Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

Chromatin, Gene, and RNA Regulation

Molecular Cancer Research

Whole Transcriptome Sequencing Reveals Extensive Unspliced mRNA in Metastatic Castration-Resistant Prostate Cancer

Adam G. Sowalsky[1], Zheng Xia[2], Liguo Wang[2], Hao Zhao[2], Shaoyong Chen[1], Glenn J. Bubley[1], Steven P. Balk[1], and Wei Li[2]

Abstract

Men with metastatic prostate cancer who are treated with androgen deprivation therapies (ADT) usually relapse within 2 to 3 years with disease that is termed castration-resistant prostate cancer (CRPC). To identify the mechanism that drives these advanced tumors, paired-end RNA-sequencing (RNA-seq) was performed on a panel of CRPC bone marrow biopsy specimens. From this genome-wide approach, mutations were found in a series of genes with prostate cancer relevance, including *AR*, *NCOR1*, *KDM3A*, *KDM4A*, *CHD1*, *SETD5*, *SETD7*, *INPP4B*, *RASGRP3*, *RASA1*, *TP53BP1*, and *CDH1*, and a novel *SND1*:*BRAF* gene fusion. Among the most highly expressed transcripts were 10 noncoding RNAs (ncRNAs), including MALAT1 and PABPC1, which are involved in RNA processing. Notably, a high percentage of sequence reads mapped to introns, which were determined to be the result of incomplete splicing at canonical splice junctions. Using quantitative PCR (qPCR), a series of genes (*AR*, *KLK2*, *KLK3*, *STEAP2*, *CPSF6*, and *CDK19*) were confirmed to have a greater proportion of unspliced RNA in CRPC specimens than in normal prostate epithelium, untreated primary prostate cancer, and cultured prostate cancer cells. This inefficient coupling of transcription and mRNA splicing suggests an overall increase in transcription or defect in splicing.

Implications: Inefficient splicing in advanced prostate cancer provides a selective advantage through effects on microRNA networks but may render tumors vulnerable to agents that suppress rate-limiting steps in splicing. *Mol Cancer Res; 13(1); 98–106. ©2014 AACR.*

Introduction

With more than 230,000 new patients and nearly 30,000 deaths annually, prostate cancer is the second most common cause of cancer-related deaths in men in the United States (1). Although greater than 75% of patients with early-stage prostate cancer can be cured with surgical and/or radiation treatment, the remainder ultimately recur with metastatic disease. Androgen deprivation therapy (surgical castration or the administration of luteinizing hormone–releasing hormone agonists) is the standard treatment for metastatic prostate cancer (2), but most tumors eventually relapse despite castrate androgen levels (castration-resistant prostate cancer, CRPC). It has now become clear that androgen receptor (AR) is substantially reactivated in a large proportion of these relapsed tumors through increased intratumoral androgen synthesis, in conjunction with other mechanisms that may enhance AR expression and activity, and many of these tumors will respond to agents that further suppress androgen synthesis (CYP17A1 inhibitors such as abiraterone) or new AR antagonists (such as enzalutamide). Unfortunately, these men generally relapse within 1 to 2 years, and increasing serum PSA in most cases suggests that AR is again active in these resistant tumors.

We reported previously on an analysis of gene expression in CRPC bone marrow metastases using Affymetrix oligonucleotide microarrays and immunohistochemistry, which showed increased expression of enzymes mediating androgen synthesis and alterations in the expression of additional genes linked to tumor progression (3). We hypothesize that additional mechanisms mediating progression to CRPC will also contribute to tumor progression after treatment with new hormonal agents including abiraterone and enzalutamide. Therefore, in this study, we have used paired-end RNA-sequencing (RNA-seq) to assess more comprehensively the transcriptome of 8 CRPC bone marrow metastases that had been examined previously on Affymetrix U133A microarrays.

Materials and Methods

Tissue samples

All tissue samples in this study were obtained with consent from patients with prostate cancer in compliance with the Beth

[1]Division of Hematology and Oncology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts. [2]Division of Biostatistics, Dan L Duncan Cancer Center and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas.

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

A.G. Sowalsky and Z. Xia contributed equally to this article.

Corresponding Authors: Steven P. Balk, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA 02215. Phone: 617-735-2065; Fax: 1-617-735-2050; E-mail: sbalk@bidmc.harvard.edu; and Wei Li, Baylor College of Medicine, 1 Baylor Plaza, Mail Stop BCM305, Cullen Building, Suite 529D, Houston, TX 77030. Phone: 713-798-7854; E-mail: wl1@bcm.edu

doi: 10.1158/1541-7786.MCR-14-0273

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

Israel Deaconess Medical Center Institutional Review Board. CRPC biopsies were obtained from the posterior iliac crest and snap frozen as previously described (3–5). Frozen sections stained with hematoxylin and eosin (H&E) were examined histologically and four to six 6-μm ribbons with >90% tumor and minimal bone marrow elements were treated with TRIzol (Invitrogen) for purification of total RNA.

To obtain nonneoplastic prostate epithelium, we examined snap-frozen samples from radical prostatectomies in patients with low volume prostate cancer and collected sections with 20% to 80% normal prostate epithelium and no evident tumor on histology. DNase-treated RNA was extracted from ten 6-μm ribbons using the RNeasy Plus Micro Kit (Qiagen). To obtain primary prostate cancer, samples from radical prostatectomies were fixed in PaxGene (Qiagen), processed into paraffin, and sectioned at 5 μm onto Arcturus polyethylene naphthalate metal-framed slides (Molecular Machines & Industries). Approximately 50,000 cells in Gleason pattern 3 and 4 glands identified by a board-certified pathologist were captured onto caps using 20-μm infrared pulses and excised from the adjacent tissue using the ultraviolet laser on an ArcturusXT Nikon Eclipse Ti-E microdissection system. DNase-treated RNA was extracted using the PaxGene Tissue RNA Kit (Qiagen).

Library preparation and data analysis

Fifty nanograms of RNA from CRPC samples was prepared for Illumina paired-end sequencing using the Ovation RNA-Seq System (NuGEN), and FastQ files were aligned to the human genome (version Hg19). Complete descriptions of library preparation methods and sequencing data analysis are provided as Supplementary Material.

