Supplementary Information

Cooperation between Polycomb and Androgen Receptor during Oncogenic Transformation

Jonathan C. Zhao[1,2], Jianjun Yu[1,2], Christine Runkle[1,2], Longtao Wu[1,2], Ming Hu[4], Dayong Wu[6], Jun Liu[4], Qianben Wang[6], Zhaohui Qin[5], Jindan Yu[1,2,3]

Contents:

Supplemental Experimental Procedures

Analysis of ARE motif frequency

To examine the frequency of ARE motifs in ChIP-seq peaks, ChIP-seq peaks were sorted by the peak height and the top 250 peaks were separated into 10 bins of decreasing peak height. For each peak, a control region with equal length was selected as the region 200 bp downstream of the peaks for AR-repressed genes. Motif score was defined as the ratio between the foreground motif model (position specific weight matrix reported by MDscan) and the background model (3-order Markov background model training from 5,000bp upstream regions). Any position with a motif score larger than 1 was considered as a motif. The total number of ARE motifs found in each bin was then determined. Two-sample t-test was used to compare the number of ARE motifs between AR-induced, -repressed, and control sequences.

Linear mixed effect model

Assume y_ij is the ChIP-seq read intensity for gene i in the j th 100 bp window. We fit a linear mixed effect model as following:

y_ij = beta0 + beta1*I(gene_i is induced) + beta2*j + gamma_i + e_ij

Here beta0 is the baseline effect, beta1 is the difference between AR-induced genes and repressed genes, beta2 is the distance effect (the distance between the j th window and TSS). gamma_i is the random effect for the i th gene since for fixed i, y_ij's are correlated. e_ij is the white noise. The hypothesis to test is whether beta1 is significantly different from zero. The p value of this difference is then determined.

Quantitative PCR

QRT-PCR or quantitative ChIP-PCR were performed using GoTaq® qPCR Master Mix 2X (Promega, Madison, WI) on an Applied Biosystems StepOne Plus Real Time PCR System as previously described (Yu, 2010; Yu et al., 2007). All primers were designed using Primer 3 and synthesized by Integrated DNA Technologies and are listed below.

Primer Name	Sequence	application
GAPDH F	TGCACCACCAACTGCTTAGC	qRT-PCR
GAPDH R	GGCATGGACTGTGGTCATGAG	qRT-PCR
PSA F	ACGCTGGACAGGGGGCAAAAG	qRT-PCR
PSA R	GGGCAGGGCACATGGTTCACT	qRT-PCR
TMPRSS2 F	CAGGAGTGTACGGGAATGTGATGGT	qRT-PCR
TMPRSS2 R	GATTAGCCGTCTGCCCTCATTTGT	qRT-PCR
AR F	CAGTGGATGGGCTGAAAAAT	qRT-PCR
AR R	GGAGCTTGGTGAGCTGGTAG	qRT-PCR
Nov F2	ACCGTCAATGTGAGATGCTG	qRT-PCR
Nov R2	TCTTGAACTGCAGGTGGATG	qRT-PCR
OPRK1 F1	aactcgctggtcatgttcgt	qRT-PCR
OPRK1 R1	ctctgaaagggcatggttgt	qRT-PCR

