PNAS
www.pnas.org

Supplementary Information for

Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13

Zhong Chen, Dayong Wu, Jennifer M. Thomas-Ahner, Changxue Lu, Pei Zhao, Qingfu Zhang, Connor Geraghty, Pearlly S. Yan, William Hankey, Benjamin Sunkel, Xiaolong Cheng, Emmanuel S. Antonarakis, Qi-En Wang, Zhihua Liu, Tim H.-M. Huang, Victor X. Jin, Steven K. Clinton, Jun Luo, Jiaoti Huang, and Qianben Wang

To whom correspondence may be addressed: Email: jluo1@jhmi.edu, jiaoti.huang@duke.edu, or qianben.wang@duke.edu.

This PDF file includes:

SI Materials and Methods
Figs. S1 to S10
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3

SI Materials and Methods

Cell lines. CWR22RV1 and LNCaP cells were obtained from the American Type Culture Collection (ATCC) and have been authenticated prior to commencing this study by short tandem repeat (STR) profiling and karyotyping. CWR22RV1 and LNCaP cells were cultured in RPMI 1640 medium with 10% fetal bovine serum (FBS). LNCaP95 cells were provided by Dr. Jun Luo (Johns Hopkins University) and were authenticated as described previously (1, 2). LNCaP-abl cells were kindly provided by Zoran Culig (Innsbruck Medical University, Innsbruck, Austria) and authenticated as described previously (3). LNCaP95 and LNCaP-abl cells were cultured in phenol red-free RPMI 1640 with 10% charcoal-stripped FBS. For all experiments, cells were maintained in phenol red-free RPMI 1640 medium with 10% charcoal-stripped FBS for 3 days to mimic human CRPC. The cell lines were passaged in our laboratory for less than 6 months after resuscitation. The cell lines were routinely tested to ensure they are free of mycoplasma contamination (VenorTMGeM Mycoplasma Detection Kit, Sigma-Aldrich).

Crosslinked IP-Mass Spectrometry. Cells were cultured in phenol red-free RPMI 1640 medium with 10% charcoal-stripped FBS for 3 days. After fixation with 1% formaldehyde for 10 min at room temperature, cells were washed and collected in ice-cold PBS. The nuclear fraction was extracted using NE-PER Reagents (Thermo Fisher) and resuspended in ChIP lysis buffer at 4 °C for 10 min. After sonication, the supernatant was incubated overnight with 4 µg of antibody against AR-V7 (AG10008, precision), full-length AR (AR-FL, C19, Santa Cruz Biotechnology) or IgG. Pre-treated Protein A-Agarose beads were added and incubated for 1 h followed by washing six times in modified RIPA buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.25% Na

deoxycholate, 1% NP-40, 0.5 M LiCl) and twice in 50 mM ammonium hydrogen carbonate solution. To isolate the purified cross-linked complex, ~10 μg protein was separated by electrophoresis in a 4-12% SDS PAGE gel. The gel region above ~70 kDa was collected and digested in-gel by trypsin at 37 °C overnight. Lyophilized peptides were resuspended in 2 μL of 0.1% formic acid and analyzed by nanoLC-MS/MS with an LTQ Orbitrap mass spectrometer (Thermo Fisher) at The Ohio State University Mass Spectrometry and Proteomics Facility. The raw MS files were searched and analyzed against the human protein database (UniProtKB/Swiss-Prot, 2017) using Peaks Studio 8.0 (Bioinformatics Solutions, Canada). The parameters were set as follows: the maximum missed cleavages were set to 2; the precursor ion mass tolerance was set to 10 ppm, and MS/MS tolerance was 0.6 Da. Only unique, high confidence peptides identified (FDR 1%) were chosen for downstream protein identification analysis.

RNA interference. ON-TARGETplus siRNAs targeting AR-V7, HoxB13 or control siRNAs were purchased from Dharmacon. For shRNA knockdown, recombinant lentiviral particles were produced by transient transfection of 293T cells. In brief, 10 μg of the shRNA plasmids, 3.5 μg of the pVSV-G plasmid, 2.5 μg of the pRev plasmid and 6.5 μg of pGag-pol plasmids were co-transfected using Lipofectamine 2000 into 293T cells. Viral supernatant was collected 48 h after transfection and filtered. Cells were infected for 48 h with viral supernatant containing 10 μg/ml polybrene, and then were examined by immunoblot or used for xenograft assay without antibiotic selection.

RNA-seq and Data Analysis. RNA-seq analysis was performed as previously described (4). For cell line RNA-seq, 72 h after siRNA transfection, total RNA was isolated using an RNeasy Mini

kit (QIAGEN). Next, mRNA was enriched using NEBNext Poly (A) mRNA Magnetic Isolation Module. After two rounds of this process, the mRNA was recovered for library generation, which was performed with NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina following the manufacturer's instructions. The cDNA molecules were amplified by 8 cycles of PCR. Non-size-selected libraries were then sequenced using an Illumina HiSeq 2500 instrument at the Ohio State University Comprehensive Cancer Center (OSUCCC) sequencing core. For CRPC tissue RNA-seq, total RNA was extracted using the RNeasy Plus Universal Mini Kit (Qiagen). Pure RNA with RIN > 7.0 was used in this study. Fifty base pairs of single-end or pair-end reads were generated on the Illumina HiSeq 2500 platform at OSUCCC sequencing core with three multiplexed samples per lane. Read alignment was conducted using TopHat 2.0.13, and relative transcript abundances and differentially expressed genes were determined using Partek Genomics Suite (v6.6) with default settings. The reads count in exon regions was used to estimate the gene expression level. All biological duplicates (for cell line RNA-seq) and technical duplicates (for CRPC tissue RNA-seq) show high correlation around 0.99 (Spearman). Hierarchical clustering was performed using genes with a false discovery rate (FDR) of 0.05 and a fold change of 2 or 1.5. For Gene Ontology analysis of AR-V7- or HoxB13-regulated genes, the top Biological Processes (Gene Ontology) or Diseases (MeSH) were selected based on the statistical significance for each category using the Genomatix Pathway System (v3.3). To analyze CRPC patient cohorts, two datasets were downloaded from the database of Genotypes and Phenotypes (dbGaP) repository (dbGaP Study Accession: phs000915.v1.p1 and phs000909.v.p1). The raw data were aligned against hg19 and analyzed by Partek Genomics Suite (v6.6) with default settings. As for AR-V7 status, patients with Exon3-CE3 junction (> 2 counts) and gapped PE-reads (> 5 counts) were manually curated as AR-V7 positive patients.

For Gene Set Enrichment Analysis (GSEA), all genes analyzed by RNA-seq were ranked and weighted by their mean log2 fold change upon siRNA treatment. Lists of genes within ± 50 kb or 20 kb of AR-V7 or HoxB13 binding sites were derived, and then analyzed using the GSEA PreRanked tool (v2.2.4) with default parameters.