Cell lines

VCaP and LNCaP cells were obtained from ATCC and passaged for fewer than 6 months after receipt. VCS2 (6) and C4-2 (7) cells were derived from VCaP and LNCaP cells, respectively. Subconfluent cultures of VCaP, LNCaP, VCS2, and C4-2 cells grown in the presence of androgen (5%–10% FBS) were used as a source of control RNA. Cell lines' identities were routinely validated by examining cell morphology, verifying *AR* mRNA expression, and sequencing for expected *AR* mutations (in LNCaP and LNCaP-derived C4-2 cells) and/or TMPRSS2:ERG translocation (in VCaP and VCaP-derived VCS2 cells). DNase-treated RNA was extracted using the RNeasy Plus Mini Kit (Qiagen).

Results

RNA-seq gene expression analysis is concordant with previous microarray analysis

We had previously analyzed on Affymetrix U133A microarrays a panel of 33 CRPC bone marrow biopsies in comparison with a series of primary prostate cancer (3). However, the additional information that can be gained by paired-end RNA-seq led us to re-analyze a subset of these CRPC samples, which were selected on the basis of very low contaminating hematopoietic or stromal cell content (>90% tumor by H&E) and availability of adequate RNA. For each of the 8 samples selected, 50 ng of total RNA was amplified into double-stranded cDNA and Illumina paired-end adaptors were ligated onto the library for 76 cycles of paired-end sequencing (samples 49 and 66) or 101 cycles of paired-end sequencing (samples 24, 28, 39, 55, 71, and 74; see Supplementary Methods).

Although RNA from the previously analyzed primary prostate cancer was not available, we were still interested in whether gene expression data from the RNA-seq and the previous Affymetrix U133A microarrays were consistent. Therefore, we re-analyzed the Affymetrix raw data to perform a transcript-level normalization and performed a correlation analysis between the intensity values of these arrays with the RPKM from our RNA-seq data (see Supplementary Methods). Considering approximately 13,000 transcripts (Supplementary Table S1), our analysis showed a statistically significant, positive correlation between gene expression values measured from the same CRPC sample on both platforms (Supplementary Fig. S1). Our observation of r values less than 0.7 may be attributed to the 3-prime bias intrinsic in the U133A microarray, whereas our random priming, whole transcriptomic RNA-seq approach resulted in consistent coverage across transcripts (8) and better detection of low-abundance transcripts (9). Spearman r values increased when only the last exon RPKM was used for correlation analysis (data not shown). Nonetheless, this result indicated that gene expression values were not platform-dependent and supported our previous conclusions regarding gene expression differences between the primary prostate cancer and CRPC samples (3).

Mutation analysis reveals potential drivers of tumor development or progression

Across the 8 CRPC samples, we found an average of 131 protein-coding, somatic mutations (either frameshift, nonsense, or missense) with at least 20% variant reads at 20× coverage that were screened against the SNP databases as described in the Supplementary Methods (Table 1 and Supplementary Table S2).

Table 1. Spectrum of genetic alterations detected in CRPC

			>10 coverage	>20 coverage	Protein coding		
Sample	Total	Somatic	>10% allele	>20% allele	Missense	Nonsense	Frameshift
24	132,923	2,133	1,074	440	122	7	25
28	120,300	2,608	1,599	740	122	1	26
39	70,115	2,139	1,248	546	86	0	27
49	101,200	2,903	781	293	87	0	3
55	102,631	2,024	1,123	496	67	0	27
66	142,647	3,584	984	318	103	1	6
71	136,153	2,460	1,468	671	95	2	42
74	108,171	2,647	1,762	799	132	7	66

NOTE: The total number of variants is indicated, with anticipated (somatic) variants filtered as present in the COSMIC database or not represented in the dbSNP, HapMap, or 1000Genomes databases. Among the higher confidence, 10% and 20% sequence read fractions are protein-coding mutations of missense, nonsense, and frameshift variants.

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

Table 2. Fusion and splice site location for three novel fusion transcripts detected by deFuse and ChimeraScan

Sample	5′ Gene	3′ Gene	Fragments	Type	5′ Splice	3′ Splice	Frame (5′/3′)
28	*SND1*	*BRAF*	27	Intrachromosomal	chr7:127361454	chr7:140487384	Coding/coding
49	*EPB41L5*	*PCDP1*	9	Intrachromosomal	chr2:120844816	chr2:120317265	Coding/coding
66	*PHF20L1*	*LRRC6*	12	Intrachromosomal	chr8:133790157	chr8:133584728	Coding/coding

NOTE: For each fusion shown, information is provided indicating the CRPC identifier, number of fusion/splice spanning fragments sequenced, as well as the chromosomal coordinates for the novel splice junction.

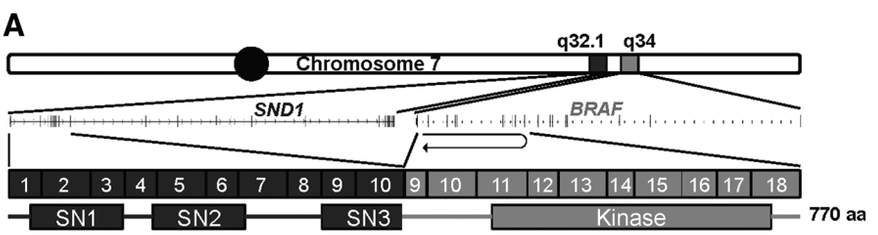
Among the mutations that were likely drivers of tumor progression, we found mutations in *AR* that we had previously reported in these tumors (4). These were an H875Y mutation in CRPC 39 and T878A mutation in CRPC 55 and 71 (Hg19 annotation; equivalent to H874Y and T877A, respectively, in the former Hg18 annotation).