DDC F1	gcccctacttcttcgcctac	qRT-PCR
DDC R1	cacagtctccagctctgtgc	qRT-PCR
MET F1	cgctgacttctccactggtt	qRT-PCR
MET R1	tacactccccattgctcctc	qRT-PCR
SI F1	ttttggcagccttatccaag	qRT-PCR
SI R1	caatcagagagatttccaatcca	qRT-PCR
FKBP5 F1	TCTCATGTCTCCCCAGTTCC	qRT-PCR
FKBP5 R1	TTCTGGCTTTCACGTCTGTG	qRT-PCR
KLK2 F1	CCATGCCTGGAGACATATCA	qRT-PCR
KLK2 R1	TCCAGCACATGTCACTCTCC	qRT-PCR
SLC43A1 F1	CTGCCTCCTGGTACCTCTTG	qRT-PCR
SLC43A1 R1	ACCTCAGGGGTAGCCTGTTT	qRT-PCR
PSA_AF3 (enhancer)	GCCTGGATCTGAGAGAGATATCATC	ChIP-PCR
PSA_AR3 (enhancer)	ACACCTTTTTTTTTCTGGATTGTTG	ChIP-PCR
TMPRSS2_pF1 (enhancer)	tggagctagtgctgcatgtc	ChIP-PCR
TMPRSS2_pR1 (enhancer)	ctgccttgctgtgtgaaaaa	ChIP-PCR
NOV pF2	ATGCACGTGCGTGTAAACAG	ChIP-PCR
NOV pR2	CACAAGGTTTCTGGGTAGGG	ChIP-PCR
KIAA0066_pF2	CTAGGAGGGTGGAGGTAGGG	ChIP-PCR
KIAA0066_pR2	GCCCCAAACAGGAGTAATGA	ChIP-PCR
OPRK1 pF1	CTGCTCCTGGCATTATCCTC	ChIP-PCR
OPRK1 pR1	TGTGGCTCTCAGCAGGAAGT	ChIP-PCR
DDC pF1	AATCCTTTGGCTGCCAGTTA	ChIP-PCR
DDC pR1	AGGTACTTCTGGGCATGGTG	ChIP-PCR
MET pF1	GCTGCTTTGGGAAATGGTTA	ChIP-PCR
MET pR1	TGGTCGCAGGTTTCAACATA	ChIP-PCR
SI pF1	TCCACATAGGGAAACAGTCCT	ChIP-PCR
SI pR1	GGAAATTGCCAAGGACAGAA	ChIP-PCR
KLK2 pF1	AGCATCTAGGTGCCAACAGG	ChIP-PCR
KLK2 pR1	GACAAGGCGATGGAGAGAAC	ChIP-PCR
PSA_EF5 (promoter)	CCTAGATGAAGTCTCCATGAGCTACA	ChIP-PCR
PSA_ER5 (promoter)	GGGAGGGAGAGCTAGCACTTG	ChIP-PCR
TMPRSS2 pF2 (promoter)	GGTAAACTCTCCCTGCCACA	ChIP-PCR
TMPRSS2 pR2 (promoter)	TACTCCAGGAAGTGGGGATG	ChIP-PCR

References:

Massie, C. E., Lynch, A., Ramos-Montoya, A., Boren, J., Stark, R., Fazli, L., Warren, A., Scott, H., Madhu, B., Sharma, N.*, et al.* (2011). The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. EMBO J *30*, 2719-2733.

Wang, Q., Li, W., Liu, X. S., Carroll, J. S., Janne, O. A., Keeton, E. K., Chinnaiyan, A. M., Pienta, K. J., and Brown, M. (2007). A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol Cell *27*, 380-392.

Yu, J., Cao, Q., Yu, J., Wu, L., Dallol, A., Li, J., Latif, F., Wu, J., Chinnaiyan, A.M. (2010). The neuronal repellent SLIT2 is a target for repression by EZH2 in prostate cancer. Oncogene *29*, 5370-5380.

Yu, J., Rhodes, D. R., Tomlins, S. A., Cao, X., Chen, G., Mehra, R., Wang, X., Ghosh, D., Shah, R. B., Varambally, S.*, et al.* (2007). A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. Cancer Res *67*, 10657-10663.

