AWARD NUMBER: W81XWH-16-1-0195

TITLE: Biological characterization and clinical utilization of metastatic prostate cancer-associated lincRNA *SChLAP1*

PRINCIPAL INVESTIGATOR: Lanbo Xiao

CONTRACTING ORGANIZATION: University of Michigan

Ann Arbor, Michgian 48109-1274

REPORT DATE: October 2018

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
October 2018	FINAL	15 June 2016 – 14 June 2018

4. TITLE AND SUBTITLE
Biological characterization and clinical utilization of metastatic prostate cancer-associated lincRNA SChLAP1

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0195

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Lanbo Xiao, Ph.D.

E-Mail: lanbox@med.umich.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Regents of the University of Michigan
3003 S. State Street
Ann Arbor, MI 48109-1274

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Long non-coding RNAs (lincRNAs) are polyadenylated RNA species that are more than 200bp in length and are implicated in development of variety of cancers including prostate cancer. In this study we have investigated the precise interaction between SChLAP1 and SWI/SNF complex components, furthermore studied the effects of SChLAP1 on genome-wide nucleosome occupancy and H3K27Me3 in metastatic prostate cancer. Finally, we have determined the potential clinical utility of targeting SChLAP1, SChLAP1-SWI/SNF interaction and/or downstream effectors as a treatment modality in metastatic prostate cancer, and found that co-targeting Androgen Receptor (AR) and Enhancer of zeste homolog 2 (EZH2) as a promising strategy for treatment of CRPC.

15. SUBJECT TERMS
Prostate Cancer, Long-coding RNA, SChLAP1, SWI/SNF, EZH2, ASO

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	99	USAMRMC

Table of Contents

1. Introduction

Prostate cancer is one of the leading causes of cancer-related deaths and the most commonly diagnosed cancer in the United States among men [6]. Most prostate cancer-related deaths are due to advanced disease. Identifying novel biomarkers for advanced prostate cancer and uncovering the molecular mechanism of disease progression will significantly benefit prostate cancer patients. Our previous studies defined *SChLAP1* (Second Chromosome Locus Associated with Prostate-1) as a prognostic biomarker of advanced prostate cancer, and also showed that *SChLAP1* promotes prostate cancer migration and invasion by interacting with the SWI/SNF complex. However, the exact mechanism of how SChLAP1 contributes to the metastasis of prostate cancer is still unclear. Thus, we hypothesize that elucidating ***the mechanism of SChLAP1-mediated abrogation of SWI/SNF function in prostate cancer progression will extend our knowledge of prostate cancer biology and, more importantly, reveal novel therapeutic targets against SChLAP1 and/or its downstream factors.***

2. KEY WORDS

LincRNA, SChLAP1, SWI/SNF, EZH2, ASO

3. ACCOMPLISHMENTS

3.1 What were the major goals of the project?

The following three aims were proposed.

Specific Aim-1: To investigate the precise interaction between *SChLAP1* and components of the SWI/SNF complex.

Specific Aim-2: To examine the effects of *SChLAP1* on genome-wide nucleosome occupancy and H3K27Me3 in metastatic prostate cancer.

Specific Aim-3: To determine the potential clinical utility of targeting *SChLAP1*, *SChLAP1*-SWI/SNF interaction, and downstream histone modification as a treatment modality in metastatic prostate cancer.

3.2 What was accomplished under these goals?

3.2.1 Precisely map the interaction between SChLAP1 and the SWI/SNF complex

In Specific Aim-1, we proposed to perform a Protein Interaction Profile sequencing (PIP-seq) in SChLAP1-expressing prostate cancer cells to precisely determine the proteinbinding sites of SChLAP1. Furthermore, PCR and an RNA-pull down experiment will be performed to validate the key regions of SChLAP1 that are bound by SWI/SNF proteins.

Major activities

1. We performed the PIP-seq in LNCaP cells to unbiasedly characterize the protein binding sites of SChLAP1 transcripts.
2. We constructed the SChLAP1 deletion constructs, tiling every 250bp, and used them to validate the interaction between SChLAP1 and the SWI/SNF complex.
3. We performed single molecular Fluorescence in situ hybridization (smFISH) to evaluate the subcellular localization of SChLAP1, co-localization of SChLAP1, and the SWI/SNF complex in prostate cancer cells

Significant Results

1. Perform PIP-seq to identify and validate potential protein binding sites of SChLAP1.

PIP-Seq is a recently developed approach to profile RNA binding protein (RBP) interaction sites on both unprocessed and mature RNA species. In PIP-seq, both RNase-sensitive and RNase-insensitive fragments are isolated and processed separately to differentiate which sequences are actually bound to RBPs and which are just insensitive to RNases. To ascertain the protein binding site of SChLAP1, we performed PIP-seq in LNCaP cells to unbiasedly characterize all potential protein binding sites of SChLAP1 transcripts. Crossing the entire *SChLAP1* transcript, as shown in **Figure 1A**, we foundmultiple, potential RBP binding sites on SChLAP1. Among those potential RBP binding sites, we found Exon 5 of *SChLAP1* to be one of the most highly protected exonic regions. This suggests that Exon 5 of SChLAP1 might be the critical protein binding region. Interestingly, we also found one highly protected intronic region, which suggests the nascent SChLAP1 RNA transcript might also be bound by a protein. To further confirm our findings, we employed qPCR in LNCaP and VCaP cells to validate the potential exonic and intronic protein binding (protecting) sites of *SChLAP1*. RNase I specifically digests single strand RNA, and RNase III specifically digests double-stranded RNA degradation. As shown in **Figure 1B**, for the

Exon 5 region, formaldehyde-fixed LNCaP and VCaP RNA could only be digested by RNase I, not RNase III, suggesting the protected Exon 5 is not due to double-stranded RNA formation, but potentially to protein-RNA interactions. Similar to the Exon 5 region, the protected intronic region of SChLAP1 was alsoconfirmed by qPCR (**Figure 1C**). Thus, we concluded that the SChLAP1 Exon 5 region might be the most important SChLAP1-protein interaction region.

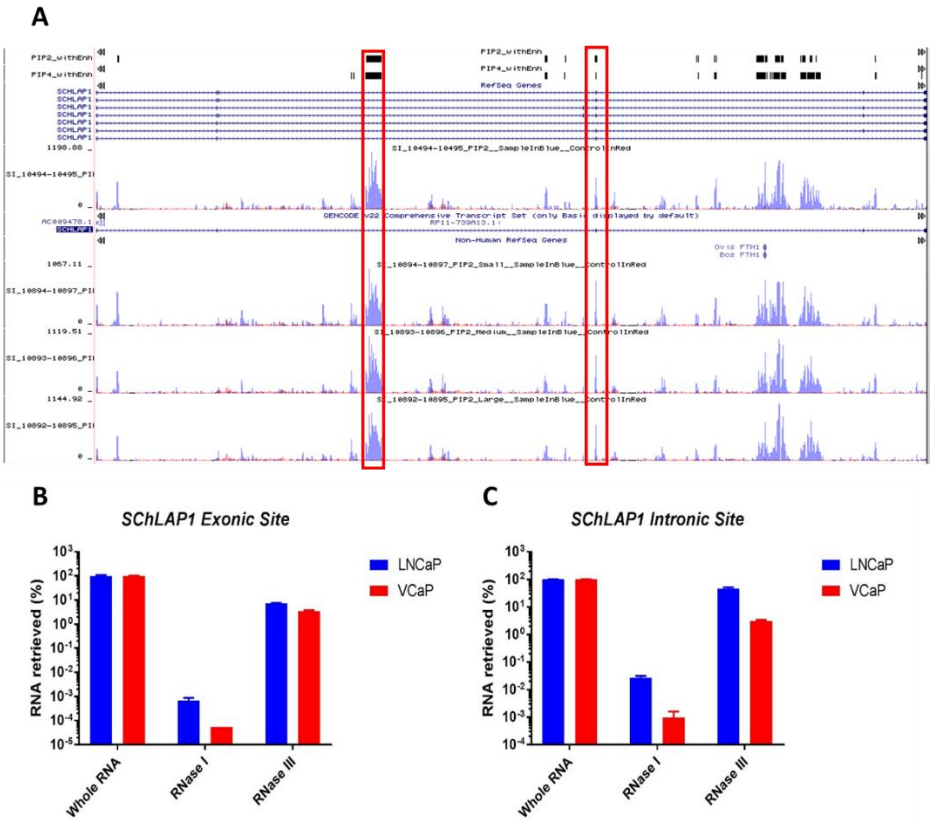


Figure 1. **Perform PIP-seq to identify and validate potential protein binding sites of SChLAP1.** (A) Genome browser view of PIP-seq result of SChLAP1 transcripts. Intronic and exonic protected regions were highlighted in red. (B) and (C) RNA of LNCaP and VCaP was treated by Rnase I and Rnase III. Levels of SChLAP1 exonic and intronic protected sites were quantified by qRT-PCR. Data are represented as normalized means +/- S.E.M.

2. Identification of key subunits of SWI/SNF complex that bind to SChLAP1.

To characterize specific regions of SChLAP1 essential for its function, we generated deletion constructs tiling every 250bp and overexpressed the constructs in RWPE cells (**Figure 2A**). Deletion of a single 250bp region (deletion construct #5, bp 1001 – 1250 for SChLAP1 isoform 1), shared by all three major isoforms of the RNA,abrogated SChLAP1-mediated cell invasion in RWPE cells (**Figure 2B**). To test whether this region is important for SWI/SNF binding, we performed RNA immunoprecipitation (RIP) for SNF5, a core subunit of the SWI/SNF complex, in RWPE cells overexpressing SChLAP1 isoform 1,

SChLAP1 isoform 2, and SChLAP1 deletion construct #5, which failed to induce cell invasion. We observed that overexpression of both SChLAP1 isoform 1 and 2 robustly bound to SNF5, whereas deletion construct #5 failed to bind to SNF5 (**Figure 2C**). *In silico* modeling with RNAfold of the *SChLAP1* RNA structure suggested the presence of a RNA hairpin in this region that is lost specifically in the deletion construct #5 (**Figure 2D**), potentially implicating this secondary structure in the function of the molecule.

The SWI/SNF complex is a multi-subunit epigenetic modifier that regulates gene expression by reorganizing nucleosomes to alter chromatin architecture. Each individual SWI/SNF complex contains either BRG1 (also known as SMARCA4) or BRM (also known as SMARCA2) as its enzymatic component, but not both. While redundancy between these ATPase subunits does exist, there are well-described functional differences between BRG1- and BRM-containing SWI/SNF complexes, and this remains an active area of research investigation. To determine whether *SChLAP1* binds specifically to either type of complex, we performed RIP for BRG1 and BRM in 22Rv1 and LNCaP prostate cancer cells. We found that endogenous *SChLAP1*, but not other prostate-cancer associated lncRNAs, robustly co-immunoprecipitated with BRG1 but not BRM (**Figure 3A**). To ensure the BRM antibody is able to retrieve intact SWI/SNF complexes, we performed immunoprecipitation followed by immunoblotting for a core component of SWI/SNF (**Figure 3B**). To extend this analysis further, we assessed the impact of *SChLAP1* expression on BRG1 and BRM genomic binding. Using ChIP-PCR for four target genes of SNF5, we found that *SChLAP1* overexpression in RWPE cells preferentially decreased BRG1 binding from SNF5 target promoters, whereas the effect on BRM was markedly more mild (**Figure 3C**). Taken together, these results suggest that SWI/SNF complexes utilizing BRG1 as the enzymatic subunit are the primary target for *SChLAP1*.

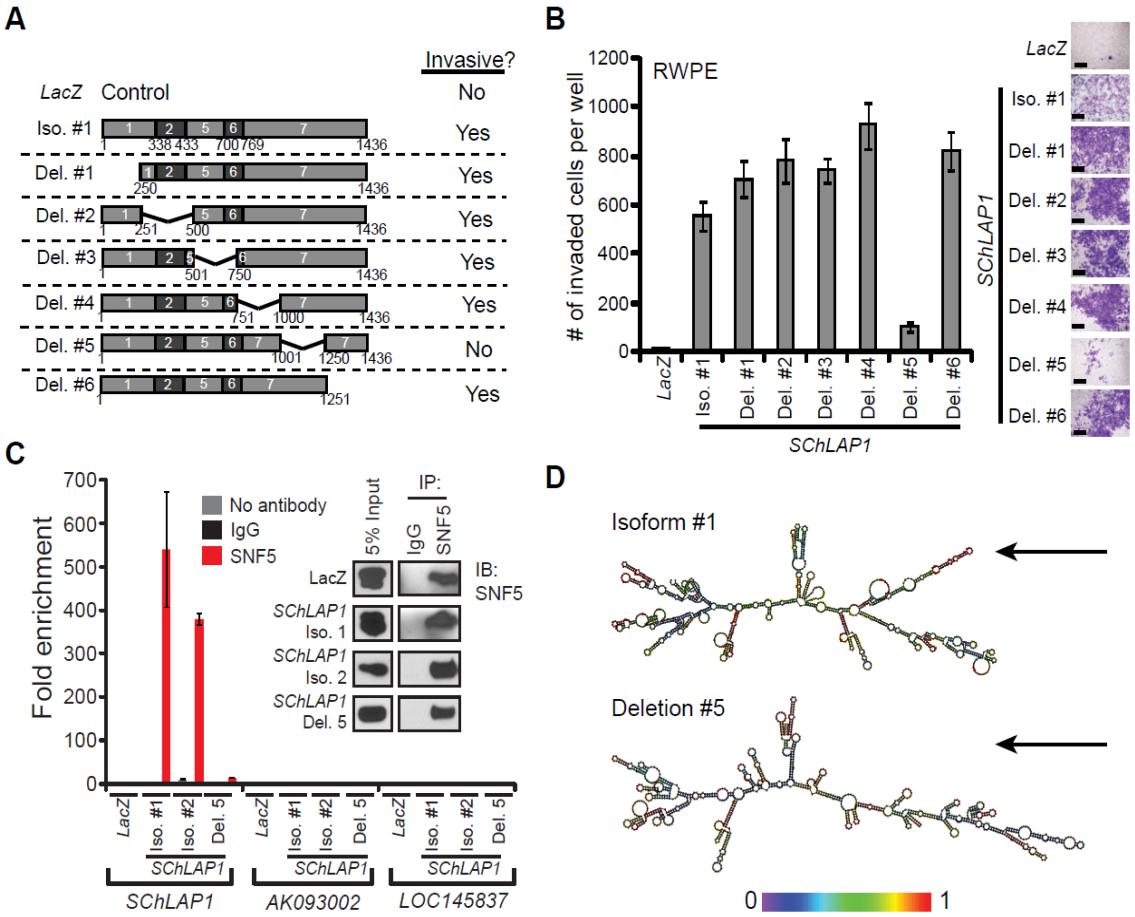


Figure 2. Identification of key subunits of SWI/SNF complex that bind to SChLAP1. A 250bp region is necessary for *SChLAP1* function. (A) Schematic of deletion constructs made for *SChLAP1* and their impact on cell invasion. (B) Deletion constructs of *SChLAP1* were overexpressed in RWPE cells, and the resulting cells were assayed for invasion in a Boyden chamber assay. Data are represented as normalized means +/- S.E.M. Images to the right show representative Boyden chamber membranes following invasion. All images were captured at the same magnification. (C) RIP analysis of SNF5 in RWPE cells overexpressing *LacZ*, *SChLAP1* isoform 1, *SChLAP1* isoform 2, or *SChLAP1* deletion construct #5. Inset protein blots showing pulldown efficiency. *AK093002* and *LOC145837* serve as negative controls. Data are represented as means +/- S.E.M. (d) *In silico* structural predictions of *SChLAP1* isoform 1 and deletion construct #5 by RNAfold. Arrows identify the structural hairpin lost in *SChLAP1* deletion construct #5.

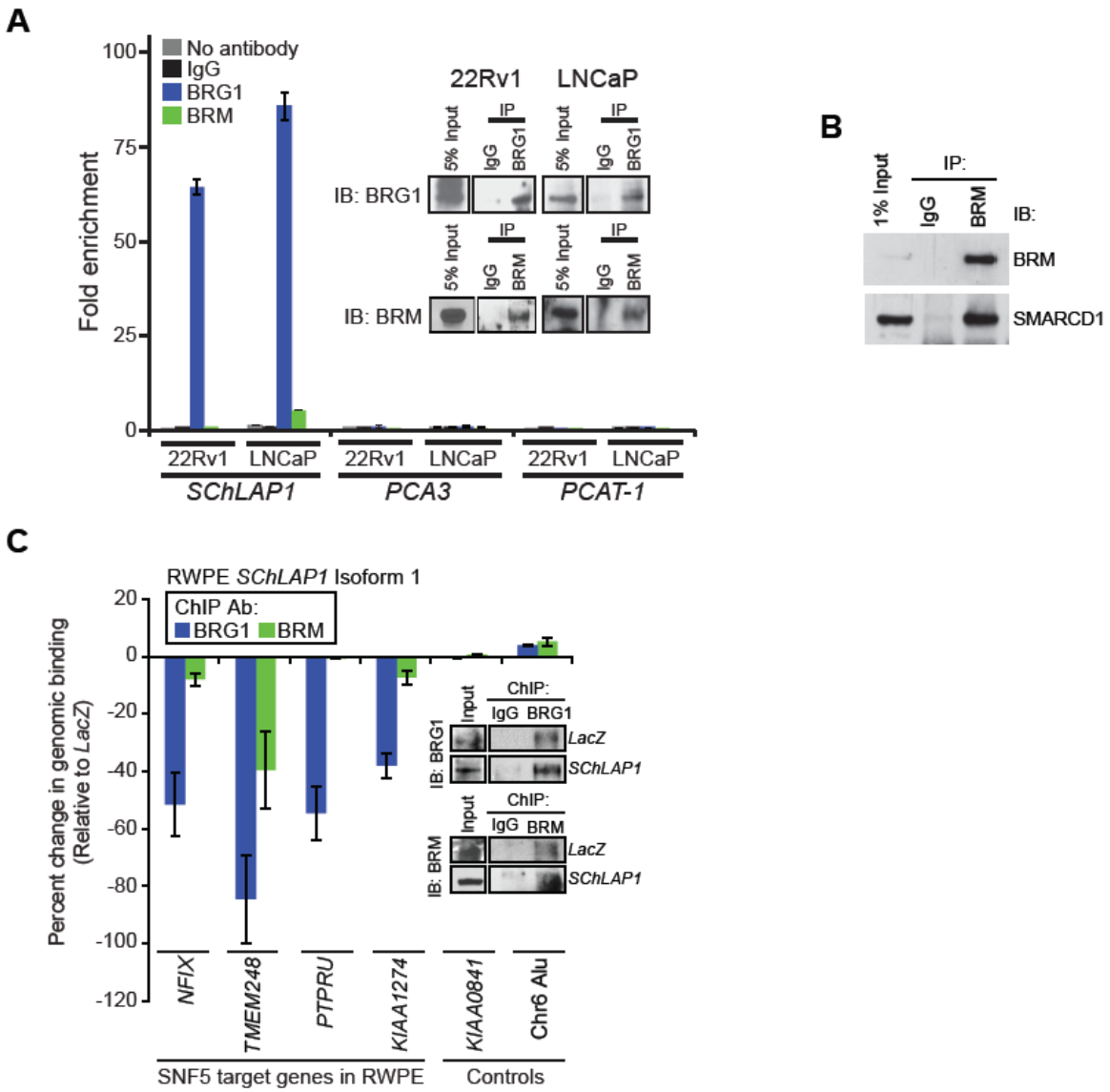


Figure 3: *SChLAP1* interacts with BRG1, exposing BRM as a therapeutic target. (A) RIP of BRG1 or BRM in 22Rv1 and LNCaP cells. Inset protein blots showing pulldown efficiency. PCA3 and PCAT-1 serve as negative controls. Data are represented as means +/- S.E.M. (B) Immunoprecipitation for BRM followed by immunoblotting in LNCaP cells. (C) ChIP for BRG1 and BRM in RWPE-SChLAP1 and RWPE-LacZ cells. ChIP-PCR for four target genes of SNF5. KIAA0841, and Chr6 Alu serve as negative controls. Data are represented as percentage changes in genomic binding relative to LacZ after being normalized to IgG controls. The inset western blot indicates immunoprecipitation efficiency for BRG1 and BRM.

3. Precise characterization of the subcellular localization of SChLAP1 and the co-localization of SChLAP1 and SWI/SNF complex in prostate cancer cells

To precisely characterize the subcellular localization of SChLAP1 in prostate cancer, we performed single molecular Fluorescence in situ hybridization (smFISH) in LNCaP cells. Compared to conventional Fluorescence in situ hybridization (FISH), the smFISH assay is an absolute quantitative method for RNA molecules at the single cell level. As shown in **Figure 4A**, we harnessed SChLAP1 exon specific smFISH probes (Red) and intron specific smFISH probes (Green) to the SChLAP1 smFISH assay, and found that SChLAP1 is a highly nuclear-specific transcript. By quantifying the absolute counts of each SChLAP1 molecules, we ascertained theabsolute expression of SChLAP1 in LNCaP cells (**Figure 4B**). Instead of using RWPE SChLAP1 overexpression models, we decided to further employ the smFISH approach to access the subcellular localization of SChLAP1 in metastatic prostate cancer biopsies. As shown in **Figure 4C-D**, we found that the subcellular localization pattern of SChLAP1 in prostate cancer biopsies is consistent in both prostate cancer cell lines.

After we determined the subcellular localization of SChLAP1 in prostate cancer cell lines and tissue, we assessed the co-localization of SChLAP1 with BRG or BRM in LNCaP cells. As shown in **Figure 4E**, we successfully performed the immunofluorescence-FISH for SChLAP1 (Red) and BRG/BRM (Green). However, since BRG and BRM are abundant in LNCaP cells, it is difficult to claim that SChLAP1 is co-localized with the SWI/SNF complex. Taken together, the results of smFISH experiments suggest that SChLAP1 localizes at a few specific nuclear loci in prostate cancer cell lines and metastatic prostate

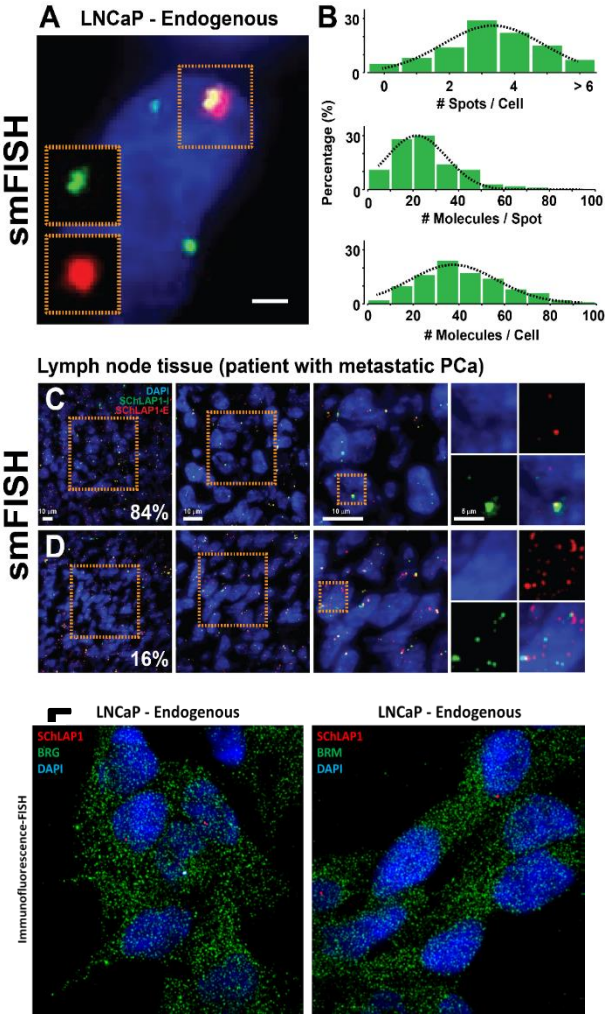


Figure 4 Localization of SChLAP1 by smFISH. (A) smFISH for endogenous SChLAP1 (E-exons, I-introns) in LNCap cells. (B) Quantification of SChLAP1 expression. (C) and (D), smFISH for endogenous SChLAP1 (E-exons, I-introns) in tissue (metastatic site). Percentage depicts fraction of samples with <6 spots (C) or > 6 spots (D) per cells. (E) Co-localization of SChLAP1 and BRG or BRM complex in LNCaP cells by immunofluorescence-FISH.

cancer biopsies; the interaction between SChLAP1 and the SWI/SNF complex might change the nucleosome positions around these loci.

Conclusion:

1. Throughthe PIP-seq experiment, we systematically profiled the potential protein binding region of SChLAP1 in prostate cancer cells. We found that the exon 5 region of SChLAP1 is the most protected region, indicating that exon 5 may be the most important RNA-protein interaction region.
2. By generating deletion constructs tiling every 250bp of SChLAP1, we further confirmed that the deletion of 250 bp from 1001 – 1250 of SChLAP1 abrogated SChLAP1-mediated cell invasion in RWPE cells;RIP experiments confirmed that overexpression of both SChLAP1 isoform 1 and 2 robustly bound to SNF5, whereas deletion construct #5 failed to bind to SNF5. Furthermore, we ascertained that the SWI/SNF complexes utilizing BRG1 as the enzymatic subunit are primary targets of SChLAP1.
3. By using smFISH, we determined the unique nuclear localization pattern of SChLAP1 in prostate cancer cells as well as metastatic prostate cancer biopsies, which might shed light on the development of clinical tests for SChLAP1 in prostate cancer. In addition, due to the unique nuclear localization pattern of SChLAP1, we cannot harness the immunofluorescence-FISH approach to examine the co-localization of SChLAP1 and BRG.

3.2.2 To examine the effects of SChLAP1 on genome-wide nucleosome occupancy and H3K27Me3 in metastatic prostate cancer

In Specific Aim-2, we proposed to perform MNase digestion followed by high-throughput sequencing (MNase-seq) to profile global nucleosome position changes on SChLAP1 overexpression in prostate cells. We then planned to harness Chromatin immunoprecipitation (ChIP) to check which H3K27Me3 changes around the region had SChLAP1-mediated nucleosome position changes.

Major activities

1. In SChLAP1-overexpressing RWPE cells, we performed the MNase-seq experiments to profile the nucleosome occupancy changes.
2. Integrative analysis was performed to annotate the SChLAP1-mediated nucleosome occupancy changed loci and significant relevant signature.
3. Series of ChIP experiments were performed to examine the effects of *SChLAP1* on H3K27Me3 levels and genome-wide binding sites of the PRC2 subunits in *SChLAP1*-expressing prostate cells.

Significant Results

1. Profiling of SChLAP1-mediated global nucleosome occupancy changes in SChLAP1-expressing prostate cells.

MNase-seq is a next-generation sequencing approach used to profile global nucleosome occupancy status. Essentially, nucleosome-associated DNA is relatively insensitive to digestion by micrococcal nuclease (MNase). Upon mild MNase treatment, the undigested nucleosomal DNA can be purified and sequenced allowing a precise localization of *in vivo* nucleosomes at a genome-wide level. To profile the SChLAP1-mediated nucleosome position changes of surrounding transcriptional start sites (TSS), we stably overexpressed two SChLAP1 isoforms in RWPE cells and performed MNase-seq in SChLAP1-overexpressing cells and parental cells. As shown in **Figure 5A**, the nucleosome positioning of all TSS was not significantly changed by SChLAP1. However, we divided genes based on their expression levels and found that the TSS of highly expressed transcripts contained nucleosome positioning changes downstream of TSS (**Figure 5B, 5C and 5D**). This data suggests that SChLAP1 overexpression does not

have a global effect on nucleosome occupancy, but the nucleosome position around certain genes might be affected by SChLAP1 overexpression in prostate cells.

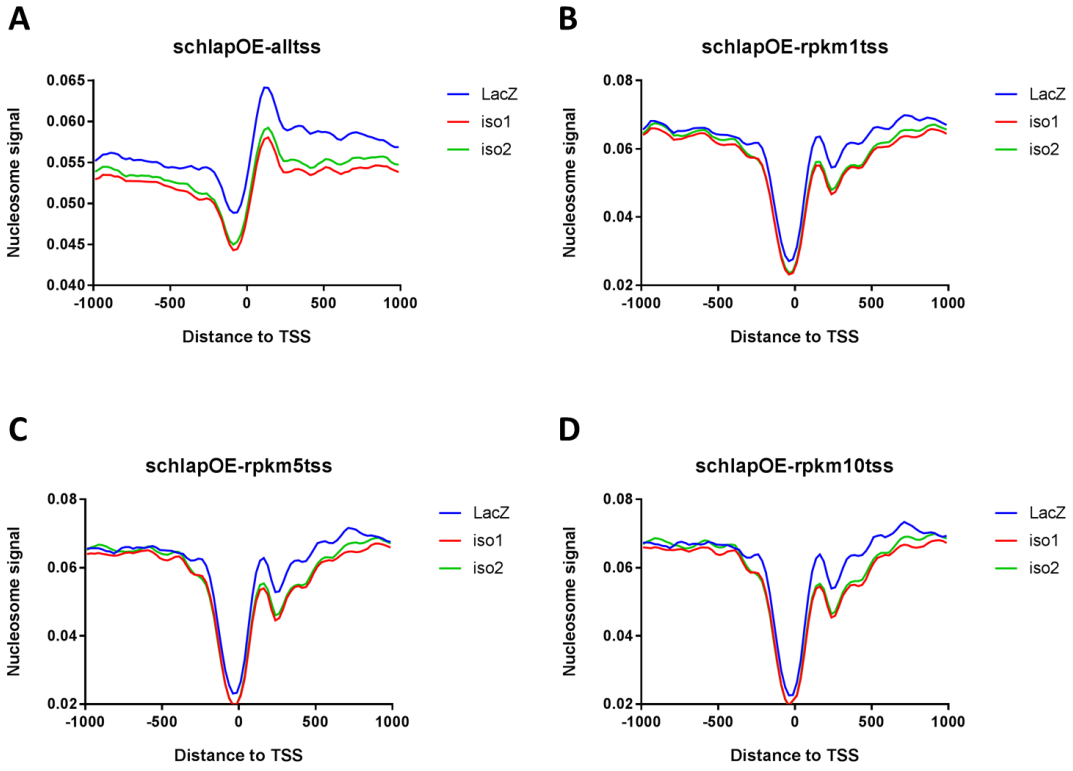


Figure 5. **Profiling of SChLAP1-mediated global nucleosome occupancy changes in SChLAP1 expressing prostate cells.** TSS-aligned overlay of nucleosome occupancy of RWPE-LacZ, RWPE-SChLAP1 iso1, and RWPE-SChLAP1 iso2 cells. Genome-wide nucleosome occupancy was determined by MNase-seq. (A) TSS-aligned overlay of nucleosome occupancy of all TSS. (B) TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=1). (C) TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=5). (D) TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=10).

Integrative analysis of SChLAP1 overexpression-mediated nucleosome and transcriptomic changes.

To further functionally annotate the SChLAP1-mediated nucleosome position changes, genes with nucleosome occupancy changes were divided into six different groups

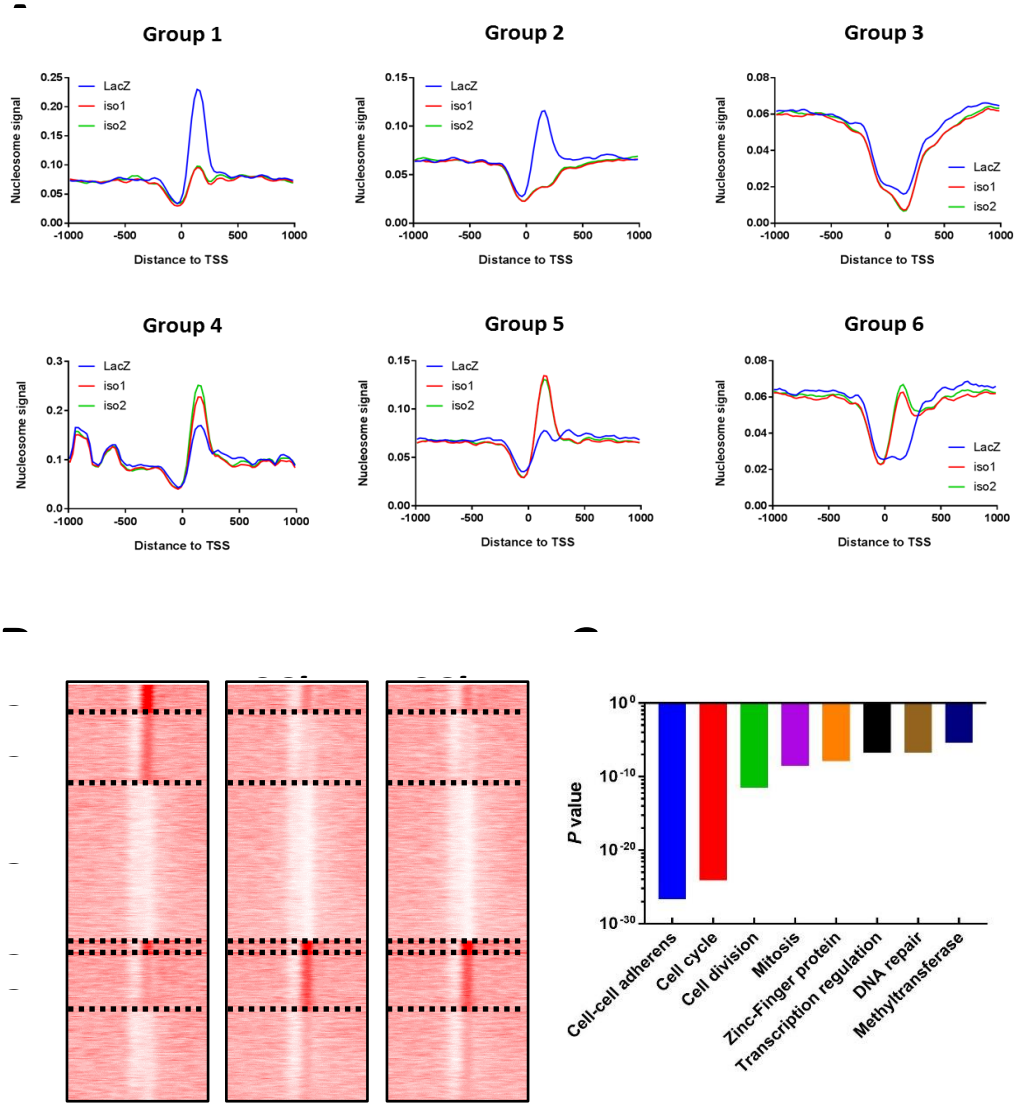


Figure 6. **Integrative analysis of SChLAP1 overexpression-mediated nucleosome and transcriptomic changes.** TSS-aligned overlay of nucleosome occupancy of RWPE-LacZ, RWPE-SChLAP1 iso1, and RWPE-SChLAP1 iso2 cells. Genome-wide nucleosome occupancy was determined by MNase-seq. (A) TSS-aligned overlay of nucleosome occupancy of all TSS. (B) TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=1). C, TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=5). D, TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=10).

based on their SChLAP1-mediated TSS nucleosome positioning patterns (**Figure 6A and 6B**). For genes belonging to groups 1, 2 and 3, the TSS of the genes contained a nucleosome-depleted region (NDR) at the 3' end of TSS; for genes belonging to groups 4, 5 and 6, the TSS of the genes contained nucleosome-gain region (NGR) at the 3' end of TSS. Functional annotation analysis was performed for

genes containing NDR or NGR, and the most significant relevant signature was shown in Figure 6C, including cell-cell adherens, cell cycle, cell division, mitosis, zinc-finger proteins, transcription regulation, DNA damage repair, and methyltransferase (**Figure 6C**).

Ascertain *SChLAP1*-mediated H3K27Me3 status changes on SChLAP1-regulated genes

Based on the integrative analysis of MNase-seq, we identified subsets of genes with SChLAP1-mediated nucleosome occupancy changes. Functional annotation suggests that these genes mainly belong to cell-cell adherens, the cell cycle, cell division, mitosis, zinc-finger proteins, transcription regulation, DNA damage repair, and methyltransferase. Our preliminary data found that overexpression of SChLAP1 in RWPE cells and knockdown of SNF5 in LNCaP cells can either increase or decrease H3K27Me3 levels, which suggests that the H3K27Me3 status and polycomb repressive complex 2 (PRC2) proteins might be involved in SChLAP1-mediated nucleosome occupancy changes in prostate cancer cells. To further ascertain the SChLAP1-mediated H3K27Me3 status and PRC2 subunits binding changes on SChLAP1-regulated genes, we selected twenty-four genes from eight different groups, and examined the H3K27Me3 status by ChIP-qPCR in SChLAP1 expressing RWPE cells and SChLAP1 knockdown in LNCaP cells. As shown in Figure 7, across all twenty-four genes, we did not find any significant changes on H3K27Me3 status and the SUZ12 binding status by either SChLAP1 overexpression or knockdown. Taken together, these results indicate that PRC2 and H3K27Me3 might not be directly involved in SChLAP1-mediated subsets of nucleosome occupancy changes.

Conclusions and discussion

1. By MNase-seq experiment, we unbiasedly profiled the SChLAP1-mediated genome-wide nucleosome occupancy changes in prostate cells. We found that SChLAP1 overexpression does not have global effects on nucleosome occupancy, but the nucleosome position around certain genes might be affected by SChLAP1 overexpression in prostate cells.
2. We performed integrative analysis on SChLAP1-mediated nucleosome and transcriptomic changes;the most significant signatures were shown in Figure 6C, including cell-cell adherens, the cell cycle, cell division, mitosis, zinc-finger proteins, transcription regulation, DNA damage repair, and methyltransferase, suggesting the SChLAP1-mediated nucleosome occupancy changes might contribute to the biological/oncogenic function of SChLAP1 in prostate cancer cells.
3. We further ascertained that PRC2 and H3K27Me3 are not involved in SChLAP1-mediated nucleosome occupancy changes; further studies might be needed to further investigate the mechanism of how long-non coding RNA mediates nucleosome occupancy changes.

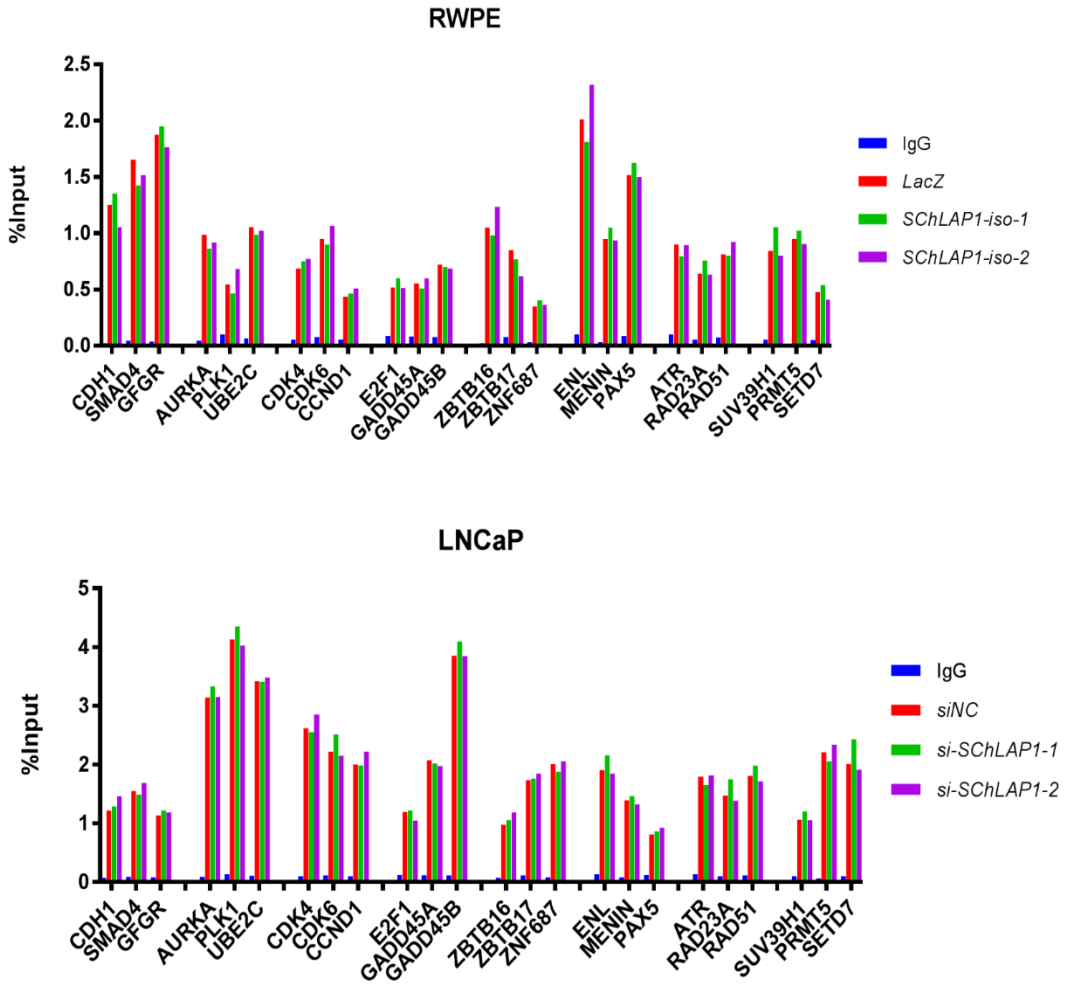


Figure 7. ChIP-qPCR validation of H3K27Me3 status at potential SChLAP1 downstream target genes' TSS. (A) ChIP-qPCR analysis of H3K27Me3 status at twenty-four different potential SChLAP1 target genes' TSS in RWPE-LacZ, RWPE-SChLAP1-iso1, and RWPE-SChLAP1-iso2 cells. Data are represented as normalized means +/- S.E.M. (B) ChIP-qPCR analysis of H3K27Me3

3.2.3 To determine the potential utility of targeting *SChLAP1*, *SChLAP1*-SWI/SNF interaction, and downstream histone modification as a clinical treatment modality in metastatic prostate cancer

In Specific Aim-3, we proposed to systematically determine the potential utility of targeting SChLAP1, SChLAP1-SWI/SNF interaction, and downstream histone modification as a clinical treatment modality in metastatic prostate cancer by using a specific antisense oligo targeting SChLAP1 and a specific inhibitor targeting EZH2, which further dismisses the global H3K27Me3.

Major activities

1. Determined the potential clinical utility of SChLAP1-ASOs *in vitro* models of prostate cancer.
2. Ascertained the therapeutic value of targeting BRM in prostate cancer cells.
3. Determined the potential clinical utility of an EZH2 inhibitor (EPZ-6438) in SChLAP1-expressing prostate cancer models.

The potential clinical utility of SChLAP1-ASOs *in vitro* models of prostate cancer.

By collaborating with Ionis Pharmaceuticals, we developed several SChLAP1-targeting antisense oligos (ASOs). As shown in **Figure 7A**, we systematically tested the knockdown efficiency of SChLAP1-targeting ASOs in different SChLAP1-expressing prostate cancer cell lines. Then we performed *in vitro* proliferation and invasion assays to evaluate the effects of SChLAP1-ASOs on LNCaP, C4-2B, and 22RV1 cells. As shown in **Figure 7C-D**, the knockdown of SChLAP1 didn not significantly inhibit cell proliferation (**Figure 7C**) but slightly decreased the invasion in LNCaP and 22RV1 cells (**Figure 7D**). Since the in vitro efficacy of SChLAP1-ASOs is not compelling, we did not further test them in *in vivo* models.

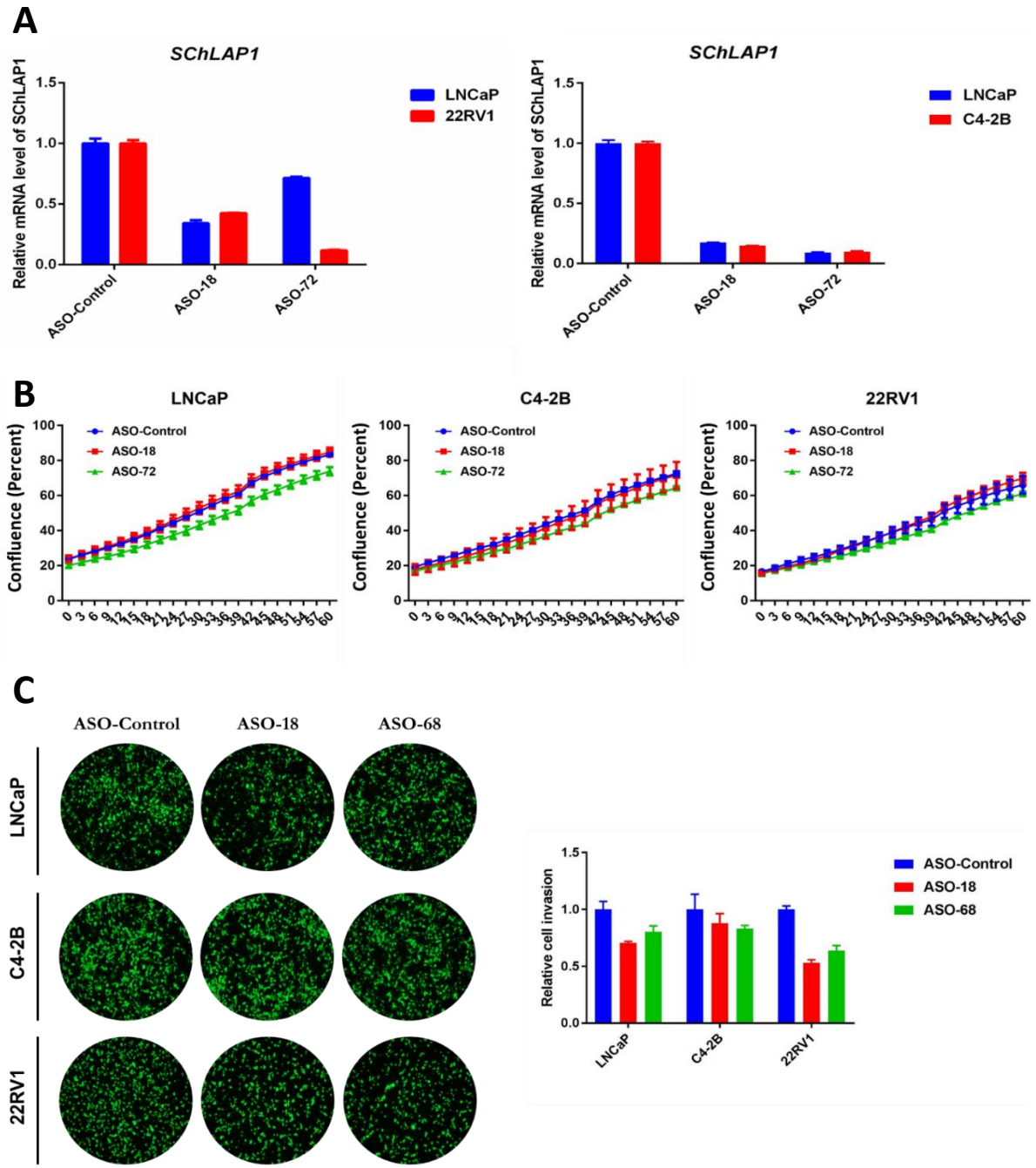


Figure 8. Validation of SChLAP1-ASO knockdown efficiency in multiple prostate cancer cell lines. (A) Left, SChLAP1-ASOs were delivered via transfection (20 nM) into LNCaP and 22RV1 cells; levels of SChLAP1 were quantified by qRT-PCR. Right, SChLAP1-ASOs were delivered via free uptake (2.5 mM) into LNCaP and 22RV1 cells; levels of SChLAP1 were quantified by qRT-PCR. **(B)** Proliferation of LNCaP, C4-2B, and 22RV1 cells after SChLAP1 knockdown by ASOs. **(C)** Invasiveness of LNCaP, C4-2B, and 22RV1 cells after SChLAP1 knockdown by ASOs.

Ascertain the therapeutic value of targeting BRM in prostate cancer cells.

Given *SChLAP1*'s specific interaction with BRG1, we hypothesized that BRM may be a similar synthetic lethal target in *SChLAP1*-expressing cells. To test this hypothesis, we performed a siRNA knockdown of BRM in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells (**Figure 9A**). We found that BRM knockdown slightly increased cell invasion in RWPE-*LacZ* cells but significantly decreased cell invasion in RWPE-*SChLAP1* cells (**Figure 9B**). Additionally, while BRM knockdown slightly increased cell proliferation in RWPE-*LacZ* cells (consistent with previous studies identifying BRM as a proliferative gatekeeper in prostate cells), BRM knockdown significantly decreased cell proliferation in RWPE-*SChLAP1* cells (**Figure 9C**). Taken together, these results suggest that BRM can be targeted in *SChLAP1*-expressing cells to decrease cell viability.

To explore the mechanistic basis of this synthetic lethality, we sought to determine whether targeting BRM in *SChLAP1*-expressing cells destabilizes residual SWI/SNF complexes, as previously suggested in BRG1-mutant cell lines. While BRM knockdown in RWPE-*LacZ* cells affected the expression of some subunits and their incorporation into the SWI/SNF complex, BRM knockdown in RWPE-*SChLAP1* cells reduced the expression and incorporation of additional components (**Figure 9D**).

