This is the post print version of the article, which has been published in British Journal of Cancer. 2018, 119, 347-356. http://dx.doi.org/10.1038/s41416-018-0172-0

TITLE: Constitutively active androgen receptor splice variants *AR-V3, AR-V7* and *AR-V9* are co-expressed in castration-resistant prostate cancer metastases

Heini M.L. Kallio[1], Reija Hieta[1], Leena Latonen[1], Anniina Brofeldt[1], Matti Annala[1], Kati Kivinummi[1], Teuvo L. Tammela[2], Matti Nykter[1], William B. Isaacs[3], Hans G. Lilja[4], G. Steven Bova[1], Tapio Visakorpi[1,5]

[1]Prostate Cancer Research Center, Faculty of Medicine and Life Sciences and BioMediTech Institute, University of Tampere, Tampere, Finland

[2]Department of Urology, University of Tampere, Tampere University Hospital, Tampere, Finland

[3]The James Buchanan Brady Urological Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[4]Departments of Surgery (Urology), Laboratory Medicine and Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK; Department of Translational Medicine, Lund University, Malmö, Sweden; Prostate Cancer Research Center, Faculty of Medicine and Life Sciences and BioMediTech Institute, University of Tampere, Tampere, Finland

[5]Fimlab Laboratories, Tampere University Hospital, Tampere, Finland

Running title: Androgen receptor splice variants are co-expressed

Corresponding author: Heini M.L. Kallio

Address: Arvo Ylpön katu 34, 33520 Tampere, Finland

Telephone: +358-50-318 5809

Email: heini.kallio@uta.fi

Keywords: androgen receptor, androgen receptor splice variant, prostate cancer, castration-resistant prostate cancer, next-generation sequencing, androgen receptor aberration

Abbreviations: ADT, androgen deprivation therapy; AR, androgen receptor; AR-FL, full-length androgen receptor; *AR*-GSR, genomic structural rearrangement of *AR*; AR-V, androgen receptor splice variant; BPH, benign prostatic hyperplasia; CNV, copy number variation; CRPC, castration-resistant prostate cancer; CTC, circulating tumor cell; PC, prostate cancer; PSA, prostate-specific antigen

Abstract

Background: Significant subset of prostate cancer (PC) patients with castration-resistant form of the disease (CRPC) show primary resistance to androgen receptor (AR)-targeting drugs developed against CRPC. As one explanation could be the expression of constitutively active androgen receptor splice variants (AR-Vs), our current objectives were to study *AR-V*s and other *AR* aberrations to better understand the emergence of CRPC. **Methods:** We analyzed specimens from different stages of prostate cancer by next-generation sequencing and immunohistochemistry. **Results:** *AR* mutations and copy number variations were detected only in CRPC specimens. Genomic structural rearrangements of *AR* were observed in 5/30 metastatic CRPC patients but they were not associated with expression of previously known *AR-V*s. The predominant *AR-V*s detected were *AR-V3*, *AR-V7* and *AR-V9*, with the expression levels being significantly higher in CRPC cases compared to prostatectomy samples. Out of 25 CRPC metastases that expressed any *AR* variant, 17 cases harbored expression of all three of these *AR-V*s. AR-V7 protein expression was highly heterogeneous and higher in CRPC compared to hormone-naïve tumors. **Conclusions:** *AR-V3*, *AR-V7* and *AR-V9* are co-expressed in CRPC metastases highlighting the fact that inhibiting AR function via regions common to all AR-Vs is likely to provide additional benefit to patients with CRPC.

Introduction

Prostate cancer (PC) is the most common malignancy and third most common cause of cancer-related death among men in Europe. Androgens are required for the normal development of prostate tissue and exert their effects through androgen receptor (AR) mediated signaling but also have important role during prostate cancer emergence and progression. Most prostate cancers grow slowly and are curable by surgery and radiation when confined to the prostate. In contrast, treatment of prostate cancers that have spread outside the prostate usually includes manipulation of the AR signaling axis; androgen deprivation therapy (ADT) either by surgical or chemical castration. However, during currently available ADT, lethal castration-resistant form of prostate cancer (CRPC) will eventually emerge after a variable period of time. Even though the exact mechanism by which CRPC develops remains to be fully understood, several mechanisms of castration resistance have been identified such as *AR* gene amplification[1,2], point mutations in *AR* gene[3,4] and induction of steroidogenesis in CRPC cells[5-7]. *AR* gene amplification has been demonstrated in approximately 30% of CRPCs[1]. Cancers with *AR* amplification have been shown to respond better to second-line maximal androgen blockade compared to tumors without the amplification, however, the response was short-lived[8]. Also *AR* mutations are rare even at CRPC stage being present in approximately 10–30% of cases[4]. These mutations are almost always associated with diverse gains-of-function and about 45% of the mutations occur in the ligand-binding domain[9]. *AR* mutations can broaden ligand specificity to alternative steroid hormones, hypersensitize the receptor to castrate levels of androgens or lead to resistance to current forms of treatment making AR active even in the presence of anti-androgens[10].

Importantly, androgen signaling remains active even in the CRPC stage[11,12]. The established concept of sustained AR signaling during CRPC has led to the clinical development of second-generation AR-

targeting drugs enzalutamide and abiraterone that target the ligand-binding domain of AR directly and indirectly, respectively. Enzalutamide is an AR antagonist whereas abiraterone is a CYP17 inhibitor approved by the US Food and Drug Administration (FDA) for the treatment of metastatic CRPC. Several studies have shown that presence of *AR* amplification or *AR* mutations in plasma samples is associated with worse outcome with enzalutamide and abiraterone[13-17]. Furthermore, a significant subset of patients show primary resistance to these agents with respect to PSA (prostate-specific antigen) level[18] and among patients who initially respond, nearly all eventually develop acquired resistance.

One potential explanation for the resistance to first- and second-generation AR-targeted therapies is the presence of AR splice variants (AR-Vs). AR-Vs are alternatively spliced isoforms of the *AR* mRNA usually resulting in truncated AR protein product. The key domains shared among wild-type full-length AR (AR-FL) and all AR-Vs are the NH_2-terminal transactivating domain (NTD) and DNA-binding domain (DBD). However, AR-Vs lack variable portions of the COOH-terminal domain including the ligand-binding domain (LBD)[19-21]. In spite of the fact that AR-Vs are unable to bind a ligand, they are constitutively active as transcription factors and capable of activating target genes[22].

To date, at least 22 AR-Vs have been discovered in CRPC specimens[23]. AR-V7 is the most clinically relevant variant as it is most frequently observed and the most abundant AR-V in clinical specimens. In addition, AR-V7 is the only variant that can be detected reproducibly at both the mRNA and protein levels. Moreover, detection of *AR-V7* mRNA in circulating tumor cells (CTCs) and peripheral whole blood from CRPC patients treated with enzalutamide or abiraterone has been implicated in primary resistance and shorter progression-free and overall survival[24-27]. Interestingly, the prevalence of *AR-V7* was shown to be higher in enzalutamide-treated men who had previously received abiraterone and in abiraterone-treated men who had previously received enzalutamide[25]. These findings were supported by

an independent study that also utilized CTC-based RT-PCR assay[28]. In this prospective study, it was shown that PSA response rate to abiraterone or enzalutamide was 7% among *AR-V7* positive patients and 63% among *AR-V7* negative patients. Another recent study demonstrated that *AR-V7* detection in plasma-derived exosomal RNA strongly predicts resistance to enzalutamide or abiraterone in CRPC patients[29]. Although these studies implicate that AR-V7 could be used as a treatment-specific biomarker, it is likely that other AR-Vs also play a role in the development of CRPC. For example, it was recently reported that *AR-V9* is often co-expressed with *AR-V7* in CRPC metastases and predicts primary resistance to abiraterone[30].

Recently, genomic structural rearrangements of *AR* (*AR*-GSRs) were established as a new class of *AR* gene alteration occurring in one third of CRPC-stage specimens[31]. This work showed that the presence of *AR*-GSRs at high variant allele frequency was associated with outlier, tumor-specific expression of rearrangement-dependent *AR-V* species that displayed androgen-independent and enzalutamide-resistant transcriptional activity. However, contrary to the prior studies in cell lines[32,33], *AR*-GSRs were not associated with the *AR-V7* expression levels in metastatic CRPC tissue[31]. Another recent study utilizing peripheral blood collected from patients with CRPC detected intra-*AR* structural variation in 15/30 patients of whom 14 expressed *AR-V*s[34]. Of note, most of the *AR-V* positive patients expressed multiple *AR-V*s with *AR-V7* being the most frequently occurring splice variant. However, *AR-V3* was the most abundantly expressed *AR* splice variant. According to this study the presence of any *AR-V* was associated with shorter progression-free survival after second-line endocrine treatment compared to patients that did not harbor *AR-V*s[34]. Furthermore, in another recent investigation *AR*-GSRs in circulating tumor DNA were shown to associate with primary resistance to enzalutamide or abiraterone also in treatment-naïve CRPC patients with metastatic disease[13].

Our aim was to study *AR* splice variants, rearrangements, mutations and copy number variations (CNVs) in different stages of prostate cancer to better understand the emergence of CRPC. We used multiple sample cohorts representing hormone-naïve prostate cancers and lymph node metastases as well as locally recurrent and metastatic CRPCs. We first employed whole genome and whole transcriptome sequencing followed by targeted *AR* sequencing panels allowing deeper sequence coverage. In particular, our aim was to confirm whether *AR-V*s are expressed in higher levels in CRPC samples compared to earlier stage cancers. In addition, we wished to elucidate to which extent *AR-V* expression is due to the aberrant splicing and, on the other hand, *AR* gene rearrangements. We also wanted to study the association between *AR-V* and *AR-FL* expression and to find out whether *AR-V* expression affects the expression of AR-regulated genes.

Materials and methods

Sample sets

Two different sample sets utilized in the study are shown in Table 1 and are described in more detail in Supplementary Table S1. The sample set 1 contained freshly frozen tissue specimens from benign prostatic hyperplasia (BPH) (n=12), hormone-naïve prostate cancer (n=30) and locally recurrent CRPC (n=13) with clinicopathological characteristics of prostate cancer cases and prior treatments of CRPC cases being shown in Supplementary Table S1. BPH samples were obtained by radical prostatectomy, cystoprostatectomy and by transurethral resection of the prostate. Hormone-naïve prostate cancer samples were obtained by radical prostatectomy and locally recurrent CRPCs by transurethral resection of the prostate. Histological evaluation and Gleason grading were performed by a pathologist based on hematoxylin/eosin stained slides. All samples contained a minimum of 70% cancerous or hyperplastic

cells. DNA and RNA were isolated simultaneously using an AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA, USA) according to manufacturer's protocol. For certain samples, additional total RNA was isolated using Trizol (Invitrogen, Carlsbad, CA, USA) extraction according to manufacturer's protocol. Three CRPC samples had RNA extracted using both Trizol and Qiagen AllPrep. Integrity was checked using Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

The sample set 2 consisted of 24 additional hormone-naïve prostate cancers removed by prostatectomy of which six specimens were also included in sample set 1, eight lymph node metastases obtained at lymphadenectomy and 30 metastatic CRPC specimens obtained at autopsy (clinicopathological characteristics of the cases are shown in Supplementary Table S1). Hormone-naïve prostate cancer samples contained a minimum of 60% cancerous or hyperplastic cells and were processed as described in the previous section.

Portions of the metastatic cancer tissue from pelvic lymphatic metastasis obtained at lymphadenectomy were used for this study. None of the eight patients had undergone androgen deprivation therapy, chemotherapy, or radiation therapy prior to this surgery. Precise histological control was achieved for all tissues studied in this group using the following protocol. Serial cryostat sectioning was used to identify portions of the sample containing a lower fraction of tumor cells. These areas were manually microdissected from the tissue block every 300 µm based on H&E stained slide visual analysis. The tumor cell fraction was 70% or greater by histologic visual estimation. DNA purification was performed as described previously[35]. Total RNA was isolated using an AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. The integrity of isolated RNA was confirmed using Fragment Analyzer (Advanced Analytical Technologies, Ankeny, IA, USA).

Metastatic CRPC specimens were obtained from 30 men who participated in the PELICAN (Project to ELIminate lethal CANcer) integrated clinical-molecular autopsy study of metastatic prostate cancer (a detailed sample list is shown in Supplementary Table S1). Androgen axis and corticosteroid clinical treatments are listed in the Supplementary Table S1. All metastases (one metastasis per patient) and noncancerous (normal, NL) control samples studied were obtained at autopsy. Isolated frozen tissue samples were serial cryostat microdissected for histological tumour purity >75%, and high-molecular-weight DNA was isolated using proteinase K digestion and phenol/chloroform extraction. Total RNA was isolated using an AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. The integrity of isolated RNA was confirmed using Fragment Analyzer (Advanced Analytical Technologies, Ankeny, IA, USA).

Low-coverage (4–6x) whole genome DNA sequencing and whole transcriptome sequencing (applied to sample set 1) have been described before[36].

Targeted *AR* DNA assay library construction and sequencing (applied to sample set 2)

A custom DNA sequencing panel was designed to cover all *AR* exons and introns. In addition, *FOXA1* exons and *SPOP* exons 6–7 were included in the panel. Targeted sequence enrichment was performed using the SureSelect^{XT} Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) according to manufacturer's instructions. Briefly, 200 ng of genomic DNA was fragmented using Covaris® (Covaris, MA, USA) to yield a fragment size of 150–200 bp. End repair, addition of the 3'-dA overhang, ligation of indexing-specific adaptors, hybridization to custom RNA baits, hybrid capture selection and index tagging were performed according to the Illumina paired-end sequencing library

protocol. All recommended quality controls were performed between steps. The multiplexed samples were sequenced on the Illumina Miseq platform using 150 bp paired-end reads.