We observed additional novel mutations in genes that have been previously reported as being mutated in prostate cancer (10–12). These included an R398W mutation in *NCoR1* (nuclear receptor corepressor 1) in CRPC 66, which may decrease its corepression of AR (13), a premature stop codon at position 546 in *KDM3A* (lysine-specific demethylase 3A) in CRPC 74, a frameshift mutation in *KDM4A* (lysine-specific demethylase 4A) in CRPC 28, frameshift mutations in the lysine methyltransferase genes *SETD5* and *SETD7* (in CRPC 71 and 74, respectively), as well as a missense mutation in *SETD5* in CRPC 49. We also found a premature stop codon in a RasGEF, *RASGRP3*, at codon 204 in CRPC 28, and an L319V mutation in a RasGAP, *RASA1* in CRPC 39. The *RASGRP3* truncation would preserve the Ras-binding REM domain and its exchange function CDC25 domain while deleting key regulatory regions in the C-terminus, which may lead to enhanced Ras activity, whereas the *RASA1* mutation in the PH domain could affect its membrane localization and thus ability to inactivate Ras (14, 15). We also detected potential loss-of-function mutations in the tumor suppressor proteins encoded by *CHD1*, *TP53BP1*, and *INPP4B*, which have been reported previously as mutated in prostate cancer (10–12). Finally, we observed an R800P mutation in *CDH1* (E-cadherin) in CRPC 74, which may interfere with the ability of the cytoplasmic domain to bind and regulate signaling through β-catenin (16).

Paired-end sequencing of metastatic CRPC reveals expression of novel fusion genes

We performed post-processing for the discovery of fusion genes using both an annotation-dependent algorithm (ChimeraScan) and an annotation-independent algorithm (deFuse; see Supplementary Methods). We found only 3 high-confidence fusions detected by both algorithms, each of which was novel (Table 2). The first of these predicted fusions, *SND1:BRAF* (Fig. 1), is a potential driver of tumorigenesis in CRPC 28, having fused the kinase domain of B-Raf (contained within exons 9–18) to the 3 staphylococcal nuclease homolog domains of Snd1. Lacking the regulatory Ras-binding domain (exons 3–7) and inhibitory serine phosphorylation site (exon 8) in wild-type B-Raf, this fusion kinase has been detected once previously in the gastric cancer cell line GTL16 and was noted to promote cancer cell growth via uncontrolled and increased activation of downstream MAP kinases (17). *BRAF* rearrangements to other genes have been observed previously in prostate cancer (18), and this particular fusion puts the B-Raf kinase domain under control of the *SND1* promoter, which is active in a majority of prostate cancers (19).

We also detected with high confidence 2 additional putative fusions genes, *EPB41L5:PCDP1* (Supplementary Fig. S2)



B

> SND1:BRAF

MASSAQSGGS SGGPAVPTVQ RGIIKMVLSG CAIIVRGQPR GGPPPERQIN LSNIRAGNLA
RRAAATQPDA KDTPDEPWAF PAREFLRKKL IGKEVCFTIE NKTPQGREYG MIYLGKDTNG
ENIAESLVAE GLATRREGMR ANNPEQNRLS ECEEQAKAAK KGMWSEGNGS HTIRDLKYTI
ENPRHFVDSH HQKPVNAIIE HVRDGSVVRA LLLPDYYLVT VMLSGIKCPT FRREADGSET
PEPFAAEAKF FTESRLLQRD VQIILESCHN QNILGTILHP NGNITELLLK EGFARCVDWS
IAVYTRGAEK LRAAERFAKE RRLRIWRDYV APTANLDQKD KQFVAKVMQV LNADAIVVKL
NSGDYKTIHL SSIRPPRLEG ENTQDLIRDQ GFRGDGGSTT GLSATPPASL PGSLTNVKAL
QKSPGPQRER KSSSSSEDRN RMKTLGRRDS SDDWEIPDGQ ITVGQRIGSG SFGTVYKGKW
HGDVAVKMLN VTAPTPQQLQ AFKNEVGVLR KTRHVNILLF MGYSTKPQLA IVTQWCEGSS
LYHHLHIIET KFEMIKLIDI ARQTAQGMDY LHAKSIIHRD LKSNNIFLHE DLTVKIGDFG
LATVKSRWSG SHQFEQLSGS ILWMAPEVIR MQDKNPYSFQ SDVYAFGIVL YELMTGQLPY
SNINNRDQII FMVGRGYLSP DLSKVRSNCP KAMKRLMAEC LKKKRDERPL FPQILASIEL
LARSLPKIHR SASEPSLNRA GFQTEDFSLY ACASPKTPIQ AGGYGAFPVH

Figure 1.
SND1:BRAF fusion transcript detected in CRPC 28. A, schematic representation of the fusion between *SND1* and *BRAF* on chromosome 7. *SND1* exons, SND1 SN domains, *BRAF* exons, and the BRAF kinase domain are indicated. B, predicted amino acid sequence for the SND1:BRAF fusion protein. Amino acids originating from SND1 are represented in dark gray, whereas amino acids contributed by BRAF are light gray.

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

and *PHF20L1*:*LRRC6* (Supplementary Fig. S3). However, it is unknown whether the fusion of their respective functional domains would confer oncogenic activity, and these genes have not been previously documented as upregulated or fused in cancer (20–23). Fusion between *TMPRSS2* and *ERG* or *ETV1*, which occur in approximately half of all prostate cancer, were notably absent from the list of predicted fusions (24). Consistent with this result, clustering of these 8 CRPC and other CRPC sets in the Affymetrix microarray dataset (GEO Accession ID GSE32269) revealed that the 8 CRPC samples we sequenced are fusion-negative (Supplementary Fig. S4). Interestingly, ChimeraScan (but not deFuse) detected with high probability a fusion between *TMPRSS2* and *ETV4* in CRPC 74 (Supplementary Table S3), which occurs with far less frequency than the *TMPRSS2*:*ERG* or *TMPRSS2*:*ETV1* fusions (24).

