Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by:

Arlou Kristina J. Angeles, M.Sc.

born in Manila, Philippines

Oral examination: 19 July 2019

The ERG-driven long non-coding RNA *LINC00920* promotes cell proliferation and migration in prostate cancer cells by modulating FOXO activity through a direct interaction with 14-3-3ε

Referees: PD Dr. Odilia Popanda

Prof. Dr. Holger Sültmann

Declaration

I hereby declare that the submitted dissertation entitled “The ERG-driven long non-coding RNA *LINC00920* promotes cell proliferation and migration in prostate cancer cells by modulating FOXO activity through a direct interaction with 14-3-3ε” is the product of my own work and that all assistance received in preparing this thesis and manuscript have been acknowledged. I have not applied to be examined at any other institution, nor have I submitted this dissertation to any other faculty. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

_____ _____

Place, Date Arlou Kristina Angeles

Summary

This thesis describes *LINC00920*, a tumor-associated lncRNA identified in the transcriptome dataset of the International Cancer Genome Consortium-Early Onset Prostate Cancer (ICGC-EOPC) cohort. SiRNA-mediated knockdown of *LINC00920* negatively affected proliferation, colony formation, and migration of PC-3 prostate cancer cells. Gene set enrichment analysis of microarray expression data revealed perturbation of pathways related to cell cycle, cell division, apoptosis, and cell movement. Focused pathway analysis of the top *LINC00920*-deregulated genes showed an inverse relationship between the lncRNA expression and FOXO signaling. Furthermore, as measured by qPCR, knockdown of *LINC00920* activated canonical FOXO targets *GADD45A*, *BCL2L11*, and *PMAIP1* while overexpression of the lncRNA reversed this effect.

In both The Cancer Genome Atlas-Prostate Adenocarcinoma (TCGA-PRAD) and ICGC-EOPC cohorts, *LINC00920* positively correlated with ERG overexpression. The regulatory influence of ERG on the lncRNA was then established using cell line models of ERG overexpression, chromatin immunoprecipitation (ChIP) of ERG at the *LINC00920* promoter, and promoter luciferase assays using wild-type and mutant promoter fragments.

To address the question of how *LINC00920* elicits its associated cellular phenotypes with consideration to its presence across cytosolic, nucleoplasmic, and chromatin compartments, chromatin isolation by RNA purification (ChIRP) followed by high throughput DNA sequencing (ChIRP-seq) and mass spectrometry (ChIRP-MS) were conducted. At the chromatin level, *LINC00920* was found primarily associating with heterochromatin regions. *LINC00920* occupancy was also be detected in a subset of promoter regions and putative enhancer loci. Interestingly, the lncRNA trace across the mappable genome bore a resemblance to that of the enhancer-associated histone mark H3K4me1, suggesting a role for *LINC00920* at enhancer elements. At the protein level, most of the identified *LINC00920* interacting partners are well established RNA binding proteins typically associated with the process of transcription. Among the *LINC00920*-precipitated proteins robustly identified in three biological replicates were two 14-3-3 isoforms—14-3-3ε and 14-3-3ζ. Binding of *LINC00920* to 14-3-3ε but not to 14-3-3ζ was validated by RNA immunoprecipitation (RIP) and affinity purification of recombinant 14-3-3ε on streptavidin beads using biotinylated *LINC00920*.

FOXO activity is mitigated by AKT phosphorylation. FOXO phosphorylation triggers 14-3-3/FOXO complex formation, leading to nuclear exportation. Current results indicate the repressive influence of *LINC00920* on FOXO signaling as well as the positive interaction between the transcript and 14-3-3ε. Considering these observations, a rational hypothesis emerged wherein *LINC00920*/14-3-3ε binding further stabilizes the 14-3-3ε/FOXO complex, resulting in a more efficient sequestration and consequent deactivation of FOXO.

Altogether, this thesis contributes a novel mechanism for a tumor-associated lncRNA in the context of ERG-overexpressing prostate cancer cells. Beginning with the transcriptome analysis of the ICGC-EOPC cohort, and later the TCGA-PRAD dataset, *LINC00920* was identified to be an ERG-driven transcript. Ultimately, molecular characterization of *LINC00920* by ChIRP-MS has revealed its apparent role in modulating FOXO in conjunction with 14-3-3ε, resulting in reduced expression of a subset of tumor suppressive FOXO targets. Since *ERG* fusions are clonal events while *PTEN* deletions are subclonal, driving *LINC00920* transcription could be a strategy, in part, for ERG-positive cells to alleviate the influence of an intact *PTEN*, paving the way for tumorigenesis.

Zusammenfassung

Diese Dissertation beschreibt *LINC00920*, eine tumorassoziierte lncRNA, die im Transkriptom-Datensatz der International Cancer Genome Consortium-Early Onset Prostate Cancer (ICGC-EOPC) Kohorte identifiziert wurde. SiRNA-vermittelter Knockdown von *LINC00920* reduzierte die Proliferation, Koloniebildung und Migration von PC-3-Zellen. Eine Gen-Set-Anreicherungsanalyse von Microarray-Expressionsdaten zeigte, dass Gene, die im Zellzyklus, sowie bei Zellteilung, Apoptose und Zellbewegung eine Rolle spielen, dereguliert waren. Die am stärksten von *LINC00920* deregulierten Gene waren invers mit Aktivität des FOXO Signalweges azzoziiert. Knockdown von *LINC00920* führte zu einer erhöhten Transkription der kanonischen FOXO Zielgene *GADD45A*, *BCL2L11* und *PMAIP1*, während die Überexpression von *LINC00920* diesen Effekt umkehrte.

Sowohl im „The Cancer Genome Atlas-Prostate Adenocarcinoma“ (TCGA-PRAD) Datensatz als auch in der ICGC-EOPC Kohorte korrelierte *LINC00920* mit dem *TMPRSS2/ERG*-Fusionssstatus der Tumoren, d.h. ERG-Überexpression. Die Regulation von *LINC00920* durch *ERG* wurde mittels Chromatin-Immunpräzipitation (ChIP) und Luciferase-Tests mit Wildtyp- und Mutanten- Promotor-Sequenzen nachgewiesen.

Um die Bindungspartner von *LINC00920* im Zyto- bzw. Nukleoplasma sowie im Chromatin zu identifizieren wurden Chromatinisolierung durch RNA-Präzipitation (ChIRP), gefolgt von Hochdurchsatz DNA-Sequenzierung (ChIRP-Seq) bzw. Massenspektrometrie (ChIRP-MS) durchgeführt. Hierbei wurde *LINC00920* primär mit Heterochromatin assoziiert gefunden und war v.a. in Promotor- und Enhancerregionen angereichert. Die Genom-weite Verteilung von *LINC00920* zeigte eine Ähnlichkeit mit der Histonmarkierung H3K4me1, was auf eine regulatorische Rolle von *LINC00920* in Enhancer-Elementen hinweist. Auf Proteinebene waren die meisten der identifizierten *LINC00920* Interaktionspartner gut etablierte RNA-Bindungsproteine, welche typischerweise mit dem Transkriptionsprozess verbunden sind. Zu den *LINC00920*-präzipitierten Proteinen gehörten zwei 14-3-3 Isoformen: 14-3-3ε und 14-3-3ζ. Die Bindung von *LINC00920* an 14-3-3ε, aber nicht an 14-3-3ζ wurde durch RNA-Immunpräzipitation (RIP) und Affinitätsreinigung von rekombinantem 14-3-3ε auf Streptavidin-Beads mittels biotinylierter *LINC00920* nachgewiesen.

Die Aktivität von FOXO wird durch AKT Phosphorylierung reduziert, welche wiederum die Bildung von 14-3-3/FOXO-Komplexen auslöst, was zu einem Kernexport führt. Die Ergebnisse deuten auf den repressiven Einfluss von *LINC00920* auf den FOXO-Signalweg durch Bindung an das 14-3-3ε Protein hin. Basierend auf diesen Beobachtungen wurde gezeigt, dass die *LINC00920*/14-3-3ε-Bindung den 14-3-3ε/FOXO-Komplex stabilisiert, was zu einem erhöhten Abbau von FOXO und erhöhter Aktivität des AKT-Signalweges führt. Da *ERG*-Fusionen klonale Ereignisse darstellen, während *PTEN*-Deletionen subklonal sind, könnte die erhöhte *LINC00920*-Transkription eine

Strategie ERG-positiver Zellen sein, um den Einfluss der Tumorsuppression durch *PTEN* zu reduzieren und den Weg für die Tumorentstehung oder -progression zu ebnen.

List of Abbreviations

Abbreviation	Definition
ADT	androgen deprivation therapy
AMPK	AMP-activated protein kinase
APA	alternative polyadenylation
AR	androgen receptor
ARE	androgen response element
ARLNC1	Androgen Receptor Regulated Long Non-coding RNA 1
ASO	antisense oligonucleotide
cDNA	complementary DNA
CHART	capture hybrid analysis of RNA targets
CHGA	chromogranin A
CHGB	chromogranin B
ChIP	chromatin immunoprecipitation
ChIRP	chromatin isolation by RNA purification
circRNA	circular RNA
CLIP	crosslinked immunoprecipitation
co-IP	co-immunoprecipitation
Cp	crossing point-PCR-cycle
CPAT	Coding Potential Assessment Tool
CPC	Coding Potential Calculator
CRPC	castration resistant prostate cancer
CTBP1-AS	C-Terminal Binding Protein 1 Antisense
DHT	5α-dihydrotestosterone
DNA	deoxyribonucleic acid
Dox	doxycycline
EAU-ESTRO-SIOG	European Association of Urology-European Society for Radiotherapy and Oncology-International Society of Geriatric Oncology
EBRT	external beam radiation therapy
EDTA	ethylenediaminetetraacetic acid
EMT	epithelial-to-mesenchymal transition
eRNA	enhancer-associated RNA
ETS	E26 transformation specific or E-twenty six
FBS	fetal bovine serum
FH	forkhead
FHD	forkhead domain
FOXO	forkhead box class O
FRE	forkhead response element
FPKM	fragments per kilobase per million mapped reads
GM-CSF	granulocyte-macrophage colony-stimulating factor
GnRH	gonadotropin-releasing hormone
GO	gene ontology
GREAT	Genomic Regions Enrichment of Annotations Tool
GSEA	gene set enrichment analysis
HIFU	high-intensity focal ultrasound

HITS-CLIP	high-throughput sequencing of RNA isolated by crosslinked immunoprecipitation
ICGC-EOPC	International Cancer Genome Consortium-Early Onset Prostate Cancer
iCLIP	individual nucleotide resolution crosslinked immunoprecipitation
IgG	immunoglobulin G
LB	Luria-Bertani
LH	luteinizing hormone
lincRNA	long intergenic non-coding RNA
lncRNA	long non-coding RNA
MACS	Model-based Analysis of ChIP-Seq
mCRPC	metastatic castration-resistant prostate cancer
miRNA	microRNA
mRNA	messenger RNA
MS	mass spectrometry
NAT	natural antisense transcript
ncRNA	non-coding RNA
NE	neuroendocrine
NEPC	neuroendocrine prostate cancer
NES	nuclear export signal
NLS	nuclear localization signal
NSE/ENO2	neuron-specific enolase
ORF	open reading frame
PAP	prostatic acid phosphatase
PAR-CLIP	photoactivatable ribonucleoside-enhanced crosslinked immunoprecipitation
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline with Tween 20
PCa	prostate cancer
PCAT1	Prostate Cancer Associated Transcript 1
PCGEM1	Prostate Cancer Gene Expression Marker 1
PCR	polymerase chain reaction
PhyloCSF	Phylogenetic Codon Substitution Frequencies
PI3K	phosphoinositide 3-kinase
PIN	prostate intraepithelial neoplasia
piRNA	PIWI-interacting RNA
pri-miRNA	primary mRNA
PRNCR1	Prostate Cancer Associated Non-coding RNA 1
PSA	prostate specific antigen
PTEN	Phosphatase and tensin homolog
PVDF	polyvinylidene difluoride
qPCR	quantitative PCR
RACE	rapid amplification of cDNA ends
RAP	RNA antisense purification
REST	RE1-silencing transcription factor
RIP	RNA immunoprecipitation
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA-mediated interference

RP	radical prostatectomy
RT	radiation therapy
SBRT	stereotactic body radiotherapy
SChLAP1	Second Chromosome Locus Associated with Prostate 1
SCNA	somatic copy number alterations
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism
ssDNA	single stranded DNA
SYP	synaptophysin
T2E	*TMPRSS2:ERG*
TAE	Tris-Acetate-EDTA
TCGA-PRAD	The Cancer Genome Atlas-Prostate Adenocarcinoma
TES	transcription end site
Tet	tetracycline
TRAMP	transgenic adenocarcinoma of the mouse prostate
tRNA	transfer RNA
TRUS	transrectal ultrasound
TSS	transcription start site
T-UCR	transcribed ultraconserved non-coding RNA
UPL	universal probe library
USPTF	U.S. Preventive Task Force
UTR	untranslated region
UV	ultraviolet
VPT	vascular-targeted photodynamic therapy
WB	western blot

List of Tables

List of Figures

Table of contents

1. Introduction

1.1. Cellular physiology of the prostate

The human prostate is an exocrine gland located at the base of the bladder which is responsible for the production of a slightly alkaline fluid that protects and nourishes the sperm in the semen [1]. Organized as a pseudostratified epithelium, the prostate is comprised of three types of terminally differentiated epithelial cells—the luminal, basal, and neuroendocrine (NE) cells (**Figure 1-1)** [2]. Luminal cells line the prostatic lumen and produce secretory proteins such as prostate-specific antigen (PSA). In addition, luminal cells characteristically express high levels of luminal cytokeratins (CK8 and CK18), NKX 3.1, and the androgen receptor (AR) [2, 3]. Basal cells are found between the basement membrane and the luminal layer. These cells express high levels of basal cytokeratins (CK14, CK5), CK19, GSTP1, and the stem cell transcription factor p63, but very low levels of AR [3]. NE cells are very rare and considered to be the least characterized prostatic cell population. These cells are androgen independent and are identified by their expression of NE differentiation markers such as neuron-specific enolase (NSE/ENO2), chromogranin A (CHGA), chromogranin B (CHGB), and synaptophysin (SYP) [4, 5]. While the physiological function of NE cells within the prostate is not yet fully understood, their secretory products imply possible roles in regulation, differentiation, and proliferation of luminal and basal cells through exocrine, endocrine, paracrine, and autocrine mechanisms [4].

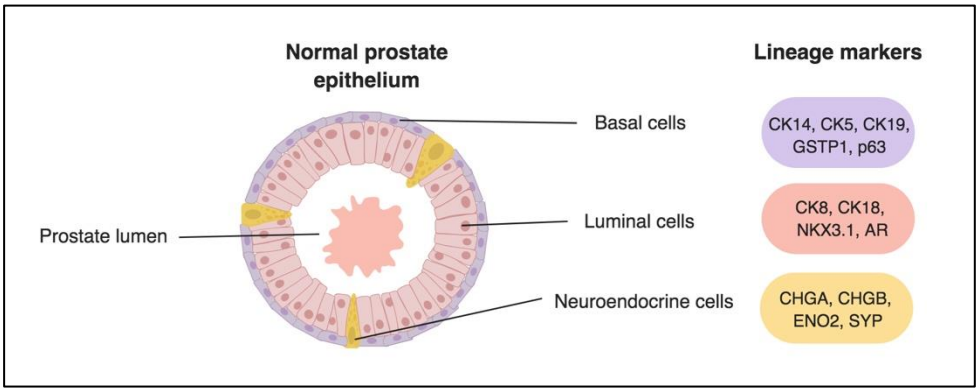


Figure 1-1. The prostate epithelium. Luminal cells produce secretory proteins and characteristically express luminal markers such as CK8, CK18, NKX3.1, and AR. Basal cells are found between the luminal layer and the basal lamina. These cells express high levels of CK14, CK5, CK19, GSTP1, and p63. A small population of neuroendocrine cells is present along the basal layer. These cells express high levels of neuroendocrine markers CHGA, CHGB, ENO2, and SYP.

Based on the expression of characteristic protein markers, prostate tumor phenotypes can be classified as either luminal or basal. However, definitive information on the cell of origin of prostate cancer (PCa) remains elusive and is the focus of active investigations. There have been efforts to correlate the cellular phenotype (i.e., luminal or basal) of tumors with clinical courses, but a clear consensus has yet

to be reached. One study reported that while luminal-type tumors can arise from basal cells, lesions of luminal origin are more aggressive and present a molecular signature that correlates with worse patient outcomes [6]. In addition, a study aiming to associate tumor phenotype with clinical response has revealed that a subset of luminal PCa exhibited the poorest diseases prognoses. Interestingly, the same luminal subset also correlated with response to androgen deprivation therapy (ADT) after prostatectomy [7]. On the other hand, it has been reported that a gene expression signature specific for human prostate basal stem cells is enriched in advanced metastatic disease, suggesting a common transcriptional program [8]. Adding to the complexity of prostatic cell dynamics are other cell types present within the surrounding stroma. These include fibroblasts, endothelial cells, immune cells, nerve fibers and associated ganglia, and smooth muscle cells—all of which could influence the maintenance and progression of the disease by molding the physical and biochemical tumor microenvironment.

1.2. Epidemiology and risk factors of prostate cancer

In western countries, PCa is the most frequently diagnosed malignancy in men. Since 2015, annual estimates of 1 million new cases and 300,000 mortalities have been reported for the disease [9]. Established non-modifiable risk factors for PCa development include age, race, and family history [10]. Among these, age is the most relevant component as reflected by the increasing standardized incidence rates in elderly men (**Figure 1-2A**) [11]. Extrapolations derived from autopsy studies suggest that most men would develop PCa if they lived longer than 100 years old [12]. Indeed, asymptomatic prostate malignancies have incidentally been found at autopsies, pointing to the tendency of localized PCa to remain undetected for many years before becoming clinically relevant.

PCa incidence is highest in more developed countries, which can partly be attributed to healthcare accessibility, specifically of screening and early detection programs [13]. Meanwhile, mortality rate is highest among men of African descent (**Figure 1-2B**). This is supported by a number of epidemiological studies with corroborating data on the increased prevalence and poorer outcomes of PCa in men of African descent compared to Caucasian and Asian males [14-16].

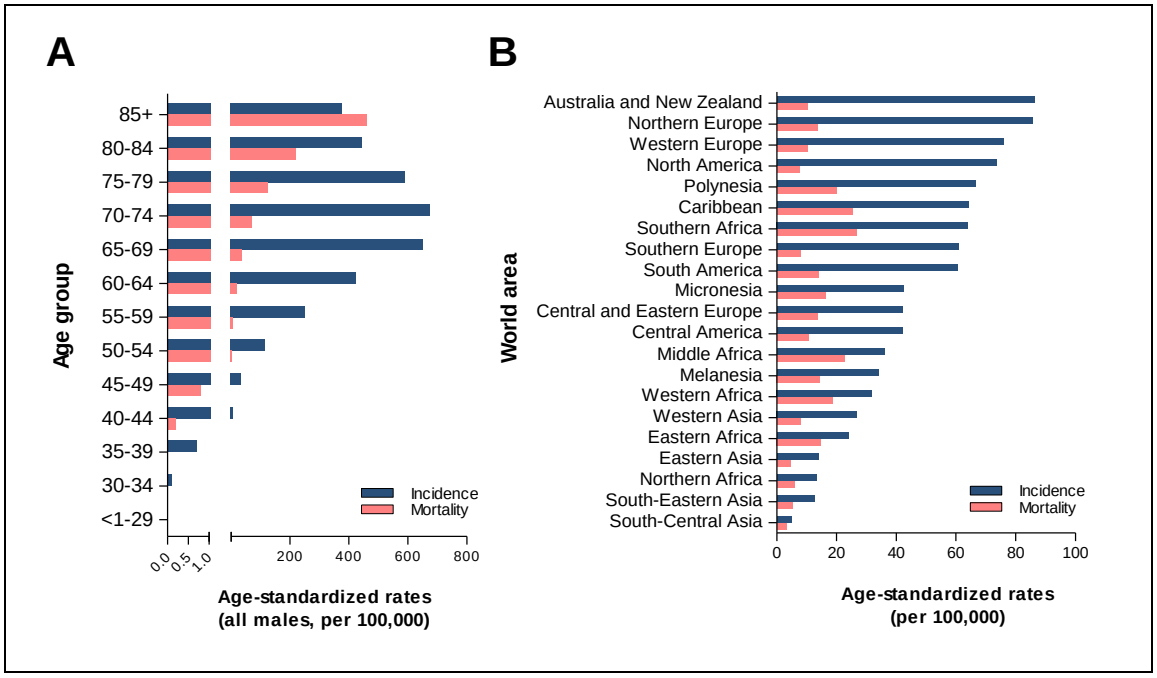


Figure 1-2. (A) Age-specific incidence and mortality rates of prostate cancer in the U.S. through 2011-2015. Data source: Surveillance, Epidemiology, and End Results (SEER) Program, National Cancer Institute [11]. (B) 2018 estimate of age-standardized incidence and mortality rates of prostate cancer in different world areas. Data source: International Agency for Research of Cancer (IARC) [9].

Multiple lines of evidence [17-20] indicate that prostate cancer is among the most heritable cancer entities together with malignancies of the stomach, colorectum, lung, and breast [17]. Multiple analyses of the Nordic twin registry have shown statistically significant heritable factors underlying PCa development. From these studies, risk attributed to heritable factors was quantified to be at 42% [17], 58% [19], and 57% [20]. In addition, an independent familial PCa study cohort in the Netherlands reported a 2.9-fold increased risk of PCa development for first-degree relatives of diagnosed patients [18]. Family-based linkage studies on European populations have identified multiple genes implicated in hereditary prostate cancer. Among them are *HPC1*, *PCAP*, *HPCX*, *CAPB*, and *HOXB13*[21-23]. In African American populations, 12q24, 1q24-5, 2p16, and 2p21 were identified linkage PCa loci [24, 25]. 1p36 was determined as a susceptibility locus by two independent studies on Japanese [26] and African American [25] cohorts. Over 100 PCa-related SNPs were also identified across populations based on genome-wide association studies [27-30]. In 2018, a genotype meta-analysis of more than 140,000 men revealed 63 novel PCa susceptibility loci [31].

Although many genetic factors have been identified to contribute to PCa development, genetic testing for hereditary PCa or clinical recommendations based on genetic information have yet to reach the clinical setting in the same way as successful genetic markers have with respect to other cancer entities. The only gene that could potentially break this translational barrier in the near future is

BRCA2. Germline mutations in *BRCA2* have shown promise as biomarkers for clinically aggressive PCa as well as for treatment response in metastatic disease [32, 33].

1.3. Development and progression of prostate lesions

Prostate cancers develop in a step-wise manner starting from prostate intraepithelial neoplasia (PIN) which is characterized by hyperproliferation of luminal cells resulting in dysplasia along the ducts (**Figure 1-3A**). The lesion then progresses to localized (confined to the prostate gland) adenocarcinoma, and becomes locally invasive adenocarcinoma once the basal cell layer has been degraded and neoplastic cells begin to penetrate through the basal lamina [2]. Metastatic seeding is initiated by cells draining into adjacent lymph nodes [34]. Metastatic colonization then occurs in distant organs, most commonly in the bone, liver, and lungs [35]. Bone metastases originating from the prostate often manifest as osteoblastic lesions which cause hypercalcemia, frequent fractures, and severe pain.

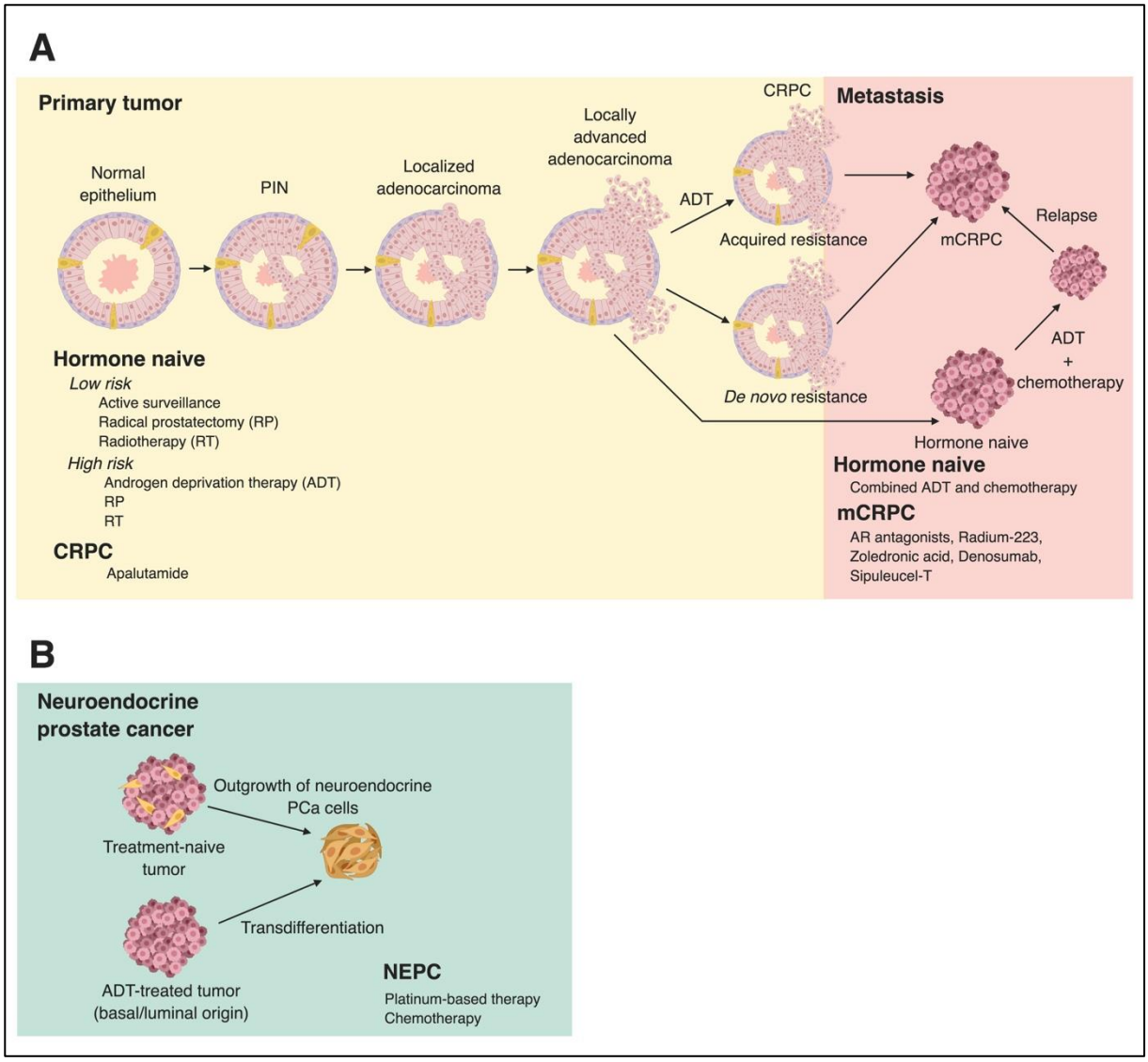


Figure 1-3. Prostate cancer development and progression. (A) Prostate intraepithelial neoplasia (PIN) initiates tumor formation. Metastasis begins upon colonization of cancer cells at proximal lymph nodes, and eventually at distant organs. At initial diagnosis, therapy options for localized PCa include active surveillance, radical prostatectomy (RP), and radiotherapy (RT). Androgen deprivation therapy (ADT) is administered to high-risk localized tumors, of which a large proportion will eventually acquire resistance (CRPC). In parallel, locally advanced adenocarcinoma can also be refractory to ADT and exhibit *de novo* resistance. Systemic therapies are administered to castrate-resistant and metastatic tumors. Hormone-naive metastatic tumors initially respond to ADT but similarly progress to metastatic CRPC (mCRPC). (B) Neuroendocrine PCa (NEPC) is a highly aggressive clinical subtype with poor prognosis. *De novo* NEPC has been suggested to originate from transformed prostatic neuroendocrine cells. Alternatively, multiple lines of evidence support the transdifferentiation model of androgen-deprived adenocarcinomas to neuroendocrine carcinomas.

1.4. PCa diagnostics

The rise of PCa incidence since the mid-1990s has been partially attributed to the implementation of the blood-based PSA screening in the U.S. and Europe. However, overtreatment of the disease—where patients undergo therapeutic courses that eventually bore no clinical benefit—became a consequence of increased PCa diagnoses. As a result, recommendation for non-discriminate PSA

screening has now been withdrawn by the U.S. Preventive Task Force (USPTF) [36] while the European Association of Urology-European Society for Radiotherapy and Oncology-International Society of Geriatric Oncology (EAU-ESTRO-SIOG) limits PSA testing to men with elevated risk of the disease [37].

PSA levels in the blood at midlife have been shown to be indicative of the risk of cancer at the time of testing. More recently, it has also been demonstrated to be useful as a predictive risk assessment tool for metastasis and cancer specific death in a subset of patients [38]. However, since PSA is not a cancer-specific marker, definitive diagnosis can only be made—following assessment of elevated PSA concentration (>4 ng/mL) and prostate enlargement—upon microscopic evaluation of prostate tissue sampled through needle biopsy. Conventionally, transrectal ultrasound (TRUS) is used to perform a systematic prostate biopsy [39]. A pathologist then scores the biopsies according to the Gleason grading system. Scores of primary and secondary patterns are combined to account for morphologic heterogeneity [40]. Ultimately, clinicians base the final diagnosis regarding the risk of tumor progression on a combined assessment of PSA levels, Gleason patterns, and clinical stage.

1.5. Management of PCa

Approximately 80-90% of newly diagnosed cases involve localized and regionally confined disease [41, 42]. In such early phases, prostatectomy and radiation therapy (RT) can be curative, with 99% of patients achieving 5-year relative survival [43]. Nevertheless, about one-third of patients who have undergone radical protastectomy (RP) and up to half of patients treated with radiation therapy will develop biochemically recurrent disease (rising PSA levels) [42]. Surgery and radiation therapy can be performed as salvage treatments. In non-responsive cases, androgen deprivation therapy (ADT) is the next line of treatment. ADT can delay the progression of disease since PCa cells are, for the most part, heavily reliant on sustained androgen signaling. Patients typically respond well to initial ADT but eventual androgen refraction inevitably occurs [44]. Once tumors become resistant to ADT, the disease has advanced to castration resistant PCa (CRPC) (**Figure 1-3A**). CRPC is treated with androgen blockade and/or systemic chemotherapy to which patient response rate has been reported to be about 50% [45]. Treatment modalities for prostate cancer have increased in recent years. However, state-of-the-art systemic therapies can only prolong late-stage patient survival for a few months [46].

1.5.1. Local therapies

The primary therapeutic options for localized PCa are expectant management, radical prostatectomy, and radiation therapy (**Figure 1-3A**). Expectant management monitors disease progression without treatment and can be further classified into observation (or watchful waiting) and active surveillance [39]. Watchful waiting involving palliative care is undertaken for frail or elderly men with more aggressive comorbidities that will most likely out-compete the localized tumor [47]. Active surveillance is intended for younger men diagnosed with low-grade cancer (Gleason score of 6 or less), with the aim of delaying treatment and its side-effects until the disease progresses [48].

Surveillance policies vary, but patients with low-risk tumors are routinely monitored using serum PSA tests, prostate biopsies, and MRI [49]. Due to longer life expectancy, men under active surveillance are followed closely, and treatment is initiated once clinical assessments worsen with the intent to reach a curative window [50].

Since considerable risks are associated with operative procedures and post-operative recovery, treatment of localized cancer with RP is an option limited to men with life expectancy of at least 10 years [51, 52]. Radiation therapy approaches for the treatment of localized PCa include external beam radiation therapy (EBRT), stereotactic body radiotherapy (SBRT), brachytherapy, and proton therapy. With the evolution of radiation therapy and imaging techniques, safer administration of higher doses of radiation at localized regions has become possible. Alternative local therapeutic strategies under development for clinical application are cryotherapy, and high-intensity focal ultrasound (HIFU). Cryotherapy or cryoablation is a minimally invasive procedure that damages tumor tissues by targeted exposure of lesions to pressurized argon and helium gases [53]. HIFU uses ultrasonic wave transmission to damage tissues by mechanical and thermal means [54]. Vascular-targeted photodynamic therapy (VPT) is an emerging treatment modality wherein a tumor vasculature photosensitizer drug is intravenously administered and a specific wavelength of light is delivered to the tissue of interest resulting in the erosion of vessel anatomy and ablation of tumor tissue [55]. These local therapies were developed with the aim of mitigating side-effects and long-term toxicities that come with RP and RT. Nonetheless, further studies are still necessary to compare the effectiveness of these measures against the standard of care [39, 49].

1.5.2. Systemic therapies

Upon progression to metastatic disease, therapy is shifted from local treatments to systemic chemotherapy (**Figure 1-3A**). Docetaxel and cabazitaxel are the first-line and second-line chemotherapeutics, respectively, administered to patients with metastatic PCa [35]. Both drugs are taxanes that induce cell cycle arrest and inhibit AR translocation into the nucleus by stabilizing microtubular structures [56]. Radium-223 (^{223}Ra) is administered intravenously to patients with bone metastases. ^{223}Ra is an alpha particle emitter that selectively targets osteoblastic metastases [57]. Sipuleucel-T therapy is an autologous cellular immunotherapy in which dendritic cells are harvested from patients and activated *ex vivo* by prostatic acid phosphatase (PAP) and granulocyte-macrophage colony-stimulating factor (GM-CSF) [58]. Activated antigen-presenting cells are then infused into the patient causing T-cell proliferation and priming, enabling the immune cells to recognize and target prostatic tumor cells. Zoledronic acid, a bisphosphonate, and denosumab, a human monoclonal antibody against RANKL, are osteoprotective agents used to manage osteoclast-mediated bone resorption in recurrent metastatic PCa [59, 60]. Finally, tumors with small cell or neuroendocrine histology were found to be sensitive to platinum-based therapies (i.e., cisplatin, carboplatin, and oxaliplatin) [61].

1.5.3. Androgen signaling as a therapeutic target in PCa

The androgen receptor (AR) is a ligand-dependent transcription factor that regulates the expression of genes that control male sexual development and differentiation [62]. AR activation and subsequent nuclear translocation is mediated by binding of its canonical ligands, 5α-dihydrotestosterone (DHT) and testosterone [44]. Androgen biosynthesis is initiated by the release of hypothalamic gonadotropin-releasing hormone (GnRH), which stimulates luteinizing hormone (LH) secretion from the anterior pituitary gland, subsequently triggering testosterone production in the testes. Upon uptake from the circulation by prostate cells, testosterone is converted to the more potent metabolite DHT which binds with high affinity to AR, leading to nuclear translocation of the receptor. Nuclear AR homodimers then directly associate with androgen response elements (AREs) found at promoter regions of target genes such as *PSA*, *TMPRSS2*, *NKX3.1*, and *PSMA* (**Figure 1-4**) [63]. Since AR activation promotes cell growth and survival, AR activity is critically linked to PCa development and progression. Thus, disrupting AR signaling is an important therapeutic strategy against the disease.

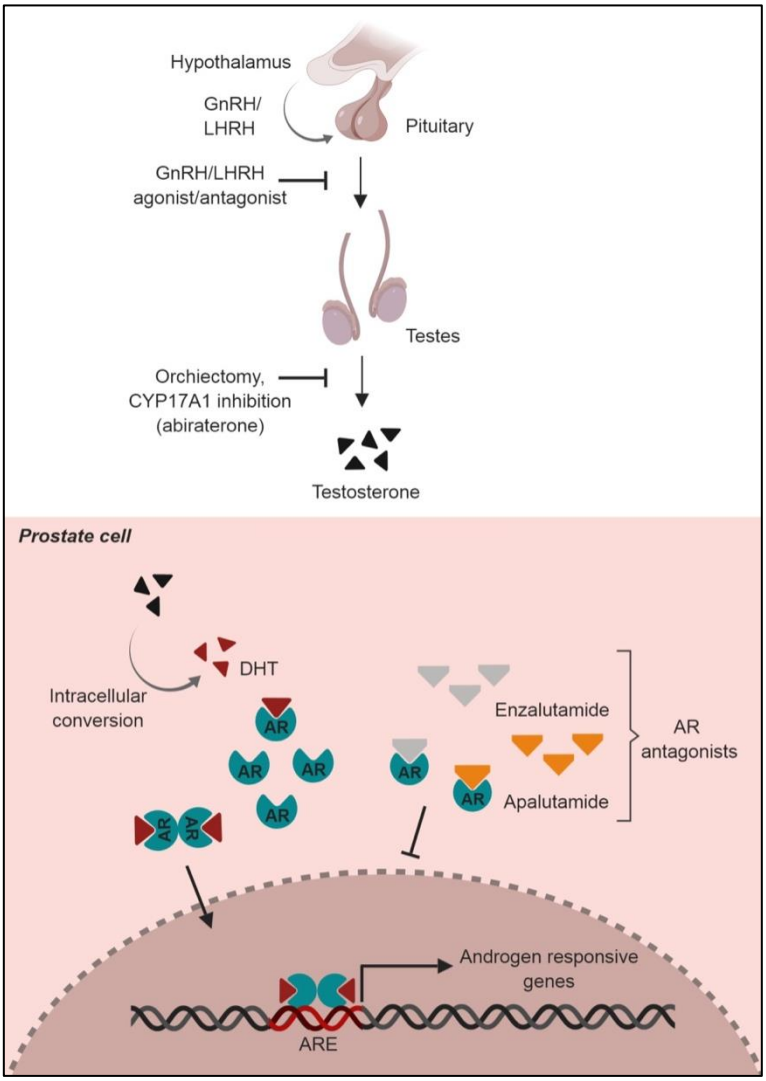


Figure 1-4. Androgen signaling in prostate cells.

Androgen deprivation therapy (ADT) is an adjuvant therapy with surgery or RT in high-risk localized disease, and a primary systemic therapy for men with metastatic PCa [47]. ADT is implemented in three ways: (i) bilateral orchiectomy (surgical castration); (ii) using luteinizing hormone-releasing hormone (LHRH, or gonadotropin-releasing hormone, GnRH) agonist or antagonist (medical castration); and (iii) administration of androgen synthesis inhibitors [64]. ADT can be augmented by androgen receptor blockade using AR antagonists [42], resulting in a regimen of complete androgen blockade [65]. It is well established that ADT provides initial clinical benefit. However, the majority of patients will become refractory to the treatment and progress to castration resistance within 2-3 years [66].

Enzalutamide, abiraterone, and apalutamide are hormonal agents that suppress androgen signaling activity. Enzalutamide and apalutamide are androgen receptor antagonists while abiraterone is an irreversible inhibitor of CYP17A1, an enzyme required in the steroidogenesis pathway that synthesizes DHT [67]. Both pre- and post-chemotherapy administration of either enzalutamide or abiraterone delayed disease progression and improved overall survival in metastatic castration-resistant prostate (mCRPC) cancer patients [68-71]. The next-generation androgen receptor inhibitor apalutamide has also been reported to prolong metastasis-free survival and time to tumor progression in non-metastatic CRPC [72].

1.6. Molecular features of PCa

While the majority of prostate tumors follow an indolent course, a significant number of cases progress with highly heterogeneous clinical trajectories that ultimately lead to lethal outcomes [73]. Conversely, high proportions of low-risk tumors are treated non-discriminately with high-risk tumors. Such overtreatment negatively impacts a patient's quality of life. The urgent need to identify clinically useful prognostic markers led to comprehensive and large-scale investigations into the genome of primary prostate tumors and mCRPC [74-80]. Numerous studies on the PCa transcriptome [75, 78, 81, 82], epigenome [83-86], and proteome [87-89] have also been undertaken with the same intent of understanding the biology of prostate oncogenesis and identifying clinically relevant targets.

The PCa genome harbors a relatively low mutational burden (approximately 1 mutation per megabase) compared to other tumor entities [75]. In contrast, multiple chromosomal gains and losses frequently occur. These DNA aberrations often lead to dysregulated processes implicated in, but not limited to, prostate development, cell-cycle regulation, cell survival, and chromatin organization [74, 90, 91]. Chromosome loss at 6p, 8p, 13q, and 16p are early events in prostate tumorigenesis, resulting in the inactivation of tumor suppressors *NKX3.1* and *RB1*. Chromosomal alterations at the *PTEN*, *TP53*, and *CDKN1* loci are also characteristic of the PCa genome. Frequent gains at chromosome 7 and 8q—along the *c-MYC* locus—are similarly common. As the disease progresses, genome-wide somatic copy number alterations (SCNA) drastically increase, with quantified CNA burden values at 4-5% in primary tumors versus 32% in metastatic tumors [92].

1.6.1. Primary tumors

A molecular taxonomy of primary prostate cancer has emerged based on integrated analyses of somatic mutations, copy number alterations, gene fusions, gene expression, and DNA methylation [75]. The tumor subtypes were defined as either harboring gene rearrangements of E-twenty-six (ETS) genes (i.e., *ERG*, *ETV1*, *ETV4*, and *FLI1*) or somatic mutations (i.e., *SPOP*, *FOXA1*, and *IDH1*). The most frequently occurring tumor class is characterized by the presence of *ERG* gene fusions (46%). These tumors overexpress the oncogenic transcription factor ERG through a translocation event on chromosome 21q that fuses ERG downstream of an androgen-regulated gene, most commonly, *TMPRSS2* [93, 94]. *TMPRSS2:ERG* (T2E)-positive tumors demonstrate characteristic transcriptional and epigenetic profiles [86]. Numerous studies have contributed to the understanding of ERG function in the context of prostate cancer cells [86, 95-97]. As a transcription factor belonging to the ETS family, ERG specifically binds to the GGA(A/T) ETS domain motif and activates target genes such as *MMP3*, *PLAT*, and *PLAU* [93, 98]. Moreover, ERG has been shown to interact with the AR and the AP-1 complex [96, 99]. More recently, ERG has been implicated in perturbing the chromatin landscape in primary PCa tumors, affecting *cis*-regulatory elements and chromatin architecture via recruitment of chromatin modifiers [86, 95, 96].

1.6.2. Advanced disease

While the relative distribution of tumor subtypes described remains comparable between primary and metastatic lesions [75], increased CNAs and somatic mutations are apparent in advanced disease. Hallmark genetic alterations in metastatic PCa target the AR, PTEN-PI3K, WNT, cell cycle, and DNA repair pathways. About 70% of mCRPC cases harbor alterations in AR signaling, of which the majority involves direct *AR* amplification or mutation [76]. Mutations in genes cooperating with AR, such as *ZBTB16*, *NCOR1*, *NCOR2*, *FOXA1*, and *SPOP* were also identified. This high incidence of aberrations implies the significance of AR signaling to the viability of metastatic cells. Genetic alterations of genes belonging to the PTEN-PI3K pathway (i.e., *PTEN*, *PIK3CA*, *PIK3CB*, *PIK3R1*, and *AKT1*) occurred in almost half of the studied cohort. Members of the WNT signaling pathway including *APC*, *CTNNB1*, *RNF43*, *ZNRF3*, and *RSPO2* were mutated in 18% of cases. Alterations in cell cycle-related genes including *RB1*, *CDKN1B*, *CDKN2A/B*, *CDKN2C*, *CCND1*, and *CDK4* were observed in 20% of cases while genes involved in the DNA repair pathway, namely *BRCA1/2*, *ATM*, *FANCA*, *RAD51B*, *RAD51C*, *MLH1*, and *MSH2* were found mutated in at least 22% of cases [76].

Neuroendocrine PCa (NEPC) is a highly aggressive clinical subtype occurring in about 2% of treatment-naïve cases (i.e., *de novo*) [100]. *De novo* NEPC has been suggested to arise from neoplastic transformation of normal prostatic neuroendocrine cells and subsequent outgrowth [101]. On the other hand, hormone-resistant tumors are hypothesized to transdifferentiate and terminally present the same histological features as *de novo* NEPC (i.e., treatment emergent) (**Figure 1-3B**) [102]. NEPC has a very poor prognosis and is characterized by rapid disease progression, recurrent bone lesions, and

increased metastatic spread, and declined survival. Aside from expression of neuroendocrine markers (e.g., SYP, CHGA, CHGB, NSE/ENO2), NEPC cells are characterized by increased concurrent loss of *RB1* and *TP53* [102, 103], upregulation of *EZH2*, *n-MYC* and *AURKA* [104], and attenuated AR and RE1-silencing transcription factor (REST) signaling [105].

1.7. Non-coding RNAs in cancer

The genesis of different cell types harboring the same genomic information relies on impeccable gene expression regulation. The resulting transcriptional programs determine and impart molecular and functional properties to cells. In the decades following the groundbreaking discovery of DNA in the 1950s, the central dogma of molecular biology, which asserts that genetic information encoded in the DNA flows via RNA towards protein synthesis, has significantly influenced the landscape of RNA research. Historically, RNA molecules were, for the most part, characterized only as temporary carriers of genetic information as mRNAs, components of the ribosome complex as rRNAs, or codon readers of the translation machinery as tRNAs [106]. In contrast, proteins were considered to be the functional terminal product of genetic information despite protein-coding exons constituting only about 2% of the human genome [107]. Pioneering works on model organisms demonstrated that non-coding RNAs (ncRNAs) such as *lin-4* [108], *let-7* [109], and *XIST* [110] exhibit dynamic expression patterns and conserved functionalities. With the advent of high throughput sequencing technologies, it became clear that, in addition to protein-coding genes, non-coding transcripts are also expressed in a cell-type-specific manner during the course of cell development [111-113].

In recent years, ncRNAs have been implicated in cellular processes involved in normal physiology, and their aberrant regulation has been associated with progression of multiple cancer entities [114]. While by no means definitive, ncRNAs are classified according to their size as short ncRNAs (18-21 nucleotides), mid-size ncRNAs (20-200 nucleotides), and long non-coding RNAs (lncRNAs; >200 nucleotides). Overall, microRNAs (miRNAs) and lncRNAs are among the most studied ncRNAs. They have been shown to be differentially expressed between tumors and their normal tissue counterparts in many cancer entities [115], suggesting functionality. Nonetheless, the biological roles of most ncRNAs remain unclear. **Table 1-1** shows selected examples of each ncRNA class and their potential functional associations with cancer.

Table 1-1. Classes of non-coding RNAs and their functional implications in cancer.

Name	Size (nt)	Examples	Described function	Cancer association	References
Short non-coding RNAs					
MicroRNAs (miRNAs)	18-24	*miR-34, miR-200, miR-375, let-7*	Initiates RISC-mediated degradation of target mRNA	Perturbed expression in cancer resulting in dysregulation of target genes related to cancer hallmarks; harbors potential utility as biomarkers	[114, 116-118]
PIWI-interacting RNAs (piRNAs)	26-31	*piR-4987, piR-932, piR-20365*	Epigenetic and post-transcriptional repression of transposons in germ line cells	Commonly overexpressed in seminomas, breast, ovarian, and cervical cancers	[119, 120]
Mid-size non-coding RNAs					
Small nucleolar RNAs (snoRNAs)	60-300	*SNORD50A, SNORD50, SNORD44, SNORD76*	Components of ribonucleoproteins; recruitment of modification enzymes to rRNAs; generation of miRNA-like ncRNAs	Perturbed expression in cancer; *in vitro* functional assays demonstrate effect on cancer cell growth	[121-124]
Transfer RNAs (tRNAs) and derivatives: tRNA halves (tiRNAs) and tRNA-dervied small fragments (tRFs)	76-90	*tRF-1, tRF-1001, CU1276*	Codon reading for protein translation; displacement of mRNA stabilizing proteins from 3' UTR	Perturbed expression of tRNA derivatives in some cancer entities; expression of certain tRNA derivatives are implicated in cell growth arrest	[125, 126]
Long non-coding RNAs					
Long intergenic non-coding RNAs (lincRNAs)	>200	*XIST, MALAT1, NEAT1, NORAD, PCAT1, ARLNC1*	Epigenetic modification; post-transcriptional modification; chromatin organization; protein scaffolding; miRNA sponging	Differentially expressed in tumors; perturbed expression leads to dysregulation of cancer hallmark genes; potential utility as biomarkers	[81, 127-130]
Natural antisense transcripts (NATs)	>200	*ANRIL, ZFAS1*	Modulation of sense transcription	Overexpressed in tumors leading to downregulation of tumor suppressive genes	[131-133]
Pseudogenes	>200	*PTENP1, BRAFP1, KRASP1*	Regulation of homologous genes through miRNA competition	Perturbed expression in tumors resulting in differential expression of important oncogenes or tumor suppressors	[134-136]
Enhancer-associated RNAs (eRNAs)	>200	*CCAT1 eRNA, ACTRT1 eRNA, PSA eRNA*	Augments enhancer function; maintains the active chromatin state of transcription locus	Functional alterations of eRNAs have been demonstrated in multiple cancer entities including breast cancer, CRPC, colorectal cancer, and basal cell carcinoma	[137-139]
Transcribed ultraconserved non-coding RNAs (T-UCRs)	>200	*Uc.8, Uc.73, Uc.300A*	Inhibition of miRNA processing; suggested to impact RNA processing and transcription	Perturbed expression in colorectal cancer and neuroblastomas; aberrant expression affects apoptosis and cellular proliferation	[140-142]
Circular RNAs (circRNAs)	>200	*Hsa_circ_0022383, Hsa_circ_0001946, Circ-ITCH, Circ-FOXO3*	Gene regulation through miRNA competition	Demonstrated to play anti-tumorigenic roles by sponging oncogenic miRNAs	[143-145]

1.7.1. MiRNAs in cancer

MiRNAs comprise a class of short ncRNAs ranging from 18 to 24 nucleotides in length [114]. These short ncRNas are well-established regulators of gene expression at the post-transcriptional level. MiRNAs target messenger RNAs either through translational repression or mRNA degradation via the RNA-induced silencing complex (RISC), with the latter being the predominant miRNA-mediated gene repression mechanism in mammals [146]. Transcription of miRNA genes by RNA polymerase II initially yields primary miRNAs (pri-miRNAs) which contain at least one stem loop structure. The DGCR8/DROSHA microprocessor complex cleaves the pri-miRNA at the base of the hairpin structure, yielding a pre-miRNA. The pre-miRNA is exported from the nucleus, through the nuclear pore complex, to the cytoplasm where the RNA is further processed by the RISC loading complex (RLC). The RLC harbors the endoribonuclease DICER1 which removes the loop of the pre-miRNA hairpin. The RNA duplex is then loaded into an Argonaut protein (AGO2 in humans) where the non-guiding strand is degraded. Upon high-complementarity but imperfect base-pair binding of a target mRNA to the mature miRNA within the RISC, the resulting RNA duplex is cleaved by AGO2 resulting in mRNA degradation [147, 148]. Because canonical miRNA target binding relies only on a short seed region along positions 2-7 [149], miRNAs can regulate multiple mRNA targets.

Aberrant expression of miRNAs in human cancers is well established [150]. In prostate cancer, microarray- [151, 152] and deep sequencing-based [153] expression studies on benign and tumorigenic tissue have revealed differential expression of miRNAs. In prostate tumors, the most upregulated miRNAs include *miR-375* [153, 154], *miR-200c* [153], and *miR-141* [154] while *miR-221*, *miR-222* [151], *miR-143*, and *miR-145* [153] were among the most downregulated. The functional consequences of such expression dysregulation were investigated using *in vitro* PCa models. For example, both *miR-375* and *miR-220c* were shown to target and repress *SEC23A* mRNA, whose protein product plays a role in shuttling proteins from the endoplasmic reticulum to the Golgi. This impairment in protein trafficking was implicated in reducing the presentation of MHC-I molecules on PCa cells, resulting in decreased tumor immunogenicity [153]. In an independent study, prostatic *miR-375* was also found to target the tumor suppressor *CBX7*, a member of the Polycomb repressive complex 1, expanding the miRNA regulatory network to include epigenetic modifiers [117]. In the same cancer entity, the tumor suppressive *miR-34b* was demonstrated to target DNA methyltransferases and histone deacetylases leading to partial demethylation and active chromatin modifications [155]. Indeed, numerous prostate cancer-associated miRNAs have already been described [156], and it is becoming apparent that certain miRNA signatures can be useful in clinical assessments. Consequently, defining these signatures and designing clinically feasible assays for miRNA quantification are currently at the forefront of active investigations.

1.7.2. The multiple functional modalities of lncRNAs

LncRNAs are defined as transcripts at least 200 nucleotides long with minimal or completely without coding potential. These RNA molecules are processed similarly as messenger RNAs (mRNAs): they are transcribed by RNA polymerase II and can undergo 5'-capping, splicing, and 3'-polyadenylation. On the other hand, lncRNA genes are generally comprised of fewer exons relative to protein coding genes while the transcripts themselves are less abundant and more cell type-specific compared to mRNAs. Interestingly, while lncRNA genes have higher tendencies to be conserved compared to neutrally evolving ancestral repeat sequences, these non-coding genes are also under weaker selective pressure compared to protein coding genes [157]. For long intergenic RNAs (lincRNAs), only about 12% of human and mouse genes are conserved in other species [158, 159]. Since lncRNA functionality presumably stems from its secondary structure in most cases, sequence alterations in lncRNA genes are more tolerated. This is in contrast to protein coding genes where base changes lead to altered codon usage, and by extension amino acid sequence, that would largely impact protein activity [160].

At present, the number of annotated human lncRNAs has reached more than 150,000 [161]. Of these, only a few have been sufficiently characterized—mostly due to technical difficulties encountered in studying these transcripts at the resolution of mechanistic detail—leaving to speculation the function of the majority. Nonetheless, the mechanisms of lncRNA action that have been described to date [162-165] have become precedent examples in establishing their roles in important cellular processes. LncRNA functions are oftentimes contingent upon the subcellular localization of the transcript. For nuclear-enriched lncRNAs, functional modalities include recruitment of chromatin modifiers to specific chromosomal loci [166-168], roles in mRNA processing [169], scaffolds for protein complexes [165, 170], and decoys for transcription factors [171]. In the cytoplasm, lncRNAs are thought to act as miRNA sponges [172, 173], protein sinks [129], and modulators of mRNA stability [174] (**Figure 1-5**).

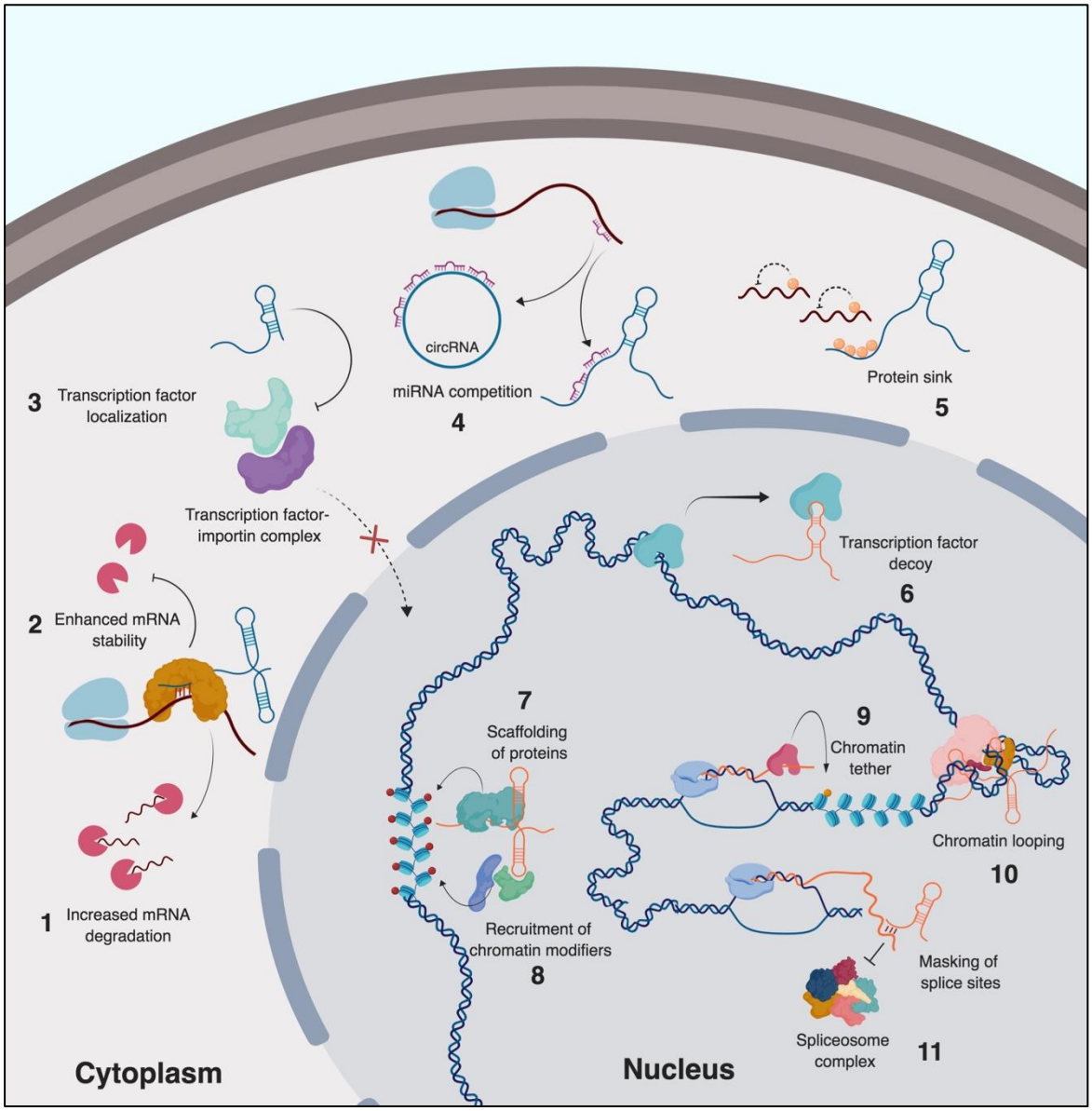


Figure 1-5. Representative lncRNA modalities described in the cytoplasmic and nuclear compartments. 1, LncRNA/mRNA binding can result in decreased mRNA stability through recruitment of destabilizing factors. **2**, Alternatively, lncRNA/mRNA binding can enhance mRNA stability through recruitment of an AU-rich element (ARE) binding protein that increases the stability of ARE-containing transcripts. **3**, By disrupting the interaction between a transcription factor and the importin complex, a lncRNA can affect the localization of the host transcription factor. **4**, Increased gene expression can result from miRNA competition between lncRNAs and cognate mRNA targets. **5**, By acting as a protein sequestering factor for RNA binding proteins that establish mRNA stability, a lncRNA may affect the regulation of cognate mRNA targets. **6**, Transcription of certain genes may be modulated by lncRNAs acting as decoys that titer away transcription factors from specific chromosomal loci. **7**, Protein complex assembly can be mediated by a transcript acting as a physical scaffold. **8**, LncRNAs have been implicated in the recruitment of chromatin remodeling complexes at certain chromosomal loci. **9**, While being transcribed, lncRNAs can act to tether chromatin modifiers and facilitate chromatin remodeling in *cis* regions. **10**, LncRNAs may mediate long-range interactions of regulatory elements such as enhancers by facilitating chromosomal looping. **11**, LncRNAs may affect splicing by masking splice sites through base complementarity binding to nascent mRNA.

1.7.3. Established roles of lncRNAs in PCa development and progression

Numerous lncRNAs exhibit differential expression in PCa, and some have been reported to regulate key functions in cells including proliferation, invasion and metastasis, induction of angiogenesis, apoptosis, and androgen signaling. Despite this, mechanistic insights on the functional roles of a large proportion of lncRNAs remain unknown. Below are examples of prominent PCa-associated lncRNAs and their elucidated functions. Comprehensive reviews of lncRNAs implicated in PCa development and progression are available elsewhere [175, 176].

1.7.3.1. Prostate Cancer Antigen 3 (PCA3)

PCA3 is a clinically relevant PCa biomarker overexpressed in at least 95% of primary tumors [177]. Shown to be involved in epithelial-to-mesenchymal transition (EMT), *PCA3* silencing in LNCaP cells increased the expression of epithelial markers *E-cadherin*, *Claudin-3*, and *CK18*, and decreased the mesenchymal marker vimentin [178]. Moreover, *PCA3* knockdown resulted in reduced AR signaling, as reflected by downregulation of AR target genes *PSA* and *PCGEM1*, and overall mitigation of cell growth and viability [178, 179].

1.7.3.2. C-Terminal Binding Protein 1 Antisense (CTBP1-AS)

CTBP1-AS is an androgen responsive lncRNA transcribed in the antisense direction of *CTBP1*, which encodes an AR corepressor. *CTBP1-AS* is upregulated in prostate cancers, enhancing hormone-dependent and castration-resistant tumor growth. The nuclear-enriched lncRNA interacts with PSF, a transcriptional repressor, and recruits the repressive HDAC-Sin3A complex to the *CTBP1* promoter resulting in histone deacetylation. *CTBP1-AS* also guides the PSF complex in *trans* to mediate histone deacetylation at regulatory regions of androgen-repressed genes, including cell cycle regulators *TP53*and *SMAD3* [180].

1.7.3.3. Prostate Cancer Associated Transcript 1 (PCAT1)

PCAT1 was first described to be upregulated in a subset of metastatic and high-grade localized PCa [81]. Lentiviral *PCAT1* overexpression in RWPE-1 benign immortalized prostate cells resulted in increased cell proliferation. In contrast, *PCAT1* knockdown reduced the proliferative capacity of the cells. Expression profiling analysis also revealed perturbed cell cycle and mitotic pathways to be the most perturbed after *PCAT1* knockdown [81]. Interestingly, reduced homologous recombination (HR) efficiency and consequent increased sensitivity to PARP inhibitors was observed in PCa cell lines expressing *PCAT1*. Subsequently, *PCAT1* was characterized to be involved in regulating the expression of *BRCA2*, a DNA repair pathway gene crucial for mediating homologous repair [181].