Targeted *AR* RNA assay library construction and sequencing (applied to sample set 2)

AR and five androgen-responsive genes (*KLK3*, *FKBP5*, *TMPRSS2*, *ACPP* and *SLC45A3*) were targeted for capture and sequencing. In addition, three house-keeping genes *TBP*, *STARD7* and *DDX1* were included for normalization purposes. This custom RNA sequencing panel was designed to cover all *AR* exons and nonrepetitive intronic regions to enable investigation of most common *AR* splicing variants (*AR-V3*, *AR-V4*, *AR-V5*, *AR-V6*, *AR-V7*, *AR-V9*, *AR-V12* and *AR-45*); other genes were covered less intensively (1 or 5 amplicons per gene). Targeted sequence enrichment was performed using the SureSelect^{XT} RNA Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) according to manufacturer's instructions. Briefly, poly(A) RNA was purified from 1 µg of total RNA and fragmented chemically. In the following steps, samples were prepared using SureSelect Strand-Specific RNA Library Prep Kit to obtain adaptor-ligated cDNA library amplicons. Finally, hybridization to custom RNA baits, hybrid capture selection and index tagging were performed. All the AMPure XP bead purification steps were conducted as instructed. The multiplexed samples were sequenced on the Illumina Miseq platform using 150 bp paired-end reads. The following modifications were made to the protocol if RNA was highly degraded (RQN < 6 determined by Fragment Analyzer) as recommended by Agilent Technologies: 1) Instead of poly(A) RNA purification from 1 µg total RNA, Ribo-Zero Gold Magnetic Kit (Illumina, San Diego, CA, USA) was used to remove rRNA from 2 µg of total RNA. 2) Instead of fragmenting the purified RNA at 94ºC for 8 min, RNA was denatured at 65ºC for 5 min. 3) All AMPure XP bead purification steps were performed using 1.8:1 bead volume to sample volume ratio. 4) Instead of 13 cycles in the pre-capture PCR, the number of cycles was increased to 14.

Validation of the targeted sequencing panels

Targeted custom SureSelect sequencing panels were validated by evaluating their performance in detecting *AR* aberrations in comparison to our previously published whole genome DNA-seq data[37,38] and whole transcriptome RNA-seq data from this study. There was a good concordance in mutation detection between SureSelect DNA panel and previously analyzed data from 22Rv1 cell line sample and metastatic CRPC samples from patients A21, A22 and A24. SureSelect DNA-seq detected the previously found H875Y mutation from 22Rv1 cell line, L702H mutation from liver metastasis from patient A21 as well as T878A mutation from a pelvic lymph node metastasis from patient A22 and from a right rib metastasis from patient A24[37,38]. The data from AR splicing variant analysis also showed good accordance between SureSelect RNA panel and whole transcriptome RNA-seq in three prostate cancer cell lines and two patient samples (Supplementary Fig. S1). It should be noted that SureSelect RNA assay was more sensitive in detecting *AR-V9* than whole transcriptome RNA-seq.

Bioinformatics

For analysis of targeted DNA-seq data, Illumina MiSeq reads were aligned to GRCh37 (hg19) genome using Bowtie2[39]. *AR*, *FOXA1* and *SPOP* variants were called using an in-house pipeline that utilizes samtools mpileup[40]. Filtered variants were annotated using the ANNOVAR software[41]. Variants in dataset 1 were analyzed from the whole transcriptome sequencing data similarly.

AR copy numbers were analyzed by calculating aligned read counts within overlapping 400 bp windows along the targeted regions using bedtools[42]. The median of all AR bait coverage ratios that were obtained by dividing each normalized bait coverage value by the median of all values, was used as the estimate of

AR copy number. Chromosomal rearrangements were called using the in-house Breakfast algorithm that looks for paired-end reads and individual mates overlapping a chromosomal breakpoint.

For *AR* splice variant analysis using targeted or WTS RNA-seq data, Illumina MiSeq reads or HiSeq reads were aligned to an indexed reference fasta file containing unique signature sequences for various *AR-V*s and *AR-FL*. The signatures consisted of 130 bp of the 3' end of upstream exon and 130 bp of the 5' end of downstream exon of a given unique splice junction (Supplementary Table S2). Relative *AR-V* expression was estimated as the percentage of all *AR* transcripts by dividing the number of reads aligned to a given *AR-V* signature by the total number of reads aligning to all the splice junctions containing the same upstream exon.

Expression levels of known AR-regulated genes were determined by aligning RNA-seq reads to GRCh37 genome using TopHat2[43]. Z-scores were calculated from the normalized read counts, and AR-signaling score was computed as the sum of the Z-scores of all AR-regulated genes. Full bioinformatics methods are described in the Supplementary methods.

Immunohistochemistry

Formalin-fixed, paraffin-embedded tumor microarrays of hormone-naïve PC, locally recurrent CRPC and metastatic CRPC (described in Leinonen et al. 2013) were used. Immunohistochemistry for AR (with N-terminal antibody recognizing full length AR and the variants) has been previously described[44]. For AR-V7-specific staining, sections were deparaffinized and antigen retrieval was performed by using Tris-EDTA buffer 0.05% Tween-20 (pH 9) at +98°C for 15 minutes. The staining was performed by Lab Vision Autostainer (ThermoFischer Scientific Inc., Waltham, MA, USA). The primary antibody Anti-

Androgen Receptor (AR-V7 specific) Rabbit Monoclonal Antibody [RM7] (RevMAb Biosciences, San Francisco, CA, USA) and secondary antibody (N-Histofine® Simple Stain MAX PO; Nichirei, Tokyo, Japan) were used. ImmPACT DAB (Vector Laboratories, Burlingame, CA, USA) was used as a chromogen. The sections were counterstained with hematoxylin and mounted with DPX mounting medium (Sigma-Aldrich). The percentage of AR-V7 positive cells between PC and CRPC groups was statistically assessed with Mann-Whitney test.

Results

AR mutations and CNVs are detected only in CRPC cases

Since the mechanisms leading to emergence of CRPC are still largely unknown, we wanted to study the expression of *AR* splicing variants and other *AR* aberrations in tandem during different stages of prostate cancer. For this purpose, we performed low-coverage whole genome DNA sequencing and whole transcriptome sequencing in sample set 1 that included BPH specimens, hormone-naïve prostate cancer from prostatectomies and locally recurrent CRPCs (Supplementary Fig. S2). In addition, we performed targeted *AR* DNA and RNA sequencing in sample set 2 that contained hormone-naïve prostate cancer from prostatectomies and lymph node metastases as well as CRPC metastases (Fig. 1).

First, we wanted to analyze the status of *AR* mutations and CNVs across widely diverse set of samples to better understand their potential link to *AR-V* expression. As expected, *AR* mutations and CNVs were detected only in locally recurrent and metastatic CRPC specimens (Fig. 1 and Supplementary Fig. S2, upper panels). T878A mutation that has been shown to confer agonist activity of flutamide on the AR[45-47] was found in 1/13 (8%) of locally recurrent CRPC specimens and in 2/23 (9%) of metastatic CRPC

specimens. L702H mutation that converts glucocorticoids to AR agonists[48,49] was found in 3/23 (13%) of metastatic CRPC specimens. Indeed, all three patients harboring L702H mutation had been treated with glucocorticoids (Supplementary Table S1; a detailed treatment history is shown for patients having *AR* mutations). Copy number gains (>1 copy of *AR*) or amplifications (>2 *AR* copies) were observed in 4/9 (44%) of locally recurrent CRPC specimens and in 19/23 (83%) of metastatic CRPC specimens, respectively (Fig. 1 and Supplementary Fig. S2). It should be noted that there were striking differences in *AR* copy numbers in metastatic CRPC specimens; for example, the lesion from patient A7 had six *AR* copies whereas the lesion from patient A4 had as many as 68 *AR* copies. In four metastatic CRPC specimens, *AR* gain or amplification co-occurred with *AR* mutation. Since a large body of data, including our current investigation, has established that there are no mutations or copy number aberrations of *AR* in untreated prostate cancers, majority of prostatectomy specimens in sample set 2 were not assayed with the targeted SureSelect DNA panel (Fig. 1). Additionally, data from targeted DNA assay is missing from those metastatic CRPC specimens of which DNA was not available (Fig. 1). The overall average coverage of the targeted regions in the samples ranged from 109X to 1829X, with the average coverage in the *AR* region being somewhat higher (114X–3358X).

The expression of *AR-V*s is highest in CRPCs and associates with expression of *AR-FL*

Next we studied the presence of known *AR-V*s that were detected from the RNA-seq data by aligning the reads against indexed *AR-V* signature sequence file containing exon-exon junction sequences unique to every *AR-V* under investigation (an example of RNA-seq read alignment of patient A17 is visualized in Supplementary Fig. S3). The *AR-V*s detected by our assays included *AR-V3*, *AR-V4*, *AR-V5*, *AR-V6*, *AR-V7* and *AR-V9*. The expression levels of *AR-V4*, *AR-V5* and *AR-V6* were negligible in comparison to *AR-V3*, *AR-V7* and *AR-V9*, and were mainly observed in CRPC metastases. In sample set 1 run by whole

transcriptome RNA-seq, BPH specimens were mainly devoid of *AR-V* expression. Instead, *AR-V3* and *AR-V7* expression were detected in both hormone-naïve prostate cancer from prostatectomies and locally recurrent CRPCs with minimal co-expression of *AR-V9* (Supplementary Fig. S2). Whereas the expression of *AR-V3* was quite similar in the two different categories of samples, higher *AR-V7* expression levels were detected in locally recurrent CRPCs as compared with hormone-naïve prostate cancer from prostatectomies. Since the depth of the whole transcriptome sequencing was not satisfactory (average per-base sequence coverage ranged from 14X–137X) in terms of reliable detection of *AR* variants, we also performed targeted RNA sequencing of the *AR* which provided an average coverage range from 95X to 2247X utilizing sample set 2 (Fig. 1). In sample set 2, the expression level of *AR-V7* but also levels of *AR-V3* and *AR-V9* were higher in metastatic lesions from CRPC cases compared to hormone-naïve prostate cancer from prostatectomy (Fig. 1). The differences were statistically significant for either variant alone (Supplementary Table S3) or when their expression fractions were combined (p=0.0006, unpaired Wilcoxon rank sum test, Table 2). In addition, metastatic CRPC cases expressed significantly more *AR-V3*, *AR-V7* and *AR-V9* compared to non-androgen deprived pelvic lymph node metastases (p=0.0282, unpaired Wilcoxon rank sum test, Table 2). We also studied whether the expression of *AR-V*s is associated with the CNV status (neutral vs. duplicated/amplified *AR*) in sample set 2. There was a modest correlation when CNV status was compared to the combined expression levels of *AR-V3*, *AR-V7* and *AR-V9* (rho=0.39, p=0.005, Spearman's rank correlation).

In sample set 2, *AR-FL* expression was three-fold higher in metastatic CRPC compared to hormone-naïve prostate cancer from prostatectomy specimens whereas in sample set 1, *AR-FL* was expressed five-fold higher in CRPC lesions than in prostatectomy samples. Notably, the expression of *AR-V3*, *AR-V7*, and *AR-V9* was strongly associated with the levels of full-length *AR* in sample set 2 (Fig. 2) and in sample set 1 (Supplementary Fig. S4) suggesting that the expression of *AR* locus drives the expression of *AR-V*s

both in hormone-naïve prostate cancer and in CRPC. Furthermore, there was strong and highly significant correlation between the expression of each individual *AR-V* compared to other *AR-V*s in sample set 2 (Fig. 3).

We also asked whether *AR* variant expression affects the expression of AR-regulated genes. This was done by calculating the summed z-score of five androgen responsive genes (*KLK3*, *FKBP5*, *TMPRSS2*, *ACPP* and *SLC45A3*) (Fig. 1, Supplementary Fig. S2). *AR-V* expression was not associated with AR-regulated gene expression in sample set 2 when the proportion of fractions, when compared to *AR-FL*, of each *AR-V* or all *AR-V*s combined, were plotted against AR signaling score (Supplementary Fig. S5). In addition, we wanted to test whether *AR-V3*, *AR-V7* or *AR-V9* expression correlates in particular to *KLK3* expression. No correlation between either variant and *KLK3* was detected in metastatic CRPC specimens (Supplementary Fig. S6). We next studied mutation status of two AR-regulating genes, *FOXA1* and *SPOP*, in sample set 2. *FOXA1* mutations were found in 3/8 (38%) lymph node metastases and in 1/23 (4%) CRPC metastases whereas *SPOP* mutations were detected in 1/8 (13%) lymph node metastases and in 2/23 (9%) CRPC metastases (Fig. 1). We did not find any association between *FOXA1* or *SPOP* mutation status and AR-regulated gene expression. All mutations found in this study and their variant allele frequencies are shown in Supplementary Table S4.

AR genomic structural rearrangements occur in the context of amplified *AR*

AR genomic structural rearrangements (*AR*-GSRs) were recently identified as a novel class of *AR* alteration using both autopsy CRPC specimens and peripheral blood collected from CRPC patients[31,34]. More importantly, the presence of *AR*-GSRs was associated with expression of *AR-V*s in both studies. To this end, we analyzed *AR* DNA-seq data with our structural variant detection pipeline to identify *AR*-

GSRs, defined as events having at least one breakpoint detected within the *AR* gene region. Average per-base sequence coverage of the *AR* gene region ranged from 114X–3358X and on average 78% of *AR* was covered by at least 10 reads (range 76%–82%). We detected putative *AR*-GSRs in 5/30 metastatic CRPC patients who all harbored a highly amplified *AR* (Supplementary Table S5). All other sample types were negative for *AR*-GSRs when cut-off of 10 supporting split reads was used. It should be noted that none of the *AR*-GSRs occurred along with *AR* missense mutations. The break fusion junctions of *AR*-GSRs were variable demonstrating several types of rearrangements including duplication, deletion, inversion and translocation events. Furthermore, all patients demonstrated unique *AR*-GSR breakpoint locations. Interestingly, Patient A27 displayed a rearrangement that deleted half of exon 4 as well as exons 5 and 6 and was the only patient whose *AR*-GSR was also detected from the RNA-seq sample (Supplementary Fig. S7). This rearrangement may lead to translation of a truncated, constitutively active protein product and could thus have some biological relevance. None of the *AR*-GSRs detected by our pipeline were associated with the expression of previously known *AR-V*s and their variant allele fractions were relatively low (range 2.6–10.9%).

