

ARTICLE

Open Access

Expression of lncRNA *MIR222HG* co-transcribed from the *miR-221/222* gene promoter facilitates the development of castration-resistant prostate cancer

Tong Sun¹, Shin-Yi Du², Joshua Armenia³, Fangfang Qu¹, Jingyu Fan⁴, Xiaodong Wang¹, Teng Fei¹, Kazumasa Komura¹, Shirley X. Liu⁴, Gwo-Shu Mary Lee¹ and Philip W Kantoff^{1,2}

Abstract

Mechanisms by which non-coding RNAs contribute to the progression of hormone-sensitive prostate cancer (PCa) (HSPC) to castration-resistant PCa (CRPC) remain largely unknown. We previously showed that *microRNA-221/222* is up-regulated in CRPC and plays a critical role in modulating androgen receptor function during CRPC development. With further investigation, we characterized a putative promoter region located 23.3 kb upstream of the *miR-221/222* gene, and this promoter is differentially activated in CRPC LNCaP-Abl cells, leading to the up-regulation of *miR-221/222*. Upon promoter activation, a set of polyadenylated long non-coding RNA (lncRNA) *MIR222HG*s was transcribed from this promoter region. Over-expression of these *MIR222HG*s increased androgen-independent cell growth and repressed the expression of androgen receptor-regulated dihydrotestosterone (DHT)-induced *KLK3*, *TMPRSS2*, and *FKBP5* in HSPC LNCaP cells, hallmarks of the CRPC phenotype. Clinically, increased expression of *MIR222HG* is associated with PCa progression to CRPC. In primary tumors, expression levels of *MIR222HG* and *miR-221/222* inversely correlate with Gleason score and androgen receptor (AR) pathway activity. Interestingly, *MIR222HG* is Argonaute 2-bound and its expression is Dicer 1-dependent, suggesting its functional association with the RNA-induced silencing complex. Further studies led to the hypothesis that *MIR222HG* may potentially affect miR-mediated expression silencing, subsequently leading to AR reprogramming. Our study highlights an essential role of a non-coding RNA in CRPC development and that differential activation of a single promoter can up-regulate two different types of non-coding RNAs, *miR-221/222* and lncRNA *MIR222HG*, in CRPC. Additionally, this study reveals a novel function of lncRNAs as a modulator of Argonaute-mediated RNA-induced silencing complex.

Introduction

Prostate cancer (PCa) is the most common non-cutaneous malignancy diagnosed in American men and the second leading cause of cancer mortality in men¹.

Androgens and the androgen receptor (AR) play crucial roles in PCa development and progression. Androgen deprivation therapy (ADT) remains a key treatment for advanced PCa. Most hormone-sensitive PCa patients initially respond to ADT; however, ultimately most patients develop resistance to ADT and progress to the lethal castration-resistant PCa (CRPC)^{2,3}. PCa cells utilize a variety of AR-dependent and AR-independent pathways to survive in an androgen-depleted environment during CRPC progression. Many studies have shown that CRPC is frequently characterized by altered AR expression and

Correspondence: G-S. Lee (gwo-shu_lee@dfci.harvard.edu) or Philip W. Kantoff (kantoff@mskcc.org)

¹Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, 450 Brookline Ave, Boston, MA 02215, USA

²Department of Medicine, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

Full list of author information is available at the end of the article

These authors contributed equally: Tong Sun, Shin-Yi Du, Joshua Armenia.

© The Author(s) 2018



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

persistent AR signaling activated by residual androgens^{4–6}. However, our understanding of the mechanisms underlying CRPC development and progression remains limited.

Non-coding RNAs (ncRNAs) are major components of the eukaryotic transcriptome⁷. Based on their size range, ncRNAs are classified into small (<200 nucleotides) and long ncRNAs (lncRNAs; >200 nucleotides). These ncRNAs play important regulatory roles in diverse biological processes including cancer development and progression⁸. Among the different classes of ncRNAs, microRNAs (miRs) are relatively conserved groups of small ncRNAs ranging between 19 and 25 nucleotides that negatively regulate gene expression. Unlike small ncRNAs, lncRNAs are less conserved among different species and are usually expressed at relatively low levels. The major functions of identified lncRNAs, thus far, include (a) regulation of gene expression, both transcriptionally or post-transcriptionally, (b) facilitation of the formation of functional transcriptional complexes at promoter or enhancer sites by induction of chromatin structural changes, and (c) recruitment of various effector molecules or direct interaction with other RNAs^{9,10}. Increasing numbers of cancer-related lncRNAs have been identified;¹¹ however, many of their biological mechanisms leading to cancer development and progression remain unknown. It has been shown that lncRNAs, *PCA3* and *PCAT-1*, are up-regulated in PCa and that the lncRNA, *SChLAP1*, is associated with aggressive PCa^{12–14}. We recently demonstrated that lncRNAs could work through a “sponge” mechanism competing for miRs that regulate phosphatase and tensin homolog (PTEN) expression, an important tumor suppressor gene in PCa progression¹⁵. We anticipated that additional novel functional mechanisms of ncRNAs may be involved in the development of CRPC.

Previously, we found that elevated expression of *miR-221/222* is associated with human CRPC. In vitro, up-regulation of *miR-221/222* expression conferred androgen-independent cell growth, reduced the transcription of a subset of androgen-responsive genes, and promoted the activation of epithelial–mesenchymal transition/tumor metastasis pathways^{16,17}. Thus, up-regulation of *miR-221/222* potentially leads to reprogramming of AR signaling, which in turn may mediate the transition to the CRPC phenotype. We continued our investigation of the structure of the *miR-221/222* locus to help understand the potential mechanisms involved in the regulation of *miR-221/222* expression in CRPC.

We present here the characterization of *miR-221/222* locus and the role of the promoter co-transcribed *MIR222HG* lncRNAs in the development of CRPC phenotype. Our study suggests that up-regulation of *miR-221/222* and lncRNA *MIR222HG* expression in CRPC are

most likely driven by the activation of the same promoter and both of them are involved in the progression from hormone-sensitive prostate cancer (HSPC) to CRPC.

