Supplementary Information for Kohli *et al*.

CONTENTS **PAGE**

Supplementary Methods

PROMOTE Trial

The "PROMOTE" (Prostate Cancer Medically-Optimized Genome-Enhanced Therapy) study, clinicaltrials.gov identifier NCT01953640, was initiated after obtaining approval from Mayo Clinic Institutional Review Board (IRB). Accrual was performed at three sites (Mayo Clinic Rochester, Mayo Clinic Florida and Mayo Clinic Arizona) between Mayo 2013 and September 2015. At that time abiraterone acetate and prednisone was the only FDA approved drug combination for treating progressive prostate cancer following failure of ADT and prior to initiating chemotherapy. The target population included patients with progressive disease while on ADT as defined by progressive measureable disease or appearance of new metastatic lesions on imaging or two serial serum prostate specific antigen (PSA) rises measured at least 1 week apart after reaching a nadir during ADT. All patients had to have sub-castrate testosterone levels (less than 50 ng/dl) and metastatic sites amenable for biopsy. A list of the complete eligibility criteria and study protocol is available in the supplementary attachment "Promote Study Protocol". The initial enrollment target of 200 metastatic CRPC patients in this biopsy trial was amended with Mayo Clinic IRB approval to 110 patients in 2015. All patients enrolled on the trial provided a written informed consent approved by the IRB and as per study procedures underwent metastatic tissue biopsy prior to initiation of abiraterone acetate-prednisone. A second metastatic site biopsy was obtained after 12-weeks of treatment with AA/P. The primary aims of the study were to determine associations between somatic tumor genome and transcriptome alterations identified pre-therapy with progression at 12-weeks. Progression was assessed at 12-weeks using a composite progression endpoint as per the recommendations of the Prostate Cancer Working Group-2 criteria (PCWG2)[1]. Briefly, serum PSA, bone and CT imaging and symptom assessments using the Functional Assessment of Cancer Therapy-Prostate (FACT-P) scale were performed at the same time (week 12)[1]. Response to therapy was defined as the absence of PSA progression (as defined by the PSA Working Group Criteria)[1]; absence of any new bone lesion on bone scan and no radiological (RECIST 1.0) progression of nodal or soft tissue metastases. Bone scans at 12 weeks in which new lesions were detected were repeated with a follow up bone scan at least six more weeks and if additional new bone lesions were observed at the second follow up scan patient was deemed to have progression. Patients who met one of these criteria for progressive disease were defined as "non-responders" after 12 weeks of drug exposure.

Collection and Analysis of Circulating Tumor Cells

Details of VERSA manufacture and operation have been described[2]. Briefly, blood specimens were collected in vacutainer tubes (BD Biosciences) with EDTA anticoagulant. Mononuclear cells were isolated with a ficoll gradient. Samples for gene expression analysis were CD45 depleted to improve purity of live cell capture of CTCs. CTCs were isolated with VERSA using an antibody to EpCAM conjugated to paramagnetic particles. mRNA isolation is integrated on the VERSA. The mRNA elution sample containing PMPs was reverse transcribed using a High Capacity cDNA Reverse Transcriptase kit (Life Tech, USA). The RT reaction (12.5 µL) was then amplified for 14 cycles using TaqMan PreAmp (Life Tech, USA) according to manufacturer's directions and diluted 1:20 in 1x TE (10 mM Tris-HCL pH8, 1 mM EDTA). For TaqMan assays, 5 µL of diluted cDNA template was mixed with 10 µL iTaq® master mix (Bio-Rad, USA), 1 µL TaqMan Gene Expression Assay (Supplementary Table 3, Life Technologies, USA) and 4 µL nuclease free water. Each reaction was amplified for 45 cycles (denatured at 95 °C for 15 seconds followed by annealing at 60°C for 1 minute) using a CFX Connect® Real-Time PCR

System (Biorad, USA). Threshold cycle (Ct) values were reported as discrete colors on a heatmap gradient.

