Capped nascent RNA sequencing reveals novel therapy-responsive enhancers in prostate cancer

Kellie A. Cotter[1,#], Sagar R. Shah[2,3,#], Mauricio I. Paramo[2,3,#], Shaoke Lou[4,#], Li Yao[3,5], Philip D. Rubin[1], You Chen[2,3], Mark Gerstein[4,*], Mark A. Rubin[1,6*], Haiyuan Yu[3,5,*]

[1]Department of Biomedical Research, University of Bern, Bern, 3008, Switzerland
[2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[3]Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14853, USA
[4]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA
[5]Department of Computational Biology, Cornell University, Ithaca, NY 14853, USA
[6]Bern Center for Precision Medicine, University of Bern and Inselspital, Bern, 3008, Switzerland

[#]Contributed equally
*Correspondence to: mark.gerstein@yale.edu (M.G), mark.rubin@dbmr.unibe.ch (M.A.R) and haiyuan.yu@cornell.edu (H.Y)

Abstract

Mounting evidence suggests that enhancer RNA (eRNA) transcription start sites (TSSs) provide higher sensitivity and specificity for enhancer identification than histone modifications and chromatin accessibility. The extent to which changes in eRNA transcription correspond to changes in enhancer activity, however, remains unclear. Here, we used precision run-on and capped RNA sequencing (PRO-cap) to assess changes in enhancer activity in response to treatment with the androgen receptor signaling inhibitor, enzalutamide (ENZ). We identified 6,189 high-confidence candidate enhancers in the human prostate cancer cell line, LNCaP; 853 of which demonstrated significant changes in activity in response to drug treatment. Notably, we found that 67% and 54% of drug-responsive enhancers did not show similar changes in activity in previous studies that utilized ChIP-seq and ATAC-seq, respectively. Strikingly, 79% of regions with increased eRNA transcription showed no other biochemical alterations, implying that PRO-cap can capture a set of precise changes in enhancer activity that classical approaches lack the sensitivity to detect. We performed *in vivo* functional validations of candidate enhancers and found that CRISPRi targeting of PRO-cap-specific drug-responsive enhancers impaired ENZ regulation of downstream target genes, suggesting that changes in eRNA TSSs mark true biological changes in enhancer activity with high sensitivity. Our study highlights the utility of using PRO-cap as a complementary approach to canonical biochemical methods for detecting precise changes in enhancer activity and, in particular, for better understanding disease progression and responses to treatment.

Main

First-line treatment for advanced metastatic prostate cancer (PCa) generally involves androgen deprivation therapy (ADT) to reduce the activity of androgen receptor (AR)[1]. Of particular clinical concern is metastases of castration-resistant PCa (CRPC), where the disease has developed resistance to both first-line and second-generation AR signaling inhibitors (ARSi, e.g., enzalutamide)[2]. CRPC is often driven by alterations of *AR* such as mutations or gene amplifications which allow AR activity even under low circulating androgen levels[3,4]. Despite recent progress in developing new treatments for metastatic CRPC, median survival at this advanced disease stage is approximately three years[5]. Thus, determining both the mechanisms behind ADT resistance and the distinct signaling pathways activated in CRPC are essential to improving existing therapies and finding new potential drug targets.

At the molecular level, PCa is generally characterized by relatively few recurrent genomic alterations[6-8], but a higher rate of large-scale structural genomic rearrangements[9,10]. The most frequent of these structural rearrangements occur in approximately 50% of PCa patients and result in the fusion of the coding region of the oncogene *ETS-related gene* (*ERG)* and other *ETS* family members downstream of the AR regulated *transmembrane protease serine 2* (*TMPRSS2*) promoter[11,12]. Significantly, genomic alterations are generally enriched in metastatic PCa[8], however, only recently has there been increased efforts to understand the complexity of the entirety of the PCa genome via whole-genome sequencing (WGS) as opposed to targeted-sequencing approaches focused mainly on coding regions[13-15]. These studies revealed several examples of alterations in gene regulatory regions. For example, in primary PCa binding sites for several transcription factors are enriched in mutations, with the most heavily mutated being AR binding sites (ARBS)[16]. Furthermore, in metastatic PCa, upwards of 80% of samples were found to have a duplication of a region upstream of *AR,* which was then shown to be a previously unidentified enhancer[14,15,17]. This enhancer duplication was shown to increase the expression of *AR* and in doing so decrease sensitivity to ARSi.

Despite the exhaustive amount of sequencing information captured by WGS, our interpretation and understanding of the data is limited due to the incomplete annotation of the non-coding genome, including transcriptional regulatory elements such as enhancers. In general, enhancer regions are currently defined by biochemical features such as chromatin accessibility (DNase I hypersensitivity or transposase accessibility) and histone modification marks (H3K27ac and H3K4me1) along with transcription factor binding profiles as determined by ChIP-seq[18]. Large-scale reporter assays (e.g., STARR-seq or MPRA) have also been used to assay the enhancer potential of candidate DNA regions[19,20]. However, these reporter assays have consistently shown that less than 50% of regions with these biochemical annotations exert enhancer activity[21,22]. Likewise, similar techniques have been used to identify thousands of potential enhancer regions in PCa by combining ChIP-seq and whole genome STARR-seq[23-26]. In addition, chromosome capture has been used to elucidate chromatin interactions genome-wide and those loci specifically associated with AR and RNA Polymerase II using ChIA-PET[27-30]. Interestingly, a recent study demonstrated that the majority of AR binding sites are not active enhancers[31]. Moreover, despite increased recruitment of AR to these regions, dihydrotestosterone (DHT) stimulation did not increase enhancer activity in over half of the AR-bound active enhancers. Perhaps most intriguing, however, CRISPR interference of some “inactive” enhancers altered the expression of the enhancer-regulated gene, at times at a similar level of interference to nearby “active” enhancers. Overall, this highlights the limitations of the datasets produced thus far and the need for alternative methods for enhancer identification given that epigenomic-mark-based approaches identify enhancers with a high false-positive rate, while reporter assays cannot reproduce much of the biological complexity of large enhancers regions in their native genomic context, in particular with regards to multiple enhancers acting on a single gene.

More recently, widespread RNA polymerase II-mediated bidirectional transcription has been observed in enhancer regions, which produce biochemically unstable transcripts known as

enhancer RNAs (eRNAs)[32-34]. Although the functional significance of eRNAs remains unclear, evidence suggests that enhancer transcription corresponds with activation[35,36], with close to 50% of short capped nascent RNAs that map to previously unannotated TSSs overlapping with episomal reporter-validated enhancers[37]. Moreover, it has been found that ~95% of putative active enhancers found within accessible chromatin drive local transcription and do so using factors and mechanisms overwhelmingly similar to those of promoters[33]. Recently, we performed systematic interrogation of enhancer elements and showed that active enhancer units are precisely marked by divergent eRNA TSSs genome-wide. Moreover, although eRNA transcription is closely correlated with histone marks, we saw that histone marks offer lower resolution and specificity for enhancer activation than transcription initiation[38]. Thus, these data support a model whereby transcription is required for distal enhancer function, challenging classical methods that rely on chromatin accessibility and histone modifications to identify active enhancers.

On average, enhancers transcribe at 5% the level of promoters[33,37]. Thus, due to their low abundance and instability, eRNA detection requires highly sensitive alternative methods to standard RNA sequencing approaches. Recently, we performed systematic comparisons of genome-wide RNA sequencing assays suitable for the identification of active enhancers and found that the nuclear run-on followed by cap-selection assays (namely, Global/Precision Run-On and capped RNA sequencing GRO/PRO-cap) provide the highest sensitivity and specificity for eRNA detection and active enhancer identification across the whole genome[39]. Importantly, PRO-cap libraries undergo a series of cap state selection reactions to modify the 5′ ends of transcripts and allow for the accurate identification of transcription initiation sites (Figure 1A). As a result, PRO-cap is highly sensitive for capturing eRNA transcription and therefore, a powerful tool for enhancer identification.

Here, we utilized changes in eRNA expression to assess changes in enhancer activity in response to short-term treatment with the ARSi enzalutamide (ENZ). We applied PRO-cap to enrich and sequence only nascent RNAs associated with engaged RNA polymerase and to

identify divergent transcription start sites marking active enhancers at base pair-resolution[38,40]. We identified over 6,000 candidate enhancers in LNCaP cells; 853 of which demonstrated significant changes in enhancer activity in response to ENZ treatment. Importantly, these results identified a large percentage of therapy-responsive enhancers, which were not previously shown to be responsive in studies utilizing other biochemical marks (H3K27ac-ChIP-seq/ATAC-seq) or reporter assays (i.e., STARR-seq). Our study highlights the utility of using eRNA transcription, in particular PRO-cap, as a generalizable and complementary approach to canonical biochemical methods for detecting precise changes in enhancer activity, specifically for applications in disease prognosis, progression, and treatment.

Results

Prioritization of active enhancer regions in LNCaP cells using PRO-cap

To measure enhancer activity in their native genomic and cellular context, we utilized PRO-cap to preferentially sequence nascent transcriptional start sites including those generating eRNAs (Figure 1A). Overall, this analysis in LNCaP cells identified 91,705 statistically significant peaks (n = 2 replicates), 68.4% of which were in gene distal regions. 6,189 of these gene distal regions contained divergent significant peaks on both the plus and minus strands and were extended by 200 bp in both directions and called as high confidence enhancer regions (Supp. Data 1). As expected, the majority of these regions contained canonical biochemical marks delineating enhancers including open chromatin (DNase I hypersensitivity, ATAC-seq), H3K27 acetylation, and H3K4 monomethylation, and are bound by prostate-enriched transcription factors (AR and FOXA1) (Figure 1B). However, 11% (n = 674) of these regions did not exhibit those classical biochemical features (at least two of the three: DNase I HS, H3K27ac, or H3K4me1), demonstrating that PRO-cap identifies a novel set of enhancer regions that would otherwise be missed.

Using publicly available H3K27ac HiChIP data, we estimated the number of chromatin connections of each candidate enhancer with other genomic regions[41]. We found that the regions

delineated by PRO-cap have a significantly greater number of loops ($P < 0.0001$, Kolmogorov-Smirnov test) than H3K27ac or DNase I peaks (Figure 1C). Next, we inquired whether these enhancers identified in LNCaP cells were expressed in patient-derived PCa samples. To that end, we analyzed published ChIP-seq data from clinical specimens of non-neoplastic prostate, primary PCa, and metastatic PCa[42]. As seen in Figure 1D, these tissues recapitulate H3K27ac and the binding pattern of prostate-enriched transcription factors (AR, FOXA1, and HOXB13) in enhancers seen in LNCaP cells. Furthermore, given that LNCaP cells are derived from an androgen-responsive metastatic lymph node lesion[43], the enhancer expression profiles are indeed strongest in metastatic PCa tissues. Overall, these results demonstrate that, unlike traditional epigenomic-based methods, PRO-cap can uncover previously unknown enhancer loci with hallmarks of PCa-relevant TF-binding patterns observed in primary and metastatic lesions.

Identification of changes in enhancer activity in response to therapy

Given our confidence in detecting enhancers by PRO-cap, we investigated whether this method could be utilized to assess changes in enhancer activity in response to treatment. To do so, we treated LNCaP cells with the ARSi enzalutamide (ENZ) for 4 or 24 hours. Using the same analysis as described above, we identified 6769 and 8513 high confidence enhancers using PRO-cap in the 4- and 24-hour time point conditions, respectively (Supp. Data 2 and 3). Approximately 4,479 of these enhancers were identified in both ENZ timepoints and 2,922 were identified irrespective of treatment (Supp. Figure 1A).

