Alternative splicing in Colon, Bladder, and Prostate Cancer Identified by Exon Array Analysis*

Kasper Thorsen‡, Karina D. Sørensen‡, Anne Sofie Brems-Eskildsen‡, Charlotte Modin‡, Mette Gaustadnes‡, Anne-Mette K. Hein‡, Mogens Kruhøffer‡, Søren Laurberg§, Michael Borre¶, Kai Wang∥, Søren Brunak∥, Adrian R. Krainer**, Niels Tørring‡, Lars Dyrskjøt‡, Claus L. Andersen‡, and Torben F. Ørntoft‡ ‡‡

Alternative splicing enhances proteome diversity and modulates cancer-associated proteins. To identify tissue- and tumor-specific alternative splicing, we used the GeneChip Human Exon 1.0 ST Array to measure whole-genome exon expression in 102 normal and cancer tissue samples of different stages from colon, urinary bladder, and prostate. We identified 2069 candidate alternative splicing events between normal tissue samples from colon, bladder, and prostate and selected 15 splicing events for RT-PCR validation, 10 of which were successfully validated by RT-PCR and sequencing. Furthermore, 23, 19, and 18 candidate tumor-specific splicing alterations in colon, bladder, and prostate, respectively, were selected for RT-PCR validation on an independent set of 81 normal and tumor tissue samples. In total, seven genes with tumor-specific splice variants were identified (ACTN1, CALD1, COL6A3, LRRFIP2, PIK4CB, TPM1, and VCL). The validated tumor-specific splicing alterations were highly consistent, enabling clear separation of normal and cancer samples and in some cases even of different tumor stages. A subset of the tumor-specific splicing alterations (ACTN1, CALD1, and VCL) was found in all three organs and may represent general cancer-related splicing events. In silico protein predictions suggest that the identified cancer-specific splice variants encode proteins with potentially altered functions, indicating that they may be involved in pathogenesis and hence represent novel therapeutic targets. In conclusion, we identified and validated alternative splicing between normal tissue samples from colon, bladder, and prostate in addition to cancer-specific splicing events in colon, bladder, and prostate cancer that may have diagnostic and prognostic implications. Molecular & Cellular Proteomics 7:1214-1224, 2008.

Alternative splicing is a key component in expanding a relatively limited number of genes into very complex proteomes. It has been estimated that about three-quarters of all human genes undergo alternative splicing (1–3), which may affect function, localization, binding properties, and stability of the encoded proteins (4). The recent results from the ENCODE (Encyclopedia of DNA Elements) consortium (5) extend and confirm the ubiquity of alternative splicing (6). Several splice variants with antagonistic functions have been described, e.g. BCL-X has an antiapoptotic long isoform and a proapoptotic short isoform (7, 8). Alternative splicing can also lead to degradation of the transcript, thereby abrogating protein expression; examples include certain Serine/Arginine-rich (SR) protein splicing factors for which the inclusion of a particular exon causes mRNA degradation by nonsense-mediated decay (9, 10).

Single nucleotide polymorphisms and somatic splice site mutations leading to aberrant splicing patterns have been described for a number of tumor suppressor genes, including APC, TP53, and BRCA1 (11). Deregulation of trans-acting proteins, such as splicing factors and heterogeneous nuclear ribonucleoproteins, may cause a more general change in RNA splicing in cancer cells. The SFRS1 gene, encoding the splicing factor 2/alternate splicing factor (SF2/ASF), was recently described as a proto-oncogene (12), indicating the importance of alternative splicing in cancer development.

Cancer-specific splice variants may potentially be used as diagnostic, prognostic, and predictive biomarkers as well as therapeutic targets, making identification of these events highly relevant (13). Splicing studies have traditionally focused on single genes, but recently several approaches using microarrays have been applied successfully (1, 14–16). The GeneChip® Human Exon 1.0 ST Array investigates the expression of virtually all known and many predicted human exons (~1 million) allowing genome-wide evaluation of splic-
ing events. \(^1\) We used this exon array to identify tissue- and
tumor-specific alternative splicing in normal and cancer tis-
ues by analyzing exon expression in more than 100 samples
from colon, bladder, and prostate. The strategy was to screen
for splice variants expressed commonly in normal tissues and
cancers, not to identify all possible splice variants.

The identified splicing alterations were validated by RT-
PCR and sequencing, which confirmed 10 genes with differ-
entially expressed splice variants between the three types of
normal tissue and seven genes between normal and cancer
tissues, respectively. Furthermore we performed an in-depth
validation using RT-PCR on a set of 83 independent samples.
A quantitative PCR was established for three of the cancer-
specific markers and proved that these could be assayed with
precision in tumor tissue, thereby serving as a biomarker for
cancer. \textit{In silico} protein predictions indicated that encoded
protein isoforms could have potentially altered function that
may serve as novel treatment targets and cancer markers.

**EXPERIMENTAL PROCEDURES**

**Clinical Samples for Array Analysis and RT-PCR Validation**—A total
of 185 tissue samples from colon, prostate, and urinary bladder were
used for exon array analysis and RT-PCR validation. Samples were
randomly divided into two independent sets: one for exon array
identification (102 samples) and one for validation of candidates by
RT-PCR (83 samples). Tumor-specific splicing candidates were vali-
dated on 81 samples from the validation set, and tissue-specific
splicing candidates were validated on a total of 32 normal samples.
The exon array identification set consisted of 10 normal colon and 20
colon adenocarcinoma samples, 10 normal prostate and 15 prostate
cancer samples, and 11 normal bladder and 36 bladder tumor samples.
The independent sample set for validation of tumor-specific splicing consisted of 26 colon samples (eight normal samples, six
adenomas, six stage II adenocarcinomas, and six stage III adenocar-
cinomas), 25 prostate samples (10 normal, seven localized tumor samples, and eight primary tumors from patients with metastases),
and 30 bladder samples (five normal samples, eight Ta tumors, eight
T1 tumors, and nine T2 tumors). Alternative splicing differences be-
tween normal tissues were RT-PCR-validated on a mixed sample set
consisting of samples from the identification set and samples from
the independent validation set (for colon five plus five, for bladder six
plus four, and for prostate 12 independent samples, respectively).
Laser microdissections were performed on normal and cancer biop-
sies from five colon cancer patients (two stage I, two stage II, and one
stage III cancer) (18). Further details about tissue samples are pro-
vided in the supplemental Experimental Procedures. Immediately af-
after surgery, tissue samples were embedded in Tissue-Tek\textsuperscript{TM}
(Optimal Cutting Temperature) Compound (Sakura Finetek) (colon
and prostate) or placed in a solution containing guanidinium isothio-
cyanate and SDS (bladder) and then snap frozen in liquid nitrogen and
stored at \(-80 °C\) (19). Informed consent was obtained from all pa-
tients, and the study was approved by the Scientific Ethics Commit-
tee of Aarhus County.

