significance of transcribed nonexonic sequences is currently undetermined, but they could represent long noncoding RNA important for cancer initiation or progression[20] or new gene variants. We selected five of these probe sets that were concomitantly significantly down-regulated in BPCs and PCa for validation by RT-qPCR, including *IDSP1* pseudogene and probe sets corresponding to *NACA*, *NR2C2*, *PDCD5*, and *DLG5* loci.

RT-qPCR Confirmation and Validation of Selected Up- and Down-Regulated Markers in BPCs

Real-time RT-qPCR was used to confirm the findings in the discovery step and to validate in independent samples. Confirmation used a portion of the microarray samples (see Materials and Methods). Except for PDGFA, the average expression levels of all markers in BPC and BP agreed with the expected trend based on the microarray data (Table 2). GPR37, SOX4, and CCNB1 were among the up-regulated genes that were confirmed and validated in the independent samples ($P \leq 0.05$). GOLIM4 was validated in the independent set ($P = 0.052$), although the lower expression of this gene in the confirmation set was not statistically significant ($P = 0.173$). DLG5, NACA, and NR2C2 were confirmed and also significantly down-regulated in independent samples, whereas PDCD5 was marginal (P = 0.063). Figure 4 is a box plot of three up- and three down-regulated markers in BPCs compared with BPs.

A Statistical Model Based on Two Markers Stratified BPCs from BPs

The ability of the markers to distinguish BPCs from BPs in the RT-qPCR data was examined by a logistic regression that included two markers. In the confirmation and validation sets, a model that included NACA and CCNB1 produced the maximum AUC in the receiver operating characteristic (ROC) plot (see *Materials and Methods*). Figure 5A is the ROC plot in the RT-qPCR validation set with an AUC of 0.84. The predictive ability of this model was also examined in the public microarray data set of Wang et al[13] (GSE17951). Model coefficients were first computed in the Mayo Clinic microarray data and then applied to the Wang et al data set. The study by Wang et al[13] focused on the expression profiles of stroma near the tumor, and samples were collected by macrodissection. We identified few samples with low epithelial content

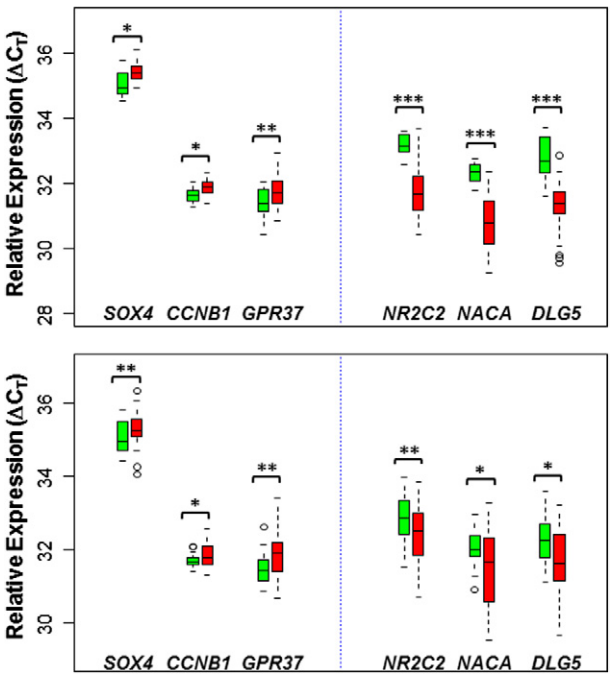

Figure 4. Box plot of six genes validated by RT-qPCR. Three up-regulated markers (GPR37, CCNB1, and SOX4) and three down-regulated markers (NR2C2, NACA, and DLG5) in BPCs compared with BPs are shown. The **top** and **bottom panels** show the results of the confirmation and validation sets, respectively. Green bars represent BP samples; red bars, BPC samples. The horizontal line in the middle of each box indicates the median, and the top and bottom borders of the box mark the 75th and 25th percentiles, respectively. The whiskers above and below the box mark 1.5× the interquartile range. The points beyond the whiskers are outliers beyond 1.5× the interquartile range. *$P \leq 0.05$, **$P < 0.01$, and ***$P < 0.001$.