In addition to detecting enhancers that are expressed at only a single treatment time point, we were interested in identifying those which were active at more than one time point, but with differential levels of expression. Thus, we compiled the total list of enhancers found in any of the two treatment timepoints and vehicles and assessed their differential expression upon treatment (versus vehicle) on at least one strand. This analysis identified 853 enhancers that were significantly activated or repressed after 24 hours of ENZ treatment (Figure 2A, Supp. Figure

1B/C, Supp. Data 4/5). While this analysis only identified 44 enhancers with significantly altered activity (FDR < 0.05) upon 4 hours of ENZ treatment (Supp. Figure 1D), these regions do demonstrate a similar trend in activity change albeit at a lower level (Figure 2A). Furthermore, 73 of these regions were significantly activated or repressed on both the plus and minus strands.

To predict putative transcription factors (TF) regulating these enhancer regions, we utilized GIGGLE[44] analysis (Figure 2B) to probe the CISTROME[45] ChIP-seq database containing published datasets with overlapping genomic regions. Of the TFs predicted to bind to the ENZ repressed elements, AR was the top candidate with the highest GIGGLE score. Interestingly, other significant TF predicted to bind the ENZ repressed enhancers included glucocorticoid receptor (*NR3C1*), HNF4G (a TF which drives ARSi resistance[46]), and the co-activator EP300, the SWI/SNF family chromatin remodeler SMARCA4, and the pioneering factor FOXA1. In addition to AR, FOXA1, and NR3C1, other TF predicted to bind the ENZ activated enhancers included the histone methyltransferase ASHL2, the AR co-regulators GRHL2 and HOXB13, the AR-regulated TF NKX3-1, the pioneering factor GATA2, ARID1A another SWI/SNF family member, and the histone deacetylase HDAC3. Motif analysis further supported these predictions. Motifs for FOXA1, AR, and MAFA were significantly enriched in ENZ-repressed enhancers, while FOXA1, HOXB13, and HOXA13 motifs were significantly enriched in the ENZ-activated enhancers (Figure 2C).

We further mined published ChIP-seq data from LNCaP cells to determine how the binding of the identified enhancer-associated TF is changing in response to ENZ treatment or androgen stimulation with dihydrotestosterone (DHT) or the synthetic androgen, R1881) (Figure 2D). Both ENZ-activated and -repressed enhancers lose AR binding in response to ENZ (AR inhibition) or the absence of hormones (charcoal-stripped, CS FCS), and binding is regained with androgen stimulation. Strikingly, FOXA1 binding at the ENZ repressed enhancers is lost with ENZ or CS FCS treatments and regained with androgen stimulation, while the binding is unaffected at ENZ activated enhancers. Similar patterns were seen with NKX3.1 and ASH2L with dynamic changes

in binding seen at the ENZ repressed enhancers with trivial changes at the activated ones. In contrast, both GRHL2 and HOXB13 show minimal binding at the ENZ repressed enhancers as compared to the activated enhancers irrespective of treatment. Similar patterns were seen between the activated and repressed groups with ChIP-seq analysis of metastatic patient samples[42] (Supp. Figure 1E).

Using a previously described method[47], we predicted which genes are regulated by each ENZ-responsive enhancer. We generated predictions for 690 of the 853 regions with an average of 6.3 genes per candidate enhancer (Supp. Figure 1F, Supp. Data 6/7). GO analysis of these gene lists demonstrated that the majority of genes are related to steroid hormone signaling (in particular androgen response) or the cell cycle (Figure 2E).

Altogether, these findings demonstrate that PRO-cap can detect a large set of precise changes in enhancer activity that other approaches lack the sensitivity to capture.

Genomic Alterations in Candidate Enhancers

We next queried if PRO-cap-detected enhancer regions could help prioritize somatic genomic variation in non-coding regions. To that end, using 286 PCa whole genomes available through ICGC, we searched for single nucleotide variants (SNVs) in the identified 853 ENZ-regulated enhancers. In total 137 variants were discovered in these regions; 20 of which were recurrent in more than one patient (Figure 3A). Interestingly, eight of these enhancers also had more than one SNV within the enhancer region. An example of this is shown in Figure 3B in which two patients have a recurrent SNV which disrupts an ESRRA motif within the enhancer, and two different patients have a recurrent SNV which creates an HNF4 motif within the enhancer.

Next, we investigated whether the PRO-cap-identified enhancers may harbor potential PCa-associated germline single nucleotide polymorphisms (SNPs) [48]. Of the 269 known PCa risk variants, none overlap with our 853 ENZ-regulated enhancers. However, 10 SNPs did overlap with our larger list of 6,189 total high confidence PCa enhancers ($P < 0.001$). A representative

example of a significantly enriched SNP at a PCa enhancer is shown in Figure 3C. It is noteworthy that none of these SNPs have been previously identified as residing in gene regulatory regions. Thus, these results highlight the power of PRO-cap in identifying and delimiting the non-coding regulatory genome to prioritize enhancer-associated mutations.

PRO-cap versus *in vitro* Episomal Measurements of Enhancer Activity

In order to demonstrate the utility of the PRO-cap approach, we sought to determine how it compares with other methods at measuring enhancer activity changes. We first surveyed published H3K27ac ChIP-seq and ATAC-seq data from LNCaP cells treated with ENZ. Many of the regions with altered enhancer activity as measured by PRO-cap could also be detected by changes in the ChIP- or ATAC-seq data (Figure 4A). Surprisingly, however, 67% and 54% of the regions did not show a similar change in the ChIP- or ATAC-seq data, respectively (highlighted in navy and burgundy). Particularly striking was that 78-79% of the regions with an increase in eRNA transcription with ENZ showed no biochemical alterations.

We further sought to compare our PRO-cap results to traditional luciferase reporter assays. We chose 7 candidate regions with the most significant changes in eRNA transcription in response to ENZ (Figure 4B). We cloned each of these regions downstream of luciferase in a reporter plasmid to assess enhancer activity. As positive controls for these experiments, we cloned the enhancer for the known AR-responsive gene *KLK3* (PSA) in the reporter plasmid and separately, cloned an AR-responsive element acting as a promoter upstream of luciferase (pGL4.36). We transfected LNCaP cells with these reporters followed by 24 hours of ENZ or DHT treatment. As expected, we observed repression and activation of luciferase signal with our two controls. Strikingly, a significant increase luciferase signal upon ENZ treatment was seen in 6 of the 7 candidate enhancers tested (Figure 4C). These candidate enhancers included four regions which were repressed by ENZ treatment in our PRO-cap data. To confirm that these eRNAs are truly AR-responsive, we quantified changes in canonical AR-responsive genes via qPCR in these

transfected cells treated with ENZ. Indeed, we saw expression changes similar to those seen in canonical AR-responsive genes (*KLK3, NKX3.1, FKBP5*) (Figure 4D).

A similar discordance was seen in the results from a STARR-seq analysis of AR binding sites[31]. Of the 139 ENZ-responsive enhancers identified by PRO-cap that were also assessed via STARR-seq, the majority (118 of 139) were deemed to either be lacking in enhancer activity or had constitutive activity independent of AR stimulation (Figure 4E).

We hypothesize that these discrepancies may be due properties of enhancer function that are lost when placed out of their endogenous genomic context.

Functional Validation of PRO-cap-Detected ENZ-Responsive Candidate Enhancers *in vivo*

In the aforementioned STARR-seq study, although the majority of AR binding sites they assessed were categorized as inactive enhancers, CRISPR interference (CRISPRi) of these enhancers altered the expression of the downstream genes[31]. These results suggest a complexity in measuring enhancer activity that is not easily captured by reporter assays alone. We therefore sought to validate the enhancer activity of our candidates identified by PRO-cap using a similar *in vivo* approach.

To that end, we genetically targeted two of the previously tested candidates (#1 and #3) that were repressed by ENZ (Figure 4B). Enhancer candidate #1 is located within in intronic region of *CACNG4*, while enhancer candidate #3 is found within the intron of *KCNMA1*, (Supp. Figures 2A and 2B). Both of these genes are known to be repressed by ENZ treatment. We designed 3 sgRNAs each against these regions centered around the TSS of the eRNAs. We then transfected dCas9-KRAB-stably expressing LNCaP cells with these sgRNAs. CRISPRi-mediated repression of the two candidate enhancers reduced the expression of both the eRNA and the predicted target genes (Figure 5A, B).

We next sought to determine whether targeting of the TSS altered the ability of the candidate enhancer to regulate gene expression in response to ENZ treatment. To do so we again

transfected dCas9-KRAB-expressing LNCaP cells with the chosen sgRNAs followed by 24-hour treatment with ENZ. Again, for both genes we demonstrated significant downregulation with ENZ (*KCNMA1, P* < 0.0001; *CACNG4, P* < 0.0001), with the sgRNA (*KCNMA1, P* < 0.0001; *CACNG4, P* < 0.0001), and a significant reduction in the ENZ downregulation in combination with the sgRNA (*KCNMA1, P* = n.s., 1.42 vs 2.05-fold change; *CACNG4, P* = 0.0007, 2.35 vs 4.48-fold change) (Figure 5C/D).

Lastly, we functionally validated the ability of PRO-cap to identify changes in enhancer activity that are not detected by ATAC-seq or H3K27ac ChIP-seq assays (Figure 4A). Given that the two above candidates both show concomitant changes in both of those biochemical marks (Supp. Figure 2A/3A), we pursued enhancer candidates which: (1) show significant changes in PRO-cap upon ENZ treatment, (2) show no changes in ATAC-seq or H3K27ac ChIP-seq after ENZ treatment, and (3) are predicted to regulate known ENZ-responsive genes.

The first enhancer candidate (#8) selected resides within an intronic region of *DENND1B* (Supp. Figure 4A), a gene which has been shown to be upregulated with ENZ (Supp. Figure 4B). While *DENND1B* was not significantly upregulated upon ENZ treatment in control sgRNA-expressing cells, a dramatic increase in its expression was seen in cells expressing sgRNAs against the candidate target #8 with ENZ treatment (*P* < 0.0001, 2.06 vs 0.946-fold change) (Supp. Figure 4C). The next enhancer candidate selected (#9) is ~80 kb upstream of *RASD1* (Figure 5E, Supp. Figure 5A), which is a known ENZ-repressed gene (Supp. Figure 5B). *RASD1* was significantly downregulated with ENZ treatment in control sgRNA-expressing cells (*P* = 0.0084). Likewise, a significant decrease in RASD1 expression was observed in cells transfected with the candidate #9 targeting sgRNAs than control sgRNA-expressing cells (*P* = 0.0214). and there was a significant reduction of ENZ downregulation in combination with the sgRNA (*P* = n.s., 1.15 vs 1.44-fold change) (Figure 5F).

Altogether, these *in vivo* functional results confirm the utility of our PRO-cap assay as a highly-sensitive approach to identify putative enhancers and detect changes in their activities

genome-wide.

Discussion

This study highlights the generalizable utility of using eRNA transcription patterns obtained from PRO-cap to detect precise changes in enhancer activity, which has broad applications in human genetics spanning development and disease. Despite previous efforts in generating chromatin accessibility and histone modification landscapes to indirectly map transcriptional regulatory networks, our knowledge of the key regulatory mechanisms that orchestrate the complexity of cellular differentiation, human development, and disease pathogenesis is still limited by our incomplete understanding and characterization of the non-coding genome, in particular with the annotation of cell-state specific transcriptional regulatory elements. Thus, the ability to delineate more precise maps of enhancer activity dynamics will facilitate the systematic examination of the transcriptional programs of developing cells across cell transition states and different differentiation lineages, including disease progression. With a significant fraction of disease-associated risk variants harbored within non-coding regions of the genome, detailed characterization of these regulatory networks that coordinate cell- and tissue-type specificity could provide insights into the molecular mechanisms that underlie dysregulation in numerous disorders and assist in the mapping of variants functional only at specific cellular states. Similarly, precise mapping of enhancer dynamics genome-wide in clinical specimens by PRO-cap can also help better understand disease mechanisms and heterogeneity across patients as well as response and resistance to treatment.

