Two-Dimensional Transcriptome Profiling: Identification of Messenger RNA Isoform Signatures in Prostate Cancer from Archived Paraffin-Embedded Cancer Specimens

Hai-Ri Li,[1,6] Jessica Wang-Rodriguez,[2] T. Murlidharan Nair,[7] Joanne M. Yeakley,[5] Young-Soo Kwon,[1] Marina Bibikova,[5] Christina Zheng,[1,4] Lixin Zhou,[5] Kui Zhang,[1] Tracy Downs,[3] Xiang-Dong Fu,[1] and Jian-Bing Fan[5]

Departments of [1]Cellular and Molecular Medicine, [2]Pathology, and [3]Surgery, [4]San Diego Supercomputer Center, University of California, San Diego, La Jolla; [5]Illumina Inc., San Diego, California; [6]Department of Medicine, Mudanjiang Medical College, Mudanjiang, China; and [7]Departments of Biological Sciences and Computer Science/Informatics, Indiana University South Bend, South Bend, Indiana

Abstract

The expression of specific mRNA isoforms may uniquely reflect the biological state of a cell because it reflects the integrated outcome of both transcriptional and posttranscriptional regulation. In this study, we constructed a splicing array to examine ~1,500 mRNA isoforms from a panel of genes previously implicated in prostate cancer and identified a large number of cell type–specific mRNA isoforms. We also developed a novel "two-dimensional" profiling strategy to simultaneously quantify changes in splicing and transcript abundance; the results revealed extensive covariation between transcription and splicing in prostate cancer cells. Taking advantage of the ability of our technology to analyze RNA from formalin-fixed, paraffin-embedded tissues, we derived a specific set of mRNA isoform biomarkers for prostate cancer using independent panels of tissue samples for feature selection and cross-analysis. A number of cancer-specific splicing switch events were further validated by laser capture microdissection. Quantitative changes in transcription/RNA stability and qualitative differences in splicing ratio may thus be combined to characterize tumorigenic programs and signature mRNA isoforms may serve as unique biomarkers for tumor diagnosis and prognosis. (Cancer Res 2006; 66(8): 4079-88)

Introduction

Prostate cancer is a leading cause of morbidity and mortality among men in the U.S. (1). Although early diagnosis based on screening for prostate-specific antigen (PSA) has led to a decline in deaths and a decrease in the prevalence of advanced disease at the time of tumor diagnosis (2), disagreement and debate continue with regard to the efficacy of PSA screening and proper strategies for treatment after initial diagnosis (3, 4). New biomarkers with diagnostic and prognostic values are thus continually sought for combating cancer, especially in the case of prostate cancer. Although classic approaches based on gene expression profiling have revealed many potential cancer biomarkers, several studies indicate that tumor-specific mRNA isoforms may further improve the specificity of tumor diagnosis (5) and/or provide novel mechanistic insights into tumor biology (6, 7).

The transcriptome in eukaryotic cells is marked by prevalent expression of mRNA isoforms. A recent estimate suggests that more than half of the human genes express mRNA isoforms via alternative splicing (8). Interestingly, alternatively spliced regions often show a high degree of sequence conservation among mammalian genomes, suggesting that many alternative splicing events may have critical biological functions (9). mRNA isoforms may dramatically enlarge the complexity of the proteome, thereby contributing to functional diversity in different cell types. Alternative splicing may also serve as a mechanism to achieve temporal and spatial regulation of gene expression as increasing evidence suggests that alternative splicing may be tightly coupled with both upstream events in transcription and downstream steps in mRNA export, degradation, and translation (10–13). Thus, the pattern of mRNA isoform expression likely reflects a highly integrated program in the regulation of gene expression in a specific cell type. Because distinct mRNA isoforms may be uniquely associated with a disease process, either as products of cellular transformation or as causative factors for a specific disease phenotype, characteristic mRNA isoforms may serve as biomarkers for disease diagnosis and prognosis as well as unique targets for disease intervention (14, 15).

Gene expression profiling by microarray has been a powerful tool for cancer biomarker discovery. Most array platforms developed to date, however, are not designed to distinguish mRNA isoforms, and only a few initial attempts have been made to examine mRNA isoforms in eukaryotic cells (8, 16–20). We previously described a multiplex mRNA isoform detection system known as the RASL assay (for RNA-mediated annealing, selection, and ligation), which is coupled with a universal array on fiberoptic bundles to allow high-throughput mRNA isoform profiling (16). More recently, we developed a parallel approach known as the DASL assay (for cDNA-mediated annealing, selection, extension, and ligation; ref. 21) and showed that this approach works with partially degraded biological samples such as RNA derived from tissue blocks that have been formalin-fixed and paraffin-embedded (22). In this report, we used the DASL assay system to identify signature mRNA isoforms that are highly characteristic of prostate cancer at both the cell and tissue levels. We developed a data analysis strategy to quantify changes in transcript abundance and mRNA isoform ratio, which reveals an extensive link between transcriptional and posttranscriptional regulation, a newly emerged paradigm in coupled processes during gene expression (23–28). The approach described here

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org).

Requests for reprints: Xiang-Dong Fu, Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093-0651. Phone: 858-534-4937; Fax: 858-534-8549; E-mail: xdfu@ucsd.edu or Jian-Bing Fan, Illumina Inc., San Diego, CA 92121. Phone: 858-202-4588; Fax: 858-202-4680; E-mail: jfan@illumina.com.

doi:10.1158/0008-5472.CAN-05-4264

establishes the foundation for future large-scale analysis of biomarkers associated with distinct disease stages and clinical outcomes in prostate cancer, which is also generally applicable to other cancer and disease types.

Materials and Methods

Cell culture. LNCaP, DU145, and PC3 cells were maintained in RPMI 1640 plus 10% fetal bovine serum in the presence of penicillin (100 units/mL) and streptomycin (100 μg/mL). LAPC4 cells were cultured in DMEM plus fetal bovine serum and antibiotics. RWPE1 and RWPE2 were cultured in keratinocyte serum-free medium (Invitrogen, Carlsbad, CA) supplemented with bovine pituitary extracts (50 μg/mL), epidermal growth factor (5 ng/mL), and gentamicin (50 μg/mL).

Tissue specimens. Two sets of formalin-fixed and paraffin-embedded prostate tissue samples were used in this study. Set 1 consisted of 10 cancerous and 6 normal tissues, and set 2 included 12 cancerous and 10 normal tissues. Radical prostatectomy specimens were processed under a routine pathologic protocol. Specimens were received from the pathology laboratory within 45 minutes of removal and fixed in 10% buffered formalin overnight. Representative sections were submitted for tissue processing and paraffin embedding. Histopathologic features of each sample were reviewed to confirm diagnosis and tumor content with Gleason scores ranging from 7 to 9, and tumor content from 25% to 95%. Specific tissue sections that included areas of carcinoma were selected for RNA extraction.

RNA isolation. For each tissue sample, RNA was extracted from five 5-μm sections using the High Pure RNA Paraffin Kit (Roche Diagnostics GmbH, Mannheim, Germany), yielding 0.5 to 3 μg of total RNA. RNA extraction, DNase treatment, and other steps were done according to the manufacturer's protocol, except that Proteinase K digestion was carried out for 12 hours. Isolated RNAs were stored at −80°C until use.

BeadArray technology. Universal bead arrays were assembled by loading pools of glass beads (3 μm in diameter) derivatized with oligonucleotides onto the etched ends of fiberoptic bundles (29). About 50,000 optical fibers are hexagonally packed to form an ∼1.4 mm diameter bundle. The fiberoptic bundles are assembled into an array matrix (Sentrix Array Matrix), comprising 96 bundles arranged in an 8 × 12 matrix that matches the dimensions of standard microtiter plates (30). This arrangement allows simultaneous processing of 96 samples. A decoding process is carried out to determine the location and identity of each bead in every array location (31).

Assay probe design. As shown in Fig. 1, two oligonucleotide probes were designed to explore each target site on the cDNA as described previously (21). The first oligo consists of the donor exon-specific sequence and a universal PCR primer sequence (P1, 5′-ACTTCGTCAGTAACGGAC-3′) at the 5′-end. The second oligo consists of the acceptor exon-specific sequence and a universal PCR primer sequence (P2, 5′-GTCTGCCTATAGTGAGTC-3′) at the 3′-end. The gene-specific sequence is designed with T_m ranging from 57°C to 62°C. To detect specific mRNA isoforms from a common precursor mRNA, a specific address sequence is linked to each isoform-specific oligo between the target sequence and the primer sequence (colored lines in Fig. 1). This address sequence, which is complementary to one of 1,536 capture sequences on the universal microarray, allows the hybridization of PCR-amplified products to the array for quantification of individual mRNA isoforms.

