Discovery and Characterization of Non-Coding RNAs with Therapeutic and Diagnostic Potential in Prostate Cancer

by

Yajia Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in the University of Michigan
2019

Doctoral Committee:

Professor Arul M. Chinnaiyan, Chair
Professor Gregory R. Dressler
Professor Elizabeth R. Lawlor
Associate Professor Scott A. Tomlins

Yajia Zhang

yajiazh@umich.edu

ORCID iD: 0000-0002-5907-4253

To mom and dad,

thank you for your unconditional love and unwavering support,

despite the distance between us

ACKNOWLEDGMENTS

Graduate school has been an unforgettable experience in my life. There were many exciting moments when I enjoyed conducting research and making scientific contributions, as well as some difficult times, when I was faced with unexpected challenges and frustration. I am fortunate to have gotten help from many individuals during this process. I would like to take this opportunity to express my gratitude for their guidance and support.

First and foremost, I would like express my deepest appreciation to my mentor, Dr. Arul M. Chinnaiyan. He guided me through my thesis study and provided a nourishing environment for me to grow as scientist. He taught me ways of scientific thinking, and equipped me with research and professional skills essential for my career path. Thank you for pushing me to strive for my personal best in research, and allowing me to pursue my interest in bioinformatics. It has truly been a privilege to have Arul as my graduate mentor.

I would also like to thank members of my thesis committee, Prof.s Gregory R. Dressler, Elizabeth R. Lawlor, and Scott A. Tomlins. I am grateful for their helpful suggestions and feedback throughout the course of my education. I would especially like to thank Scott for guiding me into the field of cancer genomic research during my first rotation, and offering

incredible support throughout my graduate training. I would like to thank Beth and Greg for their constructive input and scientific insights that helped me advance my thesis work.

I am extremely grateful to many colleagues who have contributed to this study. The work would have not been possible without their input. Rohit Malik provided guidance and taught me how to conduct research at the bench. Marcin Cieslik taught me most of what I know in the bioinformatics field. I have been fortunate to work with many collaborators with expertise: Sethuramasundaram Pitchiaya, Josh Vo, Lanbo Xiao, Yasuyuki Hosono, Jean C.-Y. Tien, Sudhanshu K. Shukla, Yashar S. Niknafs, Yuanyuan Qiao, Michelle T. Paulsen, Rohit Mehra, Lisha Wang, Mats Ljungman. I am also indebted to the many collaborators who have let me contribute to their research and value my opinions: Weiping Zou, Rolf Backofen, Cenk S. Sahinalp. These opportunities have deepened my scientific horizon and widened my training experiences.

I am thankful to all current and previous members of Chinnaiyan lab: Ron Siebenaler, Abhijit Parolia, Jae Eun Choi, Yuping Zhang, Mohan Dhanasekaran, Carl Engelke, Xiaoming Wang, Steve Kregel, Dan Robinson, Yi-Mi Wu, Sunita Shankar, Chandan Kumar, Seema Chugh, Sathiya Narayanan, Palak Shah, Fengyun Su, Rui Wang, Apel Ingrid, Jin Chen, Alec Chu, Xiaoxuan Dang, Jamie Estill, Pankaj Vats, Alexander Hopkins, Brendan Veeneman, Sudhanshu Shukla, and Anirban Sahu. It has been a great pleasure working with these talented colleagues, and sharing up and downs in our lives.

Additionally, I would like to extend my appreciation to everyone who has provided administrative support at the Michigan Center for Translational Pathology and Molecular and Cellular Pathology Program: Christine Betts, Dianna Banka, Sisi Gao, Stephanie Ellison, Jyoti Athanikar, Jamie Young, Zaneta Nikolovska-Coleska, and Laura Labut. Your timely support has largely smoothed my path of graduate training. Special thanks goes to Xuhong Cao. Thank you for coordinating everything in lab and making all kinds of works possible.

Finally, no words can adequately express my gratitude towards my parents. Thank you for encouraging me to dream big and inspiring me to follow my dreams. Despite the long distance between us, your unconditional love and unwavering support has lifted my spirits at many moments of this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Prostate cancer is the most common malignancy and the second leading cause of death in American men. Primary prostate cancer is often hormone-dependent and relies on signaling through the androgen receptor (AR); therefore, the majority of patients are responsive to front-line treatment with androgen-deprivation therapy. However, some cases progress to an incurable stage of the disease known as castration-resistant prostate cancer (CRPC), which still relies on AR signaling.

Substantial efforts have been undertaken to study sustained AR signaling in CRPC. The mechanisms identified include AR amplifications, AR splice variants, alterations in cofactor recruitment, and ligand-independent activation via crosstalk with signal transduction pathways. However, few studies have investigated the roles of non-coding RNAs (ncRNAs) in AR signaling axis. Noncoding RNAs are a class of transcripts with diverse and largely uncharacterized biological functions. Through crosstalk with chromatin, DNA, RNA, and proteins, ncRNAs function via chromatin remodeling as well as transcriptional and post-transcriptional regulation.

To discover long non-coding RNAs (lncRNAs) with therapeutic or diagnostic potential in prostate cancer, we have performed a comprehensive RNA-seq profile investigation of cancer-associated, AR-regulated lncRNAs from prostate cancer cell lines and patient tissue samples. Through this analysis, we identified ARLNC1 (AR-regulated long noncoding RNA 1), an important lncRNA that is significantly enriched in prostate lineage and associates with prostate cancer progression.

ARLNC1 is directly induced by the AR protein and modulates AR signaling. ARLNC1 knockdown suppresses AR expression, AR-responsive genes, and prostate cancer growth *in vitro* and *in vivo*. Further dissection of ARLNC1 cellular mechanism showed that the lncRNA stabilized the AR transcript via RNA-RNA interaction. This positive feedback loop, where AR transcriptionally regulates ARLNC1 and ARLNC1 regulates its master regulator (AR) post-transcriptionally, suggests that lncRNAs can serve as novel regulatory nodes on AR signaling pathway.

Because modulation of ARLNC1 expression levels resulted in a striking proliferation phenotype, it is hypothesized that ARLNC1 inhibition could be used therapeutically for the treatment of prostate cancer. Supporting this hypothesis, antisense oligonucleotides (ASOs) targeting ARLNC1 inhibited prostate cancer cell growth *in vitro* and *in vivo*. The fact that ARLNC1 acts upstream of AR signaling presents the possibility that targeting ARLNC1 may afford an additional option to patients that have *de novo* or acquired resistance to therapies targeting AR itself.

In addition, we evaluated the clinical relevance of ARLNC1 as a diagnostic biomarker. Among the prostate samples, ARLNC1 expression was significantly higher in localized and metastatic prostate cancers than in benign tissues, as assessed by RNA-seq and RNA *in situ* hybridization. Moreover, there exists an association of ARLNC1 levels with accentuated AR signaling and luminal epithelial differentiation in patient tumors, both of which are important clinical considerations for anti-androgen treatment.

Finally, we characterized expression patterns of another type of non-coding RNA, circular RNA (circRNA). A set of circRNAs were identified as candidate biomarkers for prostate cancer. They have higher stability compared to linear transcripts. We further validated the existence of a novel class of circular transcripts, termed read-through circRNAs, that involve exons originating from different genes. Additionally, circRNAs can be detected in patient urine samples and may be developed into non-invasive biomarkers.

Taken together, this work defines essential roles for ARLNC1 in prostate cancer progression, uncovers novel aspects of AR signaling and lncRNA biology, and underscores the clinical potential of non-coding RNAs as diagnostic markers or therapeutic targets in prostate cancer.

Chapter 1

Introduction: Non-coding RNAs and prostate cancer biology

Prostate cancer

Prostate cancer statistics

Prostate cancer is the most commonly diagnosed cancer in American men, accounting for nearly 1 case in 5 new cancer diagnoses. It is estimated that 174,650 new cases will be diagnosed in 2019[1]. Compared with other types of cancer, prostate cancer has the highest 5-year relative survival rates, which is 98%. The survival rate differs significantly across cancer stages. Localized and regional prostate cancer has nearly a 100% 5-year survival rate, while distant metastasis has a poor survival rate of 30%. As such, it is vitally important to discriminate dormant prostate cancer from aggressive disease, and to understand the molecular pathogenesis of metastatic prostate cancer. The mortality of prostate cancer has dropped steeply by ~50% from 1993 to 2016[1]. These reductions result from efforts to detect the disease at early stages and advances in treatments[2, 3]. Despite the substantial drop in mortality, prostate cancer is still the second leading cause of cancer-related deaths in American men, with estimated deaths of 31,620 cases in 2019[1].

The biggest risk factor for developing prostate cancer is age. Risk for prostate cancer increases with age, especially after age 50. More than 80% of prostate cancers are diagnosed in men beyond 65 years old, and less than 2% of cases are diagnosed in men under the age of 40 (SEER Cancer Statistics Review 1975-2015). Ethnicity is another risk factor. Compared to Caucasian men, African-American men are 76% more likely to acquire prostate cancer, and more than twice as likely to die from the disease[4]. Meanwhile, Hispanic or Asian men have a lower risk of developing and dying from the disease. There is also inherited risk for prostate cancer, due to a combination of shared genes (such as germline mutations on BRCA2) and shared environmental or lifestyle factors. Hereditary prostate cancer accounts for 5 to 10 percent of all prostate cancer diagnoses. A man with a close relative diagnosed with prostate cancer (especially at an early age) is approximately twice as likely to develop prostate cancer compared to a man with no family history of the disease[5-7].

Understanding prostate cancer genome

The majority of patients with primary prostate cancer can be cured, while some men develop metastatic disease and receive primary Androgen Deprivation Therapy (ADT). However, nearly all men with metastatic disease eventually develop resistance to primary ADT, reaching a disease state named metastatic castration-resistant prostate cancer (mCRPC). This advanced disease is treated by second generation ADT agents, such as

abiraterone acetate[8, 9] and enzalutamide[10, 11]. However, mCRPC patients will eventually develop resistance to these drugs as well (**Figure 1.1**).

Advances of large-scale sequencing have largely reshaped our understanding of cancer. Pan-cancer genomic analysis uncovered the causative roles of driving mutations/copy number variations/gene fusion or rearrangement in tumorigenesis[12, 13]. In the past five years, a significant part of the sequencing efforts has shifted from profiling primary cancers to metastatic disease. This is because advanced cancers are more likely to be lethal, and the clinical genomics of advanced disease produce altered patterns compared to that of primary cancer[14].

In primary prostate cancer, past studies have identified recurrent mutations, gene rearrangements, deletions of tumor suppressive genes, and amplifications of oncogenes [15-22]. Frequent point mutations were observed in *TP53* and *SPOP*, while low levels of mutations occurred on *FOXA1*, *ASHXL*, *MLL2*, and *BRAF*. E26 transformation-specific [23] fusions, which originate from the 5' regions of an androgen-responsive gene plus a member of the ETS transcription factor family, can be detected in half of prostate cancer cases. Tumor suppressive genes *PTEN*, *TP53*, and *RB1* are commonly deleted, while oncogenes like *MYC* and *PIK3CA* are amplified. Despite the comprehensive understanding of genomic aberrations, most of these results have not been translated into therapeutic impact.

Recent studies on individuals affected with mCRPC uncovered novel alterations that are enriched in mCRPC, and highlighted clinically actionable molecular alterations in specific pathways (**Figure 1.2**)[24]. Most of genetic alterations identified from primary prostate cancer also exist in mCRPC, including deletions of *TP53*, *PTEN*, *RB1*, and *CHD1*, as well as amplifications of *AR*, *PIK3CA*, *PIK3CB*, and *CCND1*.

Of note, a few genes were selectively mutated at a higher rate in mCRPC: *TP53*, *AR*, *KMT2D*, *APC*, *BRCA2*, and *GNAS*. Interestingly, the androgen receptor (AR) and GNAS were exclusively mutated in mCRPC[24]. This study highlighted the aberrations in the AR signaling pathway in mCRPC[24]: A total of 71.3% of all cases studied harbored alterations in the AR pathway. The majority (63%) of them directly affects AR (by mutations or amplifications), while the rest of the cases have aberrations of the AR-cofactors *FOXA1*, *NCOR1/2*, and *ZBTB16*. Additional deregulated pathways enriched in mCRPC include PI3K pathway, Wnt pathway, Cell-cycle pathway, and DNA repair pathway.

Central role of Androgen Receptor signaling in prostate cancer

Androgen Receptor signaling plays a pivotal role in the development of normal prostate and prostate cancer[25]. It is physiologically regulated by hypothalamic-pituitary-testicular axis, adrenal gland steroidogenesis, and prostate cell intrinsic factors (**Figure 1.3**)[25]. The androgen receptor belongs to the steroid hormone group of nuclear receptors, which is translocated to the nucleus upon ligand binding and controls the

expression of specific genes. At AR-responsive elements, AR recruits coactivator complexes that open chromatin structure and recruit RNA polymerase II to initiate transcription[26]. Additionally, AR can act through binding to other transcription factors[27] or alternative steroid receptors[28]. In the normal prostate, AR mediates differentiation of luminal epithelial cells and regulates genes that are essential for prostate function. In prostate cancer, sustained AR signaling modulates cell cycle regulation, survival signaling, and contribute to cell growth[29].

As such, androgen signaling blockade by inhibiting androgen production or components in AR signaling axis largely inhibits prostate cancer growth. Androgen deprivation therapy (ADT) is initially highly effective for treatment of locally advanced or metastatic disease. However, the clinical course is uniformly marked by progression to the metastatic castration resistant disease state (mCRPC), where residual androgens that remain after castration and AR itself continue to be crucial for the progression to CRPC and its continued growth. This sustained AR signaling is achieved through mechanisms of AR mutations [24, 30], AR amplifications[24], AR splice variants[31-33], somatically acquired enhancer of AR[34], alterations in cofactor recruitment[23, 35], and ligand-independent activation via cross-talk with signal transduction pathways[36, 37]. These findings highlight the rationale of the search for alternative nodes on AR signaling axis to block AR signaling.

The non-coding transcriptome in cancer

Overview of non-coding RNA species

RNA molecules were initially identified as a node in the "central dogma", that carries genetic information from DNA to the functional product of genetic information, protein[38, 39]. For many years, the only reported "non-coding" forms of RNA were ribosomal RNAs, which are components of the ribosome, and tRNAs, which are translators of codon sequences[40].

Two decades ago, pioneering works in *C. elegans* (*lin-4* and *let-7*) demonstrated for the first time that some small RNAs, though without protein-coding potential, are functional regulators for development[41, 42]. Since then, advances in next-generation sequencing have enabled the comprehensive identification and annotation of non-coding RNAs (ncRNAs) from various normal and cancer tissues[43-46]. Aside from the ~2% protein-coding genes within human genome, non-coding RNAs constitute ~60% of total transcriptional output[44, 45, 47]. Increasing evidence has demonstrated that ncRNAs serve as regulators for a multitude of cellular processes and pathways under developmental and pathological contexts.

Non-coding RNAs can be classified by size into three classes[48]: (a) Short ncRNAs with 18-31 nucleotides in length (e.g. microRNAs and PIWI-interacting RNAs), (b) mid-size ncRNAs with 20~200 nucleotides (e.g. small nucleolar RNAs and tRNA derivatives), and (c) long non-coding RNAs with >200 nucleotides (e.g. long intergenic long non-coding RNAs, pseudogenes, circularized RNAs).

microRNAs in cancer transcriptome

The most extensively studied type of ncRNAs are miRNAs. They are 18 to 24 nucleotides in length, and regulate target gene expression by mRNA degradation or translational repression. Primary miRNAs are initially transcribed by RNA Polymerase II (Pol II), then sequentially processed by Dgcr8/Drosha (microprocessor complex) and Dicer (riboendonuclease) to yield mature miRNA duplexes[42, 49-51]. Following the formation of the RNA-induced silencing complex (RISC), miRNAs recognize the 3' untranslated regions (3' UTR) on mRNA targets via imperfect base pairing, mediating post-transcriptional repression. The functional relevance of miRNAs in cancer was first suggested by the frequent genomic deletion or amplification observed at genomic sites coding for miRNAs in certain types of cancer[52, 53]. Since then, deregulation of miRNAs and miRNAs possessing oncogenic or tumor-suppressive functions have been reported[54-60]. miRNA-mediated feedback or feedforward loops play an integral role in gene regulatory networks[61-66].

Long non-coding RNA in cancer transcriptome

Long non-coding RNAs are capped, spliced, polyadenylated RNA transcripts ranging from 200 to several kilobases in length[67]. Transcriptional start sites (TSSs) of lncRNAs are enriched for histone 3 lysine 4 trimethylation (H3K4me3) and Pol II binding sites, both of which are markers for active promoters. The catalog of lncRNAs has grown tremendously over the past decade, largely due to the development of new computational

methods for transcriptome assembly and lncRNA annotation[44-46]. Relative locations of lncRNAs in the genome can be within intergenic regions, antisense of protein-coding gene regions[68], active enhancers[69], or promoter regions[70].

In order to comprehensively profile human lncRNAs in the context of normal tissues and cancers, our group performed *ab initio* analysis on 7,256 RNA-sequencing libraries derived from normal tissues, primary cancers, metastasis diseases, and cell lines[46]. A total of 58,648 lncRNAs were identified in the human transcriptome. Several features of lncRNAs were discovered that are in accord with other published studies: (1) LncRNAs outnumber protein-coding genes, but with a much lower average expression level. With RNA-sequencing analysis, RPKM values of protein coding genes are more than 10 fold as high as that of the majority of lncRNAs. (2) Overall, the conservation scores of lncRNAs are not as high as protein-coding genes, although they are higher than random control regions[71]. However, there does exist a small set of intergenic lncRNAs harboring ultraconserved DNA elements that are more than 200nt in length with nearly perfect sequence identity across multiple organisms. These highly conserved lncRNAs are of great research interest and may contribute to development and oncogenic processes[72]. Some lncRNAs overlap with disease-associated SNPs identified from Genome-wide association (GWAS) studies[73, 74]. This regional proximity suggests possible regulation of SNPs on allele-specific gene expression at certain genomic hotspot regions critical to cancer progression. (4) Differential expression analysis has uncovered disease associated, lineage-associated lncRNAs. Clustering of top cancer-specific

lncRNAs demonstrated unique signatures for each cancer type. These lncRNAs can be further evaluated as potential pathological biomolecules.

Circular RNAs in cancer transcriptome

Circular RNAs are a group of single-stranded RNA molecules produced from pre-mRNAs through the back-splicing process. They form covalently closed circles and thus have different chemical-physical features compared to linear RNA transcripts (for instance, resistant to exoribonuclease digestion). Only a handful of circRNAs have been studied in cancer. They have implicated roles in tumorigenesis, epithelial-mesenchymal transition, and gene regulation[75-82]. Studies in colorectal cancer unraveled the sponging effect of circular RNA ciRS-7 on the tumor suppressor miR-7[82]. Cancer-associated chromosomal translocations may generate fusion-circRNAs with oncogenic potential, as exemplified by the f-circM9-accelerated tumorigenesis in MLL-AF9 driven acute myeloid leukemia models [83].

Expression of some circular RNAs are tissue-specific, and can be found in cells, exosomes, cell-free saliva, and plasma. The spatial-temporal specific expression pattern, plus the existence in cell-free context, make circular RNAs potential biomarkers in cancer. It is thus worth investigating which circular RNAs are specifically enriched in cancers originating from varied lineages, and which are detectable in cell-free specimens.

Non-coding RNAs with diagnostic and therapeutic potential in cancer

Non-coding RNAs as potential biomarkers

Cancer genome and transcriptome analyses have revealed an extensive landscape of alterations in the non-coding genome. Regardless of the functional relevance of these alterations, characterization of the expression of alterations can significantly contribute to the development of biomarkers. Compared to normal physiological states, recurrent somatic mutations, copy-number variations, and epigenetic alterations in cancer have defined non-coding RNA expression profiles unique to pathological states. These findings highlight the opportunities for exploration of non-coding RNAs as biomarkers.

Through the differential expression analysis of RNA-Seq from normal samples and tumor samples from the same tissue, researchers have identified lncRNA biomarkers that can stratify normal tissues from cancers tissues (*eg.* PCAT1, CCAT2, HOTAIR, PCA3, lncRNA-AA174084). Through surveying clinically-relevant cancer subtypes with differing degrees of aggression, biomarkers indicating a higher grade of cancer were developed (*eg* SCHALP1). Through comparison of specimens from patients with a spectrum of survival rates and drug responsive rates, biomarkers with potential to predict therapeutic responses were identified (HOTAIR). Reported lncRNAs with diagnostic or prognostic potential are summarized in **Table 1.1**.

Non-coding RNAs as therapeutic targets

The last two decades have witnessed rapid advances in targeted therapies for malignancies. The clinical benefits achieved following imatinib treatment for BCR-ABL

leukemia, or vemurafenib treatment for BRAF(V600E) melanoma have verified the rationale of targeted therapies, which function by directly targeting the existing driver mutations in certain subtypes of cancer. This principle can be further developed and applied in the search of therapeutic options targeting non-coding RNAs.

Compared to the development of chemical compounds inhibiting protein targets, the targeting strategy for non-coding RNAs is more straight-forward. Clinical-grade antisense oligonucleotides have been developed to inhibit RNAs directly *in vivo*. ASOs bind with RNA targets by base-pair matching, resulting in the mRNA degradation, translational arrest, splicing modulation, or inhibition of RNA-binding proteins[84]. As of 2019, the FDA has approved the use of several antisense drugs: fomivirsen as a treatment for cytomegalovirus retinitis, mipomersen for homozygous familial hypercholesterolemia, eteplirsen for Duchenne muscular dystrophy, and nusinersen for spinal muscular atrophy. These practices open a new door for therapeutic management of diseases where non-coding RNAs are key drivers to disease initiation or progression.

By contributing to nearly 60% of transcriptional output, non-coding RNAs offer great opportunities for the exploration of therapeutic targets in cancer. The critical problem lies in the identification of non-coding RNAs with *bona fide* oncogenic effects. Furthermore, the clinical benefits of non-coding RNA that exhibit tissue- and/or cancer-restricted expression pattern are large. Targeting these ncRNAs should eschew potential off-target effects, while generating highly tissue-restricted anti-cancer effects. A small number of

lncRNAs with clear mechanism of oncogenic functions have been reported in the literature. Functions of two example oncogenic lncRNAs, *THOR* and *SAMMSON*, will be discussed in the next section.

Functions and mechanisms of non-coding RNAs in cancer

As discussed above, most lncRNAs are expressed at relatively low levels compared to protein-coding genes. It remains a debatable question what percentage of lncRNAs are solely by-product of genome openness and regulation, and what percentage has *bona-fide* functional roles. The well-established hallmarks of cancer contain six aspects, which includes sustaining proliferative signaling, evading growth suppressors, resisting cell death, activating invasion and metastasis, enabling replicative immortality, and inducing angiogenesis[85]. LncRNAs have been reported to contribute to each of these aspects, mainly through modulating critical pathways involved in cancer initiation and progression (**Figure 1.4**).

Functional mechanisms of lncRNAs can be categorized by the molecular complexes that interact with them (**Figure 1.5**). The first group of lncRNAs interacts with the chromatin complex and directly regulates gene transcription and chromatin openness. The second group of lncRNAs interacts with RNAs or miRNAs and mediates post-transcriptional regulation of RNA transcripts. The third group of lncRNAs directly interacts with proteins and affects protein abundance or localization. Detailed mechanisms of lncRNAs in each group are reviewed in the following sections.

The interaction of lncRNAs with the chromatin complex

LncRNAs can form architectures with chromatin complexes and mediate epigenetic changes by altering DNA methylation, histone modifications, and chromatin structures. These lncRNAs can either work *in cis*, which regulate histone modifications or chromatin structures and thus affect transcription of neighboring genes, or they can work *in trans* by recruiting or scaffolding histone modification complexes to genomically distant genes. The effect of these lncRNAs could either be localized to specific gene regions (for instance, *LUNAR1*, *ANRIL*), or cover large chromosomal regions (for instance, Xist, *H19*, *Kcnq1ot1*, *Airn*, *Snprn*, etc). The proposed functioning models of imprinting lncRNAs, enhancer lncRNAs, and other lncRNAs that bind chromatin compartments are reviewed below.

Imprinting lncRNAs

Genomic imprinting introduces parent-dependent epigenetic marking, thus ensuring the expression of genes in a parent-of-origin manner. This lncRNA-mediated allelic regulation affects ~5%-10% of genes in the mammalian genome[86]. LncRNAs coordinate *in cis* with chromatin-modifying complexes, which in turn, silence or activate genes within clusters regulated by alleles. Roles of imprinting RNAs are directly defined by endogenous genetic manipulation (truncating lncRNA by insertion of polyadenylation signals). Well characterized clusters of imprinted genes include *Igf2*/*H19*, *Igf2r*/*Airn*, *Snprn*/*Ube3A*, *Kcnq1*, *Gnas*, and *Dlk1*/*Ftl2*[87, 88].

Two types of mechanisms have been proposed regarding how imprinting is initiated by lncRNAs *in cis*[86]: (1) imprinting RNAs can function through the act of transcription (not through the transcript itself). The sense-antisense overlap transcription between lncRNA and adjacent imprinted genes results in transcriptional interference at mRNA promoter or enhancers, thus attenuating transcription from adjacent mRNA. One example supporting this model is imprinting cluster *Igf2r*/*Airn*. Even after *Airn* was truncated, silencing of *Igf2r* could still be achieved as long as *Airn* transcription overlaps with *Igf2r* promoter[89, 90]. (2) Many imprinting RNAs act through the transcripts themselves. By coating the local chromosomal structure by RNA clouds at the site of transcription, imprinting RNAs recruit repressive chromatin modifiers, thus confirming a repressive chromatin compartment[91-94]. The earliest evidence of this mechanism was acquired from *Xist*, which mediates X-chromosome inactivation via Polycomb proteins[95-98]. It is of note that the aforementioned two types of mechanisms can be employed by the same imprinting cluster under different spatiotemporal contexts. For example, in placenta, the aforementioned example *Airn* is required for allele-specific silencing of *Slc22a3* through interacting with H3K9 histone methyltransferase G9a at *Slc22a3* promoter regions[91].

Enhancer lncRNAs

Enhancer RNAs are lncRNAs transcribed by Pol II from enhancer regions[69]. The universal existence of enhancer lncRNAs in many different cell lines and species suggests that lncRNAs may be part of the structural complex that governs enhancer

function[69]. Among the limited enhancer RNAs studied with clear functions, two types of mechanisms have been reported. One type includes lncRNA recruitment of the WDR5-containing methyltransferase complex, resulting in methylation of the histone H3 on lysine 4 (H3K4me), which is a canonical chromatin signature for enhancers. lncRNAs bind to a specific pocket structure on WDR5 protein, through which the MLL complex is guided to specific stretches of DNA and confirm active chromatin [99]. Example enhancer lncRNAs in this group include NeST activation of Interferon-v Locus[100] and HOTTIP[101, 102]. The functioning mechanism of HOTTIP is described below.

HOTTIP supports active chromatin confirmation by recruiting histone methyltransferase at HOX gene locus. HOX genes are evolutionally highly conserved and encode homeodomain-containing transcription factors[103]. They control critical processes in early development, ranging from cell identity establishment, tissue pattern development, to organogenesis[104, 105]. Malfunction of HOX genes result in epigenetic dysregulation, thus causing onset of malignancy. There are four clusters of HOX genes in human (A, B, C, D), located on different chromosomes. Within each cluster, HOX genes are “spatial collinearly” arranged on the chromosome according to their expression pattern along the anterior-posterior axis[106, 107].

In 2011, The Chang Group identified for the first time that lncRNA contribute to HOX gene locus regulation[102]. HOTTIP is transcribed from the 5’ tip of the HOXA locus, and helps to enhance HOXA expression. HOTTIP directly binds with adaptor protein

WDR5, and targets the SET-domain containing WDR5/MLL complexes across HOXA, thus inducing H3K4 trimethylation and gene activation. Pro-oncogenic roles of HOTTIP have been observed in blood, prostate, lung, pancreatic, and many other cancers[108-110].

The other type of enhancer lncRNA involves modification of chromatin structures through lncRNA-mediated stabilization of Mediator complexes and RNA PolII at enhancer elements. Some studies termed lncRNAs that facilitate enhancer-promoter interaction as ncRNA-activating (ncRNA-a). ncRNA-a is able to induce long-range transcriptional activation through DNA-looping and association with the Mediator complex. Examples of these lncRNAs and their targets include *ncRNA*-a7 vs. *SNA1* and *AURKA*, *ncRNA-a7* vs. *AURKA*, *ncRNA-a3* vs. *TAL1*, and *LUNAR* vs. *IGF1R*.[111-113] The functioning mechanism of one lncRNA in this group, LUNAR1, is described below.

LUNAR1 configures chromatin structure through Mediator Complex. This lncRNA is regulated by NOTCH, and is essential for human T-cell acute lymphoblastic leukemia (T-ALL) growth *in vitro* and *in vivo*. Genome-wide chromosome conformation capture (Hi-C, a method to examine high-order chromatin structure) demonstrated that LUNAR1, together with neighboring gene IGF1R, locates in the same 500 kb topologically associating domain[114]. Expression of LUNAR1 is driven by a highly active enhancer in the last intron of IGF1R, which is occupied by NOTCH1[115].

Interestingly, the phenotypic effect and gene expression profile change caused by LUNAR1 loss largely resembles those observed when the IGF1R gene is silenced[113]. Chromosome conformation capture (3C) experiments have demonstrated that a chromatin loop places the enhancer region on IGF1R and the promoter region on LUNAR1 in close physical proximity, leading to the formation of a DNA loop. Depletion of LUNAR1 lncRNA by shRNA leads to a decreased occupancy of the Mediator complex (MED1 and MED2) and RNA Pol II occupancy at the IGF1R enhancer region, as well as significant loss of RNA Pol II binding to the IGF1R promoter region. These observations support a model where LUNAR1 lncRNA enhances IGF1R mRNA expression and maintains IGF1 signaling through *cis-activation* of chromatin high-order conformation[113].

Other lncRNAs that bind with chromatin *in cis* or *in trans*

Apart from imprinting lncRNAs and enhancer lncRNAs, there exist lncRNAs that function through chromatin, either *in cis* or *in trans*, thus repressing or enhancing target gene expression. lncRNAs in this group can bind to multi-layered components in the chromosomal structure, thus mediating DNA methylation, histone methylation, and chromatin looping. Molecules involved in this process include DNA methyltransferase (DNMTs) and demethylase (TET), histone methyltransferase or demethylase (Polycomb complex, MLL complex, LSD1/CoREST/REST Complex), as well as CTCF/cohesion and Mediator Complex[68, 116-121]. One lncRNA, *ANRIL*, that affects histone modifications is described below as an example.

ANRIL is a lncRNA located antisense to *INK4b-ARF-INK4a* gene cluster on chromosome 9p21.The *INK4b-ARF-INKa* locus encodes for tumor suppressor genes that play widespread and independent roles in tumor suppression[122]. Recurrent events of deletions, mutations, and aberrant epigenetic inactivation of this locus have been observed across human cancer tissues and cell lines[123-125]. ANRIL expression is associated with SNPs on chromosome 9p21 and suggest susceptibility to several types of cancer[126, 127]. The epigenetic status of this locus is partially controlled by methylation of histone H3 at lysine 27 (H3K27me), which is exerted by the Polycomb group proteins[128].

Elevated expression of ANRIL has been observed in cancers originated from a variety of tissues that include neoplastic epithelial tissues, oesophageal squamous cell carcinoma, and leukaemia leukocytes[126, 128, 129]. Through direct interaction with the CBX7 protein (a component of PRC1 complex) or Suz12 protein (a component of PRC2 complex), ANRIL recruits the PcG protein complex to the local chromosome, thus modulating chromatin structure and inducing stable transcription silencing of the tumor suppressor genes *INK4B* and *INK4A*.

As lncRNA functions may be lineage-specific and pathological context-dependent, lncRNAs with established functions *in cis* may possess further functional roles *in trans* in different lineages. For example, in adult T-cell leukemia models, ANRIL engages EZH2 to support proliferation of Human T-Cell Leukemia Virus Type I (HTLV-1) infected

cells. On the one hand, the ANRIL/EZH2 complex employs histone methyltransferase (HMT) activity to induce the H3K27 trimethylation to p21/CDKN1A promoter, thus inhibiting p21/CDKN1A transcription. On the other hand, ANRIL forms a ternary complex with EZH2 and p65, and activates the NF-kb pathway in ATL cells. This activation was independent of the HMT activity of EZH2[130].

lncRNAs mediating post-transcriptional regulation of RNA

Compared to the number of lncRNAs that interact with chromatin structures or the number of lncRNAs that bind proteins, a smaller amount of lncRNAs regulate mRNA at the post-transcriptional level, and can either increase or decrease the stability of mRNAs via RNA-binding proteins and miRNA species. Micro RNAs have been studied as key modulators of mRNA stability, while many lncRNAs have proposed roles of miRNA sponging.

In some cases, lncRNAs act as microRNA decoys, causing sequestration of microRNAs, favoring expression of repressed target mRNAs[131]. For instance, lnc-MD1 “sponges” both miR-133 and miR-135 to stabilize MAML1 and MEF2C. Both of them are transcription factors that control muscle differentiation by activating muscle-specific gene expression[132]. lncRNA H19 “sponges” the let-7 family of microRNAs via canonical and non-canonical binding sites, thus playing a key role in cell differentiation and cancer metabolism[133]. lncRNA-ROR “sponges” miR-145 to stabilize ZEB2, thus promoting invasion and metastasis in hepatocellular carcinoma[134].

In other cases, lncRNAs stabilize mRNA by competing with miRNAs for shared target mRNAs. For example, lncRNA BACE1-AS competes with miR-485-5p for binding to BACE1 mRNAs, leading to elevated expression level of BACE1, a critical enzyme in Alzheimer's disease pathophysiology[72].

lncRNAs interacting with proteins

LncRNAs can bind with proteins and directly modulate protein functions. This is carried out by regulating protein stability, affecting protein binding with target macromolecules (DNA, RNA, or proteins), or dictating relative localization of proteins in cell compartments, etc. For instance, the modular domain in NEAT1 lncRNA recruits NONO dimers to initiate assembly of the paraspeckle structure[135]. p53-induced lincRNA-p21 binds with hnRNP-K to modulate gene expression in the DNA damage response[119]. GAS5 lncRNA interacts with glucocorticoid receptor (GR) to prevent it from binding to the genomic glucocorticoid receptor elements, limiting the transcriptional induction of GR-responsive genes. Two lncRNAs functioning through protein binding partners are described in details below.

THOR is an example lncRNAs that binds protein and modulates downstream protein functions[136]. THOR is expressed exclusively in the testis, and has elevated expression in a broad range of cancers[136]. This lncRNA is evolutionarily conserved, with homologs existing in the mouse and zebrafish. Expression of this lncRNA potentiates

tumor growth and invasiveness[136]. These functions are carried out by a conserved domain on THOR that interacts with the IGF2BP1 protein[136]. The IGF2BP1 protein possess many functions, including regulating the stability of a set of target mRNAs and modulating IGF2 signaling. THOR regulates these functions by stabilization between the IGF2BP1 protein and its target RNAs[136]. Loss of THOR decreases the stability of IGF2BP1 targeting mRNAs, as well as alters IGF2 signaling[136].

SAMMSON is an example lncRNA that binds protein and modulates protein cellular localization[137]. This pro-oncogenic lncRNA is induced by lineage-specific transcription factor SOX10, with elevated expression levels in melanomas[137]. Loss of SAMMSON significantly decreases melanoma cell viability, whereas over-expression of SAMMSON increases clonogenic potential. These effects were achieved through a direct interaction with the p32 protein, a master regulator of mitochondrial homeostasis and metabolism[137]. Although SAMMSON silencing does not affect total p32 levels, it alters the cellular localization of p32, with a decrease in the mitochondrial fraction and an increase in nuclear fraction. Mitochondrial p32 plays a critical role in the maintenance of oxidative phosphorylation through mediating maturation of mitochondria 16S ribosomal RNA. SAMMSON-silencing causes p32 deficiency in mitochondria, and thus leads to reduced levels of mature 16S ribosomal RNA, as well as impaired enzymatic activity of respiratory complexes I and IV[137]. Through dictating protein localization in cell compartments, SAMMSON is able to potentiate mitochondria functions, and thus plays an oncogenic roles in a melanoma-specific manner[137].

Outline of this study

In this thesis, we aim to discover and characterize novel non-coding RNAs with therapeutic or diagnostic potential in prostate cancer. We begin by performing transcriptome analysis to identify long non-coding RNAs that are regulated by AR signaling and are associated with prostate cancer progression (Chapter 2). We then characterize a top lineage-specific lncRNA (ARLNC1) identified from this analysis, investigate its functional role, and evaluate its therapeutic potential in prostate cancer (Chapter 3). We further interrogate the involvement of ARLNC1 in the AR signaling regulation axis (Chapter 4). Next, we explore a novel class of non-coding RNA, circular RNAs, and evaluate their potential as biomarkers in cancer (Chapter 5). Finally, we explore and discuss future directions to study lncRNAs with functional roles and translational opportunities in cancer (Chapter 6).

Figures

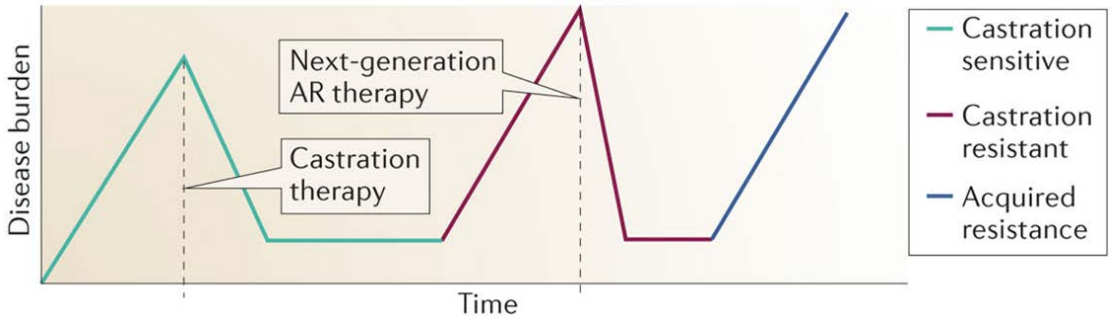


Figure 1.1 Phases in prostate cancer progression
(Reproduced with permission from Watson, P.A., V.K. Arora, and C.L. Sawyers, Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer,* 2015. 15(12): p. 701-11. License Number for Figure reprinting: 4592810956344)

Increasing disease burden following primary prostate cancer therapy is indicated by rising prostate-specific antigen (PSA) levels and/or radiographic progression, and is treated with medical castration. The castration-resistant prostate cancer (CRPC) stage follows the failure of castration therapy. Treatment with next-generation androgen receptor (AR) inhibitors (for example, abiraterone or enzalutamide) is initiated during CRPC, but acquired (or inherent) resistance mechanisms lead to disease recurrence and ultimately death.

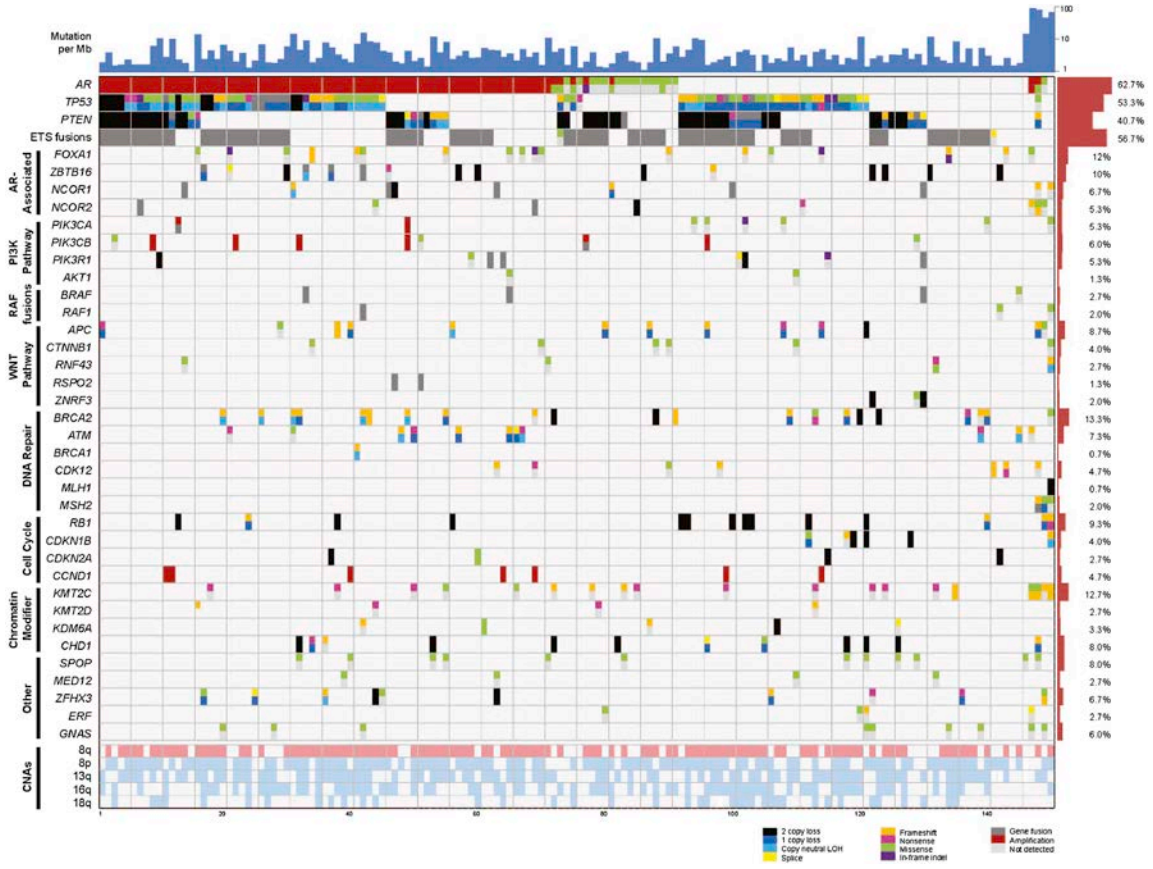


Figure 1.2 Integrative landscape analysis of somatic and germline aberrations in metastatic CRPC
(Reprinted with permission from Robinson, D., et al., Integrative clinical genomics of advanced prostate cancer. *Cell*, 2015. 161(5): p. 1215-1228. License Number for Figure reprinting: 4592810146567)

Columns represent individual affected individuals, and rows represent specific genes grouped in pathways. Mutations per Mb are shown in the upper histogram, and incidence of aberrations in the cohort is in the right histogram. Copy number variations (CNVs) common to mCRPC are shown in in the lower matrix, with pink representing gain and light blue representing loss. Color legend of the aberrations represented including amplification, two copy loss, one copy loss, copy neutral loss of heterozygosity (LOH), splice site mutation, frameshift mutation, missense mutation, in-frame indel, and gene fusion. Cases with more aberration in a gene are represented by split colors.

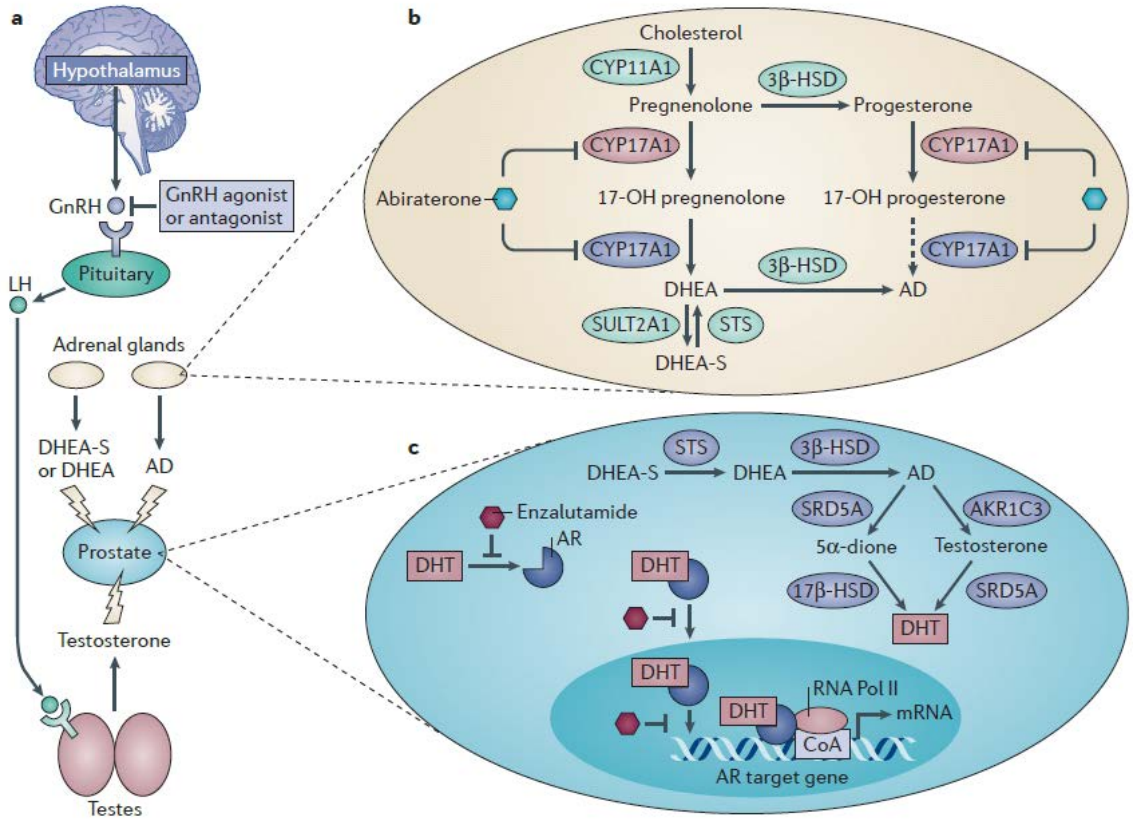


Figure 1.3 AR signaling regulation
(Reproduced with permission from Watson, P.A., V.K. Arora, and C.L. Sawyers, Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer*, 2015. 15(12): p. 701-11. License Number for Figure reprinting: 4592810956344)

(**a**) The hormones gonadotropin-releasing hormone (GnRH) and luteinizing hormone (LH) bind to their cognate receptors, resulting in testosterone secretion from Leydig cells of the testes. Chronic use of GnRH agonists leads to downregulation of the GnRH receptor (GnRH‐R), whereas antagonists provide immediate GnRH‐R blockade. Both agents suppress LH production, which causes a decline in serum testosterone to castrate levels. The adrenal glands secrete the androgens dehydroepiandrosterone sulfate (DHEA−S; predominantly), DHEA and androstenedione (AD) into the circulation.
(**b**) Adrenal androgen de novo steroidogenesis (enzymes in ovals). CYP17A1 (cytochrome P450 family 17 subfamily A polypeptide 1 has 17α‐hydroxylation (red) and 17, 20‐lyase (blue) activities; both are inhibited by abiraterone. The dashed arrow indicates a weak effect.
(**c**) Prostate conversion of adrenal androgens to dihydrotestosterone (DHT) is shown in the top right part of the cell. DHT binds to the androgen receptor (AR) in the cytoplasm, triggering a

conformational change that leads to nuclear translocation. DHT-bound AR homodimerizes and, with co-activators (CoAs) and RNA polymerase II (RNA Pol II) or co-repressors (not shown), binds to DNA at cis androgen response elements to activate or repress (not shown) AR target gene expression, respectively.Enzalutamide inhibits AR by competing with DHT for binding, blocking nuclear translocation, and blocking DNA and cofactor binding8. AKR1C3, aldo-keto reductase family 1 member C3; 3β-HSD, 3β‑hydroxysteroid dehydrogenase/Δ5‑4‑isomerase; 17β‑HSD, 17β‑hydroxysteroid dehydrogenase; SRD5A, steroid 5α‑reductase; STS, steryl-sulfatase; SULT2A1, bile salt sulfotransferase.