While most of the PRO-cap enhancers identified in LNCaP displayed canonical biochemical signatures, 11% were not captured by both DNase-seq and ChIP-seq. This disparity was amplified upon ENZ treatment, where 67% and 54% of enhancers did not show similar alterations in H3K27ac ChIP-seq or ATAC-seq, respectively, and 79% of enhancers with increased eRNA transcription showing no other biochemical alterations. In addition to recapitulating previous observations that eRNA provides distinct profiles of active enhancers,

these data suggest that PRO-cap can also capture a large set of precise changes in enhancer activity that other approaches lack the sensitivity to detect.

H3K27ac HiChIP analysis found that loops were significantly more likely to overlap with PRO-cap than with H3K27ac or DHS, evidence that the regions that PRO-cap is detecting are transcriptionally active. Transcription factor (TF) predictions found AR, among other prostate-enriched TFs, to be likely regulators of the ENZ-responsive enhancer set, with these predictions further confirmed by motif analysis finding that AR and prostate-related TF binding motifs to be significantly overrepresented. We mined published ChIP-seq data from LNCaP cells to determine how binding of these factors changed in response to ENZ treatment or androgen stimulation and saw that ENZ activated and repressed regions lost AR binding in response to AR inhibition or absence of hormones and with binding regained upon androgen stimulation. These data offer support that the regions detected by PRO-cap are transcriptionally active and are of relevance to AR signaling.

While changes in enhancer activity with androgen stimulation or inhibition measured by luciferase reported assays did not agree with PRO-cap, we confirmed that these eRNAs were indeed AR-responsive like those seen in canonical AR-responsive genes via qPCR. A similar discordance was observed in a previous STARR-seq study of AR binding sites: although most sites were categorized as inactive by STARR-seq, CRISPRi perturbation of these sites altered the expression of downstream genes[31]. We hypothesize that these discrepancies may be due to a level of enhancer-promoter compatibility not considered in these episomal assays, in addition to other properties of enhancer function that may be lost when placed out of their endogenous genomic context.

Nevertheless, in agreement with our previous studies[38], CRISPRi-mediated repression of PRO-cap candidate enhancers significantly reduced the expression of the eRNA itself and their respective downstream target genes, validating PRO-cap's capability to detect active enhancers genome-wide. More importantly, we show for the first time that CRISPRi targeting of PRO-cap-

detected ENZ-responsive enhancers also impaired the ENZ-regulation of their respective downstream target genes. Thus, we conclude that the PRO-cap-specific drug-responsive distal TSSs we identified in this study mark true biological changes in enhancer activity.

Gene ontology analysis of predicted genes regulated by 690 of the ENZ-responsive enhancers demonstrated that most were related to steroid hormone signaling, particularly androgen response, as well as the cell cycle. Analysis of 286 PCa whole genomes also identified 137 SNVs in the ENZ-regulated enhancers identified via PRO-cap, 20 of which were recurrent in more than one patient and with some that overlapped with ENZ downregulated enhancers breaking TF binding motifs. While analysis of PRO-cap enhancers and germline SNPs associated with PCa risk found no overlap with our 853 ENZ-regulated enhancers, 10 SNPs did overlap with our larger list of 6189 total high confidence enhancers, with none of these SNPs previously identified as being in gene regulatory regions.

Analyzing published ChIP-seq data from patient samples of normal prostate, primary PCa, and metastatic PCa revealed that LNCaP exhibited H3K27ac and binding of prostate-enriched TFs, signals observed the strongest in metastatic patient samples. We conclude that LNCaP closely recapitulates the transcriptional regulatory landscape of metastatic PCa and is thus a suitable model to interrogate metastatic PCa androgen-responsive enhancers. Nevertheless, we acknowledge the limitations of utilizing a cell line as a model for such complex disease states, and thus, the development of PRO-cap to interrogate patient-derived tissue specimens could yield enhancer landscapes that are of much greater clinical relevance. Hence, our study underscores the clinical value of identifying and delineating aberrant distal regulatory elements in cancer to identify potential therapeutic vulnerabilities.

Materials and Methods

Cell Lines

LNCaP cells (male, ATCC, RRID: CVCL_1379) were maintained in RPMI medium (Gibco, A1049101), supplemented with 10% FBS (Gibco, 10270106), and 1% penicillin-streptomycin

(Gibco, 11548876) on poly-L-lysine coated plates. All cell lines were grown at 37 °C with 5% CO_2. All cell lines were authenticated by STR analysis and regularly tested for mycoplasma.

ChIP-seq Data Analysis

Figure 1B- Previously aligned and normalized bigwig files were downloaded from the Cistrome Data Browser allowing for consistent and standardized analysis of data from multiple studies. For other plots previously aligned and normalized bigwig files were downloaded and analyzed in their published format from GEO. Heatmaps and summary plots were generated using the deepTools suite[49].

PRO-cap

LNCaP cells were treated with 10 µM enzalutamide or DMSO for 4 or 24 hrs. For PRO-cap, approximately 10 to 30 million cells were processed per sample. Library preparations for two biological replicates each consisting of two technical replicates per condition were processed separately. Cells were permeabilized and run-on reactions were carried out as previously described[40]. Following RNA isolation, two adaptor ligations using T4 RNA Ligase 1 (catalog no. M0204; NEB) and reverse transcription using SuperScript III Reverse Transcriptase (catalog no. 18080044; Invitrogen) were performed, with custom adaptors detailed in Supplementary Table 1. Between adaptor ligations, cap state selection reactions were performed by treating the samples with CIP (catalog no. M0290; NEB) to reduce uncapped RNAs to 5′ hydroxyls and make them incapable of ligating to 5′ adaptor and Cap-Clip (catalog no. C-CC15011H; Cambio) to remove the 5′ cap of transcripts that had undergone guanylation and allow them to be incorporated into the library through 5′ adapter ligation. RNA washes, phenol:chloroform extractions and ethanol precipitations were carried out between reactions. All steps were performed under RNase-free conditions and following manufacturer protocols. Libraries were sequenced on an Illumina HiSeq 3000 following PCR amplification and library clean-up.

ChIA-PET and Hi-C Data Processing for Downstream Gene Predictions

Similar to a previously described method[47], LNCaP RNA Pol II ChIA-PET interactions[27] (GSM3423998) were lifted over from GRCh37 to GRCh38. LNCaP Hi-C data (ENCFF676WJO) was downloaded from ENCODE and interactions were lifted over from GRCh37 to GRCh38. ChIA-PET and Hi-C interactions were merged and filtered to remove duplicate interactions. The merged dataset was used as the final 3D interaction set.

Active enhancers and active promoters were inferred based on a merged set of unidirectional and bidirectional PRO-cap peaks from LNCaP treated with either ENZ or DMSO for 4 or 24 hrs. PRO-cap peak regions were intersected with GENCODE (V28) annotated promoters to get a list of active promoter regions. Active enhancers and promoters were overlapped with the final 3D interaction set to identify potential enhancer-gene pairs. A maximum of 1 Mb distance was allowed.

Analysis of Whole Genome Patient Data

Prostate adenocarcinoma whole genome variants were downloaded from TCGA. This included SNV data from 286 samples and indel data from 200 samples after filtering and overlapping with enhancer regions. The effects of these variants were predicted using Funseq2[50] (V2.1.4). The motif break and gain events of the variants were extracted from Funseq2 annotations.

Analysis of PRO-cap Data

Differential transcription at eRNA peaks was quantified using EdgeR[51] analysis of the total read counts in the core promoter (-35 to 60 bp from the peak TSS)[38] of the plus and minus strands separately. Induced enhancers were defined as FDR adj p-value < 0.05 in at least one peak direction.

CRISPRi

Lenti-dCas9-KRAB-blast was a gift from Gary Hon (Addgene plasmid # 89567). Lentivirus was produced in HEK293T cells, and subsequent virus containing media was used to transduce LNCaP cells, followed by blasticidin selection.

gRNAs against the candidate enhancer regions were designed using CRISPick. (https://portals.broadinstitute.org/gppx/crispick/public)[52,53] (Supplementary Table 1). Custom crRNAs were ordered from IDT and were annealed with Alt-R® CRISPR-Cas9 tracrRNA (IDT, 1072532) to generate sgRNAs according to the manufacturer instructions. For each experiment LNCaP-dCas9-KRAB cells were transfected with 30 nM sgRNA complex using Lipofectamine RNAiMAX (Invitrogen, 13778030). 48 hours post transfection cells were treated with media containing 10 µM enzalutamide or DMSO for a further 24 hours.

qPCR

RNA was extracted directly from cells using the RNeasy Mini Kit (Qiagen, 74106), and genomic DNA was removed using the DNA-free kit (Ambion, AM1906). RNA was reverse transcribed using random primers and SuperScript IV Reverse Transcriptase (Invitrogen, 18090010). Quantitative real-time PCR was performed on the ViiA 7 system (Applied Biosystems) using HOT FIREPol EvaGreen qPCR mix (Solis Biodyne, 08-24-00020) following manufacturer's instruction. Primer sequences are listed in Supplementary Table 1. All quantitative real-time PCR assays were carried out using three technical replicates using HMBS as the housekeeping gene.

Reporter Assays

Candidate reporter regions were synthesized as gene blocks (IDT) with Gateway cloning attB1 and attB2 sites on either end. Regions were reduced to 442 bp with the TSSs at the center, if necessary, to accommodate the Gateway sites and the 500 bp fragment synthesis limit. Fragments were then cloned into pDEST-hSCP1-luc using the Gateway cloning system (Invitrogen). LNCaP cells were transfected with 200 ng of the above luciferase reporters and 20 ng of renilla luciferase (pRL-CMV, Promega) in triplicate in 96-well plates with Lipofectamine 3000

(Invitrogen). 24 hours post-transfection cells were treated with media containing 10 μM enzalutamide or DMSO for a further 24 hours at which point luciferase activity was assayed with the Dual-Glo Luciferase Assay System (Promega).

Data availability

PRO-cap enhancer calls are available in Supplementary Data 1-7. Raw PRO-cap data and bigWig files are available through GEO under accession no. GSE198268. Other datasets used in this study are publicly available under the following accession nos. LNCaP ATAC-seq (accession no. GSE105116). LNCaP DNase-seq (accession no. GSM816637). LNCaP H3K27ac ChIP-seq (accession no. GSE107780). LNCaP H3K4me1 ChIP-seq (accession no. GSE73783). LNCaP RNA Pol II ChIA-PET (accession no. GSM3423998). LNCaP Hi-C (accession no. ENCFF676WJO). LNCaP AR ChIP-seq, LNCaP FOXA1 ChIP-seq, and LNCaP H3K27ac ChIP-seq (accession no. GSE85558). ATAC-seq, AR ChIP-seq, FOXA1 ChIP-seq, and H3K27ac ChIP-seq (accession no. GSE137775) of LNCaP treated with ENZ. ASH2L ChIP-seq (accession no. GSE60841) of VCaP treated with R1881 or cultured in media containing CS FCS. ChIP-seq (accession nos. GSE137775, GSE125245, GSE94682, GSE70079, GSE40269, GSE28264, and GSE80256) of LNCaP treated with ENZ, R1881, DHT, or cultured in media containing CS FCS. ChIP-seq (accession no. GSE130408) of normal prostate tissue, primary prostate cancer, and metastatic prostate cancer. LNCaP STARR-seq (GSE151064) of AR binding regions treated with DHT or EtOH.