Expression of AR-V7 is heterogeneous at the protein level

To study how the detected differences in AR variant expression between prostate cancer stages are translated to the protein level, we performed immunohistochemistry against AR-V7 with tumor microarrays of hormone-naïve PC from prostatectomies (n=146), locally recurrent CRPCs (n=97), and metastatic CRPC samples (103 metastases in total from 31 patients; 1-5 metastases per patient). We also studied immunohistochemistry of AR (N-terminal antibody recognizing full length AR as well as all variants containing exon 1, including *AR-V3*, *AR-V7* and *AR-V9*). As a positive control we used a sample of 22Rv1 cell line known to contain high AR-V7 expression (Supplementary Fig. S8a). Primarily, AR-

V7 was detected in the nucleus (92% of hormone-naïve, 62% of CRPC and 75% of metastatic samples), although variable cytoplasmic staining could be detected in minority of the samples in all phases of the disease (12% of hormone-naïve, 32% of CRPC and 21% of metastatic samples) (Supplementary Fig. S8b). In contrast to AR staining, the AR-V7 staining was heterogeneous and often present in only a fraction of the cells (Supplementary Fig. S8b). For example, 89% of the positive, hormone-naïve cases had nuclear AR-V7 in less than 10% of the cells (mean value of positive cells 6.4%, median 3.2%) (Supplementary Fig. S9a,b). In CRPC, the number of AR-V7 negative cases increased as compared to hormone-naïve disease (38% vs. 8% of no nuclear AR-V7 detected, respectively) (Supplementary Fig. S9a). Interestingly, many of the AR-V7 negative CRPC samples had a strong mesenchymal phenotype, while the cells in most positive tumors had round, epithelial phenotype. In the positive CRPC cases, the percentage of AR-V7 positive cells increased as compared to hormone-naïve disease, with mean value 24.9% and median 13.6% (Supplementary Fig. S9b). As for the metastatic disease, 88% of the tumors studied had detectable AR-V7 positivity, and all 31 patients had one or more AR-V7 positive metastases. It should be noted that direct comparison of AR-V7 mRNA and protein levels is not possible in most of the cases as samples do not originate from the same tumor areas. However, general observations can be made. For example, patient A28 with highest *AR-V7* expression at the mRNA level in the metastasis subjected to sequencing analysis (Fig. 1) had AR-V7 positivity in all four metastases that were studied with immunohistochemistry.

Discussion

This study describes the *AR* aberration status in two comprehensive patient cohorts including specimens from benign prostatic hyperplasia, untreated localized and metastatic prostate cancer as well as both locally recurrent and metastatic CRPCs. We show that even though *AR-V3*, *AR-V7* and *AR-V9* are

expressed widely in different sample types, they are statistically more highly expressed in metastatic CRPCs in comparison to two hormone-naïve sample groups, prostatectomies and lymph node metastases. This further reinforces the conception that AR-Vs likely have a role in CRPC progression and development of resistance to AR-targeted therapies.

In CRPC metastases, the expression of *AR-V7* was 13% of *AR* transcript at maximum and it was present in 21/29 cases whereas the expression levels of *AR-V3* and *AR-V9* were highly similar (7% of *AR* transcript at maximum) and detected in 23/29 and 22/29 CRPC metastases, respectively. Our finding that *AR-V3*, *AR-V7* and *AR-V9* are present at varying levels also in benign prostate tissue and hormone-naïve primary prostate cancers is in line with previous reports[23,50]. Furthermore, our whole genome and targeted DNA sequencing results were in accordance with previous reports demonstrating that *AR* mutations and amplifications are rare in early stages of untreated prostate cancer but occur much more frequently in patients affected by metastatic CRPC[4,23,50]. In our study, no *AR* mutations or copy number changes were detected in untreated cases; they were observed only in locally recurrent and metastatic CRPC specimens. In CRPC metastases, 5/23 cases harbored an *AR* mutation and 19/23 cases had a copy number gain or amplification. Out of 22 cases of metastatic CRPC of which both DNA- and RNA-seq data was available, all but one patient (patient A5) had at least one *AR* aberration underlining the crucial role of AR in the disease progression. Figure 4 summarizes both genome and RNA level alterations of *AR* detected in this study during different stages of prostate cancer.

It is noteworthy that lymph node metastasis specimens from patients that had not undergone any androgen deprivation therapy did not show elevated levels of *AR-V*s. It has been demonstrated earlier using several prostate cancer cell lines that inhibition of the full-length AR protein via castration, antiandrogen treatment or siRNA induced the expression of AR-V7, although concomitant, yet lesser

increases in full-length AR were also observed[51,52]. It has also been shown that ADT does not directly regulate levels of AR-V7 but rather enhances *AR* gene transcription rate and splicing factor recruitment to *AR* pre-mRNA thus elevating AR-V7 levels[53]. Accordingly, we also showed that the expression levels of *AR-V7* as well as levels of *AR-V3* and *AR-V9* were strongly associated with the levels of full-length *AR* indicating that the *AR-V* expression is dependent on transcription rate of *AR* locus.

Interestingly, we observed that *AR-V7* was strongly co-expressed with *AR-V9* in sample set 2, which is in line with a recent report also demonstrating simultaneous expression of *AR-V7* and *AR-V9* in CRPC metastases[30]. Moreover, our data showed that *AR-V7* was co-expressed with *AR-V3* and there was also a clear positive correlation between expression of *AR-V9* and *AR-V3*. It should be noted that out of 25 CRPC metastases that expressed any *AR* variant, as many as 17 cases harbored expression of all three of these *AR-V*s. Since *AR-V3*, *AR-V7* and *AR-V9* are constitutively active, it is reasonable to expect that their combined contribution to prostate cancer progression might be greater than what could be expected when their effects are studied separately. In our data, *AR-V* expression levels were above 5% when compared to overall *AR* transcript expression levels in 11 metastatic CRPC specimens (sample set 2). Since metastatic CRPC specimens harbored three times higher expression of *AR-FL* in comparison to hormone-naïve prostatectomy samples it would mean that 5% *AR-V* fraction does not yet bring the *AR-V* levels to corresponding levels of *AR-FL* in hormone-naïve prostate cancer. However, the levels of *AR-V* required to drive an androgen-independent transcriptome are unknown.

It has been demonstrated in several cell line studies that AR-Vs are able to induce the expression of AR-controlled genes such as *KLK3*, *TMPRSS2* and *FKBP5* in the absence of androgens or AR-FL[19,20,54,55]. Therefore, we interrogated the levels of classical AR-regulated genes in our sample sets and calculated the summed z-score of five androgen responsive genes (*KLK3*, *FKBP5*, *TMPRSS2*, *ACPP* and *SLC45A3*).

There was no association between the expression levels of *AR-V3*, *AR-V7* or *AR-V9* and z-score in sample set 2. Likewise, no correlation was detected when *KLK3* expression was compared to the expression levels of these *AR-V*s in CRPC metastases. One explanation for this discrepancy could be the fact that metastatic CRPC samples expressing the highest levels of *AR-V*s were taken at autopsy making it highly unlikely that AR-regulation was anymore classical at the late stage of the disease. In addition, it has previously been demonstrated that bone metastases with high *AR-V* levels did not show high levels of *KLK3*, *KLK2*, *FKBP5*, *TMPRSS2* and *NKX3-1* whereas the levels of other transcripts known to be positively regulated by AR were elevated (including *CDK1*, *CYCLINA2*, *HSP27* and *C-MYC*)[56]. Therefore, it seems that the expression profile induced by AR-Vs can be context-dependent and might not correspond to the effects observed in cell lines.

As protein expression does not always fully correlate with mRNA expression, and as the AR-Vs may also be regulated post-transcriptionally, it is of importance to study their expression also at the protein level. We assessed the expression of AR-V7 by immunohistochemistry in hormone-naïve PC, locally recurrent CRPC and metastatic CRPC samples. Although a third of CRPCs in this cohort were found negative for AR-V7 protein, the results support the view that AR-V7 expression increases during castration resistance, and that the protein is present in most prostate cancer metastases. It is noteworthy that expression of AR-V7 is highly more heterogeneous than that of AR overall. This indicates either differences in transcriptional expression of AR-V7 between tumor cells, or heterogeneous post-transcriptional regulation of it within tumor cell populations.

Genomic structural rearrangements (GSRs) have recently been shown to define a class of *AR* aberrations occurring at a considerable frequency in CRPC material[31,34]. Henzler et al. studied *AR*-GSRs in 30 rapid autopsy CRPC soft tissue metastases obtained from 15 patients and found that 10/30 metastases (6/15

patients) displayed at least one *AR*-GSR event. Instead, De Laere et al. utilized liquid biopsies from 30 chemotherapy pretreated or chemo-naïve CRPC patients and detected at least one *AR*-GSR in 15/30 patients. We observed *AR*-GSRs in 5/30 patients with metastatic CRPC, which is considerably less when compared to these prior findings. The reason for this discrepancy is unclear, but it can be at least partly due to the fact that there were more uncovered regions in our assay (76%–82% of *AR* covered) than in Henzler et al. assay (83–89% of *AR* covered). Interestingly, *AR*-GSRs were not detected in the context of *AR* missense mutations in both our and Henzler et al. sample cohorts[31]. Furthermore, in our material, the break fusion junctions of *AR*-GSRs were variable demonstrating several types of rearrangements but none of the *AR*-GSRs were associated with the expression of previously known *AR-V*s. Patient A27 was the only one whose *AR*-GSR was also detected by RNA-seq and his variant allele fraction was also the highest being 10.9%. For other patients with *AR*-GSRs the variant allele fractions ranged from 2.6% to 8.6%. In CRPC metastases, half of the *AR*-GSR positive patients expressed *AR-V*s[31] whereas all but one of *AR*-GSR positive patients who were liquid-biopsied harbored *AR-V* expression[34]. Together, these results demonstrate that the connection of *AR*-GSRs and the expression of *AR-V*s is highly variable in different sample cohorts. It is also noteworthy that in the study of Henzler et al., the only previously reported variant that was associated with the presence of *AR*-GSRs were *AR-V7* and *AR-V12* (*ARv567es*) but the data from De Laere et al. showed that the majority of *AR*-GSR positive patients expressed multiple previously reported *AR-V*s. *AR*-GSRs were restricted to CRPC specimens in both our and Henzler et al. data suggesting that they are yet another means of CRPC to retain AR signaling.

In conclusion, the finding that *AR-V* expression levels increase in patients treated with androgen deprivation therapy might indicate that there is a clonal selection pressure on the different tumor clones in order to maintain functional AR signaling independent of the androgen levels. We provide evidence that *AR-V3*, *AR-V7* and *AR-V9* are co-expressed in metastatic CRPC highlighting the fact that targeting

of the AR ligand-binding domain might not be sufficient to achieve clinically relevant treatment responses. Consequently, inhibiting AR function via regions common to all AR-Vs is likely to provide additional benefit to patients with CRPC.

Additional information

Ethics approval and consent to participate: The use of benign prostatic hyperplasia, hormone-naïve prostate cancer and locally recurrent CRPC specimens was approved by the ethical committee of the Tampere University Hospital. Written informed consent was obtained from the subjects. Samples from non-androgen deprived pelvic lymph node metastases were obtained under Johns Hopkins Medicine IRB approval 04-05-03-02. Metastatic CRPC specimens were obtained from 30 men who participated in the PELICAN (Project to ELIminate lethal CANcer) integrated clinical-molecular autopsy study of metastatic prostate cancer. Subjects consented to participate in the Johns Hopkins Medicine IRB-approved study between 1995 and 2005. The study was performed in accordance with the Declaration of Helsinki.

Availability of data and material: All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Conflict of interest: The authors declare no conflict of interest.

Funding: The work was supported in parts by grants from the Finnish Funding Agency for Technology and Innovation, the Academy of Finland, the Cancer Society of Finland, the Sigrid Juselius Foundation and the Medical Research Fund of Tampere University Hospital.

Authors' contributions: HMLK wrote the article and conducted the NGS laboratory experiments. RH and MA conducted the computational analyses. LL performed the IHC stainings and analyses. AB helped with the NGS laboratory experiments. KK, MN, HGL and TV designed the study. TLT, WBI and GSB provided materials. HMLK and TV analyzed the data. All authors read and approved the manuscript.

Acknowledgements: We are grateful to Orion for providing the AR-V7 specific antibody. We thank Päivi Martikainen, Riina Kylätie and Marika Vähä-Jaakkola for their technical assistance.

Supplementary information is available at the British Journal of Cancer's website.

References

1. Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinanen R, Palmberg C *et al.* In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* 1995; **9**: 401-6.

2. Waltering KK, Helenius MA, Sahu B, Manni V, Linja MJ, Janne OA *et al.* Increased expression of androgen receptor sensitizes prostate cancer cells to low levels of androgens. *Cancer Res* 2009; **69**: 8141-9.

3. Brooke GN, Bevan CL. The role of androgen receptor mutations in prostate cancer progression. *Curr Genomics* 2009; **10**: 18-25.

4. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* 2012; **487**: 239-43.

5. Cai C, Balk SP. Intratumoral androgen biosynthesis in prostate cancer pathogenesis and response to therapy. *Endocr Relat Cancer* 2011; **18**: R175-82.