**RNA Extraction and Array Hybridization**—Total RNA from colon,
bladder, and prostate samples was purified from serial cryosections.
The first and last sections were hematoxylin- and eosin-stained to
evaluate the tissue composition, and samples with a high tumor cell content (>75% for bladder and colon and >50% for prostate) were
selected. Total RNA was purified using RNeasy MinElute columns
following the manufacturer's instructions (Qiagen); bladder samples
were, because of sampling procedures, extracted using the RNAzol B
RNA isolation method (Wak-Chemie Medical GmbH). The RNA quality
was verified by analysis on the 2100 Bioanalyzer (Agilent), and sam-
ple s with a 28S/18S ratio <1.0 and RNA integrity number <7 were
excluded. One microgram of total RNA was labeled according to
the GeneChip Whole Transcript (WT) Sense Target Labeling Assay as
provided by the manufacturer (Affymetrix) and hybridized to Human
Exon 1.0 ST Arrays (Affymetrix) overnight before scanning in an
Affymetrix GCS 3000 7G scanner. All 102 samples were labeled
and scanned in a randomized order to avoid batch effects.

**Samples and Method for Estimation of Variation**—The following RNA samples were used to estimate variation. (i) Three RNA samples
from three replicate cultures of the colon cancer cell line LS174T (20)
were used to assess intercell line variation. RNA sample 1 was further
split into three replicates to allow technical variation estimates. (ii) For
each of the three patients, three biopsies of different location from within
the same tumor were used to assess intratumor heterogeneity and
interumor variation. (iii) Nine colon cancer stage II tissue samples and
nine matched normal colon samples from nine patients were used to
estimate variation in normal and tumor samples. Expression values
were obtained using ArrayAssist (see “Data Analysis” below). Only
probe sets that were detected above background (defined by detec-
tion above background p < 0.05) on more than half of the arrays
within a tissue or cell line were selected for further analysis. Exon-
specific variances were calculated for each group (intracell line, in-
tercell line, intrapatient, and interpatient) of arrays. The dominant-
negative TCF1-inducible LS174T-derived cell lines were described
previously (20). Cell culture, harvesting, and RNA extraction were
performed as described previously (21).

**Data Analysis**—Exon array data files were loaded into the Array-
Assist exon software (Stratagene Software Solutions), and samples
were quantile-normalized using ExonRMA\textsuperscript{2} with core probe sets
(228,940 probe sets) and antigenomic background probes.\textsuperscript{1} For vari-
ance stabilization, 16 was added to probe set intensity values before
transformation to a log\(_2\) scale. Transcript (gene) level expression was
calculated using TranscriptRMA. A splice index (SI) was calculated for
all probe sets (SI = log\(_2\) (probe set intensity/transcript expression
level)). Normal colon sample number 10 was identified as an outlier
and excluded from subsequent analyses. Statistical testing using a t
test or ANOVA on the SI was performed, and p values were used as a
ranking tool as described previously (22). Data were subjected to
several restrictive layers of filtering as follows. Transcripts without
probe sets differing significantly (t test score >0.05) between the
sample groups (normal and cancer) were omitted from the analysis
along with genes without probe sets detected above background
(defined by detection above background p < 0.05) in at least half of
the samples. Only genes expressed in both sample groups were
included in the analysis (transcript expression level >64). The delta
splice index (\(\Delta SI = \text{mean } \log\(_2\) \text{ SI group } 1 \text{ minus mean } \log\(_2\) \text{ SI group } 2\)) was
calculated, and in all normal versus cancer analyses, it was required
that \(\Delta SI < -0.5 \) or >0.5. Probe sets fulfilling all filtering criteria
were ranked based on their SI p values, and the top 300 candidates
in each tissue were selected for manual inspection. In the analysis of

\(^1\) Affymetrix white papers: Exon Probeset Annotations and Trans-
script Cluster Groupings v1.0 and Exon Array Background Correction
v1.0.

\(^2\) The abbreviations used are: RMA, robust multichip average; SI,
splice index; ANOVA, analysis of variance; VCL, vinculin; LRRFIP2,
leucine-rich repeat (in FLII)-interacting protein; ACTN1, actinin, \(\alpha_1\);
CALD1, caldesmon 1; COL6A3, collagen, type VI, \(\alpha_3\); PIK4CB,
phosphatidylinositol 4-kinase, catalytic, \(\beta\) polypeptide; TPM1, tropo-
myosin 1.

\* Alterative Splicing in Colon, Bladder, and Prostate Cancer

\* Molecular & Cellular Proteomics 7.7 1215
normal samples, we required that the transcript was expressed in all three tissue types, that the ANOVA p value be <0.005, and that the ΔSI be <−0.8 or >0.8 resulting in 2069 transcripts fulfilling the sample criteria. The selected transcripts were finally manually filtered to evaluate the probe set expression across all samples. Probe sets with low variation, cross-hybridization potential, or low correlation to the transcript were filtered out. Finally selected transcripts were inspected in the University of California, Santa Cruz genome browser (23) to localize and describe the alternative splicing event(s). We primarily focused our search on cassette exons or, in rare cases, on genes with multiple skipped exons. No alternative splicing events in the extreme 5′- or 3′-ends of transcripts were selected for RT-PCR validation because of the experimental validation setup.