Code availability

PRO-cap enhancers were called using Peak Identifier for Nascent Transcript Starts (PINTS)[39]. All analysis was performed using common publicly available tools.

Acknowledgements

Competing interests

The authors declare that there are no competing interests.

References

1. Heinlein, C.A. & Chang, C. Androgen receptor in prostate cancer. *Endocr Rev* **25**, 276-308 (2004).
2. Scher, H.I. & Sawyers, C.L. Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. *J Clin Oncol* **23**, 8253-61 (2005).
3. Taplin, M.E. *et al.* Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N Engl J Med* **332**, 1393-8 (1995).
4. Visakorpi, T. *et al.* In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* **9**, 401-6 (1995).
5. Roviello, G. *et al.* Role of the novel generation of androgen receptor pathway targeted agents in the management of castration-resistant prostate cancer: A literature based meta-analysis of randomized trials. *Eur J Cancer* **61**, 111-21 (2016).
6. Barbieri, C.E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* **44**, 685-9 (2012).
7. Chalmers, Z.R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* **9**, 34 (2017).
8. Armenia, J. *et al.* The long tail of oncogenic drivers in prostate cancer. *Nat Genet* **50**, 645-651 (2018).
9. Baca, S.C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-77 (2013).
10. Berger, M.F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-20 (2011).
11. Mosquera, J.-M. *et al.* Prevalence of TMPRSS2-ERG Fusion Prostate Cancer among Men Undergoing Prostate Biopsy in the United States. *Clinical Cancer Research* **15**, 4706-4711 (2009).
12. Tomlins, S.A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-8 (2005).
13. Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359-364 (2017).
14. Quigley, D.A. *et al.* Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell* **174**, 758-769 e9 (2018).
15. Viswanathan, S.R. *et al.* Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. *Cell* **174**, 433-447 e19 (2018).
16. Morova, T. *et al.* Androgen receptor-binding sites are highly mutated in prostate cancer. *Nat Commun* **11**, 832 (2020).
17. Takeda, D.Y. *et al.* A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. *Cell* **174**, 422-432 e13 (2018).
18. Gasperini, M., Tome, J.M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet* **21**, 292-310 (2020).
19. Arnold, C.D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-7 (2013).
20. Kheradpour, P. *et al.* Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* **23**, 800-11 (2013).
21. Kwasnieski, J.C., Fiore, C., Chaudhari, H.G. & Cohen, B.A. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res* **24**, 1595-602 (2014).
22. Vanhille, L. *et al.* High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* **6**, 6905 (2015).
23. Kron, K.J. *et al.* TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat Genet* **49**, 1336-1345 (2017).

24. Liu, Y. *et al.* Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol* **18**, 219 (2017).
25. Sharma, N.L. *et al.* The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. *Cancer Cell* **23**, 35-47 (2013).
26. Stelloo, S. *et al.* Integrative epigenetic taxonomy of primary prostate cancer. *Nat Commun* **9**, 4900 (2018).
27. Ramanand, S.G. *et al.* The landscape of RNA polymerase II-associated chromatin interactions in prostate cancer. *J Clin Invest* **130**, 3987-4005 (2020).
28. Rhie, S.K. *et al.* A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. *Nat Commun* **10**, 4154 (2019).
29. Taberlay, P.C. *et al.* Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res* **26**, 719-31 (2016).
30. Zhang, Z. *et al.* An AR-ERG transcriptional signature defined by long-range chromatin interactomes in prostate cancer cells. *Genome Res* **29**, 223-235 (2019).
31. Huang, C.F. *et al.* Functional mapping of androgen receptor enhancer activity. *Genome Biol* **22**, 149 (2021).
32. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-7 (2010).
33. Core, L.J. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**, 1311-20 (2014).
34. Tippens, N.D., Vihervaara, A. & Lis, J.T. Enhancer transcription: what, where, when, and why? *Genes Dev* **32**, 1-3 (2018).
35. Chen, H. *et al.* A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell* **173**, 386-399 e12 (2018).
36. Chen, H. & Liang, H. A High-Resolution Map of Human Enhancer RNA Loci Characterizes Super-enhancer Activities in Cancer. *Cancer Cell* **38**, 701-715 e5 (2020).
37. Henriques, T. *et al.* Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* **32**, 26-41 (2018).
38. Tippens, N.D. *et al.* Transcription imparts architecture, function and logic to enhancer units. *Nat Genet* **52**, 1067-1075 (2020).
39. Yao, L. *et al.* A comparison of experimental assays and analytical methods for genome-wide identification of active enhancers. *Nat Biotechnol* (2022).
40. Mahat, D.B. *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455-76 (2016).
41. Giambartolomei, C. *et al.* H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility. *Am J Hum Genet* **108**, 2284-2300 (2021).
42. Pomerantz, M.M. *et al.* Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat Genet* **52**, 790-799 (2020).
43. Horoszewicz, J.S. *et al.* LNCaP model of human prostatic carcinoma. *Cancer Res* **43**, 1809-18 (1983).
44. Layer, R.M. *et al.* GIGGLE: a search engine for large-scale integrated genome analysis. *Nat Methods* **15**, 123-126 (2018).
45. Mei, S. *et al.* Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res* **45**, D658-D662 (2017).
46. Shukla, S. *et al.* Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance. *Cancer Cell* **32**, 792-806 e7 (2017).
47. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**(2018).

48. Conti, D.V. *et al.* Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet* **53**, 65-75 (2021).
49. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-5 (2016).
50. Fu, Y. *et al.* FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol* **15**, 480 (2014).
51. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
52. Doench, J.G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).
53. Sanson, K.R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416 (2018).
54. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244-251 (2020).
55. McNair, C. *et al.* Differential impact of RB status on E2F1 reprogramming in human cancer. *J Clin Invest* **128**, 341-358 (2018).
56. Kim, J. *et al.* Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. *Cell Rep* **25**, 2808-2820 e4 (2018).
57. Hwang, J.H. *et al.* CREB5 Promotes Resistance to Androgen-Receptor Antagonists and Androgen Deprivation in Prostate Cancer. *Cell Rep* **29**, 2355-2370 e6 (2019).
58. Palit, S.A. *et al.* TLE3 loss confers AR inhibitor resistance by facilitating GR-mediated human prostate cancer cell growth. *Elife* **8**(2019).
59. Paltoglou, S. *et al.* Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer. *Cancer Res* **77**, 3417-3430 (2017).
60. Pomerantz, M.M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat Genet* **47**, 1346-51 (2015).
61. Rasool, R.U. *et al.* CDK7 Inhibition Suppresses Castration-Resistant Prostate Cancer through MED1 Inactivation. *Cancer Discov* **9**, 1538-1555 (2019).
62. Tan, P.Y. *et al.* Integration of regulatory networks by NKX3-1 promotes androgen-dependent prostate cancer survival. *Mol Cell Biol* **32**, 399-414 (2012).
63. Malik, R. *et al.* Targeting the MLL complex in castration-resistant prostate cancer. *Nat Med* **21**, 344-52 (2015).
64. Ge, S.X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628-2629 (2020).

Figures

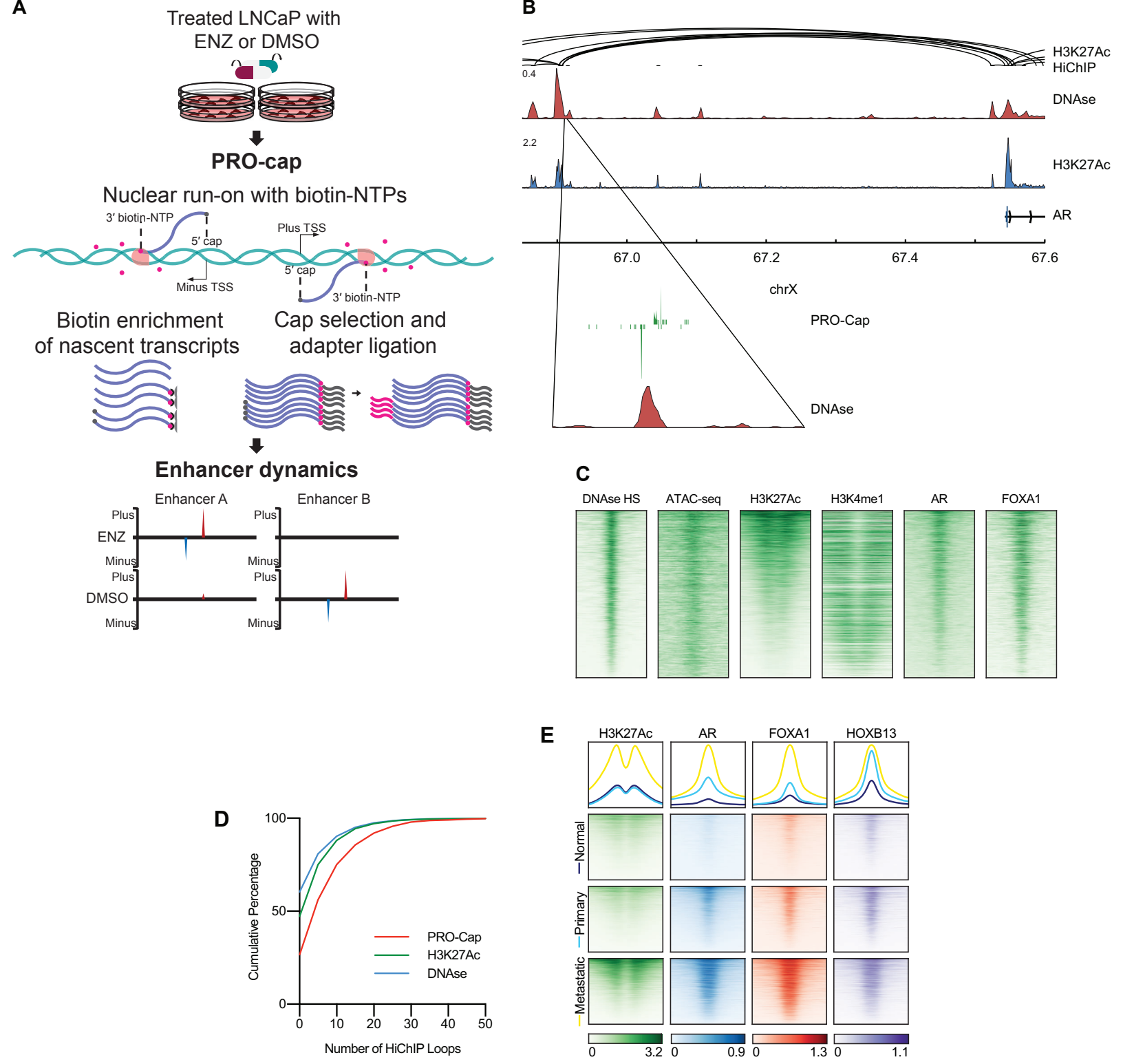

A
Treated LNCaP with
ENZ or DMSO
PRO-cap
Nuclear run-on with biotin-NTPs
3′ biotin-NTP
5′ cap
Minus TSS
Plus TSS
5′ cap
3′ biotin-NTP
Biotin enrichment
of nascent transcripts
Cap selection and
adapter ligation
Enhancer dynamics
Enhancer A
Enhancer B
Plus
ENZ
Minus
Plus
DMSO
Minus
B
H3K27Ac
HiChIP
DNAse
H3K27Ac
AR
67.0
67.2
67.4
67.6
chrX
PRO-Cap
DNAse
C
DNAse HS
ATAC-seq
H3K27Ac
H3K4me1
AR
FOXA1
D
Cumulative Percentage
Number of HiChIP Loops
PRO-Cap
H3K27Ac
DNAse
E
H3K27Ac
AR
FOXA1
HOXB13
Normal
Primary
Metastatic

Figure 1. PRO-cap sequencing of eRNAs to map active enhancers in prostate cancer. A) Overview of experimental pipeline implemented in this study. **B)** Enhancer candidates demonstrate divergent transcription of eRNAs. Shown is the upstream enhancer for androgen receptor (AR) along with published H3K27Ac HiChIP loops[41] and read tracks for DNase (GSM816637) and H3K27Ac (GSE85558) ChIP-seq[46,54] of LNCaP cells. **C)** Putative enhancers identified by PRO-Cap (n = 6,189) are bound by canonical biochemical marks and known important prostate transcription factors in LNCaP cells. Heatmaps show read density from publicly available ChIP-seq data[29,46,54-56] (DNase: GSM816637, AR/FOXA1: GSE85558, ATAC-seq: GSE105116, H3K4me1: GSE73783, and H3K27Ac: GSE107780) 1kb up- and downstream of the center of the enhancer. **D)** H3K27Ac loop anchors[41] are significantly more likely ($P < 0.0001$, Kolmogorov-Smirnov test) to overlap with PRO-cap peaks versus H3K27Ac or DNase I peaks. **E)** Putative enhancers identified by PRO-Cap are also bound by biochemical marks and prostate transcription factors in patient samples. Heatmaps show average read density from publicly available ChIP-seq[42] data (GSE130408) from normal prostate tissue, primary PCa, and metastatic PCa (n = 8 samples for each).