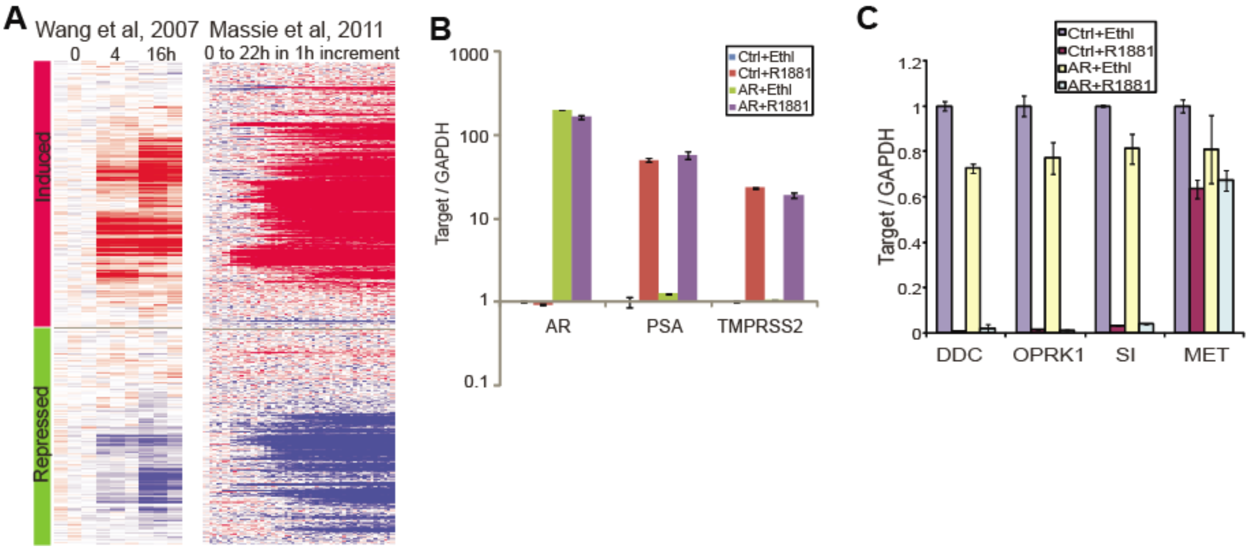

Figure S1 (related to Figure 1). Androgen represses target gene expression.

(A) The expression of our AR-induced and –repressed genes in two previously published androgen time-course treatment studies (Massie et al., 2011; Wang et al., 2007). The datasets GSE7868 and GSE18684 were downloaded from GEO database and gene expression was normalized to the 0 time-point of each dataset and viewed using Treeview.

(B) LNCaP cells were hormone-starved for 3 days followed by AR transfection in the presence or absence of R1881. QRT-PCR was used to determine the transcript levels of AR and several AR-induced genes.

(C) Expression of AR repressed genes upon AR overexpression as described in (B).