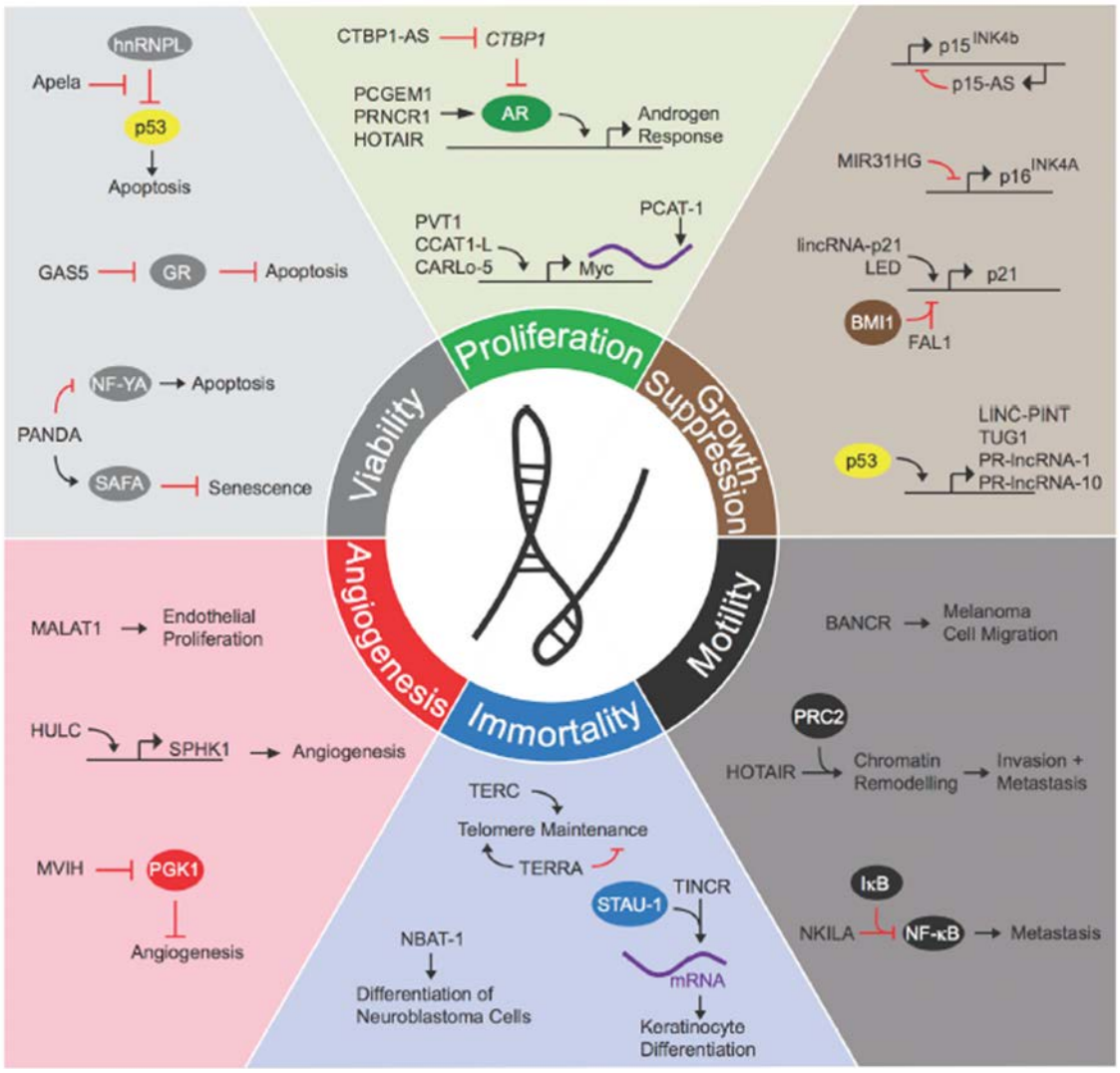


Figure 1.4 LncRNAs contribute to hallmarks of cancer
(Reproduced with permission from Schmitt, A.M. and H.Y. Chang, Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*, 2016. 29(4): p. 452-463. License Number for Figure reprinting: 4592800936029)

lncRNAs contribute to each of the six hallmarks of cancer (diagram adapted from *Hanahan and Weinberg, 2000[138]*). Selected examples of lncRNAs and their molecular partners or genomic targets are shown for proliferation, growth suppression, motility, immortality, angiogenesis, and viability cancer phenotypes.

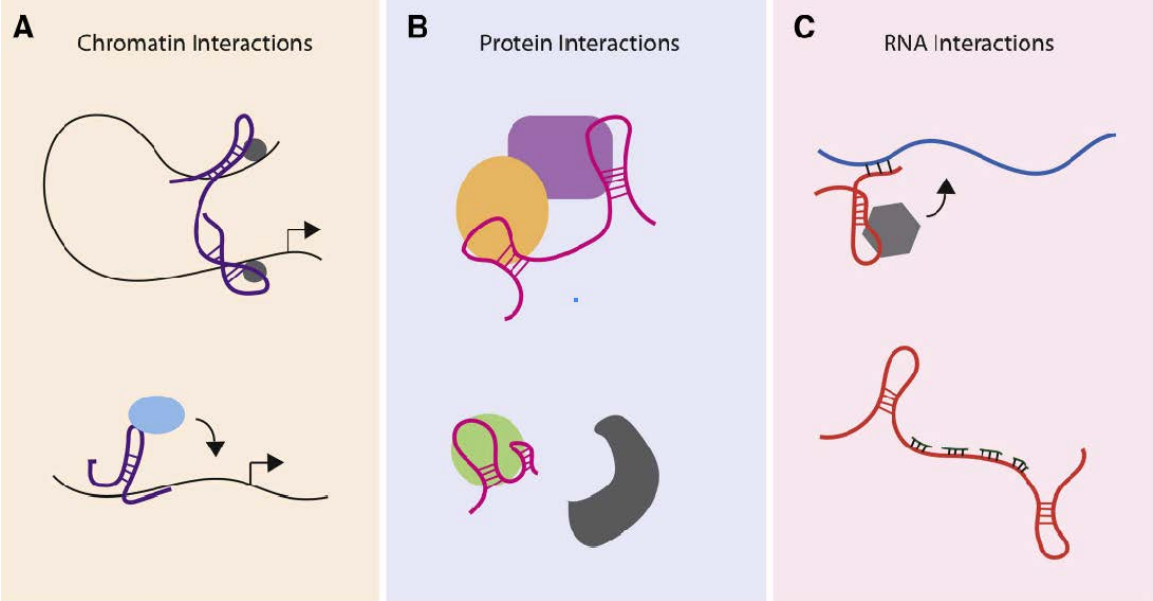


Figure 1.5 lncRNA mechanisms rely on interactions with cellular macromolecules (Reproduced with permission from Schmitt, A.M. and H.Y. Chang, Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*, 2016. 29(4): p. 452-463. License Number for Figure reprinting: 4592800936029)

(**A**) Chromatin-bound lncRNAs can regulate gene expression by controlling local chromatin architecture (above) or directing the recruitment of regulatory molecules to specific loci (below).
(**B**) lncRNA interactions with multiple proteins can promote the assembly of protein complexes (above) or impair protein-protein interactions(below).
(**C**) mRNA interactions with lncRNA can recruit protein machinery involved in multiple aspects of mRNA metabolism to affect splicing, mRNA stability, or translation (above) or sequester miRNA away from target mRNA (below).

Tables

Table 1.1 lncRNAs with diagnostic or prognostic potential
(Table adapted from Bhan, A., M. Soleimani, and S.S. Mandal, Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Res,* 2017. 77(15): p. 3965-3981. License Number for reproduction: 4593960187578)

Cancer	LncRNA Biomarker	Potential implications	Site of detection
Prostate cancer	PCA3	Diagnosis; Prognosis	Urine; Tumor
	LincRNA-p21	Diagnosis; Stratification	Urine
	PCAT-18	Metastasis	Plasma
	MALAT1	Risk of tumorigenesis;	Urine; Plasma
	PVT1	Aggressiveness	Tumor
	PCAT-14	Diagnosis	Tumor, Urine
	SCHLAP1	Aggressiveness	Urine, Tumor
	TRPM2	Early identification; aggressiveness	Tumor
Breast cancer	ZFAS1	Diagnosis	Tumor
	HOTAIR	Diagnosis	Serum
	RP11-445H22.4	Diagnosis	Serum
	HIF1A-AS2; AK124454	Recurrence	Tumor
Lung cancer	MALAT1	Early detection; Risk of metastasis	Whole blood; Tumor
	SPRY4-IT1; ANRIL; NEAT1	Early detection	Plasma
	UCA1	Diagnosis	Plasma; Tumor
	GAS5	Diagnosis	Plasma
Colorectal cancer	HOTAIR	Prognosis and metastatsis	Tumor
	HOTAIR; CCAT1; CCAT2	Diagnosis	Serum
	FER1L4	Recurrence; Metastasis	Plasma
	XLOC_006844; LOC152578	Risk of tumorigenesis	Plasma
Hepatocellular cancer	HOTAIR	Recurrence after transplant	Tumor

	uc001ncr; AX800134	Detection (especially early-stage)	Serum
	HULC; Linc00152	Diagnosis; Metastasis	Plasma
	RP11-160H22.5; XLOC014172; LOC149086, HEIH	Risk of tumorigenesis, prognostic factor for recurrence and survival	Plasma
	XLOC014172; LOC149086	Risk of metastasis	
Bladder cancer	UCA1	Diagnosis	Urine
	H19	Early recurrence	Tumor
	HOTAIR	Overall survival	Tumor
Leukemia	CRNDE	Identification of subtypes of AML (acute myeloid leukemia) (M2 or M3)	Bone marrow, Lymph nodes
Ovarian cancer	NEAT1	Invasiveness; Prognosis	Tumor
Renal cancer	LET; PVT1; PANDAR; PTENP1; LINC00963	Early detection	Serum
Cervical cancer	HOTAIR	Prognosis; Recurrence	Serum
Esophageal cancer	POU3F3; HNF1A-AS1; SPRY4-IT1	Early screening	Plasma
Gastric cancer	H19	Early screening	Plasma
	LINC00152	Detection; Invasion	Gastric juice; Tumor
	UCA1	Early detection; Prognosis prediction	Gastric juice; Tumor
	CUDR; LSINCT-5; PTENP1	Diagnosis	Serum
	AA174084	Early diagnosis	Tumor; Plasma; Gastric juice

References

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019.* CA Cancer J Clin, 2019. **69**(1): p. 7-34.
2. Etzioni, R., et al., *Quantifying the role of PSA screening in the US prostate cancer mortality decline.* Cancer Causes Control, 2008. **19**(2): p. 175-81.
3. Tsodikov, A., et al., *Reconciling the Effects of Screening on Prostate Cancer Mortality in the ERSPC and PLCO Trials.* Ann Intern Med, 2017. **167**(7): p. 449-455.
4. DeSantis, C.E., et al., *Cancer statistics for African Americans, 2019.* CA Cancer J Clin, 2019. **69**(3): p. 211-233.
5. Bratt, O., et al., *Family History and Probability of Prostate Cancer, Differentiated by Risk Category: A Nationwide Population-Based Study.* J Natl Cancer Inst, 2016. **108**(10).
6. Lesko, S.M., L. Rosenberg, and S. Shapiro, *Family history and prostate cancer risk.* Am J Epidemiol, 1996. **144**(11): p. 1041-7.
7. Matikaine, M.P., et al., *Relatives of prostate cancer patients have an increased risk of prostate and stomach cancers: a population-based, cancer registry study in Finland.* Cancer Causes Control, 2001. **12**(3): p. 223-30.
8. de Bono, J.S., et al., *Abiraterone and increased survival in metastatic prostate cancer.* N Engl J Med, 2011. **364**(21): p. 1995-2005.
9. Ryan, C.J., et al., *Abiraterone in metastatic prostate cancer without previous chemotherapy.* N Engl J Med, 2013. **368**(2): p. 138-48.
10. Beer, T.M. and B. Tombal, *Enzalutamide in metastatic prostate cancer before chemotherapy.* N Engl J Med, 2014. **371**(18): p. 1755-6.
11. Scher, H.I., et al., *Increased survival with enzalutamide in prostate cancer after chemotherapy.* N Engl J Med, 2012. **367**(13): p. 1187-97.
12. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types.* Nature, 2013. **502**(7471): p. 333-339.
13. Bailey, M.H., et al., *Comprehensive Characterization of Cancer Driver Genes and Mutations.* Cell, 2018. **174**(4): p. 1034-1035.
14. Zehir, A., et al., *Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.* Nat Med, 2017. **23**(6): p. 703-713.
15. Baca, S.C., et al., *Punctuated evolution of prostate cancer genomes.* Cell, 2013. **153**(3): p. 666-77.
16. Barbieri, C.E., et al., *Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer.* Nat Genet, 2012. **44**(6): p. 685-9.
17. Berger, M.F., et al., *The genomic complexity of primary human prostate cancer.* Nature, 2011. **470**(7333): p. 214-20.
18. Cooper, C.S., et al., *Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue.* Nat Genet, 2015. **47**(4): p. 367-372.

19. Pflueger, D., et al., *Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing.* Genome Res, 2011. **21**(1): p. 56-67.
20. Taylor, B.S., et al., *Integrative genomic profiling of human prostate cancer.* Cancer Cell, 2010. **18**(1): p. 11-22.
21. Tomlins, S.A., et al., *Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer.* Nature, 2007. **448**(7153): p. 595-9.
22. Wang, X.S., et al., *Characterization of KRAS rearrangements in metastatic prostate cancer.* Cancer Discov, 2011. **1**(1): p. 35-43.
23. Zhu, P., et al., *Macrophage/cancer cell interactions mediate hormone resistance by a nuclear receptor derepression pathway.* Cell, 2006. **124**(3): p. 615-29.
24. Robinson, D., et al., *Integrative clinical genomics of advanced prostate cancer.* Cell, 2015. **161**(5): p. 1215-1228.
25. Watson, P.A., V.K. Arora, and C.L. Sawyers, *Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer.* Nat Rev Cancer, 2015. **15**(12): p. 701-11.
26. Heinlein, C.A. and C. Chang, *Androgen receptor in prostate cancer.* Endocr Rev, 2004. **25**(2): p. 276-308.
27. Lu, S., G. Jenster, and D.E. Epner, *Androgen induction of cyclin-dependent kinase inhibitor p21 gene: role of androgen receptor and transcription factor Sp1 complex.* Mol Endocrinol, 2000. **14**(5): p. 753-60.
28. Unni, E., et al., *Changes in androgen receptor nongenotropic signaling correlate with transition of LNCaP cells to androgen independence.* Cancer Res, 2004. **64**(19): p. 7156-68.
29. Balk, S.P. and K.E. Knudsen, *AR, the cell cycle, and prostate cancer.* Nucl Recept Signal, 2008. **6**: p. e001.
30. Han, G., et al., *Mutation of the androgen receptor causes oncogenic transformation of the prostate.* Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1151-6.
31. Andersen, R.J., et al., *Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor.* Cancer Cell, 2010. **17**(6): p. 535-46.
32. Sun, S., et al., *Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant.* J Clin Invest, 2010. **120**(8): p. 2715-30.
33. Watson, P.A., et al., *Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor.* Proc Natl Acad Sci U S A, 2010. **107**(39): p. 16759-65.
34. Takeda, D.Y., et al., *A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer.* Cell, 2018. **174**(2): p. 422-432 e13.
35. Xu, J., R.C. Wu, and B.W. O'Malley, *Normal and cancer-related functions of the p160 steroid receptor co-activator (SRC) family.* Nat Rev Cancer, 2009. **9**(9): p. 615-30.

36. Chen, S., et al., *Androgen receptor phosphorylation and stabilization in prostate cancer by cyclin-dependent kinase 1.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15969-74.
37. Tan, S.H., et al., *Transcription factor Stat5 synergizes with androgen receptor in prostate cancer cells.* Cancer Res, 2008. **68**(1): p. 236-48.
38. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
39. Crick, F.H., *On protein synthesis.* Symp Soc Exp Biol, 1958. **12**: p. 138-63.
40. Eddy, S.R., *Non-coding RNA genes and the modern RNA world.* Nat Rev Genet, 2001. **2**(12): p. 919-29.
41. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.
42. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.
43. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome Res, 2012. **22**(9): p. 1775-89.
44. Djebali, S., et al., *Landscape of transcription in human cells.* Nature, 2012. **489**(7414): p. 101-8.
45. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.
46. Iyer, M.K., et al., *The landscape of long noncoding RNAs in the human transcriptome.* Nat Genet, 2015. **47**(3): p. 199-208.
47. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
48. Esteller, M., *Non-coding RNAs in human disease.* Nat Rev Genet, 2011. **12**(12): p. 861-74.
49. Han, J., et al., *The Drosha-DGCR8 complex in primary microRNA processing.* Genes Dev, 2004. **18**(24): p. 3016-27.
50. Hutvagner, G., et al., *A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.* Science, 2001. **293**(5531): p. 834-8.
51. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
52. Calin, G.A., et al., *Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15524-9.
53. Ota, A., et al., *Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma.* Cancer Res, 2004. **64**(9): p. 3087-95.
54. Chang, T.C., et al., *Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis.* Mol Cell, 2007. **26**(5): p. 745-52.

55. Cui, B., et al., *MicroRNA-155 influences B-cell receptor signaling and associates with aggressive disease in chronic lymphocytic leukemia.* Blood, 2014. **124**(4): p. 546-54.
56. Ferrajoli, A., et al., *Prognostic value of miR-155 in individuals with monoclonal B-cell lymphocytosis and patients with B chronic lymphocytic leukemia.* Blood, 2013. **122**(11): p. 1891-9.
57. Gregory, P.A., et al., *The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1.* Nat Cell Biol, 2008. **10**(5): p. 593-601.
58. Hayashita, Y., et al., *A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation.* Cancer Res, 2005. **65**(21): p. 9628-32.
59. He, L., et al., *A microRNA component of the p53 tumour suppressor network.* Nature, 2007. **447**(7148): p. 1130-4.
60. He, L., et al., *A microRNA polycistron as a potential human oncogene.* Nature, 2005. **435**(7043): p. 828-33.
61. Fazi, F., et al., *A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis.* Cell, 2005. **123**(5): p. 819-31.
62. Johnston, R.J., Jr., et al., *MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision.* Proc Natl Acad Sci U S A, 2005. **102**(35): p. 12449-54.
63. Li, X. and R.W. Carthew, *A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye.* Cell, 2005. **123**(7): p. 1267-77.
64. Li, Y., et al., *MicroRNA-9a ensures the precise specification of sensory organ precursors in Drosophila.* Genes Dev, 2006. **20**(20): p. 2793-805.
65. O'Donnell, K.A., et al., *c-Myc-regulated microRNAs modulate E2F1 expression.* Nature, 2005. **435**(7043): p. 839-43.
66. Yoo, A.S. and I. Greenwald, *LIN-12/Notch activation leads to microRNA-mediated down-regulation of Vav in C. elegans.* Science, 2005. **310**(5752): p. 1330-3.
67. Rinn, J.L. and H.Y. Chang, *Genome regulation by long noncoding RNAs.* Annu Rev Biochem, 2012. **81**: p. 145-66.
68. Canzio, D., et al., *Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin alpha Promoter Choice.* Cell, 2019.
69. Kim, T.K., et al., *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-7.
70. Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.
71. Hezroni, H., et al., *Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species.* Cell Rep, 2015. **11**(7): p. 1110-22.

72. Faghihi, M.A., et al., *Evidence for natural antisense transcript-mediated inhibition of microRNA function.* Genome Biol, 2010. **11**(5): p. R56.
73. Castellanos-Rubio, A., et al., *A long noncoding RNA associated with susceptibility to celiac disease.* Science, 2016. **352**(6281): p. 91-5.
74. Huarte, M., *RNA. A lncRNA links genomic variation with celiac disease.* Science, 2016. **352**(6281): p. 43-4.
75. Capel, B., et al., *Circular transcripts of the testis-determining gene Sry in adult mouse testis.* Cell, 1993. **73**(5): p. 1019-30.
76. Du, W.W., et al., *Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2.* Nucleic Acids Res, 2016. **44**(6): p. 2846-58.
77. Hansen, T.B., et al., *Natural RNA circles function as efficient microRNA sponges.* Nature, 2013. **495**(7441): p. 384-8.
78. Hansen, T.B., J. Kjems, and C.K. Damgaard, *Circular RNA and miR-7 in cancer.* Cancer Res, 2013. **73**(18): p. 5609-12.
79. Li, F., et al., *Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/beta-catenin pathway.* Oncotarget, 2015. **6**(8): p. 6001-13.
80. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency.* Nature, 2013. **495**(7441): p. 333-8.
81. Rybak-Wolf, A., et al., *Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed.* Mol Cell, 2015. **58**(5): p. 870-85.
82. Weng, W., et al., *Circular RNA ciRS-7-A Promising Prognostic Biomarker and a Potential Therapeutic Target in Colorectal Cancer.* Clin Cancer Res, 2017. **23**(14): p. 3918-3928.
83. Guarnerio, J., et al., *Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations.* Cell, 2016. **166**(4): p. 1055-1056.
84. Rinaldi, C. and M.J.A. Wood, *Antisense oligonucleotides: the next frontier for treatment of neurological disorders.* Nat Rev Neurol, 2018. **14**(1): p. 9-21.
85. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
86. Lee, J.T. and M.S. Bartolomei, *X-inactivation, imprinting, and long noncoding RNAs in health and disease.* Cell, 2013. **152**(6): p. 1308-23.
87. Barlow, D.P., *Genomic imprinting: a mammalian epigenetic discovery model.* Annu Rev Genet, 2011. **45**: p. 379-403.
88. Mancini-Dinardo, D., et al., *Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes.* Genes Dev, 2006. **20**(10): p. 1268-82.
89. Latos, P.A., et al., *Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing.* Science, 2012. **338**(6113): p. 1469-72.
90. Latos, P.A., et al., *An in vitro ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias.* Development, 2009. **136**(3): p. 437-48.
91. Nagano, T., et al., *The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin.* Science, 2008. **322**(5908): p. 1717-20.

92. Pandey, R.R., et al., *Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.* Mol Cell, 2008. **32**(2): p. 232-46.
93. Redrup, L., et al., *The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing.* Development, 2009. **136**(4): p. 525-30.
94. Terranova, R., et al., *Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos.* Dev Cell, 2008. **15**(5): p. 668-79.
95. Brockdorff, N., et al., *The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus.* Cell, 1992. **71**(3): p. 515-26.
96. Penny, G.D., et al., *Requirement for Xist in X chromosome inactivation.* Nature, 1996. **379**(6561): p. 131-7.
97. Pinter, S.F., et al., *Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations.* Genome Res, 2012. **22**(10): p. 1864-76.
98. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.
99. Rapicavoli, N.A., et al., *A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics.* Elife, 2013. **2**: p. e00762.
100. Gomez, J.A., et al., *The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus.* Cell, 2013. **152**(4): p. 743-54.
101. Chu, C., et al., *Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions.* Mol Cell, 2011. **44**(4): p. 667-78.
102. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.
103. McGinnis, W. and R. Krumlauf, *Homeobox genes and axial patterning.* Cell, 1992. **68**(2): p. 283-302.
104. Cerda-Esteban, N. and F.M. Spagnoli, *Glimpse into Hox and tale regulation of cell differentiation and reprogramming.* Dev Dyn, 2014. **243**(1): p. 76-87.
105. Krumlauf, R., *Hox genes in vertebrate development.* Cell, 1994. **78**(2): p. 191-201.
106. Lappin, T.R., et al., *HOX genes: seductive science, mysterious mechanisms.* Ulster Med J, 2006. **75**(1): p. 23-31.
107. Noordermeer, D., et al., *The dynamic architecture of Hox gene clusters.* Science, 2011. **334**(6053): p. 222-5.
108. Fu, Z., et al., *LncRNA HOTTIP modulates cancer stem cell properties in human pancreatic cancer by regulating HOXA9.* Cancer Lett, 2017. **410**: p. 68-81.
109. Malek, R., et al., *TWIST1-WDR5-Hottip Regulates Hoxa9 Chromatin to Facilitate Prostate Cancer Metastasis.* Cancer Res, 2017. **77**(12): p. 3181-3193.
110. Sun, Y., et al., *LncRNA HOTTIP-Mediated HOXA11 Expression Promotes Cell Growth, Migration and Inhibits Cell Apoptosis in Breast Cancer.* Int J Mol Sci, 2018. **19**(2).

111. Lai, F., et al., *Activating RNAs associate with Mediator to enhance chromatin architecture and transcription.* Nature, 2013. **494**(7438): p. 497-501.
112. Orom, U.A., et al., *Long noncoding RNAs with enhancer-like function in human cells.* Cell, 2010. **143**(1): p. 46-58.
113. Trimarchi, T., et al., *Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia.* Cell, 2014. **158**(3): p. 593-606.
114. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.
115. Medyouf, H., et al., *High-level IGF1R expression is required for leukemia-initiating cell activity in T-ALL and is supported by Notch signaling.* J Exp Med, 2011. **208**(9): p. 1809-22.
116. Chalei, V., et al., *The long non-coding RNA Dali is an epigenetic regulator of neural differentiation.* Elife, 2014. **3**: p. e04530.
117. Di Ruscio, A., et al., *DNMT1-interacting RNAs block gene-specific DNA methylation.* Nature, 2013. **503**(7476): p. 371-6.
118. Merry, C.R., et al., *DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer.* Hum Mol Genet, 2015. **24**(21): p. 6240-53.
119. Bao, X., et al., *The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters.* Cell Res, 2015. **25**(1): p. 80-92.
120. Jin, J.J., et al., *Long noncoding RNA SYISL regulates myogenesis by interacting with polycomb repressive complex 2.* Proc Natl Acad Sci U S A, 2018. **115**(42): p. E9802-E9811.
121. Neumann, P., et al., *The lncRNA GATA6-AS epigenetically regulates endothelial gene expression via interaction with LOXL2.* Nat Commun, 2018. **9**(1): p. 237.
122. Kim, W.Y. and N.E. Sharpless, *The regulation of INK4/ARF in cancer and aging.* Cell, 2006. **127**(2): p. 265-75.
123. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. **463**(7283): p. 899-905.
124. Bignell, G.R., et al., *Signatures of mutation and selection in the cancer genome.* Nature, 2010. **463**(7283): p. 893-8.
125. Freedberg, D.E., et al., *Frequent p16-independent inactivation of p14ARF in human melanoma.* J Natl Cancer Inst, 2008. **100**(11): p. 784-95.
126. Cunnington, M.S., et al., *Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression.* PLoS Genet, 2010. **6**(4): p. e1000899.
127. Pasmant, E., et al., *ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS.* FASEB J, 2011. **25**(2): p. 444-8.
128. Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.* Mol Cell, 2010. **38**(5): p. 662-74.

129. Yu, W., et al., *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.* Nature, 2008. **451**(7175): p. 202-6.
130. Song, Z., et al., *Long Noncoding RNA ANRIL Supports Proliferation of Adult T-Cell Leukemia Cells through Cooperation with EZH2.* J Virol, 2018. **92**(24).
131. Yoon, J.H., K. Abdelmohsen, and M. Gorospe, *Functional interactions among microRNAs and long noncoding RNAs.* Semin Cell Dev Biol, 2014. **34**: p. 9-14.
132. Cesana, M., et al., *A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA.* Cell, 2011. **147**(2): p. 358-69.
133. Kallen, A.N., et al., *The imprinted H19 lncRNA antagonizes let-7 microRNAs.* Mol Cell, 2013. **52**(1): p. 101-12.
134. Li, C., et al., *The lincRNA-ROR/miR-145 axis promotes invasion and metastasis in hepatocellular carcinoma via induction of epithelial-mesenchymal transition by targeting ZEB2.* Sci Rep, 2017. **7**(1): p. 4637.
135. Yamazaki, T., et al., *Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation.* Mol Cell, 2018. **70**(6): p. 1038-1053 e7.
136. Hosono, Y., et al., *Oncogenic Role of THOR, a Conserved Cancer/Testis Long Non-coding RNA.* Cell, 2017. **171**(7): p. 1559-1572 e20.
137. Leucci, E., et al., *Melanoma addiction to the long non-coding RNA SAMMSON.* Nature, 2016. **531**(7595): p. 518-22.
138. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

Chapter 2

Transcriptome analysis identifies Androgen Receptor regulated long non-coding RNAs associated with prostate cancer progression[1]

Abstract

As the most commonly diagnosed non-cutaneous cancer in American men, prostate cancer causes approximately 30,000 deaths per year. The majority of these deaths are caused by metastatic disease, which is resistance to androgen-deprivation therapy. Thus, it is of critical importance to characterize molecular mechanisms of prostate cancer progression and resistance, and to develop novel therapeutic strategies. Sustained AR signaling has been frequently observed in castration-resistant prostate cancer (CRPC). However, nearly all previous studies focused on studying protein-coding genes in the AR signaling axis.

Here, we study the long non-coding transcriptome regulated by AR in prostate cancer. We began by identifying lncRNAs that are directly regulated by AR using cell line models. We further constructed a bioinformatic pipeline to prioritize lncRNAs that have strong

[1] This chapter was previously published as part of the following manuscript: Zhang, Y., et al., Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. *Nat Genet*, 2018. 50(6): p. 814-824.

association with cancer progression. In addition, we queried the prostate-lineage specificity of AR-regulated lncRNAs. The result demonstrated that there exists an abundant AR-responsive non-coding transcriptome that are previously underappreciated. Top prostate lineage-specific, AR-regulated, cancer-associated lncRNAs are worth further investigation.

Introduction

Long noncoding RNAs (lncRNAs) are a class of transcripts with diverse and largely uncharacterized biological functions[1-3]. Through interactions with chromatin, DNA, RNA species and proteins, lncRNAs function via chromatin remodeling, transcriptional and post-transcriptional regulation[4-9]. High-throughput RNA sequencing (RNA-seq) has enabled the identification of lncRNAs with suggested oncogenic and tumor-suppressive roles, including involvement in the pathogenesis and progression of prostate cancer[7, 10-12].

Primary prostate cancer is often hormone dependent and relies on signaling through the AR; therefore, the majority of patients are responsive to front-line treatment with androgen-deprivation therapy[13-15]. However, approximately 20% of cases progress to an incurable stage of the disease known as castration-resistant prostate cancer (CRPC), which still critically relies on AR signaling[16, 17], as evidenced by the clinical benefit afforded through the use of enzalutamide[18-21] or abiraterone[22-24]. Substantial efforts have been undertaken to identify mechanisms of sustained AR signaling in CRPC. AR

mutations, amplifications, AR splice variants, and alternative activation pathways have been reported [25-33]. However, few studies have investigated the role of AR-regulated long non-coding RNAs. Therefore, we initiated a comprehensive RNA-seq profiling investigation of AR-regulated, cancer-associated lncRNAs from prostate cancer cell lines and patient tissue samples.

Results

Identification of AR-regulated lncRNAs in cell lines

To nominate AR-regulated genes (ARGs), RNA-sequencing was performed on AR dependent VCaP and LNCaP prostate cancer cell lines that were stimulated with an AR ligand, dihydrotestosterone (DHT), for 6 hours or 24 hours (**Figure 2.1A**). A total of 1,702 genes were identified to be concordantly induced or repressed in VCaP and LNCaP at both time points (**Figure 2.1B-C**). More than 500 of them are lncRNAs (**Figure 2.1D**, **Table 2.1**). These data indicate that a large portion of the AR transcriptome remains uncharacterized.

To differentiate between direct and indirect ARGs, previously published AR chromatin immunoprecipitation (ChIP)-Seq data from LNCaP and VCaP cells were analyzed to find AR-binding promoter regions [34]. As expected for direct AR targets, increased levels of AR binding at transcription start sites (TSSs) in both LNCaP and VCaP cells were observed (**Figure 2.2B**), and the binding sites revealed a *de novo* motif identical to the canonical AR response element (**Figure 2.2**D). Treatment of AR antagonist (enzalutamide) induced

reverse gene expression changes compared to DHT stimulation (**Figure 2.2A**). And the AR binding levels at promoter regions of ARGs decreased following enzalutamide treatment (**Figure 2.2C**). A total of 987 genes were categorized as direct ARGs, including 341 lncRNAs (lncARG).

We further analyzed the histone modification status of these lncARGs by querying published ChIP-seq data of various histone modifidcations. we observed an enrichment of chromatin marks associated with 'open' chromatin (H3K27ac, H3K4me1), active promoters (H3K4me3), and active transcription (H3K36me3). All of these modifications, together with Pol II occupancy, are recognized as manifestations of active gene expression (**Figure2.2E**).

Bromodomain and extra-terminal family proteins, such as BRD4, recognize acetylated histones and have been shown to promote AR transcriptional activity[34]. In accord with this, we observed the co-localization of BRD4 and AR protein at promoters of direct AR-responsive genes. Furthermore, treatment with a bromodomain inhibitor (JQ1) resulted in loss of AR binding at promoter regions of AR-responsive genes (**Figure 2.2C, E**).

Identification of AR-regulated lncRNAs from tissues

We next sought to determine whether ARGs identified from cell lines were also targeted by AR in normal prostate tissues and primary tumors. We leveraged the data set from reference [35] and queried for the presence of AR peaks within ARG promoters (**Figure**

2.3A). Remarkably, the majority of ARG promoters were TSS-proximally bound by AR in both tissues and cell lines. Conversely, AR-independent genes were distal to AR-binding sites (**Figure 2.3B-C**).

To confirm that the AR-regulated genes were also expressed in human prostate tissues, we interrogated RNA-Seq data from normal prostate, clinically localized prostate cancer (The Cancer Genome Atlas, TCGA)[36] and metastatic CRPC (Stand Up to Cancer-Prostate Cancer Foundation, SU2C-PCF)[31] (**Fig. 2.4A**). This revealed remarkable heterogeneity in the expression of ARGs during prostate cancer progression to metastatic disease. As expected, compared to protein-coding genes, non-coding ARGs were detected at lower overall levels, although ~10% of them showed robust expression of over 10 FPKM (fragments per kilobase of transcript per million mapped reads) on average across prostate cancer samples (**Fig. 2.4B**).

Prioritization of AR-responsive lncRNAs associated with cancer progression

We hypothesized that lncRNAs associated with prostate cancer progression and castration resistance should have differential expression levels between normal tissue and prostate cancer. Their expression is also expected to be AR-dependent and lineage-restricted if they are part of *bona fide* physiological feedback loops involved in AR signaling. Accordingly, a top-down strategy was developed to establish and prioritize clinically relevant, prostate cancer- and lineage-specific lncARGs.

First, we identified genes that were both directly regulated by AR in VCaP/LNCaP cell lines and upregulated in primary (**Figure 2.5A-B**) or metastatic prostate cancers (**Figure 2.5A-B**) compared to normal prostate tissues. Importantly, this approach highlighted several novel lncARGs, including *ARLNC1* (*ENSG00000260896, PRCAT47*), and validated previously identified lncARGs, such as *CTBP1-AS*[37]. Interestingly, *ARLNC1* was found to be one of the most differentially expressed AR-regulated genes in both localized and metastatic prostate cancers (**Figure 2.5B, D**).

Next, we sought to establish the prostate lineage and cancer specificity of prostate cancer-associated lncRNAs. We leveraged the MiTranscriptome assembly [10], an online resource, to interrogate lncRNA expression across a multitude of tissue and tumor types. We calculated sample set enrichment analysis (SSEA) scores, which indicate the strength of cancer and lineage association [10]. After applying an expression-level filter (10 FPKM at the 95th percentile), we identified 12 of the most prostate lineage- and prostate cancer-specific lncRNAs (**Figure 2.6A-C**); 5 of these lncRNAs were regulated by AR. Across these analyses, ARLNC1 was the top prioritized transcript and thus warranted further investigation.

Discussion

Previous study from our lab has identified 121 Prostate Cancer Associated Transcripts (PCATs), whose expression spectrums distinguished benign, localized, and metastatic stages of cancer. *PCAT-1* shows tissue-specific expression and is a transcriptional repressor

implicated in a subset of prostate cancer patients[38]. *SCHLAP1* is highly expressed in a subset of aggressive prostate cancers, and coordinates cancer cell invasion[7]. Lineage-specific lncRNA *PCAT14* predicts for prostate cancer aggressiveness and recurrence[12]. All of these studies initiated from differential expression analysis of prostate cancer samples, without focusing on critical cancer pathways.

In this study, we focused on the non-coding transcriptome regulated by AR signaling, the pivotal pathway that mediates prostate cancer progression. We found that compared to protein-coding genes, fewer number of AR-responsive lncRNAs exist, with a lower average expression level. Some AR-responsive lncRNAs are specifically expressed in prostate-lineage, and demonstrate differential expression patterns across benign, localized, and metastatic cancer stages. It is thus worth investigating whether these lncRNAs contribute to prostate cancer progression, what are the transcriptional machineries that drive their expression, and whether these lncRNAs are involved in a feedback or feedforward loop regulating AR signaling. In the following Chapters, we investigate these questions on a prioritized prostate cancer-specific, AR-responsive lncRNA, *ARLNC1*.

Materials and Methods

Cell lines

Cell lines were purchased from the American Type Culture Collection (ATCC) and maintained using standard media and conditions. All cell lines were genotyped by DNA fingerprinting analysis and tested for mycoplasma infection every two weeks. All cell lines

used in this study were mycoplasma-negative. For androgen stimulation experiments, VCaP and LNCaP cells were grown in charcoal-stripped serum containing media for 48 h and then stimulated with 10 nM DHT (Sigma-Aldrich) for 6 or 24 h.

RNA-Sequencing

Total RNA was extracted from LNCaP and VCaP cells following DHT treatment, using the miRNeasy kit (QIAGEN). RNA quality was assessed using the Agilent Bioanalyzer. Each sample was sequenced using the Illumina HiSeq 2000 (with a 100-nt read length) according to published protocols[38].

RNA-Seq data analysis to identify AR-regulated genes

RNA-Seq data were analyzed as previously described[39]. Briefly, the strand-specific paired-end reads were inspected for sequencing and data quality (for example, insert size, sequencing adapter contamination, rRNA content, sequencing error rate). Libraries passing QC were trimmed of sequencing adapters and aligned to the human reference genome, GRCh38. Expression was quantified at the gene level using the 'intersection non-empty' mode[40] as implemented in featureCounts[41] using the Gencode v22[42] and/or MiTranscriptome assemblies[10]. All pairwise differential expression analyses were carried out using the voom–limma approach[43, 44] with all default parameters. Relative expression levels (FPKMs) were normalized for differences in sequencing depth using scaling factors obtained from the calcNormFactors (default parameters) function from edgeR[45].

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

Identification of prostate cancer-associated protein-coding genes and lncRNAs

Raw RNA-Seq data for primary and metastatic patients were obtained from the TCGA/PRAD and PCF/SU2C projects, respectively. External transcriptome samples were reanalyzed using in-house pipelines (see above) to facilitate direct comparisons of expression levels and identification of DEGs. Pan-cancer analyses based on the MiTranscriptome assembly[10] were leveraged as FPKMs and enrichment scores (SSEA) were computed as part of that project.

To visualize data, fold changes were computed relative to median expression levels estimated across the combined (normal, primary, metastatic) cohorts and subjected to unsupervised hierarchical clustering separately within each cohort. Tissue lineage[20] and prostate cancer-specific genes were identified using the SSEA method as previously described[10]. Briefly, the SSEA test was used to determine whether each gene was significantly associated with a set of samples (for example, prostate cancer), or cancer

progression in a given lineage (for example, prostate normal to prostate cancer). The genes were ranked according to their strength of association.

ChIP–seq data analysis

ChIP–seq data from published external and in-house datasets, GSE56288 and GSE55064, were reanalyzed using a standard pipeline. Briefly, groomed reads (vendor quality control, adaptor removal) were aligned to the GRCh38 reference genome using STAR settings that disable spliced alignment: outFilterMismatchNoverLmax: 0.05, outFilterMatchNmin: 16, outFilterScoreMinOverLread: 0, outFilterMatchNminOverLread: 0, alignIntronMax: 1. Improperly paired alignments and non-primary alignments were discarded. Peaks were called using MACS2 (callpeak --broad --qvalue 0.05 --broad-cutoff 0.05 and callpeak --call-summits --qvalue 0.05)[46] and Q (-n 100000)[47]. ChIP enrichment plots were computed from alignment coverage files (BigWig[48]) as trimmed (trim = 0.05) smooth splines (spar = 0.05). The baseline (non-specific) ChIP signal was estimated from genomic windows furthest from the center of the queried region (peak summit, TSS) and subtracted from each signal before plotting.

Figures

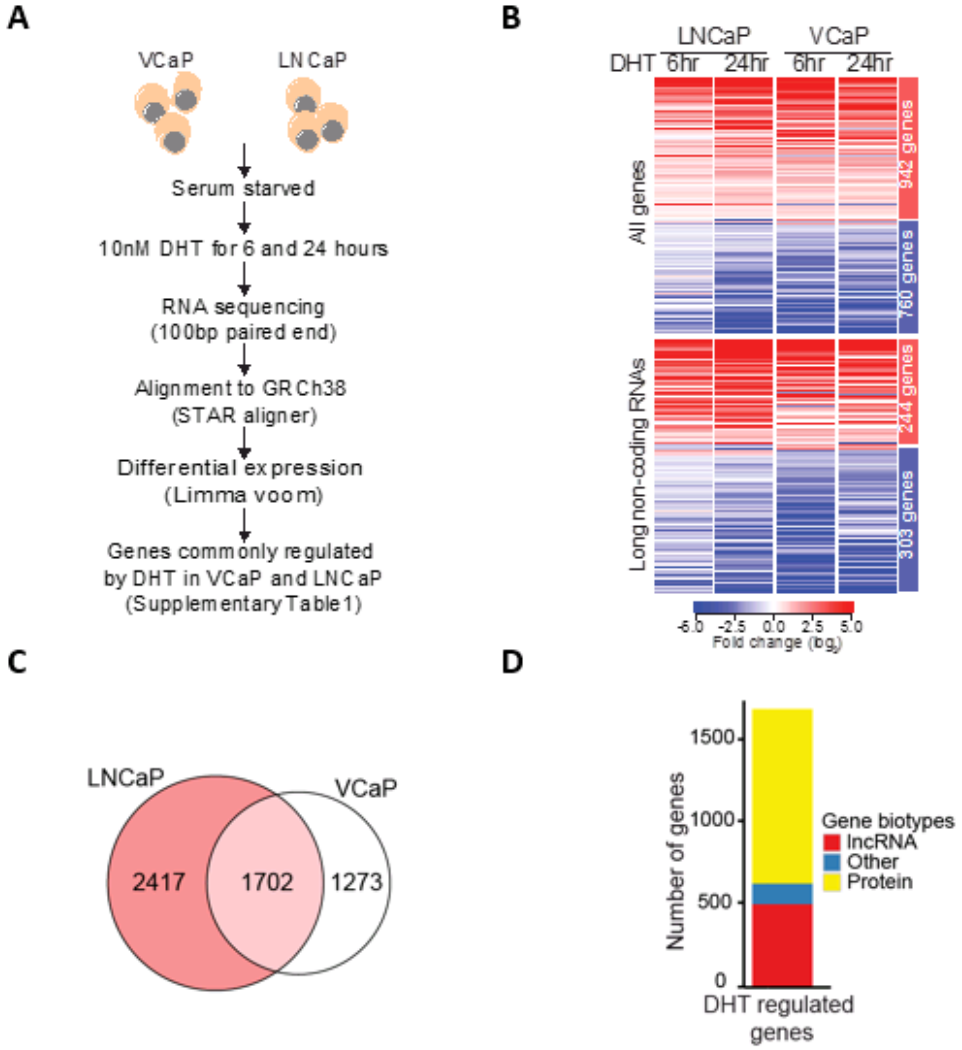


Figure 2.1. AR regulated non-coding transcriptome in prostate cancer cells
(**A**) A schematic illustration of the procedure used to discover AR-regulated genes (ARGs) in LNCaP and VCaP prostate cancer cell lines.
(**B**) The androgen-regulated transcriptome of prostate cancer cells. The heat map represents the 1,702 genes (including 547 lncRNAs) differentially regulated in LNCaP and VCaP cells following 6 and 24 h of DHT treatment
(**C**) Venn diagram indicating the overlap between AR-regulated genes in LNCaP and VCaP cells.
(**D**) Bar plot depicting the distribution of gene biotypes (protein, lncRNA, and other) of all overlapped ARGs identified in both LNCaP and VCaP cells.