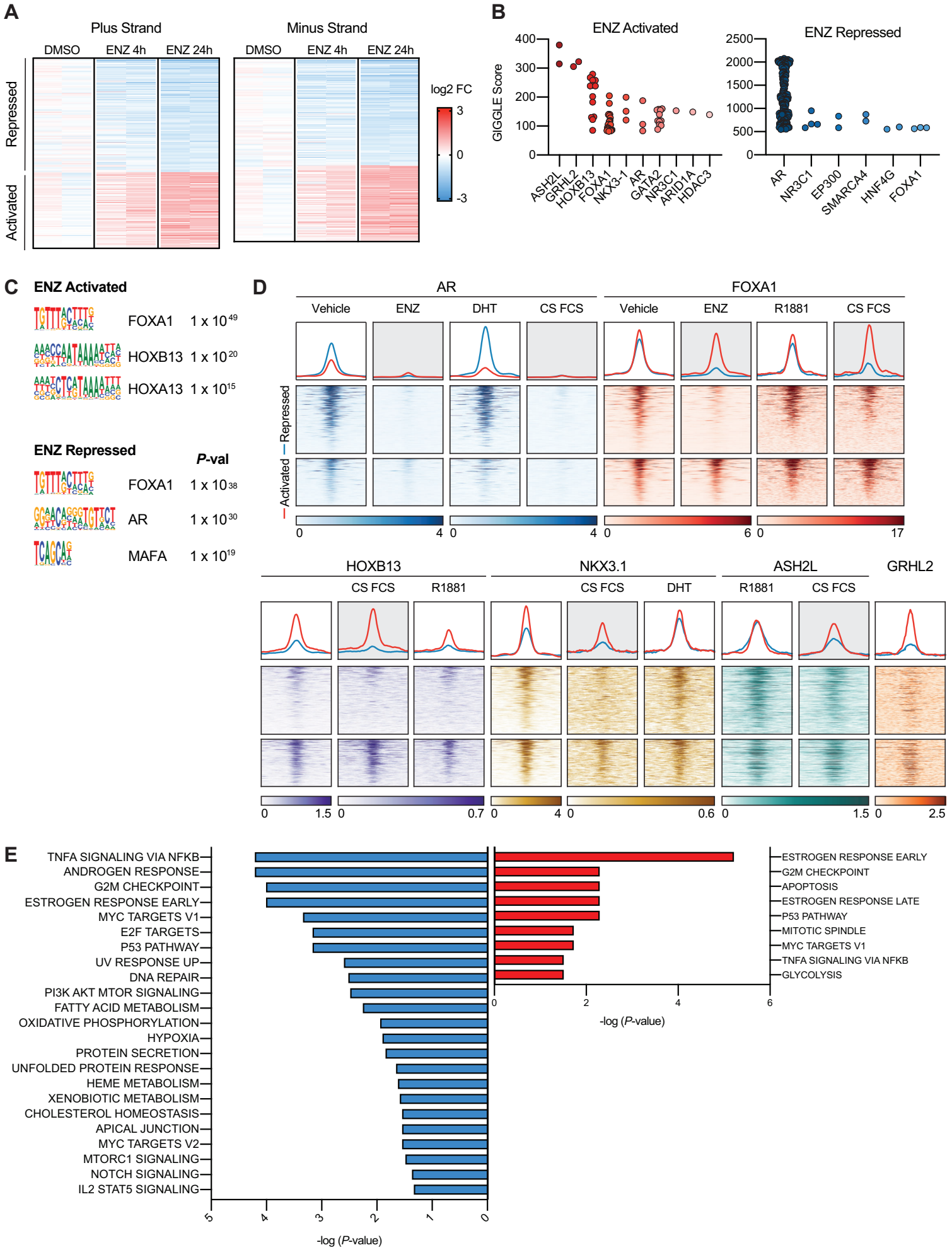

A
Plus Strand
Minus Strand
DMSO
ENZ 4h
ENZ 24h
Repressed
Activated
log2 FC
3
0
-3
B
ENZ Activated
ENZ Repressed
GIGGLE Score
400
300
200
100
0
2500
2000
1500
1000
500
ASH2L
GRHL2
HOXB13
FOXA1
NKX3-1
AR
GATA2
NR3C1
ARID1A
HDAC3
NR3C1
EP300
SMARCA4
HNF4G
C
ENZ Activated
FOXA1 1 x 10^49
HOXB13 1 x 10^20
HOXA13 1 x 10^15
ENZ Repressed
P-val
FOXA1 1 x 10^38
AR 1 x 10^30
MAFA 1 x 10^19
D
AR
FOXA1
Vehicle
ENZ
DHT
CS FCS
R1881
HOXB13
NKX3.1
ASH2L
GRHL2
Repressed
Activated
0
4
6
17
1.5
0.7
0.6
2.5
E
TNFA SIGNALING VIA NFKB
ANDROGEN RESPONSE
G2M CHECKPOINT
ESTROGEN RESPONSE EARLY
MYC TARGETS V1
E2F TARGETS
P53 PATHWAY
UV RESPONSE UP
DNA REPAIR
PI3K AKT MTOR SIGNALING
FATTY ACID METABOLISM
OXIDATIVE PHOSPHORYLATION
HYPOXIA
PROTEIN SECRETION
UNFOLDED PROTEIN RESPONSE
HEME METABOLISM
XENOBIOTIC METABOLISM
CHOLESTEROL HOMEOSTASIS
APICAL JUNCTION
MYC TARGETS V2
MTORC1 SIGNALING
NOTCH SIGNALING
IL2 STAT5 SIGNALING
ESTROGEN RESPONSE EARLY
G2M CHECKPOINT
APOPTOSIS
ESTROGEN RESPONSE LATE
P53 PATHWAY
MITOTIC SPINDLE
MYC TARGETS V1
TNFA SIGNALING VIA NFKB
GLYCOLYSIS
-log (P-value)
0
2
4
6
5
4
3
2
1
0

Figure 2. PRO-cap sequencing of eRNAs identifies enhancers affected by enzalutamide. A) PRO-Cap analysis of LNCaP cells treated with 10 µM enzalutamide (ENZ) for 24 hours identified 853 putative enhancers which demonstrated significant differential activation or repression on the plus or minus strands as determined by edgeR analysis (FDR < 0.05). Data is displayed as a heatmap of the fold-change in mapped reads with 4 or 24 hour ENZ treatments. For each treatment condition two independent biological replicates are shown. **B)** GIGGLE[44] analysis demonstrates overlap of the ENZ regulated enhancers with regions identified in published ChIP-seq datasets. Each dot represents a single study, and the GIGGLE score incorporates both enrichment and significance. **C)** Overrepresented TF motifs as determined by Cistrome MDSeqPos[45] for the ENZ regulated enhancers. **D)** Heatmaps and summary plots show read density 1kb up- and downstream of the center of the ENZ regulated enhancers. Data is from publicly available ChIP-seq data[57-62] (GSE137775, GSE125245, GSE94682, GSE70079, GSE40269, GSE28264, and GSE80256) for LNCaP cells treated with ENZ, R1881, DHT or cultured in media containing charcoal stripped (CS) FCS. ASH2L data[63] (GSE60841) is from VCaP cells treated with R1881 or cultured in media containing CS FCS. **E)** Gene ontology analysis of the genes predicted to be regulated by these enhancers generated using ShinyGO[64].

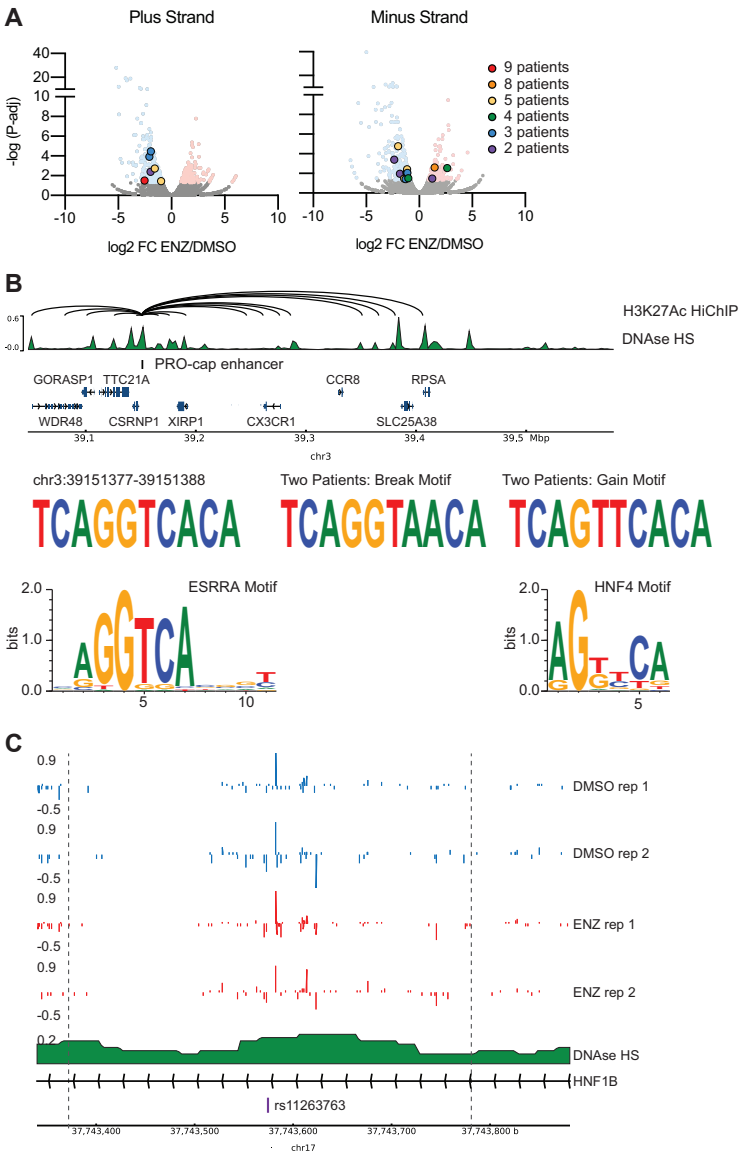

A
Plus Strand
Minus Strand
-log (P-adj)
log2 FC ENZ/DMSO
log2 FC ENZ/DMSO
9 patients
8 patients
5 patients
4 patients
3 patients
2 patients
B
H3K27Ac HiChIP
DNAse HS
PRO-cap enhancer
GORASP1
TTC21A
CCR8
RPSA
WDR48
CSRNP1
XIRP1
CX3CR1
SLC25A38
chr3
chr3:39151377-39151388
Two Patients: Break Motif
Two Patients: Gain Motif
ESRRA Motif
HNF4 Motif
bits
C
DMSO rep 1
DMSO rep 2
ENZ rep 1
ENZ rep 2
DNAse HS
HNF1B
rs11263763
chr17

Figure 3. PRO-cap helps prioritize candidate enhancer regions to search for cancer-associated germline and somatic variants. A) Analysis of 286 PCa whole genomes identified 137 single nucleotide variants (SNV) in the ENZ-regulated enhancers identified via PRO-cap, 20 of which were recurrent in more than one patient. **B)** Example of an enhancer downregulated by ENZ which demonstrates two separate recurrent SNVs, one in two patients that breaks an ESRRA motif, and another in two different patients that creates an HNF4 motif. **C)** Example of a PCa risk variant[48] which overlaps with an enhancer identified via PRO-cap.