ChIP-exo and Data Analysis. ChIP-exo was performed and analyzed as previously described (4, 5). Briefly, for cell line ChIP-exo, after fixation with 1% formaldehyde for 10 min at room temperature, chromatin was sonicated and incubated overnight with 2-4 μg antibodies against AR-V7 (AG10008, Precision Antibody), full-length AR (AR-FL, C19, Santa Cruz Biotechnology) or HoxB13 (H-80, Santa Cruz Biotechnology). For tissue ChIP-exo, the frozen CRPC samples were trimmed and chopped into small pieces on ice, and fixed immediately with 1% formaldehyde for 20 min at room temperature. T4 DNA polymerase, T4 PNK and Klenow DNA Polymerase were used together for end polishing. The ligation step was performed with 1mM dithiothreitol. Protein A Dynal magnetic beads were washed using modified RIPA buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.25% sodium deoxycholate, 1% NP-40, 0.5 M LiCl) followed by Tris pH 8.0 twice during each step. The library was amplified with only 10-12 cycles and prepared without gel-based size selection. Paired-end sequencing (50 bp) was performed using Illumina HiSeq2500 at the OSUCCC sequencing core. Raw reads were aligned to the human reference genome (hg19) using Bowtie with default parameter settings. The binding locations for AR-V7, AR-FL and HoxB13 were identified by BELT(6) and Genomics Suite (*v*6.6, Partek) from ChIP-exo Reads 2. ChIP-exo borders were called by the MALD model from ChIP-exo Reads 1, and DNA motifs were precisely defined by the BPMotif approach as we described previously (4, 5). Briefly, the enriched DNA motifs were defined by a multi-phase

cross-validation procedure. Genomics Suite (*v*6.6) and MEME Suite *v*4.9 (7) were used to find the candidate motifs. Initial motif candidates were generated using default program settings (one instance per sequence, less than 40 bp of border extension). Motifs were then clustered with the Pearson correlation coefficient. Exo signal was measured to define border patterns and classify motifs. A set of overrepresented motifs was then used to correct border extension according to the enriched motif position. Motif discovery was repeated twice. Motifs with E <1e-10 or that were found in 10% of sequences were retained as reliable predictions for the next round of analysis. Finally, we identified motifs satisfying the following extensible criteria: 1) Motif similarity compared to ARE or Homeobox motifs in the transcription factor binding databases or between core motifs defined in AR-V7, AR-FL or HoxB13 ChIP-exo data; 2) At least one common protected border exists upstream and downstream of the strand-specific motif; and 3) Same distance from borders to the most conserved nucleotides in variable motifs. For those core motifs that did not meet criteria (1), we also performed motif comparison and clustering based on criteria 2) and 3) using the exo-defined matrix.

ATAC-seq and Data Analysis. ATAC-seq library preparation and data analysis were performed as described previously (8). Briefly, cells were treated with DNase (Invitrogen) at a final concentration of 50 U/ml for 30 min. Then cells were washed, trypsinized and resuspended in cold PBS. Each biological replicate consisted of 50,000 cells which were resuspended in cold ATAC-seq resuspension buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, and 3 mM $MgCl_2$). Cell nuclei were then prepared by incubation in 50 µl of ATAC-seq resuspension buffer containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin on ice for 3 min. After centrifugation, nuclei were resuspended in 50 µl of transposition mix (25 µl 2X TD buffer, 2.5 µl Nextera Tn5

transposase (Illumina), 16.5 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20, and 5 µl water), and incubated at 37 °C for 30 min in a thermomixer with shaking at 1,000 rpm. Transposed fragments were then purified with a Zymo DNA Clean and Concentrator-5 Kit. All libraries showed sufficient amplification after the 5 pre-amplification cycles and were quantified using the KAPA Library Quantification Kit (Roche). Paired-end sequencing was performed using an Illumina HiSeq 4000 instrument at OSUCCC. Sequence reads were obtained and mapped to the human (hg19) genomes using Bowtie2 with standard parameters and a maximum fragment length of 2,000. Picard was then used to remove duplicate reads.

GST Pull-Down Assay. PCR-amplified fragments of human gene AR-V7 were cloned in frame into SgfI-PmeI sites of the Promega Flexi vector. All constructs were verified by DNA sequencing. The resulting plasmids were then transfected into KRX competent cells (Cat#L3002, Promega) and expression of GST-AR-V7 protein fusions was induced by 0.1% rhamnose overnight according to the manufacturer's protocol (Promega). The bacteria were lysed with lysis buffer and the GST fusion proteins were bound to magneGST particles. HoxB13 was synthesized *in vitro* with HoxB13 expression vector (Cat#RC209991, OriGene) in the TnT-T7 quick coupled transcription/translation reaction as described for the MagneGST pull-down system (Cat#V8870, Promega). Typically, 20 µl of TnT *in vitro* translated HoxB13 product per reaction was incubated with the immobilized GST-AR-V7 fusion proteins (5 µl of GST fusion protein immobilized magneGST particles per reaction) in a total volume of 200 µl of magneGST binding/washing buffer at 4 °C with rotation. After 1 h incubation, the beads were washed 5 times with 250 µl ice-cold binding/washing buffer. HoxB13 proteins remaining bound to the immobilized GST fusion proteins were eluted with 20 µl of 2xSDS loading buffer and the eluted

proteins were resolved by SDS-PAGE. HoxB13 western blot was then performed using an anti-HoxB13 (H-80) antibody from Santa Cruz Biotechnology.

Immunohistochemistry (IHC) Analysis. For AR-V7 IHC, antigen retrieval was performed by placing the de-paraffinized slides in Antigen Unmasking Solution, Tris-Based High pH (H-3301, Vector Lab) at 95°C for 20 min. The slides were incubated with the rabbit monoclonal anti-human AR-V7 antibody (1:100, RevMAb Biosciences, South San Francisco, CA) for 1 h at room temperature. A ready-to-use goat anti-Rabbit poly HRP IgG (MP-7451, Vector lab, Burlingame, CA) was used as the secondary antibody by 30 min incubation at room temperature. For HoxB13 IHC, the antigen retrieval was performed by placing the de-paraffinized slides in Antigen Unmasking Solution, Citric Acid based Low pH (H-3300, Vector Lab) at 95°C for 30 min. The slides were incubated with the sheep polyclonal anti-human HoxB13 antibody (1:30, R & D Systems, Minneapolis, MN) at 4°C overnight. The secondary antibody was a rabbit anti-sheep poly HRP IgG (1:200, ThermoFisher Scientific, Rockford, IL) and was incubated with the slides for 30 min at room temperature. Slides were scanned using an Aperio Digital Pathology Slide Scanner (Leica Biosystems) at 40X magnification and staining quantified using the Aperio Image Scope (v11). Positive pixel count was set to detect the fraction of pixels that exceeded set threshold limits for weak, moderate, and strong staining in the brown colorimetric channel (AR-V7: Hue value=0.1, Hue Width 0.5, Color Saturation Threshold 0.04, Iwp (high)=185, Iwp (low)=175, Ip (low)=170, Isp (low)=0, Inp (high)= -1; HoxB13: Hue value=0.1, Hue Width 0.5, Color Saturation Threshold 0.08, Iwp (high)=200, Iwp (low)=195, Ip (low)=190, Isp (low)=0, Inp (high)= -1). Data represent the ratio of positive pixels per total pixels from areas of tissue containing CRPC.

Analysis of AR-V7 and HoxB13 mRNA expression in circulating tumor cells from CRPC patients. For blood-based HoxB13 measurements, a total of 86 consecutive samples collected during this ongoing study from November 2016 to June 2017 were evaluated for the purpose of examining the correlation of AR-V7 and HoxB13 levels. Blood collection, processing, and CTC isolation steps were carried out according to the Adnatest platform (Qiagen, Hanover, Germany) as described previously (9). CTC samples used in this study (n=86), each representing a unique sampling event, were processed for mRNA analysis using quantitative RT-PCR. Expression data for AR-V7 and HoxB13 were normalized to a control gene (*RPL13A*).