RT-PCR, Real Time RT-PCR, and Sequencing—cDNA was generated by reverse transcription of 1 μg of total RNA using SuperScript II reverse transcriptase (Invitrogen) and oligo(dT) or random nonamer primer. PCR using 1 μl of 20-fold diluted cDNA was performed for 40 cycles with the Expand High Fidelity PCR System (Roche Diagnostics GmbH) or for 36 cycles with TEMPase Hot Start DNA polymerase (Ampliqon). PCR products were analyzed on 2–3% agarose gels. The RT-PCR covering exon 6 of tropomyosin 1 (TPM1) gave rise to two similar sized PCR products, depending on the presence of exon 6a or 6b, both with a length of 76 bp. To be able to discriminate between these exons, we digested the PCR product with PstI, which cleaves exon 6a, but not exon 6b, resulting in two DNA fragments of 109 and 132 bp only when exon 6a was present. Distinct gel bands were purified with the QiAquick Gel Extraction kit (Qiagen), sequenced (same primers as used for PCR) using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), and analyzed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). PCR bands that were insufficiently separated were cloned using the TOPO TA Cloning kit for sequencing (Invitrogen), and positive transformants were sequenced (using a T3 sequencing primer) as described above. Ratios between splice variants were determined by densitometry using ImageJ software (24). Real time RT-PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) on the colon samples from the independent validation set. All reactions were run in triplicate on a 7900HT Fast Real-Time PCR System (Applied Biosystems) with quantification based on standard curves. For each 10-μl reaction, 3 pmol of each primer pair and 2 μl of 20-fold diluted cDNA were used. All primer sequences are listed in supplemental Table 1.

Bioinformatics Analysis of Differentially Expressed Protein Features—Protein features (experimentally determined and predicted) were mapped onto the translated alternative transcript sequences to identify putative functional changes resulting from the alternative splicing events linked to cancer progression. First the peptides corresponding to the alternative exon (with and without the flanking peptides spanning the exon-exon junctions) were aligned against the Protein Data Bank by PSIBLAST (25). Then FeatureMap3D (26), PyMOL (27), and scripts created for this analysis were used to analyze and visualize the placement of structures and/or conserved domains. A wide range of public databases, in addition to a large data warehouse consisting of essentially all publicly available, experimentally determined protein-protein interactions, including interactions inferred by orthology, were checked in relation to the binding properties of chain, segments, and domain substructures. PSIPred (28) was used to predict secondary structures on full and alternative peptide sequences. Hydrophobicity plots were based on Kyte-Doolittle and Hopp-Woods scales to predict potential hydrophilic regions most likely exposed on the protein surface. Phosphorylation sites were predicted by NetPhos (29) in particular on the alternative exon peptides. Furthermore we made several other protein feature predictions, for example signal peptide cleavage site prediction (30), propeptide cleavage sites, and N-glycosylation and O-glycosylation sites (31).

![Fig. 1. Box-and-whiskers plot of the technical variation and variation within colon cancer cell lines, colon tumors, and normal colon mucosa in exon array measurements. A, variance estimate distributions of three technical replicates (technical), inter-cell line variation (Inter cell), three samples), intrapatient tumor variation (Intra Pt, nine samples), and interpatient tumor variation (Inter Pt, nine samples). B, variance estimate distributions of normal colon samples and colon tumor samples removed as pairs from nine patients. Boxes represent 25–75th percentiles, and whiskers represent the 0–25th and 75–100th percentiles. The median is marked by a horizontal line across each box. Note that variance estimate distributions in A and B cannot be directly compared because all 12 RNA samples used for calculations in A were labeled and scanned in a single batch; this was not the case for the 18 samples used in B.

ProtFun was used to compare protein functions that potentially change between alternative isoforms.

RESULTS

To identify tissue- and tumor-specific alternative splicing patterns, global exon expression was examined in a total of 102 normal and cancer tissue samples from three different organs: colon, urinary bladder, and prostate. We used the novel Human Exon 1.0 ST Array to measure the expression of more than one million exons in a full genome-wide screening for alternative splicing events. Initially to assess the magnitude of biological and technical variation of exon expression levels, samples from a colon cancer cell line, normal colon mucosa, and colon cancer tissues were analyzed (Fig. 1). As expected, the smallest variation was seen between technical replicates followed by the variation within a cell line, variation within a single tumor, and variation between tumors (Wilcoxon signed rank tests, \( p < 2.22 \times 10^{-16} \)) (Fig. 1A). A similar analysis of matched normal and stage II colon cancer samples (included in the exon array analysis) showed the variation between normal colon samples to be significantly smaller than the variation between colon cancer samples (\( p < 2.22 \times 10^{-16} \)) (Fig. 1B).
Alternative Splicing in Normal Tissues—Tissue-specific alternative splicing events were identified by exon array analysis of nine colon, 10 prostate, and 11 bladder normal tissue samples. By applying an ANOVA statistical analysis to splice indices along with several filtration criteria (see “Experimental Procedures”), 2069 genes with candidate alternative splicing events were identified (data not shown). After filtering and inspection to characterize the candidate splicing events, 15 splicing events, covering cassette exons, intron retention, multiple exon skipping, and mutually exclusive exons, were selected for validation in all three tissues by RT-PCR on a mixed sample set consisting of independent samples and samples used for exon array analysis. PCR products were sequenced to verify individual splice variants. For 10 of the 15 candidates, distinct alternative splicing patterns between at least two tissue types were confirmed (Fig. 2 and Table I). In half of the cases (CLSTN1, MCM7, TPM1, AKAP13, and CD44), splicing patterns were similar in prostate and bladder but differed from the pattern in colon samples. For the other five genes (NR4A1, MRRF, AUP1, TCF12, and CTNND1), unique splicing patterns were identified in each of the three tissues.