Cell Line Authentication and Mycoplasma Detection

Authentication of the CWR-R1 cell line was performed by sequence-based validation of two signature AR gene alterations: AR H874Y point mutation and 50kb intragenic deletion within AR intron 1[3]. Sequence-based authentication of CWR-R1 was performed every 5-10 passages. Aliquots of cell culture supernatants from cells in active culture were evaluated regularly for mycoplasma contamination using a PCR-based method as described[4]. All experiments with 22Rv1, CWR-R1, LNCaP, VCaP, and DU145 cells were performed within 2-3 months of resuscitation of frozen cell stocks prepared within 3 passages of receipt.

Bioinformatics Analysis of RNA-seq Data

Annotated AR and AR-V isoforms including AR-FL, AR45, AR23, ARV3, ARV7 and ARV9 have been defined previously[5]. For analysis of AR RNA-seq data, Illumina reads were aligned to hg19 using TopHat (v.2.0.11)[6] with --no-coverage-search and --micro-exon-search enabled, and the mean and standard deviation for the inner distance between mate pairs set empirically. Tumor RNA purity for 65 PROMOTE biopsies was estimated where both RNA-seq and DNA-seq data were available. This was calculated by multiplying the CNV purity by the estimated weighted mean allele frequency of RNA over the estimated weighted mean allele frequency of DNA. The mutant allele frequency for each variant was estimated using the R binom package to take into account that the number of mutant reads can often be 0 when the purity is low and the number of reads is equally low. The weighting gave more weight to variants with more reads. Annotated AR and AR-V isoforms from the UCSC annotation database were used as a guide for TopHat alignment. The TopHat junctions.bed output file was parsed to determine how many reads supported novel and canonical splice junctions within the AR locus. PROMOTE RNA-seq data was analyzed using MAP-RSeq pipeline (v1.2.1)[7]. Briefly, raw RNA-seq reads were aligned to the human reference genome (hg19/GRCh37) using Tophat (v2.0.6) and bowtie1 aligner options[6]. Pre- and post-alignment quality controls, gene level raw read count and normalized read count (i.e. FPKM), and identification of splice junctions were performed using RSeQC package (v2.3.6) with NCBI human RefSeq gene model (release August, 2014) as the reference gene model[8]. Expression of AR and AR-Vs were measured by the number of spliced reads mapped to AR-V-specific splice junctions[9]. The number of splice reads were then normalized by total splice reads to correct for sequencing depth (i.e. SRPM, splice read per million).

Bioinformatics Analysis of PacBio SMRT Iso-Seq Data

PacBio ToFU software (pbtranscript.py classify) version 2.2.3 (Pacific Biosciences, https://github.com/PacificBiosciences/cDNA_primer) was used to remove primers and classify the reads into full-length and non-full-length transcripts. Full-length reads were mapped using gmap, sorted, and collapsed into transcripts using ToFU's collapse_isoforms_by_sam.py script, using default parameters, corresponding to a minimum threshold for coverage of 0.99 and a minimum threshold for alignment accuracy of 0.85. The resulting data were then compared to known exons of AR, and plotted using ggplot2 in R to show the exon composition of each isoform detected, and the number of reads supporting each isoform.

Plasmids and Lentivirus

A plasmid encoding AR-V9 was generated using a previously-described strategy[10]. Briefly, synthetic cassettes encoding the AR-V9 COOH-terminal extension (AR-V9-FW: 5'-CTAGGAGACAACTTACCTGAGCAAGCTGCTTTTTGGAGACATTTGCACATCTTTTGGGATC ACGTTGTTAAGAAGTAGT and AR-V9-RV: 5'-CTAGACTACTTCTTAACAACGTGATCCCAAAAGATGTGCAAATGTCTCCAAAAAGCAGCTTG CTCAGGTAAGTTGTCTC) were annealed, phosphorylated, and ligated into an XbaI-mutant version of the p5HBhAR-A expression plasmid [10] prepared by digestion with XbaI. All constructs were verified by DNA sequencing. To construct lentiviral expression vectors, the AR-V9 coding sequence was liberated from this p5HBhAR-A-based plasmids using PCR with primers designed to incorporate 5' EcoRV and 3' SalI sites (forward primer: 5'-TGGGATATCCAGCCAAGCTCAAGG; reverse primer: 5'- GGGAGTCGACACAGGGATGCCA), EcoRV/SalI-digested PCR products were then ligated with EcoRV/SalI-digested pLV-GFP [11] to replace the GFP insert with AR-V9 cDNA. The AR-V9 lentivirus construct was verified by DNA sequencing, restriction mapping, transient transfection, and western blotting with an antibody specific for the AR N-terminus.