1.7.3.4. Second Chromosome Locus Associated with Prostate 1 (SChLAP1)

SChLAP1 was found to be overexpressed in approximately 25% of a PCa cohort with combined primary and metastatic disease. The lncRNA was more frequently expressed in metastatic samples,

and correlated with ETS fusion events [164]. Gain-of function and loss-of-function of *SChLAP1* both *in vitro* and *in vivo* revealed the role of the lncRNA in cell invasion and metastasis. At the molecular level, the nuclear-enriched SChLAP1 associates with and antagonizes the chromatin remodeling and tumor-suppressive SWI/SNF complex [164].

1.7.3.5. Prostate Cancer Gene Expression Marker 1 (PCGEM1)

The lncRNA *PCGEM1* was found to be upregulated in high risk PCa patients [182, 183]. RNA immunoprecipitation revealed that *PCGEM1* binds to AR at a specific post-translational modification site, and this binding enhances ligand-dependent and ligand-independent AR transcriptional programs, ultimately leading to increased cell proliferation [184].

1.7.3.6. Androgen Receptor Regulated Long Non-coding RNA 1 (ARLNC1)

ARLNC1 was initially identified to be an AR target through DHT stimulation experiments performed in both androgen-dependent LNCaP and VCaP cell lines [130]. Furthermore, *ARLNC1* upregulation was observed in both a localized PCa cohort (n=500) and a metastatic cohort (n=100) compared with benign prostate samples. Interestingly, *ARLNC1* knockdown resulted in significant repression of AR target genes, as well as both AR mRNA and protein levels, demonstrating a positive feedback loop between the lncRNA and AR signaling [130].

1.7.4. Technologies to dissect lncRNA function

Upon identification of lncRNA for further investigation, establishment of genetic models demonstrating phenotypic changes upon perturbation of candidate transcript levels is critical in initiating its functional dissection [185]. Equally important is the elucidation of molecular processes the lncRNA is involved in. To address this, a number of technologies have been developed to identify RNA interactors—whether protein, chromatin, or other RNA molecules—and these methods can be classified as either protein-centric or RNA-centric. Fundamentally, protein-centric methods rely on purifying a protein to pulldown interacting RNA molecules which would then be identified through quantitative PCR or high-throughput sequencing. Conversely, RNA-centric methods employ tagged oligonucleotide probes antisense to the lncRNA of interest to isolate the transcript and its associating factors. The RNA interactome can then be determined through immunoblotting, mass spectrometry, qPCR, and high-throughput DNA sequencing. RNA-centric methods are utilized to identify novel RNA-binding proteins [186].

1.7.4.1. Protein-centric methods

1.7.4.1.1. RNA Immunoprecipitation (RIP)

RNA immunoprecipitation is arguably the most commonly used method for RNA-protein interaction analysis. Initially developed for native purification, RIP enables capture of complexes at physiological conditions. Since native protein conformations are preserved, RIP can identify both direct and indirect

binding partners. Binding strength between the captured RNA and RNA binding protein (RBP) is quantified through quantitative PCR as enrichment of the RNA by target-specific immunoprecipitation, normalized to a control immunoprecipitation. Modifications to the method include the use of ultraviolet (UV) light or chemical crosslinkers to preserve transient RNA-protein complexes and to perform stringent washing steps and minimize false-positives [187]. RIP can also be followed by high-throughput sequencing to obtain a transcriptome-wide view of protein-RNA interaction [185].

1.7.4.1.2. Crosslinked Immuniprecipitation (CLIP)

In crosslinked immunoprecipitation, intact cells are irradiated by ultraviolet (UV) light to preserve RNA-protein complexes [188]. As a result, only RNAs directly bound to the protein are captured. Total RNA is partially digested by RNase A and the RNA-binding protein (RBP) is immunoprecipitated together with covalently crosslinked RNA. Because of the crosslinking step, strong washings can be performed to remove non-specifically bound RNA. Variations of the method include (i) high-throughput sequencing of RNA isolated by CLIP (HITS-CLIP), wherein 3' RNA adapters are ligated to captured RNA to enable reverse transcription and subsequent DNA library preparation and sequencing [189]; (ii) photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP), wherein cells are cultured in media containing nucleotide analogs 4-thiouridine (4-SU) or 6-thioguanosine (6-SG) which, upon incorporation into synthesized transcripts, can form crosslinks upon exposure to UV light [190]; and (iii) individual nucleotide resolution CLIP (iCLIP), wherein modified reverse transcription primers are used to circularize cDNA, and subsequent cleavage through the barcoded adaptor region can determine the exact position of RNA-protein interaction [191].

1.7.4.2. RNA-centric methods

1.7.4.2.1. Capture Hybrid Analysis of RNA Targets (CHART)

CHART was first developed to map the genomic occupancy of *roX2*, a 600 nt lncRNA involved in dosage compensation in *Drosophila* [192]. In this method, RNA interactions are preserved by formaldehyde-crosslinking. Capture probes are designed by identifying high-accessibility regions of crosslinked lncRNA through RNase H digestion. Briefly, 20-mer synthetic DNA oligonucleotides are mixed with sheared chromatin lysate in the presence of RNase H which digests RNA-DNA hybrids. Biotinylated antisense capture oligos are designed after identification of RNase H-sensitivity sites by qPCR. CHART enrichment is then performed by hybridizing the chromatin lysate with custom-made probes, followed by bead capture of lncRNA-chromatin complex. After crosslink reversal by proteinase K digestion, DNA isolation is performed, followed by library preparation and deep sequencing. CHART has since been successful in mapping the chromatin binding sites of *MALAT1* and *NEAT1* in human cells [193].

1.7.4.2.2. Chromatin Isolation by RNA Purification (ChIRP)

ChIRP has a similar workflow as CHART. The difference between the methods lies in the design of capture probes and the choice of crosslinker. Whereas CHART utilizes formaldehyde crosslinking and requires RNase H mapping of accessibility sites on the target lncRNA, ChIRP employs glutaraldehyde crosslinking and tiling 20-mer DNA oligos that span the non-repetitive regions of the target transcript. The chromatin occupancy of the human telomerase RNA *TERC* has been mapped through this method (ChIRP-seq) [194]. Additionally, ChIRP has also been utilized to identify the *Xist* proteome by coupling mass spectrometry downstream of the lncRNA-protein capture (ChIRP-MS) [195].

1.7.4.2.3. RNA Antisense Purification (RAP)

RAP is distinct from CHART and ChIRP by its use of long capture biotinylated probes—at least 60 nt in length—which results in very specific and stable RNA-DNA hybrids [196]. The choice of crosslinking agent depends on which molecular interaction is being investigated. Psoralens are the most suitable crosslinking reagents for identifying RNA-RNA interactions using RAP [197]. On the other hand, formaldehyde and UV crosslinking are used to analyze nucleic acid and protein interactions. RAP in tandem with deep sequencing (RAP-seq) was used to model the spreading of *Xist* across the inactive X chromosome upon initiation of X chromosome inactivation (XCI) [128]. Direct interaction of *Xist* with SHARP, resulting in HDAC3 activation and RNA polymerase II exclusion across the X-chromosome was identified through RAP-MS [127]. Through RAP, the lncRNA *Firre* whose gene locus escapes XCI, was found to localize to chromatin in *cis* and *trans*, associating with genes implicated in adipogenesis [198].

1.8. Aim and scope of the thesis

Although numerous non-coding transcripts have been reported to correlate with PCa development and progression, there remains a paucity of lncRNAs with well elucidated functional roles. Accordingly, the aim of this thesis was to identify and functionally characterize PCa-associated lncRNAs. To this end, the transcriptome dataset of the International Cancer Genome Consortium-Early Onset Prostate Cancer (ICGC-EOPC) cohort was leveraged to identify previously uncharacterized non-coding transcripts differentially expressed in PCa tumors compared to normal prostatic tissue [80].

PCa-associated lncRNAs were shortlisted and the top candidate was functionally characterized *in vitro*. Cellular processes and pathways dependent on the transcript were determined through gene expression arrays. Moreover, the gene regulatory mechanism driving the lncRNA expression in PCa cells was investigated. To further specify the molecular underpinnings of lncRNA function, the transcript interactome was identified by establishing and applying the RNA-centric technology ChIRP in tandem with high throughput sequencing and mass spectrometry. Relevant lncRNA-protein interactions were validated using parallel RNA-protein binding assessment methods such as RNA immunoprecipitation and affinity purification. Integrative analysis of lncRNA-dependent molecular

processes and protein interactome was performed to generate a mechanistic model of lncRNA function in PCa cells, and concomitantly provide novel insights into PCa and lncRNA biology.

2. Materials

2.1. Materials for cell culture

Table 2-1. Cell lines.

Cell line	Origin	Supplier
DU-145	Prostate carcinoma, derived from brain metastasis	American Type Culture Collection (ATCC)
LNCaP #126 empty	Prostate carcinoma, derived from lymph node metastasis; stably transfected with an empty expression cassette	[97]
LNCaP #126 T/E III	Prostate carcinoma, derived from lymph node metastasis; stably transfected with a tet-inducible *TMPRSS2:ERGa* expression cassette	[97]
LNCaP clone FGC	Prostate carcinoma, derived from lymph node metastasis	ATCC
PC-3	Prostate carcinoma, derived from bone metastasis	ATCC
RWPE-1	Normal prostate epithelium	ATCC
VCaP	Prostate carcinoma, derived from vertebral metastasis	ATCC

Table 2-2. Cell culture media and supplements.

Materials/Media/Reagents	Supplier
Doxycycline	Sigma-Aldrich
Dulbecco Modified Eagle Medium (DMEM)	Gibco™ Thermo Fisher Scientific
Dulbecco's Phosphate-Buffered Saline (DPBS)	Gibco™ Thermo Fisher Scientific
Fetal Bovine Serum (FBS)	Gibco™ Thermo Fisher Scientific
Hygromycin B	Gibco™ Thermo Fisher Scientific
Kaighn's Modification of Ham's F-12 Medium (F-12K Medium)	ATCC
Keratinocyte Serum-Free Medium (Keratinocyte-SFM) **Bovine Pituitary Extract (BPE)** **Epidermal Growth Factor (EGF)**	Gibco™ Thermo Fisher Scientific
Opti-MEM™ I Reduced Serum Media	Gibco™ Thermo Fisher Scientific
Roswell Park Memorial Institute 1640 Medium (RPMI 1640 Medium)	Gibco™ Thermo Fisher Scientific
Tet System Approved FBS (Tet-FBS)	Clontech
Trypsin-EDTA (0.25%), phenol red	Gibco™ Thermo Fisher Scientific

2.2. Materials for molecular biology

Table 2-3. Vectors.

Vector	Application	Supplier
pAAVsi2	*Renilla* reporter vector	[199]
pcDNA™3.1(+)	Mammalian expression vector	Invitrogen™ Thermo Fisher Scientific
pCR®2.1-TOPO®	Cloning vector	Invitrogen™ Thermo Fisher Scientific
pCR®4Blunt-TOPO®	Cloning vector	Invitrogen™ Thermo Fisher Scientific
pGL4.10[*luc2*]	Firefly *luc* reporter vector	Promega

Table 2-4. Antibodies used for western blot (WB).

Target protein	Source	Isotype	Supplier	Application	Working dilution
GAPDH	Rabbit	IgG	#2118, Cell Signaling Technology	1° WB	1:1000
H3	Rabbit	polyclonal	#9715, Cell Signaling Technology	1° WB	1:1000
HPRT1	Rabbit	IgG	ab109021, Abcam	1° WB	1:5000
FOXO1	Rabbit	IgG	#2880, Cell Signaling Technology	1° WB	1:1000
FOXO3a	Rabbit	IgG	#2497, Cell Signaling Technology	1° WB	1:1000
14-3-3ε	Rabbit	polyclonal	#9635, Cell Signaling Technology	1° WB	1:1000
Rabbit IgG	Goat		#7074, Cell Signaling Technology	2° WB	1:10000

Table 2-5. Antibodies used for chromatin immunoprecipitation (ChIP).

Target protein	Source	Isotype	Supplier	Working dilution
ERG	Rabbit	IgG	ab92513, Abcam	2 µg/IP
Rabbit IgG	Rabbit		ab172730, Abcam	2 µg/IP

Table 2-6. Antibodies used for RNA immunoprecipitation (RIP).

Target	Source	Isotype	Supplier	Working dilution
14-3-3ε	Rabbit	polyclonal	#9635, Cell Signaling Technology	3 µg/IP
14-3-3ζ	Rabbit	IgG	#7413, Cell Signaling Technology	3 µg/IP
Rabbit IgG	Rabbit		#2729, Cell Signaling Technology	3 µg/IP

Table 2-7. Sequences of short interfering RNAs (siRNAs).

Target gene	Designation	Target sequence	Working concentration	Supplier
LINC00920	siRNA-Q2	CAGGGCTTGGGAGATAAGACA	10 nM (PC-3) 35 nM (VCaP)	Qiagen
	siRNA-Q2	CTGGCCATTCCTTAAGCTGAA	10 nM (PC-3) 35 nM (VCaP)	Qiagen
	siRNA-D1	AGTAAGAACTATAAGGCTA	35 nM (PC-3)	Dharmacon
	siRNA-D3	CCACAGAGTTGAATGAATT	35 nM (PC-3)	Dharmacon
	siRNA-D4	GGAAAGGCCTATAGACACA	35 nM (PC-3)	Dharmacon
ERG	siERG-7	CAGATCCTACGCTATGGAGTA	50 nM (VCaP)	Qiagen
	siERG-8	CTCCACGGTTAATGCATGCTA	50 nM (VCaP)	Qiagen
ETV4	siETV4-1	ATGGGCTATGGCTATGAGAAA	10 nM (PC-3)	Qiagen
	siETV4-8	CCGCTCGCTCCGATACTATTA	10 nM (PC-3)	Qiagen
Scrambled control	NTC	proprietary	as above	Qiagen

Table 2-8. Primer oligonucleotides used for full-length lncRNA and promoter amplification.

Target gene	Product	Forward sequence (5'→3')	Reverse sequence (5'→3')	Supplier
LINC00920	Full-length cDNA	AATCTTCACAGGGAAG GAAGCAACAAAA	TAGAATTTACATTTTAA TAATTCTGAGACT	Sigma-Aldrich
LINC00844	Full-length cDNA	GTAGAGACAAAGGAAA CACAGAGACATA	TAGACAGACAATTCAA GCAATTTATTGT	Sigma-Aldrich
LINC01082	Full-length cDNA	AAATTGGTCCCAGTTTT CACCCTGC	TTTTCTGTTTGAGACAT ATTAAACAAGCT	Sigma-Aldrich
LINC00920	Full-length promoter	AGTAGATATCCATCTTC AGGTTATGA	TTCCCTGTGAAGATTCA CTTCCTGCC	Sigma-Aldrich

Table 2-9. Primer oligonucleotides used for quantitative PCR (qPCR).

Target	Application*	Forward sequence (5'→3')	Reverse sequence (5'→3')	Supplier
RP11-867G23.3	UPL#42	GGAAGAGCGACAC TCACGAT	GGACTTCAAGATCCG AACCA	Sigma-Aldrich
LINC00920	UPL#19	CCTGCCACACTCAA GTGGA	CCCGTGTGATGGAAG AACTC	Sigma-Aldrich
RP11-3P17.5	UPL#51	CACTGCCTTCTTGG CCTTTA	GGACCCTTTTCACAAC ATGG	Sigma-Aldrich
RP11-395L14.4	UPL#41	GCGTTGGGAACAA TATGGAA	ATGGGAACTGCGTGA ATACC	Sigma-Aldrich
SNHG18	UPL#6	CACATCCCTAAGCT GCCATC	CAGATACCCGGCTTTC CTTT	Sigma-Aldrich
LINC00844	UPL#88	GGTTTGGCTGGACT GTGAGT	CTTTCTGATTTCAATG TTCTCTGC	Sigma-Aldrich
LINC01082	UPL#57	CCACCCTGCAAGT GAGAAG	GGTTGTTTTTCTTCGG TGCT	Sigma-Aldrich
NEAT1	UPL#42	AGTGAATGTGCAC CCTTGG	AACAAACCACGGTCC ATGA	Sigma-Aldrich
MALAT1	UPL#71	GACCCTTCACCCCT CACC	TTATGGATCATGCCCA CAAG	Sigma-Aldrich
HPRT1	UPL#73	TGACCTTGATTTAT TTTGCATACC	CGAGCAAGACGTTCA GTCCT	Sigma-Aldrich

GAPDH	UPL#60	AGCCACATCGCTCAGACAC	GCCCAATACGACCAAATCC	Sigma-Aldrich
ERG	UPL#64	GGTTAATGCATGCTAGAAACACA	AGATGGTTGAGCAGCTTTCG	Sigma-Aldrich
ETV4	UPL#83	TTATGAGAAAGGCATCATGCAG	CGGGCTCACACACAAACTT	Sigma-Aldrich
BCL2L11	UPL#86	ACGGCCTATTCTCAGAGGATTAT	AAACTAAGGCAGCTTTTTAAGTTAGC	Sigma-Aldrich
PMAIP1	UPL#28	ACAGCAACAACAACAATGCAC	CCACGAGGAACAAGTGCAA	Sigma-Aldrich
GADD45A	UPL#70	GCCAAGCTGCTCAACGTC	AGCCACATCTCTGTCGTCGT	Sigma-Aldrich
FOXO3	UPL#22	CCAGCCTAACCAGGGAAGTT	AGCCCATGTTGCTGACAGA	Sigma-Aldrich
FOXO3	UPL#62	TCTGAATGATGGGCTGACTG	CTAGAGCTCCGCTGCATGA	Sigma-Aldrich
LINC00920	SYBR	AGGACATCTGAAGCTAAACATGGATC	AATTCATTCAACTCTGTGGTCTTGGAA	Sigma-Aldrich
LINC00844	SYBR	TGGCAGAATTGGGATCTGACT	AACTGGACATTGCAAACACTTT	Sigma-Aldrich
Non-genic chr 12	ChIP, ChIRP	CTGTCCCTGGTCAAGAGTGACTTCCCT	ACAGAGTCAAAAACTGCAAGGCTGC	Sigma-Aldrich
LINC00290#1	ChIP	TCACAGGGAAGGAAGCAACA	GCCCAATGCCCTGTGTCTAT	Sigma-Aldrich
LINC00290#2	ChIP	GGGGATGTTTAATGTTGTTACGC	TTCACTTCCTGCCCAGAGTC	Sigma-Aldrich
LINC00290#3	ChIP	TGTTTAATGTTGTTACGCAGGAA	AGATTCACTTCCTGCCCAGA	Sigma-Aldrich
HEXIM1	ChIRP	TTTATTGGGGTGCTCCGCTT	GCAATCTGGGGAGCTCAAGT	Sigma-Aldrich
PS2	ChIRP	AGTGAGAGATGGCCGGAAAA	TCATGAGCTCCTTCCCTTCC	Sigma-Aldrich
PNN	ChIRP	CGTGGATCGGAAGAGAAGGG	CTGTCGGACCGGGAATTCTT	Sigma-Aldrich
RNF40	ChIRP	CCGCACATGGTTAGGAGGTT	TCCCGATCTGTGCATTCGAG	Sigma-Aldrich
LINC00920#1	RIP	TCACAGGGAAGGAAGCAACA	GCCCAATGCCCTGTGTCTAT	Sigma-Aldrich
LINC00920#2	RIP	GAGCACTACATAAAGCAGCCA	CCCCGGATGACTTTCACTCT	Sigma-Aldrich
LINC00920#3	RIP	GCAGACACAGCACTAAGAACT	TCTTATCTCCCAAGCCCTGC	Sigma-Aldrich
LINC00920#4	RIP	AGGACATCTGAAGCTAAACATGGATC	AATTCATTCAACTCTGTGGTCTTGGAA	Sigma-Aldrich
LINC00920#5	RIP	AAAGCCTGCGAGAGAGAGAG	AGAAAAGGTGACCAAGGTGAC	Sigma-Aldrich
LINC00920#6	RIP	ACTGGCTGGAGGAGTAAGAAC	CAATCAACCCTTACCTTCCAGT	Sigma-Aldrich
LINC00920#7	RIP	ACTGCAACCATTTTCTAACTCTTAA	CAAGTAAGCTACATTATGTTCCTGT	Sigma-Aldrich
LINC00920#8	RIP	TTTATCTGCCTTGTCGATACTCT	AGAGCCAGTGACAAAGGAAGA	Sigma-Aldrich
LINC00920#9	RIP	TGTCACTGGCTCTGTAAATTTGA	TCTTGCAAAAGACAAAGGGTTT	Sigma-Aldrich
LINC00920#10	RIP	TGGATCAAGAGAGACAAAGTGT	TGTAGTTTTCAGCATACAGGTCC	Sigma-Aldrich
SNORA55	RIP	GTGGGGACAGATGGTGCTAC	CCCCAAGACAAATGGAAAAC	Sigma-Aldrich

*UPL: Universal Probe Library assay; SYBR: SYBR green assay; ChIP: Chromatin immunoprecipitation; ChIRP: Chromatin Isolation by RNA Purification; RIP: RNA immunoprecipitation.

Table 2-10. Primer oligonucleotides used for site-directed mutagenesis of the *LINC00920* promoter.

Target region	Application*	Forward sequence (5'→3')	Reverse sequence (5'→3')	Supplier
ETS domain 1	SDM	AGTAGATATCCATC TTCAGGTTATGA	TTCCCTGTGAAGATTCA CTTGGTGCCCAGAGTC	Sigma-Aldrich
ETS domain 2	SDM OE-PCR#1	AGTAGATATCCATC TTCAGGTTATGA	CAGCACAGCTTGGTGCG TAACAACA	Sigma-Aldrich
ETS domain 2	SDM OE-PCR#2	TGTTGTTACGCACC AAGCTGTGCTG	TTCCCTGTGAAGATTCA CTTCCTGCC	Sigma-Aldrich
ETS domain 2	SDM OE-PCR#3	AGTAGATATCCATC TTCAGGTTATGA	TTCCCTGTGAAGATTCA CTTCCTGCC	Sigma-Aldrich

*SDM: Site-directed mutagenesis; OE: overlap extension. Mutagenic bases are underlined.

Table 2-11. Primer oligonucleotides used for rapid amplification of cDNA ends (RACE).

Primer	Application	Forward sequence (5'→3')	Reverse sequence (5'→3')	Supplier
GeneRacer™ 5′ Primer	5' RACE	CGACTGGAGCACGA GGACACTGA	N/A	Invitrogen™ Thermo Fisher Scientific
GeneRacer™ 5′ Nested Primer	5' RACE	GGACACTGACATGG ACTGAAGGAGTA	N/A	Invitrogen™ Thermo Fisher Scientific
GeneRacer™ 3′ Primer	3' RACE	N/A	GCTGTCAACGATAC GCTACGTAACG	Invitrogen™ Thermo Fisher Scientific
GeneRacer™ 3′ Nested Primer	3' RACE	N/A	CGCTACGTAACGGC ATGACAGTG	Invitrogen™ Thermo Fisher Scientific
RP11-867G23.3	5' GSP RACE	N/A	GCCTGGGCAACAAG AGCAAAACTCA	Sigma-Aldrich
RP11-867G23.3 nested	5' GSP RACE	N/A	ATTGGGGCTTGGTGG TTCGGAGAC	Sigma-Aldrich
LINC00920	5' GSP RACE	N/A	CCCGTGTGATGGAA GAACTCTAAGATG	Sigma-Aldrich
LINC00920 nested	5' GSP RACE	N/A	GTTTTGCTTCAGGGC TGTTGTCACC	Sigma-Aldrich
RP11-3P17.5	5' GSP RACE	N/A	TTTTCACAACATGGC GCCGAAAG	Sigma-Aldrich
RP11-3P17.5 nested	5' GSP RACE	N/A	GGCATATGTTCGACT GGCTCCTGAT	Sigma-Aldrich
RP11-395L14.4	5' GSP RACE	N/A	TCTTTATGTTGAAGA GAATGGCTAAAAA	Sigma-Aldrich
RP11-395L14.4 nested	5' GSP RACE	N/A	ATGGGAACTGCGTG AATACCATTCT	Sigma-Aldrich
SNHG18	5' GSP RACE	N/A	GTTAGGTGAGGTCC AGGTCATGCTG	Sigma-Aldrich
SNHG18 nested	5' GSP RACE	N/A	ACCACAGATACCCG GCTTTCCTTTG	Sigma-Aldrich
LINC00844	5' GSP RACE	N/A	GTGAGTCAGATCCC AATTCTGCC	Sigma-Aldrich
LINC00844 nested	5' GSP RACE	N/A	AGCAAGGAGGTTTC TTGCATGGCTAA	Sigma-Aldrich
LINC01082	5' GSP RACE	N/A	CTGGTTGTTTTTCTT CGGTGCTG	Sigma-Aldrich

LINC01082 nested	5’ GSP RACE	N/A	ACATTCCTCGCATTC CTGACGGTTG	Sigma-Aldrich
RP11-867G23.3	3’ GSP RACE	TCCGAACCACCAAG CCCCAATTCCCAGC	N/A	Sigma-Aldrich
RP11-867G23.3 nested	3’ GSP RACE	GCTTGGCAGAGAAG GCCCCAGAAGT	N/A	Sigma-Aldrich
LINC00920	3’ GSP RACE	GGCCTCCCCAACATG CTCACCTGCT	N/A	Sigma-Aldrich
LINC00920 nested	3’ GSP RACE	TGCCCAACTCATCTG GATCTTCCTTTG	N/A	Sigma-Aldrich
RP11-3P17.5	3’ GSP RACE	TTGGCAACATCCAA AGCATCGTAAT	N/A	Sigma-Aldrich
RP11-3P17.5 nested	3’ GSP RACE	GGGTGACGTGCGGA TCTTCTTCTTT	N/A	Sigma-Aldrich
RP11-395L14.4	3’ GSP RACE	GCCGGGCTGAAGAA AAGAAGAATGG	N/A	Sigma-Aldrich
RP11-395L14.4 nested	3’ GSP RACE	GGGGTGAGAGGAAT GGGGAAATGTT	N/A	Sigma-Aldrich
SNHG18	3’ GSP RACE	TGTGGGCCATGAGT GACCTTCAAAG	N/A	Sigma-Aldrich
SNHG18 nested	3’ GSP RACE	CCACCTCACAGCCA AGTTCAAGGAA	N/A	Sigma-Aldrich
LINC00844	3’ GSP RACE	CCCAATTCTGCCATA CTGTTTCTGGTTC	N/A	Sigma-Aldrich
LINC00844 nested	3’ GSP RACE	GGCAGAATTGGGAT CTGACTCAC	N/A	Sigma-Aldrich
LINC01082	3’ GSP RACE	TCTATCGAGGCACAC AGACAGACCA	N/A	Sigma-Aldrich
LINC01082 nested	3’ GSP RACE	CAGCCTGAAATGAA GCCGGGATCAA	N/A	Sigma-Aldrich

Table 2-12. Antisense biotinylated oligonucleotides* used for chromatin isolation by RNA purification (ChIRP).

No.	lacZ pool (5'→3')	MALAT1 pool (5'→3')	LINC00920 pool (5'→3')
1	TGAATCCGTAATCATGGTCA	GCTTAAGAGGGCAGGAGAGG	TGTTGCTTCCTTCCCTGTGA
2	CGCTATTACGCCAGCTGGCG	GAGCTTCAGACCTTCTGAAC	AGGCCTTTCCCTGCTCAGCC
3	CTCAGGAAGATCGCACTCCA	AGTGGCCCACTCTGATCTGC	TCTAGGGAGGGCTGTTCTAG
4	CATCGTAACCGTGCATCTGC	TCAGGGCTTTACTTTCCATT	CACTCTTCCCTATGCTTTGC
5	ATAATTCGCGTCTGGCCTTC	ATTCGATCACCTTCCGCCGC	TGTTGTCACCAAGTTCTTAG
6	GACGGCAAACGACTGTCCTG	CATGCTACTCTTCTAAGTCT	GTCTTATCTCCCAAGCCCTG
7	TCCAGATAACTGCCGTCACT	TCACCTTCGGTTTAATCTCT	CACTTTGATTGTAGCTTTCA
8	ATCATCATTAAAGCGAGTGG	TTCCCGTACTTCTGTCTTCC	TGGAAGTGGGTCCTTCTCTA
9	GATAATTTCACCGCCGAAAG	TCTACGTAAACACCCTCATC	GCCCTCAGTTCAACGGGCTG
10	AGTTTCGGGTTTTCGACGTT	AATGCTAGTCCTCAGGATTT	TGGACTCTCTCTCTCGCAGG
11	CGACATCGCAGGCTTCTGCT	TGTGGTTGCCAAGCCAAGCC	CAGGATGTCACTTCAGTGAT
12	TGACGGTTAACGCCTCGAAT	GATTCATGAGTATAAGCCTG	CTTGGGACTCATTTATAATG
13	GTGTACCACAGCGGATGGTT	GCCTCAGTTACACATCCAAA	CTCCAGCCCTGTGTAATCCC
14	GTTCGCTCATCGCCGGTAGC	CTGTTAAGACCATCCCAAAA	TGAGAATTCCCGTGTGATGG
15	GATTAGCGCCGTGGCCTGAT	TTTGGCCTACTCAAGCTCTT	AGGCCTTCAGCTTAAGGAAT
16	GCAAATAATATCGGTGGCCG	TGCCCAAATTAATGCACTGG	GACAAACCCTGGGTTTATTA
17	TGGGCGTATTCGCAAAGGAT	GAAATCCCTTCAGGATCATT	TATAGTTCTTACTCCTCCAG
18	CAGACGAAGCCGCCCTGTAA	ATCATACTGCCAGGCTGGTT	TCAATCAACCCTTACCTTCC
19	TCGTTCGGCGTATCGCCAAA	GACATTGCCTCTTCATTGTA	CCTGTATATTTATGTTGGGA
20	TTTGCCCGGATAAACGGAAC	GAACTCCACAGCTCTTAAAA	GAGCCTGGAATATTTGTATA
21	GGAGCTCGTTATCGCTATGA	TCTGATTCTAACAGCACATC	ATGAGAAGCCTTTGGGAAAA
22	CATGCGGTCGCGTTCGGTTG	CCATGTGCCTGGAATTATTA	CTTGAAAGAGGGAATGATTT
23	TGCCAACGCTTATTACCCAG	AAGGATGAAATGCCTCTGCA	TATCGACAAGGCAGATAAAA
24	AGCGGTGCACGGGTGAACTG	TCTAATAGCAGCGGGATCAG	GAGTTGGGCAAATTAACAAA
25	GTCAATGCGGGTCGCTTCAC	AAGACTGTTGCTTGTTTGGA	ACAGAGCCAGTGACAAAGGA
26	CACGCGTGAGCGGTCGTAAT	ACTAGTGGTTCCCAATCCCC	CAAAGGGTTTCATAAAGTTG
27	GGTAATCGCCATTTGACCAC	CTTAGGATAATAGCGCTTTG	CACTTACTGTAGAATGCTTA
28	TTGCGGCCCTAATCCGAGCC	GGCGATGTGGCAGAGAAGTT	TCTCTTGATCCATCACTCAT
29	TTCGCTCGGGAAGACGTACG	GTTCCCACCCAGCATTACAG	GTTAGATGGTAACCAAGAAT
30	TGTTGACTGTAGCGGCTGAT	GTCCTGGAAACCAGGAGTGC	TTCTGAGACTTCCAATTCAT

*Manufactured by siTOOLs Biotech GmbH

Table 2-13. Index primers used for ChIRP-seq DNA library preparation.

ChIRP-seq library	Index primer	Index primer sequence (5'→3')	Index primer sequence read	Supplier
P28 lacZ	#5	CAAGCAGAAGACGGCATACGAGATC<u>ACTGTG</u>GTGACTGGAGTTCAGACGTGT GCTCTTCCGATC-s*-T	ACAGTG	New England Biolabs
P28 LINC00920	#6	CAAGCAGAAGACGGCATACGAGAT<u>A</u><u>TTGGC</u>GTGACTGGAGTTCAGACGTGT GCTCTTCCGATC-s*-T	GCCAAT	New England Biolabs
P29 lacZ	#10	CAAGCAGAAGACGGCATACGAGAT<u>A</u><u>AGCTA</u>GTGACTGGAGTTCAGACGTGT GCTCTTCCGATC-s*-T	TAGCTT	New England Biolabs
P29 LINC00920	#12	CAAGCAGAAGACGGCATACGAGAT<u>T</u><u>ACAAG</u>GTGACTGGAGTTCAGACGTGT GCTCTTCCGATC-s*-T	CTTGTA	New England Biolabs
P30 lacZ	#2	CAAGCAGAAGACGGCATACGAGAT<u>A</u><u>CATCG</u>GTGACTGGAGTTCAGACGTGT GCTCTTCCGATC-s*-T	CGATGT	New England Biolabs
P30 LINC00920	#4	CAAGCAGAAGACGGCATACGAGAT<u>T</u><u>GGTCA</u>GTGACTGGAGTTCAGACGTGT GCTCTTCCGATC-s*-T	TGACCA	New England Biolabs

*Phosphorothioate bond

Table 2-14. Enzymes.

Enzyme/Master Mix	Supplier
ABsolute Blue qPCR Master Mix (2X)	Thermo Scientific™ Thermo Fisher Scientific
ABsolute qPCR Master Mix (2X)	Thermo Scientific™ Thermo Fisher Scientific
Benzonase nuclease	Merck Millipore
CutSmart® Buffer (10x)	New England Biolabs
DNase I (RNase-Free)	Qiagen
DNase I (RNase-free)	New England Biolabs
DreamTaq Green PCR Master Mix (2X)	Thermo Scientific™ Thermo Fisher Scientific
NheI-HF	New England Biolabs
NotI-HF	New England Biolabs
Proteinase K	Invitrogen™ Thermo Fisher Scientific
RNase A	Qiagen
RNase H	New England Biolabs
SpeI-HF	New England Biolabs
T4 DNA Ligase	New England Biolabs
T7 RNA Polymerase	New England Biolabs
Taq DNA Polymerase	New England Biolabs
XbaI	New England Biolabs
XhoI	New England Biolabs

Table 2-15. Reagents, chemicals, and materials.

Reagent/Material	Supplier
1,4-Dithiothreitol (DTT)	Carl Roth GmbH
4-20% Mini-PROTEAN® TGX™ precast protein gel	Bio-Rad
Absolute ethanol	Fisher Scientific GmbH
Acetone	Fisher Scientific GmbH
Agar	Carl Roth GmbH
Agarose	Fisher Scientific GmbH
Agencourt AMPure XP beads	Beckman Coulter
Alkaline Phosphatase (Calf Intestinal)	New England Biolabs
Ampicillin	AppliChem
Biotin RNA labeling mix	Roche
Bovine Serum Albumin (BSA)	Sigma-Aldrich
Cell Proliferation Reagent WST-1	Roche
ChIP-grade protein G magnetic beads	Cell Signaling Technology
cOmplete, Mini protein inhibitor cocktail	Roche
cOmplete, Mini protein inhibitor cocktail (EDTA-free)	Roche
Crystal violet solution (1% aqueous)	Sigma-Aldrich
Dimethyl pimelimidate (DMP)	Thermo Scientific™ Thermo Fisher Scientific
Dimethyl sulfoxide (DMSO)	AppliChem
DNA Gel Loading Dye (6X)	Thermo Scientific™ Thermo Fisher Scientific
Dynabeads™ MyOne™ Streptavidin C1	Invitrogen™ Thermo Fisher Scientific

Dynabeads™ Protein G magnetic beads	Invitrogen™ Thermo Fisher Scientific
Dynabeads™ M-270 Streptavidin beads	Invitrogen™ Thermo Fisher Scientific
***E. coli* poly(A) polymerase**	New England Biolabs
EDTA (Disodium salt)	Gerbu Biotechnik GmbH
Ethanolamine	Merck Millipore
Ethidium bromide (0.5% solution)	Carl Roth GmbH
Formaldehyde (16% w/v), Methanol-free	Thermo Scientific™ Thermo Fisher Scientific
Formaldehyde (37% w/w)	Fisher Scientific GmbH
Formamide	Carl Roth GmbH
GeneRuler DNA 1 kb Ladder	Thermo Scientific™ Thermo Fisher Scientific
GeneRuler DNA 100 bp Plus Ladder	Thermo Scientific™ Thermo Fisher Scientific
GeneRuler DNA Ladder Mix	Thermo Scientific™ Thermo Fisher Scientific
Glutaraldehyde (25%)	Sigma-Aldrich
Glycerol (87%)	VWR
Glycine	Gerbu Biotechnik GmbH
GlycoBlue coprecipitant	Invitrogen™ Thermo Fisher Scientific
HEPES	Sigma-Aldrich
Human Prostate Total RNA (Lot # 0903001)	Invitrogen™ Thermo Fisher Scientific
Ipatasertib (GDC-0068)	Target Molecule
KCl	Carl Roth GmbH
LiCl	Sigma-Aldrich
Lipofectamine™ 2000	Invitrogen™ Thermo Fisher Scientific
Lipofectamine™ RNAiMAX	Invitrogen™ Thermo Fisher Scientific
Methanol	Fisher Scientific GmbH
$MgCl_2 \cdot 6H_2O$	Sigma-Aldrich
MOPS	Gerbu Biotechnik GmbH
Na_3VO_4	Sigma-Aldrich
NaCl	Fisher Scientific GmbH
NaF	Carl Roth GmbH
$NaHCO_3$	Fisher Scientific GmbH
NEBNext® Multiplex Oligos for Illumina® (Index Primer Set 1)	New England Biolabs
N-lauroylsarcosine sodium salt (20% solution)	Sigma-Aldrich
NP-40 (Igepal® CA-630)	Sigma-Aldrich
Nuclease-Free Water	Invitrogen™ Thermo Fisher Scientific
One Shot™ TOP10 chemically competent *E. coli* cells	Invitrogen™ Thermo Fisher Scientific
PageRuler Prestained Protein Ladder	Thermo Scientific™ Thermo Fisher Scientific
Phenol/chloroform/isoamyl alcohol 125:24:1 mixture pH 4.3	Fisher Scientific GmbH
Phenylmethane sulfonyl fluoride (PMSF)	Cell Signaling Technology
PhosSTOP phosphatase inhibitor cocktail	Roche
Phusion® High-Fidelity DNA Polymerase	New England Biolabs
Premix Ex Taq™ master mix	Takara Bio

Recombinant human 14-3-3ε protein	Abcam
RNAse-away	Carl Roth GmbH
Roti®-Load 1 protein loading buffer	Carl Roth GmbH
S.O.C medium	Invitrogen™ Thermo Fisher Scientific
Saline-sodium citrate buffer (SSC buffer, 20X)	Sigma-Aldrich
SC79	Sigma-Aldrich
Sodium acetate	AppliChem
Sodium deoxycholate	Sigma-Aldrich
Sodium dodecyl sulfate	Fisher Scientific GmbH
ssRNA ladder	New England Biolabs
Sucrose	Sigma-Aldrich
Superase In™ RNase inhibitor	Invitrogen™ Thermo Fisher Scientific
SuperSignal™ West Dura Extended Duration Substrate	Thermo Scientific™ Thermo Fisher Scientific
SYBR® Gold nucleic acid gel stain	Invitrogen™ Thermo Fisher Scientific
Trichloroacetic acid (TCA)	Sigma-Aldrich
Triethanolamine	Sigma-Aldrich
Tris base	Sigma-Aldrich
Tris(2-carboxyethyl)phosphine (TCEP, 0.5 M)	Sigma-Aldrich
Triton X-100	Sigma-Aldrich
Tryptone	Carl Roth GmbH
Tween® 20	Sigma-Aldrich
UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v)	Invitrogen™ Thermo Fisher Scientific
UltraPure TAE Buffer (10X)	Invitrogen™ Thermo Fisher Scientific
Universal Probe Library (UPL)	Roche
Yeast extract	Gerbu Biotechnik GmbH

Table 2-16. Kits.

Kit	Supplier
Dual-Glo® Luciferase Assay Kit	Promega
GeneRacer® Kit with SuperScript® III RT and Zero Blunt® TOPO® PCR Cloning Kit for Sequencing	Invitrogen™ Thermo Fisher Scientific
High Pure PCR Template Preparation Kit	Roche
High Sensitivity DNA Kit	Agilent Technologies
miRNeasy® Mini Kit	Qiagen
NEBNext Ultra II DNA Library Preparation Kit for Illumina	New England Biolabs
BCA Protein Assay Kit	Pierce™ Thermo Fisher Scientific
Silver Stain Kit	Pierce™ Thermo Fisher Scientific
QIAprep Spin Maxiprep Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
Qubit® dsDNA HS assay Kit	Invitrogen™ Thermo Fisher Scientific
RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Scientific™ Thermo Fisher Scientific
RNA 6000 Nano Kit	Agilent Technologies
RNeasy® Mini Kit	Qiagen

2.3. General laboratory materials and equipment.

Table 2-17. Consumables.

Consumable	Supplier
Bioruptor® microtubes (1.5 mL)	Diagenode
Black, flat- and clear-bottomed 96-well plate	Perkin-Elmer
Cell culture dish (150 mm)	TPP
Cell culture flask (175 cm²)	TPP
Cell culture flask (25 cm²)	TPP
Cell culture flask (75 cm²)	TPP
Countess Cell Counting Chamber Slides	Invitrogen™ Thermo Fisher Scientific
Cryovials (1.8-mL)	Neolab
DNA LoBind tubes (1.5 mL)	Eppendorf
Falcon™ round bottom tubes (14-mL)	Fisher Scientific GmbH
Falcon™ cell culture dish (12-well)	Fisher Scientific GmbH
Falcon™ cell culture dish (24-well)	Fisher Scientific GmbH
Falcon™ cell culture dish (6-well)	Fisher Scientific GmbH
Falcon™ cell scrapers	Fisher Scientific GmbH
Falcon™ conical tubes (50-mL, 15-mL)	Fisher Scientific GmbH
Filter tips (1250-µL, 200-µL, 20-µL, 10-µL)	Neptune Scientific
LightCycler® 480 qPCR plate (384-well)	Roche
LightCycler® 480 qPCR plate sealing foil	Roche
Microflex™ XCEED™ Powder-Free Nitrile Examination Gloves	Fisher Scientific

PCR tubes (200 µL)	Starlab
Safe-Lock tubes (5-mL, 2-mL, 1.5-mL, 0.5-mL)	Eppendorf
Serological pipettes (50-mL, 25-mL, 10-mL, 5-mL, 2-mL)	Corning
ThinCert™ well insert (8-µm)	Greiner Bio-One

Table 2-18. Laboratory equipment.

Equipment	Supplier
2100 Bioanalyzer	Agilent Technologies
Axiovert 40 CFL inverted microscope	Carl Zeiss
Bacterial incubator	Infors HAT
Bioruptor® Pico sonication device	Diagenode
ChemiDoc™ XRS+	Bio-Rad
Countess II FL Automated Cell Counter	Invitrogen™ Thermo Fisher Scientific
Electronic micropipette (200 µL)	Sartorius
HERA Safe cell culture hood	Thermo Scientific
Heracell™ VIOS 160i CO_2 incubator	Thermo Scientific
Heraeus™ Fresco 21 refrigerated microcentrifuge (2-/1.5-mL)	Thermo Scientific
Heraeus™ Megafuge™ 16 cell culture centrifuge	Thermo Scientific
Hybaid Maxi 14 hybridization oven	Thermo Scientific
Infinite M200 microplate reader	TECAN
LightCycler® 480 II	Roche
MagnaRack (1.5-mL)	Life Technologies
Magnetic stand (96-well)	Ambion
Micropipettes (1000-µL, 200-µL, 20-µL, 10-µL, 2-µL)	Gilson
Mini-PROTEAN®	Bio-Rad
Mr. Frosty cryobox	Nalgene
NanoDrop 1000	Thermo Fisher
Pipette controller (Pipetboy)	Integra biosciences
PTC-225 gradient thermal cycler	MJ Research
QUANTUM UV transilluminator	Vilber Lourmat
Qubit® Fluorometer	Life Technologies
Refrigerated centrifuge (50-/15-mL)	Sigma
Speed Vac DNA 120 vacuum concentrator	Thermo Scientific
ThermoMixer® Comfort heat block	Eppendorf
Trans-Blot Turbo	Bio-Rad
Vortex Genie	Scientific Industries

2.4. Data analysis

Table 2-19. Programs and softwares.

Software	Reference/Supplier
Coding Potential Assessment Tool Version 2.0.0	[200]
Coding Potential Calculator Version 2.0	[201]
Genomic Regions Enrichment of Annotations Tool (GREAT) Version 3.0.0	[202]
GSEAPreranked Desktop Application Version 3.0	[203]
i-control™ Version 1.6	TECAN
Image Lab™ 6.0	Bio-Rad
ImageJ 1.52h	[204]
Ingenuity Pathway Analysis (IPA) Build 486617M	Qiagen
LightCycler® 480 Software Release 1.5.0	Roche
OpenCFU Version 3.9.0	[205]
Quantum Capt UV Imaging System Version 15.10	Vilber Lourmat
Rstudio Version 1.0.136	[206]
SnapGene® Viewer Version 3.3.4	GSL Biotech

3. Methods

3.1. Cell culture

3.1.1. Propagation and maintenance

PC-3 cells were cultured in F-12K medium supplemented with 10% fetal bovine serum (FBS) at 37^{o}C in a humidified incubator with 5% CO_2. VCaP cells were cultured in Dulbecco Modified Eagle Medium (DMEM) with 10% FBS. RWPE-1 cells were cultured in Keratinocyte serum-free medium supplemented with 0.05 mg/mL bovine pituitary extract (BPE), and 5 ng/mL epidermal growth factor (EGF). DU-145 and parental LNCaP cells were cultured in RPMI 1640 with 10% FBS. Tet-inducible LNCaP #126 cell clones were cultured in RPMI 1640 supplemented with 10% tet-free FBS, and 80 µg/mL Hygromycin B [97]. Cells were maintained until 70-80% confluency and then split into new culture flasks. Briefly, attached cells were washed with 1X phosphate-buffered saline (PBS) and chemically detached using 0.25% trypsin-EDTA. Trypsin inactivation was performed by adding 2X volume of complete medium to the cell suspension. Cells were centrifuged at 400 g, room temperature for 2 min and the pellet was resuspended in complete growth medium. The resulting single-cell suspension was seeded into new flasks at the desired density.

Cells were routinely tested for mycoplasma contamination and were authenticated by single nucleotide polymorphisms (SNP)-profiling (Multiplexion GmbH).

3.1.2. Cryopreservation

Cells were trypsinized from culture flasks and washed twice with 1X PBS. For PC-3, VCaP, LNCaP, and DU-145, cells were resuspended in their respective complete growth media supplemented with 5% (v/v) cell culture-grade dimethyl sulfoxide (DMSO) at a concentration of 1 x 10^6 cells/mL. For RWPE-1, cells were resuspended in complete growth medium supplemented with 10% (v/v) DMSO and 15% FBS. One million cells were aliquoted in cryovials and stored short-term in a freezing container at -80^{o}C. Frozen cryovials were subsequently transferred into liquid nitrogen for long-term storage.

3.1.3. Gene knockdown via short interfering RNAs (siRNAs)

Cells were seeded to reach at least 60% confluency 24 h prior to transfection. On the day of transfection, maintenance medium was refreshed. For a 6-well plate format, Lipofectamine® RNAiMAX:siRNA dilutions were prepared as follows. First, 8 µL RNAiMAX was diluted with 66.6 µL Opti-MEM™ I Reduced Serum Media. For 10 nM siRNA transfections, 2.9 µL of 10 µM siRNA was similarly diluted with 66.6 µL Opti-MEM™. The RNAiMAX and siRNA dilutions were combined and mixed throroughly and incubated for 5 min at room temperature. One hundred twenty microliters of RNAiMAX:siRNA dilution was then added to the cells of one well. From these values, siRNA transfection reactions were scaled up or down depending on the surface area of the culture vessel and the siRNA concentration determined for optimal knockdown efficiency. Unless otherwise

stated, cells were harvested for RNA and/or protein isolation 48 h post transfection, or were used for subsequent functional assays. **Table 2-7** lists the siRNA sequences used in this study, and their optimal working concentrations.

3.1.4. Plasmid transfection in cell lines

Cells were seeded to reach at least 80% confluency 24 h prior to transfection. On the day of transfection, maintenance medium was refreshed. For a 6-well plate format, Lipofectamine™ 2000:plasmid DNA dilutions were prepared as follows. For each well, 5.76 µL Lipofectamine ™ 2000 was diluted in Opti-MEM™ to a final volume of 144 µL. Next, 720 ng plasmid was similarly diluted with Opti-MEM™ to a final volume of 144 µL. Both dilutions were mixed and incubated at room temperature for 5 min. Two hundred microliters of Lipofectamine™ 2000:DNA dilution was added to the cells. After 6 h, cell medium was refreshed and cells were maintained for 48-72 h. Lipofectamine™ 2000 and DNA amounts were scaled accordingly depending on the culture format.

For luciferase assays, cells in each 96-well were transfected in the same manner as described except with 0.35 µL Lipofectamine™ 2000 and 100 ng total plasmid DNA, diluted to 50 µL with Opti-MEM™ each. One hundred microliter mix was applied to each well and the cells were incubated for 48 h before proceeding to doxycycline induction.

3.1.5. Functional assays

3.1.5.1. Cell proliferation

Cells were trypsinized 24 h post transfection from the culture vessel, washed, and resuspended in complete medium at a concentration of 5 x 10^4 cells/mL. One hundred microliters of cell suspension corresponding to 5000 cells were seeded per well into 3 96-well plates (corresponding to 48-, 72-, and 96-h time points) in triplicate and incubated at 37^{o}C in a humidified incubator with 5% CO_2. To quantify cell proliferation, untreated, pre-counted cells were seeded in triplicate to generate a standard curve. Ten microliters of pre-warmed cell proliferation reagent WST-1 was added to all wells. Signals were developed for 1-3 h in the cell culture incubator, until a red to yellow color shift was observed. The absorbance value of each well at 440 nm was measured using a TECAN Infinite M200 reader with the i-control™ (version 1.6) software. Cell counts per well were calculated from the generated standard curve.

3.1.5.2. Colony formation

Transfected cells were seeded into duplicate wells in a 6-well plate format at a density of 1.3 x 10^3 cells/3 mL of complete medium. The plates were maintained for 9 days at which point the cell colonies were stained with crystal violet. Briefly, cells were washed with ice-cold 1X PBS and fixed on ice with 100% methanol for 30 min. The cells were washed with 1X PBS and afterwards stained with 0.005% crystal violet at room temperature for 1 h. The staining solution was removed and the

cells were washed again. Cell colonies were air-dried and imaged using the ChemiDoc™ XRS+ system with Image Lab™ software. Colony quantification was performed using OpenCFU [205].

3.1.5.3. *Cell migration*

Transfected cells were washed and resuspended in serum-free medium at a concentration of 5 x 10^5 cells/mL. Meanwhile, 700 µL of complete growth medium was added in wells of a 24-well plate. A 8-µm ThinCert™ for 24-well plates was then placed in each well to complete the migration assembly (**Figure 3-1**). Two hundred microliters of serum-free cell suspension, corresponding to 1 x 10^5 cells, was seeded in duplicate migration assemblies. The plates were then incubated for 24 h. To quantify cell migration, cells at the bottom of the migration inserts were either stained with crystal violet or trypsinized for WST-1 analysis. To stain cells, the insert was emptied of medium and the bottom was submerged in 100% methanol for 30 min and washed 4 times in deionized water. The interior of the insert was cleaned with cotton swabs. Each insert bottom was then submerged in 0.05% crystal violet for 30 min and washed twice with deionized water. The inserts were air-dried overnight. Stained cells were manually counted using a light microscope in three fields of view per insert. Alternatively, migratory cells were quantified using a colorimetric method. Briefly, the media within the insert was aspirated and the insert was placed in a well containing 500 µL of 0.25% trypsin-EDTA. Trypsinization was performed at 37°C for 10 min. The insert was agitated to release the loosened cells from the bottom of the membrane. Trypsin was quenched by adding an equal volume of complete medium and the detached cells were collected by centrifugation at 400 g for 5 min. The cell pellet was resuspended in 100 µL complete medium and seeded into a 96-well plate. The WST-1 method described above was then performed to determine the cell number per well.

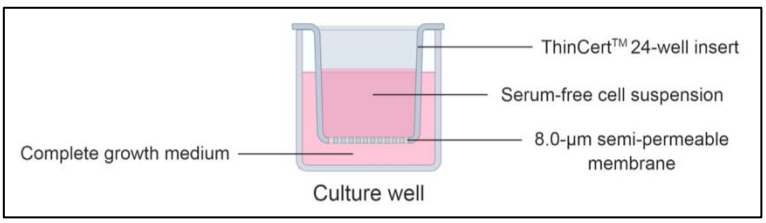


Figure 3-1. Schematic of the culture well assembly for cell migration assay.

3.1.6. Tet-induction of LNCaP #126 cells

ERG overexpression was induced in LNCaP #126 clones by shifting the medium to RPMI 1640 supplemented with 10% tet-free FBS (Clontech), and 50 ng/mL doxycycline as described [97]. Cells were harvested or further treated 48 h post induction.

3.1.7. Ipatasertib treatment of transfected PC-3 cells

Forty-eight hours post siRNA transfection, cells were treated with ipatasertib (GDC-0068) by shifting the culture medium to that supplemented with 300 nM of inhibitor [207]. Cells were harvested after 24 h for RNA isolation.

3.2. General molecular biology techniques

3.2.1. Genomic DNA isolation

For cell line contamination and authentication tests, genomic DNA was isolated using High Pure PCR Template Preparation Kit according to the manufacturer's protocol. Template DNA for *LINC00920* promoter amplification was isolated from PC-3 cells using the same procedure.

3.2.2. RNA extraction and cDNA synthesis

Unless otherwise specified, RNA from cell lines was isolated using RNeasy® Mini Kit following the manufacturer's protocol including the optional DNase-I digestion step. Prior to RNA extraction, cells were washed twice with 1X PBS. RNA from each column was eluted in 30 µL nuclease-free water, quantified by NanoDrop 1000, and stored at -80^{o}C until use. Complementary DNA was synthesized from 2000 ng RNA input using RevertAid H Minus First Strand cDNA Synthesis Kit. Briefly, 1 µL of 100 µM random hexamer primer was added to 2000 ng RNA and resulting volume was filled up to 12 µL with nuclease-free water. The reaction mixture was incubated at 65^{o}C for 5 min and placed on ice for 1-2 min. Four microliters of 5X reaction buffer, 1µL of 20 U/µL RiboLock RNase inhibitor, 2 µL of 10 mM dNTP mix, and 1 µL of 200 U/µL RevertAid H Minus M-MuLV Reverse Transcriptase were added to the reaction mixture which was then incubated at 42^{o}C for 1 h. The reaction was terminated by heating at 70^{o}C for 5 min, and the synthesized cDNA was stored at -20^{o}C until further use.

3.2.3. Protein isolation and quantification

Cells from one well of a 6-well plate were washed twice with 1 mL 1X PBS and resuspended in 60 µL 1X RIPA buffer (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate [SDS], 50 mM Tris pH 8.0) supplemented with 1X cOmplete, Mini protein inhibitor, and 1X PhosSTOP phosphatase inhibitor cocktails. Cell lysis was facilitated by incubation on ice for 45 min. The resulting protein lysate was clarified by centrifugation at 12000 rpm, 4^{o}C for 15 min. The supernatant was transferred to a clean tube and stored at -80^{o}C until use. Protein quantification was performed using the microplate procedure of the Pierce^{TM} BCA Protein Assay Kit.

3.2.4. SDS-polyacrylamide gel electrophoresis

Prior to electrophoresis, 20 ng of protein was denatured at 95^{o}C in 1X Roti®-Load 1 protein loading buffer for 5 min. The samples, together with PageRuler^{TM} prestained protein ladder, were loaded into 4-20% Mini-PROTEAN® TGX^{TM} precast protein gels. Gel electrophoresis was performed through 1X

Laemmli running buffer (0.025 M Tris, 0.192 M glycine, 0.1% SDS) at 200V using the Mini-PROTEAN® system.

3.2.5. Western blotting

Protein transfer onto polyvinylidene difluoride (PVDF) membrane was facilitated using the Trans-Blot Turbo system (Bio-Rad). Briefly, transfer stacks were saturated with 1X transfer buffer (5X Trans-Blot® Turbo™ transfer buffer: 100% ethanol: deionized water, 3:1:1) and the PVDF membrane was activated in 100% ethanol prior to assembly of the blotting sandwich. One transfer stack was placed on the bottom (+) cassette, followed by the PVDF membrane, the mini polyacrylamide gel, and another transfer stack. The top (-) cassette was secured over the assembly and protein transfer was conducted using the MIXED MW blotting program (1.5 V, 7 min). The blotted membrane was then blocked in 10 mL of blocking buffer (5% bovine serum albumin in 1X PBS-T) at room temperature for 1 h. The membrane was transferred into the primary antibody solution and incubated with rotation at 4^{o}C overnight. Next day, the blot was washed for 5-, 10-, and 15- min, respectively, with the washing buffer (1X PBS-T) and incubated with the appropriate secondary antibody solution at room temperature for 1 h. The secondary antibody is conjugated to horseradish peroxidase (HRP) for chemiluminescent detection. Excess antibody was removed by washing the membrane for 5-, 10-, and 15- min, respectively, with the washing buffer. When probing for phosphorylated proteins, blocking and washing buffers as well as antibody solutions were supplemented with 50 mM NaF and 1 mM Na_3VO_4.

Chemiluminescent immunoblot signals were developed using the SuperSignal™ West Dura Extended Duration Substrate and imaged by the ChemiDoc™ XRS+ system with Image Lab™ software. Band intensities were quantified using Fiji [204]. **Table 2-4** lists the antibodies used in this study and their respective working dilutions.

3.2.6. Silver staining

To visualize resolved proteins, polyacrylamide gels were silver-stained using the Pierce™ Silver Stain kit. Briefly, the gel was washed twice with deionized water for 5 min and then fixed with 30% ethanol: 10% acetic acid solution twice for 15 min. The gel was transferred into a proprietary sensitizer solution for 1 min and washed twice with deionized water. The gel was incubated with the stain working solution (<0.25 % w/w $AgNO_3$) for 30 min and afterwards washed twice with deionized water. Protein bands were visualized in the developer solution (1-5% w/w Na_2CO_3). Upon reaching the desired band intensity, the reaction was stopped with 5% acetic acid.

3.2.7. Agarose gel electrophoresis

Agarose gels were casted by heating and dissolving an appropriate amount of agarose in 1X TAE buffer. Prior to casting, 10 µL of 0.5% ethidium bromide solution was added to the molten agarose. Sufficient DNA loading dye stock was combined with approximately 100 ng of DNA to make a 1X

loading dye:sample solution. The samples, together with a DNA ladder mix, were loaded into the agarose gel submerged in 1X TAE. Gel electrophoresis was performed at 100 V and the gel was imaged on a UV transilluminator.

3.2.8. PCR amplification of LINC00920 cDNA and LINC00920 promoter fragments

Phusion® High-Fidelity DNA polymerase was used to amplify the full-length *LINC00920* cDNA and promoter fragments. Each reaction mixture consisted of 1X Phusion HF buffer, 0.2 mM dNTPs, 1 µM of each forward and reverse primers (**Table 2-8**), 3% v/v DMSO, 0.2 µL of Phusion DNA polymerase, 20 ng of cDNA or genomic DNA template and enough nuclease-free water to reach the final volume of 20 µL. Temperature cycling was performed on the PTC-225 gradient thermal cycler. Template DNA was initially denatured at 98^{o}C for 2.5 min, followed by 30 cycles of denaturation (30 s at 98^{o}C), annealing (30 s at 55^{o}C) and extension (2 min at 72^{o}C). The final extension step was extended to 10 min. Successful amplification of PCR products were verified by electrophoresing 2-µL aliquots through 1% agarose- 1X Tris-Acetate-EDTA (TAE) gels. Amplicons were visualized by staining with ethidium bromide and viewing under UV light. Bands with the expected fragment size were cut from the gel and purified using QIAquick gel extraction kit using the manufacturer's protocol. After elution in nuclease-free water, DNA was quantified by NanoDrop 1000, and stored at -20^{o}C.

3.2.9. Quantitative PCR

Quantitative PCR was performed in the LightCycler® 480 II system using assays described below adapted to a 384-well plate format. Crossing point-PCR-cycle (Cp) values were generated by the LightCycler® 480 software (release 1.5.0) using the second derivative method. Relative expression levels were calculated by obtaining the difference of the median Cp value of triplicate reactions corresponding to the housekeeping gene and the gene of interest (ΔCp). Next, the difference in ΔCp values of the control and treated samples was computed ($\Delta\Delta Cp$), and normalized by calculating $2^{-\Delta\Delta Cp}$, respectively. **Table 2-9** lists the qPCR primers used in this study.

3.2.9.1. Universal Probe Library (UPL) assay

Each qPCR reaction consisted of 10 µL 2X ABsolute qPCR master mix or 2X Premix Ex Taq™ master mix, 0.2 µL of 10 µM Universal Probe Library probe, 0.4 µL of 20 µM forward and reverse primer pool, 5 µL of 4ng/µL cDNA template, and nuclease-free water to reach 20 µL. Template DNA was initially denatured at 95^{o}C for 10 min, followed by 45 cycles of denaturation (10 s at 95^{o}C), annealing (30 s at 55^{o}C) and extension (1 s at 72^{o}C). The reaction was cooled down to 40^{o}C indefinitely.