DASL assay reaction and hybridization on Sentrix arrays. cDNA synthesis, DASL assay processes, array image processing, and signal extraction were as described previously (21, 22). Briefly, a 20 μL cDNA synthesis reaction was carried out with a reaction mix (MCS, Illumina, San Diego, CA) containing biotinylated random nonamers and oligo-d(T)$_{18}$, and total RNA (up to 1 μg). Pooled assay oligos were annealed to their targets on the cDNA under a controlled hybridization program. The cDNA was immobilized on paramagnetic beads and washed to remove excess oligos. Hybridized oligos were then extended and ligated to generate amplifiable templates using Illumina-supplied reagents and conditions (DASL assay system manual, Illumina). PCR was done using universal PCR primers, one labeled with Cy3. Single-stranded PCR products were prepared by denaturation, and hybridized to Sentrix arrays under a temperature gradient program. The arrays were imaged using a BeadArray Reader (Illumina). Image processing and intensity data extraction software were as described previously (32). Raw data were normalized using the cubic spline normalization strategy (33).

RT-PCR and qPCR. One microgram of total RNA was reverse-transcribed using random hexamers and AMV reverse transcriptase. cDNA (equivalent to 50 ng of RNA) was subjected to PCR. Specific primer pairs with matching T_m's were targeted to constitutive exonic sequences surrounding alternative exons or exonic regions. Most PCR primers (specific sequences for individual primers are listed in Supplementary Table S1) were designed to amplify ∼90-bp fragments. Standard PCR conditions were used, but cycle numbers varied from 25 to 35 cycles to maintain linearity. The lowest cycle number sufficient to amplify detectable products was chosen in each case. PCR products were separated on 2% agarose gels stained with ethidium bromide. Quantitative real-time PCR (qPCR) analyses were done on the ABI Prism 7900HT sequence detection system (Applied Biosystems, Foster City, CA) as described previously (21).

Data normalization and analysis. Prior to statistical analysis, raw data were first normalized against a synthetic average by using LOcally WEighted polynomial regreSSion (LOWESS) transformation (34). The data was then scaled to the same median and cluster analysis was conducted using Cluster and TreeView.[8]

Statistical significance of changes in transcript abundance (Pt) and splicing (Ps) were analyzed as follows. To calculate Pt, we first summed the level of individual isoforms in each measurement. This gave rise to three values from three biological repeats in one cell type or experimental condition, which were then compared with the three values from a different cell type or experimental condition to calculate the *P* value. To calculate Ps, we first derived the ratio of each isoform measured in two cell types or under two different experimental conditions. This gave rise to three ratio values from three biological repeats, which were then compared with the three ratio values for the alternative isoform from the same gene to calculate the *P* value.

Results

Experimental design to profile mRNA isoforms. To accurately profile mRNA isoforms by microarray, a pool of oligo pairs (in each pair, one oligo is indexed with a unique address sequence) was annealed to cDNAs generated with biotinylated random primers (Fig. 1*A*). The annealing reaction was followed by affinity selection for hybridized oligos on biotinylated cDNA, removing unhybridized oligos. Paired oligos were next ligated to become amplifiable templates for PCR. This step distinguishes our system from other array-based approaches because specificity is enforced by hybridization and ligation, rather than by hybridization alone. The PCR step significantly enhances the sensitivity of the assay and the signal amplification process is relatively unbiased because all oligo pairs contain identical universal primer landing sites for PCR and the length of amplicons is similar. Amplified products were quantified by hybridization to a universal Sentrix array consisting of a unique set of address sequences. This highly specific, sensitive, and quantitative system has made it possible to profile gene expression in some of the most demanding and widely accessible biological samples, such as RNA extracted from formalin-fixed, paraffin-embedded tissue blocks in which RNAs are heavily cross-linked and extensively fragmented (22).

To explore the value of mRNA isoforms for tumor classification, we first did isoform profiling experiments on a panel of tumor cell lines. Because specific sequence information at splice junctions is

[8] http://rana.lbl.gov/EisenSoftware.htm.

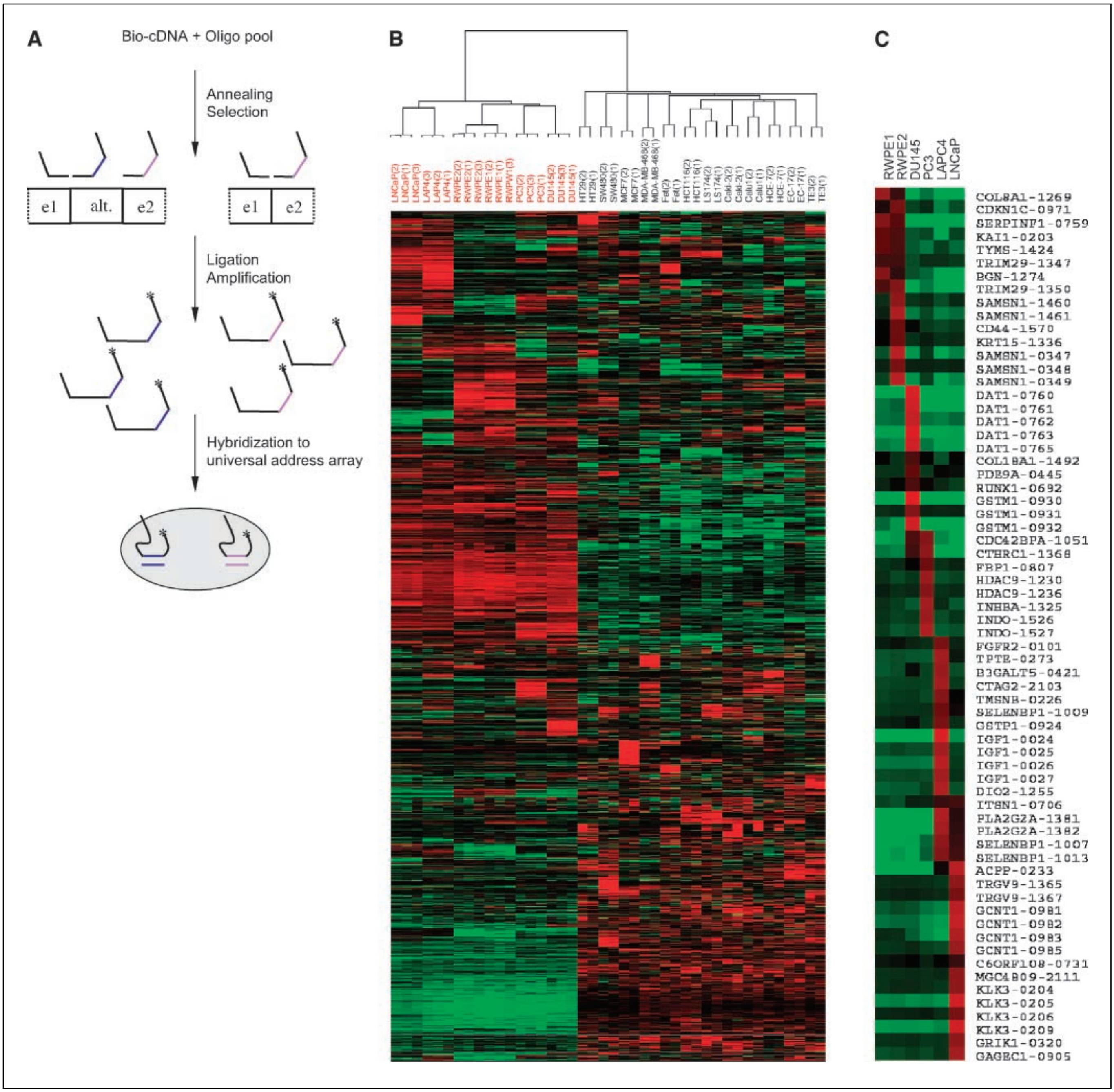


Figure 1. Profiling mRNA isoform expression in prostate cancer cell lines in comparison with other tumor cell lines. *A,* the DASL assay scheme. Total RNA is first converted to biotin-labeled cDNA using biotinylated nonamers and oligo-dT. Oligos targeted to alternative splice sites are each tagged with a specific address (*blue* and *purple*). Both indexed oligos and common oligos contain universal primer sites for PCR. cDNA is annealed to oligos, followed by affinity selection on streptavidin beads. Oligos annealed to specific splice junctions in pairs are ligated and the resulting amplicons are amplified by PCR using universal primers, one of which is Cy3-labeled. The products are hybridized to universal Sentrix bead arrays. *B,* unsupervised hierarchical clustering analysis of 1,532 isoforms expressed in 18 cancer cell lines. *Rows,* individual isoforms; *columns,* individual samples. Prostate cell lines were assayed in triplicate (biological replicates) and other cell lines were measured in duplicate (technical replicates). Dendrogram of samples shows overall similarity in isoform profile across the cell lines. Prostate cancer cells (*red*) are clustered in one group and the rest of tumor cell lines (*black*) in another. Androgen-sensitive (LNCaP and LAPC4) and -insensitive (PC3 and DU145) cell lines were also segregated. *C,* examples of prostate cell type–specific isoforms.