Transient Transfections

22Rv1 cells were electroporated with siRNAs targeted to AR Exon CE3 and CE5. Briefly, 200 pmol of siRNA were mixed with 4 X 10^6 22Rv1 cells in a 4mm gap-width electroporation cuvette in a total volume of 400 μL, and subjected to a 350V, 10 ms pulse using a Square Wave Electroporator (BTX/Harvard Apparatus) prior to plating in RPMI medium supplemented with 10% fetal bovine serum (FBS). siRNAs were designed to target the 3'-utr encoded by exon CE3[12] (siRNA CE3) or the exon CE3 open reading frame (CE3-011 target sequence: 5'-CAAAUGACCAGACCCUGA). Three siRNAs were designed to target exon CE5 (CE5 ORF: 5'-GGGAUCACGUUGUUAAGAA; CE5 3' UTR-1: 5'-CAACCAAGGCAUAGAAAUA; CE5 3' UTR-2: 5'-CAACAGAUAUCUAGGAAGU). VCap cells were seeded at a density of 5x10^5 cells/well in 6-well plates and transfected with siRNA using Lipofectamine 2000 (Invitrogen) exactly per the manufacturer's recommendations. Cells were lysed 48hr post-transfection for isolation of total RNA and protein. LNCaP cells were transfected by electroporation exactly as described[10]. Briefly, 12.42 μg of DNA (9μg -5746 PSA-LUC, 3μg SV40-Renilla, and 0.42μg AR-V expression plasmid) were mixed with 4 X 10^6 LNCaP cells in a 4mm gap-width electroporation cuvette, and subjected to a 305V, 10 ms pulse using a Square Wave Electroporator (BTX/Harvard Apparatus) prior to plating in RPMI medium supplemented with 10% charcoal-stripped, steroid-free fetal bovine serum (CSS). DU145 cells were transfected using Superfect (Qiagen) as per the manufacturer's recommendations exactly as described [11]. Briefly, 1.2μg of 4XARE-E4-LUC, 0.4μg SV40-REN, 40 ng of AR-V expression plasmid, and 8μL Superfect reagent were suspended in 68 μL of serum-free DMEM, incubated for 20 min, diluted with 480μL of DMEM supplemented with 10% CSS, and added to 1 well of a 24-well plate that had been seeded with 1X10^5 cells the day prior. For all reporter-based experiments, 24 hours post-transfection, cells were re-fed with serum free medium containing 1nM dihydrotestosterone (Sigma) or 0.1% ethanol as vehicle control in combination with 30μM enzalutamide (Selleck Chemical) or 0.1% DMSO as vehicle control for 24 hours. Cells were harvested in 1X passive lysis buffer provided in a Dual Luciferase Assay Kit (Promega). Activities of the firefly and Renilla luciferase reporters were assayed using a Dual Luciferase Assay Kit as per the manufacturer's recommendations. Transfection efficiency was normalized by dividing firefly luciferase activity by Renilla luciferase activity. Data presented represent the mean +/- s.e.m. from three independent experiments, each performed in duplicate.