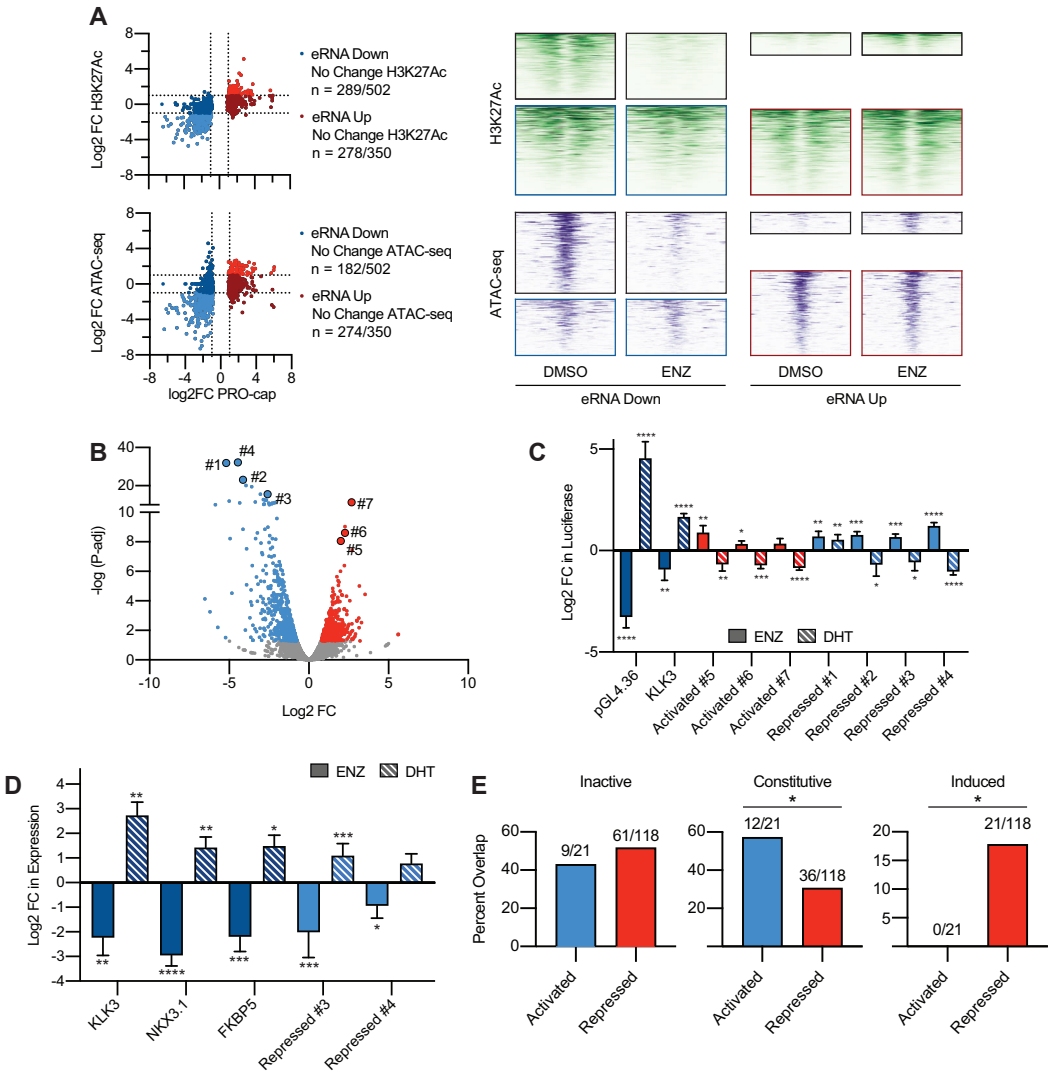

A
Log2 FC H3K27Ac
eRNA Down
No Change H3K27Ac
n = 289/502
eRNA Up
No Change H3K27Ac
n = 278/350
Log2 FC ATAC-seq
eRNA Down
No Change ATAC-seq
n = 182/502
eRNA Up
No Change ATAC-seq
n = 274/350
log2FC PRO-cap
H3K27Ac
ATAC-seq
DMSO
ENZ
DMSO
ENZ
eRNA Down
eRNA Up
B
-log (P-adj)
Log2 FC
#1
#2
#3
#4
#5
#6
#7
C
Log2 FC in Luciferase
ENZ
DHT
pGL4.36
KLK3
Activated #5
Activated #6
Activated #7
Repressed #1
Repressed #2
Repressed #3
Repressed #4
D
Log2 FC in Expression
ENZ
DHT
KLK3
NKX3.1
FKBP5
Repressed #3
Repressed #4
E
Inactive
Constitutive
Induced
Percent Overlap
9/21
61/118
12/21
36/118
0/21
21/118
Activated
Repressed
Activated
Repressed
Activated
Repressed

Figure 4. Functional validation of PRO-cap enhancers. A) Comparing changes in eRNA expression with ENZ treatment as detected by PRO-cap with changes in H3K27Ac ChIP-seq and ATAC-seq. XY charts and heatmaps show read density 1kb up- and downstream of the center of the ENZ regulated enhancers. Data is from publicly available data[57] (GSE137775). **B)** Selected candidate enhancers for functional analysis. **C)** Analysis of changes in enhancer activity with androgen stimulation or inhibition as determined by luciferase reported assays does not agree with PRO-cap data. 24 hours post transfection cells were treated with 10 µM enzalutamide or 10 nM DHT for and additional 24 hours. Shown is the average log2 fold change between DMSO and ENZ or EtOH and DHT from n = 3 experiments, each with n = 3 replicates. Statistics were determined by t-test. **D)** Analysis of eRNA expression in response to AR stimulation (10 nM DHT) or inhibition (10 µM ENZ). Expression was measured via qPCR (n = 3) and is displayed as the average log2 fold change with treatment. Statistics were determined by t-test**. E)** Categorization of ENZ-regulated enhancers identified with PRO-cap in publish STARR-seq data[31]. Activity of 139 of the 853 candidate enhancers was measured in response to androgen stimulation (DHT) and categorized as inactive (no enhancer activity), constitutively active (no change in activity with DHT) or induced by DHT. Statistics were determined using Fisher's exact test.