RNA *in situ* hybridization (RISH). RISH was performed by using RNAscope® 2.5 HD Detection Reagent – RED (Advanced Cell Diagnostics). RNAscope® probe for *AAK1* was designed to target 1111-1258 of AAK1 mRNA (NM_001143935.1); RNAscope® probe for *CROT* was designed to target 632-1620 of CROT mRNA (NM_014911.3). The RISH assays were performed following the user manual of RNAscope® 2.5 HD Detection Reagent – RED (Document Number 322360-USM, Advanced Cell Diagnostics). Formalin-fixed paraffin-embedded (FFPE) CRPC tissues were sectioned and the slides were baked at 60 °C for 1 h. The slides were de-paraffinized in xylene for 5 min twice and then dehydrated in 100% ethanol twice. The slides were allowed to air-dry, followed by multiple pretreatment steps. The slides were permeabilized by protease at 40 °C for 30 min to allow the probe access to mRNA targets. The slides were then hybridized with target probes at 40 °C for 2 h. Each tissue was stained with probes for *AAK1* and *CROT*. The probe hybridization was followed by a series of preamplifier and amplifier steps. The hybridization signals were detected by red chromogen staining, and the slides were then counterstained with 50% hematoxylin.

Mouse Xenograft Studies. Mouse xenograft studies were performed as described previously (10). All experiments were conducted in accordance with the guidelines of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) International and The Ohio State University Institutional Animal Care and Use Committee approved protocol. Briefly, male, 7-week-old, athymic nude mice were obtained from Charles River Laboratory and acclimated for 1 week in a pathogen-free enclosure before the start of the study. At 48 h post-infection of CWR22RV1 and LNCaP95 cells with lentivirus encoding specific shRNA against HoxB13 (pLKO_HoxB13_#1 and pLKO_HoxB13_#2 from Addgene #70093 and #70094 (11)), HoxB13 (11) plus AR-V7 (12), or Control shRNA, 1 million cells were inoculated subcutaneously into castrated mice. Tumor volumes were measured twice a week and were calculated using the standard formula: V =length x width^2 x 0.5. After 22-34 days, mice were euthanized and tumor tissues were weighed. For xenograft ChIP analysis, the tumor tissues were trimmed and chopped into small pieces on ice and fixed immediately with 2% formaldehyde for 20 min at room temperature. After sonication, chromatin was immunoprecipitated with specific antibodies at 4°C overnight. The ChIP DNA was purified and then analyzed by real-time PCR.

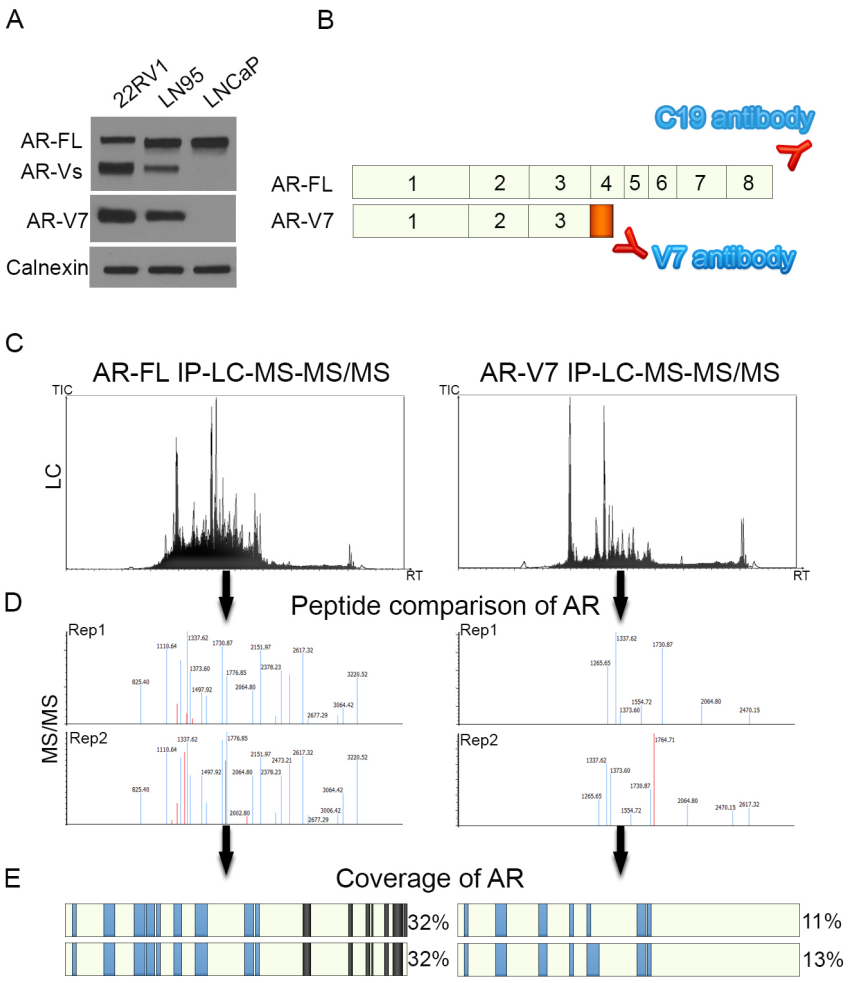

Fig. S1. AR-V7 does not interact with AR-FL in the absence of hormone. (*A*) Western blot analysis of AR-V7 and AR-FL expression in 22RV1, LN95 and LNCaP cells. (*B*) Schematic representation of the gene structures of AR-FL and AR-V7. The numbers denote exons. An antibody recognizing the C-terminus of AR was used to pull-down AR-FL but not AR-V7, whereas an antibody targeting a specific cryptic exon (CE3, Red) within the AR intron 3 was used to specifically precipitate AR-V7. (*C*) NanoLC-MS/MS base peak chromatogram of AR-FL pull-down and AR-V7 pull-down. (*D*) Mass spectrometry spectra of AR. *(E)* Coverage of AR. (FL-specific peptides were highlighted in black).

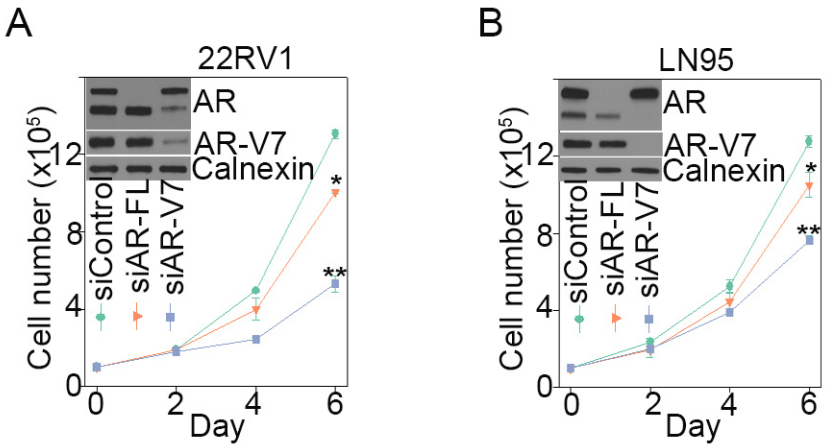

Fig. S2. Silencing of AR-V7 but not AR-FL dramatically decreases castration-resistant growth of 22RV1 and LN95 cells. (*A* and *B*) Hormone-depleted 22RV1 cells (*A*) or LN95 cells (*B*) were transfected with control siRNAs and siRNAs targeting AR-FL or AR-V7, and western blots (top-left panels) were performed using indicated antibodies. 22RV1 (*A*) or LN95 (*B*) cell proliferation was measured by direct cell count assays. The results are shown as mean ± SD (n = 2). The significance was determined by one-way ANOVA. **P*<0.01, ***P*<0.001.