Preparation of Lentivirus

Lentivirus encoding GFP or AR-V9 was prepared using a standard third-generation packaging system in 293T cells. Briefly, 293T cells were co-transfected with lentivirus vectors and packaging vectors pCMVΔR8.91 (19) and pMD.G (20) at a ratio of 4:3:1 using Lipofectamine 2000 (ThermoFisher Scientific). Medium containing lentivirus was collected from 48hr to 96hr post-transfection, pooled, and concentrated using Lenti-X Concentrator (Clontech).

Reverse Transcription PCR (RT-PCR)

RNA was converted to cDNA using a reverse transcription kit following the manufacturer's instructions (Quanta Biosciences). cDNA was subjected to quantitative PCR using PerfeCTa SYBR Green FastMix (Quanta Biosciences) with primers specific for the AR exon 3/CE3, exon 3/CE5, and exon 3/4 splice junctions to measure mRNA levels of AR-V7, AR-V9, and full-length AR, respectively. Each PCR reaction utilized a common forward primer (qSPLICE Exon3 FW: 5'-AAC AGA AGT ACC TGT GCG CC) paired with exon-specific reverse primers (qSPLICE CE3 RV): 5'-TCA GGG TCT GGT CAT TTT GA; qSPLICE CE5 RV: 5'-GCA AAT GTC TCC AAA AAG CAG C; qSPLICE Exon4 RV: 5'- TTC AGA TTA CCA AGT TTC TTC AGC. Primers for assessing actin levels were FW: 5'-.ATG CAG AAA GAG ATC ACC GC and RV: 5'-ACA TCTGCT GGA AGG TGG AC. Fold change in mRNA expression levels were calculated by the comparative Ct method using the formula $2^{-(\Delta\Delta Ct)}$ and GAPDH as calibrator. For actinomycin D experiments, serial dilutions were performed with cDNA from untreated cells, followed by PCR to generate standard curves. These standard curves were used to extrapolate the % target remaining from Ct values of actinomycin D-treated samples, which was expressed relative to vehicle-treated cells.

Generation of AR-9 Antisera and purification of AR-V9 antibodies

A synthetic peptide containing AR-V9-derived sequence NH_2-CDNLPEQAAFWRHLHIFWDHVVKK was submitted to Cocalico Biologicals (Reamstown, PA) for KLH conjugation and inoculation. Rabbits were inoculated with 500 μg of KLH-conjugated peptide mixed with Complete Freund's Adjuvant. Rabbits received boosts on days 14, 21, and 49 with 50 μg of KLH conjugated peptide/adjuvant mixtures. Rabbits were exsanguinated on day 75 and serum was stored at -80°C. Antibodies recognizing AR-V9 were isolated by affinity purification. Briefly, the same peptides used for KLH conjugation and inoculation of rabbits were coupled to SulfoLink columns (Pierce) and antisera was bound, washed, and eluted as per the manufacturer's protocol. Eluted fractions that were positive for protein content as determined by Bradford Assay (BioRad) were pooled, representing the purified AR-V9 antibody used in this study.

Supplementary References

1 Scher, H. I. *et al.* Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. *J Clin Oncol* **26**, 1148-1159 (2008).

2 Sperger, J. M. *et al.* Integrated analysis of multiple biomarkers from circulating tumor cells enabled by exclusion-based analyte isolation. *Clin Cancer Res* (2016).

3 Li, Y. *et al.* AR intragenic deletions linked to androgen receptor splice variant expression and activity in models of prostate cancer progression. *Oncogene* **31**, 4759-4767 (2012).

4 Uphoff, C. C., Denkmann, S. A. & Drexler, H. G. Treatment of mycoplasma contamination in cell cultures with Plasmocin. *Journal of biomedicine & biotechnology* **2012**, 267678 (2012).

5 Lu, C. & Luo, J. Decoding the androgen receptor splice variants. *Translational andrology and urology* **2**, 178-186 (2013).

6 Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* **12**, R72 (2011).

7 Kalari, K. R. *et al.* MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. *BMC Bioinformatics* **15**, 224 (2014).

8 Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184-2185 (2012).

9 Kohli, M. *et al.* Mutational Landscapes of Sequential Prostate Metastases and Matched Patient Derived Xenografts during Enzalutamide Therapy. *PLoS One* **10**, e0145176 (2015).