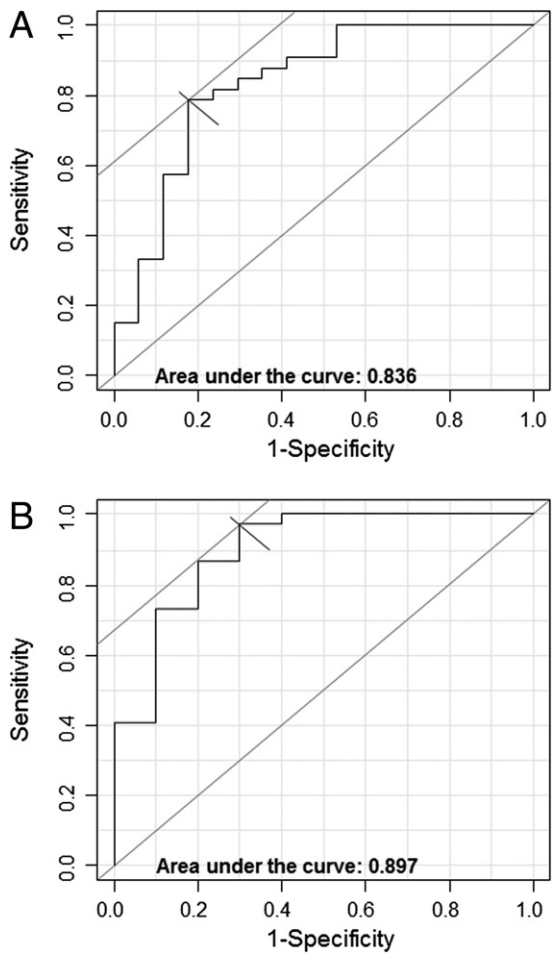

Figure 5. A: ROC plot of the logistic regression model in the RT-qPCR validation set. The model included NACA and CCNB1 (AUC = 0.84). **B:** ROC plot of the logistic regression model in the external microarray set (GSE17951) of Wang et al[13] after excluding samples with low epithelial content (see *Materials and Methods*) (AUC = 0.90).

based on prostate-specific antigen expression (see *Materials and Methods*; see also Supplemental Figure S2 at *http://ajp.amjpathol.org*). Figure 5B is the ROC plot for the model in the Wang et al data set after eliminating samples with low epithelial content. This model had an AUC of 0.90. The AUC of the model that included all the samples, irrespective of prostate-specific antigen expression, in the Wang et al database was 0.89. These results suggest that the field effect markers can discriminate between BPs and BPCs with high accuracy.

Enriched GO Categories in BPC Stroma

Recent studies have highlighted the influence of stroma in PCa behavior and survival.[21] These studies have often focused on reactive stroma immediately adjacent to tumor. We were interested in finding whether in the microarray data from bulk samples we could detect dysregulated pathways in prostate stroma at large distances from prostate tumors. We first identified genes with prominent expression in stroma by analyzing expression data from pure stroma and epithelial cell populations collected by LCM (see *Materials and Methods*). Among the identified genes, there were 218 unique genes that were also differentially expressed between BPCs and BPs in the bulk microarray data. We argued that most of the differential expression occurred in stroma as the expression levels of these genes in stroma were on average approximately fivefold higher than in epithelia. These genes were examined by the Matisse software package.[17] The DEGAS module in the package identified a network of 91 genes that included 47 of 218 unique genes (~22%) in our list (see Supplemental Figure S4 at *http://ajp.amjpathol.org*). Enriched GO categories in the network were examined by the TANGO module for multiple comparison correction. Three GO categories remained significant (Table 3). Platelet-derived growth factor (PDGF) receptor signaling is one of the most notable carcinogenesis-related categories on the list. Also, enrichment of the "regulation of epithelial cell differentiation" category in BPC stroma was unexpected. Although the number of genes limited this analysis, it was possible to identify the cancer fingerprint in the enriched GO categories in BPC stroma.

Discussion

In this study, we observed a common cancer transcriptome between histologically benign prostate tissue adjacent to PCa and PCa. A logistic regression model that included down- and up-regulated markers in BPCs compared with BPs predicted the presence of tumor in the prostate gland with high accuracy in independent RT-qPCR samples (AUC = 0.84) and in an external microarray data set from Wang et al[13] (AUC = 0.90). These findings provide strong evidence that the transcriptome profiles of the benign prostatic tissue can indicate the presence or absence of PCa. We believe that using independent PCa expression profiling to guide the selection process allowed a greater likelihood of identifying markers that would validate.