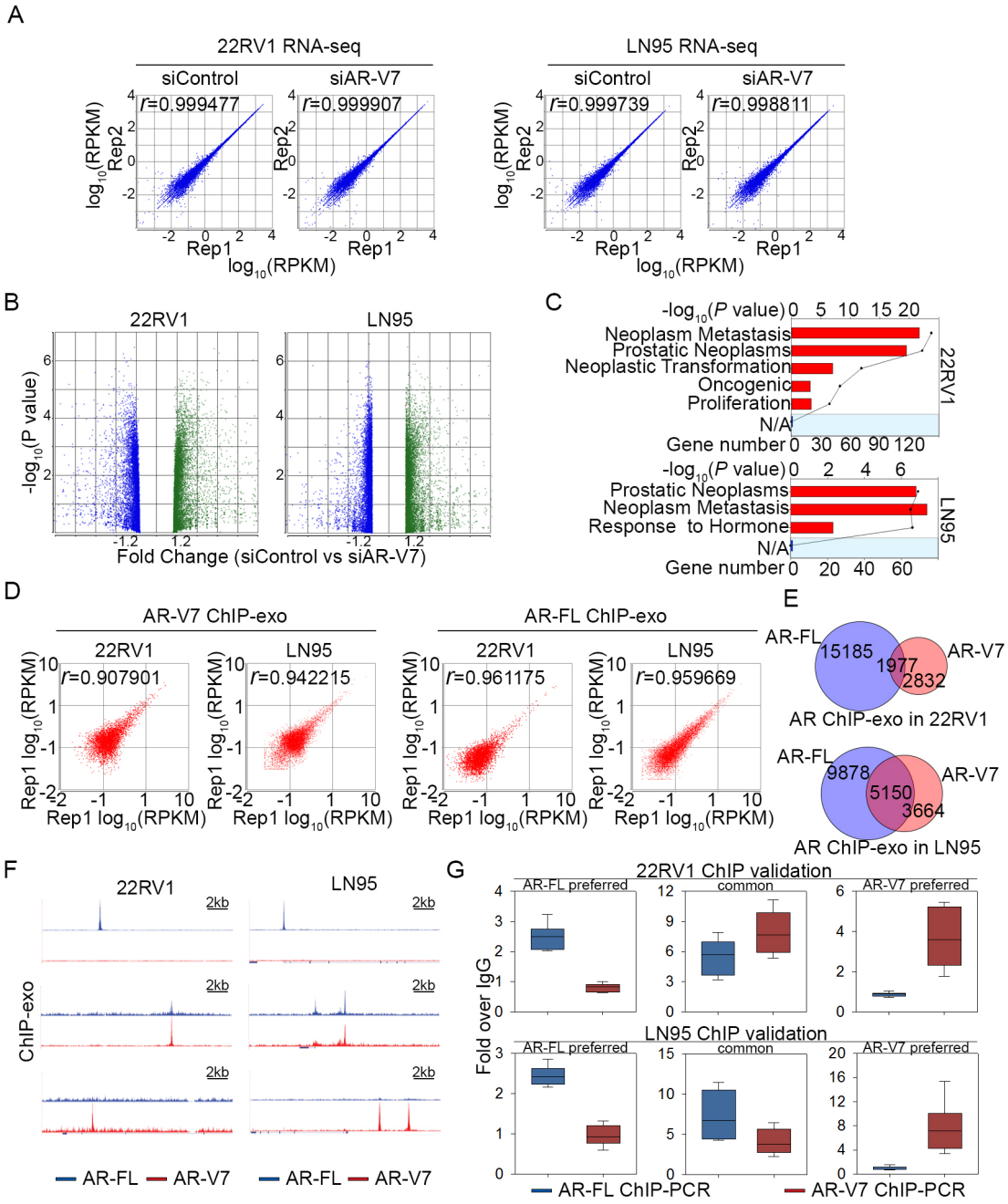

Fig. S3. Definition of AR-V7-regulated transcriptomes and AR-V7 cistromes in CRPC cells. (*A*) Correlation between biological replicates of RNA-seq in 22RV1 and LN95 cells transfected with control siRNAs and siAR-V7. (*B*) Volcano plots of pairwise gene expression changes in siControl and siAR-V7 treatment. All differentially expressed genes are highlighted in color (green for up-regulated genes and blue for down-regulated genes) with FDR<0.05. (*C*) Enriched Gene Ontology (GO) terms in AR-V7 regulated genes in 22RV1 and LN95 cells. (*D*) Correlation between biological replicates of AR-V7 and AR-FL ChIP-exo in 22RV1 and LN95 cells. (*E*) Venn diagram shows overlap between AR-V7 and AR-FL binding locations in 22RV1 cells and LN95 cells. (*F*) UCSC genome browser views of representative AR-V7 and AR-FL preferred and common binding sites in 22RV1 and LN95 cells. (*G*) AR-FL and AR-V7 ChIP assays were performed in randomly selected AR-FL preferred enhancers, AR-V7 preferred enhancers, and common enhancers in both cell lines (n=4 regions in each group).