3.2.9.2. SYBR assay

Each qPCR reaction consisted of 5 µL 2X ABsolute Blue qPCR master mix, 0.5 µL 20 µM forward and reverse primer pool, 5 µL 4ng/µL cDNA template, and nuclease-free water to reach 11 µL.

Template DNA was initially denatured at 95°C for 15 min, followed by 40 cycles of denaturation (15 s at 95°C), annealing (30 s at 60°C) and extension (60 s at 72°C). Following the amplification cycling, a melt curve analysis was performed by heating the reaction to 95°C for 30 s followed by cooling to 60°C for 30 s. A final denaturation step was held at 97°C wherein the fluorescence signals at 5°C-temperature increments were continuously acquired. The reaction was cooled down to 40°C indefinitely.

3.2.10. Cloning

3.2.10.1. TOPO insertion

TOPO technology was used to clone: (i) amplified lncRNA cDNA into pCR®4Blunt-TOPO®; and (ii) *LINC00920* promoter fragments into pCR®2.1-TOPO®. *LINC00920* promoter fragments were first A-tailed prior to cloning. To this end, a 25-µL reaction volume consisting of 10 µL of purified PCR product, 1X ThermoPol® buffer, 200 µM ATP, 0.5 U Taq DNA polymerase, and nuclease-free water was incubated at 72°C for 20 min. DNA insertion into both TOPO vectors was facilitated by mixing 5 ng TOPO vector, 0.5 µL salt solution (1.2 M NaCl, 0.06 M $MgCl_2$), and enough purified PCR product to reach a 3-µL reaction volume. The TOPO reaction was incubated at room temperature for 30 min and stored at 4°C until use.

3.2.10.2. Sticky-end ligation

Subcloning by sticky-end ligation was performed to shuttle: (i) full-length lncRNA cDNA from pCR®4Blunt-TOPO® into pcDNA™3.1(+); and (ii) *LINC00920* promoter fragments from pCR®2.1-TOPO® into pGL4.10[*luc2*]. **Table 3-1** lists the enzyme combinations for digesting the donor and acceptor vectors.

Table 3-1. Vectors and restriction enzymes used for subcloning.

Vector	Description	Restriction enzyme pair
pCR®4Blunt-TOPO® constructs	Harbors full-length lncRNA cDNA	SpeI*, NotI
pcDNA™3.1(+)	Mammalian expression vector	NheI*, NotI
pCR®2.1-TOPO® constructs	Harbors *LINC00920* promoter fragments	SpeI*, XhoI
pGL4.10[*luc2*]	Firefly luciferase reporter vector	NheI*, XhoI

*These restriction enzymes generate compatible cohesive ends.

Each restriction enzyme double digestion reaction contains 1X CutSmart buffer, 20 U of each enzyme, 1.5 µg plasmid DNA, and enough nuclease-free water to reach a 50-µL volume. The samples were incubated at 37°C for 30 min and afterwards deactivated by heating to 80°C for 20 min. Linearized acceptor vectors were dephosphorylated by adding 5 µL 10X CutSmart buffer, 5 µL 10 U/µL alkaline phosphatase, and 40 µL nuclease-free water to the digestion reaction. The mixture was incubated at

37°C for 30 min and then heated to 65°C for 5 min to terminate the reaction. DNA digests were resolved in 1% agarose gels and gel-purified.

DNA inserts were ligated into acceptor vectors by mixing 1X T4 DNA ligase buffer, 75 ng linearized and dephosphorylated acceptor vector, 25 ng DNA insert, 400 U T4 DNA ligase, and enough nuclease-free water to reach a 20-µL reaction volume. The samples were incubated overnight at 16°C and inactivated by heating to 65°C for 10 min.

3.2.10.3. Bacterial transformation and colony screening

One Shot™ TOP10 chemically competent *E. coli* cells were thawed on ice. Three microliters of ligation reaction was added into 30 µL competent cells and the mixture was incubated on ice for 30 min. Cells were heated to 42°C for 30 s and chilled for 1 min. Two hundred fifty microliters of pre-warmed S.O.C. medium was added into the cells which were allowed to recover by incubation with shaking (200 rpm) at 37°C for 1 h. Transformed cells were plated at 1:5 and 1:2 dilutions onto LB/ampicillin/X-gal plates and incubated overnight at 37°C. White colonies were further screened for the correct insert sequence by colony PCR. Briefly, colonies were separately inoculated into 10-µL volume PCR mixture containing 1X DreamTaq Green PCR master mix, 0.2 µM each of forward and reverse primers, and nuclease-free water. Template DNA from cells was initially denatured at 95°C for 5 min and 30 s, followed by 30 cycles of denaturation (30 s at 98°C), annealing (30 s at 55°C) and extension (60 s at 72°C). The final extension step was extended to 10 min. Successful amplification of PCR products were verified by electrophoresing 2-µL aliquots through 1% agarose- 1X TAE gels.

3.2.10.4. Plasmid extraction

Insert-positive colonies were inoculated into 3 mL Luria-Bertani (LB) broth (1% tryptone, 0.5% yeast extract, 0.17 M NaCl) supplemented with 100 µg/mL ampicillin and grown overnight with shaking at 37°C. Plasmid extraction was performed using QIAprep Spin Miniprep kit following the manufacturer's protocol. Plasmid DNA was resuspended in nuclease-free water and stored at -20°C. Insert sequence fidelity was assessed by Sanger sequencing (GATC, Eurofins Genomics). Scaled-up plasmid production was performed by inoculating positive transformants into 100-mL culture medium. Plasmid DNA was prepared using QIAprep Spin Maxiprep kit.

3.2.10.5. Preparation of glycerol stocks

Glycerol stocks were prepared from bacterial clones by mixing an equal volume of 87% sterile glycerol to 250-µL culture aliquots. Stocks were stored long-term at -80°C.

3.2.11. Site-directed mutagenesis by overlap-extension PCR

The 1000 bp *LINC00920* promoter amplified from PC-3-derived genomic DNA was cloned into pCR®2.1-TOPO®. To introduce the GG>CC transversions in ETS domain 1, a mutagenic reverse primer was paired with a wild-type forward primer (**Table 2-10**) in amplifying the 1000 bp promoter

region using the pCR®2.1-TOPO® construct as template. In contrast, overlap extension-PCR was performed to introduce the double nucleotide mutations in ETS domain 2 which is further upstream the *LINC00920* TSS. Intermediate fragments were amplified using Phusion® High-Fidelity DNA polymerase as described in section 3.2.8 with primers listed in **Table 2-10**. Equal amounts of the intermediates were then used as template for the third round of PCR which fused the two mutant fragments, yielding the full-length promoter region. Mutants were then cloned into pCR®2.1-TOPO® and sticky cloning ends were generated by restriction enzyme digestion in preparation for subcloning into pGL4.10[*luc2*].

3.3. Molecular characterization of lncRNAs

3.3.1. Rapid amplification of cDNA ends (RACE)

3.3.1.1. 5'-RACE

RACE experiments were performed using the GeneRacer™ Kit. 5'-RACE was initiated by dephosphorylation of total human prostate RNA. Two thousand nanograms of RNA was mixed with 1X calf intestinal phosphatase, 40 U RNaseOut™, 1 U calf intestinal phosphatase, and nuclease-free water to a final volume of 10 µL. The sample was incubated at 50°C for 1 h, and then cooled on ice. Next, 90 µL nuclease-free water and 100 µL phenol:chloroform:isoamyl alcohol (25:24:1) were added to precipitate RNA. After mixing, the sample was centrifuged at 21,000 g, room temperature for 5 min. The aqueous phase was transferred into a new tube where 2 µL 10 mg/mL mussel glycogen, 10 µL 3 M pH 5.2 sodium acetate, and 220 µL 95% ethanol were added. The sample was cooled on dry-ice for 10 min and afterwards centrifuged at 21,000 g, 4°C for 20 min. The RNA pellet was washed once with 70% ethanol and resuspended in 7 µL nuclease-free water. The mRNA cap structure was then removed by incubating the dephosphorylated RNA at 37°C for 1 hour with 1X tobacco acid pyrophosphatase buffer, 40 U RNaseOut™, and 0.5 U tobacco acid pyrophosphatase. RNA was precipitated by phenol:chloroform:isoamyl alcohol as described above. To prime 5'-ends, an RNA oligo ligation step was performed by mixing decapped RNA with 0.25 µg GeneRacer™ oligo, 1X T4 RNA ligase buffer, 1 mM ATP, 40 U RNaseOut™, and 5 U T4 RNA ligase. The sample was incubated at 37°C for 1 h, and RNA was once again precipitated. Purified RNA was reverse-transcribed as described in section 3.2.2, and 5'-ends of lncRNAs were amplified as described in section 3.2.8 using GeneRacer™ 5' primer and a gene-specific 5'-end reverse primer (**Table 2-11**). The resulting PCR product was used in a subsequent nested PCR using the GeneRacer™ 5' nested primer and a more upstream gene-specific 5'-end reverse primer.

3.3.1.2. 3'-RACE

Total human prostate RNA was reverse-transcribed by mixing 2000 ng RNA with 900 ng GeneRacer™ Oligo dT Primer, 10 nmol dNTP mix, and nuclease-free water to a final volume of 13 µL. The sample was incubated at 65°C for 5 min and chilled on ice for 1 min. Next, 1X first strand

buffer, 0.1 µmol DTT, 40 U RNaseOut™, and 200 U SuperScript™ III reverse transcriptase were added to the denatured RNA. Reverse-transcription was facilitated at 50°C for 1 h, and afterwards terminated by heating to 70°C for 15 min. The sample was chilled on ice and RNA digestion was carried out by incubating the sample with 2 U RNase H at 37°C for 20 min. Gene-specific amplification of 3'-ends was performed as described in section 3.2.8 using a gene-specific 3'-end forward primer and the GeneRacer™ 3' primer (**Table 2-11**). As with 5'-RACE, a nested PCR was subsequently performed using a more downstream gene-specific 3'-end forward primer and the GeneRacer™ 3' nested primer.

3.3.1.3. Cloning and sequence analysis

Successful amplification of 5'- and 3'-ends were verified by electrophoresing 2-µL aliquots through 1% agarose- 1X TAE gels. Amplicons were visualized by staining with ethidium bromide and viewing under UV light. DNA bands were cut from the gel and purified using QIAquick gel extraction kit using the manufacturer's protocol, and cloned into pCR™4Blunt-TOPO® as described in section 3.2.10.1. Sequences of lncRNA ends were determined by Sanger sequencing (GATC, Eurofins Genomics) and aligned with annotated tracks using the UCSC genome browser.

3.3.2. Coding potential analysis

LINC00920 sequence as determined by RACE was used to survey the coding potential of the transcript. Sequences of control coding mRNAs, lncRNAs, and *LINC00920* were loaded into Coding Potential Assessment Tool (CPAT; Version 2.0.0, URL: http://lilab.research.bcm.edu/cpat/) and the hg19 assembly was selected as reference genome. Sequences were similarly loaded into Coding Potential Calculator (CPC; Version 2.0 beta, URL: http://cpc2.cbi.pku.edu.cn/). Hg19 PhyloCSF tracks were extracted from the Track Data Hubs and visualized in the UCSC genome browser spanning the chromosomal locus of *LINC00290*.

3.3.3. Subcellular fractionation

A fractionation protocol [208] was adapted to prepare cytoplasmic, nucleoplasmic, and chromatin lysates from PC-3, VCaP, and LNCaP cells for eventual RNA extraction. Thirty million cells were resuspended in 400 µL ice-cold cytoplasmic buffer (0.15% NP-40, 10 mM Tris-HCl pH 7.5, 150 mM mM NaCl) and incubated on ice for 10 min. Five hundred microliters of ice-cold sucrose buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.7 M sucrose) was layered on the cell suspension and the tubes were centrifuged at 13000 rpm, 4°C for 10 min. The supernatant corresponding to the cytoplasmic fraction was collected, leaving approximately 100 µL liquid with the pellet to minimize nuclear contamination. The remaining nuclear pellet was resuspended in 250 µL ice-cold glycerol buffer (20 mM Tris-HCl pH 7.9, 75 mM NaCl, 0.5 mM EDTA pH 8.0, 50% glycerol, 0.85 mM DTT) and an equal volume of ice-cold nuclei lysis buffer (20 mM HEPES pH 7.6, 7.5 mM $MgCl_2$, 0.2 mM EDTA pH 8.0, 0.3 M NaCl, 1 M urea, 1% NP-40, 1 mM DTT) was added. After mixing, the samples were

incubated on ice for 1 min. The soluble nucleoplasmic and chromatin fractions were separated by centrifugation at 14000 rpm, 4^{o}C for 2 min. In collecting the nucleoplasmic fraction, approximately 100 µL liquid was left with the chromatin pellet to minimize cross-contamination. The chromatin pellet was resuspended in 50 µL ice-cold 1X PBS. RNA from all subcellular fractions was extracted using RNeasy® Mini Kit as described in section 3.2.2, followed by cDNA synthesis. Relative enrichments of *HPRT1*, *GAPDH*, *LINC00920*, *NEAT1*, and *MALAT1* transcripts in each fraction were measured by qPCR.

3.4. Gene expression profiling

3.4.1. RNA preparation and quality assessment

Three biological replicates of *LINC00920* knockdown in PC-3 cells using siRNA-Q2, siRNA-Q3, and scrambled control (

Table 2-7) were performed in a 6-well plate format. Forty-eight hours post transfection, RNA was extracted as described in section 3.2.2 and quantified by NanoDrop 1000. RNA quality was assessed using the RNA 6000 Nano kit with the 2100 Bioanalyzer. RNA samples were diluted to 50 ng/µL in 20-µL volumes and submitted to the Genomics and Proteomics Core Facility of the DKFZ where gene expression profiling was performed using the Human HT-12 v4 Expression Bead Chip from Illumina.

3.4.2. Gene Set Enrichment Analysis (GSEA)

Gene Set Enrichment Analysis tool was used to identify perturbed biological pathways upon lncRNA knockdown. Expression fold-change values of all analyzed genes generated from the biological replicates of microarray experiments were used to prepare *.rnk* files for each siRNA knockdown condition. Using *.rnk* files as input, enrichment analysis of canonical pathways and gene ontology gene sets was performed using GSEAPreranked (desktop application version 3.0) with permutations value set to 1000.

3.4.3. Ingenuity Pathway Analysis (IPA)

Top 1000 upregulated and top 1000 downregulated genes upon *LINC00920* knockdown were overlapped for the two independent siRNA experiments. For siRNA-Q2, this cutoff translated to genes having fold-change values greater than 1.223946 or less than 0.8010429. Likewise, for siRNA-Q3 the cut-off included genes with fold-change values greater than 1.349623 or less than 0.735197. Expression values of the common genes in both siRNA knockdowns were derived from the siRNA-Q3 knockdown dataset. Altogether, 315 genes (**Supplementary Table 9-2**) and their corresponding expression fold-change values were loaded into the IPA (build 486617M) core analysis tool.

3.5. Validation of *LINC00920* regulation by ERG

3.5.1. Expression correlation analysis

Fragments Per Kilobase of transcript per Million mapped reads upper quartile (FPKM-UQ+1) values of *ERG* and *LINC00920* in the TCGA-PRAD RNA-seq data were extracted from the UCSC Xena platform [209]. Expression correlation analyses were carried out for both ICGC-EOPC (n=135) and TCGA-PRAD (n=551) cohorts using the R *cor()* function executed in RStudio (version 1.0.136).

3.5.2. Promoter analysis and ERG binding site prediction

Raw H3K27ac, H3K4me3, and RNA polymerase II ChIP-seq data were downloaded from the NCBI Gene Expression Omnibus (GSE57498) [210]. All data analyses subsequently described were conducted on the Galaxy public server (URL: usegalaxy.org) [211]. Quality assessment of *.fastq* files was performed using FastQC (version 0.11.5) [212]. Overrepresented sequences corresponding to adapter sequences were clipped using the clip adapter sequences tool (version 1.0.1). Reads were filtered using the filter by quality tool (version 1.0.0) with the quality cut-off value set at 20 and the percent of bases in sequence reaching the cut-off was set at 90%. Read ends (quality score = 20) were trimmed using FASTQ quality trimmer (version 1.0.0) [213]. Afterwards, trimmed reads were mapped to the human hg19 reference genome using Bowtie2 (version 1.1.2) [214]. SAMtools [215] was used to sequentially remove multi-mapping reads, sort the resulting *.bam* files, and remove PCR duplicates. BigWig files were generated from the clean *.bam* files using bamCoverage (version 2.5.0.0) [216] with bin size set to 25 bases. Finally, the bigwig files were visualized in the UCSC genome browser [217].

Promoter sequence 1000 bp upstream the annotated *LINC00920* transcription start site was extracted from the UCSC Genome Browser. ERG binding motifs along the promoter sequence were scanned and scored using JASPAR CORE [218] at a threshold of 85%.

3.5.3. Chromatin immunoprecipitation (ChIP)

3.5.3.1. Crosslinking of VCaP cells

VCaP cells were washed with 1X PBS and fixed with 1 mL of 1% formaldehyde per 1 x 10^6 cells at room temperature for 20 min. The reaction was quenched by adding 1/10 volume of 1.25 M glycine and rocking the cell suspension for 5 min. Cells were washed twice with ice-cold 1X PBS. Cytoplasmic lysis was performed by resuspending cells in ice-cold cytoplasmic lysis buffer (150 mM NaCl, 20 mM EDTA pH 8.0, 50 mM Tris-HCl pH 7.5, 0.5% NP-40, 1% Triton X-100, 20 mM NaF) supplemented with 1X cOmplete, Mini protein inhibitor cocktail and 1X phenylmethane sulfonyl fluoride (PMSF). The remaining nuclear pellet was collected by centrifugation at 12,000 g, 4^{o}C for 2 min, snap-frozen, and stored at -80^{o}C.

3.5.3.2. *Chromatin preparation*

Chromatin lysates were prepared by first resuspending the nuclear pellet in a nuclear lysis buffer (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 50 mM Tris-HCl pH 8.0, 20 mM EDTA pH 8.0, 20 mM NaF, 0.1% SDS) supplemented with 1X cOmplete, Mini protein inhibitor cocktail and 1X PMSF, at a density of 3 x 10^6 cells/150 µL. Nuclear lysis was performed on ice for 15 min and the resulting lysate was aliquoted into 1.5 mL Bioruptor® microtubes (150 µL/tube). Chromatin preparations were then sonicated at 4^{o}C using the Bioruptor® Pico sonication device for 13 cycles applying the 30-second on/off high setting. The lysates were pooled and clarified by centrifugation at 8,000 g, 4^{o}C for 10 min.

3.5.3.3. Immunoprecipitation

Chromatin lysate volume equivalent to 4 x 10^6 cells and 30 µL of ChIP-grade protein G magnetic beads were used for each immunoprecipitation reaction. Prior to use for lysate pre-clearing, 30 µL magnetic beads were pre-washed with nuclear lysis buffer. Chromatin lysate was diluted to 600 µL with the nuclear lysis buffer and added into the washed beads and pre-cleared by rotating at 4^{o}C for 2 h. The beads were separated on a magnetic rack and the cleared lysate was transferred into a new tube. A volume corresponding to the 2% input sample was aliquoted and 2 µg of antibody (**Table 2-5**) was added to the remaining lysate. Antibody hybridization was facilitated by overnight rotation at 4^{o}C. Next day, 30 µL of pre-washed magnetic beads was added to the reaction tube and hybridized at 4^{o}C for 2 h.

3.5.3.4. Bead washing and DNA isolation

The magnetic beads were collected and washed sequentially using the following ice-cold buffers supplemented with 1X PMSF: once with the nuclear lysis buffer, once with high-salt buffer (50 mM Tris pH 7.9, 500 mM NaCl, 1 mM EDTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS), once with Li buffer (20 mM Tris pH 7.9, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.1% sodium deoxycholate), and twice with 10 mM Tris pH 7.9. Chromatin elution was performed by adding 150 µL elution buffer (50 mM Tris pH 7.9, 1 mM EDTA pH 8.0, 1% SDS, 50 mM $NaHCO_3$, 300 mM NaCl) and 2 µL RNase A to the input and beads. Reaction tubes were incubated at 65^{o}C with shaking overnight. Afterwards, 2 µL Proteinase K (20 mg/mL) was added to each reaction followed by incubation at 60^{o}C for 1 h to reverse chemical crosslinks. The magnetic beads were discarded and DNA was isolated from the eluate using UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v). Briefly, 300 µL phenol:chloroform:isoamyl alcohol was mixed with the eluate for 10 min. The samples were centrifuged at room temperature, 14000 rpm for 5 min. The aqueous layer was transferred into a new tube wherein 3 µL GlycoBlue coprecipitant, 120 µL 3M pH 5.2 sodium acetate, and 900 µL 100% ethanol were added. DNA was precipitated overnight at -20^{o}C. To collect the precipitated DNA, the samples were centrifuged at 4^{o}C, 14,000 rpm for 30 min. The pellet was washed

with 70% ethanol, air-dried, and reconstituted in 100 µL nuclease-free water. Five microliters of DNA was used in subsequent qPCR assays.

3.5.3.5. ChIP-qPCR

To quantify DNA enrichments, SYBR assays were performed as described in section 3.2.9.2. using primers in **Table 2-9**. Percent of input using Cp values were calculated as:

$$\%\, input \;=\; 100\% \times 2^{AI - Cp_{pulldown}}$$

where

$$adjusted\ input\ (AI) \;=\; Cp_{input} - log_2\left(\frac{volume_{pulldown}}{volume_{input}}\right).$$

3.5.4. Promoter luciferase assay

LNCaP #126 T/E cells were seeded into black, flat- and clear-bottomed 96-well plates (2 x 10^4 cells in 75 µL complete medium/well) for 48 h. Next, doxycycline treatment was carried out for another 48 h as previously described (section 3.1.6). Control and doxycycline-treated cells were then co-transfected using Lipofectamine® 2000 with the generated pGL4.10[*luc2*] construct containing either wild-type or mutant *LINC00920* promoter fragments and pAAVpsi2 in 10:1 ratio (90 ng:10 ng). pAAVpsi2 encodes the *Renilla* luciferase gene (*hRluc*) under the control of the SV40 promoter and enhancer [199]. The latter vector was used to uncouple the luminescence signals from the effects of varying transfection efficiencies across wells, effectively normalizing luminescence derived from the firefly *luc2* reporter gene. After 48 h, firefly and *Renilla* luminescence signals were developed using the Dual-Glo® Luciferase assay system following the manufacturer's protocol. Luminescence signals were measured using TECAN Infinite M200 with the i-control™ (version 1.6) software.

3.6. Chromatin isolation by RNA purification (ChIRP)

3.6.1. Crosslinking of PC-3 cells

To preserve RNA/DNA/protein interactions, PC-3 cells were fixed using either 1% glutaraldehyde (ChIRP-seq) or 3% formaldehyde (ChIRP-MS). The crosslinking procedures for both reagents are similar. Briefly, cells were trypsinized and washed twice with 1X PBS. Cells were resuspended in the chemical crosslinker at a concentration of 1 x 10^6 cells/ 1 mL 1% glutaraldehyde or 3% formaldehyde in 1X PBS and incubated with rocking at room temperature for 10 min (glutaraldehyde) or 30 min (formaldehyde). Crosslinkers were then quenched with 1:10 volume of 1.25 M glycine for 5 min. Cells were collected by centrifugation at 1,000 g, 4^{o}C for 5 min, and washed thrice with ice-cold 1X PBS. Cell pellets were collected in 1.5 mL tubes and snap-frozen on dry-ice.

3.6.2. Cell lysate preparation

Crosslinked cells were thawed on ice and resuspended in ChIRP cell lysis buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA pH 8.0, 1% SDS) supplemented with 1X cOmplete, Mini protein inhibitor, 1X

PMSF, and 0.05 U/µL Superase In™ RNase inhibitor at a density of 2 x 10^4 cells/mL. Cell lysis was facilitated by incubation on ice for 15 min. Next, 300-µL volumes of cell suspension were aliquoted into 1.5 mL Bioruptor® microtubes and sonicated at 4°C using the Bioruptor® Pico sonication device for 35 cycles with the 30-second on/off high setting. Sheared lysates were pooled and clarified by centrifugation at 12,000 rpm, 4°C for 10 min.

3.6.3. ChIRP-seq

For each pulldown reaction (i.e., *lacZ*, *LINC00920*), sheared cell lysate equivalent to 6 x 10^7 glutaraldehyde-fixed cells was diluted to 5 mL with ChIRP hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris-Cl pH 7.5, 1 mM EDTA, 15% formamide) supplemented with 1X cOmplete, Mini protein inhibitor, 1X PMSF, and 0.05 U/µL Superase In™ RNase inhibitor. In parallel, 100 µL slurry of Dynabeads™ MyOne™ Streptavidin C1 was pre-washed with hybridization buffer. The cell lysate was then pre-cleared by mixing with washed beads at 37°C for 30 min. The beads were magnetically separated and discarded. An aliquot corresponding to 0.5% input from the cleared lysate was set aside and stored at -80°C. Three hundred picomoles of pooled oligos (**Table 2-12**) was denatured at 75°C for 2 min, cooled on ice, and added to the clarified lysate. Hybridization was facilitated by rotation overnight at 37°C. Next day, 300 µL of bead slurry was pre-washed with the hybridization buffer and added into the hybridization reaction. Biotinylated complexes were captured on the beads by incubation with rocking at 37°C for 30 min. Afterwards, the beads were collected and the supernatant was discarded. The beads were then washed 5 times with 1 mL ChIRP wash buffer (2X saline-sodium citrate buffer, 0.5% SDS, 1X PMSF) at 37°C for 5 min. During the last wash, 25 µL of the bead slurry was aliquoted for RNA isolation. The remaining beads were then collected for DNA isolation.

3.6.3.1. RNA isolation

Prior to RNA isolation, crosslink reversal was performed on input samples and capture beads by proteinase K digestion. Briefly, samples were resuspended in 95 µL proteinase K buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, 0.5% SDS) and supplemented with 5 µL of 20 mg/mL proteinase K. Tubes were incubated at 65°C with shaking for 45 min, followed by enzyme deactivation at 95°C for 10 min. RNA from the samples were subsequently isolated using miRNeasy® Mini kit following the manufacturer's protocol, including the optional DNase-I digestion step.

3.6.3.2. DNA isolation

DNA bound to the beads was eluted by resuspending the beads in 150 µL high salt elution buffer (50 mM Tris pH 7.9, 1 mM EDTA pH 8.0, 1% SDS, 50 mM $NaHCO_3$, 300 mM NaCl) supplemented with 15 µg RNase A and 15 U RNase H and incubating at 37°C for 30 min. This step was done twice. The beads were then discarded and 15 µL proteinase K was added to the supernatant. Crosslink reversal was facilitated by incubating the samples overnight at 50°C. DNA isolation was performed using phenol:chloroform:isoamyl alcohol as previously described in section 3.5.3.4., with the final

reconstitution volume decreased to 20 µL. DNA concentrations were measured by Qubit® Fluorometer using the Qubit® dsDNA HS assay kit. Three to ten nanograms of precipitated DNA was used for subsequent library preparation.

3.6.3.3. Library preparation

The NEBNext Ultra II DNA library preparation kit for Illumina and NEBNext Multiplex Oligos for Illumina, Index Primers Set 1 were used, following the manufacturer's protocols with some modifications, to prepare sequencing DNA libraries from ChIRP-DNA. Due to the low amounts of input DNA, the NEBNext adapter was diluted 25-fold (1:25) in 10 mM Tris-HCl, pH 8.0 with 10 mM NaCl. Size-selection step prior to amplification was not performed due to the same reason of low DNA input. Index primer combination for the libraries was chosen to enable multiplex sequencing. **Table 2-13** shows the index primers matched with each library. PCR enrichment was performed using 12-15 amplification cycles. To remove primer duplexes, a final bead clean up was performed using Agencourt AMPure XP beads. Quality assessment of libraries was performed using the DNA High Sensitivity kit with 2100 Bioanalyzer. The libraries were pooled at an equimolar concentration (10 nM per library in 30 µL volume) and the resulting multiplexed sample was submitted to the Genomics and Proteomics Core Facility of the DKFZ. Single-read, 50-bp sequencing in a single lane of the HiSeq 2000 sequencing platform was employed to obtain a depth of at least 20 million reads per library.

3.6.3.4. Analyses of high-throughput sequencing data

3.6.3.4.1. Genomic enrichment analysis

ChIRP-seq data were initially obtained as raw de-multiplexed *.fastq* files. Additionally, raw H3K27ac, H3K4me3, H3K4me1, H3K4me3, and RNA polymerase II ChIP-seq data were downloaded from the NCBI Gene Expression Omnibus (GSE57498) [210]. These files were pre-processed as described in section 3.5.2 to generate clean *.bam* files. Next, the hg19 genome was split into bins of 10 kbp and read coverages were computed using multiBamSummary (version 2.5.0.0). The matrix output was processed by plotCorrelation (version 2.5.0.0) to generate the correlation heatmap.

To map ChIRP-seq and ChIP-seq reads across annotated genomic features, each dataset was first normalized with control signals (i.e., lacZ ChIRP or input ChIP) using bamCompare (version 2.5.0.0) (--binsize 50). In mapping *LINC00920* occupancy across genic regions, the *.bed* file containing genic locations in the hg19 genome build was extracted from the UCSC browser. Computed matrices were aligned with the .bed file using plotHeatmap (version 2.5.0.0) wherein gene lengths were normalized to 10 kbp with 2-kb extensions upstream the transcription start sites and downstream the transcription end sites. Promoter and enhancer enrichment analyses were performed in an analogous manner. Genomic regions annotated as promoters or enhancers in the PC-3 genome were extracted as *.bed* files

from NCBI Gene Expression Omnibus (GSE73785) [219]. Normalized matrices were then aligned with the corresponding *.bed* file using plotHeatmap in reference-point mode.

To cluster the promoter regions based on similarity in *LINC00920* score distribution, k-means clustering was activated in plotHeatmap where the number of clusters to compute was set to 4.

To determine the biological relevance of *LINC00920* enrichment in cluster 1 promoters, promoter loci were analyzed using Genomic Regions Enrichment of Annotations Tool (GREAT) [202] with species assembly set to hg19 and background region set to the whole genome.

3.6.3.4.2. Peak calling

Model-based analysis of ChIP-seq 2 (MACS2) was used to identify the chromatin binding sites of *LINC00920*. The bandwidth fragment size was set to 150 (--bw 150); lower and upper mfold boundaries were set to 10 and 30, respectively (--mfold 10 30); q-value was set to 0.001 (--qvalue 0.001); and broad peak calling was performed (--broad) using the shifting model. Peaks were called for each replicate separately using paired lacZ pulldowns as normalizing controls. High confidence peaks (n=2985) were identified by overlapping called peaks among the replicates and setting a fold-change cutoff value of 10. Peaks were annotated with genomic features using published chromatin state and characterization (ChromHMM) data generated specifically for the PC-3 genome [219].

3.6.3.4.3. Motif discovery

Genomic sequences of the top 1000 MACS2-called peaks (ranked by fold-enrichment) were extracted from the hg19 build of the human genome using the Extract Genomic DNA (version 3.0.3) function of Galaxy. The sequences were then used to identify DNA motifs using the default parameters of MEME (version 4.6.0) [220].

3.6.4. ChIRP-MS

For each pulldown reaction (i.e., *lacZ*, *LINC00920*), sheared cell lysate equivalent to 2 x 10^8 formaldehyde-fixed cells was diluted to 10 mL with ChIRP hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris-Cl pH 7.5, 1 mM EDTA, 15% formamide) supplemented with 1X cOmplete, Mini protein inhibitor, 1X PMSF, and 0.05 U/µL Superase In™ RNase inhibitor. In parallel, 200 µL slurry of Dynabeads™ MyOne™ Streptavidin C1 was pre-washed with the hybridization buffer. The cell lysate was then pre-cleared by mixing with the washed beads at 37°C with gentle rocking for 30 min. The beads were magnetically separated and discarded. An aliquot corresponding to 0.5% input from the cleared lysate was set aside and stored at -80°C. Six hundred picomoles of pooled oligos (**Table 2-12**) were denatured at 75°C for 2 min and placed on ice. Denatured oligos were added to the clarified lysate and hybridization was facilitated by rotation overnight at 37°C. Next day, 600 µL bead slurry was pre-washed with hybridization buffer and added into the hybridization reaction. Biotinylated complexes were captured on the beads by incubation with rocking at 37°C for 30 min. Afterwards, the

beads were collected and the supernatant was discarded. The beads were then washed 5 times with 1 mL ChIRP wash buffer at 37°C for 5 min. During the last wash, 25 µL of the bead slurry was aliquoted for RNA isolation. The remaining beads were then collected for protein elution.

3.6.4.1. Protein elution

Beads were resuspended in 800 µL benzonase elution buffer (20 mM Tris-HCl pH 7.9, 0.05% N-lauroylsarcosine, 2 mM $MgCl_2$, 0.5 mM Tris[2-carboxyethyl]phosphine) followed by addition of 125 U of Benzonase® non-specific nuclease. Protein elution was facilitated by overnight digestion at 37°C. Next day, the beads were discarded and proteins were precipitated by adding trichloroacetic acid to a final concentration of 25%. The samples were incubated overnight at 4°C. Precipitated proteins were collected by centrifugation at 16,000 g, 4°C for 1 h. The supernatant was removed and the pellet was washed once with 100% ice-cold acetone. Protein pellets were air-dried and snap-frozen on dry ice and submitted to the mass spectrometry-based Protein Analysis Unit of the Genomics and Proteomics Core Facility, DKFZ.

3.6.5. Analysis of identified proteins

An enrichment ratio of 1.2 (*LINC00920* ChIRP:*lacZ* ChIRP) was set to identify candidate *LINC00920*-interacting proteins. Enriched proteins common to 3 biological replicates were investigated by computing gene set overlaps in the molecular signatures database (MsigDB) [221] using the BioCarta, KEGG, Reactome, and gene ontology gene sets.

3.7. Validation of RNA-protein interaction

3.7.1. RNA immunoprecipitation (RIP)

3.7.1.1. Crosslinking of VCaP cells

VCaP cells were washed with 1X PBS and fixed with 1 mL of 1% glutaraldehyde per 1 x 10^6 cells at room temperature for 10 min. The reaction was quenched by adding 1/10 volume of 1.25 M glycine and rocking for 5 min. Cells were collected by centrifugation at 1,000 g, 4°C for 5 min, and washed thrice with ice-cold 1X PBS. Cell pellets were collected in 1.5 mL tubes and snap-frozen on dry-ice.

3.7.1.2. Cell lysate preparation

Cell lysis was performed by resuspending cells in ice-cold RIP buffer (150 mM KCl, 25 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5 mM dithiothreitol, 0.5% NP-40) supplemented with 1X cOmplete, Mini protein inhibitor, 1X PMSF, and 0.05 U/µL Superase In™ RNase inhibitor at a density of 3 x 10^6 cells/150 µL. Samples were incubated on ice for 15 min and the resulting lysate was aliquoted into 1.5 mL Bioruptor® microtubes (150 µL/tube). Cell lysates were sonicated at 4°C using the Bioruptor® Pico sonication device for 13 cycles applying the 30-second on/off high setting. The lysates were pooled into 400 µL volumes and supplemented with 28 U Superase In™ RNase inhibitor and 5.5 U

TURBO™ DNase. Chromatin digestion was facilitated by incubating the samples at 37°C for 15 min. The reaction was quenched by adding EDTA pH 8.0 to a final concentration of 15 mM. Lysates were clarified by centrifugation at 8000 g, 4°C for 10 min, transferred into new tubes, and snap-frozen on dry-ice.

3.7.1.3. Immunoprecipitation

Cell lysate volume equivalent to 7.5 x 10^6 cells and 45 µL of ChIP-grade protein G magnetic beads were used for each immunoprecipitation reaction. Prior to use for pre-clearing, 30 µL magnetic beads were pre-washed with RIP buffer. The cell lysate was diluted to 500 µL with RIP buffer and added into the washed beads and pre-cleared by rotating at 4°C for 2 h. The beads were separated on a magnetic rack and the cleared lysate was transferred into a new tube. A volume corresponding to 5% input sample was aliquoted and 3 µg of antibody (**Table 2-6**) was added to the remaining lysate. Magnetic beads from 45 µL slurry were pre-washed and added into the hybridization reaction. Samples were incubated overnight with rotation at 4°C.

3.7.1.4. Bead washing and RNA isolation

Magnetic beads were collected and washed 4 times with ice-cold RIP wash buffer (3X SSC buffer, 1 mM EDTA pH 8.0, 0.1% Tween-20, 1X PMSF) and once with ice-cold 1X PBS supplemented with 1X PMSF. Crosslink reversal of input and beads was performed by adding 95 µL proteinase K buffer and 5 µL of proteinase K (20 mg/mL). Tubes were incubated at 65°C with shaking for 45 min, followed by enzyme deactivation at 95°C for 10 min. RNA from the samples were subsequently isolated using phenol/chloroform/isoamyl alcohol (125:24:1 mixture pH 4.3). Briefly, 100 µL of the phenol/chloroform/isoamyl alcohol reagent was mixed with the samples. After centrifugation at 21,000 g, room temperature for 10 min, the upper aqueous phase was transferred into a fresh tube and mixed with 1 volume of chloroform. The samples were centrifuged at 20,000 g, room temperature for 10 min and the upper aqueous phase was transferred into a fresh tube. To precipitate RNA, 3 µL GlycoBlue, 0.1 volume of 3 M pH 5.2 sodium acetate, and 1 volume of 100% isopropanol were added to the samples. After overnight incubation at -20°C, the RNA pellet was collected by centrifugation at 20,000 g, 4°C for 30 min. The RNA was washed once with ice-cold 80% ethanol, air-dried, and resuspended in 11 µL nuclease-free water. Complementary DNA was synthesized as described in section 3.2.2 using the entirety of reconstituted RNA. Each 20 µL cDNA reaction volume was then diluted to 100 µL with nuclease-free water. Five microliters of cDNA was used in subsequent qPCR assays.

3.7.1.5. RIP-qPCR

RNA enrichments were quantified by qPCR as described in section 5.2.5 using primers in **Table 2-9**. Fold enrichment values were calculated as follows:

$$Fold\ enrichment\ = 2^{\left[(AI-Cp_{pulldown})_{14-3-3\ IP}-(AI-Cp_{pulldown})_{IgG\ IP}\right]}$$

$$=\frac{100\%\times(2^{AI-Cp_{pulldown}})_{14-3-3\ IP}}{100\%\times(2^{AI-Cp_{pulldown}})_{IgG\ IP}}$$

$$=\frac{\%\ of\ input_{14-3-3\ IP}}{\%\ of\ input_{IgG\ IP}}$$

3.7.2. Affinity purification

3.7.2.1. In vitro transcription

Biotinylated *LINC00920* RNA was generated by using T7 RNA polymerase with a biotin RNA labeling mix to transcribe the full length lncRNA cDNA previously cloned in pcDNA3.1(+). The vector harbors a T7 promoter upstream its multiple cloning site which can be used for *in vitro* transcriptional activation. Briefly, sequence-verified pcDNA3.1(+)_LINC00920 was digested with XbaI to obtain a linearized DNA template which was subsequently purified as described in section 3.2.1. A linearized vector containing the *Fluc* gene was used as positive control for the reaction. Each 20 µL transcription reaction consisted of 1X transcription buffer, 1X biotin labeling mix, 100 U T7 RNA polymerase, 500-1000 ng DNA template, and nuclease-free water. The reaction mixtures were incubated at 37^{o}C for 2 hours. To remove template DNA, the reaction volume was brought up to 50 µL with nuclease-free water, supplemented with 4 U of RNase-free DNase-I, and incubated at 37^{o}C for 15 min. The RNA product was purified using the RNeasy® Mini Kit. Optional poly-A tailing was performed by mixing 2000 ng purified RNA with 1X *E. coli* poly(A) polymerase reaction buffer, 1 mM ATP, 0.5 U/µL Superase In^{TM} RNase inhibitor, and nuclease-free water to a final volume of 20 µL. The reaction was incubated at 37^{o}C for 30 min followed by RNA purification. RNA yield was quantified using NanoDrop 1000. RNA quality was assessed by electrophoresing 100 ng of sample through 1% agarose/1X MOPS/6% formaldehyde gels. RNA bands were visualized by staining with 1X SYBR® Gold nucleic acid gel stain and viewing under UV light.

3.7.2.2. Hybridization and purification

Recombinant human 14-3-3ε protein (ab54317) was obtained from Abcam. For each affinity pulldown, 300 ng of recombinant protein and 0.5 pmol purified biotinylated RNA were hybridized. A day prior to the experiment, 50 µL Dynabeads™ M-270 Streptavidin bead slurry was washed and blocked with 0.1% BSA overnight at 4^{o}C. Next day, 0.5 pmol RNA was diluted to 50 µL with RNA structure buffer (10 mM Tris-HCl pH 7.5, 0.1 mM KCl, 10 mM $MgCl_2$) and denatured at 75^{o}C for 2 min and placed on ice. The recombinant protein was diluted to 1 mL with EMSA buffer (20 mM HEPES pH 7.6, 50 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA pH 8.0, 5 mM DTT) supplemented with 1X cOmplete, Mini protein inhibitor, 1X PMSF, and 0.05 U/µL Superase In^{TM} RNase inhibitor. RNase A treatment was performed by adding the enzyme at a final concentration of 10 µg/mL. Denatured RNA was then added to the protein solution and the tubes were incubated with rotation at room temperature

for 2 hours. Blocked beads were washed thrice with EMSA buffer and added to the hybridization solution. Bead capture was facilitated at room temperature for 10 min. The beads were magnetically separated and washed 5 times with ice-cold RIP wash buffer. Proteins were eluted by resuspending the beads in 30 µL 1X Roti®-Load 1 protein loading buffer followed by incubation at 95^{o}C for 5 min. The beads were separated and discarded while the supernatant was used for polyacrylamide gel electrophoresis and western blot as described in sections 3.2.4 and 3.2.5.

3.8. Statistical analyses

All quantitative data were presented as mean ± standard error (SE). Unpaired t-test was used to accept or reject the null hypothesis that there was no significant difference between the control and treated conditions with respect to cellular phenotypes (i.e., cell proliferation, colony formation, cell migration), expression levels (i.e., relative gene expression, luciferase assays), and enrichment values (i.e., as determined by ChIP-, RAP-, and ChIRP-qPCR).

4. Results

4.1. Long non-coding RNA candidate selection

Differentially expressed lncRNAs from the ICGC-EOPC whole transcriptome sequencing data were screened for promising candidates that may play a role in prostate cancer development or progression. This cohort consisted of 125 prostate tumor and 10 normal tissue specimens that were sampled from early onset PCa patients (i.e., under the age of 50). A DeSeq analysis was previously performed on the transcriptome data (**Figure 4-1**) and the lncRNA selection process was guided by the following criteria: (i) the non-coding transcript must be of the long intergenic RNA (lincRNA) biotype; (ii) there should not be a gene in the antisense orientation; (iii) the lncRNA must have at least 2 exons; (iv) the average transcript count (i.e., baseMean value) must be at least 500 for either the tumor or normal sample group; (v) a significant (p value<0.05) up- or downregulation must be observed between the tumor and normal sample group (log2FC >|1|); (vi) the presence of specific genetic and epigenetic elements within the 50 kbp locus of the candidate gene (e.g., active transcription marks, CpG islands, oncogenes, or tumor suppressor genes); (vii) and expression profiles in agreement with annotations in relevant databases (i.e., TCGA [75], MiTranscriptome [222]). Subsequently, 7 lncRNAs were considered for further investigation (**Figure 4-1** and **Table 4-1**).

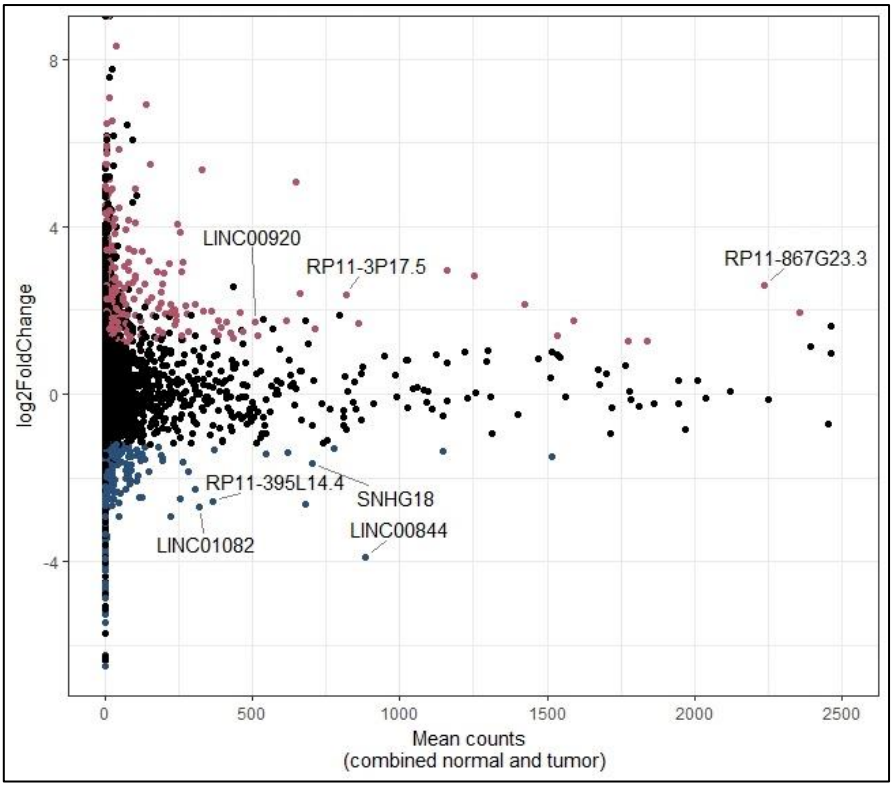


Figure 4-1. MA plot of 7,335 long intergenic RNAs (lincRNAs) analyzed by DeSeq. Red points represent significantly upregulated transcripts (p>0.05, log2[fold change]>1.2). Blue points represent significantly downregulated transcripts (p>0.05, log2[fold change]<1.2). Labelled points correspond to shortlisted lncRNA candidates further investigated in this study.

Table 4-1. Summary of DeSeq analysis and annotations of selected PCa-related lncRNAs.

Candidate	Relative expression	Highest baseMean count (tumor or normal)	Fold-change over normal tissue	p-value	Available annotation
RP11-867G23.3 ENST00000501708.1[#]	Upregulated	3903.975	5.845	0.018	None
RP11-3P17.5 ENST00000602890.1[#]	Upregulated	873.214	5.201	0.022	Upregulated in prostate tumors[§]
LINC00920 ENST00000499966.1[#]	Upregulated	538.437	3.270	0.014	Upregulated in lung, breast, colorectal carcinoma*; correlates with ERG mRNA expression[§]
LINC00844 ENST00000432535.1[#]	Downregulated	6331.756	0.067	1.11×10^{-10}	Downregulated in breast, lung, prostate tumors*
LINC01082 ENST00000601250.1[#]	Downregulated	1457.759	0.153	1.26×10^{-11}	Downregulated in colorectal, prostate tumors*
RP11-395L14.4 ENST00000416105.1[#]	Downregulated	1573.445	0.169	4.96×10^{-11}	Downregulated in prostate tumors with TP53 mutation[§]
SNHG18 ENST00000508179.1[#]	Downregulated	1885.687	0.320	8.54×10^{-5}	Downregulated in breast, lung, prostate tumors*

[#]Ensembl transcript ID; [§]TANRIC annotation; *TCGA annotation;

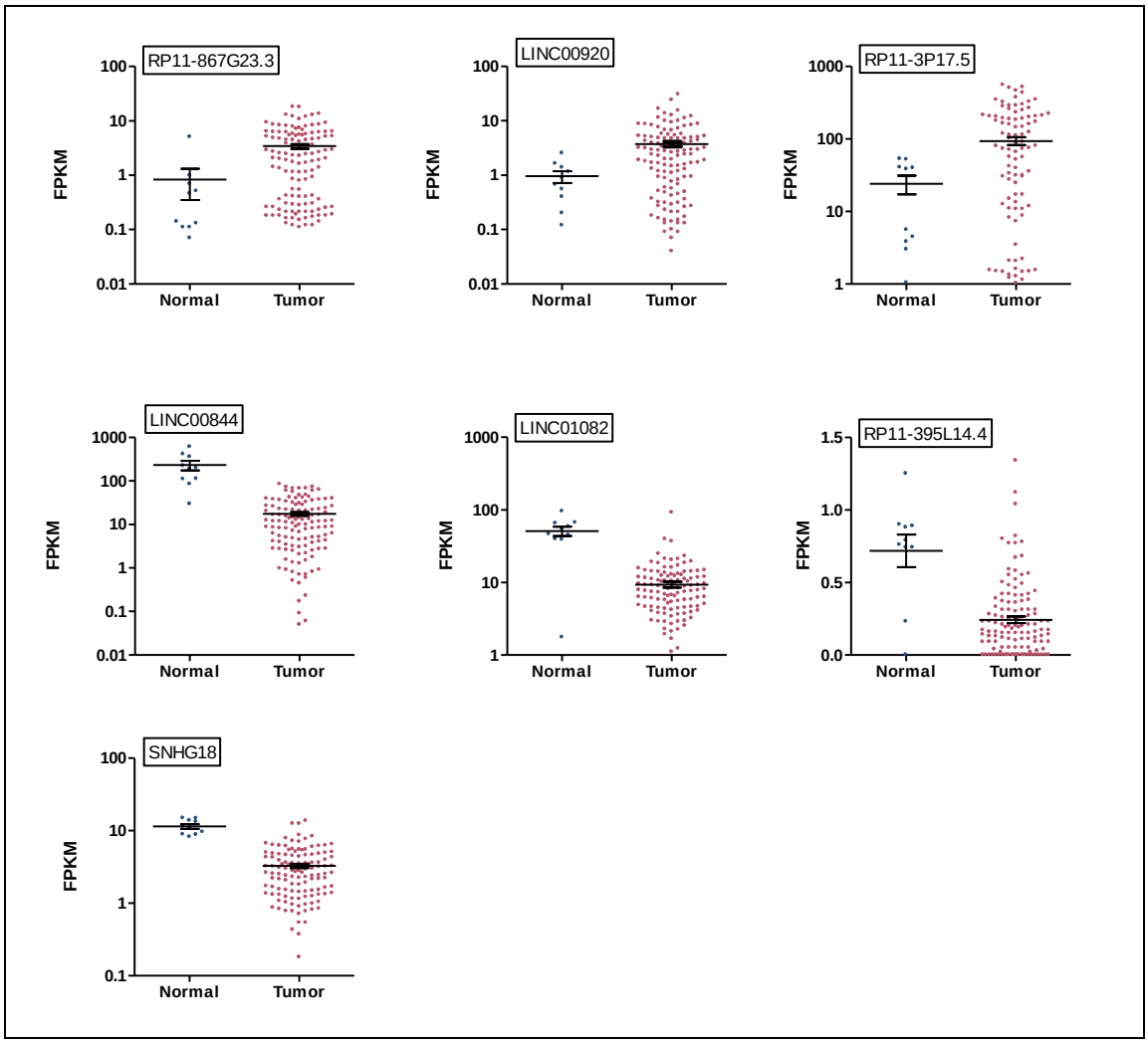


Figure 4-2. Transcript quantitation of the short-listed lncRNAs in the ICGC-EOPC dataset. FPKM: fragments per kilobase per million mapped reads.

4.2. Selected lncRNAs exhibit characteristic expression profiles in prostatic cell lines

LNCaP, VCaP, DU-145, and PC-3 cells were used as prostate cancer models to validate lncRNA expression *in vitro*. LncRNA expression levels from these metastatic lines were compared with the benign prostate epithelial cell line RWPE-1 to assess the relative expression (**Figure 4-3**). The transcription level of each lncRNA candidate varied among cell lines, which was not surprising given the unique genetic background and phenotype of the cells. At the same time, this further corroborates widespread reporting of lncRNAs being highly cell type-specific [157, 159, 223-225]. Candidates determined to be upregulated in the ICGC-EOPC cohort (i.e., *RP11-867G23.3*, *LINC00920*, and *RP11-3P17.5*) were overexpressed only in specific cell lines (**Figure 4-3A**). Briefly, *RP11-867G23.3* was found to be highly expressed in VCaP, while *LINC00920* was upregulated in both VCaP and PC-3 cells. On the other hand, *RP11-3P17.5* appeared to have comparable expression levels across all cell lines including RWPE-1. A similar generalization of cell-type specific expression can also be made for the downregulated candidates (i.e., *LINC00844*, *LINC01082*, *SNHG18*, and *RP11-395L14.4*) (**Figure 4-3B and 3C**). For *RP11-395L14.4*, amplification signals were detected only in RWPE-1, DU-145,

and PC-3. *SNHG18* was measurable only in RWPE-1 and PC-3 cells. *LINC00844* and *LINC01082* were not detected in any of the cell lines tested using a threshold C_p value of 35, but their expression was validated in normal prostate tissue using a standard reference RNA source (Invitrogen™ Thermo Fisher Scientific; **Figure 4-3C**).

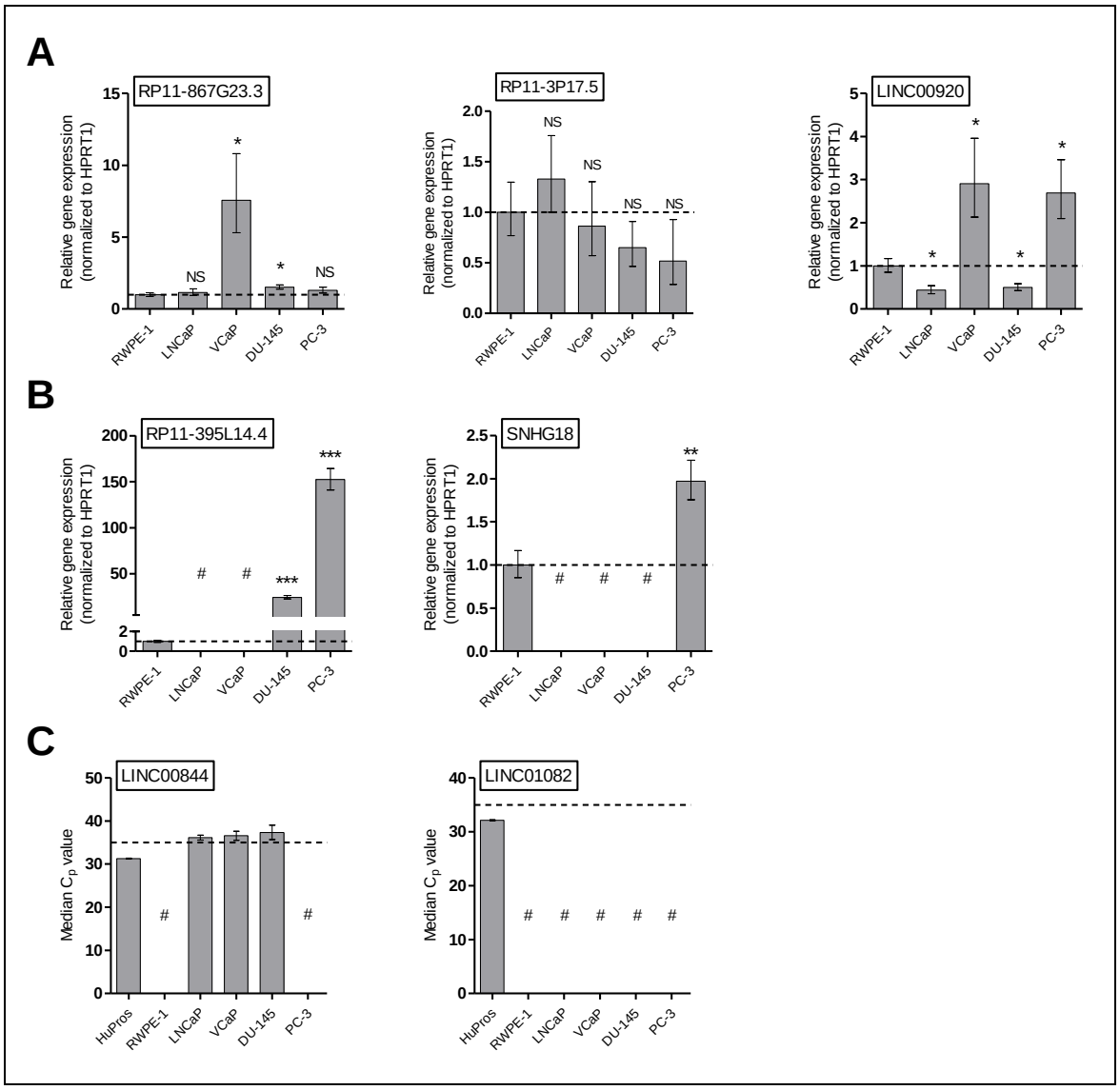


Figure 4-3. Representative expression profiles of lncRNA targets in indicated prostatic cell lines. (A) Upregulated lncRNAs. (B) Downregulated lncRNAs with measurable signals. Broken lines denote fold change = 1. Number signs (#) denote infinitely small or undetectable qPCR signals (i.e., C_p>35 or C_p=0). Changes in gene expression were analyzed using RWPE-1 as the normalizing control (*p≤0.05; **p≤0.01; ***p≤0.001). (C) Downregulated lncRNAs that were undetectable in all prostate cancer cell lines by qPCR. Broken lines denote the threshold C_p value of 35 cycles. HuPros: standard human prostate RNA sample.

4.3. Rapid amplification of cDNA ends reveals polyadenylation of candidate lncRNAs

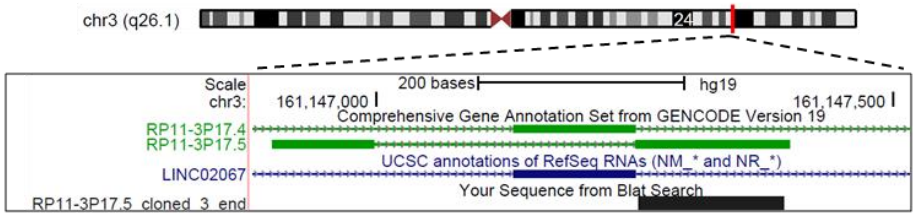
At present, lncRNA annotation remains highly dynamic and subject to frequent revisions [226, 227]. Rapid amplification of cDNA ends (RACE) [228] was performed on the selected lncRNAs to determine the correct gene transcription start site, the actual length of transcript, and the presence of polyadenylation at the 3' terminal end. Amplification of transcript cDNA was performed using

standard human prostate as RNA source. The cDNA ends of all lncRNA candidates except *RP11-867G23.3*—for which amplified end fragments consistently mapped to non-specific genomic regions—and *RP11-3P17.5*—whose 5'-end could not be primed—were successfully amplified, aligned, and cloned (**Figure 4-4**). Since an oligo dT primer was used for cDNA synthesis, 3'-end amplification was limited to transcripts harboring a polyadenylated tail. Indeed, it was possible to amplify the 3'-ends of *RP11-3P17.5*, *LINC00920*, *LINC00844*, *LINC01082*, RP11-395L14.4, and *SNHG18* these transcripts do possess this post-transcriptional modification.

For the most part, the transcript termini validated by RACE aligned correspondingly and with high fidelity to reference sequence databases. However, length divergence of a few nucleotides is apparent in many cases. This may be due to the cell-type specificity of transcript isoforms, as the RACE experiments were exclusively performed on prostatic RNA. It is also worth noting that reference annotations for *RP11-3P17.5* (**Figure 4-4A**), *LINC00844* (**Figure 4-4C**) and *LINC01082* (**Figure 4-4D**) were missing from the UCSC Genes track, while *LINC00920* annotations from UCSC Genes and GENCODE (**Figure 4-4B**) showed dissimilar 3' ends. A detailed inspection of *LINC00920* exon 2 sequence revealed eight polyadenylation signals (i.e., AATAAA or ATTAAA motifs), two of which could putatively enable transcript processing that would result in the 3'-end determined by RACE (**Figure 4-4G**). Similarly, polyadenylation signals preceeding the 3'-ends of *LINC00844* and *SNHG18* were also observed (**Figure 4-4H and 4I**). Interestingly, for transcripts with more than one splice variant, specifically *RP11-395L14.4* (**Figure 4-4E**) and *SNHG18* (**Figure 4-4F**), sequencing of cDNA ends appeared to identify the most likely predominant splice isoform in the context of prostate cells. Altogether, these observations highlight the cell-type specificity of gene transcription as well as the current limitations of reference sequence databases, specifically in annotating lncRNAs.