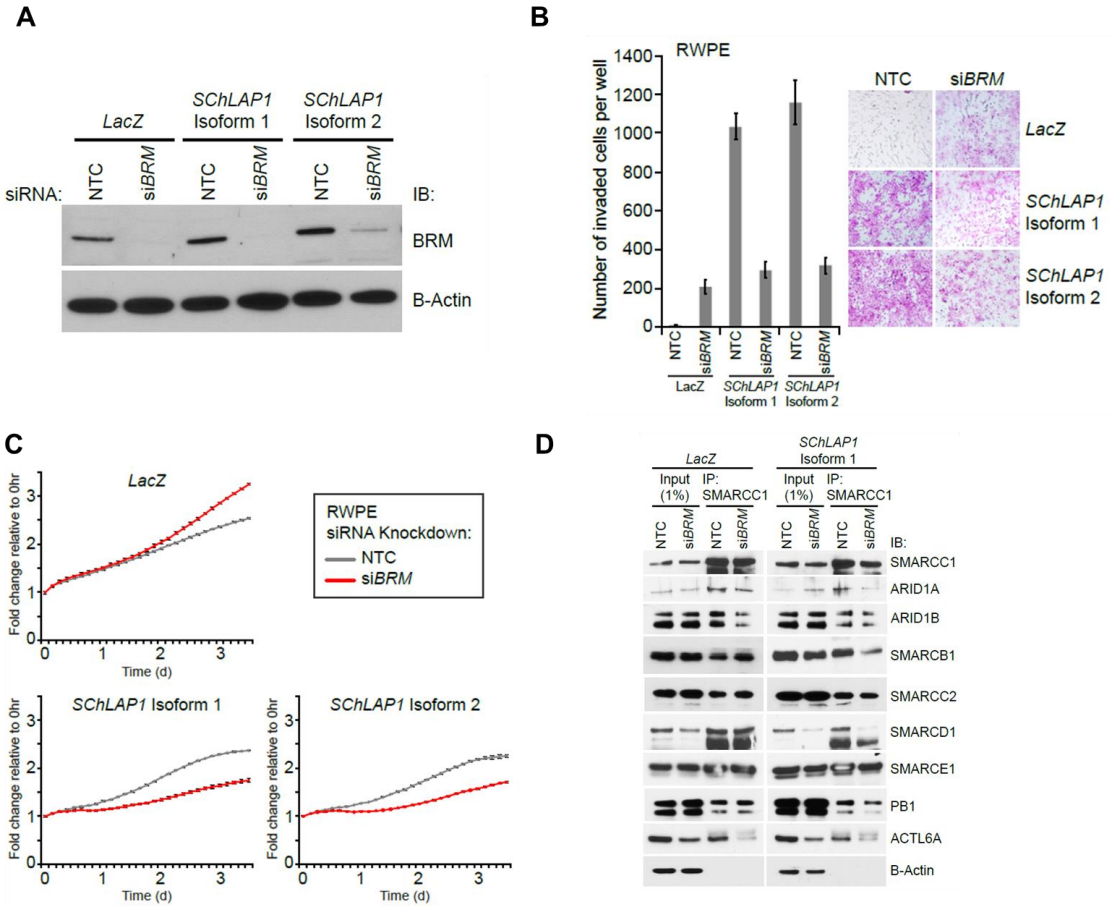


Figure 9. *SChLAP1* interacts with BRG1, exposing BRM as a therapeutic target. (A) Knockdown efficiency of BRM siRNA in RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 cells. Western blot for BRM is shown. B-Actin serves as a loading control. **(B)** Cell invasion through Matrigel in a Boyden chamber assay of RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 with or without BRM knockdown. Data are represented as normalized means +/- S.E.M. Images to the right show representative Boyden chamber membranes following invasion. All images were captured at the same magnification. **(C)** Cell proliferation of RWPE-*LacZ*, RWPE-*SChLAP1* isoform 1, and RWPE-*SChLAP1* isoform 2 with or without BRM knockdown. Cell proliferation assays were performed by Incucyte live-cell imaging. Data shown are fold changes in cell confluence vs. time at three-hour intervals. Each data point is the mean of quadruplicates +/- S.E.M. **(D)** Immunoprecipitation for SMARCC1 followed by immunoblotting in RWPE-*LacZ* and RWPE-*SChLAP1* cells treated with non-targeting control siRNA or BRM siRNA.

Determined the potential clinical utility of an EZH2 inhibitor (EPZ-6438) in SChLAP1-expressing prostate cancer models.

In our previous study, we found that SChLAP1 might act as an alternative mechanism to inactivate the tumor suppressive SWI/SNF complex function. For the cancers which harbor loss-of-function mutation or aberration of SWI/SNF, the levels of EZH2 and H3K27Me3 are usually higher than SWI/SNF wild-type cancers. Thus, we proposed to test the potential clinical utility of an EZH2 inhibitor (EPZ-6438) in SChLAP1-expressing prostate cancer models. To this end, we tested the growth inhibitory effect of the EZH2 inhibitor (EPZ-6438) in SChLAP1-expressing prostate cancer cell lines (LNCaP and 22RV1) and a SChLAP1 non-expressing prostate cancer cell line (PC3). As shown in **Figure 10A**, we found that EPZ-6438 has modest growth inhibitory effect on SChLAP1-expressing prostate cancer cells but not on non-SChLAP1 expressing PC3 cells. Furthermore, the IC50 of EPZ-6438 was examined in SChLAP1-expressing andnon-expressing prostate cancer cell lines We found that EPZ-6438 is more sensitive in SChLAP1-expressing prostate cancer cells (**Figure 10B and 10C**).

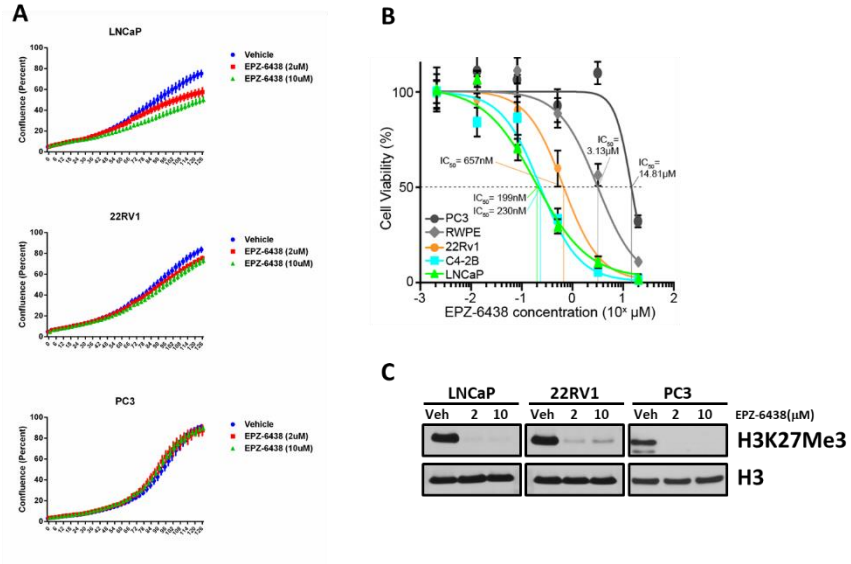


Figure 10. The anti-tumor effect of EPZ-6438 in multiple prostate cancer cell lines. (A) The growth inhibitory effect of EPZ-6438 in LNCaP, 22RV1 and PC3. (B) IC50 plot of EPZ-6438 in multiple prostate cancer cell lines. (C) Effect of EPZ-6438 on H3K27Me3 in LNCaP, 22RV1 and PC3.

Furthermore, we systematically tested the efficacy of EPZ-6438 in LNCaP-derivedC4-2B and 22RV1 xenograft models. **As shown in Figure 11A**, mice were dosed orally with either vehicle, enzalutamide or EPZ-6438 respectively. We found that EPZ-6438 efficiently decreased H3K27Me3 levels in C4-2B xenografts (**Figure 11D**), but the tumor volume and weight were not significantly affected (**Figure 11A and 11B**). In addition, we tested EPZ-6438 in 22RV1 xenografts at a higher dose. Like C4-2B xenografts, EPZ-6438 did not significantly inhibit 22RV1 xenografts growth *in vivo* (**Figure 11E, 11F and 11G**). Taken together, even though we observed a modest growth inhibitory effect *in vitro*, targeting EZH2 by EZP-6438 did not have a significantly growth inhibitory effect *in vivo*.

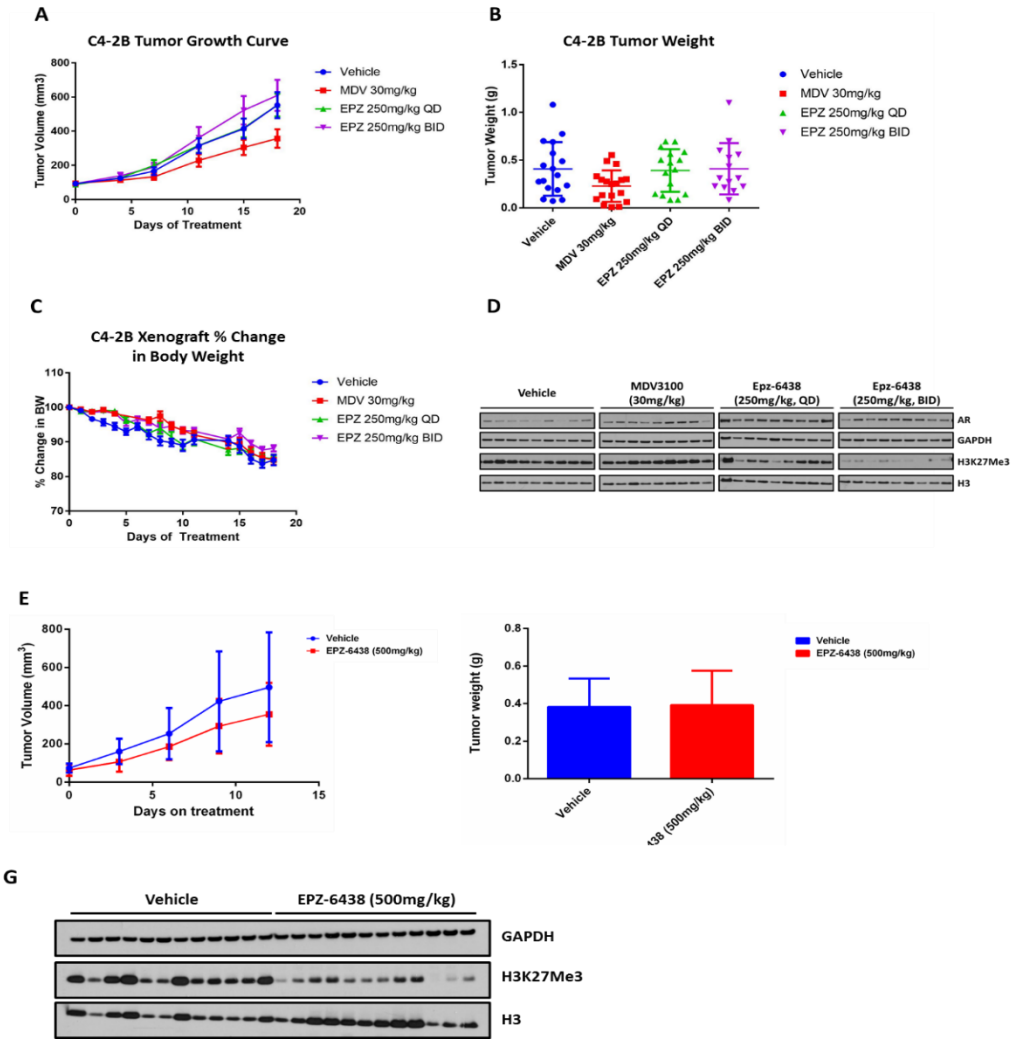


Figure 11. Tumor growth inhibitory effect of EPZ-6438 in vivo. A, B and C, Comparison of the effectsof vehicle, enzalutamide (30 mg/kg), and EPZ-6438 (250 mg/kg, QD or BID) on tumor volume (A), tumor weight (B) and animal body weight changes (C) in C4-2B xenografts. D, The effect of EPZ-6438 on H3K27Me3 in C4-2B xenografts. E and F, The effect of vehicle and EPZ-6438 (500 mg/kg) on tumor volume (E), tumor weight (F) in 22RV1 xenografts. G, The effect of EPZ-6438 on H3K27Me3 in 22RV1 xenografts.

However, in addition to what we proposed in the application, we also exploited several other therapeutic strategies around the SChLAP1 biology. Interestingly, we screened a library of epigenetic inhibitors for their ability to render SChLAP1-expressing CRPC cells sensitive to enzalutamide and found that EZH2 inhibitors specifically potentiated enzalutamide-mediated inhibition of proliferation (**Figure 10A**). Moreover, we identified antisense oligonucleotides (ASO) as a novel drug strategy to ablate EZH2 and AR expression, which may have advantageous properties in certain settings (**Figure 10B**). Combination treatment with ASOstargeting EZH2 and AR transcripts inhibited SChLAP1-expressing prostate cancer cell growth *in vitro* and *in vivo* better than single agents (**Figure 10C-D**). In sum, we identified EZH2 as a critical epigenetic regulator of ADT resistance and defined the ASO-based, co-targeting of EZH2 and AR as a promising strategy for treatment of SChLAP1-expressing CRPC.

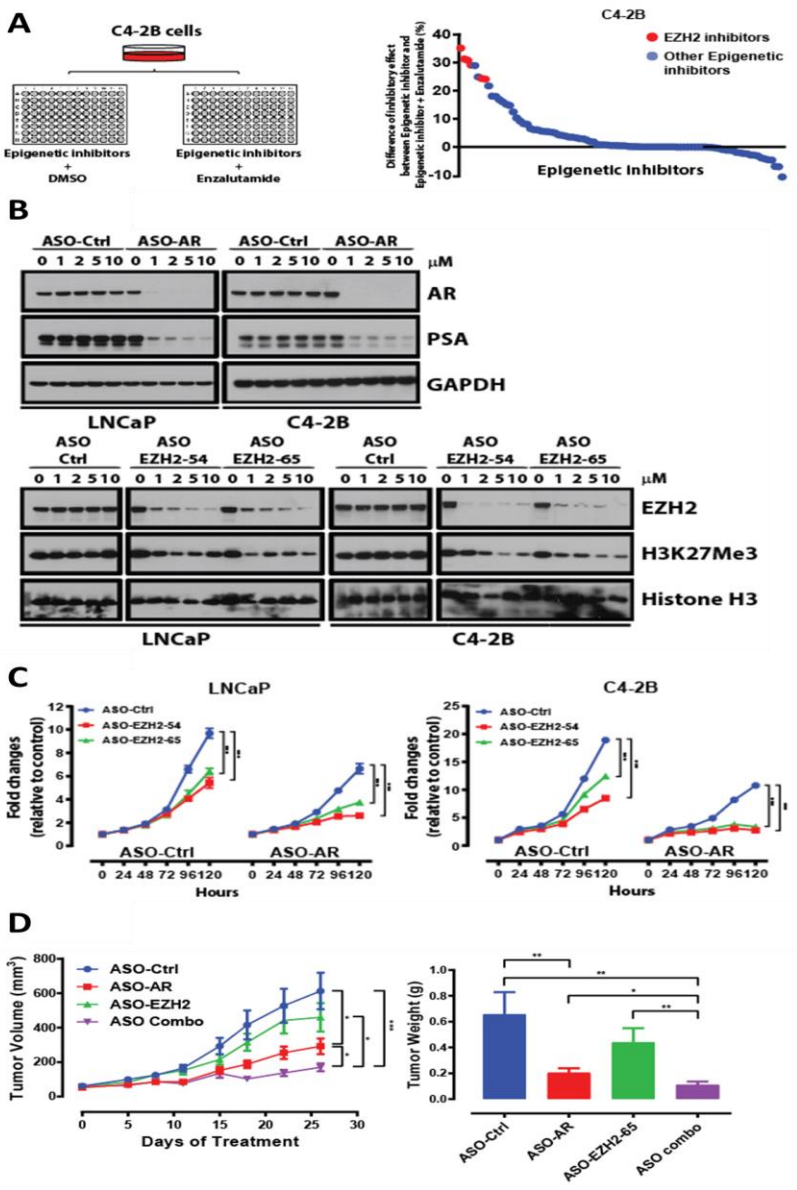


Figure 11. Identification of co-targeting AR and EZH2 as a promising strategy for treatment of SChLAP1-expressing CRPC. (A) Diagram depicting the screen performed in C4-2B cells treated with the epigenetic inhibitor panel plus enzalutamide (2 μM) or DMSO control. Degree of proliferation inhibition for each individual inhibitor (1 μM) plus enzalutamide versus each inhibitor alone is graphed (all normalized to DMSO). **(B)** Top, immunoblot analysis of AR and PSA in LNCaP and C4-2B cells treated with ASO-Ctrl or ASO-AR. ASOs were delivered to cells in all experiments via free uptake. Bottom, immunoblot analysis of EZH2 and H3K27Me3 in LNCaP and C4-2B cells treated with ASO-Ctrl and EZH2 ASOs (EZH2-54 and EZH2-65). **(C)** Growth curves of LNCaP and C4-2B cells treated with ASO-Ctrl, ASO-AR, ASO-EZH2 (EZH2-54 and EZH2-65, 1 μM), or combination of ASO-AR (1 μM) plus ASO-EZH2 (1 μM). Data represent mean ± standard error (n = 6) from one of three independent experiments. **(D)** Tumor volume time course for subcutaneous C4-2B xenografts in CB17SCID mice. Treatment with ASO-Ctrl, ASO-AR, ASO-EZH2-65, or combination of ASO-AR and ASO-EZH2-65 started at tumor volumes of 50-100 mm^3. Tumor weights from respective treatment groups at the termination of the study are also graphed.

3.2.4 Conclusion and Discussion:

1. We have successfully developed specific SChLAP1-targeting antisense oligos, and on-target efficiency has been tested in multiple prostate cancer cell lines. However, we did not observe any striking growth inhibitory and migration/invasion inhibitory effects, suggesting that, while SChLAP1 may not be an ideal target as a monotherapy strategy, it may be interesting to test the efficacy of combining SChLAP1-ASOs with other therapeutic strategies, such as androgen deprivation therapy. Because of this, we did not furthertest the efficacy of SChLAP1-ASOs in *in vivo* models.

2. We have shown that a 250bp region in exon 5 of SChLAP1 mediates its invasive phenotype and coordinates its interaction with the SWI/SNF complex. Additionally, we found that SChLAP1 interacts with BRG1-, but not BRM-, containing SWI/SNF complexes and preferentially decreases BRG1 genomic binding. Finally, SChLAP1's preference for BRG1 may expose BRM as a therapeutic target in prostate cancer. Thus, while mutations in SWI/SNF subunits have been used to reveal specific vulnerabilities in cancer, our work suggests that lncRNA-mediated BRG1-inactivation may uncover similar therapeutic opportunities in malignancy. Taken together, our findings have broad implications for cancer biology and provide additional evidence for the development and use of specific BRM inhibitors for the therapeutic treatment of cancer.

3. We have shown the growth inhibitory effect of the EZH2-specific inhibitor, EPZ-6438, in multiple SChLAP1-expressing and non-expressing prostate cancer cell lines It was foundthat EPZ-6438 has a modest growth inhibitory effect in SChLAP1-expressing prostate cancer cells *in vitro*, but has no significant growth inhibitory effect *in vivo*, suggesting that EZH2 may not be a promising target as a monotherapy strategy. More interestingly, instead of targeting SChLAP1 and EZH2, SChLAP1-experessing cells tend to be more sensitive to a co-targeting EZH2 and AR strategy. Here, we have shown that a combination treatment, with ASOtargeting EZH2 and AR transcripts, inhibited SChLAP1-expressing prostate cancer cell growth *in vitro* and *in vivo* better than single agents. In sum, we identified EZH2 as a critical epigenetic regulator of ADT resistance and defined the ASO-based, co-targeting of EZH2 and AR as a promising strategy for the treatment of SChLAP1-expressing CRPC.

3.3 What opportunities for training and professional development has the project provided?

Several opportunities for training and professional development were provided while working on this project, including:

One-on-one meetings with a mentor: Data was discussed with the mentor (Dr. Arul Chinnaiyan) during biweekly individual meetings and via monthly progress reports. Continuous input and advice were provided.

Presentations at Scientific Conferences: Data obtained was presented at scientific conferences:

1. AACR Annual Meeting, Washington, DC, 2017.
2. Prostate SPOREs, Fort Lauderdale, Florida, 2018.
3. AACR Annual Meeting, Chicago, IL, 2018.

3.4 How were the results disseminated to communities of interest?

Nothing to Report.

3.5 What do you plan to do during the next reporting period to accomplish the goals?

Nothing to Report.

4. IMPCAT

4.1 What was the impact on the development of the principal discipline(s) of the project?

The completion of the proposed project “Biological characterization and clinical utilization of metastatic prostate cancer-associated lincRNA *SChLAP1*” has substantial impacts on understanding the biological function of SChLAP1 in prostate cancer progression. We have discovered the potential protein binding sites of SChLAP1 and determined its specific localization pattern in prostate cancer cells and biopsies. We have also shown that by interacting with the SWI/SNF complex, SChLAP1 significantly mediates nucleosome occupancy changes in prostate cells. More importantly, we have identified EZH2 as a critical epigenetic regulator of ADT resistance and defined the ASO-based, co-targeting of EZH2 and AR as a promising strategy for the treatment of SChLAP1-expressing CRPC. By completing this project, we believe that it will advance our knowledge on the critical role of lincRNA in metastatic prostate cancer and offer potential translation opportunities for future progressive prostate cancer treatment.

4.2 What was the impact on other disciplines?

Nothing to Report.

4.3 What was the impact on technology transfer?

Nothing to Report.

4.4 What was the impact on society beyond science and technology?

Nothing to Report.

5. CHANGES

5.1 Changes in approach and reasons for change

Nothing to Report.

5.2 Actual or anticipated problems or delays and actions or plans to resolve them

Nothing to Report.

5.3 Changes that had a significant impact on expenditures

Nothing to Report.

5.4 Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to Report.

6. PRODUCTS

6.1 Publications, conference papers, and presentations

1. **Xiao L**, Tien JC, Vo J, Tan M, Parolia A, Zhang Y, Wang L, Qiao Y, Shukla S, Wang X, Zheng H, Su F, Jing X, Luo E, Delekta A, Juckette KM, Xu A, Cao X, Alva AS, Kim Y, Macleod AR, Chinnaiyan AM **(*equal contribution)**: Epigenetic reprogramming with antisense oligonucleotides enhances effectiveness of androgen receptor inhibition in castration-resistant prostate cancer. *Cancer Res*. 2018.
2. Zhang Y, Pitchiaya S, Cieślik M, Niknafs YS, Tien JC, Hosono Y, Iyer MK, Yazdani S, Subramaniam S, Shukla SK, Jiang X, Wang L, Liu TY, Uhl M, Gawronski AR, Qiao Y, **Xiao L**, Dhanasekaran SM, Juckette KM, Kunju LP, Cao X, Patel U, Batish M, Shukla GC, Paulsen MT, Ljungman M, Jiang H, Mehra R, Backofen R, Sahinalp CS, Freier SM, Watt AT, Guo S, Wei JT, Feng FY, Malik R, Chinnaiyan AM: Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. *Nat Genet*. 50(6):814-824, 2018.
3. Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, Iyer MK, Pitchiaya S, Malik R, Hosono Y, Prensner JR, Poliakov A, Singhal U, **Xiao L**, Kregel S, Siebenaler RF, Zhao SG, Uhl M, Gawronski A, Hayes DF, Pierce LJ, Cao X, Collins C, Backofen R, Sahinalp CS, Rae JM, Chinnaiyan AM, Feng FY: The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. Nat Commun. 7:12791, 2016.
4. Shukla S*, Zhang X*, Niknafs YS*, **Xiao L***, Mehra R, Cieślik M, Ross A, Schaeffer E, Malik B, Guo S, Freier SM, Bui HH, Siddiqui J, Jing X, Cao X, Dhanasekaran SM, Feng FY, Chinnaiyan AM, Malik R **(*equal contribution)**: Identification and validation of PCAT14 as prognostic biomarker in prostate cancer. *Neoplasia*. 18(8):489-99, 2016.
5. **Xiao L***, Shukla S*****, Zhang X*****, Niknafs Y*****, Malik R, Chinnaiyan A **(*equal contribution)**: Identification and validation of PCAT14 as prognostic biomarker in prostate cancer, AACR Annual Meeting, Washington, DC, 2017.
6. **Xiao L***, Tien JC*****, Vo J, Tan M, Parolia A, Zhang Y, Wang L, Qiao Y, Shukla S, Wang X, Zheng H, Su F, Jing X, Luo E, Delekta A, Juckette KM, Xu A, Cao X, Alva A, Kim Y, MacLeod AR, Chinnaiyan AM **(*equal contribution)**: Epigenetic reprogramming with antisense oligonucleotides

enhances effectiveness of androgen receptor inhibition in castration-resistant prostate cancer, Prostate SPOREs, Fort Lauderdale, Florida, 2018.

7. **Xiao L*****, Tien JC*****, Vo J, Tan M, Parolia A, Zhang Y, Wang L, Qiao Y, Shukla S, Wang X, Zheng H, Su F, Jing X, Luo E, Delekta A, Juckette KM, Xu A, Cao X, Alva A, Kim Y, MacLeod AR, Chinnaiyan AM **(*equal contribution)**: Epigenetic reprogramming with antisense oligonucleotides enhances effectiveness of androgen receptor inhibition in castration-resistant prostate cancer. AACR Annual Meeting, Chicago, IL, 2018.
8. Parolia A*, **Xiao L***, Vo JN, Cieslik M, Cao X, Chinnaiyan AM **(*equal contribution)**: Functional CRISPR screen towards identifying novel epigenetic co-factors of oncogenic AR-activity. Prostate SPOREs, Fort Lauderdale, Florida, 2018.
9. Parolia A*, **Xiao L***, Vo JN, Cieslik M, Cao X, Chinnaiyan AM **(*equal contribution)**: Functional CRISPR screen towards identifying novel epigenetic co-factors of oncogenic AR-activity, AACR Annual Meeting, Chicago, IL, 2018.

6.2 Website(s) or other Internet site(s)

Nothing to Report.

6.3 Technologies or techniques

Nothing to Report.

6.4 Inventions, patent applications, and/or licenses

Nothing to Report.

6.5 Other Products

Nothing to Report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

7.1 What individuals have worked on the project?

Name:	*Lanbo Xiao*
Project Role:	*Research Fellow*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	10
Contribution to Project:	*Lanbo Xiao designed and performed most of the experiments.*
Funding Support:	*Department of Defense*

Name:	*Jean Tien*
Project Role:	*Research Investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2
Contribution to Project:	*Jean Tien performed all in vivo experiments.*
Funding Support:	*Prostate Cancer Foundation*

7.2 Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to Report.

7.3 What other organizations were involved as partners?

Organization Name: Ionis Pharmaceuticals

Location of Organization: Carlsbad, CA

Partner's contribution to the project: **Collaboration**

8. SPECIAL REPORTING REQUIREMENTS

Nothing to Report.

9. APPENDICES

Attached.

Epigenetic reprogramming with antisense oligonucleotides enhances the effectiveness of androgen receptor inhibition in castration-resistant prostate cancer

Lanbo Xiao[1,2,8], Jean C. Tien[1,2,8], Josh Vo[1,3], Mengyao Tan[1,2], Abhijit Parolia[1,2], Yajia Zhang[1,2], Lisha Wang[1,2], Yuanyuan Qiao[1,2], Sudhanshu Shukla[1,2,4], Xiaoju Wang[1,2], Heng Zheng[1,2], Fengyun Su[1,2], Xiaojun Jing[1,2], Esther Luo[1,2], Andrew Delekta[1,2], Kristin M. Juckette[1,2], Alice Xu[1,2], Xuhong Cao[1,2], Ajjai S Alva[1,5], Youngsoo Kim[6], A. Robert MacLeod[6], and Arul M. Chinnaiyan[1,2,3,7]

[1]Michigan Center for Translational Pathology, Ann Arbor, MI, USA
[2]Department of Pathology, University of Michigan Health System, Ann Arbor, MI, USA
[3]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[4]Department of Biosciences and Bioengineering, Indian Institute of Technology Dharwad, Dharwad, India
[5]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA
[6]Ionis Pharmaceuticals, Carlsbad, CA, USA
[7]Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI, USA
[8]These authors contributed equally

Running title: ASO-mediated inhibition of EZH2 and AR synergize in CRPC

Corresponding author: Arul M. Chinnaiyan, Michigan Center for Translational Pathology, Rogel Cancer Center, 1400 East Medical Center Drive, Ann Arbor, MI 48109. Phone: 734-615-4062; Fax: 734-615-4498; Email: arul@umich.edu

Conflict of interest statement: Y.K. and A.R.M. are employees of Ionis Pharmaceuticals, which developed the ASOs against EZH2 and AR that were used in this study.

ABSTRACT

Advanced prostate cancer initially responds to androgen deprivation therapy (ADT), but the disease inevitably recurs as castration-resistant prostate cancer (CRPC). Although CRPC initially responds to abiraterone and enzalutamide, the disease invariably becomes non-responsive to these agents. Novel approaches are required to circumvent resistance pathways and extend survival, but the mechanisms underlying resistance remain poorly defined. Our group previously showed the histone lysine-N-methyltransferase EZH2 to be overexpressed in prostate cancer and quantitatively associated with progression and poor prognosis. In this study, we screened a library of epigenetic inhibitors for their ability to render CRPC cells sensitive to enzalutamide and found that EZH2 inhibitors specifically potentiated enzalutamide-mediated

inhibition of proliferation. Moreover, we identified antisense oligonucleotides (ASO) as a novel drug strategy to ablate EZH2 and AR expression, which may have advantageous properties in certain settings. RNA-seq, ChIP-seq, and ATAC-seq demonstrated that EZH2 inhibition altered the AR cistrome to significantly upregulate AR signaling, suggesting an enhanced dependence of CRPC cells on this pathway following inhibition of EZH2. Combination treatment with ASO targeting EZH2 and AR transcripts inhibited prostate cancer cell growth in vitro and in vivo better than single agents. In sum, this study identifies EZH2 as a critical epigenetic regulator of ADT resistance and defines ASO-based co-targeting of EZH2 and AR as a promising strategy for treatment of CRPC.

Statements of significance

Simultaneous targeting of lysine methyltransferase EZH2 and the androgen receptor with antisense oligonucleotides proves a novel and effective therapeutic strategy in castration-resistant prostate cancer patients.

INTRODUCTION

Prostate cancer remains the most commonly diagnosed cancer and the second leading cause of cancer-related deaths in American men, with most mortalities resulting from advanced metastatic tumors (1). Since its development and progression depend on signals through the androgen receptor (AR), physical and pharmacological castration became therapeutic mainstays for advanced disease (2). While initially responsive to androgen deprivation therapy (ADT), these tumors almost invariably recur in an incurable androgen-independent form, called castration-resistant prostate cancer (CRPC). The pathophysiological underpinnings of CRPC are many, but frequently involve the continued activation of AR despite low levels of androgens (2). To address this phenomenon, contemporary antiandrogen therapy, typified by the second-generation antagonist enzalutamide, seeks to directly bind and target AR, blocking its

intracellular activity. Though clinical outcomes are superior to conventional castration, enzalutamide-treated CRPC tumors still uniformly recur (3,4).

The role of epigenetic modifications in CRPC has become an important area of inquiry. For example, we and others have shown that inhibition of BET bromodomain chromatin reader proteins attenuates growth of AR-wild-type and AR-mutant CRPC (5,6). Our group has also shown that the histone-lysine N-methyltransferase EZH2 is overexpressed and associated with poor prognosis in prostate cancer (7). While EZH2 has been described as a transcriptional repressor that methylates histone H3 lysine 27 (H3K27Me3) to mediate epigenetic silencing of multiple tumor suppressors, studies also suggest that EZH2 possesses additional activities in prostate cancer (8). Recently, EZH2 has been found to be highly expressed in neuroendocrine prostate cancer (NEPC); progression to NEPC includes loss of lineage-specific AR-mediated gene expression and gain-of-function in EZH2 (9,10). Another study in a NEPC model showed that EZH2 inhibition could reactivate AR signaling and sensitivity to drugs (11), while an additional showed that NEPC may represent a late-stage with loss of plasticity, and AR expression/signaling was not amenable to EZH2 targeting (12). Whether inhibition of EZH2 can impact AR signaling during resistance to current AR targeting therapies (e.g., enzalutamide) outside of NEPC has not been defined.

In this study, we identified EZH2 as a key regulator of sensitivity to AR-targeted therapies in CRPC adenocarcinoma models through an unbiased epigenetic inhibitor screen. EZH2 inhibition had profound effects on the AR cistrome and signaling, suggesting an enhanced dependence on this pathway. Furthermore, we tested antisense oligonucleotides (ASOs) targeting EZH2 and AR as an alternative strategy to target oncogenic factors, which may be beneficial in settings where ablation of these proteins is desired compared to inhibition of select enzymatic activities or domains. ASOs have demonstrated clinical activity for the treatment of multiple diseases, including diabetes, hyperlipidemias, cardiovascular diseases, neurodegenerative diseases, and cancer; those currently approved for therapeutic use include

Kynamro®, Spinraza™, and Exondys 51™ (13). Herein, a novel ASO strategy targeting EZH2 and AR was shown to be effective in inhibiting CRPC growth *in vivo*, suggesting that targeting this axis may provide clinical benefit in patients that have developed resistance to established therapies.

MATERIALS AND METHODS

Cell culture

LNCaP, C4-2B, and CWR-R1 prostate cancer cell lines were grown in RPMI 1640 (Invitrogen) supplemented with 10% fetal bovine serum (FBS) (Invitrogen) in a 37°C incubator with 5% CO_2. The LNCaP-abl (14) and LNCaP-EnzR (15) cell lines were propagated as previously described. LNCaP cells were obtained from ATCC. C4-2B cells were generously provided by Evan Keller, Ph.D., at the University of Michigan, CWR-R1 and LNCaP-EnzR cell lines were kindly provided by Donald Vander Griend, Ph.D., at the University of Chicago, and the LNCaP-abl cell line was kindly provided by Myles Brown, Ph.D., at Harvard University. Cells were genotyped to confirm identity at the University of Michigan Sequencing Core and tested routinely for Mycoplasma contamination.

Epigenetic inhibitors screen

C4-2B cells were seeded in 96-well plates at a density of 2000 cells/well in a total volume of 50 µL media. Each epigenetic inhibitor (see **Supplementary Table S1**) in the customized panel (Selleck) was added at 1 µM concentration to an individual well containing either enzalutamide or DMSO vehicle (final concentration 2 µM). Cells were incubated for five days before quantification with CellTiter-Glo reagent (Promega), per manufacturer's instructions. CellTiter-Glo analysis was performed after transferring 100 µL of CellTiter-Glo solution from each well into a Costar 96-well clear flat bottom plate for reading with a Tecan Infinite M1000 PRO reader.

CellTiter-Glo luminescent signal was normalized to negative control wells (without epigenetic inhibitor); raw values are included in **Supplementary Table S2**.

Prostate tumor xenograft model and drug studies

Six-week-old male CB17 severe combined immunodeficiency (SCID) mice were used for experiments. C4-2B prostate cancer cells (1 X 10^6 cells) in 50% Matrigel (BD Biosciences) were injected subcutaneously into bilateral dorsal flanks. When xenografts reached an average size of 50-100 mm^3, animals were randomized into groups (each containing 10 animals) subsequently treated with one of the following: PBS, Vehicle, Enzalutamide (30 mg/kg), ASO-Ctrl (50 mg/kg), ASO-AR (25 mg/kg), ASO-EZH2-65 (50 mg/kg), or ASO combo (25 mg/kg ASO-AR, 50 mg/kg ASO-EZH2-65). For experiment in supplemental data, treatment cohorts were as follows: ASO-Ctrl (75 mg/kg), ASO-AR (25 mg/kg) + ASO-Ctrl (50 mg/kg), ASO-EZH2-65 (50 mg/kg) + ASO-Ctrl (25 mg/kg), or ASO combo (25 mg/kg ASO-AR, 50 mg/kg ASO-EZH2-65). All reagents were dosed five days per week for three weeks. Enzalutamide was delivered by gavage and ASOs by subcutaneous injection. Tumor volume was calculated from digital caliper measurements made at the study outset, then twice weekly. Tumors were resected and weighed at the end of the time course. Animal experiments were approved by the University of Michigan Institution Animal Care and Use Committee (IACUC).

Chou-Talalay combination index for Loewe additivity

Loewe additivity is a dose-effect model which states that additivity occurs in a two-drug combination if the sum of the ratios of the dose vs. the median effect for each individual drug is 1. In this model, combination index (CI) scores estimate the interaction between the two drugs. If CI < 1, the drugs have a synergistic effect and if CI > 1, the drugs have an antagonistic effect. CI = 1 means the drugs have additive effect. The CI coefficients were computed based on the Chou-Talalay median effect model as implemented in CalcuSyn v2.11

(http://www.biosoft.com/w/calcusyn.htm).

Antibodies and immunoblot analyses

Antibodies used in the immunoblotting assays are as follows: AR (Millipore, Cat. # 06-680), PSA (DAKO, Cat. # A0562), EZH2 (Cell Signaling, Cat. # 5246S), H3K27Me3 (Diagenode, Cat. # C15410069), Histone H3 (Cell Signaling, Cat. # 9715S), GAPDH (Cell Signaling, Cat. # 3683S). All antibodies were employed at dilutions suggested by the manufacturers. For Western blot analysis, 200 µg total protein extract was boiled in sample buffer and 10-20 µg aliquots were separated by SDS-PAGE and transferred onto Polyvinylidene Difluoride membrane (GE Healthcare). The membrane was incubated for one hour in blocking buffer [Tris-buffered saline, 0.1% Tween (TBS-T), 5% nonfat dry milk] followed by incubation overnight at 4°C with primary antibody. Following a wash with TBS-T, the blot was incubated with horseradish peroxidase-conjugated secondary antibody, and signals were visualized by enhanced chemiluminescence system as per manufacturer's protocol (GE Healthcare).

RNA isolation, quantitative real-time PCR, and RNA-seq

Total RNA was isolated from cells using Direct-zol kit (Zymo), and cDNA was synthesized from 1,000 ng total RNA using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific). Quantitative real-time PCR was performed in triplicate using standard SYBR green reagents and protocols on a StepOnePlus Real-Time PCR system (Applied Biosystems). The target mRNA expression was quantified using the ΔΔCt method and normalized to HMBS expression. All primers were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/) and synthesized by Integrated DNA Technologies (Coralville). Primer sequences are listed in **Supplementary Table S3**. RNA-seq was performed using the Illumina HiSeq 2000 in paired-end mode, as previously described (16). Detailed description of GSEA analysis is provided in the Supplementary Materials and Methods.

Chromatin immunoprecipitation (ChIP)-seq

ChIP assays for AR were performed using HighCell ChIP kit (Diagenode) according to manufacturer's protocol, using AR antibody (Millipore, Cat. # 06-680). For AR ChIP-seq experiments with EPZ-6438 and EZH2 ASOs, C4-2B cells were treated with 1 μM EPZ-6438 or ASO-EZH2-65 for 5 days. Next, cells were cross-linked for 10 minutes with 1% formaldehyde. Cross-linking was terminated by the addition of 1/10 volume 1.25 M glycine for five minutes at room temperature followed by cell lysis and sonication (Bioruptor, Diagenode). This resulted in an average chromatin fragment size of 200 base pairs. A chromatin equivalent of 5×10^6 cells was used for each assay. ChIP DNA was isolated (IPure Kit, Diagenode) from samples by incubation with antibody at 4°C overnight followed by wash and reversal of cross-linking. The ChIP-seq sample preparation for sequencing was performed according to manufacturer's instructions (Illumina). ChIP-enriched DNA samples (1-10 ng) were converted to blunt-ended fragments using T4 DNA polymerase, E. coli DNA polymerase I large fragment (Klenow polymerase), and T4 polynucleotide kinase (New England BioLabs, NEB). A single A-base was added to fragment ends by Klenow fragment (3′ to 5′ exo minus; NEB) followed by ligation of Illumina adapters (Quick ligase, NEB). The adapter-modified DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (Qiagen). Libraries were quantified with the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2000 Sequencer (100 nucleotide read length). Detailed description of ChIP-seq data analysis is provided in the Supplementary Materials and Methods.

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)

ATAC-seq was performed as previously described (17). Briefly, C4-2B cells were treated with 1

μM EPZ-6438 or DMSO control for 5 days (or ASO-Ctrl and ASO-EZH2), then trypsinized and pelleted. 25,000 cells were washed in cold PBS and resuspended in lysis buffer (CER-I of NE-PER kit, Invitrogen, Cat. # 78833) on ice for 10 minutes with occasional pipetting. The lysate was centrifuged at 1300g for five minutes at 4℃. Nuclei were resuspended in 2X TD buffer, then incubated with Tn5 enzyme for 30 minutes at 37℃ (Nextera DNA Library Preparation Kit, Cat. # FC-121-1031). Samples were immediately purified by Qiagen minElute column and PCR amplified with the NEBNext High-Fidelity 2X PCR Master Mix (NEB, Cat. # M0541L). qPCR was utilized to determine the optimal PCR cycles to prevent over-amplification. The amplified library was further purified by Qiagen minElute column and SPRI beads (Beckman Coulter, Cat. # A63881). ATAC-seq libraries were sequenced on the Illumina HiSeq 2500, paired-end, 50 cycles. Detailed description of ATAC-Seq data analysis is provided in the Supplementary Materials and Methods.

RNA *in situ* hybridization (RNA-ISH)

The RNAscope 2.5 HD BROWN assay (Advanced Cell Diagnostics, Cat. #322300) was performed using target probes to EZH2 or AR according to the manufacturer's instructions. EZH2 RNA probes (Cat. #405491, accession #NM_001203248.1, targeting 197 - 1337) and AR RNA probes (Cat. #40049, accession # NM_000044.3, targeting 5604 - 6660) are complementary to the target mRNA. Probes Hs-PPIB (human peptidylprolyl isomerase B) and DapB (bacterial dihydrodipicolinate reductase) were used as positive and negative controls, respectively. Paraffin sections from formalin-fixed tumor samples were incubated for one hour in a 60°C drying oven, before deparaffinization in xylene, brief incubation in 100% ethanol, and air drying. Samples were permeabilized with Protease Plus, then subjected to probe hybridization with two-hour incubation in HybEZ at 40°C. After washing, slides were processed by standard signal amplification steps. Chromogenic detection was performed using DAB,

followed by counterstaining with 50% Gill's hematoxylin I (Fisher, Cat. #26801-01).

All slides were examined for EZH2 or AR ISH signals in morphologically intact cells and scored by a blinded independent pathologist. ISH signal was defined as brown, punctate dots, and expression level scored as described in (18): 0 = no staining or less than one dot per 10 cells, 1 = one to three dots per cell, 2 = four to nine dots per cell (few or no dot clusters), 3 = 10 to 15 dots per cell (less than 10% in dot clusters), and 4 = greater than 15 dots per cell (more than 10% in dot clusters). Cumulative ISH product score was calculated for each evaluable slide as the sum of the individual products of the expression level (0–4) and percentage of cells [0–100; i.e., (A% × 0) + (B% × 1) + (C% × 2) + (D% × 3) + (E% × 4); total range = 0 to 400].

Data availability

The raw RNA-seq, ChIP-seq (AR) and ATATC-seq data have been deposited at SRA(SRP157942).

RESULTS

EZH2 inhibitors sensitize prostate cancer cells to enzalutamide

Since past studies associated epigenetic changes with CRPC development, we postulated that inhibition of key epigenetic regulators may sensitize CRPC cells to antiandrogen therapy, such as enzalutamide (5-7). An unbiased epigenetic inhibitor screening experiment was conducted, subjecting CRPC-derived C4-2B cells to treatment with a library of 92 epigenetic regulator inhibitors (**Supplementary Table S1**) (**Fig. 1A**). This screen found that EZH2 inhibitors markedly enhanced the inhibitory effect of enzalutamide on cell proliferation and ranked the highest among the compounds assessed (**Fig. 1A**). The experiment was repeated with the 10 most efficacious epigenetic inhibitors and found that six targeted EZH2 (**Fig. 1B**).

Immunoblot analysis demonstrated that these six reduced H3K27Me3 protein content, indicating diminution of EZH2 function (**Fig. 1B**).

These findings were validated by employing EPZ-6438, a small molecule EZH2 inhibitor that is effective in treating hematological malignancies and solid tumors (19). Combination index analysis indicated that co-treatment with EPZ-6438 and enzalutamide produced synergistic antiproliferative effects in C4-2B and androgen-dependent LNCaP cells (**Fig. 1C**). Co-treatment yielded the slowest growth curve compared to treatment by DMSO or single agent and resulted in near-complete inhibition of cell proliferation (**Fig. 1D**). We further confirmed that EPZ-6438 was able to sensitize cells to enzalutamide using an enzalutamide-resistant LNCaP model, LNCap-EnzR (15) (**Supplementary Fig. S1**). Selective disruption of EZH2 with CRISPR/Cas9 further established that loss of EZH2 enhanced the growth inhibitory effect of enzalutamide in prostate cancer cells and in an additional model, CWR-R1 cells (**Supplementary Fig. S2A-D and Supplementary Fig. S3A-C**).

Antisense oligonucleotides (ASOs) represent a novel technique to target AR and EZH2 in prostate cancer cells

AR is central to prostate cancer development, progression, and drug resistance (2). Among the many AR signaling perturbations in CRPC, AR mutant forms or ligand-independent variants (AR-Vs) are not effectively inhibited by currently approved therapies, such as enzalutamide (20,21). Similarly, most inhibitors of EZH2, including EPZ-6438, target the SET domain of EZH2 responsible for methyltransferase activity. Studies have, however, shown that the oncogenic activity of EZH2 in CRPC can occur independently of its Polycomb-repressive function, and targeting the non-PRC2 (polycomb repressive complex-2) function of EZH2 may have therapeutic efficacy (8). ASOs optimized for *in vivo* delivery present a novel solution by targeting all forms of EZH2 and AR at the mRNA level, leading to ablation of the protein itself rather than inhibition of enzymatic function.

In collaboration with Ionis Pharmaceuticals, we have tested clinical-grade ASOs targeting AR and EZH2. ASOs were delivered to the cells *via* free uptake, as previously described (22). As shown in **Fig. 2A** and **Supplementary Fig. S4A-B**, ASOs inhibited their targets and downstream activities efficiently, as indicated by respective dose-dependent reductions in EZH2/H3K27Me3 and AR/PSA levels. We compared the growth inhibitory effects of EZH2 ASOs to EPZ-6438 in LNCaP-abl cells, an androgen-independent cell line which relies on the PRC2-independent functions of EZH2 (8). As shown in **Supplementary Fig. S5A**, EPZ-6438 and EZH2 ASOs both inhibited LNCaP-abl cell growth, but EZH2 ASOs exhibited enhanced efficacy. Furthermore, ASO-AR treatment was able to decrease growth of enzalutamide-resistant LNCaP-EnzR cells (**Supplementary Fig. S5B**), suggesting that ablation of EZH2 and AR by ASOs may indeed be advantageous to enzymatic inhibition in certain settings.

Consequently, we determined whether the results generated in **Fig. 1** could be recapitulated with ASOs to ablate EZH2 and AR. In corroboration with data in **Fig. 1**, EZH2 ASOs sensitized LNCaP and C4-2B cells to ASO-AR, lowering the IC_{50} of ASO-AR and shifting the growth inhibitory curve to the left (**Fig. 2B**). Furthermore, both CellTiter-Glo assays (**Fig. 2C**) and soft agar colony experiments (**Supplementary Fig. S6A-B**) demonstrated that ASOs co-targeting EZH2 and AR yielded the most potent cellular growth inhibition. Combination index analysis indicated that EZH2 and AR ASOs synergized in both LNCaP and C4-2B cells (**Fig. 2D**). Similar results were obtained using CRISPR/Cas9 to target EZH2 in the presence of ASO-AR (**Supplementary Fig. S6C-D**). Combination ASO-AR and ASO-EZH2 treatment was also tested in LNCaP-EnzR cells and shown to be more effective at inhibiting growth of this model than either agent alone (**Supplementary Fig. S5B**).

EZH2 inhibition activates AR signaling in prostate cancer cells

We sought to evaluate how EZH2 inhibition may sensitize CRPC cells to AR-targeted

therapies. RNA-seq was employed to profile transcriptomic changes in C4-2B cells treated with EPZ-6438 or ASO-EZH2. Interestingly, gene set enrichment analysis (GSEA) revealed AR signaling to be the set most significantly activated by either agent (**Fig. 3A-B**). RT-PCR confirmed that multiple AR target genes were significantly upregulated upon treatment with EPZ-6438 or EZH2 ASOs in LNCaP and C4-2B cells (**Fig. 3C**).

As a subunit of PRC2, EZH2 associates with chromatin, inducing its compaction and limiting access of transcription factors and ATP-dependent remodeling machinery; therefore, EZH2 inhibition may yield global redistribution of transcription factor cistromes. Genome-wide Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) was performed to profile chromatin accessibility changes resulting from EZH2 inhibition in C4-2B cells with either EPZ-6438 or ASO-EZH2. As shown in **Fig. 3D** and **Supplementary Fig. S7A**, both treatments dramatically increased chromatin accessibility. *De novo* transcription factor binding site motif analysis was performed on differential accessible chromatin regions and revealed binding sites for AR and associated pioneer factors to be the most enriched motifs upon EZH2 inhibition or ablation (**Fig. 3D** and **Supplementary Fig. S7B**). Genome-wide AR chromatin immunoprecipitation sequencing (ChIP-seq) was employed and integrated with ATAC-seq data to directly profile EPZ-6438 and ASO-EZH2-mediated AR occupancy changes. Consistent with motif analysis, these data showed that differentially accessible regions were highly associated with AR binding (**Fig. 3D**, **Supplementary Fig. S7A and C**). Combined, these data propose that inhibition of EZH2 may lead to reprogramming of the AR cistrome and upregulation of AR signaling. These alterations could increase dependence of prostate cancer cells on this pathway and their sensitivity to AR-targeted therapies.