6. Locke JA, Guns ES, Lubik AA, Adomat HH, Hendy SC, Wood CA *et al.* Androgen levels increase by intratumoral de novo steroidogenesis during progression of castration-resistant prostate cancer. *Cancer Res* 2008; **68**: 6407-15.

7. Montgomery RB, Mostaghel EA, Vessella R, Hess DL, Kalhorn TF, Higano CS *et al.* Maintenance of intratumoral androgens in metastatic prostate cancer: a mechanism for castration-resistant tumor growth. *Cancer Res* 2008; **68**: 4447-54.

8. Palmberg C, Koivisto P, Kakkola L, Tammela TL, Kallioniemi OP, Visakorpi T. Androgen receptor gene amplification at primary progression predicts response to combined androgen blockade as second line therapy for advanced prostate cancer. *J Urol* 2000; **164**: 1992-5.

9. Gottlieb B, Beitel LK, Nadarajah A, Paliouras M, Trifiro M. The androgen receptor gene mutations database: 2012 update. *Hum Mutat* 2012; **33**: 887-94.

10. Eisermann K, Wang D, Jing Y, Pascal LE, Wang Z. Androgen receptor gene mutation, rearrangement, polymorphism. *Transl Androl Urol* 2013; **2**: 137-47.

11. Attard G, Reid AH, A'Hern R, Parker C, Oommen NB, Folkerd E *et al.* Selective inhibition of CYP17 with abiraterone acetate is highly active in the treatment of castration-resistant prostate cancer. *J Clin Oncol* 2009; **27**: 3742-8.

12. Tran C, Ouk S, Clegg NJ, Chen Y, Watson PA, Arora V *et al.* Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 2009; **324**: 787-90.

13. Annala M, Vandekerkhove G, Khalaf D, Taavitsainen S, Beja K, Warner EW *et al.* Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. *Cancer Discov* 2018; **8**: 444-57.

14. Romanel A, Gasi Tandefelt D, Conteduca V, Jayaram A, Casiraghi N, Wetterskog D *et al.* Plasma AR and abiraterone-resistant prostate cancer. *Sci Transl Med* 2015; **7**: 312re10.

15. Azad AA, Volik SV, Wyatt AW, Haegert A, Le Bihan S, Bell RH *et al.* Androgen Receptor Gene Aberrations in Circulating Cell-Free DNA: Biomarkers of Therapeutic Resistance in Castration-Resistant Prostate Cancer. *Clin Cancer Res* 2015; **21**: 2315-24.

16. Wyatt AW, Azad AA, Volik SV, Annala M, Beja K, McConeghy B *et al.* Genomic Alterations in Cell-Free DNA and Enzalutamide Resistance in Castration-Resistant Prostate Cancer. *JAMA Oncol* 2016; **2**: 1598-606.

17. Conteduca V, Wetterskog D, Sharabiani MTA, Grande E, Fernandez-Perez MP, Jayaram A *et al.* Androgen receptor gene status in plasma DNA associates with worse outcome on enzalutamide or abiraterone for castration-resistant prostate cancer: a multi-institution correlative biomarker study. *Ann Oncol* 2017; **28**: 1508-16.

18. Scher HI, Beer TM, Higano CS, Anand A, Taplin ME, Efstathiou E *et al.* Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet* 2010; **375**: 1437-46.

19. Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, Tindall DJ. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res* 2008; **68**: 5469-77.

20. Hu R, Dunn TA, Wei S, Isharwal S, Veltri RW, Humphreys E *et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res* 2009; **69**: 16-22.

21. Lu C, Luo J. Decoding the androgen receptor splice variants. *Transl Androl Urol* 2013; **2**: 178-86.

22. Jenster G, van der Korput HA, van Vroonhoven C, van der Kwast TH, Trapman J, Brinkmann AO. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. *Mol Endocrinol* 1991; **5**: 1396-404.

23. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* 2015; **161**: 1215-28.

24. Antonarakis ES, Lu C, Luber B, Wang H, Chen Y, Nakazawa M *et al.* Androgen Receptor Splice Variant 7 and Efficacy of Taxane Chemotherapy in Patients With Metastatic Castration-Resistant Prostate Cancer. *JAMA Oncol* 2015; **1**: 582-91.

25. Antonarakis ES, Lu C, Wang H, Luber B, Nakazawa M, Roeser JC *et al.* AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med* 2014; **371**: 1028-38.

26. Antonarakis ES, Lu C, Luber B, Wang H, Chen Y, Zhu Y *et al.* Clinical Significance of Androgen Receptor Splice Variant-7 mRNA Detection in Circulating Tumor Cells of Men With Metastatic Castration-Resistant Prostate Cancer Treated With First- and Second-Line Abiraterone and Enzalutamide. *J Clin Oncol* 2017; **35**: 2149-56.

27. Seitz AK, Thoene S, Bietenbeck A, Nawroth R, Tauber R, Thalgott M *et al.* AR-V7 in Peripheral Whole Blood of Patients with Castration-resistant Prostate Cancer: Association with Treatment-specific Outcome Under Abiraterone and Enzalutamide. *Eur Urol* 2017; **72**: 828-34.

28. Steinestel J, Luedeke M, Arndt A, Schnoeller TJ, Lennerz JK, Wurm C *et al.* Detecting predictive androgen receptor modifications in circulating prostate cancer cells. *Oncotarget* 2015. https://doi.org/10.18632/oncotarget.3925

29. Del Re M, Biasco E, Crucitta S, Derosa L, Rofi E, Orlandini C *et al.* The Detection of Androgen Receptor Splice Variant 7 in Plasma-derived Exosomal RNA Strongly Predicts Resistance to Hormonal Therapy in Metastatic Prostate Cancer Patients. *Eur Urol* 2017; **71**: 680-687.

30. Kohli M, Ho Y, Hillman DW, Van Etten JL, Henzler C, Yang R *et al.* Androgen Receptor Variant AR-V9 Is Coexpressed with AR-V7 in Prostate Cancer Metastases and Predicts Abiraterone Resistance. *Clin Cancer Res* 2017; **23**: 4704-4715.

31. Henzler C, Li Y, Yang R, McBride T, Ho Y, Sprenger C *et al.* Truncation and constitutive activation of the androgen receptor by diverse genomic rearrangements in prostate cancer. *Nat Commun* 2016; **7**: 13668.

32. Li Y, Alsagabi M, Fan D, Bova GS, Tewfik AH, Dehm SM. Intragenic rearrangement and altered RNA splicing of the androgen receptor in a cell-based model of prostate cancer progression. *Cancer Res* 2011; **71**: 2108-17.

33. Li Y, Hwang TH, Oseth LA, Hauge A, Vessella RL, Schmechel SC *et al.* AR intragenic deletions linked to androgen receptor splice variant expression and activity in models of prostate cancer progression. *Oncogene* 2012; **31**: 4759-67.

34. De Laere B, van Dam PJ, Whitington T, Mayrhofer M, Diaz EH, Van den Eynden G *et al.* Comprehensive Profiling of the Androgen Receptor in Liquid Biopsies from Castration-resistant Prostate Cancer Reveals Novel Intra-AR Structural Variation and Splice Variant Expression Patterns. *Eur Urol* 2017; **72**: 192-200.

35. Bova GS, Carter BS, Bussemakers MJ, Emi M, Fujiwara Y, Kyprianou N *et al.* Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer. *Cancer Res* 1993; **53**: 3869-73.

36. Annala M, Kivinummi K, Tuominen J, Karakurt S, Granberg K, Latonen L *et al.* Recurrent SKIL-activating rearrangements in ETS-negative prostate cancer. *Oncotarget* 2015; **6**: 6235-50.

37. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JM, Papaemmanuil E *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* 2015; **520**: 353-7.

38. Bova GS, Kallio HM, Annala M, Kivinummi K, Hognas G, Hayrynen S *et al.* Integrated clinical, whole-genome, and transcriptome analysis of multisampled lethal metastatic prostate cancer. *Cold Spring Harb Mol Case Stud* 2016; **2**: a000752.

39. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**: 357-9.

40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078-9.

41. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164.

42. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; **26**: 841-2.

43. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013; **14**: R36.

44. Leinonen KA, Saramaki OR, Furusato B, Kimura T, Takahashi H, Egawa S *et al.* Loss of PTEN is associated with aggressive behavior in ERG-positive prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2013; **22**: 2333-44.

45. Lallous N, Volik SV, Awrey S, Leblanc E, Tse R, Murillo J *et al.* Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. *Genome Biol* 2016; **17**: 10.

46. Zhou J, Liu B, Geng G, Wu JH. Study of the impact of the T877A mutation on ligand-induced helix-12 positioning of the androgen receptor resulted in design and synthesis of novel antiandrogens. *Proteins* 2010; **78**: 623-37.

47. 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.

48. Carreira S, Romanel A, Goodall J, Grist E, Ferraldeschi R, Miranda S *et al.* Tumor clone dynamics in lethal prostate cancer. *Sci Transl Med* 2014; **6**: 254ra125.

49. Zhao XY, Malloy PJ, Krishnan AV, Swami S, Navone NM, Peehl DM *et al.* Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. *Nat Med* 2000; **6**: 703-6.

50. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 2015; **163**: 1011-25.

51. Yu Z, Chen S, Sowalsky AG, Voznesensky OS, Mostaghel EA, Nelson PS *et al.* Rapid induction of androgen receptor splice variants by androgen deprivation in prostate cancer. *Clin Cancer Res* 2014; **20**: 1590-600.

52. Hu R, Lu C, Mostaghel EA, Yegnasubramanian S, Gurel M, Tannahill C *et al.* Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Res* 2012; **72**: 3457-62.

53. Liu LL, Xie N, Sun S, Plymate S, Mostaghel E, Dong X. Mechanisms of the androgen receptor splicing in prostate cancer cells. *Oncogene* 2014; **33**: 3140-50.

54. Guo Z, Yang X, Sun F, Jiang R, Linn DE, Chen H *et al.* A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer Res* 2009; **69**: 2305-13.

55. Sun S, Sprenger CC, Vessella RL, Haugk K, Soriano K, Mostaghel EA *et al.* Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J Clin Invest* 2010; **120**: 2715-30.

56. Hornberg E, Ylitalo EB, Crnalic S, Antti H, Stattin P, Widmark A *et al.* Expression of androgen receptor splice variants in prostate cancer bone metastases is associated with castration-resistance and short survival. *PLoS One* 2011; **6**: e19059.

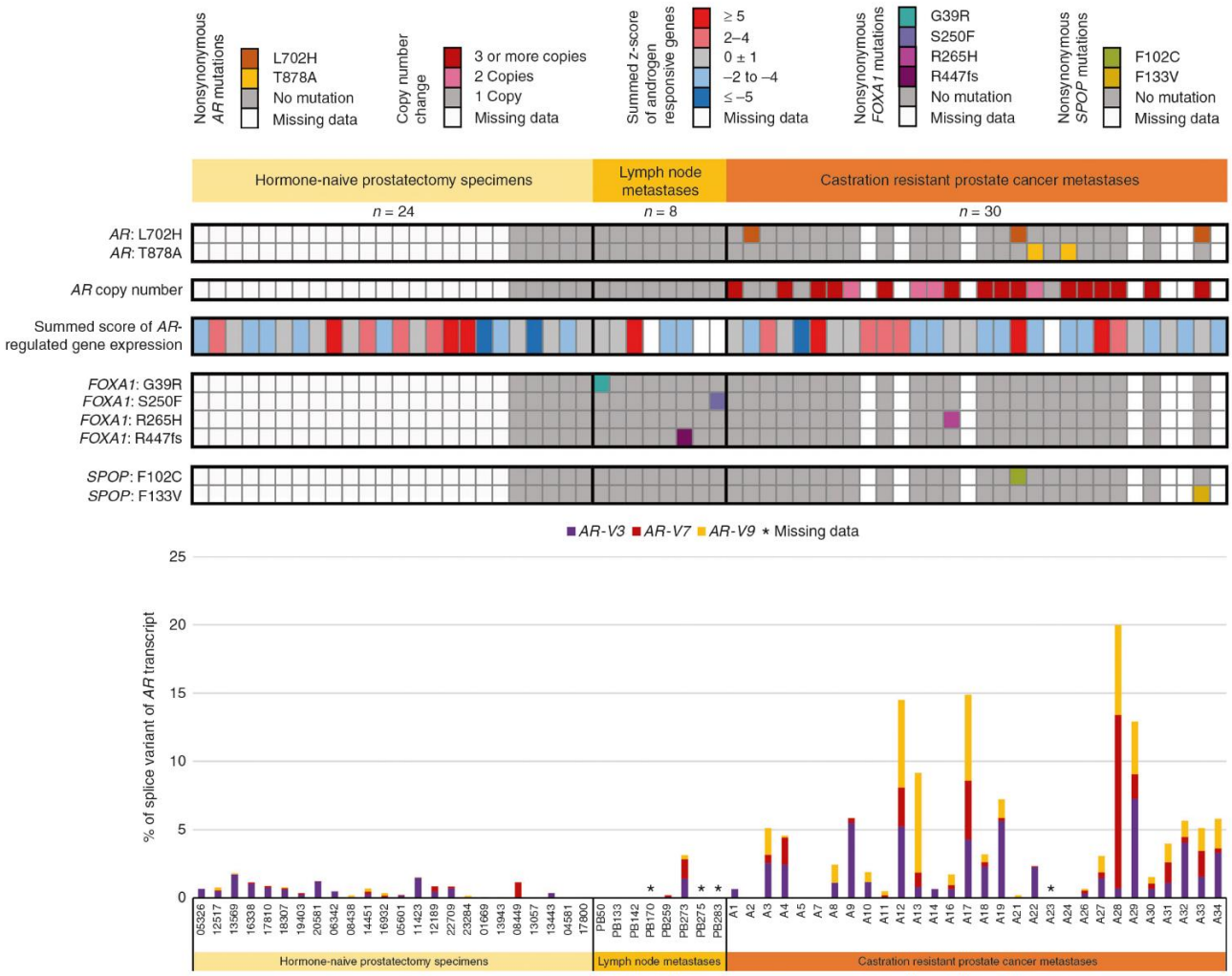


Figure 1. Combined DNA and RNA sequencing data from sample set 2 assayed by targeted SureSelect *AR* sequencing. *AR* mutations, copy number alterations, summed score of AR-regulated gene expression and *AR-V* expression level as a fraction of *AR* transcript are shown. *AR-V* fractions are shown as CI95 lower bound values. Additionally, *FOXA1* and *SPOP* mutation status is included.