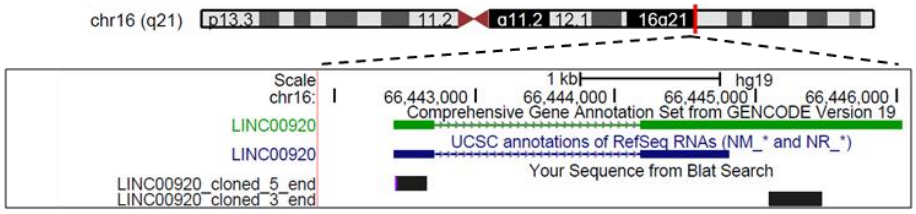
A

RP11-3P17.5



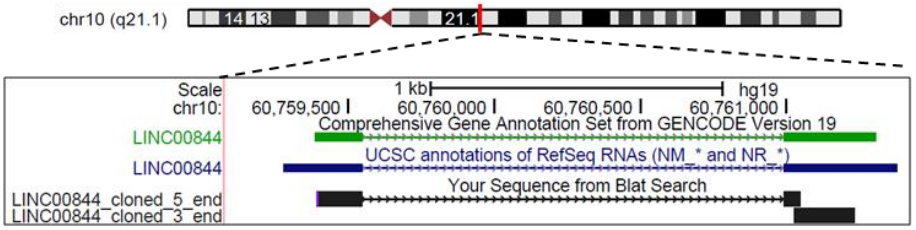
B

LINC00920



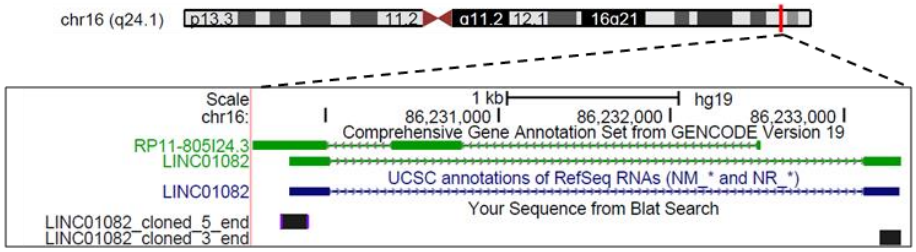
C

LINC00844



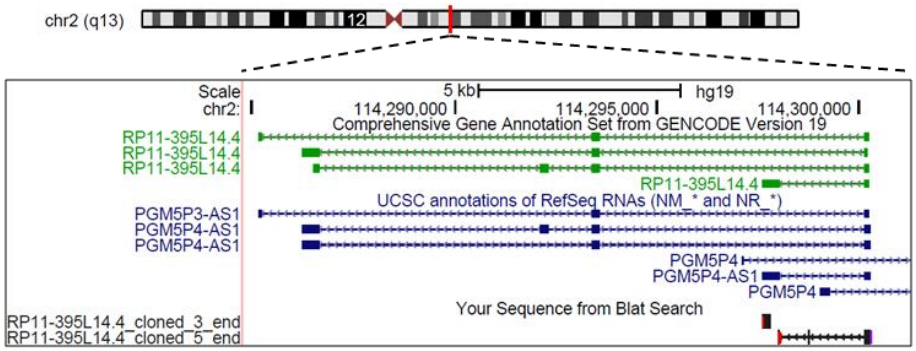
D

LINC01082



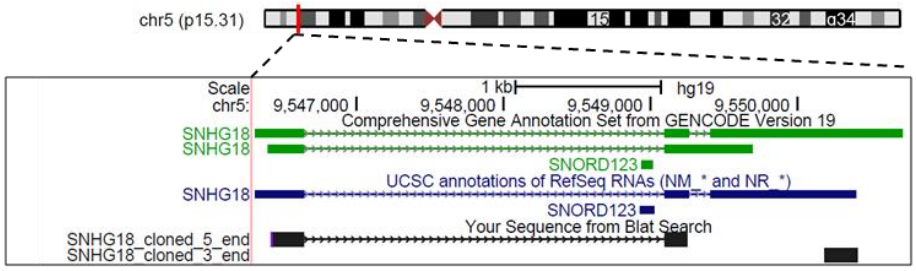
E

RP11-395L14.4



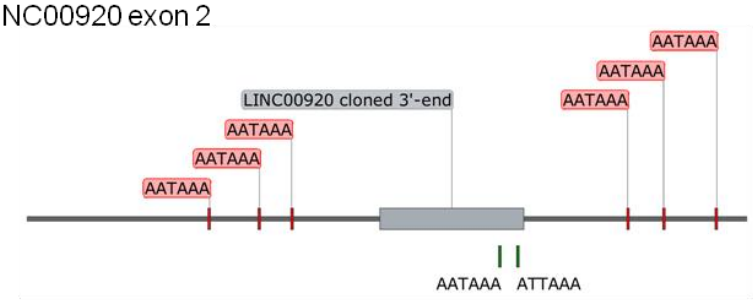
F

SNHG18



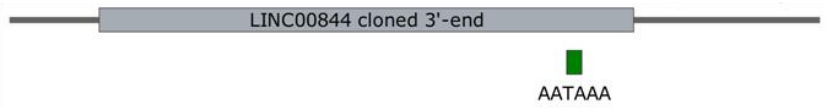
G

LINC00920 exon 2



H

LINC00844 exon 2



I

SNHG18 exon 3

Figure 4-4. Aligned sequences of cloned cDNA ends derived from RACE experiments. (A) *RP11-3P17.5*; (B) *LINC00920*; (C) *LINC00844*, (D) *LINC01082*; (E) *RP11-395L14.4*; (F) *SNHG18*. Purple tracks correspond to UCSC Genes. Green tracks correspond to GENCODE comprehensive annotation. Black tracks correspond to transcript termini identified by RACE. Detailed views of *LINC00920* exon 2 (G), *LINC00844* exon 2 (H), and *SNHG18* exon 3 (I) based on the GENCODE annotation. The labelled gray bar correspond to the cloned 3'-end of the lncRNA. Polyadenylation signals are indicated by colored bars. Putative polyadenylation signals for the RACE transcript are in green.

Due to the incomplete information on the transcript structure of *RP11-867G23.2* and *RP11-3P17.5*, the candidate list was shortened to the five remaining lncRNAs. **Table 4-2** summarizes the RACE experiment results.

Table 4-2. Summary of RACE results.

Candidate	Annotated transcript length (nt)*	5'-end	3'-end	Transcript length based on prostate RNA RACE (nt)
RP11-867G23.3 ENST00000501708.1	2792	Not cloned	Not cloned	n/a
RP11-3P17.5 ENST00000602890.1	248	Not cloned	Cloned	n/a
LINC00920 ENST00000499966.1	2147	Cloned	Cloned	1567
LINC00844 ENST00000432535.1	477	Cloned	Cloned	407
LINC01082 ENST00000601250.1	441	Cloned	Cloned	490
RP11-395L14.4 ENST00000416105.1	537	Cloned	Cloned	580
SNHG18 ENST00000508179.1	1799	Cloned	Cloned	1533

*Ensembl release 75 – Feb. 2014; n/a: not analyzed

4.4. Amplification and cloning of full-length transcripts

Amplification and cloning of full-length cDNAs were performed with the aim of generating overexpression constructs to functionally characterize the lncRNA candidates in prostate cancer cell lines. Candidate-specific primers were designed to amplify the complete transcript cDNA from a reference human prostate RNA template based on the outcome of RACE experiments. Successful amplification and sequencing of full-length *LINC00920*, *LINC00844*, and *LINC01082* were achieved (**Figure 4-5**). On the other hand, despite exhaustive attempts to amplify *SNHG18* and *RP11-395L14.4*, no PCR amplicons yielded fragments that matched the expected size, or aligned to the reference sequence (data not shown).

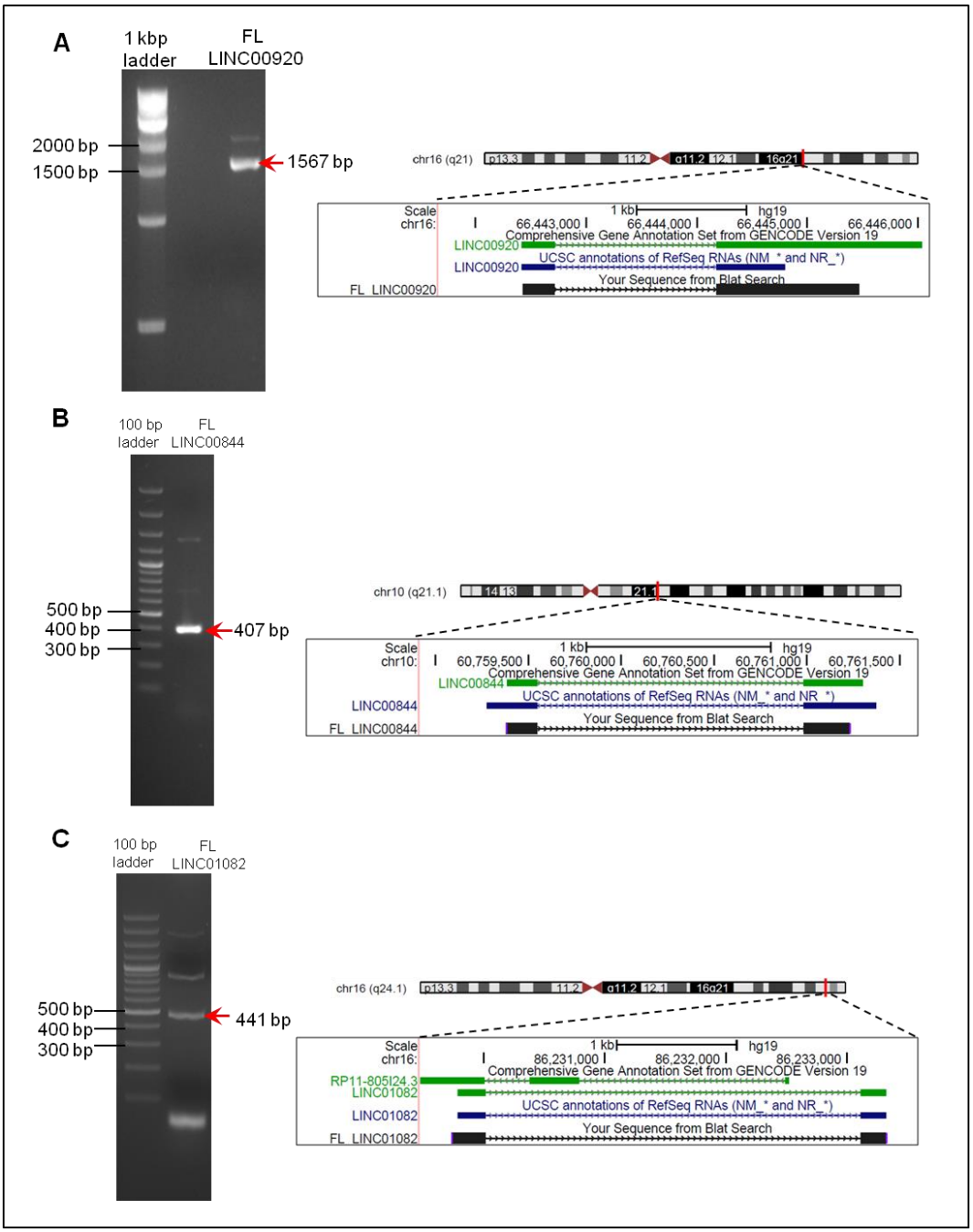


Figure 4-5. Amplification and sequencing of the full length (FL) cDNA of lncRNA targets. PCR products for (A) *LINC00920* (1567 bp), (B) *LINC00844* (407 bp), and (C) *LINC01082* (441 bp) were loaded in 2% agarose gels and stained with ethidium bromide. Alignment tracks are shown on the right of each agarose gel image.

The amplified cDNA fragments corresponding to *LINC00920*, *LINC00844*, and *LINC01082* were initially cloned into pCR™4Blunt-TOPO®, subcloned into the mammalian expression vector pcDNA™3.1(+), and sequence-verified by Sanger sequencing (**Supplementary Table 9-1**). Going forward, *LINC00920* was selected for further investigation due to its overexpression in tumors which suggested an oncogenic function. Moreover, among the remaining candidates, only *LINC00920* demonstrated robust expression in the PCa cell lines tested (**Section 4.2**), enabling the possibility of performing functional analysis through gene silencing. Lastly, the high endogeous expression of *LINC00920* would be advantageous for subsequent RNA pulldown assays.

4.5. Multiple models validate the non-coding potential of *LINC00920*

Three transcript assessment platforms—Coding-Potential Assessment Tool (CPAT) [200], Coding Potential Calculator (CPC) [201], and Phylogenetic Codon Substitution Frequencies (PhyloCSF) [229]—were utilized to confirm the non-coding potential of *LINC00920* (**Figure 4-6**). Using the transcript sequence verified by RACE, the CPAT and CPC scores were computed for *LINC00920* alongside *ACTB* (ENST00000331789.5) and *GAPDH* (ENST00000396861.1) mRNAs as coding transcript controls and *MALAT1* (ENST00000534336.1) and *NEAT1* (ENST00000501122.2) as non-coding controls. As expected, CPAT scores for *ACTB* and *GAPDH* mRNA were above the human coding threshold score of 0.364 [200] while values for *MALAT1*, *NEAT1*, and *LINC00920* transcripts were diminishingly low (**Figure 4-6A**). Evaluation of the CPC coding potential also presented a similar trend wherein *ACTB* and *GAPDH* transcripts had positive scores while the control lncRNAs and *LINC00920* produced negative values (**Figure 4-6B**). Positive PhyloCSF codon scores indicate increased likelihood of a particular region to be protein coding in a specified reading frame. Visualizing the *LINC00920* transcript alongside PhyloCSF tracks for the three forward reading frames revealed negative codon scores for regions where evaluation is possible (green tracks) (**Figure 4-6C**). Altogether, these tests affirm that *LINC00920* is a non-coding transcript.

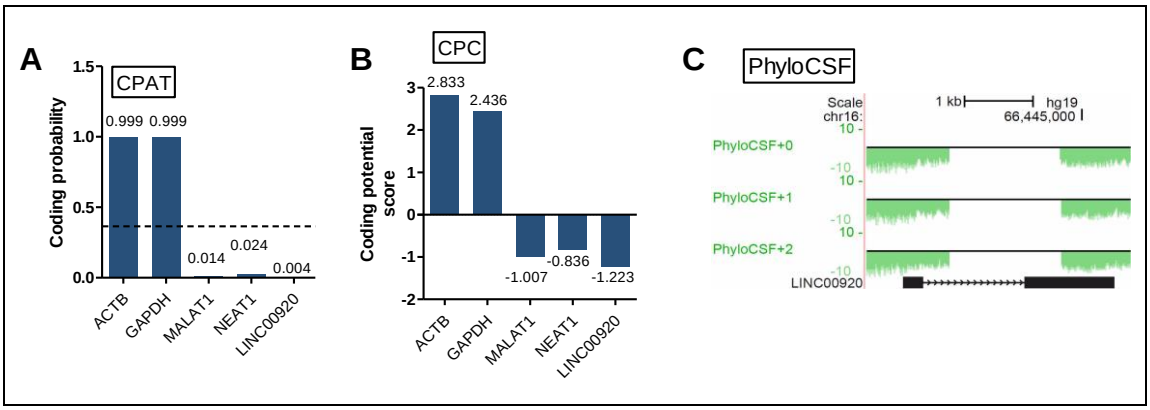


Figure 4-6. Validation of the non-coding potential of the *LINC00920* transcript. (A) Coding-Potential Assessment Tool (CPAT), (B) Coding Potential Calculator (CPC) and (C) Phylogenetic Codon Substitution Frequencies (PhyloCSF) were used to assess the non-coding potential of *LINC00920*. The dashed line in (A) indicates the human coding threshold score of 0.364. *ACTB* and *GAPDH* were used as coding transcript controls while *MALAT1* and *NEAT1* were used as non-coding transcript controls.

4.6. Silencing of *LINC00920* results in decreased cellular proliferation, migration, and colony formation of PC-3 cells

To query whether *LINC00920* has a functional role in prostate cancer cells, knockdown experiments were performed in PC-3 cells wherein the lncRNA exhibits a relatively high expression as previously described (**Section 4.2**). Two independent siRNAs were able to efficiently silence the transcript by at least 80% (**Figure 4-7A**). The proliferative capacity of cells transfected with *LINC00920*-targeting siRNAs was significantly reduced at the 72- and 96-h post-transfection time points (**Figure 4-7B**). These cells also formed fewer colonies after 9 days of cultivation (**Figure 4-7C-D**). There was also a marked decrease in their migratory potential across a Boyden chamber (**Figure 4-7E-F**). However, the transfected cells did not show remarkable changes in invasive capacity through Matrigel (data not shown).

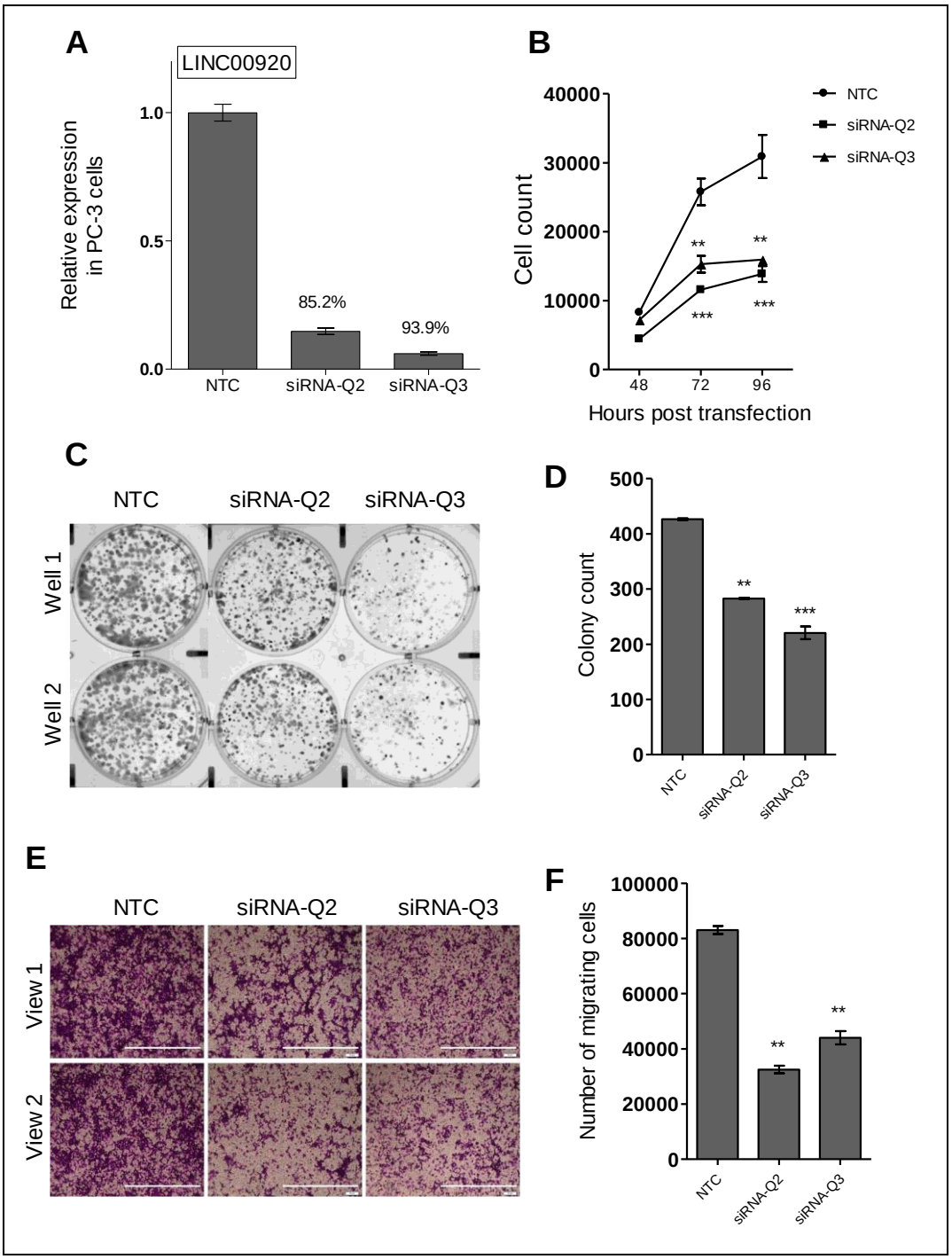
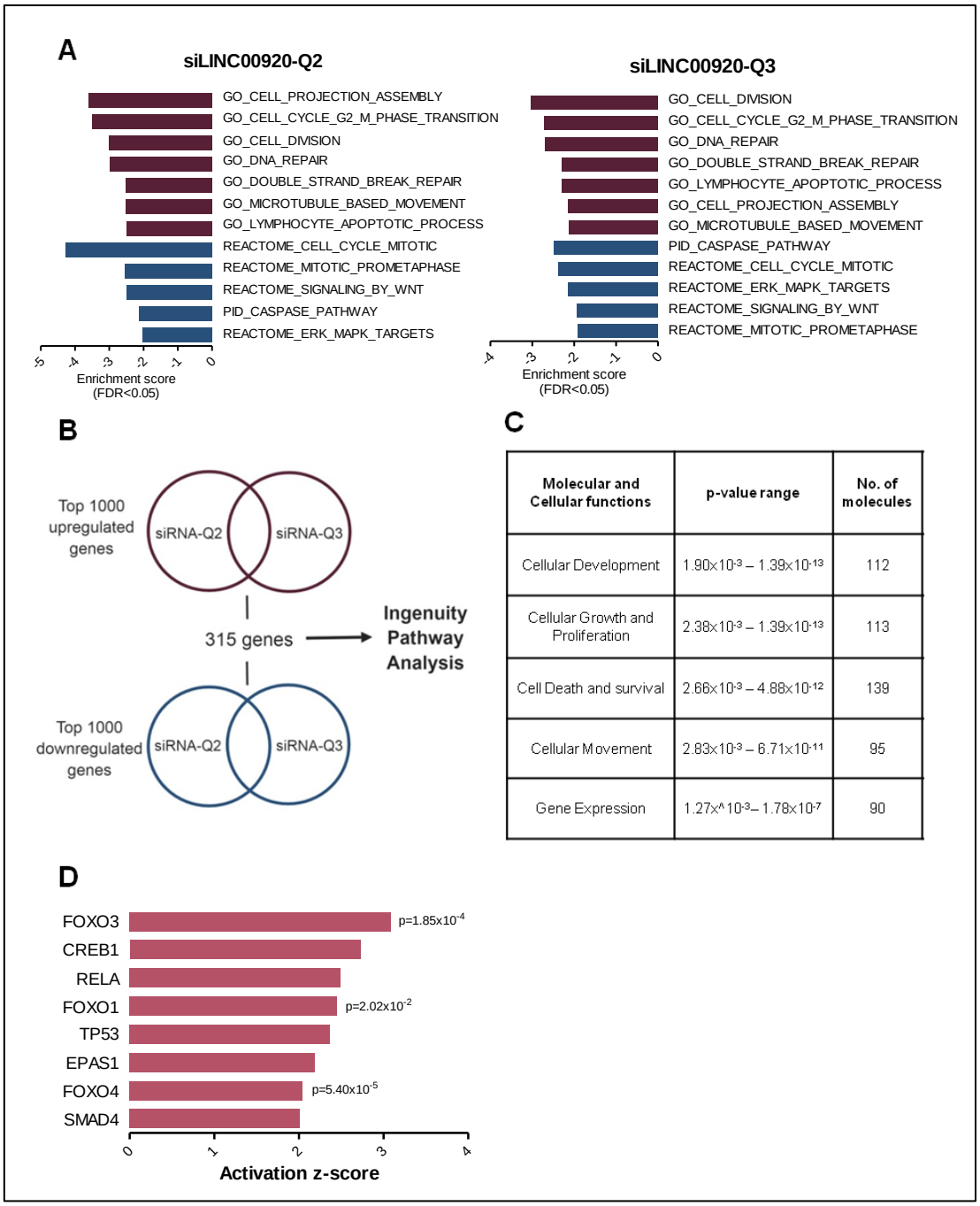


Figure 4-7. Functional assays performed on PC-3 cells upon *LINC00920* siRNA knockdown. (A) SiRNAs targeting *LINC00920* successfully reduced gene expression at the transcript level for subsequent cell assays. (B) PC-3 cells transfected with *LINC00920* siRNAs have decreased proliferative capacity beginning at 72-h post transfection. (C-D) Cells transfected with *LINC00920* siRNAs formed fewer cell colonies. (E-F) Cells transfected with *LINC00920* siRNAs had reduced migratory potential. (NTC: non-targeting control; **p≤0.01; ***p≤0.001; white bars: 300 µm)

To gain further insight on which cellular and biological processes *LINC00920* could participate in that would lead to the observed phenotypes, gene expression arrays were performed upon lncRNA knockdown. The Human HT-12 v4 Expression Bead Chip from Illumina—consisting of 47000 probes corresponding to 31000 annotated genes—was the microarray platform used. Normalized expression values were processed in two ways. First, Gene Set Enrichment Analysis (GSEA) [203] was conducted on all annotated genes ranked according to fold-change values. Knockdown of *LINC00920* by two independent siRNAs resulted in common negatively enriched curated and gene ontology (GO) gene sets. The gene sets involved processes such as cell division, cell cycle, microtubule-based movement, and apoptosis, among others (**Figure 4-8A**). Such perturbed pathways could explain the observed phenotypes upon lncRNA knockdown. A second analysis was conducted on 315 genes shared among the top 1000 upregulated and downregulated genes upon knockdown by the two siRNAs (**Figure 4-8B**). Subsequent Ingenuity Pathway Analysis (IPA) performed on these genes showed "Cellular Development", "Cellular Growth and Proliferation", "Cell Death and Survival", "Cellular Movement", and "Gene Expression" as the top molecular and cellular functions deregulated upon *LINC00920* knockdown (**Figure 4-8C**). Moreover, FOXO signaling activation was predicted in both microarray datasets particularly activation of FOXO3, FOXO1, and FOXO4 transcription factors (**Figure 4-8D**).



Molecular and Cellular functions	p-value range	No. of molecules
Cellular Development	1.90×10^{-3} – 1.39×10^{-13}	112
Cellular Growth and Proliferation	2.38×10^{-3} – 1.39×10^{-13}	113
Cell Death and survival	2.66×10^{-3} – 4.88×10^{-12}	139
Cellular Movement	2.83×10^{-3} – 6.71×10^{-11}	95
Gene Expression	$1.27\times^{\wedge}10^{-3}$ – 1.78×10^{-7}	90

Figure 4-8. Microarray analysis upon *LINC00920* knockdown in PC-3 cells. (A) Top negatively enriched curated and gene ontology gene sets based on Gene Set Enrichment Analysis (GSEA) for *LINC00920* knockdown using 2 independent siRNAs. (B) Strategy in selecting analysis genes for Ingenuity Pathway Analysis (IPA). (C) Top deregulated molecular and cellular functions and (D) top activated transcriptions factors upon *LINC00920* knockdown based on IPA.

4.7. *LINC00920* knockdown increases expression of FOXO target genes in PC-3

To first verify the effect of *LINC00920* on FOXO activity, the expression of the canonical FOXO targets *BCL2L11* [230, 231], *GADD45A* [232], and *PMAIP1* [233] was assessed upon lncRNA knockdown or overexpression in PC-3 cells. In agreement with the predicted activation of FOXO signaling, a general trend of increased FOXO target expression was observed upon *LINC00920* knockdown using five independent siRNAs (**Figure 4-9)**. Four siRNAs showed the same trend for *BCL2L11* and *GADD45a*, and 2 siRNAs for *PMAIP1*. All expression values were normalized using a scrambled siRNA control. Furthermore, episomal overexpression of *LINC00920* (pLINC00920) led to significant reduction of all FOXO targets in comparison with the empty vector control. The moderate upregulation of all FOXO targets upon *LINC00920* knockdown most likely stemmed from hyperphosphorylation of AKT due to PC-3 cells harboring a homozygous *PTEN* deletion [234, 235]. Because of this increased AKT activity, and consequentially robust FOXO phosphorylation, *LINC00920* knockdown would only be expected to have a marginal effect on the levels of FOXO targets. To test this rationale, AKT inhibition using the pan-AKT inhibitor ipatasertib [207] was performed in addition to lncRNA knockdown.

Accordingly, a compounded increase in expression of all FOXO targets was observed in *LINC00920*-knockdown cells treated with ipatasertib (**Figure 4-9**). With AKT signaling attenuated in PC-3 cells, the p-FOXO/FOXO ratio must have been within a window wherein the impact of *LINC00920* is still observable. This is in contrast to untreated cells where further *LINC00920* knockdown could not reactivate FOXO signaling due to an endogenously low FOXO activity level and an unchecked AKT pathway.

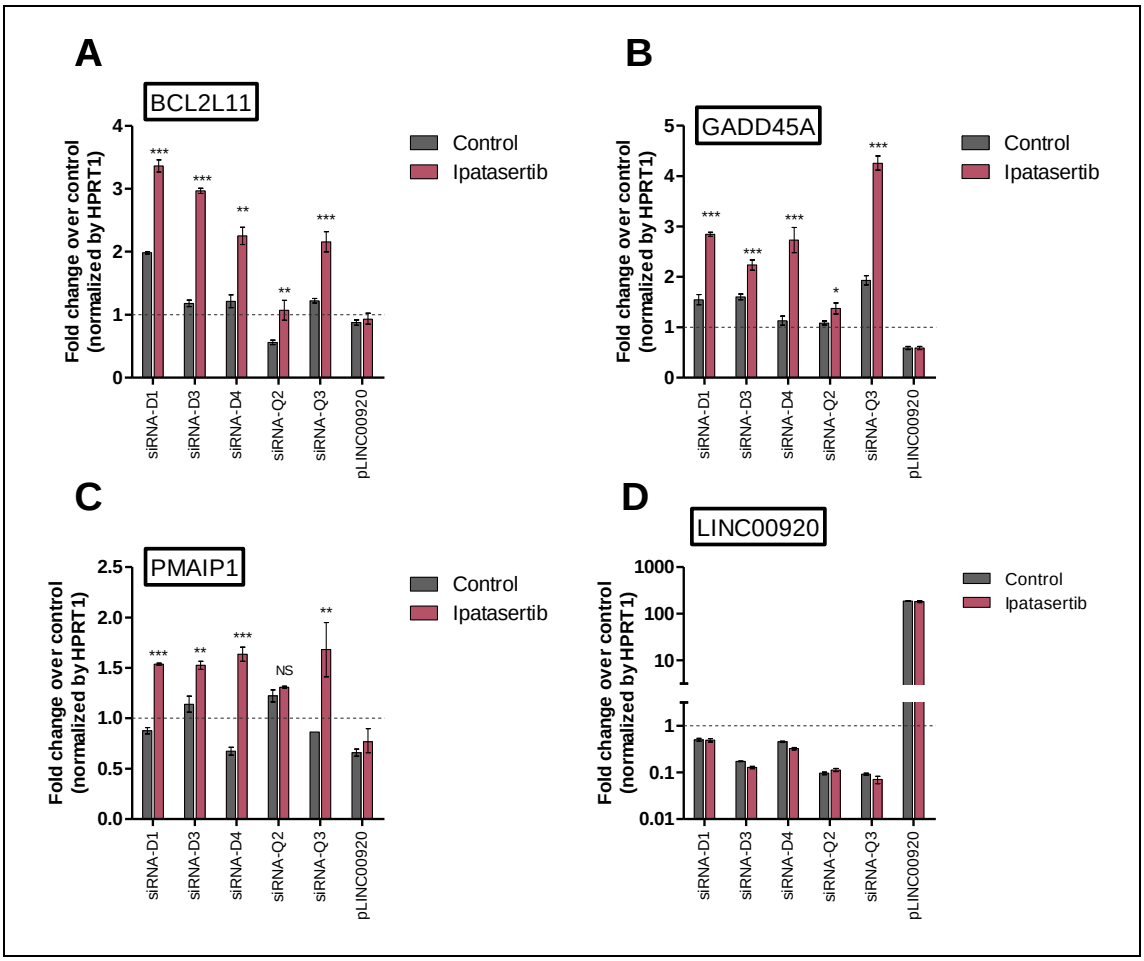


Figure 4-9. Perturbation of *LINC00920* levels in ipatasertib-treated PC-3 cells influences expression of FOXO targets. Five independent siRNAs were used to silence *LINC00920* and FOXO activity was evaluated by the expression of (A) *BCL2L11*, (B) *GADD45A*, and (C) *PMAIP1*, which are known targets of FOXO proteins. Minimal to no upregulation of FOXO targets were observed upon *LINC00920* knockdown in the control (without ipatasertib treatment). In ipatasertib-treated cells, *LINC00920* knockdown resulted in significant upregulation of all FOXO targets. Overexpressing *LINC00920* on the other hand led to the downregulation of FOXO targets in both control and ipatasertib-treated cells. (D) Quantitation of *LINC00920* levels upon siRNA-mediate knockdown and episomal overexpression. All expression values were normalized to the housekeeping gene *HPRT1*. Fold change values were calculated relative to a scrambled siRNA control. (*p≤0.05; **p≤0.01; ***p≤0.001)

Similarly, in T2E-positive VCaP cells, *LINC00920* knockdown using two independent siRNAs resulted in the increased expression of *BCL2L11*, *GADD45A*, and *PMAIP1* (**Figure 4-10**). Taken together, these experiments provide evidence for the inverse correlation between *LINC00920* transcript levels and FOXO signaling activity in prostate cancer cells.

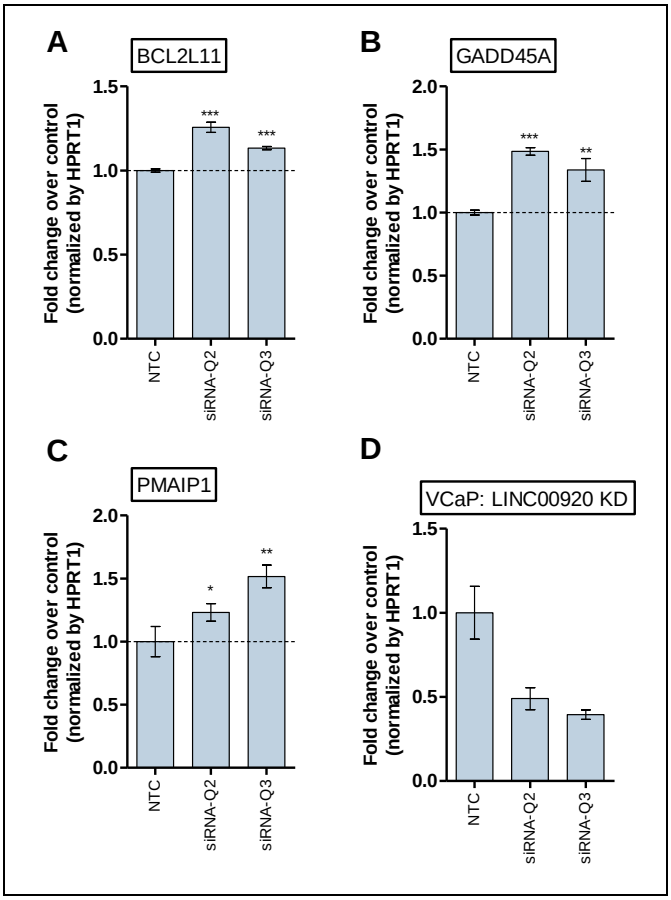


Figure 4-10. Perturbation of *LINC00920* levels in VCaP cells also influences expression of FOXO targets. Two independent siRNAs were used to silence *LINC00920* and FOXO activity was evaluated by the expression of (A) *BCL2L11*, (B) *GADD45A*, and (C) *PMAIP1*, which are known targets of FOXO proteins. Upregulation of FOXO targets were observed upon *LINC00920* knockdown using both siRNAs. (D) Quantitation of *LINC00920* levels upon siRNA-mediate knockdown. All expression values were normalized to the housekeeping gene *HPRT1*. Fold change values were calculated relative to a scrambled siRNA control. (*p≤0.05; **p≤0.01; ***p≤0.001)

4.8. The oncogenic transcription factor ERG drives *LINC00920* transcription

VCaP cells harbor a copy of the *TMPRSS2:ERG* gene fusion [236] commonly detected in TMPRSS2:ERG positive tumors (*TMPRSS2:ERGa*) [93, 94] which allows for androgen-dependent upregulation of *ERG*. ERG activity is the plausible cause for the relatively high *LINC00920* expression observed in this cell line. Indeed, expression analysis conducted for both genes in the ICGC-EOPC (n=135) and TCGA-PRAD (n=568) [75] cohorts revealed a positive correlation between *ERG* and *LINC00920*, with Pearson values of 0.57 and 0.45, respectively (**Figure 4-11**). To test the causality of this correlation, siRNA-mediated knockdown of *ERG* or *LINC00920* in VCaP cells was performed, respectively, followed by expression quantification of the other gene. While an almost 50% *ERG* knockdown resulted in a consequential downregulation of *LINC00920* (**Figure 4-12A**), knockdown of the lncRNA did not have a significant effect on *ERG* expression (**Figure 4-12B**). Furthermore, in a tet-inducible *ERG* overexpression LNCaP model [97], a concomitant increase in *LINC00920* expression was observed upon doxycycline induction in a temporal manner (**Figure 4-12C**).

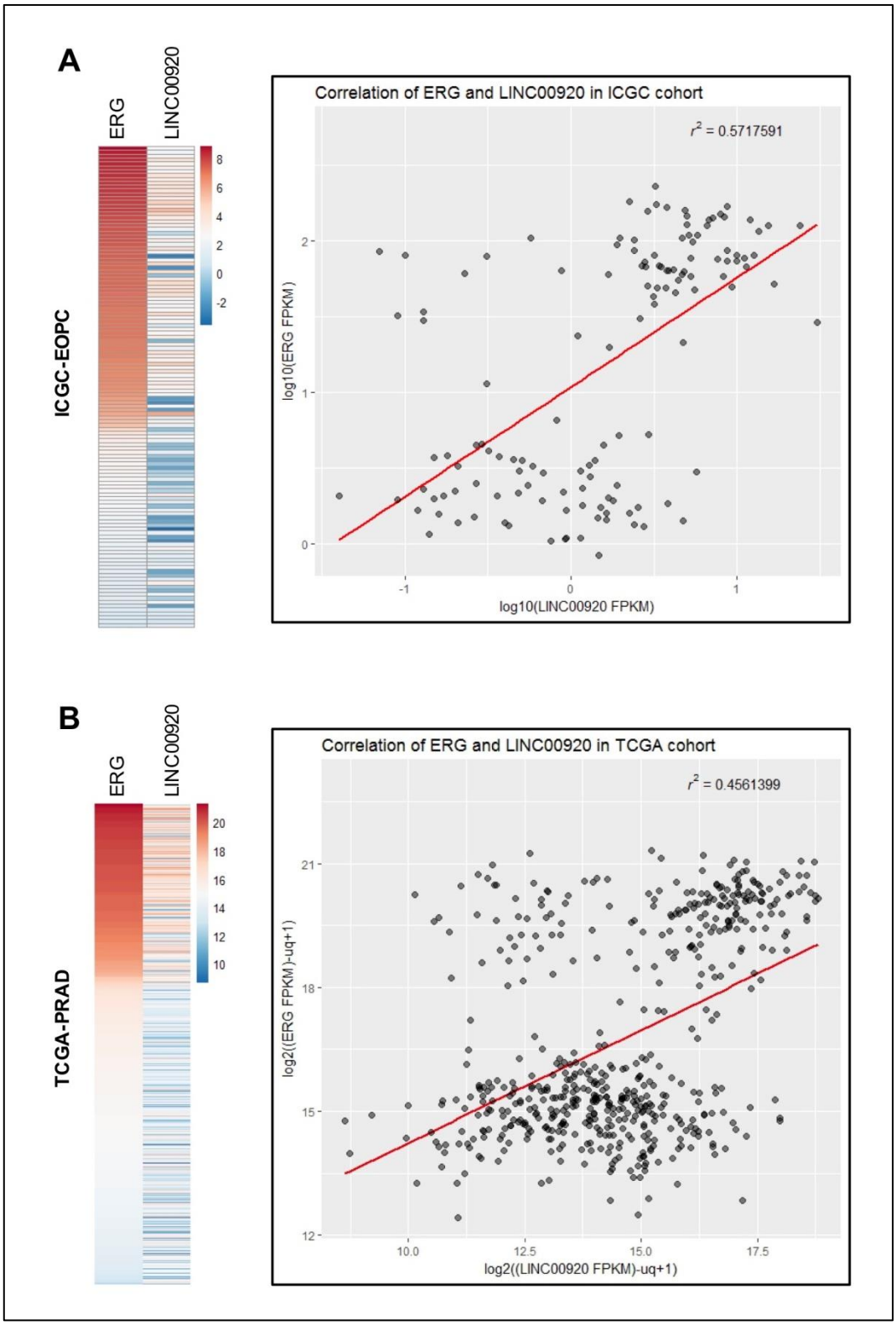


Figure 4-11. *ERG* and *LINC00920* show positive gene expression correlation. Correlation analyses performed on the (A) ICGC-EOPC and (B) TCGA-PRAD RNA-seq data. A heatmap and a scatterplot are shown for both cohorts. The Pearson correlation value is indicated at the top right corner of each scatterplot.

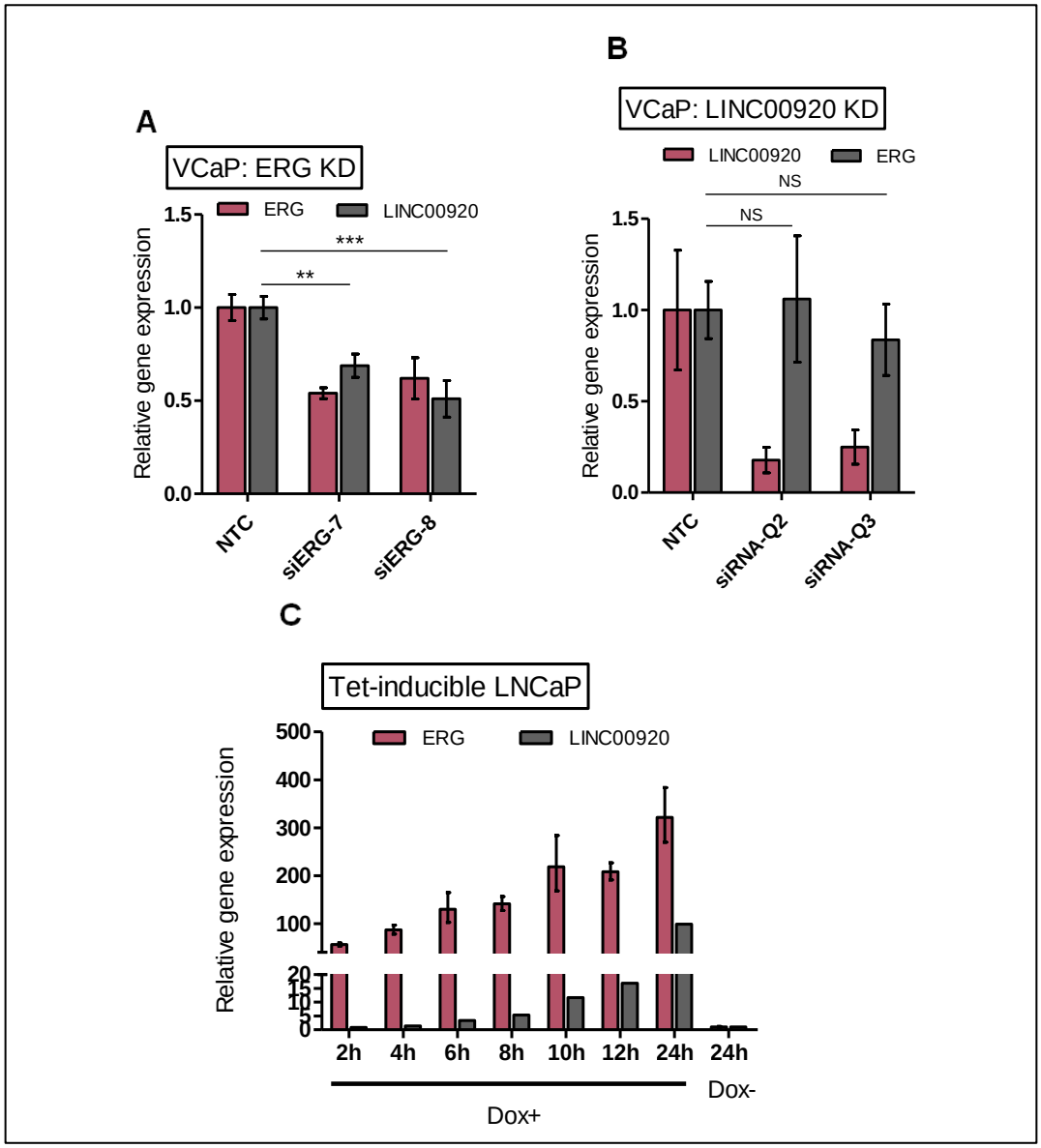
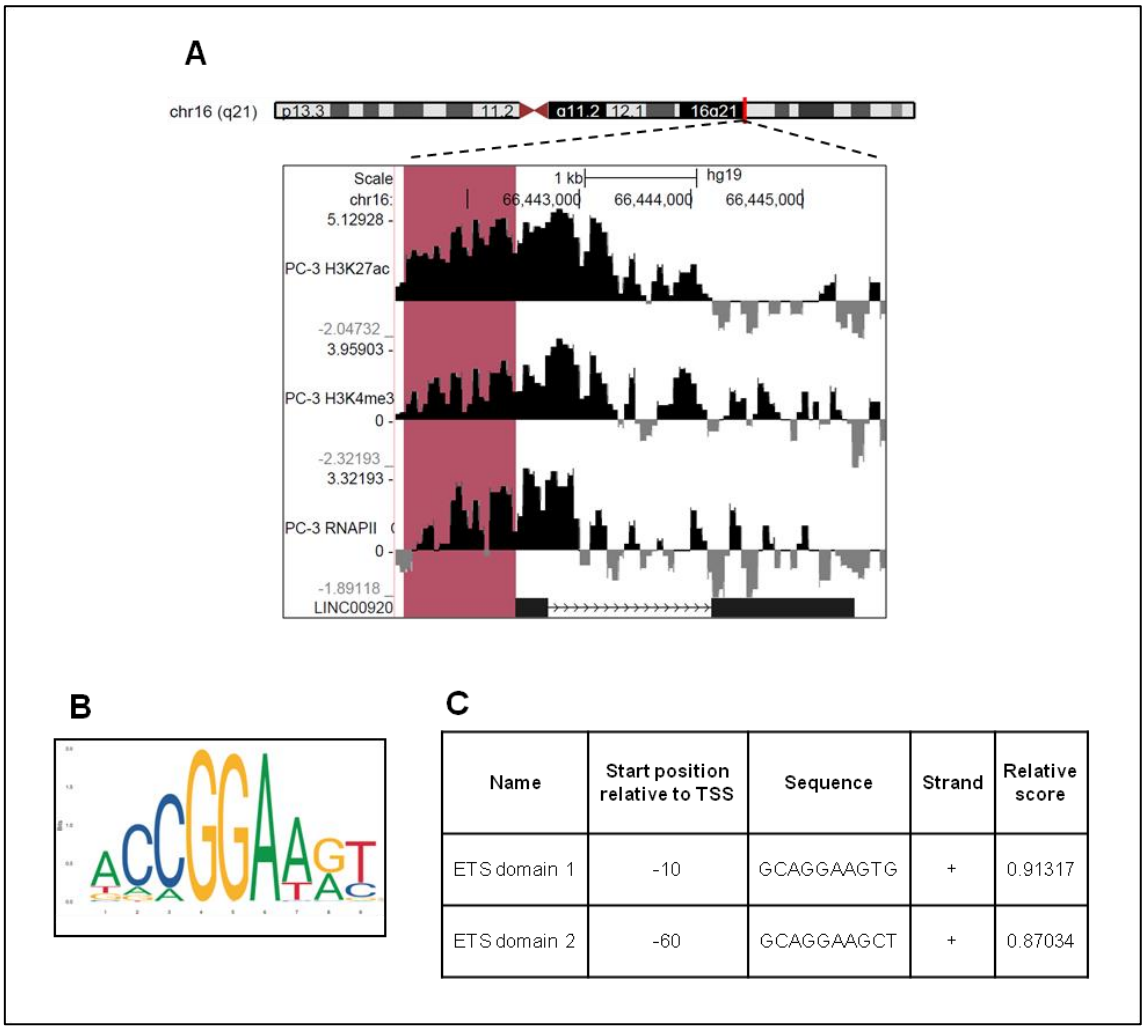


Figure 4-12. *ERG* expression directly influences *LINC00920* transcription. (A) *ERG* knockdown using two independent siRNAs significantly decreased *LINC00920* levels in VCaP cells. (B) In contrast, *LINC00920* knockdown did not perturb *ERG* expression. (C) In a tet-inducible ERG overexpression LNCaP model, progressive *LINC00920* upregulation was observed upon doxycyline induction. (**p≤0.01; ***p≤0.001)

Chromatin immunoprecipitation (ChIP) and promoter luciferase assays were performed to query the underlying mechanism behind the positive influence of ERG on *LINC00920* expression. These experiments were performed to clarify whether *LINC00920* transcription is directly controlled by ERG through its function as a transcription factor, or by a secondary ERG-dependent mechanism. Chromatin marks from published PC-3 ChIP-seq datasets [210] were examined and overlayed upstream the lncRNA TSS to guide the identification of the bona fide promoter region of *LINC00920*. The -1000 bp window showed simultaneous positive enrichments for H3K27ac, H3K3me3, and RNA polymerase II, indicating promoter activity (**Figure 4-13A**) [237]. Within this region, two putative ETS binding domains harboring the core GGA(A/T) (**Figure 4-13B**) [99] motif were predicted using

JASPAR [218] at a stringent threshold score of 85% (**Figure 4-13C**). ChIP was performed in VCaP cells to determine whether the identified ETS domains are legitimate ERG binding sites. ChIP primers were designed to be in close proximity with the two binding domains (**Figure 4-14A**). In contrast to chromatin precipitated with IgG control, ERG precipitation resulted in significant enrichment of regions amplified by all three primer pairs (**Figure 4-14B**). The specificity of enrichments was also apparent upon comparison with negative control primers targeting a non-genic region.



Name	Start position relative to TSS	Sequence	Strand	Relative score
ETS domain 1	-10	GCAGGAAGTG	+	0.91317
ETS domain 2	-60	GCAGGAAGCT	+	0.87034

Figure 4-13. *LINC00920* promoter region analysis reveals putative ERG binding sites. (A) Genomic region 1000 bp upstream of the *LINC00920* TSS (highlighted in red) show occupancy of active promoter marks H3K27ac, H3K4me3, and RNA polymerase II (RNAP II). (B) The GGA(A/T) ETS binding motif . (C) Sequence analysis of the *LINC00920* promoter region using JASPAR revealed two putative domains harboring the ETS binding motif.

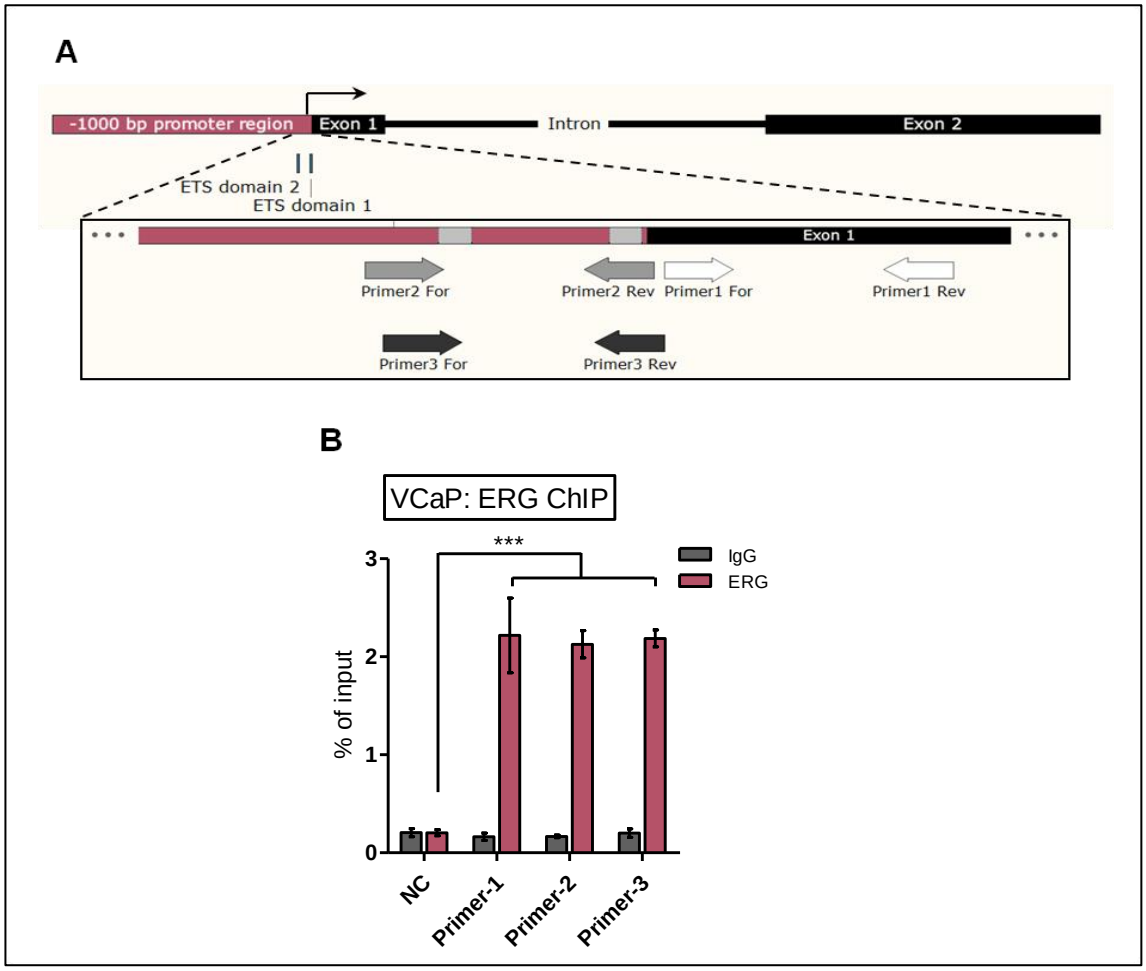


Figure 4-14. ERG chromatin immunoprecipitation at the *LINC00920* promoter in VCaP cells. (A) The gene structure of *LINC00920* with the inset showing the target regions of ChIP primers around the predicted ETS domains (gray). (B) Significant enrichment of *LINC00920* promoter fragments over a non-genic control (NC) was observed in ERG-precipitated DNA as quantified by qPCR using three primer pairs. (***$p \leq 0.001$)

Promoter luciferase assays were conducted to further characterize the ERG-mediated activation of *LINC00920* transcription. The 1000 bp promoter fragment was initially amplified from PC-3 genomic DNA. Mutagenic primers were designed to introduce a double transversion (GG>CC) in the ETS GGA(A/T) motif [98]. While a single round of PCR was sufficient to introduce the mutations within ETS domain 1, overlap-extension PCR was performed to mutate the second ETS domain which is further upstream of the 3'-end of the promoter (**Figure 4-15**). These promoter fragments, together with a 1000-bp non-genic control DNA, were cloned upstream the luciferase reporter gene *luc2* (**Figure 4-16**). The promoter constructs were cotransfected with a *Renilla luciferase* containing vector into the tet-inducible LNCaP cells described previously [97]. Luminescence was measured as the readout of promoter activity upon doxycycline induction (**Figure 4-17**). Compared with empty vector and non-genic controls, there is a significant increase in luminescence in cells transfected with the wild-type *LINC00920* promoter construct. While mutation of the ETS domain immediately upstream of the TSS (ETS domain 1) did not significantly affect the luminescence signal compared with the wild-type ($p=0.2627$), a general trend of decrease was observed in replicate experiments. On the other hand,

mutation of the ETS domain 2 resulted in significantly diminished luciferase activity compared to the wild-type construct. It is noteworthy that despite abolishing the ETS domain 2, there was still residual signal of greater intensity approaching significance ($p=0.0694$) compared to the non-genic control that could be attributed to the intact ETS domain 1. Taken together, while ETS domain 2 appears to be the preferred ERG binding region, ETS domain 1 could potentially be a supplemental binding site. These results, together with the ChIP-qPCR data, clearly show that ERG regulates *LINC00920* transcription by promoter binding, primarily at the ETS binding domain located -60 bp relative to the TSS.

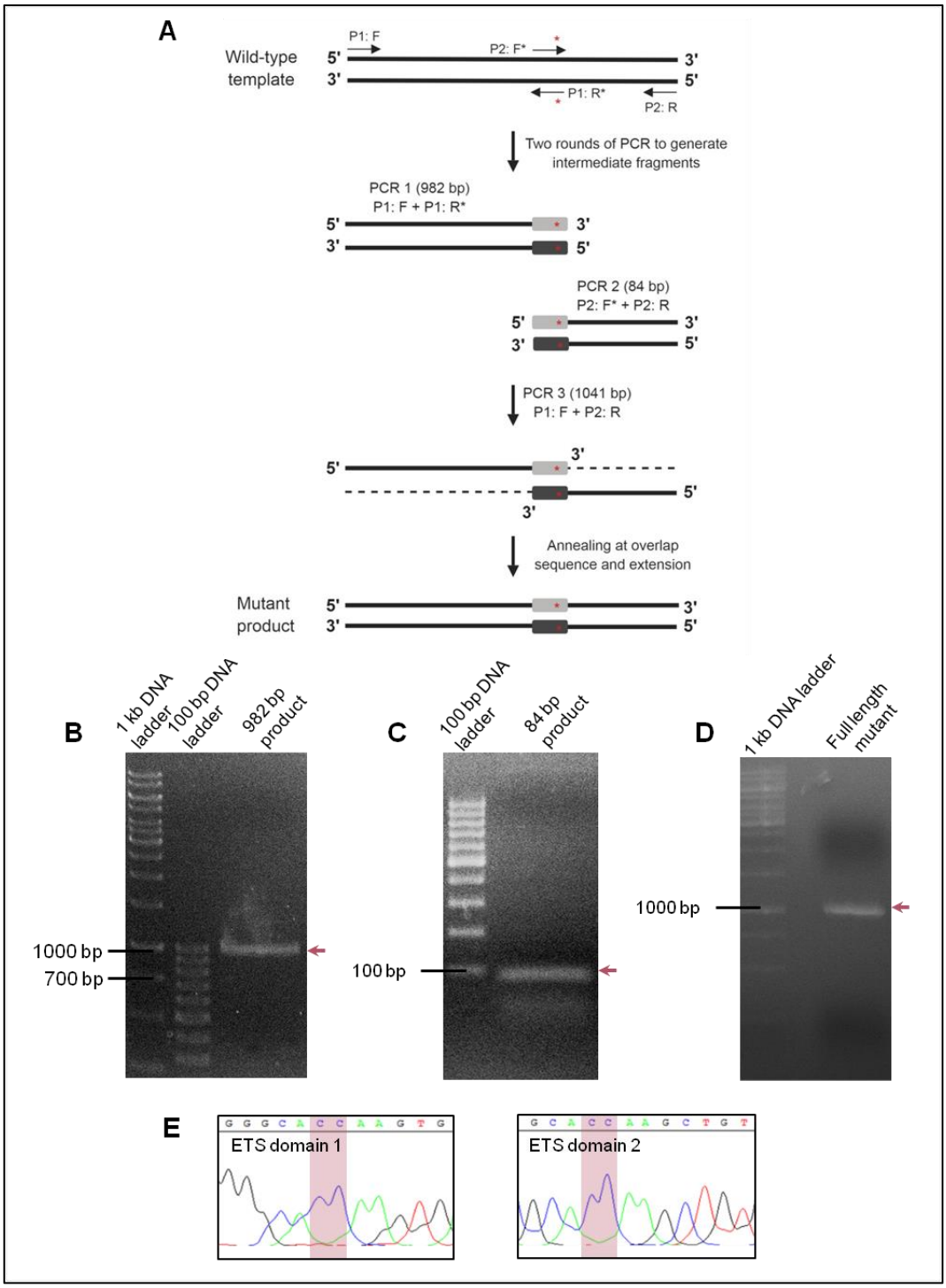


Figure 4-15. Site-directed mutagenesis of the ETS domains within the *LINC00920* promoter. (A) Overlap extension PCR was performed to mutate ETS domain 2. Briefly, a wild-type 5'-end forward primer (P1: F) and a mutant reverse primer targeting the span of the ETS domain (P1: R*) were used to generate the mutant 5'-end intermediate (PCR 1). Simultaneously, the 3'-end intermediate was generated (PCR 2) using a mutant forward primer overlapping the same region as P1: R* (P2: F*) and a wild-type 3'-end reverse primer (P2: R). The intermediate fragments were used as template in the final round of PCR (PCR 3) to generate the full-length mutant product. (B) The 982 bp, 5'-end intermediate fragment. (C) The 84 bp, 3'-end intermediate fragment. (D) The full-length mutant *LINC00920* promoter. (E) Sanger sequencing traces showing the GG>CC (highlighted) transversions introduced in the two ETS domains.

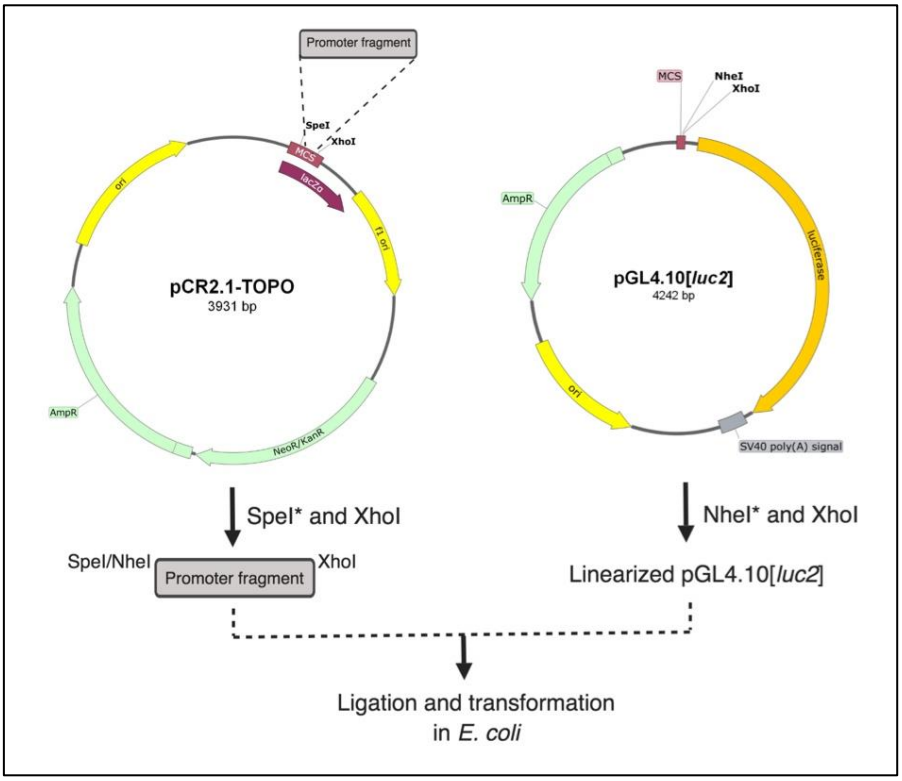


Figure 4-16. Generation of promoter constructs for luciferase assays. The wild-type and mutant *LINC00920* promoter fragments, together with a non-genic negative control, were initially cloned into pCR®2.1-TOPO® and propagated. The pCR®2.1-TOPO® constructs were digested with SpeI and XhoI to generate promoter fragments with compatible ends for subcloning. The recipient pGL4.10[*luc2*] vector was linearized using NheI and XhoI to enable ligation of promoter fragments upstream the *luciferase* gene. *SpeI and NheI produce compatible cohesive ends.