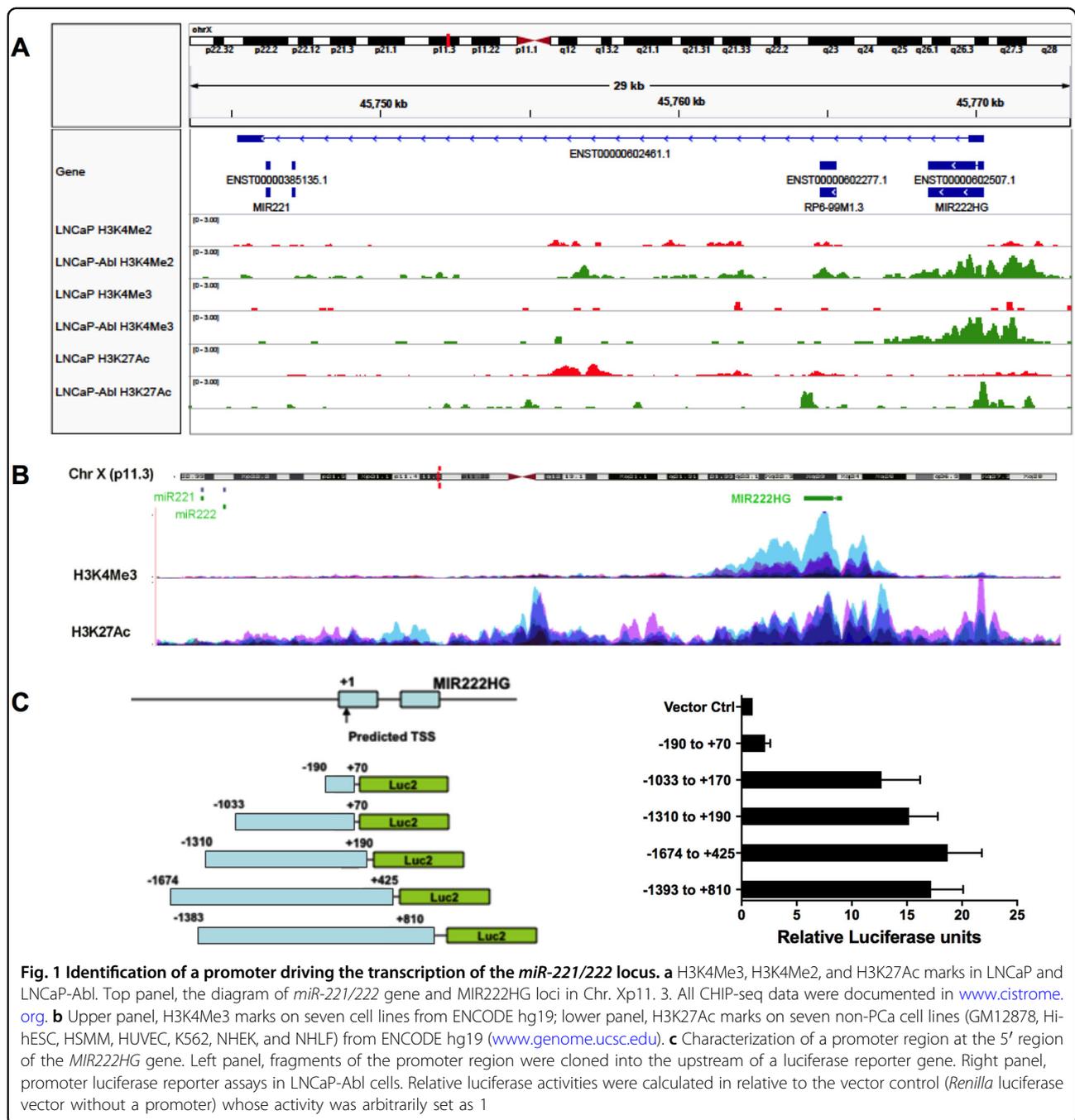
Results

lncRNA *MIR222HG* are coordinately transcribed from the *miR-221/222* gene promoter

We previously showed that *miR-221/222* is frequently over-expressed in CRPC tumors, participating in the CRPC development. *MIR-221/222* is expressed >10-fold higher in the androgen-independent LNCaP-Abl cell line compared to that of the androgen-dependent LNCaP^{16,17}. LNCaP-Abl cell line was established by a long period of adapting growth of LNCaP in medium containing charcoal-stripped fetal bovine serum (CFBS). The whole chromosome AR binding sites in LNCaP and LNCaP-Abl have been mapped and the dynamics of AR transcription complex loading on the regulatory regions are well characterized¹⁸. To address the mechanisms underlying the transcriptional regulation of *miR-221/222* in PCa cells, we identified the potential regulatory region of the *miR-221/222* gene locus by comparing histone modification marks in LNCaP and LNCaP-Abl cell lines with further confirmation using ENCODE hg19 histone modification data (Fig. 1).

We compared the H3K4Me2 (define transcription factor binding sites), H3K4Me3 (highly enriched at active promoters near transcription start sites, TSS) and H3K27Ac (associated with a higher transcriptional activation that defines an active enhancer mark) chromatin immunoprecipitation sequencing (ChIP-seq) data between LNCaP and LNCaP-Abl¹⁸. We noticed that a cluster of high-density active transcription histone marks H3K4Me3 and H3K4Me2 was located –23.3 kb upstream of the *miR-221/222* gene in LNCaP-Abl, but not in LNCaP (Fig. 1a). Consistently, the H3K27Ac mark, associated with transcriptional initiation and open chromatin, was also detected in the same location in LNCaP-Abl, but not in LNCaP (Fig. 1a). With further comparison of ENCODE data, we confirmed that the dense H3K4Me3 marks, and a wider H3K27Ac mark at –23.3 kb upstream of the *miR-221/222* gene were also present in many other non-PCa cell lines (Fig. 1b; <https://genome.ucsc.edu/>). Additionally, a DNAase I hypersensitivity cluster site (from 125 cell types) also co-localized in the same position, indicating possible transcriptional activity (<https://genome.ucsc.edu/>)¹⁹. Taken together, these data strongly suggested that a promoter region was located at –23.3 kb upstream of the *miR-221/222* gene, and, most likely, the activation of this upstream promoter leads to the differential up-regulation of *miR-221/222* expression in LNCaP-Abl.

We further analyzed the nucleotide sequence of this putative promoter of *miR-221/222* gene to identify



TSS. By employing the Promoter 2.0 Prediction Server²⁰, an RNA polymerase II TSS was predicted with a 1.046 score, indicating that a promoter was highly likely to be in this region (Fig. S1). Consistently, the predicted promoter region overlaps with H3K4me3 and H3K4me2 histone modification marks. By sequential deletion analysis of promoter activity, the essential domain of this upstream putative *miR-221/222* promoter was defined (Fig. 1c). We found that the DNA fragment -1393 to +810 (+1 was the predicted TSS) and the

-1674 to +425 fragment exhibited similar strong promoter activity. The promoter activity was slightly reduced in the reporter constructs containing promoter fragments -1310 to +190 and -1033 to +170 and was completely lost in the fragment -190 to +70. These data suggested that the core promoter could be located in the fragment between -1033 and -190. This was consistent with the transcriptional activities assayed by RNA-seq on nine cell lines from ENCODE demonstrating that baseline transcription events started ~23.3 kb upstream of the

miR-221/222 gene and proceeded to the *miR-221/222* gene¹⁹.

Interestingly, several transcripts, named *MIR222HG*s, derived from this *miR-221/222* promoter region were found in the ENCODE database (Fig. 1a). The sizes of the annotated *MIR222HG*s vary in size from ~1 to ~23 kb. A long precursor transcript spanning the entire 23.3 kb from the promoter region to the 3' end of the *miR-221/222* gene was also noted. We hypothesized that this remote upstream promoter may drive transcription of the entire 23 kb locus and the precursor *miR-221/222* and *MIR222HG*s might be generated from the same primary transcript.

***MIR222HG*s are spliced and polyadenylated lncRNAs**

We investigated whether *MIR222HG*-related transcripts are present in LNCaP-Abl, which exhibits active transcription at the 23 kb *miR-221/222* locus. We performed Northern blotting analysis of poly(A)⁺ RNA from LNCaP-Abl and identified two different length transcripts of ~330 and ~500 bp, respectively (Fig. 2a). Using a strand-specific reverse transcription-polymerase chain reaction (RT-PCR) assay, we found that these two different variants were transcribed in the same orientation as the *miR-221/222* gene. By 5'-rapid amplification of cDNA ends (5'RACE) and 3'RACE, we determined the sequence of the two *MIR222HG* isoforms of 332 and 464 bp, respectively (Fig. S1). Nucleotide sequence analysis indicated that the *MIR222HG*-332 bp isoform was spliced from the -464 bp isoform by removing one 132 bp intron. We examined the potential protein-coding capacity of *MIR222HG* sequences. Three potential open reading frames (ORFs), 37 to 80 amino acids in length, were predicted in each of the *MIR222HG* isoforms. These short ORFs lacked the Kozak sequence and had negative coding potential calculator (CPC) scores, implying that the RNAs were CPC classified as non-coding²¹. These results indicated that the 332 and 464 bp *MIR222HG*s identified in LNCaP-Abl are unlikely encoding functional peptides and they are spliced and polyadenylated lncRNAs.