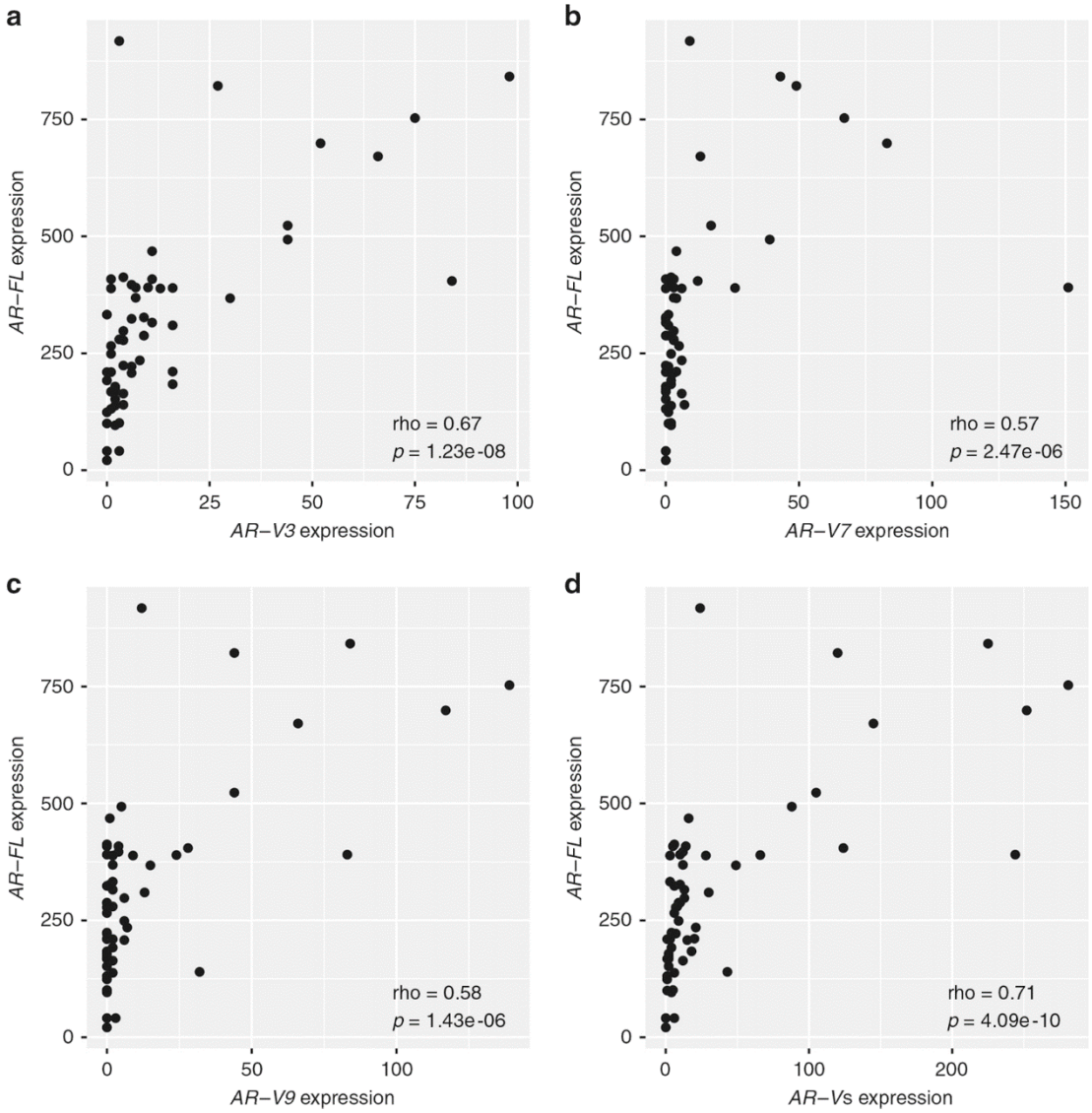


Figure 2. The correlation between *AR-FL* mRNA expression and mRNA expression of (a) *AR-V3*, (b) *AR-V7*, (c) *AR-V9*, (d) all three *AR-V*s combined utilizing specimens from sample set 2. The counts of splice junction reads indicative of *AR-FL* or *AR-V*s are plotted in the y- and x-axis, respectively. Spearman's rank correlation coefficients and p-values computed via the asymptotic t approximation are also shown in the figures.

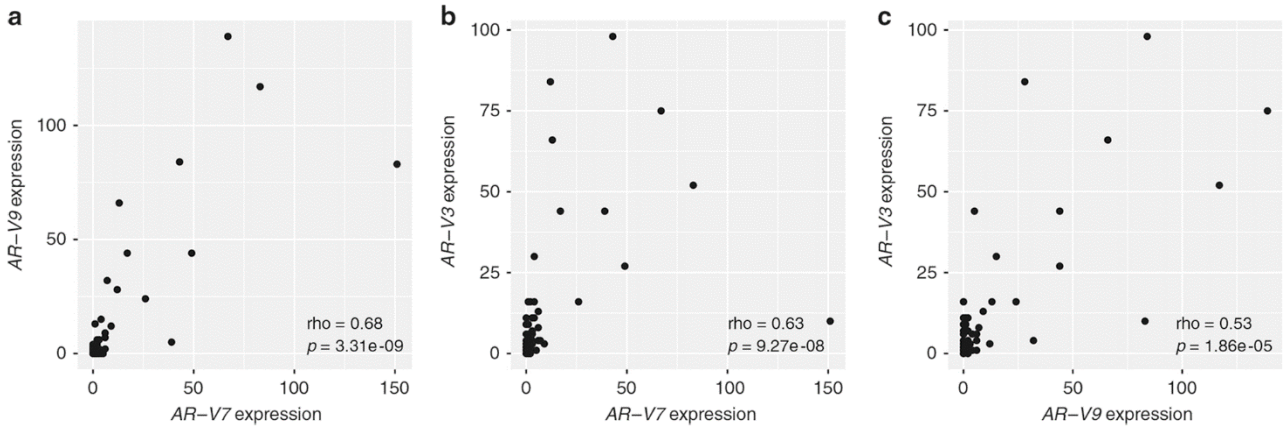
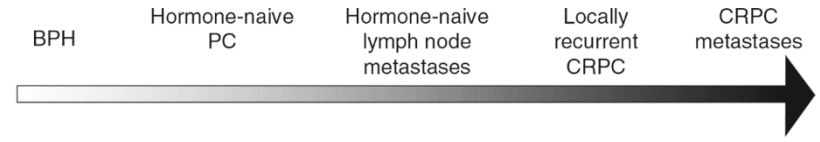


Figure 3. The correlation between (a) *AR-V7* and *AR-V9* mRNA expression, (b) *AR-V7* and *AR-V3* mRNA expression and (c) *AR-V9* and *AR-V3* mRNA expression utilizing specimens from sample set 2. The counts of splice junction reads indicative of given *AR-V*s are plotted in the y- and x-axis. Spearman's rank correlation coefficients and p-values computed via the asymptotic t approximation are also shown in the figures.



Genome level alteration					
T878A mutation	Not present	Not present	Not present	8%	9%
L702H mutation	Not present	Not present	Not present	Not present	13%
CN gain	Not present	Not present	Not present	22%	17%
CN amplification	Not present	Not present	Not present	22%	65%
AR-GSR	Not present	Not present	Not present	Not present	17%
RNA level alteration					
AR-V low	38%	79%	40%	50%	48%
AR-V high	Not present	Not present	Not present	17%	38%

Figure 4. Summary of the frequency of the genome and RNA level alterations of *AR* during different stages of prostate cancer. Copy number (CN) changes of *AR* are presented as gains (>1 copy of *AR*) and amplifications (>2 *AR* copies). *AR-V* expression levels are divided into *AR-V* low (<5% of splice variant of *AR* transcript) and *AR-V* high (>5% of splice variant of *AR* transcript) groups. The data is from MiSeq assays for all other sample groups except for BPH and locally recurrent CRPC whose data is from HiSeq assays.

Table 1. Sample sets utilised in the study

	Number of samples	Origin	Subgroups	Sequencing method
Sample set 1	55	Prostatectomy specimens of non-treated cancer, transurethral resection specimens of locally recurrent CRPC and BPH specimens obtained by several methods	BPH (12), PC (30), locally recurrent CRPC (13)	Whole genome and whole transcriptome/ HiSeq
Sample set 2	80	Prostatectomy specimens of non-treated cancer, lymph node metastases collected during lymphadenectomy, metastatic CRPC specimens and noncancerous control specimens collected from 30 individuals at autopsy	PC (24), lymph node metastasis (8), metastatic CRPC (30), noncancerous control (18)	SureSelect targeted DNA and targeted RNA/MiSeq

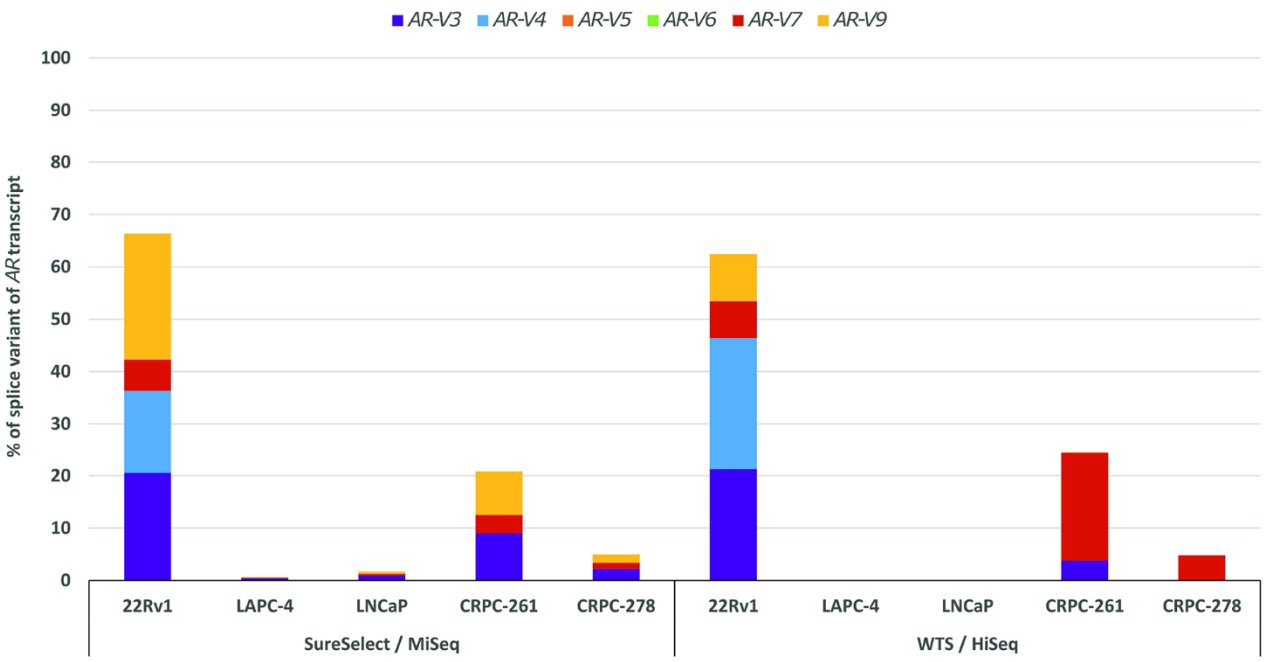
BPH benign prostatic hyperplasia, *PC* prostate cancer, *CRPC* castration-resistant prostate cancer

Table 2. Statistical comparison of combined expression of *AR-V3*, *AR-V7* and *AR-V9* in different sample types using two-tailed, unpaired Mann-Whitney *U* test