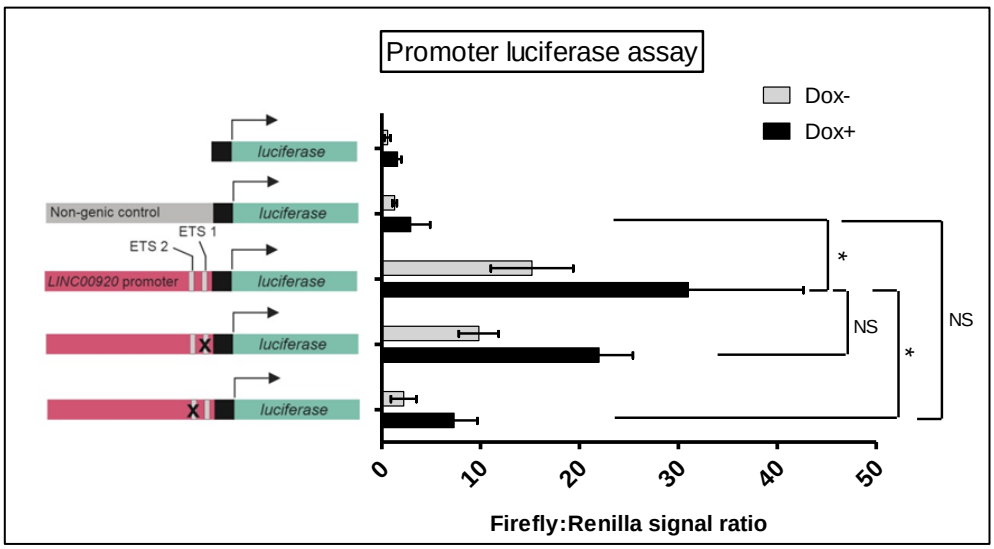


Figure 4-17. Normalized luciferase signals upon transfection of pGL4.10[*luc2*] constructs with wild-type or mutant *LINC00920* promoter fragments into tet-inducible ERG overexpression LNCaP cells. Empty vector and non-genic DNA controls did not show increased luciferase activity upon transfection. In contrast, the wild-type promoter showed a marked increase of luciferase signal. The ETS domain 1 (ETS 1) mutant tended to diminish promoter activity in comparison with the wild-type, but the shifts in signal did not reach significance. On the other hand, the ETS domain 2 (ETS 2) mutant significantly decreased promoter activation compared with the wild-type fragment, suggesting its greater relevance for ERG binding. (*$p \leq 0.05$)

4.9. The ETS family member ETV4 potentially regulates *LINC00920* expression in PC-3 cells

Since PC-3 cells do not harbor the *TMPRSS2:ERG* allele, another ETS family member most likely mediates *LINC00920* overexpression in this cell line. ETV4 has previously been described to be highly expressed in PC-3 compared to other prostatic cell lines [238]. This observation was recapitulated at the transcript level (**Figure 4-18**). SiRNA-mediated knockdown of ETV4 was performed in PC-3 cells to determine the regulatory effect of ETV4 on *LINC00920*. At least 90% ETV4 knockdown was achieved using two independent siRNAs (**Figure 4-19**). Interestingly, *LINC00920* levels were reduced to about 50% upon *ETV4* knockdown, suggesting promiscuity between ERG and ETV4 in regulating the lncRNA expression.

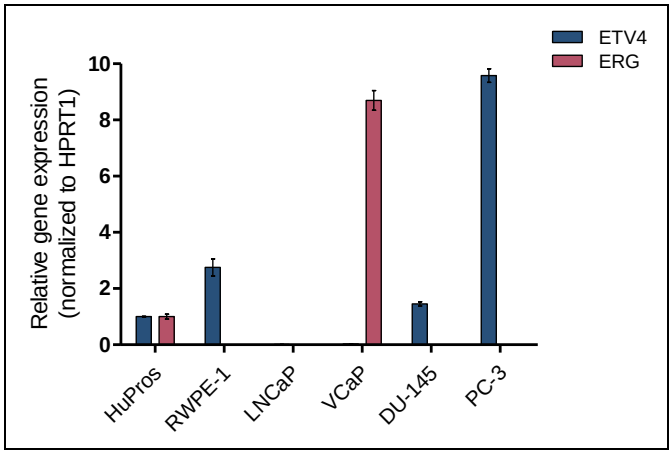


Figure 4-18. Normalized gene expression levels of ETS family members *ETV4* and *ERG* in PCa cell lines relative to normal human prostatic tissue. While the prostate epithelial cell line RWPE-1 and the metastatic line DU-145 showed modest increase ETV4 expression, PC-3 cells overexpress ETV4 approximately 10-fold higher compared with normal prostatic tissue (HuPros). Among the PCa cell lines tested, only VCaP, which harbors a *TMPRSS2:ERG* allele showed striking upregulation of *ERG*.

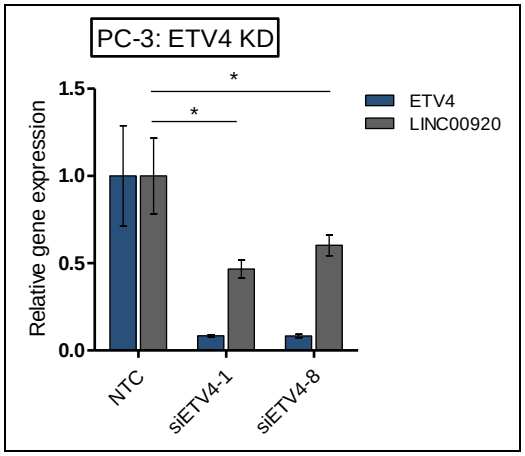


Figure 4-19. Dependence of *LINC00920* transcription level on ETV4. ETV4 knockdown using two independent siRNAs significantly decreased *LINC00920* levels in PC-3 cells. (*p≤0.05)

4.10. Mature *LINC00920* transcripts are present in the nuclear and cytosolic compartments

Similar to proteins, the function of lncRNAs is significantly tied to their subcellular localization. A selective lysis protocol which fractionates chromatin-bound, nucleoplasmic, and cytoplasmic RNA [208] was applied to PC-3, VCaP, and LNCaP cells. Quantitative PCR using primers spanning the intron of *LINC00920* was carried out to determine whether mature and processed *LINC00920* transcripts are enriched in a specific compartment. Cytosolic *HPRT1* and *GAPDH* transcripts as well as nuclear lncRNAs *NEAT1* and *MALAT1* were quantified in parallel for reference. As expected, *NEAT1* and *MALAT1* were overwhelmingly abundant in the nuclear fraction of all cell lines, with particularly high chromatin fraction enrichment in PC-3 and LNCaP cells (**Figure 4-20**). In contrast, *HPRT1* and *GAPDH* had distribution profiles of higher cytoplasmic enrichment compared to the nuclear controls in PC-3 and LNCaP cells. For all cell lines tested, *LINC00920* can be detected in all compartments, with higher nuclear distributions (nucleoplasm and chromatin) in PC-3 and VCaP. The presence of the lncRNA in the chromatin fraction, particularly in PC-3 cells, hints at potential functional interaction with chromosomal regions. The distribution of *LINC00920* across all compartments also raises the possibility of dynamically shuttling between the nucleus and the cytoplasm which could be a cause or a consequence of a particular cellular process.

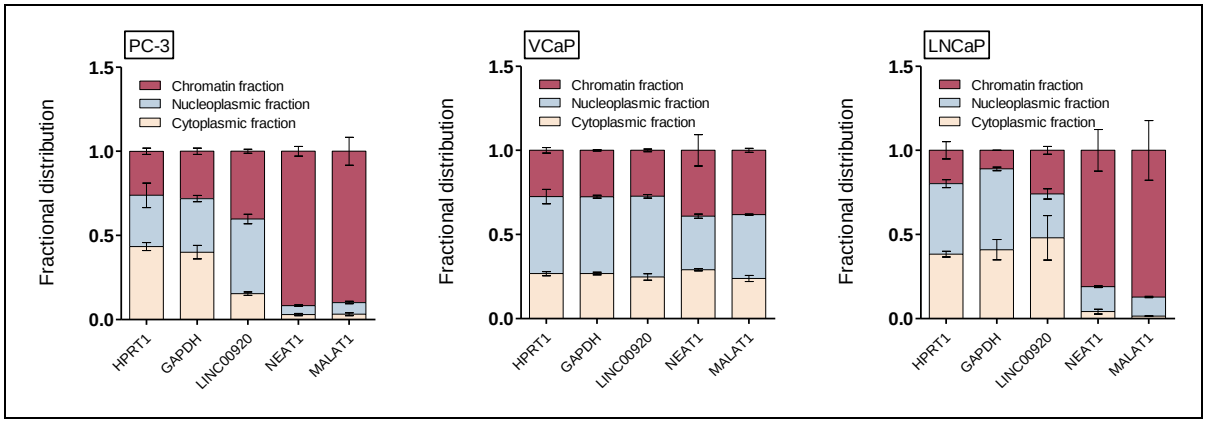


Figure 4-20. Subcellular distribution of *LINC00920* in selected PCa cell lines. *LINC00920* levels were compared with cytoplasmic-enriched (i.e., *HPRT1*, *GAPDH*) and nuclear enriched (i.e., *NEAT1*, *MALAT1*) transcript controls.

4.11. Putative role of *LINC00920* at enhancer regions as revealed by chromatin isolation by RNA purification-high throughput sequencing (ChIRP-seq)

Because a considerable fraction of mature *LINC00920* transcripts were chromatin associated, it was reasonable to hypothesize that the lncRNA could have functional consequences on chromatin structure maintenance or on gene regulation. To address this, chromatin isolation by RNA isolation (ChIRP) [194] (**Figure 4-21**) was performed on PC-3 cells where endogenous *LINC00920* expression was among the highest in the PCa cell lines tested. Chromatin shearing conditions were optimized using Bioruptor® Pico. For 6 x 10^6 cells suspended in 300 µL lysis buffer, 35 sonication cycles were required to obtain the ideal uniform RNA length distribution between 100 to 1000 nt (**Figure 4-22A**)

[128, 194]. Simultaneously, this degree of shearing also led to chromatin solubilization. A considerable fraction of the chromatin was reduced down to 100 to 300 bp, peaking around the nucleosomal length of about 150 bp (**Figure 4-22B**).

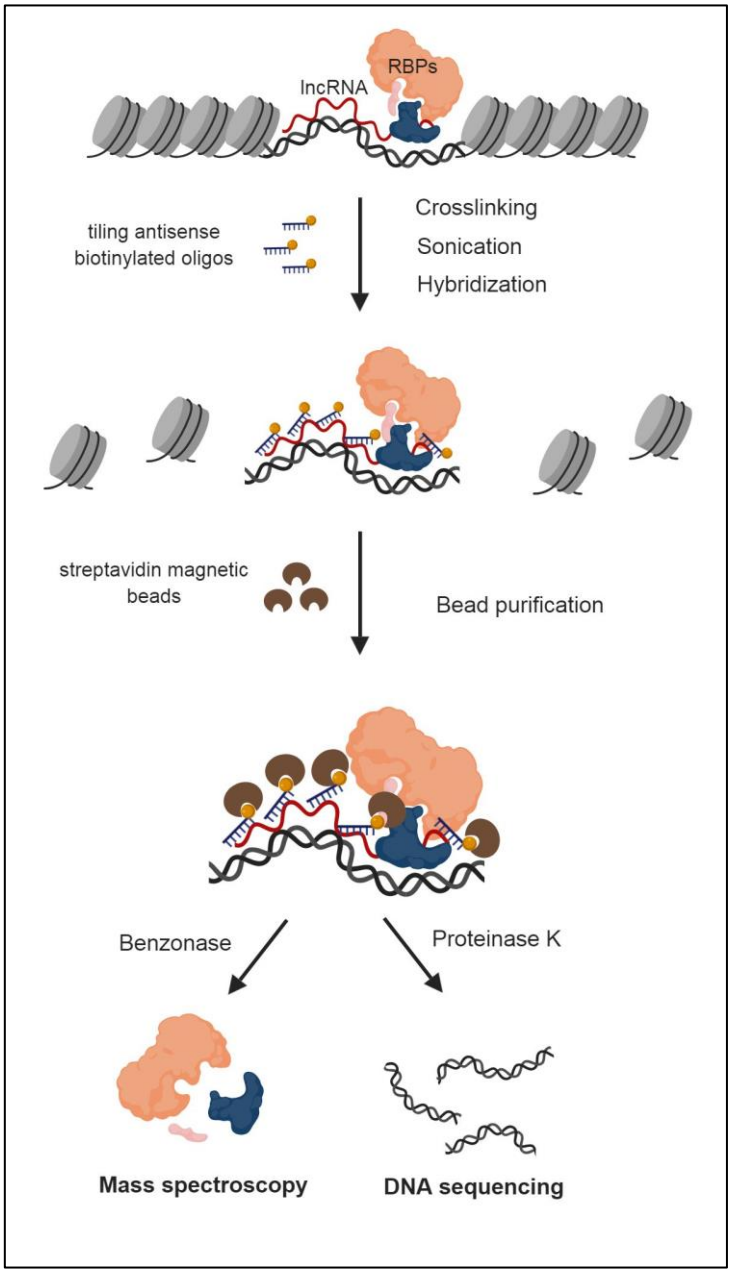


Figure 4-21. Chromatin isolation by RNA precipitation (ChIRP). Prior to pulldown, cells are crosslinked to preserve RNA interactions and then sonicated for chromatin solubilization. The resulting lysate is hybridized with a pool of tiling, biotinylated 20-mer DNA oligonucleotides antisense to the target transcript. Purification using streptavidin magnetic beads enriches target RNA complexes. For subsequent mass spectrometry, proteins are eluted by benzonase digestion. For DNA preparation, nucleic acids are eluted by proteinase K digestion, followed by DNA precipitation.

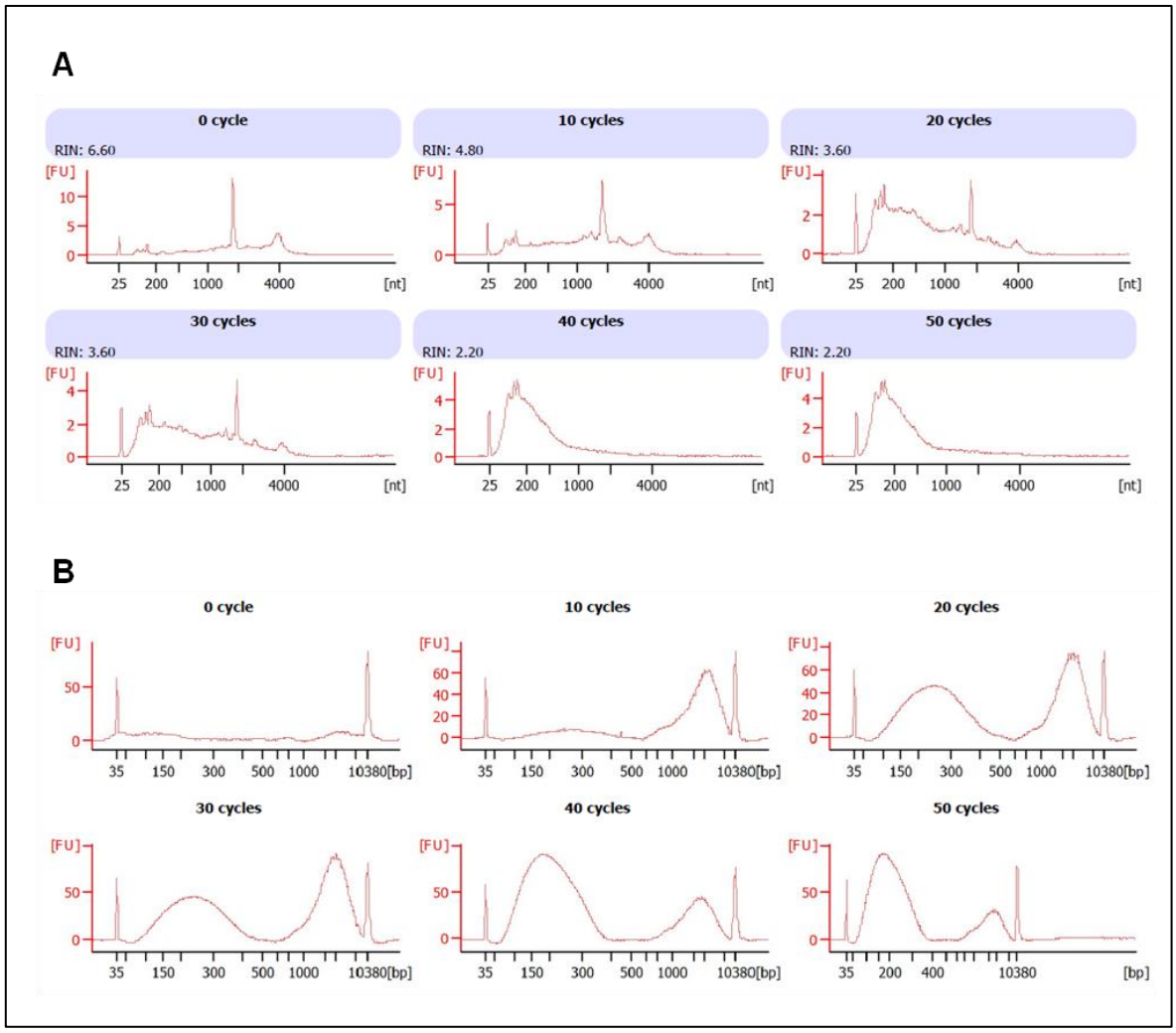


Figure 4-22. Optimization of sonication conditions for crosslinked PC-3 cells. Distribution of (A) RNA and (B) chromatin lengths were monitored at the indicated number of shearing cycles. Each shearing tube contains 6 million cells in 300 µL lysis buffer. The optimal cycle number was taken to be 35—the intermediary between 30 and 40 cycles—which would shear RNA and chromatin down to the ideal fragment range.

Thirty 20-nt biotinylated single stranded DNA (ssDNA) oligos antisense to the target transcript (**Figure 4-23A**) were used to establish the RNA pulldown protocol. In small-scale experiments consisting of 2 x 10^6 cells per pulldown, transcript enrichment was determined by quantitative PCR. Oligos targeting the *lacZ* transcript, which is normally absent in human cells, were used as negative control. The *lacZ* oligos did not enrich either *MALAT1* or *LINC00920* after RNA pulldown. In contrast, highly selective lncRNA enrichments were achieved for both *LINC00920* (**Figure 4-23B**) and *MALAT1* (**Figure 4-23C**) oligos.

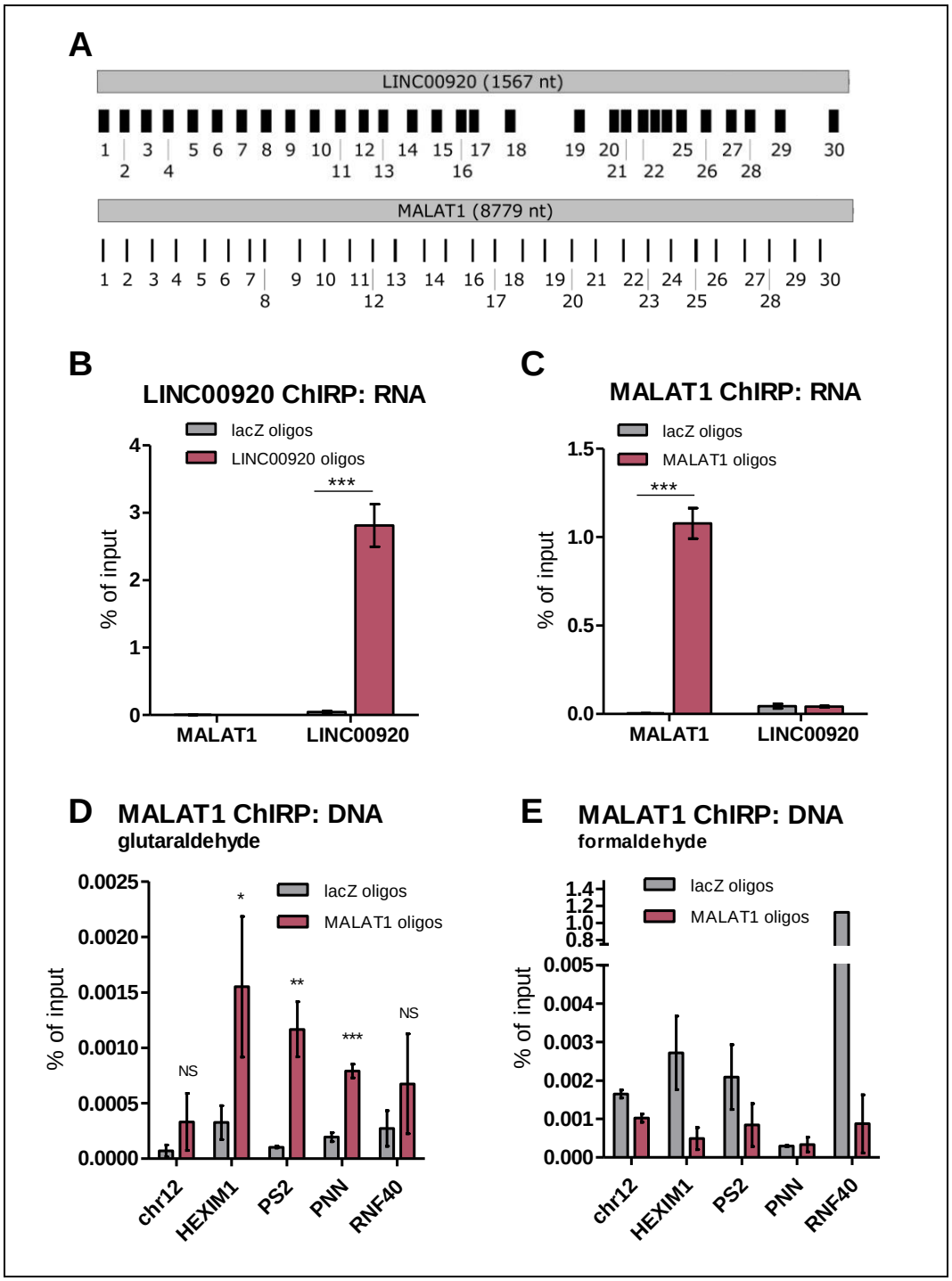


Figure 4-23. Establishment of the ChIRP protocol by targeting *LINC00920* and *MALAT1* lncRNAs. (A) Pools of 30 biotinylated, antisense, 20-mer DNA oligos were used to ChIRP *LINC00920* and *MALAT1*. Alignments of the oligos along the full transcript are shown for both lncRNAs. For sequences, see **Table 2-12**. Significant specific enrichments for *LINC00920* (B) and *MALAT1* (C) were achieved using the pooled oligos in glutaraldehyde-fixed PC-3 cells. (D) Formaldehyde crosslinking did not result in enrichments for known *MALAT1* chromatin targets. (E) Glutaraldehyde crosslinking showed considerable improvement in enriching the same *MALAT1* targets. (*p≤0.05; **p≤0.01; ***p≤0.001)

All RNA methods aiming to map chromatin binding sites begin with a cross-linking step critical in preserving the chromosomal location of the target transcript [187]. Both formaldehyde and glutaraldehyde are chemical cross-linkers that have been used to capture *in vivo* RNA interactions [128, 192, 194]. Due to its small size, formaldehyde can preserve interactions within a 2 Å range, making it an ideal reagent for identifying molecules associating in close proximity [239]. Additionally, formaldehyde cross-links are reversible upon mild heating in an appropriate buffer, allowing retrieval of the interacting components. On the other hand, glutaraldehyde mediates irreversible cross-linking at longer distances owing to the carbon spacers present between the two aldehyde groups. To determine which chemical cross-linker would work best in capturing lncRNA-chromatin interactions, ChIRP using formaldehyde or glutaraldehyde was initially performed on *MALAT1*, a well-studied lncRNA with published chromatin targets [193]. Glutaraldehyde-ChIRP showed significant enrichment of *MALAT1* targets *HEXIM1*, *PS2*, and *PNN*, with a similar trend of enrichment observed for another target, *RNF40*, compared with the lacZ control pulldown (**Figure 4-23E**). In contrast, formaldehyde-ChIRP did not enrich any of the *MALAT1* targets (**Figure 4-23F**). With these results, subsequent *LINC00920*-DNA capture experiments were performed using glutaraldehyde as the cross-linker.

Upscaled glutaraldehyde-ChIRP experiments using 8 x 10^7 cells per pulldown were performed to obtain sufficient input DNA material for library preparation using the NEBNext Ultra II DNA library preparation kit. Without prior DNA size selection, the resulting library size for all samples peaked between 300 and 400 bp (**Figure 4-24**). This can be rationalized by considering the addition of an adaptor, a linker, and index primers with a combined length of 184 bp to the approximately 150 bp-fraction of sheared chromatin discussed previously. The resulting 6 libraries—a triplicate of *lacZ*-precipitated DNA, and a triplicate of *LINC00920*-precipitated DNA—were pooled in equimolar amounts and sequenced on the HiSeq 2000 Illumina platform with a single-end 50-bp read length. **Table 4-3** shows the read counts attributed to each demultiplexed library based on unique barcodes introduced by library-specific index primers. The raw sequencing data were processed and analyzed as outlined in **Figure 4-25A**. To assess the read coverage reproducibility among the replicates, signal correlation was made for each aligned BAM file (**Figure 4-25B**). The triplicates for both *lacZ* and *LINC00920* showed very high correlation with each other, and the samples were distinctly clustered based on the precipitation condition (i.e., *lacZ* or *LINC00920* capture). This correlation can also be visualized on the Integrative Genomics Viewer (IGV) [90] where peak profiles can be seen to be similar among the replicates (**Figure 4-25C**).

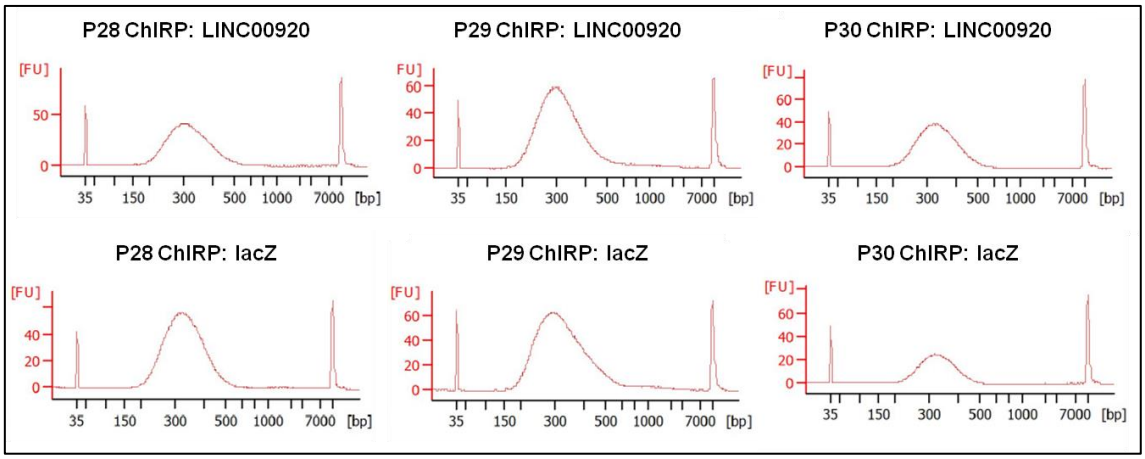


Figure 4-24. Bioanalyzer traces of sequencing libraries prepared from ChIRP DNA. NEBNext®Ultra™ II DNA Library Prep Kit was used to construct libraries from three biological replicates of *LINC00920*- and *lacZ*-precipitated DNA. Each library had a single peak centered between 300 to 400 bp corresponding to the combined lengths of the sequencing adapters, index primers, and ChIRP DNA insert.

Table 4-3. Read count statistics of sequenced ChIRP-precipitated DNA libraries.

FASTQ file	Sample ID	Base Count	Read Count	Barcode
AS-207375-LR-31363_R1.fastq.gz	P28 ChIRP: lacZ	1519529190	29794690	ACAGTG
AS-207377-LR-31363_R1.fastq.gz	P28 ChIRP: LINC00920	1279734738	25092838	GCCAAT
AS-207381-LR-31363_R1.fastq.gz	P29 ChIRP: lacZ	1839903642	36076542	TAGCTT
AS-207383-LR-31363_R1.fastq.gz	P29 ChIRP: LINC00920	2038382025	39968275	CTTGTA
AS-207385-LR-31363_R1.fastq.gz	P30 ChIRP: lacZ	1632430491	32008441	CGATGT
AS-207387-LR-31363_R1.fastq.gz	P30 ChIRP: LINC00920	1486910865	29155115	TGACCA

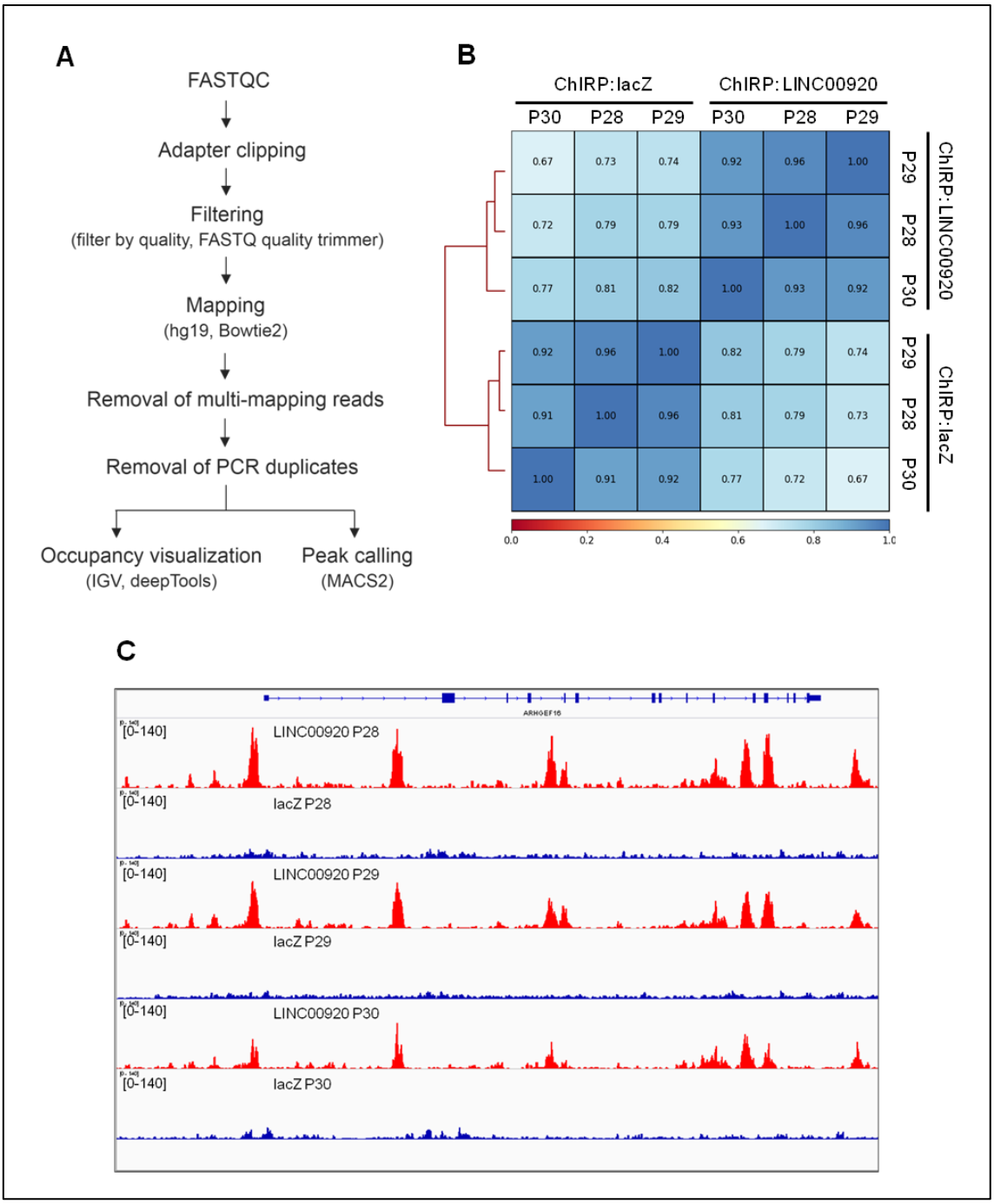


Figure 4-25. Initial processing and quality control assessment of ChIRP-seq data. (A) The workflow adapted to prepare the raw ChIRP-seq data for occupancy analysis and peak calling. (B) The read coverage correlation among samples showed high reproducibility for each pulldown. (C) A representative view of aligned reads in the IGV browser revealing robust signal reproducibility among the triplicate samples. Red tracks correspond to *LINC00920* ChIRP while blue tracks correspond to *lacZ* ChIRP.

To answer the question of whether *LINC00920* binds directly to genic regions and consequently exert proximal gene regulatory function, normalized *LINC00920* read coverage was overlapped with annotated genes (hg19). On average, *LINC00920* traces were minimal across gene bodies with notable depletion around transcription start sites (TSSs) (**Figure 4-26**).

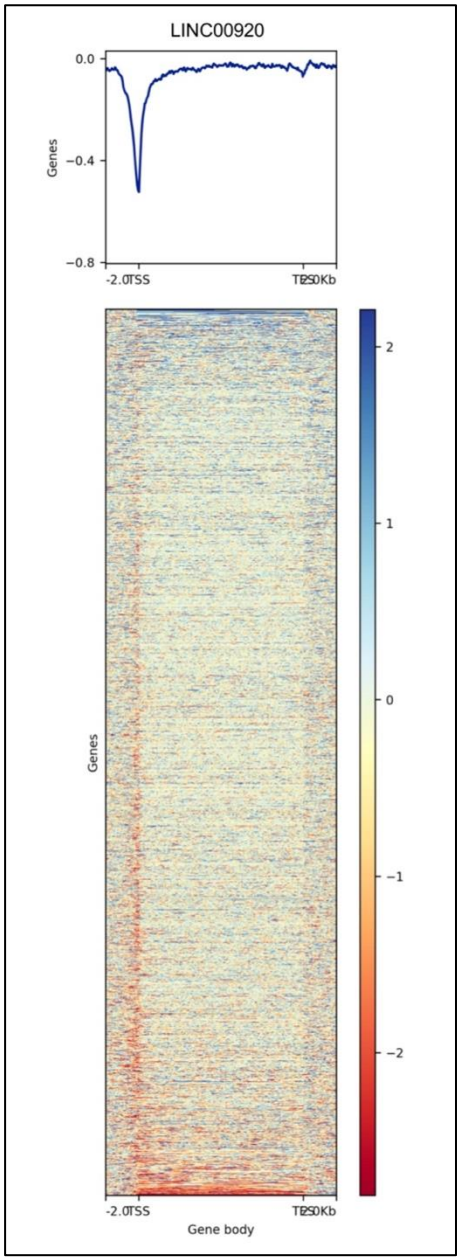
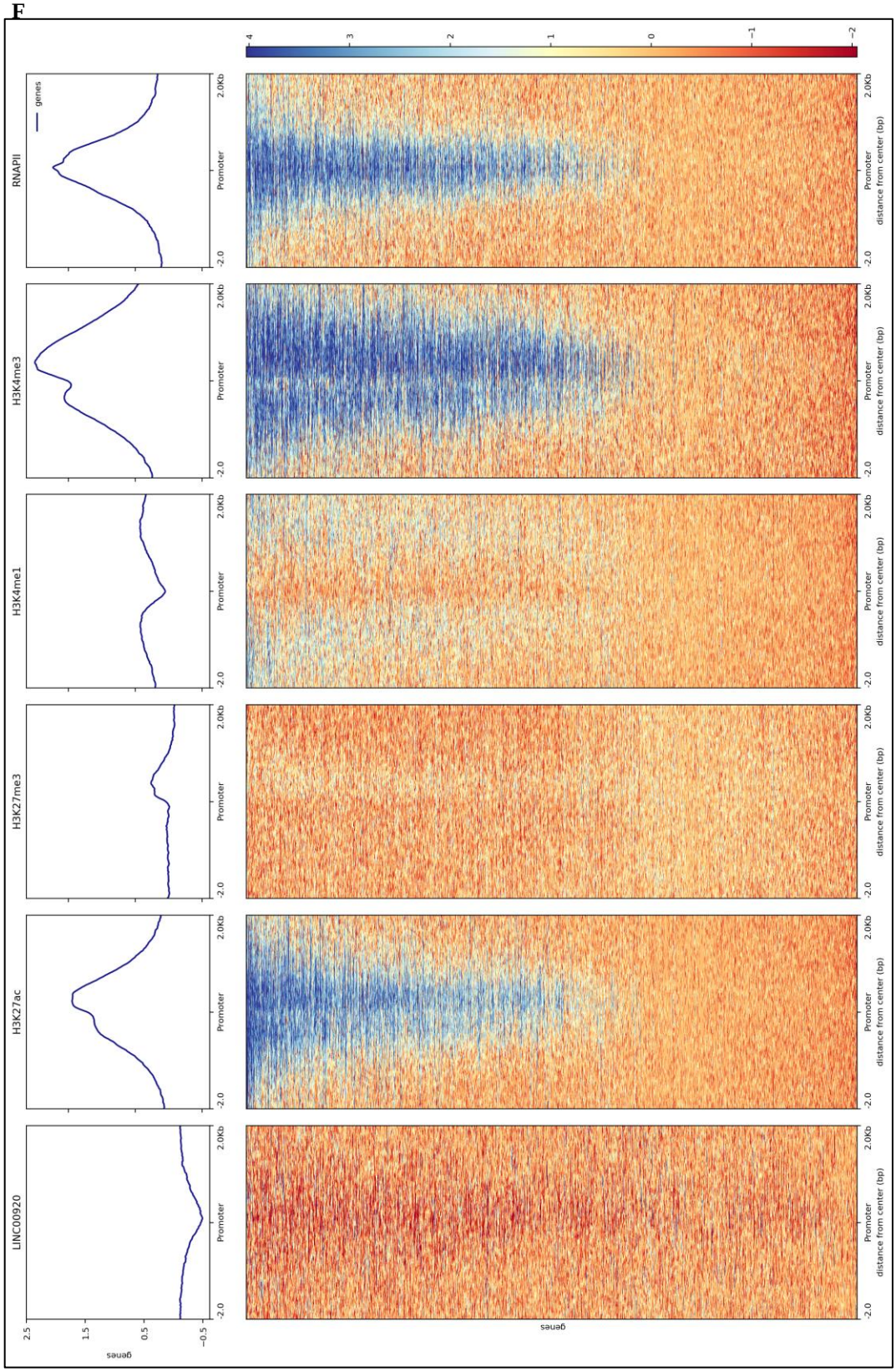


Figure 4-26. Normalized *LINC00920* read coverage across genic regions in the human genome build hg19. Top: Average profiles of *LINC00920* occupancy on genes normalized by length of 10 kbp with 2-kb extensions upstream the TSSs and downstream the TESs. Bottom: Heatmaps showing signals for individual genes (heatmap rows). Shown data was derived from a representative replicate. Blue: high read coverage, red: low read coverage.

As histone modifications are tightly linked to chromatin state and transcriptional status, read densities of *LINC00920* ChIRP-seq at TSSs were visualized vis-à-vis published PC-3 ChIP-seq datasets [210] for (i) H3K27ac, a marker of enhancers and promoters of active genes; (ii) H3K27me3, a repressive mark; (iii) H3K4me1, a marker found at transcriptional enhancers; (iv) H3K4me3, a marker of gene promoters; and (v) RNA polymerase II, a marker of active transcription (**Figure 4-27**). As expected, presumably active promoters enriched in RNA polymerase II showed positive H3K27ac and H3K4me3 occupancy without H3K27me3 enrichment. Interestingly, H3K4me1 marks showed comparably reduced signal profile resembling *LINC00920* density around TSSs.

F


LINC00920
H3K27ac
H3K27me3
H3K4me1
H3K4me3
RNAPII
genes
Promoter
2.0Kb
-2.0
distance from center (bp)

H3K4m

e1, H3K4me3, and RNA polymerase II (RNAP II) read coverages across promoter regions defined in the PC-3 genome. Top: Average profiles of *LINC00920*, H3K27ac, H3K27me3, H3K4me1, H3K4me3, and RNAP II occupancy on promoter centers (n=36,180) with 2-kb extensions upstream and downstream. Bottom: Heatmaps showing signals for individual genes (heatmap rows). Blue: high read coverage, red: low read coverage.

In an effort to isolate *LINC00920*-enriched promoter regions for further investigation, all 36,180 annotated TSSs were clustered based on read densities. Among the resulting four region clusters, cluster 1, corresponding to 5,287 loci, had the highest positive lncRNA association (**Figure 4-28**). Looking at the presence of RNA polymerase II and histone modifications in these regions, a number of interesting observations can be made (**Figure 4-29**). First, there is an overall decrease in RNA polymerase II occupancy indicating decreased transcription. Second, H3K4me3 peak density was narrower, possibly corroborating the diminished transcriptional activity in these regions. The final and most intriguing pattern was the coincident deposition of H3K4me1 in regions of high *LINC00920* occupancy. Moreover, since H3K27ac occupancy was invariantly high in this cluster, the emergence of H3K4me1 marks becomes predictive of enhancer activity [240, 241] in these *LINC00920*-associated loci.

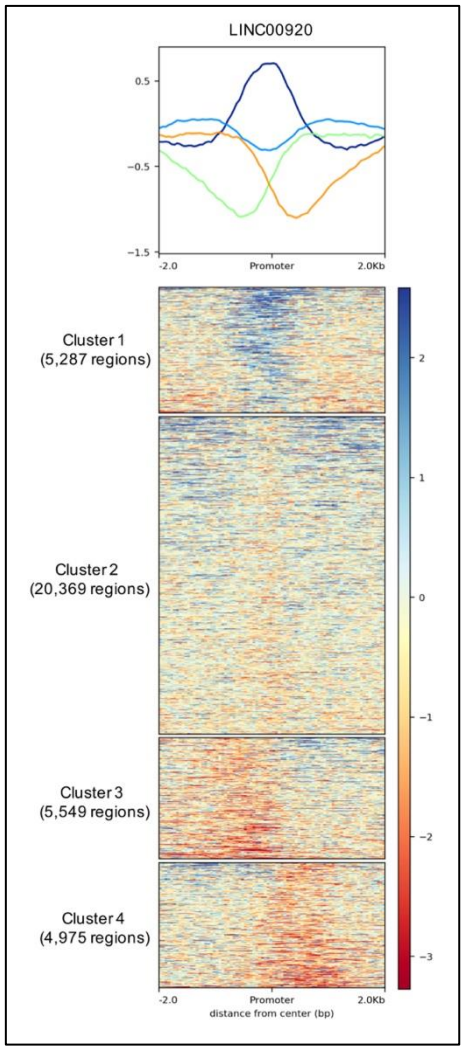


Figure 4-28. ChIRP-seq density clustering reveals *LINC00920* binding to a subset of promoters. Top: Average profiles of *LINC00920* occupancy across clustered promoters. Bottom: Segmented heatmaps showing clustered promoters based on signal density. Promoter

regions in cluster 1 exhibit positive *LINC00920* enrichment in all replicates. Cluster 2 promoters show modest *LINC00920* occupancy both upstream and downstream the promoter center. Clusters 3 and 4 promoters exhibit depletion of *LINC00920* signal upstream and downstream the promoter center, respectively. Shown data was derived from a representative replicate. Blue: high read coverage, red: low read coverage.

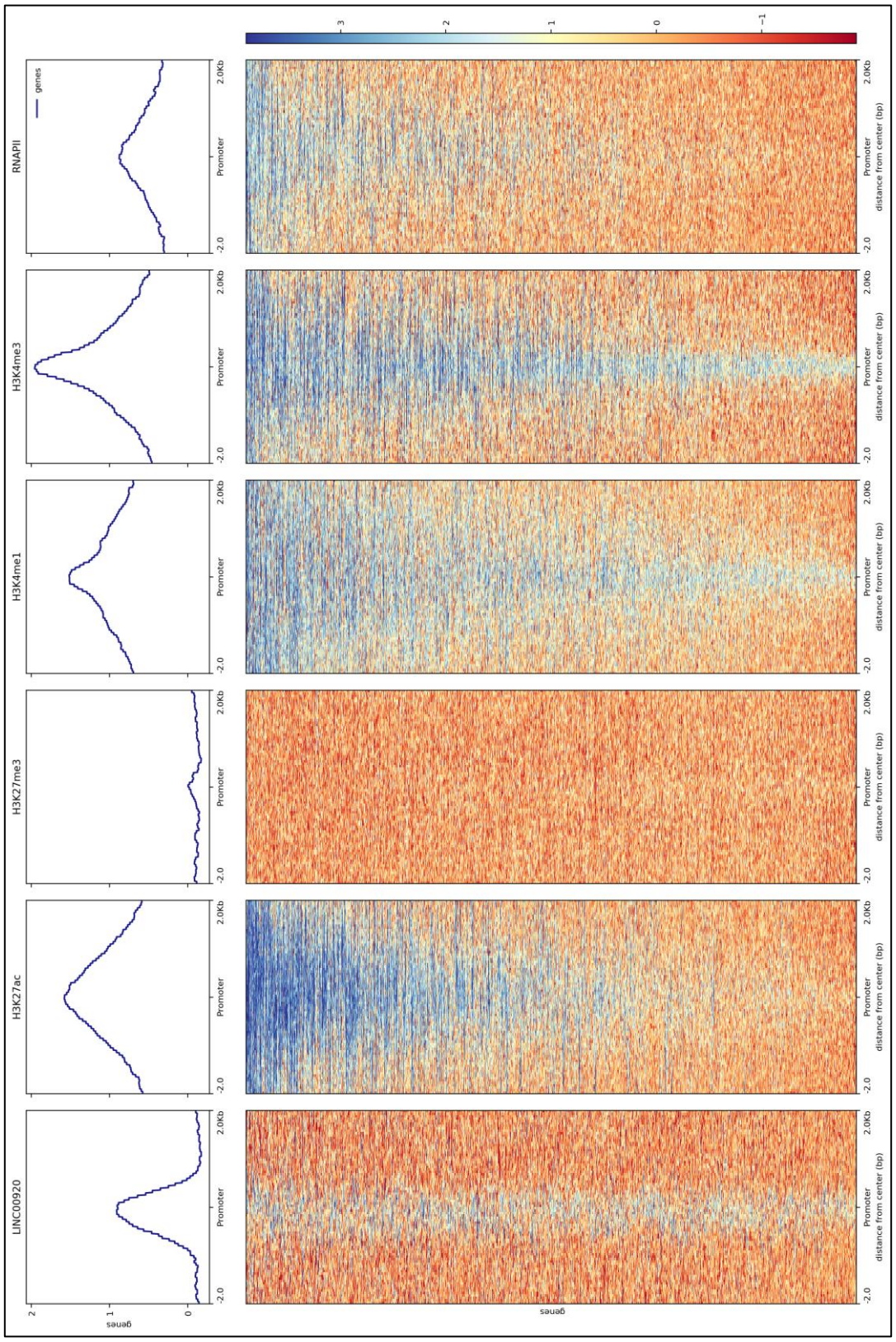


Figure 4-29. Normalized *LINC00920*, H3K27ac, H3K27me3, H3K4me1, H3K4me3, and RNA polymerase II (RNAP II) read coverages across cluster 1 promoter regions. Top: Average profiles of *LINC00920*, H3K27ac, H3K27me3, H3K4me1, H3K4me3, and RNAP II occupancy across 5,287 regions. Bottom: Heatmaps showing signals for individual promoter region (heatmap rows). Blue: high read coverage, red: low read coverage.

The biological significance of *LINC00920* interaction with these regions was queried using the Genomic Regions Enrichment of

Annotations Tool (GREAT) which assigns biological meaning to a set of genomic regions by analyzing the annotations of nearby genes [202]. The term "LKB1 signaling events" was the top-most hit among the molecular signatures included in the database (**Figure 4-30A**) [203, 221]. The *LKB1* gene (also known as *STK11*) encodes a serine/threonine kinase that phosphorylates and activates 5' AMP-activated protein kinase (*AMPK*) [242]. The LKB1-AMPK axis is involved in a complex network of metabolic pathways that ultimately control cell growth in response to environmental nutrient changes. To specify the pathways most affected within the LKB1-AMPK network, Ingenuity Pathway Analysis was performed on the *LKB1* gene set using the *LINC00920* knockdown microarray data. This analysis revealed that upon *LINC00920* silencing, PI3K/AKT and 14-3-3 signaling pathways were deactivated (**Figure 4-30B**).

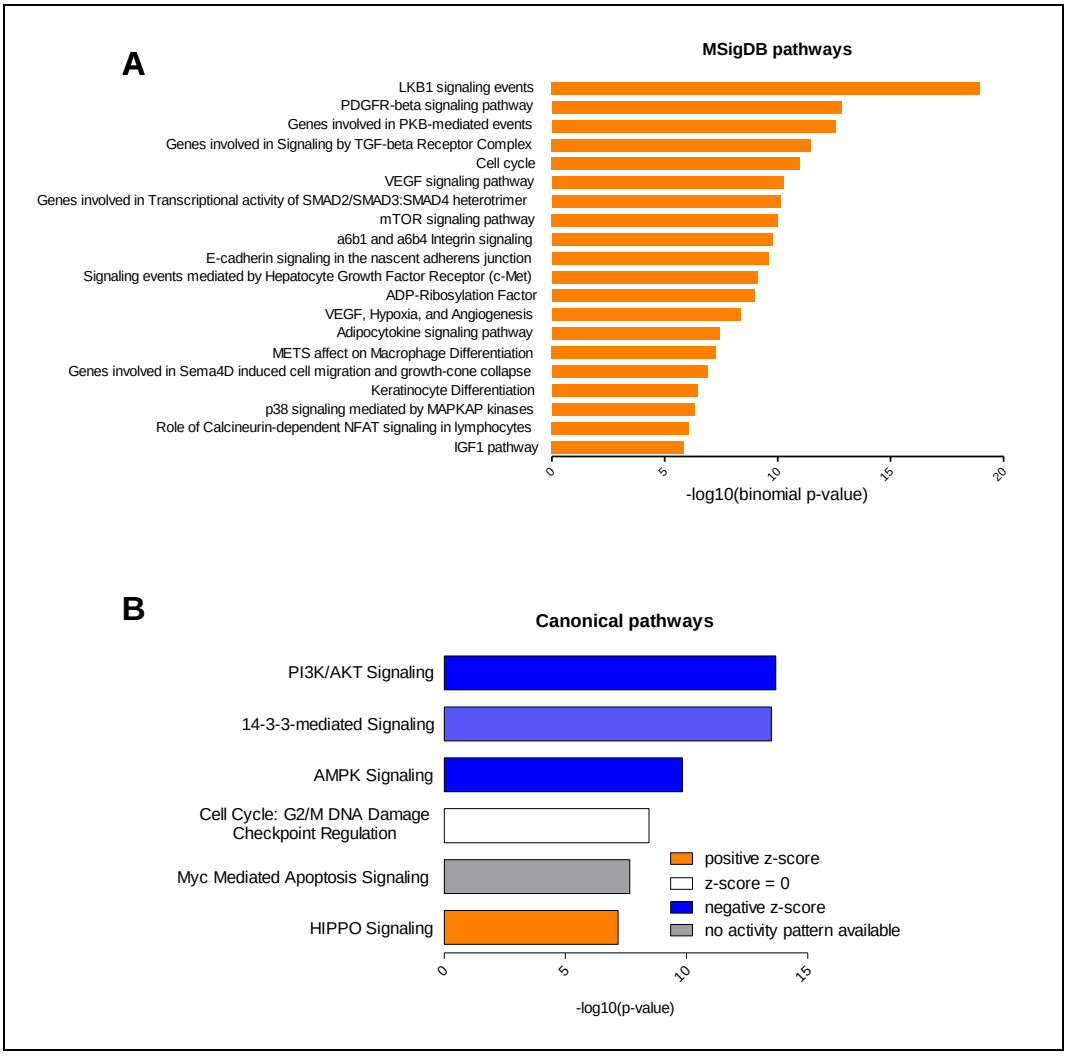


Figure 4-30. Biological pathways predicted to be affected by the promoter-associated function of *LINC00920*. (A) LKB1 signaling is the most relevant result of the Genomic Regions Enrichment of Annotations Tool (GREAT) analysis performed on cluster 1 promoter regions. (B) Ingenuity pathway analysis predicted deactivation of PI3K/AKT and 14-3-3-mediated signaling pathways from the expression profile of the LKB1 gene set in *LINC00920*-knockdown cells.

Further evidence of a putative enhancer-associated function of *LINC00920* is its enrichment in a subset of enhancer regions previously annotated in PC-3 cells [210]. Remarkably, about 29% of the 70,496 predicted enhancer regions had traces of *LINC00920* occupancy, with an almost equal distribution upstream and downstream of the enhancer center (**Figure 4-31**).

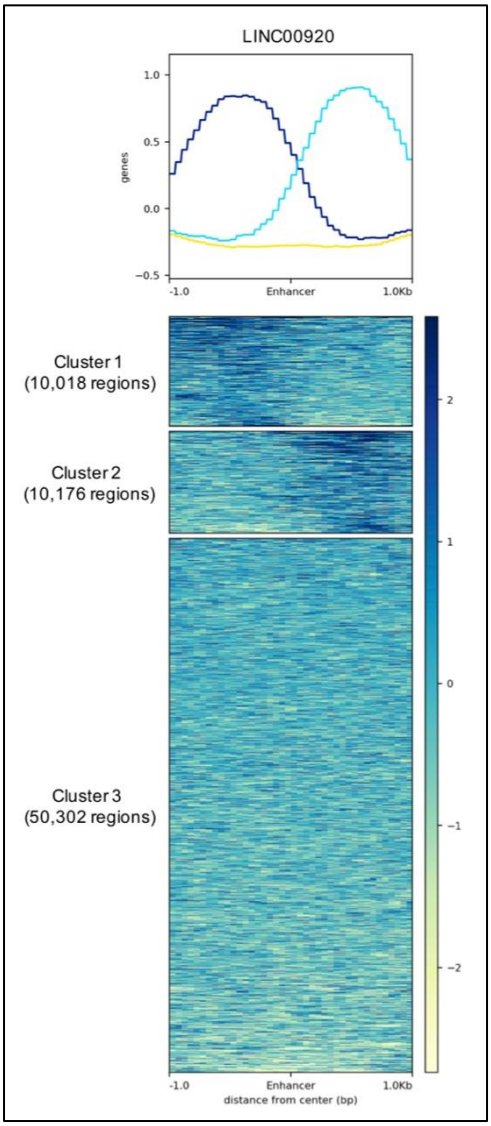


Figure 4-31. Normalized *LINC00920* read coverage across annotated enhancer regions in the PC-3 genome. Top: Average profiles of *LINC00920* occupancy on enhancers 1-kb extensions upstream and downstream the annotated center. Bottom: Segmented heatmaps showing clustered enhancer regions based on signal intensity. Cluster 1 enhancer regions show enrichment of *LINC00920* occupancy upstream the annotated center. Enhancer regions in cluster 2 show enrichment downstream of the center. The remaining cluster 3 enhancer regions do not show *LINC00920* enrichment. Shown data was derived from a representative replicate. Dark blue: high read coverage, yellow: low read coverage.

In addition to the coverage density-based analysis described above, bona fide RNA binding peaks were called using Model-based Analysis of ChIP-Seq 2 (MACS2). While this method was first developed to identify protein-bound chromatin regions [243], RNA-DNA capture by ChIRP is an analogous technology to ChIP making MACS2 an appropriate tool for RNA peak calling [193, 194].

Genome-wide, 2,985 peaks were called using a cut-off fold-change value of 10 (over the lacZ signal) and a q-value of 0.001. Genomic feature annotation of peaks using a published chromatin segmentation dataset of the PC-3 genome [210] revealed that although the majority of the peaks was identified in heterochromatic regions, *LINC00920* also associated with a variety of regulatory elements including enhancers (n=190, 6.4%) and promoters (n=64, 2.1%) (**Figure 4-32**).

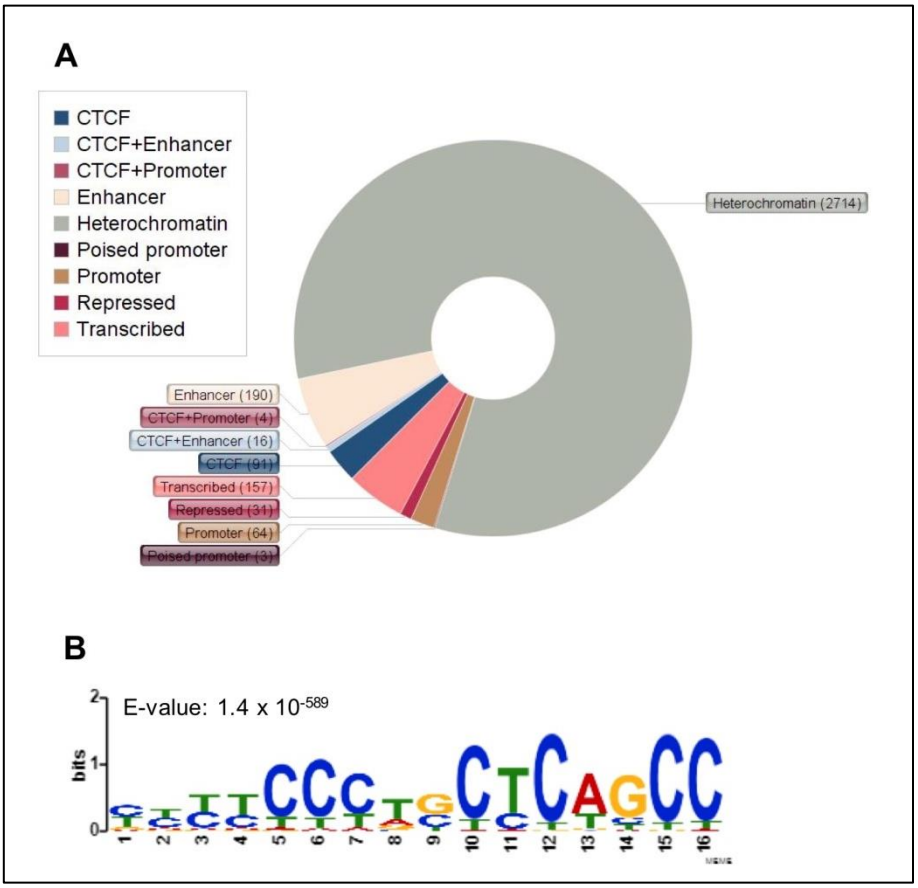


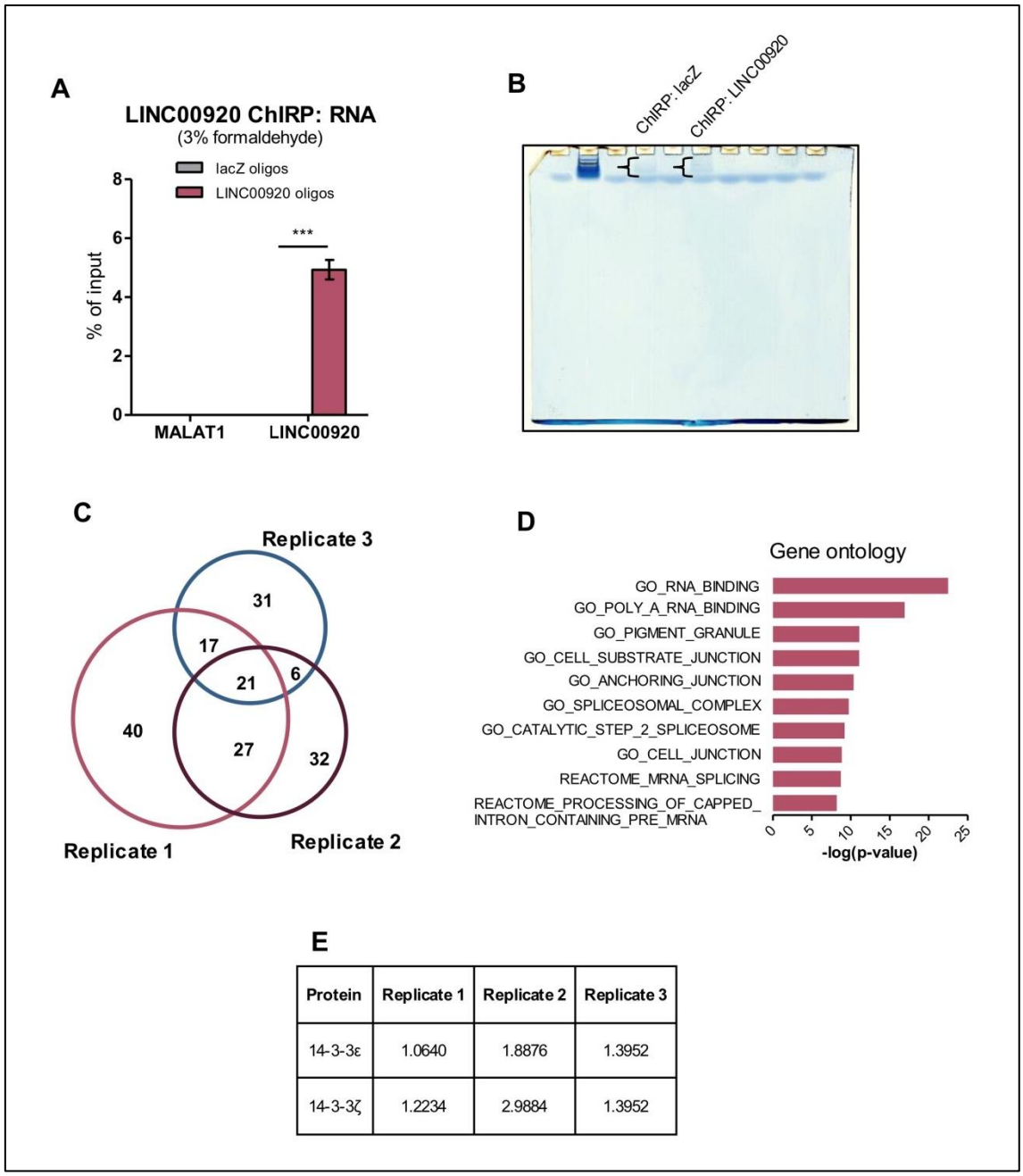
Figure 4-32. Analysis of *LINC00920* chromatin binding sites using MACS2. (A) Distribution of *LINC00920* ChIRP-seq peaks across annotated features in the PC-3 genome. (B) CT-rich homopyrimidine motif enriched in *LINC00920* binding sites.