We identified one novel splice variant of NR4A1 (containing exons 3, 4, and 5 and the intron between exons 3 and 4), which was expressed only in prostate. A novel splice variant of MRRF (containing alternative exons 4 and 5 and the intron between exons 3 and 4) was also identified. We noted that for several of the candidate genes examined more bands than predicted from known splicing events appeared on the gel; some of these could not be sequenced and most likely represented minor splice variants or RT-PCR artifacts. * designates novel NR4A1 and MRRF splice variants. The lane designated M contains a 100-bp ladder.

**Fig. 2.** RT-PCR validation of alternatively spliced genes in normal samples from bladder, prostate, and colon. RT-PCR was performed with primers hybridizing to neighboring constitutively expressed exons, and the PCR products were resolved on agarose gels. Gene symbols are listed to the left of each gel, and sequence-validated exon structures are shown to the right with primer positions indicated by arrows. TPM1 PCR products were cleaved by PstI to distinguish between two mutually exclusive exons of similar size. Unmarked bands probably represent minor splice variants or RT-PCR artifacts. * designates novel NR4A1 and MRRF splice variants. The lane designated M contains a 100-bp ladder.
We were unable to validate tissue-specific alternative splicing events for five of the 15 candidates by RT-PCR. Although three of these showed alternative splicing as indicated by multiple bands on agarose gels, no consistent changes in splicing patterns between the tissues were found (data not shown). RT-PCR analysis of the two other candidates only gave rise to a single PCR product for all samples tested (data not shown).

**Alternative Splicing in Normal Versus Cancer Samples**—Tumor-specific alternative splicing events were identified by exon array analysis of normal and tumor tissue samples from colon, bladder, and prostate. Twenty-three candidate alternative splicing events were selected for validation from colon, 19 were selected for validation from bladder, and 18 were selected for validation from prostate (supplemental Table 1). At first, these candidate alternative splicing events were subjected to validation by RT-PCR and sequencing using a subset of the same RNA samples that were also analyzed on exon arrays. In colon, six of 23 (26%) splicing events were confirmed by RT-PCR; two of these were novel, whereas four were reported previously in a study of alternative splicing in colon cancer (33) suggesting the robustness of the alternative splicing detection method used. In bladder and prostate, six of 19 (32%) and five of 18 (28%) splicing events, respectively, were confirmed. All validated splicing events were found in at least two tissue types, and of 46 unique candidate genes, a total of seven alternative splicing events could be validated by RT-PCR (Table II). Included in the seven validated candidates is a novel splice variant of leucine-rich repeat (in FLII)-interacting protein (LRRFIP2) containing exons 5 and 6 but lacking exon 4 and exons 7–15. To further test the robustness of the seven identified alternative splicing events (in actin, α1 (ACTN1), caldesmon 1 (CALD1), collagen, type VI, α3 (COL6A3), LRRFIP2, phosphatidylinositol 4-kinase, catalytic, β
polypeptide (PIK4CB), TPM1, and vinculin (VCL)), we validated them by RT-PCR and sequencing on an independent sample set consisting of 81 normal and cancer tissue samples of different stages from colon, bladder, and prostate (Fig. 3).

CALD1, encoding an actin- and calmodulin-binding protein (34), displayed an identical splicing pattern in all three types of cancers. A long CALD1 isoform, including an extended form of exon 5 (via a downstream alternative 5’/H11032 splice site) and exon 6, was absent or reduced in bladder, colon, and metastatic prostate cancer but also missing from two of five normal bladder samples. Localized prostate cancers showed an intermediate splicing pattern between that of normal and metastatic samples (Fig. 3A).

VCL encodes a cytoskeletal protein associated with cell-cell and cell-matrix junctions (35). The splicing pattern of VCL was similar to that of CALD1: the long variant was expressed in most normal samples but absent in the majority of cancers. In prostate, the long isoform was down-regulated in metastatic prostate cancer compared with localized cancer (Fig. 3A). ACTN1 encodes an actin-binding cytoskeletal protein (35) that contains two mutually exclusive exons. Both isoforms were present in virtually all samples analyzed. Alternative splicing events were therefore reflected as changes in the ratio between the two isoforms, and in all three tissue types, the long isoform predominated in cancer samples (Fig. 3A).

TPM1 encodes an actin-binding protein with numerous splice variants (36). In addition to the tissue-specific splicing patterns as identified in normal samples from colon, bladder, and prostate (Table I and Fig. 2), our exon array data indicated tumor-specific alternative splicing of TPM1 in both prostate and bladder cancers (Fig. 3A) as evidenced by two mutually exclusive exons, 6a and 6b, with opposing ΔSI values (Table II). In prostate, normal tissue expressed only the PstI-uncleaved variant (including exon 6b), whereas many localized prostate cancers expressed the PstI-cleaved variant (including exon 6a) as well. The majority of metastatic prostate cancer samples expressed the cleaved variant. In bladder, three of five normal samples only expressed the uncleaved variant, and the remaining normal and all cancer samples expressed both variants.

COL6A3 encodes a protein of the extracellular matrix (37). In colon and bladder, the long isoform was nearly absent from normal samples, whereas it was expressed in almost all tumor samples. The long isoform was present in nearly half of the metastatic prostate cancer samples (Fig. 3B).