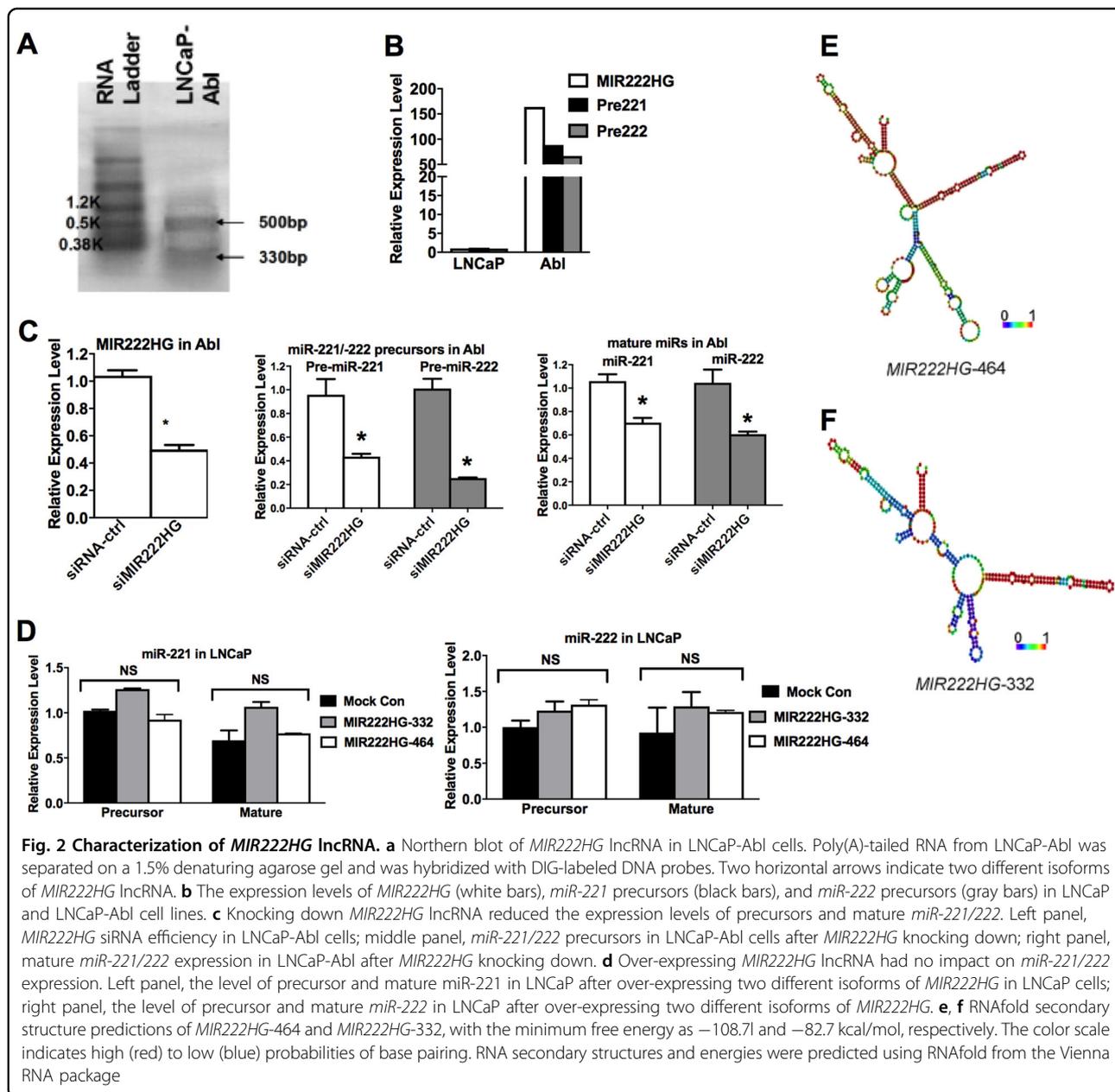
The expression level of *MIR222HG*s in PCa cell lines was determined by strand-specific RT-PCR. Consistently with our H3K4me2 ChIP-seq data, we found *MIR222HG*s' expression was high in LNCaP-Abl, while only barely detectable in LNCaP (Fig. 2b). Since it is known that some promoter associated lncRNAs play important roles in the transcription regulation of genes driven by the same promoter, we further investigated the impact of expression of lncRNA *MIR222HG*s on *miR-221/222* expression. We found that knocking down *MIR222HG*s in LNCaP-Abl significantly down-regulated *miR-221/222* gene expression, especially at the *miR-221/222* precursor level (Fig. 2c). However, over-expressing the two *MIR222HG*s had no impact on the expression of *miR-221/222* (Fig. 2d).

Taken together with the fact that the ENCODE data demonstrated the presence of a long transcript spanning from *MIR222HG* to the *miR-221/222* gene, we concluded that unlike other promoter associated lncRNAs, the *MIR222HG*s most likely were not involved in the transcriptional regulation of the *miR-221/222* locus. We surmised that *MIR222HG* and *miR-221/222* most likely originate from the same polycistronically transcribed primary transcript in LNCaP-Abl. Thus, knocking down the endogenous *MIR222HG*s subsequently led to knock-down of the long primary transcript from which *miR-221/222* was derived, resulting in the reduction of *miR-221/222*. Additional structural analysis using the RNAfold program²² indicated that *MIR222HG*-332 and *MIR222HG*-464 form stable stem-loop structures (Fig. 2e, f), suggesting that *MIR222HG*s potentially function as regulatory RNAs, which prompted us to further investigate their potential roles in PCa progression.

***MIR222HG*s are involved in the CRPC phenotype in PCa cell lines**

To explore the potential biological function of lncRNA *MIR222HG*s in PCa cells, we evaluated the importance of *MIR222HG*s' expression on the development or maintenance of CRPC, since their expression is higher in the CRPC LNCaP-Abl compared to that in the HSPC cell line LNCaP. We determined the impact of changing *MIR222HG* expression on growth efficiency of LNCaP and LNCaP-Abl in the presence or absence of androgen. Over-expressing either *MIR222HG*-332 or *MIR222HG*-464 in LNCaP did not alter cell growth either in the regular medium or the medium with CFBS and DHT, though interestingly it significantly increased the growth of LNCaP in androgen-free CFBS by ~50% at day 7 (Fig. 3a). The enhanced androgen-independent cell growth seen with over-expression of *MIR222HG*s was also observed in another androgen-dependent cell line, LAPC4 (Fig. S2). Since *MIR222HG*s are up-regulated in LNCaP-Abl, we examined the impact of knocking down *MIR222HG* on the growth efficiency of LNCaP-Abl. Knocking down *MIR222HG* significantly reduced cell growth of LNCaP-Abl in CFBS, but did not affect the growth of LNCaP-Abl in either CFBS with DHT or the regular medium (Fig. 3b). Similarly, knocking down *miR-221* expression significantly affected the growth rate of LNCaP-Abl in androgen-free medium, but not in medium containing androgen as shown in Fig. 3b and as previously reported^{16,17}. These data suggested that expression of *MIR222HG*s promotes androgen-independent growth of PCa cells, a hallmark of CRPC.