required for oligo design, we selected several hundred genes for isoform annotation. We focused on genes associated with prostate cancer from published expression profiling experiments (35–38), the cancer anatomy project at NIH,[9] and mechanistic studies reported in the literature (39). Of ∼500 genes selected, ∼70% (364 genes) were found to express multiple mRNA isoforms, which were individually annotated by aligning cDNA sequences against genomic sequences using a computer-assisted alternative splicing annotation program (40). The statistical features of the annotated genes were separately described (41). Oligos were designed and synthesized for each splice junction to examine a total of 1,532

[9] http://cgap.nci.nih.gov/Tissues/GXS.

mRNA isoforms, which correspond to a total of 721 alternative splicing events (note that many genes show more than one alternative splicing event).

Signature mRNA isoforms in prostate cancer cell lines. The total RNA extracted from three independent cultures of 6 prostate cell lines and 12 randomly selected tumor cell lines of nonprostate origin were profiled. Results from independent cultures of the same cell type (biological replicates) showed high concordances (R^2 > 0.95) in all cases (data not shown; ref. 21). Hierarchical clustering analysis revealed robust cosegregation of related cell types, with all prostate cancer cell lines clustering on one side of the dendrogram (Fig. 1*B*). Significantly, each prostate cancer cell line was associated with the expression of a group of specific RNA isoforms (Fig. 1*C*). For example, the *KLK3* gene, which encodes the PSA, was highly expressed in LNCaP cells derived from early neoplasia, but was depressed in other prostate cancer cell lines, consistent with elevated PSA expression in early neoplastic prostate tumors reported in the literature. Clustering analysis also revealed two large groups of isoforms that were either uniformly up-regulated or down-regulated in all prostate cancer cell lines compared with all other tumor cell lines (Fig. 1*B*). Together, the data show the ability of this approach to define the molecular characteristics of individual cell types by mRNA isoforms.

The six prostate cell lines surveyed were segregated into three distinct groups: (*a*) cell lines derived from normal prostate epithelial cells (RWPE-1 and RWPE-2), (*b*) androgen-dependent cells (LNCaP and LAPC4), and (*c*) androgen-independent cells (DU145 and PC3). The coclustering of the RWPE lines was expected because they were derived from a common origin. RWPE-1 was immortalized by a human papillomavirus, and RWPE-2 is a v-Ki-ras transformed derivative of RWPE-1 (42), raising the possibility that these lines may be useful for studying regulated splicing induced by ras signaling. Our array results showed differential expression between RWPE-1 and RWPE-2 cells of several isoforms from a number of genes, including CD44, KRT15, and SAMSN1 (Fig. 1*C*). Of particular interest is the case of CD44 alternative splicing, which has been shown to be regulated by ras via the RNA-binding proteins, SAM68 and hnRNP A1 (43). These results therefore validate our experimental approach for defining regulated splicing events.

Deconvoluting changes in transcript abundance and splicing. The abundance of mRNA isoforms in a specific cell type may result from regulated gene expression at the level of transcription, RNA stability, and splicing. This complexity is clearly reflected by our array data (Fig. 1*C*). In many cases, multiple isoforms from one gene were similarly elevated or depressed in a given cell type, indicating coordinated changes in transcription and/or RNA stability. On the other hand, other isoforms seemed to be uniquely expressed in a specific cell type(s), suggesting that those genes may be differentially regulated at the splicing level.

To develop a standardized data analysis strategy to differentiate changes at different regulatory levels, we reasoned that an apparent array signal (S_{app}) should be equal to $X + Y + (P + N) \times \theta$, where X is the baseline signal, Y is the measurement error, P is the true product level, N is the cross-hybridization signal, and θ is the probe efficiency. Because of the high redundancy built into our array system (31), both the baseline signal and measurement error are minimal, as indicated by the high concordance among technical replicates. In addition, the ligation reaction built into our assay practically eliminates cross-hybridization, as previously shown (16). Thus, the formula can be simplified to $S_{\mathrm{app}} \approx P \times \theta$. By taking the ratio of readings from the same probe in pairwise comparisons, the fold change can thus be expressed as $\log_2(P_1/P_2)$ for product P in cell type 1 versus cell type 2.

As illustrated in Fig. 2*A*, we were able to calculate total transcript changes by summing up the weighted fold change for isoforms A and B, where the weight was roughly estimated by the fractional contribution of individual isoforms to the total signal detected for a given gene. To calculate a splicing change, we subtracted fold changes for the two isoforms to detect "anticorrelation" as previously described (18). For instance, if two isoforms were similarly up-regulated or down-regulated, the splicing change for the gene would be close to zero. If one isoform were up-regulated, and the other down-regulated, the splicing change would be the sum of the fold changes. In cases where a pre-mRNA gives rise to more than two isoforms, we selected probes exhibiting the largest ratio difference to calculate the trend change in splicing. To determine the statistical significance of a change in transcript abundance or splicing, *t* tests were conducted based on biological repeats to calculate a *P* value for the transcript change (Pt) and splicing change (Ps). This data transformation allowed us to use the same data set to score potential changes in transcript abundance, splicing, or both in pairwise comparisons.