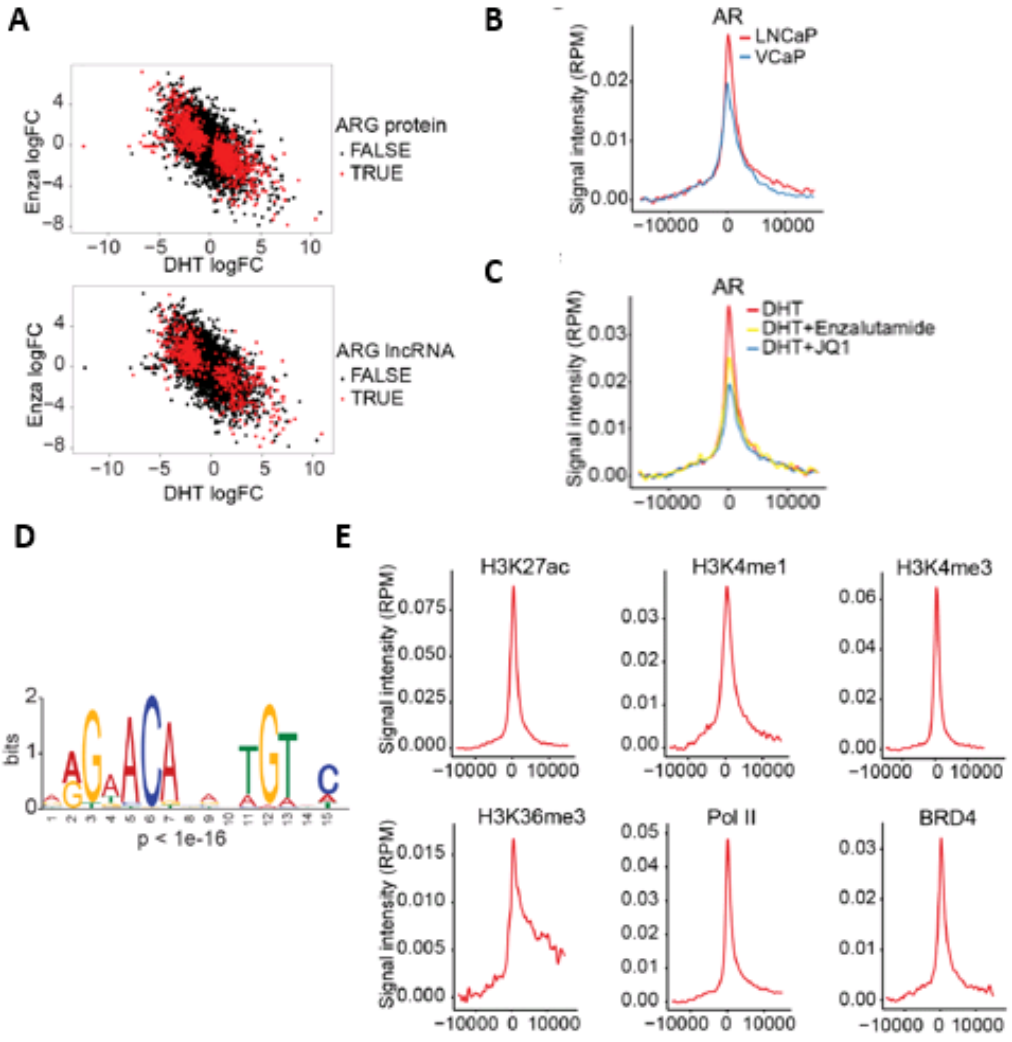


Figure 2.2 ARGs that are directly regulated by AR

(**A**) Transcriptional response to DHT and enzalutamide treatment in VCaP cells, plotting AR-regulated protein-coding genes (top) or AR-regulated lncRNAs (bottom).

(**B**)Aggregate ChIP–seq enrichment profile depicting AR ChIP–seq signaling density on ARG promoters in LNCaP and VCaP cells. Data representative of two biological replicates are shown.

(**C**) Aggregate ChIP–seq enrichment profile illustrating AR ChIP–seq signaling density on ARG promoters in LNCaP cells following DHT stimulation, AR antagonist (enzalutamide) treatment, or BRD4 inhibitor (JQ1) treatment. Data representative of two biologically replicated assays are shown.

(**D**) Motif discovery analysis of the top 250 AR ChIP–seq peaks on AR promoters identifies a binding motif similar to the canonical AR response element. P value was calculated in the MEME suite. n = 2 independent AR ChIP–seq assays.

(**E**) Aggregated ChIP–seq enrichment profiles depicting ChIP–seq signal density on direct ARG promoters for H3K27ac, H3K4me1, H3K4me3, H3K36me3, Pol II, and BRD4 in LNCaP cells.

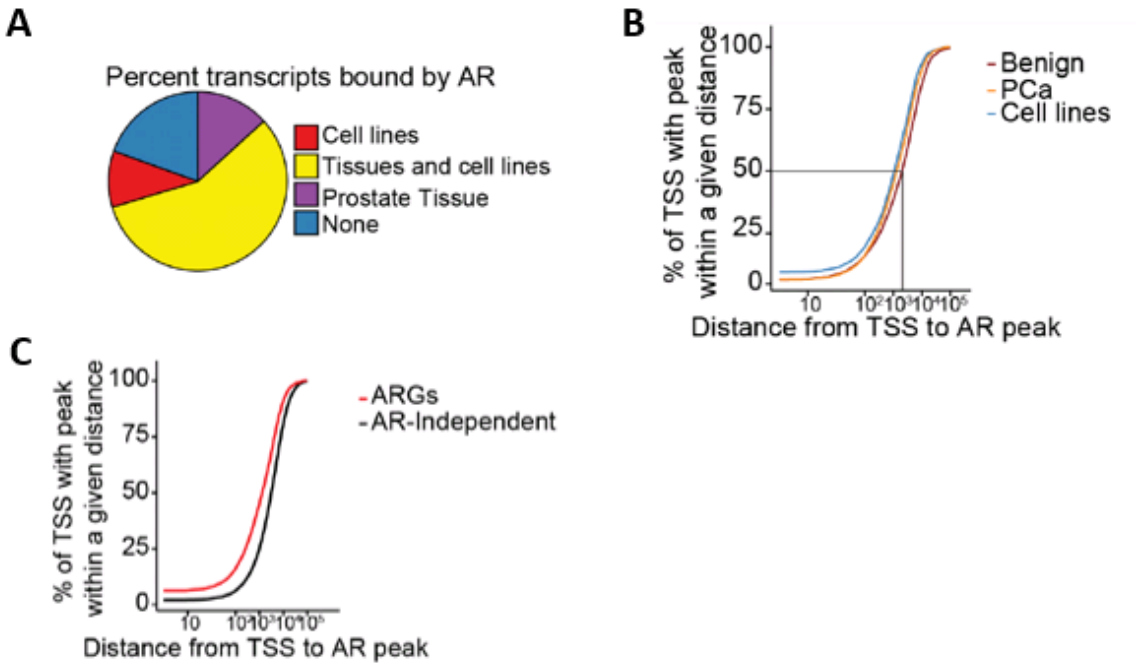


Figure 2.3 AR regulated genes in cell lines and clinical cancer samples
(**A**) Pie chart showing the prostate cell line or tissue distribution of direct ARGs with AR binding at transcription start sites (TSSs) in ChIP–Seq.
(**B-C**) Cumulative distribution plots of distances between the TSSs of genes and the nearest AR peak. **B**, AR binding near ARGs in benign prostate, prostate cancer tissues (PCa), and prostate cell lines. Approximately 50% (75%) of ARGs have an AR peak within 1 kb (10 kb) of the TSS. **C**, Comparison of distances between AR-binding sites for ARGs and genes not regulated by AR. Globally, ARGs are positioned closer to AR-binding sites.

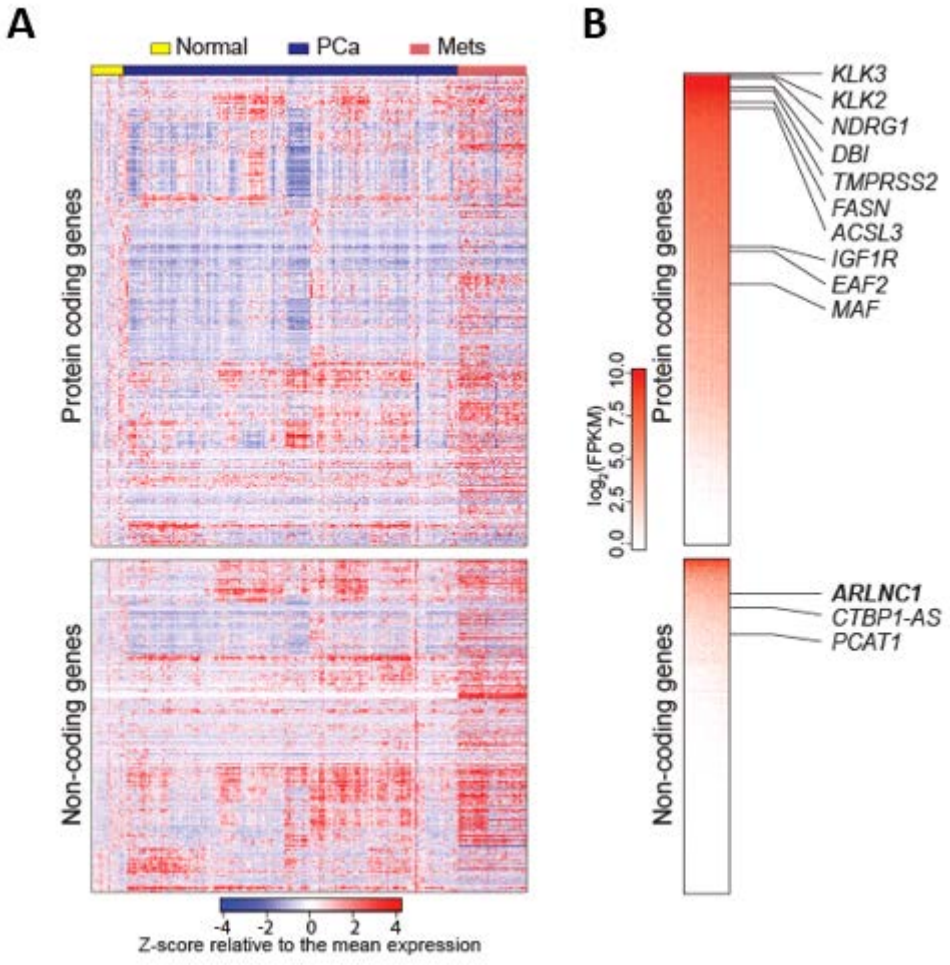
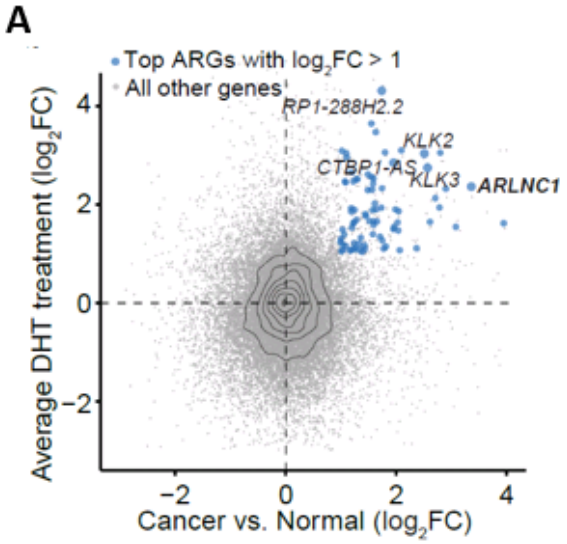


Figure 2.4 LncRNA expression in prostate cancer tissues

(**A**) The landscape of transcriptomic alterations of prostate cancer progression. The heat map depicts 1,155 protein-coding genes and 547 lncRNAs across benign prostate (normal, n = 52 samples), localized prostate cancer (PCa, n = 500 samples) and metastatic prostate cancer (Mets, n = 100 samples) in TCGA prostate and SU2C-PCF RNA-seq data, with rows representing genes and columns representing patients. Patients were grouped by clinical stage, and genes were subjected to hierarchical clustering. Expression variability is quantified for each gene as a z-score relative to the mean expression in normal prostate samples.

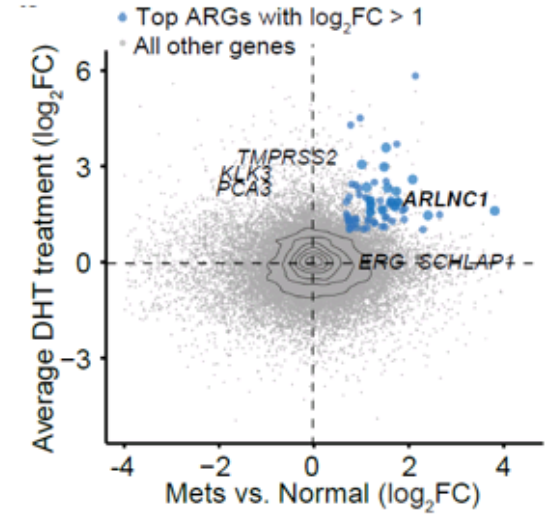
(**B**) A heat map representation of ranked median protein-coding gene and lncRNA expression levels in prostate tissues. Canonical prostate lineage and prostate cancer markers are listed.



B

Gene Name	Category	AR-peak	log_2FC
ARLNC1	LncRNA	Yes	3.373
KLK3	Protein coding	Yes	2.560
KLK2	Protein coding	Yes	2.506
CTBP1-AS	LncRNA	Yes	1.955
RP1-288H2.2	Read through	Yes	1.766
F5	Protein coding	Yes	1.759
STEAP4	Protein coding	Yes	1.710
AZGP1P1	Pseudogene	No	1.649
MYBPC1	Protein coding	Yes	1.635
SDK1	Protein coding	Yes	1.613

Cutoff: DHT stimulation >4 fold



D

Gene Name	Category	AR-peak	log_2FC
F5	Protein coding	Yes	2.330
FZD8	Protein coding	No	1.911
ARLNC1	LncRNA	Yes	1.666
RP11-244F12.3	Antisense	Yes	1.195
RP1-288H2.2	Read through	Yes	1.161
RP11-599B13.9	Pseudogene	Yes	0.449
SDK1	Protein coding	Yes	0.302
EFCAB12	Protein coding	No	0.256
LRIG1	Protein coding	Yes	0.225
GADD45G	Protein coding	Yes	0.136

Cutoff: DHT stimulation >4 fold

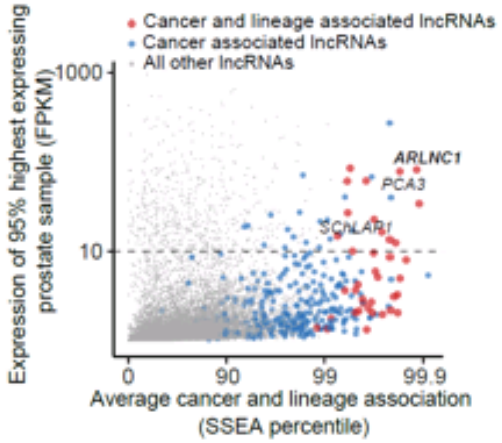
Figure 2.5 Identification of AR-regulated, prostate cancer–associated genes

(**A/C**) Identification of androgen-regulated transcripts elevated in prostate cancer progression. Scatterplots show the AR regulation and cancer association of the ARGs identified in **Figure 2.1**(**B**). The y axis depicts the log2-transformed fold change in gene expression following DHT stimulation, and the x axis indicates the log2-transformed difference in gene expression level between benign prostate (n = 52 samples) and localized prostate cancer (n = 500 samples) (**A**) and between benign prostate (n = 52 samples) and metastatic prostate cancer (n = 100 samples) (**C**). Significant genes with log2 fold change > 1 were ranked according to combined P value (limma-moderated t test).

(**B**) The top ten AR-regulated, localized prostate cancer–associated genes identified in (**A**), after applying an expression filter of at least fourfold change (log2 (FC) = 2) upon DHT stimulation and at least 1 FPKM average expression in prostate cancer tissues. Genes are listed in the order of log2 gene expression level difference between benign and localized prostate cancer tissues.

(**D**) The top ten AR regulated, metastatic prostate cancer–associated genes identified in (**C**), after applying an expression filter of at least four-fold change (log2 (FC) = 2) upon DHT stimulation and at least 1 FPKM average expression in prostate cancer tissues. Genes are listed in the order of log2 gene expression level difference between benign and metastatic prostate cancer tissues.

A



B

Gene	Coordinates (hg19)	SSEA%	FPKM95%	AR regulated
ARLNC1	chr16:80862547-80904394[-]	99.71	85.915	Yes
CAT2227	chr22:23880888-23890019[-]	99.45	75.85	No
PCA3	chr9:79379167-79413274[+]	99.15	73.283	No
PRCAT122	chr2:180977994-181006483[-]	99.42	73.189	No
PRCAT104	chr22:23869620-23890525[+]	99.38	44.449	No
PRCAT23	chr21:43005587-43005823[+]	99.89	30.741	Yes
PRCAT119	chr13:95925315-95940124[+]	99.42	21.935	Yes
PRCAT71	chr11:58805944-58806983[+]	99.69	19.574	No
SCHLAP1	chr2:181556702-181727451[+]	99.26	16.474	No
PRCAT55	chr11:58804433-58849219[+]	99.74	16.04	No
PRCAT44	chr2:172951763-172953352[-]	99.78	12.915	Yes
PRCAT38	chr21:42951648-42955549[+]	99.81	10.286	Yes

C

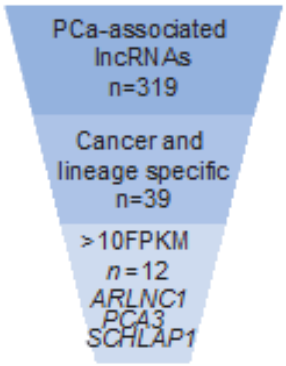


Figure 2.6 AR-regulated lncRNAs with restricted expression in prostate lineage
(**A**) Nomination of prostate cancer- and lineage-associated lncRNAs based on expression levels. The scatterplot shows the expression level, prostate tissue specificity and prostate cancer association of lncRNAs. The expression level is the FPKM value at the 95th percentile across TCGA prostate samples. Average cancer and lineage associations are represented by the percentile rank for each gene in SSEA (total n = 7,256 samples)

(**B**) The top 12 prostate tissue–specific, prostate cancer–associated lncRNAs identified in (A), after applying an expression filter of at least 10 FPKM in the prostate samples in the top 5% based on gene expression level. Genes are listed in the order of SSEA percentile (total n = 7,256 samples). (**C**) Schematic illustration of the procedure used to nominate prostate-lineage-specific, cancer-associated lncRNAs in prostate cancer.

Table

Table 2.1 AR-regulated long non-coding RNAs

gene_id	logFC	AveExpr	P.Value	Name	AR peak
G000029	-1.40273	4.512501	0.086322	RP11-54O7.16	TRUE
G000083	-0.93844	1.158834	0.09047		TRUE
G000413	2.459051	-3.02711	0.218751		FALSE
G000570	3.800678	-3.04578	0.503674	RP1-69M21.2,LOC101929181	FALSE
G000954	-2.8858	-1.53426	0.061614		FALSE
G000957	-3.28718	-3.10387	0.053328		FALSE
G000961	-2.3384	-4.10562	0.475695	LOC339505	FALSE
G000984	-2.06135	-2.16795	0.07318		TRUE
G001032	-2.60188	-3.54856	0.091678		TRUE
G001044	-1.85233	-2.28271	0.064382		FALSE
G001284	-0.03819	-1.98824	0.968998		TRUE
G001523	-2.07056	-2.70009	0.191786		FALSE
G001610	1.019695	4.203816	0.069391	TRIM62,AX747064	FALSE
G002113	-1.36784	-1.62679	0.130349		FALSE
G002293	-1.2983	-0.97782	0.059388	RP11-296A18.5	FALSE
G002951	-1.48863	-1.93426	0.069999		FALSE
G002957	-2.88601	-3.6998	0.118898		FALSE
G002979	2.291978	-4.23024	0.451315		TRUE
G003105	2.206414	-2.03532	0.46127	AC104169.1	TRUE
G003397	-2.59275	-2.23208	0.093728	PRCAT162	TRUE
G003443	2.870749	-0.38506	0.441531	CAT81	TRUE
G003511	-2.86857	-3.60086	0.041797		TRUE
G003728	-2.25656	-1.79939	0.198171		FALSE
G004008	-3.24034	-3.89471	0.037644	RP4-666F24.3	TRUE
G004771	2.057032	2.691291	0.253539		TRUE
G005055	0.938317	7.41409	0.056745	PEA15	TRUE
G005339	-3.44333	-0.17244	0.216331		FALSE
G005341	-2.17554	-4.10545	0.328533		FALSE
G005401	0.382604	6.797597	0.824107	ATP1B1,RP5-1018K9.1	FALSE
G005831	2.559296	-3.96092	0.355659	CAT142	TRUE
G006356	2.384253	-2.80508	0.326401	BRCAT400	TRUE
G006605	4.565885	-3.36451	0.33294		TRUE
G006817	4.543252	0.16941	0.069695		TRUE
G006819	6.186749	-0.34331	0.076236		TRUE
G006954	-2.23559	-3.25508	0.333857	HLA-AS1,RP11-295M18.6	FALSE
G007085	1.85509	-1.04499	0.055775		FALSE

G007098	-2.38812	-3.70988	0.114113	RP11-100E13.1	FALSE
G007159	-1.99968	-1.98837	0.03704		FALSE
G007163	-2.09422	-3.27298	0.062638		FALSE
G007383	-2.96705	-3.30843	0.105423		FALSE
G007610	-1.41944	-0.87436	0.086302	RP11-385F5.4	FALSE
G008007	-2.27686	-1.12569	0.161393	RP11-809C18.4	TRUE
G008037	5.790587	-2.77786	0.054362	HICLINC238	FALSE
G008042	5.162066	-1.63039	0.267279	PRCAT131	TRUE
G008058	3.21539	-1.70575	0.488045		TRUE
G008061	2.060401	-0.24004	0.570387		TRUE
G008276	-1.17264	3.449149	0.031712	ANKRD16	TRUE
G008509	2.108577	-2.07119	0.268352	RP11-397C18.2	TRUE
G008695	4.641644	-3.4018	0.213316		FALSE
G008815	3.978471	-3.31978	0.130609	LINC00836	TRUE
G009117	4.410848	-3.60583	0.17427	OVAT13	TRUE
G009153	-0.68824	-2.76147	0.761584	PRCAT146	TRUE
G009673	-2.27416	2.703485	0.043651	CAT1262	FALSE
G009760	2.838842	-1.11348	0.091931	LINC00844	TRUE
G009961	-0.96642	1.763986	0.08088	TET1,RP11-119F7.5	FALSE
G010213	1.614178	7.878369	0.109639	VCL,RP11-178G16.5	TRUE
G010922	-3.03621	-3.99695	0.09751		FALSE
G011318	-3.18849	-4.94246	0.167603		FALSE
G011552	3.791503	-1.79146	0.138597	HICLINC252,LINC00867	TRUE
G011599	-2.04286	-1.40846	0.21971	CAT1308	TRUE
G011673	-1.20567	2.059589	0.023432		TRUE
G012015	-1.88167	-1.00968	0.311504	RP11-140A10.3	TRUE
G012036	2.073307	-0.64926	0.561769	CAT1320	FALSE
G012133	-2.19959	-2.99893	0.158982	RP11-326C3.7	FALSE
G012156	-0.17914	0.022055	0.805833		TRUE
G012515	-2.44759	-0.49534	0.106552	CTD-2516F10.4	TRUE
G012518	-2.15444	-3.24177	0.056549	CAT1334	FALSE
G012596	-2.34963	-2.81748	0.025585	RP11-540A21.2,LOC440028	TRUE
G012600	-2.72678	-2.75043	0.107964	SBF2-AS1	FALSE
G012645	-2.18801	0.195019	0.31606	PRCAT107	FALSE
G012962	-2.70445	1.886309	0.039751	RP11-945A11.1	TRUE
G012972	-2.09999	-3.60996	0.179782	RNU6-783P	FALSE
G013149	-0.58692	5.368326	0.096565	PIGCP1,CSTF3	FALSE
G013177	-2.12665	-2.47174	0.101613	THCAT62	TRUE
G013377	-1.86659	-3.08754	0.091945		FALSE
G013394	0.5421	-2.14031	0.643857	CAT1352	FALSE
G013729	2.539392	-1.78995	0.157474	RP11-872D17.4	TRUE

G013731	-2.15881	-3.39243	0.086878		FALSE
G013895	-2.09179	-2.79967	0.136017		FALSE
G014019	1.121222	0.722989	0.08334	RP11-727F15.9	FALSE
G014366	-1.54226	2.600475	0.099778	CAT1378	TRUE
G014393	3.8621	-2.52079	0.345474		TRUE
G014395	3.510878	-2.53966	0.455489		TRUE
G014426	5.539431	-2.97535	0.18408		TRUE
G014440	2.115131	0.086668	0.061658		TRUE
G014443	-1.3306	-0.7857	0.065621	RP11-554A11.9	FALSE
G014526	2.94667	-3.72043	0.355806		TRUE
G014752	2.49067	-4.80177	0.422102		TRUE
G014814	1.174462	5.985298	0.11751	RP11-21L23.2,TSKU	TRUE
G014875	-1.53762	-1.31705	0.090784	KCTD21-AS1	FALSE
G015185	6.297588	1.293683	0.094008		TRUE
G015244	-2.06054	-2.94271	0.066042		FALSE
G015250	1.117356	1.900357	0.083539	CAT1412	FALSE
G015388	-1.64346	-1.26584	0.198222	CAT1418	FALSE
G015569	-2.26587	-3.17946	0.036842		FALSE
G015689	2.276888	-3.25057	0.166531		TRUE
G015705	-2.19318	-4.02178	0.281674		FALSE
G015801	-2.10643	-0.40878	0.247464		TRUE
G016087	2.246145	-2.93804	0.19288		TRUE
G016096	-1.955	-1.62232	0.075791	RP11-677M14.2	FALSE
G016104	-1.21398	1.317699	0.066136		FALSE
G016157	-2.75581	-1.14453	0.397584		TRUE
G016301	2.354383	-3.21515	0.145715		TRUE
G016346	-2.46747	0.864171	0.083108	BRCAT168	FALSE
G016393	4.413411	1.144751	0.01745		FALSE
G016394	5.944861	0.734599	0.036705		FALSE
G016562	-2.39152	0.138731	0.014567	LSCAT213	TRUE
G017565	-0.87589	1.147721	0.097781	DDX11-AS1	TRUE
G017746	3.083808	-4.30591	0.467076	RP11-554L12.1	TRUE
G017755	0.274128	6.02902	0.556552	KIF21A	TRUE
G017973	-3.2876	-3.92572	0.057536		FALSE
G018118	-2.089	0.477997	0.008888	RP11-133N21.7	FALSE
G018251	3.894872	-1.62045	0.215083	CAT1519	FALSE
G018313	1.618524	9.212634	0.02132	KRT18	TRUE
G018376	-1.86618	0.757625	0.416396	HOXC12	FALSE
G018377	-2.5219	-2.47677	0.09772	HOTAIR	FALSE
G018403	-2.23539	-1.19372	0.386618	PRCAT266	TRUE
G018424	-2.81916	-4.32886	0.182785	RP11-616L12.1	TRUE

G018646	-2.49969	-3.69424	0.132005	RP11-150C16.1	TRUE
G019202	-2.83338	-4.36607	0.134294		FALSE
G019273	-2.33538	2.333406	0.477252	CAT1556	TRUE
G019355	2.712525	-2.9699	0.292885	THCAT348	TRUE
G019413	-2.13267	-3.56224	0.087021	THCAT117	TRUE
G019524	3.898204	-2.64033	0.097695	RP11-1060G2.1	TRUE
G019603	-2.54518	-0.92217	0.023421		TRUE
G019709	2.126962	-2.34201	0.174618		TRUE
G019738	2.228846	-2.00067	0.316086		TRUE
G020410	-2.50675	-1.59928	0.033233	RP11-340F14.5	TRUE
G020732	-2.47975	-3.62988	0.389865	RP11-474D1.4	TRUE
G021007	-1.40868	-0.27431	0.039187		TRUE
G021008	-1.00041	2.048907	0.077388		FALSE
G021168	-2.41102	-1.49025	0.276367	RP11-556N21.1	FALSE
G021193	3.133273	-3.19062	0.201823		TRUE
G021396	-2.25568	-3.45376	0.140427		FALSE
G021791	2.545117	-1.23567	0.121584		TRUE
G021795	-2.08065	-4.73708	0.461419	PNAT43	FALSE
G022457	2.124815	-1.39562	0.602168		FALSE
G022968	-0.13757	-0.86369	0.90107	CVAT23	FALSE
G023010	3.433052	-2.38729	0.389409	PRCAT119	TRUE
G023011	2.508325	2.481916	0.026529		FALSE
G023160	1.218719	4.117183	0.092673	CAT1629	TRUE
G023293	-1.67437	-1.94169	0.095162	LINC00443, CAT1630	FALSE
G023299	-2.77178	-3.30031	0.220396		TRUE
G023358	0.244029	-2.84844	0.918259		TRUE
G023426	2.165865	0.632621	0.30537		TRUE
G023500	-1.98748	-2.16237	0.275833	KHCAT425	FALSE
G023634	3.033478	-4.16511	0.376117	CTD-2311B13.5	TRUE
G023653	3.149136	-4.8208	0.252929	RP11-244H18.1	TRUE
G024250	1.995911	-3.31086	0.411596		FALSE
G024296	-1.28195	-0.18703	0.120839	RP11-356O9.2	TRUE
G024330	1.100132	-2.91931	0.553886	RP11-545M17.1	FALSE
G024513	1.318476	3.700693	0.02454	METAZOA_SRP	TRUE
G024748	-1.5747	0.018065	0.114371	KTN1-AS1	FALSE
G024875	-2.22908	-1.92311	0.282515	CAT1672	FALSE
G024927	7.287695	-1.0292	0.0697	CAT1675	TRUE
G025353	-1.4429	-0.973	0.051318	CAT1690	FALSE
G025411	-2.45281	-3.22515	0.097509	PNAT35	TRUE
G025547	-1.51519	-2.83162	0.093512		TRUE
G025778	-1.63215	-0.8919	0.072199	RP11-1078H9.5	FALSE

G025786	-1.66493	-1.88622	0.085561	THCAT584	FALSE
G026094	1.197943	4.624136	0.034955	EML1,RNU1-47P	FALSE
G026238	-2.64345	-3.00103	0.224774		FALSE
G026470	1.839078	2.537458	0.396469	LACAT82	TRUE
G026630	3.464848	-1.11123	0.011904		TRUE
G026637	3.802638	-3.28186	0.100706	RP11-73C9.1	TRUE
G026770	-2.07999	-1.00441	0.303534		FALSE
G026838	-1.71555	0.514114	0.031741	RP11-16E12.1	FALSE
G027262	-0.80142	3.030266	0.093148		FALSE
G027289	2.148715	-3.63384	0.395701	Y_RNA	FALSE
G027363	-1.6571	-2.49369	0.087245	CMAT29	FALSE
G027732	-2.77653	0.806557	0.020331	RP11-82L7.4	TRUE
G027831	-1.38642	-1.63598	0.076946		TRUE
G028003	3.573007	-0.35107	0.107095	ITGA11,RP11-709B3.2	FALSE
G028313	-1.62027	-1.91996	0.085827	DNM1P35	TRUE
G028326	-1.67163	-0.89168	0.043935	RP11-593F23.1	FALSE
G028517	-1.20337	2.402392	0.098477	CAT1779	TRUE
G028795	-2.57497	-3.6042	0.135936	MEAT119	FALSE
G028850	-1.60316	-2.32207	0.063746	PRC1-AS1	FALSE
G029034	3.121861	-3.98315	0.207632		TRUE
G029104	1.550068	6.509867	0.007771	CAT1795	TRUE
G029150	-1.72605	-3.39836	0.219192	MIR1302-10	TRUE
G029156	-1.56768	0.278262	0.033168	WASIR2	TRUE
G029543	1.014125	1.452777	0.05825		TRUE
G029836	3.45029	0.089496	0.010351	RP11-65J21.4	TRUE
G030045	3.471665	-3.42146	0.298081	RP11-429K17.1	FALSE
G030443	2.946707	-1.68009	0.250879	ESAT93	TRUE
G030445	2.224184	-0.50135	0.180222		TRUE
G031234	0.551103	5.242305	0.503294	MT1X,DPPA2P4	TRUE
G031345	2.527321	-3.48937	0.17547	CRAT4	TRUE
G031570	-1.04742	0.727521	0.082053		FALSE
G031811	1.273835	6.310542	0.039968	ZFHX3	TRUE
G031940	1.626879	-1.48317	0.441982		TRUE
G032033	2.345185	0.803853	0.122124	PRCAT47, ARlnc1	TRUE
G032157	4.281667	0.584181	0.042052	KCCAT592	TRUE
G032161	2.338742	-1.90459	0.644724	KCCAT103	TRUE
G032162	4.097745	-2.91283	0.459008		TRUE
G032200	-1.43318	-2.86358	0.302731		FALSE
G033025	-3.87994	-3.86784	0.029689	LINC00675	TRUE
G033054	1.332873	3.876648	0.088487	ZNF18,RPL21P122	TRUE
G033966	2.897175	-2.71539	0.244139	AC015849.12	TRUE

G034483	0.18823	2.366982	0.868899	MEAT128	FALSE
G035180	1.009611	6.100407	0.063386	VMP1	TRUE
G035181	1.186456	-0.73564	0.098795	CAT1958	FALSE
G035267	-0.40137	0.294888	0.703085		FALSE
G035370	-1.36513	-0.42872	0.05856		FALSE
G035430	2.52264	-0.99598	0.153784		FALSE
G035432	2.367083	-4.24796	0.381265		FALSE
G036007	0.786606	6.303478	0.057621	RP11-353N14.5,TBC1D16	TRUE
G036055	-2.07652	-2.76204	0.121124	CTD-2561B21.3	TRUE
G036065	-2.80834	-4.19604	0.151915		TRUE
G036093	-2.33215	-3.37839	0.117876	THCAT107	TRUE
G036240	-1.39938	2.118313	0.055144	LACAT117	TRUE
G036242	2.474863	-2.18838	0.150307		TRUE
G036277	-1.86196	-2.92347	0.094757	RP11-14P20.1	TRUE
G036290	-1.06489	3.576236	0.022244		FALSE
G036314	-3.15727	-3.28551	0.127509	LINC00470,RP11-288C17.1	TRUE
G036657	-1.19448	2.681492	0.04639	RP11-861E21.1	FALSE
G036676	-1.47541	-1.37018	0.147752		TRUE
G036767	3.442996	-3.27553	0.106789	MEAT86	TRUE
G036901	-1.13203	-0.85951	0.205868	RP11-621L6.3	TRUE
G036953	5.085424	-1.76239	0.019978	AQP4-AS1	TRUE
G036958	4.50091	-2.77922	0.368466	AQP4-AS1	TRUE
G037652	2.803241	-3.5241	0.33847		FALSE
G037664	1.710822	5.437217	0.003608	RP11-108P20.1,MALT1	TRUE
G037684	1.926495	0.43449	0.209754		TRUE
G037920	-1.26768	0.437091	0.056127	DSEL,CTD-2541J13.1	FALSE
G038047	6.965749	-2.06305	0.014509	RP11-126K15.1	TRUE
G038054	2.877181	-3.36277	0.505406		TRUE
G038563	-2.03726	-3.30387	0.370064		TRUE
G038694	-0.95685	1.410002	0.059143	KCCAT275	FALSE
G039235	-1.43563	-2.0681	0.373277	CTD-2192J16.11	FALSE
G039445	3.168078	0.166061	0.309987		FALSE
G039681	-2.00896	-3.4696	0.22499		TRUE
G039774	-1.48687	-0.95995	0.295909		FALSE
G040328	-1.62976	-2.83756	0.094471		FALSE
G040376	3.069986	-3.29013	0.098239		TRUE
G040593	-0.89947	4.930078	0.08749	LRFN1,CTC-246B18.8	FALSE
G040635	0.733791	6.497617	0.086089	ZNF546,PSMC4,CTC-471F3.6	FALSE
G040900	-1.08781	0.817244	0.159522	RP11-15A1.2	FALSE
G041027	-0.13484	0.689709	0.866395	CAT2113	TRUE
G041349	1.336208	9.323637	0.280006	CTD-2568A17.1	TRUE

G041363	-3.05623	-3.44812	0.267831	CTB-147C22.8	FALSE
G041739	2.133448	-1.78797	0.140128		TRUE
G041868	0.719199	-3.57458	0.680448	CTC-444N24.13	TRUE
G041938	-1.50138	-2.54004	0.288694	CTD-2619J13.19	FALSE
G042070	-2.07643	-2.07349	0.097847	AC018685.1	FALSE
G042286	2.230012	0.578907	0.022556		FALSE
G042341	-1.53572	1.046452	0.194136	CAT207	TRUE
G042418	-0.19303	-3.23738	0.935638		TRUE
G042532	2.875899	-3.44589	0.368418	PRCAT126	TRUE
G042596	-1.40339	-1.03225	0.075586	AC079145.4	FALSE
G042696	2.075451	-2.968	0.226001	BRCAT387	FALSE
G042906	2.934558	-3.05154	0.273422		TRUE
G043040	-2.10741	-3.45717	0.204801	AC020594.5	FALSE
G043304	-1.75332	0.809875	0.052496	KHCAT367	TRUE
G043436	3.05746	-3.71492	0.232877		FALSE
G044809	-1.48959	-1.46217	0.150655	KCCAT298	FALSE
G044873	-2.45026	-3.67847	0.210015		TRUE
G045024	-2.62353	-3.26382	0.142153		TRUE
G045282	3.615251	-1.96762	0.087628	CRAT41	TRUE
G045317	-1.41141	1.836644	0.07871	CAT288	TRUE
G045585	-1.22959	-2.52117	0.499405	HICLINC62	FALSE
G045881	2.865413	-3.00706	0.435629	OVAT98	TRUE
G045939	1.981531	0.862341	0.010671	AC018804.3	FALSE
G045973	2.422984	-1.26897	0.081595		TRUE
G045991	2.85109	-3.22869	0.12512		FALSE
G046013	1.976443	-0.64193	0.05982		TRUE
G046046	-1.49022	-1.903	0.393657	RP11-725P16.2	FALSE
G046160	-1.4184	-0.6794	0.099678		FALSE
G046325	3.055415	-4.01242	0.241941		FALSE
G046493	-2.15339	-2.53194	0.598985	AC061961.2	FALSE
G046726	-1.58491	0.370166	0.06561	PRCAT52	TRUE
G046755	2.128283	-3.07751	0.175375		FALSE
G046789	-2.23952	-0.47632	0.315622		TRUE
G046869	4.20075	2.896508	0.212228	PRCAT44	TRUE
G047119	-2.12767	-3.57564	0.537926		FALSE
G047196	0.672998	-3.38796	0.71702	RP11-438L19.1	FALSE
G047232	-2.19318	-4.02178	0.281674	AC007966.1,AC008174.3	FALSE
G047342	-2.26832	-3.47495	0.058665	HNCAT25	FALSE
G047447	3.014398	-2.5332	0.388046		TRUE
G047809	1.431345	2.275725	0.091021	CCNYL1	TRUE
G047838	1.475707	8.779537	0.021354	IDH1,AC016697.2	FALSE

G048220	3.123402	-4.45069	0.237742		TRUE
G048744	-2.03864	-2.59131	0.150333		FALSE
G048958	-1.56019	-2.28737	0.090733	AC114730.5	FALSE
G049118	-2.48271	-3.04456	0.035821		FALSE
G049125	2.021664	-2.75568	0.120513		TRUE
G049439	-3.31715	-2.24029	0.357632	RP11-157E14.1	FALSE
G049641	-2.1155	-3.78793	0.160757	ESAT110	FALSE
G049678	3.583105	-2.4339	0.285866	LGAT90	FALSE
G049687	4.695585	-2.83402	0.054085		TRUE
G050039	-1.52393	-0.32925	0.185747	OVAT22	FALSE
G050719	-1.95044	-2.85918	0.096711		FALSE
G051083	2.823898	-0.54529	0.292117	CAT2184	TRUE
G051095	4.115547	-3.04448	0.224061		TRUE
G051142	-3.86036	-2.7866	0.110614	OVAT151	FALSE
G051303	-2.79881	-1.26763	0.15726		FALSE
G051309	-2.00204	-1.35536	0.065133	RP11-93B14.5	TRUE
G051383	-1.13174	2.85313	0.093204		TRUE
G051436	0.670692	6.898255	0.08565		TRUE
G051610	-1.18213	4.322191	0.099994	LINC00478	TRUE
G052017	-2.68284	-3.05531	0.160349	HUNK-AS1	TRUE
G052077	2.060642	-0.97585	0.240257	IL10RB-AS1	TRUE
G052134	5.596326	-2.93806	0.09666	LINC00160	TRUE
G052135	3.170495	-3.91367	0.274146	CAT2206	FALSE
G052222	3.065809	-3.68781	0.276546		TRUE
G052331	2.865746	-2.2517	0.199216	PRCAT17	TRUE
G052333	5.320418	2.821257	0.03574	PRCAT38	TRUE
G052334	6.057732	0.667845	0.085387	PRCAT23	FALSE
G052336	1.490611	-0.14263	0.063174	PRCAT82	TRUE
G052447	-1.85491	-1.87343	0.072066		FALSE
G052460	-2.05228	-2.84222	0.045093		FALSE
G052516	-1.69753	-0.43256	0.05663		FALSE
G052527	-1.34068	0.612529	0.063419	AL773604.8	TRUE
G052551	-1.80575	-1.43595	0.088799	KCCAT638	TRUE
G052610	-0.82027	2.094316	0.089199	MCM3AP-AS1	TRUE
G052613	-1.09912	3.297382	0.033729	MCM3AP-AS1	FALSE
G052646	2.67517	-4.00997	0.431947	POTEH-AS1	FALSE
G052797	-1.5264	-0.60521	0.061651	CAT2223	TRUE
G052848	-0.23056	-1.76755	0.91422		FALSE
G052853	-2.07322	-2.98275	0.212562	LINC00896	FALSE
G052925	4.004657	-3.69051	0.419482	KB-1592A4.15,AL117485	FALSE
G053066	-1.17104	-0.4974	0.084177		FALSE

G053485	3.940381	-1.82414	0.131197	RP1-127L4.10	TRUE
G053486	3.95116	-2.57176	0.348021		FALSE
G053607	-1.9879	-2.00466	0.062271		TRUE
G053674	-1.35937	2.732232	0.063884	ELFN2,DQ594725	FALSE
G053708	0.897047	4.523936	0.066046	ANKRD54	TRUE
G053961	-2.00003	-1.8589	0.059894	CTA-126B4.7	FALSE
G054033	-1.00494	0.609311	0.091656	KHCAT97	FALSE
G054063	2.020559	-2.36774	0.069587		TRUE
G054130	3.601571	0.372222	0.058199		TRUE
G054131	3.742133	1.48255	0.048349		FALSE
G054133	3.088103	-0.00383	0.103218	LOC730668	FALSE
G054434	-1.30771	3.073655	0.070977	RP11-97C16.1	FALSE
G054452	-1.23101	0.308185	0.054907	ITPR1-AS1	TRUE
G054486	2.114731	-5.00623	0.514695		FALSE
G054811	-1.47875	-0.72716	0.049432	CAT389	FALSE
G055301	2.193166	1.035908	0.077851	AC097359.1	FALSE
G055447	2.18231	-0.22069	0.348492		TRUE
G055462	-1.48253	-1.52168	0.084232	CAT400	FALSE
G055480	-2.09565	-3.05617	0.082517		FALSE
G055896	-2.24116	-2.82255	0.156149	CAT410	FALSE
G055988	-2.02702	0.291157	0.412161		TRUE
G056035	2.86229	-3.84176	0.338026	KCCAT260	TRUE
G056292	4.627277	-2.40154	0.239468	LVCAT46	FALSE
G056935	1.595333	4.487098	0.015232	LINC00883	TRUE
G057053	-3.32119	-1.23167	0.250988	RP11-90K6.1	TRUE
G057062	-2.24683	-1.09445	0.195453	RP11-572C15.6	TRUE
G057244	-2.22401	-3.28893	0.079555		FALSE
G057283	3.118228	-2.36212	0.294381		FALSE
G057284	4.159438	-2.31583	0.165573		FALSE
G057398	-2.02956	2.014061	0.156957	RP11-379B18.5	TRUE
G057514	0.823954	8.416813	0.084521	RPN1,POU5F1P6	FALSE
G057557	0.551371	5.91327	0.071874	,RPL32P3,BX641154	TRUE
G057593	-1.15646	3.523208	0.049336	LACAT17	FALSE
G057924	3.748474	1.53674	0.353552	CHST2	FALSE
G058210	-2.38522	-3.00162	0.040887	RP11-529G21.2	TRUE
G058215	-3.18849	-4.94246	0.173549	KCCAT39	FALSE
G058256	1.876028	-0.73964	0.280904	ESAT16	TRUE
G058489	-2.23338	4.948063	0.148988	SI,RP11-747D18.1	FALSE
G058547	-1.98705	-1.53175	0.224061		TRUE
G058643	-1.55068	-3.27553	0.189117	RP11-185E8.1	FALSE
G058722	-2.42901	-1.17365	0.164835	PRCAT24	TRUE