We further investigated whether AR signaling may be compromised due to the expression of *MIR222HG*s in LNCaP leading to the development of CRPC phenotype, by examining the level of AR-mediated transcription in



response to DHT treatment. The *MIR222HG* level in LNCaP-*MIR222HG*-over-expressing cells was confirmed by RT-PCR and a ~4-fold over-expression of *MIR222HG* was observed (Fig. S9). We examined the expression level of a series of AR-regulated genes including *KLK2*, kallikrein-related peptidase 2 gene; *KLK3*, kallikrein-related peptidase 3 gene (or PSA, a marker for PCa screening); *TMPRSS2*, transmembrane protease serine 2 gene; *FKBP5*, peptidyl-prolyl *cis-trans* isomerase (or FK506 binding protein 5) gene; *PLZF*, promyelocytic leukemia zinc finger protein, which is a putative tumor suppressor gene in PCa; and *cdc20*, cell division cycle 20 homolog. Over-expression of the two isoforms of

MIR222HG in LNCaP significantly reduced the DHT-induced expression of *KLK2*, *KLK3*, *TMPRSS2* mRNA by ~30% to ~70% at 16 and 24 h induction, and *FKBP5* by 50% at 16 h (Fig. 3c–f). However, over-expression of *MIR222HG* did not significantly affect the expression of AR-regulated *PLZF* and *cdc20* (Figs. 3g, h). Meanwhile, control transfections with an empty vector did not affect the response of LNCaP to DHT and over-expression of either the *MIR222HG* isoforms had no significant impact on AR expression in LNCaP (Fig. 3i). These results suggested that expression of *MIR222HG*s can significantly affect AR signaling leading to altered expression of a subset of AR-regulated genes in PCa cells. In sum, two

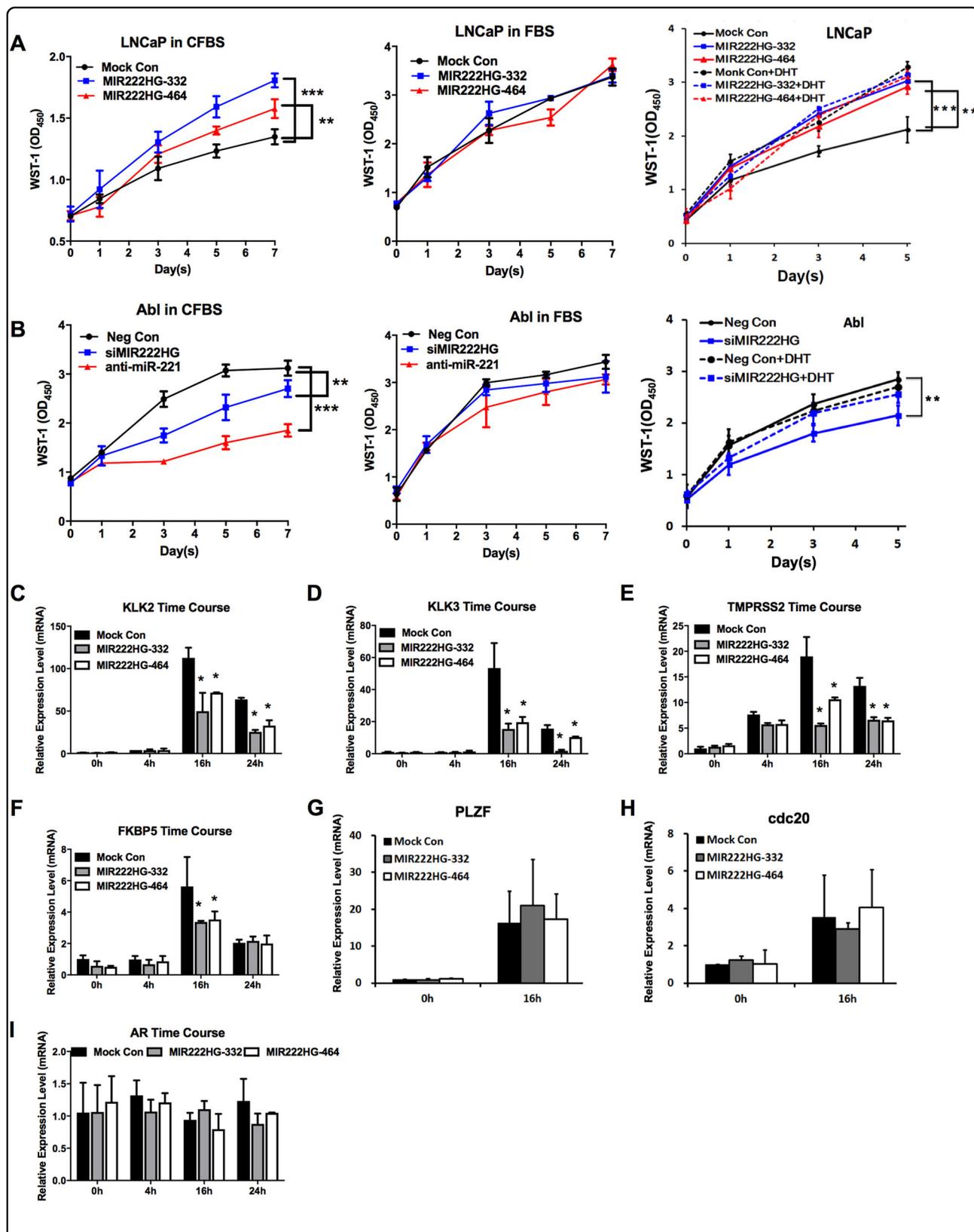


Fig. 3 (See legend on next page.)

Fig. 3 The impact of *MIR222HG* expression levels on CRPC characteristics. **a** Effect of *MIR222HG* expression level on the growth of LNCaP. LNCaP cells that were transfected with the empty vector (Mock Con, black lines), *MIR222HG*-332 bp isoform (blue lines) or 464 bp isoform (red lines) in hormone-free medium (CFBS, left panel), in regular medium (FBS, middle panel), or in CFBS plus 10 nM DHT (right panel); solid lines indicating CFBS only; dotted lines indicating CFBS + DHT. **b** Effect of *MIR222HG* expression level on the growth of LNCaP-Abl. LNCaP-Abl cells that were transfected with the negative siRNA control (black lines), anti-miR-221 (red lines), or *MIR222HG* siRNAs (blue lines) and kept in hormone-free medium (CFBS, left panel), in regular medium (FBS, middle panel), or in CFBS treated with 10 nM DHT (right panel); solid lines indicating CFBS only; dotted lines indicating CFBS + DHT). Cell growth was measured by the WST-1 assay. Triplicate experiments were performed for each set. The data represents mean \pm S.D. ($n = 3$). *The WST-1 fold changes at day 7 after transfection with a p value <0.01 (**) or p value <0.001 (***) in one-way ANOVA. **c-h** The impact of *MIR222HG* expression level on the AR-mediated transcription in response to the DHT treatment. Quantitative analysis of the expression level of *KLK2* (**c**), *KLK3* (**d**), *TMPPSS2* (**e**), *FKBP5* (**f**), *PLZF* (**g**), and *cdc20* (**h**) in LNCaP upon DHT treatment. **i** The AR mRNA expression level in *MIR222HG*-over-expressing cells. The relative expression levels of AR and AR-mediated genes in each sample were normalized with the expression level of GAPDH. Values represent the fold differences relative to those in cells without any drug treatment or transfection (Mock), which were set as 1.0. *The fold changes of those transfected samples compared with their corresponding negative controls show a p value <0.05 in one-way ANOVA

types of ncRNAs, lncRNA *MIR222HG*s and *miR-221/222*, are transcribed from the 23.3 kb *miR-221/222* locus driven by a single differentially activated promoter in LNCaP-Abl and expression of *MIR222HG*s and/or *miR-221/222* can independently promote the CRPC phenotype.

We further investigated whether *MIR222HG*s might be directly involved in the AR complex, thus affecting AR signaling. We found no association of *MIR222HG*s with the AR complex by RNA-IP pull-down assays in *MIR222HG*-over-expressing LNCaP cells under DHT treatment (Fig. S3). Additional ChIP assays comparing the AR occupancy at the *KLK3* and *KLK2* promoter regulatory regions indicated that AR binding to androgen-responsive elements at the *KLK3* or *KLK2* locus was not affected by the expression of *MIR222HG*s in LNCaP (Fig. S4). Furthermore, the efficiency of AR translocation into the nucleus with DHT treatment was not affected in *MIR222HG*-over-expressing LNCaP cells compared to control cells (Fig. S5). These data suggested that the *MIR222HG* over-expression induced down-regulation of some AR-controlled genes most likely is not due to a direct effect on AR, but may be ascribed to its effect on factors associated with the AR machinery, a hypothesis which remains to be investigated.