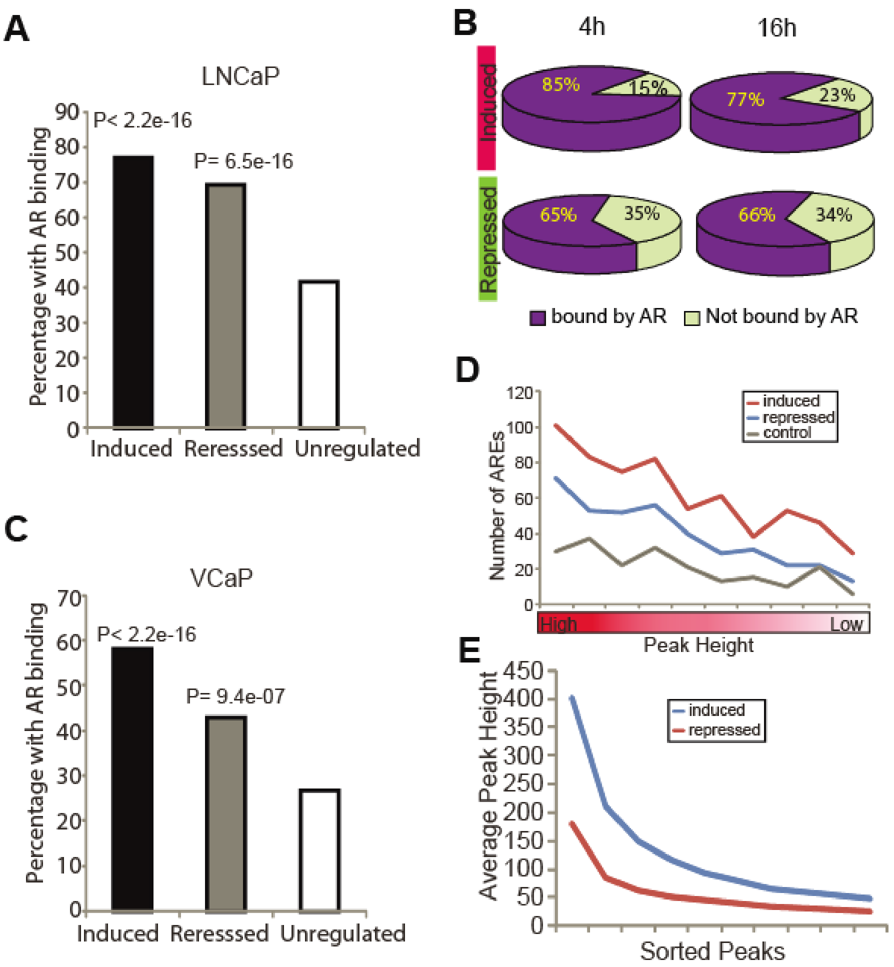

Figure S2 (related to Figure 2). AR binding events on androgen-regulated genes.

(A) The percentage of androgen-inducd and –repressed genes that contain at least one AR binding sites in LNCaP cells. Statistical differences were evaluated by comparing androgen-induced gens, -repressed genes to the control unregulated genes.

(B) Percentage of androgen-induced and –repressed genes that are bound by AR. Genes that are induced or repressed by androgen at 4 hr and 16hr of treatment were derived from previously published dataset GSE7868 (Wang et al., 2007). AR ChIP-seq data was used to determine whether these genes contain at least one AR binding events.

(C) The percentage of androgen-inducd and –repressed genes that contain at least one AR binding sites in VCaP cells. Statistical differences were evaluated by comparing androgen-induced gens, -repressed genes to the control unregulated genes.

(D) The number of consensus ARE motifs in the ChIP-seq AR bound regions. AR binding sites were first sorted by peak height and then grouped into 10 groups of equal size. For each group, the number of ARE motifs present in each sequence was determined, summed up, and normalized to the total number of sequences analyzed. This analysis was performed separately for AR-induced, AR-repressed, and a random control set of genes.

(E) The average peak height of AR binding peaks of AR-induced and –repressed genes as described in **D**.