EZH2 and AR ASO therapy additively inhibit prostate cancer xenograft growth

ASOs engineered for *in vivo* delivery are an emerging class of oncology therapeutics (13). An ASO-based strategy was employed to determine the effect of inhibiting AR and EZH2

in prostate cancer xenografts. The efficacy of AR-targeting ASO as an anti-tumor agent in C4-2B cells was first evaluated. Following a three-week time course, xenografts in both active agent groups had significantly lower volumes than those of controls (**Fig. 4A**). Xenograft volumes were significantly reduced in animals treated with ASO-AR versus enzalutamide, highlighting advantages of the ASO technology; xenograft weights confirmed volume analysis (**Fig. 4A**).

The same experimental protocol was used to determine the impact of combining ASOs targeting AR and EZH2 on xenograft growth. As observed above, ASO-AR significantly inhibited xenograft growth, while ASO-EZH2 (50 mg/kg) had no effect (**Fig. 4B**). However, combination of ASO-AR and ASO-EZH2 yielded significant reduction in tumor volume versus ASO-AR alone. Measurements of xenograft weight paralleled those of volume (**Fig. 4B**), and a clear size decrement was appreciable in the combination treatment (**Supplementary Fig. S8A**). Quantitative RNA *in situ* hybridization (RNA-ISH) verified significant reductions in mRNA encoding EZH2 and AR (**Fig. 4C** and **Supplementary Fig. S8B-C**). Tumor growth and weights in the combination group were also significantly reduced when compared to single agent arms employing equal amounts of total ASOs (i.e. ASO-AR + ASO-Ctrl and ASO-EZH2 + ASO-Ctrl). This ensured that antitumor effects of combination treatment did not result from higher total ASO load (**Supplementary Fig. S8D**). These data demonstrate the ability of EZH2 inhibition to augment the growth inhibitory effect of AR inhibition in CRPC. Additionally, the findings highlight ASO therapy as a novel method of targeting both AR and EZH2 *in vivo*.

DISCUSSION

Although advanced prostate cancer often responds initially to therapies that suppress androgen-axis signaling, resistance inevitably develops, leading to the emergence of CRPC. The clinical efficacies of therapies targeting AR, such as abiraterone and enzalutamide, have confirmed that AR signaling remains an important driver of CRPC (3,4,23). Approaches to improve the duration of response and address key pathways of resistance are needed. The data

presented here suggest that administration of therapies that inhibit EZH2 signaling may sensitize tumor cells to AR-targeting therapies in CRPC. EZH2 inhibitors lead to increased open chromatin and AR binding throughout the genome, upregulating AR signaling. Interestingly, this appears to increase dependence of CRPC cells on this pathway, as sensitivity to AR-targeting therapies is enhanced (**Fig. 4D**).

Past studies have associated epigenetic aberrations with prostate cancer progression and drug resistance and support targeting of epigenetic mediators as a potential therapeutic strategy. Our group previously found EZH2 to be overexpressed in prostate cancer and associated with progression and poor prognosis (7). Recent studies showed that EZH2 plays a critical role in NEPC development and suggested that inhibition of EZH2 is a promising therapeutic avenue for NEPC that could reactivate AR signaling and restore sensitivity to AR-targeting drugs (9-11). Here, we defined the role of EZH2 inhibition in CRPC adenocarcinoma models and demonstrated that EZH2 inhibition is a viable treatment strategy to restore sensitivity; results were similar in LNCaP cells, proposing EZH2 inhibition may also augment responses to AR-targeted drugs in androgen-dependent stages of prostate cancer. Interestingly, inhibition of EZH2 caused redistribution of AR binding and upregulated AR signaling, suggesting a greater dependence on this pathway underlies ADT sensitization. Recently, a phase 2 study showed that bipolar androgen therapy resulted in resensitization to enzalutamide in patients undergoing rechallenge (24). Although a different avenue of intervention, these data support our hypothesis that further activation of AR signaling may restore response to ADT.

Through exploration of alternative strategies to inhibit all functions of EZH2 and AR, we found ASOs, a rapidly expanding class of oncology therapeutic agents, as a novel treatment strategy for combinatorial targeting of these factors in CRPC (13). In reports involving other cancer lineages, AZD4785 (ASO targeting KRAS) has been shown to be an attractive therapeutic for treatment of KRAS-driven human cancers, and AZD9150 (ASO targeting STAT3) exhibited promising antitumor activity in patients with treatment-refractory lymphoma and non-

small cell lung cancer (22,25). Previously, AR ASOs targeting both full-length AR and AR variants have been shown as a rational approach for treatment of AR-dependent CRPC (26). Our data demonstrate that combining AR-ASO with EZH2 ASOs enhances the antitumor effect in CRPC. Furthermore, tumors treated with ASO-AR alone exhibited enhanced growth inhibition compared to enzalutamide, another advantage of ASO technology in this setting. Finally, ASOs have the power to target previously undruggable proteins or disease-associated non-coding RNAs, such as AR variants or Polycomb-independent functions of EZH2 (8,20). Altogether, data from this study suggest that clinical studies investigating ASO-AR and ASO-EZH2 therapy in CRPC patients are warranted.

ACKNOWLEDGMENTS

We thank S.M. Dhanasekaran, C. Kumar, M. Cieślik, D. Robinson, Y-M. Wu, S. Shankar, and S. Kregel for helpful discussions. We also thank S. Ellison for assistance in writing and editing of the manuscript. This work was supported in part by a NCI Prostate SPORE (P50CA186786). A.M. Chinnaiyan is a Howard Hughes Medical Institute Investigator, A. Alfred Taubman Scholar, and American Cancer Society Professor as well as is supported by the Prostate Cancer Foundation. L. Xiao is supported by a Department of Defense (DoD) Postdoctoral Award (W81XWH-16-1-0195). J.C. Tien and Y. Qiao are supported by the PCF Young Investigator Award. A. Parolia is supported by the DoD Early Investigator Research Award (W81XWH-17-1-0130). Y. Zhang is also supported by the DoD Early Investigator Research Award (W81XWH-17-1-0134).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7-30.

2. Ferraldeschi R, Welti J, Luo J, Attard G, de Bono JS. Targeting the androgen receptor pathway in castration-resistant prostate cancer: progresses and prospects. Oncogene 2015;34:1745-57.

3. Scher HI, Fizazi K, Saad F, Taplin ME, Sternberg CN, Miller K*, et al.* Increased survival with enzalutamide in prostate cancer after chemotherapy. N Engl J Med 2012;367:1187-97.

4. Beer TM, Armstrong AJ, Rathkopf DE, Loriot Y, Sternberg CN, Higano CS, *et al.* Enzalutamide in metastatic prostate cancer before chemotherapy. N Engl J Med 2014;371:424-33.

5. Asangani IA, Dommeti VL, Wang X, Malik R, Cieslik M, Yang R, *et al.* Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. Nature 2014;510:278-82.

6. Coleman DJ, Van Hook K, King CJ, Schwartzman J, Lisac R, Urrutia J, *et al.* Cellular androgen content influences enzalutamide agonism of F877L mutant androgen receptor. Oncotarget 2016;7:40690-703.

7. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 2002;419:624-9.

8. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT, *et al.* EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science 2012;338:1465-9.

9. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med 2016;22:298-305.

10. Dardenne E, Beltran H, Benelli M, Gayvert K, Berger A, Puca L, *et al.* N-Myc induces an EZH2-mediated transcriptional program driving neuroendocrine prostate cancer. Cancer Cell 2016;30:563-77.

11. Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW, *et al.* Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science 2017;355:78-83.

12. Puca L, Bareja R, Prandi D, Shaw R, Benelli M, Karthaus WR, *et al.* Patient derived organoids to model rare prostate cancer phenotypes. Nat Commun 2018;9:2404.

13. Bennett CF, Baker BF, Pham N, Swayze E, Geary RS. Pharmacology of antisense drugs. Annu Rev Pharmacol Toxicol 2017;57:81-105.

14. Culig Z, Hoffmann J, Erdel M, Eder IE, Hobisch A, Hittmair A, *et al.* Switch from antagonist to agonist of the androgen receptor bicalutamide is associated with prostate tumour progression in a new model system. Br J Cancer 1999;81:242-51.

15. Kregel S, Chen JL, Tom W, Krishnan V, Kach J, Brechka H, *et al.* Acquired resistance to the second-generation androgen receptor antagonist enzalutamide in castration-resistant prostate cancer. Oncotarget 2016;7:26259-74.

16. Cieslik M, Chugh R, Wu YM, Wu M, Brennan C, Lonigro R, *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. Genome Res 2015;25:1372-81.

17. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213-8.

18. Wang L, Harms PW, Palanisamy N, Carskadon S, Cao X, Siddiqui J*, et al.* Age and gender associations of virus positivity in merkel cell carcinoma characterized using a novel RNA in situ hybridization assay. Clin Cancer Res 2017;23:5622-30.

19. Knutson SK, Kawano S, Minoshima Y, Warholic NM, Huang KC, Xiao Y*, et al.* Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. Mol Cancer Ther 2014;13:842-54.

20. 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.

21. McCrea E, Sissung TM, Price DK, Chau CH, Figg WD. Androgen receptor variation affects prostate cancer progression and drug resistance. Pharmacol Res 2016;114:152-62.

22. Hong D, Kurzrock R, Kim Y, Woessner R, Younes A, Nemunaitis J*, et al.* AZD9150, a next-generation antisense oligonucleotide inhibitor of STAT3 with early evidence of clinical activity in lymphoma and lung cancer. Sci Transl Med 2015;7:314ra185.

23. de Bono JS, Logothetis CJ, Molina A, Fizazi K, North S, Chu L*, et al.* Abiraterone and increased survival in metastatic prostate cancer. N Engl J Med 2011;364:1995-2005.

24. Teply BA, Wang H, Luber B, Sullivan R, Rifkind I, Bruns A*, et al.* Bipolar androgen therapy in men with metastatic castration-resistant prostate cancer after progression on enzalutamide: an open-label, phase 2, multicohort study. Lancet Oncol 2018;19:76-86.

25. Ross SJ, Revenko AS, Hanson LL, Ellston R, Staniszewska A, Whalley N*, et al.* Targeting KRAS-dependent tumors with AZD4785, a high-affinity therapeutic antisense oligonucleotide inhibitor of KRAS. Sci Transl Med 2017;9:1-13.

26. Yamamoto Y, Loriot Y, Beraldi E, Zhang F, Wyatt AW, Al Nakouzi N*, et al.* Generation 2.5 antisense oligonucleotides targeting the androgen receptor and its splice variants suppress enzalutamide-resistant prostate cancer cell growth. Clin Cancer Res 2015;21:1675-87.

FIGURE LEGENDS

Figure 1. Identification of EZH2 as a key target for enzalutamide sensitization.

A. Diagram depicting screen performed in C4-2B cells treated with epigenetic inhibitor panel plus enzalutamide (2 μM) or DMSO control. Degree of proliferation inhibition for

each individual inhibitor (1 μM) plus enzalutamide versus each inhibitor alone is graphed (all normalized to DMSO).

B. Dose response curves of C4-2B cells treated with enzalutamide in combination with each of the 10 most effective epigenetic inhibitors (1 μM). Data represent mean ± standard error (n = 3) from one of three independent experiments. Immunoblot analysis of H3K27Me3 levels in C4-2B cells treated with each epigenetic inhibitor demonstrated EZH2 inhibition with red compounds.

C. Combination index analysis (CI) for the combination of EPZ-6438 and enzalutamide. Circles represent experimentally determined CI values (Chou-Talalay method).

D. Growth curves of LNCaP and C4-2B cells treated with DMSO, enzalutamide (Enza, 2 μM), EPZ-6438 (1 μM), and combination of enzalutamide plus EPZ-6438. Data represent mean ± standard error (n = 6) from one of three independent experiments.

The experiment was conducted in biological triplicate. When not otherwise indicated, data shown as representative experiments, with each point representing the mean (± SD) of technical duplicates (t test). ***, $P < 0.001$.

Figure 2. EZH2- and AR-targeting ASOs synergistically inhibit prostate cancer growth.

A. Top, immunoblot analysis of AR and PSA in LNCaP and C4-2B cells treated with ASO-Ctrl or ASO-AR. ASOs were delivered to cells in all experiments *via* free uptake. Bottom, immunoblot analysis of EZH2 and H3K27Me3 in LNCaP and C4-2B cells treated with ASO-Ctrl and EZH2 ASOs (EZH2-54 and EZH2-65).

B. Dose response curves of LNCaP and C4-2B cells treated with ASO-Ctrl or ASO-EZH2 (EZH2-54 and EZH2-65, 1 μM), then treated with ASO-AR for five days. Data represent mean ± standard error (n = 3) from one of three independent experiments.

C. Growth curves of LNCaP and C4-2B cells treated with ASO-Ctrl, ASO-AR, ASO-EZH2 (EZH2-54 and EZH2-65, 1 μM), or combination of ASO-AR (1 μM) plus ASO-EZH2 (1

μM). Data represent mean ± standard error (n = 6) from one of three independent experiments.

D. Combination index analysis (CI) for the combination of ASO-AR and EZH2 ASOs. Circles represent experimentally determined CI values (Chou-Talalay method).

The experiment was conducted in biological triplicate. When not otherwise indicated, data shown as representative experiments, with each point representing the mean (± SD) of technical duplicates (t test). ***, $P < 0.001$.

Figure 3. EZH2 inhibition enhances AR signaling and genome-wide AR binding.

A. Transcriptomic changes induced in C4-2B cells by treatment with EPZ-6438 and ASO-EZH2, as determined by RNA-seq with GO analysis. GO enrichment p values and number of genes in each GO category are respectively indicated at the x-axis and next to the bar. “Androgen Response” was the top activated gene set by either agent.

B. EPZ-6438 and ASO-EZH2-mediated elevation of AR signaling as demonstrated by GSEA. Hallmark Androgen Response gene sets shown for C4-2B cells treated with EPZ-6438 or ASO-EZH2.

C. qRT–PCR analysis of AR target gene in LNCaP and C4-2B cells treated with DMSO, EPZ-6438, or enzalutamide for 72 h (left) or designated ASOs for 72h (right). Data represent mean ± SEM (n = 3) from one of three independent experiments.

D. Heatmaps of ATAC-seq and AR ChIP-seq signals around ± 2kb of peak center (left). Each row represents ± 2kb around the center of differentially accessible regions (DAR). Top panel represents 2591 DAR that gain accessibility upon EPZ-6438 treatment compared to DMSO. Bottom panel represents 302 DAR that lose accessibility upon EPZ-6438 treatment compared to DMSO. (Right) Motif analysis showing enriched motifs in open chromatin upon EPZ-6438 treatment.

Figure 4. ASOs targeting AR and EZH2 additively inhibit prostate cancer xenograft growth.

A. Tumor volume time course for subcutaneous C4-2B xenografts in CB17SCID mice. Treatment with vehicle, ASO-Ctrl, ASO-AR, or enzalutamide started at tumor volume 50-100 mm^3. Tumor weights from respective treatment groups at day 21 were also determined. N=10 mice in all treatment groups for each panel.

B. Tumor volume time course for subcutaneous C4-2B xenografts in CB17SCID mice. Treatment with ASO-Ctrl, ASO-AR, ASO-EZH2-65, or combination of ASO-AR and ASO-EZH2-65 started at tumor volume 50-100 mm^3. Tumor weights from respective treatment groups at the termination of the study are also graphed.

C. EZH2 and AR expression (ISH product score) in indicated xenograft samples.

D. Schematic depicting proposed mechanism of EZH2 inhibition-mediated sensitization of prostate cancer cells to AR-targeted therapies. EZH2 inhibition increases activation of the AR signaling pathway, increasing their dependence on AR through reprogramming of the AR cistrome. In this setting, effectiveness of AR antagonists (e.g. enzalutamide) and ASO-AR are enhanced.

The experiment was conducted in data shown ± s.e.m. *, $P < 0.05$, **, $P< 0.01$, ***, $P < 0.001$.

Author Manuscript Published OnlineFirst on August 22, 2018; DOI: 10.1158/0008-5472.CAN-18-0941
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

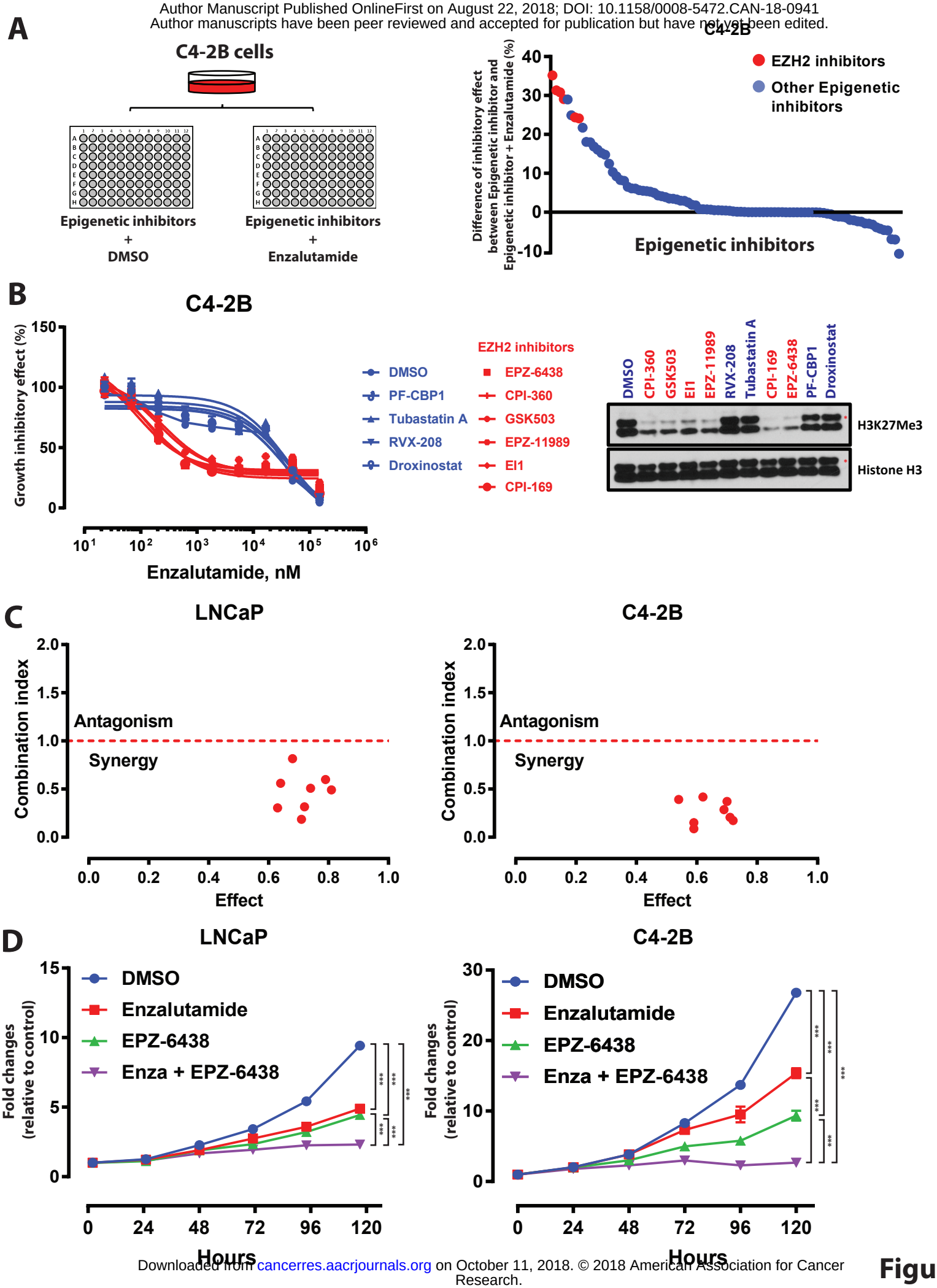


Figure 1

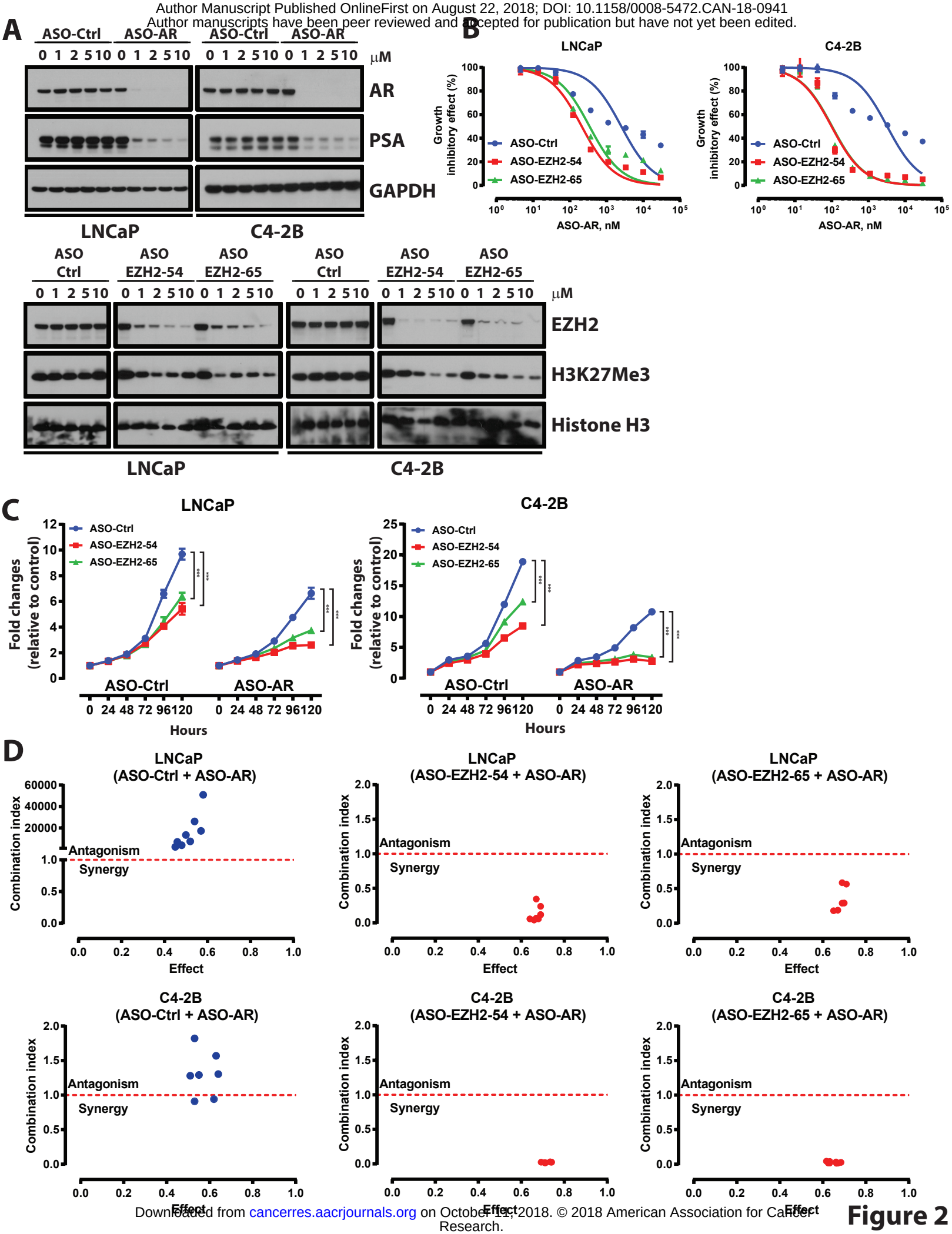


Figure 2

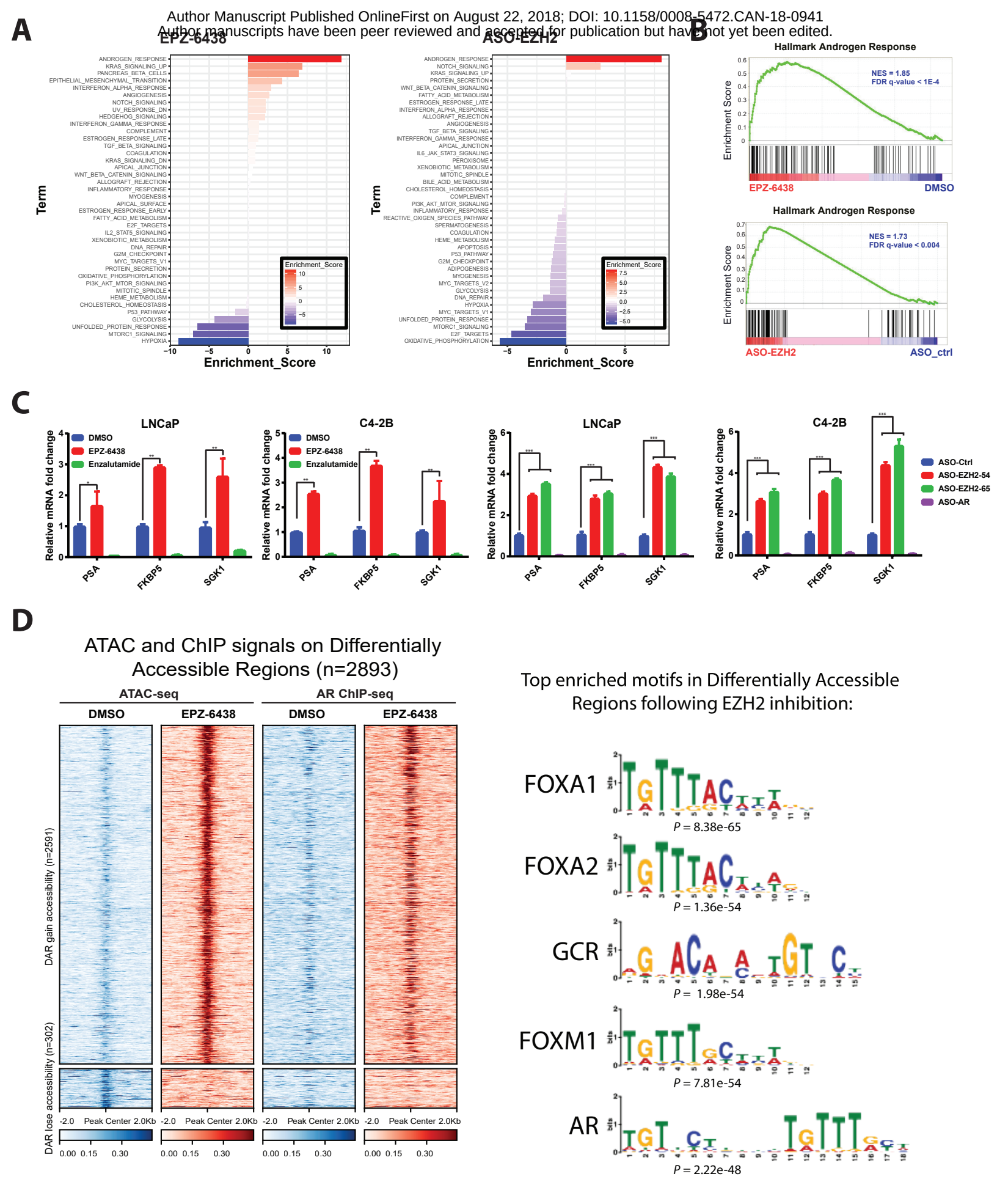


Figure 3

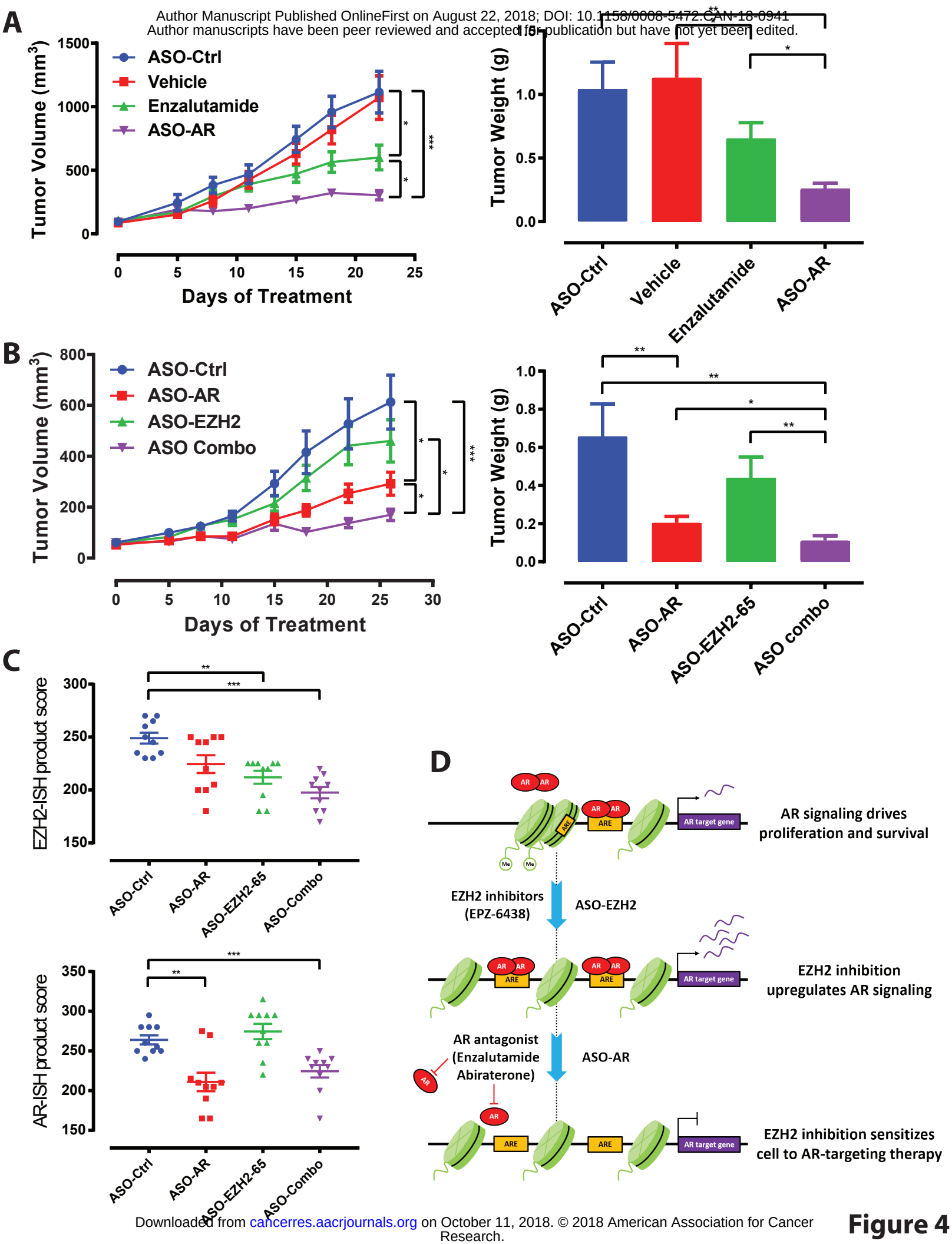


Figure 4

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Epigenetic reprogramming with antisense oligonucleotides enhances the effectiveness of androgen receptor inhibition in castration-resistant prostate cancer

Lanbo Xiao, Jean C Tien, Josh Vo, et al.

Cancer Res Published OnlineFirst August 22, 2018.

ARTICLES
https://doi.org/10.1038/s41588-018-0120-1
nature genetics

Analysis of the androgen receptor–regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression

Yajia Zhang[1,2,3,4,23], Sethuramasundaram Pitchiaya[1,2,23], Marcin Cieślik[1,2,23], Yashar S. Niknafs[1,5], Jean C.-Y. Tien[1,2], Yasuyuki Hosono[1], Matthew K. Iyer[1,4], Sahr Yazdani[1], Shruthi Subramaniam[1], Sudhanshu K. Shukla[1,20], Xia Jiang[1], Lisha Wang[1], Tzu-Ying Liu[6], Michael Uhl[7], Alexander R. Gawronski[8], Yuanyuan Qiao[1,2,9], Lanbo Xiao[1], Saravana M. Dhanasekaran[1,2], Kristin M. Juckette[1], Lakshmi P. Kunju[1,2,9], Xuhong Cao[1,10], Utsav Patel[11], Mona Batish[11,12], Girish C. Shukla[13], Michelle T. Paulsen[9,14], Mats Ljungman[9,14], Hui Jiang[6,9], Rohit Mehra[2,9,15], Rolf Backofen[7], Cenk S. Sahinalp[16,17], Susan M. Freier[18], Andrew T. Watt[18], Shuling Guo[18], John T. Wei[15], Felix Y. Feng[1,9,14,19,21], Rohit Malik[1,22,24] and Arul M. Chinnaiyan[1,2,4,5,9,10,15,24*]

The androgen receptor (AR) plays a critical role in the development of the normal prostate as well as prostate cancer. Using an integrative transcriptomic analysis of prostate cancer cell lines and tissues, we identified ARLNC1 (AR-regulated long noncoding RNA 1) as an important long noncoding RNA that is strongly associated with AR signaling in prostate cancer progression. Not only was ARLNC1 induced by the AR protein, but ARLNC1 stabilized the *AR* transcript via RNA–RNA interaction. ARLNC1 knockdown suppressed AR expression, global AR signaling and prostate cancer growth in vitro and in vivo. Taken together, these data support a role for ARLNC1 in maintaining a positive feedback loop that potentiates AR signaling during prostate cancer progression and identify ARLNC1 as a novel therapeutic target.

Long noncoding RNAs (lncRNAs) are a class of transcripts with diverse and largely uncharacterized biological functions[1–3]. Through crosstalk with chromatin, DNA, RNA species and proteins, lncRNAs function via chromatin remodeling, as well as transcriptional and post-transcriptional regulation[4–9]. High-throughput RNA sequencing (RNA-seq) has enabled the identification of lncRNAs with suggested oncogenic and tumor-suppressive roles, including involvement in the pathogenesis of prostate cancer[7,10–12]. Primary prostate cancer is often hormone dependent and relies on signaling through the AR; therefore, the majority of patients are responsive to front-line treatment with androgen-deprivation therapy[13–15]. However, approximately 20% of cases progress to an incurable stage of the disease known as castration-resistant prostate cancer (CRPC), which still critically relies on AR signaling[16,17], as evidenced by the clinical benefit afforded through the use of enzalutamide[18–21] or abiraterone[22–24]. While substantial efforts have been undertaken to identify mechanisms of sustained AR signaling in CRPC (i.e., AR alterations, *AR* splice variants and alternative activation pathways)[25–31], few studies have investigated the role of AR-regulated lncRNAs. Therefore, we initiated a comprehensive RNA-seq profiling investigation of AR-regulated, cancer-associated lncRNAs from prostate cancer cell lines and patient tissue samples.

Results

Analysis of AR-regulated transcriptome in prostate cancer. To nominate AR-regulated genes (ARGs), RNA-seq was performed on AR-dependent VCaP and LNCaP prostate cancer cell lines that were stimulated with an AR ligand, dihydrotestosterone (DHT), for 6 and

[1]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, USA. [2]Department of Pathology, University of Michigan, Ann Arbor, MI, USA. [3]Molecular and Cellular Pathology Program, University of Michigan, Ann Arbor, MI, USA. [4]Department of Computational Medicine and Bioinformatics, Ann Arbor, MI, USA. [5]Department of Cellular and Molecular Biology, University of Michigan, Ann Arbor, MI, USA. [6]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [7]Department of Computer Science and Centre for Biological Signaling Studies (BIOSS), University of Freiburg, Freiburg, Germany. [8]School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada. [9]Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA. [10]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, MI, USA. [11]New Jersey Medical School, Rutgers University, Newark, NJ, USA. [12]Department of Medical Laboratory Sciences, University of Delaware, Newark, DE, USA. [13]Department of Biological, Geological and Environmental Sciences, Center for Gene Regulation in Health and Disease, Cleveland State Univesity, Cleveland, OH, USA. [14]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA. [15]Department of Urology, University of Michigan, Ann Arbor, MI, USA. [16]School of Informatics and Computing, Indiana University, Bloomington, IN, USA. [17]Vancouver Prostate Centre, Vancouver, British Columbia, Canada. [18]Ionis Pharmaceuticals, Carlsbad, CA, USA. [19]Breast Oncology Program, University of Michigan, Ann Arbor, MI, USA. Present address: [20]Department of Biosciences and Bioengineering, Indian Institute of Technology Dharwad, Dharwad, India. [21]Departments of Radiation Oncology, Urology, and Medicine, Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA. [22]Bristol-Myers Squibb, Princeton, NJ, USA. [23]These authors contributed equally: Yajia Zhang, Sethuramasundaram Pitchiaya, Marcin Cieślik. [24]These authors jointly supervised this work: Rohit Malik, Arul M. Chinnaiyan. *e-mail: arul@umich.edu

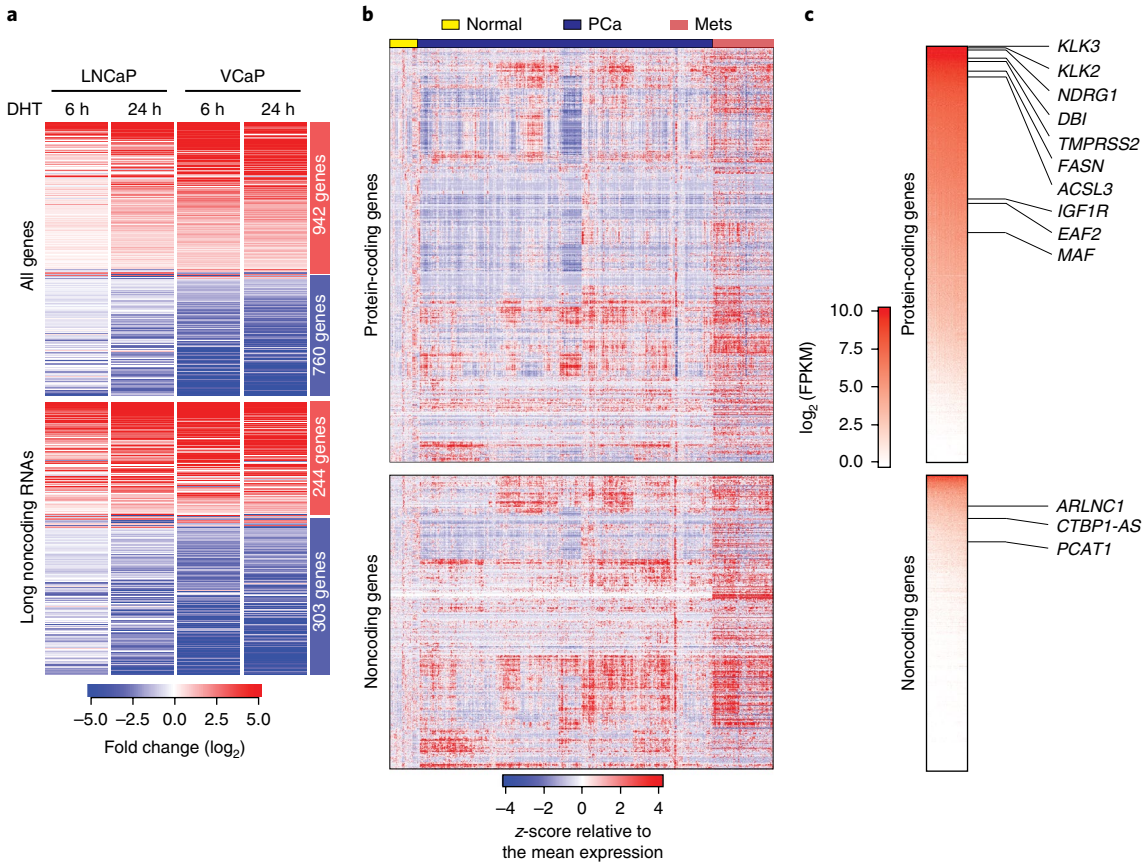


Fig. 1 | Identification of AR-regulated genes in prostate cancer. a, The androgen-regulated transcriptome of prostate cancer cells. The heat map represents the 1,702 genes (including 547 lncRNAs) differentially regulated in LNCaP and VCaP cells following 6 and 24 h of DHT treatment. **b**, The landscape of transcriptomic alterations of prostate cancer progression. The heat map depicts 1,155 protein-coding genes and 547 lncRNAs across benign prostate (normal, $n=52$ samples), localized prostate cancer (PCa, $n=500$ samples) and metastatic prostate cancer (Mets, $n=100$ samples) in TCGA prostate and SU2C-PCF RNA-seq data, with rows representing genes and columns representing patients. Patients were grouped by clinical stage, and genes were subjected to hierarchical clustering. Expression variability is quantified for each gene as a *z*-score relative to the mean expression in normal prostate samples. **c**, A heat map representation of ranked median gene expression levels in prostate tissues. Canonical prostate lineage and prostate cancer markers are listed.

24 h (Supplementary Fig. 1a). A total of 1,702 genes were identified to be concordantly induced or repressed in VCaP and LNCaP cells at both time points (Fig. 1a, Supplementary Fig. 1b,c and Supplementary Table 1), including more than 500 lncRNAs (Fig. 1a and Supplementary Fig. 1d); these data indicate that a large portion of the AR transcriptome remains uncharacterized.

To differentiate between direct and indirect ARGs, previously published and in-house AR chromatin immunoprecipitation (ChIP)–seq data from LNCaP and VCaP cells were analyzed[32]. As expected for direct AR targets, increased levels of AR binding at transcription start sites (TSSs) in both LNCaP and VCaP cells were observed (Supplementary Fig. 1e). The binding levels decreased following treatment with an AR antagonist (enzalutamide) (Supplementary Fig. 1f,g), and the binding sites revealed a de novo motif identical to the canonical AR response element[33] (Supplementary Fig. 1h). A total of 987 genes were categorized as direct ARGs, including 341 lncRNAs (lncARGs) (Supplementary Table 1). Within these genes, we observed an enrichment of chromatin marks associated with 'open' chromatin (H3K27ac, H3K4me1), active promoters (H3K4me3) and transcription (H3K36me3), which, together with RNA polymerase II (Pol II) occupancy, are recognized as manifestations of active gene expression (Supplementary Fig. 1i). Bromodomain and extra-terminal (BET) family proteins, such as BRD4, recognize acetylated histones and have been shown to promote AR transcriptional activity[32]. Consistent with this, we observed colocalization of BRD4 and AR proteins at the promoters of direct AR-responsive genes and loss of AR ChIP peaks following treatment with a bromodomain inhibitor (JQ1) (Supplementary Fig. 1f,i). We further sought to determine whether ARGs identified from cell lines were also targeted by AR in normal prostate tissues and primary tumors. We leveraged a published dataset[34] and queried for the presence of AR peaks within ARG promoters. Remarkably, the majority of ARG promoters were TSS-proximally bound by AR in both tissues and cell lines (Supplementary Fig. 1j,k); conversely, AR-independent genes were distal to AR-binding sites (Supplementary Fig. 1l).

Finally, we confirmed that the ARGs were also expressed in human prostate tissues. We interrogated RNA-seq data from normal prostate, clinically localized prostate cancer (The Cancer Genome Atlas, TCGA)[35] and metastatic CRPC (Stand Up to Cancer–Prostate Cancer Foundation, SU2C-PCF)[30] (Fig. 1b). This revealed remarkable heterogeneity in the expression of ARGs during prostate cancer progression to metastatic disease. As expected, compared to protein-coding genes, noncoding ARGs were detected at lower overall levels (Fig. 1c), although ~10% showed robust expression of over 10 FPKM (fragments per kilobase of transcript per million mapped reads) on average across prostate cancer samples.

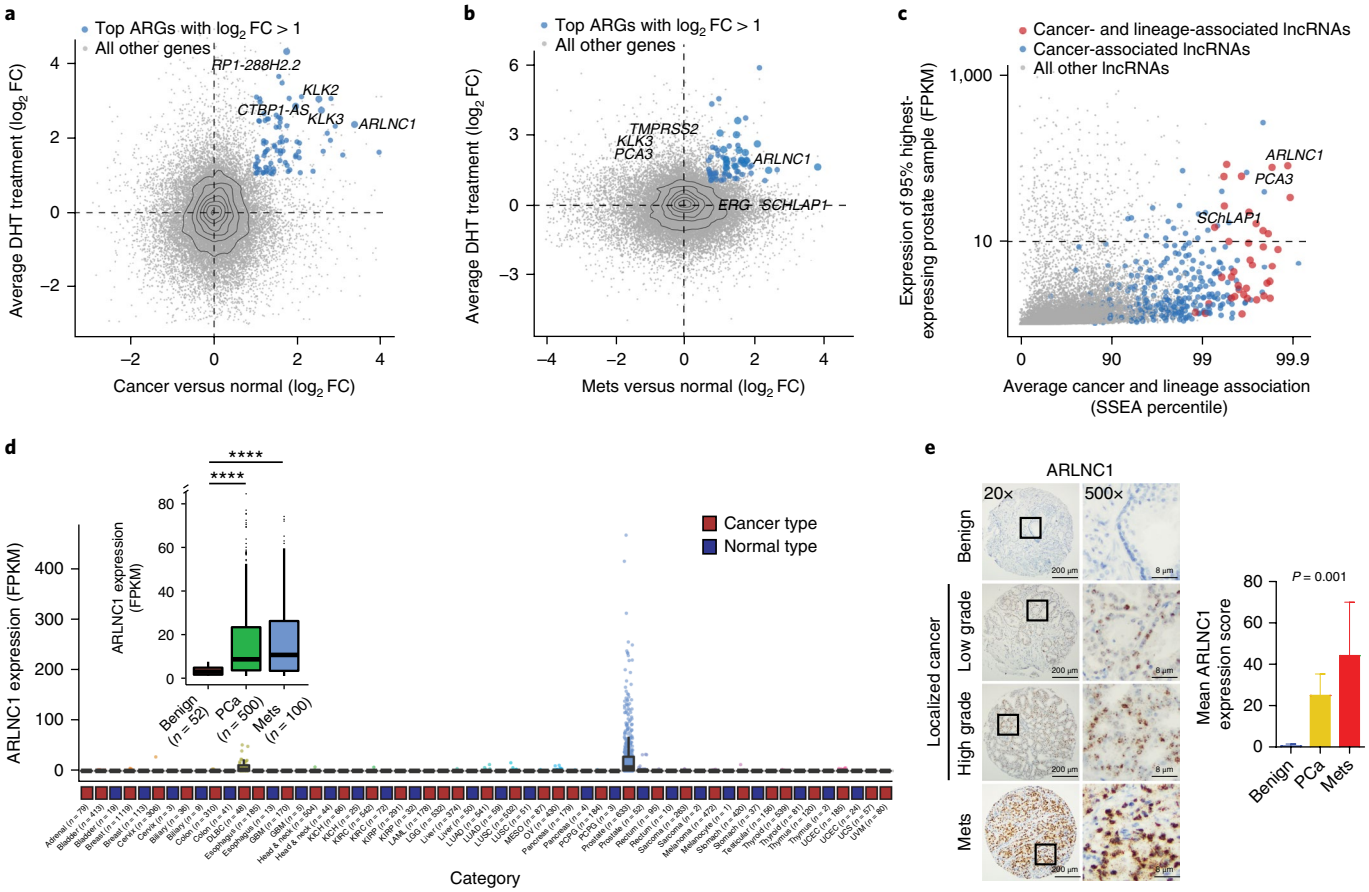


Fig. 2 | Nomination and in situ characterization of ARLNC1 in prostate cancer. a,**b**, Identification of androgen-regulated transcripts elevated in prostate cancer progression. Scatterplots show the AR regulation and cancer association of the ARGs identified in Fig. 1a. The *y* axis depicts the $\log_2$-transformed fold change in gene expression following DHT stimulation, and the *x* axis indicates the $\log_2$-transformed difference in gene expression level between benign prostate ($n=52$ samples) and localized prostate cancer ($n=500$ samples) (**a**) and between benign prostate ($n=52$ samples) and metastatic prostate cancer ($n=100$ samples) (**b**). Significant genes with $\log_2$ fold change > 1 were ranked according to combined *P* value (limma-moderated *t* test). **c**, Nomination of prostate cancer- and lineage-associated lncRNAs based on expression levels. The scatterplot shows the expression level, prostate tissue specificity and prostate cancer association of lncRNAs. The expression level is the FPKM value at the 95th percentile across TCGA prostate samples. Average cancer and lineage associations are represented by the percentile rank for each gene in SSEA (total $n = 7{,}256$ samples). **d**, Relative expression (FPKM) of ARLNC1 across different cancer types in the TCGA cohort. Inset, relative expression (FPKM) of ARLNC1 across benign prostate ($n=52$ samples), localized prostate cancer ($n=500$ samples) and metastatic prostate cancer ($n=100$ samples). PCa versus Normal: $^{****}P < 2.2\times10^{-16}$; Mets versus Normal: $^{****}P = 2.6\times10^{-7}$ (two-sided *t* test). Box-plot definitions: the center line depicts the median, the box shows the first and third quartiles, and the whiskers follow the 1.5 rule. **e**, ISH of ARLNC1 transcript in a human prostate cancer tissue microarray. Representative ARLNC1 staining is shown for benign prostate and localized and metastatic prostate cancer tissue. The bar plot represents mean ARLNC1 expression scores across benign prostate ($n=11$), localized prostate cancer ($n=85$) and metastatic prostate cancer ($n=37$) tissues, with vertical bars indicating the bootstrapped 95% confidence interval of the means. Significance was calculated by a Kruskal–Wallis rank-sum test.