G059025	0.327163	-2.93106	0.803086		TRUE
G059106	-2.55484	-3.71656	0.341642	RP11-48F14.2	TRUE
G059324	-0.41752	-4.03768	0.849696		FALSE
G059369	-1.32567	-1.29151	0.081806		FALSE
G059439	1.576466	0.487942	0.159315	LINC00969	FALSE
G059452	2.243055	2.435608	0.10459	KCCAT630	FALSE
G059663	-0.4025	2.673994	0.574967	LGAT106	TRUE
G059668	0.199979	5.311342	0.780097	KPCAT2	TRUE
G059676	3.116358	4.254516	0.005154	HV535487	TRUE
G059677	2.854194	1.876596	0.004898	CTBP1-AS	FALSE
G059682	-0.57075	-0.90487	0.570524		TRUE
G059745	-2.01705	-2.99935	0.096561		FALSE
G059823	-0.92646	2.304127	0.429881	AC226119.5	TRUE
G059855	2.776496	-2.95286	0.15434		TRUE
G059857	3.009263	0.351528	0.121889		TRUE
G059864	-2.62442	-1.71271	0.044566		FALSE
G059867	-2.04011	-3.5985	0.151226		TRUE
G059984	-1.63372	-1.84546	0.06796	HNCAT68	TRUE
G060068	1.156664	7.851356	0.057055	RP11-448G15.3	TRUE
G060215	-1.10908	1.672588	0.068804	TAPT1-AS1	TRUE
G060776	-0.22363	-2.37947	0.851793		FALSE
G061287	-2.35391	-2.29708	0.404143	RP11-725D20.1	TRUE
G061455	-2.13087	-1.86799	0.245056		TRUE
G061709	-2.06254	-3.4082	0.154557	CAT561	FALSE
G061721	2.432969	-2.11548	0.10743	AC107072.2	FALSE
G062022	2.070118	-3.44744	0.324685	RN7SKP244,7SK	TRUE
G062025	1.99341	0.323828	0.045208	FAM13A-AS1	FALSE
G062041	-1.24791	2.163784	0.043887	RP11-115D19.1	TRUE
G062057	-2.4536	-0.79629	0.032556		FALSE
G062116	3.224245	-3.74087	0.165206	CAT574	TRUE
G062224	-3.09171	-2.29607	0.161628		FALSE
G062381	-2.99751	-3.69894	0.240354		FALSE
G063265	4.238715	-3.76134	0.222735	PRCAT187	TRUE
G063348	2.462588	-3.80905	0.429552	CAT594	TRUE
G063514	-2.00217	-3.12989	0.045588	CAT600	FALSE
G063530	-1.12901	5.420321	0.06987	CVAT11	TRUE
G063618	-1.23932	0.083001	0.067538	RP11-218F10.3	FALSE
G063845	3.060114	4.933645	0.276964	RP11-440I14.2,HPGD	FALSE
G063846	2.607233	-1.14629	0.318449	RP11-440I14.3	FALSE
G064103	1.482696	1.036643	0.062366	RP11-714G18.1,SNX25	FALSE
G064346	2.365893	-4.13703	0.347141		TRUE

G064797	2.500019	4.989899	0.023953	CTD-2165H16.3	TRUE
G064843	-1.86622	-1.24572	0.097121	RP11-260E18.1	FALSE
G065393	-2.86536	-1.69694	0.069894	CTD-2116N24.1	FALSE
G065410	1.232887	2.298507	0.09064	LIFR-AS1	TRUE
G065433	-2.36333	-4.52847	0.439482		FALSE
G065553	1.230315	2.505369	0.098879	CAT655	TRUE
G065780	-2.57794	-2.46459	0.048062		FALSE
G065803	2.674788	-2.26336	0.161225		TRUE
G065808	-1.64159	-0.19811	0.075878	CTD-2031P19.3	TRUE
G065880	-2.25392	-2.65758	0.141745	LOC101928569	TRUE
G066769	1.867301	-1.61259	0.191932		TRUE
G066890	2.247371	-4.06804	0.588654		FALSE
G067016	0.673261	6.650198	0.093223	PJA2,CTD-2587M2.1	TRUE
G067899	0.212064	-3.79722	0.914224		FALSE
G068067	2.661636	-2.58203	0.06163		TRUE
G068470	-2.90086	-1.77043	0.068239	CTC-340A15.2	TRUE
G068471	-3.49602	-3.62172	0.020237		TRUE
G068966	-2.1948	-3.21672	0.138541		TRUE
G069047	3.723394	-3.41632	0.054125	CAT756	TRUE
G069069	-1.04491	1.427673	0.065648	CTC-338M12.5	TRUE
G069122	-2.09052	-2.77542	0.170315	LACAT77	TRUE
G069259	1.355817	-1.39144	0.091775		FALSE
G069472	2.082248	-4.04617	0.409378		TRUE
G070137	-2.18119	-3.50696	0.142715	HCG14	TRUE
G070488	1.302193	0.171372	0.082339	HCG25	TRUE
G070556	-2.07182	-3.51849	0.147742		TRUE
G070610	6.29957	-2.30556	0.485713	CAT795	TRUE
G070612	4.554682	9.215672	0.002644	FKBP5,LOC285847	TRUE
G070619	4.344905	-1.84839	0.105973		TRUE
G070798	-3.45864	-2.94729	0.023295	HNCAT211	FALSE
G070801	-2.21155	-2.71656	0.133769		TRUE
G070950	2.271519	-3.18429	0.182763		TRUE
G070968	3.197496	0.030825	0.046615		FALSE
G071242	1.159907	3.212114	0.036001	OVAT19	TRUE
G071271	4.587036	-3.60996	0.029314	LSCAT162	TRUE
G071600	-0.48331	-2.32494	0.875622		FALSE
G072429	-1.60635	-1.44244	0.057898		TRUE
G072446	3.689979	-4.08736	0.169976		TRUE
G072863	3.135769	-2.24069	0.077282	BRCAT246	FALSE
G072921	-0.07872	-3.29345	0.97268		FALSE
G073186	1.701079	-0.69226	0.259209	RP11-356I2.4	TRUE

G073824	3.121558	-1.8858	0.205041		TRUE
G074015	-2.34359	-3.14536	0.411657		FALSE
G074101	2.902999	-0.20088	0.032725		TRUE
G074104	2.235417	-2.53504	0.417475		FALSE
G074149	-1.58188	0.801403	0.032755	RP1-266L20.4	FALSE
G074385	2.817504	-4.57984	0.357059	KHCAT361	FALSE
G074407	-2.30898	-4.2499	0.471235	PNAT3	FALSE
G074494	-2.11097	-2.58359	0.049824		FALSE
G074539	-1.40895	-0.16751	0.076914		TRUE
G074968	2.010818	-1.18126	0.410168	PRCAT161	TRUE
G075017	2.281212	-0.78629	0.180889	HNCAT223	TRUE
G075063	-2.50601	-4.47707	0.288189	AC004947.2	TRUE
G075089	-2.01154	0.153998	0.683709	HOTTIP	TRUE
G075092	-2.61906	-1.954	0.223106	RP1-170O19.17,EVX1-AS	FALSE
G075097	-4.60832	-0.3503	0.029346		TRUE
G075676	-2.42862	-4.53335	0.264144		TRUE
G076097	-1.89262	-1.50227	0.249581		FALSE
G076184	-1.85953	-0.64316	0.021942		TRUE
G076242	-0.73731	2.600749	0.077485	LOC441242	TRUE
G076645	-2.88207	-1.82329	0.107078	PRCAT177	TRUE
G076860	7.604254	3.444154	0.06407	CTB-167B5.2,STEAP4	TRUE
G077087	-2.22722	-3.44423	0.174488		FALSE
G077380	-2.65683	-3.4608	0.071783		FALSE
G077505	-0.2167	-2.74062	0.85187	SLC26A4-AS1	TRUE
G077510	1.765936	0.129628	0.04568	AC002467.7	TRUE
G077512	1.745997	0.68737	0.031341	AC002467.7	TRUE
G077737	-2.60346	-0.66291	0.031509	ST7-AS1	FALSE
G077739	-2.31591	-1.56796	0.163094	ST7-AS1	FALSE
G077742	-2.3044	-3.58705	0.135025		FALSE
G077982	-2.30111	-4.26111	0.306287	AC000124.1	FALSE
G078277	-2.44813	-1.8145	0.016501	AC015987.1,AC015987.2	TRUE
G078281	-2.71091	-0.30828	0.019327	AC015987.1	TRUE
G078444	-1.02459	0.858999	0.069018	WEE2-AS1	TRUE
G078504	2.664191	-2.2404	0.164717		TRUE
G078673	-0.76416	5.310495	0.074907	ZNF783,RP4-800G7.2	TRUE
G078770	-2.37024	-1.39958	0.052013		TRUE
G079054	4.063248	-3.99732	0.202558	HICLINC181	TRUE
G079125	-1.57549	0.454829	0.036431		FALSE
G079134	-2.07412	-1.20131	0.064847	FAM87A	FALSE
G079148	-2.01396	-0.42951	0.040909	ERICH1-AS1	TRUE
G079291	4.130739	-3.8813	0.103485	SMAT11	TRUE

G079502	1.390094	-0.01851	0.040702	LSCAT125	FALSE
G079556	2.371455	-3.10437	0.196563		TRUE
G079682	-1.96913	0.410999	0.030334	RP11-156K13.1	TRUE
G079833	-2.0032	-1.00896	0.073681	RP11-582J16.4	FALSE
G079868	0.642972	4.433418	0.078996	RP11-1149O23.4	FALSE
G079946	0.994022	4.798586	0.017807	RP11-395I14.2,GNRH1	TRUE
G080409	2.444153	-2.29417	0.046735	PRCAT118	TRUE
G080725	3.143953	-1.93841	0.032343		TRUE
G080834	-0.26278	-3.77161	0.907039	RP11-162D9.3	TRUE
G081325	-1.66755	1.666923	0.07067	LACAT125	TRUE
G081326	-2.24103	-2.20557	0.179672		TRUE
G081606	1.413406	1.052917	0.083179	RP11-26J3.1,HEY1	TRUE
G081615	3.120709	-1.20545	0.085584		TRUE
G081661	-1.1029	-2.38219	0.736892	PRCAT127	TRUE
G082010	-1.08956	-0.79801	0.099943	RP11-267M23.4	TRUE
G082011	-1.9448	-3.3781	0.172501		FALSE
G082227	0.984013	4.15626	0.044342	CAT1073	TRUE
G082546	-2.17156	-3.61628	0.127186		FALSE
G082839	2.039805	-2.97772	0.593665	LINC00964,LOC157381	TRUE
G082910	-1.17115	1.593753	0.094349	PCAT1,CAT1096	TRUE
G082921	-3.32178	-3.42433	0.016738		TRUE
G082926	-1.44712	0.122449	0.064395	CAT1099	TRUE
G083069	1.969502	-3.42333	0.505536	CAT1104	FALSE
G083070	3.486336	-1.44583	0.180817		TRUE
G083076	6.968643	-1.26333	0.036874	CAT1105	TRUE
G083255	-0.27999	-3.24986	0.904926	RP11-128L5.1	FALSE
G083270	-1.77936	-1.86731	0.06661	CAT1110	TRUE
G083427	-1.48611	-0.68523	0.035188	AC087793.1	FALSE
G083488	2.561124	-0.89171	0.11312		FALSE
G083537	-2.10185	-3.36337	0.3318	AF186192.5,AF186192.6	FALSE
G083581	1.070449	-2.04374	0.655781	RP11-143M1.3,FOXD4	FALSE
G083669	-1.19217	3.558728	0.017932	CVAT7	TRUE
G084112	2.404886	0.009819	0.225408		TRUE
G084143	-3.39449	-2.50829	0.121207		TRUE
G084146	-2.50645	-2.8591	0.290954		FALSE
G084326	-2.0569	-2.95487	0.078675		FALSE
G084916	3.548148	-4.44986	0.188249	CAT1147	FALSE
G084971	2.645704	-1.32517	0.342742	THCAT527	TRUE
G085024	-2.18213	-0.68974	0.382319	PRCAT102	TRUE
G085136	5.214157	-3.22853	0.054979	CAT1155	TRUE
G085294	-2.2938	-5.23099	0.358887		TRUE

G085431	2.840642	-3.20847	0.213208		TRUE
G085534	1.135815	-0.23603	0.088984	PRCAT135	TRUE
G085566	-2.18521	-3.20428	0.341135		TRUE
G086306	-1.00567	4.120839	0.175835	THCAT63	FALSE
G086466	2.662952	-1.30707	0.334593		TRUE
G086627	-1.19919	-2.29797	0.496586	RP11-175D17.3	TRUE
G087016	4.472483	-3.7218	0.025129	SNOU13	TRUE
G087078	0.276445	-3.15884	0.845186	HNCAT22	TRUE
G087152	-2.70803	-5.48883	0.278044	CAT1209	TRUE
G087245	-1.26703	0.033701	0.092819		TRUE
G087273	2.027331	-3.02711	0.239193		TRUE
G087778	2.395636	-2.00543	0.053774		FALSE
G088325	-2.13257	-2.73048	0.102258	ZNF630-AS1	FALSE
G088623	-0.94804	0.098651	0.088277		FALSE
G088721	-2.18126	-3.73679	0.189908		TRUE
G088897	-4.41195	-4.24177	0.151531	XIST	FALSE
G089246	-2.32242	-3.34185	0.398295		FALSE
G089267	-2.28821	-4.83273	0.414308	PNAT116	FALSE
G089469	-1.37864	-1.83114	0.080296		FALSE

References

1. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions.* Nat Rev Genet, 2009. **10**(3): p. 155-9.
2. Rinn, J.L. and H.Y. Chang, *Genome regulation by long noncoding RNAs.* Annu Rev Biochem, 2012. **81**: p. 145-66.
3. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Mol Cell, 2011. **43**(6): p. 904-14.
4. Faghihi, M.A., et al., *Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase.* Nat Med, 2008. **14**(7): p. 723-30.
5. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-6.
6. Lee, N., et al., *EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA.* Cell, 2015. **160**(4): p. 607-618.
7. Prensner, J.R., et al., *The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex.* Nat Genet, 2013. **45**(11): p. 1392-8.
8. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.
9. Wutz, A., T.P. Rasmussen, and R. Jaenisch, *Chromosomal silencing and localization are mediated by different domains of Xist RNA.* Nat Genet, 2002. **30**(2): p. 167-74.
10. Iyer, M.K., et al., *The landscape of long noncoding RNAs in the human transcriptome.* Nat Genet, 2015. **47**(3): p. 199-208.
11. Malik, R., et al., *The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer.* Mol Cancer Res, 2014. **12**(8): p. 1081-7.
12. Shukla, S., et al., *Identification and Validation of PCAT14 as Prognostic Biomarker in Prostate Cancer.* Neoplasia, 2016. **18**(8): p. 489-99.
13. *Treatment and survival of patients with cancer of the prostate. The Veterans Administration Co-operative Urological Research Group.* Surg Gynecol Obstet, 1967. **124**(5): p. 1011-7.
14. Huggins, C. and C.V. Hodges, *Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941.* J Urol, 2002. **168**(1): p. 9-12.
15. Lu-Yao, G.L., et al., *Fifteen-year survival outcomes following primary androgen-deprivation therapy for localized prostate cancer.* JAMA Intern Med, 2014. **174**(9): p. 1460-7.
16. Chen, Y., C.L. Sawyers, and H.I. Scher, *Targeting the androgen receptor pathway in prostate cancer.* Curr Opin Pharmacol, 2008. **8**(4): p. 440-8.
17. Wong, Y.N., et al., *Evolution of androgen receptor targeted therapy for advanced prostate cancer.* Nat Rev Clin Oncol, 2014. **11**(6): p. 365-76.

18. Mukherji, D., C.J. Pezaro, and J.S. De-Bono, *MDV3100 for the treatment of prostate cancer.* Expert Opin Investig Drugs, 2012. **21**(2): p. 227-33.
19. Scher, H.I., et al., *Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study.* Lancet, 2010. **375**(9724): p. 1437-46.
20. Scher, H.I., et al., *Increased survival with enzalutamide in prostate cancer after chemotherapy.* N Engl J Med, 2012. **367**(13): p. 1187-97.
21. Tran, C., et al., *Development of a second-generation antiandrogen for treatment of advanced prostate cancer.* Science, 2009. **324**(5928): p. 787-90.
22. de Bono, J.S., et al., *Abiraterone and increased survival in metastatic prostate cancer.* N Engl J Med, 2011. **364**(21): p. 1995-2005.
23. Reid, A.H., et al., *Significant and sustained antitumor activity in post-docetaxel, castration-resistant prostate cancer with the CYP17 inhibitor abiraterone acetate.* J Clin Oncol, 2010. **28**(9): p. 1489-95.
24. Stein, M.N., S. Goodin, and R.S. Dipaola, *Abiraterone in prostate cancer: a new angle to an old problem.* Clin Cancer Res, 2012. **18**(7): p. 1848-54.
25. Antonarakis, E.S., et al., *AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer.* N Engl J Med, 2014. **371**(11): p. 1028-38.
26. Antonarakis, E.S., M. Nakazawa, and J. Luo, *Resistance to androgen-pathway drugs in prostate cancer.* N Engl J Med, 2014. **371**(23): p. 2234.
27. Attard, G., J. Richards, and J.S. de Bono, *New strategies in metastatic prostate cancer: targeting the androgen receptor signaling pathway.* Clin Cancer Res, 2011. **17**(7): p. 1649-57.
28. Chan, S.C., Y. Li, and S.M. Dehm, *Androgen receptor splice variants activate androgen receptor target genes and support aberrant prostate cancer cell growth independent of canonical androgen receptor nuclear localization signal.* J Biol Chem, 2012. **287**(23): p. 19736-49.
29. Hearn, J.W.D., et al., *HSD3B1 and resistance to androgen-deprivation therapy in prostate cancer: a retrospective, multicohort study.* Lancet Oncol, 2016. **17**(10): p. 1435-1444.
30. Li, Y., et al., *Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines.* Cancer Res, 2013. **73**(2): p. 483-9.
31. Robinson, D., et al., *Integrative Clinical Genomics of Advanced Prostate Cancer.* Cell, 2015. **162**(2): p. 454.
32. Visakorpi, T., et al., *In vivo amplification of the androgen receptor gene and progression of human prostate cancer.* Nat Genet, 1995. **9**(4): p. 401-6.
33. Watson, P.A., V.K. Arora, and C.L. Sawyers, *Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer.* Nat Rev Cancer, 2015. **15**(12): p. 701-11.
34. Asangani, I.A., et al., *Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer.* Nature, 2014. **510**(7504): p. 278-82.
35. Pomerantz, M.M., et al., *The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis.* Nat Genet, 2015. **47**(11): p. 1346-51.

36. Cancer Genome Atlas Research, N., *The Molecular Taxonomy of Primary Prostate Cancer.* Cell, 2015. **163**(4): p. 1011-25.
37. Takayama, K., et al., *Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer.* EMBO J, 2013. **32**(12): p. 1665-80.
38. Prensner, J.R., et al., *Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression.* Nat Biotechnol, 2011. **29**(8): p. 742-9.
39. Cieslik, M., et al., *The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing.* Genome Res, 2015. **25**(9): p. 1372-81.
40. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.
41. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
42. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.
43. Law, C.W., et al., *voom: Precision weights unlock linear model analysis tools for RNA-seq read counts.* Genome Biol, 2014. **15**(2): p. R29.
44. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
45. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
46. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
47. Hansen, P., et al., *Saturation analysis of ChIP-seq data for reproducible identification of binding peaks.* Genome Res, 2015. **25**(9): p. 1391-400.
48. Kent, W.J., et al., *BigWig and BigBed: enabling browsing of large distributed datasets.* Bioinformatics, 2010. **26**(17): p. 2204-7.

Chapter 3

Functions and therapeutic potential of long non-coding RNA ARLNC1 in prostate cancer[1]

Abstract

Through a top-down bioinformatic analysis, we have identified AR-regulated long non-coding RNA-1 (ARLNC1). In this chapter, we assess the clinical relevance and therapeutic potential of ARLNC1, beyond our initial nomination of this lncRNA via its prostate lineage- and prostate cancer-specific expression in patient samples. We discovered an association of ARLNC1 levels with accentuated AR signaling and luminal epithelial differentiation in patient tumors, both of which are important clinical considerations for anti-androgen treatment. Using loss-of-function experiments *in vitro* and *in vivo*, we confirmed that ARLNC1 is critical for AR-positive prostate cancer cell growth. Interestingly, AR signaling is identified as one of the pathways regulated by ARLNC1, indicating the existence of a feedforward loop. These results suggest that ARLNC1 is a

[1] This chapter was previously published as part of the following manuscript: Zhang, Y., et al., Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. *Nat Genet*, 2018. 50(6): p. 814-824.

lncRNA that contribute to prostate cancer development, at least in part through preservation of AR signaling.

Introduction

The androgen receptor signaling is the critical driver of all stages of prostate cancer. In the most lethal, metastatic castration resistant prostate cancer (mCRPC), persistent AR activity in spite of ADT and the second-generation AR-targeting agents is achieved by direct alteration to the AR signaling axis. Aberrations of AR signaling axis exist in 71.3% of mCRPC cases, with the majority being direct alterations affecting AR (amplification and mutation). As such, a better understanding of players in the AR signaling pathway may provide novel therapeutic opportunities to disrupt persistent AR activity.

In the previous chapter, we discovered lncRNAs responsive to DHT stimulation from cells, and prioritized ARLNC1 with suggestive oncogenic roles. The objective of this chapter is to carry out *in vitro* and *in vivo* studies to investigate functions and evaluate therapeutic potential of ARLNC1 in prostate cancer.

Results

Prevalence of ARLNC1 in normal and cancer tissues

Expression of ARLNC1 was interrogated across cancer and normal tissue RNA-Seq samples from TCGA and the Genotype-Tissue Expression (GTEx) project[1, 2], respectively. In the TCGA cohort, ARLNC1 exhibited a highly prostate cancer specific

expression pattern, with little to no expression in other tumor types (**Figure 3.1A**). Similarly, in the GTEx normal tissue cohort, its expression was limited to the prostate (**Figure 3.1B**).

Among prostate samples, ARLNC1 expression was significantly higher in localized and metastatic prostate cancers compared to benign tissues, as assessed by RNA-Seq (**Figure 3.1C**) and *in situ* hybridization (**Figure 3.1D**). In an extensive differential expression analysis using MiTranscriptome, ARLNC1 was found to be among the top 1% of transcripts most upregulated in prostate cancer and specific to the prostate lineage, with no significant associations in other tissues (**Figure 3.2A**). Moreover, the protein-coding genes that were most correlated with ARLNC1 were found to be associated with prostate cancer progression in ONCOMINE concept analyses performed on multiple clinical data sets[3] (**Figure 3.2B**).

Furthermore, we compared the top- and bottom-quartiles of prostate cancer samples based on ARLNC1 expression, and as expected, found that the top-samples were characterized by elevated expression of "Androgen Receptor Signaling Targets" genes and an expression profile consistent with luminal epithelial prostate cancer, which was also reflected in the down-regulation of genes associated with the epithelial-to-mesenchymal transition (**Figure 3.2E-G**). We verified that these differences were not due to tumor purity (**Figure 3.2D**). Overall, these results strongly suggest the association of ARLNC1 levels with AR signaling and luminal epithelial differentiation in patient tumors, which are important clinical

considerations for administering anti-androgen therapy. Additionally, we evaluated association between ARLNC1 level and Gleason score, the grading system used to determine prostate cancer aggressiveness. No significant association between ARLNC1 expression and disease grade was observed (**Figure 3.2C**).

Taken together, these results confirm that ARLNC1 expression is restricted to the prostate lineage, elevated in prostate cancer and associated with AR signaling throughout prostate cancer progression.

Characterization of ARLNC1

To functionally characterize ARLNC1, we first identified appropriate prostate cancer cell lines with moderate to high levels of ARLNC1 expression using in house RNA-Seq data (**Figure 3.3A**). Supporting the association of AR with ARLNC1, ARLNC1 expression was highly enriched in AR-positive cell lines, with the highest expression in MDA-PCa-2b and LNCaP cells. In addition, quantitative PCR (qPCR) analysis for the ARLNC1 transcript also demonstrated that this gene was expressed at the highest level in the MDA-PCa-2b and LNCaP cell lines (**Figure 3.3B**).

As existing annotations of ARLNC1 (located on chromosome 16) predict the presence of several transcript isoforms that differ in exon and TSS usage, we determined the exact structure in MDA-PCa-2b and LNCaP cells, by random amplification of cDNA ends (RACE) (**Figure 3.3D**). A common TSS for ARLNC1 was found in both cell lines, and the

~2.8 kilobase (kb) transcript isoform was further confirmed by northern blot analysis (**Figure 3.3C**). Single molecule fluorescent ISH (smFISH) revealed that approximately 100 molecules of ARLNC1 transcripts existed per MDA-PCa-2b cell (**Figure 3.4A-B**). Using smFISH and qPCR, we also found that ARLNC1 molecules were distributed equally between the nuclear and cytoplasmic cellular compartments (**Figure 3.4C-D**).

Phenotypic effect of ARLNC1 loss in vitro

Knockdown of ARLNC1 had a significant effect on the proliferation of AR-dependent MDA-PCa-2b, LNCaP, and LNCaP-AR cells, but had no effect on AR-negative DU145 and PC3 cells (**Figure 3.5A-D**). Knockdown of ARLNC1 also resulted in increased apoptosis in AR-positive prostate cancer cells (**Figure 3.5E-F**). Although we were not able to generate single cell-derived clones harboring ARLNC1 knockout (by CRISPR-Cas9) in AR-positive MDA-PCa-2b or LNCaP cells, we observed a growth defect of the cell population with ARLNC1 deficiency (**Figure 3.6**). Furthermore, we designed antisense oligonucleotides (ASOs) targeting the ARLNC1 transcript (**Figure 3.7A**). Transfection of ASOs exhibited strong knockdown efficiency (**Figure 3.7B**). AR-positive cells transfected with ARLNC1 ASOs exhibited retarded growth, similar to those treated with siRNAs (**Figure 3.7C**).

To elucidate ARLNC1 regulated genes in these processes, we performed gene expression profiling of wild-type and ARLNC1-knockdown MDA-PCa-2b cells (**Figure 3.8A**). We found that ASO-mediated knockdown resulted in similar effects on gene expression

profiling to siRNA (**Figure 3.8B-C**). Gene ontology (GO) pathway enrichment analysis of the differentially expressed genes revealed deregulation of four main biological activities: apoptosis, cell proliferation, DNA damage response and androgen signaling (**Figure 3.8A**). A significant decrease in AR target gene expression is particularly interesting (**Figure 3.8D-E**) given the fact that ARLNC1 expression is stimulated by AR, suggesting a positive feedback loop between ARLNC1 and AR signaling.

Phenotypic effect of ARLNC1 loss in vivo

Since modulation of ARLNC1 expression levels resulted in a striking proliferation phenotype, we hypothesized that ARLNC1 inhibition could be used therapeutically for the treatment of prostate cancer. Cells expressing shRNA targeting ARLNC1 formed smaller tumors in mice when compared to cells expressing non-targeting shRNA (**Figure 3.9A**), thus suggesting that ARLNC1 is an important survival factor for AR-dependent prostate cancer *in vivo*.

ASOs have been shown to be effective in targeting RNAs *in vivo*[4-6]. To evaluate the therapeutic potential of ARLNC1 ASOs, we first assessed the cellular free-uptake efficiency of ARLNC1 ASOs, a prerequisite for ASO therapeutic use. Importantly, several ASOs significantly reduced ARLNC1 levels through free uptake (**Figure 3.7D**). Free uptake of ARLNC1 ASOs led to a significant decrease in the proliferation capacity of MDA-PCa-2b cells in both normal cell culture and three-dimensional sphere conditions (**Figure 3.7E, 3.9B-C**). Treatment of mice bearing MDA-PCa-2b xenografts with

ARLNC1-targeting ASOs led to significant decreases in tumor growth compared to control ASO (**Figure 3.9D-I**). Taken together, these data, along with the association of ARLNC1 with aggressive androgen signaling, suggest that ARLNC1 plays a critical role in the proliferation of AR-dependent prostate cancer and can be effectively exploited as a therapeutic target.

Discussion

In this Chapter, we interrogated basic cell functions affected by ARLNC1, using a series of loss-of-functions assays (RNA interference, CRISPR-mediated knockout, and ASO-mediated knockdown). As a novel non-coding regulator of AR signaling, ARLNC1 has the potential to be not only a mechanistic biomarker, but also a therapeutic target for advanced prostate cancer. Specific antisense nucleotides targeting ARLNC1, which we demonstrate to be specifically expressed in the prostate, could circumvent undesirable side effects that occur in other tissues with exposure to androgen synthesis inhibitors or anti-androgens. The application of ASOs has ushered in an exciting era that makes it possible to target previously 'undruggable' molecules directly at the transcript level, such as ARLNC1, which is likely to yield promising opportunities in cancer treatment.

Materials and Methods

Oncomine concept analysis of the ARLNC1 signature

Genes with expression levels significantly correlated with ARLNC1 were separated into positively and negatively correlated gene lists. These two lists were then imported into

Oncomine as custom concepts and queried for association with other prostate cancer concepts housed in Oncomine. All of the prostate cancer concepts with odds ratio > 2.0 and P value < 1 × 10 were selected. Top concepts (based on odds ratios) were selected for representation. We exported these results as the nodes and edges of a concept association network and visualized the network using Cytoscape version 3.3.0. Node positions were computed using the edge-weighted force directed layout in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of Mode labels[7].

RNA ISH on tissue microarray

ISH assays were performed on tissue microarray sections from Advanced Cell Diagnostics, Inc. as described previously[8]. In total, 133 tissue samples were included (11 from benign prostate, 85 from localized prostate cancer and 37 from metastatic prostate cancer). ARLNC1 ISH signals were examined in morphologically intact cells and scored manually by a study pathologist, using a previously described expression value scoring system[9]. For each tissue sample, the ARLNC1 product score was averaged across evaluable TMA tissue cores. Mean ARLNC1 product scores were plotted in **Figure 3.1C**.

RACE

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

Northern blot analysis

NorthernMax-Gly Kit (Ambion) was used for ARLNC1 detection following the manufacturer's protocol. Briefly, 20 μg of total RNA was resolved on a 1% agarose glyoxal gel and then transferred to nylon membrane (Roche), crosslinked to the membrane (UV Stratalinker 1800; Stratagene) and the membrane was prehybridized. Overnight hybridization was performed with a ARLNC1-specific P - labeled RNA probe. Membranes were exposed to HyBlot CL autoradiography film (Denville Scientific). The primer sequences used for generating the probes are as follows:

ARLNC1-NB-F1: TCCTGAGCCGAAAATAAGGA

ARLNC1-NB-T7-R1:

GATCACTAATACGACTCACTATAGggagaggagaCCCTCATTTCCTTCCAGCTT

ARLNC1-NB-F2: CACTCTATCCAGCCCCATGA

ARLNC1-NB-T7-R2:

GATCACTAATACGACTCACTATAGggagaggagaATGGGCAAGTCCAAAATCAA

RNA isolation and cDNA synthesis

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

qRT–PCR analysis

Relative RNA levels determined by qRT–PCR were measured on an Applied Biosystems 7900HT Real-Time PCR System, using Power SYBR Green MasterMix (Applied Biosystems). All of the primers were obtained from Integrated DNA Technologies, and gene-specific sequences are listed in **Table 3.1**. GAPDH, HMBS and ACTB were used as internal controls for quantification of gene targets. The relative expression of RNAs was calculated using the ΔΔCt method.

Cytoplasmic and nuclear RNA purification

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

siRNA-mediated knockdown

Transfections with siRNA (50 nM) were performed with Lipofectamine RNAiMAX according to the manufacturer's instructions. RNA and protein were harvested for analysis 72 h after transfection.

ASO-mediated knockdown

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

transfectionufacturer's instructions. RNA and protein were harvested for analysis 72 h after transfection.

Gene expression profiling

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

Analysis of Agilent 44k microarrays was carried out using limma[10] and included background subtraction (bc.method = 'half', offset = 100) and within-array normalization (method = 'loess'). Between-array quantile normalization of average expression levels (but not log-fold changes) was performed using the function normalizeBetweenArrays (method = 'Aquantile'). Control probes and probes with missing values were excluded from further analyses. Probes were annotated to Gencode v22 genes using the mapping downloaded from Ensembl (efg_agilent_wholegenome_4 × 44k_v2). Probes originally annotated as

AK093002 were used to detect ARLNC1. Differentially expressed genes following ARLNC1 knockdown in MDA-PCA-2b cells were identified from triplicate biological repeats using adjusted P value < 0.1 and absolute log fold-change > 0.6 cutoffs. Consensus targets of ARLNC1 knockdown using siRNA and ASOs were identified using a merged linear model (all 10 samples treated replicates) and a P value < 0.001 cutoff.

Cell proliferation assay

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

Apoptosis analysis

Cells were grown in 6-well plates and transfected with nonspecific siRNA or siRNAs targeting ARLNC1. Apoptosis analysis was performed 48 h after transfection, using the Dead Cell Apoptosis Kit (Molecular Probes no. V13241) according to the manufacturer's instructions.

Immunoblot analysis

Cells were lysed in RIPA lysis and extraction buffer (Thermo Scientific no. 89900) supplemented with protease inhibitor cocktail (ROCHE no. 11836170001). Protein concentrations were quantified using the DC protein assay (BIO-RAD), and protein lysates were boiled in sample buffer. Protein extracts were then loaded and separated on SDS–

PAGE gels. Blotting analysis was performed with standard protocols using polyvinylidene difluoride membrane (GE Healthcare). Membranes were blocked for 60 min in blocking buffer (5% milk in a solution of 0.1% Tween-20 in Tris-buffered saline (TBS-T)) and then incubated overnight at 4 °C with primary antibody. After three washes with TBS-T, membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibody. Signals were visualized with an enhanced chemiluminescence system as described by the manufacturer (Thermo Scientific Pierce ECL Western Blotting Substrate). Primary antibodies used were: androgen receptor (1:1000 dilution, Millipore, no. 06-680, rabbit), GAPDH (1:5000 dilution, Cell Signaling, no. 3683, rabbit), PSA (KLK3)(1:5000 dilution, Dako, no. A0562, rabbit) and cleaved PARP (1:1,000 dilution, Cell Signaling, no. 9542, rabbit).

In vivo experiments

All experiments were approved by the University of Michigan Institutional Animal Care and Use Committee. For tumor generation with shRNA-mediated knockdown, shRNA targeting ARLNC1 was cloned in pSIH1-H1-copGFP-T2A-Puro (System Biosciences). Lentiviral particles were generated at the University of Michigan Vector Core. LNCaP-AR cells were infected with lentivirus expressing ARLNC1 shRNA for 48 h. Knockdown of ARLNC1 was confirmed by qPCR analysis. Male athymic nude mice were randomized into two groups at six to eight weeks of age. Five million cells expressing sh-ARLNC1 or sh-vector were injected into bilateral flanks of mice. Caliper measurements were taken in two dimensions twice a week by an investigator blinded to the study objective and used to

calculate tumor volume. The study was terminated when the tumor volume reached 1,000 mm. For ASO treatment *in vivo*, six- to eight-week-old male athymic nude mice were inoculated subcutaneously with MDA-PCa-2b cells suspended in a Matrigel scaffold in the posterior dorsal flank region (5 million cells per site, two sites per animal). When the mean tumor volume reached approximately 150 mm^3, mice were randomized into two groups, and respectively treated with ARLNC1-specific or control ASO. ASOs, dosed at 50 mg/kg, were subcutaneously injected between the scapulae once daily for three periods of five days on/two days off. Tumor size was measured twice per week using a digital caliper by a researcher blinded to the study design. Mouse body weights were monitored throughout the dosing period. When the average tumor size in the control group reached 1,500 mm^3, mice were euthanized and the primary tumors were excised for weight determination. One third of the resected specimen was placed in 10% formalin buffer, and the remaining tissue was snap-frozen. For *in vivo* experiments, power analysis (GPOWER software) performed for each tumor type tested to date indicates that the sample size we chose yields a statistical power >90% for detection of tumor size reduction of 40%. Sample sizes were not pre-determined for all other assays. For *in vivo* experiments, animals were randomized. Randomization was not performed for all other assays.

Figures

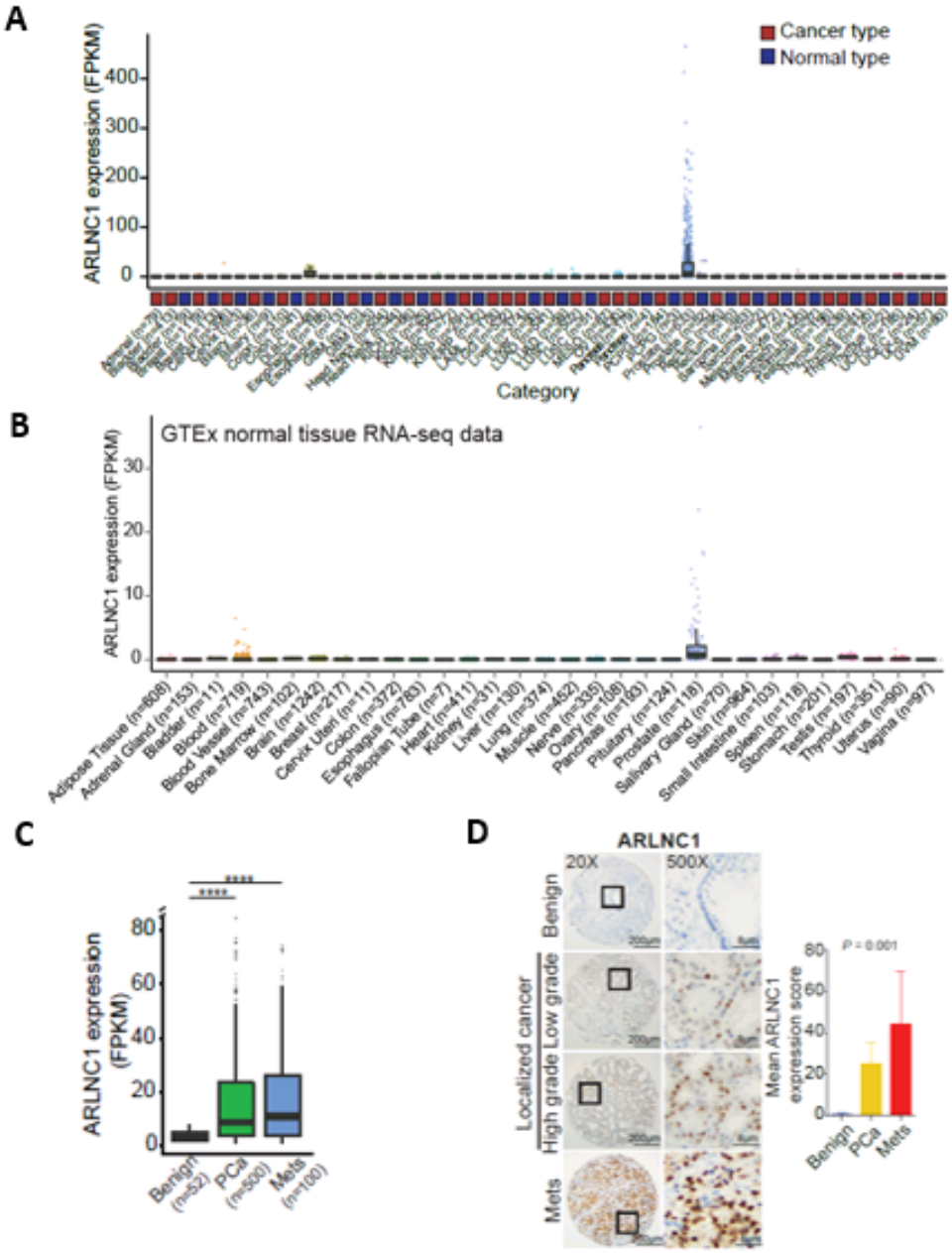


Figure 3.1 Expression of ARLNC1 across lineages and within prostate cancer
(**A**) Relative expression (FPKM) of ARLNC1 across different cancer types in the TCGA cohort. (**B**) Relative expression (FPKM) of ARLNC1 across a panel of normal tissues in the GTEx normal tissue RNA-seq cohort (n = 9,435 samples) (http://www.gtexportal.org/). Box-plot definitions: center, median; box limits, 1st and 3rd quartiles; whiskers follow the 1.5 rule.
(**C**) relative expression (FPKM) of ARLNC1 across benign prostate (n = 52 samples), localized prostate cancer (n = 500 samples) and metastatic prostate cancer (n = 100 samples). PCa versus

Normal: ****$P < 2.2 \times 10^{-16}$; Mets versus Normal: ****$P = 2.6 \times 10^{-7}$ (two-sided t test). Box-plot definitions: the center line depicts the median, the box shows the first and third quartiles, and the whiskers follow the 1.5 rule.

(**D**) ISH of ARLNC1 transcript in a human prostate cancer tissue microarray. Representative ARLNC1 staining is shown for benign prostate and localized and metastatic prostate cancer tissue. The bar plot represents mean ARLNC1 expression scores across benign prostate (n = 11), localized prostate cancer (n = 85) and metastatic prostate cancer (n = 37) tissues, with vertical bars indicating the bootstrapped 95% confidence interval of the means. Significance was calculated by a Kruskal–Wallis rank-sum test.

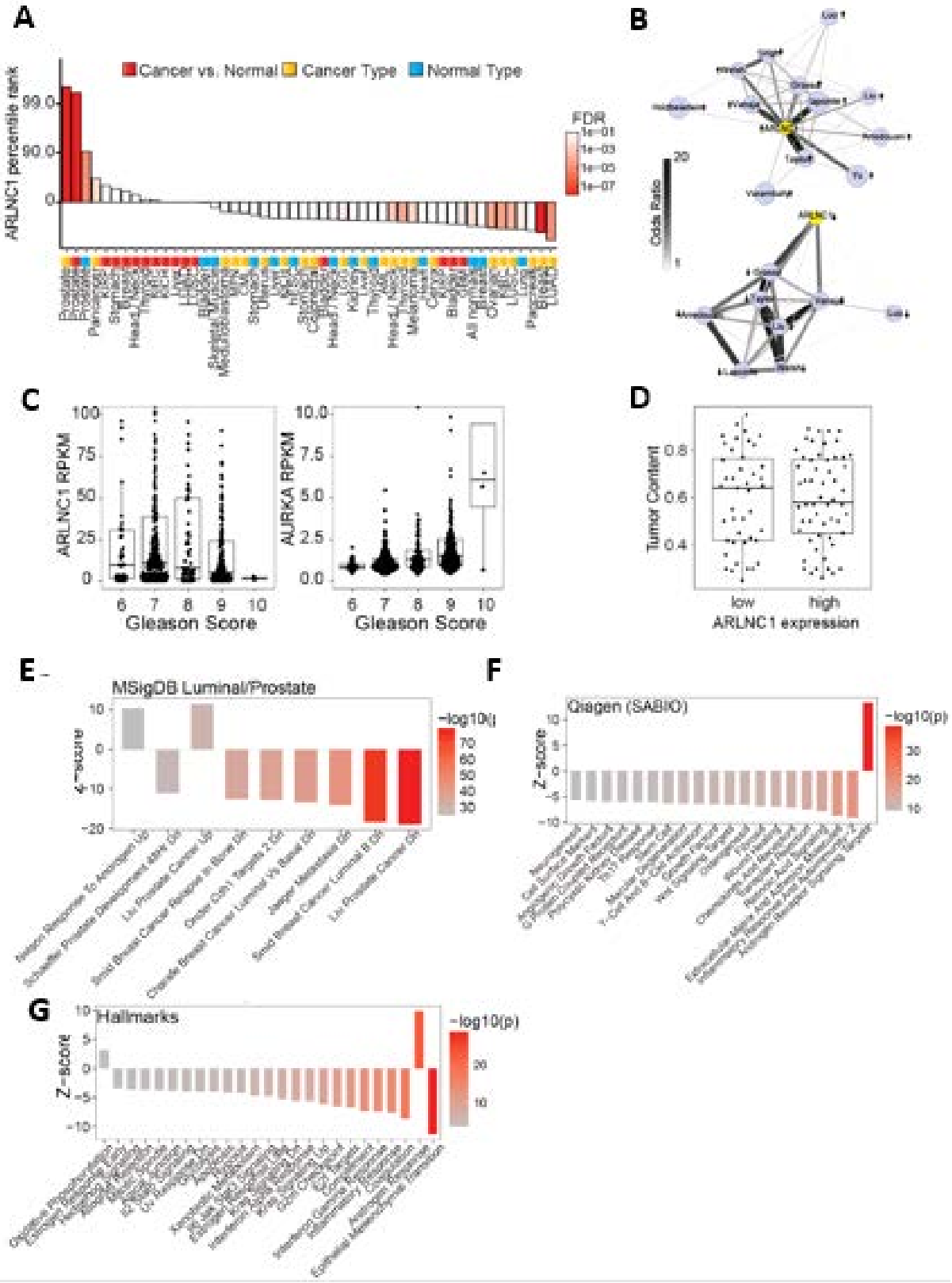

A
Cancer vs. Normal
Cancer Type
Normal Type
ARLNC1 percentile rank
99.0
90.0
0
FDR
1e−01
1e−03
1e−05
1e−07
B
Odds Ratio
20
1
C
ARLNC1 RPKM
100
75
50
25
0
6 7 8 9 10
Gleason Score
AURKA RPKM
10.0
7.5
5.0
2.5
0.0
6 7 8 9 10
Gleason Score
D
Tumor Content
0.8
0.6
0.4
low
high
ARLNC1 expression
E
MSigDB Luminal/Prostate
Z−score
10
0
−10
−20
−log10(p
70
60
50
40
30
F
Qiagen (SABIO)
Z−score
10
5
0
−5
−10
−log10(p)
30
20
10
G
Hallmarks
Z−score
10
5
0
−5
−10
−log10(p)
20
10

Figure 3.2 ARLNC1 related signatures in prostate cancer

(**A**) Tissue- and cancer-specific expression of ARLNC1 according to MiTranscriptome. The SSEA test (total n = 7,256 samples) was used to determine whether each gene was significantly associated with a set of samples (e.g., prostate cancer) or cancer progression in a given lineage (e.g., prostate normal to prostate cancer). The genes were ranked according to their strength of association. Shown is the percentile rank for ARLNC1 (enrichment-positive, depletion-negative) across the tested sample sets/comparisons.

(**B**) Oncomine concepts analysis of genes positively or negatively (bottom) correlated with ARLNC1. The network was visualized using the force-directed layout algorithm in the Cytoscape tool, with node sizes representing the number of genes in each molecular concept and node names indicating the author of the primary study.

(**C**) ARLNC1 expression levels are not associated with Gleason score. AURKA expression levels are significantly associated with Gleason score. Based on n = 500 samples from TCGA. Significance was determined by two-sided t test. Box-plot definitions: center, median; box limits, 1st and 3rd quartiles; whiskers follow the 1.5 rule.

(**D**) Tumor content estimated from whole exome sequencing is compared between high (top quartile) and low (bottom quartile) ARLNC1 expression in mCRPC samples (n = 100). Box-plot definitions: center, median; box limits, 1st and 3rd quartiles; whiskers follow the 1.5 rule.