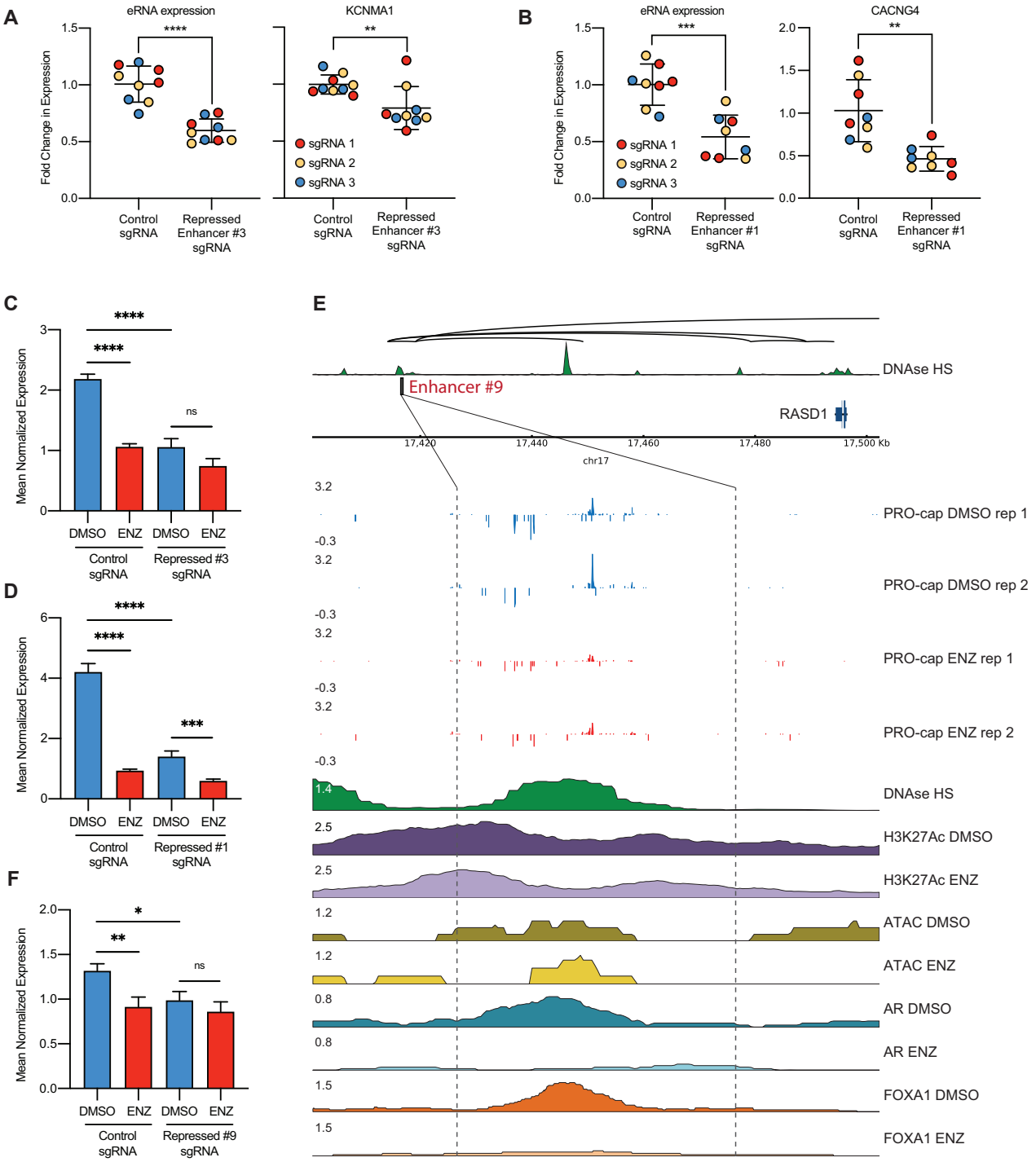
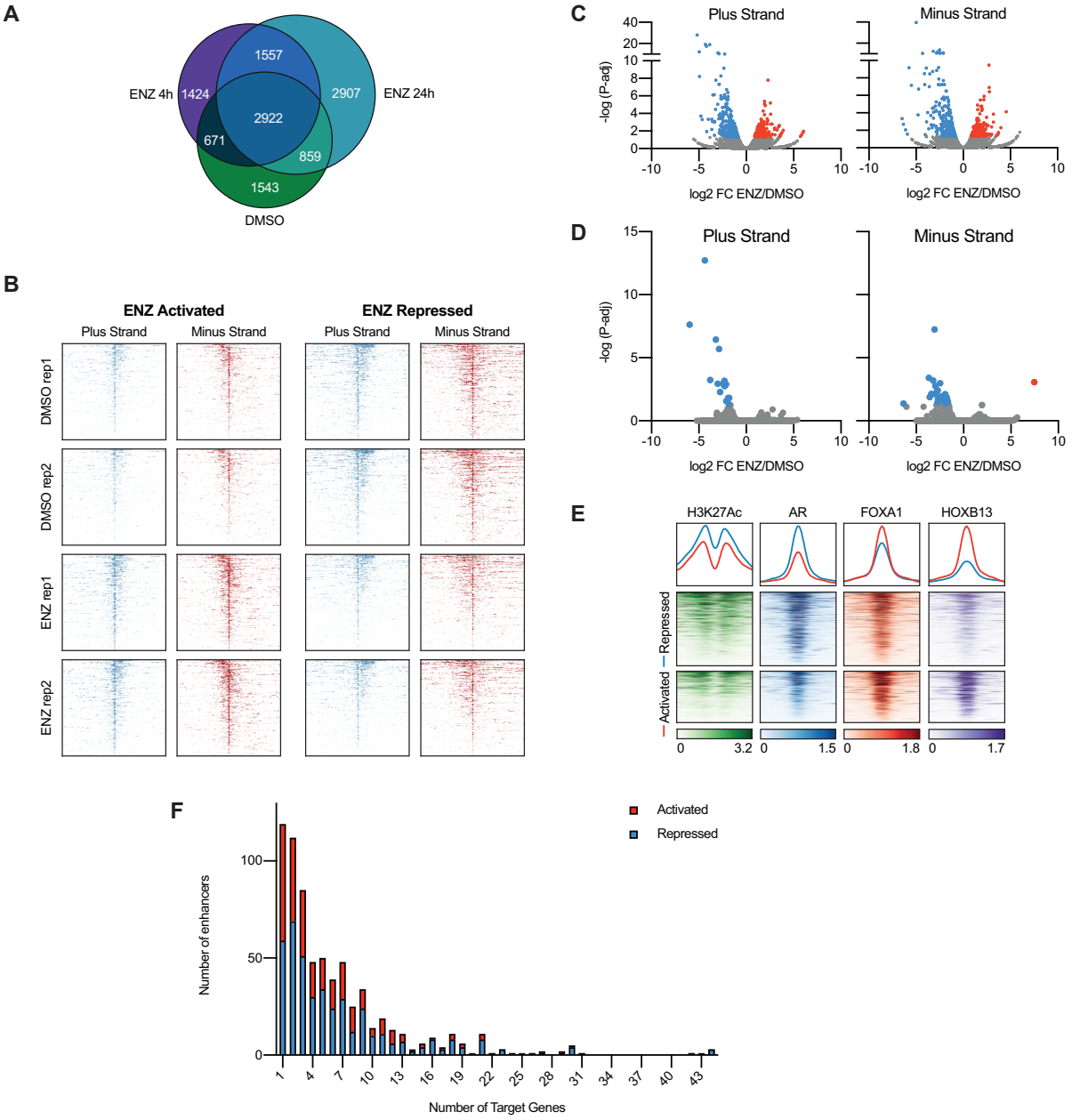
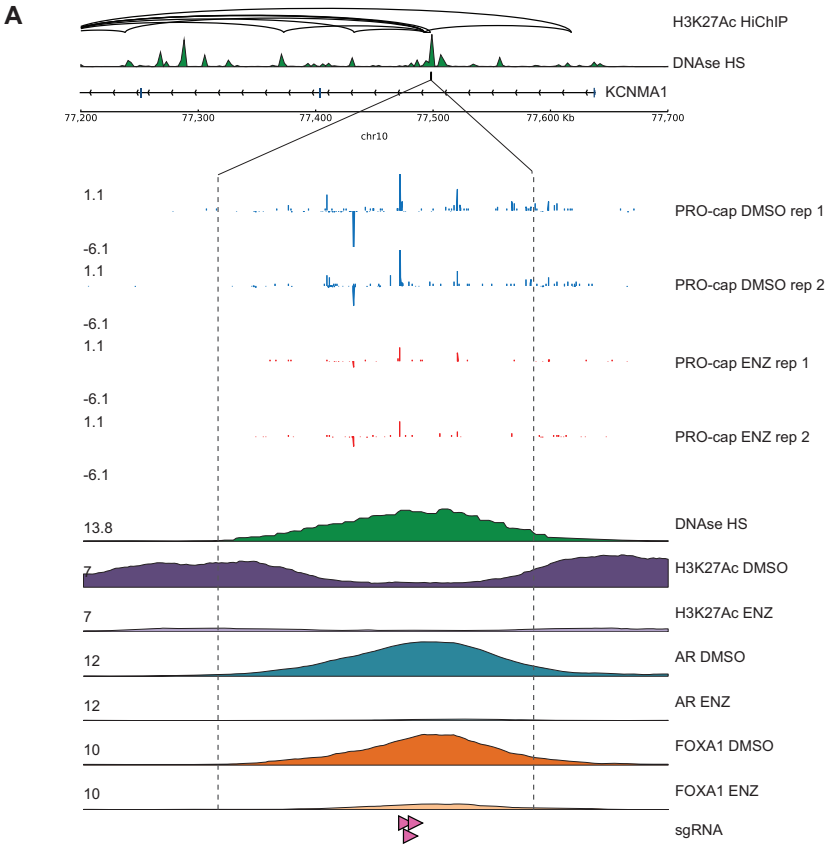
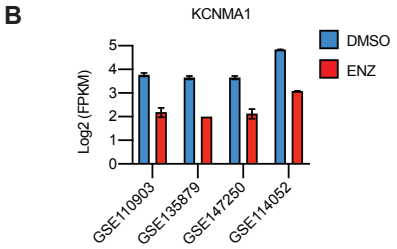

A
eRNA expression
KCNMA1
Fold Change in Expression
****
**
sgRNA 1
sgRNA 2
sgRNA 3
Control sgRNA
Repressed Enhancer #3 sgRNA
B
eRNA expression
CACNG4
***
**
Repressed Enhancer #1 sgRNA
C
Mean Normalized Expression
DMSO
ENZ
Control sgRNA
Repressed #3 sgRNA
ns
D
Repressed #1 sgRNA
E
DNAse HS
Enhancer #9
RASD1
chr17
PRO-cap DMSO rep 1
PRO-cap DMSO rep 2
PRO-cap ENZ rep 1
PRO-cap ENZ rep 2
DNAse HS
H3K27Ac DMSO
H3K27Ac ENZ
ATAC DMSO
ATAC ENZ
AR DMSO
AR ENZ
FOXA1 DMSO
FOXA1 ENZ
F
Repressed #9 sgRNA

Figure 5. CRISPRi interrogation of PRO-cap enhancer candidates. A) CRISPRi targeting PRO-cap candidate enhancer #3 with three different sgRNAs significantly reduces expression of the eRNA itself ($P < 0.0001$), and of the downstream target gene *KCNMA1* ($P = 0.0083$). **B)** CRISPRi targeting PRO-cap candidate enhancer #1 with three different sgRNAs significantly reduces expression of the eRNA itself ($P = 0.0002$), and of the downstream target gene *CACNG4* ($P = 0.0011$). **C)** CRISPRi targeting PRO-cap candidate enhancer #3 with three different sgRNAs impairs the ENZ-regulation of the downstream target gene *KCNMA1* (2.05-fold change, $P < 0.0001$ vs. 1.42-fold change). **D)** CRISPRi targeting PRO-cap candidate enhancer #1 with three different sgRNAs impairs the ENZ-regulation of the downstream target gene *CACNG4* (4.48-fold change, $P < 0.0001$ vs. 2.35-fold change, $P = 0.0007$). **E)** Shown is the candidate enhancer PRO-Cap signal along with published H3K27Ac HiChIP loops[41] and read tracks for DNase (GSM816637), ATAC-seq, and ChIP-seq with and without ENZ for H3K27Ac, AR, and FOXA1 (GSE137775)[57]. **F)** CRISPRi targeting PRO-cap candidate enhancer #9 with two different sgRNAs impairs the ENZ-regulation of the downstream target gene *RASD1* (1.44-fold change, $P = 0.0084$ vs. 1.15-fold change).

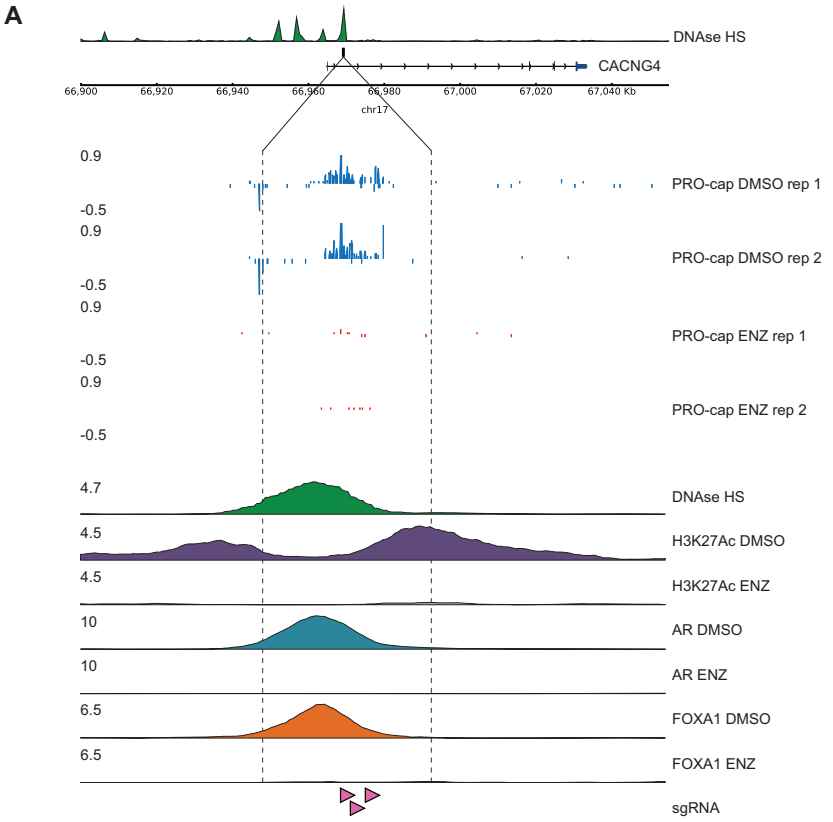

A
ENZ 4h
ENZ 24h
DMSO
1424
1557
2907
2922
671
859
1543
B
ENZ Activated
ENZ Repressed
Plus Strand
Minus Strand
DMSO rep1
DMSO rep2
ENZ rep1
ENZ rep2
C
Plus Strand
Minus Strand
-log (P-adj)
log2 FC ENZ/DMSO
D
Plus Strand
Minus Strand
-log (P-adj)
log2 FC ENZ/DMSO
E
H3K27Ac
AR
FOXA1
HOXB13
Repressed
Activated
0
3.2
0
1.5
0
1.8
0
1.7
F
Activated
Repressed
Number of enhancers
Number of Target Genes

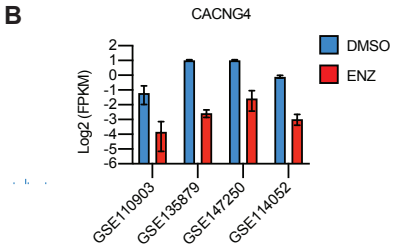
Supplementary Figure 1. PRO-cap sequencing of eRNAs identifies enhancers affected by enzalutamide. A) Venn diagram of the overlap of enhancer regions between the three treatments paradigms **B)** Heatmap of the PRO-cap signal 250 bp up- and downstream of ENZ activated or repressed peaks from two replicates of LNCaP cells treated with DMSO or 10 µM ENZ for 24 hours. **C)** PRO-Cap analysis of LNCaP cells treated with 10 µM enzalutamide (ENZ) for 24 hours identified 853 putative enhancers which demonstrated significant differential activation or repression on the plus or minus strands as determined by edgeR analysis (FDR < 0.05). **D)** PRO-Cap analysis of LNCaP cells treated with 10 µM enzalutamide (ENZ) for 4 hours identified 44 putative enhancers which demonstrated significant differential activation or repression on the plus or minus strands as determined by edgeR analysis (FDR < 0.05). **E)** Heatmaps and summary plots show read density 1kb up- and downstream of the center of the ENZ regulated enhancers in metastatic PCa. Data is from publicly available ChIP-seq data[42] (GSE130408) and is displayed as the mean signal from n = 8 metastatic PCa samples. **F)** Histogram demonstrating the number of downstream target gene predictions for the 853 ENZ-regulated enhancers identified with PRO-cap.