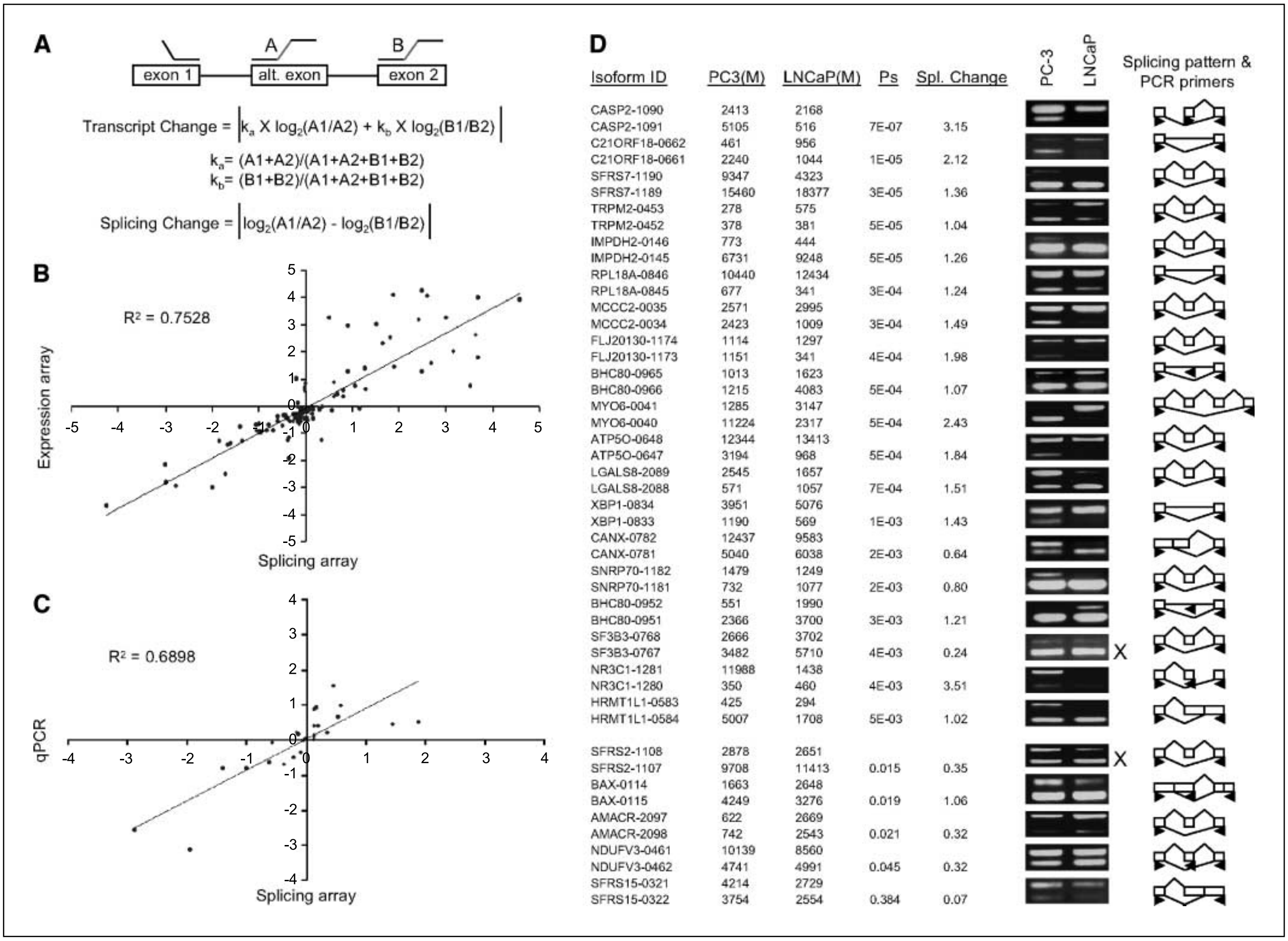
Experimental validation of the data analysis strategy. To experimentally validate our data analysis strategy, we selected 107 genes for expression profiling using an independent oligo pool in which three pairs of oligos were designed to target common regions in each transcript, a standard approach for expression array using the DASL assay (21). The ratio between mean signals for each probe set in one cell line (LNCaP) versus another (PC3) was used to derive differences in total transcript abundance (Fig. 2*B*). When the expression array data were compared with those calculated from the splicing array, we obtained a concordance of R^2 = 0.75, indicating a reasonable agreement between our approach and the standard expression array analysis. To further validate our data analysis strategy, we measured fold differences in a subset of the genes by qPCR and obtained a similar concordance (R^2 = 0.67) with splicing array results (Fig. 2*C*).

To validate the splicing changes detected in prostate cancer cell lines, we selected a panel of genes for RT-PCR analysis. The results are shown in Fig. 2*D* for 24 genes, ranked in order of *P* values (Ps) for calculated splicing changes. Although quantitative differences between fold changes detected by splicing array and those by RT-PCR are difficult to compare, the RT-PCR data clearly corroborate the array results in a qualitative manner. Among 19 genes scored at a *P* value of <0.005 for differences in splicing, 18 were detected by RT-PCR as being differentially spliced between LNCaP and PC3 cells with the exception of the *SF3B* gene (*indicated by the cross*; Fig. 2*D*). Conversely, four out of five genes showing higher *P* values with the array analysis were indeed unaltered in splicing by RT-PCR with the exception of the *SFSR2* gene (*indicated by the cross*), which showed a slight shift in splicing. We therefore chose a *P* value cutoff of <0.005 to score splicing changes. A similar *P* value cutoff was used to identify changes in transcript abundance, although this choice was somewhat arbitrary because of the lack of quantitative comparison between microarray and qPCR data. Together, these independent methods validated the data analysis strategy in simultaneously scoring changes in transcript abundance and splicing using the same data set from the splicing array.

"Two-dimensional" profiling and coordinated transcriptional and posttranscriptional regulation. Having established a data analysis strategy to simultaneously score changes in transcript abundance and splicing, we analyzed and plotted the data from different prostate cancer cell lines according to fold differences for individual genes at the level of transcript abundance and splicing (Fig. 3). To illustrate genes that were altered significantly between two given cell types, we highlighted genes with fold changes >1.5 and $P < 0.005$ with different colors: *black,* genes with no significant change in both splicing and transcript abundance [fold change <1.5 (<0.6 on the $\log_2$ scale) or $P > 0.005$]; *red,* genes showing a change in transcript abundance, but not splicing (fold change >1.5 and $P < 0.005$ only at the transcript abundance level); *green,* genes showing a change in splicing, but not transcript abundance (fold change >1.5 and $P < 0.005$ only at the splicing level); and *blue,* genes showing changes in both transcript abundance and splicing (fold change >1.5 and $P < 0.005$). For isoforms showing near-background intensities in both cell types or in cases where both isoforms had intensities near background in any cell type, a reliable splicing change would be difficult to derive. In these cases, the splicing change was set to zero, and only changes at the transcript abundance level were calculated.

Based on the above criteria, a comparison between LNCaP and PC3 cells revealed that many genes showed significant differences in transcript abundance, splicing, or both (Fig. 3*A*): a substantial number of genes showed changes only in transcript abundance (*red*, 23.3% of total), indicating that these genes may be differentially regulated at the level of transcription and/or RNA stability. We also noted many genes showing changes only in splicing (*green*, 4.2% of total), suggesting differential splicing in these prostate cell lines. Strikingly, many genes showed significant covariation in both transcript abundance and alternative splicing (*blue*, 9.7% of total), indicating that they may be regulated by coupled mechanisms. Among genes showing changes in splicing, a large fraction (60-70%) exhibited covariation (genes labeled with *blue* over those labeled with either *blue* or *green*). Similar observations were also made between PC3 and DU145 cells (Fig. 3*B*) and all pairwise comparisons between the prostate cancer cell lines (data not shown). These results strongly suggest that splicing regulation may be extensively coupled with the control of transcription/RNA stability *in vivo*.

Tumor classification based on signature mRNA isoforms. Having established the capability of the array method and the data analysis strategy, we applied this approach to the analysis of



Isoform ID	PC3(M)	LNCaP(M)	Ps	Spl. Change
CASP2-1090	2413	2168		
CASP2-1091	5105	516	7E-07	3.15
C21ORF18-0662	461	956		
C21ORF18-0661	2240	1044	1E-05	2.12
SFRS7-1190	9347	4323		
SFRS7-1189	15460	18377	3E-05	1.36
TRPM2-0453	278	575		
TRPM2-0452	378	381	5E-05	1.04
IMPDH2-0146	773	444		
IMPDH2-0145	6731	9248	5E-05	1.26
RPL18A-0846	10440	12434		
RPL18A-0845	677	341	3E-04	1.24
MCCC2-0035	2571	2995		
MCCC2-0034	2423	1009	3E-04	1.49
FLJ20130-1174	1114	1297		
FLJ20130-1173	1151	341	4E-04	1.98
BHC80-0965	1013	1623		
BHC80-0966	1215	4083	5E-04	1.07
MYO6-0041	1285	3147		
MYO6-0040	11224	2317	5E-04	2.43
ATP5O-0648	12344	13413		
ATP5O-0647	3194	968	5E-04	1.84
LGALS8-2089	2545	1657		
LGALS8-2088	571	1057	7E-04	1.51
XBP1-0834	3951	5076		
XBP1-0833	1190	569	1E-03	1.43
CANX-0782	12437	9583		
CANX-0781	5040	6038	2E-03	0.64
SNRP70-1182	1479	1249		
SNRP70-1181	732	1077	2E-03	0.80
BHC80-0952	551	1990		
BHC80-0951	2366	3700	3E-03	1.21
SF3B3-0768	2666	3702		
SF3B3-0767	3482	5710	4E-03	0.24
NR3C1-1281	11988	1438		
NR3C1-1280	350	460	4E-03	3.51
HRMT1L1-0583	425	294		
HRMT1L1-0584	5007	1708	5E-03	1.02
SFRS2-1108	2878	2651		
SFRS2-1107	9708	11413	0.015	0.35
BAX-0114	1663	2648		
BAX-0115	4249	3276	0.019	1.06
AMACR-2097	622	2669		
AMACR-2098	742	2543	0.021	0.32
NDUFV3-0461	10139	8560		
NDUFV3-0462	4741	4991	0.045	0.32
SFRS15-0321	4214	2729		
SFRS15-0322	3754	2554	0.384	0.07

Figure 2. Data analysis strategy and validation by independent quantification methods. *A,* oligo design for splicing array and data analysis strategy. Oligos A and B are targeted to alternative splice sites. The change in transcript abundance is expressed as the sum of weighted ratios, whereas the change in splicing is calculated by the fold difference of individual isoforms. *B,* concordance in transcription of 107 genes in LNCaP and PC3 cells measured by the splicing array and the standard gene expression array. *C,* concordance in the expression of 24 genes measured by the splicing array and qPCR. *D,* RT-PCR validation and determination of the *P* value cutoff for splicing switches. Listed is a panel of genes along with the raw data from the splicing array comparing LNCaP and PC3 cells. The alternatively spliced regions for each gene with the primer design for RT-PCR analysis (*right*). PCR cycle numbers were individually determined. The trend changes are largely consistent with the array results, with two outliers (×).