10 Dehm, S. M., Schmidt, L. J., Heemers, H. V., Vessella, R. L. & Tindall, D. J. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res* **68**, 5469-5477 (2008).

11 Chan, S. C., Li, Y. & Dehm, S. M. Androgen receptor splice variants activate AR target genes and support aberrant prostate cancer cell growth independent of the canonical AR nuclear localization signal. *J Biol Chem* **287**, 19736-19749 (2012).

12 Hu, R. *et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res* **69**, 16-22 (2009).

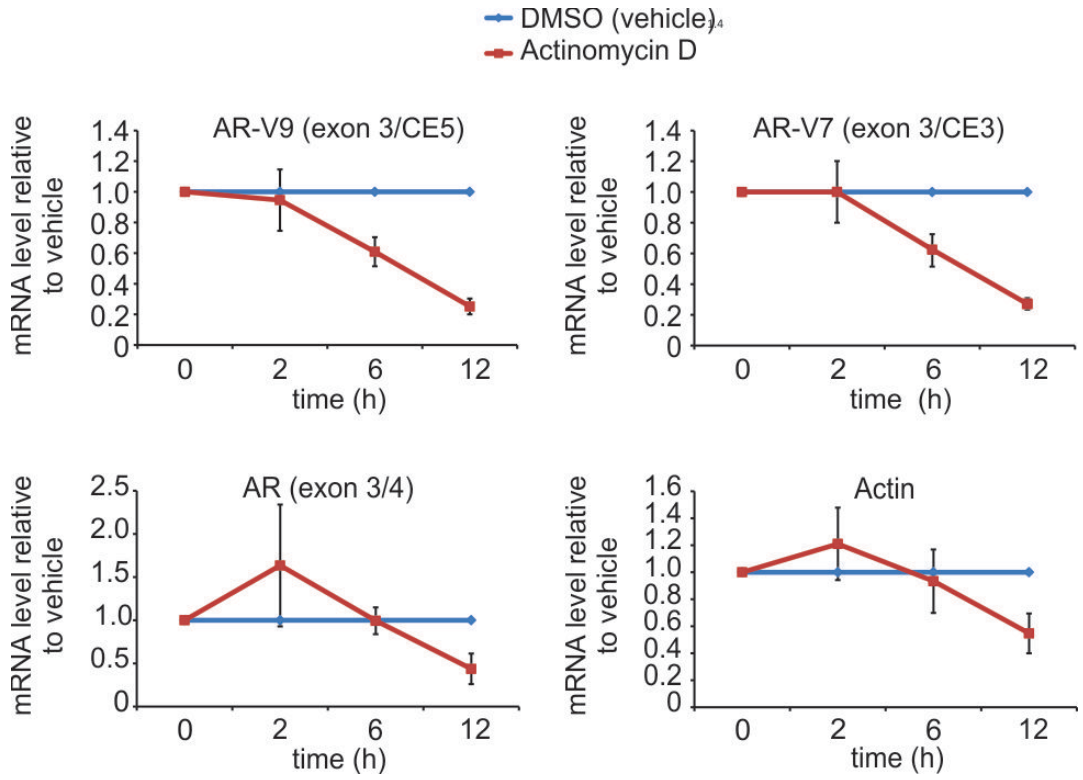
Supplementary Table 1: Baseline Characteristics of the Patients	
Characteristic	Total (N=78)
Race (N=78) – no. (%)	
White	74 (96)
Black or African American	1 (1)
Asian	1 (1)
American Indian or Alaska Native	1 (1)
Unknown	1 (1)
Age - yr	
Median	72.5
Range	39-92
Gleason score at initial diagnosis – no. (%)	
2 - 6	14 (18)
7	23 (29)
8 - 10	41 (53)
Primary radiation therapy at initial diagnosis – no. (%)	
Yes	51 (65)
No	27 (35)
Primary radical prostatectomy at initial diagnosis – no. (%)	
Yes	38 (49)
No	40 (51)
Volume of metastatic disease – no. (%)	
Low	31 (40)
High	47 (60)
Time from starting ADT to enrollment - yr	
Median	2.6
Interquartile range	1.3-4.8
Time from starting ADT to CPRC - yr	
Median	2.5
Interquartile range	1.3-4.9
Metastatic Biopsy site at study enrollment – no. (%)	
Bone	56 (72)
Lymph nodes	14 (18)
Liver / lung	3 (4)
Others	5 (6)
Metastatic Biopsy site at 12 weeks – no. (%)	
Bone	60 (81)
Lymph nodes	9 (12)
Liver / lung	3 (4)
Others	2 (3)
Unable to biopsy	4 (5)
PSA at study enrollment (N=77) – ng/ml	
Median	14.6
Interquartile range	8.0-45.6
Serum Chromogranin levels at study enrollment (N=76) – ng/ml	
Median	94.5
Interquartile range	56.5-236.5
Testosterone at study enrollment (N=76) – ng/dl	
Median	7.0
Interquartile range	6.9-10.0
Study follow-up	
Median days of follow-up (IQR)	538 (378-715)
Number progressed at 12 weeks (%)	32 (41)
Median time to progression (IQR	225 (91-534)