***MIR222HG*s are associated with the Argonaute complex**

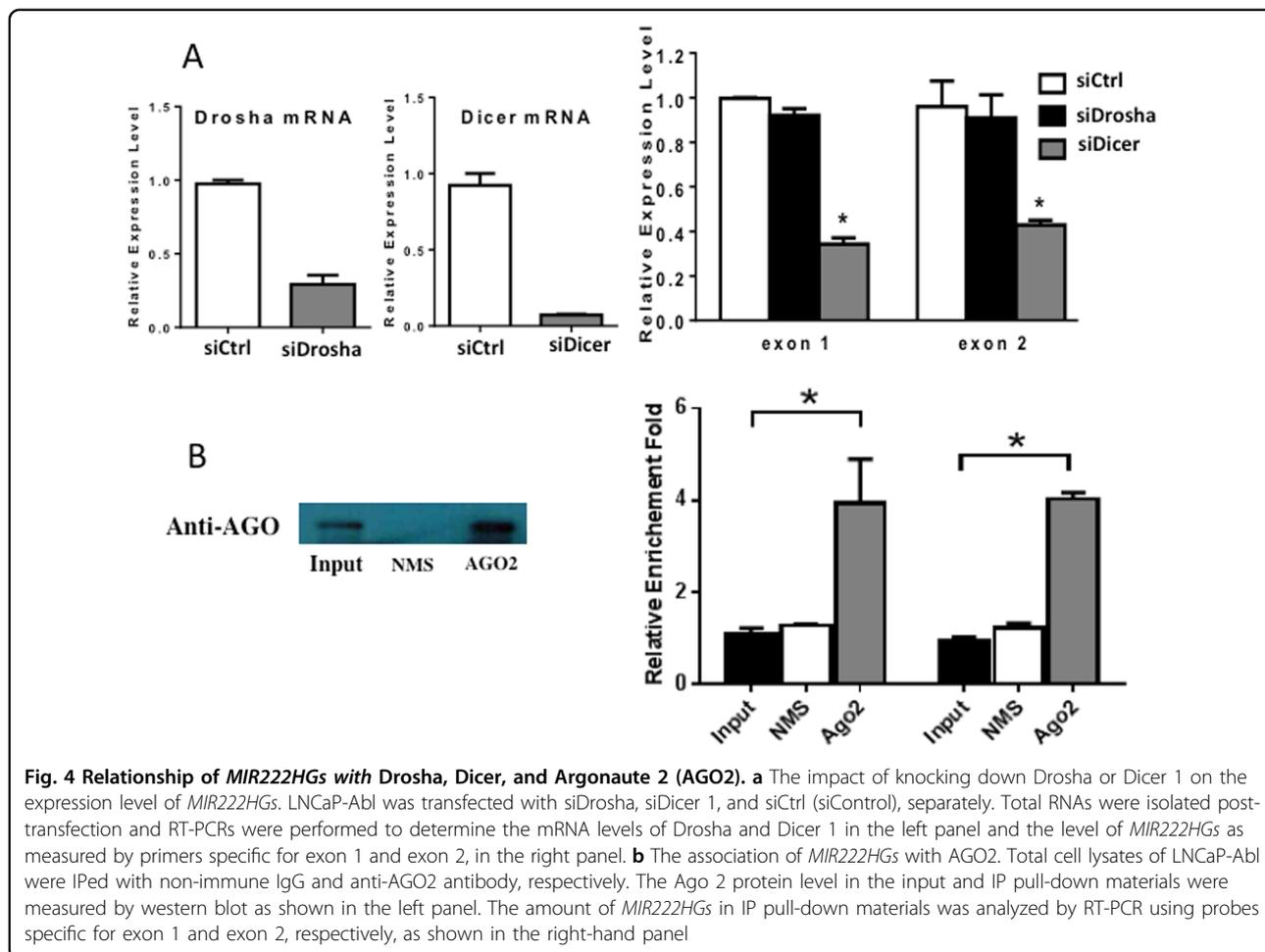
MIR222HG-332 and *MIR222HG*-464 contain regions of strong stable hairpin stem loops (Fig. 2) and potentially function as regulatory RNAs. In view of the structure of *MIR222HG*s and the biological-functional similarity between *MIR222HG*s and *miR-221/222*, we postulated the possibility that *MIR222HG*s may be functionally associated with miR-mediated processes and thus explored the potential association of *MIR222HG*s with the machinery related to the biogenesis and function of miR.

We first determined the impact of knocking down *Drosha* and *Dicer 1* on the expression level of *MIR222HG*s in LNCaP-Abl. *Drosha* is a member of ribonuclease III

and responsible for the initial step of miR-precursor processing in the nucleus. *Dicer 1*, a member of RNaseIII family, cleaves double-stranded RNA and pre-miRs into small double-stranded RNA fragments (small interfering RNA (siRNA)) and miRs in the cytoplasm. Following cleavage, *Dicer* docks the matured miR or siRNA onto the RNA-induced silencing complex (RISC), facilitating the function of the RISC catalytic component, Argonaute (AGO), for RNA interference. Knocking down *Drosha* did not significantly affect the expression level of *MIR222HG*s, as anticipated, as shown by the transcript level of exon 1 and exon 2 (Fig. 4a), since *MIR222HG* is not an miR. Surprisingly, knocking down *Dicer 1* drastically reduced the level of *MIR222HG* (to $<40\%$; Fig. 4a), suggesting that *MIR222HG*s' maturation may occur via *Dicer 1*-associated machinery. We further performed an AGO pull-down assay to determine whether *MIR222HG*s are associated with AGO complex, since *Dicer* docks its processed miRs onto the AGO-associated RISC. As shown in Fig. 4b, the level of *MIR222HG* is significantly enriched in the anti-AGO2 pull-down material (indicated by exon 1 and exon 2), compared to those in the input material and in the material pull-down by the non-immune IgG, indicating the close association of *MIR222HG* with the AGO2 complex. The fact that *MIR222HG* is AGO-bound and its expression level is *Dicer*-dependent suggested that *MIR222HG*s may be functionally involved in the efficiency of miR-mediated RNA interference.

The association of a high-level expression of *MIR222HG*s with the human CRPC

To determine the clinical significance of *MIR222HG* expression, we analyzed the expression levels of the *MIR222HG* in 86 normal prostate tissues, 34 localized HSPC primary tumors, and 17 bone metastatic CRPC samples (Fig. 5a). The expression of *MIR222HG*s was significantly higher in CRPC specimens compared to those in HSPC and normal prostate tissues 2.4-fold and



4.4-fold, respectively (both $p < 0.001$, Fig. 5a). The expression level of *MIR222HG* in HSPC tumors was 1.8-fold higher ($p < 0.01$) than those in normal prostate tissues. Interestingly, previously we found that the expression of miR-221/222 was significantly down-regulated in HSPC primary tumors compared to normal prostate tissues²³. We further performed a Pearson's correlation analysis and found that *MIR222HG* expression only modestly correlated with expression levels of miR-221/222 (Fig. 5b).

Upon further analysis of the expression data from the TCGA data set, we found that the expression of *MIR222HG* and *miR-221/222* were inversely correlated with Gleason score (GS) in HSPC primary tumors ($p < 0.0001$ Kruskal–Wallis test; Fig. 5c). Interestingly, we also observed an inverse correlation of *MIR222HG* expression with AR transcriptional activity, as inferred by the induction of AR target genes and this result was confirmed using two AR scores (Fig. 5d)^{24,25}. A higher AR score was significantly associated with Gleason 6 (Fig. 5e). Additionally, the expression of *MIR222HG* and *miR-221/222* significantly correlated with the degree of copy

number alterations in primary PCa (Fig. S7)²⁴. Expression of *MIR222HG* and *miR-221/222* appears to be inversely correlated with overall copy number burden (fraction of genome altered). Moreover, primary tumors that exhibit the highest burden of copy number alterations (more SCNA cluster)²⁴ show a significantly lower expression ($p < 0.0001$) of *MIR222HG*, *miR-221*, and *miR-222* compared to tumors with the lowest copy number burden (Quiet cluster)²⁴. Taken together, in human PCa, *MIR222HG* expression is correlated with prostate tumor progression and CRPC development, and was in part independent of miR-221/222 expression.