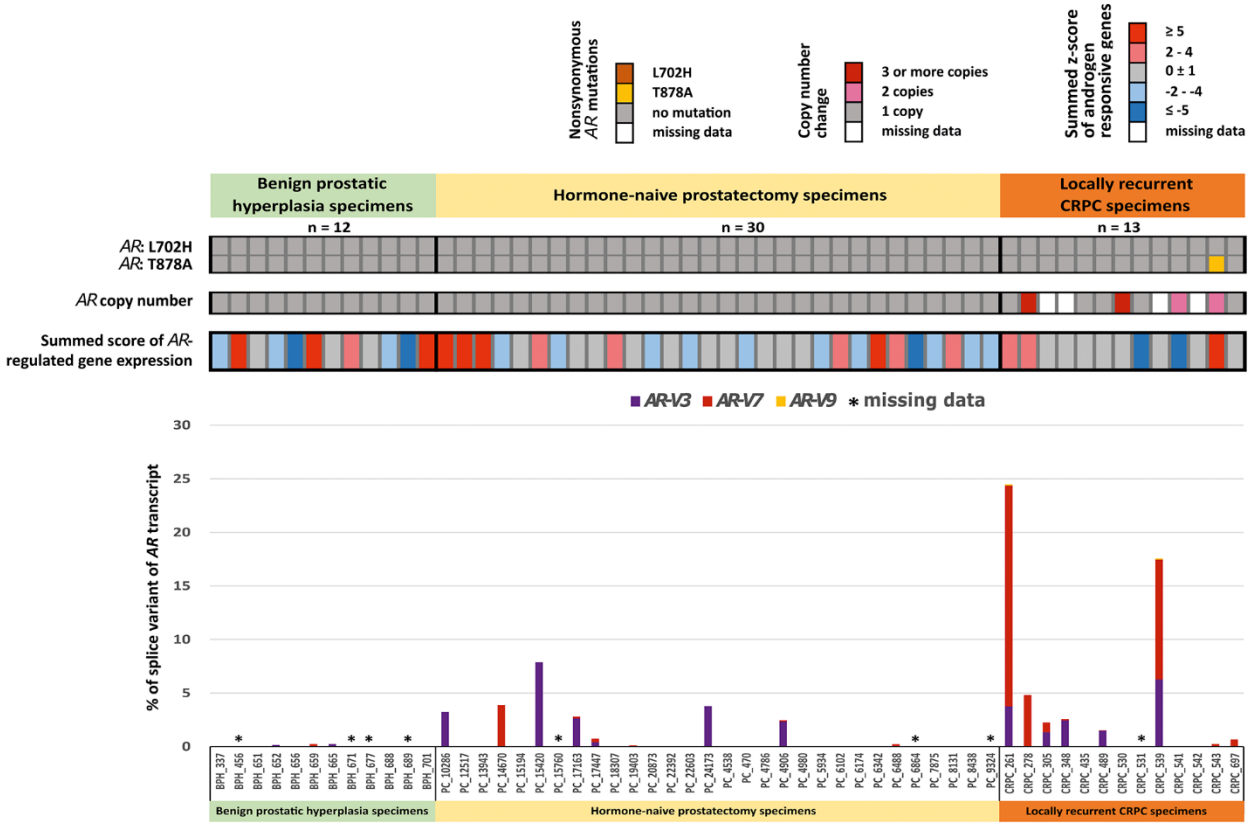
	Prostatectomy	Lymph node metastases
Prostatectomy	-	
Lymph node metastases	0.1765	-
CRPC metastases	**0.0006*** **	**0.0282***

CRPC castration-resistant prostate cancer. $^{*}p < 0.05$, $^{***}p < 0.001$. Statistically significant *p*-values are shown in bold

Supplementary files

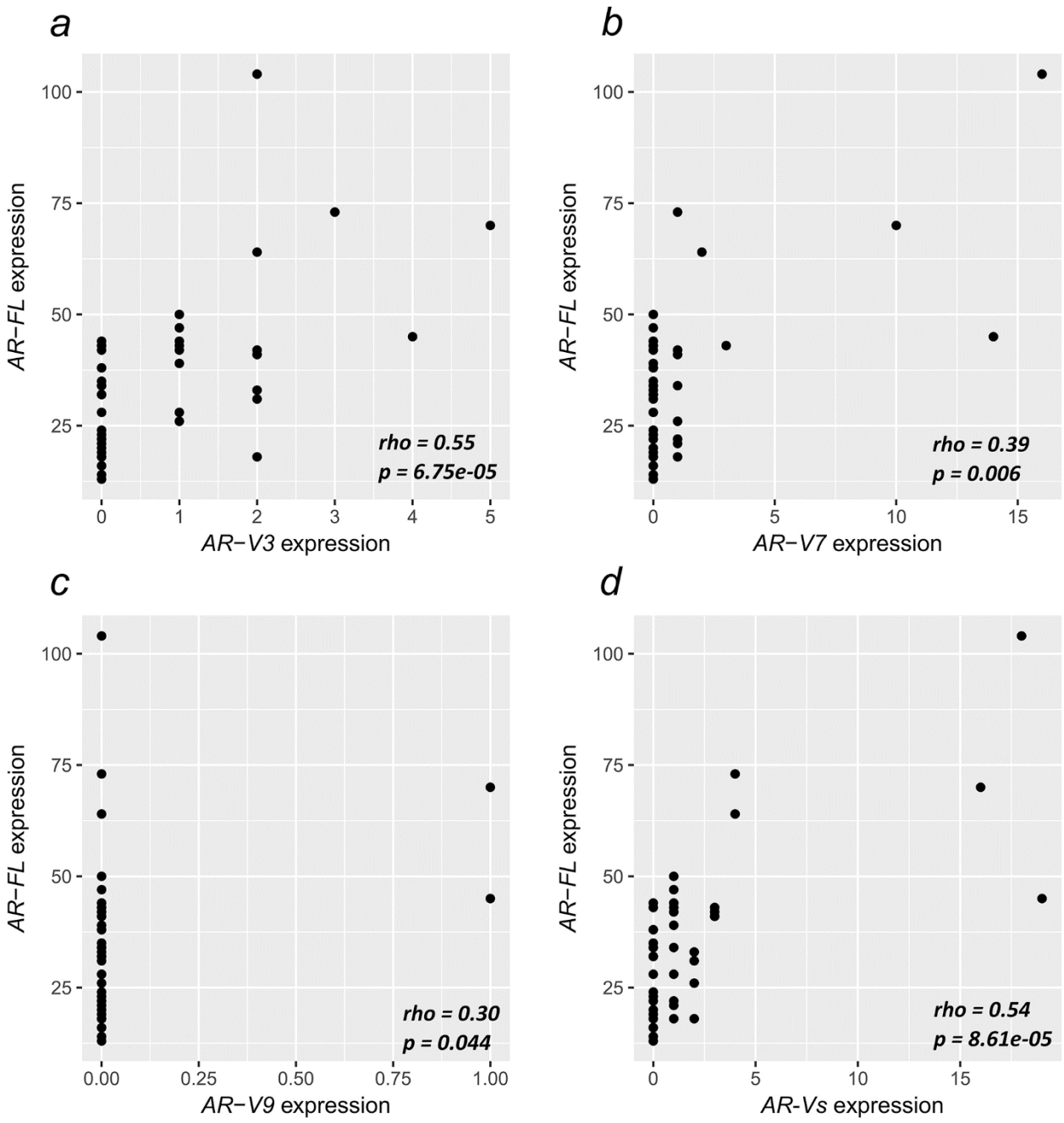


Supplementary Figure S1. Validation of targeted SureSelect *AR* splicing variant detection assay using cell lines and two CRPC samples. SureSelect assay performance was compared to whole transcriptome sequencing data from this study. *AR-V* expression level as a fraction of *AR* transcript is shown.

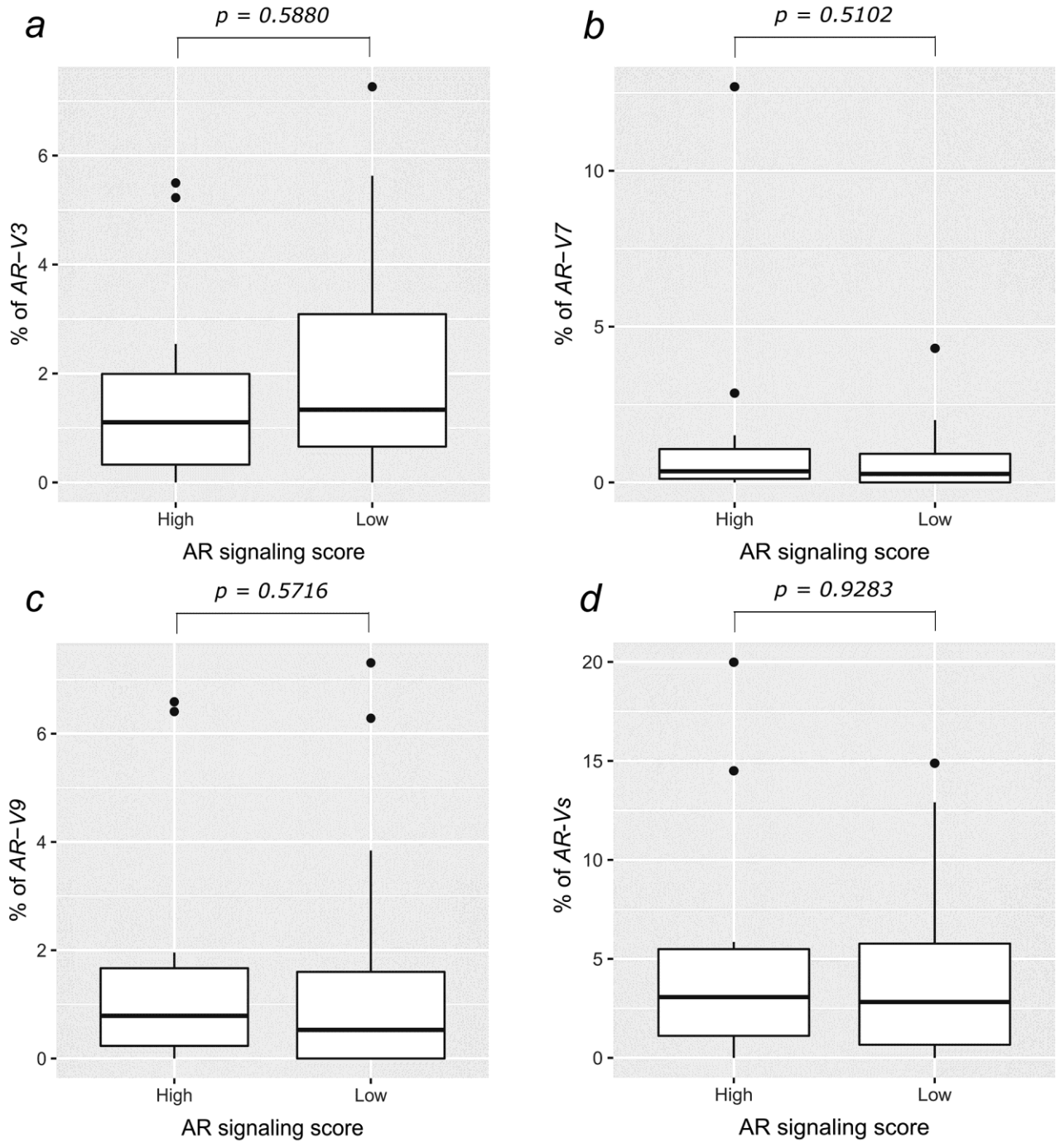


Supplementary Figure S2. Combined DNA and RNA sequencing data from sample set 1 assayed by whole genome and whole transcriptome sequencing. *AR* mutations, copy number alterations, summed score of AR-regulated gene expression and *AR-V* expression level as a fraction of *AR* transcript are shown. *AR-V* fractions are shown as CI95 lower bound values.

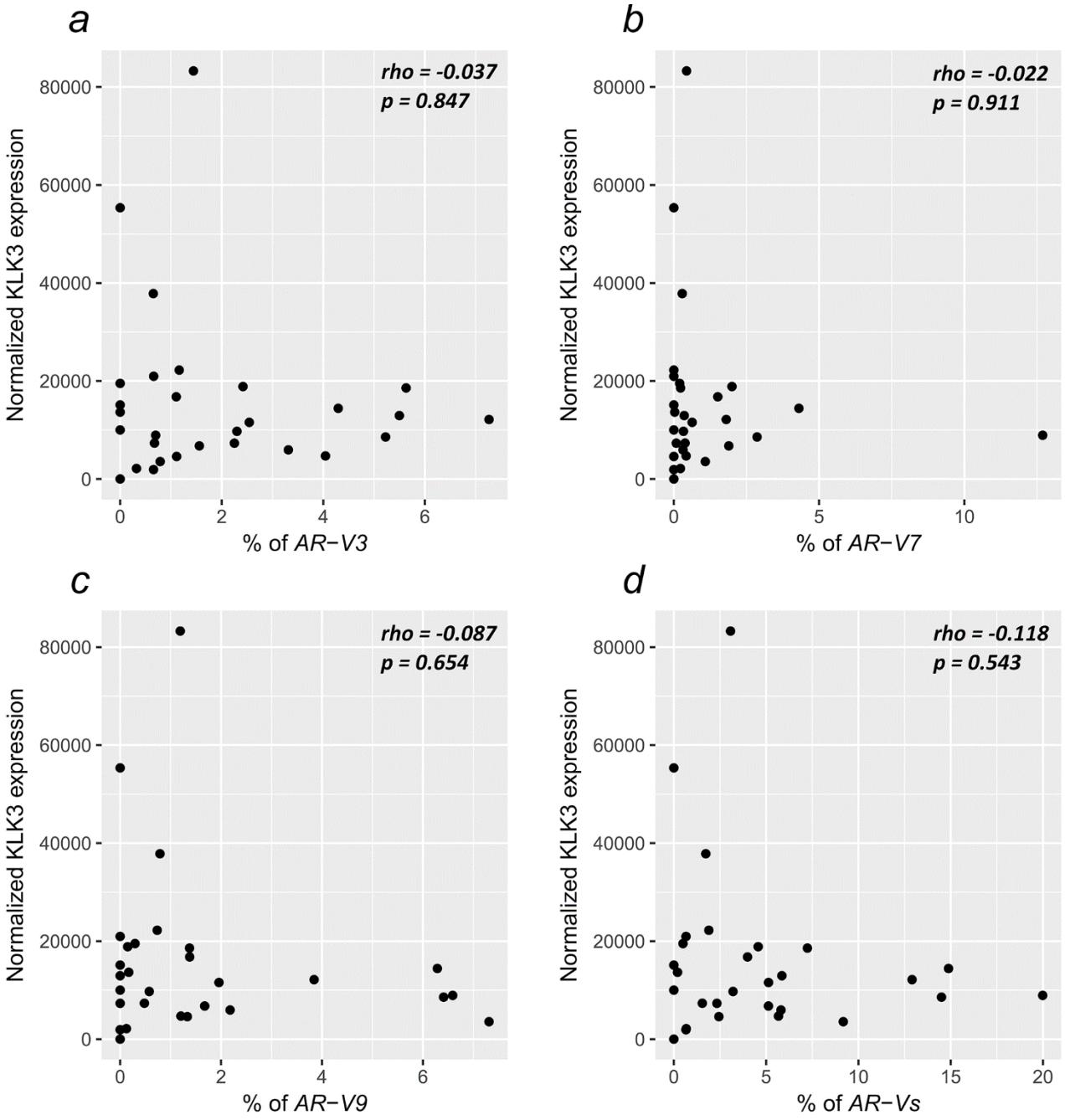
Supplementary Figure S3. RNA-seq read alignment visualization example of patient A17, showing split reads aligning to exon 3 and one of the cryptic exons CE1, CE3, CE4 and CE5. This type of reads were used for the relative quantification of the *AR-V*s by aligning them to the *AR-V* signature sequence reference.