4.12. Identification of proteins interacting with *LINC00920* via ChIRP-mass spectrometry (ChIRP-MS)

Due to the observation that *LINC00920* is not restricted within either nuclear or cytosolic compartment, *LINC00920*-ChIRP in tandem with mass spectrometry (ChIRP-MS) was performed to query the functional role of *LINC00920* beyond chromatin binding. The same set of biotinylated antisense oligos was used as in ChIRP-DNA with a few modified steps to tailor the protocol for protein retrieval. First, the number of cells was increased to 2 x 10^8 per RNA precipitation to achieve high purification yields of the lncRNA complex since proteins cannot be amplified. Second, extensive formaldehyde crosslinking was performed (3% formaldehyde crosslinking for 30 minutes) to maximize protein capture. Finally, crosslinked proteins were eluted from streptavidin beads by benzonase digestion. These modifications were not detrimental to on-bead *LINC00920* enrichment

upon pulldown (prior to elution) (**Figure 4-33A**). The eluted proteins were stained on a Coomassie gel prior to peptide digestion in preparation for mass spectrometry (**Figure 4-33B**). Three biological replicates were performed and a peptide signal intensity ratio cutoff of 1.2 between *LINC00290* and *lacZ* precipitation was implemented to qualify protein enrichment. In total, 21 identified proteins were enriched in all replicates (**Figure 4-33C, Supplementary Table 9-3**). Unsurprisingly, gene ontology analysis of these proteins revealed RNA binding and transcript splicing functions (**Figure 4-33D**).

Among the identified proteins are 14-3-3 protein isoforms: 14-3-3ε (*YWHAE*) and 14-3-3ζ (*YWHAZ*). 14-3-3 proteins are small chaperone proteins that bind to phosphorylated ligands. Such binding provides steric hindrance or elicits a conformational change that alters the biochemical properties of the 14-3-3-bound protein [244]. The presence of these adaptor proteins was notable in light of the observation that FOXO signaling is activated upon *LINC00920* knockdown. While 14-3-3ε enrichment did not reach the 1.2 cutoff in one MS replicate (**Figure 4-33E**), it was deemed worthwhile to investigate the appearance of both isoforms in the *LINC00920* pulldown due to their implicated role in FOXO signaling. One of the many functions of 14-3-3 proteins is the regulation of FOXO transcription factors by cytoplasmic sequestration [245]. AKT-mediated phosphorylation of FOXO induces 14-3-3 binding, preventing reentry of FOXO into the nucleus. FOXO transcriptionally activates genes involved in cell cycle arrest, apoptosis, and reactive oxygen species detoxification, among others. Altogether, the microarray and mass spectrometry results led to the hypothesis that *LINC00920*/14-3-3 association enhances nuclear FOXO exclusion and subsequently reduce the expression of FOXO targets.



E

Protein	Replicate 1	Replicate 2	Replicate 3
14-3-3ε	1.0640	1.8876	1.3952
14-3-3ζ	1.2234	2.9884	1.3952

Figure 4-33. Identification of the *LINC00920* protein interactome through ChIRP-MS. (A) Significant specific enrichment for *LINC00920* using pooled antisense oligos was similarly achieved in formaldehyde-fixed PC-3 cells as in glutaraldehyde-fixed cells. (B) Eluted proteins from respective *LINC00920*- and lacZ-precipitations were loaded into a polyacrylamide gel and stained with Coomassie dye. Brackets indicate stained proteins that were cut from the gel and eventually processed for mass spectrometry. (C) Proteins enriched in the *LINC00920* pulldown were identified using a signal ratio cutoff of 1.2 (*LINC00920*:lacZ). Considerable overlap of putative lncRNA binding proteins among the three biological replicates of ChIRP-MS was observed with 21 proteins common to all replicates. (D) Gene ontology analysis of the 21 proteins revealed predominantly RNA-associated processes. (E) Relative enrichment values of 14-3-3 proteins in the *LINC00920* pulldown for all replicates.

4.13. Validation of *LINC00920* transcript/14-3-3 protein interaction

RNA immunoprecipitation (RIP) using 14-3-3-specific antibodies was performed in VCaP cells to validate the association of *LINC00920* and the 14-3-3 proteins as identified from the ChIRP-MS data. Primers tiling the span of the spliced *LINC00920* transcript were designed in the attempt to map the protein-interacting portion of the lncRNA (**Figure 4-34A**). Significant *LINC00920* enrichment over the IgG control was observed for the primer pair amplifying the intronic junction of the transcript upon 14-3-3ε precipitation (**Figure 4-34B**). On the other hand, 14-3-3ζ precipitation did not enrich for any fragment of the lncRNA. These results point to the specific interaction of *LINC00920* with the 14-3-3ε protein isoform.

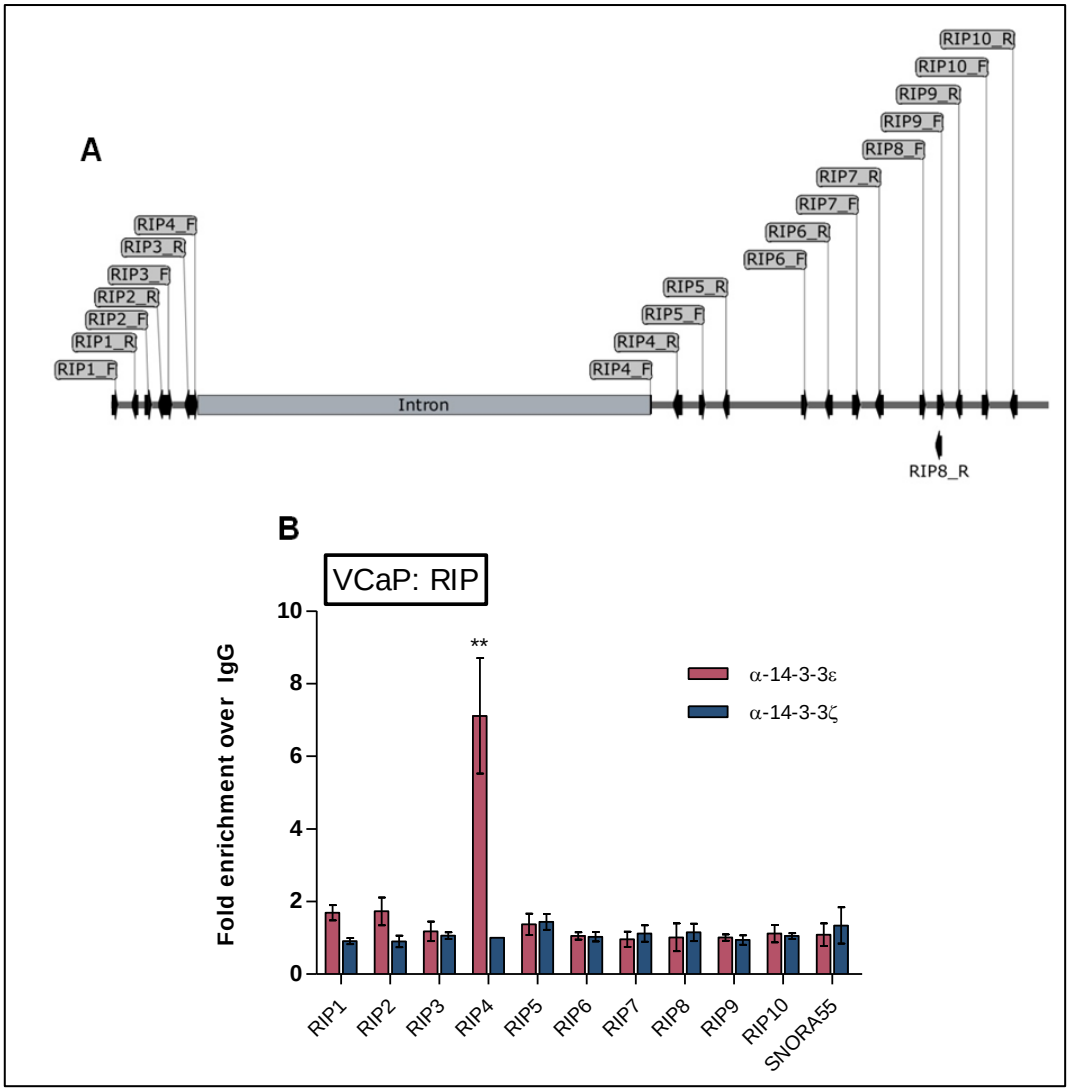


Figure 4-34. RNA immunoprecipitation (RIP) of 14-3-3 proteins. (A) RIP primers were designed to tile across the full *LINC00920* transcript. (B) RIP was performed for the 14-3-3ε and 14-3-3ζ isoforms in VCaP cells. RNA precipitated by 14-3-3ε showed significant enrichment, relative to the IgG control, for the *LINC00920* fragment amplified by the intron-spanning primer pair (RIP4). In contrast, no significant enrichment for any amplified *LINC00920* fragment was observed in the 14-3-3ζ pulldown, indicating a 14-3-3ε-specific binding of the lncRNA. The small nucleolar RNA *SNORA55* was used as negative control. (**p≤0.01)

Complementary to RNA immunoprecipitation, affinity purification on streptavidin beads using *in vitro* transcribed biotinylated *LINC00920* (bi-LINC00920) was performed to pull down recombinant 14-3-3ε from solution (**Figure 4-35A**). Because the pcDNA3.1(+) mammalian expression vector harbors the T7 promoter upstream of its multiple cloning site, the pcDNA3.1(+)-*LINC00920* construct was used to generate the DNA template to be transcribed *in vitro*. The construct was linearized at the unique XbaI restriction site downstream the cloned full-length *LINC00920* cDNA (**Figure 4-35B**). Together with a linearized plasmid harboring the firefly *luciferase* (*Fluc*) gene under the transcriptional control of the T7 promoter to serve as control template, *LINC00920* was transcribed using T7 RNA polymerase and a ribonucleotide mix with and without biotin-16-UTP. Non-biotinylated transcription reactions yielded the expected RNA lengths for *Fluc* (1800 nt) and *LINC00920* (1567 nt) (**Figure 4-35C**). Biotinylated reactions generated RNA products slightly heavier than the expected size due to the incorporation of biotinylated uracil. In some reactions, template DNA contamination was present seen as high molecular weight bands in the RNA gel. In such cases, DNase-I digestion was performed prior to use of the biotinylated products (**Figure 4-35D**).

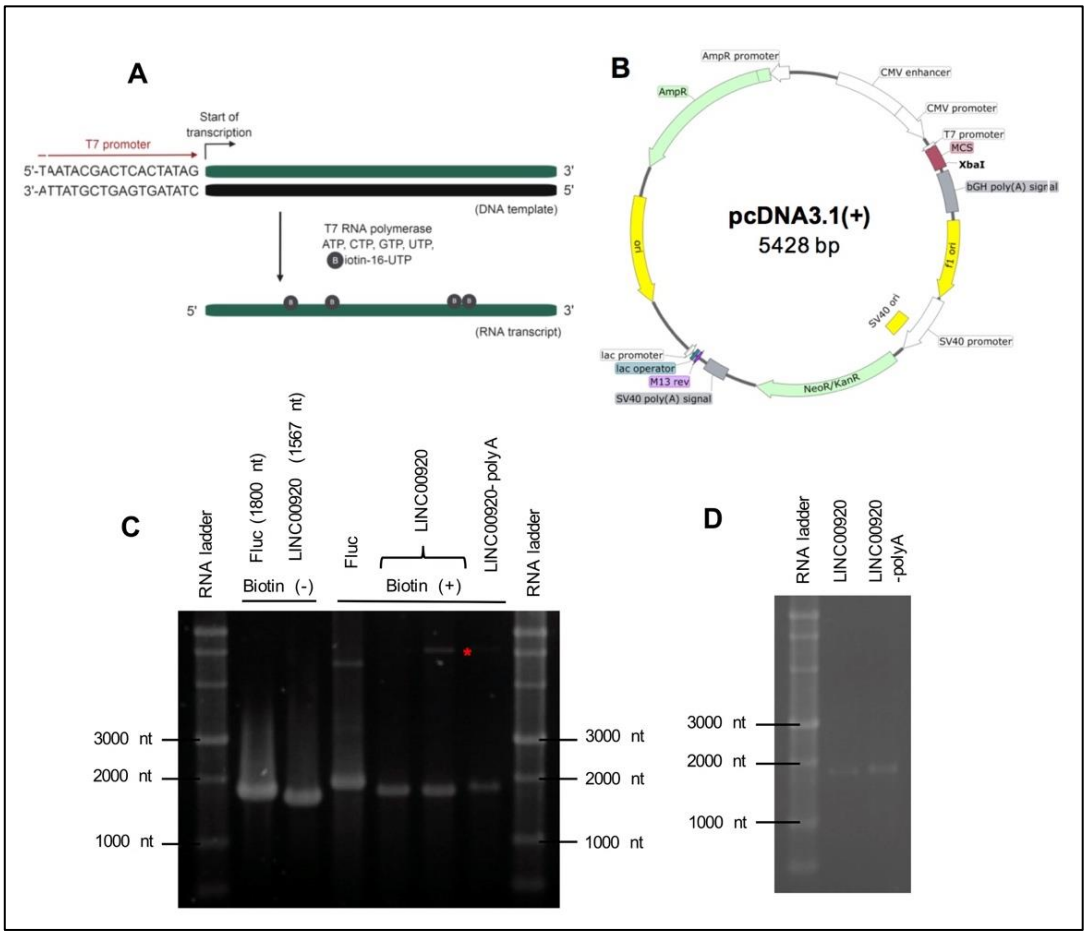


Figure 4-35. Generation of biotin-tagged *LINC00920* through *in vitro* transcription. (A) *In vitro* transcription using T7 RNA polymerase. (B) Vector map of pcDNA3.1(+) wherein the full-length *LINC00920* cDNA was cloned upstream of the XbaI restriction site. (C) *In vitro* transcription was performed for the firefly *luciferase* (*Fluc*; 1800 nt) and *LINC00920* (1567 nt) without and with biotinylation. RNA product lengths were estimated by running the samples though 1% formaldehyde/MOPS/agarose denaturing gel. The red asterisk (*) indicates a DNA template band that was removed upon additional DNase-I digestion (D).

Next, direct interaction of lncRNA and protein was investigated *in vitro* by precipitating recombinant 14-3-3ε (r14-3-3ε) with the biotinylated transcript on magnetic streptavidin beads. On a silver-stained polyacrylamide gel, r14-3-3ε had an estimated molecular weight of 35 kDa (**Figure 4-36A**). RNase A treatment was performed to verify that the precipitation of the protein is dependent on the presence of *LINC00920*. After incubation, the precipitated protein was eluted by boiling the beads in Laemmli buffer. Without RNase A treatment, r14-3-3ε could be probed via western blot while RNA digestion abrogated the band signal (**Figure 4-36B**). These observations indicate that intact *LINC00920* is required to pulldown r14-3-3ε and furthermore imply a direct interaction between the RNA and protein in solution.

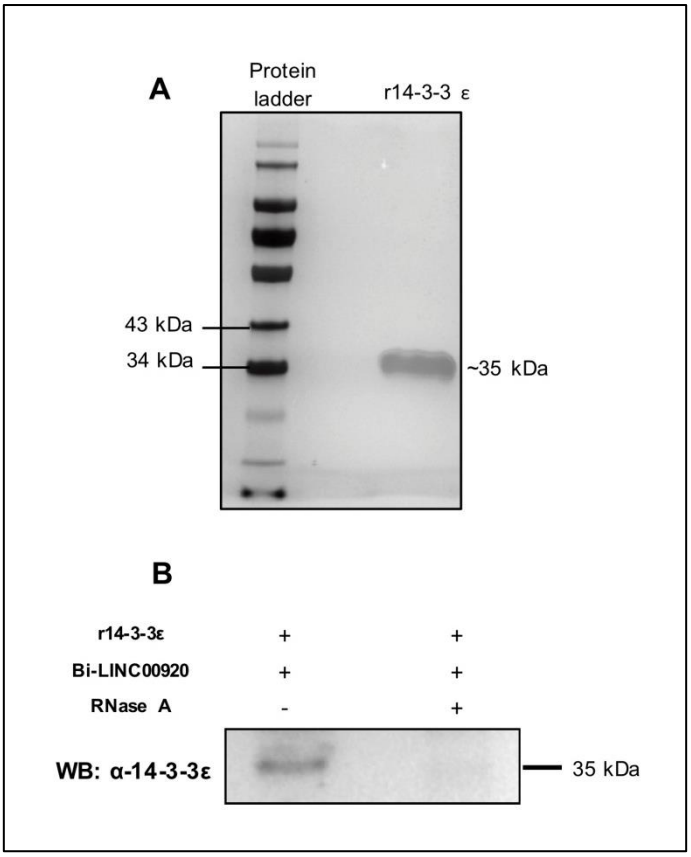


Figure 4-36. Affinity purification of recombinant 14-3-3ε (r14-3-3ε) using biotinylated *LINC00920* (bi-*LINC00920*). (A) R14-3-3ε was estimated to be approximately 35 kDa on a silver-stained 4-20% polyacrylamide gel. (B) After hybridization with r14-3-3ε, affinity purification was performed on bi-*LINC00920* using streptavidin beads without or with RNase A treatment to determine the RNA dependence of r14-3-3ε binding. Immunoblot of the eluted proteins showed a direct interaction of r14-3-3ε and bi-*LINC00920* that was abolished upon RNA digestion.

5. Discussion

5.1. Identification and selection strategy of prostate cancer-associated lncRNAs

Participation of lncRNAs in cancer hallmark processes such as cell proliferation, apoptosis, invasion, and metastasis through various molecular modalities has emerged in the past decade. Nonetheless, the numbers of differentially expressed transcripts and well-characterized lncRNAs remain severely disproportionate in all cancer entities. In the context of the low mutational burden in PCa, looking into the dysregulation of the non-coding genome can potentially yield novel insights into the molecular biology of the tumor.

Differential expression analysis performed on the ICGC-EOPC cohort revealed multiple deregulated lncRNAs in the prostate transcriptome. Transcripts selected for further investigation were limited to the multi-exonic lincRNA biotype. This criterion minimizes the possibility of selecting false-positive candidates that could have arisen from poor mapping of sequencing reads. In addition, the majority of functionally annotated lincRNAs was excluded from the short-list of candidates—despite considerable differences in expression between normal and tumor tissues—due to low read counts (<500 mean counts), which would translate to technical difficulties in cellular functional analysis. The proximity of regulatory elements around the chromosomal locus of the target was also inspected to assess whether the candidate might be actively transcribed. Lastly, the expression profile of a potential candidate was confirmed to be robust in the TCGA-PRAD [75] and TANRIC [246] datasets. Following these selection guidelines, seven PCa-associated lncRNAs were nominated for further study. *RP11-867G23.3*, *RP11-3P17.5*, and *LINC00920* were found to be significantly upregulated in prostate tumors while *LINC00844*, *LINC01082*, *RP11-395L14.4*, and *SNHG18* were significantly repressed.

The well-documented cell-type specificity of lncRNAs became apparent upon expression level validation of the seven candidates by qPCR in PCa cell lines. Although RWPE-1, LNCaP, VCaP, DU-145, and PC-3 are all prostatic-derived cell lines, each exhibits a unique genetic background with a specific transcriptional program. RWPE-1 is a model for non-tumorigenic human prostatic epithelium. LNCaP and VCaP cells are derived from metastatic prostate adenocarcinomas and are androgen-responsive. Furthermore, VCaP cells harbor an allele of the *TMPRSS2:ERG* gene fusion which results in ERG overexpression. DU-145 and PC-3 are metastatic cell lines but do not express AR nor PSA, and are consequently hormone insensitive. PC-3 cells are also *PTEN* deficient and highly aneuploidal [247]. These distinct genetic contexts would explain the inconsistent trend of lncRNA expression between the transcriptome data and normalized qPCR measurements for some cell lines. At the same time, this observation highlights the potential of lncRNAs in defining specific tumor subtypes.

RACE was performed using a standard human prostate RNA in order to characterize the predominant and most likely functional transcript isoform in prostate cells. The 5'- and 3'- ends were successfully cloned and sequenced for all lncRNA candidates except for *RP11-867G23.3* and *RP11-3P17.5*. For

LINC00920, *LINC00844*, and *SNHG18*, the sequenced 3'-ends revealed shorter transcript isoforms compared to the lengths annotated by RefSeq or GENCODE. Alternative polyadenylation (APA) is the most likely cause of this discrepancy. APA entails RNA processing yielding distinct 3' termini on RNA polymerase II transcripts, including lncRNAs [248]. APA is tissue-specific, and is recognized to be a mechanism of gene regulation widespread in eukaryotes. Endonucleolytic cleavage and subsequent polyadenylation occurs downstream of canonical hexameric polyadenylation signals [249]. Accordingly, these polyadenylation motifs were present immediately upstream the cloned 3'-ends of *LINC00920*, *LINC00844*, and *SNHG18*, implying prostate cell-specific RNA processing of these transcripts. For lncRNAs with multiple annotated isoforms (i.e., *RP11-395L14.4* and *SNHG18*), RACE results also demonstrated the utility of this technology in identifying predominant transcript isoforms.

From the remaining lncRNA candidates, only the full-length cDNAs of *LINC00920*, *LINC00844*, and *LINC01082* were successfully cloned and amplified, further narrowing down the selection list. Due to its overexpressed nature in tumors—suggestive of an oncogenic function—and its robust expression in PCa cell lines—in contrast to the limited detectability of *LINC00844* and *LINC01082—LINC00920* was ultimately selected for further investigation.

5.2. The non-coding potential of *LINC00920* RNA

Although lncRNAs and mRNAs share post-transcriptional features such as 5'- m^7Gpppn capping and polyadenylation, and the capacity to undergo splicing, non-coding transcripts can be distinguished from their coding counterparts upon evaluation of ORF size and coverage, nucleotide or codon frequencies and composition, evolutionary substitution patterns, similarity to known protein-coding transcripts, and presence of known functional domains [250]. Coding regions have the tendency to harbor ORF lengths longer than expected by chance [251]. The ORF coverage, which is the length of the longest ORF normalized to the transcript size, may also be considered as the probability of determining a long ORF increases with transcript length [200]. Moreover, nucleotide frequencies within protein coding ORFs are defined by non-random codon usage and thus are partly indicative of coding potential [250]. Protein coding-, in contrast to non-coding genes, evolve under selective pressure to maintain intact ORFs and to preserve specific amino acid residues or amino acid types at defined positions [252]. This selective pressure can be evaluated by performing multiple sequence alignments, comparing nucleotide substitution frequencies, and assessing the integrity of the ORF upon introduction of insertions and deletions (indels). Sequence similarities to known mRNAs may also be inspected to evaluate the coding potential of a transcript. Finally, encoded protein domains are commonly present in protein-coding sequences but absent in non-coding transcripts [250].

In this study, three computational tools were applied to confirm the non-coding potential of *LINC00920*. This approach enabled an evaluation of the (non-)coding potential using multiple criteria that would not have been achieved by a single tool. The CPAT coding probability score is based on a

logistic regression model built on open reading frame (ORF) size, ORF coverage, combinatorial effect of nucleotide composition and codon usage bias (Fickett score) and hexamer usage score [200], which is related to the differential usage of nucleotide hexamers observed in exons and introns [253]. The CPC relies on the ORF length and quality of the transcript. In addition, this tool considers the BLASTX output quality of the RNA query and compares the resulting information with bona fide protein-coding transcripts [201]. Finally, PhyloCSF utilizes evolutionary signatures derived from alignments of conserved coding regions to score the likelihood of a transcript to be protein-coding [229]. All three tools corroborated the non-coding capacity of *LINC00920*, yielding non-coding potential scores comparable to well-established lncRNAs *NEAT1* and *MALAT1*, and in contrast to mRNAs such as *GAPDH* and *ACTB1*.

5.3. *In vitro* functional characterization of *LINC00920*

The functional role of *LINC00920* was interrogated in PC-3 cells, wherein the transcript was found to be upregulated compared to the immortalized prostate epithelial cell line RWPE-1. Cells depleted of *LINC00920* showed decreased proliferative, migratory, and colony forming capacities. Together with the high expression observed in tumors, these results assert the oncogenic properties of *LINC00920*. Furthermore, microarray profiling of cells depleted of *LINC00920* revealed perturbed cellular pathways highly relevant to the observed cellular phenotypes. Interestingly, focused analysis of the top *LINC00920*-deregulated genes predicted enhanced FOXO signaling activity upon lncRNA knockdown.

In humans, the FOXO family of transcription factors is comprised of four members: FOXO1, FOXO3, FOXO4, and FOXO6 [254]. FOXO proteins share redundant functions as they bind the cognate (G/C)(T/A)AA(C/T)AA Forkhead response element (FRE) [255]. FOXO protein specific functions can be mediated by interaction with various coregulators [254]. Depending on the cellular context, FOXO transcriptional programs affect a variety of processes by regulating genes involved in cell cycle arrest (e.g., *GADD45A*, *CDKN1B*), apoptosis (e.g., *BCL2L11*, *PMAIP1*), differentiation (e.g., *PDGFRA*, *PRDM1*) and metabolic response (e.g., *CAT*, *SOD2*) [230-233, 256, 257]. In cancer entities such as leukemia, breast cancer, and prostate cancer [258, 259], FOXO transcription factors are considered tumor suppressors as they act as downstream effectors of *PTEN* [260]. Upon *PTEN* loss or somatic mutations in pathway-involved genes, PI3K signaling activation leads to elevated AKT survival pathway activity. Among the direct AKT substrates in the nucleus are FOXO proteins whose subsequent phosphorylation leads to deactivation and eventual nuclear exclusion (**Figure 5-1**) [261-263].

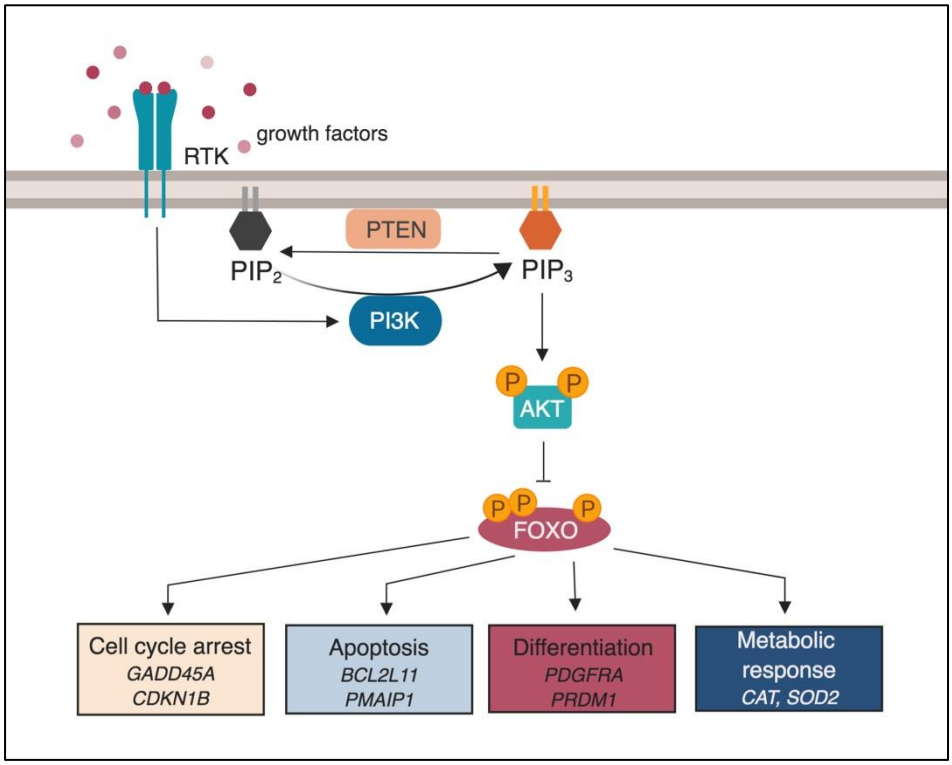


Figure 5-1. Simplified illustration of FOXO signaling regulation by the PTEN/PI3K pathway. Receptor tyrosine kinase (RTK) activates phosphoinositide 3-kinase (PI3K) which in turn phosphorylates phosphatidylinositol 4,5-bisphosphate (PIP_2) [264]. If active, PTEN antagonizes PI3K function by dephosphorylating phosphatidylinositol (3,4,5)-trisphosphate (PIP_3) [265]. PIP_3 triggers the phosphorylation and concomitant activation of AKT. Active AKT phosphorylates FOXO proteins at specific sites resulting in FOXO deactivation and subsequent downregulation of target genes related to the indicated physiological processes.

The antagonistic effect of *LINC00920* expression on FOXO signaling was further confirmed by the increased transcription of canonical FOXO target genes upon *LINC00920* knockdown in ipatasertib-treated PC-3 cells. For experimental rigor, five independent siRNAs were used to validate the effect of *LINC00920* on the canonical FOXO target genes *BCL2L11*, *GADD45A*, and *PMAIP1*. While silencing *LINC00920* alone led to minor changes in the expression of FOXO targets, simultaneous *LINC00920* knockdown and AKT inhibition through ipatasertib treatment significantly upregulated all genes in comparison with the scrambled siRNA control. These observations can be rationalized by a hyperactive AKT brought about by *PTEN* deletion inherent in PC-3 cells. Without ipatasertib, FOXO proteins are consistently phosphorylated and inactive, presumably masking the effect of *LINC00920* in FOXO target activation. On the other hand, limiting AKT activity uncoupled the influence of the hyperactive PI3K pathway on FOXO signaling. In this context, the isolated activating effect of *LINC00920* knockdown on FOXO signaling was evident. Complementing these results, *LINC00920* overexpression led to significant downregulation of *GADD45A* and *PMAIP1*. *BCL2L11* expression was similarly affected, albeit not statistically significant. The opposing effect of *LINC00920* expression to FOXO activity was also demonstrated in VCaP, a *PTEN*-intact cell line with high endogenous *LINC00920* expression.

Growth arrest and DNA damage inducible 45 alpha (*GADD45A*) encodes a tumor suppressive protein implicated in DNA repair, maintenance of genomic stability, cell cycle control, and apoptosis [266]. GADD45A facilitates cell cycle arrest in response to genotoxic stress by inhibiting CDK1/CYCLIN B1 complex formation required for G_2-M transition during cell cycle progression [267]. *GADD45A* has been shown to be downregulated in primary prostate tumors compared to nonmalignant tissue [268]. The gene products of BCL2-Like 11 (*BCL2L11* or *BIM*) and Phorbol-12-Myristate-13-Acetate-Induced Protein 1 (*PMAIP1* or *NOXA*) are essential pro-apoptotic proteins belonging to the BH3-only protein family. BH3-only proteins initiate the mitochondrial apoptotic pathway by activating Bax-like proteins or by binding and sequestering anti-apoptotic Bcl-2 proteins [269]. It is also important to note that other FOXO established targets such as *TNFSF10*, *CCNG2*, and *CDKN1B* were similarly evaluated for expression activation upon *LINC00920* perturbation. However, no remarkable alteration in expression was observed for these genes upon lncRNA knockdown in the absence or presence of ipatasertib (**Supplementary Figure 9-2**). This suggests that inactivation of FOXO signaling in this context is limited to the downregulation of select genes. Interestingly, a previous report has asserted that FOXO target genes can be classified as responsive to FOXO levels alone (class I genes) or in combination with other transcriptional activators (class II genes) depending on the multiplicity of FREs in the promoter region [270]. This classification might explain the non-responsiveness of those FOXO target genes to *LINC00920* knockdown. Taken together, *LINC00920* plays a role in modulating a subset of FOXO targets associated with cell cycle control and apoptosis in prostate cancer cells.

5.4. *LINC00920* transcription is regulated by ERG

The hypothesis that ERG drives *LINC00920* transcription stemmed from the initial observation that the lncRNA is upregulated in the T2E-positive VCaP cells. In support of this, correlations between *LINC00920* and ERG were determined in the TCGA-PRAD and ICGC-EOPC datasets. Lentiviral transduction of ERG in LNCaP cells also recently showed concomitant upregulation of LINC00920 [95]. Indeed, ERG perturbation in VCaP and tet-inducible ERG-overexpressing LNCaP cells demonstrated the dependence of *LINC00920* expression on the transcription factor. Furthermore, direct interaction of ERG with identified ETS domains within the *LINC00920* promoter was established through promoter luciferase and ChIP assays. These experiments further suggest a hierarchical multi-site binding of ERG within a single regulatory region, which has previously also been observed in the *YAP1* promoter [271].

In androgen independent and T2E-negative PC-3 cells, *LINC00920* transcription can be attributed to the overexpression of another ETS family member ETV4, which binds to the same ETS domain as ERG. *ETV4* gene fusions and the consequent overexpression of the ETV4 protein are detected in 4% of primary prostate tumors [75]. The metastasis promoting effect of ETV4, in collaboration with activated PI3K and RAS signaling pathways, has been reported in an advanced PCa mouse model [272]. Moreover, ETV4 has been established to be required for the anchorage-independent growth of

PC-3 cells [273]. Since *LINC00920* acts downstream of ERG and ETV4, it is tempting to speculate that the tumorigenic effects of the two ETS transcription factors could partly be mediated by *LINC00920* through its crosstalk with FOXO signaling. To date, *LINC00920* is only the second lncRNA reported to be regulated by ERG [82, 274], and the first lncRNA gene to be described as a direct ERG target in prostate cancer cells.

5.5. The *LINC00920* interactome

The presence of mature *LINC00920* transcripts in the chromatin, nucleoplasmic, and cytoplasmic fractions implies a number of possibilities as to how the lncRNA can negatively affect FOXO signaling/elicit the observed cellular phenotype in prostate cancer cells. Since lncRNAs do not function in isolation and instead work in complement with proteins or other nucleic acids, lncRNA interactome identification is central to understanding its modality. In this study, ChIRP, in tandem with high throughput sequencing (ChIRP-seq) and mass spectrometry (ChIRP-MS), was applied to interrogate the role of *LINC00920* in the context of both the chromatin and RNA binding proteins (RBPs). ChIRP experiments were performed with endogenous levels of *LINC00920* in PC-3 cells, minimizing potential interacting artifacts that would most likely be detected if the transcript was otherwise overexpressed.

5.5.1. The chromatin binding map of LINC00920

Due to the notable presence of *LINC00920* in the chromatin fraction, ChIRP-seq was performed to map its putative chromatin binding sites. Previously, ChIRP-seq experiments targeting *Drosophila roX2*, human *TERC* and *HOTAIR* lncRNAs have revealed important insights regarding interactions between non-coding transcripts and the chromatin [194]. ChIRP can be performed using either formaldehyde or glutaraldehyde as chemical crosslinkers as the chemistry of both reagents in theory enables sufficient preservation of RNA:DNA contacts [192, 239, 275]. However, it has been empirically determined that formaldehyde is not effective in crosslinking some lncRNA:DNA interactions, such as those mediated by *TERC* [194, 276]. Ultimately, *LINC00920* ChIRP-seq experiments were optimized using glutaraldehyde instead of formaldehyde as the fixing agent due to its superior DNA capture efficiency as revealed by *MALAT1* validation pulldowns.

ChIRP-seq displayed a low genome-wide mean coverage signal of *LINC00920* which likely reflects the limited number of transcript copies available for chromatin interaction within the cell. Nonetheless, enrichment of the lncRNA across subsets of regulatory regions, particularly of promoters and enhancers, suggests potential functionalities at the level of gene regulation. Interestingly, *LINC00920*-occupied promoters were found to be implicated in biological functions related to LKB1 signaling, and further converging on the PI3K/AKT and 14-3-3 signaling pathways. Intriguingly, perturbation of these pathways has been redundant in the context of *LINC00920* function in prostate cancer cells. Nonetheless, additional functional interrogation of these chromatin regions is required to complement and validate such predictions.

Another notable observation from the genomic mapping of *LINC00920* occupancy is its enrichment in a subset of enhancer regions. Adding to this is the apparent colocalization of *LINC00920* and the histone mark H3K4me1 signals in the same cluster of promoter regions. H3K4me1 is typically found at enhancers and large 5' segments of actively transcribed genes [277]. Enhancers are *cis*-regulatory elements defined as 100-1000 bp non-coding DNA regions that activate gene transcription regardless of their distance, location, or orientation relative to specific cognate promoters [278]. Specifically, primed enhancers are marked with H3K4me1 with simultaneous depletion of H3K4me3. On the other hand, active enhancers are enriched for H3K4me1, H3K27ac, and other histone modifications such as H4K16ac and H3K122ac [240, 241, 279]. It is currently unclear whether this histone deposition overlap has biological relevance, but the complete reversal from H3K4me1 depletion in the genome-wide context to enrichment at *LINC00920*-occupied regions is striking.

Although *LINC00920* ChIRP-MS failed to identify H3K4me1 associated proteins (such as chromatin remodelers BAZ1A/B and chromatin associated factors belonging to the SWI/SNF complex [280]), this does not rule out the colocalization of H3K4me1 and *LINC00920*. This only demonstrates the non-association of the lncRNA with the chromatin regulators. Considering that only around 15% of annotated promoters were co-occupied, it is also possible that the amount of H3K4me1-associated proteins precipitated by the *LINC00920*-antisense oligos did not reach detection sensitivity. It has been demonstrated that H3K4me1 has a role in recruiting the SWI/SNF complex to enhancers, but how enhancers are pre-marked by H3K4me1 remains an open question [280]. Histone methyltransferases such as KMT2C and KMT2D (MLL3/4) have been reported to interact with cell-type specific and signaling-dependent factors that possibly earmark regions for histone methylation [281-283]. Interestingly, the lncRNA *HOTTIP* has been reported to recruit WDR5-MLL5 complexes to the 5' *HOXA* locus, which results in the deposition of H3K4me3 mark along a broad chromatin domain, triggering gene activation [284]. It should be noted however that recruitment of WDR5-MLL5 is brought about through the nascent transcription of *HOTTIP*, which acts to tether the protein complex to the target locus in *cis*. On the other hand, the mouse lncRNA *Fendrr* has been demonstrated to recruit the chromatin regulatory Polycomb complex PRC2 both in *cis* and in *trans*, mediating trimethylation at H3K27 at target gene promoters [112]. These examples could be taken into account upon generating a hypothesis for *LINC00920* function at enhancer regions.

In parallel, corroborating evidence implicating *LINC00920* with enhancer-related function was revealed by MACS2-based peak calling and subsequent genomic feature annotation. Second only to heterochromatic regions, *LINC00920* was found most enriched at annotated enhancers in the PC-3 genome. Remarkably, DNA motif discovery revealed lncRNA binding tendency to stretches of CT-rich homopyrimidines, which are known to form triple-helical nucleic acid interactions through Hoogsteen hydrogen bonding [285, 286]. This suggests the possibility of a *LINC00920*:DNA:DNA

triplex formation at the identified chromatin sites. However, validation of this hypothesis requires further investigation.

5.5.2. The protein interaction partners of LINC00920

The majority of the captured proteins identified by MS upon *LINC00920* ChIRP are well known RNA binding proteins involved in RNA splicing and maturation (i.e., heterogeneous ribonucleoprotein particles, hnRNPs). This result does not come as a surprise since *lacZ*-targeting oligos were used for the negative control pulldown. Hence, proteins related to RNA biogenesis are likely to be enriched in the experimental condition as the *lacZ* mRNA is not endogenously transcribed in human cells. Lest a false-positive target is identified, these RNA processing proteins were excluded from candidate selection. Nonetheless, the possibility that among these is a bona fide *LINC00920*-interacting protein with functional implication beyond RNA biogenesis and maturation cannot be fully discounted and remains a limitation of this study.

Given the intersection of *LINC00920* with FOXO signaling in the microarray data, the consistent enrichment of 14-3-3 proteins in *LINC00920* ChIRP-MS experiments was a relevant observation. While little is known about the RNA-binding capacity of 14-3-3 proteins, RNA-binding activities within the 14-3-3 domain have recently been reported through a global RNA proteomics approach [287].

Canonically, 14-3-3s are chaperone proteins that bind to phosphorylated ligands, among them FOXO proteins [244]. In total, there are seven 14-3-3 isoforms expressed in mammals: β (identical to α upon phosphorylation), γ, ε, η, ζ (identical to δ upon phosphorylation), θ, and σ. Although encoded by different genes, all isoforms exhibit highly similar primary sequences [288]. Each protein is approximately 30 kDa in size. Typically functioning as homo- or heterodimers partnered with other family members, a 14-3-3 monomer consists of nine α-helices that form a conserved amphipathic region that acts as the phosphorylation-binding pocket [289]. Binding of the 14-3-3 dimer provides steric hindrance or elicits a conformational change that alters the biochemical properties of host proteins [244]. For FOXO proteins in particular, 14-3-3 binding licenses FOXO for nuclear exclusion.

FOXO proteins harbor both a nuclear export signal (NES) at the C-terminus and a nuclear localization signal (NLS) proximal to the FH domain, which permit nucleocytoplasmic shuttling [262, 290]. This shuttling mechanism is directly influenced by FOXO phosphorylation—primarily by AKT—resulting in a regulatory mechanism involving subcellular FOXO sequestration. All mammalian FOXO proteins harbor three highly conserved putative AKT recognition motifs with the consensus sequence RXRXXS/T [291]. The recognition motifs are found, one each, at the N-terminal, C-terminal, and FH domains. Phosphorylation at the N-terminal and FH domains are required for protein translocation. The phosphorylated residues act as docking points for 14-3-3 proteins whose binding initiates the formation of the nuclear export complex [261]. Upon PI3K pathway activation, dual phosphorylation

of FOXO by AKT at the C-terminal and FH domains triggers 14-3-3 binding (**Figure 5-2**) [292]. This results in DNA displacement from the FOXO FH domain. Dimer binding at the same time masks the FOXO NLS. Other kinases (i.e., CK1, DYRK1A) phosphorylate multiple residues at the C-terminus, priming the complex for association with export factors [263]. The FOXO NES is then recognized and bound by the exportin protein CRM1, followed by RAN-GTP attachment. The assembled complex is then shuttled from the nucleus through the nuclear pore complex [293]. Once in the cytoplasm, phosphorylated FOXO proteins are ubiquitinated, leading to proteasomal degradation which provides another layer of FOXO regulation [294, 295].

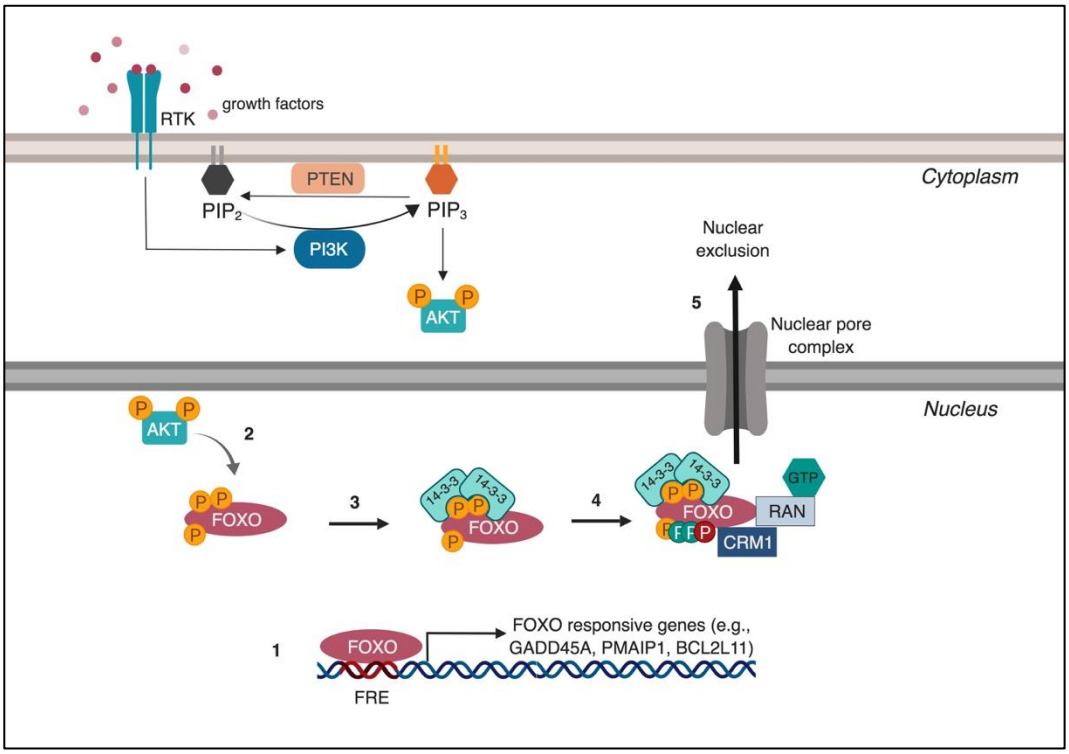


Figure 5-2. Subcellular shuttling of FOXO proteins. 1, Active nuclear FOXO proteins bind to Forkhead response elements (FREs) at gene regulatory regions, driving the transcription of FOXO responsive genes. **2**, In the presence of growth factors, the PI3K pathway activates AKT, which translocates into the nucleus and phosphorylates FOXO. **3**, 14-3-3 proteins recognize the phosphorylated residues, and as a dimer, bind at the N-terminal and Forkhead AKT sites of FOXO. **4**, Multiple residues are subsequently phosphorylated by other kinases and nuclear export proteins CRM1 and RAN-GTP interact with the nuclear export signal of FOXO. **5**, Finally, the assembled complex is transported to the cytoplasm through the nuclear pore complex.

Although both 14-3-3ε and 14-3-3ζ isoforms were identified by ChIRP-MS, only the direct association between *LINC00920* and 14-3-3ε was validated by RIP, which was further corroborated by subsequent *in vitro* affinity purification experiments. Due to the high sequence and structural similarities among 14-3-3 isoforms, it is possible that 14-3-3ε peptides were misidentified as 14-3-3ζ. This would account for the apparent specificity of *LINC00920* interaction with 14-3-3ε. Indeed, in one experimental replicate, peptide signals attributed to 14-3-3 proteins were not isoform-specific. On the other hand, it

is also likely that both isoforms were accurately identified by MS but only 14-3-3ε is capable of directly associating with the lncRNA, and 14-3-3ζ was simultaneously captured as the dimeric partner of 14-3-3ε. In support of this, 14-3-3ε/14-3-3ζ heterodimers have been reported to form in mammalian cells [296].

LINC00920 knockdown increases FOXO function while maintaining the expression of the most abundant FOXO isoform in PC-3 cells (**Supplementary Figure 9-1**). Consequently, the most rational implication of *LINC00920* binding to 14-3-3ε appears to be increasing the stability of the 14-3-3/FOXO complex, triggering the nuclear export of FOXO. With respect to FOXO signaling, *LINC00920* upregulation mimics a cellular context with an activated PI3K pathway, resulting in increased nuclear exclusion of FOXO and subsequent repression of its gene targets (**Figure 5-3**). The observed activation of FOXO target genes upon *LINC00920* silencing can then be rationalized within this molecular framework. In addition, the non-exclusive localization of *LINC00920* transcripts in the nuclear and cytoplasmic compartments is also in line with the proposed role of *LINC00920* within the nucleocytoplasmic shuttling model of FOXO.

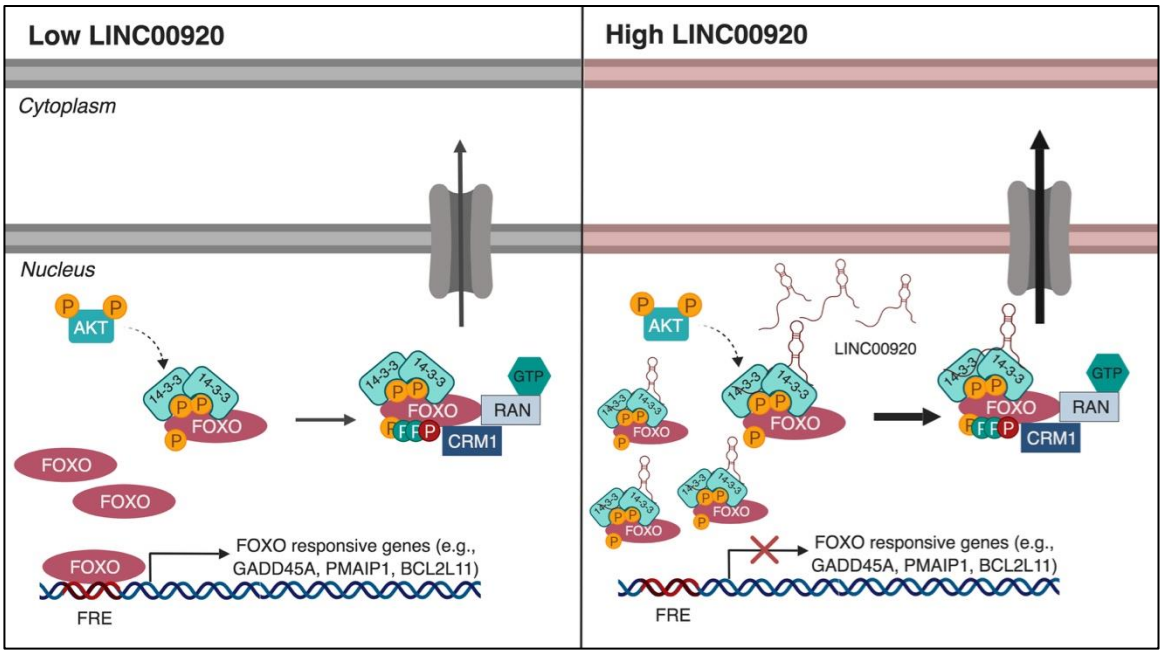


Figure 5-3. The proposed role of *LINC00920* in FOXO/14-3-3 complex assembly in PCa cells. In cells depleted of *LINC00920*, FOXO signaling is basally active. In a cellular context of *LINC00920* enrichment and AKT activation, 14-3-3 dimerization and binding to FOXO is enhanced, resulting in promoted shuttling of FOXO to the cytoplasm. This depletes nuclear FOXO levels, leading to diminished transcription of a subset of FOXO gene targets.

5.6. Bridging ERG, PTEN, and FOXO signaling through *LINC00920*

Despite the molecular functions attributed to ERG, gene fusion occurrence is considered to be an early event in carcinogenesis, and ERG overexpression alone is not sufficient to trigger cellular

transformation [297, 298]. Concomitant tumor suppressor gene (i.e., *PTEN*, *TP53*) inactivation or additional oncogene (i.e., *PI3K*/*AKT*) activation must take place to initiate tumorigenesis [299].

Hyperactivation of PI3K signaling is a common event in PCa, typically facilitated by *PTEN* loss or somatic mutations in pathway-involved genes. Consequently, diminished FOXO signaling is frequently observed in the clinical setting and PCa models. In 640 radical prostatectomy samples, nuclear FOXO1 expression was higher in normal prostate than in benign prostatic hyperplasia and prostate cancer [300], suggesting decreased FOXO1 activity in the diseased state. In a transgenic adenocarcinoma of the mouse prostate (TRAMP) mouse model, suppression of FOXO3 activity led to increased PCa progression [301]. FOXO3 function was similarly attenuated, as measured by *CDKN1B* promoter activity, in an androgen-independent PCa model derived from *in viv*o selection of LNCaP tumors in castrated mice [302]. In *PTEN*-null PC-3 cells, overexpression of a phosphorylation-resistant mutant form of FOXO1 negatively regulates the oncogenic *RUNX2* transcription factor, leading to reduced migration and invasion [303]. Adenoviral overexpression of FOXO1 and FOXO3 in LAPC4 prostate carcinoma cells induced the expresssion of pro-apoptotic and tumor suppressive genes such as *TNFSF10*, *BNIP3L*, *DAPK1*, and *SMAD4* [304]. More recently, it has been reported that loss of FOXO1 cooperates with *TMPRSS2:ERG* overexpression to drive tumor formation and cell invasion in prostate cancer [305]. Taken together, these reports illustrate the apparent red line interweaving between processes involving ERG, PTEN/PI3K/AKT, and FOXO signaling in the context of prostate cancer progression. However, a complete understanding of the diverse molecular mechanisms underpinning these complex associations remains elusive.

The functional model of *LINC00920* proposed in this study provides a novel insight on how ERG mediates its downstream effects through a lncRNA-mediated attenuation of FOXO signaling. ERG drives the transcription of *LINC00920* which, upon binding to 14-3-3ε, promotes FOXO sequestration and nuclear export of the complex. This mechanism partly rationalizes the decline of FOXO signaling through the clinical course of PCa, particularly in ERG-overexpressing cancer cells [305]. Expanding this model to an early time point of the disease, during which ETS gene fusions have just been established, *LINC00920* expression could be a way for pre-cancerous cells to circumvent the tumor suppressive influence of *PTEN*. In this context, ERG- or ETV4-overexpressing cells most likely obtain survival advantage by downregulating a subset of tumor suppressive FOXO targets. Subsequently, upon *PTEN* deactivation or PI3K hyperactivation in an ERG overexpressing background, the cell experiences concerted oncogenic pressure which initiates downstream transcriptional programs that trigger cellular transformation, and eventually maintenance of the neoplastic phenotype.

5.7. **Outlook**

5.7.1. Further exploration of LINC00920 function in the chromatin

As *LINC00920* coverage was found to colocalize with the enhancer-associated histone mark H3K4me1 in a subset of promoter regions in the PC-3 genome, it would be worthwhile to investigate the causality between *LINC00920* binding and H3K4me1 deposition at these sites. If *LINC00920* is found to be required for the establishment of H3K4me1 marks, an interesting query would be whether *LINC00920* licenses these regions as substrates for histone methyltransferases. This can be evaluated through H3K4me1 ChIP-seq upon *LINC00920* expression perturbation. If corroborating evidence is found, it can be inferred that the lncRNA impacts the activity of a number enhancer elements. However, given that enhancers can function at great distances, a perceivable challenge that must eventually be confronted is the identification of cognate target genes of these putative enhancer regions. Ultimately, querying the role of *LINC00920* would entail measurements of transcription activation of target genes regulated by the enhancer under consideration.

5.7.2. Assessment of FOXO/14-3-3ε binding affinity

In support of the proposed role of *LINC00920* in FOXO/14-3-3 complex assembly in PCa cells, further investigations should include evaluations of the binding affinity between FOXO and 14-3-3ε, in the absence or supplementation of *LINC00920* RNA. Although the FOXO1 isoform has been observed to be the most abundant FOXO protein in PC-3 cells (**Supplementary Figure 9-1A**) and presumably the most important effector of FOXO signaling in this cell line, it would of interest to query whether 14-3-3ε exhibits degenerate or preferential binding to a particular FOXO isoform. To date, different 14-3-3 isoforms have been shown to bind and regulate FOXO proteins, including 14-3-3σ [306], 14-3-3ε [307], 14-3-3θ [308] and 14-3-3ζ [261]. However, nuanced combinations and binding preferences among family members remain unclear [309] but are likely to be context- and cell type-dependent. Technologies that can be applied to assess and compare 14-3-3ε/FOXO interaction include reciprocal co-immunoprecipitation, proximity ligation assay, colocalization assay through fluorescence imaging, and immunoblotting of nuclear and cytoplasmic FOXO.

5.7.3. Expanding the in vitro-generated model to ex- and in vivo systems

Future efforts to dissect the function of *LINC00920* using preclinical models would extend the relevance of the results presented in this thesis. Genetically engineered mouse models [310, 311], xenografted human cell lines [312, 313], patient-derived xenograft models [314, 315], and patient-derived tumor organoids [316, 317] have been used to characterize cancer-associated lncRNAs in cells within intact tumor microenvironments. While each approach has inherent strengths and weaknesses, these models in general recapitulate the dynamics of multiple tumor components—such as tumor vasculature, immune cells, stromal cells, signaling molecules, and the extracellular matrix—that contribute to tumor development and evolution [318].

5.7.4. Considerations for clinical translation

LncRNAs in general could offer multiple opportunities for clinical translation. Many lncRNAs demonstrate tissue-restricted and cancer-specific expression signatures [81]. Indeed, based on tissue-wide GTEx RNA-seq analysis, the prostate and the testis are among the tissue types with the highest *LINC00920* expression (**Supplementary Figure 9-3**) [319]. Moreover, its upregulation in ERG-positive PCa tumors potentially makes *LINC00920* a suitable biomarker for both tissue-of-origin assays and tumor molecular subtyping. However, it remains to be explored whether the overall expression level of *LINC00920* would be sufficient, in the practical sense, for diagnostic applications.

In recent years, nucleic acid-based therapies have emerged to post-transcriptionally target lncRNAs. These include the application of RNA-mediated interference (RNAi), single-stranded antisense oligonucleotides (ASOs), and morpholino oligonucleotides [320]. On the basis of its oncogenicity and expression pattern, *LINC00920* could be a potential therapeutic target in ERG-positive PCa. However, despite clinical approvals for ASO drugs as treatment for spinal muscular atrophy [321] and familial hypercholesterolemia [322], and a morpholino-based splicing modulator for patients with Duchenne muscular dystrophy [323], most lncRNA-targeting therapeutics are in the very early stages of development and are still far from clinical use [320].

6. Attributions

LINC00920 promoter luciferase assays, including promoter construct generation, were performed with supervision by Niclas Flosdorf as part of a bachelor's thesis. Gene expression profiling was performed by the Microarray Unit of the Genomics and Proteomics Core Facility (GPCF), DKFZ. Similarly, high-throughput DNA sequencing was performed by the High-Throughput Sequencing Unit while mass spectrometry experiments and subsequent analyses were performed by the Mass Spectrometry-based Protein Analysis Unit of the GPCF.

7. Acknowledgements

This work came to fruition through the efforts of a number of talented and hard-working individuals. A lot of credit goes to my supervisor—Prof. Dr. Holger Sültmann—for supporting and mentoring me through this stage of my academic career, for the guidance and constant encouragement, and for teaching me how to ask the important questions. I would like to thank my thesis advisory committee members—PD Dr. Odilia Popanda and Prof. Dr. med. Stefan Dünsing—for sharing their expertise and for providing constructive comments that made the project balanced and more coherent.

Special thanks to Prof. Dr. Sven Diederichs and Minakshi Gandhi for their invaluable support during the establishment of the ChIRP protocol. I am deeply grateful to Dr. Angela Schulz, Dr. Bernd Heßling, and Martin Schneider from the Genomics and Proteomics Core Facility of the DKFZ for generously sharing their expertise on DNA sequencing (Angela), and mass spectrometry (Bernd and Martin). I am grateful to Dr. Doreen Heckmann-Nötzel for the shared experience of exploring the complex world of lncRNAs and for being an excellent sounding board for ideas. I am indebted to PD Dr. Sabine Klauck for her support in and out of the laboratory, and for always taking the extra step in providing assistance—administrative or otherwise. Many thanks for the critical reading of the manuscript.

I would like to show my greatest appreciation to Sabrina Gerhardt for her endless patience and exceptional technical assistance. My deepest thanks go to Niclas Flosdorf for the stimulating discussions and for completing a portion of the project. The biggest and warmest thank you to the members of the B063 group: Leonie, Steffen, Anja, Sabrina M., Sebastian, Mone, Anka, Florian, Lisa, Louise, and Saskia. If not for the insightful scientific chats, easygoing company, delicious cakes, and oftentimes random and silly pop-culture conversations, completing this thesis would have been a thousand-fold more daunting.

To my family in Heidelberg: Karol, Luis, Simon, Gretchen, Jagoda, Ate Ina, Friedhelm, and Tita Yolly. Thank you for providing me the comforts of home thousands of miles away from home. To my Heidelberg/Mannheim/Wuppertal Filipino core group: Paulo, Angel, Josh, Fidel, Aeiou, Paul, Joseph, and David. Thank you for the company, the laughs, and the lounging sessions.