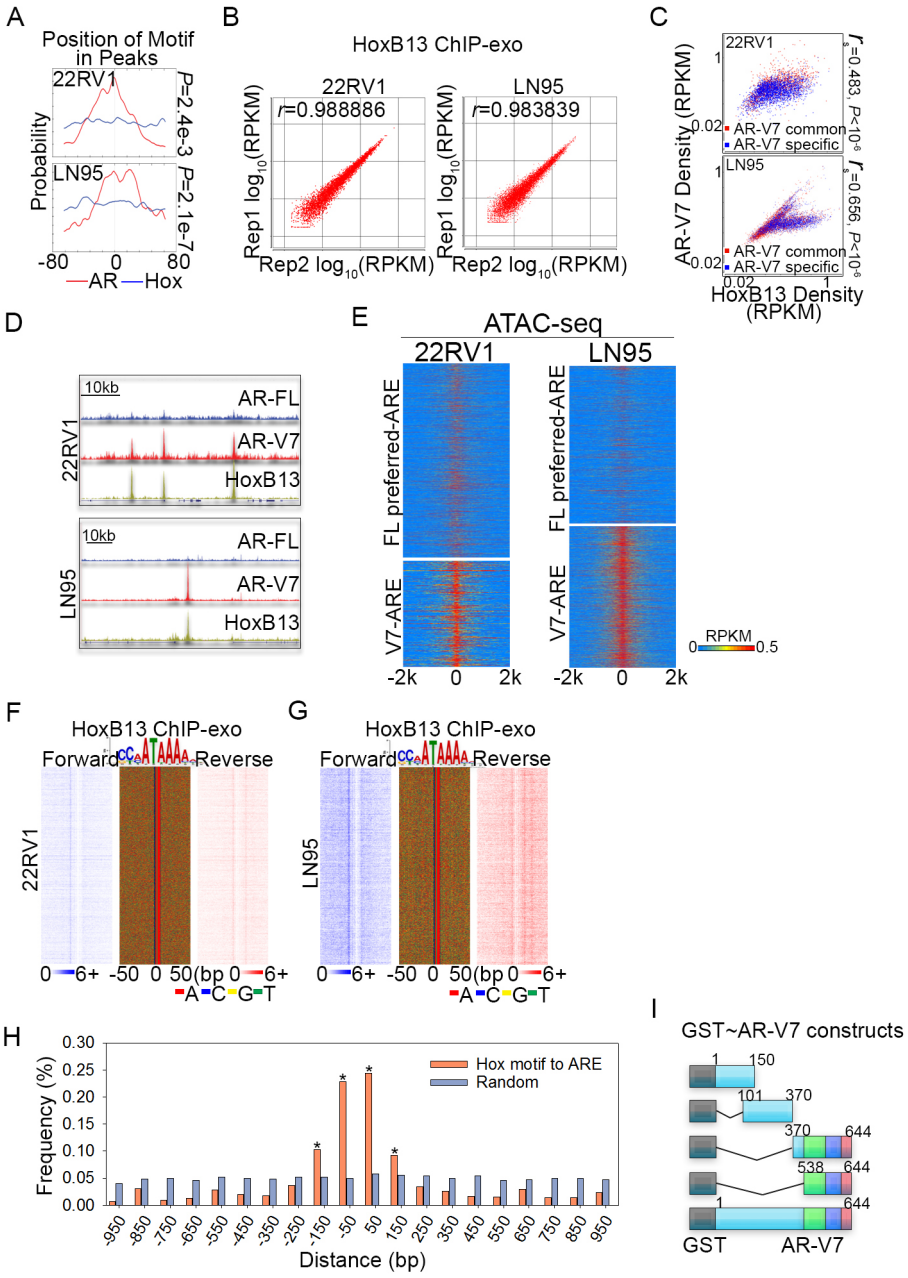

Fig. S4. HoxB13 interacts with AR-V7 on open chromatin. (*A*) Probability of ARE and Homeobox motif occurrence at AR-V7 binding peaks. The window indicates ± 80 bp regions from the ARE peak summits. (*B*) Correlation between biological replicates of HoxB13 ChIP-exo in 22RV1 and LN95 cells. (*C*) Correlation between binding signals of AR-V7 and HoxB13 in 22RV1 and LN95 cells. (*D*) UCSC genome browser views of representative AR-V7 and HoxB13 common binding sites in 22RV1 and LN95 cells. (*E*) Heatmap showing ATAC-seq tag distribution around AR-FL preferred ARE regions and AR-V7 ARE regions within ± 2 kb. (*F* and *G*) ChIP-exo raw tags distribution (1 bp-resolution) over HoxB13 motifs on the forward (blue, left panel) and reverse (red, right panel) strands, respectively in 22RV1 cells (*F*) and LN95 cells (*G*). The center panels represent the bound HoxB13 motifs ordered as in the left and right panels. (*H*) Comparison of frequencies of genomic distances between HoxB13 motifs and AR-V7-bound AREs. *P<1x10^{-5}. (*I*) Schematic representation of GST~AR-V7 fusion proteins.

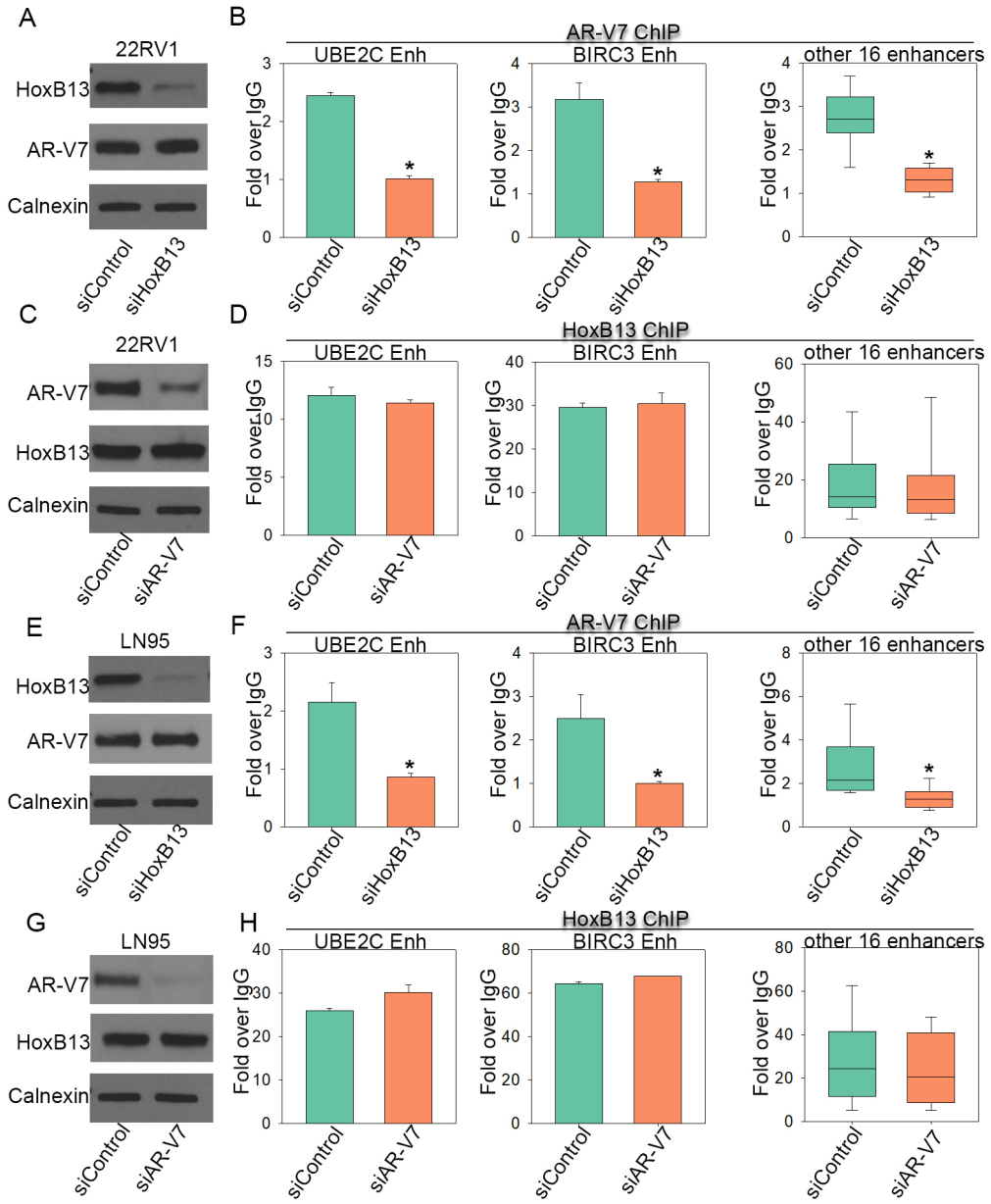

Fig. S5. HoxB13 pioneers AR-V7 binding in CRPC cells. (*A* and *E*) Hormone-depleted 22RV1 cells (*A*) or LN95 cells (*E*) were transfected with siControl or siHoxB13, and HoxB13 and AR-V7 Western blot analyses were performed. (*B* and *F*) ChIP assays were performed in siControl and siHoxB13 transfected 22RV1 (*B*) or LN95 cells (*F*) to examine AR-V7 binding to the enhancers of *UBE2C*, *BIRC3*, 16 selected 22RV1 AR-V7/HoxB13 co-binding locations and 16 selected LN95 locations. (*C* and *G*) Hormone-depleted 22RV1 cells (*C*) or LN95 cells (*G*) were transfected with siControl or siAR-V7, and AR-V7 and HoxB13 Western blot analyses were performed. (*D* and *H*) ChIP assays were performed in siControl and siAR-V7 transfected 22RV1 (*D*) or LN95 cells (*H*) to examine HoxB13 binding to the enhancers examined in panels D and H. The significance was determined by one-tailed t-test. **P*<0.01.