Discussion

Recent advances in genomics revealed that >90% of human genome is transcribed and only a small portion of the genome is translated into proteins, and the remaining is largely transcribed as ncRNAs including miRs and lncRNAs⁷. An abundance of evidence demonstrated that the expression alteration of specific miRs or lncRNAs significantly affect cell developmental processes or cancer development and progression²⁶. However, very little is

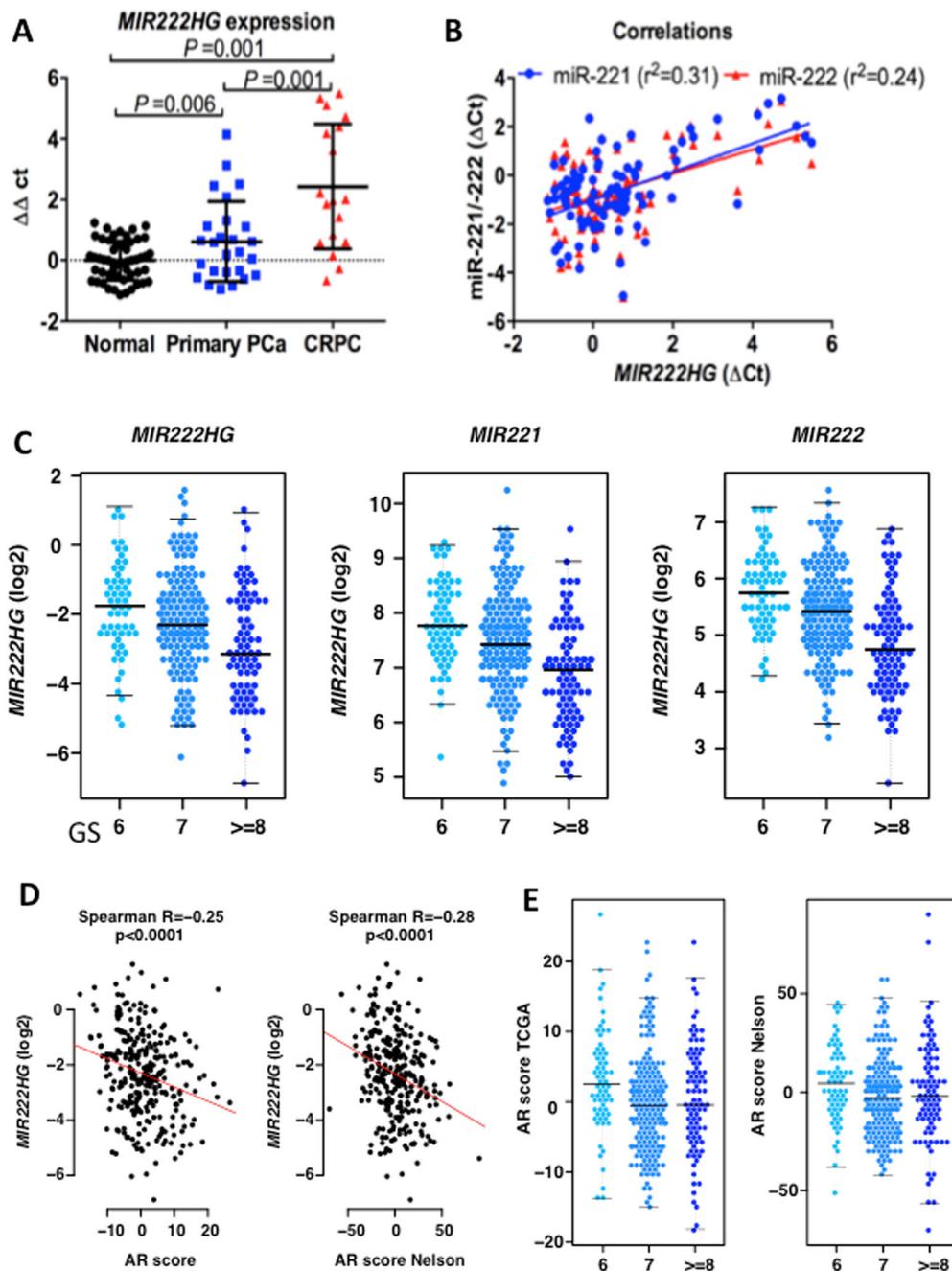


Fig. 5 *MIR222HG* expression pattern in normal prostate, primary tumors, and CRPC. **a** Comparison of the *MIR222HG* expression. Expression levels of *MIR222HG* were measured by quantitative RT-PCR on 86 normal prostate tissues (black dots), 34 hormone-sensitive primary prostate tumors (blue squares), and 17 metastatic CRPC tissues (red triangles). All expression levels were normalized by 28S ribosome RNA (RPS28). Values represent in qPCR cycle differences relative to the mean expression level of normal tissue, which was set as 0. P values were estimated by one-way ANOVA. **b** The expression correlation between *MIR222HG* and *miR-221* or *miR-222* in human prostate tissue specimen. The Pearson's correlation coefficient indexes (r^2) were calculated by comparing normalized qPCR cycle values (ΔCts) in detecting *miR-221* (blue) with ΔCt value of *MIR222HG*. **c** Association of expression levels of *MIR222HG*, *miR-221*, and *miR-222* with GS in primary tumors of the TCGA data set ($p < 0.0001$ Kruskal–Wallis test). **d** Correlation of *MIR222HG* expression level with AR activity (indicated by AR score) in primary tumors of the TCGA data set. AR activity score was inferred by the induction of AR target genes^{24,25}. AR score Nelson indicates that the AR score was re-evaluated using the AR gene index previously published by Dr. Nelson's laboratory²⁵. **e** Correlation of GS with AR activity score

known about mechanisms controlling the differential regulation of most miRs or lncRNAs at different developmental stages in normal or tumor cells.

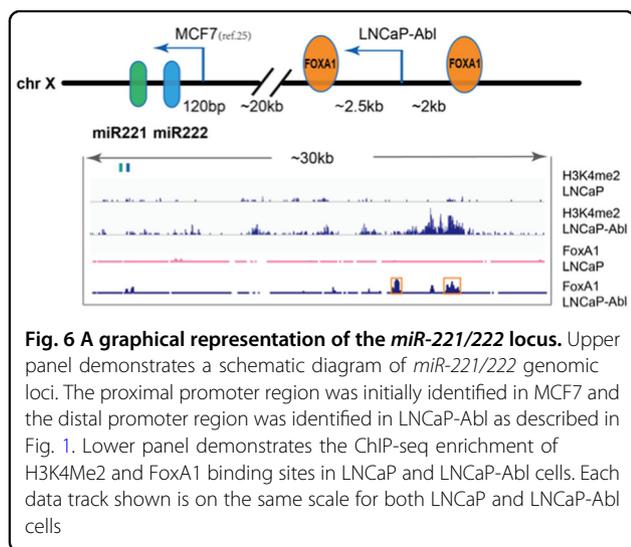
We present here a detailed analysis of the *miR-221/222* gene locus. *MiR-221* and *miR-222* are tandemly linked on the X chromosome and processed from a single pri-miR. *MiR-221/222* are highly expressed in CRPC and their expression promotes CRPC development^{16,17,23}. In examining mechanisms accounting for the differential activation of *miR-221/222* in PCa, especially during the CRPC development, we identified novel lncRNA *MIR222HG*s, derived from the immediate downstream region of a differentially activated promoter located ~23.3 kb upstream of the *miR-221/222* gene. Activation of this remote promoter in CRPC cells leads to a high production of *miR-221/222* and *MIR222HG*s. Our data suggested that *miR-221/222* and *MIR222HG*s are most likely transcribed from a single promoter and originate from the same large primary transcript in LNCaP-Abl. Apparently, in the MCF7 cell line, transcription of *miR-221/222* starts at ~120 bp upstream of *miR-222* and the minimal-promoter region was mapped to a -150 to -50 bp region²⁷. In melanoma cells, the *mir-221/222* promoter was mapped to within 500 bp upstream of the *miR-221/222* gene and could be repressed by PLZF²⁸. We referred to the promoters immediately adjacent to the 5' end of the *miR-221/222* gene as proximal promoters (Fig. 6). Our histone methylation data did not identify the proximal promoters in the *miR-221/222* locus in LNCaP and LNCaP-Abl cell lines. In contrast, in LNCaP-Abl, but not in LNCaP, a putative transcription start site was found ~23 kb upstream of *miR-221/222* gene and we referred to this putative promoter region as a distal promoter (Fig. 6). We hypothesize that the proximal promoters may drive the basal level transcription of *miR-221/222* in HSPC