**Fig. 3.** RT-PCR validation of differentially spliced genes between normal and cancer tissues. RT-PCR validations on normal and tumor tissue samples from bladder, prostate, and colon were performed with primers hybridizing to neighboring constitutively expressed exons, and the PCR products were resolved on agarose gels. Gene symbols are listed to the left of each gel, and sequence-validated exon structures are shown to the right with primer positions indicated by arrows. Red arrows show heteroduplexes confirmed by subcloning followed by sequencing. Unmarked bands probably represent minor splice variants or RT-PCR artifacts, A, alternative splicing events identified in all three tissue types. ACTN1, RT-PCR validation (upper panel) and the ratio between the intensity of the two bands (lower panel). B, alternative splicing events identified in two of the three tissue types. TPM1, PCR products were cleaved by PstI to distinguish between two mutually exclusive exons of similar size. * designates a novel LRRFIP2 splice variant. PIK4CB, RT-PCR validation (upper panel) and the ratio of the two bands (lower panel). Lanes designated M contain a 100-bp ladder. PC, localized prostate cancer; MPC, metastatic prostate cancer.
**LRRFIP2** encodes a protein that activates Wnt signaling (38). **LRRFIP2** was identified here as a candidate gene for alternative splicing in colon and prostate cancer (Table II). The long and medium isoforms containing exons 5 and 6 or exon 6, respectively, predominated in normal colon samples, whereas the ratio was shifted toward a shorter isoform lacking both of these exons in adenomas and colon cancer samples. Metastatic prostate cancer samples displayed slightly less of the long variant compared with normal and localized prostate cancer samples (Fig. 3B).

**PIK4CB** is part of the phosphatidylinositol 4-kinase family, which generates phosphatidylinositol 4-phosphate, the precursor of phosphoinositides, which are important for the activity and recruitment of many signaling proteins on cellular membranes (39). Validation by RT-PCR showed that both isoforms were present in all samples (Fig. 3B). In bladder and colon normal samples as well as colon adenomas, these isoforms were expressed at roughly equal levels, whereas the short isoform was predominant in most cancer samples from both tissues (Fig. 3B).

To assess the validity of our RT-PCR results, we further analyzed the expression of three candidate genes (**COL6A3**, **TPM1**, and **ACTN1**) using a quantitative real time RT-PCR approach. The ratios and band patterns from the RT-PCR and the quantitative real time RT-PCR data for these genes correlated remarkably well (Fig. 4) supporting that the RT-PCR approach we used for validation of alternative splicing is essentially semiquantitative.

All RNA preparations in this study were extracted from trimmed tissue sections aiming at more than 50% (prostate)
or 75% (colon and bladder) cancer cells. To verify that the alternative splicing events we observed were indeed cancer cell-specific, we extracted RNA following laser microdissection of five matched normal and cancer colon samples. Because of limited amounts of RNA, the splicing patterns of only two genes, LRRFIP2 and ACTN1 (Supplemental Fig. S1), were analyzed by RT-PCR. The splicing patterns of both genes corresponded to the pattern seen in the trimmed tissue samples indicating that they reflect cancer cell-specific splicing events.

In Situ Prognosis Predictions of Cancer-specific Splice Variants—Based on current literature, the potential effects of the seven cancer-specific splicing events on protein structure and function remain uncertain. To address this question, we analyzed individual splice variants using various in silico protein function and structure prediction tools (see “Experimental Procedures”). The analyses predicted that all identified transcript variants can be translated and that different protein isoforms are likely to differ in their functional properties.

For VCL, exon 19 was skipped in most cancer samples (Fig. 3). This cassette exon encodes a peptide in the C-terminal tail of VCL that presumably is located at the protein surface from where it could participate in protein-protein interactions (Fig. 5). In addition, VCL exon 19 encodes two strong potential phosphorylation sites that could serve important regulatory functions. In TPM1, the switch between two mutually exclusive exons (6a and 6b) alters a lysine to an arginine residue potentially weakening the bonding between associated polypeptide chains by one hydrogen bond (from two to one). Furthermore the exon 6a- and 6b-encoded peptide fragments each host a unique predicted phosphorylation site. For LRRFIP2, three splice variants differing in their inclusion or skipping of exons 5 and/or 6 were observed (Fig. 3). These exons contain five predicted putative serine phosphorylation sites and one putative O-glycosylation site, which could modulate LRRFIP2 protein function. We identified three CALD1 isoforms differing in their inclusion of an exon 5 extension and exon 6 (Fig. 3). We predicted five phosphosites in the alternative exon 5 extension along with a putative propeptide cleavage site, which has a potential critical biological function. COL6A3 exon 6 is a cassette exon that is 201 amino acids long and most likely encodes a von Willebrand factor domain. It contains two serine, two threonine, and three tyrosine predicted phosphorylation sites, which have potential regulatory effects on protein function. Exon 3 of PIK4CB is a cassette exon that contains five putative serine phosphorylation sites even though it is only 15 amino acids long. Furthermore exon 3 is hydrophilic and, thus, likely to be exposed on the surface. Neither the CALD1, COL6A3, or PIK4CB cassette exons nor the junction peptides have significant Protein Data Bank hits.

ACTN1 contains an actin-binding domain in the N-terminal part; however, the two mutually exclusive exons (19a and 19b) identified in the present study are located in the C terminus and have no significant Protein Data Bank hits. Also predictions of ACTN1 secondary structure and known phosphorylation sites provided no clear clues to possible effects of alternative splicing on ACTN1 function. In conclusion, in silico protein predictions disclosed several potential structural differences between VCL, TPM1, LRRFIP2, CALD1, COL6A3, and PIK4CB splice variants that may significantly alter their function and interaction partners and, thereby, support a possible role in carcinogenesis.

DISCUSSION

In the present study, we used a novel microarray technology that measures the expression of single exons to identify
alternative splicing events between normal tissues from colon, urinary bladder, and prostate or between normal and cancer tissues from these organs. Several other microarray platforms have been used to detect alternative splicing, including arrays that utilize exon junction probes (1, 14, 15). Exon junction arrays are powerful tools, but because junction probes usually target known splicing events, they are less suitable for detection of novel splice variants. Genome-wide alternative splicing between panels of normal tissue types have been analyzed previously using a prototype exon microarray (2) and exon junction microarrays (1). Thousands of candidate alternative splicing events were identified, and a subset was RT-PCR-validated with success rates of 86% (72 of 86) (2) and 49% (73 of 150) (1), respectively.