Supplementary Table 2
Baseline RNA-seq Expression Levels for AR/ARVs (spliced reads per million)

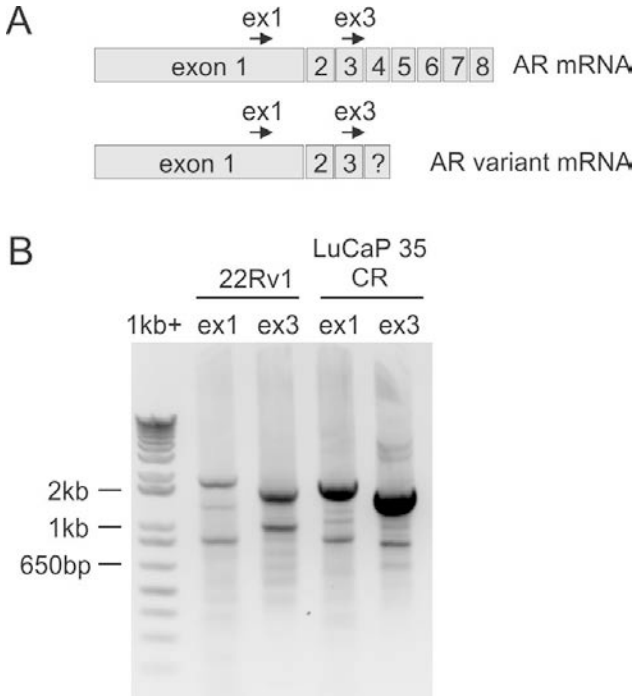
AR / AR variant (N=78)	Value
AR full length	
Mean (SD)	14.8 (28.8)
Median (IQR)	5.6 (1.5-17.7)
ARV3	
Mean (SD)	0.2 (0.4)
Median (IQR)	0.0 (0.0-0.2)
ARV7	
Mean (SD)	0.8 (1.4)
Median (IQR)	0.1 (0.0-0.9)
ARV9	
Mean (SD)	0.2 (0.6)
Median (IQR)	0.0 (0.0-0.2)
AR23	
Mean (SD)	0.1 (0.2)
Median (IQR)	0.0 (0.0-0.1)
AR45	
Mean (SD)	0.2 (0.4)
Median (IQR)	0.1 (0.0-0.3)
Chromogranin A	
Mean (SD)	9.2 (46.5)
Median (IQR)	0.1 (0.0-0.3)
ARV7 full length ratio (N=66)	
Mean (SD)	0.1 (0.1)
Median (IQR)	0.0 (0.0-0.1)
ARV9 full length ratio (N=66)	
Mean (SD)	0.0 (0.0)
Median (IQR)	0.0 (0.0-0.0)