cells, while the distal promoter functions in the transcriptional activation of *miR-221/222* in LNCaP-Abl CRPC cells. In searching for transcription factors potentially regulating the distal promoter activity, we identified two putative FoxA1 binding sites located ~5 kb region around the distal promoter in LNCaP-Abl using the previously published ChIP-seq data¹⁸. Simply knocking down FoxA1 yielded a mild reduction of the RNA level of *MIR222HG* and *miR-221/222* (Fig. S8), suggesting the potential involvement of FoxA1 in the differential transcription of the *miR-221/222* locus. However, this result remains to be further confirmed by quantitative FOXA1 ChIP analysis.

As mentioned, we identified two lncRNA *MIR222HG* isoforms of 332 and 464 bp, co-transcribed with *miR-221/222* from the distal promoter. Since the entire 23 kb could be transcribed as indicated by the ENCODE data, it is not excluded that other lncRNAs may also be generated from the large primary transcripts via processing events. It is possible that we may have missed some of them due to their low abundance and the abundance of different lncRNAs may vary in different types of cells. Apparently, an angiotensin II-up-regulated lncRNA, *Lnc-Ang362* of 1758 bp, is also derived from the *miR-221/222* distal promoter region in vascular smooth muscle cells (VSMCs)²⁹. Knocking down *Lnc-Ang362* reduced the expression of the co-transcribed *miR-221/222* and *MCM7* in VSMCs, leading to reduced cell proliferation²⁹. However, we did not find that over-expression or knockdown of the *MIR222HG*-332 bp and *MIR222HG*-464 bp had a significant impact on the expression level of *MCM7* in LNCaP and LNCaP-Abl (Fig. S6)

The most surprising observation is that expression of lncRNA *MIR222HG* itself was significantly associated with the development of CRPC phenotype in PCa cell lines and tumor tissues. Over-expression of *MIR222HG* significantly decreased transcription of some DHT-induced AR-mediated genes and promoted androgen-independent growth of LNCaP. CRPC specimens had higher *MIR222HG* lncRNA expression than did primary PCa tissue. Interestingly, increased expression level of *MIR222HG* is associated with lower GS and decreased AR activity score in the primary tumors of the TCGA data set. Assuming that tumors with high AR activity and low Gleason grade respond better to ADT as suggested by the association of GS with time to progression on ADT³⁰, then it is possible that tumors with low AR activity/high GS and high *MIR222HG* may be more resistant to ADT. This scenario is consistent with the finding of a higher level expression of *MIR222HG* in bone-derived mCRPC than that in HSPC localized tumors²³.

As opposed to miRs, lncRNAs can regulate their target genes through diverse mechanisms. Our data demonstrate that the impact of *MIR222HG* expression on promoting



the CRPC phenotype was likely not due to a direct association with AR, but may have resulted from the alteration of factors associated with the AR machinery, a hypothesis remained to be confirmed. The facts that *MIR222HG*s-332/464 are AGO-associated and their expression in LNCaP-Abl is Dicer 1-dependent suggested the possibility of the functional involvement of *MIR222HG*-332/464 in the AGO-associated RISC, which may in turn lead to modulation of AR signaling. As reported that *miR-221/222* promotes androgen independence in PCa cells in part due to down-regulation of HECTD2, a direct target of *miR-221/222*¹⁷. HECTD2 is an essential E3 ligase involved in the degradation of PIAS1, which is a potent anti-inflammatory protein and mediates the cross-talk between nuclear factor- κ B and AR signaling^{31,32}. We are currently investigating the potential involvement of *MIR222HG*s in miR function and the role of Dicer 1 in the biogenesis of *MIR222HG*s.

In summary, the *miR-221/222* locus extends ~23 kb, activation of the distal promoter leads to the transcription of large primary transcripts in CRPC cells, which are further processed to generate at least two different types of ncRNAs, the *miR-221/222* and the lncRNA *MIR222HG*s. Both the expression of *miR-221/222* and *MIR222HG*s can significantly promote the CRPC phenotype. Further studies are needed to address the specific role of this newly identified lncRNA *MIR222HG*s in AR-mediated transcription regulation and CRPC development. Nevertheless, this study reveals a new potential function of lncRNAs as a modulator of AGO-mediated RISC.

Materials and methods

Cell culture

LNCaP was obtained from ATCC and maintained in RPMI-1640 with 10% FBS. LNCaP-Abl was provided by Zoran Culig (Innsbruck Medical University, Austria)³³ and maintained in RPMI-1640 with 10% CFBS¹⁶. All cells were regularly screened for mycoplasma using a Venor GeM Mycoplasma Detection Kit (Sigma), passaged for fewer than 4 months and used at early passages (<20).

Quantitative RT-PCR

For *MIR222HG* expression, strand-specific cDNAs were generated by RT and then subjected to quantitative RT-PCR (qRT-PCR) analysis by FastStart Universal SYBR Green Master (Roche, IN, USA). *MiR-221/222* levels were quantitated by qRT-PCR, using TaqMan miR assays from Applied Biosystem¹⁶. Real-time PCR primers/probe sets for *KLK3*, *KLK2*, *TMPRSS2*, *FKBP5*, *AR*, *Drosha*, *Dicer 1*, and *GAPDH* were inventoried products of Applied Biosystem.

Northern blot analyses

Poly(A)+ RNAs were isolated by Invitrogen Dynabeads mRNA Purification Kit (Carlsbad, CA, USA). About 100 ng poly(A)+ RNA was fractionated in 1.5% (w/v) denaturing agarose gel and transferred onto nylon membrane, which was subsequently hybridized with digoxigenin (DIG)-labeled DNA probes and detected by DIG Detection Kit (Roche, IN, USA).

Rapid amplification of cDNA ends

5'RACE and 3'RACE were performed using Invitrogen Kits (Carlsbad, CA, USA). For 5'RACE, ~10 μ g of total RNA from LNCaP-Abl was sequentially treated with Calf Intestine Phosphatase and Tobacco Acid Pyrophosphatase prior to ligation with adaptor RNAs. The ligation products were subjected to RT-nested PCR. For 3'RACE, ~3 μ g total RNA was treated by 10 U poly(A) polymerase for 30 min at 37 °C. The poly(A)-tailed RNAs were amplified by RT-nested PCR. Primer information is in Supplementary Table 1. The RACE products were analyzed by 1% agarose gel and sequenced.

Dual-luciferase reporter assay

The different sized DNA fragments were cloned into pGL4 promoter luciferase reporter vector (Promega, WI, USA). LNCaP-Abl was seeded at a density of 2×10^5 cells per 12-well dish. Twenty-four hours later, 400 ng of each construct was co-transfected with 20 ng of pRL-TK as transfection efficiency control, using Lipofectamine 2000. Forty-eight hours after transfection, luciferase activity was measured by Dual-Luciferase Reporter Assay Kit (Promega).