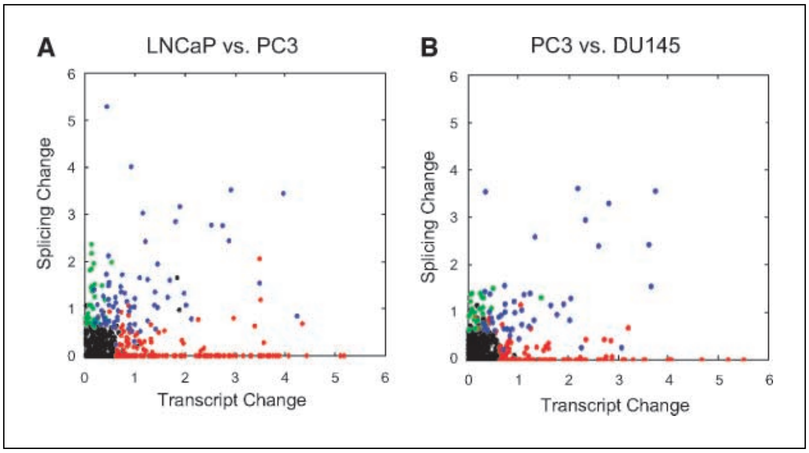


Figure 3. "Two-dimensional profiling" for changes in transcript abundance and splicing. *A,* comparison between LNCaP and PC3 cells. *B,* comparison between PC3 and DU145 cells. Individual genes and associated alternative splicing events are displayed in a two-dimensional plot according to calculated changes in transcript abundance and splicing. Colors are assigned based on fold difference (>1.5 fold or >0.6 on the $\log_2$ scale) and the *P* value (<0.005). *Black spots,* no change in both transcript abundance and splicing; *red spots,* change in transcript abundance only; *green spots,* change in splicing only; and *blue spots,* changes in both transcript abundance and splicing.

clinical samples. In this study, we took advantage of the benefits of our technology for analyzing formalin-fixed, paraffin-embedded tissue samples (21), and collected 10 prostate tumors and 6 normal tissues from the University of California, San Diego prostate tumor bank. These human tissue samples had been stored for various periods up to 11 years, and histopathologic characterization of H&E stained sections showed variable degrees of tumor content and pathologic stages (Table 1A). Total RNA was isolated from 5-μm sections of each tissue block and used in the DASL assay. Each RNA sample from the same tissue was arrayed twice to ensure assay reproducibility.

To identify isoform signatures uniquely associated with prostate cancer, we conducted feature selection by *t* test to identify isoforms that were differentially expressed in prostate cancer compared with normal prostatic tissue. The top 50 differentially expressed isoforms were of $P < 0.0003$, whereas the median number at this cutoff in 200 random permutations was 2 and the 90th percentile was 5, indicating that the features selected are highly specific (44). During the course of data analysis, we realized that the *t* test might not be ideal for feature selection because it penalizes those showing large variations in cancer samples. For example, an isoform may be a good biomarker for prostate cancer if it is expressed at a low level in normal prostatic tissue and a high but variable level in prostate cancer. Such an otherwise valuable isoform biomarker would not rank high in a *t* test. To overcome this problem, we conducted a Wilcoxon test in parallel to derive a separate list of 50 signature isoforms, which favored the significance of trend changes in comparison between normal and tumor specimens. By combining results that were scored significantly in both tests, a total of 61 unique isoform features were identified. We used these features to segregate normal tissues from tumors by unsupervised clustering analysis and determined the improvement of tumor classification by sequential elimination of low-ranking isoform features. A total of 57 isoforms were found to maximally segregate tumors from normal tissues (Fig. 4*A*).

To validate selected isoforms as prostate cancer biomarkers, we did additional profiling experiments using an independent set of normal and tumor tissues (Table 1B). The 57 isoforms selected from the first sample set were used to classify the new sample set and a good segregation was achieved with misclassification of only 1 out of 10 normal tissues, and 1 out of 12 tumor samples, the latter of which contained a lower tumor content (Fig. 4*B*). We then conducted a converse analysis by using the second sample set for feature selection and the first sample set for validation. We identified a separate list of 57 isoforms (the number is coincidental) capable of maximally segregating prostate tumors from normal prostatic tissues (Fig. 4*C* and *D*). Ten isoforms were common in both lists, which is not surprising, considering the contribution of multiple variations in tissue heterogeneity to signature ranking as previously discussed (45). Importantly, when the combined panel of 104 isoforms was displayed across all tissue samples (Fig. 4*E*), the majority of the isoforms showed differential expression between normal prostate tissue and prostate cancer. These isoform biomarkers are listed according to the *P* values from the Wilcoxon test in Supplementary Table S2.

Similar to the tumor cell line comparisons, we noticed that many of the mRNA isoform biomarkers derived from the same precursor mRNAs were similarly elevated or repressed in tumors, indicating that those genes may be altered primarily at the level of gene expression. However, many specific isoforms were also uniquely up-regulated or down-regulated in tumors, suggesting that individual mRNA isoforms from those genes may be differentially regulated at the level of splicing. These specific mRNA isoforms would thus be more powerful than total transcript abundance in molecular classification of cancer.

Comparison between cell line and tumor markers. We next asked whether the isoform biomarkers identified in tumors were characteristic of prostate cancer cell lines, and conversely, whether signature mRNA isoforms in prostate cancer cell lines could reflect characteristics of prostate cancer. As shown in Supplementary Fig. S1*A*, the use of the 104 isoform markers derived from prostate cancer tissues enabled the clustering of the prostate cancer cell lines as well, with a clear distinction from other tumor cell lines. Thus, these biomarkers are characteristic of prostate cancer at both the tissue and cell line levels. About half of the isoforms were uniformly up-regulated or down-regulated in prostate cancer cell lines in comparison with other tumor cell lines, suggesting that this group of isoforms may be useful in future cross-tumor comparisons. The remaining half were differentially expressed in subsets of prostate cancer cell lines and other tumor cell lines, suggesting that these groups may play some role in tumorigenesis, but not in a prostate cancer–specific manner.

We then did a converse analysis, asking whether we could segregate prostate cancers from normal prostatic tissues using isoform markers derived from prostate cancer cell lines. We separated the cell line markers into three groups: (*a*) those showing cell type–specific isoform expression in prostate cancer cell lines (see Fig. 1*C*); (*b*) those showing universal high expression in all prostate cancer cell lines in comparison with other cancer cell lines (*red,* Fig. 1*B*); and (*c*) those showing universal low expression in prostate cancer cell lines in comparison with other cancer cell lines (*green,* Fig. 1*B*). None of these isoform groups could be used to segregate prostate tumors from normal prostatic tissues (Supplementary Fig. S1*B* for the first group; data not shown for the second and third groups). These observations indicate that contrary to the ability of prostate cancer biomarkers to characterize prostate cancer cell lines, biomarkers derived from cell lines may not be representative of the original tumors, likely reflecting events during cell culture among other possibilities.

Splicing switch during prostate tumorigenesis. Inspection of the isoform biomarkers listed in Supplementary Table S2 indicates that many genes are alternatively spliced during prostate tumorigenesis. To identify genes that showed isoform expression change in opposite directions in prostate tumors, we selected five pairs of normal and cancerous tissues from the same patients in our sample sets and did two-dimensional

Table 1.