ARLNC1 is a prostate lineage-specific lncRNA with elevated expression in cancer. We hypothesized that lncRNAs associated with prostate cancer progression and castration resistance should be either upregulated if they enhance AR signaling or, conversely, downregulated if they attenuate AR signaling. Their expression is also expected to be AR dependent and lineage restricted if they are part of bona fide physiological feedback loops. Accordingly, a top–down strategy was developed to establish and prioritize clinically relevant, prostate cancer- and lineage-specific lncARGs. First, we identified genes that were both regulated by AR in the VCaP and LNCaP cell lines and upregulated in primary (Fig. 2a) or metastatic (Fig. 2b) prostate cancer as compared to normal prostate tissues. As expected, canonical AR targets, including *KLK3*, *KLK2* and *TMPRSS2*, were among the most differentially expressed protein-coding genes. Notably, this approach highlighted several novel lncARGs, including *ARLNC1* (*ENSG00000260896*, *PRCAT47*[10]), and validated previously identified lncARGs, such as *CTBP1-AS*[36] (Fig. 2a,b). Interestingly, *ARLNC1* was found to be one of the most differentially expressed AR-regulated genes in both localized and metastatic prostate cancer (Fig. 2a,b and Supplementary Fig. 2a,b).

Next, we sought to establish the prostate lineage and cancer specificity of prostate cancer–associated lncRNAs. We leveraged the MiTranscriptome assembly[10], an online resource, to interrogate lncRNA expression across a multitude of tissue and tumor types, and we calculated sample set enrichment analysis (SSEA) scores, which indicate the strength of cancer and lineage association[10]. After applying an expression-level filter (10 FPKM at the 95th percentile), we identified 12 of the most prostate lineage- and cancer-specific lncRNAs (Fig. 2c and Supplementary Fig. 2c,d); 5 of these lncRNAs were regulated by AR. Across these analyses, ARLNC1 was the top prioritized transcript and thus warranted further investigation.

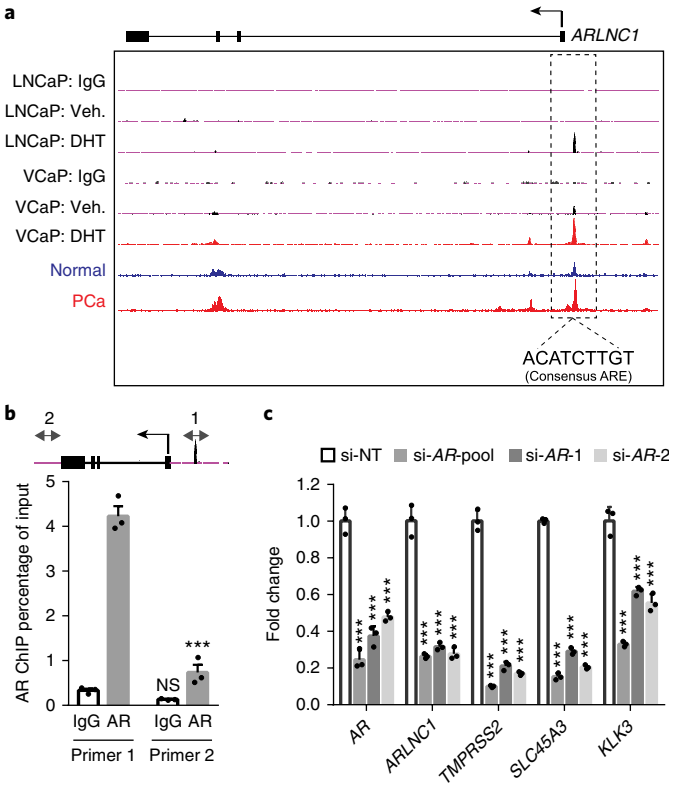


Fig. 3 | *ARLNC1* is directly regulated by AR. a, AR ChIP–seq in prostate cancer cell lines and tissues. Normalized ChIP–seq enrichment is shown. Top, AR or control (IgG) ChIP–seq results across the *ARLNC1* locus in LNCaP and VCaP cells with vehicle (ethanol) or DHT treatment. Bottom, AR ChIP–seq in benign prostate and clinically localized prostate cancer tissue. ARE, androgen response element. **b**, ChIP–qPCR in MDA-PCa-2b cells showing AR or IgG enrichment (ChIP/input) over the *ARLNC1* promoter region (primer 1) or a control region (primer 2). Data are shown as the mean ± s.e.m. ($n = 3$ biologically independent samples). ***P (adjusted) < 0.0001, NS (not significant): $P = 0.5746$, compared to the control region (primer 2) by ANOVA with Sidak correction for multiple comparisons. Top, schematic of the amplicon locations for ChIP–qPCR validation. **c**, Expression of *AR* and AR target genes (*ARLNC1*, *TMPRSS2*, *SLC45A3* and *KLK3*) in MDA-PCa-2b cells transfected with control siRNA (si-NT) or siRNAs against *AR* (si-*AR*-pool, si-*AR*-1, si-*AR*-2). Mean ± s.e.m. values are shown, $n = 3$ biologically independent samples. ***$P = 0.0001$, determined by ANOVA with Dunnett's multiple-comparisons test.

Expression of ARLNC1 was interrogated across cancer and normal tissue RNA-seq samples from TCGA and the Genotype-Tissue Expression (GTEx) project[37,38], respectively. In the TCGA cohort, ARLNC1 exhibited a highly prostate cancer–specific expression pattern, with little to no expression in other tumor types (Fig. 2d). Similarly, in the GTEx normal tissue cohort, its expression was limited to the prostate (Supplementary Fig. 2e). Among the prostate samples, ARLNC1 expression was significantly higher in localized and metastatic prostate cancers than in benign tissues, as assessed by RNA-seq (Fig. 2d, inset) and in situ hybridization (ISH; Fig. 2e). In an extensive differential expression analysis using MiTranscriptome, ARLNC1 was found to be among the top 1% of transcripts most upregulated in prostate cancer and specific to the prostate lineage, with no significant associations in other tissues (Supplementary Fig. 2f). Moreover, the protein-coding genes that were most correlated with ARLNC1 were found to be associated with prostate cancer progression in ONCOMINE concept analyses performed on multiple clinical datasets[39] (Supplementary Fig. 2g). Together, these results confirm that ARLNC1 expression is restricted to the prostate lineage, elevated in prostate cancer and associated with AR signaling throughout prostate cancer progression.

To functionally characterize ARLNC1, we first identified appropriate prostate cancer cell lines with moderate to high levels of ARLNC1 expression using in-house RNA-seq data (Supplementary Fig. 3a). Supporting the association of AR with ARLNC1, ARLNC1 expression was highly enriched in AR-positive cell lines, with the highest expression in MDA-PCa-2b and LNCaP cells. In addition, qPCR analysis for the ARLNC1 transcript also demonstrated that this gene was expressed at the highest level in the MDA-PCa-2b and LNCaP cell lines (Supplementary Fig. 3b). As existing annotations of the *ARLNC1* gene (located on chromosome 16) predict the presence of several transcript isoforms that differ in exon and TSS usage, we determined the exact structure in MDA-PCa-2b and LNCaP cells by RACE. A common TSS for *ARLNC1* was found in both cell lines, and the ~2.8-kb transcript isoform was further confirmed by northern blot analysis (Supplementary Fig. 3c). Single-molecule FISH (smFISH) revealed that approximately 100 molecules of ARLNC1 transcript existed per MDA-PCa-2b cell (Supplementary Fig. 3d,e). Using smFISH and qPCR, we also found that ARLNC1 molecules were distributed equally between the nuclear and cytoplasmic cellular compartments (Supplementary Fig. 3f,g).

***ARLNC1* transcription is directly regulated by AR.** Because ARLNC1 was identified as an AR-regulated lncRNA, we inspected the promoter region of the *ARLNC1* gene for AR occupancy and identified an androgen-induced AR peak in AR ChIP-seq data from both DHT-stimulated VCaP and LNCaP cells (Fig. 3a). Notably, this AR-binding site was also observed in prostate tissue samples and contained a canonical androgen response element[33] (Fig. 3a). These observations were corroborated by ChIP–qPCR in MDA-PCa-2b cells, which showed the highest level of ARLNC1 expression (Fig. 3b). Considering the observation that ARLNC1 expression is prostate tissue specific, while AR expression is not as much, we searched for additional regulators (transcription factors and epigenetic modifiers) of the *ARLNC1* gene (Supplementary Fig. 4a). Motif analysis of the *ARLNC1* promoter region identified several transcription factor binding sites, including a FOXA1 response element. To further validate *ARLNC1* gene regulation by AR and FOXA1, we evaluated ARLNC1 transcript levels following AR or FOXA1 knockdown. AR or FOXA1 loss resulted in decreased expression of ARLNC1, along with other canonical AR target genes that served as positive controls (Fig. 3c and Supplementary Fig. 4b). ChIP–seq and ChIP–qPCR analyses additionally confirmed the putative FOXA1-binding motif on the *ARLNC1* promoter (Supplementary Fig. 4c). Together, these observations suggest that ARLNC1 is directly regulated by AR and modestly regulated by FOXA1, which partially explains the tissue-specific expression pattern of ARLNC1, as expression of these two transcription factors overlaps nearly exclusively in prostate tissue[37,38] (Supplementary Fig. 4d).

ARLNC1 regulates AR signaling. To elucidate the function of ARLNC1 in prostate cancer, we performed gene expression profiling of wild-type and ARLNC1-knockdown MDA-PCa-2b cells (Fig. 4a). Gene ontology (GO) pathway enrichment analysis of the differentially expressed genes revealed deregulation of four main biological activities: apoptosis, cell proliferation, DNA damage response and androgen signaling (Fig. 4a). The significant decrease in AR target gene expression is particularly interesting given the fact that ARLNC1 is regulated by AR, suggesting a positive feedback loop between ARLNC1 and AR signaling. To confirm this observation, we generated an AR target gene signature from MDA-PCa-2b cells stimulated with DHT (Supplementary Fig. 5a and Supplementary Table 2) and performed gene set enrichment analysis (GSEA) using this gene signature (Fig. 4b). Knockdown

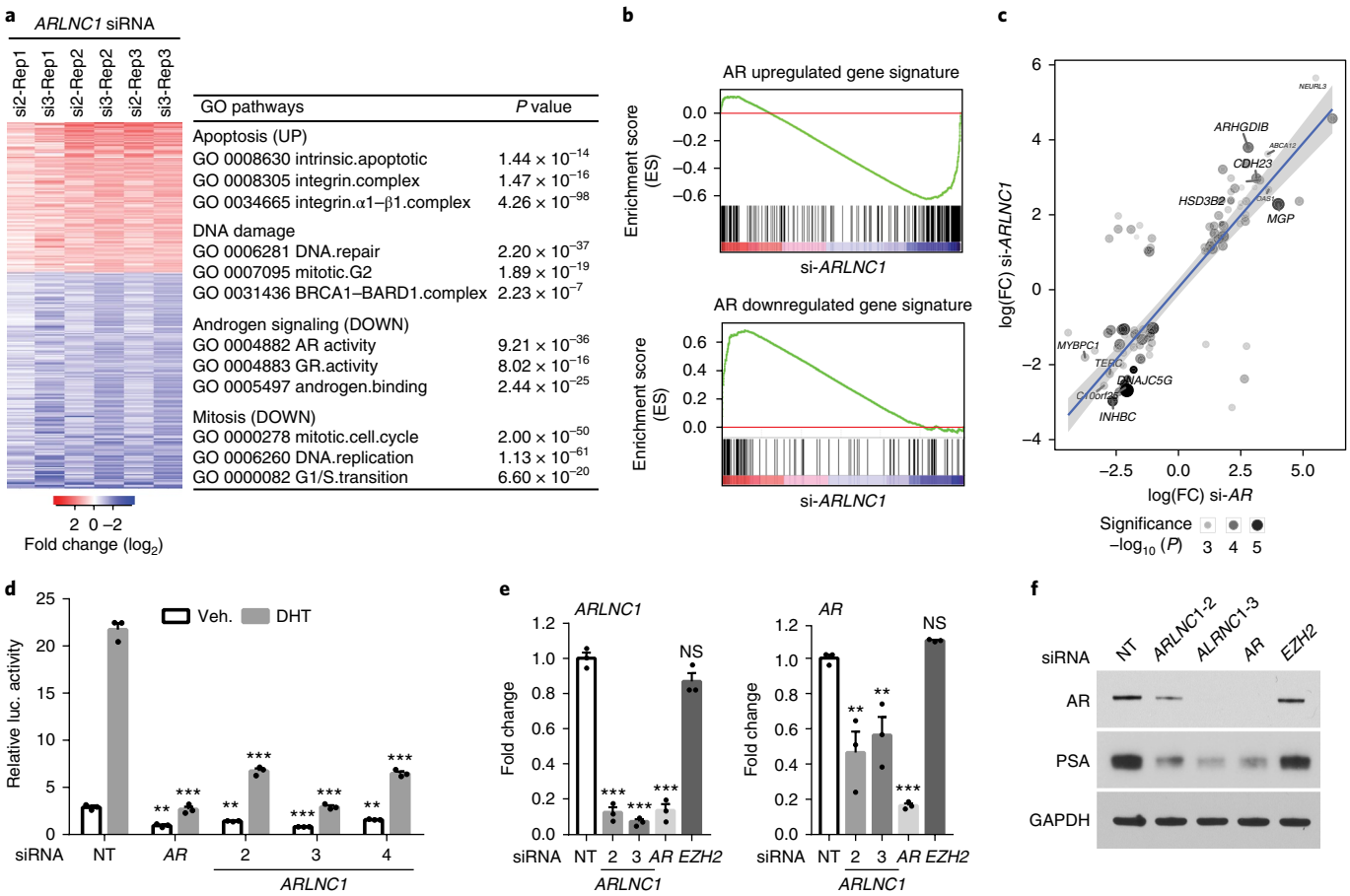


Fig. 4 | ARLNC1 loss attenuates AR signaling. a, Gene expression profiling for ARLNC1 knockdown in MDA-PCa-2b cells ($n = 3$ biologically independent cell cultures for each siRNA). The chart presents the top enriched pathways following ARLNC1 knockdown, identified using GO enrichment analysis (RandomSet test). **b**, GSEA showing significant enrichment of the ARLNC1-regulated gene set with respect to the AR target gene sets ($n = 3$ independent gene expression profiles). Shown are enrichment plots for gene sets consisting of genes positively regulated by AR (top) and genes negatively regulated by AR (bottom). **c**, Comparison of ARLNC1-regulated and AR target genes based on RNA-seq following knockdown of AR and ARLNC1. Significant genes and their log-transformed fold changes in either of the conditions are shown ($n = 2$ independent cell cultures per-condition). Combined significance levels, determined by a limma-moderated *t* test (across both knockdowns), are indicated by circle size. **d**, siRNA knockdown of *ARLNC1* in MDA-PCa-2b cells impairs AR signaling as determined by AR reporter gene assay. siRNA against *AR* serves as a positive control for inhibition of AR signaling. Mean ± s.e.m. values are shown, $n = 3$ biologically independent cell cultures. $^{**}P < 0.01$, $^{***}P = 0.0001$, determined by ANOVA with Dunnett correction. **e**, qRT-PCR analysis of *ARLNC1* and *AR* in MDA-PCa-2b cells transfected with siRNAs against *ARLNC1*, *AR*, *EZH2* or non-specific control (NT). siRNA against *AR* serves as a positive control for inhibition of AR signaling, while siRNA against *EZH2* serves as a negative control. Mean ± s.e.m. values are shown, $n = 3$. $^{**}P < 0.01$, $^{***}P = 0.0001$, determined by ANOVA with Dunnett correction. **f**, Immunoblots of AR, PSA and GAPDH in MDA-PCa-2b cells transfected with siRNAs against *ARLNC1*, *AR*, *EZH2* or non-specific control (NT). The experiments were repeated three times independently with similar results. Uncropped images are shown in Supplementary Fig. 9.

of ARLNC1 led to suppression of genes positively regulated by AR and upregulation of genes negatively regulated by AR (Fig. 4b,c and Supplementary Fig. 5b). This was further confirmed by an AR reporter activity assay (Fig. 4d and Supplementary Fig. 5c), as well as qPCR analysis of AR target genes (Supplementary Fig. 5d). Interestingly, ARLNC1 knockdown also had a significant effect on the mRNA and protein levels of AR (Fig. 4e,f), suggesting direct regulation of AR by ARLNC1. However, we found that ARLNC1 overexpression did not affect AR and its signaling cascade (Supplementary Fig. 5e).

In situ colocalization of ARLNC1 and *AR* transcripts. Noncoding RNAs have been shown to target mRNAs via direct or indirect RNA–RNA interaction[9,40–42]. To identify target mRNAs that could interact with ARLNC1, we performed unbiased prediction of RNA–RNA interactions using IntraRNA[43,44]. Interestingly, the 3′ UTR of the *AR* transcript was identified as a target of ARLNC1 (Fig. 5a and Supplementary Fig. 6a). An in vitro RNA–RNA interaction assay between the 3′ UTR of *AR* and full-length ARLNC1 confirmed this in silico prediction (Fig. 5b). To evaluate this interaction in the context of the cellular environment, multiplexed smFISH for *AR* and ARLNC1 transcripts was performed in MDA-PCa-2b cells. On co-staining of MDA-PCa-2b cells for either a combination of *AR* transcripts and a panel of lncRNAs or ARLNC1 and a panel of mRNAs, we observed specific colocalization of *AR* and ARLNC1 transcripts in the nucleus within foci that were typically larger than individual molecules (Fig. 5c–e). The extent of colocalization was much higher than that expected from coincidental colocalization with an abundant transcript, such as MALAT1 or *GAPDH* (Fig. 5c–e). More specifically, colocalization typically occurred at a stoichiometry of 2:1 ARLNC1/*AR*, which accounted for ~10–20% of all *AR* and ARLNC1 transcripts in the cell (Supplementary Fig. 6b). Furthermore, *AR*–ARLNC1 colocalization was observed in *ARLNC1*-positive prostate cancer tissues (Fig. 5f,g).

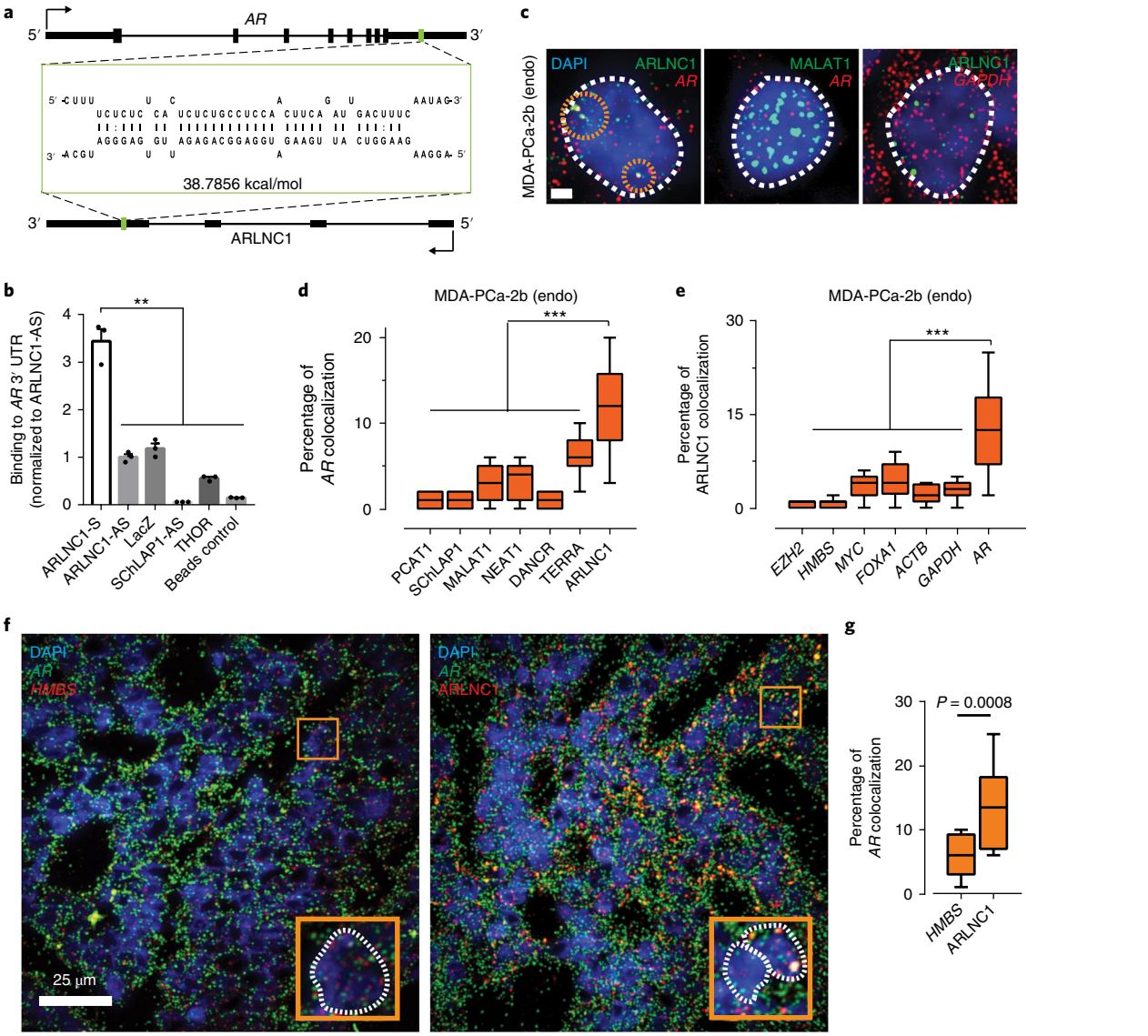


Fig. 5 | In situ colocalization of *AR* mRNA and ARLNC1 in prostate cancer cells. a, Schematic of the predicted RNA–RNA interaction between ARLNC1 and the 3′ UTR of *AR*. **b**, ARLNC1 interacts with the *AR* 3′ UTR in an in vitro RNA–RNA interaction assay. Compared to a panel of control RNAs (ARLNC1 antisense, LacZ, SChLAP1-AS, THOR), ARLNC1 binds to *AR* 3′ UTR-1–980 with high affinity. The binding affinity was quantified by qPCR analysis of the *AR* 3′ UTR. Data were normalized to the ARLNC1-AS control. Mean ± s.e.m. are shown, $n = 3$. $^{**}P < 0.001$, by two-tailed Student's *t* test. **c**–**e**, smFISH depiction of *AR*–ARLNC1 colocalization in situ. **c**, Representative pseudocolored images of MDA-PCa-2b cell nuclei stained for the appropriate endogenous (endo) transcripts (green, red) and with DAPI (nucleus, blue). Scale bar, 5 μm. The orange circles represent regions of colocalization. **d**,**e**, Quantification of the percentage of *AR* or ARLNC1 molecules colocalizing with a panel of lncRNAs (**d**) or mRNAs (**e**). The center line and whiskers depict the median and range, respectively, and the box extends from the 25th to the 75th percentile ($n = 50$ cells for each sample aggregated from 3 independent experiments). $^{***}P < 0.0001$, by two-tailed Student's *t* test. **f**, Representative pseudocolored images of ARLNC1-positive prostate cancer tissues stained with DAPI (nucleus, blue) and for *AR* (green), *HMBS* (red) or ARLNC1 (red) transcripts (smFISH). Inset, 5.5 × 5.5 μm² zoomed-in view of the box within the main panel. **g**, Quantification of the percentage of *AR* molecules colocalizing with *HMBS* transcripts or ARLNC1. The center line and whiskers depict the median and range, respectively, and the box extends from the 25th to the 75th percentile ($n = 15$ field-of-views for each sample aggregated from 3 independent tissues). $P < 0.001$, by two-tailed Student's *t* test.

Using an in vitro RNA–RNA binding assay, we identified nucleotides 700–1300 of ARLNC1 to be critical for binding to the *AR* 3′ UTR (Fig. 6a,b). To confirm this observation within the cellular context, we ectopically overexpressed different fragments of ARLNC1 together with *AR* in U2OS osteosarcoma cells. In this exogenous system, colocalization of *AR* and ARLNC1 was once again demonstrated, wherein colocalization was dependent on the presence of nucleotides 700–1300 of ARLNC1 (Fig. 6c,d and Supplementary Fig. 6c). Furthermore, incubation with antisense oligonucleotides (ASOs) that blocked the interaction site led to a significant reduction in ARLNC1–*AR* interaction in vitro and in situ (Fig. 6e,f and Supplementary Fig. 6d,e). Decreased AR signaling was also observed following blocking of this interaction (Fig. 6g and Supplementary Fig. 6f).

ARLNC1 regulates the cytoplasmic levels of *AR* transcripts. We then sought to delineate the mechanism of ARLNC1-mediated AR regulation. We first monitored the stability of these two transcripts and found that *AR* and ARLNC1 have similar half-lives of ~9 h (Supplementary Fig. 6g). As ARLNC1 depletion resulted in a

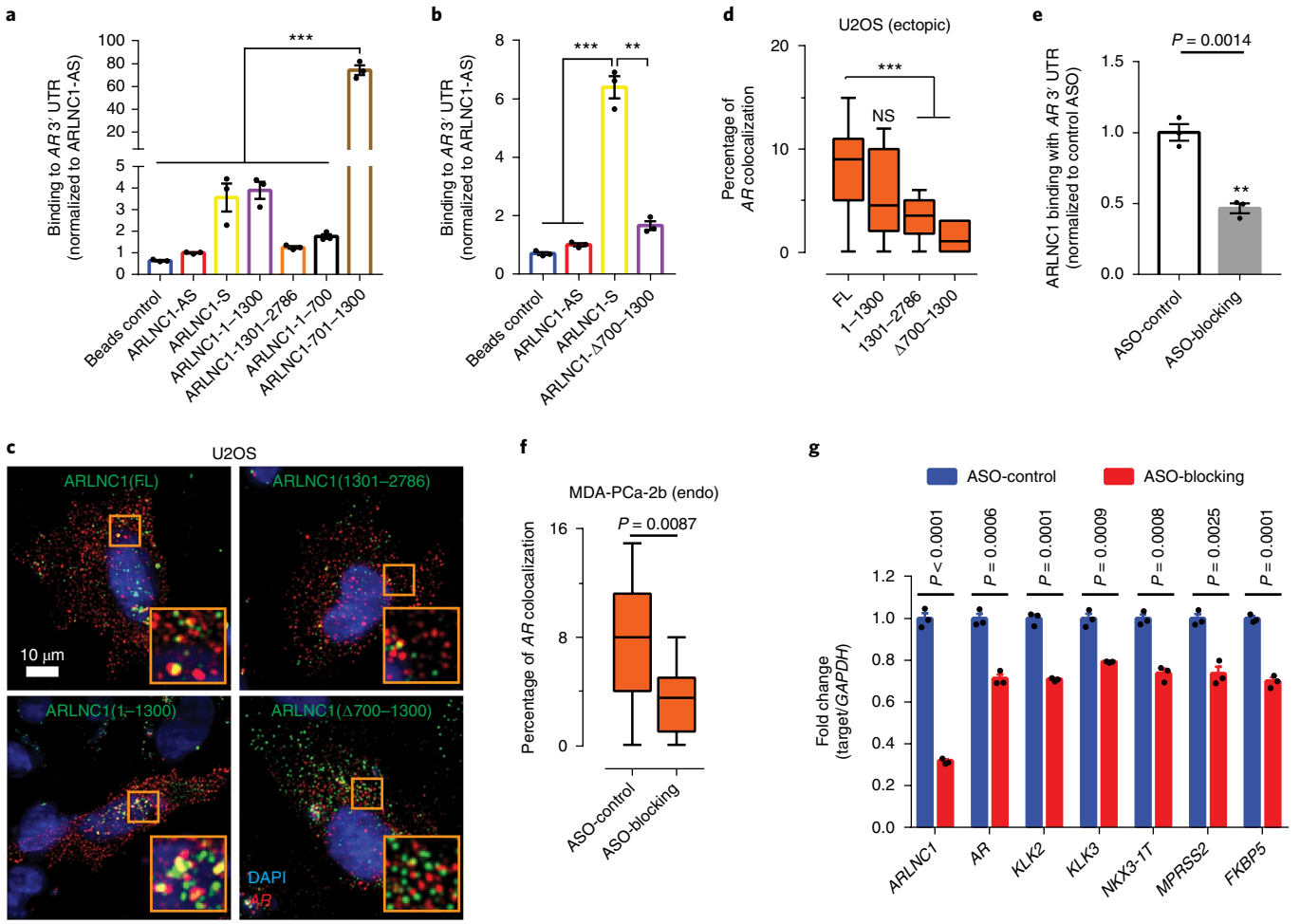


Fig. 6 | Identification of the ARLNC1 fragment mediating RNA–RNA interaction with *AR* mRNA. a, In vitro RNA–RNA interaction assay identifies nucleotides 700–1300 on ARLNC1 as critical binding site to *AR* 3′ UTR-1–980. ARLNC1 fragments covering nucleotides 700–1300 display comparable or higher *AR* 3′ UTR binding affinity than ARLNC1-S, with ARLNC1-700–1300 exhibiting the highest binding affinity. Data were normalized to the ARLNC1-AS control. Mean ± s.e.m. values are shown, $n = 3$. ***P (adjusted) = 0.0001, determined by ANOVA with Dunnett's multiple-comparisons test. **b**, Deletion of nucleotides 700–1300 on ARLNC1 results in impaired binding to the *AR* 3′ UTR, as shown by in vitro RNA–RNA interaction assay. Data were normalized to the ARLNC1-AS control. Mean ± s.e.m. values are shown, $n = 3$. ***$P = 0.0001$, **$P = 0.0003$, by two-tailed Student's t test. **c**,**d**, smFISH shows that nucleotides 700–1300 in ARLNC1 are important for colocalization in situ. **c**, Representative pseudocolored images of U2OS cells stained with DAPI (nucleus, blue) and for ARLNC1 (green) and *AR* transcripts (red). Inset, 10 × 10 μm² zoomed-in view of the orange box in the main image. **d**, Quantification of the percentage of *AR* molecules colocalizing with various ARLNC1 fragments. The center line and whiskers depict the median and range, respectively, and the box extends from the 25th to the 75th percentile ($n = 50$ cells for each sample aggregated from 3 independent experiments). ***$P < 0.0001$, by two-tailed Student's t test. NS, not significant. **e**, ASOs targeting nucleotides 700–1300 on the ARLNC1 transcript (ASO-blocking pool) inhibit ARLNC1 interaction with the *AR* 3′ UTR. In vitro RNA–RNA interaction assays were performed using ARLNC1 and the *AR* 3′ UTR, with the addition of the blocking ASO pool or control ASO. Data were normalized to the control ASO. Mean ± s.e.m. values are shown, $n = 3$. $P = 0.0014$, by two-tailed Student's t test. **f**, smFISH shows that ASOs targeting nucleotides 700–1300 on the ARLNC1 transcript (ASO-blocking) inhibit ARLNC1 colocalization with *AR* in situ. Quantification is shown of the percentage of *AR* transcripts colocalizing with ARLNC1 after various treatments in MDA-PCa-2b cells. The center line and whiskers depict the median and range, respectively, and the box extends from the 25th to the 75th percentile ($n = 50$ cells for each sample aggregated from 3 independent experiments). The P value was computed by two-tailed Student's t test. **g**, qPCR analysis of *ARLNC1*, *AR* and AR signaling genes (*KLK2*, *KLK3*, *NKX3-1*, *TMPRSS2* and *FKBP5*) in MDA-PCa-2b cells transfected with control or blocking ASOs targeting the interaction site between ARLNC1 and the *AR* 3′ UTR. Mean ± s.e.m. values are shown, $n = 3$. Significance was determined by two-tailed Student's t test.

striking reduction of AR protein levels, much more than could be explained by *AR* transcript reduction, we hypothesized that ARLNC1 could affect *AR* post-transcriptionally. To test this hypothesis, we tracked the subcellular localization of *AR* transcripts using smFISH after depleting ARLNC1. We confirmed successful in situ knockdown of ARLNC1 using siRNAs, antisense oligonucleotide (ASO) and the blocking oligonucleotides that targeted the ARLNC1–*AR* interaction (ASO-blocking) in MDA-PCa-2b cells (Supplementary Fig. 6h,i). Quantification of the subcellular distribution of ARLNC1 suggested that the nuclear fraction of ARLNC1 was enriched only in the *ARLNC1* siRNA (si-*ARLNC1*) condition (Supplementary Fig. 6j), a result expected for siRNAs that are typically more functional in the cytosol[45]. Surprisingly, ARLNC1 knockdown or obstruction of the *AR*–ARLNC1 interaction increased the nuclear *AR* fraction by dramatically decreasing cytoplasmic levels of the *AR* transcript (Fig. 7a,b and Supplementary Fig. 6k–l). This observation was further supported by BrU-seq and BrUChase-seq, two high-throughput tools that monitor transcript synthesis and stability. On ARLNC1

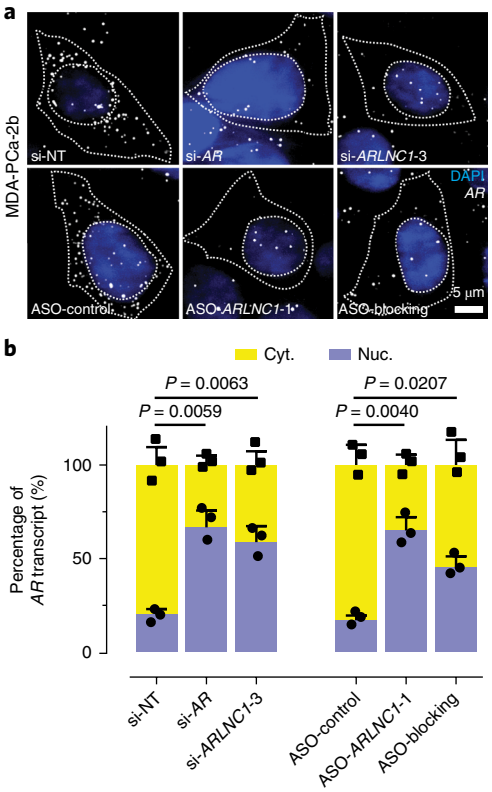


Fig. 7 | ARLNC1 regulates the cytoplasmic level of the *AR* transcript. **a**, ARLNC1 regulates *AR* post-transcriptionally by specifically affecting cytoplasmic *AR* mRNA. Representative pseudocolored images are shown of MDA-PCa-2b cells stained for DAPI (nucleus, blue) and *AR* (gray) after treatment with siRNA against *AR* (si-*AR*), siRNA against *ARLNC1* (si-*ARLNC1*-3), ASO against *ARLNC1* (ASO-*ARLNC1*-1) or blocking ASO against the *AR*–ARLNC1 colocalizing segment (ASO-blocking). Quantification of knockdown is shown in Supplementary Fig. 6k,l. **b**, Fractional column plots depicting the nucleo-cytoplasmic distribution of *AR* mRNA after the various treatment conditions in **a**, as computed using smFISH. Mean ± s.e.m. values are shown, $n = 3$ independent experiments and 60 cells analyzed for each sample. The *P* values were computed by comparing to si-NT- or ASO-control-treated cells, by two-tailed Student's *t* test.

knockdown, the synthesis rate of the *AR* transcript remained the same (Supplementary Fig. 6m), while the stability of the transcript decreased, particularly through the 3′ UTR (Supplementary Fig. 6n). Taken together, our data suggest that ARLNC1 regulates the cytoplasmic levels of *AR* transcripts. Furthermore, the transcriptional coupling between *AR* and ARLNC1 transcripts is mediated by direct interactions that are encoded in their sequences.

Inhibition of ARLNC1 delays prostate cancer growth in vitro and in vivo. Having established a role for ARLNC1 in the regulation of AR signaling, we further evaluated the biological effects of ARLNC1 in prostate cancer cell lines. GO pathway enrichment analysis of the knockdown microarray data showed that ARLNC1-regulated genes were involved in cell proliferation and apoptosis (Fig. 4a). Knockdown of ARLNC1 had a significant effect on the proliferation of AR-dependent MDA-PCa-2b and LNCaP cells, but had no effect on AR-negative DU145 and PC3 cells (Fig. 8a and Supplementary Fig. 7a,b). Knockdown of ARLNC1 also resulted in increased apoptosis in AR-positive prostate cancer cells (Fig. 8b and Supplementary Fig. 7c). Notably, these results translated to effects in vivo, as cells expressing shRNA targeting *ARLNC1* formed smaller tumors in mice when compared to cells expressing non-targeting shRNA (Fig. 8c), thus suggesting that ARLNC1 is an important survival factor for AR-dependent prostate cancer.

Because modulation of ARLNC1 expression levels resulted in a striking proliferation phenotype, we hypothesized that ARLNC1 inhibition could be used therapeutically for the treatment of prostate cancer. ASOs have recently been shown to be effective in targeting RNA in vivo[46–49]; thus, we designed ASOs targeting the ARLNC1 transcript (Supplementary Fig. 7d). Transfection of ASOs resulted in strong knockdown efficiency (Supplementary Fig. 7e), and ASO-mediated knockdown resulted in similar effects on gene expression profiling to siRNA (Fig. 8d,e and Supplementary Fig. 7f). Furthermore, AR-positive cells transfected with *ARLNC1* ASOs exhibited retarded growth, similar to those treated with siRNAs (Fig. 8f). To evaluate the therapeutic potential of *ARLNC1* ASOs in vivo, we first assessed the cellular free-uptake efficiency of *ARLNC1* ASOs, a prerequisite for ASO therapeutic use. Notably, several ASOs significantly reduced ARLNC1 levels through free uptake (Supplementary Fig. 7g). Free uptake of *ARLNC1* ASOs led to a significant decrease in the proliferation capacity of MDA-PCa-2b cells in both normal cell culture and 3D sphere conditions (Supplementary Fig. 7h–j). Treatment of mice bearing MDA-PCa-2b xenografts with *ARLNC1*-targeting ASOs led to significant decreases in tumor growth compared to control ASO (Fig. 8g,h and Supplementary Fig. 8a–e). Taken together, these data, along with the association of ARLNC1 with aggressive androgen signaling (Supplementary Fig. 8f–j), suggest that ARLNC1 plays a critical role in the proliferation of AR-dependent prostate cancer and can be effectively exploited as a therapeutic target.

Discussion

As AR signaling remains a significant driver of CRPC pathogenesis, it is imperative to generate novel strategies to target this pathway. Even with the addition of enzalutamide or abiraterone to CRPC treatment regimens, progression invariably occurs. Exploiting players other than AR itself that are pivotal to maintaining the magnitude of the androgen response is an alternative approach. Our comprehensive profiling of AR-regulated, prostate cancer–associated lncRNAs identified the top-ranking candidate ARLNC1 that we functionally characterized. We identified a positive feedback loop between ARLNC1 and AR that maintains the androgen transcriptional program in AR-positive prostate cancer cells, specifically through regulating the cellular levels of AR (Fig. 8i). The mechanism we identified echoes previous studies on lncRNAs—1/2-sbsRNAs[42], BACE1-AS[9] and TINCR[41], which highlights the role of lncRNA in increasing or decreasing RNA stability.

As a novel noncoding regulator of AR signaling, ARLNC1 has the potential to be not only a mechanistic biomarker but also a therapeutic target for advanced prostate cancer. In addition, the fact that it acts upstream of AR signaling presents the possibility that targeting ARLNC1 may afford an additional option to patients that have de novo or acquired resistance to therapies targeting AR itself (that is, enzalutamide or abiraterone). Furthermore, specific antisense nucleotides targeting ARLNC1, which we demonstrate to be specifically expressed in the prostate, could circumvent undesirable side effects that occur in other tissues with exposure to androgen synthesis inhibitors or anti-androgens.

Although we have identified a new node of the AR signaling network that can be therapeutically targeted, the molecular mechanism through which ARLNC1 regulates *AR* transcript levels remains to be fully characterized. At this time, it is unclear whether the physical interaction between the *AR* 3′ UTR and ARLNC1 functions with the aid of additional RNA-binding proteins (for example, HuR) and/or RNAs in vivo[50,51]. Nonetheless, the application of ASOs has ushered in an exciting era that makes it possible to target previously 'undruggable' molecules directly at the transcript level, such as ARLNC1, which is likely to yield promising opportunities in cancer treatment.

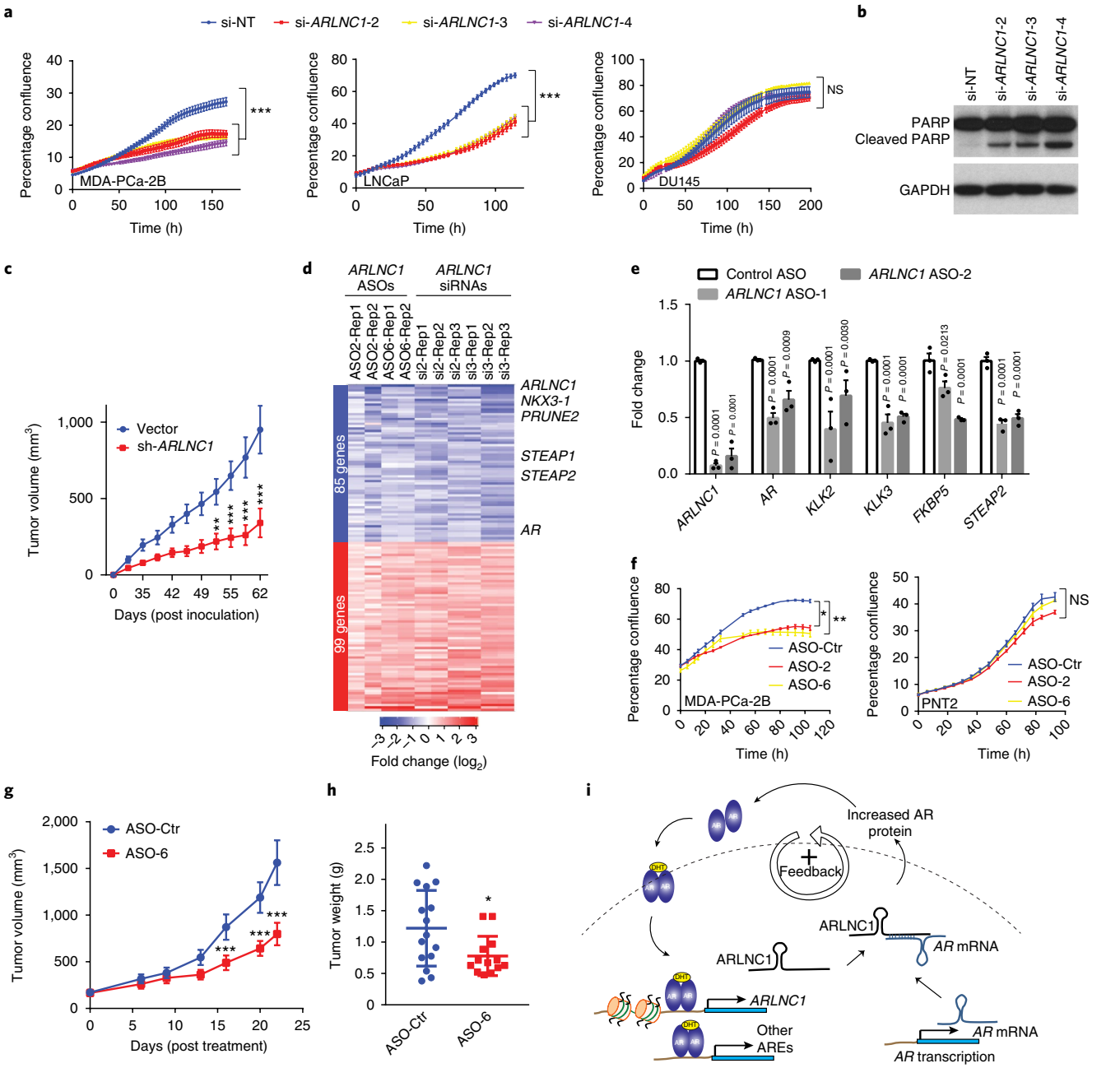


Fig. 8 | ARLNC1 as a therapeutic target in AR-positive prostate cancer models. a, siRNA-mediated knockdown of ARLNC1 in vitro in AR-positive prostate cancer cell lines (MDA-PCa-2b and LNCaP) inhibits cell proliferation. The AR-negative prostate cell line DU145 serves as a negative control. Mean ± s.d. values are shown, $n = 6$ independent cell cultures per group, $^{**}P$ (adjusted) = 0.0001 compared to si-NT-treated cells, by one-way ANOVA with Dunnett's multiple-comparisons test; NS, not significant. **b**, ARLNC1 loss leads to increased apoptosis as shown by western blot analysis of PARP and cleaved PARP in LNCaP cells following ARLNC1 knockdown. The experiment was repeated independently three times with similar results. Uncropped images are shown in Supplementary Fig. 9. **c**, Tumor growth of LNCaP-AR cells expressing shRNA targeting *ARLNC1* or shRNA vector. Mean ± s.e.m. values are shown. $n = 10$ independent tumors, $^{***}P < 0.0001$, $^{**}P = 0.0007$, as determined by two-tailed Student's t test. **d**, Gene expression profiling for siRNA-mediated or ASO-mediated ARLNC1 knockdown in MDA-PCa-2b cells. The numbers above the heat map represent the specific microarray replicates. **e**, qRT-PCR analysis of *ARLNC1*, *AR* and AR targets (*KLK2*, *KLK3*, *FKBP5* and *STEAP2*) in MDA-PCa-2b cells transfected with ASOs against *ARLNC1*. Data were normalized to a housekeeping gene, and the levels in control ASO-treated cells were set to 1. Mean ± s.e.m. values are shown, $n = 3$. Adjusted P values were determined by one-way ANOVA with Dunnett correction for multiple comparisons. **f**, Transfection of ASOs targeting *ARLNC1* in AR-positive MDA-PCa-2b cells inhibits cell proliferation. The AR-negative prostate cell line PNT2 serves as a negative control. Mean ± s.e.m. values are shown, $n = 6$ independent cell cultures per treatment group. $^{*}P$ (adjusted) = 0.0112, $^{**}P$ (adjusted) = 0.0065, NS: not significant; compared to the control-ASO group by one-way ANOVA with Dunnett correction for multiple comparisons. **g**,**h**, Effect of ASO treatment on the growth of MDA-PCa-2b xenografts in male athymic nude mice, with control ASO ($n = 15$) or *ARLNC1* ASO ($n = 13$) treatment subcutaneously at 50 mg per kg body weight, five times per week for 3 weeks. Tumors were measured by caliper biweekly (**g**) and tumor weights were measured at the end point (**h**). Mean ± s.d. values are shown. $^{*}P = 0.0251$, $^{***}P < 0.0001$; compared to control ASO by two-tailed Student's t test. **i**, A model depicting the positive feedback loop between ARLNC1 and AR that is critical for prostate cancer growth.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0120-1.

Received: 7 February 2017; Accepted: 23 March 2018;
Published online: 28 May 2018

References

1. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159 (2009).
2. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914 (2011).
3. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
4. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
5. Lee, N., Moss, W. N., Yario, T. A. & Steitz, J. A. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. *Cell* **160**, 607–618 (2015).
6. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30**, 167–174 (2002).
7. Prensner, J. R. et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* **45**, 1392–1398 (2013).
8. Gupta, R. A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
9. Faghihi, M. A. et al. Expression of a noncoding RNA is elevated in Alzheimeras disease and drives rapid feed-forward regulation of β-secretase. *Nat. Med* **14**, 723–730 (2008).
10. Iyer, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
11. Malik, R. et al. The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer. *Mol. Cancer Res.* **12**, 1081–1087 (2014).
12. Shukla, S. et al. Identification and validation of PCAT14 as prognostic biomarker in prostate cancer. *Neoplasia* **18**, 489–499 (2016).
13. Lu-Yao, G. L. et al. Fifteen-year survival outcomes following primary androgen-deprivation therapy for localized prostate cancer. *JAMA Intern. Med.* **174**, 1460–1467 (2014).
14. Huggins, C. & Hodges, C. V. Studies on prostatic cancer. I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. *J. Urol.* **167**, 948–951 (2002).
15. The Veterans Administration Co-operative Urological Research Group. Treatment and survival of patients with cancer of the prostate. *Surg. Gynecol. Obstet.* **124**, 1011–1017 (1967).
16. Chen, Y., Sawyers, C. L. & Scher, H. I. Targeting the androgen receptor pathway in prostate cancer. *Curr. Opin. Pharmacol.* **8**, 440–448 (2008).
17. Wong, Y. N., Ferraldeschi, R., Attard, G. & de Bono, J. Evolution of androgen receptor targeted therapy for advanced prostate cancer. *Nat. Rev. Clin. Oncol.* **11**, 365–376 (2014).
18. Mukherji, D., Pezaro, C. J. & De-Bono, J. S. MDV3100 for the treatment of prostate cancer. *Expert Opin. Investig. Drugs* **21**, 227–233 (2012).
19. Scher, H. I. et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. *N. Engl. J. Med.* **367**, 1187–1197 (2012).
20. Tran, C. et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**, 787–790 (2009).
21. Scher, H. I. et al. Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet* **375**, 1437–1446 (2010).
22. Stein, M. N., Goodin, S. & Dipaola, R. S. Abiraterone in prostate cancer: a new angle to an old problem. *Clin. Cancer Res.* **18**, 1848–1854 (2012).
23. Reid, A. H. et al. Significant and sustained antitumor activity in post-docetaxel, castration-resistant prostate cancer with the CYP17 inhibitor abiraterone acetate. *J. Clin. Oncol.* **28**, 1489–1495 (2010).
24. de Bono, J. S. et al. Abiraterone and increased survival in metastatic prostate cancer. *N. Engl. J. Med.* **364**, 1995–2005 (2011).
25. Watson, P. A., Arora, V. K. & Sawyers, C. L. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* **15**, 701–711 (2015).
26. Antonarakis, E. S. et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N. Engl. J. Med.* **371**, 1028–1038 (2014).
27. Attard, G., Richards, J. & de Bono, J. S. New strategies in metastatic prostate cancer: targeting the androgen receptor signaling pathway. *Clin. Cancer Res.* **17**, 1649–1657 (2011).
28. Hearn, J. W. et al. HSD3B1 and resistance to androgen-deprivation therapy in prostate cancer: a retrospective, multicohort study. *Lancet Oncol.* **17**, 1435–1444 (2016).
29. Chan, S. C., Li, Y. & Dehm, S. M. Androgen receptor splice variants activate androgen receptor target genes and support aberrant prostate cancer cell growth independent of canonical androgen receptor nuclear localization signal. *J. Biol. Chem.* **287**, 19736–19749 (2012).
30. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
31. Visakorpi, T. et al. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat. Genet.* **9**, 401–406 (1995).
32. Asangani, I. A. et al. Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* **510**, 278–282 (2014).
33. Roche, P. J., Hoare, S. A. & Parker, M. G. A consensus DNA-binding site for the androgen receptor. *Mol. Endocrinol.* **6**, 2229–2235 (1992).
34. Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
35. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
36. Takayama, K. et al. Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. *EMBO J.* **32**, 1665–1680 (2013).
37. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
38. Mele, M. et al. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
39. Rhodes, D. R. et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166–180 (2007).
40. Engreitz, J. M. et al. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* **159**, 188–199 (2014).
41. Kretz, M. et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493**, 231–235 (2013).
42. Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470**, 284–288 (2011).
43. Gawronski, A. R. et al. MechRNA: prediction of lncRNA mechanisms from RNA-RNA and RNA-protein interactions. *Bioinformatics* (2018) https://doi.org/10.1093/bioinformatics/bty208
44. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017).
45. Lennox, K. A. & Behlke, M. A. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* **44**, 863–877 (2016).
46. Meng, L. et al. Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* **518**, 409–412 (2015).
47. Wheeler, T. M. et al. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature* **488**, 111–115 (2012).
48. Hua, Y. et al. Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev.* **24**, 1634–1644 (2010).
49. Evers, M. M., Toonen, L. J. & van Roon-Mom, W. M. Antisense oligonucleotides in therapy for neurodegenerative disorders. *Adv. Drug Deliv. Rev.* **87**, 90–103 (2015).
50. Yeap, B. B. et al. Novel binding of HuR and poly(C)-binding protein to a conserved UC-rich motif within the 3'-untranslated region of the androgen receptor messenger RNA. *J. Biol. Chem.* **277**, 27183–27192 (2002).
51. Lebedeva, S. et al. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* **43**, 340–352 (2011).