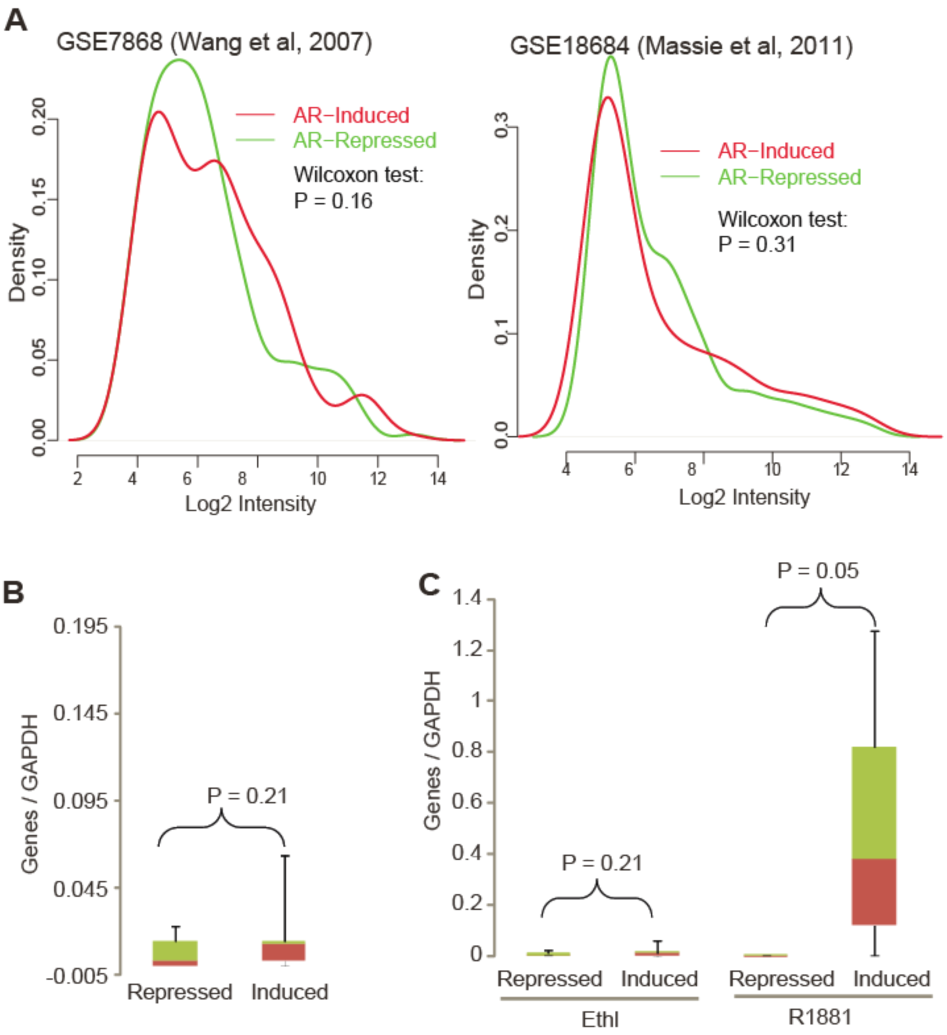

Figure S3 (related to Figure 3). Expression level of AR-induced or –repressed genes in LNCaP cells.

(A) The density plot of the basal expression levels of AR-induced and –repressed genes. The expression level of each gene at 0hr of androgen treatment was obtained from the corresponding dataset. The p value of Wilcoxon test of two groups of genes in each dataset is shown.

(**B**,**C**) Boxplots of the expression levels of 5 AR-induced and 5 AR-repressed genes in LNCaP cells without (**B**) or with (**C**) AR activation. LNCaP cells were hormone-starved for 3 days, treated with ethanol (Ethl) or 1nM androgen (R1881) for 2 days before subjected to RNA isolation. QRT-PCR was performed to determine gene expresion, which was normalized to the level of GAPDH. The differences between AR-induced and –repressed genes were analyzed by t-test and the p values are shown.