(**E**) Signatures associated with prostate cancer and luminal differentiation were selected from MSigDB and contrasted between the ARLNC1-high (top quartile) and ARLNC1-low (bottom quartile) mCRPC samples (n = 100). The z score represents the effect-size and direction for the relative signature enrichment. For each signature 'Up' signifies that a signature is expected to be upregulated in the tested condition; conversely, 'Dn' signifies that a signature is expected to be downregulated.

(**F**) Curated pathway signature analysis between ARLNC1-high (top-quartile) and ARLNC1-low (bottom-quartile) mCRPC samples (n = 100). The z score represents the effect size and direction for the relative signature enrichment.

(**G**) Cancer hallmark signature analysis between ARLNC1-high-expression (top quartile) and ARLNC1-low-expression (bottom quartile) mCRPC samples (n = 100 samples). The z score represents the effect size and direction for the relative signature enrichment, determined by two-sided random-set test.

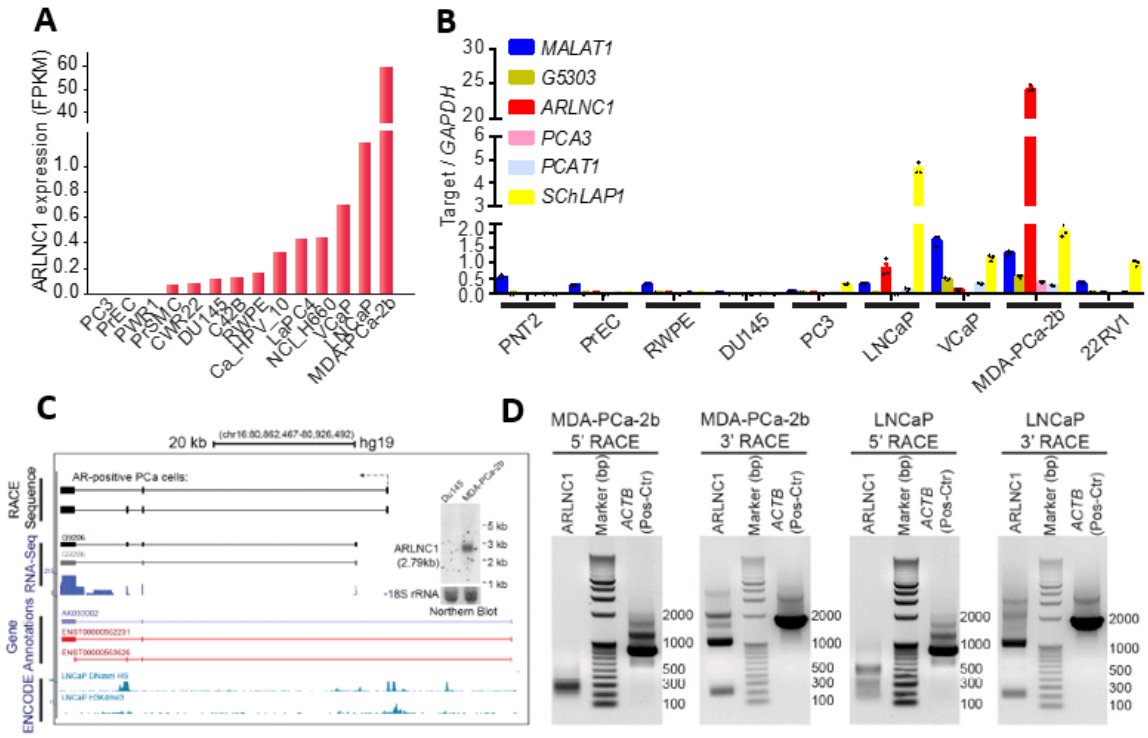


Figure 3.3 Characterization of ARLNC1 transcript

(**A**) Relative expression of ARLNC1 (FPKM) across 14 prostate cancer cell lines.

(**B**) qPCR analysis of ARLNC1 expression in nine prostate cancer cell lines. Expression levels of several known prostate cancer-associated lncRNAs are also shown. Mean ± s.e.m. values are graphed, n = 3.

(**C**) Left, representative image of ARLNC1 gene structure in AR-positive prostate cancer cells. Annotations of ARLNC1 in Ensembl and the Encyclopedia of DNA Elements (ENCODE) are also shown. Inset, expression of ARLNC1 transcripts in MDA-PCa-2b cells, validated by northern blot. ARLNC1-negative DU145 cells serve as a negative control.

(**D**) 5' RACE and 3' RACE results in MDA-PCa-2b cells and LNCaP cells. Experiments were repeated three times with similar results.

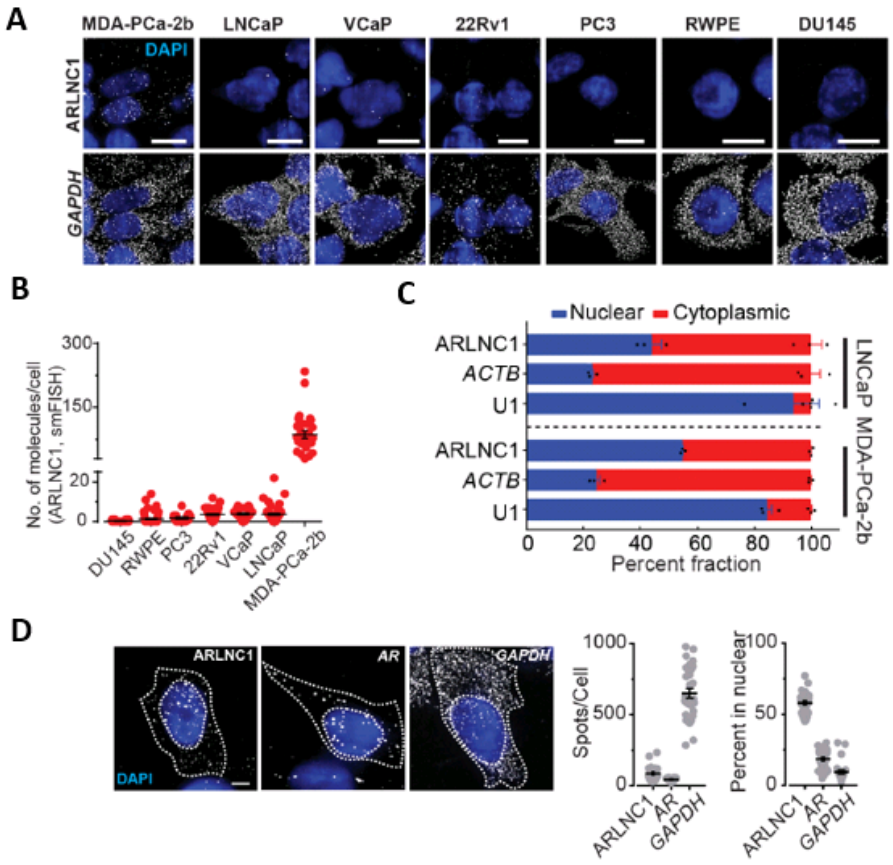


Figure 3.4 Cellular localization of ARLNC1

(**A**) smFISH images depicting localization of ARLNC1 transcripts in a panel of prostate cancer cell lines. Representative pseudocolored images are shown of MDAPCa-2b, LNCaP, VCaP, 22Rv1, PC3, RWPE, and DU145 cells probed for ARLNC1 (gray) or GAPDH (gray, control). The nucleus is stained with DAPI (blue). Scale bar, 5 μm; n = 3 independent experiments for each cell line.

(**B**) Scatterplot representing the average number of ARLNC1 transcripts per cell in a panel of prostate cancer cell lines, including MDA-PCa-2b, LNCaP, VCaP, 22Rv1, PC3, RWPE, and DU145. The black line and whiskers depict the mean and s.e.m., respectively (n = 50 cells for each cell line aggregated from three independent experiments).

(**C**) Percentage of nuclear/cytoplasmic RNA levels of ARLNC1, ACTB, and U1, measured by qRT–PCR after subcellular fractionation of MDA-PCa-2b and LNCaP cells. U1 serves as a positive control for nuclear gene expression, while ACTB RNA serves as a positive control for cytoplasmic gene expression. The graphs show the mean ± s.e.m., n = 3 independent experiments for each cell line.

(**D**) Representative grayscale images of MDA-PCa-2b cells stained for DAPI (nucleus, blue) and ARLNC1, AR, or GAPDH transcripts (smFISH, gray). Scale bar, 10 μm. Quantification of the number of molecules per cell and the nucleo-cytoplasmic distribution of each transcript are also represented (n = 60 cells for each sample aggregated from three independent experiments). The center line represents the mean, and error bars represent s.e.m.

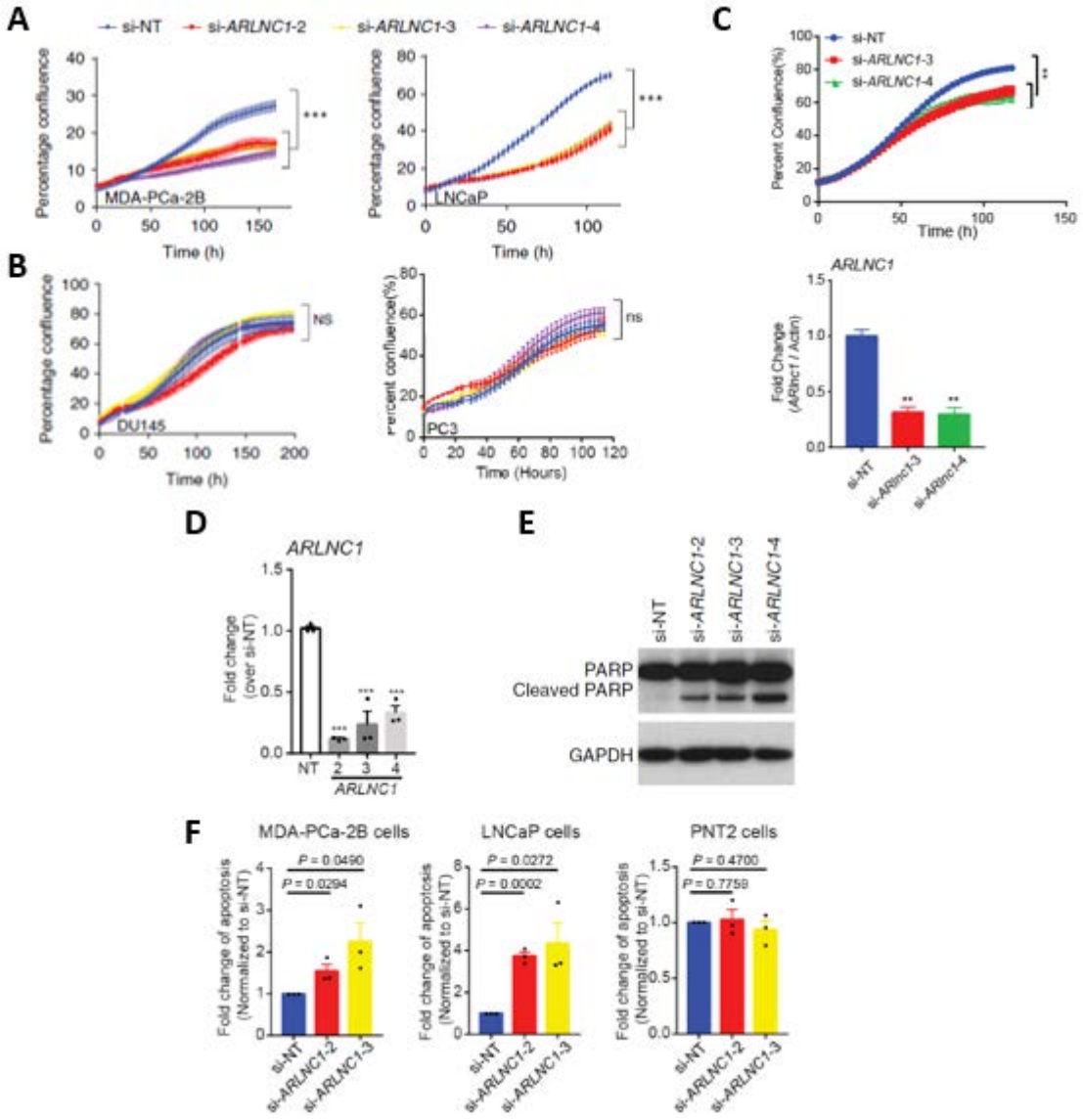


Figure 3.5 Changes on basic cell functions following siRNA-mediated ARLNC1 loss

(**A**) siRNA-mediated knockdown of ARLNC1 *in vitro* in AR-positive prostate cancer cell lines (MDA-PCa-2b and LNCaP) inhibits cell proliferation. Mean ± s.d. values are shown, n = 6 independent cell cultures per group, **P (adjusted) = 0.0001 compared to si-NT-treated cells, by one-way ANOVA with Dunnett's multiple-comparisons test.

(**B**) Loss of ARLNC1 does not affect cell growth in AR-negative prostate cell lines DU145 and PC3. P value not significant (NS) compared to si-NT-treated cells, by one-way ANOVA.

(**C**) Top: siRNA knockdown of ARLNC1 in LNCaP-AR cells inhibits cell proliferation. Mean ± s.e.m. are shown, n = 4, **Adjusted P < 0.001 compared to siNT treated cells, by one-way ANOVA analysis. Bottom: Knockdown efficacy of siRNAs targeting ARLNC1 in LNCaP-AR cells. Mean ± s.e.m. are shown, n = 3. **P < 0.001 compared to siNT group by two-tailed Student's t-test.

(**D**) Knockdown efficacy of three independent siRNAs targeting ARLNC1 in MDA-PCa-2b cells. Mean ± s.e.m. values are shown, n = 3. ***P = 0.0001 determined by ANOVA with Dunnett correction.
(**E**) ARLNC1 loss leads to increased apoptosis as shown by western blot analysis of PARP and cleaved PARP in LNCaP cells following ARLNC1 knockdown. The experiment was repeated independently three times with similar results.
(**F**) Increased apoptosis observed in MDA-PCa-2b and LNCaP cells 48 h after transfection with ARLNC1 siRNAs. ARLNC1-negative PNT2 cells serve as a negative control. Fold change of apoptosis was calculated relative to si-NT-treated samples. n = 3 independent cell cultures. Mean ± s.e.m. values are shown. P values were determined by two-tailed Student's t test.

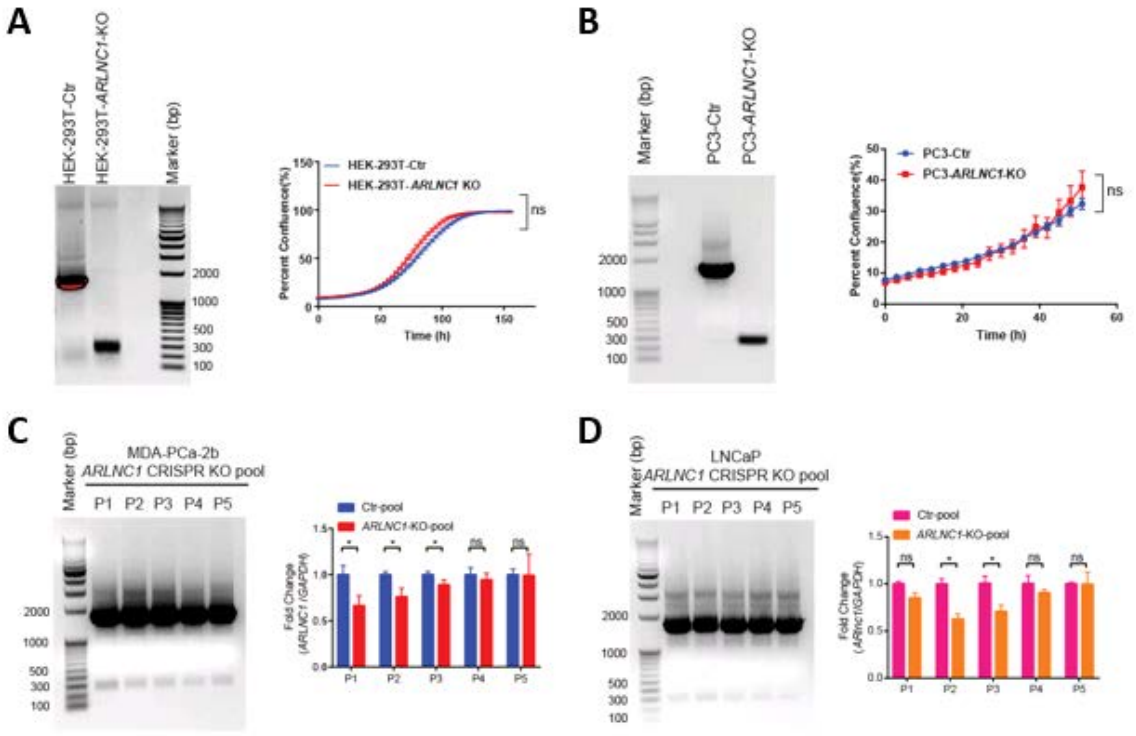


Figure 3.6 Effect of ARLNC1 loss generated by CRISPR-Cas9

(**A**, **B**) In HEK-293T cells (**A**) and PC3 cells (**B**), genomic PCR indicates loss of *ARLNC1* genomic region. Proliferation curve indicates same proliferation rate between WT clone and ARLNC1 knockout clone.

(**C**, **D**) Pool of MDA-PCa-2b cells (**C**) or LNCaP cells (**D**) were treated with guide RNA pairs targeting ARLNC1, and monitored for 5 passages. Genomic PCR analysis of *ARLNC1* genomic region and qPCR analysis of ARLNC1 transcript indicate a diminishing population with ARLNC1 loss in the total cell population.

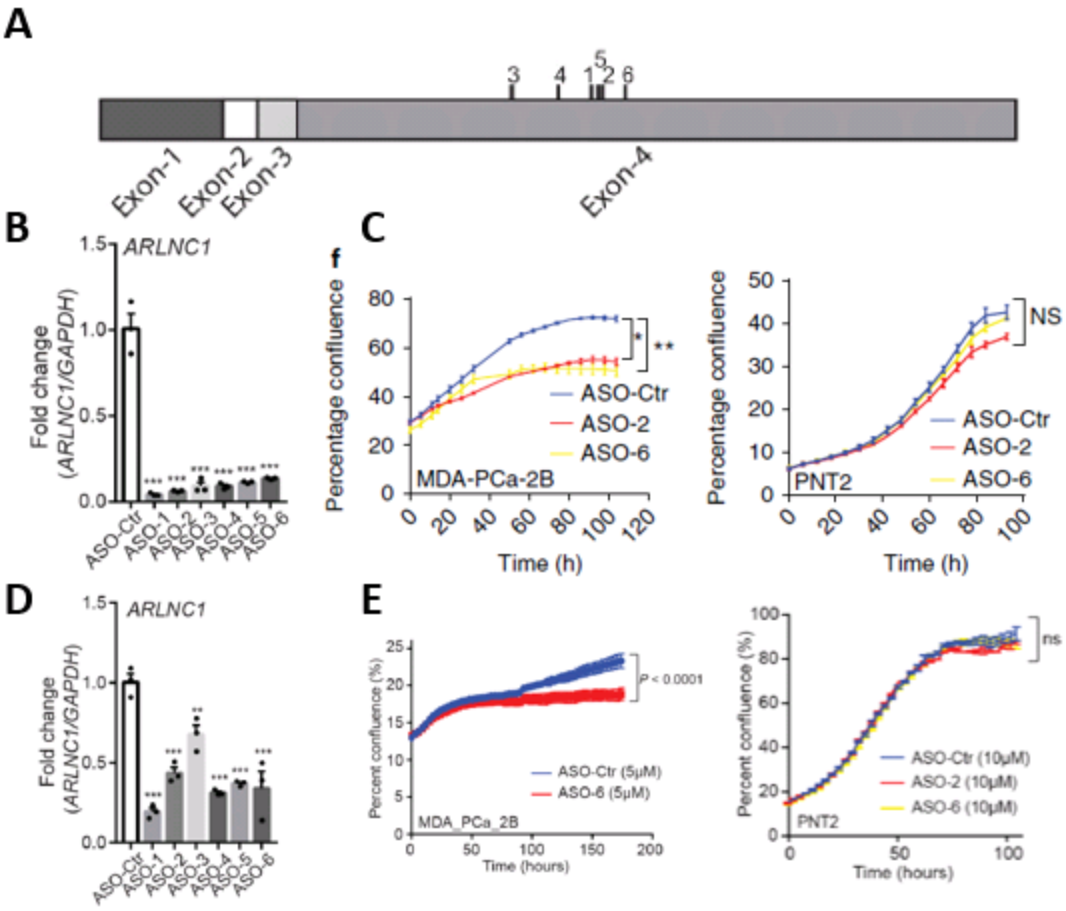


Figure 3.7 Changes on basic cell functions following ASO-mediated ARLNC1 knockdown

(**A**) The positions of ARLNC1 ASO-targeting sites (1-6) are indicated on the schematic representation of the ARLNC1 transcript.

(**B**) MDA-PCa-2b cells were transfected with six independent ASOs targeting ARLNC1. Knockdown efficacy was evaluated by qPCR analysis. Mean ± s.e.m. values are shown, n = 3. ***P = 0.0001 determined by ANOVA.

(**C**) Transfection of ASOs targeting ARLNC1 in AR-positive MDA-PCa-2b cells inhibits cell proliferation. The AR-negative prostate cell line PNT2 serves as a negative control. Mean ± s.e.m. values are shown, n = 6 independent cell cultures per treatment group. *P (adjusted) = 0.0112, **P (adjusted) = 0.0065, NS: not significant; compared to the control-ASO group by one-way ANOVA with Dunnett correction for multiple comparisons.

(**D**) The free-uptake efficacy of ARLNC1 ASOs was examined in MDA-PCa-2B cells 72h after ASO addition to the culture medium (10 μM). ARLNC1 expression was evaluated by qPCR analysis. Mean ± s.e.m. values are shown, n = 3. ***P = 0.0001, **P = 0.0032; determined by ANOVA with Dunnett correction.

(**E**) Free-uptake treatment with ASOs targeting ARLNC1 resulted in retarded growth of MDA-PCa-2b cells *in vitro*. ARLNC1-negative prostate cell line PNT2 served as a negative control. Mean ± s.e.m. values are shown, n = 6. P values were determined by two-tailed Student's t test (ns, not significant).

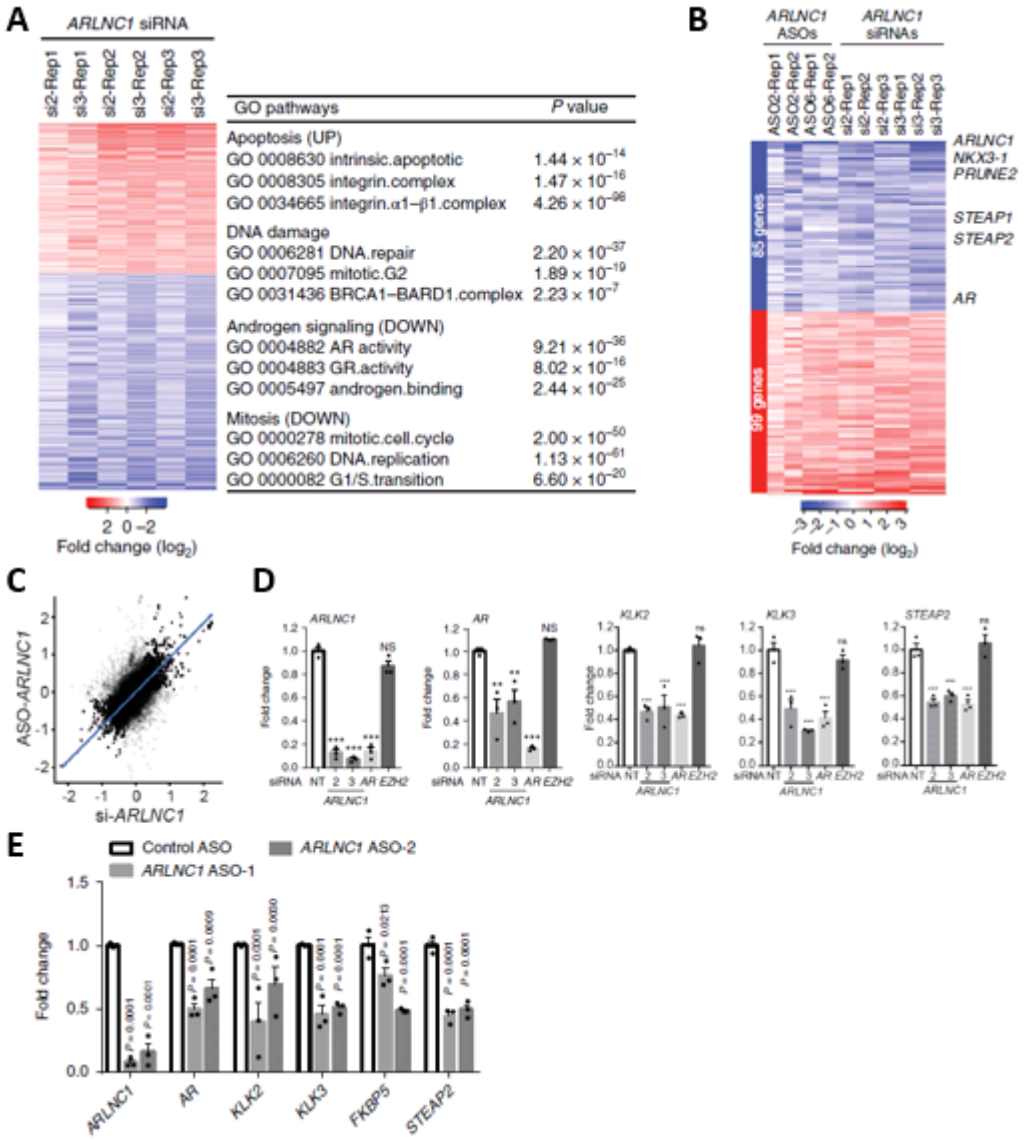


Figure 3.8 Transcriptome changes induced by ARLNC1 silencing

(**A**) Gene expression profiling for ARLNC1 knockdown in MDA-PCa-2b cells (n = 3 biologically independent cell cultures for each siRNA). The chart presents the top enriched pathways following ARLNC1 knockdown, identified using GO enrichment analysis (RandomSet test).

(**B**) Gene expression profiling for siRNA-mediated or ASO-mediated ARLNC1 knockdown in MDA-PCa-2b cells. The numbers above the heat map represent the specific microarray replicates.

(**C**) Correlation analysis of siRNA-mediated knockdown and ASO-mediated knockdown of ARLNC1 among replicated microarray experiments in MDA-PCa-2b cells (n = 2 biological replicates per ASO treatment group and n = 3 biological replicates per siRNA treatment group).

(**D**) qRT–PCR analysis of ARLNC1, AR, KLK2, KLK3, and STEAP2, in MDA-PCa-2b cells transfected with siRNAs against ARLNC1, AR, EZH2, or non-specific control. siRNA against AR serves as a positive control for inhibited AR signaling, while siRNA against EZH2 serves as a negative control. Mean ± s.e.m. values are shown, n = 3. ***P < 0.001; ns, not significant, determined by ANOVA.

(**E**) qRT–PCR analysis of ARLNC1, AR and AR targets (KLK2, KLK3, FKBP5 and STEAP2) in MDA-PCa-2b cells transfected with ASOs against ARLNC1. Data were normalized to a housekeeping gene, and the levels in control ASO-treated cells were set to 1. Mean ± s.e.m. values are shown, n = 3. Adjusted P values were determined by one-way ANOVA with Dunnett correction for multiple comparisons.

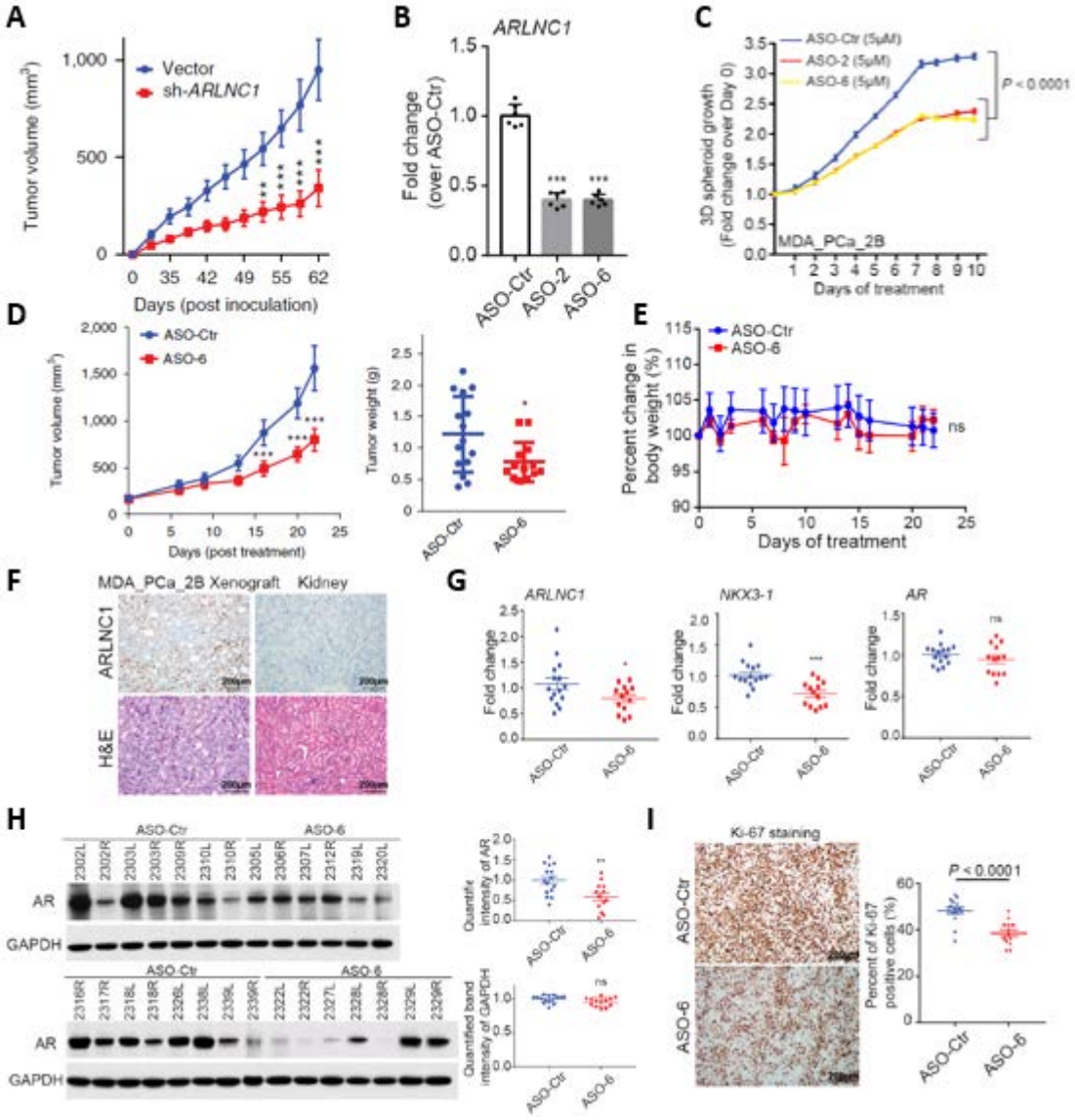


Figure 3.9 Phenotypic effect following ARLNC1 silencing in cell line derived xenograft.
(**A**) Tumor growth of LNCaP-AR cells expressing shRNA targeting ARLNC1 or shRNA vector. Mean ± s.e.m. values are shown. n = 10 independent tumors, ***P < 0.0001, **P = 0.0007, as determined by two-tailed Student's t test.
(**B-C**) ARLNC1 ASOs inhibit MDA-PCa-2b cell proliferation in the 3D sphere models. Cells were harvested at the end of the experiment, and ARLNC1 expression was evaluated by qPCR analysis. Mean ± s.d. values are shown, n = 6. ***P < 0.0001 compared to control-ASO-treated cells, by two-tailed Student's t test.
(**D**) Effect of ASO treatment on the growth of MDA-PCa-2b xenografts in male athymic nude mice, with control ASO (n = 15) or ARLNC1 ASO (n = 13) treatment subcutaneously at 50 mg per kg body weight, five times per week for 3 weeks. Tumors were measured by caliper biweekly and tumor weights were measured at the end point. Mean ± s.d. values are shown. *P = 0.0251, ***P < 0.0001; compared to control ASO by two-tailed Student's t test.

(**E**) Percentage change in mouse body weight over the time of ASO treatment in MDA-PCa-2b xenografts treated with control ASO (n = 15) or ASO targeting ARLNC1 (n = 13). Mean ± s.e.m. values are shown. Significance was determined by two-tailed Student's t test.
(**F**) Representative image of *in situ* hybridization for ARLNC1 in MDA-PCa-2b cell line–derived xenograft. H&E staining is shown for tumor tissue and murine kidney.
(**G**) qRT–PCR analysis of ARLNC1, NKX3-1, and AR in MDA-PCa-2b xenografts treated with control ASO (n = 15) or ASO targeting ARLNC1 (n = 13). Data were normalized to a housekeeping gene (GAPDH), and the average expression level in the control ASO group was set to 1. Mean ± s.e.m. values are shown. *P = 0.0483, ***P = 0.0004; compared to control group by two-tailed Student's t test.
(**H**) Left, immunoblots of AR and GAPDH in MDA-PCa-2b xenografts treated with control ASO (n = 15) or ASO targeting ARLNC1 (n = 13). Right, the relative intensity of the bands was quantified using ImageJ. Mean ± s.e.m. values are shown. **P < 0.005; ns, not significant; compared to the control-ASO-treated group by two-tailed Student's t test.
(**I**) Left, immunohistochemistry staining for Ki67 in MDA-PCa-2b xenograft treated with control ASO or ASO against ARLNC1. Right, summary of Ki67 tumor staining for control (n = 15) or ARLNC1-ASO-treated tumors (n = 13) shows significant difference in Ki67 staining intensity. Mean ± s.e.m. values are shown. P value was determined by two-tailed Student's t test.

Tables

Table 3.1 Primers used in Chapter 3

Name	Sequence
GAPDH-F	CCATCACCATCTTCCAGGAGCGA
GAPDH-R	GGTGGTGAAGACGCCAGTGGA
ACTB-F	CACCATTGGCAATGAGCGGTTC
ACTB-R	AGGTCTTTGCGGATGTCCACGT
HMBS-F	ACGGCTCAGATAGCATACAAGAG
HMBS-R	GTTACGAGCAGTGATGCCTACC
AR-F	CAGTGGATGGGCTGAAAAAT
AR-R	GGAGCTTGGTGAGCTGGTAG
ARLNC1-F	CCTTGTCCACTGGAACTCGT
ARLNC1-R	TATAACCTTGGGGGCCATGA
TMPRSS2-F	CAGGAGTGTACGGGAATGTGATGGT
TMPRSS2-R	GATTAGCCGTCTGCCCTCATTTGT
MYC-F	GCTCGTCTCAGAGAAGCTGG
MYC-F	GCTCAGATCCTGCAGGTACAA
KLK2-F	GGCTCTGGACAGGTGGTAAAGA
KLK2-R	CGGTAATGCACCACCTTGGTGT
KLK3-F	ACGCTGGACAGGGGGCAAAAG
KLK3-R	GGGCAGGGCACATGGTTCACT
FKBP5-F	GCGAAGGAGAAGACCACGACAT
FKBP5-F	TAGGCTTCCCTGCCTCTCCAAA
SLC45A3-F	TCGTGGGCGAGGGGCTGTA
SLC45A3-R	CATCCGAACGCCTTCATCATAGTGT
ETV1-F	GCAAGAAGGCTTCCTGGCTCAT
ETV1-R	CCTTCCCGATACATTCCTGGCT
STEAP2-F	AAAATTTTACATGCCCTGTAATGGA
STEAP2-R	TCTGTATGGAAAAGGATGGTAGCA
NKX3.1-F	CAGTCCCTACTGAGTACTCTTTCTCTC
NKX3.1-R	CACAGTGAAATGTGTAATCCTTGC
POU1F1-F	TCACAGTGCTGCCGAGTGTCTA
POU1F1-R	CCATAGGTTGATGGCTGGTTTCC
IRF1-F	GAGGAGGTGAAAGACCAGAGCA
IRF1-R	TAGCATCTCGGCTGGACTTCGA
FOXA1-F	GCAATACTCGCCTTACGGCTCT
FOXA1-R	GGGTCTGGAATACACACCTTGG
BRD4-F	CGCTATGTCACCTCCTGTTTGC
BRD4-R	ACTCTGAGGACGAGAAGCCCTT

EZH2-F	GACCTCTGTCTTACTTGTGGAGC
EZH2-R	CGTCAGATGGTGCCAGCAATAG
LSD1-F	CTCTTCTGGAACCTCTATAAAGC
LSD1-R	CATTTCCAGATGATCCTGCAGCAA

References

1. Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.
2. Mele, M., et al., *Human genomics. The human transcriptome across tissues and individuals.* Science, 2015. **348**(6235): p. 660-5.
3. Rhodes, D.R., et al., *Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles.* Neoplasia, 2007. **9**(2): p. 166-80.
4. Evers, M.M., L.J. Toonen, and W.M. van Roon-Mom, *Antisense oligonucleotides in therapy for neurodegenerative disorders.* Adv Drug Deliv Rev, 2015. **87**: p. 90-103.
5. Meng, L., et al., *Towards a therapy for Angelman syndrome by targeting a long non-coding RNA.* Nature, 2015. **518**(7539): p. 409-12.
6. Wheeler, T.M., et al., *Targeting nuclear RNA for in vivo correction of myotonic dystrophy.* Nature, 2012. **488**(7409): p. 111-5.
7. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape.* Nat Protoc, 2007. **2**(10): p. 2366-82.
8. Prensner, J.R., et al., *The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex.* Nat Genet, 2013. **45**(11): p. 1392-8.
9. Mehra, R., et al., *A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer.* Neoplasia, 2014. **16**(12): p. 1121-7.
10. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.

Chapter 4

ARLNC1, a lineage-specific long-noncoding RNA that regulates Androgen Receptor signaling axis[1]

Abstract

Having characterized the basic transcript features and functional phenotypes of ARLNC1, we went on to investigate the mechanisms that drive ARLNC1 expression, and explore how ARLNC1 regulates the AR-signaling axis. In this Chapter, we identified the transcription factors that drive ARLNC1 expression in a lineage-specific manner, each of which have been previously reported to interact with AR, and therefore underscores AR as the major regulator. In addition, we have dissected the cellular mechanism of ARLNC1-mediated modulation of AR transcripts, and find that the lncRNA affects the cytoplasmic level of the AR transcripts. ARLNC1 regulates AR mRNA post-transcriptionally via specific RNA-RNA associations *in situ*. our results now robustly demonstrate a feedforward loop in AR signaling axis, where AR, in concert with other lineage specific

[1] This chapter was previously published as part of the following manuscript: Zhang, Y., et al., Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. *Nat Genet*, 2018. 50(6): p. 814-824.

factors, transcriptionally regulates ARLNC1, and that ARLNC1 regulates its master regulator (AR) post-transcriptionally.

Introduction

Recently, out group identified a prostate lineage-specific, pro-oncogenic lncRNA named ARLNC1 that is overexpressed in prostate cancers. Subsequent studies found that ARLNC1 possesses diagnostic potential for both primary and metastatic prostate cancer. Phenotypically, ARLNC1 silencing inhibits cell proliferation and induces apoptosis. Interestingly, analysis of ARLNC1-regulated transcriptome revealed enrichment of AR targets. ARLNC1 loss resulted in attenuated global AR signaling.

However, several questions remain regarding ARLNC1 expression pattern and functions: (1) What are the transcription factors that determine its prostate lineage-specific expression pattern? (2) What is the mechanistic nature of ARLNC1 regulation of AR signaling? Is it mediated by RNA-chromatin, RNA-RNA, or RNA-protein interactions?

Results

ARLNC1 is directly regulated by AR and FOXA1

Since ARLNC1 was identified as an AR-regulated lncRNA, we asked the question whether ARLNC1 is directly or indirectly regulated by AR. We thus inspected the ARLNC1 promoter region for AR occupancy in AR ChIP-Seq data from both DHT-stimulated VCaP and LNCaP cells (**Figure 4.1A**). An androgen-induced AR peak was identified in both cell

lines. Importantly, this AR-binding site was also observed in prostate tissue samples and contained a canonical androgen response element[1] (**Figure 4.1A**). These observations were corroborated by ChIP–qPCR in MDA-PCa-2b cells, which showed the highest level of ARLNC1 expression (**Figure 4.1B**). Considering the observation that ARLNC1 expression is prostate tissue-specific, while AR expression is not as much, we searched for additional regulators (transcription factors and epigenetic modifiers) of this gene (**Figure 4.2**). Motif analysis of the ARLNC1 promoter region identified several transcription factor binding sites, including a FOXA1 response element. To further validate ARLNC1 gene regulation by AR and FOXA1, we evaluated ARLNC1 transcript levels following AR or FOXA1 knockdown. AR or FOXA1 loss resulted in decreased expression of ARLNC1, along with other canonical AR target genes that served as positive controls (**Figure 4.1C, Figure 4.3A**). ChIP-Seq and ChIP–PCR analysis additionally confirmed the putative FOXA1-binding motif on the ARLNC1 promoter (**Figure 4.3B**). Together, these observations suggest that ARLNC1 is directly regulated by AR and modestly regulated by FOXA1, which partially explains the tissue-specific expression pattern of ARLNC1, as expression of these two factors overlaps nearly exclusively in prostate tissue[2, 3] (**Figure 4.3C**).

ARLNC1 affects AR signaling

In the previous chapter, we performed gene expression profiling of wild-type and ARLNC1-knockdown MDA-PCa-2b cells (**Figure 3.8A**). Gene ontology (GO) pathway enrichment analysis of the differentially expressed genes revealed deregulation of

androgen signaling, and this finding was corroborated by qPCR analysis (**Figure 3.8D, E**). To confirm this observation, we generated an AR target gene signature from MDA-PCa-2b cells stimulated with DHT and performed gene set enrichment analysis (GSEA) using this gene signature (**Figure 4.4A-B**). Knockdown of ARLNC1 led to suppression of genes positively regulated by AR and upregulation of genes negatively regulated by AR (**Figure 4.4A, C**). This was further confirmed by an AR reporter activity assay (**Figure 4.4G-H**). Interestingly, ARLNC1 knockdown also had a significant effect on the messenger RNA and protein levels of AR (**Figure 4.4E-F**), suggesting direct regulation of AR by ARLNC1. However, we found that ARLNC1 overexpression did not affect AR and its signaling cascade (**Figure 4.4I**).

RNA-RNA interaction between ARLNC1 and AR in vitro and in situ

Non-coding RNAs have been shown to target mRNAs via direct or indirect RNA–RNA interaction[4-8]. To identify target mRNAs that could interact with ARLNC1, we performed an unbiased prediction of RNA–RNA interactions using IntraRNA[9, 10]. Interestingly, the 3' UTR of the AR transcript was identified as a target of ARLNC1 (**Figure 4.5A-B**). An *in vitro* RNA–RNA interaction assay between the 3' UTR of AR and full-length ARLNC1 confirmed this *in silico* prediction (**Figure 4.5C**). To evaluate this interaction in the context of the cellular environment, multiplexed smFISH for AR and ARLNC1 transcripts was performed in MDA-PCa-2b cells. On co-staining MDA-PCa-2b cells with either AR and a panel of lncRNAs, or ARLNC1 and a panel of mRNAs, we observed specific co-localization between AR and ARLNC1 transcripts in the nucleus

within foci that were typically larger than individual molecules (**Figure 4.5D, H**). The extent of colocalization was much higher than that expected from co-incidental co-localization with an abundant transcript, such as MALAT1 or GAPDH (**Figure 4.5E-F**). More specifically, co-localization typically occurred at a stoichiometry of 2:1 ARLNC1/AR, which accounted for ~10–20% of all AR and ARLNC1 transcripts in the cell (**Figure 4.5G**). Furthermore, AR–ARLNC1 co-localization was observed in ARLNC1-positive prostate cancer tissues (**Figure 4.6**).

Domain on ARLNC1 that mediates its interaction with AR

Using an *in vitro* RNA–RNA binding assay, we identified nucleotides (nt) 700–1300 of ARLNC1 to be critical for binding to the AR 3' UTR (**Figure 4.7A-B**). To confirm this observation within the cellular context, we ectopically overexpressed different fragments of ARLNC1 together with AR in U2OS osteosarcoma cells. In this exogenous system, co-localization between AR and ARLNC1 was once again demonstrated, wherein co-localization was dependent on the presence of nt 700–1300 of ARLNC1 (**Figure 4.7C-D**). Furthermore, incubation with a pool of antisense oligonucleotides (ASOs) that blocked the interaction sites led to a significant reduction in ARLNC1-AR interaction *in vitro* and *in situ* (**Figure 4.8A-E**). Decreased AR signaling was also observed following blocking of this interaction (**Figure 4.8F-G**).

ARLNC1 regulates the cytoplasmic levels of AR transcripts.

We then sought to delineate the mechanism of ARLNC1-mediated AR regulation. We first monitored the stability of these two transcripts and found that AR and ARLNC1 have similar half-lives of ~9 h (**Figure 4.9A**). As ARLNC1 depletion resulted in a striking reduction of AR protein levels, much more than could be explained by AR transcript reduction, we hypothesized that ARLNC1 could affect AR post-transcriptionally. To test this hypothesis, we tracked subcellular localization of AR transcripts using smFISH after depleting ARLNC1. We confirmed successful *in situ* knockdown of ARLNC1 using siRNAs, antisense oligonucleotide (ASO) and the blocking oligonucleotides that targeted ARLNC1-AR interaction (ASO-blocking) in MDA-PCa-2b cells (**Figure 4.9B-C**).