Noncoding RNAs expressed in CRPC

RNA-seq permitted us to examine the expression of genes for which probes were not present on the microarrays performed previously. A complete list of genes and their computed RPKM values is provided in Supplementary Table S4. Interestingly, among the top-expressing 100 transcripts by mean RPKM across all 8 CRPC samples (Supplementary Table S5) were 10 previously annotated noncoding RNAs (ncRNAs), all of which were also present in the list of the top 100 genes determined by median RPKM (Supplementary Table S6). The most highly expressed transcript, the ncRNA *MALAT1* (CR595720), is a long noncoding RNA (lncRNA) that has been implicated in regulating mRNA splicing (25) and its expression was recently found to be associated with prostate cancer progression, including CRPC (26). Also, on this list is the lncRNA *PABPC1*, which interacts with poly-A-mRNA–binding proteins and is important for RNA decay in response to poly-A shortening. Its upregulation in prostate cancer has been suggested to be in response to an increased number of improperly spliced or improperly processed transcripts (27).

We observed that our list of highly expressed ncRNAs did not contain any of the noncoding prostate cancer–associated transcripts (PCAT) recently reported such as *SChLAP1* (28) and *PCAT-1* (29), although they were expressed in a subset of samples at lower levels (see Supplementary Table S4). To identify any additional highly expressed lncRNA, we next performed novel lncRNA discovery using CuffLinks, accepting any novel unannotated transcript greater than 200 nucleotides with at least 2 exons. A complete list of novel lncRNAs and their mean RPKM values is provided in Supplementary Table S7.

Pathways upregulated in CRPC

To determine whether the coding or noncoding RNAs abundantly expressed in CRPC may play a significant physiologic role in promoting cancer progression, we performed differential expression analysis of these samples against RNA-seq performed on 240 primary prostate cancers sequenced as part of The Cancer Genome Atlas (TCGA). In a combined dataset of both the TCGA and CRPC samples, unsupervised hierarchical clustering of 1,465 transcripts with the widest range of expression across all samples separated the TCGA and CRPC samples into 2 distinct groups (Supplementary Fig. S5A). The average RPKM difference between CRPC and TCGA samples for these 1,465 transcripts are listed in Supplementary Table S8.

To determine whether these other differentially regulated transcripts indicated any disease-driving pathways, we used Gene Set Enrichment Analysis to identify pathways enriched in CRPC versus primary cancer (TCGA). Pathways identified as enriched in CRPC included cell adhesion molecules and MAP kinase signaling (Supplementary Fig. S5B), although the small number of input genes precluded reaching a statistically significant *P* value for these pathways.

Transcripts in metastatic CRPC contain high frequency of intronic reads

We anticipated that this RNA-seq analysis would also add to the previous Affymetrix U133A analysis by revealing alternatively spliced isoforms for many genes. However, while we expected the RNA-seq analysis of RNA that was not poly-A selected to yield many intronic reads, we found an unexpectedly high level of intronic coverage (Supplementary Table S9) that made discovery of novel splice variants difficult. Examination of the mapping statistics showed that the high percentage of intronic reads was not correlated with the percentage of intergenic reads (which were much lower when corrected for total intergenic DNA), indicating that the intronic reads were not gDNA contamination (Supplementary Table S9). Among the top 10 genes as determined by intronic read depth in 2 samples examined in detail (CRPC 49 and CRPC 66; Table 3), we found known markers of prostate cancer including *AR*, *KLK3*, *KLK2*, and *STEAP2*, all of which are regulated by AR (30, 31). However, these genes also had high levels of exonic reads, indicating they were highly expressed. Moreover, global assessment of intronic sequence coverage in CRPC 49 and CRPC 66 (Supplementary Tables S10 and S11, respectively) showed high levels of intronic sequence for a broad spectrum of genes, and this was correlated with their exonic read depth (see below, Supplementary Fig. S7).

Inspection of the Bowtie-mapped reads in the Integrative Genome Viewer (IGV) for all 8 CRPC samples similarly revealed substantial intronic coverage for *KLK3*, *KLK2*, and *AR* (Fig. 2A–C) and for *STEAP2* (Supplementary Fig. S6A). As anticipated from the mapping statistics, we observed much lower levels of intergenic reads between and outside of *KLK2* and *KLK3* (Supplementary Fig. S6B), further indicating only a low level of gDNA contamination. We also observed high intronic read depth in many other genes that are not AR-regulated, such as *CDK19* and *CPSF6* (Supplementary Fig. S6C and S6D), further showing that this phenomenon was not limited to AR-regulated genes. To

Table 3. Ten top-ranking genes with retained introns in CRPC 49 and 66

mCRPC 49	mCRPC 66
OR51E2	*KLK2*
KLK2	*KLK3*
TMEFF2	*AR*
AR	*HFM1*
STEAP2	*AMACR*
KLK3	*TPT1*
TPT1	*SNORA31*
SNORA31	*SHROOM1*
SAT1	*HNRNPC*
SAT	*HNRPC*

NOTE: Genes are ranked in descending order on the basis of their intronic RPB measurement.

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

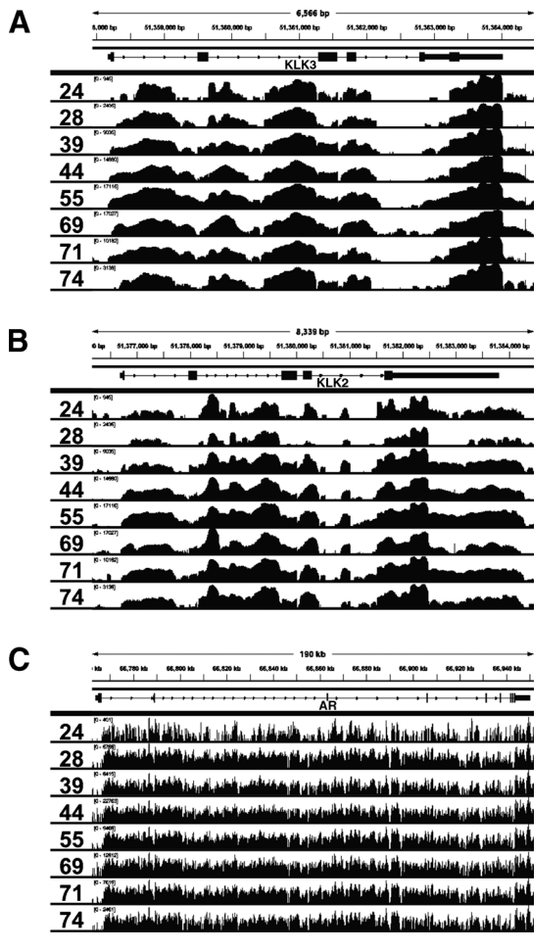


Figure 2.
Extensive intronic coverage in a subset of genes. Quality-filtered read coverage for (A) *KLK3*, (B) *KLK2*, and (C) *AR* for all 8 CRPC mRNA samples sequenced.

globally assess whether intronic read depth was related to overall gene expression, we plotted the $\log_{10}$-transformed values for the exonic RPKM versus the $\log_{10}$-transformed values for the intronic RPKM for all genes across all 8 CRPC samples (Supplementary Fig. S7). The observed strong positive correlation indicated that the level of intron reads for most genes was proportional to the overall expression of the gene.