(A) Clinicopathologic information of the first set of prostate normal and tumor tissues

RNA no.	Case no.	Age	PSA	Stage	Gleason score	Tumor (%)	BPH (%)	Atrophy (%)	Stroma (%)	Inflammation (%)
77	N1	66	3.15			0	0	10	89	1
85	N2	66	5.4			0	25	0	75	0
88	N3	61	2.23			0	10	30	60	0
113	N4	70	4.78			0	10	5	85	0
109	N5	67	7			0	5	0	90	5
56	N6	67	5.7			0	5	0	94	0
78	T1	66	3.15	$T_{2c}N_xM_x$	3 + 4 = 7	30-35	5	0	65	0
86	T2	66	5.4	$T_{3b}N_0M_x$	4 + 4 = 8	90	5	0	5	0
87	T3	61	2.23	$T_{2b}N_0M_x$	4 + 3 = 7	25-30	45	5	20	0
114	T4	70	4.78	$T_{3a}N_xM_x$	4 + 4 = 8	40	0	5	55	0
110	T5	67	7	$T_{2b}N_xM_x$	3 + 4 = 7	40	0	0	58	0
122	T6	67	7	$T_{2b}N_xM_x$	3 + 4 = 7	70	0	5	25	0
72	T7	68	8.27	$T_{3b}N_1M_x$	4 + 3 = 7	70	0	0	30	0
84	T8	60	9.99	$T_{3b}N_0M_x$	4 + 5 = 9	70-80	0	0	20	0
107	T9	68	7.4	$T_{2b}N_xM_x$	4 + 3 = 7	60	10	0	30	0
123	T10	78	17.7	NA	5 + 5 = 10	80	0	0	20	0
(B) Clinicopathologic information of the second set of prostate normal and tumor tissues										
22	N1	74	6.7			0	10	40	50	0
30	N2	55	11.68			0	10	30	68	0
44	N3	61	5.46			0	10	2	88	0
46	N4	74	8.06			0	45	20	35	0
121	N5	50	0.22			0	30	2	68	0
148	N6	67	4.68			0	35	10	55	0
155	N7	70	8.4			0	40	10	48	2
196	N8	73	4.59			0	40	5	55	0
201	N9	64	NA			0	20	5	45	0
133	N10	NA	NA			0	25	5	75	0
5	T1	67	8.48	$T_{3b}N_1M_x$	5 + 4 = 9	50	0	0	20	0
21	T2	74	6.7	$T_{2b}N_xM_x$	4 + 4 = 8	60	10	10	20	0
147	T3	78	6.9	$T_{2b}N_0M_x$	4 + 4 = 8	70	0	0	30	0
167	T4	72	18	$T_{2b}N_0M_x$	4 + 4 = 8	80	0	10	10	0
174	T5	83	15	T_4	5 + 4 = 9	70	5	0	25	0
177	T6	67	10.87	$T_{2c}N_0M_x$	4 + 4 = 8	40	0	30	30	0
189	T7	77	2.51	$T_{2b}N2M_x$	5 + 5 = 10	70	0	0	0	30
192	T8	61	5.7	$T_{3a}N_xM_x$	4 + 4 = 8	50	5	10	35	0
197	T9	67	21.82	$T_{3a}N_1M_x$	4 + 4 = 8	95	0	0	5	0
198	T10	60	4.06	$T_{3b}N_xM_x$	4 + 4 = 8	65	0	10	25	0
202	T11	67	12.34	$T_{3b}N_xM_x$	4 + 4 = 8	90	0	5	5	0
204	T12	54	3.91	$T_{3c}N_xM_x$	4 + 5 = 9	80	0	5	15	0

NOTE: Summarized are the clinicopathologic data in the two sets of prostate carcinoma and benign prostate tissues. The noncancerous, benign prostate tissues were retrieved from patients who either had diagnostic carcinomas elsewhere or from resections for benign diagnosis. The patient characteristics include age, pre-surgery PSA, tumor stage based on the 6th edition of the American Joint Committee on Cancer guideline, Gleason grades, and scores. The samples were reviewed for percentage of tumor, glandular hyperplasia, atrophy, stroma, and inflammation. Identical PSA values indicate samples from the same patients.

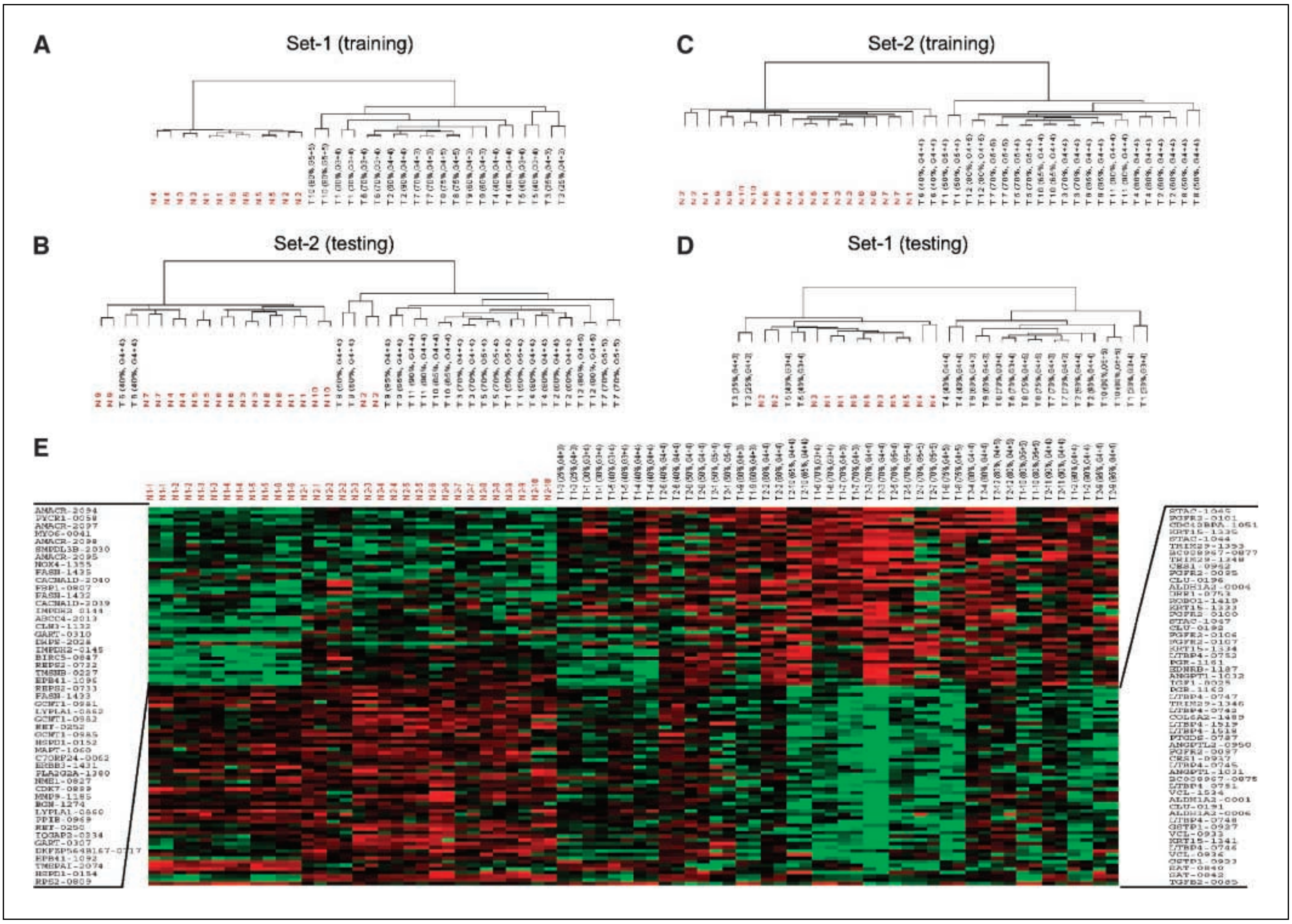


Figure 4. Unsupervised hierarchical cluster analysis of normal and prostate cancer tissues. *A,* clustering analysis of the first tissue set with 57 selected isoform markers. Normal prostatic tissues (*red*) were completely segregated from prostate cancer tissues. *B,* the 57 isoforms selected from the first set were used to cluster an independent second tissue set. One normal and one cancer tissue sample were missegregated. *C,* an independent list of 57 isoforms identified from the second tissue set was used to maximally segregate normal from tumor samples. One tumor sample was always missegregated with any combination of isoform markers. *D,* the isoform markers selected from the second tissue set were tested on the first tissue set. Two tumor samples were missegregated. All missegregated tumors contain a relatively low tumor content and/or tumor grade (see Table 1). *E,* the isoforms selected from the two independent sets (104 in total) are displayed across all tissue samples to visualize trend changes. Normal tissues (*red*) and tumors are arranged from left to right according to tumor content and Gleason scores. Up-regulated (*left*) and down-regulated (*right*) isoforms in tumors are listed according to *P* values from the Wilcoxon ranking test.

analysis. As shown in Fig. 5*A*, a number of genes showed changes with sufficient statistical significance in splicing (*green,* 11% or 1.5% of total), in both transcript abundance and splicing (*blue,* 7% or 1% of total), and transcript abundance alone (*red,* 9% or 1.2% of total).