Quantification of the subcellular distribution of ARLNC1 suggested that the nuclear fraction of ARLNC1 was enriched only in the ARLNC1 siRNA (si-ARLNC1) condition (**Figure 4.9E**), a result that was expected for siRNAs that are typically more functional in the cytosol[11]. Surprisingly, ARLNC1 knockdown or obstruction of the AR-ARLNC1 interaction increased the nuclear AR fraction by dramatically decreasing cytoplasmic levels of the AR transcript (**Figure 4.9F-G**). This observation was further supported by BrU-Seq and BrUChase-Seq, two high-throughput tools that monitor transcript synthesis and stability. On ARLNC1 knockdown, the synthesis rate of the AR transcript remained the same, while the stability of the transcript decreased, particularly through the 3' UTR region (**Figure 4.9D**). Taken together, our data suggest that ARLNC1 regulates the cytoplasmic levels of AR transcripts. Furthermore, the transcriptional coupling between

AR and ARLNC1 transcripts is mediated by direct interactions that are encoded in their sequences.

Discussion

Our results now robustly demonstrate that AR, together with other lineage specific cofactors, transcriptionally regulates ARLNC1, and that ARLNC1 regulates its master regulator (AR) post-transcriptionally via co-localizing to AR mRNA (**Figure 4.10**). While lncRNAs are commonly shown to stabilize mRNAs via genomically anti-sense mechanisms, we found that two transcripts from distinct chromosomal loci could associate inside cells to regulate mRNA abundance. The mechanism we identified echoes previous studies on lncRNAs: 1/2-sbsRNAs[6], BACE1-AS[5] and TINCR[8], which highlights the role of lncRNA in increasing or decreasing RNA stability. Moreover, acting upstream of AR signaling presents the possibility that targeting ARLNC1 may afford an additional option to patients that have *de novo* or acquired resistance to therapies targeting AR itself (enzalutamide or abiraterone, etc).

Materials and Methods

ChIP–qPCR assay

AR, FOXA1 or NKX3-1 ChIP was performed following our previous protocol[12]. (Antibodies: AR, Millipore cat. no. 06-680; FOXA1, Thermo Fisher cat. no. PA5-27157; NKX3-1, CST cat. no. 83700S.) Primers targeting the CYP2B7 promoter were purchased from CST (cat. no. 84846).

ChIP-Seq data analysis

ChIP-Seq data from published external and in-house data sets, GSE56288 and GSE55064, were reanalyzed using a standard pipeline. Briefly, groomed reads (vendor QC, adapter removal) were aligned to the GRCh38 reference genome using STAR settings that disable spliced alignment: outFilterMismatchNoverLmax: 0.05, outFilterMatchNmin: 16, outFilterScoreMinOverLread: 0, outFilterMatchNminOverLread: 0, alignIntronMax: 1. Improperly paired alignments and non-primary alignments were discarded. Peaks were called using MACS2 (callpeak --broad --qvalue 0.05 --broad-cutoff 0.05 and callpeak –callsummits --qvalue 0.05)[13] and Q (-n 100000)[14]. ChIP enrichment plots were computed from alignment coverage files (BigWig)[15] as trimmed (trim = 0.05) smooth splines (spar = 0.05). The baseline (non-specific) ChIP signal was estimated from genomic windows furthest from the center of the queried region (peak summit, TSS) and subtracted from each signal before plotting.

siRNA-mediated knockdown

siRNA oligonucleotides targeting ARLNC1, AR, FOXA1, BRD4, NKX3-1, LSD1, IRF1, POU1F1 or EZH2 and a non-targeting siRNA were purchased from Dharmacon. (si-AR-pool, cat. no. L-003400-00-0005; si-FOXA1, cat. no. LU-010319-00-0005; si-BRD4, cat. no. LU-004937-00-0002; si-NKX3-1, cat. no. LU-015422-00-0005; si-LSD1, cat. no. LU-009223-00-0002; si-IRF1, cat. no. LU-011704-00-0005; si-POU1F1, cat. no. LU-012546-00-0005; si-EZH2, cat. no. L-004218-00-0005; si-NT, cat. no. D-001810-01-05.) For AR

knockdown, two more siRNAs were purchased from Life Technologies (no. HSS179972 and no. HSS179973). Transfections with siRNA (50 nM) were performed with Lipofectamine RNAiMAX according to the manufacturer's instructions. RNA and protein were harvested for analysis 72 h after transfection.

Gene Set Enrichment Analysis

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

Androgen receptor reporter gene assay

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

Overexpression of ARLNC1

Full-length ARLNC1 was amplified from MDA-PCa-2b cells and cloned into the pCDH clone and expression vector (System Biosciences). Insert sequences were validated by Sanger sequencing at the University of Michigan Sequencing Core.

RNA in vitro transcription

Linearized DNA templates for full-length ARLNC1, ARLNC1 fragments, ARLNC1 deletion, antisense ARLNC1, LacZ, SCHLAP1-AS, THOR and AR-3' UTR-1-980 were synthesized using T7-containing primers (**Table 4.1**). *In vitro* transcription assays were performed with T7 RNA polymerase (Promega) according to the manufacturer's instructions. For BrU-labeled RNA synthesis, 5-bromo-UTP was added to the transcription mix. At the end of transcription, DNA templates were removed by Turbo DNase (ThermoFisher), and RNA was recovered using RNA Clean and Concentrator Kit (Promega). RNA size and quality was further confirmed by the Agilent Bioanalyzer.

RNA–RNA in vitro interaction assay

Twenty-five microlitres of Protein A/G Magnetic Beads were washed twice with RIP wash buffer (Millipore, cat. no. CS203177) before incubating with BrU antibody for 1 h at room temperature. After antibody conjugation, beads were washed twice with RIP wash buffer and then resuspended in incubation buffer containing RIP wash buffer, 17.5 mM EDTA (Millipore, cat. no. CS203175) and RNase Inhibitor (Millipore, cat. no. CS203219). Equal

amounts (5 pmol) of BrU-labeled RNAs (ARLNC1, ARLNC1-AS, ARLNC1-1-1300, ARLNC1-1301-2786, ARLNC1-1-700, ARLNC1-701-1300, ARLNC1-del-701-1300, LacZ, SCHLAP1-AS, THOR) were incubated with beads in Incubation Buffer for 2 h at 4 °C. Following incubation, 2.5 pmol of the AR 3′ UTR-1–980 RNA fragment was added into individual tubes and incubated overnight at 4 °C. After incubation, beads were washed six times with RIP Wash Buffer. To recover RNA, beads were digested with proteinase K buffer containing RIP Wash Buffer, 1% SDS (Millipore, cat. no.CS203174), and 1.2 μg/μl proteinase K (Millipore, cat. no. CS203218) at 55 °C for 30 min with shaking. After digestion, RNA was extracted from supernatant using the miRNeasy kit (QIAGEN), and reverse transcription was performed using the Superscript III system (Invitrogen). The amount of AR 3' UTR-1–980 recovered in each interaction assay was quantified by qPCR analysis. Data were normalized to the ARLNC1-AS control, using the ΔCt method. We designed ASOs blocking the AR–ARLNC1 interaction sites (ASO-blocking, Ionis Pharmaceuticals) and used them in the in vitro interaction assays. Data were normalized to the control ASO, using the ΔCt method.

RNA stability assay

LNCaP cells were treated with 5 μg/ml of actinomycin D for various times as indicated. RNA was extracted and qRT–PCR was carried out as described above. RNA half-life was calculated by linear regression analysis.

smFISH

smFISH and image analysis were performed as described previously[16, 17]. Probe sequences targeting ARLNC1, PCAT1, DANCR, AR, EZH2 and FOXA1 were designed using the probe design software in https://www.biosearchtech.com/stellaris-designer and are listed in **Table 4.2**. TERRA probes were designed as described previously[18]. Other probes were purchased directly from the LGC-Biosearch. U2-OS cells were seeded in 6-well dishes and transfected with ARLNC1-expression vector alone, or in combination with AR expression vector, using Fugene-HD (Promega) according to the manufacturer's protocol. Cells were incubated for 24 h, reseeded into 8-well chambered coverglasses, and were formaldehyde-fixed for smFISH (as described above) after 24 h.

BrU-Seq and BrUChase-seq

BrU-Seq and BrUChase-Seq assays were performed as previously described[19, 20], with MDA-PCa-2b cells treated with either si-NT or si-ARLNC1. BrU-labeling was performed for 30 min, and chase experiments were performed for 6 h.

Figures

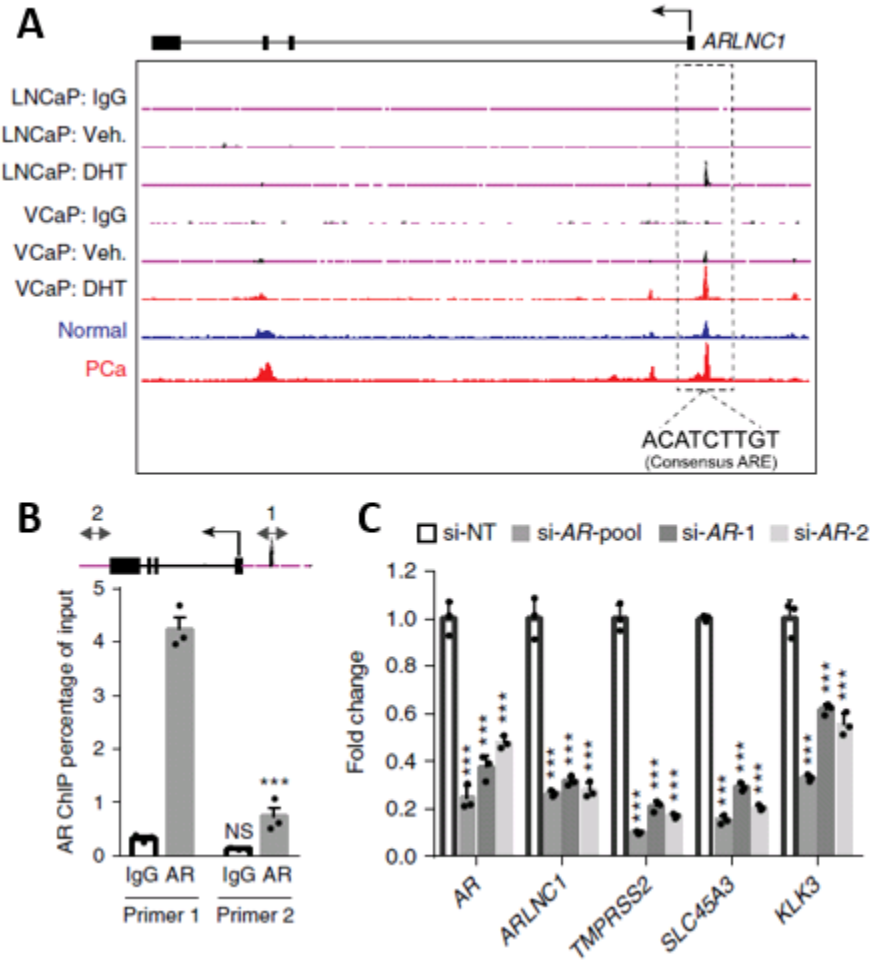


Figure 4.1 ARLNC1 is directly regulated by AR

(**A**) AR ChIP–seq in prostate cancer cell lines and tissues. Normalized ChIP–seq enrichment is shown. Top, AR or control (IgG) ChIP–seq results across the *ARLNC1* locus in LNCaP and VCaP cells with vehicle (ethanol) or DHT treatment. Bottom, AR ChIP–seq in benign prostate and clinically localized prostate cancer tissue. ARE, androgen response element.

(**B**) ChIP–qPCR in MDA-PCa-2b cells showing AR or IgG enrichment (ChIP/input) over the ARLNC1 promoter region (primer 1) or a control region (primer 2). Data are shown as the mean ± s.e.m. (n = 3 biologically independent samples). ***P (adjusted) < 0.0001, NS (not significant): P = 0.5746, compared to the control region (primer 2) by ANOVA with Sidak correction for multiple comparisons. Top, schematic of the amplicon locations for ChIP–qPCR validation.

(**C**) Expression of AR and AR target genes (ARLNC1, TMPRSS2, SLC45A3 and KLK3) in MDA-PCa-2b cells transfected with control siRNA (si-NT) or siRNAs against AR (si-AR-pool, si-AR-1, si-AR-2). Mean ± s.e.m. values are shown, n = 3 biologically independent samples. ***P = 0.0001, determined by ANOVA with Dunnett's multiple-comparisons test.

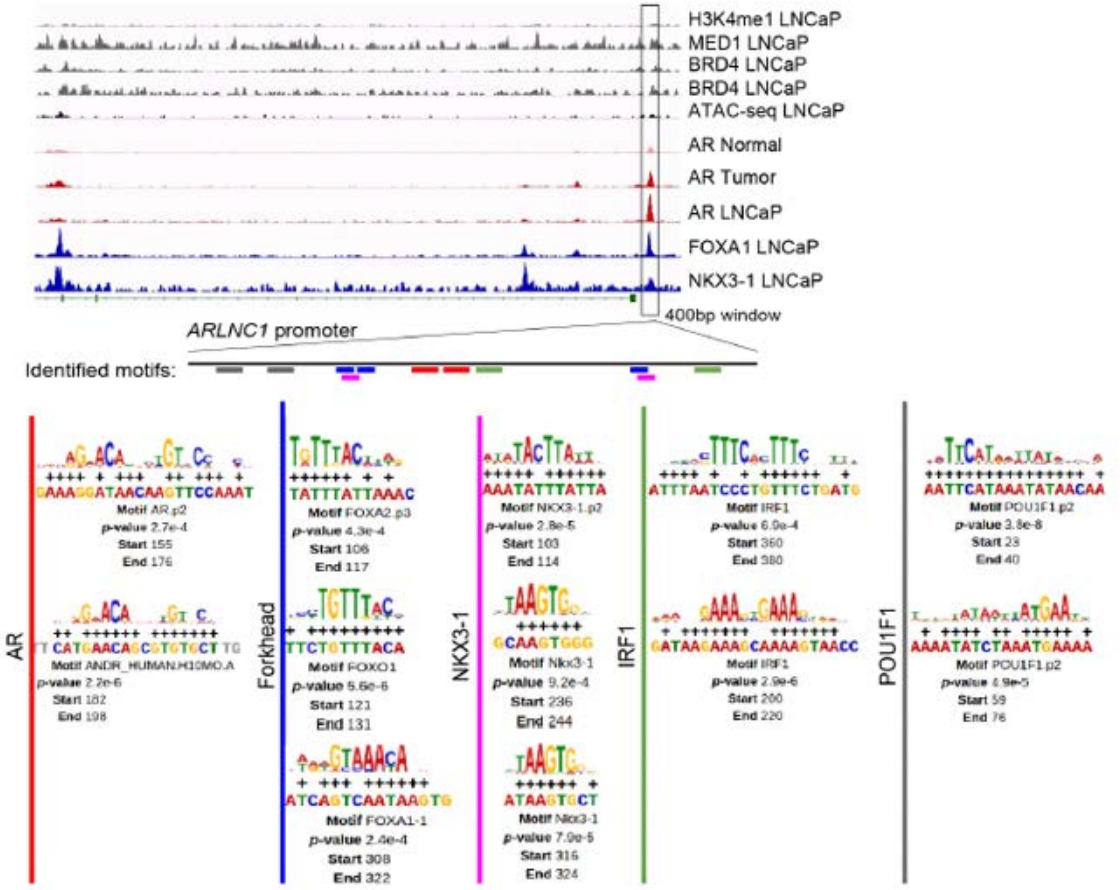


Figure 4.2 Predicted transcriptional regulation at *ARLNC1* promoter
ChIP–seq peaks of H3K4me1, MED1, BRD4, FOXA1, and NKX3-1 from LNCaP cells at the *ARLNC1* promoter region. AR ChIP–seq tracks from normal prostate and prostate cancer are also shown. Motif analysis results are summarized at the bottom, suggesting possible binding of AR, FOXA1, NKX3-1, IRF1, and POU1F1 at the *ARLNC1* promoter region.

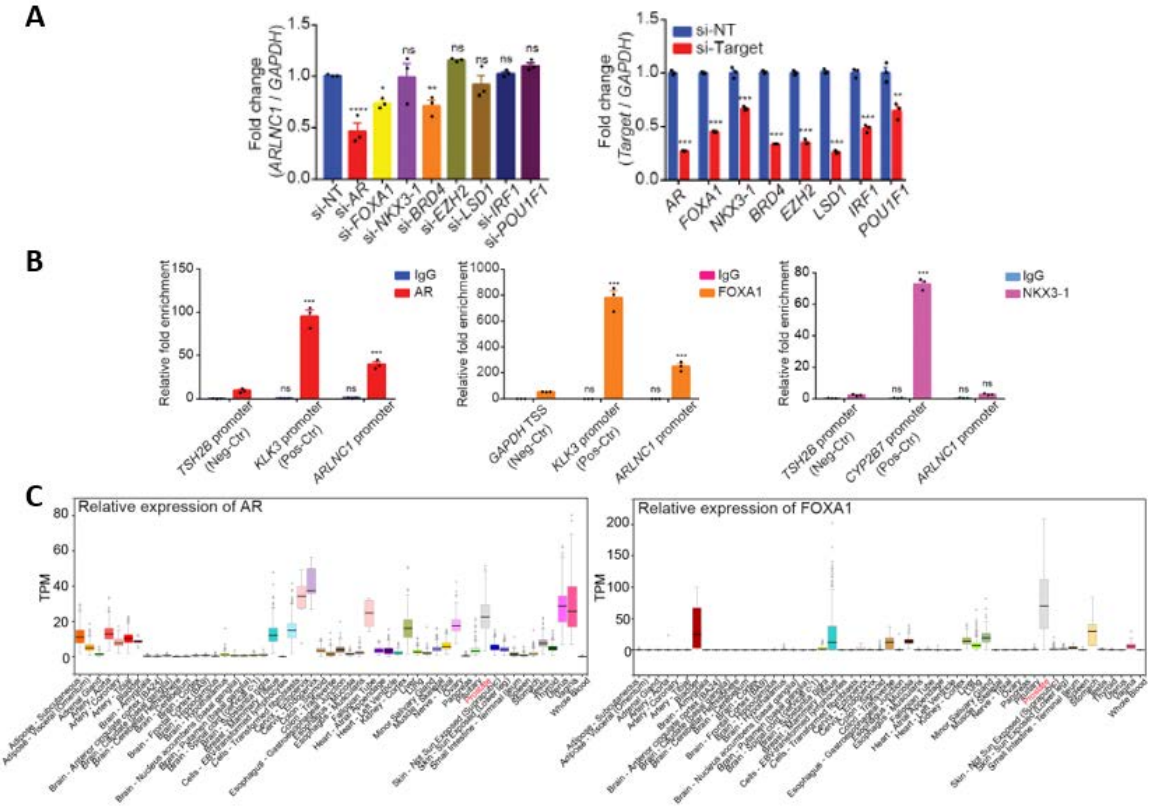


Figure 4.3 Lineage specific expression of ARLNC1 is driven by AR and FOXA1

(**A**) Top, qPCR analysis of ARLNC1 expression in LNCaP cells, following treatment with siRNAs targeting AR, FOXA1, NKX3-1, BRD4, EZH2, LSD1, IRF1, and POU1F1. Mean ± s.e.m. values are shown, n = 3. *Adjusted P = 0.0436, **adjusted P = 0.0264, ****adjusted P = 0.0001, compared to control siRNA (si-NT) by ANOVA with Dunnett correction. Bottom, the on-target effect of siRNAs was evaluated by qPCR analysis. Mean ± s.e.m. values are shown, n = 3. **P = 0.006, ***P < 0.001, compared to control siRNA (si-NT) by two-tailed Student's t test.

(**A**) ChIP-PCR analysis in MDA-PCa-2b cells showing relative enrichment (ChIP/input) of AR, FOXA1, NKX3-1, or IgG over ARLNC1 promoter region or control region. Error bars represent mean ± s.e.m. (n = 3). ***Adjusted P = 0.0001 compared to negative control, by ANOVA with Dunnett correction for multiple comparisons.

(**C**) Relative expression (TPM) of AR (left) and FOXA1 (right) across a panel of normal tissues in the GTEx normal tissue RNA-seq cohort (n = 8,745 samples). Box-plot definition: center, median; box limits, 1st and 3rd quartile; whiskers follow the 1.5 rule.

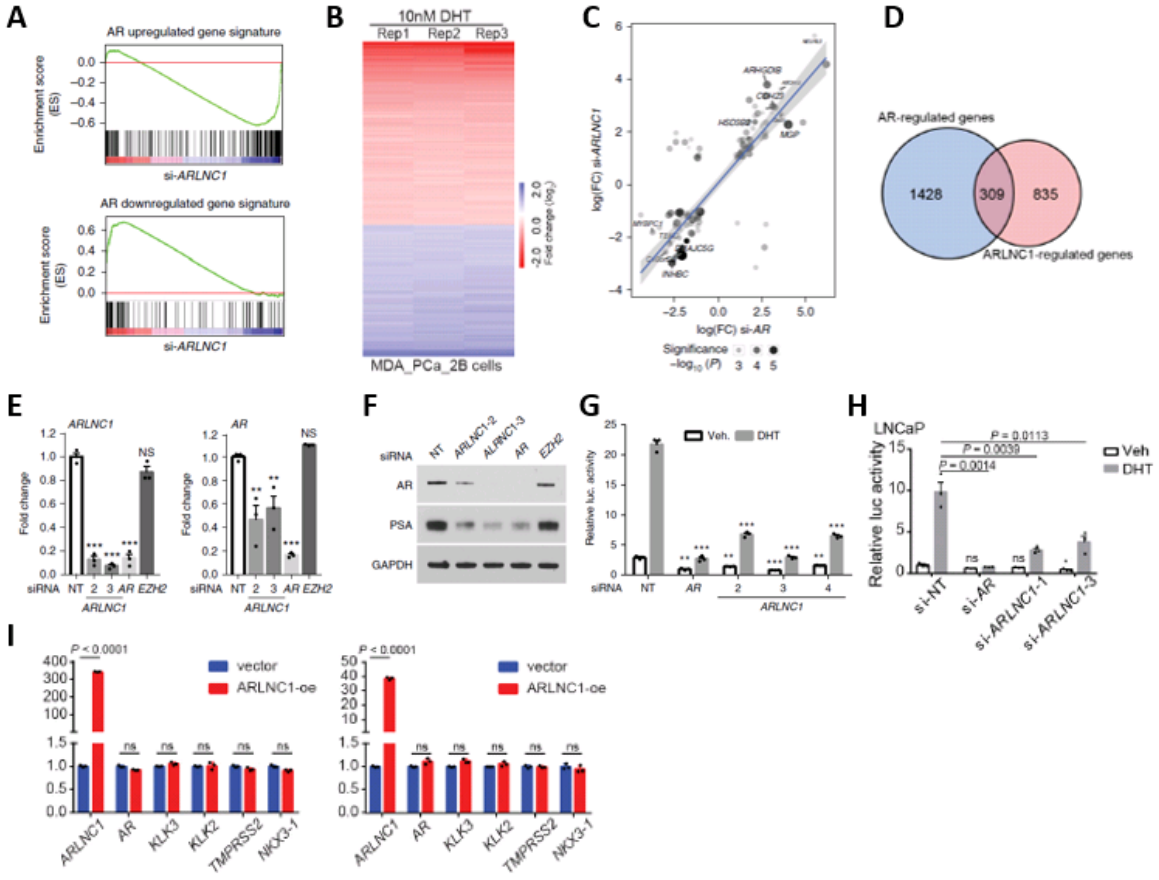


Figure 4.4 A positive feedback loop between ARLNC1 and AR signaling

(**A**) GSEA showing significant enrichment of the ARLNC1-regulated gene set with respect to the AR target gene sets (n = 3 independent gene expression profiles). Shown are enrichment plots for gene sets consisting of genes positively regulated by AR (top) and genes negatively regulated by AR (bottom).

(**B**) Reproducibility of expression profiling following 10 nM DHT treatment in MDA-PCa-2b cells. The most significant AR targets were used to derive a gene signature of the AR response.

(**C**) Comparison of ARLNC1-regulated and AR target genes based on RNA–seq following knockdown of AR and ARLNC1. Significant genes and their log-transformed fold changes in either of the conditions are shown (n = 2 independent cell cultures per-condition). Combined significance levels, determined by a limma-moderated t test (across both knockdowns), are indicated by circle size.

(**D**) Overlap between genes differentially expressed upon AR knockdown and ARLNC1 knockdown in MDA-PCa-2B cells.

(**E**) qRT–PCR analysis of ARLNC1 and AR in MDA-PCa-2b cells transfected with siRNAs against ARLNC1, AR, EZH2 or non-specific control (NT). siRNA against AR serves as a positive control for inhibition of AR signaling, while siRNA against EZH2 serves as a negative control. Mean ± s.e.m. values are shown, n = 3. **P < 0.01, ***P = 0.0001, determined by ANOVA with Dunnett correction.

(**F**) Immunoblots of AR, PSA and GAPDH in MDA-PCa-2b cells transfected with siRNAs

against ARLNC1, AR, EZH2 or non-specific control (NT). The experiments were repeated three times independently with similar results.

(**G**) siRNA knockdown of ARLNC1 in MDA-PCa-2b cells impairs AR signaling as determined by AR reporter gene assay. siRNA against AR serves as a positive control for inhibition of AR signaling. Mean ± s.e.m. values are shown, n = 3 biologically independent cell cultures. **P < 0.01, ***P = 0.0001, determined by ANOVA with Dunnett correction.

(**H**) siRNA knockdown of ARLNC1 in LNCaP cells impaired AR signaling by AR reporter gene assay. siRNA against AR served as a positive control for AR signaling inhibition. Mean ± s.e.m. values are shown, n = 3. *P = 0.0233; ns, not significant. P values determined by two-tailed Student's t test. Veh, vehicle control.

(**I**) qPCR analysis of ARLNC1 and AR signaling genes in LNCaP cells (left) and MDA-PCa-2b cells (right) transfected with ARLNC1-expressing vector or control vector. Mean ± s.d. values are shown, n = 3. ns, not significant, compared to the vector group by two-tailed Student's t test.

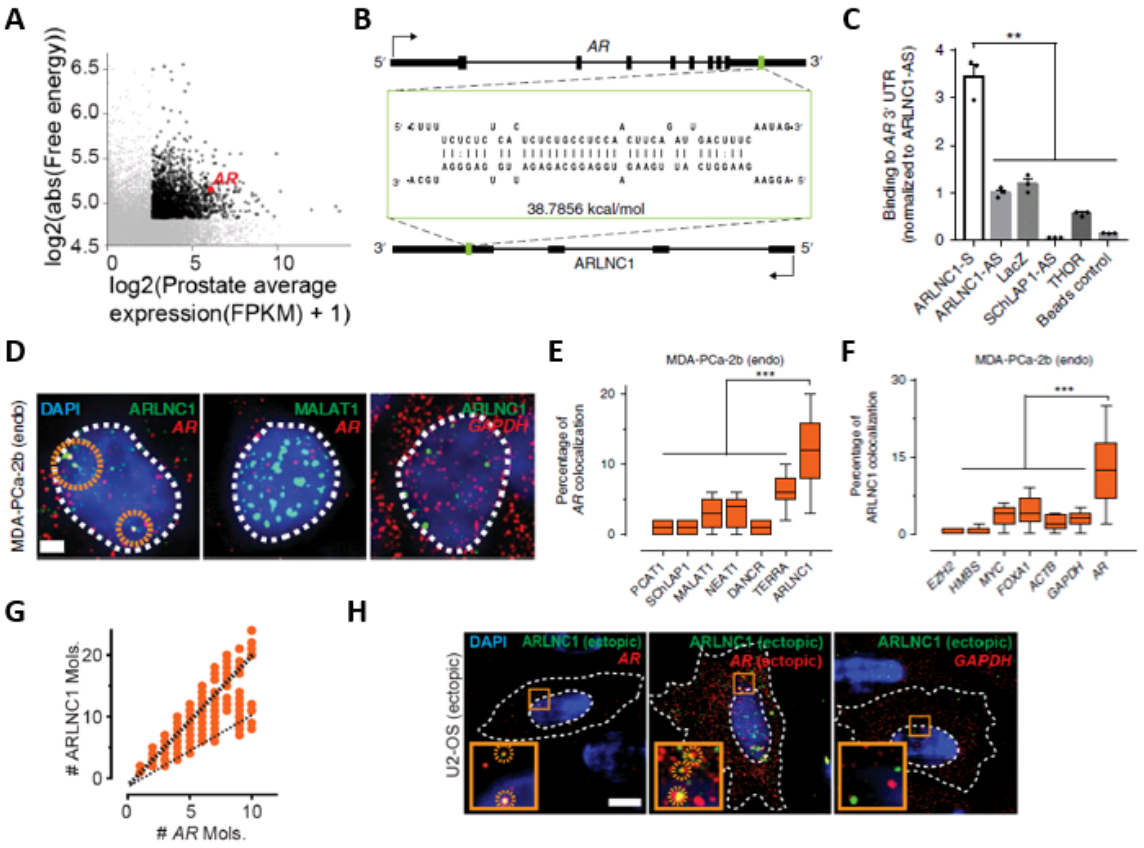


Figure 4.5 Co-localization of AR mRNA and ARLNC1 *in vitro* and in prostate cancer cells

(**A**) *In silico* prediction of ARLNC1 RNA-binding partners, with the y axis representing log2 transformed absolute RNA-binding energy between ARLNC1 and various RNA species while the x axis depicts log2-transformed average expression levels of these RNAs in prostate cancer.

(**B**) Schematic of the predicted RNA-RNA interaction between ARLNC1 and the 3′ UTR of AR.

(**C**) ARLNC1 interacts with the AR 3′ UTR in an in vitro RNA–RNA interaction assay. Compared to a panel of control RNAs (ARLNC1 antisense, LacZ, SChLAP1-AS, THOR), ARLNC1 binds to AR 3′ UTR-1–980 with high affinity. The binding affinity was quantified by qPCR analysis of the AR 3′ UTR. Data were normalized to the ARLNC1-AS control. Mean ± s.e.m. are shown, n = 3. **P < 0.001, by two-tailed Student's t test.

(**D-F**) smFISH depiction of AR–ARLNC1 colocalization in situ. (**D**), Representative pseudocolored images of MDA-PCa-2b cell nuclei stained for the appropriate endogenous (endo) transcripts (green, red) and with DAPI (nucleus, blue). Scale bar, 5 μm. The orange circles represent regions of colocalization. (**E-F**), Quantification of the percentage of AR or ARLNC1 molecules colocalizing with a panel of lncRNAs (**E**) or mRNAs (**F**). The center line and whiskers depict the median and range, respectively, and the box extends from the 25th to the 75th percentile (n = 50 cells for each sample aggregated from 3 independent experiments). ***P < 0.0001, by two-tailed Student's t test.

(**G**) Stoichiometry of ARLNC1:AR colocalization. The dotted line represents a 1:1 stoichiometry, whereas the bold dotted line represents 2:1 stoichiometry. Each dot represents the number of

molecules when AR and ARLNC1 colocalize in any given cell (n = 1,000 colocalized spots from nine independent samples).
(**H**) Representative pseudocolored images of U2-OS cells ectopically expressing ARLNC1 alone (green, left and right) or both ARLNC1 and AR (red, middle) and stained for the appropriate transcripts and DAPI (blue). Scale bar, 10 μm. The inset depicts a zoomed-in view of the orange box in the image.

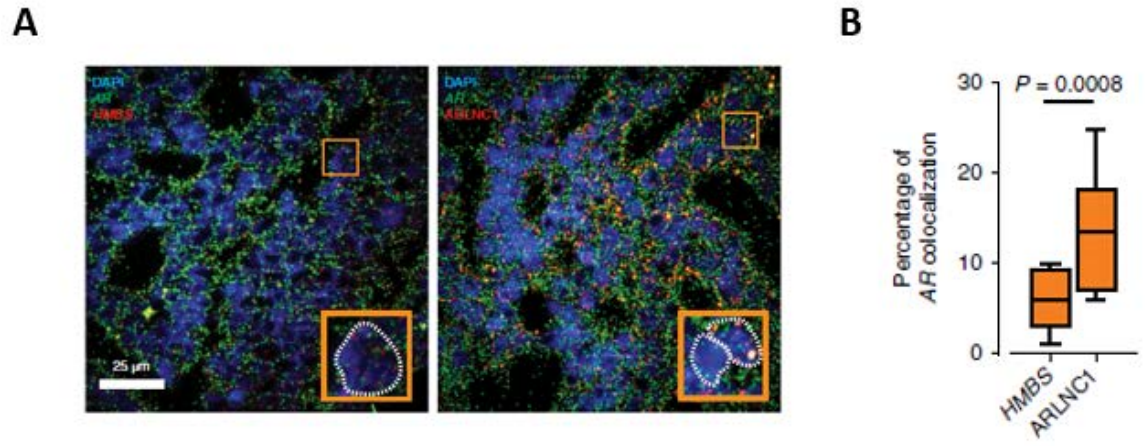


Figure 4.6 Co-localization of AR mRNA and ARLNC1 in prostate cancer tissues
(**A**) Representative pseudocolored images of ARLNC1-positive prostate cancer tissues stained with DAPI (nucleus, blue) and for AR (green), HMBS or ARLNC1 transcripts (smFISH). Inset, 5.5 × 5.5 μ m^2 zoomed-in view of the box within the main panel.
(**B**) Quantification of the percentage of AR molecules colocalizing with HMBS transcripts or ARLNC1. The center line and whiskers depict the median and range, respectively, and the box extends from the 25th to the 75th percentile (n = 15 field-of-views for each sample aggregated from 3 independent tissues). $P < 0.001$, by two-tailed Student's t test.

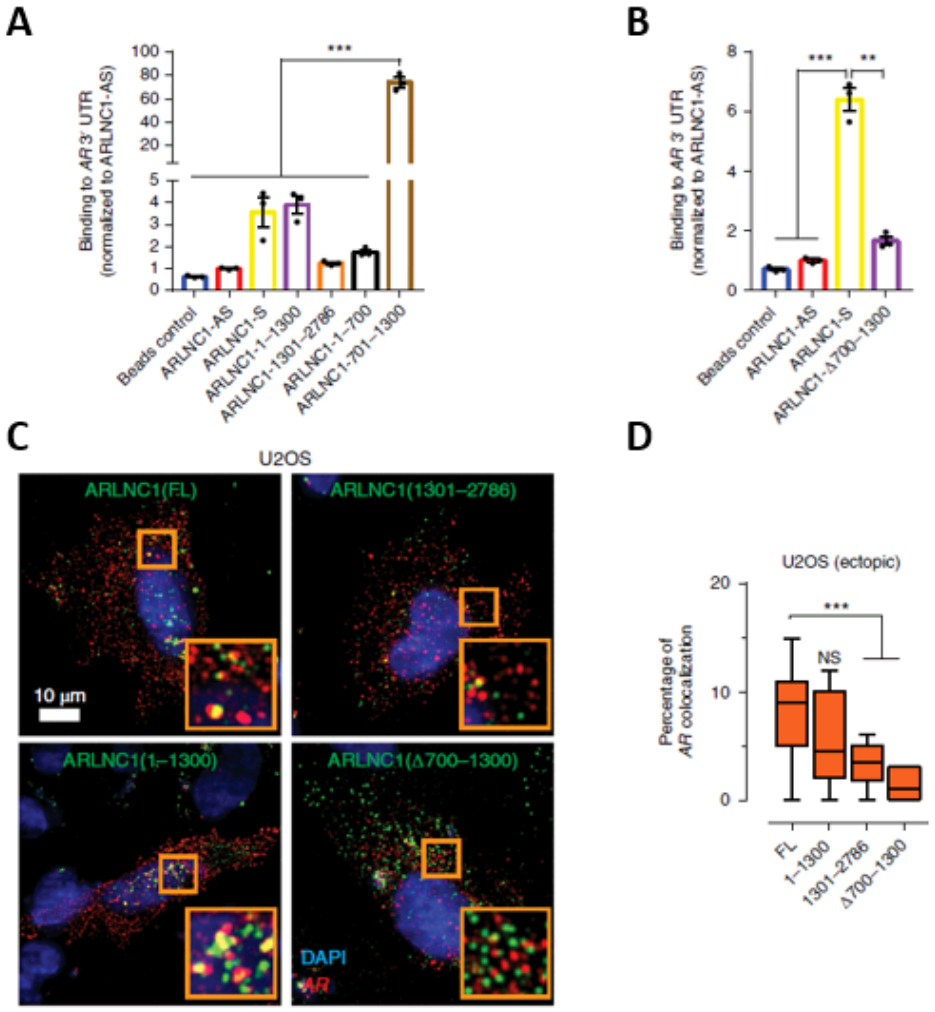


Figure 4.7 Identification of the ARLNC1 fragment mediating RNA–RNA interaction with AR mRNA

(**A**) *In vitro* RNA–RNA interaction assay identifies nucleotides 700-1300 on ARLNC1 as critical binding site to AR 3′ UTR-1-980. ARLNC1 fragments covering nucleotides 700–1300 display comparable or higher AR 3′ UTR binding affinity than ARLNC1-S, with ARLNC1-700-1300 exhibiting the highest binding affinity. Data were normalized to the ARLNC1-AS control. Mean ± s.e.m. values are shown, n = 3. ***P (adjusted) = 0.0001, determined by ANOVA with Dunnett's multiple-comparisons test.

(**B**) Deletion ofnucleotides 700–1300 on ARLNC1 results in impaired binding to the AR 3′ UTR, as shown by *in vitro* RNA–RNA interaction assay. Data were normalized to the ARLNC1-AS control. Mean ± s.e.m. values are shown, n = 3. ***P = 0.0001, **P = 0.0003, by two-tailed Student's t test.

(**C-D**) smFISH shows that nucleotides 700–1300 in ARLNC1 are important for colocalization in situ. (**C**) Representative pseudocolored images of U2OS cells stained with DAPI (nucleus, blue) and for ARLNC1 (green) and AR transcripts. Inset, 10 × 10 μ m^2 zoomed-in view of the orange box in the main image. (**D**) Quantification of the percentage of AR molecules colocalizing with

various ARLNC1 fragments. The center line and whiskers depict the median and range, respectively, and the box extends from the 25th to the 75th percentile (n = 50 cells for each sample aggregated from 3 independent experiments). ***P < 0.0001, by two-tailed Student's t test. NS, not significant.

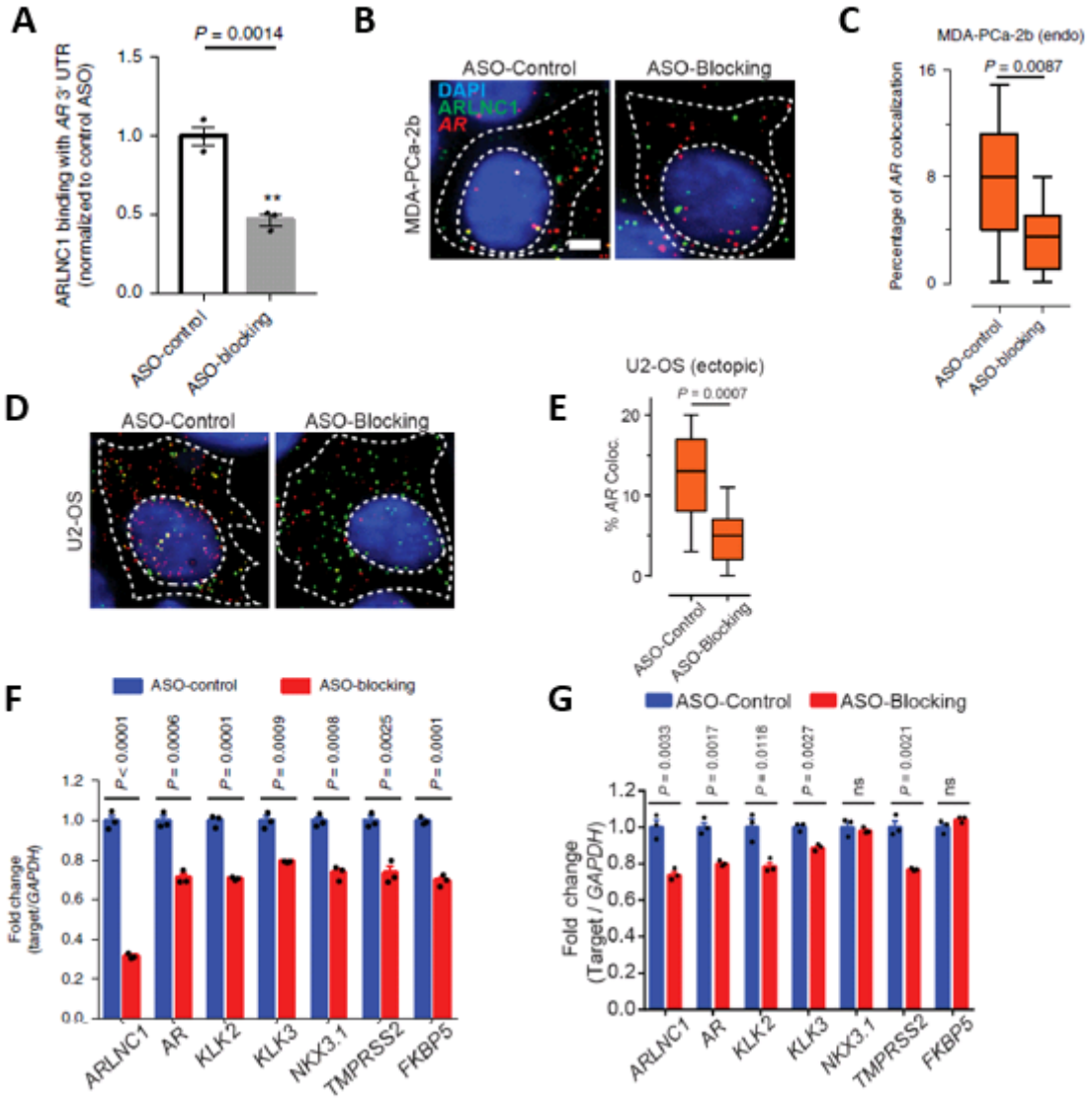


Figure 4.8 Disrupting RNA-RNA interaction between ARLNC1 and AR by antisense oligonucleotides

(**A**) ASOs targeting nucleotides 700–1300 on the ARLNC1 transcript (ASO-blocking pool) inhibit ARLNC1 interaction with the AR 3′ UTR. *In vitro* RNA–RNA interaction assays were performed using ARLNC1 and the AR 3′ UTR, with the addition of the blocking ASO pool or control ASO. Data were normalized to the control ASO. Mean ± s.e.m. values are shown, n = 3. P = 0.0014, by two-tailed Student's t test.

(**B-C**) Representative pseudocolored images of MDA-PCa-2b cells stained for DAPI (nucleus, blue) and ARLNC1 (green) and AR transcripts, following treatment of blocking ASOs targeting the ARLNC1:AR 3' UTR interaction. Quantifications of colocalization are depicted in (**C**). Center line and whiskers depict the median and range, respectively, and the box extends from the 25th to 75th percentiles (n = 60 cells for each sample aggregated from three independent experiments). P value was computed by two-tailed Student's t test.

(**D**-**E**) Representative pseudocolored images of U2-OS cells stained for DAPI (nucleus, blue) and ARLNC1 (green) and AR transcripts, following treatment of blocking ASOs targeting the ARLNC1:AR 3' UTR interaction. U2-OS cells were transfected with ARLNC1 and AR expression vector prior to blocking ASO treatment. Scale bar, 10 μm. Quantifications of colocalization in U2-OS cells are depicted in (**E**) as a box plot. Center line and whiskers depict the median and range, respectively, and the box extends from the 25th to 75th percentiles (n = 60 cells for each sample aggregated from three independent experiments). P value was computed by two-tailed Student's t test.

(**F**-**G**) qPCR analysis of ARLNC1, AR and AR signaling genes (KLK2, KLK3, NKX3-1, TMPRSS2 and FKBP5) in MDA-PCa-2b cells (**F**) or LNCaP cells (**G**) transfected with control or blocking ASOs targeting the interaction site between ARLNC1 and the AR 3′ UTR. Mean ± s.e.m. values are shown, n = 3. Significance was determined by two-tailed Student's t test.

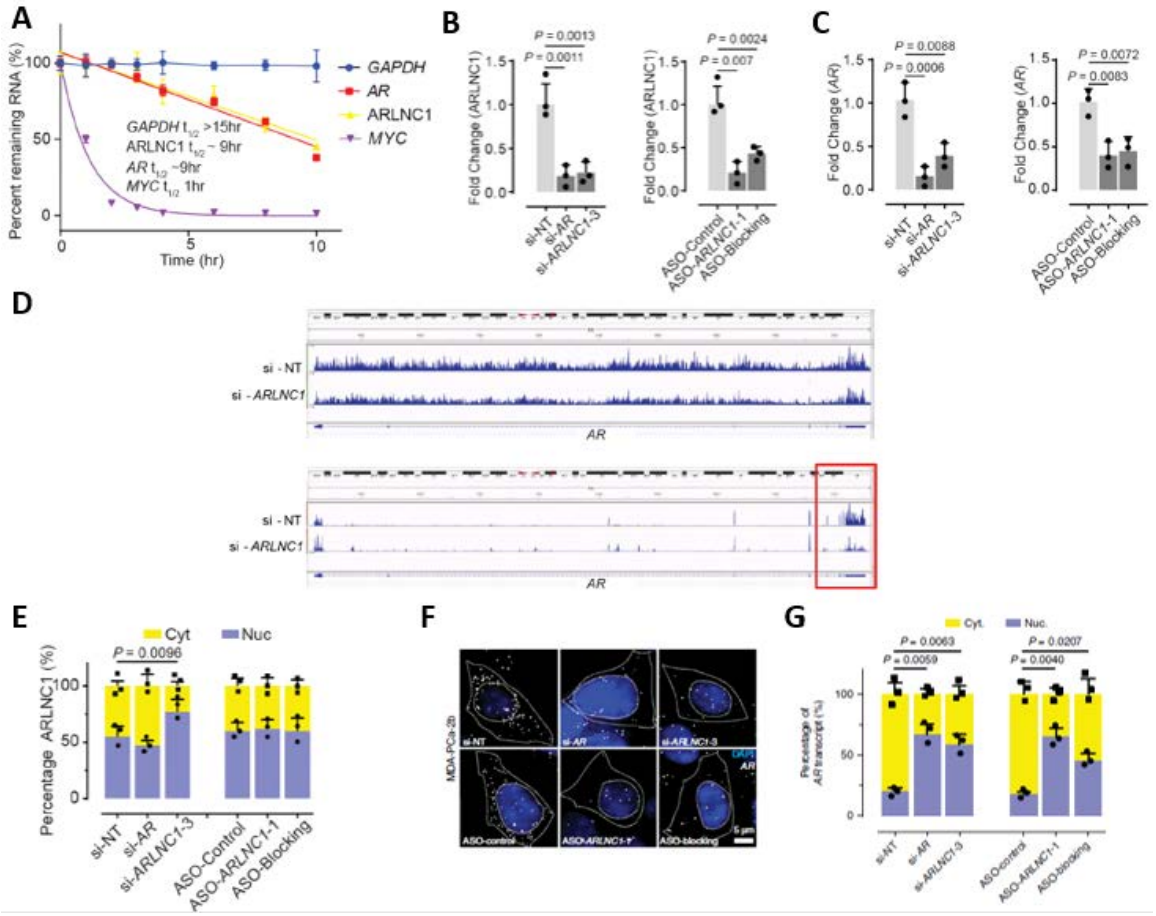


Figure 4.9 ARLNC1 regulates the cytoplasmic level of the AR transcript

(**A**) Half-life of GAPDH, AR, ARLNC1, and MYC RNA transcripts in LNCaP cells. Cells were incubated with actinomycin D for the indicated times, and target RNA quantities were evaluated by qPCR. RNA halflife was calculated by linear regression analysis. n = 3. At different time points, mean ± s.d. values are shown.