Metastatic CRPC cells undergo inefficient splicing

We next addressed whether the high frequency of intronic reads reflected unspliced introns versus introns that were spliced but not degraded. Therefore, for each splice site in every gene, we computationally counted the total number of fragments spanning the site that was spliced (exon-to-exon reads) versus fragments that were not spliced (exon-to-intron reads). We then calculated the percentage of reads corresponding to an unspliced junction out of the total number of reads for that junction (spliced plus unspliced; Supplementary Table S12). On the basis of these calculations across all samples, we determined that approximately 28% of the splice junctions were not spliced. It should be noted that the absolute number of reads that mapped completely within an exon or within an intron were approximately equal (see Supplementary Table S9). However, this is not inconsistent with the above estimate of 28% unspliced mRNA as the greater length of introns relative to exons increases the likelihood that an RNA-seq read from an unspliced transcript will map to an intron versus an exon.

We next wanted to determine the extent to which the unspliced introns reflected nascent mRNA that was not yet polyadenylated. To address this question, we isolated the polyadenylated fraction of mRNA from the total RNA pool in four samples and performed whole transcriptome amplification using the same method used for whole cellular RNA. We then used a series of PCR primer pairs in a qRT-PCR scheme (Supplementary Fig. S8) to amplify either spliced or unspliced junctions in a group of highly expressed genes that had high frequencies of intronic reads (*AR*, *KLK2*, *KLK3*, *STEAP2*, *CPSF6*, and *CDK19*; see Supplementary Table S13 for primer sequences). Similar to our computational approach above, we calculated a relative splicing index for each junction on the basis of amplification with exon–intron primers versus amplification with exon–exon plus exon–intron primers. This relative splicing index, which reflects the ratio of unspliced to total junctions (unspliced plus spliced), was then averaged for each gene and was further normalized across the samples on the basis of amplification with primers within exons. Finally, we compared the results for the poly-A versus unfractionated total cellular RNA. For *KLK3*, *STEAP2*, and *CPSF6* (Fig. 3A–C), there were no significant differences between the poly-A and total cellular RNA fractions, indicating that a substantial fraction of the poly-A mRNA for these genes is unspliced. In contrast, the splicing index values for *CDK19*, *KLK2*, and *AR* were lower in the poly-A fraction, indicating that a proportion of the unspliced junctions for these genes were contained in nonpolyadenylated nuclear RNA (Fig. 3D–F).

Splicing efficiency in CRPC is decreased relative to primary prostate cancer

It did not appear that the apparently substantial unspliced mRNA was due to biases in the whole transcriptome amplification methods we used, as we observed high levels of intronic reads and of exon–intron junctions. Moreover, examination of transcripts in the bone marrow biopsy samples that were derived from hematopoietic or stromal cells, such as *HBB* (hemoglobin beta; Supplementary Fig. S9A) and *SPP1* (osteopontin; Supplementary Fig. S9B), respectively, showed very few intronic reads or exon–intron junctions, indicating that inefficient splicing was a property of the tumor cells. Nonetheless, we next addressed possible biases by comparing cDNA generated from amplified versus unamplified RNA. For this analysis, we used RNA from CRPC 66, for which we had an adequate amount of extracted RNA. Portions of the RNA were used to generate single-stranded or double-stranded amplified libraries or to generate cDNA directly without amplification using conventional reverse transcriptase with a pool of oligo-dT and random oligonucleotide primers. We then assessed the *AR* splicing index by amplification with primers corresponding to exons 4–5, exons 5–6, exon 4 to intron 4, and exon 5 to intron 5. Significantly, we observed a higher *AR* splicing index, indicative of more unspliced mRNA, in conventionally synthesized cDNA compared with the whole transcriptome–amplified libraries (Supplementary Fig. S9C), further supporting the conclusion that a substantial proportion of transcripts in the CRPC samples was not spliced.