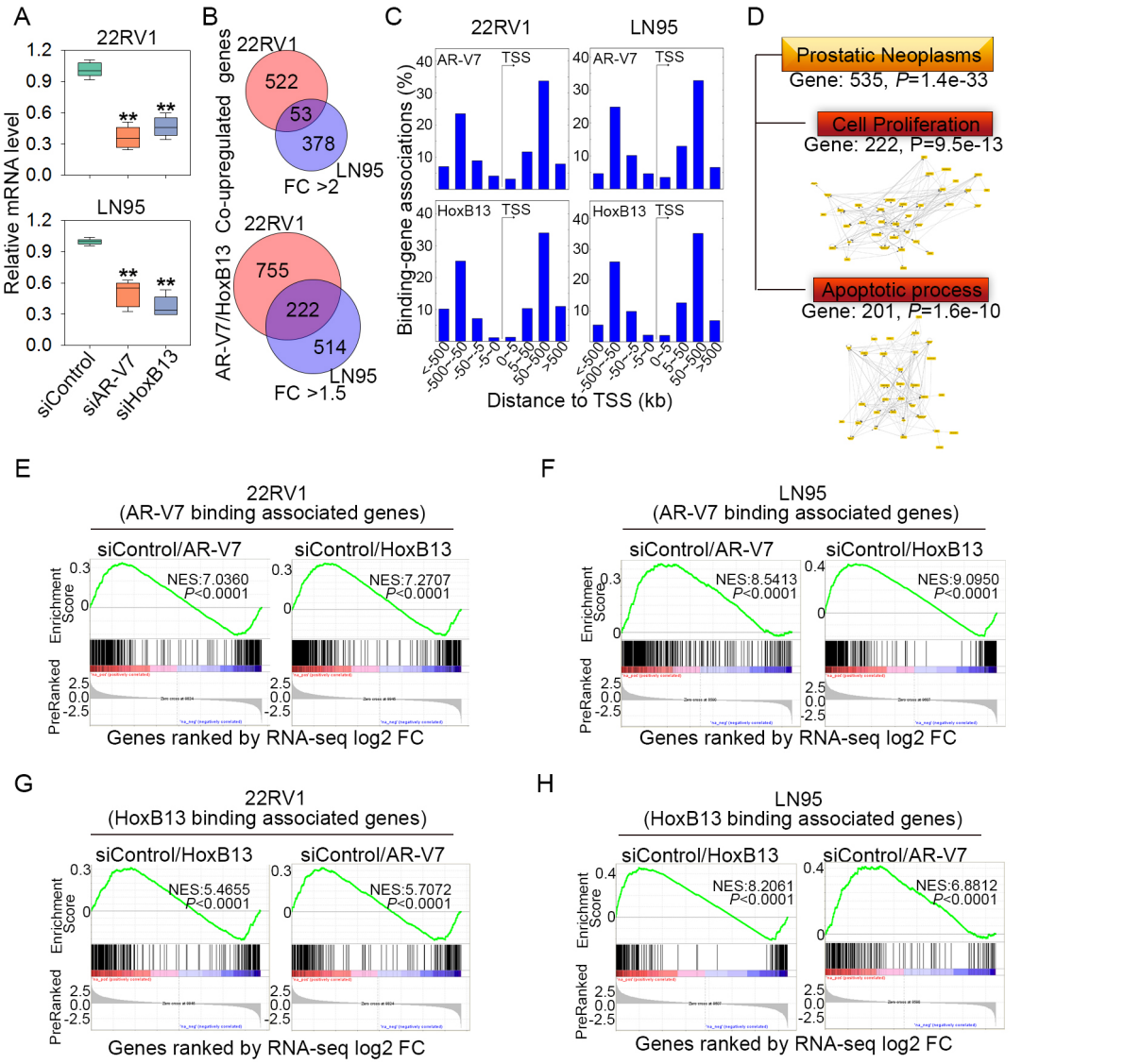

Fig. S6. HoxB13 and AR-V7 directly co-upregulate target genes in CRPC cells. (*A*) RT-PCR validation of RNA-seq data. Genes validated in 22RV1 cells are *CROT*, *ETV1*, *PDGFA*, *SOX9*, and *NKX3-1*, and genes examined in LN95 cells are *RHOU*, *PDE9A*, *RAB40B*, *CROT* and *ZBTB10*. The significance was determined by one-tailed t-test. ***P*<0.001. (*B*) Venn diagrams show overlap of AR-V7 and HoxB13 co-upregulated genes between 22RV1 and LN95 cells. (*C*) The distribution of AR-V7 and HoxB13 binding around the transcription start site (TSS) of AR-V7 and HoxB13 co-regulated genes in 22RV1 and LN95 cells is shown. (*D*) Pathways and networks constructed using AR-V7 and HoxB13 co-upregulated genes (Genomatix). (*E* and *F*) GSEA analyses compare associated genes within ± 20 kb of AR-V7 binding locations with AR-V7 or HoxB13 regulated genes determined by RNA-seq analysis in 22RV1 (*E*) and LN95 cells (*F*). (*G* and *H*) GSEA analyses compare associated genes within ± 20 kb of HoxB13 binding locations with AR-V7 or HoxB13 regulated genes determined by RNA-seq analysis in 22RV1 (*G*) and LN95 cells (*H*).