MIR222HG silencing and over-expression

Customized siRNA SmartPool targeting *MIR222HG*s was synthesized by GE Dharmacon (Lafayette, CO, USA) and their sequences were included in Supplementary Table 1. LNCaP-Abl was seeded into 6-well plates and transfected with 30 nM of SmartPool siRNA using Lipofectamine RNAiMax Reagent (Invitrogen). cDNAs of *MIR222HG*s were cloned into pcDNA3.1(+) vector (Invitrogen) and confirmed by sequencing. LNCaP was transfected with *MIR222HG*-over-expressing vectors and selected by 500 μ g/mL geneticin (G418, Invitrogen).

Cell proliferation assay

For all cell growth studies, cells were plated in 96-well plates. Twenty-four hours after seeding, cells were transfected with anti-miR-221 or *MIR222HG* siRNAs. Tetrazolium salt cell proliferation assay (WST-1, Roche Applied Science) was used to determine proliferation.

Ago RIP-CHIP analysis

Ago RIP-CHIP was performed as previously described¹⁷. Briefly, 10⁷ LNCaP-Abl were lysed. Cell lysates were collected, cleared with pre-blocked protein-G beads (Invitrogen), and used for co-IP with either anti-AGO G beads, NMS G beads (Pierce Biotechnology) at 4 °C for 90 min. After co-IP, the beads were washed and treated with DNase I. Co-immunoprecipitated RNAs were extracted using TRIzol (Invitrogen). The associated *MIR222HG* transcripts were measured by RT-PCR. The AGO2 (C34C6) Rabbit-mAb #2897 is from Cell Signaling Technology.

Prostate tissue or tumor-derived RNAs

Total RNAs isolated from primary PCa tumor tissues, their adjacent normal tissues, and bone marrow biopsies were as previously described²³.

Bioinformatic analyses

Genome-wide histone modification data from multiple cell lines was derived from ENCODE HG19 data¹⁹. The ChIP-seq data sets of LNCaP and LNCaP-Abl were as previously described^{18,34}. ORF Finder from NCBI (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) and CPC (<http://cpc.cbi.pku.edu.cn>)²¹ were used to define the non-coding potential. RNA structure was predicted using RNAfold from the Vienna RNA package²².

Statistical analysis

A two-tailed *t* test was used for statistical *p* value analysis. Error bars represent standard deviation of three independent experiments. The Pearson's correlation coefficient was used to determine the expression correlation between *miR-221/222* and *MIR222HG*s in tumors tissues. Spearman's correlation coefficient was used to determine the correlation among *miR-221/222* and *MIR222HG* expression in the TCGA data set. The significance of expression association of *MIR222HG*, *miR-221*, *miR-222*, and *GS* in the TCGA data set was evaluated using Kruskal–Wallis test. All plots and statistical analysis were performed in R (version 3.3.1). Fraction of genome altered and copy number clustering data are available at <http://www.cbioportal.org/>.

Acknowledgements

We thank Dr. Myles Brown's lab for sharing ChIP-seq information. This study was supported by the Dana-Farber PCa SPORE P50CA090381, the NIH/NCI Cancer Center Support Grant P30 CA008748 and the DOD-PCa Training Award (W81XWH-09-1-0372) to T.S.

Author details

¹Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, 450 Brookline Ave, Boston, MA 02215, USA. ²Department of Medicine, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA. ³Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ⁴Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41389-018-0039-5>).

Received: 5 October 2017 Revised: 22 January 2018 Accepted: 6 February 2018

Published online: 13 March 2018

References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J. Clin.* **65**, 5–29 (2015).
- Heinlein, C. A. & Chang, C. Androgen receptor in PCa. *Endocr. Rev.* **25**, 276–308 (2004).
- Harris, W. P., Mostaghel, E. A., Nelson, P. S. & Montgomery, B. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nat. Clin. Pract. Urol.* **6**, 76–85 (2009).
- Yuan, X. et al. Androgen receptor functions in castration-resistant PCa and mechanisms of resistance to new agents targeting the androgen axis. *Oncogene* **33**, 2815–2825 (2014).
- Bluemn, E. G. & Nelson, P. S. The androgen/androgen receptor axis in PCa. *Curr. Opin. Oncol.* **24**, 251–257 (2012).
- Karantanos, T., Corn, P. G. & Thompson, T. C. PCa progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene* **32**, 5501–5511 (2013).
- Djebali, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
- Cech, T. R. & Steitz, J. A. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
- Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504 (2009).
- Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159 (2009).
- Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. *Cancer Discov.* **1**, 391–407 (2011).
- Sun, T., Ye, H., Wu, C.-L., Lee, G.-S. M. & Kantoff, P. W. Emerging players in PCa: long non-coding RNAs. *Am. J. Clin. Exp. Urol.* **2**, 294–299 (2014).
- Prensner, J. R. et al. RNA biomarkers associated with metastatic progression in PCa: a multi-institutional high-throughput analysis of SChLAP1. *Lancet Oncol.* **15**, 1469–1480 (2014).
- Prensner, J. R. et al. The long noncoding RNA SChLAP1 promotes aggressive PCa and antagonizes the SWI/SNF complex. *Nat. Genet.* **45**, 1392–1398 (2013).
- Du, Z. et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in PCa. *Nat. Commun.* **7**, 10982 (2016).
- Sun, T. et al. The role of microRNA-221 and microRNA-222 in androgen-independent PCa cell lines. *Cancer Res.* **69**, 3356–3363 (2009).
- Sun, T. et al. MiR-221 promotes the development of androgen independence in PCa cells via downregulation of HECTD2 and RAB1A. *Oncogene* **33**, 2790–2800 (2014).
- He, H. H. et al. Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* **42**, 343–347 (2010).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Knudsen, S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* **15**, 356–361 (1999).
- Kong, L. et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349 (2007).
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA website. *Nucleic Acids Res.* **36**, W70–W74 (2008).

23. Sun, T. et al. The altered expression of MiR-221/-222 and MiR-23b/-27b is associated with the development of human castration resistant PCa. *Prostate* **72**, 1093–1103 (2012).
24. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
25. Nelson, P. S. et al. The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc. Natl. Acad. Sci. USA* **99**, 11890–11895 (2002).
26. Di Gesualdo, F., Capaccioli, S. & Lulli, M. A pathophysiological view of the long non-coding RNA world. *Oncotarget* **5**, 10976–10996 (2014).
27. Di Leva, G. et al. MicroRNA cluster 221–222 and estrogen receptor alpha interactions in breast cancer. *J. Natl. Cancer Inst.* **102**, 706–721 (2010).
28. Felicetti, F. et al. The promyelocytic leukemia zinc finger-microRNA-221/-222 pathway controls melanoma progression through multiple oncogenic mechanisms. *Cancer Res.* **68**, 2745–2754 (2008).
29. Leung, A. et al. Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. *Circ. Res.* **113**, 266–278 (2013).
30. Ross, R. W. et al. Efficacy of androgen deprivation therapy (ADT) in patients with advanced PCa: association between GS, prostate-specific antigen level, and prior ADT exposure with duration of ADT effect. *Cancer* **112**, 1247–1253 (2008).
31. Coon, T. A. et al. The proinflammatory role of HECTD2 in innate immunity and experimental lung injury. *Sci. Transl. Med.* **7**, 295ra109 (2015).
32. Malinen, M., Niskanen, E. A., Kaikkonen, M. U. & Palvimo, J. J. Crosstalk between androgen and pro-inflammatory signaling remodels androgen receptor and NF- κ B cistrome to reprogram the PCa cell transcriptome. *Nucleic Acids Res.* **45**, 619–630 (2017).
33. Culig, Z. 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* **81**, 242–251 (1999).
34. Wang, Q. et al. Androgen receptor regulates a distinct transcription program in androgen-independent PCa. *Cell* **138**, 245–256 (2009).