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

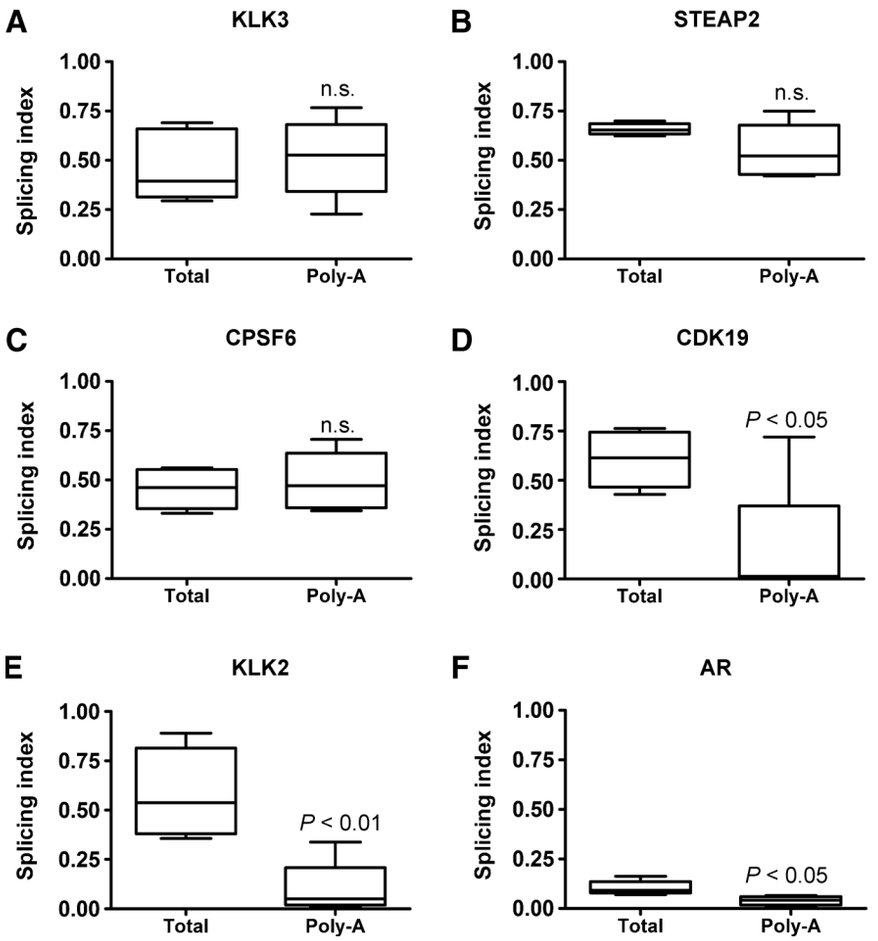


Figure 3.
Poly-adenylated RNA contains unspliced introns. The splicing index was calculated for (A) *KLK3*, (B) *STEAP2*, (C) *CPSF6*, (D) *CDK19*, (E) *KLK2*, and (F) *AR* in CPRC samples before and after OligoTex purification for poly-adenylated (Poly-A) species. Measurement was performed in triplicate, and the average values for each CRPC are depicted on box plots.

Finally, we addressed whether the inefficient splicing we observed was a general feature of prostate cancer. For this analysis, we isolated whole cellular RNA from 6 cases of laser capture microdissected untreated primary prostate cancer (Gleason score 7), 10 cases of normal prostate epithelium, and 4 prostate cancer cell lines (LNCaP, C4-2, VCaP, and VCS2). The RNA was then subjected to whole transcriptome amplification as for the metastatic CRPC samples, and we determined the splicing index for the 6-gene panel. Significantly, the median splicing index was higher for all 6 genes in the CRPC samples when compared with primary prostate cancer, normal epithelium, or cell lines, indicating that splicing is less efficient in metastatic CRPC versus normal prostate or primary prostate cancer (Fig. 4).

Discussion

This study used RNA-seq to further characterize gene expression in a series of metastatic CRPC samples and in particular to assess for mutations, gene fusions, IncRNA, and alternative splicing. We detected novel mutations in a series of genes that have been implicated previously in prostate cancer development or progression to metastatic CRPC. These included mutations in genes encoding proteins that regulate transcription (*NCOR1*, *KDM3A*, *KDM4A*, *CHD1*, *SEDT5*, and *SETD7*), PI3K pathway (*INPP4B*), and Ras pathway signaling (*RASGRP3* and *RASA1*). Although the functional significance of these mutations has not been determined, NCoR1 can function as a corepressor for AR and its loss could enhance AR activity in CRPC. Alterations in KDM3A, KDM4A, and CHD1 could also affect AR activity but would likely have broad effects on gene expression. Our observation of novel mutations to *SETD5* and *SETD7* may result in altered chromatin accessibility during co-transcriptional RNA processing and thus may also contribute to intron retention, a phenomenon recently reported in an RNA-seq analysis of clear cell renal cell carcinoma (32). Mutations we found in *RASA1* and *RASGRP3*, and a novel *SND1*:*BRAF* gene fusion, may contribute to the enhanced RAS/RAF/MAPK signaling observed with progression to CRPC (33). Interestingly, although gene fusions are common in prostate cancer, they were infrequent in these samples when we used a high stringency threshold. While we may have failed to detect some abundant fusion gene transcripts, it is also likely that many gene fusions are not drivers of tumor progression and that their expression may thereby not confer a selective advantage in these advanced tumors.

Amongst the most highly expressed genes were 10 noncoding RNAs, including *MALAT1* and *PABPC1*, and in a subset of our cases, we also observed expression of one or more of the recently reported noncoding PCATs (28, 29). In particular, the