Here we identified ~2000 transcripts containing candidate alternative splicing events between normal tissues from colon, bladder, and prostate. Following a restrictive set of filtering criteria, 15 of these were selected for validation by RT-PCR and sequencing, and the success rate (67%) was comparable to what was reported previously (1, 2). Tissue-specific alternative splicing between colon, bladder, and prostate was discovered for half of the genes (NR4A1, MRRF, AUP1, TCF12, and CTNND1). For the other half (CLSTN1, MCM7, TPM1, AKAP13, and CD44), bladder and prostate, having an embryological origin from the urogenital sinus, resembled each other but differed from normal colon mucosa. Alternative splicing in cancer samples from colon, bladder, and prostate, as compared with normal tissue, was identified for a total of seven genes (CALD1, VCL, ACTN1, TPM1, COL6A3, LRRFIP2, and PIK4CB) with a validation rate within each tissue type ranging from 26 to 32%.

The relationship between cancer and alternative splicing is well established for several mutations affecting cis-acting splicing signals (40). Altered splicing in cancer could also be due to a generalized lack of fidelity in the splicing machinery, e.g. the SFRS1 gene encoding the splicing factor 2/alternate splicing factor (SF2/ASF) has recently been described as a proto-oncogene (12). Interestingly the alternative splicing we identified in cancer tissues was not cancer type-specific: the same splicing pattern always occurred in at least two cancer tissue types, and three candidates (CALD1, VCL, and ACTN1) even displayed similar alternative splice patterns in all three cancer tissues. This indicates a general loss of splicing fidelity in the cancer samples. Although alternative splicing events may not be the driving forces in cancer development, some of these could still promote cancer development and thereby be expanded by positive clonal selection (11). In support of this, we see a clear relationship between advanced cancer stage and systematic occurrence of alternative splicing isoforms. This was most obvious in prostate cancers where differences in the splicing patterns of TPM1, ACTN1, CALD1, and VCL were clearly evident not only between normal and cancer samples but also between samples from localized tumors and tumors with metastases. In colon adenomas, the splicing patterns of the tumor-specific candidate genes, with the exception of PIK4CB and CALD1, resembled the splicing patterns of the more advanced cancer stages, suggesting that deregulation of splicing takes place early in colon cancer development. In bladder cancer, the change in splicing pattern of ACTN1 was more pronounced in T2 tumors compared with Ta tumors. These examples indicate that some of the identified splice variants could be driving forces in cancer development.

Remarkably several of the cancer-associated candidates are related to the cytoskeleton. We identified variants of CALD1 (34), which links myosin and actin filaments; TPM1, an actin-binding tumor suppressor with apoptosis-promoting function (36); and ACTN1 and VCL, both encoding cytoskeletal elements that participate in the organization of the cytoskeleton by interacting with actin and with each other (35). Whether this group of proteins is more prone to splicing variation without functional effects or whether the malignant process selects for these variants is unknown. However, they are extremely overrepresented, constituting more than half of all detected variants. The actin cytoskeleton plays a key role in cell motility processes, and a deregulated actin organization has been linked to tumor invasion and metastasis (41, 42). In silico protein predictions indicated that alternative splicing has the capability of changing the structure and, potentially, the binding capacity of VCL, CALD1, and TPM1. Especially the VCL cassette exon has the potential to affect protein interactions as it encodes a peptide that most likely is located on the surface of the protein where it interacts with other proteins. COL6A3 is involved in cell anchoring and remodeling of the extracellular matrix, and in ovarian cancer the expression of COL6A3 is associated with tumor grade and contributes to cisplatin resistance (43). LRRFIP2 has been found to activate Wnt signaling, which is crucial in colon cancer development (38). Also it has been shown that splicing factor-1 (SF1) is regulated by the Wnt pathway (44). In silico predictions indicated that the LRRFIP2 and COL6A3 alternatively spliced exons contain several phosphorylation sites that might influence protein function. The effects of the identified splice variants, however, are still unclear. A previous exon array study, which analyzed alternative splicing in colon cancer, revealed nine genes that were differentially spliced between 10 normal and 10 matched colon-cancer samples, including ACTN1, VCL, COL6A3, TPM1, and CALD1 (33), in agreement with our findings.

The total number of detected splice variants between normal and cancer samples was comparable to other reports (33, 45). The number was relatively low, which could be due to our conservative approach where only splice variants that occurred in most of the samples were identified; the stringent filtering applied to the data also lowered the number of splice variants detected. The validation rates of the cancer-specific candidates, within each tissue type ranging from 26 to 32%,
were relatively low compared with the validation rate for tissue-specific splice candidates. This reflects the larger variation among tumor samples seen in the array data caused by heterogeneous biopsies composed of cancer cells, immune cells, endothelial cells, histiocytes, etc. Furthermore, tumor heterogeneity caused by accumulation of secondary genetic alterations may also lead to a heterogeneous splicing pattern. In general, we found larger and more significant differences in splice indices between normal samples from colon, bladder, and prostate than between normal and cancer samples; this may also contribute to the higher validation rate observed here.

To minimize the proportion of different cell types being represented besides normal epithelium and cancer cells, RNA from all colon and prostate samples and the independent bladder samples was extracted from tissue sections where we observed the tissue composition from stained neighboring sections. This allowed us to crudely dissect the tissue, thereby avoiding plaques of infiltrating lymphocytes, muscle tissue, etc. Inclusion of data from five laser-microdissected normal colon and colon cancer samples also confirmed our findings for the two candidates LRRFIP2 and ACTN1. We validated the splicing candidates by RT-PCR, allowing us to amplify several isoforms in the same reaction using the same primer pairs. This approach also detected unpredicted isoforms, e.g., LRRFIP2, for which we identified an isoform that had not been described previously. In general, our RT-PCR results were robust even under different experimental setups. Analysis of COL6A3, ACTN1, and TPM1 by real time RT-PCR further confirmed the initial semiquantitative findings.