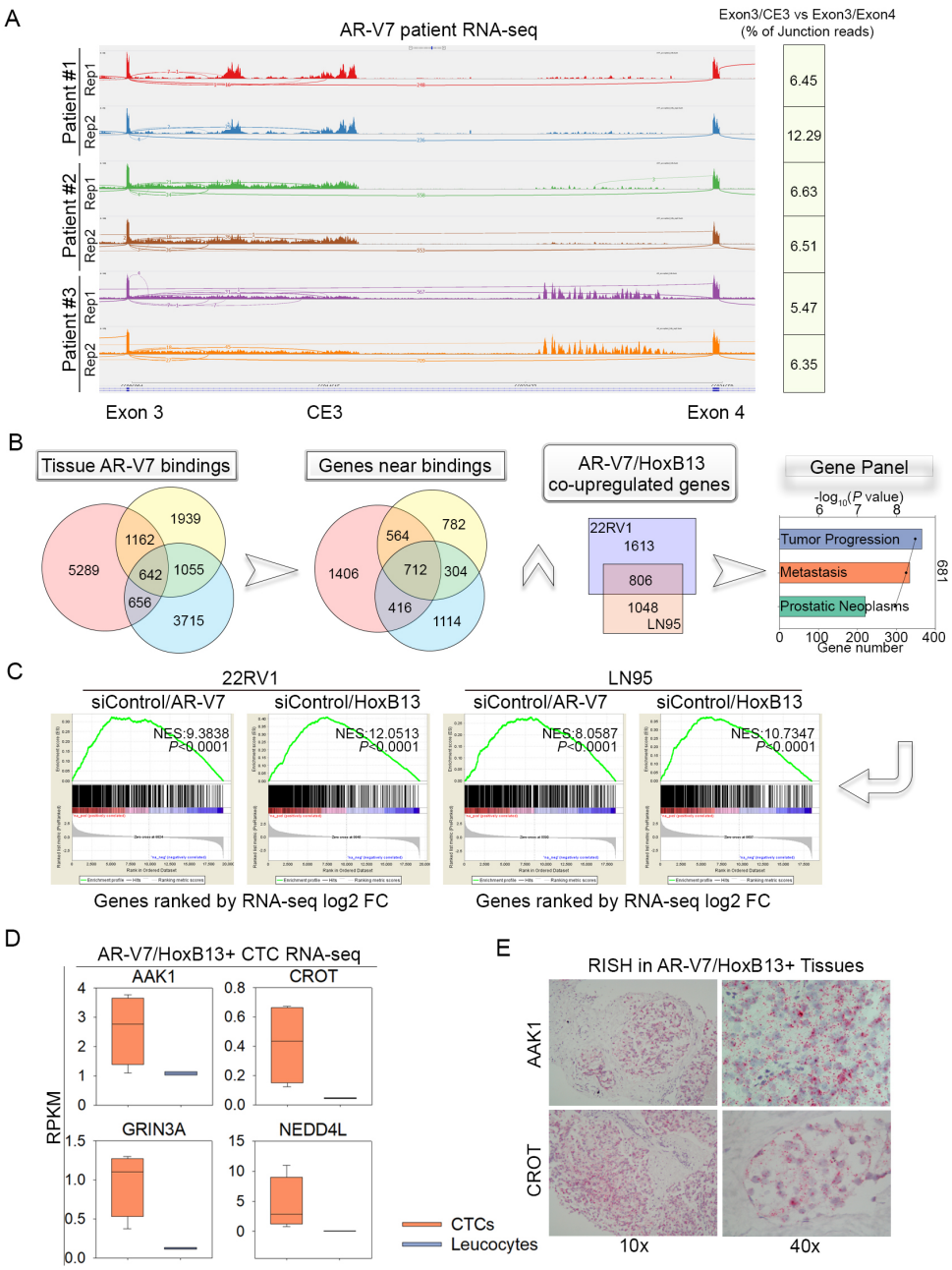

Fig. S7. HoxB13 and AR-V7 transcriptional co-regulation in CRPC patients. (*A*) Sashimi plots to visualize splice junctions. The regions between *AR* exon 3 and exon 4 are shown for each CRPC patient. The ratio of exon3/CE3 junction-PE reads *vs* exon3/exon4 junction-PE reads was calculated and shown in the right column. (*B*) A workflow for integrative analysis of AR-V7 and HoxB13 co-regulated genes in CRPC patients and CRPC cells. A 681-gene panel was generated by overlapping genes within 50 kb of CRPC tissue AR-V7 binding sites with AR-V7/HoxB13 co-upregulated genes in 22RV1 and LN95 cells (FDR<0.05, Fold>1.2), and enriched GO terms in this gene panel were shown through Genomatix pathway system analysis. (*C*) GSEA analysis compared the gene panel and AR-V7 /HoxB13 regulated genes determined by RNA-seq analysis in 22RV1 and LN95 cells. (*D*) Box plots compare expression of *AAK1*, *CROT*, *GRIN3A* and *NEDD4L* in CRPC CTC (see also Fig. 4*K*) RNA-seq datasets (CTC n=4; leucocyte n=2). (*E*) Representative images demonstrate positive detection for the *AAK1* and *CROT* mRNA expression in AR-V7/HoxB13+ CRPC tissues (see also Fig. 4*H*, *I*). Red punctate labeling was identified from the RNA *in situ* hybridization (RISH) assays.

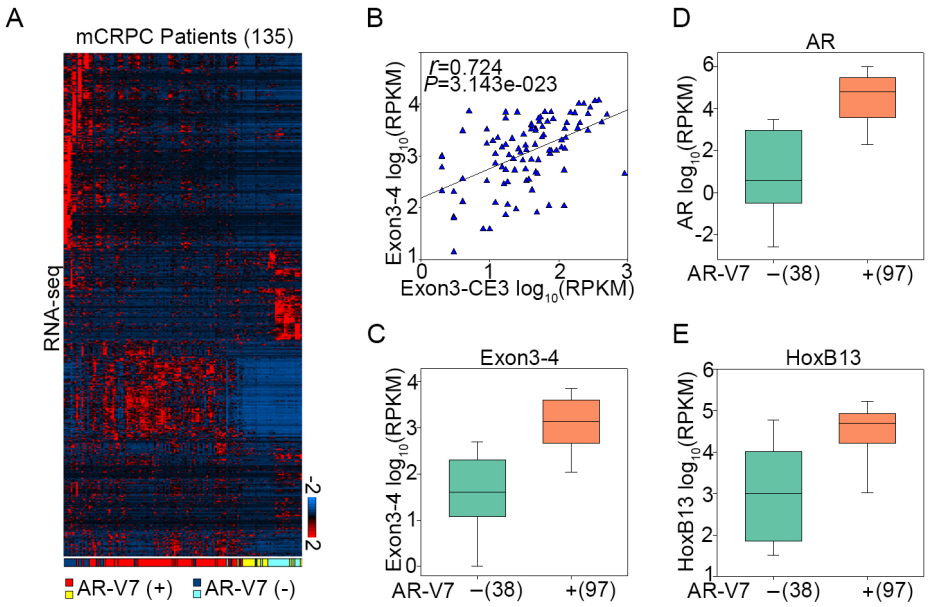

Fig. S8. HoxB13 expression is increased in AR-V7+ versus AR-V7- CRPC patients. (*A*) A heatmap shows gene expression of 135 patients. Genes that showed 2-fold difference between AR-V7+ and AR-V7- groups were used in hierarchical clustering. The scale bar is shown with the minimum expression value for each gene in blue and the maximum value in red. (*B*) Correlation of gene expression values (RPKM) between *AR* Exon3-CE3 and *AR* Exon3-4. (*C-E*) Box plots compare gene expression values (RPKM) of *AR* Exon 3-4 (*C*), *AR* (*D*), and *HoxB13* (*E*) between AR-V7+ and AR-V7- groups in 135 patients.

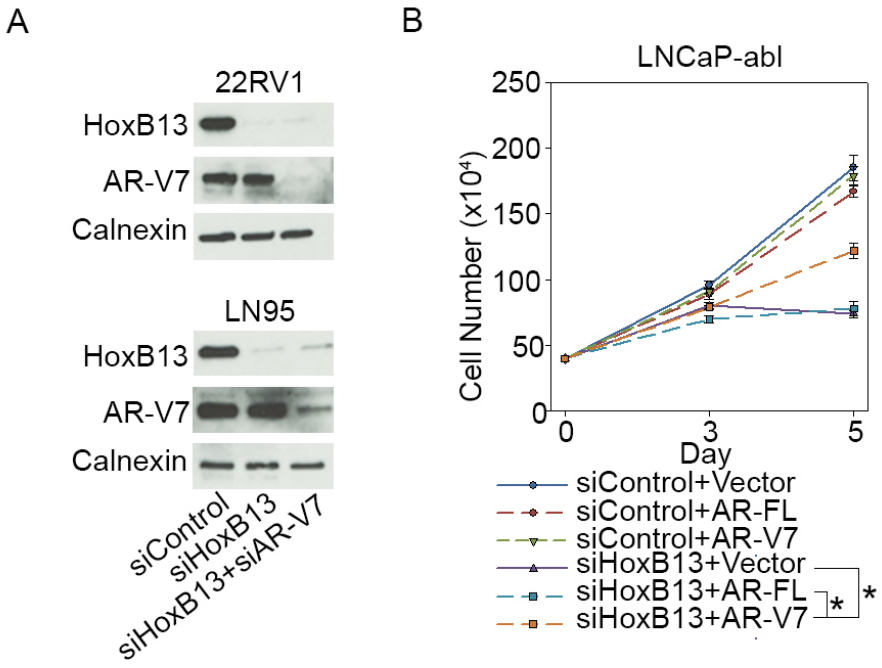

Fig. S9. AR-V7 mediates HoxB13 function in CRPC. (*A*) 22RV1 cells and LN95 cells were transfected with control siRNAs, siRNAs targeting HoxB13, and combined siRNAs targeting HoxB13 and AR-V7. Western blots were performed using indicated antibodies. (*B*) LNCaP-abl cells were co-transfected with siRNAs combined with pcDNA3.1 vector, pcDNA3.1-AR-FL (13), or pcDNA3.1-AR-V7 (1). Cell proliferation was measured by direct cell count assays. The significance was determined by one-way ANOVA. **P*<0.01.