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

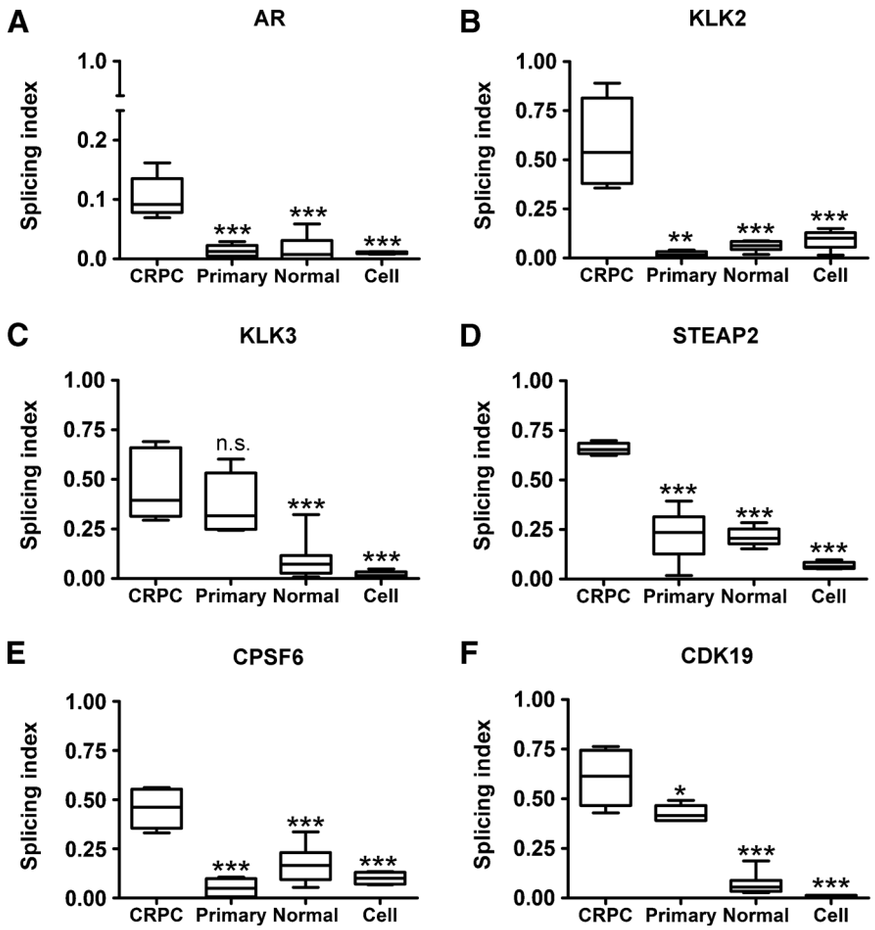


Figure 4.
CRPC samples express more unspliced mRNA than primary prostate cancers, normal prostatic epithelium, or cultured prostate cancer cell lines. Splicing indices were calculated and compared between CRPC, normal prostatic epithelial tissue, laser capture microdissected primary prostate cancer cells, and established prostate cancer cell lines. Boxplots representing the data within each set are shown for (A) *AR*, (B) *KLK2*, (C) *KLK3*, (D) *STEAP2*, (E) *CPSF6*, and (F) *CDK19*. Boxplots represent the set of average values from 3 replicate experiments for each biologic sample. Statistical significance between samples was measured by the Student unpaired *t* test (95% confidence interval), and probability of statistical difference is indicated by *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$; ns, not statistically significant.

outlier PCAT predictive of lethal disease (*PCAT-114*, also referred to as *SChLAP1*) was expressed in a subset of cases. Interestingly, many of the lncRNAs that were expressed at high levels, in addition to PCAT-114, are known to be involved in regulating transcription and may contribute to tumor progression (25, 27, 28).

An unexpected result was the large number of sequence reads that mapped to introns. This appeared to reflect incomplete splicing based on the fraction of reads that spanned exon–intron junctions compared with those that spanned exon–exon junctions. For some genes, this may reflect the use of whole-cell RNA rather than poly-A RNA, but for others, we found that the ratio of exon–intron versus exon–exon junctions was not significantly decreased when we examined poly-A RNA. In either case, this inefficient splicing was greater in the metastatic CRPC samples than in normal prostate and primary prostate cancer, indicating that it is a feature of metastatic CRPC. It is not clear why this inefficient splicing was not observed in the prostate cancer cell lines as these were derived from metastatic CRPC but possibilities include a role for the tumor microenvironment or a selective advantage *in vitro* for subclones that splice more efficiently.

Significantly, genes with the greatest levels of intron retention did not group into any specific biologic pathways but rather were those with the greatest overall expression (see Supplementary Tables S10 and S11). Therefore, we suggest that these findings reflect global increases in gene transcription in advanced CRPC and a saturation of the cellular splicing machinery, with subsequent uncoupling of transcription and splicing (34). This hypothesis is consistent with the high level and increased expression of multiple ncRNA involved in transcription and RNA processing with prostate cancer progression (25, 27, 28). It is also supported by a recent report showing that increased transcription of already upregulated genes, which correspond to changes in the methylation status of the genome, occurs during progression to CRPC (35). Finally, it is of interest that H3K27me3 levels are decreased with prostate cancer progression, which may contribute to global derepression of gene transcription (36, 37).

Alternative splicing can clearly contribute to tumor progression (34, 38, 39), and the inefficient removal of introns may provide increased substrate for alternative splicing to generate isoforms of some proteins that contribute to tumor progression. Moreover, high levels of intronic RNA also would presumably sequester many microRNA species, resulting in dysregulation of multiple miRNA-regulated protein expression networks. However, further studies are needed to determine whether inefficient splicing provides a selective advantage driving tumor progression *in vivo* and whether these tumors may be vulnerable to agents that suppress rate-limiting steps in splicing.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A.G. Sowalsky, H. Zhao, G.J. Bubley, S.P. Balk, W. Li
Development of methodology: A.G. Sowalsky, H. Zhao, S.P. Balk, W. Li
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.G. Sowalsky, H. Zhao, G.J. Bubley
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.G. Sowalsky, Z. Xia, L. Wang, H. Zhao, S. Chen, G.J. Bubley, S.P. Balk, W. Li
Writing, review, and/or revision of the manuscript: A.G. Sowalsky, Z. Xia, G.J. Bubley, S.P. Balk, W. Li
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.G. Sowalsky, S. Chen, W. Li
Study supervision: A.G. Sowalsky, S.P. Balk, W. Li

Grant Support

The study was supported by grants from the NIH (T32CA081156 to A.G. Sowalsky, R00CA135592 to S. Chen, P01CA163227-01A1 to S.P. Balk, DF/HCC-Prostate Cancer SPORE P50CA090381 to S.P. Balk, R01HG007538 to W. Li), Department of Defense Prostate Cancer Research Program (Postdoctoral Training Award W81XWH-13-1-0267 to A.G. Sowalsky, Idea Development Awards W81XWH-11-1-0295, W81XWH-08-1-0414, and W81XWH07-1-0443 to S.P. Balk, and W81XWH-10-1-0501 to W. Li), CPRIT (RP110471 to W. Li), and a Prostate Cancer Foundation Challenge Award (S.P. Balk).

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.