(**B**) Quantification of ARLNC1 levels, as measured by smFISH, after treatment of MDA-PCa-2b cells with siRNA against AR (si-AR), siRNA against ARLNC1 (si-ARLNC1-3), ASO against ARLNC1 (ASO-ARLNC1-1), or blocking ASO against the AR–ARLNC1 colocalizing segment (ASOblocking). Data were normalized to si-NT (Left) or ASO-Control (Right). Mean ± s.e.m. values are shown, n = 3 independent experiments and 60 cells analyzed for each sample. P values were computed by two-tailed Student's t test.

(**C**) Quantification of AR transcript levels, as measured by smFISH, after treatment of MDA-PCa-2b cells with siRNA against AR (si-AR), siRNA against ARLNC1 (si-ARLNC1-3), ASO against ARLNC1 (ASOARLNC1-1), or blocking ASO against the AR–ARLNC1 colocalizing segment (ASO-Blocking). Data were normalized to si-NT (Left) or ASO-Control (Right). Mean ± s.e.m. values are shown, n = 3 independent experiments and 60 cells analyzed for each sample. P values were computed by two-tailed Student's t test.

(**D**) BrU-seq alignment track (Top) and BrUChase-seq alignment track (Bottom) at the *AR* gene locus. Data acquired from MDA-PCa-2b cells, following treatment with control siRNA (si-NT) or si-ARLNC1.

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

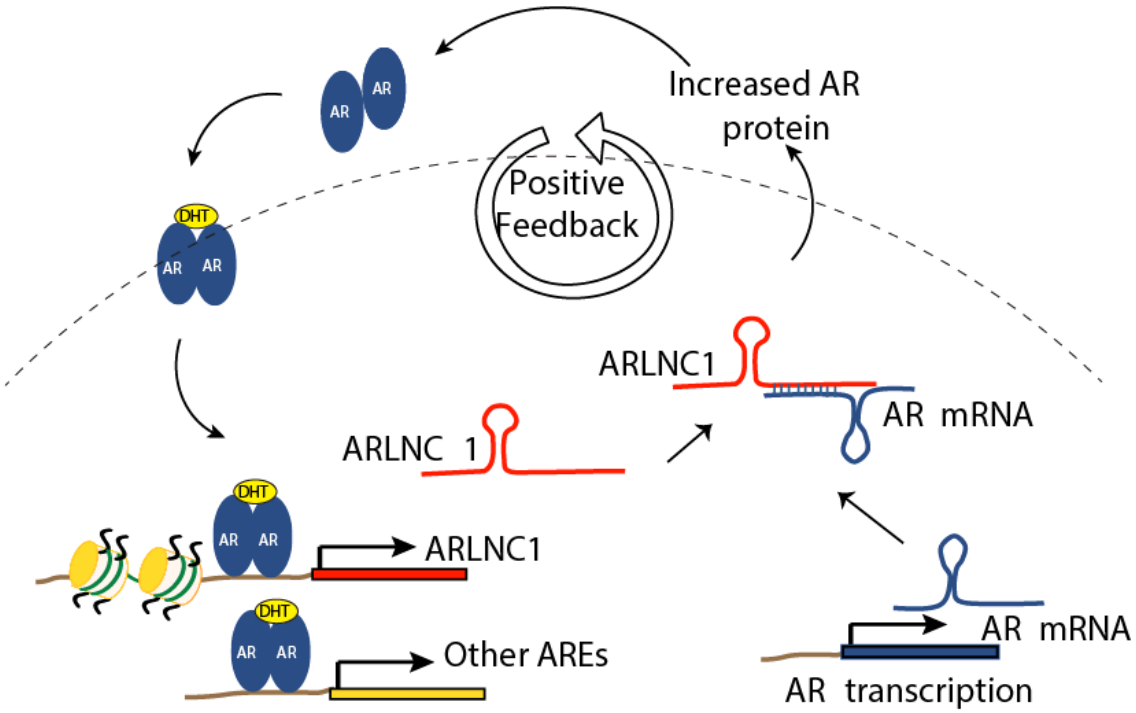


Figure 4.10 A model of ARLNC1 function in AR signaling

Tables

Table 4.1 Primers used in Chapter 4

Name	sequence
AR-3'UTR1-980-T7-F	GATCACTAATACGACTCACTATAGggagaggagaagcattggaaaccctatttcc
AR-3'UTR-1-980-R	GAACTCGAGGCCAAGTTTTGGCTGAAGAG
AR-3'UTR-1-1988-R	gagttcatgggtggcaaagtat
LacZ-T7-F	GATCACTAATACGACTCACTATAGggagaggagaGTCGTTTTACAACGTCGTGACTG
LacZ-R	GTACGGGGTATACATGTCTGACA
ARlnc1-S-T7-F	GATCACTAATACGACTCACTATAGggagaggagaAAAAGTACTGTAGCCTTCAGTAATTC
ARlnc1-S-R	TTGGCACAGAAGCAGTACACAC
ARlnc1-AS-F	AAAAGTACTGTAGCCTTCAGTAATTC
ARlnc1-S-T7-R	GATCACTAATACGACTCACTATAGggagaggagaTTGGCACAGAAGCAGTACACAC
ARlnc1-del700-1300-F	GAAATAGTGCCTTTGTGTATATTCTGCCTA
ARlnc1-del700-1300-R	TAGGCAGAATATACACAAAGGCACTATTTC
ARlnc1-Frag-700-R	CAAAGGCACTATTTCCAGAT
ARlnc1-Frag-1300-R	GAACACAGATTCACCTTTTC
ARlnc1-Frag-T7-1301-F	GATCACTAATACGACTCACTATAGggagaggagaTGTATATTCTGCCTAAGGCA
ARlnc1-Frag-T7-701-F	GATCACTAATACGACTCACTATAGggagaggagaCGGATATAAGTTAGGTAAAA

Table 4.2 smFISH probes used in Chapter 4

Name	smFISH probe sequence (5'-3')
ARLNC1 (31 probes)	ctagaaccttctattcattc
	gaacaacaacagaagccaga
	catggaaccagcacctgaaa
	tggagtttccttattttcgg
	gaacgagttccagtggacaa
	taaccttgggggccatgaag
	gtgctgaagttctagatgga
	agcagctatatgaattagga
	gtttagctctgcgggaagtg
	ggtctttatttcatatctgt
	caaccagagcaactgctgtc
	atgtcttacttccctcacag
	cttctaaagtggggctctga
	tcattatctagccagttcac
	cactgctatagcagcgttag
	aactgaaaacccatcacctc
	gggtattgatgtttgaggtc
	taaaggagggactctgttcc
	tgccatctcattctttctaa
	tcactttgcctgtttcattg
	tcaggaatggctgagtgatg
	ttctagttgcctgttttaca
	gtctccaaatacttcaagct
	caggaatgatgcagtcctta
	tacatacagtgtttcggagc
	aacttccttaattccctgag
	gcctttatgaacactgtgtt
	aacatctctgaccttgtctc
	catcctttattctccattat
	tgtttgtctactctggatgt
	gcacagaagcagtacacact
AR (62 probes)	aaagctcctcggtaggtctt
	cacgctctggaacagattct
	tgaaggttgctgttcctcat
	cagcagggacaacgtggatg

tgcttaagccggggaaagtg
aggatgtctttaaggtcagc
aagtgccccctaagtaattg
ttggcgttgtcagaaatggt
cgacactgccttacacaact
aaaagtggggcgtacatgca
caatggggcacaaggagtgg
agcagagaacctttgcattc
cagtatcttcagtgctcttg
cccttgaaaggggaatactc
ttctagccctttggtgtaac
tagacggcagttcaagtgtc
ccggacttgtagagagacag
gtagtcgcgactctggtacg
ccagagccagtggaaagttg
ttgatgcgagcgtggggatg
aagagagtgtgccaggatga
catacaactggccttcttcg
acacatcaggtgcggtgaag
gactgggatagggcactctg
ccatttcgcttttgacacaa
taaggtccggagtagctatc
caatgggcaaaacatggtcc
cttctggggtggaaagtaat
catctccacagatcaggcag
tccatagtgacacccagaag
tcagcggctcttttgaagaa
caagtttcttcagcttccgg
tcctcctgtagtttcagatt
tgtgacactgtcagcttctg
ctgacattcatagccttcaa
gcttccaggacattcagaaa
acacactacacctggctcaa
caaggctgcaaaggagtcgg
ccagttcattgaggctagag
cacttgaccacgtgtacaag
cacgtgtaagttgcggaagc
aggagtactgaatgacagcc
catggcaaacaccatgagcc

	cattggtgaaggatcgccag
	aagtagagcatcctggagtt
	attgaaaaccagatcagggg
	ctgtacatccgggacttgtg
	ttgagagaggtgcctcattc
	tgatttggagccatccaaac
	tcatgcacaggaattcctgg
	atgctgaagagtagcagtgc
	tttttgatttttcagcccat
	ccttgatgtagttcattcga
	gcatgcaatgatacgatcga
	agcaggatgtgggatttttt
	ttggtgagctggtagaagcg
	aagtgaactgatgcagctct
	ccatgtgtgacttgattagc
	atcatttccggaaagtccac
	cttgcacagagatgatctct
	tttcccagaaaggatcttgg
	ggtgtggaaatagatgggct
PCAT1 (29 probes)	gcctatgcagatatccaata
	aagggtacagatgctttctc
	gacctgtgggaattcataca
	agaggttcctttcttcatta
	taggtagctctttgtactca
	gctatgcatcttatatcctt
	tctttaattgctcaggttcc
	caagtgccagttaagtgtga
	ggccttattaagatgggatc
	cttatccattggtgtttctg
	cacttagaggcacatgggaa
	ttaaagtccagttaggttcc
	tttctcttctcacttctagt
	ttatcttgggaggttccaat
	catggtcttatgtatctgcg
	caaaggcgttggtgatgttg
	gtctttgtcgacttccaatg
	gcttcaatgattcctctcaa
	atggtcaacattgcgttctt

	cacccacttatcaagttttt
	tttggatttttgctgagctc
	ggttattgttgttgcgtaga
	cccaaatccacttttcattg
	attgctggttgccatatata
	gctcattgatttgttgagca
	aacttttgaagcgctgcttg
	aactatgtagcccaatttgt
	ttggaaagcgtgttctgcat
	gtgcgtcaggattcgacaaa
DANCR (30 probes)	cgggaagactctgggcaagg
	cgggcgcacaaaccagagag
	gcaactccagctgacaaaga
	cgaaacccgctacatagtgg
	aactcctggagctcaaggtc
	acttccgcagacgtaagaga
	tgcgctaagaactgaggcag
	ataccagcaacaggacattc
	gggatagttggcttaagtca
	ggcactttcctattgtaact
	cacgtggttgctacaagtta
	cagcattgtcactgctctag
	aacatgaagcacctgctaca
	acagcgtgaaacttgtagag
	gcttttgtaggttcatgact
	gctgagcatcttcaaagatt
	ggtcttggagaaatttcaga
	gcatgatcctgttttgttca
	tgcagcttgggtgtgtattc
	gccaaaccaaaatagggcta
	tgtggctgaagatctcatgg
	ggaagatttttatctcctgc
	acttgcagctgatgaaagct
	tcggttttctcaaatagcca
	gtcacccacagaatccaatt
	tgatgtgcaaagcggtcatc
	gaaggtaaggatgataccca
	tttgactggcacaaaaggtt

	ggtgatgacatatcaagagc
	agccaagacaagtggcaatt
EZH2 (34 probes)	ccaacaaactggtcccttct
	gtactctgattttacacgct
	aacctcttgagctgtctcag
	ctttacttcatcagctcgtc
	cttcgctgtttccattcttg
	aagtcactggtcaccgaaca
	atgggatgacttgtgttgga
	agcaactgcattcagagtct
	ataaaattctgctgtagggg
	ctgttcggtgagttctttat
	gtacattcaggaggaagtgc
	catcgcctacagaaaagcgt
	gcacttacgatgtaggaagc
	gttgggtgttgcatgaaaag
	ctgtgttcttccgcttataa
	aaggtttgttgtctagagct
	aatgctggtaacactgtggt
	gtgagagcagcagcaaactc
	tgctactgttattgggaagc
	ttccagcacattaatggtgg
	ttggtgtttgacaccgagaa
	ggactctaaacattgaggct
	taacctagcaatggcacaga
	ctgtctacatgttttggtcc
	ctggagctatgatgctagat
	ctgtatctttctgcagtgtg
	aaacatggttagaggagccg
	tggatgatcacagggttgat
	atcacacaagggcacgaact
	agatggtgccagcaatagat
	tgataaaaatcccccagcct
	tctccacagtattctgagat
	gatggctctcttggcaaaaa
	aaaaacagctcttcgccagt
FOXA1 (47 probes)	tgcccaatacaaccatccag

catcttcacagttcctaaca
tgtctgcgtagtagctgttc
aggtgttcatggagttcatg
tcgtagtcatggtgttcatg
gttggcataggacatgttga
cgcagtcatgctgttcatgg
tagctgcgcttgaacgtctt
gagatgtacgagtagggcgg
ctggatggccatggtgatga
atgatccactggtagatctc
tgccggtaataggggaagag
agccgttctcgaacatgttg
tgcttctcgcacttgaagcg
ctggagtcttcaactccgag
aacgggtggttgaaggagta
ggacatgaggttgttgatgg
ttgaagtccagcttatgctg
tattgcagtgcctgttcgta
caacgtagagccgtaaggcg
atacacaccttggtagtacg
aagtgtttaggacgggtctg
ttgcactgggggaaaggttg
aaattggtttggggttgtct
ggatcattaaacttcgcagg
gtaggggtcaggtaaggag
attgccacagacctgtaaac
cttttaagagcctctagtgt
agcaaatggctctgatgttt
gcatgtgcataattaagtcc
acggaggatgtctacacatc
gcaactcttgagaatgtatc
ttggggtccttgtaactttc
attcctgaggaattgattcc
agaagcagagttcttgaggg
atgacatgaccatggcactc
ctctcctccaacattgtaat
aaatccagctccctataact
ttgaatcttggaccacgttt
atggccactatcaataggat

	agcacacgatggcaatgatt
	tccaactgtggaaagtgcat
	gtctggctatactaacacca
	gaacattttccacggcttaa
	gtccttaactgcaaatgatc
	aaacacagaaggcttaagcc
	ttaattctatcagccacagc

References

1. Roche, P.J., S.A. Hoare, and M.G. Parker, *A consensus DNA-binding site for the androgen receptor.* Mol Endocrinol, 1992. **6**(12): p. 2229-35.
2. Carithers, L.J. and H.M. Moore, *The Genotype-Tissue Expression (GTEx) Project.* Biopreserv Biobank, 2015. **13**(5): p. 307-8.
3. Mele, M., et al., *Human genomics. The human transcriptome across tissues and individuals.* Science, 2015. **348**(6235): p. 660-5.
4. Engreitz, J.M., et al., *RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites.* Cell, 2014. **159**(1): p. 188-199.
5. Faghihi, M.A., et al., *Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase.* Nat Med, 2008. **14**(7): p. 723-30.
6. Gong, C. and L.E. Maquat, *lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements.* Nature, 2011. **470**(7333): p. 284-8.
7. Kretz, M., *TINCR, staufen1, and cellular differentiation.* RNA Biol, 2013. **10**(10): p. 1597-601.
8. Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR.* Nature, 2013. **493**(7431): p. 231-5.
9. Gawronski, A.R., et al., *MechRNA: prediction of lncRNA mechanisms from RNA-RNA and RNA-protein interactions.* Bioinformatics, 2018. **34**(18): p. 3101-3110.
10. Mann, M., P.R. Wright, and R. Backofen, *IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions.* Nucleic Acids Res, 2017. **45**(W1): p. W435-W439.
11. Lennox, K.A. and M.A. Behlke, *Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides.* Nucleic Acids Res, 2016. **44**(2): p. 863-77.
12. Asangani, I.A., et al., *Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer.* Nature, 2014. **510**(7504): p. 278-82.
13. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
14. Hansen, P., et al., *Saturation analysis of ChIP-seq data for reproducible identification of binding peaks.* Genome Res, 2015. **25**(9): p. 1391-400.
15. Kent, W.J., et al., *BigWig and BigBed: enabling browsing of large distributed datasets.* Bioinformatics, 2010. **26**(17): p. 2204-7.
16. Niknafs, Y.S., et al., *The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression.* Nat Commun, 2016. **7**: p. 12791.
17. Raj, A., et al., *Imaging individual mRNA molecules using multiple singly labeled probes.* Nat Methods, 2008. **5**(10): p. 877-9.
18. Rossiello, F., et al., *DNA damage response inhibition at dysfunctional telomeres by modulation of telomeric DNA damage response RNAs.* Nat Commun, 2017. **8**: p. 13980.

19. Paulsen, M.T., et al., *Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA.* Methods, 2014. **67**(1): p. 45-54.
20. Paulsen, M.T., et al., *Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response.* Proc Natl Acad Sci U S A, 2013. **110**(6): p. 2240-5.

Chapter 5

Circular RNAs as a novel class of non-coding transcripts with diagnostic potential in prostate cancer[1]

Abstract

Circular RNAs (circRNAs) are an intriguing class of non-coding RNA due to their covalently closed structure, higher stability compared to linear counterparts, and implicated roles in gene regulation. Recently, our lab used an exome capture RNA sequencing protocol to characterize circRNA landscape across >2,000 cancer samples. When compared against Ribo-Zero sequencing, capture sequencing significantly enhanced the enrichment of circRNAs and preserved accurate circular-to-linear ratios. Using capture sequencing, we built the most comprehensive catalog of circRNA species to date: MiOncoCirc, the first database to be composed primarily of circRNAs directly detected in tumor tissues. MiOncoCirc serves as a valuable guide for the development of circRNAs as diagnostic or therapeutic targets across cancer types.

[1] This chapter was previously published as part of the following manuscript: Vo, J.N., et al., The Landscape of Circular RNA in Cancer. *Cell*, 2019. 176(4): p. 869-881 e13.

In this chapter, we performed validation of circRNA species nominated from MiOncoCirc. We confirmed that circRNAs are resistant to treatment of RNaseR, a 3' to 5' exoribonuclease that digests all linear RNAs except lariat or circular RNA structures. Furthermore, we confirmed existence of a novel class of circular transcripts, termed read-through circRNAs (rt-circRNAs), that involve exons originating from different genes. In prostate cancer, we identified a group of circRNAs that are specifically enriched in Neuroendocrine Prostate Cancer (NEPC). In addition, we showed that circRNAs have potential to serve as non-invasive biomarkers. They are more stable than linear RNAs incubated in plasma, and some circRNAs were detectable in urine.

Introduction

Out lab has developed a poly(A)-independent, exome-capture based RNA-Seq protocol to profile circular RNAs[1]. This method is able to consistently detect more circular RNA species compared to the previously used Ribo-Zero sequencing protocol, while maintaining comparable circular-to-linear fractions to sequencing results before.

Using exome-capture RNA-Seq protocol, we have constructed a bioinformatics pipeline to generate MiOncoCirc compendium[1], consisting circRNA species detected from clinical cancer samples, cancer cell lines, and pooled normal tissues (**Figure 5.1**). Published circRNA bioinformatics tool, CIRCexplorer[2], as well as in-house computational pipeline CODAC[3], were adapted to discover backspliced circular RNA reads originated from one

gene (circRNA), or more than one genes (read-through circular RNAs, rt-circRNA, typically from adjacent genes).

Characterization of circular RNAs in MiOncoCirc revealed several features and properties associated with circRNA expression (**Figure 5.2**): (1) More than 99% of circRNAs harbor circular junction boundaries flanked by the canonical splicing motif AG-GT, supporting the theory that circRNAs are produced from back-splicing events. (2) CircRNAs generated from single gene could form several isoforms. The number of circRNA isoforms increases proportionally with number of exons per gene (**Figure 5.2A**). (3) Regarding abundance, baseline expression of the linear gene counterpart is not a reliable predictor of the corresponding levels of circRNAs originated from it (**Figure 5.2B**). Rather than driven solely by abundance of linear gene counterparts, circular RNAs expression are more likely to be a result from diverse regulatory splicing mechanisms. (4) There is a small subset of circRNAs that are consistently detectable in more than 90% of samples, with more than 5-fold higher expression levels than median abundance of all circRNAs. This is possibly due to their gnomic structures which support transcript back-splicing and circularization (**Figure 5.2C**). (5) The majority of circRNA species exist ubiquitously across different tissues, or with limited lineage-specificity (**Figure 5.2D**). A small set of circRNAs with high tissue-restricted expression (895 circRNAs from MiOncoCirc) could be explained by the tissue specificity of the parental genes that circRNAs originate from. This set of lncRNAs can be evaluated as biomarkers to discriminate different tissue types.

Results

Identification and validation of read-through circRNAs (rt-circRNA)

circRNAs produced from exons originating from different genes were previously reported to be products of gene fusions, in which each fusion partner donated their exons for backsplicing, and were named f-circRNAs[4]. We recently developed a novel annotation pipeline, CODAC, that could annotate backsplicing events involving two genes. Since pairs of homologous or paralogous genes can give rise to mapping ambiguities and false positives, we performed preliminary filtering, as well as indicated pairs with high degrees of similarity. Although we did not detect any f-circRNAs in MiOncoCirc resulting from chromosomal translocations and deletions, we discovered a novel class of circular transcripts that involved exons originating from two adjacent genes on the same strand: the rt-circRNA. Without the genomic information from matched whole-genome sequencing (WGS) and whole-exome sequencing acquired through our integrative clinical sequencing approach[5], rt-circRNAs would have appeared deceptively similar to linear transcripts resulting from tandem duplications in RNA-seq (**Figure 5.3A**). In general, rt-circRNAs comprised a small portion of all circRNAs in each sample (average 2.5%) and were detected at lower abundance (average 3.13 lower) than most other circRNAs from a single gene [1].

Some of these backspliced reads involving two genes were commonly found across different cancer types (**Figure 5.3B**) and were even detected in normal tissues or in samples with normal copy number (diploid) of the parent genes, which further suggested that they

were true common transcriptomic processes rather than rare genomic events. To further experimentally validate by quantitative reverse transcriptase PCR (qRT-PCR) that this class of read-through transcript was circularized, we searched for transcripts that were expressed in cell lines and selected a backspliced event spanning two adjacent genes, TTTY15 and USP9Y, which were less than 9 kb apart on chromosome Y and detected in several prostate cancer tissue samples (**Figure 5.3C**). The product of outward-facing primers involving exon 3 of TTTY15 and exon 3 of USP9Y were detected via qRT-PCR in LNCaP prostate cancer cells, as well as the product of inward-facing primers involving the same pair of exons. However, only the product of out-ward-facing amplification was resistant to RNase R degradation (**Figure 5.3C**), and the backspliced exon-exon junctions of USP9Y and TTTY15 were validated by Sanger sequencing (**Figure 5.3E**), confirming that the target was a circular molecule.

Finally, even though read-through circularization was largely widespread across cancer types (**Figure 5.3B**), we were able to nominate a small set of select rt-circRNAs that were tissue specific in the MiOncoCirc compendium (**Figure 5.3D**). Their tissue specificity could be explained by the tissue-specific expression. of the genes involved in the generation of the corresponding rt-circRNAs.

Differentially expressed circRNAs in Neuroendocrine Prostate Cancer

Because circRNAs are “spliced out” from mRNAs, any cellular process or transformation that has a profound impact on the transcriptome should likewise alter the circRNA

landscape of the cell. Thus, we characterized the circRNA landscape of prostate cancers undergoing neuroendocrine differentiation. Neuroendocrine prostate cancer (NEPC) is a rare, aggressive subtype of prostate cancer that can arise from post-hormonal therapy for PRAD[6] and has a poor prognosis[7]. In our cohort, pathologists diagnosed eight NEPC cases based on cell morphology.

We then performed differential circRNA expression analyses (**Figure 5.4A**) and uncovered 34 upregulated and 48 downregulated circRNAs with statistical significance ($p < 0.01$). In NEPC, the most significantly upregulated and downregulated circRNAs were circ-AURKA and circ-AMACR, respectively (**Figure 5.4B**). This finding is consistent with the change in parental gene expression of AURKA and AMACR. We carried out qRT-PCR in RNase R-treated NCI-H660, a neuroendocrine cell line, and confirmed that circ-AURKA was generated from exon 6 backspliced to exon 3 (**Figure 5.4C, E**). Finally, we confirmed that circ-AURKA was expressed more highly in the NEPC cell line NCI-H660 than in the non-NEPC prostate cell lines, LNCaP and VCaP, a result that was consistent with the expression of its parent gene (**Figure 5.4D**).

CircRNAs Are More Stable Than Cognate Linear Transcripts

Due to their lack of open ends, circRNAs are resistant to exoribonuclease (RNase R treatment) and are potentially more stable than their cognate linear transcripts, thus making them ideal candidates for biomarker development. To evaluate the stability of circRNAs identified in MiOncoCirc, total RNA was first isolated from LNCaP prostate cancer cells

and incubated with RNase R for 30 min. The ratio of circular-to-linear RNA species was then quantified by qRT-PCR. All circRNA species tested showed resistance to exoribonuclease and were thus significantly more stable than their linear counterparts (**Figure 5.5A**). In an orthogonal method to assess the stability of circRNAs, we compared concentrations of circular and linear transcripts in LNCaP cells that were treated with actinomycin D, a transcription inhibitor, over time. In LNCaP cells harvested at 0, 2, 4, 8, and 24 h after actinomycin D treatment, relative circRNA levels increased whereas relative mRNA levels decreased (**Figure 5.5B**), thus demonstrating the relatively higher stability of circular transcripts. Identification of biomarker species that are resistant to degradation is desirable for clinical settings. To determine whether circRNAs are more stable than their cognate linear RNAs in biospecimens, we analyzed the stability of circRNA in human blood plasma. VCaP RNAs were incubated in the plasma to simulate an environment of circulating RNAs. Indeed, as assessed by the ratio of circular-to-linear transcripts of select candidates, circRNAs were more stable than linear RNA in plasma after incubation (**Figure 5.5C**).

Detection of circRNAs in urine from prostate cancer patients

Because noninvasive methods of detection are more ideal for screening assays in the clinic, we assessed whether circRNAs could be reliably detected in urine samples. We generated three libraries with exome capture RNA-seq and detected 1,092 circRNAs in urine samples from prostate cancer patients that completely overlapped with circRNAs identified in PRAD tissue samples from the MiOncoCirc compendium (**Figure 5.6A**). Furthermore,

analysis of circRNAs by qRT-PCR showed that circRNA species could be detected in urine from prostate cancer patients (**Figure 5.6B**). Furthermore, these data demonstrate that, even with low starting amounts of RNA (50 ng), exome capture RNA-seq of urine samples is a promising assay for profiling circRNAs of prostate cancer patients in a noninvasive manner.

Discussion

Read-through chimeric transcripts are widespread phenomena and may represent a mechanism for the evolution of protein complexes[8], which may explain the detection of some rt-circRNA at high frequency in our consortium. Circularized read-through events may have several important implications. First, whether the process of circularization ("backsplicing") is a co-transcriptional or post-transcriptional process has remained an ongoing debate[9-14]. However, recent evidence provided by Liang et al. (2017) shows that depleting CPSF3, a 3' end processing endonuclease, could increase the intergenic read-through as well as circularization at one specific locus. Further, the formation of rt-circRNAs confirms that in some pairs of genes, circularization must occur prior to cleavage and polyadenylation. Our data, therefore, contribute to clarifying the timing of the circularization process and provides evidence for the co-transcriptional model. Second, the backspliced reads from those resulting from RNA circularization, genomic tandem duplications, or some structural rearrangements in RNA-seq appear identical (**Figure 5.3A**). Indeed, one of our most commonly detected backspliced events, from exon 3 of USP9Y to exon 3 of TTTY15, was previously proposed to be a "fusion" or "translocation"

in prostate cancer[15]. Our validation via RNase R treatment and Sanger sequencing (**Figure 5.3C, E**), however, proved that it is a rt-circRNA. The MiOncoCirc resource will thus serve as a highly valuable tool for cancer genomic researchers who wish to filter out rt-circRNA transcripts from a list of potential structural rearrangement candidates.

In conclusion, we have performed validation for selected circRNA species catalogued in MiOncoCirc. The database serves as an important resource for scientists who wish to explore the lineage-specific and expression patterns of circRNAs in cancer, as well as the intriguing mechanisms of read-through splicing. Such studies may shed light into the function of circRNAs and help develop the use of circRNAs in diagnostic medicine.

Materials and Methods

Cell lines

LNCaP, VCaP, 22Rv1 (male, prostate adenocarcinoma), and NCI-H660 (male, prostate epithelial neuroendocrine) cell lines were obtained from the American Type Culture Collection (ATCC). LNCaP and 22Rv1 cells were cultured in ATCC-formulated RPMI1640 medium supplemented with 10% fetal bovine serum (FBS; Invitrogen). VCaP cells were maintained in ATCC-formulated Dulbecco's Modified Eagle's Medium supplemented with 1% penicillin/streptomycin (Invitrogen) and 10% FBS (Invitrogen). NCIH660 cells were maintained in ATCC-formulated RPMI-1640 medium, supplemented with 0.005 mg/mL insulin, 0.01 mg/mL transferrin, 30 nM sodium selenite, 10 nM hydrocortisone, 10 nM beta-estradiol, 2 mM L-glutamine, and 5% FBS. All cell lines were

genotyped to confirm their identity at the University of Michigan Sequencing Core. We maintained cell lines at 37C in a 5% CO2 cell culture incubator and tested all cell lines routinely for Mycoplasma contamination.

RNase R treatment

Total RNA was isolated by TRIZOL lysis followed by purification using the miRNeasy Mini Kit (QIAGEN) with DNase digestion step. 2 mg of total RNA was either treated with 0 units (control) or 20 units of RNase R (Lucigen) in reaction buffer consisting 20 mM Tris-HCl (pH 8.0), 100 mM KCl, and 0.1 mM $MgCl_2$, respectively. Treatment was conducted at 37 °C for 1 hour, followed by RNase R inactivation at 65°C for 20 min. RNA was then extracted using miRNeasy Mini Kit (QIAGEN) and eluted in 15 ml of water. Reverse transcription was performed using SuperScript III Reverse Transcriptase (Invitrogen) and random primers (Invitrogen) following manufacturer's standard protocol.

RT-qPCR and validation of circRNA

To assess relative expression of circRNA candidates, quantitative Real-time PCR (qRT PCR) assays were performed using Power SYBR Green Master Mix (Applied Biosystems) and were carried out with the StepOne Real-Time PCR System (Applied Biosystems). Sequences of oligonucleotide primers were included in **Table 5.1**, with the following abbreviations used- li: linear RNAs; circ: circular RNAs; in: inward facing direction; out: outward facing direction; F: forward; R: reverse. Linear version of a housekeeping gene, GAPDH, were amplified as control. Expression of targets were calculated relative to the

housekeeping gene. Fold changes following RNase R treatment were calculated relative to the control untreated samples. The genomic sequence of qPCR products from the circRNA backspliced junction was further validated with Sanger Sequencing at the University of Michigan Sequencing Core.

Actinomycin D treatment

To validate the stability of RNAs, LNCaP cells were plated in 6-well plates and incubated for 12 hours. After incubation, cells were treated with 2.5 mg/mL of actinomycin D (Sigma) for 0-24 hours. Cells were harvested in Qiazol at 0, 2, 4, 8, and 24 hours post-treatment. RNA was isolated using the miRNeasy mini kit (QIAGEN). RNA was quantified, and 1 mg of RNA was used to make cDNA using SuperScript III First-Strand Synthesis System for RT-qPCR (Invitrogen) using random primers. We then performed RT-qPCR and analyzed data with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) used as a normalization control.

RNA stability in blood plasma

To check the stability of various linear and circular transcripts in plasma, we first isolated the blood plasma from fresh blood taken from healthy individual (male, age 30). In short, a total of 15 mL blood was collected in a vacutainer tube containing EDTA as the anticoagulant and mixed well before centrifugation at 2,000 rcf for 20 min at room temperature. The plasma layer was then carefully aspirated and stored at –80°C in cryovials. Next, we incubated 1 mg of VCaP RNA with 100 mL of plasma for 0, 15, 30,

45, 60, and 75 min. After incubation, total RNA was isolated and various linear and circular transcripts were quantified using qRT-PCR. The expression of transcripts at the zero-minute time point was considered as the control. Relative levels of circular and linear transcripts were calculated and shown.

Urine RNA extraction for RNA-seq and qRT-PCR

Post-digital rectal examination (Post-DRE) urine was collected from 13 prostate cancer patients presenting for diagnostic prostate biopsy using standardized protocols at University of Michigan Rogel Cancer Center. Urine was collected in an equal volume of RNA Protection Reagent and then frozen at 80°C until extraction of RNA was performed. Urine RNA was isolated by MagMAX mirVana Total RNA Isolation Kit (Invitrogen), which allows for recovery of total RNA (both intra- and extracellular) in urine. Capture sequencing was performed on three urine samples (**Figure 5.6A**), and qRT-PCR was performed on another 10 urine samples (**Figure 5.6B**).

Figures

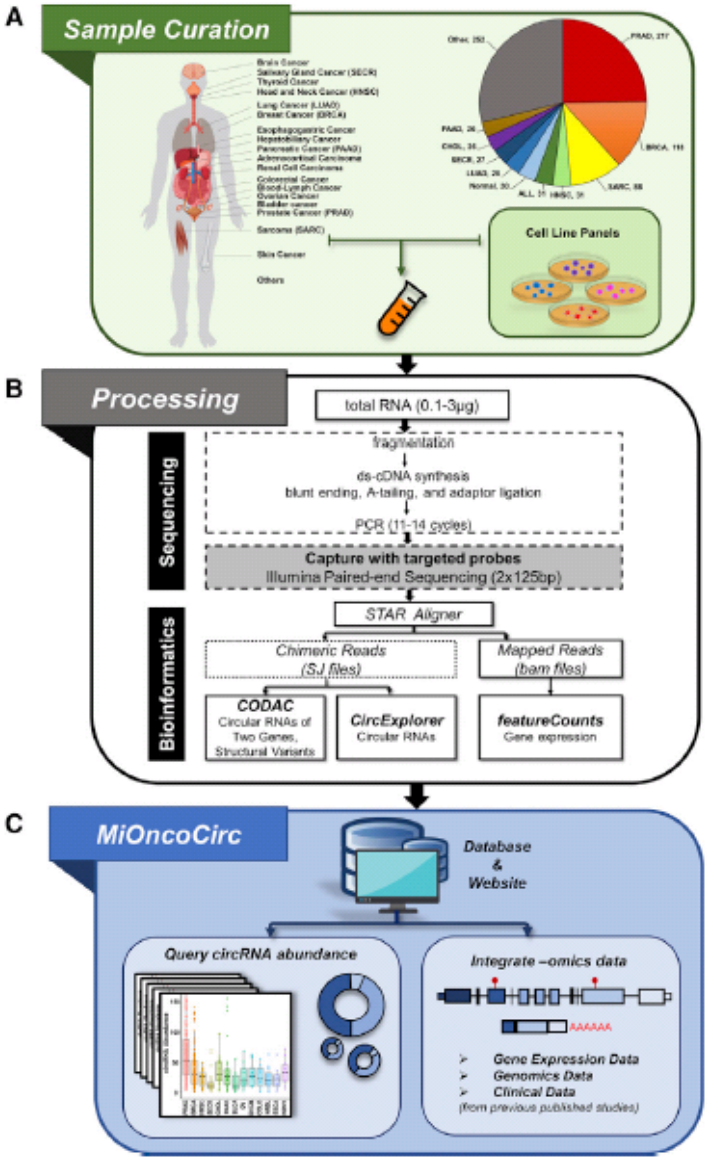


Figure 5.1 Overview of the MiOncoCirc Compendium
(**A**) Summary of high-depth, paired-end RNA-seq samples from previously published datasets, cell line panels, and normal tissues.
(**B**) Exome capture RNA-seq protocol and the bioinformatics pipeline for creation of MiOncoCirc. The unmapped reads from chimeric aligner (STAR) were annotated against the exon junctions. CIRCexplorer was used to call circRNA transcripts, and CODAC was used to annotate circRNAs involving two genes. FeatureCounts was used to quantify gene expression.
(**C**) MiOncoCirc, an online database that enables querying and downloading of circRNAs abundance across different tissues.

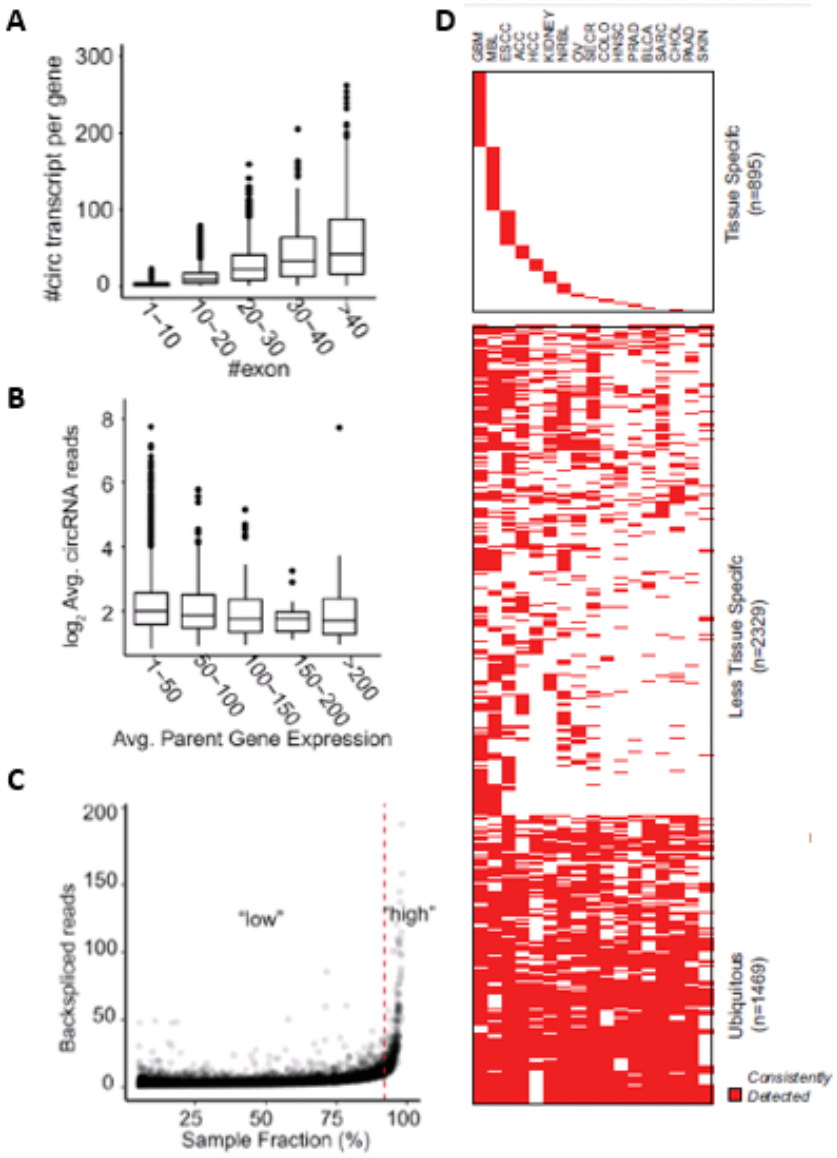


Figure 5.2 Properties and prevalence of circRNAs in MiOncoCirc

(**A**) Genes can form multiple circRNA transcripts. The number of circular transcripts increased proportionally with the number of exons per gene.

(**B**) Average expression of circRNA abundance (in normalized backspliced reads) versus average expression of parental expression (in FPKM). Parent gene expression was grouped into bins of 50. Overall, there was no difference in the mean of the bins (ANOVA p = 0.12), indicating that the correlation was weak (Spearman's r = 0.12).

(**C**) Circular RNA abundance (in normalized backspliced reads) versus sample fraction (%). There was a small portion of circRNAs (<2% of all circRNAs, generated from 589 genes, marked as "high") that were detected in more than 90% of all samples. They also had higher expression compared to the median of all circRNAs.

(**D**) Tissue-specific heatmap of genes that can generate circRNAs, as demonstrated in 17 cancer cohorts from the MiOncoCirc compendium. A gene was considered to be consistently detected if it generated at least one high-confidence circRNA in more than 30% of samples of any given lineage.

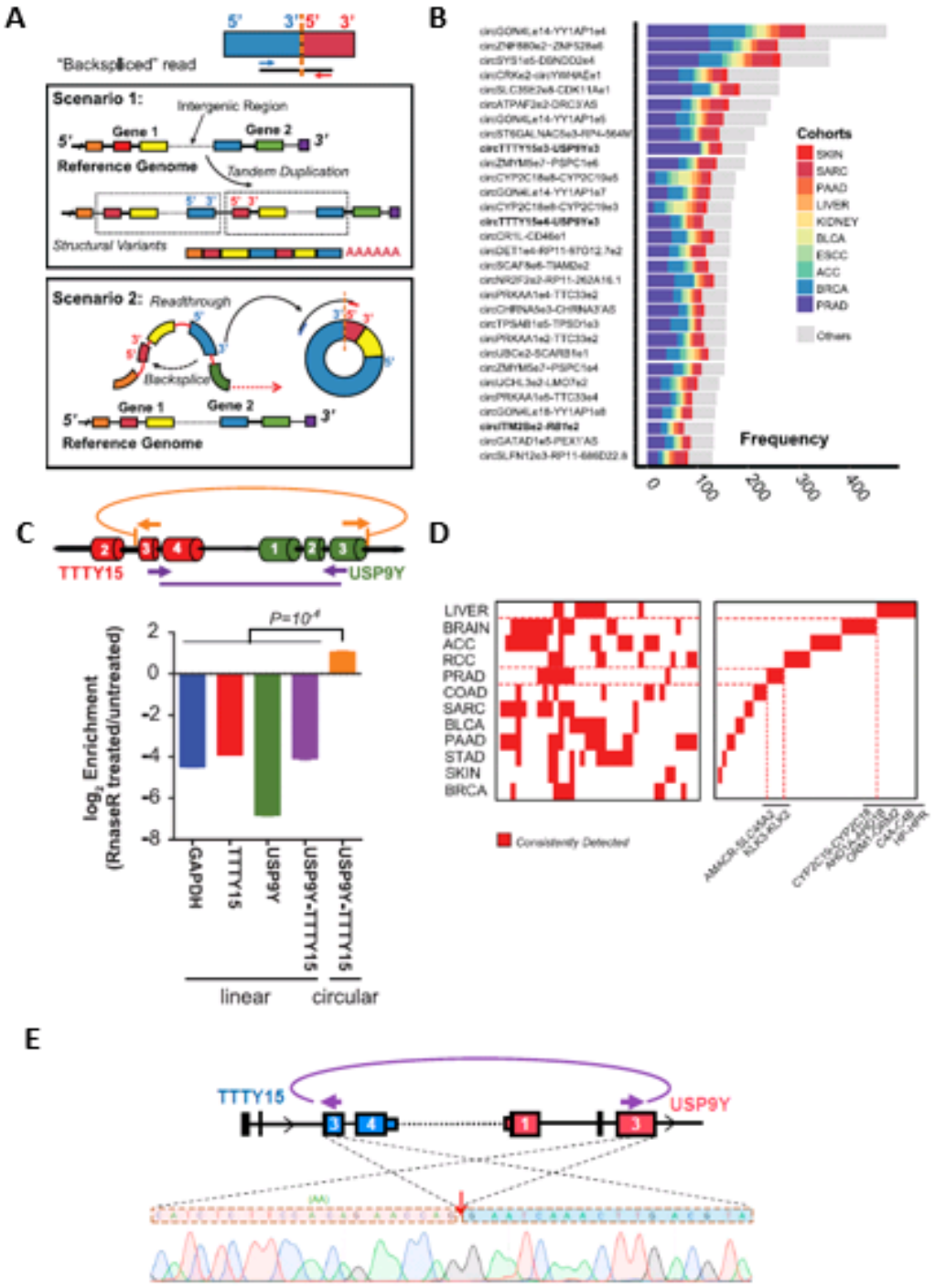


Figure 5.3 Existence of novel circularized, read-through transcripts involving two genes
(**A**) Schematic showing that genomic tandem duplications and circRNAs involving two genes can appear similar in paired-end RNA-seq. Specifically, when mates of a paired-end read were aligned in divergent orientation to exons of two adjacent genes, the result could be interpreted as either a

duplication of a group of exons from two genes (Scenario 1) or a circularization from the downstream gene back to the upstream gene (Scenario 2).
(**B**) The frequency and distribution of the top 30 most abundant back-spliced events involved neighboring genes in miOncoCirc compendium.
(**C**) The circular read-through event involving exon 3 of TTTY15 and exon 3 of USP9Y was chosen for validation in LNCaP cells. Post-RNase R treatment, only the qRT-PCR product of outward-facing primers was resistant to exoribonuclease degradation.
(**D**) The tissue specificity heatmap of rt-circRNAs. The rt-circRNAs specific to liver and prostate cancer are labeled.
(**E**) Schematic depicting the circular read-through event that generated circTTTY15e3-USP9Ye3 transcript detected in the MiOncoCirc compendium. The Sanger sequencing result of circTTTY15-USP9Y RT-qPCR product showed the correct sequence spanning the backsplice exon-exon junctions of USP9Y and TTTY15.