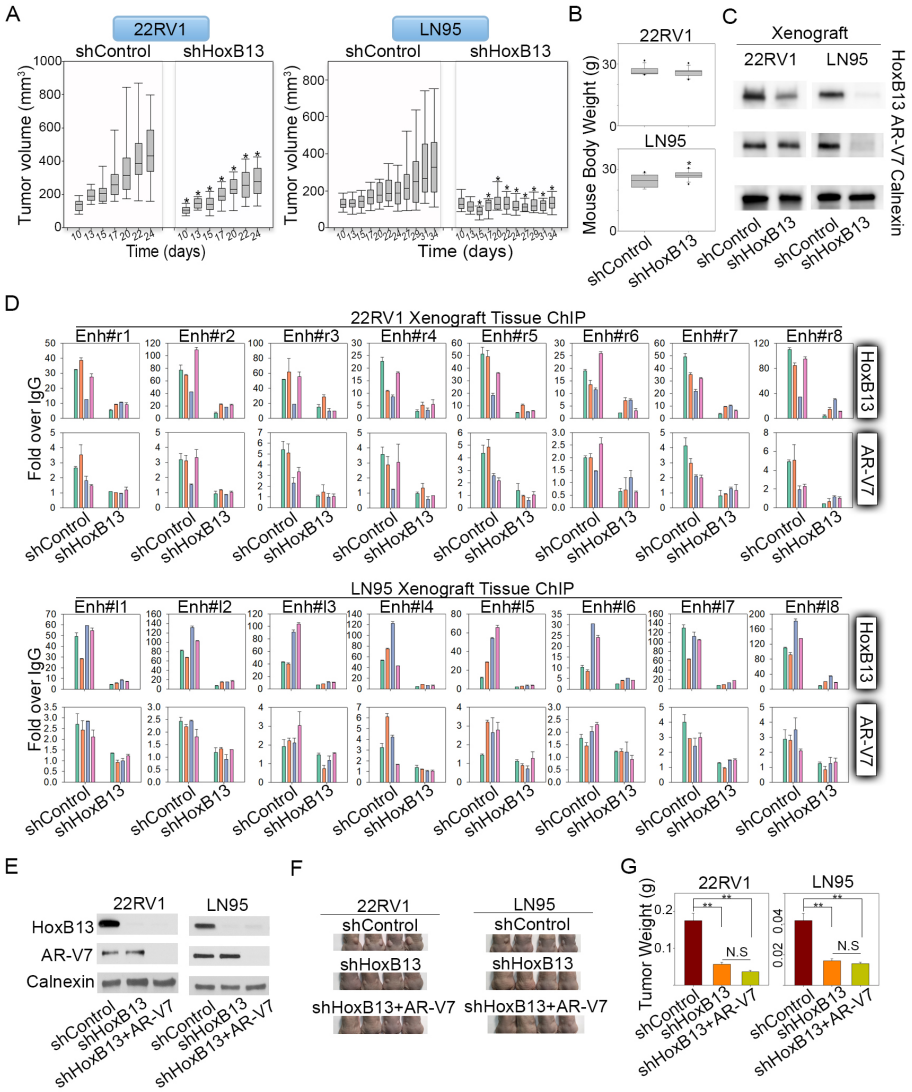

Fig. S10. HoxB13 silencing decreases CRPC growth *in vivo* through inhibition of AR-V7 binding and AR-V7 oncogenic function. (*A*) Forty-eight h post infection with shControl or shHoxB13 virus, 22RV1 (left) and LN95 (right) cells were injected into castrated mice and tumor volumes were measured twice per week. *P* values were determined by one-tailed t-test. **P*<0.001. (*B*) Average body weights for control (n=12) and HoxB13 silenced groups (n=12) of mice harboring the 22RV1 xenograft tumors at time of collection (day 24; upper panel), and for control (n=12) and HoxB13 silenced groups (n=12) of mice harboring the LN95 xenograft tumors at time of collection (day 34; lower panel). The results are shown as mean ± SD. The significance was determined by one-tailed t-test. **P*<0.01. (*C*) HoxB13 and AR-V7 expression in 22RV1 or LN95 combined tumors was measured by Western blots. (*D*) HoxB13 and AR-V7 tissue ChIP was performed using engrafted 22RV1 tumors (shControl group n = 4; shHoxB13 group n = 8, combined n = 4) and LN95 tumors (shControl group n =4; shHoxB13 n = 16, combined n = 4). The results for each column are shown as mean ± SD (n = 2). (*E*) HoxB13 and AR-V7 protein expression in 22RV1 cells (left) and LN95 cells (right) infected with lentivirus encoding HoxB13 shRNA, HoxB13 shRNA plus AR-V7 shRNA, or a control shRNA after 48 h. Western blots were performed using the same cells used for injection into castrated mice. (*F*) Images of the dorsal flank containing subcutaneous xenografts from representative animals per group. (*G*) Average tumor weight of 22RV1 (left) or LN95 (right) tumors for control (n=24), HoxB13 silenced groups (n=24), and HoxB13 and AR-V7 co-silenced groups (n=24) at time of collection (day 22). The results are shown as mean ± SE. The significance was determined by one-way ANOVA ***P*<0.001.

Additional data Datasets S1-S3 (separate files)

Dataset S1. Gene lists for AR-V7 and HoxB13 regulated genes

Dataset S2. Primers for ChIP, RT-PCR, GST constructs, and RNAi.

Dataset S3. Expression of the Homeobox genes in 22RV1 and LN95 cells.

References

1. Hu R, *et al.* (2012) Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer research* 72(14):3457-3462.
2. Kato M, *et al.* (2016) Cotargeting Androgen Receptor Splice Variants and mTOR Signaling Pathway for the Treatment of Castration-Resistant Prostate Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 22(11):2744-2754.
3. Culig Z, *et al.* (1999) Switch from antagonist to agonist of the androgen receptor bicalutamide is associated with prostate tumour progression in a new model system. *Br J Cancer* 81(2):242-251.
4. Chen Z, *et al.* (2015) Ligand-dependent genomic function of glucocorticoid receptor in triple-negative breast cancer. *Nature communications* 6:8323.
5. Chen Z, *et al.* (2015) Agonist and antagonist switch DNA motifs recognized by human androgen receptor in prostate cancer. *EMBO J* 34(4):502-516.
6. Lan X, Bonneville R, Apostolos J, Wu W, & Jin VX (2011) W-ChIPeaks: a comprehensive web application tool for processing ChIP-chip and ChIP-seq data. *Bioinformatics* 27(3):428-430.
7. Bailey TL, *et al.* (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* 37(Web Server issue):W202-208.
8. Corces MR, *et al.* (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature methods* 14(10):959-962.
9. Antonarakis ES, *et al.* (2014) AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *The New England journal of medicine* 371(11):1028-1038.
10. Wang H, *et al.* (2011) CCI-779 Inhibits Cell-Cycle G2-M Progression and Invasion of Castration-Resistant Prostate Cancer via Attenuation of UBE2C Transcription and mRNA Stability. *Cancer research* 71(14):4866-4876.
11. Pomerantz MM, *et al.* (2015) The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nature genetics* 47(11):1346-1351.
12. Guo Z, *et al.* (2009) A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer research* 69(6):2305-2313.
13. Wang Q, Udayakumar TS, Vasaitis TS, Brodie AM, & Fondell JD (2004) Mechanistic relationship between androgen receptor polyglutamine tract truncation and androgen-dependent transcriptional hyperactivity in prostate cancer cells. *The Journal of biological chemistry* 279(17):17319-17328.