I owe a very important debt to Prof. Dr. Rey Garcia for letting my scientific interests flourish and for encouraging me to pursue advanced studies abroad. Likewise, I would like to extend my heartfelt thanks to Lorenz and Joanne: I am truly blessed by your friendships.

And to my family, thank you. I dedicate this achievement to you. I love you 3000.

8. References

1. Zhang, D., et al., *Prostate Luminal Progenitor Cells in Development and Cancer.* Trends Cancer, 2018. **4**(11): p. 769-783.
2. Wang, G., et al., *Genetics and biology of prostate cancer.* Genes Dev, 2018. **32**(17-18): p. 1105-1140.
3. Wang, Y., et al., *Cell differentiation lineage in the prostate.* Differentiation, 2001. **68**(4-5): p. 270-9.
4. Szczyrba, J., et al., *Neuroendocrine Cells of the Prostate Derive from the Neural Crest.* J Biol Chem, 2017. **292**(5): p. 2021-2031.
5. Abrahamsson, P.A., et al., *The course of neuroendocrine differentiation in prostatic carcinomas. An immunohistochemical study testing chromogranin A as an "endocrine marker".* Pathol Res Pract, 1989. **185**(3): p. 373-80.
6. Wang, Z.A., et al., *Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity.* Nat Cell Biol, 2013. **15**(3): p. 274-83.
7. Zhao, S.G., et al., *Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy.* JAMA Oncol, 2017. **3**(12): p. 1663-1672.
8. Smith, B.A., et al., *A basal stem cell signature identifies aggressive prostate cancer phenotypes.* Proc Natl Acad Sci U S A, 2015. **112**(47): p. E6544-52.
9. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
10. Leitzmann, M.F. and S. Rohrmann, *Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates.* Clin Epidemiol, 2012. **4**: p. 1-11.
11. Noone AM, H.N., Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). *SEER Cancer Statistics Review, 1975-2015, National Cancer Institute*. April 2018 [cited 2018; Available from: https://seer.cancer.gov/csr/1975_2015/.
12. Gronberg, H., *Prostate cancer epidemiology.* Lancet, 2003. **361**(9360): p. 859-64.
13. Torre, L.A., et al., *Global Cancer Incidence and Mortality Rates and Trends--An Update.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(1): p. 16-27.
14. Jalloh, M., et al., *Evaluation of 4,672 routine prostate biopsies performed in six African countries.* Journal Africain du Cancer / African Journal of Cancer, 2013. **5**(3): p. 144-154.
15. Hsing, A.W., et al., *High prevalence of screen detected prostate cancer in West Africans: implications for racial disparity of prostate cancer.* J Urol, 2014. **192**(3): p. 730-5.
16. Rebbeck, T.R. and G.P. Haas, *Temporal trends and racial disparities in global prostate cancer prevalence.* Can J Urol, 2014. **21**(5): p. 7496-506.
17. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland.* N Engl J Med, 2000. **343**(2): p. 78-85.
18. Verhage, B.A., et al., *Site-specific familial aggregation of prostate cancer.* Int J Cancer, 2004. **109**(4): p. 611-7.
19. Hjelmborg, J.B., et al., *The heritability of prostate cancer in the Nordic Twin Study of Cancer.* Cancer Epidemiol Biomarkers Prev, 2014. **23**(11): p. 2303-10.
20. Mucci, L.A., et al., *Familial Risk and Heritability of Cancer Among Twins in Nordic Countries.* JAMA, 2016. **315**(1): p. 68-76.
21. Berry, R., et al., *Linkage analyses at the chromosome 1 loci 1q24-25 (HPC1), 1q42.2-43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer.* Am J Hum Genet, 2000. **66**(2): p. 539-46.
22. Schleutker, J., et al., *A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: frequent HPCX linkage in families with late-onset disease.* Clin Cancer Res, 2000. **6**(12): p. 4810-5.

23. Xu, J., et al., *HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG).* Hum Genet, 2013. **132**(1): p. 5-14.
24. Ledet, E.M., et al., *Suggestive evidence of linkage identified at chromosomes 12q24 and 2p16 in African American prostate cancer families from Louisiana.* Prostate, 2012. **72**(9): p. 938-47.
25. Brown, W.M., et al., *Hereditary prostate cancer in African American families: linkage analysis using markers that map to five candidate susceptibility loci.* Br J Cancer, 2004. **90**(2): p. 510-4.
26. Matsui, H., et al., *Genomewide linkage analysis of familial prostate cancer in the Japanese population.* J Hum Genet, 2004. **49**(1): p. 9-15.
27. Eeles, R.A., et al., *Identification of seven new prostate cancer susceptibility loci through a genome-wide association study.* Nat Genet, 2009. **41**(10): p. 1116-21.
28. Schumacher, F.R., et al., *Genome-wide association study identifies new prostate cancer susceptibility loci.* Hum Mol Genet, 2011. **20**(19): p. 3867-75.
29. Haiman, C.A., et al., *Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21.* Nat Genet, 2011. **43**(6): p. 570-3.
30. Kote-Jarai, Z., et al., *Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study.* Nat Genet, 2011. **43**(8): p. 785-91.
31. Schumacher, F.R., et al., *Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci.* Nat Genet, 2018. **50**(7): p. 928-936.
32. Mateo, J., et al., *DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer.* N Engl J Med, 2015. **373**(18): p. 1697-708.
33. *TOPARP: A Phase II Trial of Olaparib in Patients With Advanced Castration Resistant Prostate Cancer*. Available from: https://ClinicalTrials.gov/show/NCT01682772.
34. Datta, K., et al., *Mechanism of lymph node metastasis in prostate cancer.* Future Oncol, 2010. **6**(5): p. 823-36.
35. Sartor, O. and J.S. de Bono, *Metastatic Prostate Cancer.* N Engl J Med, 2018. **378**(17): p. 1653-1654.
36. Force, U.S.P.S.T., et al., *Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement.* JAMA, 2018. **319**(18): p. 1901-1913.
37. Mottet, N., et al., *EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent.* Eur Urol, 2017. **71**(4): p. 618-629.
38. Vickers, A.J., et al., *Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study.* BMJ, 2013. **346**: p. f2023.
39. Litwin, M.S. and H.J. Tan, *The Diagnosis and Treatment of Prostate Cancer: A Review.* JAMA, 2017. **317**(24): p. 2532-2542.
40. Gordetsky, J. and J. Epstein, *Grading of prostatic adenocarcinoma: current state and prognostic implications.* Diagn Pathol, 2016. **11**: p. 25.
41. Sun F, O.O., Fontanarosa J, et al. *Therapies for Clinically Localized Prostate Cancer: Update of a 2008 Systematic Review [Internet]*. 2014 Dec. (Comparative Effectiveness Reviews, No. 146) Introduction:[Available from: https://www.ncbi.nlm.nih.gov/books/NBK269309/.
42. Fakhrejahani, F., R.A. Madan, and W.L. Dahut, *Management Options for Biochemically Recurrent Prostate Cancer.* Curr Treat Options Oncol, 2017. **18**(5): p. 26.
43. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2018.* CA Cancer J Clin, 2018. **68**(1): p. 7-30.
44. Dai, C., H. Heemers, and N. Sharifi, *Androgen Signaling in Prostate Cancer.* Cold Spring Harb Perspect Med, 2017. **7**(9).
45. Mahon, K.L., et al., *Pathways of chemotherapy resistance in castration-resistant prostate cancer.* Endocr Relat Cancer, 2011. **18**(4): p. R103-23.
46. Galazi, M., et al., *Precision medicine for prostate cancer.* Expert Rev Anticancer Ther, 2014. **14**(11): p. 1305-15.

47. Network, N.C.C. *NCCN Guidelines Version 2.2018 Prostate Cancer*. 2018 [cited 2018 27 July 2018]; Available from: https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf.
48. Briganti, A., et al., *Active Surveillance for Low-risk Prostate Cancer: The European Association of Urology Position in 2018.* Eur Urol, 2018. **74**(3): p. 357-368.
49. Attard, G., et al., *Prostate cancer.* The Lancet, 2016. **387**(10013): p. 70-82.
50. Salari, K., et al., *Active Surveillance of Prostate Cancer is a Viable Option in Men Younger Than 60 Years.* J Urol, 2019.
51. Bill-Axelson, A., et al., *Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial.* J Natl Cancer Inst, 2008. **100**(16): p. 1144-54.
52. Bill-Axelson, A., et al., *Radical prostatectomy or watchful waiting in early prostate cancer.* N Engl J Med, 2014. **370**(10): p. 932-42.
53. Theodorescu, D., *Cancer cryotherapy: evolution and biology*. Rev Urol, 2004. **6 Suppl 4**: p. S9-S19.
54. Ahmed, H.U., et al., *Focal therapy for localised unifocal and multifocal prostate cancer: a prospective development study.* Lancet Oncol, 2012. **13**(6): p. 622-32.
55. Azzouzi, A.R., et al., *TOOKAD(R) Soluble focal therapy: pooled analysis of three phase II studies assessing the minimally invasive ablation of localized prostate cancer.* World J Urol, 2015. **33**(7): p. 945-53.
56. Montero, A., et al., *Docetaxel for treatment of solid tumours: a systematic review of clinical data.* Lancet Oncol, 2005. **6**(4): p. 229-39.
57. Parker, C., et al., *Alpha emitter radium-223 and survival in metastatic prostate cancer.* N Engl J Med, 2013. **369**(3): p. 213-23.
58. Kantoff, P.W., et al., *Sipuleucel-T immunotherapy for castration-resistant prostate cancer.* N Engl J Med, 2010. **363**(5): p. 411-22.
59. Saad, F., J. McKiernan, and J. Eastham, *Rationale for zoledronic acid therapy in men with hormone-sensitive prostate cancer with or without bone metastasis.* Urol Oncol, 2006. **24**(1): p. 4-12.
60. Fizazi, K., et al., *Denosumab versus zoledronic acid for treatment of bone metastases in men with castration-resistant prostate cancer: a randomised, double-blind study.* Lancet, 2011. **377**(9768): p. 813-22.
61. Humeniuk, M.S., et al., *Platinum sensitivity in metastatic prostate cancer: does histology matter?* Prostate Cancer Prostatic Dis, 2018. **21**(1): p. 92-99.
62. Matsumoto, T., et al., *The androgen receptor in health and disease.* Annu Rev Physiol, 2013. **75**: p. 201-24.
63. Tan, M.H., et al., *Androgen receptor: structure, role in prostate cancer and drug discovery.* Acta Pharmacol Sin, 2015. **36**(1): p. 3-23.
64. Carroll, P.H. and J.L. Mohler, *NCCN Guidelines Updates: Prostate Cancer and Prostate Cancer Early Detection.* J Natl Compr Canc Netw, 2018. **16**(5S): p. 620-623.
65. Koshkin, V.S. and E.J. Small, *Apalutamide in the treatment of castrate-resistant prostate cancer: evidence from clinical trials.* Ther Adv Urol, 2018. **10**(12): p. 445-454.
66. Chandrasekar, T., et al., *Mechanisms of resistance in castration-resistant prostate cancer (CRPC).* Transl Androl Urol, 2015. **4**(3): p. 365-80.
67. Scott, L.J., *Abiraterone Acetate: A Review in Metastatic Castration-Resistant Prostrate Cancer.* Drugs, 2017. **77**(14): p. 1565-1576.
68. de Bono, J.S., et al., *Abiraterone and increased survival in metastatic prostate cancer.* N Engl J Med, 2011. **364**(21): p. 1995-2005.
69. Ryan, C.J., et al., *Abiraterone in metastatic prostate cancer without previous chemotherapy.* N Engl J Med, 2013. **368**(2): p. 138-48.
70. Scher, H.I., et al., *Increased survival with enzalutamide in prostate cancer after chemotherapy.* N Engl J Med, 2012. **367**(13): p. 1187-97.
71. Beer, T.M., et al., *Enzalutamide in metastatic prostate cancer before chemotherapy.* N Engl J Med, 2014. **371**(5): p. 424-33.

72. Smith, M.R., M.K. Yu, and E.J. Small, *Apalutamide and Metastasis-free Survival in Prostate Cancer.* N Engl J Med, 2018. **378**(26): p. 2542.
73. Angeles, A.K., et al., *Genome-Based Classification and Therapy of Prostate Cancer.* Diagnostics (Basel), 2018. **8**(3).
74. Fraser, M., et al., *Genomic hallmarks of localized, non-indolent prostate cancer.* Nature, 2017. **541**(7637): p. 359-364.
75. Cancer Genome Atlas Research, N., *The Molecular Taxonomy of Primary Prostate Cancer.* Cell, 2015. **163**(4): p. 1011-25.
76. Robinson, D., et al., *Integrative clinical genomics of advanced prostate cancer.* Cell, 2015. **161**(5): p. 1215-1228.
77. Weischenfeldt, J., et al., *Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer.* Cancer Cell, 2013. **23**(2): p. 159-70.
78. Ren, S., et al., *Whole-genome and Transcriptome Sequencing of Prostate Cancer Identify New Genetic Alterations Driving Disease Progression.* Eur Urol, 2017.
79. Quigley, D.A., et al., *Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer.* Cell, 2018. **175**(3): p. 889.
80. Gerhauser, C., et al., *Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories.* Cancer Cell, 2018. **34**(6): p. 996-1011 e8.
81. 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.
82. Ylipaa, A., et al., *Transcriptome Sequencing Reveals PCAT5 as a Novel ERG-Regulated Long Noncoding RNA in Prostate Cancer.* Cancer Res, 2015. **75**(19): p. 4026-31.
83. Bhasin, J.M., et al., *Methylome-wide Sequencing Detects DNA Hypermethylation Distinguishing Indolent from Aggressive Prostate Cancer.* Cell Rep, 2015. **13**(10): p. 2135-46.
84. Kim, J.H., et al., *Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer.* Genome Res, 2011. **21**(7): p. 1028-41.
85. Brocks, D., et al., *Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer.* Cell Rep, 2014. **8**(3): p. 798-806.
86. Kron, K.J., et al., *TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer.* Nat Genet, 2017. **49**(9): p. 1336-1345.
87. Latonen, L., et al., *Integrative proteomics in prostate cancer uncovers robustness against genomic and transcriptomic aberrations during disease progression.* Nat Commun, 2018. **9**(1): p. 1176.
88. Muller, A.K., et al., *Proteomic Characterization of Prostate Cancer to Distinguish Nonmetastasizing and Metastasizing Primary Tumors and Lymph Node Metastases.* Neoplasia, 2018. **20**(2): p. 140-151.
89. Sinha, A., et al., *The Proteogenomic Landscape of Curable Prostate Cancer.* Cancer Cell, 2019. **35**(3): p. 414-427 e6.
90. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.
91. Rubin, M.A. and F. Demichelis, *The Genomics of Prostate Cancer: emerging understanding with technologic advances.* Mod Pathol, 2018. **31**(S1): p. S1-11.
92. Hieronymus, H., et al., *Copy number alteration burden predicts prostate cancer relapse.* Proc Natl Acad Sci U S A, 2014. **111**(30): p. 11139-44.
93. Tomlins, S.A., et al., *Role of the TMPRSS2-ERG gene fusion in prostate cancer.* Neoplasia, 2008. **10**(2): p. 177-88.
94. Tomlins, S.A., et al., *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer.* Science, 2005. **310**(5748): p. 644-8.
95. Sandoval, G.J., et al., *Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis.* Mol Cell, 2018. **71**(4): p. 554-566 e7.
96. Yu, J., et al., *An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression.* Cancer Cell, 2010. **17**(5): p. 443-54.
97. Ratz, L., et al., *TMPRSS2:ERG gene fusion variants induce TGF-beta signaling and epithelial to mesenchymal transition in human prostate cancer cells.* Oncotarget, 2017. **8**(15): p. 25115-25130.

98. Dryden, N.H., et al., *The transcription factor Erg controls endothelial cell quiescence by repressing activity of nuclear factor (NF)-kappaB p65.* J Biol Chem, 2012. **287**(15): p. 12331-42.
99. Adamo, P. and M.R. Ladomery, *The oncogene ERG: a key factor in prostate cancer.* Oncogene, 2016. **35**(4): p. 403-14.
100. Beltran, H., et al., *Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets.* Cancer Discov, 2011. **1**(6): p. 487-95.
101. Terry, S. and H. Beltran, *The many faces of neuroendocrine differentiation in prostate cancer progression.* Front Oncol, 2014. **4**: p. 60.
102. Beltran, H., et al., *Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer.* Nat Med, 2016. **22**(3): p. 298-305.
103. Aparicio, A.M., et al., *Combined Tumor Suppressor Defects Characterize Clinically Defined Aggressive Variant Prostate Cancers.* Clin Cancer Res, 2016. **22**(6): p. 1520-30.
104. Dardenne, E., et al., *N-Myc Induces an EZH2-Mediated Transcriptional Program Driving Neuroendocrine Prostate Cancer.* Cancer Cell, 2016. **30**(4): p. 563-577.
105. Svensson, C., et al., *REST mediates androgen receptor actions on gene repression and predicts early recurrence of prostate cancer.* Nucleic Acids Res, 2014. **42**(2): p. 999-1015.
106. Eddy, S.R., *Non-coding RNA genes and the modern RNA world.* Nat Rev Genet, 2001. **2**(12): p. 919-29.
107. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
108. 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.
109. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.
110. Brown, C.J., et al., *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome.* Nature, 1991. **349**(6304): p. 38-44.
111. Ulitsky, I., et al., *Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution.* Cell, 2011. **147**(7): p. 1537-50.
112. Grote, P., et al., *The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse.* Dev Cell, 2013. **24**(2): p. 206-14.
113. Sauvageau, M., et al., *Multiple knockout mouse models reveal lincRNAs are required for life and brain development.* Elife, 2013. **2**: p. e01749.
114. Lin, C.-P. and L. He, *Noncoding RNAs in Cancer Development.* Annual Review of Cancer Biology, 2017. **1**(1): p. 163-184.
115. Schmitt, A.M. and H.Y. Chang, *Long Noncoding RNAs in Cancer Pathways.* Cancer Cell, 2016. **29**(4): p. 452-463.
116. Pickl, J.M., et al., *Novel RNA markers in prostate cancer: functional considerations and clinical translation.* Biomed Res Int, 2014. **2014**: p. 765207.
117. Pickl, J.M., et al., *Ago-RIP-Seq identifies Polycomb repressive complex I member CBX7 as a major target of miR-375 in prostate cancer progression.* Oncotarget, 2016. **7**(37): p. 59589-59603.
118. Pichler, M. and G.A. Calin, *MicroRNAs in cancer: from developmental genes in worms to their clinical application in patients.* Br J Cancer, 2015. **113**(4): p. 569-73.
119. Ng, K.W., et al., *Piwi-interacting RNAs in cancer: emerging functions and clinical utility.* Mol Cancer, 2016. **15**: p. 5.
120. Moyano, M. and G. Stefani, *piRNA involvement in genome stability and human cancer.* J Hematol Oncol, 2015. **8**: p. 38.
121. Appaiah, H.N., et al., *Persistent upregulation of U6:SNORD44 small RNA ratio in the serum of breast cancer patients.* Breast Cancer Res, 2011. **13**(5): p. R86.
122. Mannoor, K., J. Liao, and F. Jiang, *Small nucleolar RNAs in cancer.* Biochim Biophys Acta, 2012. **1826**(1): p. 121-8.
123. Chen, L., et al., *SNORD76, a box C/D snoRNA, acts as a tumor suppressor in glioblastoma.* Sci Rep, 2015. **5**: p. 8588.

124. Siprashvili, Z., et al., *The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer.* Nat Genet, 2016. **48**(1): p. 53-8.
125. Goodarzi, H., et al., *Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement.* Cell, 2015. **161**(4): p. 790-802.
126. Green, D., W.D. Fraser, and T. Dalmay, *Transfer RNA-derived small RNAs in the cancer transcriptome.* Pflugers Arch, 2016. **468**(6): p. 1041-7.
127. McHugh, C.A., et al., *The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.* Nature, 2015. **521**(7551): p. 232-6.
128. Engreitz, J.M., et al., *The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome.* Science, 2013. **341**(6147): p. 1237973.
129. Lee, S., et al., *Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins.* Cell, 2016. **164**(1-2): p. 69-80.
130. 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.
131. Pasmant, E., et al., *Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF.* Cancer Res, 2007. **67**(8): p. 3963-9.
132. Askarian-Amiri, M.E., et al., *SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer.* RNA, 2011. **17**(5): p. 878-91.
133. Latge, G., et al., *Natural Antisense Transcripts: Molecular Mechanisms and Implications in Breast Cancers.* Int J Mol Sci, 2018. **19**(1).
134. Yu, G., et al., *Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression.* Mol Cancer Ther, 2014. **13**(12): p. 3086-97.
135. Karreth, F.A., et al., *The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo.* Cell, 2015. **161**(2): p. 319-32.
136. Poliseno, L., et al., *A coding-independent function of gene and pseudogene mRNAs regulates tumour biology.* Nature, 2010. **465**(7301): p. 1033-8.
137. Liu, Y., et al., *Current Advances on the Important Roles of Enhancer RNAs in Gene Regulation and Cancer.* Biomed Res Int, 2018. **2018**: p. 2405351.
138. McCleland, M.L., et al., *CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer.* J Clin Invest, 2016. **126**(2): p. 639-52.
139. Zhao, Y., et al., *Activation of P-TEFb by Androgen Receptor-Regulated Enhancer RNAs in Castration-Resistant Prostate Cancer.* Cell Rep, 2016. **15**(3): p. 599-610.
140. Terracciano, D., et al., *The role of a new class of long noncoding RNAs transcribed from ultraconserved regions in cancer.* Biochim Biophys Acta Rev Cancer, 2017. **1868**(2): p. 449-455.
141. Olivieri, M., et al., *Long non-coding RNA containing ultraconserved genomic region 8 promotes bladder cancer tumorigenesis.* Oncotarget, 2016. **7**(15): p. 20636-54.
142. Mestdagh, P., et al., *An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours.* Oncogene, 2010. **29**(24): p. 3583-92.
143. Dragomir, M. and G.A. Calin, *Circular RNAs in Cancer - Lessons Learned From microRNAs.* Front Oncol, 2018. **8**: p. 179.
144. Li, X., L. Yang, and L.L. Chen, *The Biogenesis, Functions, and Challenges of Circular RNAs.* Mol Cell, 2018. **71**(3): p. 428-442.
145. Bachmayr-Heyda, A., et al., *Correlation of circular RNA abundance with proliferation--exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues.* Sci Rep, 2015. **5**: p. 8057.
146. Guo, H., et al., *Mammalian microRNAs predominantly act to decrease target mRNA levels.* Nature, 2010. **466**(7308): p. 835-40.
147. Roberts, T.C., *The MicroRNA Biology of the Mammalian Nucleus.* Mol Ther Nucleic Acids, 2014. **3**: p. e188.
148. Liu, J., et al., *Argonaute2 is the catalytic engine of mammalian RNAi.* Science, 2004. **305**(5689): p. 1437-41.

149. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.
150. Lu, J., et al., *MicroRNA expression profiles classify human cancers.* Nature, 2005. **435**(7043): p. 834-8.
151. Wach, S., et al., *MicroRNA profiles of prostate carcinoma detected by multiplatform microRNA screening.* Int J Cancer, 2012. **130**(3): p. 611-21.
152. Tong, A.W., et al., *MicroRNA profile analysis of human prostate cancers.* Cancer Gene Ther, 2009. **16**(3): p. 206-16.
153. Szczyrba, J., et al., *The microRNA profile of prostate carcinoma obtained by deep sequencing.* Mol Cancer Res, 2010. **8**(4): p. 529-38.
154. Brase, J.C., et al., *Circulating miRNAs are correlated with tumor progression in prostate cancer.* Int J Cancer, 2011. **128**(3): p. 608-16.
155. Majid, S., et al., *miRNA-34b inhibits prostate cancer through demethylation, active chromatin modifications, and AKT pathways.* Clin Cancer Res, 2013. **19**(1): p. 73-84.
156. Vanacore, D., et al., *Micrornas in prostate cancer: an overview.* Oncotarget, 2017. **8**(30): p. 50240-50251.
157. 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.
158. Church, D.M., et al., *Lineage-specific biology revealed by a finished genome assembly of the mouse.* PLoS Biol, 2009. **7**(5): p. e1000112.
159. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.* Genes Dev, 2011. **25**(18): p. 1915-27.
160. Ulitsky, I. and D.P. Bartel, *lincRNAs: genomics, evolution, and mechanisms.* Cell, 2013. **154**(1): p. 26-46.
161. Zhao, Y., et al., *NONCODE 2016: an informative and valuable data source of long non-coding RNAs.* Nucleic Acids Res, 2016. **44**(D1): p. D203-8.
162. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.* Cell, 2010. **142**(3): p. 409-19.
163. Ji, P., et al., *MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer.* Oncogene, 2003. **22**(39): p. 8031-41.
164. 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.
165. 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.
166. 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.
167. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.
168. 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.
169. Tsuiji, H., et al., *Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC repeats and associates with splicing factor-1.* Genes Cells, 2011. **16**(5): p. 479-90.
170. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.
171. Hung, T., et al., *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.* Nat Genet, 2011. **43**(7): p. 621-9.
172. Wang, J., et al., *CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer.* Nucleic Acids Res, 2010. **38**(16): p. 5366-83.
173. 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.

174. Matsui, K., et al., *Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes.* Hepatology, 2008. **47**(2): p. 686-97.
175. Aird, J., et al., *Carcinogenesis in prostate cancer: The role of long non-coding RNAs.* Noncoding RNA Res, 2018. **3**(1): p. 29-38.
176. Walsh, A.L., et al., *Long noncoding RNAs and prostate carcinogenesis: the missing 'linc'?* Trends Mol Med, 2014. **20**(8): p. 428-36.
177. Bussemakers, M.J., et al., *DD3: a new prostate-specific gene, highly overexpressed in prostate cancer.* Cancer Res, 1999. **59**(23): p. 5975-9.
178. Lemos, A.E., et al., *PCA3 long noncoding RNA modulates the expression of key cancer-related genes in LNCaP prostate cancer cells.* Tumour Biol, 2016. **37**(8): p. 11339-48.
179. Ozgur, E., et al., *PCA3 Silencing Sensitizes Prostate Cancer Cells to Enzalutamide-mediated Androgen Receptor Blockade.* Anticancer Res, 2017. **37**(7): p. 3631-3637.
180. Takayama, K., et al., *Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer.* EMBO J, 2013. **32**(12): p. 1665-80.
181. Prensner, J.R., et al., *PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer.* Cancer Res, 2014. **74**(6): p. 1651-60.
182. Petrovics, G., et al., *Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients.* Oncogene, 2004. **23**(2): p. 605-11.
183. Chung, S., et al., *Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility.* Cancer Sci, 2011. **102**(1): p. 245-52.
184. Yang, L., et al., *lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs.* Nature, 2013. **500**(7464): p. 598-602.
185. Barra, J. and E. Leucci, *Probing Long Non-coding RNA-Protein Interactions.* Front Mol Biosci, 2017. **4**: p. 45.
186. McHugh, C.A., P. Russell, and M. Guttman, *Methods for comprehensive experimental identification of RNA-protein interactions.* Genome Biol, 2014. **15**(1): p. 203.
187. Nainar, S., C. Feng, and R.C. Spitale, *Chemical Tools for Dissecting the Role of lncRNAs in Epigenetic Regulation.* ACS Chem Biol, 2016. **11**(8): p. 2091-100.
188. Ule, J., et al., *CLIP identifies Nova-regulated RNA networks in the brain.* Science, 2003. **302**(5648): p. 1212-5.
189. Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.* Nature, 2009. **460**(7254): p. 479-86.
190. Garzia, A., et al., *Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins.* Methods, 2017. **118-119**: p. 24-40.
191. Konig, J., et al., *iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution.* J Vis Exp, 2011(50).
192. Simon, M.D., et al., *The genomic binding sites of a noncoding RNA.* Proc Natl Acad Sci U S A, 2011. **108**(51): p. 20497-502.
193. West, J.A., et al., *The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites.* Mol Cell, 2014. **55**(5): p. 791-802.
194. 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.
195. Chu, C., et al., *Systematic discovery of Xist RNA binding proteins.* Cell, 2015. **161**(2): p. 404-16.
196. Engreitz, J., E.S. Lander, and M. Guttman, *RNA antisense purification (RAP) for mapping RNA interactions with chromatin.* Methods Mol Biol, 2015. **1262**: p. 183-97.
197. 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.
198. Hacisuleyman, E., et al., *Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre.* Nat Struct Mol Biol, 2014. **21**(2): p. 198-206.
199. Borner, K., et al., *Robust RNAi enhancement via human Argonaute-2 overexpression from plasmids, viral vectors and cell lines.* Nucleic Acids Res, 2013. **41**(21): p. e199.
200. Wang, L., et al., *CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model.* Nucleic Acids Res, 2013. **41**(6): p. e74.

201. Kong, L., et al., *CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W345-9.
202. McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nat Biotechnol, 2010. **28**(5): p. 495-501.
203. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
204. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
205. Geissmann, Q., *OpenCFU, a new free and open-source software to count cell colonies and other circular objects.* PLoS One, 2013. **8**(2): p. e54072.
206. Team, R., *RStudio: Integrated Development for R.* 2015, RStudio Inc.: Boston, MA, USA.
207. Lin, J., et al., *Targeting activated Akt with GDC-0068, a novel selective Akt inhibitor that is efficacious in multiple tumor models.* Clin Cancer Res, 2013. **19**(7): p. 1760-72.
208. Lai, F., E. Blumenthal, and R. Shiekhattar, *Detection and Analysis of Long Noncoding RNAs.* Methods Enzymol, 2016. **573**: p. 421-44.
209. Goldman, M., et al., *The UCSC Xena platform for public and private cancer genomics data visualization and interpretation.* bioRxiv, 2019: p. 326470.
210. Taberlay, P.C., et al., *Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer.* Genome Res, 2014. **24**(9): p. 1421-32.
211. Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W3-W10.
212. Andrews, S. *FastQC: A quality control tool for high throughput sequence data*. 2018; Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
213. Blankenberg, D., et al., *Manipulation of FASTQ data with Galaxy.* Bioinformatics, 2010. **26**(14): p. 1783-5.
214. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
215. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
216. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.
217. Haeussler, M., et al., *The UCSC Genome Browser database: 2019 update.* Nucleic Acids Res, 2019. **47**(D1): p. D853-D858.
218. Khan, A., et al., *JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework.* Nucleic Acids Res, 2018. **46**(D1): p. D1284.
219. Taberlay, P.C., et al., *Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations.* Genome Res, 2016. **26**(6): p. 719-31.
220. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Proc Int Conf Intell Syst Mol Biol, 1994. **2**: p. 28-36.
221. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-40.
222. Iyer, M.K., et al., *The landscape of long noncoding RNAs in the human transcriptome.* Nat Genet, 2015. **47**(3): p. 199-208.
223. Werner, M.S., et al., *Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription.* Nat Struct Mol Biol, 2017. **24**(7): p. 596-603.
224. Ransohoff, J.D., Y. Wei, and P.A. Khavari, *The functions and unique features of long intergenic non-coding RNA.* Nat Rev Mol Cell Biol, 2018. **19**(3): p. 143-157.
225. Washietl, S., M. Kellis, and M. Garber, *Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals.* Genome Res, 2014. **24**(4): p. 616-28.
226. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.

227. Xu, J., et al., *A comprehensive overview of lncRNA annotation resources.* Brief Bioinform, 2017. **18**(2): p. 236-249.
228. Frohman, M.A., M.K. Dush, and G.R. Martin, *Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer.* Proc Natl Acad Sci U S A, 1988. **85**(23): p. 8998-9002.
229. Lin, M.F., I. Jungreis, and M. Kellis, *PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions.* Bioinformatics, 2011. **27**(13): p. i275-82.
230. Dijkers, P.F., et al., *Expression of the pro-apoptotic Bcl-2 family member Bim is regulated by the forkhead transcription factor FKHR-L1.* Curr Biol, 2000. **10**(19): p. 1201-4.
231. Dijkers, P.F., et al., *Forkhead transcription factor FKHR-L1 modulates cytokine-dependent transcriptional regulation of p27(KIP1).* Mol Cell Biol, 2000. **20**(24): p. 9138-48.
232. Tran, H., et al., *DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein.* Science, 2002. **296**(5567): p. 530-4.
233. Valis, K., et al., *Hippo/Mst1 stimulates transcription of the proapoptotic mediator NOXA in a FoxO1-dependent manner.* Cancer Res, 2011. **71**(3): p. 946-54.
234. Vlietstra, R.J., et al., *Frequent inactivation of PTEN in prostate cancer cell lines and xenografts.* Cancer Res, 1998. **58**(13): p. 2720-3.
235. McMenamin, M.E., et al., *Loss of PTEN expression in paraffin-embedded primary prostate cancer correlates with high Gleason score and advanced stage.* Cancer Res, 1999. **59**(17): p. 4291-6.
236. Mertz, K.D., et al., *Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model.* Neoplasia, 2007. **9**(3): p. 200-6.
237. Negre, N., et al., *A cis-regulatory map of the Drosophila genome.* Nature, 2011. **471**(7339): p. 527-31.
238. Hollenhorst, P.C., et al., *The ETS gene ETV4 is required for anchorage-independent growth and a cell proliferation gene expression program in PC3 prostate cells.* Genes & cancer, 2010. **1**(10): p. 1044-1052.
239. Machyna, M. and M.D. Simon, *Catching RNAs on chromatin using hybridization capture methods.* Brief Funct Genomics, 2018. **17**(2): p. 96-103.
240. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.
241. Heintzman, N.D., et al., *Histone modifications at human enhancers reflect global cell-type-specific gene expression.* Nature, 2009. **459**(7243): p. 108-12.
242. Shackelford, D.B. and R.J. Shaw, *The LKB1-AMPK pathway: metabolism and growth control in tumour suppression.* Nat Rev Cancer, 2009. **9**(8): p. 563-75.
243. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
244. Hermeking, H., *The 14-3-3 cancer connection.* Nat Rev Cancer, 2003. **3**(12): p. 931-43.
245. Zhang, X., et al., *Akt, FoxO and regulation of apoptosis.* Biochim Biophys Acta, 2011. **1813**(11): p. 1978-86.
246. Li, J., et al., *TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer.* Cancer Res, 2015. **75**(18): p. 3728-37.
247. Cunningham, D. and Z. You, *In vitro and in vivo model systems used in prostate cancer research.* J Biol Methods, 2015. **2**(1).
248. Tian, B. and J.L. Manley, *Alternative polyadenylation of mRNA precursors.* Nat Rev Mol Cell Biol, 2017. **18**(1): p. 18-30.
249. Hu, J., et al., *Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation.* RNA, 2005. **11**(10): p. 1485-93.
250. Housman, G. and I. Ulitsky, *Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs.* Biochim Biophys Acta, 2016. **1859**(1): p. 31-40.
251. Dinger, M.E., et al., *Differentiating protein-coding and noncoding RNA: challenges and ambiguities.* PLoS Comput Biol, 2008. **4**(11): p. e1000176.

252. Clamp, M., et al., *Distinguishing protein-coding and noncoding genes in the human genome.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19428-33.
253. Claverie, J.M., I. Sauvaget, and L. Bougueleret, *K-tuple frequency analysis: from intron/exon discrimination to T-cell epitope mapping.* Methods Enzymol, 1990. **183**: p. 237-52.
254. van der Vos, K.E. and P.J. Coffer, *FOXO-binding partners: it takes two to tango.* Oncogene, 2008. **27**(16): p. 2289-99.
255. Furuyama, T., et al., *Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues.* Biochem J, 2000. **349**(Pt 2): p. 629-34.
256. Vogel, M.J., et al., *FOXO1 repression contributes to block of plasma cell differentiation in classical Hodgkin lymphoma.* Blood, 2014. **124**(20): p. 3118-29.
257. Mei, Y., et al., *Regulation of neuroblastoma differentiation by forkhead transcription factors FOXO1/3/4 through the receptor tyrosine kinase PDGFRA.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 4898-903.
258. Paik, J.H., et al., *FoxOs are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis.* Cell, 2007. **128**(2): p. 309-23.
259. Dansen, T.B. and B.M. Burgering, *Unravelling the tumor-suppressive functions of FOXO proteins.* Trends Cell Biol, 2008. **18**(9): p. 421-9.
260. Ghaffari, S., et al., *Cytokines and BCR-ABL mediate suppression of TRAIL-induced apoptosis through inhibition of forkhead FOXO3a transcription factor.* Proc Natl Acad Sci U S A, 2003. **100**(11): p. 6523-8.
261. Brunet, A., et al., *Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor.* Cell, 1999. **96**(6): p. 857-68.
262. Brownawell, A.M., et al., *Inhibition of nuclear import by protein kinase B (Akt) regulates the subcellular distribution and activity of the forkhead transcription factor AFX.* Mol Cell Biol, 2001. **21**(10): p. 3534-46.
263. Rena, G., et al., *Two novel phosphorylation sites on FKHR that are critical for its nuclear exclusion.* EMBO J, 2002. **21**(9): p. 2263-71.
264. Carracedo, A. and P.P. Pandolfi, *The PTEN-PI3K pathway: of feedbacks and cross-talks.* Oncogene, 2008. **27**(41): p. 5527-41.
265. Maehama, T. and J.E. Dixon, *The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate.* J Biol Chem, 1998. **273**(22): p. 13375-8.
266. Zhan, Q., *Gadd45a, a p53- and BRCA1-regulated stress protein, in cellular response to DNA damage.* Mutat Res, 2005. **569**(1-2): p. 133-43.
267. Jin, S., et al., *The GADD45 inhibition of Cdc2 kinase correlates with GADD45-mediated growth suppression.* J Biol Chem, 2000. **275**(22): p. 16602-8.
268. Schulz, W.A., et al., *Factor interaction analysis for chromosome 8 and DNA methylation alterations highlights innate immune response suppression and cytoskeletal changes in prostate cancer.* Mol Cancer, 2007. **6**: p. 14.
269. Doerflinger, M., J.A. Glab, and H. Puthalakath, *BH3-only proteins: a 20-year stock-take.* FEBS J, 2015. **282**(6): p. 1006-16.
270. Wang, F., et al., *Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment.* Proc Natl Acad Sci U S A, 2012. **109**(16): p. 6078-83.
271. Kim, T.D., S. Shin, and R. Janknecht, *ETS transcription factor ERG cooperates with histone demethylase KDM4A.* Oncol Rep, 2016. **35**(6): p. 3679-88.
272. Aytes, A., et al., *ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer.* Proc Natl Acad Sci U S A, 2013. **110**(37): p. E3506-15.
273. Hollenhorst, P.C., et al., *The ETS gene ETV4 is required for anchorage-independent growth and a cell proliferation gene expression program in PC3 prostate cells.* Genes Cancer, 2011. **1**(10): p. 1044-1052.
274. Smolle, M.A., et al., *Current Insights into Long Non-Coding RNAs (LncRNAs) in Prostate Cancer.* Int J Mol Sci, 2017. **18**(2).

275. Quinn, J.J., et al., *Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification.* Nat Biotechnol, 2014. **32**(9): p. 933-940.
276. Chu, C., R.C. Spitale, and H.Y. Chang, *Technologies to probe functions and mechanisms of long noncoding RNAs.* Nat Struct Mol Biol, 2015. **22**(1): p. 29-35.
277. Calo, E. and J. Wysocka, *Modification of enhancer chromatin: what, how, and why?* Mol Cell, 2013. **49**(5): p. 825-37.
278. Long, H.K., S.L. Prescott, and J. Wysocka, *Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution.* Cell, 2016. **167**(5): p. 1170-1187.
279. Pradeepa, M.M., et al., *Histone H3 globular domain acetylation identifies a new class of enhancers.* Nat Genet, 2016. **48**(6): p. 681-6.
280. Local, A., et al., *Identification of H3K4me1-associated proteins at mammalian enhancers.* Nat Genet, 2018. **50**(1): p. 73-82.
281. Kawabe, Y., et al., *Carm1 regulates Pax7 transcriptional activity through MLL1/2 recruitment during asymmetric satellite stem cell divisions.* Cell Stem Cell, 2012. **11**(3): p. 333-45.
282. Lee, S., et al., *Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15392-7.
283. Mo, R., S.M. Rao, and Y.J. Zhu, *Identification of the MLL2 complex as a coactivator for estrogen receptor alpha.* J Biol Chem, 2006. **281**(23): p. 15714-20.
284. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.
285. Bacolla, A., G. Wang, and K.M. Vasquez, *New Perspectives on DNA and RNA Triplexes As Effectors of Biological Activity.* PLoS Genet, 2015. **11**(12): p. e1005696.
286. Szabat, M., E. Kierzek, and R. Kierzek, *Modified RNA triplexes: Thermodynamics, structure and biological potential.* Sci Rep, 2018. **8**(1): p. 13023.
287. Castello, A., et al., *Comprehensive Identification of RNA-Binding Domains in Human Cells.* Mol Cell, 2016. **63**(4): p. 696-710.
288. Babula, J.J. and J.Y. Liu, *Integrate Omics Data and Molecular Dynamics Simulations toward Better Understanding of Human 14-3-3 Interactomes and Better Drugs for Cancer Therapy.* J Genet Genomics, 2015. **42**(10): p. 531-547.
289. Pennington, K.L., et al., *The dynamic and stress-adaptive signaling hub of 14-3-3: emerging mechanisms of regulation and context-dependent protein-protein interactions.* Oncogene, 2018. **37**(42): p. 5587-5604.
290. Biggs, W.H., 3rd, et al., *Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1.* Proc Natl Acad Sci U S A, 1999. **96**(13): p. 7421-6.
291. Alessi, D.R., et al., *Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p70 S6 kinase.* FEBS Lett, 1996. **399**(3): p. 333-8.
292. Obsil, T. and V. Obsilova, *Structural basis for DNA recognition by FOXO proteins.* Biochim Biophys Acta, 2011. **1813**(11): p. 1946-53.
293. Van Der Heide, L.P., M.F. Hoekman, and M.P. Smidt, *The ins and outs of FoxO shuttling: mechanisms of FoxO translocation and transcriptional regulation.* Biochem J, 2004. **380**(Pt 2): p. 297-309.
294. Plas, D.R. and C.B. Thompson, *Akt activation promotes degradation of tuberin and FOXO3a via the proteasome.* J Biol Chem, 2003. **278**(14): p. 12361-6.
295. Matsuzaki, H., et al., *Insulin-induced phosphorylation of FKHR (Foxo1) targets to proteasomal degradation.* Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11285-90.
296. Chaudhri, M., M. Scarabel, and A. Aitken, *Mammalian and yeast 14-3-3 isoforms form distinct patterns of dimers in vivo.* Biochem Biophys Res Commun, 2003. **300**(3): p. 679-85.
297. Carver, B.S., et al., *Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate.* Nat Genet, 2009. **41**(5): p. 619-24.
298. King, J.C., et al., *Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis.* Nat Genet, 2009. **41**(5): p. 524-6.

299. Zong, Y., et al., *ETS family transcription factors collaborate with alternative signaling pathways to induce carcinoma from adult murine prostate cells.* Proc Natl Acad Sci U S A, 2009. **106**(30): p. 12465-70.
300. Li, R., et al., *Forkhead protein FKHR and its phosphorylated form p-FKHR in human prostate cancer.* Hum Pathol, 2007. **38**(10): p. 1501-7.
301. Shukla, S., et al., *Deregulation of FoxO3a accelerates prostate cancer progression in TRAMP mice.* Prostate, 2013. **73**(14): p. 1507-17.
302. Lynch, R.L., et al., *The progression of LNCaP human prostate cancer cells to androgen independence involves decreased FOXO3a expression and reduced p27KIP1 promoter transactivation.* Mol Cancer Res, 2005. **3**(3): p. 163-9.
303. Zhang, H., et al., *FOXO1 inhibits Runx2 transcriptional activity and prostate cancer cell migration and invasion.* Cancer Res, 2011. **71**(9): p. 3257-67.
304. Modur, V., et al., *FOXO proteins regulate tumor necrosis factor-related apoptosis inducing ligand expression. Implications for PTEN mutation in prostate cancer.* J Biol Chem, 2002. **277**(49): p. 47928-37.
305. Yang, Y., et al., *Loss of FOXO1 Cooperates with TMPRSS2-ERG Overexpression to Promote Prostate Tumorigenesis and Cell Invasion.* Cancer Res, 2017. **77**(23): p. 6524-6537.
306. Su, Y.W., et al., *14-3-3sigma regulates B-cell homeostasis through stabilization of FOXO1.* Proc Natl Acad Sci U S A, 2011. **108**(4): p. 1555-60.
307. Shen, Y., et al., *FoxO1 inhibits transcription and membrane trafficking of epithelial Na+ channel.* J Cell Sci, 2015. **128**(19): p. 3621-30.
308. Arimoto-Ishida, E., et al., *Inhibition of phosphorylation of a forkhead transcription factor sensitizes human ovarian cancer cells to cisplatin.* Endocrinology, 2004. **145**(4): p. 2014-22.
309. Tzivion, G., M. Dobson, and G. Ramakrishnan, *FoxO transcription factors; Regulation by AKT and 14-3-3 proteins.* Biochim Biophys Acta, 2011. **1813**(11): p. 1938-45.
310. Zhang, B., et al., *The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult.* Cell Rep, 2012. **2**(1): p. 111-23.
311. Mello, S.S., et al., *Neat1 is a p53-inducible lincRNA essential for transformation suppression.* Genes Dev, 2017. **31**(11): p. 1095-1108.
312. 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.
313. Gutschner, T., et al., *The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells.* Cancer Res, 2013. **73**(3): p. 1180-9.
314. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-6.
315. Leucci, E., et al., *Melanoma addiction to the long non-coding RNA SAMMSON.* Nature, 2016. **531**(7595): p. 518-22.
316. Arun, G., et al., *Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss.* Genes Dev, 2016. **30**(1): p. 34-51.
317. Diermeier, S.D., et al., *Mammary Tumor-Associated RNAs Impact Tumor Cell Proliferation, Invasion, and Migration.* Cell Rep, 2016. **17**(1): p. 261-274.
318. Balkwill, F.R., M. Capasso, and T. Hagemann, *The tumor microenvironment at a glance.* J Cell Sci, 2012. **125**(Pt 23): p. 5591-6.
319. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.
320. Arun, G., S.D. Diermeier, and D.L. Spector, *Therapeutic Targeting of Long Non-Coding RNAs in Cancer.* Trends Mol Med, 2018. **24**(3): p. 257-277.
321. Finkel, R.S., et al., *Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study.* Lancet, 2016. **388**(10063): p. 3017-3026.
322. Geary, R.S., B.F. Baker, and S.T. Crooke, *Clinical and preclinical pharmacokinetics and pharmacodynamics of mipomersen (kynamro((R))): a second-generation antisense oligonucleotide inhibitor of apolipoprotein B.* Clin Pharmacokinet, 2015. **54**(2): p. 133-46.
323. Goemans, N.M., et al., *Systemic administration of PRO051 in Duchenne's muscular dystrophy.* N Engl J Med, 2011. **364**(16): p. 1513-22.

9. Appendix

9.1. Full-length sequences of cloned lncRNAs

Supplementary Table 9-1. Sequences of lncRNAs as verified by Sanger sequencing.

Full-length LINC00844 cDNA sequence (405 bp)

Source: Normal human prostate

AGAGACAAAGGAAACACAGAGACATAGACATGGATCTGGGAAATACACCTTTTGCTACTCGTTCA
GTTTTAGCAAGGAGGTTTCTTGCATGGCTAAGCAAAACTTAAACTTCCTCTGAGAATTACAGGAAT
TACAGGACCTGACAAAGCTATGAAGATTAAAACCTATAGGAAGAAAATCTGAACCAGAAACAGTA
TGGCAGAATTGGGATCTGACTCACAGAGGGAAGAACTTATAATTCTTCACAGGTCACATAGAAGCA
TGAGAATTTGGGTTCAAGCAAGTAAATTCTAAATCAGAATCCATACATAAAGTGTTTGCAATGTCC
AGTTATATCTCCATGATATTTTCTTTGTGGAAGTTGATTGTTCTTCCTTACAATAAATTGCTTGAATT
GTCTGTCTA

Full-length LINC01082 cDNA sequence (441 bp)

Source: Normal human prostate

GGTTTAGATTAGCCGTGGCCTAGGCCGTTTGACGGGGTGACACGAGCCTGCAGGGCCGAGTCCAAG
GCCCGGAGATAGGACCAACCGTCAGGAATGCGAGGAATGTTTTTCTTCGGACTCTATCGAGGCACA
CAGACAGACCATGGGGATTCTGTCTACAGTGACAGCCTTAACATTTGCCAGAGCCCTGGACGGCTG
CAGAAATGGCATTGCCCACCCTGCAAGTGAGAAGCACAGACTCGAGAAATGTAGGGAACTCGAGA
GCAGCCACTCGGCCCCAGGATCAACCCAGCACCGAAGAAAAACAACCAGAAGAAATTATTCTTCA
GCCTGAAATGAAGCCGGGATCAAATGGTTGCTGATCAGAGCCCATATTTAAATTGGAAAAGTCAA
ATTGAGCATTATTAAATAAAGCTTGTTTAATATGTCTCAAACAGAAAA

Full-length LINC00920 cDNA sequence (1567 bp)

Source: Normal human prostate

ATCTTCACAGGGAAGGAAGCAACAAAACTCTGCCTTTGGCTTTTGCTGGCTGAGCAGGGAAAGGCC
TATAGACACAGGGCATTGGGCAGGAGCTAGAACAGCCCTCCCTAGAGCACTACATAAAGCAGCCA
ATATTTTGCAAAGCATAGGGAAGAGTGAAAGTCATCCGGGGCATTTGCAGACACAGCACTAAGAA
CTTGGTGACAACAGCCCTGAAGCAAAACAGCAGCATGTACTGGGCAGGGCTTGGGAGATAAGACA
GGACATCTGAAGCTAAACATGGATCCCCTCTGAAAGCTACAATCAAAGTGTCATCCACAAAATCTT
ATCTCAAGCCTTGACTAGAGAAGGACCCACTTCCAAGACCACAGAGTTGAATGAATTCAGTCCTTG
CAGCCCGTTGAACTGAGGGCCTCCCCAACATGCTCACCTGCTTCATCAAAGCCTGCGAGAGAGAGA
GTCCACTAGCAAGAGGACATTGCAGTCTTATCTAATGCAATCACTGAAGTGACATCCTGTCACCTT
GGTCACCTTTTCTATTCTATTCATTATAAATGAGTCCCAAGTCCTGCCACACTCAAGTGGAGGGGAT
TACACAGGGCTGGAGTACCAGCGGTGGGGATAATTTGGGGTCATCTTAGAGTTCTTCCATCACACG
GGAATTCTCAGCTCACCAAATCTGGGATTCCGCATCTGGCCATTCCTTAAGCTGAAGGCCTGGCAT
ATTTTTGAGTGTCCATTTGGATCAGCTAATAAACCCAGGGTTTGTCTACTGGCTGGAGGAGTAAGA
ACTATAAGGCTAATTGAAATGAATCTACTTAAAATAGTGACCTGATTTTTCTAATAATTACTGGAA
GGTAAGGGTTGATTGAGACTTTAAAATAAAACCAAAAATTATTCTAAATTTTTCATATTTTATATAA
GAAAGTTTTGATTTTTACTGCAACCATTTTCTAACTCTTAAAATAAAGAAAGGATAATTCAAGTGTT
GATATTTTCCCAACATAAATATACAGGAACATAATGTAGCTTACTTGTATTTTATCTTTGATTTACA
CAAGAGAATTTTTATACAAATATTCCAGGCTCATTAGTTTTCCCAAAGGCTTCTCATAATCCTTTGA
TATTTAAATCATTCCCTCTTTCAAGTCATTTTTTATCTGCCTTGTCGATACTCTTTTTGTTAATTTGCC
CAACTCATCTGGATCTTCCTTTGTCACTGGCTCTGTAAATTTGAGTATTTCTCCAATAATGCTCCTGT
CAACTTTATGAAACCCTTTGTCTTTTGCAAGATTTAAAAATTCCCTTTATAATAAGCATTCTACAGT
AAGTGAAGACTCACTAGCAAATATATGAGTGATGGATCAAGAGAGACAAAGTGTTAAAAATTGAC
AGATGCAACTTTTAAAAATGAAATTCTTGGTTACCATCTAACAAAGTAAGTACAGGACCTGTATGC
TGAAAACTACAAAACACTAATGAAGGAATCAAAGAAAGTGTAAATAAATGGAGAGATATACCATG
TTCATGAATTGGAAGTCTCAGAATTATTAAAATGTAAATTCTA

9.2. Top deregulated genes upon *LINC00920* knockdown in PC-3 cells

Supplementary Table 9-2. *LINC00920*-deregulated genes (n=315) analyzed with Ingenuity Pathway Analysis (IPA)

Expr Fold Change	ID	Entrez Gene Name	Location	Type(s)
5.696	IL8	C-X-C motif chemokine ligand 8	Extracellular Space	cytokine
4.399	PDLIM7	PDZ and LIM domain 7	Cytoplasm	other
3.911	RDH10	retinol dehydrogenase 10	Nucleus	enzyme
3.699	PI3	peptidase inhibitor 3	Extracellular Space	other
3.2	CXCL1	C-X-C motif chemokine ligand 1	Extracellular Space	cytokine
3.077	GEM	GTP binding protein overexpressed in skeletal muscle	Plasma Membrane	enzyme
2.954	SOD2	superoxide dismutase 2	Cytoplasm	enzyme
2.924	DSC2	desmocollin 2	Plasma Membrane	other
2.913	DUSP1	dual specificity phosphatase 1	Nucleus	phosphatase
2.892	DHRS9	dehydrogenase/reductase 9	Cytoplasm	enzyme
2.683	CLDN11	claudin 11	Plasma Membrane	other
2.659	CFDP1	craniofacial development protein 1	Extracellular Space	other
2.625	TACSTD1	epithelial cell adhesion molecule	Plasma Membrane	other
2.445	RASD1	ras related dexamethasone induced 1	Cytoplasm	enzyme
2.315	P4HA2	prolyl 4-hydroxylase subunit alpha 2	Cytoplasm	transporter
2.256	CRABP2	cellular retinoic acid binding protein 2	Cytoplasm	transporter
2.231	FAM91A1	family with sequence similarity 91 member A1	Cytoplasm	other
2.165	ZNF385D	zinc finger protein 385D	Nucleus	other
2.109	FGFBP1	fibroblast growth factor binding protein 1	Extracellular Space	other
2.081	LAMB3	laminin subunit beta 3	Extracellular Space	transporter
2.069	PLOD2	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	Cytoplasm	enzyme
2.055	TFF2	trefoil factor 2	Extracellular Space	other
2.044	FTHL2	ferritin heavy chain 1 pseudogene 2	Other	other
2.02	BAIAP2L1	BAI1 associated protein 2 like 1	Cytoplasm	other
1.97	HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5	Cytoplasm	enzyme
1.961	FTHL16	ferritin heavy chain 1 pseudogene 16	Other	other
1.953	ID1	inhibitor of DNA binding 1, HLH protein	Nucleus	transcription regulator
1.953	OAS1	2'-5'-oligoadenylate synthetase 1	Cytoplasm	enzyme
1.948	VPS4B	vacuolar protein sorting 4 homolog B	Cytoplasm	transporter
1.91	FUT11	fucosyltransferase 11	Cytoplasm	enzyme
1.901	EGFR	epidermal growth factor receptor	Plasma Membrane	kinase
1.894	SGK	serum/glucocorticoid regulated kinase 1	Cytoplasm	kinase
1.869	PTGER2	prostaglandin E receptor 2	Plasma Membrane	G-protein coupled receptor
1.838	FTHL11	ferritin heavy chain 1 pseudogene 11	Other	other
1.826	ATP1B1	ATPase Na+/K+ transporting subunit beta 1	Plasma Membrane	transporter
1.803	AIF1L	allograft inflammatory factor 1 like	Plasma Membrane	other
1.787	PABPC4	poly(A) binding protein cytoplasmic 4	Cytoplasm	translation regulator
1.785	GSK3B	glycogen synthase kinase 3 beta	Nucleus	kinase

1.784	EIF2AK2	eukaryotic translation initiation factor 2 alpha kinase 2	Cytoplasm	kinase
1.774	MOXD1	monooxygenase DBH like 1	Cytoplasm	enzyme
1.773	AP3S1	adaptor related protein complex 3 subunit sigma 1	Cytoplasm	transporter
1.769	GNG5	G protein subunit gamma 5	Plasma Membrane	other
1.764	CYR61	cysteine rich angiogenic inducer 61	Extracellular Space	other
1.758	TMEM167A	transmembrane protein 167A	Other	other
1.754	IFIT1	interferon induced protein with tetratricopeptide repeats 1	Cytoplasm	other
1.752	FAIM3	Fc fragment of IgM receptor	Plasma Membrane	other
1.731	SH3BP4	SH3 domain binding protein 4	Cytoplasm	other
1.728	IFNGR1	interferon gamma receptor 1	Plasma Membrane	transmembrane receptor
1.728	MATN2	matrilin 2	Extracellular Space	other
1.723	TMEM27	collectrin, amino acid transport regulator	Plasma Membrane	other
1.722	IL13RA1	interleukin 13 receptor subunit alpha 1	Plasma Membrane	transmembrane receptor
1.717	FBXL20	F-box and leucine rich repeat protein 20	Cytoplasm	enzyme
1.716	RIOK3	RIO kinase 3	Cytoplasm	kinase
1.709	MAP4K5	mitogen-activated protein kinase kinase kinase kinase 5	Cytoplasm	kinase
1.705	HLTF	helicase like transcription factor	Nucleus	transcription regulator
1.7	PLCL2	phospholipase C like 2	Cytoplasm	enzyme
1.693	FTHL3	ferritin heavy chain 1 pseudogene 3	Other	other
1.681	GRB7	growth factor receptor bound protein 7	Plasma Membrane	other
1.677	IER3	immediate early response 3	Cytoplasm	other
1.671	PPPDE1	desumoylating isopeptidase 2	Cytoplasm	enzyme
1.671	ERN1	endoplasmic reticulum to nucleus signaling 1	Cytoplasm	kinase
1.667	WWP2	WW domain containing E3 ubiquitin protein ligase 2	Cytoplasm	enzyme
1.664	ELF3	E74 like ETS transcription factor 3	Nucleus	transcription regulator
1.66	STK38	serine/threonine kinase 38	Nucleus	kinase
1.646	CAV2	caveolin 2	Plasma Membrane	other
1.643	USP36	ubiquitin specific peptidase 36	Nucleus	peptidase
1.642	LIPA	lipase A, lysosomal acid type	Cytoplasm	enzyme
1.638	C1orf86	FA core complex associated protein 20	Nucleus	other
1.629	ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif 1	Extracellular Space	peptidase
1.623	GBE1	1,4-alpha-glucan branching enzyme 1	Cytoplasm	enzyme
1.618	CLGN	calmegin	Cytoplasm	peptidase
1.618	GLT25D1	collagen beta(1-O)galactosyltransferase 1	Cytoplasm	enzyme
1.608	PEG10	paternally expressed 10	Nucleus	other
1.608	TFF1	trefoil factor 1	Extracellular Space	other
1.605	CTSC	cathepsin C	Cytoplasm	peptidase
1.598	NAT13	N(alpha)-acetyltransferase 50, NatE catalytic subunit	Cytoplasm	enzyme
1.596	DOCK10	dedicator of cytokinesis 10	Cytoplasm	other
1.595	CENTG2	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1	Cytoplasm	enzyme
1.594	FAM83H	family with sequence similarity 83 member H	Other	other
1.593	RNF149	ring finger protein 149	Cytoplasm	enzyme

1.59	AXUD1	cysteine and serine rich nuclear protein 1	Nucleus	transcription regulator
1.582	PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha	Cytoplasm	kinase
1.581	L2HGDH	L-2-hydroxyglutarate dehydrogenase	Cytoplasm	enzyme
1.575	SUMO2	small ubiquitin-like modifier 2	Nucleus	enzyme
1.562	CD24	CD24 molecule	Plasma Membrane	other
1.559	IDS	iduronate 2-sulfatase	Cytoplasm	enzyme
1.548	SPRY2	sprouty RTK signaling antagonist 2	Plasma Membrane	other
1.545	TIMP2	TIMP metallopeptidase inhibitor 2	Extracellular Space	other
1.539	HEATR1	HEAT repeat containing 1	Nucleus	other
1.534	FAM102A	family with sequence similarity 102 member A	Other	other
1.531	RPS15A	ribosomal protein S15a	Cytoplasm	other
1.524	SEPN1	selenoprotein N	Cytoplasm	other
1.521	SAT1	spermidine/spermine N1-acetyltransferase 1	Cytoplasm	enzyme
1.52	SUV420H1	lysine methyltransferase 5B	Nucleus	enzyme
1.518	ACP1	acid phosphatase 1	Cytoplasm	phosphatase
1.518	EPHX1	epoxide hydrolase 1	Cytoplasm	peptidase
1.509	TUBB	tubulin beta class I	Cytoplasm	other
1.496	FTH1	ferritin heavy chain 1	Cytoplasm	enzyme
1.496	KIAA1143	KIAA1143	Other	other
1.495	GJC2	gap junction protein gamma 2	Plasma Membrane	transporter
1.495	PTK2	protein tyrosine kinase 2	Cytoplasm	kinase
1.492	GPR180	G protein-coupled receptor 180	Cytoplasm	other
1.491	CLDN1	claudin 1	Plasma Membrane	other
1.488	LAPTM4B	lysosomal protein transmembrane 4 beta	Cytoplasm	other
1.486	AVPI1	arginine vasopressin induced 1	Other	other
1.486	GADD45A	growth arrest and DNA damage inducible alpha	Nucleus	other
1.484	FKBP1A	FK506 binding protein 1A	Cytoplasm	enzyme
1.481	TMEM185A	transmembrane protein 185A	Nucleus	other
1.48	LHFP	LHFPL tetraspan subfamily member 6	Other	other
1.476	ADAM19	ADAM metallopeptidase domain 19	Plasma Membrane	peptidase
1.476	HPRT1	hypoxanthine phosphoribosyltransferase 1	Cytoplasm	enzyme
1.476	NCRNA00161	long intergenic non-protein coding RNA 161	Other	other
1.474	COL5A1	collagen type V alpha 1 chain	Extracellular Space	other
1.472	GMDS	GDP-mannose 4,6-dehydratase	Cytoplasm	enzyme
1.467	CDA	cytidine deaminase	Nucleus	enzyme
1.466	ASPH	aspartate beta-hydroxylase	Cytoplasm	enzyme
1.461	ELMOD2	ELMO domain containing 2	Other	other
1.46	TMEM87A	transmembrane protein 87A	Cytoplasm	other
1.46	VAPA	VAMP associated protein A	Plasma Membrane	other
1.452	LHPP	phospholysine phosphohistidine inorganic pyrophosphate phosphatase	Cytoplasm	phosphatase
1.445	CTSH	cathepsin H	Cytoplasm	peptidase
1.443	CTGF	connective tissue growth factor	Extracellular Space	growth factor
1.44	MAX	MYC associated factor X	Nucleus	transcription regulator
1.438	EBPL	EBP like	Cytoplasm	enzyme
1.438	ZC3HAV1	zinc finger CCCH-type containing, antiviral 1	Plasma Membrane	other
1.435	MOSC1	mitochondrial amidoxime reducing component 1	Cytoplasm	enzyme