Acknowledgements

We thank A. Poliakov, A. Parolia, V. Kothari and J. Siddiqui for helpful discussions, the University of Michigan Sequencing Core for Sanger sequencing, H. Johansson (LGC-Biosearch) for initial assistance with smFISH probe design, and S. Ellison, S. Gao and K. Giles for critically reading the manuscript and submitting documents. This work was supported in part by NCI Prostate SPORE (P50CA186786 to A.M.C.) and EDRN (U01 CA214170 to A.M.C.) grants. A.M.C. is also supported by the Prostate Cancer Foundation and by the Howard Hughes Medical Institute. A.M.C. is an American Cancer Society Research Professor and a Taubman Scholar of the University of Michigan. R. Malik was supported by a Department of Defense Postdoctoral Award (W81XWH-13-1-0284). Y.Z. is supported by a Department of Defense Early Investigator Research Award (W81XWH-17-1-0134). R. Malik, M.C., Y.S.N., J.C.-Y.T. and Y.Q. were supported by the Prostate Cancer Foundation Young Investigator Award. R. Mehra was supported by a Department of Defense Idea Development Award (W81XWH-16-1-0314). Y.S.N. is supported by a University of Michigan Cellular and Molecular Biology National Research Service Award Institutional Predoctoral Training Grant. S.P. was supported by an AACR-Bayer Prostate Cancer Research Fellowship (16-40-44-PITC). L.X. is supported by a US Department of Defense Postdoctoral Fellowship (W81XWH-16-1-0195). M.B. was supported by NIH DP5 grant OD012160. G.C.S. was supported by the Department of Defense awards W81XWH-14-1-0508 and

W81XWH-14-1-0509. M.U. was funded by the German Research Foundation (DFG grant BA2168/11-1 SPP 1738).

Author contributions

R. Malik, Y.Z., M.C., S.P. and A.M.C. conceived the study and designed the research. Y.Z. and R. Malik performed most of the cellular and molecular biology experiments with the assistance of Y.H., S.Y., S.S., S.K.S., L.X., X.J., S.M.D., X.C., J.T.W. and F.Y.F. M.C. performed most of the bioinformatics analyses with the help of Y.S.N. and M.K.I. S.P., U.P. and M.B. performed all smFISH work, and S.P. performed the mechanistic work-up. J.C.-Y.T. and K.M.J. carried out the in vivo mouse xenograft studies, and Y.Q. performed the 3D sphere model work. L.P.K. performed the histopathological analyses. L.W. and R. Mehra carried out RNA ISH on tissue microarrays, and T.-Y.L. and H.J. performed the statistical analysis for this technique. M.U., A.R.G., R.B. and C.S.S. performed the in silico binding predictions. S.M.F., A.T.W. and S.G. provided ASOs. G.C.S. provided the AR expression construct. M.T.P. and M.L. performed BrU and BrUChase sample preparation. Y.Z., M.C., R. Malik, S.P. and A.M.C. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests

The University of Michigan has filed a patent on lncRNAs as biomarkers of cancer, and A.M.C., R. Malik, Y.Z., M.C. and S.P. are named as co-inventors. A.M.C. is a co-founder of LynxDx, which is developing lncRNA biomarkers. S.M.F., A.T.W. and S.G. are employees of Ionis Pharmaceuticals, which developed the ASOs against *ARLNC1* that were used in this study.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0120-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.M.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Cell lines. Cell lines were purchased from the American Type Culture Collection (ATCC) and maintained using standard media and conditions. All cell lines were genotyped by DNA fingerprinting analysis and tested for mycoplasma infection every 2 weeks. All cell lines used in this study were mycoplasma negative. For androgen stimulation experiments, VCaP and LNCaP cells were grown in medium supplemented with charcoal-stripped serum for 48 h and then stimulated with 10 nM DHT (Sigma-Aldrich) for 6 or 24 h.

RNA-seq. Total RNA was extracted from LNCaP and VCaP cells following DHT treatment, using the miRNeasy kit (Qiagen). RNA quality was assessed using an Agilent Bioanalyzer. Each sample was sequenced using the Illumina HiSeq 2000 (with a 100-nt read length) according to published protocols[52].

RNA-seq data analysis to identify AR-regulated genes. RNA-seq data were analyzed as previously described[53]. Briefly, the strand-specific paired-end reads were inspected for sequencing and data quality (for example, insert size, sequencing adaptor contamination, rRNA content, sequencing error rate). Libraries passing quality control were trimmed of sequencing adaptors and aligned to the human reference genome, GRCh38. Expression was quantified at the gene level using the 'intersection non-empty' mode[54] as implemented in featureCounts[55] using the Gencode v22[56] and/or MiTranscriptome[10] assemblies. All pairwise differential expression analyses were carried out using the voom–limma approach[57,58] with all default parameters. Relative expression levels (FPKM) were normalized for differences in sequencing depth using scaling factors obtained from the calcNormFactors (default parameters) function from edgeR[59].

ARGs were identified from expression data for VCaP and LNCaP cells treated with DHT for 6 and 24 h using three linear models: separate models for each of the cell lines treating the two time points as biological replicates and a merged model with all treated samples as replicates. ARGs were defined as genes that were significant (P value < 0.1 and absolute log fold change > 2) in both separate models and/or the merged model.

Identification of prostate cancer–associated protein-coding genes and lncRNAs. Raw RNA-seq data for patients with primary and metastatic prostate cancer were obtained from the TCGA/PRAD and PCF/SU2C projects, respectively. External transcriptome samples were reanalyzed using in-house pipelines (see above) to facilitate direct comparisons of expression levels and identification of differentially expressed genes. Pan-cancer analyses based on the MiTranscriptome assembly[10] were leveraged as FPKMs, and enrichment scores (SSEA) were computed as part of that project. Tissue lineage (prostate) and prostate cancer–specific genes were identified using the SSEA method as previously described[10]. Briefly, the SSEA test was used to determine whether each gene was significantly associated with a set of samples (for example, prostate cancer) or cancer progression in a given lineage (for example, prostate normal to prostate cancer). The genes were ranked according to their strength of association.

Oncomine concept analysis of the ARLNC1 signature. Genes with expression levels significantly correlated with ARLNC1 were separated into positively and negatively correlated gene lists. These two lists were then imported into Oncomine as custom concepts and queried for association with other prostate cancer concepts housed in Oncomine. All of the prostate cancer concepts with odds ratio > 2.0 and P value < 1×10^{-4} were selected. Top concepts (based on odds ratios) were selected for representation. We exported these results as the nodes and edges of a concept association network and visualized the network using Cytoscape version 3.3.0. Node positions were computed using the edge-weighted force-directed layout in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of Mode labels[60].

ChIP–seq data analysis. ChIP–seq data from published external and in-house datasets, GSE56288 and GSE55064, were reanalyzed using a standard pipeline. Briefly, groomed reads (vendor quality control, adaptor removal) were aligned to the GRCh38 reference genome using STAR settings that disable spliced alignment: outFilterMismatchNoverLmax: 0.05, outFilterMatchNmin: 16, outFilterScoreMinOverLread: 0, outFilterMatchNminOverLread: 0, alignIntronMax: 1. Improperly paired alignments and non-primary alignments were discarded. Peaks were called using MACS2 (callpeak --broad --qvalue 0.05 --broad-cutoff 0.05 and callpeak --call-summits --qvalue 0.05)[61] and Q (-n 100000)[62]. ChIP enrichment plots were computed from alignment coverage files (BigWig[63]) as trimmed (trim = 0.05) smooth splines (spar = 0.05). The baseline (non-specific) ChIP signal was estimated from genomic windows furthest from the center of the queried region (peak summit, TSS) and subtracted from each signal before plotting.

AR binding motif search. An unsupervised motif search was carried out using MEME[64]. DNA sequences (GRCh38) from the uni-peak ChIP–seq regions overlapping promoters (5 kb upstream, 1 kb downstream of the assembled or known TSS) of ARGs were used as input to MEME (default parameters).

ChIP–qPCR assays. AR, FOXA1 or NKX3-1 ChIP was performed following our previous protocol[32]. (Antibodies: AR, Millipore cat. no. 06-680; FOXA1, Thermo Fisher cat. no. PA5-27157; NKX3-1, CST cat. no. 83700S.) qPCR analysis was performed using the primers listed in Supplementary Table 3. Primers targeting the *CYP2B7* promoter were purchased from CST (cat. no. 84846).

RNA ISH on tissue microarray. ISH assays were performed on tissue microarray sections from Advanced Cell Diagnostics as described previously[7]. In total, 133 tissue samples were included (11 from benign prostate, 85 from localized prostate cancer and 37 from metastatic prostate cancer). ARLNC1 ISH signals were examined in morphologically intact cells and scored manually by a study pathologist, using a previously described expression value scoring system[65]. For each tissue sample, the ARLNC1 product score was averaged across evaluable tissue microarray cores. Mean ARLNC1 product scores are plotted in Fig. 2e.

RACE. 5′ and 3′ RACE were performed to determine the transcriptional start and termination sites of *ARLNC1*, using the GeneRacer RLM-RACE kit (Invitrogen), according to the manufacturer's instructions.

Northern blot analysis. The NorthernMax-Gly Kit (Ambion) was used for ARLNC1 detection following the manufacturer's protocol. Briefly, 20 µg of total RNA was resolved on a 1% agarose glyoxal gel and then transferred to nylon membrane (Roche), cross-linked to the membrane (UV Stratalinker 1800; Stratagene) and the membrane was prehybridized. Overnight hybridization was performed with an ARLNC1-specific ^{32}P-labeled RNA probe. Membranes were exposed to HyBlot CL autoradiography film (Denville Scientific). The primer sequences used for generating the probes are given in Supplementary Table 3.

RNA isolation and cDNA synthesis. Total RNA from cell lines was isolated using QIAzol Lysis reagent (Qiagen) and the miRNeasy kit (Qiagen) with DNase digestion according to the manufacturer's instructions. cDNA was synthesized using Superscript III (Invitrogen) and random primers (Invitrogen).

qRT–PCR analysis. Relative RNA levels determined by qRT–PCR were measured on an Applied Biosystems 7900HT Real-Time PCR System, using Power SYBR Green MasterMix (Applied Biosystems). All primers were obtained from Integrated DNA Technologies, and gene-specific sequences are listed in Supplementary Table 3. *GAPDH*, *HMBS* or *ACTB* was used as an internal control for quantification of gene targets. The relative expression of RNAs was calculated using the ΔΔCt method.

Cytoplasmic and nuclear RNA purification. Cell fractionation was performed using the NE-PER nuclear extraction kit (Thermo Scientific) according to the manufacturer's instructions. RNA was extracted using the previously mentioned protocol.

siRNA-mediated knockdown. siRNA oligonucleotides targeting *ARLNC1*, *AR*, *FOXA1*, *BRD4*, *NKX3-1*, *LSD1*, *IRF1*, *POU1F1* or *EZH2* and a non-targeting siRNA were purchased from Dharmacon. (si-*AR*-pool, cat. no. L-003400-00-0005; si-*FOXA1*, cat. no. LU-010319-00-0005; si-*BRD4*, cat. no. LU-004937-00-0002; si-*NKX3-1*, cat. no. LU-015422-00-0005; si-*LSD1*, cat. no. LU-009223-00-0002; si-*IRF1*, cat. no. LU-011704-00-0005; si-*POU1F1*, cat. no. LU-012546-00-0005; si-*EZH2*, cat. no. L-004218-00-0005; si-NT, cat. no. D-001810-01-05.) siRNA sequences for *ARLNC1* knockdown are listed in Supplementary Table 3. For *AR* knockdown, two more siRNAs were purchased from Life Technologies (no. HSS179972 and no. HSS179973). Transfections with siRNA (50 nM) were performed with Lipofectamine RNAiMAX according to the manufacturer's instructions. RNA and protein were harvested for analysis 72 h after transfection.

ASO-mediated knockdown. ASOs targeting *ARLNC1* were obtained from Ionis Pharmaceuticals. Transfections with ASOs (50 nM) were performed with Lipofectamine RNAiMAX according to the manufacturer's instructions. RNA and protein were harvested for analysis 72 h after transfection.

Gene expression profiling. Total RNA was extracted following the aforementioned protocol. RNA integrity was assessed using an Agilent Bioanalyzer. Microarray analysis was carried out on the Agilent Whole Human Oligo Microarray platform, according to the manufacturer's protocol. siRNA-mediated knockdown experiments were run in triplicate, comparing knockdown samples treated with two independent *ARLNC1* siRNAs to samples treated with non-targeting control siRNA. ASO-mediated knockdown experiments were run in replicate, comparing knockdown samples treated with two *ARLNC1* ASOs to samples treated with non-targeting control. An AR signature was generated using MDA-PCa-2b cells treated with 10 nM DHT in triplicate.

Analysis of Agilent 44k microarrays was carried out using limma and included background subtraction (bc.method = 'half', offset = 100) and within-array normalization (method = 'loess'). Between-array quantile normalization of average expression levels (but not log-transformed fold change) was performed using the function normalizeBetweenArrays (method = 'Aquantile'). Control probes and

probes with missing values were excluded from further analyses. Probes were annotated to Gencode v22 genes using the mapping downloaded from Ensembl (efg_agilent_wholegenome_4×44k_v2). Probes originally annotated as AK093002 were used to detect ARLNC1. Differentially expressed genes following *ARLNC1* knockdown in MDA-PCA-2b cells were identified from triplicate biological repeats using adjusted P value < 0.1 and absolute log fold change > 0.6 cutoffs. Consensus targets of *ARLNC1* knockdown using siRNA and ASOs were identified using a merged linear model (all ten samples treated as replicates) and a P value < 0.001 cutoff.

GSEA. Enrichment analyses for custom and experimentally derived signatures (that is, AR targets, genes upregulated and downregulated following DHT treatment) were carried out using the non-parametric GSEA software with all default settings. For GO term enrichment, we applied the parametric randomSet[66] enrichment statistic to voom–limma-estimated fold changes (see above).

Overexpression of ARLNC1. Full-length ARLNC1 was amplified from MDA-PCa-2b cells and cloned into the pCDH clone and expression vector (System Biosciences). Insert sequences were validated by Sanger sequencing at the University of Michigan Sequencing Core. The full-length sequence for ARLNC1 expression is listed in Supplementary Table 4.

smFISH. smFISH and image analysis were performed as described previously[67,68]. Probe sequences targeting ARLNC1, PCAT1, DANCR, *AR*, *EZH2* and *FOXA1* were designed using the probe design software at https://www.biosearchtech.com/stellaris-designer and are listed in Supplementary Table 5. TERRA probes were designed as described previously[69]. Other probes were purchased directly from LGC-Biosearch. U2OS cells were seeded in six-well dishes and transfected with ARLNC1-expression vector alone or in combination with AR expression vector, using Fugene-HD (Promega) according to the manufacturer's protocol. Cells were incubated for 24 h, reseeded into eight-well chambered coverglasses, and fixed in formaldehyde for smFISH (as described above) after 24 h.

RNA in vitro transcription. Linearized DNA templates for full-length ARLNC1, ARLNC1 fragments, ARLNC1 deletion, antisense ARLNC1, LacZ, SChLAP1-AS, THOR and AR-3′ UTR–1–980 were synthesized using T7-containing primers. In vitro transcription assays were performed with T7 RNA polymerase (Promega) according to the manufacturer's instructions. For BrU-labeled RNA synthesis, 5-bromo-UTP was added to the transcription mix. At the end of transcription, DNA templates were removed by Turbo DNase (Thermo Fisher), and RNA was recovered using the RNA Clean and Concentrator Kit (Promega). RNA size and quality were further confirmed by Agilent Bioanalyzer.

RNA–RNA in vitro interaction assays. Twenty-five microlitres of Protein A/G Magnetic Beads (Pierce) was washed twice with RIP wash buffer (Millipore, cat. no. CS203177) before incubating with BrU antibody for 1 h at room temperature. After antibody conjugation, beads were washed twice with RIP wash buffer and then resuspended in incubation buffer containing RIP wash buffer, 17.5 mM EDTA (Millipore, cat. no. CS203175) and RNase Inhibitor (Millipore, cat. no. CS203219). Equal amounts (5 pmol) of BrU-labeled RNAs (ARLNC1, ARLNC1-AS, ARLNC1-1–1300, ARLNC1-1301–2786, ARLNC1-1–700, ARLNC1-701–1300, ARLNC1-del-701–1300, *LacZ*, SChLAP1-AS, THOR) were incubated with beads in Incubation Buffer for 2 h at 4 °C. Following incubation, 2.5 pmol of the *AR* 3′ UTR–1–980 RNA fragment was added into individual tubes and incubated overnight at 4 °C. After incubation, beads were washed six times with RIP Wash Buffer. To recover RNA, beads were digested with proteinase K buffer containing RIP Wash Buffer, 1% SDS (Millipore, cat. no. CS203174) and 1.2 µg/µl proteinase K (Millipore, cat. no. CS203218) at 55 °C for 30 min with shaking. After digestion, RNA was extracted from supernatant using the miRNeasy kit (Qiagen), and reverse transcription was performed using the Superscript III system (Invitrogen). The amount of *AR* 3′ UTR–1–980 recovered in each interaction assay was quantified by qPCR analysis. Data were normalized to the ARLNC1-AS control, using the ΔCt method. We designed ASOs blocking the *AR*–ARLNC1 interaction site (ASO-blocking, Ionis Pharmaceuticals) and used them in the in vitro interaction assays. Data were normalized to the control ASO, using the ΔCt method.

RNA stability assays. LNCaP cells were treated with 5 µg/ml actinomycin D for various times as indicated. RNA was extracted and qRT–PCR was carried out as described above. RNA half-life ($t_{1/2}$) was calculated by linear regression analysis.

Cell proliferation assays. Cells treated with siRNAs or ASOs were seeded into 24-well plates and allowed to attach. Cell proliferation was recorded by IncuCyte live-cell imaging system (Essen Biosciences), following the manufacturer's instructions.

Apoptosis analysis. Cells were grown in six-well plates and transfected with nonspecific siRNA or siRNAs targeting *ARLNC1*. Apoptosis analysis was performed 48 h after transfection, using the Dead Cell Apoptosis Kit (Molecular Probes no. V13241) according to the manufacturer's instructions.

Immunoblot analysis. Cells were lysed in RIPA lysis and extraction buffer (Thermo Scientific no. 89900) supplemented with protease inhibitor cocktail (Roche no. 11836170001). Protein concentrations were quantified using the DC protein assay (Bio-Rad), and protein lysates were boiled in sample buffer. Protein extracts were then loaded and separated on SDS–PAGE gels. Blotting analysis was performed with standard protocols using polyvinylidene difluoride membrane (GE Healthcare). Membranes were blocked for 60 min in blocking buffer (5% milk in a solution of 0.1% Tween-20 in Tris-buffered saline (TBS-T)) and then incubated overnight at 4 °C with primary antibody. After three washes with TBS-T, membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibody. Signals were visualized with an enhanced chemiluminescence system as described by the manufacturer (Thermo Scientific Pierce ECL Western Blotting Substrate). Primary antibodies used were as follows: androgen receptor (1:1,000 dilution, Millipore, no. 06-680, rabbit), GAPDH (1:5,000 dilution, Cell Signaling, no. 3683, rabbit), PSA (KLK3) (1:5,000 dilution, Dako, no. A0562, rabbit) and cleaved PARP (1:1,000 dilution, Cell Signaling, no. 9542, rabbit).

Androgen receptor reporter gene assays. Dual-luciferase reporter assays were performed using the Cignal Androgen Receptor Reporter Kit (Qiagen) according to the manufacturer's instructions. Briefly, cells were cotransfected with siRNAs (nonspecific, targeting *AR* or *ARLNC1*) and reporter vectors (negative control or AR reporter), using Lipofectamine 2000 transfection reagent (Thermo Fisher Scientific). Forty hours after transfection, DHT (or ethanol vehicle control) was added to induce AR signaling. The Dual-Luciferase assay was conducted 8 h after DHT stimulation, using the Dual-Luciferase Reporter Assay System from Promega (cat no. 1910). Reporter activity was analyzed on the basis of the ratio of firefly/*Renilla* activity to normalize for cell number and transfection efficiency.

In vivo experiments. All experiments were approved by the University of Michigan Institutional Animal Care and Use Committee. For tumor generation with shRNA-mediated knockdown, shRNA targeting *ARLNC1* was cloned into pSIH1-H1-copGFP-T2A-Puro (System Biosciences). Lentiviral particles were generated at the University of Michigan Vector Core. LNCaP-AR cells were infected with lentivirus expressing *ARLNC1* shRNA for 48 h. Knockdown of ARLNC1 was confirmed by qPCR analysis. Male athymic nude mice were randomized into two groups at 6 to 8 weeks of age. Five million cells expressing sh-*ARLNC1* or sh-vector were injected into bilateral flanks of mice. Caliper measurements were taken in two dimensions twice a week by an investigator blinded to the study objective and used to calculate tumor volume. The study was terminated when the tumor volume reached 1,000 mm^3. For ASO treatment in vivo, 6- to 8-week-old male athymic nude mice were inoculated subcutaneously with MDA-PCa-2b cells suspended in a Matrigel scaffold in the posterior dorsal flank region (5 million cells per site, two sites per animal). When the mean tumor volume reached approximately 150 mm^3, mice were randomized into two groups and treated with *ARLNC1*-specific or control ASO. ASOs, dosed at 50 mg per kg body weight, were subcutaneously injected between the scapulae once daily for three periods of 5 d on/2 d off. Tumor size was measured twice per week using a digital caliper by a researcher blinded to the study design. Mouse body weights were monitored throughout the dosing period. When the average tumor size in the control group reached 1,500 mm^3, mice were euthanized and the primary tumors were excised for weight determination. One-third of the resected specimen was placed in 10% formalin buffer, and the remaining tissue was snap-frozen.

BrU-seq and BrUChase-seq. BrU-seq and BrUChase-seq assays were performed as previously described[70,71] with MDA-PCa-2b cells treated with either si-NT or si-*ARLNC1*. BrU labeling was performed for 30 min, and chase experiments were performed for 6 h.

Statistical analysis. For in vivo experiments, power analysis (GPOWER software) performed for each tumor type tested to date indicates that the sample size we chose yields a statistical power >90% for detection of tumor size reduction of 40%. Sample sizes were not predetermined for all other assays. For in vivo experiments, animals were randomized. Randomization was not performed for all other assays. Statistical analyses were performed using GraphPad Prism software or using R. Data are presented as either means ± s.e.m. or means ± s.d. All of the experiments were performed in biological triplicate unless otherwise specified. Statistical analyses shown in figures represent two-tailed t tests, one-way ANOVA, two-way ANOVA or Kruskal–Wallis rank-sum tests as indicated. $P < 0.05$ was considered to be statistically significant. Details regarding the statistical methods employed during microarray, RNA-seq and ChIP–seq data analysis were included in the aforementioned methods for bioinformatics analyses.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Software for transcriptome meta-assembly and lncRNA discovery is available at https://tacorna.github.io/.

Data availability. RNA-seq and microarray data sets generated from this study have been deposited into the Gene Expression Omnibus, with accession GSE110905. Other data supporting the findings of this study are included in the Supplementary Information.

References

52. Prensner, J. R. et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* **29**, 742–749 (2011).
53. Cieslik, M. et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372–1381 (2015).
54. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
55. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
56. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
57. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
58. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
59. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
60. Cline, M. S. et al. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2**, 2366–2382 (2007).
61. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
62. Hansen, P. et al. Saturation analysis of ChIP-seq data for reproducible identification of binding peaks. *Genome Res.* **25**, 1391–1400 (2015).
63. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
64. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
65. Mehra, R. et al. A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia* **16**, 1121–1127 (2014).
66. Newton, M. A., Quintana, F. A., Den Boon, J. A., Sengupta, S. & Ahlquist, P. Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann. Appl. Stat.* **1**, 85–106 (2007).
67. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
68. Niknafs, Y. S. et al. The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. *Nat. Commun.* **7**, 12791 (2016).
69. Rossiello, F. et al. DNA damage response inhibition at dysfunctional telomeres by modulation of telomeric DNA damage response RNAs. *Nat. Commun.* **8**, 13980 (2017).
70. Paulsen, M. T. et al. Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl Acad. Sci. USA* **110**, 2240–2245 (2013).
71. Paulsen, M. T. et al. Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* **67**, 45–54 (2014).

nature research

Corresponding author(s): Arul M. Chinnaiyan

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	RNA seq, Microarray and qRT-PCR data were collected using vendor's software on Illumina HiSeq 2000, Agilent Whole Human Oligo Microarray, Applied Biosystems 7900HT Real-Time PCR platforms respectively. Microscopy images were acquired using Metamorph.
Data analysis	Software for transcriptome meta-assembly and lncRNAs discovery is available at https://tacorna.github.io/. Gene signatures were obtained using GSEA software. Image analysis was performed using custom-written macros in Image J and can be shared upon request. Statistical analysis was performed using Graphpad-Prism 6.0 and R.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq and microarray datasets generated from this study have been deposited into Gene Expression Omnibus, with accession number: GSE110905. Other data supporting the finding of this study are included in the Supplementary Information files. Software for transcriptome meta-assembly and lncRNAs discovery is available at https://tacorna.github.io/. We have no restrictions on data availability and data can be shared upon request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For in vivo experiments, power analysis (GPOWER software) performed for each tumor type tested to date indicates this animal number yields a statistical power >90% for detection of tumor size reduction of 40%. Sample sizes were not pre-determined for all other assays.
Data exclusions	No data were excluded.
Replication	All experiments were carried out at least 3 independent times for statistical reproducibility, unless otherwise stated, as represented by p-values.
Randomization	For in vivo experiments, animals were randomized. Randomization was not performed for all other assays.
Blinding	For in vivo experiments, tumor size was measured twice per week using a digital caliper by a researcher blinded to the study design. Blinding was not performed for all other assays.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☐	☒ Research animals
☒	☐ Human research participants

Antibodies

Antibodies used	Primary antibodies used in this study were: Androgen Receptor (1:1000 dilution, Millipore, #06-680, rabbit), GAPDH (1:5000 dilution, Cell Signaling, #3683, rabbit), PSA (1:5000 dilution, Dako, #A0562, rabbit), FOXA1 (Thermo Fisher Cat# PA5-27157) NKX3.1 (CST Cat# 83700S) and cleaved PARP (1:1000 dilution, Cell Signaling, #9542, rabbit).
Validation	All antibodies were validated by the vendors. Androgen receptor and PSA antibodies were also validated by siRNA treatment and androgen signaling assays respectively.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	All cell lines were purchased form ATCC.
Authentication	All cell lines were genotyped by STR profiling in house based on ATCC markers.

Mycoplasma contamination	All cells were tested for mycoplasma every two weeks.
Commonly misidentified lines (See ICLAC register)	Study does not include misidentified lines.

Research animals

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Animals/animal-derived materials	Male athymic nude mice were used in our in vivo studies.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging

nature COMMUNICATIONS

ARTICLE

Received 25 May 2016 | Accepted 1 Aug 2016 | Published 26 Sep 2016

DOI: 10.1038/ncomms12791 OPEN

The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression

Yashar S. Niknafs[1,2,*], Sumin Han[3,*], Teng Ma[3,4], Corey Speers[3,5,6], Chao Zhang[3], Kari Wilder-Romans[3], Matthew K. Iyer[1,7], Sethuramasundaram Pitchiaya[1], Rohit Malik[1], Yasuyuki Hosono[1], John R. Prensner[1], Anton Poliakov[1], Udit Singhal[1,8], Lanbo Xiao[1], Steven Kregel[1], Ronald F. Siebenaler[1], Shuang G. Zhao[3], Michael Uhl[9], Alexander Gawronski[10], Daniel F. Hayes[5,6,11], Lori J. Pierce[3,5,6], Xuhong Cao[1,8], Colin Collins[13], Rolf Backofen[9], Cenk S. Sahinalp[10,12,13], James M. Rae[5,6,11], Arul M. Chinnaiyan[1,2,5,6,8,14,15,**] & Felix Y. Feng[1,3,5,6,**,†]

Molecular classification of cancers into subtypes has resulted in an advance in our understanding of tumour biology and treatment response across multiple tumour types. However, to date, cancer profiling has largely focused on protein-coding genes, which comprise <1% of the genome. Here we leverage a compendium of 58,648 long noncoding RNAs (lncRNAs) to subtype 947 breast cancer samples. We show that lncRNA-based profiling categorizes breast tumours by their known molecular subtypes in breast cancer. We identify a cohort of breast cancer-associated and oestrogen-regulated lncRNAs, and investigate the role of the top prioritized oestrogen receptor (ER)-regulated lncRNA, *DSCAM-AS1*. We demonstrate that *DSCAM-AS1* mediates tumour progression and tamoxifen resistance and identify hnRNPL as an interacting protein involved in the mechanism of *DSCAM-AS1* action. By highlighting the role of *DSCAM-AS1* in breast cancer biology and treatment resistance, this study provides insight into the potential clinical implications of lncRNAs in breast cancer.

[1] Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA. [2] Department of Cellular and Molecular Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. [3] Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109, USA. [4] Department of Radiation Toxicology and Oncology, Beijing Key Laboratory for Radiobiology (BKLRB), Beijing Institute of Radiation Medicine, Beijing 100850, P. R. China. [5] Breast Oncology Program, University of Michigan, Ann Arbor, Michigan 48109, USA. [6] Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan 48109, USA. [7] Department of Computational Medicine and Bioinformatics, Ann Arbor, Michigan 48109, USA. [8] Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan 48109, USA. [9] Department of Computer Science and Centre for Biological Signaling Studies (BIOSS), University of Freiburg, Freiburg 79110, Germany. [10] School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6. [11] Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA. [12] School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405, USA. [13] Vancouver Prostate Centre, Vancouver, British Columbia, Canada V6H 3Z6. [14] Department of Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA. [15] Department of Urology, University of Michigan, Ann Arbor, Michigan 48109, USA. * These authors contributed equally to this work. ** These authors jointly supervised this work. † Present address: Departments of Radiation Oncology, Urology, and Medicine, Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, California 94115, USA. Correspondence and requests for materials should be addressed to A.M.C. (email: arul@umich.edu) or to F.Y.F. (email: felix.feng@ucsf.edu).

Long noncoding RNAs (LncRNAs) have recently been implicated in a variety of biological processes, including carcinogenesis and tumour growth[1–6]. Operating through a myriad of mechanisms[2], lncRNAs have challenged the central dogma of molecular biology as prominent functional RNA molecules. To investigate the role of lncRNAs in breast cancer, we interrogated the expression of lncRNAs across an RNA-sequencing (RNA-seq) breast tissue cohort comprised of 947 breast samples[7,8]. Previously, in a large-scale *ab initio* meta-assembly study from 6,503 RNA-seq libraries, we discovered ~45,000 of unannotated human lncRNAs[7], and this assembly was utilized for the present study. Building on prior work that has begun to investigate the role of lncRNAs in breast cancer[9], we set out to perform a comprehensive analysis of breast cancer tissue RNA-seq data to identify the lncRNAs potentially involved in breast cancer.

Patients with oestrogen receptor (ER)-positive breast cancer have better prognosis than those with ER-negative disease, based on both a more indolent natural history but perhaps more importantly due to effective anti-oestrogen, also designated 'endocrine,' therapy[10]. Despite the efficacy of endocrine therapy, however, the majority of breast cancer deaths occur in women with ER-positive breast cancers, because the incidence of ER-positive versus-negative disease is much higher (approximately 80 versus 20%), and because a substantial fraction of women either have inherent or acquired endocrine therapy-resistant disease[11].

Taken together, these considerations highlight the pressing need to understand the biology of the ER-driven breast cancers and their mechanism of resistance to endocrine therapy. The mechanism through which ER mediates cancer initiation and progression is an area of intense scientific investigation[12–14] that remains incompletely understood. In this regard, while substantial research has been focused on ER abnormalities, such as mutations in the gene encoding for ER (*ESR1*)[14,15] and on the co-existing activation pathways that might mediate resistance, such as HER2[16], few studies exist that interrogate ER-regulated noncoding RNAs[17–21]. Therefore, we set out to perform a comprehensive discovery and investigation of those lncRNAs that are driven by oestrogen in breast cancers drawing from a large human tissue RNA-seq cohort.

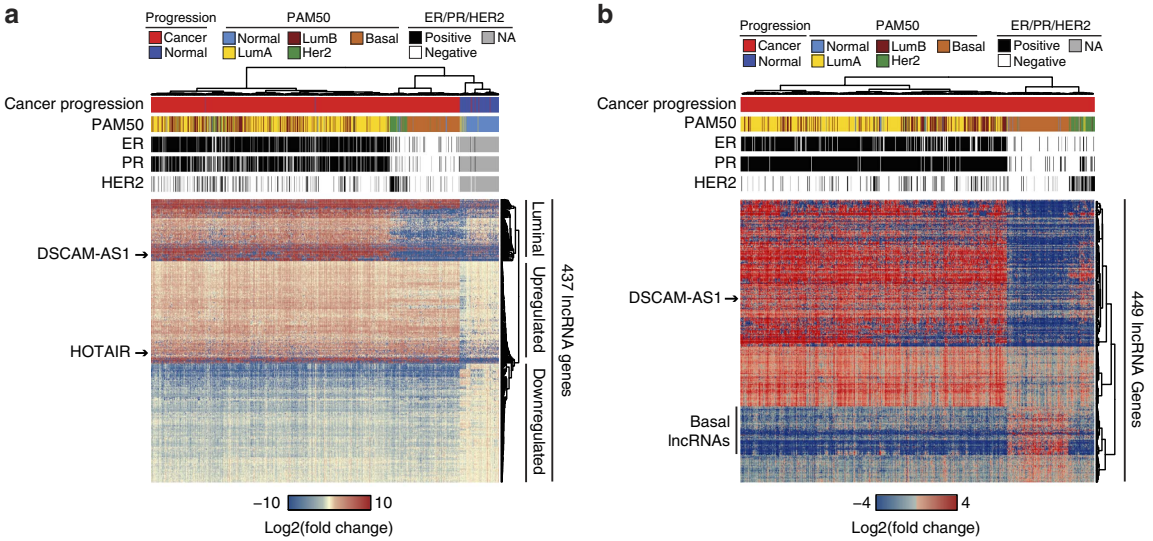
Results

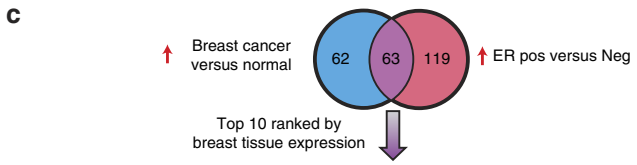
Identification of ER- and breast cancer-associated lncRNAs. We initially focused on those lncRNAs most differentially expressed in breast cancers in comparison to benign adjacent tissue (Supplementary Data 1), utilizing a non-parametric differential expression tool for RNA-seq called Sample Set Enrichment Analysis (SSEA)[7]. After applying an expression filter (at least one fragments per kilobase of transcript per million mapped reads (FPKM) expression in the breast samples in the top 5% based on gene expression level), we identified 437 of the most differentially expressed lncRNAs in breast cancer (Supplementary Data 2). Interestingly, unsupervised hierarchical clustering of the samples based on expression of these lncRNAs across all breast cancer samples (Methods section) largely separated out the breast cancer samples by PAM50 subtypes[22,23], suggesting that lncRNAs may be contributing to the distinct biology of these subtypes (Fig. 1a). While lncRNA expression was unable to distinguish between the ER-driven luminal A and luminal B subtypes, the luminal subtypes were well separated from the HER2, basal and normal subtypes (Fig. 1a). In addition to separating out the clinical subtypes of breast cancer, the lncRNAs themselves separated into three distinct clusters. The first cluster (Fig. 1a, 'Luminal') contains lncRNAs overexpressed mostly in luminal A and luminal B samples, with little expression in samples of the other subtypes, and little expression in normal samples. The next cluster contains lncRNAs upregulated across all breast cancer samples (Fig. 1a, 'Upregulated'), and this cluster included the known breast cancer lncRNA, HOTAIR. The third cluster (Fig. 1a, 'Downregulated') contains lncRNAs downregulated in breast cancers. The lncRNAs in the luminal cluster present a particularly intriguing class of potentially oestrogen-responsive lncRNAs.

Using the 947 breast tumour RNA-seq samples (Supplementary Data 1), we identified lncRNAs differentially expressed in ER-positive versus ER-negative breast tumours (Fig. 1b, Supplementary Data 2). As expected, the expression of lncRNAs differentially expressed in ER-positive tumours separated the luminal tumours from the basal and HER2 on unsupervised hierarchical clustering (Fig. 1b). Quite interestingly, a number of lncRNAs that were downregulated in ER-positive samples exhibited increased expression in the basal samples (Fig. 1b, 'Basal lncRNAs'). While these basal lncRNAs were identified in an ER-positive versus ER-negative cancer analysis, a number of them also exhibit low expression in normal breast tissue (Supplementary Fig. 1). Given that a paucity of known driver genes exist for basal breast cancers and that these tumours are the most clinically aggressive, these basal-specific lncRNAs may represent an exciting future area for basal breast cancer biology.

We set out to investigate potentially oncogenic ER-regulated lncRNAs by intersecting the lncRNAs upregulated in both the cancer versus normal (Fig. 1a) and ER-positive versus ER-negative (Fig. 1b) analyses. Sixty-three lncRNAs were upregulated in both the cancer versus normal analysis and the ER-positive versus ER-negative analysis (Supplementary Data 2, Fig. 1c). To prioritize the most biologically and clinically relevant lncRNAs, we focused on lncRNAs most highly expressed in breast cancer tissues, and those most directly regulated by ER, based on ER binding to the targets' promoter as well as the degree of induction of expression following oestrogen stimulation in breast cancer cells (Fig. 1c and Supplementary Fig. 2a). This approach nominated *DSCAM-AS1* as a lncRNA expressed at a very high level in breast cancer tissues, containing ER promoter binding, and exhibiting the strongest oestrogen induction in MCF7 and T47D cells by both RNA-seq and quantitative PCR (qPCR) validation (Fig. 1c and Supplementary Fig. 2a). We thus selected *DSCAM-AS1* for further investigation.

Characterization of *DSCAM-AS1*. *DSCAM-AS1* has been previously reported to be involved in the proliferation of a luminal breast cancer cell line[20]. It exhibits a highly cancer-specific expression pattern, mostly in breast cancer and lung adenocarcinoma, in transcriptome sequencing data from a cohort of 6,503 cancer and normal tissues and cell lines from the TCGA and the Michigan Center for Translation Pathology[7] (Fig. 2a). Supporting its association with ER biology, *DSCAM-AS1* expression is highly enriched (Student's *t*-test, P value $< 10E^{-5}$) in ER-positive tumours among the breast cancer samples in this RNA-seq cohort with ER status determined by IHC (Fig. 2b and Supplementary Data 1). In addition, analysis of RNA-seq performed on 50 breast cancer cell lines[24] revealed that expression of *DSCAM-AS1* is highly specific to ER-positive cell lines (Fig. 2c and Supplementary Fig. 2b). Further supporting the association of ER with *DSCAM-AS1*, ER chromatin immunoprecipitation-sequencing (ChIP-seq) in both MCF7 and T47D identified ER binding to the *DSCAM-AS1* promoter following oestrogen stimulation (Fig. 2d), and this finding was confirmed by ChIP-qPCR of the *DSCAM-AS1* promoter





MiTranscriptome name	Genomic coordinates	Breast Ca versus norm SSEA %ile	ER+ versus ER- SSEA %ile	Breast expression 95th %ile	ER promoter binding	ER response
DYNLL1-AS1	chr12:120899317-120933832(–)	99.03	99.89	231.53	T47D	↑
CERS6-AS1	chr2:169627826-169678566(–)	99.71	99.93	115.62	⦸	↑
DSCAM-AS1	chr21:41754824-41757280(+)	98.02	97.51	45.29	BOTH	↑↑↑
BRCAT1	chr10:8768317-9084377(+)	99.71	99.53	42.65	BOTH	—
CAT1222	chr10:5500354-5543763(–)	98.71	99.06	35.10	MCF7	↑
BRCAT54	chr5:44742417-44809099(–)	97.97	99.82	30.10	MCF7	—
BRCAT32	chr6:151913411-151937977(–)	98.27	99.86	29.33	⦸	—
CAT146	chr1:197122504-197127797(+)	98.16	97.91	26.80	⦸	↑
CAT659	chr5:44826244-44828803(–)	97.68	98.87	25.55	⦸	—
CAT505	chr4:780400-792957(+)	97.54	99.73	15.71	BOTH	—

Figure 1 | Identification of ER and breast cancer-associated lncRNAs. (**a**) Heatmap depiction of the top cancer versus normal differentially expressed lncRNAs among the TCGA breast RNA-seq cohort ($n=946$). 437 lncRNAs were differentially expressed with an SSEA FDR<1e-5 and an SSEA percentile cutoff of 0.975 (Methods section). Expression values are depicted as log2 of the fold change over the median of the normal samples ($n=104$). Unsupervised hierarchical clustering was done on both lncRNAs and patients. Cancer progression, PAM50 classification, and ER, PR, and HER2 status are shown above heatmap. LncRNAs clustered into 3 distinct categories, 'Luminal', 'Upregulated', and 'Downregulated'. Two representative lncRNAs are highlighted. (**b**) Heatmap depiction of the top ER-positive versus ER-negative lncRNAs. 449 lncRNAs met the SSEA criteria described in **a**. Unsupervised clustering was performed for samples and lncRNAs. Expression values depicted as log2 of the fold change over the median of the ER-negative samples ($n=538$). Cancer progression, PAM50 classification, and ER, PR and HER2 status are shown above heatmap. One representative lncRNA is highlighted along with a group of lncRNAs with basal-specific expression. (**c**) Venn diagram of the intersection of the breast cancer versus normal and ER-positive versus ER-negative analyses. Intersection is shown for the overexpressed lncRNAs in both categories. The top 10 lncRNAs based on expression level in breast cancer tissues (expression value of 95th percentile sample) are depicted in table. ER promoter binding determined via ChIP-seq is depicted (in either MCF7, T47D cell lines, or both) along with expression response from RNA-seq following 3 h of oestrogen stimulation in MCF7 cells (one arrow represents >1.5 fold increase, three arrows represents >2.5 fold increase).

(Supplementary Fig. 2c). The isoforms of *DSCAM-AS1* in MCF7 cells were identified using 3′ and 5′ RACE (Fig. 2d and Supplementary Table 1). *DSCAM-AS1* expression is induced in both MCF7 and T47D cells after oestrogen stimulation, and this induction is reversed with addition of tamoxifen, corroborating that ER is in fact regulating the expression of this lncRNA (Fig. 2e). Expression of known ER-regulated protein-coding genes *GREB1* and *PGR* follow the same pattern of response to oestrogen, while the lncRNA *MALAT1*, serving as a negative control, is not induced by oestrogen (Fig. 2e). In addition to being oestrogen-responsive, *DSCAM-AS1* expression is present in both the cytoplasm and nucleus at nearly identical fractions in both MCF7 and T47D cells (Supplementary Fig. 2d), and the identity of *DSCAM-AS1* as a noncoding gene was corroborated using the CPAT tool[25] (Supplementary Fig. 2e). We used single-molecule fluorescence *in situ* hybridization (ISH) to further dissect the subcellular localization and gene expression levels of DSCSM-AS1 in breast cancer cells. To this end, we designed probes that targeted all potential isoforms of the transcript predicted by RACE. On staining, we found that each MCF7 cell expressed ~800 copies of the DSCAM-AS1 transcript, almost half as much as the expression level of GAPDH (Supplementary Fig. 2f,g), additionally the similar nuclear and cytoplasmic localization was corroborated by ISH (Supplementary Fig. 2h). While the abundance of DSCAM-AS1 was lower in T47D cells (~260 molecules per cell, Supplementary Fig. 2i,j), the relative expression level (compared with GAPDH) and the subcellular localization pattern were very

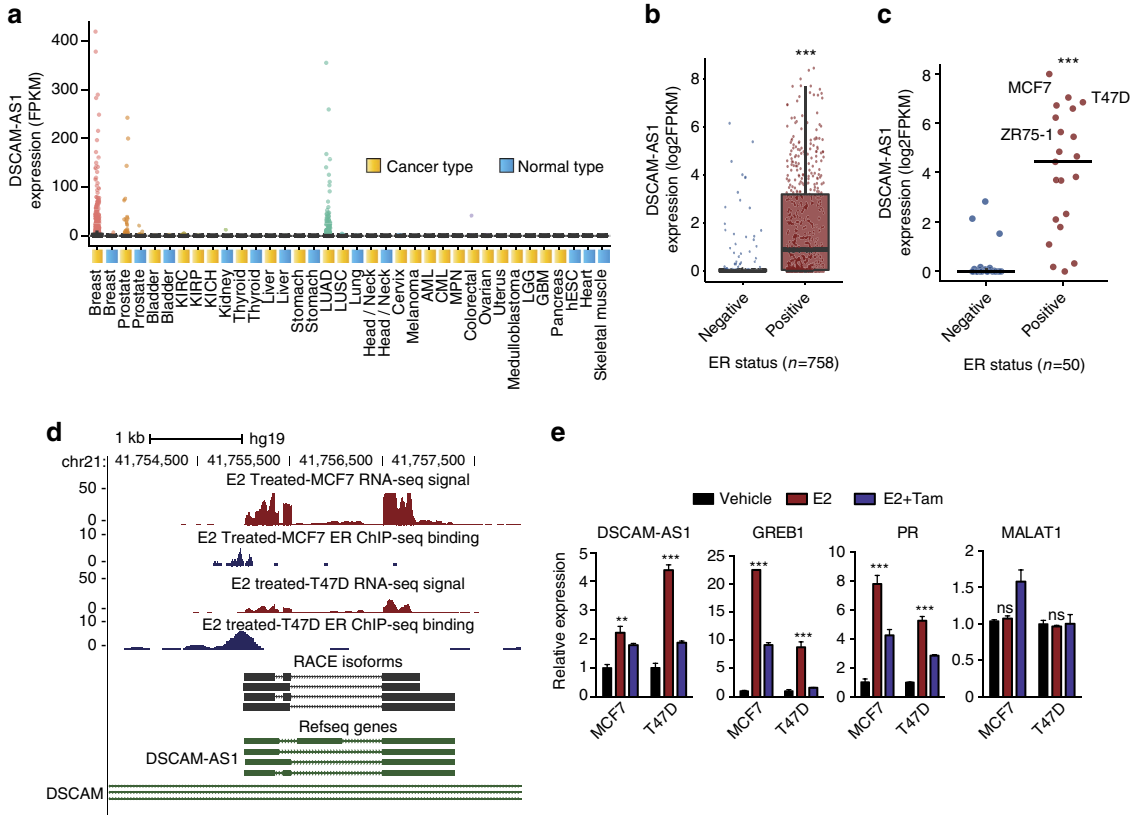


Figure 2 | Characterization of *DSCAM-AS1*. (**a**) Plot highlighting the expression in FPKM of *DSCAM-AS1* in the 6,503 sample MiTranscriptome RNA-seq compendium[7] categorized by the different cancer/tissue types. Each point represents one RNA-seq tissue sample. (**b**) Expression of *DSCAM-AS1* is significantly higher in ER-positive breast cancer tissue samples ($n=584$) compared with ER-negative samples ($n=174$). Expression was analysed in samples for which ER IHC was performed. Each point represents one RNA-seq sample. ***$P<0.0001$, comparing ER-positive with -negative. (**c**) Expression of *DSCAM-AS1* by RNA-seq in breast cancer cell lines categorized by ER status. *DSCAM-AS1* expression is significantly higher in ER-positive cell lines ($n=21$) versus ER-negative cell lines ($n=29$). Each point represents one cell line. ***$P<0.0001$, comparing ER-positive to –negative via Student's *t*-test. (**d**) UCSC genome browser depiction of *DSCAM-AS1* region on chromosome 21. RNA-seq expression track shown in red, and ER ChIP-seq shown in blue. Refseq transcripts shown in green. RACE verified transcript structure shown in black. (**e**) qPCR expression of *DSCAM-AS1*, *GREB1*, *PGR*, and *MALAT1* 8 h following addition of DMSO vehicle (black), 10 nM estrogen (red), and 10 nM estrogen and 1 μM tamoxifen (blue) in MCF7 and T47D cell lines. Error bars represent s.e.m. for three biological replicates. **$P<0.001$, ***$P<0.0001$, NS: $P>0.01$ comparing with vehicle for each condition via Student's *t*-test. NS, not significant.

similar to those observed in MCF7 cells (Supplementary Fig. 2k).