Received May 14, 2014; revised July 30, 2014; accepted August 24, 2014; published OnlineFirst September 4, 2014.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin 2013;63:11–30.
2. Shen MM, Abate-Shen C. Molecular genetics of prostate cancer: new prospects for old challenges. Genes Dev 2010;24:1967–2000.
3. Stanbrough M, Bubley GJ, Ross K, Golub TR, Rubin MA, Penning TM, et al. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer Res 2006;66:2815–25.
4. Taplin ME, Bubley GJ, Ko YJ, Small EJ, Upton M, Rajeshkumar B, et al. Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. Cancer Res 1999;59:2511–5.
5. Taplin ME, Bubley GJ, Shuster TD, Frantz ME, Spooner AE, Ogata GK, et al. Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. N Engl J Med 1995;332:1393–8.
6. Cai C, He HH, Chen S, Coleman I, Wang H, Fang Z, et al. Androgen receptor gene expression in prostate cancer is directly suppressed by the androgen receptor through recruitment of lysine-specific demethylase 1. Cancer Cell 2011;20:457–71.
7. Wu HC, Hsieh JT, Gleave ME, Brown NM, Pathak S, Chung LW. Derivation of androgen-independent human LNCaP prostatic cancer cell sublines: role of bone stromal cells. Int J Cancer 1994;57:406–12.
8. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods 2013;10:623–9.
9. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 2014;9:e78644.
10. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 2012;44:685–9.
11. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. Cell 2013;153:666–77.
12. Lindberg J, Mills IG, Klevebring D, Liu W, Neiman M, Xu J, et al. The mitochondrial and autosomal mutation landscapes of prostate cancer. Eur Urol 2013;63:702–8.
13. Linja MJ, Porkka KP, Kang Z, Savinainen KJ, Janne OA, Tammela TL, et al. Expression of androgen receptor coregulators in prostate cancer. Clin Cancer Res 2004;10:1032–40.
14. Aiba Y, Oh-hora M, Kiyonaka S, Kimura Y, Hijikata A, Mori Y, et al. Activation of RasGRP3 by phosphorylation of Thr-133 is required for B cell receptor-mediated Ras activation. Proc Natl Acad Sci U S A 2004;101:16612–7.
15. Pamonsinlapatham P, Hadj-Slimane R, Lepelletier Y, Allain B, Toccafondi M, Garbay C, et al. p120-Ras GTPase activating protein (RasGAP): a multi-interacting protein in downstream signaling. Biochimie 2009;91:320–8.
16. Nelson WJ, Nusse R. Convergence of Wnt, beta-catenin, and cadherin pathways. Science 2004;303:1483–7.
17. Lee NV, Lira ME, Pavlicek A, Ye J, Buckman D, Bagrodia S, et al. A novel SND1-BRAF fusion confers resistance to c-Met inhibitor PF-04217903 in GTL16 cells through [corrected] MAPK activation. PLoS One 2012;7:e39653.
18. Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med 2010;16:793–8.
19. Kuruma H, Kamata Y, Takahashi H, Igarashi K, Kimura T, Miki K, et al. Staphylococcal nuclease domain-containing protein 1 as a potential tissue marker for prostate cancer. Am J Pathol 2009;174:2044–50.
20. Gosens I, Sessa A, den Hollander AI, Letteboer SJ, Belloni V, Arends ML, et al. FERM protein EPB41L5 is a novel member of the mammalian CRB-MPP5 polarity complex. Exp Cell Res 2007;313:3959–70.
21. DiPetrillo CG, Smith EF. Pcdp1 is a central apparatus protein that binds Ca(2+)-calmodulin and regulates ciliary motility. J Cell Biol 2010;189:601–12.
22. Shimojo H, Sano N, Moriwaki Y, Okuda M, Horikoshi M, Nishimura Y. Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. J Mol Biol 2008;378:987–1001.
23. Kott E, Duquesnoy P, Copin B, Legendre M, Dastot-Le Moal F, Montantin G, et al. Loss-of-function mutations in LRRC6, a gene essential for proper axonemal assembly of inner and outer dynein arms, cause primary ciliary dyskinesia. Am J Hum Genet 2012;91:958–64.
24. Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, Helgeson BE, et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res 2006;66:3396–400.
25. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 2010;39:925–38.
26. Ren S, Liu Y, Xu W, Sun Y, Lu J, Wang F, et al. Long noncoding RNA MALAT-1 is a new potential therapeutic target for castration resistant prostate cancer. J Urol 2013;190:2278–87.
27. Yang C, Ströbel P, Marx A, Hofmann I. Plakophilin-associated RNA-binding proteins in prostate cancer and their implications in tumor progression and metastasis. Virchows Arch 2013;463:379–90.
28. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet 2013;45:1392–8.
29. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol 2011;29:742–9.
30. Porkka KP, Helenius MA, Visakorpi T. Cloning and characterization of a novel six-transmembrane protein STEAP2, expressed in normal and malignant prostate. Lab Invest 2002;82:1573–82.
31. Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. Cell 2009;138:245–56.
32. Simon JM, Hacker KE, Singh D, Brannon AR, Parker JS, Weiser M, et al. Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res 2014;24:241–50.

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

33. Mulholland DJ, Kobayashi N, Ruscetti M, Zhi A, Tran LM, Huang J, et al. Pten loss and RAS/MAPK activation cooperate to promote EMT and metastasis initiated from prostate cancer stem/progenitor cells. Cancer Res 2012;72:1878–89.
34. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev 2010;24: 2343–64.
35. Friedlander TW, Roy R, Tomlins SA, Ngo VT, Kobayashi Y, Azameera A, et al. Common structural and epigenetic changes in the genome of castration-resistant prostate cancer. Cancer Res 2012;72:616–25.
36. Pellakuru LG, Iwata T, Gurel B, Schultz D, Hicks J, Bethel C, et al. Global levels of H3K27me3 track with differentiation in vivo and are deregulated by MYC in prostate cancer. Am J Pathol 2012;181:560–9.
37. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT, et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science 2012;338:1465–9.
38. Rajan P, Elliott DJ, Robson CN, Leung HY. Alternative splicing and biological heterogeneity in prostate cancer. Nat Rev Urol 2009;6:454–60.
39. Sette C. Alternative splicing programs in prostate cancer. Int J Cell Biol 2013;2013:458727.

Published OnlineFirst September 4, 2014; DOI: 10.1158/1541-7786.MCR-14-0273

Molecular Cancer Research

Whole Transcriptome Sequencing Reveals Extensive Unspliced mRNA in Metastatic Castration-Resistant Prostate Cancer

Adam G. Sowalsky, Zheng Xia, Liguo Wang, et al.

Mol Cancer Res 2015;13:98-106. Published OnlineFirst September 4, 2014.