SD = standard deviation, IQR = Interquartile range

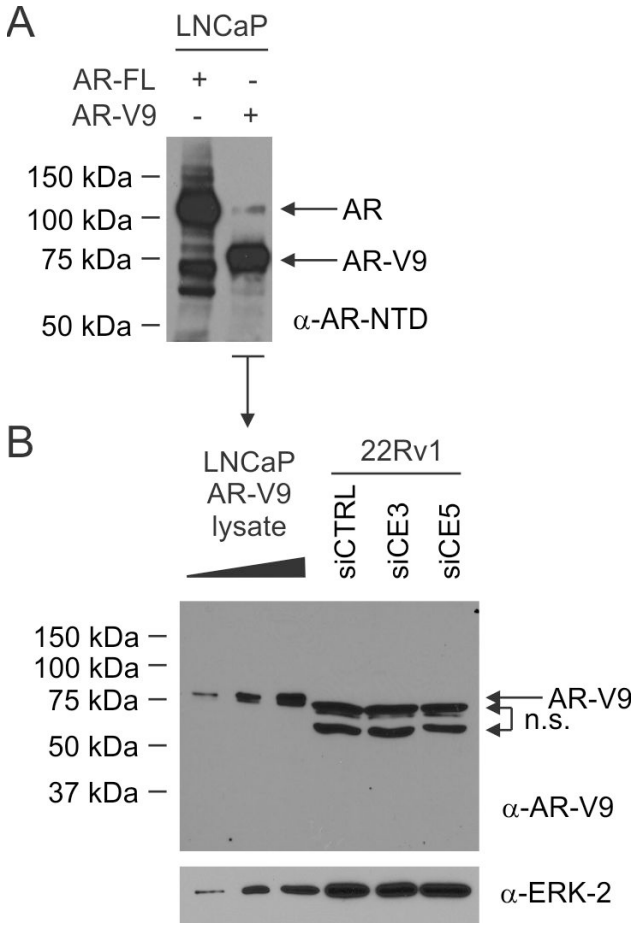
Supplementary Table 3 **TaqMan Gene Expression Assays with Circulating Tumor Cells**		
Gene name	Exon junction	Thermo Fisher Catalog number
AR_Total	1-2	Hs00907242_m1
AR_Full length	4-5	Hs00171172_m1
AR-V7	3-CE3	Hs04260217_m1
AR-V9	3-CE5	Custom (AI6RPM5)
KLK2	2-3	Hs00428383_m1
KLK3	1-2	Hs02576345_m1
TMPRSS2	13-14	Hs01120965_m1
PSMA	5-6	Hs00379515_m1
NKX3.1	2-3	Hs00171834_m1
PTPRC	2-3	Hs04189704_m1
RPLOP0		4333761F