***DSCAM-AS1* is implicated in cancer aggression**. We next investigated the clinical relevance of *DSCAM-AS1*. Given that *DSCAM-AS1* is a lncRNA, its expression is not measured by most traditionally used microarrays, which are the primary high-throughput platforms annotated with reliable clinical outcomes in breast cancer[26]. As a surrogate, we employed a guilt-by-association analysis to interrogate the clinical relevance of those genes most correlated to *DSCAM-AS1*. Given that *DSCAM-AS1* is an ER-regulated lncRNA, correlation was performed using only ER-positive breast cancers, to ascertain clinical relevance in the breast cancer samples in which *DSCAM-AS1* would be enriched and most relevant. We obtained a number of breast cancer clinical data sets from Oncomine[27] containing gene expression sets associated with the presence of cancer (versus normal tissue), high clinical grade, recurrence, survival and metastasis[22,23,26–40] (Methods section). We assessed for the overlap between these gene sets with the genes most positively or negatively correlated to *DSCAM-AS1*. *DSCAM-AS1* positively correlated genes were significantly associated with clinical signatures associated with increased cancer aggression, tamoxifen resistance, higher grade, stage and metastasis (Fig. 3a,b, Supplementary Data 3 and 4). Similarly, the *DSCAM-AS1* negatively correlated genes associated with clinical signatures that portended a more favourable clinical outcome (Supplementary Fig. 3a,b, Supplementary Data 3 and 4). For many of the clinical concepts, *DSCAM-AS1* positively correlated genes displayed a clinical association comparable to those genes most correlated to *EZH2*, a gene known to be a marker of clinical aggressiveness in breast cancer[41], while genes correlated to other lncRNAs expressed in breast tissue, such as *HOTAIR*, *MALAT1* and *NEAT1*, showed modest-to-no association (Fig. 3b, Supplementary Fig. 3b, Supplementary Data 3 and 4). In addition, performing a Gene Set Enrichment Analysis (GSEA)[42] on all genes correlated to *DSCAM-AS1* yielded significant association with a myriad of breast cancer, cancer aggressiveness, and ER- and tamoxifen-associated gene signatures (Supplementary Fig. 3c). While ER-positive breast cancers typically result in better clinical outcomes[23], among the luminal breast cancers, *DSCAM-AS1* is expressed significantly higher in luminal B, a clinical subtype containing most of the clinically aggressive ER-positive breast cancers[22,23] (Fig. 3c). Despite these associations of clinical aggression with *DSCAM-AS1*, in a survival analysis of the ER-positive TCGA breast samples, expression of *DSCAM-AS1* was not significantly associated with clinical outcome (Supplementary Fig. 3d). Definitive assessment of

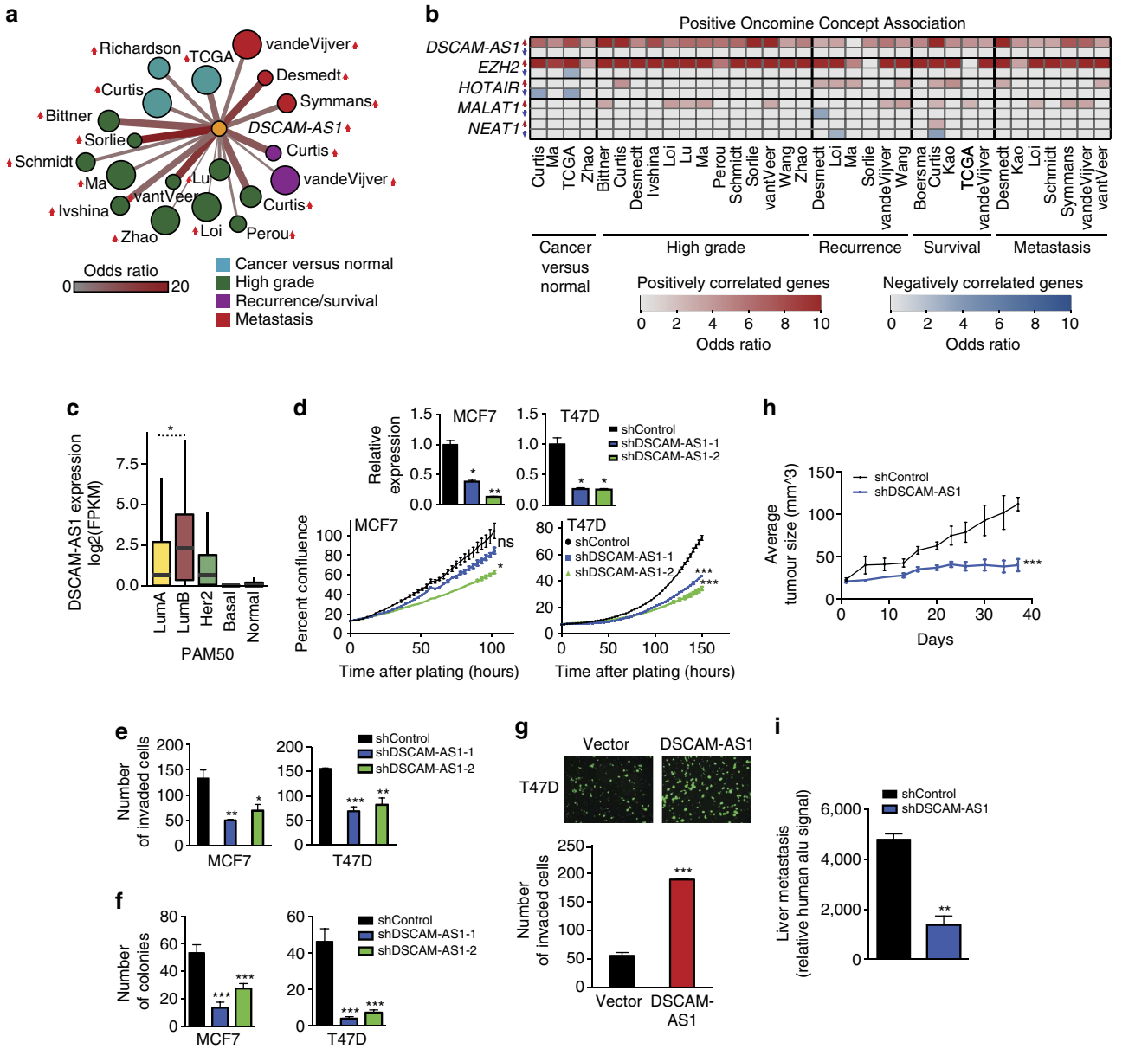


Figure 3 | *DSCAM-AS1* is implicated in cancer aggression clinically and in cell lines. (**a**) Cytoscape depiction of the overlap between the 150 genes most positively correlated with *DSCAM-AS1* and clinical signatures from Oncomine[27] for breast cancer clinical outcomes (i.e., recurrence, survival and metastasis), high cancer grade, and cancer versus normal. All significant associations with an odds ratio >6 are shown (Fisher's *P* value $<1E^{-4}$). Size of node reflects the size of the gene signature, and the thickness/redness of the line represents the magnitude of the odds ratio. (**b**) Heatmap displaying the overlap between the top 150 genes correlated to *DSCAM-AS1*, *EZH2*, *HOTAIR*, *MALAT1* and *NEAT1* and the genes positively associated with various breast cancer clinical signatures (see above). For each gene, the top row depicts the odds ratio for the positively correlated genes (red), and the bottom row represents the odds ratio for the negatively correlated genes (blue). The first name of the author for each clinical study is listed. (**c**) Expression of *DSCAM-AS1* from breast cancer RNA-seq by PAM50 classification ($n=946$). Luminal B expression is significantly greater than Luminal A (Student's *t*-test, $P=0.006$) (**d**) Incucyte proliferation assay performed following knockdown of *DSCAM-AS1* using two independent shRNAs. Degree of knockdown determined by qPCR shown above. Error bars represent the s.e.m. for three biological replicates. $^{*}P<0.01$, $^{**}P<0.001$, $^{***}P<0.0001$, NS: $P>0.01$ comparing to shControl for each condition via Student's *t*-test. (**e**) Invasion assay following shRNA knockdown of *DSCAM-AS1* using Matrigel coated Boyden chamber assay. Error bars represent the s.e.m. for three biological replicates. $^{*}P<0.01$, $^{**}P<0.001$, $^{***}P<0.0001$ comparing to shControl for each condition via Student's *t*-test. (**f**) Soft agar colony formation assay following shRNA knockdown of *DSCAM-AS1*. Error bars represent the s.e.m. for three biological replicates. $^{***}P<0.0001$ comparing to shControl for each condition via Student's *t*-test. (**g**) Invasion assay following overexpression of *DSCAM-AS1* and *LacZ* control. Error bars represent the s.e.m. for three biological replicates. Representative invasion images shown above. $^{***}P<0.0001$, comparing to vector overexpression via Student's *t*-test. (**h**) Mouse xenograft study of tumour growth for T47D cells with shRNA knockdown of *DSCAM-AS1*. Xenografts with shRNA knockdown of *DSCAM-AS1* ($n=24$) exhibited reduced growth when compared to control shRNA knockdown ($n=26$). Error bars represent the s.e.m. for all xenografts used. $^{***}P<0.0001$, comparing to shControl via Student's *t*-test. (**i**) Assessment of xenograft metastasis to liver by human Alu PCR, which detects the human cancer cells in the mouse liver. Error bars represent the s.e.m. for three biological replicates. $^{**}P<0.001$ comparing with shControl via Student's *t*-test. NS, not significant.

survival in this cohort, however, will likely require more robust and longer-term clinical curation of the TCGA breast samples.

We then studied the role of *DSCAM-AS1* on oncogenic phenotypes in ER-positive breast cancer cell lines. In MCF7 and T47D cells, stable knockdown of *DSCAM-AS1* was achieved using shRNA approaches. *DSCAM-AS1* knockdown reduced the proliferative ability of both cell lines (Fig. 3d), diminished the ability of these cells to invade in a Boyden chamber invasion assay (Fig. 3e), and substantially abolished the ability of these cells to form colonies in soft agar (Fig. 3f). While ER regulates levels of *DSCAM-AS1*, *ER* expression and protein levels are not dependent on level of *DSCAM-AS1* (Supplementary Fig. 4a), ruling out the possibility that the phenotype observed could be explained through changes in the level of *ER*. In addition, knockdown of *DSCAM-AS1* exhibited no affect on RNA or protein levels of the *DSCAM* gene, in which *DSCAM-AS1* resides antisense and intronic (Supplementary Fig. 4b). To further demonstrate the impact of *DSCAM-AS1* on aggressive cancer phenotypes, we overexpressed *DSCAM-AS1* in T47D (Supplementary Fig. 4c) and ZR75-1 (Supplementary Fig. 4d), two ER-positive breast cancer cell lines with moderate *DSCAM-AS1* expression (Fig. 2c), and observed an increase in the invasion phenotype (Fig. 3g and Supplementary Fig. 4e). MCF7 cells were not included in the overexpression studies as *DSCAM-AS1* is already expressed at a very high level in these cells (Fig. 2c). Overexpression was also tested in MDA-MB-231 cells (Supplementary Fig. 4f), a common ER-negative cell line. However, exogenous *DSCAM-AS1* was unable to confer oncogenicity via proliferation (Supplementary Fig. 4g) and invasion (Supplementary Fig. 4h). This phenomenon may be explained by a requisite genetic and epigenetic milieu provided by ER-positive cells in order for *DSCAM-AS1* to confer its cancer phenotype, and more investigation into the precise mechanisms through which it acts will shed light on this finding. Furthermore, the simple presence of *DSCAM-AS1* alone is not sufficient to make cells highly aggressive, as evidenced by its high expression in ER-positive cell lines that are moderately invasive (for example, MCF7). To further characterize the impact of *DSCAM-AS1* on cancer phenotype, we performed a mouse xenograft tumour growth assay, showing that loss of *DSCAM-AS1* reduces the growth of implanted T47D cells *in vivo* (Fig. 3h). The metastatic potential of these implanted cells were also reduced with *DSCAM-AS1* knockdown, as evidenced through decreased liver metastasis following xenograft (Fig. 3i).

Role of hnRNPL in *DSCAM-AS1* mechanism. LncRNAs have been shown to be functional through their binding interactions with other RNAs, DNA, and with proteins[2]. Thus, identifying protein binding partners for *DSCAM-AS1* is a crucial step in determining the mechanism through which it confers oncogenicity. To identify *DSCAM-AS1* binding partners, we performed pull-down of *DSCAM-AS1* and performed mass spectrometry on the pull-down product to identify proteins bound to *DSCAM-AS1* (Fig. 4a). The protein hnRNPL was observed to have the highest spectral counts for the sense form of *DSCAM-AS1* with zero spectral counts in the antisense pull-down (Fig. 4b). In addition, PCBP2, a protein known to complex with hnRNPL[43], was also among the top proteins bound to *DSCAM-AS1*. We thus investigated the interaction between *DSCAM-AS1* and hnRNPL further. HnRNPL is a protein widely expressed in many tissue types (Supplementary Fig. 5a) and has been implicated in regulating RNA stability and processing with subsequent effects on gene expression[44–47]. The binding of hnRNPL to *DSCAM-AS1* was confirmed by RNA pull-down followed by western blot, with no binding of hnRNPL to the negative control antisense transcript (Fig. 4c). Other RNA-binding proteins did not bind *DSCAM-AS1*, however, suggesting that *DSCAM-AS1* does not promiscuously bind to RNA-binding proteins in general (Fig. 4c). To further confirm this binding interaction and its specificity, RNA immunoprecipitation (RIP) was performed with using antibodies directed against hnRNPL. *DSCAM-AS1* was highly enriched by anti-hnRNPL RIP in both MCF7 and T47D cells, while control coding and noncoding genes exhibited modest binding (Fig. 4d). In addition, anti-snRNP70 and anti-HuR RIP failed to pull-down *DSCAM-AS1*, further suggesting the specificity of the *DSCAM-AS1*-hnRNPL interaction (Supplementary Fig. 5b).

To more specifically investigate the functional relationship of *DSCAM-AS1* and hnRNPL, we performed rescue studies assessing the impact of *hnRNPL* knockdown on the invasive advantage conferred by *DSCAM-AS1* overexpression, observing that reduction of *hnRNPL* levels entirely reversed the increase in invasion observed on *DSCAM-AS1* overexpression (Fig. 4e, Supplementary Fig. 6a). Because there was only slight, non-significant reduction in invasion with *hnRNPL* knockdown in control cells, the marked reduction in invasion observed in the *DSCAM-AS1* overexpressing cells with *hnRNPL* knockdown may be the result of *hnRNPL* affecting invasion in a mechanism exclusive to *DSCAM-AS1*. So, to further characterize the functional relationship between *DSCAM-AS1* and hnRNPL, we set out to localize the binding site of hnRNPL within the *DSCAM-AS1* lncRNA. Using *in silico* prediction drawing from prior studies of hnRNPL crosslinking-immunoprecipitation sequencing (CLIP-seq)[48], a single strong predicted binding peak was identified near the 3′-end of *DSCAM-AS1* (Fig. 4f). HnRNPL has been shown to bind CACA-rich RNA sites[45], and the predicted binding region possessed a 10 base pair CACA stretch. To identify if this predicted region does in fact account for the hnRNPL binding, multiple mutant forms of *DSCAM-AS1* were created with or without the binding site. *DSCAM-AS1-5* and *DSCAM-AS1-3* are large deletion mutants containing only the 5′- and 3′-end, respectively, with only *DSCAM-AS1-3* possessing the predicted binding site, and *DSCAM-AS1-D* is a mutant form with the 27 nucleotides comprising the predicted binding site deleted (Fig. 4f,g, red). The various mutant forms of *DSCAM-AS1* were expressed in HEK293, a cell line that lacks endogenous *DSCAM-AS1* expression while still expressing hnRNPL (Supplementary Fig. 6b). While both the full-length and *DSCAM-AS1-3* mutant retained hnRNPL binding, loss of the predicted binding region was effective in abrogating hnRNPL binding via both Western blot following RNA pull-down (Fig. 4g) and by qPCR following hnRNPL RIP (Supplementary Fig. 6c). All deletion mutants were expressed at comparable levels, ruling out the possibility of falsely diminished binding due to failed expression of the mutant construct (Supplementary Fig. 6d). RNA secondary structure is a crucial component of RNA functionality and is a key player in RNA-protein interactions. While the 27 nucleotide deletion in the *DSCAM-AS1-D* mutant is a small fraction of the total number of bases in the transcript, to ensure that this deletion was not causing a marked RNA secondary structure change, we investigated the impact of this deletion on RNA secondary structure via the RNAfold structure prediction tool[49]. Evidenced by a minimal free energy prediction, the posited secondary structure of *DSCAM-AS1* is largely similar to that of *DSCAM-AS1-D* (Supplementary Fig. 6e), suggesting that the loss of hnRNPL binding observed with the *DSCAM-AS1-D* mutant is not due to a dramatic secondary structure rearrangement. Quite interestingly, overexpression of the *DSCAM-AS1-D* mutant in T47D cells failed to recapitulate the increased invasion observed when overexpressing full-length *DSCAM-AS1* (Fig. 4h). This finding, in combination with the rescue studies following *hnRNPL* knockdown (Fig. 4e), strongly

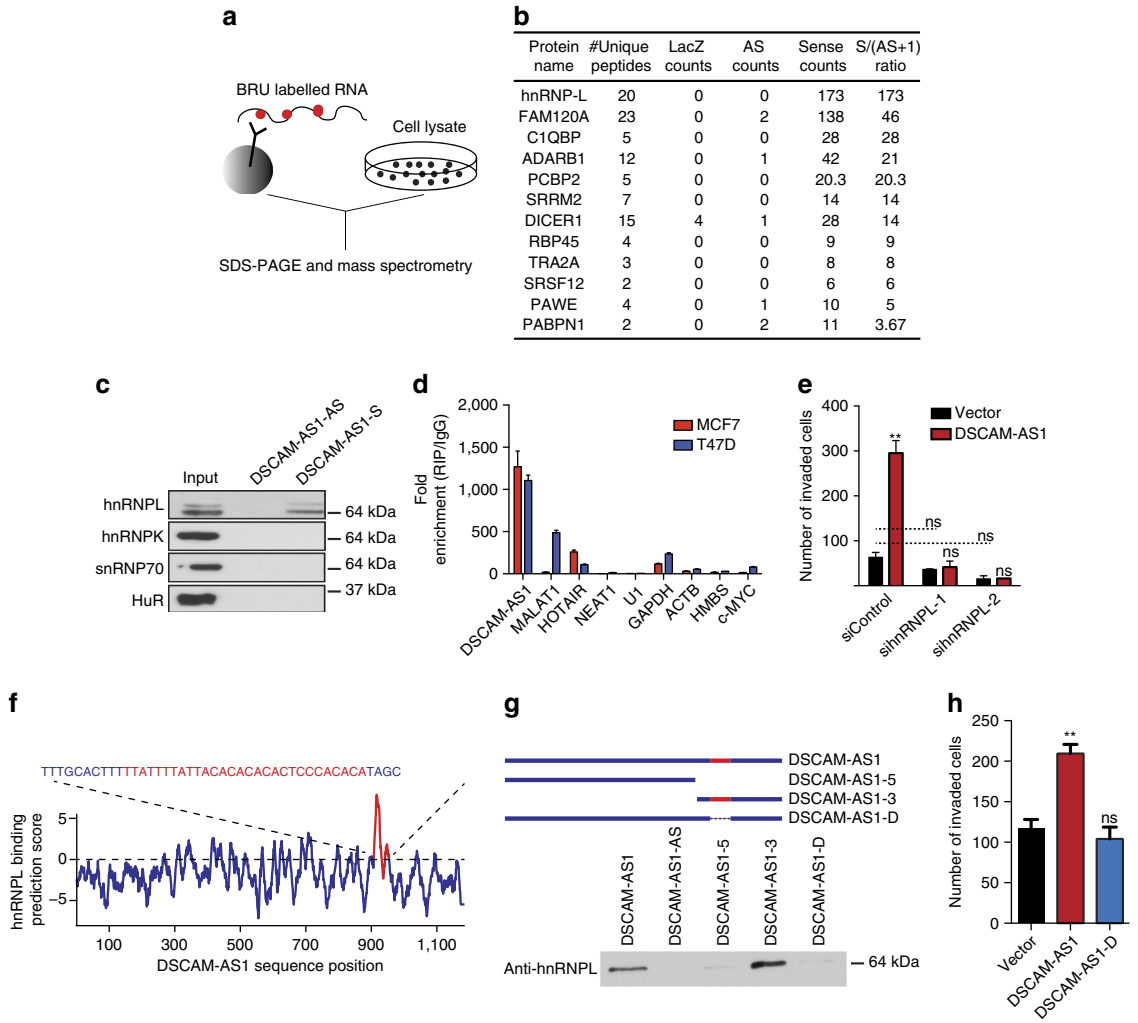


Figure 4 | Physical and functional relationship of *DSCAM-AS1* with hnRNPL. (**a**) Schematic representation of the RNA pull-down technique used to identify protein binding partners of *DSCAM-AS1*. The BRU labeled RNA transcripts are incubated with cell lysate from T47D cells and the eluted protein is resolved by SDS–PAGE. RNA-bound protein product is then processed by mass spectrometry. (**b**) Top protein binding partners for *DSCAM-AS1*. Pull-down of *LacZ* and antisense *DSCAM-AS1* used as control. S/AS ratio determined as sense counts divided by 1+antisense counts. (**c**) Western blot of hnRNPL, hnRNPK, snRNP70, and HuR following pull-down of BRU labelled *DSCAM-AS1* and antisense *DSCAM-AS1*. (**d**) qPCR following RIP for hnRNPL performed in MCF and T47D cells. Data represented as fold-enrichment over IgG RIP. Error bars represent the s.e.m. for three biological replicates. (**e**) Invasion assay for T47D cells overexpressing *LacZ* control or *DSCAM-AS1* following siRNA-mediated knockdown of *hnRNPL*. Error bars represent the s.e.m. for two biological replicates. $^{**}P<0.001$, NS: $P>0.01$ comparing to siControl (unless otherwise specified with dotted line) for each condition via Student's *t*-test. (**f**) Per base *in silico* prediction for binding of hnRNPL to *DSCAM-AS1*. One strong predicted binding peak exists in the 3′ region of *DSCAM-AS1* shown in red. (**g**) Schematic depicting the mutant forms of *DSCAM-AS1* generated with or without the predicted binding site (top). Western blot for hnRNPL shown following pull-down of each mutant form of *DSCAM-AS1* in HEK293 cells (bottom). (**h**) Invasion assay in T47D cells overexpressing *LacZ* control, full-length *DSCAM-AS1*, and the *DSCAM-AS1-D* mutant. Error bars represent the s.e.m. for three biological replicates. $^{**}P<0.001$, NS: $P>0.01$ comparing to vector overexpression for each condition via Student's *t*-test. NS, not significant; siRNA, small interfering RNA.

suggest that *DSCAM-AS1* promotes oncogenicity via its interaction with hnRNPL in these ER-positive breast cancer cells.

Role of *DSCAM-AS1* in tamoxifen resistance. A substantial number of patients with ER-positive breast cancer eventually develop resistance to endocrine therapy and present with clinical recurrence and metastasis[11,50,51]. Thus, as *DSCAM-AS1* is implicated in poor-prognosis ER-positive breast cancer (Fig. 3a-c and Supplementary Fig. 3), we set out to investigate its potential role in subverting oestrogen dependence and promoting resistance to anti-oestrogen therapies. We continuously passaged MCF7 cells in 1 uM tamoxifen for 6 months until we attained a subpopulation of MCF7 cells that were able to grow in in tamoxifen and termed these tamoxifen-resistant MCF7 cells (TamR-MCF7). Interestingly, although expression of canonical ER targets (*GREB1* and *PGR*) was decreased compared to the parental MCF7 cells, *DSCAM-AS1* expression was significantly upregulated despite already being expressed at very high levels in MCF7 cells (Fig. 5a). The levels of *ER* were also increased, which is likely a compensatory upregulation in response to the continual anti-oestrogen effects of tamoxifen. Additionally, short-term tamoxifen treatment of parental MCF7 cells transiently reduced *DSCAM-AS1* levels at 8hrs following tamoxifen treatment, with a rise back to pre-treatment levels after 24 h (Supplementary Fig. 7a). In contrast, canonical ER target, *GREB1*, exhibited pronounced expression reduction at both the short- and long-term timescale (Supplementary Fig. 7b). To interrogate whether

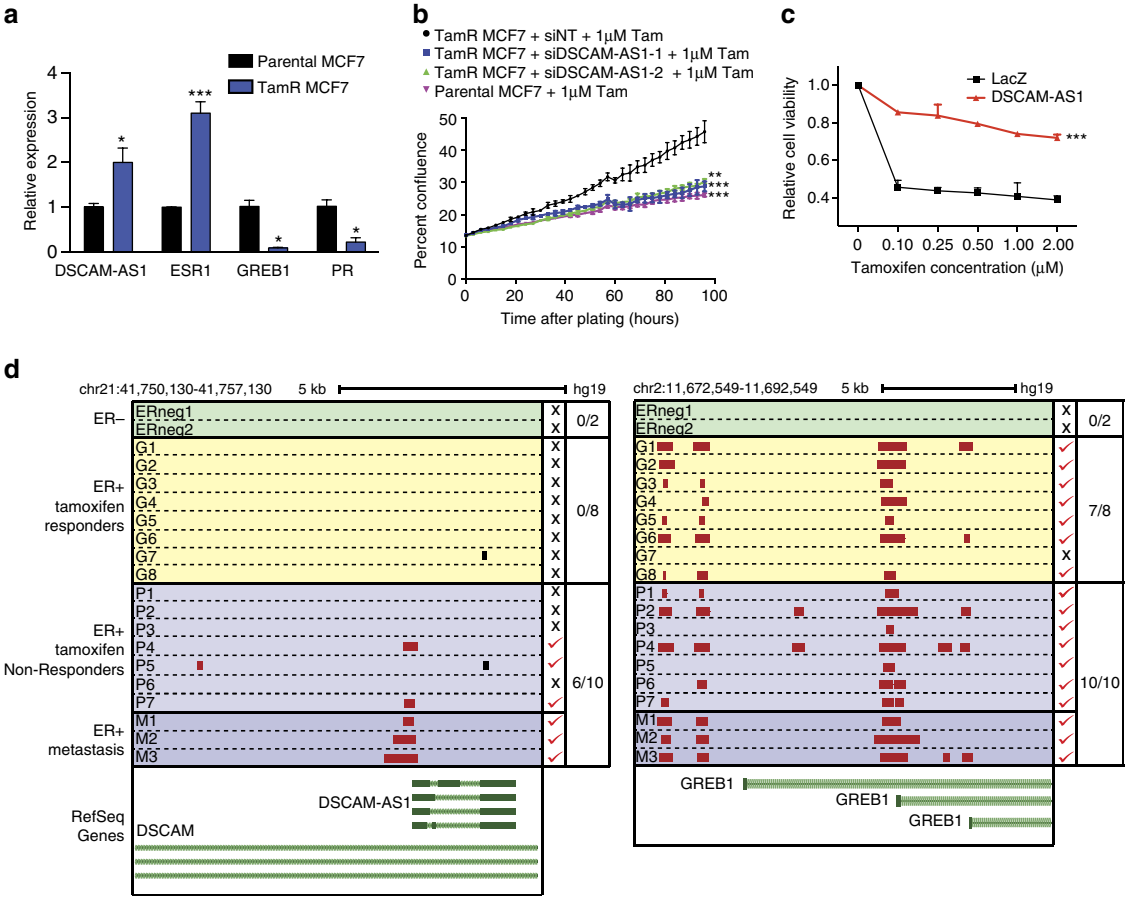


Figure 5 | *DSCAM-AS1* is implicated in tamoxifen resistance. (**a**) qPCR expression of *DSCAM-AS1*, *ESR1*, *GREB1* and *PGR* in tamoxifen-resistant MCF7 cells relative to parental MCF7. Error bars represent the s.e.m. for three biological replicates. $*P<0.01$, $***P<0.0001$, comparing to parental MCF7 for each condition via Student's *t*-test. (**b**) Proliferation assay in parental MCF7 cells and in TamR-MCF7 cells following siRNA-mediated knockdown of *DSCAM-AS1* via two independent siRNAs. Error bars represent the s.e.m. for three biological replicates. $**P<0.001$, $***P<0.0001$, comparing to parental TamR siNT for each condition via Student's *t*-test. (**c**) WST viability assay following 10 days of culture in varying levels of tamoxifen performed for T47D cells overexpressing LacZ control and *DSCAM-AS1*. Error bars represent the s.e.m. for three biological replicates. $***P<0.0001$, comparing to LacZ overexpression via Student's *t*-test. (**d**) Depiction of oestrogen receptor binding to the *DSCAM-AS1* (left) and *GREB1* (right) promoters via ChIP-seq performed in primary and metastatic breast cancer tumour tissues. ER status and response to tamoxifen treatment detailed to left. ER binding peaks (determined using MACS software) are depicted in red (for promoter binding) and black (for non-promoter binding). Promoter defined as 5KB upstream of any transcriptional start site. ER promoter binding indicated by red check or 'x' to the right. Genomic coordinates in hg19 listed above. siRNA, small interfering RNA.

this upregulation of *DSCAM-AS1* in the TamR-MCF7 cells is functionally significant, we assessed the proliferative capacity of these cells following *DSCAM-AS1* knockdown. With knockdown levels of *DSCAM-AS1* comparable to the endogenous levels in parental MCF7 cells (Supplementary Fig. 7c), knockdown of *DSCAM-AS1* in TamR-MCF7 cells led to a loss of their baseline proliferative advantage when cultured in tamoxifen, exhibiting a proliferation profile nearly identical to that of the parental MCF7 cells (Fig. 5b). Additionally, knockdown of *hnRNPL* in these cells produced a similar loss of proliferative capacity in the TamR-MCF7 cells (Supplementary Fig. 7d,e), suggesting that both *DSCAM-AS1* and hnRNPL may be playing a role in promotion of the tamoxifen resistance developed by these cells.

We also interrogated the ability of *DSCAM-AS1* to confer tamoxifen resistance in native T47D cells via overexpression of *DSCAM-AS1*. *DSCAM-AS1* overexpression was also associated with tamoxifen-resistant growth in a dose-dependent manner, with a striking increase in cell viability at levels of tamoxifen as low was 100 nM (Fig. 5c). Additionally, in line with the ability of *DSCAM-AS1* to provide oestrogen-independent growth advantage, cells overexpressing *DSCAM-AS1* also exhibited a proliferative advantage when grown in oestrogen-deprived medium compared to normal serum (Supplementary Fig. 7f). This growth advantage was abolished with the addition of oestrogen, and returned with the subsequent addition of tamoxifen (Supplementary Fig. 7f). Conversely corroborating the relationship of *DSCAM-AS1* on oestrogen dependence in these cells, we witnessed an increased oestrogen dependence of T47D cells following *DSCAM-AS1* knockdown (Supplementary Fig. 7g).

To corroborate our *in vitro* findings in a tissue model, we obtained data previously generated performing ChIP-seq for ER in primary and metastatic breast tumour tissue[13]. These tumours were grouped into the following categories as previously described[13]: primary ER-negative ($n=2$), primary ER-positive, tamoxifen-responder ($n=8$), primary ER-positive, tamoxifen-non-responder ($n=7$), metastatic ER-positive ($n=3$). Strikingly, investigation of the *DSCAM-AS1* promoter revealed that ER preferentially binds to the *DSCAM-AS1* promoter in tumours with clinical aggression (ie, metastatic and tamoxifen non-responders; Fig. 5d), while a canonical ER target, *GREB1*, exhibits ER-biding to its promoter in nearly all ER-positive tumours, lacking preference for the more clinically aggressive

tumours. Altogether, these data suggest that the association between *DSCAM-AS1* expression with clinical aggressiveness in ER-positive breast cancer samples may be explained, in part, by the ability of *DSCAM-AS1* to facilitate oestrogen-independent oncogenicity, thus potentially promoting resistance to endocrine therapy with tamoxifen.

Discussion

Further investigation and study of the mechanisms through which ER-dominant breast cancers become aggressive and eventually evade traditional clinical therapies is of intense clinical interest. In this study, we identify a myriad of potentially ER-associated lncRNAs, and functionally and mechanistically characterize one of the most intriguing candidates. Nevertheless, further investigation of some of these other lncRNAs may also contribute to our understanding of ER biology and ER-driven oncogenesis. LncRNAs have been shown to function through multiple mechanisms, and the study of the interaction of *DSCAM-AS1* with hnRNPL is a promising step towards understanding the ways through which this molecule executes its oncogenic function. While we show that the binding of hnRNPL to *DSCAM-AS1* is responsible for at least some of its oncogenicity, a further understanding of how the interaction between hnRNPL and *DSCAM-AS1* is mediating this phenotype is necessary.

Novel mediators of tumour aggression, such as *DSCAM-AS1*, can provide insight into the mechanism of endocrine therapy resistance. This increased understanding may in turn lead to more effective strategies to overcome this resistance, which is one of the last, great clinical challenges in treatment of ER-positive breast cancer. In addition, there is little known regarding the role of noncoding RNAs in developing resistance to anti-oestrogen therapy, with a small number of studies implicating some of the more prominent, well characterized breast cancer lncRNAs[52,53]. *DSCAM-AS1* is just one of many potentially relevant ER-regulated lncRNAs in breast cancer, and further investigation of the other candidates is likely to yield a greater understanding of ER-mediated cancer biology. Ultimately, this study provides key insight into the role of lncRNAs in ER breast cancer biology, and is an important step in better understanding this common disease.

Methods

Cell lines and cell culture. All cell lines were obtained from the American Type Culture Collection (Manassas, VA). Cell lines were maintained using standard media and conditions. Specifically, T47D cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin and 5 mg ml^{-1} insulin. ZR75-1 cells were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin. HEK293 cells were maintained in DMEM plus 10% fetal bovine serum (FBS) plus 1% penicillin-streptomycin. MCF7 cells were cultured in Dulbecco's modified Eagle's medium plus GlutaMAX (DMEM, Invitrogen) containing 10% fetal bovine serum (Hyclone) and 1% penicillin-streptomycin. To establish the tamoxifen-resistant cell line, MCF7 cells were grown in IMEM phenol-red free medium with 10% Charcoal-stripped FBS in the presence of 1 uM (Z)-4-hydroxytamoxifen (Sigma) for 6 months. All cell lines were grown at 37 °C in a 5% CO_2 cell culture incubator, and were genotyped for identity at the University of Michigan Sequencing Core and tested routinely for *Mycoplasma* contamination.

Cell proliferation assay. Cells were seeded in a 48-well plate at 3×10^4 cells per well. Plates were added to Incucyte machine (Essen Bioscience) 16–20 h following seeding. Growth curves were constructed by imaging plates using the Incucyte system, where the growth curves are generated from confluence measurements acquired during continuous kinetic imaging. Four wells were measured per condition. For tamoxifen treatment, 16–20 h after seeding, the medium was changed to RPMI phenol-red free medium containing 10% charcoal-treated FBS in the presence of 1 uM tamoxifen or ethanol. Growth curves were obtained using Incucyte system as described above.

Cell viability assay. Cells were seeded in 96-well plates at 5,000 cells per well in a total volume of 100 µl media containing 10% FBS. Serially diluted tamoxifen in 100 µl of media was added to the cells 12 h after seeding. Medium containing tamoxifen was replenished every 2–3 days. Following 10 days of incubation, cell viability was assessed by WST assay (WST-8, Dojindo). All assays were performed in triplicate and repeated at least three times. The relative cell viability was expressed as a percentage of the control that was treated with vehicle solutions.

Soft agar colony formation assay. 10,000 cells were suspended in medium containing 0.3% agar, 10% FBS, and layered on medium containing 0.6% agar and 10% FBS in six-well plate. Colonies were stained for 18–24 h with iodonitrotetrozolium chloride (Sigma #18377) following 3 weeks of incubation. Colonies from three replicate wells were quantified.

Quantitative RT-PCR assay. The miRNeasy mini kit was utilized to isolate RNA from cell lysates. From 1 µg of isolated RNA, SuperScript III (Invitrogen) and Random Primers (Invitrogen) were used to generate cDNA according to the manufacturer's protocol. The ABI7900 HT Fast Real time system (Applied Biosystems) was utilized for quantitiative reverse transcriptase–PCR (qRT–PCR) reactions. Gene-specific primer were designed using the Primer3 software and were subsequently synthesized by IDT Technologies. A relative quantification method was used in analysing the qRT–PCR data and data were depicted as average fold change versus the control (as internal reference, GAPDH and actin were utilized). All primers used for qPCR are detailed in Supplementary Table 2. Three technical replicates were used in each assay, and all data shown was performed with at least three biological replicates.

Oestrogen and tamoxifen treatment. To evaluate the effect of oestrogen stimulation, cells were first hormone depleted via growth in phenol-red free medium containing 10% charcoal-treated FBS for 72 h and then treated with ethanol vehicle, 10 nM β-estradiol, or 10 nM β-estradiol plus 1 uM tamoxifen. After 10 h, RNA was isolated as described above and qPCR was performed as described above using Power SYBR Green Mastermix (Applied Biosystems).

Subcellular fractionation. Cellular fractionation was performed using a RiboTrap Kit (MBL International), according to manufacturer's instructions. RNA was isolated and qRT–PCR was performed as described above.

Knockdown and overexpression studies. Knockdown of *DSCAM-AS1* and *hnRNPL* in T47D and MCF7 cells was accomplished by small interfering RNA from Dharmacon. Transfections were performed with OptiMEM (Invitrogen) and RNAi Max (Invitrogen) per manufacturer instruction. Target sequences used for shRNA or small interfering RNA knockdown are listed in Supplementary Table 3. For stable knockdown of *DSCAM-AS1*, MCF7 and T47D cells were transfected with lentiviral constructs containing 2 different *DSCAM-AS1* shRNAs or no targeting shRNAs in the presence of polybrene (8 µg ml^{-1} Supplementary Table 3). After 48 h, transduced cells were grown in culture media containing 2 µg ml^{-1} puromycin.

For *DSCAM-AS1* overexpression, the predominant isoform (isoform 2, Supplementary Table 1) was cloned into the pLenti6.3 vector (Invitrogen) using PCR8 non-directional Gateway cloning (Invitrogen) as an initial cloning vector and shuttling was then done to pLenti6.3 using LR clonase II (Invitrogen) according to the manufacturer's instructions. As control, *LacZ* was also cloned into the same vector system. The primers for making *DSCAM-AS1* mutation and truncations are listed in Supplementary Table 2. Lentiviral particles were made and T47D and ZR75.1 cells were transduced as described above. Stable cell lines were generated by selection with 3 µg ml^{-1} blasticidin. Transient transfection of *DSCAM-AS1* and its derivative mutants was done in HEK293 cells was performed with Lipofectamine LTX (Invitrogen) following manufacturer's instruction. Cells were collected at 48 h post transfection.

***In vitro* RNA-binding assay.** The RNA-binding assay was performed according to the protocol of the RiboTrap Kit (MBL International). In brief, 5-bromo-UTP (BrU) was randomly incorporated into sense *DSCAM-AS1*, antisense *DSCAM-AS1*, and *LacZ* control via PCR-based transcription. The primers are shown in Supplementary Table 2. The the BrU labelled RNA transcripts were bound to beads conjugated with anti-BrdU antibodies. Then, the cytoplasmic or nuclear fractions from MCF7 or T47D cells were mixed for 2 h. Samples were washed four times with Wash Buffer II before elution. The samples were sent to the Michigan Center for Translational Pathology proteomic core facility for mass spectrometry.

Mass spectrometry. The samples were treated with SDS-PAGE loading buffer supplied with 10 mM DTT for 5 min at 85 °C. The proteins were alkylated by the addition of iodoacetamide to the final concentration of 15 mM. The samples were subjected to SDS–PAGE and the whole lanes were cut out and digested with trypsin in-gel for 2 h. The resulting peptides were extracted, dried and resuspended in 0.1% formic acid with 5% acetonitrile before loading onto a 2 cm EASY-column

(Thermo Scientific) coupled to an in-house made nano HPLC column (20 cm × 75 um) packed with LUNA C18 media (Phenomenex). Analysis was performed on a Velos Pro mass spectrometer (Thermo Scientific) operated in data-dependent mode using 120-min gradients in EASY-LC system (Proxeon) with 95% water, 5% acetonitrile (ACN), 0.1% formic acid (FA) (solvent A), and 95% ACN, 5% water, 0.1% FA (solvent B) at a flow rate of 220 nl min^{-1}. The acquisition cycle consisted of a survey MS scan in the normal mode followed by 12 data-dependent MS/MS scans acquired in the rapid mode. Charge state was not recorded. Dynamic exclusion was used with the following parameters: exclusion size 500, repeat count 1, repeat duration 10 s, exclusion time 45 s. Target value was set at 10^4 for tandem MS scan. The precursor isolation window was set at 2 *m/z*. The complete analysis comprised two independent biological replicates.

Mass spectrometry data analysis. The resulting spectrum files were transformed into MGF format by MSConvert software and interrogated by MASCOT 2.4 search engine using human UniProt database version 15 concatenated with reverse sequences for estimation of false discovery rate (FDR) and with a list of common contaminants (40,729 entries in total). The search parameters were as follows: full tryptic search, 2 allowed missed cleavages, peptide charges +2 and +3 only, MS tolerance 1 Da, MS/MS tolerance 0.5 Da. Permanent post-translational modification was: cysteine carbamidomethylation. Variable post-translational modifications were: protein N-terminal acetylation, Met oxidation and N-terminal Glutamine to pyro-Glutamate conversion. The remaining analysis was performed as previously described[54]. To summarize, the minimal ion score threshold was chosen such that a peptide FDR below 1% was achieved. The peptide FDR was calculated as: 2 × (decoy_hits)/(target + decoy hits). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[55] via the PRIDE partner repository with the data set identifier PXD002421 and 10.6019/PXD002421. Spectral counts for all detected proteins were assembled using an in-house written Python script. The adjustment of spectral counts was done as previously described[54].

RNA immunoprecipitation. RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit (Millipore, #17-700) according to the manufacturer's instructions. RIP-PCR was performed as qPCR, as described above, using total RNA as input controls. 1:150th of RIP RNA product was used per PCR reaction. Antibodies used for RIP are listed in Supplementary Table 4. All RIP assays were performed in biological duplicate.

Invasion assay. 3×10^5 cells were seeded in a 24-well corning FluoroBlok chamber pre-coated with Matrigel (BD Biosciences). Medium containing 10% FBS in the lower chamber served as chemoattractant. After 48 h, cells remaining on the lower side of the membrane were stained with calcein AM (C34852 invitrogen). The invasive cells adhering to the bottom surface of the filter were quantified under a fluorescent microscope (× 2).

Antibodies and immunoblot analyses. Western immunoblot assays were performed by running cell lysates on 4–12% SDS polyacrylamide gels (Novex) to separate proteins. Proteins were then transferred to a nitrocellulose membrane (Novex) via wet transfer at 30 V overnight. Blocking buffer incubation was then performed for 1 h (Tris-buffered saline, 0.1% Tween (TBS-T), 5% nonfat dry milk). Indicated antibodies were then added to membrane and incubated at 4 °C overnight. Enhanced chemiluminescence (ECL Prime) was utilized to develop blots via the manufacturer's protocol. All the antibodies used in this study are described in Supplementary Table 4. Representative full blot images are shown in Supplementary Fig. 8.

Chromatin immunoprecipitation. HighCell# ChIP kit (Diagenode) was utilized to perform ChIP assays via the manufacturer's protocol. Briefly, MCF7 cells were grown in charcoal-stripped serum media (described above) for 72 h and then stimulated 10 nM estradiol for 12 h. Cells were then crosslinked using 1% formaldehyde for 10 min, and crosslinking was quenched for 5 min at room temperature using a 1/10 volume of 1.25 M glycine. Cells were then lysed and sonicated (Bioruptor, Diagenode), yielding an average chromatin fragment size of 300 bp. An equivalent amount of chromatin equivalent to 5×10^6 cells was used for the ChIP for all antibodies. DNA bound to immunoprecipitated product was isolated (IPure Kit, Diagenode) via overnight incubation with antibody at 4 °C. Samples were then washed, and crosslinked reversed.

ChIP-seq library construction and sequencing analysis. DNA was purified for library preparation using the IPure Kit (Diagenode). The ChIP-seq sample preparation for sequencing was performed according to the manufacturer's instructions (Illumina). ChIP-enriched DNA samples (1–10 ng) were converted into blunt-ended fragments using T4 DNA polymerase, *Escherichia coli* DNA polymerase I large fragment (Klenow polymerase) and T4 polynucleotide kinase (New England BioLabs (NEB)). A single adenine base was added to fragment ends by Klenow fragment (3′ to 5′ exo$^-$; NEB) followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-modified DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (QIAGEN). Libraries were quantified with the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2000 Sequencer (100-nucleotide read length). ChIP-seq data were mapped to human genome version hg19 using BWA[56]. The MACS program[57] was used to generate coverage map files to visualize the raw signal on the UCSC genome browser[58]. Hpeak[59], a hidden Markov model (HMM)-based peak-calling software program designed for the identification of protein-interactive genomic regions, was used for ChIP-seq peak determination.

ChIP-seq peak promoter overlap. Overlap of ChIP-seq peaks with gene promoters was performed using the BEDTools 'coverage' tool. Intervals of ± 5–10 kilobases surrounding unique transcriptional starts were used to assess promoter overlap.

Coding potential scoring. Coding potential for all lncRNA transcripts was determined as described previously[4]. The alignment-free Coding Potential Assessment Tool (CPAT)[25] was used to determine coding probability for each transcript. CPAT determines the coding probability of transcript sequences using a logistic regression model built from ORF size, Fickett TESTCODE statistic, and hexamer usage bias.

Xenograft analysis. All experimental procedures were approved by the University of Michigan Committee for the Use and Care of Animals (UCUCA) and conform to all regulatory standards. A total of 5×10^6 cells of T47D control or T47D shM41 cells were suspended in 100 ul of PBS/Matrigel (1:1) were injected subcutaneously in 5-week-old pathogen-free female CB-17 severe combine immunodefiecient mice (CB-17 SCID) which simultaneously received a 60-day slow release pellet containing 0.18 mg of 17b-estradiol (Innovative Research of America). Tumours were measured weekly using a digital caliper. Growth in tumour volume was recorded using digital calipers and tumour volumes were estimated using the formula $(\pi/6)$ $(L \times W^2)$, where L = length of tumour and W = width. In addition, mouse livers were collected to determine spontaneous metastasis by measuring human Alu sequence. Briefly, genomic DNA from livers were prepared using Puregene DNA purification system (Qiagen), followed by quantification of human Alu sequence by human Alu specific Fluorogenic Taqman qPCR probes.

RNA-seq data processing. Sequence quality control was done using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Next, reads mapping to mitochondrial DNA, ribosomal RNA, poly-A, poly-C, Illumina sequencing adaptors, and the spiked-in phiX174 viral genome were filtered. Sequences were downloaded from the Illumina iGenomes server (2012, March 9). Mapping was performed using bowtie2 (2.0.2). Reads were mapped using TopHat2 (2.0.6 and 2.0.8) using default parameters. A human genome reference was constructed from UCSC version Feb 2009 (GRCh37/hg19) chromosomes 1–22, X, Y and mitochondrial DNA, and references from alternate haplotype alleles were omitted. Bowtie-build and bowtie2-build were used to build genome reference for Bowtie versions 0.12.8 and 2.0.2 were, respectively. The Ensembl version 69 transcriptome was used as a reference gene set. Using the --transcriptome-index option in TopHat version 2.0.6 (ref. 60), alignment index files were prepared from this reference for Bowtie versions 0.12.8 and 2.0.2.

RNA-seq transcript expression estimation. Cufflinks version 2.1.1 (ref. 61) was used with the following parameters to estimate transcript abundance from RNA-seq data: '--max-frag-multihits = 1', '--no-effective-length-correction', '--max-bundle-length 5000000', '--max-bundle-frags 20000000'. To convert FPKM abundance estimates (generated by Cufflinks) to approximate fragment count values we multiplied each FPKM by the transcript length (in kilobases) and by the 'Map Mass' value (divided by 1.0E^6) found in the Cufflinks log files.

Breast cancer tissue expression heatmap generation. The 'gplots' R-package was used to generate heatmaps using the *heatmap.2* function. For the cancer versus normal heatmap, expression was normalized as log2 of the fold change over the median of the normal samples for each transcript. For the ER-positive versus ER-negative heatmap, expression was normalized to the median of the ER-negative samples. Unsupervised heirarchical clustering was performed with the *hclust* function, using Pearson correlation as the clustering distance, using the 'ward' agglomeration method.

RNA-seq differential expression testing. Differential expression testing was performed using the SSEA tool described previously[7]. Briefly, following count data normalization, SSEA performs the weighted KS-test procedure described in GSEA[42]. The resulting enrichment score statistic describes the enrichment of the sample set among all samples being tested. To test for significance, SSEA enrichment tests are performed following random shuffling of the sample labels.