We did not observe large-scale or systematic differences in the expression profiles of BPCs and BPs. Potential changes in gene expression due to ischemia were balanced in the BP and BPC samples (Figure 1). We expect that this partly stems from protocols followed at the Mayo Clinic, where a unique rapid grossing, frozen section, and tissue preservation strategy after removal of tissue from the body is followed that significantly minimizes ischemia time and reduces variation in ischemia time. Furthermore, previous studies did not identify a significant association between ischemia time and total operating time with the magnitude of the gene expression changes in ischemia.[18] Therefore, even if some differ-

Table 3. GO Categories Enriched in the Network Identified by DEGAS

Enriched GO category	*P* value*
PDGF receptor signaling	0.028
Regulation of epithelial cell differentiation	0.017
Peptidyl-tyrosine phosphorylation	0.014

**P* value corrected for multiple comparisons by TANGO.

ences in ischemia time existed, they would not have significantly impacted the findings.

Some of the overlapping gene alterations between BPCs and PCa are noteworthy. Cyclin B1 (CCNB1) is essential for control of the cell cycle at the G2/M transition and, in complex with CDC2, is part of the master mitotic regulator. Compounds such as BZL101, which inhibit PCa growth, block CCNB1-related pathways.[22] Increased expression of CCNB1 in BPCs found in the present experiments may suggest that a small but detectable group of morphologically benign prostate cells from patients with PCa incurred compromised cell cycle regulatory mechanisms. Sex-determining region Y-box 4 (SOX4) is involved in the regulation of embryonic development. There is evidence that in PCa, SOX4 is involved in the cross talk between Wnt, Notch, and PI3K pathways.[23] The influence of GPR37 and GOLIM4 in PCa is currently unknown. We also validated down-regulation of nonexonic or misspliced markers that potentially code for long (>200-nucleotide) noncoding RNA or novel transcript variants. Recently dubbed as human cell "dark matter" for being a major contributor to nonribosomal RNA, long noncoding RNAs are emerging as important contributors to tumorigenesis.[20,24] Further research is required to determine whether the down-regulated markers represent long noncoding RNA or previously unknown transcripts and to determine their role in the initiation or progression of PCa.

Other investigators have used expression profiling to identify PCa field effects. Chandran et al[6] used an earlier version of the Affymetrix platform (U95A) for discovery of gene expression changes between tumor-adjacent nonneoplastic prostate glands and benign prostate glands from organ donors. This study found increased expression of proliferation-related genes; however, these findings were not independently validated. Haaland et al[7] used a sample of six tumor and adjacent normal pairs and six benign prostates from organ donors for profiling on a Qiagen platform containing 37,123 transcripts. They reported overexpression of FAS, TTP, EGR1, and SPOCK1, but their validation set included six BPC–PCa pairs and no independent organ donor prostate samples. Schlomm et al[25] selected 11 markers from expression profiling and 18 markers by searching in the Oncomine database and tested them by RT-qPCR on a set of 114 biopsy specimens from 3 risk groups. Only one of the microarray markers (FOS) was significant. In the list of markers selected from Oncomine, EGR1 was down-regulated and MYC, TFRC, and FOLH1 were up-regulated in negative biopsy specimens from patients who were later identified with PCa in a follow-up biopsy. Although this study included a relatively large sample of clinically relevant biopsies, further multi-institutional investigations are required before these results can be applied to the clinic. Finally, a study by Risk et al[8] analyzed expression profiles of prostate biopsies from patients with and without PCa. This study was unique in profiling pure populations of epithelial cells collected by LCM. The investigators reported differential expression of ERG, HOXC4, HOXC5, FOLH1, and MME confirmed by qPCR, two of which (FOLH1 and MME) were also confirmed by immunohistochemical analysis. However, this study did not include independent validation.

There is no high concordance between the biomarkers identified in the aforementioned studies or with the study presented herein for several reasons. Aside from different detection chemistry, discrepancies in gene loci target regions in different array platforms can be an important contributor. For example, the down-regulated nonexonic elements identified in this study are not featured on many other platforms. More specific to the study of PCa field effects is the consideration that the alterations in gene expression in the benign prostate tissue are generally of much lower magnitude than in PCa (Figure 3). Also, it is often difficult to acquire a sufficient number of "normal" samples (prostate glands without cancer), which has led to limited statistical power and a lack of validation experiments in previous studies. In the present study, we used prostate tumor expression profiles to guide the selection of markers that were more robust and had a reasonable prospect for validation in independent samples. In addition, we took advantage of the relatively high number of samples collected from cystoprostatectomy specimens at our institution to perform discovery and validation experiments.