To validate these changes, we prepared laser-captured normal prostatic epithelia and prostate cancer samples from independent frozen tissues (see representative captures in Fig. 5*B*) and conducted RT-PCR analysis (Fig. 5*C*). As predicted by the array result, two mRNA isoforms expressed from the *MAPT* gene (encoding the microtubule-associated tau protein) showed opposite changes in normal versus tumor, indicating a splicing switch during tumorigenesis. The *CACNA1D* gene (which encodes a Ca^{2+} channel) was up-regulated in prostate cancer, but the larger isoform was more elevated than the smaller isoform, indicating that the gene was differentially regulated at the levels of both transcript abundance and splicing. The *AMACR* gene (which encodes α-methylacyl-CoA racemase) has recently emerged as a robust biomarker for prostate cancer (46, 47). We analyzed the two isoforms resulting from the alternative use of the last exon coupled with alternative polyadenylation (5), and found that whereas one isoform showed a quantitative up-regulation in prostate cancer compared with normal prostatic tissues, the other seemed to be expressed only in prostate cancer. The short isoform may thus serve as a better indicator for prostate cancer, which agrees with a recent analysis of a large number of expression profiling results (5). Together, these data not only validate our profiling results, but also illustrate signature mRNA isoforms as unique biomarkers for prostate cancer.

Discussion

Changes in gene expression may not only serve as diagnostic and prognostic tumor markers, but also provide potential targets for the development of new therapeutic strategies (48). Although a variety of approaches have been used to search for DNA-, RNA-, and protein-based biomarkers associated with specific tumor types and/or stages (49), our current effort represents the first systematic attempt to identify tumor-specific mRNA isoforms. Our analysis is based on the rationale that mRNA isoforms may better reflect the biological state of specific cell types or tissues, and thus may serve as more robust biomarkers for disease diagnosis and prognosis. This rationale is supported by several recent studies which indicate that mRNA isoforms are the products of independent regulatory pathways (18–20).

By using the two-dimensional profiling strategy, we found a large number of genes that showed coordinated changes in transcript abundance and splicing, indicating that many distinct steps in gene expression from transcription, stability control, splicing, and transport may be distinctly coupled in different cell types. The network in coupling multiple steps in gene expression represents a new paradigm in the regulation of gene expression in eukaryotic cells, although most evidence collected to date is from transfected cells (50). Our observations reveal frequent coupling events *in vivo* and provide physiologically relevant models to pursue potential coupling mechanisms.

Our isoform profiling experiments suggest that signature mRNA isoforms may be more powerful in distinguishing between normal prostatic epithelia and prostate cancer than total transcript alone. For example, although the gene products of *AMACR* were recently reported to be highly diagnostic for prostate cancer, a particular *AMACR* isoform seems uniquely associated with prostate cancer, which is consistent with a recent analysis based on a large amount of profiling data (5). In addition, our studies revealed many other candidate splicing switch events associated with prostate tumorigenesis, and a number of representative events were further validated by laser capture microdissection. These data should allow the development of prostate cancer biomarkers by combining quantitative (up-regulation and down-regulation) and qualitative (splicing switch) differences in gene expression.

When expression profiling is used for biomarker discovery, the requirement for cross-validation using independent experimental data sets is of critical importance (51). We accomplished this by using two different panels of tissue samples. In addition, we profiled available prostate tumor cell lines in comparison with other tumor cell lines. When these independent data sets were cross-analyzed, a highly specific panel of mRNA isoforms emerged. The identified isoform biomarkers could be used to segregate prostate cancer from normal prostatic tissues as well as prostate cancer cells from other tumor cell types. The molecular signatures are thus characteristic of prostate cancer at both the cell and tissue levels. Although this approach was only applied to normal/tumor comparison in the current study, the strategy is readily applicable for the characterization of tumor stage,

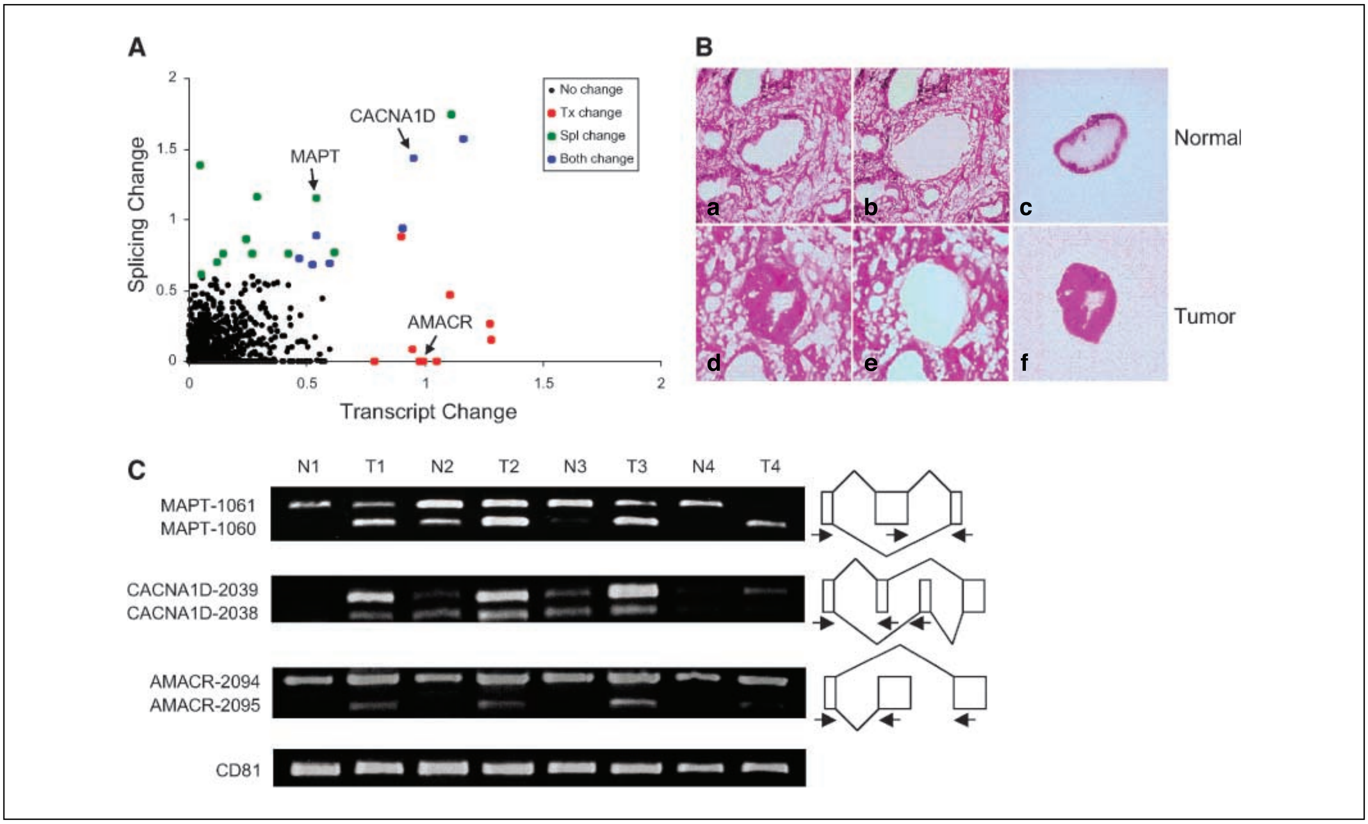


Figure 5. Two-dimensional analysis of normal prostate cancer tissue samples and validation by laser capture microdissection. *A,* profiling data from five pairs of normal and prostate cancer samples shown in Table 1A were analyzed by two-dimensional plot. Genes showing changes in transcript abundance and/or splicing with the fold change cutoff of >0.6 and the *P* value cutoff of <0.05 are labeled with individual colors as described in Fig. 3. *B,* examples of laser-captured materials. A section of frozen normal prostatic tissue (*a*) or cancerous prostate tissue (*d*) was used for laser microdissection. Images of cut sections (*b* and *e*) and captured samples (*c* and *f*). *C,* RT-PCR validation of laser microdissected materials. *N,* normal prostatic epithelia; *T,* prostate tumor. Four independent pairs were analyzed; gene names (*left*); splicing pattern and specific PCR primers (*right*). *MAPT* shows an isoform switch between normal and prostate cancer tissues. *CACNA1D* exhibits changes in transcript abundance and splicing (note that the upper isoform was significantly elevated than the lower isoform in prostate tumors). *AMACR* displays changes in transcript abundance with the upper isoform elevated in cancer in a quantitative manner and the lower isoform elevated in cancer in a qualitative manner. CD81 was analyzed as a control.

metastatic potential, and differential drug response by scaling the splicing array analysis to large numbers of samples in further studies.