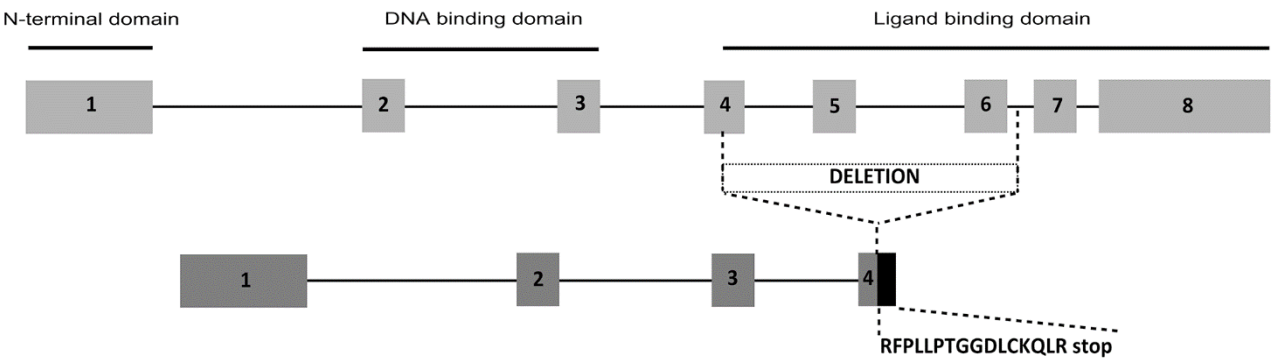
Supplementary Figure S4. The correlation between *AR-FL* mRNA expression and mRNA expression of (a) *AR-V3*, (b) *AR-V7*, (c) *AR-V9*, (d) all three *AR-V*s combined utilizing specimens from sample set 1. Spearman's rank correlation coefficients and p-values computed via the asymptotic t approximation are also shown in the figures.



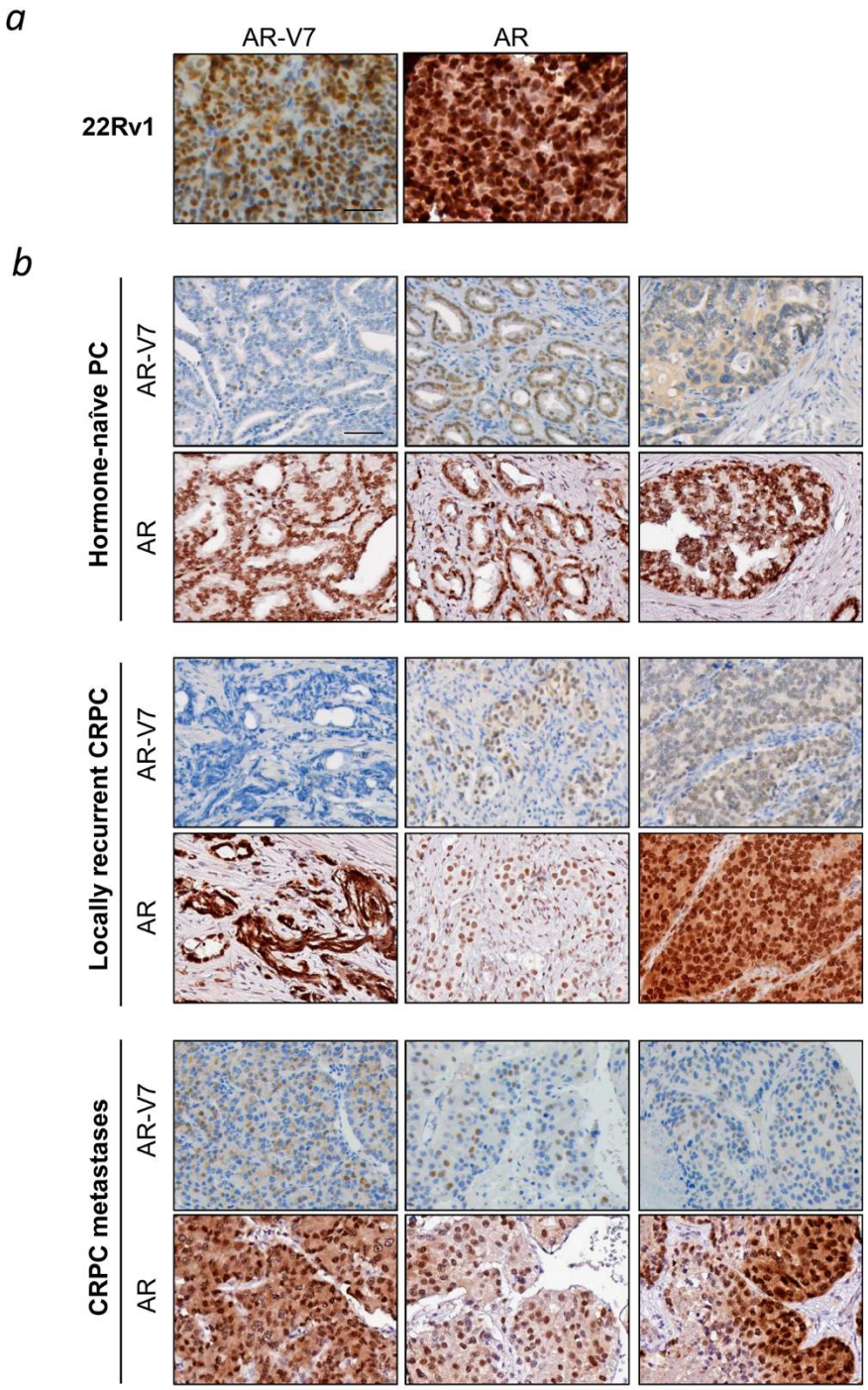
Supplementary Figure S5. The association between AR signaling score and the fraction of (a) *AR-V3*, (b) *AR-V7*, (c) *AR-V9*, (d) all three *AR-V*s combined utilizing specimens from sample set 2. "High" group: summed Z-score is above the mean of all samples. "Low" group: summed Z-score is below the mean of all samples.



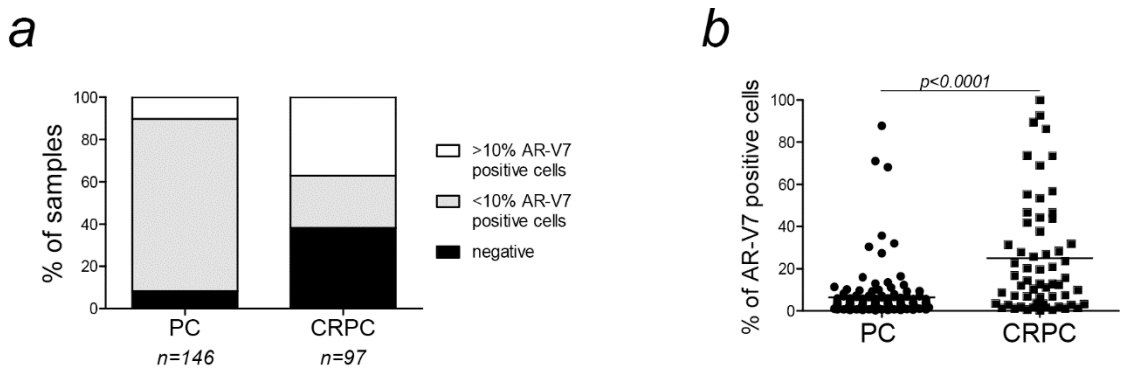
Supplementary Figure S6. The correlation between normalized *KLK3* expression and fraction of (a) *AR-V3*, (b) *AR-V7*, (c) *AR-V9*, (d) all three *AR-V*s combined utilizing specimens from sample set 2. Spearman's rank correlation coefficients and p-values computed via the asymptotic t approximation are also shown in the figures.



Supplementary Figure S7. *AR* genomic structural rearrangement (*AR*-GSR) detected in metastatic CRPC sample from patient A27. The rearrangement resulted in a truncated variant that harbored exons 1-3 and half of exon 4.



Supplementary Figure S8. Example microscopic images of AR and AR-V7 immunohistochemistry. (a) 22Rv1 cells known to contain high AR-V7 were used as a positive control for the stainings of AR-V7 (left panel) and AR (right panel). (b) Three representative tumors with different staining patterns are shown for samples of hormone-naïve PC, locally recurrent CRPC and CRPC metastases with AR-V7 (upper panels) and AR (lower panels). Scale bars 50 μm (a) and 100 μm (b).



Supplementary Figure S9. Results of AR-V7 immunohistochemistry in hormone-naïve PC and locally recurrent CRPC. (a) Percentage of cases with no AR-V7 positive cells (negative) and cases with AR-V7 positive tumor cells below or above 10% of all tumor cells. (b) Percentage of AR-V7 positive tumor cells within the positive samples in (a).

Supplementary Table S1. Clinicopathological characteristics of the prostate cancer cases and treatments of CRPC cases from sample set 1 and 2. Detailed treatment modalities are shown for metastatic CRPC patients that harbored AR mutations.

Sample set 1

Sample ID	PSA at diagnosis	pT stage	Prostatectomy Gleason score	Hormonal treatment	NGS method
BPH_337					whole genome, whole transcriptome
BPH_456					whole genome, whole transcriptome
BPH_651					whole genome, whole transcriptome
BPH_652					whole genome, whole transcriptome
BPH_656					whole genome, whole transcriptome
BPH_659					whole genome, whole transcriptome
BPH_665					whole genome, whole transcriptome
BPH_671					whole genome, whole transcriptome
BPH_677					whole genome, whole transcriptome
BPH_688					whole genome, whole transcriptome
BPH_689					whole genome, whole transcriptome
BPH_701					whole genome, whole transcriptome
PC_10286	5,4	T2b	<7		whole genome, whole transcriptome
PC_12517	5,9	T2b	7		whole genome, whole transcriptome, targeted RNA
PC_13943	4,9	T2b	7		whole genome, whole transcriptome, targeted RNA
PC_14670	7,5	T3b	8		whole genome, whole transcriptome
PC_15194	9,5	T3b	7		whole genome, whole transcriptome
PC_15420	12,6	T3a	7		whole genome, whole transcriptome
PC_15760	14	T3b	7		whole genome, whole transcriptome
PC_17163	19,8	T2b	7		whole genome, whole transcriptome
PC_17447	13,2	T3b	9		whole genome, whole transcriptome
PC_18307	5,1	T2b	7		whole genome, whole transcriptome, targeted RNA
PC_19403	7	T3b	7		whole genome, whole transcriptome, targeted RNA
PC_20873	8,3	T3a	7		whole genome, whole transcriptome
PC_22392	48,1	T3a	7		whole genome, whole transcriptome
PC_22603	7,54	T3a	<7		whole genome, whole transcriptome
PC_24173	8,5	T2b	<7		whole genome, whole transcriptome
PC_4538	8.9	T2b	7		whole genome, whole transcriptome
PC_470	17,6	T3b	8		whole genome, whole transcriptome
PC_4786	10,2	T2b	7		whole genome, whole transcriptome
PC_4906	4	T2b	<7		whole genome, whole transcriptome
PC_4980	3,5	T2b	<7		whole genome, whole transcriptome
PC_5934	9,6	T2b	<7		whole genome, whole transcriptome
PC_6102	4.4	T2b	7		whole genome, whole transcriptome
PC_6174	6	T3b	8		whole genome, whole transcriptome
PC_6342	13,4	T2b	7		whole genome, whole transcriptome, targeted RNA
PC_6488	4,2	T2b	7		whole genome, whole transcriptome
PC_6864	4,8	T3a	10		whole genome, whole transcriptome
PC_7875	14,7	T3b	8		whole genome, whole transcriptome
PC_8131	6.7	T2b	<7		whole genome, whole transcriptome
PC_8438	10,1	T3b	8		whole genome, whole transcriptome, targeted RNA
PC_9324	10,3	T3a	<7		whole genome, whole transcriptome
CRPC_261				ADT	whole genome, whole transcriptome
CRPC_278				ADT	whole genome, whole transcriptome
CRPC_305				ADT	whole genome, whole transcriptome
CRPC_348				ADT	whole genome, whole transcriptome
CRPC_435				ADT	whole genome, whole transcriptome
CRPC_489				ADT	whole genome, whole transcriptome
CRPC_530				ADT	whole genome, whole transcriptome
CRPC_531				ADT	whole genome, whole transcriptome
CRPC_539				ADT	whole genome, whole transcriptome
CRPC_541				ADT	whole genome, whole transcriptome
CRPC_542				ADT	whole genome, whole transcriptome
CRPC_543				ADT	whole genome, whole transcriptome
CRPC_697				ADT	whole genome, whole transcriptome