An integrative pathway analysis by the Matisse software package, which included gene expression values and protein interactome data, was used to identify enriched GO categories in BPC stroma. Two of the three identified categories were "PDGF receptor signaling" and "regulation of epithelial cell differentiation." Overexpression of PDGFs and their receptors have been reported in many cancers, including PCa,[26] and PDGF receptors are being actively pursued as targeting agents in various tumors.[27] Also, paracrine tumor-stroma PDGF signaling has been reported.[28–30] By and large, these reports have focused on signaling events in stroma near the tumor. The present finding suggests that changes in the PDGF receptor pathway are present in wide areas of peripheral zones in prostates that harbor cancer. The additional identification of the category "regulation of epithelial cell differentiation" in BPC stroma is also intriguing. Whereas epithelial to mesenchymal transition has been actively researched in many solid tumors, the significance of dysregulated epithelial cell differentiation signaling in prostate stroma from patients with PCa is largely unknown. If validated in larger studies, these findings can provide important clues regarding the initiation of prostate tumors.

Potential limitations warrant discussion. As in all studies of this nature, the presence of incidental cancers in BP samples cannot be completely excluded. Although the resected prostate specimens from radical cystoprostatectomy specimens were thoroughly analyzed grossly and microscopically, the prostate gland was not entirely inspected to exclude the possibility of tiny incidental cancer. It is also important to note that we focused these analyses on PCa with a Gleason score ≥7. These tumors are clinically most relevant and, therefore, the most important to identify in men. Also, most of the BPC samples were from patients with tumors with Gleason scores ≥8. Further work is required to determine whether insignificant cancers with a lower grade (Gleason score of 6) can

be separated from high-grade cancer based on PCa field effect. Finally, for clinical utility, the present findings require validation in prostate needle biopsy specimens.

In conclusion, this study focused on transcriptomic alterations that occur in morphologically benign prostate tissue in prostate glands that harbor cancer. We confirmed the presence of field effect and that it occurs at some distance from prostate tumors. Validated field effect biomarkers can be valuable in prevention trials as a tool for choosing the correct intervention strategy or in clinical assays aimed at identifying men who potentially have PCa but whose prostate needle biopsy results are negative. Larger studies are warranted before these findings can be applied to clinical practice.