Acknowledgments

Received 11/30/2005; revised 2/6/2006; accepted 2/17/2006.

Grant support: National Cancer Institute grant to X-D. Fu.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Grace Shen and Webster Cavenee for advice during the course of this study and critical comments on the manuscript, Wei Jiang for providing RNA samples from tumor cell lines, Sean Hu, Yu-Ting Jia, and Xi-Wei Wu for discussion and suggestion on microarray data analysis.

References

1. Wingo PA, Cardinez CJ, Landis SH, et al. Long-term trends in cancer mortality in the United States, 1930–1998. Cancer 2003;97:3133–275.

2. Miller DC, Hafez KS, Stewart A, Montie JE, Wei JT. Prostate carcinoma presentation, diagnosis, and staging: an update form the National Cancer data base. Cancer 2003;98:1169–78.

3. Thompson I, Leach RJ, Pollock BH, Naylor SL. Prostate cancer and prostate-specific antigen: the more we know, the less we understand. J Natl Cancer Inst 2003;95:1027–8.

4. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med 2003;349:215–24.

5. Shen-Ong GL, Feng Y, Troyer DA. Expression profiling identifies a novel α-methylacyl-CoA racemase exon with fumarate hydratase homology. Cancer Res 2003;63:3296–301.

6. Stearns ME, Wang M, Hu Y, Kim G, Garcia FU. Expression of a flt-4 (VEGFR3) splicing variant in primary human prostate tumors. VEGF D and flt-4t(Δ773–1081) overexpression is diagnostic for sentinel lymph node metastasis. Lab Invest 2004;84:785–95.

7. Narla G, DiFeo A, Yao S, et al. Targeted inhibition of the KLF6 splice variant, KLF6 SV1, suppresses prostate cancer cell growth and spread. Cancer Res 2005;65:5761–8.

8. Johnson JM, Castle J, Garrett-Engele P, et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 2003;302:2141–4.

9. Sorek R, Shamir R, Ast G. How prevalent is functional alternative splicing in the human genome? Trends Genet 2004;20:68–71.

10. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A 2003;100:189–92.

11. Green RE, Lewis BP, Hillman RT, et al. Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. Bioinformatics 2003;19 Suppl 1:i118–21.

12. Maniatis T, Reed R. An extensive network of coupling among gene expression machines. Nature 2002;416:499–506.

13. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 2004;5:89–99.

14. Brinkman BM. Splice variants as cancer biomarkers. Clin Biochem 2004;37:584–94.

15. Venables JP. Aberrant and alternative splicing in cancer. Cancer Res 2004;64:7647–54.

16. Yeakley JM, Fan JB, Doucet D, et al. Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol 2002;20:353–8.

17. Clark TA, Sugnet CW, Ares MJ. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 2002;296:907–10.

18. Le K, Mitsouras K, Roy M, et al. Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. Nucleic Acids Res 2004;32:e180.

19. Pan Q, Shai O, Misquitta C, et al. Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol Cell 2004;16:929–41.

20. Ule J, Ule A, Spencer J, et al. Nova regulates brain-specific splicing to shape the synapse. Nat Genet 2005;37:844–52.

21. Fan JB, Yeakley JM, Bibikova M, et al. A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome Res 2004;14:878–85.

22. Bibikova M, Talantov D, Chudin E, et al. Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. Am J Pathol 2004;165:1799–807.

23. Cramer P, Caceres JF, Cazalla D, et al. Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol Cell 1999;4:251–8.

24. Cramer P, Pesce CG, Baralle FE, Kornblihtt AR. Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci U S A 1997;94:11456–60.

25. de la Mata M, Alonso CR, Kadener S, et al. A slow RNA polymerase II affects alternative splicing *in vivo*. Mol Cell 2003;12:525–32.

26. Auboeuf D, Dowhan DH, Kang YK, et al. Differential recruitment of nuclear receptor co-activators may determine alternative RNA splice site choice in target genes. Proc Natl Acad Sci U S A 2004;101:2270–4.

27. Auboeuf D, Dowhan DH, Li X, et al. CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. Mol Cell Biol 2004;24:442–53.

28. Auboeuf D, Honig A, Berget SM, O'Malley BW. Coordinate regulation of transcription and splicing by steroid receptor coregulators. Science 2002;298:416–9.

29. Walt DR. Techview: molecular biology. Bead-based fiber-optic arrays. Science 2000;287:451–2.

30. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 2003;68:69–78.

31. Gunderson KL, Kruglyak S, Graige MS, et al. Decoding randomly ordered DNA arrays. Genome Res 2004;14:870–7.

32. Galinsky VL. Automatic registration of microarray images. II. Hexagonal grid. Bioinformatics 2003;19:1832–6.

33. Workman C, Jensen LJ, Jarmer H, et al. A new nonlinear normalization method for reducing variability in DNA microarray experiments. Genome Biol 2002;3:RESEARCH0048(1–16).

34. Cleveland W. Robust locally weighted regression and smoothing scatter plots. J Am Stat Assoc 1979;74:829–36.

35. Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001;412:822–6.

36. Welsh JB, Sapinoso LM, Su AI, et al. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res 2001;61:5974–8.

37. Stuart RO, Wachsman W, Berry CC, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. Proc Natl Acad Sci U S A 2004;101:615–20.

38. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;101:811–6.

39. Lara PN, Jr., Kung HJ, Gumerlock PH, Meyers FJ. Molecular biology of prostate carcinogenesis. Crit Rev Oncol Hematol 1999;32:197–208.

40. Zheng CL, Kwon YS, Li HR, et al. MAASE: an alternative splicing database designed for supporting splicing microarray applications. RNA 2005;11:1767–76.

41. Zheng CL, Fu XD, Gribskov M. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. RNA 2005;11:1777–87.

42. Bello D, Webber MM, Kleinman HK, Wartinger DD, Rhim JS. Androgen responsive adult human prostatic epithelial cell lines immortalized by human papillomavirus 18. Carcinogenesis 1997;18:1215–23.

43. Matter N, Herrlich P, Konig H. Signal-dependent regulation of splicing via phosphorylation of Sam68. Nature 2002;420:691–5.

44. Schadt EE, Li C, Su C, Wong WH. Analyzing high-density oligonucleotide gene expression array data. J Cell Biochem 2000;80:192–202.

45. Ein-Dor L, Kela I, Getz G, Givol O, Domany E. Outcome signature genes in breast cancer: is there a unique set? Bioinformatics 2005;21:171–8.

46. Zha S, Ferdinandusse S, Denis S, et al. α-Methylacyl-CoA racemase as an androgen-independent growth modifier in prostate cancer. Cancer Res 2003;63:7365–76.

47. Mubiru JN, Shen-Ong GL, Valente AJ, Troyer DA. Alternative spliced variants of the α-methylacyl-CoA racemase gene and their expression in prostate cancer. Gene 2004;327:89–98.

48. Quinn DI, Henshall SM, Sutherland RL. Molecular markers of prostate cancer outcome. Eur J Cancer 2005;41:858–87.

49. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5:845–56.

50. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. Multiple links between transcription and splicing. RNA 2004;10:1489–98.

51. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 2004;4:309–14.