Sample set 2_prostatectomies

Sample ID	PSA at diagnosis	pT stage	Prostatectomy Gleason score	NGS method
5326	5,7	T2b	7	targeted RNA
12517	5,9	T2b	7	targeted RNA, whole genome, whole transcriptome
13569	5,6	T2	<7	targeted RNA
16338	10,3	T2b	7	targeted RNA
17810	4,1	T3b	9	targeted RNA
18307	5,1	T2b	7	targeted RNA, whole genome, whole transcriptome
19403	7,0	T3b	7	targeted RNA, whole genome, whole transcriptome
20581	6,6	T2b	7	targeted RNA
6342	13,4	T2b	7	targeted RNA, whole genome, whole transcriptome
8438	10,1	T3b	8	targeted RNA, whole genome, whole transcriptome
14451	6,0	T2	7	targeted RNA
16932	5,8	T3a	9	targeted RNA
5601	12,3	T3a	7	targeted RNA
11423	12,2	T3a	9	targeted RNA
12189	21,0	T3a	7	targeted RNA
22709	5,0	T2	<7	targeted RNA
23284	9,3	T2	<7	targeted RNA
1669	8,4	T2b	7	targeted RNA
13943	4,9	T2b	7	targeted RNA, whole genome, whole transcriptome
8449	4,2	T2b	<7	targeted DNA, targeted RNA
13057	4,5	T2b	7	targeted DNA, targeted RNA
13443	6,0	T2b	<7	targeted DNA, targeted RNA
4581	6,6	T2c	<7	targeted DNA, targeted RNA
17800	9,9	T3b	7	targeted DNA, targeted RNA

Sample set 2_lymph node metastases

Patient ID	Sample ID	Sample type	Age	PSA prior to pelvic node dissection	Primary tumor Gleason score	Right pelvic lymph nodes (positive lymph nodes for carcinoma/all lymph nodes excised)	Left pelvic lymph nodes (positive lymph nodes for carcinoma/all lymph nodes excised)	NGS method
PB50	21688	lymph node metastasis	69	21,0	7	1/10	4/13	targeted DNA, targeted RNA
PB133	21691	lymph node metastasis	70	69,7	9	2/6	2/3	targeted DNA, targeted RNA
PB142	21747	lymph node metastasis	61	26,2	9	3/8	2/7	targeted DNA, targeted RNA
PB170	21080	lymph node metastasis	57	3,3	<7	0/3	1/11	targeted DNA
PB259	21752	lymph node metastasis	69	32,3	7	1/3	1/3	targeted DNA, targeted RNA
PB273	21694	lymph node metastasis	53	29,0	7	1/3	none removed	targeted DNA, targeted RNA
PB275	21087	lymph node metastasis	66	123,0	<7	3/6	1/5	targeted DNA
PB283	21089	lymph node metastasis	68	unknown	unknown	0/8	3/20	targeted DNA

Sample set 2_mCRPCs

Patient ID	Sample ID	Nucleic acid	Sample type	Organ site	NGS method
A1	21040	DNA	metastasis	subdural	targeted DNA
A1	21039	DNA	normal	spleen	targeted DNA
A1	21626	RNA	metastasis	subdural	targeted RNA
A1	21830	RNA	normal	adrenal gland	targeted RNA
A2	21042	DNA	metastasis	liver	targeted DNA
A2	21041	DNA	normal	kidney	targeted DNA
A2	21673	RNA	metastasis	liver	targeted RNA
A2	21838	RNA	normal	testicle	targeted RNA
A3	21044	DNA	metastasis	pelvic paraaortic lymph node	targeted DNA
A3	21043	DNA	normal	kidney	targeted DNA
A3	21743	RNA	metastasis	pelvic paraaortic lymph node	targeted RNA
A3	21834	RNA	normal	liver	targeted RNA
A4	21046	DNA	metastasis	liver	targeted DNA
A4	21045	DNA	normal	EBV (Epstein-Barr virus) transformed lymphocytes	targeted DNA
A4	21632	RNA	metastasis	liver	targeted RNA
A4	21810	RNA	normal	prostate	targeted RNA
A5	21048	DNA	metastasis	left iliac lymph node	targeted DNA
A5	21047	DNA	normal	spleen	targeted DNA
A5	21675	RNA	metastasis	left iliac lymph node	targeted RNA
A5	21836	RNA	normal	liver	targeted RNA
A7	20681	DNA	metastasis	right subdural	targeted DNA
A7	21254	DNA	normal	liver	targeted DNA
A7	21677	RNA	metastasis	right subdural	targeted RNA
A7	21808	RNA	normal	liver	targeted RNA
A8	21050	DNA	metastasis	liver	targeted DNA
A8	21049	DNA	normal	kidney	targeted DNA
A8	21636	RNA	metastasis	liver	targeted RNA
A8	21818	RNA	normal	testicle	targeted RNA
A9	20685	DNA	metastasis	periportal lymph node	targeted DNA
A9	20686	DNA	normal	liver	targeted DNA
A9	21679	RNA	metastasis	periportal lymph node	targeted RNA
A9	21721	RNA	normal	liver	targeted RNA
A10	21804	RNA	metastasis	periportal lymph node	targeted RNA

A10	21820	RNA	normal	liver	targeted RNA
A11	21052	DNA	metastasis	linguinal lymph node	targeted DNA
A11	21051	DNA	normal	liver	targeted DNA
A11	21639	RNA	metastasis	linguinal lymph node	targeted RNA
A11	21726	RNA	normal	liver	targeted RNA
A12	21745	RNA	metastasis	mediastinal lymph node	targeted RNA
A12	21828	RNA	normal	kidney	targeted RNA
A13	21054	DNA	metastasis	L3 vertebral bone	targeted DNA
A13	21053	DNA	normal	liver	targeted DNA
A13	21681	RNA	metastasis	L3 vertebral bone	targeted RNA
A13	21731	RNA	normal	liver	targeted RNA
A14	20389	DNA	metastasis	liver	targeted DNA
A14	21055	DNA	normal	liver	targeted DNA
A14	21734	RNA	metastasis	liver	targeted RNA
A14	21642	RNA	normal	liver	targeted RNA
A16	21057	DNA	metastasis	left pulmonary hilar lymph node	targeted DNA
A16	21056	DNA	normal	liver	targeted DNA
A16	21504	RNA	metastasis	left pulmonary hilar lymph node	targeted RNA
A16	21824	RNA	normal	liver	targeted RNA
A17	21736	RNA	metastasis	cranial subdural cavity	targeted RNA
A17	21741	RNA	normal	kidney	targeted RNA
A18	21059	DNA	metastasis	left cervical lymph node	targeted DNA
A18	21058	DNA	normal	kidney	targeted DNA
A18	21644	RNA	metastasis	left cervical lymph node	targeted RNA
A19	21061	DNA	metastasis	left pelvic lymph node	targeted DNA
A19	21060	DNA	normal	liver	targeted DNA
A19	21506	RNA	metastasis	left pelvic lymph node	targeted RNA
A19	21806	RNA	normal	liver	targeted RNA
A21	20401	DNA	metastasis	liver	targeted DNA
A21	20186	RNA	metastasis	liver	targeted RNA
A21	21826	RNA	normal	kidney	targeted RNA
A22	21062	DNA	metastasis	right pelvic lymph node	targeted DNA
A22	21683	RNA	metastasis	right pelvic lymph node	targeted RNA
A23	20398	DNA	metastasis	liver	targeted DNA
A23	21063	DNA	normal	spleen	targeted DNA
A23	21514	RNA	metastasis	liver	targeted RNA
A23	21719	RNA	normal	spleen	targeted RNA
A24	21064	DNA	metastasis	right rib 7	targeted DNA
A24	21496	RNA	metastasis	right rib 7	targeted RNA
A24	21500	RNA	normal	spleen	targeted RNA
A26	21066	DNA	metastasis	L4 vertebral bone with hemorrhagic features	targeted DNA
A26	21065	DNA	normal	spleen	targeted DNA
A26	21508	RNA	metastasis	L4 vertebral bone with hemorrhagic features	targeted RNA
A26	21778	RNA	normal	spleen	targeted RNA
A27	21068	DNA	metastasis	right axillary lymph node	targeted DNA
A27	21067	DNA	normal	spleen	targeted DNA
A27	21510	RNA	metastasis	right axillary lymph node	targeted RNA
A27	21780	RNA	normal	spleen	targeted RNA
A28	21070	DNA	metastasis	left superficial inguinal lymph node	targeted DNA
A28	21069	DNA	normal	spleen	targeted DNA
A28	21512	RNA	metastasis	left superficial inguinal lymph node	targeted RNA
A28	21783	RNA	normal	spleen	targeted RNA
A29	21769	RNA	metastasis	right superficial inguinal lymph node	targeted RNA
A29	21789	RNA	normal	spleen	targeted RNA
A30	20399	DNA	metastasis	liver	targeted DNA
A30	21071	DNA	normal	spleen	targeted DNA
A30	21647	RNA	metastasis	liver	targeted RNA
A30	21822	RNA	normal	spleen	targeted RNA
A31	21773	RNA	metastasis	left adrenal	targeted RNA
A31	21791	RNA	normal	liver	targeted RNA
A32	21775	RNA	metastasis	right humerus	targeted RNA
A32	21794	RNA	normal	spleen	targeted RNA
A33	21073	DNA	metastasis	left axillary lymph node	targeted DNA
A33	21072	DNA	normal	liver	targeted DNA
A33	21685	RNA	metastasis	left axillary lymph node	targeted RNA
A33	21814	RNA	normal	liver	targeted RNA
A34	21802	RNA	metastasis	liver	targeted RNA
A34	21799	RNA	normal	spleen	targeted RNA
A35	21650	RNA	normal	iver	targeted RNA

Supplementary Table S2. Unique splice junctions for *AR-V* detection.

AR-V	**Detected splice junction**
AR-V3	exon 2 – CE4
AR-V4	exon 3 – CE4
AR-V5	exon 3 – CE2-1*
AR-V6	exon 3 – CE2-2*
AR-V7	exon 3 – CE3
AR-V9	exon 3 – CE5
AR-V12	exon 4 – exon 8

*Alternative 5' splicing site

Supplementary Table S3. Statistical comparison of individual AR variants in different sample types from sample set 2 using two-tailed, unpaired Mann-Whitney U test.

AR-V3

	Prostatectomy	Lymph node metastases
Prostatectomy		
Lymph node metastases	0,2947	
CRPC metastases	**0.0019****	**0.0302***

AR-V7

	Prostatectomy	Lymph node metastases
Prostatectomy		
Lymph node metastases	0,7458	
CRPC metastases	**0.0054****	0,2372

AR-V9

	Prostatectomy	Lymph node metastases
Prostatectomy		
Lymph node metastases	0,3532	
CRPC metastases	**0.0007*****	**0.0201***

Supplementary Table S4. AR, FOXA1 and SPOP mutations detected in sample set 2.

PATIENT ID	MUTATION NAME	CHROM	POSITION	REF	ALT	FUNCTION	EXONIC_FUNCTION	AA_CHANGE	READS	VAF
A2	AR: L702H	chrX	66931463	T	A	exonic	nonsynonymous SNV	AR:NM_000044:exon4:c.	74:169	0,44
A21	AR: L702H	chrX	66931463	T	A	exonic	nonsynonymous SNV	AR:NM_000044:exon4:c.	57:975	0,06
A33	AR: L702H	chrX	66931463	T	A	exonic	nonsynonymous SNV	AR:NM_000044:exon4:c.	47:180	0,26
A22	AR: T878A	chrX	66943552	A	G	exonic	nonsynonymous SNV	AR:NM_000044:exon8:c.	27:146	0,18
A24	AR: T878A	chrX	66943552	A	G	exonic	nonsynonymous SNV	AR:NM_000044:exon8:c.	564:578	0,98
PB50	FOXA1: G39R	chr14	38061874	C	G	exonic	nonsynonymous SNV	FOXA1:NM_004496:exor	44:270	0,16
PB283	FOXA1: S250F	chr14	38061240	G	A	exonic	nonsynonymous SNV	FOXA1:NM_004496:exor	83:265	0,31
A16	FOXA1: R265H	chr14	38061195	C	T	exonic	nonsynonymous SNV	FOXA1:NM_004496:exor	80:139	0,58
PB273	FOXA1: R447fs	chr14	38060648	C	-	exonic	frameshift deletion	FOXA1:NM_004496:exor	97:250	0,39
A21	SPOP: F102C	chr17	47696643	A	C	exonic	nonsynonymous SNV	SPOP:NM_001007228:ex	108:252	0,43
PB142	SPOP: F102C	chr17	47696643	A	C	exonic	nonsynonymous SNV	SPOP:NM_001007228:ex	162:358	0,45
A33	SPOP: F133V	chr17	47696426	A	C	exonic	nonsynonymous SNV	SPOP:NM_001007228:ex	74:210	0,35

Supplementary Table 5. *AR* genomic structural rearrangements detected in metastatic CRPC specimens.

Patient ID	***AR* copy number**	**Tumour type**	**Tissue**	***AR*-GSR**	**Break fusion junction coordinates (hg19)**	**Remarks**	**# supporting split reads**	**Variant allele fraction**
A4	68	CRPC metastasis	Liver	Inversion	chrX:66,189,781 / chrX:66,863,033	5' breakpoint located in a *LINE-1* element between *EDA2R* and *AR*, region containing exon 1, 1b and most of intron 1 inverted	22	8.6%
A7	6	CRPC metastasis	Subdural	Inversion	chrX:66,783,645 / chrX:66,784,142	Inverted ~500bp region in *AR* intron 1 between exon 1 and 1b	53	5.4%
A8	59	CRPC metastasis	Liver	Translocation	chr7:35,711,637 / chrX:66,911,250	5' breakpoint located in a *LINE-1* element in *HERPUD2* intron 3, 3' breakpoint in *AR* intron 3 -> deletion of cryptic exons CE3/CE5 + exons 4-8	18	2.6%
A11	7	CRPC metastasis	Inguinal lymph node	Duplication	chrX:116,819,638 / chrX:66,769,235	*AR* intron 1 breakpoint between exons 1 and 1b (exon 1 deleted)	32	6.2%
A27	9	CRPC metastasis	Axillary lymph node	Deletion	chrX:66,931,389 / chrX:66,942,180	Deletes half of exon 4 + exons 5 and 6; detected also from RNA-seq sample	37	10.9%