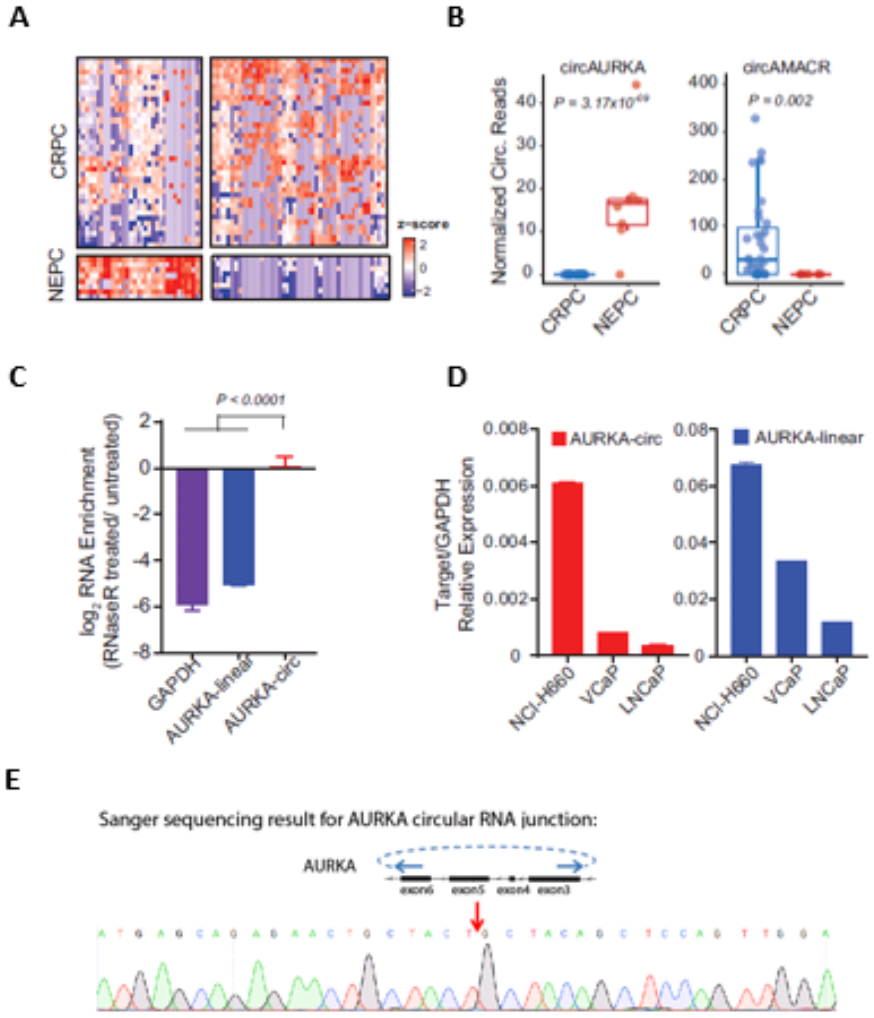


Figure 5.4 Circular RNAs enriched in Neuroendocrine Prostate Cancer

(**A**) The heatmap of 34 upregulated and 48 downregulated circRNAs with statistical significance ($p < 0.01$) in NEPC compared to CRPC cases.

(**B**) Comparing NEPC to CRPC, the most significantly upregulated circRNA was circ-AURKA (Mann–Whitney U test $p < 0.001$); the most significantly downregulated circRNA was circ-AMACR (Mann–Whitney U test $p = 0.002$).

(**C**) qRT-PCR of outward-facing primers of AURKA (backspliced from exon 6 to exon 3) in RNase R-treated NCI-H660, a NEPC cell line, confirmed the circular structure of this molecule. $p < 0.0001$ calculated from one-way ANOVA. Error bars show mean ± SD for triplicates.

(**D**) qRT-PCR of circular and linear AURKA in prostate cancer cell lines. Both circ-AURKA and linear-AURKA were expressed higher in NCI-H660 than in two non-NEPC cell lines, LNCaP and VCaP. Error bars show mean ± SD for triplicates.

(**E**) Schematic depicting the circular RNA generated from AURKA transcript detected in the MiOncoCirc compendium. Bottom: The Sanger sequencing result of circAURKA RT-qPCR product showed the correct sequence spanning the backsplice exon-exon junctions within AURKA.

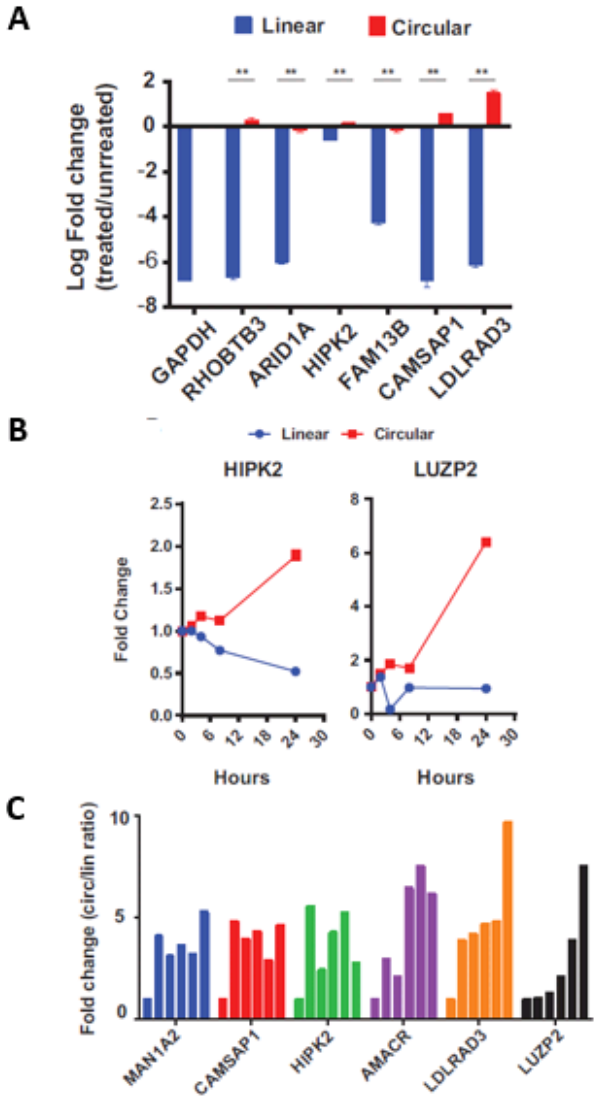


Figure 5.5 Stability of circular RNAs compared to linear counterparts

(**A**) Compared to their linear counterparts, circRNAs were resistant to RNase R degradation. Linear transcripts were detected by inward-facing qRT-PCR primers, while circular transcripts were detected by outward-facing qRT-PCR primers (**p < 0.0001, calculated from Student's t test). Error bars show mean ± SD for triplicates**.**

(**B**) After transcription inhibition by actinomycin D in LNCaP cells, linear transcripts (Linear) degraded faster than their corresponding circular transcripts (Circular). Samples were harvested at 0, 2, 4, 8, and 24 h post-treatment. GAPDH was used as the control. The fold changes were calculated relative to the starting time point. circHIPK2 was selected to represent ''high'' class circRNAs. circLUZP2 represented "low" class circRNAs but with elevated expression in prostate cancer compared to normal.

(**C**) After incubating VCaP RNAs in plasma, the circular-to-linear ratio of circRNAs increased over time. Samples were harvested at 0, 15, 30, 45, 60, and 75 min.

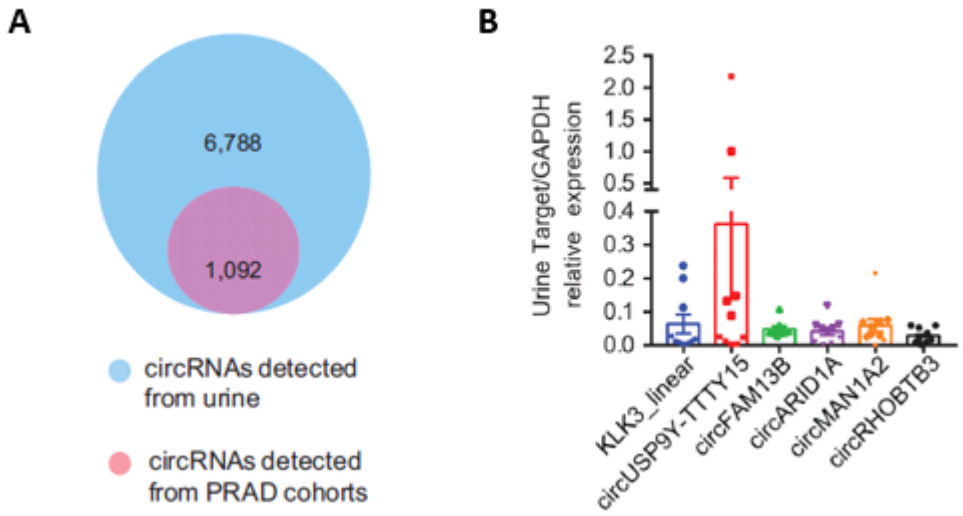


Figure 5.6 Detectable circRNAs in urine samples
(**A**) circRNAs were detected by exome capture RNA-seq of three urine samples from prostate cancer patients. These circRNAs greatly overlapped with circRNAs identified in prostate cancer tissues from the MiOncoCirc cohorts.
(**B**) The relative expression of circRNAs detected in urine from 10 prostate cancer patients. Linear GAPDH expression was used for normalization, and KLK3 was included as a positive control. Error bars show mean ± SD.

Tables

Table 5.1 Primers used in Chapter 5

Target Name	Primer Sequence
USP9Y-TTTY15-in-F	CGTATGGAGCCAAGAAAGCC
USP9Y-TTTY15-in-R	CTGTCGTTCCCTCCTACTGG
USP9Y-TTTY15-circ-out-F	CAGTAGGAGGGAACGACAGC
USP9Y-TTTY15-circ-out-R	CCAGTTTTTCCAAGGGCTTT
USP9Y-li-F	AGCCAGATGACCAGGATGC
USP9Y-li-R	TTCTTGTGTTCGTTGGCCTG
TTTY15-li-F	GCGCCATATTGAAGAGGACG
TTTY15-li-R	AGACACTCCTATCTCCCGGT
GAPDH-li-F	CCATCACCATCTTCCAGGAGCGA
GAPDH-li-R	GGTGGTGAAGACGCCAGTGGA
RHOBTB3-circ-F	TCAGTGGGAAGAATTGGAAGA
RHOBTB3-circ-R	ACACACTGGCAGCAGAACAG
RHOBTB3-li-F	TGAACTCCACAGCCTTGATG
RHOBTB3-li-R	TCCATGAAAGGAGTTGCTCA
ARID1A-circ-F	CCCCTCAATGACCTCCAGTA
ARID1A-circ-R	GTCCCTGTTGCTGCGAGTAT
ARID1A-li-F	CATGGCAGAAGGAGGAGACT
ARID1A-li-R	GAGTCATGGAATTCCGCTTC
HIPK2-circ-F	TTATCCACGCTGACCTCAAA
HIPK2-circ-R	TGAGGGGAGAAAACTTGCAC
HIPK2-li-F	GACGAGGAGGAGGAACAGAA
HIPK2-li-R	GAGTCGGAGTAGGGGGAGTC
FAM13B-circ-F	GACTGTTCAAAACCTGTGGCTA
FAM13B-circ-R	AGAGGCTGGTAGGATGCTGA
FAM13B-li-F	GCAAATTCTTTGGCTGCTGT
FAM13B-li-R	TTTCCAGAAGTCCAGCCATT
CAMSAP1-circ-F	ACGTTCAGTGCCTCGAAAGA
CAMSAP1-circ-R	GCTGGACAGGAGAAGCTTGA
CAMSAP1-li-F	ATCTTCCCTGGCCTCAGTG
CAMSAP1-li-R	CTTGTGGGGTTCGTTCACTT
LDLRAD3-circ-F	CAGAATGCGTCGGAAGTAGG
LDLRAD3-circ-R	CACCACCAGCACAAAAATGA
LDLRAD3-li-F	GCAGAATGCGTCGGAAGTAG
LDLRAD3-li-R	TCGGCTTGGTTCAGAGATTC
MAN1A2-circ-F	GGAGGCCTACTTGCAGCATA

MAN1A2-circ-R	TGCTCGAATTTCCTCTCTTGA
MAN1A2-li-F	GCATTCTGGCTGAATTTGGT
MAN1A2-li-R	CCATTTGGACGATCCATTTT
AURKA-circ-F	TCCATGATGCTACCAGAGTCT
AURKA-circ-R	GGAATGCGCTGGGAAGAATT
AURKA-li-F	GCAACCAGTGTACCTCATCCTG
AURKA-li-R	AAGTCTTCCAAAGCCCACTGCC

References

1. Vo, J.N., et al., *The Landscape of Circular RNA in Cancer.* Cell, 2019. **176**(4): p. 869-881 e13.
2. Zhang, X.O., et al., *Complementary sequence-mediated exon circularization.* Cell, 2014. **159**(1): p. 134-147.
3. Liang, D., et al., *The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting.* Mol Cell, 2017. **68**(5): p. 940-954 e3.
4. Guarnerio, J., et al., *Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations.* Cell, 2016. **166**(4): p. 1055-1056.
5. Robinson, D.R., et al., *Integrative clinical genomics of metastatic cancer.* Nature, 2017. **548**(7667): p. 297-303.
6. Beltran, H., et al., *Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets.* Cancer Discov, 2011. **1**(6): p. 487-95.
7. Conteduca, V., et al., *Neuroendocrine differentiation in prostate cancer: current and emerging therapy strategies.* Crit Rev Oncol Hematol, 2014. **92**(1): p. 11-24.
8. Akiva, P., et al., *Transcription-mediated gene fusion in the human genome.* Genome Res, 2006. **16**(1): p. 30-6.
9. Ashwal-Fluss, R., et al., *circRNA biogenesis competes with pre-mRNA splicing.* Mol Cell, 2014. **56**(1): p. 55-66.
10. Ebbesen, K.K., T.B. Hansen, and J. Kjems, *Insights into circular RNA biology.* RNA Biol, 2017. **14**(8): p. 1035-1045.
11. Kramer, M.C., et al., *Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins.* Genes Dev, 2015. **29**(20): p. 2168-82.
12. Liang, D. and J.E. Wilusz, *Short intronic repeat sequences facilitate circular RNA production.* Genes Dev, 2014. **28**(20): p. 2233-47.
13. Zhang, Y., et al., *The Biogenesis of Nascent Circular RNAs.* Cell Rep, 2016. **15**(3): p. 611-624.
14. Zhang, Y., L. Yang, and L.L. Chen, *Characterization of Circular RNAs.* Methods Mol Biol, 2016. **1402**: p. 215-227.
15. Ren, S., et al., *RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings.* Cell Res, 2012. **22**(5): p. 806-21.

Chapter 6

Concluding remarks and future directions for investigating non-coding RNAs in cancer

Summary of this study

For many years, researchers have been searching for oncogenes and tumor suppressors that contribute to cancer development. Most of these are protein-coding genes, which only consist of ~2% of human genome. During the last decade, advances in next-generation sequencing enabled the discovery of non-coding transcripts, a new category of molecules that function in a spatiotemporal manner in normal tissue development and oncogenesis process. Through interactions with chromatin, RNA, and protein, long non-coding RNAs mediate diverse cellular functions, including epigenetic regulation, RNA transcription, post-transcriptional regulation, and translation control. Under the context of cancer, the lncRNA expression landscape has been comprehensively profiled by several studies, but their clinical utilities, functioning mechanisms, and therapeutic potential remain largely unknown.

In prostate cancer, malfunction of Androgen Receptor (AR) signaling has been shown to drive prostate cancer progression. However, nearly all of the genes identified in this

pathway are protein-coding targets. To systematically identify lncRNAs regulated by AR signaling, we analyzed RNA-seq data from AR-positive prostate cancer cell lines before and after DHT stimulation. To prioritize lncRNAs that contribute to prostate cancer progression, we intersected the lncRNAs identified above with a list of lncRNAs that exhibit differential abundance between benign prostate and primary/metastatic prostate cancer tissues. Through this analysis, we discovered a lncRNA, ARLNC1, that is strongly induced by AR and has elevated expression in localized and metastatic prostate cancer.

We further characterized the transcript structure, cellular localization, and expression pattern of this lncRNA. We found that ARLNC1 expression is largely confined to the prostate lineage, and this pattern is dictated by the direct regulation of AR and cofactors (FOXA1) at ARLNC1 promoter region. Loss-of-function studies using RNAi-, ASO-, and CRISPR-based technologies demonstrated that ARLNC1 silencing inhibits cell proliferation and induces apoptosis *in vitro* and *in vivo*. Interestingly, we also observed attenuated AR signaling following ARLNC1 loss. We further dissected the cellular mechanism of ARLNC1-mediated modulation of AR signaling and found that it regulates AR mRNA post-transcriptionally via specific RNA-RNA associations in cells. These results echo previous observations, where lncRNAs serve as pathway regulators in a feedforward loop (**Figure 6.1**). These results have shed light on ARLNC1's mechanism of action and contributed to the field of lncRNA biology.

This study also suggested the clinical relevance of ARLNC1 expression: (1) ARLNC1 has potential to serve as a diagnostic biomarker. It could be readily detected using RNA *in-situ* hybridization (RNA-ISH) on formalin-fixed paraffin-embedded (FFPE) tissue samples, with expression significantly higher in localized and metastatic prostate cancer tissues compared to benign prostate. (2) Orthogonal targeting of the lncRNA specifically modulates AR and its aberrant signaling in tumors. This could be achieved using antisense technologies in several models *in vitro* and *in vivo*.

Finally, we investigated another class of non-coding transcript, circular RNAs. We confirmed the existence of a novel class of circRNA, read-though circRNAs. Stability of circRNAs is higher compared to their linear counterparts, and circRNAs with high abundancy can be detected in urine. The existence of circRNA in cell-free specimens, plus their increased stability compared to linear RNAs, make circular RNAs another attractive pool for biomarker exploitation.

Unexplored areas of this study

We have shown that AR transcriptionally regulates ARLNC1 and that ARLNC1 regulates its master regulator (AR) post-transcriptionally. The co-localization between ARLNC1-AR was supported by multilayered *in vitro* and *in situ* evidences from RNA-RNA interaction assays and RNA *in situ* hybridization. However, it remains a question whether this interaction is facilitated by protein binding partners, and how exactly does ARLNC1 affect AR mRNA stability. Future studies on the ARLNC1 interactome and ARLNC1-

bound miRNAs may shed light on these questions. Furthermore, the clinical relevance of ARLNC1 as a diagnostic marker can be further evaluated on a large cohort of FFPE or urine samples. Perhaps ARLNC1 can join a panel of established biomarkers to provide improved sensitivity and specificity for prostate cancer detection.

ARLNC1 interactome

Since lncRNAs rarely function alone with no protein co-factors, it is very likely that RNA-binding proteins are involved in the mechanistic function of a lncRNA. We have performed preliminary experiments to dissect the ARLNC1 interactome. In order to identify the protein binding-partners of ARLNC1, we carried out a RiboTrap assay followed by mass spectrometry analysis (**Figure 6.2A**). We used synthesized BrU-labeled lncRNA to incubate with cell lysate from the nucleus or cytoplasm of MDA-PCa-2B cells. In both cell nucleus and cytoplasm, we identified two proteins that strongly interact with ARLNC1: HuR(ELAV1) and TIAR. Both HuR and TIAR are known to bind to AU-rich elements and regulate the stability of messenger RNAs, including AR mRNA[1-5]. It is thus worth investigating whether the lncRNA-protein binding indeed exists in the cell, and whether these proteins play a role in ARLNC1-mediated functions, especially in the context of AR signaling.

By using RNA-immunoprecipitation (antibodies against HuR and TIAR) followed by qPCR analysis, we validated ARLNC1-binding with protein partners HuR and TIAR from cells (**Figure 6.2B**). Interestingly, HuR (not TIAR) also binds strongly with AR mRNA,

which is in accord with published results[5]. In order to investigate the functional significance of the interaction between ARLNC1 and HuR, we performed gene expression profiling following siRNA-mediated knockdown of ARLNC1 or HuR. We observed a significant overlap of genes regulated in each setting (**Figure 6.2C**), indicating that these two molecules might be involved in common RNA regulating processes. Furthermore, using the AR signature previously defined in MDA-PCa-2B cells, gene set enrichment analysis detected enrichment of AR-regulated gene set following HuR depletion (**Figure 6.2D**). This result confirms the regulatory effect by HuR on AR signaling, while suggesting a phenotypically-relevant link between HuR and ARLNC1. Additionally, HuR depletion resulted in a decreased level of AR protein (**Figure 6.2E**), an effect resembling the observation following ARLNC1 knockdown.

Now that we have established the direct binding between HuR and ARLNC1, and the phenotypic effect of HuR on AR signaling, we can test the hypothesis whether ARLNC1 sustains AR mRNA level by recruiting HuR protein to AR 3'UTR. Moreover, given the significant overlap between the HuR- and AR-regulated transcriptomes, it is worth investigating whether ARLNC1 facilitates HuR-mediated RNA stabilization of a group of mRNA targets.

Involvement of miRNA in ARLNC1 functions

Our current model suggests that mechanistic ARLNC1 function is at least partially driven by lncRNA-mRNA interaction via the sequence information encoded in the RNA

molecules. This base pair matching resembles the well-established miRNA-RNA binding in principle. MicroRNAs have been studied as key modulators of mRNA stability, while many lncRNAs have proposed roles of miRNA sponging/decoy. In some cases, lncRNAs act as microRNA decoys, causing sequestration of microRNAs favoring expression of repressed target mRNAs[6]. For instance, lnc-MD1 “sponges” both miR-133 and miR-135 to stabilize MAML1 and MEF2C. Both of them are transcription factors that control muscle differentiation by activating muscle-specific gene expression[7]. lncRNA H19 “sponges” the let-7 family of microRNAs via canonical and non-canonical binding sites, thus playing a key role in cell differentiation and cancer metabolism[8]. lncRNA-ROR “sponges” miR-145 to stabilize ZEB2, thus promoting invasion and metastasis in hepatocellular carcinoma[9]. In other cases, lncRNAs stabilize mRNA by competing with miRNAs for shared target mRNAs. For example, lncRNA BACE1-AS competes with miR-485-5p for binding to BACE1 mRNAs, leading to elevated expression level of BACE1, a critical enzyme in Alzheimer’s disease pathophysiology[10].

As for AR mRNA, multiple miRNAs have been reported to bind with its 3’UTR region and modulate transcript abundance[11]. Based on our results so far, one hypothesis could be that ARLNC1 competes with miRNAs for binding with the 3’UTR region in AR mRNA. This hypothesis could be first tested *in silico*, by predicting binding sites involved in miRNA-lncRNA-mRNA interplay. The specific binding domains on AR 3’UTR and ARLNC1 could then by validated by binding assays, and the effect of miRNAs could be read out through reporter genes assays.

Clinical Significance of ARLNC1

Through RNA-seq analysis, we observed elevated ARLNC1 expression in prostate cancers compared to benign tissues, thus indicating that it may serve as a useful diagnostic marker. As expected, the ROC analysis to differentiate cancers from normal showed an AUC of 0.819, which was only slightly lower than PCA3, a well-known diagnostic marker of prostate cancer that currently being clinically utilized (**Figure 6.3**). In addition, we confirmed the differential expression of ARLNC1 using *in situ* hybridization (ISH)-based screening of prostate tissues. As expected, expression of ARLNC1 was high in both localized and metastatic prostate cancer cells, and low in normal prostate cells. Furthermore, we discovered an association of ARLNC1 levels with accentuated AR signaling and luminal epithelial differentiation in patient tumors, both of which are important clinical considerations for anti-androgen treatment.

Our results call for a more thorough assessment of clinical utility of ARLNC1 as a prognostic or diagnostic biomarker. Comprehensive analyses to interrogate ARLNC1 expression with prognosis and treatment resistance could be conducted using Affemetrix Human Exon ST 1.0 microarray data generated from clinical cohorts after radical prostatectomy. The ARLNC1 RNA-ISH assay developed here could be applied to larger cohorts of formalin-fixed paraffin-embedded tissues represented on tissue microarrays [12] to determine the correlation between ARLNC1 abundance with the progression of prostate cancer stages. In addition, expression of ARLNC1 can be further assessed in whole urine

samples collected after digital rectal examination and analyzed with clinical-grade transcription-mediated amplification assays[13, 14]. Association between ARLNC1 levels and clinicopathological variables can be evaluated to determine whether ARLNC1 levels serve as indicators for clinical aggressiveness or treatment responses. If proved useful, ARLNC1 could be further tested in a multi-gene biomarker panel that serves as a part of clinical decision-making algorithms for routine patient care.

Future directions of non-coding RNA research in cancer

Correct annotations of non-coding RNAs

While we appreciate the contribution of next-generation sequencing in nominating genomic alterations and transcriptome aberrations, intrinsic limitations in library preparation and defects in sequencing analysis pipelines have resulted in incorrect annotations of transcriptomic structures. In our recent work to profile circRNAs, we have found that some back-spliced read-through circRNAs were previously annotated as gene-rearrangements. For instance, the circRNA ESR1-CCDC170 was reported as gene rearrangement in an aggressive subset of estrogen-receptor positive breast cancers[15]. It is thus of pivotal importance to revisit gene alteration datasets and update the annotations with our latest understanding of RNA species.

Differentiation of functional long non-coding RNAs from transcription by-product

Although lncRNA species have been comprehensively profiled in normal lineages of tissues, as well as in cancer tissues, only a handful of lncRNAs have been characterized

with clear functions to play oncogenic roles in cancer. The huge knowledge gap calls for scalable functional investigation methods towards identifying lncRNAs with *bona fide* roles contributing to cancer progression.

CRISPR screens provide novel opportunities to address this challenge. The CRISPR-Cas9 system for genome editing is a powerful tool for functional screens *in vitro* and *in vivo*. Co-expressing Cas9 endonuclease and a single short guide RNA molecule (gRNA) is sufficient to generate double-stranded DNA breaks in eukaryotic cells. These double-stranded DNA breaks are primarily repaired through the error-prone non-homologous end-joining pathway, often generating small indels at the target site and introducing frame-shift alterations. Thus, the CRISPR-Cas9 system provides a simple way of disrupting the open reading frame of protein-coding genes to produce loss of function alleles. However, when it comes to the study of lncRNAs, these transcripts are generally insensitive to reading frame alterations. In order to knockout lncRNAs reliably, we need to generate deletions of larger fragments.

Three types of CRISPR systems have been developed to fulfill this need. (a) CRISPR-Cas9 system featured by paired guide RNAs [16]; (b) CRIPSRi-system that represses the transcription of any gene via the targeted recruitment of the nuclease-dead dCas9-KRAB repressor fusion protein to the transcriptional start site (TSS) by a guide RNA (sgRNA)[17-20]; (c) CRISPR-Cas13d-system that directly targets mRNAs [21, 22]. The CRISPR-Cas13d system possesses unique advantages over the first two systems, in that it acts

directly on RNA molecules without affecting chromatin structures and local DNA at the genomic region harboring lncRNA. Meanwhile, CasRx-mediated knockdown exhibits high efficiency and specificity compared to RNA interference across diverse endogenous transcripts [21, 22].

These CRIPSR systems can be applied to the study of lncRNAs that are either highly conserved between species, or specifically expressed in different lineages of cancer. Proof-of-concept screens can be carried out in cancer cell line models. By recording the relative abundance of guide RNAs at different time points, we will be able to identify oncogenic lncRNAs that accelerate cancer cell growth and tumor-suppressive lncRNAs that delay cancer cell growth. Cellular location, regulated pathways, effected cell functions, as well as acting mechanisms of the identified functional lncRNAs could be further studied using loss-of-function and gain-of-function assays.

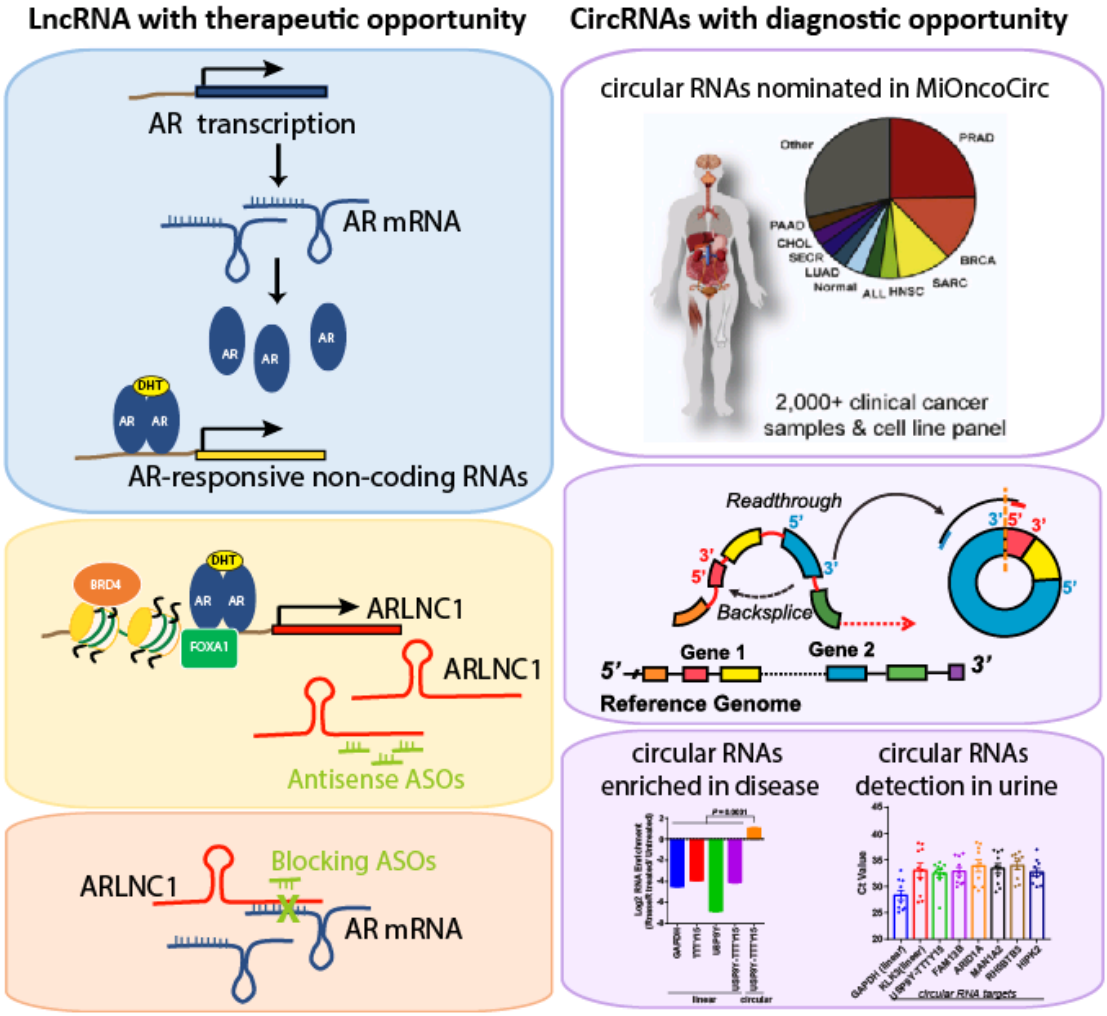
Conclusion

High-throughput sequencing technology has revolutionized our understanding of non-coding transcriptome. A growing number of non-coding RNA species has been discovered in well-known and novel classes of non-coding RNAs, including microRNAs, long non-coding RNAs, and circular RNAs. Despite functional relevance, their expression spectrum is largely specific to lineages, and deregulated in the processes of neoplasia growth, cancer progression, and drug response. Functional studies of non-coding RNAs have identified a small set of them as critical nodes in oncogenic or tumor suppressive regulation networks.

Through physical interactions with chromatin structures, DNA, RNA, or proteins, ncRNAs could either boost or suppress key pathways in cancer. Therefore, ncRNAs provide novel therapeutic and diagnostic opportunities for cancer management.

My thesis work has examined the androgen-responsive lncRNAs as well as circular RNAs in prostate cancer. We defined the AR-regulated non-coding transciptome and characterized the pro-oncogenic role of a lineage-specific lncRNA, ARLNC1. The involvement of ARLNC1 in AR signaling regulation echoes the roles of some other lncRNAs that function through feedback loops. Our work also suggests that the RNA-targeting antisense technology is likely to provide benefit for cancer types with oncogenic lncRNA expression. Additionally, the existence of circular RNAs species has been validated in prostate cancer. The disease-associated expression pattern of circular RNAs, plus their comparatively stable structures in cell-free bio-fluid, makes circular RNAs a new frontier in cancer biomarker development. Future directions of non-coding RNA research include curation of lncRNA annotations, development of large-scale screens to identify functional non-coding RNAs, as well as application of antisense technologies for precision medicine.

Figures



Figures 6.1 Summary of the thesis

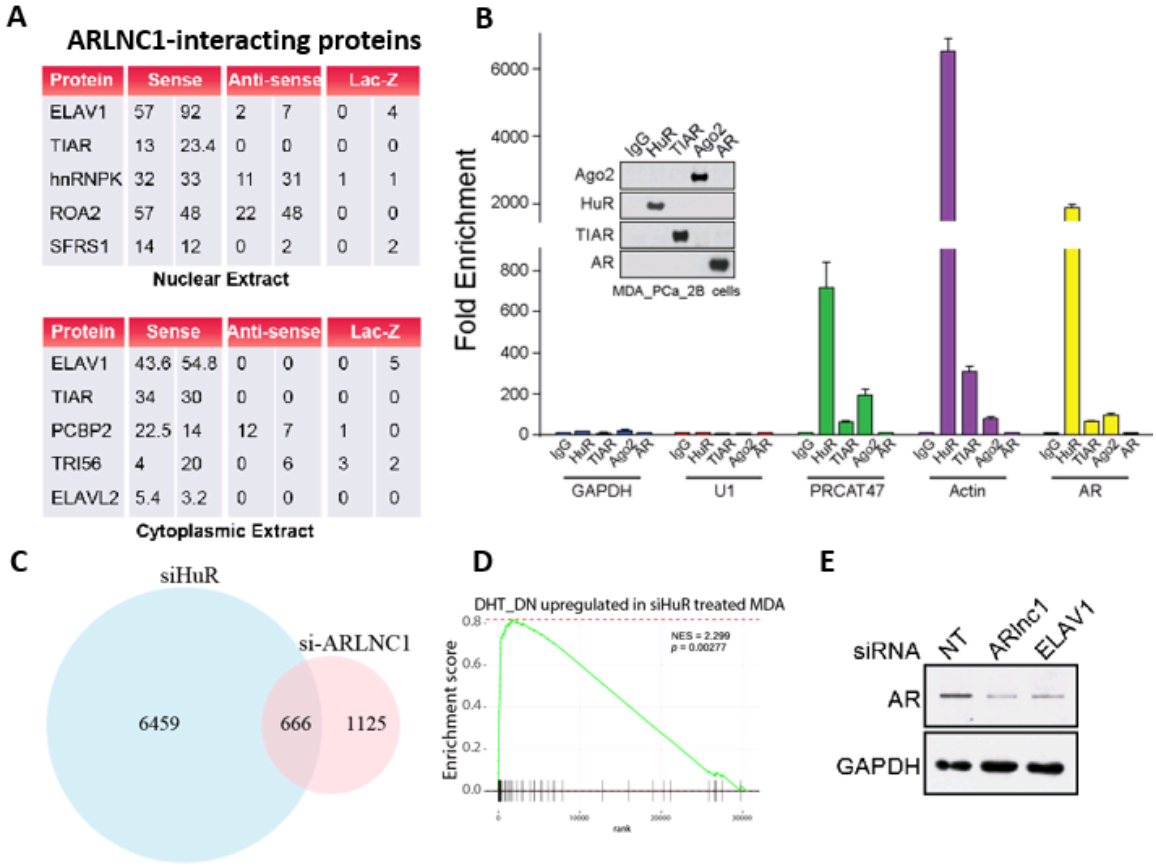


Figure 6.2 ARLNC1 interactome

(**A**) RNA immunoprecipitation followed by Mass-Spectrometry to identify proteins binding with ARLNC1.

(**B**) RNA immunoprecipitation (RIP) to validate the binding of ARLNC1 with candidate protein. Inset: Western blotting analysis of proteins pulled down by immunoprecipitation.

(**C**) Venn Diagram indicating overlap between HuR-regulated genes ARLNC1-responsive targets identified from RNA-Seq analysis following HuR loss or ARLNC1 loss.

(**D**) Gene Set Enrichment Analysis showing enrichment of AR-regulated gene set following HuR depletion.

(**E**) Immunoblot assay showing relative protein levels of AR and GAPDH in MDA-PCa-2B cells 72 hours following transfection of siRNA targeting ARLNC1 or ELAV1 (HuR).

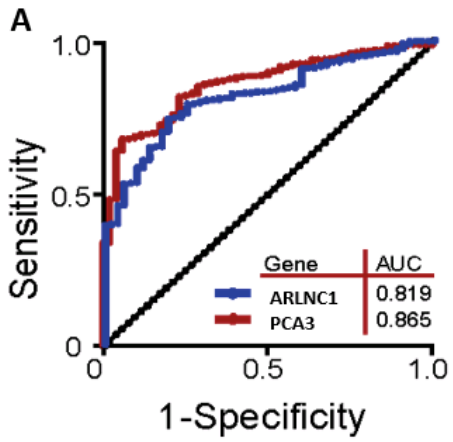


Figure 6.3 Clinical significance of ARLNC1
AUC value of ARLNC1 and PCA3 calculated from TCGA RNA-Seq dataset.

References

1. Brennan, C.M. and J.A. Steitz, *HuR and mRNA stability.* Cell Mol Life Sci, 2001. **58**(2): p. 266-77.
2. Garcia-Maurino, S.M., et al., *RNA Binding Protein Regulation and Cross-Talk in the Control of AU-rich mRNA Fate.* Front Mol Biosci, 2017. **4**: p. 71.
3. Kim, H.S., et al., *Different modes of interaction by TIAR and HuR with target RNA and DNA.* Nucleic Acids Res, 2011. **39**(3): p. 1117-30.
4. Mazan-Mamczarz, K., et al., *Translational repression by RNA-binding protein TIAR.* Mol Cell Biol, 2006. **26**(7): p. 2716-27.
5. Yeap, B.B., et al., *Novel binding of HuR and poly(C)-binding protein to a conserved UC-rich motif within the 3'-untranslated region of the androgen receptor messenger RNA.* J Biol Chem, 2002. **277**(30): p. 27183-92.
6. Yoon, J.H., K. Abdelmohsen, and M. Gorospe, *Functional interactions among microRNAs and long noncoding RNAs.* Semin Cell Dev Biol, 2014. **34**: p. 9-14.
7. Cesana, M., et al., *A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA.* Cell, 2011. **147**(2): p. 358-69.
8. Kallen, A.N., et al., *The imprinted H19 lncRNA antagonizes let-7 microRNAs.* Mol Cell, 2013. **52**(1): p. 101-12.
9. Li, C., et al., *The lincRNA-ROR/miR-145 axis promotes invasion and metastasis in hepatocellular carcinoma via induction of epithelial-mesenchymal transition by targeting ZEB2.* Sci Rep, 2017. **7**(1): p. 4637.
10. Faghihi, M.A., et al., *Evidence for natural antisense transcript-mediated inhibition of microRNA function.* Genome Biol, 2010. **11**(5): p. R56.
11. Ostling, P., et al., *Systematic analysis of microRNAs targeting the androgen receptor in prostate cancer cells.* Cancer Res, 2011. **71**(5): p. 1956-67.
12. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape.* Nat Protoc, 2007. **2**(10): p. 2366-82.
13. Tomlins, S.A., et al., *Urine TMPRSS2:ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA.* Sci Transl Med, 2011. **3**(94): p. 94ra72.
14. Tomlins, S.A., et al., *Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment.* Eur Urol, 2016. **70**(1): p. 45-53.
15. Veeraraghavan, J., et al., *Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers.* Nat Commun, 2014. **5**: p. 4577.
16. Vidigal, J.A. and A. Ventura, *Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries.* Nat Commun, 2015. **6**: p. 8083.
17. Gilbert, L.A., et al., *Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.* Cell, 2014. **159**(3): p. 647-61.
18. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.
19. Liu, S.J., et al., *CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells.* Science, 2017. **355**(6320).

20. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
21. Konermann, S., et al., *Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors.* Cell, 2018. **173**(3): p. 665-676 e14.
22. Yan, W.X., et al., *Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein.* Mol Cell, 2018. **70**(2): p. 327-339 e5.

APPENDIX: Author contributions

Chapter 1

This chapter was written by Yajia Zhang. The Figures were reprinted or adapted with permission from the following sources:

Figure 1.1 and **Figure 1.3**: Philip A. Watson, Vivek K. Arora, Charles L. Sawyers. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nature Reviews Cancer*, Nov 13, 2015. (License Number for Figure reprinting: 4592810956344)

Figure 1.2: Dan Robinson, Eliezer M. Van Allen, Yi-Mi Wu, Nikolaus Schultz, Robert J. Lonigro, Juan-Miguel Mosquera, Bruce Montgomery, Mary-Ellen Taplin, Colin C. Pritchard, Gerhardt Attard, Himisha Beltran, Wassim Abida, Robert K. Bradley, Jake Vinson, Xuhong Cao, Pankaj Vats, Lakshmi P. Kunju, Maha Hussain, Felix Y. Feng, Scott A. Tomlins, Kathleen A. Cooney, David C. Smith, Christine Brennan, Javed Siddiqui, Rohit Mehra, Yu Chen, Dana E. Rathkopf, Michael J. Morris, Stephen B. Solomon, Jeremy C. Durack, Victor E. Reuter, Anuradha Gopalan, Jianjiong Gao, Massimo Loda, Rosina T. Lis, Michaela Bowden, Stephen P. Balk, Glenn Gaviola, Carrie Sougnez, Manaswi Gupta, Evan Y. Yu, Elahe A. Mostaghel, Heather H. Cheng, Hyojeong Mulcahy, Lawrence D. True, Stephen R. Plymate, Heidi Dvinge, Roberta Ferraldeschi, Penny Flohr, Susana

Miranda, Zafeiris Zafeiriou, Nina Tunariu, Joaquin Mateo, Raquel Perez-Lopez, Francesca Demichelis, Brian D. Robinson, Andrea Sboner, Marc Schiffman, David M. Nanus, Scott T. Tagawa, Alexandros Sigaras, Kenneth W. Eng, Olivier Elemento, Andrea Sboner, Elisabeth I. Heath, Howard I. Scher, Kenneth J. Pienta, Philip Kantoff, Johann S. de Bono, Mark A. Rubin, Peter S. Nelson, Levi A. Garraway, Charles L. Sawyers, Arul M. Chinnaiyan. Integrative clinical genomics of advanced prostate cancer. *Cell*, Volume 162, Issue 2, 16 July 2015, Pages 454. (License Number for Figure reprinting: 4592810146567)

Figure 1.4 and **Figure 1.5**: Adam M. Schmitt, Howard Y. Chang. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*, 11 April 2016. (License Number for Figure reprinting: 4592800936029)

Chapters 2, 3, 4

These chapters were previously published in the article below:

Yajia Zhang[*], Sethuramasundaram Pitchiaya[*], Marcin Cieślik[*], Yashar S. Niknafs, Jean C.-Y. Tien, Yasuyuki Hosono, Matthew K. Iyer, Sahr Yazdani, Shruthi Subramaniam, Sudhanshu K. Shukla, Xia Jiang, Lisha Wang, Tzu-Ying Liu, Michael Uhl, Alexander R. Gawronski, Yuanyuan Qiao, Lanbo Xiao, Saravana M. Dhanasekaran, Kristin M. Juckette, Lakshmi P. Kunju, Xuhong Cao, Utsav Patel, Mona Batish, Girish C. Shukla, Michelle T. Paulsen, Mats Ljungman, Hui Jiang, Rohit Mehra, Rolf Backofen, Cenk S. Sahinalp, Susan M. Freier, Andrew T. Watt, Shuling Guo, John T. Wei, Felix Y. Feng, Rohit Malik[#], Arul M. Chinnaiyan[#] ([*]Co-first authors, [#]co-senior authors). Analysis of the androgen receptor–

regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression. *Nature Genetics*, volume 50, pages 814–824 (2018).

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

Chapter 5

This chapter was previously published as parts of the article below, and was reproduced with permission from the publisher (Elsevier):

Josh N. Vo, Marcin Cieslik, Yajia Zhang, Sudhanshu Shukla, Lanbo Xiao, Yuping Zhang, Yi-Mi Wu, Saravana M. Dhanasekaran, Carl G. Engelke, Xuhong Cao, Dan R. Robinson,

Alexey I. Nesvizhskii, Arul M. Chinnaiyan. The Landscape of Circular RNA in Cancer. *Cell*, 176, 869–881, Feb 7, 2019.

J.N.V., M.C., D.R.R., A.I.N., and A.M.C. designed the study. J.N.V., and M.C. ran the bioinformatics pipelines, performed data analysis, and prepared the figures with the help of A.I.N, D.R.R., and A.M.C. Yajia Z., S.S., and L.X. performed validation of circular RNA candidates, RNase R treatments, actinomycin D treatments, plasma incubation experiments, and prepared corresponding Figures. S.S., Yajia Z., L.X., and C.G.E. collected clinical urine samples, performed quality controls, RNA extraction, and qPCR analysis. Yuping Z. assisted with additional bioinformatics analysis and figure preparation. S.M.D., X.C., Y.-M.W., and D.R.R. carried out sequencing and coordinated patient cohorts for the cited studies. D.R.R., A.I.N., and A.M.C. supervised the study.

Chapter 6

This chapter was written by Yajia Zhang and data was acquired in collaboration with Rohit Malik and Anton Poliakov. Y.Z., R.M., and A.P. performed the RNA immunoprecipitation assays and Mass Spectrometry experiments. Y.Z performed the RNA-Seq analysis and Gene Set Enrichment Analysis. Y.Z and R.M evaluated the clinical significance of ARLNC1.