Cancer-specific alternatively spliced mRNAs and protein isoforms may be used as cancer biomarkers. The long isoform of COL6A3 is expressed almost exclusively in cancer samples and could potentially serve as a new cancer marker. In addition, TPM1 exon 6a is exclusively included in the prostate cancer samples and preferentially included in bladder cancer samples, making it a potential cancer marker in these tissues. Detection of these cancer-specific isoforms in urine or feces could help early diagnosis of bladder or colon cancer, respectively. Further characterization of the identified splice variants, along with other markers, could also improve the staging of cancer samples as several of the identified splice variants change their expression pattern with progression of the disease. Indeed splice variants have been used in prostate tumor classification (17). One may also hypothesize that the cancer-specific alternatively spliced isoforms could be used as cancer-specific drug targets. This could be achieved by targetting the protein or silencing the specific mRNA variant by RNA interference. The use of the normal, non-cancer-related splice forms as therapeutic proteins is also a possibility. In conclusion, we detected novel tissue- and cancer-related splicing in a number of genes and validated these on independent sample sets, indicating the potential importance of the variants for cancer cell biology and the potential utilization of the cancer-specific variants as biomarkers and drug targets.

Acknowledgments—The excellent technical assistance of Hanne Steen, Bente Devanté, Pamela Celis, Inge-Lis Thorsen, Lisbet Kjeldsen, Karen Bihl, and Gitte Hoj is gratefully appreciated. We thank Dr. Hans Clevers (The Hubrecht Laboratory, Utrecht, The Netherlands) for the inducible LS174T-derived cell lines.

* This work was supported by The John and Birthe Meyer Foundation, The Danish Cancer Society, The Danish Research Council for Health, The Hede Nielsen Foundation, The Novo Nordisk Foundation, and the European Union Biosapiens project. 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.

‡‡ To whom correspondence should be addressed. Tel.: 45-89495100; Fax: 45-89496018; E-mail: orntoft@ki.au.dk.

REFERENCES

1. Johnson, J. M., Castle, J., Barrett-Engle, P., Kan, Z., Loehr, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R., and Shoemaker, D. D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302, 2141–2144.

2. Clark, T. A., Schweitzer, A. C., Chen, T. X., Staples, M. K., Lu, G., Wang, H., Williams, A., and Blume, J. E. (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol. 8, R64.

3. Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., Tammana, H., and Gingeras, T. R. (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. 14, 331–342.

4. Stamm, S., Ben-Ari, S., Rafaelska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A., and Soreq, H. (2005) Function of alternative splicing. Gene (Amst.) 344, 1–20.

5. The ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816.

6. Tress, M. L., Martelli, P. L., Frankish, A., Reeves, G. A., Wesselink, J. J., Yeats, C., Olason, P. L., Albrecht, M., Hegyi, H., Giorgetti, A., Raimondo, D., Lagarde, J., Laskowska, R. A., Lopez, G., Sadowski, M. I., Watson, J. D., Farissi, P., Rores, I., Nagy, A., Kai, W., Stirling, Z., Orsini, M., Assenov, Y., Blankenburg, H., Huthmacher, C., Ramirez, F., Schlicker, A., Deneoud, F., Jones, P., Kerrien, S., Orchard, S., Antonarakis, S. E., Reymond, A., Birney, E., Brunak, S., Casadio, R., Guigo, R., Harrow, J., Herijjakob, H., Jones, D. T., Lengauer, T., Orengo, C. A., Patthy, L., Thornton, J. M., Tramontano, A., and Valencia, A. (2007) The implications of alternative splicing in the ENCODE protein complement. Proc. Natl. Acad. Sci. U. S. A. 104, 5495–5500.

7. Boise, L. H., Gonzalez-Garcia, M., Postema, C. E., Ding, L., Lindsten, T., Turk, L. A., Mao, X., Nunez, G., and Thompson, C. B. (1993) bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell 74, 597–608.

8. Schwcr, C., and Schulze-Osthoff, K. (2005) Regulation of apoptosis by alternative pre-mRNA splicing. Mol. Cell 19, 1–13.

9. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., and Brenner, S. E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446, 926–929.

10. Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O’Brien, G., Shiue, L., Clark, T. A., Blume, J. E., and Ares, M., Jr. (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. 21, 708–718.

11. Skotheim, R. I., and Nees, M. (2007) Alternative splicing in cancer: noise, functional, or systematics? Int. J. Biochem. Cell Biol. 39, 1432–1449.

12. Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., and Krainer, A. R. (2007) The gene encoding the splicing factor SF2/ASF is a proto-
Alternative Splicing in Colon, Bladder, and Prostate Cancer

13. Pajares, M. J., Ezponda, T., Catena, R., Calvo, A., Pio, R., and Montuenga, L. M. (2007) Alternative splicing: an emerging topic in molecular and cellular oncology. Lancet Oncol 8, 349–357

14. Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A. L., Mohammad, N., Babak, T., Stiu, H., Hughes, T. R., Morris, Q. D., Frey, B. J., and Blencowe, B. J. (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol Cell 16, 929–941

15. Fehlbaum, P., Guihal, C., Bracco, L., and Cochet, O. (2005) A microarray configuration to quantify expression levels and relative abundance of splice variants. Nucleic Acids Res 33, e47

16. Yeakley, J. M., Fan, J. B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M. S., and Fu, X. D. (2002) Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol 20, 353–358

17. Li, H. R., Wang-Rodriguez, J., Nair, T. M., Yeakley, J. M., Kwon, Y. S., Bibikova, M., Zheng, C., Zhou, L., Zhang, K., Downs, T., Fu, X. D., and Fan, J. B. (2006) Two-dimensional transcriptome profiling: identification of messenger RNA isoform signatures in prostate cancer from archived paraffin-embedded cancer specimens. Cancer Res 66, 4079–4088