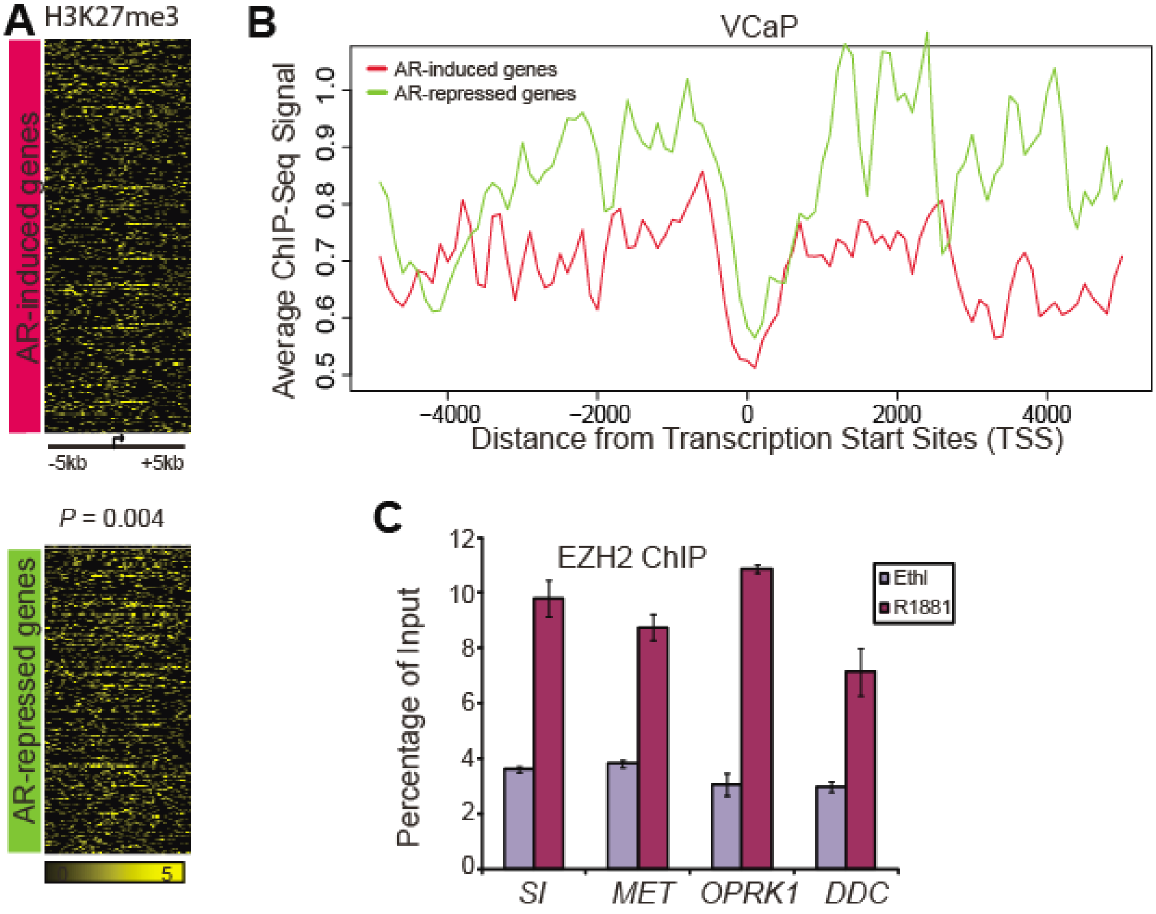

Figure S4 (related to Figure 4). EZH2 occupancy on AR-repressed genes is increased upon AR activation.

- **(A)** A yellow-blue heatmap view of Figure 4A to demonstrate the differences of H3K27me3 ChIP-seq read density between AR-induced and –reperssed genes.
- **(B)** Average H3K27me3 ChIP-seq signal around AR-induced (red) and AR-repressed (green) genes in VCaP. The H3K27me3 ChIP-seq signal is significantly (*$P<0.001$*) higher for AR-repressed than AR-induced genes.
- **(C)** EZH2 ChIP followed by PCR analysis in LNCaP cells in the absence of present of androgen. EZH2 occupancy significantly (*$P<0.03$*) increases upon AR activation.