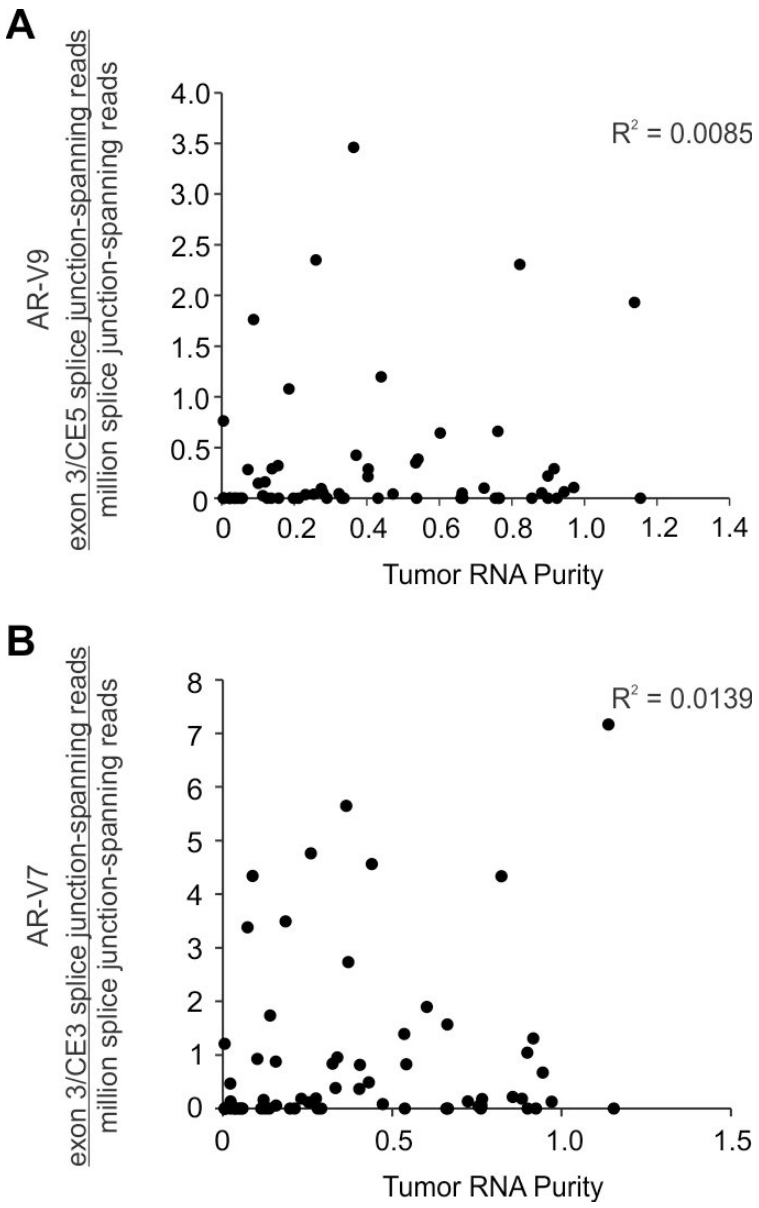
Supplementary Figure 1 | Actinomycin D responses of AR-V9 and AR-V7 mRNA in 22Rv1 cells. Cells were treated with 10μM actinomycin D or vehicle (DMSO) for indicated times. Total RNA was analyzed by quantitative RT-PCR using primers specific for AR-V9 (exon 3/CE5 splice junction), AR-V7 (exon 3/CE3 splice junction), AR (exon 3/4 splice junction) or actin (control transcript, long half-life) as indicated. Data represent mean +/- SD from 3 biological replicates (n=3).



Supplementary Figure 2 | 3' rapid amplification of cDNA ends (RACE) for single molecule real-time (SMRT) sequencing. A, Schematic of forward primer design strategies for 3' RACE reactions. **B,** Agarose gel electrophoresis of 3' RACE products using RNA from 22Rv1 cells and LuCaP 35-CR patient derived xenograft tissue.



Supplementary Figure 3 | Purified AR-V9 polyclonal antibodies detect overexpressed AR-V9 in western blot with overexpressed AR-V9. A, Western blot of LNCaP cells overexpressing full-length AR (AR-FL) or AR-V9. The blot was probed with a polyclonal antibody specific to the AR NTD. **B,** Western blot of increasing amounts of lysate from LNCaP cells overexpressing AR-V9 and lysates from 22Rv1 cells transfected with a control siRNA (siCTRL) and siRNAs targeted to exon CE3 (siCE3) or CE5 (siCE5). Blots were probed with a polyclonal antibody specific for the AR-V9 COOH-terminal extension or ERK-2 as loading control.



Supplementary Figure 4 | Relationship between tumor RNA purity, AR-V7, and AR-V9 expression levels in CRPC metastases. Plots of **A,** AR-V9 and **B,** AR-V7 as a function of tumor RNA purity for 65 PROMOTE biopsies where both RNA-seq and DNA-seq data were available. Tumor RNA purity was calculated by multiplying the copy number variant (CNV) purity by the estimated weighted mean allele frequency of RNA over the estimated weighted mean allele frequency of DNA.