These shuffled enrichment tests are used to derive a set of null enrichment scores (1,000 null enrichment scores computed). The nominal *P* value reported is the relative rank of the observed enrichment score within the null enrichment scores. Multiple hypothesis testing is performed by comparing the enrichment score of the test to the null normalized enrichment score distributions for all transcripts in a sample set. This null normalized enrichment score distribution is used to compute FDR *Q* values in the same manner used by GSEA[42].

Associations with oncomine clinical signatures. We identified the top 150 positively and negatively correlated genes (Spearman's correlation) to *DSCAM-AS1* among the ER-positive breast cancer samples. These gene lists were imported into Oncomine[27] as custom concepts. We then nominated significantly associated breast cancer concepts with odds ratio >4.0 for negatively associated concepts and >6.0 for positively associated concepts and *P* value $<1\times10^{-6}$. Nodes and edges of these associations were exported and a concept association network was generated using Cytoscape version 3.2.1. Node positions were computed using the Force-Directed Layout algorithm in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of node labels.

Association of correlation signatures with oncomine concepts. Correlation analysis described above was performed for *DSCAM-AS1*, *EZH2*, *HOTAIR*, *MALAT1*, and *NEAT1*. For each gene, we created a signature of the top 150 most positively and top 150 most negatively correlated genes. We performed a Fisher's exact test of overlap for each of the above gene signatures with Oncomine clinical signatures for cancer versus normal, clinical recurrence, clinical survival, metastasis, and high clinical grade. The following studies were utilized: Curtis Breast[26], Ma Breas[62], TCGA Breast[28], Zhao Breast[29], Bittner Breast[63], Desmedt Breast[30], Ivshina Breast[31], Loi Breast[32], Lu Breast[33], Perou Breast[22], Schmidt Breast[34], Sorlie Breast[23], vantVeer Breast[64], Wang Breast[36], Boersma Breast[37], Kao Breast[38], Symmans Breast[39] and vandeVijver Breast[40]. For each Oncomine concept, overlap was tested for the top 1, 5 and 10% of genes up- and downregulated, and the gene signature with the greatest odds ratio was selected for each study. Signature comparisons were performed using a one-sided Fisher's exact test.

Survival analysis with TCGA breast data. Association of DSCAM-AS1 levels on clinical outcomes was assessed using the TCGA breast cohort. Survival data was obtained from the TCGA data portal. ER-positive samples were used for survival analysis as indicated by the TCGA clinical metadata via IHC status. Samples with DSCAM-AS1 expression >10 FPKM were grouped into the 'DSCAM-AS1 high' category and samples with expression <1 FPKM were grouped into the 'DSCAM-AS1 low' category. Kaplan-Meier analysis was performed, and log-rank test was performed to assess statistical significance.

Tissue expression level percentile metric. To generate a metric to summarize the expression of each lncRNA in breast cancer tissues, we identified the expression level of the 95th percentile sample among all breast RNA-seq samples including cancers tissue, normal tissue, and cell lines.

RNA-sequencing library preparation. Total RNA was obtained from cancer cell lines, and RNA quality was determined using an Agilent Bioanalyzer. Poly-A transcriptome libraries from the mRNA fractions were generated following the Illumina RNA-seq protocol. Each sample was sequenced in a single lane with the Illumina HiSeq 2000 (100-nucleotide read length) as previously described[3,65]. The dUTP method of second-strand marking was used for strand-specific library preparation as described previously[66].

Gene set enrichment analysis. Expression levels of *DSCAM-AS1* were correlated (Spearman) to the expression of all protein-coding genes across all ER-positive breast cancers. The protein-coding genes were then ranked by the Spearman Rho value, and used in a weighted, preranked GSEA analysis against MSigDB gene sets V5.0 (ref. 67).

***In silico* binding prediction.** To obtain potential HNRNPL binding sites on DSCAM-AS1, we utilized GraphProt[68] to learn a predictive model from genome-wide HNRNPL binding sites identified by iCLIP-seq[48]. For training data generation, we extracted the genomic binding positions (GSE37560) with BED table scores $>=10$, followed by an extension of ±20 nt resulting in 41 nt long binding sites. After mapping the sites to annotated RefSeq genes obtained from UCSC, an equally-sized set of negative sites was selected such that the sites were on the same RefSeq genes and did not overlap with any of the identified positive sites from the initial table. The GraphProt sequence model trained on these data was then used to identify high-scoring sites in the *DSCAM-AS1* sequence (NCBI GenBank NR_038899.1). The highest-scoring site centred at RNA position 923 contains a CA-repeat motif known for its affinity towards HNRNPL and was thus used for subsequent analysis.

Rapid amplification of cDNA ends (RACE). 5′ and 3′ RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen) according to the manufacturer's protocols. RACE PCR products obtained using Platinum Taq high-fidelity polymerase (Invitrogen), were resolved on a 1.5% agarose gel. Individual bands were gel purified using a Gel Extraction kit (Qiagen), and cloned into PCR4 TOPO vector, and sequenced using M13 primers.

Single-molecule fluorescence *in situ* hybridization. Single-molecule fluorescence *in situ* hybridization was performed as described[69], with some minor modifications. Cells were grown on 8-well chambered coverglasses, formaldehyde fixed and permeablized overnight at 4 °C using 70% ethanol. Cells were rehydrated in a solution containing 10% formamide and $2\times$ SSC for 5 min and then treated with 10 nM fluorescence *in situ* hybridization probes for 16 h in $2\times$ SSC containing 10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNAse-free BSA, 1 μg μl^{-1} *E. coli* transfer RNA and 10% formamide at 37 °C. After hybridization, cells were washed twice for 30 min at 37 °C using a wash buffer (10% formamide in $2\times$ SSC). Cells were then mounted in solution containing 10 mM Tris/HCl pH 7.5, $2\times$ SSC, 2 mM trolox, 50 μM protocatechiuc acid and 50 nM protocatechuate dehydrogenase. fluorescence *in situ* hybridization samples were imaged in three dimensions using HILO illumination as described[70]. Images were processed using custom-written macros in ImageJ. Analysis routines comprises 3 major steps: background subtraction, Laplacian of Gaussian (LoG) filtering and thresholding. Spots with intensity above set threshold are represented in images. Probes were designed to target all isoforms of the DSCAM-AS1 transcript. Probe sequences targeting DSCAM-AS1 (21 probes per transcript) are as follows: 5′-cctatccctttctctaagaa-3′, 5′-acttctgcaaaaacgtgctg-3′, 5′-ggttccactccatt ttaatt-3′, 5′-ctatagcgtcttatcagctg-3′, 5′-catgtgtccggatatcattt-3′, 5′-tcagtgagtggataact ggt-3′, 5′-aattctagtggaggcaccta-3′, 5′-ctaagtagcttcatctttcc-3′, 5′-caactgcgtgtttccta gtc-3′, 5′-agcattctctgttttaacca-3′, 5′-ttagcaactgccttgctctg-3′, 5′-gctgtccagttttagta aca-3′, 5′-cgttgtgagcctgagagatc-3′, 5′-agaacttccctagaggagtg-3′, 5′-atggggagtgagaccaa aca-3′, 5′-tggaggagggacagagaagg, 5′-tgtgggtgattggtactttt-3′, 5′-atggatgagtatgtcat gcc-3′, 5′-tattgccatggttagcatga-3′, 5′-aatgcatgcttgatggagct-3′.

Data availability. Sequence data that support the findings of this study have been deposited in the Short Read Archive with the accession code SRP078392. Tissue ChIP-seq data referenced in this study are available in the Gene Expression Omnibus with the accession code GSE32222. All remaining data are contained within the Article and Supplementary Information files or available from the author on request.

References

1. Prensner, J. R. & Chinnaiyan, A. M. The Emergence of lncRNAs in Cancer Biology. *Cancer Discov.* **1,** 391–407 (2011).
2. Ulitsky, I. & Bartel, D. P. LincRNAs: Genomics, evolution, and mechanisms. *Cell* **154,** 26–46 (2013).
3. Prensner, J. R. *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* **29,** 742–749 (2011).
4. Prensner, J. R. *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* **45,** 1392–1398 (2013).
5. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142,** 409–419 (2010).
6. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464,** 1071–1076 (2010).
7. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47,** 199–208 (2015).
8. Cancer Genome Atlas Research, N. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2015).
9. Sun, M., Gadad, S. S., Kim, D. S. & Kraus, W. L. Discovery, annotation, and functional analysis of long noncoding RNAs controlling cell-cycle gene expression and proliferation in breast cancer cells. *Mol. Cell* **59,** 698–711 (2015).
10. Sotiriou, C. *et al.* Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl Acad. Sci. USA* **100,** 10393–10398 (2003).
11. Haque, R. *et al.* Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol. Biomarkers Prev.* **21,** 1848–1855 (2012).
12. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38,** 1289–1297 (2006).
13. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481,** 389–393 (2012).
14. Robinson, D. R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* **45,** 1446–1451 (2013).
15. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat. Genet.* **45,** 1439–1445 (2013).

16. Osborne, C. K. & Schiff, R. Mechanisms of endocrine resistance in breast cancer. *Annu. Rev. Med.* **62,** 233–247 (2011).
17. Qiu, J.-J. *et al.* Expression and clinical significance of estrogen-regulated long non-coding RNAs in estrogen receptor α-positive ovarian cancer progression. *Oncol. Rep.* **31,** 1613–1622 (2014).
18. Chakravarty, D. *et al.* The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.* **5,** 5383 (2014).
19. Bhan, A. *et al.* Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. *J. Mol. Biol.* **425,** 3707–3722 (2013).
20. Miano, V. *et al.* Luminal long non-coding RNAs regulated by estrogen receptor alpha in a ligand-independent manner show functional roles in breast cancer. *Oncotarget* **7,** 3201–3216 (2016).
21. Hah, N. *et al.* A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145,** 622–634 (2011).
22. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
23. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98,** 10869–10874 (2001).
24. Robinson, D. R. *et al.* Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat. Med.* **17,** 1646–1651 (2011).
25. Wang, L. *et al.* CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41,** 1–7 (2013).
26. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486,** 346–352 (2012).
27. Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform1. *Neoplasia* **6,** 1–6 (2004).
28. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).
29. Zhao, H. *et al.* Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol. Biol. Cell* **15,** 2523–2536 (2004).
30. Desmedt, C. *et al.* Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin. Cancer Res.* **14,** 5158–5165 (2008).
31. Ivshina, A. V. *et al.* Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* **66,** 10292–10301 (2006).
32. Loi, S. *et al.* Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J. Clin. Oncol.* **25,** 1239–1246 (2007).
33. Lu, X. *et al.* Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res. Treat.* **108,** 191–201 (2008).
34. Schmidt, M. *et al.* The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res.* **68,** 5405–5413 (2008).
35. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).
36. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365,** 671–679 (2005).
37. Boersma, B. J. *et al.* A stromal gene signature associated with inflammatory breast cancer. *Int. J. Cancer* **122,** 1324–1332 (2008).
38. Kao, K.-J., Chang, K.-M., Hsu, H.-C. & Huang, A. T. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* **11,** 143 (2011).
39. Symmans, W. F. *et al.* Genomic index of sensitivity to endocrine therapy for breast cancer. *J. Clin. Oncol.* **28,** 4111–4119 (2010).
40. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009 (2002).
41. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl Acad. Sci. USA* **100,** 11606–11611 (2003).
42. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
43. Kim, J. H., Hahm, B., Kim, Y. K., Choi, M. & Jang, S. K. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* **298,** 395–405 (2000).
44. Liu, X. & Mertz, J. E. HnRNP L binds a cis-acting RNA sequence element that enables intron- independent gene expression. *Genes Dev.* **9,** 1766–1780 (1995).
45. Hui, J., Stangl, K., Lane, W. S. & Bindereif, A. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat. Struct. Biol.* **10,** 33–37 (2003).
46. Rothrock, C. R., House, A. E. & Lynch, K. W. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *EMBO J.* **24,** 2792–2802 (2005).
47. Hung, L.-H. *et al.* Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* **14,** 284–296 (2008).
48. Rossbach, O. *et al.* Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* **11,** 146–155 (2014).
49. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31,** 3429–3431 (2003).
50. Mouridsen, H. *et al.* Letrozole therapy alone or in sequence with tamoxifen in women with breast cancer. *N. Engl. J. Med.* **361,** 766–776 (2009).
51. Winer, E. P. *et al.* American Society of Clinical Oncology technology assessment on the use of aromatase inhibitors as adjuvant therapy for postmenopausal women with hormone receptor-positive breast cancer: status report 2004. *J. Clin. Oncol.* **23,** 619–629 (2005).
52. Xue, X. *et al.* LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. *Oncogene* **35,** 2746–2755 (2015).
53. Godinho, M., Meijer, D., Setyono-Han, B., Dorssers, L. C. J. & Van Agthoven, T. Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. *J. Cell. Physiol.* **226,** 1741–1749 (2011).
54. Poliakov, A. *et al.* Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. *Mol. Cell. Proteomics* **10,** M110.007039 (2011).
55. Vizcaíno, J., Deutsch, E. & Wang, R. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32,** 223–226 (2014).
56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
57. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
58. Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* **42,** 764–770 (2014).
59. Qin, Z. S. *et al.* HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data. *BMC Bioinformatics* **11,** 369 (2010).
60. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
61. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–578 (2012).
62. Ma, X.-J. *et al.* Gene expression profiles of human breast cancer progression. *Proc. Natl Acad. Sci. USA* **100,** 5974–5979 (2003).
63. McLendon, R. *et al.* Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
64. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).
65. Maher, C. A. *et al.* Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl Acad. Sci. USA* **106,** 12353–12358 (2009).
66. Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7,** 709–715 (2010).
67. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).
68. Maticzka, D., Lange, S. J., Costa, F. & Backofen, R. GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol.* **15,** R17 (2014).
69. Raj, A., van den Bogaard, P., Rifkin, S. a., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5,** 877–879 (2008).
70. Pitchiaya, S., Androsavich, J. R. & Walter, N. G. Intracellular single molecule microscopy reveals two kinetically distinct pathways for microRNA assembly. *EMBO Rep.* **13,** 709–715 (2012).

Acknowledgements

We thank K.W.-R. for help with the mouse Xenograft work, the University of Michigan DNA Sequencing Core for Sanger sequencing and K. Giles for critically reading the manuscript and for the submission of documents. This work was supported in part by US National Institutes of Health Prostate Specialized Program of Research Excellence grant P50CA186786, Early Detection Research Network grant UO1 CA111275, US National Institutes of Health grants R01 CA132874 and RO1 CA154365 (A.M.C.), and US Department of Defense grant PC100171 (A.M.C.). A.M.C. is supported by the Prostate Cancer Foundation and the Howard Hughes Medical Institute. A.M.C. is an American Cancer Society Research Professor and a Taubman Scholar of the University of Michigan. R.M. was supported by a Prostate Cancer Foundation Young Investigator Award and by US Department of Defense Post-Doctoral Fellowship W81XWH-13-1-0284. Y.S.N. is supported by a University of Michigan Cellular and Molecular Biology National Research Service Award Institutional Predoctoral Training Grant.

Author contributions

Y.S.N., S.H., A.M.C. and F.Y.F. conceived the study and analyses. Y.S.N. performed all bioinformatics analyses and guided study design with the assistance of M.K.I. and S.G.Z. S.H. performed cellular and molecular biology experiments with the assistance T.M., R.M., C.Z., J.R.P., Y.H., S.R., S.K., R.F.S., L.X. and U.S., K.W.-R. performed mouse xenograft work. X.C. generated RNA-seq libraries and performed the sequencing. A.P. performed the mass spectrometry. M.U., A.G., C.C., R.B. and C.S.S. performed the *in silico* binding predictions. J.M.R. produced the tamoxifen resistance cell line. D.F.H. and L.J.P. provided guidance and analysis for the clinical work. Y.S.N., A.M.C.

and F.Y.F. wrote the manuscript. All authors discussed results and commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: Oncomine is supported by ThermoFisher, Inc. (Previously Life Technologies and Compendia Biosciences). A.M.C. was a co-founder of Compendia Biosciences and served on the scientific advisory board of Life Technologies before it was acquired. D.F.H. receives research funding from Astra-Zeneca, Eli Lilly, Janssen, Pfizer and Puma. He has stock options in OncImmune and in InBiomotion, and he is named as a primary inventor on two patents issued to the University of Michigan, one of which is licensed to Janssen and for which he receives royalties. These patents are not pertinent to the work presented here. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Niknafs, Y. S. *et al.* The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. *Nat. Commun.* 7:12791 doi: 10.1038/ncomms12791 (2016).

NEOPLASIA
www.neoplasia.com

Volume 18 Number 8 August 2016 pp. 489–499 489

Identification and Validation of *PCAT14* as Prognostic Biomarker in Prostate Cancer[1]

Sudhanshu Shukla[*,‡,2], Xiang Zhang[*,†,2], Yashar S. Niknafs[*,2], Lanbo Xiao[*,‡,2], Rohit Mehra[*,‡], Marcin Cieślik[*,‡], Ashley Ross[§], Edward Schaeffer[§], Bhavna Malik[¶], Shuling Guo[#], Susan M. Freier[#], Huynh-Hoa Bui[#], Javed Siddiqui[*,‡], Xiaojun Jing[*,‡], Xuhong Cao[*,‡], Saravana M. Dhanasekaran[*,‡], Felix Y. Feng[*,¶,3], Arul M. Chinnaiyan[*,‡,,††,‡‡,3] and Rohit Malik[*,‡,3]**

*Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, USA; [†]Department of Urology, Qilu Hospital of Shandong University, Jinan, China; [‡]Department of Pathology, University of Michigan, Ann Arbor, USA; [§]James Buchanan Brady Urological Institute, Johns Hopkins University, USA; [¶]Department of Radiation Oncology, University of Michigan, Ann Arbor, USA; [#]Ionis Pharmaceutical, CA; [**]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, USA; [††]Department of Urology, University of Michigan Medical School, Ann Arbor, MI, USA; [‡‡]Comprehensive Cancer Center, University of Michigan, Ann Arbor, USA

Abstract

Rapid advances in the discovery of long noncoding RNAs (lncRNAs) have identified lineage- and cancer-specific biomarkers that may be relevant in the clinical management of prostate cancer (PCa). Here we assembled and analyzed a large RNA-seq dataset, from 585 patient samples, including benign prostate tissue and both localized and metastatic PCa to discover and validate differentially expressed genes associated with disease aggressiveness. We performed Sample Set Enrichment Analysis (SSEA) and identified genes associated with low versus high Gleason score in the RNA-seq database. Comparing Gleason 6 versus 9+ PCa samples, we identified 99 differentially expressed genes with variable association to Gleason grade as well as robust expression in prostate cancer. The top-ranked novel lncRNA *PCAT14,* exhibits both cancer and lineage specificity. On multivariate analysis, low *PCAT14* expression independently predicts for BPFS ($P = .00126$), PSS ($P = .0385$), and MFS ($P = .000609$), with trends for OS as well ($P = .056$). An RNA in-situ hybridization (ISH) assay for *PCAT14* distinguished benign vs malignant cases, as well as high vs low Gleason disease. *PCAT14* is transcriptionally regulated by AR, and endogenous *PCAT14* overexpression suppresses cell invasion. Thus, Using RNA-sequencing data we identify *PCAT14,* a novel prostate cancer and lineage-specific lncRNA. *PCAT14* is highly expressed in low grade disease and loss of *PCAT14* predicts for disease aggressiveness and recurrence.

Neoplasia (2016) 18, 489–499

Address all correspondence to: Arul M. Chinnaiyan, M.D., Ph.D., Investigator, Howard Hughes Medical Institute, American Cancer Society Professor, S. P. Hicks Endowed Professor of Pathology, Professor of Pathology and Urology, Comprehensive Cancer Center, University of Michigan Medical School, 1400 E. Medical Center Dr. 5316 CCGC, Ann Arbor, MI 48109-0602.
E-mail: arul@med.umich.edu

[1]Funding/Support: This work was supported in part by the Prostate Cancer Foundation (F.Y.F, A.M.C), the National Institutes of Health Prostate SPORE (P50CA186786 to A.M.C) and the Early Detection Research Network (U01CA111275 and U01CA113913 to A.M.C). A.M.C. is supported by the Alfred A. Taubman Institute, Howard Hughes Medical Institute and the American Cancer Society. R.M is supported by a Department of Defense postdoctoral award (W81XWH-13-1-0284). R.M. and M.C. are supported by a Prostate Cancer Foundation Young Investigator award. The sponsors played no role in the design and conduct of the study.
[2]Equal Contributions.
[3]Co-senior authors.
Received 5 July 2016; Accepted 6 July 2016

1476-5586
http://dx.doi.org/10.1016/j.neo.2016.07.001

Introduction

Early detection of prostate cancer, largely facilitated by the advent of PSA screening, has also been attributed to over-diagnosis and overtreatment of this disease [1–3]. While coupling PSA screening with other biomarkers such as the long non-coding RNA (lncRNA) transcript *PCA3* or gene fusions events (such as TMPRSS2-ERG) have increased specificity of cancer diagnosis, these biomarkers have limited utility in stratifying patients in terms of prognosis [4,5]. While stratifying patients into risk groups based on clinicopathologic features is currently used to guide treatment decisions [6], it is clear that current stratification approaches need to be further refined to allow better personalization of therapy. Thus, identifying molecular biomarkers to distinguish indolent versus aggressive disease would address an unmet need in the clinical management of prostate cancer.

Advances in next-generation sequencing technologies have enabled thorough characterization of cancer transcriptomes, especially in unraveling the realm of non-coding RNAs (ncRNAs) [7,8]. In particular, lncRNAs, a class of ncRNAs, have gained increasing attention as biomarkers due to their tissue- and cancer-specific expression profile [9]. In this study, we assembled and analyzed a large RNA-seq compendium compiled from recent publications from consortiums such as The Cancer Genome Atlas (TCGA), the Prostate Cancer Foundation/Stand Up to Cancer international team, and others to identify differentially expressed genes (both protein coding and non-coding genes), that are associated with indolent versus aggressive disease [10,11]. Our results identify *PCAT14*, a prostate cancer- and lineage-specific lncRNA, as a top differentially expressed gene in this context. We characterize *PCAT14* preclinically and demonstrate that it correlates inversely in expression with disease aggressiveness and adds to conventional clinicopathologic risk factors in predicting prognosis in prostate cancer patients. Finally, we develop a novel in-situ hybiridation (ISH)-based approach for detecting *PCAT14* in clinical samples.

Material and Methods

RNA-Seq Data Set

Prostate RNA-seq cohort (n = 585) containing 52 benign prostate tissues, 501 primary prostate cancers, and 132 metastatic prostate cancers was used in this study. For nomination of Gleason associated genes, we compared low Gleason tumors (Gleason 6, n = 45) to high Gleason tumors (Gleason 9+, n = 140).

RNA-seq Data Processing

TCGA prostate Fastq files were obtained from the CGhub. Reads were aligned using STAR version 2.4.2 [12] and read abundance was calculated using FeatureCounts version 1.4.6 [13].

RNA-Seq Differential Expression Testing

Differential expression testing was performed using the Sample Set Enrichment Analysis (SSEA) tool described previously [7]. Briefly, following count data normalization, SSEA performs the weighted KS-test procedure described in GSEA [14]. The resulting enrichment score (ES) statistic describes the enrichment of the sample set among all samples being tested. To test for significance, SSEA enrichment tests are performed following random shuffling of the sample labels. These shuffled enrichment tests are used to derive a set of null enrichment scores (1000 null enrichment scores computed). The nominal p value reported is the relative rank of the observed enrichment score within the null enrichment scores. Multiple hypothesis testing is performed by comparing the enrichment score of the test to the null normalized enrichment score (NES) distributions for all transcripts in a sample set. This null NES distribution is used to compute FDR q values in the same manner used by GSEA [14]. SSEA percentile score determined by ranking the genes in each analysis by their NES score.

Tissue Expression Heatmap Generation

The "gplots" R-package was used to generate heatmaps using the *heatmap.2* function. Expression was normalized as log2 of the fold-change over the median of the normal samples for each transcript. Unsupervised hierarchical clustering was performed with the *hclust* function, using Pearson correlation as the clustering distance, using the "ward" agglomeration method.

Identification of Genes Differentially Expressed in Prostate Cancer of Varying Gleason Score

Differentially expressed Gleason associated genes were identified as any gene with an SSEA FDR< 0.01 when comparing Gleason 6 primary tumors to Gleason 9+ primary tumors. Filtering for expression levels in tissues was done by enforcing that each gene had >5FPKM expression in the top 5% of prostate tumor samples. Filtering for overexpression in cancers versus normal was done by enforcing an SSEA FDR of <0.0001 in an analysis comparing the TCGA prostate cancer vs normal tissues. Tissue specificity percentile was determined as the SSEA percentile for each gene in an SSEA analysis comparing the TCGA prostate samples to all other TCGA tumors in our multi-tissue compendium [7].

Clinical Analysis

To assess the prognostic value of *PCAT14*, microarray data was obtained from the Johns Hopkins University (JHU) (N = 355). Patients were treated with prostatectomy and subsequently received no adjuvant or salvage treatment until metastasis. Microarray processing and normalization was performed as described previously [15]. *PCAT14* expression was calculated by taking the mean expression of probe sets mapping to exons. High/low *PCAT14* was determined by splitting on the median expression level. Kaplan–Meier curves are shown and statistical inference was performed using the Log-rank test. Multivariate analysis was performed using Cox regression. Age was treated as a continuous variable. PSA was grouped into low (<10 ng/ml), intermediate (10–20 ng/ml), and high (>20 ng/ml). Surgical margin status (SMS), seminal vesicle invasion (SVI), extracapsular extension (ECE), and lymph node invasion (LNI) were treated as binary variables. Gleason score was grouped into low (≤7) or high (8–10). Association of *PCAT14* and clinicopathologic variables was evaluated using a *t*-test for continuous variables, and a chi-squared test for categorical variables. Statistical significance was set as a two-sided *p*-value <0.05. All analyses were performed in R 3.1.2.

ISH Analysis

PCAT14 ISH was performed on thin (approximately 4 μm thick) TMA sections (Advanced Cell Diagnostics, Inc., Hayward, CA), as described previously [16,17]; in parallel, *PCAT14* ISH was performed on previously identified positive and negative control index formalin-fixed paraffin embedded (FFPE) tissue sections. All slides were examined for *PCAT14* ISH signals in morphologically intact cells and scored manually by a study pathologist (Rohit Mehra). Specific *PCAT14* ISH signal was identified as brown, punctate dots, and expression level was scored as follows: 0 = no staining or less than 1 dot per 10 cells, 1 = 1 to 3 dots per cell, 2 = 4 to 9 dots per cell (few

or no dot clusters), 3 = 10 to 14 dots per cell (less than 10% in dot clusters), and 4 = greater than 15 dots per cell (more than 10% in dot clusters). For each evaluable tissue core, a cumulative ISH product score was calculated as the sum of the individual products of the expression level (0 to 4) and percentage of cells (0 to 100) (i.e., [A% × 0] + [B% × 1] + [C% × 2] + [D% × 3] + [E% × 4]; total range = 0 to 400). For each tissue sample, the ISH product score was averaged across evaluable TMA tissue cores. All quantitative data were shown as mean ± S.D. To obtain significance in the difference between two groups was performed by two-sided t test using Graph Pad Prism 6.02 software.

Cell Lines, Tissues and Reagents

All prostate cell lines used in this study were purchased from the American Type Culture Collection (ATCC), cultured according to their recommendations and were periodically checked for mycoplasma contamination and genotyped to confirm identity. For androgen treatment experiments, VCaP cells were pre-cultured in androgen-free charcoal-stripped medium for 48 hours and treated with 10 nM dihydrotestosterone (DHT) or 10 μM MDV3100 or vehicle (ethanol) for indicated time points before cells were harvested for RNA isolation. For drug treatment experiments, LNCaP cells were treated with the 5–20 μM DNA methylation inhibitor 5-aza-2′-deoxycytidine (5-aza) (catalog: A3656-5MG, Sigma), or DMSO for 5 days. RNA was isolated 24 h after drug treatment and expression was analyzed by qRT-PCR.

Prostate specimens were acquired from the patients who underwent radical prostatectomy and from the Rapid Autopsy Program at the tissue core of University of Michigan as part of the University of Michigan Prostate Cancer Specialized Program Of Research Excellence (S.P.O.R.E.). Informed consents were obtained from each patient.

RNA Isolation and qPCR Analysis

Total RNA was extracted using Trizol reagent and an RNeasy Micro Kit (Qiagen) with DNase I digestion according to the manufacturer's protocols. RT-PCR was performed from total RNA using Superscript III (Invitrogen) with random primers (Invitrogen). Quantitative PCR (qPCR) was performed using Fast SYBR Green Master Mix (Applied Biosystems) on a 7900HT Fast Real-Time PCR system (Applied Biosystems). All oligonucleotide primers were purchased from Integrated DNA Technologies (Coralville, IA) are sequence of each primer is listed in Supplementary Table 4. Primer specificity was determined by sequence verifying the PCR products using the University of Michigan Sequencing Core facility.

Rapid Amplification of cDNA Ends (RACE)

5′ and 3′ RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen) according to the manufacturer's protocols. RACE PCR products obtained using Platinum Taq high-fidelity polymerase (Invitrogen), were resolved on a 1.5% agarose gel. Individual bands were gel purified using a Gel Extraction kit (Qiagen), and cloned into PCR4 TOPO vector, and sequenced using M13 primers.

Knock Down Studies

MDA-PCa-2b and VCaP cells were seeded in biocoated 6-well plates at 60% confluency, incubated overnight, and transfected with 50 nM siRNAs targeting different exons of *PCAT14* or non-targeting siRNAs, using RNAi MAX reagent (Invitrogen) per manufacturer's instructions. RNA was harvested 48 h after transfection. Functional experiments were performed at indicated time points. Sequence of all the siRNA used in shown in Supplementary Table 4.

Nuclear-Cytoplasmic Subcellular Fractionation

Nuclear-cytoplasmic fraction of MDA-PCa-2b and VCaP cells was performed using an NE-PER Nuclear and Cytoplasmic Extraction kit (Thermo Scientific) following manufacturer's instructions, followed by RNA isolation and qPCR analysis.

CRISPR Based Overexpression of PCAT14

Stable cell lines overexpressing *PCAT14* endogenously were made using previously published protocol [18]. Briefly, guide RNAs targeting promoter region of *PCAT14* (Supplementary Table 4) were designed using online tool at http://crispr.mit.edu/ and cloned into sgRNA-MS2 vector using lenti sgRNA(MS2) zeo backbone. Lentiviral particles expressing *PCAT14* sgRNA-MS2 were generated by the University of Michigan vector core. To generate LNCaP or PC3 cell over expressing *PCAT14*, first cells were seeded into 100 mm dish and transduced with Lenti dCAS-VP64 (blasticidin) and Lenti-MS2-p65-HSF1 (hygromycin) vectors. After 2 days, cells were selected with 4 μg/ml Blasticidin and 200 μg/ml Hygromycin. Cells stably expressing dCAS-VP64 and MS2-p65-HSF1 cells were then seeded in 6-well plates and infected with *PCAT14* sgRNA-MS2 lentivirus. 24 hours later, cells were selected with triple antibiotics: 4 μg/ml Blasticidin, 200 μg/ml Hygromycin and 800 μg/ml Zeomycin for 1 week. Expression of *PCAT14* in these cells was verified by qPCR.

In Vitro FluoroBlok Tumor Invasion Assay

The In vitro FluoroBlok Tumor Invasion Assay (BD) was performed as previously described [19]. Briefly, after rehydration of the BD FluoroBlok membrane, 500 ul of serum-free RPMI medium resuspended prostate cancer cells (PC3, 50,000 cells per well, or LNCaP, 100,000 cells per well) were seeded into the apical chambers. 750 ul RPMI medium containing 10% FBS were added to the basal chamber as chemoattractant. Then plates were incubated at 37 °C, 5% CO_2 for 24 hours. Following incubation, medium from the apical chambers were removed, and the inserts were transferred to a 24-well plate containing 500ul/well of 4ug/mL Calcein AM (Invitrogen) in Hanks buffered saline. Plates were incubated for 1 hour at 37 °C, 5% CO_2, then pictures of invaded cells were taken by using inverted fluorescence microscope (Olympus), and quantified by ImageJ software [20].

Oncomine Concepts Analysis of the PCAT14 Signature

Gene that positively correlated ($R^2 > 0.35$, n = 591) with *PCAT14* in TCGA RNA-seq data were selected and uploaded into Oncomine database [21] as custom concepts (Supplementary Table 2). All the prostate cancer concepts with odds ratio> 2.0 and *p*-value $<1 \times 10^{-4}$ were selected. For simplicity, top 4 concepts (based on odds ratios) were selected for representation. We exported these results as the nodes and edges of a concept association network and visualized the network using Cytoscape version 3.3.0. Node positions were computed using the Edge-weighted force directed layout in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of node labels.

Statistics

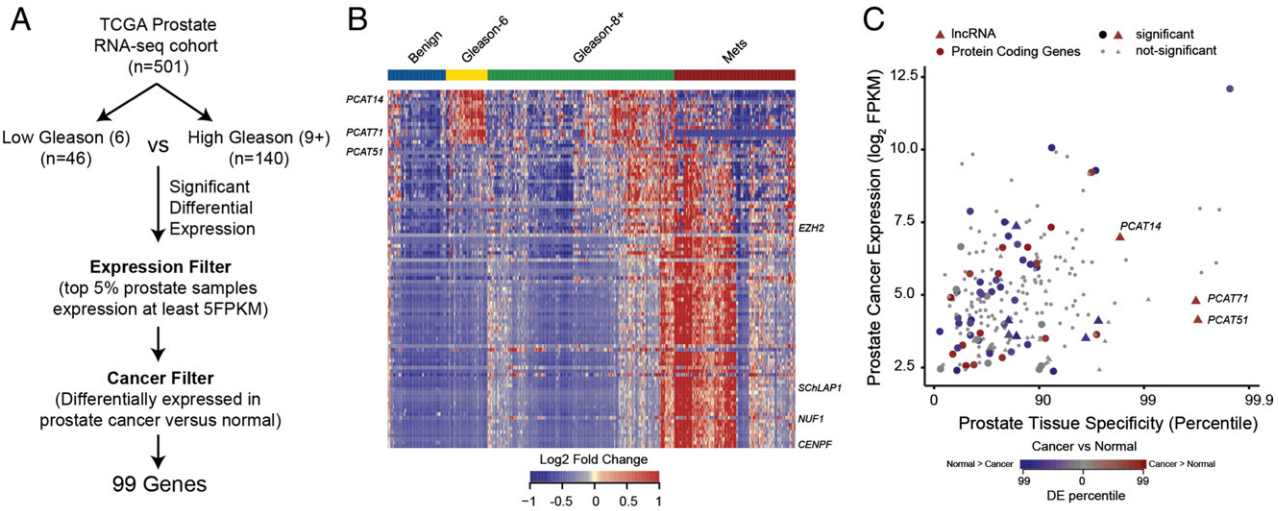
All quantitative data were shown as mean ± S.D. To obtain significance in the difference between two groups was performed by two-sided t test or ANOVA using Graph Pad Prism 6.02 software.

Results

Identification of Genes Associated With Gleason Grade in Prostate Cancer

Comprehensive molecular characterization of common cancer types has become feasible with the recent availability of large next generation sequencing datasets on tumor tissues. To identify genes (both coding and non-coding) associated with aggressive prostate

cancer, we assembled a large prostate RNA-seq cohort (n = 585) containing 52 benign prostate tissues, 501 primary prostate cancers, and 132 metastatic prostate cancers. We performed differential expression testing utilizing a non-parametric tool we developed for RNA-seq data called Sample Set Enrichment Analysis [7]. In order to nominate the most intriguing biomarkers associated with aggressive disease, we compared low Gleason tumors (Gleason 6, n = 45) to high Gleason tumors (Gleason 9+, n = 140) and applied filters for substantial expression in prostate tumor tissue (>5PKM in the top 5% of prostate samples), and significant differential expression in



D

Gene Name	Category	Gleason 9+ Percentile	Prostate Expression (95%ile FPKM)	Cancer vs Normal Percentile	Tissue Sepcificity Percentile
PCAT14	lncRNA	99.57	125.589	97.93	98.40
TSPAN1	Protein coding	99.63	594.854	98.05	96.92
PTRHD1	Protein coding	99.38	160.925	99.03	92.41
DUS1L	Protein coding	99.57	99.683	99.02	87.25
GMDS	Protein coding	99.91	99.117	98.31	77.82

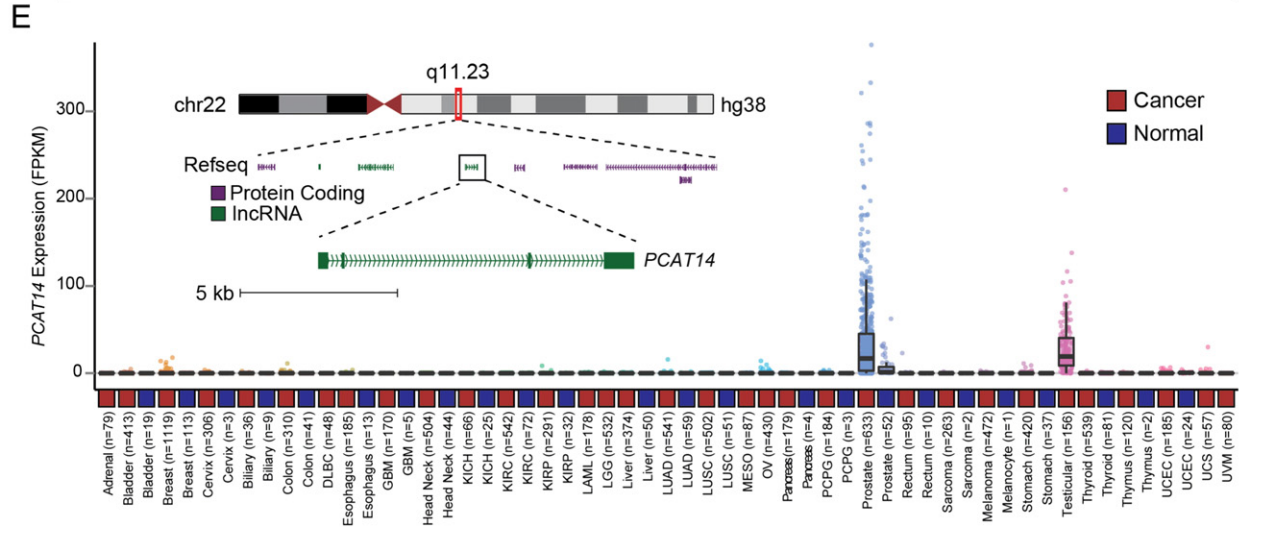


Figure 1. Identification of lncRNA *PCAT14* as a prostate cancer biomarker. A. Schematic representation of the workflow utilized to identify highly-expressed, prostate cancer specific genes associated with low-Gleason disease. B. Heatmap depiction of the lncRNA and protein coding genes differentially expressed (n = 99) between Gleason 6 versus 9+ analysis in TCGA prostate RNA-seq data. Relative expression of these genes in benign and metastatic prostate cancer tissues [11] are also displayed alongside for comparison. Expression is depicted as log2 of the fold-change over the median of the Gleason 6 samples for each gene. Patients grouped by cancer progression/Gleason score. Rows represent genes and columns represent samples. C. Scatterplot showing the expression level, prostate tissue specificity, and prostate cancer association of protein coding (solid circle) and lncRNA (solid triangles) genes identified in 1 A. Expression is represented by the FPKM value for the 95th percentile prostate cancer sample. Cancer versus normal and prostate tissue specificity are represented by the percentile score for each gene in an SSEA analysis. D. The top five, Gleason 6 associated genes listed in the order of prostate tissue specificity. E. Expression of *PCAT14* across all cancer and normal tissue type represented in the TCGA. Inset shows genome browser view of *PCAT14* genomic location.

prostate cancers versus normal (SSEA, FDR <0.0001) leaving a total of 99 candidates genes (Figure 1*A*, Supplementary Table 1). Interestingly, clustering analysis revealed signature expression patterns, specifically associated with low, high Gleason and metastatic status and included both novel and previously characterized genes (Figure 1*B*). *CENPF* and EZH2, protein coding genes with a known association with high grade prostate cancer were rediscovered through this analysis [22,23]. Similarly, we rediscovered *SChLAP1* a long non-coding RNA (lncRNA) associated with aggressive prostate cancers [15,17] in our analysis (Figure 1*B*). With the goal of identifying potential biomarkers that distinguish indolent prostate cancers, we focused on genes enriched in low grade disease that are expressed highly in prostate tissue and that also show prostate cancer and tissue specificity (Figure 1*C*). Interestingly, a lncRNA, *PCAT14* appeared to be one of the top low-Gleason-associated genes with robust prostate tissue expression, substantial prostate tissue specificity, and significant overexpression in prostate cancers versus normal (Figure 1*D*). In fact, among all genes (coding and non-coding), *PCAT14* ranked among the top 5 in terms of expression level, Gleason 6 versus 9+ association, and cancer versus normal association (Figure 1*D*). Additionally, among the top 5 candidate genes, *PCAT14* was the only gene to exhibit striking prostate tissue specificity, a particularly relevant metric for a potential biomarker (Figure 1*E*). The remaining 4 genes exhibited variable prostate tissue specificity (Supplementary Figure 1). *PCAT14* is a poly-exonic gene found within a gene desert on chromosome 22, with a striking prostate cancer and lineage specific expression pattern across the >10,000 TCGA cancer and normal tissue samples (Figure 1*E*). For these reasons, we elected to pursue *PCAT14* as a promising biomarker that can identify low grade prostate cancer.

Genomic Organization and Regulation of PCAT14

We collected multiple lines of evidence from both experimental data and available annotations to consolidate the genomic organization of *PCAT14*. Based on assembled reads from RNA-seq data assembled in the MiTranscriptome [7], we predicted the structure of the *PCAT14* transcript variants (Supplementary Figure 1A). Additionally, as an independent approach to define the exon structure of *PCAT14,* we performed rapid amplification of cDNA ends (RACE) in two prostate cancer cell lines VCaP and MDA-PCa-2b that express *PCAT14* at high levels (Supplementary Figure 1B and C). Our analyses show that the *PCAT14* gene is located on chr22-q11.2 and contains 4 exons. Among the four transcript isoforms, the 2.3 kb variant-1 demonstrates the highest expression (Supplementary Figure 1D). Next, using published ChIP-seq data in VCaP cells [24], we show that *PCAT14* has all the histone marks (H3K4me3, H3K36me3, H3K27ac) associated with actively transcribed genes (Figure 2A). We further performed subcellular fractionation followed by qPCR to show that *PCAT14* is distributed equally between nuclear and cytoplasmic compartments (Figure 2*B*).

Androgen receptor plays a major role To identify any potential regulation of *PCAT14* gene by androgen, we assessed the presence of AR peaks in *PCAT14* genomic region using AR-ChIP-seq data generated in VCaP cells [24] and saw significant AR peaks in *PCAT14* loci. Some of these peaks were also enhanced upon treatment with DHT and were suppressed upon treatment with AR antagonist MDV3100 or bicalutamide (Figure 2*C*). To corroborate this finding, we assessed the expression of *PCAT14* mRNA in VCaP cells upon AR stimulation. Similar to the canonical AR targets such as *KLK3* and *FKBP5*, *PCAT14* expression was also significantly elevated (four fold in 24 hours) upon DHT stimulation (Figure 2*D*) and suppressed by MDV3100 treatment (Figure 2*E*). In another line of investigation, we queried if epigenetic regulation might play a role in the prostate cancer and lineage specific expression of *PCAT14* observed in tissue samples (Figure 1*E*). Using a prostate cancer cell line (LNCaP) model we show significant elevation of *PCAT14* expression when treated with 5-azacytidine (5-Aza), a DNA demethylation agent, suggesting a potential role for promoter methylation in regulation of *PCAT14* (Figure 2*F*). However, our attempt to capture this event in TCGA tissue samples where Infinium 450 K DNA methylation array data is available was inconclusive, due to the lack of probes in *PCAT14* promoter region. Taken together we show *PCAT14* is an AR target gene that may also be subjected to epigenetic regulation in prostate cancer.

Clinical Association of PCAT14

Having observed an inverse correlation of *PCAT14* with Gleason Score (GS) in our RNA-seq cohort, we next assessed the association of *PCAT14* expression with clinical outcomes in prostate cancer. For this analysis we first divided samples into 7 groups (benign, GS-6, GS-7 (3 + 4), GS7 (4 + 3), GS-8, GS-9 and Mets) and examined the expression of *PCAT14* using two different datasets (TCGA and Taylor et al.). We identified a significant decrease in *PCAT14* expression as Gleason grade increased in both cohorts (Figure 3*A* and *B*). Importantly, in the large TCGA dataset, expression was significantly different between GS6 and all other groups except GS7 (3 + 4). We next assessed the diagnostic ability of *PCAT14* to identify prostate cancers versus normal. In both the TCGA and Taylor prostate cancer cohorts, *PCAT14* expression was able to significantly distinguish cancer from normal with an AUC of 0.837 and 0.823 respectively (Figure 3*C*) supporting its utility as a diagnostic biomarker.

Using an alternate approach to further characterize the clinical associations of *PCAT14*, we performed a "guilt-by-association" analysis, assessing the clinical significance of the protein-coding genes most correlated with *PCAT14* (Supplementary Table 2) in the TCGA prostate cancer cohort, leveraging cancer microarray data from the Oncomine resource [21]. As expected, genes positively correlated with *PCAT14* were upregulated in cancer vs normal analysis and were downregulated in clinically advanced prostate cancer (Figure 3*D*). Interestingly, we found a striking association of *PCAT14* correlated genes with concepts related to better prognosis (Figure 3*D*), and these genes were under-expressed in recurrent and hormone refractory prostate cancer suggesting that *PCAT14* may be a marker of better clinical outcomes in prostate cancer. In contrast, genes that positively correlated with *SChLAP1*, a lncRNA known to be associated with clinically aggressive prostate cancer, were found to be overexpressed in advanced prostate cancer as well as in cancer with poor outcomes [15,17].

To further investigate the association of *PCAT14* with favorable clinical outcomes in prostate cancer, we performed Cox regression analysis on a cohort of 355 patients (John Hopkins University (JHU) cohort) who did not receive treatment prior to metastasis (median follow-up 9 years). Univariate analysis showed that, patients with high *PCAT14* expression were significantly associated with better BPFS (P = .000062; HR = 0.59 [0.45–0.76]), MFS (P = .00016; HR = 0.46 [0.32–0.66]), PSS (P = .0067; HR = 0.47[0.27–0.82]) and OS (P = .022; HR = 0.57 [0.35–0.93]) (Figure 4*A-D*). In a Cox multivariate analysis including clinicopathologic variables, *PCAT14* stands out as a significant independent predictor of PSS (P = .0385; HR = 0.55 [0.31–0.97]), MFS (P = .000609; HR = 0.52[0.36–0.76])

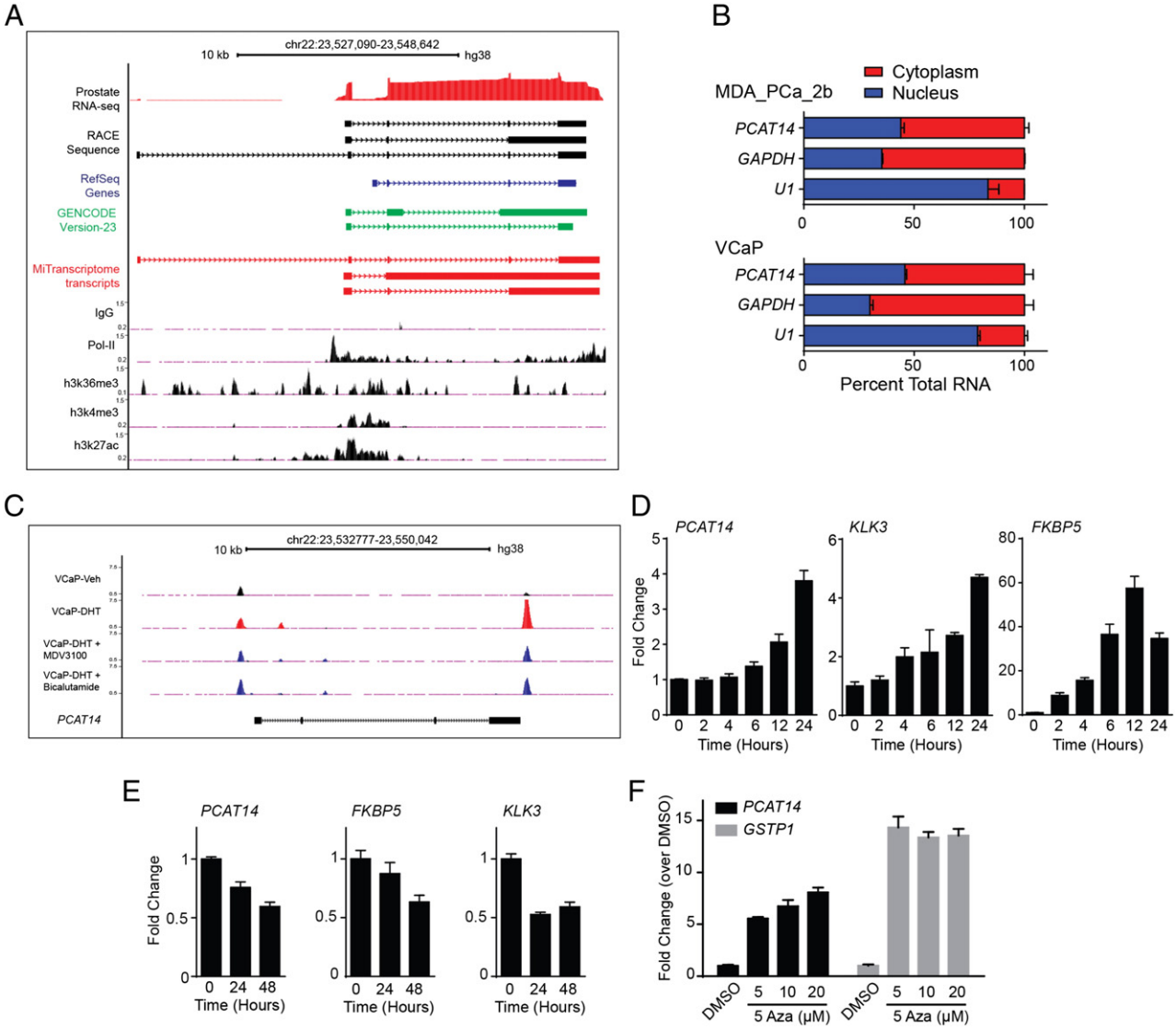


Figure 2. Subcellular localization and regulation of *PCAT14*. A. Genome browser view of *PCAT14* locus. ChIP-seq tracks for H3K4me3, H3K27ac, H3K36me3 and Pol-II generated in prostate cancer VCaP cells are shown. Prostate RNA-seq reads, transcript schematic based on RACE results and Refseq, GENCODE, MiTranscriptome assembly annotations are also provided. Solid blocks indicate exons while thin lines intron and arrows indicate the genomic orientation. B. Bar plots represent the subcellular localization of *PCAT14* in prostate cell lines. *PCAT14* transcript was equally found in both cytoplasmic (red) and nuclear (blue) compartments in both MDA-PCa-2b and VCaP cell lines. GAPDH and U1 RNA were used as controls. C. Genome browser view of the *PCAT14* genomic locus for AR ChIP-seq data tracks obtained from VCaP cells treated with either vehicle (black) or dihydrotestosterone (DHT) alone (Red) or combinations (dark blue) including DHT + MDV3100 and DHT + Bicalutamide. Significant AR binding observed in each data track are represented as peaks. D-E. Histograms represent the expression of *PCAT14*, *TMPRSS2* and *KLK3* in VCaP cells after treatment with 10 nM DHT or with MDV3100 for indicated time points. F. Bar plots represent re-expression of *PCAT14* and *GSTP1* in LNCaP cells after treatment with 5-Aza deoxycytidine (5-Aza) for 5 days at indicated concentrations.

and BRFS (P = .00126, HR = 0.64 [0.49–0.84]), with borderline significance for OS (Table 1, Supplementary Table 3). In addition, we also analyzed the association of *PCAT14* expression with clinical outcome in two independent data sets of 140 (Taylor et al) and 377 (TCGA) patients using the statistical approaches mentioned above [25]. Similar to JHU cohort, high *PCAT14* expression predicted for better BRFS (Figure 4*E*) and MFS (Figure 4*F*). We also show that high PCAT14 expression was predictor of better prognosis in lower Gleason grade samples (Supplementary Figure 3B).