A
H3K27Ac HiChIP
DNAse HS
KCNMA1
77,200
77,300
77,400
77,500
77,600 Kb
77,700
chr10
1.1
-6.1
PRO-cap DMSO rep 1
PRO-cap DMSO rep 2
PRO-cap ENZ rep 1
PRO-cap ENZ rep 2
13.8
DNAse HS
7
H3K27Ac DMSO
H3K27Ac ENZ
12
AR DMSO
AR ENZ
10
FOXA1 DMSO
FOXA1 ENZ
sgRNA

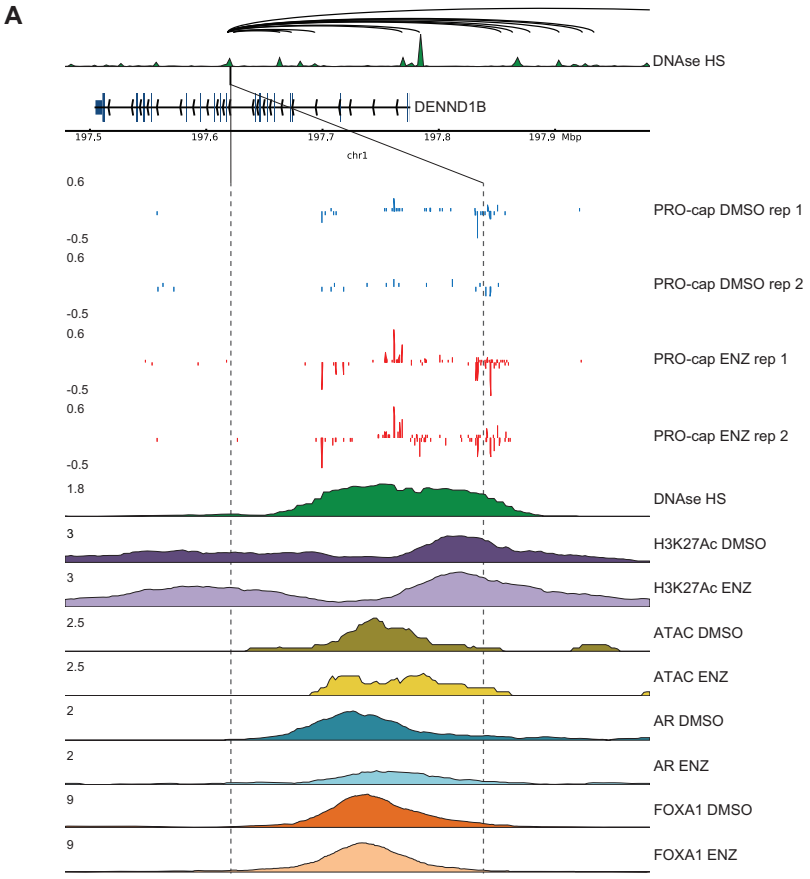

B
KCNMA1
DMSO
ENZ
Log2 (FPKM)
0
1
2
3
4
5
GSE110903
GSE135879
GSE147250
GSE114052

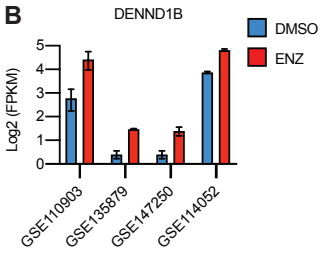
Supplementary Figure 2. Candidate Enhancer #3. A) Shown is the candidate enhancer PRO-Cap signal along with published H3K27Ac HiChIP loops[41] and read tracks for DNase (GSM816637), and ChIP-seq with and without ENZ for H3K27Ac, AR, and FOXA1 (GSE137775)[57]. **B)** Expression of the candidate downstream target gene *KCNMA1* in response to ENZ from multiple published RNA-seq data sets, accession number as indicated.

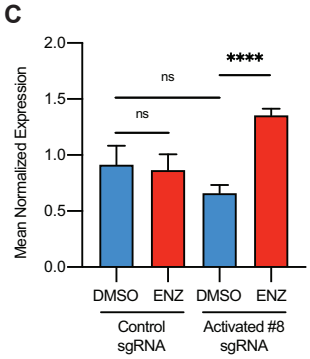

A
DNAse HS
CACNG4
66,900
66,920
66,940
66,960
66,980
67,000
67,020
67,040 Kb
chr17
0.9
-0.5
PRO-cap DMSO rep 1
PRO-cap DMSO rep 2
PRO-cap ENZ rep 1
PRO-cap ENZ rep 2
4.7
DNAse HS
4.5
H3K27Ac DMSO
H3K27Ac ENZ
10
AR DMSO
AR ENZ
6.5
FOXA1 DMSO
FOXA1 ENZ
sgRNA


B
CACNG4
DMSO
ENZ
Log2 (FPKM)
2
1
0
-1
-2
-3
-4
-5
-6
GSE110903
GSE135879
GSE147250
GSE114052

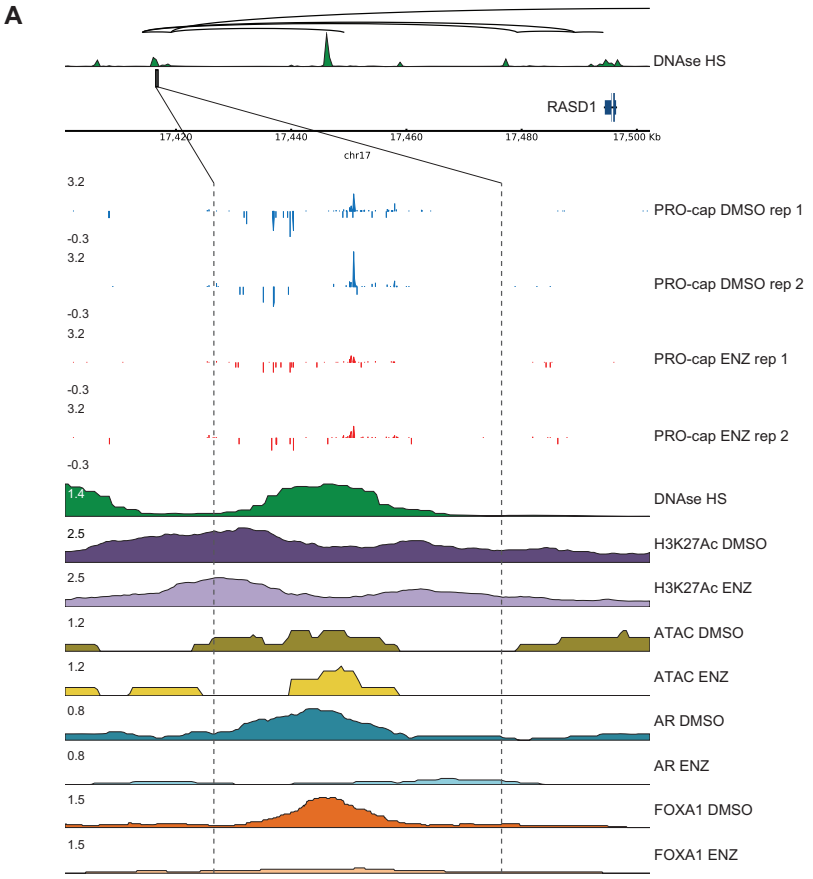
Supplementary Figure 3 Candidate Enhancer #1. A) Shown is the candidate enhancer PRO-Cap signal along with published H3K27Ac HiChIP loops[41] and read tracks for DNase (GSM816637), and ChIP-seq with and without ENZ for H3K27Ac, AR, and FOXA1 (GSE137775)[57]. **B)** Expression of the candidate downstream target gene *CACNG4* in response to ENZ from multiple published RNA-seq data sets, accession number as indicated.

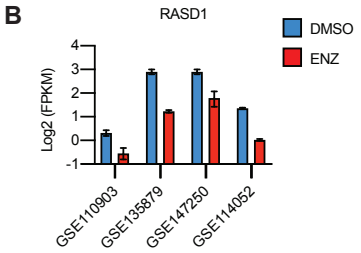

A
DNAse HS
DENND1B
197.5
197.6
197.7
197.8
197.9 Mbp
chr1
0.6
-0.5
PRO-cap DMSO rep 1
PRO-cap DMSO rep 2
PRO-cap ENZ rep 1
PRO-cap ENZ rep 2
1.8
DNAse HS
3
H3K27Ac DMSO
H3K27Ac ENZ
2.5
ATAC DMSO
ATAC ENZ
2
AR DMSO
AR ENZ
9
FOXA1 DMSO
FOXA1 ENZ


B
DENND1B
DMSO
ENZ
Log2 (FPKM)
0
1
2
3
4
5
GSE110903
GSE135879
GSE147250
GSE114052


C
Mean Normalized Expression
0.0
0.5
1.0
1.5
2.0
ns
ns
****
DMSO
ENZ
DMSO
ENZ
Control sgRNA
Activated #8 sgRNA

Supplementary Figure 4. Candidate Enhancer #8. A) Shown is the candidate enhancer PRO-Cap signal along with published H3K27Ac HiChIP loops[41] and read tracks for DNase (GSM816637), ATAC-seq, and ChIP-seq with and without ENZ for H3K27Ac, AR, and FOXA1 (GSE137775)[57]. **B)** Expression of the candidate downstream target gene *DENND1B* in response to ENZ from multiple published RNA-seq data sets, accession number as indicated. **C)** CRISPRi targeting PRO-cap candidate enhancer #8 with three different sgRNAs impairs the ENZ-regulation of the downstream target gene DENND1B (0.946-fold change vs. 2.06-fold change $P < 0.0001$).


A
DNAse HS
RASD1
17,420
17,440
17,460
17,480
17,500 Kb
chr17
3.2
-0.3
PRO-cap DMSO rep 1
PRO-cap DMSO rep 2
PRO-cap ENZ rep 1
PRO-cap ENZ rep 2
1.4
DNAse HS
2.5
H3K27Ac DMSO
H3K27Ac ENZ
1.2
ATAC DMSO
ATAC ENZ
0.8
AR DMSO
AR ENZ
1.5
FOXA1 DMSO
FOXA1 ENZ


B
RASD1
DMSO
ENZ
Log2 (FPKM)
4
3
2
1
0
-1
GSE110903
GSE135879
GSE147250
GSE114052

Supplementary Figure 5. Candidate Enhancer #9. A) Shown is the candidate enhancer PRO-Cap signal along with published H3K27Ac HiChIP loops[41] and read tracks for DNase (GSM816637), ATAC-seq, and ChIP-seq with and without ENZ for H3K27Ac, AR, and FOXA1 (GSE137775)[57]. **B)** Expression of the candidate downstream target gene *RASD1* in response to ENZ from multiple published RNA-seq data sets, accession number as indicated.