References

1. Patel AR, Jones JS, Rabets J, DeOreo G, Zippe CD: Parasagittal biopsies add minimal information in repeat saturation prostate biopsy. Urology 2004, 63:87–89
2. Stewart CS, Leibovich BC, Weaver AL, Lieber MM: Prostate cancer diagnosis using a saturation needle biopsy technique after previous negative sextant biopsies. J Urol 2001, 166:86–91; discussion 91–82
3. Montironi R, Diamanti L, Pomante R, Thompson D, Bartels PH: Subtle changes in benign tissue adjacent to prostate neoplasia detected with a Bayesian belief network. J Pathol 1997, 182:442–449
4. Dhir R, Vietmeier B, Arlotti J, Acquafondata M, Landsittel D, Masterson R, Getzenberg RH: Early identification of individuals with prostate cancer in negative biopsies. J Urol 2004, 171:1419–1423
5. Troyer DA, Lucia MS, de Bruine AP, Mendez-Meza R, Baldewijns MM, Dunscomb N, Van Engeland M, McAskill T, Bierau K, Louwagie J, Bigley JW: Prostate cancer detected by methylated gene markers in histopathologically cancer-negative tissues from men with subsequent positive biopsies. Cancer Epidemiol Biomarkers Prev 2009, 18:2717–2722
6. Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J: Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. BMC Cancer 2005, 5:45
7. Haaland CM, Heaphy CM, Butler KS, Fischer EG, Griffith JK, Bisoffi M: Differential gene expression in tumor adjacent histologically normal prostatic tissue indicates field cancerization. Int J Oncol 2009, 35: 537–546
8. Risk MC, Knudsen BS, Coleman I, Dumpit RF, Kristal AR, Lemeur N, Gentleman RC, True LD, Nelson PS, Lin DW: Differential gene expression in benign prostate epithelium of men with and without prostate cancer: evidence for a prostate cancer field effect. Clin Cancer Res 2010, 16:5414–5423
9. 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
10. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999, 286: 531–537
11. Leav I, McNeal JE, Ho SM, Jiang Z: α-Methylacyl-CoA racemase (P504S) expression in evolving carcinomas within benign prostatic hyperplasia and in cancers of the transition zone. Hum Pathol 2003, 34:228–233
12. Ananthanarayanan V, Deaton RJ, Yang XJ, Pins MR, Gann PH: α-Methylacyl-CoA racemase (AMACR) expression in normal prostatic glands and high-grade prostatic intraepithelial neoplasia (HGPIN): association with diagnosis of prostate cancer. Prostate 2005, 63:341–346
13. Wang Y, Xia XQ, Jia Z, Sawyers A, Yao H, Wang-Rodriquez J, Mercola D, McClelland M: In silico estimates of tissue components in surgical samples based on expression profiling data. Cancer Res 2010, 70:6448–6455
14. Kube DM, Savci-Heijink CD, Lamblin AF, Kosari F, Vasmatzis G, Cheville JC, Connelly DP, Klee GG: Optimization of laser capture microdissection and RNA amplification for gene expression profiling of prostate cancer. BMC Mol Biol 2007, 8:25
15. Storey JD, Tibshirani R: Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003, 100:9440–9445
16. Ulitsky I, Krishnamurthy A, Karp RM, Shamir R: DEGAS: de novo discovery of dysregulated pathways in human diseases. PLoS One 2010, 5:e13367
17. Ulitsky I, Shamir R: Identification of functional modules using network topology and high-throughput data. BMC Syst Biol 2007, 1:8
18. Lin DW, Coleman IM, Hawley S, Huang CY, Dumpit R, Gifford D, Kezele P, Hung H, Knudsen BS, Kristal AR, Nelson PS: Influence of surgical manipulation on prostate gene expression: implications for molecular correlates of treatment effects and disease prognosis. J Clin Oncol 2006, 24:3763–3770
19. Dash A, Maine IP, Varambally S, Shen R, Chinnaiyan AM, Rubin MA: Changes in differential gene expression because of warm ischemia time of radical prostatectomy specimens. Am J Pathol 2002, 161: 1743–1748
20. Gibb EA, Brown CJ, Lam WL: The functional role of long non-coding RNA in human carcinomas. Mol Cancer 2011, 10:38
21. Tammi RH, Kultti A, Kosma VM, Pirinen R, Auvinen P, Tammi MI: Hyaluronan in human tumors: pathobiological and prognostic messages from cell-associated and stromal hyaluronan. Semin Cancer Biol 2008, 18:288–295
22. Marconett CN, Morgenstern TJ, San Roman AK, Sundar SN, Singhal AK, Firestone GL: BZL101, a phytochemical extract from the Scutellaria barbata plant, disrupts proliferation of human breast and prostate cancer cells through distinct mechanisms dependent on the cancer cell phenotype. Cancer Biol Ther 2010, 10:397–405
23. Moreno CS: The Sex-determining region Y-box 4 and homeobox C6 transcriptional networks in prostate cancer progression: crosstalk with the Wnt, Notch, and PI3K pathways. Am J Pathol 2010, 176:518–527
24. Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen PH, Reaman G, Milos P, Arceci RJ, Thompson JF, Triche TJ: The majority of total nuclear-encoded non-ribosomal RNA in a human cell is "dark matter" un-annotated RNA. BMC Biol 2010, 8:149
25. Schlomm T, Hellwinkel OJ, Buness A, Ruschhaupt M, Lubke AM, Chun FK, Simon R, Budaus L, Erbersdobler A, Graefen M, Huland H, Poustka A, Sultmann H: Molecular cancer phenotype in normal prostate tissue. Eur Urol 2009, 55:885–890
26. Homsi J, Daud AI: Spectrum of activity and mechanism of action of VEGF/PDGF inhibitors. Cancer Control 2007, 14:285–294
27. Ostman A, Heldin CH: PDGF receptors as targets in tumor treatment. Adv Cancer Res 2007, 97:247–274
28. Skobe M, Fusenig NE: Tumorigenic conversion of immortal human keratinocytes through stromal cell activation. Proc Natl Acad Sci U S A 1998, 95:1050–1055
29. Lederle W, Stark HJ, Skobe M, Fusenig NE, Mueller MM: Platelet-derived growth factor-BB controls epithelial tumor phenotype by differential growth factor regulation in stromal cells. Am J Pathol 2006, 169:1767–1783
30. Anderberg C, Li H, Fredriksson L, Andrae J, Betsholtz C, Li X, Eriksson U, Pietras K: Paracrine signaling by platelet-derived growth factor-CC promotes tumor growth by recruitment of cancer-associated fibroblasts. Cancer Res 2009, 69:369–378