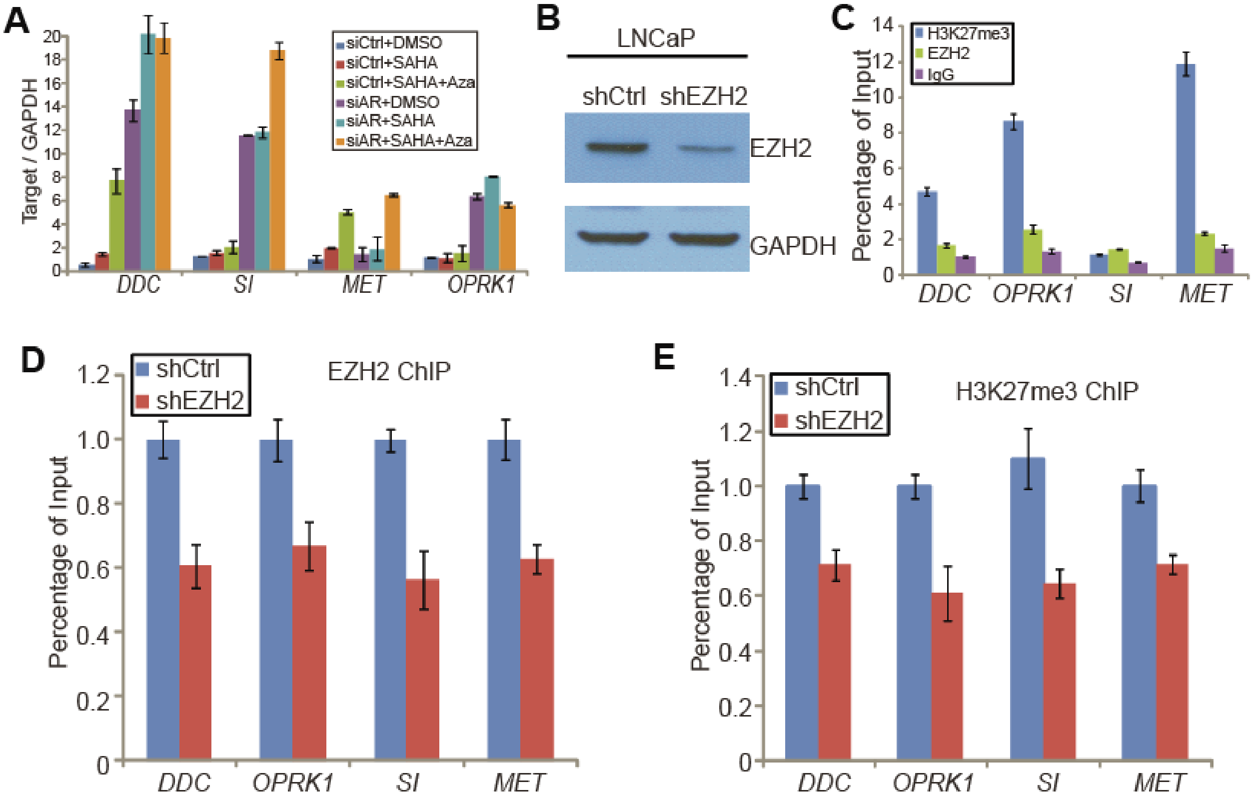

Figure S5 (related to Figure 4). EZH2 collaborates with AR on transcriptional repression.

- **(A)** HDAC inhibitor SAHA (5uM) and AR knockdown synergistically de-repress AR-repressed genes. LNCaP cells were treated with RNA interference of AR or 5uM SAHA/1uM AZA or both. Gene expression changes of AR-repressed genes were monitored by qRT-PCR. Error bars: n=3, mean ± SEM.
- **(B)** Western blot confirms EZH2 knockdown by shRNA in LNCaP cells.
- **(C)** ChIP followed by PCR showing EZH2 occupancy and H3K27me3 on the AR-repressed gene promoters. ChIP was performed using antibodies against EZH2 and H3k27me3 in LNCaP cells. The differences between H3K27me3 ChIP and IgG, and EZH2 ChIP compared to IgG are statistically significant with P values less than 0.03 and 0.02, respectiely.
- **(D)** ChIP using an EZH2 antibody in LNCaP cells with or without EZH2 knockdown. The levels of EZH2 enrichment are significantly ($P<0.02$) decreased in EZH2-knockdown cells.
- **(E)** ChIP using an H3K27me3 antibody in LNCaP cells with or without EZH2 knockdown. The levels of H3K27me3 modification are significantly ($P<0.42$) decreased in EZH2-knockdown cells.