PCAT14 Expression In-Situ

LncRNA detection in cancer tissue sections by RNA in-situ hybridization (RNA-ISH) technology has similar clinical utility as immunohistochemical evaluation of protein biomarkers [16,26]. Hence we evaluated *PCAT14* transcript levels in PCa FFPE tissues using specific probes to perform a RNA-ISH. We first probed a panel of FFPE sections derived from either murine prostate, kidney, lung (negative controls) or xenografts from MDA-PCa-2b cells, a cell line that expresses *PCAT14* at high levels (positive control). As expected,

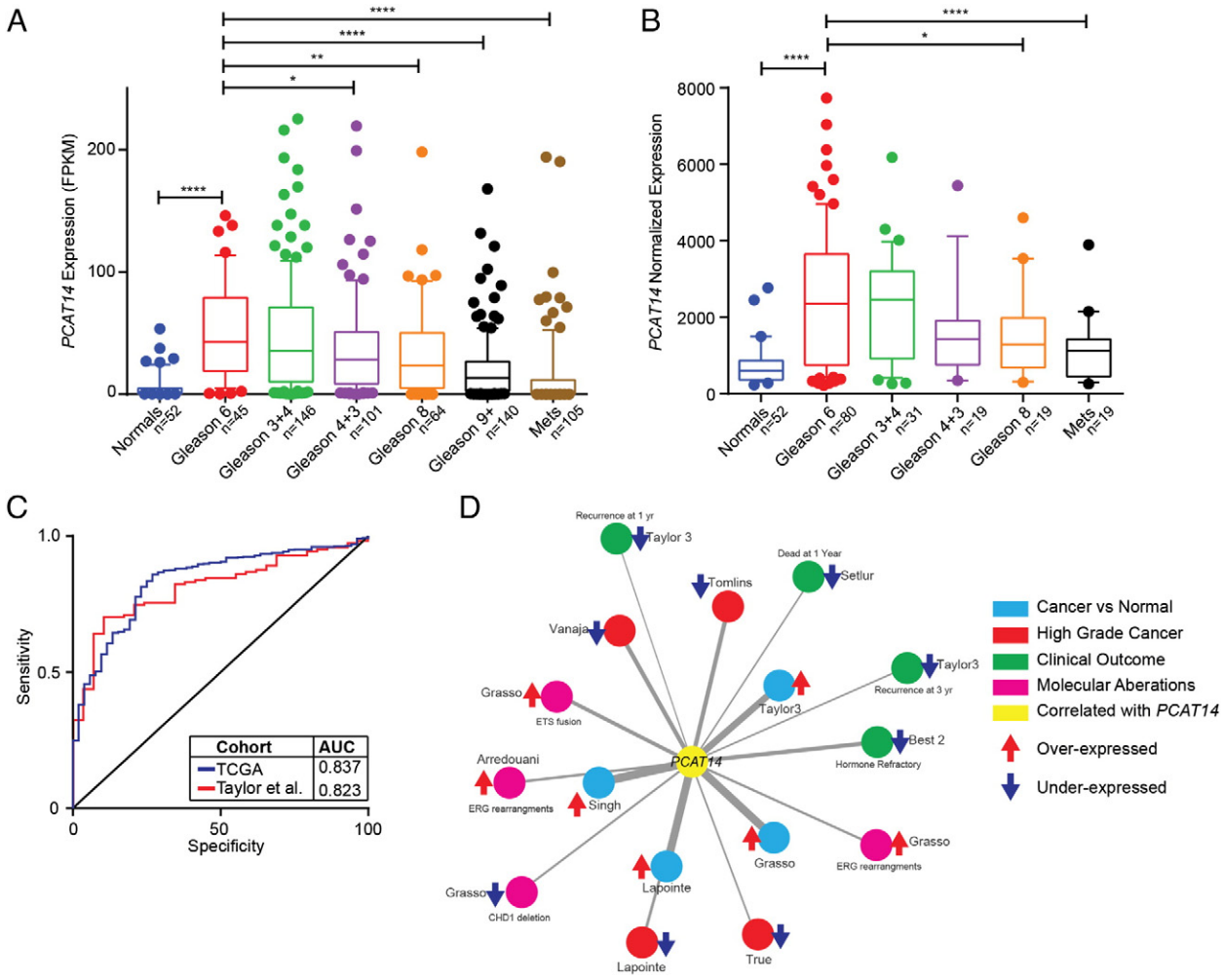


Figure 3. *PCAT14* is marker of low grade tumros. **A-B**. Expression of *PCAT14* in samples distinguished by Gleason grade in TCGA (A), Taylor (B) cohorts. (* = $P < .05$, ** = $P < .01$, **** = $P < .0001$; compared to Gleason 6). C. ROC analysis of *PCAT14* expression in the TCGA and Taylor cohorts. D. Network representation of genes positively correlated with *PCAT14* in localized prostate cancers using Oncomine concepts analysis and visualized with the Force-Directed Layout algorithm in the Cytoscape tool [29]. Node names are assigned according to the author of the primary study [25,30–38]. Nodes are colored according to the concept categories indicated in the figure legend. Thickness of the edges implies higher odds ratio.

high levels of specific signal was present in MDA-PCa-2b xenografts while no expression/staining was seen in the negative control murine tissues (Supplementary Figure 5A and B). Consistent with the cell fractionation data, expression of *PCAT14* was seen in both nuclear and cytoplasmic compartments. Next we obtained frozen and matched formalin fixed paraffin embedded (FFPE) tissues sections derived from a patient radical prostatectomy specimen with Gleason score 3 + 3 = 6 disease. q-PCR analysis on cDNA from frozen tissues derived from this specimen shows a 7–8 fold increase in *PCAT14* expression in cancer compared to the adjacent benign tissue (Figure 5*A*). RNA-ISH also demonstrated that *PCAT14* is differentially expressed in PCa as we saw striking difference of transcript expression with high signals located in the prostatic adenocarcinoma glands and with no/minimum staining in the benign section (Figure 5*B*). To further expand these results, we performed RNA-ISH on a PCa tissue microarray (TMA, n = 129) (Figure 5*C*) and found that *PCAT14* expression was able to distinguish tumor from normal (AUC 0.863) (Figure 5*D*) and was high in Gleason-6 with minimal expression noted in benign tissue or Gleason 8 disease (Figure 5*E*).

Functional Evaluation of PCAT14

Since expression of *PCAT14* was lower in high grade prostate cancer and its expression predicted better outcomes, we hypothesize that *PCAT14* may have tumor suppressive effects. To test this hypothesis, we performed overexpression studies in PC3 and LNCaP cells, prostate cancer cell lines that do not express *PCAT14* (Supplementary Figure 2B, C). To overexpress *PCAT14*, we used a CRISPR (clustered regularly interspaced short palindromic repeat)-Cas9 Synergistic Activation Mediator (SAM) complex [18]. This method allows endogenous overexpression of a gene by recruiting artificial transcriptional factors to the promoter using single-guide RNA (sgRNA-MS2) (See method section for details). We designed 6 sgRNAs targeting the *PCAT14* promoter and tested their ability to induce *PCAT14* expression using HEK293 cells stably expressing transcription factors. We found three sgRNAs that significantly increased *PCAT14* expression in HEK293 cells (Supplementary Figure 5A). We next used these sgRNA to construct PC3 and LNCaP cells stable expressing *PCAT14* (Figure 6*A*). Using two independent sgRNAs we were able to achieve 500 to 1000-fold endogenous overexpression of

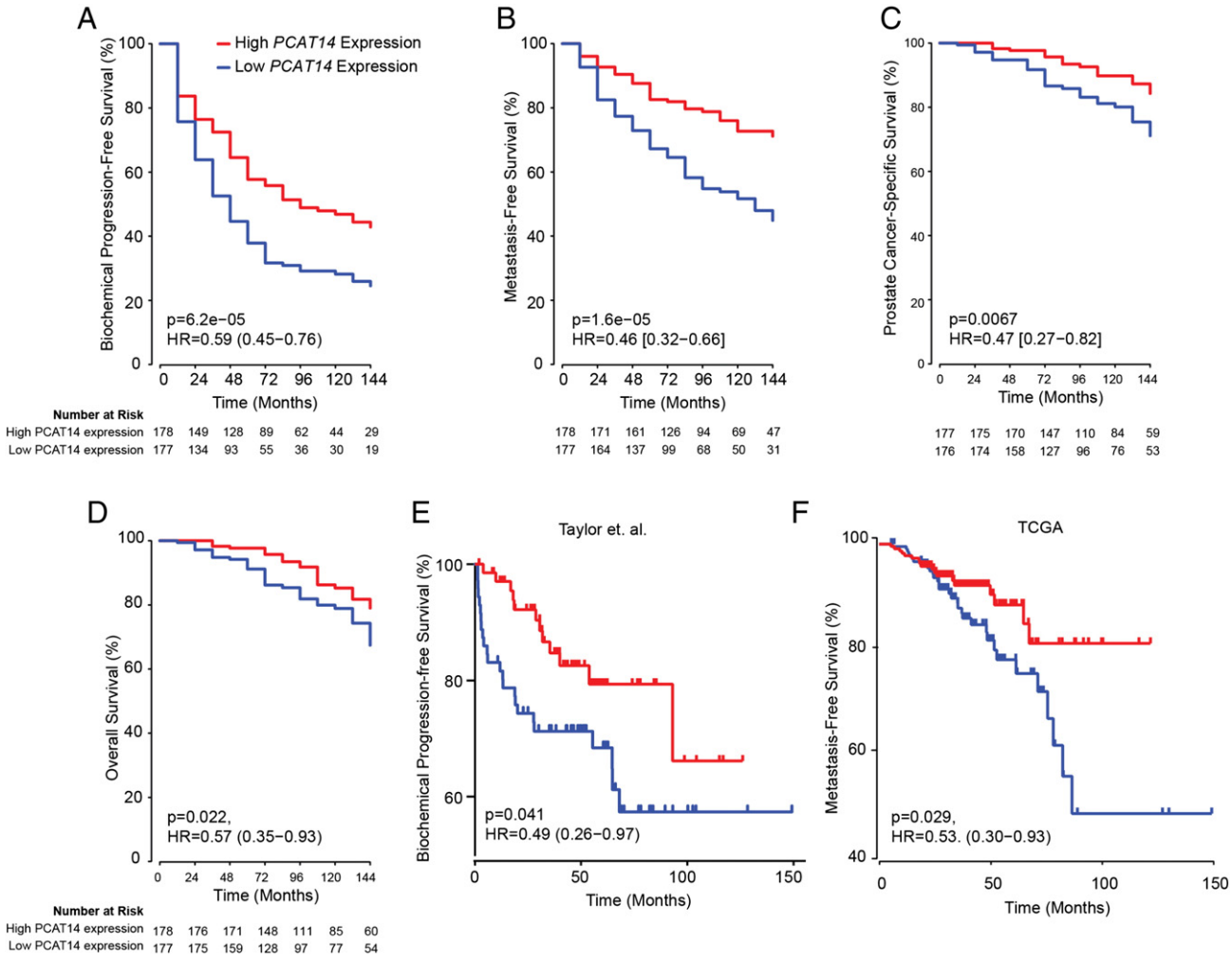


Figure 4. *PCAT14* is a prognostic biomarker. A-D. Kaplan–Meier analyses of prostate cancer outcomes in the John Hopkins cohort. *PCAT14* expression was measured using Affymetrix exon arrays, and subjects were stratified according to their *PCAT14* expression level. Subject outcomes were analyzed for biochemical progression (D) and Metastasis free survival (E), Prostate cancer-specific survival (F) and overall survival (G). Subject outcomes were analyzed for Kaplan–Meier curves, *P* values determined using a log-rank test. E-F. Kaplan–Meier analyses of biochemical progression free survival in the Taylor (E) and Metastasis Free survival in the TCGA (F) cohorts of prostate cancer. Patients were divided into two groups based on the expression level of *PCAT14*. *P* values for Kaplan–Meier curves were determined using a log-rank test.

PCAT14 in PC3 cells (Figure 6*B*) and 20–100 fold overexpression in LNCaP cells (Supplementary Figure 5*B*). While we observed no significant effect of *PCAT14* overexpression on proliferation of PC3 or LNCaP cells (Figure 6*C* and Supplementary Figure 5*C*), overexpression of *PCAT14* lead to suppression of invasion capacity of both PC3 and LNCaP cells (Figure 6*C*, *D*; Supplementary Figure 5*E*, *F*), in line with its prior identified association with clinically indolent disease. We then looked at the effects of *PCAT14* knockdown on cell expressing

Table 1. Multivariate Analysis in JHU Cohort

	Biochemical Recurrence Free Survival		Metastasis Free Survival		Prostate Cancer Free Survival		Overall Survival	
	P-Value	HR [95% CI]	*P*-Value	HR [95% CI]	*P*-Value	HR [95% CI]	*P*-Value	HR [95% CI]
PCAT14 High vs. Low	.00126	0.64 [0.49–0.84]	.000609	0.52 [0.36–0.76]	.0385	0.55 [0.31–0.97]	.0567	0.62 [0.38–1.01]
Age	.818	1 [0.98–1.02]	.65	0.99 [0.96–1.02]	.338	0.98 [0.93–1.02]	.151	0.97 [0.93–1.01]
PSA Int vs. Low	.241	0.83 [0.62–1.13]	.353	0.83 [0.55–1.24]	.385	0.75 [0.4–1.42]	.366	0.77 [0.44–1.35]
PSA High vs. Low	.916	0.98 [0.63–1.52]	.574	0.84 [0.47–1.52]	.463	0.73 [0.31–1.7]	.582	0.81 [0.39–1.7]
Gleason High vs. Low	2.98E-05	1.83 [1.38–2.43]	1.00E-08	3.08 [2.1–4.52]	.000224	3.1 [1.7–5.65]	.000988	2.38 [1.42–3.99]
Seminal vesicle invasion	.0042	1.52 [1.14–2.03]	.453	1.16 [0.79–1.69]	.774	0.92 [0.51–1.66]	.82	0.94 [0.56–1.59]
Surgical margin status	.000533	1.78 [1.28–2.47]	.000276	2.15 [1.42–3.25]	.0487	1.93 [1–3.7]	.0825	1.67 [0.94–2.99]
Extracapsular extension	.456	1.14 [0.81–1.58]	.459	1.21 [0.73–2.03]	.636	0.83 [0.39–1.77]	.816	0.93 [0.48–1.77]
Lymph node invasion	8.98E-12	3.23 [2.31–4.52]	.000164	2.21 [1.46–3.35]	.0616	1.86 [0.97–3.57]	.254	1.42 [0.78–2.6]

HR: Hazard Ratio

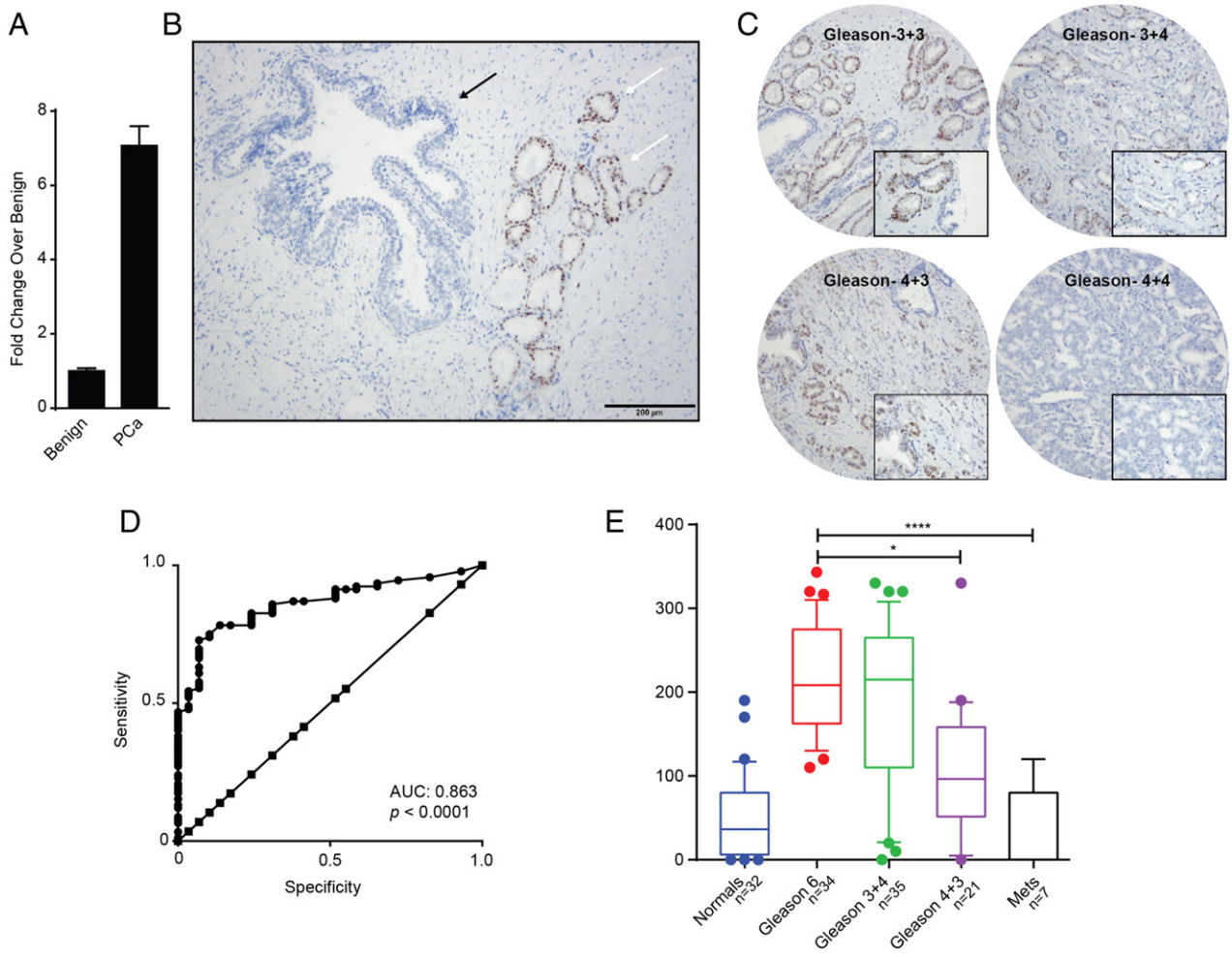


Figure 5. *PCAT14* RNA-ISH in prostate cancer tissues. A. Barplot to show the expression of *PCAT14* in tumor tissue and adjacent benign by qRT-PCR. B. A representative *PCAT14* RNA in-situ hybridization image. White arrows indicate Gleason score 6 disease and black arrows indicate benign glands. C. Representative *PCAT14* In situ hybridization images of human prostate cancer samples of different Gleason grades. D. ROC analysis of *PCAT14* expression in the prostate TMAs. E. Representation of mean *PCAT14* ISH product score for benign prostatic glands (benign), Gleason score 6, Gleason score 3 + 4 = 7, Gleason score 4 + 3 = 7 and Gleason score 8+ clinically localized prostate cancer in a TMA cohort. (** = $P < .01$; compared to benign).

PCAT14 at high levels (VCaP and MDA-PCa-2B). In both MDA-PCa-2b and VCaP cells using 2 independent siRNA as well as 8 independent ASOs we were able to achieve more than 80% knockdown efficiency (Supplementary Figure 5G-J). However, we did not observe a consistent effect on cell proliferation as well as cell invasion (Supplementary Figure 5K-N and data not shown).

Discussion

In this study, we perform a large-scale RNA-sequencing-based analysis of biomarkers associated with indolent versus aggressive prostate cancer and identify the long noncoding RNA *PCAT14* as a marker of low grade and indolent disease. We define the exon structure of *PCAT14* and demonstrate that *PCAT14* is an AR-regulated lncRNA. Using two independent data sets, we show that *PCAT14* is highly upregulated in prostate cancer compared to benign tissue and is able to distinguish prostate cancer from normal tissue with high sensitivity and specificity, suggesting that *PCAT14* can be an excellent diagnostic biomarker. Moreover, we demonstrate that expression of *PCAT14* is prognostic of outcome and is associated with better biochemical progression-free survival, metastases-free survival, and prostate cancer-specific survival. Importantly, we find that *PCAT14* expression is a prognostic biomarker which adds to standard clinicopathologic variables.

As such, *PCAT14* represents a unique biomarker. Most diagnostic biomarkers, such as *PCA3*, can distinguish cancer from normal tissue, but are not prognostic [4]. Conversely, many prognostic biomarkers, such as Ki-67, hold little diagnostic value. It is unclear why *PCAT14* increases significantly in expression during the initial formation of cancer, but then subsequently decreases in expression in disease aggressiveness; this observation requires follow up with further mechanistic studies but is also a feature that gives *PCAT14* value as a biomarker across multiple clinical contexts. Of note, *PCAT14* was also found to be expressed in testicular cancer samples along with prostate cancer, suggesting the role of *PCAT14* in the testicular cancer pathogenesis. However, due to lack of normal testis samples in the TCGA database, it is unclear, at this point, whether *PCAT14* is differentially regulated in testicular cancer compared to normal testis. Recently, the Genotype-Tissue Expression (GTEx) program has

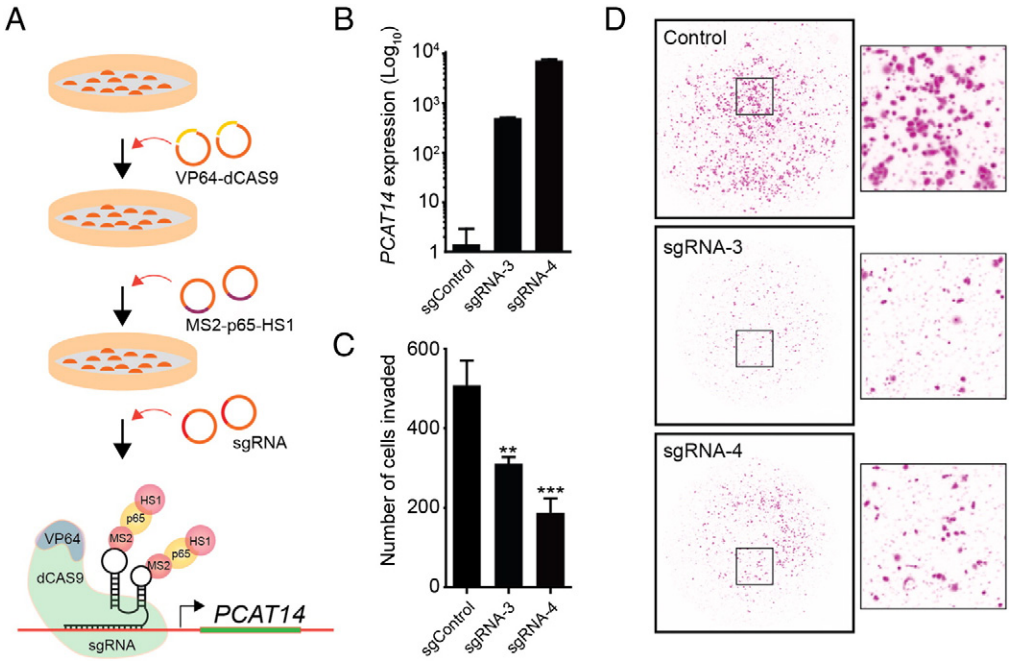


Figure 6. Functional analysis is *PCAT14*. A. Schematic representation of the workflow to endogenously overexpress *PCAT14* in prostate cancer cells using CRISPR/SAM system. B. Bar plots represent fold increase in *PCAT14* level in PC3 cells expressing dCas9–VP64 and MS2–p65–HSF1 with control or 2 independent *PCAT14* sgRNAs. C. Bar plot represent quantification of invaded PC3 cells with or without *PCAT14* expression. D. Representative images of invaded PC3 cells with or without *PCAT14* expression.

generated a large amount of high throughput sequencing data on normal tissue including testis [27]. This data would be useful to look at the role of PCAT14 in testicular carcinoma.

In an attempt to develop a clinical grade assay to detect expression of *PCAT14*, we developed a novel assay, using ISH probes, which can be applied to formalin fixed paraffin-embedded tissues. This ISH assay provides an opportunity to validate our findings in larger cohorts with associated clinical data in the future. Ultimately, an optimized approach for predicting indolent versus aggressive disease will include both clinicopathologic parameters integrated with molecular biomarkers. It is likely that this molecular assay will involve multiplexing multiple biomarkers, and may require combining both tissue-based and urine-based biomarkers. Potential intriguing subsequent studies include the assessment of *PCAT14* and other candidate lncRNAs, in addition to *PCA3*, as urine biomarkers.

There are a several limitations to our study. While we demonstrate the potential value of *PCAT14* expression as a biomarker, it is unclear how *PCAT14* is modulating oncogenic phenotypes, from a mechanistic perspective. Additionally, while we demonstrate the relative specificity of *PCAT14* for both prostate and testicular cancers, the molecular basis underlying this specificity remains to be elucidated. It is known that AR can regulate expression of genes in both prostatic and testicular tissues, but we do not know whether the relative cancer-specificity can be attributed to AR. Clearly, these are important areas for future study.

Overall, our study highlights the need to look at both conventional protein-coding genes and noncoding genes in the search for optimal biomarkers. To our knowledge, there are approximately 20,000 protein coding genes [28], which comprise 2% of the genome. Given our recent study demonstrating that there are close to 60,000 long noncoding RNAs (lncRNAs) [7], many of which are specific to certain cancers, it is clear that these lncRNAs present a relatively underexplored frontier for biomarker development, and that *PCAT14* may represent an initial candidate to be further explored along this frontier.

Conclusion

By performing differential expression analysis between prostate cancer with low vs high Gleason scores, we identified lncRNA *PCAT14* as a prostate cancer- and lineage- specific biomarker of indolent disease. We show that *PCAT14* is an AR-regulated transcript and its overexpression suppresses invasion of prostate cancer cells. Moreover, in multiple independent datasets, *PCAT14* expression associates with favorable outcomes in prostate cancer and adds prognostic value to standard clinicopathologic variables.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neo.2016.07.001.

Acknowledgments

We thank Sethuramasundaram Pitchiaya, Xia Jiang, Fengyun Su and Ingrid Apel for technical assistance; K. Giles for critically looking over the manuscript and the submission of documents; the University of Michigan Viral Vector Core for generating the lentiviral constructs.

References

[1] Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, and Crawford ED, et al (2004). Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* **350**, 2239–2246.

[2] Andriole GL, Crawford ED, Grubb 3rd RL, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, and Reding DJ, et al (2009). Mortality

results from a randomized prostate-cancer screening trial. *N Engl J Med* **360**, 1310–1319.

[3] Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, and Zappa M, et al (2009). Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* **360**, 1320–1328.

[4] Leyten GH, Hessels D, Jannink SA, Smit FP, de Jong H, Cornel EB, de Reijke TM, Vergunst H, Kil P, and Knipscheer BC, et al (2014). Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *Eur Urol* **65**, 534–542.

[5] Tomlins SA, Aubin SM, Siddiqui J, Lonigro RJ, Sefton-Miller L, Miick S, Williamsen S, Hodge P, Meinke J, and Blase A, et al (2011). Urine TMPRSS2: ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Sci Transl Med* **3**, 94ra72.

[6] Mohler JL, Armstrong AJ, Bahnson RR, D'Amico AV, Davis BJ, Eastham JA, Enke CA, Farrington TA, Higano CS, and Horwitz EM, et al (2016). Prostate Cancer, Version 1.2016. *J Natl Compr Canc Netw* **14**, 19–30.

[7] Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, and Zhao S, et al (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* **47**, 199–208.

[8] Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, and Kominsky HD, et al (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**, 742–749.

[9] Sahu A, Singhal U, and Chinnaiyan AM (2015). Long noncoding RNAs in cancer: from function to translation. *Trends Cancer* **1**, 93–109.

[10] Cancer Genome Atlas Research Network. Electronic address scmo, Cancer Genome Atlas Research N (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025.

[11] Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, and Attard G, et al (2015). Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228.

[12] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

[13] Liao Y, Smyth GK, and Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.

[14] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, and Lander ES, et al (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550.

[15] Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekaran SM, Magi-Galluzzi C, Mehra R, Sahu A, and Siddiqui J, et al (2014). RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. *Lancet Oncol* **15**, 1469–1480.

[16] Mehra R, Shi Y, Udager AM, Prensner JR, Sahu A, Iyer MK, Siddiqui J, Cao X, Wei J, and Jiang H, et al (2014). A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia* **16**, 1121–1127.

[17] Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Erho N, and Ghadessi M, et al (2013). The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet* **45**, 1392–1398.

[18] Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, and Nishimasu H, et al (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588.

[19] Partridge J and Flaherty P (2009). An in vitro FluoroBlok tumor invasion assay; 2009. J Vis Exp.

[20] Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675.

[21] Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, and Chinnaiyan AM (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6.

[22] Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, Eastham JA, Gopalan A, Pienta KJ, and Shen MM, et al (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638–651.

[23] Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, and Otte AP, et al (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629.

[24] Asangani IA, Dommeti VL, Wang X, Malik R, Cieslik M, Yang R, Escara-Wilke J, Wilder-Romans K, Dhanireddy S, and Engelke C, et al (2014). Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* **510**, 278–282.

[25] Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, and Reva B, et al (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22.

[26] Mehra R, Udager AM, Ahearn TU, Cao X, Feng FY, Loda M, Petimar JS, Kantoff P, Mucci LA, and Chinnaiyan AM (2015). Overexpression of the Long Non-coding RNA SChLAP1 Independently Predicts Lethal Prostate Cancer. *Eur Urol.* pii: S0302-2838(15)01211-7. doi: 10.1016/j.eururo.2015.12.003. [Epub ahead of print].

[27] Consortium GT (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660.

[28] Consortium EPBirney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, and Dermitzakis ET, et al (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816.

[29] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504.

[30] Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, and Brenner JC, et al (2012). The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243.

[31] Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, and Bergerheim U, et al (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* **101**, 811–816.

[32] Vanaja DK, Cheville JC, Iturria SJ, and Young CY (2003). Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res* **63**, 3877–3882.

[33] Arredouani MS, Lu B, Bhasin M, Eljanne M, Yue W, Mosquera JM, Bubley GJ, Li V, Rubin MA, and Libermann TA, et al (2009). Identification of the transcription factor single-minded homologue 2 as a potential biomarker and immunotherapy target in prostate cancer. *Clin Cancer Res* **15**, 5794–5802.

[34] Best CJ, Gillespie JW, Yi Y, Chandramouli GV, Perlmutter MA, Gathright Y, Erickson HS, Georgevich L, Tangrea MA, and Duray PH, et al (2005). Molecular alterations in primary prostate cancer after androgen ablation therapy. *Clin Cancer Res* **11**, 6823–6834.

[35] Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, and Richie JP, et al (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**, 203–209.

[36] True L, Coleman I, Hawley S, Huang CY, Gifford D, Coleman R, Beer TM, Gelmann E, Datta M, and Mostaghel E, et al (2006). A molecular correlate to the Gleason grading system for prostate adenocarcinoma. *Proc Natl Acad Sci U S A* **103**, 10991–10996.

[37] Setlur SR, Mertz KD, Hoshida Y, Demichelis F, Lupien M, Perner S, Sboner A, Pawitan Y, Andren O, and Johnson LA, et al (2008). Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J Natl Cancer Inst* **100**, 815–825.

[38] Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, and Pienta KJ, et al (2007). Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* **39**, 41–51.

Lanbo Xiao, Ph.D.
Research Fellow
University of Michigan
Department of Pathology
1400 E Medical Center Drive
Room 5110, Comprehensive Cancer Center
Ann Arbor, MI 48105
Phone: (734) 223-7146
E-mail: lanbox@med.umich.edu

Education and Training

Education

09/2004-06/2008	BS, Biotechnology, Hunan Agriculture University, Hunan, China
09/2008-06/2014	PhD, Physiopathology, Central South University, Hunan, China

Postdoctoral Training

07/2014-present	Research Fellow, Michigan Center for Translational Pathology, University of Michigan

Research Interests

- Defined EZH2 as a driver of castration-resistant prostate cancer. By employing unbiased screen approaches, I identified EZH2 as a key epigenetic mediator of anti-androgen therapy resistance in prostate cancer cells. To address this, I employed a novel anti-sense oligonucleotide-based strategy to co-target AR and EZH2. This approach inhibited castration-resistant prostate cancer growth *in vitro* and *in vivo*, which provides fundamental data for the following clinical trial.
- Identified and characterized prostate cancer-associated long non-coding RNAs. By assembling and analyzing a large RNA-seq dataset of prostate cancer patient samples, I identified a novel lncRNA *PCAT14*, exhibiting both cancer and lineage specificity. *PCAT14* is highly expressed in low grade disease and loss of *PCAT14* predicts increased disease aggressiveness and recurrence. In addition, I am characterizing the role of the known prostate cancer-associated lncRNA *SchLAP1* in metastasis.
- Applied CRISPR/Cas9 genome editing techniques in cancer research. I established whole genome-wide/focused CRISPR screening pipelines in MCTP and have used these to investigate mechanisms of resistance to anti-cancer drugs. I generate CRISPR-based knockout cell lines for multiple ongoing projects in the MCTP.
- Developing novel urine-based prostate cancer diagnostic assays. I am applying a capture-based RNA sequencing pipeline to develop a novel biomarker panels for prostate cancer diagnostic screening. I will subsequently interrogate this biomarker panel by using a CRISPR/Cas-based highly sensitive RNA detection approach. This work will significantly improve the clinical diagnosis of prostate cancer.

Grants

Current Grants

1U01 CA214170: Discovery and qualification of transcriptomic biomarkers for the early detection of aggressive prostate cancer, NIH/NCI
Xiao, Lanbo, Co-Investigator
09/2016-08/2021. $2,867,413 ($372,001)

Past Grants

W81XWH-16-1-0195: Biological characterization and clinical utilization of metastatic prostate cancer-associated lincRNA SChLAP1, Department of Defense
Xiao, Lanbo, PI
06/2016-06/2018. $193,750 ($61,477)

Honors and Awards

International

2008	Excellent Undergraduate Student, Hunan Agriculture University
2013	Excellent Graduated Student, Central South University
2017	Team Science Award, Michigan Center for Translational Pathology, University of Michigan

Memberships in Professional Societies

2013-present	Associate, American Association for Cancer Research (AACR)

Editorial Positions, Boards, and Peer-Review Service

Editorial Boards

2016-present	Review Editor, Frontiers in Genetics and Molecular Biosciences

Journal Reviewer

2014-present	Cancer Letter
2014-present	PLOS ONE
2014-present	Medicine
2014-present	Molecular Carcinogenesis

Teaching

Undergraduate Student

07/2014-present	Celeb Cheng, BS, University of Michigan
09/2015-present	Esther Luo, BS, University of Michigan
02/2017-12/2017	Jonathan Yang, BS, University of Michigan

Bibliography

Peer-Review Journals and Publications

1. Li L, Guo L, Tao Y, Zhou S, Wang Z, Luo W, Hu D, Li Z, **Xiao L**, Tang M, Yi W, Tsao SW, Cao Y: Latent membrane protein 1 of Epstein-Barr virus regulates p53 phosphorylation through MAP kinases. *Cancer Lett*. 255(2):219-31, 2007.
2. Li L, Li Z, Zhou S, **Xiao L**, Guo L, Tao Y, Tang M, Shi Y, Li W, Yi W, Cao Y: Ubiquitination of MDM2 modulated by Epstein-Barr virus encoded latent membrane protein 1. *Virus Res*. 130(1-2):275-80, 2007.
3. Yang L, **Xiao L**, Ma X, Tang M, Weng X, Chen X, Sun L, Cao Y: Effect of DNAzymes targeting Akt1 on cell proliferation and apoptosis in nasopharyngeal carcinoma. *Cancer Biol Ther*. 8(4):366-71, 2009.
4. Yang L, Ma X, **Xiao L**, Tang M, Weng X, Sun L, Cao Y: Uniquely modified RNA oligonucleotides targeting STAT3 suppress melanoma growth both in vitro and in vivo. *Cancer Biol Ther*. 8(21):2065-72, 2009.

5. Luo XJ, Li LL, Deng QP, Yu XF, Yang LF, Luo FJ, **Xiao LB**, Chen XY, Ye M, Liu JK, Cao Y: Grifolin, a potent antitumour natural product upregulates death-associated protein kinase 1 DAPK1 via p53 in nasopharyngeal carcinoma cells. *Eur J Cancer*. 47(2):316-25, 2011.
6. Ma X, Yang L, **Xiao L**, Tang M, Liu L, Li Z, Deng M, Sun L, Cao Y: Down-regulation of EBV-LMP1 radio-sensitizes nasal pharyngeal carcinoma cells via NF-κB regulated ATM expression. *PLoS One*. 6(11):e24647, 2011.
7. Luo XJ, Li W, Yang LF, Yu XF, **Xiao LB**, Tang M, Dong X, Deng QP, Bode AM, Liu JK, Cao Y: DAPK1 mediates the G1 phase arrest in human nasopharyngeal carcinoma cells induced by grifolin, a potential antitumor natural product. *Eur J Pharmacol*. 670(2-3):427-34, 2011.
8. Guo L, Tang M, Yang L, **Xiao L**, Bode AM, Li L, Dong Z, Cao Y: Epstein-Barr virus oncoprotein LMP1 mediates survivin upregulation by p53 contributing to G1/S cell cycle progression in nasopharyngeal carcinoma. *Int J Mol Med*. 29(4):574-80, 2012.
9. Li L, Li W, **Xiao L**, Xu J, Chen X, Tang M, Dong Z, Tao Q, Cao Y: Viral oncoprotein LMP1 disrupts p53-induced cell cycle arrest and apoptosis through modulating K63-linked ubiquitination of p53. *Cell Cycle*. 11(12):2327-36, 2012.
10. Shi Y, Tao Y, Jiang Y, Xu Y, Yan B, Chen X, **Xiao L**, Cao Y: Nuclear epidermal growth factor receptor interacts with transcriptional intermediary factor 2 to activate cyclin D1 gene expression triggered by the oncoprotein latent membrane protein 1. *Carcinogenesis*. 33(8):1468-78, 2012.
11. Xu J, Deng X, Tang M, Li L, **Xiao L**, Yang L, Zhong J, Bode AM, Dong Z, Tao Y, Cao Y: Tyrosylprotein sulfotransferase-1 and tyrosine sulfation of chemokine receptor 4 are induced by Epstein-Barr virus encoded latent membrane protein 1 and associated with the metastatic potential of human nasopharyngeal carcinoma. *PLoS One*. 8(3):e56114, 2013.
12. Deng Q, Yu X, **Xiao L**, Hu Z, Luo X, Tao Y, Yang L, Liu X, Chen H, Ding Z, Feng T, Tang Y, Weng X, Gao J, Yi W, Bode AM, Dong Z, Liu J, Cao Y: Neoalbaconol induces energy depletion and multiple cell death in cancer cells by targeting PDK1-PI3-K/Akt signaling pathway. *Cell Death Dis*. 4:e804, 2013.
13. Ma X, Xu Z, Yang L, **Xiao L**, Tang M, Lu J, Xu S, Tang Y, Wen X, Deng X, Sun L, Cao Y: EBV-LMP1-targeted DNAzyme induces DNA damage and causes cell cycle arrest in LMP1-positive nasopharyngeal carcinoma cells. *Int J Oncol*. 43(5):1541-8, 2013.
14. Li W, Peng C, Lee MH, Lim D, Zhu F, Fu Y, Yang G, Sheng Y, **Xiao L**, Dong X, Ma W, Bode AM, Cao Y, Dong Z: TRAF4 is a critical molecule for Akt activation in lung cancer. *Cancer Res*. 73(23):6938-50, 2013.
15. Hu ZY, **Xiao L**, Bode AM, Dong Z, Cao Y: Glycolytic genes in cancer cells are more than glucose metabolic regulators. *J Mol Med (Berl).* 92(8):837-45, 2014 Aug.
16. Yi L, Song C, Hu Z, Yang L, **Xiao L**, Yi B, Jiang W, Cao Y, Sun L: A metabolic discrimination model for nasopharyngeal carcinoma and its potential role in the therapeutic evaluation of radiotherapy. *Metabolomics* 10:697–708, 2014.
17. **Xiao L**, Hu ZY, Dong X, Tan Z, Li W, Tang M, Chen L, Yang L, Tao Y, Jiang Y, Li J, Yi B, Li B, Fan S, You S, Deng X, Hu F, Feng L, Bode AM, Dong Z, Sun LQ, Cao Y: Targeting Epstein-Barr virus oncoprotein LMP1-mediated glycolysis sensitizes nasopharyngeal carcinoma to radiation therapy. *Oncogene*. 33(37):4568-78, 2014.
18. Yu X, Deng Q, Li W, **Xiao L**, Luo X, Liu X, Yang L, Peng S, Ding Z, Feng T, Zhou J, Fan J, Bode AM, Dong Z, Liu J, Cao Y: Neoalbaconol induces cell death through necroptosis by regulating RIPK-dependent autocrine TNFα and ROS production. *Oncotarget*. 6(4):1995-2008, 2015.
19. Yang L, Liu L, Xu Z, Liao W, Feng D, Dong X, Xu S, **Xiao L**, Lu J, Luo X, Tang M, Bode AM, Dong Z, Sun L, Cao Y: EBV-LMP1 targeted DNAzyme enhances radiosensitivity by inhibiting tumor angiogenesis via the JNKs/HIF-1 pathway in nasopharyngeal carcinoma. *Oncotarget*. 6(8):5804-17, 2015.
20. Zhao L, Tang M, Hu Z, Yan B, Pi W, Li Z, Zhang J, Zhang L, Jiang W, Li G, Qiu Y, Hu F, Liu F, Lu J, Chen X, **Xiao L**, Xu Z, Tao Y, Yang L, Bode AM, Dong Z, Zhou J, Fan J, Sun L, Cao Y: miR-504 mediated down-regulation of nuclear respiratory factor 1 leads to radio-resistance in nasopharyngeal carcinoma. *Oncotarget*. 6(18):15995-6018, 2015.
21. Luo X, Yang L, **Xiao L**, Xia X, Dong X, Zhong J, Liu Y, Li N, Chen L, Li H, Li W, Liu W, Yu X, Chen H, Tang M, Weng X, Yi W, Bode A, Dong Z, Liu J, Cao Y: Grifolin directly targets ERK1/2 to epigenetically suppress cancer cell metastasis. *Oncotarget*. 6(40):42704-16, 2015.

22. Tan Z, Luo X, **Xiao L**, Tang M, Bode AM, Dong Z, Cao Y: The role of PGC1α in cancer metabolism and its therapeutic implications. *Mol Cancer Ther*. 15(5):774-82, 2016 May.
23. Shukla S*, Zhang X*, Niknafs YS*, **Xiao L***, Mehra R, Cieślik M, Ross A, Schaeffer E, Malik B, Guo S, Freier SM, Bui HH, Siddiqui J, Jing X, Cao X, Dhanasekaran SM, Feng FY, Chinnaiyan AM, Malik R **(*equal contribution)**: Identification and validation of PCAT14 as prognostic biomarker in prostate cancer. *neoplasia*. 18(8):489-99, 2016.
24. Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, Iyer MK, Pitchiaya S, Malik R, Hosono Y, Prensner JR, Poliakov A, Singhal U, **Xiao L**, Kregel S, Siebenaler RF, Zhao SG, Uhl M, Gawronski A, Hayes DF, Pierce LJ, Cao X, Collins C, Backofen R, Sahinalp CS, Rae JM, Chinnaiyan AM, Feng FY: The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. Nat Commun. 7:12791, 2016.
25. Zhang Y, Pitchiaya S, Cieślik M, Niknafs YS, Tien JC, Hosono Y, Iyer MK, Yazdani S, Subramaniam S, Shukla SK, Jiang X, Wang L, Liu TY, Uhl M, Gawronski AR, Qiao Y, **Xiao L**, Dhanasekaran SM, Juckette KM, Kunju LP, Cao X, Patel U, Batish M, Shukla GC, Paulsen MT, Ljungman M, Jiang H, Mehra R, Backofen R, Sahinalp CS, Freier SM, Watt AT, Guo S, Wei JT, Feng FY, Malik R, Chinnaiyan AM: Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. *Nat Genet*. 50(6):814-824, 2018.
26. Tan Z, **Xiao L**, Tang M, Bai F, Li J, Li L, Shi F, Li N, Li Y, Du Q, Lu J, Weng X, Yi W, Zhang H, Fan J, Zhou J, Gao Q, Onuchic JN, Bode AM, Luo X, Cao Y: Targeting CPT1A-mediated fatty acid oxidation sensitizes nasopharyngeal carcinoma to radiation therapy: *Theranostics*. 8(9): 2329–2347, 2018.
27. **Xiao L**, Tien JC, Vo J, Tan M, Parolia A, Zhang Y, Wang L, Qiao Y, Shukla S, Wang X, Zheng H, Su F, Jing X, Luo E, Delekta A, Juckette KM, Xu A, Cao X, Alva AS, Kim Y, Macleod AR, Chinnaiyan AM **(*equal contribution)**: Epigenetic reprogramming with antisense oligonucleotides enhances effectiveness of androgen receptor inhibition in castration-resistant prostate cancer. *Cancer Res*. 2018.

Abstracts

1. **Xiao L***, Shukla S**, Zhang X**, Niknafs Y**, Malik R, Chinnaiyan A **(*equal contribution)**: Identification and validation of PCAT14 as prognostic biomarker in prostate cancer, AACR Annual Meeting, Washington, DC, 2017.
2. **Xiao L***, Tien JC**, Vo J, Tan M, Parolia A, Zhang Y, Wang L, Qiao Y, Shukla S, Wang X, Zheng H, Su F, Jing X, Luo E, Delekta A, Juckette KM, Xu A, Cao X, Alva A, Kim Y, MacLeod AR, Chinnaiyan AM **(*equal contribution)**: Epigenetic reprogramming with antisense oligonucleotides enhances effectiveness of androgen receptor inhibition in castration-resistant prostate cancer, Prostate SPOREs, Fort Lauderdale, Florida, 2018.
3. **Xiao L***, Tien JC**, Vo J, Tan M, Parolia A, Zhang Y, Wang L, Qiao Y, Shukla S, Wang X, Zheng H, Su F, Jing X, Luo E, Delekta A, Juckette KM, Xu A, Cao X, Alva A, Kim Y, MacLeod AR, Chinnaiyan AM **(*equal contribution)**: Epigenetic reprogramming with antisense oligonucleotides enhances effectiveness of androgen receptor inhibition in castration-resistant prostate cancer. AACR Annual Meeting, Chicago, IL, 2018.
4. Parolia A*, **Xiao L***, Vo JN, Cieslik M, Cao X, Chinnaiyan AM **(*equal contribution)**: Functional CRISPR screen towards identifying novel epigenetic co-factors of oncogenic AR-activity. Prostate SPOREs, Fort Lauderdale, Florida, 2018.
5. Parolia A*, **Xiao L***, Vo JN, Cieslik M, Cao X, Chinnaiyan AM **(*equal contribution)**: Functional CRISPR screen towards identifying novel epigenetic co-factors of oncogenic AR-activity, AACR Annual Meeting, Chicago, IL, 2018.