18. Andersen, C. L., Schepeler, T., Thorsen, K., Birkenkamp-Demtroder, K., Mansilla, F., Aaltonen, L. A., Lauberg, S., and Orntoft, T. F. (2007) Clusterin expression in normal mucosa and colorectal cancer. Mol Cell Proteomics 6, 1039–1048

19. Birkenkamp-Demtroder, K., Christensen, L. L., Olesen, S. H., Frederiksen, C. M., Laiho, P., Aaltonen, L. A., Lauberg, S., Sorensen, F. B., Hagemann, R., and Orntoft, T. F. (2002) Gene expression in colorectal cancer. Cancer Res 62, 4326–4363

20. van de Wetering, M., Sancho, E., Wurtz, T., de Nies, N., van der Woude, W., Krijgsman, A., Luijten, P. R., Mensink, H., de Rond, S., van der Wel, M., de Vries, M., de Vries, W., and van der Velden, P. J. (2003) Clusterin expression in normal colon and colorectal cancer. Cancer Res 63, 344–351

21. Schepeler, T., Mansilla, F., Aaltonen, L. A., Laurberg, S., and Orntoft, T. F. (2007) Gene expression in colorectal cancer. Mol Cell Proteomics 6, 1039–1048

22. Goncalves, J., and Edwards, D. J. (2007) Use of an exon software package to identify global gene splicing and gene expression regulation. American Biotechnology Laboratory, Vol. 25, No. 4, International Scientific Communications Inc., Shelton, CT

23. Kent, W. J., Sgouros, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002) The human genome browser at UCSC. Genome Res 12, 996–1006

24. Abramoff, M. D., Magelhaes, P. J., and Ram, S. J. (2004) Image processing with ImageJ. Biophotonics Int 11, 36–42

25. Attschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402

26. Wernersson, R., Rapacki, K., Staaerfeldt, H. H., Sackett, P. W., and Molgaard, A. (2006) FeatureMap3D—a tool to map protein features and sequence conservation onto homologous structures in the PDB. Nucleic Acids Res 34, W84–88

27. DeLano, W. L. (2002) The PyMOL Molecular Graphics System. Delano Scientific, LLC, Palo Alto, CA

28. Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S., and Jones, D. T. (2005) Protein structure prediction servers at University College London. Nucleic Acids Res 33, W36–38

29. Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol 294, 1351–1362

30. Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2, 953–971

31. Julenius, K., Molgaard, A., Gupta, R., and Brunak, S. (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 15, 153–164

32. Zacharias, D. A., Garaszmegi, N., and Strehler, E. E. (1994) Characterization of persistent artifacts resulting from RT-PCR of alternatively spliced mRNAs. Biotechniques 17, 652–655

33. Gardina, P. J., Clark, T. A., Shimada, B., Staples, M. K., Yang, Q., Veitch, J., Schwartz, A., Awad, T., Sugnet, C., Dee, S., Davies, C., Williams, A., and Turpaz, Y. (2005) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics 7, 325

34. Li, Y., Lin, J. L., Reiter, R. S., Daniels, K., Soll, D. R., and Lin, J. J. (2004) Caldesmon mutant defective in Ca2+-calmodulin binding interferes with assembly of stress fibers and affects cell morphology, growth and motility. J Cell Sci 117, 3593–3604

35. Bois, P. R., Borgon, R. A., Vonrhein, C., and Izard, T. (2005) Structural analysis of α-actinin-vinculin interactions. Mol Cell Biol 25, 6112–6122

36. Varga, A. E., Stourman, N. V., Zheng, Q., Safina, A. F., Quan, L., Li, X., Sosey-Alaoui, K., and Bakin, A. V. (2005) Silencing of the Tropomyosin 1 gene by DNA methylation alters tumor suppressor function of TGF-β. Oncogene 24, 5043–5052

37. Zanussi, S., Doliana, R., Segat, D., Bonaldo, P., and Colombatti, A. (1992) The human type VI collagen gene. mRNA and protein variants of the α3 chain generated by alternative splicing of an additional 5′end exon. J Biol Chem 267, 24082–24089

38. Liu, J., Bang, A. G., Kintner, C., Orth, A. P., Chanda, S. K., Ding, S., and Schultz, P. G. (2006) Identification of the Wnt signaling activator leucine-rich repeat in Flightless interaction protein 2 by a genome-wide functional analysis. Proc Natl Acad Sci U S A 102, 1927–1932

39. Balla, A., and Balla, T. (2006) Phosphatidylinositol 4-kinesins: old enzymes with emerging functions. Trends Cell Biol 16, 351–361

40. Srebrow, A., and Komblittht, A. R. (2006) The connection between splicing and cancer. J Cell Sci 119, 2635–2641

41. Lambrecht, A., Van Troyes, M., and Ampe, C. (2004) The actin cytoskeleton in normal and pathological cell motility. Int J Biochem Cell Biol 36, 1890–1909

42. Yamazaki, D., Kurisu, S., and Takenawa, T. (2005) Regulation of cancer cell motility through actin reorganization. Cancer Sci 96, 379–386

43. Sherman-Baust, C. A., Weeraratna, A. T., Rangel, L. B., Pizer, E. S., Cho, K. R., Schwartz, D. R., Shock, T., and Morin, P. J. (2003) Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cspatin resistance in ovarian cancer cells. Cancer Cell 3, 377–386

44. Shitashige, M., Naishiro, Y., Idogawa, M., Honda, K., Ono, M., Hirohashi, S., and Yamada, T. (2007) Involvement of splicing factor-1 in β-catenin/T-cell factor-4-mediated gene transactivation and pre-mRNA splicing. Gastroenterology 132, 1039–1054

45. French, P. J., Peeters, J., Horsman, S., Duijm, E., Siccama, I., van den Bent, M. J., Luider, T. M., Kros, J. M., van der Spek, P., and Sillevis Smitt, P. A. (2007) Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. Cancer Res 67, 5632–5642