1.43	RGMB	repulsive guidance molecule BMP co-receptor b	Plasma Membrane	other
1.428	PRPS2	phosphoribosyl pyrophosphate synthetase 2	Cytoplasm	kinase
1.425	TGM2	transglutaminase 2	Cytoplasm	enzyme
1.424	ABLIM1	actin binding LIM protein 1	Cytoplasm	other
1.415	SLC22A5	solute carrier family 22 member 5	Plasma Membrane	transporter
1.406	DENND1A	DENN domain containing 1A	Plasma Membrane	other
1.406	MYO19	myosin XIX	Cytoplasm	enzyme
1.406	NAPG	NSF attachment protein gamma	Cytoplasm	transporter
1.404	ARL14	ADP ribosylation factor like GTPase 14	Other	other
1.403	CLDND1	claudin domain containing 1	Plasma Membrane	other
1.403	DNAJB6	DnaJ heat shock protein family (Hsp40) member B6	Nucleus	transcription regulator
1.4	LTBP3	latent transforming growth factor beta binding protein 3	Extracellular Space	other
1.395	PSPC1	paraspeckle component 1	Nucleus	transcription regulator
1.394	IRF7	interferon regulatory factor 7	Nucleus	transcription regulator
1.394	C16orf13	methyltransferase like 26	Other	other
1.393	FTHL8	ferritin heavy chain 1 pseudogene 8	Other	other
1.392	C7orf42	transmembrane protein 248	Other	other
1.392	UGDH	UDP-glucose 6-dehydrogenase	Nucleus	enzyme
1.39	CCDC85B	coiled-coil domain containing 85B	Cytoplasm	other
1.39	DNER	delta/notch like EGF repeat containing	Plasma Membrane	transmembrane receptor
1.387	IL1RL1	interleukin 1 receptor like 1	Plasma Membrane	transmembrane receptor
1.387	SPINT2	serine peptidase inhibitor, Kunitz type 2	Extracellular Space	other
1.38	MID2	midline 2	Cytoplasm	other
1.377	HAS3	hyaluronan synthase 3	Plasma Membrane	enzyme
1.376	SAMD9	sterile alpha motif domain containing 9	Cytoplasm	other
1.371	DENND5B	DENN domain containing 5B	Cytoplasm	other
1.367	MOSC2	mitochondrial amidoxime reducing component 2	Cytoplasm	enzyme
1.366	HIST1H3D	histone cluster 1 H3 family member d	Nucleus	other
1.366	MAGED4B	MAGE family member D4B	Other	other
1.357	HPCAL1	hippocalcin like 1	Cytoplasm	other
1.357	IGFBP4	insulin like growth factor binding protein 4	Extracellular Space	other
1.354	ATP6V1A	ATPase H+ transporting V1 subunit A	Plasma Membrane	transporter
1.354	CCDC47	coiled-coil domain containing 47	Extracellular Space	other
1.353	ARFGEF2	ADP ribosylation factor guanine nucleotide exchange factor 2	Cytoplasm	other
1.351	SERF1B	small EDRK-rich factor 1A	Other	other
1.35	MMP23A	matrix metallopeptidase 23A (pseudogene)	Extracellular Space	other
-1.367	ARHGEF19	Rho guanine nucleotide exchange factor 19	Cytoplasm	other
-1.368	ENC1	ectodermal-neural cortex 1	Nucleus	peptidase
-1.371	MAP4K4	mitogen-activated protein kinase kinase kinase kinase 4	Cytoplasm	kinase
-1.376	CRIP2	cysteine rich protein 2	Nucleus	other
-1.377	SKP2	S-phase kinase associated protein 2	Nucleus	enzyme
-1.377	STXBP5	syntaxin binding protein 5	Plasma Membrane	other
-1.378	SLC37A4	solute carrier family 37 member 4	Cytoplasm	transporter

-1.38	SFRS6	serine and arginine rich splicing factor 6	Nucleus	other
-1.386	CCND1	cyclin D1	Nucleus	transcription regulator
-1.387	F2R	coagulation factor II thrombin receptor	Plasma Membrane	G-protein coupled receptor
-1.39	COL13A1	collagen type XIII alpha 1 chain	Plasma Membrane	other
-1.392	MIPOL1	mirror-image polydactyly 1	Nucleus	other
-1.393	PTGFR	prostaglandin F receptor	Plasma Membrane	G-protein coupled receptor
-1.396	TMEM45B	transmembrane protein 45B	Extracellular Space	other
-1.405	COL18A1	collagen type XVIII alpha 1 chain	Extracellular Space	other
-1.407	RXRB	retinoid X receptor beta	Nucleus	ligand-dependent nuclear receptor
-1.408	ROBO3	roundabout guidance receptor 3	Plasma Membrane	transmembrane receptor
-1.409	LEPROTL1	leptin receptor overlapping transcript like 1	Plasma Membrane	other
-1.41	PDDC1	glutamine amidotransferase like class 1 domain containing 1	Cytoplasm	other
-1.41	KCTD5	potassium channel tetramerization domain containing 5	Cytoplasm	other
-1.411	RBBP9	RB binding protein 9, serine hydrolase	Nucleus	other
-1.411	ZNF837	zinc finger protein 837	Other	other
-1.413	NIPA1	NIPA magnesium transporter 1	Plasma Membrane	transporter
-1.421	TM7SF2	transmembrane 7 superfamily member 2	Cytoplasm	enzyme
-1.422	TMEM30A	transmembrane protein 30A	Cytoplasm	transporter
-1.423	NSMCE4A	NSE4 homolog A, SMC5-SMC6 complex component	Nucleus	other
-1.425	ACTN4	actinin alpha 4	Cytoplasm	transcription regulator
-1.43	POLR3G	RNA polymerase III subunit G	Nucleus	enzyme
-1.435	ADK	adenosine kinase	Nucleus	kinase
-1.437	PFTK1	cyclin dependent kinase 14	Nucleus	kinase
-1.438	MAPK3	mitogen-activated protein kinase 3	Cytoplasm	kinase
-1.444	SDSL	serine dehydratase like	Cytoplasm	enzyme
-1.445	ARSB	arylsulfatase B	Cytoplasm	enzyme
-1.449	ACCS	1-aminocyclopropane-1-carboxylate synthase homolog (inactive)	Other	enzyme
-1.449	RPL23A	ribosomal protein L23a	Cytoplasm	other
-1.45	CECR7	cat eye syndrome chromosome region, candidate 7	Other	other
-1.457	ABCA13	ATP binding cassette subfamily A member 13	Extracellular Space	transporter
-1.457	C20orf177	family with sequence similarity 217 member B	Other	other
-1.462	NDRG3	NDRG family member 3	Cytoplasm	other
-1.467	TRPM4	transient receptor potential cation channel subfamily M member 4	Plasma Membrane	ion channel
-1.472	EIF2S2	eukaryotic translation initiation factor 2 subunit beta	Cytoplasm	translation regulator
-1.474	CTDSP2	CTD small phosphatase 2	Nucleus	phosphatase
-1.474	HIBADH	3-hydroxyisobutyrate dehydrogenase	Cytoplasm	enzyme
-1.477	NUDT1	nudix hydrolase 1	Extracellular Space	phosphatase
-1.478	ZFYVE20	rabenosyn, RAB effector	Cytoplasm	other
-1.479	NPEPL1	aminopeptidase like 1	Nucleus	peptidase
-1.487	APOBEC3G	apolipoprotein B mRNA editing enzyme catalytic subunit 3G	Nucleus	enzyme
-1.487	COPS7B	COP9 signalosome subunit 7B	Cytoplasm	other
-1.49	C20orf56	long intergenic non-protein coding RNA 261	Other	other

-1.49	NBPF20	NBPF member 11	Other	other
-1.491	ARL5A	ADP ribosylation factor like GTPase 5A	Other	enzyme
-1.494	ANKFY1	ankyrin repeat and FYVE domain containing 1	Cytoplasm	transcription regulator
-1.494	HCP5	HLA complex P5	Other	other
-1.498	PHF14	PHD finger protein 14	Nucleus	other
-1.508	CDC2L6	cyclin dependent kinase 19	Nucleus	kinase
-1.508	CXADR	CXADR, Ig-like cell adhesion molecule	Plasma Membrane	transmembrane receptor
-1.511	PNRC2	proline rich nuclear receptor coactivator 2	Nucleus	other
-1.518	TMX1	thioredoxin related transmembrane protein 1	Cytoplasm	enzyme
-1.525	SDC1	syndecan 1	Plasma Membrane	enzyme
-1.529	TCEA2	transcription elongation factor A2	Nucleus	transcription regulator
-1.53	WDR42A	DDB1 and CUL4 associated factor 8	Nucleus	other
-1.531	CAST	calpastatin	Cytoplasm	peptidase
-1.532	SKP1	S-phase kinase associated protein 1	Nucleus	transcription regulator
-1.571	ETV5	ETS variant 5	Nucleus	transcription regulator
-1.572	ITPRIPL2	ITPRIP like 2	Other	other
-1.575	IL17RC	interleukin 17 receptor C	Plasma Membrane	transmembrane receptor
-1.576	GPR4	G protein-coupled receptor 4	Plasma Membrane	G-protein coupled receptor
-1.582	SCD	stearoyl-CoA desaturase	Cytoplasm	enzyme
-1.586	MUC1	mucin 1, cell surface associated	Plasma Membrane	other
-1.587	KIAA1641	ankyrin repeat domain 36B	Extracellular Space	other
-1.588	SLC35F2	solute carrier family 35 member F2	Other	other
-1.59	SMARCD1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1	Nucleus	transcription regulator
-1.593	PLCXD3	phosphatidylinositol specific phospholipase C X domain containing 3	Plasma Membrane	other
-1.598	CALM3	calmodulin 1	Cytoplasm	other
-1.599	LANCL1	LanC like 1	Plasma Membrane	other
-1.611	ICK	intestinal cell kinase	Cytoplasm	kinase
-1.612	KCNT2	potassium sodium-activated channel subfamily T member 2	Plasma Membrane	ion channel
-1.628	ACLY	ATP citrate lyase	Cytoplasm	enzyme
-1.628	GSTZ1	glutathione S-transferase zeta 1	Cytoplasm	enzyme
-1.633	FAR1	fatty acyl-CoA reductase 1	Cytoplasm	enzyme
-1.637	HNRNPUL2	heterogeneous nuclear ribonucleoprotein U like 2	Nucleus	other
-1.64	OSBP	oxysterol binding protein	Cytoplasm	transporter
-1.648	IFRD2	interferon related developmental regulator 2	Nucleus	other
-1.657	AES	amino-terminal enhancer of split	Nucleus	transcription regulator
-1.661	CNPY4	canopy FGF signaling regulator 4	Other	other
-1.661	RNASEH1	ribonuclease H1	Nucleus	enzyme
-1.664	ATG5	autophagy related 5	Cytoplasm	other
-1.664	CYB561D1	cytochrome b561 family member D1	Other	other
-1.669	HMGCL	3-hydroxy-3-methylglutaryl-CoA lyase	Cytoplasm	enzyme
-1.669	RYBP	RING1 and YY1 binding protein	Nucleus	transcription regulator
-1.673	PDCD6IP	programmed cell death 6 interacting protein	Cytoplasm	other
-1.679	SOCS2	suppressor of cytokine signaling 2	Cytoplasm	other
-1.681	DDX17	DEAD-box helicase 17	Nucleus	enzyme

-1.685	UBE2Z	ubiquitin conjugating enzyme E2 Z	Nucleus	enzyme
-1.695	ARL6IP5	ADP ribosylation factor like GTPase 6 interacting protein 5	Cytoplasm	other
-1.707	C7orf55	formation of mitochondrial complex V assembly factor 1 homolog	Cytoplasm	other
-1.708	GBA2	glucosylceramidase beta 2	Cytoplasm	enzyme
-1.709	HINT2	histidine triad nucleotide binding protein 2	Cytoplasm	other
-1.71	CTTN	cortactin	Plasma Membrane	other
-1.714	ZNF827	zinc finger protein 827	Other	other
-1.716	PHB2	prohibitin 2	Cytoplasm	transcription regulator
-1.723	MRPL52	mitochondrial ribosomal protein L52	Cytoplasm	other
-1.735	SERF2	small EDRK-rich factor 2	Other	other
-1.739	TRIB3	tribbles pseudokinase 3	Nucleus	kinase
-1.743	CBX6	chromobox 6	Nucleus	other
-1.756	SLC35B4	solute carrier family 35 member B4	Cytoplasm	transporter
-1.758	CDC25B	cell division cycle 25B	Nucleus	phosphatase
-1.772	HOXB13	homeobox B13	Nucleus	transcription regulator
-1.774	PTP4A2	protein tyrosine phosphatase type IVA, member 2	Cytoplasm	phosphatase
-1.79	BAG3	BCL2 associated athanogene 3	Cytoplasm	other
-1.791	TCTEX1D2	Tctex1 domain containing 2	Cytoplasm	other
-1.807	BMPR2	bone morphogenetic protein receptor type 2	Plasma Membrane	kinase
-1.812	PHACTR3	phosphatase and actin regulator 3	Nucleus	other
-1.814	RBM3	RNA binding motif protein 3	Cytoplasm	other
-1.819	ACP6	acid phosphatase 6, lysophosphatidic	Cytoplasm	phosphatase
-1.82	C19orf60	required for excision 1-B domain containing	Other	other
-1.834	DYNC1LI2	dynein cytoplasmic 1 light intermediate chain 2	Cytoplasm	other
-1.875	ADM2	adrenomedullin 2	Extracellular Space	other
-1.887	ABCC3	ATP binding cassette subfamily C member 3	Plasma Membrane	transporter
-1.895	TAGLN3	transgelin 3	Extracellular Space	other
-1.899	PGD	phosphogluconate dehydrogenase	Cytoplasm	enzyme
-1.947	ENSA	endosulfine alpha	Cytoplasm	transporter
-1.95	USP4	ubiquitin specific peptidase 4	Nucleus	peptidase
-1.955	RAB8B	RAB8B, member RAS oncogene family	Cytoplasm	enzyme
-1.957	HSBP1	heat shock factor binding protein 1	Nucleus	other
-1.961	WARS	tryptophanyl-tRNA synthetase	Cytoplasm	enzyme
-1.986	EPRS	glutamyl-prolyl-tRNA synthetase	Cytoplasm	enzyme
-1.987	FBLN1	fibulin 1	Extracellular Space	other
-1.998	PRKCA	protein kinase C alpha	Cytoplasm	kinase
-2.01	KIAA1310	KAT8 regulatory NSL complex subunit 3	Nucleus	other
-2.023	RAG1AP1	solute carrier family 50 member 1	Plasma Membrane	transporter
-2.076	VRK3	vaccinia related kinase 3	Nucleus	kinase
-2.128	HAPLN1	hyaluronan and proteoglycan link protein 1	Extracellular Space	other
-2.149	KLC1	kinesin light chain 1	Cytoplasm	other
-2.151	ZNF664	zinc finger protein 664	Nucleus	transcription regulator
-2.188	RPS23	ribosomal protein S23	Cytoplasm	translation regulator
-2.233	IDH1	isocitrate dehydrogenase (NADP(+)) 1, cytosolic	Cytoplasm	enzyme
-2.253	NFATC2IP	nuclear factor of activated T cells 2 interacting protein	Nucleus	other
-2.264	UBE3C	ubiquitin protein ligase E3C	Nucleus	enzyme

-2.276	FIBP	FGF1 intracellular binding protein	Nucleus	other
-2.282	MED20	mediator complex subunit 20	Nucleus	transcription regulator
-2.362	ASNS	asparagine synthetase (glutamine-hydrolyzing)	Cytoplasm	enzyme
-2.413	FZD4	frizzled class receptor 4	Plasma Membrane	G-protein coupled receptor
-2.428	TTC19	tetratricopeptide repeat domain 19	Cytoplasm	other
-2.492	PHGDH	phosphoglycerate dehydrogenase	Cytoplasm	enzyme
-2.503	GPX8	glutathione peroxidase 8 (putative)	Cytoplasm	enzyme
-2.509	URM1	ubiquitin related modifier 1	Cytoplasm	other
-2.573	ZNF598	zinc finger protein 598	Extracellular Space	enzyme
-2.706	CCNY	cyclin Y	Nucleus	other
-3.158	CD59	CD59 molecule (CD59 blood group)	Plasma Membrane	other
-3.249	PPP1CB	protein phosphatase 1 catalytic subunit beta	Cytoplasm	phosphatase
-3.662	FAM62B	extended synaptotagmin 2	Plasma Membrane	other
-4.242	TSPO	translocator protein	Cytoplasm	transmembrane receptor

9.3. Protein interaction partners of *LINC00920* as identified by ChIRP-MS

Supplementary Table 9-3. Proteins interacting with *LINC00920* as identified by ChIRP-MS*

Replicate AA0074		Replicate AA2864		Replicate AA0097	
Protein	**Enrichment fold-change (*LINC00920* vs. lacZ)**	**Protein**	**Enrichment fold-change (*LINC00920* vs. lacZ)**	**Protein**	**Enrichment fold-change (*LINC00920* vs. lacZ)**
CLMN	54.256	CLMN	132.02	PURA	97
HNRNPH1	**8.4646**	**HNRNPA2B1**	**8.5829**	RBMX; RBMXL1; RBMXL2	2.70
HNRNPF	**8.1897**	**HNRNPH1**	**7.7861**	HNRNPK	2.50
HNRNPH3	7.7915	**HNRNPA1; HNRNPA1L2**	**7.3625**	**HNRNPA2B1**	**2.49**
HNRNPM	6.4508	**HNRNPA3**	**6.4529**	**HNRNPH1**	**2.41**
HNRNPA3	**6.3449**	**HNRNPF**	**5.1694**	YBX1	2.35
HNRNPA2B1	**6.3104**	RAB1B;RAB1A; RAB1C	4.8331	**HNRNPF**	**2.35**
HNRNPL	6.3095	**EEF1G**	**3.8563**	CALR	2.24
HNRNPK	6.1481	TPM3	3.6192	HNRNPH3	2.21
HNRNPA1; HNRNPA1L2	**6.0889**	NME1-NME2; NME2; NME1; NME2P1	3.6111	APOBEC3C; APOBEC3D; bK150C2.9; APOBEC3F	2.21
SFPQ	5.993	**PDIA3**	**3.4972**	**HNRNPA1; HNRNPA1L2**	**2.16**
HNRNPU	5.9732	PRKCSH	3.4643	RPLP2	2.12
HNRNPC	5.791	PRDX1	3.31	HNRNPU	2.08
FUS	5.6598	CLTC;CLTCL1	3.2165	**HNRNPA3**	**2.07**
HNRNPDL	5.5687	PRDX3	3.1902	HNRNPD	1.90
GRSF1	5.4929	**CLIC1**	**3.1075**	ELAVL1	1.89
NONO	5.4928	PFN1	3.0106	HNRNPC	1.89
RBMX; RBMXL1	5.1306	**YWHAZ**	**2.9884**	HSPA5	1.86
DDX5; DDX17	4.9802	SLC25A5; SLC25A4; SLC25A6	2.978	HNRNPAB	1.86
HNRNPR; SYNCRIP	4.959	ANXA7	2.9746	**HSP90B1**	**1.71**
HNRNPAB	4.6391	HSPA8; HSPA2	2.9432	DDX39A	1.71
HNRNPD	4.1919	PPIA	2.9307	PPA1	1.67
KRT6B	4.0641	VDAC2	2.9271	SNRPD3	1.65
KHSRP	3.8019	LDHA	2.9054	ERP29	1.65
SRSF1	3.5474	TXNRD1	2.8951	**NPM1**	**1.64**
YBX1;YBX3	3.008	ACTG1; ACTB	2.8943	HNRNPM	1.59
ILF3	2.7134	**LDHB**	**2.8848**	**PDIA3**	**1.57**
PCBP1	2.6766	PRDX2	2.8059	KHSRP	1.57
RPS3	2.6245	GOT2	2.8055	DDX5	1.55
EIF4A1; EIF4A2	2.5963	PHGDH	2.8049	TPI1	1.52
PCBP2; PCBP3	2.5653	ATP5B	2.8049	**HSP90AB1**	**1.51**
DDX39A; DDX39B	2.26	**P4HB**	**2.7907**	**CFL1**	**1.45**

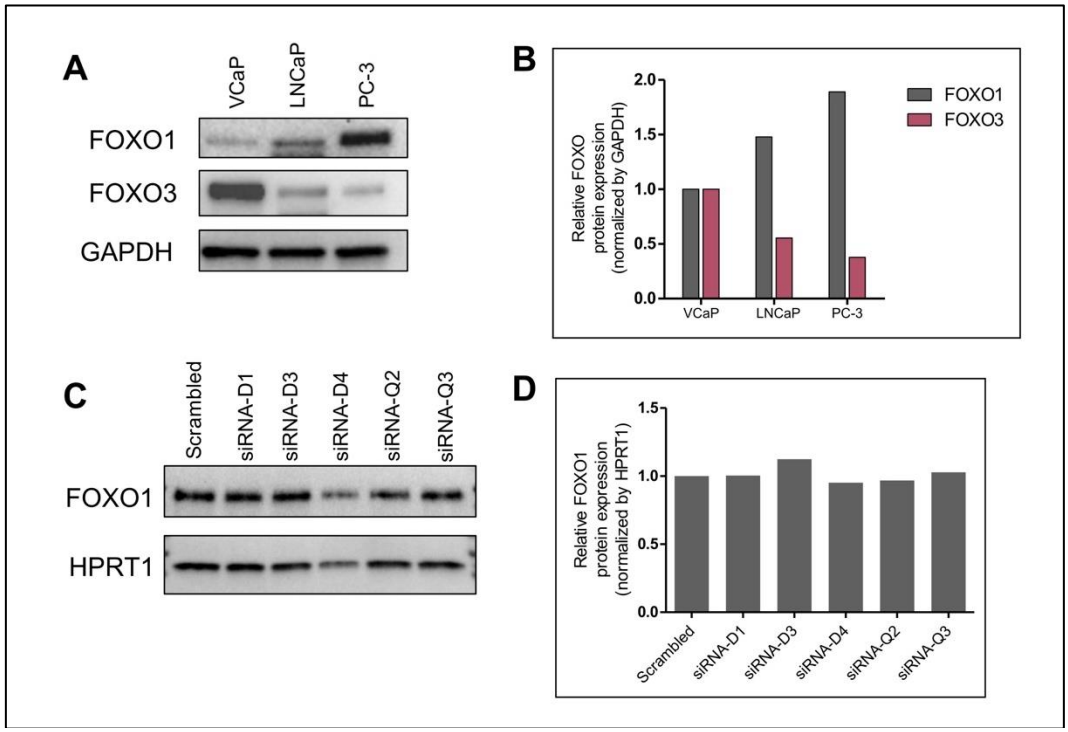
MCCC1	2.1346	PIP	2.7742	HSPA8	1.45
PABPC1	2.1112	PKM	2.76	PEBP1	1.42
PCCA	2.0685	**HSP90AB1**	**2.7244**	**HSP90AA1**	**1.40**
PC	2.0549	**CFL1**	**2.7101**	CCT3	1.40
TUFM	1.9764	ALDOA	2.7009	**YWHAZ; SFN; YWHAE; YWHAB; YWHAQ; YWHAH; YWHAG**	**1.40**
HSPA9	1.9733	LZIC	2.6976	ACTN1	1.37
ALDH1A3; ALDH1A2	1.9722	ARHGDIA	2.6818	ARHGDIA	1.37
PRMT1	1.9707	**TKT**	**2.6757**	MYDGF	1.36
HSPD1	1.9293	**FSCN1**	**2.6534**	GSTP1	1.34
ACACA	1.914	PRDX5	2.6518	**FLNA**	**1.34**
PCCB	1.9079	HSPA5	2.6481	ATP5F1	1.34
CS	1.8781	MDH2	2.6161	PABPC1	1.33
MCCC2	1.8265	EEF1A1P5; EEF1A1; EEF1A2	2.612	NME2;NME1-NME2;NME1; NME2P1	1.33
NCL	1.8248	ANXA2; ANXA2P2	2.5998	EIF4A1; EIF4A2	1.32
PGK1	1.7991	TAGLN2	2.5738	CCT8	1.32
FH	1.7838	TRAP1	2.5682	**EEF1G**	**1.31**
HSP90B1	**1.7819**	**EEF1D**	**2.5573**	PPIB	1.31
ACTN1; ACTN4; ACTN3	1.7777	VDAC1	2.4503	**P4HB**	**1.30**
TRAP1	1.7722	**HSP90AA1**	**2.3703**	TALDO1	1.30
TUBB4A; TUBB4B; TUBB8	1.7394	TUFM	2.3645	PCBP1	1.30
EEF2;EFTUD2	1.7279	HSPE1; HSPE1-MOB4	2.35	PAICS	1.30
PPIA	1.7205	VCL	2.3464	**TKT**	**1.29**
ENO1	1.6963	PC	2.3431	**CLIC1**	**1.28**
MSN; RDX; EZR	1.6942	HSPA9	2.3428	HIST1H2AJ; HIST1H2AH; H2AFJ; HIST2H2AC; HIST2H2AA3; HIST1H2AD; HIST1H2AG; H2AFV; H2AFZ; HIST1H2AC; HIST3H2A; HIST1H2AB; HIST1H2AA; H2AFX	1.25
LDHA	1.6931	CS	2.3132	TAGLN2	1.25
PDIA3	**1.6926**	**HSP90B1**	**2.3132**	CALM2; CALM3; CALM1	1.25
FLNA	**1.6869**	PCCB	2.2846	RPLP0; RPLP0P6	1.25
HSP90AB1	**1.6708**	MCCC1	2.2676	RAN	1.25

RAN	1.6384	PGK1	2.2278	ACTG1	1.24
TUBB;TUBB3; TUBB2B; TUBB2A	1.6338	TUBB; TUBB2B; TUBB2A	2.2164	**EEF1D**	**1.23**
HSP90AA1	**1.6267**	ENO1	2.2084	**FSCN1**	**1.23**
HSPA8	1.6203	PCCA	2.1811	PRDX2	1.23
CALM2; CALM3; CALM1	1.6025	MCCC2	2.181	S100A7; S100A7A	1.22
DSTN	1.5932	HSPD1	2.1802	**LDHB**	**1.22**
EEF1G	**1.5884**	TUBB3	2.1721	**PLEC**	**1.22**
FSCN1	**1.576**	FH	2.1462	XRCC6	1.21
ANXA7	1.5655	TUBA1B; TUBA1C; TUBA1A; TUBA4A; TUBA3C; TUBA8; TUBA3E	2.1278	ANXA5	1.20
RPSA	1.5472	CALR	2.0788		
VIM	1.5124	**NPM1**	**2.0308**		
MDH2	1.5084	**PLEC**	**1.9837**		
EEF1A1P5; EEF1A1; EEF1A2	1.4991	GAPDH	1.9614		
PLEC	**1.494**	ACACA	1.9443		
TUBA1B; TUBA1C; TUBA1A; TUBA3C; TUBA4A; TUBA3E	1.4835	LCN1;LCN1P1	1.9188		
FLNB	1.474	**YWHAE**	**1.8876**		
ACTG1	1.4729	ACTN1; ACTN4	1.838		
PRDX6	1.4636	**FLNA**	**1.8187**		
P4HB	**1.4508**	TIMM44	1.6966		
ATP5B	1.4363	FLT1	1.6598		
LMNA	1.4301	IGHG1	1.6398		
RPS20	1.4183	FABP5	1.6301		
NME1;NME2; NME1-NME2	1.4033	IARS2	1.5835		
PKM	1.3952	ALDH7A1	1.5545		
ATP5A1	1.3892	EEF2	1.3601		
CLIC1	**1.3881**	UBB;RPS27A; UBC;UBA52	1.2394		
PFN1	1.3663				
VCL	1.3548				
SLC25A5	1.3258				
CFL1	**1.3214**				
GSTP1	1.3181				
CLTC	1.3164				
HSPE1	1.3123				
ALDOA	1.2998				
TKT	**1.2964**				

PEBP1	1.2784		
LDHB	**1.2746**		
NPM1	**1.2577**		
AHCY	1.2531		
PRDX3	1.2525		
EEF1D	**1.2356**		
RPS12	1.2338		
YWHAZ	**1.2234**		
ANXA2; ANXA2P2	1.2142		
RAB1B; RAB8B; RAB1A; RAB10; RAB1C; RAB13; RAB8A; RAB15	1.2111		

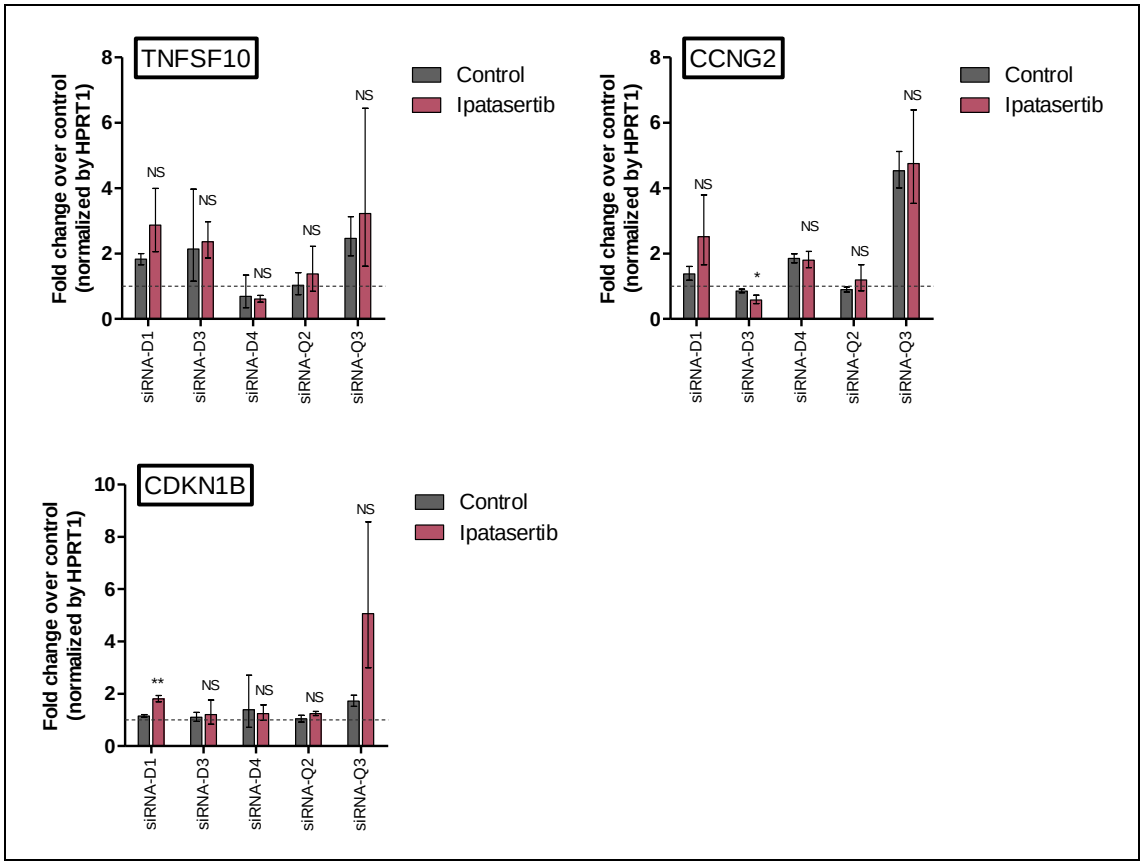
***** With enrichment fold-change values greater than 1.2. Proteins in **bold** are common to all replicates.

9.4. FOXO1 is the predominant FOXO isoform in PC-3 cells and remains unchanged upon *LINC00920* knockdown



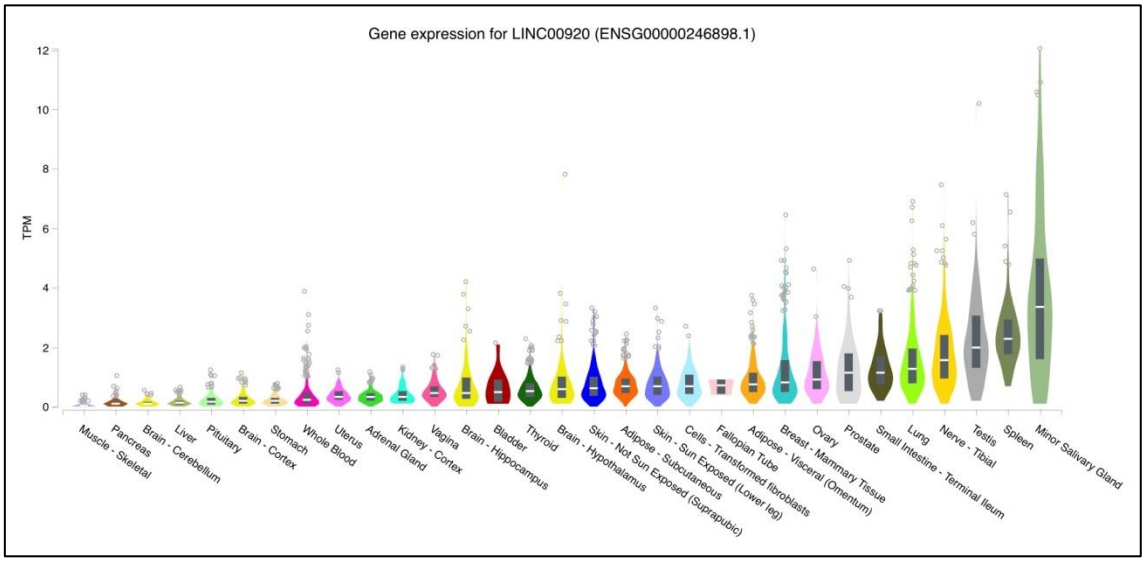
Supplementary Figure 9-1. FOXO1 and FOXO3 protein expression analysis in prostate cancer cell lines. (A) Representative immunoblots of FOXO1 and FOXO3 showing their relative expression in VCaP, LNCaP, and PC-3 cells. (B) Normalized quantification of protein band intensities in *A*. (C) Knockdown of *LINC00920* does not affect FOXO1 expression in PC-3 cells. (D) Normalized quantification of protein band intensities in *C*.

9.5. *LINC00920*-insensitive FOXO targets



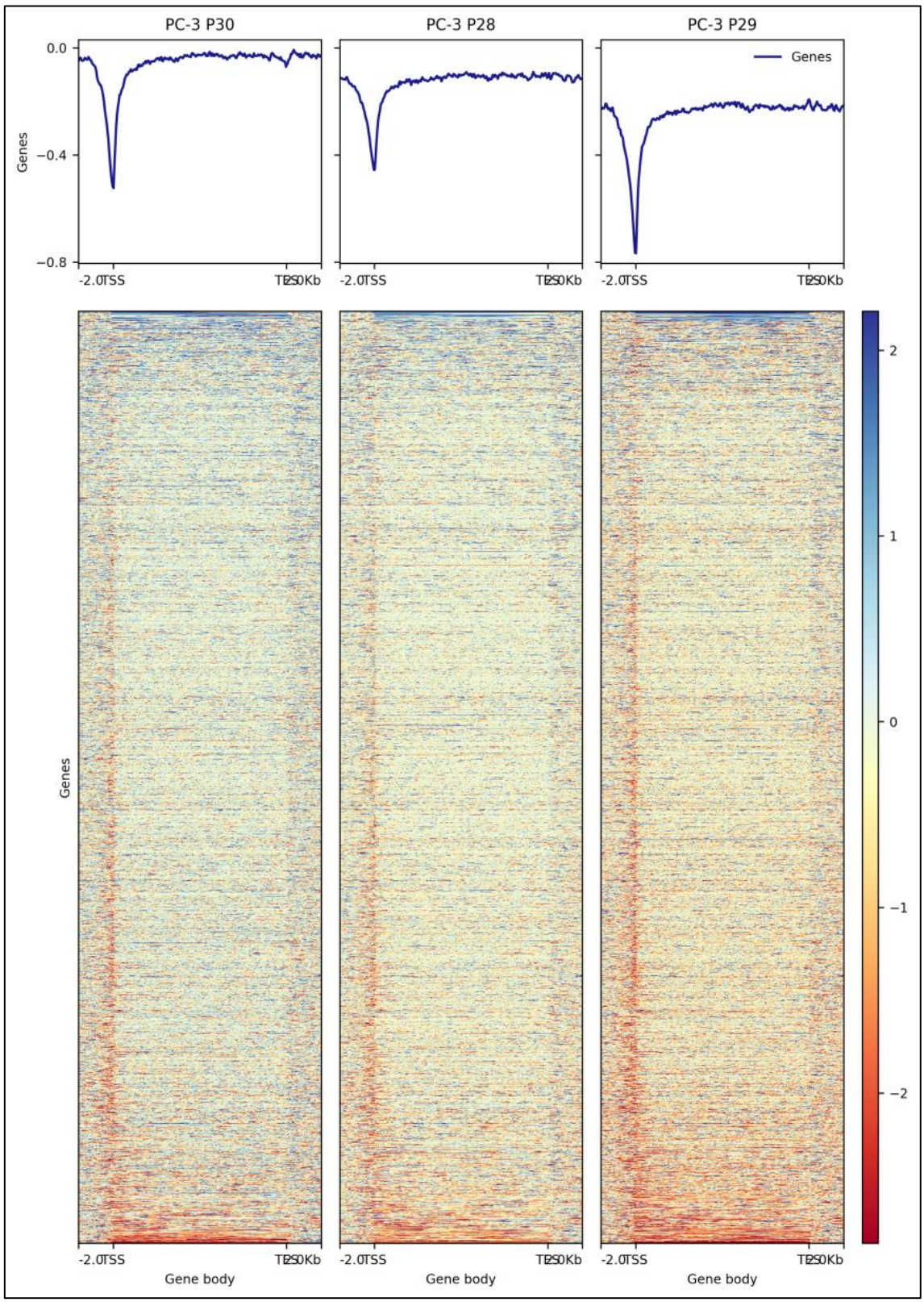
Supplementary Figure 9-2. FOXO targets indifferent to combined *LINC00920* knockdown and AKT inhibition.

9.6. Tissue-specific expression of *LINC00920*

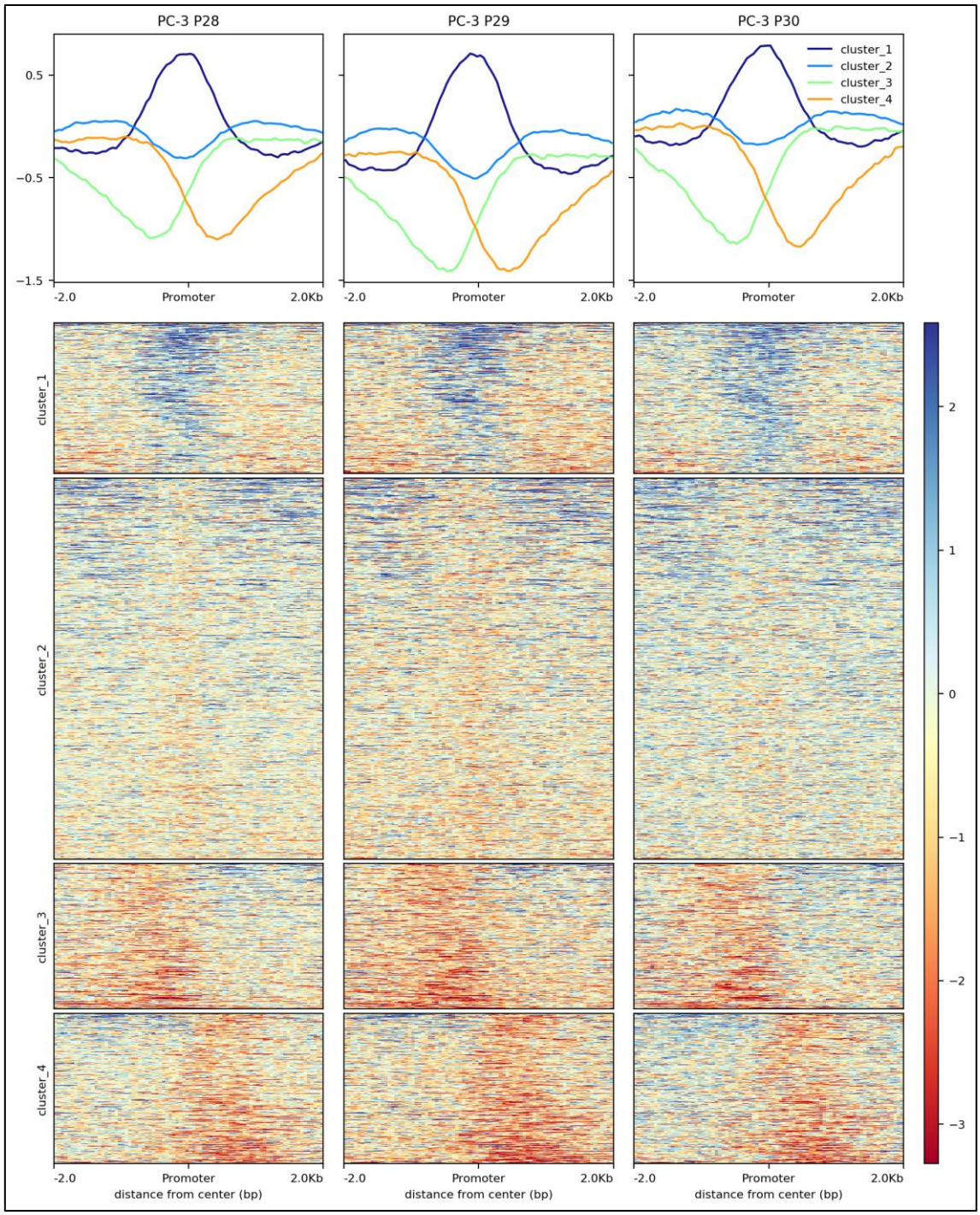


Supplementary Figure 9-3. RNA-seq quantitation of *LINC00920* expression across human tissues [319]. TPM: Transcripts Per Kilobase Million.

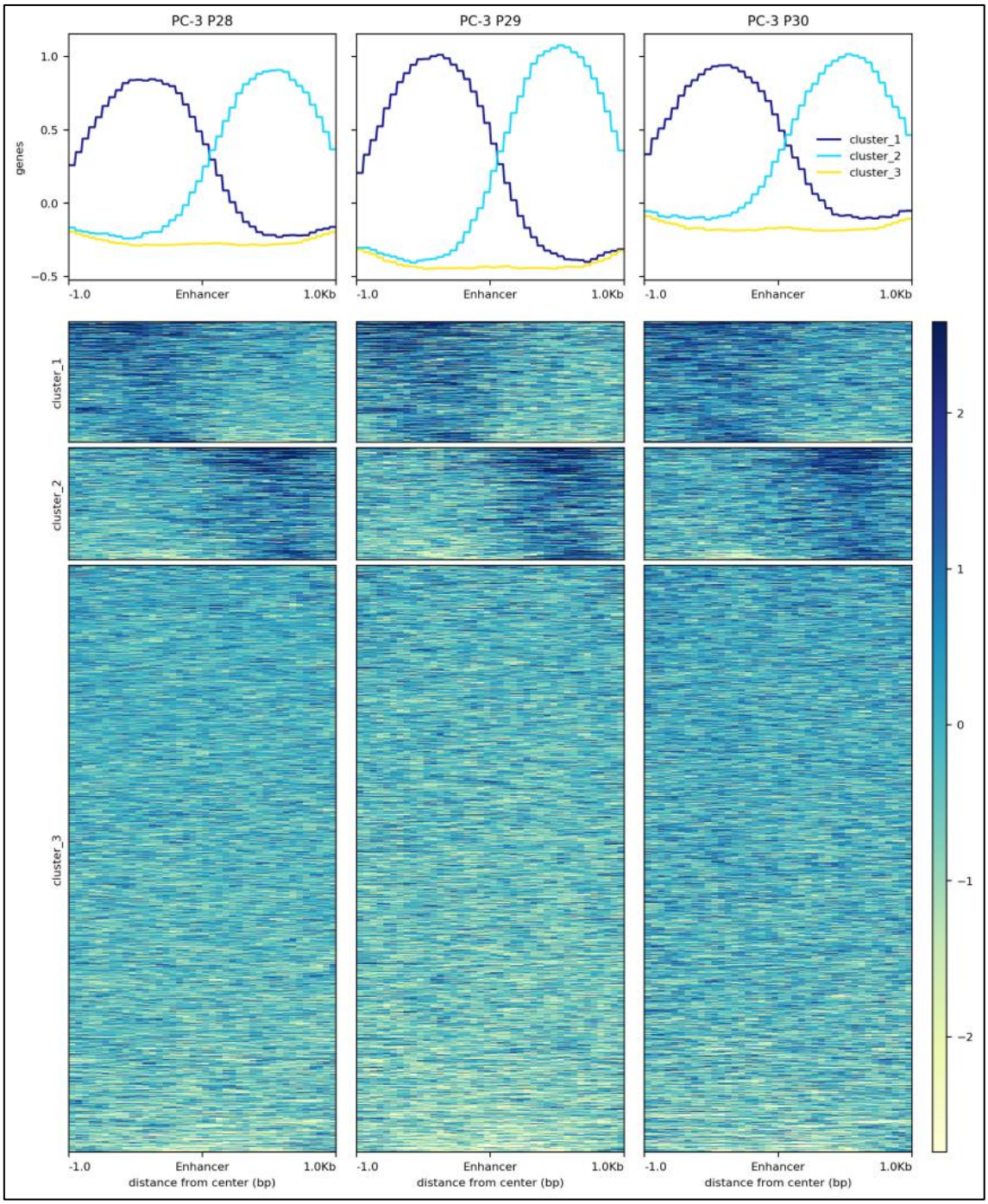
9.7. Replicates of ChIRP-seq coverage alignment



Supplementary Figure 9-4. Normalized *LINC00920* read coverage across genic regions in the human genome build hg19, shown in triplicate. Top: Average profiles of *LINC00920* occupancy on genes normalized by length of 10 kbp with 2-kb extensions upstream the TSSs and downstream the TESs. Bottom: Heatmaps showing signals for individual genes (heatmap rows). Shown data was derived from a representative replicate. Blue: high read coverage, red: low read coverage.

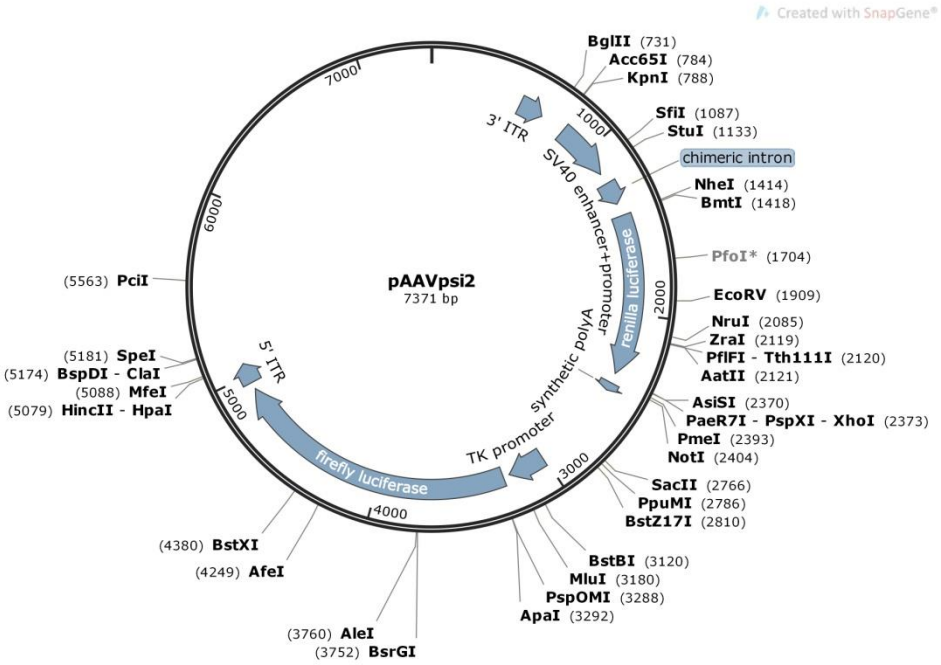


Supplementary Figure 9-5. ChIRP-seq density clustering reveals *LINC00920* binding to a subset of promoters (in triplicate). Top: Average profiles of *LINC00920* occupancy across clustered promoters. Bottom: Segmented heatmaps showing clustered promoters based on signal density. Promoter regions in cluster 1 (n=5,287) exhibit positive *LINC00920* enrichment in all replicates. Cluster 2 (n=20,369) promoters show modest *LINC00920* occupancy both upstream and downstream the promoter center. Clusters 3 (n=5,549) and 4 (n=4,975) promoters exhibit depletion of *LINC00920* signal upstream and downstream the promoter center, respectively. Blue: high read coverage, red: low read coverage.

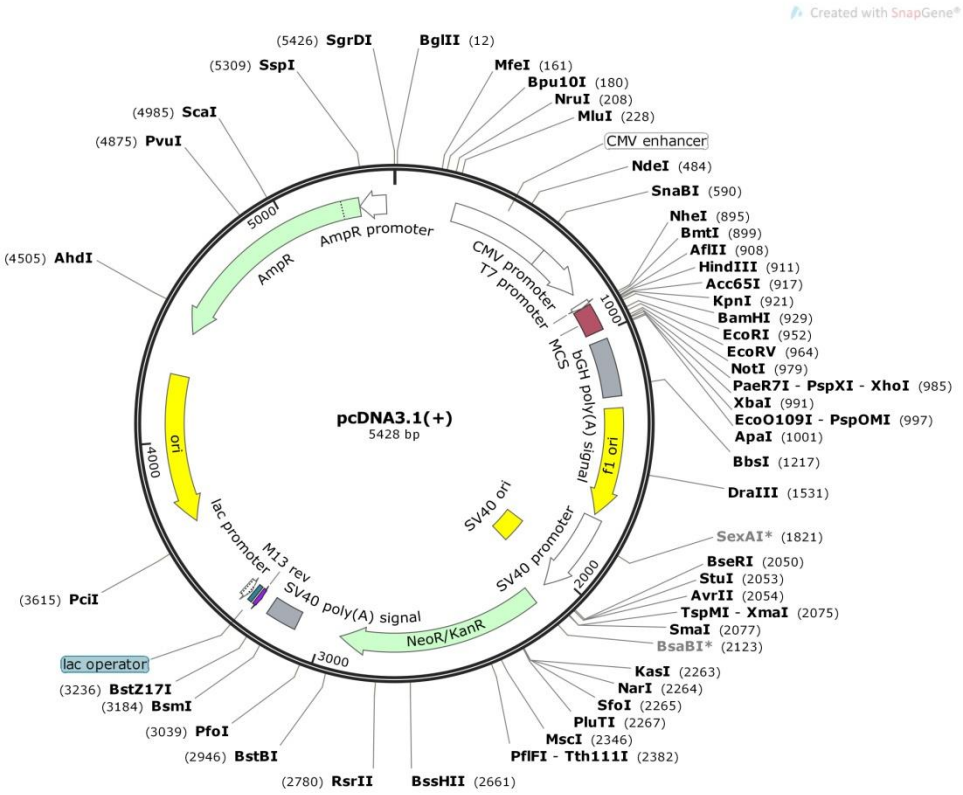


Supplementary Figure 9-6. Normalized *LINC00920* read coverage across annotated enhancer regions in the PC-3 genome, shown in triplicate. Top: Average profiles of *LINC00920* occupancy on enhancers 1-kb extensions upstream and downstream the annotated center. Bottom: Segmented heatmaps showing clustered enhancer regions based on signal intensity. Cluster 1 (n=10,018) enhancer regions show enrichment of *LINC00920* occupancy upstream the annotated center. Enhancer regions in cluster 2 (n=10,176) show enrichment downstream of the center. The remaining cluster 3 (n=50,302) enhancer regions do not show *LINC00920* enrichment. Dark blue: high read coverage, yellow: low read coverage.

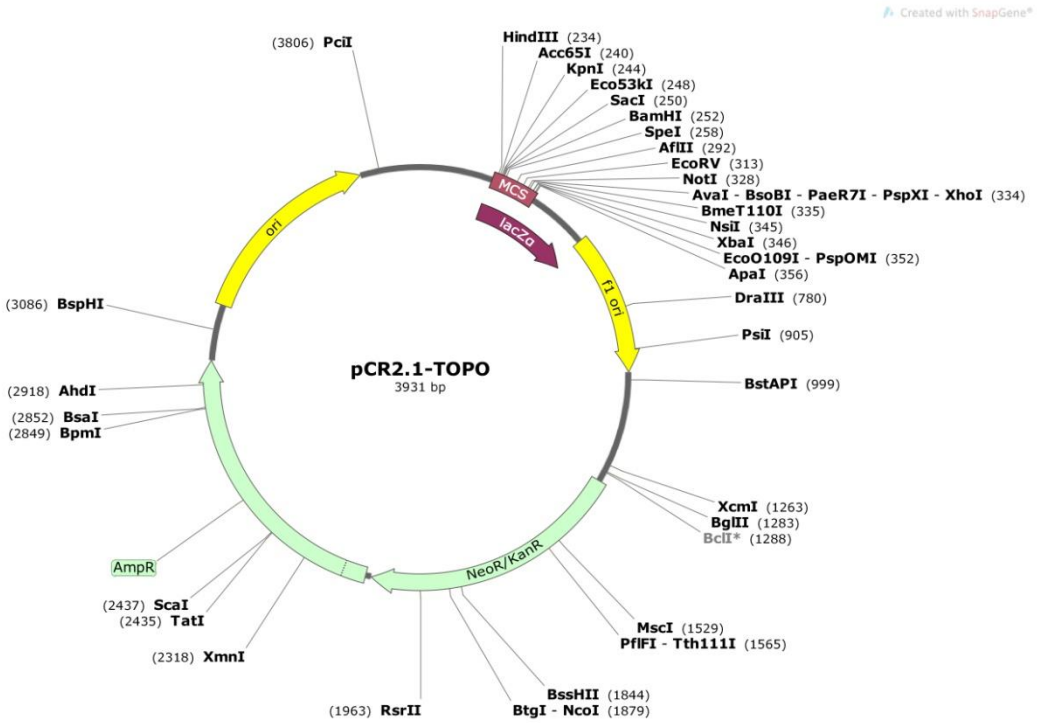
9.8. Vector maps



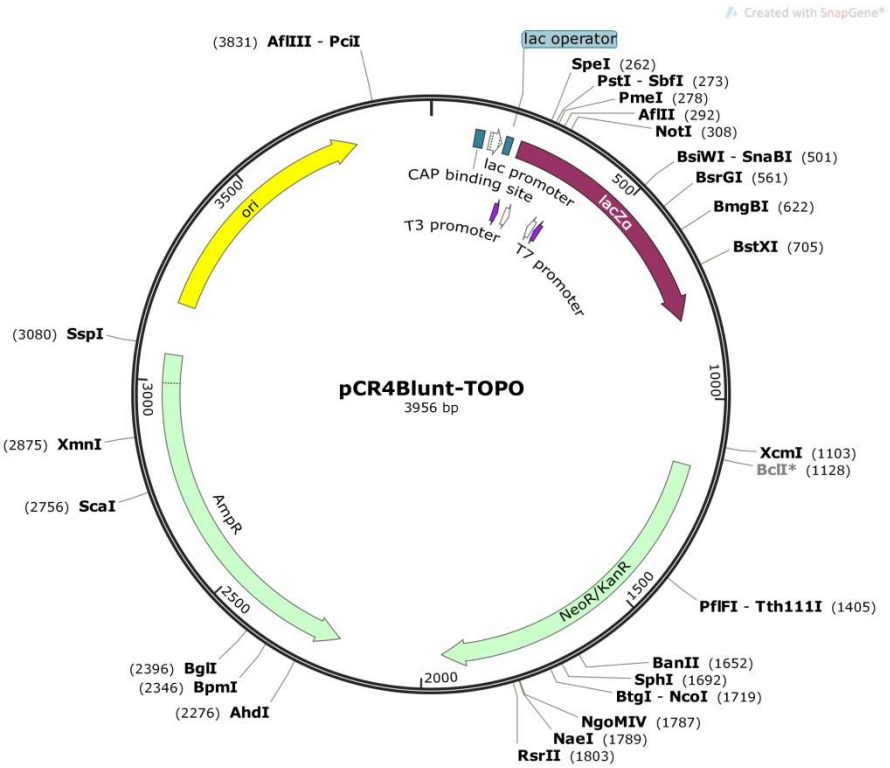
Supplementary Figure 9-7. Vector map of pAAVpsi2.



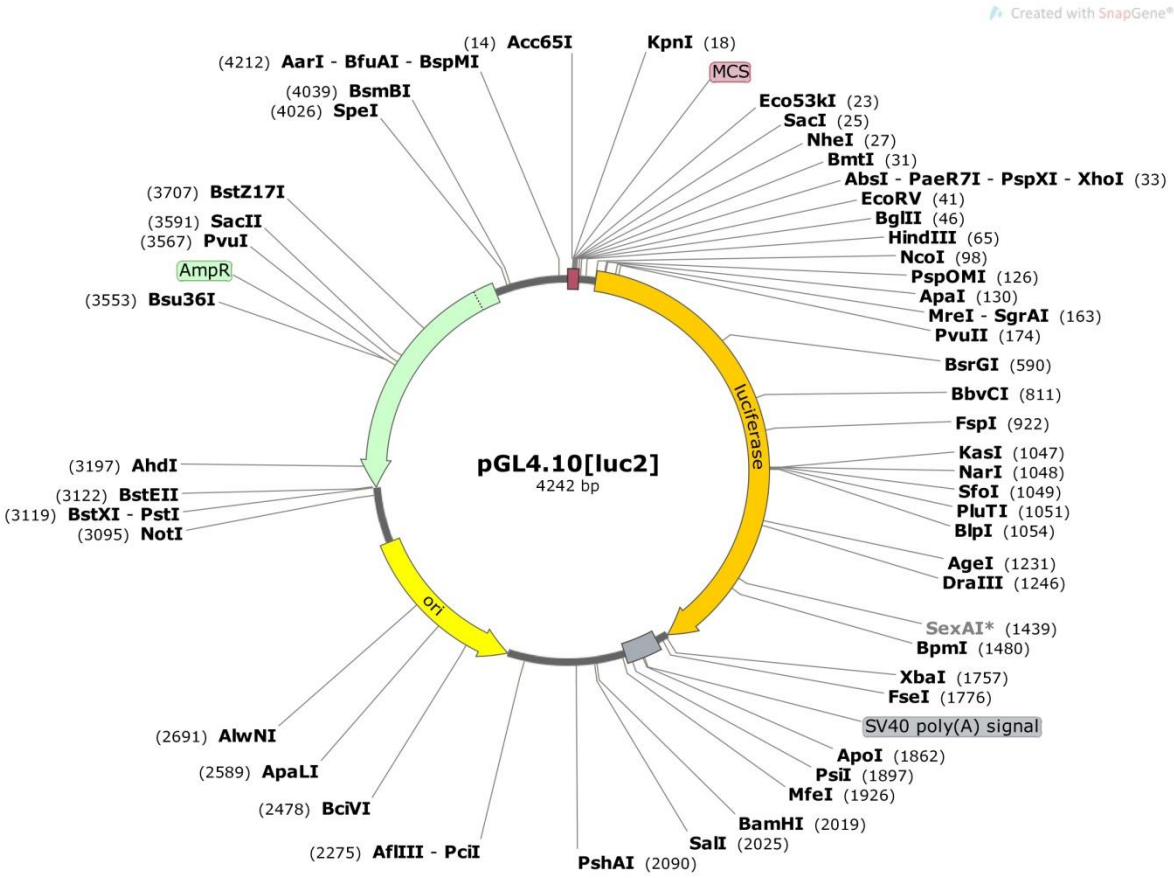
Supplementary Figure 9-8. Vector map of pcDNA^{TM}3.1(+).



Supplementary Figure 9-9. Vector map of pCR®2.1-TOPO®.



Supplementary Figure 9-10. Vector map of pCR®4Blunt-TOPO®.



Supplementary Figure 9-11. Vector map of pGL4.10[*luc2*].